Glycopeptide Ligation

DOI: 10.1002/ange.200801097

Uncovering a Latent Ligation Site for Glycopeptide Synthesis**

*Ryo Okamoto and Yasuhiro Kajihara**

Glycosylation, one of the most important posttranslational modifications, plays an important role in a variety of biological events.[1] Oligosaccharides on glycoproteins exhibit structural heterogeneity, which makes it difficult to elucidate the relationship between the oligosaccharide structure and the function of the glycoprotein.

Chemical synthesis is one of the powerful approaches for obtaining homogeneous glycoproteins.[2] We have already reported the synthesis of a glycoprotein with a homogeneous N-linked complex-type oligosaccharide.[3] This synthesis employed native chemical ligation (NCL) to perform peptide-segment coupling. NCL relies on the thiol-exchange reaction between a peptide with an α-thioester group at the C terminus and another peptide with a cysteine residue at the N terminus and on the subsequent intramolecular acyl transfer.[4] However, occasionally, the cysteine residue is not properly located or does not exist in the target protein. To take this potential difficulty into consideration, a long glycopeptide sequence that is 30–50 amino acids from one cysteine site to another cysteine site occasionally needs to be synthesized for glycoprotein synthesis by the NCL method. The synthesis of such a glycopeptide with an N-linked glycopeptide is not easy to perform[2,5] and requires an appropriate amount of N-linked complex-type oligosaccharides; therefore, there is greater difficulty in glycoprotein synthesis than in simple protein synthesis.

To examine NCL without a cysteine residue in a long target peptide, reduction methods changing the sulfhydryl group of cysteine to a hydrogen atom after NCL and utilizing an auxiliary group have been developed.[6] In the latter method, the amino acid sequence at the ligation site is limited for performance. For the development of a widely usable method in glycopeptide synthesis, we have also explored suitable NCL approaches; this endeavor enabled us to find a new ligation position at the serine site in the consensus sequence NXS (X: any amino acid except for proline), by which an asparagine residue is generally incorporated in an N-linked oligosaccharide. This sequence is found in glycoproteins along with the NXT sequence.[1] In order to use the serine site for a new NCL, we have examined the new concept and attempted the conversion of a cysteine residue into a serine residue after NCL. For such a technique, it was necessary to explore concise reaction sequences. As a result, we found the possibility of using a CNBr cleavage method at a methylcysteine site, which could be obtained by specific methylation of cysteine.[7] Herein, we report a new chemical ligation approach at serine sites, which relies on the conversion of a cysteine residue into a serine residue after NCL.

The strategy is shown in Scheme 1. After NCL (product **A**), the conversion of cysteine into serine was performed by the following reactions: S methylation of cysteine with methyl 4-nitrobenzenesulfonate (product **B**) and intramolecular rearrangement by activation with CNBr in 80% HCOOH solution followed by an *O*- to *N*-acyl shift. Activation of the *S*-methyl group by CNBr results in intramolecular attack by the neighboring carbonyl oxygen atom on the β-carbon atom of the methylcysteine residue and generates an *O*-ester peptide intermediate (product **C**). This intermediate can be converted into the desired peptide (product **D**) through the *O*- to *N*-acyl shift under slightly basic conditions (pH 7–8).

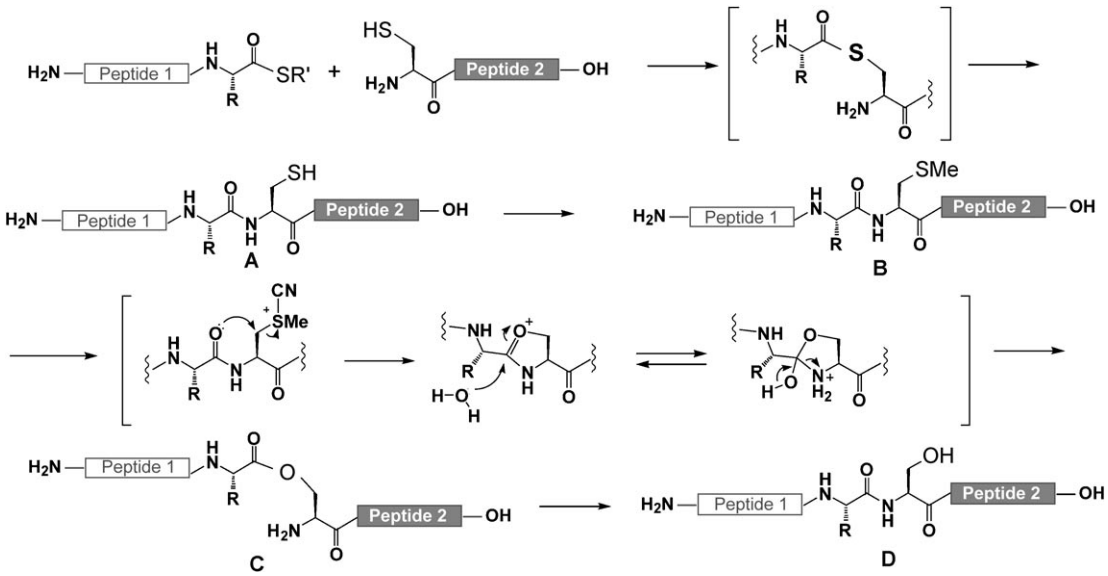
In order to examine this strategy, we first demonstrated the utility of the reaction by means of a model tetrapeptide with a cysteine residue (Table 1, entry 1). As shown in entry 1 in Table 1, in the case of tetrapeptide Ac-ACGL-OH, we could achieve the conversion of cysteine into serine in moderate yield. To confirm the optical purity of the peptide thus prepared, we compared it with authentic peptide samples containing D-amino acids, such as Ac-^{D}ASGL-OH, Ac-A^{D}SGL-OH, Ac-^{D}A^{D}SGL-OH, and Ac-ASGL-OH, by HPLC and NMR analysis (Figure 1 and the Supporting Information). These results showed that product **2** is identical to Ac-ASGL-OH and the conversion method did not cause any epimerization in the peptide. We also examined this method by using octa- and undecapeptides (Table 1, entries 2 and 3), and each of the conversion reactions was found to afford the desired peptides in moderate yield.

It is known that CNBr has also been used for cleavage at the methionine site in proteins. In order to distinguish methionine from methylcysteine residues, we introduced methionine in the sulfoxide form. Due to the fact that the sulfoxide form of methionine is inactive for the CNBr reaction,[8] we expected that an oxidation/reduction protocol[9] would enable us to use this new approach for the synthesis of peptides with methionine residues. We examined the strategy by using pentapeptide **7**, which contained cysteine and the sulfoxide form of methionine. S methylation of this pentapeptide afforded **9**. Conversion of *S*-methylcysteine to a serine residue and subsequent reduction of the sulfoxide group by NH_4I, SMe_2, and trifluoroacetic acid (TFA)[9] were performed as a one-pot reaction and afforded the desired pentapeptide **8** in good yield (73% yield of

[*] R. Okamoto, Prof. Dr. Y. Kajihara
International Graduate School of Arts and Sciences
Yokohama City University
22-2, Seto, Kanazawa-ku, Yokohama, 236-0027 (Japan)
Fax: (+81) 45-787-2413
E-mail: kajihara@yokohama-cu.ac.jp

[**] Financial support from the Japan Society for the Promotion of Science (Grant-in-Aid for Creative Scientific Research no. 17GS0420) is acknowledged.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200801097.



Scheme 1. Reaction mechanism of the conversion of cysteine into serine. For details see text.

Table 1: Conversion of cysteine into serine in model peptides.[a]

No.	Cys peptide →Ser peptide	Cys methylation [%]	CNBr conversion [%]	*O*- to *N*-acyl shift [%]
1	Ac-A**C**GL-OH (**1**) →Ac-A**S**GL-OH (**2**)	quant (>90)	75 (62)	quant (70)
2	VDKAV**C**GL-OH (**3**) →VDKAV**S**GL-OH (**4**)	97 (90)	85 (63)	90 (82)
3	LFRVY**C**NFLRG-OH (**5**) →LFRVY**S**NFLRG-OH (**6**)	90 (81)	51 (41)	quant (77)
4	Ac-G**C**GM(O)A-OH (**7**) → Ac-G**S**GMA-OH (**8**)	quant (80)	82 (73)[b]	

[a] Yields were estimated from HPLC peak areas, with yields after isolation given in parenthesis. [b] One-pot reaction.

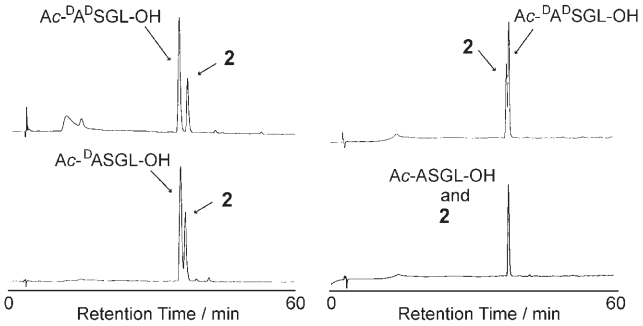
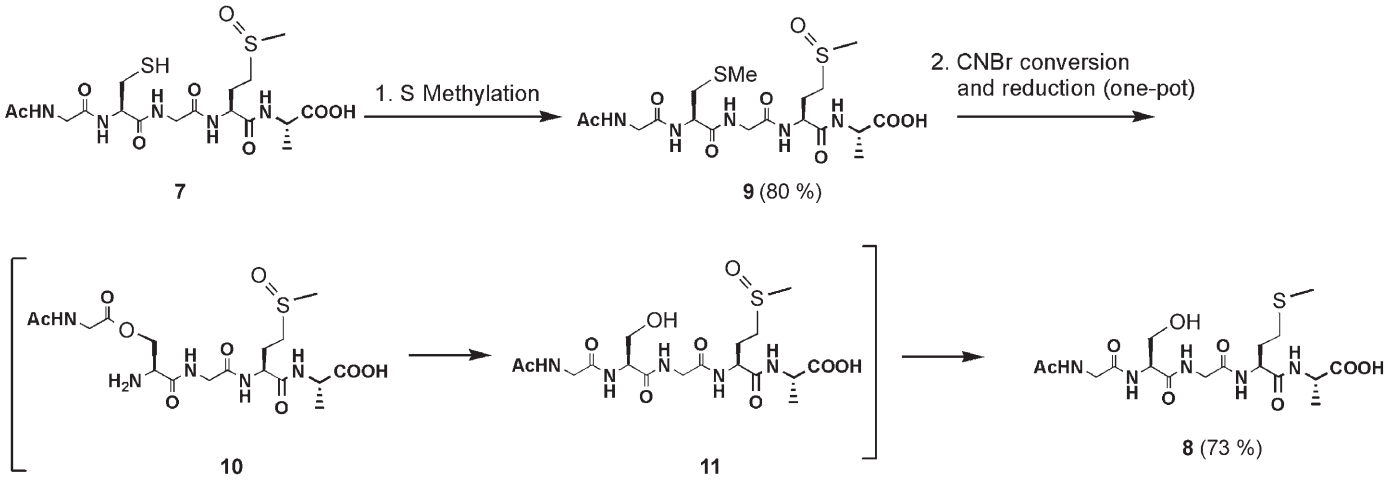


Figure 1. Analytical HPLC profiles (absorbance at 220 nm) of mixtures of synthetic tetrapeptide **2** and authentic tetrapeptides with D-amino acids. The ratios of the injected amounts of synthetic sample and authentic sample were about 1:2.

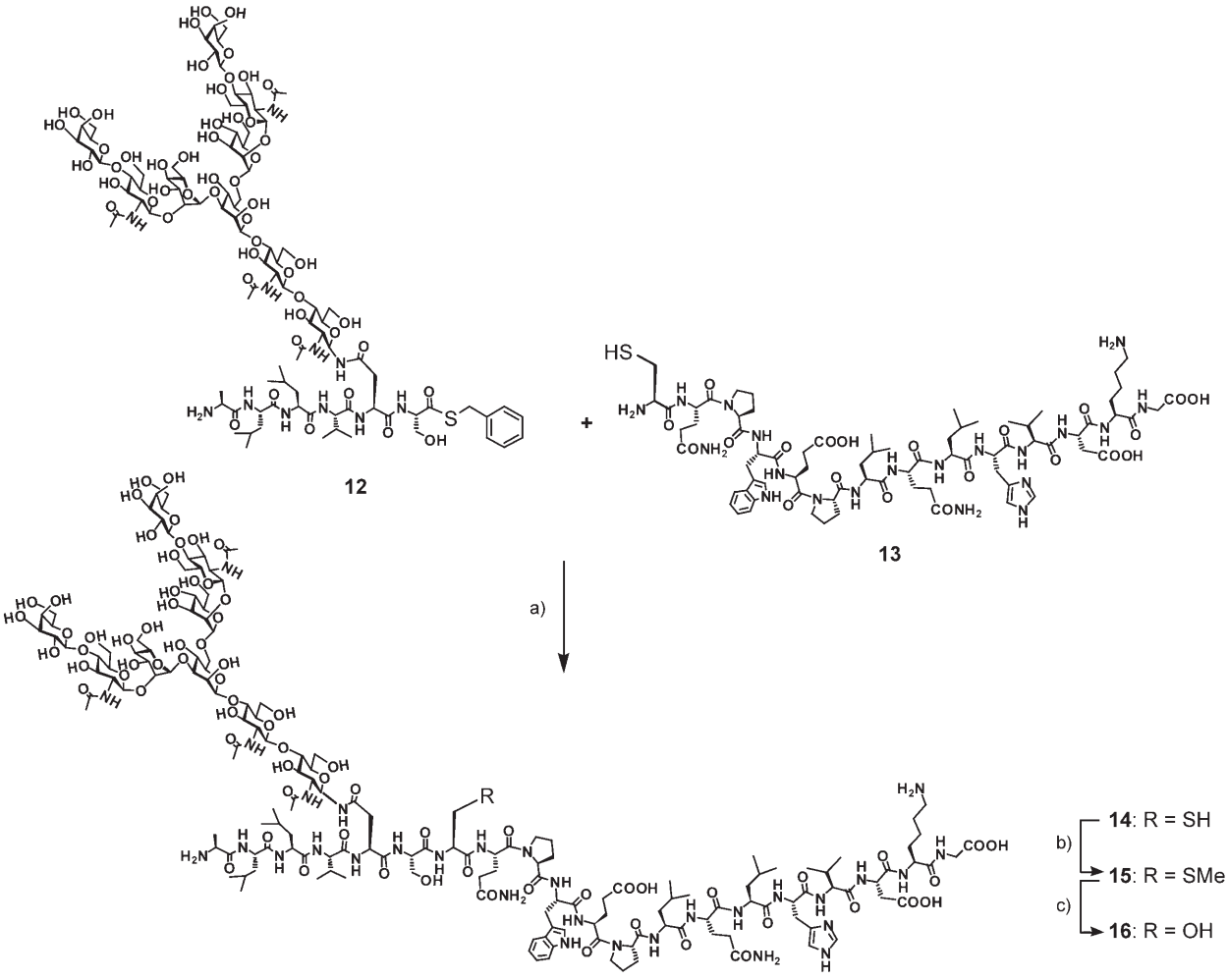
isolated product; Scheme 2 and data shown in the Supporting Information).

We then applied the new NCL method for the synthesis of an N-linked glycopeptide that is a fragment of erythropoietin (residues 79–98; Scheme 3). The NCL, between a glycosyl hexapeptide thioester with a complex-type N-linked asialo-oligosaccharide, **12**, prepared by a reported method,[10] and a tetradecapeptide with a cysteine residue, **13**, was performed by the conventional method. In this case, many of the hydroxy groups of the oligosaccharide were free (see the Supporting Information). The glycosyl icosapeptide **14** thus obtained was subjected to S-selective methylation (Figure 2a–c) followed by activation of the *S*-methylcysteine residue by CNBr (Figure 2d). The product was observed with broad peaks by reversed-phase HPLC (RP-HPLC; Figure 2e). We concluded that this was because of random formylation of the sugar hydroxy groups during the CNBr activation reaction under the formic acid and CNBr conditions. After lyophilization, the residue was dissolved in 5% hydrazine hydrate solution (or a slightly basic solution: <pH 10) for 10 min (Figure 2f) in order for the peptide to undergo the *O*- to *N*-acyl shift and to remove the formyl groups from the sugar hydroxy groups. As expected, this treatment afforded the target glycosyl icosapeptide with the serine residue, **16** (Figure 2g). The structure and purity of this glycopeptide was confirmed by comparison with an authentic sample synthesized by SPPS (see the Supporting Information).

This finding encouraged us to undertake the synthesis of an O-linked glycopeptide, the MUC1 repeat segment of the variable-number tandem-repeat region,[11] a segment in which there are abundant proline, threonine, and serine residues. The NCL between two O-linked glycosyl icosapeptides with mono *N*-acetylgalactosamine (GalNAc) segments, **17** and **18**, afforded the desired glycosyl tetracontapeptide **19** with a cysteine residue (Scheme 4 and the Supporting Information). The cysteine residue that was the ligation position in glycopeptide **19** was converted into a serine residue through the approach described above. This conversion afforded the desired glycopeptide **21** in good yield (43% overall conversion yield estimated by HPLC peak area, 16% yield of isolated product; Scheme 4 and the Supporting Information).



Scheme 2. Conversion of a cysteine into a serine residue in a pentapeptide with a methionine residue. S methylation was performed from **7**, which was synthesized by solid-phase peptide synthesis (SPPS).



Scheme 3. Synthesis of an N-linked glycosyl icosapeptide through ligation at the serine site. a) 6 M Guanidine hydrochloride, 0.1 M sodium phosphate buffer (pH 7.1) containing thiophenol (0.5% v/v) and phenylmethanthiol (0.5% v/v), 70% yield after isolation; b) 6 M guanidine hydrochloride, 0.25 M tris(hydroxymethyl)aminomethane/HCl (Tris-HCl), 3.3 mM ethylenediaminetetraacetate sodium salt (EDTA-2 Na) buffer (pH 8.6), CH_3CN, methyl 4-nitrobenzenesulfonate, 85% conversion yield estimated by HPLC peak area (67% yield after isolation); c) 1. CNBr, 80% HCOOH, 2. 5% hydrazine hydrate solution, 80% conversion yield estimated by HPLC peak area (70% yield after isolation).

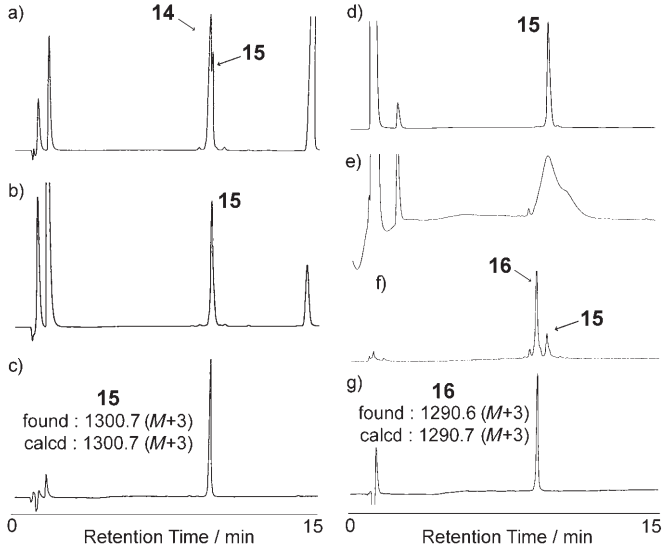
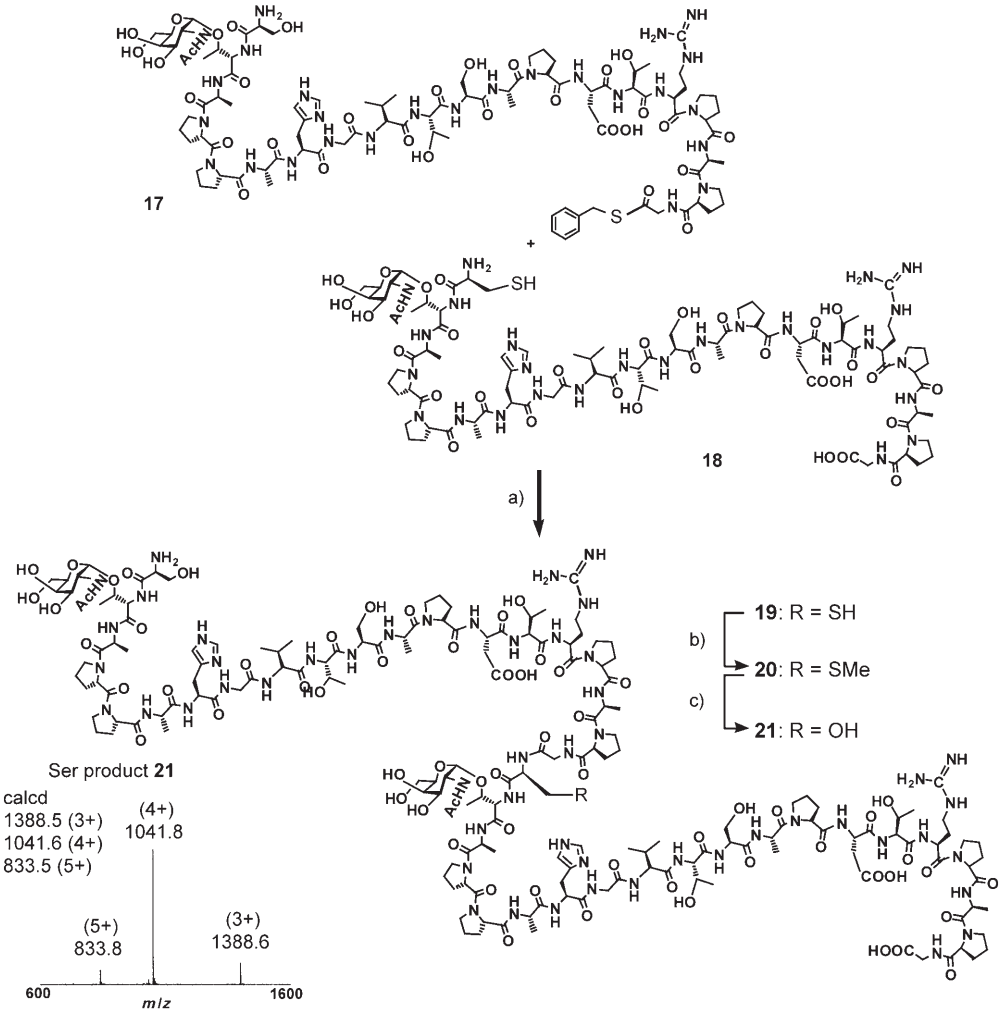


Figure 2. HPLC profiles (absorbance at 220 nm) of the conversion of a cysteine into a serine residue in the N-linked glycosyl icosapeptide **14**. S Methylation: a) at the start ($t < 1$ min), b) after 20 min, and c) after purification. Conversion of the *S*-methylcysteine into a serine residue: d) at the start ($t < 1$ min), e) after 38 h, f) after hydrazine treatment, and g) after purification.

In order to confirm the structure of the final glycopeptide, we examined the peptide digestion of compound **21** by actinase E (*Streptomyces griseus*, Kaken pharma, Tokyo), and this nonspecific digestion afforded a glycosyl octadecapeptide involving a ligation site (SAPDTRPAPGST(GalNAc)APPAHG). Comparison of this peptide fragment with the authentic glycosyl octadecapeptide, synthesized by SPPS, proved that the conversion reaction successfully afforded the desired serine residue from the cysteine residue (see the Supporting Information).

In summary, this report presents a new ligation strategy that uses a method of converting cysteine into serine residues. This method was successfully used for the synthesis of both N-linked and O-linked glycopeptides. Although the synthesis of large glycopeptides has been troublesome, NCL at a serine residue between a short glycopeptide thioester segment and a nonglycosylated long peptide segment could solve this problem, even if there is no cysteine in the target glycopeptide. We also demonstrated that this methodology could be used for peptides containing methionine, histidine, or tryptophan residues, which are potentially problematic residues during the CNBr conversion or methylation step. The conversion yield by this procedure may be dependent on the peptide sequence and number of amino acids in the peptide backbone. As far as we have monitored the reactions, we could not find β-elimination products or peptide–hydrazide derivatives. Research is underway to utilize this approach for the threonine site in peptides, in addition to the serine site, and to synthesize a number of glycoproteins.



Scheme 4. Conversion of a cysteine into a serine residue for the synthesis of the MUC1 repeat segment. a) 6 M Guanidine hydrochloride, 0.1 M sodium phosphate, 60 mM 4-mercaptophenylacetic acid, 20 mM tris(2-carboxyethyl)phosphine buffer (pH 7.2); b) methyl 4-nitrobenzenesulfsonate, 6 M guanidine hydrochloride, 0.25 M Tris-HCl 3.3 mM EDTA-2Na buffer (pH 8.6), CH_3CN; c) 1. CNBr, 80% HCOOH, 2. TFA, NH_4I, Me_2S, 3. 5% hydrazine hydrate solution.

Experimental Section

S Methylation of **14**: Compound **14** (1 equiv) was dissolved in the 0.25 M Tris-HCl buffer (pH 8.6) containing 6 M guanidine hydrochloride and 3.3 mM EDTA-2Na (peptide concentration was adjusted to 1 mM). Methyl 4-nitrobenzenesulfonate (20 equiv) in acetonitrile (peptide concentration was adjusted to 3 mM) was added, and the whole mixture was stirred for 20 min. The mixture was then neutralized by 10% TFA solution, and the mixture was washed with Et_2O to remove the excess methylation reagent. The solution was concentrated in vacuo. The residue was purified by RP-HPLC or chromatography on a short ODS (octadecylsilyl) column to afford the desired S-methylated peptide **15**.

Conversion of *S*-methylcysteine in **15** to a serine residue by CNBr: Peptide **15** (1 equiv) was dissolved in 80% HCOOH solution (peptide concentration was adjusted to 1 mM). CNBr (100 equiv) was added to the solution, and the mixture was stirred for 38 h in the dark and under an Ar atmosphere at 37 °C. The mixture was concentrated in vacuo or lyophilized. The residue was dissolved in 5% hydrazine hydrate solution and left for 10 min before the solution was neutralized with AcOH. The mixture was purified by RP-HPLC to afford the desired peptide with the serine residue, **16**.

Received: March 6, 2008
Published online: June 11, 2008

Keywords: chemical ligation · glycopeptides · glycoproteins · oligosaccharides

[1] R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683 – 720.
[2] a) L. Liu, C. S. Bennett, C. H. Wong, *Chem. Commun.* **2006**, 21 – 33; b) Y. Nakahara, *Biopolymers* **2007**, *88*, 308 – 324; c) Z. Guo, N. Shao, *Med. Res. Rev.* **2005**, *25*, 655 – 678.
[3] N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson, Y. Kajihara, *J. Am. Chem. Soc.* **2008**, *130*, 501 – 510.
[4] P. E. Dawson, M. J. Churchill, M. R. Ghadiri, S. B. H. Kent, *J. Am. Chem. Soc.* **1997**, *119*, 4325 – 4329.
[5] N. Yamamoto, A. Takayanagi, A. Yoshino, T. Sakakibara, Y. Kajihara, *Chem. Eur. J.* **2007**, *13*, 613 – 625.
[6] a) L. E. Canne, S. J. Bark, S. B. H. Kent, *J. Am. Chem. Soc.* **1996**, *118*, 5891 – 5896; b) D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent, P. Botti, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554 – 6559; c) T. Kawakami, K. Akaji, S. Aimoto, *Org. Lett.* **2001**, *3*, 1403 – 1405; d) J. Offer, C. N. C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642 – 4646; e) D. Macmillan, D. W. Anderson, *Org. Lett.* **2004**, *6*, 4659 – 4662; f) G. Chen, J. D. Warren, J. Chen, B. Wu, Q. Wan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, *128*, 7460 – 7462; g) L. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526 – 533; h) B. L. Pentelute, S. B. H. Kent, *Org. Lett.* **2007**, *9*, 687 – 690; i) A. Brik, S. Ficht, Y. Y. Yang, C. S. Bennett, C. H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 15026 – 15033; j) D. Crich, A. Benerjee, *J. Am. Chem. Soc.* **2007**, *129*, 10064 – 10065; k) P. Botti, S. Tchertchian, WO/2006/13962, **2006**; l) Q. Wan, S. J. Danishefsky, *Angew. Chem.* **2007**, *119*, 9408 – 9412; *Angew. Chem. Int. Ed.* **2007**, *46*, 9248 – 9252.
[7] E. Gross, J. L. Morell, *Biochem. Biophys. Res. Commun.* **1974**, *59*, 1145 – 1150.
[8] R. Kaiser, L. Metzka, *Anal. Biochem.* **1999**, *266*, 1 – 8.
[9] C. P. R. Hackenberger, *Org. Biomol. Chem.* **2006**, *4*, 2291 – 2295.
[10] Y. Kajihara, A. Yoshihara, K. Hirano, N. Yamamoto, *Carbohydr. Res.* **2006**, *341*, 1333 – 1340.
[11] a) S. J. Gendler, C. A. Lancaster, J. Taylor-Papadimitriou, T. Duhig, N. Peat, J. Burchell, L. Pemberton, E.-N. Lalani, D. Wilson, *J. Biol. Chem.* **1990**, *265*, 15286 – 15293; b) F.-G. Hanisch, S. Müller, *Glycobiology* **2000**, *10*, 439 – 440; c) T. R. E. Stadie, W. Chai, A. M. Lawson, P. G. Byfield, F.-G. Hanisch, *Eur. J. Biochem.* **1995**, *229*, 140 – 147.

