Annu. Rev. Biochem. 2002. 71:593–634
DOI: 10.1146/annurev.biochem.71.110601.135334

HOMOGENEOUS GLYCOPEPTIDES AND GLYCOPROTEINS FOR BIOLOGICAL INVESTIGATION

Michael J. Grogan,[1] Matthew R. Pratt,[1]
Lisa A. Marcaurelle,[1] and Carolyn R. Bertozzi[1–4]

Departments of Chemistry[1] and Molecular and Cell Biology,[2] University of California; Center for Advanced Materials,[3] Materials Sciences Division, Lawrence Berkeley National Laboratory; Howard Hughes Medical Institute,[4] Berkeley, California, 94720; e-mail: groganmike2001@yahoo.com, mpratt@uclink.berkeley.edu, lamarcau@mit.edu, bertozzi@cchem.berkeley.edu

Key Words N-linked, O-linked, GPI anchor, oligosaccharide, chemoselective ligation

■ **Abstract** Protein glycosylation is widely recognized as a modulator of protein structure, localization, and cell-cell recognition in multicellular systems. Glycoproteins are typically expressed as mixtures of glycoforms, their oligosaccharides being generated by a template-independent biosynthetic process. Investigation of their function has been greatly assisted by sources of homogeneous material. This review summarizes current efforts to obtain homogeneous glycopeptide and glycoprotein materials by a variety of methods that draw from the techniques of recombinant expression, chemical synthesis, enzymatic transformation, and chemoselective ligation. Some of these techniques remove obstacles to glycoprotein synthesis by installing nonnative linkages and other modifications for facilitated assembly. The end purpose of the described approaches is the production of glycosylated materials for experiments relevant to the biological investigation of glycoproteins, although the strategies presented apply to other posttranslational modifications as well.

CONTENTS

INTRODUCTION 594
MUCIN-TYPE O-LINKED GLYCOSYLATION 595
Chemical Synthesis of O-Linked Glycopeptides 597
Chemoenzymatic Synthesis of O-Linked Glycopeptides 600
Chemoselective Ligations for the Synthesis of O-Linked Glycopeptides 605
Synthesis of O-Linked Glycoproteins by Native Chemical Ligation 609
N-LINKED GLYCOSYLATION 611
Recombinant Methods for the Production of N-Linked Glycoproteins 615
Chemoenzymatic Synthesis of N-Linked Glycopeptides 616

Chemical Synthesis of N-Linked Glycopeptides and Proteins 618
Chemoselective Ligations for the Synthesis of N-Linked Glycopeptides and Proteins 621
Synthesis of N-Linked Glycoproteins by Native Chemical Ligation 625
GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED PROTEINS 625
SUMMARY 630

INTRODUCTION

Oligosaccharides linked to proteins or lipids are characteristic features of all cell surfaces. Combinations of natural hexoses joined by their possible linkages and branch points produce a bewildering array of sugar oligomers, and their known structural diversity reflects this complexity. It can be speculated that polysaccharides have evolved from structural components of the cell to mediators of complex, multicellular processes with the emergence of higher organisms. Conserved features of oligosaccharides serve basic housekeeping functions within cells, such as protein folding and trafficking. But their widely diverse tissue-specfic elaborations have been adapted to serve in such processes as development and immune recognition (1, 2).

Glycosylation provides an opportunity to vastly augment the information of a very concise genome. For example, a single gene product like tissue plasminogen activator (tPA) can be expanded into over 100 discrete glycoforms with a distribution of bioactivities (3). A meaningful analysis of eukaryotic genomes for global differences in glycoprotein expression is by nature difficult to conduct, for posttranslational modifications are not under direct genetic control. Furthermore, genomic comparison of complexity among yeast, flies, worms, and humans is probably misleading (they are estimated to have 6,000, 13,000, 18,000, and 32,000 genes each, respectively), as humans have a perplexingly low gene number that produces a relatively high organismal complexity (4–6). Glycosylation represents a level of variability that may be necessary for complex processes in higher organisms. As an example, the *N*-acetylglucosaminyltransferase Fringe of *Drosophila melanogaster* was recently found to glycosylate the key developmental switch Notch and thereby modulate developmental patterning of wing cells (7–10). The details of this process are still being defined and emphasize the importance of glycoproteins to multicellular phenomena.

Progress toward understanding the functions of oligosaccharides has been hampered by their complexity and heterogeneity (11). Since they are products of a template-independent biosynthesis, genetic methods for the alteration and homogeneous expression of defined carbohydrate structures do not exist. At best, some oligosaccharide sites can be deleted or installed in order to probe their effects on function. Biochemical modifications to the termini of glycans can produce some defined structural changes, but, as with the native structures, the modified glycopeptides generated remain heterogeneous and difficult to purify and characterize.

The assembly of glycoproteins from readily available components represents an attractive route to homogeneous materials for structural and functional studies. Unlike peptide and nucleotide synthesis, carbohydrate chemistry is complicated by the structural branching and varied stereochemistry of large glycans. Furthermore, the formation of glycosidic bonds requires strictly anhydrous conditions that are incompatible with unprotected peptides and proteins. As a testament to synthetic creativity, glycopeptides and glycoproteins that vary in complexity of the sugars from monosaccharides to complex glycans have been prepared with an amalgam of methods drawn from chemical synthesis, chemoselective ligation reactions, and chemoenzymatic transformations. This review summarizes the tools and approaches that are now available to the biochemist for preparation of homogeneous glycopeptides and glycoproteins. Some of the methods discussed have been reviewed separately (12–16); this review focuses on emerging strategies that will provide glycosylated structures previously unattainable and directly relevant to structural and functional studies of glycoproteins.

MUCIN-TYPE O-LINKED GLYCOSYLATION

In mammals and other eukaryotes, the most prevalent form of O-linked protein glycosylation is mucin-type glycosylation, where *N*-acetylgalactosamine (GalNAc) is α-O-linked to the β-hydroxyl group of either a serine or a threonine residue of the polypeptide (Figure 1) (17–21). Mucins are glycoproteins with dense clusters of such O-linked sites bearing up to 20 monosaccharides per glycan, yet isolated mucin-type glycosylation also occurs at singular positions on glycoproteins. Glycosylaminoglycans such as heparin and chondroitin sulfate are also O-linked, but through a core β-*O*-xylose residue linked to serine; these structures are found on proteoglycans principal components of the extracellular matrix (Figure 1). Other forms of O-linked glycosylation include β-*O*-*N*-acetylglucosamine (GlcNAc) commonly found on nuclear proteins (22) and α-*O*-fucose that has been discovered on both serine and threonine residues of epidermal growth factor (EGF)-like domains (23).

The α-*O*-GalNAc-Ser/Thr structure, commonly referred to as the Tn antigen (Figure 1), forms the biosynthetic foundation for a diverse array of core structures generated by glycosylation at the C-3 and/or C-6 hydroxyl groups of GalNAc (Figure 2). Of the eight identified core structures, cores 1 and 2 are the most abundant and widely distributed in both mucin and nonmucin glycoproteins (24). Other structures, such as cores 3 and 4, occur less frequently and are typically confined to mucins. The core structures of O-linked glycans may be elongated by the addition of sialic acid, fucose, and/or repeating units of Galβ1,4GlcNAc. Together with further modifications, such as sulfation, these elaborations give rise to highly complex structures, often containing important recognition elements such as the Lewis and blood group antigens (25).

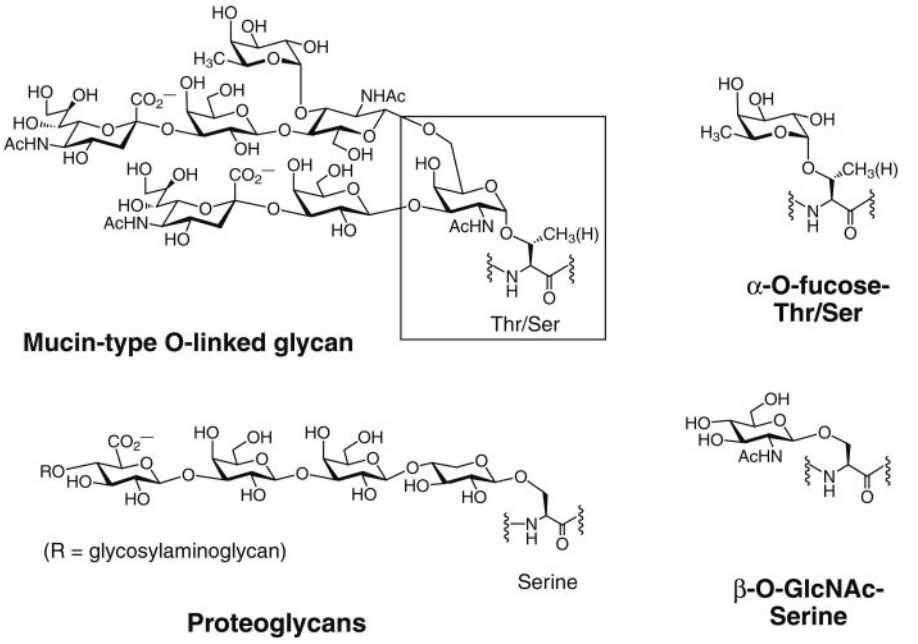


Figure 1 Four major classes of O-linked glycopeptides. The Tn antigen of mucin-type O-linked glycans is outlined in a box.

Mucin-type glycosylation is found on many cell-surface proteins and plays a vital role in interactions of the cell with its environment, particularly for immune responses (26). Because the signal sequence for O-linked glycosylation of proteins is poorly understood, the presence and function of these glycans is generally difficult to determine for a particular protein. The dense glycosylation of mucin domains has been suggested to effect an extended polypeptide conformation, consistent with structural details from electron micrographs, atomic force microscopy, and light scattering of cell-surface mucin glycoproteins (27–31). To address the structural effects of mucin-type glycosylation, several studies have been performed with small synthetic glycopeptide fragments. Notably, the Danishefsky group has reported nuclear magnetic resonance (NMR) solution structures of a mucin fragment from CD43 (32). Comparison of unglycosylated peptide to four different glycoforms with clusters of mono-, di-, or trisaccharides illustrated that the peptide backbone is extended and ordered upon glycosylation with α-GalNAc, which is consistent with other conformational studies of O-linked glycopeptides and glycoproteins (33–36). Replacement of the α-GalNAc residues with a nonnative β-O-linked GalNAc residue yielded a disordered structure similar to that of the unglycosylated peptide. In addition, the indifference of the peptide conformation to distal sugars added to the core α-GalNAc residue further suggests that it is the clusters of α-GalNAc on the

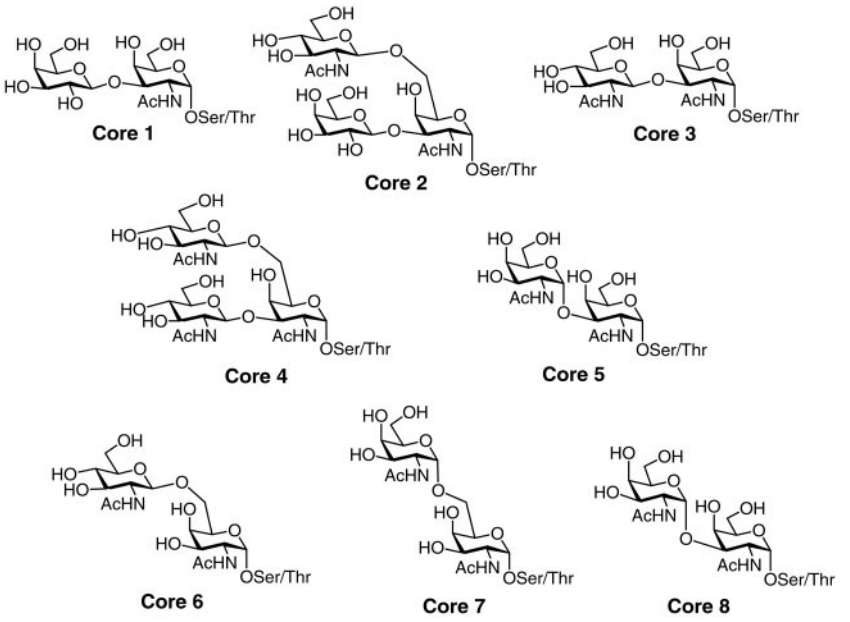


Figure 2 Eight known O-linked core structures of the mucin type.

mucin scaffold, and not the precise glycan structures, that provide a persistently extended backbone for the multivalent display of various branched oligosaccharides.

Chemical Synthesis of O-Linked Glycopeptides

Synthesis of mucin-type glycopeptides is commonly accomplished by incorporation of a suitably protected *O*-glycosyl amino acid into a polypeptide by solid-phase peptide synthesis (SPPS). Of the two standard methods, 9-fluorenylmethoxycarbonyl (Fmoc)-based chemistry is more often employed than *tert*-butyloxycarbonyl (Boc)-based chemistry for the SPPS of glycopeptides. The sequential removal of base-labile Fmoc protecting groups for peptide elongation is compatible with the presence of acid-sensitive glycosidic bonds and avoids repeated exposure to trifluoroacetic acid and final deprotection with hydrogen fluoride, common to Boc-based methods. The hydroxyl groups of the carbohydrates are typically protected as acetyl or benzoyl esters which, upon cleavage of the completed glycopeptide from the solid support, are readily removed by treatment with sodium methoxide or hydrazine. Obtaining the appropriate *O*-glycosyl amino acid is the major challenge to glycopeptide synthesis, and the size limit of the synthetic glycopeptides is imposed by the technical constraints of SPPS.

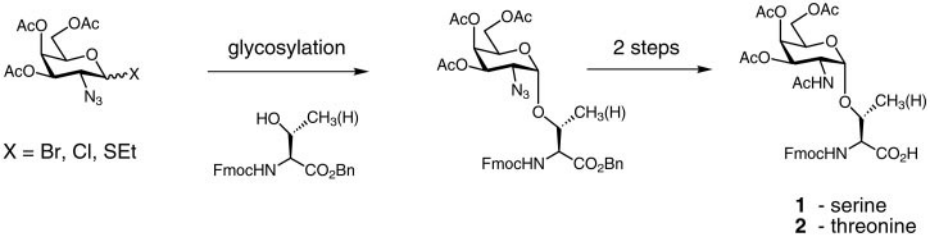


Figure 3 Synthesis of α-*O*-GalNAc-serine (**1**) and threonine (**2**) building blocks for SPPS of O-linked glycopeptides. (SEt = SCH_2CH_3)

The α-*O*-GalNAc-Ser/Thr building blocks required for SPPS of mucin-type glycopeptides are generally prepared by glycosylation on the appropriate serine or threonine alcohol acceptor. The use of a 2-azido halo- or thioglycoside donor ensures high α-selectivity in the glycosylation reaction (37) (Figure 3). Conversion of the 2-azido group to an *N*-acetamido group provides the Fmoc-protected GalNAc-Ser/Thr amino acids **1** and **2**, respectively. Since these building blocks are now commercially available, simple O-linked glycopeptides are readily accessible to those with experience in SPPS. For example, we have synthesized the 82-residue glycoprotein diptericin, an insect-derived antibacterial molecule bearing two O-linked glycans (38). Glycosyl amino acid **2** was synthesized on gram scale, allowing the production of full-length glycosylated protein and each of its two individual domains in quantities sufficient to probe diptericin's mechanism of action.

Although the assembly of peptides containing simple α-GalNAc residues is now relatively straightforward, the routine construction of glycopeptides bearing more elaborate O-linked glycans remains a tremendous challenge. The chief obstacle in the synthesis of complex *O*-glycosyl amino acids is obtaining high α-selectivity in the formation of the *O*-Ser/Thr mucin-type linkage. Even with simple monosaccharide donors, establishing α-selective conditions for the glycosylation reaction can be trying, and this variability is exaggerated when dealing with large oligosaccharide donors to prepare complex structures found on native mucins (39). Winterfeld & Schmidt have recently found that nitroglycals of mono- and disaccharides react selectively to form the α-*O*-Ser/Thr linkage to access Tn and sialyl Tn (STn) antigens as well as core 7 and core 1 structures. This method should become broadly applicable to larger oligosaccharides (40).

This selectivity problem for complex structures is typically bypassed by formation of the desired α-*O*-Ser/Thr linkage prior to addition of branching residues onto the core GalNAc residue of an α-glycosyl amino acid, referred to as an α-O-linked cassette (Figure 4). Rather than uniform masking with benzoyl or acetyl esters, having orthogonally removable protecting groups on the C-3, C-4, and C-6 hydroxyl groups of an α-*O*-GalNAc cassette permits the controlled, sequential addition of branching sugars. Applying cassette strategies, a variety of

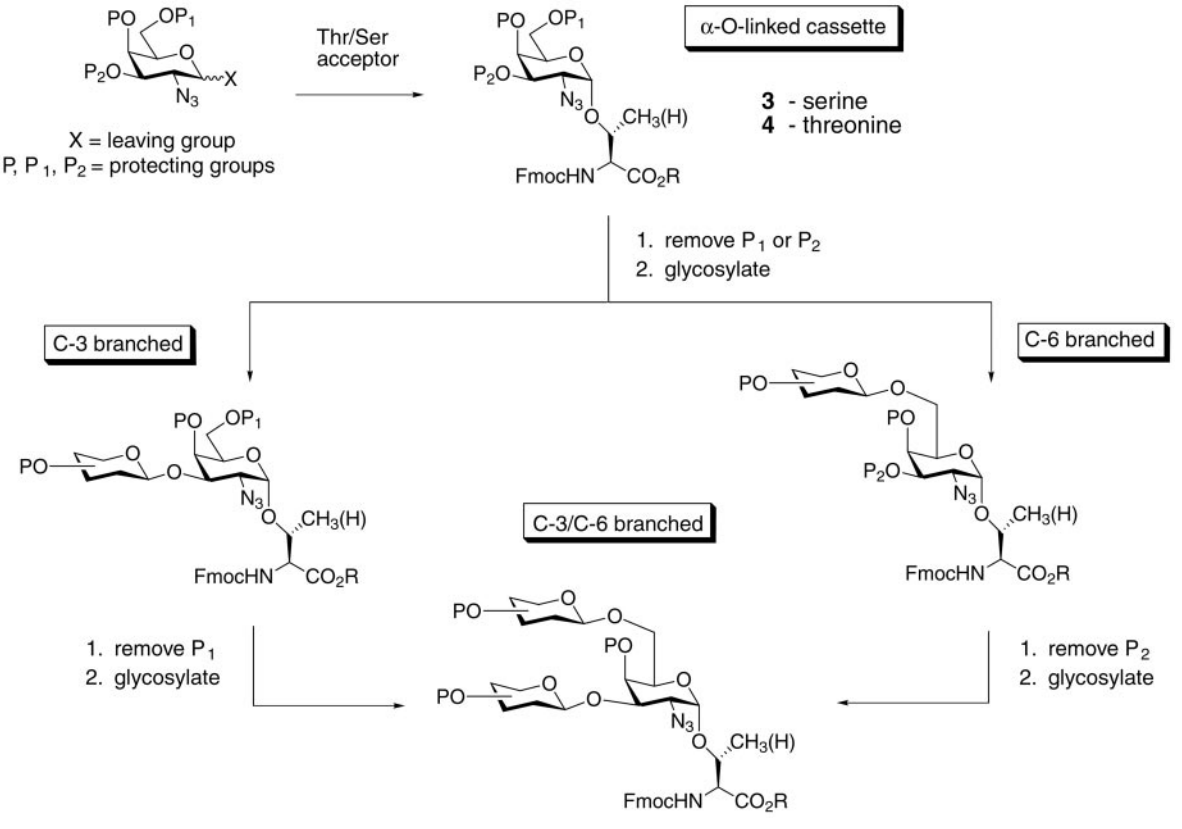


Figure 4 Synthesis of complex O-linked glycosyl amino acids for SPPS from serine and threonine cassettes **3** and **4**. Sequential deprotection and glycosylation of the cassette hydroxyl groups produces core O-linked glycans.

complex glycosyl amino acids have been incorporated into synthetic mucin-related glycopeptides. The seminal work of Meldal & coworkers (41, 42) first used α-O-linked cassettes for the synthesis of four O-linked structures, the common cores 1–4 (Figure 2). Only two sugar donors and one protected α-*O*-GalNAc-Thr cassette were necessary to prepare the four core glycans, establishing a highly efficient route to core O-linked glycosyl amino acids for SPPS. The core glycan building blocks that resulted were used for the construction of a series of decapeptides corresponding to domains of the mucins MUC-2 and MUC-3 (41).

Danishefsky & coworkers have employed related cassette-based approaches for preparing a diverse set of complex glycosyl amino acids bearing tumor-related antigens (39, 43, 44). Amino acids with protected Tn, TF, 2,6-STF, and Lewis y (Le^y) antigens (Figure 5) were generated from appropriate threonine and serine cassettes for the syntheses of clustered O-linked glycopeptides. Variants of fragment **5** bearing the Tn, TF, and 2,6-STF antigens were designed for generating antitumor vaccines (45) and were also used for NMR structure determination of the CD43 mucin domain mentioned earlier (32). Mucin carbohydrate patterns change to characteristic glycoforms during the malignant transformation of cells, exposing a potential target for cancer immunotherapy (46). Glycopeptides represented by **5** generate robust antibody responses that cross-react with the same antigen expressed on tumor cells (44). Although it is difficult to determine with certainty the causalities behind immunogenesis, it appears that a mucin-like oligomeric display of the carbohydrate antigens on a peptide backbone contributes significantly to correct antigen presentation. Unless presented in a native context, carbohydrate-based antigens frequently fail to produce immune responses. For example, only antibodies raised against the peptide **6** (Figure 5) bearing clustered Le^y antigen elicited antibody reactivity to both monomeric Le^y-ceramide and clustered Le^y-mucin glycoproteins (43), whereas antibodies to monovalent forms of Le^y did not recognize clustered Le^y. The clustered Tn/TF glycoforms of **5** as well as other glycoconjugates have been evaluated as vaccines for prostate cancer, and **6** has advanced further into ovarian cancer trials (45).

Cassette-based routes to glycopeptides have proven effective for the generation of biologically relevant mucin-type glycopeptides bearing a range of defined and impressively complex O-linked glycans. Recent progress in oligosaccharide synthesis, notably the use of automated solid-phase synthesis (47, 48) and programmable one-pot methods (49), should accelerate the production of elaborate glycosyl amino acids for synthesis of O-linked glycopeptides.

Chemoenzymatic Synthesis of O-Linked Glycopeptides

The incorporation of complex O-linked glycan cassettes into larger peptides and native-length proteins remains a significant synthetic challenge, from the standpoints of both carbohydrate complexity and the material requirements of linear peptide synthesis (49). The enzymatic transfer of individual monosaccharides

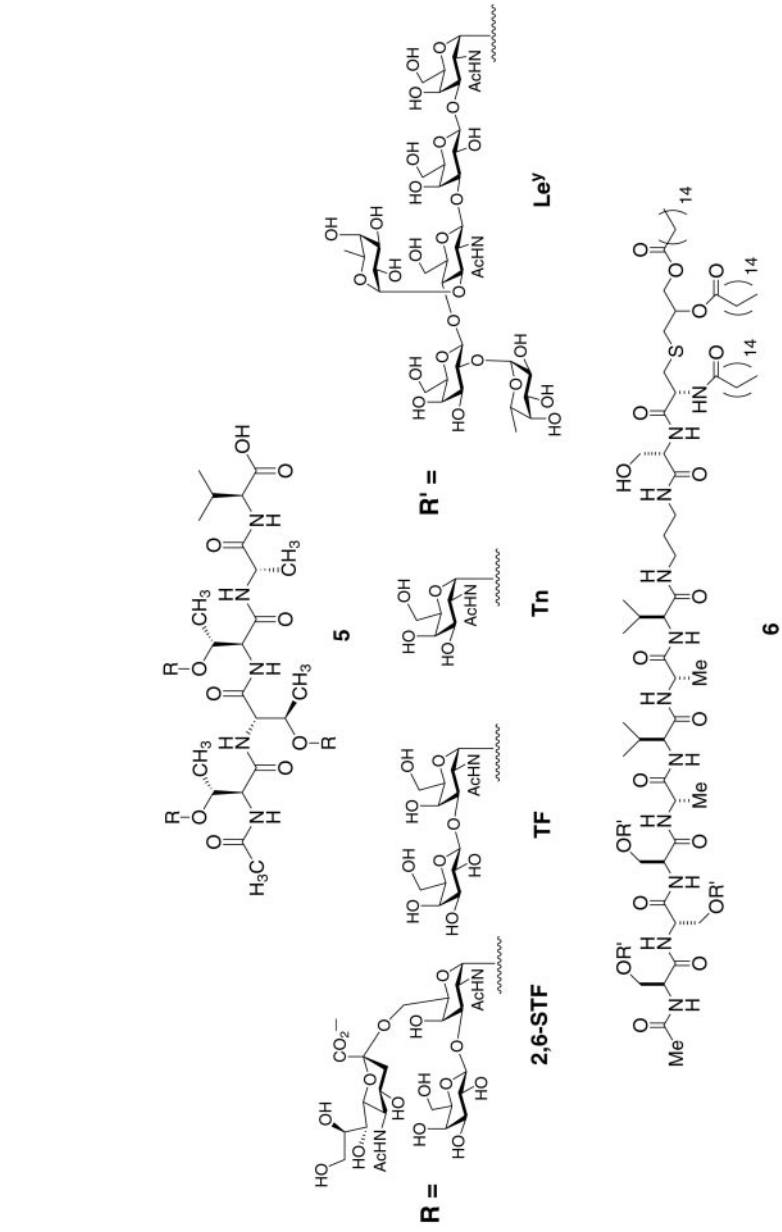


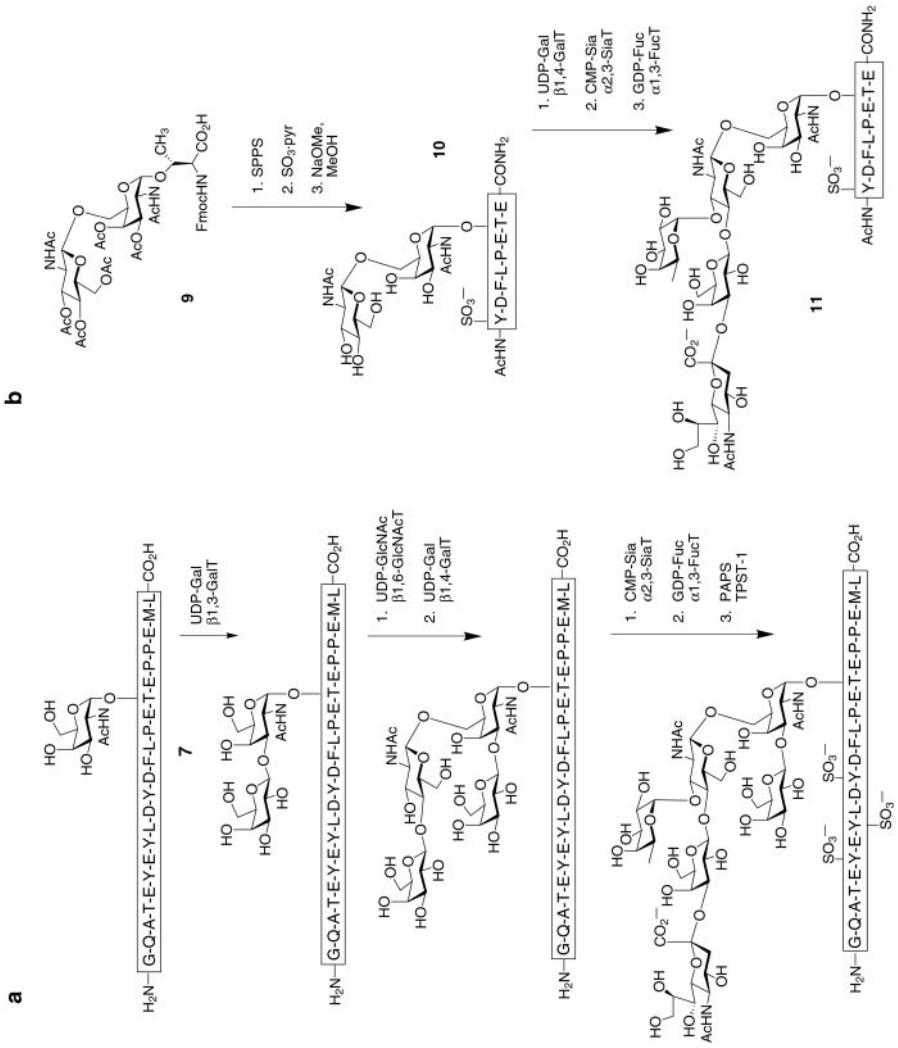
Figure 5 Glycopeptides prepared by cassette synthesis bearing the tumor-associated antigens Tn, TF, and 2,6-sialyl TF (STF) on **5**, a mucin domain of CD43, and the Lewis y (Le^y) antigen on **6**.

to synthetic glycopeptides containing simple O-linked glycans is a seductive alternative to cumbersome stereoselective glycosylations and protecting-group strategies for building complex structures. An increasing number of glycosyltransferases and glycosidases are becoming available from recombinant expression or natural sources for reactions with unprotected substrates to produce more complex glycopeptides and proteins.

Enzymatic methods have been strikingly highlighted by recent investigations into the binding of E- and P-selectin to P-selectin glycoprotein ligand-1 (PSGL-1), an event central to leukocyte recruitment to sites of inflammation. The selectins are a family of cell-surface glycoproteins that bind, by lectin domains, to counter-receptors such as GlyCAM-1, CD34, ESL-1, and PSGL-1 (50). The molecular basis of these interactions is under intense scrutiny, as they are validated targets for anti-inflammatory therapeutics: PSGL-1-deficient mice are severely lacking in early neutrophil recruitment and P-selectin-mediated leukocyte rolling (51), and soluble PSGL-1 fragments are being investigated as inhibitors of the inflammation process. The binding of P-selectin to PSGL-1, a mucin glycoprotein, requires the anionic N terminus of PSGL-1 bearing an O-linked sialyl Lewis x (sLe^x) glycan at Thr57 and one or more sulfated Tyr residues (Tyr46, Tyr48, or Tyr51) (52).

To elucidate the contributions of these modifications to binding, the chemoenzymatic synthesis of a collection of PSGL-1 fragments was established (Figure 6*a*) (53, 54). Glycopeptide **7** was prepared by SPPS, and the appropriate glycosyltransferases were used to elaborate the core monosaccharide, to give a hexasaccharide with the sLe^x motif. Enzymatic sulfation of the three tyrosine residues with a recently cloned tyrosylprotein sulfotransferase (TPST-1) (55) produced the fully modified peptide **8**, which binds P-selectin nearly as well as full-length PSGL-1. In a more recent report, the tyrosine sulfate residues were installed on the peptide prior to enzymatic glycosylations using the SPPS building block Fmoc-Tyr(SO_3)-OH, to generate the possible combinations of mono-, di-, and trisulfated peptides O-linked to sLe^x (54). These and related PSGL-1 glycopeptides were compared for binding to P-selectin, revealing the individual contributions of carbohydrate and sulfate modifications. A functional model for selectin binding to PSGL-1 has been constructed from these results and

Figure 6 Chemoenzymatic syntheses of a PSGL-1 N terminus modified with sulfate and sialyl Lewis x (sLe^x). The syntheses were performed by the Cummings (*a*) and Wong (*b*) groups. Synthetic glycopeptides (**7** or **10**) were sequentially treated with the appropriate sugar donors and glycosyltransferases. Abbreviations: galactosyl transferase (GalT), *N*-acetylglucosaminyltransferase (GlcNAcT), sialyltransferase (SiaT), fucosyltransferase (FucT), 3′-phosphoadenosine 5′-phosphosulfate (PAPS), tyrosylprotein sulfotransferase-1 (TPST-1).


a
b
UDP-Gal β1,3-GalT
UDP-GlcNAc β1,6-GlcNAcT
UDP-Gal β1,4-GalT
CMP-Sia α2,3-SiaT
GDP-Fuc α1,3-FucT
PAPS TPST-1
1. SPPS
2. SO3·pyr
3. NaOMe, MeOH
1. UDP-Gal β1,4-GalT
2. CMP-Sia α2,3-SiaT
3. GDP-Fuc α1,3-FucT
7
8
9
10
11
H2N-G-Q-A-T-E-Y-E-Y-L-D-Y-D-F-L-P-E-T-E-P-P-E-M-L-CO2H
AcHN-Y-D-F-L-P-E-T-E-CONH2

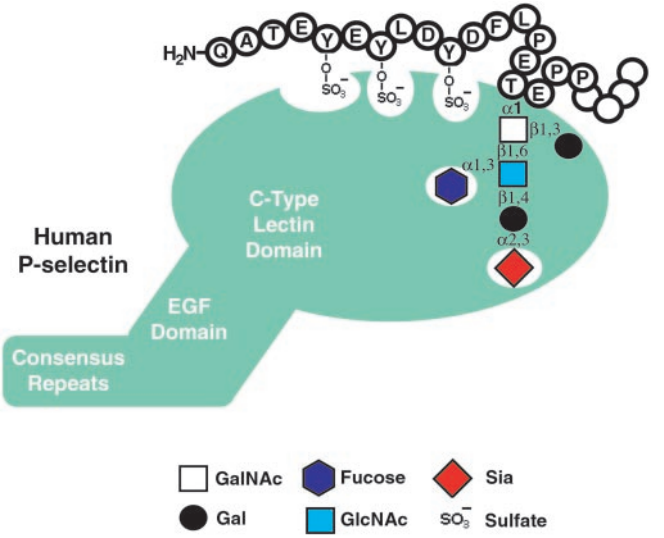


Figure 7 A soluble form of P-selectin bound to the N terminus of PSGL-1. P-selectin is an endothelial adhesion molecule comprising a C-type lectin domain, an epidermal growth factor (EGF) like domain, and a series of consensus repeats similar to complement regulatory proteins. P-selectin binding to the N terminus of PSGL-1 requires sulfotyrosine residues and an sLe^x glycan. The sialic acid and fucose residues of sLe^x interact strongly with the lectin domain.

from the recent X-ray structures of the P- and E-selectin lectin domains bound to an expressed fragment of PSGL-1 similar to **8** (56) (Figure 7). For these studies, the preparation of pure homogeneous sulfated glycopeptides was essential for uncovering the molecular basis for this binding interaction.

The chemoenzymatic synthesis of a related sulfoglycopeptide fragment of PSGL-1 (**11**, Figure 6*b*) has been accomplished by Wong and coworkers (57, 58). The enzymatic glycosylations started with a monosulfated glycopeptide (**10**) carrying an α-O-linked disaccharide rather than a simple monosaccharide. The presence of tyrosine sulfation strongly decelerated the activity of rat liver α2,3-sialyltransferase (SiaT), but the authors discovered that an expressed α2,3-SiaT from *Neisseria meningitidis* was able to produce the sialylated glycan. The route does require preparation of a complex disaccharide cassette (**9**). Yet this strategy advantageously does not require either the core 1 β1,3-galactosyltransferase (GalT) or the core 2 β1,6-*N*-acetylglucosaminyltransferase (GlcNAcT), neither of which is currently available commercially, to elaborate the sLe^x moiety from the core GalNAc residue. The use of amino acid **9** for generating C-6 branched *O*-glycans demonstrates the flexibility of a combined enzymatic and chemical route in overcoming obstacles to glycoprotein synthesis.

Chemoselective Ligations for the Synthesis of O-Linked Glycopeptides

In order to study glycoprotein function and the role of carbohydrates in a biologically relevant context, homogeneous complex glycoproteins are required. The chemical methods to prepare glycopeptides described above have been most successfully applied to short oligopeptide substrates and less commonly to protein-like structures. Conventional methods to form glycosidic bonds require strictly anhydrous conditions and polypeptides with protected side chains, and furthermore, the size limit of SPPS is well below the length of many glycoproteins of interest.

Chemoselective ligations for joining carbohydrates, peptides, and glycopeptides present powerful alternate strategies for assembling homogeneous glycoproteins (15, 16, 59). By introducing nonnative electrophilic and nucleophilic functional groups into the peptide or carbohydrate scaffold, chemical reactions orthogonal to the surrounding molecules can be designed to form stable bonds efficiently, in aqueous environments and at moderate temperatures. The high selectivity of these chemoselective reactions dispenses with the need for installing and removing protecting groups, whereby the convergently assembled glycopeptides are not subjected to additional material-depreciating steps. With regard to synthetic design, application of these ligations to the rapid construction of glycopeptides enjoys both the structural control and broad substrate scope of chemical synthesis without forsaking robust selectivity and mild reaction conditions offered by enzymatic synthesis.

Toward the fabrication of O-linked glycopeptides by chemoselective ligation, our group has developed several methods to introduce ketone, aldehyde, aminooxy, or thiol groups into peptides and glycopeptides and their carbohydrate partners for chemoselective ligations (60–64). To prepare an analog of the bacteriostatic glycopeptide drosocin by this strategy, we designed an oxime linkage between aminooxy-sugars and ketone-bearing peptides as a surrogate for the mucin-type O-linkage to Ser/Thr (Figure 8*a*). (2*S*)-Aminolevulinic acid, an unnatural amino acid possessing a ketone side chain, was incorporated into the peptide as a replacement for Thr11 by means of Fmoc-based SPPS. Aminooxy-GalNAc was coupled to the unprotected nonadecapeptide to cleanly provide the drosocin analog with an oxime-glycoside linkage proximal to the peptide backbone. This oxime-linked glycopeptide was fourfold more potent in blocking bacterial growth ($IC_{50} = 0.16 \pm 0.04\ \mu M$) than unglycosylated drosocin ($IC_{50} = 0.63 \pm 0.05\ \mu M$) and similar in potency to native drosocin ($IC_{50} = 0.10 \pm 0.02\ \mu M$), which indicates that the native sugar-peptide linkage in drosocin is not essential to achieve glycosylation-dependent enhancement in potency. The extent to which this observation extends to other O-linked glycopeptides is a matter of ongoing investigation in our laboratory. The ligation of ketone-containing peptides and aminooxy-sugars is not limited to single sites or simple monosaccharides, as we have conjugated aminooxy-Tn and aminooxy-STn to multiple ketone sites on mucin fragments of GlyCAM-1 (Figure 8*b*) (65, 66). Peptides **12** and **13**,

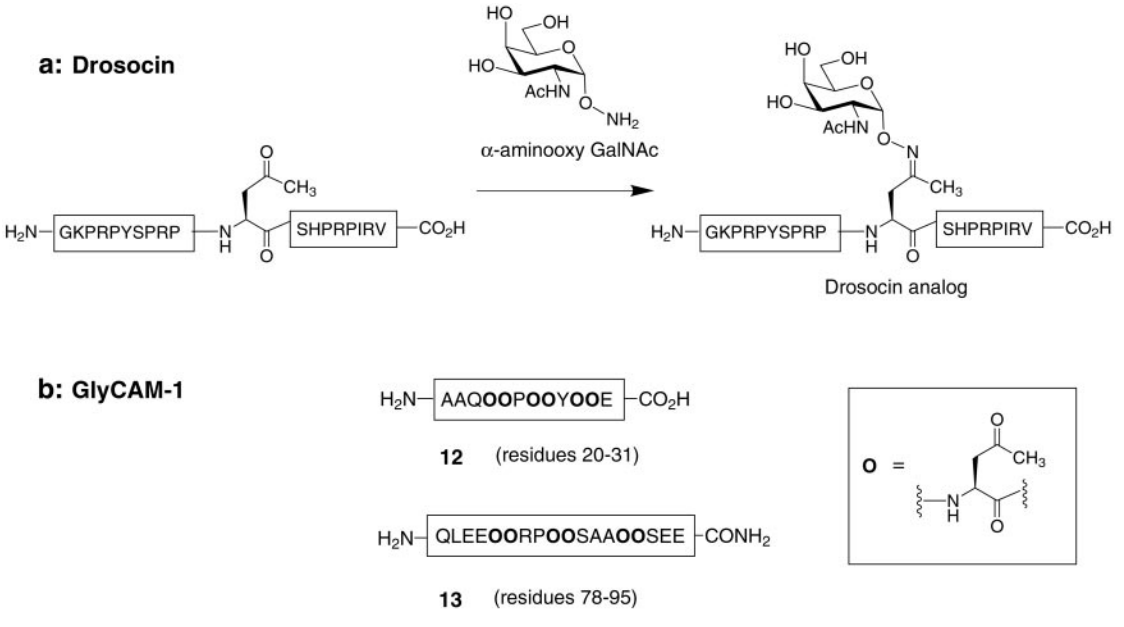


Figure 8 Synthesis of two glycopeptides, (*a*) drosocin and (*b*) GlyCAM-1 fragments **12** and **13**, by chemoselective ligation of aminooxy-sugars to ketone side chains of (2*S*)-aminolevulinic acid (designated **O**).

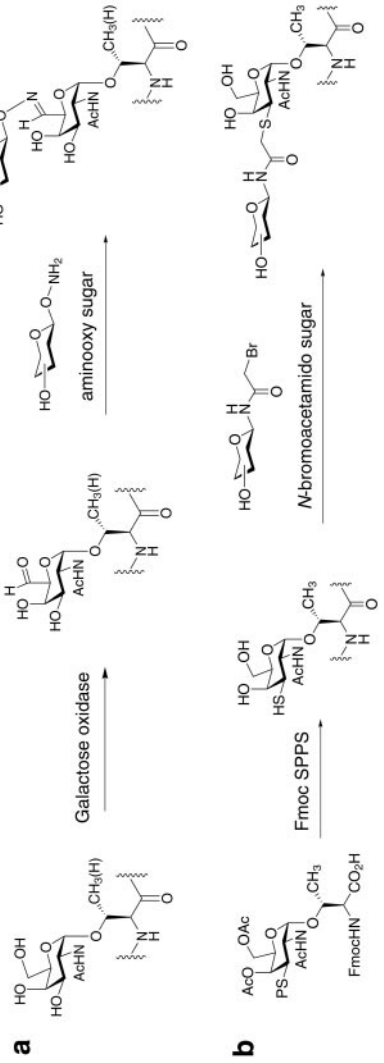


Figure 9 Chemoselective ligations to a core *O*-GalNAc residue by (*a*) ligation of aminooxy-sugars to C-6 aldehydes generated by galactose oxidase and (*b*) alkylation of 3-thioGalNAc with bromoacetamido sugars.

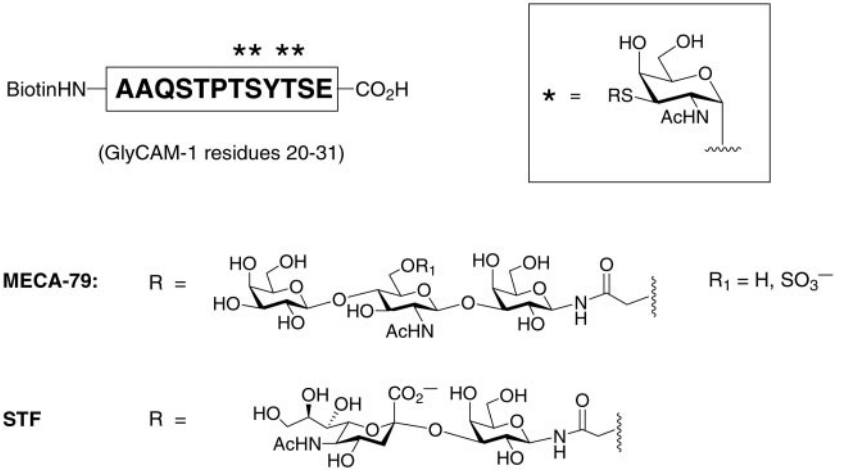


Figure 10 GlyCAM-1 mucin domains bearing MECA-79 and sialyl TF (STF) antigens installed by bromoacetamide ligation to 3-thioGalNAc. MECA-79 is expressed on endothelial mucins at sites of inflammation, and STF antigen is a marker for types of breast cancer cells.

generated by SPPS, have six Ser or Thr residues within the 12- or 17-amino-acid sequences replaced with (2*S*)-aminolevulinic acid. Reaction of either peptide with α-aminooxy-GalNAc cleanly gave the respective oxime-linked glycosylated peptides. Aminooxy-sLe^x has likewise been displayed on synthetic peptide backbones, using a similar synthetic strategy (67).

To retain the natural glycopeptide linkage shown to be important for structural integrity, additional chemoselective ligation methods have recently been employed for the synthesis of O-linked glycopeptide mimetics possessing unnatural bonds at the branch points (C-3 and C-6) of the core O-linked GalNAc residue. As depicted in Figure 9, an unprotected glycopeptide bearing a proximal GalNAc residue was oxidized with commercially available galactose oxidase to produce the C-6 aldehyde (64). The aldehyde group was reacted with an aminooxy-sugar to afford a higher-order glycan with an oxime linkage replacing the β1,6 glycosidic bond. To generate O-linked glycopeptides with oligosaccharides extended at C-3 of the core GalNAc residue, a protected 3-thioGalNAc amino acid was incorporated into a glycopeptide fragment of GlyCAM-1 by Fmoc-based SPPS. Selective alkylation of the deprotected thiol with appropriate *N*-bromoacetamido di- or trisaccharides produced structures that correspond to the 2,3-sialyl TF (STF) and MECA-79 antigens on a mucin-like domain of GlyCAM-1 (Figure 10) (60, 68). The MECA-79 antigen was recently identified by the Fukuda group (69) and is uniquely expressed on endothelial mucins in lymph nodes and at sites of chronic inflammation (70, 71). Mucin fragments displaying the tumor antigen STF are also being investigated as breast cancer

vaccines (44, 72). The late-stage ligation of sulfated and sialylated carbohydrates to the glycopeptide avoids subjecting these sensitive groups to acidic deprotection after SPPS. To date, no native glycopeptides with sulfated oligosaccharides have been synthesized chemically, highlighting the significance of this ligation chemistry for producing biologically important structures. The competence of thioether-linked antigen structures for binding their corresponding antibodies is a topic of current investigation.

Synthesis of O-Linked Glycoproteins by Native Chemical Ligation

The methods described above for the construction of O-linked glycopeptides permit access to structurally diverse but relatively short glycopeptide fragments (~20 amino acids). Naturally occurring mucin-type oligosaccharides are typically present on proteins that far exceed this size. To surpass the size limits inherent in linear SPPS, the coupling of peptide fragments by native chemical ligation (NCL) technology has found widespread use (73). The ligation between two unprotected peptide segments, one bearing a *C*-terminal thioester and the other an N-terminal cysteine residue, affords the product peptide with a native amide bond at the ligation site. NCL is efficient and highly chemoselective, and the reaction conditions are entirely compatible with carbohydrates and native proteins. For these reasons, the extension of NCL to glycoprotein synthesis presents an ideal solution for accessing large glycoprotein structures.

Lymphotactin (Lptn) is a 93-amino-acid chemokine that serves as a potent chemoattractant for both T cells and natural killer cells (74, 75). With a small mucin-like domain located at its C terminus, Lptn is unusual, for relatively few chemokines are extensively O-glycosylated. The structural organization of Lptn is readily dissected by the NCL strategy into two synthetic peptides: a 47-residue peptide α-thioester (**14**) and a 46-residue glycopeptide (**15**) with eight α-GalNAc residues (Figure 11*a*). The thioester (**14**) was synthesized using traditional Boc-based SPPS methods (73), and Fmoc-based synthesis with α-*O*-GalNAc-Ser/Thr afforded the mucin domain of lymphotactin. Ligation of the two fragments cleanly gave the glycosylated chemokine, which was biologically active in a standard calcium mobilization assay (74). Synthesis of this glycoprotein by NCL has provided milligram quantities of homogeneous Lptn for structural and functional studies.

A chemically defined version of diptericin, an 82-residue glycoprotein, has also been prepared by NCL (Figure 11*b*). Containing a proline-rich sequence similar to drosocin and an attacin-like domain, this modular antimicrobial peptide carries potential O-linked glycosylation sites at Thr11 and Thr54 (75). Diptericin entirely lacks cysteine, and so Gly25 was strategically changed to the cysteine required for the NCL. Positioned between the drosocin- and attacin-like domains, this disconnection also made possible investigation of the isolated domains for biological activity. For generation of the acid- and base-sensitive N-terminal glycopeptide thioester, conventional Boc- and Fmoc-based methods to prepare

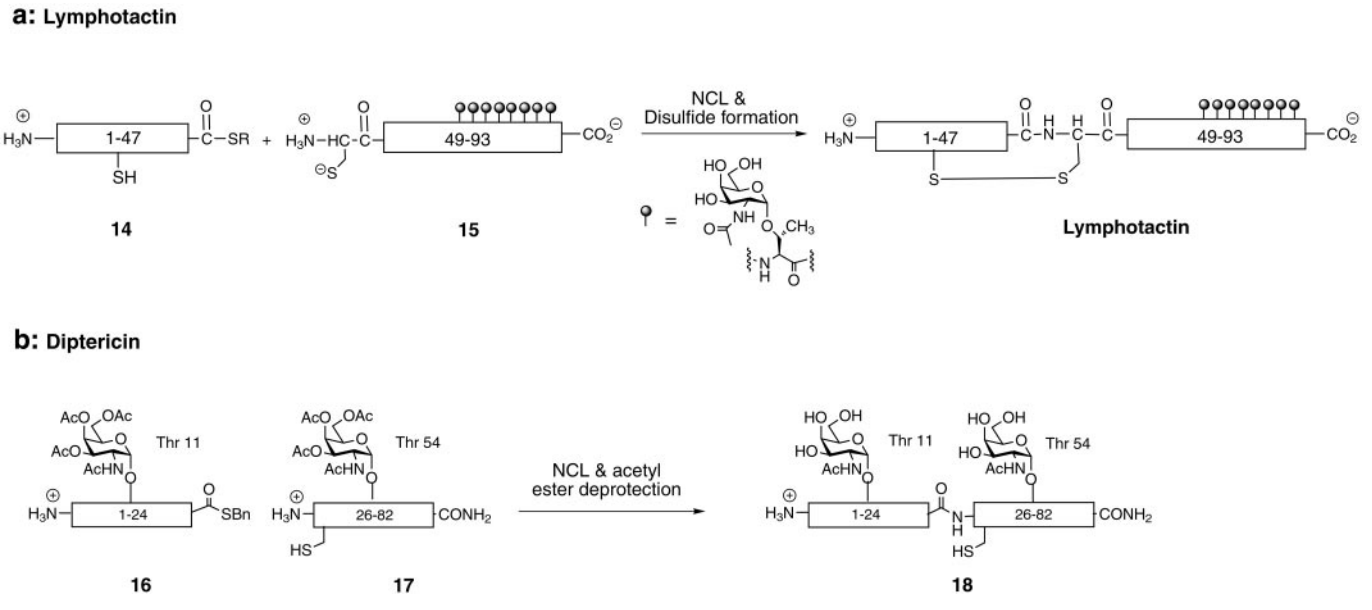


Figure 11 Native chemical ligation (NCL) of peptide thioesters to N-terminal cysteinyl peptides provides full-length glycoproteins: (*a*) lymphotactin, a chemokine for T cells and natural killer cells, and (*b*) diptericin, an antimicrobial peptide from insects.

the thioester could not be used. To overcome this obstacle, we utilized Fmoc-based SPPS on a sulfonamide "safety-catch" resin developed by the Ellman group (76) that allowed the release of peptide thioesters, under mild conditions, by nucleophilic addition of thiols (77). Glycopeptide thioesters for NCL can be routinely prepared by this method. Removal of side-chain protecting groups yielded glycopeptide-α-thioester **16**, which was ligated to the glycopeptide fragment **17** provided by Fmoc-based SPPS. NCL efficiently produced the full-length glycoprotein **18**, which inhibited bacterial growth with an IC_{50} of 2.70 ± 0.30 μM, similar to the potency of synthetic diptericin previously prepared in our laboratory (38).

Even by NCL, proteins larger than 20 kilodaltons (kDa) in size are difficult to prepare, for the serial concatenation of synthetic fragments by NCL does not entirely escape size limitations imposed by SPPS. Fortunately, recombinantly expressed peptides exceed these limits and can be used for NCL to prepare medium to large glycoproteins. The naturally occurring process of protein splicing (78) has been modified for the recombinant expression of proteins with the C-terminal thioesters necessary for NCL. The target protein or peptide is expressed from *Escherichia coli*, fused to an intein and a chitin-binding domain. This intein rearranges the peptide backbone at the site of fusion to an internal thioester, which by attack of exogenous thiols releases a protein thioester available for NCL. Termed expressed protein ligation (EPL), this technology is commercially available as an expression system and has been applied frequently for the semisynthesis of large, biologically active proteins (80, 81).

Using EPL, we have recently ligated expressed C-terminal thioesters with synthetic glycopeptides to construct glycoproteins such as GlyCAM-1, the 132-residue endothelial-derived ligand for L-selectin in murine lymph nodes (Figure 12). For illuminating the importance of the two mucin domains (**19** and **20**), a panel of the domains was produced by SPPS with various numbers and positions of O-linked GalNAc residues. By EPL of the recombinant peptide thioester fragment between the terminal fragments (**19** and **20**), a series of GlyCAM-1 glycoforms was constructed (82). Masking the N-terminal cysteine of the expressed fragments by fusion with a cleavage sequence from Factor Xa protease permitted the timed exposure of the cysteine for ligation to fragment **20**. Serial ligation of the three fragments yielded the target GlyCAM-1 containing 12 O-linked GalNAc residues. By extending NCL and EPL strategies to more complicated glycopeptide fragments generated by cassette, chemoenzymatic, and expression methods, previously inaccessible full-length proteins with complex but defined glycans should be more frequently attainable for biological investigations.

N-LINKED GLYCOSYLATION

An older and more prevalent form of glycosylation, N-linked glycosylation, is found in a wide range of organisms ranging from Archaea to mammals and other eukaryotes (83, 84). N-glycosylation is a modification performed cotranslation-

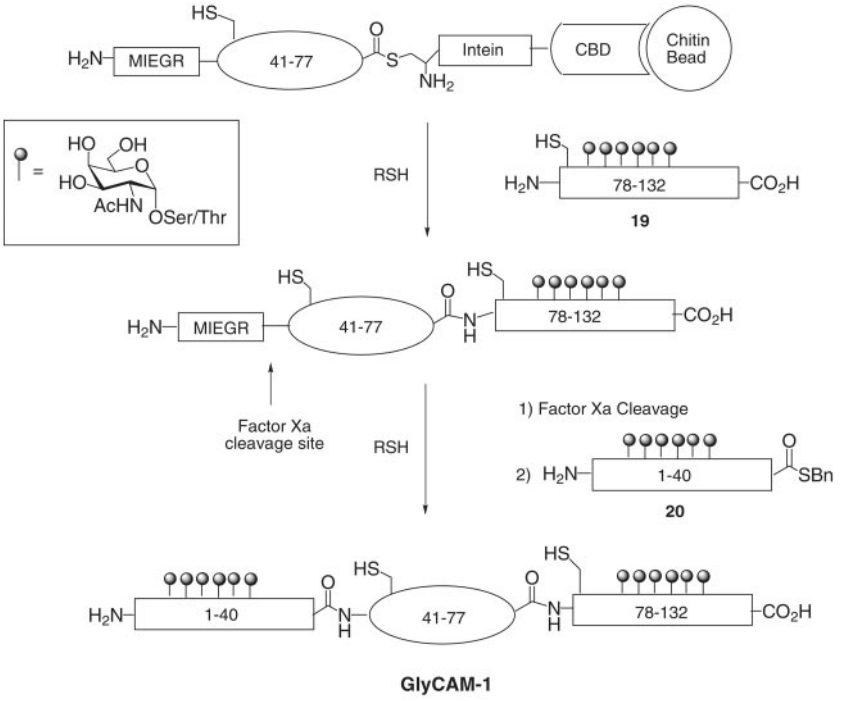


Figure 12 Synthesis of glycosylated GlyCAM-1 by expressed protein ligation (EPL) to synthetic mucin domains **19** and **20**. Factor Xa proteolysis allows late-stage exposure of an N-terminal cysteine for the timed NCL to **20** (CBD, chitin-binding domain).

ally and available to any secreted or membrane-bound protein containing the triplet amino acid sequence AsnXaaSer/Thr (where Xaa is any amino acid except Pro). An oligosaccharide is transferred to the amide side chain of Asn, via a dolichol phosphate glycosyl donor, by the action of oligosaccharyl transferase located in the membrane of the endoplasmic reticulum (ER) (11, 85). The fully translated glycoprotein is then subjected to trimming, processing, and further glycosylation in the ER and Golgi apparatus. Between 70 and 90% of potential N-glycosylation sites appear to be occupied (86). The $Man_3GlcNAc_2(\beta\text{-}N)$Asn structure is the ubiquitous pentasaccharide core found in all N-linked glycans (Figure 13). This core oligosaccharide can be extended in three broadly defined fashions to produce high-mannose, complex, and hybrid type glycans. As for O-linked glycans, differing numbers of repeating units of galactose/GalNAc and GlcNAc, addition of sialic acid and/or fucose, and other modifications such as sulfation lead to a vast array of highly complex structures.

The functions of N-linked glycosylation are wide-ranging and not understood in the case of every protein. However, they can be functionally divided into two

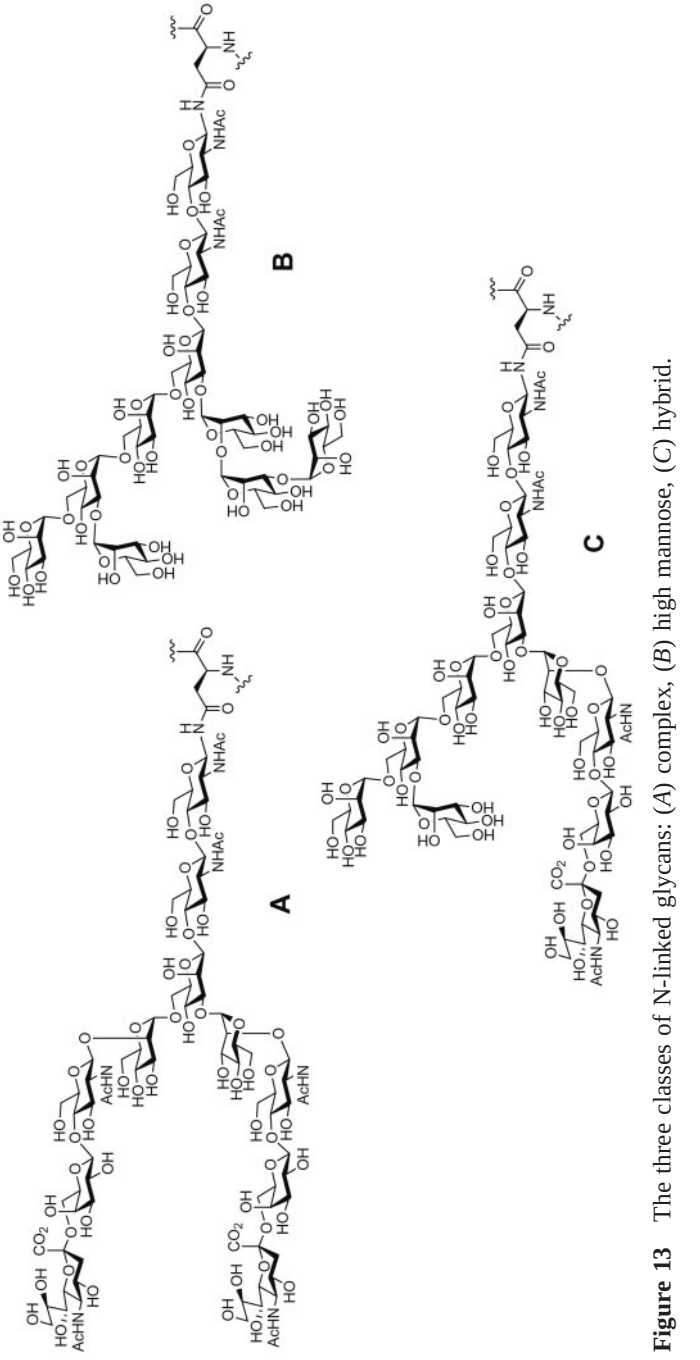


Figure 13 The three classes of N-linked glycans: (*A*) complex, (*B*) high mannose, (*C*) hybrid.

types: intra- and extracellular. Intracellularly, the broad function of N-linked glycans is protein trafficking (85). This is elegantly exemplified by the quality control mechanism for protein folding in the cell that involves a chaperone system found in the ER of nearly all eukaryotes, the calnexin-calreticulin cycle. Calnexin and calreticulin are related ER lectins that interact with N-linked structures bearing a single glucose residue exposed by the trimming action of glucosidases I and II (87). Calnexin and calreticulin binding slows the trafficking process and allows for proper protein folding and disulfide bond isomerization (88). While the glycoprotein is retained in the ER, glucosidase II removes the last glucose residue of N-linked glycans and causes release of the protein from the lectin-like chaperones. In a poorly understood process, a glucosyltransferase simultaneously acts as a folding sensor, apparently replacing a glucose residue only on improperly folded proteins, which causes their binding to calnexin and calreticulin and return to the folding cycle of the ER (89).

Extracellularly, N-linked glycans can function as structural elements and as ligands for receptors. Structurally, the large and flexible N-linked glycans can increase protein stability by restricting the conformational flexibility of the underlying protein without sacrificing net entropy of the system. In a thermodynamic study of an ovomucoid protein domain, DeKoster & Robertson demonstrated that two glycans found on Asn10 and Asn52 increased the melting temperature of the 68-residue polypeptide by 4.8°C (90). An N-linked glycan can also affect the local structure and stability of a protein, as revealed by an NMR structure of a soluble form of human CD2 solved by the Wagner group (91). CD2 is a cell-surface glycoprotein present on T lymphocytes and natural killer cells, and the attachment of an N-linked glycan at Asn65 is necessary for its binding to CD58. The glycan was found to have conformational rigidity similar to a folded polypeptide. As measured from nuclear Overhauser effects (NOEs), the protein-proximal GlcNAc-GlcNAc disaccharide was in close contact with a cluster of charged and polar residues located on one face of a β-sheet. The proper orientation of this sheet was required for folding of the CD58 binding site. In this case, as well as others (92), the glycan acts in concert with the polypeptide to orchestrate the overall structure and function of the protein.

N-Linked glycans can also function extracellularly as ligands for carbohydrate receptors, as highlighted by the glycoprotein growth factors and hormones. For example, erythropoietin (EPO) is synthesized by the kidneys and circulated in the blood to stimulate red cell proliferation and differentiation in bone marrow. The carbohydrates on EPO consist of one O-linked and three N-linked glycans to make up 40% of the protein's total weight. The Fukuda group has shown that variation in the carbohydrate content of the N-linked glycans alters the serum half-life of EPO and thus alters its activity in vivo (93). Asialo-erythropoietin has no measurable activity in vivo because of its rapid clearance from the bloodstream. EPO glycoforms containing *N*-acetyllactosamine (LacNAc) repeats were similarly cleared from serum circulation in a rapid fashion. In both cases, exposed galactose residues were recognized by the hepatic asialoglycoprotein

receptor, and the hormone was internalized by endocytosis and degraded in the lysosome. Thus, EPO must be glycosylated in a very specific fashion to retain activity in vivo.

Another example of glycan-specific hormonal regulation is provided by equine lutropin (eLH) and chorionic gonadotropin (eCG). These are heterodimeric glycoproteins synthesized in the placenta and pituitary, respectively. Both the α and β subunits of eLH and eCG are coded by a single gene and differ only in the makeup of their N-linked glycan (94). Tissue-specific expression of a GalNAc transferase generates the terminal structure found on the eLH glycan, 4-(SO_3)-GalNAc-β1,4-GlcNAc. eLH is bound by a hepatic endothelial cell receptor specific for the carbohydrate sequence and is removed from circulation 5.7-fold more rapidly than eCG. As this glycosylation is the only structural difference between eLH and eCG, the concomitant difference in circulatory half-life probably generates the functional distinctions between the two hormones.

Recombinant Methods for the Production of N-Linked Glycoproteins

The recombinant production of homogeneous N-linked glycoproteins would provide material for important structural and functional studies. This has been accomplished with some success by enzymatic modification of purified glycoproteins in order to increase glycan homogeneity. For example, Raju and coworkers (95) expressed tumor necrosis factor receptor (TNFR) as an immunoadhesin molecule (TNFR-IgG) for anti-inflammatory therapy. The glycoprotein contained four N-linked glycosylation sites bearing heterogeneous glycans. Terminal galactose or GlcNAc residues on the glycans were shown to produce rapid clearance of the glycoprotein as a result of interactions with galactose-binding receptors of the liver. Subjection of the recombinant glycoprotein to in vitro galactosylation using bovine β1,4-galactosyltransferase (GalT) and sialylation using rat liver α2,3-sialyltransferase (SiaT) in a one-pot procedure produced TNFR-IgG with a homogeneous terminal sialylation pattern. As intended, in vitro glycosylation increased the serum half-life of TNFR-IgG and should be applicable to other therapeutic glycoproteins as well.

With the goal of producing recombinant glycoproteins with the longest serum half-lives possible, the Krummen group engineered CHO cells to increase the bulk sialic acid content of overexpressed proteins (96). Two CHO cell lines were engineered to express high levels of human β1,4-GalT and/or α2,3-SiaT. Overexpression of the SiaT led to sialylation of $>$90% of N-linked glycans on TNFR-IgG, whereas the overexpression of GalT gave a significant reduction in exposed terminal GlcNAc residues. When both transferases were overexpressed, an increase in sialylation and a decrease in the overall length of the N-linked glycans were observed. These results are encouraging with regard to the production of glycoprotein therapeutics with optimized glycans.

Chemoenzymatic Synthesis of N-Linked Glycopeptides

Chemical synthesis performed in concert with additional enzymatic elaborations is now a widely used approach for the synthesis of N-linked glycopeptides. The transglycosylation reaction enacted by *Mucor hiemalis* endo-*β*-*N*-acetylglucosaminidase (Endo-M) has attracted particular attention (Figure 14). Haneda & coworkers applied the enzyme in their synthesis of glycosylated versions of calcitonin, a 32-amino-acid hormone that regulates calcium and is used as a therapeutic agent for hypercalcemia, Paget's disease, and osteoporosis (97). To study the consequences of glycosylation at a potential N-linked site, a synthetic glycopeptide was prepared containing a single GlcNAc residue on the Asn of the triplet sequence. Transglycosylation by Endo-M using a di-sialo transferrin glycosyl amino acid (STF-GP) as the glycosyl donor gave the desired glycosylated calcitonin in 8.5% isolated yield (Figure 14). The glycosylated version of the hormone had increased biological activity in vitro.

To understand the influence of an N-linked glycan on local peptide conformation, the Imperiali and Yamamoto groups undertook the synthesis of several differentially glycosylated fragments of the extracellular domain of the nicotinic acetylcholine receptor (nAChR) (Figure 15) (98). In another powerful example of the Endo-M methodology, the nAChR domain **21** and its simple glycoforms **22** and **23** were prepared by traditional SPPS. Glycopeptide **24** was synthesized by Endo-M-mediated transglycosylation as described above, with **22** serving as the glycosyl acceptor. The distal sugar residues of glycopeptide **24** were then enzymatically cleaved to produce glycopeptides **25** and **26**. NMR studies revealed several NOEs between the Asn residue and the proximal GlcNAc residue in glycopeptide **24**. These signals were much weaker in glycopeptides **22** and **23**, which suggests that distal sugar residues may rigidify the proximal core disaccharide. The addition of glycans also increased disulfide bond formation in the peptide, and the rate of *cis*/*trans* proline isomerization of **21** was twofold faster than those of glycosylated derivatives **22–26**. This study implies that N-linked glycosylation can alter both protein stability and the dynamic process of protein folding.

Unverzagt has applied an alternative chemoenzymatic approach to the synthesis of N-linked glycopeptides with impressive success (99–103). A synthetic core glycopeptide or glycosyl amino acid was elaborated at its terminus through the successive use of glycosyltransferases. This technique takes advantage of commercially available enzymes to ease the task of synthesizing a large oligosaccharide. In one example, synthetic glycopeptide **27**, containing a septasaccharide, was treated with a galactosyltransferase and UDP-Gal, followed by CMP-Sia and an α2,6-sialyltransferase (Figure 16). This produced in 91% yield the fully elaborated glycopeptide **28** with an undecasaccharide moiety identical to complex biantennary glycans found in human ribonuclease. The extension of this technique to glycoproteins is within the bounds of the foreseeable future.

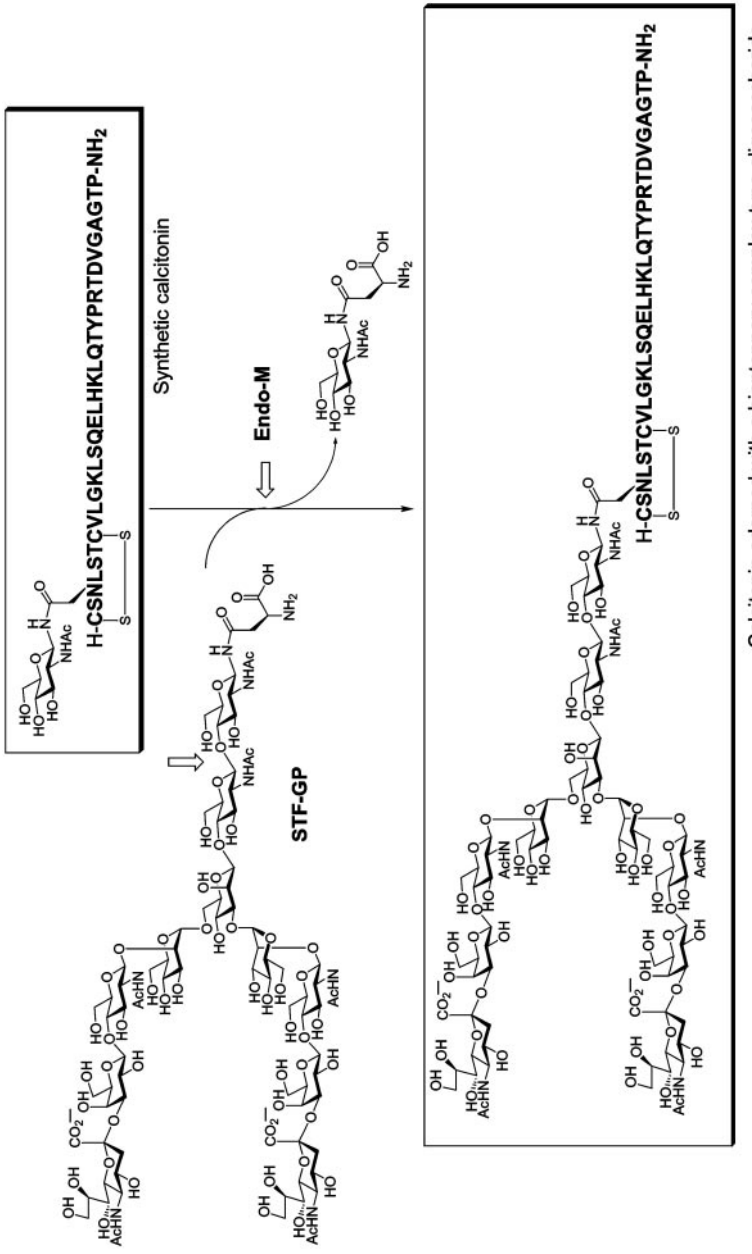


Figure 14 Synthesis of a calcitonin derivative with a complex-type N-linked oligosaccharide by transglycosylation. Endo-M transfers the biantennary glycan from the donor glycosyl amino acid (STF-GP) derived from transferrin, liberating GlcNAc-Asn.

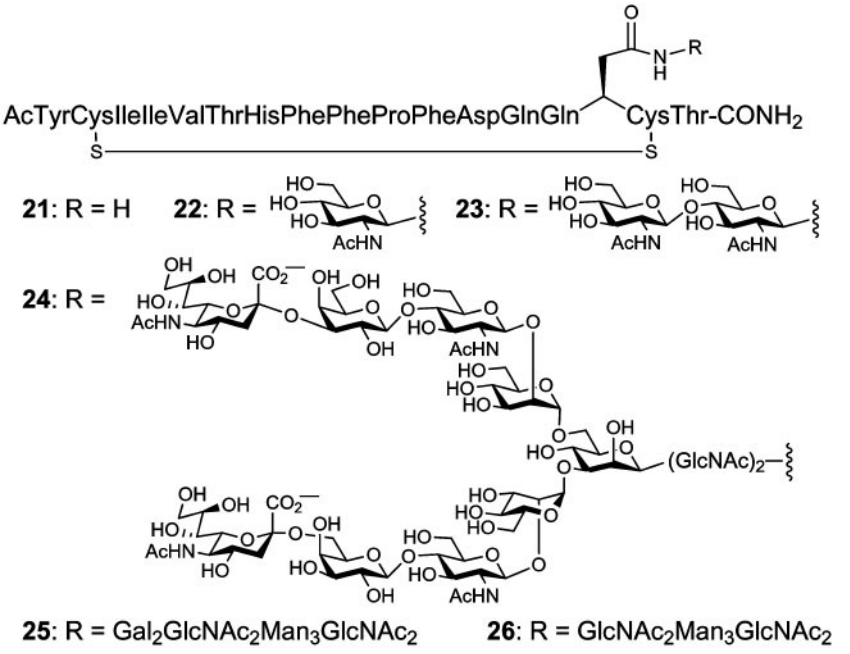


Figure 15 Peptide fragments from the extracellular domain of the nicotinic acetylcholine receptor (nAChR) (**21**, **22,** and **23**) prepared by SPPS. Several glycoforms were constructed by the sequential action of Endo-M on **22** to produce **24** followed by glycosidase digestion to afford **25** and **26**.

Chemical Synthesis of N-Linked Glycopeptides and Proteins

Complex N-linked glycopeptides are commonly synthesized by a cassette method similar to the strategy for O-linked glycopeptide synthesis, where a complex glycosyl amino acid is incorporated into a growing peptide chain using standard Fmoc-based SPPS. The reactivity considerations and protecting group manipulations are similar to the O-linked syntheses. Glycosyl-β-*N*-Asn, building block **31**, is generally prepared by coupling of a reactive glycosylamine (**29**) with an activated ester of aspartic acid (**30**, Figure 17). The glycosylamine can be prepared by reduction of a glycosylazide under mild conditions (104–107). Ogawa and coworkers have successfully applied this method to produce a fully synthetic CD52 glycopeptide with a core pentasaccharide glycan (Figure 17) (104, 105). Monosaccharide building blocks were combined to produce the fully protected azido-pentasaccharide **32**, which was then converted to the glycosylamine and coupled to Fmoc-Asp-*t*Bu to produce the building block for incorporation into peptide **33**.

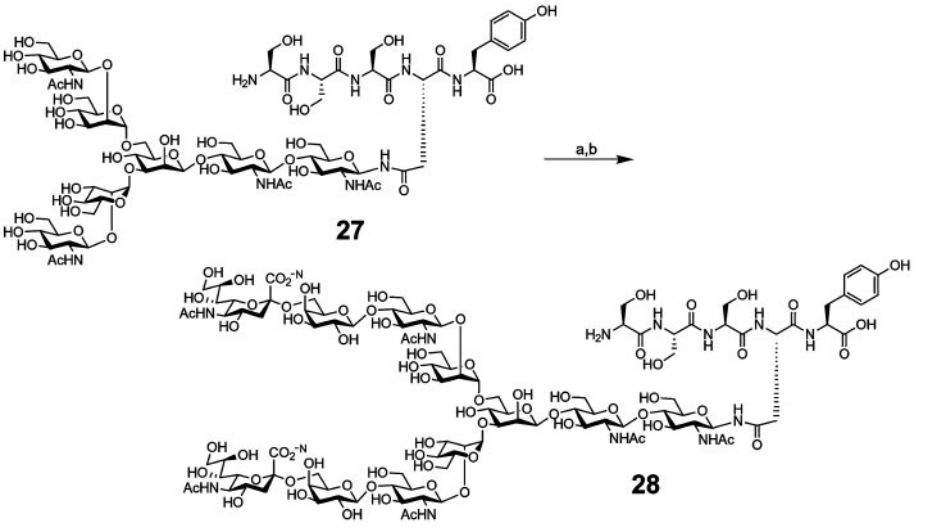


Figure 16 Chemoenzymatic synthesis of a complex-type N-linked glycopeptide (**28**) in two steps: treatment with (*a*) UDP-Gal and β1,4-galactosyltransferase, and with (*b*) CMP-Sia and α2,6-sialyltransferase (91% overall yield).

The condensation of a free reducing sugar with ammonium bicarbonate, commonly known as the Kochetkov method, also produces a glycosylamine (Figure 18) (108). Several groups have used resulting glycosylamines in the synthesis of glycosyl amino acid cassettes (109–112). Meldal has extended this method to the production of N-linked glycopeptides using glycans retrieved from natural sources (109). To this end, bovine fetuin containing three complex-type oligosaccharides and ribonuclease B containing one high-mannose oligosaccharide were subjected to hydrazinolysis to give β-glycosylhydrazines (Figure 18). These glycosylhydrazines were transformed into glycosylamines for coupling to aspartic acid by way of the Kochetkov method. The resulting cassettes were purified by size exclusion chromatography and incorporated into peptides using SPPS, which produced homogeneous glycopeptides with extremely large and complex glycans.

The convergent coupling of glycosylamines to the aspartyl side chains of a protected peptide has also been achieved. Danishefsky and coworkers have applied the approach to short peptides bearing very large N-linked oligosaccharides (110–113). Pentasaccharide **35** was synthesized using the "glycal assembly method" (111), converted to glycosylamine **36**, and then coupled to a pentapeptide to produce glycopeptide **37** in good yield (Figure 19). These synthetic achievements have enabled landmark studies of the stereochemical communication between the carbohydrate and peptide domains. NMR studies undertaken by Live and coworkers compared two glycopeptides, differing only in the absolute stereochemistry of the amino acids (L-peptide versus D-peptide) (110). Both

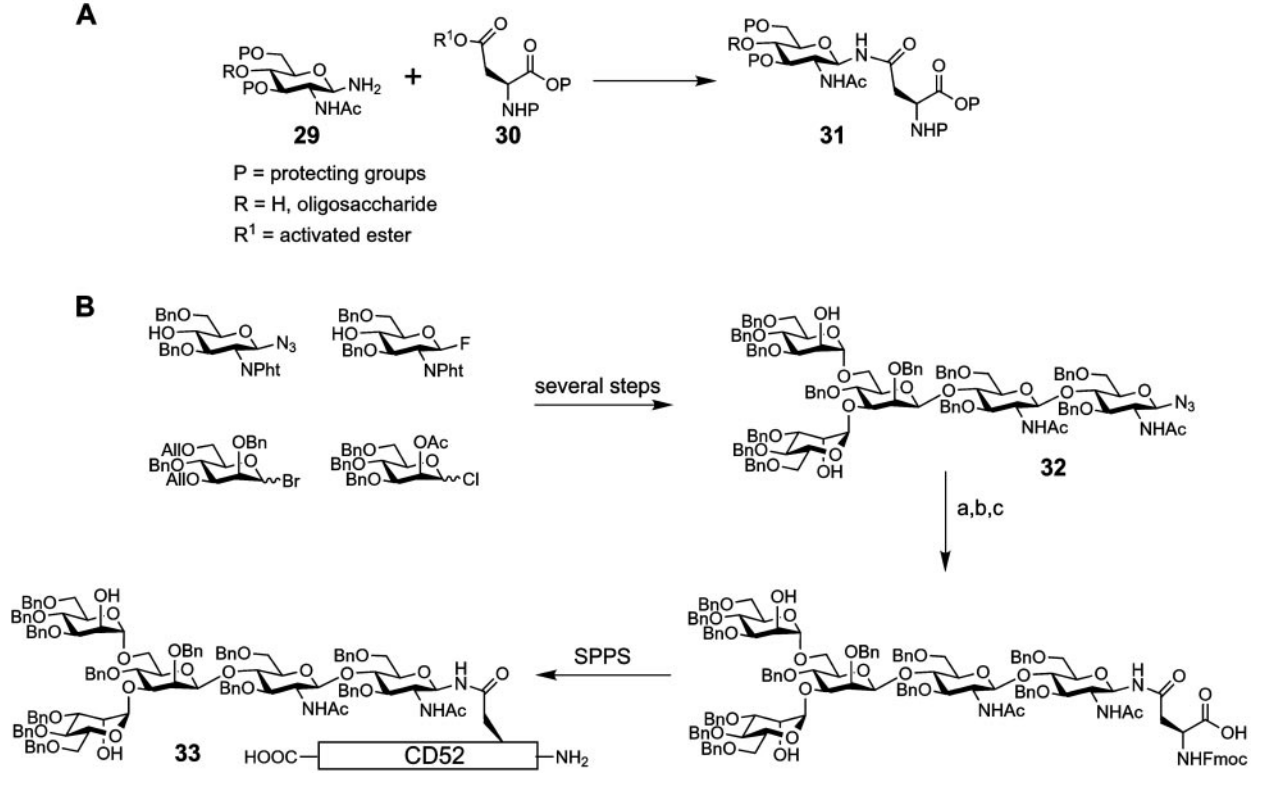


Figure 17 (*A*) General synthesis of N-linked glycosyl amino acids by coupling of glycosylamines to aspartic acid. The resulting cassettes can be used in SPPS. (*B*) Synthesis of soluble CD52 from a pentasaccharide-modified amino acid includes: (*a*) azide reduction with Lindlar's catalyst; (*b*) amide formation with Fmoc-aspartic acid-*t*-butyl ester (75% yield); (*c*) trifluoroacetic acid/CH_2Cl_2 (98% yield).

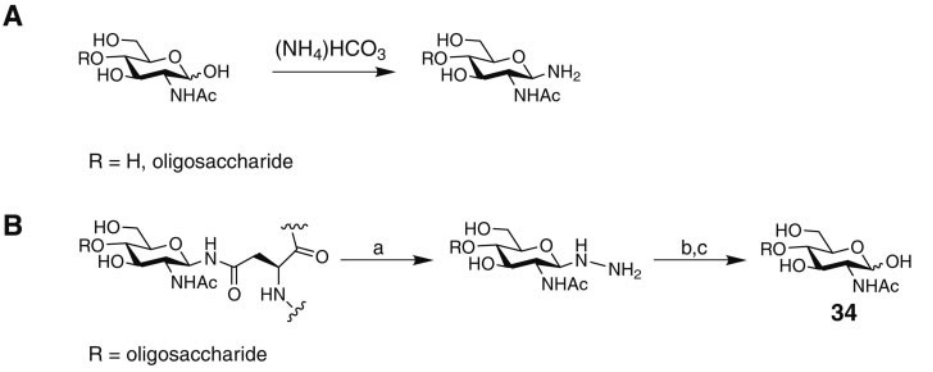


Figure 18 (*A*) Synthesis of glycosylamines by the Kochetkov method. (*B*) Production of free reducing oligosaccharides from glycoproteins: (*a*) NH_2NH_2, 85°C; (*b*) Ac_2O; (*c*) (i) copper(II)acetate, H_2O, (ii) 25 mM H_2SO_4, 80°C. The free sugars can be converted to glycosylamines via the Kochetkov method.

peptides adopted a type I β-turn, but there were measurable differences between the stereochemically "matched" peptide and the "mismatched" one. Thus, communication between the carbohydrate and polypeptide domains of a glycoprotein is not solely based on the bulk of the carbohydrate. Rather, specific interactions between the polypeptide and the carbohydrate are governed by their precise structures. Finally, in a synthetic tour de force, Danishefsky and coworkers undertook the synthesis of the H-type 2 blood group antigen on a pentapeptide scaffold (Figure 20) (111). The glycopeptide **38** was shown to react with an antibody against H-type 2 determinants in an enzyme-linked immunosorbant assay (ELISA).

Chemoselective Ligations for the Synthesis of N-Linked Glycopeptides and Proteins

With the difficulties inherent in oligosaccharide synthesis and the limitations imposed by SPPS technology, the above synthetic techniques have been applied only to short peptides. Although enzymatic methods have found utility in constructing glycoproteins, the enzymatic remodeling of a glycan is difficult to execute on a large scale, and the enzymatic construction of natural epitopes is limited by the availability of enzymes and natural substrates. As for the construction of O-linked glycopeptides and glycoproteins, chemoselective ligations are powerful methods for accessing larger and more complex N-linked glycopeptides (16, 59). Toward the goal of synthesizing a neoglycopeptide containing an N-linked glycan mimic, our group performed the synthesis of thiosemicarbazide chitobiose derivative **39** from chitobiosylamine obtained by the Kochetkov method (Figure 21) (61). A ketopeptide scaffold (**40**) was prepared by SPPS, incorporating (2*S*)-aminolevulinic acid, and was ligated with **39** to access the neoglycopeptide **41**.

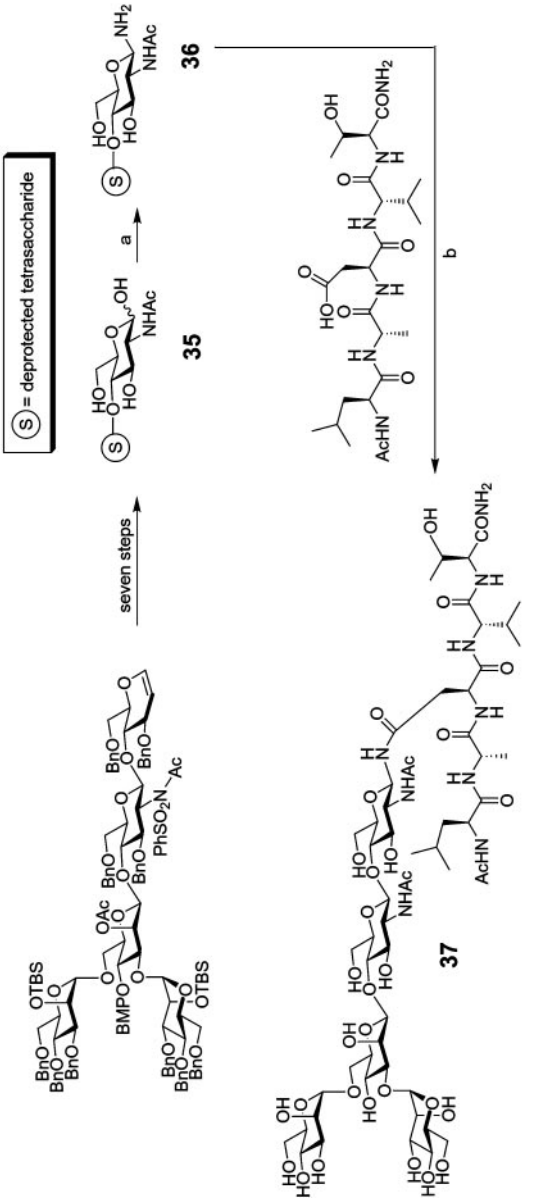


Figure 19 Synthesis of a pentasaccharide-modified glycopeptide (**37**): (*a*) $(NH_4)HCO_3$, H_2O (95%); (*b*) amide formation: HOBt and HATU in DMF (40% yield).

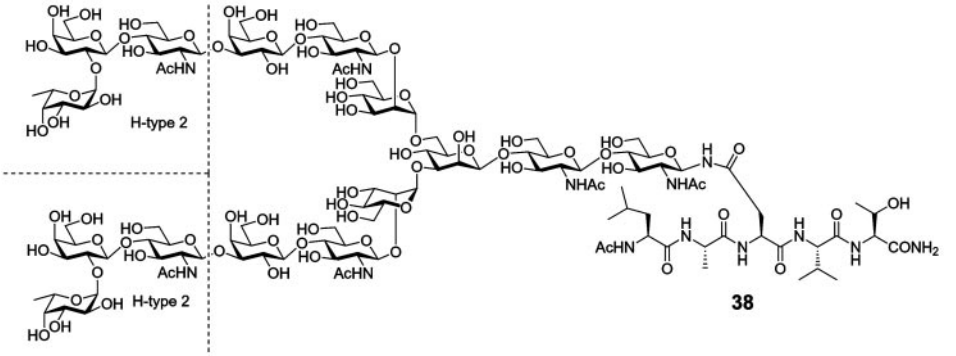


Figure 20 Fully synthetic glycopeptide **38** bearing a pentadecasaccharide with H-type 2 termini, prepared by coupling of the relevant glycosylamine to an aspartyl peptide.

In a unique entry to N-linked glycoproteins, the laboratories of Flitsch (114, 115) and others (116) have reversed the nucleophilic and electrophilic partners for chemoselective ligation by reacting α-haloacetamido sugars with introduced thiol nucleophiles. The paramount example of this approach is Flitsch's semi-synthesis of homogeneous glycoforms of human erythropoietin. EPO contains three natural N-glycosylation sites at Asn24, Asn38, and Asn83. Asparagine to cysteine mutants at each site were expressed recombinantly in *E. coli* and oxidatively refolded to produce EPO with the two native disulfide bonds and a remaining cysteine residue. Treatment with α-iodoacetamido-GlcNAc produced the neoglycoprotein (Figure 22). Analysis by protein digestion and mass spectrometry revealed the protein to be alkylated only at the mutant cysteine residue. The introduction of the nonnative linkage produced no noticeable structural perturbations when compared to native EPO by ^{1}H-NMR analysis. This impressive example highlights the relative ease with which chemoselective ligations can be carried out and their relevance to very large proteins.

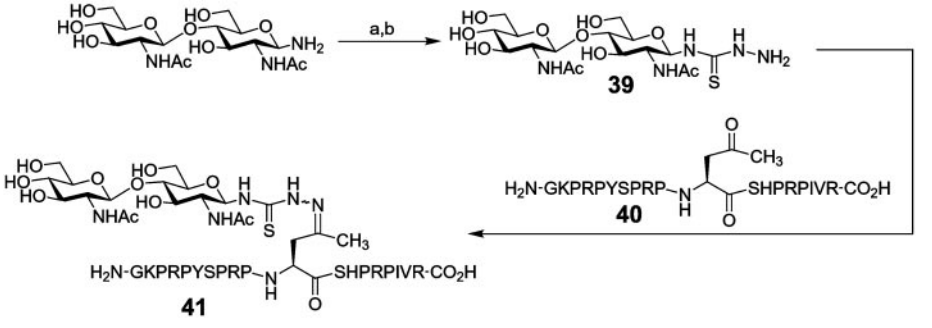


Figure 21 Synthesis of neoglycopeptide **41** with a thiosemicarbazide linkage to chitobiose: (*a*) Cl_2CS, 0.3 M $NaHCO_3$; (*b*) $NH_2NH_2{\cdot}H_2O$ (70% yield).

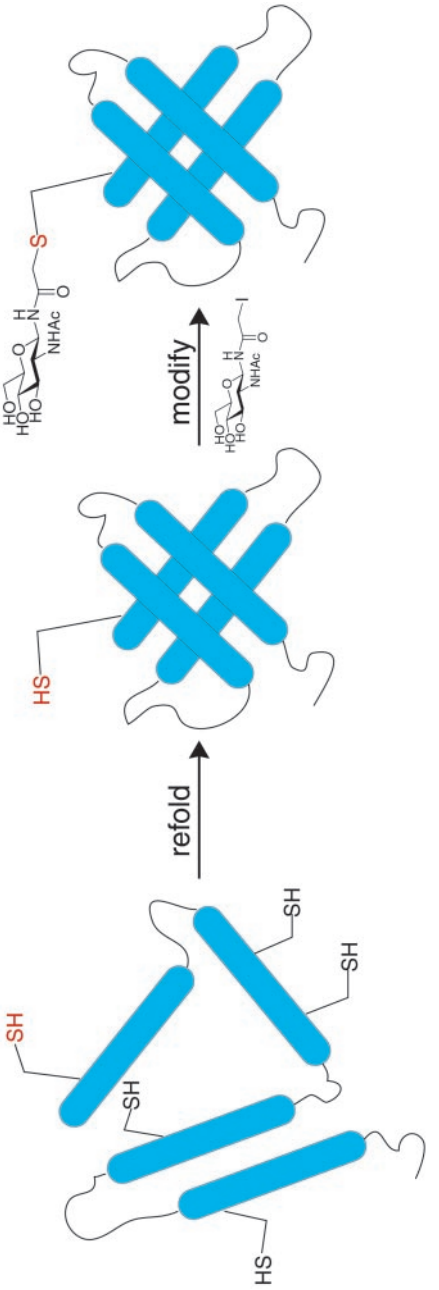


Figure 22 Cysteine mutants of erythropoietin (EPO) were refolded to expose a single free thiol, which was alkylated with *N*-iodoacetamido GlcNAc.

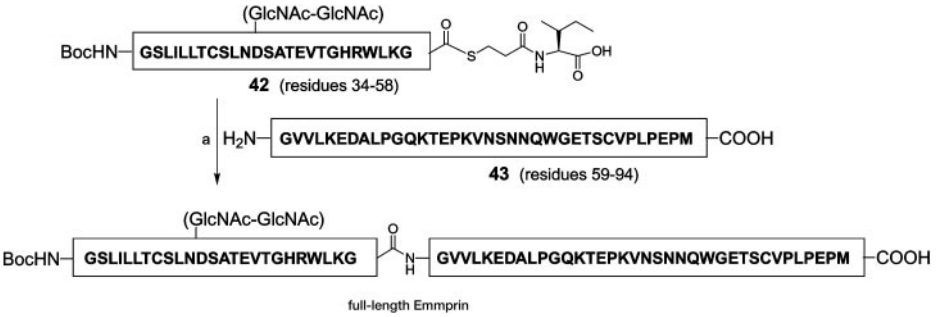


Figure 23 Synthesis by native chemical ligation (NCL) of Emmprin containing a chitobiose unit, by activation of thioester **42** to attack by amine **43**: (*a*) AgCl and HOBt in dimethyl sulfoxide (77% yield).

Synthesis of N-Linked Glycoproteins by Native Chemical Ligation

To meet the challenge of preparing large N-linked glycoproteins, NCL and EPL (117) have again arisen as the answer. Tolbert & Wong ligated a glycosylated N-terminal cysteine dipeptide, Cys-Asn(GlcNAc), to expressed maltose binding protein containing a C-terminal thioester (118). The methodology could in principle be extended to ligations with biologically relevant N-linked glycopeptide fragments. In a recent extension of NCL to N-linked glycoproteins (Figure 23), Hojo and coworkers synthesized the extracellular immunoglobulin domain I of Emmprin (119), a glycoprotein located on the surface of human tumor cells that stimulates nearby fibroblasts to produce matrix metalloproteinases. Asparagine bearing a benzyl-protected chitobiose disaccharide (GlcNAc-β1,4-GlcNAc) was used for the preparation by Boc-based SPPS of peptide thioester **42**, a fragment corresponding to residues 34–58 of Emmprin (73). A C-terminal peptide (**43**, residues 59–94) was also prepared by SPPS. The two segments were condensed by activation of the thioester with silver chloride to obtain the fully protected form of Emmprin, residues 34–94, in 77% yield. The protein was then deprotected and subjected to oxidative disulfide bond formation to furnish the desired glycoprotein with a chitobiose unit. Biological studies by the same group are under way.

GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED PROTEINS

Lipid modifications to proteins are ubiquitous like O- and N-linked glycosylations, yet are less easily addressed by chemical and biochemical methods. The chemoselective methods used for glycoprotein construction are nonetheless broadly relevant to accessing lipoproteins and examining the roles of lipid and

glycolipid posttranslational modifications. As an example, the obstacles to studying glycosylphosphatidylinositol (GPI) proteins by traditional methods highlight the importance of preparing pure, lipid-modified proteins. In species that range from Archaea to humans, many proteins are attached to the outer leaflet of the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor, a complex lipid-oligosaccharide attached to the protein C terminus (**44**, Figure 24) (120). The anchor structure can differ between species, but the protein C terminus is invariably amide-linked via a phosphoethanolamine bridge that is attached to the conserved GPI core glycan: -6-Man-α1,2-Man-α1,6-Man-α1,4-GlcN-α1,6-*myo*-inosityl-1-phospholipid. The membrane-associated lipid is typically composed of saturated 14- to 18-carbon esters and ethers to glycerol, although *Saccharomyces cerevisiae* does adopt sphingomyelin rather than glycerides. Some anchor forms carry an additional 2-inosityl fatty ester that blocks phosphatidylinositol-specific phospholipase C (PI-PLC) release of the GPI glycan from the phospholipid. The core pentasaccharide can bear additional phosphoethanolamines (PEtN) and mannose (Man), GalNAc, or sialic acid residues, depending on the particular protein and organism (121). In mammalian cells, GPI-anchored proteins are involved in fundamental processes such as signal transduction, cell membrane processing and homeostasis, and immune responses, as well as the pathobiology of cancers and infectious disease (122). This section describes how the structural complexity and curious biology of GPIs are yielding to investigation by methods of glycoprotein synthesis.

GPI anchors are biosynthesized in the endoplasmic reticulum (ER), where they are attached to an expressed protein before exit from the ER and transport to the cell surface through the secretory pathway. GPI-linked proteins freely diffuse in the lipid membrane, but many are found on mammalian cells associated with other membrane-bound proteins in cholesterol- and sphingolipid-enriched microdomains, termed lipid rafts (123). Their clustering has been detected by depolarization-FRET (fluorescence resonance energy transfer) (124) and chemical cross-linking (125). Diffusion measurements using an optical trap revealed that individual GPI proteins associate with rafts for up to several minutes (126). The structural basis for this association is unclear, but the preponderance of saturated versus unsaturated lipids on GPI anchors probably drives the localization into these higher-melting, detergent-insoluble lipid domains, even in model membranes (127). Raft-associated GPI proteins are essential factors for many cellular phenomena such as immune synapse formation (128), virus/host infection (129, 130), and the pathogenesis of human prion, a GPI protein (131).

The ostensible function of the GPI anchor is to physically position proteins on the cell surface and perhaps in membrane microdomains, but the general purposes or benefits of the complex pentasaccharide structure are less easily delineated. To access defined anchor structures, the total synthesis of the GPIs has been accomplished by several groups (132–134). Although arduous, these syntheses provide materials for elucidating specific interactions of the GPI

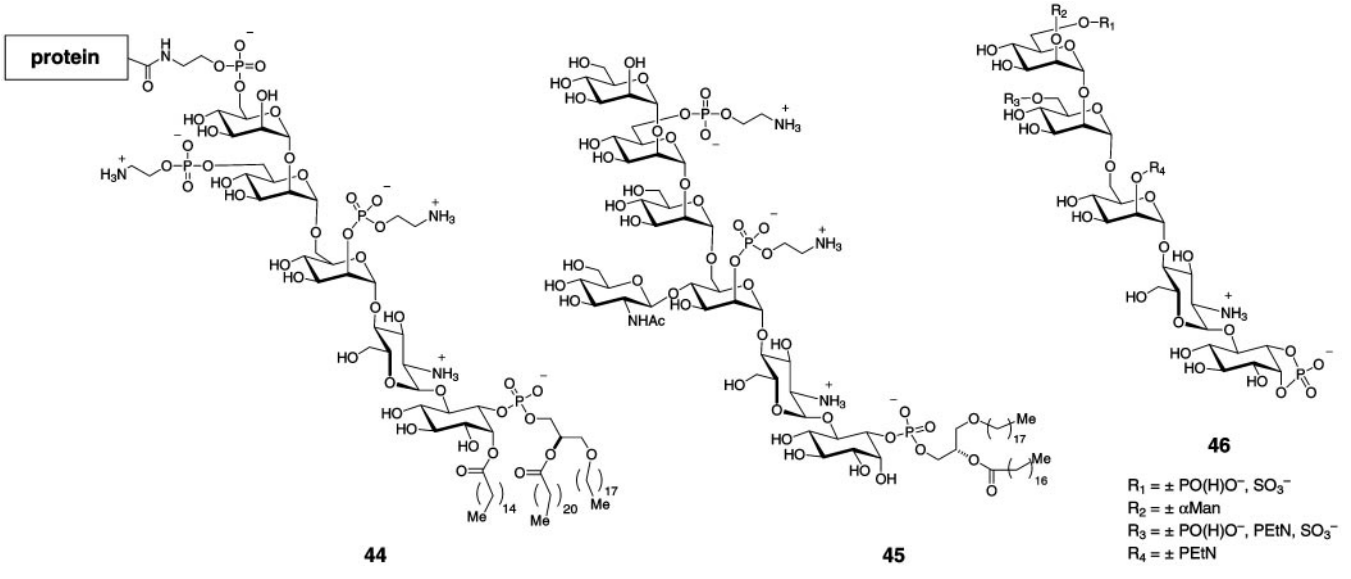


Figure 24 Glycosylphosphatidylinositol (GPI) anchor structures: **44**, a typical mammalian core structure; **45**, synthetic rat Thy-1 anchor; **46**, representative phosphoinositolglycan (PIG) structures synthesized to probe GPI-dependent stimulation of insulin metabolism. Abbreviations: PEtN, phosphoethanolamine; Man, mannose.

structure with other cellular factors. In studies to better understand immune recognition of parasites, protozoal pathogen GPI anchors were found to be bound and presented on natural killer T (NKT) cells by CD1d, a protein similar to major histocompatibility complex I (MHC-I) (135). Synthetic rat Thy-1 GPI anchor (**45;** Figure 24) and anchor structures isolated from parasites were so presented and examined for their ability to activate NKT cells for production of an MHC-independent immunoglobulin G (IgG) response. The structural requirements and mechanism of NKT cell activation suggest that the CD1d display of pathogen anchors may provide a general, MHC-independent antibody response to parasite infection. In a second study, an impressive array of synthetic phosphoinositol glycans (PIGs) such as **46** (Figure 24), representing cyclic phosphates produced by the delipidating action of PI-PLC on GPI proteins, were synthesized and examined for stimulation of glucose and lipid metabolism (136). This study demonstrated that PIG structures are specific and physiologically potent signaling molecules that mimic insulin activity, which confirmed previous hypotheses. The chemical syntheses of defined GPI structures were essential for these studies because pure anchors with native and nonnative structures are not easily extracted from cells.

GPI-protein deposition onto cell membranes is a promising avenue for exploring GPI anchor function without succumbing to the limitations of genetic methods. Expressed and purified GPI proteins spontaneously incorporate into cell membranes and impart their specific activities and functions (137), a useful feature for exploring the biology of GPI proteins in a native context. In an effort to engineer organs for xenotransplantation, decay-accelerating factor (DAF, CD55) was deposited onto sheep erythrocytes in culture to protect cells from human complement-mediated lysis (138). The incorporation of GPIs into membranes can be controlled to give defined numbers of pure, active proteins on the cell surface, a parameter difficult to control by recombinant expression (139). A remarkable use of this control was demonstrated by the deposition of MHC-peptide complexes onto cytotoxic T-cell targets (140). By preloading recombinant MHC-GPI with synthetic peptides from hepatitis B virus, defined antigen structures and numbers could be presented on cells and examined individually for ability to activate cytotoxic T cells.

In a landmark example of combining protein semisynthesis with cellular experiments to dissect the role of posttranslational lipidation, a truncated form of Ras was expressed and chemoselectively conjugated by a C-terminal cysteine to the maleimides of several synthetic lipopeptides (Figure 25*a*) (141). The two native sites of farnesylation and palmitoylation on the C-terminal peptide fragment were synthetically varied to display native, deleted, or unnatural lipids, and the Ras constructs were microinjected into cells to initiate cell differentiation. The synthetically lipidated Ras was fully active, and results with nonhydrolyzable hexadecyl lipids suggested an important role for lipid turnover in Ras activity.

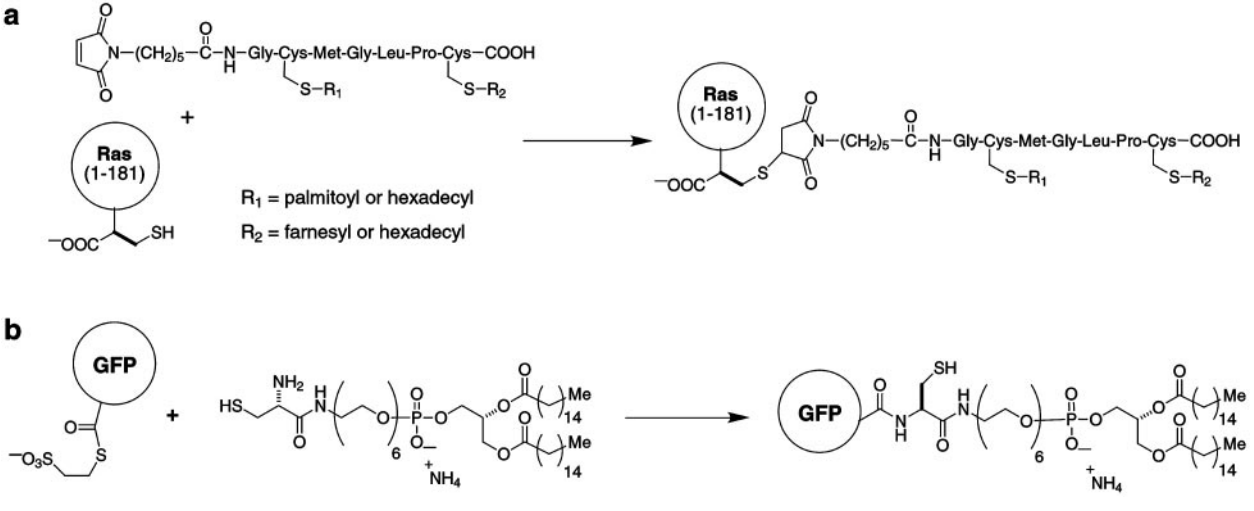


Figure 25 Synthesis of lipoproteins by chemoselective ligation: (*a*) maleimide alkylation of truncated Ras with lipidated C-terminal peptides; (*b*) expressed protein ligation (EPL) of a GPI anchor analog to green fluorescent protein (GFP) thioester.

A parallel study for GPI proteins has not been performed, because the complex anchor structures are typically synthesized in small quantities and without means for ligation to proteins. To address GPI proteins with chemical control, we have developed the means to ligate synthetic lipids to the C termini of proteins by use of EPL (Figure 25*b*) (M.J. Grogan, R.M. Conrad, C.R. Bertozzi, in preparation). In a model starting with simple cysteine-bearing lipids as GPI surrogates, the native chemical ligation to expressed green fluorescent protein (GFP) thioester cleanly produces lipo-GFP. Essential to the ligation was the addition of a mild detergent, β-octylglucoside, for maintaining the solubility of the lipid and product lipoprotein during ligation. This synthetic lipo-GFP was deposited onto the surface of mammalian cells, and we are extending this process to lipid anchors with native features of the conserved GPI core. These methods complement genetic approaches that permit only the exchange or deletion of lipids from expressed proteins without controlled modifications to the lipid structures. The semisynthesis of lipidated proteins with native and modified GPI anchors will allow the functional roles of the complex GPI anchor to be examined at a new level of detail.

SUMMARY

Although carbohydrate structures are not conventionally assembled and handled under conditions compatible with protein preparation and synthesis, the described methods to prepare pure glycoproteins have opened a door to meaningful investigations of protein biochemistry and glycobiology. Many of these emerging techniques complement each other, and their forthcoming integration should yield larger and more complex structures for further discovery and for applications to human health.

ACKNOWLEDGMENTS

We thank HC Hang for helpful suggestions. The authors are grateful for generous support from Glaxo Wellcome, Merck, and DuPont. MJG thanks the American Cancer Society for a postdoctoral fellowship. Research from the authors' lab was supported by grants to CRB from the National Institutes of Health (GM58867), the National Science Foundation (CHE-9734439), and the Director, Office of Science, Office of Basic Energy Sciences, Division of Materials and Engineering, and the Office of Energy Biosciences of the U.S. Department of Energy under Contract No. DE-AC03–76SF00098.

The *Annual Review of Biochemistry* is online at http://biochem.annualreviews.org

LITERATURE CITED

1. Drickamer K, Taylor ME. 1998. *Trends Biochem. Sci.* 23:321–24
2. Varki A, Gagneux P. 1999. *Glycobiology* 9:747–55
3. Parekh RB, Dwek RA, Thomas JR, Opdenakker G, Rademacher TW, et al. 1989. *Biochemistry* 28:7644–62
4. Int. Hum. Genome Seq. Consort. 2001. *Nature* 409:860–921
5. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. *Science* 291: 1304–51
6. Rubin GM, Yandell MD, Wortman JR, Miklos GLG, Nelson CR, et al. 2000. *Science* 287:2204–15
7. Fortini M. 2000. *Nature* 406:357–58
8. Moloney DJ, Panin VM, Johnston SH, Chen J, Shao L, et al. 2000. *Nature* 406:369–75
9. Brückner K, Perez L, Clausen H, Cohen S. 2000. *Nature* 406:411–15
10. Munro S, Freeman M. 2000. *Curr. Biol.* 10:813–20
11. Rudd PM, Dwek RA. 1997. *Crit. Rev. Biochem. Mol. Biol.* 32:1–100
12. Herzner H, Reipen T, Schultz M, Kunz H. 2000. *Chem. Rev.* 100:4495–537
13. St. Hilaire PM, Meldal M. 2000. *Angew. Chem. Int. Ed. Engl.* 39:1162–79
14. Koeller KM, Wong CH. 2000. *Chem. Rev.* 1000:4465–93
15. Marcaurelle LA, Bertozzi CR. 1999. *Chem. Eur. J.* 5:1384–90
16. Hang HC, Bertozzi CR. 2001. *Acc. Chem. Res.* 34:727–36
17. Shimizu Y, Shaw S. 1993. *Nature* 366: 630–31
18. Hilkens J, Ligtenberg MJL, Vos HL, Litvinov SV. 1992. *Trends Biochem. Sci.* 17:359–63
19. Strous GJ, Dekker J. 1992. *Crit. Rev. Biochem. Mol. Biol.* 27:57–92
20. Carraway KL, Hull SR. 1991. *Glycobiology* 1:131
21. Hanisch F-G. 2001. *Biol. Chem.* 382: 143–49
22. Comer FI, Hart GW. 2000. *J. Biol. Chem.* 275:29179–82
23. Harris RJ, Spellman MW. 1993. *Glycobiology* 3:219–24
24. Hanisch FG, Chai WG, Rosankiewicz JR, Lawson AM, Stoll MS, Feizi T. 1993. *Eur. J. Biochem.* 217:645–55
25. Varki A, Cummings RD, Esko J, Freeze H, Hart G, Marth J, eds. 1999. *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Lab.
26. Tsuboi S, Fukuda M. 2001. *BioEssays* 23:46–53
27. Fong AM, Erickson HP, Zachariah JP, Poon S, Schamberg NJ, et al. 2000. *J. Biol. Chem.* 275:3781–86
28. Cyster JG, Shotton DM, Williams AF. 1991. *EMBO J.* 10:893–902
29. Li F, Erickson HP, James JA, Moore KL, Cummings RD, McEver RP. 1996. *J. Biol. Chem.* 271:6342–48
30. McMaster TJ, Berry M, Corfield AP, Miles MJ. 1999. *Biophys. J.* 77:533–41
31. Shogren R, Gerken TA, Jentoft N. 1989. *Biochemistry* 28:5525–36
32. Live DH, Williams LJ, Kuduk SD, Schwartz JB, Glunz PW, et al. 1999. *Proc. Natl. Acad. Sci. USA* 96:3489–93
33. Andreotti AH, Kahne D. 1993. *J. Am. Chem. Soc.* 115:3352–53
34. Huang X, Barchi JJ, Lung F-DT, Roller PP, Nara PL, et al. 1997. *Biochemistry* 36:10846–56
35. Liang R, Andreotti AH, Kahne D. 1995. *J. Am. Chem. Soc.* 117:10395–96
36. Gerken TA, Butenhof KJ, Shogren R. 1989. *Biochemistry* 28:5536–43
37. Paulsen H, Kolár C, Stenzel W. 1978. *Chem. Ber.* 111:2358–69
38. Winans KA, King DS, Rao VR, Bertozzi CR. 1999. *Biochemistry* 38:11700–10
39. Schwarz JB, Kuduk SD, Chen X-T, Sames D, Glunz PW, Danishefsky SJ. 1999. *J. Am. Chem. Soc.* 121:2662–73
40. Winterfeld GA, Schmidt RR. 2001. *Angew. Chem. Int. Ed. Engl.* 40:2654–57

41. Mathieux N, Paulsen H, Meldal M, Bock K. 1997. *J. Chem. Soc. Perkin Trans. 1* 16:2359–68
42. Meinjohanns E, Meldal M, Schleyer A, Paulsen H, Bock K. 1996. *J. Chem. Soc. Perkin Trans.* 110: 985–93
43. Glunz PW, Hintermann S, Willams LJ, Schwarz JB, Kuduk SD, et al. 2000. *J. Am. Chem. Soc.* 122:7273–79
44. Kuduk SD, Schwarz JB, Chen X-T, Glunz PW, Sames D, et al. 1998. *J. Am. Chem. Soc.* 120:12474–85
45. Danishefsky SJ, Allen JR. 2000. *Angew. Chem. Int. Ed.* 39:837–63
46. Springer GF. 1984. *Science* 224:1198–206
47. Seeberger PH, Haase W-C. 2000. *Chem. Rev.* 100:4349–93
48. Plante OJ, Palmacci ER, Seeberger PH. 2001. *Science* 291:1523–27
49. Koeller KM, Wong CH. 2000. *Glycobiology* 10:1157–69
50. Vestweber D, Blanks JE. 1999. *Physiol. Rev.* 79:181–213
51. Yang J, Hirata T, Croce K, Merrill-Skoloff G, Tchernychev B, et al. 1999. *J. Exp. Med.* 190:1769–82
52. McEver RP, Cummings RD. 1997. *J. Clin. Invest.* 100:485–92
53. Leppanen A, White SP, Helin J, McEver RP, Cummings RD. 2000. *J. Biol. Chem.* 275:39569–78
54. Leppanen A, Mehta P, Ouyang Y-B, Ju T, Helin J, et al. 1999. *J. Biol. Chem.* 274:24838–48
55. Ouyang Y-B, Lane WS, Moore KL. 1998. *Proc. Natl. Acad. Sci. USA* 95:2896–901
56. Somers WS, Tang J, Shaw GD, Camphausen RT. 2000. *Cell* 103:467–79
57. Koeller KM, Smith ME, Wong CH. 2000. *Bioorg. Med. Chem.* 8:1017–25
58. Koeller KM, Smith ME, Huang R-F, Wong CH. 2000. *J. Am. Chem. Soc.* 122:4241–42
59. Lemieux GA, Bertozzi CR. 1998. *Trends Biotech.* 16:506–13
60. Marcaurelle LA, Bertozzi CR. 2001. *J. Am. Chem. Soc.* 123:1587–95
61. Rodriguez EC, Marcaurelle LA, Bertozzi CR. 1998. *J. Org. Chem.* 63:7134–35
62. Marcaurelle LA, Rodriguez EC, Bertozzi CR. 1998. *Tetrahedron Lett.* 39:8417–20
63. Marcaurelle LA, Bertozzi CR. 1998. *Tetrahedron Lett.* 39:7279–82
64. Rodriguez EC, Winans KA, King DS, Bertozzi CR. 1997. *J. Am. Chem. Soc.* 119:9905–6
65. Marcaurelle LA, Shin Y, Bertozzi CR. 2001. *Org. Lett.* 3:3691–94
66. Lasky LA, Singer MS, Dowbenko D, Imai Y, Henzel WJ, et al. 1992. *Cell* 69:927–38
67. Marcaurelle LA. 2001. *Synthesis of glycoproteins and glycopeptide mimetics.* PhD thesis. Univ. Calif., Berkeley
68. Marcaurelle LA, Pratt MR, Bertozzi CR. 2002. Synthesis of thioether-linked analogs of the 2,3-STF and MECA-79 antigens: mucin type oligosaccharides associated with cancer and inflammation. Submitted
69. Yeh J, Hiraoka N, Petryniak B, Nakayama J, Ellies LG, et al. 2001. *Cell* 105:957–69
70. Streeter PR, Rouse BT, Butcher EC. 1988. *J. Cell Biol.* 107:1853–62
71. Hemmerich S, Butcher EC, Rosen SD. 1994. *J. Exp. Med.* 180:2219–26
72. Hanisch FG, Stadie TR, Deutzmann F, Peter-Katalinic J. 1996. *Eur. J. Biochem.* 236:318–27
73. Hojo H, Aimoto S. 1991. *Bull. Chem. Soc. Jpn.* 64:111–17
74. Marcaurelle LA, Mizoue LS, Wilken J, Oldham L, Kent SBH, et al. 2001. *Chem. Eur. J.* 7:1129–32
75. Bulet P, Hegy G, Lambert J, Van Dorsselaer A, Hoffman JA, Hetru C. 1995. *Biochemistry* 34:7394–400
76. Backes BJ, Ellman JA. 1999. *J. Org. Chem.* 64:2322–30
77. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR. 1999. *J. Am. Chem. Soc.* 121:11684–89

78. Noren CJ, Wang JM, Perler FB. 2000. *Angew. Chem. Int. Ed.* 39:450–66
79. Deleted in proof
80. Muir TW, Sondhi D, Cole PA. 1998. *Proc. Natl. Acad. Sci. USA* 95:6705–10
81. Muir TW. 2001. *Synlett* 6:733–40
82. Macmillan D, Bertozzi CR. 2000. *Tetrahedron* 56:9515–25
82a. Macmillan D, Bertozzi CR. 2002. Glycoprotein semi-synthesis: the use of expressed protein ligation and native chemical ligation for the construction of Glycam-1 glycoforms. In preparation
83. Kornfeld R, Kornfeld S. 1985. *Annu. Rev. Biochem.* 54:631–64
84. Lechner J, Wieland F. 1989. *Annu. Rev. Biochem.* 58:173–94
85. Helenius A, Aebi M. 2001. *Science* 291: 2364–69
86. Gavel Y, von Heijne G. 1990. *Protein Eng.* 3:433–42
87. Hammond C, Braakman I, Helenius A. 1994. *Proc. Natl. Acad. Sci. USA* 91:913–17
88. Huppa JB, Ploegh HL. 1998. *Cell* 92:145–48
89. Cannon KS, Helenius A. 1999. *J. Biol. Chem.* 274:7537–44
90. DeKoster GT, Robertson AD. 1997. *Biochemistry* 36:2323–31
91. Wyss DF, Choi JS, Li J, Knoppers MH, Willis KJ, et al. 1995. *Science* 269:1273–78
92. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, et al. 1996. *Science* 274:209–19
93. Fukuda MN, Sasaki H, Lopez L, Fukuda M. 1989. *Blood* 73:84–89
94. Smith PL, Bousfield GR, Kumar S, Fiete D, Baenziger JU. 1993. *J. Biol. Chem.* 268:795–802
95. Raju TS, Briggs JB, Chamow SM, Winkler ME, Jones AJS. 2001. *Biochemistry* 40:8868–76
96. Weikert S, Papac D, Briggs J, Cowfer D, Tom S, et al. 1999. *Nat. Biotechnol.* 17:1116–21
97. Haneda K, Inazu T, Mizuno M, Iguchi R, Yamamoto K, et al. 1998. *Bioorg. Med. Chem. Lett.* 8:1303–6
98. O'Connor SE, Pohlmann J, Imperiali B, Saskiawan I, Yamamoto K. 2001. *J. Am. Chem. Soc.* 123:6187–88
99. Seifert J, Unverzagt C. 1996. *Tetrahedron Lett.* 37:6527–30
100. Unverzagt C. 1997. *Carbohydr. Res.* 305:423–31
101. Seifert J, Unverzagt C. 1997. *Tetrahedron Lett.* 38:7857–60
102. Unverzagt C. 1996. *Angew. Chem. Int. Ed. Engl.* 35:2350–53
103. Unverzagt C. 1997. *Tetrahedron Lett.* 38:5627–30
104. Guo ZW, Nakahara Y, Ogawa T. 1997. *Bioorg. Med. Chem.* 5:1917–24
105. Guo ZW, Nakahara Y, Ogawa T. 1997. *Angew. Chem. Int. Ed. Engl.* 36: 1464–66
106. Unverzagt C. 1997. *Angew. Chem. Int. Ed. Engl.* 36:1989–92
107. Unverzagt C. 1994. *Angew. Chem. Int. Ed. Engl.* 33:1102–4
108. Likhosherstov LM, Novikova OS, Dervitskaya VA, Kochetkov NK. 1986. *Carbohydr. Res.* 146:C6
109. Meinjohanns E, Meldal M, Paulsen H, Dwek RA, Bock K. 1998. *J. Chem. Soc. Perkin Trans. 1* 3:549–60
110. Live DH, Wang ZG, Iserloh U, Danishefsky SJ. 2001. *Org. Lett.* 3:851–54
111. Wang Z-G, Zhang XF, Visser M, Live D, Zatorski A, et al. 2001. *Angew. Chem. Int. Ed. Engl.* 40:1728–32
112. Wang Z-G, Zhang XF, Live D, Danishefsky SJ. 2000. *Angew. Chem. Int. Ed. Engl.* 39:3652–56
113. Danishefsky SJ, Hu S, Cirillo PF, Eckhardt M, Seeberger PH. 1997. *Chem. Eur. J.* 3:1617–28
114. Macmillan D, Bill RM, Sage KA, Fern D, Flitsch SL. 2001. *Chem. Biol.* 8:133–45
115. Davis NJ, Flitsch SL. 1991. *Tetrahedron Lett.* 32:6793–96
116. Bengtsson M, Broddefalk J, Dahmén J,

Henriksson K, Kihlberg J, et al. 1998. *Glycoconjugate J.* 15:223–31
117. Dawson PE, Kent SBH. 2000. *Annu. Rev. Biochem.* 69:923–60
118. Tolbert TJ, Wong C-H. 2000. *J. Am. Chem. Soc.* 122:5421–28
119. Hojo H, Watabe J, Nakahara Y, Ito Y, Nabeshima K, Toole BP. 2001. *Tetrahedron Lett.* 42:3001–4
120. McConville MJ, Ferguson MAJ. 1993. *Biochem. J.* 294:305–24
121. Ferguson MAJ. 1999. *J. Cell. Sci.* 112: 2799–809
122. Nosjean O, Briolay A, Roux B. 1997. *Biochim. Biophys. Acta* 1331:153–86
123. Simons K, Ikonen E. 1997. *Nature* 387: 569–72
124. Varma R, Mayor S. 1998. *Nature* 394: 798–801
125. Friedrichson T, Kurzchalia TV. 1998. *Nature* 394:802–5
126. Pralle A, Keller P, Florin E-L, Simons K, Hörber JKH. 2000. *J. Cell Biol.* 148: 997–1007
127. Dietrich C, Volovyk ZN, Levi M, Thompson NL, Jacobson K. 2001. *Proc. Natl. Acad. Sci. USA* 98:10642–47
128. Bromley SK, Burack WR, Johnson KG, Somersalo K, Sims TN, et al. 2001. *Annu. Rev. Immunol.* 19:375–96
129. Liao Z, Cimakasky LM, Hampton R, Nguyen DH, Hildreth JEK. 2001. *AIDS Res. Hum. Retroviruses* 17:1009–19
130. Dykstra ML, Longneckker R, Pierce SK. 2001. *Immunity* 14:57–67
131. Kaneko K, Vey M, Scott M, Pilkuhn S, Cohen FE, Prusiner SB. 1997. *Proc. Natl. Acad. Sci. USA* 94:2333–38
132. Campbell AS, Fraser-Reid B. 1995. *J. Am. Chem. Soc.* 117:10387–88
133. Mayer TG, Schmidt RR. 1999. *Eur. J. Org. Chem.* 1153–65
134. Baeschlin DK, Chaperon AR, Charbonneau V, Green LG, Ley SV, et al. 1998. *Angew. Chem. Int. Ed. Engl.* 37:3423–28
135. Schofield L, McConville MJ, Hansen D, Campbell AS, Fraser-Reid B, et al. 1999. *Science* 283:225–29
136. Frick W, Bauer A, Bauer J, Wied S, Müller G. 1998. *Biochemistry* 37: 13421–36
137. Medof ME, Nagarajan S, Tykocinski ML. 1996. *FASEB J.* 10:574–86
138. Medof ME, Kinoshita T, Nussenzweig V. 1984. *J. Exp. Med.* 160:1558–78
139. Zhang F, Schmidt WG, Hou Y, Williams AF, Jacobson K. 1992. *Proc. Natl. Acad. Sci. USA* 89:5231–35
140. Huang J-H, Getty RR, Chisari FV, Fowler P, Greenspan NS, Tykocinski ML. 1994. *Immunity* 1:607–13
141. Bader B, Kuhn K, Owen DJ, Waldmann H, Wittinghofer A, Kuhlmann J. 2000. *Nature* 403:223–26