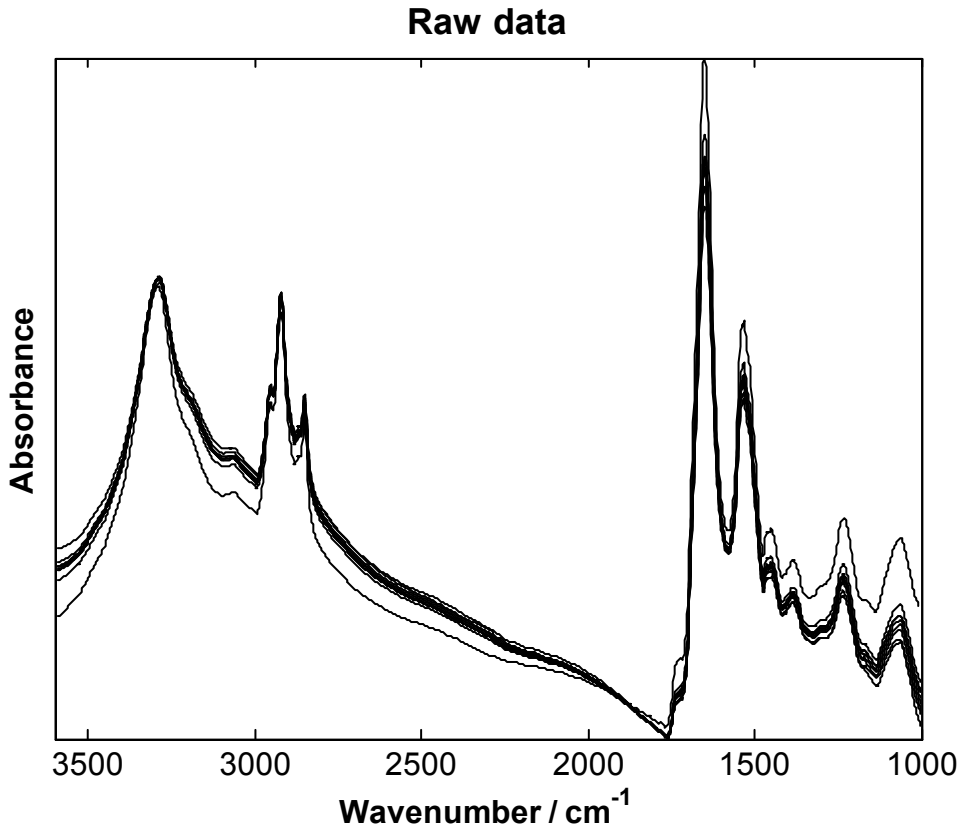
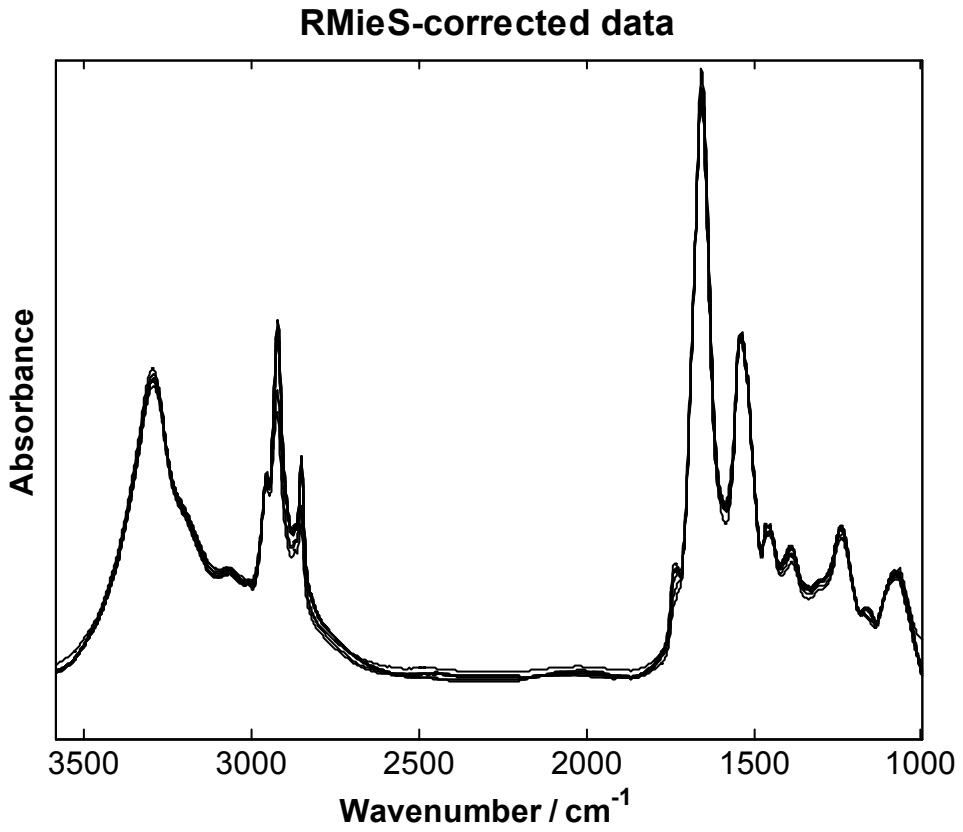
Electronic Supplementary Material (ESI) for Analyst.

Characterising cytotoxic agent action as a function of the cell cycle using Fourier Transform Infrared Mircospectroscopy

Supplementary Information

The following figures show the raw and RMieS-corrected FTIR dataset of the mean spectrum per cell cycle phase in the control and drug-treated samples.





Description regarding the optimization and validation of the models described in this paper.

Support Vector Machines

The Support Vector Machines were optimized, constructed and validated via the utilization of the LibSVM(*1*) toolbox. The spectral dataset, (765 individual spectra) encompassing data from both control and drug-treated cells, was labelled in order to identify both drug-treatment/control and cell cycle phase dependency. The newly created set of data was inputted into the toolbox in order to establish the parameters that would therefore yield a robust SVM with the maximum predictive power. It is important to highlight that, when initially installed, this toolbox uses 3-fold cross validation as a default mode of optimization; thus, the results presented originally in this study were retrieved by using this cross-validation method. Based upon the referee's suggestion for performing a more exhaustive validation approach, the data was subjected to a 10-fold cross validation and the results are shown in Table 1. Optimizing, constructing and validating a 10-fold cross validated support vector machine was, on average, 4.7 times more time consuming; yet, the results improved by 6.7% when the cell cycle information was taken into consideration for the purposes of classification. Such improvement in the prediction accuracy may be due to the fact that the SVM employs a wider range of parameters in the optimization process, which inherently takes longer, to find out which are the best "variables" that would create the hyperplanes of classification.

Table 1. K-fold Cross validation comparison for the optimized SVM's.

	Class	**3-fold cross validation**	**10-fold validation**
		Classification accuracy,%	
Control Vs Drug Treatments	Control	98	98
	5FU	89	89
	Paclitaxel	90	91
	Average accuracy	**92.3**	**92.7**
	Time, s	252.3	858.5
Cell cycle phases from both Control and Drug-treated cells	G_0-G_1	73	80
	S	79	82
	G_2-M	81	89
	Average accuracy	**77**	**83.7**
	Time, s	249.35	1493.43

1. Chang, C.-C., and Lin, C.-J. (2011) LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol. 2*, 1-27.