REVIEW

Glycopeptides as versatile tools for glycobiology

Therese Buskas, Sampat Ingale, and Geert-Jan Boons
Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA

This review describes the recent advances in the field of glycopeptide and small glycoprotein synthesis. The strategies covered include chemical and chemoenzymatic synthesis, native chemical ligation (NCL), and expressed chemical ligation. The importance of glycopeptide synthesis is exemplified by giving the reader an overview of how versatile and important these well-defined glycopeptides are as tools in glycobiology.

Accepted on April 27, 2006

Key words: glycobiology/glycopeptides/glycoproteins/oligosaccharide/synthesis

Introduction

Naturally occurring types of protein glycosylation

Glycosylation is a common co- or posttranslational modification that is estimated to be present in more than half of all proteins in nature (Apweiler et al., 1999). The carbohydrate moiety of glycoproteins can exert a wide range of biological functions such as fertilization, embryogenesis, neuronal development, hormone activities, proliferation of cells and their organization into specific tissues, immune surveillance, and inflammatory reactions (Varki, 1993; Bertozzi and Kiessling, 2001). Remarkable changes in cell-surface carbohydrates occur with tumor progression, which appears to be intimately associated with the dreaded state of metastasis (Kim and Varki, 1997; Dennis et al., 1999). Furthermore, the carbohydrates of host cells are often employed by pathogens for cell entry or immunological evasion (Rostand and Esko, 1997; Moncada et al., 2003; van Kooyk et al., 2004). Almost all naturally occurring protein glycosylations can be classified either as N-glycosides, whereby N-acetylglucosamine is linked to the amide side chain of an asparagine, or as O-glycosides, whereby a saccharide is linked to the hydroxyl of serine, threonine, or tyrosine. Recently, several unusual glycoproteins have been identified such as O-glycosides of hydroxylysine and hydroxyproline and the C-glycosides of tryptophan (Figure 1).

The biosynthesis of N-linked oligosaccharides occurs in the endoplasmic reticulum (ER) and Golgi complex (Ellgaard and Helenius, 2001; Helenius and Aebl, 2001). In the ER, a dolichol-linked Glc3Man9GlcNAc2 oligosaccharide precursor is biosynthesized and transferred en bloc to an Asn-X-Ser/Thr sequon on newly synthesized polypeptides through the action of a multi-subunit oligosaccharide transferase complex. Subsequent trimming and processing of the transferred oligosaccharide results in a Man3GlcNAc2 core structure, which is transported to the medial stacks of the Golgi complex where maturation of the oligosaccharide gives rise to an extreme structural diversity. This complexity of N-glycan structures is largely based on the cell-specific expression of a collection of glycosyltransferases that specify the extension of oligosaccharide structures onto the trimmed Man3GlcNAc2 core structure. The switch from structural uniformity in the ER to diversification in the Golgi complex coincides with a marked change in glycan function. In the early secretory pathway, the glycans have a common role in the promotion of protein folding, quality control, and certain sorting event. This is in contrast to their roles in the Golgi complex, in which they are modified to perform a wide spectrum of functions displayed by the mature glycoproteins.

Modification of the side chain of serine or threonine by N-acetyl-α-D-galactosamine is the most common form of O-linked glycoproteins (Perez-Vilar and Hill, 1999; Roussel and Delmotte, 2004; Hang and Bertozzi, 2005). The biosynthesis of these glycans occurs in the Golgi apparatus where the GalNAc moiety of UDP-GalNAc is transferred to the hydroxyl of serine or threonine catalyzed by polypeptide GalNAc transferase. In contrast to N-glycosylation, a consensus sequence for α-D-GalNAc addition has not been found, although predictive algorithms do exist. Many O-glycans are extended into long biantennary oligosaccharide chains with variable termini that may be similar in structure to those of N-linked glycoproteins. Glycoproteins, which are modified by dense clusters of O-glycosylation, have been coined mucins (for core structures of the mucin type, see Figure 2). Generally, they are found at the apical surface of epithelial cells where they function as protective barriers and provide lubrication because of their hydration capacity. Beyond bulk properties, mucins can modulate protein functions such as intracellular trafficking and the regulation of half-lives of chemokines and hormones as well as determining aggregation states of membrane-bound glycoproteins. The carbohydrate moiety of mucins can also serve as ligands for cell adhesion events, and perhaps the best-studied example is the homing of leukocytes to the site of inflammation, a process dependent on the interaction oligosaccharides of mucins with selectins (Mcever et al., 1995; Ley and Kansas, 2004). Mucins of tumor cells exhibit
Glycopeptides as versatile tools for Glycobiology

striking alterations in the level of expression and glycosylation profile (Baldus et al., 2004; Byrd and Bresalier, 2004). For example, aberrations of glycosylation are often associated with tumor progression and prognosis. Mucins are also involved in host–pathogen interactions (Hicks et al., 2000; Theodoropoulos et al., 2001), and for example, many pathogens exploit host mucins for cell adhesion or cell entry or alternatively parasite-derived mucin-like compounds can be involved in the attachment of microbes to host cells.

In the cytosol and nucleus, many proteins are glycosylated by a single β-linked N-acetyl-β-D-glucosamine (β-D-GlcNAc) linked to the side chain hydroxyl of serine or threonine (Slawson and Hart, 2003; Wells and Hart, 2003). This type of O-glycosylation is rendered highly dynamic by the action of a unique set of enzymes that add or remove the monosaccharide. Functionally, β-D-GlcNAc modification appears to regulate protein stability, sub-cellular localization, and protein–protein interactions. It often acts in a reciprocal manner to phosphorylation of proteins and together they can synergistically control the activity of many cellular processes. It has been demonstrated that β-D-GlcNAc addition plays a role in diseases such as diabetes, cancer, and neurodegeneration. For example, the increased levels of β-D-GlcNAc additions characteristic of diabetes are associated with decreased insulin responsiveness in adipocytes.

The attachment of α-D-GlcNAc to hydroxyproline is another type of O-glycosylation that is performed in the cytoplasm (West et al., 2004). This type of modification is only present on a small protein, named Skp1, which is expressed universally in the cytoplasmic and nuclear compartments of eukaryotes. Skp1, produced by the slime mold Dictyostelium discoideum, is modified by a pentasaccharide, which is introduced by the sequential action of a soluble prolyl hydroxylase and five soluble glycosyltransferases.

Apart from GlcNAc and GalNAc, proteins can be modified by other monosaccharides such as xylosides, fucosides, mannosides, and glucosides. For example, the side chains of selected serine residues of heparan sulfate (HS) and chondroitin sulfate (CS) core proteins are glycosylated in the Golgi compartment by a xyloside (Esko and Selleck, 2002; Grobe et al., 2002). Through the action of three different glycosyltransferases, the xyloside is extended to a tetrasaccharide with the following structure: GlcAβ(1,3)Galβ(1,3)Galβ(1,4)Xylβ-Ser. The first committed step in HS biosynthesis is the addition of GlcNAcα(1,4) residue followed by polymerization by the alternating

Fig. 1. Examples of naturally occurring protein glycosylation.

Fig. 2. Core structures of mucin-type O-linked glycans.
addition of GlcAβ(1,4) and GlcNAcα(1,4) residues. Subsequently, the disaccharide repeating units are modified by enzymatic N-deacetylizations, epimerizations of glucuronic acid to iduronic acid moieties, and selective O- and N-sulfations. Ultimately, these modifications result in the formation of an IdoS(2-OSO$_3$)-GlcNSO$_3$(6-OSO$_3$)$_2$ sequence. Structural studies have, however, shown that HS contains many other disaccharide subunits arising from incomplete or additional enzymatic modifications. The first committed step in the biosynthesis of HS involves the addition of a GalNAcβ(1,4) to the protein-linked tetrasaccharide. The resulting pentasaccharide is then further extended by the addition of multiple GlcAβ(1,4)GalNAcβ(1,3) disaccharide repeating units. Subsequently, two sulfotransferases modify the disaccharides by the addition of sulfate esters to the C-4 and C-6 positions of GalNAc. Control over the addition of α-GlcNAc leading to HS or β-GalNAc resulting in CS is controlled by the amino acid sequence of the protein. HS has been implicated in many biological processes, and many enzymes, growth factors, enzyme inhibitors, chemokines, extracellular matrix components, cell adhesion molecules, and microbial proteins require HS for their functions. Among others, HS–protein complexes have been implicated in the modulation of embryonic development (by growth factor modulation), inhibition of blood coagulation, organization of the extracellular matrix, angiogenesis, anchorage of cells, presentation of enzymes and cytokines on cell surfaces and as “co-receptors” in viral infections.

Fucosides linked to serine and threonine have been observed in many proteins including uPA, tissue-type plasminogen activator, and several clotting factors (Haltiwanger, 2002; Haltiwanger and Stanley, 2002).

A comparison of the sequences surrounding the sites of O-fucosylation showed it to be localized to a putative sequence within EGF repeats. O-fucosylation appears to be involved in signal transduction. For example, binding of uPA to the uPA receptor results in the activation of several signaling cascades. Although removal of the fucosides from the EGF repeat did not affect binding to the receptor, it did abrogate uPA receptor activation. This example demonstrates that the presence or absence of a simple monosaccharide can regulate signal transduction events. The Notch cell-signaling pathway is also regulated by the alteration of a carbohydrate modification. The extracellular domain of Notch is composed of many EGF repeats, several of which contain consensus sites for O-fucosylation modification. Fringe, which is an N-acetylgalcosaminyltransferase, modulates Notch activity by altering fucosides on EGF repeats by the addition of β(1,3)-GlcNAc. Subsequent addition of a β(1,4)galactoside results in the inhibition of Jagged1-dependent Notch activation.

Although α-mannosylation of serine and threonine moieties of proteins was originally observed in fungi, it has been established that this type of protein modification is widespread and is, for example, found in mammals (Strahl-Bolsinger et al., 1999; Willer et al., 2003). These glycosides are beginning to attract considerable interest because a defect in their biosynthesis has been linked to several forms of congenital muscular dystrophies, which are often associated with brain abnormalities.

Recently, a novel structural class of glycoproteins was identified in human RNase 2, where an α-t-mannopyranose is connected to the C-2 of a tryptophan residue via a C-glycosidic linkage (Vliegenthart and Casset, 1998; Gonzalez de Peredo et al., 2002). Subsequent studies have revealed that C-mannosylation involves the attachment of a mannose residue to the indole moiety of Trp-Xaa-Xaa-Trp (Xaa is any amino acid) as a consensus recognition site. It has been established that dolichyl-phosphate mannose is the glycosyl donor for the biosynthesis of C-mannosylation. Although C-mannosyltransferase activity is found in most mammals, its biological function remains unclear. The most striking example of C-mannosylation is that of the protein FRGP, a positive regulator of complement, which contains 20 tryptophan moieties of which as many as 17 are C-mannosylated.

Recently, several new classes of glycoproteins have been identified in which an oligosaccharide is linked to serine or threonine moieties via a phosphodiester linkage. This new type of protein modification has been referred to as protein phosphoglycosylation (Haynes, 1998; Ilg, 2000). The first reported example of a protein modified by a phosphoglycoside was an endopeptidase isolated from the slime mold D. discoideum. It was demonstrated that this glycoprotein contains a serine moiety modified by a phosphodiester linked to N-acetylgalcosamine (α-D-GlcNAc-1-PO$_4$3-Ser). It has also been established that two other cysteine proteinases isolated from D. discoideum carry GlcNAc-1-PO$_4$ modifications. Protein phosphoglycosylation is the predominant type of protein glycosylation in the parasite Leishmania, and for example, a secreted acid phosphatase of Leishmania mexicana is modified by α-mannosidic phosphodiester linkages, which either can be monomeric or may consist of a series of neutral α(1–2)-linked oligomannannanes or phosphorylated oligosaccharides composed of PO$_4$3-Galβ(1–4)Man and PO$_4$3-Glc(β(1–3))Galβ(1–4)Man repeating units capped with a neutral mannosyl oligosaccharide (McConville et al., 1990; Ilg et al., 1994, 1996). Phosphoglycosylation has also been observed in other microorganisms. In Trypanosoma cruzi, fucosyl-, fucose-, xylose-, and galactose-containing phosphoglycans are linked to several proteins via phosphothreonine. These components may be important for the development of the parasite in the insect stage. The recently identified glucosylphosphatidylinositol (GPI) anchor of Entamoeba histolytica carries linear α(1–6) glucans linked to serine phosphate and is thought to be a virulent factor and vaccine candidate. In mammals, as in other vertebrates, phosphoglycosylation appears to be absent, therefore implying that protein phosphoglycosylation is an attractive target for drug development.

Previously, glycosylation was thought to be a phenomenon exhibited exclusively by eukaryotes. Since the first report of a glycoprotein in halobacteria, there has been an overwhelming response in relation to the discovery of bacterial glycoproteins (Schmidt et al., 2003; Upreti et al., 2003). Most of the bacteria reported to possess glycosylation belong to the Archaea bacteria, but evidence is emerging that shows it also occurs in Eubacteria.

Production of Glycoproteins and Glycopeptides

As protein glycosylation is not under direct genetic control, it results in the formation of a heterogeneous range of
glycoforms that possess the same peptide backbone but differ in the nature and site of glycosylation. In general, it is difficult to control glycoform formation in cell culture, which hampers progress in glycobiology and is a major obstacle to the development of therapeutic glycoproteins.

The production of recombinant proteins that are N-glycosylated in their native states requires a mammalian expression host that has the ability to mimic human glycosylation. Prokaryotic hosts such as Escherichia coli do not glycosylate their proteins, and lower eukaryotic expression systems such as yeast and insect cells are typically unable to provide mammalian-type glycosylation. In particular, Chinese hamster ovary (CHO) cells have emerged as the most widely used expression system for the production of human glycoproteins, although it is recognized that glycan structures from CHO cell lines differ from those produced in human cells (Fussenegger et al., 1999; Warner, 1999). To overcome this problem, in vitro enzymatic methods have been employed to modify glycosylation after purification of the protein. In an alternative strategy, genetic engineering of the glycosylation pathway has been employed to provide more human-like carbohydrate additions. For example, the rat α(2,6)-sialyl transferase, a gene that is normally nonfunctional in CHO cells, has been coexpressed along with the human tissue plasminogen activator factor, tPA, and as expected the resulting protein contained 2,6-linked sialic acid residues.

The growing demand for improved biopharmaceutical expression hosts has led several research groups and companies to glyco-engineer insect, yeast, and fungal expression systems to produce humanized glycoproteins (Tomiya et al., 2003; Gerngross, 2004; Nevalainen et al., 2005). Early N-glycan processing in the ER is highly conserved among many different organisms. However, after transport of the glycoprotein to the Golgi compartment, glycosylation pathways diverge considerably. For example, at this junc

ture, mammalian cells rely on α(1,2)-mannosidases to trim mannosides before complex oligosaccharides are being biosynthesized, whereas yeasts initiate a series of mannosyl-transferase reactions to give high mannose structures. Thus, a critical step in obtaining complex glycoproteins in a fungal expression host is to engineer α(1,2)-mannosidase activity. Ultimately, a strain of Pichia pastoris displaying relevant α(1,2)-mannosidase activity in vivo was obtained by the generation of large-scale combinatorial genetic libraries of catalytic domains of α(1,2)-mannosidases fused to yeast type II leader sequences (Chen et al., 2005; Gomord et al., 2005). Screening was performed by purifying a recombinant reported protein secreted by each strain and analyzing its glycans by matrix-assisted laser desorption time-of-flight (MALDI-TOF) mass spectrometry after release of the glycans by N-glycanase. Over 90% of the glycans were of the Manα type providing a basis for further humanization of the glycosylation pathway in P. pastoris. Next, after knocking out α(1,6)-mannosyltransferase activity, and introducing active GnTI, GnTII, GaIT, and Golgi mannosidase II genes, a UDP-GlcNAc transporter, and a UDP-galactose-4-epimerase, a P. pastoris strain was created that produced GlcNAcManαGalβ glycans. Unlike glycoproteins obtained from mammalian cell culture, it has been found that the humanized yeast strains secrete glycoproteins with exceptional glycan uniformity.

Efforts are also underway to utilize transgenic plants to produce humanized glycoproteins (Chen et al., 2005; Gomord et al., 2005). The advantages of this expression host are significantly lower cost to produce recombinant proteins; minimal risk of contamination by human or animal pathogens, oncogenic DNA, and endotoxins; and the production of properly folded and glycosylated proteins. In respect of the latter, plants are able to produce glycoproteins modified by complex N-linked glycans. However, plant saccharides differ from those of humans by lacking galatosides and terminal sialic acids and the addition of α(1,3)-fucosides and β(1,2)-xylosides. Not surprisingly, concerns have been raised that plant-derived glycoproteins can induce immunological responses. To address these problems, transgenic Arabidopsis strains have been created by knocking out relevant α(1,3)-fucoside and β(1,2)-xyloside genes. Furthermore, efforts are underway to introduce a β(1,4)galactosyltransferase in plant to produce humanized glycoproteins.

Recently, a highly innovative approach has been reported in which well-defined O-linked glycoproteins were produced in E. coli by evolving an orthogonal synthetase–tRNA pair that genetically encodes a glycosylated amino acid in response to the amber code, TAG (Zhang et al., 2004). Thus, a series of positive and negative selections were used to isolate from a library of active mutants, an orthogonal Methanococcus jannaschii tyrosyl tRNA synthetase that specifically charges the corresponding M. jannaschii suppresses with β-GlcNAc-serine. Tri-O-acetyl-β-D-GlcNAc-serine was used as a β-D-GlcNAc-serine source because acetylation increases cellular uptake. Subsequently, intracellular esterases can remove the acetyl esters to reveal β-D-GlcNAc-serine, which can then be used to charge tRNA. By employing this approach, myoglobin was expressed containing a β-D-GlcNAc-serine at a defined position.

A combination of site-directed mutagenesis and chemical attachment of synthetic saccharides offers another unique approach for the preparation of well-defined glycoproteins (Davis, 2004; Khmelnitsky, 2004). This strategy involves the introduction of a cysteine in a protein, which can site-selectively be coupled with a synthetic saccharide containing an electrophilic moiety. This approach exploits the fact that cysteines in proteins are usually present as disulfides. An additional cysteine moiety with a free thiol can, however, be introduced at any point of a polypeptide chain by site-directed mutagenesis. The resulting highly nucelophilic thiol can then be reacted with a sugar bearing an electrophilic moiety to give a well-defined neoglycoprotein. In a seminal study (Macmillan et al., 2001), Flitsch and coworkers have coupled a monosaccharide bearing an iodoacetamide group to a human erythropoietin cysteine mutant. In an alternative strategy, we have engineered an IgG1-Fc fragment that has its glycan-linked Asn replaced by Cys-297 (Watt et al., 2003). Subsequently, synthetic monoo-, di-, tri-, and pentasaccharides derived from N-linked oligosaccharides were coupled to the thiol of the cysteine moiety by asymmetric disulfide formation. The resulting neoglycoproteins were tested for their ability to interact with human FcγRI by inhibiting superoxide production by interferon-γ (IFNγ)-stimulated U937 cells. These studies showed, for the

---

Glycopeptides as versatile tools for Glycobiology
first time, that the biological activity of a glycoprotein lost by the removal of its natural glycosylation site could be restored partially by the in vitro attachment of synthetic saccharides. In vitro glycosylation of recombinant proteins offers several distinct advantages including quantitative conjugation of oligosaccharides, the incorporation of oligosaccharides that are of low natural abundance but which are known to display high bioactivities, and the possibility of using more efficient bacterial or yeast protein expression systems.

While recombinant methods for the production of glycoproteins have advanced dramatically, so has the synthesis of glycopeptides. Routine procedures can now be employed for the chemical synthesis of glycopeptides carrying many simple glycans (Seitz, 2000; Davis, 2002; Grogan et al., 2002). Enzymatic approaches have been utilized to introduce more complex glycans into synthetic glycopeptides. The advent of native chemical ligation (NCL) now makes it possible to synthesize glycoproteins from well-designed peptide and glycopeptide building blocks. In Chemical Synthesis of Glycopeptides, recent progress in the chemical and chemoenzymatic synthesis of glycopeptides will be discussed in detail.

Chemical Synthesis of Glycopeptides

A crucial step in the chemical synthesis of glycopeptides is the incorporation of the saccharide into the peptide. To accomplish this, two approaches can be considered: the direct glycosylation of a properly protected full-length peptide and the use of a preformed glycosylated amino acid building block for the stepwise synthesis of the peptide backbone (Scheme 1). An advantage of the direct glycosylation method is that the route is more convergent, permitting fast access to glycopeptides differing in glycan structure. However, direct O-glycosylation is often plagued by low yields due to the low reactivity of the side chain hydroxyls and the low solubility of the peptides under conditions commonly employed for chemical glycosylation. To reach N-glycosylated peptides, the method of direct condensation of a glycosylamine and an aspartic acid-containing peptide is severely hampered by the formation of intramolecular aspartimides (Bodanszky and Natarajan, 1975). This side reaction eventually leads to peptides linked through the α- and β-carboxyl groups.

Currently, the most general synthetic methodology for glycopeptides employs preformed glycosylated amino acids for the stepwise solid-phase peptide synthesis (SPPS).

It is important to realize that in developing a methodology for solid-phase glycopeptide synthesis, the additional complexity and lability conferred by the carbohydrate group must be considered. Thus, many methods and reaction conditions that are commonly employed in SPPS are not suitable for the preparation of glycopeptides (Seitz, 2000; Davis, 2002; Grogan et al., 2002). Therefore, special care must be taken in choosing protecting groups for the glycosylated amino acids. The choices are rather limited, as the O-glycosidic bond and particularly the α-fucosidic linkage are acid labile (Kunz and Unverzagt, 1988; Unverzagt and Kunz, 1994; Peters et al., 1995), and the O-linked glycopeptide can undergo β-elimination (Sjolin et al., 1996) upon treatment with strong bases (Scheme 2). Strong bases can also racemize the stereogenic centers of peptides.

Presently, the use of acetyl esters as hydroxyl protection for the oligosaccharide part and Nα-Fmoc protected amino acids (Carpino and Han, 1972), together with coupling reagents such as DIC/HOBt (Albericio and Carpino, 1997), PyBOP/HOBt (König and Geiger, 1970), HBTU/HOBt (Coste et al., 1990), and HATU/HOAt (Carpino, 1993), has emerged as a standard technique in solid-phase glycopeptide synthesis. Typical linkers (Songster and Barany, 1997) used for glycopeptide synthesis are shown in Figure 3. In the Fmoc protocol, the iterative removal of the Nα-Fmoc group needed for peptide elongation can be achieved using mild bases such as piperidine and morpholine, without affecting the O-acetyl groups of the saccharide or inducing β-elimination of the glycan (Kihlberg and Vuljanic, 1993; Vuljanic et al., 1996). However, in the Boc/benzyl protocol, the recurring cleavage of the Nα-Boc group is accomplished by trifluoroacetic acid (TFA) treatment, and such strongly acidic conditions are incompatible with the presence of acid-labile glycosidic bonds.

The use of the electron withdrawing O-acetyl and O-benzoyl protecting groups stabilizes the glycosidic linkage during TFA cleavage from the solid support. Acetyl groups can be removed by treatment with dilute sodium methoxide in methanol or hydrazine hydrate (Schultheissreimann and Kunz, 1983) or saturated ammonia in methanol (Paulsen et al., 1985). These reaction conditions, if carefully monitored, will not affect the glycopeptide structure. However, the more severe basic conditions required for the removal of O-benzoate render them less suitable for use in glycopeptide
synthesis (Paulsen et al., 1985; Reimer et al., 1993; Sjolin et al., 1996).

The use of O-acylated glycosyl amino acids in glycopeptide synthesis is, however, not free of complications. One such complication is that the basic conditions needed for deprotection can affect the side chains of cysteine (Vuljanic et al., 1996) and asparagine (Peters et al., 1995). Also, during introduction of the O-acylated glycosyl amino acid, an acyl migration to the terminal amino group can occur if the coupling reaction is exceedingly slow, leading to capping of the growing peptide backbone (Elofsson et al., 1997). In light of the complications, alternative protecting groups for the carbohydrate moiety of glycopeptides have been explored. Benzyl ethers, which are commonly employed protecting groups in carbohydrate chemistry, have a limited use as their removal is typically affected by catalytic hydrogenolysis, making it particularly difficult in the presence of cysteine and methionine side chains (Bodanszky and Martinez, 1981). Recently, it has been reported (Hojo et al., 2003; Takano et al., 2004) that de-O-benzylaion of the glycan part of a glycopeptide can be affected by using the low TfOH (Tam et al., 1986) approach, which offers milder and less acidic SN2-type deprotection conditions.

Synthetic routes to properly protected O- and N-glycosyl amino acids that emphasize on the installment of the glycosidic linkage have garnered excellent reviews and, therefore, will not be discussed in detail (Arsequell and Valencia, 1997; Taylor, 1998). This review intends to focus on the various synthetic routes that are commonly employed to obtain glycopeptides and show how important these glycopeptides are as tools for biological studies.

The α-GalNAc-Ser/Thr glycosyl amino acid commonly known as the Tn antigen is a key structure for the mucin-type O-linked glycopeptides. The O-glycosidic bond of the Tn antigen is commonly installed by glycosylation of the hydroxyl group of a properly protected Ser or Thr using a 2-azido-2-deoxygalactose halo-donor (Scheme 3) (Paulsen and Holck, 1982; Kunz et al., 1990; Nakahara et al., 1990;
Paulsen et al., 1994; Liebe and Kunz, 1997; Chen et al., 1998; Kuduk et al., 1998). A typical route employed for glycosylation is the silver salt-promoted activation of fully acetylated 2-azido-galactose bromide (Winans et al., 1999) in the presence of the amino acid building block, which gives the glycosyl amino acid in a mixture of anomers. Also, properly protected trichloroacetimidate (Chen et al., 1998; Kuduk et al., 1998), \(n\)-pentenyl (Svarovsky and Barchi, 2003), and thioglycoside (Paulsen et al., 1988; Elofsson et al., 1997) donors have also been utilized for the installment of the \(\alpha\)-GalN\(_3\)-Ser/Thr bond. Recently, it was found that the use of an acetylated 2-azido-2-deoxy thiophenyl donor and the Ph\(_2\)SO/Tf\(_2\)O promotor system for the glycosylation of Fmoc-Thr(OH)-OBn gave complete stereoselectivity with only the \(\alpha\)-glycoside formed (Scheme 3) (Cato et al., 2005). The reduction of the 2-azido-2-deoxy functionality using, for example, thioacetic acid or Zn in acetic acid followed by acetylation will provide the \(N\)-acetylated Gal\(_{\alpha}\)NAc-Ser/Thr building block.

With respect to the \(N\)-linked glycoproteins, glycans are attached to the Asn residue of the Asn-X-Thr/Ser sequence. The preparation of glycosyl amino acid building blocks for \(N\)-glycopeptides requires a completely different strategy because an anomeric amide must be introduced instead of a glycosidic linkage (Arsequell and Valencia, 1999). The favored approach is the reaction of a glycosylamine and a suitably protected Asp derivative (Likhosherstov et al., 1986; Clark et al., 1990). The glycosylamine is typically obtained by the reaction of a reducing sugar with ammonia or ammonium hydrogen carbonate, a method known as the Kochetkov amination (Likhosherstov et al., 1986), or by the reduction of a corresponding anomeric azide derivative (Scheme 4). However, both these methods suffer from disadvantages. Anomerization of the glycosylamine can occur under reduction conditions, and the glycosylamine is not stable during the aminolysis reaction and can form trehalose derivatives. A direct method to obtain an \(N\text{\textbeta}\)-glycosylated asparagine derivative, which avoids glycosylamine anomerization, is the reaction of a glycosyl azide and an \(N\text{\textalpha}\)-protected aspartic acid \(\alpha\)-monoester in the presence of trialkylphosphine at low temperature (Scheme 4) (Mizuno, Haneda, et al., 1999).

Since the discovery of glycoproteins where a mannose residue is connected to tryptophan via a \(C\)-glycosidic linkage, several approaches to reach \(C^2\text{-}\alpha\text{\textalpha\textbeta}\)-C-Man-L-Trp building blocks have been reported (Manabe and Ito, 1999; Nishikawa et al., 2001). The group of Ito (Manabe et al., 2003) recently published a route that differs from conventional approaches to \(C\)-glycosylation (Beau and Gallagher, 1997). In their approach, the \(C\)-\(C\) bond between the 2-position of the indole ring of tryptophan and the anomeric carbon of mannose was installed by reaction of a C-2-lithiated \(N\text{-arylsulfonated}\) indole derivative with 1,2-anhydro-mannose (Scheme 5). After protecting group manipulations, the alcohol of the latent tryptophan moiety was directly oxidized to the corresponding carboxylic acid without concomitant oxidation of the indole moiety using a combination of TEMPO and iodosobenzene diacetate.

A novel approach was reported by the authors to reach glycopeptide structures from the recently discovered glycoprotein class, which carries glycans through a phosphodiester linkage to serine and threonine moieties (Majumdar et al., 2005). Anomeric \(\alpha\text{-D-mannosyl}\) and \(\alpha\text{-N-acetyl-D-glucosamine}\) phosphoramidite derivatives were coupled to the hydroxyl of various amino acids and peptides by activation with tetrazole (Scheme 6). The phosphites were then oxidized \textit{in situ} to give the corresponding phosphotriesters. Deprotection could be performed using conventional methods.
It was found that glycophosphorylation of preassembled and properly protected peptides presents a more versatile methodology for phosphoglycopeptide synthesis than the use of glycosphospho serine derivatives in linear stepwise peptide assembly.

Synthetic routes to an array of properly protected glycosyl amino acid carrying simple O- and N-glycans are well established, although in some cases fairly laborious. Indeed, an array of building blocks is commercially available today. However, the use of an amino acid carrying a large oligosaccharide in SPPS can result in an undesirable rapid drop-off of coupling yields as the oligosaccharide increases in size and branching. Also, glycosylation of an amino acid alcohol with a complex oligosaccharide donor to prepare a glycosyl amino acid building block can be problematic. For example, it has been found that the glycosylation of a Thr/Ser building block with the relatively simple monosaccharide GalN$_3$ donor generally results in anomic mixtures (Taylor, 1998). Using more complex oligosaccharide donors in this glycosylation can be unpredictable in regard to both yields and stereoselectivity. To this end, Meldal and coworkers (Mathieux et al., 1997) introduced what was later termed “the cassette-method” (Chen et al., 1998), an approach that utilizes a common GalN$_3$-Thr building block upon which the glycans can be extended by successive glycosylation and deprotection (Scheme 7). Glycosyl amino acids carrying core structures 1, 2, 3, 4, and 6

Scheme 5. Synthesis of C$_2$-α-D-Man-Trp. (i) BuLi, BF$_3$-OEt, THF, 63%. (ii) iodosobenzene diacetate, TEMPO, CH$_3$CN, H$_2$O, 91%.

Scheme 6. Synthesis of a α-D-Man-1-PO$_4$-Ser building block. (i) (Pr$_3$N)$_2$PCl, DIPEA, DCM. (ii) BnOH, 1H-tetrazole. (iii) 1H-tetrazole, CH$_3$CN. (iv) t-BuOOH, –40°C.

Scheme 7. Strategic principle of the cassette method.
have been synthesized according to this method. Using this approach in conjunction with multiple-column solid-phase synthesis (Peters et al., 1992), a series of O-linked glycopeptides derived from the mucins MUC-2 and MUC-3 could be synthesized, permitting a systematic glycosylation.

A way to obtain complex N-linked oligosaccharides is by isolation and purification of glycans from natural glycoproteins. The glycans can subsequently be utilized in glycopeptide synthesis. This procedure thus circumvents the tedious and time-consuming chemical synthesis of the glycans. N-linked glycans can be released from natural glycoproteins by mild hydrazinolysis and subsequently transformed into the corresponding glycosylamines using the Kochetkov amination outlined in Scheme 4 and reacted with a properly protected Asn derivative. After acetylation of the free hydroxyl groups of the oligosaccharide, these building blocks can be utilized in glycopeptide synthesis (Meinjohanns et al., 1998). Recently, this approach was employed in the synthesis of a glycopeptide carrying a sialyl-undeca- and an asialo-nonasaccharide (Scheme 8) (Kajihara et al., 2004). In their semi-synthetic method, they combined the isolation of a sialyl glycopeptide from egg yolk with the use of branch-specific exo-glycosidases for the cleavage of distal sugar residues in the preparation of a wide range of homogeneous Asn-linked oligosaccharides, which later could be incorporated into glycopeptides. Impressively, the glycans were introduced and carried through the glycopeptide synthesis unprotected.

In efforts to minimize hydroxyl protecting group manipulations, several groups have utilized glycosyl amino acids carrying unprotected glycans in glycopeptide synthesis (Otvos et al., 1990; Unverzagt, 1997; Mizuno, Muramoto, et al., 1999; Ichiyanagi et al., 2002; Xue and Guo, 2003; Kitamura et al., 2004; Takamura et al., 2004; Guo and Shao, 2005). One example is Guo’s chemical synthesis of CD52 (Scheme 9) (Shao et al., 2003). This GPI-anchored glycopeptide is expressed on virtually all human lymphocytes and sperm cells and is believed to play an important role not only in the human immune system but also in the reproduction process. The peptide backbone of CD52, which is short at a length of only 12 amino acids, carries an N-glycan containing an α-linked fucose residue. The oligosaccharide linked to an Asn residue was carried unprotected through the glycopeptide synthesis, which for comparison was performed in both solution phase and solid phase. Also, the hydrophilic properties induced by the free oligosaccharide hydroxyls allowed easy purification by precipitation (Wen and Guo, 2001). It was found that the unprotected glycan containing an extremely acid-sensitive fucosidic linkage exhibited better stability toward acidic cleavage than the corresponding benzyl protected glycan. Surprisingly, the fucosidic linkage could withstand the treatment of 18% TFA in dichloromethane (DCM) necessary for deprotection of the glycopeptides.

Scheme 8. Synthesis of a glycopeptide carrying a sialyl-undeca- and an asialo-nonasaccharide. An N-glycan was isolated from egg yolk and trimmed using branch-specific exo-glycosidases. The Asn-linked glycans were N-Fmoc protected and used in SPPS. The carboxylic groups of NeuAc were protected as benzyl esters. The oligosaccharide hydroxyl groups were carried through the synthesis unprotected.

Scheme 9. Synthesis of a CD52 glycopeptide carrying an unprotected oligosaccharide through the synthesis. (i) DCC, HOBt, NMP. (ii) HOAc/TFE/DCM, 1/1/8. (iii) TFA/DCM/Et3SiH, 2/8/1.
Chemoenzymatic Synthesis of Glycopeptides

The incorporation of amino acids carrying large and complex glycans into long peptides remains a significant challenge. The use of enzymatic transfer of individual monosaccharides or large oligosaccharides to glycopeptides containing simple glycans, thus, offers an attractive alternative to the tedious chemical synthesis of complex oligosaccharides.

The power of chemoenzymatic synthesis was, for example, elegantly demonstrated by the construction of the glycopeptide P-selectin ligand-1 (PSGL-1). The P-, E-, and L-selectins are a family of adhesion proteins that function as cell-surface receptors, which are involved in the inflammatory cascade (Rosen and Bertozzi, 1994). P-selectin is present on the cell surface of epithelial cells, and the binding to its counter receptor PSGL-1, which is expressed on leukocytes, initiates the inflammatory cascade by capturing the leukocytes to sites of inflammation. PSGL-1 is a dimeric membrane-bound mucin whose binding to P-selectin occurs through the N-terminal, at which it carries a sialyl Lewis x (SLex) moiety O-linked to the peptide backbone at Thr57 through the core-1 saccharide. The glycopeptide is sulfated at Tyr46, Tyr48, and Tyr51. To discern detailed information about the interaction of P-selectin with the O-glycan and sulfated Tyr residues on PSGL-1, Cummings and coworkers have synthesized an array of sulfated glycopeptides corresponding to the N-terminus of PSGL-1 by chemoenzymatic methods (Scheme 10) (Leppanen et al., 1999, 2000). Synthetic glycopeptides were assembled using standard Fmoc-chemistry protocol in SPPS, in which the GalNAc-Thr residue was incorporated using a tri-O-acetyl-GalNAc-Fmoc-Thr-OH cassette. Sulfation of the Tyr residues was introduced chemically (Leppanen et al., 2000) or enzymatically (Leppanen et al., 1999), producing many glycopeptides differing in sulfation pattern. The purified glycopeptides then functioned as acceptors for the specific glycosyltransferases required for the stepwise elongation of the hexasaccharide. Binding studies revealed that sulfation of all three tyrosine residues of PSGL-1 increases the binding affinity to P-selectin by 40-fold. Crystal structures of P-selectin and fragments of PSGL-1 provided additional insight into the binding. The research group of Wong has also synthesized sulfated analogs of PSGL-1 (Koeller et al., 2000).

Instead of building a glycan through the stepwise addition of monosaccharide residues by enzymatic transfers, another approach is the single transfer of a fully assembled glycan. Endo-β-N-acetylglucosaminidases (ENGases) are a family of enzymes responsible for the hydrolytic cleavage of the glycosidic bond in the N,N′-diacetylchitobiose moiety of N-linked oligosaccharides (Yamamoto, 2001). Several of these ENGases have also been shown to possess transglycosylation activity and are able to transfer a releasing N-glycan to a GlcNAc-peptide acceptor, forming a new β-1,4-glycosidic bond (Fan et al., 1995). The use of ENGases is appealing since one enzyme transfers a complex oligosaccharide donor to an acceptor in a single step, thus circumventing the tedious use of several glycosyltransferases to successively build up the glycan. Especially, the ENGase from Arthrobacter protophormiae, the Endo-A, and Endo-M isolated from Mucor hiemalis have been utilized for glycopeptide synthesis. Endo-A specifically transfers high-mannose type N-glycans (Wang et al., 1997; Singh et al., 2003), whereas Endo-M displays broader substrate specificity with a preference for complex-type N-glycans (Kadowaki et al., 1991; Yamamoto et al., 1994). Endo-M was used to transfer a disialo complex-type oligosaccharide from human transferrin to a GlcNAc glycopeptide acceptor (Scheme 11) in the synthesis of eel calcitonin, a 32-amino acid residue calcium-regulating hormone (Mizuno, Haneda, et al., 1999). The power of this method was recently demonstrated by the synthesis of C34, a 34mer glycopeptide that is derived from the transmembrane envelope glycoprotein gp41 of human immunodeficiency virus (HIV) type 1 (Wang et al., 2005). Using Endo-A, oligosaccharide donor ManαGlcNAcβAsn isolated from soybean flour was transferred to the GlcNAc-C34 glycopeptide acceptor in a yield of 11% (Scheme 12). The glycopeptide was used to study the effect of glycosylation on the inhibitory activity and conformation of C34.

Despite its attractiveness, the transglycosylation approach is hampered by low transglycosylation yields, risk of product hydrolysis, and the fact that only natural N-glycans can be used as donor substrates. To overcome these problems, Wang and coworkers recently showed that ENGase-catalyzed transglycosylation could indeed be achieved using oligosaccharide oxazolines as glycosyl donors.
donors (Scheme 13) (Li et al., 2005). The oxazolines were used to mimic a presumed oxazolinium ion intermediate formed during the transfer reaction. Wang’s approach gave improved yields in the range of 70–80%, but also a broader range of substrates could be used as glycosyl donors, indicating that oligosaccharide oxazolines are more kinetically favored substrates than the natural \( N \)-glycans. The method using oxazoline oligosaccharides in ENGase-catalyzed transglycosylations shows excellent promise for total glycoprotein synthesis and remodeling.

Native Chemical Ligation

Despite the tremendous power of SPPS, the routine synthesis is limited to <50 amino acids peptides. In the quest to reach synthetic glycoproteins, several new approaches have been developed.

NCL provides a powerful tool for stitching together unprotected glycopeptide and peptide fragments (Dawson and Kent, 2000). It is a chemoselective reaction that occurs at physiological pH between an \( N \)-terminal cysteine residue and a C-terminal peptide thioester (Scheme 14). The method was developed by Kent and coworkers in the 1990s (Dawson et al., 1994) and has since been further refined to
enhance its utility (Yeo et al., 2004). In the first step of NCL, a reversible trans-thioesterification reaction between the C-terminal thioester and the sulfhydryl group from the N-terminal cysteine residue takes place. The ligated single peptide thioester then undergoes a rapid, irreversible, and spontaneous intramolecular S → N shift, which generates the thermodynamically favored native amide bond at the ligation junction. NCL occurs uniquely at an N-terminal cysteine residue regardless of any additional internal cysteine residues and, as this ligation method is compatible with both carbohydrates and peptides, provides access to glycoprotein structures.

Preparation of peptide thioesters initially relied on the Boc strategy due to the inherent base sensitivity of the thioester (Hojo and Aimoto, 1991; Schwabacher and Maynard, 1993; Canne et al., 1995; Hackeng et al., 1997; Zhang and Tam, 1997). However, the widespread use of the Fmoc strategy for peptide synthesis and the acid sensitivity of glycopeptides have stimulated the development of new methodologies for the preparation of peptide thioesters. For example, milder α-N-Fmoc-deprotection cocktails have been used to suppress the base-induced decomposition of thioesters (Li et al., 1998). Other approaches include the use of novel linkers combined with the generation of the α-thioester after complete peptide synthesis (Alsina et al., 1999; Ishii et al., 2002; Brask et al., 2003; Camarero et al., 2004). Also, Lewis acid-catalyzed thiolysis of ester linkers has been used (Swinnen and Hilvert, 2000; Sewing and Hilvert, 2001). One of the most attractive methods is based on Kenner’s safety-catch sulfonamide linker (Kenner et al., 1971), which allows the assembly of the peptide by standard Fmoc chemistry (Scheme 15) (Ingenito et al., 1999; Shin et al., 1999). In their synthesis of dipterisin, an 82 amino acids glycoprotein with anti-microbial properties, Bertozzi and coworkers employed this approach (Scheme 16) (Shin et al., 1999). Using the sulfonamide linker, the N-terminal glycopeptide thioester was assembled by applying Fmoc chemistry. A two-step procedure was then used to cleave the glycopeptide from the resin. First, the linker is activated by an alkylating reagent such as iodoacetanilide (Backes et al., 1996; Backes and Ellman, 1999) and then treated with 6 M Gn-HCl, 0.1 M sodium phosphate, pH 7.5, 4% thiophenol.
with a thiol nucleophile, in this case benzyl mercaptan, which liberates a benzyl α-thioester peptide. The thioester is thus formed after the peptide chain has been fully assembled avoiding repeated base exposure. The thioester fragment was then ligated with the 58 amino acids C-terminal glycopeptide that carried an N-terminal Cys residue in aqueous guanidinium buffer. Thiophenol was used as a reaction rate-enhancing additive (Dawson et al., 1997), allowing for an isolation of the glycopeptide in 55% yield.

The inertia of the sulfonamide linker and the necessity of two-step time-consuming protocol render control of a growing glycopeptide laborious. To overcome this problem, Unverzagt recently developed an orthogonal double linker based on the sulfonamide linker (Mezzato et al., 2005). In addition to the thioester-generating linker, a second handle, the Rink-amide linker, was introduced. The Rink-amide linker allowed easy detachment of the glycopeptide chain for step-by-step quality control during synthesis, and activation and thiolysis of the sulfonamide linker gave an α-thioester, which was used in NCL to furnish a fragment of RNase B carrying a complex-type N-glycan.

Tandem NCL was introduced to prepare even larger proteins. By introducing a peptide derivative that carries a C-terminal thioester and an N-terminal cysteine with an orthogonal thiol-protecting group, a serial ligation reaction can be performed (Scheme 17) (Clayton et al., 2004). The multiple ligation routes have been exploited with great success for the total chemical synthesis of proteins or protein modifications. In a noteworthy example, a synthetic derivative (synthetic erythropoiesis protein [SEP]) of human erythropoiesis protein (Epo) with 166 amino acid residues was synthesized from four peptide blocks by tandem NCL (Scheme 18) (Kochendoerfer et al., 2003). The glycans of human Epo, which are critical for the in vivo half-life of the protein (Elliott et al., 2003; Koury, 2003), were exchanged for synthetic polyethylene glycol chains to simplify the synthesis. Gratifyingly, it was shown that the appended polymer chains prolonged the half-life in vivo and hence the biological activity.

Easy access to α-thioesters is crucial to NCL, and although several methods have been developed for their preparation, it is not always a simple task. Furthermore, despite the promise of tandem NCL, the inherent size limitations associated with glycopeptide preparation pose a limiting factor affecting the size of glycopeptides and glycoproteins that can be reached by fully synthetic means.

Expressed protein ligation (EPL) is a technique that produces recombinant α-thioesters (Muir, 2003). EPL is based on the biochemical mechanism of protein splicing, which shares some critical features with NCL. In protein splicing, a precursor protein undergoes an intramolecular rearrangement, which results in the excision of an internal segment, an intein, and the joining of the two flanking C- and N-terminal regions, the exteins (Scheme 19). This process, which is catalyzed by inteins and results in a native amide bond, has been shown to involve a thioester intermediate. Genetically modified inteins have been designed which block the final step in protein splicing. Thiolysis of a protein that has been expressed fused with a modified intein thus results in a partially processed recombinant protein that is tagged with an α-thioester at the C-terminal. The produced thioester protein can be submitted to routine NCL and be ligated with a synthetic glycopeptide to give a desired glycoprotein.

In essence, this combined “semi-synthetic” approach increases the size of glycoproteins that are accessible by NCL.

Recently, EPL was utilized for the production of three distinct glycoforms of GlyCAM-1, which is a ligand for the leukocyte adhesion molecule L-selectin (Macmillan and Bertozzi, 2004). In a most impressive semi-synthetic approach, a cysteine-containing synthetic peptide, two synthetic α-thioester glycopeptides (one of which carried a masked N-terminal cysteine thiol), and a recombinant α-thioester peptide which carried an N-terminal cysteine thiol masked by a protease recognition peptide were stitched together in a serial fashion to produce a defined glycoform of GlyCAM-1 carrying 13 GalNAc residues (Scheme 20).

Since the discovery of NCL, several glycoproteins have been chemically synthesized. The reactions that involve NCL are mild, selective, and compatible with the presence of glycans. NCL and EPL provide powerful tools for the incorporation of nonnatural amino acids or biophysical probes into glycoproteins (Dawson and Kent, 2000; Muir, 2003). In addition, it is possible to incorporate isotopic labeling at specific locations. These ligation reactions can aid in the preparation of homogeneous posttranslationally modified proteins, which otherwise are inaccessible with rDNA techniques. By choosing to use EPL, it is possible to overcome some of the synthetic challenges associated with the synthesis of large glycoproteins and a glycoprotein of large bulk can be accessed.

However, another limiting factor for NCL is its intrinsic reliance on a cysteine residue at the ligation juncture. Cysteine
comprises only 1.7% of all residues in proteins, rendering the majority of proteins as unable to be prepared in their native form by NCL. For example, in the semi-synthesis of GlyCAM-1 outlined in Scheme 20, cysteine moieties had to be introduced into the peptide sequence at strategic positions to allow NCL. Efforts to find reliable alternative acyl transfer auxiliaries that would replace cysteine have been reported and are continuously being constructed (Kimmerlin and Seebach, 2005). Also, several alternative ligation methods resulting in natural and unnatural bond formations have been developed (Lemieux and Bertozzi, 1998; Tam et al., 1999, 2001; Kimmerlin and Seebach, 2005). The Staudinger ligation (Saxon et al., 2000; Saxon and Bertozzi, 2000), which has been called a gift to chemical biology, is based on the classical Staudinger reaction (Kohn and Breinbauer, 2004). It is a reaction between an azide and a phosphine, which yields an aza-ylide that in turn can undergo hydrolysis in the presence of water to give an amine and a phosphine oxide. The aza-ylide is highly electrophilic and susceptible to attack by a range of nucleophiles.

The Staudinger reaction serves as a powerful tool for the ligation of glycopeptides because a cysteine residue is not required at the ligation site, and a natural amide bond is formed at the ligation junction (Nilsson et al., 2000, 2001). The Staudinger reaction and NCL have also been utilized in tandem ligation reactions (Nilsson et al., 2003). However, as is the case with NCL, these methods have yet to be extensively tested for their compatibility with the semi-synthesis of proteins and glycoproteins.

Applications of Synthetic Glycopeptides in Biological and Structural Studies

It is unclear whether glycosylation greatly influences both the overall structure of a protein and a more local conformation. By synthesizing a range of N-linked hemagglutinin-derived glycopeptides differing in molecular composition, Imperiali...
and coworkers could show that the two \(N\)-acetamido groups of the conserved GlcNAcGlcNAc core in \(N\)-linked glycans induce a major conformational alteration in the peptide backbone (Figure 4) (O’Connor and Imperiali, 1998). Indeed, the two \(N\)-acetamido groups were shown to be essential for inducing a well-defined secondary structure with a type I \(\beta\)-turn in the peptide backbone (Figure 5). \(\beta\)-Turn formation is believed to be critical for the protein folding process (Wright et al., 1988), and ~30% of all glycosylation sites occur at sites that fold ultimately to \(\beta\)-turns (Beintema, 1986). In a more recent study, Imperiali and Woods compared the conformational effect induced by corresponding \(\alpha\)- and \(\beta\)-linked glycans (Bosques et al., 2004). Two glycopeptides carrying an \(\alpha\)- or a \(\beta\)-linked chitobiosyl moiety, and with an incorporated Asn-Gly-Thr motif, were compared with the corresponding unglycosylated peptide. NMR studies revealed that the \(\alpha\)-linked glycopeptide adopted a conformation that was similar to the unglycosylated peptide, which has an Asx-turn type structure, while the \(\beta\)-linked glycopeptide adopted a type I \(\beta\)-turn in the peptide sequence around the saccharide (Figure 5). The three glycopeptides were in parallel subjected to computational analysis involving molecular dynamic (MD) simulations using explicit water solvation. The MD simulations were in excellent agreement with the results from the NMR studies. The results from this study provide valuable information for the design of glycopeptide mimetics as well as a tested computational approach useful for the prediction of the influence saccharides exert on the peptide backbone.

Since their discovery in the mid-1970s, the endogenous opioid peptides, endorphins, have attracted much attention from medical researchers (Polt et al., 2005), as it was recognized that their therapeutic potential is tremendous. Opioid peptide receptor could potentially affect selective pharmacological intervention without the side effects, such as tolerance, physical dependence, and addiction that are associated with the classical opiates such as morphine. However, the instability of these peptides in serum and their inability to effectively penetrate the blood–brain barrier and thus enter the brain and spinal cord have severely hampered their use as analgesics. It has been demonstrated that glycosylation of these neuropeptides, which renders them more hydrophobic, in fact provides a means to transport the peptides across the lipophilic blood–brain barrier (Polt et al., 2005). It has also been shown that if properly placed, the glycosyl moiety will not interfere with the receptor binding of the native peptide. Polt and coworkers synthesized and studied \textit{in vivo} the opioid peptide sequence DTLES, which is a \(\delta\)-selective agonist, and compared the effect of glycosylation of the C-terminal serine using different monosaccharides and disaccharides (Figure 6). The glycopeptide analogs are quite potent \(\delta\)-agonists, but they still possess an appreciable \(\mu\)-activity. It was found that glycosylation with \(\beta\)-D-glucose led to a significant increase in enzymatic stability in both serum and brain and increased blood–brain barrier permeability compared to the unglycosylated peptide (Bilsky et al., 2000; Egleton et al., 2001). In addition, the glycosylated analogs were a 100- to 200-fold more potent than morphine (Elmagbari et al., 2004). Further studies in mice showed that analgesia induced by the glycopeptides did not result in side effects that are commonly observed with morphine-induced analgesia. Thus, the strategy of glycosylating endogenous neuropeptides shows great potential for the development of receptor selective drugs that can be transported across the blood–brain barrier.

Glycopeptides are useful as mimetics of natural carbohydrate ligands. In fact, in many cases, glycopeptides bind with higher affinity to a receptor than the natural carbohydrate. For example, glycopeptides that mimic carbohydrate ligands and can bind with high affinity have been found for the P-, E-, and L-selectins (Sears and Wong, 1999). A glycopeptide that mimics SLe\(^x\) has been found to have a 10-fold increase in affinity for binding to E-selectin (Lin et al., 1996). Also, the assembly of glycopeptides is a considerably easier task than the assembly of complex oligosaccharides,
and it is feasible to synthesize glycopeptides in a combinatorial library format to identify high-affinity ligands.

In their seminal work, Meldal and coworkers synthesized a one-bead-one-compound library of 300,000 glycopeptides that could be screened directly on-bead to find high-affinity ligands for the C-type lectin from *Lathyrus odoratus* (St Hilaire et al., 1998). The library was assembled using a combination of “chemical encoding and ladder synthesis” (Youngquist et al., 1995) together with a photolytic linker on a PEG-based resin. In the ladder approach, a capping agent is introduced in the stepwise synthesis of the glycopeptide, which gives sequentially terminated glycopeptides together with the full-length glycopeptide ligand on one single bead. Different carboxylic acid capping agents that specifically encoded for different glycosyl amino acids were used for easy identification of the glycans. As opposed to polystyrene-based resins, the PEG-based resins (Meldal, 1997) show excellent swelling properties in aqueous media, characteristics that allow biomolecules access to the entire bead. Consequently, it was possible to use rapid solid-phase screening of the lectin binding. A fluorescent label was used in the binding assay, and the active glycopeptides could be released from the beads by irradiation with the MALDI-TOF laser. The ladder of glycopeptide fragments and the chemical encodings facilitated the structural elucidation of the active glycopeptide ligands.

Glycopeptide antigens are unique targets for the development of immunotherapy against cancer, given the recognition that in the event of cell malignancies, dramatic changes occur in the nature and abundance of protein- and lipid-linked cell-surface oligosaccharides (Feizi and Childs, 1985; Lloyd, 1987; Springer, 1997; Hakomori, 1998). This abnormal glycosylation is associated with tumor progression and strongly correlates with poor survival rates. Mucins, which carry truncated glycans such as the tumor-associated T antigens, are associated with breast, colon, and prostate cancers (Springer, 1997). Immunizations with synthetic glycopeptides could evoke an immune response that is directed toward cancer cells, which otherwise are ignored by the immune system. The antibodies raised in such an immune response may be exploited for the treatment of a “minimal residual disease.” In this respect, they could target a small
number of metastasized cells that may have persisted after primary therapies such as surgery or chemotherapy. This add-on immunotherapy could protect cancer patients against relapse and, thus, enhance survival rates. This approach is, however, associated with several complicating factors. Synthetic antigens are commonly of a low molecular weight, which renders them poor immunogens. Also, the inherently T-cell-independent nature of carbohydrates complicates their use in vaccines (Kuberan and Linhardt, 2000). To overcome these problems, research has focused on the conjugation of glycopeptides and saccharides to carrier proteins (Danishefsky and Allen, 2000). This approach is, however, also not free of complications and has so far failed to induce sufficiently strong helper T-cell responses in human patients. To this end, Kunz and coworkers have reported the chemical synthesis of several well-defined glycopeptides based on the Tn, TF, and STn mucin core structures (Brocke and Kunz, 2002; Dziadek and Kunz, 2004), which in addition to a B-epitope glycopeptide antigen, contain a promiscuous helper T-cell epitope that can induce a tumor-selective T-cell-dependent immune response (Keil et al., 2001). A construct based on the tumor-associated MUC-1 glycopeptide antigen and a Tetanus toxin T-cell epitope was assembled by block synthesis on solid support (Scheme 21). The MUC-1 glycopeptide was synthesized using the allylic HYCRON anchor (Seitz and Kunz, 1997) and a combination of Fmoc and Boc chemistry. The STn-Thr building block with its fully acetylated saccharide was incorporated into the peptide backbone using TBTU and HBTU as coupling reagents. The T-epitope peptide was assembled separately with a flexible ethylene glycol spacer at the C-terminus and was then coupled to the N-terminus of the HYCRON-linked glycopeptide in a HOAt/HATU-mediated condensation. The purified conjugate was then applied to peripheric blood lymphocytes for immunological evaluation. IFNγ was produced by cells from three of four donors, and a measurement of cell proliferation showed that it proceeded only in the presence of antigen-presenting cells, offering a distinct proof of an antigen-specific reactivity. In a more recent article, the same group reports the synthesis and immunological evaluation of a similar vaccine candidate (Dziadek et al., 2005). A different helper T-cell
epitope derived from ovalbumin was incorporated in the vaccine construct linked to the STn-glycosylated MUC-1 peptide B-epitope. The vaccine candidate was subsequently used to immunize transgenic mice with a receptor specific for the ovalbumin helper T-cell epitope. High titers of anti-MUC-1 antibodies were detected in one of the mice, and it was shown by neutralization experiments that the elicited antibodies had a high specificity for the glycosylated MUC-1 peptide and low specificity for the corresponding unglycosylated peptide or a glycosylated or unglycosylated peptide derived from MUC-4.

In several elegant studies, multiple antigen glycopeptides (MAGs) have been explored for the use in cancer immunotherapy (Lo-Man et al., 1999; Vichier-Guerre et al., 2003). The tetraantennary vaccine candidates, in which each arm incorporates a cluster of the Tn antigen and a helper T-cell peptide, were synthesized in solid phase on a nonimmunogenic core based on β-alanine and lysine (Figure 7) (Lo-Man et al., 2001). In a recent study, it was shown that the MAG strategy resulted in higher titers of elicited specific immunoglobulin G (IgG) antibodies in a mild adjuvant setting and thus has better efficacy compared to the traditional protein-conjugate method (Lo-Man et al., 2004). The vaccine candidate induced, in both mice and non-human primates, strong tumor-specific anti-Tn antibodies that recognized and mediated antibody-dependent cell cytotoxicity against human tumor cells.

In a program aimed at the development of cancer vaccines (Buskas et al., 2004; Buskas, Ingale, et al., 2005; Buskas, Li, et al., 2005), we synthesized a lipidated glycopeptide that incorporates the minimal structural features required for a focused immune response (Buskas, Ingale, et al., 2005). The vaccine candidate contained the tumor-associated Tn antigen, which functioned as a B-epitope, and a peptide derived from E. coli, which provided a universal T-epitope for T-cell activation. The lipopeptide Pam3Cys functioned as a built-in immunoadjuvant. The lipidated glycopeptide required a highly convergent synthesis and could be assembled by a combination of solution and solid phase synthesis (Scheme 22). The fully protected lipidated peptide was assembled on solid support using the hyper-acid-sensitive linker HMPB-MBHA. The tumor-associated Tn antigen was then coupled via a linker to the C-terminus of the fully protected lipidated peptide in solution phase by means of DIC and HOAt. The lipidated glycopeptide was then incorporated into phospholipid-based liposomes for the immunization of mice. To explore the adjuvant properties of the built-in Pam3Cys, the vaccine candidate was administered with and without the powerful external immunoadjuvant QS-21. Specific IgG antibodies were elicited against the tumor antigen, both in the presence and in the absence of the external adjuvant, which provides proof-of-principle evidence that lipidated glycopeptides can evoke a T-cell-dependent immune response and thus may be effective as minimal subunit vaccines.

Scheme 21. Synthesis of a cancer-vaccine candidate consisting of a T-cell epitope derived from Tetanus toxin and a tumor-associated glycopeptide antigen from MUC-1. (i) HOAt, HATU, NMM, DMF. (ii) Ac2O, pyridine 1/3. (iii) Pd(PPh3)4, morpholine, DMF/DMSO, 1/1. (iv) Pd/C, H2, MeOH. (v) DCM/TFA/thioanisole/EDT, 10/10/1/1. (vi) NaOMe, MeOH, pH 8.5.

Fig. 7. Schematic representation of a multiple antigen glycopeptide cancer-vaccine candidate containing a tumor-associated Tn-antigen cluster and a universal T-cell epitope.
Concluding Remarks

The methods for the chemical synthesis of glycopeptides have advanced considerably, and peptides carrying relatively simple saccharides can now be prepared by routine procedures. Enzymatic procedures have been employed to extend these simple saccharides into complex oligosaccharides. Furthermore, glycopeptides containing elaborate oligosaccharides can also be obtained by the chemical synthesis of a properly protected amino acid containing a complex oligosaccharide, which can then be used in traditional polymer support glycopeptide synthesis. This approach, however, requires highly specialist knowledge of chemical synthesis of complex oligosaccharides. The advent of NCL makes it possible to synthesize glycoproteins from well-designed peptide and glycopeptide building blocks. Synthetic glycopeptides and glycoproteins are becoming increasingly interesting derivatives for biological and structural studies.

Acknowledgments

The authors are grateful for financial support for this work by the National Cancer Institute of the National Institutes of Health (Grant RO1 CA88986).

Conflict of interest statement

None declared.

Abbreviations

β-D-GlcNAc, N-acetyl-β-D-glucosamine; CHO, Chinese hamster ovary; CS, chondroitin sulfate; DCM, dichloromethane; ENGase, endo-β-N-acetylglucosaminidase; EPL, expressed protein ligation; ER, endoplasmic reticulum; GPI, glucosylphosphatidylinositol; HS, heparan sulfate; NCL, native
chemical ligation; PSGL-1, glycopeptide P-selectin ligand-1; SPPS, solid-phase peptide synthesis; TFA, trifluoroacetic acid.

References


Haltiwanger, R.S. (2002) Regulation of signal transduction pathways in 133R T. Buskas


Haltiwanger, R.S. (2002) Regulation of signal transduction pathways in 133R T. Buskas


