BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **219,** 157–162 (1996)
ARTICLE NO. 0198

Effect of Serine *O*-Glycosylation on *Cis–Trans* Proline Isomerization

Ya-Lan Pao, Mark R. Wormarld, Raymond A. Dwek, and Annemarie C. Lellouch

Department of Biochemistry, Oxford Glycobiology Institute, Oxford University, South Parks Road, Oxford OX1 3QU, United Kingdom

Received December 28, 1995

In order to examine the ability of an *O*-glycosylatedserine residue preceding proline to stabilize a *cis* amide conformation in a fashion similar to that observed with aromatic amino acid residues, we prepared a series of glycosylated analogs of a small linear peptide which we have previous reported to contain a *cis* conformation of an amide bond between tyrosine and proline. The glycopeptides were prepared by incorporating glycosylated N^{α}-(fluoren-9-yl) methoxycarbonyl (Fmoc) amino acids into a standard solid phase peptide synthesis protocol. The peptides and glycopeptides were analyzed using 2-dimensional NMR spectroscopy. Unlike the case where the residue preceding proline was tyrosine, no signals corresponding to a *cis* proline conformation were detected in the spectra of the two glycopeptides containing serine *O*-glycosylated with either β-linked *N*-acetyl glucosamine or α-linked *N*-acetyl galactosamine in the position preceding proline.

Proline residues have often been associated with conformational changes in peptides and proteins, in part because of their unique ability to undergo *cis-trans* isomerization about the amide bond linking proline and the preceding amino acid residue. Crystallographic and NMR studies of proteins and peptides have revealed that *cis* proline amide bonds, though rare, are indeed present in a variety of molecules such as staphylococcal nuclease [1], insulin [2], and salmon calcitonin [3]. While proline is commonly known to terminateα-helices, or define the limits of β-sheets, *cis*-proline amide bonds have been shown to alter the backbone chain direction through type VI turns [4, 5]. In addition, the dynamic process of the *cis-trans* isomerization has been proposed to play a role in the slow phase of protein folding [6]. Studies of small linear peptides [7, 8] have shown that the aromatic residues, Tyr, Phe, His, and Trp, in the position preceding proline, are more efficient than other natural occurring amino acids in stabilising the *cis*-amide bond conformation, a property proposed to be due to the steric bulk of their side chains.

Glycosylation is an important post translational modification in eukaryotes that can take a variety of forms [9]. In *N*-linked glycosylation, the oligosaccharide is attached to Asn via a β-*N*-glycosidic linkage at an Asn-Xxx-Ser/Thr consensus sequence, where Xxx can be any amino acid except proline. In *O*-linked glycosylation, such as that found commonly on mucoproteins and mucin-like protein domains, the large and often heterogeneous oligosaccharide chains are linked to serine or threonine through an α-anomeric linkage of a core *N*-acetyl galactosamine (GalNAc) residue. An additional form of *O*-glycosylation has been identified in which a single residue of *N*-acetyl glucosamine (GlcNAc) is attached via a β-*O*-glycosidic linkage to serines and threonines of cytoplasmic and nuclear proteins [10]. Both mucin-like and cytoplasmic *O*-glycosylation are often found in protein sequences rich in proline. Furthermore, such *O*-glycosylation has been reported to be found on residues immediately preceding a proline residue, such as threonine 21 of mammalian neurofilament NF-L [11] or in the mucin like hinge domain of human IgA1 [12].

Because glycosylation would impart a greater degree of bulk to a serine or threonine side chain, we chose to examine the possibility that a glycosylated serine residue, such as Ser(β-GlcNAc) or Ser(α-GalNAc), in the position preceding proline could also stabilize a *cis* conformation in a proline amide bond, in a fashion analogous to an aromatic amino acid. Such a finding might help define the role of *O*-glycosylation in influencing protein conformational properties.

We have previously reported the synthesis and study of analogues of the Sendai Viral Nucleoprotein antigen, in which we found a significant population of *cis*-amide conformation between the

0006-291X/96 $18.00

Tyr and Pro residues in a peptide K3 (FAPSNYPAL), but none in the Ala-Pro linkage [13]. Using glycosylated Fmoc amino acids, Fmoc(Ac_3-α-D-GalNAc)OPfp and Fmoc-Ser(Ac_3-β-GlcNAc)OH [13], we prepared analogues of the K3 sequence in which we replaced the Tyr residue with either Ser, Ser(α-GalNAc) or Ser(β-GlcNAc) (fig 1). Here we report the synthesis of these compounds as well as the results of 2-D ^{1}H-NMR experiments in which we found that the all the proline residues exist only in the *trans* amide conformation.

MATERIALS AND METHODS

N-Acetyl glucosamine, peracetylated *N*-acetyl glucosamine, *N*-acetyl galactosamine and 2,3,6-tri-*O*-acetyl galactal were purchased from Sigma. All Fmoc amino acids, Fmoc-Leu Wang Resin and (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP) were purchased from Novabiochem. Hydroxybenztriazole (HOBt), N,N′-dicyclohexylcarbodiimide (DCCI), silver trifluoromethanesulfonic acid, N,N-diisopropylethylamine (DIEA), and penta-f-luorophenol were purchased from Fluka. Peptide synthesis grade dimethylformamide (DMF) was purchased from Rathburn and used directly for synthesis. Immediately before use, dichloromethane (DCM), ethyl acetate and acetonitrile (CH_3CN) were dried by distillation from P_2O_5 and kept over molecular sieves. Triethylamine was distilled from potassium hydroxide and kept over molecular sieves. Thioacetic acid was multiply distilled under reduced pressure. Concentrations were performed under reduced pressure at temperatures <40°C. TLC was performed on Silica Gel 60 F_{254} (Merck) with detection by UV light and charring with H_2SO_4. Column chromatography was performed on a Merck Silica Gel 60 (230–400 mesh, 40–60 μm) under dry conditions with distilled solvents.

Fmoc-glycosyl amino acid preparation. The full experimental detail of the synthesis of Fmoc-Ser(Ac_3-β-GlcNAc)OH and Fmoc(Ac_3-α-D-GalNAc)OPfp have been described by us recently elsewhere [13]. In brief, Fmoc-Ser(Ac_3-β-GlcNAc)OH was prepared in a one-step reaction by glycosylation of Fmoc-Ser-OH with peracetylated GlcNAc using boron-trifluoride-diethyl ether as a Lewis acid catalyst. Fmoc-Ser(Ac_3-α-D-$GalN_3$)OPfp was achieved by the glycosylation of Fmoc-Ser-OPfp with tri-*O*-acetyl-2-azido-2-deoxy-α-D-galactosyl bromide using silver triflate as a catalyst to afford a mixture of α and β glycosides of Fmoc(Ac_3-α-D-$GalN_3$)OPfp. The 2-azidogalactosyl bromide derivative was prepared from tri-O-acetyl-D-galactal by using the azidonitration and bromination methods described by Lemieux and Ratcliffe [14]. The pentafluorophenylester derivative was prepared by esterifaction of Fmoc-Ser-OH with pentafluorphenol in the presence of N,N′-diisopropyl-carbodiimide (DIC). The desired α-anomer, Fmoc-Ser(Ac_3-α-D-$GalN_3$)OPfp, was purified from the crude by column chromatography on silica gel and by HPLC. The transformation of the azido group into acetamide was achieved after the completion of the peptide chain on the solid support by using multiply distilled thioacetic acid.

Glycopeptide synthesis and characterization. The peptide and glycopeptides were prepared as described previously. All analytical and preparative reverse phase HPLC was performed on a Waters HPLC system. Analytical HPLC was performed using a Vydac C4 column (214TP54, 5 mm, 250mm × 4.6 mm) run at 1.0 ml/minute. HPLC purification of each compound was performed on a Vydac C4 column (214TP54, 5 mm, 250mm × 10 mm) run at 4.0 ml/min. Solvent system used was A: 0.1% TFA in H_2O and B: 0.1% TFA in CH_3CN. Matrix-assisted Laser Desorption/Ionization mass spectrometry (MALDI) was performed on either a Fisons VG Autospec QFPD using 2,5-dihydroxybenzoic acid as a matrix or on Finnegan Lasermat Time of Flight mass spectrometer (Finnigan MAT), using a-cyano-4-hydroxy cinnamic acid and Renin (m/z = 1760 M + H) as internal standard. For hexosamine and amino acid analysis, glycopeptide samples were hydrolysed in vapor phase HCl (6N HCl) for either 6h at 100°C (hexosamine analysis) or 24 h at 110°C (amino acid composition).(*d*,*l*)-2-aminobutyric acid (Aldrich 16,266-3) was used as an internal standard in both cases. The hydrozylates were derivatized with phenylisothiocyanate according to the Waters Picotag manufacturer instructions. The phenylthiocarbamyl derivatives were separated by reverse phase HPLC on a 5μm Spherisorb ODS2 column (0.46 × 25 cm) at 50°C using modifications of the Picotag solvent and gradient system.

L-Phenylalanyl-L-alanyl-L-prolyl-L-seryl-L-asparagyl-L-tyrosyl-L-prolyl-L-alanyl-L-leucine acid **K3** peptide was synthesized (32 mmol scale) and cleaved from the resin as described previously. Purification by HPLC (Vydac C4 semi-preparative column gradient 90:10 to 40:60 (A:B) in 30 min, retention time, 20.2 min), gave K3 (14 mg, 46% overall yield). MALDI m/z = 999 (M + Na) (Calcd. 979.08). Amino acid analysis was as expected. ^{1}H-NMR see Table I.

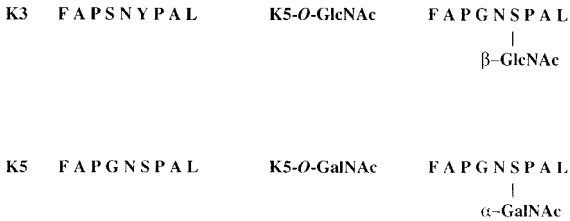

FIG. 1. Amino acid sequences of the peptides and glycopeptides used in this study.

L-Phenylalanyl-L-alanyl-prolyl-L-glycyl-L-asparagyl-L-seryl-L-prolyl-L-alanyl-L-leucine acid **K5** peptide was synthesized (22 μmol scale) and cleaved from the resin as described previously. HPLC (Vydac C4 semi-preparative column gradient 90:10 to 50:50 (A:B) in 30 min, retention time 13.3 min) gave K5 (2.1 mg, 11% yield). MALDI m/z = 901 (M + Na) (Calcd. 873). Amino acid analysis as expected. ^{1}H-NMR see Table I.

L-Phenylalanyl-L-alanyl-prolyl-L-glycyl-L-asparagyl-O-(2-N-acetyl-β-D-glucopyranosyl)-L-seryl-L-prolyl-L-alanyl-L-leucine **K5-*O*-GlcNAc** glycopeptide was synthesized (22 μmol scale) and cleaved from the resin as described previously. The crude cleavage product was purified by HPLC (Vydac C4 semi-preparative column gradient 95:5 to 50:50 (A:B) in 30 min, retention time 21.6 min), to give the *O*-acetylated glycopeptide. MALDI m/z = 1230 (M + 23). Deacetylation and purification by gel filtration gave K5-*O*-GlcNAc (10.4 mg, 39% overall yield). LD-MS m/z = 1076.8 (M + H) (Calcd. 1077.16). Amino acid and hexosamine analysis as expected. ^{1}H-NMR see Table I.

L-Phenylalanyl-L-alanyl-L-prolyl-L-glycyl-L-asparagyl-O-(2-N-acetyl-α-D- galactopyranosyl)-L-seryl-L-prolyl-L-alanyl-L-leucine **K5-*O*-GlcNAc** glycopeptide was synthesized (22 μmol scale), treated with thioacetic acid and cleaved from the resin as described. The crude cleavage product was purified by HPLC (Vydac C4 semi-preparative column gradient 95:5 to 50:50 (A:B) in 30 min, retention time 21.0 min), to give the *O*-acetylated glycopeptide. MALDI m/z = 1222 (M + Na). Deacetylation and purification by gel filtration gave K5-*O*-GalNAc (9.4 mg, 35% overall yield). LD-MS m/z = 1098.5 (M + Na) (Calcd. 1077.16). Amino acid and hexosamine analysis as expected. ^{1}H-NMR see Table I.

1H-NMR spectroscopy. The ^{1}H-NMR spectra were recorded with a Varian Unity 500 spectrometer. First-order chemical shifts were obtained from one-dimensional spectra. Proton resonances were assigned from two dimensional (2D) COSY, TOCSY (80 msec mixing time) and ROESY (200 msec mixing time) spectra obtained for each sample at 30°C. 2D spectra were processed with sine bell and cosine bell weighting functions as necessary. Glycopeptide samples were prepared for ^{1}H-NMR spectroscopy by lyophilizing several times from ^2H_2O. Samples used were typically 5 mg glycopeptide in 500 μl unbuffered ^2H_2O with an internal standard (0.75% 3-(trimethylsilyl)propionic-2,2,3,3-d_4-acid sodium salt).

RESULTS AND DISCUSSION

Proline is the only amino acid residue found in eukaryotes for which the *cis* amide bond is energetically accessible. The lack of favourable hydrogen bonds stabilizing the *trans* conformation as well as the presence of unfavorable steric interactions of the type CαH-CαH or CαH-CδH in both the *cis* and the *trans* conformations, result in a sufficiently similar stability in the two isoforms to allow both the *cis* and the *trans* conformations of amide bonds involving proline to exist in solution [15]. Studies of dipeptides and small peptides have shown that the nature of the amino acid which precedes proline has a greater influence on the amide bond conformation than the residue following proline (except when the proline is in the C-terminal position in which case the free carboxyl-group may stabilise certain conformations) [16, 17]. Such studies have also shown that the relative quantity of *cis* conformation is greatest when the residue X in an X-Pro amide bond is one of the aromatic amino acid residues (Trp, Tyr, Phe or even His). This finding has been proposed to be due to the steric hindrance imposed by these relatively large side chains.

We wished to examine the ability of an *O*-glycosylated amino acid residue in the position preceding proline to stabilize a *cis* amide conformation. To that end, two glycopeptides whose sequences are derived from a peptide we have shown previously to contain a *cis* proline amide bond were obtained by the incorporation of Fmoc glycosyl-amino acid building blocks into a standard solid phase peptide synthesis protocol. In our previous work, we found, as have others, that the chemical shifts of the proline ring signals in the *cis* conformation are shifted upfield from those resulting from the proline ring in the *trans* conformation (fig 2A) and that characteristic patterns of sequential NOEs are observed for the two conformations [13]. In the COSY spectra of either the K5 peptide or the K5-GalNAc (fig 2B) or K5-GalNAc glycopeptides, we observed spin system patterns from only one form of proline, the resonance chemical shifts being consistent with the *trans* conformation. The characteristic pattern of sequential NOEs observed in the ROESY spectra confirmed that all the proline residues were indeed in the *trans*-amide conformation. In addition, we observed in the COSY spectra of K5 GlcNAc and K5 GalNAc that the Ser αβ cross peaks were resolved as two separate resonances as compared to a single peak as in the spectrum of K5 (see Table I), indicating that the Ser side chain is rotating more slowly when glycosylated.

When compared to the side chains of Phe and Tyr, Ser (β- GlcNAc) and Ser(α-GalNAc) are large and if steric bulk were the only determinant, we might expect these glycosylated side-chains

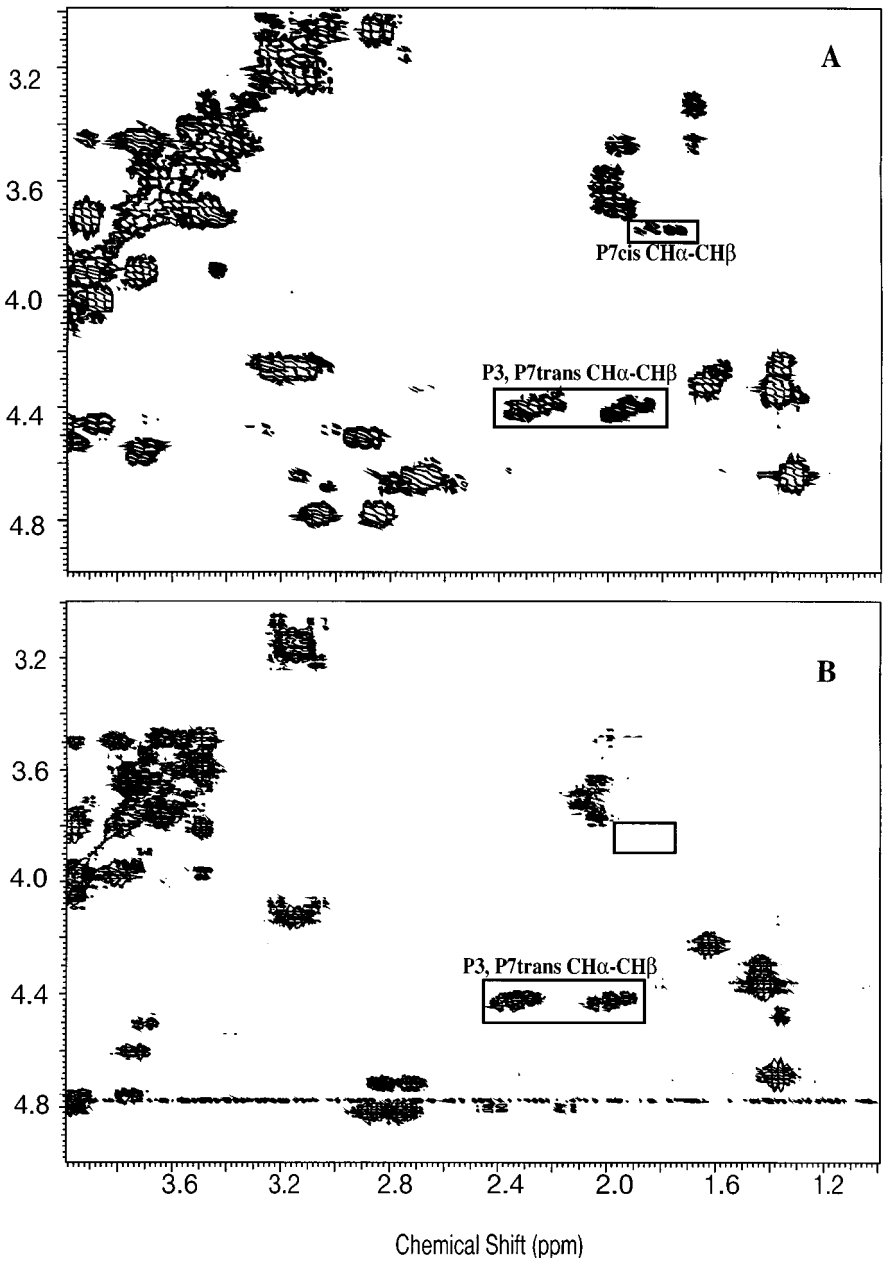

FIG. 2. (A) Selected region of the COSY spectrum of the peptide K3. Pro 3 gives a single set of CαH-CβH cross peaks, whereas Pro 7 shows two sets of cross peaks from the *cis*- and *trans*- conformers [13]. (B) Selected region of the COSY spectrum of the glycopeptide K5-GlcNAc. Both Pro 3 and Pro 7 show a single set of CαH-CβH cross peaks. The expected position for *cis*-proline cross-peaks has also been highlighted.

to cause *cis-trans* proline isomerization. The absence of such isomerization cannot simply be due to extra flexibility of the serine side chain compared to the aromatic amino acids, as the Ser 6 flexibility is reduced on glycosylation. Thus, our results suggest that bulk alone is not sufficient to stabilize a *cis*-amide conformation and that more specific and perhaps hydrophobic interactions are required. This is consistent with the observation that while an uncharged His residue will cause

TABLE 1
^{1}H NMR Chemical Shifts (ppm) in ^2H_2O Solution, T = 30°C, for the Peptides K3 and K5 and the Glycopeptides K5-GalNAc and K5-GlcNAc

Residue	Proton	K3	Residue	Proton	K5	K5-O-GalNAc	K5-O-GlcNAc
Phe-1	CαH	4.26	Phe-1	CαH	4.31	4.17	4.10
	CβH	3.19,3.19		CβH	3.27,3.20	3.18,3.15	3.12,3.12
	Aromatic	7.39,7.29		Aromatic	7.44,7.33	7.40,7.29	7.40,7.29
Ala-2	CαH	4.65	Ala-2	CαH	4.69	4.66	4.66
	CβH	1.33		CβH	1.40	1.35	1.34
Pro-3	CαH	4.42	Pro-3	CαH	4.45	4.40	4.38
	CβH	2.33,1.98		CβH	2.33,2.04	2.28,1.92	2.29,1.92
	CγH	2.02		CγH	2.04	2.06	2.06
	CδH	3.66		CδH	3.73	3.64,3.70	3.69
Ser-4	CαH	4.43	Gly-4	CαH	4.00	N.D.	N.D.
	CβH	3.85,3.85					
Asn-5	CαH	4.70	Asn-5	CαH	4.81	4.77	4.78
	CβH	2.77,2.68		CβH	2.87,2.78	2.82,2.77	2.82,2.70
Tyr-6	CαH	4.80t	Ser-6	CαH	4.74	4.85	4.74
		4.51c					
	CβH	3.03,2.83t		CβH	3.87,3.87	3.97,3.75	4.01,3.94
		2.89,2.89c					
	Aromatic	7.13,6.86					
Pro-7	CαH	4.39t	Pro-7	CαH	4.44	4.39	4.40
		3.78c					
	CβH	2.24,1.91t		CβH	2.33,2.04	2.32,1.97	2.34,1.98
		1.86,1.77c					
	CγH	1.96t		CγH	2.04	1.99	2.00
		1.70c					
	CδH	3.71,3.50t		CδH	3.73,3.73	3.74,3.64	3.73,3.61
		3.34,3.48c					
Ala-8	CαH	4.27	Ala-8	CαH	4.34	4.32	4.33
	CβH	1.31		CβH	1.40	1.39	1.39
Leu-9	CαH	4.29	Leu-9	CαH	4.41	4.18	4.19
	CβH	1.63		CβH	1.71	1.59	1.60
	CγH	1.56		CγH	1.71	1.59	1.60
	CδH	0.89,0.89		CδH	0.92,0.98	0.92,0.89	0.92,0.89
			HexNAc	C1H		4.92	4.57
				C2H		4.17	3.71
				C3H		3.89	3.57
				C4H		3.98	3.46
				C5H		3.64	3.77
				C6H		3.70	3.94
				C6′H		3.73	3.93

c, resonances from the conformer with *cis*-Pro 7; t, resonances from the conformer with *trans*-Pro 7.

proline isomerization, the protonated form will not [17]. It has been reported that for both *N*-linked oligosaccharides in peptides [18] and *O*-linked oligosaccharides in mucins [19] only the inner most oligosaccharide residues influence the peptide structure. However, our results indicate that Ser *O*-glycosylation containing a single saccharide residue, such as that found naturally with cytoplasmic Ser *O*-β-GlcNAc glycosylation, does not affect the *cis-trans* isomerisation of a following proline.

REFERENCES

1. Fox, R. O., Evans, P. A., and Dobson, C. M. (1986) *Nature* **320,** 192–194.
2. Higgins, K. A., Craik, D. J., Hall, J. G., and Andrews, P. R. (1988) *Drug. Des. Delivery* **3,** 159–170.

3. Amodeo, P., Morelli, M. A. C., and Motta, A. (1994) *Biochemistry* **33,** 10754–10762.
4. Chou, P. Y., and Fasman, G. D. (1978) *Ann. Rev. Biochem.* **47,** 251–276.
5. Frommel, C., and Preissner, R. (1990) *FEBS Lett.* **277,** 159–163.
6. Garel, J. R. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 795–798.
7. Grathwohl, C., and Wuthrich, K. (1976) *Biopolymers* **15,** 2043–2057.
8. Grathwohl, C., and Wuthrich, K. (1981) *Biopolymers* **20,** 2623–2633.
9. Rademacher, T. W., Parekh, R. B., and Dwek, R. A. (1988) *Ann. Rev. Biochem.* **57,** 785–838.
10. Holt, G. D., and Hart, G. W. (1986) *J. Biol. Chem.* **261,** 8049–8057.
11. Dong, D. L., Xu, Z. S., Chevrier, M. R., Cotter, R. J., Cleveland, D. W., and W., H. G. (1993) *J. Biol. Chem.* **268,** 16679–16687.
12. Mattu, T., *personal communication.*
13. Arsequell, G., Haurum, J. S., Elliott, T., Dwek, R. A., and Lellouch, A. C. (1995) *J. Chem. Soc. Perkin. Trans.* **1,** 1739–1745.
14. Lemieux, R. U., and Ratcliffe, R. M. (1979) *Can. J. Chem.* **57,** 1244–1251.
15. Zimmerman, S. S., and Scheraga, H. A. (1976) *Macromolecules* **9,** 408–416.
16. Juy, M., Hung, L.-T., Lintner, K., and Fermandjian, S. (1983) *Int. J. Peptide Protein Res.* **22,** 437–449.
17. Unkefer, C. J., Walker, R. D., and London, R. E. (1983) *Int. J. Peptide Protein Res.* **22,** 582–589.
18. Wormald, M. R., Wooten, E. W., Bazzo, R., Edge, C. J., Feinstein, A., Rademacher, T. W., and Dwek, R. A. (1991) *Eur. J. Biochem.* **198,** 131–139.
19. Gerken, T. A., Butenhof, K. J., and Shogren, R. (1989) *Biochemistry* **28,** 5536–5543.