Chemoselective Approaches to Glycoprotein Assembly

HOWARD C. HANG[†] AND
CAROLYN R. BERTOZZI*

Departments of Chemistry and Molecular and Cell Biology, University of California, Center for Advanced Materials, Materials Sciences Division, Lawrence Berkeley National Laboratory, and Howard Hughes Medical Institute, Berkeley, California 94720

Received February 23, 2001

ABSTRACT

Oligosaccharides on proteins and lipids play central roles in human health and disease. The molecular analysis of glycoconjugate function has benefited tremendously from new methods for their chemical synthesis, which provides homogeneous material not attainable from biosynthetic systems. Still, glycoconjugate synthesis requires the manipulation of multiple stereocenters and protecting groups and remains the domain of a few expert laboratories around the world. This Account summarizes chemoselective approaches for assembling homogeneous glycoconjugates that attempt to reduce the barriers to their synthesis. The objective of these methods is to make glycoconjugate synthesis accessible to a broader community, thereby accelerating progress in glycobiology.

Introduction

Oligosaccharides on proteins provide a mechanism to modulate their behavior in a complex multicellular environment.[1] The glycosylation of proteins directs protein folding, endows conformational stability, bestows resistance to proteolytic degradation, regulates the protein's serum half-life, and provides unique epitopes for molecular recognition. The effects of protein glycosylation are manifested in many physiological events such as cell–cell communication and cell growth and differentiation, as well as viral infection. Interestingly, aberrant glycosylation of proteins or lipids has often been correlated with specific disease states. This has motivated the development of therapeutic agents designed to interfere with carbohydrate biosynthesis or molecular recognition[2] and the development of carbohydrate-based anticancer vaccines.[3]

Despite the significance of glycoconjugates in physiology and medicine, there remain substantial gaps in our appreciation of glycobiology at the molecular level, especially when compared to our knowledge of protein and nucleic acid-mediated processes. This discrepancy in knowledge stems from the complex nature of oligosaccharides as branched biopolymers connected through glycosidic linkages of variable stereochemistry. To complicate matters even further, the microheterogeneity of oligosaccharide structures on a protein or lipid yields an ensemble of glycoconjugates differing only in their glycan structure.[4] While the structural diversity of glycoconjugates is ideal for encoding information from cells, these unwieldy properties of glycoconjugates have challenged existing analytical techniques for separation and purification, restricting access to sufficient quantities of homogeneous material for structural and functional analysis.

Unlike the biosynthesis of nucleic acids and proteins, oligosaccharide biosynthesis is neither template-driven nor under direct transcriptional control, precluding direct access to homogeneous material by recombinant DNA technology. The overexpression of recombinant glycoproteins has been the primary means of obtaining material for fundamental research and pharmaceutical therapeutics. However, glycoproteins obtained from recombinant overexpression are heterogeneous and organism-dependent, complicating the analysis of their structural and biological properties.[5] Genetic engineering of host organisms has allowed the production of glycoproteins with more defined glycan structures by addition or deletion of glycosyltransferases.[6] Unfortunately, these genetically engineered systems still afford heterogeneous glycoproteins.

In principle, total chemical synthesis could provide well-defined and homogeneous glycoproteins. While tremendous advances have been made in oligosaccharide[7] and glycopeptide[8] syntheses, they remain labor-intensive, requiring extensive protecting group manipulations for regio- and stereoselective glycosylations, and are still far from routine. The technical challenges are further exacerbated by the protecting group demands of solid-phase synthesis of the peptide component (SPPS). As a testament to the power of organic synthesis, a few groups have completed the total syntheses of some impressive *N*-linked[9] and *O*-linked[10] glycopeptides (Figure 1 shows prototypical *N*-linked and *O*-linked oligosaccharides). Enzymatic synthesis presents a means of circumventing the protecting group demands of chemical methods for the construction of large glycoproteins.[11] At present, costly nucleotide–monosaccharide donors and the lack of a complete repertoire of glycosidases and glycosyltransferases available commercially still limit this strategy.[12] Nonetheless, the combination of synthetic and enzymatic techniques has produced glycoprotein fragments of considerable complexity.[13] Thus, the prospect of glycoprotein total synthesis remains a possibility only in the laboratories of a handful of experts.

To facilitate glycobiology research within laboratories lacking such expertise, new methods are required that involve general and even automated procedures. In

Howard C. Hang was born in Saigon, Vietnam, in 1976. He received his B.S. degree in chemistry from the University of California, Santa Cruz, where he worked with Prof. Joseph P. Konopelski. He is currently pursuing a Ph.D. degree in chemistry at the University of California, Berkeley.

Carolyn R. Bertozzi was born in Boston, Massachusetts, in 1966. She received her undergraduate degree in chemistry from Harvard University and her Ph.D. in chemistry from the University of California, Berkeley, working with Prof. Mark Bernarski. After a postdoctoral stint with Prof. Steven Rosen at the University of California, San Francisco, she joined the faculty of the University of California, Berkeley, in 1996. Her research interests focus on applications of chemistry to biological problems.

* Corresponding author. Tel.: (510) 643-1682. Fax: (510) 643-2628. E-mail: bertozzi@cchem.berkeley.edu.
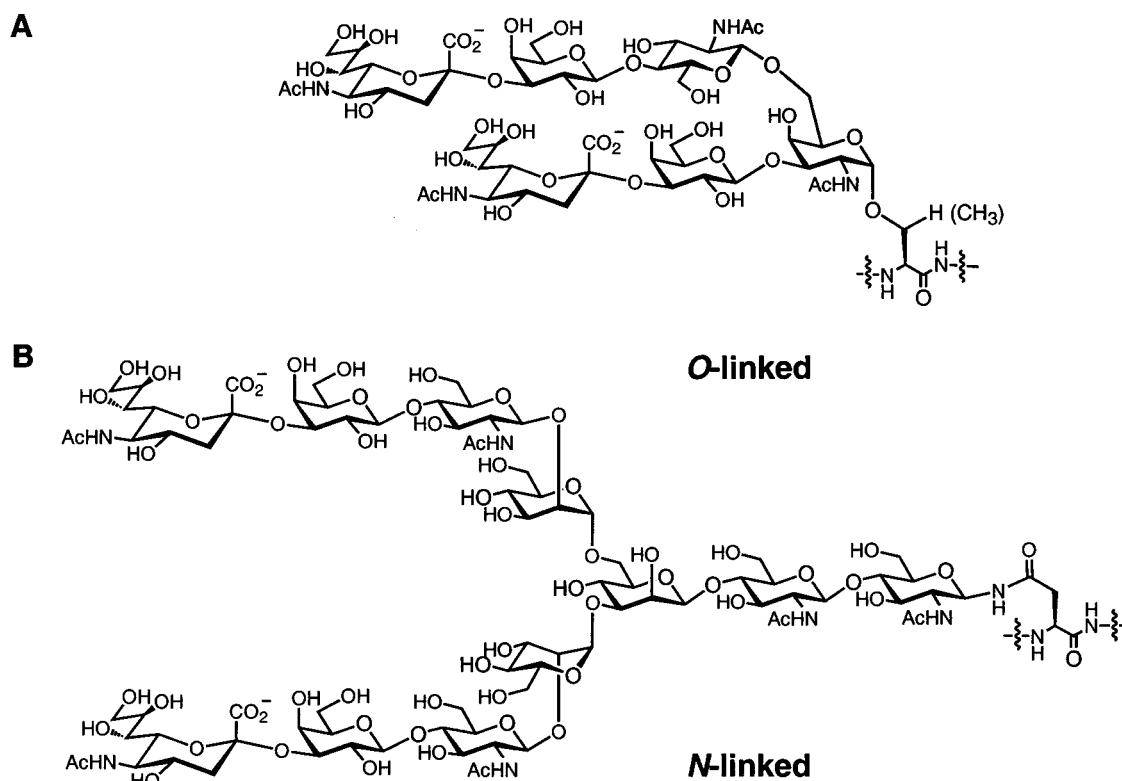† Department of Chemistry, University of California, Berkeley.

A

B

O-linked

N-linked

FIGURE 1. Examples of (A) *O*-linked and (B) *N*-linked glycans found on mammalian glycoproteins.

contrast to nucleic acid and peptide synthesis, which have been automated for decades, automation of oligosaccharide synthesis has only recently been addressed. Breakthrough technologies such as "programmable one-pot oligosaccharide synthesis" [14] and automated solid-phase oligosaccharide synthesis[15] promise to increase the availability of large oligosaccharides for structural and functional evaluation. The integration of these structures, soon to be available to any biologist or chemist, into full-length glycoproteins remains the next major synthetic challenge. Herein, we review applications of the principle of chemoselective ligation for the convergent assembly of glycopeptides and glycoproteins. Other approaches to the synthesis of glycopeptide mimetics[16] and neoglycoconjugates[17] have been reviewed elsewhere and will not be discussed.

Chemoselective Ligation Reactions

Chemoselective ligation reactions were first described by protein chemists as the coupling of two mutually and uniquely reactive functional groups in an aqueous environment. As the name of the technique implies, these uniquely reactive functional groups are selective for each other and also tolerate a diverse array of other functionality, which renders the use of protecting groups unnecessary.[18] Chemoselective ligation reactions thus offer advantages similar to those of enzymatic reactions, with the potential of a much broader range of substrates for use as coupling partners. A survey of known chemical reactions yields several that fulfill the criteria for a chemoselective ligation reaction (Figure 2). The absence of aldehydes and ketones on the side chains of the naturally

occurring amino acids renders them privileged electrophiles that can react with a variety of nucleophiles. For example, aldehydes and ketones react with hydrazides to form *N*-acyl hydrazones (Figure 2A), with aminooxy groups to yield oximes (Figure 2B), and with thiosemicarbazides to give the corresponding thiosemicarbazones (Figure 2C). The superior nucleophilicity of sulfhydryl groups has been exploited for selective alkylation with α-halocarbonyl electrophiles, even in the presence of other proteinaceous nucleophiles (Figure 2D). Such alkylations of cysteine side chains have been widely used for site-specific protein labeling. In addition, peptide C-terminal thiocarboxylates can be selectively ligated in the presence of cysteine residues due the pK_a difference of the thiocarboxylic acid (p$K_a \sim 3$) and sulfhydryl group (p$K_a \sim 8$).[19]

Aldehydes can also react with β-amino thiols or alcohols to form thiazolidines and oxazolidines, respectively.[20] Tam and co-workers have exploited this reactivity to ligate C-terminal glycol aldehyde ester peptides with peptides bearing N-terminal cysteine or serine/threonine residues; the intermediate thiazolidines or oxazolidines then rearrange to generate proteins with pseudoproline linkages (Figure 2E).[21] In addition, this group has developed a Pictet–Spengler ligation, which involves the reaction of a C-terminal glycine aldehyde peptide with an N-terminal tryptophan (Trp) peptide to generate an unnatural tetrahydro-β-carboline linkage (Figure 2F).[22] These reactions have found widespread use in the convergent assembly of peptide fragments to make larger proteins bearing unnatural internal linkages at the ligation site. While such orthogonal ligation strategies have enabled the assembly

A

B

C

D

X = Cl, Br, I

E

X = O, S

F

G

H

X = S, Se

I

J

X = O, S
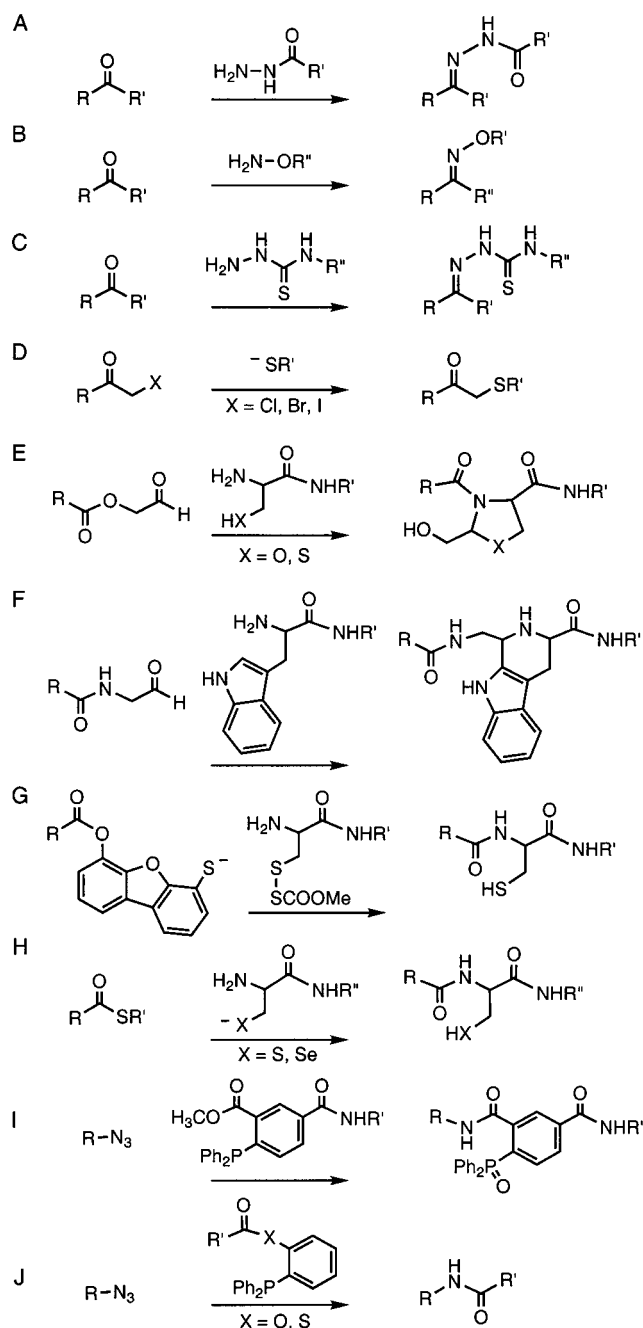
FIGURE 2. Chemoselective ligation reactions used in the convergent assembly of biopolymers and their mimetics, and in the modification of biopolymers and cells. The products of the reactions are (A) *N*-acyl hydrazones, (B) oxime ethers, (C) thiosemicarbazones, (D) thioethers, (E) pseudoproline linkages, (F) tetrahydro-β-carbolines, (G) native amide bonds after reduction of disulfide bonds, (H) native amide bonds, (I) amide phosphine oxide, and (J) simple amide bonds.

of larger peptides or proteins, the introduction of non-native linkages into these biomolecules could potentially perturb their structure and function.

The desire to generate proteins with native amide bonds by a chemoselective ligation reaction inspired Kemp and co-workers to develop the "thiol capture" ligation of C-terminal 4-mercaptodibenzofuran peptides with N-terminal Cys-Scm (Scm = SCOOMe) peptides (Figure 2G).[23] This methodology allowed the coupling of

small peptides that formed native amide bonds at the ligation site following reduction of the disulfide bond. The subsequent development of "native chemical ligation" simultaneously by Kent and Tam and their co-workers provided a more general approach for coupling unprotected peptides bearing C-terminal thioesters with unprotected peptides bearing N-terminal Cys to yield native amide bonds at the ligation site (Figure 2H, vide infra).[24] More recently, Raines and van der Donk and their co-workers have extended native chemical ligation to include the use of selenocysteine (Figure 2H).[25] The increased nucleophilicity of RSe$^-$ versus RS$^-$ and differences in the pK_a values of selenols versus thiols allowed selenocysteine ligations to occur at pH 5 rather than pH 8. These differences in reactivity should also provide chemoselectivity for the alkylation of selenols versus thiols. The introduction of selenium into proteins also provides a ^{77}Se probe for NMR spectroscopy as well as a heavy atom for phasing in X-ray structure determination of proteins.

Synthesis of Glycopeptide Mimetics Using Chemoselective Ligation Reactions

Chemoselective ligation reactions present an attractive means for assembling glycopeptide mimics by the convergent coupling of unprotected oligosaccharide analogues with cognate functionalized peptides. Several groups have exploited the unique reactivity of cysteine residues to chemoselectively ligate α-haloacetamido (Figure 3A)[26] or bromoethyl glycosides (Figure 3B),[27] generating glycopeptide mimetics with non-native sugar—peptide linkages. An alternative approach reported by Boons[28] and Davis[29] utilized the sulfhydryl group on the side chain of cysteine residues to form a disulfide bond to the carbohydrate fragment (Figure 3C). If no other thiols or disulfide bonds are present in the protein, these strategies permit the site-specific attachment of glycans.

To create orthogonal coupling partners that do not cross-react with native functionality present on proteins, we have utilized a ketone electrophile on an unnatural amino acid side chain that reacts with oligosaccharides bearing aminooxy, hydrazide, or thiosemicarbazide nucleophiles at the anomeric position. Using these chemoselective coupling partners, we have assembled *O*-linked (Figure 4A) and *N*-linked (Figure 4B) glycoprotein mimetics from unprotected oligosaccharides and peptides.[30] The non-native sugar—peptide linkages embodied in these structures might negatively impact biologically relevant structural elements, and therefore functional activity. In one functional study, however, an oxime-linked analogue of the antimicrobial peptide drosocin demonstrated activity comparable to that of native glycosylated drosocin.[31] In a similar approach, with the reversal of electrophile and nucleophile, Mutter and co-workers stereoselectively coupled carbohydrates with free-reducing ends onto peptides with *N*-methyl aminooxy side chains to produce β-*N*-linked glycopeptide mimics (Figure 4C).[32] The
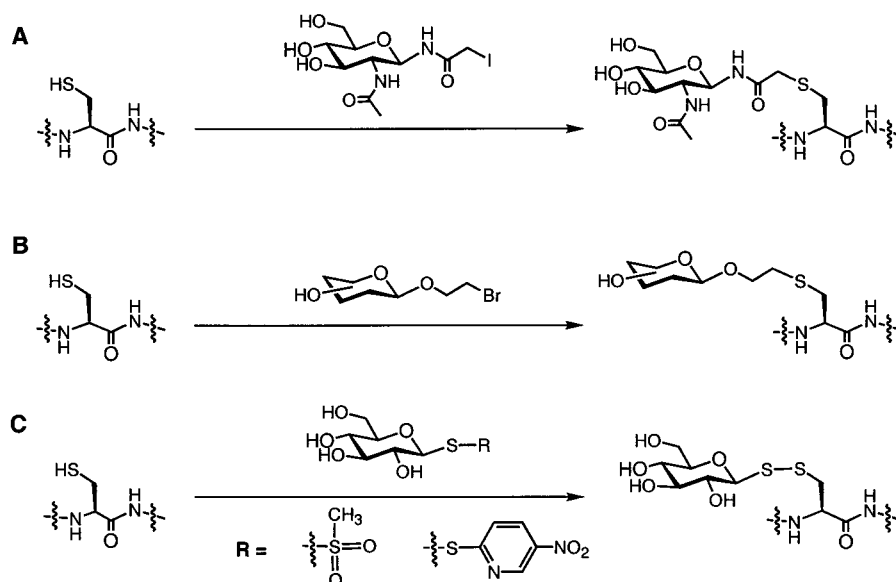
FIGURE 3. Chemoselective modification of cysteine side chains in the assembly of glycopeptide mimetics. (A) Alkylation of cysteine with α-haloacetamido sugars. (B) Alkylation of cysteine with bromoethyl glycosides. (C) Disulfide bond formation with thioglycosides.
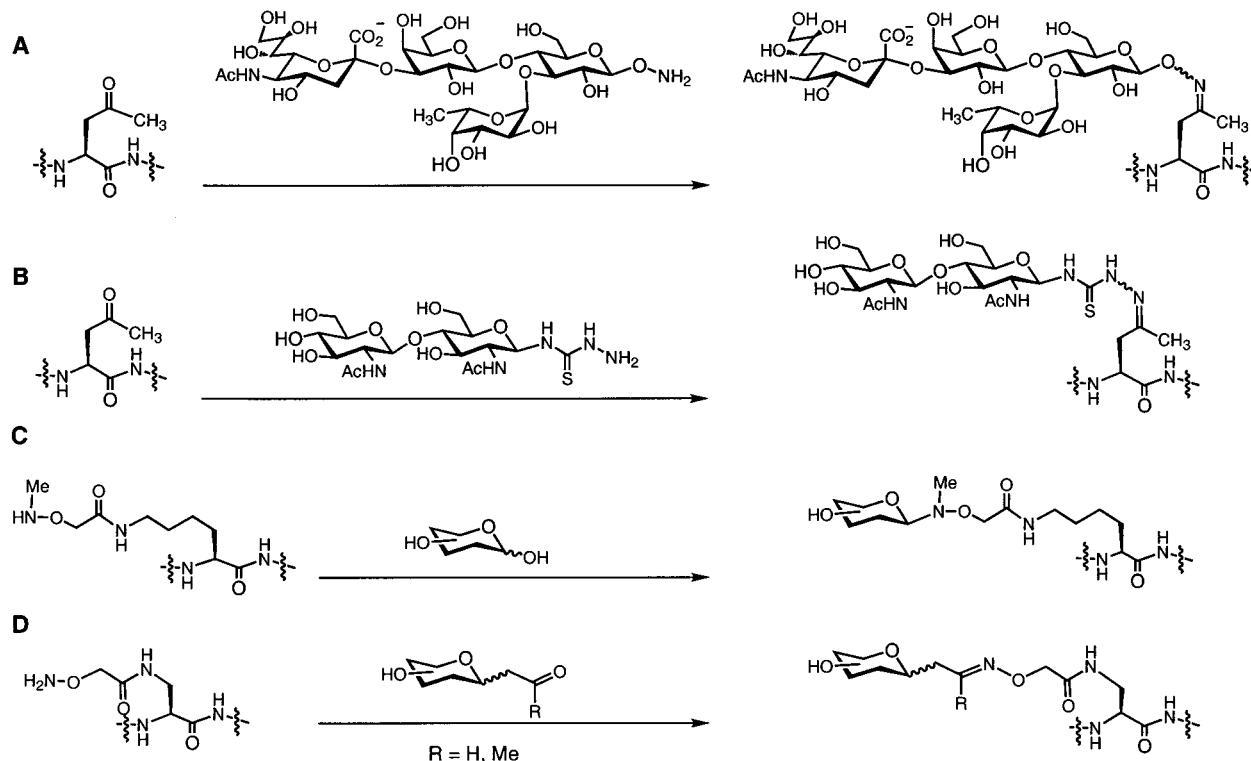

FIGURE 4. Assembly of glycopeptide mimetics by chemoselective reaction of aldehydes and ketones. (A) Oxime-linked analogue of an *O*-linked glycopeptide bearing the tetrasaccharide epitope sialyl Lewis x. (B) Thiosemicarbazone-linked analogue of an *N*-linked glycopeptide bearing the core chitobiose motif. (C) Aminooxy glycoside-linked glycopeptide analogue with an intact pyranose generated from a free reducing sugar. (D) Oxime-linked glycopeptide analogue with a *C*-glycosyl sugar.

aminooxy functionality has also been installed at the N-terminus of a peptide for coupling with free reducing oligosaccharides.[33] More recently, *C*-glycosides bearing ketones or aldehydes have been chemoselectively ligated to peptides adorned with aminooxy groups on amino acid side chains (Figure 4D).[34] The introduction of orthogonal functionality such as ketone or aminooxy groups was accomplished by SPPS for these small peptides but would

require unnatural amino acid mutagenesis[35] or expressed protein ligation[36] (vide infra) for larger proteins.

Synthetically difficult glycosidic linkages within the oligosaccharide portion of glycoproteins might be ideal sites for chemoselective linkage substitution. We have applied this strategy in the chemoselective synthesis of *O*-linked glycopeptide mimetics that retain the native sugar–peptide linkage[37] but instead possess unnatural
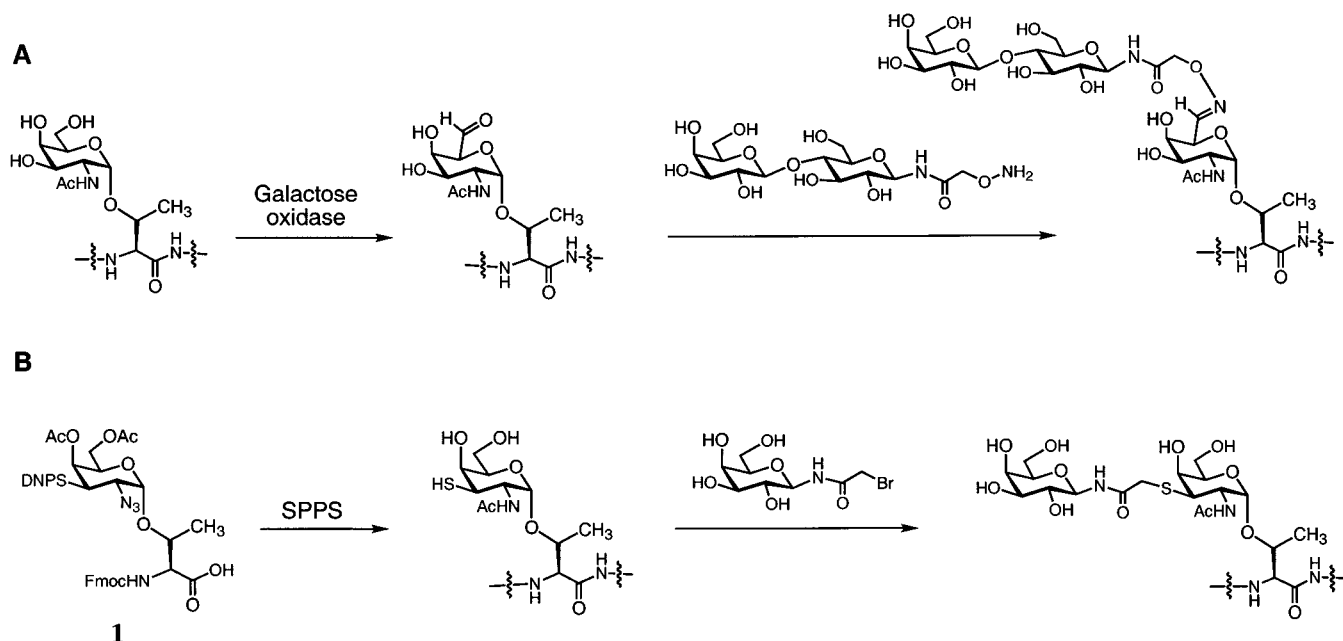
FIGURE 5. Syntheses of glycopeptide mimetics that substitute glycosidic linkages with chemoselective linkages. (A) Aldehydes can be generated at the C-6 position of GalNAc using the enzyme galactose oxidase. Coupling with aminooxy sugars produces oxime-linked products. (B) Introduction of **1** into SPPS, followed by reductive acetylation of the azide and deprotection, produces glycopeptides with pendant 3-thioGalNAc residues. Selective alkylation of the thiol extends the glycan with a chemoselective linkage rather than a glycosidic linkage at that position. DNP = dinitrophenyl.

bonds at the branch points (C-6 and C-3) of the core GalNAc residue (Figure 5). Using protected GalNAc-Thr, now commercially available, as a building block in SPPS, simple glycopeptides bearing single GalNAc residues can be readily generated. Selective oxidation of these glycopeptides using commercially available galactose oxidase converted the C-6 hydroxyl group of GalNAc to the corresponding aldehyde. Chemoselective ligation with aminoxy-functionalized oligosaccharides produced higher order glycans through an unnatural oxime linkage (Figure 5A).[38] To introduce an orthogonal branching point at C-3 of GalNAc-Thr, we have recently designed and synthesized glycosylated amino acid **1,** which contains a protected sulfhydryl group in place of the C-3 hydroxyl group of GalNAc (Figure 5B).[39] Following incorporation of **1** into a glycopeptide by SPPS, the deprotected sulfhydryl group was selectively elaborated with α-haloacetamido sugars.

These two orthogonal ligation reactions could potentially be used for tamdem coupling of aminooxy and α-haloacetamido sugars to a bis-functionalized core Gal-NAc residue bearing a C-6 aldehyde and C-3 thiol, allowing the one-pot assembly of complex glycopeptide mimics. The process of constructing these glycosidic linkage-substituted mimetics is considerably more facile than the process required for the corresponding native glycopeptides. The price one pays for synthetic facility is the presence of unnatural linkages that may perturb the three-dimensional structure of the oligosaccharide, and perhaps its molecular recognition activity. How these unnatural glycans compare to their native counterparts at the structural level is a topic of current investigation.

Synthesis of Glycoproteins Using Native Chemical Ligation

Even without bound oligosaccharides, the linear, stepwise synthesis of proteins becomes intractable when their size exceeds 50–60 residues. The majority of proteins and glycoproteins of biological interest are considerably larger. The transition from glycopeptides to glycoproteins has been addressed by using commercially available glycosidases and proteases to digest heterogeneous glycoproteins, which can then be reassembled with peptidyl ligases and glycosyltransferases to regenerate homogeneous glycoproteins.[40] Unfortunately, this approach is not general due to the limited availability and restricted substrate specificity of the required enzymes.

Recent developments in the field of protein chemistry have now made the chemical synthesis of larger proteins achievable. In particular, the technique of "native chemical ligation" developed by the laboratories of Kent and Tam[24] permits the convergent coupling of unprotected peptide fragments to generate proteins over 100 residues in length. One peptide fragment possesses a C-terminal thioester, and the other fragment bears an N-terminal cysteine residue; their selective and reversible transthioesterification reaction is followed by an irreversible rearrangement that forms an amide bond at the site of ligation (Figure 2H). The requirement of a Cys residue at the ligation site limits the position within a protein for retrosynthetic disconnection. If the protein of interest does not have any Cys residues, a mutation to Cys must be made to allow its synthesis by native chemical ligation. The development of conditions for desulfurization of
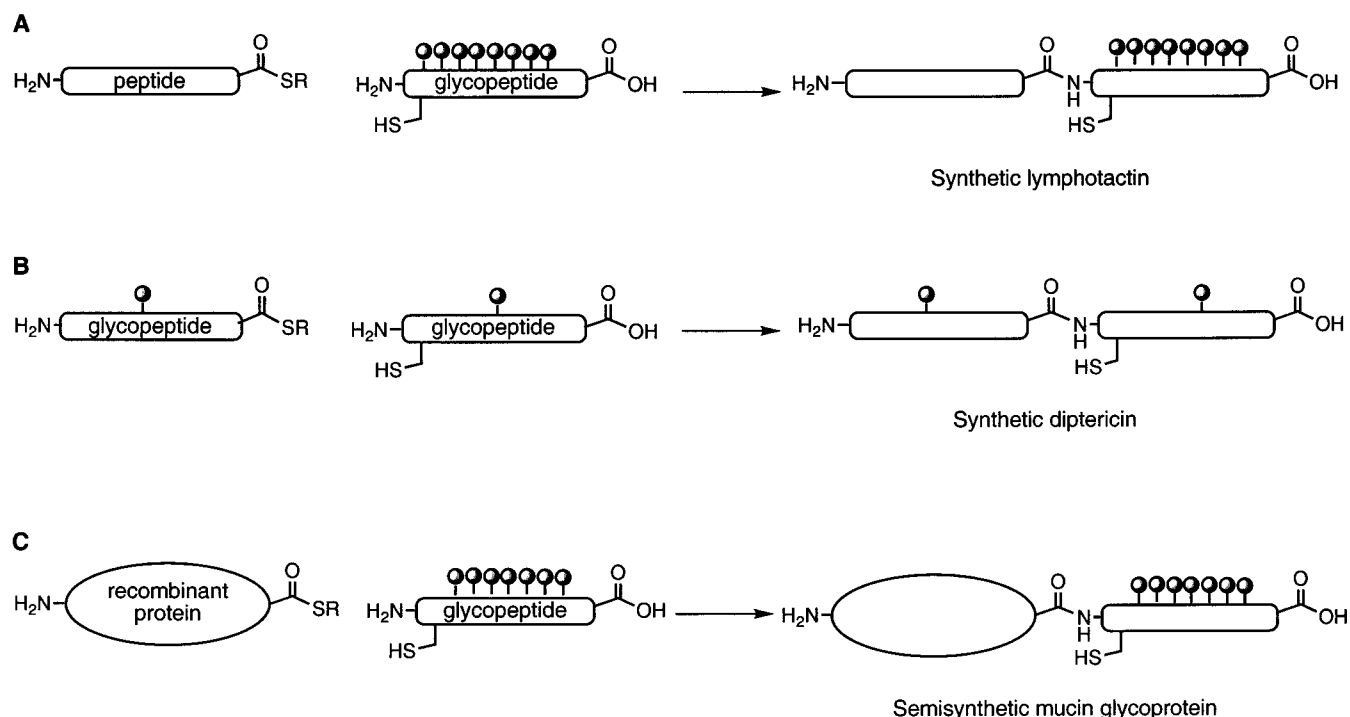
FIGURE 6. Applications of native chemical ligation and expressed protein ligation to the assembly of glycoproteins from glycopeptide fragments. (A) Synthesis of the glycosylated chemokine lymphotactin by native chemical ligation. The glycopeptide fragment bearing a mucin-like domain was generated by Fmoc-based SPPS, and the peptide−$^\alpha$thioester fragment was generated using Boc-based methods. (B) Synthesis of the antimicrobial glycoprotein diptericin by native chemical ligation. Both glycopeptide fragments were generated using Fmoc-based methods. (C) Semisynthesis of a mucin glycoprotein by expressed protein ligation. A recombinant protein thioester fragment expressed as an intein fusion was ligated to a synthetic mucin glycopeptide.

proteins has allowed the transformation of Cys into Ala after ligation, which may circumvent any deleterious effects of introducing Cys into a protein.[41] However, this strategy cannot be applied in the presence of additional Cys or Met residues elsewhere in the protein.

Just as it transformed the field of protein synthesis, so has the native chemical ligation technique impacted glycoprotein research. Indeed, several full-length, functional glycoproteins have now been assembled from synthetic glycopeptide and peptide fragments. We have applied this methodology in the total chemical synthesis of homogeneously glycosylated lymphotactin (Lptn) in order to study the effect of glycosylation on the structure of an unusual C-terminal mucin domain (Figure 6A).[42]

The synthesis of C-terminal peptide−$^\alpha$thioesters was initially developed using Boc-based SPPS; the basic reagents of Fmoc-based SPPS were incompatible with the C-terminal thioester linkers used to anchor the peptide to solid support. Due to acid-lability of *O*-linked glycosidic linkages, Fmoc-based SPPS is generally used for the synthesis of *O*-linked glycopeptides. Thus, glycosylation and thioesterification may be in conflict with each other. In the case of Lptn, the mucin domain resides in the C-terminal rather than the N-terminal fragment. Consequently, the N-terminal peptide−$^\alpha$thioester fragment could be generated in the traditional Boc-based manner, while the glycosylated C-terminal fragment was synthesized using Fmoc-based SPPS. However, some glycoproteins possess glycans near the N-terminus, and their synthesis

using native chemical ligation would require access to glycopeptide−$^\alpha$thioesters. To address this problem, we have employed Ellman's modification of Kenner's "safety catch-linker"[43] to generate C-terminal thioesters using Fmoc-based SPPS, which is compatible with the presence of glycosidic linkages. This breakthrough enabled the total chemical synthesis of biologically active diptericin, an antimicrobial glycoprotein with *O*-linked glycosylation sites throughout the protein (Figure 6B).[44]

Synthesis of Large Glycoproteins by Expressed Protein Ligation

The application of native chemical ligation to glycoprotein engineering has allowed access to significant amounts of homogeneous glycoproteins of moderate size (around 100 amino acids in length), but many larger glycoproteins in nature are still not accessible. The development of "expressed protein ligation" has allowed the union of recombinant protein expression with peptides generated by SPPS.[36,45] Expressed protein ligation utilizes commercially available protein expression systems that produce recombinant proteins with C-terminal protein−$^\alpha$thioesters as a result of intein-mediated protein splicing.[46] Recombinant proteins with C-terminal thioesters can be ligated to other proteins or peptides with N-terminal cysteine residues. This technique has revolutionized protein engineering and presents an opportunity for generating large homogeneous glycoproteins through semisynthesis. Recently, Wong and co-workers demonstrated that modified N-
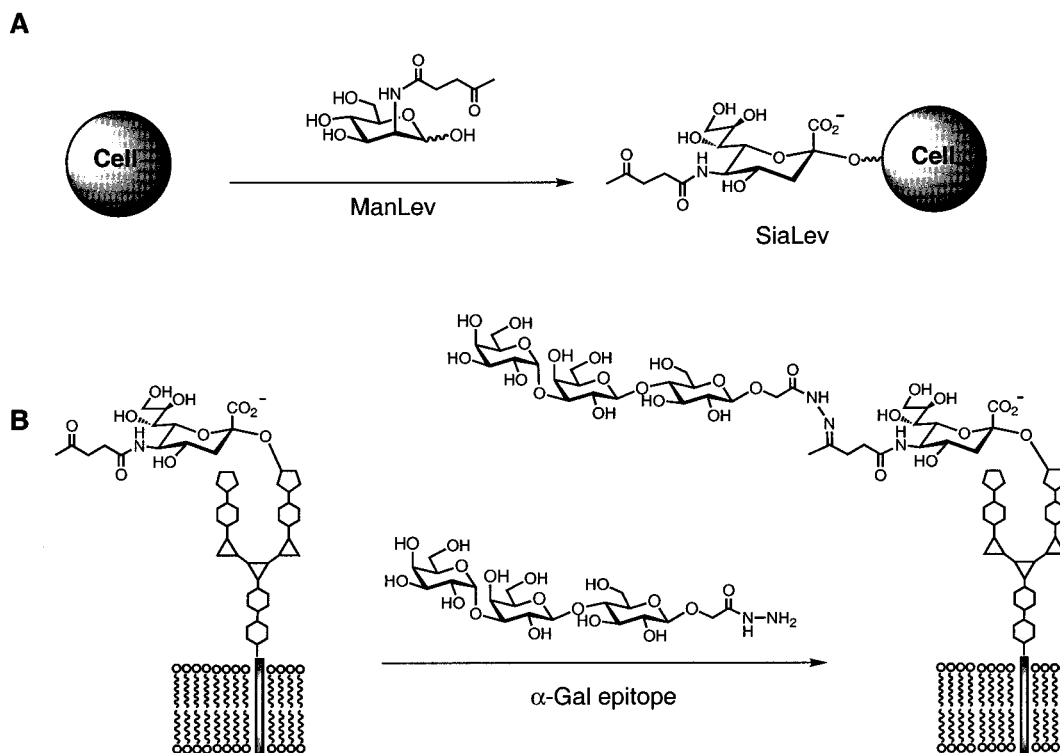
FIGURE 7. Cell surface glycoform remodeling by chemoselective ligation. (A) Metabolic delivery of orthogonal ketone groups to cell surface glycoconjugates using the sialic acid biosynthetic pathway. (B) Elaboration of the ketone with chemically defined oligosaccharides creates a new glycosylation pattern on the cell surface. The α-Gal epitope (Galα1,3Galβ1,4Glc) is a xenotransplant antigen that elicits a vigorous immune response in humans, resulting in tissue rejection.

terminal cysteine peptides could be ligated to recombinantly expressed maltose binding protein with a C-terminal thioester.[47] Our group has exploited expressed protein ligation in the semisynthesis of homogeneous *O*-linked mucin-type glycoproteins (Figure 6C).[48] These new developments in glycoprotein synthesis and semisynthesis will undoubtedly dissolve the barriers for obtaining sufficient quantities of homogeneous glycoproteins for structural and functional studies.

Metabolic Engineering of Cell Surface Oligosaccharides

The increasing molecular complexity from prokaryotes to higher eukaryotes has required cells to evolve new mechanisms to thrive in a multicellular environment. As complex and heterogeneous biopolymers, oligosaccharides on secreted and cell surface glycoconjugates are well-suited to present and receive enormous amounts of information from the environment. However, the complex and heterogeneous nature of glycoconjugates presents tremendous challenges for scientists who aim to understand the precise mechanisms by which cells communicate with each other and the extracellular matrix. In an effort to generate cell surfaces with chemically defined glycoforms, our group has exploited the unnatural substrate tolerance exhibited by some carbohydrate biosynthetic enzymes. Unnatural monosaccharides bearing pendant orthogonal functional groups can be metabolized by cells and incorporated into cell surface glycoconjugates,

where the functional groups can then engage in chemoselective ligation reactions (Figure 7).[49] For example, *N*-levulinoylmannosamine (ManLev), an unnatural variant of *N*-acetylmannosamine (ManNAc), is converted to the corresponding unnatural sialic acid (*N*-levulinoyl sialic acid or SiaLev) on the surface of human cells (Figure 7A).[50] The ketone group within the levulinoyl side chain serves as a unique chemical handle for reaction with complementary nucleophiles. This metabolic cell surface engineering technique enabled the remodeling of cells with alternative glycoforms,[51] such as the α-Gal xenograph transplant antigen (Figure 7B).[52] The technology has also been applied to the selective targeting of toxins[50] and diagnostic agents[53] to tumor cells, facilitated retroviral gene transfer,[54] modulated cell surface immunoreactivity,[55] and allowed selections of mammalian cells to uncover defects in carbohydrate biosynthetic pathways relevant to human disease.[56] While the sialic acid biosynthetic pathway tolerates a number of unnatural ManNAc analogues, we have shown that the amide functionality of ManNAc is essential for metabolic transformation and that there are limitations to the length and steric bulk of unnatural substituents tolerated at the *N*-acyl position.[57]

Our interest in performing chemical reactions in a cellular environment has led to the development of a new chemoselective ligation reaction based on the classic Staudinger reaction (Figures 2I and 8).[58] This "Staudinger ligation" proceeds by reaction of a modified triarylphosphine with an alkyl (or aryl) azide to form an intermediate aza-ylide, which undergoes an intramolecular cyclization

A **Staudinger Ligation**

B

**C** **Traceless Staudinger Ligation**
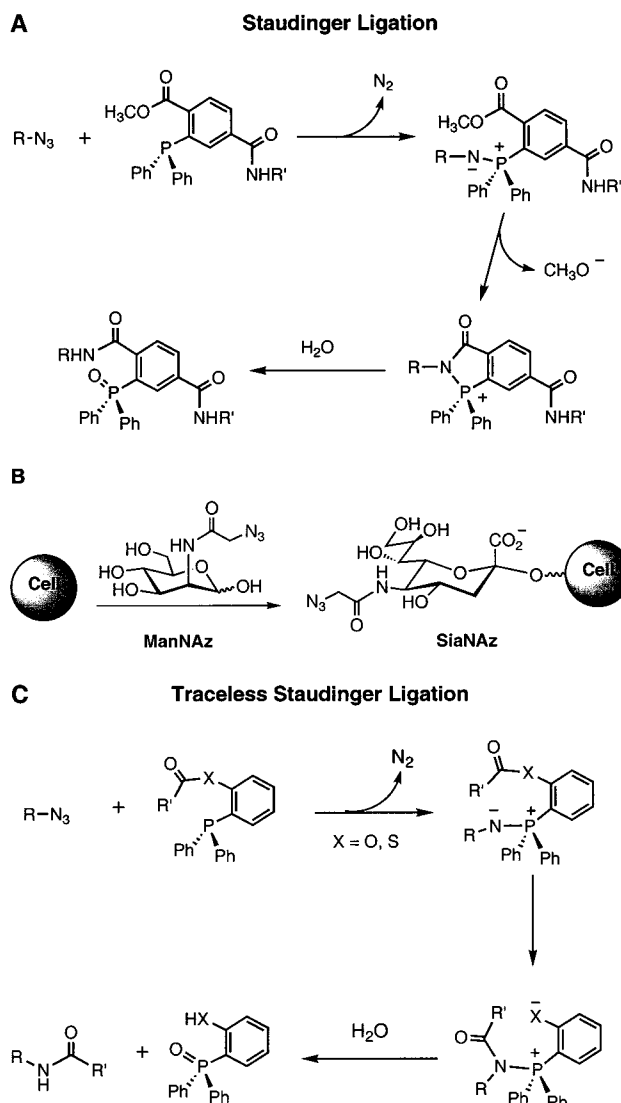
FIGURE 8. Cell surface remodeling using a new chemoselective ligation reaction called the Staudinger ligation. (A) The Staudinger ligation proceeds by the reaction of modified triphenylphosphine analogues with azides to form initially an aza-ylide intermediate. This reacts rapidly in an intramolecular sense to produce, after hydrolysis, an amide-linked product. The reaction is orthogonal to the complex functionality of the cell surface. (B) Azides are delivered to cell surface glycoproteins via the sialic acid biosynthetic pathway. (C) A "traceless Staudinger ligation" produces a simple amide bond between the two coupling partners. Applications to native chemical ligation of peptide fragments at sites other than cysteine may be possible.

onto the adjacent methyl ester to yield a new amide bond and oxidized phosphine upon hydrolysis (Figure 8A). Using the Staudinger ligation, we have recently demonstrated that *N*-azidoacetylmannosamine (ManNAz) was metabolically converted to the corresponding sialic acid and incorporated into cell surface glycoconjugates (Figure 8B); the azide was available on the cell surface for Staudinger ligation with exogenous reagents. The presentation of ketones and azides on cell surfaces provides two orthogonal means of chemically remodeling glycoproteins within their native biological context.
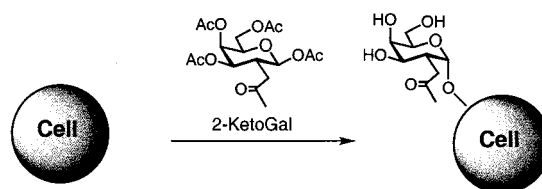
FIGURE 9. Metabolic incorporation of 2-KetoGal, an isosteric analogue of GalNAc, into *O*-linked glycoproteins on cell surfaces.
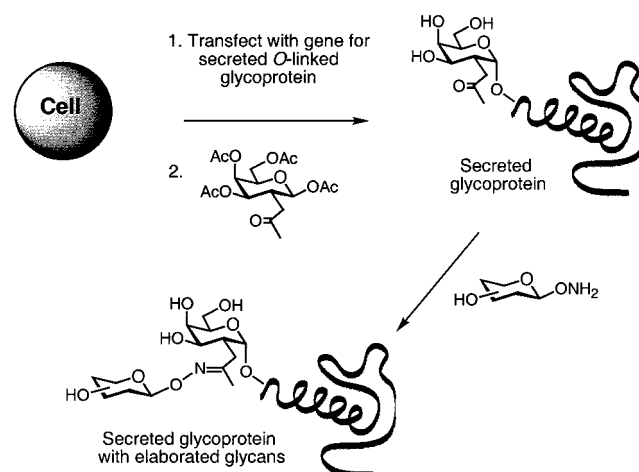
FIGURE 10. Application of 2-KetoGal metabolism to "posttranslational glycoprotein engineering". Overexpression of *O*-linked glycoproteins in the presence of 2-KetoGal is expected to produce material bearing the unnatural sugar at the core position (i.e., proximal to the polypeptide). The ketone group can be exploited for chemoselective elaboration with chemically defined oligosaccharides, generating homogeneous complex glycans on a recombinant protein.

The adduct formed by the Staudinger ligation includes a rather abiotic triarylphosphine oxide which we sought to eliminate for future biological applications. Thus, a second-generation variant of the Staudinger ligation was developed to generate a simple amide bond from azide and phosphine reagents. Termed the "traceless" Staudinger ligation (Figures 2J and 8C), the reaction utilizes phosphines bearing a transferable acyl group. Reaction with azides generates, after rearrangement of the intermediate aza-ylide and hydrolysis, the amide-linked product and a liberated phosphine oxide. Preliminary results from our laboratory[59] and Raines's laboratory[60] using a variety of acyl phosphines appear promising for application of the traceless Staudinger ligation to peptide couplings as well as cell surface modification.

The remarkable promiscuity exhibited by the enzymes of sialic acid biosynthesis prompted us to explore the unnatural substrate tolerance of other carbohydrate biosynthetic pathways. In particular, the 2-*N*-acetamido sugars present in many glycoconjugates seemed abundant targets for metabolic replacement with unnatural, chemically functionalized variants. We synthesized C-2 ketone isosteres as mimics of 2-*N*-acetamidosugars and evaluated their metabolism in a variety of mammalian cell lines that possess different levels of *N*- and *O*-linked glycosylation.[61] The C-2 ketone isostere of GalNAc (2-KetoGal, Figure 9) was incorporated via the GalNAc salvage pathway into *O*-linked glycoproteins and proteoglycans on the

surface of Chinese hamster ovary (CHO) cells. 2-KetoGal replaces the core GalNAc residue attached to serine or threonine residues at sites of *O*-linked glycosylation. Its residence therein provides a mechanism to build well-defined glycoforms from the "bottom up", rather than adding oligosaccharide structures to the periphery of the glycocalyx as transpires when unnatural sialosides present the chemical handle for chemoselective ligation.

Future Directions

The coalescence of techniques for chemoselective glycoprotein engineering with those of metabolic engineering should provide novel routes to homogeneous glycoproteins. For example, highly productive techniques for incorporating unnatural amino acids into proteins in vivo[62] may be adopted to install keto- or azido-amino acids into proteins for chemoselective attachment of oligosaccharides. An alternative approach which we term "posttranslational glycoprotein engineering" utilizes recombinant glycoproteins overexpressed in the presence of unnatural metabolic substrates. If these substrates, like 2-KetoGal, are incorporated into core positions on the protein, their chemical handles can be elaborated with defined oligosaccharides to afford homogeneous glycoprotein mimetics bearing complex glycan structures (Figure 10).

Conclusions

Chemoselective approaches to glycoprotein assembly have provided access to sufficient quantities of homogeneous glycoproteins and glycoprotein mimetics for structural and biological studies. They are likely to impact our molecular understanding of how specific glycoprotein isoforms mediate physiological and pathological events. Furthermore, these methods may allow, for the first time, the production of homogeneous glycoproteins for biotechnology and pharmaceutical applications.

References

(1) Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J. *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1999.

(2) (a) Bertozzi, C. R.; Kiessling, L. L. Chemical Glycobiology. *Science* **2001**, *291*, 2357–2364. (b) Koeller, K. M.; Wong, C.-H. Emerging themes in medicinal glycoscience. *Nat. Biotechnol.* **2000**, *18*, 835–841. (c) Sears, P.; Wong, C.-H. Carbohydrate Mimetics: A New Strategy for Tackling the Problem of Carbohydrate-Mediated Biological Recognition. *Angew. Chem., Int. Ed.* **1999**, *38*, 2300–2334.

(3) Danishefsky, S. J.; Allen, J. R. From the Laboratory to the Clinic: A Retrospective on Fully Synthetic Carbohydrate-Based Anticancer Vaccines. *Angew. Chem., Int. Ed.* **2000**, *39*, 836–863.

(4) Schachter, H. Biosynthetic controls that determine the branching and microheterogeneity of protein-bound oligosaccharides. *Biochem. Cell Biol.* **1986**, *64*, 163–181.

(5) Cummings, D. A. Glycosylation of recombinant protein therapeutics: control and functional implications. *Glycobiology* **1991**, *1*, 115–130.

(6) Grabenhorst, E.; Schlenke, P.; Pohl, S.; Nimtz, M.; Conradt, H. S. Genetic engineering of recombinant glycoproteins and the glycosylation pathway in mammalian host cells. *Glycoconjugate J.* **1999**, *16*, 81–97.

(7) Nicolau, K. C.; Mitchell, H. J. Adventures in Carbohydrate Chemistry: New Synthetic Technologies, Chemical Synthesis, Molecular Design, and Chemical Biology. *Angew. Chem., Int. Ed.* **2001**, *40*, 1576–1624.

(8) Herzner, H.; Reipen, T.; Shultz, M.; Kunz, H. Synthesis of Glycopeptides Containing Carbohydrate and Peptide Recognition Motifs. *Chem. Rev.* **2000**, *100*, 4495–4537.

(9) (a) Wang, Z.-G.; Zhang, X.; Visser, M.; Live, D.; Zatorski, A.; Iserloh, U.; Lloyd, K. O.; Danishefsky, S. J. Toward Fully Synthetic Homogeneous Glycoproteins: A High Mannose Core Containing Glycopeptide Carrying Full H-Type 2 Human Blood Group Specificity. *Angew. Chem., Int. Ed.* **2001**, *40*, 1728–1732. (b) Guo, Z.-W.; Nakahara, Y.; Ogawa, T. Solid-phase synthesis of the CD52 glycopeptide carrying an *N*-linked core pentasaccharide structure. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1464–1466.

(10) (a) Glunz, P. W.; Hintermann, S.; Schwarz, J. B.; Kuduk, S. D.; Chen, X.-T.; Williams, L. J.; Sames, D.; Danishefsky, S. J.; Kudryashov, V.; Lloyd, K. O. Probing Cell Surface "Glyco-Architecture" through Total Synthesis, Immunological Consequences of a Human Blood Group Determinant in a Clustered Mucin-like Context. *J. Am. Chem. Soc.* **1999**, *121*, 10636–10637. (b) Schwarz, J. B.; Kuduk, S. D.; Chen, X.-T.; Williams, L. J.; Sames, D.; Glunz, P. W.; Danishefsky, S. J. A Broadly Applicable Method for the Efficient Synthesis of α-*O*-linked Glycopeptides and Clustered Sialic Acid Residues. *J. Am. Chem. Soc.* **1999**, *121*, 2662–2673.

(11) Gijsen, H. J. M.; Qiao, L.; Fitz, W.; Wong, C.-H. Recent Advances in the Chemoenzymatic Synthesis of Carbohydrates and Carbohydrate Mimetics. *Chem. Rev.* **1996**, *96*, 443–473.

(12) The development of "superbeads" for sugar nucleotide regeneration by Wang and co-workers may reduce the cost of enzymatic syntheses of sugar nucleotides: Chen, X.; Fang, J.; Zhang, J.; Liu, Z.; Shao, J.; Kowal, P.; Andreana, P.; Wang, P. G. Sugar Nucleotide Regeneration Beads (Superbeads): A Versatile Tool for the Practical Synthesis of Oligosaccharides. *J. Am. Chem. Soc.* **2001**, *123*, 2081–2082.

(13) (a) Leppanen, A.; White, S. P.; Helin, J.; McEver, R. P.; Cummings, R. D. Binding of Glycosulfopeptides to P-selectin Requires Stereospecific Contributions of Individual Tyrosine Sulfate and Sugar Residues. *J. Biol. Chem.* **2000**, *275*, 39569–39578. (b) Koeller, K. M.; Smith, M. E. B.; Huang, R.-F.; Wong, C.-H. Chemoenzymatic Synthesis of a PSGL-1 N-terminal Glycopeptide Containing Tyrosine Sulfate and α-*O*-linked Sialyl Lewis X. *J. Am. Chem. Soc.* **2000**, *122*, 4241–4242. (c) Mizuno, M.; Haneda, K.; Iguchi, R.; Muramoto, I.; Kawakami, T.; Aimoto, S.; Yamamoto, K.; Inazu, T. Synthesis of a Glycopeptide Containing Oligosaccharides: Chemoenzymatic Synthesis of Eel Calcitonin Analogues Having Natural *N*-linked Oligosaccharides. *J. Am. Chem. Soc.* **1999**, *121*, 284–290.

(14) Koeller, K. M.; Wong, C.-H. Synthesis of Complex Carbohydrates and Glycoconjugates: Enzyme-Based and Progammable One-Pot Strategies. *Chem. Rev.* **2000**, *100*, 4465–4493.

(15) Plante, O. J.; Palmacci, E. R.; Seeberger, P. H. Automated Solid-Phase Synthesis of Oligosaccharides. *Science* **2001**, *293*, 1523–1527.

(16) (a) St. Hilaire, P. M.; Meldal, M. Glycopeptide and Oligosaccharides Libraries. *Angew. Chem., Int. Ed.* **2000**, *39*, 1162–1179. (b) Marcaurelle, L. A.; Bertozzi, C. R. New Directions in the Synthesis of Glycopeptide Mimetics. *Chem. Eur. J.* **1999**, *5*, 1384–1389.

(17) Lee, Y. C., Lee, R. T., Eds. *Neoglycoconjugates: Preparation and Applications*; Academic Press: London, 1994.

(18) Lemieux, G. A.; Bertozzi, C. R. Chemoselective ligation reactions with proteins, oligosaccharides and cells. *Trends Biotechnol.* **1998**, *16*, 506–513.

(19) Schnolzer, M.; Kent, S. B. H. Constructing Proteins by Dovetailing Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease. *Science* **1992**, *256*, 221–225.

(20) Kallen, R. G. Equilibria for the Reaction of Cysteine and Derivatives with Formaldehyde and Protons. *J. Am. Chem. Soc.* **1971**, *93*, 6227–6235.

(21) Tam, J. P.; Yu, Q.; Miao, Z. Orthogonal Ligation Strategies for Peptide and Protein. *Biopolymers* **1999**, *51*, 311–332.

(22) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. A new ligation method for N-terminal tryptophan-containing peptides using the Pictet–Spengler reaction. *Tetrahedron Lett.* **2000**, *41*, 4069–4073.

(23) Kemp, D. S.; Carey, R. I. Synthesis of a 39-Peptide and a 25-Peptide by Thiol Capture Ligations: Observation of a 40-fold Rate Acceleration of the Intramolecular *O*,*N*-Acyl-Transfer Reaction between Peptide Fragments Bearing Only Cysteine Protective Groups. *J. Org. Chem.* **1993**, *58*, 2216−2222.

(24) (a) Dawson, P. E.; Kent. S. B. H. Synthesis of Native Proteins by Chemical Ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923−960. (b) Dawson, P. E.; Muir, T. M.; Clark-Lewis, I.; Kent. S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266*, 776−779. (c) Tam, J. P.; Lu, Y.-A.; Liu, C.-F.; Shao, J. Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 12485−12489.

(25) (a) Hondal, R. J.; Nilsson, B. L.; Raines, R. T. Selenocysteine in Native Chemical Ligation and Expressed Protein Ligation. *J. Am. Chem. Soc.* **2001**, *123*, 5140−5141. (b) Gieselman, M. D.; Xie, L.; van der Donk, W. A. Synthesis of a Selenocysteine-Containing Peptide by Native Chemical Ligation. *Org. Lett.* **2001**, *3*, 1331−1334.

(26) MacMillan, D.; Bill, R. M.; Sage, K. A.; Fern, D.; Flitsch, S. L. Selective in vitro glycosylation of recombinant proteins: semisynthesis of novel homogeneous glycoforms of human erythropoietin. *Chem. Biol.* **2001**, *8*, 133−145.

(27) Bengtsson, M.; Broddefalk, J.; Dahmen, J.; Hendrikson, K.; Kihlberg, J.; Lohn, H.; Srinivasa, B. R.; Stenvall, K. Convergent synthesis of neoglycoconjugates by coupling of 2-bromoethyl glycosides to cysteine and homocysteine residues in T cell stimulating peptides. *Glycoconjugate J.* **1998**, *15*, 223−231.

(28) Macindoe, W. M.; van Oijen, A. H.; Boons, G.-J. A unique and highly facile method for synthesizing disulfide linked neoglycoconjugates: a new approach for remodeling of peptides and proteins. *Chem. Commun.* **1998**, 847−848.

(29) Davis, B. G.; Lloyd, R. C.; Jones, J. B. Controlled Site-Selective Glycosylation of Proteins by a Combined Site-Directed Mutagenesis and Chemical Modification Approach. *J. Org. Chem.* **1998**, *63*, 9614−9615.

(30) Rodriguez, E. C.; Marcaurelle, L. A.; Bertozzi, C. R. Aminooxy-, Hydrazide-, and Thiosemicarbazide-Functionalized Saccharides: Versatile Reagents for Glycoconjugate Synthesis. *J. Org. Chem.* **1998**, *63*, 7134−7135.

(31) Marcaurelle, L. A.; Bertozzi, C. R. Direct Incorporation of Unprotected Ketone Groups into Peptides during Solid-phase Synthesis: Application to the One-step Modification of Peptides with Two Different Biophysical Probes for FRET. *Tetrahedron Lett.* **1998**, *39*, 7279−7282.

(32) Peri, F.; Dumy, P.; Mutter, M. Chemo- and Stereoselective Glycosylation of Hydroxylamino Derivatives: A Versatile Approach to Glycoconjugates. *Tetrahedron* **1998**, *54*, 12269−12278.

(33) Cervigni, S. E.; Dumy, P.; Mutter, M. Synthesis of glycopeptides and lipopeptides by chemoselective ligation. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1230−1232.

(34) Peri, F.; Cipolla, L.; Ferla, B. L.; Dumy, P.; Nicotra, F. A highly convergent approach to *O*- and *N*-linked glycopeptide analogues. *Glycoconjugate J.* **1999**, *16*, 399−404.

(35) Cornish, V. W.; Shultz, P. G. Probing Protein Structure and Function with an Expanded Genetic Code. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 621−633.

(36) Ayers, B.; Blaschke, U. K.; Camarero, J. A.; Cotton, G. J.; Holford, M.; Muir, T. W. Introduction of Unnatural Amino Acids into Proteins Using Expressed Protein Ligation. *Biopolymers* **1999**, *51*, 343−354.

(37) Danishefsky and co-workers have shown that this native linkage is critical to the conformations of the underlying peptide backbone, which yields extended conformations of mucins: Live, D. H.; Williams, L. J.; Kuduk, S. D.; Schwartz, J. B.; Glunz, P. W.; Chen, X.-T.; Sames, D.; Kumar, R. A.; Danishefsky, S. J. Probing cell-surface architecture through synthesis: An NMR-determined structural motif for tumor-associated mucins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3489−3493.

(38) Rodriguez, E. C.; Winans, K. A.; King, D. S.; Bertozzi, C. R. A Strategy for the Chemoselective Synthesis of *O*-linked Glycopeptides with Native Sugar-Peptide Linkages. *J. Am. Chem. Soc.* **1997**, *119*, 9905−9906.

(39) Marcaurelle, L. A.; Bertozzi, C. R. Chemoselective Elaboration of *O*-linked Glycopeptide Mimetics by Alkylation of 3-ThioGalNAc. *J. Am. Chem. Soc.* **2001**, *123*, 1587−1595.

(40) Witte, K.; Sears, P.; Martin, R.; Wong, C.-H. Enzymatic Glycoprotein Synthesis: Preparation of Ribonuclease Glycoforms via Enzymatic Glycopeptide Condensation and Glycosylation. *J. Am. Chem. Soc.* **1997**, *119*, 2114−2118.

(41) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001**, *123*, 526−533.

(42) Marcaurelle, L. A.; Mizoue, L. S.; Wilken, J.; Oldham, L.; Kent, S. B. H.; Handel, T. M.; Bertozzi, C. R. Chemical Synthesis of Lymphotactin: A Glycosylated Chemokine with a C-terminal Mucin-like Domain. *Chem. Eur. J.* **2001**, *7*, 1129−1132.

(43) Backes, B. J.; Virgilio, A. A.; Ellman, J. A. Activation Method to Prepare a Highly Reactive Acylsulfonamide Safety-Catch Linker for Solid-Phase Synthesis. *J. Am. Chem. Soc.* **1996**, *118*, 3055−3056.

(44) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. Fmoc-Based Synthesis of Peptide-αThioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *J. Am. Chem. Soc.* **1999**, *121*, 11684−11689.

(45) Cotton, G. J.; Muir, T. M. Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem. Biol.* **2000**, *7*, 253−261.

(46) Noren, C.; Wang, J.; Perler, F. B. Dissecting the Chemistry of Protein Splicing and Its Application. *Angew. Chem., Int. Ed.* **2000**, *39*, 450−466.

(47) Tolbert, T. J.; Wong, C.-H. Intein-Mediated Synthesis of Proteins Containing Carbohydrates and Other Molecular Probes. *J. Am. Chem. Soc.* **2000**, *122*, 5421−5428.

(48) Macmillan, D.; Bertozzi, C. R. New Directions in Glycoprotein Engineering. *Tetrahedron* **2000**, *56*, 9515−9525.

(49) Goon, S.; Bertozzi, C. R. Metabolic substrate engineering a tool for glycobiology. In *Glycobiology: Principles, Synthesis, and Applications*; Wang, P. G., Bertozzi, C. R., Eds.; Marcel Dekker: New York, 2001; pp 641−674.

(50) Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. Engineering Chemical Reactivity on Cell Surfaces Through Oligosaccharide Biosynthesis. *Science* **1997**, *276*, 1125−1128.

(51) Yarema, K. J.; Mahal, L. K.; Bruehl, R. E.; Rodriguez, E. C.; Bertozzi, C. R. Metabolic Delivery of Ketone Groups to Sialic Acid Residues. *J. Biol. Chem.* **1998**, *273*, 31168−31179.

(52) Mahal, L. K. Biosynthetic Modulation of Cell Surface Sialosides. Ph.D. dissertation, University of California, Berkeley, 2000.

(53) Lemieux, G. A.; Yarema, K. J.; Jacobs, C. L.; Bertozzi, C. R. Exploiting differences in sialoside expression for selective targeting of MRI contrast agents. *J. Am. Chem. Soc.* **1999**, *121*, 4278−4279.

(54) Lee, J. H.; Baker, T. J.; Mahal, L. K.; Zabner, J.; Bertozzi, C. R.; Wiemer, D. F.; Welsh, M. J. Engineering Novel Cell Surface Receptors for Virus-mediated Gene Transfer. *J. Biol. Chem.* **1999**, *274*, 21878−21884.

(55) Lemieux, G. A.; Bertozzi, C. R. Modulating cell surface immunoreactivity by metabolic induction of unnatural carbohydrate antigens. *Chem. Biol.* **2001**, *8*, 265−275.

(56) Yarema, K. J.; Goon, S.; Bertozzi, C. R. Metabolic Selection of Glycosylation Defects in Human Cells. *Nat. Biotechnol.* **2001**, *19*, 553−558.

(57) Jacobs, C. J.; Goon, S.; Yarema, K. J.; Hang, H. C.; Chai, D.; Bertozzi, C. R. Exploration of the Mechanism for Substrate Selection by the Sialic Acid Biosynthetic Pathway. **2001**, submitted.

(58) Saxon, E.; Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287*, 2007−2010.

(59) Saxon, E.; Armstrong, J. A.; Bertozzi, C. R. A "Traceless" Staudinger Ligation for Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2000**, *2*, 2141−2143.

(60) Nilsson, B. L.; Kiessling, L.; Raines, R. T. High-Yielding Staudinger Ligation of a Phosphinothioester and Azide To Form a Peptide. *Org. Lett.* **2001**, *3*, 9−12.

(61) Hang, H. C.; Bertozzi, C. R. Ketone Isosteres of 2-*N*-Acetamidosugars as Substrates for Metabolic Cell Surface Engineering. *J. Am. Chem. Soc.* **2001**, *123*, 1242−1243.

(62) (a) van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. Efficient Incorporation of Unsaturated Methione Analogues into Proteins in Vivo. *J. Am. Chem. Soc.* **2000**, *122*, 1282−1288. (b) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of *Escherichia coli. Science* **2001**, *292*, 498−500.

AR9901570