

Probing cell-surface architecture through synthesis: An NMR-determined structural motif for tumor-associated mucins

DAVID H. LIVE*, LAWRENCE J. WILLIAMS†, SCOTT D. KUDUK†, JACOB B. SCHWARZ†, PETER W. GLUNZ†, XIAO-TAO CHEN‡, DALIBOR SAMES‡, R. AJAY KUMAR§, AND SAMUEL J. DANISHEFSKY†¶

*Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota Medical School, Minneapolis, MN 55455; †Laboratory for Bioorganic Chemistry, Sloan-Kettering Institute for Cancer Research, 1275 York Avenue, Box 106, New York, NY 10021; ‡Department of Chemistry, Columbia University, Havemeyer Hall, New York, NY 10027; and §Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598

Contributed by Samuel J. Danishefsky, February 3, 1999

ABSTRACT Cell-surface mucin glycoproteins are altered with the onset of oncogenesis. Knowledge of mucin structure could be used in vaccine strategies that target tumor-associated mucin motifs. Thus far, however, mucins have resisted detailed molecular analysis. Reported herein is the solution conformation of a highly complex segment of the mucin CD43. The elongated secondary structure of the isolated mucin strand approaches the stability of motifs found in folded proteins. The features required for the mucin motif to emerge are also described. Immunocharacterization of related constructs strongly suggests that the observed epitopes represent distinguishing features of tumor cell-surface architecture.

The exterior of most cells is dominated by glycolipids, proteoglycans, and glycoproteins, including mucin-like proteins. These display polyvalent α -O-linked carbohydrates on proximal serine and threonine residues (1–4). Altered expression of cell-surface mucin character is often characteristic of malignant cells (1, 3, 5). Accordingly, tumor-associated mucins are good targets for a vaccination strategy (3, 5–9). We have surmounted the synthetic challenges of constructing polypeptides bearing clustered glycodomains (7, 8), and one such construct is currently in human clinical trials. Given this access, we probed the effects of clustered glycosylation patterns on peptide conformation and recognition (9–18). We have conducted extensive NMR and restrained molecular dynamics calculations (19, 20) on fully synthetic clustered carbohydrate tumor antigens (1–4) corresponding to a fragment of CD43, a glycoprotein aberrantly expressed on the surface of acute myelogenous leukemia cells (21–24). Our findings demonstrate how clustered glycosylation induces the peptide backbone into an unprecedented rigid scaffold corresponding to a polypeptide secondary conformation, which is consistent with elongated mucin glycoprotein structure and function (1, 25–29). Remarkably, the glycosylation-induced structure approaches the stability of motifs found in globular proteins.

The CD43, or leukosialin, protein presented an ideal candidate from which to select a substructure for synthesis. In consequence of its possible role in inducing immunologic response, the system is relatively well characterized (22–24). The sequence STTAV is a glycosylation locus found in the amino terminus of the protein. We have accessed through chemical synthesis the clustered 2,6-STF trisaccharide **1** (7), which is present on CD43 when expressed on acute myelogenous leukemia cells (24). In addition, we have synthesized the TF disaccharide **2** and the monosaccharide Tn antigen **3**. Through our synthetic methods, we also gained access to the β -linked stereoisomer of the TF antigen **4** (Fig. 1) (8). As will

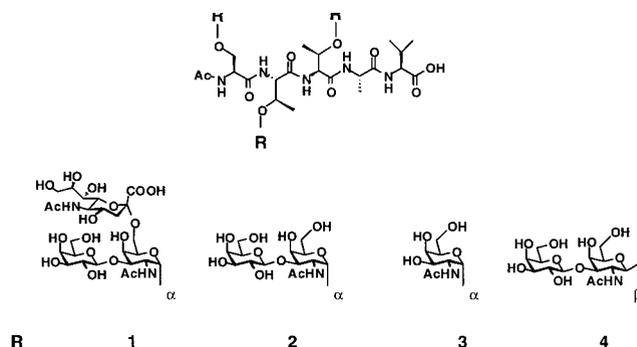


FIG. 1. Polypeptides **1–4** were prepared by total synthesis and represent clustered α - and β -O-linked glycodomains. The sequence STTAV is in the amino terminus of the protein CD43. The glycans, indicated as R, are known tumor antigens. The STF trisaccharide antigen of **1** is present on CD43 when expressed on acute myelogenous leukemia cells. The TF antigen of **2** and Tn antigen of **3** are expressed on certain malignant carcinomas, particularly of the prostate and colon. Constructs **1–3** contain the natural mucin α -O-GalNAc core. Construct **4** was prepared for comparison and has the unnatural β -linkage.

be seen, the latter served as a critical control to assess the sensitivity of the glycosylation-induced structure to anomeric stereochemistry.

At the outset, we were mindful that, to date, mucin proteins have resisted detailed NMR analysis and that spectral resonances of highly glycosylated proteins lack the dispersion necessary to permit sequence assignment (16). The main difficulties arise from the presence of tandemly repeated peptide segments rich in serine and threonine. The first carbohydrate of O-linked mucins is α -O-GalNAc, and most of the serine and threonine residues are thought to be glycosylated (1, 4). Accordingly, it was of great interest to find that the amide region of the $^1\text{H-NMR}$ spectrum of constructs **1–3** revealed a dramatic alteration in the structure of the peptide backbone as a consequence of the α -linked carbohydrates (Fig. 2a). The structural change of the α -linked series further manifests itself in the significant increase in the lifetime of exchangeable peptide backbone amide protons (NH) relative to the free peptide. By contrast, the β -linked isomer **4** showed rather meager changes in amide chemical shifts relative to the α -isomer **2**. Furthermore, the exchange lifetimes of the NH of the GalNAc residues are very sensitive to anomeric stereochemistry. For example, where the β -linked GalNAc NH in

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, Biology Department, Brookhaven National Laboratory, Upton, NY 11973 (PDB ID code 1sia).

¶To whom reprint requests should be addressed. e-mail: s-danishefsky@ski.mskcc.org.

compound **4** exchange in minutes, the corresponding NH of **1** persist for more than 12 hr. *These results demonstrate that sequential O-glycosylation by means of α -linked GalNAc induces transition to a highly stable structure—a remarkable result in such a relatively small peptide.*

To elucidate the conformation of the α -linked STF cluster in detail, extensive two-dimensional NMR homo- and heteronuclear NMR experiments, as well as a three-dimensional Total correlation spectroscopy (TOCSY)–nuclear Overhauser effect spectroscopy (NOESY) experiments were conducted at 600 and 800 MHz in H₂O and D₂O^{||**}. Whereas the presence of the identical pendant glycans present formidable challenges in resonance assignment, we were able to fully assign the amino acid residues and have identified the H1, H2, H3, NH, and methyl groups of each proximal GalNAc. The proton and carbon heteronuclear multiple quantum coherence resonances for the peripheral galactose and sialic acid residues were found to be degenerate. Remarkably, we were able to correlate over 100 proton NOEs. These measurements reveal a strong interaction between the methyl protons of the GalNAc acetyl groups and the peptide for **1** (Fig. 2*b*). Similar interactions were observed for **2** and **3**. By contrast, no such interactions were observed in the “nonnatural” β -linked construct **4**. (For an in-depth analysis of the conformational symbiosis between β -O-linked Lewis X glycans and a segment of MAd-CAM-1, see ref. 30. For earlier papers on this subject, see refs. 31–33.)

Fig. 3 depicts the solution structures of **1** calculated by restrained molecular dynamics in torsion angle space guided by data from multidimensional NMR (19, 20)^{||**}. The anomeric torsions are consistent with expected values based on the exoanomeric effect. Whereas the peptide backbone torsions do

not match secondary structural motifs common in globular proteins, they fall in allowed regions of the Ramachandran map (see below). Comparison of the calculated structures revealed a remarkably well ordered core (with a rms deviation of only 1.17 ± 0.55 Å) consisting of the peptide backbone and the proximally linked glycans (Table 1). *Thus, compound 1 displays an unprecedented degree of order for a pentapeptide that approaches the structural stability of motifs found in folded proteins.*

The structure, as organized, displays two faces, one of which is primarily a carbohydrate surface, whereas the other presents a comparatively smaller peptide component. This structure is consistent within the larger mucin context, where the carbohydrate is directed to the exterior while the polypeptide is elongated to maximize accessibility of the glycodomain (1). The paucity of NOE interactions between the peripheral sugars and the core glycodomain suggests that distal glycan components play little role in determining the core mucin structure. Indeed, when **2** and **3** were examined in comparison with **1**, the NOE patterns corresponding to the core residues were virtually identical to those of the trisaccharide cluster. This homology of **1–3**, which does not extend to **4**, indicates the specific role of the α -linkage, but not the β -linkage, in inducing the secondary structure observed. The strong NOEs between the methyls of the GalNAc acetyl and the peptide indicate the GalNAc acetyls are probably necessary to maintain the observed structure, consistent with *N*-acetyl dependent conformation of monoglycosylated peptides (10, 13, 18). Thus, the core glycodomain, comprised of

^{||}Compounds **1–4** were synthesized as described (see refs. 7 and 8). Samples of synthetic peptide and the synthetic glycopeptides were dissolved in D₂O or 90% H₂O/10% D₂O with 10 mM phosphate buffer for NMR analysis. The pH was adjusted in each case to ≈ 4.5 . Sample concentrations were between 5 and 20 mM. NMR experiments were run at 18°C on Varian INOVA 600 and 800 spectrometers. 1D proton, two-dimensional heteronuclear multiple quantum coherence, heteronuclear multiple bond correlation, and double quantum correlation spectroscopy experiments were run in D₂O. One-dimensional proton, two-dimensional TOCSY (45-ms mixing period), NOESY (350-ms mixing period), and three-dimensional TOCSY–NOESY (45-ms and 500-ms mix times) were acquired in the H₂O/D₂O solvent by using the WATERGATE method for suppressing the solvent signal. The NOESY and TOCSY–NOESY experiments were used to determine NOE restraints for structure calculations, the one-dimensional experiment to determine the backbone couplings, and the other experiments for resonance assignments. Interproton distances were obtained from analysis of cross-peak intensities in NOESY spectra of **1** and were classified into three categories with corresponding bounds: strong (2.4 ± 0.6 Å), medium (2.9 ± 1.1 Å), and weak (3.4 ± 1.6 Å). Initial structural model of **1** was constructed in extended geometry (with sugars in chair form) by using INSIGHT II/BIOPOLYMER [Molecular Simulations (MSI), San Diego, CA]. Atom types and force field parameters were assigned corresponding to PARALLHDG.PRO (Version 4.02, M. Nilges) for the pentapeptide. For the glycans, ideal geometry values were taken from CHARMM19 (MSI) with bond, angle, and improper force constants set to match those in PARALLHDG.PRO. Nonbonded interactions were computed with the van der Waals term; the electrostatic term was turned off. The structure was minimized with the covalent energy terms to idealize the covalent geometry in the present force field. Peptide backbone and sugar linkage torsions were randomized to generate starting conformations. Structure refinement was carried out with distance and three-bond *J*-coupling restrained torsion angle dynamics in X-PLOR Version 98.0 (MSI) (see ref. 20) by using an optimized version (MSI) of the TAD protocol (see ref. 19), during which the conformation of each sugar ring was maintained rigid. The simulated annealing protocol consisted of 15 ps of dynamics computed with a time step of 15 fs at 50,000 K followed by cooling to 0 K in 15 ps and 2,000 steps of energy minimization. From a total of 100 computations, 20 final structures were chosen based on the criterion of least restraint violation. Color figures were prepared with INSIGHTII.

**Coordinates have been deposited with the Protein DataBank

Table 1. Statistical analysis of NMR and computed structural data for **1**

Distance restraints	Observed
Total	116
Intraresidue	
Pentapeptide	23
Glycans	10
Sequential ($ k-j = 1$)	
Pentapeptide	28
Glycans	8
Medium range ($2 \leq i-j \leq 4$)	
Pentapeptide	1
Glycans	2
Pentapeptide–Glycans	
Self*	29
Other†	13
3-bond <i>J</i> -coupling restraints	
Pentapeptide	5
Glycans	6
Structure Statistics	Value
NOE violations	
Number >0.2 Å	2
Number >0.5 Å	0
Three-bond <i>J</i> -coupling violations	
Number >0.25 Hz	0
Deviations from ideal covalent geometry	
Bond length, Å	0.013 ± 0.005
Bond angle, Å	2.7 ± 0.4
Improper, deg	1.3 ± 0.4
Pairwise rms deviation among 20 final structures, Å	
Peptide backbone + (S1G1, T2G1, T3G1) rings	1.17 ± 0.55
Pentapeptide + (S1G1, T2G1, T3G1) heavy atoms	1.36 ± 0.60

*Between peptide residue and its attached glycans. These NOEs were limited to the proximally linked GalNAc *N*-acetyl-methyl group.

†Between peptide residue and glycans on other peptide residues.

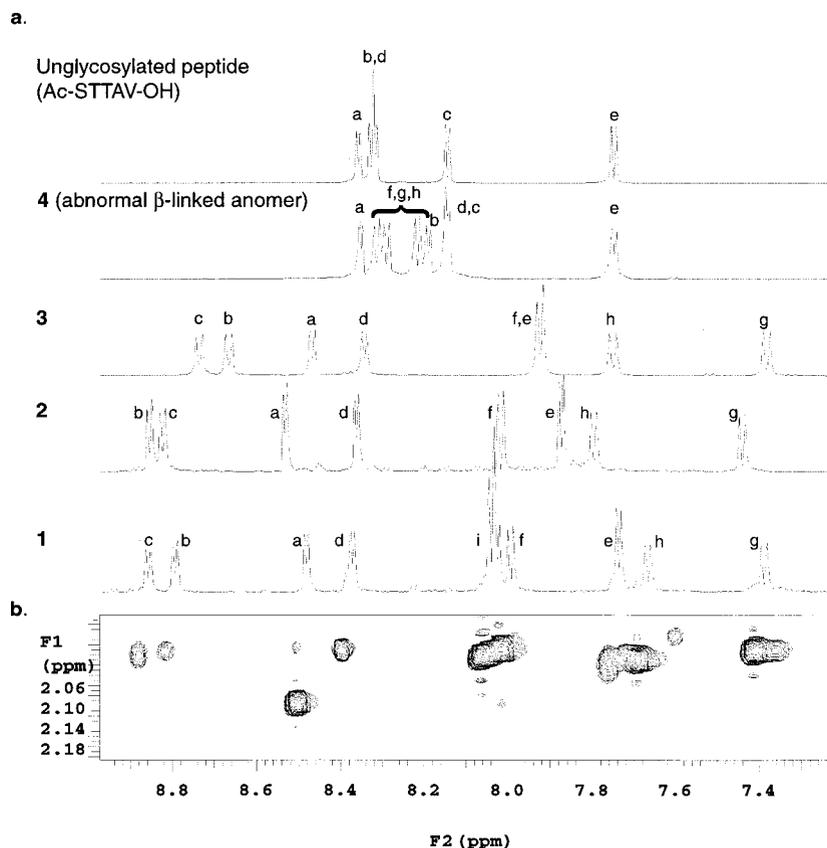


FIG. 2. One- and two-dimensional NMR analysis^{||} demonstrates that sequential *O*-glycosylation of α -linked GalNAc induces transition to a highly stable structure. (a) The amide region of the ¹H-NMR for the peptide Ac-S TTAV-OH, the β -linked **4**, and the α -linked series **3**, **2**, and **1**. The amide proton resonances of the peptide backbone, S TTAV, are labeled a, b, c, d, and e, respectively. The GalNAc amide resonances on the S TT segments of **1–4** are labeled f, g, and h, respectively, and the sialic acid amide signals are labeled i. Comparison of these spectra shows that NH patterns in abnormal (β -linked) anomer **4** differ little from the unglycosylated peptide in contrast to the spectra of α -linked clusters, which reveal dramatic dispersion of resonances. (b) Two-dimensional NOESY measurements reveal strong interactions between the methyl protons of the GalNAc acetyl groups and the peptide amide protons for **1**. Similar interactions were observed for α -linked **2** and **3**. No such interactions were observed in the β -linked **4**.

an amino acid and an α -*O*-GalNAc, dictates the organization of the mucin glycopeptide backbone into a scaffold on which the carbohydrate extensions are mounted, relatively unhindered in their conformational disposition, allowing the display of antennary glycans.

The stability of the core conformation is apparent from our data and the peptide backbone angles fall in allowed regions. However, the overall fold does not fit into one of the canonical classes of polypeptide secondary structure.** This organization is apparently because of conformational accommodations necessary to form a compact structure that also incorporates large branching sidegroups (starting with the proximal GalNAc), which have no counterpart in a nonglycosylated peptide of comparable size. Whereas the molecular details of the motif are novel, the elongated peptide dimensions are consistent with dimensions derived from electron micrographs of cell-surface mucin proteins (26–28). Indeed, the persistence of the backbone fold in the series of analogues we have examined demonstrates that the elongated secondary structure is energetically stable and suggests that *this may be a common motif in the nonglobular structure of mucin glycoproteins.*

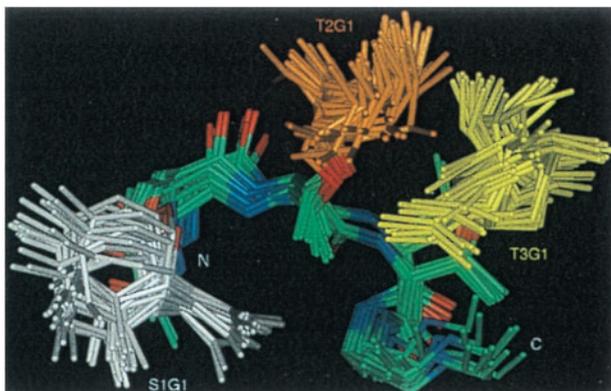
We have reported immunological characterization of constructs related to **1**, **2**, and **3** and have shown that they elicit robust antibody responses that crossreact with tumor cells displaying the corresponding antigen (8). Hence, Fig. 3 also represents the epitope recognized by antibodies stimulated by our potential vaccines and the probable epitope of other vaccine candidates and related structures (6, 21). Tumor-associated mucins have the same GalNAc core as normal

mucins, and the structure we observed is independent of antennary glycans. Thus, normal mucins should present the same scaffold as tumor-related mucins, except that they are more highly glycosylated, effectively concealing the carbohydrate scaffold and the proximal peptides (1, 5).

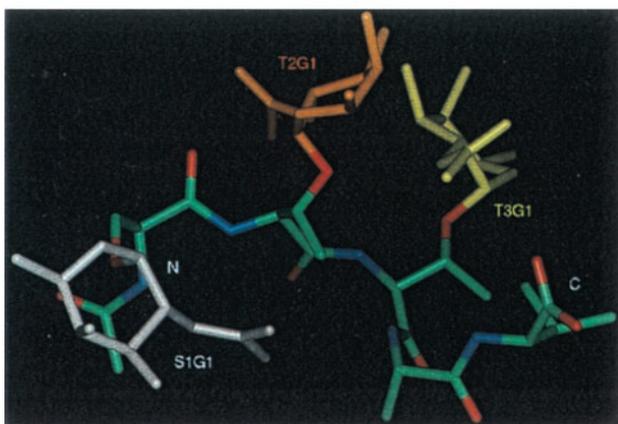
In combination with other conformational studies of *O*-glycosylated peptides, our findings suggest a progression from flexible peptide to the stable elongated structure of mucin proteins. Peptide flexibility is reduced on monoglycosylation with GalNAc (12, 18) and further restricted on addition of a second GalNAc (3, 9). These low degrees of glycosylation result in modest β -turn formation usually in the vicinity of the glycodomain and appear to be sequence sensitive. On formation of the clustered triad, the structure converges to a stable elongated motif. Incomplete biosynthetic elaboration of the glycan core, like that associated with certain carcinomas (1, 3, 5, 24), results in less ordered cell-surface architectures. Subsequent glycosylation in the normal course, leading to fully mature mucins, yields a stable elongated protein that polyvalently displays the glycans necessary for functional recognition (1, 4, 25–29).

In summary, sequential *O*-glycosylation of α -linked GalNAc induces a transition to an unprecedented secondary structure that approaches the stability of motifs found in folded proteins. We have shown that the α -linkage and sequential placement of GalNAc are required for this mucin motif to emerge. It is likely that the acetyl group on the GalNAc residue is also necessary to support structural coherence. Furthermore, installation of the initial α -*O*-

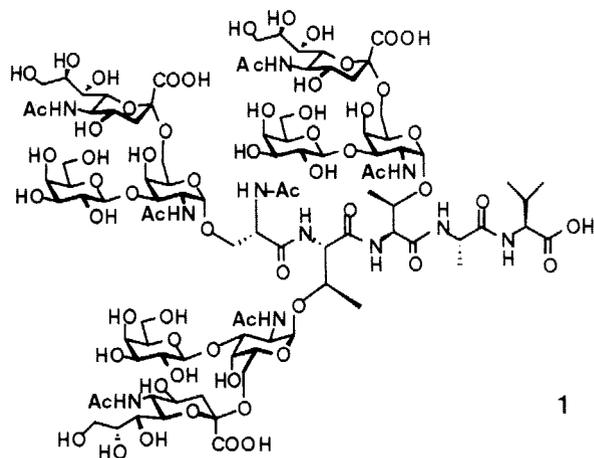
a



b



c



1

FIG. 3. Structure of the mucin glycopeptide 1^{ll}. (a) Superimposed view of the 20 final structures. The superposition was performed on the backbone atoms of the peptides and the ring atoms of the displayed sugars. The relatively unrestrained and unstructured distally linked glycans are omitted for clarity. The amino and acid termini are labeled N and C, respectively. The α -O-GalNAc residues are labeled S1G1, T2G1, and T3G1 for the STT segment. Carbon, nitrogen, and oxygen atoms are colored green, blue, and red, respectively, on the pentapeptide. (b) View of the structure that has lowest rms deviation to the average of the 20 structures shown in a. (c) STF cluster 1.

GalNAc residue in a cluster domain creates a stable scaffold that can accept, without intrinsic change, increased glycosylation. We note that a variety of carbohydrate structures can be accommodated in this way so that the same protein backbone can display a variety of glycans, the nature of which reflect the physiological state of the cell.

We thank G. Sukenick for mass spectroscopic assistance. This work was supported by the National Institutes of Health (N.I.H.) [grant numbers AI-16943, CA-28824 (S.J.D.) and CA-08748 (Sloan Kettering Institute to S.J.D.)], the National Science Foundation (N.S.F.) [grant number NSF/BIR-9601477 (University of Minnesota), and Graduate School Initiative in Structural Biology (University of Minnesota) (D.H.L.)]. Postdoctoral research fellowships are gratefully acknowledged by L.J.W. [N.I.H. (1F32CA79120-01)]; J.B.S. [N.I.H. (F3218804)]; S.D.K. [U.S. Army Breast Cancer Grant (DAMD 17-98-1-8154)]; P.W.G. [American Cancer Society (PF 98026)]; and D.S. [The Irvington Institute and M.R. Bloomberg].

- van den Steen, P., Rudd, P. M., Dwek, R. A. & Podenakker, G. (1998) *Crit. Rev. Biochem. Mol. Biol.* **33**, 151–208.
- Rudd, P. M. & Dwek, R. A. (1997) *Crit. Rev. Biochem. Mol. Biol.* **32**, 1–100.
- Koganty, R. R., Reddish, M. A. & Longenecker, B. M. (1997) in *Glycopeptides and Related Compounds: Synthesis, Analysis and Application*, eds. Large, D. G. & Warren, C. D. (Dekker, New York), pp. 707–743.
- Carlstedt, I. & Davies, J. R. (1997) *Biochem. Soc. Trans.* **25**, 214–219.
- Lloyd, K. O., Burchell, J., Kudryashov, V., Yin, B. W. T. & Taylor-Papadimitrou, J. T. (1996) *J. Biol. Chem.* **271**, 33325–33334.
- Toyokuni, T. & Singhal, A. K. (1995) *Chem. Soc. Rev.* **24**, 231–242.
- Sames, D., Chen, X.-T. & Danishefsky, S. J. (1997) *Nature (London)* **389**, 587–591.
- Kuduk, S. D., Schwarz, J. B., Chen, X.-T., Glunz, P. W., Sames, D., Ragupathi, G., Livingston, P. O., Danishefsky, S. J. (1998) *J. Am. Chem. Soc.* **120**, 12474–12485.
- Liu, X., Sejbál, J., Kotovych, G., Kohanty, R. R., Reddish, M. A., Jackson, L., Gandhi, S. S., Mandenca, A. J. & Longenecker, B. M. (1995) *Glycoconjugate J.* **12**, 607–617.
- O'Connor, S. E. & Imperiali, B. (1998) *Chem. Biol.* **5**, 427–437.
- van Zuylen, C. W. E. M., de Beer, T., Leeflang, B. R., Boelens, R., Kaptein, R., Kamerling, J. P. & Vliegthart, J. F. G. (1998) *Biochemistry* **37**, 1933–1940.
- Huang, X., Barchi, J. J., Jr., Lung, F.-D. T., Roller, P. P., Nara, P. L., Muschik, J. & Garrity, R. R. (1997) *Biochemistry* **36**, 10846–10856.
- Elofsson, M., Walse, B. & Kihlberg, J. (1996) *Int. J. Pept. Protein Res.* **47**, 340–347.
- Mer, G., Hietter, H. & Lefevre, H. J.-F. (1996) *Nat. Struct. Biol.* **3**, 45–53.
- Weller, C. T. *et al.* (1996) *Biochemistry* **35**, 8815–8823.
- Pieper, J., Ott, K.-H. & Meyer, B. (1996) *Nat. Struct. Biol.* **3**, 228–232.
- Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R. N., Smolyar, A., Reinherz, E. L. & Wagner, G. (1995) *Science* **269**, 1273–1278.
- Liang, R., Andreotti, A. H. & Kahne, D. (1995) *J. Am. Chem. Soc.* **117**, 10395–10396.
- Stein, E. G., Rice, L. M. & Brunger, A. T. (1995) *J. Magn. Res.* **124**, 154–164.

20. Brunger, A. T (1992) X-PLOR Version 3.1, A System for X-Ray Crystallography and NMR (Yale University Press, New Haven, CT).
21. Nakada, H., Inoue, M., Numata, Y., Tanaka, N., Funakoshi, I., Fukui, S., Mellors, A. & Yamashina, I. (1993) *Proc. Nat. Acad. Sci. USA* **90**, 2495–2499.
22. Saitoh, O., Gallagher, R. E. & Fukuda, M (1991) *Cancer Res.* **51**, 2854–2862.
23. Pallant, A., Eskenazi, A., Mattei, M. G., Rournier, R. E. K., Carlsson, S. R., Fukuda, M. & Frelinger, J. G. (1989) *Proc. Nat. Acad. Sci. USA* **86**, 1328–1332.
24. Fukuda, M., Carlsson, S. R. & Klock, J. C. D. (1986) *J. Biol. Chem.* **261**, 12796–12806.
25. Varki, A. (1993) *Glycobiology* **3**, 97–130.
26. Cyster, J. G., Shotton, D. M. & Williams, A. F. (1991) *EMBO J.* **10**, 893–902.
27. Li, F. *et al* (1996) *J. Biol. Chem.* **271**, 6342–6348.
28. Rose, M. C., Voter, W. A., Sage, H., Brown, C. F. & Kaufman, B. (1984) *J. Biol. Chem.* **259**, 3167–3172.
29. Gerken, T. A., Butenhof, K. J. & Shogren, R. (1989). *Biochemistry* **28**, 5536–5543.
30. Wu, W.-G., Pasternack, L., Huang, D.-H., Koeller, K. M., Lin, C.-C., Seitz, O. & Wong, C.-H. *J. Am. Chem. Soc.*, in press.
31. Simanek, E. E., Huang, D.-H., Pasternack, L., Machajewski, T. D., Seitz, O., Millar, D. S., Dyson, H. J. & Wong, C.-H. (1998) *J. Am. Chem. Soc.* **120**, 11567–11575.
32. DeFrees, S. A., Kosch, W., Way, W., Paulson, J. C., Sabesan, S., Halcomb, R. L., Huang, D.-H., Ichikawa, Y. & Wong, C.-H. (1995) *J. Am. Chem. Soc.* **117**, 66–79.
33. DeFrees, S. A., Gaeta, F. C. A., Lin, Y.-C., Ichikawa, Y. & Wong, C.-H. (1993) *J. Am. Chem. Soc.* **115**, 7549–7550.