analytical chemistry

ARTICLE

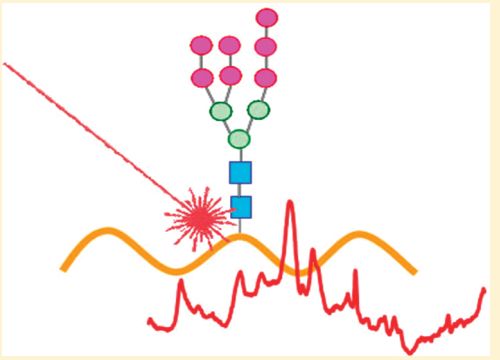
pubs.acs.org/ac

Monitoring the Glycosylation Status of Proteins Using Raman Spectroscopy

Victoria L. Brewster, Lorna Ashton, and Royston Goodacre*

School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, M1 7DN, U.K.

ABSTRACT: Protein-based biopharmaceuticals are becoming increasingly widely used as therapeutic agents, and the characterization of these biopharmaceuticals poses a significant analytical challenge. In particular, monitoring posttranslational modifications (PTMs), such as glycosylation, is an important aspect of this characterization because these glycans can strongly affect the stability, immunogenicity, and pharmacokinetics of these biotherapeutic drugs. Raman spectroscopy is a powerful tool, with many emerging applications in the bioprocessing arena. Although the technique has a relatively rich history in protein science, only recently has Raman spectroscopy been investigated for assessing posttranslational modifications, including phosphorylation, acetylation, trimethylation, and ubiquitination. In this investigation, we develop for the first time Raman spectroscopy combined with multivariate data analyses, including principal components analysis and partial least-squares regression, for the determination of the glycosylation status of proteins and quantifying the relative concentrations of the native ribonuclease (RNase) A protein and RNase B glycoprotein within mixtures.



Protein-based biopharmaceuticals are becoming increasingly popular therapeutic agents, and the production of these drugs is rapidly growing within the pharmaceutical industry. There are currently over 50 recombinant proteins approved for use and over 500 more under development.[1] Recent relaxation in the legislation governing stem cell technologies[2,3] is expected to increase further developments in this field.

Over one-third of these biopharmaceuticals are glycoproteins,[4] in which short chains of carbohydrates are covalently linked to the peptide chain of a protein following translation. Glycosylation is the most common posttranslational modification (PTM), occurring in up to one-half of all gene products.[4] The glycosylation status of a protein therapeutic is of great importance because it can affect the stability; pharmacokinetics; and, perhaps most importantly, immunogenicity of the drug.[5] Consequently, with biopharmaceuticals, it is necessary to determine not only whether a protein is glycosylated, but also that the correct glycan (series of polysaccharides or oligosaccharides) has been attached completely and linked to the correct amino acid.

Raman spectroscopy is a promising technique for the characterization and quantification of therapeutic proteins. The technique has been long used for therapeutic protein analysis,[6] both in solid state[7] and in solution.[8] Many studies have investigated structural differences in protein molecules by taking advantage of the sensitivity of the amide I region to conformational changes.[9−11] Raman spectroscopy has also been successful in monitoring therapeutic protein, aggregation,[12] lyophilization,[7] and side chain confirmation.[13] Furthermore, Raman spectroscopy is ideal for the online monitoring of bioprocesses because it is nondestructive, inexpensive, rapid, and quantitative. In addition, its confocal nature makes it possible to focus through transparent vessels, allowing direct analysis from fermentation broths.[14] Previous studies have successfully applied Raman spectroscopy, with appropriate multivariate analysis, to the online monitoring of glucose, lactate, ammonium, and ethanol levels.[15−17] Raman spectroscopy has been proved useful in many other areas of biotechnology, including microorganism identification;[18,19] bioprocesses monitoring;[20,15] and analysis of smaller biological molecules, such as antibiotics.[21]

Raman spectroscopy has been used to characterize and quantify various carbohydrates[22−26] as well as to provide structural information about glycoproteins, in particular, glycoprotein-C of the herpes simplex virus[27] and α_1-acid glycoprotein from blood plasma.[28] Raman, surface enhanced Raman scattering (SERS), and Raman optical activity (ROA) have been used to study interactions between pharmaceutical molecules, in particular, anticancer drugs,[29] and also to monitor the binding of the glycoproteins found in antifreeze.[30] However, despite the fact that Raman spectroscopy has an extensive history in protein and glycoprotein analysis, it is relatively underutilized in the monitoring of PTMs, particularly glycosylation. The few PTMs previously studied by Raman spectroscopy include phosphorylation,[31] acetylation, trimethylation, and ubiquitination.[32] Investigations of phosphorylation status of α-casein demonstrated how the ratio of phosphorylated and nonphosphorylated states could be determined by a combination of Raman spectroscopy and multivariate analysis.

The aim of this study was to develop Raman spectroscopy as a rapid approach to characterize the glycosylation status of proteins.

Received: May 11, 2011
Accepted: June 24, 2011
Published: June 24, 2011

Bovine pancreatic ribonuclease proteins, RNaseA and B, were chosen for the initial investigation because they provide a simple model system. Although both proteins have identical amino acid sequences and reported secondary and tertiary structure, only RNase B is glycosylated.[33] Although RNase B is glycosylated at only one site, unlike the majority of glycoproteins, which have numerous and more complex arrangements of glycans, the availability of a native nonglycosylated form makes it an ideal glycoprotein for initial investigation. By directly comparing RNase A and B spectra as well as further investigations with appropriate glycan standards and chemical and enzymatic deglycosylation of RNase B, we have been able to demonstrate the potential of Raman spectroscopy for characterizing the glycosylation status of this protein. Furthermore, through the implementation of chemometric approaches, we have identified the most selective vibrational modes for the quantification of glycosylation in this system. Although the chemometric models discussed here are specific to this protein, the identification of the most informative vibrational modes is applicable to all Raman glycoprotein investigations. By further developing these methods and building up a knowledge base of glycan standards, it should be possible to adapt these chemometric models to facilitate the use of Raman spectroscopy for characterization of glycosylation status in far more complex glycoproteins.

MATERIALS AND METHODS

Materials. Ribonucleases A and B, monosaccharides, PNGase F enzymes, trifluoromethanesulfonic acid, and other deglycosylation reagents were obtained from Sigma-Aldrich (Dorset, U.K.). MALDI matrixes and calibration standards were also obtained from Sigma-Aldrich. SpectraRIM slides were purchased from Tienta Sciences Inc. (Tienta Sciences Inc., Indianapolis, IN, USA).

Deglycosylation Methods. Deglycosylation methods and protein recovery protocols have been adapted from those found in the literature.[34−37]

Chemical Deglycosylation Method. Precooled trifluoromethansulfonic acid (TFMS) and anisole were mixed to form a solution of 10% anisole in TFMS (15 μL anisole in 140 μL TFMS). A 150 μL portion of anisole/TFMS solution was then added to 1 mg of precooled lyophilized RNase B in a reaction vial and shaken until all the protein was dissolved. The sample reaction vial was incubated on ice for 3 h, with occasional shaking, and 4 μL of 0.4% bromophenol blue was then added as an indicator dye; the color of the solution turned deep red. The sample reaction vial and a 60% pyridine solution were then cooled to −15 °C in a methanol dry ice bath. The cooled pyridine solution was then added dropwise to the sample, with mixing and cooling between drops, until the color changed to yellow and then to blue. A total of ∼300 μL of pyridine solution was added. A 10-fold excess of diethyl ether with 10% hexane was then added to the reaction mixture, mixed, and left to stand at −80 °C for 1 h. The sample vial was centrifuged at 8765*g* for 5 min, and the supernatant containing the pyridinium salts was removed. The deglycosylated protein was then recovered by precipitation with ethanol, and 500 μL of ethanol was added to the reaction vial. The mixture was mixed and stored at −20 °C for 1 h and centrifuged at 10 956*g* for 15 min, and the supernatant was removed. The resulting protein pellet was resuspended in water. Control samples were created by subjecting RNase B samples to the same conditions as the deglycosylated protein with the exception of adding the deglycosylation agent.

Enzymatic Deglycosylation Method. RNase B was prepared as a 1 mg/mL solution using 20 mM ammonium bicarbonate reaction buffer. A 90 μL portion of glycoprotein solution was then added to a reaction vial along with 5 μL of denaturant solution (2% octyl β-D-glucopyranoside with 100 mM 2-mercaptoethanol), and the vial was placed in a heating block at 100 °C for 10 min. The vial was allowed to cool to room temperature, and the 5 μL of reaction buffer was added.The vial was spun briefly, at 503*g* for 15 s, in a microcentrifuge. Ten microliters of PNGase F enzymes (500 unit/mL) was then added, mixed, spun, and incubated at 37 °C for 24 h. The reaction was stopped by heating to 100 °C for 10 min. PNGase F enzymes were removed by precipitation according to methods described in the literature.[37]Deglycosylated RNase B was recovered using the ethanol precipitation method described in the Chemical Deglycosylation Method section. RNase B control samples were created, subjecting the protein to identical conditions, minus the addition of PNGase F enzymes.

Raman Spectroscopy. All Raman data in this study were collected on Tienta Spectra RIM slides. These slides have a hydrophobic coating, which causes the liquid samples to "bead up" on the surface, increasing the concentration of protein in the spot. This can increase the Raman intensity by as much as 4 times[38] and can also help to reduce fluorescence emission.[38] As a data quality control measure, the consistency of data acquired from these slides was tested. Using principal components analysis (PCA), spectra taken from protein and glycoprotein powders were compared to those taken from spots on Spectra RIM slides. In addition, spectra were taken for six different positions within each spot to assess the variation within a protein spot. Figure 1 summarizes these results, showing that there was very little variation among the spectra, strongly suggesting that the spectral output is independent of the sampling position within a protein spot.

Raman data were collected using a Renishaw 2000 Raman microscope (Renishaw Plc., Old Town, Wotton-under-Edge, Gloucestershire, U.K.) with a low-power (27 mW), near-infrared, 785 nm diode laser with power at the sampling point between 2 and 4 mW and a spectral resolution 6 cm^{-1}. The instrument was wavelength-calibrated with a silicon wafer focused under the 50× objective and collected as a static spectrum centered at 521 cm^{-1} for 1 s. The GRAMS WiRE software package (Galactic Industries Corp., 395 Main St., Salem, NH) running under Windows 95 was used for instrument control and data capture. Experimental parameters varied among samples: all were single accumulation, extended scans between 200 and 2000 cm^{-1}, and exposure time varied between 30 and 120 s. For glycosylation status studies, 2 μL aliquots of solution were spotted onto a hydrophobic SpectraRIM slide and allowed to dry out at room temperature for approximately 1 h.

Mass Spectrometry. MALDI-MS was performed on an Axima CFR*plus* MALDI-TOF mass spectrometer (Shimadzu Biotech, Manchester, UK), equipped with a nitrogen pulsed UV laser (337 nm), in the positive ion mode. The instrument was calibrated before each use using apomyoglobin, aldolase, and albumin as calibration standards. Intact protein samples were analyzed in linear time-of-flight (TOF) mode, whereas protein digests were analyzed in reflectron TOF mode. A total of 10 shots were recorded per profile, and 1000 profiles were averaged per sample. Data were collected over a mass-to-charge (m/z) range of 5000−20000 with a typical laser power of 125 mW for proteins and 1−3000 m/z, laser power 75 mW, for protein digests. Protein digests were performed using the enzyme trypsin. Fifteen microliters of digestion buffer (50 mM ammonium bicarbonate) and 1.5 μL of

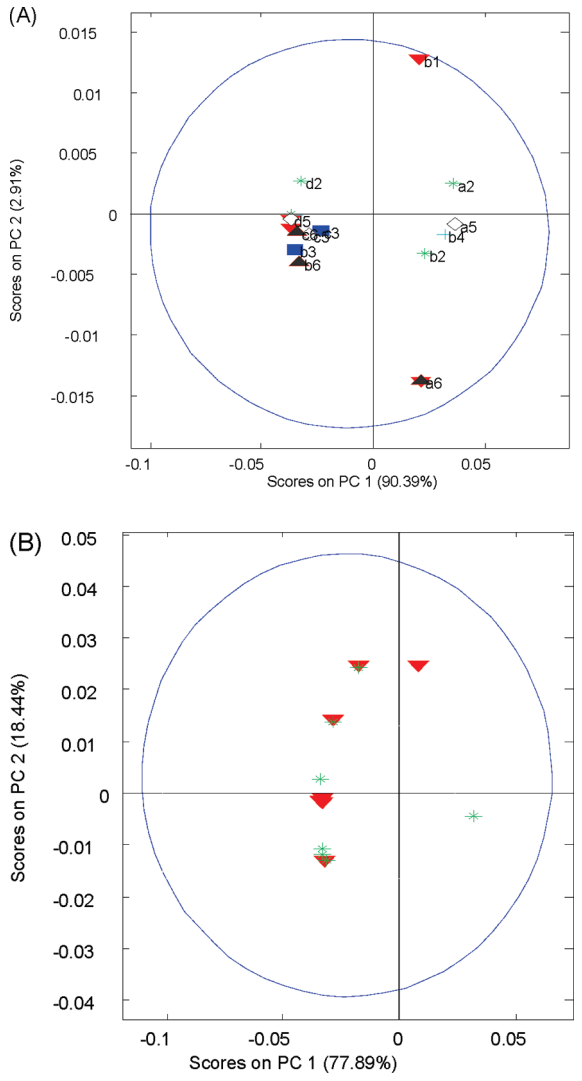


Figure 1. (A) PCA scores plot (PC1 vs PC2) showing the variation in the Raman spectra of the RNase B, recorded from different positions on a protein spot on a Spectra RIM slide. Each symbol encodes for a different location of the protein spot. The blue line shows 95% significance level. (B) PCA scores plot (PC1 vs PC2) showing random distribution of RNase A spectra recorded from both powder (red triangles) and Spectra RIM slides (green asterix).

reducing buffer (100 mM DDT) were added to 10 μL of protein and incubated at 95 °C for 5 min. Three microliters of alkylation buffer (100 mM idoaceamide) was then added and incubated at room temperature for 20 min. One μL of trypsin (0.1 μL/μL) was then added, and the mixture was incubated at 37 °C for 3 h. Finally, an additional 1 μL of enzyme was added, and the sample was incubated at 30 °C overnight.[39] A 1 μL portion of sample was spotted onto a MALDI target plate and allowed to dry at room temperature; 1 μL of 10 mg/mL matrix (sinapinic acid for proteins and α-cyano-4-hydroxycinnamic acid for tryptic digests) was then spotted on top of each sample, and the sample was dried at room temperature prior to analysis.

Data Analysis. Raman spectroscopic data were exported from the instrument software into Matlab 7.6 (The MathWorks, Inc., Natick, MA., USA), where data preprocessing was performed, to allow direct comparison of these data. Initial preprocessing (Savitzy-Golay smoothing and weighted least-squares baseline

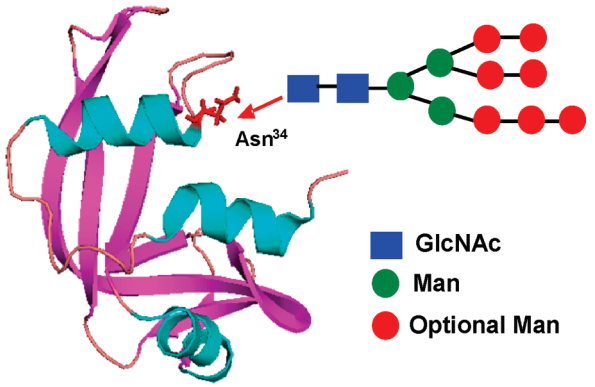


Figure 2. Cartoon representation of the native state of bovine RNase drawn from atomic coordinates in the PDB (5RSA) using PyMOL showing the Asn[34] residue and the RNase B glycan. Optional mannose (Man, indicated by red circle) refers to the variation in number and possible arrangements of mannose between the different glycoforms that occur in RNase B.

corrections) was applied using eigenvector PLS Toolbox 4.0 (Eigenvector Research, Inc., Wenatchee, WA, USA). Further preprocessing (scaling and normalization) and multivariate analysis was performed using PyChem 3.0.5 software.[40]

Principal components analysis (PCA)[41] was used for exploratory cluster analysis. PCA is a well-known technique for reducing the dimensionality of multivariate data while preserving most of the variance, and PyChem was employed to perform PCA using the NIPALS algorithm.[42] 2D correlation moving windows calculations[43,44] were performed using 2D Shige freeware (http://science.kwansei.ac.jp/~ozaki/index-e.html) to assess in a multivariate manner any changes in the Raman spectra with the level of RNase B in mixtures with RNase A.

Partial least-squares regression (PLSR)[45] was used to correlate the Raman data with the level of RNase B in mixtures with RNase A, and predictions of these levels were generated using training, cross validation, and a hold-out test set. The data were first smoothed (Savitzy-Golay using a filter width of 9), baseline-corrected (asymmetric least-squares), row-scaled (min 0, max +1), and normalized (extended multiplicative scatter correction using a polynomial of order 10). This order of preprocessing was determined after comparing the accuracy of PLSR plots at each stage of the preprocessing to maintain reliable and constant results. In all PLSR, the model was calibrated using the five replicates samples from 0, 10, 20, ..., 90, and 100% RNase B in RNase A and the cross validation data contained mixtures of RNase B in RNase A at levels of 5, 15, ..., 85, and 95%. The test data used the same series as in the cross validation data. In the cross validation data, the first two replicates were used at each concentration, and for the test set, the remaining three replicates were used.

RESULTS AND DISCUSSION

Glycosylation Status. The ability of Raman spectroscopy to distinguish between glycosylated and deglycosylated proteins was tested using bovine pancreatic RNase proteins. RNase exists in many forms, and of particular relevance to studying glycosylation status are two forms of the same RNase: the native nonglycosylated RNase A (EC 3.1.27.5) and glycosylated RNase B. Both RNase A and B are globular proteins with identical amino

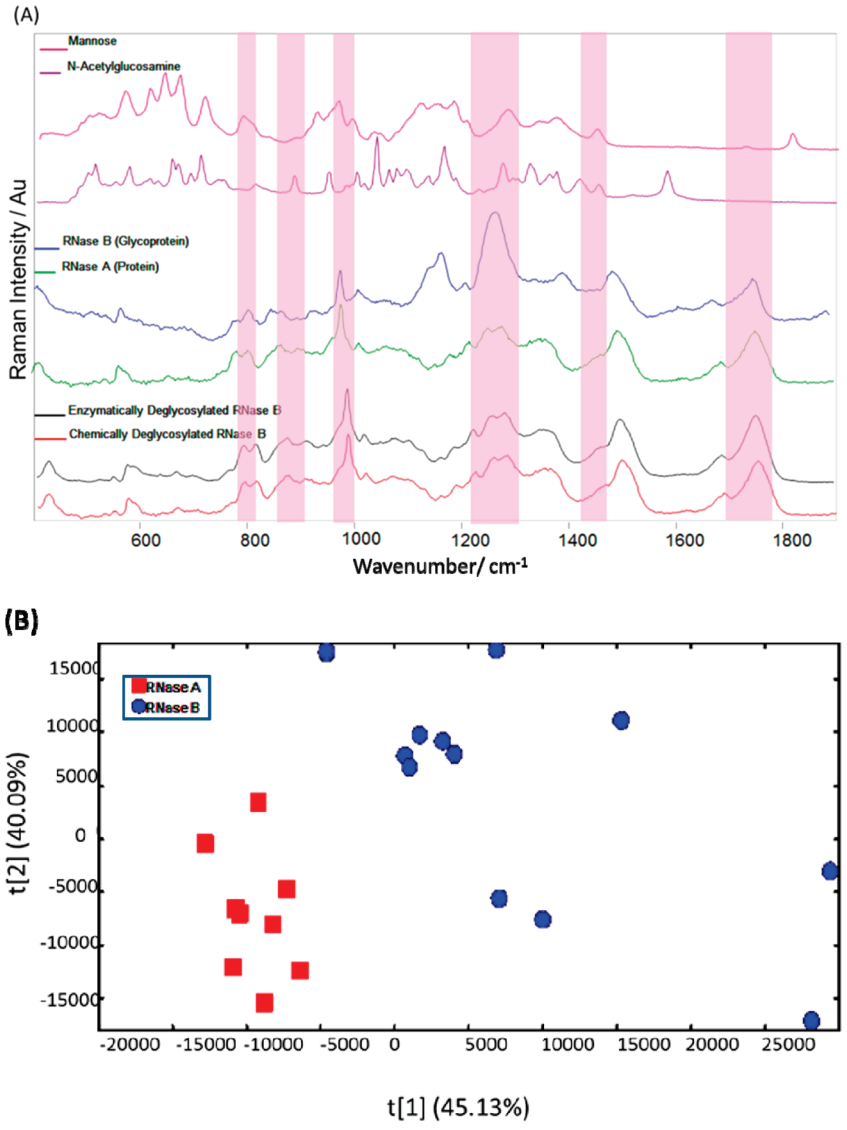


Figure 3. (A) Average Raman spectra of mannose, GlcNAc, RNase A and B, and chemically and enzymatically deglycosylated RNase B, obtained at 785 nm. Areas highlighted are regions of significance, as indicated by the PLS loading in Figure 5B. Band assignments for these regions are given in Table 1. (B) PCA Scores plot (PC1 vs PC2) of RNase data, showing RNase A and B spectra resolved into separate clusters.

acid sequences and secondary structure and with near-identical tertiary structures. RNase B contains one single N-linked glycan at asparagine 34 (Asn34). The composition of the RNase glycan varies; however, it will always contain two *N*-acetylglucosamine (GlcNAc) residues and between three and nine mannose residues (Figure 2).[33]

Solutions of RNase A and RNase B were prepared in deionised water (1 mg/mL) and spotted onto Tienta Spectra RIM slides. Raman spectra were recorded at 785 nm from six different positions on each protein spot. Figure 3a shows the resultant Raman spectra of RNase A and RNase B. The spectra show one of the main differences to be in the amide III region centered around ~1254 cm^{-1}, where a broad doublet appears in RNase A, and a more intense single peak, in RNase B. The loss of the amide III band at ~1246 cm^{-1}, previously associated with less-ordered protein structure,[46,47] could be assigned to the disordered loops of the RNase protein in proximity to Asn34 (Figure 2), which are stabilized by the addition of a carbohydrate moiety. Alternatively, this difference can be attributed to bands from the sugar molecules that mask the protein signal. Both the sugars in the glycan of RNase B exhibit bands in this region: NH_2 twisting in GlcNAc at ~1253 cm^{-1} and ring stretching in mannose at ~1259 cm^{-1}.[25]

Also highlighted in Figure 3a is a small upward shift in an amide I band from ~1665 cm^{-1} in RNase A to ~1676 cm^{-1} in RNase B. The band at ~1665 cm^{-1} is assigned to β-structure,[47] and bands occurring at 1675 cm^{-1} have been associated with turn structure.[48] Consequently, the upward shift in peak position could be attributed to conformational changes in the tertiary structure, specifically, the turn structure of RNase brought about by the addition of a carbohydrate group. PCA was able to separate these data into two distinct clusters of glycoprotein and protein (Figure 3b), where the separation was largely accounted for in the first principal component score (PC1 = $t[1]$) that accounts for 45.5% of the total explained variance. Inspection of the loadings matrix (data not shown) revealed the two major areas of significance: the amide I and III regions as highlighted in Figure 3a.

dx.doi.org/10.1021/ac2012009 |*Anal. Chem.* 2011, 83, 6074–6081

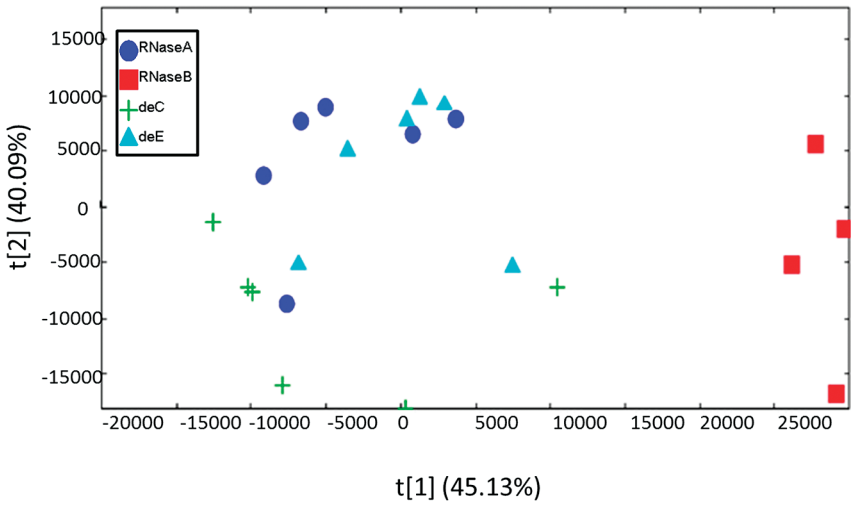


Figure 4. PCA scores plot (PC1 vs PC2) of Raman data from deglycosylated RNase B. deC refers to samples deglycosylated using the chemical approach, and deE, the enzymatic approach.

Deglycosylation of RNase B. The ability of Raman spectroscopy to differentiate between deglycosylated and native glycoproteins was investigated using both chemical and enzymatic deglycosylation methods (see above for optimized methods). Following deglycosylation, RNase B samples were first analyzed using MALDI-MS (data not shown) to confirm that the deglycosylation had been successful. All MALDI-MS spectra showed an average shift of 1460 m/z (average from 10 measurements; which equates to six mannose sugars in the glycan plus two GlcNAc). For the tryptic digest peptide RNL cleaved at positions 33 and 36, a m/z of 450 Da can be observed in the deglycosylated protein, as opposed to an average m/z of 2130 in the control protein, confirming the loss of the glycan.

Raman spectra were recorded for both controls (RNase A and B) and for the deglycosylated RNase B samples using Spectra RIM slides as described previously. The Raman spectra of deglycosylated RNase B in Figure 3a show that deglycosylated RNase follows the same trends in amide I and amide III bands as observed in the RNase A spectra.

To assess the changes in the Raman spectra from the deglycosylated proteins further, PCA was performed, and Figure 4 shows the PCA scores plot for both chemically (deC) and enzymatically (deE) deglycosylated protein spectra along with the two controls (RNase A and RNase B). This scores biplot shows glycosylated and nonglycosylated proteins in two distinct clusters separated in PC1. The spectra from chemically and enzymatically deglycosylated proteins fall very close to the RNase A spectra. This suggests that the differences observed are, indeed, due to the removal of sugars rather than any minor changes in tertiary structure. On closer inspection of this PCA scores plot, it is clear that the enzyme-treated samples more closely match RNase A than the chemically deglycosylated samples. This may be due to structural changes in the protein induced by the extreme pH needed to perform this deglycosylation of RNase B; this is likely to be structural because MALDI-MS showed no degradation of the intact deglycosyated RNase protein when analyzed in linear TOF mode (data not shown).

Quantification of Glycosylation. The next step was to determine whether it was possible to predict the extent of glycosylation using Raman spectroscopy. To test this hypothesis, RNase A and B mixtures were used to create a model system.

Raman data were acquired from 21 different mixtures of RNase A and B at 5% concentration intervals. The total protein concentration in each sample was kept constant at 1 mg/mL (~70 nM). Raman spectra were recorded from Spectra RIM slides using a Renishaw Raman microscope, again with excitation at 785 nm.

Data were preprocessed, as described previously, and PLSR was then applied to the data using PyChem software. Because PLSR is a supervised learning method that uses both X-data (Raman spectra) and Y-data (RNase B percent), alternate samples were used for training and cross-validation or testing (0, 10, 20 ..., 90, 100% were used for training, and 5, 15, 25%, etc. for cross-validation and testing, as described in the Materials and Methods Data Analysis section). The minimal number of factors suggested by the cross-validation data set was five (latent variables (LVs)), and hence, five factors were used for modeling. Figure 5a shows the PLSR plot of predicted versus actual concentrations of RNase B for this data, in which the holdout test set error was 5.56%. This plot clearly shows that Raman spectroscopic data can accurately measure the relative concentrations of the glycoprotein within a mixture of the native nonglycan protein.

The PLS loadings from the first two LVs ($w \times c[1]$ vs $w \times c[2]$, which account for the majority of the variance in the model), are plotted against each other in Figure 5b. In this depiction, each point represents a different wavenumber, with each symbol coding for a different spectral region. The outer circle indicates a 95% confidence level and only points outside this 95% confidence boundary have been plotted. The spectral regions that have been indicated by the loadings plot as being of importance to this model have been highlighted in the spectra of the RNase proteins and sugars shown in Figure 3a. It is clear from this representation that each of the six major areas of importance correspond to visible differences in the Raman spectra of RNase A and B, including band-broadening and changes in peak intensities and position. In addition, the spectra of both sugars that form the glycan (*N*-acetylglucosamine and mannose) show features in the majority of these regions.

Assignments for the amide I and III regions have been discussed previously. Changes in the other regions (~830, 880, 1000, and 1450 cm^{-1}) could be attributed to conformational protein changes, such as changes in the local solvent environment of the aromatic side chains.

dx.doi.org/10.1021/ac2012009 |*Anal. Chem.* 2011, 83, 6074–6081

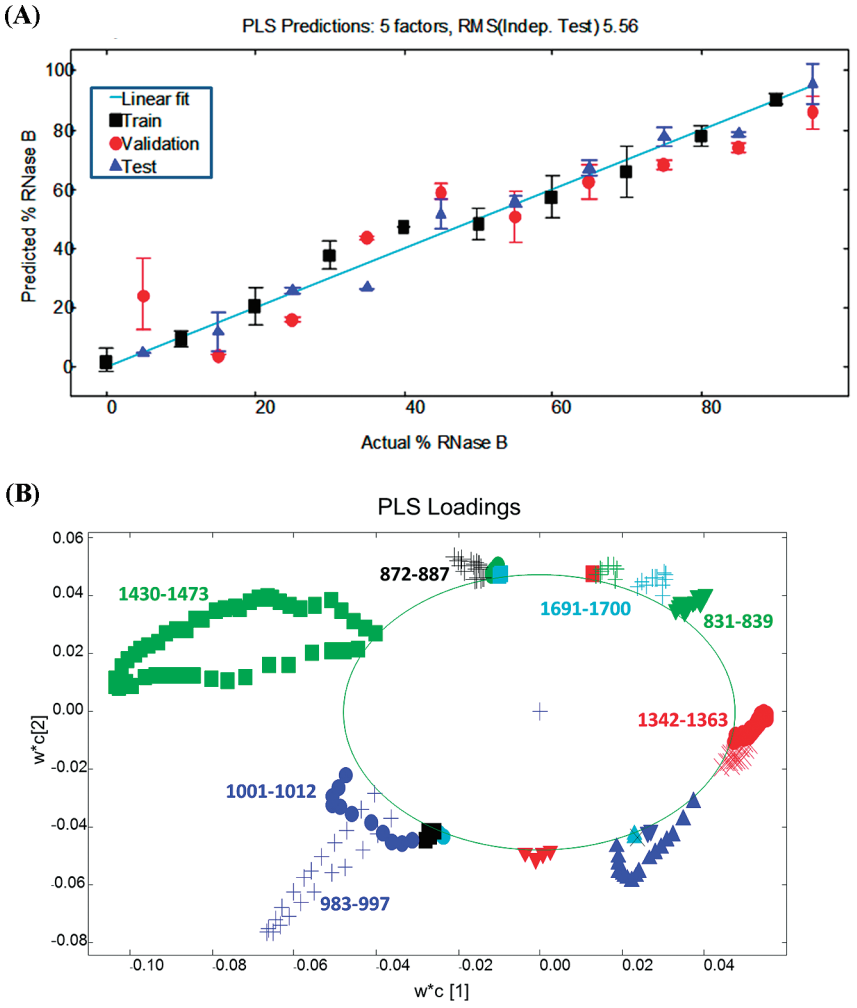


Figure 5. (A) PLS predictions from Raman data of RNase mixtures (mean predictions (from $n = 5$) are plotted with standard error bars); data preprocessing as described in Methods. (B) PLS loading plot of the first two latent variables (LVs; $w \times c[1]$ vs $w \times c[2]$); the green circle indicates 95% confidence.

Table 1. Band Assignments and Summary of Correlations Found between Peak Parameters and RNase B Concentration

~wavenumber (cm^{-1})	protein assignment	sugar assignment	peak parameter	R^2
830	Tyr ring[51]	glycosidic ring[52]	area	0.403
880	C−C backbone[52]	C−O−C stretch[52]	area	0.872
1000	Phe ring[10]	glycosidic ring[22]	area	0.924
1350	amide III[52]	NH_2 twist[25]	area	0.894
1450	alanine CH_3[52]	glycosidic ring[26]	area	0.732
1690	amide I[10]	n/a	center	0.910

However, because there are no aromatic residues close to the glycosylation site, it is more likely that these spectral differences are brought about by the presence of sugar bands in the spectra of RNase B. Figure 5b is of particular value because they not only inform us which vibrational modes are the most selective for determining glycosylation status (see Table 1) but also assist in confirming that the quantification model is based on real spectral features, as opposed to artifacts in the baseline and noise.

Further confirmation of the importance of these regions was gained by performing a Gaussian curve fit on the preprocessed data (using GRAMS software) to located peak areas and centers. Positive correlations were found between peak parameters and RNase B concentration in five out of the six spectral regions highlighted as important from PLSR; the assignments and correlation values (R^2) are detailed in Table 1.

To investigate these correlations further, a variation of 2D-correlation analysis,[49] using moving windows, was also applied as an alternative method of displaying the changes that occur within the data. 2D-correlation analysis is a cross-correlation technique that simplifies a data set by displaying relative similarities and relative differences as contour plots. Moving windows, in particular, relates spectral variation to the perturbation by separating the data into smaller data sets to locate key transition points.[43,44,50] Although this is not strictly a quantitative technique, it has been

dx.doi.org/10.1021/ac2012009 |*Anal. Chem.* 2011, 83, 6074–6081

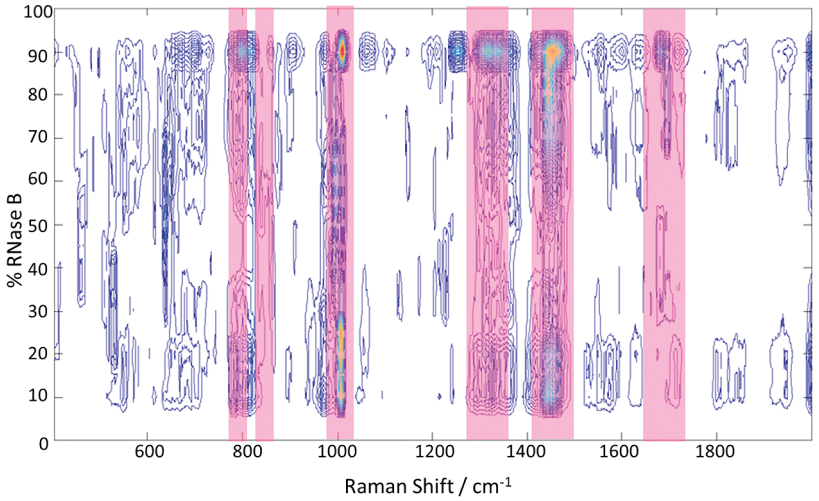


Figure 6. 2D-correlation moving windows plot as a function of spectral wavenumber and average translating window concentration of the RNase data.

performed to compare the regions in which the most change occurs with those highlighted by the PLS loadings plot.

A 2D moving windows plot for the RNase A and B Raman data is shown in Figure 6. The spectral regions where the most change occurs across the full % range, indicated by the largest number of contours, are highlighted. When compared with the spectra (Figure 3a) and the PLS loadings plot (Figure 5b), it can easily be seen that the results from both data analysis methods correlate as to which vibrational modes are the most important. It should also be noted that from the moving windows plot contours can be observed to form two distinctive groupings at 0−30% and 80−100%. This suggests that the majority of the spectral changes are not continuous, with many occurring in two stages, from 0 to 30% and 80 to 100%.

The final step for this analysis was to challenge PLS model by adding the spectra from the deglycosylation experiments into the PLS model. Although the model was able to predict the chemically and enzymatically deglycosylated protein spectra correctly, there was a larger error margin than with the standard mixtures. This error was 10.50%, compared with 5.56% generated previously.

In conclusion, the results in this paper clearly show the potential for Raman spectroscopy to be developed as a technique for monitoring (de)glycosylation. We have demonstrated that Raman spectroscopy with appropriate chemometric strategies is capable of distinguishing between the glycoprotein RNase B and the nonglycan containing RNase A protein as well as between native RNase B and deglycosylated forms, and future work will be to expand the range of proteins and gylcoproteins studied. This work has also illustrated the potential for Raman spectroscopic data to be combined with multivariate analysis methods for the successful quantification of the glycosylation status of target proteins.

AUTHOR INFORMATION

Corresponding Author

*Phone: +44(0) 161 3064480. E-mail: roy.goodacre@manchester.ac.uk.

ACKNOWLEDGMENT

We are very grateful to the UK BBSRC, EPSRC, Avacta group plc., and industrial members of the Bioprocessing Research Industry Club (BRIC) for funding this work. V.L.B. thanks Alastair Smith and Simon Webster of Avacta Plc. for additional CASE funding of her BBSRC studentship and for the use of their Raman spectrometer. V.L.B. also thanks Dr. Chris Sellick for advice on enzymatic deglycosylation and Dr. Roger Jarvis for introducing her to multivariate analysis and PyChem.

REFERENCES

(1) Greer, F. *Eur. BioPharm. Rev.* **2008**, *1*, 64–67.
(2) Hovatta, O.; Stojkovic, M.; Nogueira, M.; Varela-Nieto, I. *Stem Cells* **2010**, *28*, 1005–1007.
(3) Scott, C. T.; McCormick, J. B.; DeRouen, M. C.; Owen-Smith, J. *Nat. Methods* **2010**, *7*, 866–867.
(4) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta, Gen. Subj.* **1999**, *1473*, 4–8.
(5) Greer, F. *Eur. BioPharm. Rev.* **2007**, *6*, 106–111.
(6) Wen, Z. Q. *J. Pharm. Sci.* **2007**, *96*, 2861–2878.
(7) Sane, S. U.; Wong, R.; Hsu, C. C. *J. Pharm. Sci.* **2004**, *93*, 1005–1018.
(8) Tuma, R. *J. Raman Spectrosc.* **2005**, *36*, 307–319.
(9) Chi, Z. H.; Asher, S. A. *Biochemistry* **1998**, *37*, 2865–2872.
(10) Tuma, R.; Russell, M.; Rosendahl, M.; Thomas, G. J. *Biochemistry* **1995**, *34*, 15150–15156.
(11) Huang, K.; Maiti, N. C.; Phillips, N. B.; Carey, P. R.; Weiss, M. A. *Biochemistry* **2006**, *45*, 10278–10293.
(12) Zheng, R.; Zheng, X. J.; Dong, J.; Carey, P. R. *Protein Sci.* **2004**, *13*, 1288–1294.
(13) Wen, Z. Q.; Cao, X. C.; Vance, A. *J. Pharm. Sci.* **2008**, *97*, 2228–2241.
(14) Brewster, V.; Jarvis, R.; Goodacre, R. *Eur. BioPharm. Rev.* **2009**, *14*, 4–9.
(15) Shaw, A. D.; Kaderbhai, N.; Jones, A.; Woodward, A. M.; Goodacre, R.; Rowland, J. J.; Kell, D. B. *Appl. Spectrosc.* **1999**, *53*, 1419–1428.
(16) Abu-Absi, N. R.; Kenty, B. M.; Cuellar, M. E.; Borys, M. C.; Sakhamuri, S.; Strachan, D. J.; Hausladen, M. C.; Zheng, J. L. *Biotechnol. Bioeng.* **2010**, *108*, 1215–1221.
(17) Maquelin, K.; Kirschner, C.; Choo-Smith, L. P.; van den Braak, N.; Endtz, H. P.; Naumann, D.; Puppels, G. J. *J. Microbiol. Methods* **2002**, *51*, 255–271.
(18) Jarvis, R.; Goodacre, R. *Chem. Soc. Rev.* **2008**, *31*, 931–936.
(19) Jarvis, R.; Goodacre, R. *Anal. Chem.* **2004**, *76*, 40–47.
(20) McGovern, A. C.; Broadhurst, D.; Taylor, J.; Kaderbhai, N.; Winson, M. K.; Small, D. A.; Rowland, J. J.; Kell, D. B.; Goodacre, R. *Biotechnol. Bioeng.* **2002**, *78*, 527–538.
(21) Lopez-Diez, E. C.; Winder, C. L.; Ashton, L.; Currie, F.; Goodacre, R. *Anal. Chem.* **2005**, *77*, 2901–2906.

(22) Oleinikov, V.; Kryukov, E.; Kovner, M.; Ermishov, M.; Tuzikov, A.; Shiyan, S.; Bovin, N.; Nabiev, I. Prague, Czech Republic, Aug 23−28, 1998; Elsevier Science Bv; pp 475–480.
(23) Zhu, F.; Isaacs, N. W.; Hecht, L.; Tranter, G. E.; Barron, L. D. *Chirality* **2006**, *18*, 11.
(24) Zhu, F. J.; Isaacs, N. W.; Hecht, L.; Barron, L. D. *J. Am. Chem. Soc.* **2005**, *127*, 6142–6143.
(25) Arboleda, P. H.; Loppnow, G. R. *Anal. Chem.* **2000**, *72*, 2093–2098.
(26) Mrozek, M. F.; Zhang, D.; Ben-Amotz, D. *Carbohydr. Res.* **2004**, *339*, 141–145.
(27) Kikuchi, G. E.; Baker, S. A.; Merajver, S. D.; Coligan, J. E.; Levine, M.; Glorioso, J. C.; Nairn, R. *Biochemistry* **1987**, *26*, 424–431.
(28) Kopecky, V.; Ettrich, R.; Hofbauerova, K.; Baumruk, V. *Biochem. Biophys. Res. Commun.* **2003**, *300*, 41–46.
(29) Fleury, F.; Ianoul, A.; Baggetto, L.; Jardillier, J. C.; Alix, A. J. P.; Nabiev, I. SPIE-Int. Soc. Opt. Eng.; San Jose, CA, Jan 25−26, 1999; pp 80–89.
(30) Cui, Y.; Turner, G.; Roy, U. N.; Guo, M.; Pan, Z.; Morgan, S.; Burger, A.; Yeh, Y. *J. Raman Spectrosc.* **2005**, *36*, 1113–1117.
(31) Jarvis, R. M.; Blanch, E. W.; Golovanov, A. P.; Screen, J.; Goodacre, R. *Analyst* **2007**, *132*, 1053–1060.
(32) Sundararajan, N.; Mao, D. Q.; Chan, S.; Koo, T. W.; Su, X.; Sun, L.; Zhang, J. W.; Sung, K. B.; Yamakawa, M.; Gafken, P. R.; Randolph, T.; McLerran, D.; Feng, Z. D.; Berlin, A. A.; Roth, M. B. *Anal. Chem.* **2006**, *78*, 3543–3550.
(33) Taylor, M. E.; Drickamer, K. *Introduction to Glycobiology*; Oxford University Press: Oxford, 2006.
(34) Edge, A. S. B.; Faltynek, C. R.; Hof, L.; Reichert, L. E.; Weber, P. *Anal. Biochem.* **1981**, *118*, 131–137.
(35) Edge, A. S. B. *Biochem. J.* **2003**, *376*, 339–350.
(36) Tarentino, A. L.; Gomez, C. M.; Plummer, T. H. *Biochemistry* **1985**, *24*, 4665–4671.
(37) Hansen, R.; Dickson, A. J.; Goodacre, R.; Stephens, G. M.; Sellick, C. A. *Biotechnol. Bioeng.* **2010**, *107*, 902–908.
(38) Kreimer, D. I.; Ben-Amotz, D.; Zhang, D.; Xie, Y.; Ortiz, C.; DeGrella, R. F.; Adar, F.; Davisson, V. J. *Protein Sci.* **2004**, *13*, 125–127.
(39) ThermoScientific, In-Solution Tryptic Digestion and Guanidination Kit, http://www.piercenet.com/browse.cfm?fldID=973D9FA3-3C0B-452D-9633-AC1A8168D21D.
(40) Jarvis, R. M.; Broadhurst, D.; Johnson, H.; O'Boyle, N. M.; Goodacre, R. *Bioinformatics* **2006**, *22*, 2565–2566.
(41) Jolliffe, I. T. *Principal Components Analysis*; Springer-Verlag: New York, 1986.
(42) Wold, H. In *Multivariate Analysis*; Krishnaiah, K. R., Ed.; Academic Press: New York, 1966; pp 391–420.
(43) Thomas, M.; Richardson, H. H. *Vib. Spectrosc.* **2000**, *24*, 137–146.
(44) Morita, S.; Shinzawa, H.; Tsenkova, R.; Noda, I.; Ozaki, Y. *J. Mol. Struct.* **2006**, *799*, 111–120.
(45) Martens, H.; Naes, T. In *Multivariate Calibration*; John Wiley & Sons: Chichester, 1989; pp 73–236.
(46) Ellepola, S. W.; Choi, S.-Z.; Phillips, D. L.; Ma, C.-Y. *J. Cereal Sci.* **2006**, *43*, 85–93.
(47) Ashton, L.; Barron, L. D.; Hecht, L.; Hyde, J.; Blanch, E. W. *Analyst* **2007**, *132*, 468–479.
(48) Takekiyo, T.; Takeda, N.; Isogai, Y.; Kato, M.; Taniguchi, Y. B. *Biopolymers* **2006**, *85*, 185–188.
(49) Noda, I.; Ozaki, Y. *Two-Dimensional Correlation Spectroscopy: Applications in Vibrational and Optical Spectroscopy*; John Wiley and Sons Ltd.: Chichester, 2004.
(50) Ashton, L.; Blanch, E. W. *Appl. Spectrosc.* **2008**, *62*, 469–475.
(51) Siamwiza, M. N.; Lord, R. C.; Chen, M. C.; Takamatsu, T.; Harada, I.; Matsuura, H.; Shimanouchi, T. *Biochemistry* **1975**, *14*, 4870–4876.
(52) Socrates, G. *Infrared and Raman Characteristic Group Frequencies*, 3rd ed.; John Wiley & Sons Ltd: Chichester, 2001.