Carbohydrate Research, 251 (1994) 285–301
Elsevier Science B.V., Amsterdam

Chemical and enzymatic synthesis of multivalent sialoglycopeptides

Carlo Unverzagt [a,b,*], Soerge Kelm [a,c] and James C. Paulson [a,d]

[a] *Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, CA 90024-1737 (USA)*
[b] *Organisch-Chemisches Institut der Technischen Universität München, Lichtenbergstr. 4, 85748 Garching (Germany)*
[c] *Biochemisches Institut II der Universität Kiel, Olshausenstr. 40, 24098 Kiel (Germany)*
[d] *Cytel Corp. and Department of Chemistry, Scripps Research Institute, 3525 John Hopkins Court, La Jolla, CA 92121 (USA)*

(Received August 7th, 1992; accepted in revised form June 9th, 1993)

ABSTRACT

Linear and branched glycopeptides containing multiple sialyl-*N*-acetyllactosamine side chains have been synthesized using a combined chemical and enzymatic approach. Peptide backbones in which β-GlcNAc-Asn residues were incorporated were obtained in good yields by optimized solid-phase synthesis following the Boc strategy. The resulting multivalent glycopeptides were galactosylated in near-quantitative yields using bovine galactosyltransferase, UDP-galactose, and calf alkaline phosphatase that destroys the inhibiting side product UDP. Subsequent enzymatic sialylation yielded the desired glycopeptides containing asparagine-linked sialyl-*N*-acetyllactosamine side chains. The compounds were characterized by ^{1}H NMR and FABMS. Recombinant sialyltransferase and CMP–sialate synthetase were used for the enzymatic synthesis of sialosides on a preparative scale.

The synthetic glycopeptides were tested as inhibitors of influenza virus to cells, revealing that most of the multivalent sialoglycopeptides exhibit increased binding that depends on the spacing when compared to monovalent compounds. A possible mechanism for increased binding is proposed.

INTRODUCTION

Carbohydrates are increasingly recognized as playing multiple roles in biological recognition[1]. For example, they mediate targeting of enzymes to the lysosomes via a mannose-6-phosphate receptor, the clearance of glycoproteins from the blood by carbohydrate-binding proteins in the liver, and adhesion of cells, bacteria, and viruses to animal cells.

A well studied example is the binding of influenza virus, which recognizes terminal sialic acid present on glycoproteins and glycolipids[2]. With a high number of receptors on its surface, influenza viruses can bind their target cells very tightly,

* Corresponding author.

SSDI 0008-6215(93)E0220-U

even though the individual receptor–ligand bond[3] is weak. Adhesion occurs via a sialic acid binding protein, the hemagglutinin, a trimeric transmembrane glycoprotein whose crystal structure, together with sialic acid in the receptor binding pocket, has been determined[4]. A very potent inhibitor of virus binding was found in the glycoprotein α_2-macroglobulin. By comparing the inhibition of α_2-macroglobulin from different species[5] and from other glycoproteins, it was proposed that the overall number of sialic acid residues present, as well as their spacial arrangement, are of particular importance for the inhibitory potency of a glycoprotein. To systematically examine the influence of multivalency and spacing of sialosides on their binding to the influenza virus hemagglutinin, we synthesized a series of divalent glycopeptide model compounds. We chose glycopeptides as a model system because they can be easily synthesized, and their structures are readily analyzed by NMR techniques.

Understanding of multivalent interaction of carbohydrate binding proteins may lead to drugs to combat influenza virus infection and may provide insights to applications in the regulation of carbohydrate-dependent[1f,g] activities of cells within the immune sytem. (Note that after completion of this work, a number of reports using α-sialosides linked to polyacrylamide as influenza virus inhibitors have appeared[6].)

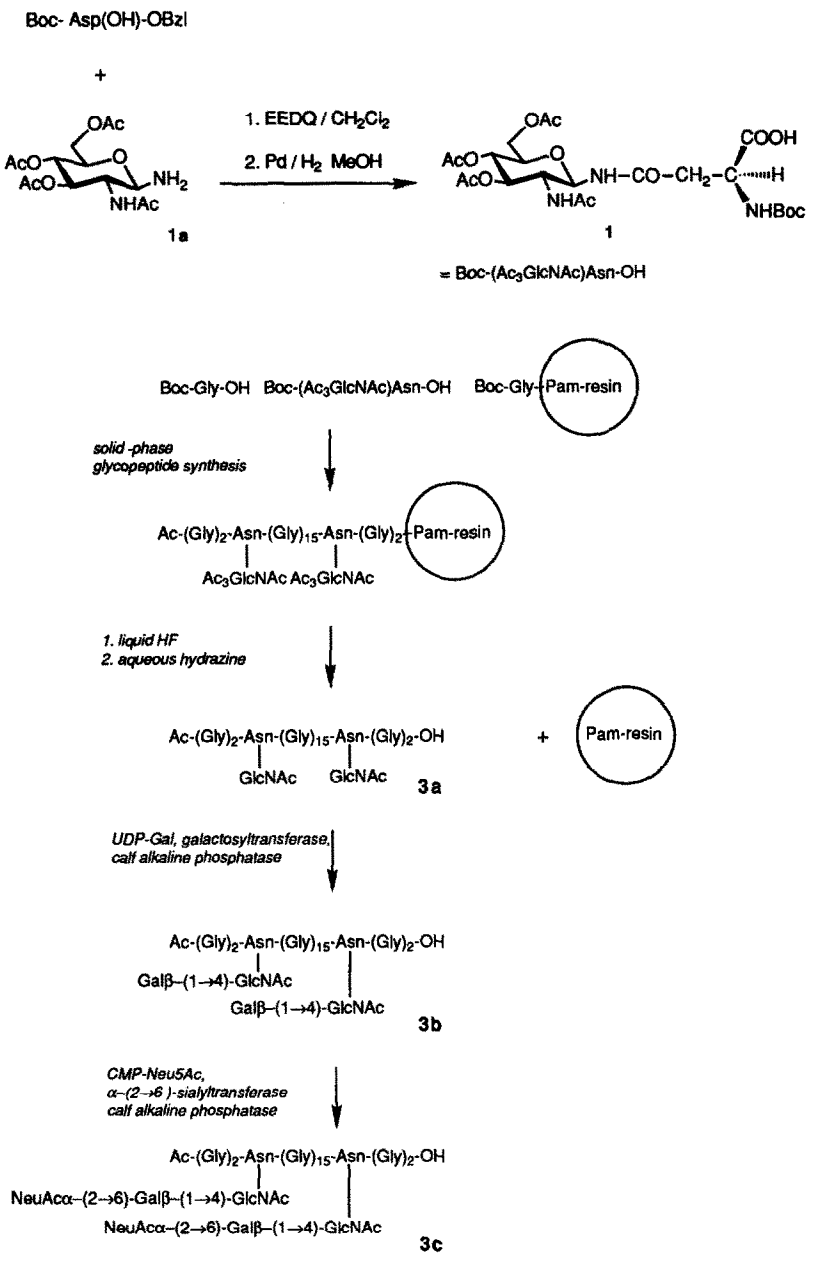
EXPERIMENTAL

Materials—Boc–glycyl–Pam resin and *N*-methylpyrrolidone was purchased from Applied Biosystems. Other solvents needed for solid-phase peptide synthesis were from Fisher Scientific Co. Boc amino acids were obtained from Bachem Biochemica GmbH. Ethyl-2-ethoxy-1,2-dihydroquinoline-1-carboxylate (EEDQ), diisopropylcarbodiimide (DIC), 1-hydroxybenzotriazole (HOBT), *p*-cresol, and *p*-thiocresol were from Aldrich Chemical Co. The reaction vessel and the shaker for solid-phase peptide synthesis were from Milligen-Biosearch. Thin-layer chromatography was performed on plates of Silica Gel-60-F254 (E. Merck, Darmstadt) and visualized by spraying with N H_2SO_4 in EtOH containing 0.1% orcinol. Sephadex G-25–50, UDP-glucose, UDP-glucose-4′-epimerase (EC 5.1.3.2), bovine galactosyltransferase (EC 2.4.1.22) and bovine serum albumin were obtained from Sigma Chemical Co. Calf intestinal alkaline phosphatase (EC 3.13.1) and inorganic pyrophosphatase (EC 3.6.1.1) were obtained from Boehringer Mannheim. Recombinant CMP–sialate synthetase[7] (EC 2.7.7.43) was a generous gift from Dr. James Rasmussen of Genzyme. The β-D-Gal-(1 → 4)-D-GlcNAc-(2 → 6)-α-sialyltransferase (E C 2.4.99.1) was isolated as previously described[8a]. Recombinant sialyltransferase was obtained according to the published procedure[8b]. DEAE Sephadex A-25 and Sephacryl S-200 HR were from Pharmacia LKB Biotechnology, Inc. ^{1}H NMR spectra were recorded in D_2O on a Bruker WP 360. FABMS spectra (negative-ion mode) were recorded on a Finnigan MAT 95Q using a thioglycerol–acetic acid matrix. Sialoglycopeptide **2c** was synthesized as previously described[9].

Glycopeptide synthesis.—Boc-(β-D-GlcNAcAc$_3$)Asn-OBzl was synthesized from β-D-GlcNAcAc$_3$NH$_2$ (**1a**) and Boc-Asp-OBzl using EEDQ as the coupling reagent[10]. The benzyl ester was removed by catalytic hydrogenation[11] to yield Boc-(β-D-GlcNAcAc$_3$)Asn-OH (**1**), (Scheme 1). Glycopeptides were obtained by solid-phase synthesis following the Boc strategy. Boc–glycine–Pam resin (0.5 mmol) was deprotected and neutralized as described[12]. Coupling was performed in *N*-methylpyrrolidone using 4 equiv of amino acid, hydroxybenzotriazole, and diisopropylcarbodiimide. The amino acids were preactivated at 0°C for 30 min. Coupling efficiency was monitored by the Kaiser test[12]. Boc-serine was protected with a benzyl ether in the side chain. After introducing the branching *Nα*,*Nε*(Boc)$_2$lysine in compound **11**, the amount of coupling reagents was doubled. For the β-D-GlcNAc-Asn derivative **1**, 1.2 equiv. were used, and coupling to the resin was allowed to proceed until the reaction was complete (1–24 h). Similar reaction times were employed for proline couplings. N-Terminal acetylation was performed with Ac_2O in pyridine (1:2 vol). The glycopeptides (0.5 mmol) were cleaved from the resin with liquid HF (20 mL) in a Teflon apparatus using *p*-thiocresol and *p*-cresol (0.5 mL each) as scavengers. After 0.5 h at −15°C and 0.5 h at 0°C, the HF was removed under high vacuum. The residue was repeatedly washed with diethyl ether and subsequently extracted with 50 mL of 0.1 M NH_4HCO_3. After lyophilization of the aqueous solution, the crude glycopeptides were deacetylated[13] with 20% aq hydrazine for 1 h at ambient temperature. Hydrazine was removed by addition of acetone at 0°C and subsequent distillation under reduced pressure. The remainder was purified on a Sephadex G-25–50 column (2.6 × 30 cm) with 0.1 M NH_4HCO_3 as eluent. Glycopeptides were detected by TLC (1.5:1 2-propanol–M NH_4OAc) using the orcinol reagent for visualization. The yields for the pooled and lyophilized glycopeptides ranged between 50 and 80%.

Enzymatic glycosylation of glycopeptides.—(*a*) *Galactosylation.* The glycopeptides **3a–11a** (20 μmol) and bovine serum albumin (1 mg) were dissolved in 1 mL of 50 mM sodium cacodylate, pH 7.4, containing 50 μmol of UDP-glucose, 1 mmol of $MnCl_2$, 7 mmol of NaN_3, 400 milliunits of galactosyltransferase, 1 unit of UDP-glucose-4′-epimerase, and 12 units of calf intestinal alkaline phosphatase. The mixture was placed in a 37°C incubator, and the pH was maintained at 7.4 by periodic addition of 0.25 M NaOH. After 48 h the precipitate that formed was removed by centrifugation. The supernatant was separated on a Sephadex G-25–50 column (2.6 × 30 cm) using 0.1 M NH_4HCO_3 as eluent. The glycopeptides were detected by TLC (1.5:1 2-propanol–M NH_4OAc), and subsequent lyophilization of the positive fractions gave the desired digalactoglycopeptides **3b–11b**. The yields are listed in Table I.

(*b*) *Sialylation.* The digalactoglycopeptides **3b–11b** (10 μmol) and bovine serum albumin (1 mg) were dissolved in 1 mL of distilled water containing 30 μmol of CMP-NeuAc, 7 mmol of NaN_3, 100 milliunits of sialyltransferase, and 12 units of calf intestinal alkaline phosphatase. The reaction was performed at 37°C, and pH



Scheme 1.

7.4 was maintained by periodic addition of 0.25 M NaOH. After two days, the mixture was applied to a column of Sephacryl S-200 HR (1.6 × 56 cm) and eluted with 0.1 M NH_4HCO_3. The product was detected by TLC (1.5 : 1 2-propanol–M NH_4OAc), pooled and lyophilized. Further purification was achieved by ion-ex-

change chromatography as follows: 5 μmol of the crude sialoglycopeptide were dissolved in 5 mL of distilled water and loaded onto a column of DEAE Sephadex A-25 (formate form, 1.5 × 4 cm). After washing with 15 mL of distilled water, a 40 mL gradient of 0–0.5 M NH_4HCO_3 was passed over the column. Monosialylated glycopeptides eluted faster than the disialylated compounds. The fractions were examined by TLC (1.5:1 2-propanol–M NH_4OAc) and checked for purity by NMR spectroscopy. Fractions containing the disialosides were pooled and lyophilized. Yields are listed in Table I.

Compound **9c** was also obtained in equal yield by using recombinant (2 → 6)-α-sialyltransferase[8b] instead of the enzyme purified from rat liver. The sialic acid content of glycopeptides **2c**–**11c** was determined by the thiobarbituric acid method[14] after mild acid hydrolysis[15].

Synthesis of CMP-Neu5Ac with recombinant CMP-Neu5Ac-synthetase.—A 620-mg portion (2 mmol) of *N*-acetylneuraminic acid and 1.1 g (2 mmol) of cytidinetriphosphate disodium salt were added to 65 mL of 100 mM $MgCl_2$ containing 3 mmol of tris(hydroxymethyl)aminomethane, 33 mg of bovine serum albumin, 13 μL of β-mercaptoethanol and 0.3 mmol of sodium azide. After adjusting the solution to pH 8 by addition of 0.25 M NaOH, the reaction was started by addition of 2 mL of recombinant CMP-sialate synthetase (Genzyme, the 3.1 U/mL supension in 60% $(NH_4)_2SO_4$ was shaken prior to use) and 40 U of inorganic pyrophosphatase. The pH of the reaction was maintained by periodic addition of 0.25 M NaOH. After no further drop of pH was observed (24–30 h), 200 U of calf alkaline phosphatase were added, followed by the addition of 0.25 M NaOH to keep the pH above pH 8. When the reaction was terminated (3–4 h), the crystalline phosphate precipitate was removed by centrifugation in a clinical centrifuge, and the supernatant was decanted. After triturating the precipitate with 20 mL of 20 mM NH_4OH and centrifugation, the aqueous phases were combined and lyophilized. The remainder was dissolved in a small volume of 20 mM NH_4OH, loaded onto a column of Biogel P-2 (1.6 × 80 cm) and eluted with the same buffer. Fractions (5 mL) containing sialic acid (orcinol–Fe^{3+} assay) were pooled, adjusted to pH 9 with 0.25 M NaOH, and lyophilized. The yield of crude CMP-SA was 85% (determined by A_{271}), which was sufficiently pure for the sialyltransferase reactions.

Bioassays.—Virus adsorption inhibition using the recombinant influenza virus X-31 (H_3N_2) was performed as described[16]. Relative inhibitory potencies were calculated from the concentration of inhibitor necessary for 50% inhibition. Compound **2** as the monovalent reference substance was given as the relative inhibitory potency of 1.0.

RESULTS

Synthesis of glycopeptides.—Glycopeptides can be synthesized using the classical solution techniques (for review, see ref 17), or by solid-phase methods[9,11,18], which have recently become increasingly useful through the development of mildly

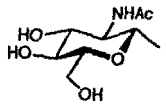
TABLE I

Structures, yields, and antiviral activity of compounds **2–11**

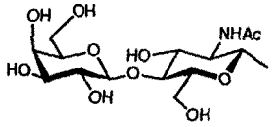
Compound	Structure	Yield of galactosylated Glycopeptide (%)	Yield of sialylated Glycopeptide (%)	Inhibitory potency of slalosides relative to **2c**	Concentration of glycopeptide for 50% inhibition [mM]
2a,b,c	H_2N-G-G-N(CHO)-G-G-OH		–[9]	1.0	2
3a,b,c	Ac-G-G-N(CHO)-G-G-G-G-G-G-G-G-G-G-G-G-G-G-N(CHO)-G-G-OH	n.d. [a]	n.d.	8.0	0.25
4a,b,c	Ac-G-G-N(CHO)-G-G-A-G-G-G-S-G-G-N(CHO)-G-G-OH	91	97	4.0	0.5
5a,b,c	Ac-G-G-N(CHO)-G-G-S-G-G-A-G-G-G-S-G-G-N(CHO)-G-G-OH	96	97	2.0	1
6a,b,c	Ac-G-G-N(CHO)-G-G-A-G-G-S-G-G-A-G-G-G-S-G-G-N(CHO)-G-G-OH	92	88	2.0	1
7a,b,c	Ac-G-G-N(CHO)-G-G-S-G-G-A-G-G-S-G-G-A-G-G-G-S-G-G-N(CHO)-G-G-OH	92	58	1.3	1.5

8a,b,c	Ac-N(CHO)-G-P-P-P-P-P-P-P-N(CHO)-G-G-OH	80	98	4.0	0.5
9a,b,c	Ac-N(CHO)-G-P-P-P-P-P-P-P-P-P-P-P-P-N(CHO)-G-G-OH	95	83	< 0.7	> 3.0
10a,b,c	Ac-N(CHO)-G-P-P-P-P-P-P-P-P-P-P-P-P-P-P-N(CHO)-G-G-OH	80	87	< 0.7	> 3.0
11a,b,c	Ac-N(CHO)-G-P-P-P-P-P-P-P K(Ac-N(CHO)-G-P-P-P-P-P-P-P)-G-OH	90	73	4.0	0.5

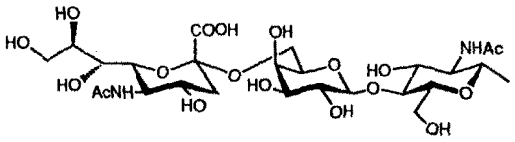
A= Alanine
G= Glycine
K=Lysine
N= Asparagine
P= Proline
S= Serine



a: CHO = (GlcNAc-)



b: CHO = (LacNAc-)



c: CHO = (Neu5AcLacNAc-)

[a] n.d., not determined.

TABLE II

^{1}H NMR data of synthetic glycopeptides **3–11** [a]

Compound	H-1 ($J_{1,2}$)	H-1′ ($J_{1',2'}$)	H-α ($J_{\alpha,\beta}$)	H-β Asn (J_{vic}, J_{gem})	H-2″*eq* ($J_{3eq'',3ax''}$; $J_{3eq'',4''}$)	NAc	H-2″*ax* (J_{vic})	CH_3 Ala ($J_{\alpha,\beta}$)
3c	5.2 (9.7)	4.44 (7.6)		2.88 m	2.67 (12.0; 4.3)	2.06; 2.03; 2.02	1.71 (12)	
4a	5.07 (9.6)		4.51 (5) Ser	2.87 m		2.07 Gly		1.42 (7.2)
	5.06 (9.6)		4.38 (7.2) Ala			2.02 GlcNAc		
4b	5.1 (9.7)	4.49 (7.8)	4.51 m Ser	2.87 m		2.08 Gly		1.43 (7.2)
			4.38 (8.2) Ala			2.08 GlcNAc		
4c	5.11 (9.8)	4.45 (7.7)	4.51 (5) Ser	2.87 m	2.67 (12.2; 4.5)	2.07 Gly	1.72 (12.2)	1.42 (7.3)
	5.11 (8.1)		4.38 (7.3) Ala			2.04; 2.03; Neu; GlcNAc		
5a	5.07 (9.6)		4.5 (4.4) Ser	2.87 m		2.08 Gly		1.42 (7.3)
	5.06 (9.6)		4.38 (7.3) Ala			2.02 GlcNAc		
5b	5.08 (9.8)	4.49 (7.7)	4.51 m Ser	2.87 m		2.07 Gly		1.42 (7.2)
	5.09 (9.8)		4.38 (7.2) Ala			2.02 GlcNAc		
5c	5.11 (9.9)	4.45 (7.7)	4.51 (5) Ser	2.87 m	2.68 (12.2; 4.3)	2.07 Gly	1.72 (12.2)	1.42 (7.1)
	5.11 (9.0)		4.38 (7.1) Ala			2.04; 2.03; Neu; GlcNAc		
6a	5.06 (9.6)		4.52 (5) Ser	2.87 m		2.08 Gly		1.42 (7.2)
	5.07 (9.6)		4.37 (7.2) Ala			2.02 GlcNAc		
			4.38 (7.2) Ala					
6b	5.1 (9.9)	4.49 (7.5)	4.51 m Ser	2.87 m		2.07 Gly	1.42 (7.2)	
			4.37 (7.2) Ala			2.02 GlcNAc		
			4.38 (7.2) Ala					
6c	5.11 (9.8)	4.45 (7.8)	4.51 (4.9) Ser	2.87 m	2.68 (12.0; 4.4)	2.07 Gly	1.72 (12.0)	1.42 (7.2)
	5.11 (9.0)		4.37 (7.2) Ala			2.04; 2.03; Neu; GlcNAc		
7a	5.08 (9.7)		4.52 (4.9) Ser	2.87 m		2.08 Gly		1.42 (7.3)
	5.07 (9.7)		4.38 (7.2) Ala			2.03 GlcNAc		

7b	5.09 (9.9)	4.49 (7.7)	4.51 m Ser 4.38 (7.1) Ala	2.87 m		2.07 Gly 2.02 GlcNAc	1.42 (7.1)	
7c	5.11 (9.8; 9.0) 5.11 (9.0)	4.45 (7.9)	4.51 (4.9) Ser 4.37 (7.3) Ala	2.87 m	2.68 (12.2; 4.4)	2.07 Gly 2.04; 2.03; Neu; GlcNAc	1.72 (12.2)	1.42 (7.3)
8a	5.06 (9.7) 5.04 (9.7)			2.86–2.82 m 3H 2.71 (7.7; 16.0) 1H		2.05; 2.03; 2.01		
8b	5.07 (9.6) 5.05 (9.6)	4.47 (7.5)		2.86–2.81 m 3H 2.72 (8.1; 15.9) 1H		2.03; 1.99		
8c	5.12 (9.4) 5.10 (9.5) 5.03 (9.7)	4.45 (8.1)		2.88–2.82 m 3H 2.86–2.82 m 3H 2.71 (8.0; 16.0) 1H	2.68 (4.5; 12.2)	2.05; 2.04; 2.03 2.03; 2.02; 1.99	1.72 (12.2)	
9b	5.08 (9.6) 5.06 (9.6)	4.47 (7.6)		2.88–2.82 m 3H 2.73 (7.9; 16.1) 1H		2.05; 2.04; 2.01		
9c	5.12 (9.6) 5.10 (9.6)	4.45 (7.8)		2.87–2.83 m 3H 2.74 (7.6; 15.9) 1H	2.67 (4.6; 12.1)	2.04; 2.02	1.71 (12.1)	
10a	5.06 (9.8) 5.04 (9.6)			2.87–2.83 m 3H 2.71 (7.6; 15.7) 1H		2.05; 2.04; 2.01		
10b	5.07 (9.6) 5.05 (9.6)	4.46 (7.5)		2.86–2.81 m 3H 2.71 (8.1; 15.8) 1H		2.04;2 2.03; 2.0		
10c	5.12 (9.2) 5.11 (9.6)	4.45 (8.0)		2.87 m 3H 2.75 (8.0; 16.0) 1H	2.68 (4.6; 12.2)	2.06; 2.04; 2.03	1.72 (12.2)	
11a	5.07 (9.8)		4.16; 4.03 (17.1) Gly	2.86 (5.0; 16.0) 2.74 (8.2; 16.0)		2.05; 2.02		
11b	5.12 (9.7)	4.51 (8.1)	4.18; 4.07 (17) Gly	2.89 (5.0; 16.0) 2.77 (8.0; 16.0)		2.08; 2.05		
11c	5.15 (9.4)	4.48 (8.2)	4.19; 4.05 (17.4) Gly	2.89 (5.0; 16.0) 2.77 (7.7; 16.0)	2.7 (4.8; 12.3)	2.07; 2.06	1.74 (12.3)	

[a] Spectra were determined for solution in D_2O using the HOD signal (4.81 ppm) as a reference standard.

cleavable resin anchors[18a,c] that are compatible with labile O-glycosidic bonds. However, the N-glycosidic linkage between *N*-acetylglucosamine and asparagine is fairly stable towards acids and bases. The β-D-GlcNAc-Asn building block **1** was incorporated into peptides by solid-phase peptide synthesis using Boc strategy and a final hydrogen fluoride cleavage[11] (Scheme 1). After removal of the acetates on the sugar moiety with hydrazine, the carbohydrate side chains in the glycopeptides were elongated by glycosyltransferases. Our previous work[9] showed that sialic acid can be introduced in high yields into glycopentapeptides containing one β-D-GlcNAc-Asn moiety by using first (1 → 4)-β-D-galactosyltransferase and subsequently (2 → 6)-α-sialyltransferase in a one-flask reaction. In Scheme 1 the modified reaction sequence leading to multivalent sialoglycopeptides like **3c** is outlined.

The glycosyl amino acid Boc-(GlcNAcAc$_3$)Asn-OH (**1**) needed for the solid-phase synthesis was obtained on a 20-g scale by coupling the glycosylamine **1a** with Boc-Asp(OH)-OBzl, followed by catalytic hydrogenation of the benzyl ester. As a model glycopeptide, the 21-mer **3a**, containing 19 glycine and two β-D-GlcNAc-Asn residues was synthesized. The polyglycine spacer was chosen because of expected high conformational flexibility and favorable water solubility. When the synthesis of **3a** was performed on an automatic peptide synthesizer using the symmetric anhydrides of Boc amino acids and the Merrifield resin as a polymeric support, the reactions were incomplete and a significant amount of peptide was lost during the multiple acidic deprotection steps. Therefore, we changed to a resin with higher stability under the acidic conditions (Pam resin[19]) and conducted the synthesis manually. The use of *N*-methylpyrrolidone (NMP) as a solvent instead of *N*,*N*-dimethylformamide improved the efficiencies for the coupling of the fifteen consecutive glycine residues in **3a**. Generally 4 equivalents of activated Boc-amino acids were used in the coupling steps. To limit the consumption of the precious glycosylamino acid **1**, the use of a smaller excess was examined. We found, that couplings of the β-D-GlcNAc-Asn derivative **1** in NMP can be conducted with as low as 1.2 equivalents, which also gave complete reaction after allowing prolonged reaction times (1–24 h). The final glycopeptides were liberated from the Pam resin by HF cleavage, which did not affect the N-glycosidic bond or the carbohydrate acetates[11]. Subsequent mild O-deacetylation with aqueous hydrazine[13] gave the fully deprotected glycopeptides **3a**–**11a** shown in Table I. The crude glycopeptides were purified by gel filtration on Sephadex G-25. Typically, a synthesis with 0.5 mmol of starting amino acid yielded several hundred milligrams of glycopeptide. No further attempts were made to increase the purity of the glycopeptides at this stage because the later addition of sialic acid allowed better separation by use of ion-exchange chromatography.

To increase the coupling efficiency, a glycine residue was introduced prior to the N-terminal β-D-GlcNAc-Asn in the polyprolyl-glycopeptides **8**–**11**. This procedure also facilitates monitoring the coupling with the more sensitive ninhydrin reaction of the primary amino group.

The synthesis of the branched glycopeptide **11a** was shortened by using a

$N\alpha,N\epsilon(Boc)_2$-protected lysine derivative. After introduction of the branching point, the simultaneous extension of two peptide chains reduces the number of couplings by half. Lysine residues serving as branching points have been useful in the synthesis of multiple antigen peptides[20] and may also lead to glycopeptides with higher valency.

Enzymatic modification of synthetic glycopeptides.—Enzyme-catalyzed synthesis of sialyloligosaccharides[21] has been shown in numerous examples, but only a few reports exist concerning the action of glycosyltransferases on synthetic asparagine-linked glycosylaminoacids[22] and glycopeptides[9]. As a model compound for the synthetic pathway outlined in Scheme 1, a derivative of **3a** was synthesized without the terminal N-acetyl group. This divalent glycopeptide reacted poorly with the galactosyltransferase. ^{1}H NMR data showed that only one galactose unit had been transferred to the acceptor molecule. The low reactivity was presumably caused by the poor solubility of the acceptor in the reaction mixture. Acetylation of the N-terminus gave **3a** which exhibited improved solubility and gave higher yields in the galactosyltransferase reaction. Both *N*-acetylglucosamine residues in **3a** could be converted into *N*-acetyllactosamine units that are substrates for the $(2 \rightarrow 6)$-α-sialyltransferase[23]. Double sialylation of the digalactopeptide **3b** gave **3c**, a first example for a synthetic multivalent sialoglycopeptide.

To increase the solubility of the glycopeptides **4a–11a**, the N-terminus was generally acetylated, and the polyglycine sequences were interrupted by alanine and serine residues which appeared to reduce the self-aggregating tendency observed for **3a**.

The galactosyltransferase is subject to substrate and product inhibition[24]. Typically, the yields obtained in preparative enzymatic galactosylations were $\sim 60\%$ for good substrates like *N*-acetylglucosamine. Assuming a 60% yield per site for the glycopeptides **3a–11a**, the maximum yield would decrease to 36% for a bivalent compound. Higher yields in this crucial step were obtained by using the alkaline phosphatase-assisted glycosylation method[9], which allows the generation of sialyl-*N*-acetyllactosamine structures very efficiently by subsequently adding galactose and sialic acid in a one-flask reaction. In contrast to compound **2a**, which can be sialylated by this procedure (86% yield from **2a**), these reaction conditions had to be modified for the glycopeptides **3a–11a**. The one-flask procedure applied on a divalent glycopeptide results in salt concentrations higher than 100 mM for the sialylation step. The $(2 \rightarrow 6)$-α-sialyltransferase is significantly inhibited by high buffer concentrations[25] requiring the purification of the digalactosyl intermediates **3b–11b** before sialylation. Enzymatic galactosylation assisted by alkaline phosphatase gave excellent yields in the synthesis of the glycopeptides **4b–11b** as shown in Table I. The enzymatic glycosylation appears to be of high synthetic value especially for compounds with multiple acceptor sites.

Separation of the enzymatically modified glycopeptides from the mixture containing BSA and low-molecular-weight compounds such as the donor substrate was accomplished by gel filtration, which provided efficient separation and a high

TABLE III

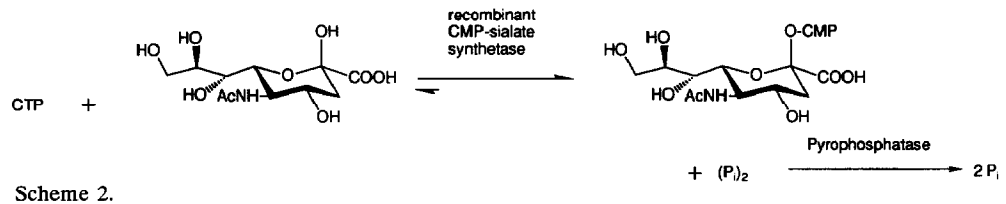
FABMS data of sialoglycopeptides **4c–11c** [a]

Compound	Formula	Nominal mass [b]	Found
4c	$C_{88}H_{139}N_{21}O_{56}$	2385.867	2386.7
5c	$C_{95}H_{150}N_{24}O_{60}$	2586.942	2587.2
6c	$C_{102}H_{161}N_{27}O_{63}$	2772.022	2773.3
7c	$C_{109}H_{172}N_{30}O_{67}$	2973.097	2973.2
8c	$C_{106}H_{161}N_{19}O_{53}$	2548.048	2549.3
9c	$C_{136}H_{203}N_{25}O_{59}$	3130.365	3131.1
10c	$C_{151}H_{224}N_{28}O_{62}$	3421.253	3422.4
11c	$C_{154}H_{233}N_{29}O_{63}$	3494.576	3495.1

[a] Spectra were determined in the negative-ion mode using thioglycerol–acetic acid as the matrix.
[b] The nominal mass is calculated for the major isotopes. The accuracy of the mass spectra was estimated at ±1 mass unit.

recovery rate. The transfer of sialic acid increases the charge on the acceptor and ion-exchange chromatography can be used to separate glycopeptides containing one, two, or more negative charges. The sialoglycopeptides were isolated on a DEAE Sephadex column using a gradient of ammonium hydrogencarbonate buffer. A minor fraction was eluted first, followed by the main fraction containing the desired multivalent sialoglycopeptides. Theoretically, the fast-eluting fraction could result from either incomplete glycosylation of the divalent compounds or a faulty peptide backbone that contains only one β-D-GlcNAc-Asn unit. Surprisingly, the ^{1}H NMR spectrum indicated that the minor fraction was composed almost exclusively of compounds that contained only one sialyl-*N*-acetyllactosamine unit per molecule as indicated by the missing β-methylene signal at δ 2.7 for the β-D-GlcNAc-Asn located near the N-terminus (see Table II for NMR data of compounds **3c–10c**). This finding underscores the high efficiency of the enzymatic synthesis since no significant amounts of incompletely glycosylated compounds were detected. Additional information about the structure of compounds **4c–11c** was obtained by negative-ion FABMS. Due to the low solubility of the glycopeptides in the matrix, the accuracy of the mass spectra was estimated at ±1 mass unit. The molecular ions found are in good agreement with the calculated nominal masses (see Table III for FABMS data of compounds **4c–11c**).

Recombinant CMP-sialate synthetase in preparative-scale synthesis of CMP-sialic acid.—The limiting factors in enzymatic oligosaccharide synthesis have heretofore been the availability of nucleotide sugars and glycosyltransferases. Chemical synthesis of nucleotide sugars is tedious and difficult[21,26]. By using enzymes, the synthesis of donor substrates can be achieved in a more elegant way. Since commerically available CMP-sialic acid is one of the most costly donor substrates, a multigram synthesis from sialic acid and CTP was developed[27,28]. Crude CMP-sialate synthetase isolated from calf brain[29] contains phosphatases and therefore requires the use of excess CTP. However, with pure, recombinant *E. coli* CMP-sialate synthetase[7] (Genzyme), CMP-sialic acid could be synthesized very effi-



Scheme 2.

ciently (Scheme 2). It is known that the synthetase also catalyzes the reverse reaction. This problem can be abolished by enzymatically degrading the pyrophosphate to phosphate; however, the use of pyrophosphatase as proposed by Whitesides and co-workers[27] did not increase the yields in reactions conducted with the synthetase isolated from calf brain. In contrast, with recombinant CMP-sialate synthetase, the final yield was ~10% higher when pyrophoshatase was added, although no increase of reaction velocity could be observed until the reaction had reached 80% conversion. Equimolar amounts of CTP and sialic acid gave 85% yield of crude CMP-sialic acid, which was conveniently separated from the mixture by a single gel chromatography run on Biogel P-2, and was pure enough for enzymatic sialylation.

Virus inhibition studies with synthetic sialoglycopeptides.—The purified multivalent sialoglycopeptides **3c–11c** were tested for their ability to inhibit the binding of influenza virus to partially resialylated erythrocytes. Monovalent glycopeptide **2c** represents a partial structure of the multivalent compounds **3c–11c** and was therefore preferred over Neu5Acα2Me as a reference. This takes into consideration that viral hemagglutinin shows differential binding to sialosides with (2 → 6)-α- or (2 → 3)-α-linkages, respectively[30]. The results of the inhibition experiments are shown in Table I and are expressed in relative inhibitory potencies on the basis of glycopeptide concentration. Among the glycine-rich compounds **3c–7c**, the highest inhibition was found for **3c**, which was eight times more effective than the standard **2c**. A fourfold increased potency was found for **4c**, containing a short, nine amino acid (9 aa) spacer. When the length of the spacer was increased as in **5c** (12 aa) and **6c** (15 aa), the relative inhibitory potency dropped to two. Further increase in spacing reduced the inhibition to 1.3 for **7c** (18 aa).

For the glycopeptides **8c–11c** containing polyproline sequences, the highest inhibition (4.0) was found for **8c** (8 Pro), the shortest linear compound and for **11c** (2 × 8 Pro), which was synthesized from the branching lysine. In contrast, for the longer linear polyproline compounds **9c** (14 Pro) and **10c** (17 Pro), the inhibitory potency was lower than for monovalent reference **2c**.

DISCUSSION

Multivalent sialoglycopeptides could be obtained in good yields by a combination of chemical and enzymatic synthesis. Glycosyl amino acid **1** could be incorpo-

rated into peptides very efficiently by manual solid-phase peptide synthesis under optimized conditions. Glycopeptides containing up to 24 amino acid residues (**7a**) or synthetically difficult polyproline sequences **8a–11a** were obtained. After chemical deprotection, the glycopeptides were glycosylated using a modified procedure that allows the enzymatic transfer of carbohydrates onto compounds with multiple acceptor sites in high yields. This reaction sequence combines the advantages of highly developed chemical peptide synthesis and the selectivity of enzymatic transfer of carbohydrates. With increasing numbers of glycosyltransferases being cloned[31], the expression of those enzymes will make the enzymatic oligosaccharide synthesis a very powerful tool in carbohydrate chemistry.

In this study, we used recombinant (2 → 6)-α-sialyltransferase and found for the sialylation of **9b** that the recombinant enzyme was equal to the enzyme purified from rat liver (data not shown).

The sialoglycopeptides **3c–11c** were designed to bind to the influenza virus hemagglutinin, thereby blocking the sites that allow the virus to adhere to the target cells. Monovalent sialosides have only low affinity to the hemagglutinin and constitute poor inhibitors[3,14]. Divalent sialosides are expected to bind at two sites simultaneously, thereby increasing the affinity to the virus by cooperational effects. However, the binding of an inhibitor to multiple sites can be disturbed by sterical hindrance leaving the sialic acid unaccessible to the hemagglutinin and by entropic effects such as a high conformational flexibility in the spacer part. The degree of simultaneous binding should correlate with the inhibitory potency revealed by the virus adsorption assay.

In our experiments we used two classes of spacers: the glycine-rich type (**3–7**) and the polyproline type (**8–11**). Polyglycine-peptides are known to have a compact conformation (as determined by the chain end distances)[32]. Heteropolymers between glycine and alanine with less than 50 mole-percent alanine were shown to exhibit only slightly larger chain dimensions[33]. For glycopeptides **4c–7c**, the flexible spacer is expected to generate a compact, folded conformation of the peptide part with similar distances of the sialyllactosamines on the chain ends. Our finding, that the inhibitory potency of **4c–7c** drops gradually with the extension of the spacer, could be explained by the increasingly unfavourable simultaneous binding of the two sialic acid residues. Presumably, with growing chain length, the flexibility of the spacer influences binding.

Poly-L-proline exhibits a very rigid, helical conformation due to the severely hindered rotation of the imide bonds within the peptide chain[32,34]. The same helical structure can be found in collagen. In water, the polyproline type II helix is dominant with all imide links in the *trans* conformation. This results in a rigid, extended chain, where the distance of the chain ends is proportional to the number of proline residues. The polyproline type II helix spans a distance of 34 Å with only 12 proline residues[32,34].

For compounds **8c–10c** the increasing number of proline residues extends the rigid spacer helix and moves the two sialic acid containing side chains apart. This

could interfere with the steric accessability of the sialic acids and prevent simultaneous binding of both sugar chains.

In the case of bivalent glycopeptide **11**, the branching lysine was used for two reasons: (*i*) the number of peptide coupling steps needed can be reduced by half, and (*ii*) the lysine introduces a hinge into the very rigid polyproline backbone. The conformational flexibility of the lysine side chain (ω) should allow the attached polyproline arm to bend easily and to take on different angles to the main chain (α) resulting in a variable spacing.

According to the X-ray crystallographic analysis of the hemagglutinin (HA) trimer[4], the sialic acid binding pockets are ~ 40 Å apart. The polyproline spacer in **8c** as calculated for the polyproline type II helix provides ~ 23 Å between the two asparagine residues. One would expect this to be too short to allow the two sialic acids residues at the end of the asparagine side chains to reach simultaneously into two binding pockets of the same hemagglutinin trimer. In contrast, **9c** and **10c** (calculated spacing Asn_x-Asn_3: 40 and 48 Å, respectively) could theoretically bind two sites on the same trimer. However, only **8c** shows increased inhibition, whereas **9c** and **10c** exhibit even less affinity to the hemagglutinin than monovalent **2c**. This drop in inhibitory potency from **8c** to **9c** and **10c** is evidence that the increased length of the spacer in **9c** and **10c** is placing the sialic acids too far apart for simultaneous binding. This assumption is supported by **11c**, where the flexible lysine residue allows the sialic acids to get closer than in linear **10c**. Compound **11c** inhibits more than five times better than **10c**, although the number of proline residues of both substances is nearly equal.

If we rule out peptide–protein attraction, we can assume that the increased inhibitory potencies of **3c**, **4c**, and **8c** result from the simultaneous binding of the second sialic acid by the virus. These compounds are unlikely to occupy two binding sites on the same HA trimer due to conformational restrictions. This seems reasonable if we compare the data within the two homologous series **4c–7c** and **8c–10c**. However, a bridging of two neighbouring hemagglutinin trimers by one inhibitor molecule is possible and is entropically favored over the bridging of two viruses. Together with the structural assumptions for the different spacers, the formation of interhemagglutinin bridges can explain the inhibition data obtained. This is in accordance with the mechanism of inhibition proposed by the group of Knowles and co-workers[6d].

The presented work shows that synthetic sialoglycopeptides are useful as model compounds for the study of sialic acid binding proteins. They may also serve as starting structures in the development of anti-influenza drugs.

ACKNOWLEDGMENTS

We are especially grateful to Dr. Janis Young for her help with the solid-phase peptide synthesis. We thank Drs. J. Rasmussen and S. Shames of Genzyme for generously providing the recombinant CMP-sialate synthetase. Gratitude is ex-

pressed to Dr. M. Vogel of Hoechst AG for recording the FABMS spectra. This work was supported by fellowships from NATO (Deutscher Akademischer Austauschdienst) (to C.U.) and the Deutsche Forschungsgemeinschaft (to C.U. and S.K.) and the National Science Foundation Grant DMB 88-21049 (to J.C.P.).

REFERENCES

1 (a) G. Ashwell and J. Hartfort, *Annu. Rev. Biochem.*, 51 (1982) 531–554; (b) S. Kornfeld, *FASEB J.*, 1 (1987) 462–468; (c) T. Feizi, *Nature*, 314 (1985) 53–57; (d) J.C. Paulson, in M. Conn (Ed.), *The Receptors*, Vol 2, Academic Press, New York, 1985, pp 131–219; (e) T.W. Rademacher, R.B. Parekh and R.A. Dwek, *Annu. Rev. Biochem.*, 57 (1988) 785–838; (e) J.C. Paulson, *Trends Biochem. Sci.*, 14 (1989) 272–276; (f) T.A. Springer, *Nature*, 346 (1990) 425–434; (g) T. Feizi, *Trends Biochem. Sci.*, 16 (1991) 84–86.

2 D.C. Wiley and J.J. Skehel, *Annu. Rev. Biochem.*, 56 (1987) 365–394.

3 N.K. Sauter, M.D. Bednarski, B.A. Wurzburg, J.E. Hanson, G.E. Whitesides, J.J. Skehel, and D.C. Wiley, *Biochemistry*, 28 (1989) 8388–8396.

4 W. Weiss, J.H. Brown, S. Cusak, J.C. Paulson, J.J. Skehel, and D.C. Wiley, *Nature*, 333 (1988) 426–431.

5 (a) T.J. Pritchett and J.C. Paulson, *J. Biol. Chem.*, 264 (1989) 9850–9858; (b) K. Hanaoka, T.J. Pritchett, S. Takasaki, N. Kochibe, S. Sabesan, J.C. Paulson, and A.J. Kobata, *J. Biol. Chem.*, 264 (1989) 9842–9849.

6 (a) M.N. Matrosovich, L.V. Mochalova, V.P. Marinina, N.E. Byramova, and N.V. Bovin, *FEBS Lett.*, 272 (1990) 209–212; (b) A. Spaltenstein and G.M. Whitesides, *J. Am. Chem. Soc.*, 113 (1991) 686–687; (c) A. Gamian, M. Chomik, C.A. Laferrière, and R. Roy, *Can. J. Microbiol.*, 37 (1991) 233–237; (d) G.D. Glick and J.R. Knowles, *J. Am. Chem. Soc.*, 113 (1991) 4701–1703; (e) S. Sabesan, J.Ø Duus, P. Domaille, S. Kelm, and J.C. Paulson, *J. Am. Chem. Soc.*, 113 (1991) 5865–5866.

7 S.L. Shames, E.S. Simon, C.W. Christopher, W. Schmid, G.M. Whitesides, and L.-L. Yang, *Glycobiology*, 1 (1990) 187–191.

8 (a) J. Weinstein, U. de Souza-e-Silva, and J.C. Paulson, *J. Biol. Chem.*, 257 (1982) 1383–13844; (b) K.J. Colley, E.U. Lee, B. Adler, J.K. Browne, and J.C. Paulson, *J. Biol. Chem.*, 264 (1989) 17619–17622.

9 C. Unverzagt, H. Kunz, and J.C. Paulson, *J. Am. Chem. Soc.*, 112 (1990) 9308–9309.

10 H. Waldmann and H. Kunz, *Liebigs Ann. Chem.*, (1983) 1712–1725.

11 S. Lavielle, N.C. Ling, and R. Guillemin, *Carbohydr. Res.*, 89 (1981) 221–228.

12 J.M. Stewart and J.D. Young, *Solid-Phase Peptide Synthesis*, Pierce Chemical Company, Rockford, Illinois, 1984.

13 P. Schultheiss-Reimann and H. Kunz, *Angew. Chem.*, 95 (1983) 64; *Angew. Chem. Intl. Ed. Engl.*, 22 (1983) 62–63.

14 L. Warren, *J. Biol. Chem.*, 234 (1959) 1971–1975.

15 R. Schauer, *Adv. Carbohydr. Chem. Biochem.*, 40 (1982) 131–234.

16 T.J. Pritchett, R. Brossmer, U. Rose, and J.C. Paulson, *Virology*, 160 (1987) 502–506.

17 (a) H.G. Garg and R.W. Jeanloz, *Adv. Carbohydr. Chem. Biochem.*, 43 (1985) 135–201; (b) H. Kunz, *Angew. Chem.*, 99 (1987) 297–311; *Angew. Chem. Intl. Ed. Engl.*, 26 (1987) 294–308.

18 (a) H. Kunz and B. Dombo, *Angew. Chem.*, 100 (1988) 732–734; *Angew. Chem. Intl. Ed. Engl.*, 27 (1988) 711–713; (b) H. Paulsen, G. Merz, and U. Weichert, *Angew. Chem.*, 100 (1988) 1425–1427; *Angew. Chem. Intl. Ed. Engl.*, 27 (1988) 1365–1367; H. Paulsen, G. Merz, S. Peters, and U. Weichert, *Liebigs Ann. Chem.*, (1990) 1165–1173; (c) B. Luning, T. Norberg, and T. Teijbrant, *J. Chem. Soc., Chem. Commun.*, (1989) 1267–1268; (d) M. Meldal and K.J. Jensen, *J. Chem. Soc., Chem. Commun.*, (1990) 483–485; M. Meldal and K. Bock, *Tetrahedron Lett.*, 31 (1990) 6987–6990; A.M. Jansson, M. Meldal, and K. Bock, *ibid.*, 6991–6994; (e) L. Otvos, Jr., K. Wroblewski, E. Kollat, A. Perczel, M. Hollosi, G.D. Fasman, H.C.J. Ertl, and J. Thurin, *Pept. Res.*, 2 (1989) 362–366; (f) F. Filira, L. Biondi, F. Cavaggion, b. Scolaro, and R. Rocchi, *Int. J. Pept. Protein Res.*, 36 (1990) 86–96; (g) E. Bardaji, J.L. Torres, P. Clapés, F. Albericio, G. Barany, and G. Valencia, *Angew. Chem.*, 102 (1990) 311–313; *Angew. Chem. Intl. Ed. Engl.*, 29 (1990) 291–292.

19 A.R. Mitchell, S.B.H. Kent, M. Engelhard, and R.B. Merrifield, *J. Org. Chem.*, 43 (1978) 2845–2852.
20 P.J. Tam, *Proc. Natl. Acad. Sci. USA*, 85 (1988) 5409–5413.
21 (a) E.J. Toone, E.S. Simon, M.D. Bednarski, and G.M. Whitesides, *Tetrahedron*, 45 (1989) 5365–5422; (b) D.G. Drueckhammer, W.J. Hennen, R.L. Pederson, C.F. Barbas III, C.M. Gautheron, T. Krach, and C.H. Wong, *Synthesis* (1991) 499–525; (c) S. Sabesan and J.C. Paulson, *J. Am. Chem. Soc.*, 108 (1986) 2068–2080; (d) J. Thiem and W. Treder, *Angew. Chem.*, 98 (1986) 1100–1101; *Angew. Chem. Intl. Ed. Engl.*, 25 (1986) 1096–1097; (e) H.T. De Heij, M. Kloosterman, P.L. Koppen, J.H. van Boom, and D.H. van den Eijnden, *J. Carbohydr. Chem.*, 7 (1988) 209–222; (f) K.G. Nilsson, *Carbohydr. Res.*, 188 (1989) 9–17; (g) M.M. Palcic, A.P. Venot, R.M. Ratcliffe, and O. Hindsgaul, *Carbohydr. Res.*, 190 (1989) 1–11; (h) Y. Ichikawa, G.J. Shen, and C.H. Wong, *J. Am. Chem. Soc.*, 113 (1991) 4698–4700.
22 (a) C. Augé, C. Gautheron, and H. Pora, *Carbohydr. Res.*, 193 (1989) 288–293; (b) J. Thiem and T. Wiemann, *Angew. Chem.*, 102 (1990) 78–80; *Angew. Chem. Intl. Ed. Engl.*, 29 (1990) 80–82.
23 J. Weinstein, U. de Souza-e-Silva, and J.C. Paulson, *J. Biol. Chem.*, 257 (1982) 13845–13853.
24 K. Brew, T.C. Vanaman, and R.L. Hill, *Proc. Natl. Acad. Sci. USA*, 59 (1968) 491–497.
25 J. Weinstein, personal communication.
26 (a) R.R. Schmidt, B. Wegmann and K.H. Jung, *Liebigs Ann. Chem.*, (1991) 121–124; (b) U.B. Gokhale, O. Hindsgaul, and M.M. Palcic, *Can. J. Chem.*, 68 (1990) 1063–1071.
27 E.S. Simon, M.D. Bednarski, and G.M. Whitesides, *J. Am. Chem. Soc.*, 110 (1988) 7159–7163.
28 (a) J. Thiem and P. Stangier, *Liebigs Ann. Chem.*, (1990) 1101–1105; (b) Y. Ichikawa, G.-J. Shen, and C.H. Wong *J. Am. Chem. Soc.*, 113 (1991) 4698–4700.
29 H. Higa and J.C. Paulson, *J. Biol. Chem.*, 260 (1985) 8838–8849.
30 G.N. Rogers and J.C. Paulson, *Virology*, 127 (1983) 361–373.
31 (a) J.C. Paulson and K.C. Colley, *J. Biol. Chem.*, 264 (1989) 17615–17618; (b) D.H. Joziasse, N.L. Shaper, L.S. Salyer, D.H. Van den Eijnden, A.C. Van der Spoel, and J.H. Shaper, *Eur. J. Biochem.*, 191 (1990) 75–83; (c) F.I. Yamamoto, H. Clausen, T. White, J. Marken, and S.I. Hakomori, *Nature*, 345 (1990) 223–233; (d) J.B. Lowe, J.F. Kukowska-Latallo, R.P. Nair, R.D. Larsen, R.M. Marks, B.A. Macher, R.J. Kelly, and L.K. Ernst, *J. Biol. Chem.*, 266 (1991) 17615–17618; (e) M. Sakar, E. Hull, Y. Nishikawa, R.J. Simpson, R.L. Moritz, R. Dunn, and H. Schachter, *Proc. Natl. Acad. Sci. USA*, 88 (1991) 234–238.
32 P.R. Schimmel and P.J. Flory, *Proc. Natl. Acad. Sci. USA*, 58 (1967) 52–59.
33 W.G. Miller, D.A. Brant, and P.J. Flory, *J. Mol. Biol.*, 23 (1967) 67–80.
34 L. Stryer and R.P. Haugland, *Proc. Natl. Acad. Sci. USA*, 58 (1967) 719–726.