

Enzymes for Glycoprotein Synthesis

Chi-Huey Wong*

Abstract: More than 90% of human proteins are glycosylated and since protein glycosylation is understood to play a role in folding, trafficking, stability, immunogenicity, and function there is a need to generate pure protein glycoforms. Pure single protein glycoforms are difficult to obtain because the cellular machinery produces complex mixtures for any given protein. In this article an overview is given of various approaches used to generate specific single glycoforms of proteins, studies of the role of glycosylation in protein stability and function, and the imaging and identification of new glycoproteins related to cancer.

Keywords: Enzymes · Glycopeptides · Synthesis

Introduction

A major challenge in the study of glycoproteins is to characterize the effects of glycans on glycoprotein folding and function, in pure form. Pure, single protein glycoforms are difficult to obtain, since the cellular machinery produces complex mixtures of glycoforms for any given protein. Given that more than 90% of human proteins are glycosylated and protein glycosylation is understood to play a role in folding, trafficking, stability, immunogenicity, and function,^[1–8] there is a need for methods of generating pure protein glycoforms. To date, biophysical, structural, and functional characterizations have generally been limited to mixed populations of inseparable glycoforms which have been compared to deglycosylated protein which was either enzymatically treated to remove all glycans, expressed in a bacterial host which lacks the protein glycosylation machinery found in eukaryotes, treated with glycosidase inhibitors during expression, or mutated to remove specific glycosylation sites.^[9–11]

Due to the importance of this problem, several approaches have been taken towards generating specific single glycoforms of proteins. An early approach was *in vitro* enzymatic remodeling of native glycoprotein mixtures with a combination of glycosidases to trim outer saccharides, and glycosyltransferases to rebuild new glycan structures.^[12–14] While conceptually simple, this approach is difficult in practice because it requires each enzymatic step to go to completion in order to obtain single glycoforms. If individual steps do not reach 100% yield, the result will be a mixture of glycoforms that is similar in complexity to the native mixture. A new approach, developed collaboratively between the Wong and Schultz labs, takes advantage of bacterial expression of protein with incorporation of an unnatural glycosyl amino acid.^[15,16] This could be followed by *in vitro* enzymatic elaboration of the glycan to produce desired structures. Improvements in yield are needed to make this process practical, and once again the need for several enzymatic steps to convert the initial monosaccharide to a full-length complex glycan would lead to glycoform mixtures unless each enzymatic step goes to completion. In addition, it is difficult to incorporate more than one glycosyl amino acid into a protein using this method. A more convergent route toward fully elaborated glycans, with fewer individual steps, is necessary to prevent this problem.

Another approach that holds great promise is engineering of production organisms, such as yeast, to attach specified N-linked glycans in pure form to proteins that they are overexpressing.^[17–19] The significant commercial potential of this approach for biopharmaceuticals is exemplified by the 2006 purchase of GlycoFi, a company developing this technology, by Merck. While the yeast engineering approach is promising for production of

single glycoform biopharmaceuticals bearing specific human glycans, it is not ideal for production of O-linked glycoproteins and analogues for biochemical research, since any modification of a specific glycan would require re-engineering of a new yeast strain. For this purpose, a modular chemical approach in which different glycans can be swapped out to make a series of related glycoforms using the same set of reactions, is more attractive.

Since the initial report of native chemical ligation (NCL),^[20] many advances towards total chemical synthesis of proteins have been achieved.^[21–28] These methods have been applied to answer basic questions in protein folding and function, which were otherwise difficult to answer with expressed proteins containing native sequences and naturally occurring amino acids.^[29–31] Earlier versions of NCL were limited by the requirement for cysteine at the ligation junction, as well as a strong dependence on peptide sequence at the ligation junction, primarily due to steric influence on ligation yield. The invention of removable thiol auxiliaries^[27,32–35] has expanded the sequence repertoire that is accessible to NCL, and recently the new methods have been applied to synthesis of full-length glycoproteins^[36] or fragments thereof.^[37,38] However, ligation junction sequence dependence remains a serious issue, as these auxiliary-based systems rely on an amine that is less nucleophilic and more hindered than a native peptide's N-terminal primary amine. Thus, these ligations are generally slower and proceed in lower yields, and with more racemization, than traditional cysteine-based NCL ligations. A recent approach to NCL is expression of N-terminal cysteine containing protein fragments, with subsequent native chemical ligation (NCL) of glycopeptide thioesters. TEV protease specifically cleaves its target sequence to produce an N-terminal cysteine. This approach was

*Correspondence: Prof. C.-H. Wong
President, Academia Sinica
128 Academia Road
Section 2, Nankang
Taipei, Taiwan
and Department of Chemistry
The Scripps Research Institute
10550 N. Torrey Pines Rd.
La Jolla, CA 92037
Tel.: +886 2 2789 9400
Fax: +886 2 2785 3852
E-mail: chwong@gate.sinica.edu.tw
or wong@scripps.edu

used to assemble N-linked glycosylated IL-2.^[14]

Sugar Assisted Glycopeptide Ligation

In 2006 our lab reported that a native N-linked or O-linked saccharide attached to asparagine or serine/threonine, respectively, when appended with a thiol handle, was an excellent auxiliary for promoting ligation with a peptide thioester, in what we call sugar-assisted ligation (SAL) (Fig. 1).^[39-41]

In SAL, the ultimate nucleophile is a primary amine, as is the case for NCL (Fig. 1). It is therefore more reactive than auxiliaries attached directly to the nucleophilic amine, as shown in Fig. 2. The convergent nature and high efficiency of SAL, as well as the improvements to the original method which are outlined below, make it a promising technique for practical synthesis of homogenous glycoproteins. The scope and limitations of the first generation SAL reaction were determined on peptide substrates,^[39,40] with a Gly-Gly junction being the fastest, Ala-Gly half as rapid, and Val-Gly an order of magnitude slower. Isolated yields for slower reactions are lower, due to the competing thioester hydrolysis reaction. Therefore, sterically hindered junctions such as Ala-Val are not practical from a synthetic standpoint: although the desired product can be detected, yields are poor. One of the goals of subsequent efforts has been to improve ligation rates for these difficult junctions, and/or suppress the competing hydrolysis reaction.

One of the limitations of SAL and other thiol auxiliary-based approaches is the need to remove the thiol function to obtain fully native protein or glycoprotein. When additional cysteine residues are present in the protein, the thiol auxiliary must be removed selectively using reductive desulfurization, with orthogonal protection of internal cysteine residues. We recently reported a shorter alternative route, the second generation SAL.^[42] The thiolacetic acid moiety was shifted from the 2-position of glucosamine to the 3-position, and was connected through an ester rather than an amide linkage. Thus, after ligation, the thiol auxiliary could be removed by mild hydrazinolysis rather than desulfurization (Fig. 3).

Through efforts to suppress competing thioester hydrolysis rates when faced with difficult ligation junctions, a mixed aqueous/organic solvent system was discovered, which allowed slow ligations to reach synthetically useful yields. The solvent consists of 4:1 by volume N-methylpyrrolidinone (NMP) to standard aqueous ligation buffer, containing 6M guanidinium chloride as denaturant. Using

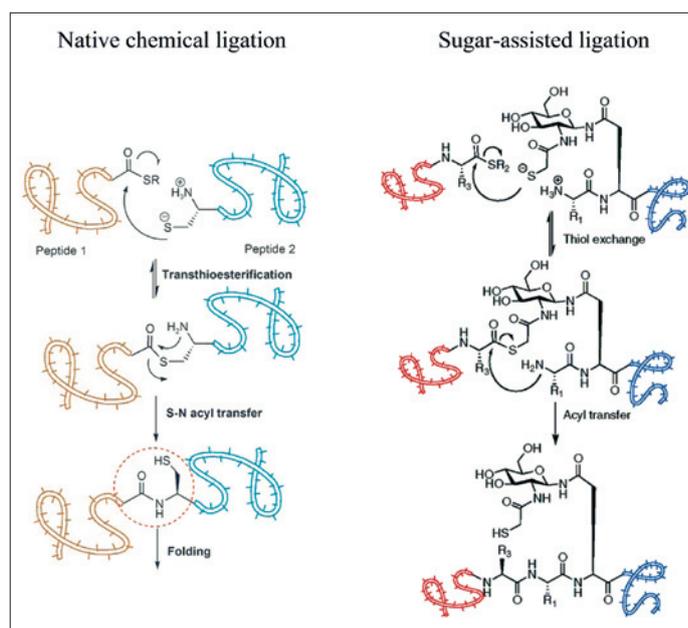


Fig. 1. Comparison of proposed mechanisms for the original native chemical ligation (NCL) and sugar-assisted ligation (SAL). In NCL, an obligatory N-terminal cysteine residue of peptide 2 undergoes a transthioesterification reaction with a C-terminal thioester of peptide 1. A subsequent intramolecular S-N acyl transfer results in peptide bond formation. In SAL, a thiol auxiliary that plays the role of cysteine in NCL is appended from an asparagine-linked carbohydrate. After acyl transfer to form the native glycopeptide, the thiol can be reduced to provide the native N-acetylglucosamine residue. In SAL, the N-terminal residue is not limited to cysteine.

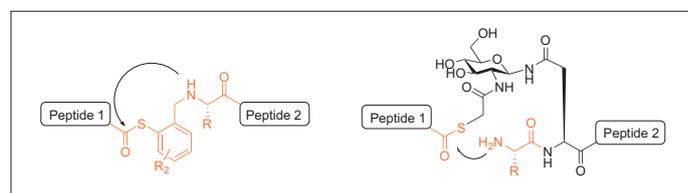


Fig. 2. Comparison of thioester intermediates in peptide ligations mediated by aryl thiol auxiliaries (*i.e.* ref. [27]), and SAL. The ultimate amine nucleophile in the former case is an α -branched (except for glycine) benzylic secondary amine which suffers considerable steric hindrance. In the case of SAL, it is a peptide N-terminal primary amine, as found in NCL. While the transition state ring size for intramolecular S-N acyl transfer is larger for SAL, ligation rate and sequence tolerance at the ligation junction remains comparable to other ligation methods.

these conditions, it was found that ligation occurred even when the thiol auxiliary was further removed from the ligation site (extended SAL).^[42,43] Extended SAL greatly increased the junction sequence repertoire that is accessible to ligation. If the amino acids neighboring the sugar residue were difficult to ligate due to steric bulk, the ligation site could be moved over one or more residues until a satisfactory junction was found (Fig. 4).

The hydrolysis-suppressing mixed solvent system was found to enable direct bimolecular coupling of peptide thioesters with peptide N-terminal amines, in the absence of a thiol auxiliary or cysteine residue.^[44] The direct bimolecular aminolysis has an obvious limitation relative to thiol

exchange-based intramolecular ligations, because it is much more concentration-dependent. As peptide substrates become larger, it becomes more difficult to achieve productive concentrations. Therefore, this method may be limited to smaller peptide and glycopeptide targets rather than full-length glycoproteins. An additional caveat is the requirement of protecting groups for lysine residues. If left unprotected, peptides containing lysines would produce a mixture of desired peptide product and isomeric material acylated at the lysine side chain amine. Despite these technical limitations, we were encouraged by the successful synthesis of a MUC1 tandem repeat consisting of 60 amino acids and six glycans by the direct bimolecular

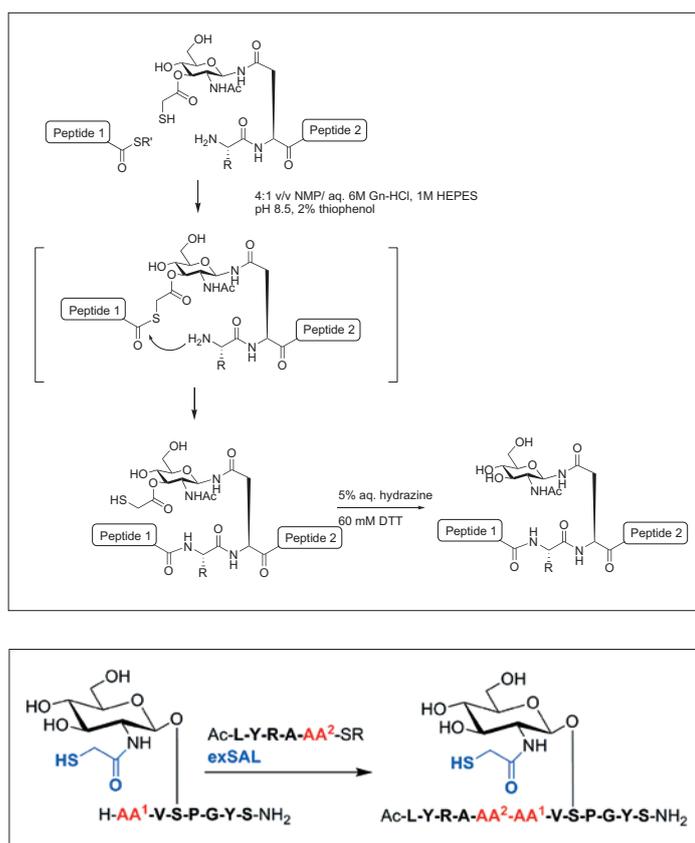


Fig. 4. Extended SAL. In comparison to the original SAL reaction, the thiol auxiliary-appended sugar is two amino acid residues removed from the ligation junction rather than one. This allows for synthesis of glycopeptide sequences in which the residues adjacent to the glycosylation site are poor substrates (*i.e.* valine, as shown). The range of sequences amenable to SAL is greatly expanded by the ability to skip over difficult amino acid residues.

aminolysis reaction (Fig. 5). The presence of glycans improves the solubility of the peptide substrates, and may allow for efficient ligation of larger fragments than would have been expected under these conditions.

When the sugar-bound thiol auxiliary was four, five, or six amino acids removed from the ligation junction, the transition states for the proposed intramolecular acyl transfers in these cases are >20-membered rings, and the relative rates of these super-extended SAL reactions are similar to those of direct aminolysis reactions at similar concentrations. However, ligation rates for single extended SAL reactions (as in Fig. 4) are faster than direct aminolysis rates, and comparable to those of the original SAL reaction. Therefore we believe in these cases the sugar-bound thiol auxiliary is participating in the reaction mechanism. As previously mentioned, direct aminolysis reactions are likely to be limited in utility to glycopeptides and smaller glycoproteins because they are bimolecular and therefore concentration-dependent. However, the single extended SAL, in which the auxiliary is two amino acids removed from the ligation junction,

should be useful for synthesis of larger targets in analogy to NCL and SAL. The scope and limitations of the methods described here were explored on small peptide substrates for the sake of ease of synthesis, since a large number of sequences had to be prepared to fully explore sequence dependence. The goal of developing these methods, however, is to enable chemical synthesis of full-length homogenous glycoproteins, not just glycopeptides. As a proof of principle, the original SAL method was applied to the synthesis of a naturally occurring glycoprotein, dipterucin, an antimicrobial protein from *Drosophila* which consists of 82 amino acids, including two α -linked N-acetylgalactosaminyl (GalNAc) threonine residues at positions 10 and 54, and no cysteine residues. This simple model glycoprotein has been synthesized by NCL in the past (with subsequent conversion of the non-native cysteine to alanine). Therefore this target is useful for gauging the efficiency of SAL in comparison to earlier methods. The synthetic approach toward dipterucin is illustrated in Fig. 6.

The C-terminal 30mer peptide bearing the SAL auxiliary was synthesized

Fig. 3. Second-generation SAL, with the thiolacetic acid auxiliary linked as an ester to the 3-position of GlcNAc, and subsequent mild hydrazinolysis to generate native glycopeptides containing internal cysteine residues.

by standard Fmoc-based solid-phase peptide synthesis. The central 16mer peptide thioester was prepared by standard Boc-based solid-phase methods. The native alanine 37 of dipterucin was temporarily mutated to cysteine for subsequent use in NCL, and synthesized in protected form. As previously suggested by model peptide studies, the Gly-Val ligation junction at amino acids 52-53 was amenable to sugar-assisted ligation, resulting in formation of the 37-82 fragment which was deprotected with mercuric acetate to unmask the N-terminal cysteine residue. The C-terminal glycopeptide thioester was synthesized by Fmoc-based solid-phase synthesis via a side-chain anchoring strategy,^[45,46] using *p*-methoxybenzyl ethers to orthogonally protect the sugar hydroxyls instead of the traditional acetate esters. Traditional NCL proceeded smoothly, and was followed by desulfurization to concomitantly remove the sugar thiol auxiliary and mutate Cys37 back to the native alanine. Thus, using a combination of ligation methods (NCL and SAL) and suitable protecting group manipulations, a highly convergent route to a native glycoprotein was achieved.

All thiol-based ligation methods, including NCL and SAL, require access to peptide thioesters. As larger and more complex glycoproteins become targets for synthesis, the technical challenges of preparing large, complex peptide thioesters become more limiting. Therefore, development of high-yielding, racemization-free methods for solid-phase synthesis of these key intermediates has been an area of intensive research.^[45-50] Our approach^[46] relies on side-chain anchoring of the growing peptide rather than the traditional C-terminal anchoring, to allow for synthesis by Fmoc methodology. This approach was first demonstrated for synthesis of peptide thioesters by Wang and Miranda,^[45] through linkage to either asparagine or glutamine. We developed new linkage strategies to extend the applicability of the method, which uses a C-terminal allyl ester as an orthogonally protected precursor for subsequent generation of the thioester on-resin. Peptide thioesters were successfully prepared through side-chain linkages to Asp, Glu, Ser, Thr, Tyr, and Cys, greatly expanding the range of sequences that were amenable to solid-phase Fmoc synthesis of corresponding peptide thioesters. The chemistry was shown to tolerate the presence of glycosyl amino acids in the peptide. The utility of the side-chain anchoring strategy was demonstrated by synthesis of a 28mer glycopeptide thioester, which was subsequently ligated by NCL to a N-terminal cysteine-containing glycopeptide to produce a 40mer N-terminal fragment of

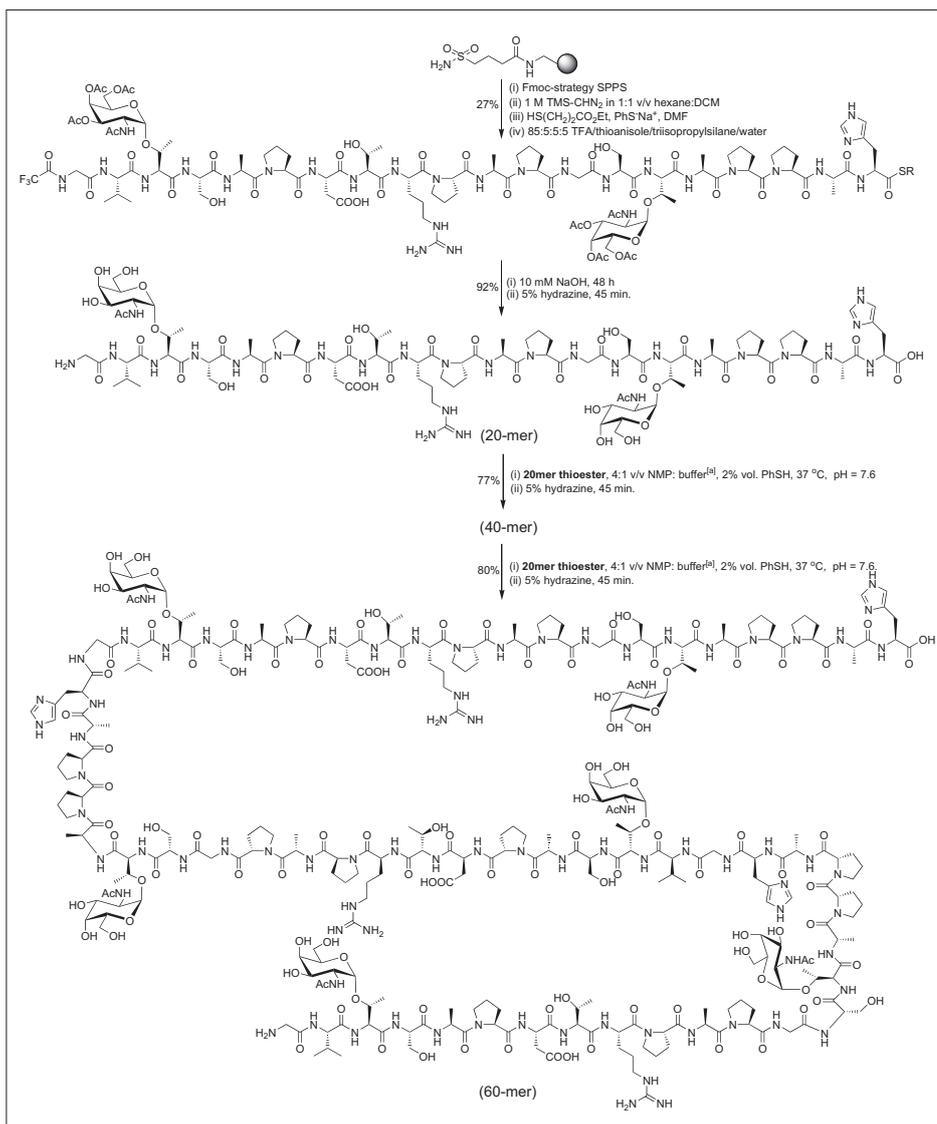


Fig. 5. Synthesis of a 60mer glycopeptide of the MUC1 tandem repeat by iterative direct bimolecular aminolysis of a 20mer glycopeptide thioester with a 20mer and 40mer N-terminal glycopeptides. Aberrant O-glycosylation of the extracellular MUC1 tandem repeat domain is associated with cancer, and these glycopeptides may be candidates for cancer vaccine development.

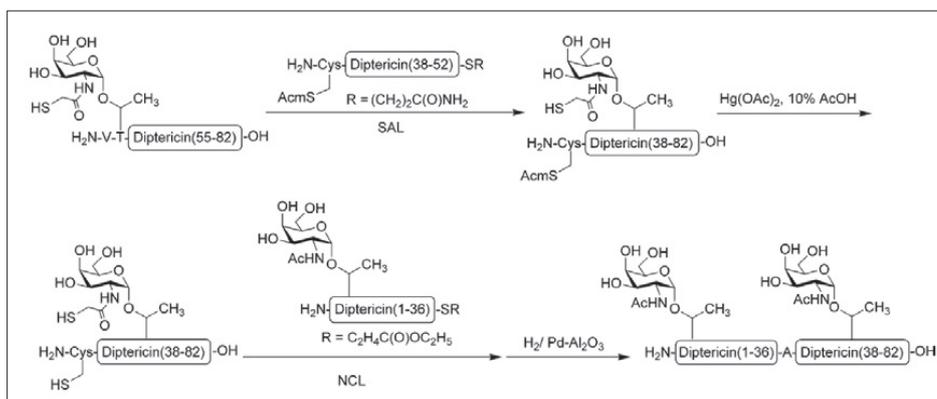


Fig. 6. Synthetic approach to SAL-mediated synthesis of diptericin, an 82-amino acid glycoprotein. Two peptide ligation disconnections were designed. The first was a sugar-assisted ligation between the C-terminal 30-amino acid peptide containing a α -linked GalNAc bearing the thiol auxiliary, and a central 16mer thioester, bearing a protected cysteine at the N-terminal. After cysteine deprotection, the second ligation, *via* traditional NCL, was effected between the resultant 46mer glycopeptide and a 36mer glycopeptide thioester. Desulfurization concomitantly generated the native GalNAc and converted the cysteine used in the NCL reaction into the native alanine residue found in diptericin.

erythropoietin, containing two N-linked glycosylation sites (Fig. 7).

The methods described thus far have been shown to be amenable to synthesis of glycopeptides and proteins containing monosaccharides attached through amino acid side chains. Naturally occurring glycoproteins contain a range of more complex glycans. We set out to investigate whether SAL could be performed with extended, native glycan structures rather than just monosaccharides, since the method would be most valuable if it could be applied to convergent synthesis of native glycoproteins, rather than just truncated variants.^[51] A series of glycopeptides were synthesized by a combination of solid-phase peptide synthesis and enzymatic glycosylation to systematically assess the effect of extended glycan structures on SAL within the same peptide sequence context (Fig. 8). Notably, the presence of a free thiol auxiliary on the first GlcNAc does not interfere with enzymatic transfer by any of the glycosyltransferases that were tested.

SAL-mediated ligations with the peptides bearing extended glycans on the 4-position proceeded in comparable yields to those performed with monosaccharides (Fig. 9). Glycosylation at both the 4- and 3-positions (Sialyl Lewis X), however, prevented ligation from proceeding, most likely due to steric occlusion of the thiol auxiliary by the neighboring fucose. Fucosylation at the 6-position did not prevent ligation from occurring, however. Glycosylation at the 3-position appears to be the only modification that prevents SAL from occurring.

In addition to building up oligosaccharides one sugar at a time with glycosyltransferases, endoglycosidases have proven to be useful tools for block transfer of certain oligosaccharides when used in the reverse (synthetic) direction with oligosaccharides activated as oxazoline derivatives.^[52–56] We took advantage of this elegant methodology developed by Wang and coworkers, to synthesize and ligate a peptide containing the core trisaccharide common to all N-linked glycans, Man- β -1,4-GlcNAc- β -1,4-GlcNAc- β -Asn (Fig. 10).

The various synthetic methods described above were developed to address a range of challenges in chemical synthesis of glycoproteins. It remains to apply the methods to preparation of single glycoforms of therapeutically important targets. If successful, it will allow us to probe the role of glycosylation in protein stability and function with pure samples rather than mixtures. These efforts will lead to discovery of optimal glycoforms with respect to a number of therapeutic parameters, including biological half life, potency, and immunogenicity.

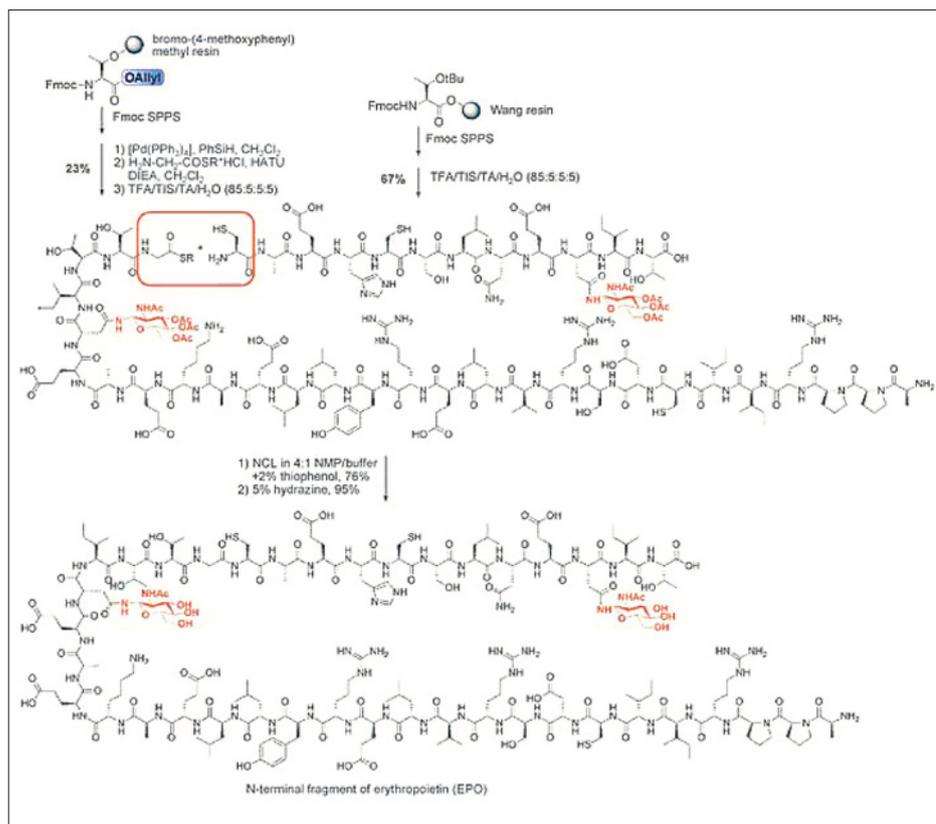


Fig. 7. Synthesis of a 40mer N-terminal glycopeptide domain of erythropoietin (EPO) using a new side-chain anchoring strategy for Fmoc-based solid-phase synthesis of a 28mer glycopeptide thioester. The growing peptide was linked through the side chain of a C-terminal threonine, protected orthogonally as an allyl ester. After assembly of the 27mer glycopeptide chain, the allyl ester was deprotected and coupled on-resin to a glycine thioester made in solution. After acidic resin cleavage/side chain deprotection, the resulting glycopeptide thioester was ligated by NCL to a 12mer glycopeptide made by standard solid phase methods.

Study of Glycosylation of Human CD2 Folding

The technical challenges for preparation of single glycoforms for biochemical characterization have opened the door to recent computational efforts to dissect the role of glycosylation in protein stability and function.^[57] With improved methods for chemical synthesis of single glycoforms, experimental validation of the computational results will be possible.

The cell adhesion and signaling molecule CD2, which is expressed on T-lymphocytes and natural killer cells, has been extensively studied as a model for the role of glycosylation in protein folding and function.^[2] The functional 105 amino acid extracellular adhesion domain bears a single glycosylation site at Asn65 in the human homolog. While the glycosylation site is unique to the human protein (hCD2)^[58] (there is no glycosylation site in the rat homolog, for instance), it is essential not only for functional binding to the counter-receptor hCD58, but is in fact essential for proper folding of the extracellular adhesion domain of hCD2. Since the study of effects on stability and function can be reduced to a single glycosylation site, this system is attractive for molecular-level analysis of the origin of stabilization arising from protein glycosylation. However, despite being a relatively simple glycoprotein, a limited number of variants have been accessible with existing technologies, and even fewer variants can be obtained in pure form. Wagner and collaborators studied the role of the single glycan on function and folding of the extracellular adhesion domain hCD2ad by NMR spectroscopy.^[59] The material used in their study was expressed in Chinese hamster ovary cell culture as a mixture of high-mannose glycoforms $\text{Man}_{5,8}\text{-GlcNAc}_2\text{-Asn}$. Samples were enzymatically treated with PNGase F to remove the entire glycan, or with endoglycosidase H to reduce the glycan to a single GlcNAc. Thus, studies were performed on a mixture of high-mannose

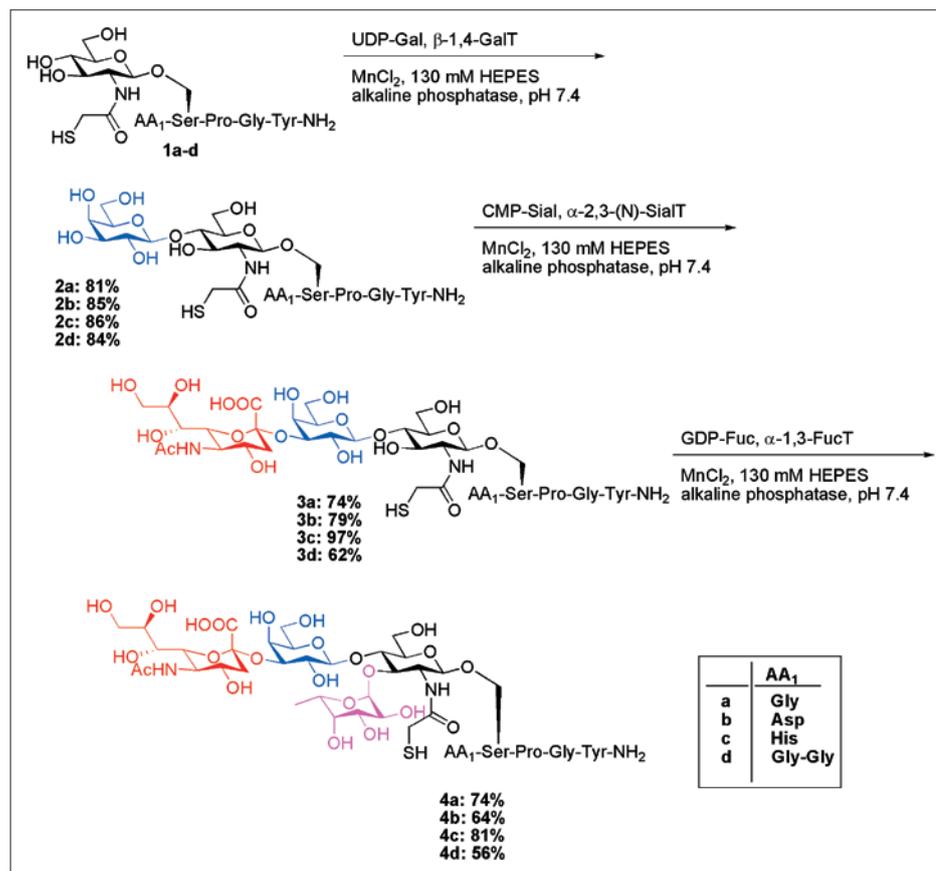


Fig. 8. Synthesis of glycopeptide substrates for SAL experiments with extended glycans. A series of pentapeptides was synthesized on solid phase bearing a thiolacetate on the O-linked GlcNAc residue, and Gly, Asp, His, or Gly-Gly (i.e. extended SAL) as the N-terminal nucleophilic residue. The peptide glycans were elaborated by glycosyltransferase-catalyzed transfer of sugars in good yields, producing LacNAc (2a-d), Sialyl LacNAc (3a-d), and Sialyl Lewis X (4a-d). α -1,6-Fucosyl GlcNAc (not shown) was also prepared by chemical synthesis, so that the effect of glycosylation at the 3, 4, and 6 positions on ligation rates could all be assessed.

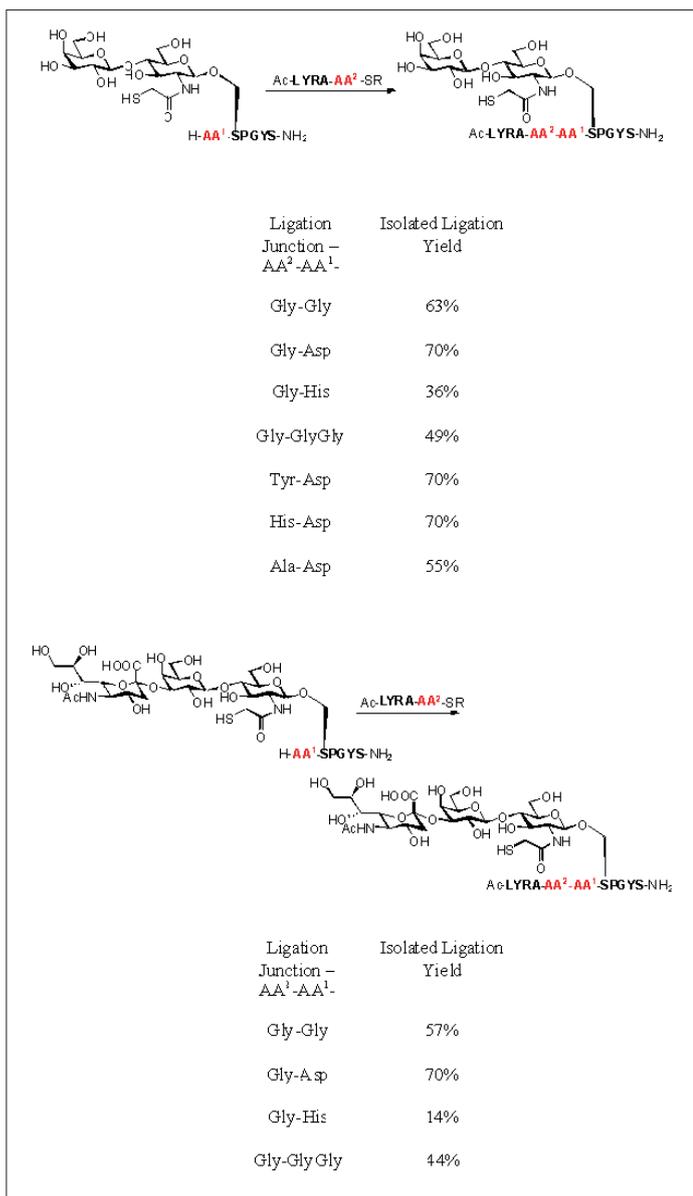


Fig. 9. SAL reactions with extended glycans. Elaboration of the thiol-containing glycan at the 4-position to a disaccharide or trisaccharide did not adversely affect the isolated yields of SAL products. Fucosylation at the 3-position did prevent ligation from occurring, however. Fucosylation at the 6-position slowed the reaction relative to the monosaccharide, although an isolated yield of 44% was obtained for a Gly-Gly junction in this case.

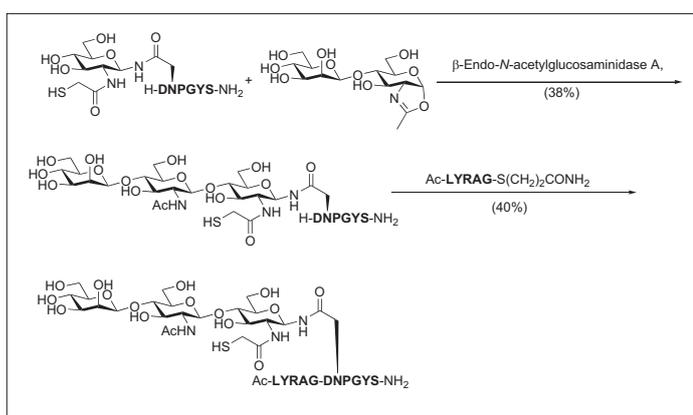


Fig. 10. Endoglycosidase-mediated transfer of a disaccharide oxazoline to thiol auxiliary-containing glycopeptide, and sugar-assisted ligation of the resulting glycopeptide containing the common trisaccharide core of all N-linked glycans.

glycoforms, deglycosylated hCD2ad, and a single monosaccharide truncated variant. Existing glycan remodeling methods did not allow for ready access to other glycoforms, such as complex oligosaccharides or partially truncated variants.

With this set of CD2 variants, it was determined that the glycan was essential

for binding to CD58, although it does not make direct contacts with the counter-receptor. It is also essential for proper folding. The de-glycosylated variant was unstable and tended to aggregate. The rat sequence, which is stable and functional in the absence of any glycan, has additional compensating amino acid mutations to sta-

bilize the protein. Further analysis showed much of the glycan was more highly ordered than it would be in the free oligosaccharide, as were the neighboring amino acids of the polypeptide. An extensive network of contacts was deduced between glycan and polypeptide, especially to the first two GlcNAc residues (Fig. 11). The mannose residues were less well ordered, although the presence of mixtures at these positions likely complicated the analysis. The importance of contacts with the first GlcNAc residue was further supported by characterization of the Endo H-treated variant bearing only the single GlcNAc at Asn65. Unlike the fully deglycosylated variant, this one was stably folded, exhibited a comparable CD spectrum relative to the high-mannose mixture of variants, and functionally bound to CD58 and anti-C2 antibodies, albeit with poorer affinity. These studies suggested a key role of the N-linked glycan in folding and function of CD2ad, and began to dissect the relative contributions of specific sugar moieties. However, the limited and heterogenous data set prohibited more detailed characterization.

We have been interested in generating a more extensive library of glycosylated variants of hCD2ad, to further dissect the role of each carbohydrate residue in folding, and also in stability and function. The roles in folding and function may be separated *in vivo*, and modulated by different sugars, since different glycosylation patterns are present during folding and processing in the Golgi (high-mannose type) and in the mature, functional secreted protein (usually complex type in mammals).^[60] Some glycan variants can be accessed through enzymatic remodeling of expressed mixtures of glycoforms, others can only be accessed practically by chemical synthesis. The conserved oligomannose saccharide structure that is initially block transferred to all N-linked glycosylation sites is shown in Fig. 12. The outer three glucose residues are involved in calnexin/calreticulin (CNX/CRT)-assisted folding and ER-associated degradation (ERAD) of misfolded proteins, and are removed early in the maturation cycle of newly synthesized glycoproteins.^[60] Likewise, outer mannose residues on the B and C arms are frequently trimmed in mammals, to be replaced by other sugars in hybrid and complex type glycans. However, the core Man₃GlcNAc₂ (paucimannose) pentasaccharide is conserved and common to all N-linked glycans, and is most likely to play the key role in N-linked glycan effects on glycoprotein folding, stability, and function.

Using a combination of different expression hosts and enzymatic remodeling techniques, we have generated a more extensive library of hCD2ad glycoforms, al-

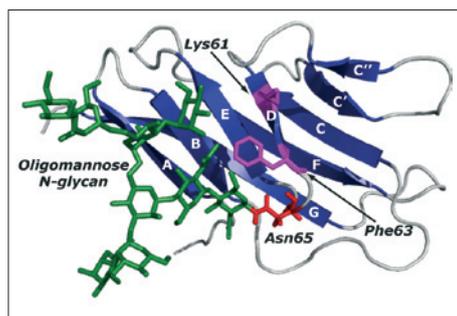


Fig. 11. Structure of the extracellular domain of hCD2^[59] showing stabilizing contacts with neighboring amino acid residues. Although structural and biochemical studies have suggested the primary role of the single N-linked glycan in CD2 folding and function, access to only a limited set of glycan variants has prevented a detailed molecular analysis of the roles of individual saccharides.

beit still as mixtures of related structures. Specifically, variants were expressed in *E. coli*, HEK 293 cells (human kidney cell line), and SF9 insect cell line to generate unglycosylated, mixed complex and hybrid, and pauci- and high-mannosylated glycoform mixtures, respectively. Enzymatic remodeling of these mixtures with mannosidase or endoglycosidase produced further mixtures of truncated variants (Fig. 13).^[61] These sets of glycoform mixtures were then analyzed with respect to folding kinetics and energetics, using stopped-flow fluorescence spectroscopy. Refolding of urea-denatured glycoprotein and unfolding of glycoprotein in buffer upon mixing with urea were monitored by the intrinsic fluorescence of tryptophan. Monoexponential folding and unfolding kinetics in these glycoform mixtures suggested that partial glycan heterogeneity did not appreciably complicate the experiments, and it was possible to draw a number of conclusions based on comparison of these samples. However, we expect with access to pure single glycoforms by chemical synthesis, including those bearing glycans not accessible by expression and/or enzymatic remodeling, a clearer picture will be achieved. Availability of pure glycoforms of the hCD2ad will also facilitate more detailed structural studies by both X-ray crystallography and NMR spectroscopy.

Results of folding experiments showed that the deglycosylated human sequence was not properly folded under standard conditions, as expected based on previous reports. It could be induced to fold by addition of the osmolyte proline, however. In contrast, all glycosylated variants, whether truncated or extended, were able to fold into a similar stable structure, with similar kinetic parameters. It was concluded that the glycans provided both a kinetic contribution (increased folding rate) and a ther-

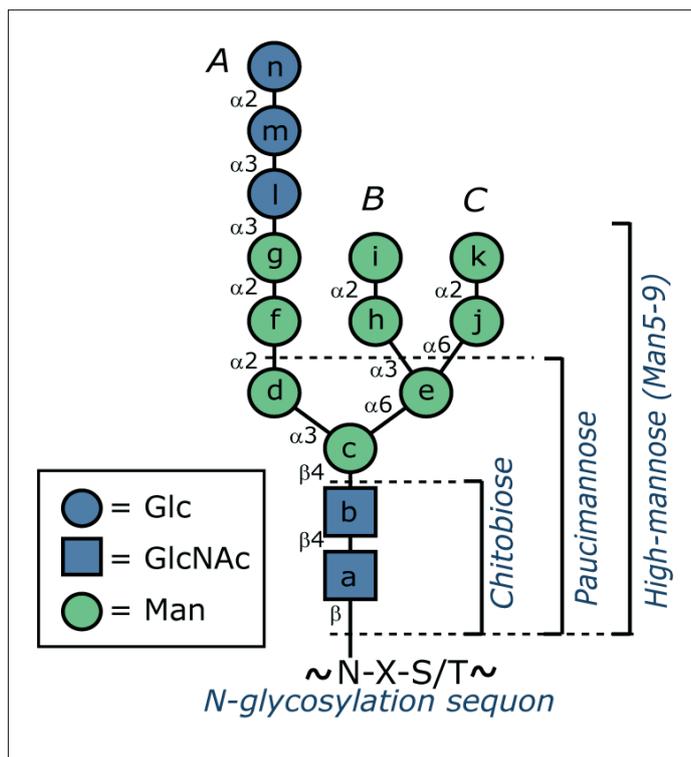


Fig. 12. Schematic of the structure of the evolutionarily conserved N-linked oligosaccharide that is initially transferred to nascent glycoproteins. Glucose residues 1-n are enzymatically clipped early in the maturation process, and a linear combination of mannose residues f-k may be retained, removed, replaced, or appended to generate a complex mixture of glycoforms. The paucimannose core a-e, containing the asparagine-linked chitobiose disaccharide, is conserved, and is most likely to be important in N-glycan mediated effects on glycoprotein structure and function.

modynamic one (stabilization of the folded state). Comparison of glycoform samples 3–7 suggested that all of the hastening of folding component resides in the first GlcNAc residue, since rates of all samples were similar. A total of 3 kcal/mol of thermodynamic stabilization was dissected to arise from the first GlcNAc (2/3 contribution) and the second and third sugars of the core trisaccharide (1/3 contribution), with negligible effect arising from outer sugars of extended glycan structures. The analysis may be complicated by the presence of mixtures however, as well as fucosylation of the key 1st GlcNAc in significant fractions of samples 4–6.

The results suggest that the core trisaccharide (A–C in Fig. 12) is sufficient for intrinsic acceleration of folding and stabilization in hCD2ad, and the presence of this trisaccharide as the common and conserved core in all N-linked glycans may not be coincidental. To further investigate this conclusion, and assist in structural studies, we intend to synthesize a series of pure glycoforms of this protein, containing mono-, di-, and trisaccharides, including non-native sequences only available through chemical and enzymatic synthesis.

Imaging and Identification of New Glycoproteins Related to Cancer

The cellular enzymatic machinery for biosynthesis of oligosaccharides and glycoconjugates has been shown to be tolerant of structural modifications to the natural

substrates. This has allowed for *in vivo* incorporation of chemical handles into glycoproteins. Selected modifications on the C(6) of fucose (as azidofucose FucAz or alkyne-modified Fucyne) are tolerated by fucose kinase, GDP fucose phosphorylase, and fucosyltransferase, allowing for incorporation of these probes into growing glycoproteins in the Golgi.^[62,63] Likewise, N-acylmannosamines modified with an alkyne on the 2-acyl group (ManNAcyne) were successfully incorporated by the complex sialic acid biosynthetic machinery and expressed in sialylated glycoproteins. By enzymatically incorporating azido sugars or alkyne sugars, the copper-catalyzed azide alkyne cycloaddition (CuAAC) was used to click on fluorogenic reporter groups (Fig. 14) to enable imaging and quantitation of fucosylation and sialylation in different cell types.^[63,64] For instance, increased fucosylation and sialylation was detected in prostate cancer cells relative to normal prostate cells.

Enzymatic incorporation of chemically modified probes into cellular glycans has also led to a new method of glycoproteomic analysis, GIDmap.^[64] Combining the metabolic oligosaccharide engineering methods with proteomic and glycomic tools such as multidimensional mass spectrometry has allowed for proteome-wide mapping of glycoproteins (Fig. 15). The sugar probes developed by our group as described here and by the groups of Bertozzi^[65] and Wilson^[66] will lead to identification of new glycoproteins, and potentially to discovery of new glycans associated

	hCD2ad variant	Expression host	Enzyme	N-glycan composition	Rel. abund.
Non-glycosylated	1	<i>E. coli</i>	N/A	No glycan; wild type sequence	
	2a	<i>E. coli</i>	N/A	No glycan; Lys61Glu	
	2b	<i>E. coli</i>	N/A	No glycan; Lys61Glu/Phe63Leu	
Glycosylated	3	HEK293	N/A		R ₁ GlcNAc Man ₂ R ₃ None >0.50
					Man Man ₂ None 1.0
					LacNAc Man ₂ None 0.67
					LacNAc Man ₂ None 0.60
					LacNAc LacNAc None 0.56
Glycosylated	4	Sf9	N/A		R ₁ Fuc (71%) None (8%)
					R ₂ None (20%) Man (1%)
Remodeled	5	Sf9	Endo H,D		R ₁ Fuc (74%) None (26%)
	6	Sf9	Jack bean mannosidase		R ₁ Fuc None 44%
					R ₂ Fuc Man 25%
	7	Sf9	Endo D (uncut)		R ₁ None (90%) Man (10%)

Fig. 13. Glycoform mixtures produced by expression of hCDad in different hosts, for study of the effect of glycan structure on folding kinetics and energetics. Glycan structures and abundances were determined by LC/MS, with electrospray ionization. Variant 3 is a complex mixture of hybrid and complex type glycans expressed in human kidney cell line HEK293; only the six most abundant structures are shown. 4 was expressed in insect cells, resulting in a mixture dominated by core-fucosylated paucimannose (70%), with 20% high-mannose glycoforms also present. 5, 6, and 7 were generated by enzymatic truncation of insect cell-expressed glycoforms. 5 contains a 3:1 mixture of fucosylated GlcNAc and GlcNAc monosaccharide. 6 is a mixture of fucosylated and unfucosylated core tri- or tetrasaccharides, with one or two mannoses attached to the chitobiose core. 7 is a high-mannose mixture containing Man₆₋₇GlcNAc₂, similar to that used in ref. [59].

with specific disease states. In particular, cancer-specific glycans may be targets for vaccine development, and also be useful biomarkers for diagnostics. The new glycoproteins identified will be useful targets for synthesis and for the study of glycosylation effect on the structure and function of these glycoproteins.

Received: March 6, 2009

- [1] A. Helenius, M. Aebi, *Annu. Rev. Biochem.* **2004**, 73, 1019.
- [2] B. Imperiali, S. E. O'Connor, *Curr. Opin. Chem. Biol.* **1999**, 3, 643.
- [3] M. Molinari, *Nature Chem. Biol.* **2007**, 3, 313.

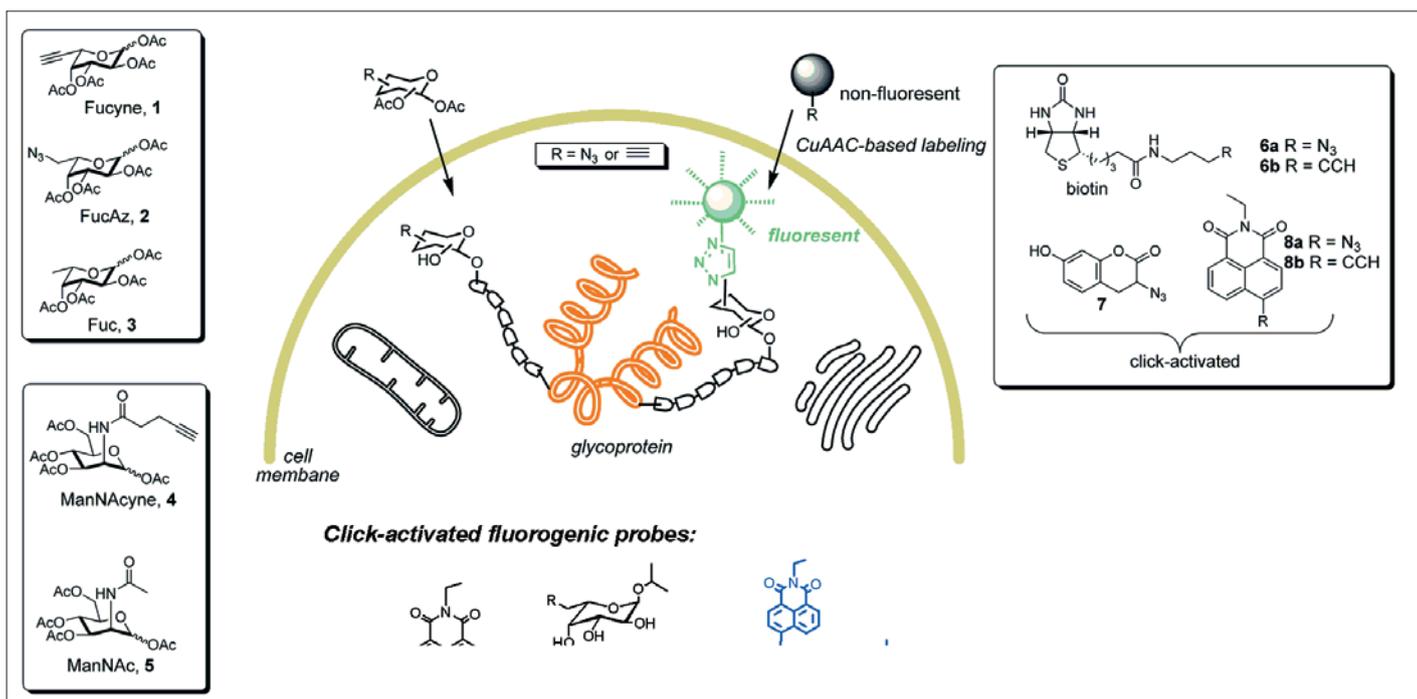


Fig. 14. Metabolic oligosaccharide engineering of fluorescent probes.

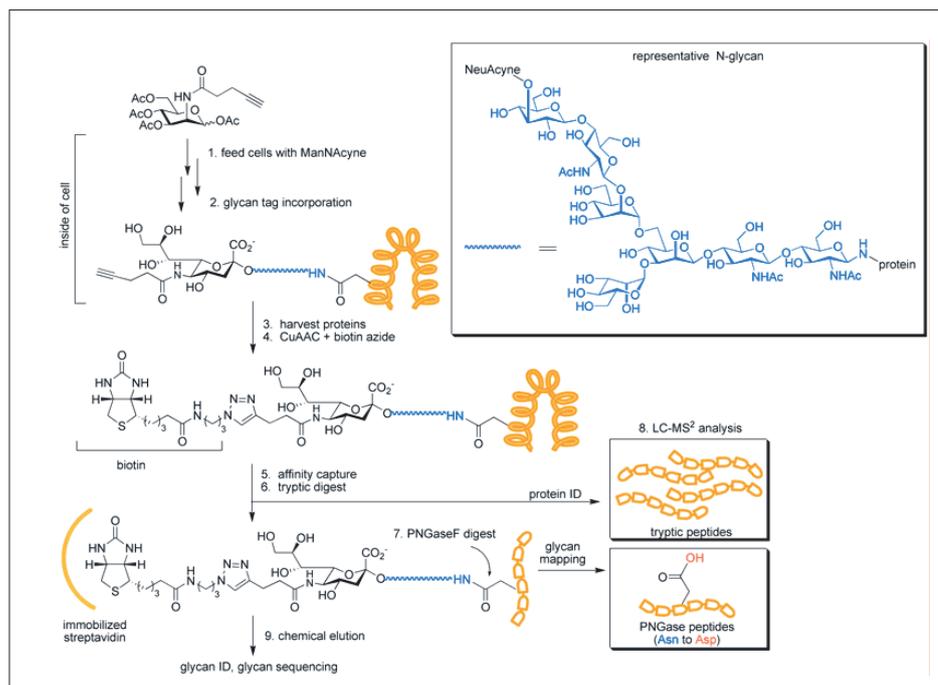


Fig. 15. Scheme for GIDmap glycoproteomic analysis

- [4] A. J. Petrescu, M. R. Wormald, R. A. Dwek, *Curr. Opin. Struct. Biol.* **2006**, *16*, 600.
- [5] N. Mitra, S. Sinha, T. M. C. Ramya, A. Surolia, *Trends Biochem. Sci.* **2006**, *31*, 156.
- [6] P. Scheiffele, J. Peranen, K. Simons, *Nature* **1995**, *378*, 96.
- [7] K. Fiedler, K. Simons, *Cell* **1995**, *81*, 309.
- [8] E. E. Simanek, D.-H. Huang, L. Pasternack, T. D. Machajewski, O. Seitz, D. S. Millar, H. J. Dyson, C.-H. Wong, *J. Am. Chem. Soc.* **1998**, *120*, 11567.
- [9] C. Wang, M. Eufemi, C. Turano, A. Giartosio, *Biochemistry* **1996**, *35*, 7299.
- [10] C. Hammond, I. Braakman, A. Helenius, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 913.
- [11] S. Banerjee, P. Vishwanath, J. Cui, D. J. Kelleher, R. Gilmore, P. W. Robbins, J. Samuelson, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11676.
- [12] K. Witte, P. Sears, R. Martin, C.-H. Wong, *J. Am. Chem. Soc.* **1997**, *119*, 2114.
- [13] M. J. Grogan, M. R. Pratt, L. A. Marcaurelle, C. R. Bertozzi, *Annu. Rev. Biochem.* **2002**, *71*, 593.
- [14] C. S. Bennett, C.-H. Wong, *Chem. Soc. Rev.* **2007**, *36*, 1227.
- [15] Z. W. Zhang, J. Gildersleeve, Y. Y. Yang, R. Xu, J. A. Loo, S. Uryu, C.-H. Wong, P. G. Schultz, *Science* **2004**, *303*, 371.
- [16] R. Xu, S. R. Hanson, Z. W. Zhang, Y. Y. Yang, P. G. Schultz, C.-H. Wong, *J. Am. Chem. Soc.* **2004**, *126*, 15654.
- [17] S. R. Hamilton, R. C. Davidson, N. Sethuraman, J. H. Nett, Y. Jiang, S. Rios, P. Bobrowicz, T. A. Stadheim, H. Li, B.-K. Choi, D. Hopkins, H. Wischniewski, J. Roser, T. Mitchell, R. R. Strawbridge, J. Hoopes, S. Wildt, T. U. Gerngross, *Nature Biotechnol.* **2006**, *24*, 210.
- [18] H. Li, N. Sethuraman, T. A. Stadheim, D. Zha, B. Prinz, N. Ballew, P. Bobrowicz, B. K. Choi, W. J. Cook, M. Cukan, N. R. Houston-Cummings, R. Davidson, B. Gong, S. R. Hamilton, J. P. Hoopes, Y. Jiang, N. Kim, R. Mansfield, J. H. Nett, S. Rios, R. Strawbridge, S. Wildt, T. U. Gerngross, *Nature Biotechnol.* **2006**, *24*, 210.
- [19] T. U. Gerngross, *Nature Biotechnol.* **2004**, *22*, 1409.
- [20] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776.
- [21] T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249.
- [22] L. E. Canne, S. J. Bark, S. B. H. Kent, *J. Am. Chem. Soc.* **1996**, *118*, 5891.
- [23] P. E. Dawson, M. J. Churchill, M. R. Ghadiri, S. B. H. Kent, *J. Am. Chem. Soc.* **1997**, *119*, 4325.
- [24] T. M. Hackeng, J. H. Griffin, P. E. Dawson, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068.
- [25] Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman, C. R. Bertozzi, *J. Am. Chem. Soc.* **1999**, *121*, 11684.
- [26] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526.
- [27] J. Offer, C. N. C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642.
- [28] D. Bang, B. L. Pentelute, S. B. H. Kent, *Angew. Chem., Int. Ed.* **2006**, *45*, 3985.
- [29] N. Metanis, E. Keinan, P. E. Dawson, *J. Am. Chem. Soc.* **2006**, *128*, 16684.
- [30] M. E. Cremeens, H. Fujisaki, Y. Zhang, J. Zimmermann, L. B. Sagle, P. E. Dawson, J. E. Straub, F. E. Romesberg, *J. Am. Chem. Soc.* **2006**, *128*, 6028.
- [31] L. B. Sagle, J. Zimmermann, P. E. Dawson, F. E. Romesberg, *J. Am. Chem. Soc.* **2006**, *128*, 14232.
- [32] B. Wu, J. H. Chen, J. D. Warren, G. Chen, Z. H. Hua, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2006**, *45*, 4116.
- [33] P. Botti, M. R. Carrasco, S. B. H. Kent, *Tetrahedron Lett.* **2001**, *42*, 1831.
- [34] G. Chen, Q. Wan, Z. Tan, C. Kan, Z. H. Hua, K. Ranganathan, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2007**, *46*, 7383.
- [35] G. Chen, J. D. Warren, J. H. Chen, B. Wu, Q. Wan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, *128*, 7460.
- [36] N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson, Y. Kajihara, *J. Am. Chem. Soc.* **2008**, *130*, 501.
- [37] B. Wu, Z. P. Tan, G. Chen, J. H. Chen, Z. H. Hua, Q. Wan, K. Ranganathan, S. J. Danishefsky, *Tetrahedron Lett.* **2006**, *47*, 8009.
- [38] J. H. Chen, G. Chen, B. Wu, Q. Wan, Z. P. Tan, Z. H. Hua, S. J. Danishefsky, *Tetrahedron Lett.* **2006**, *47*, 8013.
- [39] A. Brik, Y.-Y. Yang, S. Ficht, C.-H. Wong, *J. Amer. Chem. Soc.* **2006**, *128*, 5626.
- [40] A. Brik, S. Ficht, Y.-Y. Yang, C. Bennett, C.-H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 15026.
- [41] Y.-Y. Yang, S. Ficht, A. Brik, C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 7690.
- [42] S. Ficht, R. J. Payne, A. Brik, C.-H. Wong, *Angew. Chem., Int. Ed.* **2007**, *46*, 5975.
- [43] R. J. Payne, S. Ficht, S. Tang, A. Brik, Y.-Y. Yang, D. A. Case, C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 13527.
- [44] R. J. Payne, S. Ficht, W. A. Greenberg, C.-H. Wong, *Angew. Chem., Int. Ed.* **2008**, *47*, 4411.
- [45] P. Wang, L. P. Miranda, *Int. J. Pept. Res. Therap.* **2005**, *11*, 117.
- [46] S. Ficht, R. J. Payne, R. T. Guy, C.-H. Wong, *Chem. Eur. J.* **2008**, *14*, 3620.
- [47] T. Kawakami, M. Sumida, K. Nakamura, T. Vorherr, S. Aimoto, *Tetrahedron Lett.* **2005**, *46*, 8805.
- [48] E. C. B. Johnson, S. B. H. Kent, *Tetrahedron Lett.* **2007**, *48*, 1795.
- [49] D. Lelievre, P. Barta, V. Augagne, A. F. Delmas, *Tetrahedron Lett.* **2008**, *49*, 4016.
- [50] J. B. Blanco-Canosa, P. E. Dawson, *Angew. Chem., Int. Ed.* **2008**, *47*, 6851.
- [51] C. S. Bennett, S. M. Dean, R. J. Payne, S. Ficht, A. Brik, C.-H. Wong, *J. Am. Chem. Soc.* **2008**, *130*, 11945.
- [52] B. Li, Y. Zeng, S. Hauser, H. J. Song, L. X. Wang, *J. Am. Chem. Soc.* **2005**, *127*, 9692.
- [53] Y. Zeng, J. S. Wang, S. Hauser, H. G. Li, L. X. Wang, *Chem. Eur. J.* **2006**, *12*, 3355.
- [54] B. Li, H. J. Song, S. Hauser, L. X. Wang, *Org. Lett.* **2006**, *8*, 3081.
- [55] T. W. D. E. Rising, C. D. Heidecke, J. W. D. Moir, Z. L. Ling, A. J. Fairbanks, *Chem. Eur. J.* **2008**, *14*, 6444.
- [56] H. Ochiai, W. Huang, L.-X. Wang, *J. Am. Chem. Soc.* **2008**, *130*, 13790.
- [57] D. Shental-Bechor, Y. Levy, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8266.
- [58] M. A. Recny, M. A. Luther, M. H. Knoppers, E. A. Neidhardt, S. S. Kandekar, M. F. Concino, P. A. Schimke, M. A. Francis, U. Moebius, B. B. Reinhold, V. N. Reinhold, E. L. Reinherz, *J. Biol. Chem.* **1992**, *267*, 22428.
- [59] D. F. Wyss, J. S. Choi, J. Li, M. H. Knoppers, K. J. Willis, A. R. N. Arulanandam, A. Smolyar, E. L. Reinherz, G. Wagner, *Science* **1995**, *269*, 1273.
- [60] A. Helenius, M. Aebi, *Science* **2001**, *291*, 2364.
- [61] S. R. Hanson, E. K. Culyba, T. L. Hsu, C.-H. Wong, J. W. Kelly, E. T. Powers, *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 3131.
- [62] M. Sawa, T.-L. Hsu, T. Itoh, M. Sugiyama, S. R. Hanson, P. K. Vogt, C.-H. Wong, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12371.
- [63] T.-L. Hsu, S. R. Hanson, K. Kishikawa, S.-K. Wang, M. Sawa, C.-H. Wong, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2614.
- [64] S. R. Hanson, T. L. Hsu, E. Weerapana, K. Kishikawa, G. M. Simon, B. F. Cravatt, C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 7266.
- [65] J. A. Prescher, C. R. Bertozzi, *Cell*, **2006**, *126*, 851.
- [66] N. Khidekel, S. B. Ficarro, E. C. Peters, L. C. Hsieh-Wilson, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13132.