NIH Public Access
Author Manuscript
J Am Chem Soc. Author manuscript; available in PMC 2008 September 24.

Published in final edited form as:
J Am Chem Soc. 2007 June 20; 129(24): 7690–7701. doi:10.1021/ja0708971.

Sugar-Assisted Ligation in Glycoprotein Synthesis

Yu-Ying Yang[1], **Simon Ficht**[1], **Ashraf Brik**[1,*,†], and **Chi-Huey Wong**[1,2,*]

1 *Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA*

2 *Genomic Research Center, Academia Sinica, 128 Sec. 2, Academia Road, Nankang, Taipei 115, Taiwan*

Abstract

Sugar-assisted ligation (SAL) presents an attractive strategy for the synthesis of glycopeptides, including the synthesis of cysteine-free β-O-linked and N-linked glycopeptides. Here we extended the utility of SAL for the synthesis of α-O-linked glycopeptides and glycoproteins. In order to explore SAL in the context of glycoprotein synthesis, we developed a new chemical synthetic route for the α-O-linked glycoprotein diptericin ε. In the first stage of our synthesis, diptericin segment Cys(Acm)^{37}-Gly^{52} and segment Val^{53}-Phe^{82} were assembled by SAL through a Gly-Val ligation junction. Subsequently, after Acm deprotection, diptericin segment Cys^{37}-Phe^{82} was ligated to segment Asp^{1}-Asn^{36} by means of native chemical ligation (NCL) to give the full sequence of diptericin ε. In the final synthetic step, hydrogenolysis was applied to remove the thiol handle from the sugar moiety with the concomitant conversion of mutated Cys^{37} into the native alanine residue. In addition, we extended the applicability of SAL to the synthesis of glycopeptides containing cysteine residues by carrying out selective desulfurization of the sulfhydryl-modified sugar moiety in the presence of acetamidomethyl (Acm) protected cysteine residues. The results presented here demonstrated for the first time that SAL could be a general and useful tool in the chemical synthesis of glycoproteins.

Introduction

Glycosylation represents the most prevalent post-translational modification of proteins. More than 50% of proteins in humans are glycosylated.[1] However, efforts toward understanding the effects of carbohydrates on their attached proteins are greatly hampered by the difficulty in obtaining sufficient amounts of pure single glycoform for studies. Unlike protein or DNA, oligosaccharide assembly is not controlled by a defined template.[2] As a result, glycoproteins usually exist as heterogeneous mixtures of glycoforms, which possess various glycosylation patterns on the same protein backbone. Recent advances in the development of glycoprotein synthesis, however, have created opportunities for researchers to obtain homogeneous glycoproteins in the laboratory.[3] For instance, Gerngross and co-workers recently have genetically engineered yeast to produce human therapeutic glycoproteins in single glycoforms.[3c–d]

In comparison to biological methods, synthetic approaches provide researchers more flexibility and control in manipulating the glycosylation pattern on the desired glycoprotein.[3a,4] The synthetic strategies for glycopeptides and glycoproteins generally involve several techniques, including carbohydrate synthesis, peptide assembly, and enzymatic elaboration.[3a, 4a,5] To successfully synthesize the desired glycopeptides or glycoproteins, every element in each

E-mail: abrik@bgu.ac.il, wong@scripps.edu.
†Current address: *Department of Chemistry, Ben Gurion University, Beer Sheva 84105, Israel*

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

technique has to be carefully coordinated. Native chemical ligation (NCL) represents the most widespread and powerful peptide ligation method for protein synthesis.[5a,5c,6] It allows the assembly of two unprotected synthetic peptides through a chemoselective reaction between peptide bearing N-terminal cysteine and a peptide bearing C-terminal thioester. This strategy was further strengthened by the development of expressed protein ligation (EPL) to extend its utility to the synthesis of large-sized proteins.[5d,7] However, in both NCL and EPL, the requirement of N-terminal cysteine for ligation imposes an intrinsic limitation to their application. For instance, to take advantage of these ligation methods, in many cases, introducing cysteine mutations into protein sequences is necessary, which could lead to the formation of mismatched disulfide bonds during protein folding.

To tackle this problem, different strategies were developed to expand the repertoire of ligation junctions of NCL into non-cysteinyl residues.[8] One of the most notable approaches is the removable auxiliary-assisted ligation strategy.[8c–d,9] In this strategy, a removable thiol auxiliary mimicking the function of N-terminal cysteine was anchored onto the N-terminus of a peptide to promote the transthioesterification and the subsequent intra-molecular S→N acyl transfer. The auxiliary was then removed after the ligation to regenerate the unmodified peptide chain. Various elegant auxiliaries have been developed to enable the ligation of peptides or glycopeptides through non-cysteinyl junctions[8c–d,9] Nevertheless, so far, most of these auxiliaries demand at least one Gly residue at its ligation site.[8c–d,9a–9e]

Recently, our group introduced a new chemoselective ligation method named sugar-assisted ligation (SAL) for assembling glycopeptides through non-cysteinyl residues and successfully demonstrated its utility in the synthesis of cysteine-free β-O-linked and N-linked glycopeptides.[10] SAL adopts similar concepts presented in NCL and removable auxiliary-assisted ligation. However, instead of anchoring the auxiliary onto the N-terminus of the peptide, we take advantage of the sugar by modifying its acetamido moiety at the C2 position with a sulfhydryl group to enable its transthioesterification with peptide thioester (Scheme 1). Notably, we believe the restricted conformation of the sugar moiety plays an important role in the S→N acyl transfer by bringing the N^{α}-amino group of glycopeptide and the electrophilic carbonyl group of the thioester intermediate into close proximity to facilitate the intra-molecular rearrangement via 14- or 15-membered ring.[11] After the ligation, the sulfhydryl group can be removed by hydrogenolysis to regenerate the unmodified sugar.[8b,10] Alternatively, the sulfhydryl-modified sugar can be further elaborated with glycosyltransferases to extend its glycan structure, or alkylated with labeling reagents such as fluorophores.[10b] One important feature about SAL, when compared to removable auxiliaries, is that the nucleophile that participates in the S→N rearrangement is a primary amine. This less sterically-hindered nucleophile is one of the attributing factors that allow SAL to accept a broader spectrum of amino acid residues at its ligation site. Preferences are for those amino acids possessing less hindered side chains (Gly and Ala), and those containing a side chain that can serve as a general base (His and Asp) in the ligation pathway.[10b]

This broad sequence tolerance exhibited in the ligation sites of SAL represents the most attractive advantage in this method for chemical glycoprotein synthesis. Given most naturally-occurring glycoproteins bear multiple glycosylation sites, SAL could be quite useful if these appending glycans of glycoproteins are eligible to participate in glycopeptide ligation. To enable SAL to be a general method for the preparation of homogeneous glycoproteins, herein, we made several advancements of this method, including extending its utility to the synthesis of α-*O*-linked glycopeptides, demonstrating selective desulfurization to enable this method for the synthesis of cysteine-containing glycopeptides, and more importantly, demonstrating its utility in the synthesis of the complete α-O-linked glycoprotein diptericin ε.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Diptericin ε is an 82-residue, cysteine-free α-O-linked antibacterial glycoprotein.[12] It contains two glycosylation sites, Thr^{10} and Thr^{54}, in which each threonine carries an N-acetyl-galactosamine (GalNAc) moiety (Scheme 3A).[5a,12,13] Previously, Bertozzi and coworkers have successfully applied NCL to chemically synthesize diptericin ε.[5a] The synthesis was accomplished through the assembly of two glycopeptide segments, in which the glycopeptide thioester was prepared by Fmoc-based solid-phase peptide synthesis (SPPS) using the Kenner linker. Notably, since diptericin ε contains no cysteine residue, the Bertozzi group mutated Gly^{25} to cysteine to allow NCL to be amenable to diptericin synthesis. In this report, we took advantage of the glycosylation site Thr^{54} of diptericin ε to apply SAL for the first time to the synthesis of a complete glycoprotein.

Results and Discussion

Synthesis of Building Blocks for SAL

Diptericin ε bears two GalNAc moieties.[12] GalNAc is the core structure of the mucin-type O-linked glycans, which are typically α-linked to serine or threonine residues.[14] Mucin-type O-linked glycans exist abundantly in naturally-occurring glycoproteins. To apply SAL for the synthesis of this type of α-O-linked glycopeptides and glycoproteins, first, we prepared the N^{α}-Fmoc-Thr[Ac_3-α-GalNAc(SH)] **4a** and N^{α}-Fmoc-Ser[Ac_3-α-GalNAc(SH)] **4b** as the building blocks for SPPS. As shown in Scheme 2, the synthetic strategy for building block **4a** is straightforward. The synthesis was started with TMSOTf-mediated glycosylation of compound **1**[15] with Fmoc-Thr-OAll at −78 °C to afford glycosylated product **2a** in α and β mixture (α/β = 4/3). Compound **2aα** was separated from the mixture using column chromatography to give the desired pure α-isomer in 45% yield. Following purification, the azido moiety was reduced to amine by treatment with Zn/AcOH.[16] The amino group was then coupled with *S*-trityl-2-mercaptoacetic acid[17] under standard HBTU/DIEA coupling condition[18] to give the compound **3a**. Subsequently, the allyl group was removed by treatment with catalytic amount of $Pd(PPh_3)_4$ in the presence of *N*-methylaniline[19] to furnish the building block **4a**. Building block **4b** was also prepared using a similar synthetic strategy.

Synthesis of α-O-Linked Glycopeptides using SAL

To investigate the ligation efficiency of SAL in α-O-linked glycopeptides, we synthesized four short glycopeptides **5a–b** and **6a–b** as well as peptide thioester $Cys(Acm)^{37}$-Gly^{52} for model studies (Table 1). Peptides containing sequence from diptericin ε were designed to explore the optimal ligation conditions for diptericin ε synthesis. Having the building blocks **4a** and **4b** in hand, we synthesized Ser[α-GalNAc(SH)]-containing glycopeptides **5a–b** and Thr[α-GalNAc (SH)]-containing glycopeptides **6a–b** by means of Fmoc-based SPPS[10b,20] on Rink-amide resin. Peptide thioester $Cys(Acm)^{37}$-Gly^{52} was prepared by Boc-based SPPS[10a,21] on MBHA resin. Herein, two ligation junctions Gly-Gly and Gly-Val were examined, in which Gly-Val represents the ligation junction found in diptericin ε (Scheme 3A). As the results show in Table 1, α-O-linked glycopeptides were successfully synthesized by SAL with the resulting ligation efficiencies similar to β-O-linked[10a] and *N*-linked glycopeptides[10b] (Table 1, entries 1–4).[22] GalNAc(SH)-Thr-containing glycopeptides exhibited ligation rates similar to GalNAc(SH)-Ser-containing glycopeptides, despite the additional steric factor contributed by the methyl group of threonine (Table 1, entries 1–4). The similarity of the ligation rates among various glycopeptides indicates that the feature of the glycosidic linkage as well as the glycan structure may have a negligible effect on the ligation efficiency. Notably, in this report, thiophenol was excluded from the ligation media in most of the experiments. As our earlier studies have shown that the rate-determining step of SAL is S→N acyl transfer[10b], this suggests that thiophenol may deteriorate the ligation efficiency, especially in cases exhibiting a slower ligation rate. As shown in entry 5 (Table 1), the ligation rate at the Gly-Val junction decreases when thiophenol was present in the reaction solution (entry 4 vs 5). On the other hand, to investigate the optimal

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

ligation conditions for the Gly-Val junction, a series of ligation studies were carried out using glycopeptide **6b** and peptide thioester Cys(Acm)^{37}-Gly^{52}. As shown in Table 1, the ligation efficiency at the Gly-Val junction is affected by the reaction pH (Table 1, entries 4 and 6). An ideal reaction pH for SAL is around 7.2 to 7.4, as measured by a microelectrode.[23] In this range of pH, SAL reaches its highest efficiency while giving a minimum amount of hydrolyzed thioester. As reactant concentrations were high, the best ligation efficiency was observed at high phosphate buffer concentration, to maintain the pH through the reaction (Table 1, entries 4 and 6). Moreover, the ligation efficiency at the Gly-Val junction can be increased by adding more equivalents of peptide thioester (Table 1, entries 4, 7 and 8).

Synthetic Strategy of Diptericin

Our synthetic strategy for diptericin ε is depicted in Scheme 3. In our synthetic design, the whole protein sequence was dissected into three segments, which were assembled sequentially from C to N-terminus by SAL and then NCL (Scheme 3). The residues Gly^{52}-Val^{53} next to the glycosylation site Thr^{54} represent an ideal ligation site to demonstrate the general utility of SAL. The sequence of the C-terminal segment Cys(Acm)^{37}-Phe^{82} and middle segment Cys (Acm)^{37}-Gly^{52} covers nearly all the standard L-amino acids, except Met and Glu. In addition, the relatively slow Gly-Val junction is difficult enough to represent a typical challenge in glycoprotein synthesis by SAL. Performing SAL gave us the 46-residue glycopeptide Cys (Acm)^{37}-Phe^{82}. After removing Acm, Cys^{37}-Phe^{82} was coupled to N-terminal segment Asp^{1}-Asn^{36} by NCL to generate the full-sized diptericin ε. By incorporating NCL into our synthetic strategy, the potential difficulty associated with the synthesis of a 52-residue glycopeptide thioester was bypassed. To take advantage of NCL, Ala^{37} was temporarily mutated to cysteine. In the final synthetic step, both the cysteine residue and the thiol handle on the sugar were desulfurized[8b,10] to complete the synthesis of diptericin ε.

Synthesis of Diptericin Segment Cys(Acm)^{37}-Phe^{82} by SAL

The C-terminal segment of diptericin, the glycopeptide Val^{53}-Phe^{82}, was prepared by Fmoc-SPPS[20] on 2-ClTrt resin using the building block **4a**. For the synthesis of the middle segment, the peptide thioester Cys(Acm)^{37}-Gly^{52}, the acetamidomethyl (Acm) group was used to protect N-terminal cysteine from self-ligation and intra-molecular cyclization. The coupling of these two segments by SAL was carried out in 6M Gn·HCl, 200 mM phosphate, pH = 8.5, where the concentration of glycopeptide Val^{53}-Phe^{82} and the peptide thioester Cys(Acm)^{37}-Gly^{52} was 6 mM and 12 mM, respectively. Due to the change of pH (the measured pH was 6.4) after mixing both peptides, the reaction pH was adjusted to 7.2 by carefully adding triethanolamine (~0.2 μL per 100 μL reaction volume).[23] The ligation was carried out at 37°C and monitored by LC/MS. After 48 h, LC/MS analysis showed that over 70% of the glycopeptide Val^{53}-Phe^{82} had been consumed[24], and there was only a trace amount of peptide thioester remaining. The ligated product was purified by reversed-phase HPLC and lyophilized to afford the glycopeptide Cys(Acm)^{37}-Phe^{82} in 28 % isolated yield. To increase the product yield, we added an extra 0.5 equivalent of peptide thioester into the reaction mixture after the ligation had proceded for 24 h. After 48 h, the reaction was quenched and subjected to HPLC purification. Using this method, the reaction was able to yield 36% of isolated product (Figure 1A). The glycopeptide Cys(Acm)^{37}-Phe^{82} was characterized by electrospray mass spectrometry (the reconstructed mass was 5408 Da, the calculated average mass: 5409 Da)

To explore the effect of the N-terminal amino acid residue of the glycopeptide Val^{53}-Phe^{82} on its ligation efficiency, we replaced Val^{53} with either Gly or His. Glycopeptide Gly^{53}-Phe^{82} and glycopeptide His^{53}-Phe^{82} were then ligated to the peptide thioester Cys(Acm)^{37}-Gly^{52} (1.5 equiv), respectively. These ligations were monitored by LC/MS. Interestingly, in the ligation of glycopeptide His^{53}-Phe^{82} with peptide thioester, the reaction proceeded smoothly to completion after 24 h even at pH 6.4 (Figure 1B). Given the pKa of the imidazole side chain

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of histidine is around 6.5, this experimental result further supports our previous hypothesis on His and Asp, in which the side chain of these residues can serve as a general base in the ligation pathway[9b]. On the other hand, the ligation of glycopeptide Gly^{53}-Phe^{82} with peptide thioester also proceeded to completion in 24 h, when the reaction pH was adjusted to 7.2. Together, these ligation results show that SAL is chemoselective and amenable for the synthesis of large sized glycopeptides.

Synthesis of the N-Terminal Diptericin Segment Glycopeptide Thioester Asp^1-Asn^{36}

Synthesis of glycopeptide thioesters in the context of glycoprotein synthesis remain a challenge. While Boc-based SPPS has been intensively used for the synthesis of peptide thioesters, its synthetic conditions are too harsh to be compatible with the presence of carbohydrates. As a result, many efforts have been aimed at adapting Fmoc-based SPPS to the synthesis of glycopeptide thioesters. [5a,5d,9e,25] One of the central problems with the Fmoc strategy is that the basic conditions used for removing acetate esters commonly used as protecting groups for sugars are not compatible with the presence of the thioester moiety. Consequently, the acetyl groups cannot be deprotected prior to the ligation step, which could complicate the synthesis or lead to a loss of ligated product in the O-acetate removal step.[5a,5d,25b] To facilitate the synthesis of fully unprotected glycopeptide thioesters Asp^1-Asn^{36}, here we introduced the acid-labile *p*-methoxybenzyl (PMB) group for the protection of the hydroxyl groups of GalNAc. In addition, since the C-terminal amino acid residue of the glycopeptide thioester is asparagine,[26] we were allowed to operate the whole synthesis of glycopeptide thioester Asp^1-Asn^{36} on solid support by using the side-chain anchoring strategy[27].

The preparation of building block **10** was started with the reaction of compound **7**[15] with *p*-thiocresol in the presence of $BF_3{\cdot}OEt_2$, which afforded the corresponding thioglycoside in 85% isolated yield (Scheme 4).[28] Following this step, the acetyl groups on the sugar were replaced with PMB groups to give compound **8**. The glycosylation of compound **8** with Fmoc-Thr-OAllyl was carried out at −15 °C to give compound **9** in α and β mixture (α/β = 3/2). The desired α isomer was then purified by column chromatography to afford compound **9α** in 45% yield. The azido group of compound **9α** was converted to the acetamido group by reductive acetylation using thioacetic acid in pyridine[29] Subsequently, the allyl group was removed by a catalytic amount of $Pd(PPh_3)_4$ in the presence of *N*-methylaniline[17] to furnish the compound **10**.

For the synthesis of glycopeptide thioester Asp^1-Asn^{36}, we employed the side-chain anchoring strategy by following the procedure reported by Wang et al. (Scheme 4).[27] The unprotected side-chain of Fmoc-Asp-OAllyl was first anchored onto Rink-amide resin, afterwards, the N-terminal amine was deblocked for glycopeptide synthesis. Fmoc-based SPPS was conducted by using standard HBTU/DIEA coupling conditions,[15,30] and the building block **10** was attached to the peptide as the 10th residue. Once the full sequence of the glycopeptide was completed, the C-terminal carboxylic acid was unmasked and transformed into C^{α}-thioester by treatment with 3-mercaptopropionic acid in the presence of DIC/HOBt and DIEA in DMF/DCM (4/1) [27]. Following this thioesterification step, the glycopeptide thioester was cleaved from the resin with concomitant full deprotection by TFA/triethylsilane/thioanisole/H_2O (85/5/5/5). The crude product was then purified by preparative reversed-phase HPLC and lyophilized to afford the glycopeptide thioester Asp^1-Asn^{36} in 9% isolated yield. This compound was characterized by electrospray mass spectrometry (the reconstructed mass was 3975 Da, the calculated average mass: 3975 Da).

Acm Removal, NCL and Desulfurization

Having the glycopeptide segments in hand, we turned our focus to generation of the full sequence of diptericin ε by means of NCL. The N-terminal cysteine of glycopeptide Cys

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(Acm)^{37}-Phe^{82} was first deblocked by treatment with $Hg(OAc)_2$/AcOH (pH = 4.0)[31] to give the glycopeptide Cys^{37}-Phe^{82} in 83% isolated yield (see supporting information). The glycopeptide Cys^{37}-Phe^{82} (3 mM) and glycopeptide thioester Asp^{1}-Asn^{36} (4.5 mM) were then ligated in 6.0 M Gn·HCl, 0.2 M phosphate buffer, pH = 7.9 containing 2% (v/v) PhSH and 2% (v/v) BnSH at 37 °C. After 16 h, the ligation was nearly complete and the ligated product appeared as the major peak by HPLC. After reversed-phased HPLC purification and lyophilization, the pure glycopeptide Asp^{1}-Phe^{82} was obtained in 47% yield (see supporting information). Finally, the ligated product was subjected to hydrogenolysis under the conditions of H_2/(Pd/Al_2O_3), to reduce the thiol handle attached to GalNAc, as well as the cysteine residue.[8b] After HPLC purification and lyophilization, the full-sized diptericin was obtained in 54% isolated yield (254 μg). The final product was characterized by electrospray mass spectrometry (the reconstructed mass was 9113 Da, the calculated average mass: 9114 Da).

Selective Desulfurization

To allow SAL to be amenable for the synthesis of glycopeptides or glycoproteins containing cysteine residues, the sulfhydryl-modified sugar moiety has to be desulfurized[8b] selectively in the hydrogenolysis. To investigate desulfurization selectivity, we synthesized a model glycopeptide **11** bearing one Cys(Acm) residue and a sulfhydryl-modified GalNAc α-linked to a Ser residue. Previously, we have shown that the thiol handle of the sugar moiety can be easily removed under H_2 in the presence of 10–20 mg of Pd/Al_2O_3 per 1mg of glycopeptide. Herein, the hydrogenolysis of compound **11** was carried out in the same condition previously reported[10], 6M Gn·HCl, 0.1 M phosphate buffer, pH 5.8 containing 10 mM TCEP, however, in order to obtain desulfurization selectivity, we performed this reaction by varying the amount of Pd/Al_2O_3 added. As a consequence, the reaction containing 1 mg glycopeptide was reacted with 10, 15 and 20 mg of Pd/Al_2O_3, respectively. The progress of the desulfurization in these reactions was carefully monitored by LC/MS. After 2.5 h, the desired product **12**, in which the mass corresponds to a loss of 32 Da from compound **11**, was observed as the major peak on the HPLC chromatograph (Figure 3). The analytical data from these three reactions showed that the use of 15 mg Pd/Al_2O_3 gave the best selectivity and afforded the product in 77% isolated yield. The reaction using 20 mg of Pd/Al_2O_3 completed within 1 hr, however, it gave poor desulfurization selectivity [**12**/**13**, ~3:1]. To examine the molecular structure of the selectively desulfurized product **12**, we removed its Acm group and then subjected the product to NCL with the peptide thioester Cys(Acm)^{37}-Phe^{82} (see supporting material). After 4 h, the clean ligated product as shown in the analytical HPLC chromatograph suggested the selective desulfurization would allow the SAL product to be sequentially ligated to another peptide or glycopeptide. In addition, we also tested the desulfurization selectivity in the glycopeptide containing two Cys(Acm) residues and a 46-residue glycopeptide Cys(Acm)^{37}-Phe^{82} (see supporting material). Similar desulfurization selectivity was obtained in these model studies, however, we noticed that the desulfurization efficiency in long glycopeptide decreased, which in turn affected its desulfurization selectivity (see supporting information).[32]

Conclusion

We have shown that SAL can be extended to the synthesis of α-O-linked glycopeptides, in which the ligation rates are similar to β-O-linked and N-linked glycopeptides, making it a general method for various glycopeptide syntheses. For the first time, we have demonstrated that SAL could be a potentially useful approach for the synthesis of glycoproteins, as we demonstrated by the synthesis of diptericin ε. Our synthetic scheme combined the advantages given by SAL, NCL, and desulfurization. Here, SAL aided us to prepare the 46-resdiue C-terminal glycopeptide, which after Acm removal, was subsequently ligated to the 36-residue N-terminal segment by NCL to achieve the full-sized diptericin ε. In the final step, the desulfurization reaction was used to remove the thiol handle from the sugar moiety and to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

convert the cysteine into the native alanine residue. Moreover, we have demonstrated the sulfhydryl-modified sugar moiety can be selectively desulfurized in the presence of Cys(Acm), which expands the utility of SAL into the synthesis of cysteine-containing glycopeptides or glycoproteins. Based on these results, we believe that SAL, in conjunction with other ligation methods, should facilitate the synthesis of homogeneous glycoproteins, which will help to elucidate the effects of specific glycans on protein structure and function.[33]

Experimental Procedures

General Methods and Materials

Tetrahydrofuran (THF) was distilled over sodium/benzophenone, and methylene chloride (CH_2Cl_2) was distilled over calcium chloride. Reagents of commercial quality were purchased and used without further purification. Glycosylation experiments were performed using molecular sieves (AW 300), which were flame-dried under high vacuum right before the reaction. Analytical thin-layer chromatography (TLC) was performed using silica gel 60 F_{254} glass plates, compound spots were visualized by UV light (254 nm) and by staining with acidic ceric ammonium molybdate. Flash chromatography was performed on silica gel 60 Geduran (35–75 μm, EM Science). ^{1}H, ^{13}C-NMR spectra were recorded on a Bruker AMX-500 MHz spectrometer. Coupling constants (*J*) are reported in hertz (Hz), and chemical shifts are reported in parts per million (ppm) relative to tetramethylsilane (TMS, 0.0 ppm). Fmoc-amino acids, Boc-amino acids, HBTU, PyBOP, pre-loaded 2-ClTrt resins, MBHA resin LL (100–200 mesh) · HCl and Rink-amide resin were purchased from Novabiochem. Boc-His(3-Bom)-OH was purchase from Bachem. HOBt, *N*,*N*-Diisopropylethylamine (DIPEA), *N*,*N*-dimethylformamide (Biotech grade), HPLC grade acetonitrile, trifluoroacetic acid were purchased from Sigma-Aldrich and Fisher.

Mass Spectrometry

MALDI-TOF/MS (Applied Biosystems DE) and ESI-TOF (IonSpec Ultima FTMS) mass spectra were obtained from the routine analyses service at the Scripps Center of Mass Spectrometry. LC/MS mass spectrometric analyses were performed on the Agilent 1100 MSD LC/MS system which utilizes the electrospray ionization method.

Solid-Phase Peptide Synthesis

Solid-phase peptide syntheses were carried out manually with syringes (Torviq) which were equipped with teflon filters. Preparative, semi-preparative and analytical reversed-phase HPLC purification was performed on a Hitachi D-7000 HPLC system. All C18 HPLC columns were purchased from Grace Vydac. The flow rates used for HPLC were 1.5 ml/min (analytical), 4.0 ml/min (semipreparative) and 8.0 ml/min (preparative). **Fmoc-SPPS:** *Fmoc Cleavage*: After treatment with 20% piperidine in DMF (2 × 10 min), the resin was washed with DMF (4 ×), CH_2Cl_2 (4 ×) and DMF (4 ×). *Coupling*: After preactivation of Fmoc-amino acid (5 equiv) for 5 min using HBTU (4 equiv, final concentration 0.5 M in DMF), HOBt (2 equiv), and DIPEA (10 equiv), the solution was added to the resin. After 1 hr, the resin was washed with DMF (4 ×), CH_2Cl_2 (2 ×), and DMF (2 ×). Glycosylated amino acid was coupled to the resin for 24 h. Double coupling were performed for some bulky amino acids. *Acetyl group Removal*: The resin was first washed with MeOH (10 ×), and then treated with hydrazine/MeOH (1/6) (3 × 2h). The resulting resin was washed with MeOH (10 ×), DMF (5 ×), and CH_2Cl_2 (5 ×). **Boc-SPPS:** *Boc Cleavage*: After treatment with 5% *m*-cresol/TFA (2 × 4 min), the resin was washed with DCM (4 ×) and with DMF (4 ×). *Coupling*: After preactivation of Boc-amino acid (5 equiv) for 5 min using PyBOP (4 equiv, final concentration 0.5 M in DMF) and *N*-methylmorpholine (8 equiv), the solution was added to the resin. After 45 min, the resin was washed with DMF (3 ×), DCM (4 ×). *Cleavage*: HF with 10% (v/v) anisole was used as cleavage solution. Once the HF was removed, cold ethyl ether was added to the crude reaction product

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

and stirred for 10 min. The precipitate was collected and re-dissolved in H_2O/CH_3CN (9/1) for HPLC purification.

Compound 2aα—Compound **1** (5.0 g, 10.5 mmol) and Fmoc-Thr-OAllyl (4.83 g, 12.7 mmol) were mixed and dried under high vacuum overnight prior to the reaction. The mixture was then dissolved in anhydrous CH_2Cl_2 (30 mL) under N_2. 3.0 g of fresh flame-dried molecule sieves were added and the mixture was stirred at room temperature for 1 h. The reaction was then cooled down to −78 °C, and TMSOTf (203.9 μL, 1.066 mmol) was added slowly into the reaction solution. The mixture was kept at −78 °C and stirred for 45 min. Once the reaction was down, it was quenched by adding Et_3N and the temperature was allowed to rise to room temperature. Molecule sieves were removed by filtration through Celite, and the filtrate was concentrated under vacuum. The crude product was purified by flash column chromatography (20–50% EtOAc in hexane) to give the pure α product (3.31 g, 4.77 mmol, 45.4 %). ^{1}H-NMR ($CDCl_3$, 500 MHz): δ 7.76 (d, 2H, *J* = 7.4 Hz), 7.62–7.64 (m, 2H), 7.36–7.41 (m, 2H), 7.30–7.34 (m, 2H), 5.90–5.99 (m, 1H), 5.68 (d, 1H, *J* = 9.6 Hz), 5.46 (brs, 1H), 5.26–5.39 (m, 3H), 5.05 (d, 1H, *J* = 3.3 Hz), 4.69 (d, 1H, *J* = 5.2 Hz), 4.41–4.48 (m, 3H), 4.36 (dd, 1H, *J* = 10.3, 7.3 Hz), 4.26–4.29 (m, 2H), 4.09 (d, 1H, *J* = 6.6 Hz), 3.67 (dd, 1H, *J* = 11.4, 3.7 Hz), 2.14 (s, 3H), 2.07 (s, 3H), 2.03 (s, 3H), 1.36 (d, 3H, *J* = 6.2 Hz); ^{13}C-NMR ($CDCl_3$, 125 MHz):δ170.2, 169.9, 169.7, 156.7, 143.7, 143.6, 141.2, 141.1, 131.2, 127.6, 127.0, 125.2, 125.1, 119.9, 119.8, 119.2, 99.3, 76.9, 68.1, 67.4, 67.3, 66.9, 66.4, 61.7, 58.7, 57.6, 47.0, 20.6, 20.5, 20.4, 18.4; HRMS (ESI-TOF) calcd for $C_{34}H_{38}N_4O_{12}$ $[M+H]^+$: 695.2559. Found: 695.2553.

Compound 3a—Compound **2aα** (2.68 g, 3.86 mmol) was dissolved in AcOH (25 mL), followed by adding Zinc powder (15.0 g). The reaction mixture was stirred for 4 h at room temperature. Once the reaction was done, Zinc was filtered off through Celite and washed with CH_2Cl_2. The filtrate was concentrated under vacuum. The concentrated residue was then purified by flash column chromatography (20% hexane in EtOAc) to give the pure amine product (2.40 g, 3.62 mmole, 94%). ^{1}H-NMR ($CDCl_3$, 500 MHz): δ 7.76 (d, 2H, *J* = 7.4 Hz), 7.64 (d, 2H, *J* = 5.9 Hz), 7.38–7.41 (m, 2H), 7.30–7.35 (m, 2H), 5.90–5.98 (m, 1H), 5.77 (d, 1H, *J* = 9.2 Hz), 5.34–5.38 (m, 2H), 5.27 (d, 1H, *J* = 10.7 Hz), 4.97 (d, 1H, *J* = 3.0 Hz), 4.88–4.92 (m, 1H), 4.65–4.74 (m, 2H), 4.41–4.49 (m, 3H), 4.23–4.33 (m, 3H), 4.08–4.10 (m, 2H), 3.15 (dd, 1H, *J* = 10.7, 3.0 Hz), 2.11 (s, 3H), 2.04 (s, 3H), 2.03 (s, 3H), 1.35 (d, 3H, *J* = 6.2 Hz); ^{13}C-NMR ($CDCl_3$, 125 MHz):δ170.5, 170.3, 170.2, 170.1, 156.5, 143.8, 143.7, 141.2, 131.3, 127.6, 127.0, 125.1, 125.1, 119.9, 119.9, 119.3, 101.8, 77.3, 71.8, 67.4, 67.3, 67.2, 66.4, 62.1, 58.6, 49.7, 47.0, 20.7, 20.6, 20.5, 18.1; HRMS (ESI-TOF) calcd for $C_{34}H_{40}N_2O_{12}$ $[M+H]^+$: 669.2654. Found: 669.2641. The amine (650 mg, 0.973 mmol) was dissolved in DMF (15 mL). TrtS-CH_2COOH (651 mg, 1.95 mmol), HBTU (740 mg, 1.95 mmol) and DIPEA (678 μL, 3.89 mmol) were added to the amine subsequently. The reaction mixture was stirred at room temperature for 2 h. Once the reaction was done, the reaction solution was concentrated under vacuum. The concentrated residue was then purified by flash column chromatography (20–40% EtOAc in hexane) to give pure compound **3a** (877.7 mg, 0.892 mmol, 91%). ^{1}H-NMR ($CDCl_3$, 500 MHz): δ 8.00 (brs, 1H), 7.77 (d, 2H, *J* = 7.3 Hz), 7.63 (d, 2H, *J* = 7.3 Hz), 7.16–7.47 (m, 19H), 5.95–6.01 (m, 1H), 5.72–5.82 (m, 2H), 5.39 (brs, 1H), 5.20–5.25 (m, 2H), 5.03–5.08 (m, 1H), 4.86–4.89 (m, 1H), 4.37–4.50 (m, 6H), 4.19–4.28 (m, 3H), 4.04–4.13 (m, 2H), 2.14 (s, 3H), 2.03 (s, 3H), 1.95 (s, 3H), 1.28 (d, 3H, *J* = 6.2 Hz); ^{13}C-NMR ($CDCl_3$, 125 MHz): δ 171.2, 170.7, 170.6, 170.4, 169.2, 163.0, 156.9, 144.5, 144.2, 144.1, 141.7, 131.4, 129.9, 128.4, 128.1, 127.5, 127.3, 125.5, 125.4, 120.5, 120.4, 120.1, 99.8, 68.6, 67.7, 67.6, 67.6, 66.6, 62.5, 58.8, 48.2, 47.6, 39.0, 37.1, 36.9, 31.8, 21.2, 21.1, 21.0, 18.5; HRMS (ESI-TOF) calcd for $C_{55}H_{56}N_2O_{13}S$ $[M+Na]^+$: 1007.3395. Found: 1007.3390.

Compound 4a—Compound **3a** (750 mg, 0.762 mmol) was dissolved in THF (20 mL). Pd $(PPh_3)_4$ (88.0 mg, 0.076 mmol) and *N*-methylaniline (827 μL, 7.62 mmol) were added, and

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

then the reaction mixture was stirred at room temperature for 45 min. Once the reaction was done, the solution was concentrated under vacuum. The crude product was purified by flash column chromatography (50% EtOAc in hexane then 10% MeOH in CH_2Cl_2) to give pure compound **4a** (690 mg, 0.731mmol, 95%). ^{1}H-NMR (MeOD, 500 MHz): δ 7.75 (d, 2H, *J* = 7.7 Hz), 7.57–7.63 (m, 3H), 7.47–7.50 (m, 1H), 7.31–7.33 (m, 6H), 7.12–7.26 (m, 11H), 5.38 (brs, 1H), 5.09 (dd, 1H, *J* = 11.4, 2.9 Hz), 4.96 (brs, 1H), 4.24–4.45 (m, 6H), 4.13–4.16 (m, 2H), 4.03–4.09 (m, 3H), 2.99 (d, 1H, *J* = 13.2 Hz), 2.91 (d, 1H, *J* = 13.2 Hz), 2.10 (s, 3H), 1.96 (s, 3H), 1.91 (s, 3H), 1.19 (d, 3H, *J* = 7.0 Hz); ^{13}C-NMR (MeOD, 125 MHz):δ172.2, 172.1, 171.9, 171.9, 158.9, 145.7, 145.5, 145.1, 142.7, 142.7, 133.8, 133.8, 133.2, 133.1, 130.8, 130.1, 130.0, 129.1, 128.9, 128.3, 128.1, 126.2, 126.1, 121.1, 121.1, 100.3, 77.6, 69.7, 68.9, 68.3, 68.2, 67.8, 63.4, 61.6, 60.3, 37.6, 21.0, 21.0, 20.7, 20.7, 19.2, 14.6; HRMS (ESI-TOF) calcd for $C_{52}H_{52}N_2O_{13}S$ $[M+Na]^+$: 967.3082. Found: 967.3059.

Compound 2bα—Compound **1** (1.33 g, 2.81 mmol) and Fmoc-Ser-OAllyl (858.8 mg, 2.34 mmol) were mixed together and dried under high vacuum overnight prior to the reaction. The mixture was then dissolved in anhydrous CH_2Cl_2 (15 mL) under N_2. To this reaction mixture, 1.5 g of fresh flame-dried molecule sieves were added, afterwards, it was stirred at room temperature for 1 h. The reaction mixture was then cooled down to −78 °C, followed by slowly adding TMSOTf (54.3 μL, 0.281 mmol) into the reaction solution. The reaction was kept at −78 °C and stirred for 45 min. Once the reaction was done, it was quenched by adding Et_3N and the temperature was allowed to rise to room temperature. Molecule sieves were then filtered off through Celite, and the filtrate was concentrated under vacuum. The crude product was purified by flash column chromatography (20–50% EtOAc in hexane) to give pure α product (872 mg, 1.28 mmol, 45%). ^{1}H-NMR ($CDCl_3$, 500 MHz): δ 7.76 (d, 2H, *J* = 7.7 Hz), 7.62 (dd, 2H, *J* = 7.0, 3.7 Hz), 7.40 (dd, 2H, *J* = 7.7, 7.7 Hz), 7.32 (td, 2H, *J* = 7.7, 3.3 Hz), 5.94 (m, 2H), 5.44 (brs, 1H), 5.28–5.38 (m, 3H), 4.94 (d, 1H, *J* = 3.3 Hz), 4.69–4.71 (m, 2H), 4.59–4.60 (m, 1H), 4.41 (d, 2H, *J* = 7.0 Hz), 4.25 (t, 1H, *J* = 7.3 Hz), 4.19 (t, 1H, *J* = 6.6 Hz), 4.13 (dd, 1H, *J* = 10.6, 2.9 Hz), 4.00-.4.04 (m, 2H), 3.63 (dd, 1H, *J* = 11.0, 3.3 Hz), 2.14 (s, 3H), 2.06 (s, 3H), 1.97 (s, 3H); ^{13}C-NMR ($CDCl_3$, 125 MHz):δ 170.3, 169.8, 169.6, 169.1, 155.7, 143.7, 141.2, 131.2, 127.7, 127.6, 127.0, 125.0, 119.9, 119.2, 99.2, 69.7, 67.8, 67.4, 67.2, 67.1, 66.5, 61.6, 57.3, 54.4, 47.0, 20.5, 20.5, 20.4; HRMS (ESI-TOF) calcd for $C_{33}H_{36}N_4O_{12}$ $[M+Na]^+$: 703.2222. Found: 703.2208.

Compound 3b—The amine (822 mg, 1.26 mmol) [HRMS (ESI-TOF) calcd for $C_{33}H_{38}N_2O_{12}$ $[M+H]^+$: 655.2497. Found: 655.2499] was dissolved in DMF (10 mL). TrtS-CH_2COOH (840 mg, 2.51 mmol), HBTU (953 mg, 2.51 mmol) and DIPEA (875 μL, 5.03 mmol) were added to the amine subsequently. The reaction was stirred at room temperature for 2 h. Once the reaction was done, the solution was concentrated under vacuum. The concentrated residue was then purified by flash column chromatography (20–40% EtOAc in hexane) to give pure compound **3b** (1.03 g, 1.06 mmol, 84%). ^{1}H-NMR ($CDCl_3$, 500 MHz): δ 7.76 (d, 2H, *J* = 7.3 Hz), 7.60 (d, 2H, *J* = 5.2 Hz), 7.36–7.41 (m, 8H), 7.25–7.32 (m, 8H), 7.19–7.22 (m, 3H), 5.90–5.96 (m, 2H), 5.78–5.86 (m, 1H), 5.35 (d, 1H, *J* = 2.9 Hz), 5.27 (d, 1H, *J* = 17.3 Hz), 5.23 (d, 1H, *J* = 10.3 Hz), 4.98 (dd, 1H, *J* = 11.4, 2.6 Hz), 4.71 (brs, 1H), 4.41–4.56 (m, 6H), 4.22 (t, 1H, *J* = 6.6 Hz), 4.06–4.11 (m, 2H), 3.98–4.01 (m, 2H), 3.85–3.87 (m, 1H), 2.13 (s, 3H), 1.96 (s, 3H), 1.93 (s, 3H); ^{13}C-NMR ($CDCl_3$, 125 MHz):δ 170.4, 170.2, 170.0, 169.2, 168.5, 155.6, 143.8, 143.6, 141.2, 131.0, 129.3, 128.0, 127.6, 127.0, 124.9, 119.9, 119.3, 98.8, 69.8, 68.0, 67.6, 67.2, 67.0, 66.2, 61.8, 54.3, 47.6, 46.9, 38.5, 36.3, 20.6, 20.5, 20.4; HRMS (ESI-TOF) calcd for $C_{54}H_{54}N_2O_{13}S$ $[M+Na]^+$:993.3239. Found: 993.3240.

Compound 4b—Compound **3b** (420 mg, 0.433 mmol) was dissolved in THF (10 mL), followed by adding $Pd(PPh_3)_4$ (50.0 mg, 0.043 mmol) and *N*-methylaniline (470 μL, 4.33 mmol). The reaction mixture was stirred at room temperature for 45 min. Once the reaction

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

was done, the reaction solution was concentrated under vacuum. The crude product was purified by flash column chromatography (50% EtOAc in hexane then 10% MeOH in CH_2Cl_2) to give pure compound **4b** (390 mg, 0.420 mmol, 96%). ^{1}H-NMR (MeOD, 500 MHz): δ 7.76–7.78 (m, 2H), 7.60–7.65 (m, 2H), 7.16–7.41 (m, 19H), 5.36 (dd, 1H, *J* = 2.6 Hz), 5.11–5.14 (m, 1H), 4.82 (brs, 1H), 4.35 (dd, 1H, *J* = 11.4, 3.3 Hz), 4.27–4.32 (m, 3H), 4.22 (t, 1H, *J* = 6.6 Hz), 4.17 (t, 1H, *J* = 6.6 Hz), 4.05 (dd, 1H, *J* = 11.0, 6.3 Hz), 3.99 (dd, 1H, *J* = 11.0, 7.0 Hz), 3.91–3.93 (m, 1H), 3.82 (dd, 1H, *J* = 10.3, 5.1 Hz), 2.82–2.93 (m, 2H), 2.13 (s, 3H), 1.92 (s, 3H), 1.89 (s, 3H); ^{13}C-NMR (MeOD, 125 MHz):δ 172.3, 172.2, 171.9, 171.8, 158.4, 146.0, 145.7, 145.4, 145.3, 142.7, 133.9, 133.2, 133.1, 130.8, 130.1, 130.0, 129.9, 129.2, 129.1, 128.9, 128.8, 128.3, 128.1, 128.0, 126.3, 121.1, 99.8, 70.4, 69.8, 68.8, 68.4, 68.2, 68.1, 63.0, 57.2, 37.5, 20.9, 20.7; HRMS (ESI-TOF) calcd for $C_{51}H_{50}N_2O_{13}S$ $[M+Na]^+$: 953.2926. Found: 953.2892.

Compound 8α—Compound **7** (10.28 g, 27.55 mmol) and *p*-thiocresol (5.1 g, 41.3 mmol) were dissolved in CH_2Cl_2 (40 mL) under $N_{2\,(g)}$. The reaction solution was cooled down to 0 °C in ice bath. To this solution, $BF_3{\cdot}OEt_2$ (6.92 mL, 55.1 mmol) was added slowly, afterwards, the reaction temperature was allowed to gradually increase to room temperature. After 8 h, the reaction was quenched by slowly adding saturated $NaHCO_3$ and diluted with CH_2Cl_2. The product was washed with saturated $NaHCO_3$ and extracted with CH_2Cl_2. The organic layers were collected and concentrated under vacuum. The concentrated residue was then purified by flash column chromatography (20–40% EtOAc in hexane) to give the pure product (α and β mixtures, 10.3 g, 23.6 mmol, 85%). The purified thioglycoside were dissolved in MeOH. NaOMe (25% wt in MeOH) was added drop wisely to the reaction solution until the pH reached around 11–12. The reaction mixture was stirred at room temperature for 1 h and neutralized with acidic resin (Dowex 50WX2-200(H)). The resin was then filtered off and washed with CH_2Cl_2. The filtrate was concentrated under vacuum. The concentrated residue was then purified by flash column chromatography (5–10% MeOH in CH_2Cl_2). The pure thioglycoside (7.83g, 25.2 mmol) was then dissolved in 60 mL of dry DMF at 0 °C and fully protected with PMB groups by adding PMBCl (20.5 mL, 151 mmol) and NaH (3.62 g, 151 mmol). The reaction mixture was stirred at room temperature for 2 h and quenched with MeOH. After removing MeOH, it was concentrated under vacuum. Then the product was extracted with EtOAc and washed with H_2O. The organic layers were collected and washed again with brine. After concentrating the organic layer under vacuum, the crude residue was purified by flash column chromatography (10–40% EtOAc in hexane) to give the pure PMB protected thioglycoside **8α** and **8β**, respectively. [15.5 g, 23.2 mmol(α + β), 92%]. ^{1}H-NMR ($CDCl_3$, 500 MHz): δ 7.37 (d, 2H, *J* = 8.1 Hz), 7.33 (d, 2H, *J* = 8.8 Hz), 7.19 (d, 2H, *J* = 8.5 Hz), 7.18 (d, 2H, *J* = 8.8 Hz), 7.02 (d, 2H, *J* = 7.8 Hz), 6.90 (d, 2H, *J* = 8.4 Hz), 6.86 (d, 2H, *J* = 8.5 Hz), 6.81 (d, 2H, *J* = 8.5 Hz), 5.50 (d, 1H, *J* = 5.5 Hz), 4.80 (d, 1H, *J* = 11.0 Hz), 4.63–4.67 (m, 2H), 4.32–4.47 (m, 5H), 3.98 (brs, 1H), 3.80 (s, 3H), 3.79 (s, 3H), 3.77 (s, 3H), 3.74 (dd, 1H, *J* = 10.6, 2.6 Hz), 3.55 (dd, 1H, *J* = 9.2, 7.0 Hz), 3.47 (dd, 1H, *J* = 9.2, 5.9 Hz), 2.29 (s, 3H); ^{13}C-NMR ($CDCl_3$, 125 MHz): δ 159.4, 159.3, 159.2, 137.6, 132.7, 130.4, 129.9, 129.8, 129.7, 129.6, 129.5, 129.4, 129.3, 113.9, 113.7, 113.6, 87.9, 78.7, 74.4, 73.1, 73.0, 72.0, 70.4, 68.3, 60.3, 55.3, 55.2, 55.1, 21.0; HRMS (ESI-TOF) calcd for $C_{37}H_{41}N_3O_7S$ $[M+Na]^+$: 694.2557. Found: 694.2556

Compound 8β—^{1}H-NMR ($CDCl_3$, 500 MHz): δ 7.43 (d, 2H, *J* = 7.7 Hz), 7.27 (d, 2H, *J* = 8.4 Hz), 7.18 (d, 2H, *J* = 8.4 Hz), 7.13 (d, 2H, *J* = 8.4 Hz), 6.96 (d, 2H, *J* = 7.7 Hz), 6.80–6.87 (m, 6H), 4.75 (d, 1H, *J* = 11.0 Hz), 4.60 (d, 1H, *J* = 11.0 Hz), 4.54 (d, 1H, *J* = 11.3 Hz), 4.43 (d, 1H, *J* = 11.0 Hz), 4.40 (d, 1H, *J* = 11.3 Hz), 4.33 (d, 1H, *J* = 11.8 Hz), 4.31 (d, 1H, *J* = 10.3 Hz), 3.86 (d, 1H, *J* = 1.8 Hz), 3.73–3.78 (m, 11H), 3.57 (d, 1H, *J* = 6.6 Hz), 3.49 (dd, 1H, *J* = 6.3, 6.3 Hz), 3.34 (dd, 1H, *J* = 9.5, 2.2 Hz), 2.26 (s, 3H); ^{13}C-NMR ($CDCl_3$, 125 MHz): δ 159.2, 159.1, 158.8, 137.7, 133.0, 130.5, 129.7, 129.4, 129.3, 129.2, 129.1, 128.0, 113.7, 113.6,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

113.3, 86.3, 82.0, 77.2, 73.8, 73.0, 71.7, 71.5, 68.0, 61.3, 55.0, 20.9; HRMS (ESI-TOF) calcd for $C_{37}H_{41}N_3O_7S$ $[M+Na]^+$: 694.2557. Found: 694.2556.

Fmoc-Thr-O-Allyl—Cesium carbonate (2.28 g, 7.0 mmol) was added to a suspension of Fmoc-Thr-OH (4.75 g, 13.9 mmol) in dry MeOH (40 mL) under $N_{2\ (g)}$. The reaction mixture was stirred at room temperature for 2 h and then evaporated to dryness under vacuum. The concentrated residue was further dried under high vacuum for another 2 h. Afterwards, the dried mixture was re-dissolved in dry DMF (40 mL) under N_2 and added with allylbromide (1.45 mL, 16.7 mmol). The reaction mixture was stirred at room temperature for 8 h. The white precipitate was filtered off through Celite and washed with CH_2Cl_2. The filtrate was concentrated under vacuum, and then was purified by flash column chromatography (25–50% EtOAc in hexane) to give the pure white powder (4.64 g, 12.2 mmol, 87%). ^{1}H-NMR ($CDCl_3$, 500 MHz): δ 7.72 (d, 2H, *J* = 7.4 Hz), 7.59 (d, 1H, *J* = 5.9 Hz), 7.58 (d, 1H, *J* = 5.9 Hz), 7.36 (d, 1H, *J* = 7.3 Hz), 7.35 (d, 1H, *J* = 7.7 Hz), 7.27 (d, 1H, *J* = 7.3 Hz), 7.25 (d, 1H, *J* = 7.4 Hz), 5.92 (d, 1H, *J* = 9.2 Hz), 5.82–5.89 (m, 1H), 5.29 (d, 1H, *J* = 17.2 Hz), 5.20 (d, 1H, *J* = 10.3 Hz), 4.62–4.64 (m, 2H), 4.36–4.41 (m, 4H), 4.20 (t, 1H, *J* = 7.0 Hz), 1.23 (d, 3H, *J* = 5.9 Hz); ^{13}C-NMR ($CDCl_3$, 125 MHz): δ 170.8, 156.7, 143.6, 143.5, 141.1, 131.2, 127.5, 126.9, 125.0, 119.8, 118.7, 67.7, 67.0, 66.0, 59.1, 46.9, 19.7; HRMS (ESI-TOF) calcd for $C_{22}H_{23}NO_5$ $[M+Na]^+$: 404.1468. Found: 404.1468.

Compound 9α—Compound **8** (917 mg, 1.366 mmol) and Fmoc-Thr-OAllyl (624.7 mg, 1.639 mmol) were mixed and dried under high vacuum overnight prior to the reaction. The reaction mixture was dissolved in dry CH_2Cl_2 under N_2. The fresh flame-dried molecular sieves (AW 300) (3 g) was added to the reaction mixture and stirred for about 2 h, afterwards, the reaction was cooled to −20 °C, followed by adding *N*-iodosuccinimide (338.1 mg, 1.503 mmol). To this mixture, freshly prepared TfOH (0.05 equiv) was slowly added at −20 °C. The reaction was stirred at −15 °C for 1 h. Once the reaction was done, it was quenched by adding saturated $Na_2S_2O_{3\ (aq)}$ and saturated $NaHCO_{3\ (aq)}$. The molecular sieves was filtered off through Celite and washed with CH_2Cl_2. The filtrate was extracted with CH_2Cl_2 and washed with saturated $Na_2S_2O_{3\ (aq)}$ and saturated $NaHCO_{3\ (aq)}$. After drying over Na_2SO_4, the filtrate was evaporated to dryness under vacuum. The concentrated residue was purified by flash column chromatography (25–50% EtOAc in hexane) to give α product (506.7 mg, 0.546 mmol, 40%). ^{1}H-NMR ($CDCl_3$, 600 MHz): δ 7.76 (d, 2H, *J* = 7.9 Hz), 7.64 (d, 1H, *J* = 7.4 Hz), 7.63 (d, 1H, *J* = 7.4 Hz), 7.39 (d, 1H, *J* = 7.4 Hz), 7.38 (d, 1H, *J* = 7.4 Hz), 7.33 (d, 2H, *J* = 8.3 Hz), 7.31 (d, 1H, *J* = 7.4 Hz), 7.30 (d, 1H, *J* = 7.4 Hz), 7.20 (d, 2H, *J* = 8.3 Hz), 7.17 (d, 2H, *J* = 8.3 Hz), 6.90 (d, 2H, *J* = 8.3 Hz), 6.86 (d, 2H, *J* = 8.3 Hz), 6.81 (d, 2H, *J* = 8.3 Hz), 5.89–5.96 (m, 1H), 5.74 (d, 1H, *J* = 9.2 Hz), 5.35 (d, 1H, *J* = 17.1 Hz), 5.25 (d, 1H, *J* = 10.1 Hz), 4.90 (d, 1H, *J* = 3.5 Hz), 4.78 (d, 1H, *J* = 10.9 Hz), 4.66–4.68 (m, 2H), 4.61 (d, 1H, *J* = 11.0 Hz), 4.39–4.49 (m, 5H), 4.31–4.36 (m, 2H), 4.28 (dd, 1H, *J* = 7.0, 7.0 Hz), 4.01 (brs, 1H), 3.95 (dd, 1H, *J* = 6.6, 6.6 Hz), 3.88 (dd, 1H, *J* = 10.5, 2.2 Hz), 3.79 (s, 3H), 3.78 (s, 3H), 3.77 (s, 3H), 3.53 (dd, 1H, *J* = 9.2, 7.5 Hz), 3.48 (dd, 1H, *J* = 8.8, 5.7 Hz), 1.30 (d, 3H, *J* = 5.5 Hz); ^{13}C-NMR ($CDCl_3$, 150 MHz): δ169.9, 163.1, 159.4, 159.3, 159.2, 156.7, 143.9, 143.6, 141.2, 131.3, 130.3, 129.8, 129.7, 129.5, 129.4, 129.3, 127.6, 127.1, 127.0, 125.2, 125.1, 119.9, 119.3, 113.8, 113.7, 113.6, 99.5, 76.0, 74.7, 73.4, 73.1, 71.8, 69.9, 68.5, 67.3, 66.5, 59.5, 58.8, 55.1, 47.0, 18.7; HRMS (ESI-TOF) calcd for $C_{52}H_{56}N_4O_{12}$ $[M+H]^+$: 929.3967. Found: 929.3954.

Compound 10—Compound **9α** (200 mg, 0.215 mmol) was dissolved in pyridine (1.2 mL) at 0 °C, followed by adding AcSH (1.2 mL). After stirring for 1 min, the reaction temperature was raised to room temperature and the reaction mixture was continued to stir for 4 h. Once the reaction was done, the mixture was evaporated to dryness under vacuum. The concentrated residue was purified by flash column chromatography (20%-50% EtOAc in hexane then 30% hexane in EtOAc) to give pure product for the next allyl removal step. (174.7 mg, 0.185 mmol,

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

86%). ^{1}H-NMR ($CDCl_3$, 600 MHz): δ 7.77 (d, 2H, *J* = 7.4 Hz), 7.63 (d, 1H, *J* = 6.5 Hz), 7.61 (d, 1H, *J* = 6.1 Hz), 7.40 (d, 1H, *J* = 7.9 Hz), 7.38 (d, 1H, *J* = 7.9 Hz), 7.33 (d, 1H, *J* = 7.9 Hz), 7.30 (d, 1H, *J* = 7.4 Hz), 7.20–7.25 (m, 8H), 6.80–6.87 (m, 4H), 5.81 (m, 1H), 5.23–5.32 (m, 4H), 4.88 (d, 1H, *J* = 11.0 Hz), 4.80 (d, 1H, *J* = 3.5 Hz), 4.55–4.64 (m, 4H), 4.43–4.53 (m, 4H), 4.33–4.39 (m, 2H), 4.27 (dd, 1H, *J* = 6.6, 6.6 Hz), 4.16 (m, 1H), 3.94 (brs, 1H), 3.87 (dd, 1H, *J* = 6.6, 6.6 Hz), 3.79 (s, 3H), 3.77 (s, 3H), 3.73 (s, 3H), 3.47–3.55 (m, 2H), 1.94 (s, 3H), 1.22 (d, 3H, *J* = 6.5 Hz); ^{13}C-NMR ($CDCl_3$, 150 MHz): δ 170.5, 169.8, 162.9, 159.2, 159.1, 158.9, 156.2, 143.7, 143.4, 141.2, 141.1, 130.8, 130.5, 129.9, 129.8, 129.7, 129.6, 129.2, 129.1, 127.6, 127.0, 124.8, 124.7, 119.9, 119.8, 119.6, 113.7, 113.6, 113.4, 76.4, 76.3, 73.8, 73.1, 72.0, 71.0, 70.2, 68.5, 66.7, 66.0, 58.5, 55.1, 48.8, 47.1, 23.3, 18.3; HRMS (ESI-TOF) calcd for $C_{54}H_{60}N_2O_{13}$ $[M+Na]^+$: 967.3987. Found: 967.3984. In the allyl removal step, the product from previous step (720 mg, 0.762 mmol) was dissolved in THF (10 mL), followed by adding $Pd(Ph_3P)_4$ (88.1 mg, 0.076 mmol) and *N*-methylaniline (827 μL, 7.62 mmol) subsequently. The reaction mixture was stirred at room temperature for 45 min. Once the reaction was done, the reaction solution was evaporated to dryness under vacuum. The concentrated residue was then purified by flash column chromatography (50% EtOAc in hexane, then 10% MeOH in CH_2Cl_2) to give the pure compound 10 (655 mg, 0.724 mmol, 95%) ^{1}H-NMR (d^6-DMSO, 600 MHz): δ 7. 87 (d, 2H, *J* = 7.5 Hz), 7.71 (d, 2H, *J* = 7.0 Hz), 7.39 (d, 1H, *J* = 7.0 Hz), 7.37 (d, 1H, *J* = 7.0 Hz), 7.29 (t, 2H, *J* = 7.0 Hz), 7.20–7.25 (m, 4H), 7.14 (d, 2H, *J* = 8.3 Hz), 6.87–6.89 (m, 4H), 6.83 (d, 2H, *J* = 8.3 Hz), 4.64 (dd, 2H, *J* = 19.7, 11.0 Hz), 4.33–4.46 (m, 6H), 4.17–4.27 (m, 3H), 3.92–3.96 (m, 3H), 3.71 (s, 9H), 3.66–3.70 (m, 2H), 3.42–3.50 (m, 2H), 1.90 (s, 3H), 1.09 (d, 3H, *J* = 6.1 Hz); ^{13}C-NMR (d^6-DMSO, 125 MHz): δ 172.8, 170.1, 158.7, 158.6, 158.6, 156.2, 143.9, 143.8, 140.8, 130.9, 130.9, 130.1, 129.3, 129.2, 128.8, 127.6, 127.0, 125.2, 125.2, 124.9, 120.1, 113.6, 113.5, 113.4, 98.7, 79.3, 79.0, 78.8, 77.2, 75.2, 73.5, 72.0, 71.0, 69.4, 68.8, 65.4, 59.3, 55.0, 54.9, 48.7, 46.8, 23.1, 18.4; HRMS (ESI-TOF) calcd for $C_{51}H_{56}N_2O_{13}$ $[M+Na]^+$: 927.3674. Found: 927.3680.

Solid-phase synthesis of glycopeptide Val^{53}-Phe^{82}

The synthesis of the C-terminal glycopeptide segment (53–82) was started with H_2N-Phe-2-ClTrt-resin and carried out by using Fmoc-based strategy. For coupling the glycosylated amino acid onto the resin, we used 2 equivalents of building block **4a** and left the reaction to shake for 24 h. Once the SPPS was completed, the acetyl groups on the sugar were removed on the solid support by treating with hydrazine/MeOH (1/6) for 6 h. Following this O-acetate removal step, the resin was washed with MeOH, DMF and CH_2Cl_2. The resulting glycopeptide was deprotected and cleaved from the resin by treating with TFA/H_2O/Et_3SiH/thioanisole (17/1/1/1) for 50 min at room temperature. The crude glycopeptide solution was evaporated to remove all the cleavage cocktail solution, and then the dried residue was re-dissolved in H_2O/CH_3CN (1/1) for HPLC purification. The purified glycopeptide was lyophilized to give pure glycopeptide. (18% yield). ESI-MS: 3528 Da. (the peptide mass was reconstructed from the experimental mass-to-charge (m/z) ratios from all of the observed protonation states of the peptide)

Synthesis of thioester peptide Cys(Acm)^{37}-Gly^{52}

The synthesis of the thioester peptide was begun with the preloading of 3-(tritylthio) propanoic acid onto MBHA resin LL. The detailed preloading procedure can be referred to the supporting material in *J. Am. Chem. Soc.* **2006**, *128*, 5626–5627.

Solid-phase peptide synthesis was carried out by using Boc strategy. After the full sequence was completed on the solid support, we removed the N-terminal Boc group to minimize side reactions during HF cleavage. The resin was then dried under high vacuum for 4 h prior to the final cleavage by HF with 10% (v/v) of anisole. The crude thioester peptide was purified by HPLC to give the pure product in 43% yield. ESI-MS: 1982 Da. (the peptide mass was

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

reconstructed from the experimental mass-to-charge (m/z) ratios from all of the observed protonation states of the peptide)

Synthesis of glycopeptide thioester Asp^{1}-Asn^{36}

The synthesis was based on the procedure reported by Wang et al. in *Int. J. Pept. Res. Therap*. **2005**, *11*, 117–123. The synthesis started with Rink amide resin. After removing Fmoc group by treatment with 20% piperidine (in DMF) for 20 min (× 2), the resin was loaded with Fmoc-Asp-OAllyl (4 equiv) by mixing together with the coupling solution containing HBTU (4 equiv) and DIEA (10 equiv) in DMF for 2 h. Afterwards, the resin was washed with DMF (4mL × 4), followed by acetylation of free amines on the resin with Ac_2O/DIEA/DMF (1:2:17) for 20 min. Afterwards, the resin was washed with DMF (4mL × 4), DCM (4mL × 4), and DMF (4mL × 4). Fmoc-SPPS was performed using HBTU/DIEA coupling condition. To couple the glycosylated amino acid, 2 equivalents of building block **16** was used in the synthesis, and the coupling time was elongated to 1 day. Once the full sequence of the glycopeptide was completed on the resin, the allyl group was removed by treating with Pd $(PPh_3)_4$ (25 mg/0.1 mmol resin) and $PhSiH_3$ (10 equiv) in DCM for 30 min (× 2). After washing with DCM (4mL × 4), DMF (4m × 4), and then DCM (4mL × 4), we transformed the free acid to thioester by mixing the resin with the reaction solution containing ethyl 3-mercaptopropionate (24 equiv), DIEA (37.5 equiv), anhydrous HOBt (30 equiv), DIC (30 equiv) and DCM/DMF (1/4) for 6 h (× 2). Afterwards, the resin was washed with DCM (4mL × 4), DMF (4m × 4), and then DCM (4mL × 4). The glycopeptide was cleaved from the resin with concomitant full deprotection by treatment with TFA/H_2O/thioanisole/TES (17/1/1/1). The crude product was subjected to HPLC purification to give the pure product in 9% yield. ESI-MS: 3977 Da. (the peptide mass was reconstructed from the experimental mass-to-charge (m/z) ratios from all of the observed protonation states of the peptide)

General procedure for SAL

The ligation of unprotected peptide segments was performed as follows: Prior to the reaction, the ligation solution (6 M Gn·HCl, 0.2 M phosphate buffer, pH 8.5) was degassed for 10 min. Subsequently, glycopeptide and thioester peptide were dissolved in the ligation solution under Ar. The reaction was performed at 37 °C and was vortexed periodically to equilibrate the reaction mixture. Before HPLC or LC/MS analysis, TCEP (60 mM) or 10% (v/v) of 2-mercaptoethanol was added to reduce any formed disulfide bonds.

General procedure for desulfurization

Desulfurization was performed at room temperature in 6 M Gn·HCl, 0.2 M phosphate buffer, pH 5.8 containing 15 mM TCEP. The buffer was degassed by bubbling Ar through the solution for 10 min prior to use. Following the addition of Pd/Al_2O_3 (15 times of the weight of glycopeptide), the reaction mixture was kept under H_2 using H_2 balloon. The desulfurization reaction was monitored by analytical HPLC chromatography and LC/MS. Once the reaction was complete, Pd/Al_2O_3 was spun down by centrifuge, and the supernatant was collected for HPLC purification.

Removing Acm group from glycopeptide segment $Cys(Acm)^{37}$-Phe^{82}

$Cys(Acm)^{37}$-Phe^{82} (3 mg) was dissolved in 1 mL of 10% AcOH (pH 4.0) containing 30 equivalents of $Hg(OAc)_2$. The reaction mixture was mixed well and left to sit at room temperature for 1 h. Once the reaction was done, 120 equivalents of DTT were added and the mixture was allowed to react for 12 h to precipitate all $Hg(OAc)_2$. The black precipitate was spun down afterwards, and the supernatant was collected for HPLC purification.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NCL of glycopeptide Cys^{37}-Phe^{82} and glycopeptide thioester Asp^{1}-Asn^{36}

The ligation of glycopeptide segment Cys^{37}-Phe^{82} (2.9 mM) and glycopeptide thioester segment Asp^{1}-Asn^{36} (4.4 mM) was carried out in a solution of 6M Gn·HCl, 200 mM phosphate buffer, pH 7.9 containing 2% (v/v) of thiophenol and 2% (v/v) benzylmercaptan at 37 °C. After 16 h, we observed that the ligation reaction was complete and the ligation product was confirmed by ESI-MS. After HPLC purification, the pure product was obtained in 47%.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the NIH and Skaggs Institute for Chemical Biology. S. F. is grateful for a postdoctoral fellowship provided by the Deutsche Akademische Austauschdienst (DAAD). We would like to thank Dr. Asad Chavochi-Negad and Prof. M. Reza Ghadiri for assistance in HF operation. We would also like to give our appreciation to Dr. William Greenberg for editing the manuscript.

References

1. Apweiler R, Hermjakob H, Sharon N. Biochim Biophys Acta, Gen Subj 1999;1473:4–8.
2. (a) Dwek RA, Butters TD, Platt FM, Zitzmann N. Nature Rev Drug Discov 2002;1:65–75. [PubMed: 12119611] (b) Spiro RG. Glycobiology 2002;12:43–56. (c) Graffenried CLD, Bertozzi CR. Curr Opin Cell Biol 2004;16:356–363. [PubMed: 15261667] (d) Sinclair AM, Elliott S. J Pharm Sci 2005;94:1626–1635. [PubMed: 15959882]
3. (a) Sears P. Science 2001;291:2344–2350. [PubMed: 11269314] (b) Xu R, Hanson SR, Zhang Z, Yang YY, Schultz PG, Wong CH. J Am Chem Soc 2004;126:15654–15655. [PubMed: 15571382] (c) Hamilton SR, et al. Science 2006;313:1441–1443. [PubMed: 16960007] (d) Li H, et al. Nature Biotechnol 2006;24:210–215. [PubMed: 16429149]
4. (a) Davis BG. Chem Rev 2002;102:579–602. [PubMed: 11841255] (b) Guo Z, Shao N. Med Res Rev 2005;25:655–678. [PubMed: 15895471] (c) Pratt MR, Bertozzi CR. Chem Soc Rev 2005;34:58–68. [PubMed: 15643490] (d) Brik A, Ficht S, Wong C-H. Curr Opin Chem Biol 2006;10:638–644. [PubMed: 17046315]
5. (a) Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR. J Am Chem Soc 1999;21:11684–11689. (b) Koeller KM, Smith MEB, Wong CH. J Am Chem Soc 2000;122:742–743. (c) Mcmillan D, Bertozzi CR. Angew Chem Int Ed 2004;43:1355–1359. (d) Hackenberger CPR, Friel CT, Radford SE, Imperiali B. J Am Chem Soc 2005;127:12882–12889. [PubMed: 16159282] (e) Li B, Song H, Hauser S, Wang LX. Org Lett 2006;8:3081–3084. [PubMed: 16805557]
6. (a) Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. Science 1994;266:776–779. [PubMed: 7973629] (b) Dawson PE, Kent SBH. Annu Rev Biochem 2000;69:923–960. [PubMed: 10966479] (c) Marcaurelle LA, Mizoue LS, Wilken J, Oldhm L, Kent SBH, Hndel TM, Bertozzi CR. Chem Eur J 2001;7:1129–1132.
7. (a) Muir TW, Sondhi D, Cole PA. Proc Nat Acad Sci USA 1998;95:6705–6710. [PubMed: 9618476] (b) Muralidharan V, Muir TW. Nature Methods 2006;3:429–438. [PubMed: 16721376] (c) Tolbert TJ, Franke D, Wong C-H. Bioorg Med Chem 2005;13:909–915. [PubMed: 15653356]
8. (a) Beligere GS, Dawson PE. J Am Chem Soc 1999;121:6332–6333. (b) Yan LZ, Dawson PE. J Am Chem Soc 2001;123:526–533. [PubMed: 11456564] (c) Bang D, Makhatadze VT, Kossiakoff AA, Kent SB. Angew Chem Int Ed 2005;44:3852–3856. (d) Macmillan D. Angew Chem Int Ed 2006;45:7668–7672.
9. (a) Cane LE, Bark SJ, Kent SBH. J Am Chem Soc 1996;118:5891–5896. (b) Low DW, Hill MG, Carrasco MR, Kent SBH, Botti P. Proc Natl Acad Sci USA 2001;98:6554–6559. [PubMed: 11390992] (c) Offer J, Boddy C, Dawson PE. J Am Chem Soc 2002;124:4642–4646. [PubMed: 11971712] (d) Macmillan D, Anderson DW. Org Lett 2004;6:4659–4662. [PubMed: 15575654] (e) Wu B, Chen J, Warren JD, Chen G, Hua Z, Danishefsky SJ. Angew Chem Int Ed 2006;45:4116–4125. (f) Chen J,

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Chen G, Wu B, Wan Q, Tan Z, Hua Z, Danishefsky SJ. Tetrahedron Lett 2006;47:8013–8016. [PubMed: 17426819]

10. (a) Brik A, Yang Y-Y, Ficht S, Wong C-H. J Am Chem Soc 2006;128:5626–5627. [PubMed: 16637616] (b) Brik A, Ficht S, Yang Y-Y, Bennett CS, Wong C-H. J Am Chem Soc 2006;128:15026–15033. [PubMed: 17105315]
11. Molecular simulation of transition states have been performed, and will be presented in a following publication. The simulations support the hypothesis that the restricted conformation of the sugar helps to pre-organize the GalNAc-linked thioester intermediate and N-terminal amine for nucleophilic attack.
12. Bulet P, Hegy G, Lambert J, Dorsselaer AV, Hoffmann JA, Hetru C. Biochemistry 1995;34:7394–7400. [PubMed: 7779781]
13. (a) Winans KA, King DS, Rao VR, Bertozzi CR. Biochemistry 1999;38:11799–11710. (b) Cudic M, Bulet P, Hoffmann R, Craik DJ, Otvos L Jr. Eur J Biochem 1999;266:549–558. [PubMed: 10561597]
14. (a) Hanisch FG. Biol Chem 2001;382:143–149. [PubMed: 11308013] (b) Hang HC, Bertozzi CR. Bioorg Med Chem 2005;13:5021–5034. [PubMed: 16005634]
15. Koeller KM, Smith MEB, Wong C-H. Bioorg Med Chem 2000;8:1017–1026. [PubMed: 10882013]
16. Schultz M, Kunz H. Tetrahedron Lett 1992;33:5319–5322.
17. Kupihar Z, Schmel Z, Kele Z, Penke B, Kovacs L. Bioorg Med Chem 2001;9:1241–1247. [PubMed: 11377182]
18. Albericio F, Carpino LA. Methods Enzymol 1997;289:104–126. [PubMed: 9353720]
19. Ciommer M, Kunz H. Synlett 1991;8:593–595.
20. Wellings DA, Atherton E. Methods Enzymol 1997;289:44–67. [PubMed: 9353717]
21. (a) Alewood P, Alewood D, Miranda L, Love S, Meutermans W, Wilson D. Methods Enzymol 1997;289:14–29. [PubMed: 9353715] (b) Stewart JM. Methods Enzymol 1997;289:29–44. [PubMed: 9353716]
22. For instance: $t_{1/2}$(h) for the Gly-Gly junction for which the ligation efficiency is less susceptible to varying the reaction conditions: 5h for β-O-linked glycopeptides, 4 h for both N-linked and α-O-linked glycopeptides. On the other hand, $t_{1/2}$(h) for the Gly-Val junction for which the ligation efficiency is more susceptible to varying the reaction conditions: 18 h for N-linked glycopeptides (6 M Gn·HCl, 0.1 M PBS, 2% PhSH, glycopeptide: 1.2 equiv., peptide thioester: 1 equiv.); 14 h for α-O-linked glycopeptides (6 M Gn·HCl, 0.1 M PBS, glycopeptide: 1 equiv., peptide thioester: 1.2 equiv.).
23. The initial pH of the buffer used for SAL is 8.5. After adding the peptides, the pH drops due to the residual TFA present in HPLC purified peptides. As a consequence, in some cases, triethanolamine was added to adjust the reaction pH to the ideal range 7.2~7.4.
24. The glycopeptide was consumed by two reaction paths: reacting with the peptide thioester Cys(Acm)^{37}-Gly^{52}, and by generating a degraded fragment. The mass of this degraded fragment (observed mass = 3901 Da) corresponds to a loss of the first two N-terminal residues of glycopeptide Val^{53}-Phe^{82}. The same fragment mass was also observed in the cases of glycopeptide His^{53}-Phe^{82} and glycopeptide Gly^{53}-Phe^{82} by LC/MS. The first observation of this degraded fragment was in the crude glycopeptide cleaved from the resin. In the beginning of the ligation, the amount of this degraded fragment is less than 5%. However, its amount increased during the course of the reaction. A similar degradation phenomenon was not observed for the short glycopeptides (Table 1).
25. (a) Warren JD, Miller JS, Keding SJ, Danishefsky SJ. J Am Chem Soc 2004;126:6576–6578. [PubMed: 15161285] (b) Mezzato S, Schaffrath M, Unverzagt C. Angew Chem Int Ed 2005;44:1650–1654.
26. It caught our attention that Asp or Glu on the C-terminus of the peptide thioester has the potential to form an intra-molecular anhydride during NCL, which could lead to the generation of a branched ligation product. To avoid this problem, the original Asp^{36} of diptericin was mutated to Asn^{36}. Alternatively, an orthogonal protecting group could be used to avoid this side reaction (Villain M, Gaertner H, Botti P. Eur J Org Chem 2003:3267–3272.).
27. Wang P, Miranda LP. Int J Pept Res Therap 2005;11:117–123.
28. Ye X-S, Wong C-H. J Org Chem 2000;65:2410–2431. [PubMed: 10789453]
29. Takahashi S, Kuzuhara H, Nakajima M. Tetrahedron 2001;57:6915–6926.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

30. Since Asp-Gly sequence is prone to aspartimide formation during peptide synthesis, we mutated Asp^{29} to Asn residue. On the other hand, our earlier efforts toward solving this problem by using the commercially available Fmoc-Asp(O^tBu)-(Hmb)Gly-OH in SPPS failed due to the difficulty of complete removal of the Hmb group in the final global deprotection step.

31. Ingale S, Buskas T, Boons G-J. Org Lett 2006;8:5785–5788. [PubMed: 17134272]

32. While we were preparing this manuscript, Kent's group reported selective desulfurization of cysteine in peptides containing Cys(Acm) residues (Pentelute BL, Kent SBH. Org Lett 2007;9:687–690. [PubMed: 17286375]). In their report, Raney nickel was used as the catalyst. In our studies, we found that the thiol handle of the sugar moiety reacts much more rapidly towards desulfurization than the cysteine residue (data not shown), when Pd/Al_2O_3 was used as the catalyst. This indicates that even better desulfurization selectivity might be achieved if different catalysts, like Raney nickel, or other conditions were used to selectively desulfurize the thiol handle in the presence of Cys(Acm) residues.

33. Hsieh-Wilson LC. Nature 2007;445:31–33. [PubMed: 17203048]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

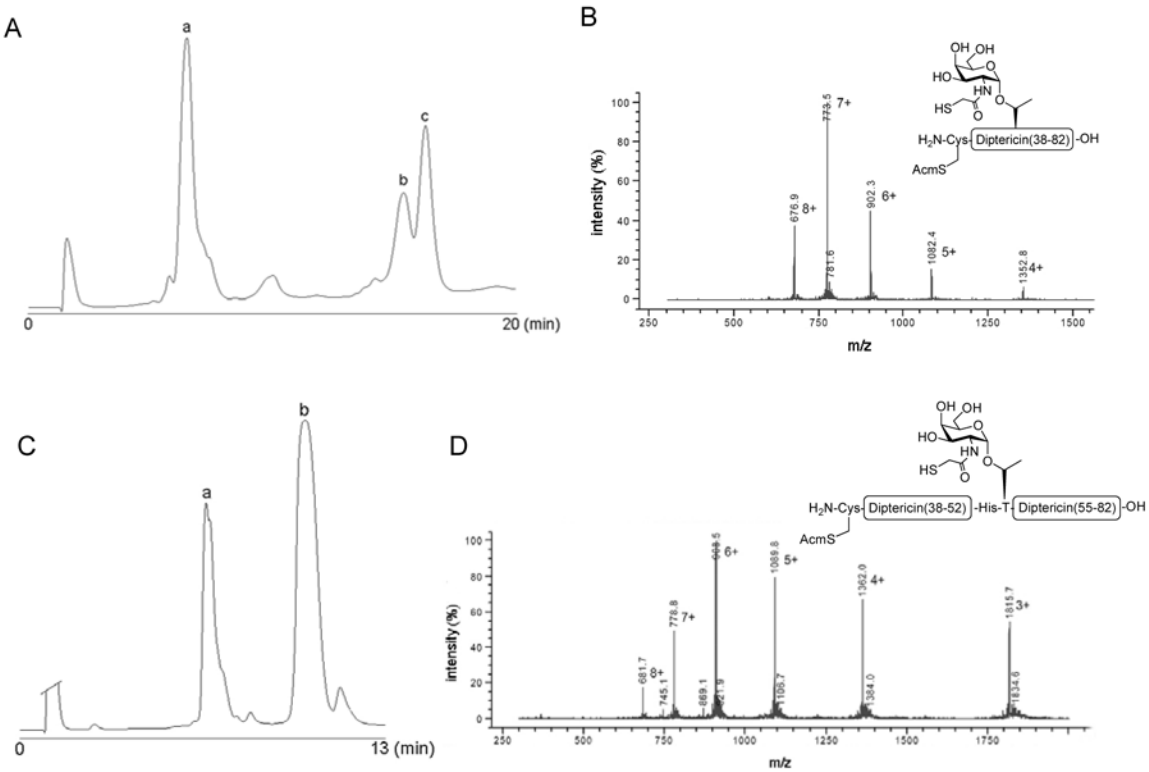


Figure 1.

(A) Analytical HPLC chromatograph of SAL of glycopeptide Val^{53}-Phe^{82} and peptide thioester Cys(Acm)^{37}-Gly^{52} at 48 h. Peak a, hydrolyzed peptide thioester; Peak b, the remaining glycopeptide Val^{53}-Phe^{82} and its degraded fragment (the ratio is around 2:1);[24] Peak c, the ligated product. HPLC gradient: 5–30% of CH_3CN (+ 0.1 % TFA) over 30 min at a flow rate of 1.5 ml/min (B) ESI-MS spectrum of the glycopeptide Cys(Acm)^{37}-Phe^{82} (the reconstructed mass was 5408, the calculated average mass: 5409 Da). (C) Analytical HPLC chromatograph of SAL of glycopeptide His^{53}-Phe^{82} and peptide thioester Cys(Acm)^{37}-Gly^{52} at the time point of 24 h. Peak a, the hydrolyzed thioester; peak b, the ligated product and the degraded fragment from glycopeptide His^{53}-Phe^{82} (the ratio is around 9:1)[24]. HPLC gradient: 5–50% of CH_3CN (+ 0.1 % TFA) over 30 min at a flow rate of 1.5 ml/min (D) ESI-MS spectrum of the ligated product of glycopeptide His^{53}-Phe^{82} and peptide thioester Cys (Acm)^{37}-Gly^{52} (the reconstructed mass was 5446 Da, the calculated average mass: 5447 Da).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

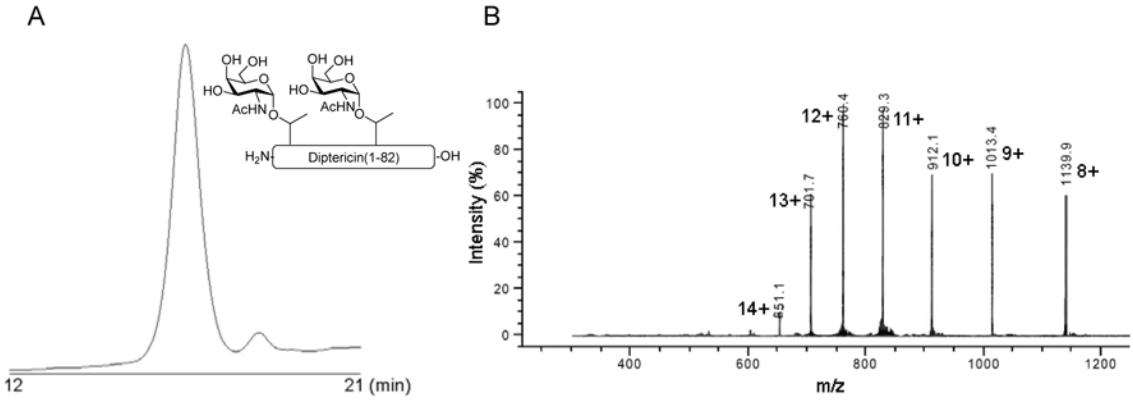


Figure 2.

(A) HPLC analysis of the crude product from the desulfurization step, which was eluted with a gradient of 5–30% of CH_3CN (+ 0.1 % TFA) over 30 min at a flow rate of 1.5 ml/min. (B) ESI-MS spectrum of the pure product of the desulfurization step (the reconstructed mass is 9113 Da)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

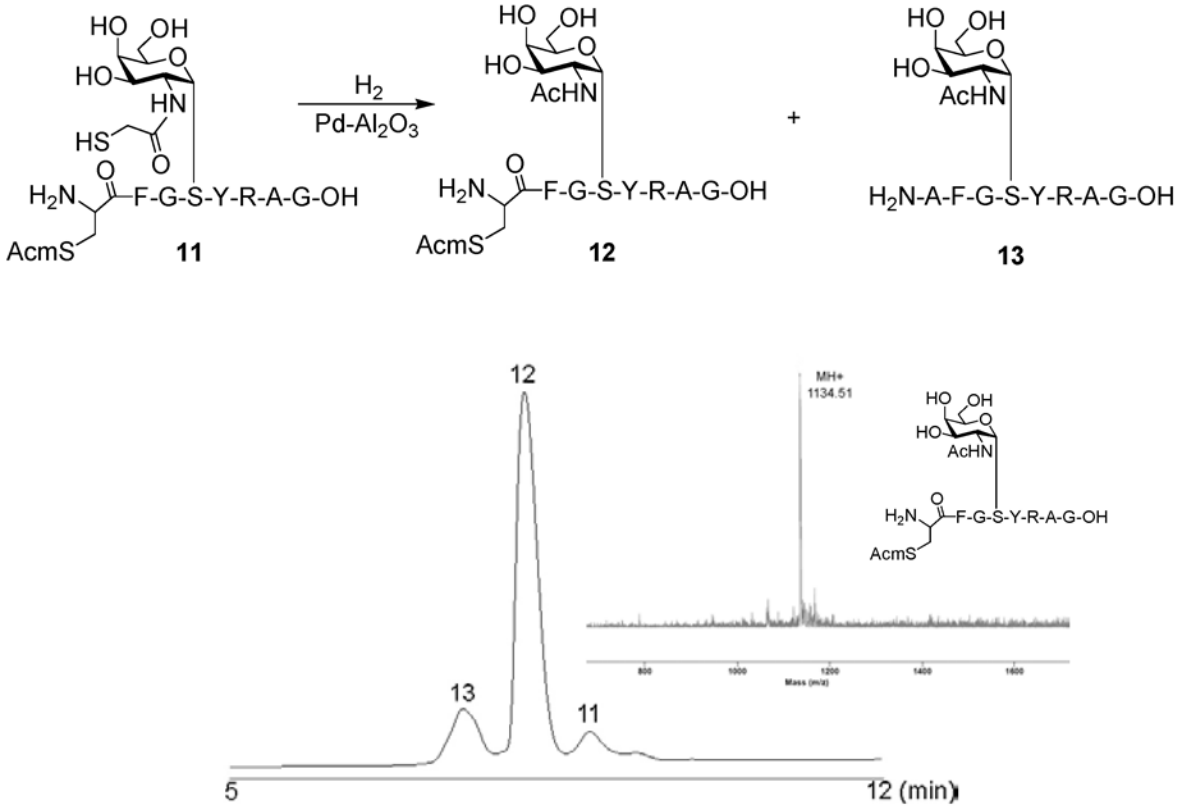


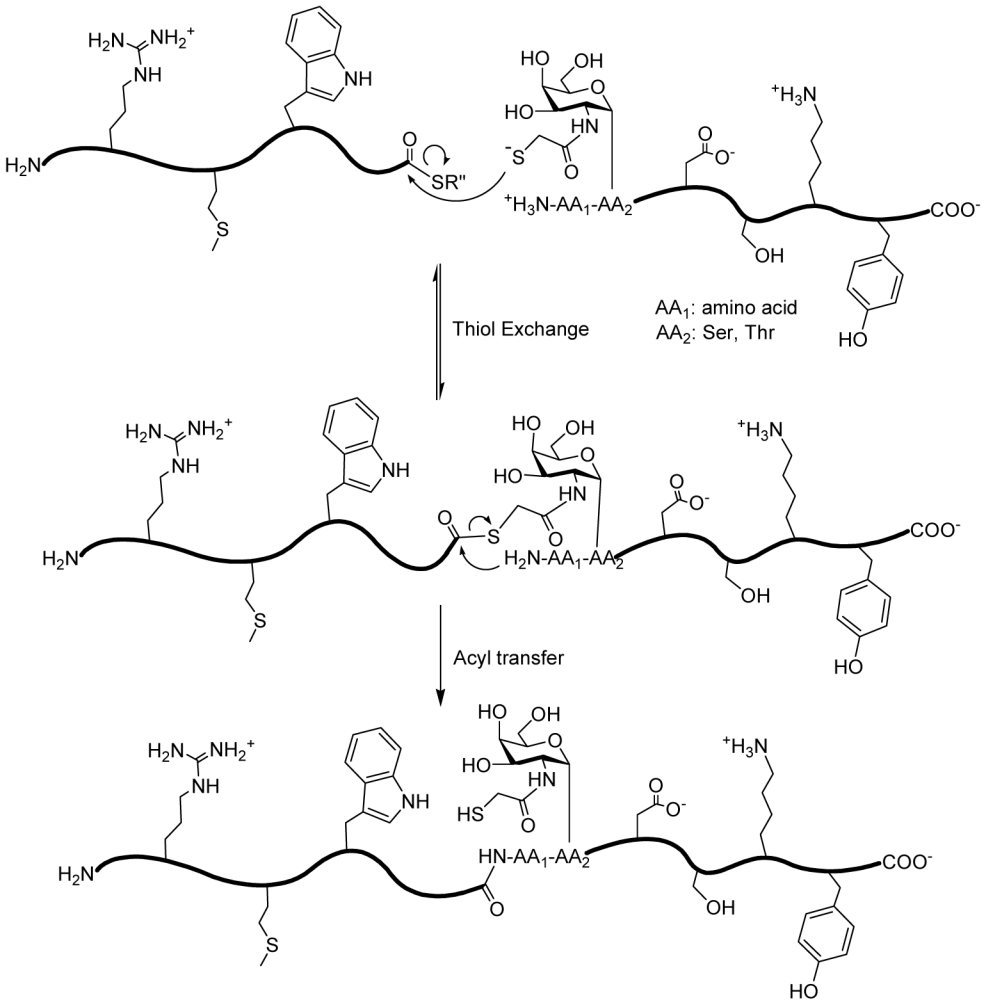
Figure 3.

HPLC analysis of the selective desulfurization of glycopeptide **11** and MALDI-TOF/MS spectrum of glycopeptide **12** (desired product).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

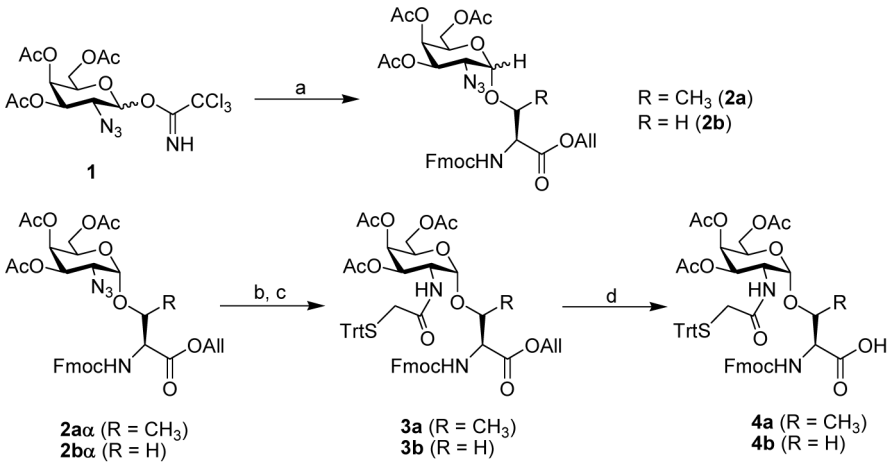


Scheme 1.
Proposed Mechanism of the Sugar-Assisted Ligation for α-linked Glycopeptide Synthesis

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



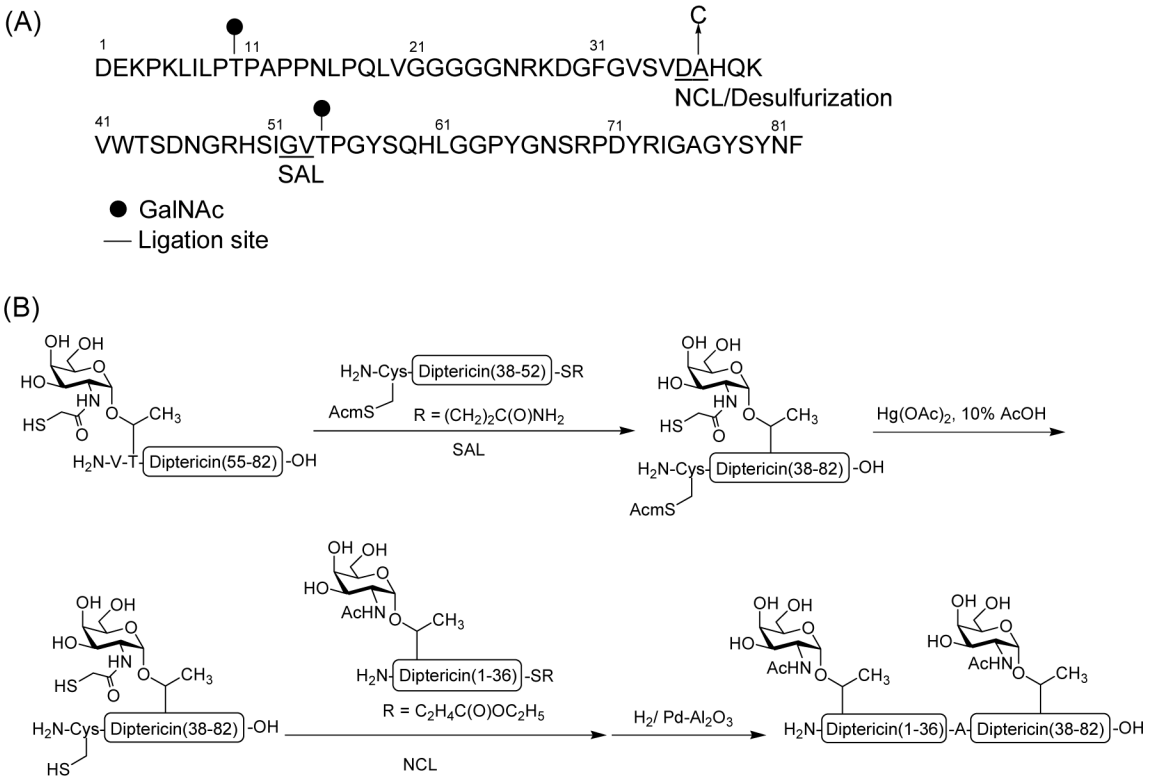
Scheme 2. Synthesis of the Building Blocks 4a and 4b

Reagents and Conditions: (a) Fmoc-Thr-OAll (R = CH_3 for **2a**) or Fmoc-Ser-OAll (R = H for **2b**), TMSOTf, molecular sieves AW 300, CH_2Cl_2, −70 °C, α/β = 4/3 (**2aα**: 45% yield, **2bα**: 45% yield); (b) Zn/AcOH; (c) TrtS-CH_2COOH, HBTU, DIEA, DMF, 84% (2 steps); (d) Pd $(PPh_3)_4$, NMA, THF, 95%. (Trt = trityl, TMSOTf = trimethylsilyl trifluoromethanesulfonate)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



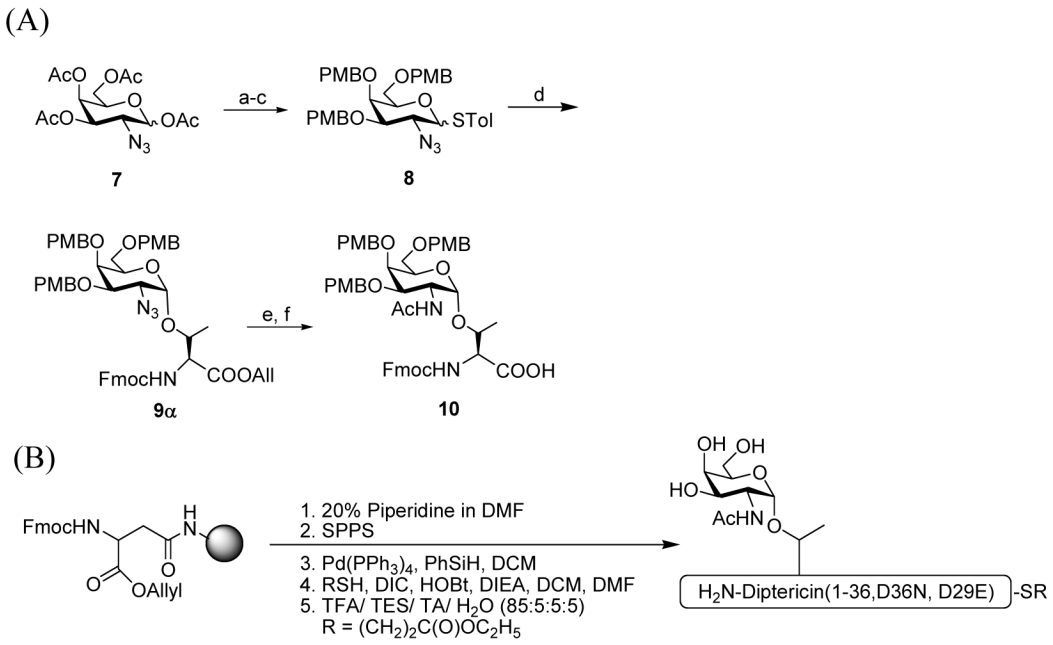
Scheme 3.

(A) The Sequence of Diptericin and Ligation Sites. (B) The Synthetic Strategy of Diptericin

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



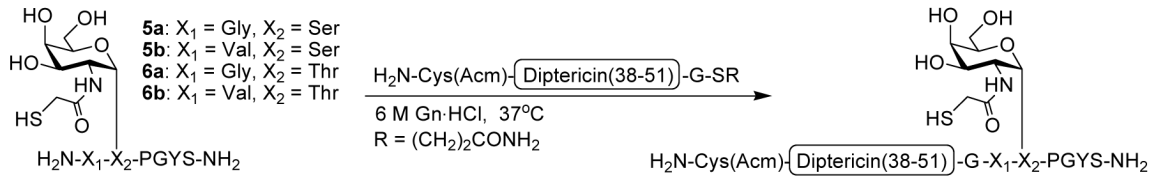
Scheme 4. (A) Synthesis of the Building Block 10. (B) Synthesis of Glycopeptide Thioester Asp^{1}-Asn^{36} Using the Side-Chain Anchoring Strategy

Reagents and Conditions: (a) $BF_3{\cdot}OEt_2$, TolSH, CH_2Cl_2, 0 °C then rt, 85%; (b) NaOMe, MeOH, pH~11; (c) PMBCl, NaH, DMF, 0 °C then rt., 68% (3 steps); (d) Fmoc-Thr-OAll, NIS, cat. TfOH, molecular sieves AW 300, CH_2Cl_2, −15 °C, 83%, α/β = 3/2; (e) AcSH, pyridine, 86%; (f) $Pd(PPh_3)_4$, NMA, THF, 95%. TolSH = *p*-thiocresol, NIS = N-iodosuccinimide

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Sugar-Assisted Ligation of α-Linked Glycopeptides



Entry	-X_1-	-X_2-	Ligation Junction	$t_{1/2}$(h)[a]	Thioester (equiv.)[c]	Reaction Medium	Product Mass (Da) obsd[b]	Product Mass (Da) calcd
1	Gly	Ser	Gly-Gly	4	1.2	0.2 M phosphate buffer	2677.4[f]	2677.4
2	Val	Ser	Gly-Val	11	1.2	0.2 M phosphate buffer	2720.4[g]	2719.5
3	Gly	Thr	Gly-Gly	4	1.2	0.2 M phosphate buffer	2691.5[h]	2691.4
4	Val	Thr	Gly-Val	11	1.2	0.2 M phosphate buffer[d]	2733.6[i]	2733.5
5	Val	Thr	Gly-Val	14	1.2	0.2 M phosphate buffer + 2% PhSH	2733.6[j]	2733.5
6	Val	Thr	Gly-Val	14	1.2	0.1 M phosphate buffer[e]	2733.6[k]	2733.5
7	Val	Thr	Gly-Val	9	1.5	0.2 M phosphate buffer[d]	2733.6[l]	2733.5
8	Val	Thr	Gly-Val	7	2.0	0.2 M phosphate buffer[d]	2733.6[m]	2733.5

[a] $t_{1/2}$ indicates the required reaction time for the reactant glycopeptide to reach 50% consumption.

[b] Characterized by LC/MS.

[c] The concentration of the reactant glycopeptide in each reaction is 6 mM.

[d] Reaction pH is around 7.2~7.4.

[e] Reaction pH is around 7.0.

[f] isolated yield = 78%.

[g] isolated yield = 56%.

[h] isolated yield = 73%.

[i] isolated yield = 59%.

[j] isolated yield = 45%.

[k] isolated yield = 47%.

[l] isolated yield = 63%.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[m] isolated yield = 63%.