Bioconjugate Chemistry

pubs.acs.org/bc Article

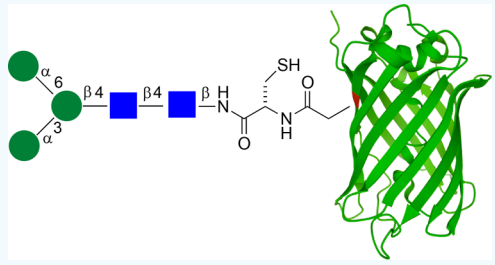
Site-Specific Synthesis of Cysteine-Bridged Glycoproteins via Expressed Protein Glycoligation

Nicholas Holloran, Daniel Collins, Upendra Rathnayake, Bixia Zhang, Minseob Koh, ChulHee Kang, and Philip Garner*

Cite This: *Bioconjugate Chem.* 2020, 31, 2362−2366 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Site-specific glycosylation of a functional recombinant protein thioester is reported. The thioester functionalized protein sfGFP-Y151ThioD, prepared by genetic code expansion, underwent native chemical ligation with the cysteine-conjugated glycans H-Cys-NH-GlcNAc and H-Cys-NH-$(GlcNAc)_2(Man)_3$ to give the corresponding cysteine-bridged glycoproteins. The intact glycoproteins, which retained their fluorescence, were characterized by top-down mass spectrometry and gel electrophoresis. The bridging cysteine provided a convenient handle for affinity chromatography purification of the glycoproteins via a removable biotin tag. Given the influence that specific glycoforms can have on a protein's function, the ability to attach a homogeneous glycan to an intact protein in a functional group controlled yet sequon-independent manner could find widespread application. These preliminary results set the stage for development of the expressed protein glycoligation (EPG) concept.



INTRODUCTION

Protein glycosylation is a ubiquitous post- (and co-) translational modification that occurs across all three domains of life. The attachment of a particular glycan to a protein can stabilize its tertiary structure, mediate protein recognition, and modulate the protein's function.[1] At the organismic level, protein glycosylation patterns influence developmental processes, help orchestrate immune responses and, when aberrant, can lead to various pathologies.[2] Glycosylation also plays a role in therapeutic scenarios such as biologic drug development, for example, where specific Fc glycoforms exert control over the effector functions of therapeutic antibodies[3] and where hyperglycosylation improves the pharmacokinetics of darbepoetin alpha resulting in a longer serum half-life for the drug.[4]

In spite of advances in the field, much remains to be learned about the biological roles and therapeutic utilization of protein glycosylation. Progress has been hampered by insufficient access to homogeneous glycoproteins. The biosynthesis of glycoproteins is a complex process that involves multiple enzymes and typically produces proteins that are glycosylated at specific sites but that bear various glycoforms (termed microheterogeneity). While glycan mixtures may be important for the expression of a particular biological effect, they are not conducive to research that requires pure, characterized substances or for use as drugs, making the case for new methodology that would produce defined glycoproteins.

Strategies that are currently employed for the preparation of homogeneous glycoproteins include chemical synthesis via native chemical ligation (NCL) of preformed glycopeptide segments (Figure 1A)[5,6] as well as chemoenzymatic synthesis

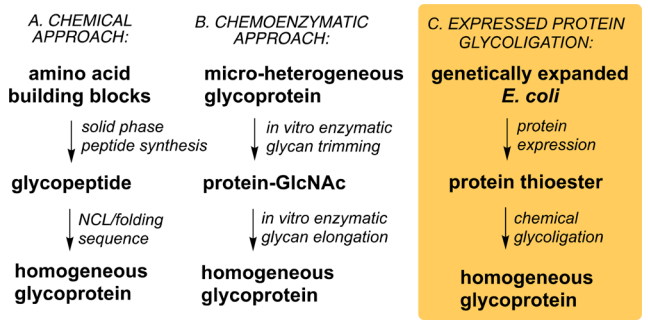
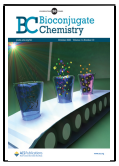


Figure 1. Synthetic approaches to glycoproteins.

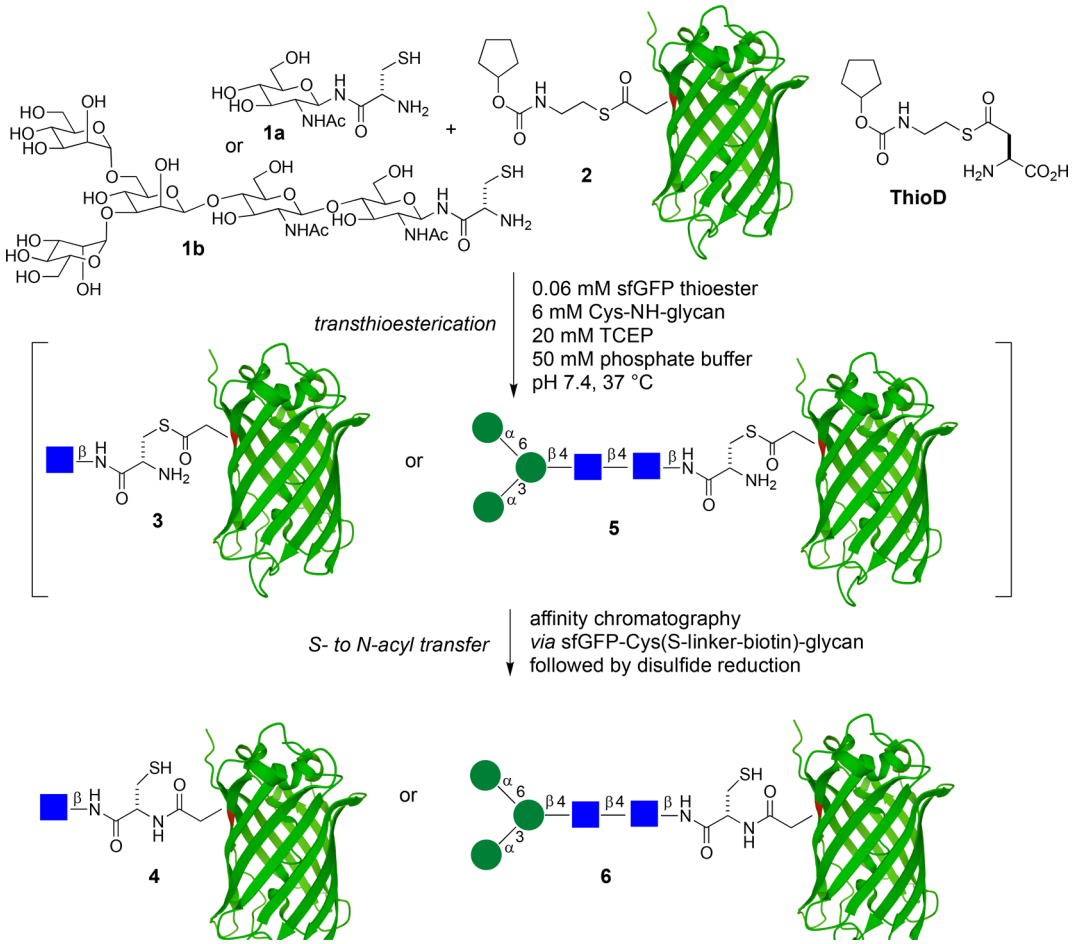
(Figure 1B)[7] via the trimming of native microheterogeneous glycan mixtures and glycan remodeling using mutant N-endoglycosidases (ENGases) and oxazolines. Progress has also been made reengineering the glycoprotein biosynthesis machinery to function in a cell-free environment.[8] However, each of these approaches have drawbacks associated with them. The chemical approach is labor-intensive and requires the folding of a denatured glycoprotein, while the mutant ENGases

Received: August 3, 2020
Revised: August 30, 2020
Published: September 15, 2020



https://dx.doi.org/10.1021/acs.bioconjchem.0c00437
Bioconjugate Chem. 2020, 31, 2362−2366

Scheme 1. Expressed Protein Glycoligation



that are involved in the chemoenzymatic approach often lack substrate generality.

Ideally, one would like to be able to attach a discrete glycan to any specific site on the surface of any folded recombinant protein via an easy-to-perform chemical reaction that involves the union of complementary orthogonal functional groups positioned on each partner. Examples of this approach to glycoprotein synthesis include introduction of an "aldehyde tag" by mutagenesis and its reaction with aminoxy glycans to give oxime ether-bridged glycoprotein mimics.[9] The "aldehyde tag" takes the form of a formylglycine (fG) residue, which is introduced by the conversion of a C*x*P*x*R sequence (*x* is variable) to fG*x*P*x*R by coexpressed formylglycine generating enzyme (FGE). The site-specific incorporation of unnatural amino acids (UAAs) bearing orthogonal functional groups into a protein via codon reassignment provides a more general means to synthesize glycoproteins. For example, the UAA *p*-acetylphenylalanine has been used to prepare oxime ether-bridged glycoprotein mimics.[10] Azide- and alkyne-containing proteins, available by usurping the Met codon, together with their complimentarily functionalized glycans actuate a Cu(I)-catalyzed Huisgen cycloaddition to link protein and glycan.[11] Azidoproteins also participate in traceless Staudinger ligations producing glycoproteins with a reverse amide linker.[12] However, this reaction is incompatible with both free and alkylated cysteine residues. With the exception of the last example, the resulting linkers are more hydrophobic in character than the native construct, though the implications of this structural difference on biological activity remains speculation.

All of these examples—as well as our work described in this paper—produce glycoprotein mimics that do not possess the native glycan–protein linkage. The fact that such structural differences can influence the biological activity of the glycoprotein helps to define a longer-range goal of this research: a general chemical ligation that produces glycoproteins with native linkers. This is an undeveloped area, though we do note that codon reassignment can be used for direct incorporation of O-glycosylated amino acids to give O-glycoproteins.[13]

We now report that the convergent synthesis of glycoproteins is feasible via the native chemical ligation of a cysteine terminated glycan and a recombinant protein functionalized with an internal thioester (Figure 1C). The approach takes advantage of our recently disclosed ability to genetically incorporate a thioester-containing aspartic acid derivative into proteins using amber suppression[14] and marries this technology with NCL.[15] The resulting expressed protein glycoligation (EPG) builds on—but is distinguished from—the work of Tolbert and Wong,[16] who used NCL to attach H-Cys-GlcNAc to an intein-derived C-terminal protein thioester. Seeberger and co-workers used an analogous approach in their semisynthesis of a cysteine-bridged GPI-anchored protein.[17] The glycoligation protocol described here (Scheme 1) produces a homogeneous N-glycoprotein with a protein–

https://dx.doi.org/10.1021/acs.bioconjchem.0c00437
Bioconjugate Chem. 2020, 31, 2362–2366

cysteine−glycan linkage. In theory, the glycan can be incorporated at any permissible site on the surface of a protein.

RESULTS AND DISCUSSION

Superfolder green fluorescent protein (sfGFP)[18] was chosen for this demonstration in part because it allows one to visually assess that the tertiary structure of the protein is retained as fluorescence is lost upon denaturing. It also shows that the glycoligation technology does not require the presence of an N-glycosylation sequon. Protein thioester **2** was produced according to the reported procedure,[9] which is described here in more detail. Basically, *E. coli* DH10B carrying two plasmids (one for an amino acyl tRNA synthetase/$tRNA_{CUA}$ pair, and one for sfGFP-Y151TAG) was grown in the presence of carbenicillin to OD_{600} of 0.6. Expression of the plasmids was induced by adding isopropyl-β-D-1-thiogalactopyranoside (IPTG) and ThioD to produce sfGFP-Y151ThioD. The cells were lysed by sonication, and the protein was purified by Ni-NTA affinity chromatography. Further purification could be achieved by ion exchange chromatography.

The glycan−cysteine conjugates **1a** and **1b** were synthesized from the corresponding glycosylamines in a manner similar to that described by Wong.[16] The pentasaccharide glycan Manα3(Manα6)Manβ4GlcNAcβ4GlcNAc-OH (the core sequence found in eukaryotic N-glycoproteins) needed for **1b** was synthesized following a route reported by Walczak and Danishefsky.[19] The free glycans with a hydroxyl at the anomeric position were subjected to Kotchetkov amination conditions[20] to give the glycosylamines. The respective glycosylamines were coupled to a Boc/Trityl protected cysteine via DIC coupling, followed by global deprotection using TFA to give **1a** and **1b** as TFA salts.

Compound **1a** was subjected to native chemical ligation conditions with the crude protein thioester **2** in 50 mM phosphate buffer at pH 7.4 in the presence of 20 mM TCEP at 37 °C. The glycoligation was allowed to proceed for 20 h, during which time the product **4** was produced via rearrangement of the intermediate thioester **3**. Although the crude starting protein thioester contained a minor amount of the corresponding glutathione thioester, it was unnecessary to remove this side product, since it can also undergo transthioesterification to give **3**. However, a better yield of glycoprotein was obtained when pure **2** was used. The glycoligation reaction was monitored by top-down mass spectrometry (Figure 2)[21] which showed the main competing reaction to be hydrolysis of the protein thioester(s). This was not unexpected in light of sfGFP-Y151ThioD stability studies (ref 14, Supporting Information, Figure S3), which showed that **2** had a half-life of ∼20 h at pH 7.4. Thus, the rate of hydrolysis was on the same order as a typical NCL reaction. The hydrolysis side product was removed by affinity chromatography. The reaction mixture was dialyzed and buffer exchanged into water to remove all nonprotein compounds (especially the excess **1a**, which would interfere with the affinity chromatography-based purification described below).

The bridging cysteine provides a convenient handle for separation of nonglycosylated protein byproducts (principally hydrolyzed thioester) by affinity chromatography. Thus, the protein mixture was treated with *N*-[6-(biotinamido)hexyl]-3′-(2′-pyridyldithio)propionamide (biotin HPDP) in 20 mM phosphate buffer at pH 8.0 and the resulting biotin-containing glycoprotein adsorbed onto an avidin column yielding homogeneous Cys-bridged glycoprotein upon reductive release

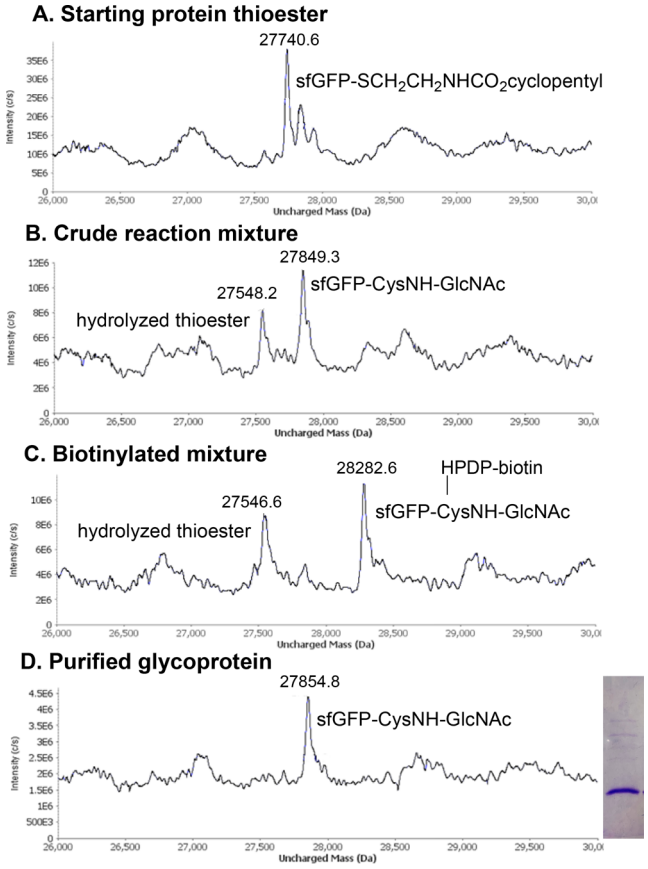


Figure 2. Expressed protein glycoligation and product purification as assessed by top-down ESI-MS. Panel A: Unpurified starting protein thioester **2**. Panel B: Crude glycoligation reaction mixture. Panel C: Reaction mixture after biotinylation. Panel D: MS and SDS-PAGE of purified glycoprotein **4** after affinity chromatography. The gel was visualized by Coomassie blue staining.

as evidenced by top-down MS and gel electrophoresis (Figure 2, Panel D). The Cys-bridged glycoprotein **4** was obtained in 78% overall yield from pure **2**. The product maintained its green fluorescence after the reaction, indicating that it remained folded and intact throughout this process. In a similar manner, compound **1b** was ligated to **2** to give the glycoprotein **6** via the intermediate thioester **5**.

CONCLUSIONS

We have described a protocol that enables the site-specific attachment of a defined glycan onto a recombinant protein. The linchpin glycoligation proceeds via reaction of a cysteine-conjugated glycan and a protein thioester prepared by genetic code expansion. The site-specific mutation of a single amino acid residue to ThioD may be advantageous for modification of internal amino acids in the context of maintaining protein tertiary structure when compared to the "aldehyde tag" method which requires the minimal introduction of a five-amino-acid sequon. While ThioD has a relatively hydrophobic side chain, the hydrophobic thioester is replaced by an amide during the glycoligation resulting in a more hydrophilic, native-like linker. Since the glycoligation involves a native chemical ligation mechanism, the reaction is compatible with the presence of ancillary cysteine residues. In addition to these features, the bridging cysteine residue provides a convenient handle for further functionalization of the glycoprotein. The

https://dx.doi.org/10.1021/acs.bioconjchem.0c00437
Bioconjugate Chem. 2020, 31, 2362−2366

cysteine thiol could be used to attach a payload or a spectroscopic tag to the glycoprotein. In this work, it served as a point of attachment for a biotin tag which facilitated product purification. The described method is versatile. For example, an alternative purification scheme involving lectin affinity chromatography using resin-bound concanavalin A could be used to avoid potential interference by native cysteine residues.

EXPERIMENTAL PROCEDURES

sfGFP-Y151ThioD 2. Lysogeny broth (LB, 5 mL—see Supporting Information for details) was inoculated with 5 μL of a glycerol stock solution of *E. coli* strain DH-10B carrying 2 plasmids (pUltra-ThioD and pET22b-sfGFP-Y151TAG—obtained from the Schultz lab at TSRI). Carbenicillin and spectinomycin were added as 1000× stock solutions to a final concentration of 100 μg/mL and 50 μg/mL, respectively. The cell culture was incubated with shaking at 37 °C for 20 h. The saturated culture was then diluted to 1 L with LB containing the same antibiotic concentrations. This cell culture was incubated at 37 °C with shaking until OD_{600} reached 0.6 (this was achieved after about 4 h). The culture was cooled to 23 °C, and ThioD was added, followed by IPTG (2 mM and 1 mM, respectively). The culture was incubated in an orbital shaker incubator with 220 rpm at 23 °C for another 16 h. The cells were pelleted by centrifugation (8000 × g for 30 min at 4 °C), resuspended in lysis buffer (20 mL—see Supporting Information for details) and lysed by sonication. The lysate was purified by Ni-NTA affinity chromatography by washing with a solution made up of 50 mM phosphate, 300 mM NaCl, and 20 mM imidazole (pH = 7.4) and eluting with a solution made up of 50 mM phosphate, 300 mM NaCl, and 250 mM imidazole (pH = 7.4). The protein solution was buffer exchanged by dialysis into 25 mM phosphate buffer (pH = 7.2) and concentrated by centrifugal filtration (10,000 MWCO). Protein concentration was determined by Bradford assay, and the solution was aliquoted and stored as a 20% glycerol stock at −80 °C. This procedure yielded between 4 and 6 mg of crude protein per liter of cell culture. Fluorescence indicates that the tertiary structure of the protein is intact. Deconvolution of the ESI mass spectrum showed the presence of the desired thioester (MW 27,740.6), as well as a glutathione thioester (MW 27,841.8). Only trace amounts of hydrolyzed thioester were observed (MW 27,546.6) in agreement with the literature.[14] Pure **2** could be obtained by subjecting the crude material to anion exchange chromatography (Waters Q 8HR, 10 × 100 mm column, 1000 Å pore size), using 20 mM Tris pH 7.5 as mobile phase for 5 min, followed by 0−300 mM NaCl gradient over 40 min followed by dialysis (Supporting Information, Figure S1). Our observed mass values vary slightly from the literature probably due to variable protein−ion adducts and error during deconvolution. Since this error is under the value of any reasonable fragmentation, and changes with sample purity, we are confident in our assignment of **2**.

Protein sequence: (***X*** denotes ThioD): MSKGEELF-TGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLK-FICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHD-FFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDT-LVNRIELKGIDFKEDGNILGHKLEYNFNSHNV***X***ITA-DKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGD-GPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTA-AGITHGMDELYKGSHHHHHH

GFP-Cys-NH-GlcNAc 4. To a solution of **1a** (1.6 mg, 3.6 μmol) in ligation buffer (50 mM phosphate, 20 mM TCEP, pH = 7.4, 1.0 mL) was added protein thioester **2** (1.0 mg, 36 nmol) as a stock solution in 20% glycerol. The mixture was incubated at 37 °C for 20 h and was then buffer-exchanged at 4 °C into degassed 50 mM phosphate, 5 mM EDTA (pH = 8.0). This buffer exchange must be done quickly, and at low temperature to prevent cysteine disulfide formation. To this solution was added HPDP−biotin as a stock solution (10 mM in DMSO, 100 μL). The mixture was incubated at RT for 2 h, and then buffer-exchanged into the avidin loading buffer. This mixture was purified by avidin affinity chromatography eluting with 10 mM TCEP added to the loading buffer. The purified protein solution was concentrated by centrifugal filtration and desalted for analysis by MS and SDS-PAGE (Supporting Information, Figures S2). It should be noted that the crude **4** observed in the reaction mixture and the purified **4** display slightly different masses. This is again likely due to multiple protein−ion adducts and error during deconvolution based on sample purity. Protein concentration was determined by Bradford assay and showed that 0.78 mg of **4** was obtained (78% yield over 3 steps). The product maintains fluorescence indicating that the tertiary structure of the protein remains intact through this procedure.

GFP-Cys-NH-GlcNAc$_2$Man$_3$ 6. The pentasaccharide conjugate **6** was synthesized as described above with two exceptions. First, crude **2** was used in this reaction. Second, the excess glycan was removed by ion-exchange chromatography (DEAE-cellulose eluting with a NaCl gradient from 0 to 100 mM in 50 mM tris buffer, pH = 7.4) after the ligation. This column produced crude **6** with an extra glycan attached through a disulfide bond to the cysteine linker. The disulfide could be reduced with 10 mM TCEP in 50 mM phosphate buffer (pH = 7.4). This mixture could then be buffer-exchanged into degassed 50 mM phosphate with 5 mM EDTA (pH = 8.0). The procedure was then continued as described above monitoring the reaction by LC/MS (Supporting Information, Figure S3). Due to the presence of inseparable Cys-NH-GlcNAc$_2$Man$_2$ in the starting material, glycoprotein **6** was contaminated with the corresponding tetrasaccharide glycoprotein. Compound **6** also displayed variance in the mass spectra as noted above. Bradford assay of the affinity purified **6** showed that 0.011 mg of glycoprotein was obtained. The product retained its fluorescence, indicating that the tertiary structure of the protein remains intact through this procedure.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.bioconjchem.0c00437.

General experimental information, detailed procedures, characterization data (PDF)

AUTHOR INFORMATION

Corresponding Author

Philip Garner − *Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, United States;* orcid.org/0000-0002-6503-9550; Email: ppg@wsu.edu

https://dx.doi.org/10.1021/acs.bioconjchem.0c00437
Bioconjugate Chem. 2020, 31, 2362−2366

Authors

Nicholas Holloran − *Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, United States*

Daniel Collins − *Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, United States*

Upendra Rathnayake − *Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, United States*

Bixia Zhang − *Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, United States*

Minseob Koh − *Department of Chemistry and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, United States*

ChulHee Kang − *Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.bioconjchem.0c00437

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by National Science Foundation grant CHE-1665274. Thanks to Peter Schultz (TSRI) for his helpful assistance on this project and to Pete Reilly (WSU) for fruitful discussions about top-down mass spectrometry.

■ REFERENCES

(1) *Essentials of Glycobiology*, Varki, A., Ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009.

(2) Lauc, G., Pezer, M., Rudan, I., and Campbell, H. (2016) Mechanisms of disease: The human N-glycome. *Biochim. Biophys. Acta, Gen. Subj. 1860*, 1574−1582.

(3) Schwab, I., and Nimmerjahn, F. (2013) Intravenous immunoglobulin therapy: How does IgG modulate the immune system? *Nat. Rev. Immunol. 13*, 176−189.

(4) Sinclair, A. M., and Elliott, S. (2005) Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *J. Pharm. Sci. 94*, 1626−1635.

(5) Wang, P., Dong, S., Shieh, J.-H., Peguero, E., Hendrickson, R., Moore, M. A. S., and Danishefsky, S. J. (2013) Erythropioetin derived by chemical synthesis. *Science 342*, 1357−1360.

(6) Murakami, M., Kiuchi, T., Nishihara, M., Tezuka, K., Okamoto, R., Izumi, M., and Kajihara, Y. (2016) Chemical synthesis of erythropoietin glycoforms for insights into the relationship between glycosylation pattern and bioactivity. *Sci. Adv. 2*, e1500678.

(7) Li, C., and Wang, L.-X. X. (2018) Chemoenzymatic methods for the synthesis of glycoproteins. *Chem. Rev. 118*, 8359−8413.

(8) Jaroentomeechai, T., Stark, J. C., Natarajan, A., Glasscock, C. J., Yates, L. E., Hsu, K. J., Mrksich, M., Jewett, M. C., and Delisa, M. P. (2018) Single-pot glycoprotein biosynthesis using a cell free transcription-translation system enriched with glycosylation machinery. *Nat. Commun. 9*, 1−11.

(9) Hudak, J. E., Yu, H. H., and Bertozzi, C. R. (2011) Protein glycoengineering enabled by the versatile synthesis of aminooxy glycans and the genetically encoded aldehyde tag. *J. Am. Chem. Soc. 133*, 16127−16135.

(10) Liu, H., Wang, L., Brock, A., Wong, C. H., and Schultz, P. G. (2003) A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc. 125*, 1702−1703.

(11) Van Kasteren, S. I., Kramer, H. B., Jensen, H. H., Campbell, S. J., Kirkpatrick, J., Oldham, N. J., Anthony, D. C., and Davis, B. G. (2007) Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature 446*, 1105−1109.

(12) Bernardes, G. J. L., Linderoth, L., Doores, K. J., Boutureira, O., and Davis, B. G. (2011) Site-selective traceless Staudinger ligation for glycoprotein synthesis reveals scope and limitations. *ChemBioChem 12*, 1383−1386.

(13) Fahmi, N. E., Dedkova, L., Wang, B., Golovine, S., and Hecht, S. M. (2007) Site-specific incorporation of glycosylated serine and tyrosine derivatives into proteins. *J. Am. Chem. Soc. 129*, 3586−3597.

(14) Xuan, W., Collins, D., Koh, M., Shao, S., Yao, A., Xiao, H., Garner, P., and Schultz, P. G. (2018) Site-specific incorporation of a thioester containing amino acid into proteins. *ACS Chem. Biol. 13*, 578−581.

(15) Conibear, A. C., Watson, E. E., Payne, R. J., and Becker, C. F. W. (2018) Native chemical ligation in protein synthesis and semi-synthesis. *Chem. Soc. Rev. 47*, 9046−9068.

(16) Tolbert, T. J., and Wong, C.-H. (2000) Intein-mediated synthesis of proteins containing carbohydrates and other molecular probes. *J. Am. Chem. Soc. 122*, 5421−5428.

(17) Becker, C., Liu, X., Olschewski, D., Castelli, R., Seidel, R., and Seeberger, P. (2008) Semisynthesis of a glycosylphosphatidylinositol-anchored prion protein. *Angew. Chem., Int. Ed. 47*, 8215−8219.

(18) Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol. 24*, 79−88.

(19) Walczak, M. A., and Danishefsky, S. J. (2012) Solving the convergence problem in the synthesis of triantennary N-glycan relevant to prostate-specific membrane antigen (PSMA). *J. Am. Chem. Soc. 134*, 16430−16433.

(20) Likhosherstov, L. M., Novikova, O. S., Derevitskaja, V. A., and Kochetkov, N. K. (1986) A new simple synthesis of amino sugar β-D-glycosylamines. *Carbohydr. Res. 146*, C1−C5.

(21) Cui, W., Rohrs, H. W., and Gross, M. L. (2011) Top-down mass spectrometry: Recent developments, applications and perspectives. *Analyst 136*, 3854−3864.