Sugar-Assisted Ligation for the Synthesis of Glycopeptides

ARTICLE *in* CHEMISTRY · SEPTEMBER 2007

Impact Factor: 5.73 · DOI: 10.1002/chem.200700330 · Source: PubMed

2 AUTHORS, INCLUDING:

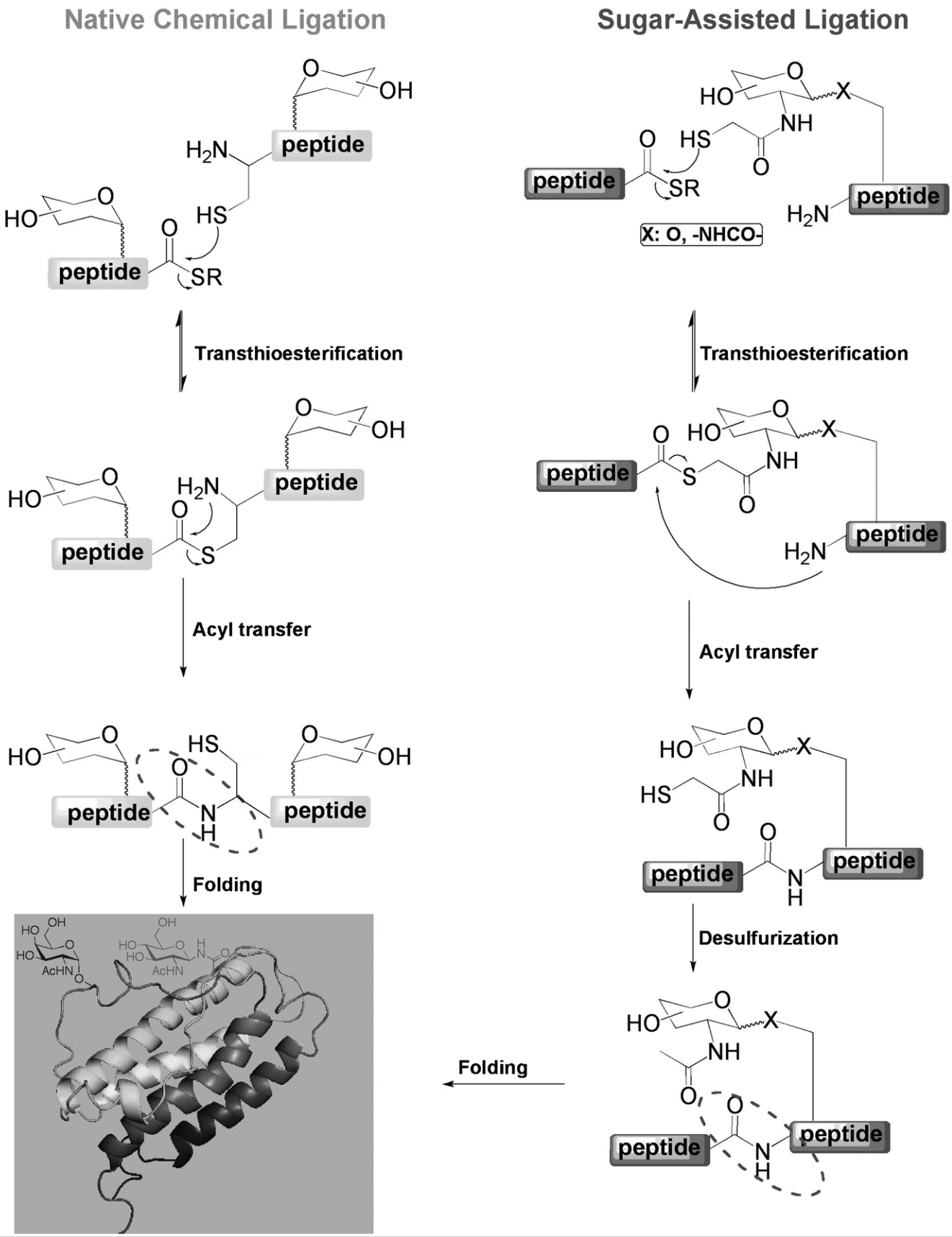
Ashraf Brik

Ben-Gurion University of the Negev

93 PUBLICATIONS **2,682** CITATIONS

DOI: 10.1002/chem.200700330

Glycopeptide and Glycoprotein Synthesis



Sugar-Assisted Ligation for the Synthesis of Glycopeptides

Ashraf Brik*[a] and Chi-Huey Wong*[b]

Abstract: Chemical synthesis of glycoproteins from readily available materials is a powerful method for obtaining a pure product with full control of its atomic structure. Sugar-assisted ligation (SAL) is an emerging approach that allows the synthesis of a large glycopeptide from two unprotected fragments. Contrary to other ligation methods that are limited to the use of a cysteine residue or depend on external auxiliary, SAL takes advantage of the existing sugars in glycopeptides to promote proximity between the two peptides to facilitate an amide bond formation.

Keywords: amine capture strategy · chemical ligation · glycopeptides · glycoproteins · removable auxiliarys

Introduction

Glycoproteins play an important role in several important biological processes, including protein folding, secretion, cell targeting and adhesion, stability in circulation, signal transduction, and many intercellular processes.[1] Moreover, in the drug-discovery arena, glycoproteins are one of the most important biotechnology products. As a result, the chemical synthesis of glycoproteins with a well-defined structure at the atomic level has been the driving force for the development of several synthetic strategies.[2] By their nature, glycoproteins are perhaps the most complex natural product, not only because of their large size, but also due to the complexity of the glycan moiety that decorate the polypeptide scaffold. Protein glycosylation is a complex posttranslational modification[3] that often leads to a heterogeneous glycoprotein mixture. Therefore, to understand the discrete effect of the glycan portion on the protein structure and function becomes a formidable task.

The synthetic scheme of the targeted glycoprotein is challenged by the synthesis of the peptidic component and the glycan constituents. The synthesis of the glycan structure is a tedious task and often requires heroic efforts.[4] On the other hand, the assembly of the peptidic scaffold is also very challenging considering the large size of the chain as well as the limitations that are usually imposed by the glycan on the assembly process. Solid-phase peptide synthesis (SPPS), in most cases, is limited to a peptide chain of $\approx$50 amino acids,[5] which corresponds to a small number of proteins and glycoproteins. As a result, ligation methods have been sought to produce the desired sequence from smaller peptide fragments. In this account, we highlight some of the recent developments in glycopeptide synthesis based on ligation methods with emphasis on a new concept that utilizes the sugar moiety to assist the ligation to form the desired glycopeptide.

Entropic-Activation-Induced Intramolecular Acyl Transfer

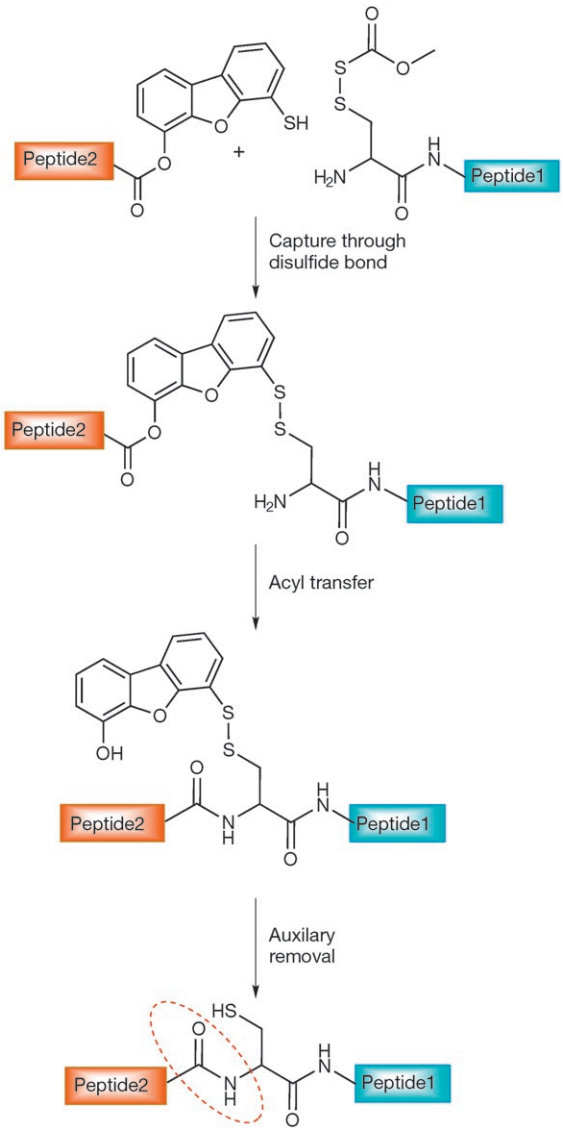
Amide bond formation is one of the most common reactions used by nature. The peptide assembly process in natural systems is carried out by the ribosomal and nonribosomal peptide synthetases (NRPS). The synthesis is powered by using entropic activation through positioning the C-terminal acyl group (mildly activated as ester) of the growing peptide chain into close proximity to the α-amino group on the aminoacyl *t*RNA. Chemists also take advantage of the entropic-activation/proximity effect to achieve various chemical transformations. For example, Wieland,[6] Benkovic,[7] and Brenner[8] laid the early concepts in the field of peptide

[a] Dr. A. Brik
Department of Chemistry, Ben Gurion University
Beer Sheva, 84105 (Israel)
Fax: (+972) 8-647-2943
E-mail: abrik@bgu.ac.il

[b] Prof. Dr. C.-H. Wong
Department of Chemistry
and the Skaggs Institute for Chemical Biology
The Scripps Research Institute, 10550 North Torrey Pines Road
La Jolla, California, 92037 (USA)
Fax: (+1) 858-784-2409
E-mail: wong@scripps.edu

bond formation by using entropic activation to induce intramolecular acyl transfer.

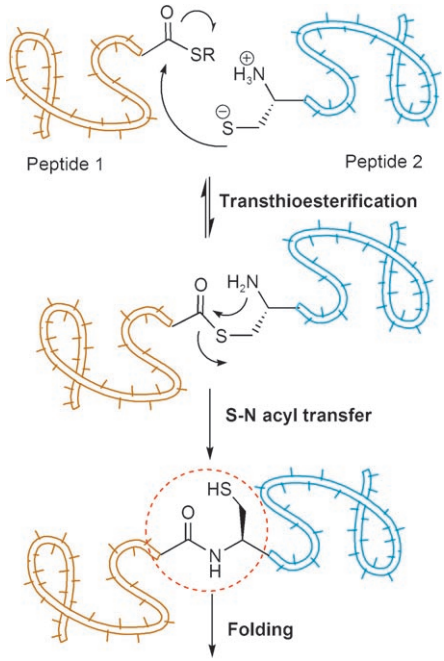
Research done by Kemp and co-workers on peptide ligation by intramolecular aminoacyl transfer, also known as "the prior-capture strategy", brought the above concepts to a certain level of practicality in the area of peptide assembly.[9] In these studies, a rigid scaffold such as the dibenzofuran molecule (Scheme 1) was used to bring the two peptides bearing the electrophilic center (ester) and nucleophilic amine to a close proximity to promote O→N acyl transfer (calculated effective molarity for the dibenzofuran system is 4.6 M). This strategy inspired the development of several other scaffolds to facilitate amide bond formation and was eventually applied in the synthesis of large peptides.[10]



Scheme 1. Amine-capture strategy using the dibenzofuran scaffold to induce proximity and facilitate amide bond formation.

Native Chemical Ligation

The breakthrough in the chemical synthesis of protein occurred in 1994 when Kent and co-workers introduced the native chemical ligation (NCL) method.[11] The ingenuity of this approach is the lack of an external scaffold to promote amide bond formation at the ligation site. The method relies on the thiol side chain of a cysteine residue to bring the N-terminal amine of the first peptide to a close proximity to the thioester C-terminal peptide. Once the intermediate is formed, it spontaneously rearranges through a five membered-ring intermediate to form the native amide bond (Scheme 2). In NCL the rate-determining step is the transthioesterification step, contrary to all the other amine capture strategies. The size of the ring in the S→N acyl transfer intermediate and the high nucleophilicity of the thiolate, in the presence of the protonated N-terminal amine, contribute to fast intermolecular thioester formation followed by an intramolecular aminolysis to form the amide bond.



Scheme 2. Proposed mechanism for NCL. After completing the polypeptide assembly a folding step follows to generate the active protein.

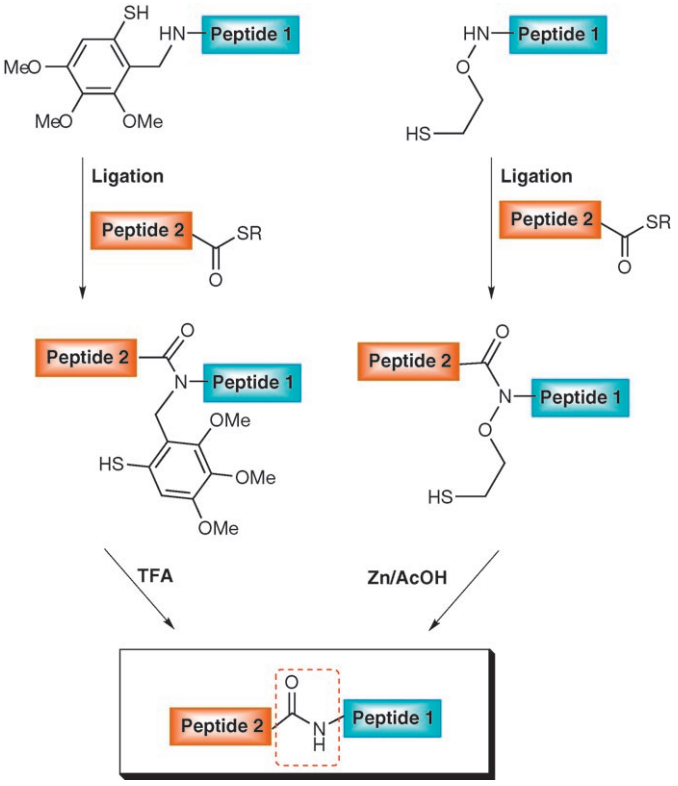
Currently, NCL is viewed as one of the most useful chemical methods enabling protein manipulation for protein studies. Yet, the technology suffers from a few restrictions that hamper its use in a variety of protein systems. The size of the targeted protein could restrict the use of NCL particularly when only chemical means are applied. However, thanks to protein-expressed ligation (EPL), NCL could be applied to larger proteins. EPL is semisynthetic protein approach that allows the thioester fragment or/and the peptide bearing the N-terminal cysteine to be expressed in biological

systems, while the actual ligation step is still a chemical process.[12]

Extending NCL by Using Removable Auxiliaries

A major limitation of NCL/EPL is the reliance of the ligation step on the cysteine residue to promote the acyl transfer. Although a number of proteins contain native cysteine residues in synthetically useful positions (ideally at every 30–40 amino acids of the protein sequence), the majority of proteins do not have the cysteine residue (cysteine is the second least common of the 20 amino acids in proteins). Moreover, in many cases the cysteine location within the peptide sequence is not beneficial for the ligation scheme.

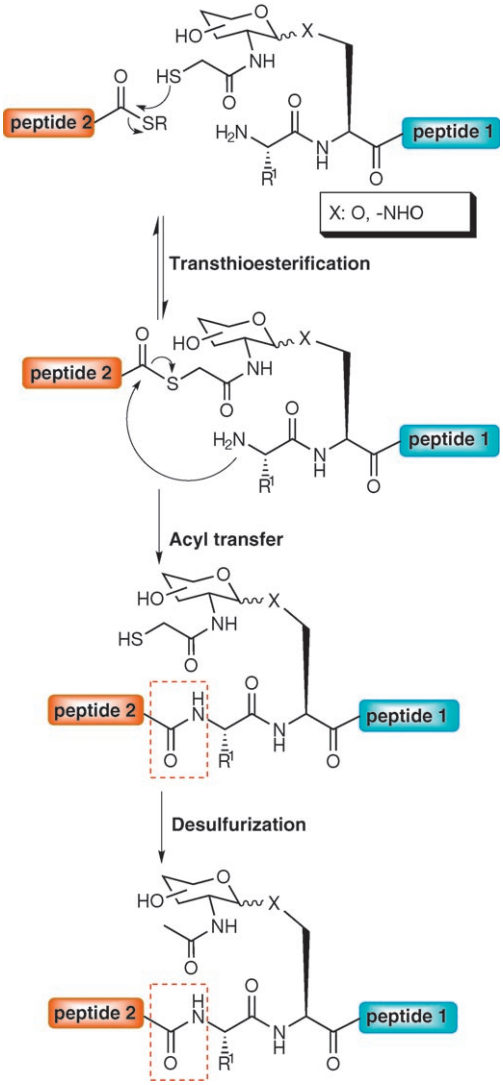
To overcome the requirement of a cysteine residue at the ligation site, several strategies were developed. One such approach is the use of removable auxiliaries to mimic the cysteine function at the ligation site. Peptides with N-linked auxiliaries, such as 1-phenyl-ethanethiol,[13] 2-mercaptoethyl,[14] and 2-mercaptobenzyl,[15] have been investigated and successfully applied to the synthesis of large peptides and complex glycopeptides (Scheme 3).[16] In these elegant examples, the attachment of the auxiliary to the N terminus of the peptide generates a secondary amine that is engaged in the S→N acyl transfer. Due to the increased steric hindrance, only ligation at Gly–Gly or Gly–AA (AA represent unhindered amino acids) junctions delivered the desired products, limiting their scope in the synthesis of proteins and glycoproteins.[17]



Scheme 3. 2-Mercaptobenzyl and ethanethiol auxiliaries were developed to assist NCL.

Sugar-Assisted Ligation

Recently, our group introduced a new ligation approach named sugar-assisted ligation (SAL) for a non-cysteinyl ligation junction. The SAL method was successfully applied to the synthesis of β-O-[18] and N-linked[19] glycopeptides. Our method uses a similar principle of NCL and auxiliary assisted ligation in terms of the entropic activation/proximity effect. However, SAL uses the thiol that is placed on the sugar rather than on the side chain of the amino acid, for example, cysteine. In SAL, instead of anchoring an auxiliary to the N terminus of the peptide, which has to be removed once the ligation is completed, we take advantage of the already existing sugar by slightly modifying its acetamido group on the C2 position to facilitate the ligation (Scheme 4). Apparently, the sugar moiety imposes its re-

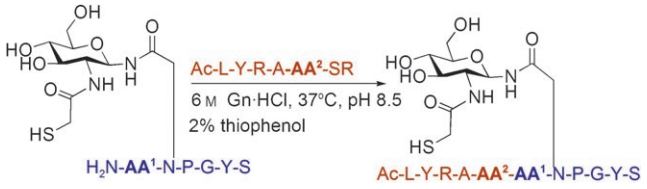


Scheme 4. Proposed mechanism of SAL.

stricted conformation on the reacting groups of the thioester intermediate, thus acting as a rigid scaffold to facilitate the S→N rearrangement via 14- (O-linked glycopeptide) or 15-membered ring intermediates (N-linked glycopeptide).

Notably, the ring size of the intermediate in SAL is similar to the one in the dibenzofuran template reported by Kemp and co-workers (Scheme 1). Despite being much larger than the size of the intermediate produced by the 1-phenyl-2-mercaptoethyl, 2-mercaptobenzyl, and ethanethiol auxiliaries (six-membered ring), the efficiency of SAL in terms of rate and the sequence tolerance at the ligation junction is superior. In the previous auxiliaries, the N-terminal nucleophile is secondary; however, in our strategy the nucleophile is a primary amine. This leads to a less hindered transition state, which may explain the tolerance of the ligation junction to a variety of amino acids (Table 1). Another

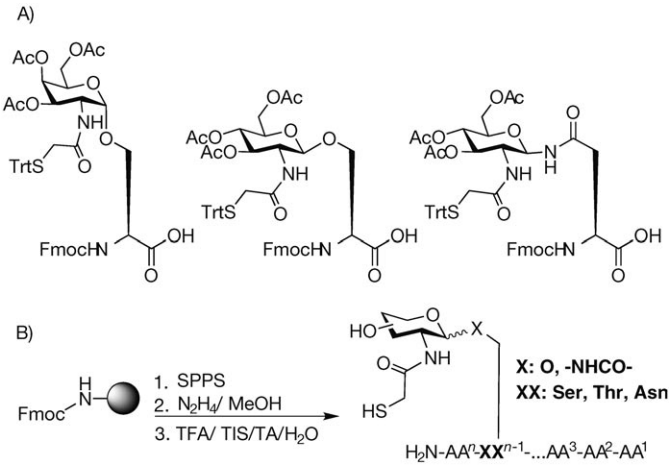
Table 1. Scope of SAL in N-linked glycopeptide synthesis. Isolated yields for the ligation reactions were 40–60%.



	-AA^2-	-AA^1-	-AA^2-AA^1-	t [h]
1	Gly	Gly	Gly-Gly	12
2	Gly	Ala	Gly-Ala	24
3	Gly	Val	Gly-Val	48
4	Ala	Ala	Ala-Ala	>48
5	Gly	Asn	Gly-Asn	36
6	Gly	Asp	Gly-Asp	12
7	Ala	Asp	Ala-Asp	36
8	His	His	His-His	8
9	Gly	His	Gly-His	6

important observation in SAL is the effect of the basicity of some of the side chain amino acids on the ligation rate (Table 1). For example, the side chain of asparagine is as hindered as the aspartic acid, but the rate of the ligation is three times slower. This can be explained by the ability of the carboxylate side chain serving as a general base in the reaction pathway.

Once the ligation reaction is completed, the thiol handle can be removed by desulfurization[20] to regenerate the unmodified sugar. Alternatively, the modified sugar can be further elaborated with glycosyl transferases to extend its glycan structure, or alkylated with labeling reagents such as fluorescent dyes. Interestingly, in the three most prevalent occurring glycopeptides; N-, β-O-, and α-O-linked, the sugar that is directly attached to the peptide is equipped with the *N*-acetyl moiety at the C2 position (Scheme 5). This would allow the introduction of the thiol handle, regardless of the glycoform type, to assist the ligation. Our recent work has demonstrated that the thiol handle can be attached to the sugar oxygen groups, allowing for the ligation of glycopeptides containing sugars without a nitrogen group. Since our approach permits the ligation at difficult junctions, the method could be applied to the synthesis of the large peptides or glycopeptides through enzymatic removal or addition of the sugar moiety.[19]



Scheme 5. A) Building blocks for glycopeptide synthesis that are used in SAL. B) Fmoc solid-phase peptide synthesis of glycopeptide used in SAL. After completing the peptide assembly, the acetate protecting groups on the sugar are removed using N_2H_2/MeOH mixture (1:6). Final deprotection and cleavage from the solid support are achieved by treating the peptide resin with TFA/trisopropylsilane/H_2O (85:5:5) mixture. The final glycopeptide is purified on the reverse phase HPLC column.

A few challenges remain in SAL that have to be encountered before it can be applied broadly in the context of glycoprotein synthesis. These include:

1) Carrying SAL followed by desulfurization in glycopeptides that contain cysteine residues. In principle, finding orthogonal protecting groups for the cysteine residue could circumvent this problem.
2) The effect of attaching glycan unit(s) to the sugar containing the thiol handle on the ligation rate.
3) The effect of the size of glycopeptides on the ligation efficiency. The peptide length has not been a problem in NCL and we expect the same in SAL.

We are currently investigating these presented challenges and the results will be reported in a near future.

Acknowledgements

We thank the NIH, the Skaggs Institute for Chemical Biology (C.H.W.) and The Marc Rich Foundation for financial support (A.B.).

[1] R. W. Dwek, *Chem. Rev.* **1996**, *96*, 683–720.
[2] a) A. Brik, S. Ficht, C.-H. Wong, *Curr. Opin. Chem. Biol.* **2006**, *10*, 1–7; b) M. R. Pratt, C. R. Bertozzi, *Chem. Soc. Rev.* **2005**, *34*, 58–68; c) B. G. Davis, *Chem. Rev.* **2002**, *102*, 579–602; d) O. Seitz, *ChemBioChem* **2000**, *1*, 214–246.
[3] a) R. G. Spiro, *Glycobiology* **2002**, *12*, 43–56; b) A. Helenius, M. Aebi, *Science* **2001**, *291*, 2364–2369.

[4] C.-H. Wong, *J. Org. Chem.* **2005**, *70*, 4219–4225.
[5] S. B. H. Kent, *Annu. Rev. Biochem.* **1988**, *57*, 957–989.
[6] T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang, H. Lau, W. Schafer, *Justus Liebigs. Ann. Chem.* **1953**, *583*, 129–149.
[7] T. C. Bruce, S. J. Benkovic, *J. Am. Chem. Soc.* **1963**, *85*, 1–8.
[8] M. Bernner, *Pept. Proc. Eur. Pept. Symp. 8th* **1967**, 1–7.
[9] a) D. S. Kemp, *Biopolymers* **1981**, *20*, 1793–1804; b) R. I. Carey, D. S. Kemp, *J. Org. Chem.* **1993**, *58*, 2216–2222.
[10] D. C. Coltart, *Tetrahedron* **2000**, *56*, 3349–3491.
[11] a) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776–779; b) P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923–960.
[12] T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
[13] L. E. Canne, S. J. Bark, S. B. H. Kent, *J. Am. Chem. Soc.* **1996**, *118*, 5891–5896.
[14] D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent, P. Botti, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554–6559.
[15] J. Offer, C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642–4646.
[16] B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua, S. J. Danishefsky, *Angew. Chem.* **2006**, *118*, 4222–4231; *Angew. Chem. Int. Ed.* **2006**, *45*, 4416–4425.
[17] D. Macmillan, D. W. Anderson, *Org. Lett.* **2004**, *6*, 4659–4662.
[18] A. Brik, Y.-Y. Yang, S. Ficht, C.-H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 5626–5627.
[19] A. Brik, S. Ficht, Y.-Y. Yang, C.-H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 15026–15033.
[20] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526–533.
[21] L. C. Hsieh-Wilson, *Nature* **2007**, *445*, 31–33.

Published online: May 16, 2007

