Synthesis of Glycoproteins

Benjamin G. Davis†

Dyson Perrins Laboratory, University of Oxford, South Parks Road, Oxford, OX1 3QY, U.K.

Received August 9, 2001

Contents

I. Introduction

It is becoming ever clearer that the presence of carbohydrate units in naturally occurring structures and their mimetics has a dramatic effect on their physical, chemical, and biological properties. The ubiquity of glycoproteins in nature reflects their broad functions as markers in cell−cell communication events that determine microbial virulence,[1] inflammation,[2,3] and host immune responses.[4−6] In addition, the correct glycosylation of proteins is critical to their expression and folding[7−9] and increases their thermal and proteolytic stability.[10] Access to well-defined scaffolds to probe the nature of these processes is essential. Their manipulation is a dominant primary goal in glycoscience[11,12] and has driven and continues to drive the synthesis of glycoconjugates and in particular glycoproteins.

† E-mail: Ben.Davis@chem.ox.ac.uk. Phone: +44 (0)1865 275652. Fax: +44 (0)1865 275675.

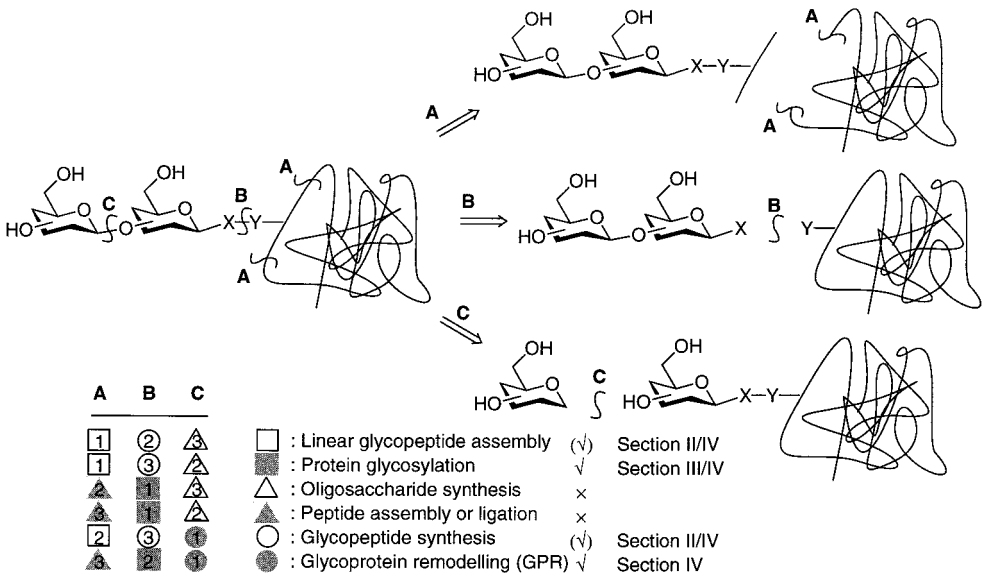
Ben Davis got his B.A. (1993) and D.Phil. (1996) from the University of Oxford. During this time he learned the beauty of carbohydrate chemistry under the supervision of Professor George Fleet. He then spent 2 years as a postdoctoral fellow in the laboratory of Professor Bryan Jones at the University of Toronto, exploring protein chemistry and biocatalysis. In 1998 he returned to the U.K. to take up a lectureship at the University of Durham. In the autumn of 2001 he moved to the Dyson Perrins Laboratory, University of Oxford and received a fellowship at Pembroke College, Oxford. His group's research centers on chemical biology with an emphasis on carbohydrates and proteins. In particular, the group's interests encompass synthesis and methodology, inhibitor design, biocatalysis, enzyme mechanism, protein engineering, drug delivery, molecular modeling, molecular biology, and glycoscience. This work has received the RSC Meldola Medal and Prize, AstraZeneca Strategic Research Award, DTI Smart Award, and Mitzutani Foundation for Glycoscience Award.

A. Coverage

Trivially, glycoproteins may be considered to be polyamino acid−poly/oligosaccharide conjugates. As such, synthetic strategies may be dissected as shown in Scheme 1. Many of the terms that have been used to describe the construction of linkages A, B, C are arbitrary ones that have been coined over time by various workers in the field but are nonetheless

10.1021/cr0004310 CCC: $39.75
Published on Web 01/29/2002

Scheme 1. Disconnective Analysis of Glycoprotein Synthesis[a]



[a] The order in which the three disconnective processes A, B, and C can be performed has a critical bearing on the strategies and techniques to be used. In the table, the various permutations and corresponding techniques (see symbols and key) are listed. The corresponding coverage of this review is noted by full, √, partial, (√), and none, ×, and relevant sections are broadly noted.

useful and shall be used where appropriate in this review. Many of these terms also describe extremes, and as for all syntheses, they simply describe points on a continuum of potential disconnective strategies. For example, some bond formations used for protein glycosylation have been or may be equally applied to peptides or vice versa, and often the distinction drawn is one of scale and not bonding type. On such occasions a distinction made between, for example, "protein glycosylation" and "glycopeptide synthesis" may in fact be unhelpful in drawing useful chemical parallels. On the other hand, different strategic considerations for macromolecules or current limitations in methodology might make such distinctions valuable on a practical level.

It is not the intention of this review to provide a complete coverage (coverage is noted by full, √, partial, (√), and none, ×, in Scheme 1, and relevant sections are broadly noted) of all methods for the construction of carbohydrate−amino acid conjugates but instead to focus on the formation of glycosylated polypeptide structures. Consequently, many aspects of glycopeptide synthesis and assembly will only be covered in passing and where it is of potential or direct relevance to glycoproteins. Many excellent reviews have appeared on the topic of synthesizing shorter-length glycopeptides, and the reader is referred to these.[13−26] Similarly, although they are crucial prerequisite building blocks for glycoprotein synthesis, the synthesis of proteins[27−31] or oligosaccharides[32−39] will not be covered unless of direct relevance. The reader is also referred to other excellent reviews that have previously covered aspects of glycoprotein synthesis either explicitly or as part of larger reviews on glycoconjugates.[11,40−44]

Glycoscience is by necessity broad in the range of techniques that it encompasses, and it is clear that in this context the oft-applied and somewhat artificial distinction between "chemical" and "biological" techniques is unhelpful and has been avoided where possible. Furthermore, because all such glycoproteins have potential function, there will be no distinction made between synthetic analogues, so-called *neo*glycoproteins, and those that occur naturally. Indeed, with the advent of an array of techniques adapted from nature, the distinction between *neo* (which simply means new) and natural has already blurred.

B. Carbohydrates as Communication Molecules and the "Cluster" Effect

Carbohydrate structures are unrivalled in the density of information that they can convey. Precise differences in the nature of the linkages between two residues, e.g., 1−2, 1−3, 1−4, 1−6 for two pyranoses, contrast with the linear nature of proteins and nucleic acids. A comparison of the permutations of hexamer formation illustrates this point well. Whereas DNA (with a basis set of 4) and amino acids (with a basis set of 20) may construct a biological language for information transfer of 4096 and 6.4×10^7 "words", respectively, carbohydrates have access to greater than 1.05×10^{12} variations.[45] Add to this the additional variety afforded by anomeric stereochemistry, ring size, and subunit modification (e.g., sulfation, phosphorylation or acylation), and it can be quickly seen that this greater variety of possible combinations gives the language of carbohydrates exquisite eloquence. This language has been christened *glycocode*, a term that well represents the potential level of complex information that carbohydrate structures are able to convey. It should also be noted that this vast number of potential permutations represents a technological barrier and means that no longer can the oligosaccharidic portions of glycoproteins be made on an iterative basis because there are far too many possible synthetic targets. It is therefore crucial that the design of new glycoproteins is guided by the identification of the associated functions and activities of existing structures.

The decipherers of glycocode are typically sugar-binding proteins called lectins, which, despite their very shallow binding sites, show a remarkable specificity in their binding of multivalent complex carbohydrate structures.[1] The monosaccharide−lectin interaction stands out as an unusually weak and relatively undiscriminating one (K_d on the order of millimolar for monosaccharides)[46] when compared to others in nature. This is largely due to the solvent-exposed nature of the lectin binding sites, which make few direct ligand contacts. The large difference in affinity shown by these shallow sites compared with the affinity of deep sites is amply illustrated by the influenza haemagglutinin lectin, which binds sialic acids with an approximately 1000-fold lower affinity than is shown by a neuraminidase found in the same virus.[47] However, when more than one saccharide of the right type and in the right orientation are clustered together, there is a rapid increase in both affinity and specificity by the corresponding lectin.[48] This increase is more than would be expected because of the increase in local concentration (statistical effect) alone and has been termed the "cluster" or "multivalent effect".[49]

The reasons for the cluster effect are yet to be rigorously determined,[50,51] but their implications are profound. First, the steady biological reservoirs of soluble monosaccharides are negligible inhibitors of any process that lectins mediate. Second, the specificity of this type of binding is exquisitely fine-tuned. It relies not only on the complementarity of the individual binding sites with a particular sugar ligand but also on the relative arrangement of the binding sites to each other in space and therefore by necessity the corresponding display of each sugar ligand relative to the next. Third, the kinetics of such binding are different from those of monovalent binding and may afford faster "on rates". Fourth, multipoint attachment is more resistant to shear stresses.

C. Why Conjugate? Why Glycoproteins?

Nature ably exploits the tertiary structure of proteins as a scaffold for multivalent display. Yet more complex third-order patterns are then in turn the product of the arrangement of these glycoproteins on cell surfaces. There are numerous examples of the important role that the protein that displays a glycan has in determining activity. As early as 1929, it was appreciated that immunological activity toward sugars may be greatly enhanced through conjugation to proteins (for further details see section I.E).[52] Furthermore, the specificity of such immunological responses to sugars varies greatly with the nature of the protein to which they are conjugated. For example, synthetically prepared glycoproteins bearing the oligosaccharide blood group determinant Lewis-y (Le^y) do not generate antibodies that can react with naturally occurring Le^y-bearing structures.[53] This could be due to a number of factors such as different densities of carbohydrate on the conjugate, as well as the influence of the protein or the carbohydrate-to-protein attachment structure upon conformation or accessibility. However, as yet, the reasons for such striking differences are undefined; an effect is observed, and it is one that requires the preparation of glycoproteins for further investigation.

There are also many examples of proteins and peptides whose biological activities are enhanced by conjugation to carbohydrates. For example, the activity of the antidiuretic nonapeptide arginine-Vasopressin is almost doubled through galactosylation.[54] Also, different ribonuclease B (RNase-B) glycoforms that were carefully separated using capillary electrophoresis show 4-fold different hydrolysis activities. They also show decreased flexibility and greater protease resistance, possibly through the action of the glycan as a "steric shield" for protease cleavage sites.[55]

Although in nature *O*-(Ser/Thr) and *N*-(Asn) glycosides dominate, the link between the glycan and the protein to which it is to be conjugated can, in theory and given the appropriate chemistry, be of virtually unlimited variety, and therefore, a detailed discussion of different spacer arms is outside the scope of this review. However, certain important structural features should be borne in mind. Several studies have investigated the effect of spacer arm length on the affinity of some proteins (e.g., selectins[56] and the asialoglycoprotein receptor[57]) for glycoproteins. The clear consensus result is that an optimal length is required that is long enough to allow accessibility but short enough that the loss of entropy upon binding is not a prohibitive cost in the binding equilibrium.

D. Need for Homogeneity/Single Glycoforms

Glycoproteins occur naturally in a number of forms (glycoforms)[58] that possess the same peptide backbone but differ in both the nature and site of glycosylation. The different properties exhibited[55,59] by each component within these microheterogeneous mixtures present regulatory difficulties[44] and problems in determining exact function through structure−activity relationships. It has even been suggested that these naturally occurring mixtures of glycoforms provide a spectrum of activities that can be biased in one direction or another as a means of fine-tuning.[55] Consequently, the few studies that have compared single glycoforms successfully[55] have required abundant sources and extensive chromatographic separation. There is therefore an urgent need for alternative sources of homogeneous glycoproteins. The development of highly successful small-molecule carbohydrate-containing ligands has often involved careful structure−activity relationship (SAR) refinements.[60] If we are to achieve the same successes with glycoproteins, then homogeneity must be one of the first priorities. This goal is one that the field of glycoprotein synthesis is now beginning to address.

Typically syntheses of glycoproteins adopt one of two strategies. The first is the formation of the putative glycan−protein link early to form glycopeptide building blocks that may then be assembled. The second is the formation of the link late in the synthesis once the protein scaffold for its presentation is in place. Given the instability that may be associated with the link[13] and the requirements for protection that need to be considered in the use of glyco-

sylated building blocks, it is clear why the latter has often seemed the most attractive option. While the construction of the protein−carbohydrate link is the focus of this review, the importance of a well-defined homogeneous source of glycan should not be understated. The indispensable syntheses of oligosaccharides[32−39] must continue hand-in-hand with methods for their conjugation to proteins and peptides.

E. Glycoprotein Properties and Potential Applications

The future of carbohydrate science will be honed by the application of its products; the applications of glycoproteins are therefore an important backdrop to this review and are the context in which their syntheses should rightly be judged.[11,43,61,62] The functions and versatility of glycosylated biomolecules are, as a consequence of their ability to transmit sophisticated information, incredibly broad, and the potential applications of glycoproteins are therefore correspondingly wide. For example, glycoproteins have been implicated in physiological processes ranging from receptor-mediated endocytosis and protein quality control to the interaction and subsequent invasion of pathogens and the triggering of effects that lead to the release of biomodulators. A thorough understanding of these processes is essential to their successful exploitation in pharmaceutical therapies that block the lectin-mediated process through inhibition or that exploit this binding to target-designed glycoconjugates to lectin-expressing cells. Indeed, the recent failure of a number of carbohydrate-based drugs may be attributed to a poor understanding of their supposed mechanism of action rather than due to any inherent flaws associated with carbohydrate therapeutics.[63] When good understanding has been achieved, the results have been impressive.[64,65]

In addition to their critical role in communication events, glycosylated proteins have long been known to have greater resistance to thermolysis and proteolysis.[66] A much less widely explored aspect is the use of glycosylated enzymes as tailor-made catalysts.[67] Examples include carbohydrate−protease conjugates, which show greater stabilities at high temperatures[68] and in organic solvents,[69,70] and those that catalyze high-yielding peptide syntheses[71] with sometimes greatly enhanced synthetic utility, altered stereospecificity,[72] and increased activities.[73] When RNase A, a non-glycosylated RNase form, is chemically mono- and diglycosylated, it shows slightly lower (80% of un-glycosylated) catalytic activity but greatly enhanced thermostability.[74]

An interesting aspect of N-linked protein glycosylation is its post-translational role in the "quality control" of protein synthesis. Without correct glycosylation, many proteins fail to fold properly. This suggests a novel role for added glycans as indicators of correct protein structure. In N-linked glycoprotein biosynthesis,[75] a 14-residue oligosaccharide core is added as a first co-translational step and then trimmed down to size. It has been suggested that if nascent protein fails to fold properly, these glycans are incorrectly displayed and cannot be processed in these trimming steps, leading to rejection and degradation. Therefore, these apparently superfluous trimming steps may not simply be a means to glycan structure but steps along a "quality-controlled" protein production line.[7−9,76] It has also been suggested that glycans aid the folding and transport of proteins by protecting them from proteolysis.[10]

The conformational effects of protein glycosylation may also be important,[26,77] as highlighted by several glycopeptide models. In aqueous solution a turn is induced in a SYSPTSPSYS segment of the C-terminal domain of RNA polymerase II when it contains a threonine O-linked to an α-D-GlcNAc monosaccharide, whereas the corresponding non-glycosylated peptide adopts a randomly coiled structure.[78] This result tallies well with recent suggestions that reversible glycosylation of this site might act as a regulatory mechanism for the control of transcription akin to phosphorylation.[79] Similarly, NMR studies on pentapeptide ALN[Galβ(1,6)Galβ(1,6)GlcNAcβ]LT suggest that glycosylation induces conformational bias[80] and those on YN[$Man_5GlcNAc_2$]LTS indicate a stabilizing effect on peptide conformation.[81] These striking differences in structure in aqueous solution are a clear indication of the importance of glycosylation, even by a single saccharide, upon local peptidic structure and open the door to a host of such crucial structural subtleties in glycosylated proteins. However, it should be noted that the method of glycosylation may also play a significant role; conformational changes observed for the glycosylation of lysine side chains through amide bond formation appeared to have been largely due to charge neutralization rather than to glycosylation per se.[82]

Glycoproteins that act as antifreeze in the serum of deep-sea fish allow them to survive at temperatures as low as −2 °C. Their ability to lower the freezing point is not proportional to concentration and is not accompanied by altered melting points. This noncolligative effect is thought to arise from a mechanism that inhibits ice nucleation and crystal growth. Oligomeric glycopeptide analogues of such proteins have been prepared and show significant antifreeze properties.[83]

The specificity of the hepatic asialoglycoprotein Gal/GalNAc-specific receptor[84] has been widely exploited for liver targeting and in many cases has been the model for targeted drug delivery[85] and gene delivery[86,87] therapies. For example, covalent attachment of drugs through various degradable linkers to lactosaminylated and galactosylated human serum albumin has allowed the targeted delivery[88] of anti-inflammatory agents, such as Naproxen,[89,90] and antivirals, such as arabinoside-AMP,[91] which are absorbed into hepatic cells through receptor-mediated endocytosis.[92] Furthermore, the immunogenicities of such glycoproteins are low if prepared with a high degree of homogeneity.[89]

Helicobacter pylori are the bacteria that cause gastric ulcers. They attach themselves to gut cells by binding to extracellular sialylated glycoproteins. This adhesion has been effectively inhibited, as part of an antiulceritic strategy, by albumin glycosylated with 3′-sialyllactosyl residues and illustrates how glycoproteins might serve a useful role as antiadhe-

sives.[93] It is interesting to think that the array of sialylated conjugates in human milk might serve the same antiadhesive function as these synthetic glycoconjugates, and once again, this highlights the importance of studying nature's tactics as models for our own.

The use of complex oligosaccharides as haptens for the induction of antibodies has a rich history dating back to the discovery in 1929 that oligosaccharides may be rendered immunogenic through their attachment to proteins.[52,94] As early as 1936, their use as a strategy to combat pneumococci was described.[95] However, until the 1970s, this work was limited by the often minute amounts of oligosaccharides available from natural sources. In a seminal series of four papers in 1975,[96] Lemieux and co-workers completed the total synthesis of the Lewis-a (Le^a) trisaccharide using newly developed glycosidation techniques and conjugated it to BSA using the acyl azide method (see section III), which they developed for the purpose. Remarkably it took until 1987 for the first carbohydrate−protein vaccine to be licensed, and since then, the development of bacterial capsular polysaccharide−protein conjugate vaccines has blossomed.[97]

As an approach to potential anticancer vaccines, Danishefsky and co-workers[98] have used, for example, reductive amination to construct a hexasaccharide-keyhole limpet hemocyanin (KLH) protein conjugate[99] at an approximate ratio of sugar to protein of 150:1. The hexasaccharide moiety, termed globo H, which was constructed using glycal methodology, was originally isolated from a glycoceramide associated with breast cancer. The synthetic globo H−KLH conjugate was successfully used to induce high anti-globo H antibody titers and induced cell lysis, in the presence of human complement, at levels approaching those of monoclonal antibodies raised against cancerous cells. Other potential anticancer vaccines have been reported on the basis of other tumor-associated epitopes[100] such as the sialyl-Tn motif, which as a KLH conjugate has shown promisingly enhanced survival rates in clinical trials.[101] The importance of carbohydrates in anticancer strategies has recently been reviewed.[102]

T cells mediate intercellular immune responses and are particularly useful because they allow even infants under the age of 2 to be effectively vaccinated. While oligosaccharides alone do not typically elicit T-cell responses, the discovery[103] that glycopeptides do has allowed immune responses to sugars to be probed systematically. The recent developments in the synthesis of more complex glycans bound to peptides and proteins in a site-specific manner described in this review should allow these aspects to be probed in ever-greater detail in the future.

The approaches used to target drugs described above may also be used to target physiologically beneficial enzymes in so-called enzyme replacement therapy (ERT). For example, the mannosylation of enzymes has allowed the targeting of enzymes to particular diseased cells. Replacement β-glucocerebrosidase, an enzyme that is lacking in Gaucher disease,[104] and the beneficial antioxidant effects of superoxide dismutase (SOD)[105] have both been directed to macrophages. SOD has also been conjugated with sodium hyaluronate, a polymer of the dimeric motif GlcNAc−glucuronic acid. This combined the ability of SOD to catalyze superoxide anion decomposition with the hydroxyl radical scavenging capacity of hyaluronate in a potentially dual-action antiinflammatory.[106]

Novel analytical techniques have also utilized glycoproteins. For example, layer-by-layer deposition of a mannose-specific lectin and mannosylated enzymes (glucose oxidase and lactate oxidase) on a platinum surface allowed the preparation of a sensitive active-enzyme electrode.[107]

II. Glycopeptide Assembly

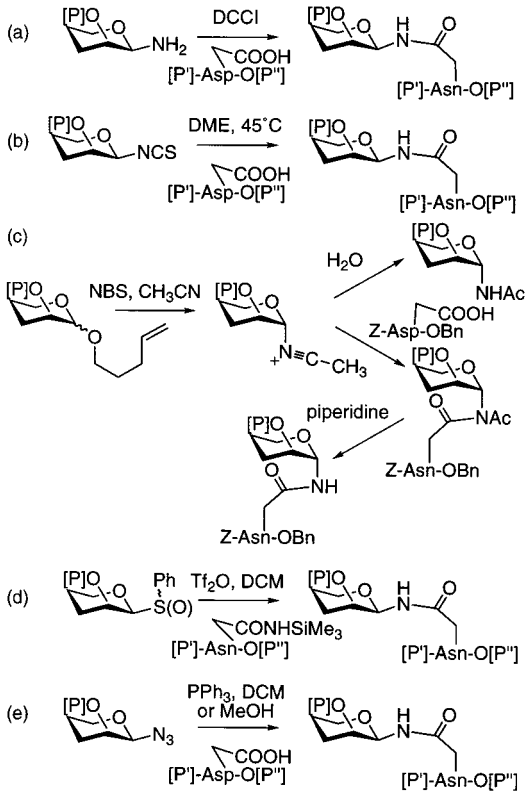
The need for homogeneous samples (single glycoforms) that was outlined above has resulted in great effort in the field of de novo synthesis of glycoproteins. The linear assembly of glycosylated amino acids in particular has, from the very first examples (such as the use of *N*-acetylglucosaminylasparagine in the synthesis of a partial sequence of fibroblast interferon[108]), provided well-defined products. Thus, the required carbohydrate structure is attached to an amino acid residue (typically serine and threonine for O-linked glycopeptides and asparagine for N-linked glycopeptides). An excellent review of methods for the formation of the glycosidic link between peptides and glycans ("Glycopeptide Synthesis", Scheme 1) has been published.[19] Arsequell's reviews also contain good summaries of approaches to N-[21] and O-linked[20] glycopeptide systems. Suitably protected, the glycoamino acid/peptide is then used as a building block in strategies that often rely heavily on standard solid-phase peptide synthesis (SPPS) techniques. That these elegant approaches are still some way off synthesizing peptides of lengths approaching those of proteins[30] is a testament to the inherent difficulties of this approach. Two factors limit the work: first, the need not only for extensive carbohydrate protection but also for amino acid protection regimes and, second, the acid and base lability of glycosylated amino acid residues.[13] Indeed, this feature of glycoproteins has long been exploited to strip glycans from protein surfaces. The necessary protection and deprotection regimes, the use of particular supports, including the introduction of specific linkers, and coupling methods have all been tailored to be compatible with the presence of carbohydrates. Several excellent reviews[13−26] cover these aspects in detail, but some recent examples that are intended to illustrate these strategies follow.

A. Strategies for Forming the Glycosidic Linkage

A.1. Linkage Variety

A vast majority of the glycan−protein linkages observed in nature are *O*-(Ser/Thr) and *N*-(Asn) glycosides. Most common among the motifs are N-linked motifs (GlcNAcβ-Asn) and O-linked motifs (GlcNAcβ-Ser/Thr and GalNAcα-Ser/Thr), although other unusual linkage forms are also observed,[109] such as

Scheme 2. Synthesis of the N-Link in Glycosylasparagine Amino Acids[a]



[a] (a) See refs 118−121. (b) See ref 122. (c) See refs 123−125. (d) See ref 126. (e) See refs 127 and 128.

GlcNAc-α-Asn,[110] Glc-Asn,[111] GalNAc-Asn,[112] Rha-Asn,[113] and even C-linked Manα-Trp.[114] The Complex Carbohydrate Structure Database (CCSD or Carb-Bank) [http://bssv01.lancs.ac.uk/gig/pages/gag/carb-bank.htm] acts as the oligosaccharide equivalent to GenBank by cataloguing many of these structures. It allows access to information about structure, author, etc. and the compilation of records that may be submitted to the CCSD. A guide to its use has recently been published.[115,116] O-GLYCBASE[117] [http://www.cbs.dtu.dk/databases/OGLYCBASE] is a database of over 170 glycoproteins with experimentally verified O-linked glycosylation sites, compiled from protein sequence databases and literature. It contains information about the glycan, information about the sequence, and literature references and is http-linked to other databases. The information in this database has also provided the basis for a predictive neural network that may be applied to predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins and is available at http://www.cbs.dtu.dk/services/NetOGlyc/. Given the apparent lack of consensus sequences for O-glycosylation, which contrasts with the fairly good predictability in N-glycosylation, such tools are very useful.

A.2. Natural Linkage Types

The formation of N-linked glyco-Asn motifs through diimide-mediated coupling of glycosylamines (Scheme 2a) with an Asp side chain carboxylate was first demonstrated as early as 1961[118] and has been widely exploited. Although by far the most popular method, it should be noted that amide bond formation of an anomeric glycosylamine to an aspartate carboxylate side chain is often plagued by the ready anomerization of the glycosylamine, and this may lead to the formation of anomeric mixtures of glycopeptides.[119,120] Some reports suggest that such anomerization is reduced by the use of propanedithiol as a reductant.[121] The glycosylamine approach may also be complicated by intramolecular aspartimide formation if the Asp to be coupled is one residue toward the N terminus of a glycinyl or alaninyl residue.

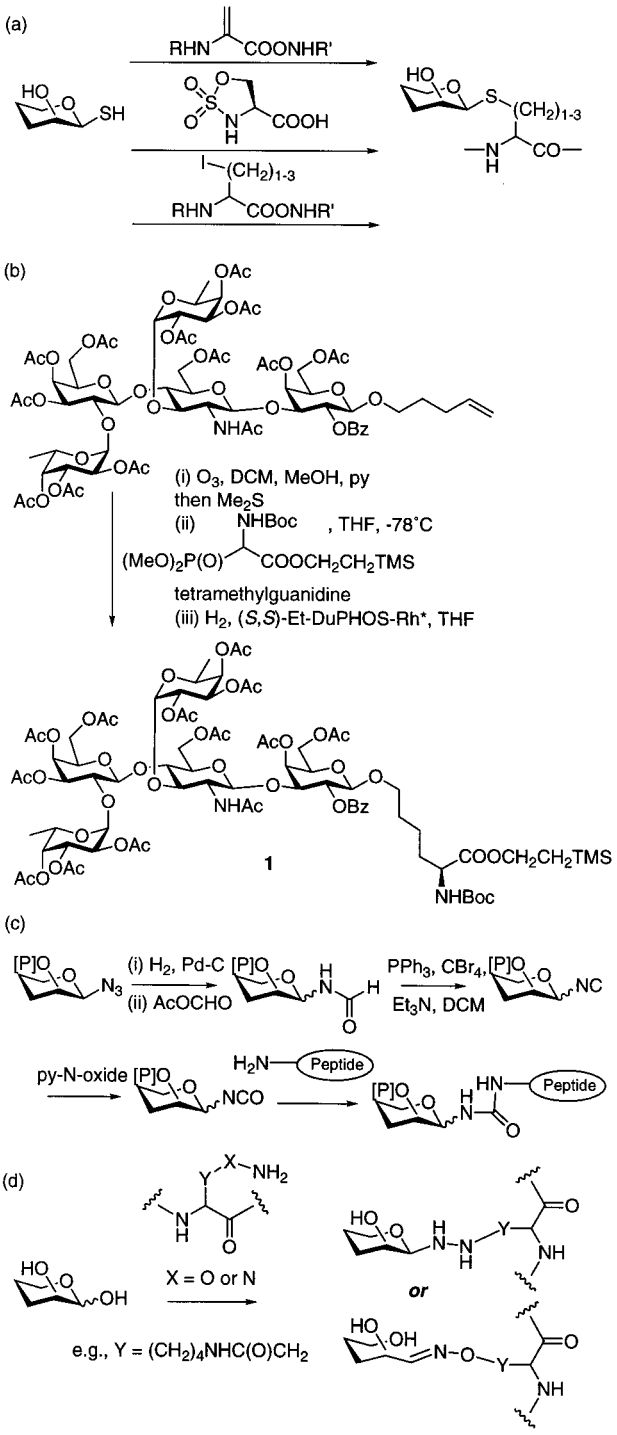
Other valuable methods exist. As a landmark achievement, a protected Manβ(1,4)GlcNAcβ(1,4)-GlcNAcβ-Asn motif was first synthesized from the reaction of the isothiocyanate of the protected trisaccharide with the carboxylate side chain of a partially protected Asp (Scheme 2b).[122] In the key amidation, *N*,*N*-bisglycosylthiourea formation also competes. Use of the Ritter reaction has allowed synthesis of the key GlcNAcβ-Asn linkage (Scheme 2c).[123] Thus, reaction of certain nitriles followed by hydrolysis of the nitrilium intermediates results in direct amide formation. More complex amides, such as side chain N-linked asparagine derivatives, were formed indirectly from attack of the appropriate carboxylate side chain on a β-acetonitrilium followed by deacetylation. Neighboring group participation by a C-2 phthalimide is an important factor to prevent preferential α-amide formation.[124] Subsequent glycosylation of the GlcNAc introduced in this way allowed GlcNAcβ(1,4)-GlcNAcβ-Asn motif synthesis. Other aromatic carboxylic acids and effects of variations in glycosyl substituents/protection on the glycosyl nitrilium have also been investigated.[125] A silylated amide asparagine side chain has been directly N-glycosylated with glycosyl sulfoxide donors (Scheme 2d).[126] Reaction of glycosyl azides with Asp side chain carboxylates in the presence of PPh_3 [127] in DCM or MeOH allows good yields of β-Asn *N*-glycosides (Scheme 2e).[128] While a majority of the above examples have been applied to glyco-Asn synthesis prior to assembly, convergent N-glycosylation of peptides also succeeds (section II.B.2).

A majority of O-linked glycoamino acids have been synthesized by standard glycosylation techniques with glycosyl donors.[20] For example, O-glycosylation of Ser or Thr with, for example, a trisaccharide trichloroacetimidate glycosyl donor allows O-linked trisaccharide−amino acid building blocks to be synthesized.[129] Recently, the glycosylation of hexafluoroacetone-protected hydroxyl-bearing Ser, hydroxyPro, and Tyr allowed preparation of glycosylated-dipeptides in just three steps from unprotected amino acid, e.g., T[Ac_4-*O*-β-Glc]Y in 70% overall yield from T and Y.[130] As yet, no convergent O-glycosylation of a peptide has been shown, and the failure of, for example, pentenyl glycosides in this regard has been noted.[131]

A.3. Unnatural Linkages

Some valuable techniques for forming unnatural glycan−peptide linkages (sometimes called "neoglycopeptides") may usefully be considered in the context of glycoprotein synthesis. Some of the techniques that

Scheme 3. Synthesis of Unnatural Link Glycopeptides[a]



[a] (a) See refs 133−136. (b) See ref 131. (c) See ref 138. (d) See refs 139−146.

follow have only been applied or indeed are limited strategically to peptide systems alone. However, others may be usefully adapted to larger protein systems; indeed, few of the examples described in section III ("Chemical glycoprotein synthesis") contain natural linkages.

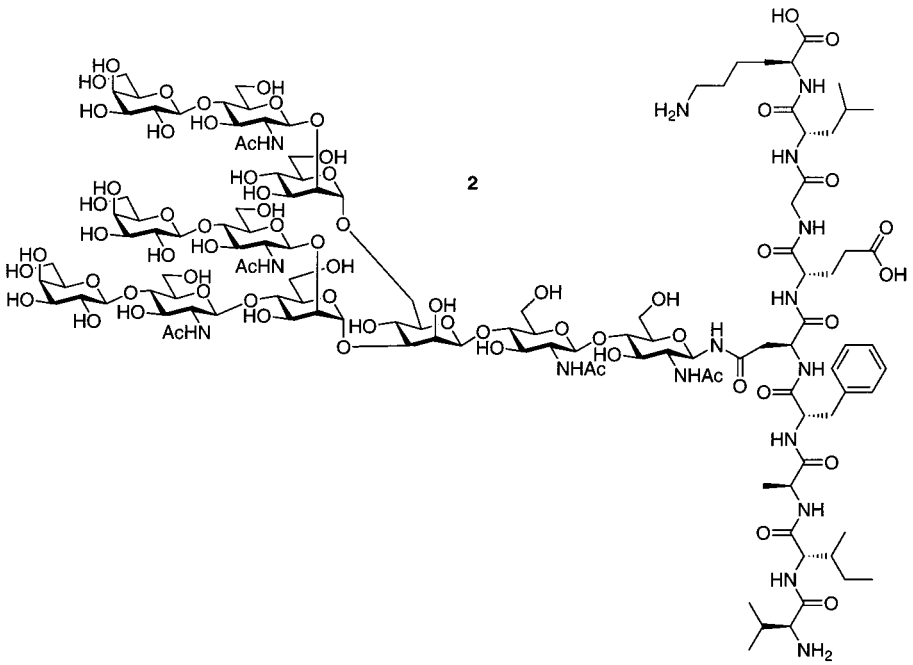
The use of a serine-derived sulfamidate (Scheme 3a), based on previous methods,[132] allows reaction with thiohexoses to create S-linked glycoamino acids in water, and this is a method that could potentially be adapted to convert N-terminal serines to glycosylcysteines in peptides, although some amino acid epimerization was noted and this may limit its general applicability.[133] Oxidative elimination of phenylselenocysteine gives dehydroalanine in various small peptides that can act as α,β-unsaturated conjugative addition acceptors for various protected and unprotected thiosugars, although with poor stereoselectivity (Scheme 3a).[134] Solid-supported unprotected thioglucose can also be used as a nucleophile to substitute a number of side chain amino acid iodides to give S-linked glycoamino acids.[135] Similarly, iodide displacement of an iodoserine or Mitsunobu coupling using 1-deoxy-1-thio-GlcNAc gave Fmoc-Cys[GlcNAc], although in this study racemization was observed during iodide displacement because of in situ elimination and subsequent conjugate addition.[136]

Danishefsky and co-workers have created unnaturally O-glycosylated glycoamino acids bearing various tumor-associated antigens.[131] The method uses a pentenylglycoside-derived aldehyde and a glycine phosphonate based on a Horner−Emmons method first devised by Toone and co-workers.[137] The resultant $-(CH_2)_4O$-glycoside side chain amino acids, e.g., **1** (Scheme 3b), were then peptide-coupled to create a "multiantigenic peptide". It should be noted in the context of attempts to form natural *O*-glycoside linkages to amino acids that despite various attempts, convergent O-glycosylation by using the *O*-pentenylglycosides as donors to glycosylate Ser or Thr failed and was abandoned as a strategy.

Glycosylisocyanates prepared by oxidation of isocyanides, in turn prepared from anomeric formamides, react with amines to give urea linkages, and this has allowed one example of a glycosylamino acid conjugate to be formed (Scheme 3c).[138]

Chemoselective ligation[139] approaches have been variously applied to unnaturally linked glycopeptide synthesis. Both α-amino and lysine ϵ-amino groups were derivatized with aminooxyacetyl groups before reaction of the free amine with reducing sugars to form oxime linkages (Scheme 3d). The use of orthogonal N-protection also allowed regioselective glycosylation.[140] In a very similar manner, reaction of a 12-aa peptide containing an N-terminal aminooxyacetyl function has also been described.[141] The use of a basic peptide allows this to be used as a method for enhancing the MALDI-MS sensitivity of oligosaccharides for carbohydrate sequencing work. Good aminooxy selectivity over lysine N-selectivity, attributable to α-effect-enhanced nucleophilicity, was observed. Aminooxyacetyl Lys capping and oxime formation with the reducing end of lactose has also been applied to a sulfopeptide.[142] This type of oxime formation may be plagued by low E versus Z stereocontrol and equilibria between open-chain oxime carbohydrate and cyclic hemiaminal pyrano or furano forms. In contrast, *N*-methylaminooxyacetyl-capped lysines react with the reducing ends of various glucosylsaccharides to give cyclic β-N-linked glycopyranosides with moderate yield and good selectivity;[143] indeed, as observed previously,[141] Lys side chains in the peptide do not react. Imperiali has recently adapted these ideas by creating nonnatural tripeptides con-

Scheme 4. Preparative-Scale Hydrazinolysis, Glycosylamine Formation, and Then Carbodiimide-Mediated Coupling Gave 2[150]



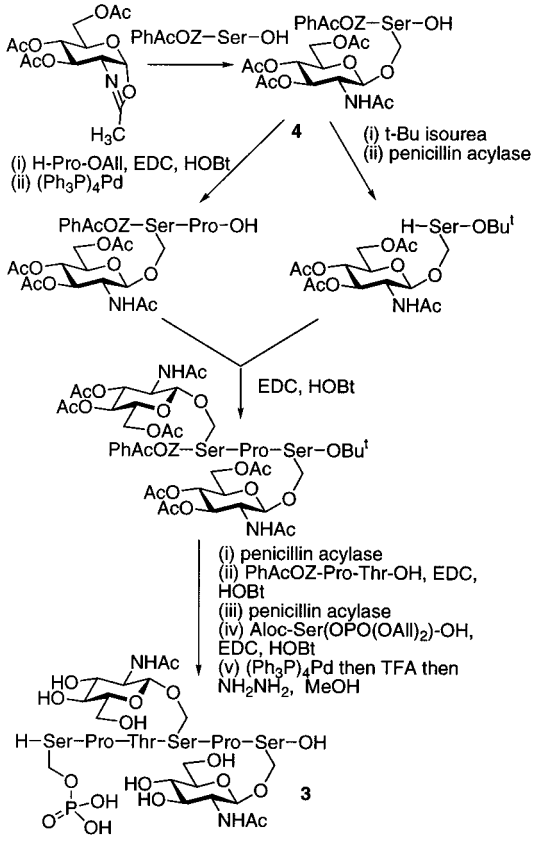
taining β-hydroxylamine and alanine-β-hydrazide (Scheme 3d) and reacted them with *N*-acetylglucosamine, also via its reducing end.[144] Double chemoselective thioether and hydrazone formation in SPPS-derived lysine clusters allowed glycosylation of lysine side chain termini using thioethyl glycosides while extending the keto-functionalized C terminus with an antigenic hydrazide peptide sequence.[145,146] Glycosylmaleimides can be used to modify cysteinyl-containing peptides[147] (see also section III.B), and in a reverse sense maleimidopeptides react with unprotected 2-thioethyl-*N*-glycosides; the latter compounds were easily prepared from the unprotected parent carbohydrate.[148]

B. Assembly Strategies

B.1. Linear Assembly

These techniques rely heavily on SPPS methodology. Impressive examples include the synthesis of the RNA polymerase II C-terminal domain heptapeptide repeat containing a single GlcNAcβSer and variously N-acetylglucosamidated-Asn 11-aa sequence from the C-terminal domain of mammalian neurofilaments.[149] Preparative scale hydrazinolysis allows *N*-glycans to be stripped from N-linked glycoproteins such as fetuin on scales of up to 500 mg for use in the construction of glyco-Asn building blocks for use in linear assembly. In this way, glycononapeptide **2** (Scheme 4) was synthesized from carbodiimide-mediated coupling of oligosaccharylglycosylamines, derived from the corresponding glycosylhydrazide-stripping product after purification, with a protected Asp.[150] This building block was used in linear assembly; product yields were lower for the synthesis of the bulkier undecasaccharide-bearing glycoamino acid than for a monosaccharide-bearing glycoamino acid building block (35% and 78%, respectively).

Scheme 5. First Synthesis of a Glycophosphopeptide 3[151]



The first example of the synthesis of a glycophosphopeptide **3** is shown in Scheme 5. The enzyme labile group PhAcOZ was used to protect the N terminus of a serine residue, which was glycosylated with GlcNAc. This formed the key building block **4** in the synthesis of **3**. Penicillin acylase allowed the removal of PhAcOZ, and the free amine was coupled

using carbodiimide to a glycosylated dipeptide, itself prepared from key block **4**. Again, enzyme cleavage allowed the N terminus to be freed before further couplings with peptides and finally a phosphoserine residue bearing a protected phosphorylated side chain. Global deprotection then afforded **4**.[151]

Protease-mediated peptide ligation is as yet not capable of general coupling of the glycopeptide block because of the often stringent specificities of these enzymes (see section IV.E).[152,153]

A hybrid strategy in which a glycosylated tripeptide was oligomerized up to 12 repeating units has been reported. The tripeptide Z-AAT-Bn was glycosylated with Galβ(1,3)GalNAc, deprotected, and then treated with diphenylphosphoryl azide as an activator to form an oligoglycopeptide with significant antifreeze properties.[83] Block coupling strategies may also be applied to the construction of non-proteolyzable (and therefore with potentially higher oral bioavailability) peptide mimics (glycopeptoids)[154] such as the LN(GlcNAc)FKA mimic **5**,[155,156] (Figure 1),

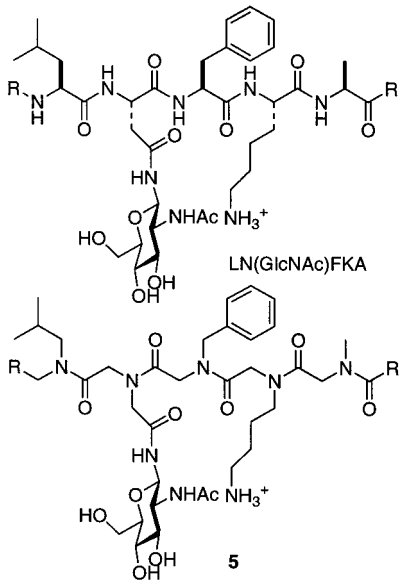


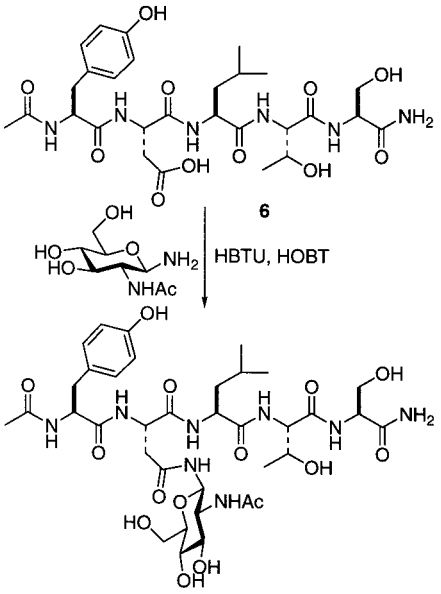
Figure 1. "Glycopeptoid" mimic **5** of LN(GlcNAc)FKA.[155]

which may also show interesting conformational restriction as a result of rotamer formation. Similarly, a linear glycopeptoid mimic of the Tn-antigen (GalNAc-α-Ser/Thr) was prepared using a reiterative TBTU-mediated coupling of a single orthogonally protected aminoester building block.[157] This method has been elegantly extended to the concept of glycopeptoids in which both the interresidue and side chain distance may be varied. Through the incorporation of different aromatic, amine, and peptoid spacer units, the spatial presentation of *C*-glycosides in these structures may be optimized.[158]

B.2. Convergent Assembly

As has been shown above, a linear strategy in glycopeptide synthesis is more usual because direct peptide glycosylation is often unsuccessful, given the variety of functional groups that would be required to protect a given oligopeptide of any significant length. However, Lansbury and co-workers have adapted the use of glycosylamine coupling with carboxylates in a convergent approach to glycopeptide synthesis. For example, HBTU-mediated coupling of GlcNAc glucosamine with the side chain aspartate carboxylate in a pentapeptide **6** allowed the formation of an Asn-linked *N*-acetylglucosaminyl containing glycopeptide (Scheme 6).[159] This method was

Scheme 6. Glycosylamine Coupling with the Side Chain of Aspartate in Peptides Allows Convergent N-Linked Glycopeptide Synthesis[159]



successfully extended to peptides containing more complex glycans such as the high-mannose core of N-linked glycoproteins, $Man_5(GlcNAc)_2$.[160] Furthermore, this method has been expanded to encompass solid-phase bound glycosylamines, which are then coupled to side chain carboxylates in pentapeptides before the peptide chain is further extended.[161,162]

The same convergent strategy has been applied to the synthesis of a high-mannose core glycopeptide; again, anomerization of the pentasaccharide glycosylamine used led to the formation of a mix of α and β Asn-GlcNAc linkages.[120] In a remarkable demonstration of this methodology a 15-mer glycosylamine corresponding to a high-mannose H-type 2 blood group determinant was coupled to the carboxylate of a LADVT pentapeptide. Although the low yield for this step (20%) indicates the apparent difficulties of this type of convergent coupling, this is a remarkable testament to the types of structures that can now be constructed. [163] Despite these successful examples of convergent N-linked glycopeptides, it should be noted that the convergent O-glycosylation of peptides to create O-linked (e.g., Ser/Thr) glycopeptides has not been achieved.

B.3. Elaborative and Mixed-Assembly Strategies

A mixed strategy of an initial linear assembly coupled with convergent elaboration may also be successful. For example, in an elegant combination of both chemical and glycosyltransferase-mediated glycosylation, the synthesis of an undecasaccharide-linked asparagine residue **7** (Figure 2) demonstrated that highly glycosylated glycopeptide building blocks

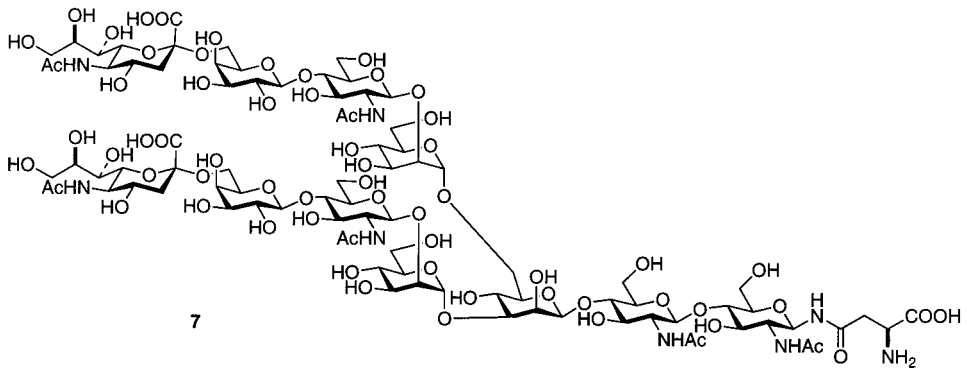
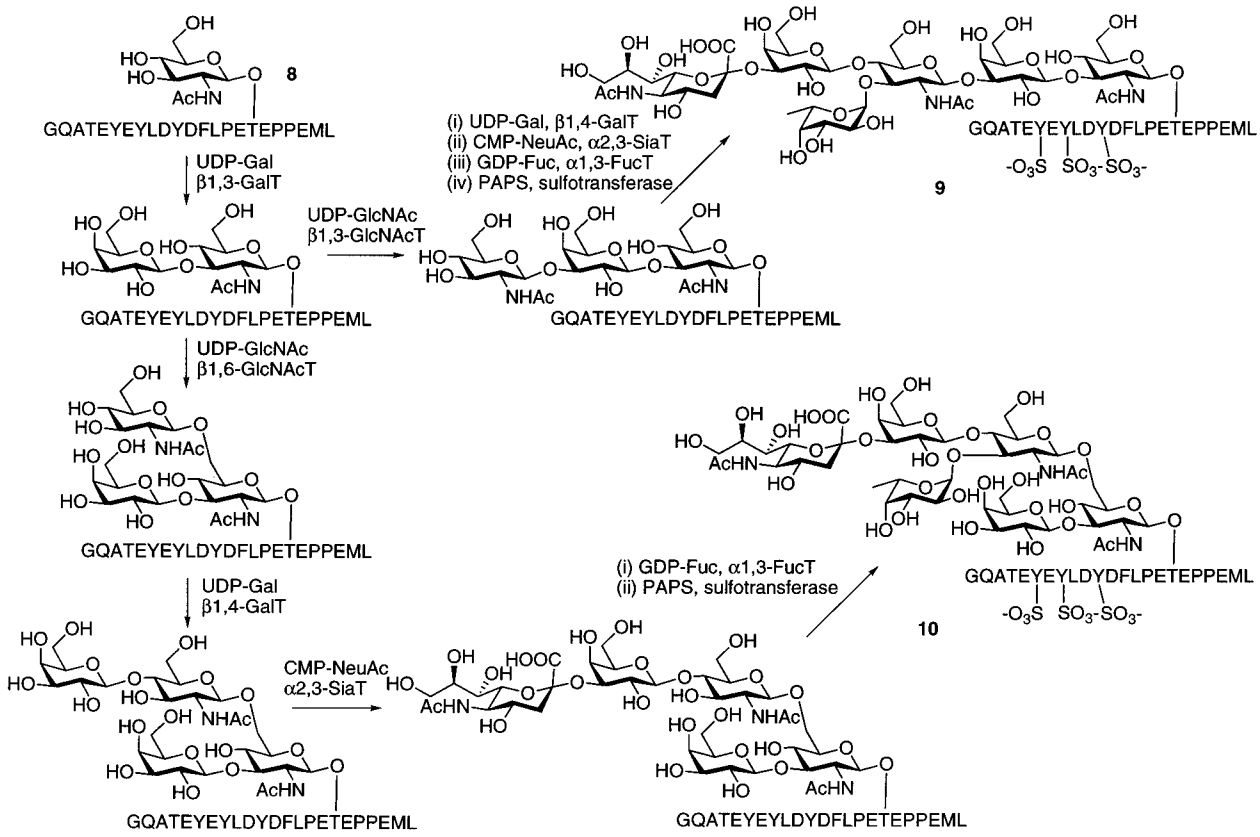


Figure 2. Undecasaccharide-asparagine **7** was formed through a combination of both chemical and glycosyltransferase-mediated glycosylation.[164]

Scheme 7. Glycosyltransferase- and Sulfotransferase-Mediated Synthesis of the Sulfated, Differently Glycosylated, N-Terminal Domain Segments of PSGL-1[167]



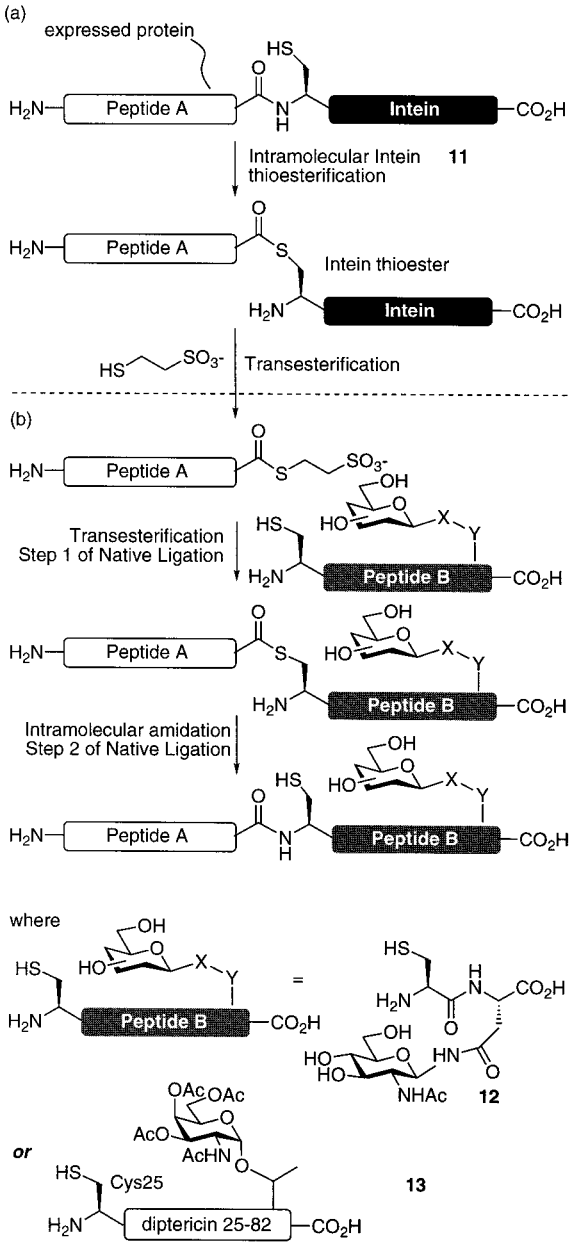
were readily accessible.[164] This was an important step toward the first synthesis of a glycopeptide fragment bearing this same full N-linked oligosaccharide, which was achieved through a linear assembly of a heptasaccharide-bearing Asn into a pentapeptide followed by galactosyltransferase and sialyltransferase-catalyzed elaboration.[165]

A sulfated N-terminal octapeptide from P-selectin glycoprotein ligand (PSGL-1) containing a pentasaccharide amino acid α(1,6)GalNAcαThr has been synthesized using a combination of initial SPPS with a disaccharide−amino acid building block and chemical tyrosine sulfation followed by glycosyltransferase-mediated elaboration.[166] The presence of sulfotyrosine makes this a particularly difficult glycopeptide to elaborate, and in this study alternative glycosyltransferases more tolerant than those used previously had to be found. In an excellent display of the power of glycosyltransferases in glycopeptide elaboration (Scheme 7), the sulfated N-terminal domain of PSGL-1 has been synthesized on a small scale and characterized by HPLC and MS in two glycoforms that crucially differ in their binding to P-selectin as a result of only a subtle alteration of internal glycan structure.[167] Isolation of the required six glycosyltransferases and one sulfotransferase allowed **9** and **10** to be synthesized from the linearly SPPS-assembled 23-aa monosaccharide glycopeptide **8**.

B.4. Native Ligation Assembly

Coupling of a 24-aa mono-GalNAcα-bearing N-terminal portion thioester with a 58-aa mono-GalNAcα-bearing C-terminal portion that has a cysteinyl N terminus allowed the synthesis of antimicrobial O-linked glycoprotein diptericin.[168] This ligation was based on the principle of native chemical

Scheme 8. Native Ligation Assembly of Glycoproteins: Native Chemical Ligation (b) of C-Terminal Thioesters with Cysteinyl N-Terminus Glycopeptides[168,170] and Expressed Ligation, [(a) Then (b)][171,172]



ligation of C-terminal thioesters with N-terminal cysteinyl peptides as developed by Kent.[169] This essentially proceeds by a transthioesterification by **13** and a subsequent S → N shift (Scheme 8b). Normally peptide thioesters may be synthesized through SPPS using Boc-based protecting strategies that do not cleave the normally employed thioester linkage to the support. Fmoc methods cannot be used because this leads to thioester degradation. However, because of the acidity of the TFA treatment required, Boc-based methods are incompatible with glycosidic linkages. The adaptation of a novel sulfonamide safety-catch linker allowed the synthesis of the required glycopeptide thioester via Fmoc strategy. To allow the syntheses, additional modifications were also needed. Critically, alteration of residue 25 in **13** from Gly → Cys was necessary to allow native chemical ligation by incorporating a Cys at the N terminus of the C-terminal portion. In addition, to prevent intramolecular aspartimide formation during SPPS, two Asp-Gly and Asp-Asn motifs were knocked out by alteration of Asp → Glu. Finally, after ligation, hydrazine treatment allowed deprotection of the acetylated glycans.

Linear assembly of a 46-aa C-terminal segment containing eight Ser/Thr-O-α-GalNAc residues followed by native chemical ligation with a 47-aa N-terminal peptide fragment has also allowed the first synthesis of a mucin-type glycoprotein, the chemokine lymphotactin.[170] Interestingly, expressed ligation, i.e., recombinant preparation (expressed protein ligation; see below and section V.D) of the N-terminal thioester, failed in this system and lower yields (38%) were also obtained from the native ligation because of the presence of a valinyl residue at the C terminus of the N-terminal fragment. In this case, glycosylation appeared to have little or no effect on the structure or activity of lymphotactin.

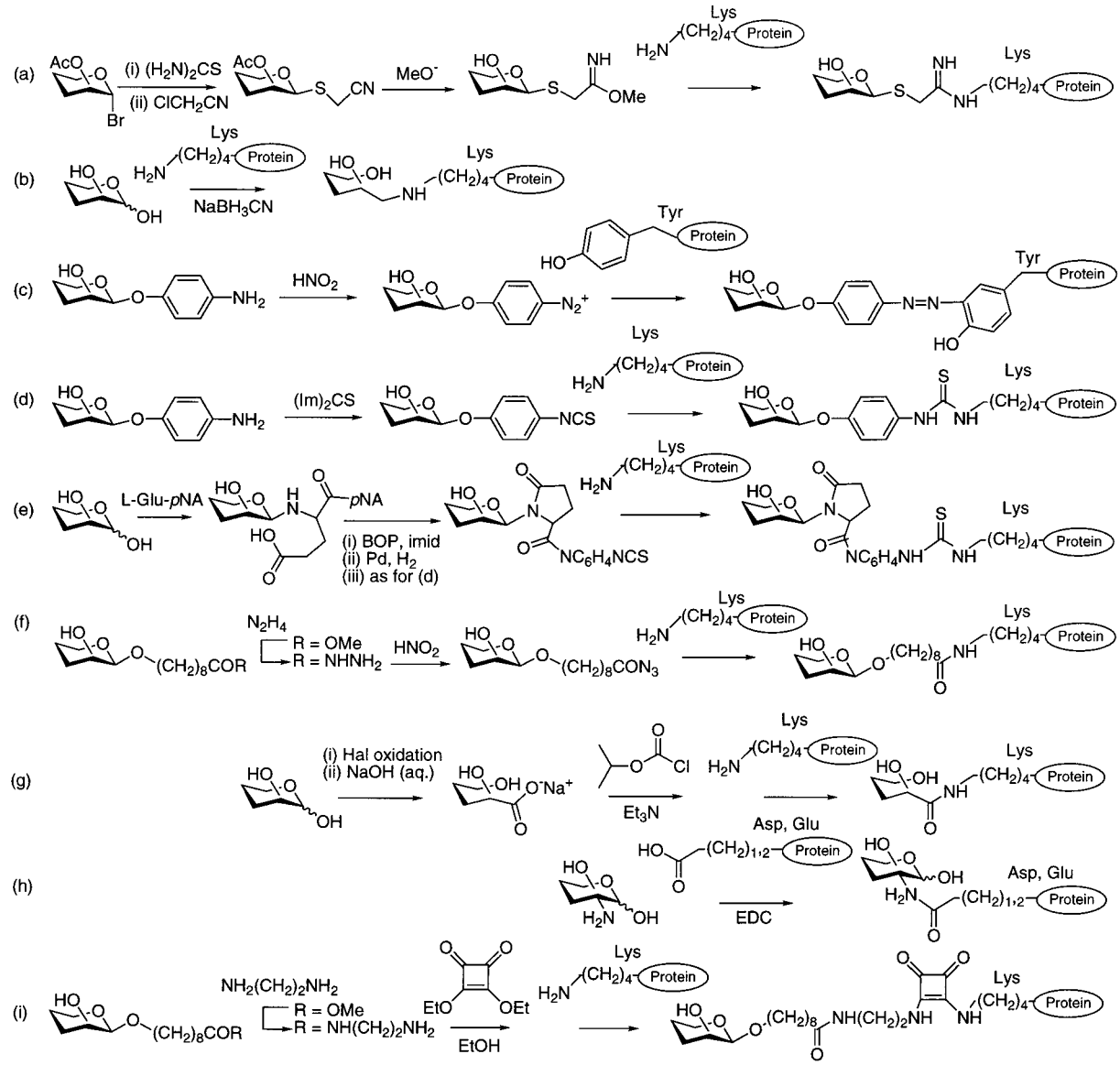
Expressed protein ligation (section V.D) has been used to incorporate variously modified cysteines at the C terminus of bacterially expressed mannan-binding protein (MBP) including Cys-Asn(*N*-*β*-GlcNAc) **12**.[171] Thus, MBP was expressed in *E. coli* as a fusion of **11** (Scheme 8a) to the N terminus of a widely used intein from *Saccharomyces cerevisiae*; this intein also bears a chitin-binding domain at its C terminus. Inteins are protein motifs that are self-spliced from protein sequences. The first step of splicing is an N → S acyl shift at the Cys N-terminal amino acid of the intein with a suitable C terminus of the MBP (Scheme 8a). The resultant thioester can be purified on chitin beads before being transthioesterified to give a soluble thioester of MBP. MBP can then be natively ligated[169] (Scheme 8b) with a cysteinyl N-terminus peptide such as **12**.

Bertozzi and co-workers have also used expressed protein ligation to construct two model nonnatural glycoproteins.[172] First, a 15-residue mucin-like glycopeptide stretch from human lymphotactin bearing five peracetylated GalNAc[α]-Thr residues and one peracetylated GalNAc[α]-Thr residue was constructed by Fmoc SPPS. Second, a GlyCAM-1 central fragment, the stretch that is not glycosylated in naturally-occurring GlyCAM-1, was expressed as an intein−chitin binding domain fusion protein in bacterial culture and purified on chitin beads. These two portions, the 15-aa glycopeptide, as a segment with a cysteinyl N-terminus, and the unglycosylated 42-aa GlyCAM segment, released as a C-terminal thioester, were then ligated with approximately 70% conversion simply through incubation for 48 h and then deacetylation using hydrazine hydrate before HPLC purification. A second unglycosylated 77-aa domain was also ligated to the same glycopeptide in this way.

III. Chemical Glycoprotein Synthesis

While in certain cases, the convergent glycosylation of oligopeptides may be successful, it is limited by a lack of suitable functional groups (either by a lack of reactivity or an inability to differentiate groups) when applied to proteins. For this reason, alternative

Scheme 9. Indiscriminate Glycoprotein Syntheses[a]



[a] (a) See ref 174. (b) See ref 175. (c) See refs 52 and 179. (d) See ref 180. (e) See ref 181. (f) See ref 96. (g) See refs 183 and 184. (h) See refs 74, 187 and 188. (i) See ref 189.

glycoprotein synthesis techniques have been more widely applied.[11,40,44,62,66,67,173]

A. Indiscriminate Glycosylation[279]

The use of 2-iminomethoxymethyl thioglycosides[174] and reductive amination methods[175] are still, after 20 years, the most frequently used strategies for glycoprotein preparation. The former may be readily prepared by the action of methoxide on cyanomethylthioglycosides that may in turn be derived from 1-thioaldoses (Scheme 9a). For the latter, Gray originally modified albumin with lactose through $NaBH_3$-CN-mediated reduction[175] (Scheme 9b), although borane may also be used. This method is amenable to other sources of aldehyde functionality such as those generated by ozonolysis of unsaturated spacer arms,[176] through periodate cleavage of diols, or the hydrolysis of acetal-containing spacer arms.[177] Conjugations through reductive amination are often accompanied by low protein loading levels, which in some cases are due to steric hindrance caused by short spacer arms. In a pragmatic approach to circumventing this problem, a second hydrazide spacer arm can be used to extend an existing aldehyde-terminated spacer.[178] Reaction of the maleimido terminus of the resulting longer spacer arm with thiols introduced to the surface of the protein keyhole limpet hemocyanin (KLH) allowed 5-fold greater loading of the sialyl-GalNAc disaccharide, sTn.

The use of glycosidic aromatic diazonium salts, derived from the corresponding *p*-aminoarylglycosides, as electrophiles to functionalize proteins was first demonstrated as early as 1929.[52] They modify a wide range of electron-rich side chains within protein structures, such as those of aromatic tyrosinyl and tryptophanyl or nucleophilic lysinyl and histidinyl residues (Scheme 9c).[179] *p*-Aminoarylglycosides may also be elaborated to phenyl isothiocyanates, which react more selectively with amino groups alone (Scheme 9d).[180] A one-pot, two-step preparation of anomeric *p*-nitroanilide (*p*NA) pyroglutamates from unprotected carbohydrates also provides an alterna-

tive route to aromatic isothiocyanates.[181] Following glycosylamine formation with the α-amino group of *p*NA glutamic acid, the side chain γ-carboxylic acid readily reacts with the resulting secondary amine to give a pyroglutamate that may then be elaborated (Scheme 9e). Glycosyl isothiocyanates have also been used in this way as protein-labeling reagents to elucidate carbohydrate transport mechanisms.[182]

In 1975, Lemieux and co-workers described the use of highly activated acyl azides for the formation of amides from proteinaceous amines and carboxylate ester terminus spacer arm carbohydrates.[96] The esters were converted to acyl hydrazides before oxidation with nitrous acid to give the corresponding acyl azides (Scheme 9f). Mixed-anhydride methods are well established for the activation of carboxylic acids to form carboxyl derivatives, and in this way aldonic acids may similarly be coupled to proteinaceous amines (Scheme 9g).[183,184] Similarly, carbodiimide chemistry[185] and the use of *N*-carboxyanhydrides[186] have also allowed the coupling of aldonates and glycosylated amino acids as sources of glycans bearing carboxylic acids. In a reverse sense, carbodiimide chemistry has also been employed in an attempt to activate, and therefore glycosylate, protein carboxylates (Scheme 9h).[187] RNase-A was mono- and diglycosylated (mixtures of mono 45%, di 34% and unglycosylated 21% were obtained) by EDC-mediated coupling of D-glucosamine via N-2 to what are suggested to be Asp 53 and Glu 49 side chain carboxylates by analogy with previous side chain carboxylate reactions in RNase but are probably a distribution among several of the >10 available carboxylates.[74] Treatment of the protein with hydroxylamine was assumed to have reversed concomitant tyrosine modification. A similar modification of α-chymotrypsin led to broad heterogeneity and the formation of mixtures bearing glucosaminyl residues on 1−7 of the available 17 carboxylates; long reaction times (18−21 h) led, in some cases, to up to 75% autohydrolysis.[188]

Hindsgaul and co-workers have described the use of diethyl squarate for the coupling of carbohydrates bearing amino-terminus spacer arms to amines in BSA (Scheme 9i).[189] 2-Chloroethyl-1-thioglycosides have also been used to indiscriminately alkylate protein amino and hydroxyl groups;[190] an essentially analogous technique using 2-bromoethylglycosides with short peptide sequences containing cysteine or homocysteine and no basic residues allowed some selectivity for S-alkylation.[191]

A novel high-temperature "baking" method has been described for the modification of proteins that are lyophilized with reducing oligosaccharides.[192] Remarkably, despite being heated with the carbohydrate in air at 95−120 °C for up to 40 min, both trypsin and an IgG antibody survived with little loss of biological activity. The results of tryptic digests and conjugate hydrolyses suggest that the mechanism of conjugation involves an Amadori rearrangement with protein lysines, which destroys the integrity of the reducing end residue.

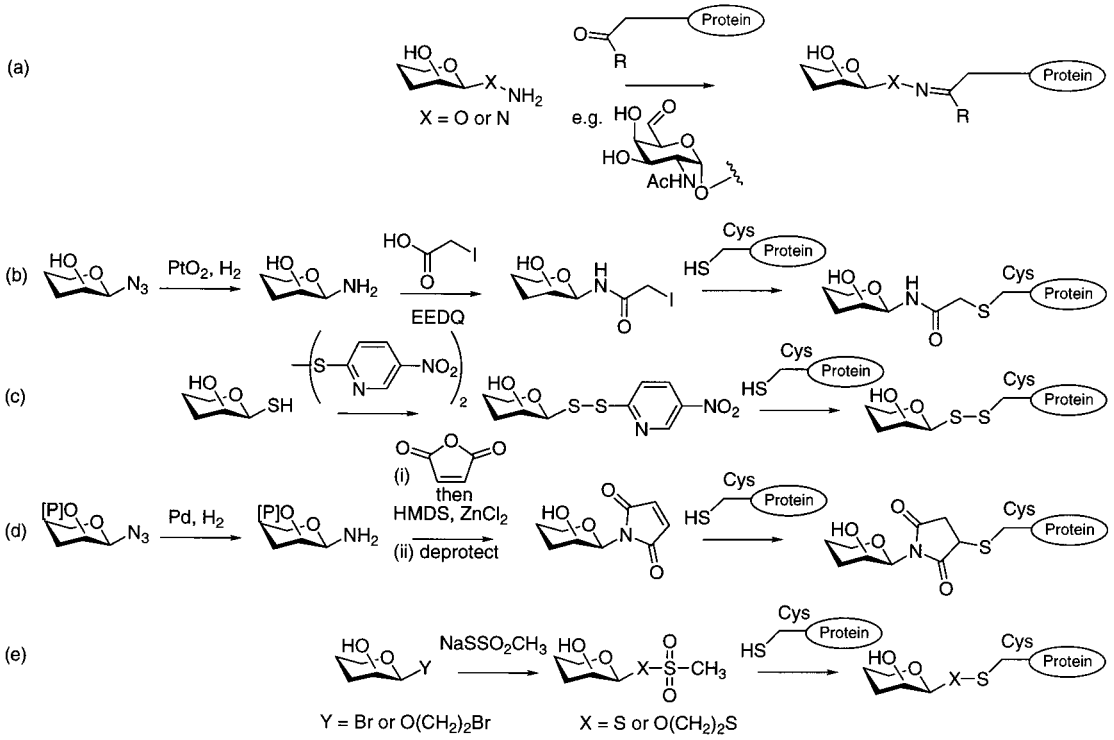
Although high levels of functionalization are thus easily accessible using the above methods, a lack of residue selectivity is often a drawback in the synthesis of well-defined conjugates. In addition, these techniques may alter the overall charge of the protein or destroy the cyclic nature of glycans introduced.

B. Chemoselective and Site-Specific Glycosylation

In an attempt to increase the selectivity and predictability of protein glycosylation, various novel approaches have been described, all of which exploit the chemoselectivities of different enzymatic and traditional methods. Bertozzi and co-workers have employed the selectivity of galactose oxidase to introduce an aldehyde tag to the C-6 of a GalNAc residue in the antimicrobial 19-residue peptide drosocin.[193] (This type of strategy has been christened chemoselective ligation, and its application to a wide range of bioconjugates has recently been reviewed.[139]) As Scheme 10a illustrates, this tag was then selectivity reacted with aminooxyglycosides to introduce further saccharides via the formation of an oxime, in a manner previously demonstrated for the conjugation of spacer-arm hydrazides with cell surface aldehydes.[194] That this nonnative glycopeptide shows biological activity comparable to the native form illustrates that certain unnatural linkages can in certain circumstances be tolerated. A similar chemoselective ligation approach has been applied to various glycopeptide syntheses (see section II.A.3).[140−142,144] Unfortunately, these approaches still require the linear construction of an initial glycopeptide and so suffer from some of the same disadvantages of protection and lability outlined in section II. However, they do hold the advantage that its application may be coupled with other methods for the introduction of a ketone tag (see below and section V.A). Thus, ligation of aminooxy-α-Gal with a keto-containing analogue of drosocin, allowed an alternative convergent glycosylation through oxime formation.[195] This nonnatively linked glycopeptide was almost as active in bioassays as linearly assembled naturally glycosylated drosocin. Similarly, the ready synthesis of hydrazides, thiosemicarbazides, and aminooxyglycosides of lactose, Gal, GalNAc, GlcNAc, and sLe^x allowed their attachment to the same ketone-containing peptide.[196] Bertozzi and co-workers have cleverly adapted Flitsch's iodoacetamide methodology[197] (see Scheme 10b and below) to convergently ligate Glc- and Gal-iodoacetamides to a C-3 thiol in a GalNAcα-Thr peptide.[198] The required thiol-protected glycosylated amino acid was constructed and assembled using SPPS to give a glycosylated 17-aa glycopeptide that corresponds to the N-terminal region of P-selectin glycoprotein (PSGL-1). Deprotection of the thiol and treatment with Glc- and Gal-iodoacetamide allowed the synthesis of a Glc/Galβ-(1,3)GalNAcα-Thr mimetic. No selectivity was observed if a cysteine was present in the peptide sequence.

In a sense, although chemoselective, these approaches have avoided the crux of glycoprotein synthesis (the formation of the carbohydrate−protein link) by relying on the presence of an existing glycan in the peptide/protein structure or an artificially introduced reactive group as a tag for reaction.

Scheme 10. Chemoselective and Site-Specific Glycoprotein Syntheses[a]



[a] (a) See refs 140−142,144,193,195,196,198. (b) See refs 197-199, 205. (c) See ref 200. (d) See ref 147. (e) See refs 67,73, 201−203.

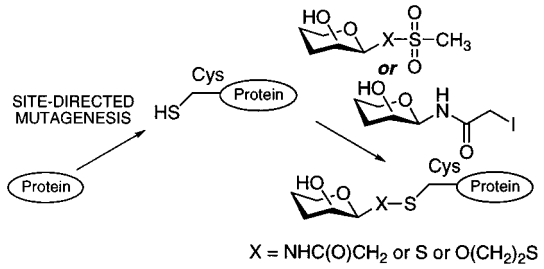
Similarly, the elegant enzymatic methods of Wong and co-workers and of Takegawa and co-workers described in section IV.C still require that a protein−*N*-glycan link be present from the start as a point of recognition for the enzymes concerned. Because these methods alter one glycan structure for another, they are therefore better described as glycoprotein remodeling (GPR) and they afford the glycoscientist no choice over the site of glycosylation.

Several methods have been proposed that tackle this central issue by exploiting specificity for functionality already found in proteins. Among the first approaches was that of Flitsch and co-workers, who reacted the α-iodoacetamide of *N*-acetyl-D-glucosamine with oxidized bovine serum albumin (BSA) to modify the single free cysteine present (Scheme 10b).[197] Later, this method was applied by Wong and co-workers to introduce chitotriose and a heptasaccharide stripped from the surface of horseradish peroxidase to BSA.[199] Boons has used dithiopyridyl methodology to make disulfide-linked BSA−*N*-acetyl-D-glucosamine constructs (Scheme 10c).[200] More recently, glycosylmaleimides have also been described as cysteine-specific glycosylation reagents for the functionalization of a cysteinyl-containing 11-aa peptide and BSA (Scheme 10d)[147] and are similar in concept to earlier spacer-arm maleimide-terminated reagents.[178]

C. Site-Selective Glycosylation

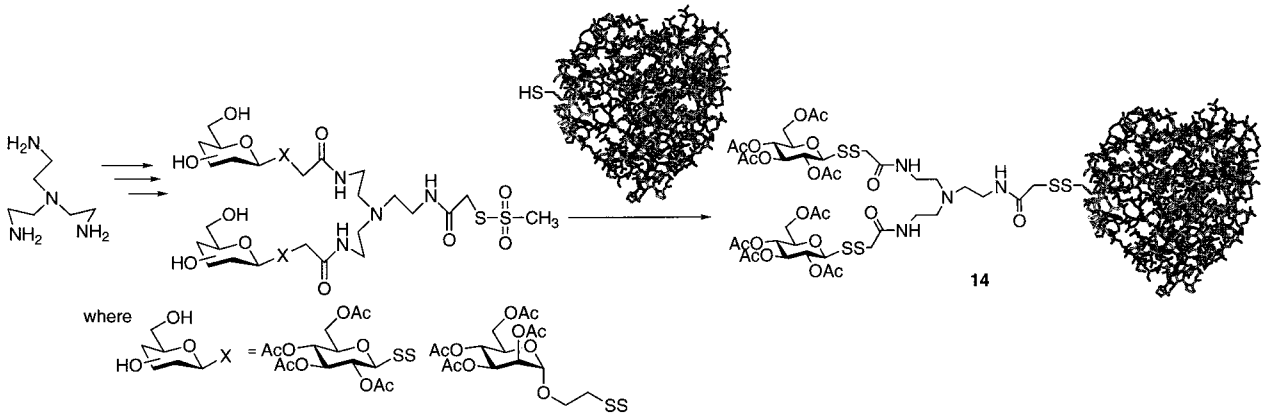
For full control of glycosylation both choice of site (site selectivity) and glycan are needed. A combined site-directed mutagenesis and chemical modification approach has solved this problem.[67,73,201−203] This approach provides a general method that allows both regio- and glycan-specific glycosylation of proteins. The strategy involves the introduction of cysteine as a chemoselective tag at preselected positions within a given protein and then reaction of its thiol group with glycomethanethiosulfonate (glycoMTS) reagents (Scheme 11). GlycoMTS reagents react specifically

Scheme 11. Site-Selective Glycoprotein Syntheses through a Combination of Site-Directed Mutagenesis and Chemical Modification[67,73,201−203,205]



and quantitatively with thiols[202] and allow the controlled formation of neutral disulfide linkages (Scheme 10e). Four sites on the representative serine protease subtilisin *Bacillus lentus* (SBL), which does not naturally contain cysteine, at different locations and of different characteristics were selected for mutation to cysteine in order to provide a broad test of the glycosylation methodology. Broad applicability with respect to the sugar moiety was evaluated by using the representative library of tethered or untethered, protected or deprotected, mono- and disaccharide methanethiosulfonates. The homogeneous glycoproteins formed allowed the first systematic determinations of the properties of novel glycoforms; detailed glycan structure−hydrolytic activity relationships for a library of 48 glycosylated forms of SBL were

Scheme 12. Glycodendriproteins May Provide a More Rapid Way of Synthesising Functional Glycoproteins Bearing Complex Multivalent Carbohydrate Structures[206]



determined.[203] Interestingly, internal glucosylation of one binding pocket increased k_{cat}/K_M 8.4-fold and the ratio of amidase to esterase activity, $(k_{cat}/K_M)_{esterase}/(k_{cat}/K_M)_{amidase}$ (E/A) 17-fold, relative to the unglycosylated enzyme.[73] These glycosylated enzymes also displayed enhanced utility as catalysts in peptide synthesis and allowed syntheses of dipeptides that were not possible using the unglycosylated catalyst.[72]

The use of a *p*-nitrophenylester linked thiogalactoside allowed an interesting site-selective glycosylation in 60% yield of one lysine among four potential lysines in a designed helix−loop−helix 42-aa polypeptide. The selectivity of the amide formation was confirmed by peptide digest mapping and was achieved by fine-tuning of conditions to allow the formation of an initial histidinyl-11-ester intermediate followed by intramolecular acyl transfer to the $i + 4$ lysine-15 to compete successfully with direct background amidations with all lysines (the rates are such that under these optimal conditions 94% of amidation should be intramolecular).[204] Some conformational effects of this glycosylation have also been noted.[82]

In an impressive, recent example of site-selective glycosylation through combined mutagenesis and modification, Flitsch and co-workers have used their iodoacetamide methodology (Scheme 10b) to glycosylate three cysteine mutants of erythropoietin (EPO) (Scheme 11).[205] The three natural Asn glycosylation sites chosen (sites 24, 38, and 83) have been shown previously to be critically N-glycosylated with the minimal N-linked oligosaccharide core for EPO to have activity; therefore, the introduction of iodoacetamide-GlcNAc to the N24C, N38C, and N83C mutants is an important starting step along this road. Reactions of EPO mutants were monitored by ESMS, and although only proceeding to approximately 60% conversion with 500-fold excess and after 24 h reaction time, it was possible to purify glycosylated EPO from unglycosylated EPO using a lectin-affinity column. Under these forcing conditions, additional nonspecific glycosylation of histidine residues (probably in the His10 tag) was also observed, thereby indicating that such glycosyliodoacetamides may not be absolutely selective; this lack of selectivity was circumvented by modification in the presence of excess imidazole. It was possible to confirm glycosylation of N83C unambiguously by proteolytic digest. It should be noted that, as for BSA, the three disulfide bonds in these EPO mutants were untouched by glycosylation of the single free cysteine.

More recently, the glycoMTS method (Scheme 10e) has also allowed the synthesis of the first examples of a homogeneous protein bearing symmetrically branched multivalent glycans in which both the site of glycosylation and the structure of the glycan introduced have been predetermined.[206] **14** (Scheme 12) represents the first of a new class of glycoconjugate: the glycodendriprotein. Since terminal glycan residues are typically the most important in glycoprotein interactions, this may provide a more rapid way of synthesizing functional glycoproteins bearing complex multivalent carbohydrate structures than the total synthesis of entire glycan structures.

IV. Enzymatic Glycoprotein and Glycopeptide Synthesis

The two distinct sets of strategies outlined in the two previous sections, linear assembly or convergent glycosylation, are equally available to enzyme-catalyzed techniques.

A. Elaboration of Glycans

Enzyme-catalyzed techniques for elaboration of existing glycans on glycoamino acids/peptides/proteins have proved particularly successful. One of the very first examples of the use of glycosyltransferases in glycoprotein synthesis was demonstrated by Paulson and co-workers who used a sialyltransferase and CMP-*N*-Ac-neuraminic acid to restore 95% of the sialic acids to a fully desialylated protein.[207] Enzymatic approaches have even been used to directly modify cell surface proteins.[208] The use of a milk fucosyltransferase that displays a broad substrate specificity allowed modification of glycans by adding fucosyl residues bearing a range of substituents at C-6.

Glycosyltransferases are often used to elaborate glycopeptide structures either prior to or after linear assembly (see section II.B.3). Indeed, highly glycosylated glycopeptide building blocks, such as **7**, for use in linear glycopeptide assembly were first made accessible through the use of glycosyltransferases.[164,165] Much of this work has strong resonance with the use of glycosyltransferases in oligosaccharide synthesis.[39,209−214]

Glycosyltransferase-mediated elaboration as a chemoselective technique is particularly useful for challenging structures rich in functionality such as sulfopeptides,[166] although strategic considerations are important with regard to the often stringent substrate specificity of glycosyltransferases. The presence of sulfation in a PSGL-1 octapeptide made this a particularly difficult glycopeptide to elaborate, and alternative glycosyltransferases more tolerant than those used previously had to be found. It should be noted that enzyme availability is sometimes a stumbling block in such techniques; in particular, the number of readily available branching GlcNAc-transferases is particularly limited. In an excellent display of the power of glycosyltransferases in glycopeptide elaboration (Scheme 7), the sulfated N-terminal domain of PSGL-1 has been synthesized on a small scale and characterized by HPLC and MS for two glycoforms that crucially differ in their binding to P-selectin as a result of only a subtle alteration of internal glycan structure.[167] This was achieved by isolation of the required six glycosyltransferases and one sulfotransferase and allowed **9** and **10** to be synthesized from a linearly SPPS-assembled 23-aa monosaccharide-bearing glycopeptide.

In a rare early use of enzymatic glycosylation on the solid phase, the sLe^x-Asn-Phe dipeptide was synthesized using aminopropylsilica as a support and subsequently was cleaved from a glycine linker by the peptidase chymotrypsin.[215] In this context, it is interesting to note that the ability of glycosyltransferases to modify glycans on glycopeptides is greatly influenced by the peptide backbone and the sites that are glycosylated within it.[216] For example, prolines, negatively charged residues, and the presence of a disaccharide all inhibited the glycosylation of an adjacent GlcNAc-Thr residue.

B. Trimming of Glycans

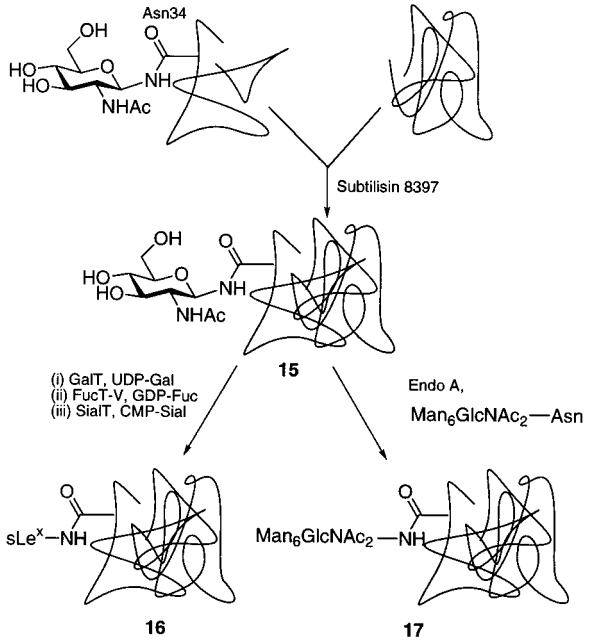
It is also possible to purify mixtures of glycoforms through selective enzymatic degradation of unwanted glycoforms.[217] Endoglycosidase-mediated wholesale trimming of glycan structures also plays a role as the first step in "glycoprotein remodeling" techniques (see section IV.C). The glycosylation of antibodies affects activity and function,[218] and their manipulation often provides some good case studies in glycoprotein synthesis methods. For example, terminal deglycosylation of human IgGs with β-galactosidase exposed GlcNAc residues and thereby introduced interactions with MBP.[219]

An in vitro enzyme cycle that trims, using a specific glucosidase, and adds, using a glucosyltransferase, a Glcα(1,3) unit to the tip of the glucosylated branch in the initial N-linked glycoprotein glycan core has been created.[220] This cycle plays a critical role in binding to lectin chaperones, calnexin, and calreticulin during folding in the endoplasmic reticulum (ER).[8]

C. Alteration of Glycans (Glycoprotein Remodeling GPR)

The combined trimming of existing glycan structures followed by elaboration to alternative ones has become known as "glycoprotein remodeling" (GPR). Key to this technique and as a result of the difficulty in making the glycan−protein link, at least one glycan must remain to serve as a tag for specific elaboration. For example, a synthesis of a single unnatural glycoform of ribonuclease B (RNase-B) **16** was achived by endoH degradation down to a single GlcNAcβ-Asn followed by elaboration with the well-established sequential system of galactosyltransferase, fucosyltransferase, and sialyltransferase to construct an sLe^x glycoform (Scheme 13).[221] Takeg-

Scheme 13. Glycoprotein Remodeling of RNase-B Using Glycosyltransferases[221] or Endoglycosidases[223]



awa and co-workers have applied endoglycosidase-mediated transglycosylation (the synthetic utility of which was demonstrated using endoF and endoH as early as 1986[222]) to the same partially deglycosylated RNase-B **15** in the synthesis of the $Man_6GlcNAc_2$ glycoprotein **17**.[223] EndoA-catalyzed transglycosylation also allowed the addition of $Man_9GlcNAc$ from naturally derived $Man_9GlcNAc_2$-Asn onto a chemically synthesized native N-link GlcNAc-Asn containing pentapeptide,[224] and a nonnative C-linked analogue[225] in yields of 25−26%.[226] This transglycosylation activity of endoA is usefully improved by the use of partial organic solvent systems, such as 35% aqueous acetone.[227] Such EndoA transglycosylation has also been cleverly combined with chemical glycoprotein synthesis techniques.[228] Thus, transglycosylation using $Man_9GlcNAc_2$-Asn as a donor onto *p*-isothiocyanatophenyl-β-D-Glc as an acceptor gave a high mannose isothiocyanate $Man_9GlcNAcGlc$ reagent that was used to glycosylate the lysines of RNase-A, lysozyme, and α-lactalbumin. EndoM endoglycosidase shows a fairly broad substrate specificity toward a variety of oligosaccharides in transfers to GlcNAc-bearing peptides (up to pentapeptides) but only in yields of up to 20%.[229,230] Complex-type glycans were transferred better than high-mannose-type. That synthetic yields are inversely related to the ability

to hydrolyze indicates that post-transglycosylation hydrolysis of product may be an important determining factor in overall yield. EndoM also catalyzes the transfer of $(NeuAcGalGlcNAcMan)_2ManGlcNAc$ to an octapeptide bearing a single GlcNAc in HPLC-determined yields of approximately 10%[231] and allows the synthesis of two different glycoforms of substance P, an undecapetide,[232] all formed using the same combined SPPS and transglycosylation strategy. The use of a dimethylphosphinothioic anhydride coupling (which avoided the need for protection of the glycan) and Ag-catalyzed condensation of an N-terminal fragment thioester with a C-terminal portion (which required only side chain Lys and Cys protection) was coupled with endoM-catalyzed transglycosylation to readily yield a C-terminal sequence of eel calcitonin in a nonnatural glycoform that bears a disialo biantennary-type undecasaccharide: CSN[(NeuAc-Gal-GlcNAc-Man)$_2$Man-GlcNAc-GlcNAc]LSTCVLGKS-GELHKLGTYPRTDVGAGTP-NH_2.[233]

Interestingly, nature employs a similar mechanism in the case of *Trypanosoma cruzi*, the protozoan that causes Chagas' disease. This parasite does not synthesize sialic acid but instead expresses a transsialidase that catalyzes the transfer of sialic acid from glycoconjugates found in the host to its own surface proteins. The resulting sialylated glycoproteins are then bound by host sialic acid binding receptors, thereby allowing cellular invasion.[234]

D. Enzymatic Formation of the Glycan–Protein/Peptide Link

An attractive approach to enzymatic glycoprotein synthesis is to exploit the enzymes responsible for the formation of the sugar-protein link in N-linked glycoprotein biosynthesis.[75] The enzyme responsible, oligosaccharyltransferase (OST), co-translationally transfers a high mannose core oligosaccharide from a fatty acid pyrophosphate carrier to the side chain amide of an asparagine (Asn) residue in the consensus sequence Asn-X-Thr/Ser of the nascent glycoprotein, although other sequences are also rarely glycosylated (e.g., Asn-Ala-Cys).[235] The use of this enzyme in isolated form in in vitro glycoprotein synthesis has, however, met with only modest success. While transfer of carbohydrates to a 17-residue peptide containing an unusual Asn-Asn-Thr-Ser sequence was possible, direct glycan transfer to RNase-A failed.[236] In addition, transfer to sequences in which X = Pro are not possible, and those in which X = Trp, Asp, Glu, Leu are inefficient.[237] Thus, such site-specific glycosylation of proteins still remains an elusive goal, and as a result, there is still no truly general enzymatic method for the synthesis of homogeneous glycoproteins. Interestingly, an endogalactosaminidase has recently been used to transfer Galβ(1,3)GalNAcα to the side chain hydroxyl of a serinyl residue in a hexapeptide.[238] Furthermore, the use of a microbial transglutaminase to transamidate the side chain γ-carboxamide group in the dipeptide Z-Gln-Gly with $-O(CH_2)_3S(CH_2)_2NH_2$ glycosides (shorter spacers were unsuccessful) to form an unnaturally N-linked glycopeptide has been described.[239,240] Both of these methods hint that general enzyme-catalyzed strategies for forming some key glycan–protein linkages may not be far off.

E. Glycopeptide Ligation

Subtilisin peptidases have been elegantly used to catalyze the synthesis of glycopeptides,[152,153] although the natural specificity of these enzymes has limited these peptide ligations to those in which the glycosylated residues are typically at least one residue distant (peptide site P_2,P_3... or P_2',P_3'...) from the amide bond formed. Thus, while ligation of Z-Gly-OBz with H-Gly-Ser(Ac$_3$GlcNAcβ)-NH_2 was successful, no yield of product was obtained with H-Ser(Ac$_3$-GlcNAcβ)-NH_2. The use of this ligation method coupled with other enzyme-mediated strategies culminated in a truly elegant synthesis of a single unnatural glycoform of ribonuclease B (RNase-B) **16** using a protease-catalyzed ligation of fragments of the protein backbone, including a fragment containing a single GlcNAcβ-Asn residue, followed by glycosyltransferase-catalyzed elaboration reactions of that glycan (Scheme 13).[221]

V. Molecular and Cell Biological Techniques

In vivo methods that alter the natural machinery of glycosylation offer promising opportunities.[241–243] Although, prokaryotic, e.g., bacterial systems, do not typically glycosylate proteins, the use of eukaryotic systems can circumvent this problem. Unfortunately, as yet, this leads to heterogeneous products.[244] The task is made difficult by the daunting array of biosynthetic glycosylation products and thus the corresponding array of pathways that need to be controlled or adapted.[245] These pathways, the levels of enzymes that drive them and their activities, differ subtly according to species, cell type, and protein. By expression of a particular glycoprotein in one organism, glycosylation patterns may arise that are different from those found by expression of the same enzyme in another. In this way glycosylation patterns may be guided in a particular direction. Thus, expression in, for example, plants[246] or mice[247] may allow the production of patterns that are similar but subtly altered compared to those in mammalian systems. The use of yeast to study the N-linked biosynthetic pathway and associated congenital diseases has been reviewed.[248] However, as yet, such biosynthetic "glycosylation engineering" still also produces mixtures.

A. Biosynthesis Augmentation

Broad changes in glycosylation patterns can be achieved through "glycosylation engineering", which regulates levels of the glycosyltransferases involved in post-translational elaboration of protein-linked glycans. For example, cell lines or cultures in which extra glycosyltransferase-expressing genes have been introduced may be used to enhance the presence of particular sugars in glycan structures. The addition of a sialyltransferase to a Chinese hamster ovary (CHO) cell line resulted in the increased "misglycosylation" of N-linked glycoproteins to give glycans bearing α(2–6)Gal- instead of α(2–3)Gal-linked sialic

acid terminated residues.[249] Similarly, tetracycline-regulated expression of the GlcNAc-transferase involved in "bisecting" (i.e., glycosylation of OH-4 of the β-Man residue in N-linked glycans) the core of N-linked glycans in CHO cells also expressing an antitumor IgG created an increase in bisected glycoforms from 25% to 50%, thereby increasing the cytotoxicity of the IgG.[250] Transfection of CHO cells with GlcNAc-transferase and fucosyltransferase activity using either two separate plasmids vectors[251] or one polycistronic vector[252] allows the construction of sLe^x on PSGL-1. The use of one plasmid prevented disparate expression levels.

Baculovirus vectors allow introduction of mammalian glycosyltransferases into insect cell systems that normally produce N-linked glycoproteins with limited glycan diversity.[253] For example, introduction of human GlcNAc-transferase I allows a high degree of further processing that is normally lacking in insect cells;[254] similarly, early expression of bovine β1,4-galactosyltransferase gave galactosylated *N*-glycans.[255] Plants have also proved to be suitable hosts; again, additional β1,4-galactosyltransferase activity is the key to altering plant N-glycosylation patterns toward those more like mammalian systems.[256] Thus, expression of human β1,4-galactosyltransferase in tobacco plants resulted in N-linked glycoproteins of which 15% bore terminal β1,4-galactosyl residues. Excitingly, crossing of this transgenic plant with one expressing mouse antibody ("plantibody") allowed in planta glycosylation engineering to produce galactosylated antibodies.

The prospects are also good for the glycosylation of larger biomolecular complexes by taking advantage of the often-relaxed specificities of biosynthetic pathways. Indeed, the use of an unnatural *N*-levulinoylmannosamine (Man-Lev) as a precursor in preference to the natural precursor ManAc can be achieved simply by feeding it to cells, and this has allowed the introduction of a unique ketone tag into sialic acid residues found at cell surfaces.[257] This allowed the selective introduction of further glycans through reaction with aminooxy and hydrazide-functionalized carbohydrates to form imines and hydrazones according to the manner described in sections II and III (Scheme 10a).[258] This strategy is similar to one previously applied to aldehydes introduced chemically to cell surfaces.[194] In a similar manner, neural cell surfaces have been also engineered by introducing an unnatural *N*-propanoylneuraminic acid precursor.[259] There is also good evidence that in certain cases the level of glycosylation is also influenced by levels of the intracellular nucleotide−mono/diphosphate donor substrates for glycosyltransferases that may in turn be increased simply by increasing the concentration of their biosynthetic precursors in incubation media. Thus, increasing glucosamine concentration increased GlcNAc-UDP levels in CHO cells and hence levels of bisecting GlcNAc structures.[260]

The powerful Man-Lev technique was also used to display biotin on cell surfaces.[261] This "cell surface engineering" utilized a biotin hydrazide to give cells that were far more readily transfected by adenovirus when pretreated with an avidin-to-(anti adenovirus antibody) conjugate. In addition, an aminooxy-functionalized Eu^{3+} complex MRI contrast reagent was constructed and localized through ligation.[262] The use of peracetylated *N*-azidoacetylmannosamine as a precursor allowed the cell surface display of azidosialic acids through hydrolysis of the acetate groups and then biosynthetic processing.[263] Staudinger-type reaction with a biotinylated phosphine ester results in intramolecular amide formation and hence biotinylation of the cell surface. The great power of this method is that azides and phosphines are not normally found in nature and therefore are unlikely to react in the absence of each other. This method, therefore, could allow *intra*cellular ligation as well. A traceless variant that extrudes the phosphine oxide formed has also been published to allow amide formation.[264] Very recently, a salvage pathway in CHO cells has allowed the incorporation of a 2-keto analogue of GalNAc into cell surface glycoprotein but not a 2-keto analogue of GlcNAc.[265]

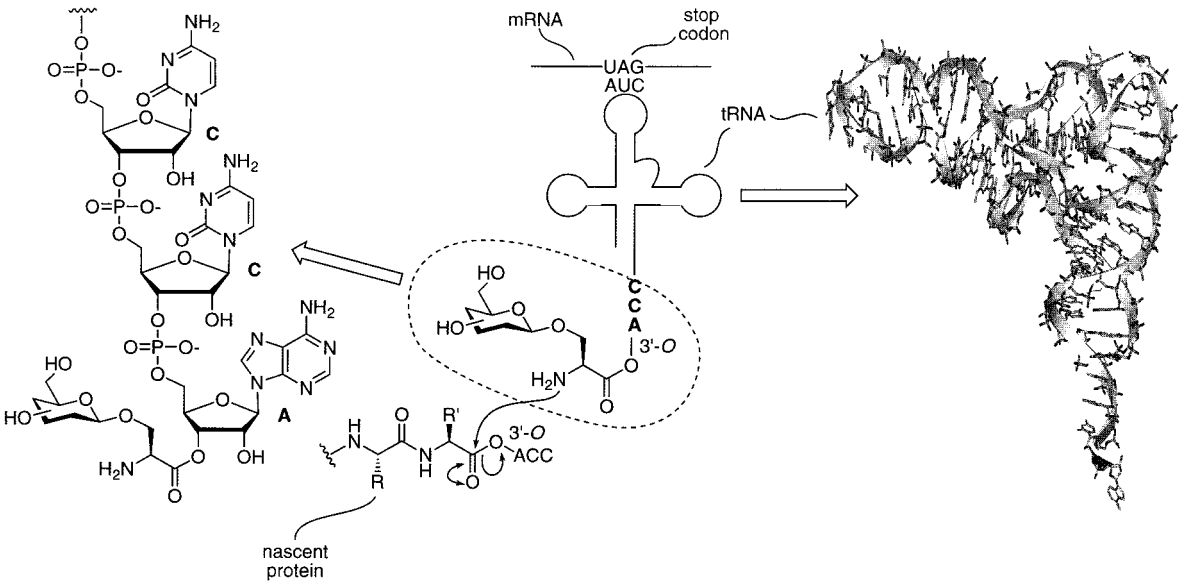
B. Biosynthesis Inhibition

The inhibition of enzymes involved in the biosynthesis[75] of glycoproteins offers an alternative way of controlling their structure. For example, the glycosyltransferase inhibitor tunicamycin inhibits the synthesis of the lipid-linked pyrophosphate oligosaccharide precursor that is used as a glycosyl donor in the formation of N-linked glycoproteins.[266] The resulting lack of donor prevents formation of Asn-linked glycans and results in only O-glycosylated proteins. Less drastic inhibition of later trimming steps, which are mediated by glycosidases in the ER and Golgi, can be used to create smaller than natural ranges of Asn-linked glycoforms rather than none at all. For example, the use of the glucosidase inhibitor *N*-butyldeoxynojirimycin (NBDNJ) resulted in a reduction in the number of glycoforms of the HIV surface protein gp120 that were produced from more than 100 to 3.[5] NBDNJ shuts down the early glucosidase trimming in HIV glycoprotein biosynthesis and results in poorly processed *N*-glycans of gp120.[267] Inhibition of trimming glycosidases also alters the function of IgGs by altering the glycosylation pattern.[268] Swainsonine, a branch-trimming α-mannosidase II inhibitor, reduces the formation of the GlcNAcβ(1,6) branch in N-linked glycans.[269] Mutation of the gene that encodes for α-mannosidase II in mice leads to lupus-like automimmune disease, thereby showing the importance of N-glycosylation.[270]

C. Noncoded Methods

Hecht[271,272] and Schmidt[273] have both proposed the adoption of in vitro use of misacylated tRNAs in nonsense codon suppression read-through techniques[274] to exploit the natural mechanism of protein biosynthesis (translation). Unlike natural protein glycosylation, which occurs co- and post-translationally, this method requires the synthesis of ^{AUC}tRNA acylated with glycosylated amino acids (Scheme 14). In this way GlcβSer was incorporated in place of Ser at position 286 of firefly luciferase.[271,272] Schmidt and Wieland have also described the preparation of a

Scheme 14. Glycoprotein Synthesis Using Misacylated tRNAs in Non-Sense Codon Suppression Read-Through[271−273]



hARF protein altered to contain GlcNAc(Ac)$_3$-α-Ser instead of a Lys in this way.[273] To this end, Hecht and co-workers have recently demonstrated the synthesis of tRNAs acylated with Gal, Glc, Man or GlcNAc-Ser, albeit in very low overall yields due to difficulties in achieving efficient acylation and deprotection steps.[275] Once it is a routine method, it will be interesting to see how glycoproteins made in this way through pretranslational glycosylation will differ from natural glycoproteins, which are co-/post-translationally glycosylated especially given the apparent role of glycosylation in correct protein folding. It is also interesting too to see the further possibilities for combining read-through techniques with those of chemical glycoprotein synthesis. For example, Schulz and co-workers have exploited the read-through technique to introduce unnatural ketone "handles" into proteins.[276] A combination of such a step with, for example, the oxime methodology of Bertozzi and co-workers (Scheme 10a)[193] would create another potential site-selective glycosylation technique.[280]

D. Expressed Protein Ligation

Expressed protein ligation (Scheme 8) has been used to incorporate variously modified cysteines at the C terminus of bacterially expressed mannan-binding protein (MBP) including Cys-Asn(GlcNAcβ).[171] Thus, MBP was expressed in *E. coli* as a fusion to the N terminus of a widely used intein from *Saccharomyces cerevisiae*; this intein also bears a chitin-binding domain at its C terminus for purification. Once expressed, this portion self-spliced the binding domain, and the resulting thioester used in native chemical ligation.

Bertozzi and co-workers have also used expressed protein ligation to construct two model nonnatural glycoproteins:[172] a 42-aa GlyCAM-1 central fragment stretch that is not glycosylated in full GlyCAM-1 was also expressed as an intein−chitin binding domain fusion protein in bacterial culture and purified on chitin beads. This was released as a C-terminal thioester and then coupled with a 15-aa glycopeptide through native chemical ligation. A second unglycosylated 77-aa domain was also ligated to the same glycopeptide in this way.

VI. Summary and Perspective

This review has illustrated the remarkable diversity of techniques available for the construction of glycoproteins. The complexity of synthesis is still an issue but by judicious and combined use of some of the complementary techniques described in this review, practical syntheses of complex glycoproteins will be achieved.[12,168] Now that semisynthesis is possible[172] using some of the variety of techniques described above and total synthesis from glycopeptides may soon be a realistic prospect,[162] we will have to ask ourselves which method should we employ to obtain *truly natural* glycoproteins. Since glycosylation has an important role in correct protein folding, it is unlikely to be enough to "just make it", and in this regard the timing of glycosylation and assembly strategy in any synthesis may too become a vital issue.

As well as acting as vital probes of carbohydrate communication, these methods may be exploited for great biomacromolecular structural diversity. Thus, in addition to the preparation of naturally occurring glycoprotein structures, we may now also start to use glycosylation on sites that are not normally glycosylated as tools for creating novel protein function and architecture. For example, glycosylation of enzymes allows the development of novel synthetic catalysts; site-selective glycosylation of their binding pockets can broaden synthetic utility.[72]

It is a sobering thought that while the protein backbone sequence can be confidently predicted through genomics, since protein glycosylation occurs co- and post-translationally, the latter is only known empirically. Only by examining the protein after glycosylation and relating its properties to its struc-

ture (glycomics[62]) can we begin to understand some of the relationships and principles. Some striking and intriguing processes mediated by protein glycosylation are emerging. It appears that O-GlcNAc glycosylation plays a role in a yin−yang type relationship with protein phosphorylation in signal transduction[277] and protein glycosylation has been heavily implicated as a potential trigger in prion diseases.

Finally, in order that the application of glycoconjugates in therapeutic strategies becomes more widespread, certain features must be addressed. Any of the glycoprotein structures described above represent potential sources of immunogenicity, and hence, much attention has been paid to the development of biocompatible polymers as alternatives. By building on the advances of glycoprotein applications, certain large (>100kDa) biocompatible polymers may offer all the advantages of glycoproteins but with reduced immunogenicity.[278] Unfortunately many glycoproteins are also poorly characterized, and so future efforts must also focus on homogeneity as a goal. The time is rapidly passing when mixtures of glycoforms will be viewed as "pure" by regulatory authorities just because they have the same peptide backbones.[44] In this context, the precision and rigor of chemistry have a crucial role to play. Collaborations between those making glycoproteins and those applying them will no longer be viewed as multi- or interdisciplinary; they are part of a new unified discipline intent on making exciting, well-defined molecules with enormous potential for the treatment of disease, intent on the elucidation of biochemical and physiological mechanisms, and that is able to powerfully fine-tune protein properties.

VII. Note Added after ASAP Posting

Squarate coupling (ref 189) was in fact first used for glycoprotein synthesis by Tietze and co-workers, see: Tietze, L. F.; Schröter, C.; Gabius, S.; Brinck, U.; Goerlach-Graw, A.; Gabius, H.-J. *Bioconjugate Chem.* **1991**, *2*, 148. I am grateful to readers for bringing this to my attention.

VIII. References

(1) Sharon, N.; Lis, H. *Essays Biochem.* **1995**, *30*, 59.
(2) Lasky, L. A. *Annu. Rev. Biochem.* **1995**, *64*, 113.
(3) Weis, W. I.; Drickamer, K. *Annu. Rev. Biochem.* **1996**, *65*, 441.
(4) Varki, A. *Glycobiology* **1993**, *3*, 97.
(5) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683.
(6) Rudd, P. M.; Elliot, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291*, 2370.
(7) Helenius, A. *Mol. Biol. Cell* **1994**, *5*, 253.
(8) Trombetta, E. S.; Helenius, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 587.
(9) Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364.
(10) Opdenakker, G.; Rudd, P. M.; Ponting, C. P.; Dwek, R. A. *FASEB J.* **1993**, *7*, 1330.
(11) Davis, B. G. *J. Chem. Soc., Perkin Trans. 1* **1999**, 3215.
(12) Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357.
(13) Kunz, H. *Angew. Chem., Int. Ed. Engl.* **1987**, *26*, 294.
(14) Paulsen, H. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 823.
(15) Meldal, M. *Curr. Opin. Struct. Biol.* **1994**, *4*, 710.
(16) Meldal, M.; Bock, K. *Glycoconjugate J.* **1994**, *11*, 59.
(17) Garg, H. G.; von dem Bruch, K.; Kunz, H. *Adv. Carbohydr. Chem. Biochem.* **1994**, *50*, 277.
(18) Meldal, M.; St. Hilaire, P. M. *Curr. Opin. Chem. Biol.* **1997**, *1*, 552.
(19) Taylor, C. M. *Tetrahedron* **1998**, *54*, 11317.
(20) Arsequell, G.; Valencia, G. *Tetrahedron: Asymmetry* **1997**, *8*, 2839.
(21) Arsequell, G.; Valencia, G. *Tetrahedron: Asymmetry* **1999**, *10*, 3045.
(22) Osborn, H. M. I.; Khan, T. H. *Tetrahedron* **1999**, *55*, 1807.
(23) Herzner, H.; Reipen, T.; Schultz, M.; Kunz, H. *Chem. Rev.* **2000**, *100*, 4495.
(24) Mizuno, M. *Trends Glycosci. Glycotechnol.* **2001**, *13*, 11.
(25) St Hilaire, P. M.; Meldal, M. *Angew. Chem., Int. Ed.* **2000**, *39*, 1162.
(26) Seitz, O. *ChemBioChem* **2000**, *1*, 214.
(27) Muir, T. W.; Dawson, P. E.; Kent, S. B. H. *Methods Enzymol.* **1997**, *289*, 266.
(28) Wilken, J.; Kent, S. B. H. *Curr. Opin. Biotechnol.* **1998**, *9*, 412.
(29) Kochendoerfer, G. G.; Kent, S. B. H. *Curr. Opin. Chem. Biol.* **1999**, *3*, 665.
(30) Robertson, N.; Ramage, R. *J. Chem. Soc., Perkin Trans. 1* **1999**, 1015.
(31) Borgia, J. A.; Fields, G. B. *TIBTECH* **2000**, *18*, 243.
(32) Paulsen, H. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 155.
(33) Boons, G. J. *Contemp. Org. Synth.* **1996**, *3*, 173.
(34) Boons, G. J. *Tetrahedron* **1996**, *52*, 1095.
(35) Toshima, K.; Tatsuta, K. *Chem. Rev.* **1993**, *93*, 1503.
(36) Whitfield, D. M.; Douglas, S. P. *Glycoconjugate J.* **1996**, *13*, 5.
(37) Davis, B. G. *J. Chem. Soc., Perkin Trans. 1* **2000**, 2137.
(38) Sears, P.; Wong, C.-H. *Science* **2001**, *291*, 2344.
(39) Koeller, K. M.; Wong, C.-H. *Chem. Rev.* **2000**, *100*, 4465.
(40) Stowell, C. P.; Lee, Y. C. *Adv. Carbohydr. Chem. Biochem.* **1980**, *37*, 225.
(41) Lee, Y. C.; Lee, R. T. *Neoglycoconjugates: Preparation and Applications*; Academic Press: San Diego, 1994.
(42) *Methods Enzymol.* **1994**, *242*, 257 (dedicated issue).
(43) Wong, S. Y. C. *Curr. Opin. Chem. Biol.* **1995**, *5*, 599.
(44) Bill, R. M.; Flitsch, S. L. *Chem. Biol.* **1996**, *3*, 145.
(45) Laine, R. A. *Glycobiology* **1994**, *4*, 759.
(46) Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *28*, 321.
(47) Janakiraman, M. N.; White, C. L.; Laver, W. G.; Air, G. M. A.; Luo, M. *Biochemistry* **1994**, *33*, 8172.
(48) Kiessling, L. L.; Pohl, N. L. *Chem. Biol.* **1996**, *3*, 71.
(49) Lee, Y. C.; Towsend, R. R.; Hardy, M. R.; Lönngren, J.; Arnarp, J.; Haraldsson, M.; Lönn, H. *J. Biol. Chem.* **1983**, *258*, 199.
(50) Dimick, S. M.; Powell, S. C.; McMahon, S. A.; Moothoo, D. N.; Naismith, J. H.; Toone, E. J. *J. Am. Chem. Soc.* **1999**, *121*, 10286.
(51) Corbell, J. B.; Lundquist, J. J.; Toone, E. J. *Tetrahedron: Asymmetry* **2000**, *11*, 95.
(52) Goebel, W. F.; Avery, O. T. *J. Exp. Med.* **1929**, *50*, 521.
(53) Kitamura, K.; Stockert, E.; Garinchesa, P.; Welt, S.; Lloyd, K. O.; Armour, K. L.; Wallace, T. P.; Harris, W. J.; Carr, F. J.; Old, L. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12957.
(54) Susaki, H.; Suzuki, K.; Ikeda, M.; Yamada, H.; Watanabe, H. K. *Chem. Pharm. Bull.* **1998**, *46*, 1530.
(55) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33*, 17.
(56) Kretzschmar, G.; Sprengard, U.; Kunz, H.; Bartnik, E.; Schmidt, W.; Toepfer, A.; Horsch, B.; Krause, M.; Seiffge, D. *Tetrahedron* **1995**, *51*, 13015.
(57) Kichler, A.; Schuber, F. *Glycoconjugate J.* **1995**, *12*, 275.
(58) Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. *Annu. Rev. Biochem.* **1988**, *57*, 785.
(59) Parekh, R. B.; Dwek, R. A.; Rudd, P. M.; Thomas, J. R.; Rademacher, T. W.; Warren, T.; Wun, T. C.; Hebert, B.; Reitz, B.; Palmier, M.; Ramabhadran, T.; Tiemeier, D. C. *Biochemistry* **1989**, *28*, 7670.
(60) Simanek, E. E.; McGarvey, G. J.; Jablonowski, J. A.; Wong, C. H. *Chem. Rev.* **1998**, *98*, 833.
(61) Yarema, K. J.; Bertozzi, C. R. *Curr. Opin. Chem. Biol.* **1998**, *2*, 49.
(62) Davis, B. G. *Chem. Ind.* **2000**, 134.
(63) Michael, A. *Nat. Biotechnol.* **1997**, *15*, 1233.
(64) Kitov, P. I.; Sadowska, J. M.; Mulvey, G.; Armstrong, G. D.; Ling, H.; Pannu, N. S.; Read, R. J.; Bundle, D. R. *Nature* **2000**, *403*, 669.
(65) Mulvey, G.; Kitov, P. I.; Marcato, P.; Bundle, D. R.; Armstrong, G. D. *Biochimie* **2001**, *83*, 841.
(66) Aplin, J. D.; Wriston, J. C. *CRC Crit. Rev. Biochem.* **1981**, *10*, 259.
(67) Davis, B. G.; Jones, J. B. *Synlett* **1999**, 1495.
(68) Hill, T. G.; Wang, P.; Huston, M. E.; Wartchow, C. A.; Oehler, L. M.; Smith, M. B.; Bednarski, M. D.; Callstrom, M. R. *Tetrahedron Lett.* **1991**, *32*, 6823.
(69) Wang, P.; Hill, T. G.; Wartchow, C. A.; Huston, M. E.; Oehler, L. M.; Smith, M. B.; Bednarski, M. D.; Callstrom, M. R. *J. Am. Chem. Soc.* **1992**, *114*, 378.
(70) Wartchow, C. A.; Wang, P.; Bednarski, M. D.; Callstrom, M. R. *J. Org. Chem.* **1995**, *60*, 2216.
(71) Wang, P.; Hill, T. G.; Bednarski, M. D.; Callstrom, M. R. *Tetrahedron Lett.* **1991**, *32*, 6827.
(72) Matsumoto, K.; Davis, B. G.; Jones, J. B. *Chem. Commun.* **2001**, 903.
(73) Lloyd, R. C.; Davis, B. G.; Jones, J. B. *Bioorg. Med. Chem.* **2000**, *8*, 1537.

(74) Baek, W. O.; Vijayalakshmi, M. A. *Biochim. Biophys. Acta* **1997**, *1336*, 394.
(75) Kornfield, R.; Kornfield, S. *Annu. Rev. Biochem.* **1985**, *54*, 631.
(76) Park, H.; Suzuki, T.; Lennarz, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11163.
(77) Imperiali, B.; O'Connor, S. E. *Curr. Opin. Chem. Biol.* **1999**, *3*, 643.
(78) Simanek, E. E.; Huang, D.-H.; Pasternack, L.; Machajewski, T. D.; Seitz, O.; Millar, D. S.; Dyson, H. J.; Wong, C.-H. *J. Am. Chem. Soc.* **1998**, *120*, 11567.
(79) Eisele, F.; Owen, D. J.; Waldmann, H. *Bioorg. Med. Chem.* **1999**, *7*, 193.
(80) Live, D. H.; Kumar, R. A.; Beebe, X.; Danishefsky, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12759.
(81) Bailey, D.; Renouf, D. V.; Large, D. G.; Warren, C. D.; Hounsell, E. F. *Carbohydr. Res.* **2000**, *324*, 242.
(82) Andersson, L. K.; Dolphin, G. T.; Kihlberg, J.; Baltzer, L. *J. Chem. Soc., Perkin Trans. 2* **2000**, 459.
(83) Tsuda, T.; Nishimura, S.-I. *J. Chem. Soc., Chem. Commun.* **1996**, 2779.
(84) Ashwell, G.; Harford, J. *Annu. Rev. Biochem.* **1982**, *51*, 531.
(85) Wadhwa, M. S.; Rice, K. G. *J. Drug Targeting* **1995**, *3*, 111.
(86) Hassan, N.; Wu, G. Y. *Nat. Med.* **1995**, *1*, 210.
(87) Wu, G. Y.; Wu, C. H. *Adv. Drug Delivery Rev.* **1993**, *12*, 159.
(88) Molema, G.; Jansen, R. W.; Visser, J.; Herdewijn, P.; Moolenaar, F.; Meijer, D. K. F. *J. Med. Chem.* **1991**, *34*, 1137.
(89) Meijer, D. K. F.; Molema, G.; Moolenaar, F.; deZeeuw, D.; Swart, P. J. *J. Controlled Release* **1996**, *39*, 163.
(90) Franssen, E. J. F.; Jansen, R. W.; Vaalburg, M.; Meijer, D. K. F. *Biochem. Pharmacol.* **1993**, *45*, 1215.
(91) Jansen, R. W.; Kruijt, J. K.; van Berkel, T. J. C.; Meijer, D. K. F. *Hepatology* **1993**, *18*, 146.
(92) Molema, G.; Meijer, D. K. F. *Adv. Drug Delivery Rev.* **1994**, *14*, 25.
(93) Simon, P. M.; Goode, P. L.; Mobasseri, A.; Zopf, D. *Infect. Immun.* **1997**, *65*, 750.
(94) Avery, O. T.; Goebel, W. F. *J. Exp. Med.* **1929**, *50*, 533.
(95) Francis, T.; Tillet, W. S. *J. Exp. Med.* **1930**, *52*, 573.
(96) Lemieux, R. U.; Bundle, D. R.; Baker, D. A. *J. Am. Chem. Soc.* **1975**, *97*, 4076.
(97) Sood, R. K.; Fattom, A.; Pavliak, V.; Naso, R. B. *Drug Discovery Today* **1996**, *1*, 381.
(98) Danishefsky, S. J.; Allen, J. R. *Angew. Chem., Int. Ed.* **2000**, *39*, 836.
(99) Ragupathi, G.; Park, T. K.; Zhang, S.; Kim, I. J.; Graber, L.; Adluri, S.; Lloyd, K. O.; Danishefsky, S. J.; Livingston, P. O. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 125.
(100) Toyokuni, T.; Singhal, A. K. *Chem. Soc. Rev.* **1995**, 231.
(101) Koganty, R. R.; Reddish, M. A.; Longenecker, B. M. *Drug Discovery Today* **1996**, *1*, 190.
(102) Barchi, J. J. *Curr. Pharm. Des.* **2000**, *6*, 485.
(103) Harding, C. V.; Kihlberg, J.; Elofsson, M.; Magnusson, G.; Unanue, E. R. *J. Immunol.* **1993**, *151*, 2419.
(104) Brady, R. O.; Barton, N. W. In *Lectins and Cancer*; Gabius, H.-J., Gabius, S., Eds.; Springer: New York, 1991.
(105) Takakura, Y.; Masuda, S.; Tokuda, H.; Nishikawa, M.; Hashida, M. *Biochem. Pharmacol.* **1994**, *47*, 853.
(106) Sakurai, K.; Miyazaki, K.; Kodera, Y.; Nishimura, H.; Shingu, M.; Inada, Y. *Glycoconjugate J.* **1997**, *14*, 723.
(107) Anzai, J.-I.; Kobayashi, Y.; Nakamura, N. *J. Chem. Soc., Perkin Trans. 2* **1998**, 461.
(108) Fields, S.; Winter, G. *Nature* **1981**, *290*, 213.
(109) Vliegenthart, J. F. G.; Casset, F. *Curr. Opin. Struct. Biol.* **1998**, *8*, 565.
(110) Shibata, S.; Takeda, T.; Natori, Y. *J. Biol. Chem.* **1988**, *263*, 12483.
(111) Wieland, F.; Heizer, R.; Schaefer, W. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 5470.
(112) Paul, G.; Lottspeich, F.; Wieland, F. *J. Biol. Chem.* **1986**, *261*, 1020.
(113) Messner, P.; Sleytr, U. B. *FEBS Lett.* **1988**, *228*, 317.
(114) DeBeer, T.; Vliegenthart, J. F. G.; Löffler, A.; Hofsteenge, J. *Biochemistry* **1995**, *34*, 11785.
(115) Nagai, K.; Ishizuka, I. *Trends Glycosci. Glycotechnol.* **1998**, *10*, 257.
(116) Nagai, K.; Ishizuka, I. *Trends Glycosci. Glycotechnol.* **1998**, *10*, 472.
(117) Gupta, R.; Birch, H.; Rapacki, K.; Brunak, S.; Hansen, J. E. *Nucleic Acids Res.* **1999**, *27*, 370.
(118) Marks, G. S.; Neuberger, A. *J. Chem. Soc.* **1961**, 4872.
(119) Teshima, T.; Nakajima, K.; Takahashi, M.; Shiba, T. *Tetrahedron Lett.* **1992**, *33*, 363.
(120) Danishefsky, S. J.; Hu, S.; Cirillo, P. F.; Eckhardt, M.; Seeburger, P. H. *Chem.—Eur. J.* **1997**, *3*, 1617.
(121) Unverzagt, C. *Carbohydr. Res.* **1998**, *305*, 423.
(122) Günther, W.; Kunz, H. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 1050.
(123) Handlon, A. L.; Fraser-Reid, B. *J. Am. Chem. Soc.* **1993**, *115*, 3796.
(124) Ratcliffe, A. J.; Konradsson, P.; Fraser-Reid, B. *J. Am. Chem. Soc.* **1990**, *112*, 5665.
(125) Nair, L. G.; Fraser-Reid, B.; Szardenings, A. K. *Org. Lett.* **2001**, *3*, 317.
(126) Kahne, D.; Walker, S.; Cheng, Y.; van Engen, D. *J. Am. Chem. Soc.* **1989**, *111*, 6881.
(127) Inazu, T.; Kobayashi, K. *Synlett* **1993**, 869.
(128) Mizuno, M.; Muramoto, I.; Kobayashi, K.; Yaginuma, H.; Inazu, T. *Synthesis* **1999**, 162.
(129) Komba, S.; Meldal, M.; Werdelin, O.; Jensen, T.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1999**, 415.
(130) Burger, K.; Kluge, M.; Fehn, S.; Koksch, B.; Hennig, L.; Müller, G. *Angew. Chem., Int. Ed.* **1999**, *38*, 1414.
(131) Allen, J. R.; Harris, C. R.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2001**, *123*, 1890.
(132) Baldwin, J. E.; Spivey, A. C.; Schofield, C. J. *Tetrahedron: Asymmetry* **1990**, *1*, 881.
(133) Cohen, S. B.; Halcomb, R. L. *Org. Lett.* **2001**, *3*, 405.
(134) Zhu, Y.; van der Donk, W. A. *Org. Lett.* **2001**, *3*, 1189.
(135) Jobron, L.; Hummel, G. *Org. Lett.* **2000**, *2*, 2265.
(136) Ohnishi, Y.; Ichikawa, M.; Ichikawa, Y. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1289.
(137) Debenham, S. D.; Debenham, J. S.; Burk, M. J.; Toone, E. J. *J. Am. Chem. Soc.* **1997**, *119*, 9897.
(138) Ichikawa, Y.; Nishiyama, T.; Isobe, M. *Synlett* **2000**, 1253.
(139) Lemieux, G. A.; Bertozzi, C. R. *TIBTECH* **1998**, *16*, 506.
(140) Cervigni, S. E.; Dumy, P.; Mutter, M. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1230.
(141) Zhao, Y.; Kent, S. B. H.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1629.
(142) Durieux, P.; Fernandez-Carneado, J.; Tuchscherer, G. *Tetrahedron Lett.* **2001**, *42*, 2297.
(143) Peri, F.; Dumy, P.; Mutter, M. *Tetrahedron* **1998**, *54*, 12269.
(144) Peluso, S.; Imperiali, B. *Tetrahedron Lett.* **2001**, *42*, 2085.
(145) Grandjean, C.; Rommens, C.; Gras-Masse, H.; Melnyk, O. *Angew. Chem., Int. Ed.* **2000**, *39*, 1068.
(146) Grandjean, C.; Gras-Masse, H.; Melnyk, O. *Chem.—Eur. J.* **2001**, *7*, 230.
(147) Shin, I.; Jung, H.-J.; Lee, M.-R. *Tetrahedron Lett.* **2001**, *42*, 1325.
(148) Hansen, P. R.; Olson, C. E.; Holm, A. *Bioconjugate Chem.* **1998**, *9*, 126.
(149) Meinjohanns, E.; Vargas-Berenguel, A.; Meldal, M.; Paulsen, H.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1995**, 2165.
(150) Meinjohanns, E.; Meldal, M.; Paulsen, H.; Dwek, R. A.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1998**, 549.
(151) Pohl, T.; Waldmann, H. *J. Am. Chem. Soc.* **1997**, *119*, 6702.
(152) Wong, C. H.; Schuster, M.; Wang, P.; Sears, P. *J. Am. Chem. Soc.* **1993**, *115*, 5893.
(153) Witte, K.; Seitz, O.; Wong, C. H. *J. Am. Chem. Soc.* **1998**, *120*, 1979.
(154) Barkley, A.; Arya, P. *Chem.—Eur. J.* **2001**, *7*, 555.
(155) Saha, U. K.; Roy, R. *Tetrahedron Lett.* **1995**, *36*, 3635.
(156) Roy, R.; Park, W. K. C.; Wu, Q.; Wang, S.-N. *Tetrahedron Lett.* **1995**, *36*, 4377.
(157) Kim, J. M.; Roy, R. *Tetrahedron Lett.* **1997**, *38*, 3487.
(158) Arya, P.; Kutterer, K. M. K.; Qin, H. P.; Roby, J.; Barnes, M. L.; Kim, J. M.; Roy, R. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1127.
(159) Ansfield, S. T.; Lansbury, P. T. *J. Org. Chem.* **1990**, *55*, 5560.
(160) Cohen-Ansfield, S. T.; Lansbury, P. T. *J. Am. Chem. Soc.* **1993**, *115*, 10531.
(161) Roberge, J. Y.; Beebe, X.; Danishefsky, S. J. *Science* **1995**, *269*, 202.
(162) Wang, Z.-W.; Zhang, X.; Visser, M.; Live, D.; Zatorski, A.; Iserloh, U.; Lloyd, K. O.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2001**, *40*, 1728.
(163) Roberge, J. Y.; Beebe, X.; Danishefsky, S. J. *J. Am. Chem. Soc.* **1998**, *120*, 3915.
(164) Unverzagt, C. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2351.
(165) Unverzagt, C. *Tetrahedron Lett.* **1997**, *32*, 5627.
(166) Koeller, K. M.; Smith, M. E. B.; Huang, R.-F.; Wong, C.-H. *J. Am. Chem. Soc.* **2000**, *122*, 4241.
(167) Leppänen, A.; Mehta, P.; Ouyang, Y.-B.; Ju, T.; Helin, J.; Moore, K. L.; van Die, I.; Canfield, W. M.; McEver, R. P.; Cummings, R. D. *J. Biol. Chem.* **1999**, *274*, 24838.
(168) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684.
(169) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *226*, 776.
(170) Marcaurelle, L. A.; Mizoue, L. S.; Wilken, J.; Oldham, L.; Kent, S. B. H.; Handel, T. M.; Bertozzi, C. R. *Chem.—Eur. J.* **2001**, *7*, 1129.
(171) Tolbert, T. J.; Wong, C.-H. *J. Am. Chem. Soc.* **2000**, *122*, 5421.
(172) Macmillan, D.; Bertozzi, C. R. *Tetrahedron* **2000**, *56*, 9515.
(173) Marcaurelle, L. A.; Bertozzi, C. R. *Chem.—Eur. J.* **1999**, *5*, 1384.
(174) Lee, Y. C.; Stowell, C. P.; Krantz, M. J. *Biochemistry* **1976**, *15*, 3956.
(175) Gray, G. R. *Arch. Biochem. Biophys.* **1974**, *163*, 426.
(176) Bernstein, M. A.; Hall, L. D. *Carbohydr. Res.* **1980**, *78*, C1.

(177) Zhang, J.; Yergey, A.; Kowalak, J.; Kovac, P. *Tetrahedron* **1998**, *54*, 11783.
(178) Ragupathi, G.; Koganty, R. R.; Qiu, D. X.; Lloyd, K. O.; Livingston, P. O. *Glycoconjugate J.* **1998**, *15*, 217.
(179) McBroom, C. R.; Samanen, C. H.; Goldstein, I. J. *Methods Enzymol.* **1972**, *28*, 212.
(180) Buss, D. H.; Goldstein, I. J. *J. Chem. Soc. C* **1968**, 1457.
(181) Quétard, C.; Bourgerie, S.; Normand-Sdiqui, N.; Mayer, R.; Strecker, G.; Midoux, P.; Roche, A. C.; Monsigny, M. *Bioconjugate Chem.* **1998**, *9*, 268.
(182) Taverna, R. D.; Langdon, R. G. *Biochim. Biophys. Acta* **1973**, *298*, 412.
(183) Arakatsu, Y.; Ashwell, G.; Kabat, E. A. *J. Immunol.* **1966**, *97*, 858.
(184) Ashwell, G. *Methods Enzymol.* **1972**, *28*, 219.
(185) Lonngren, J.; Goldstein, I. J.; Niederhuber, J. E. *Arch. Biochem. Biophys.* **1976**, *175*, 661.
(186) Rude, E.; Westphal, O.; Hurwitz, E.; Fuchs, S.; Sela, M. *Immunochemistry* **1966**, *3*, 137.
(187) Lebashov, A. V.; Rariy, R. V.; Martinek, K.; Klyachko, N. L. *FEBS Lett.* **1993**, *336*, 385.
(188) Jiang, K.-Y.; Pitiot, O.; Anissimova, M.; Adenier, H.; Vijayalakshmi, M. A. *Biochim. Biophys. Acta* **1999**, *1433*, 198.
(189) Kamath, V. P.; Diedrich, P.; Hindsgaul, O. *Glycoconjugate J.* **1996**, *13*, 315.
(190) Ticha, M.; Cerny, M.; Trnka, T. *Glycoconjugate J.* **1996**, *13*, 681.
(191) Bengtsson, M.; Broddefalk, J.; Dahmén, J.; Henriksson, K.; Kihlberg, J.; Lönn, H.; Srinivasa, B. R.; Stenvall, K. *Glycoconjugate J.* **1998**, *15*, 223.
(192) Boratynski, J.; Roy, R. *Glycoconjugate J.* **1998**, *15*, 131.
(193) Rodriguez, E. C.; Winans, K. A.; King, D. S.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1997**, *119*, 9905.
(194) Orr, G. A.; Rando, R. R. *Nature* **1978**, *272*, 722.
(195) Marcaurelle, L. A.; Rodriguez, E. C.; Bertozzi, C. R. *Tetrahedron Lett.* **1998**, *39*, 8417.
(196) Rodriguez, E. C.; Marcaurelle, L. A.; Bertozzi, C. R. *J. Org. Chem.* **1998**, *63*, 7134.
(197) Davis, N. J.; Flitsch, S. L. *Tetrahedron Lett.* **1991**, *32*, 6793.
(198) Marcaurelle, L. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2001**, *123*, 1587.
(199) Wong, S. Y. C.; Guile, G. R.; Dwek, R. A.; Arsequell, G. *Biochem. J.* **1994**, *300*, 843.
(200) Macindoe, W. M.; van Oijen, A. H.; Boons, G.-J. *Chem. Commun.* **1998**, 847.
(201) Davis, B. G.; Lloyd, R. C.; Jones, J. B. *J. Org. Chem.* **1998**, *63*, 9614.
(202) Davis, B. G.; Maughan, M. A. T.; Green, M. P.; Ullman, A. *Tetrahedron: Asymmetry* **2000**, *11*, 245.
(203) Davis, B. G.; Lloyd, R. C.; Jones, J. B. *Bioorg. Med. Chem.* **2000**, *8*, 1527.
(204) Andersson, L.; Stenhagen, G.; Baltzer, L. *J. Org. Chem.* **1998**, *63*, 1366.
(205) Macmillan, D.; Bill, R. M.; Sage, K. A.; Fern, D.; Flitsch, S. L. *Chem. Biol.* **2001**, *8*, 125.
(206) Davis, B. G. *Chem. Commun.* **2001**, 351.
(207) Paulson, J. C.; Hill, R. L.; Tanabe, T.; Ashwell, G. *J. Biol. Chem.* **1977**, *252*, 8624.
(208) Tsuboi, S.; Isogai, Y.; Hada, N.; King, J. K.; Hindsgaul, O.; Fukuda, M. *J. Biol. Chem.* **1996**, *271*, 27213.
(209) Wong, C. H.; Halcomb, R. L.; Ichikawa, Y.; Kajimoto, T. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 521.
(210) Kren, V.; Thiem, J. *Chem. Soc. Rev.* **1997**, *26*, 463.
(211) Ichikawa, Y. *Trends Glycosci. Glycotechnol.* **1997**, *9*, S47.
(212) Crout, D. H. G.; Vic, G. *Curr. Opin. Chem. Biol.* **1998**, *2*, 98.
(213) Palcic, M. M. *Curr. Opin. Biotechnol.* **1999**, *10*, 616.
(214) Wymer, N.; Toone, E. J. *Curr. Opin. Chem. Biol.* **2000**, *4*, 110.
(215) Schuster, M.; Wang, P.; Paulson, J. C.; Wong, C.-H. *J. Am. Chem. Soc.* **1994**, *116*, 1135.
(216) Granovsky, M.; Bielfeldt, T.; Peters, S.; Paulsen, H.; Meldal, M.; Brockhausen, J.; Brockhausen, I. *Eur. J. Biochem.* **1994**, *221*, 1039.
(217) Friedman, B.; Hubbard, S. C.; Rasmussen, J. R. *Glycoconjugate J.* **1993**, *10*, 257.
(218) Wright, A.; Morrison, S. L. *TIBTECH* **1997**, *15*, 26.
(219) Malhotra, R.; Wormald, M. R.; Rudd, P. M.; Fischer, P. B.; Dwek, R. A.; Sim, R. B. *Nat. Med.* **1995**, *1*, 237.
(220) Herbert, D. N.; Foellmer, B.; Helenius, A. *Cell* **1995**, *81*, 425.
(221) Witte, K.; Sears, P.; Martin, R.; Wong, C.-H. *J. Am. Chem. Soc.* **1997**, *119*, 2114.
(222) Trimble, R. B.; Atkinson, P. H.; Tarentino, A. L.; Plummer, T. H.; Maley, F.; Tomer, K. B. *J. Biol. Chem.* **1986**, *261*, 12000.
(223) Takegawa, K.; Tabuchi, M.; Yamaguchi, S.; Kondo, A.; Kato, I.; Iwahara, S. *J. Biol. Chem.* **1995**, *270*, 3094.
(224) Deras, I. L.; Takegawa, K.; Kondo, A.; Koto, I.; Lee, Y. C. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1763.
(225) Wang, L.-X.; Fan, J.-Q.; Lee, Y. C. *Tetrahedron Lett.* **1996**, *37*, 1975.
(226) Wang, L.-X.; Tang, M.; Suzuki, T.; Kitajima, K.; Inoue, Y.; Inoue, S.; Fan, J.-Q.; Lee, Y. C. *J. Am. Chem. Soc.* **1997**, *119*, 11137.
(227) Fan, J. Q.; Takegawa, K.; Iwahara, S.; Kondo, A.; Kato, I.; Abeygunawardana, C.; Lee, Y. C. *J. Biol. Chem.* **1995**, *270*, 17723.
(228) Fujita, K.; Takegawa, K. *Biochem. Biophys. Res. Commun.* **2001**, *282*, 678.
(229) Yamamoto, K.; Kadowaki, S.; Watanabe, J.; Kumagai, H. *Biochem. Biophys. Res. Commun.* **1994**, *203*, 244.
(230) Haneda, K.; Inazu, T.; Yamamoto, K.; Kumagai, H.; Nakahara, Y.; Kobata, A. *Carbohydr. Res.* **1996**, *292*, 61.
(231) Yamamoto, K.; Fujimori, K.; Haneda, K.; Mizuno, M.; Inazu, T.; Kumugai, H. *Carbohydr. Res.* **1998**, *305*, 415.
(232) Haneda, K.; Inazu, T.; Mizuno, M.; Iguchi, R.; Tanabe, H.; Fujimori, K.; Yamamoto, K.; Kumagai, H.; Tsumori, K.; Munekata, E. *Biochim. Biophys. Acta* **2001**, *1526*, 242.
(233) Mizuno, M.; Haneda, K.; Iguchi, R.; Muramoto, I.; Kawakami, T.; Aimoto, S.; Yamamoto, K.; Inazu, T. *J. Am. Chem. Soc.* **1999**, *121*, 284.
(234) Cross, G. A. M.; Takle, G. B. *Annu. Rev. Microbiol.* **1993**, *47*, 385.
(235) Stenflo, J.; Fernlund, P. *J. Biol. Chem.* **1982**, *257*, 12180.
(236) Liu, Y.-L.; Hoops, G.; Coward, J. K. *Bioorg. Med. Chem.* **1994**, *2*, 1133.
(237) Shakin-Eshleman, S. H.; Spitalnik, S. L.; Kasturi, L. *J. Biol. Chem.* **1996**, *271*, 6363.
(238) Ajisaka, K.; Miyasato, M.; Ishii-Karakasa, I. *Biosci., Biotechnol., Biochem.* **2001**, *65*, 1240.
(239) Ramos, D.; Rollin, P.; Klaffke, W. *Angew. Chem., Int. Ed.* **2000**, *39*, 396.
(240) Ramos, D.; Rollin, P.; Klaffke, W. *J. Org. Chem.* **2001**, *66*, 2948.
(241) Cumming, D. A. *Glycobiology* **1991**, *1*, 115.
(242) Stanley, P. *Glycobiology* **1992**, *2*, 99.
(243) Grabenhorst, E.; Schlenke, P.; Pohl, S.; Nimtz, M.; Conradt, H. S. *Glycoconjugate J.* **1999**, *16*, 81.
(244) Jenkins, N.; Parekh, R. B.; James, D. C. *Nat. Biotechnol.* **1996**, *14*, 975.
(245) Sears, P.; Wong, C.-H. *Cell. Mol. Life Sci.* **1998**, *54*, 223.
(246) Moffat, A. S. *Science* **1995**, *268*, 658.
(247) Hennet, T.; Ellies, L. G. *Biochim. Biophys. Acta* **1999**, *1473*, 123.
(248) Aebi, M.; Hennet, T. *Trends Cell. Biol.* **2001**, *11*, 136.
(249) Lee, E. U.; Roth, J.; Paulson, J. C. *J. Biol. Chem.* **1989**, *264*, 13848.
(250) Umaña, P.; Jean-Mairet, J.; Moudry, R.; Amstutz, H.; Bailey, J. E. *Nat. Biotechnol.* **1999**, *17*, 176.
(251) Li, F. G.; Wilkins, P. P.; Crawley, S.; Weinstein, J.; Cummings, R. D.; McEver, R. P. *J. Biol. Chem.* **1996**, *271*, 3255.
(252) Dinter, A.; Zeng, S.; Berger, B.; Berger, E. G. *Biotechnol. Lett.* **2000**, *22*, 25.
(253) Jarvis, D. L.; Kawar, Z. S.; Hollister, J. R. *Curr. Biol.* **1998**, *9*, 528.
(254) Wagner, R.; Liedtke, S.; Kretzschmar, E.; Geyer, H.; Geyer, R.; Klenk, H. D. *Glycobiology* **1996**, *6*, 165.
(255) Jarvis, D. L.; Finn, E. E. *Nat. Biotechnol.* **1996**, *14*, 1288.
(256) Bakker, H.; Bardor, M.; Molthoff, J. W.; Gomord, V.; Elbers, I.; Stevens, L. H.; Jordi, W.; Lommen, A.; Faye, L.; Lerouge, P.; Bosch, D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2899.
(257) Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. *Science* **1997**, *276*, 1125.
(258) Yarema, K. J.; Mahal, L. K.; Bruehl, R. E.; Rodriguez, E. C.; Bertozzi, C. R. *J. Biol. Chem.* **1998**, *273*, 31168.
(259) Schmidt, C.; Stehling, P.; Schnitzer, J.; Reutter, W.; Horstkorte, R. *J. Biol. Chem.* **1998**, *273*, 19146.
(260) Baker, K. N.; Rendall, M. H.; Hills, A. E.; Hoare, M.; Freedman, R. B.; James, D. C. *Biotechnol. Bioeng.* **2001**, *73*, 188.
(261) Lee, J. H.; Baker, T. J.; Mahal, L. K.; Zabner, J.; Bertozzi, C. R.; Wiemer, D. F.; Welsh, M. J. *J. Biol. Chem.* **1999**, *274*, 21878.
(262) Lemieux, G. A.; Yarema, K. J.; Jacobs, C. L.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 4278.
(263) Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007.
(264) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. *Org. Lett.* **2000**, *2*, 2141.
(265) Hang, H. C.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2001**, *123*, 1242.
(266) Tkacz, J. S.; Lampen, J. O. *Biochem. Biophys. Res. Commun.* **1975**, *65*, 218.
(267) Karlsson, G. B.; Butters, T. D.; Dwek, R. A. *J. Biol. Chem.* **1993**, *268*, 570.
(268) Rothman, R. J.; Perussia, B.; Herlyn, D.; Warren, L. *Mol. Immunol.* **1989**, *26*, 1113.
(269) Dennis, J. W.; Laferte, S.; Waghorne, C.; Breitman, M. L.; Kerbel, R. S. *Science* **1987**, *236*, 582.
(270) Chui, D.; Sellakumar, G.; Green, R. S.; Sutton-Smith, M.; McQuistan, T.; Marek, K. W.; Morris, H. R.; Dell, A.; Marth, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1142.

(271) Mamaev, S. V.; Laikther, A. L.; Arslan, T.; Hecht, S. M. *J. Am. Chem. Soc.* **1996**, *118*, 7243.
(272) Arslan, T.; Mamaev, S. V.; Mamaeva, N. V.; Hecht, S. M. *J. Am. Chem. Soc.* **1997**, *119*, 10877.
(273) Schmidt, R. R.; Castro-Palomino, J. C.; Retz, O. *Pure Appl. Chem.* **1999**, *71*, 729.
(274) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182.
(275) Fahmi, N. E.; Golovine, S.; Wang, B.; Hecht, S. M. *Carbohydr. Res.* **2001**, *330*, 149.
(276) Cornish, V. W.; Hahn, K. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 8150.
(277) Wells, L.; Vosseller, K.; Hart, G. W. *Science* **2001**, *291*, 2376.
(278) Dintzis, R. Z.; Okajima, M.; Middleton, M. H.; Greene, G.; Dintzis, H. M. *J. Immunol.* **1989**, *143*, 1239.
(279) A valuable point was raised by a reviewer. In this manuscript protein glycosylation methods have been divided into classes of chemoselective, site-specific, and site-selective methods. This distinction is arbitrarily based on the overall glycosylation *strategy* and does not necessarily reflect the inherent specificity or selectivity of a given reaction. For example, use of the same chemoselective (e.g., thiol-reactive) reaction will give rise only to chemoselectivity in a protein with multiple existing cysteine thiols, to site specificity when used to glycosylate a single, already existing cysteine thiol, and to site selectivity when used to modify a single cysteine thiol that is introduced at a preselected position in a previous step of the strategy.
(280) This wonderful idea was suggested by a reviewer.

CR0004310