Pergamon

Bioorganic & Medicinal Chemistry Letters, Vol. 7, No. 12, pp. 1537-1542, 1997

0960-894X/97 $17.00 + 0.00

PII: S0960-894X(97)00255-2

STEREOSELECTIVE SYNTHESIS OF NEO-C-GLYCOPEPTIDE BUILDING BLOCKS: TOWARDS A FLEXIBLE AND CONTROL-ORIENTED DESIGN AS PROBES FOR CARBOHYDRATE-PROTEIN INTERACTIONS[1]

Prabhat Arya,* Sylvain Dion, and George K. H. Shimizu

Steacie Institute for Molecular Sciences
National Research Council of Canada, Ottawa, Ontario, Canada, K1A 0R6

Abstract: Neo-C-glycopeptides (**1–4**) have been synthesized as building blocks to obtain higher neo-C-glycopeptides as probes for studying carbohydrate–protein interactions. A convergent approach for the synthesis of **4** has been developed, in which two galactose units are attached to a glycine derivative in a stepwise procedure (reductive amination followed by amide coupling) and finally coupling to the protected dipeptide having a free amino group on the side chain. © 1997 Elsevier Science Ltd.

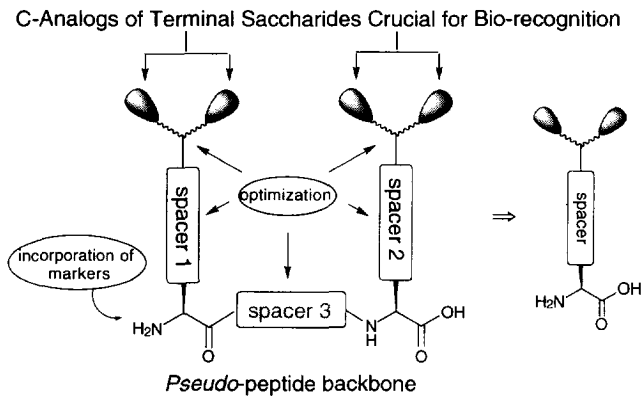
Individually weak, noncovalent interactions between protein receptors and carbohydrate ligands form the basis of recognition events that are fundamental to a vastly diverse range of intercellular processes.[2] Molecules capable of modulating or displacing these interactions offer a plethora of potential applications ranging from anti-viral agents, growth regulators to synthetic vaccines. Examples that are currently the focus of numerous zealous research efforts include: (i) mimics of Sialyl Lewis X, a cell surface tetrasaccharide that is a ligand for the endothelial leucocyte adhesion molecule, E-Selectin;[3] (ii) syntheses of compounds capable of blocking influenza virus haemagglutinin binding to cell surface sialic acid residues, the first step in infection of a cell by the virus;[4] (iii) the design of oligosaccharide constructs, which inhibit mammalian gamete adhesion in the zona pellucida (ZP).[5] It has been postulated that, in mouse ZP, the terminal saccharides involved in this recognition are galactose units α-linked to the penultimate sugars.[6]

In order to compensate for the weak forces that characterize carbohydrate–protein interactions,[2] both Nature,[7] and subsequent attempts by various groups to mimic Nature,[8] have employed multivalency as a tool to augment individual ligand–receptor interactions, the so-called "cluster effect".[2c] However, equally significant to enhancing the binding avidity of carbohydrate mimics is the elucidation of the interactions that dictate the specificity of a given recognition event. In certain cases, after synthesis of a series of neoglycoconjugates, binding of the saccharide mimics has been used to infer structural details of the protein receptor and to illustrate the subtle nature of the complementary interaction.[9]

After several years of efforts in mimicking cell surface carbohydrates, the development of flexible and appropriate designs of scaffolds to present carbohydrate units in multivalent orientations appears to be an important milestone.[2c,8j] Towards this goal, we are developing a flexible and control-oriented model for multivalent presentations of the terminal saccharides that couples the intrinsic structural variation provided by a peptide backbone[10] with the enhanced enzymatic stability offered by C-glycosidic linkages.[11] In several cases, it has been illustrated that the terminal saccharide represents the minimum recognition motif for the protein receptor and that the penultimate sugar's main function is as a scaffolding to properly orient the terminal saccharide to maximize the binding interaction.[10a] Our strategy is to display the stable mimics of terminal

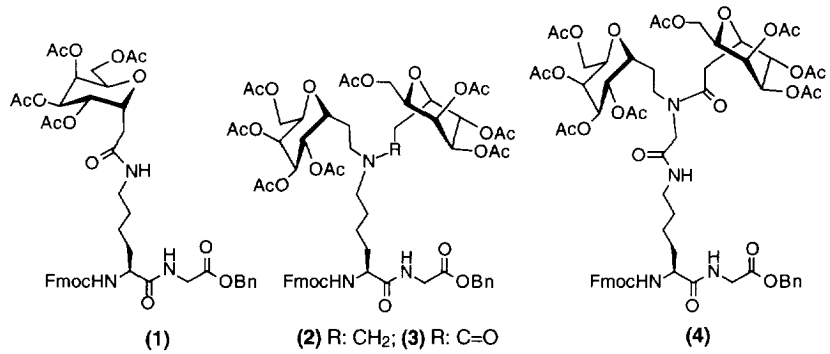
saccharides (i.e., replacement of anomeric oxygen by carbon) as recognition elements on a peptide backbone as shown schematically in Figure 1. *Optimization in the design of probes for studying carbohydrate–protein interactions can be achieved by the variation of spacers 1, 2 and 3, which allows for diversity in the display of terminal saccharides.* Our approach offers: (i) systematic building block assembly using solid-phase chemistry; (ii) highly controlled, yet broad, structural variation; (iii) secondary interactions with adjacent amino acid residues, which may further augment binding; and (iv) the incorporation of markers (e.g., biotin) either at the N- or C-terminal of the peptide backbone. This model can also exploit the growing library of techniques available for glycopeptide synthesis[12] and be applied to the generation of a combinatorial library of stable neoglycoconjugates.[13]

Figure 1

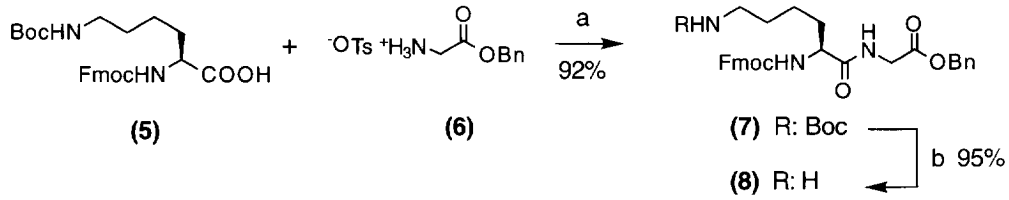


In this communication, we illustrate our approach by reporting the synthesis of four neo-C-glycopeptides as building blocks, shown in Figure 2, required for the synthesis of higher neo-C-glycopeptides as probes for studying carbohydrate–protein interactions. For the initial studies, we have selected α-galactose as the terminal sugar in the neo-C-glycopeptide building blocks (**1**–**4**) to demonstrate our control-oriented approach.

Figure 2

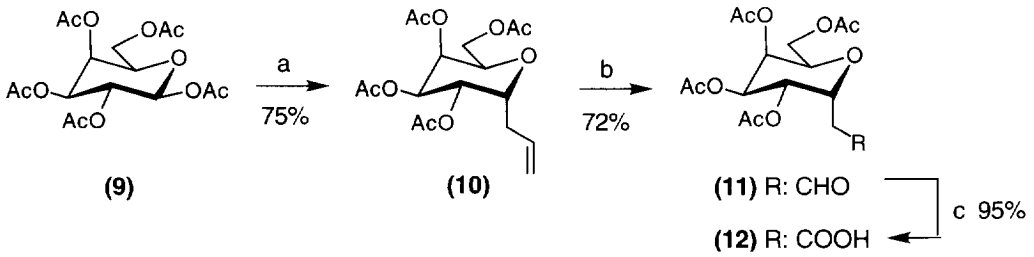


The peptide unit of compounds **1–4** was obtained from the dipeptide α-N(Fmoc)-ε-N(tBoc)-L-Lys-Gly-OBn (Scheme 1, **7**), which was prepared from the commercially available α-N(Fmoc)-ε-N(tBoc)-L-Lys-OH (**5**) and the *p*-toluenesulfonic acid salt of Gly-OBn (**6**) by a DCC/HOBt coupling method in 92% isolated yield after purification over silica gel by flash column chromatography. The fully protected dipeptide (**7**), on brief exposure to trifluoroacetic acid (TFA), gave **8** with the ε-side chain amino group as the TFA salt. Compound **8** was directly used for coupling with the glycoside units, α-C-galactose derivatives **11** and **12**.



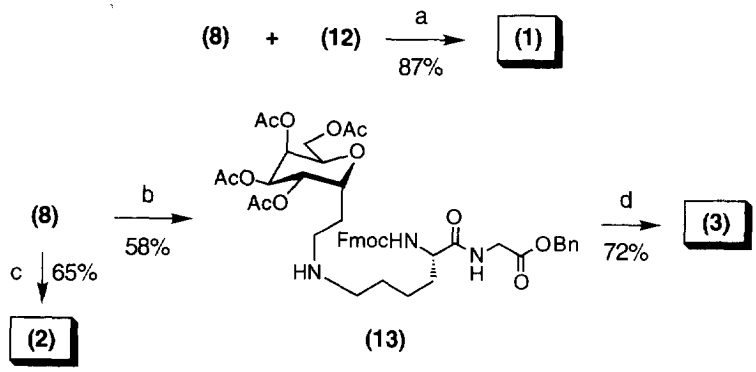
Scheme 1: (a) DCC, HOBt, DIEA, CH_2Cl_2; (b) TFA, CH_2Cl_2.

The α-C-galactose derivatives (Scheme 2, **11** and **12**), were obtained from β-D-galactose pentaacetate (**9**) through a number of steps. Stereoselective α-C-allylation of **9** by allyltrimethylsilane and $BF_3{\cdot}Et_2O$ as a Lewis acid gave **10** in 75% isolated yield after purification.[14a] The α-C-Allyl group of **10** was converted to the aldehyde derivative, **11**, by ozonolysis (O_3, CH_2Cl_2, -78 °C) followed by the reduction of the ozonide (Zn/AcOH, 0 °C). Subsequently, **11** was oxidized to the carboxylic acid derivative **12**, via a Masamune oxidation.[14b]



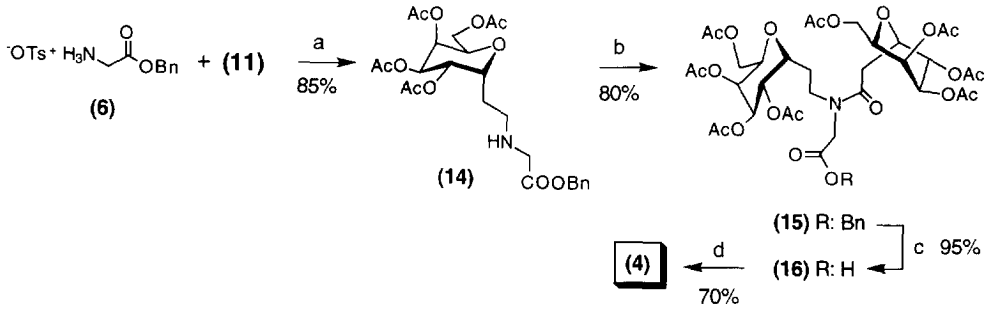
Scheme 2: (a) Allyltrimethylsilane, $BF_3{\cdot}Et_2O$, CH_2Cl_2, 4 °C; (b) (i) O_3, CH_2Cl_2, -78 °C; (ii) Zn/AcOH, 0 °C; (c) $KMnO_4$, *t*-BuOH-5% NaH_2PO_4.

With the completion of the synthesis of the peptide and glycoside subunits, the neo-glycopeptide building blocks **1–3** (Figure 1) were assembled as shown in Scheme 3. Neo-glycopeptide **1** was synthesized by the coupling of dipeptide **8** with glycoside **12** under DCC/HOBt reaction conditions. Purification over silica gel gave **1** in 87% isolated yield.[15] Mono-reductive amination of **8** with glycoside **11** gave **13** in 58% isolated yield after purification over silica gel. Under similar reaction conditions, but using 2.2 equiv of **11**, di-reductive amination of **8** was performed to give the neo-glycopeptide **2** in 65% isolated yield after purification.[15] Neo-glycopeptide **3** was synthesized from **13** in 72% yield after purification by coupling with **12** under DCC/HOBt reaction conditions.[15]



Scheme 3: (a) DCC, HOBt, DIEA, CH_2Cl_2; (b) 1.5 equiv **11**, $NaB(OAc)_3H$, *p*H 5, THF, AcOH, 0 °C; (c) 2.2 equiv **11**, $NaB(OAc)_3H$, *p*H 5, THF, AcOH, 0 °C. (d) 1.3 eq **12**, DCC, HOBt, DIEA, CH_2Cl_2.

A convergent strategy was used for the synthesis of neo-C-glycopeptide **4** (Scheme 4). The two subunits consist of the dipeptide (**8**) and a di-glycosylated amino acid (**16**). Di-glycosylated amino acid (**16**) was obtained through the following steps. Reductive amination of the *p*-toluenesulfonic acid salt of Gly-OBn (**6**) with the aldehyde (**11**) gave **14** as the mono-reductive product. This was followed by coupling with acid **12** using the DCC/HOBt method to obtain **15**. The final step consisted of deprotection by hydrogenation with 10% Pd/C in 95% EtOH (-COOBn to -COOH) yielding the desired subunit **16**. Coupling of **16** with dipeptide **8** gave neo-glycopeptide **4** in 70% isolated yield after purification over silica gel.[15]



Scheme 4: (a) DIEA, $NaB(OAc)_3H$, *p*H 5, THF, AcOH, 0 °C; (b) **12**, DCC, HOBt, DIEA, CH_2Cl_2; (c) H_2, 10% Pd/C, 95% EtOH; (d) **8**, DCC, HOBt, DIEA, CH_2Cl_2.

To summarize, the synthesis of neo-C-glycopeptides as building blocks required for the synthesis of higher neo-C-glycopeptides has been achieved employing (i) stereoselective C-allylation of a galactose derivative, (ii) oxidation of the allyl group to a -CHO group, and (iii) reductive amination of the sugar derivative with a protected dipeptide having a free amino group on the side chain. A convergent approach to branched side chains, in which two galactose units are attached to a glycine derivative in a stepwise procedure (reductive amination and amide coupling), followed by coupling to the protected dipeptide (Scheme 4) has also been developed. As a continuation, using the deprotected (-COOBn to -COOH) forms of neo-C-glycopeptides (**1–4**),

systematic synthesis of higher neo-C-glycopeptides is under study.[16] Binding studies with various derivatives of neo-C-glycopeptides employing cell agglutination assays and surface plasmon resonance technique using plant lectin, *Maclura Pomifera,* (specific for α-galactose) will be reported in due time.[16]

Acknowledgments: Dr. Kristina Kutterer is thanked for the constructive comments and for editing the manuscript. We thank Dr. Jim Richards (IBS, NRCC) and Prof. René Roy (Univ. of Ottawa) for helpful discussions; MS and NMR group (IBS, NRCC) for providing excellent support. NSERC Canada is thanked for providing financial support for S.D.

References and Notes:

1. NRCC publication no. 40823.

2. (a) Varki, A. *Glycobiology* **1993**, *3*, 97. (b) Sharon, N.; Lis, H. *Sci. Am.* **1993** (January), 82. (c) Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *26*, 323. (d) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683. (e) Sears, P.; Wong, C.-H. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12086.

3. (a) Uchiyama, T.; Vassilev, V. P.; Kajimoto, T.; Wong, W.; Huang, H.; Lin, C.-C.; Wong, C.-H. *J. Am. Chem. Soc.* **1995**, *117*, 5395. (b) Graves, B. J.; Crowther, R. L.; Chandran, C.; Rumberger, J. M.; Li, S.; Huang, K.-S.; Presky, D. H.; Familletti, P. C.; Wolitzky, B. A.; Burns, D. K. *Nature (London)*, **1994**, *367*, 532. (c) Sprengard, U.; Schudok, M.; Schmidt, W.; Kretzschmar, G.; Kunz, H. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 321.

4. (a) Sigel, G. B.; Mamman, M.; Dahmann, G.; Whitesides, G. M. *J. Am. Chem. Soc.* **1996**, *118*, 3789. (b) Toogood, P. L.; Galliker, P. K.; Glick, G. D.; Knowles, J. R. *J. Med. Chem.* **1991**, *34*, 3140. (c) Spevak, W.; Nagy, J. O.; Charych, D. H.; Schaefer, M. E.; Gilbert, J. H.; Bednarski, M. D. *J. Am. Chem. Soc.* **1993**, *115*, 1146. (d) Sabesan, S.; Duus, J. Ø.; Neira, S. ; Domaille, P.; Kelm, S.; Paulson, J. C.; Bock, K. *J. Am. Chem. Soc.* **1992**, *114*, 8363. (e) Roy, R.; Zanini, D.; Meunier, S. J.; Romanowska, A. *Chem. Commun.* **1993**, 1869.

5. (a) Miller, D. J.; Macek, M. B.; Shur, B. D. *Nature (London)*, **1992**, *357*, 589. (b) Litscher, E. S.; Juntunen, K.; Seppo, A.; Penttilä, L.; Niemelä, R.; Renkonen, O.; Wassarman, P. M. *Biochemistry*, **1995**, *34*, 4662. (c) Seppo, A.; Penttilä, L.; Niemelä, R.; Maaheimo, H.; Renkonen, O.; Keane, A. *Biochemistry*, **1995**, *34*, 4655.

6. Bleil, J. D.; Wassarman, P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6778. It has also been suggested that the terminal saccharide is GlcNAc in β-linkage. See ref. 5a.

7. (a) Weis, W. I.; Drickamer, K.; Hendrickson, W. A. *Nature (London)*, **1992**, *360*, 127. (b) Goldthar, J. *Methods Enzymol.* **1994**, *236*, 211.

8. (a) Mortell, K. H.; Gingras, M.; Kiessling, L. L. *J. Am. Chem. Soc.* **1994**, *116*, 12053. (b) Roy, R.; Saha, U. K. *Chem. Commun.* **1996**, 201. (c) Lindhorst, T. K.; Kieburg, C. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1953. (d) Ashton, P. R.; Boyd, S. E.; Brown, C. L.; Jayaraman, N.; Nepogodiev, S. A.; Stoddart, J. F. *Chem. Eur. J.* **1996**, *2*, 1115. (e) Kingery-Wood, J. E.; Williams, K. W.; Sigal, G. B.; Whitesides, G. M. *J. Am. Chem. Soc.* **1992**, *114*, 7303. (f) Aoi, K.; Itoh, K.; Okada, M. *Macromolecules*, **1995**, *28*, 5391. (g) Roy, R. *Trends Glycosci. Glycotech.* **1996**, *8*, 79. (h) Roy, R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 692. (i) Roy, R. *Polymer News*, **1996**, *21*, 226. (j) Kiessling, L. L.; Pohl, N. L. *Chem. Biol.* **1996**, *3*, 71. (k) Brüning, J.; Kiessling, L. L. *Tetrahedron Lett.* **1996**, *37*, 2907 and references cited therein.

9. (a) Glick, G. D.; Knowles, J. R. *J. Am. Chem. Soc.* **1991**, *113*, 4701. (b) Mortell, K. H.; Weathermann, R. S.; Kiessling, L. L. *J. Am. Chem. Soc.* **1996**, *118*, 2297.

10. (a) Meldal, M.; Christiansen-Brams, I.; Christensen, M. K.; Mouritsen, S.; Bock. K. In *Complex Carbohydrates in Drug Research;* Bock, K.; Clausen, H., Eds.; Munksgaard: Copenhagen, 1994; pp 153–167. (b) Voyer, N.; Lamothe, J. *Tetrahedron* **1995**, *51*, 9241.

11. Bertrozzi, C. R.; Hoeprich Jr., P. D.; Bednarski, M. D. *J. Org. Chem.* **1992**, *57*, 6092.

12. (a) Meldel, M. In *Neoglycoconjugates: Preparation and Application*; Lee, Y. C.; Lee, R. T., Eds.; Academic: San Diego, 1994; pp 145–198. (b) Christiansen-Brams, I.; Meldal, M.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1993**, 1461. (c) Cervigni, S. E.; Dumy, P.; Mutter, M. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1230. (d) Saha, U. K.; Roy, R. *Chem. Commun.* **1995**, 2571. (e) Peters, S.; Bielfeldt, T.; Meldal, M.; Bock, K.; Paulsen, H. *J. Chem. Soc., Perkin Trans. 1* **1992**, 1163.

13. For a discussion on the applications of oligosaccharide and glycoconjugat libraries see (a) Arya, P. Ben, R. N. *Angew. Chem. Int. Ed. Engl.* **1997**, in press and references cited therein. (b) Sofia, M. J. *DDT* **1996**, *1*, 27.

14. (a) Ginnis, A.; Sandhoff, K. *Tetrahedron Lett.* **1985**, *26*, 1479. (b) Abiko, A.; Roberts, J. C.; Takemasa, T.; Masamune, S. *Tetrahedron Lett.* **1986**, *27*, 4537. (c) Abdel-Magid, A. F.; Maryanoff, C. A.; Carson, K. G. *Tetrahedron Lett.* **1990**, *31*, 5595.

15. **Compound 1:** FABMS, calcd for $C_{46}H_{53}N_3NaO_{15}^+$ 910.3; found 910.4 ($M + Na^+$); ^1H NMR (600 MHz, $CDCl_3$) δ 1.40–1.90 (m, 6H), 2.04–2.10 (2s, 12H), 2.37–2.60 (ddd, J = 4.1, 9.8, 15.3 Hz, 2H), 3.25 (m, 2H), 4.00–4.29 (m, 6H), 4.40 (d, J = 6.2 Hz, 2H), 4.70 (m, 1H), 5.00–5.60 (m, 6H), 5.64 (d, J 7.4Hz, 1H), 6.63 (bs, 1H), 6.68 (bs, 1H), 7.14–7.32 (m, Ar-H, 13H); ^{13}C NMR ($CDCl_3$) δ 172.3, 170.7, 169.9, 169.8, 169.6, 156.3, 143.7, 141.2, 135.0, 129.0, 128.5, 127.7, 127.0, 126.7, 125.2, 125.0, 119.9, 69.2, 69.0, 68.7, 67.8, 67.0, 61.2, 54.3, 47.0, 41.2, 38.8, 31.2, 31.1, 28.6, 22.2, 20.6. **Compound 2:** FABMS, calcd for $C_{62}H_{78}N_3O_{23}^+$ 1232.5; found 1232.1 ($M + H^+$); Electrospray, 1232.4 ($M + H^+$); ^1H NMR (600 MHz, $CDCl_3$) δ 1.40–2.00 (m, 12H), 2.00–2.10 (4s, 24H), 2.40–2.53 (m, 6H), 4.00–4.43 (m, 12H), 5.06–5.10 (m, 1H), 5.17 (s, 2H), 5.20–5.28 (m, 4H), 5.40–5.49 (m, 4H), 7.26–7.78 (m , 13H); ^{13}C NMR ($CDCl_3$) δ 172.1, 170.9, 170.3, 169.7, 169.6, 169.1, 156.2, 143.8, 143.6, 141.1, 135.0, 128.4, 128.3, 128.1, 127.5, 126.9, 125.0, 121.9, 119.0, 68.0, 67.9, 67.1, 66.9, 61.1, 60.7, 54.1, 49.7, 49.1, 46.9, 33.5, 32.2, 28.4, 25.3, 22.4, 21.8, 21.2, 21.1, 20.4, 20.1. **Compound 3:** FABMS, calcd for $C_{62}H_{75}LiN_3O_{24}^+$ 1252.5; found 1252.5 ($M + Li^+$); Electrospray, 1246.4 ($M + H^+$); ^1H NMR (600 MHz, $CDCl_3$) δ 1.06–1.19 (m, 3H), 1.26–1.39 (m, 5H), 1.53–1.71 (m, 6H), 1.92–1.98 (m, 2H), 2.04, 2.05, 2.06, 2.09, 2.11, 2.12, 2.15 (7s, 24H), 2.46–2.60 (m, 2H), 3.31–3.56 (m, 3H), 3.71 (s, 2H), 3.98–4.31 (m, 7H), 4.66–4.77 (m, 1H), 5.10-5.47 (m, 7H),7.34–7.80 (m, 13H); ^{13}C NMR ($CDCl_3$) δ 171.8, 170.5, 170.0, 169.7, 169.6, 168.5, 156.1, 143.8, 143.7, 141.2, 128.9, 128.4, 128.2, 128.1, 127.6, 126.9, 125.0, 119.4, 70.1, 69.4, 69.1, 68.9, 68.4, 67.7, 67.4, 67.2, 67.0, 66.9, 61.4, 61.3, 53.5, 48.0, 47.0, 42.9, 41.1, 32.7, 31.8, 30.7, 28.4, 24.8, 23.6, 22.1. **Compound 4:** FABMS, calcd for $C_{64}H_{79}N_4O_{25}^+$ 1303.5; found 1303.1 ($M + H^+$); Electrospray, 1303.4 ($M + H^+$); ^1H NMR (600 MHz, $CDCl_3$) δ 1.04–1.76 (m, 10H), 1.80–1.99 (m, 4H), 2.0, 2.02, 2.03, 2.06, 2.08, 2.10, 2.12 (7s, 24H), 2.92–3.61 (m, 7H), 3.14–3.72 (m, 14H), 5.13–5.17 (m, 3H), 5.22–5.28 (m, 1H), 5.30–5.35 (m, 1H), 5.39–5.41 (m, 2H), 7.24–7.76 (m, 13H); ^{13}C NMR ($CDCl_3$) δ 172.4, 170.4, 169.9, 169.6, 169.5, 168.9, 156.1, 143.6, 141.1, 135.1, 128.4, 128.1, 127.5, 126.9, 126.2, 125.0, 119.8, 70.7, 69.1, 68.6, 68.3, 68.2, 67.7, 66.3, 66.2, 62.3, 61.3, 60.4, 54.2, 51.2, 48.9, 46.9, 45.2, 41.1, 33.6, 32.0, 31.7, 28.3, 24.7, 23.5, 22.1, 20.5.

16. Arya, P.; Shimizu, G. K. H.; Kutterer, K. M. K.; Roby, J. *unpublished results.*

(*Received in USA* 3 *February* 1997; *accepted* 13 *May* 1997)