Glycoprotein Synthesis

DOI: 10.1002/anie.201205038

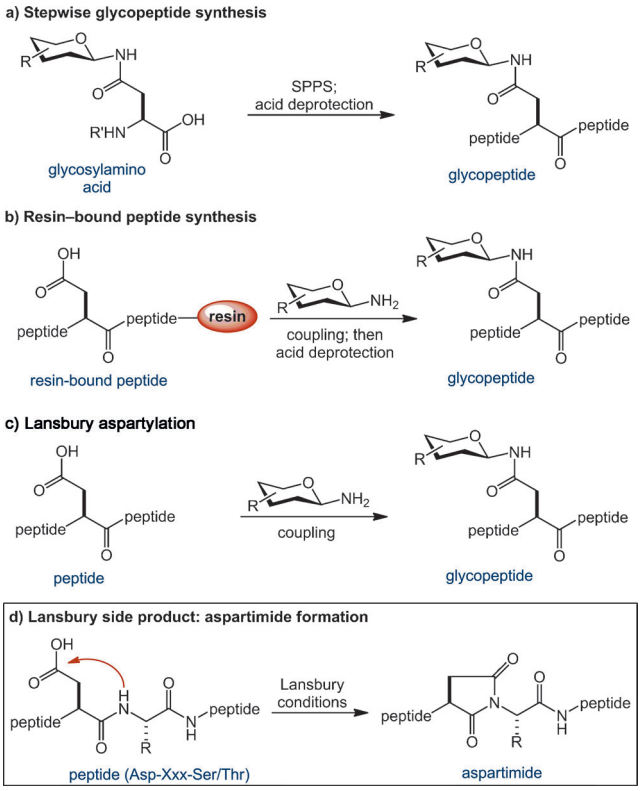
An Advance in the Chemical Synthesis of Homogeneous N-Linked Glycopolypeptides by Convergent Aspartylation**

*Ping Wang, Baptiste Aussedat, Yusufbhai Vohra, and Samuel J. Danishefsky**

Our research group has a longstanding interest in the development of improved methods for the chemical synthesis of structurally complex proteins and glycoproteins.[1] Our efforts started with the study of new departures in oligosaccharide synthesis,[2] as well as in polypeptide synthesis.[1a] Building from the foundations of these key constituents, we have recently turned our attention to merging such subunits to produce glycoproteins.[3] Toward these pursuits, the complex, multiply glycosylated protein, erythropoietin (EPO),[4] has served as a defining synthetic target.[5] Our focusing goal, that is, a total synthesis of homogeneous EPO, has served to prompt a variety of ventures directed to enabling advances in glycoprotein synthesis. We disclose herein a valuable new aspartylation technology, developed in the context of our EPO program, which allows for the highly convergent synthesis of complex glycopeptides from fully elaborated peptide and carbohydrate fragments.

Several general strategies are commonly employed for the assembly of N-linked glycoprotein and glycopeptide targets,[6] though each approach suffers from significant limitations in scope or efficiency. According to the stepwise glycosylamino acid approach (Scheme 1a),[7] the carbohydrate domain is first appended to an 9-fluorenylmethyloxycarbonyl- or *tert*-butoxycarbonyl-protected Asp residue. The resultant glycosylamino acid is subsequently used directly in solid-phase peptide synthesis (SPPS). When the carbohydrate component is complex, the overall efficiency of this linear strategy is significantly compromised by low reaction yields obtained during the glycosylamino acid coupling step and the subsequent elongation of the next amino acid. Moreover, any sialic acid motifs in the glycan must be protected during SPPS. Glycopeptides of up to 20 amino acids in length may be generated through this method.[8]

A second SPPS-based approach offers enhanced convergence. According to the on-resin α-linked glycopeptide strategy[9] (Scheme 1b), the resin-bound peptide, assembled through SPPS, is selectively deprotected to reveal the Asp residue. Coupling of the glycan domain, followed by TFA-mediated cleavage from the resin, delivers the glycopeptide fragment. Two factors serve to mitigate the broad utility of this approach. In certain peptide sequences, particularly those incorporating aggregation-prone sectors, the SPPS-derived resin-bound peptide can suffer from significant impurities. Appendage of a high-value glycan domain to impure peptide substrate thus results in significant loss of precious material and formation of difficultly separable mixtures of glycopeptide product. Moreover, release from the resin delivers glycopeptide presenting a C-terminal carboxylic acid, which must then be converted into an activated thioester functionality prior to subsequent native chemical ligation[10] with a peptide or glycopeptide coupling partner.



Scheme 1. Strategies for aspartylation.

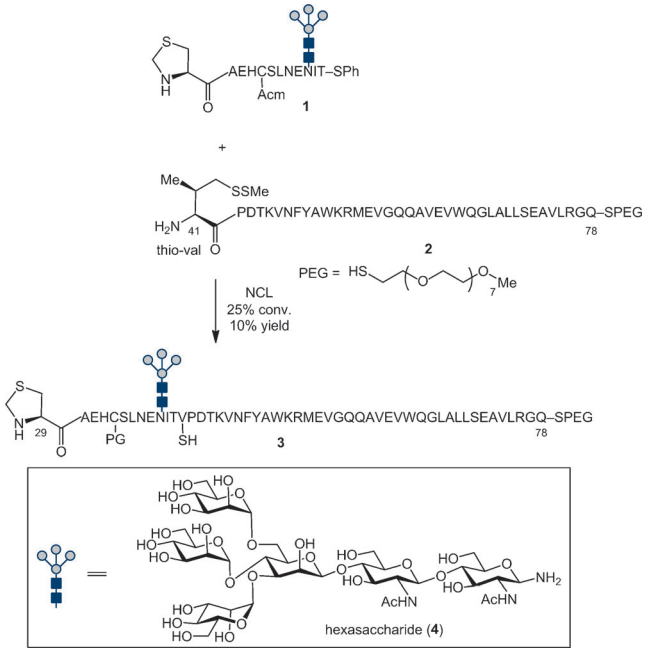
Our research group has been employing[1] the convergent aspartylation approach pioneered by Lansbury and co-workers[11] (Scheme 1c). In this protocol, a moderately sized, partially protected peptide, bearing the free aspartyl residue, is merged with the glycosyl amine to generate a glycopeptide fragment. While useful for producing short peptide fragments, this method is often compromised when long sequences are being joined with glycosylamine. Coupling yields may be badly undermined by peptide decomposition pathways, pre-

[*] Dr. P. Wang, Dr. B. Aussedat, Y. Vohra, Prof. S. J. Danishefsky
Laboratory for Bioorganic Chemistry
Sloan-Kettering Institute for Cancer Research
1275 York Avenue, New York, NY 10065 (USA)
E-mail: s-danishefsky@ski.mskcc.org
Prof. S. J. Danishefsky
Department of Chemistry, Columbia University
3000 Broadway, New York, NY 10027 (USA)

[**] This research was supported by NIH grant HL25848 (S.J.D.).
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201205038.

dominantly by aspartimide formation (Scheme 1 d).[12] Accordingly, our general approach to glycoprotein synthesis has involved assembly of short glycopeptide fragments through Lansbury aspartylation, followed by ligation with a long peptide domain, which may itself contain a glycosylation site. Although not maximally convergent, this strategy is apt to offer ease of use and a high measure of control over glycopeptide purity.

In the context of our ongoing EPO synthesis, we sought to apply this less convergent strategy to the assembly of the EPO(29–78) glycopeptide fragment. Thus, glycopeptide **1**, encompassing the EPO(29–40) peptide sector bearing hexasaccharide **4**,[13] was prepared through Lansbury aspartylation. However, all efforts to couple **1** with peptide **2** [EPO(41–78)] delivered prohibitively low yields of **3** (Scheme 2). We

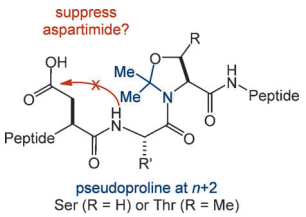


Scheme 2. Attempted synthesis of EPO fragment **3**. Reaction conditions: Gn·HCl (6 M), Na_2HPO_4 (0.1 M), TCEP·HCl (50 mM), pH 6.8. Gn = guanidine, NCL = native chemical ligation, TCEP = tris(2-carboxyethyl)phosphine.

attribute the failure of this attempted aspartylation to the low solubility of peptide **2**, which contains a large hydrophobic domain. Even appendage of a C-terminal PEGylated thioester did not improve the solubility of the peptide to an acceptable level.[14]

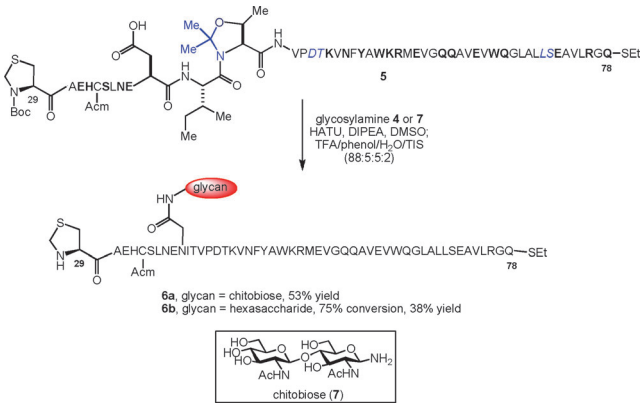
An alternative approach, which would allow us to circumvent issues of peptide solubility and would provide maximum efficiency, envisions merger of the full peptide sequence with the glycan through aspartylation. The successful implementation of this strategy would require conditions under which the problematic aspartimide formation (see Scheme 1 d) could be suppressed or even eliminated.

The consensus sequence for N-glycosylation is Asn-Xaa-Ser/Thr. With this generic structural pattern in mind, we evaluated a potential general solution to the challenge of aspartimide avoidance in the Lansbury coupling. Clearly, placement of a pseudoproline motif, derived from Ser or Thr, at the $(n+1)$ position necessarily blocks aspartimide formation at the n position.[15] Driven by curiosity, rather than by a convincing rationale, we wondered whether temporary installation of the pseudoproline motif at the $(n+2)$ position might also somehow serve to suppress the undesired aspartimide formation (Scheme 3). Because a Ser or Thr residue is universally located at the $(n+2)$ position (relative to Asp) in native protein sequences, success in this regard could bring with it major enabling progress in the chemical synthesis of homogeneous complex glycopolypeptides.



Scheme 3. Pseudoproline dipeptide.

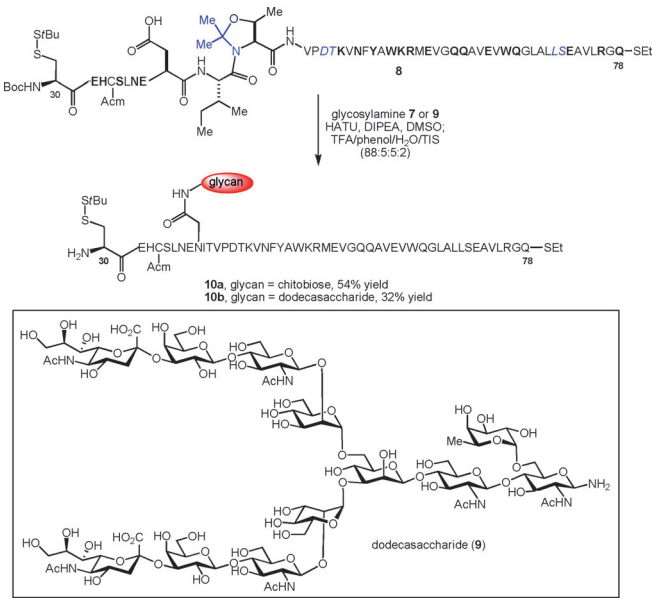
We examined this possibility in the context of the EPO(29–78) segment. Thus, through recourse to SPPS, we prepared partially protected peptide **5**, incorporating several pseudoproline dipeptides[16] (highlighted in blue, Scheme 4), including one at the $(n+2)$ position relative to Asp. In the crucial transformation, peptide **5** readily underwent coupling with chitobiose (**7**) under Lansbury conditions. Subsequent addition of TFA cocktail (TFA/phenol/water/TIS; 88:5:5:2) served to unmask the pseudoproline motifs and remove the peptide protecting groups, thereby delivering the target glycopeptide, **6a**, with quantitative conversion and 53% yield upon isolation. Peptide **5** also underwent one-flask aspartylation/deprotection with the more complex hexasac-



Scheme 4. Synthesis of glycopeptides **6a** and **6b**. Pseudoproline dipeptides are depicted in blue. Amino acids protected with acid-labile protecting groups are shown in bold. Amino acid protecting groups are: E(*t*Bu), H(Trt), S(*t*Bu), N(Trt), K(Boc), Y(*t*Bu), W(Boc), R(Pbf), Q(Trt). Boc = *tert*-butyloxycarbonyl, DIPEA = diisopropylethylamine, DMSO = dimethylsulfoxide, HATU = O-(7-azabenzotriazol-1-yl)-*N*,*N*,*N*',*N*'-tetramethyluronium hexafluorophosphate, TFA = trifluoroacetic acid, TIS = triisopropylsilane, Trt = trityl.

charide, **4**, to generate **6b** with 75% conversion, albeit in somewhat lower yield upon isolation (38%).

In a further demonstration of the potential utility of this transformation, we accomplished the aspartylation of the modified EPO(30–78) peptide, **8**, both with chitobiose and with the challenging dodecasaccharide, **9**.[17] As outlined in Scheme 5, under our one-flask aspartylation/deprotection

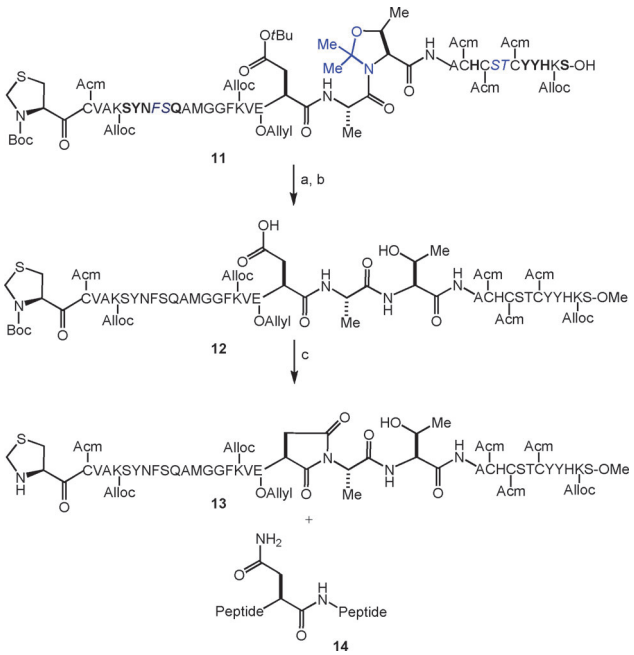


Scheme 5. Synthesis of glycopeptides **10a** and **10b**. Pseudoproline dipeptides are depicted in blue. Amino acids protected with acid-labile protecting groups are shown in bold. Amino acid protecting groups are: E(*t*Bu), H(Trt), S(*t*Bu), N(Trt), K(Boc), Y(*t*Bu), W(Boc), R(Pbf), Q(Trt).

conditions, glycopeptide **10a** was obtained in 54% yield and glycopeptide **10b** was isolated in 32% yield. Notably, the fucose and sialic acid motifs of the dodecasaccharide glycan survived under these conditions, despite the potential sensitivity of these functionalities to acid-mediated decomposition.

Drawing encouragement from these early experiments, we sought to further establish the role of the $(n+2)$ pseudoproline in minimizing nonproductive peptide aspartimide formation. For these studies, we selected as a model peptide scaffold the 34-mer, **12**, incorporating the aspartimide-prone Asp-Ala-Thr sequence. It is of note that the successful preparation of peptide **12**, itself, through SPPS necessitated the installation of the Thr-based pseudoproline motif at the $(n+2)$ position (see **11**). In the absence of pseudoproline protection, serious competition from aspartimide-containing peptide resulted, even at the SPPS stage (see the Supporting Information for details). Apparently, the $(n+2)$ pseudoproline functionality effectively suppresses formation of aspartimide in SPPS, particularly at the stage of 1,8-diazabicyclo-[5.4.0]undec-7-ene (DBU)/piperidine-mediated deprotection.

In our control experiment, we evaluated partially protected **12**, wherein the amine and acid functionalities are masked by allyl and Alloc groups and the $(n+2)$ residue is not protected as a pseudoproline (Scheme 6). As anticipated,



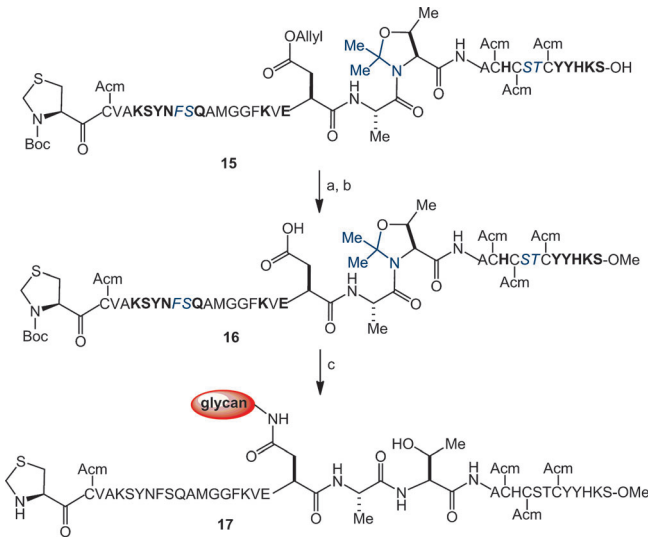
Scheme 6. Attempted aspartylation of **12**. Reaction conditions: a) TMS-diazomethane CH_2Cl_2/MeOH; b) TFA/PhOH/H_2O/TIS, 12% over 3 steps; c) glycosylamine **7**, HATU, DIPEA, DMSO. Pseudoproline dipeptides are depicted in blue. Amino acids protected with acid-labile protecting groups are shown in bold. Amino acid protecting groups are: E(allyl), H(Trt), S(*t*Bu), N(Dmcp), K(Alloc), Y(*t*Bu), R(Pbf), Q(Dmcp). Acm = acetamidomethyl, Alloc = allyloxycarbonyl, Dmcp = dimethylcyclopropyl, Pbf = 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl.

exposure of peptide **12** to chitobiose under Lansbury conditions resulted in little or no glycopeptide formation. The only observed products were those incorporating aspartimide (**13**) and Asn (**14**) in place of the Asp residue. The latter product presumably arises from addition of trace ammonia[18] to the activated aspartate.

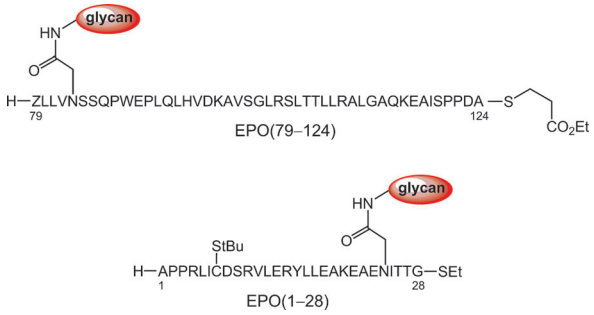
In contrast, the fully protected peptide **15**, incorporating the $(n+2)$ pseudoproline, was prepared through SPPS (Scheme 7). C-Terminal methyl ester formation, followed by palladium-mediated removal of the allyl group, yielded peptide **16**. The latter readily underwent one-flask aspartylation with deprotection to deliver the desired chitobiose-containing glycopeptide **17** in 45% overall yield (starting from SPPS). The results of this comparison study (Scheme 6 vs. 7) clearly reveal the capacity for incorporation of a pseudoproline motif in the $(n+2)$ position to mitigate aspartimide formation at the n position.

In addition to the examples provided above, this protocol enabled the efficient and convergent syntheses of two key glycopeptide fragments en route to homogeneous EPO, namely, EPO(79–124) and EPO(1–28) (Scheme 8).[19] Indeed, in the absence of the pseudoproline functionalities, the precursor peptide domains were not amenable to preparation through SPPS, presumably as a result of the propensity of the peptides to undergo aspartimide formation.

In summary, through incorporation of a pseudoproline motif at the $(n+2)$ Ser or Thr residue, it proved possible to suppress otherwise competitive aspartimide-based peptide



Scheme 7. Synthesis of glycopeptide **17**. a) TMS-Diazomethane CH_2Cl_2/MeOH; b) $[Pd(PPh_3)_4]$, CH_2Cl_2, $PhSiH_3$; c) glycosylamine **7**, HATU, DIPEA, DMSO; then TFA/PhOH/H_2O/TIS, 45% over 4 steps. Pseudoproline dipeptides are depicted in blue. Amino acids protected with acid-labile protecting groups are shown in bold. Amino acid protecting groups are: E(*t*Bu), H(Trt), S(*t*Bu), N(Dmcp), K(Boc), Y(*t*Bu), R(Pbf), Q(Dmcp).



Scheme 8. EPO(79–124) and EPO(1–28).

decomposition pathways. This strategy is also effective for minimizing aspartimide formation during SPPS. Moreover, the aspartylation approach has been successfully applied to convergent syntheses of the EPO glycopeptide fragments. Though the reasons for this observed phenomenon are presently matters of conjecture, its consequences on this field of research are apt to be quite important. Indeed, if generalizable, it could well have solved one of the most serious problems in the assembly of complex glycopolypeptides by chemical means.[20]

Received: June 27, 2012
Revised: July 17, 2012
Published online: September 25, 2012

Keywords: aspartimide · aspartylation · glycopeptides · pseudoproline dipeptide · solid-phase synthesis

[1] For recent reviews from our research group, see: a) C. Kan, S. J. Danishefsky, *Tetrahedron* **2009**, *65*, 9047; b) R. M. Wilson, J. L. Stockdill, X. Wu, X. Li, P. A. Vadola, P. K. Park, P. Wang, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 2888; *Angew. Chem. Int. Ed.* **2012**, *51*, 2834.

[2] S. J. Danishefsky, M. T. Bilodeau, *Angew. Chem.* **1996**, *108*, 1482; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1380.

[3] a) S. Shang, Z. Tan, S. J. Danishefsky, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4297; b) P. Nagorny, N. Sane, B. Fasching, B. Aussedat, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 999; *Angew. Chem. Int. Ed.* **2012**, *51*, 975; c) B. Aussedat, B. Fasching, E. Johnston, N. Sane, P. Nagorny, S. J. Danishefsky, *J. Am. Chem. Soc.* **2012**, *134*, 3532.

[4] a) A. J. Sytkowski, *Erythropoietin*, Wiley-VCH, Weinheim, **2004**; b) P. H. Lai, R. Everett, F. F. Wang, T. Arakawa, E. Goldwasser, *J. Biol. Chem.* **1986**, *261*, 3116.

[5] a) C. Kan, J. D. Trzupek, B. Wu, Q. Wan, G. Chen, Z. Tan, Y. Yuan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 5438; b) Y. Yuan, J. Chen, Q. Wan, Z. Tan, G. Chen, C. Kan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 5432; c) Z. Tan, S. Shang, T. Halkina, Y. Yuan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 5424; d) J. A. Brailsford, S. J. Danishefsky, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7196.

[6] For recent reviews on glycoprotein synthesis, see: a) B. G. Davis, *Chem. Rev.* **2002**, *102*, 579; b) D. Gamblin, E. M. Scanlan, B. G. Davis, *Chem. Rev.* **2009**, *109*, 131; c) R. J. Payne, C. H. Wong, *Chem. Commun.* **2010**, *46*, 21.

[7] a) Y. Kajihara, A. Yoshihara, K. Hirano, N. Yamamoto, *Carbohydr. Res.* **2006**, *341*, 1333; b) N. Yamamoto, Y. Tanabe, R. Okamoto, P. Dawson, Y. Kajihara, *J. Am. Chem. Soc.* **2008**, *130*, 501; c) I. Sakamoto, K. Tezuka, K. Fukae, K. Ishii, K. Taduru, M. Maeda, M. Ouchi, K. Yoshida, Y. Nambu, J. Igarashi, N. Hayashi, T. Tsuji, Y. Kajihara, *J. Am. Chem. Soc.* **2012**, *134*, 5428; d) T. Takemura, H. Hojo, Y. Nakahara, T. Ishimuzu, S. Hase, *Org. Biomol. Chem.* **2004**, *2*, 133; e) C. Unverzagt, *Tetrahedron Lett.* **1997**, *38*, 5627; f) M. Murakami, R. Okamoto, M. Izumi, Y. Kajihara, *Angew. Chem.* **2012**, *124*, 3627; *Angew. Chem. Int. Ed.* **2012**, *51*, 3567; g) H. Kunz, C. Unverzagt, *Angew. Chem.* **1988**, *100*, 1763; *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 1697; h) M. R. Pratt, C. R. Bertozzi, *J. Am. Chem. Soc.* **2003**, *125*, 6149; i) Z. W. Guo, N. Shao, *Med. Res. Rev.* **2005**, *25*, 655; j) C. Piontek, D. Varón Silva, C. Heinlein, C. Pöhner, S. Mezzato, P. Ring, A. Martin, F. X. Schmid, C. Unverzagt, *Angew. Chem.* **2009**, *121*, 1974; *Angew. Chem. Int. Ed.* **2009**, *48*, 1941.

[8] For recent improvements in the efficiency of the SPPS approach to glycopeptide synthesis, see: C. Heinlein, D. V. Silva, A. Tröster, J. Schmidt, A. Gross, C. Unverzagt, *Angew. Chem.* **2011**, *123*, 6530; *Angew. Chem. Int. Ed.* **2011**, *50*, 6406.

[9] a) R. Chen, T. J. Tolbert, *J. Am. Chem. Soc.* **2010**, *132*, 3211; b) T. Conroy, K. A. Jolliffe, R. J. Payne, *Org. Biomol. Chem.* **2010**, *8*, 3723.

[10] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776.

[11] a) S. T. Anisfeld, P. T. Lansbury, *J. Org. Chem.* **1990**, *55*, 5560; b) S. T. Cohen-Anisfeld, P. T. Lansbury, *J. Am. Chem. Soc.* **1993**, *115*, 10531.

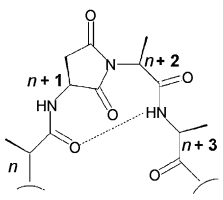
[12] a) M. Bodanszky, J. C. Tolle, S. S. Deshmane, A. Bodanszky, *Int. J. Pept. Protein Res.* **1978**, *12*, 57; b) M. Bodanszky, G. F. Sigler, A. Bodanszky, *J. Am. Chem. Soc.* **1973**, *95*, 2352.

[13] P. Wang, X. Li, J. Zhu, J. Chen, Y. Yuan, X. Wu, S. J. Danishefsky, *J. Am. Chem. Soc.* **2011**, *133*, 1597.

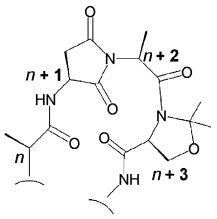
[14] Attempts to improve reactivity by increasing the solubility of peptide **2** through the use of dodecylphosphocholine were unsuccessful.

[15] T. Haack, M. Mutter, *Tetrahedron Lett.* **1992**, *33*, 1589.

[16] Installation of the Leu-Ser pseudoproline dipeptide at the C-terminus served to minimize peptide aggregation: F. García-Martín, P. White, R. Steinauer, S. Côté, J. Tulla-Puche, F. Albericio, *Biopolymers* **2006**, *84*, 566.
[17] P. Nagorny, B. Fasching, X. Li, G. Chen, B. Aussedat, S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 5792.
[18] Small amounts of ammonia are often present in the glycosylamine component, even following multiple lyophilizations.
[19] P. Wang, S. Dong, J. A. Brailsford, S. D. Townsend, Q. Zhang, K. Iyer, R. C. Hendrickson, J. H. Shieh, M. Moore, S. J. Danishefsky, *Angew. Chem.* **2012**, DOI: 10.1002/ange.201206090; *Angew. Chem. Int. Ed.* **2012**, DOI: 10.1002/anie.201206090.
[20] The basis for the $(n+2)$ pseudoproline effect in suppressing aspartimide formation both in peptide synthesis as well as in asparagine glycosylation invites various interpretations. It has been argued, with computational support, that type II′ β turn, stabilized by an intramolecular hydrogen bond (see structure **A**) of a type which cannot be present in the pseudoproline case (see structure **B**). This effect could be operative at the kinetic level, leading to suppression of imide formation relative to biomolecular acylation. See: a) S. Capasso, L. Mazzarella, A. Zagari, *Chirality* **1995**, *7*, 605; b) R. Subirós-Funosas, A. El-Faham, F. Albericio, *Tetrahedron* **2011**, *67*, 8595. Another line of conjecture focuses on the possible inhibition of the amidic NH deprotonation step necessary for aspartamide formation. Perhaps the hydrogen of the aspartamide NH group is transferred to the adjacent amide of the $(n+1)$ amino acid through a hydrogen bond (to N or O). The tertiary amide character of the pseudoproline engagement, would impede such a transfer to either the N or O atom. Clearly the question of the basis for the pseudoproline effect will be pursued by concerned research groups, including our own.



Structure A



Structure B

