HHS Public Access

Author manuscript

Methods Enzymol. Author manuscript; available in PMC 2019 June 27.

Published in final edited form as:

Methods Enzymol. 2019 ; 621: 213–229. doi:10.1016/bs.mie.2019.02.030.

Chemical Biology of Glycoproteins: From Chemical Synthesis to Biological Impact

Yaohao Li[1,2], **Amy H. Tran**[2], **Samuel J. Danishefsky**[3,*], and **Zhongping Tan**[1,2,*]

[1]Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100050, China.

[2]Department of Biochemistry and BioFrontiers Institute, University of Colorado, Boulder, CO 80309, United States

[3]Chemical Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY, 10065, United States

Abstract

Recent advances have demonstrated the feasibility and robustness of chemical synthesis for the production of homogeneously glycosylated protein forms (glycoforms). By taking advantage of the unmatchable flexibility and precision provided by chemical synthesis, the quantitative effects of glycosylation were obtained using chemical glycobiology approaches. These findings greatly advanced our fundamental knowledge of glycosylation. More importantly, analysis of these findings has led to the development of glycoengineering guidelines for rationally improving the properties of peptides and proteins. In this chapter, we present the key experimental steps for chemical biology studies of protein glycosylation, with the aim of facilitating and promoting research in this important but significantly underexplored area of biology.

Keywords

Chemical Synthesis; NCL/MFD; Glycoproteins; Chemical Glycobiology; Glycoengineering

1. INTRODUCTION

Protein glycosylation is the most common form of posttranslational modification (PTM) on excreted and extracellular membrane-associated proteins.(Spiro, 2002) It involves the covalent attachment of many different types of glycans (also called carbohydrates, saccharides, or sugars) to a protein. Based on the amino acid side-chain atoms to which glycans are linked, protein glycosylation can be divided into two major categories: N-glycosylation and O-glycosylation. In N-glycosylation, glycans are attached to the side-chain nitrogen atoms of Asn residues in a conserved consensus sequence Asn–Xaa–Ser/Thr (Xaa ≠ Pro), whereas in O-glycosylation, glycans are attached to the side-chain oxygen atoms of hydroxyl amino acids, primarily serine and threonine residues (Fig. 1).

*Corresponding authors: danishes@mskcc.org, zhongping.tan@imm.pumc.edu.cn.

Over the past several decades, research from many disciplines has established the importance of protein glycosylation (Varki, 2017). More and more evidence has pointed to the fact that glycosylation can greatly diversify the functions of therapeutically important proteins(Anthony, Wermeling, & Ravetch, 2012) and industrial enzymes(Greene, Himmel, Beckham, & Tan, 2015), possibly achieved by adjusting their physicochemical and/or biological properties (Moremen, Tiemeyer, & Nairn, 2012). Aberrant glycosylation has been implicated as a significant contributor to many human diseases and to reduced (or even completely abolished) enzyme activity (Dwek, 1995; Meany & Chan, 2011). However, despite its significance, protein glycosylation has long remained understudied, and thus underutilized. This lagging behind is mainly due to the lack of efficient strategies to obtain large collections of glycosylated protein forms (glycoforms) with defined, systematically varied structures in a sufficient amount for characterization (Rich & Withers, 2009). The absence of such strategies is a result of the complexity of glycosylation biosynthetic pathways and the inseparable nature of glycoforms (Rudd & Dwek, 1997).

Depending on the exact glycan structure, glycosylation is initiated and proceeds variously at different cellular locations and/or different stages of protein synthesis (Ohtsubo & Marth, 2006). Like most enzymatic processes, it is regulated by many different host and environmental factors, such as conformation at potential glycosylation sites, local amino acid sequence, and availability of activated sugar donors and enzymes. Under the combined influence of these highly variable factors, glycoproteins synthesized by cells usually exist as complex mixtures of up to hundreds of glycoforms (or glycosylated variants, glyco-variants). These variants differ in glycosylation site occupancy and glycan structures (giving rise to macroheterogeneity and microheterogeneity, respectively) and are often impossible to separate from one another. As a result, most previous studies had to rely on mixture samples with unknown compositions (Rudd & Dwek, 1997). Not surprisingly, the uncertainty in sample composition has led to inconsistent and even conflicting conclusions about protein glycosylation, which hampered further progress in this area (Varki, 1993, 2017).

The study of homogenous glycoforms with defined structures is key to drawing strong and valid conclusions about protein glycosylation (P. Wang et al., 2013). Similar to the studies of other biopolymers like nucleic acids and proteins, the effects of glycosylation are expected to be precisely established by comparing the properties of a series of representative glycoforms with systematically varied glycans (P. K. Chaffey, Guan, Li, & Tan, 2018). Since it is currently not possible to separate and isolate individual glycoforms from biological sources, alternative methods must therefore be developed. Many biochemical methods, including glycosylation pathway engineering, glycan remodeling and enzymatic glycosylation, have been explored for the purpose of generating homogeneous glycoforms (Rich & Withers, 2009). However, although these methods are more practical means to prepare large glycoproteins, and to produce glycoproteins on preparative scales, they are severely limited with respect to the generation of large collections of glycoforms with structurally closely related glycan structures. Their intrinsic lack of flexibility thus makes them not suitable to address the issues in protein glycosylation research.

In recent years, chemical synthesis has evolved to become a core method for preparing homogeneous glycoforms (Wilson, Dong, Wang, & Danishefsky, 2013). Unlike

biosynthesis, chemical glycosylation is not subject to the restrictions imposed by substrate specificity. Therefore, chemical synthesis is a highly flexible and precise means by which to access glycoforms with diverse structural variations. Such advantages make chemical synthesis an attractive method to help reach a comprehensive and definitive understanding of protein glycosylation. While useful and promising, chemical synthesis also has its drawbacks. Chemical synthesis generally involves labor-intensive, sometimes tedious preparation of starting materials and intermediates. In order to obtain a large collection of homogenous glyco-variants in a reasonable amount of time through chemical synthesis, innovation in the form of synthetic strategy, tactics and method invention is required.

Chemical synthesis of glycoproteins generally involves the ligation of a set of chemically prepared peptide and glycopeptide fragments into a long polyglycopeptide chain, which then folds up to form targeted glycoproteins (Fig. 2). Peptide fragments can be readily prepared using well-developed Solid-Phase Peptide Synthesis (SPPS) methods. The most challenging tasks in the synthesis of glycoproteins are the preparation of glycopeptide fragments and their assembly. In order to achieve highly effective synthesis of glycoproteins, the Danishefsky group has systematically optimized the synthetic strategies and methodologies in these two key areas (Tan, Shang, & Danishefsky, 2011;P. Wang et al., 2012). The success of this effort paves the way for the production of glycoforms with diverse structural variations and forms the essential basis for developing a molecular-level understanding of protein glycosylation.

With the newly increased sophistication of glycoprotein synthesis, the Tan group was able to apply a chemical glycobiology approach to investigate the role and mechanism of glycosylation in regulating the properties and function of proteins (P. K. Chaffey et al., 2018). Their efforts established the general workflow of this approach as well as demonstrated its feasibility and effectiveness to provide new definitive insights into the biochemical and biological impacts of protein glycosylation. Most importantly, the results from these efforts have greatly advanced the application of rational glycoengineering to the optimization of therapeutic peptides and proteins.

This chapter will review the above-mentioned innovative developments in protein glycosylation research. We will first review the chemical synthesis of glycoproteins, and then describe the implementation of chemical glycobiology approaches to explore and exploit the beneficial effects of protein glycosylation.

2. CHEMICAL SYNTHESIS OF HOMOGENEOUS GLYCOFORMS

There are two general strategies for the synthesis of glycopeptides: one is a convergent strategy, where sugars are directly coupled to appropriately-protected peptides to generate glycopeptides; the other is a cassette-based strategy, where pre-formed glycoamino acids (so-called “cassette”) are used as building blocks in synthesis (Fig. 2). Most N-linked glycopeptides can be prepared convergently by joining glycosylamines (glycans bearing anomeric amines) to aspartate residues (Fig. 2A), while most O-linked glycopeptides need to be prepared using the stepwise cassette approaches to ensure that the stereo-purity of the products is rigorously controlled (Fig. 2B) (Tan, Shang, Halkina, Yuan, & Danishefsky,

2009). However, regardless of the strategy selected, the chemical synthesis of glycans and glycoamino acids is the most challenging, but unavoidable step. The Danishefsky group did a significant amount of optimization and provided better synthetic routes for many N-linked glycans as well as O-GalNAc glycoamino acids (Sames, Chen, & Danishefsky, 1997; Wu et al., 2006). With these notable achievements, it became possible to conduct chemical synthesis of highly complex glycopeptide fragments.

The Danishefsky group exploited a number of synthetic advances toward the creation of glycopeptides (Wilson et al., 2013). Their endeavors along this direction led to the development of a range of broadly useful methods. They established that the unprotected oligosaccharides could undergo anomeric β-amination under Kochetkov conditions to produce glyco sylamines, which in turn can be directly added to peptides using the Lansbury aspartylation protocol to yield N-glycopeptide fragments (Fig. 2A) (Z. G. Wang, Zhang, Live, & Danishefsky, 2000). They also creatively addressed the challenges met in the synthesis of O-linked glycopeptides using an optimized “mask-off” approach by exploiting differences in reaction rates caused by steric effects of different carboxylic groups in unprotected O-glycoamino acids (Fig. 3) (Tan et al., 2009). Through these optimized preparation methods, they were able to complete the synthesis of many different glycopeptides, including both N- and O-linked glycopeptides (Wilson et al., 2013).

After the synthesis of glycopeptide fragments, the production of multiply glycosylated protein targets would require the development of capabilities in one more key research area. Through more than 10 years of efforts, the Danishefsky group introduced to this area a generally applicable and powerful new method: native chemical ligation coupled to metal-free desulfurization (NCL/MFD) (Shang, Tan, Dong, & Danishefsky, 2011; Tan, Shang, & Danishefsky, 2010; Wan & Danishefsky, 2007). Traditionally, NCL is a process in which two peptide or glycopeptides fragments are chemoselectively joined together by reactions of C-terminal thioesters with N-terminal cysteine residues. One of its major limitations is the requirement of cysteine residues at ligation junctions, which are rare or absent in most proteins and glycoproteins. To address this challenge, the Danishefsky group greatly extended the application scope of NCL to many abundant amino acid residues (P. K. Chaffey et al., 2018; Shang, Tan, & Danishefsky, 2011; Wilson et al., 2013). To overcome the synthesis issues associated with the removal of thiol groups that were introduced into the side chains of N-terminal amino acids to catalyze the ligation reactions, they contributed one of the most impactful developments in protein/glycoprotein synthesis, the metal-free desulfurization strategy (Fig. 4) (Wan & Danishefsky, 2007).

The optimized methods described above have been applied to the synthesis of many different glycoproteins (Fernandez-Tejada et al., 2015). Depending on the structural features of the synthetic targets, there are small differences in experimental details. They can be easily found in previously published research. Here, we only briefly describe the most commonly used experimental techniques and tips that are critical for experimental success.

Equipment

1. NMR Instruments: Bruker Advance DRX 500 MHz and Bruker Advance II 600 MHz

2. Analytical TLC: Merck silica gel 60 F254
3. Flash Chromatography: Merck silica gel 60 (40–63 mm)
4. Solvent Drying System
5. Applied Biosystems Pioneer continuous flow peptide synthesizer
6. LC-MS System: Waters 2695 Separations Module, Waters Micromass ZQ mass spectrometer, Waters 996 Photodiode Array Detector, and Varian Microsorb columns 100–5, C18, 150 × 2.0 mm and 300–5, C4, 250 × 2.0 mm
7. UPLC-MS System: Waters Acquity™ Ultra Performance LC system and Waters Acquity UPLC® BEH columns C18, 1.7 μl, 2.1 × 100 mm, C8, 1.7 μl, 2.1 × 100 mm, and C4, 1.7 μl, 2.1 × 100 mm
8. Preparative HPLC system: Ranin HPLC solvent delivery system, Rainin UV-1 detector, and Varian Dynamax Microsorb columns 100–5, C18, 250 × 21.4 mm and 300–5, C4, 250 × 21.4 mm
9. VirTis Bench Top 4.0 K ZL Freeze Dryer (SP Industries, NY, USA)
10. Eppendorf 5430R Microcentrifuge
11. Analytical Balance Mettler Toledo XS205DU
12. Bath Sonicator

Chemicals

1. Carbohydrate Building Blocks
2. Amino Acid Building Blocks
3. Pseudoproline Dipeptides
4. Preloaded NovaSyn ®TGTresin
5. Peptide Coupling Reagents
6. Desulfurization Reagents

2.1 Synthesis of N-linked glycopeptides

The preparation of N-linked glycopeptides began with the synthesis of N-glycans, such as a complex-type glycosylamine **13** (Walczak & Danishefsky, 2012; Wu et al., 2006). As shown in Figure 5, this undecasaccharide can be synthesized convergently from a pentasaccharide **9** and a trisaccharide **10**. Briefly, the synthesis commenced with the preparation of a trisaccharide **7** through the application of the Crich–Kahne direct mannosylation protocol. Next, attachment of the mannose thioethyl donor **8** to **7** provided pentasaccharide **9**. This intermediate was then subjected to a double glycosylation with anomeric fluoride to afford the fully protected undecasaccharide **11**, which in turn was deprotected through a six-step sequence to generate the unprotected glycan **12**. The Kochetkov amination of 12 followed by repetitive lyophilization gave the glycosylamine **13**. Finally, the synthetic glycosyl amines

were coupled to peptides in the presence of HATU to afford the desired glycopeptides (Fig. 2A).

Tip: The coupling efficiency of the sialylated trisaccharide **10** to the pentasaccharide **9** or other oligosaccharides is dependent on the reactivity of the anomeric donor group. If the yield of this step is low, the first thing for troubleshooting should be to test more donor groups (Walczak & Danishefsky, 2012).

Tip: The deprotection of fucosylated N-linked oligosaccharides can result in partial cleavage of glycan chains. Special attention needs to be paid to the NaOMe/NaOH-mediated hydrolysis step (Nagorny et al., 2009).

Tip: The purity of glycosyl amine products is important for their reactivity in Lansbury aspartylation reaction. In order to completely remove impurities, the lyophilization needs to be performed many times until a constant weight is obtained.

Tips: Glycosyl amine products are sensitive to H2O. When dissolved in H2O and frozen for the next lyophilization cycle, they may decompose. Therefore, this step must be done very quickly, in less than 10 seconds.

Tip: The Lansbury aspartylation reaction is a bimolecular reaction and its rate is dependent on the concentrations of both reactants. Therefore, high reactant concentrations are crucial for the success of this reaction.

2.2 Synthesis of O-linked glycopeptides

The cassette approach was applied to prepare O-linked glycopeptides. In a typical synthesis, fully-protected glycoamino acids were convergently prepared under optimized reaction conditions (Fig. 3). After deprotection, the HATU-mediated peptide coupling reaction between unprotected glycoamino acids and short peptides proceeded very selectively to afford the desired O-linked glycopeptide products (Fig 2B) (Tan et al., 2009).

Tip: The HATU-mediated peptide coupling reaction occurs quite rapidly. It is normally completed within 1 minute.

Tip: Under HATU-mediated conditions, the carboxylic groups in glycoamino acids can be clearly differentiated on the basis of the substitution levels of their β-carbons. The formation of lactone between the sialic acid carboxyl group and the hydroxyl group in position 4 of the adjacent galactose is another reason for the selectivity.

Tip: The lactone formation can also beneficially prevent the cleavage of the labile α−2,3-glycosidic bond between sialic acid and galactose. The lactone ring can be opened by dilute sodium bicarbonate solution after the completion of glycoprotein synthesis.

Tip: The by-products formed by reacting with sialic acid carboxyl groups can be identified through the analysis of their fragmentation patterns in electrospray ionization-mass spectrometry.

2.3 Ligation and Metal-Free Desulfurization

The chemical assembly of the synthetic fragments was carried out in the direction of C-terminus to N-terminus as shown in Figure 2C. Briefly, the to-be-ligated peptides and glycopeptides were mixed together and lyophilized. The resulting powder was dissolved in buffer solution to initiate the ligation reaction. After ligation was completed, the thiol functional groups were removed under metal-free desulfurization conditions to yield the full-length glycoproteins, which were then oxidatively folded using a redox buffer (Shang, Tan, & Danishefsky, 2011; P. Wang et al., 2013).

Tip: Native chemical ligation reactions are bimolecular reactions. The reaction rates are influenced by concentration and other factors like the availability of free thiol groups at the N-terminal sides of peptides and glycopeptides, pH, and temperature. To ensure high concentrations of reactants, the lyophilized reactant powder mixtures should be centrifuged to the bottom of reaction vessels before a minimum required amount of buffer solution is added. Since the rate-limiting step in most NCL is the transthioesterification with the side-chain thiol groups of the N-terminal amino acids, it is important to include soluble reducing agents, like TCEP in degassed ligation buffer, to cleave disulfide bonds at the beginning of ligation and to prevent the oxidation of thiols to disulfides during the reaction. Furthermore, for intermolecular transthioesterification to occur, the proton on the thiol group must be transferred to a putative “base” or to the bulk solvent. Therefore, the desired pH of the buffer solution should be slightly alkaline (~7.5). Ligation rates also increase with increasing temperature. If the ligation is slow, the reaction temperature can be increased to 37°C.

Tip: After ligation, the free thiol groups in the middle of the resulting peptides/glycopeptide products can further react with excess thioesters to form higher molecular weight species. The presence of these species can significantly complicate the analysis of reaction progress. This issue can be overcome by adding the alkyl thiolsodium 2-mercaptoethanesulfonate (MESNa) to reaction mixtures.

Tip: The presence of excess thiol additives and large amount of guanidine prevents direct desulfurization of ligation products. An efficient MFD can only be realized after removing these substances with a rapid and simple HPLC separation.

Tip: A large portion of ligation product can be lost during HPLC purification. Therefore, less purification steps are required for a more efficient and higher yielding synthesis.

Tip: Correctly folded proteins/glycoproteins normally eluted earlier on HPLC than other folding isomers. Correct folding can be verified by mass spectrometry analysis of enzymatically digested glyoprotein fragments.

3. BIOLOGICAL IMPACT OF PROTEIN GLYCOSYLATION

The development of chemical synthesis of multiply glycosylated proteins by the Danishefsky group provided a sufficient scientific and technical base for overcoming challenges associated with research on protein glycosylation. It enabled the Tan group to apply chemical glycobiology approaches to further the understanding of the roles of glycans

in regulating protein biophysical and biological properties (Patrick K. Chaffey, Chi, & Tan, 2017; P. K. Chaffey et al., 2018).

The predictability and flexibility of chemical synthesis allowed the generation of large collections of structurally well-defined glycoproteins with subtle differences in their glycan structures and amino acid sequences. The availability of such collections makes it possible to get a clearer and more in-depth view of the biological impact of glycosylation (Chen et al., 2014). By systematically analyzing the impact, the Tan group was able to dissect detailed structure-property relationships and to gain insights into the structural basis for the inferred relationships (Xiaoyang Guan et al., 2015). More importantly, the knowledge developed from these studies helps to formulate a directional strategy that can guide future studies towards other posttranslational modifications, and to improve the properties and performance of peptides/proteins through rational glycoengineering (X. Guan, Chaffey, Chen, et al., 2018; X. Guan, Chaffey, Wei, et al., 2018; Li, Clark, & Tan, 2018).

In the following section, we briefly summarize these studies and the findings relating to the biological impact of protein glycosylation.

Equipment

1. Chirascan™- plus CD Spectrometer
2. Voyager-DE™ STR MALDI-TOF Mass Spectrometer
3. Waters Synapt G2 HDMS (q-TOF) Mass Spectrometer
4. Proteomelab XL-I Protein Characterization System
5. 48-Well Micro Chemotaxis Chamber
6. QuantaMaster-4/2005 SE Spectrofluorometer

Chemicals

1. Proteases
2. Chemicals for biochemical and biological buffers

3.1 Impact on Folding

The process of protein folding is a notoriously complicated process. For proteins containing disulfide bonds, it is a combination of two distinct, but interrelated processes: conformational folding of the protein sequence into secondary and tertiary structures and formation of disulfide bonds. By taking advantage of the relatively slow kinetics in disulfide bond formation, the Tan group was able to confirm the importance of glycosylation in regulating the folding of a family 1 carbohydrate-binding module (CBM). Additionally, they were able to elucidate the glycosylation site-, glycan size- and structure-specific effects of O-glycosylation on the folding using a mass spectrometry-based approach (P. K. Chaffey, X. Guan, X. Wang, et al., 2017).

Via comparison of the folding rates of a large collection of synthetic glycoforms, it was found that many factors, including glycosylation site, glycan structure and/or size, and

amino acid sequence, contributed to the effects of glycosylation. Small changes in either glycans or amino acids could lead to large differences in the ability of glycosylation to affect CBM folding. It was also revealed that glycosylation had different effects on the rates of each step in the folding process. Additionally, only the attachment of specific glycans, such as mono-mannose, to glycosylation sites nearest the N-terminus of CBM could significantly increase the overall folding rate. Together, the results pointed to the possibility that the interactions between glycan and local amino acids might be responsible for the increased folding rate.

3.2 Impact on Stability

The term “protein stability” includes many meanings, including proteolytic stability, thermal stability, physical stability (aggregation), et al. Thermal stability is the stability of a molecule at elevated temperatures. Proteolytic stability refers to the resistance toward the action of proteolytic enzymes. Aggregation represents the most vexing physical instability of proteins, which is a process by which misfolded proteins form insoluble precipitates.

Through the characterization and comparison of the properties of a series of synthetic glycoforms, the Tan group was able to reveal how glycans, and their site occupancy, can affect the thermal and proteolytic stability, as well as resistance to aggregation. Data from these studies showed that the effects of glycosylation are strongly dependent on the glycosylation sites and the structures of glycans directly attached to proteins (Chen et al., 2014; X. Guan, Chaffey, Chen, et al., 2018; Xiaoyang Guan et al., 2015; Prates et al., 2018). Increasing the sizes of glycans could further increase or maintain the stability conferred by original glycan structures. Again, the observations from these studies also suggested that the interactions between the α-linked glycans and the local amino acids might be the molecular basis underlying the increased stability.

3.3 Impact on Structure

Glycosylation tends to significantly increase the solubility of proteins. This makes crystallization of glycoproteins difficult. Solution NMR was thus chosen by the Tan group to investigate the effect of glycosylation on protein structure and the molecular basis underlying the increase in protein folding rate and stability. By systematically comparing the gathered high-resolution NMR structures, the results confirmed that glycosylation did not significantly alter the structures of proteins. Its main role is to introduce direct interactions between glycan residues and side chains of local amino acids (P. K. Chaffey, X. Guan, C. Chen, et al., 2017). If these interactions can be propagated throughout the rest of the backbone, they will be able to strengthen the global network of intramolecular interactions and lead to a decrease in the backbone dynamics of proteins. Lower backbone flexibility is most likely to be the reason for the increased folding rate and stability.

3.4 Impact on Biological Activity

In addition to examining the detailed effects of glycans on the properties related to folding, stability and structure, the Tan group also used the CBM as a model molecule to investigate the effect of glycosylation on the binding affinity (Chen et al., 2014; Xiaoyang Guan et al., 2015; Payne et al., 2013). By systematically comparing the properties of glycosylated

variants, the results showed that the effect of glycosylation on the binding affinity was not dependent on glycosylation sites nor glycan structures. Small glycan residues at each of the glycosylation sites may increase the binding affinity. The molecular mechanism by which glycans affect the binding of CBM to its substrates was investigated by NMR studies and molecular simulations (Happs et al., 2015). The results revealed that the interactions between glycans and substrates could be responsible for increased binding.

4. SUMMARY

In summary, through many years of efforts and development, the Danishefsky group was able to pave the way for the synthesis of multiply glycosylated proteins. The broad applicability of the developed methodologies was demonstrated in the studies of the impacts of glycosylation on protein properties. Through systematic chemical glycobiology studies, it was revealed that the effects of glycosylation were determined by a combination of factors, including glycosylation site, glycan size, structure, and amino acid sequence. The combined observations led to a hypothesis that is important for simplifying protein glycoengineering: “glycoforms with better overall properties can be generated by collaboratively varying glycan structures and adjacent amino acids within unstructured regions crucial to biological function and/or stability”. This hypothesis was confirmed by studying two molecules, human insulin and RANTES, otherwise known as CCL5 (X. Guan, Chaffey, Chen, et al., 2018; X. Guan, Chaffey, Wei, et al., 2018).

ACKNOWLEDGMENT

The authors are grateful to the National Science Foundation CAREER Award (Award number: CHE-1454925), the National Institutes of Health (Award number: R01EB025892) and the National Natural Science Foundation of China (Grant Number: 91853120) for their support.

REFERENCES

Anthony RM, Wermeling F, & Ravetch JV (2012). Novel roles for the IgG Fc glycan. Annals of the New York Academy of Sciences, 1253, 170–180. [PubMed: 22288459]

Chaffey PK, Chi L, & Tan Z (2017). Chemical biology of protein O-glycosylation Chemical Biology of Glycoproteins (pp. 48–93): The Royal Society of Chemistry.

Chaffey PK, Guan X, Chen C, Ruan Y, Wang X, Tran AH, et al. (2017). Structural insight into the stabilizing effect of O-glycosylation. Biochemistry, 56(23), 2897–2906. [PubMed: 28494147]

Chaffey PK, Guan X, Li Y, & Tan Z (2018). Using chemical synthesis to study and apply protein glycosylation. Biochemistry, 57(4), 413–428. [PubMed: 29309128]

Chaffey PK, Guan X, Wang X, Ruan Y, Li Y, Miller SG, et al. (2017). Quantitative effects of O-linked glycans on protein folding. Biochemistry, 56(34), 4539–4548. [PubMed: 28745859]

Chen L, Drake MR, Resch MG, Greene ER, Himmel ME, Chaffey PK, et al. (2014). Specificity of O-glycosylation in enhancing the stability and cellulose binding affinity of Family 1 carbohydrate-binding modules. Proc Natl Acad Sci U S A, 111(21), 7612–7617. [PubMed: 24821760]

Dwek RA (1995). Glycobiology: more functions for oligosaccharides. Science,269(5228), 1234–1235. [PubMed: 7652569]

Fernandez-Tejada A, Brailsford J, Zhang Q, Shieh JH, Moore MA, & Danishefsky SJ (2015). Total synthesis of glycosylated proteins. Topics in Current Chemistry, 362, 1–26. [PubMed: 25805144]

Greene ER, Himmel ME, Beckham GT, & Tan Z (2015). Glycosylation of cellulases: engineering better enzymes for biofuels. Advances in Carbohydrate Chemistry and Biochemistry, 72, 63–112. [PubMed: 26613815]

Guan X, Chaffey PK, Chen H, Feng W, Wei X, Yang LM, et al. (2018). O-GalNAcylation of RANTES improves its properties as a human immunodeficiency virus type 1 entry inhibitor. Biochemistry, 57(1), 136–148. [PubMed: 29202246]

Guan X, Chaffey PK, Wei X, Gulbranson DR, Ruan Y, Wang X, et al. (2018). Chemically precise glycoengineering improves human insulin. ACS Chemical Biology, 13(1), 73–81. [PubMed: 29090903]

Guan X, Chaffey PK, Zeng C, Greene ER, Chen L, Drake MR, et al. (2015). Molecular-scale features that govern the effects of O-glycosylation on a carbohydrate binding module. Chemical Science, 6(12), 7185–7189. [PubMed: 28966766]

Happs RM, Guan X, Resch MG, Davis MF, Beckham GT, Tan Z, & Crowley MF (2015). O-glycosylation effects on family 1 carbohydrate-binding module solution structures. FEBS Journal, 282(22), 4341–4356. [PubMed: 26307003]

Li Y, Clark KA, & Tan Z (2018). Methods for engineering therapeutic peptides, Chinese Chemical Letters, 29(7), 1074–1078.

Meany DL, & Chan DW (2011). Aberrant glycosylation associated with enzymes as cancer biomarkers. Clinical Proteomics, 8(1), 7. [PubMed: 21906357]

Moremen KW, Tiemeyer M, & Nairn AV (2012). Vertebrate protein glycosylation: diversity, synthesis and function. Nature Reviews Molecular Cell Biology, 13(7), 448–462. [PubMed: 22722607]

Nagorny P, Fasching B, Li X, Chen G, Aussedat B, & Danishefsky SJ (2009). Toward fully synthetic homogeneous beta-human follicle-stimulating hormone (beta-hFSH) with a biantennary N-linked dodecasaccharide. synthesis of beta-hFSH with chitobiose units at the natural linkage sites. Journal of the American Chemical Society, 131(16), 5792–5799. [PubMed: 19341309]

Ohtsubo K, & Marth JD (2006). Glycosylation in cellular mechanisms of health and disease. Cell, 126(5), 855–867. [PubMed: 16959566]

Payne CM, Resch MG, Chen L, Crowley MF, Himmel ME, Taylor LE 2nd, et al. (2013). Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. Proceedings of the National Academy of Sciences of the United States, 110(36), 14646–14651.

Prates ET, Guan X, Li Y, Wang X, Chaffey PK, Skaf MS, et al. (2018). The impact of O-glycan chemistry on the stability of intrinsically disordered proteins. Chemical Science, 9(15), 3710–3715. [PubMed: 29780502]

Rich JR, & Withers SG (2009). Emerging methods for the production of homogeneous human glycoproteins. Nature Chemical Biology, 5(4), 206–215. [PubMed: 19295526]

Rudd PM, & Dwek RA (1997). Glycosylation: heterogeneity and the 3D structure of proteins. Critical Reviews in Biochemistry and Molecular Biology, 32(1), 1–100. [PubMed: 9063619]

Sames D, Chen XT, & Danishefsky SJ (1997). Convergent total synthesis of a tumour-associated mucin motif. Nature, 389(6651), 587–591. [PubMed: 9335496]

Shang S, Tan Z, & Danishefsky SJ (2011). Application of the logic of cysteine-free native chemical ligation to the synthesis of Human Parathyroid Hormone (hPTH). Proceedings of the National Academy of Sciences of the United States, 108(15), 5986–5989.

Shang S, Tan Z, Dong S, & Danishefsky SJ (2011). An advance in proline ligation. Journal of the American Chemical Society, 133(28), 10784–10786. [PubMed: 21696220]

Spiro RG (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology, 12(4), 43R–56R.

Tan Z, Shang S, & Danishefsky SJ (2010). Insights into the finer issues of native chemical ligation: an approach to cascade ligations. Angewandte Chemie International Edition, 49(49), 9500–9503. [PubMed: 21053233]

Tan Z, Shang S, & Danishefsky SJ (2011). Rational development of a strategy for modifying the aggregatibility of proteins. Proceedings of the National Academy of Sciences of the United States, 108(11), 4297–4302.

Tan Z, Shang S, Halkina T, Yuan Y, & Danishefsky SJ (2009). Toward homogeneous erythropoietin: non-NCL-based chemical synthesis of the Gln78-Arg166 glycopeptide domain. Journal of the American Chemical Society, 131(15), 5424–5431. [PubMed: 19334683]

Varki A (1993). Biological roles of oligosaccharides: all of the theories are correct. Glycobiology, 3(2), 97–130. [PubMed: 8490246]

Varki A (2017). Biological roles of glycans. Glycobiology, 27(1), 3–49. [PubMed: 27558841]

Walczak MA, & Danishefsky SJ (2012). Solving the convergence problem in the synthesis of triantennary N-glycan relevant to prostate-specific membrane antigen (PSMA). Journal of the American Chemical Society, 134(39), 16430–16433. [PubMed: 22954207]

Wan Q, & Danishefsky SJ (2007). Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angewandte Chemie International Edition, 46(48), 9248–9252. [PubMed: 18046687]

Wang P, Dong S, Brailsford JA, Iyer K, Townsend SD, Zhang Q, et al. (2012). At last: erythropoietin as a single glycoform. Angewandte Chemie International Edition, 51(46), 11576–11584. [PubMed: 23012228]

Wang P, Dong S, Shieh JH, Peguero E, Hendrickson R, Moore MAS, & Danishefsky SJ (2013). Erythropoietin derived by chemical synthesis. Science, 342(6164), 1357–1360. [PubMed: 24337294]

Wang ZG, Zhang X, Live D, & Danishefsky SJ (2000). From glycals to glycopeptides: A convergent and stereoselective total synthesis of a high mannose N-linked glycopeptide. Angewandte Chemie International Edition, 39(20), 3652–3656. [PubMed: 11091428]

Wilson RM, Dong SW, Wang P, & Danishefsky SJ (2013). The winding pathway to erythropoietin along the chemistry-biology frontier: A success at last. Angewandte Chemie International Edition, 52(30), 7646–7665. [PubMed: 23775885]

Wu B, Hua Z, Warren JD, Ranganathan K, Wan Q, Chen G, et al. (2006). Synthesis of the fucosylated biantennary N-glycan of erythropoietin. Tetrahedron Letters, 47(31), 5577–5579.

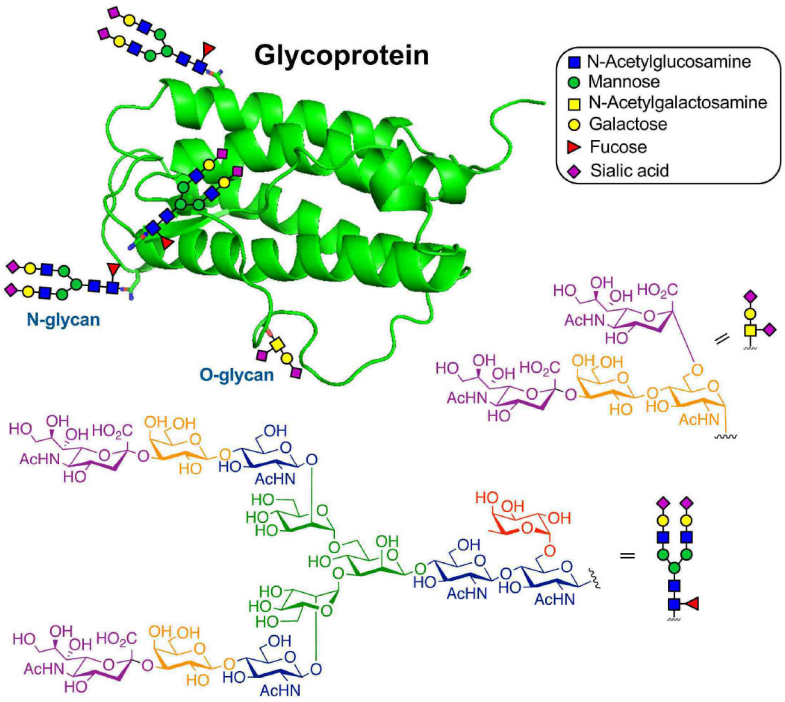

Fig. 1.

The structure of a typical glycoprotein, human EPO. The ribbon diagram of the tertiary structure of human EPO protein is shown in green. It contains three N-linked glycosylation sites, N24, N38, and N83, and one O-linked glycosylation site, S126.

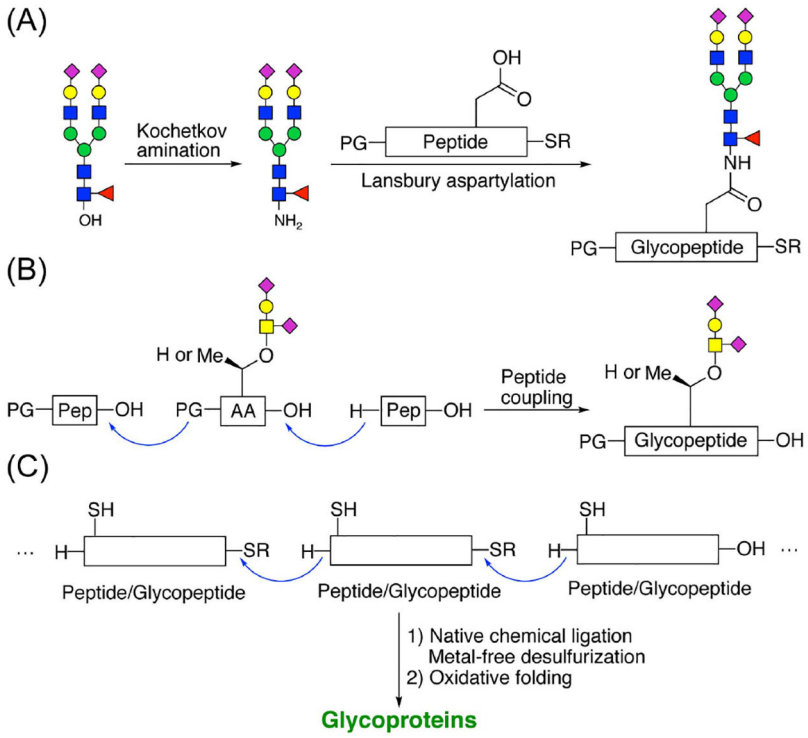

Fig. 2.

Chemical synthesis of glycoproteins. (A) Convergent synthesis of N-linked glycopeptides; (B) Stepwise synthesis of O-linked glycopeptides; (C) Native chemical ligation/metal-free desulfurization-based strategy for the synthesis of glycopeptides. PG, protecting groups; SR, thioesters; Pep, peptides; AA, amino acids.

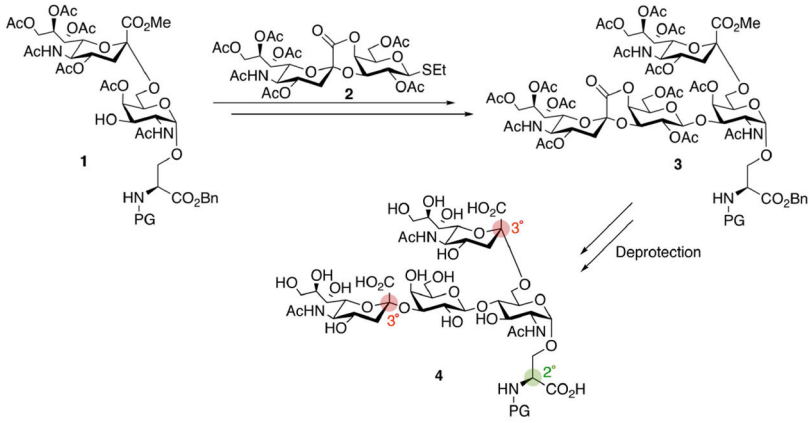

Fig. 3.

Chemical synthesis of O-linked glycoamino acid with unmasked carbohydrate moiety. The reactivity of different carboxyl groups in 4 can be effectively controlled by the substitution level (steric hindrance) of their neighboring carbons. The one that is directly attached to a secondary carbon (2°, green colored) has better reactivity to the N-terminal amine groups of peptides. PG, protecting groups.

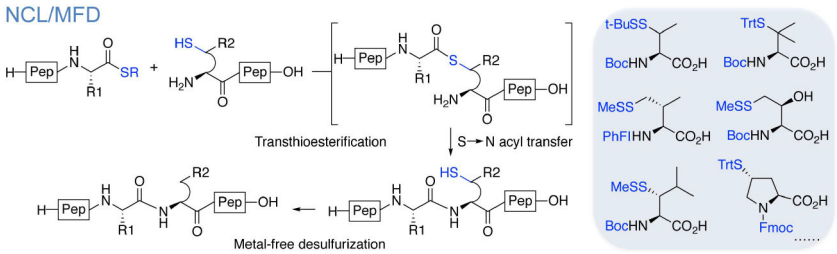

Fig. 4.

Ligation at noncysteine amino acid residues by the native chemical ligation/metal-free desulfurization strategy. SR, thioester; Pep, peptides; R1 and R2, functional groups in amino acid side chains.

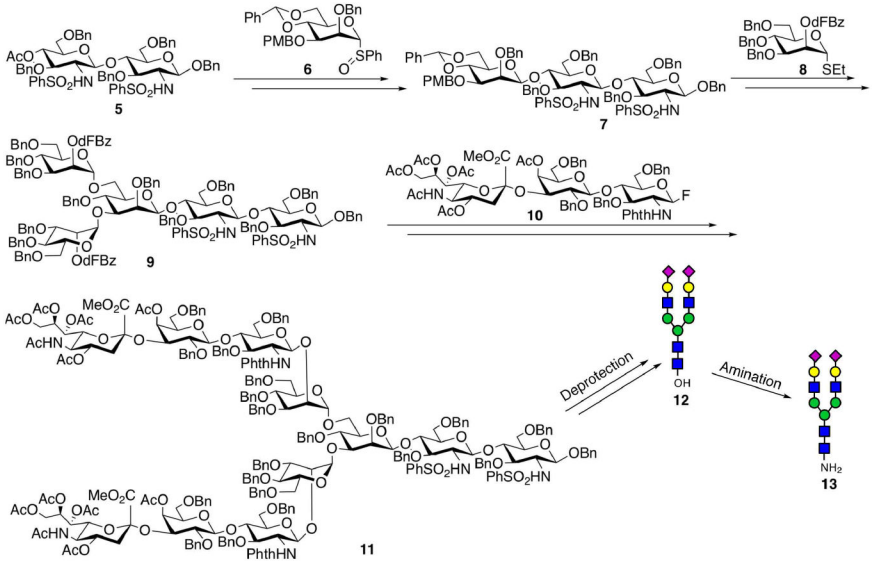

Fig. 5.
Chemical synthesis of glycosyl amines.