Pergamon

Bioorganic & Medicinal Chemistry Letters 8 (1998) 1127–1132

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

DIVERSITY OF *C*-LINKED NEOGLYCOPEPTIDES FOR THE EXPLORATION OF SUBSITE-ASSISTED CARBOHYDRATE BINDING INTERACTIONS[1]

Prabhat Arya,* Kristina M. K. Kutterer, Huiping Qin,
Johanne Roby and Michael L. Barnes

Steacie Institute for Molecular Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario, K1A 0R6, Canada

Jin M. Kim, and René Roy

Department of Chemistry, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada

Received 9 January 1998; accepted 5 March 1998

Abstract: Diversity of α-galactose based *C*-linked neoglycopeptides (**1b, 2b, 3c, 4d,** and **5d**) has been developed to explore the importance of subsite-assisted carbohydrate binding interactions. Deprotected *C*-linked neoglycopeptides (**1b, 2b, 3c, 4d,** and **5d**) were synthesized and tested in competitive inhibition assays using a model enzyme-linked lectin (e.g., *Maclura pomifera*). Compound **2b**, with two α-galactoside units on the side chain of the lysine residue of the dipeptide backbone, exhibited a remarkable effect with a 2.82-fold increase in its inhibitory properties (IC_{50} 1.48 mM) in comparison to **1b** (IC_{50} 4.18 mM).

Molecular interactions between cell surface carbohydrate ligands and protein receptors are responsible for initiating a variety of biological and pathological processes (i.e., fertilization, microbial infections, inflammations and tumor metastasis).[2] Due to the weak nature of individual carbohydrate–protein interactions, understanding of such interactions, at the molecular level, are difficult to achieve. In natural systems, these weak interactions are usually employed in a cooperative manner which accounts for both the specificity and strength of the bindings.[3] To understand the importance of cooperativity in strengthening weak carbohydrate-protein interactions, well defined cooperative cell surface carbohydrate synthetic ligands could be used as probes for studying these interactions. Such probes could also be used for mapping the specific binding sites of a given protein receptor and could be developed further as inhibiting agents to prevent these interactions. Several synthetic approaches for the presentation of cooperative carbohydrate ligands on a variety of scaffolds have been developed and have exhibited similar or stronger bindings to the natural systems.[4,5] In a few cases, structural information on the complimentary binding site of a protein receptor has been obtained by using well defined neoglycoconjugates.[6]

To obtain molecular level understanding of carbohydrate–protein interactions, much emphasis in the past few years has been paid to the synthesis of small molecules as analogs of cell surface carbohydrate ligands.[7] With a similar goal, we are developing a flexible and control-oriented model for the bivalent presentation of terminal/exposed *C*-saccharide derivatives on a peptide/pseudo-peptide template (Figure 1).[8] This model has been designed to allow for variation in the presentation of the saccharide derivatives. Using this model, our aim is to synthesize short, preferably bivalent, *C*-linked neoglycopeptides in a highly controlled and flexible manner,

0960-894X/98/$19.00
PII: S0960-894X(98)00182-6

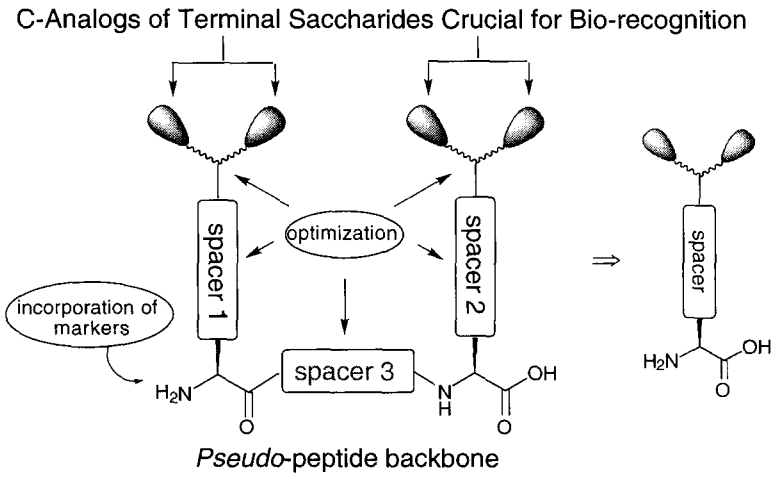
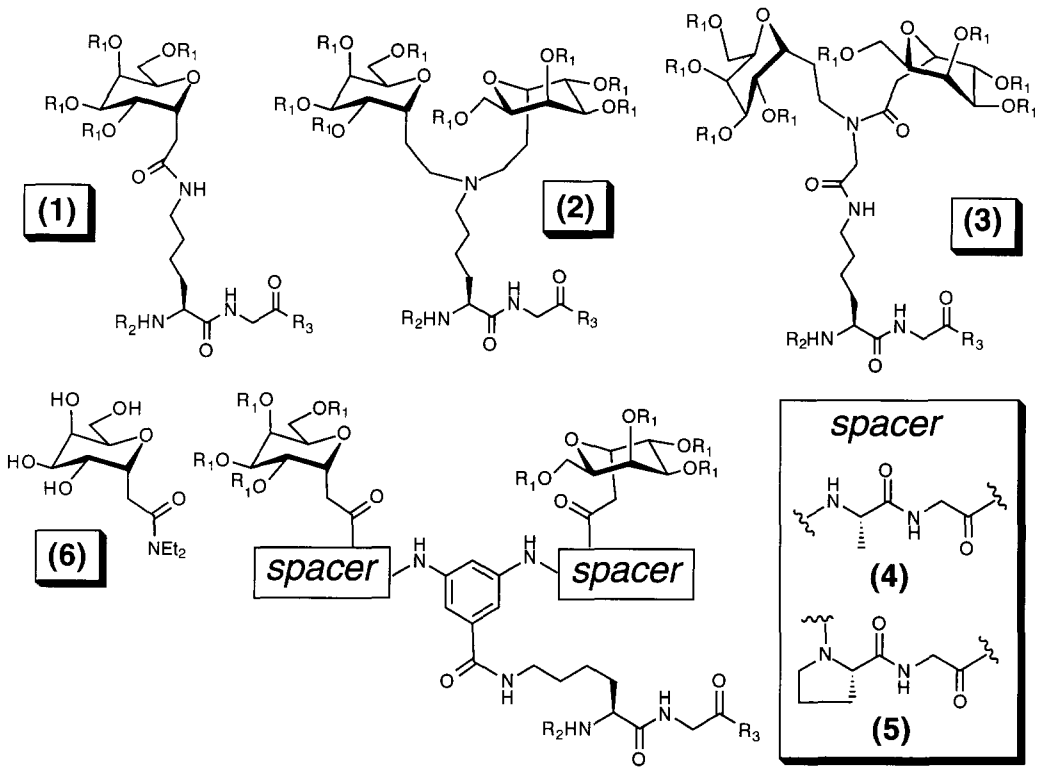


Figure 1: Model for the Presentation of Terminal/Exposed Saccharides on a Peptide/Pseudo-peptide Template

and to demonstrate the importance of optimization of the presentation of the saccharide derivatives. The peptide/pseudo-peptide portion of the *C*-linked neoglycopeptides may assist in secondary interactions with adjacent amino acid residues of protein receptors. The model shown in Figure 1 could be developed using solution phase synthesis of the building blocks, followed by building block assembly on solid phase.[9] As an alternative to the solid phase building block assembly approach, coupling of the *C*-glycoside derivative to resin-bound peptides/pseudo-peptides could also be utilized.[10] Initial studies in our group are focused on the bivalent presentation of saccharide derivatives and the examination of the effect of variable presentation in a model enzyme-linked lectin based competitive inhibition assay, using the plant lectin *Maclura pomifera*.[11]

For the synthesis of *C*-linked neoglycopeptides, α-galactose was selected as a saccharide derivative since it is the terminal saccharide of several natural ligands.[12] One such example is globotriosylceramide (GbOse3), which is responsible for the binding of Shiga and Shiga-like toxin receptors to host cells.[13] As a terminal saccharide, α-galactose is also shown to be present on the ovulated egg extra-cellular coat, the zona pellucida (ZP), and seems to be involved in recognition between mouse sperm and the zona pellucida.[14]

Unlike *O*-linked glycopeptides that are sensitive to acidic and basic media, *C*-linked neoglycopeptides are relatively stable and easy to synthesize.[9,15] In the past, synthesis of the fully protected derivatives of *C*-neoglycopeptides (**1a–3a**) has been reported by our group.[8a] Herein, we report a solution-phase synthesis of a few short *C*-neoglycopeptides (**2–5**) possessing variable bivalent presentation of α-galactose derivatives. Fully deprotected *C*-Neoglycopeptides (**1b, 2b, 3c, 4d,** and **5d**) were purified on a C-18 column by reverse-phase HPLC and characterized by MS-electrospray and NMR spectroscopy.[16]



(a) R_1: Ac; R_2: Fmoc; R_3: OBn; (b) R_1: H; R_2: Ac; R_3: NH_2;
(c) R_1: H; R_2: Ac; R_3: $N(Et)_2$; (d) R_1: H; R_2: Ac; R_3: $NHCH_2Ph$

Figure 2: Molecular Diversity of C-Linked Neoglycopeptides (1–5)

To explore the importance of subsite-assisted carbohydrate binding interactions, a model enzyme-linked lectin assay (ELLA) was examined in competitive inhibition assays using *C*-linked neoglycopeptides (**1b, 2b, 3c, 4d,** and **5d**). Horseradish peroxidase-labeled plant lectin *Maclura pomifera,* isolated from the seeds of Osage orange tree, was allowed to bind to a water-soluble poly(acrylamide-co-allyl α-D-galactopyranoside) used as a lectin-specific coating ligand in microtiter plates.[17] The effective inhibitory concentrations (IC_{50} values) were then measured and compared with methyl α-D-galactopyranoside as a reference standard (IC_{50} 2.18 mM) (Table 1).[18]

Table 1: *Results from the inhibition of binding of horseradish peroxidase-labelled Maclura Pomifera to poly(acrylamide-co-allyl α-Galactopyranoside) by C-Neoglycopeptides (**1b, 2b, 3c, 4d,** and **5d**)*

Compound (MW)	**IC_{50} (mM)**
1b (448)	4.18
2b (624)	1.48

3c (751)	4.31
4d (1132)	2.84
5d (1184)	1.50
6, control (277)	2.57
Me-α-D-Gal (194)	2.18

The replacement of an *O*- by *C*-glycoside from compound **6** (IC_{50} 2.57 mM) resulted in a marginal effect. However, the addition of a dipeptide lysine-glycine backbone (**1b**) greatly diminished the binding affinity for the lectin (IC_{50} 4.18 mM). Compound **2b**, with two α-galactoside units on the side chain of the lysine residue of the dipeptide backbone, exhibited a most remarkable effect with a 2.82–fold increase in its inhibitory properties (IC_{50} 1.48 mM). On the contrary, divalent derivative **3c**, which present the two α-D-galactoside residues, in a more rigid manner showed greatly reduced inhibitory properties. Interestingly, a minor change in the peptide residue in *C*-neoglycopeptide **4d** to **5d**, in which alanine was replaced by proline, resulted in an enhanced inhibitory effect by a factor of 2. To rule out the possibility of cross linking, the assay was performed in a concentration lower than required to observe such effects. Cooperativity, is not likely playing a role in the cases with high binding affinity, since the distances between the saccharide moieties in the divalent species are probably too close to reach two different binding sites from the same lectin.

To summarize, solution–phase synthesis of *C*-neoglycopeptides for the variable presentation of α-galactose derivatives has been achieved. The preliminary inhibition results strongly reflect upon the importance of the optimization of spacers for the presentation of saccharide derivatives. The enhanced binding affinity shown by compound **2b**, created by attaching two α-D-galactoside residues to the side chain of a lysine moiety, is highly surprising. Although the exact explanation of this effect is not clear at this stage. The results strongly suggest that the second unbound glycopeptide residue acts synergistically to increase the overall binding, possibly through subsite-assisted binding interactions.

References and notes:

1. (a) E-mail address for PA: Prabhat.Arya@nrc.ca. (b) NRCC publication no. 50682. (c) Presented at the 15th American Peptide Symposium, Nashville, Tennessee, U.S.A. June 14-19, 1997 and 2nd Canadian Combinatorial Meeting, Montreal, October 6-7, 1997.

2. (a) Varki, A. *Glycobiology* **1993**, *3*, 97. (b) Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *28*, 321. (c) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683. (d) Sears, P. ; Wong, C.-H. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 12086.

3. (a) Kiessling, L. L.; Pohl, N. L. *Chem. Biol.* **1996**, *3*, 71 and references therein. (b) Liang, R. ; Loebach, J.; Horan, N.; Ge, M.; Thompson, C.; Yan, L.; Kahne, D. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10554.

4. (a) Roy, R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 692. (b) Kanai, M.; Mortell, K. H.; Kiessling, L. L. *J. Am. Chem. Soc.* **1997**, *119*, 9931 and references therein. (c) Kingery-Wood, J. E. ; Williams, K. W.; Sigal, G. B.; Whitesides, G. M. *ibid.* **1992**, *114*, 7303.

5. (a) Sigal, G. B.; Mammen, M.; Dahmann, G.; Whitesides, G. M. *J. Am. Chem. Soc.* **1996**, *118*, 3789. (b) Spevak, W.; Nagy, J. O.; Charych, D. H.; Schaefer, M. E.; Gilbert, J. H.; Bednarski, M. D. *ibid.* **1993**, *115*, 1146. (c) Toogood, P. L.; Galliker, P. K.; Glick, G. D.; Knowles, J. R. *J. Med. Chem.* **1991**, *34*, 3140.

6. (a) Glick, G. D.; Knowles, J. R. *J. Am. Chem. Soc.* **1991**, *113*, 4701. (b) Glick, G. D.; Toogood, P. L.; Wiley, D. C.; Skehel, J. J.; Knowles, J. R. *J. Biol. Chem.* **1991**, *266*, 23660.

7. (a) Hansen, H. C.; Haataja, S.; Finne, J.; Magnusson, G. *J. Am. Chem. Soc.* **1997**, *119*, 6974. (b) Wong, C.-H.; Moris-Varas, F.; Hung, S.-C.; Marron, T. G.; Lin, C.-C.; Gong, K. W.; Weitz-Schmidt, G. *ibid.* **1997**, *119*, 8152.

8. (a) Arya, P.; Dion, S.; Shimizu, G. K. H. *Bioorg. & Med. Chem. Lett.* **1997**, *7*, 1537. (b) Roy, R.; Saha, U. K. *Chem. Commun.* **1996**, 201. (c) Kim, J. M.; Roy, R. *Tetrahedron Lett.* **1997**, *38*, 3487.

9. Arya, P.; Roby, J.; Kim, J. M.; Roy, R. *J. Org. Chem., submitted for publication.*

10. Kutterer, K. M. K.; Barnes, M. L.; Arya, P. *J. Org. Chem., submitted for publication.*

11. *The Lectins*; Liener, I.; Sharon, N.; Goldstein, I. J. Eds.; Academic Press: New York, 1986.

12. Kihlberg, J.; Magnusson, G. *Pure Appl. Chem.* **1996**, *68*, 2119.

13. (a) Nyholm, P.-G.; Magnusson, G.; Zhang, Z.; Norel, R.; Boyd, B. B.; Lingwood, C. A. *Chem. Biol.* **1996**, *3*, 263 and references therein. (b) St. Hilaire, P. M.; Boyd, M. K.; Toone, E. J. *Biochemistry*, **1994**, *33*, 14452.

14. Litscher, E. S.; Juntunen, K.; Seppo, A.; Penttilä, L.; Niemelä, R.; Renkonen, O.; Wassarman, P. M. *Biochemistry*, **1995**, *34*, 4662.

15. Kihlberg, J.; Elofsson, M. *Curr. Med. Chem.* **1997**, *4*, 85 and references therein.

16: **Compound 1b:** ^{1}H NMR (400 MHz, D_2O): δ 1.30–1.47 (m, 2H), 1.47–1.57 (m, 2H), 1.65–1.88 (m, 2H), 2.04 (s, 3H), 2.53–2.71 (m, 2H), 3.19 (t, J = 6.5 Hz, 2H), 3.64–3.86 (m, 4H), 3.89 (d, J = 9.2 Hz, 2H), 3.94–4.06 (m, 2H), 4.19–4.25 (m, 1H), 4.43–4.51 (m, 1H); ^{13}C NMR (100 MHz, D_2O): δ 21.9, 22.7, 28.3, 30.7, 32.7,

39.5, 42.4, 54.6, 61.2, 68.0, 69.1, 70.0, 72.8, 73.3, 173.9, 174.5, 175.1, 175.5; LRMS (electrospray, H_2O, positive ion mode, m/z) for $C_{18}H_{32}N_4O_9$: 449 $(MH)^+$ 471 $(M^+ + Na)$.

Compound 2b: ^1H NMR (400 MHz, D_2O): δ 1.40–1.56 (m, 2H), 1.72–1.94 (m, 4H), 2.02–2.22 (m, 4H), 2.06 (s, 3H), 3.20–3.44 (m, 6H), 3.60–3.82 (m, 8H), 3.90-4.06 (m, 6H), 4.11–4.18 (m, 2H), 4.25–4.31 (m, 1H); ^{13}C-NMR (100 MHz, D_2O): δ 19.7, 22.0, 22.5, 23.2, 30.6, 42.4, 51.1, 53.3, 54.2, 61.4, 68.1, 69.1, 70.0, 73.1, 73.3, 174.4, 175.0, 175.1; LRMS (electrospray, H_2O, positive ion mode, m/z) for $C_{26}H_{48}N_4O_{13}$: 625.2 $(MH)^+$, 647 $(M^+ + Na)$.

Compound 3c: ^1H NMR (400 MHz, D_2O): δ 0.97 (t, J = 7.1 Hz, 3H), 1.08 (t, J = 7.1 Hz, 3H), 1.24–1.30 (m, 2H), 1.37–1.43 (m, 2H), 1.58–1.76 (m, 2H), 1.83–1.93 (m, 2H), 1.93 (s, 3H), 2.58–2.80 (m, 2H), 3.07–3.14 (m, 2H), 3.24 (q, J = 7.1 Hz, 2H), 3.26 (q, J = 7.1 Hz, 2H), 3.42–3.52 (m, 2H), 3.54–3.74 (m, 9H), 3.83–3.98 (m, 4H), 3.94 (s, 2H), 3.98 (s, 2H), 4.18 (dd, J = 5.2, 8.8 Hz, 1H), 4.47 (t, J = 6.1 Hz, 1H); ^{13}C NMR (100 MHz, D_2O): δ 12.4, 13.2, 22.0, 22.7, 23.2, 28.3, 30.5, 31.0, 39.4, 41.1, 41.4, 42.2, 47.1, 50.2, 54.1, 61.0, 61.7, 68.1, 68.8, 69.5, 70.0, 70.2, 72.4, 72.6, 73.2, 73.4, 73.8, 169.3, 171.1, 174.3, 174.8, 174.9; LRMS (electrospray, H_2O, positive ion mode, m/z) for $C_{32}H_{57}N_5O_{15}$: 752.3 $(MH)^+$, 774.2 $(M^+ + Na)$.

Compound 4d: ^1H NMR (400 MHz, D_2O): δ 1.45 (br d, J = 7.2 Hz 8H), 1.59–1.67 (m, 2H), 1.72–1.85 (m, 2H), 1.95 (s, 3H), 2.62–2.68 (dd, J = 14.8 Hz, 2H), 2.74–2.80 (dd, J = 14.8 Hz, 2H), 3.38 (t, J = 6.8 Hz, 2H), 3.62–3.79 (m, 6H), 3.88–3.92 (m, 4H), 3.98 (br s, 2H), 4.01–4.04 (m, 2H), 4.05–4.11 (m, 4H), 4.25 (t, J = 7.1 Hz, 1H), 4.26–4.39 (m, 4H), 4.47–4.53 (m, 2H), 7.23 (d, J = 7.4 Hz, 2H), 7.31 (t, J = 7.3 Hz, 1H), 7.37 (t, J = 7.5 Hz, 2H), 7.65 (d, J = 1.7 Hz, 2H), 7.83–7.89 (m, 1H). ^{13}C NMR (100 MHz, D_2O): δ 17.2, 22.0, 23.1, 28.7, 32.0, 33.0, 40.5, 43.6, 43.9, 51.3, 55.1, 62.1, 73.7, 118.0, 120, 127.9, 128.3, 129.6, 136.0, 141.0, 170.7, 176.8. LRMS (electrospray, H_2O, positive ion mode, m/z) for $C_{50}H_{72}N_{10}O_{20}$: 1132.3 (M^+), 1155.5 $(M^+ + Na)$.

Compound 6: ^1H NMR (400 MHz, D_2O): δ 0.98 (t, J = 7.2Hz, 3H), 1.09 (t, J = 7.2Hz, 3H), 2.62, 2.71 (ABX, J = 4.4, 9.3H, 15.9 Hz, 2H), 3.26 (q, J = 7.2 Hz, 2H), 3.34 (q, J = 7.2 Hz, 2H), 3.56 (ABX, J = 4.4, 7.6, 11.6 Hz, 2H), 3.69 (dd, J = 3.2Hz, 1H), 3.72–3.74 (m, 1H), 3.85–3.90 (m, 1H), 3.90 (dd, J = 9.6 Hz, 1H), 4.46 (dt, J = 5.6, 1H); ^{13}C NMR (100 MHz, D_2O): δ 12.4, 13.7, 29.6, 41.4, 43.4, 61.0, 68.3, 68.9, 70.1, 72.7, 73.4, 172.8; LRMS (electrospray, H_2O, positive ion mode, m/z) for $C_{12}H_{23}NO_6$: 278.1 $(MH)^+$, 300.1 $(M^+ + Na)$.

17. The coating of copolyacrylamide, containing a galactoside to acrylamide ratio of 1:5 was prepared from allyl α-*D*-galactopyranoside according to Horejsi, V.; Kocourek, J. *Methods Enzymol.* **1974**, *34*, 361.

18. IC_{50} (mM) for all the compounds shown in table 1 is an average value of three experiments.