



Technological University Dublin
ARROW@TU Dublin

Articles

DIT Biophotonics and Imaging

2017

Prediction of Viral Loads for Diagnosis of Hepatitis C Infection in Human Plasma Samples Using Raman Spectroscopy Coupled with Partial Least Squares Regression Analysis

Hugh J. Byrne

Technological University Dublin, hugh.byrne@tudublin.ie

Franck Bonnier

Technological University Dublin, Franck.Bonnier@tudublin.ie

H. Nawaz

N. Rashid

M. Saleem

Follow this and additional works at: <https://arrow.tudublin.ie/biophonart>

 Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Nawaz, H., Rashid, N., & Byrne, H. (2017). Prediction of viral loads for diagnosis of hepatitis C infection in human plasma samples using Raman spectroscopy coupled with Partial Least Squares Regression analysis. *Journal of Raman Spectroscopy*, vol. 48, no. 5, pp. 697–704. doi:10.1002/jrs.5108

This Article is brought to you for free and open access by the DIT Biophotonics and Imaging at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#)



Title:

Prediction of viral loads for diagnosis of Hepatitis C infection in human plasma samples using Raman spectroscopy coupled with Partial Least Squares Regression analysis.

Short Title:

Raman Spectroscopy for the diagnosis of the Hepatitis C.

Authors:

H. Nawaz^a, N. Rashid^b, M. Saleem^d, I. Amin^f, M. Iqbal^c, M. Rahman^c, Ola Ibrahim^d, S.M. Baig,^c M. Ahmed^d, F. Bonnier^g, H.J. Byrne^e

^a Department of Chemistry, University of Agriculture, Faisalabad, Pakistan.

^b Faisalabad Institute of Research Science and Technology (FIRST), Abbaspura, Faisalabad, Pakistan.

^c National Institute for Biotechnology and Genetic Engineering (NIBGE), P.O.Box 577, Jhang Road Faisalabad, Pakistan.

^d National Institute of Lasers and Optronics (NILOP), Islamabad, Pakistan.

^e FOCAS Research Institute, Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland.

^f PINUM Hospital, Faisalabad, Pakistan.

^g EA 6295 Nano-médicaments and nano-sondes, Université François-Rabelais (Tours, France).

*Corresponding Author: Dr. Haq Nawaz

E-mail: haqchemist@yahoo.com

Abstract:

Raman spectroscopy has been used to identify the biochemical changes associated with the presence of the Hepatitis C virus (HCV) in infected human blood plasma samples as compared to healthy samples, as control. The aim of the study was to establish the Raman spectral markers of hepatitis infection, which could be used for diagnostic purposes. Moreover, multivariate data analysis techniques, including Principal Component Analysis (PCA), coupled with Linear Discriminant Analysis (LDA), and Partial Least Square Regression (PLSR) are employed to further demonstrate the diagnostic capability of the technique. The PLSR model is developed to predict the viral loads of the HCV infected plasma on the basis of the biochemical changes caused by the viral infection.

Specific Raman spectral features are observed in the mean spectra of HCV plasma samples which are not observed in the control mean spectra. PCA differentiated the “normal” and “HCV” groups of the Raman spectra and PCA-LDA was employed to increase the efficiency of prediction of the presence of HCV infection, resulting in a sensitivity and specificity 98.8% and 98.6%, with corresponding Positive Predictive Value of 99.2%, and Negative Predictive Value of 98%. PLSR modelling was found to be 99 % accurate in predicting the actual viral loads of the HCV samples, as determined clinically using the Polymerase Chain Reaction (PCR) technique, on the basis of the Raman spectral changes caused by the virus during the process of the development of Hepatitis C.

Key words: Raman spectroscopy, Hepatitis C Virus (HCV) infection, blood plasma, Principal Components Analysis, Partial Least Squares Regression

Introduction:

Hepatitis C is a virally infectious disease, which affects 3% of the world and 4.8 % of the Pakistani population ^[1-3], highlighting the demand for rapid, routine, reliable and cost effective diagnostic techniques. Presently, screening methods depend upon the detection of antibodies and pathogens in Hepatitis C virus (HCV) infected blood ^[4]. The serology based method, Enzyme Linked Immunosorbent Assay (ELISA), targets specific antibodies arising in the blood from the response of the immune system due to HCV infection ^[5]. However, although ELISA is very commonly employed, it is not a very reliable method for diagnosis of HCV infection. False positive and false negative anti-HCV results pose severe difficulties in medical practice and in blood screening ^[6-8].

More direct virus detection based methods employ Polymerase Chain Reaction (PCR), in which HCV RNA, considered the most reliable disease marker, is detectable in serum or plasma as early as one week after infection and the technique is considered the gold standard for diagnosis of active HCV infection ^[9-10]. The accuracy of the molecular based detection method depends upon the quality of the chemical reagents used and the sensitivity of the instrumentation. It involves very complex processes, and therefore an expert operator is required to carry out the test, which can otherwise lead to significant errors. More recently, “real-time” PCR techniques have been developed ^[11-15]. A comparison of ELISA and PCR methods has been carried out ^[16-17], leading to the conclusion that PCR based techniques are the most reliable, although they are very expensive, time consuming, laborious and less commonly available. There is a need, therefore, to develop a new, reliable, efficient, rapid, automated, cost effective, non-invasive and biochemical label free diagnosis method to screen the whole community for effective disease management.

Over the last decade, Raman spectroscopy has shown great potential for the characterization of biological tissue ^[18] and the effect of external agents on the cell ^[19-24]. In addition, Raman spectroscopy can provide high content information, as it can examine spectral changes in any bio-macromolecules including DNA, proteins, lipids simultaneously ^[25-27] and can lead towards the identification of the Raman spectral features which could be utilized as diagnostic markers of the disease. Moreover, Raman micro-spectroscopy has been employed to explore sub-cellular biochemical structure, as demonstrated by its use in studies investigating the action of various agents on biological macromolecules as well as their interaction with cancer cells ^[20,28-37]. Recently, Raman spectroscopy has been successfully employed for the diagnosis and characterization of viral infections such as dengue ^[38], HPV ^[39-40] and others ^[41-42].

Although the diagnosis of HCV in the sera samples from virally infected patients by employing Raman spectroscopy has been reported ^[43], there is no work published to date in which the technique has been applied to study the correlation of the biochemical signatures of the virus with the viral loads. In the current work, Raman spectroscopy is employed to identify the biochemical changes associated with the HCV virus infection by studying infected human plasma samples, along with healthy samples as control, in order to establish the Raman spectral markers associated with the HCV infection. Furthermore, the multivariate techniques of Principal Components Analysis (PCA), coupled with Linear Discriminants Analysis (LDA), and Partial Least Squares Regression (PLSR) were employed for Raman spectral analysis. PCA was used to discriminate the Raman spectra from the diseased blood sera from the healthy ones, and coupled with LDA, to improve the diagnostic potential, whereas, PLSR models are developed based on the Raman spectra of diseased and healthy samples to predict the viral loads of the HCV infected plasma on the basis of the biochemical changes caused by the viral infection.

Materials and Methods

Sample preparation

Blood plasma samples from 10 healthy (control) and 10 confirmed Hepatitis C patients with different viral load (VL) values, given in the **Table 1**, were collected from PINUM Hospital, Faisalabad, Pakistan and approved for study by the ethical committee. These patients were diagnosed as infected with HCV by the Polymerase Chain Reaction (PCR) technique, currently a gold standard for the diagnosis of Hepatitis, which quantified the HCV loads in the patient samples (**Table 1**). A drop of each plasma sample (20 μ l - around 5 mm in diameter) was placed on an Aluminium slide at room temperature. The Raman spectra were acquired from samples ensuring that the blood plasma samples did not dry during the measurement time. Blood plasma samples were handled according to the US National Institute of Health (NIH) recommendations.

Raman spectral acquisition

Raman spectral acquisition from all of the 10 control and 10 HCV positive plasma samples was performed using a Raman spectrometer (Peak Seeker Pro-785; Agiltron, USA). The Peak Seeker Pro-785 utilizes a 785 nm diode laser as the excitation source, delivering a laser power of ~70 mW at the sample. The laser was delivered to the sample through a 10X. The CCD used for recording the Raman scattering was cooled by a thermoelectric cooler to minimize the electrical noise. The acquisition of the Raman spectra for all the samples was executed from 400 to 1800 cm^{-1} and 30-35 Raman spectra per sample were acquired with an acquisition time of 30 seconds each.

Data pre-processing

All data processing of the Raman spectra was performed using MatLab 7.2 and established protocols ^[44]. Data pre-processing included smoothing, baseline correction, vector normalization and substrate removal. All spectra, including substrate backgrounds, were vector normalized and smoothed using a Savitzky-Golay smoothing method (order 5, 13 point window). A rubber band correction for baseline removal for all the spectra was carried out and the substrate spectra were subtracted from each spectrum.

Data analysis

The Raman spectral data have been analysed by comparing the mean spectra of each patient, including control/healthy persons as well as HCV patients. The assignments of the Raman spectral features used in interpretation of the results were taken from the literature ^[24,35,45-46] and are described in **Table 2**.

In order to further elucidate the biochemical basis of the development of HCV as compared to the healthy/control samples, PCA was performed. PCA is a mathematical procedure involving the transformation of possibly correlated variables into a smaller number of uncorrelated variables, known as principal components (PC), basically to reduce the dimensionality of the data whilst maintaining their variability. The first principal component accounts for the dominant source of variability in the data, and each succeeding principal component accounts for the next highest source of the remaining variability. The loadings of the PC can be understood as the orthogonal dimensions of biochemical differences which facilitate separation of different groups of spectra of Raman data along their variability as each spectrum scores along these dimensions. To improve the diagnostic potential, PCA-LDA was then performed on the whole dataset to produce confusion matrices using the scores ^[47]. Specifically, the algorithm was performed using a stratified 10-fold cross validation, splitting the data into calibrations and test sets. For each virus load tested, 90% of the Raman spectra were used for calibration while the remaining 10% constituted the test sets. Ultimately, the confusion matrix built results from the 10 iterations and gives the overall classification from the 10 test sets. The classification which resulted in the maximum accuracy while keeping the number of latent variables to a minimum was chosen for all successive models. In the case of all HCV versus Control samples, the maximum accuracy was found when 3 principal components were used in the classification. Further addition of principal components only increased the complexity of the model without further improving the performance.

PLSR is a common chemometric method which aims to construct a generic model to describe the response variables (i.e. analyte concentration) in terms of the observed variables (spectra) from a set of training data (viral load). The least squares model is given by an equation system whose

dependent result is given by the measured variables (i.e. viral load in this case). The independent variables (the spectra) are regressed towards those dependent variables with the regression parameter, leaving residuals (differences between measured and predicted variables) which are suitable to access the quality of the predictive model and the parameters for model optimisation. PLSR is similar to PCA, which produces factors based on variance solely within the data matrix, whereas PLSR considers both target and the data matrix, ensuring that the factors correlate the data matrix (spectra) to the target variable (viral load). In PLSR, the variables are regressed simultaneously, maximizing the covariance between both matrices to allowing direct correlation between the spectra and target. PLSR models in this work were constructed using the SIMPLS algorithm with root mean squared error of cross validation (RMSE) as the fitness value. Leave-one-out cross validation was used to select the minimum number of latent variables (LV's comparable to PC's) to retain in order to create a robust and sufficiently ubiquitous model but still avoid over fitting.

Results and Discussion:

Mean Raman spectra of HCV samples

As spectral signatures of the HCV infection are likely to be correlated with viral load, the mean Raman spectrum of the patient sample with maximum viral load was initially compared to that of the Normal plasma samples. Mean Raman spectra of plasma samples of the Normal (healthy) and maximally HCV infected patients (V.L. = 1,718,359 IU/ml- positive-10 of **Table-1**) and their difference spectra, calculated by subtracting HCV spectra from normal spectra, are presented in **Figure 1** and **Figure 2**, respectively. The spectra are off-set for clarity. Notably, although plasma is derived from blood, it does not contain red blood cells (RBCs), white blood cells (WBCs) or platelets. It contains dissolved proteins (serum albumins, globulins, and

fibrinogen), glucose and clotting factors. The healthy human plasma spectra thus contain the signatures of constituent proteins, lipids, antigens, antibodies. On the other hand, plasma from the HCV patients will also contain virus particles as an infection and hence Raman spectral signatures of the protein part of the virus or viral RNA (HCV is RNA virus) may contribute differentiating spectral features. The Raman signatures of the external micro-organism (HCV in the current study) are expected only in the infected samples and not in the control/healthy samples. Close examination of **Figure 1** and **Figure 2** reveals Raman spectral bands which are significantly changed during the development of the disease from normal to HCV. The most significant changes are highlighted as solid lines and other ones with dotted lines.

Notably, among the Raman spectral features labelled with the solid lines, there are a few peaks, including those at 874, 880, 1054, 1065, 1072 and 1211 cm^{-1} , which are encircled to show their significance as these are not observed in the mean control spectra but are very significant in the HCV ones. Of these, the features at 874 and 880 cm^{-1} (tryptophan ring breathing), 1054 cm^{-1} (C-N stretching) and 1211 cm^{-1} (amide-III beta sheet or phenyl stretching in amino acids) are associated with proteins and may be associated with Raman spectral markers of the proteinic part of the HPV itself, or its physiological impact.

Raman spectral changes associated with proteins were observed. Among those peaks, it is important to mention that the peak ratio of the 748 cm^{-1} (CH_2 bending) and 758 cm^{-1} (Tryptophan ring breathing) features is changed in the HCV positive sample as compared to the Normal and the peak height of the peak at 748 cm^{-1} is increased in the HCV positive sample as compared to control. Moreover, a shoulder is observed at 606 cm^{-1} (CH_2 bending), near 620 cm^{-1} (C-C twisting). In addition, the peak ratio of the 1090 and 1099 cm^{-1} (C-N stretching) peaks is also changed and a decrease in the peak height is observed in the HCV positive sample as

compared to the Normal. These changes may be attributed to some kind of structural changes in the proteins which might be involved in the development of the HCV infection, and hence can be taken as Raman spectral markers of the HCV infection.

Other Raman spectral features observed, related to the proteins, consist of those at 422, 426 and 860 cm^{-1} (C-C stretching), 620 cm^{-1} (C-C twisting), 720, 758, 1360 and 1378 cm^{-1} (Tryptophan ring breathing), 838 cm^{-1} (Tyrosine ring breathing), 880, 898, 1150, 1197 cm^{-1} (C-C skeletal stretching), 934 cm^{-1} (C-C alpha helix of protein), 1090, 1099, 1157, and 1164 cm^{-1} (C-N stretching), 1222, 1228, 1242 and 1254 cm^{-1} (Amide-III beta sheet) and 1655 cm^{-1} (amide-I). These changes related to protein can be associated with the development of the disease.

Moreover, the peaks at 1065 and 1072 cm^{-1} , both associated with the fatty acids (lipids), may also be associated with the viral infection. Other lipid related changes observed include the features at 711 cm^{-1} (C-N stretching), 984 cm^{-1} (C-C head groups) and 1393 cm^{-1} (CH_3 deformation). The observed changes may be the result of an increase in the levels of the lipids as well as conformational changes in some lipidic structures, indicating their involvement in the process of the viral infection. It is reported that lipid droplets provide sites for the attachment of the HCV, necessary for replication in the infected host ^[48].

In the case of the study of cervical cancer tissue samples, in which the Human Papilloma virus (HPV) is considered a primary aetiological factor, it was observed with Raman spectroscopy that proteinic indicators of the presence of the virus are readily identifiable, although the DNA/RNA signatures of the virus itself are not so prominent ^[37,40,49]. It can hence be concluded that the effects of the virus are more evident than the virus itself.

Principal Component analysis (PCA)

In order to further explore the ability of the PCA to differentiate between one normal and HCV positive sample, PCA was applied to their Raman spectral data. **Figure 3** and **Figure 4** show the scatter plot and loadings of the PCA of Normal (one group) versus HCV plasma (one patient) of highest viral load (V.L. = 1,718,359 IU/ml, positive-17 of Table-1). There is clear differentiation of the two groups of Raman spectra observed by PC-1, explaining 87.7% of the variance in the data. The loadings of PC1, **Figure 4**, confirm the results discussed earlier in the case of the difference of the mean spectra, although they are inverted (**Figure 2**). Notably, the Raman spectra of the “normal plasma samples” are clustered in the negative side and HCV positive ones in the positive side of the PC1 axis of the PCA scatter plot and hence the spectral features associated with the HCV plasma are loaded positively, while those of the Normal plasma are loaded negatively^[50]. The features identified in **Figure 2**, also feature in the Loading of PC1, and are labelled with solid lines which are encircled to show their significance. Of these, those at 874 cm^{-1} and 880 cm^{-1} (tryptophan ring breathing) and 1054 cm^{-1} (C-N stretching) and 1211 cm^{-1} , 1222 cm^{-1} , 1228 cm^{-1} , 1242 cm^{-1} and 1254 cm^{-1} (amide-III beta sheet or phenyl stretching in amino acids) are feature prominently. PCA also confirms the significance of the lipidic derived features at 1065 cm^{-1} and 1072 cm^{-1} , 1157 , and 1164 cm^{-1} (C-N stretching) which can be associated with the manifestation of the viral infection as lipid droplets. Moreover, PCA also confirms the change in the peak ratio of the 748 cm^{-1} (CH_2 bending) and 758 cm^{-1} (Tryptophan ring breathing) features and that the peak height of the peak at 748 cm^{-1} is increased in the HCV positive sample as compared to control, observed as positive loadings, indicating its association with the HCV sample. Other positive features, indicating association with the HCV sample, include those at 426 cm^{-1} (C-C stretching) 711 cm^{-1} (C-N stretching), 720 cm^{-1} (Tryptophan ring breathing), 838

cm^{-1} (Tyrosine ring breathing), 880 cm^{-1} (C-C skeletal stretching), 1150 cm^{-1} (C-C skeletal stretching), 1360 and 1378 cm^{-1} (Tryptophan ring breathing) and 1393 cm^{-1} (CH_3 deformation). Moreover, a shoulder is observed at 606 cm^{-1} (CH_2 bending), near 620 cm^{-1} (C-C twisting) as a loading in the positive again associated with HCV sample. In addition, the peak ratio of the 1090 and 1099 cm^{-1} (C-N stretching) peaks is also changed and a decrease in the peak height is observed in the HCV positive sample as compared to the control indicated by their observation as negative loadings. These changes may be attributed to some kind of structural changes in the proteins which might be involved in the development of the HCV infection, and hence can be taken as Raman spectral markers of the HCV infection.

Other Raman spectral features observed as negative loadings (associated with normal sample), related to the proteins, including those at and 860 cm^{-1} (C-C stretching), 620 cm^{-1} (C-C twisting), 898 cm^{-1} (C-C skeletal stretching), 934 cm^{-1} (C-C alpha helix of protein), 1090 , 1099 , (C-N stretching), may also be associated with structural changes in proteins. Notably, the Raman spectral feature at 1655 cm^{-1} (amide-I) is observed at 1659 cm^{-1} in **Figure 4** (loading plot of PCA) which indicates a shift in this band and can be associated with the changes in the secondary structure of the proteins which may be involved in the development of the disease.

PCA-LDA

Having demonstrated the capability of Raman spectroscopy to identify spectral differences consistent with and indicative of the presence of HVC in blood plasma samples, the diagnostic potential to identify the presence of viral infection in patient samples over the range of viral loadings can be assessed using PCA- LDA. This approach becomes quite relevant when facing

large data sets scattered over multidimensional plots such as PCA scatter plot. While the graphical representation can be useful to visualise patterns in the data distribution, the specificity of the classification is not quantifiable. PCA-LDA allows taking into account the whole data set (783 spectra from 10 patients of varying viral load) and computing a confusion matrix according to a 10-fold stratified cross validation model. As a result, the accuracy of the classification between control and HCV samples can be estimated based on the variations in the spectral features highlighted by the PCA analysis. **Table 3** shows the confusion matrix built using the 3 first PC, which was found to be the optimal. It can be seen that, out of 293 spectra collected from control samples, only 4 are misclassified as HCV infected, while 6 of the 490 HCV spectra are misclassified as control. The overall sensitivity and specificity of the model are thus, respectively, 98.8% and 98.6%, resulting in a Positive Predictive Value of 99.2%, and Negative Predictive Value of 98% ^[51].

Partial Least Square Regression (PLSR) model for prediction of viral load

Raman micro spectroscopy can clearly fingerprint the biochemical changes associated with the development of Hepatitis C due to viral infection. PCA has been used to identify specific spectral changes which are correlated with the viral load values in the plasma of the HCV patients. PCA-LDA can be employed to improve the diagnostic capability of the technique. Here, it is demonstrated that PLSR modelling can be employed for the prediction of the viral load values based on the Raman spectroscopic response. Once established, such predictive models could be used to evaluate and monitor the viral loads of the patients before and after treatment of the disease.

PLSR models of Raman spectra acquired from the plasma of the control and HCV patients versus HCV loads were constructed to determine the ability of the spectral data to predict the level of viral infection. Spectra of all normal and HCV infected patients were compiled into a matrix, and were randomly selected for modelling. A total of 60% of the spectra were used to train the PLSR regression model and 40% of the total was retained as an independent test set to assess the performance of the model in predicting the viral loads, with unseen data. Leave-one out cross validation with the calibration set was used to determine the optimal model complexity for use in testing ^[52]. This process was performed with randomization of the data matrix and splitting of the data to prevent data bias ^[53]. Control of over fitting was achieved using a procedure previously described by Martens and Naes ^[54]. The procedure involves selection of the optimal number of latent variables (LV) to retain within the PLSR model via cross-validation (in this case 10 fold) with the calibration data set. The optimal number of LV's was then selected on the basis of the number which provided the lowest root mean squared error after cross validation.

Figure 5 represents the PLSR model development and the optimal number of LVs, which appears to be 20 (**Figure 5-A**), which is the most applicable prediction of the HCV viral loads (with an accuracy of 99.8 % as shown in **Figure 5-B**). In other words, the Root Mean Square Error of Prediction (RMSEP) is 0.2 %. Notably, the models were built with up to 20 latent variables which gave the minimum prediction error (predicted viral loads), as shown in **Figure 6** (every circle represents the prediction response vs. fit response of a training spectrum). However, the actual viral load of a sample can be predicted sufficiently, according to its spectral response, with an accuracy of 99.8 % with a model that has 20 latent variables. The R^2 value was determined to be 0.96. Low numbers of latent variables improve the likelihood that the model will be transferrable to unseen data and minimise the possibility of over fitting ^[54].

Conclusion

Although the analysis of the available data can only be understood as a preliminary proof of concept study, Raman spectroscopy in conjunction with multivariate data analysis has been shown to be a potentially powerful technique for the identification of the biochemical changes associated with the development of HCV infection in human blood plasma samples. Specific Raman spectral features are identifiable in HCV infected patients, which are absent in normal patients. Peaks at 874 and 880 (tryptophan ring breathing) and 1054 cm^{-1} (C-N stretching) and 1211 cm^{-1} (amide-III beta sheet or phenyl stretching in amino acids) are associated with proteins and may be considered as Raman spectral markers of the proteinic part of the HCV. Moreover, peaks at 1065, 1072 cm^{-1} , both associated with the fatty acids (lipids), may be associated with the viral infection. The combination of PCA and LDA improves the diagnostic power of analysis, and for the combined set of normal and variable viral load samples, the diagnostic sensitivity and specificity is determined to be 98.8% and 98.6%, with corresponding Positive Predictive Value of 99.2%, and Negative Predictive Value of 98%. The PLSR model successfully predicts the viral loads of the HCV infected plasma on the basis of the biochemical changes caused by the viral infection with 99.8 % accuracy. These models can potentially be employed for monitoring the HCV loads in the samples before and after the treatment of the disease, and their robustness can be further enhanced by extending the patient sample base.

Research Grant: This work has been supported by the Research grant of worth 10,000 US \$ (Grant number: O/Ref: I.C.P.S.R/ 3.8.2.1.1/11, by Islamic Educational, Scientific and Cultural Organisation (ISESCO), Morocco awarded to the Corresponding author for this Project for the

period May, 2015-May, 2017. The corresponding author (PI of this project/research grant) is very much thankful to the donor funding agency and highly acknowledge here in this manuscript.

Acknowledgment

The authors are very thankful to Mr.Imran Ameen from PINUM hospital (PAEC), Faisalabad for providing the HCV plasma samples for this study.

References

- [1] N. Ford, C. Kirby, K. Singh, E. J. Mills, G. Cooke, A. Kamarulzaman, P. duCros, *Bull. W.H.O.* **2012**; 90, 540.
- [2] D. Lavanchy, *Liver Int.* **2009**; 29, 74.
- [3] C. W. Shepard, L. Finelli, M. J. Alter, *Lancet Infect Dis.* **2005**; 5, 558.
- [4] R. Kesli, H. Polat, Y. Terzi, M. G. Kurtoglu, Y. Uyar, *J. Clin. Microbiol.* **2011**; 49, 4089.
- [5] C. Cao, C. Shi, P. Li, Y. Tong, Q. Ma, *Clin. Diagn. Virol.* **1996**; 6, 137.
- [6] D. R. Gretch, *Hepatology.* **1997**; 26.
- [7] P. Martin, F. Fabrizi, V. Dixit, S. Quan, M. Brezina, E. Kaufman, K. Sra, R. DiNello, A. Polito, G. Gitnick, *J. Clin. Microbiol.* **1998**; 36, 387.
- [8] E. R. Schiff, M. de Medina, R. S. Kahn, *Semin. Liver Dis.* 1998; Vol. 19, p 3.
- [9] J.-M. Pawlotsky, *Clin Liver Dis.* **2003**; 7, 45.
- [10] L. Raeymaekers, *Anal. Biochem.* **1993**; 214, 582.
- [11] E. Farma, E. Boeri, P. Bettini, C. Repetto, J. McDermott, F. Lillo, O. Varnier, *J. Clin. Microbiol.* **1996**; 34, 3171.
- [12] C. A. Heid, J. Stevens, K. J. Livak, P. M. Williams, *Genome Res.* **1996**; 6, 986.
- [13] R. Higuchi, C. Fockler, G. Dollinger, R. Watson, *Bio Tech.* **1993**; 11, 1026.
- [14] M. Martell, J. Gómez, J. I. Esteban, S. Sauleda, J. Quer, B. Cabot, R. Esteban, J. Guardia, *J. Clin. Microbiol.* **1999**; 37, 327.
- [15] T. Morris, B. Robertson, M. Gallagher, *J. Clin. Microbiol.* **1996**; 34, 2933.
- [16] M. Enomoto, S. Nishiguchi, S. Shiomi, M. Tanaka, K. Fukuda, T. Ueda, A. Tamori, D. Habu, T. Takeda, Y. Yano, *J. Gastroenterol. Hepatol.* **2001**; 16, 904.
- [17] M. Swellam, M. S. Mahmoud, A. A. F. Ali, *IUBMB life.* **2011**; 63, 430.
- [18] F. M. Lyng, E. Ó. Faoláin, J. Conroy, A. Meade, P. Knief, B. Duffy, M. Hunter, J. Byrne, P. Kelehan, H. J. Byrne, *Exp. Mol. Pathol.* **2007**; 82, 121.
- [19] J. Chan, S. Fore, S. Wachsmann-Hogiu, T. Huser, *Laser Photon Rev.* **2008**; 2, 325.
- [20] Z. Farhane, F. Bonnier, A. Casey, H. Byrne, *Analyst.* **2015**; 140, 4212.
- [21] T. Huser, C. A. Orme, C. W. Hollars, M. H. Corzett, R. Balhorn, *J. of Biophotonics.* **2009**; 2, 322.
- [22] P. Knief, C. Clarke, E. Herzog, M. Davoren, F. M. Lyng, A. D. Meade, H. J. Byrne, *Analyst.* **2009**; 134, 1182.
- [23] A. D. Meade, C. Clarke, F. Draux, G. D. Sockalingum, M. Manfait, F. M. Lyng, H. J. Byrne, *Anal. Bioanal. Chem.* **2010**; 396, 1781.
- [24] A. D. Meade, F. M. Lyng, P. Knief, H. J. Byrne, *Anal. Bioanal. Chem.* **2007**; 387, 1717.
- [25] H. Nawaz, F. Bonnier, P. Knief, O. Howe, F. M. Lyng, A. D. Meade, H. J. Byrne, *Analyst.* **2010**; 135, 3070.
- [26] H. Nawaz, F. Bonnier, A. D. Meade, F. M. Lyng, H. J. Byrne, *Analyst.* **2011**; 136, 2450.
- [27] H. Nawaz, A. Garcia, A. D. Meade, F. M. Lyng, H. J. Byrne, *Analyst.* **2013**; 138, 6177.
- [28] J. Dorney, F. Bonnier, A. Garcia, A. Casey, G. Chambers, H. J. Byrne, *Analyst.* **2012**; 137, 1111.
- [29] F. Draux, P. Jeannesson, A. Beljebbar, A. Tfayli, N. Fourre, M. Manfait, J. Sulé-Suso, G. D. Sockalingum, *Analyst.* **2009**; 134, 542.
- [30] E. Efeoglu, M. Keating, J. McIntyre, A. Casey, H. J. Byrne, *Anal. Methods.* **2015**; 7, 10000.

- [31] Z. Farhane, F. Bonnier, M. A. Maher, J. Bryant, A. Casey, H. J. Byrne, *J. of Biophotonics*. **2016**.
- [32] M. E. Keating, F. Bonnier, H. J. Byrne, *Analyst*. **2012**; *137*, 5792.
- [33] M. E. Keating, H. Nawaz, F. Bonnier, H. J. Byrne, *Analyst*. **2015**; *140*, 2482.
- [34] J. Ling, S. D. Weitman, M. A. Miller, R. V. Moore, A. C. Bovik, *Appl. Opt.* **2002**; *41*, 6006.
- [35] I. Notingher, *Sensors*. **2007**; *7*, 1343.
- [36] C. A. Owen, J. Selvakumaran, I. Notingher, G. Jell, L. L. Hench, M. M. Stevens, *J. Cell. Biochem.* **2006**; *99*, 178.
- [37] N. Rashid, H. Nawaz, K. W. Poon, F. Bonnier, S. Bakhiet, C. Martin, J. J. O'Leary, H. J. Byrne, F. M. Lyng, *Exp. Mol. Pathol.* **2014**; *97*, 554.
- [38] M. Saleem, M. Bilal, S. Anwar, A. Rehman, M. Ahmed, *Laser Phys. Lett.* **2013**; *10*, 035602.
- [39] K. M. Ostrowska, A. Garcia, A. D. Meade, A. Malkin, I. Okewumi, J. J. O'Leary, C. Martin, H. J. Byrne, F. M. Lyng, *Analyst*. **2011**; *136*, 1365.
- [40] K. M. Ostrowska, A. Malkin, A. Meade, J. O'Leary, C. Martin, C. Spillane, H. J. Byrne, F. M. Lyng, *Analyst*. **2010**; *135*, 3087.
- [41] A. Sakudo, Y. Sukanuma, T. Kobayashi, T. Onodera, K. Ikuta, *Biochem. Biophys. Res. Commun.* **2006**; *341*, 279.
- [42] A. Sakudo, R. Tsenkova, T. Onozuka, K. Morita, S. Li, J. Warachit, Y. Iwabu, G. Li, T. Onodera, K. Ikuta, *Microbiol. Immunol.* **2005**; *49*, 695.
- [43] J. Saade, M. T. T. Pacheco, M. R. Rodrigues, *J. of Spectroscopy*. **2008**; *22*, 387.
- [44] P. Knief, *BioPhysic, Dublin Institute of Technology, Dublin*. **2010**.
- [45] J. De Gelder, K. De Gussem, P. Vandenabeele, L. Moens, *J. Raman Spectroscopy*. **2007**; *38*, 1133.
- [46] P. Jess, V. Garcés-Chávez, D. Smith, M. Mazilu, L. Paterson, A. Riches, C. Herrington, W. Sibbett, K. Dholakia, *Opt. Express*. **2006**; *14*, 5779.
- [47] B. A. Gutman, X. Hua, P. Rajagopalan, Y.-Y. Chou, Y. Wang, I. Yanovsky, A. W. Toga, C. R. Jack, M. W. Weiner, P. M. Thompson, *NeuroImage*. **2013**; *70*, 386.
- [48] J. McLauchlan, *Biochim. biophys. acta, mol. cell. biol. lipids*. **2009**; *1791*, 552.
- [49] M. Diem, M. Miljković, B. Bird, T. Chernenko, J. Schubert, E. Marcsisin, A. Mazur, E. Kingston, E. Zuser, K. Papamarkakis, *J. of Spectroscopy*. **2012**; *27*, 463.
- [50] F. Bonnier, H. Byrne, *Analyst*. **2012**; *137*, 322.
- [51] D. G. Altman, P. Royston, *Stat. Med.* **2000**; *19*, 453.
- [52] A. Meade, C. Clarke, H. Byrne, F. Lyng, *Radiat. Res.* **2010**; *173*, 225.
- [53] K. Varmuza, P. Filzmoser, *Introduction to multivariate statistical analysis in chemometrics*; CRC press, **2009**.
- [54] H. Martens, T. Naes, *Multivariate calibration*; John Wiley & Sons, **1992**.

Sr.No	Sample name	Viral load value (IU/ml)
1	Positive-1	67,070
2	Positive-2	125,236
3	Positive-3	166,260
4	Positive-4	185,297
5	Positive-5	390,058
6	Positive-6	442,365
7	Positive-7	846,668
8	Positive-8	873,417
9	Positive-9	907,874
10	Positive-10	1,718,359

Table 1: Viral load values of the HCV positive human blood plasma samples used for the Raman spectral acquisition.

Wave number (cm ⁻¹)	Raman Assignments
422	C-C stretching
426	C-C stretching
606	CH ₂ bending
620	C-C twisting
711	C-N stretching
720	Tryptophan ring breathing
748	CH ₂ rocking
758	Tryptophan ring breathing
838	Tyrosine ring breathing
851	Tyrosine ring breathing
860	C-C stretching
874	C-C-N symmetrical stretching
880	C-C skeletal stretching
898	C-C skeletal stretching
934	C-C skeletal stretching; alpha helix
984	C-C head groups
1054	C-N stretching
1065	fatty acids (lipids)

1072	fatty acids (lipids)
1090	C-N stretching
1099	C-N stretching
1150	C-C skeletal stretching
1157	C-N stretching
1164	C-N stretching
1197	C-C stretching
1211	amide-III beta sheet or phenyl stretching in amino acids
1222	Amide-III beta sheet
1228	Amide-III beta sheet
1242	Amide-III beta sheet
1254	Amide-III beta sheet
1360	Tryptophan ring breathing
1378	Tryptophan ring breathing
1393	CH ₃ deformation
1655-1680	Amide I: α helix

Table 2: Raman peak assignments for proteins, DNA, RNA and lipids.

		Predicted Class	
		Control	Positive
Actual Class	Control	289	4
	Positive	2	276

Table 3: Confusion matrix of PCA-LDA of control (10 samples) versus HCV (10 samples) infected patients (Table 1).







