Incorporation of N-heterocyclic cations into proteins with a highly directed chemical modification†

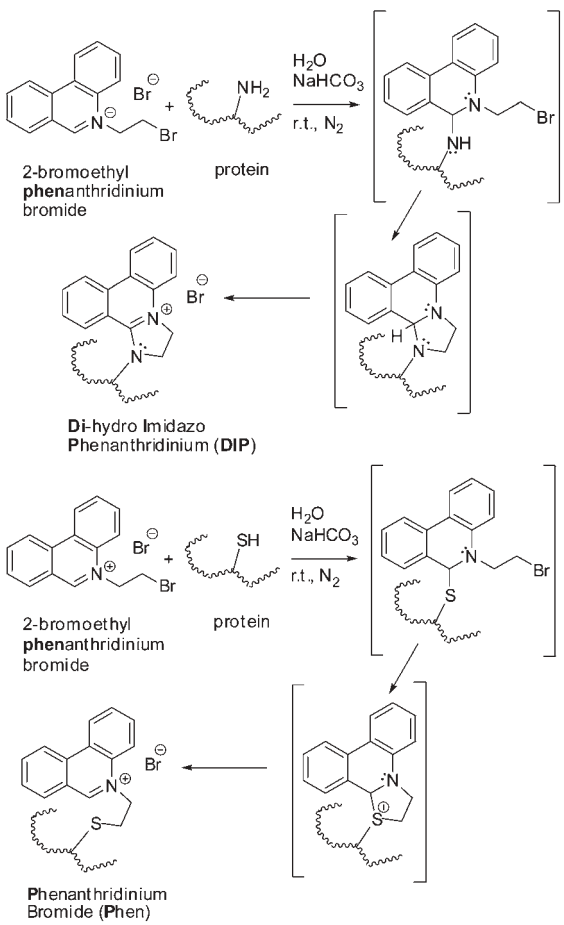
Nicola McMillan,[a] Louise V. Smith,[a] Jesus M. de la Fuente,[b] Alexis D. C. Parenty,[a] Nikolaj Gadegaard,[c] Andrew R. Pitt,[d] Katrina Thomson,[d] Cameron MacKenzie,[a] Sharon M. Kelly[d] and Leroy Cronin*[a]

Received (in Cambridge, UK) 19th February 2007, Accepted 8th May 2007
First published as an Advance Article on the web 29th May 2007
DOI: 10.1039/b702575k

N-Heterocyclic cations are incorporated into proteins using 5-(2-bromoethyl)phenanthridinium bromide, which selectively reacts with either cysteine or lysine residues, resulting in ethylphenanthridinium (Phen) or highly stable cyclised dihydro-imidazo-phenanthridinium (DIP) adducts respectively; these modifications have been found to manipulate the observed structure of lysozyme and bovine serum albumin by AFM.

Unnatural site directed protein modification has great value due to the many potential biological and biomedical applications.[1–4] For instance, proteins have been modified for uses as fluorophores,[1] cofactors for catalysis,[2] electrochemical groups for detection of ligand binding in biosensors,[3] and for immobilisation on solid surfaces.[4] Although many classes of protein modification exist within nature, pre-, co- and post-translational modification[5] routes to the introduction of unnatural functional groups and new scaffolds would be valuable in the research field of protein engineering, and in the development of post-translational modification mimics,[6] but examples of these are limited.[7] Despite many developments in synthetic methodology in recent years, only a few methods are available that allow the *direct* and unnatural chemical modification of proteins selectively. A classic example uses ketone–hydrazine chemistry[8] to attach non-natural molecules to native proteins. Modification of proteins with FITC (fluorescein-5′-isothiocyanate) is an established method used to generate fluorescent proteins that can be detected within the cell.[9] Tyrosine residues can also be modified by the Mannich reaction[10] and diazonium coupling,[10] and recently Suzuki cross-coupling has been used to give protein modification with retention of native structure and function.[2]

We are interested in the direct modification of proteins that can yield high selectivity, retain the secondary structure and provide a route to manipulate the tertiary structure of the protein in solution to present a polyvalent system with new properties, as well as using the protein as a biologically compatible carrier of potential drugs and diagnostic probes. To achieve this, we opted to utilise a simple 'one-pot–three-step' reaction system that targets primary amine residues producing a novel, cyclised dihydro-imidazo-phenanthridinium (DIP) heterocyclic framework or other nucleophiles such as thiolates, which yield an uncyclised ethylphenanthridinium adduct.[11,12] Here we aimed to examine the reaction of proteins with 5-(2-bromoethyl)phenanthridinium bromide since we envisaged that this would selectively target cysteine and lysine residues specifically, yielding ethylphenanthridinium (Phen) and dihydro-imidazo-phenanthridinium (DIP) frameworks in each case



Scheme 1 Description of the reaction and basic mechanism of the reaction yielding DIP protein adducts by reaction with primary amines (top)[11] and Phen protein adducts by reaction with thiolates (bottom).[12]

[a]*WestCHEM, Department of Chemistry, The University of Glasgow, University Avenue, Glasgow, UK G12 8QQ. E-mail: L.Cronin@chem.gla.ac.uk; Fax: +44-141-330-4888; Tel: +44-141-330-6650*
[b]*Instituto de Nanociencia de Aragon, Universidad de Zaragoza. C/ Pedro Cerbuna 12, Zaragoza 50009, Spain*
[c]*Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow, UK G12 8LT*
[d]*Department of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, Joseph Black Building, University of Glasgow, Glasgow, UK G12 8QQ*

† Electronic supplementary information (ESI) available: Full experimental preparations, MS details, circular dichroism and AFM data. See DOI: 10.1039/b702575k

(Scheme 1). This is because previous work strongly indicates that the precursor is highly reactive with primary amines but sterically hindered non-primary amines are very much less reactive.[11,12]

Such a system would not only be water soluble, but could be used to exploit the protein as a "biological carrier", through protein–cell membrane interactions, and to direct Phen/DIP molecules to specific cell types. Herein we describe the covalent modification of chicken egg white lysozyme (LSZ) and bovine serum albumin (BSA) using the one-pot methodology described in Scheme 1.‡[11]

Previous studies have described the modification of lysozyme by acetylation using acetic anhydride[13] and, more commonly, glycation by heat treatment/incubation of the protein sample with glucose or fructose.[14] In this instance the targets for modification are the 6 lysine residues; however, their reactivities have previously been shown to be primarily dependent on the relative surface accessibilities.[13] Importantly, only 4 of the 6 total number of lysine residues present in lysozyme have been modified previously and this required several chemical steps, with double glycosylation most probable.[9,14,15] However, in this work we have been able to access and modify all 6 residues in one step, depending on the overall concentration of the reagents, see Fig. 1.

To react with all the residues, the synthetic methodology previously reported using DMF[11] had to be altered so that a weakly basic aqueous solution (pH 9) compatible with most proteins (Scheme 1) was used. An excess of 2 equivalents of 5-(2-bromoethyl)phenanthridinium bromide for every amine group present was used in the case of LSZ, enabling maximum modification of lysine residues. Since only reaction between 5-(2-bromoethyl)phenanthridinium bromide and a primary amine (*i.e.* lysine) will result in cyclisation, and a mass difference of 204 Da, rather than 206 Da for the uncyclised product, both the native and modified proteins were studied using electrospray mass spectrometry and deconvoluted using Analyst BioTools. These studies of the native and modified lysozyme are shown in Fig. 1 and demonstrate that the native protein can be modified in steps of 204 Da up to six times. The observation of the 204 mass difference for the 6 lysine residues in lysozyme indicates that the reaction is totally selective for lysine with this protein. The observation that all 6 lysine residues have been aromatically modified by the cyclisation reaction is interesting since it is very hard to modify lysine K96 due to its low accessibility caused by H-bonding to the carbonyl group of H-15.[12] Further, only when 68 equivalents of the reagent are used is the uncyclised modification of the protein observed. This presumably indicates reaction at the N-terminus (and cyclisation is prevented on steric grounds).[11]

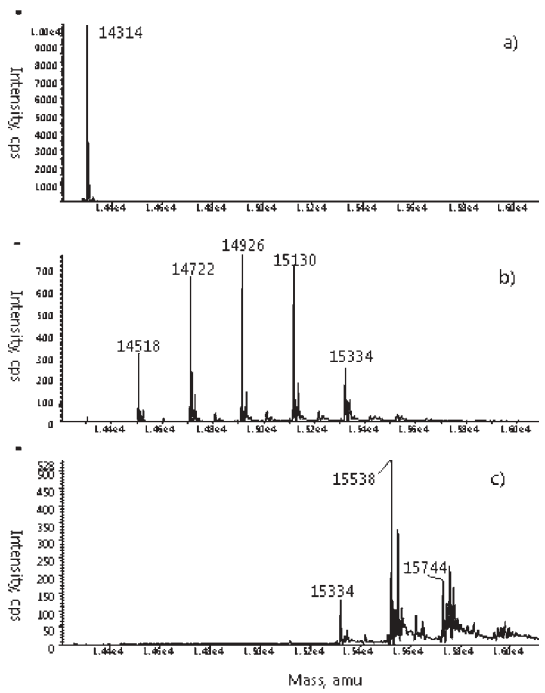


Fig. 1 (a–c) ESI$^-$ of lysozyme (a: native) modified with increasing equivalents of reagent (b–c). A statistical distribution of 0 to 6 modifications (204 D increase per cyclised modification) was observed using 7 equivalents of 5-(2-bromoethyl)phenanthridinium bromide, whereas (c) shows 6 modifications (15 538) resulting from conditions that included a large excess of 68 equivalents of 5-(2-bromoethyl)phenanthridinium bromide, including a peak at 15 744 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 peaks are associated with cation adducts *etc.*

In the case of BSA, which contains 59 lysine and 35 cysteine residues, we set out to explore the relative reactivity in a preliminary study by modifying less than 10 residues per protein, since attempts to target a higher number caused the protein to precipitate; presumably the most accessible surface residues are modified. Matrix assisted laser desorption ionisation (MALDI) for BSA (see ESI†), shows the triply charged molecular ion of the unmodified protein (66 354 Da) at *m/z* 22 119 and modified protein (67 704 Da) at *m/z* 22 569. Analysis of all peaks in the spectrum shows that the modified and unmodified proteins differ by, on average, 1350 Da, corresponding to *ca.* 7 modifications. Subsequent analysis to identify sites of modification was then undertaken and the modified and unmodified proteins were analysed by trypsin digest and peptide mass fingerprinting, which revealed several lysine (140, 228, and 235) and cysteine (58, 147 and either 581, 582 or 590) sites that are unambiguously modified, demonstrating the selectivity of the reaction.

We subsequently analysed the effect of the modifications on the structure of the proteins using circular dichroism (CD). CD analysis of both compounds (see ESI†) confirms that the proteins have been modified: the DIP moiety interferes with the near-UV CD spectra whereas the far-UV remain consistent with the native protein. This comparison allows us to suggest that the secondary structure remains largely unaltered. This is particularly exciting in the case of BSA, which is a "soft" globular protein and can easily change its internal structure.[16]

AFM studies of the proteins adsorbed onto a silicon substrate showed distinct variations before and after chemical modification of the proteins. Tapping mode AFM images show that the contrast between modified and unmodified BSA (Fig. 2-top) is striking, with only small globular features around 100–120 nm in size visible for unmodified BSA.

The difference between the unmodified and modified BSA samples is interesting since the samples containing modified BSA show large circular objects uniformly covering the silicon surface. These objects possess similar dimensions, with an inner region seen to protrude from an outer shell, giving the structure a

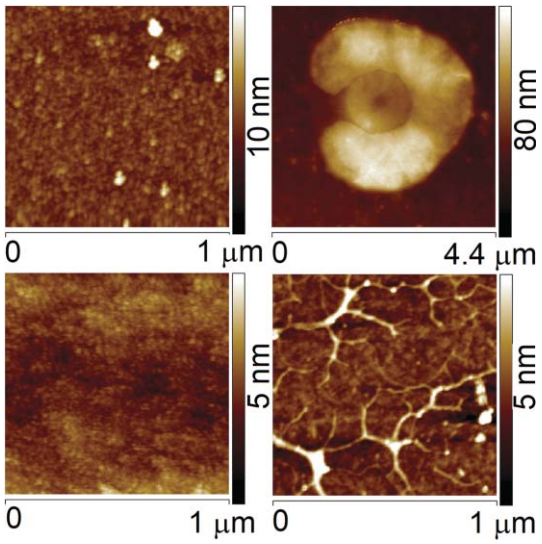


Fig. 2 Top: left: height AFM images of native BSA; right: height AFM images of modified BSA. Bottom: left: height AFM images of native LSZ; right: height AFM images of modified LSZ.

"mushroom"-like appearance. It would appear that the architectures are formed due to inter-protein interactions (due to structural changes induced by modification or even between surface modified lysine residues) and further work is ongoing to investigate this. The differences between the AFM images of modified and native LSZ (Fig. 2-bottom) are also striking.

Images of unmodified LSZ show a flat, uniform surface with no distinct features, whereas LSZ modified with DIP shows long, non-uniform, fibril-like networks covering the surface. These features are reminiscent of the lysozyme amyloid oligomers and fibrils that can be observed to be formed by lysozyme under acidic conditions or as the result of mutations. Therefore, the observation of fibril-like networks here is interesting since fibrillar aggregates can be extremely toxic *in vivo*.[17]

Clear differences can be seen in the patterns made by deposition of both proteins on the surface after modification compared to those of the native protein and these are interesting since a relatively minor chemical modification is causing massive structural changes when the proteins are deposited onto a surface.

We attribute these differences due to changes in the tertiary structure of the proteins since our CD studies indicate strongly that the secondary structures of both proteins are unchanged after modification.

In summary, we have developed a one-pot, three-step methodology for use in the synthesis of modified proteins, which targets lysine residues *via* a cyclisation reaction and cysteine residues *via* an addition to give DIP and Phen-based heterocyclic cations respectively. Compared to conventional methods, the synthesis is undemanding and utilises biocompatible conditions to obtain pure protein by straightforward dialysis. Further, mass spectroscopic studies clearly demonstrate that the reaction of LSZ with 5-(2-bromoethyl)phenanthridinium bromide results in the incorporation of all 6 lysine residues in the imidazole moiety of DIP, and BSA with 5-(2-bromoethyl)phenanthridinium bromide gives 7 modifications spread over lysine and cysteine sites. Importantly, the reaction is selective for both lysine and cysteine but crucially the modification made to each lysine or cysteine is chemically different, giving a different class of heterocyclic cation. Also, the indirect effects of these modifications can be inferred from AFM studies, showing dramatic differences between the modified and unmodified proteins. This new methodology represents an extremely efficient route to target amine and thiolate groups. The resulting products are highly water-soluble and, in the case of the lysine modifications, represent a rare example of direct 'one-pot' post-translational synthetic modification mimics affected by a cyclisation process.

In further work we will examine the extension of this methodology to other proteins, assess the activity of the modified proteins, as well as structural studies to examine how the DIP-protein modification causes changes in the tertiary structure of the protein in more detail.

Notes and references

‡ *Synthesis of modified LSZ:* chicken egg white was purchased from Sigma. A 7.5% $NaHCO_3$ solution (3 g in 40 ml H_2O) was prepared and, to this, LSZ (200 mg; 1.4×10^{-5} mol) and 5-(2-bromoethyl)phenanthridinium bromide 0.062 g; 1.7×10^{-4} mol) were added. The reaction was stirred at 0 °C, under N_2, for 2 days. The crude product was purified by dialysis and filtration using Sigma cellulose tubing; size 23 mm × 15 mm, with the water changed every 3 h. After 1 day, dialysis was stopped and pure product obtained by freeze drying. *Synthesis of modified BSA:* bovine serum albumin was purchased from Sigma. A 7.5% $NaHCO_3$ solution (3 g in 40 ml H_2O) was prepared and, to this, BSA (500 mg; 7.58×10^{-6} mol) and 5-(2-bromoethyl)phenanthridinium bromide (0.028 mg; 7.6×10^{-5} mol) were added. The reaction was stirred at 0 °C, under N_2, for 2 days. The crude product was purified by dialysis and filtration using Sigma cellulose tubing; size 23 mm × 15 mm, with the water changed every 3 h. After 1 day, dialysis was stopped and pure product obtained by freeze drying. Both proteins were examined using MALDI, CD, and AFM; see ESI for full details.†

1 E. T. Kaiser, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 913; C. M. Tann, D. Qi and M. D. Distefano, *Curr. Opin. Chem. Biol.*, 2001, **5**, 696.
2 D. E. Benson, D. W. Conrad, R. M. de Lorimier, S. A. Trammell and H. W. Hellinga, *Science*, 2001, **293**, 1641.
3 P. L. Domen, J. R. Nevens, A. K. Mallia, G. T. Hermanson and D. C. Klenk, *J. Chromatogr.*, 1990, **510**, 293.
4 J. J. Smith, D. W. Conrad, M. J. Cuneo and H. W. Hellinga, *Protein Sci.*, 2005, **14**, 64.
5 A. Ojida, H. Tsutsumi, N. Kasagi and I. Hamachi, *Tetrahedron Lett.*, 2005, **46**, 3301.
6 B. G. Davis, *Science*, 2004, **303**, 482.
7 J. S. Marvin, E. E. Corcoran and R. O'Kennedy, *Analyst*, 1997, **94**, 4366.
8 J. M. Hooker, E. W. Kovacs and M. B. Francis, *J. Am. Chem. Soc.*, 2004, **126**, 3718; N. S. Joshi, L. R. Whitaker and M. B. Francis, *J. Am. Chem. Soc.*, 2004, **126**, 15942.
9 V. Schnaible and M. Przybylski, *Bioconjugate Chem.*, 1999, **10**, 861.
10 R. E. Schweppe, C. E. Haydon, T. S. Lewis, K. A. Resing and N. G. Ahn, *Acc. Chem. Res.*, 2003, **36**, 453.
11 A. D. C. Parenty, L. V. Smith, A. L. Pickering, D.-L. Long and L. Cronin, *J. Org. Chem.*, 2004, **69**, 5934.
12 A. D. C. Parenty, L. V. Smith and L. Cronin, *Tetrahedron*, 2005, **61**, 8410.
13 F. K. Yeboah, I. Alli, V. A. Yaylayan, K. Yasuo, S. F. Chowdhury and E. O. Purisima, *Bioconjugate Chem.*, 2004, **15**, 27.
14 W. Fiedler, C. Borchers, M. Macht, S.-O. Deininger and M. Pryzybylski, *Bioconjugate Chem.*, 1998, **9**, 236.
15 Y. Shu, S. Maki, S. Nakamura and A. Kato, *J. Agric. Food Chem.*, 1998, **46**, 2433.
16 Y. M. Efimova, B. Wierczinski, S. Haemers and A. A. van Well, *J. Radioanal. Nucl. Chem.*, 2005, **264**, 91.
17 A. L. Gharibyan, V. Zamotin, K. Yanamandra, O. S. Moskaleva, B. A. Margulis, I. A. Kostanyan and L. A. Morozova-Roche, *J. Mol. Biol.*, 2007, **365**, 1337.