J|A|C|S
ARTICLES
Published on Web 10/13/2007

Extended Sugar-Assisted Glycopeptide Ligations: Development, Scope, and Applications

Richard J. Payne,[†] Simon Ficht,[†] Sishi Tang,[‡] Ashraf Brik,[†,#] Yu-Ying Yang,[†] David A. Case,[‡] and Chi-Huey Wong*[,†,§]

Contribution from the Departments of Chemistry and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and The Genomics Research Center, Academia Sinica, 128 Section 2, Academia Road, Nankang, Taipei 115, Taiwan

Received May 21, 2007; E-mail: wong@scripps.edu

Abstract: Recently, we reported the development of sugar-assisted ligation (SAL), a novel peptide ligation method for the synthesis of glycopeptides. After screening a large number of glycoprotein sequences in a glycoprotein database, it became evident that a large proportion (approximately 53%) of O-glycosylation sites contain amino acid residues that will not undergo SAL reactions. To overcome these inherent limitations and broaden the scope of the method we report here the development of an extended SAL method. Glycopeptides containing up to six amino acid extensions N-terminal to the glycosylated residue were shown to facilitate ligation reactions with peptide thioesters, and these products were isolated in good yields. Kinetic analysis was used to show that as glycopeptides were extended by further amino acid residues, ligation reactions became slower. This finding was rationalized by molecular dynamics simulations using AMBER9. These studies suggested a general trend whereby the proximal distance between the reactive sites of the thioester intermediate (the N-terminal amine and the carbonyl carbon of the thioester) increased as glycopeptides were extended, thus slowing down the ligation rate. Each of the extended SAL methods showed broad tolerance to a number of different amino acid combinations at the ligation junction. Re-evaluation of the glycoprotein database suggested that 95% of the O-linked glycosylation sites can now be utilized to facilitate SAL or extended SAL reactions. As such, this method represents an extremely valuable tool for the synthesis of naturally occurring glycopeptides and glycoproteins. To demonstrate the applicability of the method, extended SAL was successfully implemented in the synthesis of the starting unit of the cancer-associated MUC1 glycoprotein.

Introduction

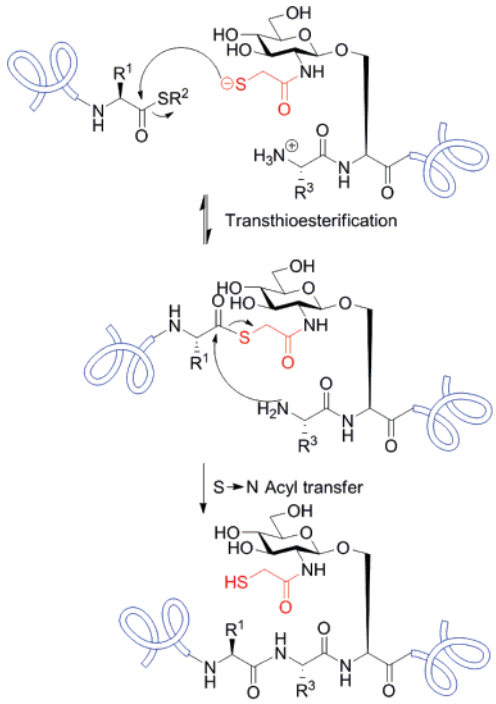
Protein glycosylation is a ubiquitous post-translational modification which introduces enormous structural diversity to proteins.[1] Estimations suggest that more than fifty percent of all human proteins are glycosylated.[2] The glycan modification is of paramount importance for a variety of biological recognition events such as cell adhesion, cell differentiation, and cell growth.[3,4] Aberrant glycosylation of proteins often modifies intracellular recognition and has been associated with a number of serious illnesses including autoimmune diseases, infectious diseases, and cancer.[5,6] To understand the role of glycosylation at a molecular level, it is imperative to have access to homogeneous glycopeptides and glycoproteins. In contrast to the protein part, the glycosylation pattern of a given glycoprotein is not under the control of a coding template, but is dictated by the relative activities of a number of glycosyltransferase enzymes. The resulting glycoproteins are usually produced as inseparable mixtures of glycoforms, and as such, biological expression systems cannot be used for the production of homogeneous glycoproteins. It is currently accepted that chemical and chemoenzymatic intervention can be used to solve the availability problem. Peptide ligation methods have emerged as powerful tools in this respect.[7−10]

One such method, native chemical ligation (NCL), has proven to be very useful for peptide chemistry, where it has been implemented in the synthesis of hundreds of proteins to date.[11] More recently, NCL has gained the spotlight as an effective method for the preparation of glycopeptides and glycoproteins.[11−17] The obvious limitation of this method is the

[†] Department of Chemistry, The Scripps Research Institute.
[‡] Department of Molecular Biology, The Scripps Research Institute.
[§] The Genomics Research Center.
[#] Current address: Department of Chemistry, Ben Gurion University, Beer Sheva 84105, Israel.

(1) Spiro, R. G. *Glycobiology* **2002**, *12*, 43−56.
(2) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4−8.
(3) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683−720.
(4) Varki, A. *Glycobiology* **1993**, *3*, 97−130.
(5) Miller, L. H.; Good, M. F.; Milon, G. *Science* **1994**, *264*, 1878−1883.
(6) Dube, D. H.; Bertozzi, C. R. *Nat. Rev. Drug Discovery* **2005**, *4*, 477−488.
(7) Davis, B. G. *Chem. Rev.* **2002**, *102*, 579−601.
(8) Pratt, M. R.; Bertozzi, C. R. *Chem. Soc. Rev.* **2005**, *34*, 58−68.
(9) Liu, L.; Bennett, C. S.; Wong, C. H. *Chem. Commun.* **2006**, 21−33.
(10) Brik, A.; Ficht, S.; Wong, C. H. *Curr. Opin. Chem. Biol.* **2006**, *10*, 638−644.
(11) Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69*, 923−960.
(12) Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H. *Justus Liebigs Ann. Chem.* **1953**, *583*, 129−149.

Scheme 1. Proposed Mechanism of Sugar-Assisted Ligation (SAL)



requirement for an N-terminal cysteine or alanine residue.[18] Unfortunately, many glycoprotein targets do not bear these amino acids at strategically useful positions in the protein sequence, and as such these methods cannot be implemented. This fact has led to the development of cysteine-free ligation techniques, which incorporate a thiol containing auxiliary at the N-terminus.[19,20] This method has proven useful for the synthesis of glycopeptides; however, the use of these auxiliaries appears to be limited to ligation sites containing amino acid side chains of low steric bulk.[21,22]

Our laboratory has recently reported the development of a sugar-assisted ligation (SAL) method for the synthesis of O- and N-linked glycopeptides.[23,24] This method utilizes a glycopeptide in which the carbohydrate (*N*-acetyl glucosamine) is derivatized with a mercaptoacetate auxiliary at the 2-position. In the presence of a peptide thioester, and under suitable ligation conditions, thioester exchange is followed by an S → N acyl transfer affording a ligated product with a native peptide backbone (Scheme 1). The reaction cascade showed high sequence tolerance at the ligation junction and was shown to be chemoselective with all naturally occurring amino acid side chains.[24] The thiol auxiliary could then be removed by desulfurization to afford the natural glycan motif. Initial concerns over the orthogonality of this reaction in the presence of other thiol containing amino acid residues was recently addressed by Kent and co-workers[18] and by our laboratory.[25] Both reports describe the use of a cysteinyl acetamidomethyl (Acm) protecting group strategy to facilitate selective desulfurization reactions. This method can be adopted for desulfurization of SAL products, as demonstrated recently in the total synthesis of the antibacterial glycoprotein diptericin.[25] This has opened the door for the synthesis of more complex glycopeptides and potentially glycoprotein targets by this technology.[26]

The continuing goal of our laboratory was to apply the existing SAL method to the total synthesis of a native glycoprotein. A detailed search of over two hundred O-linked glycoproteins was conducted using the database O-GlycBase v6.00.[27] This screen indicated that a large number of glycoproteins bear sterically hindered amino acid residues N-terminal to the sugar-carrying amino acid (i.e., R^1 and R^3 in Scheme 1). As a result, SAL cannot be used in these cases. To determine the extent of this finding, each putative ligation junction adjacent to naturally occurring O-glycosylation sites ($n = 535$, see Supporting Information) was scanned. This was conducted by assessing the nature of the required thioester (two amino acids N-terminal to the glycosylated residue, i.e., R^1 in Scheme 1), since it is established that the steric nature of the thioester component is the most influential factor in SAL (and NCL) ligation efficiency.[24] Amino acid thioesters known to be extremely difficult to ligate include proline, valine, isoleucine, leucine, and threonine.[28] Other thioesters cannot be used without side-chain protection. These include glutamic acid and aspartic acid due to the rapid formation of mixed cyclic anhydrides with the carboxylate side chains[29] and lysine due to the formation of cyclic lactams.[30] Additionally, in contrast to NCL, the use of cysteine thioesters could be problematic for SAL reactions because of cyclic thioester formation and promoted hydrolysis. Remarkably, 53% of all O-glycosylation sites studied contained such amino acids (indicated in red in Figure 1). A further 22% of glycosylation sites contained thioesters that would facilitate ligations, however, would do so at a sluggish rate (indicated in orange). Only 25% of the O-glycosylation sites studied contained thioesters that are known to undergo SAL at reasonable rates (indicated in green). It is evident from this informatics study that only a small proportion of the glycoproteome can be feasibly synthesized by the existing ligation methodology.

To circumvent this problem, we proposed a modification of the existing method whereby SAL could be conducted on glycopeptides that had been extended by the addition of extra amino acids N-terminal to the glycosylated residue. If successful in this endeavor, it would represent a flexible new method, providing access to a significantly greater number of biologically relevant glycopeptide and glycoprotein targets. The first step

(13) Dawson, P. E.; Muir, T. W.; Clarklewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779.
(14) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684−11689.
(15) Mezzato, S.; Schaffrath, M.; Unverzagt, C. *Angew. Chem., Int. Ed.* **2005**, *44*, 1650−1654.
(16) Marcaurelle, L. A.; Mizoue, L. S.; Wilken, J.; Oldham, L.; Kent, S. B. H.; Handel, T. M.; Bertozzi, C. R. *Chem. Eur. J.* **2001**, *7*, 1129−1132.
(17) Hackenberger, C. P. R.; Friel, C. T.; Radford, S. E.; Imperiali, B. *J. Am. Chem. Soc.* **2005**, *127*, 12882−12889.
(18) Pentelute, B. L.; Kent, S. B. H. *Org. Lett.* **2007**, *9*, 687−690.
(19) Offer, J.; Boddy, C. N. C.; Dawson, P. E. *J. Am. Chem. Soc.* **2002**, *124*, 4642−4646.
(20) Low, D. W.; Hill, M. G.; Carrasco, M. R.; Kent, S. B. H.; Botti, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6554−6559.
(21) Macmillan, D.; Anderson, D. W. *Org. Lett.* **2004**, *6*, 4659−4662.
(22) Wu, B.; Chen, J. H.; Warren, J. D.; Chen, G.; Hua, Z. H.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2006**, *45*, 4116−4125.
(23) Brik, A.; Yang, Y. Y.; Ficht, S.; Wong, C. H. *J. Am. Chem. Soc.* **2006**, *128*, 5626−5627.
(24) Brik, A.; Ficht, S.; Yang, Y. Y.; Bennett, C. S.; Wong, C. H. *J. Am. Chem. Soc.* **2006**, *128*, 15026−15033.
(25) Yang, Y. Y.; Ficht, S.; Brik, A.; Wong, C. H. *J. Am. Chem. Soc.* **2007**, *129*, 7690−7701.
(26) Hsieh-Wilson, L. C. *Nature* **2007**, *445*, 31−33.
(27) Gupta, R.; Birch, H.; Rapacki, K.; Brunak, S.; Hansen, J. E. *Nucleic Acids Res.* **1999**, *27*, 370−372.
(28) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10068−10073.
(29) Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, 3267−3272.
(30) De Koning, M. C.; Filippov, D. V.; van der Marel, G. A.; van Boom, J. H.; Overhand, M. *Eur. J. Org. Chem.* **2004**, 850−857.

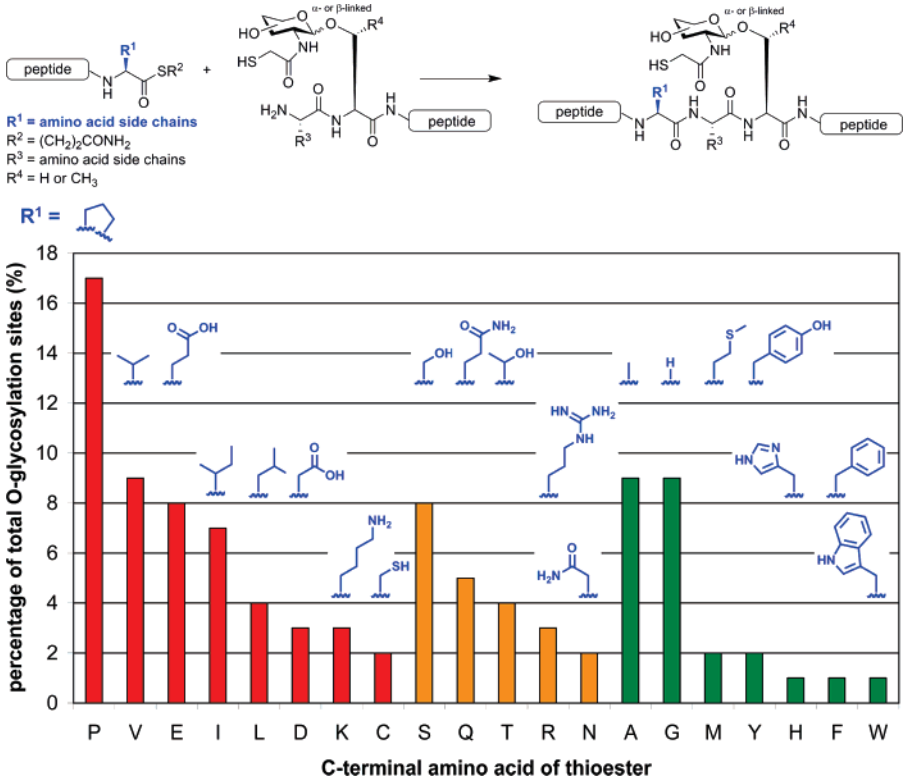


Figure 1. Histogram representing the number of O-glycosylation sites of O-GlycBase v6.00 that can be utilized for sugar-assisted ligation (SAL). The assessment was based on the C-terminal thioester component that would be required (two amino acid residues N-terminal to the glycosyl amino acid, **R**[1]): (red) residues that cannot be ligated by SAL; (orange) residues that can potentially be ligated by SAL but represent slow ligations; (green) residues that can be ligated by SAL.

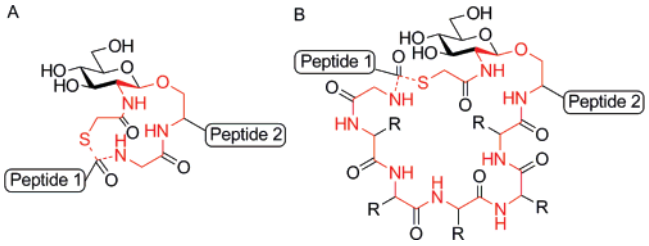


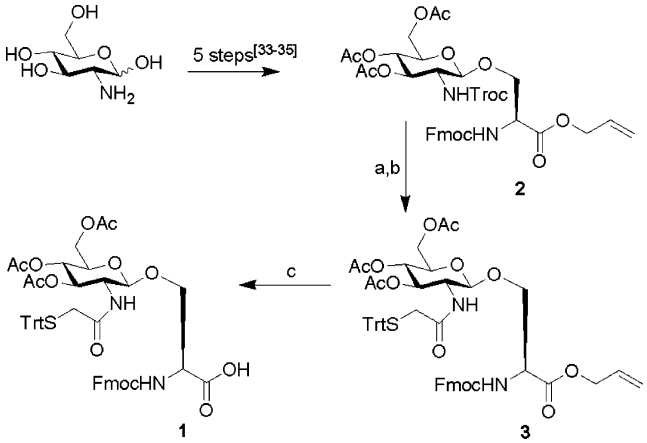
Figure 2. Proposed transition states of (a) SAL (14-membered ring) and (b) extended SAL with six amino acid extensions N-terminal to the glycosylated residue (29-membered ring); R = amino acid side-chain functionality.

of the extended SAL, the transthioesterification, would be identical to that of SAL (Scheme 1). However, as a result of adding extra amino acids to the N-terminus, the proposed S → N acyl transfer would now proceed through 17−29-membered-ring transition states for single extension to six amino acid extensions, respectively, compared with a 14-membered-ring transition state for SAL (Figure 2). Precedence for S → N acyl shifts proceeding through comparable ring sizes encouraged us in pursuing this strategy.[31,32]

Results and Discussion

Synthesis of Glycosyl Amino Acid Building Block. The initial phase of the research involved the synthesis of the glycosylated serine building block **1**, which was essential for solid-phase peptide synthesis (SPPS) of the desired auxiliary containing glycopeptides. The synthesis began from the known *β*-linked glycosyl amino acid **2**, synthesized from D-glucosamine in five steps (Scheme 2).[33−35] Subsequent introduction of the thiol auxiliary was achieved in two steps by liberation of the C-2 amine under reductive conditions, followed by coupling to *S*-trityl-2-mercaptoacetic acid to give **3** in 79% yield.[23] With

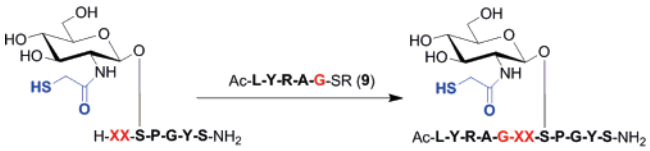
Scheme 2. Synthesis of Glycosyl Amino Acid Building Block **1** Containing a Thiol Auxiliary at the 2-Position[a]



[a] Conditions: (a) Zn dust, acetic acid, 88%; (b) *S*-trityl-2-mercaptoacetic acid, HBTU, DIEA, DMF, 90%; (c) $Pd(PPh_3)_4$, NMA, THF, 95%. HBTU = 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate, DIEA = *N*, *N*-diisopropylethylamine, NMA = *N*-methylaniline.

(31) Sasaki, S.; Shionoya, M.; Koga, K. *J. Am. Chem. Soc.* **1985**, *107*, 3371−3372.
(32) Gennari, C.; Molinari, F.; Piarulli, U.; Bartoletti, M. *Tetrahedron* **1990**, *46*, 7289−7300.
(33) Saha, U. K.; Schmidt, R. R. *J. Chem. Soc., Perkin Trans. 1* **1997**, 1855−1860.
(34) Dullenkopf, W.; CastroPalomino, J. C.; Manzoni, L.; Schmidt, R. R. *Carbohydr. Res.* **1996**, *296*, 135−147.
(35) Paulsen, H.; Helpap, B. *Carbohydr. Res.* **1991**, *216*, 289−313.

Table 1. Isolated Yields of Extended Sugar-Assisted Ligation (SAL) Reactions of Glycopeptides **4**−**8** and C-Terminal Glycine Thioester **9**[a]



ligation method	glycopeptide extension XX	ligation junction	isolated ligation yield	ligation half-life $t_{1/2}$
SAL	Gly	Gly-Gly	91%	9 h
exSAL	**4**: GlyVal	Gly-Gly	86%	12 h
dexSAL	**5**: GlyValLeu	Gly-Gly	70%	19 h
texSAL	**6**: GlyArgValLeu	Gly-Gly	60%	19 h
qexSAL	**7**: GlySerArgValLeu	Gly-Gly	38%[b]	19 h
pexSAL	**8**: GlyAlaSerArgValLeu	Gly-Gly	49%	>24 h

[a] Conditions: 4:1 v/v NMP:6 M Gn·HCl 1 M HEPES pH 8.5, 2% PhSH, 37 °C, 96 h. [b] The HPLC trace of this ligation suggests that the yield was similar to the texSAL (see Figure 3). The reduced isolated yield is a result of difficult product isolation by preparative HPLC.

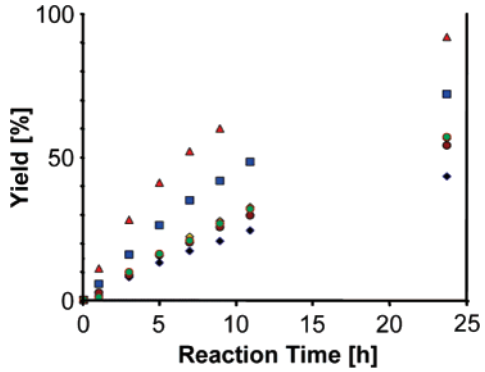
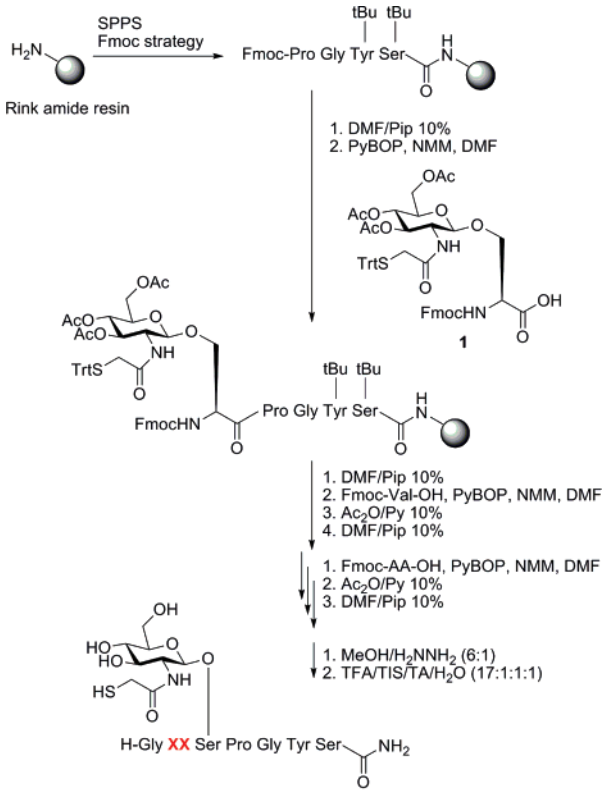


Figure 3. Kinetics for ligation reactions of glycopeptides **4**−**8** and C-terminal glycine peptide thioester **9** using SAL (red triangle), exSAL (blue square), dexSAL (yellow diamond), texSAL (brown circle), qexSAL (green circle), and pexSAL (black diamond).

the desired thiol auxiliary in place, the final step involved palladium-catalyzed removal of the allyl group to afford the desired β-glycosyl amino acid building block **1** in 95% yield (Scheme 2).

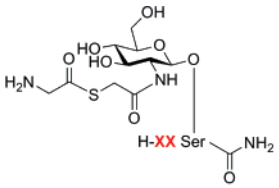
Solid-Phase Peptide Synthesis of Glycopeptides. Synthesis of the N-terminal extended glycopeptides **4**−**8** was achieved by SPPS following the Fmoc-strategy, starting from Fmoc-protected Rink amide resin. To avoid unnecessary expenditure of monomer **1**, only 1 equiv was coupled and the reaction time was increased to 6 h. The UV absorption of the Fmoc/piperidine adduct at 302 nm suggested that these conditions provided acceptable coupling yields. Amino acids possessing sterically hindered side chains were incorporated N-terminal to the glycosylated amino acid to demonstrate the applicability of the method to challenging glycopeptide and glycoprotein sequences. Specifically, we chose N-terminal amino acid sequences which were not ligatable by the normal SAL methodology. For example, glycopeptide **8** contained the sequence Ala-Ser-Arg-Val-Leu before incorporation of an N-terminal glycine. After the desired amino acid sequence was coupled, acetate groups were first removed by hydrazinolysis and the resin treated with trifluoroacetic acid/thioanisole/triisopropylsilane/water (17:1:1:1) to release the fully unprotected oligomers from the solid support. After purification by reverse phase

Scheme 3. SPPS (Fmoc Strategy) of Extended Glycopeptides **4**−**8**[a]



glycopeptide	XX	isolated yield (overall, after HPLC)
4	Val	52%
5	ValLeu	46%
6	ArgValLeu	54%
7	SerArgValLeu	42%
8	AlaSerArgValLeu	28%

[a] PyBOP = benzotriazole-1-yl-oxy-*tris*-pyrrolidino-phosphonium hexafluorophosphate; Pip = piperidine; TFA = trifluoroacetic acid; TIS = triisopropylsilane; TA = thioanisole; NMM = *N*-methylmorpholine.



glycopeptide	XX
10	-
11	Gly
12	GlyVal
13	GlyValLeu
14	GlyArgValLeu
15	GlySerArgValLeu
16	GlyAlaSerArgValLeu

Figure 4. Structures of model thioester intermediates **10**−**16** for molecular dynamics simulations.

HPLC, glycopeptides **4**−**8** were obtained in yields between 28 and 54%. Peptide thioesters bearing the amino acids glycine,

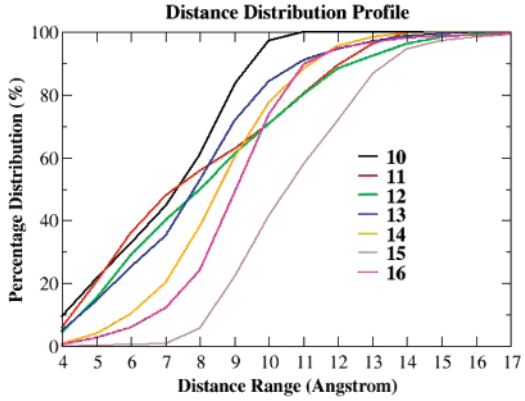


Figure 5. Distribution profile for the distance between the amine of the N-terminus and the thioester carbonyl in the model thioester intermediates.

histidine, alanine, and tyrosine on the C-terminus were synthesized using SPPS via the Boc strategy (see Supporting Information).

Extended Sugar-Assisted Ligation Screen. Initial ligation reactions of the extended glycopeptides and thioester **9** bearing a C-terminal glycine residue were conducted under the standard ligation conditions (6 M Gn·HCl, 100 mM potassium dihydrogen phosphate, pH = 8.5, 2% PhSH, 37 °C). In our hands, these conditions caused significant quantities of hydrolyzed thioester, leading to diminished ligation yields. To circumvent this problem, a range of mixed solvent ligation conditions were analyzed. The most effective solvent system proved to be a mixture of *N*-methyl pyrrolidinone (NMP) and HEPES buffer (4:1 v/v NMP/6 M Gn·HCl, 1 M HEPES, pH 8.5, 2% PhSH, 37 °C) which allowed for facile ligation reactions, coupled with minimal thioester hydrolysis.[36] Gratifyingly, the reaction of all extended glycopeptides (**4**−**8**) and thioester (**9**) under these conditions gave ligated products which were isolated in good yields when compared to the original SAL reaction (Table 1). The extended SAL (exSAL) gave an isolated yield of 86%, higher than the double-extended SAL (dexSAL) and triple-extended SAL (texSAL) which reacted in 70% and 60% yields, respectively. The quadruple-extended and penta-extended SAL (qexSAL and pexSAL) reactions proceeded in lower yields (38% and 49%, respectively).

Kinetic Studies. It is clear from the above study that there is a significant reduction in ligation yield as amino acids are added to the N-terminus of the glycopeptides. The next phase of the research focused on the analysis of the ligation kinetics in the hope that this would shed some light on the isolated yields obtained. To this end, a direct rate comparison of the extended SAL methods with the SAL method reported previously was conducted.[23] The rate of product formation was monitored by HPLC every 2 h for the first 11 h, and an endpoint was determined after 24 h (Figure 3). The half-life of the traditional SAL reaction was 9 h, slightly faster than the exSAL reaction which displayed a half-life of 12 h (Table 1). In agreement with the ligation yields shown in Table 1, ligation rates became more sluggish as additional amino acids were added to the N-terminus. Interestingly, dexSAL, texSAL, and qexSAL reactions exhibited similar rates of ligation with half-lives of 19 h (Table 1). Addition of a sixth amino acid residue in pexSAL resulted in a

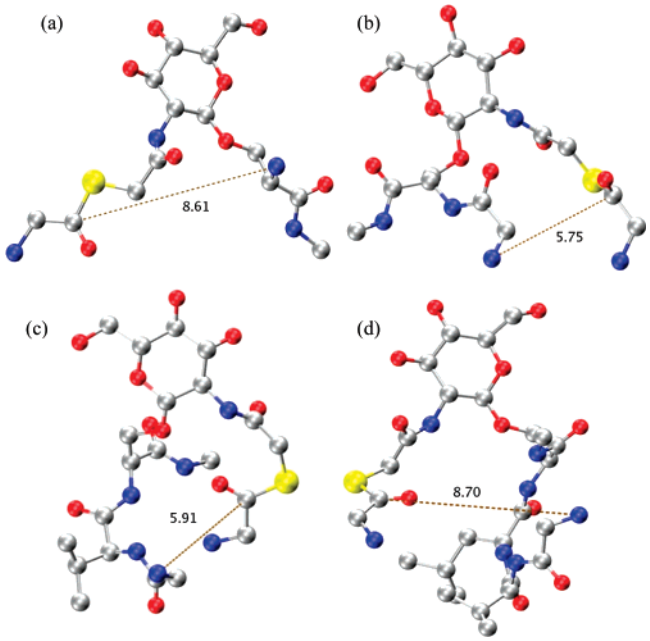


Figure 6. Snapshots of molecular dynamics simulations, showing the local maxima distances (in Å) between the N-terminal amine and the thioester carbonyl carbon for (a) **10**, (b) **11**, (c) **12**, and (d) **13**.

significant drop in ligation rate, whereby the reaction had only proceeded in 42% after 24 h. These kinetic studies clearly indicate that SAL is more facile than its extended counterparts; however, these rates differ by less than an order of magnitude, and as such, exSAL, dexSAL, texSAL, qexSAL, and pexSAL represent synthetically useful methods.

Molecular Dynamics Simulations. To rationalize the observed kinetics, molecular dynamics studies were conducted. The simulations were used to calculate distances between the two proposed reaction centers in the thioester intermediate, that is, the N-terminal amine and the carbonyl carbon of the thioester intermediate. On the assumption that the S → N acyl shift is the rate-limiting step of the reaction (as was observed in the kinetic studies), this information could then be used to gain an understanding of the relative rates of reactions observed. To this end, six glycopeptides were built using Maestro[37] and molecular dynamics simulations performed using AMBER 9.[38] These included a control glycopeptide **10**, with no extensions N-terminal to the glycosylated serine residue and **11**−**16** (Figure 4), which contain the same extensions as glycopeptides **4**−**8** described in the ligation and kinetic studies above (Table 1 and Figure 3). Equilibration of the system was necessary prior to collection of the data. The density profiles of the glycopeptides with 0−6 extensions show that all the systems are equilibrated after 20 ps of constant pressure simulations (see Supporting Information). Over the subsequent 10 ns of simulation, the distance profile of all glycopeptides showed random fluctuations of HN−CO distance and spontaneous transitions between long and short HN−CO distances (see Supporting Information).

Figure 5 describes the distribution of HN−CO distances that fall into the range of 3 to 17 Å. It is clear from these studies that glycopeptide **11** with one glycine extension exhibits the highest percentage of snapshots with the proximity distance

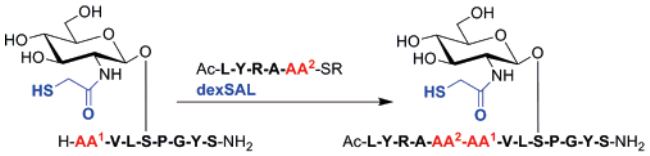
(36) Ficht, S.; Payne, R. J.; Brik, A.; Wong, C. H. *Angew. Chem., Int. Ed.* **2007**, *46*, 5975−5979.
(37) Schrödinger Inc. http://www.schrodinger.com (accessed 08/02/2007).
(38) Case, D. A.; et al., *Amber 9;* University of California: San Francisco, CA, 2006.

Table 2. Scope of the Extended Sugar-Assisted Ligation (exSAL)[a]

entry	thioester	glycopeptide	ligation junction -AA^2-AA^1-	isolated ligation yield
1	Gly	Gly	Gly-Gly	86%
2	His	Gly	His-Gly	70%
3	Ala	Gly	Ala-Gly	44%
4	Gly	His	Gly-His	77%
5	His	His	His-His	73%
6	Ala	His	Ala-His	77%
7	Gly	Asp	Gly-Asp	91%
8	His	Asp	His-Asp	91%
9	Ala	Asp	Ala-Asp	64%

[a] Conditions: 4:1 v/v NMP:6 M Gn·HCl 1 M HEPES pH 8.5, 2% PhSH, 37 °C, 96 h.

Table 3. Scope of the Double Extended Sugar-Assisted Ligation (dexSAL)[a]



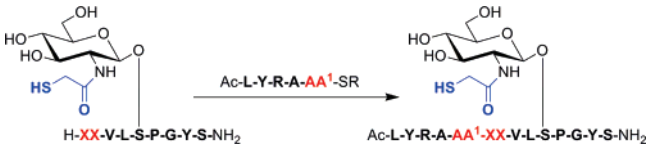
entry	thioester	glycopeptide	ligation junction -AA^2-AA^1-	isolated ligation yield
1	Gly	Gly	Gly-Gly	70%
2	His	Gly	His-Gly	65%
3	Ala	Gly	Ala-Gly	53%
4	Tyr	Gly	Tyr-Gly	44%
5	Gly	His	Gly-His	48%
6	His	His	His-His	44%
7	Ala	His	Ala-His	28%
8	Tyr	His	Tyr-His	24%
9	Gly	Asp	Gly-Asp	76%
10	His	Asp	His-Asp	60%
11	Ala	Asp	Ala-Asp	49%
12	Tyr	Asp	Tyr-Asp	68%

[a] Conditions: 4:1 v/v NMP/6 M Gn·HCl 1 M HEPES, pH 8.5, 2% PhSH, 37 °C, 96 h.

between 4 and 6 Å. A significant percentage of the total population (36%) exhibits HN−CO distances within 6 Å, which suggests a favorable S → N acyl transfer reaction. This is indeed the experimental observation, whereby the glycopeptide containing the identical extension to **11** underwent the most rapid ligation (Figure 3). As the number of amino acids in the extension increases from one to six, the percentage of snapshots that satisfy N−C proximity decreases monotonously, with the maximum distribution shifting to a longer distance range. When the length of the glycopeptide is increased to include six amino acid extensions in **16**, the total population within 6 Å HN−CO distance range is reduced to 6%. An anomaly exists for glycopeptide **15** containing five extensions, which shows the longest NH to CO distance out of all the model systems examined.

To visualize the results of the molecular simulation studies, Figure 6 shows the snapshot solutions where the NH to CO distance showed maximal frequency in model compounds **10**−**13** (see Supporting Information). For compounds **11**−**12** this distance is 5.75−5.91 Å, presumed to be favorable for the S → N acyl transfer. In contrast, control peptide **10** and extended glycopeptide **13** exhibited maximal frequency at distances of 8.61 and 8.70 Å, respectively, suggesting a less favorable acyl shift and hence a slower ligation rate, as was observed in the kinetic studies (Figure 3).

Table 4. texSAL, qexSAL, and pexSAL of Glycopeptides **6**−**8** with C-Terminal Glycine and Alanine Peptide Thioesters[a]



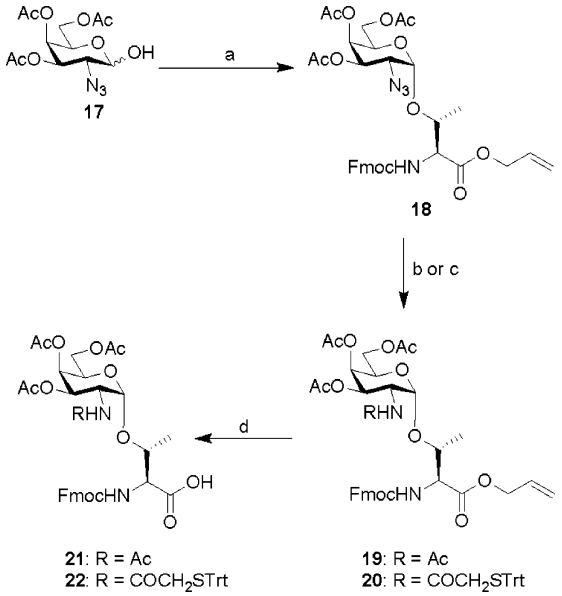
entry	thioester	glycopeptide XX	ligation junction	isolated ligation yield
1	Gly	GlyArg	Gly-Gly	60%
2	Gly	GlySerArg	Gly-Gly	38%
3	Gly	GlyAlaSerArg	Gly-Gly	49%
4	Ala	GlyArg	Ala-Gly	48%
5	Ala	GlySerArg	Ala-Gly	31%
6	Ala	GlyAlaSerArg	Ala-Gly	34%

[a] Conditions: 4:1 v/v NMP/6 M Gn·HCl 1 M HEPES, pH 8.5, 2% PhSH, 37 °C, 96 h.

From the distance profile obtained from the molecular simulation studies of **10** we would still expect a ligation to take place, albeit at a significantly slower rate (**10** displays a maximum population of 22.4% between 8 and 9 Å, see Supporting Information). However, it has previously been established that glycopeptides lacking an amino acid extension N-terminal to the glycosylated amino acid are unable to ligate using SAL.[23] It is therefore presumed that ring strain in the transition state is an additional factor that must be considered. The transition state for glycopeptide **10** would be an 11-membered ring, compared with a 14-membered ring for the single extended glycopeptide **11**, which is able to ligate rapidly with a range of thioesters. The 11-membered ring transition state is therefore presumed to possess significant ring strain, which results in an inhibited ligation rate.

Scope of exSAL. To study the effect of other amino acid residues at the ligation junction in exSAL, the glycopeptide−peptide thioester pairs shown in Table 2, entries 2−9 were examined. Glycopeptides containing N-terminal histidine and aspartic acid residues were synthesized by SPPS (Fmoc strategy) in an analogous fashion to those described in Scheme 3 (see Supporting Information). These were reacted with peptide thioesters containing C-terminal glycine, histidine and alanine residues using the mixed solvent system conditions. The ligation between glycopeptide **4** and the peptide thioester **9** bearing a C-terminal glycine gave an isolated yield of 86% (entry 1). The peptide thioester containing a C-terminal histidine residue also ligated in good yield (70%, entry 2), consistent with the previously reported SAL method.[23,24] As a result of changing the C-terminal amino acid of the thioester to an alanine residue, the ligation efficiency decreased (44%, entry 3) in accord with the established reactivity of peptide thioesters in SAL and NCL reactions.[28] Remarkable ligation yields were achieved for glycopeptides which bear an N-terminal aspartic acid or histidine residue. The isolated ligation yields for these glycopeptides exceeded those for the glycine glycopeptide (64−91% yield,

Scheme 4. Synthesis of Galactosamine-α-*O*-threonine Building Blocks **21** and **22**[a]



[a] Conditions: (a) (i) CCl_3CN, K_2CO_3, CH_2Cl_2, (ii) Fmoc-Thr-OAllyl, TMSOTf, DCM, ether, −30 °C, 55% over two steps 2:1 α:*β*; (b) Zn, Ac_2O, AcOH, 96%; (c) (i) Zn, AcOH, (ii) $TrtSCH_2CO_2H$, HBTU, DIEA, DMF, 58% over two steps; (d) $Pd(Ph_3)_4$, NMA, THF, 76% for **21** (R = Ac), quantitative for **22** (R = $COCH_2STrt$).

entries 7−9). Surprisingly, even sterically challenging ligation junctions such as His-His (73%, entry 5), Ala-His (77%, entry 6), and His-Asp (91%, entry 8) gave excellent yields.

Subsequent desulfurization of the ligated products described in Table 2 furnished glycopeptides containing the native *N*-acetyl functionality at the 2-position. This was achieved by hydrogenation conditions using palladium on alumina.[23,24] Products were generally isolated (after HPLC purification) in quantitative yields (see Supporting Information).

Scope of dexSAL. Encouraged by the implementation of exSAL to a range of different ligation partners, we examined the scope of a further extension by application of the dexSAL method to a variety of amino acids at the ligation junction (Table 3). The glycopeptides were synthesized *via* SPPS (Fmoc strategy, see Supporting Information). These glycopeptides were reacted with peptide thioesters bearing C-terminal glycine, alanine, histidine, and tyrosine residues using the previously described mixed solvent system. Reactions using glycine glycopeptides afforded ligation products in good yields (44−70%, entries 1−4). In contrast, ligations of glycopeptides containing an N-terminal histidine residue proceeded in significantly lower yields when compared to the exSAL method described in Table 2 (24−48%, entries 5−8). This can be rationalized by a slower rate of ligation in this reaction, validated by the kinetic and molecular dynamics studies (Figures 3 and 5). Gly-His and His-His ligations were isolated in satisfactory yields; however, incorporation of alanine and tyrosine peptide thioesters resulted in a significant decline in ligation yield (28% and 24%, respectively, entries 7 and 8). Aspartate extended glycopeptides ligated in good yields (49−76%, entries 9−12) for all cases; however, these were slightly lower than the exSAL cases shown in Table 2. This is once again presumed to be due to a more sluggish rate of ligation for these glycopeptides as described in the kinetic studies. The ability to conduct dexSAL reactions with sterically challenging tyrosine peptide thioesters clearly demonstrates the potential of this method.

Encouraged by the observed flexibility of the exSAL and dexSAL methods to a variety of amino acids at the ligation junction, we wondered if amino acids other than glycine could also be ligated to glycopeptides **6**, **7**, and **8** *via* texSAL, qexSAL, and pexSAL, respectively. To this end, glycopeptides **6**−**8** were reacted with a peptide thioester containing a C-terminal alanine residue. Gratifyingly, these glycopeptides ligated in satisfactory yields (48%, 31%, and 34% for texSAL, qexSAL, and pexSAL, respectively) after 96 h reaction time. As such, although reaction rates are significantly slower, these examples still represent potentially useful ligation reactions (Table 4, entries 4−6). Future research will aim to elucidate how many N-terminal extensions can be tolerated by the SAL method before the isolated yields become unsatisfactory.

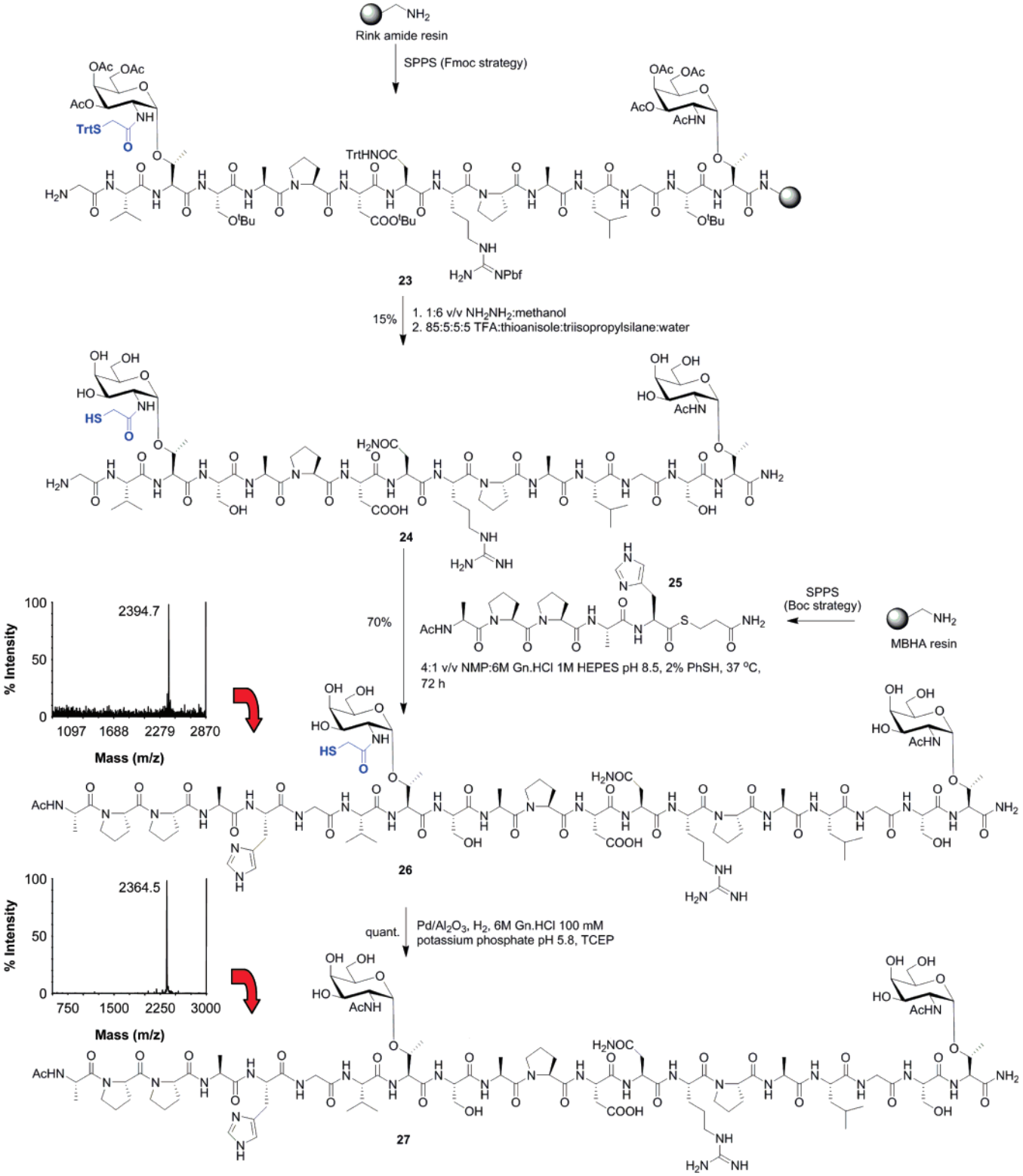
Synthesis of MUC1 Repeat Starting Unit by exSAL. After demonstrating the effectiveness of the extended SAL methodology for the synthesis of model glycopeptides, we next turned our attention to the synthesis of a glycopeptide of biological significance, the MUC1 repeating unit. MUC1 is a heavily O-glycosylated glycoprotein that is present at the interface between epithelial cells and their extracellular matrix.[39] The extracellular domain consists of tandem repeating units composed of twenty amino acid residues. In tumor cells, the expression of MUC1 is drastically increased, which is coupled with down regulation of certain glycosyltransferases and concomitant overexpression of several sialyltransferases, resulting in prematurely sialylated glycans linked to the MUC1 protein.[40] This aberrant glycosylation results in a significant alteration in the conformation of the peptide backbone thus leading to the exposure of distinct peptide epitopes, which become accessible to the immune system.[41] As such, the incorporation of specific glycans to the MUC1 repeating unit is considered as a promising target for the production of immunostimulating antigens.[42]

As part of our long-term goal to synthesize repeating oligomers of cancer-associated glycopeptides we chose to embark on the synthesis of the C-terminal starting unit of the MUC1 tandem repeat glycoprotein to demonstrate the applicability of exSAL to this system. In both the C-terminal starting unit and MUC1 tandem repeat the amino acid adjacent to the N-terminal glycan is the sterically encumbered valine residue. This would make the ligation of the corresponding C-terminal glycine thioester difficult by standard SAL methods. Extension by a further amino acid would now lead to a glycopeptide bearing an N-terminal glycine that could be reacted with a C-terminal histidine peptide thioester. This ligation was shown to be efficient in the model exSAL studies (Table 2, entry 2, 70% isolated yield).

To prepare the proposed glycopeptide target, two galactosamine building blocks, α-linked to Fmoc−threonine, were first synthesized (Scheme 4). Formation of the trichloroacetimidate and subsequent glycosylation from the known alcohol

(39) Dziadek, S.; Kunz, H. *Synlett* **2003**, 1623−1626.
(40) Dziadek, S.; Brocke, C.; Kunz, H. *Chem. Eur. J.* **2004**, *10*, 4150−4162.
(41) Dziadek, S.; Griesinger, C.; Kunz, H.; Reinscheid, U. M. *Chem. Eur. J.* **2006**, *12*, 4981−4993.
(42) Dziadek, S.; Kowalczyk, D.; Kunz, H. *Angew. Chem., Int. Ed.* **2005**, *44*, 7624−7630.

Scheme 5. Synthesis of the MUC1 Starting Unit Diglycopeptide by SPPS Followed by exSAL and Desulfurization



17[43] gave a 2:1 mixture of α/β isomers in 55% yield. It is important to note that the use of a mixed solvent system in the glycosylation reaction (1:1 v/v dichloromethane/diethyl ether) gave higher selectivity for the α-anomer than that previously reported.[25] Reduction of the desired α-anomer **18** with zinc dust in the presence of acetic anhydride and acetic acid gave glycosyl amino acid **19** containing an *N*-acetyl group in 96% yield. Alternatively, reduction of **18** with zinc dust and acetic acid, followed by coupling to *S*-trityl-2-mercaptoacetic acid gave glycosyl amino acid **20** containing the thiol auxiliary at the 2-position.[25] Finally, the allyl groups of **19** and **20** were cleaved using $Pd(PPh_3)_4$ and NMA to afford the desired glycosyl amino acid building blocks **21** and **22** in 76% and quantitative yield, respectively.

(43) Koeller, K. M.; Smith, M. E. B.; Wong, C. H. *Bioorg. Med. Chem.* **2000**, *8*, 1017−1026.

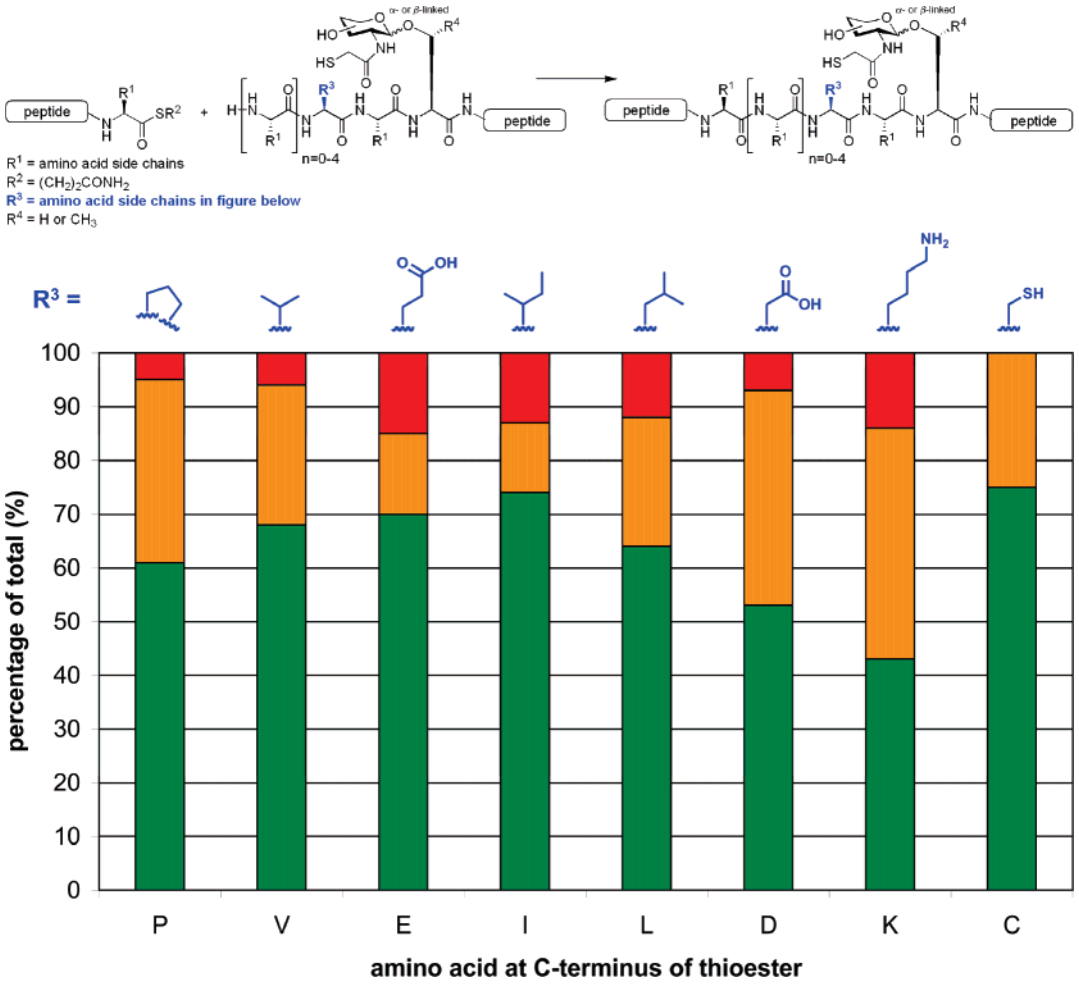


Figure 7. Histogram representing the O-glycosylation sites from O-GlycBase that could not be synthesized using sugar-assisted ligation (SAL). These sites have been re-evaluated for applications of the extended SAL methods (exSAL, dexSAL, texSAL, qexSAL, or pexSAL, two to six amino acid residues N-terminal to the glycosyl amino acid) Key: (red) residues that cannot be ligated by exSAL, dexSAL, texSAL, qexAL, or pexSAL; (orange) residues that can potentially be ligated by exSAL, dexSAL, texSAL, qexAL, or pexSAL but represent slow ligations; (green) residues that can be ligated by exSAL, dexSAL, texSAL, qexAL, or pexSAL.

Synthesis of the desired 15-mer diglycopeptide **23** was achieved by SPPS via the Fmoc strategy (see Supporting Information). The *N*-acetylgalactosamine building block **21** was incorporated as the first residue, and glycosyl amino acid **22**, containing the thiol auxiliary at the 2-position, was coupled as the third residue from the N-terminus of the glycopeptide (Scheme 5). After acetate removal, side-chain deprotection, and cleavage from the resin, the resulting diglycopeptide **24** was isolated in 15% overall yield. Diglycopeptide **24** was ligated with the corresponding peptide thioester (Ac-APPAH-S$(CH_2)_2$-$CONH_2$, **25**, synthesized by the Boc strategy) under the mixed-solvent conditions. This furnished the desired 20-mer glycopeptide **26** in 70% yield. Finally, desulfurization of the thiol auxiliary gave the starting unit of the MUC1 repeat **27** in quantitative yield (Scheme 5). The successful implementation of the exSAL method for the synthesis of this MUC1 starting unit will now serve as a platform to sequentially ligate amino acid units of the MUC1 repeat in future research. It is hoped that the resulting oligomers may serve as new immunogenic constructs for the production of immunostimulating antigens for the development of cancer vaccines.

Revised Scope of SAL for the Assembly of Naturally Occurring Glycoproteins. The ability to utilize the SAL methodology for ligation of glycopeptides extended by up to six amino acid residues N-terminal to the glycosylated amino acid should significantly increase the synthetic accessibility of glycopeptides (and potentially glycoproteins) by this method. To demonstrate the proposed generality and potential scope of the extended SAL methods with respect to the synthesis of native glycoproteins, we chose to re-examine all targets from O-GlycBase v6.00 that could not be synthesized using the traditional SAL methodology. These glycoproteins were re-evaluated by scanning amino acid residues N-terminal to the O-glycosylation sites (by up to six amino acids) until a feasible ligation junction was found. As with the previous informatics study, a feasible ligation junction was qualified by assessing the nature of the thioester component of the proposed ligation. The application of extended SAL disconnections to these O-glycosylation sites is depicted in Figure 7.

It is clear from Figure 7 that the probability that a given glycopeptide or glycoprotein target can be synthesized is greatly increased by the application of exSAL, dexSAL, texSAL, qexSAL, and pexSAL methodologies. Of the O-glycosylation sites that could not be synthesized by SAL, 65% are now accessible by these methods (indicated in green in Figure 7). A further 26% of O-glycosylation sites represent slower ligations,

but remain synthetically accessible (indicated in orange in Figure 7). Notably, a combination of the results of this study and those described in Figure 1 suggest that 95% of all O-glycosylation sites reported in O-Glycbase can be utilized for either SAL or extended SAL reactions. Of these O-glycosylation sites, 62% represent rapid ligations, while 38% represent slower ligation reactions.

Conclusion

The inherent limitations in our previously developed SAL method for the total synthesis of native glycoproteins inspired the development of extended sugar-assisted ligation (SAL) methods, namely exSAL, dexSAL, texSAL, qexSAL, and pexSAL. Ligation products were isolated in high yields when extended glycopeptides were reacted with peptide thioesters bearing a C-terminal glycine residue by using a mixed-solvent system, which served to protect the thioester from hydrolysis. Kinetic studies indicated that ligation reactions proceeded at progressively slower rates as further amino acid extensions were incorporated into the glycopeptide. These observations were rationalized by the use of molecular dynamics simulations which suggested that the distances between the required reaction sites (the N-terminus and the carbonyl carbon of the thioester) in the proposed thioester intermediate also increased as glycopeptides were extended. As such, the rate is presumed to be a function of the proximity of these two entities and hence the ease at which the glycopeptide thioester intermediate can rearrange to afford the ligation product. All five of the extended SAL methods were shown to exhibit broad tolerance for a variety of amino acid residues at the ligation junction, including challenging residues such as alanine and tyrosine. In addition, the auxiliaries of several ligation products were desulfurized efficiently to afford glycopeptides displaying the native *N*-acetylglucosamine moiety. As a preliminary application, exSAL was effectively implemented in the synthesis of a naturally occurring glycopeptide, the starting unit of the MUC1 repeat. A detailed search of over five hundred O-glycosylation sites using the glycoprotein database O-GlycBase v6.00 reveals that many more sequences are accessible (when compared to SAL) using exSAL, dexSAL, texSAL, qexSAL, or pexSAL. Specifically, 95% of all O-glycosylation sites reported in O-Glycbase v6.00 can potentially be constructed by applications of the SAL or extended SAL technologies. As such, this method, either alone, or in combination with other ligation strategies should aid in the efficient syntheses of native glycopeptide and glycoprotein targets that were not previously attainable. Current efforts in our laboratory are focused on the application of this method to ligate glycopeptides bearing pre-assembled glycans, which can be installed using synthetic or enzymatic methods.

Experimental Section

General Procedure for Extended SAL Reactions. Glycopeptides (1.5 equiv, approximately 3 μmol) were dissolved in 150 μL of deoxygenated ligation buffer [4:1 v/v *N*-methyl-2-pyrrolidinone (NMP)/6 M guanidine hydrochloride, 1 M HEPES, pH = 8.5]. This solution was transferred to an eppendorf tube containing the thioester (ca. 2 μmol). Thiophenol (2% by volume, 3 μL) was added, and the reaction was mixed gently. The ligation mixture was incubated at 37 °C with gentle mixing every 12 h until the reaction was confirmed to be complete by LC−MS. The ligation reactions were quenched by the addition of TCEP solution (0.6 mL of a 10 mg/mL solution), and the product was purified by HPLC.

General Procedure for the Desulfurization of the Thiol Auxiliary. Ligated glycopeptides (1.0 mg) were dissolved in a degassed solution of 6 M Gn·HCl, 100 mM potassium dihydrogen phosphate, 10 mM TCEP buffer, pH = 5.8 (1.0 mL). Palladium on alumina (5 wt %, 16 mg) was added, and the reaction was stirred under an H_2 atmosphere for 1 h. The reaction was filtered, and the product was purified by HPLC.

Acknowledgment. This work was supported by the NIH and the Skaggs Institute for Chemical Biology. R.J.P. is grateful for funding provided by the Lindemann Trust Fellowship. S.F. is grateful for a postdoctoral fellowship provided by the Deutsche Akademische Austauschdienst (DAAD). We would also like to thank Dr. William Greenberg for proof reading this manuscript.

Supporting Information Available: Complete ref 38; detailed molecular dynamics simulation procedures and results, experimental procedures, product characterization, and other detailed results. This material is available free of charge via the Internet at http://pubs.acs.org.

JA073653P