Glycopeptide Synthesis

DOI: 10.1002/anie.200700546

Second-Generation Sugar-Assisted Ligation: A Method for the Synthesis of Cysteine-Containing Glycopeptides**

*Simon Ficht, Richard J. Payne, Ashraf Brik, and Chi-Huey Wong**

Glycosylation is a very common post- or co-translational modification of proteins that has extensive biological significance.[1] Indeed, it is estimated that over fifty percent of all human proteins are glycosylated.[2] Protein glycosylation plays an important role for a variety of biological recognition events such as cell adhesion, cell differentiation, and cell growth.[3,4] Additionally, some parasites use heavily glycosylated membrane-bound proteins as port of entry.[5] Aberrant glycosylation of proteins often modifies intracellular recognition and is linked with several serious illnesses including autoimmune diseases, infectious diseases, and cancer.[6] In order to understand the role of the glycosylation at a molecular level, it is important to have access to homogeneous glycopeptides and glycoproteins. The glycosylation pattern of a given glycoprotein, unlike the protein element, is not under the control of a coding template, but rather is dictated by the relative activities of the constituent enzymes. The use of biological expression systems for production and study of glycoproteins has proved difficult and is hampered by the heterogeneity of the resulting products.

The necessity for homogeneous glycoproteins can be met by chemical and chemoenzymatic intervention.[7–10] In particular, a number of ligation methods have recently gained significant attention as techniques to facilitate the synthesis of such targets. Native chemical ligation (NCL), a chemoselective condensation reaction between a peptide thioester and a peptide bearing an N-terminal cysteine has proven very useful in this regard.[11–13] This method has been successfully implemented in the synthesis of hundreds of proteins to date.[13] The success of this method for peptide and protein synthesis has inspired many laboratories to employ this technology for studies towards the synthesis of glycopeptides and glycoproteins.[14–17] Although NCL has proved to be extremely powerful, certain limitations still exist with this method. The obvious limitation of NCL is the requirement for a cysteine residue at the ligation junction.[18] Cysteine has a relatively low abundance in nature (ca. 1.7%), and as such, there is a high probability that the target does not have a cysteine at a synthetically useful position. This led to the development of a number of cysteine-free ligation techniques,[19,20] some of which have been successfully implemented in the synthesis of glycopeptides.[21,22] However, the use of these methods, which rely upon the incorporation of an N-terminal auxiliary, is restricted to ligation sites containing amino acids of low steric bulk.

Our laboratory has recently reported a ligation method for the synthesis of cysteine-free O- and N-linked glycopeptides.[23,24] This method, dubbed sugar-assisted ligation (SAL), 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. The reaction cascade showed high sequence tolerance at the ligation junction, therefore expanding the number of potential targets accessible by this method. Additionally, the reaction was shown to be chemoselective even in the presence of nucleophilic amino acid side chains such as lysine.[23] This means that ligations could be conducted on glycopeptides free of protecting groups. For these reasons, SAL has recently gained the spotlight as a feasible method for the total synthesis of glycoproteins.[25] The major pitfall of this method, however, is the incompatibility of the conditions used for the removal of the auxiliary, which requires hydrogenation, with other thiol-containing residues. As mentioned above, the abundance of cysteine is low; however, a large proportion of naturally occurring glycoproteins contain this residue in their sequence. In addition many glycoproteins contain cysteine residues in nonstrategic positions, and as such, NCL cannot be implemented.

To circumvent this problem of chemoselective auxiliary removal, we embarked on the development of a second-generation SAL. This method relies on the incorporation of an auxiliary which can be removed in the presence of other cysteine residues after ligation. An acid-labile auxiliary cannot be used, as strongly acidic conditions are used in the solid-phase glycopeptide synthesis. In contrast, a base-labile auxiliary, in this case a mercaptoacetic acid moiety bound through an ester on the 3-position of the bridgehead sugar, would fulfill the required orthogonality. The mechanism of the proposed second-generation SAL is depicted in Scheme 1.

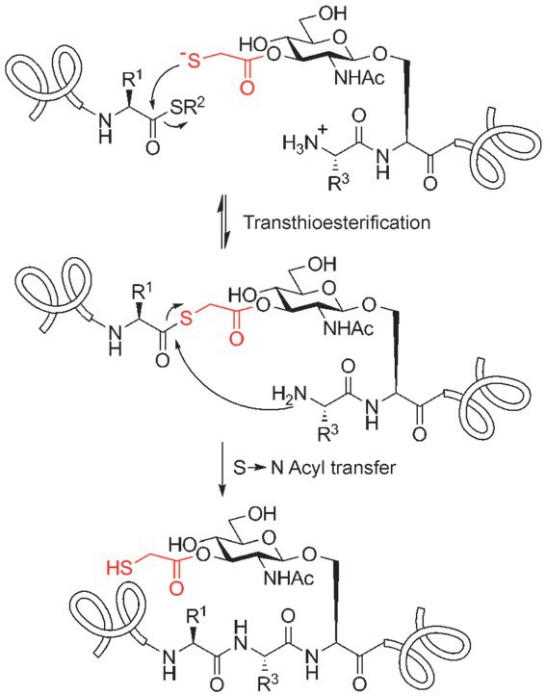
[*] S. Ficht,[+] R. J. Payne,[+] Prof. C.-H. Wong
The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
Fax: (+1) 858-748-2409
E-mail: wong@scripps.edu
and
Academia Sinica
Taipei (Taiwan)
Dr. A. Brik
Department of Chemistry
Ben Gurion University
Beer Sheva (Israel)

[+] S. Ficht and R. J. Payne contributed equally.

[**] This work was supported by the NIH and the Skaggs Institute for Chemical Biology. S.F. is grateful to the Deutsche Akademische Austauschdienst (DAAD) for a postdoctoral fellowship. R.J.P. is grateful for funding provided by the Lindemann Trust Fellowship.

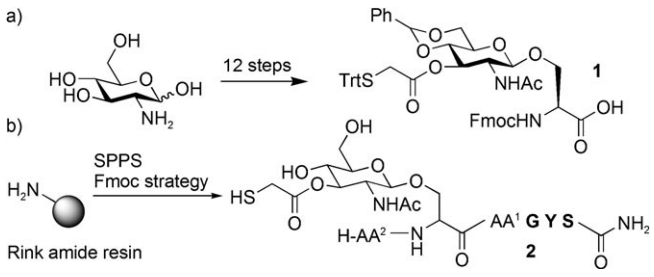
Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

The first step would involve nucleophilic attack of the thioester by the thiol functionality of the auxiliary leading to a new thioester. This would be followed by an intramolecular S→N acyl transfer forming the native peptide backbone. Subsequent auxiliary removal would yield the native glycopeptide without modification of the unprotected peptide backbone.



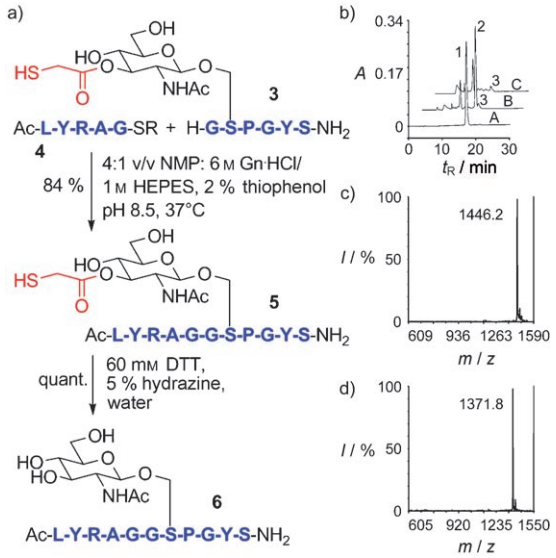
Scheme 1. Proposed mechanism for the second-generation sugar-assisted ligation mediated by a base-labile auxiliary. R^1, R^3: amino acid side chains; R^2: phenyl.

The initial phase of the research involved the synthesis of monomer **1**, which was essential for solid-phase peptide synthesis (SPPS) of the desired auxiliary-containing glycopeptides. Monomer **1** was synthesized from D-glucosamine in 12 steps (Scheme 2a, see the Supporting Information). Synthesis of the desired glycopeptide series **2** was achieved following the Fmoc strategy (Scheme 2b, see the Supporting Information). To assess the effect of different residues at the ligation junction, glycine, histidine, and aspartic acid were coupled as N-terminal amino acids. A number of peptide thioesters were also required for these studies and were synthesized by SPPS following the Boc strategy (see the Supporting Information).



Scheme 2. a) Synthesis of glycosyl amino acid building block **1** from D-glucosamine. Trt = trityl. b) SPPS of glycopeptides carrying the mercaptoacetate auxiliary. AA^1: Pro or Cys; AA^2: Gly, Asp, or His.

As an initial screen to determine the viability of the strategy, cysteine-free glycopeptide **3** and peptide thioester **4** were combined and dissolved in the standard ligation buffer at 10 mM concentration (6 M guanidine(Gn)·HCl, 100 mM NaH_2PO_4, pH 8.5, 2% thiophenol, 37 °C, 1.5 equiv glycopeptide per equiv peptide thioester).[24] Analysis of the ligation mixture after 24 h revealed the desired ligation product along with significant quantities of hydrolyzed thioester and glycopeptide where the ester linkage of the thiol auxiliary had been hydrolyzed. Modifications of the ligation conditions were therefore sought which would minimize these side products, thereby leading to higher ligation yields. The use of buffers of lower pH (6.0–7.5) led to a significantly reduced ligation rate.[26] We therefore proposed the use of a cosolvent at a higher pH to minimize hydrolysis of the starting materials. The most important requirement, the ability to dissolve entire glycoproteins, was fulfilled by a mixture of *N*-methyl pyrrolidinone and guanidine/HEPES buffer (4:1 v/v NMP:6 M Gn·HCl/1 M HEPES, 2% thiophenol, pH 8.5) in which ribonuclease A was found to be clearly soluble at a concentration of 5 mM. When glycopeptide **3** and thioester **4** were reacted under these conditions the desired ligation product **5** was isolated in 84% yield after a reaction time of 60 h (Scheme 3). These conditions prevented both the hydrolysis of the thioester and the ester-bound thiol auxiliary.



Scheme 3. a) Second-generation SAL between glycopeptide **3** and peptide thioester **4** and subsequent auxiliary removal (R = $(CH_2)_2CONH_2$). b) HPLC traces of the auxiliary removal (λ = 280 nm). Trace A: **5**, trace B: the crude reaction mixture of **5** in a solution of DTT containing 5% hydrazine after 30 min, and trace C: a solution of DTT containing 5% hydrazine. c) MALDI-TOF/MS analysis of peak 1 ($[M+H]^+$ *m/z* 1446.6, **5**). d) MALDI-TOF/MS analysis of peak 2 ($[M+H]^+$ *m/z* 1372.5, **6**). Matrix: α-cyano-4-hydroxy cinnamic acid.

With the successful implementation of the second-generation SAL in hand, the next stage of the research involved cleavage of the thiol handle from the sugar moiety. To this end, ligation product **5** was dissolved in an aqueous solution of dithiothreitol (DTT, 60 mM) containing 5% hydrazine. Scheme 3b shows the HPLC trace of starting material **5** (16.9 min, trace A, peak a). After 30 min, a new peak was observed (15.7 min, trace B, peak b) and was verified to be the auxiliary-free ligation glycopeptide **6** by mass spectrometry (Scheme 3d). The small peak c was shown not to correspond to unreacted ligation product **5**, but rather is part of the DTT/hydrazine mixture as shown by blank trace C. Encouragingly, this study showed that the auxiliary hydrazinolysis represents a peak-to-peak conversion and the yield of the isolated auxiliary-free native glycopeptide was quantitative.

To examine whether the ligation reaction tolerates an additional thiol functionality, a glycopeptide bearing a cysteine in its sequence and the amino acid glycine on the N terminus was reacted with a C-terminal glycine peptide thioester (Table 1, entry 2). After a reaction time of 60 h, the

Table 1: Scope of the second-generation SAL (R = $(CH_2)_2CONH_2$).

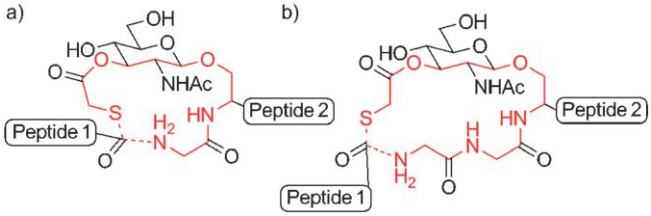
Entry	Thioester AA^1	Glycopeptide AA^2 (AA^3)	Ligation junction AA^1-AA^2	Yield [%]
1	Gly	Gly (Pro)	Gly-Gly	84[a]
2	Gly	Gly (Cys)	Gly-Gly	69[a]
3	His	Gly (Cys)	His-Gly	62[a]
4	Ala	Gly (Cys)	Ala-Gly	28[b]
5	Phe	Gly (Cys)	Phe-Gly	41[b]
6	Gly	His (Cys)	Gly-His	28[b]
7	His	His (Cys)	His-His	27[b]
8	Ala	His (Cys)	Ala-His	40[b]
9	Tyr	His (Cys)	Tyr-His	32[b]
10	Gly	Asp (Cys)	Gly-Asp	38[b]
11	His	Asp (Cys)	His-Asp	22[b]
12	Ala	Asp (Cys)	Ala-Asp	59[b]
13	Tyr	Asp (Cys)	Tyr-Asp	40[b]

[a] After a reaction time of 60 h. [b] After a reaction time of 4 d.

product was isolated in 69% yield. Although this represents a slightly lower yield than in the cysteine-free case, the second-generation SAL clearly tolerates additional thiol groups. Upon removal of the auxiliary under the conditions identical to those described in Scheme 3, the desired cysteine-containing glycopeptide was obtained quantitatively.

To study the effect of other amino acids at the ligation junction, the glycopeptide–peptide thioester pairs shown in entries 3–13 of Table 1 were examined. The highest yields were obtained when the glycopeptide bore an N-terminal glycine (Table 1, entries 1–5). An N-terminal aspartic acid also gave reasonable yields (Table 1, entries 10–13), especially when reacted with the sterically encumbered amino acid alanine (59% yield, entry 12). Histidine on the N terminus of the glycopeptide led to 27–40% ligation yields (Table 1, entries 6–9). Removal of the thiol handle was performed for all ligation products under the conditions shown in Scheme 3 and provided the native glycopeptides in yields in excess of 95% (see the Supporting Information).

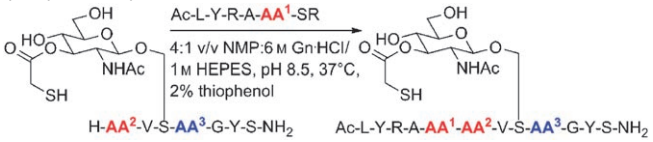
When screening libraries of glycoprotein sequences, we found that a large number of glycoproteins bear sterically hindered residues on the N-terminal side of the sugar-carrying amino acid. As with NCL and cysteine-free NCL, the presence of these sterically encumbered residues, in particular valine, leucine, isoleucine, and proline, would complicate the synthesis of these targets using second-generation SAL. Our concept to circumvent this problem was to extend the glycopeptide by an extra amino acid. The first step of a ligation involving such a glycopeptide (the transthioesterification reaction) would be identical to that depicted in Scheme 1. However, the following intramolecular S→N acyl shift would proceed not through a 15-membered-ring transition state (Scheme 4a) as for the second-generation SAL,



Scheme 4. Proposed transition states of a) the second-generation SAL (15-membered-ring transition state) and b) the extended second-generation SAL (18-membered-ring transition state).

but rather an 18-membered-ring transition state (Scheme 4b). Precedence for successful S→N acyl shifts proceeding through transition states of comparable ring sizes encouraged us to pursue this strategy.[27,28] The glycopeptides shown in Table 2 were synthesized in an analogous fashion to those in Scheme 2, and the ligations were conducted under the mixed-solvent conditions as described above. To our delight, ligations proceeded efficiently for all glycopeptides tested. Glycopeptides bearing the amino acid glycine on the N terminus were ligated in good yields (48–68%, Table 2, entries 1–7) even when sterically hindered peptide thioesters were used as reaction partners (Table 2, entries 2, 5, and 7). Again, the presence of a cysteine in the glycopeptide sequence did not show adverse effects on the ligation reaction. Glycopeptides carrying an N-terminal aspartic acid residue also reacted in satisfactory yields (Table 2, entries 12–15). Ligations with glycopeptides bearing an N-terminal histidine were slightly less efficient, especially when reacted with peptide thioesters bearing sterically hindered amino acids on the C terminus (Table 2, entries 8–11). To determine whether the internal cysteine residue could concomitantly facilitate the ligation reaction by an NCL mechanism, the auxiliary of the glycopeptide (used in entries 4–7) was cleaved and the product submitted to the ligation conditions. Interestingly, a slight background reaction was observed, details of which will be explored in future research.

Table 2: Scope of the extended second-generation SAL (R = $(CH_2)_2CONH_2$).



Entry	Thioester AA^1	Glycopeptide AA^2 (AA^3)	Ligation junction AA^1-AA^2	Yield [%]
1	Gly	Gly (Pro)	Gly-Gly	68[a]
2	Ala	Gly (Pro)	Ala-Gly	65[b]
3	His	Gly (Pro)	His-Gly	58[b]
4	Gly	Gly (Cys)	Gly-Gly	66[c]
5	Ala	Gly (Cys)	Ala-Gly	48[b]
6	His	Gly (Cys)	His-Gly	60[d]
7	Tyr	Gly (Cys)	Tyr-Gly	45[b]
8	Gly	His (Cys)	Gly-His	55[e]
9	Ala	His (Cys)	Ala-His	23[b]
10	His	His (Cys)	His-His	40[b]
11	Tyr	His (Cys)	Tyr-His	39[b]
12	Gly	Asp (Cys)	Gly-Asp	81[c]
13	Ala	Asp (Cys)	Ala-Asp	50[f]
14	His	Asp (Cys)	His-Asp	60[f]
15	Tyr	Asp (Cys)	Tyr-Asp	37[f]

[a] After a reaction time of 12 h. [b] After a reaction time of 4 d. [c] After a reaction time of 24 h. [d] After a reaction time of 36 h. [e] After a reaction time of 56 h. [f] After a reaction time of 3 d.

The extended glycopeptide ligations appeared to be more facile than their unextended counterparts. To get an indication of the relative rates of the originally presented SAL,[24] second-generation SAL, and the extended second-generation SAL, the reaction kinetics were determined by analyzing aliquots of reaction mixtures combining glycopeptides **7**, **8**, and **3** and glycine peptide thioester **4** (Figure 1). The second-generation SAL required approximately 24 h to reach a yield of 50%. The same level of completion was obtained after only 12 h using the extended glycopeptide **8**. The previously reported SAL had a $t_{1/2}$ of 9 h, slightly faster than both the second-generation SAL and the extended second-generation SAL.

Encouraged by these results, we wondered if further N-terminal extensions of the glycopeptide would be possible. To this end, four glycopeptides carrying up to five amino acids on the N terminus of the sugar-bound serine were synthesized (Table 3). These were submitted to the mixed-solvent ligation conditions. Notably, reactions would now proceed through 21- to 28-membered-ring transition states during the S→N acyl shifts. Remarkably, yields between 52 and 81% were obtained for these ligations. This finding further extends the synthetic accessibility of glycopeptides and potentially glycoproteins using this method. Significantly, this extension to the SAL methodology enables one to readily find a ligatable junction by extending in the N-terminal direction from the glycosylated amino acid residue. Future studies will focus on how far the glycopeptide can be extended using this methodology.

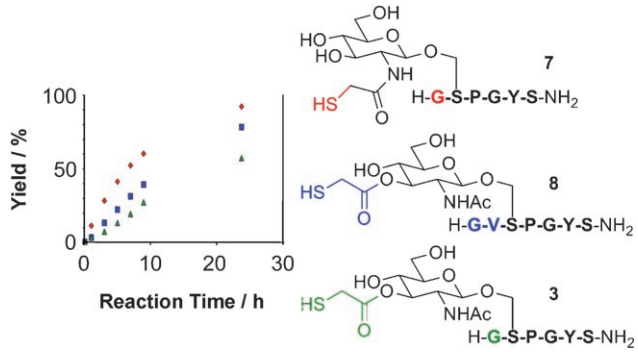
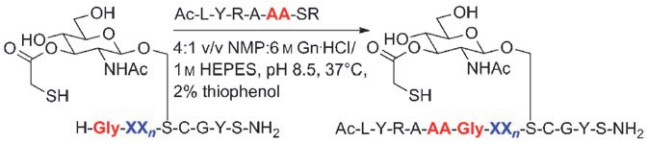


Figure 1. Early slope of the SAL (◆), extended second-generation SAL (■) and the second-generation SAL (▲).

Table 3: Scope of the extended second-generation SAL bearing multiple amino acids on the N terminus (R = $(CH_2)_2CONH_2$).



Entry	Thioester AA	Glycopeptide XX_n	Ligation junction AA-Gly	Yield [%]
1	Gly	Val Leu	Gly-Gly	62[a]
2	His	Val Leu	His-Gly	81[b]
3	Gly	Gly Val Leu	Gly-Gly	56[a]
4	Gly	Gly Gly Val Leu	Gly-Gly	52[c]

[a] After a reaction time of 24 h. [b] After a reaction time of 36 h. [c] After a reaction time of 42 h.

In summary, we have reported an extremely effective method for the assembly of cysteine-containing and cysteine-free glycopeptides, the second-generation sugar-assisted ligation (SAL). The ability to "walk" along the peptide backbone in the N-terminal direction of the glycosylated amino acid to find a suitable ligation site clearly expands the number of glycopeptides accessible by this method. The ligation reactions and conditions for the auxiliary removal are chemoselective, and thus, protection of the amino acid side chains is unnecessary. In addition to application in total chemical synthesis, this method opens opportunities for semisynthetic approaches toward the synthesis of large glycoproteins by using peptide thioesters expressed by means of the intein technique. We envisage that further elaboration of the bridgehead glycan can be achieved using enzymatic transfer. Applications of this method for the synthesis of a homogeneous glycoprotein are currently in progress, and details will be reported in due course.

Received: February 6, 2007
Published online: July 2, 2007

Keywords: acyl transfer · auxiliaries · carbohydrates · glycopeptides · ligation

[1] R. G. Spiro, *Glycobiology* **2002**, *12*, 43R–56R.
[2] R. Apweiler, H. Hermjakob, N. Sharon, *Biochim. Biophys. Acta Gen. Subj.* **1999**, *1473*, 4–8.

[3] R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683 – 720.
[4] A. Varki, *Glycobiology* **1993**, *3*, 97 – 130.
[5] L. H. Miller, M. F. Good, G. Milon, *Science* **1994**, *264*, 1878 – 1883.
[6] D. H. Dube, C. R. Bertozzi, *Nat. Rev. Drug Discovery* **2005**, *4*, 477 – 488.
[7] B. G. Davis, *Chem. Rev.* **2002**, *102*, 579 – 601.
[8] M. R. Pratt, C. R. Bertozzi, *Chem. Soc. Rev.* **2005**, *34*, 58 – 68.
[9] L. Liu, C. S. Bennett, C. H. Wong, *Chem. Commun.* **2006**, 21 – 33.
[10] A. Brik, S. Ficht, C. H. Wong, *Curr. Opin. Chem. Biol.* **2006**, *10*, 638 – 644.
[11] T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang, H. Lau, *Justus Liebigs Ann. Chem.* **1953**, *583*, 129 – 149.
[12] P. E. Dawson, T. W. Muir, I. Clarklewis, S. B. H. Kent, *Science* **1994**, *266*, 776 – 779.
[13] P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923 – 960.
[14] Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman, C. R. Bertozzi, *J. Am. Chem. Soc.* **1999**, *121*, 11684 – 11689.
[15] S. Mezzato, M. Schaffrath, C. Unverzagt, *Angew. Chem.* **2005**, *117*, 1677 – 1681; *Angew. Chem. Int. Ed.* **2005**, *44*, 1650 – 1654.
[16] L. A. Marcaurelle, L. S. Mizoue, J. Wilken, L. Oldham, S. B. H. Kent, T. M. Handel, C. R. Bertozzi, *Chem. Eur. J.* **2001**, *7*, 1129 – 1132.
[17] C. P. R. Hackenberger, C. T. Friel, S. E. Radford, B. Imperiali, *J. Am. Chem. Soc.* **2005**, *127*, 12882 – 12889.
[18] Very recently it has been reported that thiol side chains can be desulfurized to give an alanine after ligation in the presence of other (Acm-protected) cysteines. Thereby, the restrictions for the N terminus are extended to cysteine and alanine. B. L. Pentelute, S. B. H. Kent, *Org. Lett.* **2007**, *9*, 687 – 690.
[19] J. Offer, C. N. C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642 – 4646.
[20] D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent, P. Botti, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554 – 6559.
[21] D. Macmillan, D. W. Anderson, *Org. Lett.* **2004**, *6*, 4659 – 4662.
[22] B. Wu, J. H. Chen, J. D. Warren, G. Chen, Z. H. Hua, S. J. Danishefsky, *Angew. Chem.* **2006**, *118*, 4222 – 4231; *Angew. Chem. Int. Ed.* **2006**, *45*, 4116 – 4125.
[23] A. Brik, S. Ficht, Y. Y. Yang, C. S. Bennett, C. H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 15026 – 15033.
[24] A. Brik, Y. Y. Yang, S. Ficht, C. H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 5626 – 5627.
[25] L. C. Hsieh-Wilson, *Nature* **2007**, *445*, 31 – 33.
[26] C. Dose, O. Seitz, *Org. Biomol. Chem.* **2004**, *2*, 59 – 65.
[27] S. Sasaki, M. Shionoya, K. Koga, *J. Am. Chem. Soc.* **1985**, *107*, 3371 – 3372.
[28] C. Gennari, F. Molinari, U. Piarulli, M. Bartoletti, *Tetrahedron* **1990**, *46*, 7289 – 7300.

