Chemoenzymatic approaches to glycoprotein synthesis

Clay S. Bennett and Chi-Huey Wong*

Received 9th April 2007
First published as an Advance Article on the web 14th May 2007
DOI: 10.1039/b617709c

The construction of homogeneous glycoproteins presents a formidable challenge to the synthetic chemist. Over the past few years there has been an explosion in the number of methods developed to address this problem. These methods include the development of novel ligation technologies for the synthesis of the protein backbone, as well chemical and enzymatic approaches for introducing complex glycans into the peptide backbone. This *tutorial review* discusses the application of these techniques to the synthesis of peptides and proteins possessing well defined glycans.

Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California, 92037, USA. E-mail: wong@scripps.edu; Fax: +1-858-784-2409; Tel: +1-858-784-2487

1 Introduction

Posttranslational modification of proteins allows eukaryotic organisms to greatly expand the functional diversity of their proteomes. Of all the posttranslational modifications identified

Clay S. Bennett

Dr Clay S. Bennett received his BA from Connecticut College in 1999, and his PhD in chemistry from the University of Pennsylvania with Prof. Amos B. Smith, III in 2005. He is now working as a postdoctoral fellow with Prof. Chi-Huey Wong. His current research is concerned with the development of new methods for glycoprotein synthesis and the study of the effect of glycosylation on protein function.

Dr Wong received his BS and MS degrees from National Taiwan University, and PhD in Chemistry (with George M. Whitesides) from Massachusetts Institute of Technology in 1982. He then moved with Professor Whitesides to Harvard University as a postdoctoral fellow for another year. He started his independent career as Assistant Professor of Chemistry at Texas A&M University in 1983, became Associate Professor in 1986 and Professor in 1987. He was Professor and Ernest W. Hahn Chair in Chemistry at the Scripps Research Institute (1989–2006) and Director of the Genomics Research Center at Academia Sinica, Taipei (2003–2006). Since October 2006, he has been Professor of Chemistry at The Scripps Research Institute and President of Academia Sinica.

Professor Wong is a recipient of The Searle Scholar Award in Biomedical Sciences (1985), the Presidential Young Investigator Award in Chemistry (1986), the American Chemical Society A.C. Cope Scholar Award (1993), the Roy Whistler Award of the International Carbohydrate Organization (1994), the American Chemical Society Harrison Howe Award in Chemistry (1998), the American Chemical Society Claude S. Hudson Award in Carbohydrate Chemistry (1999), the International Enzyme Engineering Award (1999), the Presidential Green Chemistry Challenge Award (2000), and the American Chemical Society Award for Creative Work in Synthetic Organic Chemistry (2005).

Chi-Huey Wong

He is a member of Academia Sinica, Taipei (1994), the American Academy of Arts and Sciences (1996), and the U.S. National Academy of Sciences (2002). He serves as Editor-in-Chief of Bioorganic & Medicinal Chemistry and is a Chairman of the Executive Board of Editors of Tetrahedron Publications. He is also an editorial advisor of the Journal of American Chemical Society, ChemBiochem, Advanced Synthesis and Catalysis, Current Opinion in Chemical Biology, and Chemistry—An Asian Journal. He was head of the Frontier Research Program on Glycotechnology at RIKEN (Institute of Physical and Chemical Research, Japan, 1991–1999), and a board member of the U.S. National Research Council on Chemical Sciences and Technology (2000–2003). He is currently holding a joint appointment as Distinguished Professor of Chemistry and Biochemical Sciences at National Taiwan University, a scientific advisor of the Max-Planck Institute, and is a founding scientist and scientific advisory board chairman of Optimer Pharmaceuticals, Inc. His research interests are in the areas of bioorganic and synthetic chemistry and biocatalysis, including development of new synthetic chemistry based on enzymatic and chemo-enzymatic reactions, synthesis of complex carbohydrates, glycoproteins and small-molecule probes for the study of post-translational glycosylation and carbohydrate-mediated biological recognition, drug discovery, development of oligosaccharide microarrays for high-throughput screening and study of reaction mechanism. He is the author and co-author of over 500 publications, 60 patents, and four books (Enzymes in Synthetic Organic Chemistry, Combinatorial Chemistry in Biology, Catalysis from A to Z, and Carbohydrate-Based Drug Discovery).

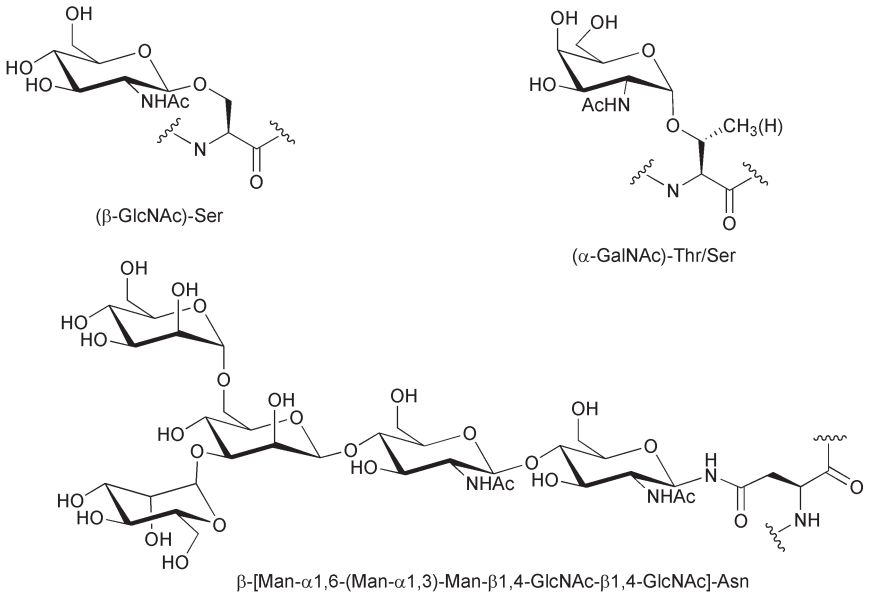

Fig. 1 Common *O*-linked serine (Ser) and threonine (Thr) linked glycosides and *N*-linked core asparagine (Ans) glycans found in glycoproteins.

to date, the attachment of oligosaccharides to proteins, referred to as protein glycosylation, is the most complex. Glycosylation has been implicated in a variety of different biological processes, including macromolecular recognition, cell adhesion, and protein folding.[1,2] In addition aberrant glycosylation has been associated with a number of pathophysiological conditions ranging from autoimmune disease to cancer.[2,3] Consequently, protein glycosylation could play a major role in the development of new drug and vaccine-based therapeutics.

Despite the importance of protein glycosylation, little is known about the molecular basis of its function. This is due in large part to the inherent complexity of the glycan structure. Oligosaccharides attached to proteins consist of highly branched biopolymers most commonly attached to the peptide backbone through the amide nitrogen of asparagine (*N*-linked) or the oxygen on serine or threonine (*O*-Linked) (Fig. 1). Branching arises because attachment of each additional monosaccharide to the glycan can take place at one of 4 unsubstituted hydroxyl groups of the previous sugar in the chain with either α- or β-stereochemistry. As a result, a simple library of tetrasaccharides formed from the nine most common monosaccharides yields over 15 million possible combinations. For purposes of comparison, the maximum possible numbers of tetramers of linear peptides or nucleic acids are 160,000 and 256, respectively.

The situation becomes even more complicated when one considers that unlike peptides and nucleic acids, there is no biological template for oligosaccharides. Instead, the type and extent of protein glycosylation is dependent on a variety of enzymes, which often compete with each other for a given acceptor. Additionally, the production of these enzymes is highly dependent on a number of factors including the local physiological environment and extracellular stress. As such, glycoproteins consist of heterogeneous mixtures known as glycoforms, in which the same peptide backbone can present dozens of different glycan structures. Small differences in molecular weight and charge between different glycoforms frequently make isolation of homogeneous glycoproteins difficult to impossible. Accordingly, synthesis is currently the only avenue for the production of homogeneous glycoproteins for study.

Synthesis of homogeneous glycoproteins presents three major obstacles: 1) obtaining homogeneous glycans, 2) efficient production of the protein backbone, and 3) coupling of the glycan to the protein backbone. Since synthetic approaches to complex oligosaccharides has been reviewed extensively,[4] this review will focus on methods for construction of the protein backbone and attachment of complex glycans through native linkages. Over the past several years a number of methods employing both chemical and enzymatic approaches have been developed to address these problems.[5,6] This review will focus on recent major developments in glycoprotein synthesis.

2. Synthesis of the protein backbone

2.1 Glycopeptide synthesis

A significant challenge in glycoprotein synthesis is the construction of peptide backbones containing glycosylation at all of the desired positions. The most common approach for glycopeptide synthesis utilizes solid phase peptide synthesis (SPPS) with preformed glycosyl amino acids. When performing SPPS on glycopeptides, the proper choice of protecting groups is essential, as *O*-glycosidic linkages are labile to strongly acidic or basic conditions. Fortunately, glycosidic linkages are stable to the weakly basic conditions employed in 9-fluorenylmethoxycarbonyl (Fmoc) based SPPS chemistry, allowing for routine preparation of glycopeptides containing small glycans.

The situation becomes more complex when attempting to carry out SPPS with amino acids carrying large oligosaccharides. As a result of the increased steric bulk of these glycans,

SPPS requires the use of large excesses of the glycosylated amino acids and much longer reaction times. One strategy that is commonly employed to address this problem is to use amino acids that contain mono- or disaccharides in SPPS and then enzymatically elaborate the glycans following cleavage of the glycopeptide from the resin (*vida infra*). While this is a very powerful method for the construction of peptides containing complex glycans, there are certain glycosidic linkages for which the requisite enzyme is not readily available. An alternative employs microware irradiation to accelerate SPPS with amino acids containing complex glycans. Using this technique, Nishimura and colleagues were able to efficiently carry out SPPS using Thr and Ser amino acids possessing protected trisaccharides.[7]

2.2 Native chemical ligation (NCL)

Construction of larger glycoproteins is a challenging problem. The maximum size peptide that can be practically synthesized using SPPS is around 50 amino acids. Above this length, epimerization, incomplete coupling and accumulation of side products results in decreased yields and an increased difficulty of final product purification. This has led to methods for the convergent coupling of large peptide fragments in order to form proteins. Due to the limited solubility of protected peptides in both aqueous and organic media, it is desirable to carry out the couplings using unprotected peptides, which are often freely soluble in water. The most common method for effecting such transformations is the native chemical ligation (NCL).[8]

NCL takes advantage of the fact that thioesters undergo rapid reversible exchange with the free thiol of cysteine at room temperature in aqueous media. In the presence of an *N*-terminal cysteine, the resulting thioester intermediate can then undergo a spontaneous and irreversible S → N acyl transfer, resulting in the formation of a new peptide bond (Fig. 2). Because the initial thioester formation is reversible, it is not necessary to protect other free cysteines in the peptides; only the *N*-terminal cysteine will undergo the S → N acyl transfer.

While NCL is a powerful method for glycoprotein synthesis, the inherent instability of peptide thioesters makes them difficult to synthesize and handle. Although these compounds can be routinely prepared using *tert*-butoxycarbonyl (Boc) based SPPS, the acidic conditions required for removal of the Boc protecting group (HF) makes this approach incompatible with glycosidic linkages. An interesting solution to this problem was introduced by Bertozzi and coworkers in their synthesis of the 82-residue antimicrobial glycoprotein diptericin.[9]

Bertozzi's approach to diptericin involved breaking the protein into a 24-mer *N*-terminal glycopeptide thioester and a 58-mer *C*-terminal glycopeptide. Since native diptericin does not contain any cysteine residues, it was necessary to introduce one in place of glycine 25 in order to furnish a handle for NCL. While the *C*-terminal sequence was readily synthesized using SPPS, synthesis of the *N*-terminal fragment proved to be problematic due to thioester hydrolysis during conventional Fmoc based SPPS. In order to circumvent this problem Bertozzi and coworkers used Ellman's sulfonamide

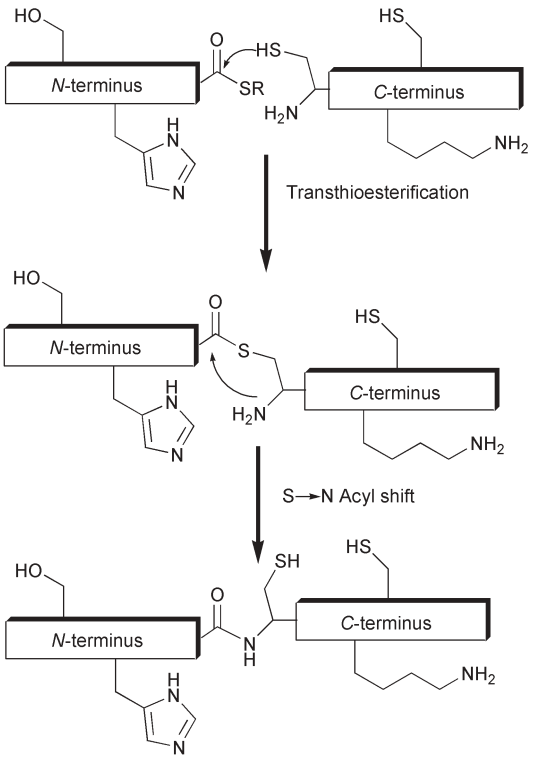

Fig. 2 Mechanism of NCL.

"safety-catch" linker, which prevents release of the peptide from the resin until it is activated by alkylation of the sulfonamide nitrogen (Fig. 3). Using this technique the Bertozzi group was able to efficiently prepare the desired glycosylated thioester. With both coupling partners in hand, NCL proceeded smoothly in the presence of thiophenol, an additive known to accelerate the reaction,[8] to furnish the desired target in good yield. More recently Unverzagt and coworkers have demonstrated that the "safety-catch" approach is compatible with thioesters bearing complex *N*-linked glycans.[10]

Danishefsky and coworkers have developed an alternative method for overcoming the problem of thioester instability. This approach utilizes a peptide phenolic ester with an *ortho* disulfide moiety to act as a protected thioester (Fig. 4).[11] The phenolic ester can be generated through solution-phase coupling of a peptide (generated using SPPS) with a protected 2-mercaptophenol. Upon treatment with a reducing agent such as sodium 2-mercaptoethanesulfonate (MES-Na) or triscarboxyethyl phosphine (TCEP), the disulfide moiety is reduced to reveal a free thiol capable of reversible O → S acyl transfer to generate a thioester *in situ*. In the presence of an appropriate acceptor, the thioester intermediate can then undergo NCL. Danishefsky and coworkers have used this method extensively in the synthesis of glycopeptides bearing multiple glycosylation sites.[12]

A second limitation of NCL is the need for cysteine at the ligation junction. This can be problematic as cysteine is relatively rare and frequently there is no cysteine in the target protein, or the cysteines are not located in a strategic position along the peptide backbone. In order to extend the scope of the reaction, Dawson and coworkers developed a method where

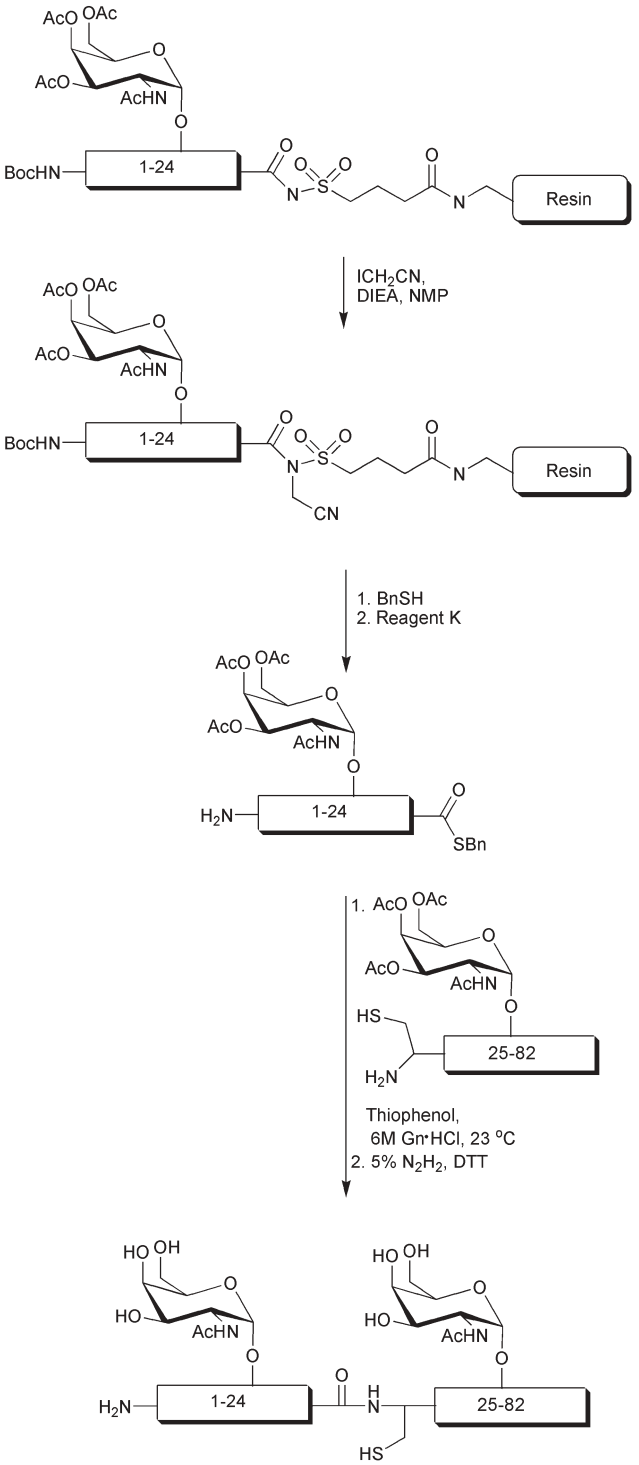

Fig. 3 NCL-mediated synthesis of the 82 amino acid glycoprotein diptericin.

cysteine is used as a masked alanine residue.[13] In this approach, following NCL the cysteine is converted to an alanine by hydrogenolysis using Rainey nickel. Although this tactic did extend the scope of NCL it was thought to suffer from the limitation that the target protein could not possess any cysteines. Very recently however, the Kent lab reported that protecting internal cysteines with acetamidomethyl (Acm) protecting groups allowed for selective desulfurization at the ligation junction. The Acm groups could then be removed using silver acetate to afford the target protein.[14]

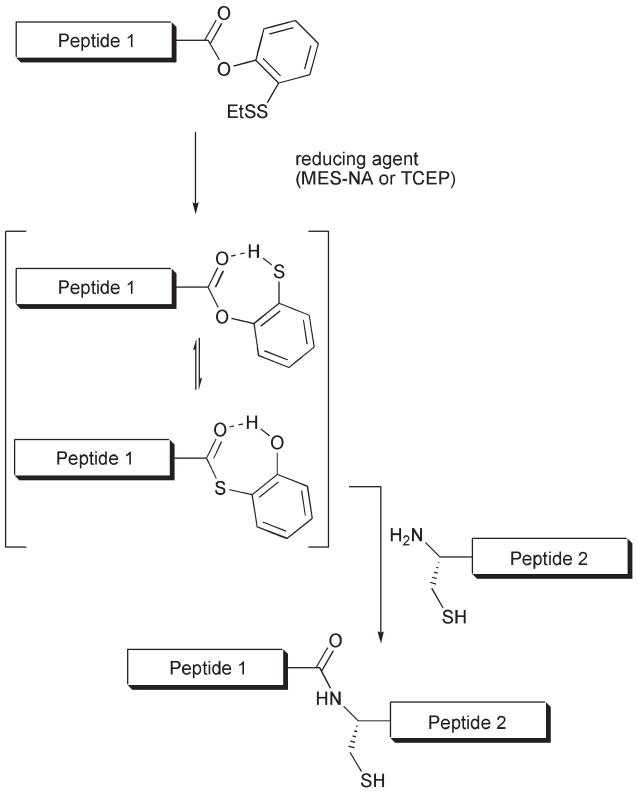

Fig. 4 Mechanism of Danishefsky's modified NCL using a masked glycopeptide thioester.

2.3 Auxiliary assisted NCL

In an effort to increase the scope of NCL to include ligation junctions which do not possess cysteine several groups have introduced removable auxiliaries which can act as cysteine surrogates. The most commonly used approach to auxiliary assisted NCL employs a benzylamine based auxiliary, which after NCL can be hydrolyzed under acidic conditions following methylation of the thiol to prevent N → S acyl transfer (Fig. 5).[15,16] Although these auxiliaries function in a fashion analogous to NCL, the increased steric bulk about the *N*-terminal amine nucleophile of the acceptor results in a decrease in the rate of the S → N acyl shift when the auxiliary is attached to an amino acid other then glycine.[15] As a result, in order to successfully employ the benzylamine based auxiliary in aqueous media, the ligation junction must possess one glycine (recently however, it was reported that it is possible to perform ligations at more hindered junctions by carrying out the reaction in organic solvent).[11] To date, both the Macmillan and Danishefsky groups have used auxiliaries of this type in glycopeptide synthesis.[12,16]

In studies aimed at further increasing the generality of the NCL, Wong and coworkers introduced a new auxiliary based on a *N*-(2-thioacetyl)-glucosamine scaffold.[17] This thiosugar, which can be attached to the peptide through either an *N*- or *O*-linkage, acts as a cysteine surrogate when placed on an amino acid one residue removed from the ligation junction

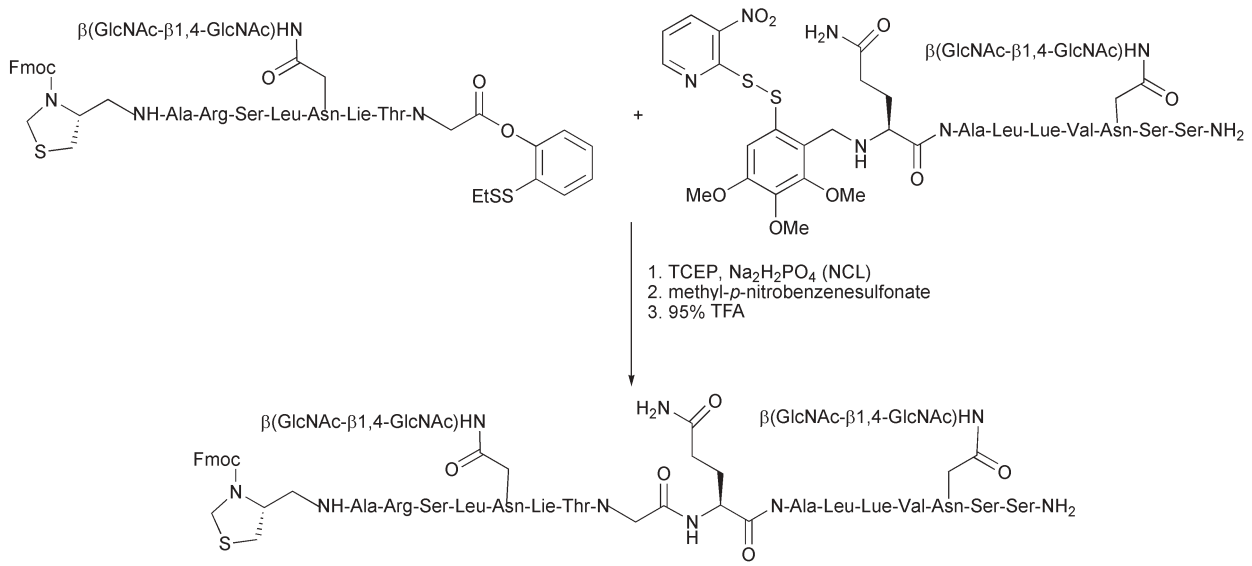

Fig. 5 Application of the benzylamine based auxiliary to glycopeptide synthesis.

(Fig. 6a). Despite the greatly increased ring size of the transition state (14–15 membered ring *vs*. 5–6 membered ring found in the NCL and benzylamine based auxiliary approaches), the ligations proceed efficiently and in good yield. Importantly, this is the only auxiliary-based method developed to date that has been shown to perform ligations in aqueous media without Gly at the ligation junction. The thiol handle can be removed using reductive desulfurization, following the ligation, to furnish the desired glycopeptide. Although the reductive desulfurization is incompatible with the presence of free cysteines, it has been demonstrated that cysteines which are protected with Acm groups are not affected by these conditions.[18]

While it is possible to protect the cysteines in peptides derived from SPPS, this approach can become impractical when dealing with large proteins possessing multiple cysteines. Additionally, protection of coupling partners derived though bacterial expression systems (*vide infra*) is generally not feasible. In order to address this issue the Wong lab has introduced an alternate scaffold for their sugar assisted ligation (SAL) based on appending a mercaptoacetate moiety on the 3-position of an *N*-acetylglucosamine attached to a peptide backbone (Fig. 6b).[19] In the presence of a mixture of *N*-methyl pyrrolidinone and HEPES buffer, this auxiliary undergoes smooth ligation with a thioester to afford the desired product in good yield. As with the *N*-(2-thioacetyl)-glucosamine scaffold, this system is tolerant of a number of possible ligation junctions. Even more importantly, it has been demonstrated that this auxiliary can be located up to five amino acids away from the ligation junction without deleterious effects on the efficiency of the reaction. As a result it is possible to use this method with sequences where the glycosylation site is flanked by bulky non-ligatable amino acids such as valine, leucine and proline.

2.4 Expressed protein ligation

For the synthesis of larger glycoproteins it is often advantageous to produce one of the two coupling partners recombinantly, in a process known as expressed protein ligation (EPL).[20] When applicable, this method has an advantage over NCL in that it is possible to obtain much larger coupling partners than with SPPS. Furthermore, the relatively facile production of many recombinant proteins renders this a cost effective approach. Methods have been developed for the recombinant production of both *C*-terminal and *N*-terminal coupling fragments, referred to as *N*-terminal engineering and *C*-terminal engineering, respectively.

In *N*-terminal engineering, a recombinant protein possessing an *N*-terminal protease recognition sequence followed by a cysteine is expressed in bacteria. After cleavage of the protease recognition sequence, the cysteine can undergo NCL with an appropriate thioester.[21] This proteolysis/mutagenesis strategy is useful for the synthesis of glycoproteins containing *N*-terminal glycosylation. In a recent example, the Wong group produced a chemically defined version of human interleukin-2 (Il-2) by first expressing the *C*-terminus as a fusion protein possessing a HIS tag and the tobacco etch virus protease (TEV protease) recognition sequence ENLYFQ.[22] Following Ni^{2+} affinity purification, the protein was treated with TEV protease to unmask the *N*-terminal cysteine. The protein then underwent NCL with a glycopeptide thioester under denaturing conditions to afford the desired product (Fig. 7a).

In order to introduce the glycopeptide into the *C*-terminus of the protein it is necessary to express a protein which possesses a *C*-terminal thioester. The intein technique developed independently by the Evans and the Muir groups is the current state of the art for achieving this transformation.[23,24] Inteins are natural protein splicing elements, analogous to the introns found in nucleic acids. The intein catalyzes its own excision through a series of acyl-transfer reactions in which a cysteine thioester is a key intermediate. By introducing an affinity sequence into the intein it is possible to isolate the thioester intermediate through immobilization. Upon treatment with a cysteine derivative in the presence

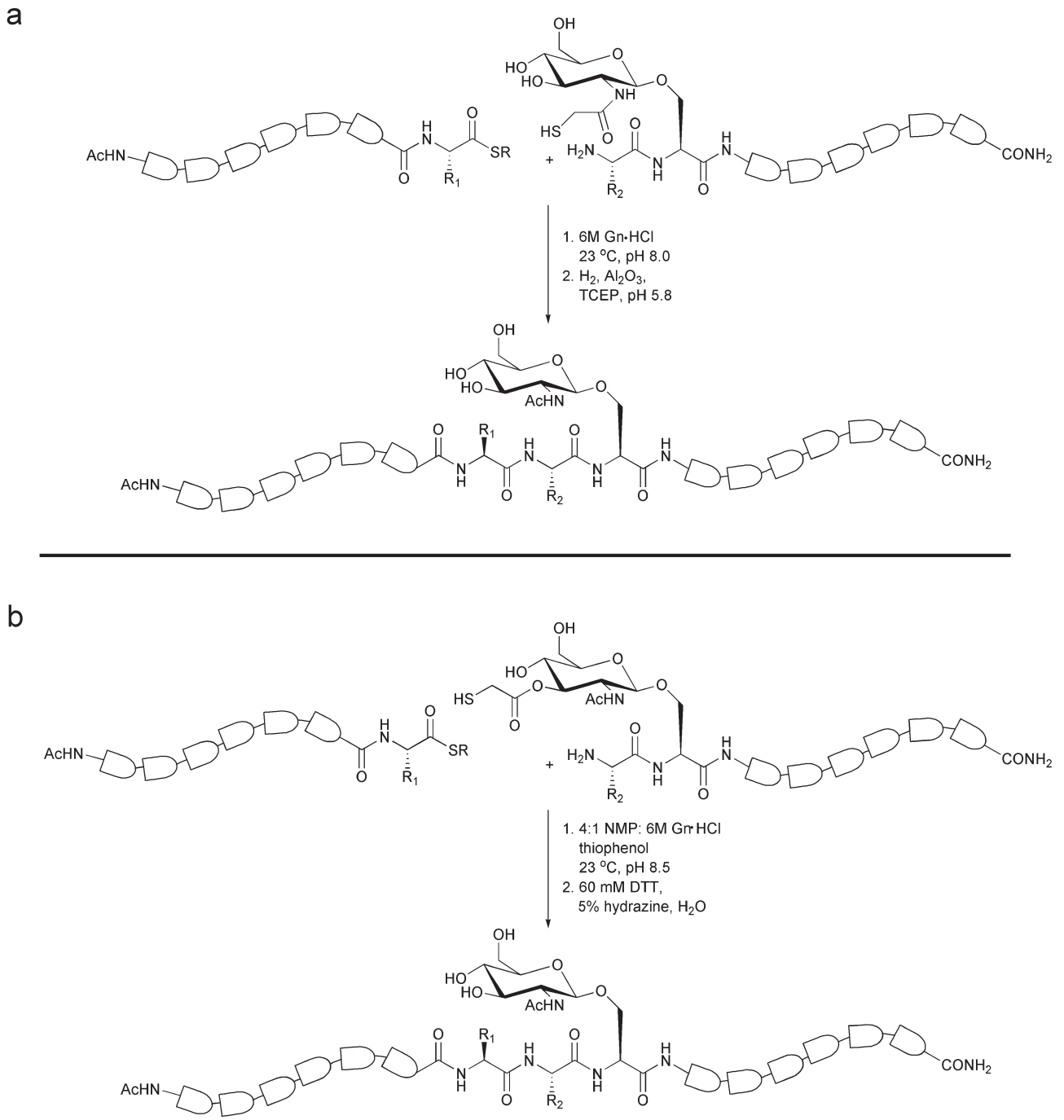

Fig. 6 Sugar assisted ligation a) first generation ligation; b) second generation ligation.

of a large excess of soluble thiol, the immobilized protein will undergo NCL to furnish the desired product. An example of this technology is the synthesis of a homogeneous glycoprotein variant of maltose binding protein (MBP) (Fig. 7b).[25] Both *N*-terminal and *C*-terminal engineering have found extensive use in protein synthesis. It is interesting to note that the two methods are orthogonal, allowing for the introduction of glycopeptide fragments at both ends of a glycoprotein, a factor which Macmillan and Bertozzi took advantage of in a recent synthesis of the glycoprotein GlyCAM-1.[26]

2.5 Protease catalyzed ligation

The use of native and engineered proteases as peptide ligases under kinetically controlled conditions has been known for quite some time.[27] This method commonly uses an activated peptide ester with a protected *N*-terminal amine as the donor in the presence of an organic co-solvent such as DMF or DMSO. While the addition of organic solvent helps to minimize unwanted hydrolysis of the peptide ester, these solvents can be problematic due to the intrinsic instability and low catalytic activity of enzymes in such environments. In an

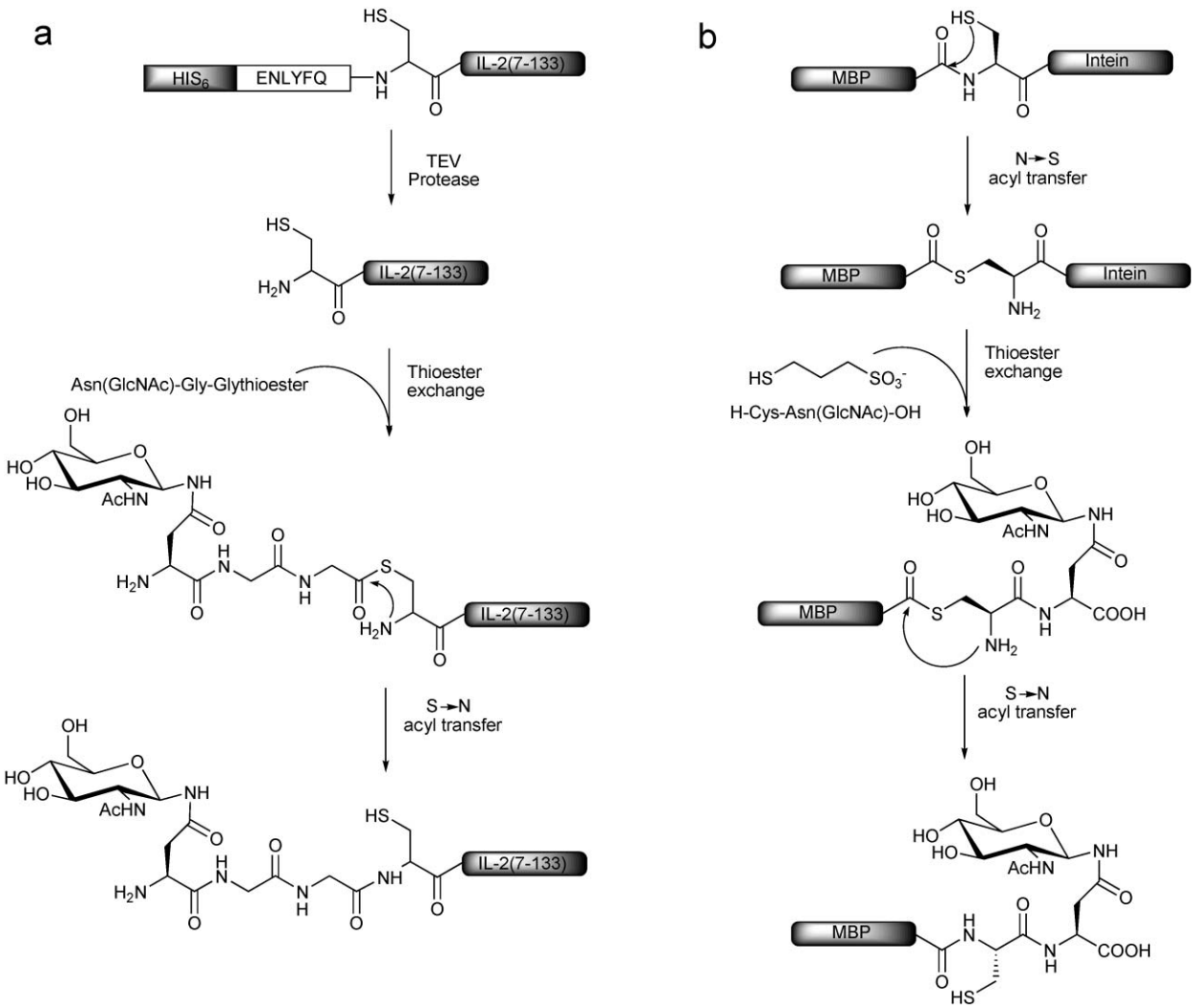

Fig. 7 Expressed protein ligation: a) *N*-terminal engineering using TEV proteases; b) Intein mediated synthesis of glycosylated MBP.

effort to overcome these problems, Wong and coworkers developed two mutant variants of the serine protease subtilisin BPN', a stable variant, 8397, and a thiosubtilisin.[28] The first of these variant enzymes, which was engineered for enhanced stability in organic solvents, possesses a half-life in DMF of 14 days at 25 °C, as opposed to a half–life of 30 min for the wild type. In contrast the thiosubtilisin, in which the catalytic serine residue of the enzyme was mutated to cysteine, was designed for activity in aqueous media. In this latter mutant the reaction proceeds through a thioester intermediate, which, similar to the thioester intermediate in NCL, preferentially undergoes aminolysis over hydrolysis. The advantage of using these enzymes is that they don't require a cysteine at the ligation junction or chemical modification of the ligation product.

In the context of glycopeptide synthesis, the enzyme appears to be tolerant of both *N*-linked and *O*-linked monosaccharides near the active site. The reaction is, however, sensitive to the exact position of the glycosylated amino acid relative to the ligation junction.[29] Many of the areas surrounding the active site were found to be tolerant of peptide glycosylation. By contrast, glycosylation was not tolerated in the active site, and hence at the ligation junction. Recently however, the Davis group reported that chemical modification of the of the subtilisin SBL active site alters the enzyme in such a way that glycosylation is tolerated at the ligation junction.[30]

The power of this method is exemplified by a relay synthesis of RNase B (Fig. 8).[31] RNase B was first treated with Endo H, an enzyme known to hydrolyze the glycans of *N*-linked glycoproteins between the first two GlcNAc residues, to afford GlcNAc-RNase A. Upon treatment with subtilisin BNP' this protein was converted into a 20 amino acid peptide and a 104 amino acid protein. The fragments were isolated then ligated using subtilisin 8397 in the presence of 9 volumes of glycerol to restore the protein backbone of RNase. The glycan was then elaborated using the glycosyltransferases to afford a homogeneous glycoprotein.

2.6 Traceless Staudinger ligation

A third method for the union of complex glycopeptides is the traceless Staudinger ligation. This method, which was first described by Bertozzi and coworkers as a method for conjugating probes to azide bearing proteins,[32] is initiated by

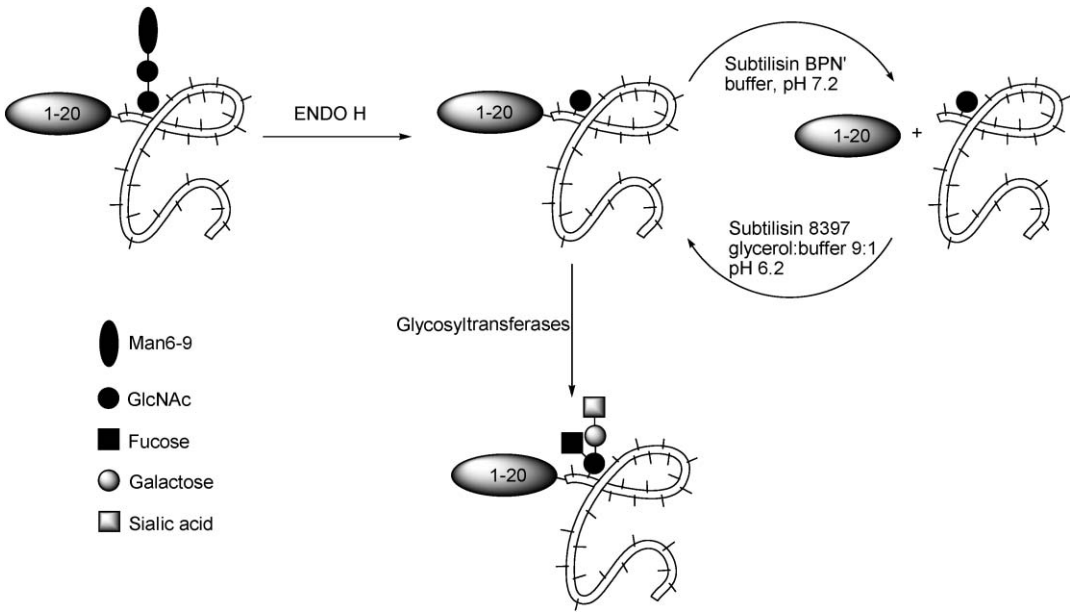

Fig. 8 Glycoprotein remodelling and subtilisin catalyzed synthesis of RNase B.

the reaction of an azide with a phosphothioester to form an imminophosphorane intermediate. This intermediate undergoes an S → N acyl transfer to form an amidophosphonium salt, which is readily hydrolyzed to afford the amide product (Fig. 9). This method has been shown by Raines and coworkers to efficiently form peptide bonds at room temperature in aqueous or wet organic solvents when there is a glycine residue at the ligation junction.[33] Subsequent studies have shown that glycosylation is tolerated on both the phosphothioester and the azidopeptide provided that the glycosylation is at least two residues removed from the ligation site.[34] Presumably this limitation is a result of the increased steric bulk that arises when one of the amino acids at the ligation junction is glycosylated. It may be possible to overcome this limitation, however, as Raines and coworkers have shown that ligations at more sterically encumbered junctions is possible if the electron density on the phosphine is increased.[35] In addition, while both coupling partners are normally obtained synthetically, it has been demonstrated that azidodipeptide esters can be ligated to unprotected peptides by subtilisin. This finding implies that it may be possible to use the Staudinger ligation with expressed proteins in a fashion analogous to EPL.

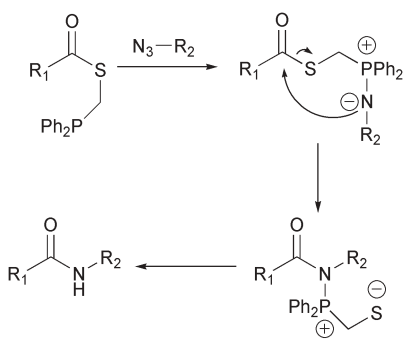

Fig. 9 Mechanism of the Staudinger ligation.

3. Methods for introducing complex glycans

A second major challenge in the synthesis of glycoproteins is the introduction of fully elaborated glycans into the protein backbone. While a number of groups have shown that it is possible to attach the oligosaccharide to the peptide backbone through a non-native linkage to afford neoglycoproteins,[5] the synthesis of glycoproteins with native glycan linkages remains a significant challenge. This is due in large part to the fact that the steric encumbrance of amino acids possessing large glycans greatly reduces the efficiency of the peptide coupling reactions in SPPS. As a result it is often desirable to take a more convergent approach where the glycan is attached to the peptide backbone subsequent to SPPS.

In the case of *O*-linked oligosaccharides this is often achieved by introducing a single sugar residue into the peptide backbone during SPPS. Following peptide synthesis, this sugar can serve as a handle for further enzymatic elaboration of the carbohydrate using glycosyltransferases. These enzymes catalyze the transfer of a glycan from a sugar nucleotide diphosphate to an appropriate acceptor. While these enzymes are highly regio- and stereoselective, glycosyltransferase catalyzed reactions suffer from the drawback that both the enzyme and the nucleotide sugar are expensive. In addition, the reaction is complicated by feedback inhibition by the nucleoside phosphate byproducts. This latter problem can be alleviated by adding alkaline phosphatase to destroy the nucleoside phosphate. Alternatively, enzymatic cofactor regeneration systems can be used to convert the nucleoside phosphate back into a sugar nucleotide, thereby bypassing feedback inhibition and greatly decreasing the cost of the process (Fig. 10). By using several glycosyltransferases in succession it is possible to build up a complex glycan on an unprotected peptide.

In the case of *N*-linked glycoproteins it is not possible to build up native glycans on the peptide backbone using glycosyltransferases as the requisite enzymes are not available.

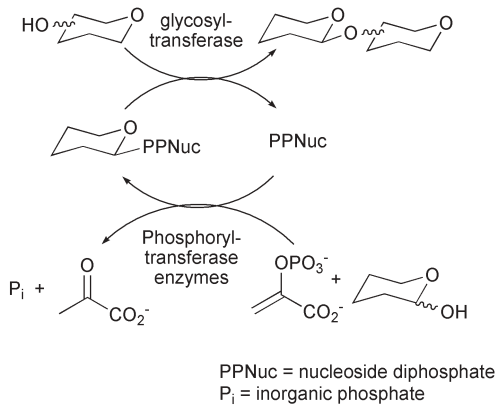

Fig. 10 General details of the sugar nucleotide regeneration system for large scale synthesis.

As a result it is necessary to couple the fully elaborated glycan to the peptide backbone in a single operation. Over the past few years, two approaches have emerged as the methods of choice for the convergent synthesis of *N*-linked glycoproteins: 1) the Lansbury aspartylation, and 2) the use of *endo*-β-*N*-acetylglucosaminidases.

The Lansbury aspartylation involves the coupling of a β-glycosylamine with an aspartate residue of a peptide under peptide coupling conditions to afford an asparagine linked glycan (Fig. 11).[36] This method is advantageous in that the reaction is highly convergent and is not dependent on the structural features of the glycan coupling partner. The major drawback of the method is the need for extensive protection of the peptide backbone in order to prevent unwanted side reactions. Despite this limitation, the method is highly versatile and has found extensive use in glycopeptide synthesis, as exemplified by a recent report from Danishefsky and co-workers where it was used to synthesize three complex glycoforms of a fragment of prostate specific antigen (PSA).[37]

An alternative to chemical glycosylation is to use *endo*-β-*N*-acetylglucosaminidases to couple complex glycans to

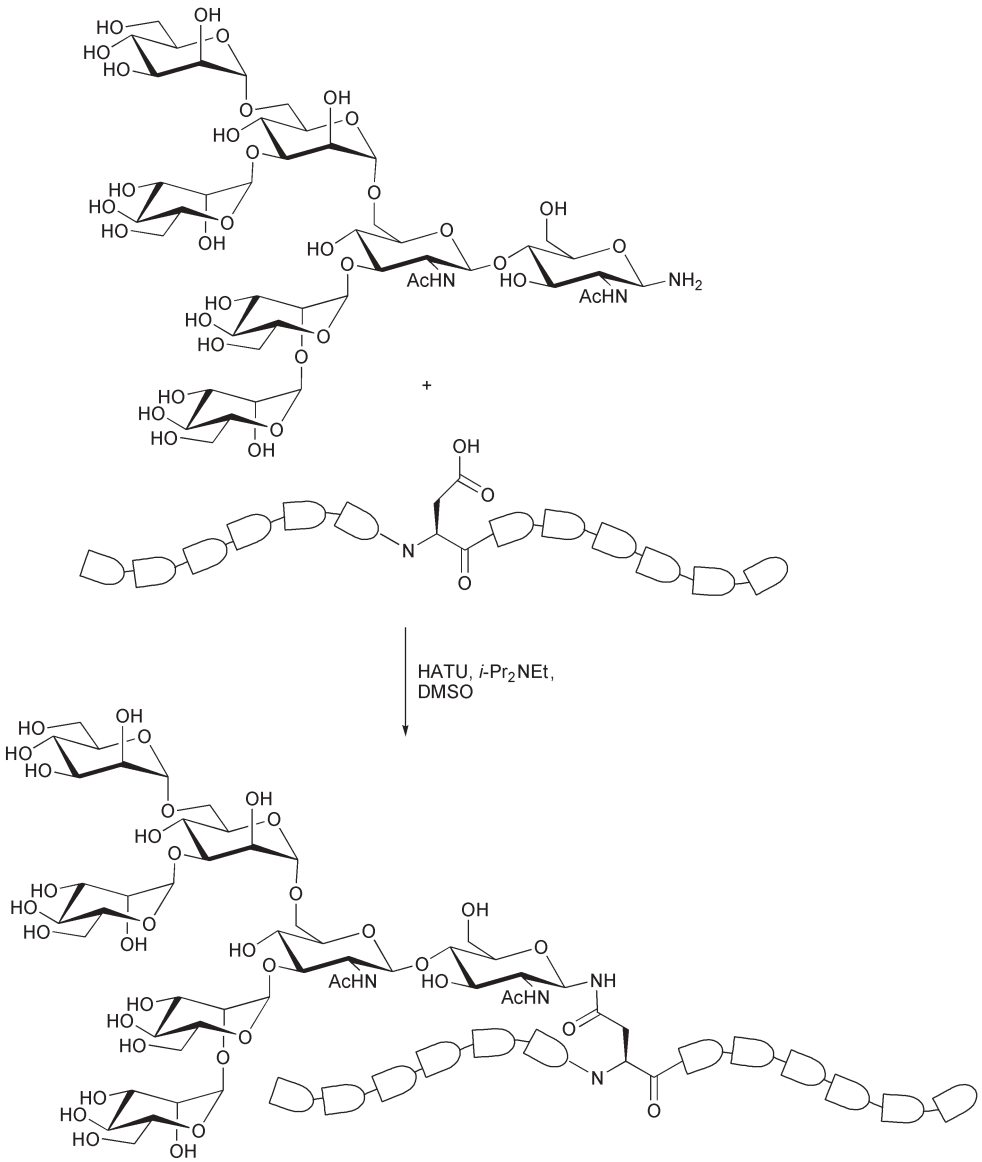

Fig. 11 Synthesis of *N*-linked glycopeptides using the Lansbury aspartylation.

peptides. Under normal conditions these enzymes hydrolyze the chitobiose (GlcNAc-GlcNAc) core of *N*-linked glycoproteins to produce a free oligosaccharide and a protein possessing a single GlcNAc at the reaction site. Importantly, it has been shown that several members of this family, especially *endo*-β-*N*-acetylglucosaminidase A (Endo A) from *Anthrobacter protophormiae* and *endo*-β-*N*-acetylglucosaminidase M (Endo M) from *Mucor hiemalis* can also catalyze transglycosylation, the transfer of an oligosaccharide from GlcNAc to another sugar residue. If the acceptor sugar is bound to a peptide the result is a glycopeptide bearing a complex *N*-linked glycan. A particularly impressive application of this methodology was the synthesis of a homogeneous glycoform of RNase B using a process known as glycoprotein remodeling.[38] Following treatment of a mixture of RNase B glycoforms with Endo H to produce a RNase containing a single GlcNAc residue, the enzyme was treated with Endo A and $(Man)_6GlcNAc_2Asn$ (obtained from ovalbumin through exhaustive pronase digestion) to afford $(Man)_6(GlcNAc)_2$RNase B. The yield of these reactions is low, and although it was shown that the presence of organic solvents could increase the synthetic utility of the process,[39] a huge excess of one of the reactants is necessary to obtain moderate yields of the desired product.

A significant improvement to this methodology was developed utilizing sugar oxazolines as the donor component. The oxazolines, which were designed based on putative transition state mimetic, greatly enhance the rate of transglycosylation. This results in high product yields using only a small excess (*ca.* 3-fold) of the donor sugar. Using this method, Wang and coworkers were able to efficiently synthesize an homogeneous glycoform of RNase B in high yield (Fig. 12).[40] Recent work has suggested that this method may not be practical for glycoprotein synthesis when the product glycan is a good substrate for Endo-mediated hydrolysis.[41] The scope of the reaction, however, has yet to be determined.

4 *In vivo* suppressor tRNA technology

The chemoenzymatic methods described above provide a variety of options for the construction of homogeneous glycopeptides; however, they all require a number of manipulations to arrive at the desired product. An alternative approach makes use of *in vivo* suppressor tRNA technology to incorporate unnatural amino acids into the peptide backbone during protein synthesis in *Escherichia coli*.[42] The technique makes use of an orthogonal tRNA synthetase–tRNA pair from *Methanococcus jannaschii* that has been evolved to charge

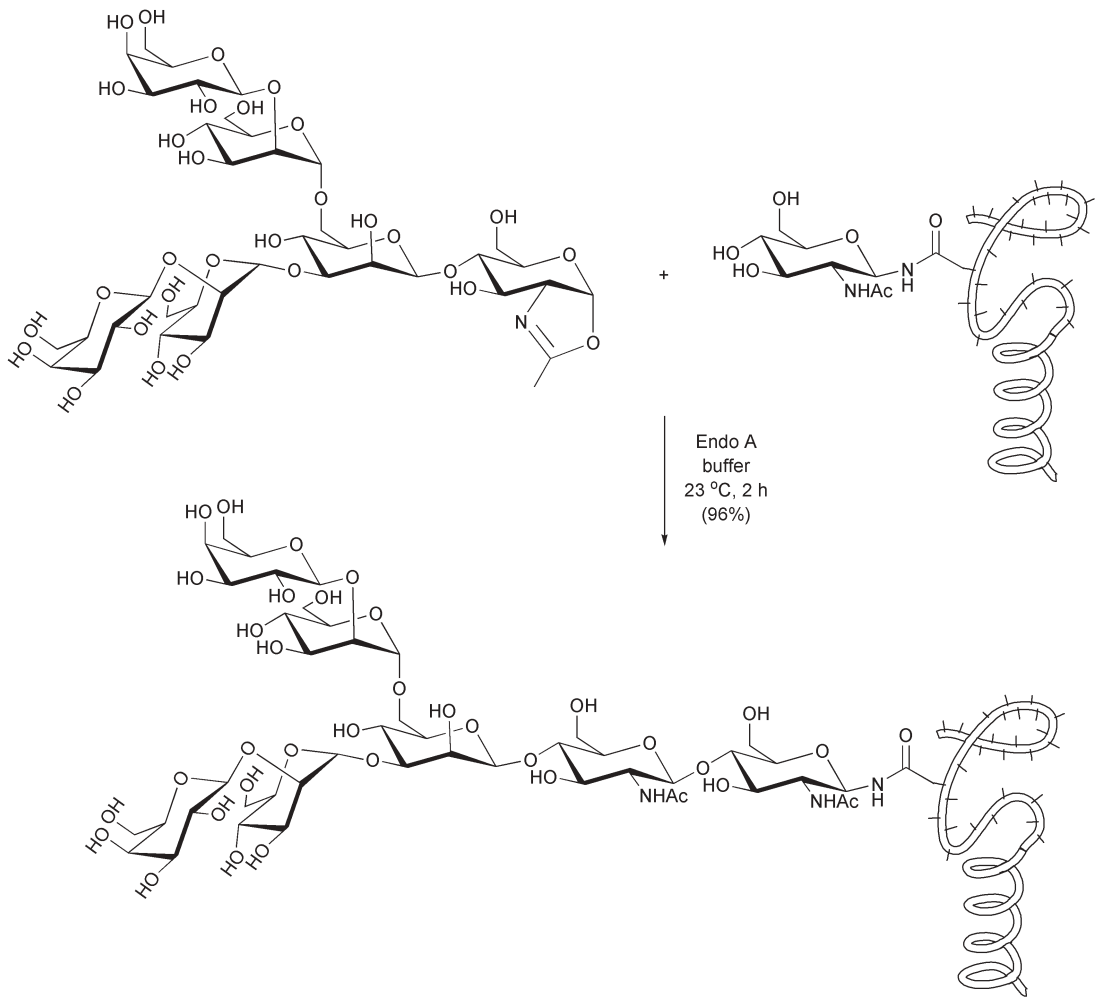

Fig. 12 Endo A-mediated synthesis of a homogeneous glycoform of RNase B.

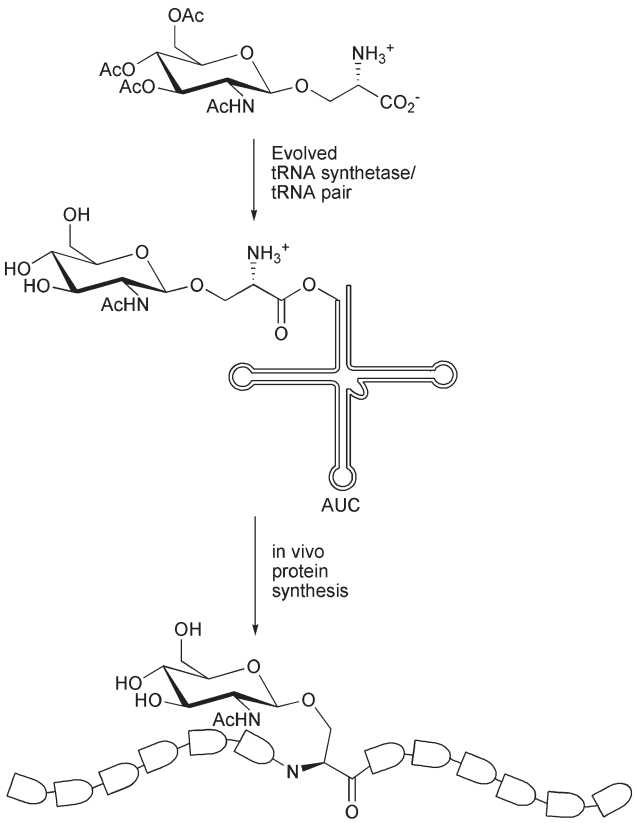

Fig. 13 Synthesis of a glycoprotein containing GlcNAc-Ser using *in vivo* suppressor tRNA technology.

amber suppressing tRNAs with unnatural amino acids. When an amber stop codon (TAG) is inserted into the gene encoding for a protein of interest, the unnatural amino acid is site-specifically incorporated into the protein backbone. This method has been used to introduce the glycosyl amino acids (GlcNAc)-Ser and (GalNAc)-Thr into proteins expressed in *E. coli* (Fig. 13).[43,44] In this approach, the sugar moiety is peracetylated to facilitate cellular uptake. Once inside the cell, the sugar protecting groups are removed by nonspecific esterases and the unprotected glycosyl amino acid can charge the tRNA. These monoglycosylated glycoproteins can be elaborated enzymatically to furnish glycoproteins more complex in nature.

5 Conclusions

The preparation of homogeneous glycoproteins is still a formidable challenge. New technologies to facilitate the process are constantly being developed, however, they still fall short of routine preparations of glycoproteins that are greatly needed to understand the scope of this complex protein modification. Still, the current array of chemical and enzymatic methods covered herein for the construction of homogeneous glycans and glycoproteins has permitted access to increasingly complex targets. These advances should prove much needed new tools for the study of glycobiology and will lay the groundwork for understanding the molecular basis of the function of protein glycosylation and the development of new avenues of therapeutic investigation.

Acknowledgements

We thank the NIH and Skaggs Institute for support of our research. C.S.B. is grateful for a Ruth L. Kirchenstein NRSA postdoctoral fellowship provided by the NIH (GM073500-02). We also thank Sarah R. Hanson for proofreading the manuscript.

References

1 A. Varki, *Glycobiology*, 1993, **3**, 97–130.
2 R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683–720.
3 D. H. Dube and C. R. Bertozzi, *Nat. Rev. Drug Discovery*, 2005, **4**, 477–488.
4 *Glycoscience*, ed. B. Fraser-Reid, K. Tatsuta and J. Thiem, Springer, Berlin, 2001.
5 B. G. Davis, *Chem. Rev.*, 2002, **102**, 579–601 and references therein.
6 M. R. Pratt and C. R. Bertozzi, *Chem. Soc. Rev.*, 2005, **34**, 58–68 and references therein.
7 T. Matsushita, H. Hinou, M. Fumoto, M. Kurogochi, N. Fujitani, H. Shimizu and S. I. Nishimura, *J. Org. Chem.*, 2006, **71**, 3051–3063.
8 P. E. Dawson and S. B. H. Kent, *Annu. Rev. Biochem.*, 2000, **69**, 923–960.
9 Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman and C. R. Bertozzi, *J. Am. Chem. Soc.*, 1999, **121**, 11684–11689.
10 S. Mezzato, M. Schaffrath and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2005, **44**, 1650–1654.
11 G. Chen, J. D. Warren, J. H. Chen, B. Wu, Q. Wan and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2006, **128**, 7460–7462.
12 B. Wu, J. H. Chen, J. D. Warren, G. Chen, Z. H. Hua and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2006, **45**, 4116–4125 and references therein.
13 L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526–533.
14 B. L. Pentelute and S. B. H. Kent, *Org. Lett.*, 2007, **9**, 687–690.
15 J. Offer, C. N. C. Boddy and P. E. Dawson, *J. Am. Chem. Soc.*, 2002, **124**, 4642–4646.
16 D. Macmillan and D. W. Anderson, *Org. Lett.*, 2004, **6**, 4659–4662.
17 A. Brik, S. Ficht, Y. Y. Yang, C. S. Bennett and C. H. Wong, *J. Am. Chem. Soc.*, 2006, **128**, 15026–15033.
18 Y.-Y. Yang, S. Ficht, A. Brik and C.-H. Wong, *J. Am. Chem. Soc.*, submitted.
19 S. Ficht, R. J. Payne, A. Brik and C.-H. Wong, *Angew. Chem., Int. Ed.*, DOI: 10.1002/anie.200700546.
20 T. W. Muir, *Annu. Rev. Biochem.*, 2003, **72**, 249–289.
21 D. A. Erlanson, M. Chytil and G. L. Verdine, *Chem. Biol.*, 1996, **3**, 981–991.
22 T. J. Tolbert, D. Franke and C. H. Wong, *Bioorg. Med. Chem.*, 2005, **13**, 909–915.
23 T. C. Evans, J. Benner and M. Q. Xu, *Protein Sci.*, 1998, **7**, 2256–2264.
24 T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705–6710.
25 T. J. Tolbert and C. H. Wong, *J. Am. Chem. Soc.*, 2000, **122**, 5421–5428.
26 D. Macmillan and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2004, **43**, 1355–1359.
27 F. Bordusa, *Chem. Rev.*, 2002, **102**, 4817–4867.
28 C. H. Wong, M. Schuster, P. Wang and P. Sears, *J. Am. Chem. Soc.*, 1993, **115**, 5893–5901.
29 T. J. Tolbert and C.-H. Wong, *Methods Mol. Biol. (Totowa, N. J.)*, 2004, **283**, 267–279.
30 K. J. Doores and B. G. Davis, *Chem. Commun.*, 2005, 168–170.
31 K. Witte, P. Sears, R. Martin and C. H. Wong, *J. Am. Chem. Soc.*, 1997, **119**, 2114–2118.
32 E. Saxon, J. I. Armstrong and C. R. Bertozzi, *Org. Lett.*, 2000, **2**, 2141–2143.
33 B. L. Nilsson, M. B. Soellner and R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 91–118.
34 L. Liu, Z. Y. Hong and C. H. Wong, *ChemBioChem*, 2006, **7**, 429–432.

35 M. B. Soellner, A. Tam and R. T. Raines, *J. Org. Chem.*, 2006, **71**, 9824–9830.
36 S. T. Cohenanisfeld and P. T. Lansbury, *J. Am. Chem. Soc.*, 1993, **115**, 10531–10537.
37 V. Y. Dudkin, J. S. Miller and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2004, **126**, 736–738.
38 K. Takegawa, M. Tabuchi, S. Yamaguchi, A. Kondo, I. Kato and S. Iwahara, *J. Biol. Chem.*, 1995, **270**, 3094–3099.
39 J. Q. Fan, K. Takegawa, S. Iwahara, A. Kondo, I. Kato, C. Abeygunawardana and Y. C. Lee, *J. Biol. Chem.*, 1995, **270**, 17723–17729.
40 B. Li, H. J. Song, S. Hauser and L. X. Wang, *Org. Lett.*, 2006, **8**, 3081–3084.
41 T. Rising, T. D. W. Claridge, J. W. B. Moir and A. J. Fairbanks, *ChemBioChem*, 2006, **7**, 1177–1180.
42 L. Wang and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2005, **44**, 34–66.
43 Z. W. Zhang, J. Gildersleeve, Y. Y. Yang, R. Xu, J. A. Loo, S. Uryu, C. H. Wong and P. G. Schultz, *Science*, 2004, **303**, 371–373.
44 R. Xu, S. R. Hanson, Z. W. Zhang, Y. Y. Yang, P. G. Schultz and C. H. Wong, *J. Am. Chem. Soc.*, 2004, **126**, 15654–15655.

