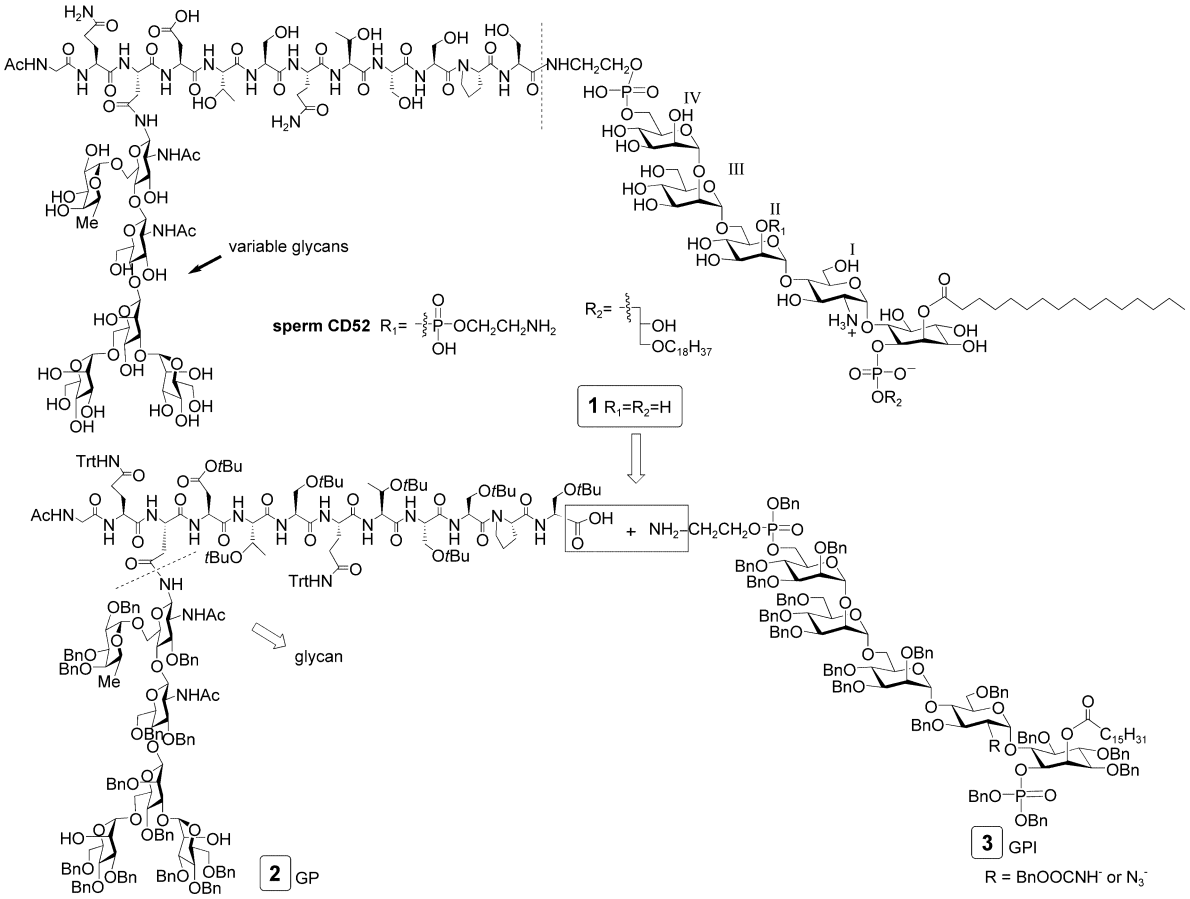
Glycopeptides

Chemical Synthesis of a Skeleton Structure of Sperm CD52—A GPI-Anchored Glycopeptide**

*Ning Shao, Jie Xue, and Zhongwu Guo**

Glycosylphosphatidylinositol (GPI)-anchored glycopeptides or glycoproteins are ubiquitously present on eukaryotic cells.[1] For example, CD52 antigen, which is expressed by virtually all human lymphocytes and sperms,[2,3] is a simple GPI-anchored glycopeptide, but it plays a critical role in the human immune and reproduction systems.[4–6] Sperm CD52 is involved in the human sperm–egg recognition and binding, while antibodies against lymphocyte CD52 have been used to treat several immune system-related diseases.

The structure of CD52 is rather typical as it has the phospholipid and the conserved GPI glycan linked to the inositol C1–*O* and C6–*O* positions and the glycopeptide to the GPI glycan nonreducing end C6–*O* position through a phosphoethanolamine bridge (Scheme 1).[2,3] Additional biomodifications of sperm CD52 GPI include a phosphoethanolamine group to the mannose II (Man^{II}) C2–*O* position and a large acyl group to the inositol C2–*O* position.[2] The peptide chain of CD52 is very short, consisting of 12 amino acids, and it has only one *N*-glycosylation site, to which complex and heterogeneous glycans are attached. The heterogeneity of CD52 glycoforms constitutes a major obstacle in accessing homogeneous antigens from biological sources.



Scheme 1. Retrosynthetic plan for **1**.

Due to its short peptide, simple glycosylation pattern, and intriguing biological activities, sperm CD52 is a useful model for structural and structure–activity relationship studies of GPI-anchored glycopeptides and glycoproteins. In this context, a practical synthetic method for it should be of great significance. However, there have been very few reports concerning the synthesis of GPI-anchored structures, even though a number of GPIs[7–16] and glycopeptides,[17,18] including CD52 GPI[19,20] and glycopeptides,[21,22] have been synthesized by chemical methods. The only synthetic studies towards GPI-anchored glycopeptides or proteins either used simple

[*] N. Shao, Dr. J. Xue, Prof. Z. Guo
Department of Chemistry, Case Western Reserve University
Cleveland, OH 44106-7078 (USA)
Fax: (+1) 216-368-3006
E-mail: zxg5@case.edu

[**] This work was supported in part by the Research Corporation (Research Innovation Award 0663), the American Chemical Society (PRF 37743-G), and the NIH/NCI (1R01 CA95142). We are also grateful to Dr. S. Chen and Mr. J. Faulk for the MS measurements.

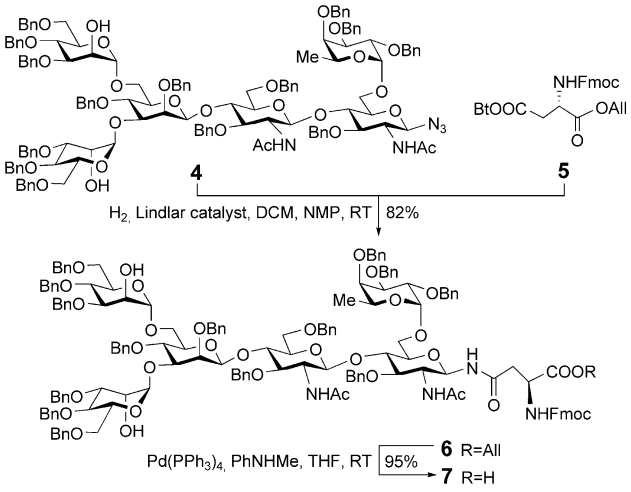
Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

 DOI: 10.1002/ange.200353251

molecular models[19,23] or had an artificial linkage between the GPI and protein.[16] Chemical synthesis of complex, native GPI-anchored glycopeptides, which involves several areas of chemistry, remains a formidable challenge.

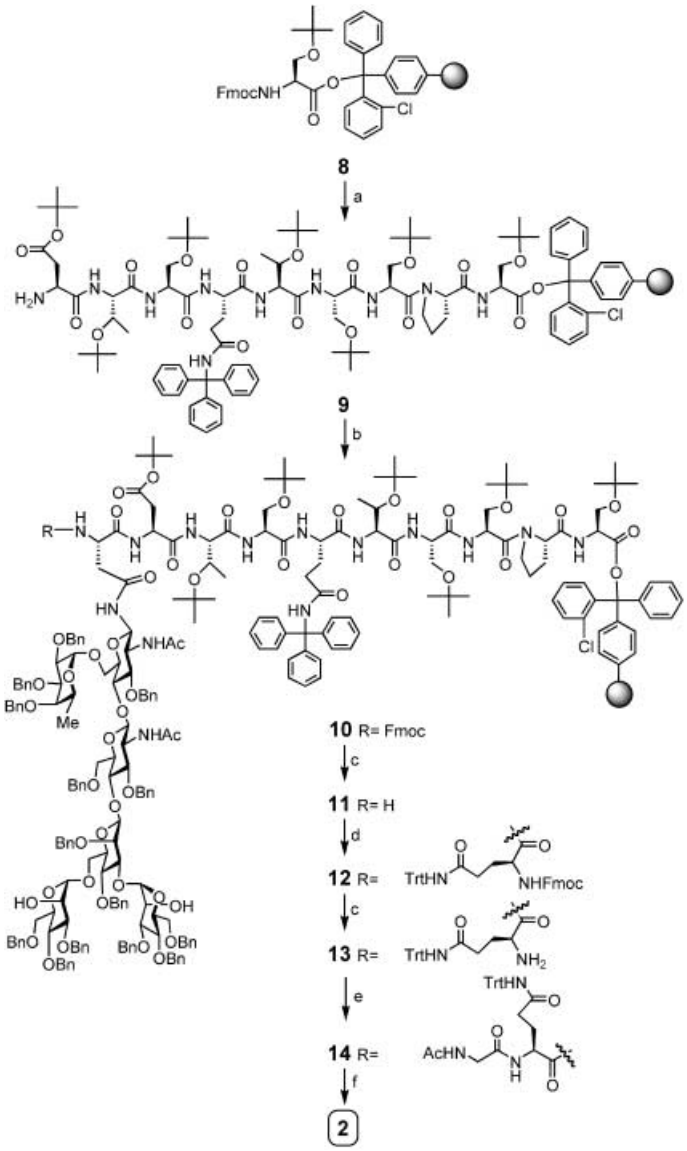
The work presented herein aims at the construction of natively linked GPI-glycopeptide conjugates, especially focusing on the strategic synthetic designs. For this purpose, a skeleton structure **1** (Scheme 1) of sperm CD52 without the glycerol lipid and the extra phosphoethanolamine of its GPI was established as the synthetic target. The omission of two functional groups can save us a few steps and problems in the GPI preparation.[20] For **1**, since an amide can be easily formed by the condensation reaction between a free amino group and a free carboxyl group, the amido bond between the GPI and the glycopeptide was cleaved first to offer two logical building blocks **2** and **3** (Scheme 1) that had very different properties. Moreover, because this linkage is shared by all GPI-anchored glycopeptides and glycoproteins, such a synthetic strategy may be of general significance. Thus, an important issue is the coupling between two complex segments, namely, the GPI and the glycopeptide. For **2** and **3**, we employed benzyl group as a permanent protection for the carbohydrate hydroxyl groups, as it can be easily removed under mild conditions later on. In the meantime, glycopeptide **2** was planned to have a fully protected peptide to prevent the potential interference of peptide side chains with the coupling reaction and the potential influence of a free peptide on the solubility and other properties of involved intermediates. Another important issue of this synthesis was thus preparing **2** and **3** in the properly protected form.

The *N*-glycan **4** with an azido group attached to the reducing end was prepared by a modified literature procedure,[24] which will be reported elsewhere. After the azido group of **4** was selectively reduced to a primary amine by Lindlar catalyst-catalyzed hydrogenation, it was coupled to the side chain of asparagine by the reaction with an active ester **5**. Because the glycan contained a benzylated fucosidic linkage that is extremely sensitive to acids,[25] the α-carboxyl group of **5** was protected by an allyl, instead of a *tert*-butyl (*t*Bu) group typically used in peptide synthesis, to circumvent acidic treatment during the deprotection.[26] Then, the allyl group was smoothly removed by $Pd(PPh_3)_4$ to afford **7** (Scheme 2), which was ready as a key building block for the glycopeptide assembly.



Scheme 2. Synthesis of the glycosyl asparagine **7**.

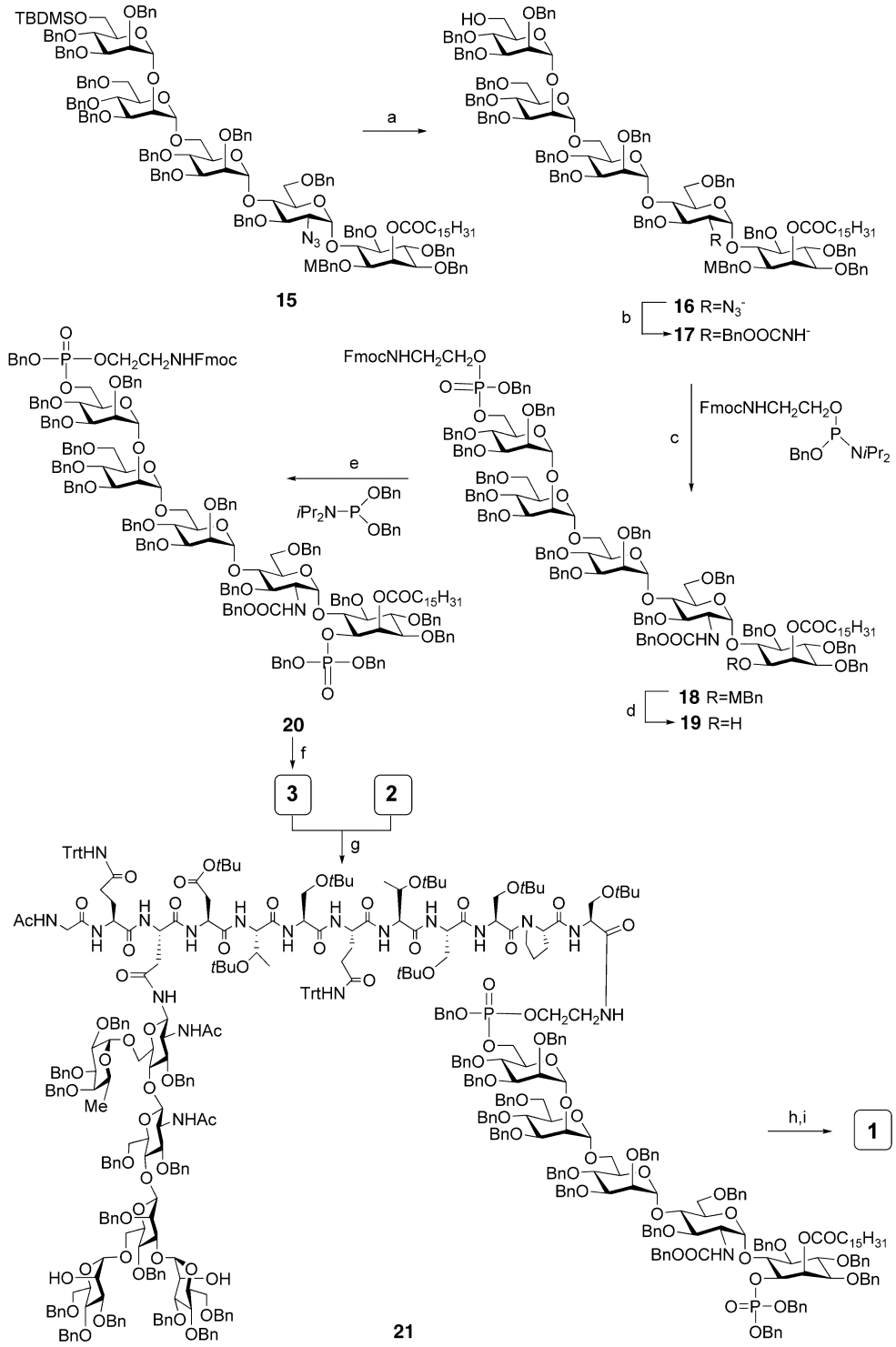
As mentioned earlier, except for the terminal carboxyl group, the peptide chain of glycopeptide **2** needed to be fully protected. In this case, the traditional solid-phase glycopeptide synthesis by using Wang resin[21] would not be useful, since the acidic conditions, such as 95% aqueous trifluoroacetic acid (TFA), used to retrieve glycopeptides from Wang resin would deprotect the amino acid side chains and affect the α-fucosidic linkage.[25] To deal with this problem, we employed the extremely acid-sensitive 2-chorotrityl resin[27] to synthesize **2** (Scheme 3). The ester bond between 2-



Scheme 3. Reagents and conditions: a) solid-phase peptide elongation by conventional Fmoc chemistry on an automatic peptide synthesizer;[30] b) **7**, HOBt, DCC, 0°C to RT, 24 h; c) 20% piperidine in NMP, RT, 2 h; d) Fmoc–Gln(Trt)–OH (Gln = glutamine), HOBt, DCC, NMP, RT, 2 h; e) Ac–Gly–OH (Gly = glycine), HOBt, DCC, NMP, RT, 2 h; f) HOAc, TFE, DCM (1:1:8), RT, 2 h, 70% (from **9**).

chorotrityl resin and the peptide *C*-terminus could be cleaved by 1–20% acetic acid, the use of which did not affect the side-chain protecting groups of amino acids, including the trityl (Trt) group on glutamine.[28]

The glycopeptide assembly started from a commercially available resin, **8**, already loaded with a serine. The peptide was first elongated by the conventional Fmoc (9-fluorenylmethoxycarbonyl) chemistry on an automatic peptide synthesizer by using 2-*O*-(1*H*-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) and *N*,*N*-diisopropylethylamine (DIEA) as the condensation reagents[29] to afford **9**. All amino acids employed were commercially



Scheme 4. Reagents and conditions: a) Tetrabutyl ammonium fluoride (TBAF), HOAc, DCM, 35–40°C, 3 days, 95%; b) PEt_3, CBz_2O, MeOH, DCM (1:2), RT, overnight, 70%; c) tetrazole, DCM, MeCN (3:1), RT, 6 h; then *t*BuOOH, 84%; d) TFA, DCM (1:9), RT, 0.5 h, >99%; e) tetrazole, DCM, MeCN (3:1), RT, 0.5 h; then *t*BuOOH; 57%. f) DBU, DCM, RT, 2 min, 90%; g) HOBt, EDC, DCM, NMP (2:1), RT, 1 day, 70%; h) 10% Pd/C, H_2, $CHCl_3$, MeOH, H_2O (10:10:3), 2 days; i) TFA in DCM (15%, containing 10% of $HSiEt_3$); then HPLC, 85% (two steps).

available. Next, the glycosyl asparagine **7** was introduced manually to minimize the use of **7**. In this coupling, only 2 equivalents of **7** were used with *N*,*N*-dicyclohexylcarbodiimine (DCC) and *N*-hydroxybenzotriazole (HOBt) as the condensation reagents. Thereafter, manual *N*-deprotection of **10** with 20% piperidine in *N*-methylpyrrolidinone (NMP), followed by further peptide elongation by using Fmoc–amino acids and the standard protocols of solid-phase peptide synthesis,[30] resulted in the glycopeptide–resin conjugate **14.** It was finally treated with a mixture of acetic acid, trifluroethanol (TFE) and dichloromethane (DCM) (1:1:8) to offer a white solid, which was purified by preparative TLC. In addition to the expected glycopeptide **2** (13 mg), a non-glycosylated peptide (2 mg) was also obtained as a side product. According to this result, the coupling yield between **7** and **9** was about 70%. The NMR spectroscopy and MS data of **2** agreed well with the proposed structure. One possible concern about this synthesis was whether asparagine suffered from racemization during the coupling between **7** and **9**. Since no other product was obtained and the NMR spectra of **2** showed only one set of signals, we assumed that this side reaction, if any, was minimal.

Scheme 4 outlines the GPI synthesis as well as the final coupling and deprotection steps. First, the GPI core **15** was prepared according to a recently reported procedure.[19] Then, its nonreducing end C6-*O* position was deprotected to give **16**. The phosphorylation of **16** turned out to be problematic.[20] In contrast to all previous GPI syntheses,[7–16] the azido group was affected by the reaction to result in complex products. In fact, the acyl group at the inositol 2-*O* position caused a number of problems in the synthesis of sperm CD52 GPI, which was discussed separately.[20] To overcome this complication, the azido group was reduced with triethylphosphine and the resulted free amine was protected by a benzyloxycarbonyl (CBz) group to afford **17** that was phosphorylated smoothly. The Fmoc group was used to protect the amino group of phosphoethanolamine in **17** to facilitate the selective deprotection by a mild base later on. After the *p*-methoxylbenzyl (MBn) group of **18** was removed by TFA, the second phosphate was attached to the exposed inositol 1-hydroxyl group (57%). All intermediates were conveniently purified by column chromatography. Finally, treatment of **20** with diluted 1,8-diazabicyclo[5,4,0]undec-7-ene (DBU) in DCM for a brief period (2 min) gave the GPI building block **3**, which was practically pure and was thus directly used without further purification during the construction of the target molecule.

The coupling reaction between **2** and **3** was performed in a mixture of DCM and NMP (2:1) under a nitrogen atmosphere by using 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (EDC) and HOBt (10 equivalents) as the condensation reagents,[31] which turned out to be very efficient. After one day of stirring at room temperature, the reaction mixture was worked up and the product was purified by column chromatography to give **21** in a 70% isolated yield. Compound **21** was well separated from both substrates by TLC and column chromatography. Its 1D, 2D NMR, and MS data were consistent with the expected structure. The NMR signals of its reporter groups, such as that of the carbohydrate anomeric protons, carbon atoms and some amino acid α-protons, were easily assigned. Moreover, from the NMR spectra, we could not identify the potential recemization product, which was consistent with our previous results with a simple molecular model.[19] Global deprotection of **21** was achieved in two steps, namely hydrogenolytic debenzylation in a hydrogen atmosphere with 10% Pd/C as the catalyst followed by removal of the peptide side-chain protecting groups by 15% TFA in DCM. It was important to observe this deprotection sequence, as the benzylated fucosidic linkage is much more labile to TFA than the free sugar. For example, 15% TFA could destroy the former within an hour but had little influence on the latter after several hours of treatment.[28] The final synthetic target **1**[32] was then thoroughly purified by reversed-phase HPLC by using a C_{18} column (250 mm × 10 mm). The presence of a long lipid chain resulted in its strong retention on the column. Therefore, with 40% *iso*-propanol/water (2 mL min^{-1}) as the eluent, **1** displayed a retention time of 37.5 min, while free CD52 glycopeptides were washed out from the same column by 2% *iso*-propanol/water in less than 25 min. The structure of **1** was supported by its high field NMR spectroscopy and MS.

In summary, this paper described the convergent synthesis of a skeleton structure of the sperm CD52 antigen, which represents the first chemical synthesis of a complex and all natively linked glycopeptide–GPI conjugate. Our general synthetic design was to separately prepare the protected glycopeptide and the GPI first. 2-Chlorotrityl resin that has an extremely acid-labile linker was used in the solid-phase synthesis of the glycopeptide to obtain its fully protected form. Next, the glycopeptide *C*-terminus and the GPI nonreducing end phosphoethanolamine group were selectively deprotected. Finally, the two segments were joined through an amido bond by using HOBt/EDC as the condensation reagents, to give the glycopeptide-GPI conjugate. The coupling reaction proved to be very efficient. Because this amido linkage is shared by various GPI-anchored glycopeptides or glycoproteins, we assume that the methods described herein may be useful for the synthesis of other similar structures.

Received: November 4, 2003 [Z53251]

Keywords: antigens · carbohydrates · glycopeptides · glycoproteins · solid-phase synthesis

[1] M. A. J. Ferguson, A. F. Williams, *Annu. Rev. Biochem.* **1988**, *57*, 285–320.
[2] S. Schroter, P. Derr, H. S. Conradt, M. Nimtz, G. Hale, C. Kirchhoff, *J. Biol. Chem.* **1999**, *274*, 29862–29873.
[3] A. Treumann, M. R. Lifely, P. Schneider, M. A. J. Ferguson, *J. Biol. Chem.* **1995**, *270*, 6088–6099.
[4] G. Hale, M.-Q. Xia, H. P. Tighe, J. S. Dyer, H. Waldmann, *Tissue Antigens* **1990**, *35*, 118–127.
[5] Y. Tsuji in *Immunology of Human Reproduction* (Eds.: M. Kurpisz, N. Fernandez), BIOS Sci. Pub., Oxford, **1995**, pp. 23–32.
[6] E. D. Eccleston, T. W. White, J. B. Howard, D. W. Hamilton, *Mol. Reprod. Dev.* **1994**, *37*, 110–119.
[7] A. Crossman, Jr., J. B. Brimacombe, M. A. J. Ferguson, T. K. Smith, *Carbohydr. Res.* **1999**, *321*, 42.

[8] U. E. Udodong, R. Madsen, C. Roberts, B. Fraser-Reid, *J. Am. Chem. Soc.* **1993**, *115*, 7886 – 7887.
[9] A. S. Campbell, B. Fraser-Reid, *J. Am. Chem. Soc.* **1995**, *117*, 10387 – 10388.
[10] D. K. Baeschlin, A. R. Chaperon, V. Charbonneau, L. G. Green, S. V. Ley, U. Lücking, E. Walther, *Angew. Chem.* **1998**, *110*, 3609 – 3614; *Angew. Chem. Int. Ed.* **1998**, *37*, 3423 – 3428.
[11] N.-C. Reichardt, M. Martín-Lomas, *Angew. Chem.* **2003**, *115*, 4822 – 4825; *Angew. Chem. Int. Ed.* **2003**, *42*, 4674 – 4677.
[12] C. Murakata, T. Ogawa, *Carbohydr. Res.* **1992**, *235*, 95 – 114.
[13] T. G. Mayer, B. Kratzer, R. R. Schmidt, *Angew. Chem.* **1994**, *106*, 2289 – 2293; *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 2177 – 2181.
[14] K. Pekari, R. R. Schmidt, *J. Org. Chem.* **2003**, *68*, 129 – 1308.
[15] M. C. Hewitt, D. A. Snyder, P. H. Seeberger, *J. Am. Chem. Soc.* **2002**, *124*, 13434 – 13436.
[16] L. Schofield, M. C. Hewitt, K. Evans, M.-A. Siomos, P. H. Seeberger, *Nature* **2002**, *418*, 785 – 789.
[17] See review: O. Seitz, *ChemBioChem* **2000**, *1*, 214 – 246, and references therein.
[18] See review: H. Herzner, T. Reipen, M. Schultz, H. Kunz, *Chem. Rev.* **2000**, *100*, 4495 – 4537.
[19] J. Xue, N. Shao, Z. Guo, *J. Org. Chem.* **2003**, *68*, 4020 – 4029.
[20] J. Xue, Z. Guo, *J. Am. Chem. Soc.* **2003**, *125*, 16334 – 16339.
[21] Z. Guo, Y. Nakahara, Y. Nakahara, T. Ogawa, *Angew. Chem.* **1997**, *109*, 2442 – 2444; *Angew. Chem. Int. Ed.* **1997**, *36*, 1464 – 1467, and references therein.
[22] M. R. Pratt, C. R. Bertozzi, *J. Am. Chem. Soc.* **2003**, *125*, 6149 – 6159.
[23] Tanaka, Y. Nakahara, H. Hojo, Y. Nakahara, *Tetrahedron* **2003**, *59*, 4059 – 4067.
[24] Z. Guo, Y. Nakahara, Y. Nakahara, T. Ogawa, *Bioorg. Med. Chem.* **1997**, *5*, 1917 – 1924, and references therein.
[25] C. Unverzagt, H. Kunz, *Bioorg. Med. Chem.* **1994**, *2*, 1189 – 1201.
[26] M. Rosch, H. Herzner, W. Dippold, M. Wild, D. Vestweber, H. Kunz, *Angew. Chem.* **2001**, *113*, 3954 – 3957; *Angew. Chem. Int. Ed.* **2001**, *40*, 3836 – 3839.
[27] K. Barlos, D. Gatos, S. Kapolos, G. Papaphotiu, W. Schaefer, W. Yao, *Tetrahedron Lett.* **1989**, *30*, 3947 – 3950.
[28] N. Shao, J. Xue, Z. Guo, *J. Org. Chem.* **2003**, *68*, 9003 – 9011.
[29] L. A. Carpino, *J. Am. Chem. Soc.* **1993**, *115*, 4397 – 4398.
[30] M. Bodanszky, *Peptide chemistry: A practical textbook*, 2nd ed., Springer, Berlin, **1993**.
[31] L. A. Carpino, A. El-Faham, F. Albericio, *J. Org. Chem.* **1995**, *60*, 3561 – 3564.
[32] Selected data. **6**: $[\alpha]_D = -3.48$ ($c = 1.2$, $CHCl_3$). ^{1}H NMR (600 MHz, $CDCl_3$): $\delta = 7.73$–7.55 (m, 4H, Fmoc), 5.83 (m, 1H, All), 5.26 (m, 2H, N-H, All), 5.17 (dd, $J = 1.2$, 10.2 Hz, 1H, All), 5.13 (s, 1H, Man-1), 4.90 (s, 1H, Man 1(anomeric)), 4.82 (s, 1H, Fuc 1(Fuc = fucose), 4.72 (d, $J = 8.4$ Hz, 1H, GlcNAc 1), 4.70 (d, $J = 8.4$ Hz, 1H, GlcNAc 1 (GlcN = glucosamine), 4.51 (s, 1H, Man 1), 2.85 (dd, $J = 3.6$, 15.6 Hz, 1H, Asn β (Asn = asparagine), 2.62 (dd, J 3.0, 15.6 Hz, 1H, Asn β), 1.73 (s, 3H, Ac), 1.66 (s, 3H, Ac), 1.03 ppm (d, $J = 6.0$ Hz, 3H, Fuc 6). MALDI-TOF MS: Calcd for $C_{160}H_{173}O_{34}N_4Na$, 2717.183, found 2717.181 [M+H+Na$^+$]. **21**: TLC: $R_f = 0.41$ (DCM and MeOH 10:1). ^{1}H NMR ($CDCl_3$, 600 MHz): $\delta = 5.81$ (br, 1H, Ino 2 (Ino = inositol)), 5.31 (d, $J = 3$ Hz, 1H, GlcN 1), 5.25 (d, $J = 7$ Hz, 1H, GlcNAc 1), 5.11 (br, 2H, 2 Man 1), 4.85 (1H, Man 1), 4.76 (1H, Fuc 1), 4.71 (1H, Man 1), 4.69 (1H, GlcNAc 1), 4.67 (1H, Man 1), 4.45 (b, 1H, Man 1), 2.80–2.62 (m, 4H, Asp β(Asp = aspartic acid), Asn β), 1.76 (s, 3H, Ac), 1.72 (s, 3H, Ac), 1.01, 0.99 (2 d, $J = 6.0$ Hz, 6H, Fuc 6, Thr γ(Thr = threonine)), 0.87 ppm (t, $J = 7.2$ Hz, 3H, lipid–Me). ^{13}C (150 MHz, $CDCl_3$, from HMQC): $\delta = 101.4$ (2 Man 1), 101.2 (Man 1), 100.6 (GlcNAc 1), 100.0 (Man 1), 99.3 (GlcN 1, 2 Man 1), 98.0 (Fuc 1), 79.5 (GlcNAc 1), 68.0 ppm (Ino 2). ^{31}P NMR ($CDCl_3$): $\delta = -1.07$, -1.31 ppm. MALDI-TOF-MS: Calcd for $C_{426}H_{505}N_{19}O_{87}P_2$ 7346, found 7385 [M+K$^+$]. **1**: HPLC: Retention time = 37.5 min, C_{18} column (Discovery 250 mm × 10 mm), eluent: 40% *i*PrOH in H_2O (2.0 mL min^{-1}). ^{1}H NMR (D_2O, 600 MHz): $\delta = 5.24$ (br, 1H, Man 1), 5.14 (s, 1H, Man 1), 5.11 (br, 2H, 2 Man 1), 5.04 (br, 3H, Man 1, GlcNAc 1, Fuc 1), 4.93 (s, 1H, Man 1), 4.88 (br, 1H, Ser α(Ser = serine)), 4.83 (m, 1H, Asp α), 4.79 (bs, 1H, Ino 2), 4.75 (m, 1H, Asn α), 4.70 (d, 1H, GlcNAc 1), 4.26 (s, 1H, GlcN 1), 2.90–2.64 (m, 4H, Asn β, Asp β), 2.09, 2.07, 2.03 (3 s, 3 × 3H, 3 Ac), 1.30–1.15 (m, Fuc 6, Thr γ), 0.88 ppm (br, lipid-Me). MALDI-TOF MS: Calcd for $C_{135}H_{229}N_{19}O_{85}P_2$ 3538, found 3561 [M+Na$^+$].

