Glycoprotein Variants

Modular Assembly of Glycoproteins: Towards the Synthesis of GlyCAM-1 by Using Expressed Protein Ligation**

Derek Macmillan and Carolyn R. Bertozzi*

Glycosylation is a vital protein modification for the normal growth and development of organisms.[1] However, establishing the biological significance of specific covalently bound

[*] Dr. D. Macmillan
School of Chemistry, University of Edinburgh
Kings Buildings, West Mains Road
Edinburgh, EH9 3JJ (GB)
Fax: (+44) 131 650 4743
E-mail: derek.macmillan@ed.ac.uk

Prof. C. R. Bertozzi
Center for New Directions in Organic Synthesis
Departments of Chemistry and Molecular and Cell Biology and
Howard Hughes Medical Institute, University of California
Berkeley, CA 94720 (USA)

[**] The center for New Directions in Organic Synthesis is supported by Bristol–Myers Squibb as Sponsoring Member and Novartis as Supporting Member. The authors acknowledge financial support from The National Science Foundation and the BBSRC.

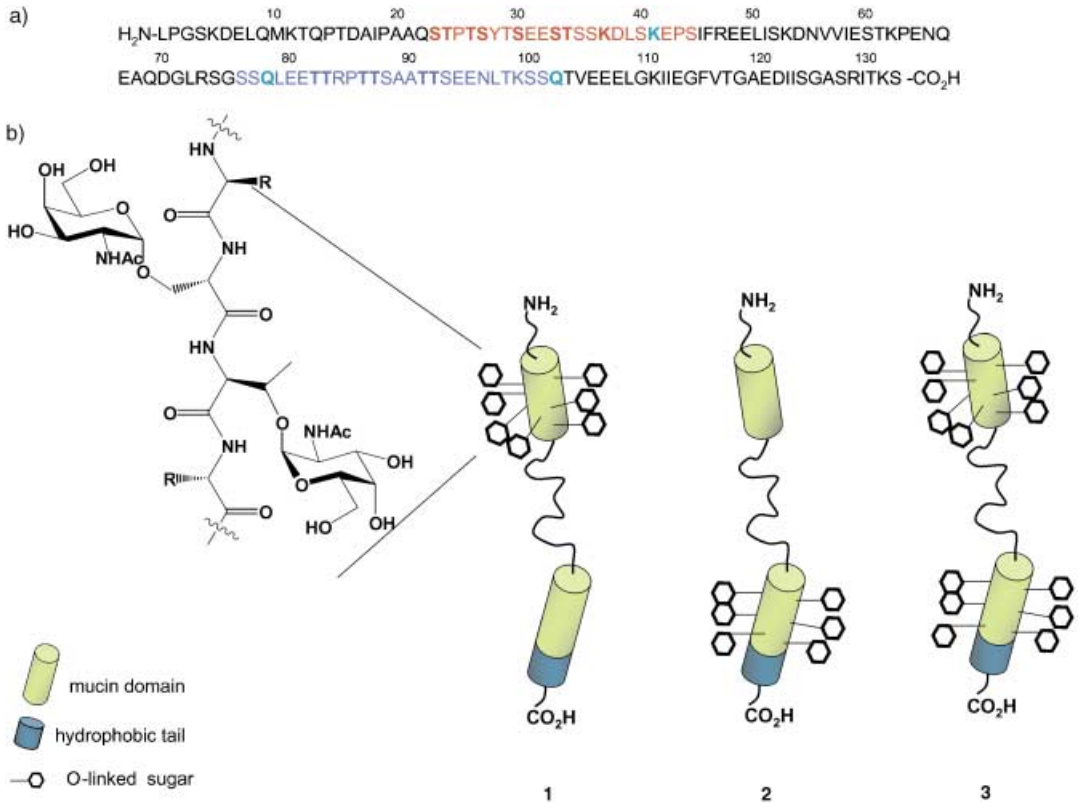
Supporting information for this article (synthetic procedures, HPLC traces, and ESI-MS data) is available on the WWW under http://www.angewandte.org or from the author.

 DOI: 10.1002/ange.200352673

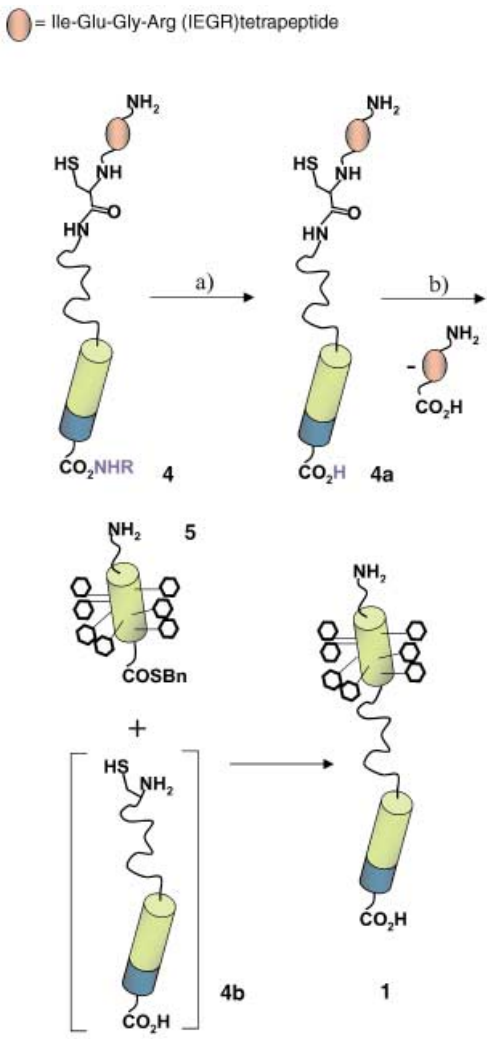
oligosaccharides is difficult owing to glycoprotein microheterogeneity,[2] a phenomenon arising from the fact that protein glycosylation is not under direct genetic control. Consequently, a single glycoprotein can exist as a complex mixture of glycosylated forms termed "glycoforms" and analytical difficulties arising from this property have been a driving force for the development of new methods for the synthesis of glycoproteins with predefined oligosaccharides. Although bacteria lack the protein-glycosylation machinery of mammalian cells, they can be employed to express large quantities of recombinant proteins which allows subsequent chemical modifications to be monitored by using standard biophysical techniques.[3] More recently it has been shown that non-native N-linked glycans can be assembled in *Escherichia coli* after transformation with the N-linked glycosylation apparatus from *Campylobacter jejuni* and raises the possibility of glycan engineering in bacteria.[4] Furthermore, bacteria can produce useful protein fragments such as C-terminal thioesters which can be used in convergent protein-coupling techniques, such as native chemical ligation[5] and expressed protein ligation (EPL).[6] Glycoprotein assembly is particularly suited to EPL since synthetically derived glycopeptides[7] can be fused with bacterially derived protein fragments of large molecular weight to give glycoproteins.[8]

We have focused our efforts on GlyCAM-1, a mucin-like glycoprotein that functions as a ligand for the leukocyte adhesion molecule L-selectin. GlyCAM-1 comprises two mucin domains separated by a central, unglycosylated domain (Scheme 1). The mucin domains, which consist of clustered oligosaccharides characterized by α-*O*-glycosidic linkages between *N*-acetyl galactosamine (GalNAc) and the hydroxy groups of Ser/Thr residues of the protein backbone, are important for binding L-selectin through a 6-sulfo sialyl Lewis^X motif.[9] We aim to study the relevance of O-linked glycosylation to the structure and function of GlyCAM-1 through the synthesis of three separate semisynthetic variants: **1** contains the N-terminal mucin domain, **2** has a glycosylated C-terminal mucin domain, and **3** is glycosylated in both mucin domains (see Scheme 1). In this study we produced glycosylated forms of these targets with single GalNAc residues at the sites of glycosylation. It is possible to elaborate these GalNAc residues enzymatically[10] to generate fully functional GlyCAM-1.

First, we set out to prepare glycoform **1** from a synthetic glycopeptide thioester and a bacterially derived protein fragment. The unglycosylated GlyCAM-1 protein fragment **4** comprising GlyCAM-1 residues 41–132 (Scheme 2) was expressed from the commercially available IMPACT CN intein-fusion vector (pTYB-1).[8a] After cleavage from the intein-chitin binding domain (CBD) fusion protein with dithiothreitol (DTT, 50 mM), the target polypeptide **4a** was readily isolated in 5–10 mg quantities from one litre of bacterial cell culture and characterized by SDS-PAGE and LC-MS. A factor Xa protease recognition peptide, Ile-Glu-Gly-Arg (IEGR), was included as a protecting group for a



Scheme 1. a) Amino acid sequence of GlyCAM-1. The mucin domains are highlighted in red and blue and sites at which O-linked GalNAc was incorporated are shown in bold. Amino acid substitutions (to cysteine) were conducted at positions highlighted in green; b) target GlyCAM-1 glycoforms for semisynthesis by using expressed protein ligation.



Scheme 2. Assembly of glycoform **1**. a) 50 mM DTT, 100 mM $NaHPO_3$; pH 8.0, 100 mM NaCl; b) Factor Xa protease, 5 mM $CaCl_2$, 100 mM $NaHPO_3$; pH 8.0, 100 mM NaCl, 2% MESNA. R = intein-CBD fusion protein, DTT = 1,4-dithiothreitol, MESNA = 2-mercaptoethanesulfonic acid.

latent N-terminal cysteine residue to prevent reaction with the intermediate peptide α-thioester formed during the cleavage of the intein-CBD fragment. The IEGR sequence was readily cleaved from precursor **4a** to form **4b** by using factor Xa protease, which liberated the N-terminal cysteine residue for subsequent ligation. In practice, the IEGR sequence was cleaved in situ and **4b** was not isolated. With recombinant peptide **4a** in hand, we embarked on the synthesis of glycopeptide **5** corresponding to the N-terminal mucin domain of GlyCAM-1.

The synthesis of glycopeptide α-thioester **5** was particularly challenging since seven glycoamino acids in close proximity had to be installed. The synthesis was conducted by using the 4-sulfamylbutyryl resin, as described by Shin et al.,[11] employing five equivalents of amino acid in each coupling step, and HBTU/HOBt (HBTU = *O*-benzotriazol-1-yl-*N*,*N*,*N*′,*N*′-tetramethyluronium-hexafluoro-phosphate, HOBt = 1-hydroxybenzotriazole) as the coupling reagents. The coupling time was seven hours for glycoamino acid building blocks Fmoc-Thr(α-*O*-GalNAc$(OAc)_3$-OH and Fmoc-Ser(α-*O*-GalNAc$(OAc)_3$-OH (Fmoc = 9-fluorenylmethyloxycarbonyl).[12] After synthesis, the resin was activated for cleavage by alkylation with iodoacetonitrile, and release of the fully protected glycopeptide α-thioester was effected with benzyl mercaptan and catalytic NaSPh.[13] After TFA-mediated (TFA = trifluoroacetic acid) removal of the remaining peptide backbone protecting groups, ether precipitation, and HPLC purification, the synthetic glycopeptide thioester **5** was obtained and characterized by electrospray mass spectrometry (Figure 1).

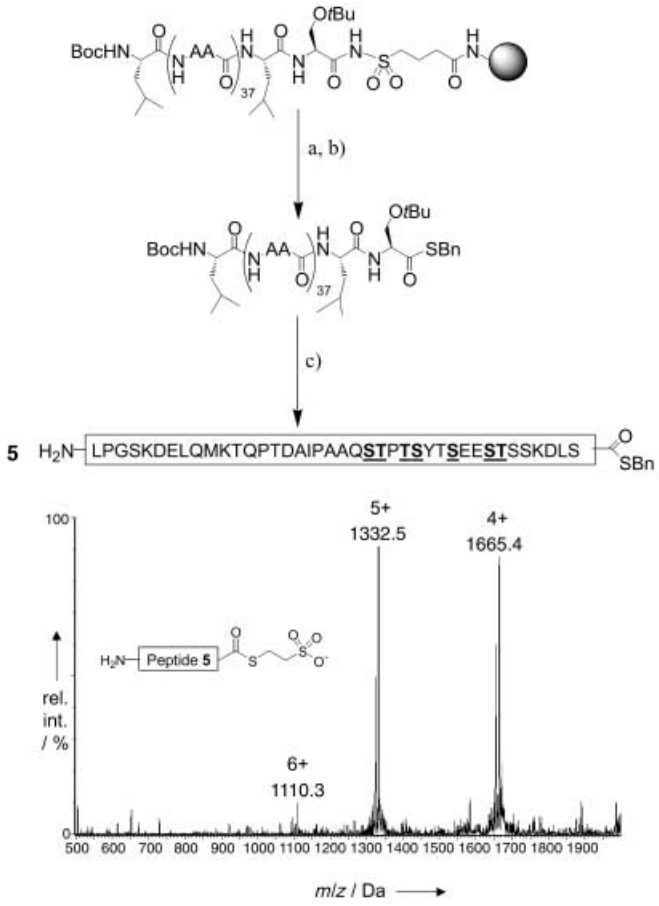


Figure 1. Synthesis and MS characterization of glycopeptide thioester **5**. Glycoamino acids were incorporated at positions indicated in bold, underlined type. a) ICH_2CN, DIPEA, DMF, 16 h; b) BnSH, NaSPh, THF, 16 h; c) 85% TFA, 5% EDT, 5% thioanisole, 5% H_2O, 4 h. 10% overall yield based on resin loading. Calculated mass (MESNA thioester) = 6654.8 Da, observed mass = 6656.9 Da. DIPEA, *N*,*N*-diisopropylethylamine; EDT = 1,2-ethanedithiol.

The ligation reaction between the fragments **4b** and **5** was carried out under nondenaturing conditions and was usually complete within 48 hours; the reaction was monitored by using LC-MS and was quantitative with respect to the limiting reagent (synthetic thioester). Excess reagents could also be recovered and nonspecific proteolysis was not observed. After ligation, the 21 acetyl esters originally present on the GalNAc residues of glycopeptide **5** were removed by using

2% aqueous hydrazine to afford the target glycoprotein **1** (Figure 2) on a multimilligram scale. Note that the promiscuity of factor Xa for sequences similar to IEGR is well documented and this "one-pot" factor Xa cleavage and ligation procedure is likely to be protein dependant to some degree.

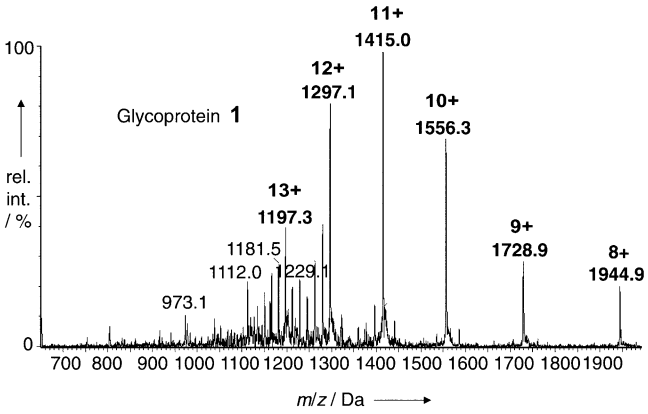
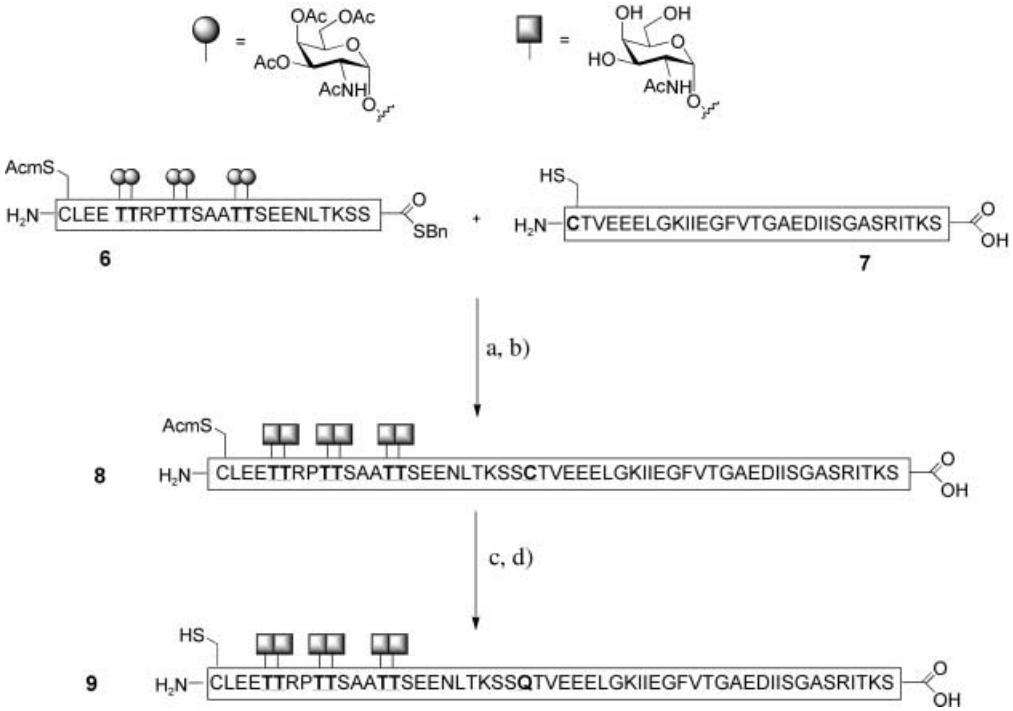
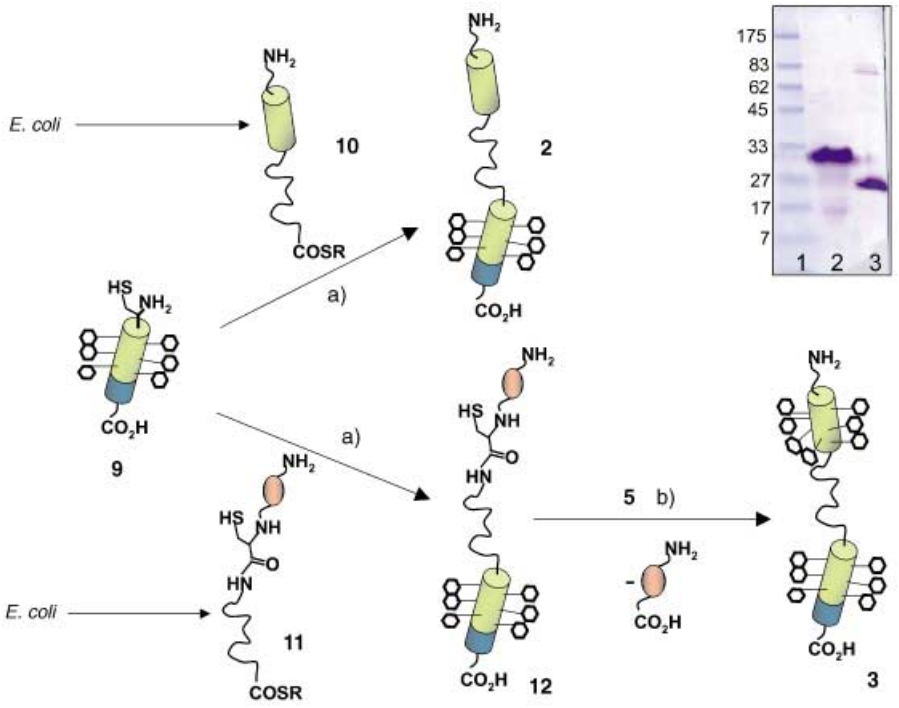


Figure 2. Electrospray MS analysis of HPLC purified glycoform **1**, calculated mass = 15550.3 Da, observed mass = 15550 Da.

The synthesis of glycoform **2** required a reversal of the above strategy, that is, a bacterially derived thioester and a synthetic glycopeptide bearing an N-terminal cysteine residue. Glycopeptide **9** (Scheme 3) was prepared in two sections that were joined by native chemical ligation. This approach was chosen since the C-terminal peptide was particularly hydrophobic, which hampered monitoring of the glycoamino acid coupling reactions by using the Kaiser ninhydrin test. Residues 78–101 (**6**) comprising the mucin domain were prepared as described above for thioester **5** except the N-terminal cysteine was temporarily protected with an *S*-acetamidomethyl group that was stable to the conditions required for cleavage from the resin. This thioester was then ligated to C-terminal residues 102–132 (**7**), prepared by standard Fmoc-based solid-phase peptide synthesis (SPPS) on NovaSyn-TGT resin, by native chemical ligation in sodium phosphate buffer (pH 8.0) and in the presence of 4% MESNA. The reaction was complete after 36 h, and following HPLC purification, the internal cysteine residue was capped by treatment with iodoacetamide added to a final concentration of 10 mM for 0.5 hours. The capping of this cysteine residue appeared essential for acceptable reaction rates and yields in subsequent ligations. Glycopeptide **8** was then purified by semipreparative HPLC and finally dissolved in 10% acetic acid. Upon overnight treatment with Hg^{II} acetate and excess 1,4-dithiothreitol (DTT) under argon, the desired C-terminal glycopeptide **9** was obtained (calculated molecular weight = 7067.4 Da, observed molecular weight = 7067.7 Da). This synthetic glycopeptide was then ligated to the bacterially derived thioester **10**[8a] (Scheme 4). The ligation was performed in sodium phosphate buffer (100 mM, pH 8.0) and yielded 25% (determined by HPLC) of the desired glycoform **2** with an observed molecular weight of 15510 Da (calculated molecular weight = 15509.7 Da). Though low yielding, the product and unreacted glycopeptide could be copurified by HPLC and resubjected to bacterially derived thioester **10** to drive the reaction to completion. In practice this ligation was conducted a total of three times to obtain reasonable yields (> 70%, determined by LC-MS) and



Scheme 3. Synthesis of the C-terminal mucin domain **9**. a) 200 mM $NaHPO_3$ buffer; pH 8.0, 4% MESNA, 24 h; b) 5% hydrazine hydrate, 10% DTT, 2 h; c) 2.5 mM iodoacetamide, 10 mM $NaHPO_3$ buffer; pH 8.0, 1 h; d) $Hg(OAc)_2$, 10% AcOH, 50 mM DTT, 16 h.



Scheme 4. Schematic representation of the semisynthetic assembly of glycoforms **2** and **3**. a) 2% MESNA, 100 mM $NaHPO_3$; pH 8.0, 100 mM NaCl; b) Factor Xa protease, 5 mM $CaCl_2$, 100 mM $NaHPO_3$; pH 8.0, 100 mM NaCl, 2% MESNA. Analysis by coomassie blue stained gel electrophoresis (inset) 1) molecular weight markers; 2) glycoform **3**; 3) full-length unglycosylated GlyCAM-1.[16]

conducting this ligation under denaturing conditions may aid future attempts.

Glycoform **3** was then prepared by ligation of synthetic glycopeptide **9** to bacterially derived peptide thioester **11**[8a] resulting in intermediate **12** with an observed molecular weight of 11731.5 Da (calculated molecular weight = 11732.3 Da). As for the synthesis of glycoform **2**, this ligation reaction proceeded sluggishly and in low yield but could be driven to completion by copurification of **9** and **12** and resubjection of the mixture to bacterial thioester **11** as above. Intermediate **12** was then converted into glycoform **3** in the same manner as for the synthesis of glycoform **1** (Scheme 2). Finally, GlyCAM-1 has no native cysteine residues, therefore additional cysteine residues introduced for ligations (Cys41 for **1**, and Cys78 for **2**) were capped to resemble glutamine residues[14] in quantitative yield (determined by LC-MS) within 0.5 hours by the addition of iodoacetamide to glycoprotein samples.

In summary, we have applied protein semisynthesis to the preparation of three defined GlyCAM-1 glycoforms with as many as 13 *N*-acetylgalactosamine residues at predetermined positions. There have been a few recent reports of the application of native chemical ligation to the total synthesis of glycoproteins.[15] However, this is the first demonstration of the modular assembly of a physiologically relevant glycoprotein by using the expressed protein ligation technology and serves as a platform for the launch of biological investigations which will be reported in due course.

Received: August 18, 2003 [Z52673]

Keywords: expressed protein ligation · glycopeptides · glycoproteins · protein engineering · semisynthesis

[1] a) A. Varki, *Glycobiology* **1993**, *3*, 97; b) R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683; c) A. Helenius, M. Aebi, *Science* **2001**, *291*, 2364; d) R. G. Spiro, *Glycobiology* **2002**, *12*, 43R.

[2] R. S. Rush, P. L. Derby, D. M. Smith, C. Merry, G. Rogers, M. F. Rohde, V. Katta, *Anal. Chem.* **1995**, *67*, 1442.

[3] B. G. Davis, *Chem. Rev.* **2002**, *102*, 579.

[4] M. Wacker, D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren, M. Aebi, *Science* **2002**, *298*, 1790.

[5] P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923.

[6] a) T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705; b) R. M. Hofmann, T. W. Muir, *Curr. Opin. Biotechnol.* **2002**, *13*, 297; c) I. Giriat, T. W. Muir, *J. Am. Chem. Soc.* **2003**, *125*, 7180.

[7] H. Herzner, T. Reipen, M. Schultz, H. Kunz, *Chem. Rev.* **2000**, *100*, 4495.

[8] a) D. Macmillan, C. R. Bertozzi, *Tetrahedron* **2000**, *56*, 9515; b) T. J. W. Tolbert, C.-H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 5421.

[9] a) D. Crommie, S. D. Rosen, *J. Biol. Chem.* **1995**, *270*, 22614; b) S. Hemmerich, S. D. Rosen, *Biochemistry* **1994**, *33*, 4830; c) S. Hemmerich, C. R. Bertozzi, H. Leffler, S. D. Rosen, *Biochemistry* **1994**, *33*, 4820; d) S. Hemmerich, H. Leffler, S. D. Rosen, *J. Biol. Chem.* **1995**, *270*, 12035.

[10] a) E. C. Rodriguez, L. A. Marcaurelle, C. R. Bertozzi, *J. Org. Chem.* **1998**, *63*, 7134; b) A. Leppanen, S. P. White, J. Helin, R. P. McEver, R. D. Cummings, *J. Biol. Chem.* **2000**, *275*, 39569.

[11] Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman, C. R. Bertozzi, *J. Am. Chem. Soc.* **1999**, *121*, 11684.

[12] K. A. Winans, D. S. King, V. R. Rao, C. R. Bertozzi, *Biochemistry* **1999**, *38*, 11700.

[13] R. Ingenito, E. Bianchi, D. Fattori, A. Pessi, *J. Am. Chem. Soc.* **1999**, *121*, 11369.

[14] G. G. Kochendoerfer, S.-Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent, J. A. Bradburne, *Science* **2003**, *299*, 884.

[15] a) L. A. Marcaurelle, L. S. Mizoue, J. Wilken, L. Oldham, S. B. H. Kent, T. M. Handel, C. R. Bertozzi, *Chem. Eur. J.* **2001**, *7*, 1129; b) H. Hojo, E. Haginoya, Y. Matsumoto, Y. Nakahara, K. Nabeshima, B. P. Toole, Y. Watanabe, *Tetrahedron Lett.* **2003**, *44*, 2961c) J. S. Miller, V. Y. Dudkin, G. J. Lyon, T. W. Muir, S. J. Danishefsky, *Angew. Chem.* **2003**, *115*, 447; *Angew. Chem. Int. Ed.* **2003**, *42*, 431.

[16] Both proteins migrate with a higher apparent molecular mass than predicted, indeed almost twice that determined by mass spectrometry (full-length bacterially derived GlyCAM-1, observed mass (ESI-MS) = 14409 Da and 14451 Da for the acetylated species; semisynthetic GlyCAM-1 observed molecular mass (ESI-MS) = 16776 Da).

