Electronic Supplementary Material (ESI) for Analyst.
This journal is © The Royal Society of Chemistry 2020

Quantification of Protein Glycation Using Vibrational Spectroscopy

Bethan S. McAvan[1], Aidan P. France[1], Bruno Bellina[1], Perdita E. Barran[1], Royston Goodacre[2] and Andrew J. Doig[3*]

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

[2] Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK.

[3] Manchester Institute of Biotechnology and Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9PT,

SUPPORTING INFORMATION

Supplementary Methods

Materials and Reagents

Recombinant human lysozyme expressed in rice, lyophilised powder, ≥100,000 units/mg protein (Lysobac) was obtained from Sigma Aldrich. D-(+)-glucose ≥99.5% anhydrous was purchased from Thermo Fischer. Recombinant human albumin expressed in rice, lyophilised powder was purchased from Sigma Aldrich. Amicon Ultra 0.5 mL centrifugal filter units, 10 KDa cut off were used for buffer exchange. For the quantification, incubation of lysozyme and albumin were buffer exchanged to remove free glucose using Centripure P100 columns (Emp Biotech GmbH). Satorius Stedim Biotech ministart 0.2 μm syringe filters were used to filter the incubation buffers. NaCl and Ammonium Acetate were purchased from Thermo Fisher Scientific. MOPS ≥99.5% was obtained from Sigma-Aldrich. Formic acid 98.0-100% analytical grade was purchased from Honeywell Fluka™. Bio spin-6 columns were purchased from Bio-Rad.

Sample Preparation

All incubations were carried out under sterile conditions as a 5 mg/mL solution in 25 mM MOPS, 115 mM NaCl, adjusted to pH 7.4. These were subsequently divided into two equal volumes to create two groups of standards and glucose incubations. To the glucose incubation group, glucose 0.5 M was added to create an overall sugar concentration of 0.5 M. All samples were statically incubated at 37 °C. The sterility of the samples was determined by plating an aliquot of the incubation solution onto LB agar (5 g NaCl, 5 g Tryptone, 2.5 g Yeast Extract, 7.5 g Agar, H_2O 500 mL) which was incubated at 37 °C for 48 h. No bacterial growth was seen across all incubations for both albumin and lysozyme.

General Mass Spectrometry Methods

All experiments were performed using nano-electrospray ionisation (nESI) in positive ionisation mode. Samples were infused into emitters prepared in-house from thin walled (O.D. 1.2 mm, I.D. 0.9 mm; WPI, UK) and thicker walled (O.D.: 1.2 mm, I.D.: 0.69 mm) fire polished borosilicate glass capillaries using a Sutter P1000 or P2000/F micropipette puller (Sutter Instrument Company, USA). In order to facilitate more facile spraying, all solution loaded emitters were centrifuged at 1957 rcf for 15 s prior to loading into the instrument tip holders. Platinum wire (diameter 0.125 mm, Goodfellow Cambridge Ltd, UK), was inserted into the capillaries to permit efficient ionisation. All Experimental masses were deconvoluted using UniDec V3.2 mass deconvolution software.[1]

Time Course Mass Spectrometry

A Waters Synapt G2 (Waters, Manchester UK) instrument was used to collect mass spectrometry data on the glycated lysozyme samples. Experiments were carried out in resolution mode. The instrument parameters were as follows: Capillary voltage 1.2 kV, sampling cone 20 V, extraction cone 1 V and trap collision energies of 4 V and 20 V. Default parameters were employed for all the other instrument settings. Lysozyme samples were solubilised in acidified ammonium acetate solution (200 mM AmAc 0.01 % formic acid) and

diluted to a concentration of ~5 μM protein whereby acidification was employed to improve signal. All subsequent analysis was undertaken using the MassLynx.V4.1 software whilst peak integration was carried out using OriginPro 8.5.1. The % of glycation for lysozyme was calculated *via* integrating the area under the MS peaks of the glycated/non-glycated species for the most abundant charge states of the protein (the 7^+ and 8^+ ions). Each addition of glucose to lysozyme was noted by a mass increase of 162 Da (addition of glucose – H_2O). Errors were calculated as standard errors (SE) across triplicate measurements.

Lysozyme Quantification and Albumin glycation Mass spectrometry.

12 mg/mL glycated and non-glycated lysozyme in MOPS were diluted at a protein to solvent volume ratio of 1:200 ratio into 1 M ammonium acetate solution pH 7.2 and analysed at a final protein concentration of ~4 μM. High concentrations of ammonium acetate were employed so as to reduce the impact of non-volatile salts (from MOPS) on the MS resolution.[2] 12 mg/mL rice albumin in MOPS was buffer exchanged 2x, using Bio spin-6 columns (BioRad, USA), into 1 M ammonium acetate solution pH 7.2. Optimal MS resolution was achieved for the glycated form of albumin by diluting the buffer exchanged sample at a 1:40 ratio into a water/formic acid solution (99.9:0.1, %v:v) at which the final protein concentration was ~5 μM prior to analysis. For consistency the non-glycated form of albumin was also treated in the same manner. MS Experiments were performed on a Q Exactive UHMR hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher scientific). Instrument parameters utilised were as follows: Capillary voltage 0.6-1.5 kV, inlet capillary temperature 400 °C and the collision gas within the HCD cell was nitrogen. The gas pressure within the HCD cell was set as follows: 1 (lysozyme) & 4 (albumin) corresponding to ultra HV pressures of $\sim 1.2\times10^{-10}$, and $\sim 5.2\times10^{-10}$ Torr respectively. A full list of the Q Exactive settings utilised are outlined in the associated Table S3. Mass spectra were recorded at a resolving power of 200,000 (lysozyme) and 3,125 (albumin). The lower resolution setting was employed for albumin so as to permit the characterisation of the main features within the glycated form of the protein. All subsequent analysis was undertaken using the Thermo Xcalibur 2.2 software and peak integration was carried out using OriginPro 8.5.1.

Raman Spectroscopy

Raman measurements were undertaken on a Renishaw inVia Raman microscope (Renishaw Plc., Gloucestershire, UK) using a 785 nm laser with a laser power on the sample of ~30 mW. Calibration was carried out by using a x50 objective (Renishaw), 10% power and 1 s acquisition focused onto a silicon plate centred at 520 cm^{-1}. The experimental parameters used for all data collection were 10 s exposure and 12 accumulations resulting in an overall acquisition time of 120 s per measurement. A 15 mm long distance objective (Renishaw) was used with a grating of 600 l/mm, centred at 1500 cm^{-1}. The set-up consisted of a 96 well quartz plate (Hellma™) using randomised sample collection orders with three repeats per well. Raw and pre-processed spectra are shown in Figure S3.

Data Pre-processing and Chemometrics.

All data pre-processing and subsequent analysis was performed using Matlab R2018a with in-house toolboxes. For FTIR-ATR the data were baselined using EMSC with a polynomial order of 2.[3] The Raman spectra were pre-processed using EMSC with a polynomial order of 3 and smoothed using Gaussian moving window with a width of 4.[4] The data set was then divided into two sets in order to generate a PLSR linear predictive model. The data set included 19 concentrations of glycated lysozyme spiked in standard lysozyme creating a ~6 % gradient from 0 to 100% glycated samples. This produced 19 datasets of which three were removed due to being outliers (see Figure S6). The training set and test set therefore included 8 datasets each with three spectral repeats per concentration. Several models were created and the optimum number of latent variables (PLS factors) was chosen as the lowest latent variable number with the lowest root-mean-squared error on the cross-validation set (RMSEcv) which was 3; i.e., the least number of PLS factors needed to decrease the RMSEcv. Leave-one-out cross validation was used to test each model. The chosen model with LV = 3 was used to run the test set and generate an R^2 and Q^2 for the model.

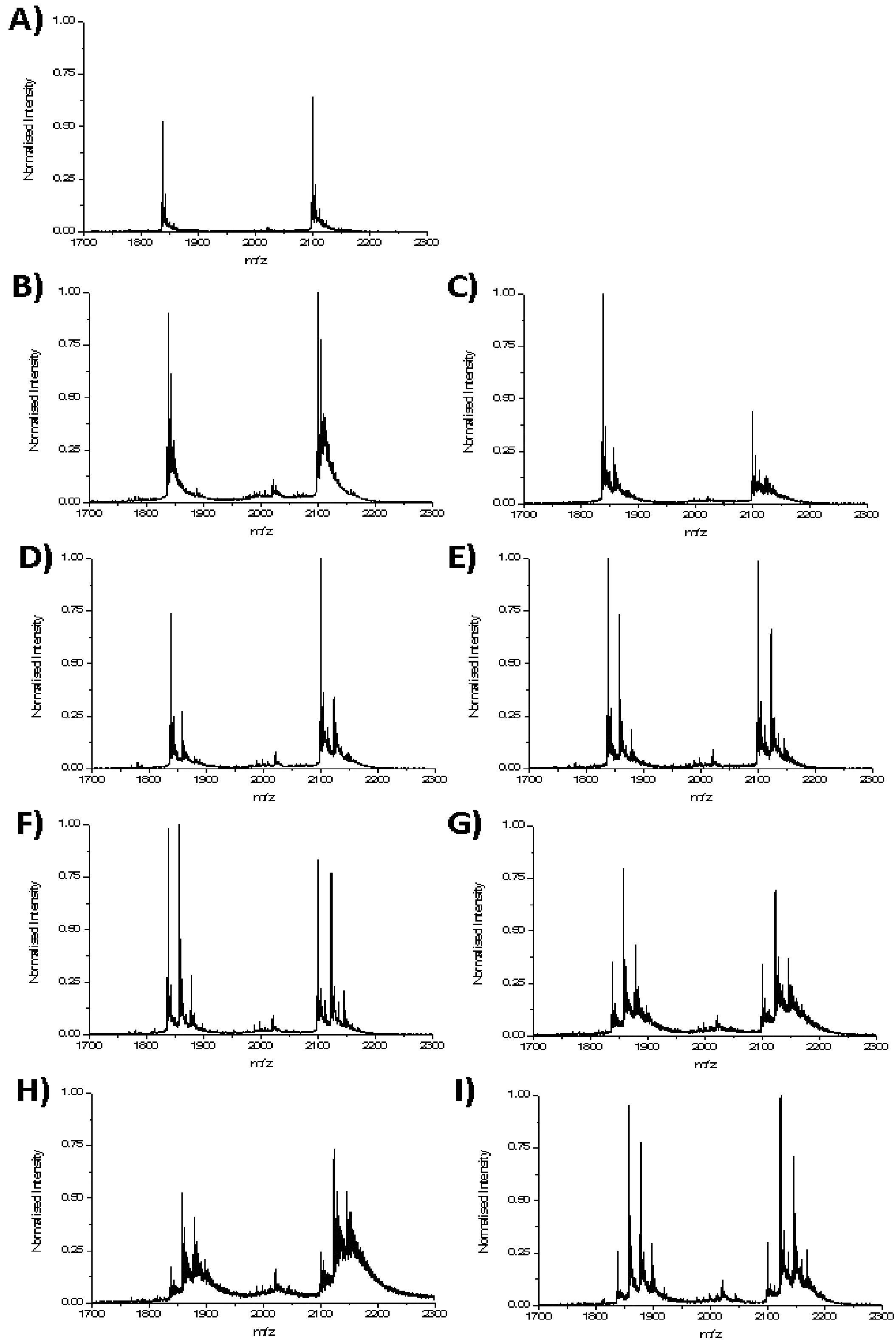

Figure S1. Normalised 7+ and 8+ mass spectra of lysozyme analysed across the time course on the Synapt G2 at 4 V. A) 5 μM lysozyme (alone), B) 5 μM lysozyme spiked with 70 μM glucose with no incubation, and 5 μM Lysozyme incubated with glucose for C) 1 Day, D) 2 Days E) 5 Days, F) 7 Days, G) 14 Days, H) 21 Days and I) 30 Days.

Table S1. Summary of deconvoluted masses for lysozyme analysed on the Synapt G2.

	Deconvoluted Mass (Da)	Mass Difference (Da)
Lysozyme	14,693.0	
lysozyme + 1 Glucose	14,855.2	162.2
Lysozyme + 2 Glucose	15,017.3	162.1
Lysozyme + 3 Glucose	15,179.5	162.2
Lysozyme + 4 Glucose	15,341.6	162.1
	Average mass differences (Da)	162.2
	Stdev	0.1

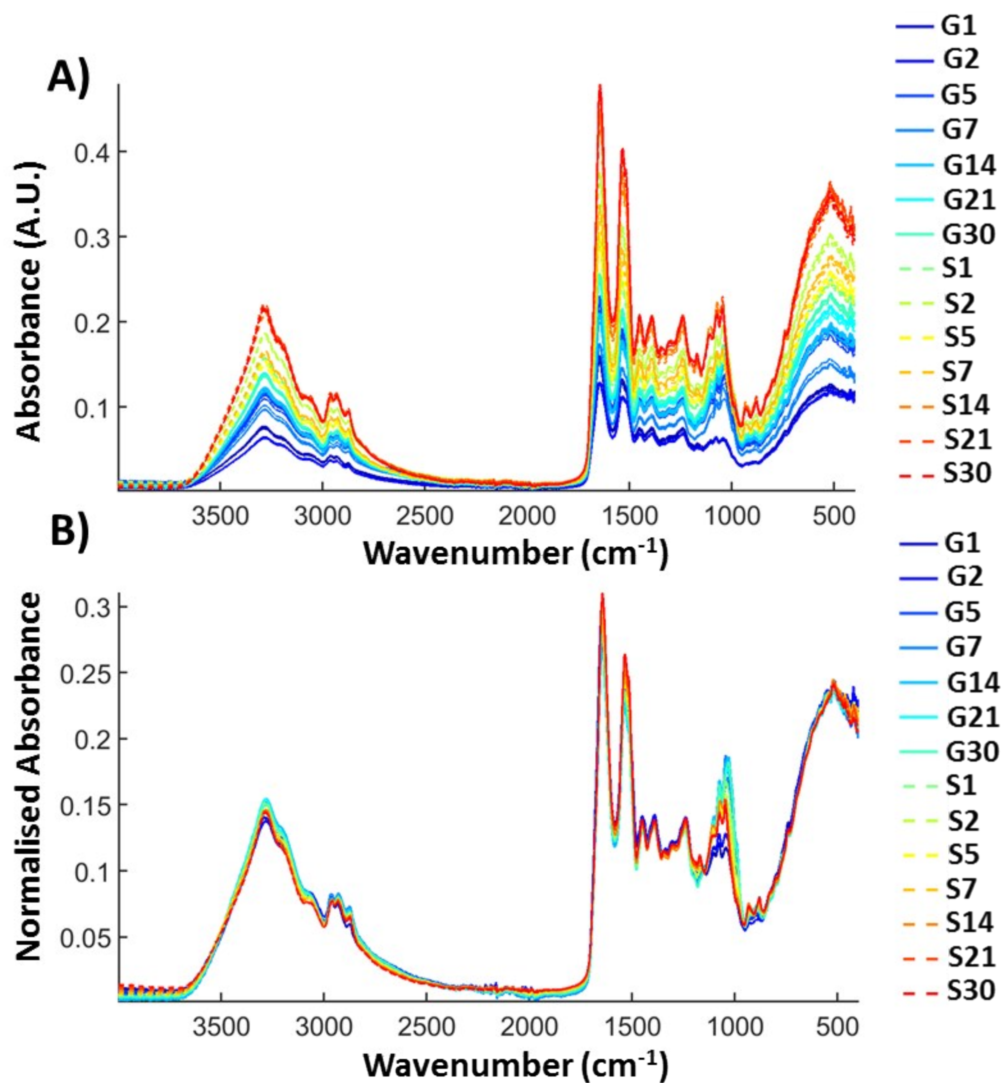

Figure S2. FTIR-ATR spectra of lysozyme time course incubated with 0.5 M glucose over a course of 30 days with samples taken at days 1, 2, 5, 7, 14, 21 and day 30. A) Raw spectra shown as 5 spectral repeats per sample and B) Pre-processed data using EMSC as described in supplementary methods. G stands for glycated samples and S for standard samples.

Table S2. Tentative IR assignments for D-(+)-Glucose and Lysozyme taken from literature.[5–8] Table adapted from reference.[9]

IR assignment			
Wavenumber (cm^{-1})	**D-(+)-Glucose**	**Wavenumber (cm^{-1})**	**Lysozyme**
3392	O-H	3279	O-H
3296	O-H	3208	O-H
3243	O-H	3068	O-H
2943	CH_2 C-6	2961	C-H
2912	CH C-3	2930	C-H
2891	CH_2 C-1	2873	C-H
1457	CH_2 Def	1695	C=O (Amide I β-Sheet)
1371	CH_2	1683	C=O Amide I Unordered
1338		1672	C=O Amide I Unordered
1296		1657	C=O Amide I α-Helix
1224		1647	Amide I C=O Unordered
1203		1639	Amide I C=O β-Sheet
1146	C-O, C-C	1627	Amide I C=O β-Sheet
1109	C-C, C-O, C-OH	1537	Amide II N-H and C-N
1048		1515	Amide II N-H and C-N (Tyr)
1014		1452	C-H Def
993	C-CH, C-CO	1389	CH_3
915		1324	Amide III α-Helix CH
837	C-C	1310	Amide III α-Helix CH
775	C-CC, C-CO,	1289	Amide III α-Helix
725		1280	Amide III Unordered
615		1262	Amide III Unordered
555		1248	Amide III Unordered
537		1233	Amide III β-Sheet
		1175	CH_2
		1102	CH_2
		1075	(C–O–C), polysaccharides
		1046	(C–O–C), polysaccharides
		990	C-N
		933	C-H
		879	C-H
		741	C-H

Table S3. Q Exactive UHMR tuning parameters for glycation analysis of lysozyme and albumin.

Scan parameters		
	Lysozyme	**Albumin**
Scan type	SIM	SIM
Scan range (*m/z*)	500 – 4,000	500 – 10,000
Fragmentation	none	none
Resolution	200,000	3,125
Polarity	positive	positive
Microscans	5	10
Lock masses	Off	Off
AGC target	5×10^5	5e5
Maximum inject time	150	5
nESI source Voltages and Temperatures		
Spray voltage (kV)	0.6 – 1.5	
Capillary temperature (°C)	400	
S-lens RF level	200.0	
Instruments settings		
Extended trapping (eV)	-1	
Averaging	0	
Detector *m/z* Optimization	Low *m/z*	High *m/z*
Noise threshold	4.64	
AGC mode	Fixed	Pre-scan
Ion transfer		
Source DC Offset (V)	21	21
In-source Trapping	Off	Off
Desolvation Voltage (V)	0	0
Inj.FI.RF Amplitude (V)	150	700
Bent Flatpole FI. RF Amplitude (V)	300	940
Trans MP and HCD-cell RF (V)	250	900
C-Trap RF Amplitude (V)	2300	2950
Injection Flatpole DC (V)	5.0	
Inter Flatpole Lens (V)	1	
Bent Flatpole DC (V)	4	
Transfer Multipole DC (V)	0	
C-trap Entrance (V)	1.8	
HCD event and Vacuum system		
Purge time (ms)	5.0	
C-trap exit (V)	2	
HCD Field Gradient	20.0	
Trapping gas pressure settings	1.0	4.0
C-Trap Charge Detector Support	On	
HV	Enabled	
High Vacuum	Turned off	
Ultra High Vacuum	1.22×10^{-10}	5.26×10^{-10}

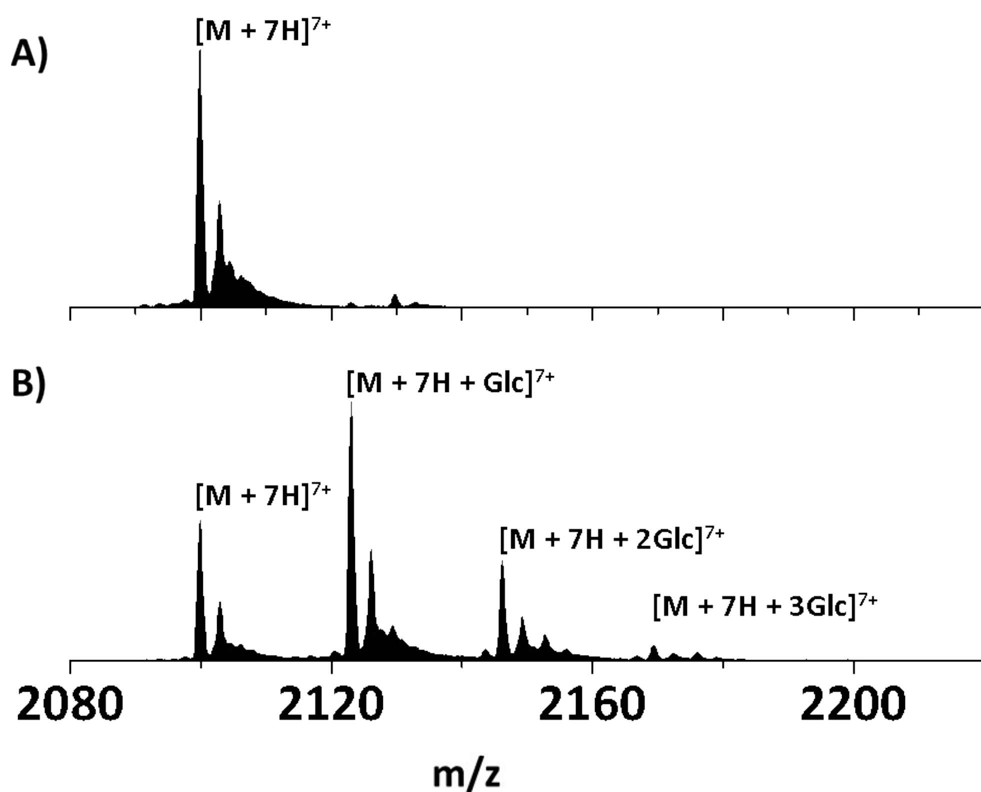

Figure S3. Mass spectra of 7^+ charge state of lysozyme obtained on the Q Exactive UHMR: A) after 30 days incubation with no glucose and B) after 30 days incubation with 0.5 M glucose.

Table S4. Summary of deconvoluted masses calculated for the glycated lysozyme sample. Mass difference shows an average value of 162.4 Da which corresponds to the mass of Glc – H_2O.

	Deconvoluted Mass (Da)	Mass difference (Da)
Lysozyme	14690.8	
Lysozyme + 1 glucose	14853.9	163.1
Lysozyme + 2 glucose	15016.3	162.4
Lysozyme + 3 glucose	15178.0	161.7
	Average mass differences (Da)	162.4
	Stdev	0.7

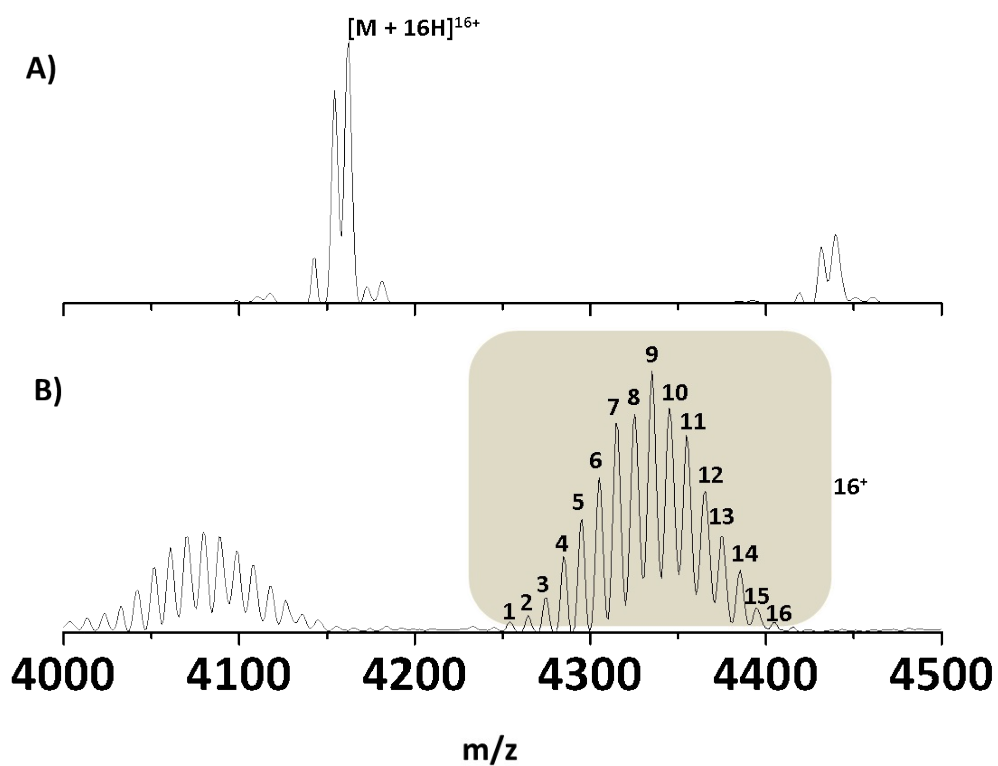

Figure S4. Mass spectra showing the 16^{+} charge state of albumin obtained on the Q Exactive UHMR: A) after 30 days incubation with no glucose and B) after 30 days incubation with 0.5 M. The number of bound glucose molecules is denoted numerically within the shaded box.

Table S5. Summary of the 16^+ charge state peaks found for the glycated albumin sample shown in Figure S4. The mass difference between the peaks shows an average value of 160.98 Da which corresponds to the subsequent addition of Glc – H_2O (within experimental error). Peak numbers match the labels in Figure S4.

Peak	# Glycations	Mass (Da)	Mass Difference Between Adjacent peaks (Da)
1	9	68047.1	
2	10	68211.1	164.0
3	11	68375.9	164.8
4	12	68537.5	161.6
5	13	68699.9	162.4
6	14	68862.3	162.4
7	15	69025.5	163.2
8	16	69183.9	158.4
9	17	69345.2	161.28
10	18	69508.7	163.52
11	19	69670.3	161.6
12	20	69833.5	163.2
13	21	69983.9	150.4
14	22	70148.7	164.8
15	23	70310.3	161.6
16	24	70468.7	158.4
		Average Mass Difference	160.98
		Stdev	3.67

Table S6. Calculation to determine the amount of glucose bound to albumin from the masses given in Table S5. The mass difference between D and E is 17.5 Da and could possibly be the association of a single NH_3 adduct. Accordingly we believe each peak in the series to contain an NH_3 adduct as well as a previously assigned number of bound glucose molecules.

	Mass Of Albumin (Calculated from standard)	**66571.2 Da**
A	Most abundant potentially glycated mass **(6 from Table S5)**	69345.2 Da
B	Glucose bound mass shift (Glc – H_2O)	162.1 Da
C	Mass of NH_3	17.0 Da
D	Mass of most abundant glycated (A) – Standard mass of albumin	2774.0 Da
E	[(D) – (C)] / 17 [# of most abundant glycations]	162.2 = Mass of Glucose
F	number of glucose molecules added*	17.1
G	Mass of glucose × the number of glycations (whole number = 17)	2756.4 Da
H	Mass difference between C-F	17.5 Da = ~Mass of NH_3

*Due to the amount of glucose that are able to bind to albumin, including the amount of salts and H^+ that can be also associated, analysis of the results is difficult and so we tentatively assign the mass difference in (H) as an NH_3 adduct.

Table S7. Set-up of the concentration gradient from 0% glycated lysozyme to 100% in ~6% intervals. Both samples (stnadard and glycated lysozyme) were concentrated to 12 mg/mL and free glucose from incubation was buffer exchanged out of the glycated samples.

Standard (µL)	**Glycated (µL)**	**% Glycated**	**Sample Number**
360	**0**	**0**	**1**
340	**20**	**6**	**2**
320	**40**	**11**	**3**
300	**60**	**17**	**4**
280	**80**	**22**	**5**
260	**100**	**28**	**6**
240	**120**	**33**	**7**
220*	**140**	**39**	**8**
200	**160**	**44**	**9**
180	**180**	**50**	**10**
160	**200**	**56**	**11**
140	**220**	**61**	**12**
120	**240**	**67**	**13**
100	**260**	**72**	**14**
80	**280**	**78**	**15**
60	**300**	**83**	**16**
40	**320**	**89**	**17**
20	**340**	**94**	**18**
0	**360**	**100**	**19**

*Concentrations in red are the ones which were subsequently found to be spectral outliers shown in Figure S5A) and using PCA in Figure S6.

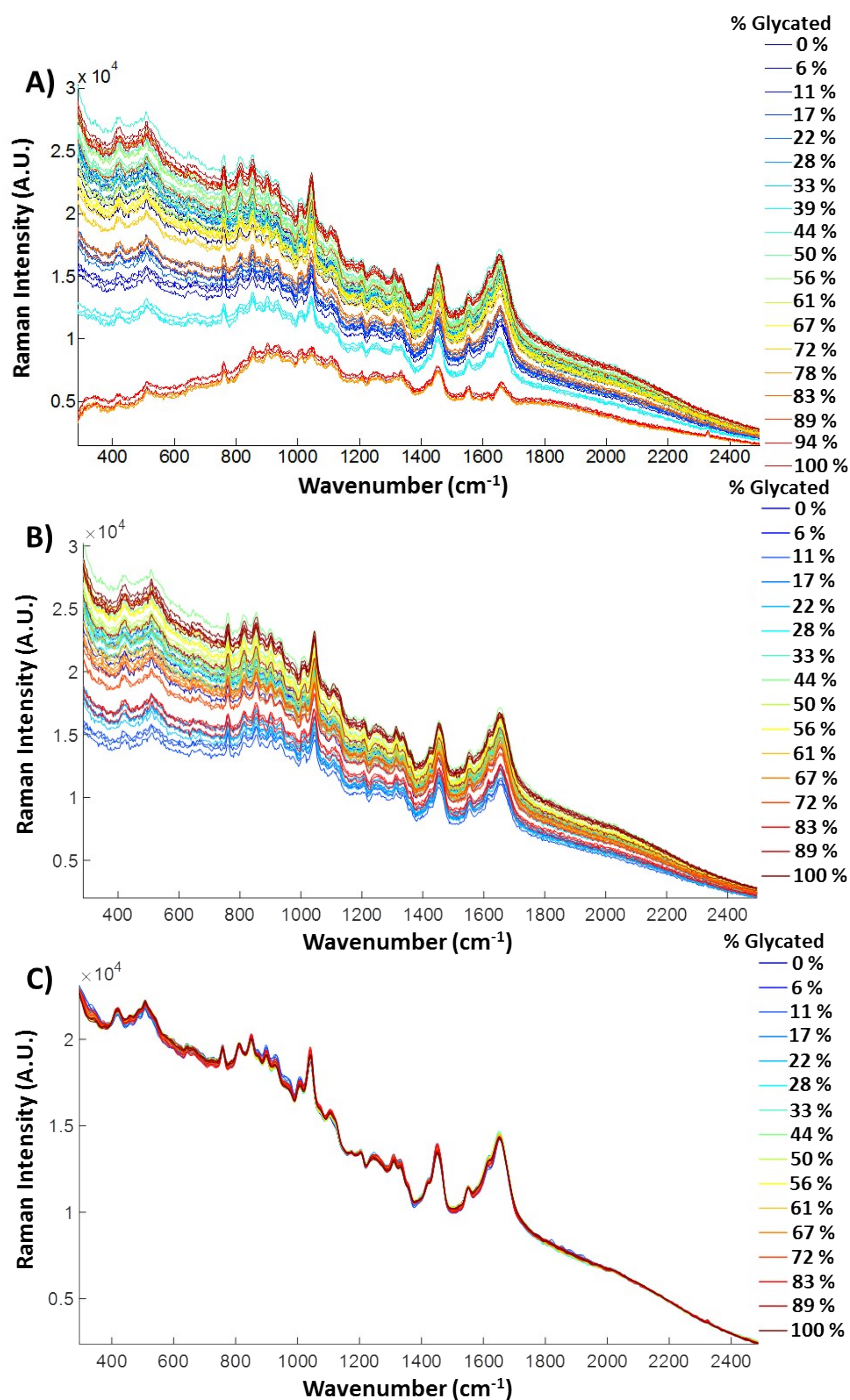

Figure S5. Raman spectra of Lysozyme concentration gradient shown as three spectral repeats per sample A) Raw Raman spectra, B) Raw Raman spectra with outliers removed (39%, 78% and 94% glycated) and C) Pre-processed Raman spectra which was baseline corrected and normalised using EMSC scaling outlined in the supplementary methods (Data Pre-processing and chemometrics).

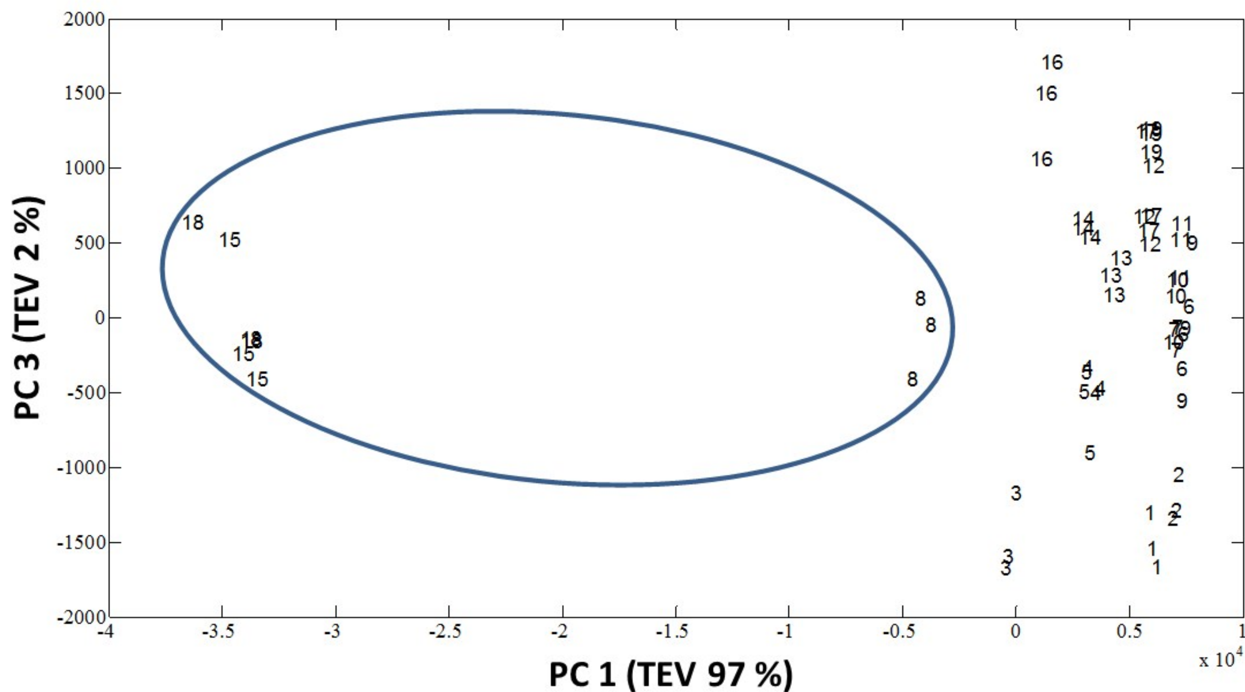

Figure S6. PCA of pre-processed Raman spectra of lysozyme concentration gradient with 3 spectral repeats for each sample. Samples 15, 18 and 8 were outliers and therefore PC 1 with a TEV = 97% meaning that the PCA mainly explains the difference between the sample outliers and therefore masked the spectra information between different concentration of glycated lysozyme and therefore were removed. TEV = total explained variance.

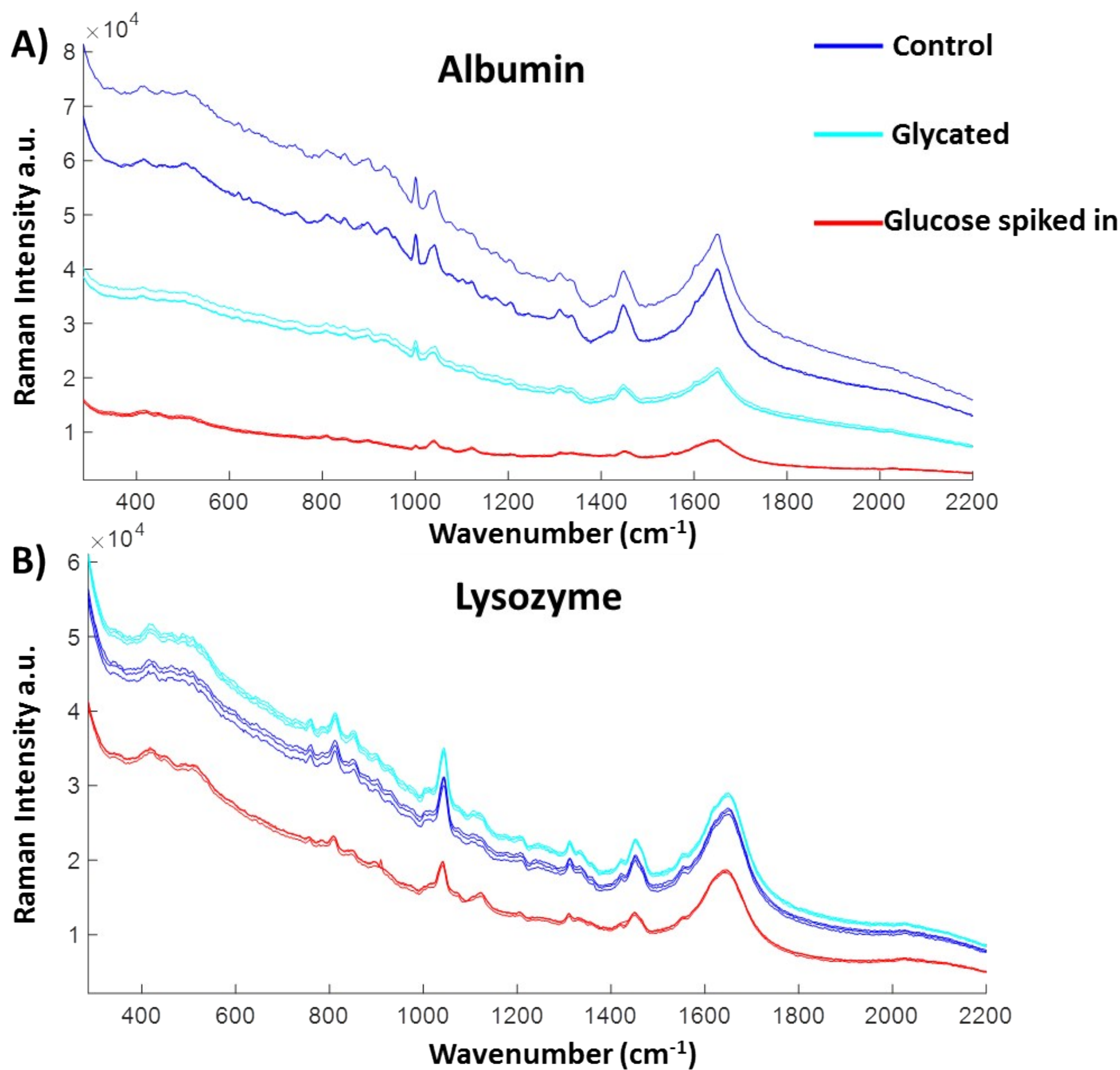

Figure S7. Raw Raman spectra of A) Albumin and B) Lysozyme showing glycated sample incubated for 30 days with glucose, control samples incubated for 30 days without glucose and C) proteins at 12 mg/mL with glucose 10 mg/mL spiked in (glucose spiked in). All spectra are shown as three spectral repeats per sample.

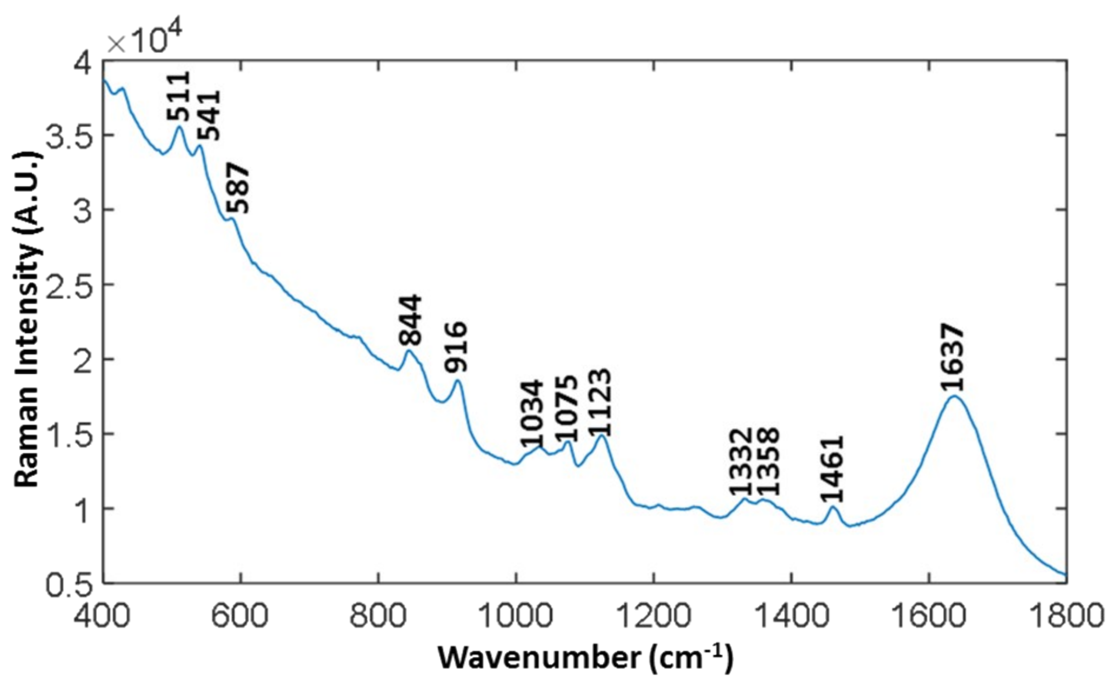

Figure S8. Raman spectrum of glucose 10 mg/mL in MOPS 25 mM, NaCl 115 mM at pH 7.4, Shown as an average spectrum of three repeats pre-processed using EMSC as described in supplementary methods (Data Pre-processing and Chemometrics).

Table S8. Tentative Raman assignments for D-(+)-Glucose, Lysozyme and Albumin taken from literature.[9–15]

Raman Spectroscopy assignment			
Wavenumber (cm^{-1})	**D-(+)-Glucose**	**Wavenumber (cm^{-1})**	**Lysozyme/Albumin**
1637	Water peak	1653	Amide I α-helix, C=O
1461	CH_2	1618	Tyr, Trp, Phe
1358	CH_2	1552	Indole ring Trp
1332	CH_2	1452	C-H def
1123	C-N, C-C, C-O, C-OH	1422	Trp
1075		1334	Trp C-H (Indole breathing)
1034		1310	Amide III
916	C-CH, C-CO	1247	Amide III
844	C-C	1202	Phe
587	C-CC, C-CO, O-CO	1175	C–C,C-O
541		1106	C–C, C-O
511		1042	C–C, C-O
		1008	Phe
		929	N-C-C (protein backbone)/ Amide IV
		896	Trp
		850	Tyr
		811	Tyr
		757	Trp
		508	S-S

References

1. M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch and C. V. Robinson, *Anal. Chem.*, 2015, **87**, 4370–4376.

2. H. Hernández and C. V Robinson, *Nat. Protoc.*, 2007, **2**, 715–726.

3. H. Martens and E. Stark, *J. Pharm. Biomed. Anal.*, 1991, **9**, 625–635.

4. T. O'haver, A Pragmatic Introduction to Signal Processing with applications in Chemical Analysis, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.608.8863&rep=rep1&type=pdf, (accessed 9 July 2019).

5. A. Barth, *Biochim. Biophys. Acta*, 2007, **1767**, 1073–1101.

6. S. Barsberg, *J. Phys. Chem*, 2010, **114**, 11703–11708.

7. K. Fu, K. Griebenow, L. Hsieh, A. M. Klibanov and Robert Langera, *J. Control. Release*, 1999, **58**, 357–366.

8. N. B. Colthup, L. H. Daly, S. E. Wiberley, N. B. Colthup, L. H. Daly and S. E. Wiberley, *Major septra-structure Correlations by Spectral Regions*, Academic Press, 1990.

9. E. Wiercigroch, E. Szafraniec, K. Czamara, M. Z. Pacia, K. Majzner, K. Kochan, A. Kaczor, M. Baranska and K. Malek, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, 2017, **185**, 317–335.

10. N. C. Maiti, M. M. Apetri, M. G. Zagorski, P. R. Carey and V. E. Anderson, 2004, **126**, 2399–2408.

11. Z. Wen, *J. Pharm. Sci.*, 2007, **96**, 2861–2878.

12. A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk and M. Baranska, *J. Raman Spectrosc.*, 2013, **44**, 1061–1076.

13. R. Tuma, *J. Raman Spectrosc. Spectrosc.*, 2005, **36**, 307–319.

14. G. J. Thomas, *Biopolymers*, 2002, **67**, 214–225.

15. R. P. Kengne-Momo, P. Daniel, F. Lagarde, Y. L. Jeyachandran, J. F. Pilard, M. J. Durand-Thouand and G. Thouand, *Int. J. Spectrosc.*, 2012, **2012**, 1–7.