The Influence of Flanking Sequence on the O-Glycosylation of Threonine in Vitro*

(Received for publication, August 5, 1992)

Brian C. O'Connell†, Fred K. Hagen‡, and Lawrence A. Tabak†

From the Departments of Dental Research and Biochemistry, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642

To investigate the influence of flanking amino acid sequence on the O-glycosylation of a single threonine residue in vitro, we have examined a series of 52 related peptides. The substrates were based upon a sequence from human von Willebrand factor which is known to be glycosylated in vitro ("PHMAQVT-VGPGL*†). Each residue of the parent peptide was substituted, in turn, with isoleucine, alanine, proline, glutamic acid, or arginine. Peptides were glycosylated using a UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase purified 15,000-fold from bovine colostrum by chromatography on DEAE-Sephacel, SP-Sephadex, Sepahcryl S-300, Affi-Gel Blue, and 5-mercuri-UDP-GalNAc thiopropyl-Sepharose. Single amino acid changes in the sequences flanking the threonine could profoundly alter the glycosylation of the substrate peptides. Substitution of any amino acid tested at positions +3, −3, and −2 markedly decreased O-glycosylation, as did the presence of a charged residue at position −1. The substitution of amino acids at the other positions of the peptide substrate had little effect on the incorporation of GalNAc.

Statistical analysis of sequences flanking known glycosylated threonine and serine residues suggests that they should be glycosylated with equal efficiency in the same sequence context (O'Connell et al., 1991). However, the bovine colostrum transferase failed to glycosylate a peptide derived from human erythropoietin which contains a serine that is glycosylated ("PPDAASAAPLR*†). When a threonine was substituted for the serine in this peptide ("PPDAATA-APLR*†), the substrate proved to be an excellent acceptor of GalNAc. These observations indicate that although flanking amino acid sequence is important for the O-glycosylation of specific hydroxyamino acids, discrete threonine- and serine-specific transferases may exist.

The initial step of mucin-type O-glycosylation is the attachment of N-acetylgalactosamine to a hydroxyamino acid. The addition of this monosaccharide extends the conformation of the protein backbone and is the point at which other sugars may be added to build an oligosaccharide side chain (Shogren et al., 1989; Gerken et al., 1989). In contrast to N-glycosylation, there is no consensus sequence for glycosylation by UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. A persistent problem with the detailed analysis of O-glycosylation is that the number of O-linked glycans which have been assigned to specific hydroxyamino acids has remained rather small, presumably due to the difficulty of directly sequencing O-linked glycoproteins (O'Connell et al., 1991).

Previous attempts at defining an O-glycosylation consensus signal have emphasized the importance of proline near the glycosylation site (Aubert et al., 1976; Wilson et al., 1991; Gooley et al., 1991). In particular, the observation that the Thr-Pro-Pro-Pro sequence in the bovine myelin protein A1 will glycosylate in vitro has dominated later work in this area (Hagopian et al., 1971; Young et al., 1979; Briand et al., 1981; Hughes et al., 1988; Cottrell et al., 1992) and has led to the proposal that O-glycosylation requires a turn in the substrate (Aubert et al., 1976; Eckhardt et al., 1987). However, the bovine myelin protein A1 is not glycosylated in vivo, and the tripolyol motif has been found in only a few glycoproteins (Smyth and Utsumi, 1967; Oppenheim et al., 1985). Moreover, there are several examples of O-glycosylated sites which lack proline within the surrounding flanking region (Honna et al., 1980; Rall et al., 1982).

Previously, we compiled a database of unambiguously-defined O-glycosylation sites and compared their flanking amino acid sequences with those of nonglycosylated serine and threonine residues (O'Connell et al., 1991). Our analysis suggested that both the type and position of amino acids near a potential glycosylation site could be important for glycosylation by the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. In particular, the presence of a proline, alanine, serine, or threonine at positions +3, −1, −6, and −3 was often associated with glycosylation of the hydroxyamino acid, whereas a charged residue at these positions was often associated with a serine or threonine which was not O-glycosylated (+ or − refers to residues that are N-terminal or C-terminal to a hydroxyamino acid). Other positions flanking the potential glycosylation site appeared to have no influence on glycosylation.

In the present study we have investigated the influence of each residue from −6 to +5, relative to threonine, on the glycosylation of a peptide substrate. The sequence of the substrates is based on a region of human von Willebrand factor that contains a single threonine which is glycosylated in vivo (Titani et al., 1986). Five amino acids were chosen to substitute each position of the parent peptide, representing residues with different properties, i.e. hydrophobic residues, charged residues, small side chains, and association with turns. The substrates for the in vitro glycosylation assays

* This work was supported in part by National Institutes of Health Grants DE-08108 and DE-086511 (to L. A. T.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
† Supported by Dentist-Scientist Award K16 DE-00159; these studies constitute work toward fulfillment of the Ph.D. degree.
‡ Supported by National Institutes of Health Grant T32 DE-07202.
§ To whom correspondence should be addressed. Tel.: 716-275-0770; Fax: 716-473-2679.
Amino acid substitutions in the sequence of the parent peptide are shown in bold type. The GalNAc incorporated by the purified transferase pool 1Al is given as the total counts/min under standard assay conditions (see "Experimental Procedures").

| Peptide designation | Peptide sequence | GalNAc incorporation |
|---------------------|------------------|---------------------|
| 1                   | PHMAQVTVGPGL     | 3,727               |
| 2                   | -6I   | PHMAQVTVGPGL     | 203                |
| 3                   | -6A   | PHMAQVTVGPGL     | 2,856              |
| 4                   | -6E   | PHMAQVTVGPGL     | 11,337             |
| 5                   | -6R   | PHMAQVTVGPGL     | 3,027              |
| 6                   | -5I   | PHMAQVTVGPGL     | 430                |
| 7                   | -5A   | PHMAQVTVGPGL     | 39                 |
| 8                   | -5P   | PHMAQVTVGPGL     | 4,801              |
| 9                   | -5E   | PHMAQVTVGPGL     | 7,218              |
| 10                  | -5R   | PHMAQVTVGPGL     | 596                |
| 11                  | -4I   | PHMAQVTVGPGL     | 22                 |
| 12                  | -4A   | PHMAQVTVGPGL     | 1,082              |
| 13                  | -4P   | PHMAQVTVGPGL     | 1,624              |
| 14                  | -4E   | PHMAQVTVGPGL     | 6,296              |
| 15                  | -4R   | PHMAQVTVGPGL     | 1,505              |
| 16                  | -3I   | PHMAQVTVGPGL     | 215                |
| 17                  | -3P   | PHMAQVTVGPGL     | 228                |
| 18                  | -3E   | PHMAQVTVGPGL     | 133                |
| 19                  | -3R   | PHMAQVTVGPGL     | 103                |
| 20                  | -2I   | PHMAQVTVGPGL     | 0                  |
| 21                  | -2A   | PHMAQVTVGPGL     | 502                |
| 22                  | -2P   | PHMAQVTVGPGL     | 413                |
| 23                  | -2E   | PHMAQVTVGPGL     | *                  |
| 24                  | -2R   | PHMAQVTVGPGL     | 280                |
| 25                  | -1I   | PHMAQVTVGPGL     | 559                |
| 26                  | -1A   | PHMAQVTVGPGL     | 2,281              |
| 27                  | -1P   | PHMAQVTVGPGL     | 6,446              |
| 28                  | -1E   | PHMAQVTVGPGL     | 577                |
| 29                  | -1R   | PHMAQVTVGPGL     | 229                |
| 30                  | +1I   | PHMAQVTVGPGL     | 520                |
| 31                  | +1A   | PHMAQVTVGPGL     | 3,102              |
| 32                  | +1P   | PHMAQVTVGPGL     | 13,501             |
| 33                  | +1E   | PHMAQVTVGPGL     | 6,035              |
| 34                  | +1R   | PHMAQVTVGPGL     | 1,648              |
| 35                  | +2I   | PHMAQVTVGPGL     | 305                |
| 36                  | +2A   | PHMAQVTVGPGL     | 18,723             |
| 37                  | +2P   | PHMAQVTVGPGL     | 12,854             |
| 38                  | +2E   | PHMAQVTVGPGL     | 6,384              |
| 39                  | +2R   | PHMAQVTVGPGL     | 9,434              |
| 40                  | +3I   | PHMAQVTVGPGL     | 247                |
| 41                  | +3A   | PHMAQVTVGPGL     | 295                |
| 42                  | +3E   | PHMAQVTVGPGL     | 302                |
| 43                  | +3R   | PHMAQVTVGPGL     | 256                |
| 44                  | +4I   | PHMAQVTVGPGL     | 230                |
| 45                  | +4A   | PHMAQVTVGPGL     | 835                |
| 46                  | +4P   | PHMAQVTVGPGL     | 18,179             |
| 47                  | +4E   | PHMAQVTVGPGL     | 3,049              |
| 48                  | +4R   | PHMAQVTVGPGL     | 1,713              |
| 49                  | +5I   | PHMAQVTVGPGL     | 965                |
| 50                  | +5A   | PHMAQVTVGPGL     | 1,839              |
| 51                  | +5P   | PHMAQVTVGPGL     | 371                |
| 52                  | +5E   | PHMAQVTVGPGL     | 585                |
| 53                  | +5R   | PHMAQVTVGPGL     | 2,504              |

*The peptide -2E could not be recovered in sufficient amounts to complete the assay.

The composition of each peptide was verified on at least two occasions by amino acid analysis. The peptides were stored lyophilized at -20 °C before use. Peptides substituted with proline, alanine, and glutamic acid were dissolved in 10 mM ammonium hydroxide at 10 nmol/μl. The arginine-containing peptides were solubilized in 1 mM acetic acid at 2 nmol/μl. The peptides containing isoleucine were the most difficult to solubilize, requiring the addition of 10% dimethyl-
formamide in water and briefly heated (65 °C) to dissolve at 1 mol/l.

Based on the sequence flanking a glycosylated serine residue in human erythropoietin (EPO), a set of six peptides was synthesized ("FPDASAAPLR"). In this case, the hydroxyamino acid was changed from the serine residue to threonine or a combination of serine and threonine. These peptides were named EPO-S, EPO-T, EPO-SS, EPO-TT, EPO-ST, and EPO-TS, depending on the hydroxyamino acids they contained (see Table II). The EPO series of substrates were made individually and purified by reverse-phase HPLC. The peptides were not acetylated so that the acceptor residue could be identified by direct sequencing.

**Purification of UDP-GalNAc-Polypeptide**

**N-Acetylgalactosaminyltransferase from Bovine Colostrum**

Buffers used in the enzyme purification were Buffer A, 25 mM imidazole, 6 mM MgCl2, 30 mM NaCl, pH 7.2. Buffer B, 25 mM MES, 5 mM MnCl2, 0.06% Triton X-100, 0.1 mM dithiothreitol, pH 6.5. Buffer C, Buffer B plus 0.1 M NaCl. Buffer D, Buffer C plus 0.5 mM uridine 5'-diphosphate. Buffer E, Buffer B plus 1 M NaCl. Buffer F, 1.5 M NaCl, 20 mM NaPO4, pH 7.5. Buffer G, 1.5 M NaSCN, 20 mM NaPO4, pH 7.5. Buffer H, Buffer C plus 20% glycerol. Buffer I, Buffer E plus 20% glycerol. Buffer J, 25 mM MES, 5 mM MnCl2, 0.1% Triton X-100, 0.1 mM dithiothreitol, 20% glycerol, pH 6.5. Buffer K, Buffer J plus 0.5 M NaCl. Buffer L, 25 mM MES, 5 mM MnCl2, 0.1% Triton X-100, 20% glycerol, 0.02% NaN3, pH 6.6. Buffer M, Buffer L plus 50 mM NaCl. Buffer N, Buffer L plus 30 mM NaCl. Buffer O, Buffer N plus 20 mM UDP-GalNAc.

**Step 1: Separation of Cellular Debris and Lipid Globules—Whole bovine colostrum (288 ml) was centrifuged for 10 h at 10,000 × g. The lipid layer and cellular pellet was separated from the supernatant by pouring through a few layers of cheese cloth. A minitangential flow apparatus (Millipore) was used to perform a buffer exchange twice into 10 volumes of Buffer A. The liquid phase of the colostrum was concentrated to 25 ml and then centrifuged at 80,000 × g for 90 min to remove any remaining lipid and insoluble material.

**Steps 2 and 3: DEAE-Sepacel and SP-Sephadex Chromatography—**The 80,000 × g supernatant was applied to a column (10 × 30 cm) of DEAE-Sepacel resin equilibrated with Buffer A. The flow-through fractions were detected by absorbance measurements at 280 nm and were pooled and applied directly to an SP-Sephadex column (10 × 30 cm) also equilibrated with Buffer A. The SP-Sephadex flow-through fractions were pooled and concentrated 3.5-fold in a minitangential flow apparatus.

**Step 4: Sephacryl S-300 Chromatography—**The SP-Sephadex flow-through fractions were pooled and concentrated to 25 mg/ml. The pool was incubated with 200 ml of Buffer E and collected in 6-ml fractions. The eluted bound protein (including transferase) was eluted from the column with 200 ml of Buffer E and collected in 6-ml fractions. The eluted material was pooled (150 ml) and immediately diluted to 750 ml with Buffer B. The resin was regenerated with 200 ml of Buffer F followed by 200 ml of Buffer G before proceeding to Step 6.

**Step 5: Affi-Gel Blue Chromatography—**Material from step 5 was loaded onto a column (25 × 25 cm) of Affi-Gel Blue, equilibrated with Buffer B. The column was eluted stepwise with 250 ml of Buffer B, 250 ml of Buffer C, and 200 ml of Buffer D containing 0.2 mM cymidine 5'-monophosphate. The remaining chromatography in this and subsequent steps was done with glycerol-containing buffers (2×) of Buffer H followed by the column and the enzyme was eluted with a linear gradient of 0.1 M NaCl to 1 M NaCl (125 ml of Buffer H and 125 ml of Buffer I), followed by 50 ml of Buffer I.

**Fractions of 3.8 ml were collected, monitored for absorbance at 280 nm and tested for transferase activity using a threonine and a serine-containing substrate. The eluant was divided into pool 1 (59 ml) and pool 2 (80 ml), whereupon each pool was added to an equal volume of Buffer J and then dialyzed against 2 liters of Buffer J in Spectra/Por 1 dialysis tubing (6000–8000 M, cut-off, 25.5 mm diameter). The Affi-Gel Blue resin was regenerated with Buffers F and G as in Step 5.

**Step 7: Affi-Gel Blue Resin Chromatography III—**Pool 1 from Step 6 was applied to 120 ml of Buffer J and then applied to a column (1 × 19 cm) of Affi-Gel Blue resin, equilibrated with Buffer J. The column was rinsed with 60 ml of Buffer J, and the transferase activity was eluted with a linear gradient of NaCl using 50 ml of Buffer J and 50 ml of Buffer K. The gradient was followed by 50 ml of Buffer K and finally 50 ml of Buffer J. The 2-ml fractions were assayed in the usual way as the previous steps. The transferase activity appeared to elute as two peaks, which were pooled separately and designated pool 1A (6 ml) and pool 1B (10 ml).

Pool 2 from step 6 (200 ml) was applied to the column (1 × 19 cm) of Affi-Gel Blue and then treated exactly the same as the first pool. Two peaks of transferase activity were again seen and designated pool 2A (5.5 ml) and 2B (6.5 ml). The four pools collected from Step 7 were dialyzed against 2 liters of Buffer L.

**Step 8: 5-Mercury-UDP-N-acetylgalactosamine Thiopropyl-Sepharose Chromatography—**The affinity resin used for this step was prepared essentially as described by Bendix and Schachter (1987) for 5-Hg-UDP-GlcNAc thiopropyl-Sepharose, except the UDP-GalNAc nucleotide sugar was substituted. The extent of mercuration of UDP-GalNAc was calculated by passage of an aliquot of the unincorporated nucleotide mixture (1.6 μmol) through 0.5 ml of reduced thiopropyl-Sepharose and then measuring the absorbance of the eluted material at 267 nm. The yield of 5-Hg-UDP-GalNAc was 69%.

The 5-Hg-UDP-GalNAc thiopropyl-Sepharose resin (25 μmol of ligand/ml of gel) was divided into two columns (0.7 × 3.9 cm), which were equilibrated with 15 ml of Buffer L. Pools 1A and 2B from Step 7 were applied to the columns and then to each was added 7.5 ml of Buffer L, 15 ml of Buffer M, and 7.5 ml of Buffer N. The transferase was recovered with 6 ml of Buffer O, collected as 600-μl fractions into microcentrifuge tubes. The fractions were assayed and the activity from each column was found to be in one major peak. The peak tubes of each column were combined as pools 1A and 2B (2.3 ml). An aliquot of each pool was dialyzed against Buffer L and used subsequently to assay the substrates. The enzyme was stored in Buffer L at −70 °C and was stable for several months.

**UDP-GalNAc-Polypeptide N-Acetylgalactosaminyltransferase Assays**

The enzyme activity of the UDP-GalNAc-polypeptide N-acetylgalactosaminyltransferase in bovine colostrum and purified fractions was determined by monitoring the transfer of [14C]-labeled UDP-GalNAc to a polypeptide substrate.

In the purification of GalNAc transferase Step 1 to 4, enzyme activity was assayed using the tetrapeptide acceptor Ac-TPPP (50 nmo/assay) in a final volume of 50 μl of 125 mM Tris, pH 7.1, 0.5% Triton X-100, 10 mM MnCl2, and 7.3 mM UDP-[14C]GalNAc (40,000 cpm/assay). The amount of enzyme used was such that <5% of the substrate was consumed in the reaction. After a 1-h incubation at 37 °C, the reaction was terminated by addition of 10 μl of 150 mM EDTA, followed by immediate separation of unincorporated UDP-[14C]GalNAc on a 200-μl Dowex 1-X8 anion exchange resin (in formate form). Pools from each step of the chromatography were also assayed in this way to determine the fold purification of the enzyme and the assays were repeated on three separate occasions.
During the transferase purification, column fractions and pooled material were assayed for activity toward other serine- and threonine-containing substrates. These substrates were based on a naturally occurring glycosylated serine residue in human erythropoietin, EPO-S ("PDPASAALPR") and EPO-T ("PDPATAALPR"). The assays consisted of 20 nmol of peptide, 40 mM sodium cacodylate, 0.1% Triton X-100, 4 mM 2-mercaptoethanol, 10 mM MnCl$_2$, and 14.6 µM UDP-[14C]GalNAc, at pH 6.5 in a volume of 50 µl, to which 2 µl of the column fraction to be assayed was added. The reactions were incubated in a 37 °C heat block for 45 min to 2 h and then terminated by the addition of EDTA.

Glycosylated peptide was separated from unincorporated nucleotide by glass fiber filter that was pretreated with BioBrene (Applied Bio-Systems, Inc.). The filter was subjected to 12 normal (Edman) cycles acid sequencing. The peptides EPO-T, EPO-TT, and EPO-ST were glycosylated essentially as outlined above, except the UDP[14C]GalNAc concentration was increased to 73 PM, and the reaction time of the glycosylated peptide (approximately 10 nmol) was applied to a C18 columns (Varian) using a Vac Elut (Analytichem) apparatus. Briefly, each column was wetted with 1 ml of methanol, followed by 1 ml of water containing 0.1% trifluoroacetic acid. Another 1 ml of water/trifluoroacetic acid was placed in the reservoir of the column, and the entire assay mixture was added and mixed by pipetting. The sample was then drawn through the columns, which was rinsed with 1 ml of water/trifluoroacetic acid. The glycosylated substrate was eluted directly into 7-ml glass scintillation vials using 1 ml of 30% acetonitrile, 0.1% trifluoroacetic acid in water. Universal ES (ICN) scintillation fluid was mixed with the eluted material and counted on a Beckman LS1801 liquid scintillation counter. The efficiency of counting was 97%.

**Determination of the Apparent $K_m$ of UDP-GalNAc and of Peptide Acceptors**

Assays containing 250 nmol of Ac-TPP, 40 mM sodium cacodylate, 0.1% Triton X-100, 4 mM 2-mercaptoethanol, 10 mM MnCl$_2$, and either 0.2 µl (1.4 ng) of enzyme 1AI or 0.5 µl (0.4 ng) of 2BI at pH 6.8. The concentration of UDP-[14C]GalNAc was varied from 0 to 36.5 µM, and the final volume was 50 µl.

For $K_m$ determinations of the peptides HVF, EPO-T, and EPO-TT, the UDP-[14C]GalNAc concentration was 36.5 µM. The acceptor concentration range was 0.4–8 mM. All the reactions were incubated for 15 min at 37 °C, after which they were immediately terminated and applied to the C18 columns as before. The assays were performed in triplicate and values for the apparent $K_m$ of UDP-GalNAc were calculated with the k*cat" program (BioMetallics, Inc.) using a constant relative error.

**In Vitro Glycosylation of Peptide Substrates**

In order to achieve the optimal reaction conditions for in vitro glycosylation, two peptide (HVF and EPO-T) were assayed using four buffer systems. Each reaction contained 50 mM MnCl$_2$, 14.6 µM UDP-[14C]GalNAc, 20 nmol of peptide, 7 ng of transferase 1AI, 0.1% Triton X-100, and one of the following: 125 mM Tris, pH 7.1, 40 mM sodium cacodylate, 4 mM 2-mercaptoethanol, pH 6.5, 40 mM MES, 4 mM 2-mercaptoethanol, pH 6.5, or 40 mM imidazole, 4 mM 2-mercaptoethanol, pH 6.5.

The substrate specificity of the two most purified transferase preparations, 1AI and 2BI, was tested with each of the 52 peptide substrates. The reactions contained 20 nmol of peptide, 40 mM sodium cacodylate, 0.1% Triton X-100, 4 mM 2-mercaptoethanol, 10 mM MnCl$_2$, and 18.25 µM UDP-[14C]GalNAc, at pH 6.5, in a volume of 50 µl. Either 1 µl (7 ng) of enzyme 1AI or 3 µl (24 ng) of 2BI was used in each reaction. The incubation time was 45 min at 37 °C, after which the reactions were terminated and separated on the C18 columns as described. Each set of assays was performed with a negative control (no peptide added) to determine the background number of counts/min found in the material that was eluted from the C18 columns. The background count (less than 5% of the maximum substrate value) was subtracted from the total counts obtained for each substrate.

**Amino Acid Sequencing of Peptides**

In order to determine which hydroxyamino acid(s) of the multisite substrates was glycosylated, the peptides were subjected to amino acid sequencing. The peptides EPO-T, EPO-TT, and EPO-ST were glycosylated essentially as outlined above, except the UDP-[14C] GalNAc concentration was increased to 73 µM, and the reaction time was 20 h. The unincorporated nucleotide was removed, and an aliquot of the glycosylated peptide (approximately 10 nmol) was applied to a glass fiber filter that was pretreated with BioBrene (Applied Biosystems, Inc.). The filter was subjected to 12 normal (Edman) cycles in an Applied Biosystems 473A sequence. The product of each cycle was collected, dried, and the counts/min determined.

**Protein Determination and Amino Acid Analysis**

To quantitate the test peptides and to verify their amino acid content, the peptides were hydrolyzed under standard conditions (vapor phase HCl, 106 °C, 20 h). The hydrolyzed material was then analyzed on a Hewlett-Packard amino acid analyzer with Amino Quant II software. Protein determination of the various purification pools was done in the same way as for the peptides, except that the buffers were first exchanged for water before hydrolysis, using Micron 10 ultrafiltration device (Amicon).

For rapid estimation of protein concentration during purification Step 4, the Lowry assay was used.

**Suppliers**

The HVF series of peptides was obtained as crude mixtures from the Nuro Corp. and the EPO series from Multiple Peptide Synthesis. The other reagents were from the following sources: crude bovine colostrum, Pel-Freeze; UDP-[14C]GalNAc (55 mCi/mmol, 20 µCi/ml), Du Pont-New England Nuclear; UDP-GalNAc, Sigma; Dowex 1X-8, Affi-Gel Blue and Chelex-100, Bio-Rad; DEAE-Sephadex, SP-Sephadex, and Sephacyr 5-100, Pharmacia LKB Biotechnology Inc.

**RESULTS**

**Substrate Purification—Synthetic peptides were designed to study the effects of changes in the amino acid sequence surrounding a threonine residue on its ability to be glycosylated. The sequence of the substrate ("PHMAQVTVG-FGL") was changed by the substitution at each residue in turn with isoleucine, alanine, proline, glutamic acid, and arginine (Table I). In order to facilitate the synthesis of many peptides, those which were altered at the same position were generally made as a mixture. The mixtures of crude peptides were separated into individual peptides by reverse phase chromatography under the same conditions as (panel I).**

**Fig. 1.** 9 mg of Mixture 7 elution on a Vydak (2.2 x 25-cm) C18 column. The gradient was 10-40% Buffer B in 5 min and then 40-60% Buffer B in 30 min (panel I). It shows peak +1E(AAP) elution on a Waters µBondapak (0.39 x 30-cm) C18 column, 10-40% Buffer B in 5 min and then 0-25% Buffer C in 20 min. From panel II peak +1P (panel III), peak +1A (panel IV), and +1P (panel V) were rechromatographed under the same conditions as (panel II).
HPLC (Fig. 1). The composition and quantity of each peptide was verified by amino acid analysis. The peptides that contained substitutions of arginine and isoleucine were most easily purified from the mixture, whereas those containing alanine, proline, and glutamic acid could only be separated by recycling the mixture through analytical columns under near-isocratic conditions.

The substrates used to compare the glycosylation of serine to threonine in this study were based on part of the sequence of human erythropoietin (EPO) containing a glycosylated serine residue "PPDAASAALPR+" (EPO-S) and its threonine-containing analogue "PPDAATAAPLR+" (EPO-T). To investigate the interaction, if any, between two potential acceptor sites, we also synthesized EPO-SS ("PPDAASS-AAPLR+"), EPO-TT ("PPDAATTAAPLR+"), EPO-ST ("PPDAASTAAPLR+"), and EPO-TS ("PPDAATSAAPLR+"), each substrate in this series was synthesized and purified separately.

**UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase Purification from Bovine Colostrum**—A summary of the enzyme purification is given in Table III. Fractions were assayed for protein by absorbance at 280 nm and for transferase activity using both threonine- and serine-containing substrates. The amount of UDP-GalNAc incorporated into EPO-S was consistently about 5% of the amount incorporated into its analogue, EPO-T, for a given enzyme fraction. Steps 1–4 were employed to remove solid material, lipid, and some contaminating proteins (Fig. 2). The adsorption of the transferase to Affi-Gel Blue was used to concentrate the activity.

The enzyme was passed through columns of Affi-Gel Blue three times and eluted with successively shallower gradients of NaCl (Figs. 3–5). The repeated passage of the transferase through Affi-Gel Blue resin resulted in separation of the enzyme into two peaks of activity, A and B. Following the Affi-Gel Blue chromatography, the transferase activity was purified 30–60-fold, compared with crude colostrum. An affinity chromatography, the transferase activity was purified 479- and 69-fold, respectively (Fig. 6). Following dialysis to eliminate UDP-GalNAc from the elution buffer, the final two purified pools of transferase, 1A1 and 2B1, were used to assay the peptide substrates.

The apparent $K_m$ of UDP-GalNAc was determined for each of the purified enzyme pools. Using the tetrapeptide Ac-TPPP at saturating conditions, the $K_m$ (app) of UDP-GalNAc for transference pool 1A1 was 4.4 μM and for pool 2B1 was 7.6 μM (Fig. 7). Apparent $K_m$ values were obtained for the peptides HVF (1.8 μM) and -6E (3.1 μM) in the presence of 36.5 μM UDP-GalNAc. Under the same conditions, peptide +3A was not glycosylated (Fig. 8).

**In Vitro Glycosylation of Peptide Substrates**—The HVF peptide and its derivatives (Table I) were used as substrates for in vitro glycosylation reactions with the purified transferase. Peptide -2E could not be recovered in sufficient amounts to be assayed, giving a total of 52 substrates. We found that reactions performed in the presence of sodium cacodylate proceeded more rapidly than in buffers containing either Tris, MES, or imidazole (Table IV). The peptides were assayed three times, with essentially the same results. There was no apparent difference between the two purified enzyme fractions, 1A1 and 2B1, in their ability to glycosylate any of the peptides. The result of substrate assays using the transferase fraction 1A1 is shown in Table I and Fig. 9. The HVF peptide was glycosylated at the rate of 0.1 μmol/min/mg under standard assay conditions. Although the other substrates differed from the HVF peptide by only one amino acid residue, they varied considerably in their ability to be glycosylated. Total GalNAc incorporation for 19 substrates was at least 50% that of the parent peptide. Thirteen peptides were poor substrates for transferase, having incorporated 10% or less [14C]GalNAc than the HVF peptide.

Amino acid substitutions at certain positions of the substrate appeared to have more influence on glycosylation than other positions (Fig. 9). For example, the transferase seemed to be much less tolerant to changes at positions -2 and -3 than at -4, -5, or -6 of the peptide. The importance of the position of a residue (relative to the threonine) is demonstrated by proline, valine, and glycine, since they occur at two positions in the substrate. Any substitution of the proline residue at position +3 resulted in almost no glycosylation of the substrate, whereas changing the proline at position -6 had little effect. The 2 valine residues in the parent peptide are at positions -1 and +1. Glutamic acid and arginine almost eliminate glycosylation of the substrate when they are substituted for valine at position -1, although they have no such effect at position +1. In contrast, glycosylation of the peptide is influenced equally by -1 and +1 substitutions when valine is replaced by isoleucine, alanine, or proline.

No amino acid had the same effect on the substrate at every position (Fig. 9). Glutamic acid substitution, for example, increased glycosylation when it was present at five positions but almost abolished glycosylation at four positions. The

### Table III

| Step | Volume | Total protein | Total activity | Yield | Specific activity | Step purification | Overall purification |
|------|--------|---------------|----------------|-------|------------------|-------------------|---------------------|
|      | ml     | mg            | units*         | %     | units/mg         | -fold            |                     |
| 1.   | 80,000 × g supernatant | 130 | 3,447 | 110 | 41 | 0.032 | 2 | 0.7 |
| 2.   | DEAE-Sepharose | 351 | 2,710 | 179 | 66 | 0.066 | 2 | 1.4 |
| 3.   | SP Sephadex | 134 | 1,574 | 165 | 61 | 0.105 | 1.6 | 2.2 |
| 4.   | Sephacryl S-300 | 210 | 395 | 129 | 48 | 0.396 | 3 | 6.8 |
| 5.   | Affi-Gel Blue I | 750 | 88.2 | 78.8 | 29 | 0.893 | 2.7 | 15.6 |
| 6.   | Affi-Gel Blue II-pool 1 | 32 | 23.5 | 25.4 | 9.4 | 1.076 | 1.2 | 22.4 |
| 7.   | Affi-Gel Blue III-pool 1A | 7 | 12.5 | 19.5 | 7.2 | 1.555 | 1.4 | 32.4 |
| 8.   | Hg-UDP-GalNAc-pool 1A | 2.3 | 0.015 | 11.3 | 4.1 | 744.4 | 479 | 15,508 |
| 9.   | Affi-Gel Blue II-pool 2 | 80 | 30.6 | 54.4 | 20.1 | 1.775 | 2 | 37 |
| 10.  | Affi-Gel Blue III-pool 2B | 7.5 | 3.3 | 9.6 | 3.6 | 2.882 | 1.6 | 60 |
| 11.  | Hg-UDP-GalNAc-pool 2B1 | 2.3 | 0.019 | 3.8 | 1.4 | 199.2 | 69 | 4,154 |

* One unit is defined as 1 nmol of UDP-[14C]GalNAc transferred to 50 nmol of Ac-TPPP/min under standard assay conditions.
Influence of Sequence on O-Glycosylation

FIG. 2. Gel filtration of SP-Sephadex flow-through on a column (5 x 120 cm) of Sephacryl S-300 resin (Step 4). Fractions were monitored for absorbance at 280 nm and assayed for transferase activity (●) as described under “Experimental Procedures.” The pooled fractions are indicated by the solid bar.

FIG. 3. Elution profile of UDP-GalNAc polypeptide:N-acetylgalactosaminyltransferase on a column of Affi-Gel Blue (Step 5) using Buffer E. Fractions were monitored for absorbance at 280 nm and assayed for transferase activity against EPO-T (●) and EPO-S (▲), as described under “Experimental Procedures.” The solid bar indicates the pooled activity. No transferase activity was found in the material eluted with Buffers B, C, or D.

FIG. 4. Elution of UDP-GalNAc polypeptide:N-acetylgalactosaminyltransferase on a column of Affi-Gel Blue (Step 6) using a linear gradient of NaCl, as described under “Experimental Procedures.” Fractions were monitored for absorbance at 280 nm and assayed for transferase activity against EPO-T (●) and EPO-S (▲). The enzyme activity was collected as two pools, as shown by the solid bars. No transferase activity was eluted by Buffers B and C or Buffer C with cytidine 5′-monophosphate.

FIG. 5. Step 7, chromatography of Pool 1 (A) and Pool 2 (B) from Step 6 on Affi-Gel Blue resin (see “Experimental Procedures”). The transferase was eluted with a shallower NaCl gradient than in Step 6, and each of the major peaks was collected separately, as indicated by the solid bars. Fractions were monitored for absorbance at 280 nm and assayed for transferase activity against EPO-T (●) and EPO-S (▲).

FIG. 6. Purification of UDP-GalNAc polypeptide:N-acetylgalactosaminyltransferase from Step 7 pools 1A and 2B, on columns of 5-mercuri-UDP-GalNAc thiopropyl-Sepharose resin (Step 8), as described under “Experimental Procedures.” No transferase activity was found in the material eluted with Buffers L, M, or N. The transferase was eluted with Buffer O, and fractions were assayed using EPO-T for pool 1A (●) and 2B (▲). The major peak of transferase activity was collected from each column and designated 1AI and 2BI. Neither of the purified pools was able to glycosylate the serine-containing peptide EPO-S.

effects of arginine substitutions showed a similar distribution to those of glutamic acid, although the enhancement of glycosylation was less marked. The alanine-substituted peptides were most like those containing arginine, except that alanine at position −1 still yielded a functional substrate but at −5 did not. Proline appeared to have the most inductive effect
Influence of Sequence on O-Glycosylation

**FIG. 7.** Apparent $K_m$ determination for UDP-GalNAc using the purified enzyme pools 1AI (○) and 2BI (○). Reaction conditions are described under "Experimental Procedures."

**TABLE IV**

| Buffer  | HVF (cpm) | EPO-T (cpm) |
|---------|-----------|-------------|
| Cacodylate | 5,039  | 24,595  |
| Imidazole  | 3,812  | 15,151  |
| MES       | 3,949   | 16,932   |
| Tris      | 2,678   | 10,329   |

*Details of the buffer composition and reaction conditions are under "Experimental Procedures."*

**FIG. 8.** Apparent $K_m$ determinations for the peptides -6E (●), HVF (●), and +3A (○) using the transferase pool 1AI. The $K_m$ for peptide +3A could not be calculated because the GalNAc incorporation was too low.

**DISCUSSION**

Ideally, studies of flanking sequence requirements for O-glycosylation would include all possible combinations of residues surrounding a hydroxylamino acid. Recently, novel techniques have been developed to synthesize and screen completely degenerate mixtures of peptides to identify sequences that bind the heat shock protein BiP (Flynn et al., 1991). Pilot experiments with this approach indicated that the presence of nonglycosylating peptides could effectively inhibit the glycosylation of normally good substrates when they were incubated together (data not shown). A similar approach, but using the peptides still attached to resin, was also found to be unsuitable; that is, substrates that were glycosylated well in...
solution could not be glycosylated on a solid support (Lam
et al., 1991). Recombinant DNA techniques present an attrac-
tive way to synthesize a single degenerate mixture of sub-
strates, but the effort involved in cloning, expressing, and then
purifying a large number of individual substrates would be con-
siderable. Synthetic peptides have been successfully used as
substrates for GalNAc transferase for many years. Peptides
can be made predictably in sufficient amounts and can be
readily purified. The main drawback of producing large num-
bers of synthetic peptides is the cost involved; previous in
vitro glycosylation studies have typically featured fewer than
a dozen peptides (Young et al., 1979; Briand et al., 1981;
Hughes et al., 1988; Cottrell et al., 1992). Our experimental
approach was to study a single 12-amino acid substrate that
is based on the sequence surrounding a naturally occurring
glycosylated site and is typical of many known glycosylation
sites. The strategy for the substrate synthesis was to make
mixtures of the peptides having substitutions at the same
positions. For example, the five peptides that had isoleucine,
alanine, proline, glutamic acid, and arginine at position −5
were synthesized as a mixture. Each peptide was individually
separated from the mixtures by reverse phase HPLC. In this
fashion, we were able to obtain substitutions at every position
of a peptide, except threonine, while minimizing the actual
number of syntheses.

The UDP-GalNAc:polypeptide N-acetylgalactosaminy-
transferase from bovine colostrum was purified 15,000-fold to
assay the peptide substrates. We purified the two peaks of
enzyme activity eluted from the Affi-Gel Blue columns sepa-
ately in the later steps, since we could not exclude the
possibility that they represented two different transferases.
However, the two enzyme pools described here, 1A1 and 2B1,
showed similar substrate specificities with respect to the
peptides in this study. The disparity in the $K_{\text{m(\text{pep})}}$ for UDP-
GalNAc between the two pools (4.4 and 7.6 μM) may be due
to the difference in the purity of the pools (15,508-fold and
4151-fold). It is also possible that one pool represents a
modification, such as a cleavage product, of the other. The
purification of this enzyme has been reported before, although
the extent of its substrate specificity was not described (Sug-
iiura et al., 1982; Elhammer and Kornfeld, 1986). Previous
purification of this enzyme relied on affinity chromatography
with apomucin-Sepharose. Given the difficulties in isolating
and deglycosylating homogeneous salivary gland mucin for an
affinity resin, we chose to use the nucleotide sugar as an
affinity ligand. The transferase was first purified 30–60-fold
by standard chromatographic methods, including three passes
through Affi-Gel Blue columns. The major purification
was achieved by affinity chromatography on 5-mercuri-UDP-
GalNAc thiopropyl-Sepharose; this step alone enhanced the
purification almost 500-fold. While this manuscript was in
preparation, another report was published that detailed the use
of 5-mercuri-UDP-GalNAc thiopropyl-Sepharose to pu-
lify a transferase from porcine submandibular gland (Wang
et al., 1992).

Glycosylation assays of the HVF-based peptide and its
derivatives confirm that single amino acid changes in the
substrate can dramatically alter its capacity to be O-glycosy-
lated (Fig. 9). At position +3, the substitution of isoleucine,
alanine, glutamic acid, or arginine almost precludes glyco-
sylation of the peptide. These results suggest that for this sub-
strate, the proline at +3 is essential for glycosylation. When
a second proline residue is placed at −1, +1, +2, or +4, the
level of glycosylation of the peptide is markedly increased.
Given their frequency near glycosylated residues, it seems
that proline residues are particularly effective at facilitat-
glycosylation, but since the enhancement is not seen at every
position, there may be limitations on the range of proline’s
influence (Gooley et al., 1991). Since most glycosylated sites
do not have proline at position +3, other sequences are able
to provide the necessary context for glycosylation to occur.
Position −1 was also sensitive to amino acid substitutions
with the charged residues causing a substantial reduction in
glycosylation. It is noteworthy that the charged residues do
not inhibit glycosylation when present at many other posi-
tions of this peptide. The peptides that were substituted at
position −3 were all poor substrates for the transferase, which
is consistent with the prediction that this position is impor-
tant for O-glycosylation, although there was no apparent
selectivity among the different types of amino acids tested.
Experimentally, amino acid substitutions at the −2 position
of this peptide all substantially reduced glycosylation, an
effect that was not predicted from the databases of known
glycosylated sites. Nonetheless, most changes at other points
in the flanking sequence (−6, −5, −4, +1, +2, +4, and +5)
were generally less able to influence glycosylation of the
threonine residue, indicating that transferase does not have
strict requirements for all positions of the flanking sequence.
From sequence analysis, it was expected that substitutions at
the −6 position would not be tolerated, but perhaps its loca-
tion at the N terminus of the peptide prevented it from
exerting its usual effect. In general, the same types of amino
acids were found to have the most influence on substrate
glycosylation by statistical and experimental methods (proline
and charged residues) (O’Connell et al., 1991). Alanine sub-
stitutions caused somewhat less pronounced effects, whereas
isoleucine produced the least change overall.

Surveys of the sequences surrounding glycosylated sites have
revealed no difference in patterns related to serine and
threonine (Wilson et al., 1991; O’Connell et al., 1991).
However, some reports have stated that peptides with serine
residues were not glycosylated in vitro (Hughes et al., 1988;
O’Connell et al., 1991). Recently, it was shown that three
serine-containing peptides, whose sequences were similar to
porcine submandibular gland apomucin, could not be glyco-
sylated by a transferase purified from porcine submandibular
gland or by homogenates of porcine, bovine, and ovine sub-
mandibular glands (Wang et al., 1992). In our experiments,
the transferase derived from bovine colostrum did not glyco-
sylate a peptide based on the sequence about a glycosylated
serine residue in human erythropoietin (Table II). Moreover,
the level of serine-specific glycosylation was very low in the
crude colostrum and throughout the enzyme purification
(Figs. 3–6). The EPO-based peptide was not glycosylated
when 2 serine residues were placed in tandem nor was an
adjacent threonine residue able to induce glycosylation of a
serine. Hence, the evidence suggests that the activity which
glycosylates serine residues is either unstable, inhibited, or
separate from the threonine-specific activity. The observation
that a threonine analogue of the erythropoietin-based peptide
was readily glycosylated indicates that the flanking sequence
requirements for the hydroxyamino acids may be the same.

It is unclear whether all of the findings presented here are
due specifically to amino acid sequence or to changes in the
substrate conformation. It has been proposed that the main
determinants for O-glycosylation are the accessibility and
local conformation of the acceptor site (Hagopian et al., 1979;
Aubert et al., 1976; Hill et al., 1977; Eckhardt et al., 1987).
Circular dichroism spectroscopy of peptides indicates that a
random-type secondary structure is necessary for the sub-
strate to be O-glycosylated (O’Connell et al., 1991). It is
possible that the transferase has both structural and sequence
concerns.
requirements superimposed on the need to be physically accessible to the transferases. The precise topography of potential glycosylation sites within proteins has yet to be determined.

It is important to note that the great majority of known O-glycosylated residues occur in clusters. The order, if any, in which clustered residues are glycosylated is unknown, so it is difficult to determine if modification of particular sites must precede the modification of other sites. The EPO-based peptide containing 2 threonine residues was glycosylated much more than the same substrate with only 1 threonine (Table II). The increased incorporation of GalNAc took place almost exclusively at the C-terminal threonine (Fig. 10), which is consistent with recently published data (Wang et al., 1992). However, in naturally occurring glycoproteins there is at least one example where only the N-terminal threonine of a pair is O-glycosylated and many cases where both threonines of a pair are acceptors (Murayama et al., 1982; Tomita and Marchesi, 1975; Putnam et al., 1981). Glycosylated residues are customarily identified by a blank cycle in the amino acid sequence of a protein, which does not accurately determine the extent to which a particular residue is glycosylated. It is possible that glycosylation is not initiated to the same extent at all the nominally modified sites of a protein (Gooley et al., 1991). Incomplete initiation of glycosylation could account for some of the heterogeneity that glycoproteins typically display. It is conceivable that more than one threonine-specific transferase exists having separate or overlapping substrate specificities (Hagopian et al., 1971). The presence of several transferases could explain why deglycosylated proteins are poorly glycosylated by a purified enzyme in vitro and why no consensus sequence has emerged (Sugiura et al., 1982; Hill et al., 1977). Examination of more complex multisite substrates will help to fully define the range of hydroxyamino acids that can be glycosylated by GalNAc transferases. These and other issues must be addressed before O-linked glycoproteins can be synthetized with fidelity either in vitro or in heterologous systems.

It would be difficult to make broad conclusions about flanking sequence requirements for O-glycosylation from any study of peptides in vitro. A logical progression of this work is to examine the effects on glycosylation of systematic mutagenesis of substrates in vivo. Future studies should compare the tolerance for amino acid changes in the sequence surrounding glycosylation sites and should include multiple substitutions. The data presented here suggest that particular combinations of amino acids at sensitive positions are more likely to promote O-glycosylation of a threonine residue.

Acknowledgments—We thank Brian VanWuyckhuys for the amino acid analysis and Pat Noonan for preparing the manuscript.

REFERENCES

Aubert, J.-P., Biserte, G., and Louchet-Lefebvre, M.-H. (1976) Arch. Biochem. Biophys. 175, 410-418

Bendaš, B., and Schachter, H. (1987) J. Biol. Chem. 262, 5775-5783

Braend, J. P., Andrews, S. P., Jr., Cahill, E., Conway, N. A., and Young, J. D. (1981) J. Biol. Chem. 256, 12205-12207

Cotterell, J. M., Hall, R. L., Sturton, R. G., and Kent, P. W. (1992) Biochem. J. 283, 399-399

Eckhardt, A. E., Timpte, C. S., Abernethy, J. L., Tonnadze, A., Johnson, W. C., Jr., and Hill, R. L. (1987) J. Biol. Chem. 262, 11338-11344

Elhammer, A., and Kornfeld, S. (1986) J. Biol. Chem. 261, 5240-5255

Flynn, G. C., Pohl, J., Flocco, M. T., and Rothman, J. E. (1991) Nature 353, 726-729

Gerken, T. A., Butenhoer, K. J., and Shogren, R. (1989) Biochemistry 28, 5536-5543

Gooley, A. A., Classon, B. J., Marschalek, R., and Williams, K. L. (1991) Biochem. Biophys. Res. Commun. 176, 1194-1201

Hagopian, A., and Eyler, E. H. (1989) Arch. Biochem. Biophys. 129, 515-524

Hagopian, A., Westall, F. C., Whitehead, J. S., and Eyler, E. H. (1991) J. Biol. Chem. 266, 2513-2523

Hill, H. D., Jr., Schwyzer, M., Steinman, H. M., and Hill, R. L. (1977) J. Biol. Chem. 252, 3799-3804

Honma, K., Tomita, M., and Hamada, A. (1990) J. Biochem. (Tokyo) 88, 1679-1689

Hughes, R. C., Bradbury, A. F., and Smyth, D. G. (1988) Carbohydr. Res. 178, 255-269

Lai, P.-H., Everett, R., Wang, F.-F., Arakawa, T., and Goldwasser, E. (1986) J. Biol. Chem. 261, 3116-3121

Lam, K. S., Salzon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) Nature 354, 82-86

Murayama, J.-I., Tomita, M., and Hamada, A. (1982) J. Membr. Biol. 64, 205-225

O’Connell, B., Tahak, L. A., and Ramsayubua, N. (1991) Biochem. Biophys. Res. Commun. 180, 1034-1030

Oppenheim, F. G., Offer, G. D., and Troxler, R. F. (1985) J. Biol. Chem. 260, 10671-10679

Pant, J. M., Takahashi, N., Tetzlaff, D., Debureih, B., and Lin, L.-C. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 6168-6172

Rall, S. C., Jr., Weisgraber, K. H., and Mahley, R. W. (1982) J. Biol. Chem. 257, 4174-4178

Shogren, R., Gerken, T. A., and Jenntoft, N. (1989) Biochemistry 28, 5536-5543

Smyth, D. S., and Utsumi, S. (1992). Biochemistry 31, 3171-3184

Tomita, M., and Marchesi, V. T. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 2964-2968

Wang, Y., Abernethy, J. L., Eckhardt, A. E., and Hill, R. L. (1992) J. Biol. Chem. 267, 12709-12715

Wilson, I. B. H., Gevel, Y., and von Heijne, G. (1991) Biochem. J. 275, 529-534

Young, J. D., Tsuchiya, D., Sandlin, D. E., and Holroyde, M. J. (1979) Biochemistry 19, 4444-4448