Supporting Information

The detection of protein glycosylation using tip enhanced Raman scattering

David P. Cowcher[a], Tanja Deckert-Gaudig[b], Victoria L. Brewster[a], Lorna Ashton[a,c], Volker Deckert[b,d*] and Royston Goodacre[a*].

[a] School of Chemistry and Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.

[b] Institute of Photonic Technology – IPHT, Albert-Einstein- Str. 9, 07745 Jena, Germany.

[c] Department of Chemistry, Faraday Building, Lancaster University, Lancaster, LA1 4YB, UK.

[c] Institut f. Physikalische Chemie, and Abbe Center of Photonics Friedrich-Schiller Universität, Helmholtzweg 4, 07743 Jena, Germany

[*] Correspondence to: roy.goodacre@manchester.ac.uk and volker.deckert@ipht-jena.de

---

Instrumentation and TERS measurement

NT-MDT non-contact AFM tips were evaporated with 25 nm silver and stored under argon until used. AFM images were acquired on a Nanowizard I (JPK Instruments AG, Germany), mounted on an inverted microscope (Olympus IX70, Japan) and equipped with an oil immersion objective, 60×, NA = 1.45 (Olympus, Japan). The objective was piezo-mounted (PIFOC, Physik Instrumente, Germany), thereby adjusting the laser focus to the z-movement of the TERS tip.

For excitation a krypton ion laser was used (Innova 300c, wavelength 530.9 nm, U.S.A.). Spectra were collected by a confocal Raman spectrometer (LabRam HR, Horiba Jobin Yvon, France) with a liquid nitrogen-cooled CCD camera (ISA Spectrum One, Horiba Jobin Yvon, France). The power on the sample was 820-850 μW.

The lateral steps of the tips were controlled by an additional 100 μm × 100 μm sample scanning stage (P-734, Physik Instrumente, Germany).

Spectra were recorded with acquisition times of 2 s, 5 s or 10 s (depending on the tip)

After each grid measurement a reference spectrum next to the gold plate was acquired to exclude tip contamination. Grids of various sizes with point-to-point distances of 2-10 nm were set on the gold

nanoplates. In total, 1977 spectra were recorded for the RNaseA sample and 2345 spectra for RNase B, respectively.

Control samples of the individual sugars that make up the glycans on RNase B, mannose and GlcNAc, were also prepared and measured. 1.5 μL of 1.00 mM solutions of the sugars were dropped onto cleaned glass slides and dried under vacuum. Raman spectra were recorded from large micrometre sized crystals. On areas appearing 'clean' under the microscope, TERS spectra were also recorded. The nanometre sized crystals were approximately 100-300 nm in width and 10-20 nm in height. The point-to-point distance for the measurements was mostly 5 nm. It was more difficult to obtain good spectra from mannose than GlcNAc and a total of 489 and 2650 spectra were recorded from each sugar respectively.

Data analysis

Data analysis was carried out using Matlab version R2013a (The Mathworks, Natick, MA, USA). In initial experimentation we did try to process the data using derivatives (first and second) and this failed to remove the baseline variation adequately; therefore we used the following approach: First, the spectra that showed no features except for the silicon background were removed from the data set. Cosmic rays were removed by inserting a linear fit between the start and end positions of the peak. The spectra were subsequently baseline-corrected using the asymmetric least-squares (ALS) method[1] with the smoothing parameter, λ = 1000 and the asymmetry parameter, p = 0.001.

Three different approaches were applied independently to negate accurately the influence of the Si peaks from the TERS tip:

(i), A Gaussian was fitted under the Si fundamental peak at 520 cm^{-1} and the parameters that described this Gaussian (centre position, width and height) were used to plot a predicted first overtone peak. The fitted fundamental Gaussian and predicted overtone Gaussian were then subtracted from the relevant part of each spectrum.

(ii) Spectra of the gold nanoplates without protein were averaged (mean) and then each spectrum was normalised against this mean.

(iii) A secondary ALS baseline correction was applied to the region of the spectrum containing the Si overtone feature, with the parameters λ = 100 and p = 0.01. The region of the spectrum containing the Si fundamental peak was set to zero.

For each of the three datasets now with the Si spectrum removed, each spectrum was row normalised to 1 of its total signal.

At each stage of the data processing, principal components analysis (PCA) was performed on the data, with 10 principal components (PCs) projected in each case.

Table S1 Tentative Raman band assignments for the regions of the TERS spectra that indicate differences between RNase A and RNase B

Region	Protein Assignment	Sugar Assignment
780 – 820 cm^{-1}	Tyrosine ring[2]	Glycosidic ring[2]
850 – 900 cm^{-1}	C-C backbone[2]	C-O-O stretch[2]
950 – 1000 cm^{-1}	β-sheet,[3] disordered secondary structure,[4] phenylalanine ring[2]	Glycosidic ring[2]
1220 – 1300 cm^{-1}	Amide III[4]	NH_2 twist[2]
1420 – 1490 cm^{-1}	CH, CH_2, CH_3[2]	Glycosidic ring[2]
1700 – 1800 cm^{-1}	Side-chain C=O[2]	n/a

Supplementary References

(1) Eilers P. H. C.; Boelens, H. F. M., Leiden University: Leiden, 2005

(2) Brewster, V. L.; Ashton, L.; Goodacre, R., *Anal. Chem.* **2011,** *83*, 6074-6081

(3) Howell, N.; Li-Chan, E., *Int. J. Food Sci. Technol.* **1996,** *31*, 439-451.

(4) Ashton, L.; Barron, L. D.; Hecht, L.; Hyde, J.; Blanch, E. W., *Analyst* **2007,** *132*, 468-479.

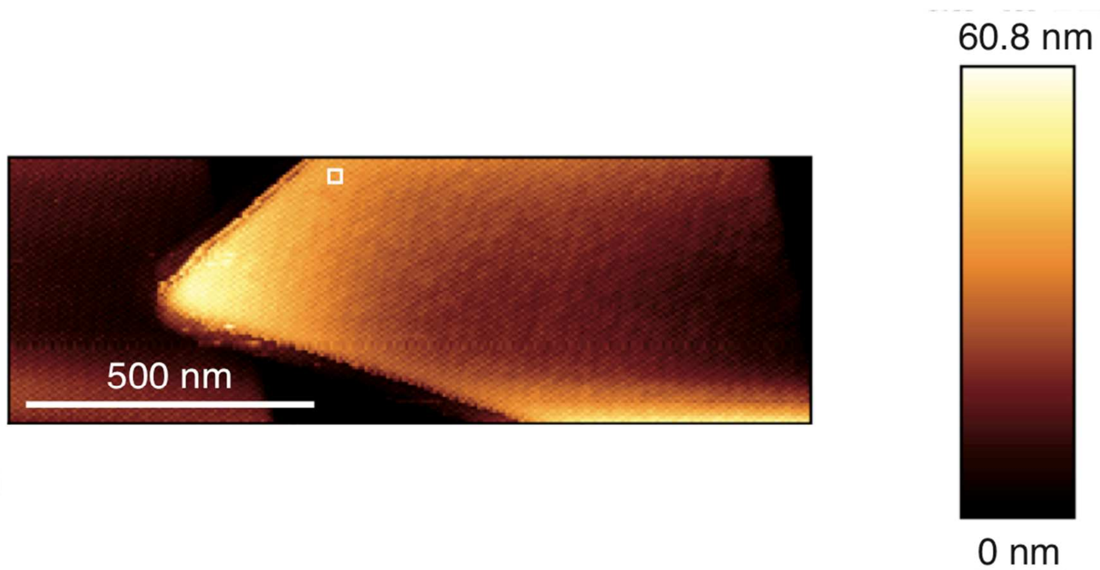

Figure S1. AFM topography of a typical gold nanoplate. The area used to collect a grid of 62 TERS spectra from this example is shown by the white square.

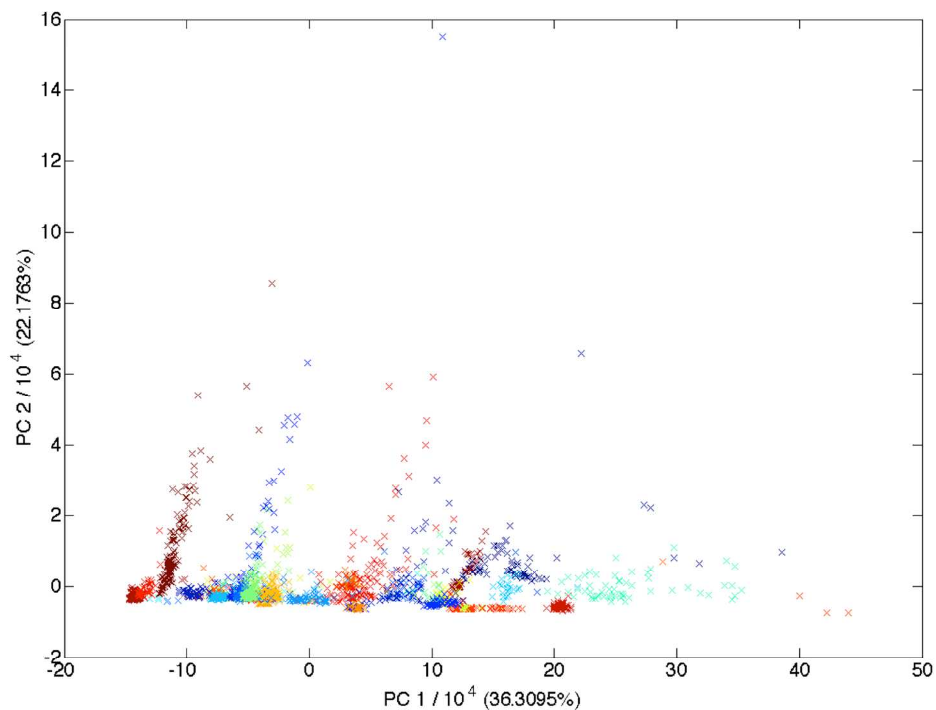

Figure S2. PCA scores plot of all data following cosmic ray removal and baseline correction. This is the same plot that is shown in Figure 2(b) but this time coloured to show how the PCA is grouping the data by measurement grid, based mainly on the spectral features of the silicon background, as shown by the loadings plot (Figure 2(c)).

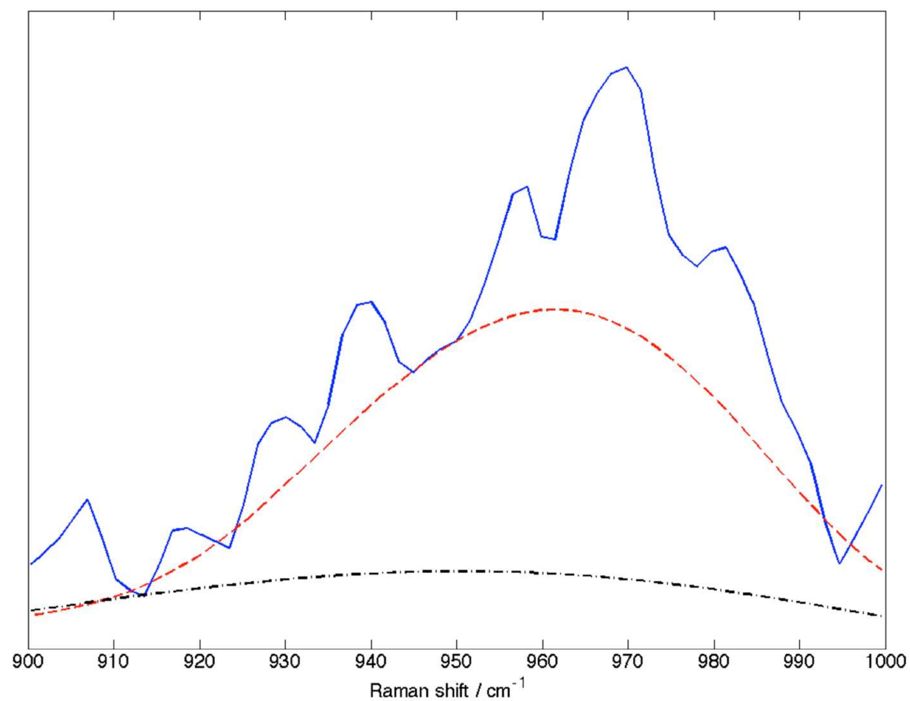

Figure S3. The fitting of asymmetric least-squares baselines under the Si overtone peak. The dot-dashed black line shows the fitted baseline as used to correct the whole spectrum (σ = 1000, p = 0.001) and the dashed red line shows the fitted baseline used to remove the Si overtone from this region (σ = 100, p = 0.01). The original TERS spectrum is shown in solid blue.

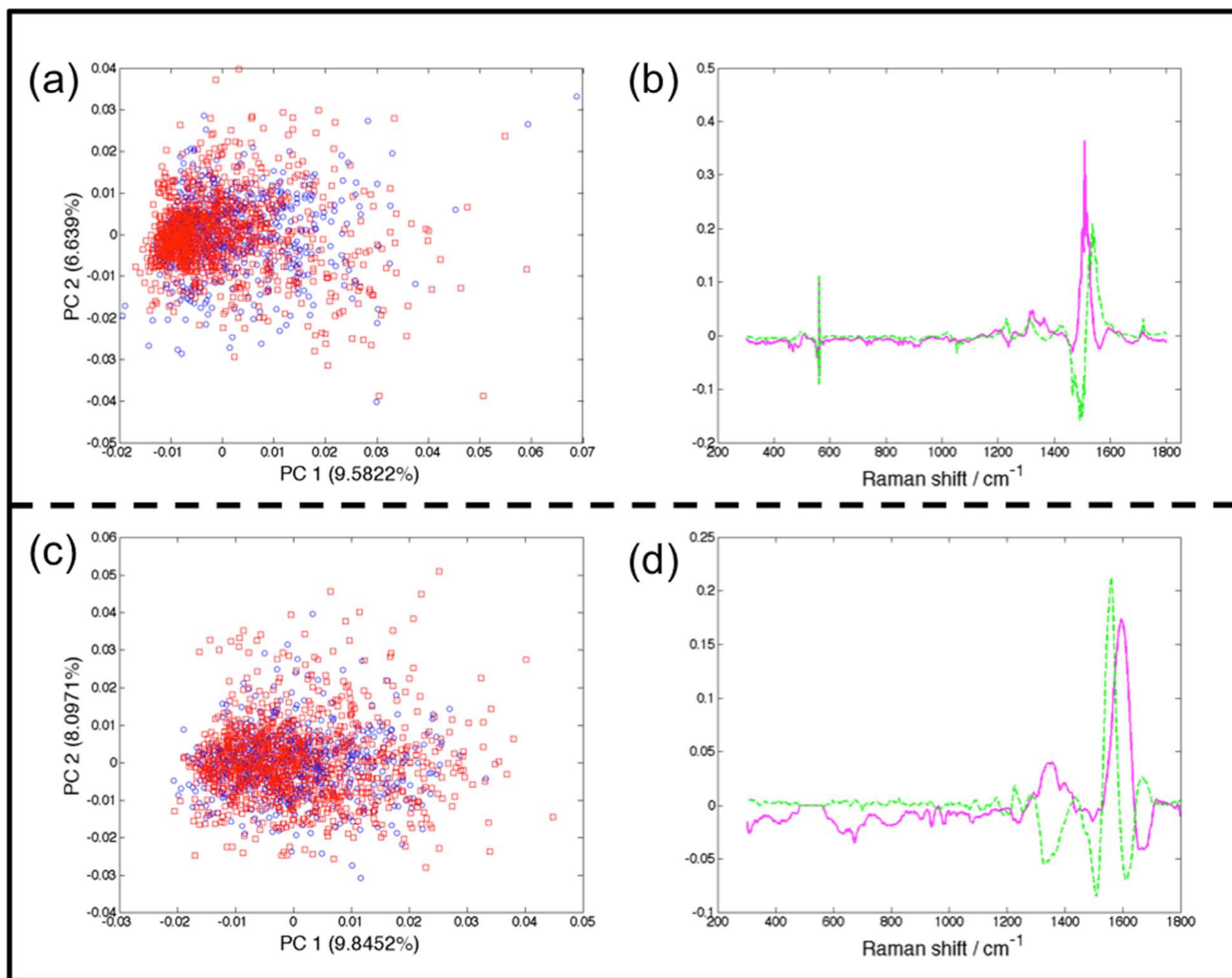

Figure S4. PCA plots of all data following normalisation against total signal. (a) and (c) Scores plots with RNase A in blue circles and RNase B in red squares, with % explained variance in brackets. (b) and (d) loadings plots of PC1 (solid purple) and PC2 (dashed green). (a) and (b) following the normalistion against blank Si and (c) and (d) following the ALS Si-removal route.

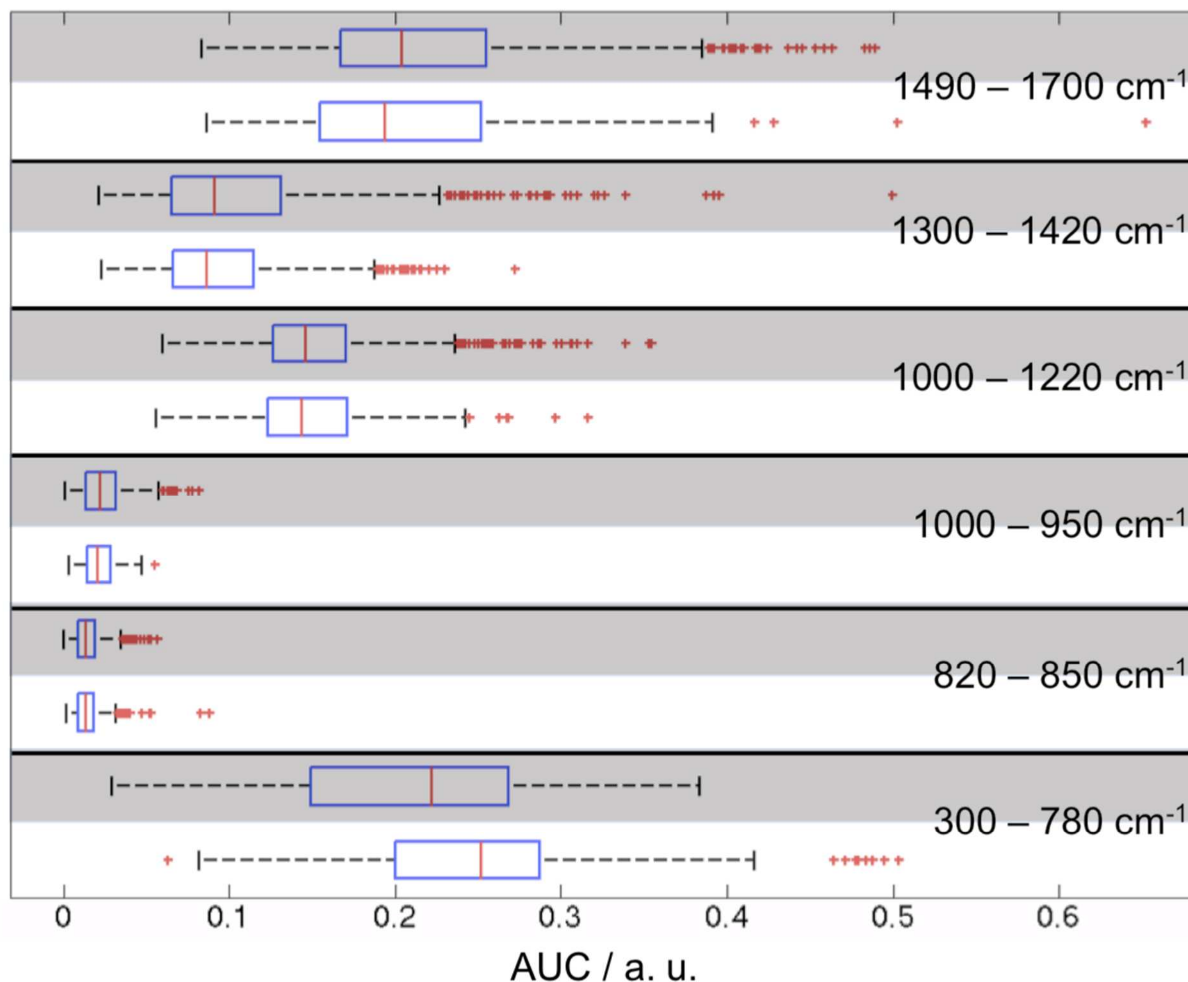

Figure S5 Box and whisker plot of the area under the preprocessed TERS spectra in the wavenumber regions that did not show differences between the two protein glycoforms using normal Raman spectroscopy. For each region, the shaded box and whisker represents RNase B. In each case the blue box represents the interquartile range (IQR), the red line the median, the black dashed whiskers the remaining data, and any outliers are indicated by the red crosses. In this example, the data were preprocessed using the normalisation against the blank Si method.

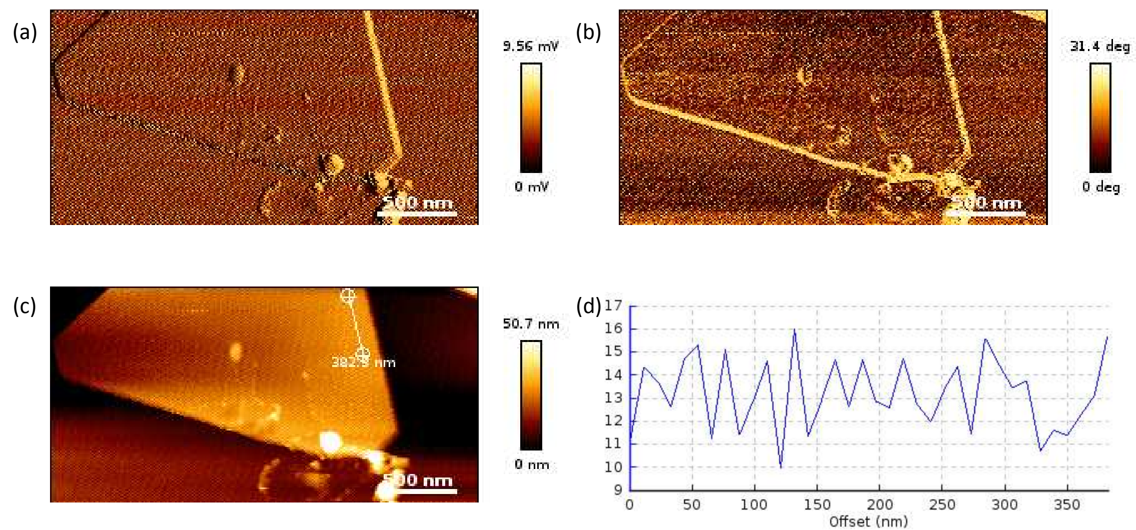

Figure S6 AFM topography of a typical gold nanoplate covered in proteins showing: (a) amplitude trace; (b) phase trace; (c) height trace indicating line used in (d); (d) AFM height trace showing presence of objects with an approximate height of 3 nm.