Toward Fully Synthetic Glycoproteins by Ultimately Convergent Routes: A Solution to a Long-Standing Problem

J. David Warren,[†] Justin S. Miller,[†] Stacy J. Keding,[†] and Samuel J. Danishefsky*,[†,‡]

*Laboratory for Bioorganic Chemistry, Sloan-Kettering Institute for Cancer Research, 1275 York Avenue,
New York, New York 10021, and Department of Chemistry, Columbia University, Havemeyer Hall, 3000 Broadway,
New York, New York 10027*

Glycoproteins are important biomacromolecules that are biosynthesized through posttranslational glycosylation of newly fashioned proteins emerging from the ribosome. A long-term goal of the growing field of chemistry-based glycobiology is the delineation of the consequences of glycosylation on critical properties such as protein folding, proteolytic stability, and cell adhesion.[1] Such insights could explain why nature bothers to glycosylate otherwise functional proteins. Moreover, glycoproteins have potentially important clinical roles in the contexts of vaccines,[2] diagnostics,[3] and therapeutics.[4] Indeed, erythropoietin (**1**) (see Figure 1), albeit a heterogeneous glycoprotein,[5] is clinically valuable as a treatment for anemia, among other indications.[4a]
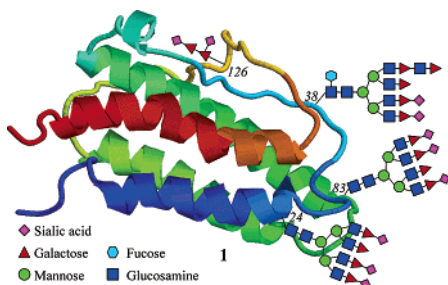

Figure 1. Erythropoietin contains an *O*-linkage at Ser[126] and *N*-linkages at Asn[24], Asn[38], and Asn[83].[6] Coordinates obtained from the Protein Data Bank (PDB accession no. 1BUY)[7,8] were used to render the protein backbone with RasMol; schematic depictions of the glycans were added by the authors. The carbohydrate structures are representative,[9] as glycan compositions can vary significantly in different glycoforms of erythropoietin.[5]

The isolation of homogeneous glycoproteins from natural sources in significant quantity is extremely difficult.[10] Formidable as the challenges are, we felt it worthwhile to explore the prospects of reaching homogeneous glycopeptides and, in time, glycoproteins through the medium of chemical synthesis. Since many naturally occurring, medicinally important glycoproteins (cf., for example, erythropoietin[4a] and gp120[11]) display multiple glycosylation sites containing large oligosaccharide domains, we insisted on stringent standards of convergence and stereoselectivity.

Following extensive earlier studies, wherein our chemistry was interfaced with critical findings from other laboratories,[12] we demonstrated the ability to ligate fully synthetic *N*-linked glycopeptides containing various complex oligosaccharides to larger polypeptide domains.[13] We were thus able to obtain major glycosylated subsections of prostate specific antigen (PSA)[3a] and gp120.[14] Our next goal, with an eye toward erythropoietin, was the achievement of a convergent approach to reach polypeptides that are multiply glycosylated at fixed sites.[15]

In Scheme 1 we sketch out a milestone challenge en route to these long-term goals. In this synthesis, we hope to generate glycopeptides **4** and **7** in a convergent way from their respective glycan and peptide precursors, and to join them with full convergence to reach **8**.

Scheme 1. Convergent Synthesis of Bifunctional Glycopeptides
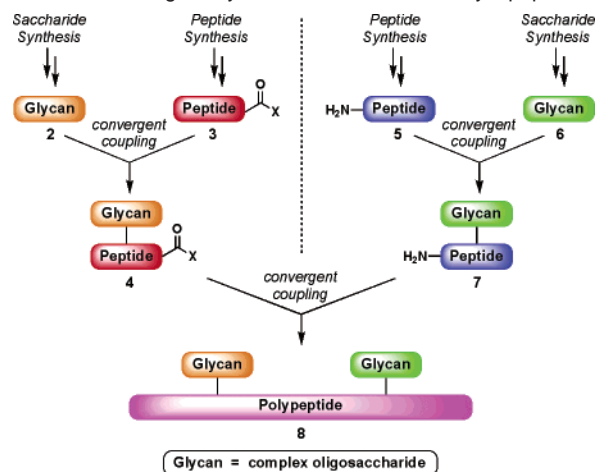
A series of technical obstacles would have to be overcome to accomplish such a goal. A particularly serious issue is the synthesis of a suitable glycopeptide acyl donor for ligation with **7**. The first possibility that comes to mind as an appropriate acyl donor is one in which **4** is a thioester. However, many difficulties would have to be surmounted to reach a thioester such as **4** by convergent means.[16]

It was in pondering this problem that an attractive possibility presented itself (Scheme 2). Our subgoal would be the synthesis of a phenolic ester equipped with an unsymmetrical disulfide (cf. **10**). The phenolic ester linkage would in itself probably not manifest sufficient reactivity to serve as a viable acylating agent.[17] However, equipped as it would be with an ortho disulfide moiety, the phenolic ester could operate as an incipient acyl donor. Thus, reduction of the disulfide in **12**, as shown, might well set the stage for in situ elaboration into a thioester (cf. **13→14**) by O→S acyl migration. Thioester **14**, present in an unfavorable, but *dynamic* equilibrium, could perhaps be interdicted by the machinery appropriate to native chemical ligation (NCL),[12c] ultimately culminating in **8**.

While the proposal for the intermediacy of a discrete thioester **14** is certainly attractive, it is not necessarily an obligatory event en route to ligation. Thus, as suggested in structure **13**, the presence of a free ortho benzene thiol function might well enhance the acylating facility of the phenolic ester,[18] perhaps through in situ

† Sloan-Kettering Institute for Cancer Research.
‡ Columbia University.

hydrogen bonding or anchimeric assistance. In either case, the sequence is set into motion only upon reductive cleavage of the disulfide.

The synthesis of the masked glycopeptide acyl donor was accomplished as follows (Scheme 2). Commercially available 2-mercaptophenol (9) was oxidized to its symmetrical disulfide. Subsequent exchange with excess ethyl disulfide[19] furnished an unsymmetrical, aryl–alkyl disulfide. Acylation of the phenolic hydroxyl group with Boc-phenylalanine followed by acidic cleavage of the Boc protecting group afforded phenylalanine derivative 10. Standard peptide coupling of 10 with a fully protected peptide acid[20] followed by global deprotection provided 11, which contains a free aspartate residue. Aspartylation of a chitobiose-derived glycosyl-amine with 11 afforded the virtual acyl donor 12, corresponding to our target structure 4 (Scheme 1).

Scheme 2. Structure, Synthesis, and Proposed Mechanism of a Masked Glycopeptide Acyl Donor[a]

Glycan = chitobiose in 12-15; Peptide = FmocRNRSG- in 12-15; R = CH₂CH₂SO₃Na
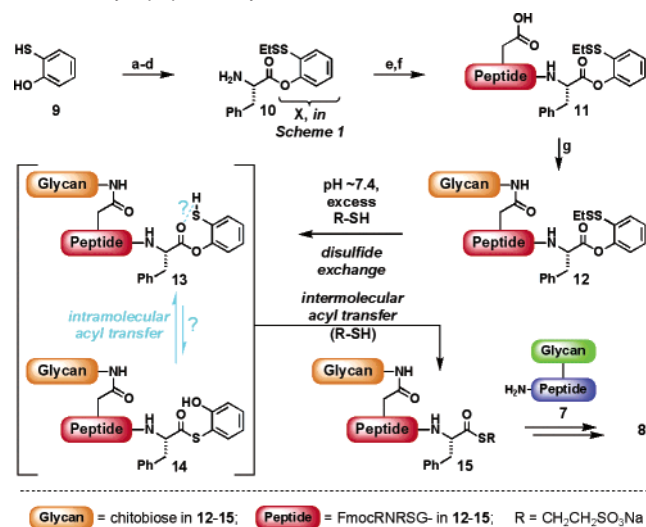
[a] Reagents and conditions: (a) I₂, MeOH, H₂O; (b) BF₃·OEt₂, EtSSEt, CH₂Cl₂, 99%, 2 steps; (c) Boc-Phe-OH, EDCI, DMAP, CH₂Cl₂/THF, 93%; (d) 4 N HCl/dioxane, 94%; (e) Fmoc-Arg(Pbf)-Asp(ᵗBu)-Arg(Pbf)-Ser(ᵗBu)-Gly-OH, HATU, DIEA, DMF, 61%; (f) TFA/phenol/Et₃SiH/H₂O, 35:2:1:1, 60%; (g) GlcNAcβ1→4GlcNAcβ1-NH₂, HATU, DIEA, DMSO, 52%.

As a preliminary probe of the practicability of this scheme, phenylalanine derivative 10 was subjected to the neutral, aqueous, reducing reaction conditions described above, along with free cysteine (Scheme 3). The desired free dipeptide Phe-Cys (17) was obtained in 78% isolated yield. Only a single product was observed by LCMS and ¹H NMR, strongly suggesting that epimerization at the α-carbon of phenylalanine did not occur.

Scheme 3. Demonstration of the Latent Acyl Donor Function
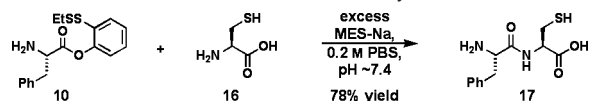
Given this encouraging result, we continued toward our goal of bifunctional glycopeptides. Thus, 12 was treated with excess sodium 2-mercaptoethanesulfonate (MES-Na) at neutral pH in aqueous phosphate buffered saline (PBS) in the presence of 18,[21] (Scheme 4). As observed by LCMS during the reaction, 12 was almost instantaneously (less than 3 min) converted to the thioester derived from MES-Na (cf. 15); very little hydrolysis to the free carboxylic acid was noted. Over the course of the next several hours, we observed a gradual decrease in the amount of this thioester, with a

Scheme 4. Convergent Coupling of Functionalized Glycopeptides
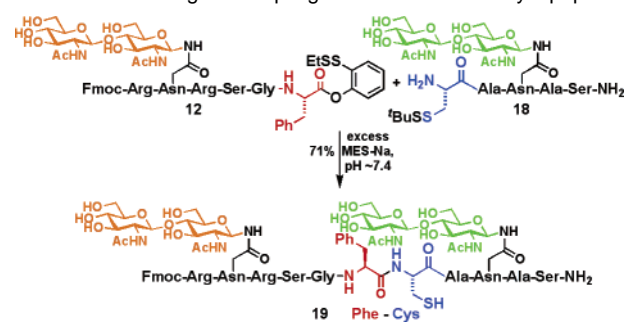
concomitant increase in the desired product 19. Only one compound of the appropriate mass was observed by LCMS and ¹H NMR, again indicating a lack of epimerization.

We then directed our attention to evaluation of the generality of our findings by first establishing applicability to incorporation of O-linked domains,[22] as shown through the synthesis of 20. Notably, we also assembled the histidine-containing system 21 (see Figure 2). Since activated C-terminal histidine residues are known to be particularly susceptible to epimerization,[23] we employed a prophy-lactic dinitrophenyl (DNP) protecting group[24] for the imidazole τ-nitrogen during the synthesis. As expected, the DNP group was cleaved concurrently during the mildly nucleophilic coupling reaction. Gratifyingly, no indications of epimerization were observed by LCMS or ¹H NMR.
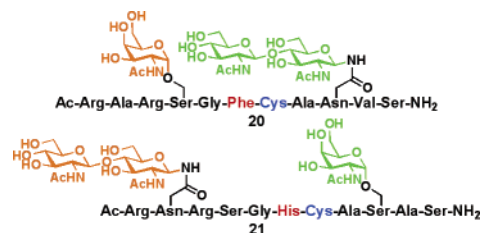

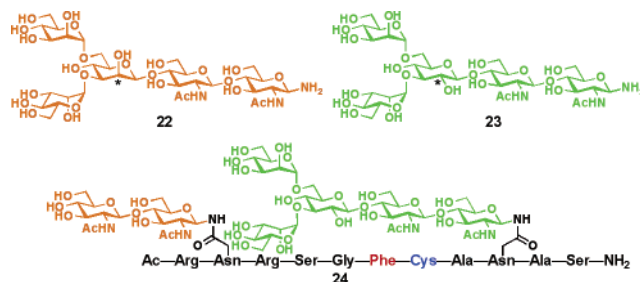**Figure 2.** Bifunctional glycopeptides containing O-linkages.

Figure 3. Anomeric β-glycosylamines (22 and 23) that differ at a single stereocenter (asterisks), and a bifunctional glycopeptide (24) containing one of these structurally complex glycans.

The next phase of our investigation involved validation of the highly convergent methodology in the context of complex glycans. Previously we had disclosed syntheses of glycosylamines 22[13] and 23[25] (Figure 3). It will be noted that these compounds differ in a single stereogenic locus in the interior C-ring (see asterisks). Thus, 22 corresponds to a fully stereochemically competent N-linked high mannose oligosaccharide. In contrast, its C-ring C-2 epimer is an unnatural stereochemical mutant.

Preparation of pentasaccharide-derived glycopeptide coupling partners proceeded via Lansbury aspartylation[12b] of 22 and 23 with 25 and 26, affording 27 and 28, respectively. As a prelude to the finale, coupling of 28 with a relatively simple chitobiose-containing acyl donor gave 24 (Figure 3) in 77% yield.

Scheme 5. Convergent Coupling of Two Functionalized Glycopeptides Containing Distinct, Structurally Complex Glycans[a]
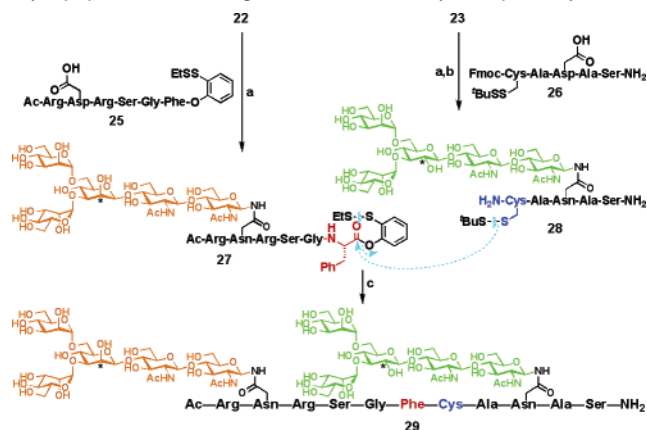
[a] Reagents and conditions: (a) **25** or **26**, HATU, DIEA, DMSO, 50% for **27** (from **22**), 71% for **28** (from **23**); (b) 20% piperidine in DMF, 63%; (c) 0.2 M phosphate, 0.2 M NaCl, pH ~7.4, excess MES-Na, 75%.

With this demonstration secure, we now addressed the critical coupling of **27** and **28**, shown in Scheme 5. This reaction was accomplished smoothly upon reductive cleavage of the disulfide linkage in **27**, affording **29** as shown. Aside from validating the methodology in a striking way, the synthesis of **29** serves to pinpoint the power of the total synthesis approach. Since one of its carbohydrates is unnatural, **29** cannot readily be obtained from natural sources or via enzymatic manipulations.

In summary, the goal set forth in Scheme 1—a convergent method for the synthesis of bifunctional glycopeptides—has been met. The mechanistic rationale set forth in Scheme 2 is supported by the identification of the MES-Na thioester as a reaction intermediate. Furthermore, potential hydrolysis and C-terminal epimerization of the glycopeptide acyl donor have been suppressed. It seems likely that the method and logic set forth above will enjoy application in the building of complex glycopeptides of biological and even medicinal consequence.

Supporting Information Available: Experimental procedures and compound characterization data, including LCMS and NMR data (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Imperiali, B.; O'Connor, S. E.; Hendrickson, T.; Kellenberger, C. *Pure Appl. Chem.* **1999**, *71*, 777−787. (b) Lis, H.; Sharon, N. *Eur. J. Biochem.* **1993**, *218*, 1−27. (c) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291*, 2370−2376. (d) Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357−2364.

(2) (a) Calarese, D. A.; Scanlan, C. N.; Zwick, M. B.; Deechongkit, S.; Mimura, Y.; Kunert, R.; Zhu, P.; Wormald, M. R.; Stanfield, R. L.; Roux, K. H.; Kelly, J. W.; Rudd, P. M.; Dwek, R. A.; Katinger, H.; Burton, D. R.; Wilson, I. A. *Science* **2003**, *300*, 2065−2071. (b) von Mensdorff-Pouilly, S.; Snijdewint, F. G.; Verstraeten, A. A.; Verheijen, R. H.;

Kenemans, P. *Int. J. Biol. Markers* **2000**, *15*, 343−356.

(3) (a) Dudkin, V. Y.; Miller, J. S.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 736−738. (b) Peracaula, R.; Tabares, G.; Royle, L.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M.; de Llorens, R. *Glycobiology* **2003**, *13*, 457−470.

(4) (a) Ridley, D. M.; Dawkins, F.; Perlin, E. *J. Natl. Med. Assoc.* **1994**, *86*, 129−135. (b) Durand, G.; Seta, N. *Clin. Chem.* **2000**, *46*, 795−805. (c) Koeller, K. M.; Wong, C. H. *Nat. Biotechnol.* **2000**, *18*, 835−841.

(5) Rush, R. S.; Derby, P. L.; Smith, D. M.; Merry, C.; Rogers, G.; Rohde, M. F.; Katta, V. *Anal. Chem.* **1995**, *67*, 1442−1452.

(6) Lai, P. H.; Everett, R.; Wang, F. F.; Arakawa, T.; Goldwasser, E. *J. Biol. Chem.* **1986**, *261*, 3116−3121.

(7) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(8) Cheetham, J. C.; Smith, D. M.; Aoki, K. H.; Stevenson, J. L.; Hoeffel, T. J.; Syed, R. S.; Egrie, J.; Harvey, T. S. *Nat. Struct. Biol.* **1998**, *5*, 861−866.

(9) Rahbek-Nielsen, H.; Roepstorff, P.; Reischl, H.; Wozny, M.; Koll, H.; Haselbeck, A. *J. Mass Spectrom.* **1997**, *32*, 948−958.

(10) (a) Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631−664. (b) Roth, J. *Chem. Rev.* **2002**, *102*, 285−303.

(11) Geyer, H.; Holschbach, C.; Hunsmann, G.; Schneider, J. *J. Biol. Chem.* **1988**, *263*, 11760−11767.

(12) (a) Likhosherstov, L. M.; Novikova, O. S.; Derevitskaja, V. A.; Kochetkov, N. K. *Carbohydr. Res.* **1986**, *146*, C1−C5. (b) Cohen-Anisfeld, S. T.; Lansbury, P. T. *J. Am. Chem. Soc.* **1993**, *115*, 10531−10537. (c) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779.

(13) Miller, J. S.; Dudkin, V. Y.; Lyon, G. J.; Muir, T. W.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2003**, *42*, 431−434.

(14) (a) Mandal, M.; Dudkin, V. Y.; Geng, X.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2004**, *43*, 2557−2561. (b) Geng, X.; Dudkin, V. Y.; Mandal, M.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2004**, *43*, 2562−2575.

(15) For the state of the art in biological approaches toward multiply glycosylated proteins, see: Zhang, Z.; Gildersleeve, J.; Yang, Y. Y.; Xu, R.; Loo, J. A.; Uryu, S.; Wong, C. H.; Schultz, P. G. *Science* **2004**, *303*, 371−373.

(16) Significant advances in this type of problem arose through the use of Ellman's Fmoc-based sulfonamide linker (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), which has been employed in the synthesis of glycopeptide thioesters. The method we practice here, however, insists on maximal convergence, as opposed to a "cassette" approach to glycan incorporation. We note that certain glycosidic linkages (e.g., fucosidic linkages in erythropoietin) are particularly unstable toward the acidic conditions (TFA) required for protecting group cleavage. Other "cassette"-based Fmoc solid-phase peptide synthesis (SPPS) techniques potentially leading to glycopeptide thioesters also include acidic conditions at some point. These methods involve alteration of the Fmoc deblocking conditions [(a) Li, X. Q.; Kawakami, T.; Aimoto, S. *Tetrahedron Lett.* **1998**, *39*, 8669−8672. (b) Clippingdale, A. B.; Barrow, C. J.; Wade, J. D. *J. Pept. Sci.* **2000**, *6*, 225−234. (c) Hojo, H.; Haginoya, E.; Matsumoto, Y.; Nakahara, Y.; Nabeshima, K.; Toole, B. P.; Watanabe, Y. *Tetrahedron Lett.* **2003**, *44*, 2961−2964], or direct conversion into thioesters of C-terminal acids (von Eggelkraut-Gottanka, R.; Klose, A.; Beck-Sickinger, A. G.; Beyermann, M. *Tetrahedron Lett.* **2003**, *44*, 3551−3554), esters (Swinnen, D.; Hilvert, D. *Org. Lett.* **2000**, *2*, 2439−2442), or trithioortho esters (Brask, J.; Albericio, F.; Jensen, K. J. *Org. Lett.* **2003**, *5*, 2951−2953). Boc-based SPPS employs strongly acidic cleavage conditions (liquid HF) that are incompatible with many glycosidic linkages.

(17) A phenolic ester may function as an acylating agent when employed in conjunction with an intramolecularly disposed amine equivalent that is generated in situ via Staudinger chemistry. For examples, see: (a) Nilsson, B. L.; Kiessling, L. L.: Raines, R. T. *Org. Lett.* **2000**, *2*, 1939−1941. (b) Saxon E.: Armstrong, J. I.; Bertozzi, C. R. *Org Lett.* **2000**, *2*, 2141−2143. (c) Bianchi, A.; Bernardi, A. *Tetrahedron Lett.* **2004**, *45*, 2231−2234.

(18) See: Verhaeghe, J.; Lacassie, E.; Bertrand, M.; Trudelle, Y. *Tetrahedron Lett.* **1993**, *34*, 461−464 and references therein.

(19) Caserio, M. C.; Fisher, C. L.; Kim, J. K. *J. Org. Chem.* **1985**, *50*, 4390−4393.

(20) The protected peptide acid was synthesized by Fmoc SPPS.

(21) Glycopeptide **18** was prepared as described previously (ref 13).

(22) *O*-Linked glycopeptide precursors to **20** and **21** were synthesized using a cassette approach, in which the Fmoc serine monomers used in SPPS contained pendant saccharides.

(23) Kemp, D. S. In *The Peptides: Analysis, Synthesis, Biology*; Gross, E., Meienhofer, J., Eds.; Academic Press: New York, 1979; Vol. 1, Part A, pp 315−381.

(24) Shaltiel, S.; Fridkin, M. *Biochemistry* **1970**, *9*, 5122−5127.

(25) Wang, Z. G.; Zhang, X. F.; Live, D.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2000**, *39*, 3652−3656.

JA0491836