Supplementary Material (ESI) for Chemical Communications

Supporting Information :

Incorporation of N-Heterocyclic Cations into Proteins with a Highly Directed Chemical Modification

Nicola McMillan[1], Louise V. Smith[1], Jesus M. de la Fuente[2], Alexis D. C. Parenty[1], Nikolaj Gadegaard[3], Andrew R. Pitt[4], Katrina M. Thomson[4] Cameron MacKenzie[1], Sharon M. Kelly[4] and Leroy Cronin[1]*.

[1]WestCHEM, Department of Chemistry, University of Glasgow, Glasgow, G12 8QQ, U.K. [2]Laboratory of Glyconanotechnology, IIQ-CSIC Sevilla 41092, Spain. [3]Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow, G12 8LT, U.K. [4]Department of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, Joseph Black Building, University of Glasgow, Glasgow, G12 8QQ, U.K.

[1]Email: L.Cronin@chem.gla.ac.uk

Synthesis of modified proteins:

Phenanthridinium bromide is used as a reagent and oxidising agent in the reaction with lysine residues and so 2 equivalents are needed for very primary amine present, in terms of the thiolate modification phenanthridinium bromide is only the reactant and so only 1 equivalent is required.

Synthesis of modified BSA:

Bovine Serum Albumin was purchased from Sigma. A 7.5% $NaHCO_3$ solution (3g in 40mL H_2O) was prepared and to this BSA (500mg; 7.58 x 10^{-6}mol) and 2-bromoethyl phenanthridinum bromide (0.028mg; 7.6 x 10^{-5}mol) added. The reaction was stirred at 0˚C, under N_2, for 2 days. Crude product was purified by dialysis and filtration using Sigma cellulose tubing; size 23mm x 15mm, with the water changed every 3 hrs. After 1 day dialysis was stopped and pure product obtained by freeze drying.

Synthesis of modified LSZ:

Chicken egg white lysozyme was purchased from Sigma. This reaction targets only the 6 lysine residues present on lysozyme. A 7.5% $NaHCO_3$ solution (3g in 40mL H_2O) was prepared and to this LSZ (200mg; 1.4 x 10^{-5}mol) and 2-bromoethyl phenanthridinium bromide (0.062g; 1.7 x 10^{-4}mol, 12 equivalents). The reaction was stirred at 0˚C, under N_2, for 2 days. Crude product was purified by dialysis and filtration using Sigma cellulose tubing; size 23mm x 15mm, with the water changed every 3 hrs. After 1 day dialysis was stopped and pure product obtained by freeze drying.

The reaction was carried out using a range of equivalents of phenanthridinium bromide to demonstrate the selectivity of the reaction as seen by mass spectrometry (Figure 1). Using 7 equivalents of starting material results in a statistical distribution of modifications and a large excess of 68 equivalents demonstrates that regardless of the excess only the 6 lysine residues present are modified.

Mass Spectrometry

Samples were desalted on a C18 ZipTip using standard procedures, eluting with 2.0-2.5uL of 50% acetonitrile/water, 0.5% formic acid. The sample was loaded directly in to a single shot nanospray tip (Proxeon), and the tip mounted on a nanospray source (Protana) on a QStar Pulsar i running AnalystQS 1.1 (Applied Biosystems). Data was collected in the 800-4000Da range until an acceptable signal to noise was obtained (1-5 minutes). The raw data was summed across the whole run and deconvoluted using Analyst BioTools, selecting a mass range of 14-16kDa, a s/n of 5 and a step size of 0.5 Da, running 20 iterations.

Lysozyme Mass Spectrometry

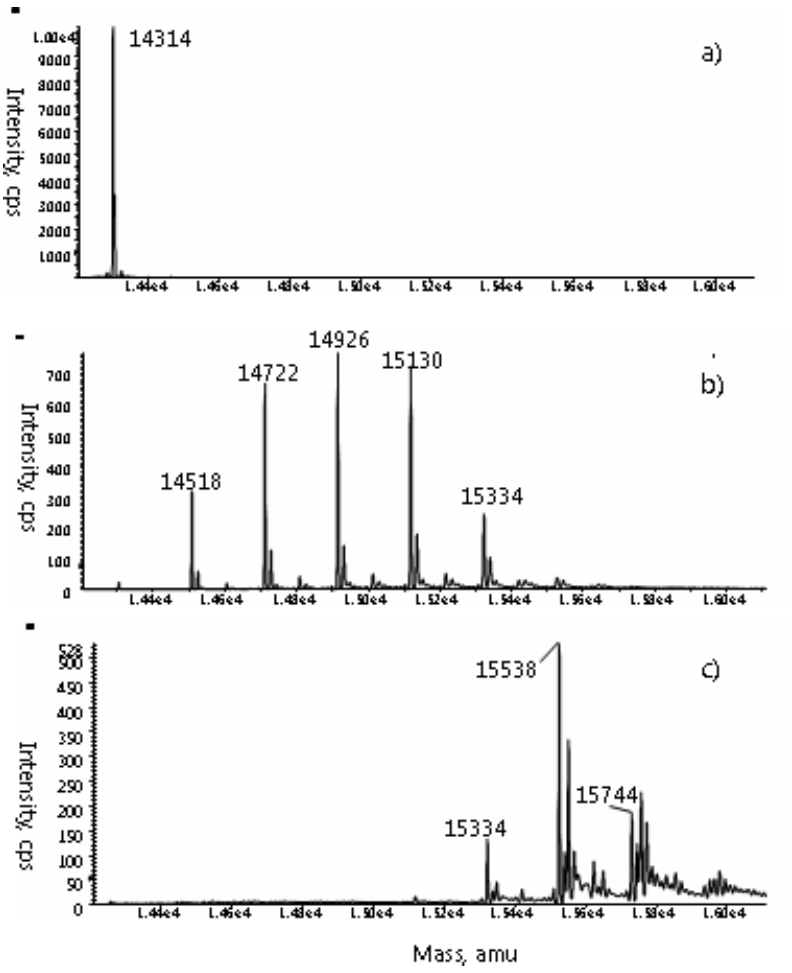


Figure 1. ESI- of lysozyme (a – native) modified with increasing equivalents of reagent (b-c). In a statistical distribution of 0 to 6 modifications (204 D increase per cyclised modification), using 7 equivalents of 2-bromoethyl phenanthridinum bromide was observed whereas c. shows 6 modifications (15538) resulting from conditions which included a large excess of 68 equivalents of 2-bromoethyl phenanthridinum bromide including a peak at 15744 which probably corresponds to 6 modifications plus reaction with the N-terminus which will not undergo cyclisation for steric reason (206 D difference). Thus this shows that the modification is dose dependent, and the preponderance for the modification of 6 residues on treatment with excess reagent correlates well with the protein containing 6 lysines. Minor associated peaks are associated with cation adducts etc.

BSA mass Spectrometry

Matrix was prepared by making a saturated solution of alpha-cyano-4-hydroxycinnamic acid (Fluka) in 1:1 acetonitrile water. Small quantities of lyopholised proteins and conjugates were dissolved in deionised water and mixed with matrix in 1:4 ratios. 1uL drops of this solution were applied to a stainless steel MALDI plate (Applied Biosystems) and allowed to dry. Analysis was on a Voyager DE pro MALDI-TOF (Applied Biosystems) operating in linear mode, and close external calibration with the Sequazyme mass standards kit (Applied Biosystems) was performed before analysing each spot. Mass spectra were collected in the positive ion mode at an accelerating voltage of 25 kV and a delay time of 320 ns. All spectra were averaged over 300 shots. Instrument control and data analysis was carried out with the Voyager and DataExplorer software. Spectra were smoothed with a Guasian window 0f 101 points and mass determined by centroiding the top 50% of the peak

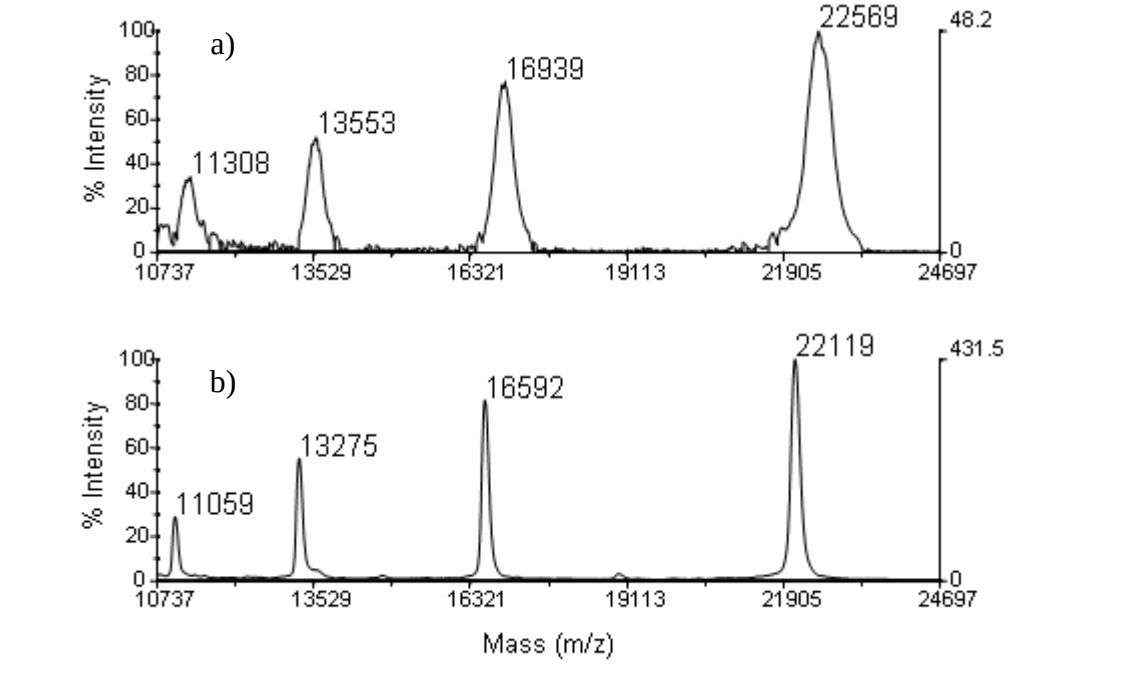


Figure 2. MALDI-ToF analysis of a. DIP modified BSA, and b. native showing 3+ to 6+ ions. Modified protein shows an average shift to higher mass of 1350, equivalent to approximately 7 DIP adducts. The significant broadening of the peaks suggests a statistical distribution of modifications around this average.

Sequence analysis

To identify sites of modification, minimally modified (average 7 modifications by mass spectrometry) and unmodified BSA (35 cysteines and 59 lysines) were analyzed by peptide mass fingerprinting. The samples were treated with dithiothreitol to reduce disulfides and then free cysteines blocked by treatment with iodoacetamide. The samples were digested with trypsin and peptides mass fingerprints obtained by MALDI-ToF MS on a 4700 Proteomic Analyser (Applied Biosystems). The raw data was searched using MASCOT (Matrix Science) with fixed modification: carbamidomethyl, variable modifications: oxidation of methionine and the uncyclised ethylphenanthridinium adduct or DIP modification on lysine, searching the SwissProt mammalian database at 100ppm mass accuracy. Only peaks appearing in the modified, but not in the unmodified sample, were considered significant. Peptide were identified with DIP modifications on lysine 140, 228, and 235, and cysteine 58, 147 and either 581, 582 or 590. Other peptides were identified where modifications were present but could not be localized to lysine or cysteine residues. These included two modifications on a peptide containing K299, K304, C301 and C302, and a one on a peptide containing K88 and C86.

a)

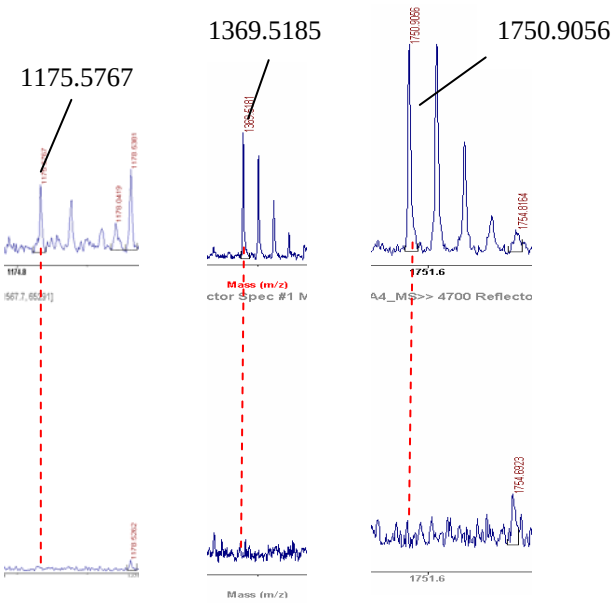
Matched peptides shown in **Bold Red**

```
  1 MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
 51 FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
101 VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF
151 KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC
201 LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
251 FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL
351 GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
451 RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC
501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
551 ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
601 STQTALA
```

b)

Start-End	Observed	Mr (expt)	Mr (calc)	Delta	Miss	Sequence
223-232	1369.5181	1368.5108	1368.6648	-0.1540	1	R.CASIQKFGER.A Phenan (K)
233-241	1175.5767	1174.5694	1174.6651	-0.0957	1	R.ALKAWSVAR.L Phenan (K)

c)



139-151	1750.9056	1749.8983	1749.8463	0.0547	0	K.LKPDPNTLCDEFK.A Phenan (K)

Figure 3. MALDI-ToF analysis of DIP modification. MASCOT analysis of data from BSA. a) Sequence coverage shown in bold red (86%), b) three of the peptides identified as containing DIP modified lysine in MASCOT analysis, c) corresponding regions of MALDI-ToF MS spectrum showing presence in modified (top) and absence in unmodified (bottom) BSA digests respectively.

Circular Dichroism

Solutions of lysozyme and BSA (modified and native) were prepared in 7.5% sodium hydrogen carbonate and concentrations analysed using the Bradford Assay. Samples were investigated using a

Jasco J-810 spectropolarimeter and curves corrected for protein concentration, cell path length, mean residue weight and baseline to remove effects from the buffer and additives (see Kelly, S.M.; Jess, T.J.; Price, N.C.; *Biochimica et Biophysica Acta*, **2005**, 119-139).

LSZ

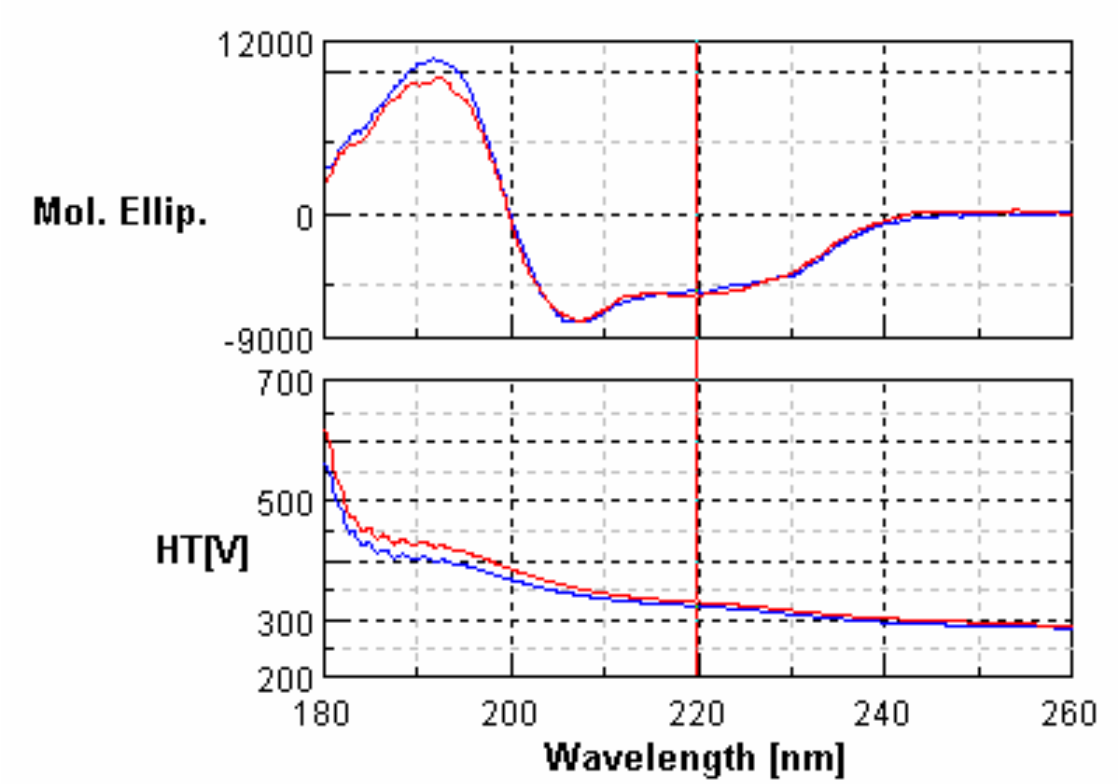


Figure 4. Far uv CD spectra of unmodified lysozyme (blue) and modified lysozyme (red). This indicates the modified lysozyme has retained its secondary structure compared with native lysozyme, as the trace remains almost consistent throughout.

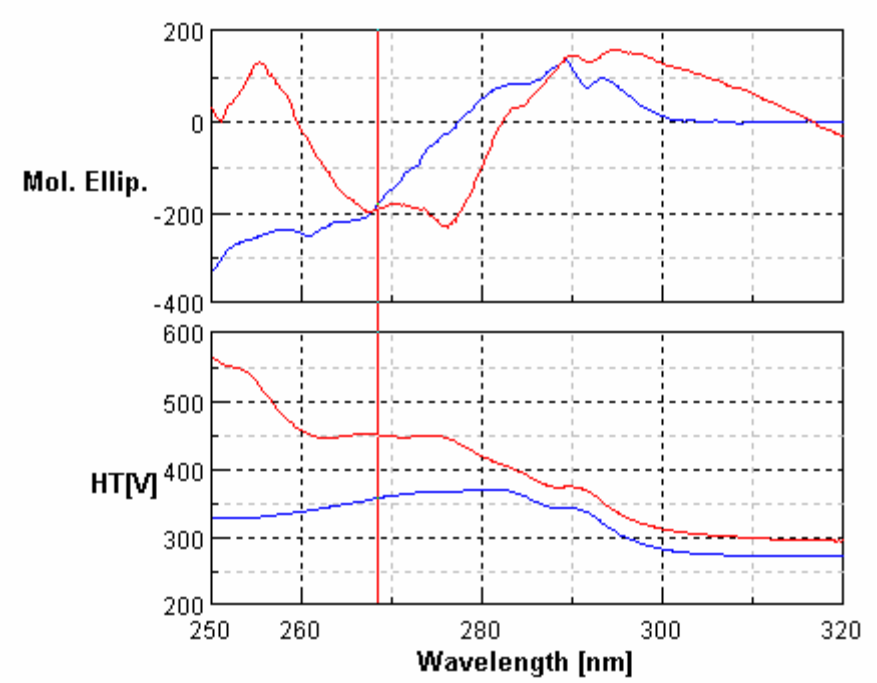


Figure 5. Near uv CD spectra of unmodified lysozyme (blue) and modified lysozyme(red). The trace of the modified lysozyme is not consistent with the near uv trace of the native protein. This is as a result of the interference of the DIP chromophore at this region of the spectrum.

BSA

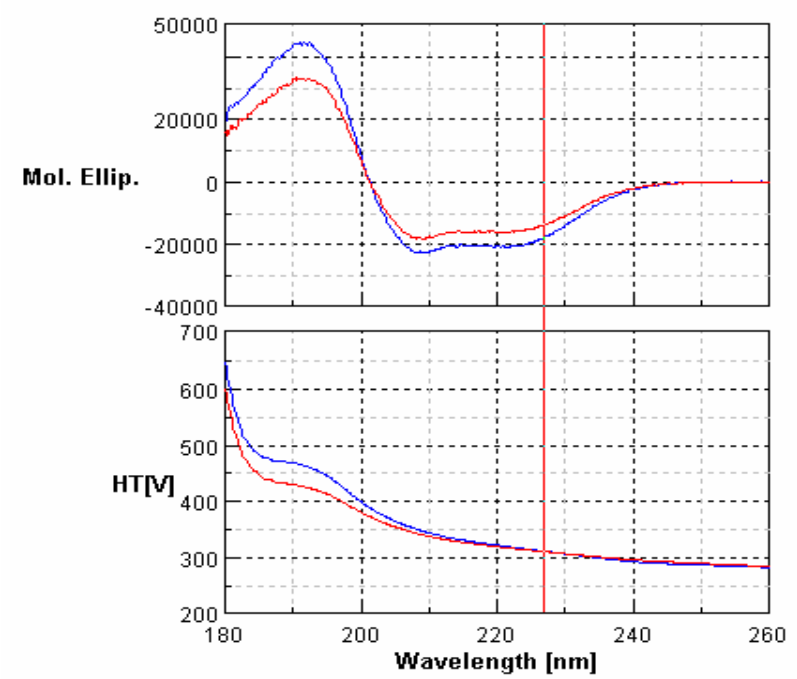


Figure 6. Far uv CD spectra of unmodified BSA (blue) and modified BSA (red). This indicates the modified BSA has retained its secondary structure compared with native BSA, as the trace remains almost consistent throughout.

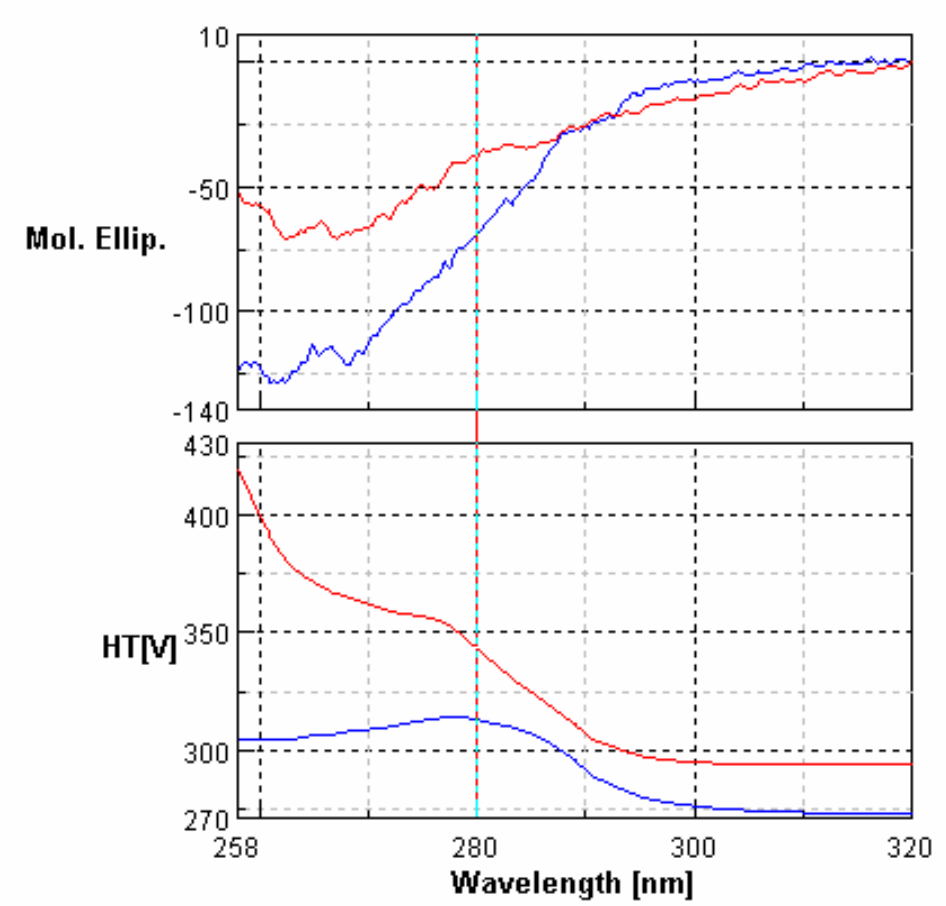


Figure 7. Near uv CD spectra of unmodified BSA (blue) and modified BSA (red). The trace of the modified BSA is not consistent with the near uv trace of the native protein. This is as a result of the interference of the DIP chromophore at this region of the spectrum.

Atomic Force Microscopy

AFM images were taken using Digital Instruments Multimode IIIa AFM in a tapping mode setting and the samples were prepared by applying a 1mg/ml solution of each protein sample to a silicon substrate drop-wise that had been sonicated in ethanol. The substrates were then dried overnight to remove the remaining solvent. Control experiments were also prepared with the unmodified proteins: 1mg/ml solutions, in water, of each protein were prepared and added to the silicon wafers in the same manner as above, and each analysed by AFM.

Lysozyme AFM

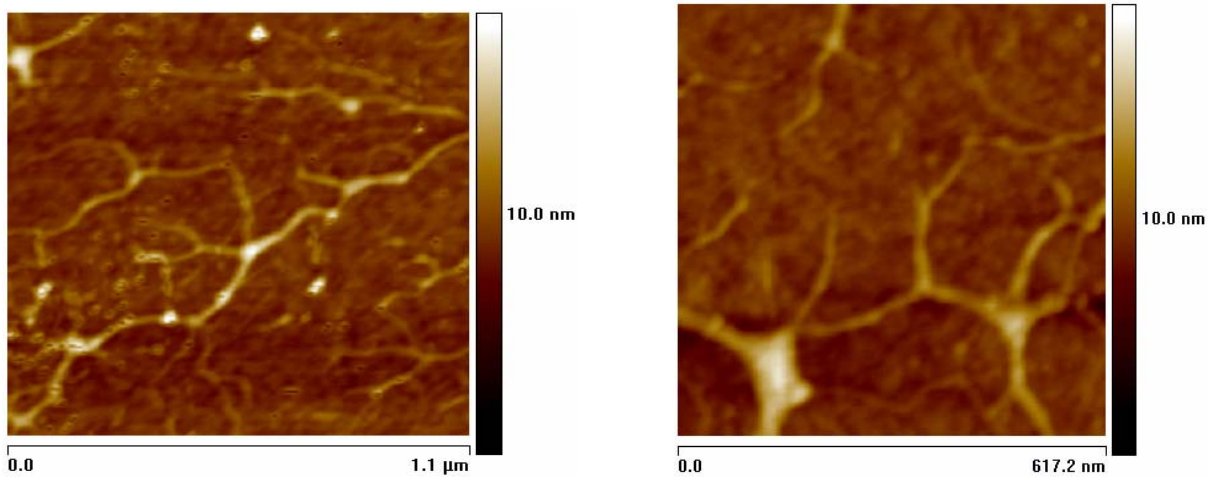


Figure 8. Tapping mode images of 1mg/ml lysozyme-DIP complex on silicon. Long vein like networks can be seen after modification compared with the flat uniform surface that can be seen from the unmodified protein.

BSA AFM

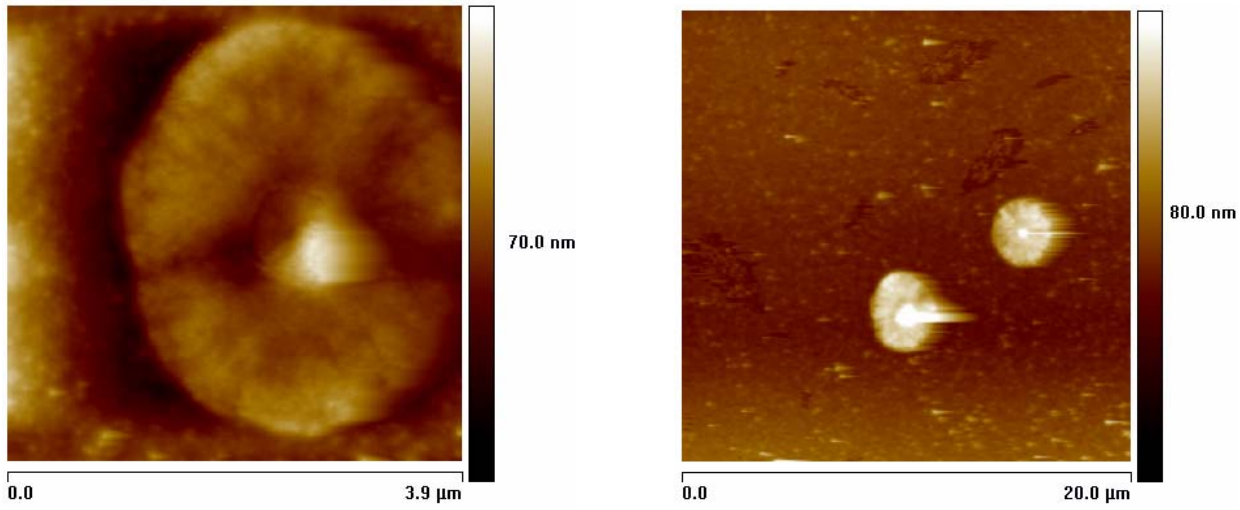


Figure 9. Tapping mode AFM images of 1mg/ml BSA – DIP complex on a silicon surface, again the large circular moiety with the central compartment can be seen, in stark contrast to the undecorated surface of the unmodified protein.