Organization of the human β-1,2-*N*-acetylglucosaminyltransferase I gene (*MGAT1*), which controls complex and hybrid N-glycan synthesis

Betty YIP*, Shi-Hao CHEN*, Hans MULDER†§, Jo W. M. HÖPPENER‡ and Harry SCHACHTER*‖

*Research Institute, Hospital for Sick Children, 555 University Avenue, and Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada M5G 1X8, †Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, The Netherlands, and ‡Department of Pathology, AZU Hospital, Utrecht University, The Netherlands.

UDP-GlcNAc:α-3-D-mannoside β-1,2-*N*-acetylglucosaminyltransferase I (EC 2.4.1.101; GlcNAc-T I) is a medial-Golgi enzyme which catalyses the first step in the conversion of oligomannose-type to *N*-acetyl-lactosamine- and hybrid-type N-glycans and is essential for normal embryogenesis in the mouse. Previous work indicated the presence of at least two exons in the human GlcNAc-T I gene *MGAT1*, exon 2 containing part of the 5′ untranslated region and the complete coding and 3′ untranslated regions, and exon 1 with the remainder of the 5′ untranslated region. We now report the cloning and sequencing of a human genomic DNA fragment containing exon 1, which is between 5.6 and 15 kb upstream of exon 2. Transient transfection, ribonuclease protection and reverse transcriptase-mediated PCR indicated the absence of transcription start sites in intron 1 between exons 1 and 2. Northern analysis, ribonuclease protection, primer extension analysis and rapid amplification of 5′-cDNA ends showed that there are multiple transcription start sites for exon 1 compatible with the expression by several human cell lines and tissues of two transcripts, a broad band ranging in size from 2.7 to 3.0 kb and a sharper band at 3.1 kb. The 5′ flanking region of exon 1 has a GC content of 81 % and has no canonical TATA or CCAAT boxes but contains potential binding sites for transcription factors Sp1, GC-binding factor and epidermal growth factor receptor-specific transcription factor. Chloramphenicol acetyltransferase (CAT) expression was observed on transient transfection into HeLa cells of a fusion construct containing the gene for CAT and a genomic DNA fragment from the 5′ flanking region of exon 1. It is concluded that *MGAT1* is a typical housekeeping gene although there is, in addition, tissue-specific expression of the larger 3.1 kb transcript.

INTRODUCTION

The conversion of $Man_5GlcNAc_2Asn$ into *N*-acetyl-lactosamine (complex type) and hybrid N-glycans is controlled by UDP-GlcNAc:α-3-D-mannoside β-1,2-*N*-acetylglucosaminyltransferase I (EC 2.4.1.101; GlcNAc-T I) [1]. The gene for GlcNAc-T I has been cloned from rabbit [2], human [3,4], mouse [5,6] and rat [7]. Only a single copy of the human gene (*MGAT1*) was detected by Southern blot analysis and the gene was mapped to chromosome 5q35 [3,8]. Mice lacking a functional GlcNAc-T I gene die at approx. 10 days of embryonic life with multiple developmental abnormalities [9,10], indicating that *N*-acetyl-lactosamine- and hybrid-type N-glycans are important in cell–cell interactions. The organization of the GlcNAc-T I gene and the control of its expression are therefore of considerable interest.

Comparison of the human GlcNAc-T I genomic DNA sequence [3] with the cDNA sequence [4] indicated the presence of at least two exons, a 2.5 kb exon (exon 2) containing 127 bp of the 5′ untranslated region and the complete coding and 3′ untranslated regions, and a non-coding exon (exon 1) more than 5 kb upstream of exon 2. It was observed that 34 bases at the 5′ end of the human cDNA sequence were not present in exon 2 [3]. In the present study, an oligodeoxynucleotide probe containing part of this 34 bp sequence was used to isolate a human genomic DNA fragment containing GlcNAc-T I exon 1. Preliminary results had suggested that intron 1 between exons 1 and 2 might contain a functional promoter [3]; however, we have now proved that there are no transcription start sites in this region of the gene. We have also shown that there are multiple transcription start sites for exon 1 compatible with the expression of two transcripts in several human cell lines and tissues, a broad band ranging in size from 2.7 to 3.0 kb and a sharper band at 3.1 kb. The 5′ flanking region of exon 1 can stimulate transcription of the gene for chloramphenicol acetyltransferase (CAT) on transient transfection into HeLa cells, has a GC content of 81 %, has no canonical TATA or CCAAT boxes but contains potential binding sites for transcription factors Sp1, GC-binding factor (GCF) and the epidermal growth factor receptor-specific transcription factor (ETF). It is concluded that the human gene for GlcNAc-T I (*MGAT1*) is a typical housekeeping gene although there is, in addition, tissue-specific expression of the larger 3.1 kb transcript.

MATERIALS AND METHODS

Cultured cell lines

All cultured cell lines were grown in a humidified incubator at 37 °C in CO_2/air (1:19). Lec 1 cells, a mutant clonal derivative of Chinese hamster ovary (CHO) clone Pro-5 (provided by Dr. Pamela Stanley, Albert Einstein College of Medicine, New York,

Abbreviations used: CAT, chloramphenicol acetyltransferase; CHO, Chinese hamster ovary; ETF, epidermal growth factor receptor-specific transcription factor; FBS, fetal bovine serum; GCF, GC-binding factor; GlcNAc-T I, UDP-GlcNAc:α-3-D-mannoside β-1,2-*N*-acetylglucosaminyltransferase I (EC 2.4.1.101); MEM, minimal essential medium; *MGAT1*, human gene for GlcNAc-T 1; 5′-RACE, rapid amplification of 5′-cDNA ends; RT–PCR, reverse transcriptase-mediated PCR; SD buffer, 80 % formamide/0.1 % Bromophenol Blue/0.1 % xylene cyanol/1 × TBE; SSPE, 0.18 M NaCl/0.01 M sodium phosphate (pH 7.7)/0.001 M EDTA; TAE, 0.04 M Tris/acetate (pH 8.0)/0.001 M EDTA; TBE, 0.089 M Tris/borate (pH 8.3)/0.025 M EDTA; TE, 10 mM Tris/HCl (pH 8.0)/1 mM EDTA.

§ Present address: Glaxo–Wellcome Medicines Research Centre, Glycobiology Research Unit, Gunnels Wood Road, Stevenage, Herts. SG1 2NY, U.K.
‖ To whom correspondence should be addressed.

The nucleotide sequence reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number L77081.

NY, U.S.A.), have no detectable GlcNAc-T I activity [4,11,12] and are ideal for the analysis of promoter activity after transfection with the GlcNAc-T I gene [13]. HL-60 (provided by Laura Cory, Hospital for Sick Children, Toronto, Ontario, Canada) is a human promyelocytic leukaemia cell line grown in suspension culture in RPMI 1640 (Gibco/BRL) with 10 % (v/v) fetal bovine serum (FBS). HeLa, a human cervical epitheloid carcinoma cell line, and HepG2, a human hepatocellular carcinoma cell line, were provided by Dr. Adam Chen (Hospital for Sick Children, Toronto, Ontario, Canada) and were grown in α-minimal essential medium (αMEM) supplemented with 10 % (v/v) FBS. LS-180, derived from a human colonic adenocarcinoma, was provided by Dr. Fabienne Vavasseur (Hospital for Sick Children, Toronto, Ontario, Canada) and was grown in Eagle's MEM (Gibco/BRL), 0.1 M non-essential amino acids (Gibco/BRL) and 10 % (v/v) FBS. A-431 (American Type Culture Collection, Rockville, MD, U.S.A.) is a human epidermoid carcinoma cell line propagated in Dulbecco's MEM (Gibco/BRL) with 4.5 g/l glucose and 10 % (v/v) FBS (Gibco/BRL).

Molecular biology procedures

Unless otherwise stated, all molecular biology procedures were performed with standard techniques [14,15]. Restriction enzymes were obtained from Pharmacia, New England Biolabs, Gibco/BRL, Boehringer Mannheim or Stratagene. Oligodeoxynucleotides were synthesized on a Pharmacia DNA synthesizer and purified by the cartridge method (Hospital for Sick Children–Pharmacia Biotechnology Centre, Toronto, Ontario, Canada). DNA sequencing was performed by the dideoxy method [16] with the T7 DNA Sequencing Kit from Pharmacia with at least 4 μg of double-stranded DNA and α-^{35}S-labelled dATP at 10 mCi/ml (Amersham).

Plasmid DNA preparation

Competent *Escherichia coli* DH5α (Gibco/BRL) or MC1061:P3 (Invitrogen) transformed with the appropriate plasmid (pGEM plasmids into DH5α or pCDNA I/lacZ plasmids into MC1061:P3) were inoculated into Luria broth media containing the appropriate antibiotics (50 μg/ml ampicillin for DH5α; 12.5 μg/ml ampicillin and 7.5 μg/ml tetracycline for MC1061:P3) and grown overnight at 37 °C [15]. Plasmid DNA was isolated by alkaline lysis/SDS followed by column purification with the Qiagen DNA Purification Kit in accordance with the manufacturer's protocol. The purity of the DNA was determined by the ratio of absorbances at 260 and 280 nm; a ratio of 1.8 or above was considered pure.

Preparation of total RNA

Cultured cells were washed with PBS and total RNA was prepared either with the Pharmacia RNA Extraction Kit [17] or by a single-step method [18] with TRIzol Reagent (Gibco/BRL) in accordance with the manufacturer's instructions.

Synthesis of GlcNAc-T I probe

A probe was made by digesting plasmid Δ-1855 (see below) with *Pst*I to excise a 0.8 kb fragment containing almost the entire coding region for GlcNAc-T I. The DNA was purified by agarose-gel electrophoresis and Gene Clean (Bio 101), heat-denatured and labelled by the random priming method with the Pharmacia T7 QuickPrime Kit and [α^{32}P]dCTP (Amersham) in accordance with the manufacturer's protocol. The probe was purified with a Sephadex G-50 DNA Nick Column (Pharmacia) before use.

Northern blot analysis

Total RNA (15–20 μg) was separated in a formaldehyde/agarose gel [1 × Mops [14]/1 % (w/v) agarose/10 % (v/v) formaldehyde] submerged in 1 × Mops running buffer at 2.5 V/cm, 4 °C for 24 h. RNA was transferred by capillary action to Hybond-N membrane (Amersham). The efficiency of transfer was monitored by staining with 0.02 % Methylene Blue/0.3 M sodium acetate (pH 5.5). After the complete removal of the dye by 0.2 × SSPE/1 % SDS [where 1 × SSPE is 0.18 M NaCl/0.01 M sodium phosphate (pH 7.7)/0.001 M EDTA], the membrane was probed with 2×10^6 c.p.m./ml randomly primed labelled GlcNAc-T I probe in hybridization solution in accordance with the Amersham instructions for Hybond-N membranes. The membrane was sealed in a plastic bag and autoradiographed at −70 °C.

Screening of human leucocyte genomic DNA library for exon 1

A human leucocyte genomic DNA library in λEMBL-3 SP6/T7 was purchased from Clontech (catalogue no. HL1111j) and screened [8] with oligonucleotide probe SCH4 (5′-AAGTTCG-GGGCCAGGACGTCGGGAGGA-3′) derived from the 34 bp at the 5′ end of the cDNA sequence not present on exon 2 [4]. SCH4 (30 pmol) was end-labelled with T_4 polynucleotide kinase (Pharmacia) and [γ-^{32}P]ATP (Amersham) at 10 mCi/ml (6000 Ci/mmol) [15] and purified on a G-25 spin column (Pharmacia). Screening of membranes with the oligonucleotide probe [(1–2) × 10^5 c.p.m./ml] was performed as described in unit 6.4 of Ausubel et al. [14]. Hybridization was performed at 65 °C for 3 days. The membrane was sealed in a plastic bag and autoradiographed for 3 days to 1 week at −70 °C.

Southern blot analysis and subcloning of exon 1

Phage DNA (5 μg) was prepared from eight hybridization-positive clones (see above) and digested for 6 h with one of *Xba*I, *Eco*R1, *Sfi*I, *Bam*HI, *Sac*I or *Xho*I. The digests were fractionated in a TAE/0.7 % agarose gel [where TAE is 0.04 M Tris/acetate (pH 8.0)/0.001 M EDTA] at 1.5 V/cm overnight. Southern blot transfer was done in accordance with the protocol specified for Hybond N and Hybond N^+ nylon membranes (Amersham). The membrane was probed with labelled SCH4 oligonucleotide (see above), stripped in boiling 0.5 % SDS and re-probed with a 3.8 kb fragment generated by digesting plasmid pHG13 (from a human myeloid leukaemia genomic DNA library [3]) with *Sph*I and *Sal*I. This probe is derived from intron 1 (−1.9 to −5.7 kb relative to the ATG translation start site). The smallest restriction enzyme fragments that hybridized to the SCH4 probe were 3.1 kb *Sac*I and 3.0 kb *Xba*I fragments, which were respectively subcloned into pGEM5Zf(+) and pGEM7Zf(+) (Promega) to yield 13 *Sac*I subclones and 10 *Xba*I subclones from eight different phage colonies (1, 3–6, 8, 10, 11). These plasmids were subjected to limited restriction enzyme analysis. Four plasmids (*Sac*I subclones S1.1 and S6.1, and *Xba*I sublcones X1.2 and X5.3 representative of phage colonies 1, 5 and 6) were subjected to detailed restriction-enzyme analysis followed by Southern blotting with the SCH4 probe and to DNA sequencing.

Ribonuclease protection analysis

The location of transcription start sites was investigated by ribonuclease protection analysis. A riboprobe template was made for the analysis of exon 1. Plasmid S1.1 was cut with *Aat*II (cut

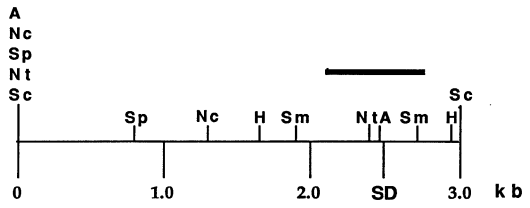

Figure 1 Restriction endonuclease map of plasmid S1.1 (see text) containing the exon 1 sequence

The vector was pGEM5Zf(+). Abbreviations: A, *Aat*II; H, *Hin*dIII; Nc, *Nco*I; Nt, *Not*I; Sc, *Sac*I; Sm, *Sma*I; Sp, *Spe*I. The splice donor junction (SD) between exon 1 and intron 1 is indicated. The solid bar indicates the sequence shown in Figure 2.

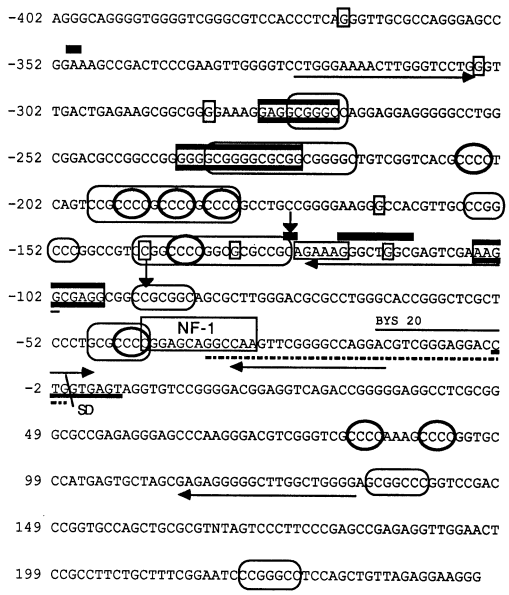

Figure 2 Nucleotide sequence of exon 1 of human GlcNAc-T I (from plasmid S1.1)

The sequence spans 648 bp, including 246 bp of intron 1 at the 3′ end. The splice donor (SD) consensus sequence between exon 1 and intron 1 is underlined. The splice junction is marked with a backslash. Indicated on the figure are: four GC boxes beginning at nucleotide positions −279, −238, −233 and −105 (bold rectangles); 16 GCF recognition domains (three between −235 and −219 and between −198 and −182; four between −143 and −126) (elongated ovals); eight putative binding sites for the ETF (bold ovals); and one NF-1-binding site (rectangle), which is so identified. Three of the CCCC sequences (starting at nucleotide positions −195, −190 and −139) form part of the AP2-binding sites. The major transcription start points deduced from ribonuclease protection are indicated by solid bars, from primer extension by boxes around the nucleotide(s) and from 5′-RACE by vertical arrows. The broken line denotes the 34 bp present in the 5′ untranslated region of human GlcNAc-T I cDNA and absent from exon 2. The horizontal arrows designate synthetic oligodeoxynucleotide primers used for sequencing.

site at −12 relative to the exon 1–intron 1 junction; Figures 1 and 2) and *Sac*I (the subcloning site) to excise a 2.5 kb band, which was subcloned into pGEM5Zf(+). The resulting plasmid was linearized with *Hin*dIII at 800 bp upstream of the *Aat*II restriction site (Figure 1). Plasmid Δ-1855 (Figure 3) was digested with *Pst*I and *Sph*I to obtain a 2.1 kb fragment (+20 to −1855 bp relative to the ATG translation start site [3]), which was subcloned downstream of the SP6 promoter of pGEM5Zf(+) to create subclone C2PS. Three riboprobe templates were synthesized for analysis of intron 1 as follows (Figure 4): (i) C2PS was cut at position −512 with *Hpa*I; (ii) C2PS was cut at position −255

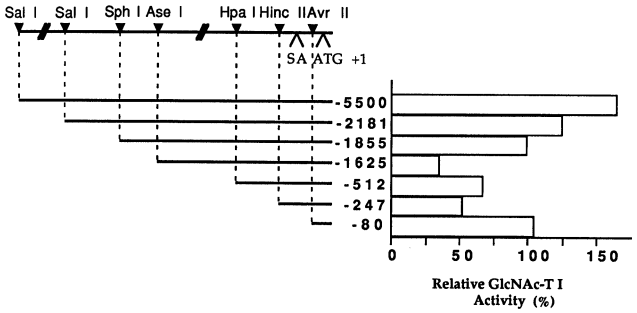

Figure 3 5′-Deletion analysis of promoter activity of intron 1

The figure shows the 5′-deletion constructs used for analysis of the promoter activity of intron 1. DNA fragments upstream of the GlcNAc-T I coding region are shown to the left with deletion end points indicated relative to the ATG start site at +1 (*y*-axis). All constructs contained the entire coding region and the native polyadenylation signal. Restriction enzyme sites, the splice acceptor site (SA) between intron 1 and exon 2 and the ATG start site are indicated [3]. The constructs were transiently transfected into Lec 1 mutants lacking GlcNAc-T I activity. Three days after the transfection, the cells were harvested, homogenized and assayed for GlcNAc-T I activity. The relative enzyme activities (mean percentages from three experiments, normalized to β-galactosidase activity) are shown on the right.

with *Sac*I, re-ligated and cut with *Ssp*I at position −866; (iii) C2PS was digested with *Dra*I (−480) and *Hin*cII (plasmid multiple cloning site), re-ligated and cut at −1128 with *Bal*I. The linearized riboprobe templates were extracted twice with 1 vol. of phenol/chloroform (1:1, v/v), twice with 1 vol. of chloroform, precipitated with 2 vol. of ethanol and suspended in RNase-free 10 mM Tris/HCl (pH 8.0)/1 mM EDTA to 1 μg/μl.

RNA run-off transcripts were made with 10 units of SP6 RNA polymerase (1 unit is 1 nmol of ATP incorporated into product in 60 min at 37 °C) at 40 °C for 2 h in extension buffer [40 mM Tris (pH 7.5)/6 mM $MgCl_2$/2 mM spermidine/10 mM NaCl/10 mM dithiothreitol/30 units of RNase Guard (Pharmacia) (1 unit is the amount of protein needed to inhibit 5 ng of RNase A by 50% using cCMP as substrate)/100 μg/ml BSA/0.5 mM each of UTP, ATP and GTP/10 μM [α-^{32}P]CTP (400 Ci/mmol; Amersham)/1 μg of linearized template]. The riboprobe was treated with 10 units of RNase-free DNase I (Pharmacia) (1 unit causes an increase in absorbance at 260 nm of 0.001/min per ml, using DNA as substrate, at 25 °C) at 37 °C for 15 min, extracted with phenol/chloroform (1:1, v/v), purified with a G-25 (Pharmacia) spin column, precipitated with ethanol, resuspended in SD buffer [80% formamide/0.1% Bromophenol Blue/0.1% xylene cyanol/1 × TBE, where TBE is 0.089 M Tris/borate (pH 8.3)/0.025 M EDTA] and separated by electrophoresis on a denaturing SDS/PAGE sequencing gel (6% gel) (National Diagnostics). The gel was covered with plastic wrap and exposed to X-ray film for 1 min to locate the RNA transcript, which was cut out of the gel and eluted at 37 °C for 4 h with 0.5 M ammonium acetate, 1 mM EDTA, 0.1% SDS and 0.4 mg/ml tRNA. Riboprobe (10^5 c.p.m.) was added to 50 μg of total RNA or 20 μg of yeast tRNA (as a negative control), co-precipitated with ethanol and resuspended in 80% (v/v) deionized formamide/0.4 M NaCl/40 mM Pipes (pH 6.7)/1 mM EDTA. The solution was heated to 85 °C for 5 min, hybridized at 50 °C overnight and digested with 1.75 μg of RNase A (Pharmacia) and 35 units of RNase T1 (Boehringer Mannheim) in 0.2 M NaCl/0.1 M LiCl/30 mM Tris (pH 7.5)/3 mM EDTA at 30 °C for 1 h. The digestion was terminated by the addition of 0.5% SDS and 50 μg of proteinase K (Gibco/BRL) at 37 °C for 15 min and extracted with phenol/chloroform (1:1, v/v). RNA

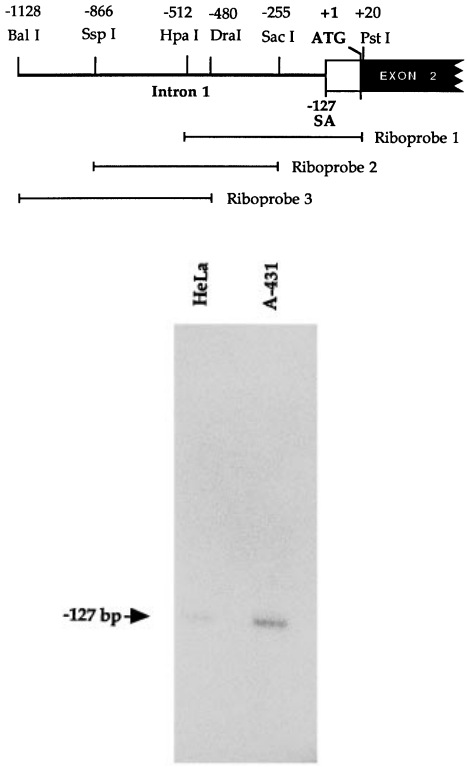

Figure 4 Ribonuclease protection of intron 1

Upper panel: locations of the three riboprobes used to analyse intron 1. The restriction sites used to create the boundaries for the riboprobes are indicated. The splice acceptor site (SA) is located at −127 bp relative to the ATG translation start site at +1. The black box is the coding region; the white box is the 5′ untranslated region. Lower panel: results of ribonuclease protection analysis with riboprobe 1. Total RNA (50 μg from HeLa or A-431 cells) was hybridized with riboprobe 1 and digested with RNase; the incubation was fractionated by electrophoresis and then autoradiographed. The size of the protected fragment was determined from a simultaneously run DNA sequencing ladder (results not shown) with adjustment for the differences in the mobilities of RNA and DNA. Yeast tRNA was substituted for total RNA in a control experiment and no sequence was protected (results not shown). The −127 bp fragment indicates the predicted site of the intron 1–exon 2 splice junction [3]. No other protected fragment was observed with riboprobe 1 or with either riboprobe 2 or 3 (results not shown), indicating the absence of a transcription start site in intron 1.

was precipitated with ethanol, resuspended in SD buffer and separated on a SDS/PAGE sequencing gel (6 % gel). The gel was dried and autoradiographed.

Primer extension analysis

Exon 2 anti-sense oligodeoxynucleotide 5′-GGGATGCCTCC-TCTGGACTATGGGATTAGG-3′ (nucleotide positions −83 to −112 relative to the ATG translation start site) [3] was end-labelled with [γ-^{32}P]ATP (6000 Ci/mmol; Amersham) [15], purified and resuspended in 30 μl of RNase-free water. The probe (10^5 c.p.m.) was added to 50 μg of total RNA in 0.4 M NaCl and 10 mM Pipes, pH 7.4, denatured at 85 °C for 5 min, kept at 42 °C for 1 h and extended in 50 mM Tris (pH 8.3)/75 mM KCl/3 mM $MgCl_2$/10 mM dithiothreitol/0.5 mM each of dATP, dTTP, dGTP and dCTP/33 units of RNase Guard (Pharmacia)/50 μg/ml actinomycin D (Pharmacia)/200 units of Superscript II RNase H-negative reverse transcriptase (BRL) at 48 °C for 1 h [1 unit incorporates 1 nmol of deoxyribonucleotide into product/10 min at 37 °C using poly(A)·oligo$(DT)_{12–18}$ as template primer]. The reactions were terminated in 1.2 mM EDTA, and RNA was degraded with 30 ng of DNase-free RNase A at 37 °C for 30 min. After extractions in 1 vol. of phenol/chloroform (1:1, v/v) followed by 1 vol. of chloroform, the samples were precipitated with ethanol, resuspended in SD buffer and analysed on a SDS/6 % polyacrylamide gel with a sequencing ladder as a standard, followed by autoradiography.

Reverse transcriptase-mediated PCR (RT–PCR) analysis

RT–PCR was performed to detect the presence of GlcNAc-T I transcripts. The following oligodeoxynucleotides were synthesized (nucleotide positions are relative to the ATG translation start site) [3]: BYA2906, 5′-CGTCCACCATCTGCTCC-TTGCCGTTGTCAT-3′ (exon 2 anti-sense primer for reverse transcription, +754 to +725); BYA2582, 5′-GCTAACGATG-ATGGGGAAGA-3′ (exon 2 nested anti-sense PCR primer, +420 to +401); BYS20, 5′-ACGTCGGGAGGACCTGGTGC-3′ (sense PCR primer containing the 3′ end of exon 1 and the first four bases at the 5′ end of exon 2; Figure 2); BYS1786, 5′-ACCTGGAGGGGAAATGAAAT-3′ (intron 1 sense PCR primer, −396 to −377). Total RNA (50 μg in 25 μl) was treated with 2.5 units of RNase-free DNase I (Pharmacia) in 20 mM Tris (pH 8.0)/20 mM $MgCl_2$ at 37 °C for 20 min to destroy contaminating DNA that might interfere with the RT–PCR analysis. After inactivation of the DNase at 90 °C for 5 min, the reaction was chilled on ice and 20 pmol of anti-sense primer BYA2906 was added to a final volume of 29 μl. The mixture was heated to 70 °C for 10 min and quickly placed on ice. Extension cocktail [50 mM Tris (pH 8.3)/75 mM KCl/3 mM $MgCl_2$/10 mM dithiothreitol/0.5 mM each of dATP, dTTP, dGTP and dCTP/0.15 mg/ml RNase-free BSA/200 units of Superscript RNase H-free reverse transcriptase (Gibco/BRL)] was added to a final volume of 50 μl and the mixture was incubated at room temperature for 10 min, then at 42 °C for 50 min, and inactivated at 90 °C for 5 min. The RNA template was destroyed by incubation with 5 units of RNase H (Promega) at 37 °C for 1 h. The cDNA product (10 μl) was added to 10 mM Tris/HCl, pH 8.3, containing 1.5 mM $MgCl_2$, 50 mM KCl, 0.001 % gelatin, 50 μM each of dATP, dTTP, dGTP and dCTP (Cetus), and two sets of PCR primers (each primer at a concentration of 0.2 mM), BYS20–BYA2582 or BYS1786–BYA2582. After Ampliwax (Perkin–Elmer) had been layered on the samples, the reaction was heated to 96 °C for 5 min and cooled to 55 °C before 1 unit of AmpliTaq DNA polymerase (250000 units/mg; see the Perkin–Elmer catalogue) was added. The samples were subjected to step cycles of 96 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min for a total of 30 cycles. The product was precipitated with ethanol and subjected to electrophoresis in a TAE/1 % agarose gel.

Rapid amplification of 5′-cDNA ends (5′-RACE)

The 5′ end of the *MGAT1* transcript was determined with the Marathon cDNA Amplification Kit (Clontech) by procedures recommended by the manufacturer. First-strand cDNA synthesis was conducted with LS-180 total RNA as template. The second strand was synthesized, blunt-ended and ligated at both ends to Marathon cDNA adaptor. The ligation mixture (10 μl) was heated at 70 °C for 5 min, diluted 5-fold with Tricine/EDTA buffer and subjected to three rounds of PCR as follows: (i) 1.5 μl of diluted adaptor-ligated double-stranded cDNA was used as template with 10 pmol each of Marathon adaptor primer AP1 and gene-specific (exon 2) anti-sense primer 5′-GGGATGCCT-CCTCTGGACTATGGG-3′ (nucleotide positions −83 to −106 relative to the ATG start site) [3]; (ii) 5 μl of first-round PCR reaction product was subjected to an identical second round of

PCR; (iii) 2 μl of second-round PCR reaction product was subjected to a third round of PCR with 20 pmol each of nested Marathon adaptor primer AP2 with an *Eco*RI site at its 5′ end and nested gene-specific (exon 1) anti-sense primer 5′-GCGCGAATTCGACGTCCTGG CCCCGAACTT-3′ (nucleotide positions −12 to −31 relative to the junction between exon 1 and intron 1, Figure 2, with an *Eco*RI site at its 5′ end, underlined). PCR conditions were as recommended by the manufacturer. The final PCR products were subcloned into the *Eco*RI site of pGEM7Zf(+) (Promega) and sequenced.

Construction of plasmids for GlcNAc-T I expression studies

The presence of a promoter region in intron 1 of the gene for GlcNAc-T I was studied by the construction of plasmids carrying DNA fragments containing the entire exon 2 sequence (approx. 2.5 kb with the complete coding and 3′ untranslated regions and the polyadenylation signal) but varying lengths of the upstream intron 1 sequence. A 6.5 kb DNA band was obtained by digestion of plasmid pHG15 (5 μg) [3] with *Sph*I. This fragment was subcloned into the *Sph*I site of pGEM5Zf(+) (Promega) and the product was named Δ-1855 (the *Sph*I site is 1855 bp upstream of the ATG translation initiation codon). Clones Δ-5500 and Δ-2181 are respectively equivalent to pHG13 and pHG15 [3], which were renamed to correspond to the above nomenclature. Smaller constructs were made by the deletion of various parts of the 5′ flanking region of the Δ-1855 clone at convenient restriction enzyme sites (Figure 3) as follows. Δ-1855 was digested with *Ase*I and *Sph*I to obtain a 6.3 kb band, which was subcloned into pGEM5Zf(+) cut with *Nde*I and *Sph*I to create the Δ-1625 clone. The Δ-512 clone was made by digestion of Δ-1855 with *Hpa*I and *Eco*RV followed by re-ligation with T4 DNA Ligase (Gibco/BRL). The Δ-247 clone was made by digestion of Δ-1855 with *Hin*cII and *Eco*RV followed by re-ligation. Δ-1855 was digested with *Avr*II and a 4.7 kb fragment was obtained and ligated into pGEM5Zf(+) cut with *Spe*I to create clone Δ-80. All constructs were confirmed by DNA sequencing.

Expression of GlcNAc-T I in Lec 1 cells

Lec 1 CHO cells (2.5×10^5 per dish) were seeded on 60 mm diameter culture dishes in 2 ml of αMEM, 10% (v/v) FBS, at least 24 h before transfection. Each dish was washed twice with 3 ml of PBS (Sigma) before the addition of the transfection cocktail [10% Nu Serum-V (Collaborative Research)/350 μg/ml DEAE-dextran (Pharmacia)/100 μM chloroquine (Sigma)/50 mM Tris (pH 7.3)/10 μg of GlcNAc-T I-encoding plasmid DNA/10 μg of β-D-galactosidase control plasmid (pCDNA I/lac Z)]. After a 90 min incubation at 37 °C in 5% CO_2, the dishes were washed twice with 3 ml of αMEM. To promote the production of protein, the cells were shocked in 10% (v/v) DMSO/PBS for 1 min, followed by two washes in 3 ml of PBS. Finally the cells were grown in 4 ml of αMEM/10% (v/v) FBS for another 72 h before harvesting. As negative controls, wild-type CHO cells and Lec 1 mutant cells were mock-transfected and Lec 1 cells were transfected with pGEM5Zf(+). Transfected cells were scraped from the culture dish, washed twice in 10 ml of PBS and resuspended in 100 μl of PBS with 2% (v/v) Triton X-100, 0.15 M NaCl, 50 μg/ml aminoethylbenzenesulphonyl fluoride (ICN), 0.5 μg/ml leupeptin (Sigma), 0.7 μg/ml pepstatin (ICN) and 5 μg/ml aprotinin (Sigma). Cells were lysed at 4 °C with an Eppendorf hand-held homogenizer. After microcentrifugation for 15 min at 4 °C, GlcNAc-T I activity was assayed [19,20] with Manα1-6(Manα1-3)Manβ-octyl as an acceptor (a gift from Dr. Hans Paulsen, University of Hamburg, Hamburg, Germany) and UDP-*N*-[^{14}C]acetyl-D-glucosamine [21] as the donor substrate. The transfection efficiency was standardized by co-transfection with the β-galactosidase gene. A control dish of Lec 1 cells from each transfection experiment was fixed with 0.1 M sodium phosphate pH 7.3/0.2% (v/v) glutaraldehyde and stained with 0.1 M sodium phosphate, pH 7.3, containing 0.3 mM $K_3Fe(CN)_6$, 0.3 mM $K_4Fe(CN)_6$ and 100 μg/ml 5′-bromo-4-chloro-3-indolyl β-D-galactopyranoside (Gibco/BRL) at 37 °C overnight. The ratio of blue-stained to unstained cells served as an index for transfection efficiency.

Analysis of transcriptional regulation by using the CAT reporter system

Plasmid S1.1 (20 μg; Figure 1) was sequentially digested with *Spe*I and *Aat*II and a 1.6 kb DNA fragment was purified by agarose-gel electrophoresis and the QIAquick Gel Extraction Kit (Qiagen). The DNA fragment and *Xba*I-digested pCAT-Basic Vector (2.1 μg, without promoter or enhancer; Promega) were blunt-ended with T4 DNA polymerase (New England Biolabs) with 0.1 mM of each of the four dNTPs at 12 °C for 20 min. The DNA preparations were purified with phenol/chloroform/isoamyl alcohol (25:24:1, by vol.), precipitated with ethanol in the presence of ammonium acetate, ligated with T4 DNA Ligase (Gibco/BRL) at 16 °C overnight, and competent *E. coli* DH5α cells were transformed by standard procedures. Plasmid DNA was extracted from ampicillin-resistant colonies by using the QIAprep Plasmid Kit (Qiagen), digested with *Nco*I and analysed by agarose-gel electrophoresis. DNA sequencing identified plasmid pCAT-GnT I containing the gene for CAT downstream of a DNA fragment in the sense direction from nucleotide positions −1600 to −12 relative to the exon 1–intron 1 junction (Figure 2).

For transfection, 6.8×10^5 HeLa cells were seeded in each 60 mm tissue culture plate in 3.5 ml of αMEM with 10% (v/v) FBS. Adherent cells were transfected at 70–80% confluence. Plasmids pCAT-GnT I (2 μg) and pSV-β-galactosidase (2 μg; Promega) and Lipofectamine reagent (8 μl; Gibco/BRL) were diluted separately with 0.18 ml of Opti-MEM I Reduced Serum Medium (Gibco/BRL). The three solutions were combined and incubated at room temperature for 45 min, after which 1.4 ml of αMEM (free of serum and antibiotics) was added and the solution was layered over the cells, which had been washed with 3.5 ml of the same medium. Transfections were performed with pCAT-Basic as a negative control and with pCAT-Promoter (containing the CAT gene and SV40 promoter but no enhancer) as a positive control. The cells were incubated for 5 h at 37 °C in a CO_2 incubator, and 1.7 ml of αMEM/20% (v/v) FBS were added. Medium was replaced at 21 h and, after a further 51 h incubation, cells were harvested and extracted by using the Tris buffer freeze–thaw protocol as described in the Promega technical bulletin. Cells were washed with Mg^{2+}- and Ca^{2+}-free PBS buffer, 1 ml of 40 mM Tris/HCl (pH 7.5)/1 mM EDTA/150 mM NaCl was added, then cells were incubated for 5 min at room temperature, centrifuged, resuspended in 0.1 ml of 0.25 M Tris/HCl, pH 8.0, and subjected to three cycles of rapid freeze–thaw. Extracts to be used for CAT assays were heated at 60 °C for 10 min before assay.

CAT assays were conducted on transfected and untransfected HeLa cell extracts by using [^{14}C]chloramphenicol (58.4 mCi/mmol; DuPont–NEN) and n-butyryl-CoA (Sigma), a 30 min incubation time, and the mixed xylenes (Aldrich) phase-separation assay (Liquid Scintillation Counting assay), in accordance with the Promega protocol. At least four independent transfections were performed for each experiment. CAT activity

was determined from a linear standard curve with CAT supplied by Promega (1 unit of enzyme activity corresponds to 1 nmol of acetate transferred to chloramphenicol per min at 37 °C). Assays of β-galactosidase activity were conducted at 28 °C with 1.2 ml incubation mixtures at pH 7.0 containing cell extract, 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-mercaptoethanol and 0.8 mg of *o*-nitrophenyl β-galactoside (ONPG). When the incubation developed a faint yellow colour, 0.5 ml of 1 M sodium carbonate was added and the absorbance at 420 nm was determined (1 unit of enzyme activity corresponds to the hydrolysis of 1 nmol of ONPG per min at 28 °C, with a molar extinction coefficient of 4500 for *o*-nitrophenol).

RESULTS

There are two GlcNAc-T I transcripts in human cultured cells and tissues

Northern analysis of RNA from mouse [5,6] and rat [7] has previously shown two GlcNAc-T I transcripts at 2.9–3.0 and 3.3 kb. Analysis of total RNA from five human cell lines (Figure 5) and seven human tissues (Figure 6) with a GlcNAc-T I probe showed two transcripts at 2.7–3.0 and 3.1 kb. Similarly to previous findings in mouse [5,6] and rat [7], human brain expressed only the larger 3.1 kb transcript. The latter transcript is clearly seen in heart, lung, skeletal muscle and pancreas but is weakly expressed in kidney and is difficult to distinguish from the shorter transcript in placenta and liver. The relative expression in different tissues varies with the species, e.g. expression is low in human brain and mouse muscle.

MGAT1 has an upstream non-coding exon

We have previously isolated two human genomic DNA clones (pHG13 and pHG15) containing a single 2.5 kb exon (designated exon 2) with the entire GlcNAc-T I-coding and 3′ untranslated regions and 127 bp of the 5′ untranslated region [3]. The 5′ end of exon 2 does not contain 34 bp found at the 5′ end of human GlcNAc-T I cDNA [4], indicating at least one non-coding exon (designated exon 1) upstream of exon 2. A λEMBL-3 human leucocyte genomic DNA library was therefore screened with a 27-mer oligonucleotide probe (SCH4) derived from these 34 bases, and 10 positive bacteriophage clones were identified. Southern analysis was performed on restriction enzyme digests of eight phage DNA preparations with probe SCH4 and with a 3.8 kb probe from intron 1 (-1.9 to -5.7 kb relative to the ATG translation start site). Four phage colonies yielded bands that hybridized to both SCH4 and the intron 1 probe (results not shown). The smallest of these bands was an *Eco*RI 10 kb fragment indicating that the maximum size of intron 1 is approx. 15 kb. We have previously shown that intron 1 is at least 5.6 kb long [3].

Sequencing of four SCH4-reactive subclones (*Sac*I subclones S1.1 and S6.1, and *Xba*I subclones X1.2 and X5.3) showed that the sequence between nucleotide positions -130 to $+90$ (relative to the exon 1–intron 1 junction; Figure 2) was the same for all four subclones. Plasmid S1.1 (Figure 1) was sequenced more extensively (Figure 2). Sequencing was initiated with the sense primer BYS20, on the basis of previously published human GlcNAc-T I cDNA data [4], and was completed with other synthetic primers as shown (Figure 2). There is a 8/9 match 5′-CTG/GTGAGT-3′ donor splice site at the exon 1–intron 1 junction [22]. A search of the GenBank database has revealed that the sequence shown in Figure 2 has a 53.6% identity in a 466 nt overlap with the human α-1,3-fucosyltransferase involved in the synthesis of the ligand for ELAM-1 (E-selectin) [23].

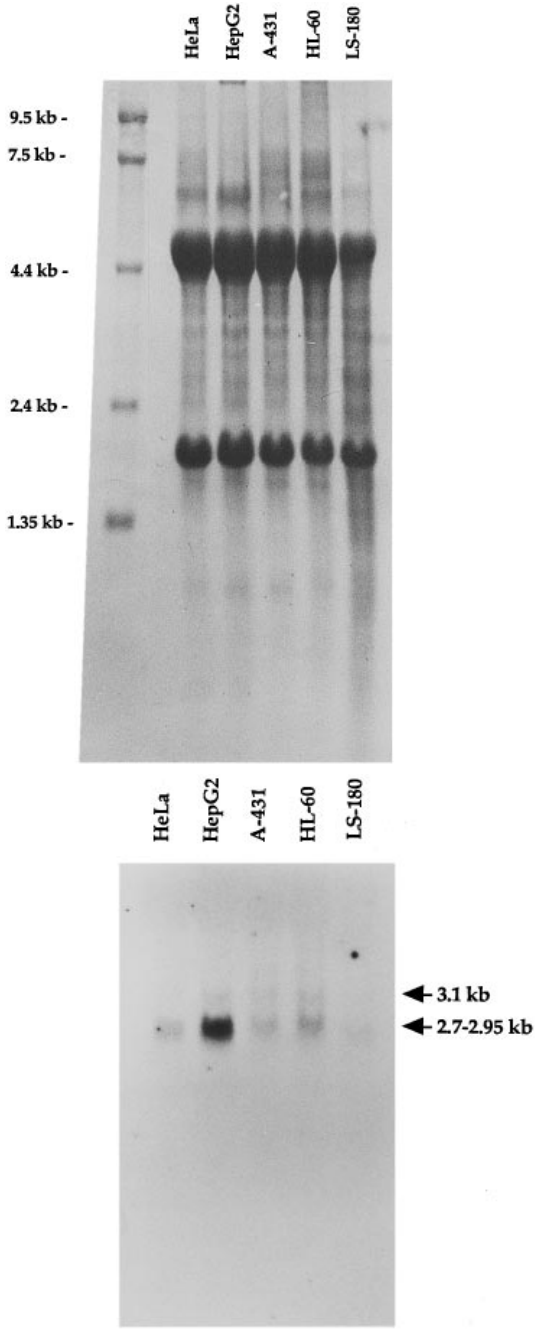

Figure 5 Northern analysis of GlcNAc-T I expression in human cell lines

Total RNA (10 μg for LS-180, 20 μg for the other four cell lines) was subjected to electrophoresis and Northern blotting. Upper panel: the membrane was stained with Methylene Blue to determine the quality and quantity of total RNA and the positions of the RNA markers (BRL). Lower panel: the membrane was hybridized with the 0.8 kb *Pst*I human GlcNAc-T I probe. Two transcripts are indicated by arrows, a broad band at 2.7–3.0 kb and a fainter band at 3.1 kb.

Ribonuclease protection

A ^{32}P-labelled RNA probe was generated from -12 to -800 bp relative to the exon 1–intron 1 junction, hybridized to total RNA from HepG2 and HeLa (results not shown) cells and subjected to RNase digestion. Protected fragments were detected from both cell lines (Figures 2 and 7) at -113 to -120, -125 and at

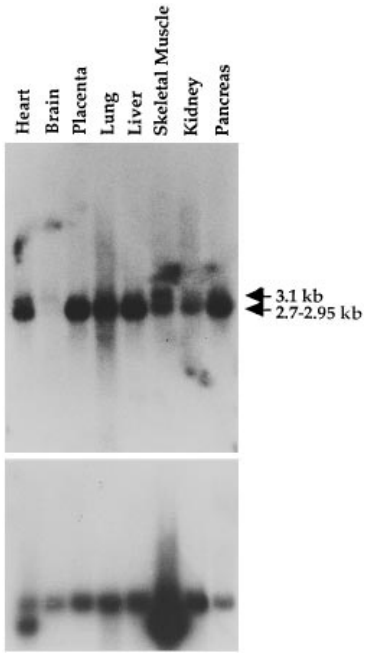

Figure 6 Northern analysis of GlcNAc-T I expression in human tissues

A human tissue Northern blot (Clontech) containing 2 μg of poly(A)-rich RNA from each of the indicated tissues was probed with the 0.8 kb *Pst*I human GlcNAc-T I probe. Arrows indicate two transcripts at 2.7-3.0 and 3.1 kb. The membrane was stripped and re-probed with a 2 kb human β-actin cDNA probe (Clontech); the results are shown in the bottom panel

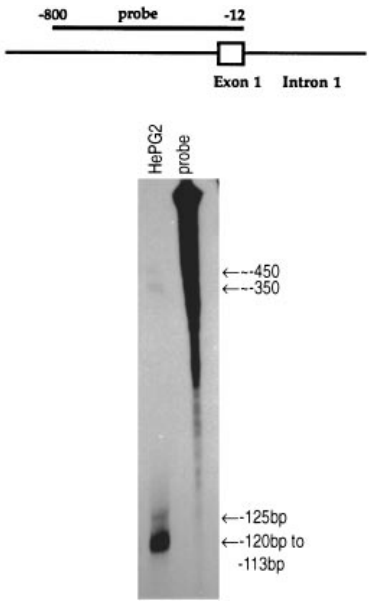

Figure 7 Ribonuclease protection analysis of exon 1

Total RNA (50 μg) from HepG2 cells was hybridized with an anti-sense RNA probe (top panel) corresponding to nucleotide positions -12 to approx. -800 bp relative to the exon 1–intron 1 splice junction (Figure 2) and digested with RNase; the resulting protected fragments were fractionated by electrophoresis and autoradiographed (bottom panel). The size of the protected fragments was determined from a simultaneously run DNA sequencing ladder (results not shown) with adjustment for the differences in the mobilities of RNA and DNA. Yeast tRNA was substituted for total RNA in a control experiment and no sequence was protected (results not shown). Protected bands indicated strong transcription start sites at -113 to -120 bp and -125 bp, and weaker signals at -350 and -450 bp. Undigested probe is shown.

approx. -350 and -450 bp relative to the exon 1–intron 1 junction. No protection was observed with yeast tRNA.

Primer extension analysis

A ^{32}P-labelled anti-sense oligonucleotide primer (corresponding to exon 2 nucleotide positions -83 to -112 relative to the ATG start codon) was hybridized with total RNA from HeLa, A-431, HepG2, HL-60 (results not shown) and LS-180 cell lines and extended by reverse transcription. The major extension end points in all cell lines were mapped to -120 to -125, -132, -142, -166, and -285 to approx. -500 bp, relative to the exon 1–intron 1 junction (Figures 2 and 8). No products were observed with a yeast tRNA template.

5′-RACE

5′-RACE was performed by subjecting a library of LS-180 cell cDNA ligated with Marathon cDNA adaptor to three rounds of PCR. The third PCR was performed with a nested Marathon adaptor primer and a nested gene-specific anti-sense primer from exon 1 (-12 to -31 relative to the exon 1–intron 1 junction). The PCR products were subcloned into pGEM7Zf(+) and four subclones were sequenced. Two distinct transcription start sites were detected at positions -92 and -126. Upstream transcription start sites were not detected by 5′-RACE, possibly because of the low expression of the longer 3.1 kb transcript.

CAT assays

HeLa cells were transiently co-transfected with plasmid pCAT-GnT I containing the gene for CAT downstream of a DNA fragment from nucleotide positions -1600 to -12 relative to the exon 1–intron 1 junction (Figure 2) and a plasmid encoding β-galactosidase. The CAT activities of cell extracts were normalized for transfection efficiency on the basis of the β-galactosidase assays. The CAT activity obtained with pCAT-GnT I (0.47 units of CAT per unit of β-galactosidase, S.D. 0.07, $n = 4$) was 12 times greater than the activity obtained on transfection of control plasmid pCAT-Basic (0.04 units of CAT per unit of β-galactosidase, S.D. 0.005, $n = 5$). This difference was significant at a probability of $P < 0.0005$. The CAT activity with the pCAT-Promoter vector (positive control) was 15 times greater than the activity obtained with pCAT-Basic. The CAT and β-galactosidase activities of non-transfected HeLa cells were respectively 8.5% and 10.3% of the values obtained with HeLa cells transfected with pCAT-Basic.

MGAT1 is a housekeeping gene with multiple transcription start sites

The multiple transcription start sites for exon 1 determined by ribonuclease protection, primer extension analysis and 5′-RACE (Figure 2) predict transcripts ranging in size from 2.7 kb (starting at position -92) to 3.1 kb (starting at position -450). On the basis of ribonuclease protection and 5′-RACE, transcripts from 2.7 to 3.1 kb can be accounted for by exons 1 and 2.

The region 5′ to the putative transcription initiation site for the 2.7 kb transcript promotes CAT activity and has a structural organization characteristic of housekeeping gene promoters (Figure 2). The sequence -92 to -302 bp has an 81% GC content and no canonical TATA or CCAAT boxes. The lack of a TATA element in a promoter has been documented to cause

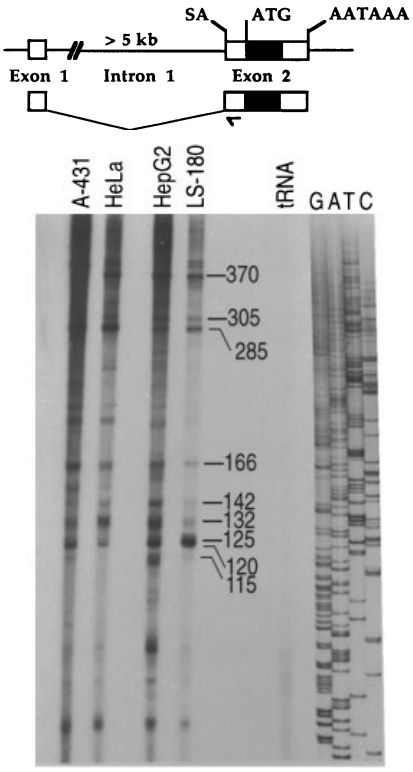

Figure 8 Mapping of the transcription start sites by primer extension analysis

A 5′ end-labelled synthetic anti-sense oligodeoxynucleotide primer (top panel, arrow, not drawn to scale) complementary to nucleotide positions −83 to −112 of exon 2 (relative to the ATG start codon) was annealed to 50 μg of total RNA from A-431, HeLa, HepG2 and LS-180 cells. The splice acceptor site (SA) is located at −127 bp relative to the ATG translation start site at +1. The black box is the coding region; the white boxes are exon 1 and the 5′ and 3′ untranslated regions of exon 2. A similar experiment was performed with HL-60 cells (results not shown). The mixtures were reverse-transcribed with Superscript II RNase H-negative reverse transcriptase and the extended products were subjected to electrophoresis on a denaturing SDS/PAGE sequencing gel (6% gel) and autoradiographed. A control was done in which total RNA was replaced with yeast tRNA. Three separate extension experiments gave similar results. One of these experiments is shown in the bottom panel. The major extension products are indicated as bp upstream of the exon 1–intron 1 splice junction (Figure 2). A DNA sequencing ladder is shown.

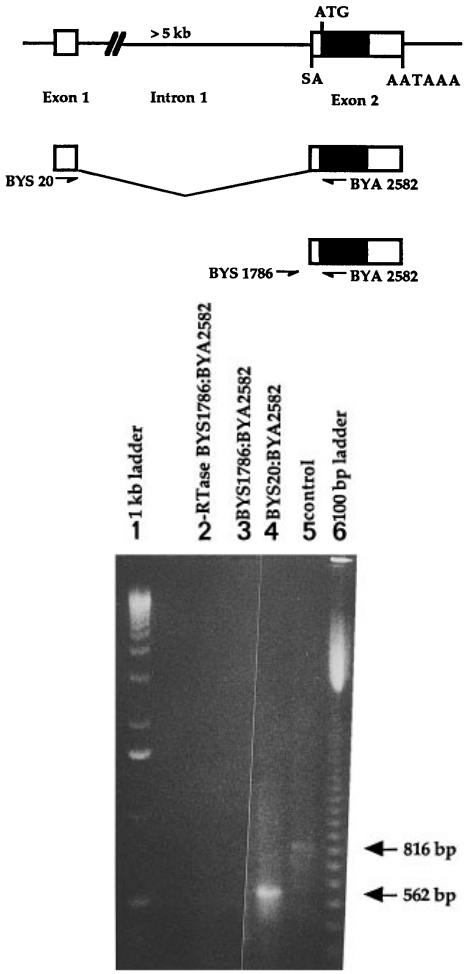

Figure 9 RT–PCR analysis of GlcNAc-T I transcripts

Top panel: human gene for GlcNAc-T I and the locations of PCR primers (not drawn to scale). The black box is the coding region; the white boxes show exon 1 and the 5′ and 3′ untranslated regions of exon 2. SA is the splice acceptor site between intron 1 and exon 2; AATAAA shows the polyadenylation site. Bottom panel: PCR was performed on reverse-transcribed HeLa cell cDNA with primer sets BYS20–BYA2582 (lane 4, expected to amplify 562 bp containing fragments of both exons 1 and 2) and BYS1786–BYA2582 (lane 3, expected to amplify 816 bp containing fragments of intron 1 and exon 2). Plasmid Δ-1855 containing the sequence required to produce the 816 bp PCR fragment was used as a positive control with primer set BYS1786–BYA2582 (lane 5). HeLa cell total RNA not incubated with reverse transcriptase was amplified with primer set BYS1786–BYA2582 as a negative control to assess contamination by genomic DNA (lane 2). Lane 1 is a 1 kb DNA ladder; Lane 6 is a 100 bp DNA ladder.

promiscuous initiation by RNA polymerase II, resulting in multiple initiation sites [24]. There are several putative binding sites for transcription factors that recognize GC-rich sequences (Figure 2), i.e. four GC boxes homologous with the Sp1 consensus binding sequence (5′-(G/T)(G/A)GGC(G/T)(G/A)(G/A)(G/T)-3′ [25] or 5′-N(G/A)GGCGN(G/A)N-3′ [26]) and eight binding sites 5′-CCCC-3′ for the ETF [27,28]. There are also 16 sites homologous with the GCF transcription repressor consensus sequence (5′-(C/G)CG(C/G)(C/G)(C/G)C-3′ [25,29]). The recognition sequences for GCF and Sp1 overlap between −276 to −271 and −235 to −225, which might be a mechanism by which positive and negative factors interact to regulate gene expression by binding to identical or overlapping DNA sequences [29].

There is no promoter in intron 1 of *MGAT1*

Plasmids pHG13 and pHG15 were made by subcloning segments of the human GlcNAc-T I gene into pGEM5Zf(+) [3]. The inserts of both plasmids contain large segments of intron 1 and all of exon 2. Transient transfection of either pHG13 or pHG15 into GlcNAc-T I-deficient Lec 1 CHO cell mutants resulted in expression of GlcNAc-T I enzyme activity, suggesting the presence of promoter activity in intron 1 between exons 1 and 2 [3]. This putative promoter was investigated by measuring GlcNAc-T I activity after transfection into Lec 1 cells of plasmid constructs with the entire exon 2 sequence and various lengths of upstream intron 1 sequence (Figure 3). Although enzyme activity decreased as nucleotides were deleted (from −5500 to −1625 bp relative to the ATG initiation start site), activity increased after further deletions and was still high with only 80 bp of intron 1 remaining. Transfections with either the pGEM5Zf(+) vector or mock transfections without plasmid did not result in expression of GlcNAc-T I activity, suggesting that there might be promoter

activity in pGEM5Zf(+) possibly due to cross-reactivity between the bacteriophage T7 and SP6 promoters on the plasmid and the mammalian transcription system [30]. The Δ-80 and Δ-5500 constructs were under control of the T7 promoter, whereas the other constructs (Δ-2181, Δ-1855, Δ-1625, Δ-512 and Δ-247) were under control of the SP6 promoter.

Ribonuclease protection was conducted to determine whether there was a transcription start site in intron 1. Total RNA from HeLa and A-431 cells was hybridized to an anti-sense radioactive RNA probe (riboprobe 1) corresponding to nucleotide positions +20 to −512 relative to the ATG translation start site (Figure 4) and digested with RNase. A single protected band was obtained (Figure 4), indicating termination of the transcript at −127 bp, the location of the intron 1–exon 2 junction [3]. No band was seen when total RNA was replaced with 20 μg of yeast tRNA (results not shown). Ribonuclease protection was also conducted with riboprobes 2 and 3 (Figure 4). Neither probe protected HepG2 and HeLa total RNA from RNase digestion (results not shown). It was concluded that there was no transcription start site in the region of intron 1 from −127 to −1128 bp relative to the ATG codon. Transcripts smaller than 3.5 kb cannot therefore be due to transcript initiation within intron 1.

The above conclusion was further supported by RT–PCR analysis. Single-strand cDNA was reverse-transcribed from HeLa cell total RNA and amplified with two sets of PCR primers. Primer set BYS20–BYA2582 served as a positive control and produced the 562 bp product expected from amplification between −16 bp relative to the exon 1–intron 1 junction and +420 bp relative to the ATG codon (Figure 9). If there had been transcript initiation in intron 1, primer set BYS1786–BYA2582 should have produced a 816 bp product owing to amplification between intron 1 at −396 bp and exon 2 at +420 (both positions relative to the ATG codon). However, no PCR product was detected (Figure 9). Therefore transfections with deletion constructs, ribonuclease protection and RT–PCR all failed to provide evidence for a transcription start site in intron 1.

DISCUSSION

GlcNAc-T I is essential for the processing of oligomannose-type N-glycans into complex and hybrid N-glycans [1]. Mice lacking a functional GlcNAc-T I gene die at about 10 days after fertilization with various developmental abnormalities [9,10]. Patients with a newly discovered inborn error of metabolism, carbohydrate-deficient glycoprotein syndrome type II, have a defect in UDP-GlcNAc:α-6-D-mannoside β-1,2-*N*-acetylglucosaminyltransferase II, an enzyme essential for complex N-glycan synthesis [31–33]; these patients suffer from severe developmental abnormalities. Complex N-glycans therefore play essential roles in embryogenesis and it is important to understand the factors which control GlcNAc-T I expression. Previous studies on the human gene for GlcNAc-T I *MGAT1* have indicated the presence of at least two exons, exon 2 containing the entire coding and 3′ untranslated regions and 127 bp of the 5′ untranslated region and exon 1 containing the remainder of the 5′ untranslated region [3,4]. The mouse gene for GlcNAc-T I has a similar organization except that there are at least two upstream non-coding exons [6]. The mouse and human cDNA sequences are 73 % identical over the 127 bp of the 5′ untranslated region immediately upstream of the translation initiation site (exon 2) but diverge completely in the upstream exon 1 (Figure 2) [3–6,34]. Several other glycosyltransferase genes have upstream non-coding exons [34,35].

In the present study we have used a probe designed from the known cDNA sequence for human GlcNAc-T I to clone exon 1 and we have also analysed the transcription of the gene for GlcNAc-T I. Northern analysis of human RNA from various cell lines and tissues showed two transcripts at 2.7–3.0 and 3.1 kb. The broad major transcript at 2.7–3.0 kb is due to several transcription start sites at nucleotide positions −92 to −350 bp upstream of the exon 1–intron 1 splice donor junction (Figures 2 and 7). The much weaker signal of the 3.1 kb transcript might be due to a transcription start site detected at nucleotide position −450. We have shown that our earlier hypothesis of a transcription start site in intron 1 [3] was incorrect.

Similarly to previous findings in mouse [5,6] and rat [7], human brain expressed only the larger 3.1 kb transcript. Data on the mouse and human genes for GlcNAc-T I differ in two respects: there are at least two non-coding upstream exons in the mouse gene, whereas we have so far identified only one such exon in the human gene, and the mouse transcripts are at approx. 2.9 and 3.3 kb respectively so that the 3.3 kb band can be more readily separated and identified on Northern analysis than in the human messages. The increased length of the mouse brain 3.3 kb transcript was shown to reside entirely in the 5′-untranslated region and evidence was obtained for the presence of two distinct promoters for the short and long mouse transcripts [36]. A probe specific for the mouse 3.3 kb transcript was developed [36]. Although the evidence we have obtained for a human tissue-specific GlcNAc-T I transcript is not as conclusive as that obtained for the mouse [36], the data suggest that there are tissue-specific promoters for the longer transcript in both species. Because the proteins produced by the two transcripts are the same, it is likely that their purpose is to allow differential regulation of enzyme levels by controlling transcription or translation rates or message stability.

The sequence upstream of exon 1 stimulates transcription of the CAT gene and shows an 81 % GC content and the absence of canonical TATA or CCAAT boxes. There are several putative binding sites for transcription factors that recognize GC-rich sequences, i.e. four GC boxes homologous with the Sp1 consensus binding sequence [37], eight binding sites for the ETF [27,28] and 16 sites homologous with the GCF transcription repressor consensus sequence [25,29]. The above properties are commonly associated with promoters for housekeeping genes.

Many housekeeping genes lacking a TATA box have been described [38–40], e.g. HMG-CoA reductase [41], dihydrofolate reductase [42] and hypoxanthine phosphoribosyltransferase [43,44]. These genes are expressed ubiquitously to provide basic and essential cell functions. The TATA box controls the accuracy of transcription through site-specific initiation [45]. The lack of a TATA element does not affect the rate of transcription but is often associated with multiple transcription initiation sites [24]. TATA-less promoters have regions with an unusually high GC content which may provide potential binding sites for Sp1 [37], ETF [27,28] and GCF [29]. GCF acts as a repressor to down-regulate gene expression. Frequently, the binding sites of GCF and Sp1 overlap, as in the human GlcNAc-T I gene (Figure 2). This overlap might provide a mechanism by which positive and negative factors interact to regulate gene expression [29] and might be responsible for the down-regulation of the smaller 2.7–3.0 kb transcripts in human brain. Alternatively, the 3.1 kb transcript might be under the control of a different promoter as has been suggested for the mouse GlcNAc-T I gene [36].

In conclusion, we have shown that the human GlcNAc-T I gene has at least one up-stream non-coding exon and that it is a typical housekeeping gene, although the minor 3.1 kb transcript might be under tissue-specific control.

We thank Dr. Pamela Stanley (Albert Einstein College of Medicine, New York, NY, U.S.A.), Laura Cory (Hospital for Sick Children, Toronto, Ontario, Canada), Dr. Adam Chen (Hospital for Sick Children, Toronto, Ontario, Canada) and Dr. Fabienne Vavasseur (Hospital for Sick Children, Toronto, Ontario, Canada) for kindly supplying us with cell lines. H. M. was on a leave of absence from graduate studies at Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, The Netherlands (under the supervision of Dr. J. P. Kamerling) during his stay in Toronto and was financially supported by the Netherlands Organization for Chemical Research (NWO). This work was supported by a grant to H. S. from the Medical Research Council of Canada.

REFERENCES

1 Schachter, H. (1986) Biochem. Cell Biol. **64**, 163–181
2 Sarkar, M., Hull, 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
3 Hull, E., Sarkar, M., Spruijt, M. P. N., Höppener, J. W. M., Dunn, R. and Schachter, H. (1991) Biochem. Biophys. Res. Commun. **176**, 608–615
4 Kumar, R., Yang, J., Larsen, R. D. and Stanley, P. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 9948–9952
5 Pownall, S., Kozak, C. A., Schappert, K., Sarkar, M., Hull, E., Schachter, H. and Marth, J. D. (1992) Genomics **12**, 699–704
6 Kumar, R., Yang, J., Eddy, R. L., Byers, M. G., Shows, T. B. and Stanley, P. (1992) Glycobiology **2**, 383–393
7 Fukada, T., Iida, K., Kioka, N., Sakai, H. and Komano, T. (1994) Biosci. Biotechnol. Biochem. **58**, 200–201
8 Tan, J., D'Agostaro, G. A. F., Bendiak, B., Reck, F., Sarkar, M., Squire, J. A., Leong, P. and Schachter, H. (1995) Eur. J. Biochem. **231**, 317–328
9 Metzler, M., Gertz, A., Sarkar, M., Schachter, H., Schrader, J. W. and Marth, J. D. (1994) EMBO J. **13**, 2056–2065
10 Ioffe, E. and Stanley, P. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 728–732
11 Stanley, P., Narasimhan, S., Siminovitch, L. and Schachter, H. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 3323–3327
12 Stanley, P., Caillibot, V. and Siminovitch, L. (1975) Cell **6**, 121–128
13 Kumar, R. and Stanley, P. (1989) Mol. Cell. Biol. **9**, 5713–5717
14 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1993) Current Protocols in Molecular Biology, John Wiley and Sons, New York
15 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
16 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
17 Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yokota, T. and Arai, K. (1987) Methods Enzymol. **154**, 3–28
18 Chomczynski, P. (1993) BioTechniques **15**, 532–535
19 Nishikawa, Y., Pegg, W., Paulsen, H. and Schachter, H. (1988) J. Biol. Chem. **263**, 8270–8281
20 Palcic, M. M., Heerze, L. D., Pierce, M. and Hindsgaul, O. (1988) Glycoconjugate J. **5**, 49–63
21 Brockhausen, I., Matta, K. L., Orr, J. and Schachter, H. (1985) Biochemistry **24**, 1866–1874
22 Mount, S. M. (1982) Nucleic Acids Res. **10**, 459—472
23 Goelz, S. E., Hession, C., Goff, D., Griffiths, B., Tizard, R., Newman, B., Chi-Rosso, G. and Lobb, R. (1990) Cell **63**, 1349–1356
24 Hansen, U. and Sharp, P. A. (1983) EMBO J. **2**, 2293–2303
25 Faisst, S. and Meyer, S. (1992) Nucleic Acids Res. **20**, 3–26
26 Berg, J. M. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 11109–11110
27 Kageyama, R., Merlino, G. T. and Pastan, I. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 5016–5020
28 Kageyama, R., Merlino, G. T. and Pastan, I. (1989) J. Biol. Chem. **264**, 15508–15514
29 Kageyama, R. and Pastan, I. (1989) Cell **59**, 815–825
30 Sandig, V., Lieber, A., Bähring, S. and Strauss, M. (1993) Gene **131**, 255–259
31 Charuk, J. H. M., Tan, J., Bernardini, M., Haddad, S., Reithmeier, R. A. F., Jaeken, J. and Schachter, H. (1995) Eur. J. Biochem. **230**, 797–805
32 Jaeken, J., Schachter, H., Carchon, H., Decock, P., Coddeville, B. and Spik, G. (1994) Arch. Dis. Childhood **71**, 123–127
33 Tan, J., Dunn, J., Jaeken, J. and Schachter, H. (1995) Glycoconjugate J. **12**, 478 (abstr.)
34 Schachter, H. (1994) in Molecular Glycobiology (Fukuda, M. and Hindsgaul, O., eds.), pp. 88–162, Oxford University Press, Oxford
35 Joziasse, D. H. (1992) Glycobiology **2**, 271–277
36 Yang, J., Bhaumik, M., Liu, Y. and Stanley, P. (1994) Glycobiology **4**, 703–712
37 Berg, E. L., Robinson, M. K., Mansson, O., Butcher, E. C. and Magnani, J. L. (1991) J. Biol. Chem. **266**, 14869–14872
38 Dynan, W. S. (1986) Trends Genet. **2**, 196–197
39 Dynan, W. S. (1989) Cell **58**, 1–4
40 Kadonaga, J. T., Jones, K. A. and Tjian, R. (1986) Trends Biochem. Sci. **11**, 20–23
41 Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G. G., Brown, M. S., Goldstein, J. L. and Luskey, K. L. (1984) Cell **38**, 275—285
42 Dynan, W. S., Sazer, S., Tjian, R. and Schimke, R. T. (1986) Nature (London) **319**, 246–248
43 Kim, S. H., Moores, J. C., David, D., Respess, J. G., Jolly, D. J. and Friedmann, T. (1986) Nucleic Acids Res. **14**, 3103–3118
44 Patel, P. I., Framson, P. E., Caskey, C. T. and Chinault, A. C. (1986) Mol. Cell. Biol. **6**, 393–403
45 McKnight, S. L. and Kinsbury, R. (1982) Science **217**, 316–324

Received 7 June 1996/5 September 1996; accepted 17 September 1996