3D-QSAR Analysis of Sialyltransferase Inhibitors

Xiaofang Wang,[a] Youhong Niu,[a,b] Xiaoping Cao,[b] Liangren Zhang,[a] Li-He Zhang[a]
and Xin-Shan Ye[a,*]

[a]*The State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Science, Peking University,
Xue Yuan Road #38, Beijing 100083, China*
[b]*National Laboratory of Applied Organic Chemistry, Lanzhou University, Lanzhou 730000, China*

Abstract—A predictive 3D-QSAR model that correlates the biological activities with the chemical structures of a series of sialyl-transferase inhibitors, exemplified by the sugar:nucleotide derivatives, was developed by means of comparative molecular field analysis (CoMFA). The resulting cross-validated value ($q^2 = 0.629$), non-cross-validated value ($r^2 = 0.965$) and standard error of estimate (SEE = 0.288) indicate that the obtained pharmacophore model indeed mimics the steric and electrostatic environment where inhibitors bind to the enzyme. The developed model also possesses promising predictive ability as discerned by the testing on the external test set, and should be useful to further understand the molecular nature of inhibitor–enzyme interactions and to aid in the design of more potent sialyltransferase inhibitors.
© 2003 Elsevier Ltd. All rights reserved.

Introduction

Sialic acids (SA) attached to the exposed terminus of the oligosaccharide chains of glycoconjugates have long being predicted to be cell-type specific markers and critical determinants in a variety of cellular recognition processes, metastasis phenomena and immune responses.[1,2] The transfer of sialic acid from its active nucleotide donor CMP-SA to glycoproteins or glycolipids is mediated by sialyltransferases, a family of at least 18 distinct glycosyltransferases. It has been shown that sialic acid residues on cell surface are involved in the cell–virus interaction in influenza infection and increased sialylation or sialyltransferase activity correlates with the growth or metastatic potential of tumor cells.[3–5] Recently, some interesting examples, such as Galβ(1-4)GlcNAcα(2-6)-sialyltransferase that regulates B lymphocyte activation and immune function, and Galβ(1-3)GlcNAcα(2-3)-sialyltransferase that controls CD8+ T lymphocyte homeostasis, have also been reported[6,7] Since the inhibitors with sialyltransferase specificity can regulate sialyltransferase activities in cells, subsequently affect the oligosaccharide structure produced, they can be useful in further elucidating the biological functions of sialic acid-containing glyco-conjugates as well as future therapeutical applications as new potent anti-inflammatory, immunosuppressive and antimetastatic agents.

The exploration of sialyltransferase inhibitors has progressed promisingly. There have been a variety of reported sialyltransferase inhibitors, including sugar-donor analogues, transition-state mimetics, sugar-acceptor analogues as well as natural products, endogenous macroglobulin, and antisense-oligodeoxynucleotides.[8–10] However, a little has been known for the structure–activity relationships due to the absence of information on their detailed mechanism. In the rational drug design, 3D-QSAR analysis is one of the important steps, especially when three-dimensional structures of bioreceptors remain unclear. Since no X-ray data of sialyltransferases are available so far, 3D-QSAR analysis becomes such a useful tool that can predict structural insights to the level of enzyme-substrate binding sites and contributes to further design of novel potent sialyltransferase inhibitors. We present here the 3D-QSAR studies using comparative molecular field analysis (CoMFA) approach on a set of sialyl-donor-based sialyltransferase inhibitors (38 compounds, including transition-state analogues of sugar-donor) by considering the steric and electrostatic influences. The model deduced from this investigation provides

*Corresponding author. Tel.: +86-10-6209-1570; fax: +86-10-6201-4949; e-mail: xinshan@mail.bjm.edu.cn

underlying structural requirements and good predictive ability, which could aid the design of new sialyltransferase inhibitors prior to their synthesis.

Materials and Methods

Compounds and biological data

Thirty-eight sugar-donor based sialyltransferase inhibitors including transition-state analogues of sugar-donor listed in Table 1 were collected from the published results by Schmidt and coworkers.[11–16] These compounds are carbohydrate mimetics derived from the donor substrate (CMP-Neu5Ac) of sialyltransferase and are mostly composed of saccharide residue, CMP nucleotide moiety, and a intermediate linker. The structural diversity is basically demonstrated on the scaffold of sugar portion and the linker between sugar portion and CMP nucleotide moiety. Their inhibitory activities in vitro for α-(2,6)-sialyltransferase from rat liver (EC2.4.99.1), measured as inhibition constant K_i values, cover a wide range from 2 mM to 29 nM. Eight inhibitors were randomly selected from Table 1 marked with an asterisk (*) as the external test set to carry out the CoMFA study while the remaining 30 compounds were used as the training set.

Conformational analysis

The determination and superimposition of the bioactive conformations of the studied compounds is a crucial step in CoMFA analysis, which relies on either X-ray structures of enzyme-substrate complexes, experiments of docking an active ligand into its receptor, or conformational analysis of active molecules. Since there is no structural information available for sialyltransferase currently and all inhibitors under studying are inherently flexible, molecular modeling of sialyltransferase inhibitors were performed.

The molecular modeling studies were carried out on a SGI O$_2$ workstation with SYBYL 6.4 software package. As most of the compounds have the common structural features, CMP-Neu5Ac (1), the donor substrate of sialyltransferase, and compound 19 with the typical sugar:nucleotide structure, were selected as the template molecules in order to explore the common bioactive conformation among all competitive inhibitors having the same binding site. Other compounds were generated by simply modifying some substituents of the template molecules with standard fragment or the sketch option under built module and subsequently minimized using Maximin2 molecular mechanics method (Tripos force field) and MOPAC AM1 semi-empirical quantum chemical method.

Both CMP-Neu5Ac (1) and compound 19 can be split into three building blocks: sugar residue, cytidine moiety and phosphate linker. The sugar residue (Neu5Ac or Neu5Ac2en) and cytidine moiety were built from the refined structures of sialic acid (or Neu5Ac2en)–neuraminidase complex[17,18] and standard fragment library, while the energetically favorable orientation of the phosphate dihedrals was determined by systematic search and subsequent geometry optimization with 30° step size in the range from 0 to 360° for all rotatable bonds as listed in Figure 1.

Starting from the template structures, all the compounds were prepared with a comparable conformation. Superimpositions were then performed by fitting atoms of common core using compound 35, the most potent sialyltransferase inhibitor in this analysis, as the template. Three alignments with rules A, B, and C in Figure 2 were investigated, respectively, in order to best match the common structure features and maintain a similar orientation in space for the final construction of the optimized CoMFA model.

CoMFA model generation

CoMFA was performed using the QSAR module of SYBYL. With compound 35 as the template, all the compounds were aligned using the atoms marked with an asterisk (*). The molecules were then presented by their steric and electrostatic fields, using a sp^3 hybridized carbon atom probe carrying a +1 charge, sampled at the intersection of a 3D lattice of 2 Å grid spacing and 4 Å boundary extension of the region beyond the largest structure in all directions. The energetic cutoff values of steric and electrostatic field were 5 and 10 kcal/mol, respectively.

Correlations of biological activity with variations at the steric and electrostatic field lattice points were calculated by Partial Least Squares (PLS) method, in which the optimal number of principal components (N_{opt}) obtained from leave-one-out cross-validation procedure was employed in the subsequent non-cross-validation calculation. The model was evaluated by the cross-validated correlation coefficient q^2, non-cross-validated statistical results [r^2, standard error of estimate (SEE) and F-test value] and also the prediction for the compounds in the test set. The validated CoMFA model can be represented by StDev*COEFF contour plots which reveals the important regions for biological activity in colored polyhedra.

Results and Discussion

CoMFA model

The CoMFA studies were performed with different superimposition as shown in Figure 2. The best model (Table 2) was developed from the alignment C with cross-validated $q^2 = 0.629$, $N_{opt} = 5$, non-cross-validated $r^2 = 0.965$, SEE = 0.288, $F_{(n1=5,n2=24)} = 165.209$. These statistical results suggest a good correlation between the CoMFA fields and the biological activities of the compounds studied and present a predictive model with the relative contributions from steric and electrostatic molecular fields as 68.8 and 31.2%, respectively.
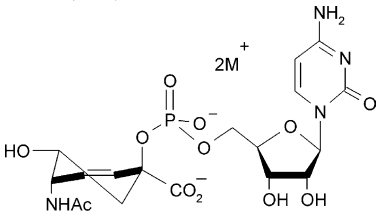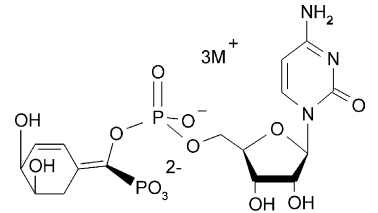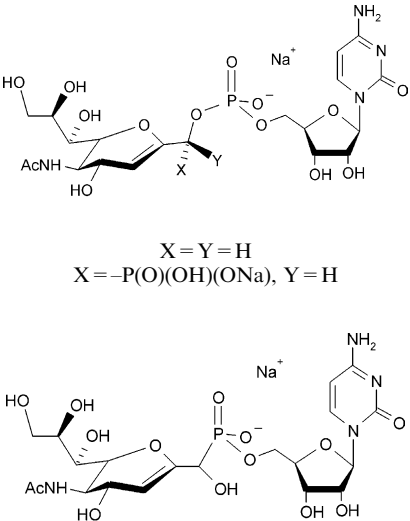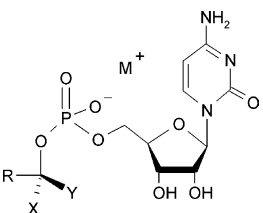
The predicted PK_i values versus the experimental values from the best CoMFA model are graphically represented

Table 1. Structures and biological activity of sialyltransferase inhibitors

N.	Structure	Experiment values		Prediction	Residual
		K_i	Pk_i	PK_i	
1. CMP-Neu5Ac		46 μM (K_m)	4.34	4.22	0.11
2*		44 μM	4.36	4.33	0.03
3	Y = CO$_2^-$	2 mM	2.70	2.3	0.40
4	Y = CH$_2$CO$_2^-$	270 μM	3.57	3.25	0.32
5	Y = CH$_2$PO$_3$H	750 μM	3.12	3.38	−0.26
6	Y = CH$_3$C(=CH$_2$)CO$_2^-$	250 μM	3.60	3.84	−0.24
7*		250 μM	3.60	3.77	−0.17
8		780 μM	3.11	3.05	0.06
9	R = Y = H, X = Z = OH	44 μM	4.36	4.33	0.03
10*	R = X = H, Y = NHAc, Z = OH	84 μM	4.08	4.58	−0.50
11*	R = X = H, Y = NH3+, Z = OH	1400 μM	2.85	3.47	−0.62
12	R = X = H, Y = NH3+, Z = NHAc	200 μM	3.70	3.77	−0.07
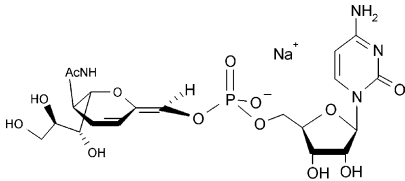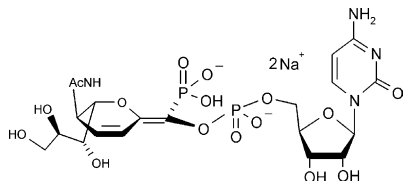
Table 1 (*continued*)

N.	Structure	Experiment values		Prediction	Residual
		K_i	Pk_i	PK_i	
13*	R = (CH$_2$)$_2$OH, X = Z = OH, Y = H	20 μM	4.70	4.48	0.22
14		15 μM	4.82	4.51	0.29
15		10 μM	5.00	4.64	0.36
16*	X = –P(O)(OH)(ONa), Y = H	1.60 μM	5.80	5.85	−0.05
17*	X = H, Y = –P(O)(OH)(ONa)	0.27 μM	6.57	6.32	0.25
18	X = Y = H	2 mM	2.70	2.52	0.18
19	X = –P(O)(OH)(ONa), Y = H	0.35 μM	6.64	6.70	−0.05
20		400 μM	3.40	3.16	0.24

Table 1 (*continued*)

N.	Structure	Experiment values		Prediction	Residual
		K_i	Pk_i	PK_i	
21*	R = Ph, X = H, Y = CO$_2$Na	10 μM	5.00	5.06	−0.06
22	R = Ph, X = CO$_2$Na, Y = H	7 μM	5.14	5.40	−0.26
23	R = Bn, X = H, Y = CO$_2$Na	15 μM	4.82	4.51	0.31
24	R = Bn, X = CO$_2$Na, Y = H	23 μM	4.64	4.60	0.04
25	R = Ph, X = H, Y = −P(O)(OH)(ONa)	0.20 μM	6.70	6.73	−0.03
26	R = Ph, X = −P(O)(OH)(ONa), Y = H	1.0 μM	6.00	5.99	0.01
27	R = 2-furyl, X = H, Y = −P(O)(OH)(ONa)	0.28 μM	6.55	6.49	0.06
28	R = 2-furyl, X = −P(O)(OH)(ONa), Y = H	1.0 μM	6.00	5.80	0.20
29		6.0 μM	5.22	5.29	−0.07
30		40 nM	7.40	7.21	0.19
31	(*E*):	158 μM	3.80	3.94	−0.14
32	(*Z*):	25 μM	4.60	4.74	0.14
33	(*E*):	2.4 μM	5.62	5.01	0.61
34	(*Z*):	3.5 μM	5.40	5.32	0.08
35	h	29 nM	7.54	7.47	0.07
36	l	690 nM	6.16	6.16	0

Table 1 (*continued*)

N.	Structure	Experiment values		Prediction	Residual
		K_i	Pk_i	PK_i	

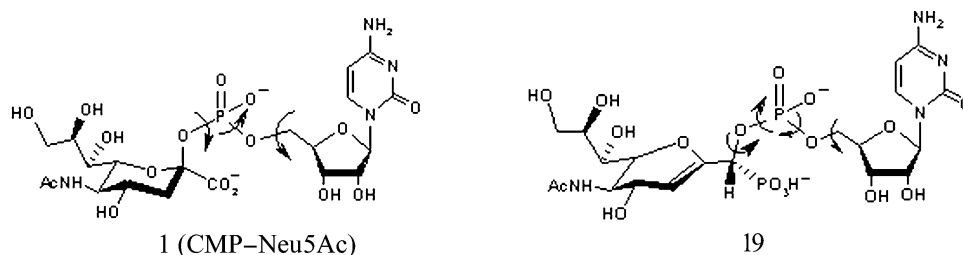|					
37	h	59nM	7.23	7.31	−0.08
38	l	38nM	7.42	7.34	0.08

Figure 1. Structures of compound **1** (CMP-Neu5Ac) and **19**.

Table 2. PLS results based on three different superimpositions

	Cross-validation		Non-cross-validation		
	q^2	N_{opt}	r^2	SEE	$F_{(n1 = 5, n2 = 24)}$
Rule A	0.463	4	0.912	0.448	62.305
Rule B	0.579	5	0.930	0.401	79.375
Rule C	0.629	5	0.965	0.288	165.209

in Figure 3. There are no obvious outliers found, validating the effectivity of the model and also suggesting that the compounds indeed share the same binding mechanism.

The superimposition of all molecules with compound **35** as a template and the StDev*COEFF contour plots of the CoMFA analysis are shown in Figures 4 and 5. In the CoMFA contour plots, sterically favored or disfavored areas are color coded by green or yellow polyhedra, while positive or negative charge favored areas are color coded by blue or red polyhedra, respectively. As shown in Figure 4, CMP moieties of all molecules are superposed well while the sugar portions are not. Figure 5 shows regions around sugar portion where variations of steric or electrostatic nature of different molecules are associated with increase or decrease in the activity, but shows less regions around CMP moiety where the molecules share the common structural feature. So the CoMFA method relates the differences in the biological properties to differences in the chemical structures or steric and electrostatic fields around the studied molecules. There may be a binding pocket related to the sugar portion of the sugar/nucleotide sub-
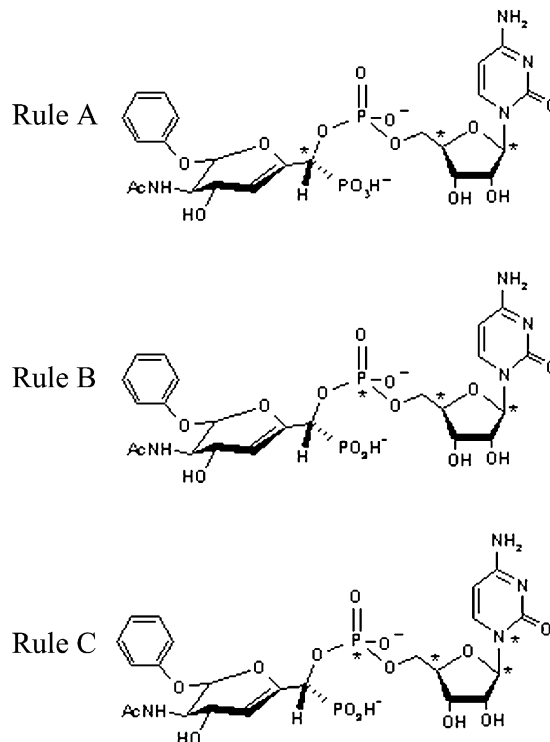

Figure 2. Three different rules of superimposition selected for CoMFA analysis.

strate and the internal orientation of the sugar residue is important for the affinity. The disfavored steric region around the Neu5Ac residue of CMP-Neu5Ac may suggest an increased distance between the anomeric carbon
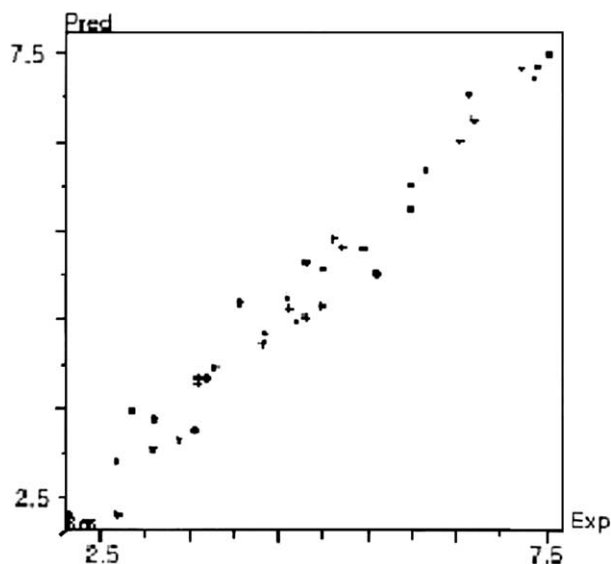
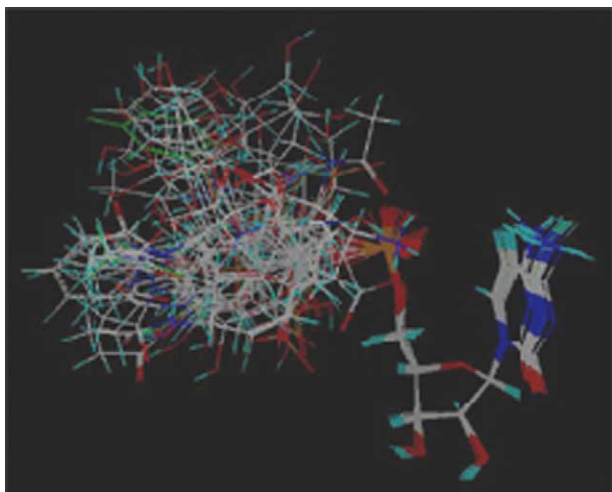**Figure 3.** Predicted values versus experimental values for the training set.

Figure 4. Superimposition of all molecules with compound **35** as a template.

and the CMP leaving group associated with higher activity, as have evidenced by the traditional SAR studies.[13] Compared with CMP-Neu5Ac, compound **19** with flatter sugar-mimetic residues displayed much higher inhibitory activities for α-(2,6)-sialyltransferase, while it was believed in the traditional SAR that the planarity at the anomeric center contributes to the high activity.[13] In addition, compounds **25–28** have a simplified flat aryl or hetary ring instead of the Neu5Ac residue, but with similar high affinities. The location of the aryl ring in phenyl-derivatives is also upper than Neu5Ac residue in CMP-Neu5Ac, implying a possible hydrophobic binding site in the sialyltransferase that is feasible for the substrate-enzyme binding and leads to enhanced activity. The phenyl group as the Neu5Ac2en side chain in the currently most potent inhibitor **35** may be also involved in this interaction as compound **35** exhibits about 10-fold higher affinity than its parent compound **19**.[16]
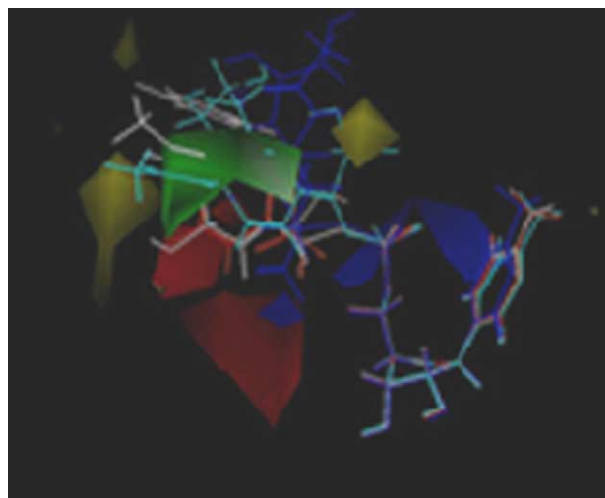

Figure 5. CoMFA steric and electrostatic STDEV*COEFF contour plot from PLS analysis based on the alignment rule C. Compounds **1** (CMP-Neu5Ac), **25**, **30**, **35** colored with blue, red, light green and white shown inside the field, respectively.

Predictive ability of the model

The predictive ability of the CoMFA model was further validated by the prediction for the external test set. The predicted PK_i values of eight compounds in the test set as well as their experimental values and residual values are given in Table 1. The predictions are close to the experimental values with a maximum deviation of 0.61 in the log scale. The results indicate that this model possesses a promising predictive ability for the compounds randomly selected and furthermore could guide the design of the potent compounds and avoid the synthesis of compounds with very low or without activity.

Conclusions

By conformational analysis using molecular mechanics and quantum chemical methods, we rationally deduced a pharmacophore model for donor-based α-(2,6)-sialyltransferase inhibitors and their transition-state analogues. The obtained CoMFA model indeed mimics the steric and electrostatic environment around inhibitor binding site to the enzyme and possesses promising predictive ability. The model described here could be potentially helpful in the design of novel and more potent sialyltransferase inhibitors.

Acknowledgements

References and Notes

1. Varki, A. *Glycobiology* **1992**, *2*, 25.
2. Powell, L. D.; Varki, A. *J. Biol. Chem.* **1995**, *270*, 14243.
3. Harduin-Lepers, A.; Krzewinski-Recchi, M. A.; Hebbar,

M.; Samyn-Petit, B.; Vallejo-Ruiz, V.; Julien, S.; Peyrat, J. P.; Delannoy, P. *Recent Res. Dev. Cancer* **2001**, *3*, 111.

4. Burchell, J.; Poulsom, R.; Hanby, A.; Whitehouse, C.; Cooper, L.; Clausen, H.; Miles, D.; Taylor-Papadimitriou, J. *Glycobiology* **1999**, *9*, 1307.

5. Lo, N. W.; Dennis, J. W.; Lau, J. T. *Biochem. Biophys. Res. Commun.* **1999**, *264*, 619.

6. Hennet, T.; Chui, D.; Paulson, J. C.; Marth, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4504.

7. Priatel, J. J.; Chui, D.; Hiraoka, N.; Simmons, C. J.; Richardson, K. B.; Page, D. M.; Fukuda, M.; Varki, N. M.; Marth, J. D. *Immunity* **2000**, *12*, 273.

8. Compain, P.; Martin, O. V. *Bioorg. Med. Chem.* **2001**, *9*, 3077.

9. Kiefel, M. J.; von Itzstein, M. *Chem. Ber.* **2002**, *102*, 471.

10. Wang, X.; Zhang, L.-H.; Ye, X.-S. *Med. Res. Rev.* **2003**, *23*, 32.

11. Schaub, C.; Müller, B.; Schmidt, R. R. *Eur. J. Org. Chem.* **2000**, 1745.

12. Müller, B.; Martin, T. J.; Schaub, C.; Schmidt, R. R. *Tetrahedron Lett.* **1998**, *39*, 509.

13. Amann, F.; Schaub, C.; Müller, B.; Schmidt, R. R. *Chem. Eur. J.* **1998**, *4*, 1106.

14. Schaub, C.; Müller, B.; Schmidt, R. R. *Glycoconjugate J.* **1998**, *15*, 345.

15. Müller, B.; Schaub, C.; Schmidt, R. R. *Angew. Chem. Int. Ed.* **1998**, *37*, 2893.

16. Schwörer, R.; Schmidt, R. R. *J. Am. Chem. Soc.* **2002**, *124*, 1632.

17. Burmeister, W. P.; Ruigrok, R. W. H.; Cusack, S. *EMBO J.* **1992**, *11*, 49.

18. Smith, B. J.; Colman, P. M.; Von-Itzstein, M.; Danylec, B.; Varghese, J. N. *Protein Sci.* **2001**, *10*, 689.