Glycoconjugate Journal (1998) **15:** 301–308

EGC2
Transcriptional regulation of the human UDP-GlcNAc: *α*-6-D-mannoside *β*-1-2-*N*-acetylglucosaminyltransferase II gene (*MGAT2*) which controls complex N-glycan synthesis*

Shi-Hao Chen, Sihong Zhou, Jenny Tan and Harry Schachter†

Research Institute, Hospital for Sick Children, and Department of Biochemistry, University of Toronto, Toronto, Ontario, M5G 1X8 Canada

UDP-GlcNAc: *α*-6-D-mannoside *β*-1,2-*N*-acetylglucosaminyltransferase II (GnT II; EC 2.4.1.143) is essential for the normal assembly of complex Asn-linked glycans. Northern analysis showed a major transcript at 2.0 kb and a minor band at ~ 2.9 kb in five different human cell lines. The gene (*MGAT2*) has three AATAAA polyadenylation sites at 68, 688 and 846 bp downstream of the translation stop codon. 3′-RACE (rapid amplification of cDNA ends) using RNA from the human cell line LS-180 indicated that all three sites were utilized for transcription termination. 5′-RACE and RNase protection analyses showed multiple transcription initiation sites at −440 to −489 bp relative to the ATG translation start codon (+1). The data show that the entire GnT II gene is on a single exon. The gene has a CCAAT box at −587 bp but lacks a TATA box and the 5′-untranslated region is GC-rich and contains consensus sequences suggestive of multiple binding sites for Sp1; these properties are typical for housekeeping genes. A series of chimeric constructs containing different lengths of the 5′-untranslated region fused to the chloramphenicol acetyltransferase (CAT) reporter gene were tested in transient transfection experiments using HeLa cells. The CAT activity of the construct containing the longest insert (−1076 bp relative to the ATG start codon) showed a ~ 38-fold increase as compared to that of the control. Removal of the region between −636 and −553 bp caused a dramatic decrease in CAT activity indicating this to be the main promoter region of the gene.

***Keywords*: *N*-glycan synthesis; GlcNAc-transferase; transcription; promoter; housekeeping gene**

***Abbreviations*: bp, base pair; CAT, chloramphenicol acetyltransferase; FBS, fetal bovine serum; GnT, *N*-acetylglucosaminyltransferase; LB, Luria broth; kb, kilobases; MEM, minimal essential medium; MOPS, 3-[N-morphiolino]propanesulfonic acid; PBS; phosphate-buffered saline; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RNase, ribonuclease; SDS, sodium dodecyl sulfate; TE, 10 mM Tris–HCl, pH 8.0, 1 mM EDTA**

***Enzymes*: UDP-GlcNAc: *α*-6-D-mannoside *β*-1, 2-*N*-acetylglucosaminyltransferase, II, GnT II (EC 2.4.1.143)**

Introduction

Analysis of complex oligosaccharide structures indicates the existence of well over 100 different glycosyltransferases [1, 2]. The family of *N*-acetylglucosaminyltransferases (GnT) acting on the N-glycan core consists of GnT I to VI [3] which are responsible for the initiation of complex N-glycan antennae. UDP-GlcNAc:α-6-D-mannoside *β*-1,2-*N*-acetylglucosaminyltransferase II (GnT II; EC 2.4.1.143), a medial Golgi enzyme, catalyses incorporation of a GlcNAc residue in *β*-1,2 linkage to the Man-α-1,6 arm of the N-glycan core and is therefore an essential step in the biosynthetic pathway leading from hybrid to complex N-linked glycans. Carbohydrate-deficient glycoprotein syndrome type II has been shown to be a recessive autosomal disease in which the GnT II gene is inactivated [4–6]. These children have severe developmental abnormalities of the central nervous system. The organization of the GnT II gene and the control of its transcription are therefore of great interest.

The genes encoding rat [7] and human [8] GnT II have been cloned. The human GnT II protein sequence showed the domain structure typical of all previously cloned glycosyltransferases, i.e. a short nine-residue putative cytoplasmic N-terminal domain, a 20-residue hydrophobic non-cleavable putative signal–anchor domain and a 418-residue

* Dedicated to Professor Dr. Hans Paulsen, University of Hamburg, Hamburg, Germany, on the occasion of his 75th birthday. This paper was posted to the internet as part of the Second Electronic Glycosciences Conference held during September, 1996.

† To whom correspondence should be addressed.

0282–0080

C-terminal catalytic domain. Southern blot analysis indicated only a single copy of the human GnT II gene (*MGAT2*) which mapped to chromosome 14q21. The 1341 bp open reading frame of the human GnT II gene is flanked by a GC-rich 5′-untranslated region and a long AT-rich 3′-untranslated region containing three canonical polyadenylation signals (AATAAA) at 68, 688 and 846 bp downstream from the translation stop codon. Northern analysis of RNA from several human tissues showed a major signal at about 3 kb and a minor band at about 2 kb [8]. In the present work, we have used human cell lines to study the transcription of the human GnT II gene. Northern analysis showed a major transcript at 2.0 kb and a minor band at $\sim$ 2.9 kb in five different human cell lines. 3′-RACE (rapid amplification of cDNA ends) analysis of human cell line LS-180 showed that all three AATAAA sequences are utilized for transcription termination. 5′-RACE, S1 nuclease protection and RNase protection analyses showed multiple transcription initiation sites and also indicated that the entire GnT II gene is on a single exon. A region between -636 and -553 bp relative to the ATG start codon ($+1$) was identified as the main promoter region of the gene. The gene appears to be a typical housekeeping gene.

Materials and methods

Cultured cell lines

All cultured cell lines were grown in a humidified incubator at 37 °C in 5% carbon dioxide. HL-60 (provided by Laura Cory, Hospital for Sick Children, Toronto) is a human promyelocytic leukaemia cell line grown in suspension culture in RPM1 1640 (Gibco/BRL) with 10% 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) and were grown in α-MEM supplemented with 10% FBS. LS-180, derived from a human colonic adenocarcinoma, was provided by Dr Fabienne Vavasseur (Hospital for Sick Children, Toronto) and was grown in Eagle's MEM (Gibco/BRL), 0.1 M non-essential amino acids (Gibco/BRL) and 10% FBS. A-431 (American Type Culture Collection, Rockville MD) is a human epidermoid carcinoma cell line propagated in Dulbecco's MEM (Gibco/BRL) with 4.5 g/l glucose and 10% foetal bovine serum (Gibco/BRL).

Molecular biology procedures

Unless otherwise stated, all molecular biology procedures were carried out by standard techniques [9, 10]. Restriction enzymes were obtained from Pharmacia. Oligodeoxynucleotides were synthesized on a Pharmacia DNA synthesizer and purified by the cartridge method (Hospital for Sick Children–Pharmacia Biotechnology Centre, Toronto, Canada). All GnT II gene-specific primers are numbered according to the ATG translation initiation codon at $+1$; the letters f and r in primer names represent the forward and reverse directions respectively. Cultured cells were washed with PBS and total RNA was prepared either with the Pharmacia RNA Extraction Kit or by a single-step method [11] using TRIzol Reagent (Gibco/BRL) according to the manufacturer's instructions. DNA sequencing was carried out by the double-strand dideoxy method [12] using the T7 DNA Sequencing Kit (Pharmacia) with 2 μg of double stranded DNA and [α-^{35}S] dATP at 10 mCi/ml (Amersham).

Plasmid DNA preparation

Competent *E. coli* DH5α cells (Gibco/BRL) transformed with plasmid were inoculated into LB media containing 0.050 mg ml^{-1} ampicillin and grown overnight at 37 °C. Plasmid DNA was purified using the QIAprep Spin Plasmid Kit (Qiagen) following the manufacturer's protocol. Purity of the DNA was determined by the ratio of absorbances at 260 and 280 nm; a ratio of 1.8 was considered pure.

Synthesis of GnT II probe

A DNA probe from the 5′-end of the GnT II open reading frame was prepared as follows. Plasmid pHG36 (accession numbers U15128 and L36537) [8] was amplified by PCR using primers P119f (5′-CGTTGCTGGACG CCGAACC CG-3′) and P625r (5′-CCAATTTCAAAGCGGCATT-3′) and Taq polymerase (Perkin-Elmer). The resulting 507 bp product was purified from a 1.4% agarose gel using the GeneClean II kit (Bio 101 Inc) and labelled using the T7 DNA polymerase Quick Primer Kit (Pharmacia) with 0.050 mCi [α-^{32}P]dCTP (Amersham). The ^{32}P-labelled DNA was purified by passage through a Sephadex G-50 DNA Nick column (Pharmacia) and eluted with 0.4 ml TE (pH 8.0) buffer.

Northern analysis of RNA from human cell lines

Total RNA (0.020 mg) from each human cell line and RNA standard (Gibco BRL) were fractionated in a formaldehyde agarose gel (1 × MOPS, 1% agarose, 10% formaldehyde). RNA was transferred by capillary action to a 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. The membrane was hybridized with GnT II probe (1.5×10^6 cpm/ml), washed and exposed to X-ray film.

Rapid amplification of cDNA ends (RACE)

The 3′- and 5′-ends of the GnT II transcripts were determined with the Marathon cDNA Amplification Kit (Clontech) using procedures recommended by the manufacturer. First strand cDNA synthesis was carried out with LS-180 total RNA as template.

For the 3′-RACE procedure, GnT II cDNA was first amplified with a forward gene-specific primer P900f

(5′-CTATGGCATGGCTGACAAGGTAGA-3′) and reverse adapter primer AP1 (Clontech). The PCR product was re-amplified using primer P900f and nested reverse adapter primer AP2 (Clontech). The P900f-AP2 PCR product was further amplified with the following three primer pairs respectively (*Eco*RI sites are underlined and were introduced for subcloning):

AP2-P922f (5′-GATGTGAAAACTTGGAAATC-3′)

AP2-P1888f (5′-CGCGGAATTCGTGTTTTGAGTGT GTTTTGG-3′) and

AP2-P2109f (5′-CGCGGAATTCAGTGAAGGCATTC TACAAG-3′).

In 5′-RACE, forward adapter primer AP1 and reverse GnT II specific primer P21r (5′-TTTGTAGAT-GCGGAACCTCA-3′) were first used to amplify the GnT II cDNA 5′-end. The resulting PCR product was further amplified twice with primers *Eco*RI-AP2 (AP2 with an EcoRI site at the 5′-end) and P-47r (5′-CGCGGAATTCGCTG CTTCTGCACGGAA-3′).

All PCR products were purified from agarose gels using the GeneClean II kit (Bio 101 Inc). PCR product from the primer pair P922f-AP2 was subcloned into the *Pst*I (at nucleotide position 1336 in the human GnT II sequence) and *Not*I (generated by the Marathon RACE procedure) sites of pGEM5Zf(+) (Promega). The other PCR products were subcloned into the *Eco*RI site of pGEM7Zf(+) (Promega).

S1 nuclease protection analysis

A single strand DNA probe (+ 21 to − 342 bp relative to the ATG initiation codon at + 1) was synthesized by asymmetric PCR using 0.2 μM reverse primer (P21r, see above), 0.004 μM forward primer (P-342f, 5′-GAAGCTAGGG TGCGGTTGGG-3′), human GnT II genomic DNA as template (plasmid pHG36, see above), 50 μM each of dATP, dGTP and dTTP, 10 μM dCTP, 30 μCi [α-^{32}P]dCTP (3000 Ci/mmol, Amersham), 1.5 mM $MgCl_2$, 1 × PCR buffer (Perkin Elmer Cetus), 10% dimethylsulphoxide and 2 units Taq polymerase in 50 μl. The probe was purified using polyacrylamide gel electrophoresis and hybridized to total RNA prepared from HepG2 and LS-180 cells or to yeast tRNA as a control. S1 nuclease analysis was performed on the DNA-RNA complexes as previously described [9].

RNase protection analysis

Total RNA was prepared from LS-180 cells using the RNeasy Total RNA Purification Kit (Qiagen). The ratio of absorbances at 260 and 280 nm was 1.76 and the RNA was shown to be undegraded by denaturing agarose gel electrophoresis. Plasmids p-636CAT, p-553CAT and p-472CAT (see below) were digested with *Sa*lI and the inserts (− 636 to − 276, − 553 to − 276, − 472 to − 276 bp upstream of the ATG start codon) were subcloned into pGEM5Zf(+) to yield plasmids p-636Ribo, p-553Ribo and p-472Ribo respectively. Plasmid sequences were checked by double-strand sequencing. The plasmids were linearized with *Nde*I and used as templates to prepare labelled riboprobes [13] by incubation with T7 RNA polymerase (Boehringer-Mannheim), ATP, GTP, UTP, [α-^{32}P]CTP (800 Ci/mmol, Amersham), RNase inhibitor (Boehringer-Mannheim) and 10 × T7 RNA polymerase buffer (Boehringer-Mannheim). DNA was removed from the riboprobes by digestion with RNase-free DNase I. The riboprobes were purified by SDS–6% PAGE (sequencing gel) and hybridized either to LS-180 RNA or to yeast tRNA as a control. The RNA–RNA complexes were digested with RNase A and RNase T1 (Pharmacia) as previously described [13] and analysed on a sequencing gel.

Analysis of transcriptional regulation using the CAT reporter system

PCR was used to synthesize chimeric constructs containing sections of the 5′-untranslated region of the GnT II gene fused to the CAT reporter gene with pHG36 [8] as GnT II DNA template and Vent DNA polymerase (New England Biolabs). The ∼ 2.5 kb GnT II 5′-untranslated region in pHG36 was amplified using forward primer CAT/FL (5′-GGCCGCATGCCGATAAGCTTGATATCGAAT-3′) complementary to the plasmid pBluescript II(+) (Invitrogen) and reverse primer P-276Xr (5′-GCGCTCTAGATC-CTCTCCATTCAGCACCAC-3′, *Xba*I site underlined). The PCR product was digested with *Xba*I and a ∼ 0.7–0.8 kb DNA fragment (from − 276 bp to about − 1000 bp relative to the ATG start codon at + 1) was purified and subcloned into pCAT-Basic vector (Promega) to produce the p-1000CAT chimera. Chimeras p-680CAT, p-636CAT, p-553CAT, p-472CAT and p-390CAT were constructed with amplified GnT II gene fragments from − 680 to − 276, − 636 to − 276, − 553 to − 276, − 472 to − 276 and − 390 to − 276 (bp upstream of the ATG start codon) cloned into the pCAT-Basic vector, respectively. The forward primers used in these PCR amplifications (*Sal*I site underlined) were:

P-680f (5′-GCATGTCGACCTTCGCACGTCTCGCCTT TC-3′) P-636f (5′-ATATGTCGACGAAGAAAGAGG CGGAAGTGG-3′)

P-553f (5′-GCTAGTCGACTCGGTCGCGTCTGGAAA GCA-3′)

P-472f (5′-GCTAGTCGACGGGCAGTTGCGGGTT GTCAT-3′)

P-390f (5′-GCAGGTCGACTAAGGATGAGAGCGC AGAGG-3′).

The reverse primer was P-276Sr (5′-ATATGTCGACTC-CTCTCCATTCAGCACCAC-3′).

The constructs were transformed into competent *E. coli* DH5α cells and ampicillin-resistant colonies were screened

for the presence of the appropriate inserts by PCR with pCAT-Basic pUC/M13 primer (5′-CAGGAAACAGCTATGAC-3′) and a GnT II promoter region specific antisense primer. Plasmid DNA was extracted using QIAprep Plasmid Kit (Qiagen) and DNA sequences were verified by double-stranded DNA sequencing.

Adherent HeLa cells (6.8×10^5 cells were seeded) were transfected at 70–80% confluence (~21 h after seeding). pCAT chimeras (2 μg) and pSV-β-galactosidase (2 μg, Promega) were co-diluted with 0.18 ml Opti-MEM I Reduced Serum Medium (Gibco BRL). For each transfection, 8 μl Lipofectamine reagent (Gibco BRL) was diluted in a separate tube into 0.18 ml Opti-MEM I Reduced Serum Medium. The two solutions were combined and incubated at room temperature for 45 min, 1.4 ml α-MEM (free of serum and antibiotics) were added and the solution was layered over the cells which had been washed by 3.5 ml of the same medium. Transfections were carried out either without DNA or with pCAT-Basic as negative controls 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 carbon dioxide incubator and 1.7 ml α-MEM-20% FBS was added. Medium was replaced at 21 h after transfection and, after a further 51 h incubation, cells were harvested and extracted using the Tris buffer freeze/thaw protocol as described in the Promega technical bulletin. Extracts to be used for CAT assays were heated at 60 °C for 10 min prior to assay. CAT activity assays were carried out on transfected and untransfected HeLa cell extracts using [^{14}C]chloramphenicol (58.4 mCi/mmol, DuPont NEN) and *n*-butyryl coenzyme A (Sigma), a 30 min incubation time, and the mixed xylenes (Aldrich) phase separation assay (Liquid Scintillation Counting assay), according to the Promega protocol. Three–five independent transfections were carried out for each experiment. CAT activity was determined from a linear standard curve using a CAT standard (Promega, 1 unit enzyme activity corresponds to 1 nmole of acetate transferred to chloramphenicol per min at 37 °C). β-Galactosidase was assayed with *o*-nitrophenyl β-galactoside (ONPG); 1 unit of enzyme activity corresponds to the hydrolysis of 1 nmole of ONPG per min at 28 °C using a molar excinction coefficient of 4500 for *o*-nitrophenol at 420 nm.

Results

Expression of GnT II in human cell lines

Northern analysis of RNA from five human cell lines revealed that all cell lines showed the same pattern: a major transcript at ~2.0 kb and a minor band at ~2.9 kb (Figure 1). This contrasts with Northern analysis previously carried out on human tissues and lymphoblasts which showed a major transcript at about 3.0 kb and a minor band at 2.0 kb [8].

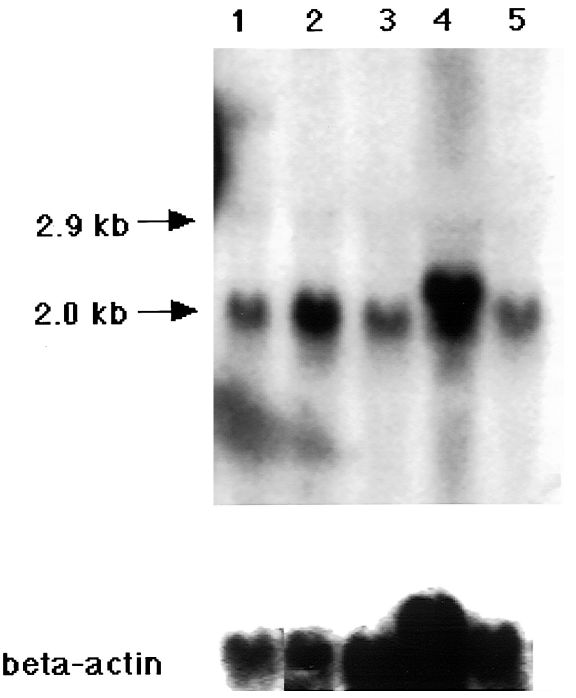

Figure 1. Northern analysis of human cell lines. Total RNA (0.020 mg) from (1) LS-180, (2) HepG2, (3) HeLa, (4) HL-60, and (5) A-431 cell lines was fractionated by electrophoresis in a denaturing 1.0% agarose gel (10% formaldehyde) and transferred to a membrane. The membrane was hybridized with a probe covering 507 bp of GnT II open reading frame (119–625 bp downstream of the ATG start codon). One major signal at ~ 2.0 kb and one minor band at ~ 2.9 kb were visible in all cell lines tested. The identity of a possible additional major band at 2.2 kb lane 4 (HL-60 cells) has not been determined. The membrane was stripped with hot 0.1% SDS and probed again with a 2 kb human β-actin cDNA probe (Clontech) (shown below the GnT II probed autoradiograph).

3′-Rapid amplification of cDNA ends (3′-RACE)

Electrophoretic analysis of the PCR product formed with the primer pair P922f and AP2 indicated a sharp band at ~600 bp which was subcloned into pGEM5Zf(+). Sequencing showed an exact sequence match to a section of the 3′-untranslated region at 1341–1430 bp relative to the ATG start codon with a poly(A) sequence commencing at position 1430, 20 bp downstream from the first AATAAA sequence. Similar experiments with PCR primer pairs P1888f-AP2 and P2109f-AP2 revealed, respectively, matches to 3′-untranslated region sequences 1888–2049 and 2109–2211 with poly(A) sequences commencing at positions 2050 (20 bp downstream from the second AATAAA element) and 2211 (23 bp downstream from the third AATAAA element). These experiments prove that all three polyadenylation sites found in the GnT II genome are functional in gene transcription termination at least in the human cell line tested.

5′-RACE

Adapter-ligated GnT II double strand cDNA template was amplified by PCR with AP1 as the forward primer and P21r, which overlapped the ATG start codon, as the reverse primer. This PCR product was reamplified twice with nested primers AP2 and P-47r to yield a ~500 bp DNA band which was subcloned. Eight subclones with a ~500 bp DNA insert were selected and sequenced. Sequences were obtained commencing at the following bp positions relative to the ATG start codon (+1): −489, −486, −477, −450, −443, −442 and −440.

S1 nuclease and RNase protection analyses

S1 nuclease protection analysis using a single strand DNA probe extending from nucleotide positions +21 to −342 (ATG initiation codon at +1, Figure 2) indicated full protection of mRNA from both HepG2 and LS-180 cells following digestion with S1 nuclease (data not shown). RNase protection analysis was performed on LS-180 total RNA. All three RNA probes showed protected bands at −442, −443, −450, −454 to −458, −462, −463, −467 to −469 and −482 (bp relative to the ATG initiation codon at +1) in excellent agreement with the results obtained by 5′-RACE (Figures 2 and 3). Control S1 nuclease and RNase digestions using tRNA instead of mRNA to protect the probe resulted in complete digestion of the probe (Figure 2).

Functional analysis of the presumptive human GnT II promoter

In order to characterize the regions regulating transcription of the GnT II gene, chimeric constructs containing different lengths of the 5′-flanking region fused to the CAT reporter gene were tested in transient transfection experiments using HeLa cells (Table 1). CAT activities of cell extracts were normalized for transfection efficiency based on β-galactosidase assays. Negative controls were performed using either untransfected HeLa cells or cells transfected with pCAT-Basic plasmid. The positive control (pCAT-Promoter) had a CAT activity 9.3-times higher than that of the pCAT-Basic control. The CAT activity of the construct containing the longest insert (p-1000CAT) showed a 38-fold increase relative to that of the pCAT-Basic control. Deletion of the 5′-flanking region from ~ −1000 to −680 bp resulted in a small reduction in CAT activity. There was no significant difference in CAT activity between p-680CAT and p-636CAT. A further deletion of 83 bp (p-553CAT) resulted in a seven-fold reduction in CAT activity. Further deletions showed activities similar to the p-553CAT plasmid. It is concluded that *cis*-acting regulatory elements between −636 and −553 bp relative to the ATG start codon play central roles in regulating the expression of the human GnT II gene.

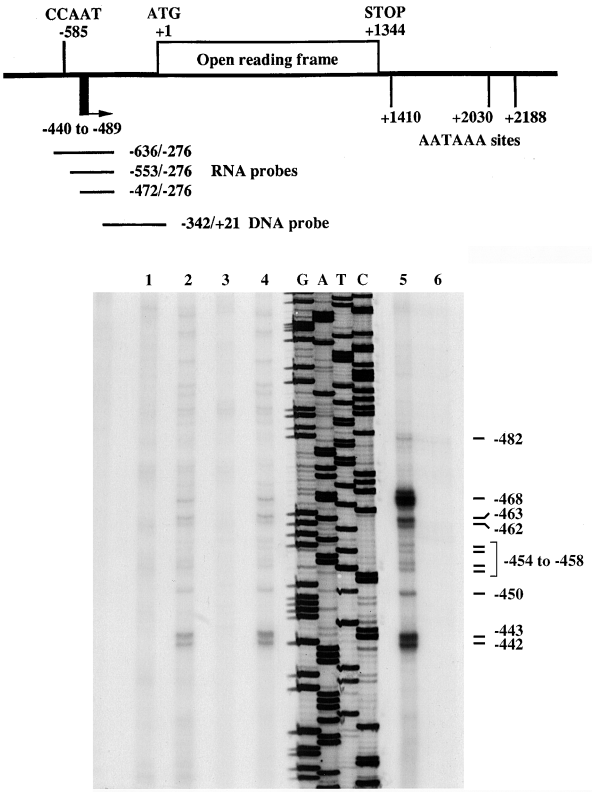

Figure 2. RNase protection analysis of the human GnT II gene. Top: Scheme of the human GnT II gene showing positions of the ATG start codon, STOP codon, open reading frame, CCAAT box and three functional polyadenylation sites. There are multiple transcription initiation sites between positions −440 and −489 (black rectangle). The three riboprobes used for RNase protection analysis and the DNA probe used for S1 nuclease protection analysis are indicated by horizontal lines. Bottom: Autoradiograph of sequencing gel electrophoretogram of products of RNase digestion. Lanes 1 and 2, riboprobe −472/−276 with tRNA and mRNA respectively. Lanes 3 and 4, riboprobe −553/−276 with tRNA and mRNA respectively. Lanes 5 and 6, riboprobe −636/−276 with mRNA and tRNA respectively. G, A, T, C, sequencing ladder. The positions of protected riboprobe fragments are indicated as base pairs relative to the ATG codon at +1.

Discussion

We have used 5′-RACE and S1 nuclease and RNase protection analyses to show that there are multiple transcription start sites within a 50 bp region (Figure 3) involved in expression of human GnT II transcripts. The data also indicate that the human GnT II gene is on a single exon and there are no upstream untranslated exons. Two transcription initiation sites at about 450 and 435 bp upstream of the translation start codon in the rat GnT II gene were previously identified by RNase protection experiments [7].

3′-RACE was used to show that there are three distinct functional polyadenylation sites. Assuming 200 bp of poly(A) tail, the shortest transcript terminating at the first poly(A) site should be about 2.0 kb, in agreement with the

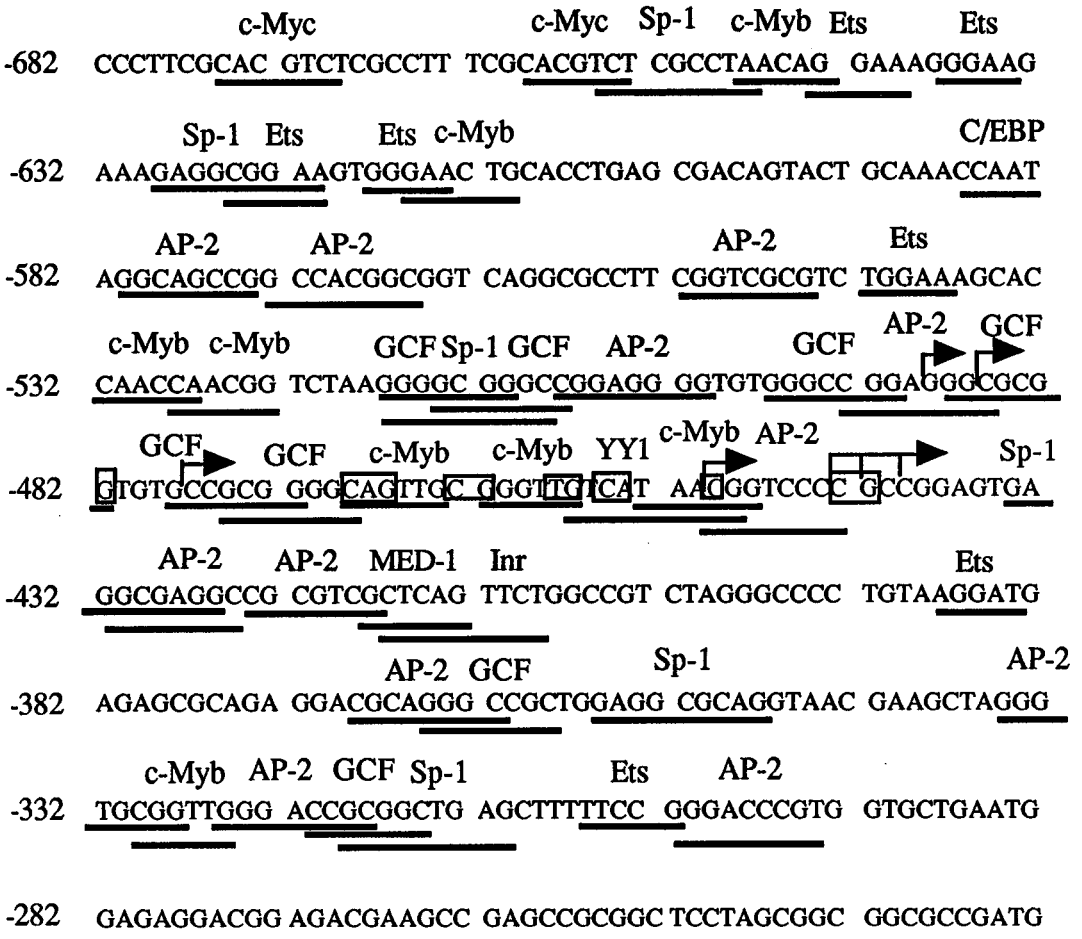

Figure 3. Nucleotide sequence of the 5′-end of the human GnT II gene. The DNA sequence upstream of the ATG translation initiation codon (at position +1) is shown. The transcription initiation sites were determined by 5-′RACE (arrows) and by RNase protection (boxes around the nucleotides). Putative transcription factor binding sites are underlined and named above the sequence. These binding sites are either exact fits to published consensus sequences (see text) or differ from these sequences by no more than a single nucleotide. Some sequences are present in the reverse orientation.

Northern analysis. The transcripts terminating at the second and third poly(A) sites should be about 2.7 and 2.9 kb, respectively. However, only one weak signal at 2.9 kb was detected in our Northern blots (Figure 1). Northern analysis of rat tissues with a rat GnT II probe showed a major transcript at about 2.8 kb and minor bands at 1.7 and 2.1 kb [7]. The two minor transcripts at 1.7 and 2.1 kb in rat tissues could not hybridize with a probe prepared from the 3′-untranslated region supporting the idea that multiple GnT II transcripts may result from the differential utilization of distinct poly A sites [7].

Many genes have been shown to have multiple mRNA transcripts due to differential polyadenylation site utilization and it has been suggested that this may be one of the mechanisms for the control of development and differentiation [14–16]. It is not clear why human tissues express primarily the longest (3 kb) of the three GnT II transcripts [8] whereas the major transcript expressed by the cell lines studied in this report is the shortest (2 kb). Further work needs to be done to resolve this question, particularly detailed analysis of the GnT II transcripts in human tissues.

Analysis of the GnT II 5′-untranslated region shows a GC content of 69% in the 682 bp upstream of the translation start codon. The observed over expected (O/E) frequency of the putative CpG doublet is 0.96, consistent with the presence of a promoter in this region since the frequency is larger than 0.6 [17]. A number of potential *cis*-elements for transcription factor binding have been identified within the GnT II gene 5′-flanking region (Figure 3) [18]. Although there is no TATA box, an initiator (Inr)-like element (YYCANTYYY) [19, 20] was identified at − 417 bp. Also detected (Figure 3) were a CCAAT box (CCAAT/enhancer binding protein family, C/EBP) at − 587 bp, an imperfect conserved multiple start site element downstream (MED-1, GCTCCC/G), a ying-yang 1 (YY1) element (core conserved sequence 5′-NNNCATNNNN 3′) [21] and several putative

Table 1. CAT assays on HeLa cells transfected with chimeric pCAT plasmids. Chemeric constructs containing sections of the region flanking the 5′-end of the GnT II open reading frame fused to the CAT reporter gene were transiently transfected into HeLa cells. Transfections were carried out either without DNA or with pCAT-Basic as negative controls and with pCAT-Promoter as a positive control. CAT activity assays were normalized for transfection efficiencies by carrying out co-transfections with a β-galactosidase encoding plasmid. Results are expressed as a percent of the CAT activity of the p-1000CAT plasmid.

Plasmid	*Nucleotide position of 5′-flanking region*	*CAT activity (%)*	*SD*	*n*
No plasmid		2.1	0.2	4
pCAT-Basic		2.6	0.4	5
pCAT-Promoter		24.2		1
p-1000CAT	− 276 to about −1000	100	13.2	3
p-680CAT	− 276 to − 680	74.5	5.4	4
p-636CAT	− 276 to − 636	75.8	3.6	4
p-553CAT	− 276 to − 553	10.6	1.1	4
p-472CAT	− 276 to − 472	14.6	5.1	4
p-390CAT	− 276 to − 390	7.0	2.8	4

n, Number of transfections; SD, standard deviation.

Sp1 binding sites [(G/T)(G/A)GGC(G/T)(G/A)(G/A)(G/T)]. Both YY1 and Sp1 are known to guide transcription initiation in some TATA-less promoters [22]. Also present are putative binding sites for AP2 [(G/C)(G/C)(G/C)A(G/C)(G/C)(G/C)(G/C)] [23, 24], the Ets family of transcription factors [(C/A)GGA(A/T)] [25], c-*myc* [26], c-*myb* [(T/C)AAC(T/C)(G/C)] [27] and GCF [(G/C)CG(G/C)(G/C)(G/C)C] [28, 29].

Serial deletion analyses of the 5′-end of the human GnT II gene showed that the main promoter region is between − 636 and − 553 pb relative to the translation start codon (Table 1). An Sp1 binding site is present within this 83 bp region (Figure 3). Although Sp1 is usually associated with housekeeping genes [30], it may be involved in regulation of gene expression in the nervous system [31]. This may be of importance in view of the severe abnormality in brain development in children with carbohydrate-deficient glycoprotein syndrome type II associated with a defective GnT II gene [5, 6]. Two AP2-like elements are found in this region. AP2 has been associated with the control of differentiation [23]. A CCAAT element is present immediately upstream of the AP2 binding sites. An unusual feature of this GnT II promoter region is the presence of proto-oncogene transcription factor targeting sites for Ets proteins which have been implicated in development and differentiation [25]. Since GnT II action is required for GnT V action, it is of interest that binding sites for Ets have also been reported in the promoter region of GnT V [32]; the N-glycan branching initiated by GnT V has been related to malignant transformation [33].

In conclusion, the absence of a TATA box, the multiple transcription start sites and the GC-rich nature of the promoter region all indicate that GnT II is a typical housekeeping gene. However, some of the properties of both the 5′- and 3′-flanking regions suggest that regulation of GnT II gene expression may play a role in development and differentiation.

Acknowledgements

We thank Ms Betty Yip for help in the preparation of RNA and Ms Laura Cory, Dr Adam Chen and Dr Fabienne Vavasseur for supplying cell lines. This work was supported by a grant from the Canadian Medical Research Council to HS.

References

1 Schachter H (1994) In *Molecular Glycobiology*, (Fukuda M, Hindsgaul O, eds) pp 88–162. Oxford: Oxford University Press.
2 Field MC, Wainwright LJ (1995) *Glycobiology* **5**: 463–72.
3 Schachter H (1986) *Biohem Cell Biol* **64**: 163–81.
4 Jaeken J, Schachter H, Carchon H, Decock P, Coddeville B, Spik G (1994) *Arch Dis Child* **71**: 123–7.
5 Charuk JHM, Tan J, Bernardini M, Haddad S, Reithmeier RAF, Jaeken J, Schachter H (1995) *Eur J Biochem* **230**: 797–805.
6 Tan J, Dunn J, Jaeken J, Schachter H (1996) *Am J Hum Genet* **59**: 810–17.
7 D'Agostaro GAF, Zingoni A, Moritz RL, Simpson RJ, Schachter H, Bendiak B (1995) *J Biol Chem* **270**: 15211–21.
8 Tan J, D'Agostaro GAF, Bendiak B, Reck F, Sarkar M, Squire JA, Leong P, Schachter H (1995) *Eur J Biochem* **231**: 317–28.
9 Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1993) *Current Protocols in Molecular Biology*. New York.: John Wiley and Sons.
10 Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
11 Chomczynski P (1993) *BioTechniques* **15**: 532–5.
12 Sanger F, Nicklen S, Coulson AR (1977) *Proc Natl Acad Sci USA* **74**: 5463–7.
13 Yip B, Chen SH, Mulder H, Höppener JWM, Schachter H (1997) *Biochem J* **321**: 465–74.
14 Sheets MD, Wu M, Wickens M (1995) *Nature* **374**: 511–16.
15 Stebbins-Boaz B, Hake LE, Richter JD (1996) *EMBO J* **15**: 2582–92.
16 Wahle E, Keller W (1996) *TIBS* **21**: 247–250.
17 Bohm SK, Gum JR, Erickson RH, Hicks JW, Kim YS (1995) *Biochem J* **311**: 835–43.
18 Faisst S, Meyer S (1992) *Nucleic Acids Research* **20**: 3–26.
19 Madden MJ, Morrow CS, Nakagawa M, Goldsmith ME, Fairchild CR, Cowan KH (1993) *J Biol Chem* **268**: 8290–7.
20 Pugh BF, Tjian R (1990) *Cell* **61**: 1187–97.
21 Lopez-Bayghen E, Vega A, Cadena A, Granados SE, Jave LF, Gariglio P, Alvarez-Salas LM (1996) *J Biol Chem* **271**: 512–20.
22 Shiffman D, Brooks EE, Brooks AR, Chan CS, Milner PG (1996) *J Biol Chem* **271**: 12199–204.

23 Williams T, Admon A, Lüscher B, Tjian R (1988) *Genes Dev* **2**: 1557–69.
24 Williams T, Tjian R (1991) *Genes Dev* **5**: 670–82.
25 Wasylyk B, Hahn SL, Giovane A (1993) *Eur J Biochem* **211**: 7–18.
26 Eilers M, Schirm S, Bishop JM (1991) *EMBO J* **10**: 133–41.
27 Nakagoshi H, Nagase T, Kanei-Ishii C, Ueno Y, Ishii S (1990) *J Biol Chem* **265**: 3479–83.
28 Kageyama R, Pastan I (1989) *Cell* **59**: 815–25.
29 Angotti E, Mele E, Costanzo F, Avvedimento EV (1994) *J Biol Chem* **269**: 17371–3.
30 Rajput B, Shaper NL, Shaper JH (1996) *J Biol Chem* **271**: 5131–42.
31 Cibelli G, Schoch S, Pajunk H, Brand IA, Thiel G (1996) *Eur J Biochem* **237**: 311–17.
32 Saito H, Gu JG, Nishikawa A, Ihara Y, Fujii J, Kohgo Y, Taniguchi N (1995) *Eur J Biochem* **233**: 18–26.
33 Dennis JW, Laferte S, Waghorne C, Breitman ML, Kerbel RS (1987) *Science* **236**: 582–5.

Posted to the internet September 1996, accepted in revised form 20 June 1997