J|A|C|S
COMMUNICATIONS

Published on Web 01/22/2003

A Method for the Generation of Glycoprotein Mimetics

Haitian Liu,[†] Lei Wang,[†] Ansgar Brock,[‡] Chi-Huey Wong,*[,†] and Peter G. Schultz*[,†,‡]

Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and The Genomics Institute of the Novartis Research Foundation, 3115 Merryfield Row, San Diego, California 92121

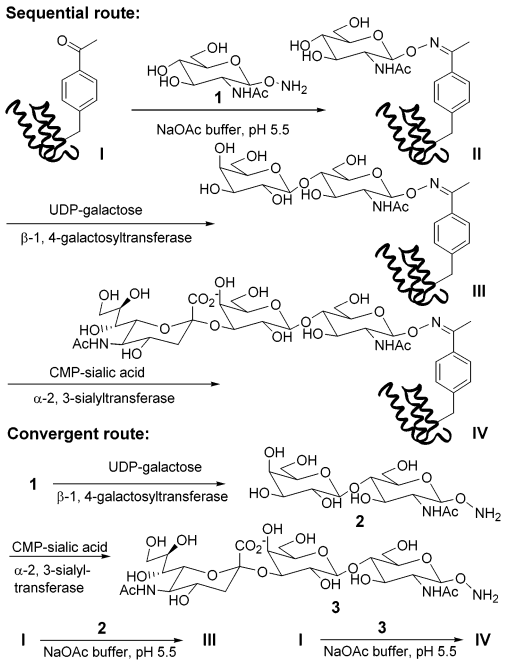
Received November 21, 2002; E-mail: *wong@scripps.edu, schultz@scripps.edu

The posttranslational modification of proteins by glycosylation can affect protein folding and stability, modify the intrinsic activity of proteins, and modulate their interactions with other biomolecules.[1] Natural glycoproteins are often present as a population of many different glycoforms, which makes analysis of glycan structure and the study of glycosylation effects on protein structure and function difficult. Therefore, methods for the synthesis of natural and unnatural homogeneously glycosylated proteins are needed for the systematic understanding of glycan function and for the development of improved glycoprotein therapeutics. One approach makes use of glycosidases to convert a heterogeneous natural glycoprotein to a simple homogeneous core, onto which saccharides can then be grafted sequentially with glycosyltransferases.[2] A limitation of this approach is that the primary glycosylation sites are predetermined by the cell line in which the protein is expressed. Alternatively, a glycopeptide containing the desired glycan structure can be synthesized by solid-phase peptide synthesis. This glycopeptide can be coupled to other peptides or recombinant protein fragments to afford a larger glycoprotein by native chemical ligation,[3] expressed protein ligation,[4] or engineered proteases.[5] Both native chemical ligation and expressed protein ligation are most effective with small proteins and necessitate a cysteine residue at the N-terminus of the glycopeptide. When a protease is used to ligate peptides together, the ligation site must be placed far away from the glycosylation site for good coupling yields.[5] A third approach is to modify proteins with saccharides directly using chemical methods. Good selectivity can be achieved with haloacetamide saccharide derivatives, which are coupled to the thiol group of cysteine,[6,7] but this method may become problematic with proteins that have more than one cysteine residue.

The availability of a nonproteinogenic functional group with a unique reactivity would greatly facilitate the selective chemical modification of proteins. The keto group is absent from the side chains of natural amino acids and reacts readily and selectively with hydrazide and hydroxylamine derivatives under mild conditions in the presence of the common amino acids.[8] It has been included in peptides by solid-phase peptide synthesis and coupled with nucleophilic saccharide derivatives to construct neoglycopeptides.[9] We recently developed a general method that allows for the site-specific incorporation of unnatural amino acids into proteins directly in living cells.[10] A keto-containing amino acid, *p*-acetyl-L-phenylalanine, has been successfully incorporated in response to the amber nonsense codon with translation fidelity greater than 99.8%.[11] We report here the preparation of homogeneous glycoprotein mimetics using the genetically encoded keto functionality together with aminooxy saccharide derivatives.

Two routes were explored to generate the glycoprotein mimetics (Scheme 1). In the first approach, one saccharide derivatized with an aminooxy group is first coupled to the keto group, and additional saccharides are attached enzymatically with glycosyltransferases. In a more convergent second route, a glycan with defined structure is prepared as an aminooxy derivative and is coupled directly to the protein in one step. The Z domain of staphylococcal protein A was used as the model protein,[12] because its relatively small size (molecular weight 7.9 kD) facilitates mass spectrometric characterization with very high accuracy.

Scheme 1



The seventh codon of the corresponding gene was mutated to amber stop codon TAG and a His6 tag was added to the C-terminus to facilitate protein purification. *p*-Acetyl-L-phenylalanine was incorporated at the amber position to afford the mutant Z domain protein by previously reported protocols.[11] Approximately 3.6 mg/L protein was obtained after nickel affinity chromatography. The β-linked aminooxy analogue of *N*-acetylglucosamine (GlcNAc) **1** was then synthesized by following published procedures.[13] The mutant Z domain protein (10 mg/mL) and aminooxy saccharide **1** (21 mM) were mixed in aqueous 100 mM sodium acetate buffer (pH 5.5) and incubated at 37 °C for 7 to 26 h. The reaction mixture was analyzed by reverse phase high-performance liquid chromatography (HPLC) by monitoring absorbance at 280 nm (Figure 1). Only two major peaks were observed, and the corresponding eluents were characterized by matrix-assisted laser desorption/ionization−Fourier transform ion cyclotron resonance mass spectrometry (MALDI−FTICR MS) (Figure 2). The monoisotopic masses obtained indicate that one peak (t_R = 44.8 min) corresponds to the

[†] The Scripps Research Institute.
[‡] The Genomics Institute of the Novartis Research Foundation.

10.1021/ja029433n CCC: $25.00

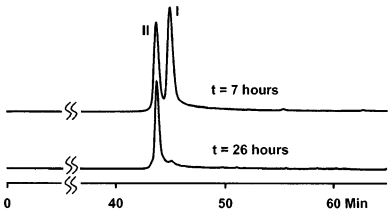


Figure 1. HPLC analysis of the coupling reaction between aminooxy saccharide **1** and mutant Z domain protein **I** containing *p*-acetyl-L-phenylalanine at 7 and 26 h.

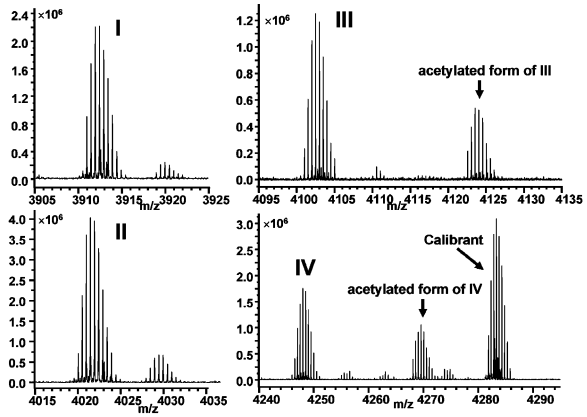


Figure 2. High-resolution MALDI−FTICR MS spectra of mutant Z domain protein **I** and glycoprotein mimetics **II**, **III**, and **IV**.[14] The 2^+ isotopic cluster of each spectrum is shown.

unreacted mutant Z domain (M_{theor} = 7818.833 Da, M_{exptl} = 7818.836 Da)[14] and the other peak (t_R = 43.2 min) corresponds to the mutant Z domain derivatized with the aminooxy saccharide **1** (**M_{theor}** = 8036.924 Da, **M_{exptl}** = 8036.914 Da**)**. As a control, when tyrosine is incorporated at the seventh position of Z domain, no saccharide derivatized protein is observed. This fact, together with the high-accuracy mass (error < 1.2 ppm) observed for the saccharide-modified Z domain, confirmed that the aminooxy saccharide **1** is attached to the keto group, selectively. The coupling efficiency increases with time (determined from the areas of the HPLC peaks corresponding to starting material and product): the conversion of starting material to product was 42% after 7 h and greater than 95% after 26 h (Figure 1).

We next determined whether a second saccharide could be coupled to the first enzymatically. The purified adduct **II** (5 mg/mL) was incubated with UDP-galactose (UDP-Gal) (16 mM) and β-1,4-galactosyltransferase (0.4 units/mL) in 150 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) buffer (pH 7.4) for 48 h at ambient temperature. β-1,4-Galactosyltransferase is known to transfer galactose from the sugar nucleotide to the 4 position of a GlcNAc moiety to form Galβ1,4GlcNAc.[15] After separation by HPLC, a new peak was identified (t_R = 42.5 min). The monoisotopic mass (**M_{theor}** = 8198.977, **M_{exptl}** = 8198.969**)** of the eluent confirmed that the galactose was coupled to GlcNAc to yield adduct **III** (Figure 2). The coupling efficiency determined by HPLC analysis was about 60%, a value close to that reported previously for β-1,4-galactosyltransferase.[2] This result indicates that the non-native linkage between the first saccharide and the protein does not significantly affect the glycosyltransferase reaction. Further reaction of this disaccharide-labeled protein with CMP-sialic acid and α-2, 3-sialyltransferase[16] resulted in the addition of sialic acid to galactose to afford **IV** (t_R = 41.7 min), as confirmed by MALDI−FTICR MS (**M_{theor}** = 8490.072, **M_{exptl}** = 8490.014**)** (Figure 2). The coupling efficiency for conversion of **III** to **IV** was 65% on the basis of HPLC analysis.

Glycoprotein mimetics **III** and **IV** were also prepared using a convergent route. Aminooxy GlcNAc (0.05 M) was converted to **2** using β-1,4-galactosyltransferase (0.75 units/mL) and the glycosyl donor UDP-galactose in 70% overall yield in 150 mM HEPES buffer (pH 7.4). After purification by aminopropyl silica gel HPLC, sialic acid was added to **2** (0.03 M) to afford **3** in the same buffer mentioned above in approximately 80% yield using α-2,3-sialyltransferase (0.22 units/mL) and CMP-sialic acid (0.03 M). Purified aminooxy analogue **2** and **3** (13 and 7.2 mM, respectively) were coupled to the Z domain protein (5 mg/mL) containing *p*-acetyl-L-phenylalanine in 100 mM aqueous sodium acetate buffer (pH 5.5) at ambient temperature to afford glycoprotein mimetics **III** and **IV**, respectively. The resultant **III** and **IV** were identical to the corresponding adducts prepared by the first sequential route, as confirmed by HPLC and MALDI−FTICR MS analysis. The coupling efficiencies of **2** to **I** and **3** to **I** under the same reaction conditions for 26 h were about 76% and 60%, respectively. The yields were lower than that for the coupling of **1** to **I** (95%) likely due to the increasing steric effect as the glycan becomes more complicated.

In summary, we have demonstrated a general method for the synthesis of homogeneous glycoprotein mimetics containing well-defined saccharide substituents. This method should be applicable to most proteins that can be expressed in *Escherichia coli*. It remains to be determined whether site-specific saccharide attachment via the non-native linkage contributes to protein stability and function in the same way as that via the natural linkage. We are currently evolving synthetases that can incorporate glycosylated amino acids directly into proteins in vivo so as to create a saccharide core with a native linkage to the protein.

Acknowledgment. We thank financial support from NIH (GM44154 and GM62159). H.L. is a Skaggs predoctoral fellow and acknowledges a Louis R. Jabinson fellowship.

Supporting Information Available: Experimental details for the preparation of compounds **2**, **3**, and glycoprotein mimetics **II**, **III**, **IV**, for MALDI−FTICR MS and spectral data for compound **2** and **3**. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Varki, A. *Glycobiology* **1993**, *3*, 97−130.
(2) Witte, K.; Sears, P.; Martin, R.; Wong, C.-H. *J. Am. Chem. Soc.* **1997**, *119*, 2114−2118.
(3) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684−11689.
(4) Tolbert, T. J.; Wong, C.-H. *J. Am. Chem. Soc.* **2000**, *122*, 5421−5428.
(5) Witte, K.; Seitz, O.; Wong, C.-H. *J. Am. Chem. Soc.* **1998**, *120*, 1979−1989.
(6) Davis, N. J.; Flitsch, S. L. *Tetrahedron Lett.* **1991**, *32*, 6793−6796.
(7) Macmillan, D.; Daines, A. M.; Bayrhuber, M.; Flitsch, S. L. *Org. Lett.* **2002**, *4*, 1467−1470.
(8) Cornish, V. W.; Hahn, K. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 8150−8151 and references therein.
(9) Rodriguez, E. C.; Marcaurelle, L. A.; Bertozzi, C. R. *J. Org. Chem.* **1998**, *63*, 7134−7135.
(10) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498−500.
(11) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56−61.
(12) Nilsson, B.; Moks, T.; Jansson, B.; Abrahmsen, L.; Elmblad, A.; Holmgren, E.; Henrichson, C.; Jones, T. A.; Uhlen, M. *Protein Eng.* **1987**, *1*, 107−113.
(13) Cao, S.; Tropper, F. D.; Roy, R. *Tetrahedron* **1995**, *51*, 6679−6686.
(14) When expressed in *E. coli*, Z domain protein has three forms: the intact protein, protein without the first methionine, and the acetylated form of the protein without methionine. The intact protein can be separated from the other two forms using reverse phase HPLC. To simplify mass spectrometric analysis, purified fraction containing Z domain without first methionine and its acetylated form were used in this study. Two molecular peaks can be observed which correspond to these two forms in all mass spectra, as labeled in spectra for **III** and **IV** in Figure 2.
(15) Schanbacher, F. L.; Ebner, K. E. *J. Biol. Chem.* **1970**, *245*, 5057−5061.
(16) Kitagawa, H.; Paulson, J. C. *J. Biol. Chem.* **1994**, *269*, 1394−1401.

JA029433N