Purification and cDNA Cloning of a Human UDP-N-acetyl-α-D-galactosamine:polypeptide N-Acetylgalactosaminytransferase*

(Received for publication, May 31, 1995)

Thayer White‡§, Eric Paul Bennett‡, Koji Takio, Tina Sørensen†, Nina Bondingt, and Henrik Clausent**

From the Faculty of Health Sciences, School of Dentistry, University of Copenhagen, DK-2200 Copenhagen N, Denmark, the §Biomembrane Institute, Seattle, Washington 98119, and the *Institute of Physical and Chemical Research (RIKEN), Wako, Saitama, 351-01, Japan

A UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (GalNAc-transferase) from human placenta was purified to apparent homogeneity using a synthetic acceptor peptide as affinity ligand. The purified GalNAc-transferase migrated as a single band with an approximate molecular weight of 52,000 by reducing sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Based on a partial amino acid sequence, the cDNA encoding the transferase was cloned and sequenced from a cDNA library of a human cancer cell line. The cDNA sequence has a 571-amino acid coding region indicating a protein of 64.7 kDa with a type II domain structure. The deduced protein sequence showed significant similarity to a recently cloned bovine polypeptide GalNAc-transferase (Homa, F.L., Hollanders, T., Lehman, D.J., Thomsen, D.R., and Elhammer, Å.P. (1993) J. Biol. Chem. 268, 12609–12616). A polymerase chain reaction construct was expressed in insect cells using a baculovirus vector. Northern analysis of eight human tissues differed clearly from that of the bovine GalNAc-transferase. Polymerase chain reaction cloning and sequencing of the human version of the bovine transferase are presented, and 98% similarity at the amino acid level was found. The data suggest that the purified human GalNAc-transferase is a novel member of a family of polypeptide GalNAc-transferases, and a nomenclature GalNAc-T1 and GalNAc-T2 is introduced to distinguish the members.

Mucin-type O-linked glycosylation is one of the dominant forms of glycosylation of glycoproteins. Mucin-type glycosylation is initiated by the addition of the monosaccharide N-acetylgalactosamine to the hydroxyl group of serine and threonine amino acids (GalNAcα1-O-Ser/Thr). GalNAc O-glycosylation is found on a variety of glycoproteins but is more prominent on high molecular weight secretory glycoproteins such as mucins, where they may constitute up to 80% of the total mass. O-Linked glycosylation also appears to have a role in the conformation and protease resistance of “stilk regions” of membrane proteins necessary for correct exposure and accessibility of a functional domain of membrane-bound protein (Jentoft, 1990; Sadler, 1984; Lis and Sharon, 1993; Varki, 1993). Recently, O-glycans have been selectively implicated in carbohydrate-lectin cell adhesion phenomena (Springer, 1994).

A polypeptide GalNAc-transferase1 initiating mucin-type glycosylation was recently purified to apparent homogeneity from bovine colostrum and porcine submaxillary glands (Elhammer and Kornfeld, 1986; Wang et al., 1992; O’Connell et al., 1991). The purification strategy relied on affinity chromatography using either ovine submaxillary amucin (acceptor substrate) or 5-mercury-UDP-GalNAc (donor substrate). Using amucin affinity chromatography, Homa et al. (1993) succeeded in obtaining amino-terminal sequence information of a bovine colostrum GalNAc-transferase sufficient for cDNA cloning from a bovine small intestine library. More recently, O’Connell et al. (1991) isolated the same GalNAc-transferase from bovine colostrum by 5-mercury-UDP-GalNAc affinity chromatography (Hagen et al., 1993).

One pertinent question has been whether one or several GalNAc-transferases are involved in mucin-type O-linked glycosylation initiation. The main evidence supporting multiple transferases has been the lack of activity toward serine substrates in purified transferase preparations (Wang et al., 1992; O’Connell et al., 1991). Elhammer et al. (1993) have, however, recently been able to identify serine transferase activity albeit apparently 35-fold lower than threonine transferase activity in purified bovine colostrum transferase using the peptide sequence PPASSAPG. Furthermore, Wang et al. (1993) showed both activities in a purified porcine GalNAc-transferase preparation. Further evidence is the identification of an apparently unique GalNAc-transferase activity associated with fetal and tumor tissue responsible for the synthesis of the oncofetal fibronectin epitope defined by monoclonal antibodies FDC6 and 5C10 (Matsuura et al., 1988, 1989). We have not been able to demonstrate GalNAc-transferase activity toward the reported fibronectin-derived acceptor peptide VTHPGY (Mandel et al., 1994). There is thus no clear evidence suggestive of more than one transferase to date.

Here we report the purification of a novel human placenta GalNAc-transferase using a defined synthetic acceptor peptide as affinity ligand. Based on partial amino acid sequences cDNA

*This work was supported by the Danish Medical Research Council, the Lundbeck Foundation, and Ingeborg Roikjer’s Foundation. The first two authors contributed equally to this work.

**To whom correspondence should be addressed: Dept. of Oral Diagnostics, School of Dentistry, Nørre Alle 20, DK-2200 Copenhagen N, Denmark. Tel.: 45-3532-6835; Fax: 45-3532-6505.
was doned and sequenced. The human GalNAc-transferase doned in the present paper possesses domains homologous with the bovine GalNAc-transferase and its 98% similar hu-
man counterpart. As will be described in detail in the accom-
panying paper, affinity chromatography with a defined syn-
thetic peptide in fact separated two distinct transferase
activities in agreement with the potential existence of multiple
GalNAc-transferases (Sørensen et al., 1995). 

EXPERIMENTAL PROCEDURES

Materials

Human placenta were collected (6–24 h postdelivery) after informed
consent at the Herlev Hospital, Copenhagen, Denmark. UDP-GalNAc
was purchased from Sigma, and UDP-[3H]GalNAc (Amersham, United
Kingdom) was mixed to final specific activities as indicated. Peptides
were synthesized on an Applied Biosystem synthesizer (model 380B) at
the State Serum Institute (Copenhagen, Denmark). Polymerase chain
reaction (PCR) was performed on a Perkin-Elmer thermocycler model
384B (Bio-Rad protein assay) using bovine serum albumin as standard.

Restriction enzymes were purchased from Promega. All chemicals
were synthesized on an Applied Biosystem synthesizer (model 380B) at
the State Serum Institute (Copenhagen, Denmark). Polymerase chain
reaction (PCR) was performed on a Perkin-Elmer thermocycler model
480 using Taq polymerase (Cetus). Reverse transcription PCR (RT-
PCR) was performed using a RT-PCR kit (Perkin-Elmer), and all se-
quencing reactions used the Ampli

was cloned and sequenced. The human GalNAc-transferase
activity was determined in reaction mix-
tures containing 10 mM Tris buffer (pH 7.4), 0.05 mM UDP-[3H]GalNAc (4,000 cpm/
nmol) in a total volume of 100 μl including enzyme. Product was
sequenced by scintillation counting after Dowex 1 formic
acid cycle chromatography, but initially all transferase preparations
were synthesized on an Applied Biosystem synthesizer (model 380B) at
the State Serum Institute (Copenhagen, Denmark). Polymerase chain
reaction (PCR) was performed on a Perkin-Elmer thermocycler model
480 using Taq polymerase (Cetus). Reverse transcription PCR (RT-
PCR) was performed using a RT-PCR kit (Perkin-Elmer), and all se-
quencing reactions used the Ampli-Taq sequencing kit (Perkin-Elmer).

Protein Determination

Protein concentrations were determined by the method of Bradford
(Bio-Rad protein assay) using bovine serum albumin as standard.

Purification of Human Polypeptide GalNAc-transferase

All steps of purification were performed at 4°C. The buffer com-
positions were as follows. Buffer A was 25 mM BisTris (pH 6.5), 0.25% Triton X-100, 100 mM NaCl. Buffer B was 25 mM BisTris (pH 7.4), 0.25% Triton X-100, 100 mM NaCl. Buffer C was 25 mM BisTris (pH 6.5), 0.25% Triton X-100, 7 mM 0.7 mM UDP. Buffer D was 25 mM BisTris (pH 6.5), 0.25% Triton X-100, 100 mM NaCl. Buffer E was 25 mM BisTris (pH 6.5), 0.25% Triton X-100, 100 mM NaCl, 5 mM 2-mercaptoethanol, 5 mM
MgCl2, 50 mM KCl, 125 mM MnCl2, 5 mM CDP-
choline, 0.25% Triton X-100, 0.05 mM UDP-[3H]GalNAc (4,000 cpm/mmole) in a total volume of 100 μl including enzyme. Product was
sequenced by scintillation counting after Dowex 1 formic
acid cycle chromatography, but initially all transferase preparations
were synthesized on an Applied Biosystem synthesizer (model 380B) at
the State Serum Institute (Copenhagen, Denmark). Polymerase chain
reaction (PCR) was performed on a Perkin-Elmer thermocycler model
480 using Taq polymerase (Cetus). Reverse transcription PCR (RT-
PCR) was performed using a RT-PCR kit (Perkin-Elmer), and all se-
quencing reactions used the Ampli-Taq sequencing kit (Perkin-Elmer).

Restriction enzymes were purchased from Promega. All chemicals were
of high purity and purchased from Sigma. 

Polyamide Gel Electrophoresis

Peptide sequencing either directly or after SDS-PAGE blotting. Peak
fractions (5–10 μl) from step 5 were applied directly to an S-12 PC
3.2/30 Superose column (SMART System), equilibrated, and run in
buffer A without detergent. Fractions (50 μl) were collected and assayed
for activity as well as assayed by SDS-PAGE (10 μl/fraction after
aceton precipitation).

Protein Sequence Analysis

Protein was hydrolyzed at 110°C for 24 h or 74 h in 6 n HCl under vacuum
and applied to an amino acid analyzer (Hitachi L-8500). The
amino-terminal sequence of the protein was obtained directly from the
Mono S purified preparation (step 5) and from Coomassie Blue-stained
protein from this same step on a PVDF blot. Approximately 10 μg of
GalNAc-transferase was reduced and pyridylethylated (Ru¨egg and
Masaki, 1991) and purified further by gel permeation chromatography on
two SynChropak GPC G300 columns (4.6 × 250 mm each) in 10 mM
sodium phosphate (pH 6.0) containing 6 mM guanidine hydrochloride. Reduced and pyridylethylated protein was
digested with Aspergillus oryzae protease 1, a lysyl bond-specific
endopeptidase (Aebbersold et al., 1987) in 50 mM Tris-HCl (pH 9.0)
containing 15 μl urea at 37°C for 15 h. Alternatively, protein blotted on
nitrocellulose was digested (Masaki et al., 1981) with A. lyticus protease 1.
Peptides generated were separated by reverse phase HPLC on an
RP300 column (2.1 × 100 mm, Perkin-Elmer) in 0.09% trifluoroacetic
acid at a flow rate of 200 μl/min with a linear gradient of 0–60% acetonitrile in 20 min. 

Preparation of PCR Probe

Based on the combined amino-terminal sequence (see Table II) the fol-
lowing oligonucleotide primers were synthesized (I indic-
es inosine).

EBHC23: 5′-AACGG(C/A)TΓG(A/G)(A/G)(A/G)(A/G)(C/T)CΓA3′
EBHC24: 5′-AA(G/A)AA(G/A)AA(G/A)AA(G/A)CT/GITT ICCA(T/C)ICCA3′
EBHC25: 5′-CG(G/A)GΓG(T/A)ΓG ΓG ΓG ΓΓΓΓTΓC(G/T)ΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓΓGamma

Total RNA from the gastric tumor cell line MKN45 was extracted by
standard procedures (Chomczynski and Sacchi, 1987), and polyadeny-
lated RNA was purified using a Promega poly(A)tract mRNA purification
kit. cDNA was synthesized using EBHC25 and Moloney murine
leukemia virus-reverse transcriptase. PCR was performed on 5 μl of the
reverse transcriptase reaction in a 25 μl solution containing 0.5 μM
EBHC25 and EBHC22, 2 μM MgCl2, 50 mM KCl, 125 mM dNTP, and 2.5
units of Taq polymerase (Cetus) and subjected to the following therm-
cycle conditions: 95°C, 45 s; 43°C, 5 s; 72°C, 15 s, for 40 cycles.

Generated products were digested with EcoRI and ligated into the
EcoRI site of pT73U19 (Pharmacia). Competent Escherichia coli cells,
TABLE I
Purification of GalNAc-transferase from human placenta

| Human placenta             | Volume | Protein | Activity | Yield | Specific activity | Purification |
|----------------------------|--------|---------|----------|-------|------------------|--------------|
| Triton X-100 extractb      | 3,000  | 26,700  | 13.9     | 100   | 0.00052          | 1            |
| Step 1: Cibacron 3GA       | 1,500  | 3,450   | 5.9      | 42    | 0.0017           | 3.3          |
| Step 2: DEAE               | 2,000  | 2,800   | 5.5      | 40    | 0.002            | 3.8          |
| Step 3: S-Sepharose (NaCl) | 150    | 900     | 2.8      | 20    | 0.003            | 6            |
| Step 4: Muc2 peptide (UDP) | 20     | 0.3     | 0.038    | 0.27  | 0.13             | 242          |
| Step 5: Mono S (NaCl)      | 0.15   | 0.02    | 0.015    | 0.11  | 0.75             | 1,442        |

* One unit of enzyme is defined as the amount of enzyme that will transfer 1 μmol of GalNAc in 1 min using the standard reaction mixture as described under "Experimental Procedures" with 25 μM of GalNAc as acceptor substrate.

Abstract

The purification of GalNAc-transferase from human placenta was solubilized in 1.5% Triton X-100 yielding a specific activity of 0.00052 unit/mg using the Muc2 peptide substrate. Little or no activity was found in the first water extraction, and the specific activity of the crude detergent homogenate was approximately 0.00052–0.00010 unit/mg. The degree of measurement of activity in crude homogenates was quite inconsistent, however, and the activity was highly labile. The purification is thus calculated from the soluble Triton X-100 extract excluding a preceding approximately 5-fold factor.

RESULTS

1. Purification to Homogeneity of a Polypeptide GalNAc-transferase from Human Placenta

Step 1: Table I summarizes the purification of the GalNAc-transferase from 2.5 kg of human placenta tissue. The GalNAc-transferase was solubilized in 1.5% Triton X-100 yielding a specific activity of 0.00052 unit/mg using the Muc2 peptide substrate. Little or no activity was found in the first water extraction, and the specific activity of the crude detergent homogenate was approximately 0.00052–0.00010 unit/mg. The degree of measurement of activity in crude homogenates was quite inconsistent, however, and the activity was highly labile. The purification is thus calculated from the soluble Triton X-100 extract excluding a preceding approximately 5-fold factor.

Step 2: Cibacron Blue Chromatography—This step was found to be sufficient for further affinity chromatography using the peptide acceptor substrate. Although this step only yielded 3.3-fold purification, it was a simple, fast, and highly concentrating step. Approximately 50% of the transferase activity measured with the Muc2 peptide bound to the column. Attempts to show differences in substrate specificity between detergent extract and the pass-through and eluate of the Cibacron column were unsuccessful with the different synthetic acceptor peptide substrates (not shown). Attempts to improve the purification by gradient NaCl elution resulted in a significantly lower yield, leaving one-step elution with 1.5 M KCl optimal.

Step 3: S-Sepharose—This step was employed to reduce the volume necessary for the subsequent affinity chromatography in the presence of UDP, but it was not necessary for binding. In separate experiments the S-Sepharose was also used as a detergent exchange method replacing Triton X-100 with octyl glucoside for running the Muc2 peptide affinity column in octyl glucoside rather than Triton X-100. Detergent exchange was necessary to purify the apparent membrane-bound form of the transferase further as described later.

Step 4: Muc2 Peptide Acceptor Substrate Chromatography—The peptide column was run successfully using peptide coupled to TNB-thiol Sepharose (Pierce) to the carboxyl-terminal cysteine residue as well as peptide coupled to cyanogen bromide-activated Sepharose 6B (Pharmacia). The purification reported in Table I utilized cyanogen bromide-coupled peptide run in Triton X-100. The peptide column only bound a minor fraction of the total transferase activity measured by the Muc2 peptide substrate (approximately 5–10%), but re-running the column eluate yielded essentially no further binding.

Further binding to the Muc2 column could be achieved when a detergent exchange from Triton X-100 to n-octyl glucoside was performed. Detergent exchange was performed on S-Sepharose of the pass-through of the first Muc2 chromatog-
raphy (step 4), and this resulted in more than 50% of the residual transferase activity binding another Muc2 peptide column run in n-octyl glucoside detergent. Importantly, however, the enzyme eluted in n-octyl glucoside was not active when the detergent was removed in contrast to the enzyme activity eluted from the Triton X-100 Muc2 chromatography. Thus, n-octyl glucoside-purified enzyme could only be further purified on Mono S chromatography in the presence of detergent. Omission of detergent at this step resulted in a complete loss of enzyme activity in eluted fractions. Furthermore, S-12 gel filtration of the enzyme activity eluted from the Mono S column run in n-octyl glucoside resulted in a spread of activity throughout the eluate, suggesting tight interaction between protein and detergent (not shown). The purification and protein analyses reported were all performed on the enzyme purified by Muc2 chromatography run in Triton X-100 even though this constituted a minor fraction.

Step 5. Mono S Cation Exchange—The eluate of the first affinity chromatography run in Triton X-100 with a specific activity of 0.126 unit/mg was essentially pure, as estimated by SDS-PAGE, but further concentration and purification as well as evaluation of purity were achieved by Mono S chromatography (Fig. 1). Essentially one major band of approximately 52,000 by SDS-PAGE under reducing conditions was obtained, which coincided with the transferase activity. It is important to note that although the affinity chromatography was run in Triton X-100, the Mono S chromatography was run without detergent without loss of transferase activity. Invariable amounts of a small peptide of approximately 10,000 were found to elute at higher ionic strength (see Fig. 1B), and this was identified as lysozyme by amino-terminal amino acid sequencing (not shown). The peak fractions 17 and 18 (50-μl fractions) (Fig. 1) were pooled and subjected to protein sequence analysis. Protein was analyzed either directly from the pooled fractions or after SDS-PAGE blotting to PVDF membrane.

S-12 Gel Filtration Chromatography

Analytical assessment of the purity of the GaINAc-transferase was obtained by S-12 gel filtration chromatography run in buffer without detergent. Ten μl, containing approximately 1 μg of protein with a specific activity of 0.7 unit/mg from the pooled peak fractions of the Mono S chromatography were applied. One major peak of approximately 50,000–55,000 coinciding with the transferase activity was obtained, and SDS-PAGE of the peak fraction also gave the 52-kDa band by SDS-PAGE (Fig. 2). Careful measurement of the A_214nm peak compared with a standard of bovine serum albumin by integration showed a calculated specific activity of 0.76 unit/mg, identical to that of the Mono S fractions. The ability to run both the Mono S as well as the S-12 chromatography without detergent and obtaining almost quantitative yield of transferase activity as well as one distinct peak of protein and activity strongly indicated that the transferase was in a soluble form. The residual transferase activity obtained by detergent exchange to n-octyl glucoside is believed to be the membrane form requiring detergent for solubility.

Properties of the Purified GaINAc-transferase

In the accompanying paper (Sørensen et al., 1995) the kinetic parameters of the soluble transferase with different peptide substrates as well as structural characterization of products are presented. Importantly, the pure transferase was capable of glycosylating the Muc2 peptide containing nine threonine sites as well as the hCG peptide containing exclusively serine sites (Fig. 2).

**Protein Sequence Analysis of the Purified GaINAc-transferase**

Table II summarizes the amino acid sequences of the purified transferase. The amino-terminal region of the GaINAc-transferase was obtained by direct sequencing of the pooled peak of the Mono S chromatography (Fig. 1A). Furthermore, sequencing was performed directly from the PVDF-blotted
positive clones were obtained from 0.75 million plaques primed human cancer cell line cDNA library from which 40 TEB1 fragment was used to screen a random full amino-terminal amino acid sequence of the GalNAc-transferase. The TEB1 fragment was sequenced and found to correspond to the sense and antisense oligonucleotide primers. The PCR-generating overlapping peptide sequences was used for the design of variation caused by varying tissue thawing time before extraction reflects differences in proteolytic degradation induced by batch residues, respectively (Table II). This difference most likely (approximately 40%) starting from SNG differing 10 and 7 (approximately 60%) started from EEK with a minor sequence essentially the same sequence. However, the major sequence 52,000 band derived from Mono S chromatography, yielding at 214nm (and 280 nm, not shown). Fractions (100 µl) of the peak fraction from a Mono S chromatography, and run in phosphate-buffered saline at 40°C were partially sequenced, and all were found to contain 5'-untranslated sequences (approximately 900 and 50 bp, respectively). Clone 5551 contained 2.3 kb including the 5'-end of the coding region but was 100 bp short of the 3' end. The combined nucleotide sequence of the GalNAc-transferase cDNA clones is shown in Fig. 3. Overlapping parts of the two isolated clones were identical in sequence apart from one nucleotide “insertion” in the 3' end of clone 2782 (position 1534, Fig. 3). Since internal peptide sequences apparently were located out of frame past the 3' termination introduced by the insertion in clone 2782, the purified GalNAc-transferase was longer than predicted by this clone. The apparent insertion in clone 2782 was concluded to be an artifact in the library as clone 5551 lacked this insertion, and more importantly, RT-PCR of this area using a variety of total RNA sources including the MKN45 RNA only yielded the sequence identified in clone 5551 (not shown). Furthermore, genomic sequencing of cloned PCR products covering this position only yielded the sequence found in clone 5551. The remaining nine cDNA clones were partially sequenced, and all were found to contain 5' intron sequences and were limited to coding sequences in the 5'-TEB1 probe area. Primary Structure of the Human GalNAc-transferase (GalNAc-T2) The GalNAc-T2 coding sequence has 1713 bp and codes for a protein of 571 amino acids (64,729 Da). Initiation is presumed to occur at the ATG codon at position 1 as an adenosine at position 3 fulfills the requirement for an initiation codon. GalNAc-T2 is predicted to be a type II transmembrane protein with a strongly hydrophobic domain (amino acids 7–24) and not flanked by charged amino acids (Prosis software, Hitachi). As shown in Fig. 3 all of the peptide sequences obtained (Table II) with minor discrepancies at weakly assigned amino acid residues are accounted for in the coding sequence. In
FIG. 3. Nucleotide sequence and predicted amino acid sequence of the two cDNA clones 2782 and 5551. For further details, see “Experimental Procedures.” The amino acid sequence is shown in a single-letter code. The hydrophobic segment representing the putative transmembrane domain is double underlined. Peptide sequences corresponding to the sequences shown in Table II are indicated by underlining. A line missing immediately under individual amino acids indicates discrepancies between the residues predicted from amino acid sequencing (upper case letter; x denotes not identifiable) and those derived from the nucleotide sequence.
agreement with the proposed soluble nature of the isolated transferase protein, the amino-terminal sequence obtained from this protein is found at amino acid position 51 carboxy-terminal to the putative transmembrane-anchoring domain. The predicted amino acid sequence of the human GalNAc-T2 has one consensus sequence (-Asn-Xaa-Ser/Thr-) for N-glycosylation (amino acid 516). N-Glycanase digestion of the soluble transferase did not alter SDS-PAGE mobility (not shown), and as the amino acid sequence of one of the A. lyticus protease I-released peptides included this site and gave positive identification of asparagine, it may be suggested that this site is not utilized or is only partially utilized. The molecular mass of the deduced amino acid sequence of the soluble transferase is 59,205 Da, which is slightly higher than the experimentally determined molecular weight but within the limits of accuracy of the SDS-PAGE and gel filtration systems used.

Expression of GalNAc-transferase (GalNAc-T2)

Expression of the pAGP67-GalNAc-T2-sol construct resulted in an almost 50-fold increase in GalNAc-transferase activity using the Muc2 substrate-peptide in the culture medium of infected cells compared with uninfected controls as well as cells infected with the blood group A glycosyltransferase (Bennett et al., 1995) (Table III). Similarly, a 50-fold increase was observed using the hCG peptide. Background levels in uninfected cell medium were always higher than in control infected cell medium, probably as a result of endogenous Sf9 produced GalNAc-transferase activity being higher due to a larger number of cells.

Distant Homology with Bovine GalNAc-transferase (GalNAc-T1)

Shown in Fig. 4 is an amino acid sequence comparison of GalNAc-T1 and -T2 (DNASIS, Hitachi) (Higgins and Sharp, 1988). The overall similarity is 44%. The amino-terminal sequences of the two enzymes show no similarity, whereas more significant similarity is found in the central area of the putative catalytic domains. The size of the two enzymes is quite similar at the amino acid level. The bovine GalNAc-transferase cDNA indicates a 560-amino acid coding region, and the amino terminus of the purified soluble enzyme was found at amino acid 41, yielding a soluble protein of 520 amino acids. The human GalNAc-transferase described here is 11 amino acids total longer, and the amino terminus of the soluble enzyme was found at amino acid 52 thus yielding a soluble protein of 520 amino acids. The discrepancy in reported molecular weight of the bovine colostrum transferase (approximately 70,000) and the human transferase reported here (approximately 52,000) must be due partly to differences in glycosylation. N-Glycanase treatment of the soluble bovine colostrum transferase reduced the SDS-PAGE molecular weight to approximately 64,000 (predicted molecular mass of 59,358 Da), suggesting that some of the three N-linked consensus glycosylation sites are utilized in contrast to the human placenta transferase which appears not to be N-glycosylated. Fig. 4 also shows that both transferases have an abundance of cysteine residues (GalNAc-T2 has 13; bovine and human GalNAc-T1 has 16) and that 12 of these are aligned (Homology plot, Clustal). Some of these aligned cysteines are positioned in highly similar regions, but others are in areas of more divergent sequences. The apparent conserved number and spacing of cysteines indicate that the GalNAc-transferases are highly disulfide-bonded proteins maintaining a similar tertiary structure. Interestingly, Wang et al. (1992) noted that a reducing agent increased the transferase activity of porcine submaxillary gland GalNAc-transferase, and we observed the same for the human placenta transferase. Drickamer (1993) originally observed a similar conserved pattern of cysteine residues in sialyltransferases with generally limited similarity (Livingston and Paulson, 1993). It therefore seems likely that the two genes (T2 and T1) possess homologous domains originating from a common ancestral gene.

PCR Cloning of the Human Version of GalNAc-T1

To verify that the purified and cloned GalNAc-T2 was distinct from the previously cloned bovine transferase, we have cloned the human equivalent of the bovine enzyme. Human GalNAc-T1 was PCR amplified and sequenced as two overlapping fragments. Since the PCR primers utilized were placed in the coding region ends (positions 5' - 2 to 20; and 3', 1661 to 1681), the 15-20 bp at the initiation and termination codons are unknown. The DNA sequence of the residual major part of the gene was found to be 95% similar to that of the bovine sequence, giving rise to only six amino acid substitutions (99%) (Fig. 4).

Northern Analysis

Northern analysis of the human gastric cancer cell line MKN45 using a 2782 PvuII/HindIII fragment (TEB2) as a probe revealed a discrete band of 4.5 kb as well as multiple lower bands of 2-3 kb in the poly(A)+ mRNA lane (Fig. 5). It is unknown which represents the mature mRNA. No hybridization was observed with total RNA, and this is most likely due to lower representation of target RNA.

Analysis of a human multiple tissue blot from Clontech showed hybridization to a 4.5-kb mRNA in all tissues (Fig. 6). Similar to the MKN45 cell line, several of the tissues also expressed the smaller size 2-3-kb mRNAs. The same Northern blot from Clontech was analyzed by Homa et al. (1993) using the bovine GalNAc-transferase-T1. 4.2-kb mRNA was detected using the GalNAc-T1 probe in all tissues but kidney, and the level of expression differed significantly from that found in Fig. 6.

DISCUSSION

Mucin-type GalNAc-transferase activity has previously been purified to apparent homogeneity from bovine colostrum and porcine submaxillary glands using either apomucin acceptor substrate chromatography or 5-mercury-UDP-GalNAc donor substrate chromatography as the principal affinity purification step (Elhammer and Kornfeld, 1986; Wang et al., 1992; O'Connell et al., 1991). Here we report the purification of a unique GalNAc-transferase from human placenta using an acceptor substrate chromatography that is based on a defined synthetic peptide derived from the human intestinal mucin Muc2 (Gum et al., 1989). The rationale for selecting this strategy was a hypothesis that multiple GalNAc-transferases would
exist. Initially, we attempted to use acceptor peptides with few threonine or serine sites to limit the number of different substrate sites, but to date only the Muc2 sequence with multiple threonine residues yielded significant purification. As shown in the following paper, our affinity chromatography apparently resulted in the separation of at least two distinct GalNAc-transferase activities (Sørensen et al., 1995).

The identity of the isolated putative GalNAc-transferase cDNA was established by functional expression using the baculovirus system (Table III). The cloned GalNAc-transferase is predicted to be a type II transmembrane protein in concordance with all other cloned mammalian glycosyltransferases (Paulson and Colley, 1989; Kleene and Berger, 1993). The purified enzyme using Triton X-100 in the affinity chromatography steps was found to be soluble by gel filtration analysis, and this was confirmed by comparing the amino-terminal amino acid sequence of the purified protein with the coding region predicted from the cloned cDNA (Fig. 3). Purification of other glycosyltransferases using the detergent Triton X-100 in the affinity chromatography step has also resulted in selective

FIG. 4. Multiple sequence alignment analysis (DNASIS, Hitachi) of human (GalNAc-T2, top) and bovine (GalNAc-T1, bottom) GalNAc-transferases. Subscript letters indicate changes in the human version of GalNAc-T1 (bovine/human GalNAc-T1 similarity 99% amino acid level and 95% at nucleotide level).

FIG. 5. Northern blot analysis of MKN45. Twenty μg of total RNA or 10 μg of poly(A)+ mRNA isolated from the MKN45 cell line was probed with 32P-labeled PvuII/HindIII 2782 probe fragment, TEB2.

FIG. 6. Northern blot analysis of human tissues. Multiple human tissue blot from Clontech as labeled was probed with 32P-labeled TEB2 probe.
isolation of the soluble forms of the enzymes (Weinstein et al., 1987; Clausen et al., 1990; Sarkar et al., 1991). An exception to this is the purification of the Galβ1-3GlcNAc α2-3-sialyltransferase, where both the membrane and soluble forms were isolated (Gillespie et al., 1992). The soluble forms of glycosyltransferases like GalNAc-T2 appear to be in the monomer form. Recently, the apparent membrane form of the β1,4-galactosyltransferase was purified and found to be a high molecular weight, probably multimeric, complex (Bendiak et al., 1993).

We found here that a detergent exchange prior to the final affinity chromatography (step 4) to n-octyl glucoside apparently resulted in purification of the membrane-bound form of the GalNAc-transferase. Initially, we purified the placenta transferase by this method because significantly more activity was recovered. However, validation of the purification was impaired by difficulties with further purification steps including ion exchange and/or gel filtration. Fractions isolated by SDS-PAGE with a molecular weight of approximately 60,000 were found to be amino-terminally blocked for sequencing, and gel filtration failed to yield distinct peaks (not shown). Because of the possibility of multiple copurified GalNAc-transferases we chose to focus on the soluble enzyme. Leaving placenta tissue thawing for 3–4 days at 4 °C prior to extraction appeared to increase the relative amount of soluble enzyme.

Most of the glycosyltransferases characterized to date have been found to be N-linked glycoproteins (Klee and Berger, 1993), although the β1-2-N-acetylglucosaminyltransferases I and II appear to be exceptions (Sarkar et al., 1991; Kumar et al., 1990; Klee and Berger, 1993). GalNAc-T1 has three N-linked glycosylation consensus sites (human GalNAc-T1 has four), and some of these appear to be utilized (Homa et al., 1993). GalNAc-T2 has one consensus site, but both N-glycanase digestion (not shown) and amino acid sequencing (Table II) indicate that the soluble enzyme lacks N-linked glycosylation, or at least that the site is only partially utilized.

The isolated and cloned human GalNAc-transferase is predicted to be a novel member of a family of polypeptide GalNAc-transferases. The similarity between the human and bovine GalNAc-transferases (44% at the amino acid level) is close to that found for the different members of the sialyltransferase family (Wen et al., 1993). Species similarities of glycosyltransferases at the amino acid level have generally been found to be within 95–98% (Klee and Berger, 1993), and here we show that the human counterpart of the bovine GalNAc-transferase (Homa et al., 1993) is 99% similar at the amino acid level (Fig. 4). We therefore suggest that a nomenclature be introduced to identify the GalNAc-transferases: GalNAc-T1 and GalNAc-T2, where the latter is the novel human transferase cloned in the present paper. Comparison of the two amino acid sequences shows two to three regions of high similarity (>80%) which could be targets for a PCR cloning strategy to identify potentially additional members of this family of transferases. Such an approach has been applied with success to the sialyltransferase family comprising to date six members with a highly conserved 55-amino acid segment in the putative catalytic domain of the transferases (Livingston and Paulson, 1993). Our preliminary results indicate that this strategy will be successful for GalNAc-transferases as well.

Homology among different glycosyltransferases is generally very limited with the exception of members of the α1-3/4 fucosyltransferases (FUT 3, 5, 6) (Weston et al., 1992) and the human blood group A/B with the α1-3 galactosyltransferase (J oziasse et al., 1991). Drickamer (1993) showed that despite lacking homology among three members of the sialyltransferase family, cysteine residues were conserved. The present findings that 12 of 13 cysteine residues in GalNAc-T2 align with GalNAc-T1 and that these are distributed throughout the proteins suggest that the enzymes have a similar overall structure. Potential disulfide bonding is expected to occur intramolecularly as evidenced by the molecular weight of GalNAc-T2 estimated by gel filtration (Fig. 2). Interestingly, GalNAc-transferase activity appears to be increased 2–3-fold in the presence of reducing agents (Wang et al., 1992), possibly suggesting an advantage of some “opening” of the predicted globular catalytic domain for substrate accessibility and catalytic activity at least as measured in vitro using peptide substrates.

The finding that multiple GalNAc-transferases are involved in O-glycosylation is important for the understanding of this predominant post-translational modification. Detailed studies of the fine specificity of the individual transferase members are necessary for defining the role of each enzyme, and a preliminary study on this presented in the accompanying paper (Sørensen et al., 1995) dearly shows that distinct differences in substrate specificity may be expected. Here we show that Northern blotting with our human cDNA sequence (Fig. 5) revealed a slightly larger mRNA and a different organ distribution than that published by Homa et al. (1993) for GalNAc-T1, although most organs express both transcripts. Differential cell/organ expression of different GalNAc-transferases may result in different GalNAc O-glycosylation processing between cells and species. This could have a significant impact on interpretations of the peptide specificity of O-glycosylation inferred from identified in vivo glycosylation (O’Connell et al., 1991; Elhammer et al., 1993) and may eventually result in the identification of more defined consensus sequences of the acceptor peptide sites of individual transferases.

In conclusion, the present data provide evidence that mucin-type GalNAc O-glycosylation is controlled by at least two distinctly different GalNAc-transferases that are expressed differentially in cells and organs. Further studies of the substrate specificity of these and potentially additional enzymes are required, but it may be anticipated that this understanding will be significant for our insight into O-glycosylation processing and of practical use, for example, for designing appropriate mammalian expression systems for recombinant glycoproteins in drug use.

Acknowledgments—We thank Drs. Thomas Kjeldsen, Peter Leth Jørgensen, and Alan Chester for helpful advice. We acknowledge gratefully Dr. S. Hakomori for help and support in the initial phases of the project.

REFERENCES

Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. A., and Kent, S. B. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 6970–6974

Bendiak, B., Ward, L. D., and Simpson, R. J. (1993) Eur. J. Biochem. 216, 405–417

Bennett, E. P., Steffenes, R., Clausen, H., Weghuis, D. O., and van Kessel, D. G. (1993) Biochem. Biophys. Res. Commun. 206, 318–325

Birken, S., Fetherston, J., Canfield, R., and Boime, I. (1981) J. Biol. Chem. 256, 1816–1823

Chanchoy, L., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159

Clausen, H. H., White, T., Takio, K., Titi, K., Stroud, M., Holmsen, H., Karkov, J., Thim, L., and Hakomori, S. (1990) J. Biol. Chem. 265, 1139–1145

Drickamer, K. (1993) Glycobiology 3, 2–7

Elhammer, Å., and Kornfeld, S. (1986) J. Biol. Chem. 261, 5249–5255

Elhammer, Å., Poo, R. M., Brown, E., Magiora, L., Hooijerheide, J., and Keesey, F. J. (1993) J. Biol. Chem. 268, 10029–10038

Friedman, M., Krull, L. H., and Cavins, J. F. (1970). J. Biol. Chem. 245, 3868–3871

Gillespie, W., Kelm, S., and Paulson, J. C. (1992) J. Biol. Chem. 267, 21004–21010

Gum, J., R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lampert, D. T. A., and Kim, Y. S. (1986) J. Biol. Chem. 261, 6480–6487

Hagen, F. K., Van Wuyckhuyse, B., and Tabak, L. A. (1993) J. Biol. Chem. 268, 18960–18965

Higgins, D. G., and Sharp, P. M. (1988) Gene (Amst.) 73, 237–244

Homa, F. L., Hollander, T., Lehman, D. J., Thomsen, D. R., and Elhammer, Å. P. (1993) J. Biol. Chem. 268, 12609–12616

Jenoff, N. (1990) Trends Biochem. Sci. 15, 291–294

Joziasse, D. H., Shaper, J. H., and J., Abs, E. W., and Shaper, N. L. (1991) J. Biol. Chem. 266, 6991–6998

Klee, R., and Berger, E. G. (1993) Biodrill. Biophys. Acta 1154, 283–325

Kumar, R., Yang, J., Larsen, R. D., and Stanley, P. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 9948–9952

J. Biol. Chem. 268, 10029–10038
Purification and Cloning of GalNAc-transferase

Lis, H., and Sharon, N. (1993) Eur. J. Biochem. 218, 1-27
Livingston, B. D., and Paulson, J. C. (1993) J. Biol. Chem. 268, 11504-11507
Mandel, U., Gaggero, B., Reibel, J., Hamilton Therkildsen, M., Dabelsteen, E., and Clausen, H. (1994) Acta Pathol. Microbiol. Immunol. Scand. 102, 695-702
Masaki, M., Tanabe, K., Nakamura, K., and Soejima, M. (1981) Biochim. Biophys. Acta 660, 44-50
Matsuura, H., Takio, K., Titani, K., Greene, T., Levery, S. B., Salyan, M. E. K., and Hakomori, S. (1988) J. Biol. Chem. 263, 3314-3322
Matsuura, H., Greene, T., and Hakomori, S. (1989) J. Biol. Chem. 264, 10472-10476
O'Connell, B., Tabak, L. A., and Ramasubbu, N. (1991) Biochem. Biophys. Res. Commun. 180, 1024-1030
Paulson, J. C., and Colley, K. J. (1989) J. Biol. Chem. 264, 17615-17618
Sadler, J. E. (1984) in Biology of Carbohydrates (Ginsburg, V., and Robbins, P. W., eds) pp. 199-331, John Wiley & Sons, New York
Sarkar, M., Huo, E., Nishikawa, Y., Simpson, R. J., Moritz, R. L., Dunn, R., and Schachter, H. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 234-238
Sørensen, T., White, T., Wandall, H. H., Kristensen, A. K., Roepstorff, P., and Clausen, H. (1995) J. Biol. Chem. 270, 24166-24173
Springer, T. A. (1994) Cell 76, 301-314
Suguri, M., Kawasaki, T., and Yamashina, I. (1982) J. Biol. Chem. 257, 9501-9507
Varki, A. (1993) Glyobiology 3, 97-130
Wang, Y., Abernethy, J. L., Eckhardt, A. E., and Hill, R. L. (1992) J. Biol. Chem. 267, 12709-12716
Wang, Y., Agrawal, N., Eckhardt, A. E., Stevens, R. D., and Hill, R. L. (1993) J. Biol. Chem. 268, 22979-22983
Weinstein, J., Lai, P.-H., McEntee, K., and Paulson, J. C. (1987) J. Biol. Chem. 262, 17735-17743
Weston, B. W., Smith, P. L., Kelly, R. J., and Lowe, J. B. (1992) J. Biol. Chem. 267, 24575-24584
Yamamoto, F., Nishikawa, Y., Tsuji, T., White, T., Clausen, H., and Hakomori, S. (1990) J. Biol. Chem. 265, 1146-1151