A trisulfide-linked glycoprotein†

Gonçalo J. L. Bernardes,[a] Justin P. Marston,[b] Andrei S. Batsanov,[b] Judith A. K. Howard[b] and Benjamin G. Davis*[a]

Received (in Cambridge, UK) 3rd May 2007, Accepted 11th June 2007
First published as an Advance Article on the web 4th July 2007
DOI: 10.1039/b706682a

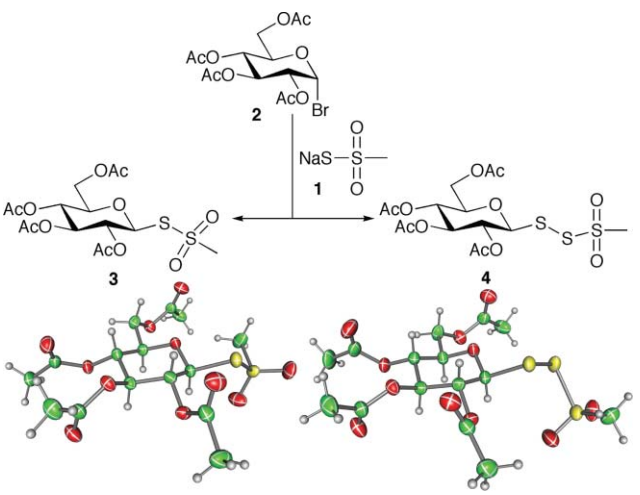
The first member of a novel class of chemoselective reagents, glycosyl methanedithiosulfonates, has been synthesized, identified and employed in the first examples of chemical, site-selective construction of a trisulfide-modified protein with complete conversion.

Trisulfides are rare but important naturally occurring linkages in biomolecules with functions ranging from subtle modulation of conformation[1–3] and stability in proteins,[4–8] such as human superoxide dismutase[4,5] and human growth hormone[6,7] and peptides,[9–11] through to enediyne antitumor antibiotics.[12,13] The trisulfide motif can convey intriguing biological properties such as fine-tuning of receptor interaction[14] in hormones such as vasopressin[14] and oxytocin[10] through to the trisulfide bridge 'trigger' that relieves strain and ultimately initiates[15] radical-mediated site-specific double-stranded DNA cleavage[16] in calicheamicins and esperamicins.[12,13]

Glycosylation is the most diverse of all post-translational modifications in Nature.[17–19] It is estimated that over 50% of all mammalian proteins carry carbohydrate appendages,[20] which affect biopharmacological properties, stability and signalling functions of the proteins.[21] As glycosylation is not under direct genetic control, glycoproteins tend to exist as complex mixtures of glycoforms;[22] proteins with the same protein backbone, but varying in the nature and sites of glycosylation. This complexity has plagued the structure–activity studies for specific glycoforms as current separation procedures do not allow for the attainment of sufficient quantities of glycoproteins for detailed studies. Chemical glycosylation of proteins has therefore been an attractive strategy for accessing single glycoform targets.[18] Various approaches exist for the precise incorporation of glycans at selected positions in protein backbones.[18] One such method relies on the attachment of carbohydrate through a disulfide linkage onto the thiol of the cysteine (Cys) sidechain;[23–29] cysteines are particularly attractive target 'tags' due to their low natural abundance, especially in their free-thiol form.[30]

Despite the availability of some methods for the formation of small molecules containing polysulfides[2,14,31,32] these are neither selective for trisulfides, complete nor compatible with or applicable to proteins as substrates. Moreover, the isolation of the small amounts of naturally-occurring protein containing, for example, Cys–CH_2SSSCH_2–Cys trisulfides is either not possible or difficult and low yielding (<1%).[4–8] To our knowledge, until now, no examples of either intermolecular protein trisulfides or trisulfide glycoproteins exist. Here we present the discovery of a novel thiol-selective modification agent for the direct and quantitative formation of protein trisulfides.

Methanethiosulfonates (MTS, **R**–S–SO_2CH_3)[33] including glycosyl methanethiosulfonates (glyco-MTS)[25] such as Glc-MTS **3** are powerful reagents for protein modification through disulfide formation. However, the use, by logical extension, of methane-*di*thiosulfonates (MDTS, **R**–S–S–SO_2CH_3) has surprisingly not been investigated, perhaps due to their apparent instability.[34] Glyco-MTS reagents are readily synthesized[23,25,35] by displacement of halide from the appropriate peracetohalohexose derivative with methanethiosulfonate, typically sodium methanethiosulfonate.‡[36] The efficiency and stereoselectivity of this displacement can be quite dramatically affected by conditions.[35] During the course of an investigation of the effect of solvent, temperature and additives upon the reaction of glucosyl bromide **2**[37] with NaMTS to form Glc-MTS **3** we observed and isolated, initially, very small amounts (<2%) of an interesting by-product, tentatively assigned as the glucosyl methane*di*thiosulfonate (Glc-MDTS) **4**, of very similar polarity to the expected product Glc-MTS **3**. Variation of conditions allowed some improvement of the yield to a workable level (13%) capable of producing sufficient Glc-MDTS (Scheme 1) using increased equivalents of NaMTS (5 equiv.) and dioxane at 70 °C for 70 h. Prolonged treatment or the use of other solvents was found to give substantially lower yields.



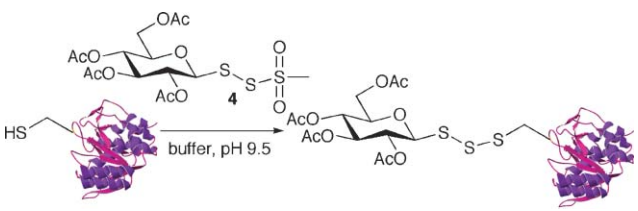
Scheme 1 Synthesis and X-ray crystal structures of glyco-MTS **3** and glyco-MDTS **4**.

[a]*Department of Chemistry, University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford, UK OX1 3TA. E-mail: Ben.Davis@chem.ox.ac.uk; Fax: +44 1865 275674; Tel: +44 1865 275652*
[b]*Department of Chemistry, University of Durham, South Road, Durham, UK DH1 3LE*

† Electronic supplementary information (ESI) available: Syntheses, characterizations, modifications X-ray crystallographic data for **3** and **4**. See DOI: 10.1039/b706682a

We are still investigating the mechanism behind the unusual formation of **4**. Although conditions that increased the yield of **4** led to lower concomitant yield of **3**, we do not observe direct conversion of **3** to **4** when re-exposed to the reaction conditions. Not only was this formation of this rare compound type surprising, the isolation of **4** is notable given the facile route for ejection of molecular sulfur from polythiosulfonates.[34] We speculate that isolation and subsequent successful use of **4** as a reagent may due to enhanced stability that is a consequence of the presence of a more hindered, secondary (glycosyl) thiyl moiety.[38] Regardless, the identities of closely related compounds **3** and **4**, was unambiguously confirmed by full characterization and X-ray crystallography§ (Scheme 1); the latter is the first known 3D structure of an MDTS. Interestingly, the MDTS moiety appears to be a powerful chirochromophore with significant ORD at 589 nm in a manner consistent with Hudson's isorotation rules:[39] the specific rotation of Na D-line light ($[\alpha]_D^{25} = -235$ ($CHCl_3$)) by β-D-**4** is in the same direction but an order of magnitude greater than that of β-D-**3** despite only differing by one sulfur atom. In the infrared spectrum of **4** two additional strong absorbances are also observed (1258, 1226 cm^{-1}) due to S–S stretching.

Next, the potential of **4** as a novel class of trisulfide-forming, protein modification reagent was investigated. In aqueous buffer at pH 9.5, **4** cleanly and completely (>95%) converted cysteine mutant protein SBL-S156C (Scheme 2 and Fig. 1) into trisulfide glycoprotein S156C–SSS–**4**, an intriguing and novel class of glycoconjugate. This first, complete and site-selective chemical construction of a trisulfide-containing protein has also allowed us to clarify aspects of the long-standing speculation surrounding proteins that contain 'excess sulfur atoms'.[4] Such proteins[4,5,40] can display unusual spectral properties due to the presence of an intramolecular Cys–CH_2SSSCH_2–Cys trisulfide. S156C–SSS–**4** displayed a broad increase in absorption over unmodified protein in the region ~250–270 nm (Fig. 2) consistent with the additional presence of an *unstrained* SS bond that gives rise to n → σ* transition but no absorption at 325 nm. These observations confirm that transitions at 325 nm observed in other trisulfide–containing proteins[4,5,40] are due primarily to the dihedral strain induced in an *intramolecular* trisulfide strained S–S[5] and not due



Scheme 2 Conditions: 70 mM CHES, 5 mM MES, 2 mM $CaCl_2$, pH 9.5, 50 equiv. of **4**,¶ protein concentration 2 mg mL^{-1}.

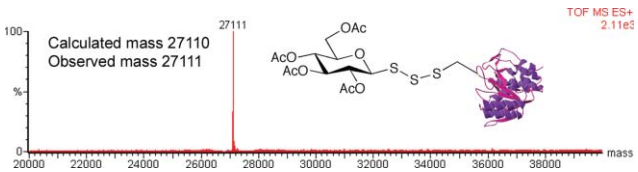


Fig. 1 ESI-MS spectrum of SBL–SSS–**4**. (Observed mass 27111, calculated mass 27110).

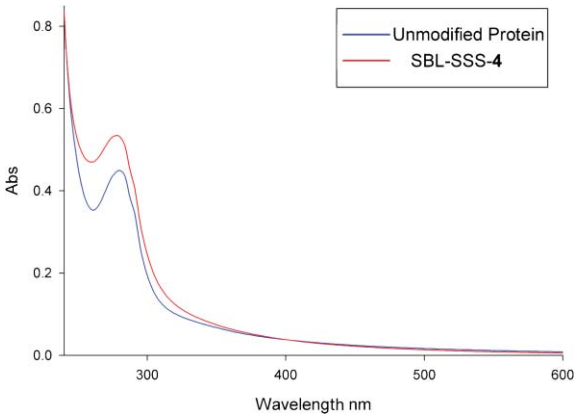


Fig. 2 UV spectrum of SBL–SSS–**4**. Both protein samples were recorded at 1 mg mL^{-1}.

simply to the presence of trisulfide.[4] We are currently exploring the spectral profiles of intra- and inter-trisulfide proteins for the direct monitoring of their presence, stability and kinetics (interaction and reaction).

In summary, this communication describes the synthesis of a novel and potent chemoselective reagent, Glc-MDTS **4** and its successful use in the first site-selective construction of a trisulfide-linked (glyco)protein. This method not only suggests strategies for accessing this class of rare natural products but allows the creation of novel (glyco)protein probes with unique reactivities and proteins with manipulable spectral signatures, the full properties of which are currently being determined.

We gratefully acknowledge the Fundação para a Ciência e a Tecnologia, Portugal for the award of a doctoral fellowship (G.J.L.B., grant BD/16997/2004), and Sander I. van Kasteren and Dr Elizabeth J. Grayson for technical support.

Notes and references

‡ Sodium methanethiosulfonate **1** for reaction with glycosyl halide was prepared through direct reaction of sodium sulfide with methanesulfonyl chloride in refluxing water.[36]

§ X-Ray data collection: Siemens SMART 1 K CCD area detector, Mo-Kα radiation ($\lambda = 0.71073$ Å), narrow-frame ω scans covering full sphere of the reciprocal space up to $2\theta \leqslant 58°$; structure solution (direct methods) and least-squares refinement against F^2 using SHELXTL 6.14 programs (Bruker AXS, Madison, Wisconsin, USA, 2003). The absolute structures were determined from anomalous dispersion of sulfur by Flack parameter (x).[41] *Crystal data*: **3**, $C_{15}H_{22}O_{11}S_2$, $M = 442.45$, $T = 108$ K, monoclinic, space group $P2_1$ (no. 4), $a = 8.790(1)$, $b = 12.774(2)$, $c = 9.607(1)$ Å, $\beta = 113.60(2)°$, $U = 988.4(2)$ Å^3, $Z = 2$, $\mu = 0.33$ mm^{-1}, 12 148 reflections (2731 unique + 2448 Friedels, $R_{int} = 0.028$), $R(F) = 0.026$ on 5038 data with $I \geqslant 2\sigma(I)$, $wR(F^2) = 0.069$ on all data, $x = 0.06(4)$. CCDC 645866.

4, $C_{15}H_{22}O_{11}S_3$, $M = 474.51$, $T = 120$ K, monoclinic, space group $P2_1$ (no. 4), $a = 5.6609(6)$, $b = 20.850(2)$, $c = 9.2785(10)$ Å, $\beta = 105.45(3)°$, $U = 1055.6(2)$ Å^3, $Z = 2$, $\mu = 0.41$ mm^{-1}, 11 166 reflections (2886 unique + 2405 Friedels, $R_{int} = 0.019$), $R(F) = 0.028$ on 5119 data with $I \geqslant 2\sigma(I)$, $wR(F^2) = 0.070$ on all data, $x = -0.07(4)$. CCDC 645867. For crystallographic data in CIF or other electronic format see DOI: 10.1039/b706682a

¶ This level of excess compares well with typical excesses used in protein modification of 500–1000 equiv. 50 equiv. was chosen here to ensure complete conversion and we are currently exploring the use of even less reagent.

1 C. Andersson, P. O. Edlund, P. Gellerfors, Y. Hansson, E. Holmberg, C. Hult, S. Johansson, J. Kordel, R. Lundin, I. Mendel-Hartvig, B. Noren, T. Wehler, G. Widmalm and J. Ohnman, *Int. J. Pept. Protein Res.*, 1996, **47**, 311–321.

2 D. A. Erlanson and J. A. Wells, *Tetrahedron Lett.*, 1998, **39**, 6799–6802.
3 E. Strandberg, J. Koerdel, R. Lundin, T. Wehler and G. Widmalm, *J. Pept. Res.*, 1997, **49**, 254–260.
4 J. M. McCord and I. Fridovich, *J. Biol. Chem.*, 1969, **244**, 6049–6055.
5 R. G. Briggs and J. A. Fee, *Biochim. Biophys. Acta*, 1978, **537**, 100–109.
6 A. M. Jespersen, T. Christensen, N. K. Lausen, P. F. Nielsen and H. H. Sorensen, *Eur. J. Biochem.*, 1994, **219**, 365–373.
7 M. K. Thomsen, B. S. Hansen, P. Nilsson, J. Nowak, P. B. Johansen, P. D. Thomsen and J. Christiansen, *Pharmacol. Toxicol.*, 1994, **74**, 351–358.
8 E. Canova-Davis, I. P. Baldonado, R. C. Chloupek, V. T. Ling, R. Gehant, K. Olson and B. L. Gillece-Castro, *Anal. Chem.*, 1996, **68**, 4044–4051.
9 M. Moutiez, G. Lippens, C. Sergheraert and A. Tartar, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 719–724.
10 L. Chen, I. Zoulikova, J. Slaninova and G. Barany, *J. Med. Chem.*, 1997, **40**, 864–876.
11 V. Windisch, F. DeLuccia, L. Duhau, F. Herman, S.-Y. Tang and M. Vuilhorgne, *J. Pharm. Sci.*, 1997, **86**, 359–364.
12 K. C. Nicolaou and W.-M. Dai, *Angew. Chem., Int. Ed. Engl.*, 1991, **30**, 1387–1416.
13 J. S. Thorson, E. L. Sievers, J. Ahlert, E. Shepard, R. E. Whitwam, K. C. Onwueme and M. Ruppen, *Curr. Pharm. Des.*, 2000, **6**, 1841–1879.
14 M. Moutiez, G. Lippens, C. Sergheraert and A. Tartar, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 719–724.
15 J. J. De Voss, C. A. Townsend, W. D. Ding, G. O. Morton, G. A. Ellestad, N. Zein, A. B. Tabor and S. L. Schreiber, *J. Am. Chem. Soc.*, 1990, **112**, 9669–9670.
16 N. Zein, A. M. Sinha, W. J. McGarren and G. A. Ellestad, *Science*, 1988, **240**, 1198–1201.
17 A. Varki, *Glycobiology*, 1993, **3**, 97–130.
18 B. G. Davis, *Chem. Rev.*, 2002, **102**, 579–601.
19 R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683–720.
20 R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta*, 1999, **1473**, 4–8.
21 P. M. Rudd, T. Elliott, P. Cresswell, I. A. Wilson and R. A. Dwek, *Science*, 2001, **291**, 2370–2376.
22 C. Neususs, U. Demelbauer and M. Pelzing, *Electrophoresis*, 2005, **26**, 1442–1450.
23 B. G. Davis, R. C. Lloyd and J. B. Jones, *J. Org. Chem.*, 1998, **63**, 9614–9615.
24 W. M. Macindoe, A. H. van Oijen and G.-J. Boons, *Chem. Commun.*, 1998, 847–848.
25 B. G. Davis, M. A. T. Maughan, M. P. Green, A. Ullman and J. B. Jones, *Tetrahedron: Asymmetry*, 2000, **11**, 245–262.
26 D. P. Gamblin, P. Garnier, S. J. Ward, N. J. Oldham, A. J. Fairbanks and B. G. Davis, *Org. Biomol. Chem.*, 2003, **1**, 3642–3644.
27 D. P. Gamblin, P. Garnier, S. van Kasteren, N. J. Oldham, A. J. Fairbanks and B. G. Davis, *Angew. Chem., Int. Ed.*, 2004, **43**, 828–833.
28 G. J. L. Bernardes, D. P. Gamblin and B. G. Davis, *Angew. Chem., Int. Ed.*, 2006, **45**, 4007–4011.
29 S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony and B. G. Davis, *Nature*, 2007, **446**, 1105–1109.
30 M. N. Fodge and S. Al-Karadaghi, *Protein Eng.*, 2002, **15**, 353–358.
31 R. H. L. Lundin, B. E. Noren and P. O. Edlund, *Tetrahedron Lett.*, 1994, **35**, 6339–6342.
32 L. Chen and G. Barany, *Lett. Pept. Sci.*, 1996, **3**, 283–292.
33 G. L. Kenyon and T. W. Bruice, *Methods Enzymol.*, 1977, **47**, 407–430.
34 D. N. Harpp, D. K. Ash and R. A. Smith, *J. Org. Chem.*, 1979, **44**, 4135–4140.
35 E. J. Grayson, S. J. Ward, A. L. Hall, P. M. Rendle, D. P. Gamblin, A. S. Batsanov and B. G. Davis, *J. Org. Chem.*, 2005, **70**, 9740–9754.
36 Z. Shaked, R. P. Szajewski and G. M. Whitesides, *Biochemistry*, 1980, **19**, 4156–4166.
37 P. G. Scheurer and F. Smith, *J. Am. Chem. Soc.*, 1954, **76**, 3224.
38 O. H. Drummer, L. Routley and N. Christophidis, *Biochem. Pharmacol.*, 1987, **36**, 1197–1201.
39 J. K. Dale and C. S. Hudson, *J. Am. Chem. Soc.*, 1930, **52**, 2534.
40 A. Okado-Matsumoto, Z. Q. Guan and I. Fridovich, *Free Radical Biol. Med.*, 2006, **41**, 1837–1846.
41 H. D. Flack, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1983, **39**, 876.