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

by

Shihao Chen

A thesis submitted in conformity with the requirements
for the Degree of Masters of Science in the
Graduate Department of Biochemistry
University of Toronto

Table of Contents

List of Figures

List of Tables

List of Abbreviations

APH	Aqueous prehybridization/hybridization solution
ARE	AU-rich element
ATP	Adenosine triphosphate
AUBP	AU-rich sequence-binding protein
bp	Base pairs
^{o}C	Degree Celsius
C2GnT	Core 2 β6 N-Acetylglucosaminyltransferase
cAMP	Cyclic adenosine monophosphate
CAT	Chloramphenicol acetyltransferase
CBP	CREB binding protein
CD	Cluster of differentiation (antigens)
CDGS II	Carbohydrate-deficient glycoprotein syndrome type II
cDNA	Complementary DNA
C/EBP	CCAAT/enhancer binding protein
CF	Cleavage factors
CFTR	Cystic fibrosis transmembrane conductance regulator
Ci	Curie
CPE	Cytoplasmic polyadenylation element
CPEB	CPE binding protein
CPSF	Cleavage and polyadenylation specificity factor
CREB	cAMP response element binding protein
CstF	Cleavage stimulation factor
CTD	Carboxyl terminal domain
dATP	Deoxyadenosine triphosphate
DMSO	dimethylsulfoxide

DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
Dol-P-Glc	Dolicol-phosphate-glucose
dpm	Disintegrations per minute
ds	Double stranded
ER	Endoplasmic Reticulum
ERCC	Excision repair cross-complementing
EtBr	Ethidium bromide
FBS	Fetal bovine serum
Fuc	Fucose
g	Gram
Gal	Galactose
GalNAc	N-Acetylgalactosamine
Glc	Glucose
GlcNAc	N-Acetylglucosamine
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GnT	N-Acetylglucosaminyltransferase
GPI	Glycosyl phosphatidyl inositol
GTF	General transcription factors
HEMPAS	Hereditary erythroblastic multinucleatity associated with a positive acidified serum lysis test
HIP1	Housekeeping initiator protein 1
HPFH	Hereditary persistence of fetal haemoglobin
IBP	Initiator binding protein
IGnT	Blood group I β6 N-Acetylglucosaminyltransferase
Inr	Initiator
IRE	Iron-responsive element

IRE-BP	IRE binding protein
kb	Kilobase
kDa	Kilodalton
l	Litre
LB	Luria-Bertani Medium
LBP	Leader binding protein
LS-180	Colonic adenocarcinoma
M	Molar
Man	Mannose
MED-1	Multiple start site element downstream
MEM	Minimal essential medium
mg	Milligram
µg	Microgram
ml	Millilitre
MOPS	3-[N-morpholino]propanesulfonic acid
mRNA	Messenger ribonucleic acid
NF-1	Nuclear factor 1
nt	Nucleotide
OD	Optical Density
ONPG	O-nitrophenyl-β-D-galactosidase
ORF	Open reading frame
PABII	Poly(A) polymerase binding protein II
PAP	Poly(A) polymerase
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
pgp 1	P-glycoprotein 1
Poly(A)	Polyadenylic acid or polyadenylate

RACE	Rapid amplification of cDNA ends
RNAP II	RNA polymerase II
RNase	Ribonuclease
RNP C	Ribonucleoprotein C
rpm	Revolutions per minute
SA	Sialic acid
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
snRNP	Small nuclear ribonucleoprotein particles
Sp1	Promoter-specific transcription factor
ss	Single stranded
SSC	Sodium chloride/sodium citrate buffer
TAFs	TATA-binding protein associated factors
TBP	TATA-binding protein
TE	Tris-EDTA buffer at pH 8.0
TEN	NaCl in TE buffer
Tris	Tris (hydroxymethyl)-aminomethane
U	Unit
UDP	Uridine diphosphate
USF	Upstream stimulatory factor
UTR	Untranslated region
UV	Ultraviolet
YY1	Ying-Yang 1

Acknowledgements

A number of persons have contributed towards the completion of this thesis and I would like to express my heartfelt thanks to them.

I wish to thank Dr. Harry Schachter, my supervisor, for his patient guidance throughout the course of the entire project, and his encouragements at times of difficulties.

I appreciate the valuable suggestions and advices from Dr. Inka Brockhausen and Dr. Rashid Anwar, members of my supervisory committee.

I would like to thank the past and present members in our lab, especially Betty Yip and Dr. Jenny Tan for the supply of essential materials and assistance in basic molecular biology techniques. Thanks are also extended to Dr. Mohan Sarkar, Dr. Leigh Revers, Dr. William J. Kuhns, Dale Toki, Ulug Ünligil, Jimmy Yang, Laura Cory, and Homa Kameh.

My special thanks are expressed to my wife for her love and support.

ABSTRACT

UDP-GlcNAc:α-6-D-mannoside β-1,2-*N*-acetylglucosaminyl-transferase II (GnT II) is essential for the normal assembly of complex Asn-linked glycans. Three GnT II transcripts, with 2.0 kb being the major signal, have been identified in five different human cell lines by Northern analysis and 3' RACE. 5' RACE results indicated multiple putative transcription initiation sites, separated over 50 bp. No untranslated upstream exon was found. It is highly likely that the utilization of distinct cleavage/polyadenylation sites leads to the production of the multiple messages. Transient transfection experiments revealed that the main promoter activity of GnT II gene was localized within the 83 bp region between -636 and -553 upstream of the ATG translation start codon. This GC-rich region contains consensus sequences suggestive of multiple binding sites for Sp1 and other transcription factors.

Chapter One

Introduction: Literature review and research objectives

Part 1.1: Glycoproteins and glycosyltransferases

Part 1.2: Transcriptional and post-transcriptional regulation of eukaryotic gene expression

Part 1.3: Objectives of the present study

1.1 Glycoproteins and glycosyltransferases

Glycosyltransferases are enzymes involved in the biosynthesis of oligosaccharides. In general, they catalyze the reaction:

$$\text{Acceptor-OH} + \text{Donor-P-Sugar} \rightarrow \text{Acceptor-O-Sugar} + \text{Donor-P}$$

The acceptor can be a free saccharide, a saccharide linked to an aglycone (e.g. a protein or lipid) or a protein or lipid. All donor monosaccharides are utilized in an activated form, either as a nucleotide sugar, e.g. GDP-mannose, or as a lipid-linked donor such as dolicol-P-glucose (Dol-P-Glc) [Schachter, 1994].

The "central dogma of glycobiology: one glycosidic linkage - one glycosyltransferase" has resulted from extensive research on the biosynthesis of the sugar-sugar linkages. This dogma is valid in most cases. A specific gene encodes a specific glycosyltransferase which in turn, synthesizes a specific sugar-sugar linkage. However, exceptions to this rule have also been reported over the past years. For example, the Lewis blood group-dependent fucosyltransferase catalyzes the synthesis of both α 1-3 and α 1-4 linkages. On the other hand, at least five distinct human α 1-3 fucosyltransferases synthesize, at least *in vitro*, the Gal β 1-4 (Fuc α 1-3) GlcNAc moiety [Schachter, 1994].

Proteins modified by glycosylation are present ubiquitously in all kinds of living forms including microorganisms, plants, animals and humans. Glycoproteins are composed of one or more carbohydrate chains attached to protein by way of either an N-glycosyl linkage (N-glycans, e.g. N-acetylglucosaminyl-asparagine) or O-glycosyl linkage (O-glycans, e.g. GalNAc-Serine/Threonine) [Schachter, 1992a]. Various other linkages

between amino acid and sugar have been reported in these two categories. An example is a glucose/GalNAc-asparagine linkage that exists in bacterial glycoproteins [Vliegenthart, 1995]. The third family of protein-sugar linkage is the glycophosphatidylinositol (GPI) anchor which has been identified in many cell surface proteins [Englund, 1993].

The complex nature of oligosaccharide structures and the complicated protein glycosylation processes explain the fact that many glycoproteins exist as a set of glycoforms. Different patterns of glycosylation can occur at either the same or different positions along the polypeptide chain, so that the carbohydrates at a particular position can differ in structure. This phenomenon is generally termed micro-heterogeneity. Factors controlling protein glycosylation include (i) the primary peptide structure which determines the number and location of potential glycosylation sites, (ii) the three dimensional structure of the protein which affects the availability of the glycosylation position, (iii) cell types which influence species and tissue specific glycosylation, and (iv) the relative activities of different glycosyltransferases and the interaction of the glycosyltransferases with the local protein structures [Dwek, 1995a].

1.1.1 N-glycans: Structure and Biosynthesis

Unlike the nucleic acid and protein macromolecules which are made via the pre-existing template molecules, the carbohydrates are highly branched and synthesized through an "assembly line" as will be discussed below. High mannose-, hybrid-, and complex-glycans constitute the family of N-acetylglucosaminyl-asparagine (GlcNAc-Asn) type N-glycans (fig1.1) [Kornfeld & Kornfeld, 1985]. The common feature of these glycans lies in

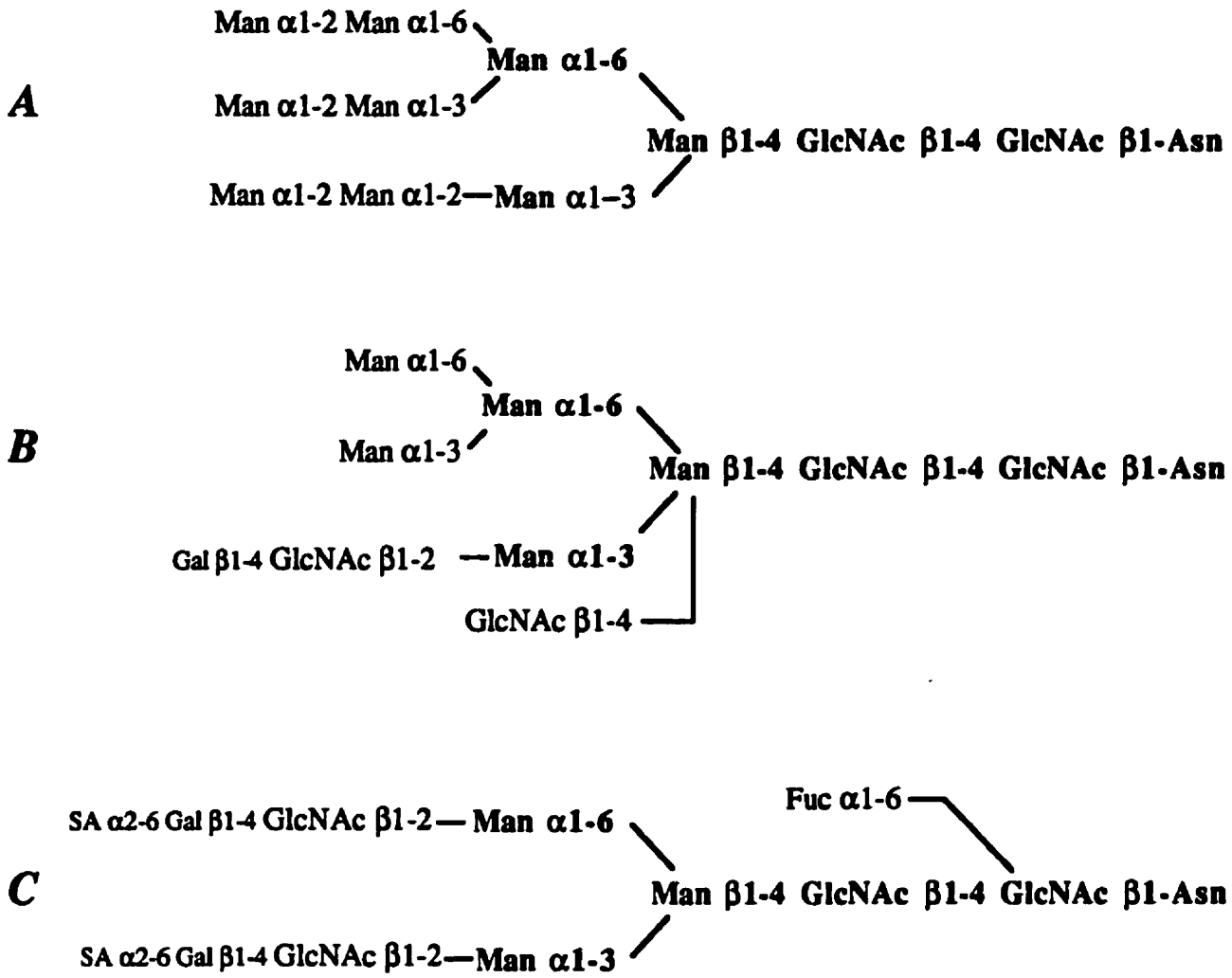


Figure 1.1 Structures of the major types of GlcNAc-Asn N-glycans

Typical primary structures of (a) high mannose, (b) hybrid and (c) complex asparagine-linked glycans. The oligosaccharides with bold face type indicate the pentasaccharide core that is common to all GlcNAc-Asn N-linked glycans. Only the Asn residue of the polypeptide is shown (far right). [Adapted from Kornfeld and Kornfeld, 1985]

that they all share a conserved core structure known as $Man_3GlcNAc_2$. Processing of N-glycans starts in the lumen of the rough endoplasmic reticulum (ER) with a co-translational 'en bloc' transfer of $Glc_3Man_9GlcNAc_2$ from the pre-assembled lipid-linked $Glc_3Man_9GlcNAc_2$ to an acceptor asparagine residue in an Asn-X-Ser/Thr glycosylation sequon of the nascent polypeptide chain (X can be any amino acid but proline) [Kornfeld & Kornfeld, 1985]. Once transferred onto the protein, the oligosaccharides are trimmed in a stepwise manner. Removal of 3 outer glucose residues results in the high mannose isoform (fig1.1). The high mannose glycan is further trimmed by ER and Golgi glycosidases to the $Man_5GlcNAc_2$ structure, which is the physiological substrate for GlcNAc-transferase I (GnT-I). In the medial Golgi, GnT-I catalyzes the addition of a GlcNAc residue from a UDP-GlcNAc donor to the Man α 1-3 arm of the $Man_5GlcNAc_2$ acceptor in a β 1-2 glycosidic linkage, initiating the synthesis of hybrid and complex N-glycans. Removal of 2 mannose residues from the Man α 1-6 arm of the above product by α-mannosidase II followed by an addition of a GlcNAc residue forms the complex N-glycan isoform. This latter step is catalyzed by GlcNAc-transferase II (GnT-II). Further modifications of the "antennae" result in a diversity of branched carbohydrate structures. The above processing pathway is summarized in figure 1.2 [Voet, 1990]. The major characteristics of this assembly line are that a series of membrane-bound glycosidases and glycosyltransferases act sequentially on the growing oligosaccharide as it moves along the lumen of the endomembrane system, and a variety of "substrate level" factors (i.e. enzyme substrate specificity, relative enzyme activities) direct the metabolic pathway at branch points. For example,

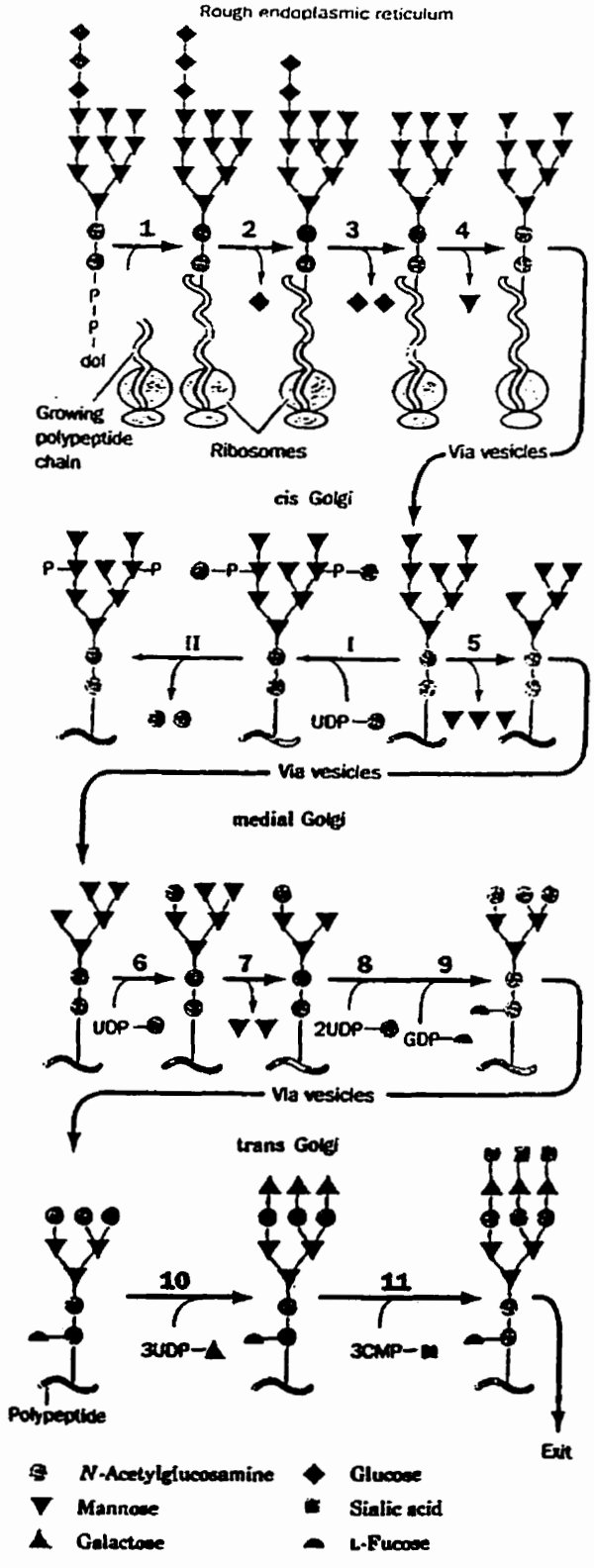


Figure 1.2 The Biosynthesis of N-linked glycans

The processing of oligosaccharide on newly synthesized glycoprotein takes place through rough endoplasmic reticulum (ER) to *trans* Golgi apparatus. The reactions are sequentially catalyzed by

(1) oligosaccharyltransferase,
(2) α-glucosidase I,
(3) α-glucosidase II,
(4) ER α-1,2-mannosidase,
(5) Golgi α-mannosidase,
(6) N-acetylglucosaminyl-transferase I (GnT I),
(7) Golgi α-mannosidase II,
(8) GnT II and GnT V,
(9) fucosyltransferase,
(10) galactosyltransferase, and
(11) sialyltransferase.
(I, II) enzymes involved in the synthesis of mannose-6-phosphate lysosome targeting signal.
[Adapted from Voet & Voet, 1990]

prior GnT-I action is essential for GnT-II, III and IV action and other processing enzymes; in turn, prior action of GnT-II is required, at least *in vitro*, for the succeeding actions of GnT-V and VI. On the other hand, if GnT-III acts on the product of the GnT I reaction before α-mannosidase II, the pathway is committed to hybrid structures because α-mannosidase II can not act on the bisected oligosaccharides, while the reverse order of action leads to the complex N-glycans [Schachter, 1995a].

1.1.2 N-glycans: Functions

In general, there is no single unifying function for protein-bound oligosaccharides. They play many different roles in biological processes [for reviews see Dwek, 1995a; Dwek, 1995b]. For example, they can modify the intrinsic properties of proteins by altering the stability, protease resistance or quaternary structure. Furthermore, the large size of oligosaccharides confers on them the ability to cover functionally important domains of proteins. The N-glycans in the adhesion domains of human CD2 on T lymphocytes and natural killer cells were shown to stabilize the interaction of CD2 with its counter receptor CD58 by counterbalancing an unfavorable clustering of positive charges centered between the CD2-CD58 interface [Wyss *et al*, 1995]. N-glycans play a critical role in the proper folding of the nascent polypeptide [Tifft *et al*, 1992] , in the sorting of protein from ER to Golgi apparatus and from the Golgi to the cell surface [Fiedler & Simons, 1995], and in the maintenance of the proper conformation of the overall mature glycoproteins [Edge *et al*, 1993]. Recently, complex N-glycans have been reported to be involved in post-implantation development of the mouse embryo [Metzler *et al*, 1994]. Knockout mouse embryos lacking a functional GnT-I gene, which catalyzes

the first committed step leading to the synthesis of hybrid and complex N-glycans, died at day 10 after fertilization with various disturbances of morphogenesis, particularly of the nervous system. Although N-glycans have diverse functions, their primary function is probably to serve as recognition markers, that is, cells "talk" with their environments by a "language" inherent with sugars. For example, circulating leukocyte trafficking into lymphatic tissues or inflammatory sites has been shown to be mediated by selectin-cell surface carbohydrate interactions [McEver *et al*, 1995; Norman *et al*, 1995; Finger *et al*, 1996].

Abnormal glycosylation is characteristic of numerous diseases, including rheumatoid arthritis, cancer and various hereditary diseases. The association of β 1-6 branches of N-linked oligosaccharides, initiated by GnT V, with metastatic potential has been demonstrated by the over-expression of GnT III. GnT III channels the biosynthetic path to the formation of a bisecting structure by competing for substrate with GnT V, and as a consequence, the metastases of tumor cells could be suppressed [Yoshimura *et al*, 1995]. The absence of, or reduction in, α-mannosidase II or GnT-II activities in a human autosomal recessive genetic disease named hereditary erythroblastic multinucleatity associated with a positive acidified serum lysis test (HEMPAS) is associated with the accumulation of N-linked hybrid oligosaccharides [Fukuda *et al*, 1987; Fukuda *et al*, 1990a; Fukuda, 1990b]. Of great interest, patients with Carbohydrate-Deficient Glycoprotein Syndrome (CDGS) type II, another related autosomal recessive disease, were shown to have no detectable GnT-II activity [Jaeken *et al*, 1994; Charuk *et al*, 1995]. Two children diagnosed with this disease displayed a severe disturbance of the nervous system [Jaeken *et al*, 1994]. DNA analysis demonstrated that the inactivation of GnT-II enzyme in these

patients was caused by a point mutation [Tan *et al*, 1996]. The complex N-glycans therefore appear to play important roles in the development and differentiation of the nervous system, although they were shown to be non-essential for the survival or proliferation of mammalian cells cultured *in vitro* [Stanley *et al*, 1975].

1.1.3 O-Glycans and GPI anchor

The major Ser/Thr-GalNAc type O-glycan-containing glycoproteins are the mucins, which contain 50-80% carbohydrate by weight. Mucins are found in mucous secretions that carry many of the blood group antigens, intercellular and molecular recognition signals, and cancer-associated and differentiation antigens [Brockhausen, 1995].

Eight O-glycan core structures, core 1 to core 8, have been discovered thus far (fig 1.3) [Schachter & Brockhausen, 1992]. The processing of O-glycans is poorly understood partly because fewer glycosyltransferases involved in the pathway have been cloned. In contrast to the N-glycosylation pathway, O-glycan formation does not require an oligosaccharide-lipid sugar donor or glycosidases. Instead, individual monosaccharides are added sequentially.

The GPI anchor is an alternative linkage between a protein and a sugar. Unlike N- and O-glycans, GPI links carbohydrates at its non-reducing end via ethanolamine to the C-terminus of a polypeptide backbone [Englund, 1993]. GPI anchored proteins have been shown to participate in transmembrane signal transduction pathways [Doering, 1993].

Core 1: Gal β1-3 GalNAc α1-O-Ser/Thr

Core 2: GlcNAc β1-6 (Gal β1-3) GalNAc α1-O-Ser/Thr

Core 3: GlcNAc β1-3 GalNAc α1-O-Ser/Thr

Core 4: GlcNAc β1-6 (GlcNAc β1-3) GalNAc α1-O-Ser/Thr

Core 5: GalNAc α1-3 GalNAc α1-O-Ser/Thr

Core 6: GlcNAc β1-6 GalNAc α1-O-Ser/Thr

Core 7: GalNAc α1-6 GalNAc α1-O-Ser/Thr

Core 8: Gal α1-3 GalNAc α1-O-Ser/Thr

Figure 1.3 Major core structures found on O-glycans
Core 1 and core 2 are most predominant in mucins. Core 3 and core 4 have only been described on mucins. Core 5 has been discovered in several tissues and species, including human meconium and in adenocarcinoma; Core 6 has been found on some human glycoproteins; Core 7 occurs as the disaccharide on bovine submaxillary mucin; and Core 8 is present on mucins from several species, including human respiratory mucins. All core structures but core 7 may be further modified. Only the Ser or Thr residues are shown.

1.1.4 UDP-N-acetylglucosamine:α-6-D-mannoside-β-1,2-N acetylglucosaminyltransferase II (GnT-II, EC 2.4.1.143)

Advances in the elucidation of complex oligosaccharide structures indicate the existence of well over 100 different glycosyltransferases (for reviews see Kleene & Berger, 1993; Narimatsu, 1994; Schachter, 1994; Field &Wainwright, 1995). The family of GlcNAc-transferases consists of GnT-I to VI as well as C2GnT and IGnT. The C2GnT and IGnT form the core 2 O-glycan branch and the blood group I antigen, respectively. GnT I - VI are responsible for the biosynthesis of highly branched (multi-antennary) complex N-glycans. Each enzyme catalyzes the transfer in a specific glycosidic linkage of a GlcNAc residue from the donor substrate UDP-GlcNAc to a mannose residue of the trimannosyl core pentasaccharide (fig 1.4) [Kornfeld & Kornfeld, 1985]. GnT-II, a medial Golgi enzyme, catalyzes the reaction shown in figure 1.5, initiating the formation of a bi-antennary structure which is an essential step in the biosynthetic pathway leading from hybrid to complex N-linked glycans [Schachter,1994].

SDS-PAGE analysis of GnT-II enzyme purified from rat liver revealed a polypeptide having a molecular weight of 42 kDa. Screening of rat liver cDNA libraries yielded a cDNA clone encoding a 442 amino acid protein with the type II transmembrane protein domain structure typical of glycosyltransferases (fig 1.6) [D'Agostaro *et al*, 1995]. A construct of a cDNA segment coding for the C-terminal 389 amino acids of rat GnT-II was expressed with high activity under the control of the Rous sarcoma virus promoter in COS-7 cells, demonstrating that the C-terminal region of the protein includes the entire catalytic domain.

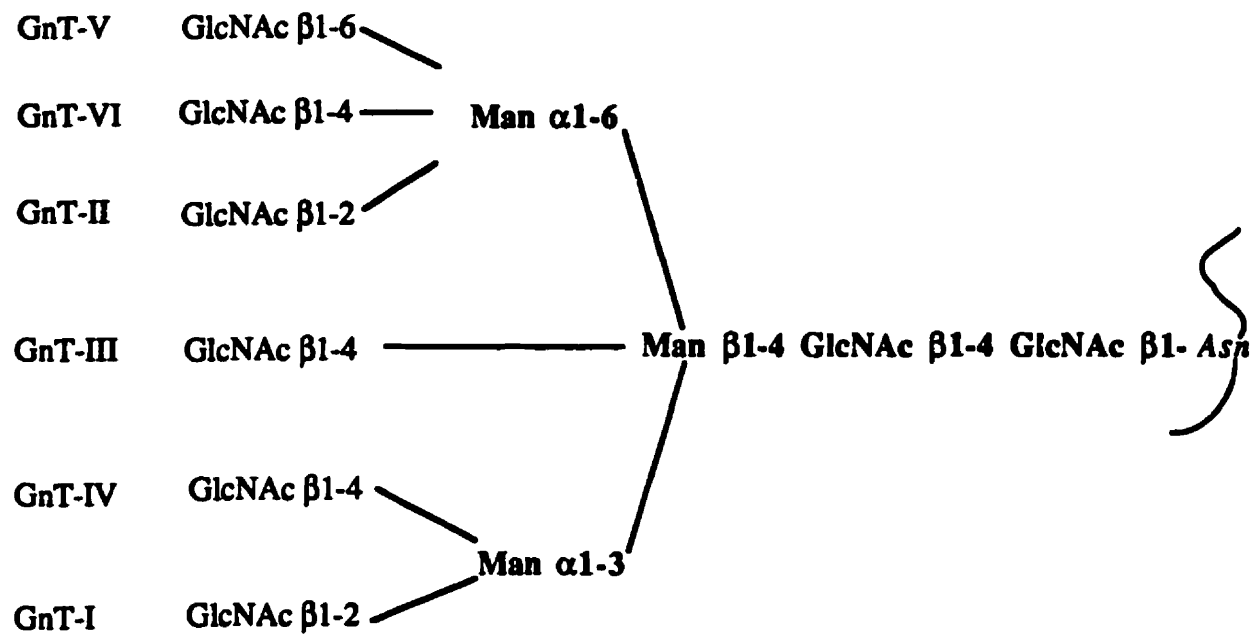


Figure 1.4 The monosaccharide attachment to the core structure catalyzed by GnT I through GnT VI

The conserved pentasaccharide core structure of N-glycans is indicated in bold. The amino acid attached to oligosaccharide is indicated in italics. GnT I, II, IV, V and VI catalyze the initiation of the antennae on the core. GnT III forms a bisecting structure which terminates further modification by GnT II, IV or V.

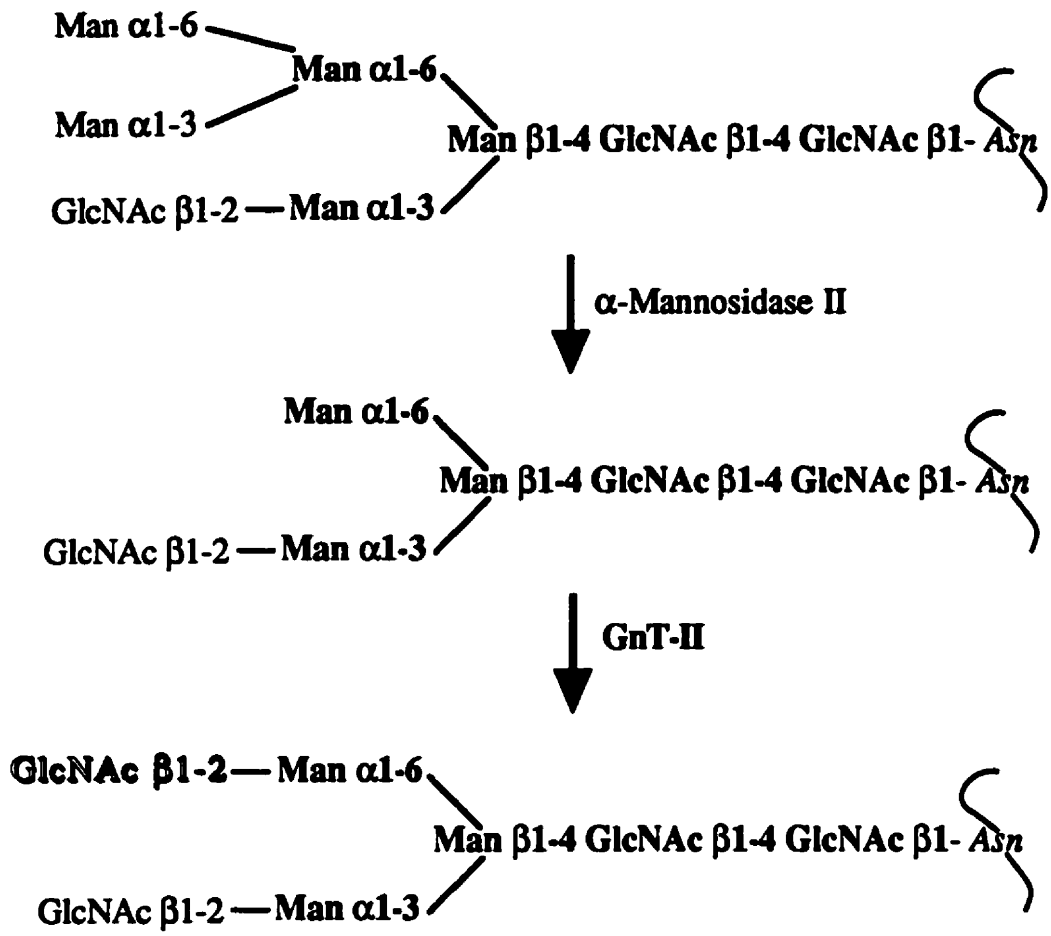


Figure 1.5 The conversion of hybrid N-glycans to complex N-glycans by the reactions of α-mannosidase II and GnT II

Two outside mannoside residues on the α1-6 arm of the hybrid N-glycan formed by GnT I are removed by the action of α-mannosidase II. The addition of a GlcNAc residue to the truncated Man α 1,6 arm by GnT II results in a complex N-glycan structure. Shown in bold is the core structure of N-glycans.

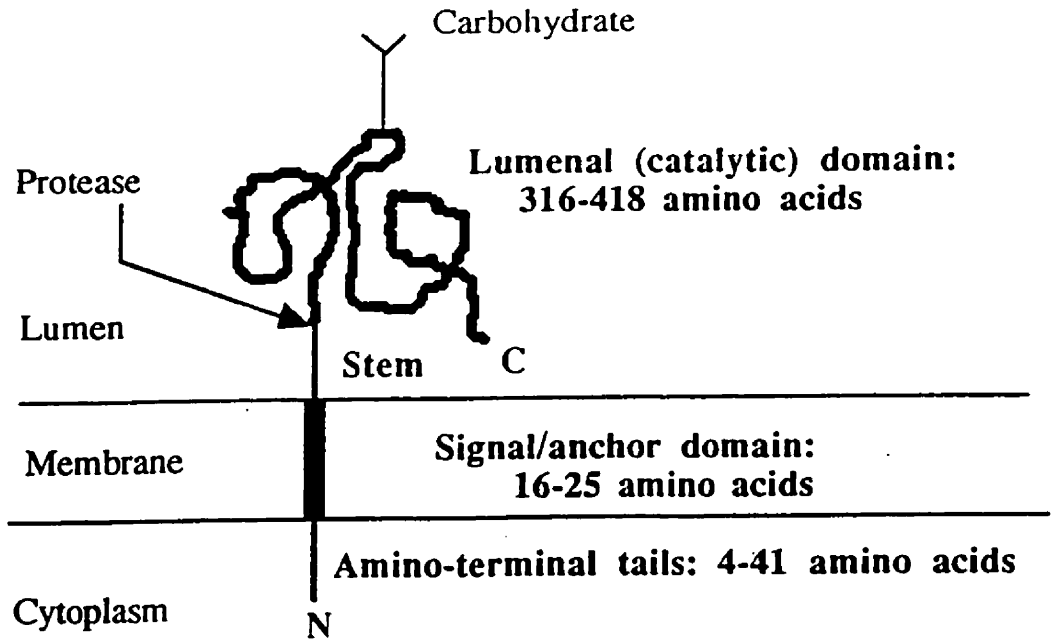


Figure 1.6 Schematic domain structure of mammalian glycosyltransferases

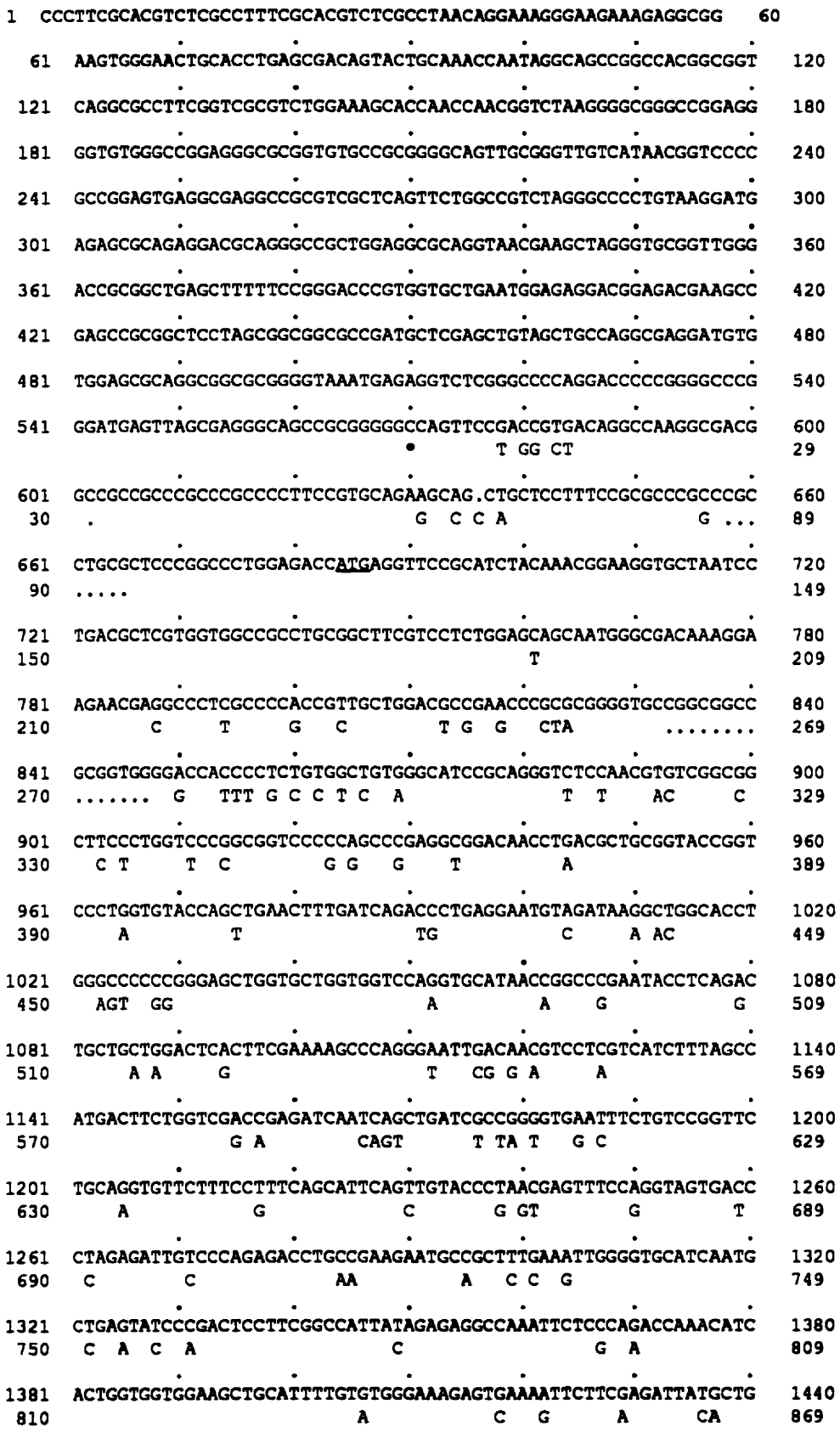
This illustration represents the typical domain structure of all mammalian glycosyltransferases cloned to date, showing the orientation of the catalytic, stem, transmembrane, and N-terminal domains. Cleavage at the stem region gives rise to a C-terminal soluble fragment with glycosyltransferase activities, indicating that the C-terminus functions as the catalytic domain and the N-terminus, transmembrane and stem domains serve as a tether and are not required for enzymatic activity. Since it is generally thought that glycosyltransferases themselves are glycoproteins, the presence of an N-glycan (Y) is also indicated.

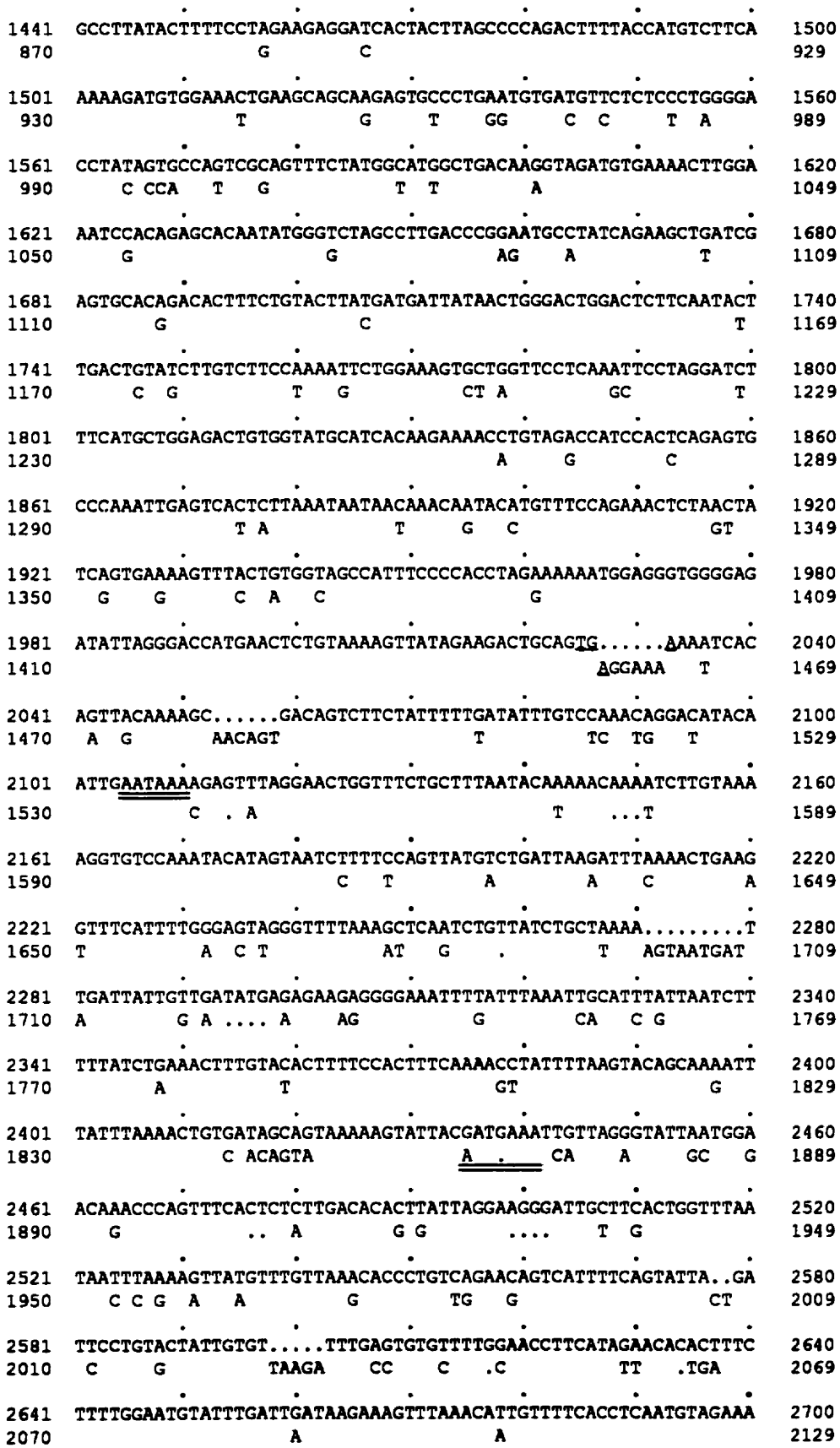
A 1.2 kb probe from a rat liver cDNA encoding GnT-II [D'Agostaro *et al*, 1995] was used to screen a human genomic DNA library in λEMBL 3. Overlapping clones encoding full length human GnT-II were obtained. The deduced human GnT II protein sequence showed the domain structure typical of all previously cloned glycosyltransferases, i.e. a short 9-residue putative cytoplasmic N-terminal domain, a 20-residue hydrophobic non-cleavable putative signal-anchor domain and a 418-residue C-terminal catalytic domain (fig 1.6) [Tan *et al*, 1995].

Southern blot analysis of human genomic DNA showed distinctive single bands after digestion with EcoR I, Hind III, BamH I, Pst I and Bgl II respectively, suggesting that there exists only a single copy of the GnT II gene in the human genome. Fluorescence *in situ* hybridization mapped the gene to chromosome 14q21, unlinked to other GlcNAc-transferases cloned to date: GnT-I gene at Chromosome 5q35 [Kumar *et al*, 1992], GnT-III at chromosome 22q13.1 [Ihara *et al*, 1993], GnT-V at 2q21 [Saito *et al*, 1994], and IGnT and C2GnT at 9q21 [Bierhuizen *et al*, 1993]. The genomic organization of mammalian glycosyltransferases can be of two types: (i) single-exonic and (ii) multiple-exonic. Human β-1,6-GlcNAc-transferase V [Saito *et al*, 1995], human and rat α-2,6-sialyltransferase, human and murine β-1,4-Gal-transferase and murine α-1,3-Gal-transferase contain coding regions with 5 or more exons [Schachter, 1995b]. Human IGnT has 3 exons for the coding region [Bierhuizen *et al*, 1995]. In contrast, the coding regions of human C2GnT [Bierhuizen *et al*, 1995], human α-1,3-Fuc-transferases III to VI, human and mouse β-1,2-GlcNAc-transferases I, and human and rat β-1,2-GlcNAc-transferases II include a single exon with no interruptions [Schachter, 1995b]. The 1344 bp open reading frame of the human GnT-II gene is flanked by a GC-rich 5'-untranslated region and

a lengthy AT-rich 3'- untranslated region containing three canonical polyadenylation signals (AATAAA). The same array of multiple poly(A) sites was also found in the rat GnT-II gene. Although GnT-II shares no sequence similarity with other glycosyltransferases cloned to date, it is highly conserved among different species (human, rat, mouse, and frog). The identity between human and rat GnT-II cDNA sequence is 89% in the coding region and 86% in the 3' untranslated region (fig 1.7). There is 92% identity between the amino acid sequences of the catalytic domains [D'Agostaro *et al*, 1995; Tan *et al*, 1995].

Northern analysis of rat GnT-II from different tissues revealed the existence of multiple transcripts. A major signal at 2.8 kb was detected in liver, brain, thymus and spleen. It was detected with probes prepared from 5' untranslated, open reading frame and 3' untranslated regions. In addition, two minor transcripts of 1.7 and 2.1 kb hybridized only to probes from the 5' untranslated and coding regions of GnT-II cDNA. The inability of two short transcripts to hybridize with 3'-untranslated region probe indicated that they might result from different 3' end processing. Although the overall levels of GnT-II messages vary in different rat tissues, the relative abundance of the major to minor messages remained unchanged [D'Agostaro *et al*, 1995]. Similarly, when probing with the coding region probe of GnT-II, multiple transcripts were also found in human heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas [Tan *et al*, 1995]. Furthermore, the transcript patterns showed no variation in relative abundance among the tissues examined. However, in contrast to rat, the major signal in human tissues was 3 kb in length with a minor mRNA at 2.0 kb. The reason for these differences are not known as


1 CCCTTCGCACGTCTCGCCTTTCGCACGTCTCGCCTAACAGGAAAGGGAAGAAAGAGGCGG 60
61 AAGTGGGAACTGCACCTGAGCGACAGTACTGCAAACCAATAGGCAGCCGGCCACGGCGGT 120
121 CAGGCGCCTTCGGTCGCGTCTGGAAAGCACCAACCAACGGTCTAAGGGGCGGGCCGGAGG 180
181 GGTGTGGGCCGGAGGGCGCGGTGTGCCGCGGGGCAGTTGCGGGTTGTCATAACGGTCCCC 240
241 GCCGGAGTGAGGCGAGGCCGCGTCGCTCAGTTCTGGCCGTCTAGGGCCCCTGTAAGGATG 300
301 AGAGCGCAGAGGACGCAGGGCCGCTGGAGGCGCAGGTAACGAAGCTAGGGTGCGGTTGGG 360
361 ACCGCGGCTGAGCTTTTTCCGGGACCCGTGGTGCTGAATGGAGAGGACGGAGACGAAGCC 420
421 GAGCCGCGGCTCCTAGCGGCGGCGCCGATGCTCGAGCTGTAGCTGCCAGGCGAGGATGTG 480
481 TGGAGCGCAGGCGGCGCGGGGTAAATGAGAGGTCTCGGGCCCCAGGACCCCCGGGGCCCG 540
541 GGATGAGTTAGCGAGGGCAGCCGCGGGGGCCAGTTCCGACCGTGACAGGCCAAGGCGACG 600
601 GCCGCCGCCCGCCCGCCCCTTCCGTGCAGAAGCAG.CTGCTCCTTTCCGCGCCCGCCCGC 660
661 CTGCGCTCCCGGCCCTGGAGACCATGAGGTTCCGCATCTACAAACGGAAGGTGCTAATCC 720
721 TGACGCTCGTGGTGGCCGCCTGCGGCTTCGTCCTCTGGAGCAGCAATGGGCGACAAAGGA 780
781 AGAACGAGGCCCTCGCCCCACCGTTGCTGGACGCCGAACCCGCGCGGGGTGCCGGCGGCC 840
841 GCGGTGGGGACCACCCCTCTGTGGCTGTGGGCATCCGCAGGGTCTCCAACGTGTCGGCGG 900
901 CTTCCCTGGTCCCGGCGGTCCCCCAGCCCGAGGCGGACAACCTGACGCTGCGGTACCGGT 960
961 CCCTGGTGTACCAGCTGAACTTTGATCAGACCCTGAGGAATGTAGATAAGGCTGGCACCT 1020
1021 GGGCCCCCCGGGAGCTGGTGCTGGTGGTCCAGGTGCATAACCGGCCCGAATACCTCAGAC 1080
1081 TGCTGCTGGACTCACTTCGAAAAGCCCAGGGAATTGACAACGTCCTCGTCATCTTTAGCC 1140
1141 ATGACTTCTGGTCGACCGAGATCAATCAGCTGATCGCCGGGGTGAATTTCTGTCCGGTTC 1200
1201 TGCAGGTGTTCTTTCCTTTCAGCATTCAGTTGTACCCTAACGAGTTTCCAGGTAGTGACC 1260
1261 CTAGAGATTGTCCCAGAGACCTGCCGAAGAATGCCGCTTTGAAATTGGGGTGCATCAATG 1320
1321 CTGAGTATCCCGACTCCTTCGGCCATTATAGAGAGGCCAAATTCTCCCAGACCAAACATC 1380
1381 ACTGGTGGTGGAAGCTGCATTTTGTGTGGGAAAGAGTGAAAATTCTTCGAGATTATGCTG 1440



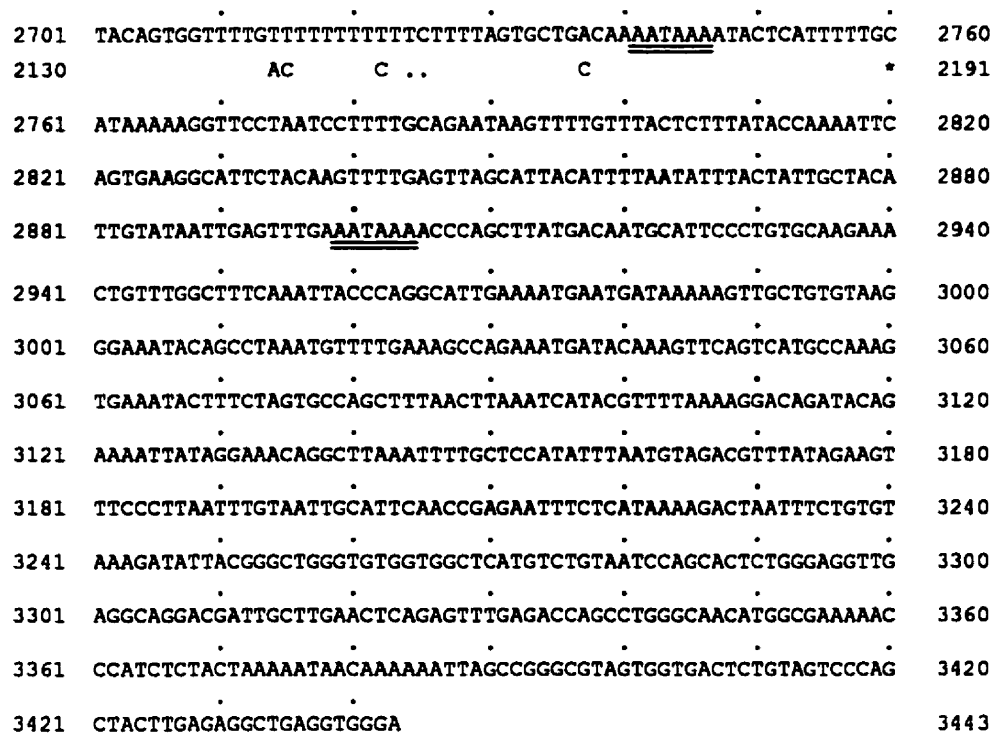


Figure 1.7 Nucleotide sequence comparison of the human and rat GnT II genes.

Human (top) and rat (bottom) GnT II nucleotide sequences are obtained from and numbered according to Genbank under accession numbers HSU15128 and RNU21662, respectively. The conserved translational initiation and termination codons are underlined. The cleavage/polyadenylation signals AATAAA are double underlined. For the sake of simplicity, the identical bases are omitted, and only different ones are shown in the rat GnT II sequence (bottom). The "*" indicates the first and last nucleotides of rat GnT II DNA sequence. The gaps, indicated with (.), were introduced in order to optimize matching.

the transcription initiation site(s) and the transcription termination site(s) of the rat and human GnT-II genes have not been identified. Little is known about the transcriptional regulation. GnT-II is probably a housekeeping gene, but it also remains to be determined whether it has additional specialized tissue-specific functions. Further the exact role(s) for GnT-II in development and differentiation are also not known.

1.2 Transcriptional and post-transriptional regulation of eukaryotic gene expression

Cells respond to environmental stimulation by modulating the extent of actively transcribing genes and by turning specific genes on or off. The initiation of mRNA transcription, which is performed by RNA polymerase II (RNAP II) and its complexes in eukaryotes, plays important roles in the regulation of gene expression. Many proteins may be involved in this process, most providing a regulatory aspect to gene transcription.

Pioneering articles have contributed to the understanding of the mechanisms by which genes are transcribed and regulated. They have been summarized in several excellent reviews which will be quoted extensively throughout the following sections.

1.2.1 Assembly of transcription initiation complex

Eukaryotic chromosomal DNAs, the physiological transcription templates, are packaged into highly ordered nucleosomal structures which affect the transcription activity by blocking the access of the transcription apparatus to promoters. Therefore, nucleosome remodeling is required prior to transcription. One mechanism to destabilize chromatin structure is charge neutralization via acetylating core histones. Nuclear histone acetyl-transferase activities have been found in components associated with RNAP II holoenzyme [Wilson *et al*, 1996; Wolffe & Pruss, 1996; Struhl, 1996].

The minimal set of protein factors necessary for accurate transcription *in vitro* include TFIIB, TFIID, TFIIE, TFIIF, TFIIH, RNAP II, and to a lesser extent, TFIIA. These proteins or complexes can assemble

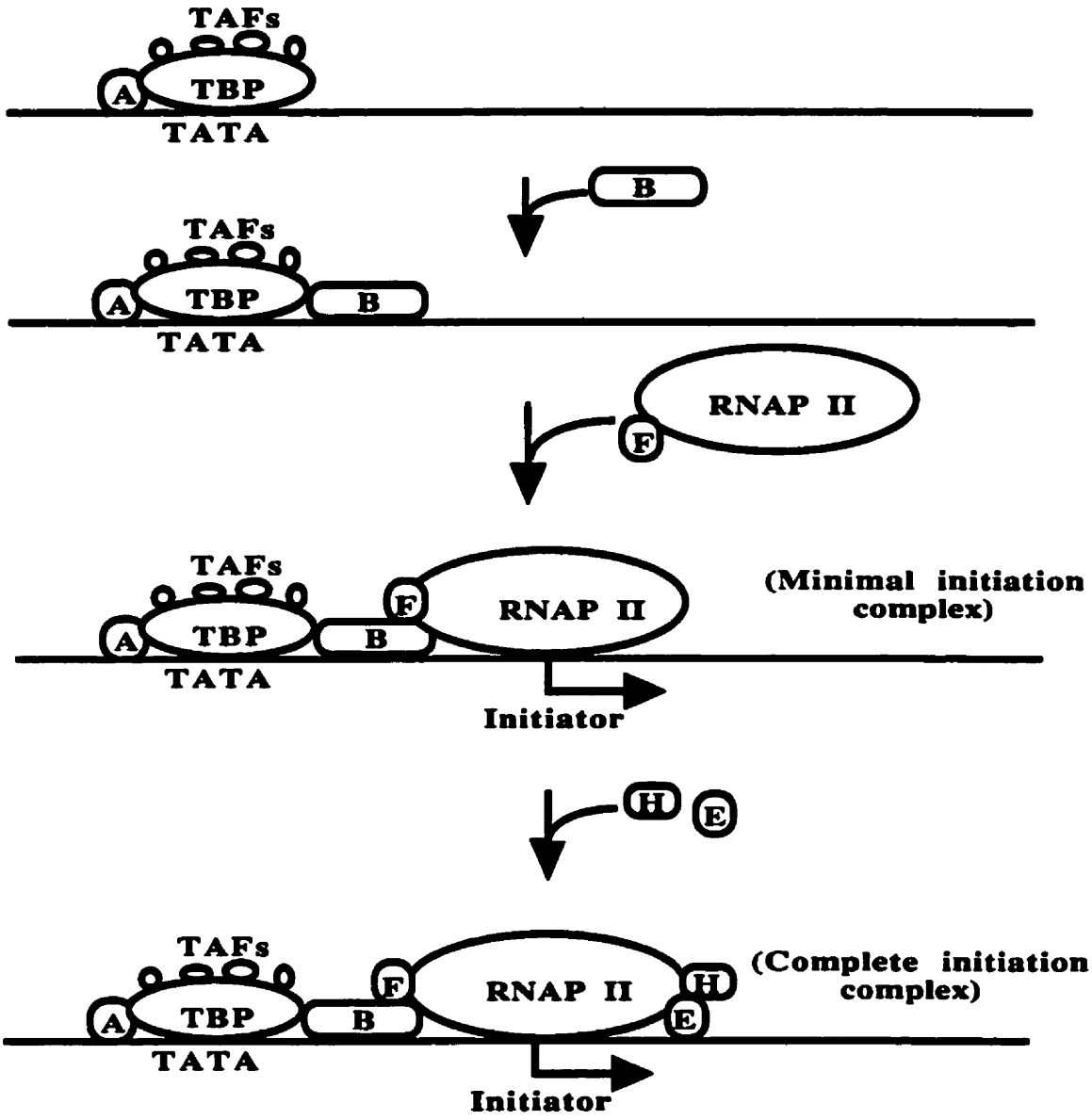


Figure 1.8 Schematic model of transcription initiation complex assembly

General transcription factors (GTFs) are represented by their letter designations (A, B, E, F, and H). TFIID consists of TBP and TAFs (see text for detail).

in a stepwise manner on a promoter (fig 1.8) [for reviews see Buratowski, 1994; Zawel & Reinberg, 1995; Struhl, 1996]. The TFIID complex contains TATA-binding protein (TBP) and the TBP-associated factors (TAFs). TBP binds specifically to the TATA sequence, rendering the TFIID complex the only general transcriptional factor (GTF) which has sequence specific recognition activity. The binding of the TBP as a molecular saddle to the promoter deforms the DNA by introducing a sharp bend and a dramatic widening of the minor groove [Werner *et al*, 1996]. The severely bent DNA accommodates the concave undersurface of the TBP saddle, leaving the entire upper surface of TBP exposed and available for protein-protein interactions. TFIIA, a coactivator mediating activated transcription, is not required for *in vitro* transcription, but has been proved to stabilize the TBP/DNA complex by contacting specifically to the limited regions (the amino-terminal stirrup) of TBP and non-specifically to the phosphate backbone of the DNA upstream of the TATA box [Geiger *et al*, 1996; Jacobson & Tjian, 1996]. This disposition of TFIIA renders itself to be recognized by other activation regulatory proteins. TFIIB binds to the carboxyl-terminal stirrup of the TBP molecular saddle, and interacts with DNA downstream of the TATA element, where it acts as a precise spacer or bridge between TBP and the RNAP II position at the transcription initiation site [Burley, 1996; Leuther *et al*, 1996]. TFIIB also interacts with TFIIF which collectively recruit RNAP II to the initiation complex [Zawel & Reinberg, 1995].

The combined activities of TFIID (TBP as well as TAFs), TFIIB and TFIIF position RNAP II stably and accurately on the promoter [Ruppert & Tjian, 1995]. However, in order to access the bases of the template strand, RNAP II requires DNA strand separation prior to initiation of

transcription. This is achieved by the actions of TFIIE and TFIIH [Zawel & Reinberg, 1995]. Incorporation of TFIIE and TFIIH leads to the complete initiation complex. TFIIE serves to sequester TFIIH to the pre-initiation complex, and further regulates TFIIH activities. TFIIH is the only basal transcription factor known to possess ATP-dependent enzymatic activities: a C-terminal domain (CTD) kinase and two DNA helicases (ERCC-2 and ERCC-3) [Guzder *et al*, 1994a, Guzder *et al*, 1994b; Zawel & Reinberg, 1995]. The DNA helicase activities of TFIIH initially open the DNA duplex in an ATP-dependent manner upstream of the transcription start site. The continuing promoter opening expands either by the action of TFIIH or by RNAP II itself [Holstege *et al*, 1996]. The CTD kinase of TFIIH specifically phosphorylates the carboxyl-terminal domain of the largest subunit of RNAP II leading to the dissociation of RNAP II from TBP subunit of TFIID [Drapkin & Reinberg, 1994a; Buratowski, 1994; Drapkin *et al*, 1994b]. This converts the initiation complex into an active elongation form capable of initiating transcription, an event known as promoter clearance. The phosphorylated RNAP II form catalyzes RNA chain elongation. Recently, the discovery of RNAP II holoenzyme containing all components required for promoter-specific transcription initiation raised an alternative way to initiate transcription, that is, via recruitment of an integral holoenzyme to the promoter [Koleske & Young, 1995; Ossipow *et al*, 1995].

A series of experiments have found that not all genes transcribed by RNAP II contain a TATA element in their promoter regions, and therefore, can not form a transcription initiation complex as discussed above. In these cases, an initiator sequence (Inr) instead of the TATA box positions RNAP II in the promoter site [for review see Weis & Reinberg,

1992]. The Inr element was first identified in the TATA-less terminal deoxynucleotidyl transferase gene promoter [Smale & Baltimore, 1989], and has since then been found in many promoters, especially housekeeping gene promoters [Weis & Reinberg, 1992; Pang *et al*, 1996; Shiffman *et al*, 1996; Vihinen *et al*, 1996]. Initiators and Inr-like elements are not well conserved although some of them show some sequence homology. The initiator is a short pyrimidine-rich element (YYANYYY) that encompasses the transcription start sites. However, the positions of Inr-like elements are not necessarily fixed and may be found distal to the initiation sites [Nakatani *et al*, 1990, Lo & Lau, 1996]. The TATA-mediated and Inr-mediated initiation complex assembly pathways are not necessarily exclusive, as it has been shown that Inr can direct complex formation even in the presence of the TATA motif, and that the Inr and the TATA motifs function synergistically [Carcamo *et al*, 1991; Roeder, 1991; Weis & Reinberg, 1992]. Several models have been proposed to explain the formation of an initiation complex mediated by the Inr. First, RNAP II alone has a low intrinsic affinity for Inr. Through the interactions with GTFs (e.g. TBP, TFIIB and TFIIF), RNAP II can recognize and form a stable complex capable of transcription initiation on the Inr, as shown for the promoters of the adenovirus ML and IVa2 genes [Weis & Reinberg, 1992]. A second model comes from the finding that $TAF_{II}150$ could recognize and bind to sequences overlapping the initiator of the Ad2 major late promoter [Verrijzer *et al*, 1994; Verrijzer *et al*, 1995]. The interaction of $TAF_{II}150$ with Inr recruits other GTFs and RNAP II to the promoter region. The identification of a number of transcription factors capable of recognizing the Inr element gives rise to a third model. The Inr binding protein (IBP, TFII-I, USF, YY1, HIP1, NP-1 and NP-D) binds specifically

to Inr elements, creating the binding sites for GTFs and RNAP II [Roy *et al*, 1991; Weis & Reinberg, 1992; Zenzie-Gregory *et al*, 1993; Pang *et al*, 1996; Goodrich *et al*, 1996]. None of these three models are mutually exclusive, and in all cases TFIID is required and TBP is anchored to the initiation complex through protein-protein interactions. Thus, TFIID contains at least two DNA-binding proteins, TBP for the TATA box and $TAF_{II}150$ for initiator.

Many TATA-less promoters initiate transcription at multiple start sites. The selection of many start sites could be a random or default process due to the lack of the TATA-box, or each start site may be independently regulated by an individual initiator-like motif. The latter possibility was ruled out based on the investigation of the thymidylate synthase promoter which initiated at multiple sites but lacked multiple initiator sequences [Geng, 1993]. A survey of several promoters with multiple start sites implicated the existence of a downstream protein-binding sequence, Multiple start site Element Downstream (MED-1) within a maximal distance of ~100 bp downstream of the 5'-end but within 20-45 bp downstream of the transcription initiation window [Ince & Scotto, 1995]. The conserved sequence is GCTCC(C/G). The activity of the MED-1 in the P-glycoprotein 1 (pgp1) promoter was further confirmed by mutations which dramatically reduced pgp1 expression [Ince & Scotto, 1995]. It has been concluded that this conserved downstream element represents a novel family of RNAP II promoters. However, how this motif recruits GTFs and RNAP II to the promoter region remains unclear.

1.2.2 Transcriptional activation

The assembly of a functional initiation complex at the promoter region is the first step in transcription of active genes. Moreover, transcription from natural RNAP II promoters is tightly controlled by the combined actions of positive and negative regulatory factors which have the ability to recognize specific sequences located upstream or downstream of the core promoter region [for reviews see Tjian & Maniatis, 1994; Buratowski, 1995]. These site-specific DNA binding proteins (specific transcription factors) can regulate the levels of gene expression and direct tissue- and/or developmental-stage-specific transcription [Orkin, 1995]. Transcriptional activators are modular in character, typically containing a functionally separable DNA targeting domain that recognizes specific sequences within its binding motif and an activation domain that is required for transcription stimulation [Tjian & Maniatis, 1994]. Transcription activation can occur at different levels. Several lines of evidence indicate that activation of some enhancer elements appears to require the assembly of a highly specific three-dimensional nucleoprotein complex among transcription activators -- a stereospecific complex [Tjian & Maniatis, 1994]. However, it is highly likely that a direct contact between an activator and GTFs represents an early step in the process of transcriptional activation. In this case, activator can function by recruiting a GTF to the promoter directly, by altering the conformation of the preformed pre-initiation complex which in turn facilitates the binding of other GTFs, or by stimulating covalent modification of the GTFs which subsequently affects the extent of transcription at the target promoters [Zawel & Reinberg, 1995; Goodrich *et al*, 1996]. Transcription activation can also occur through the protein-protein interaction between activators and TAFs. For example, the glutamine-rich domain of TAF$_{II}$110 has been

shown to bind selectively to the glutamine-rich activation domain of Sp1, and $TAF_{II}40$ is able to interact with the C-terminal domain of VP16 [Tjian & Maniatis, 1994]. Furthermore, multiple TAFs have been shown to be required for the synergistic activation of transcription by transcription activators [Buratowski, 1995; Sauer *et al*, 1995a; Sauer *et al*, 1995b]. Finally, an activator bound to its cognate site can direct RNAP II holoenzyme to the promoter region [Struhl, 1996].

Although activators act primarily at the transcription initiation level as described above, it has been demonstrated that activators can also enhance transcriptional elongation or processivity [Blair *et al*, 1996]. Efficiency of transcription processivity increases proportionally to the quantity rather than the quality of activators targeted to the promoter. On the other hand, transcription initiation is detectable even with low levels of activators.

1.2.3 Processing of pre-mRNA 3' end

The primary transcripts (pre-mRNAs) must be processed prior to translation. Distinguishing features of eukaryotic gene transcription are modification of pre-mRNA by 3'-end processing and other modifications such as 5'-capping, internal methylation, splicing and transport from the nucleus to the cytoplasm. For example, pre-mRNA introns are spliced out and exons are ligated by the spliceosome, a large complex composed of numerous small nuclear ribonucleoprotein (snRNP) particles and other factors [Ast & Weiner, 1996; Mount, 1996]. The addition of poly(A) tails to the upstream RNA fragment by template-independent poly(A) polymerase (PAP) following the endonucleolytic cleavage of the pre-mRNA has important roles in mRNA transport from nucleus to cytoplasm

and in mRNA stability and translation [Jackson & Standart, 1990; Eckner *et al*, 1991; Sachs, 1993a; Sachs & Wahles, 1993b; Beelman & Parker, 1995; Dehlin *et al*, 1996]. It has been shown that splicing and polyadenylation may be functionally coupled [Gunderson *et al*, 1994; Lutz *et al*, 1996]. Moreover, the termination of transcription by RNAP II is tightly coupled to the assembly of a functional 3'-processing complex on the nascent RNA, thus preventing premature termination and ensuring the synthesis of full-size pre-mRNAs [Keller, 1995]. Surprisingly, adenylation has been recently reported in prokaryotic RNA and affects transcript stability [Cohen, 1995].

The cleavage and polyadenylation of eukaryotic pre-mRNA require an invariantly conserved hexamer signal AAUAAA, located 10-30 nucleotides upstream of the cleavage and subsequent site for A addition followed by a U- or GU- rich flanking region [McLauchlan *et al*, 1985; Guo & Sherman, 1995; for reviews see Wahle & Keller, 1992; Keller, 1995; Wahle, 1995a (the following description is mainly based on these reviews)]. Biochemical reconstitution studies have identified six separable factors that are required for efficient polyadenylation *in vitro*. These factors form the basal polyadenylation machinery in mammalian cells. The cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF) function synergistically to bind the pre-mRNA as the first step in specifying the poly(A) site (fig 1.8). CPSF is a multisubunit protein that plays a central role. It binds specifically to the AAUAAA signal through its 160 kDa subunit (and perhaps also via its 30 kDa subunit), a process critical to the formation of polyadenylation complex [Murthy & Manley, 1995]. PAP, by itself, has a low affinity for RNA and is incapable of recognizing a pre-mRNA specifically. Binding of CPSF to substrate RNA activates the poly(A) polymerase by holding it onto the

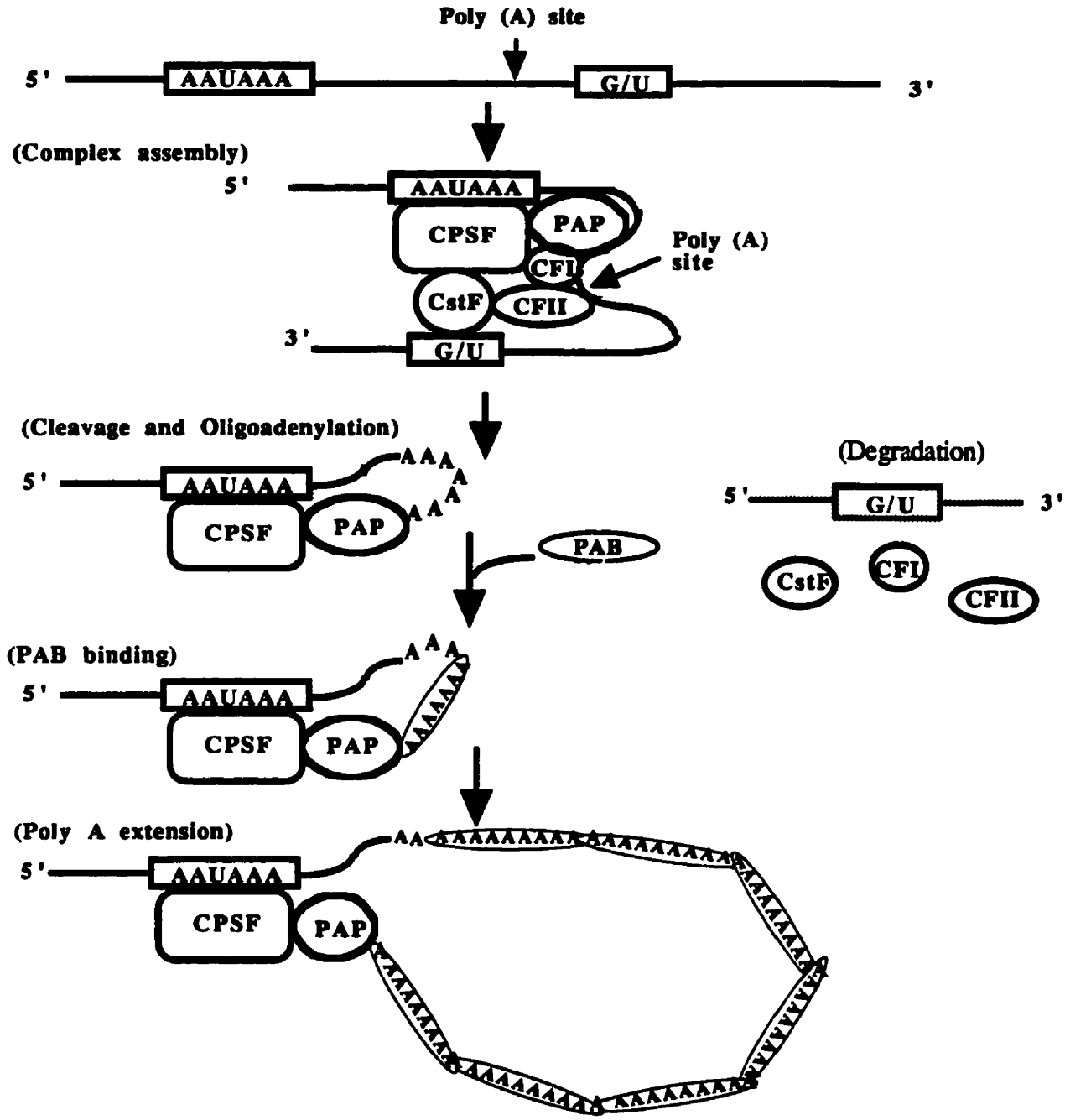


Figure 1.9 Model of the 3'-processing complex assembly

Major reaction steps and factors involved in the mammalian pre-mRNA 3' end processing are shown in this model (see text for details) [Adapted from Keller, 1995].

mRNA via a protein-protein interaction. The binding of the heterotrimeric CstF to the G/U-rich downstream sequences is mediated by its 64 kDa subunit through an RNP-type RNA-binding domain [Murthy & Manley, 1995]. Although CPSF and CstF interact separately with RNA very weakly, a strong cooperative effect has been observed through the interaction of the 160 kDa subunit of CPSF with the CstF 77 kDa subunit. As a result, CPSF and CstF, together with substrate pre-mRNA, form a stable complex conferring specificity to the 3' end processing reaction. Recruitment of two cleavage factors (CFI and CFII) to the complex leads to the cleavage of the RNA in an ATP-dependent fashion at the poly(A) site. However, the endonuclease activity has not been identified. CFI has also been shown to stabilize the binding of CPSF to pre-mRNA [Rüegsegger *et al*, 1996]. The cleavage by endonuclease results in an upstream fragment terminating in a 3'-OH and a downstream fragment starting with a 5'-phosphate. The latter is rapidly degraded and CstF and CFs leave the 3'-processing complex. Poly(A) polymerase bound to the upstream RNA then synthesizes a short tract of poly(A) of about ten adenosine residues.

While the addition of these ~10 nucleotides is thought to occur with low processivity due to the low affinity of PAP for RNA even with CPSF, a more efficient elongation of poly(A) tails is then stimulated by the incorporation of poly(A) polymerase binding protein II (PABII) [Wahle, 1995a; Keller, 1995]. Binding of PABII to the oligo(A) tail causes a rapid burst (typically 300-fold increase) of processive synthesis of a poly(A) tail of approximately 250 nucleotides. Consistent with the rapid processivity, the quaternary complex of pre-mRNA, CPSF, PAP, and PABII is more stable than the ternary complex formed in the absence of PABII. However, further elongation of poly(A) tails beyond 250 nucleotides also becomes

sluggish. It has been suggested that the length of the poly (A) tail is controlled by the number of PABII molecules in the polyadenylation complex (approximately one PABII per 30 nucleotides) and by the simultaneous stimulation of PAP by CPSF and PABII [Wahle, 1995b]. Once ~250 A residues have been polymerized, it was proposed that the polyadenylation complex was disrupted and processive polyadenylation was terminated. The nature of this event, however, is unclear.

1.3 Objectives of the present study

GnT II initiates the biosynthesis of a bi-antennary carbohydrate structure, an essential step in converting high mannose and hybrid N-glycans to complex N-glycans. Since complex N-glycans are distributed extensively in eukaryotes, GnT II is probably a housekeeping gene.

Multiple transcripts of GnT II have been detected in rat and human tissues. Since the open reading frame of GnT II gene is organized in a single exon without interruption by introns, both in the rat and human genomes, the differences between the multiple transcripts must lie in either the 5' UTR or 3' UTR, or both. The existence of multiple canonical polyadenylation sites in the 3' UTR makes it possible that transcription terminates at more than one distinct site. In contrast to the major long rat transcript (2.7 kb), the two minor short transcripts (1.7 kb and 2.1 kb) do not hybridize with a genomic probe containing only the second and third polyadenylation signals (AAUAAA), suggesting some difference in their 3' UTRs [D'Agostaro *et al*, 1995]. By the length distinction, this finding also implies that these shorter transcripts both terminate at the first polyadenylation signal, and therefore, may differ in their 5' UTRs. However, this hypothesis is currently not supported by the finding that only two transcription initiation sites at ~450 and 435 bases upstream of the translation initiation codon (AUG) have been detected by RNase protection experiments [D'Agostaro *et al*, 1995]. The reported 3' end of the major long rat transcript is also currently unreliable because of the absence of a terminal poly(A) tract in the cDNA clones analyzed [D'Agostaro *et al*, 1995]. Furthermore, even less is known about the human GnT II transcripts. Due to the difficulties involved in obtaining fresh human tissues, RNAs from cultured human cell lines were used in this project to

examine the mechanism(s) by which human GnT II transcription is regulated. The following objectives were undertaken:

(1) To survey the expression of GnT-II in different human cell lines by Northern blot analysis.

(2) To determine the transcription termination site(s) of human GnT-II by 3' Rapid Amplification of cDNA Ends (3' RACE).

(3) To search for the transcription initiation site(s) of human GnT-II by 5' RACE.

(4) To determine the promoter activities of a series of 5' genomic fragments by reporter gene (CAT) analysis.

Chapter Two

Materials and methods

2.1 Northern analysis of GnT II in human cell lines

All GnT II gene specific primers used in this thesis are numbered relative to the position of the first nucleotide of the translation initiation codon ATG, defined as +1. "f" denotes forward primer and "r" denotes reverse primer. All primers were synthesized on a Pharmacia DNA synthesizer and purified by a reversed phase chromatography cartridge method (Hospital for Sick Children-Pharmacia Biotechnology Center).

2.1.1 Preparation of DNA hybridization probe

A DNA fragment that could be used as a probe from the 5'-end of the GnT II open reading frame (ORF) was prepared by PCR as follows; 3.08 ng of plasmid pHG36 (provided by Jenny Tan in our lab, fig 2.1) was amplified in a total volume of 50 µl using primers P119f (5' CGT TGC TGG ACG CCG AAC CCG 3') and P625r (5' CCA ATT TCA AAG CGG CAT T 3') in the presence of 1x PCR buffer (Perkin-Elmer), 10% DMSO (Fisher Biotech), 1.5mM MgCl2, 50µM dNTP (Perkin-Elmer), and 1.5 U Taq polymerase (Perkin-Elmer). Hot start PCR (including incubation of the reaction mixture at 80°C for 5min prior to adding Taq enzyme) was performed with 1 cycle of 94°C for 5min, 55°C for 1min, 72°C for 1min, followed by 34 cycles of 94°C for 1min, 55°C for 1min, 72°C for 1min, and a final extension at 72°C for 15min. The resulting 507 bp product was purified from a 1.4% agarose gel (Mandel) using the GeneClean II kit (Bio 101 Inc), according to the recommended protocol.

The purified DNA fragment, ~50 ng, was labeled at 37°C for 20min in the presence of 1x Reagent mix, 1U T7 DNA polymerase of the T7 Quick Primer Kit (Pharmacia), and 50µCi [α-^{32}P]dCTP (Amersham). The

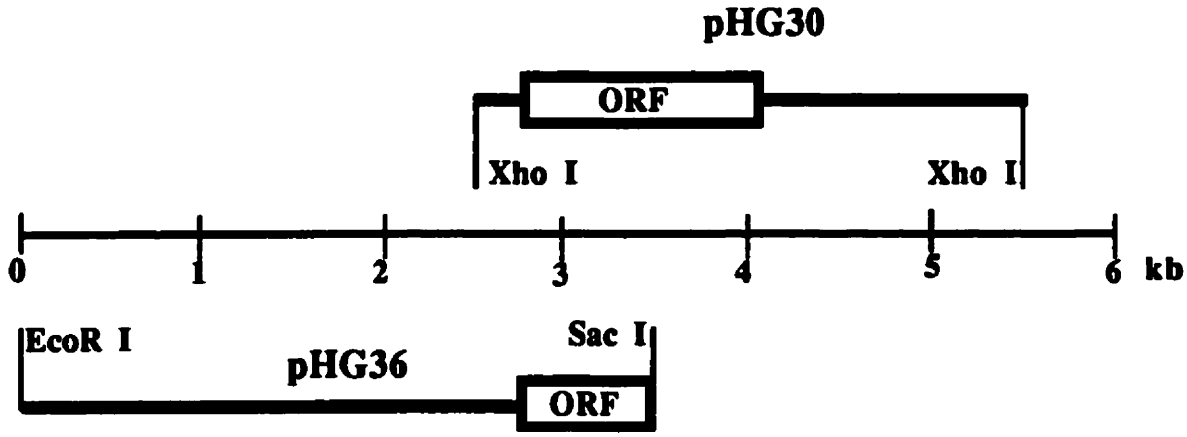


Figure 2.1 Maps of plasmids pHG30 and pHG36
pHG30 is a plasmid that contains a 3.0 kb fragment of human genomic DNA subcloned into the Xho I site of pBlueScript II KS (+) vector. Subcloning of a human genomic 3.5 kb Sac I/EcoR I DNA fragment into pBlueScript II KS (+) vector yielded pHG36. pHG30 and pHG36 overlap. The plasmid segments are not shown. pHG30 contains the entire GnT II ORF (1341 bp), 250 bp of 5' untranslated region (UTR) and 1.4 kb of 3' UTR. pHG36 includes 2.75 kb of 5' UTR and 750 bp 5' of ORF. The maps are drawn to scale, as indicated. [Adapted from Tan *et al*, 1995]

^{32}P-labeled DNA was purified by passing the reaction mixture through a Nick^{TM} column (Pharmacia), and finally eluted in 400µl of TE (pH 8.0) buffer.

2.1.2 Separation of RNAs by electrophoresis

Total RNAs from human cell lines LS-180 (colonic adenocarcinoma), HepG2 (hepatocellular carcinoma), HeLa (cervical epitheloid carcinoma), HL-60 (promyelocytic leukemia) and A431 (epidermoid carcinoma) were provided by Betty Yip in our lab. 20 µg of each RNA sample, together with RNA standards (Gibco BRL), was separated by electrophoresis in a formaldehyde agarose gel (1% agarose, 10% formaldehyde) running in 1x MOPS buffer at 30v/cm, 4°C for 24 hours. RNA was transferred by capillary action to Hybond-N membrane (Amersham) according to the manufacturer's specification. The RNAs were UV-cross-linked by illuminating the membrane on a transilluminator (long wave) for 5min.

The quality of RNA was confirmed by staining the filter with 0.02% methylene blue (Fisher).

2.1.3 Hybridization

The membrane was probed with 1.5 x 10^6 cpm/ml of the above ^{32}P-randomly-primed probe. Hybridization was performed in 5x SSPE, 5x Denhardt's solution, 0.5% SDS, 50% formamide at 42°C overnight. The membrane was then washed once for 30min duration with 2x SSC, 0.1% SDS at room temperature and three times for 20min durations with 0.2x SSC, 0.1% SDS at 50°C. After the final washing, the membrane was sealed

in a plastic bag and exposed to X-ray film for 5 days at -70°C, in the presence of intensifying screens.

2.2. 5' and 3' RACE

Marathon cDNA amplification kit was purchased from Clontech Laboratories Inc., and all procedures were performed as recommended. Briefly, 1μg of total RNA from the LS 180 cell line was reverse transcribed at 42°C for 1 hour with 200U MMLV reverse transcriptase (RNAse H^-) and the oligo(dT) cDNA synthesis primer (Table 1) in the presence of 1x First-strand buffer, 1mM dNTPs, and 1μCi [α-^{32}P]dCTP (total volume: 10μl). The second-strand cDNA was synthesized at 16°C for 2 hours by the Second-strand enzyme cocktail (containing RNase H, *E. coli* DNA polymerase I, and *E. coli* DNA ligase) in the presence of 1x Second-strand buffer and 200μM dNTP (final volume: 80μl). The resulting double stranded cDNA was blunted by addition of 1μl of 5U/μl T4 DNA polymerase with incubation at 16°C for 30min. The reaction was terminated by adding 4μl of 0.2M EDTA. ds-cDNAs were extracted by phenol-chloroform-isoamyl alcohol (25:24:1, Gibco BRL) three times at room temperature, and precipitated with addition of 1/2 volume of 4M ammonium acetate and 2.5 volumes of 95% ethanol. After centrifugation at ~14,000 rpm at room temperature for 30min, the DNA pellet was washed once with 400μl 80% ethanol. The incorporated radioactivity was assessed to confirm successful cDNA synthesis. The air dried pellet was resuspended in 10μl of H_2O. Finally, the ligation of the cDNA adapter (Table 1) to both ends of ds-cDNAs was catalyzed at 16°C overnight with 1U T4 DNA ligase with 1x DNA ligation buffer in the presence of 2μM Marathon cDNA adapter.

2.2.1 Southern blot analysis

ds cDNA of GnT II gene was amplified by a gene specific primer P900f and the cDNA synthesis primer (table 1). The PCR product was subjected to electrophoresis in a 1.2% agarose/ethidium bromide (EtBr) gel in 1x TAE buffer. Southern blot transfer was conducted according to the protocol specified for Hybond N^+ nylon membrane (Amersham). Hybridization and washing were performed according to the standard protocols [Current Protocols in Molecular Biology, Wiley Interscience & Sons Inc]. Briefly, the membrane was pre-hybridized for >4 hours at 65°C in the APH solution (6x SSC, 5x Denhardt's, 1% SDS, and 1 mg denatured salmon sperm DNA), and hybridized at 65°C overnight with a GnT-II probe covering 420 bp of the 3' portion of the open reading frame and 271 bp of 3' UTR. The filter was sequentially washed twice with 2x SSC (0.1% SDS) at room temperature for 15min durations, then once with 1x SSC (0.1% SDS) at 55°C for 30min duration, then once with 0.1x SSC (0.1% SDS) at room temperature for 30min duration, then once with 0.1x SSC (0.1% SDS) at 42°c for 30min duration, and finally twice with 0.1x SSC (0.1% SDS) at 65°C for 30min durations. The filter was exposed to X-ray film overnight at room temperature.

The probe was prepared by PCR amplification in the presence of 50μM of each dATP, dGTP and dTTP, 10μM of cold dCTP and 30μCi of [α-^{32}P]dCTP using template plasmid pHG30 (fig 2.1) and the primer sets P922f and P1612r (5' AAA ATT TCC CCT CTT CTC TC 3'). Amplication reactions were carried out as indicated in section 2.1.1.

2.2.2 3' RACE

In 3' RACE, GnT-II cDNA was amplified by PCR with a forward gene specific primer, P900f (Table 1) and a reverse adapter primer AP1. This primary 3' RACE product was re-amplified using nested adapter primer (AP2) and P922f under the conditions: 1 cycle of 94°C for 5min, 50°C for 1min, 72°C for 2min; followed by 34 cycles of 94°C for 1min, 50°C for 1min, 72°C for 2min, and a final extension at 72°C for 15min. The primary 3' RACE PCR products were also subjected to further amplification by P1888f-AP2 and P2109f-AP2, respectively (Table 1).

2.2.3 5' RACE

In 5'RACE, adapter primers and reverse GnT-II specific primers were used to amplify the GnT-II 5' end. The first round of PCR was performed using AP1 and P21r, as indicated in section 2.2.2. This primary 5' RACE PCR product was subsequently amplified twice with primers EcoR I-AP2 (AP2 with an EcoR I site at the 5'-end) and P-47r (Table 1).

2.3 Subcloning of RACE PCR products

All PCR products were purified from agarose gel using the GeneClean II kit (Bio 101 Inc). The purified products and plasmids [pGEM5Zf(+) or pGEM7Zf(+)] (Promega, fig 2.2) were digested with the same combination of restriction enzymes. In the case of subcloning the P922f-AP2 3' RACE DNA fragment (the Pst I site at 1336 was used), the purified PCR product and plasmid pGEM5Zf(+) were sequentially digested with Pst I and Not I (Pharmacia). For subcloning other PCR DNAs, both the PCR products and pGEM7Zf(+) were digested with EcoR I only (Pharmacia). The cohesive ends of the PCR product and plasmid were ligated at 16°C overnight by T4 DNA ligase (Gibco BRL). This ligated

Non-specific primers used in mRNA reverse transcription and cDNA adapter ligation

Marathon cDNA synthesis primer (52-mer):

5′ TTC TAG AAT TCA *GCG GCC GC* $(T)_{30}N_{-1}N$ 3′

Marathon cDNA adaptor:

5′ CTAATACGACTCACTATAGGGCTCGA*GCGGCCGC*CCGGGCAGGT 3′

3′ H_2N-CCCGTCCA-PO_4 5′

Non-specific primers used in both 3' and 5' RACE

AP1 5′ CCA TCC TAA TAC GAC TCA CTA TAG GGC 3′

AP2 5′ ACT CAC TAT AGG GCT CGA GCG GC 3'

EcoR I-AP2 5' CGC GGA ATT C-AP2

Gene-specific primers used in 3' RACE

P900f 5' CTA TGG CAT GGC TGA CAA GGT AGA 3'

P922f 5' GAT GTG AAA ACT TGG AAA TC 3'

P1888f 5' CGC GGA ATT CGT GTT TTG AGT GTG TTT TGG 3'

P2109f 5' CGC GGA ATT CAG TGA AGG CAT TCT ACA AG 3'

Gene-specific primers used in 5' RACE

P21r 5' TTT GTA GAT GCG GAA CCT CA 3'

P-47r 5' CGC GGA ATT CGC TGC TTC TGC ACG GAA 3'

Table 1 A summary of primers used in RACE

The EcoR I sites in gene specific or non-specific primers are indicated by underlining, Not I sites are indicated in italics. In Marathon cDNA synthesis primer, N_{-1} denotes a G, A, or C, while N denotes any of the four dNTPs.

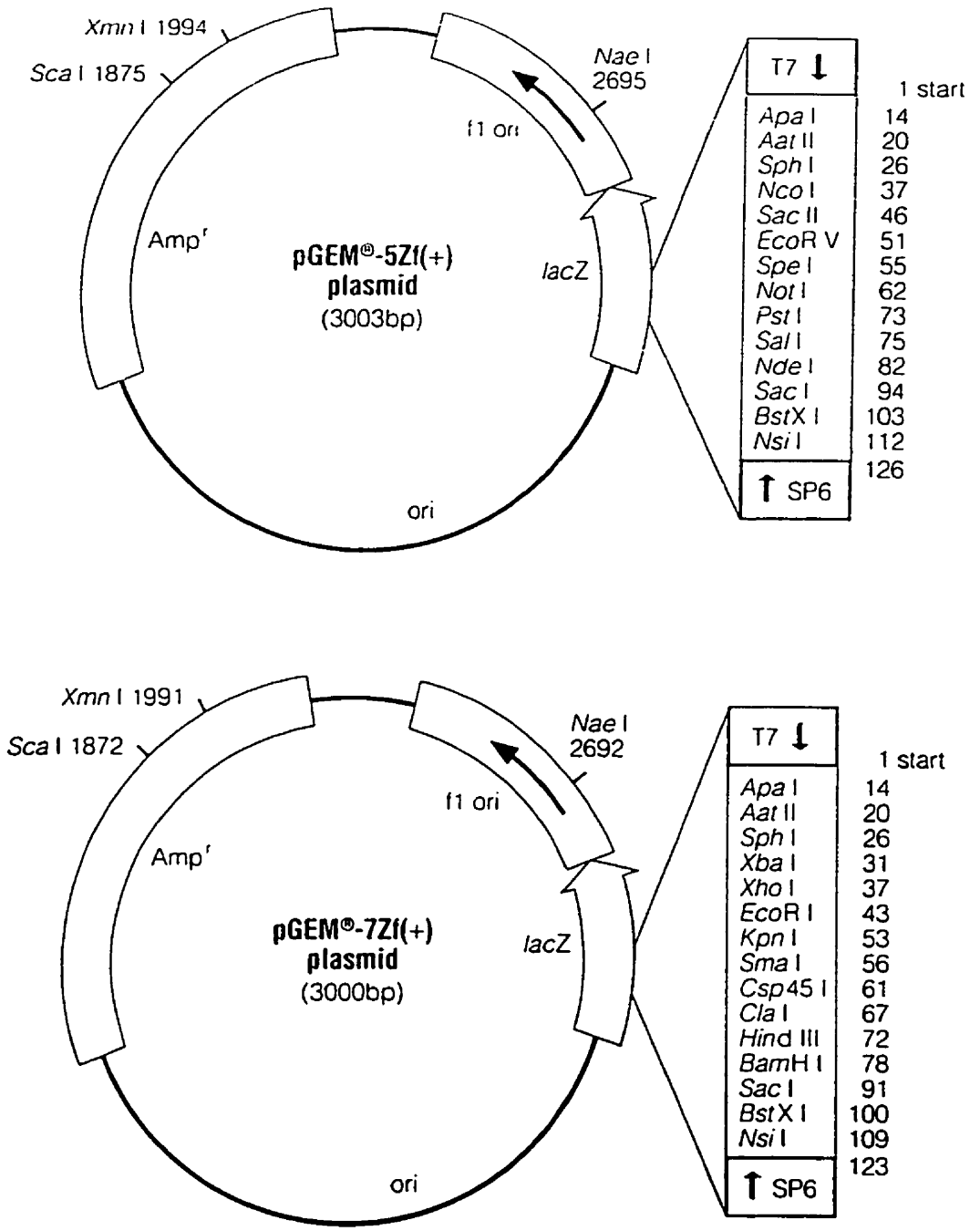


Figure 2.2 The circle maps of pGEM-5Zf(+) and pGEM-7Zf(+) vectors (Promega)

DNA was then introduced into DH5α *E. coli* competent cells as recommended (Gibco, BRL). Bacterial colonies resistant to ampicillin were further screened by blue/white selection. Finally, plasmid DNA carrying inserts was extracted and purified using the QIAprep Spin Plasmid Kit (Qiagen) following the manufacturer's protocol.

2.4 Double stranded DNA sequencing

Double stranded DNAs from all PCR product subclones were sequenced using the dideoxy sequencing method of Sanger using the T7 DNA sequencing kit (Pharmacia) following the manufacturer's protocol. 1.5-2μg of plasmid DNA was denatured by 0.2M NaOH for 10min at room temperature followed by the addition of a final concentration of 1.5M NaOAc (pH 4.8) to neutralize the reaction. After DNA was precipitated in ethanol and air dried, it was resuspended in 10μl of water, and annealed to the desired sequencing primer (3.75μM per reaction). Labeling reaction was conducted using [α-^{35}S]dATP (10mCi/ml, Amersham), and terminated by ddNTPs. The final reaction mixture was separated on a denaturing (8M urea) 6% SDS-acrylamide sequencing gel (DiaMed) in 1x TBE (100mM Tris-HCl, pH8.0, 83mM boric acid, 1mM EDTA). Following electrophoresis at 55 watts, the gel was blotted to 3MM chromatography paper (Whatman) and dried under vacuum on a gel dryer (BioRad). The gel was then exposed to X-ray film overnight at room temperature.

2.5. Restriction map of plasmid pHG 36

1.54μg aliquots of plasmid pHG 36 (fig 2.1) were digested with different combinations of EcoR I, Pst I, Sac I, Sph I, and Xba I restriction

enzymes (Pharmacia). The digested DNA fragments were analyzed by electrophoresis with 0.8% agarose gels and visualized following EtBr intercalation by UV radiation.

2.6 Assays for determination of promoter activity

2.6.1 Plasmid constructions

All PCRs were performed in a total volume of 50µl in the presence of 1x PCR buffer, 10% DMSO, 2.5mM $MgCl_2$, 400µM dNTP, 6ng pHG 36 DNA template, 1µM each primer, and 2U Vent DNA polymerase (New England Biolabs). The ~2.5 kb GnT II 5' untranslated region in pHG 36 was PCR-amplified using primer CAT/FL (Table 2), complementary to the plasmid pBluescript II(+) (Stratagene), and primer P-276Xr with an Xba I site linker (Table 2): 1 cycle of 94°C for 5min, 60°C for 1min, 72°C for 3min; followed by 34 cycles of 94°C for 1min, 60°C for 1min, 72°C for 3min, and a final extension at 72°C for 15min. The PCR product was digested with 15U of Xba I (Pharmacia), and the ~900 bp DNA fragment was subcloned into pCAT-Basic vector (fig 2.3, Promega), yielding the parent p-1200CAT chimeric construct. The chimeras p-680CAT, p-636CAT, p-553CAT, p-472CAT, and p-390CAT contain PCR-amplified fragments from -680 to -276, -636 to -276, -553 to -276, -472 to -276, and -390 to -276 cloned into the Sal I site of the pCAT-Basic vector, respectively. The corresponding forward primers used in the PCR reactions to generate these fragments were P-680f, P-636f, P-553f, P-472f, and P-390f, respectively, together with the reverse primer P-276Sr (Table 2). The above constructs were transformed into competent *E. coli* DH5α cells by standard procedures. The ampicillin-resistant colonies were screened by PCR with a pCAT-Basic primer (pUC/M13 reverse primer,

CAT/FL	5' GGC CGC ATG CCG ATA AGC TTG ATA TCG AAT 3'
P-680f	5′ GCA **TGT CGA** CCT TCG CAC GTC TCG CCT TTC 3′
P-636f	5′ ATA **TGT CGA** CGA AGA AAG AGG CGG AAG TGG 3′
P-553f	5′ GCT **AGT CGA** CTC GGT CGC TGC TGG AAA GCA 3′
P-472f	5′ GCT **AGT CGA** CGG GCA GTT GCG GGT TGT CAT 3′
P-390f	5′ GCA G**GT CGA** CTA AGG ATG AGA GCG CAG AGG 3′
P-276Sr	5′ ATA T**GT CGA** CTC CTC TCC ATT CAG CAC CAC 3′
P-276Xr	5' GCG C**TC TAG A**TC CTC TCC ATT CAG CAC CAC 3'
PUC/M13 reverse primer	5′ CAG GAA ACA GCT ATG AC 3′

Table 2 Primers used in constructing pCAT chimeric constructs In order to construct pCAT plasmids, the gene specific primers were attached to restriction enzyme site sequences. The Sal I site (indicated as S) is shown in bold, while Xba I (indicated as X) is in bold and underlined. GnT II promoter region specific primers are named corresponding to their 5' end positions. "f" denotes forward primer and "r" denotes reverse primer. Primer CAT/FL is complementary to the plasmid pBluescript II (+) sequence. PUC/M13 reverse primer was used to screen the plasmids for the presence of the correct inserts by PCR (see text).

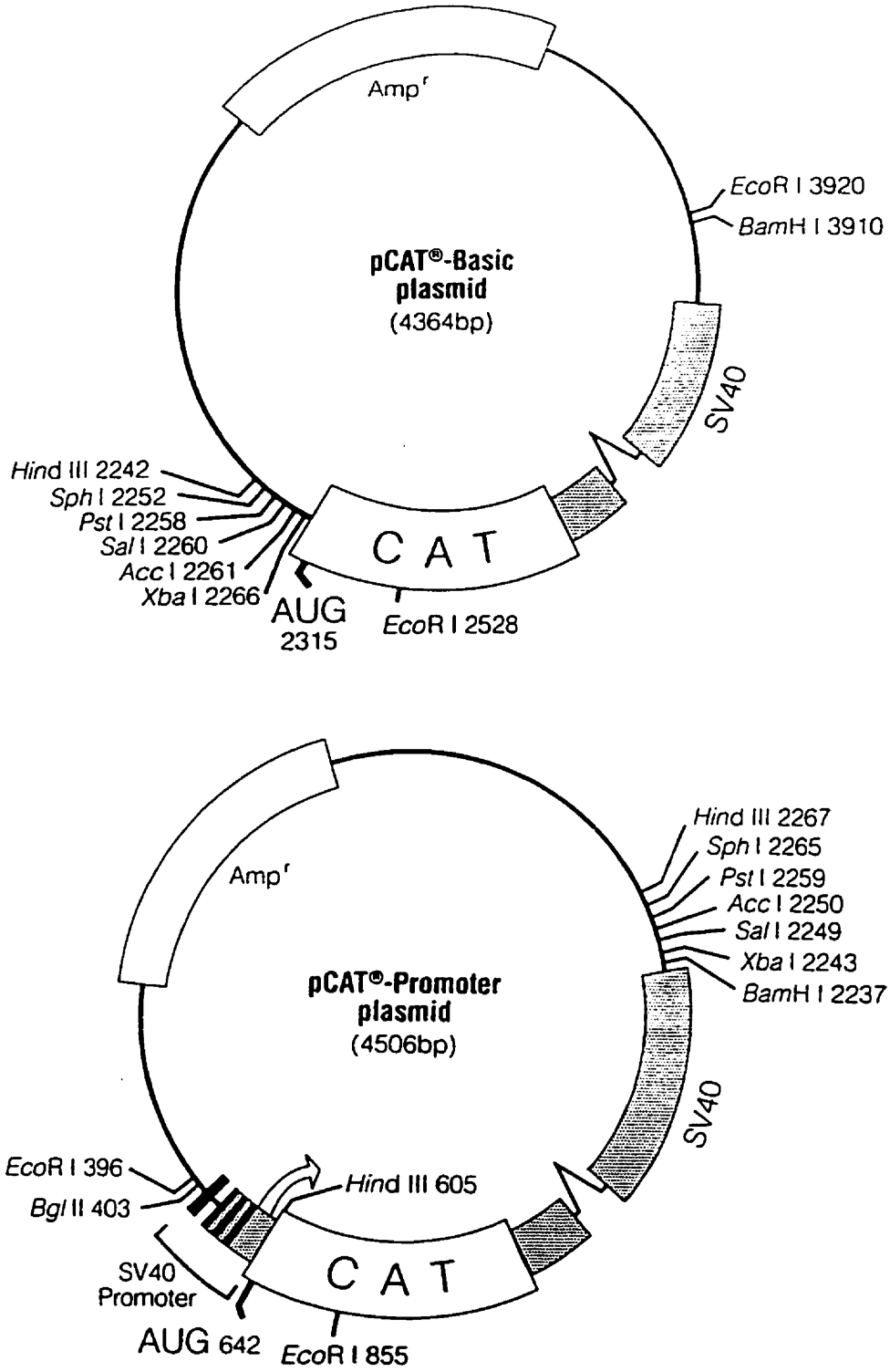


Figure 2.3 The plasmid maps of pCAT-Basic and pCAT-Promoter vectors (Promega)

Table 2) and a GnT II promoter region specific antisense primer using whole bacteria as templates. Plasmid DNAs from cultures yielding expected PCR products were extracted using QIAprep Plasmid Kit (Qiagen), and finally, the DNA sequences were verified by the double-stranded DNA sequencing method.

2.6.2 Cell culture and transient transfection

HeLa, a human cervical epitheloid carcinoma cell line, was provided by Betty Yip (Hospital for Sick Children, Toronto). The cells were grown in αMEM supplemented with 10% FBS (Gibco BRL) in a humidified incubator at 37°C with 5% CO_2.

6.8 x 10^5 HeLa cells were seeded per 60mm tissue culture dish in 3.5ml αMEM with 10% FBS. Adherent cells were transfected at 70-80% confluence (~21 hours after seeding). pCAT chimeric plasmids (2μg) and 2μg of plasmid pSV-β-galactosidase (fig 2.4, Promega) were co-diluted with 0.18ml Opti-MEM I Reduced Serum Medium (Gibco BRL). For each transfection, 8μl of Lipofectamine reagent (Gibco BRL) was diluted in a separate tube into 0.18ml Opti-MEM I Reduced Serum Medium (Gibco BRL). The two solutions were combined, mixed gently, and incubated at room temperature for 45min. The lipid-DNA complexes were further diluted with 1.4ml antibiotic-free αMEM without serum, and layered over the cells which had been washed with 3.5ml of the same medium. Control transfections were carried out either without DNA, or with pCAT-Basic, or with pCAT-Promoter vectors (fig 2.3, Promega). The cells were incubated for 5 h at 37°C in a CO_2 incubator and then 1.7ml αMEM-20% FBS were added. Medium was replaced at 21h after transfection, and, after a further 51h of incubation, cells were harvested and extracted using the

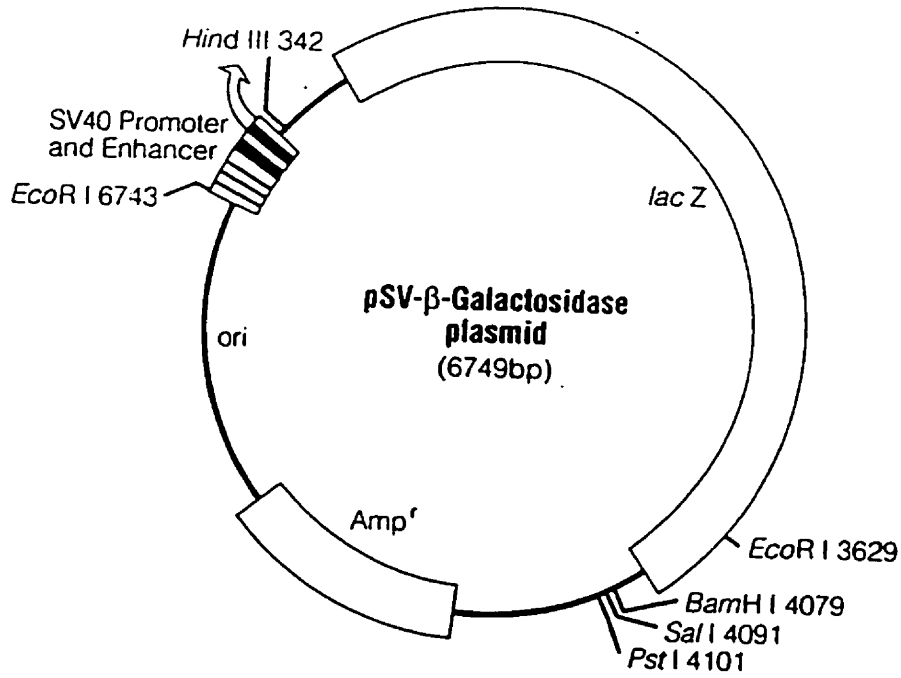


Figure 2.4 The circle map of pSV-β-galactosidase vector
The arrow indicates the direction of transcription of the *lac* Z gene (Promega).

Tris-HCl buffer freeze/thaw protocol, as described in the Promega technical bulletin (Promega). Briefly, cells were washed 3 times with 5ml of Mg^{2+} and Ca^{2+}-free PBS buffer (Sigma), then 1ml of TEN buffer (40mM Tris-HCl, pH 7.5, 1mM EDTA, pH 8.0, 150mM NaCl) was added. Cells were incubated for 5min at room temperature, scraped, centrifuged, resuspended in 0.1ml 0.25M Tris-HCl, pH 8.0, and subjected to three cycles of rapid freezing-thawing. Extracts to be used for CAT assays were heated at 60°C for 10min prior to assay.

2.6.3 CAT and β-Galactosidase Assays

CAT activity assays were carried out on transfected HeLa cell extracts using [^{14}C]chloramphenicol (58.4 mCi/mmol, DuPont NEN) and n-butyryl coenzyme A (Sigma), a 30min incubation time, and the mixed xylenes (Aldrich) phase separation assay (Liquid Scintillation Counting assay), according to the Promega protocol. At least three independent transfections were carried out for each experiment. CAT activity was determined from a standard curve using CAT enzyme supplied by Promega (1 unit enzyme activity corresponds to 1nmole of acetate transferred to chloramphenicol per minute at 37°C). β-galactosidase activity was quantitated according to the protocol provided by Invitrogen. Briefly, a mixture of HeLa cell extract and 1ml of Z Buffer (pH 7.0, 60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$, 50mM β-mercaptoethanol) were incubated at 28°C. Temperature pre-equilibrated orthonitrophenyl β-galactoside (0.2 ml, 4mg/ml ONPG, pH 7.0) was then added to initiate the reaction. When a faint yellow color developed (within 2 hr), 0.5ml of 1M sodium carbonate was added to stop the reaction and the optical density (OD) at 420nm was determined (1 unit is defined as the

amount of enzyme that will hydrolyze 1nmole of ONPG per minute at 28°C).

Chapter Three

Results

3.1 Expression of human GnT II in human cell lines

Northern analysis of RNAs from different human cell lines revealed a major transcript at ~2.0 kb and one minor band at ~2.9 kb (fig 3.1). Although the amount of mRNA in the HL60 cell line seems to some degree higher than those in the other cell lines examined, the pattern of the multiple transcripts in each cell line remains the same. This is consistent with findings in rat and human tissues, as well as in EBV-transformed human lymphoblasts [Tan *et al*, 1995, D'Agostaro *et al*, 1995]. There are no variations in expression patterns between tissues of GnT-II, probably because it is a housekeeping gene and is therefore anticipated to be present in all tissues. However, the major transcript was 2.0 kb in human cell lines, 2.6 kb in EBV-transformed human lymphoblasts, 2.8 kb in rat tissues, and 3.0 kb in human tissues. The mechanism of these differences remains unclear. Since the open reading frame of GnT II is encoded in a single exon, the multiple transcripts might result from multiple transcription initiation sites or multiple transcription termination signals, or both.

3.2 Definition of the transcription initiation and termination sites of human GnT II by 5' and 3' RACE

To address the above question, several methods could be employed. However, the rapid amplification of cDNA ends (RACE) method has an advantage over the primer extension and RNase protection methods in that RACE provides for obtaining DNA sequence information while the latter two methods only provide the lengths of the tested RNA. In RACE, mRNA is first reverse transcribed by priming with an oligo (dT) primer under the activity of reverse transcriptase, resulting in a cDNA-mRNA hybrid. The

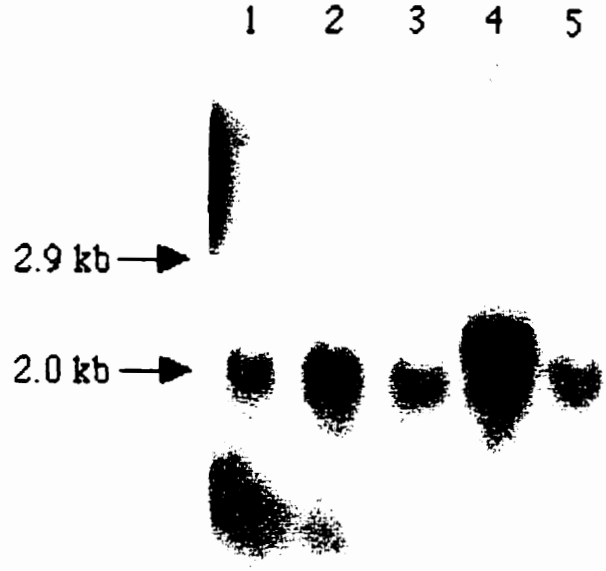


Figure 3.1 Northern blot analysis of human cell lines
20 μg aliquots of total RNA from (1) LS180, (2) HepG2, (3) HeLa, (4) HL 60, and (5) A431 cell lines were fractionated by electrophoresis in a denaturing 1.0% agarose gel (10% formaldehyde). The filter was hybridized with a probe covering 507 bp of GnT II ORF (nucleotide position: from +119 to 625). One major signal at ~2.0 kb and one minor band at ~2.9 kb were observed in all cell lines tested. The pattern indicated that there was no tissue-specific expression of human GnT II.

second strand cDNA is synthesized by replacing the mRNA with a DNA strand (fig 3.2) [Sambrook *et al*, 1989]. Shown in figure 3.3 is the generated cDNA template and primers used in the Marathon RACE PCR reactions. Three advantages are present in this method. First, the first-stranded cDNA synthesis uses a modified lock-docking oligo (dT) primer, which contains two degenerate nucleotides at the 3' end. These nucleotides position the primer at the start of the poly(A) tail and thus eliminate the 3' heterogeneity inherent with only oligo (dT) priming. Second, blunt-end ligation of the adapter to both ends of ds-cDNA by T4 DNA ligase is much more efficient than homopolymeric tailing or ligation of adapter to ss-cDNA. Lastly, both 5' and 3' RACE PCRs are primed with a gene-specific internal primer and the adapter primer, AP1 or AP2. The AP1 and AP2 primers, as shown in fig. 3.3, are colinear with the single-stranded portion of the adapter, and thus the adapter-ligated cDNA does not contain binding sites for AP1 and AP2 primers in the first PCR synthesis cycle. Furthermore, the 3' end of the shorter adapter strand is blocked with an amine group ('x' as indicated in the scheme), and therefore can not be extended to create AP1 and AP2 binding sites in the starting cDNA population. During the first round of amplification, the gene-specific primer is elongated, creating AP binding sites at the 5' (5' RACE) or 3' (3' RACE) ends of cDNA. In subsequent cycles, both APs and the gene-specific primer can bind allowing exponential amplification of the cDNA of interest. As a result, non-specific products are greatly reduced. This method works efficiently to create a library of adapter-ligated ds-cDNAs.

3.2.1 Qualification of the Marathon RACE synthesized ds-cDNAs

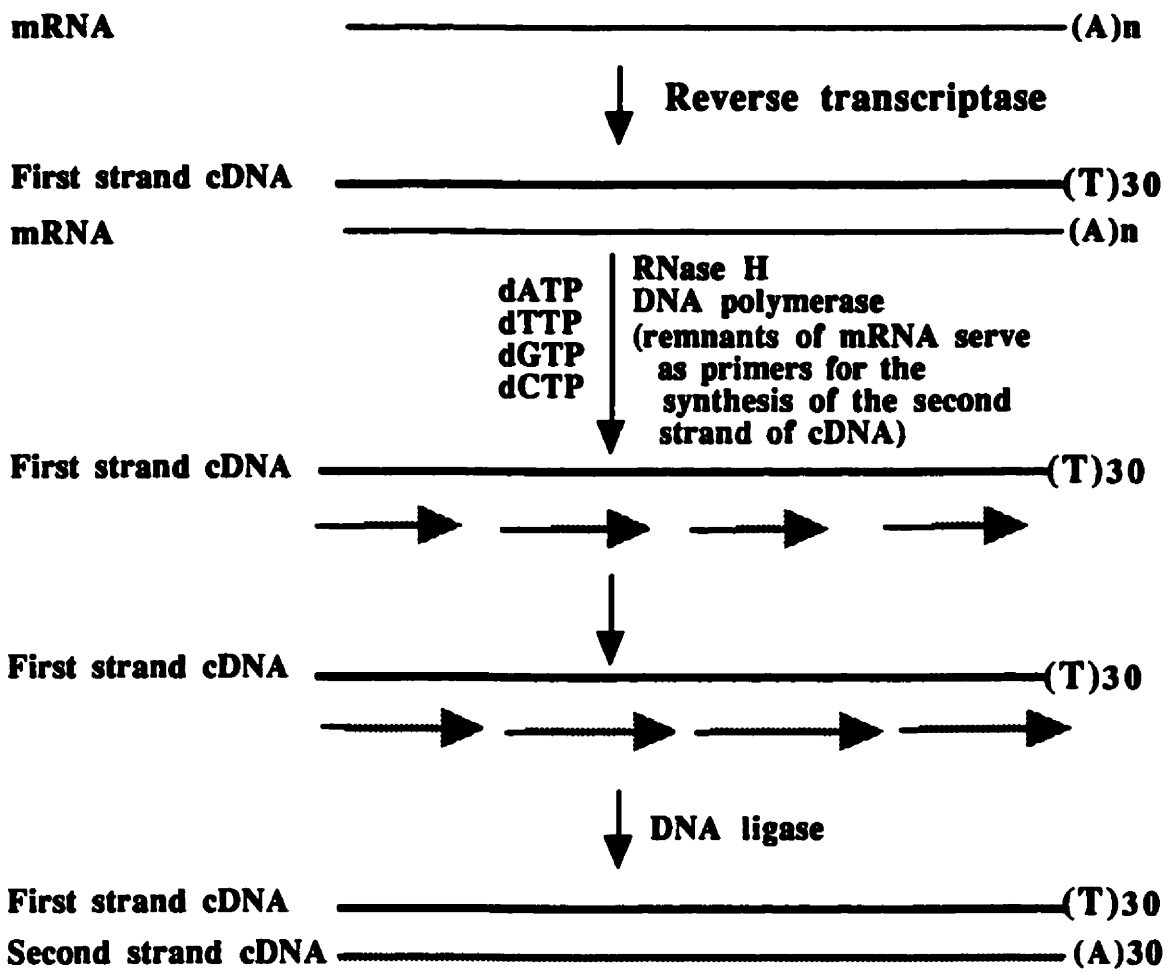
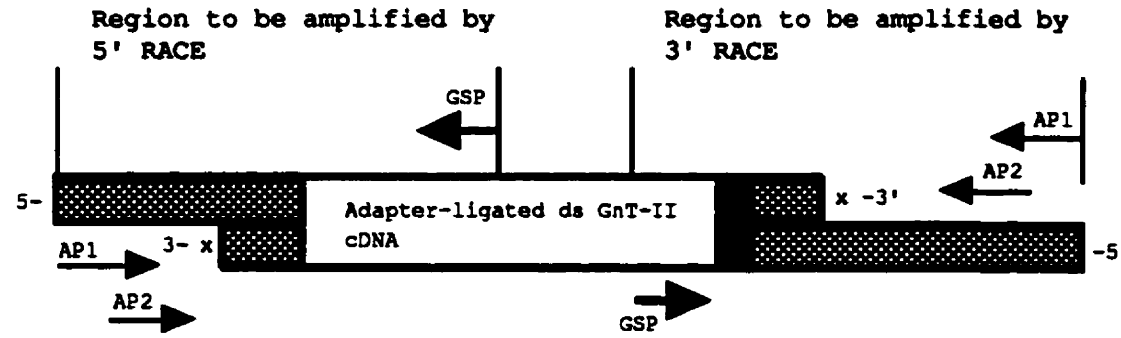


Figure 3.2 Replacement synthesis of double-stranded cDNA

The first strand cDNA was synthesized by reverse transcriptase primed with an oligo (dT) primer. The resulting cDNA-mRNA hybrid was then used as a template for a nick-translation reaction. RNase H produced nicks and gaps in the mRNA strand creating a series of RNA primers that were used by *E.coli* DNA polymerase I for the synthesis of the second strand of cDNA. Finally, fragments of second strand cDNA were ligated by *E. coli* DNA ligase. [Adapted from Sambrook *et al*, 1989]



Legends

cDNA synthesis primer * (plus complementary strand)

cDNA adapter * (ligated to either end of cDNA)

x: NH2 blocking group
N: degenerate nucleotide
GSP: GnT-II gene specific primer
*: supplied

Figure 3.3 Scheme of the template and primers used in the Marathon RACE reactions

cDNA synthesis and adaptor ligation creates a population of cDNAs with the structure depicted above. PCRs using sets of gene-specific primers indicated in bold arrows, in combination with adapter primers (thin arrows), can then be employed to amplify gene fragments of interest (see text for details).

Prior to proceeding to the PCR reactions of 5' and 3' RACE, the quality of the cDNA preparation was tested by Southern blot analysis. GnT II cDNA amplified by the gene specific primer (P900f) and the cDNA synthesis primer was analyzed in an agarose gel. Although a notably prominent band was seen at ~1.3 kb, several DNA bands between ~500 bp and ~1.3 kb could be detected by EtBr staining (fig 3.4B, lane 1). Hybridization of the southern blot of these products revealed multiple signals at ~600 bp, ~700 bp, ~1.2 kb and ~1.3 kb (fig 3.4C, lane 1). No positive signals were detectable with the PCR reaction products when using only the cDNA synthesis primer (fig 3.4C, lane 2). These data demonstrate that the cDNA constructed by this method is suitable for further experiments. The data here also suggest that the use of multiple transcription stop sites may be producing the multiple signals detected by Northern blot.

3.2.2 Transcription termination of human GnTII gene

The adapter-ligated GnT-II ds-cDNA was amplified by the gene-specific primer P900f with AP1. The resulting primary PCR product was then re-amplified using nested primers: P922F and AP2. Agarose gel analysis indicated a strong sharp band at ~600 bp and a smear at about 1.0 kb, while controls with only one of the primers or both primers in the absence of template were negative (fig 3.5B, lanes 1-4, respectively). The 600 bp DNA was subcloned into pGEM5Zf(+) plasmid at the Not I and Pst I sites. Sequencing analysis of the insert showed that priming of the cDNA occurs at the poly(A) site located 20 bp downstream from the first AATAAA sequence of the GnT II 3' UTR (fig 3.6 and fig 1.7). This is at

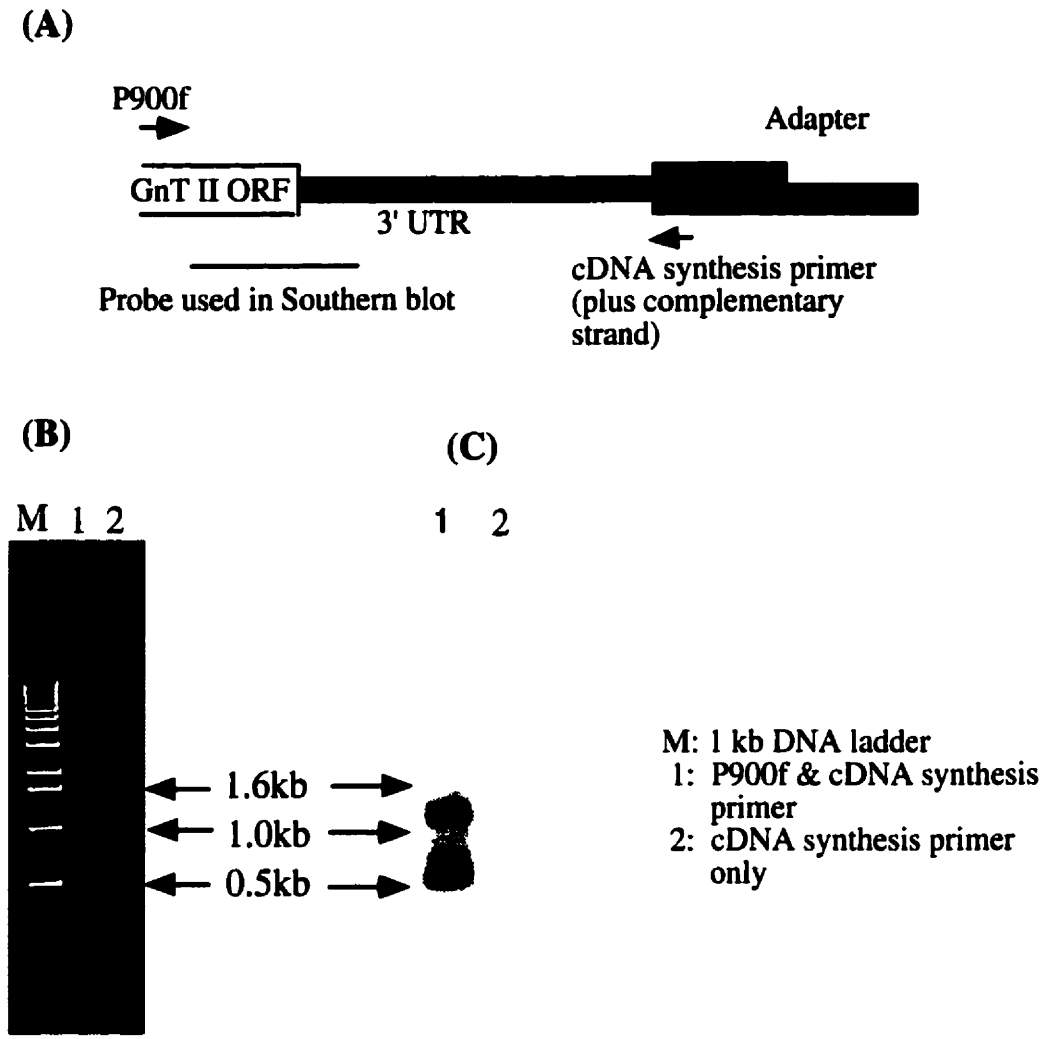


Figure 3.4 Analyses of amplified GnT II cDNA by EtBr-staining and Southern blot

(A) Template and primers used in PCR amplification of GnT II 3' UTR are indicated in this picture. Probe used in Southern hybridization is also shown beneath the template. (B) The P900f-cDNA synthesis primer PCR product was fractionated on a 1.2% agarose gel containing 1μg/ml EtBr. The gel was visualized under UV light. There were several DNA bands between 0.5 kb and ~1.3 kb (lane 1). (C) Southern blot of the same gel. The membrane was hybridized with a 691 bp probe covering 420 bp of GnT II ORF (nucleotide position: 922 to 1341) and 271 bp of 3' UTR (nucleotide position: 1342 to 1612). Four signals at ~600 bp, ~700 bp, ~1.2 kb and ~1.3 kb were detectable (lane 1). No signals were shown by hybridization in the negative control PCR reaction when using cDNA synthesis primer only (lane 2).

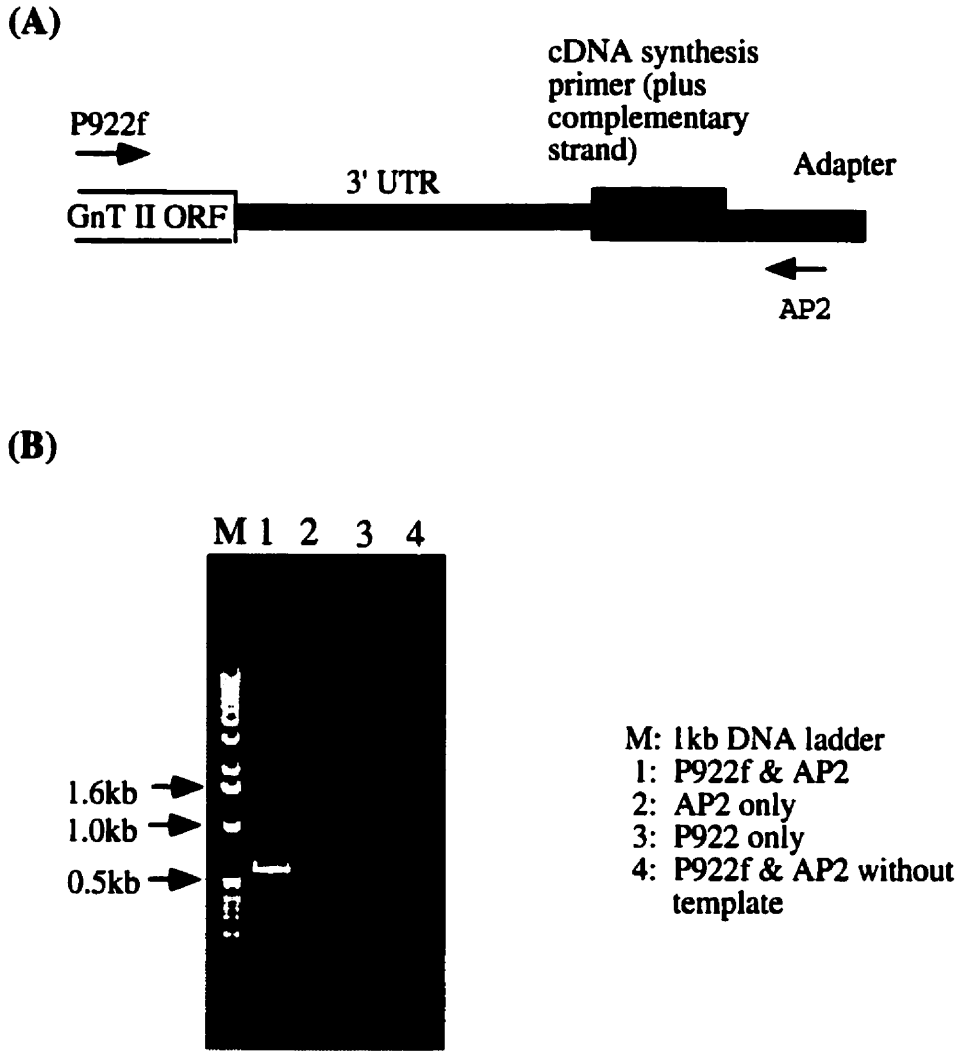


Figure 3.5 Analysis of human GnT II cDNA containing the first poly(A) site

(A) Template and primers used in PCR amplification of GnT II 3' UTR are indicated in this picture. (B) Primary 3' RACE PCR product was amplified by PCR with adapter primer 2 (AP2) and GnT-II gene specific primer P922f. This product was analyzed on a 1.2% agarose gel stained with EtBr. There was a sharp band at ~600 bp and a smear at about 1kb (lane 1).

3' ACATT ATGCT GAGTG ATATC CCGCT TAACC CGGGC TGCAG CGTAC

5' TAA TACGA CTCAC TATA 3'--->

T7 Sequence primer

GAGGG CCGGC GGTAC CGGCG CCCTA TAGTG ATCAC GCCGGCGAAA

Not I

AAAAA AAAAA AAAAA AAAAA AAAA AATG **TCAAG GATTT GAGAA**

AATAA GTTAA CATAC AGGAC AAACC TGTTT ATAGT TTTTA TCTTC

TGACA GCGAA AACAT TGACA CTAAA AGTG ACGTC CAGCT GGTAT

Pst I

ACCCT CTCGG AGGGT TGCGC AACCT ACGTA TCGAA CTCAT AAGAT ATCA 5'

Figure 3.6 Sequence of GnT II cDNA containing the first poly(A) site

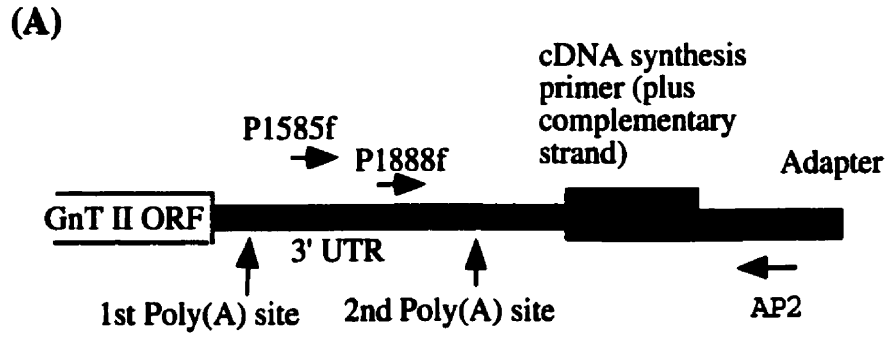
Primary 3' RACE PCR product was amplified by PCR with adapter primer 2 (AP2) and GnT-II gene specific primer P922f. The DNA was subcloned into pGEM5Zf(+) at the Pst I and Not I sites (underlined). Sequences 5' of the Pst I and 3' of the Not I sites correspond to pGEM5Zf(+) plasmid sequences. The insert is shown in bold and italics. The poly(A) tract, together with the two nucleotides originally from the cDNA synthesis primer (italics), starts at 20bp downstream from the AATAAA element (bold and underlined).

the appropriate spacing to indicate the first poly(A) site in the GnT-II genome is used in transcription.

To determine whether the second poly(A) site is functional, two gene-specific primers between the first and second poly(A) sites, either P1585f or P1888f, together with AP2, were used to re-amplify the primary 3' RACE PCR product. A very weak product at ~550 bp could be seen in the P1585-AP2 reaction, and a ~250 bp product was observed in the P1888f-AP2 reaction. These bands were extracted from the agarose gels and subjected to a third round of PCR amplification under the same conditions. This time, strongly staining products appeared at the expected sizes, supporting that they were specific, and not by-products (fig 3.7B). Subcloning and sequencing indicated that the synthesized cDNA did indeed end 20 bp downstream from the second AATAAA element (figs 3.8 and 1.7). This data demonstrates that the second poly(A) site of the GnT-II gene is also active in transcription.

Similar experiments were carried out to establish a role for the third poly(A) site by using a gene-specific primer between the second and third poly(A) sites (P2109f) and AP2. The cDNA was found to terminate with a poly(A) tail initiating at 23 bp downstream from the AATAAA moiety of the third putative polyadenylation site (figs 3.9 and 1.7).

These experiments verify that all three poly(A) sites found in the GnT-II genome are functional in gene transcription, at least in the human cell line tested. Two minor transcripts in rat tissues could not hybridize with the probe prepared from the 3' untranslated region, supporting the idea that multiple transcripts may result from the differential utilization of distinct poly(A) sites during the maturation of the primary transcript. 3' end heterogeneity, due to the use of different poly(A) sites, was also



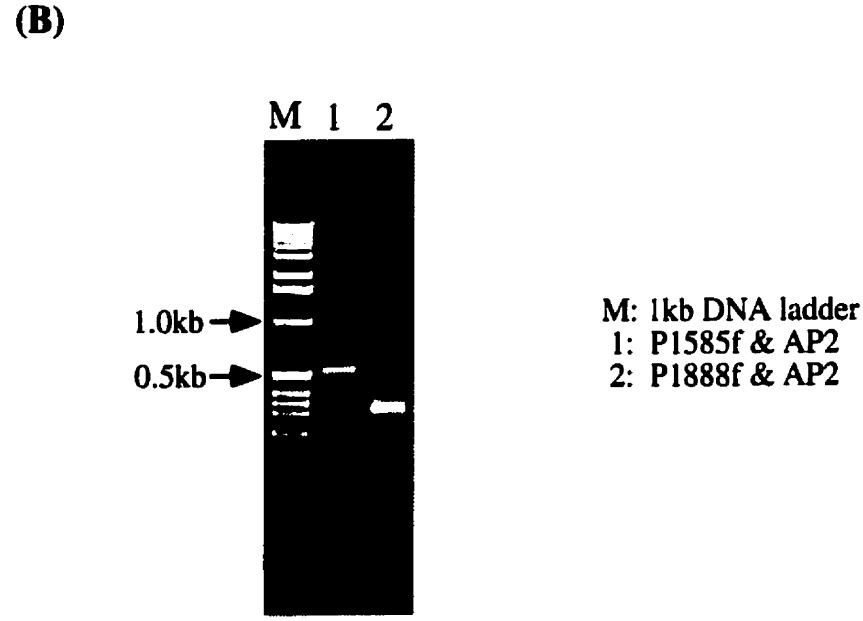


Figure 3.7 Human GnT II cDNA containing the second poly(A) site

(A) Template and primers used in PCR amplification of GnT II 3' UTR are indicated in this picture. (B) Primary 3' RACE PCR product was re-amplified using adapter primer 2 (AP2), either with GnT II gene specific primer P1585f (lane 1), or with P1888f (lane 2). These products were analyzed on a 1.8% agarose gel stained with EtBr. There were sharp bands at ~550 bp (lane 1) and ~250 bp (lane 2), respectively.

5' TGTAA TACGA CTCAC TATAG GGCGA ATTGG GCCCG ACGTC

T7 primer

GCATG CTCCT CTAGA CTCGA GGAAT TC GTGTT TTGAG TGTGT TTTGG

EcoR I

AACCT TCATA GAACA CACTT TCTTT TGGAA TGTAT TTGAT TGATA

AGAAA GTTTA AACAT TGTTT TCACC TCAAT GTAGA AATAC AGTGG

TTTTG TTTTT TTTTT TCTTT TAGTG CTGAC AAAAT AAAAT ACTCA

TTTTT GC *CT AAAAA AAAAA AAAAA AAAAA AAAAA AAAA* GCGGC CGCTG

AATTC GGTAC CCCGG GTTCG AAATC 3'

EcoR I

Figure 3.8 Sequence of GnT II cDNA containing the second poly(A) site

Primary 3' RACE PCR product was amplified twice by PCR with adapter primer 2 (AP2) and GnT-II gene specific primer P1888f. The DNA was subcloned into pGEM7Zf(+) at the EcoR I site (underlined). The insert is shown in bold and italics. The poly(A) tract, together with the two nucleotides originally from the cDNA synthesis primer (italics), occurs 20bp downstream from the AATAAA element (bold and underlined). The sequences 5' and 3' of the EcoR I sites correspond to pGEM7Zf(+) plasmid sequences.

5' TGTAA TACGA CTCAC TATAG GGCGA ATTGG GCCCG ACGTC

T7 primer

GCATG CTCCT CTAGA CTCGA GGAAT T CAGTG AAGGC ATTCT ACAAG

EcoR I

TTTTG GGTTA GCATT ACATT TTAAT ATTTA CTATT GCTAC ATTGT

ATAAT TGAGT TTGAA ATAAA ACCCA GCTTA TGACA AT *GTAAA*

*AAAAA AAAAA AAAAA AAAAA AAAAA AA*GCG GCCGC TGAATT C GGTAC

EcoR I

CCCGG GT TCG AAATC GATAA GCTTG GATCC GGAGA GCTCC CAACG

CGTTG GATGC ATAGC TTGAG TATTC TATAG TGTCA CCTAA ATAGC TTG 3'

Figure 3.9 Sequence of GnT II cDNA containing the third poly(A) site

Primary 3' RACE PCR product was amplified by PCR with adapter primer 2 (AP2) and GnT-II gene specific primer P2109f. The DNA was subcloned into pGEM7Zf(+) at the EcoR I site (underlined) The insert is shown in bold and italics. The poly(A) tract, together with the two nucleotides originally from the cDNA synthesis primer (italics), occurs 23bp downstream from the AATAAA element (bold and underlined). The sequences 5' and 3' of the EcoR I sites correspond to pGEM7Zf(+) plasmid sequences.

identified in murine α-mannosidase II transcripts; the enzyme which provides the substrate for GnT-II [Moreman & Robbins, 1991; Lal *et al*, 1994]. Although this type of extended 3' untranslated region is common among the genes of glycosylation processing enzymes, the functional significance(s) remains to be determined. It is believed that 3' UTR plays a variety of roles in the regulation of gene expression, and indirectly, in the control of cell growth, differentiation and development of at least some genes. This topic will be discussed in the following section.

3.2.3 Transcription initiation of human GnT II gene

Although we have established that multiple GnT II transcripts are produced as a result of variation of length of the 3' UTR, the role of the 5' end was still unclear. The adapter-ligated GnT-II ds-cDNA template was amplified by PCR with AP1 as the forward primer and P21r, which overlapped the ATG start codon, as the reverse primer. 5μl of this PCR product was reamplified with nested primers, i.e. AP2 and P-47r under the same conditions. The same procedure was conducted for a third round of amplification. There were no visible DNA fragments after the 1st and 2nd round PCR reactions in an EtBr-stained agarose gel, while two bands at ~500 bp and ~700 bp could be seen in the 3rd round of PCR. When the annealing temperature was raised from 55°C to 65°C, the ~700 bp DNA band was replaced with an ~900 bp band, whereas the ~500 bp DNA remained unchanged (fig 3.10). Both the ~500 bp and ~900 bp DNA bands were subcloned and sequenced. The ~900 bp DNA, not surprisingly, appeared to be due to another gene with high homology. It might result from non-specific or homologous priming during PCR. Eight subclones carrying the ~500 bp DNA insert were selected and sequenced. The 5' end

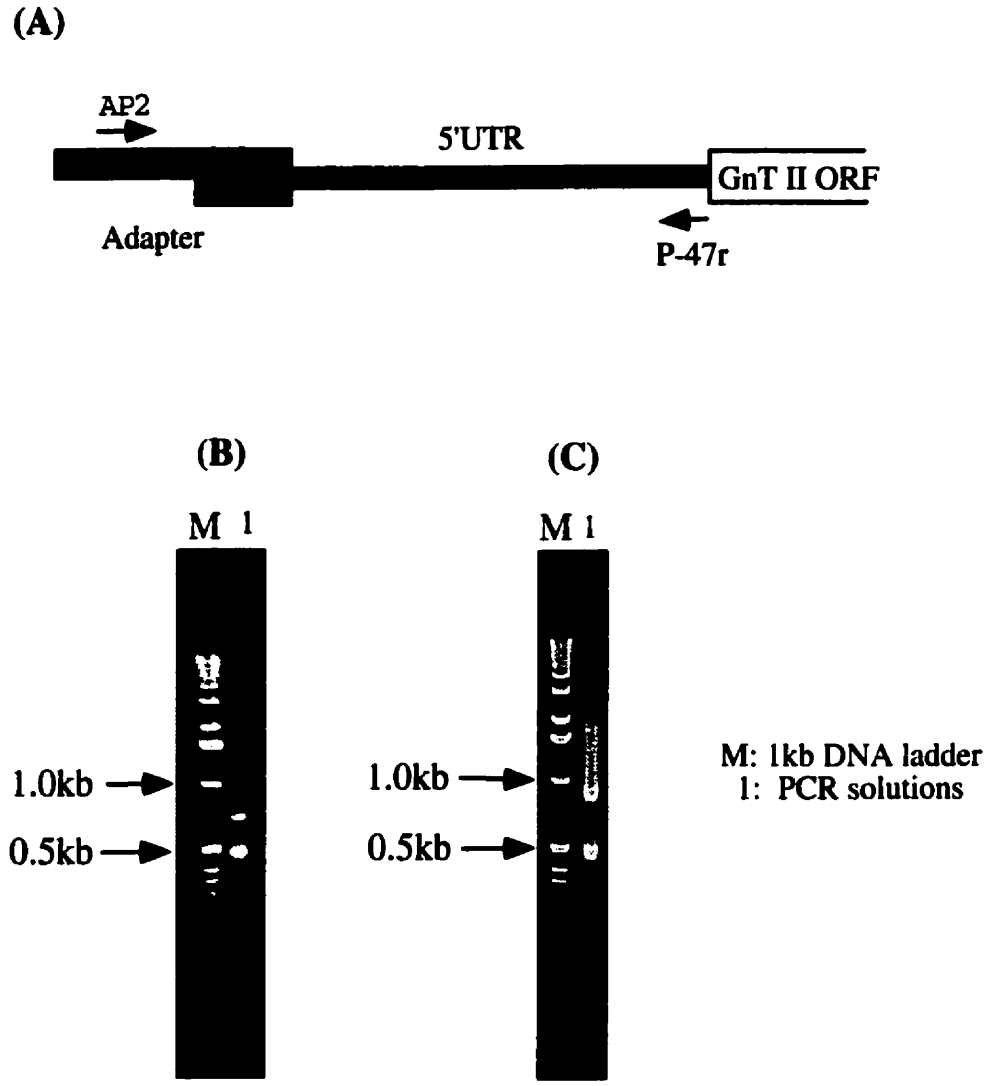


Figure 3.10 EtBr-staining of 5' RACE PCR products

(A) Template and primers used in PCR amplification of GnT II 5' UTR are indicated in this picture. 5' RACE PCR products were subjected to a 1.0% agarose electrophoresis in 1x TAE buffer, and the gel with EtBr was visualized under UV light. The PCR annealing temperature was raised from 55°C (B) to 65°C (C), otherwise other PCR conditions remained unchanged. Two bands could be detected in each gel. The longer DNA fragment was a non-specific PCR product based on sequencing. Sequencing of the ~500 bp DNA confirmed that it was the 5'-flanking region of GnT II gene.

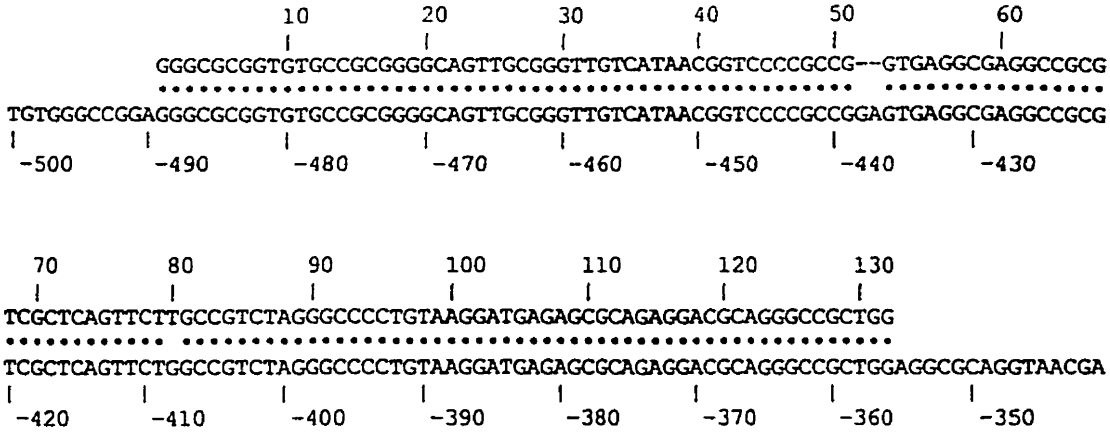
of subclone 21 corresponded to nucleotide -489 relative to the first nucleotide (+1) of the ATG translation start codon, subclone 6 to -486, subclone 15 and 16 to -477, subclone 5 to -450, subclone 10 to -443, subclone 7 to -442, and subclone 20 to -440 (fig 3.11 and fig 3.12). The putative transcription initiation sites were separated over 50 bp. Multiple initiation sites are common among housekeeping genes lacking a TATA box. This data also suggested that there were no untranslated exon(s) upstream of ATG start codon as sequence could be aligned directly with genomic sequence. Two putative transcription initiation sites at ~450 and 435 bp upstream of the translation start codon in rat GnT II gene were identified by RNase protection experiments [D'Agostaro *et al*, 1995], supporting our 5' RACE results.

Results of 3' and 5' RACE are summarized in figure 3.13. Assuming 200 bp of poly(A) tail, the shortest transcript employing the first poly(A) site should be about 2.0 kb (460 bp of average 5' untranslated region + 1341bp of open reading frame + 68bp of 3' untranslated region + 200bp of poly(A) tail). This correlates with our Northern analysis. The other two transcripts terminating at the 2nd and 3rd poly(A) sites should be about 2.7 and 2.9 kb, respectively. However, only a very weak signal at 2.9 kb was detected in our Northern blots. This might be due to the relatively low amounts of these transcripts.

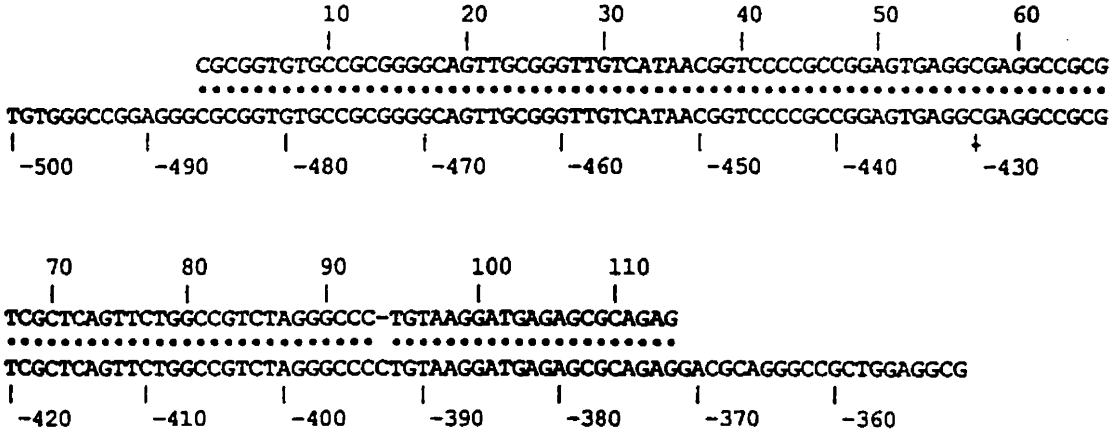
3.3 Structural analysis of human GnT II gene 5' UTR

Genomic DNA structural analysis revealed that the (G+C) content in the 682 bp region sequenced to date, upstream of the translation start codon ATG, is 69% (>50%). The observed over expected (O/E) frequency of the putative CpG doublet (scanned manually) is 0.96 (>0.6). It could be

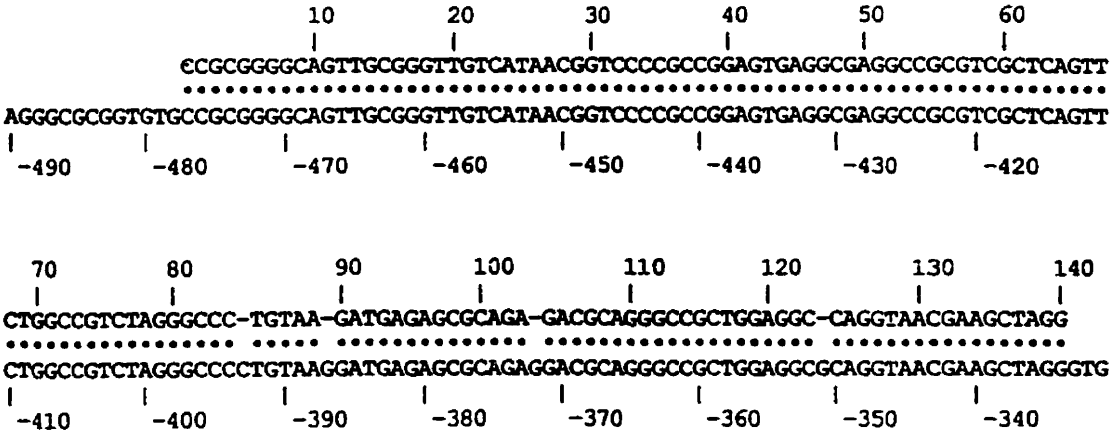
(A)


10 20 30 40 50 60
GGGCGCGGTGTGCCGCGGGGCAGTTGCGGGTTGTCATAACGGTCCCCGCCG--GTGAGGCGAGGCCGCG
TGTGGGCCGGAGGGCGCGGTGTGCCGCGGGGCAGTTGCGGGTTGTCATAACGGTCCCCGCCGGAGTGAGGCGAGGCCGCG
-500 -490 -480 -470 -460 -450 -440 -430
70 80 90 100 110 120 130
TCGCTCAGTTCTTGCCGTCTAGGGCCCCTGTAAGGATGAGAGCGCAGAGGACGCAGGGCCGCTGG
TCGCTCAGTTCTGGCCGTCTAGGGCCCCTGTAAGGATGAGAGCGCAGAGGACGCAGGGCCGCTGGAGGCGCAGGTAACGA
-420 -410 -400 -390 -380 -370 -360 -350

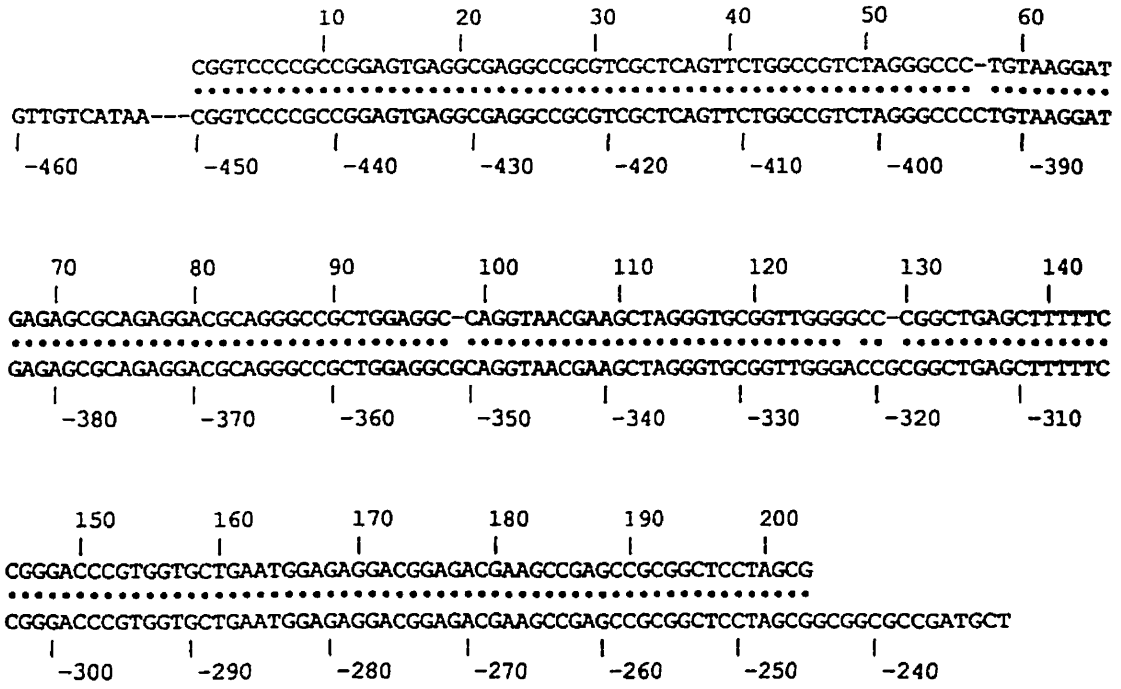
(B)


10 20 30 40 50 60
CGCGGTGTGCCGCGGGGCAGTTGCGGGTTGTCATAACGGTCCCCGCCGGAGTGAGGCGAGGCCGCG
TGTGGGCCGGAGGGCGCGGTGTGCCGCGGGGCAGTTGCGGGTTGTCATAACGGTCCCCGCCGGAGTGAGGCGAGGCCGCG
-500 -490 -480 -470 -460 -450 -440 -430
70 80 90 100 110
TCGCTCAGTTCTGGCCGTCTAGGGCCC-TGTAAGGATGAGAGCGCAGAG
TCGCTCAGTTCTGGCCGTCTAGGGCCCCTGTAAGGATGAGAGCGCAGAGGACGCAGGGCCGCTGGAGGCG
-420 -410 -400 -390 -380 -370 -360

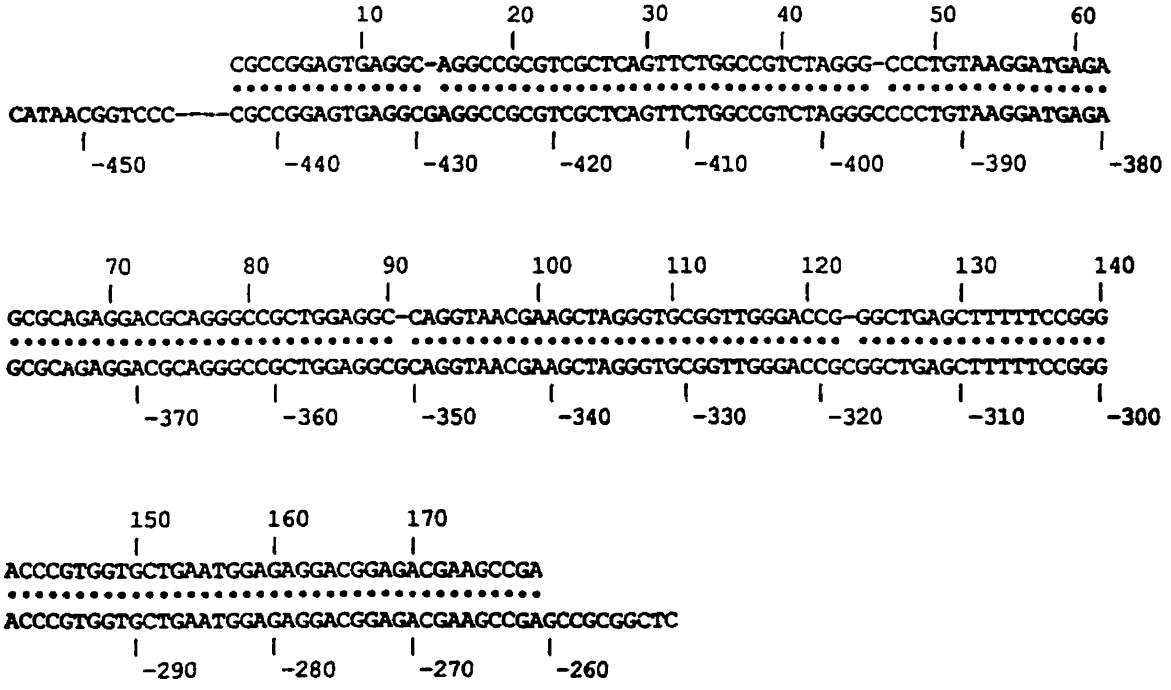
(C)


10 20 30 40 50 60
CCGCGGGGCAGTTGCGGGTTGTCATAACGGTCCCCGCCGGAGTGAGGCGAGGCCGCGTCGCTCAGTT
AGGGCGCGGTGTGCCGCGGGGCAGTTGCGGGTTGTCATAACGGTCCCCGCCGGAGTGAGGCGAGGCCGCGTCGCTCAGTT
-490 -480 -470 -460 -450 -440 -430 -420
70 80 90 100 110 120 130 140
CTGGCCGTCTAGGGCCC-TGTAA-GATGAGAGCGCAGA-GACGCAGGGCCGCTGGAGGC-CAGGTAACGAAGCTAGG
CTGGCCGTCTAGGGCCCCTGTAAGGATGAGAGCGCAGAGGACGCAGGGCCGCTGGAGGCGCAGGTAACGAAGCTAGGGTG
-410 -400 -390 -380 -370 -360 -350 -340

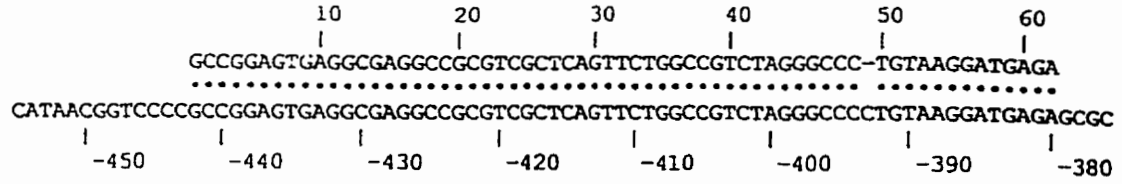
(D)


10 20 30 40 50 60
CGGTCCCCGCCGGAGTGAGGCGAGGCCGCGTCGCTCAGTTCTGGCCGTCTAGGGCCC-TGTAAGGAT
GTTGTCATAA---CGGTCCCCGCCGGAGTGAGGCGAGGCCGCGTCGCTCAGTTCTGGCCGTCTAGGGCCCCTGTAAGGAT
-460 -450 -440 -430 -420 -410 -400 -390
70 80 90 100 110 120 130 140
GAGAGCGCAGAGGACGCAGGGCCGCTGGAGGC-CAGGTAACGAAGCTAGGGTGCGGTTGGGGCC-CGGCTGAGCTTTTTC
GAGAGCGCAGAGGACGCAGGGCCGCTGGAGGCGCAGGTAACGAAGCTAGGGTGCGGTTGGGACCGCGGCTGAGCTTTTTC
-380 -370 -360 -350 -340 -330 -320 -310
150 160 170 180 190 200
CGGGACCCGTGGTGCTGAATGGAGAGGACGGAGACGAAGCCGAGCCGCGGCTCCTAGCG
CGGGACCCGTGGTGCTGAATGGAGAGGACGGAGACGAAGCCGAGCCGCGGCTCCTAGCGGCGGCGCCGATGCT
-300 -290 -280 -270 -260 -250 -240

(E)


10 20 30 40 50 60
CGCCGGAGTGAGGC-AGGCCGCGTCGCTCAGTTCTGGCCGTCTAGGG-CCCTGTAAGGATGAGA
CATAACGGTCCC----CGCCGGAGTGAGGCGAGGCCGCGTCGCTCAGTTCTGGCCGTCTAGGGCCCCTGTAAGGATGAGA
-450 -440 -430 -420 -410 -400 -390 -380
70 80 90 100 110 120 130 140
GCGCAGAGGACGCAGGGCCGCTGGAGGC-CAGGTAACGAAGCTAGGGTGCGGTTGGGACCG-GGCTGAGCTTTTTCCGGG
GCGCAGAGGACGCAGGGCCGCTGGAGGCGCAGGTAACGAAGCTAGGGTGCGGTTGGGACCGCGGCTGAGCTTTTTCCGGG
-370 -360 -350 -340 -330 -320 -310 -300
150 160 170
ACCCGTGGTGCTGAATGGAGAGGACGGAGACGAAGCCGA
ACCCGTGGTGCTGAATGGAGAGGACGGAGACGAAGCCGAGCCGCGGCTC
-290 -280 -270 -260

(F)



(G)

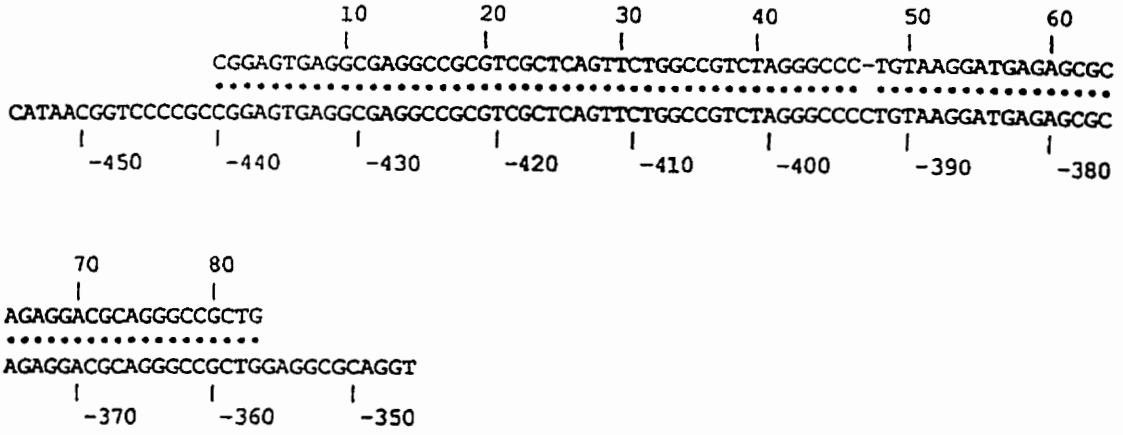


Figure 3.11 Comparison of the 5' end sequences of 5' RACE subclones with that of human genomic GnT II 5' UTR

The lower sequences correspond to the human GnT II genomic DNA. The upper sequences are: (A) subclone 21; (B) subclone 6; (C) subclones 15 and 16; (D) subclone 5; (E) subclone 10; (F) subclone 7; and (G) subclone 20. Gaps, indicated as (-), were introduced in order to optimize matching.

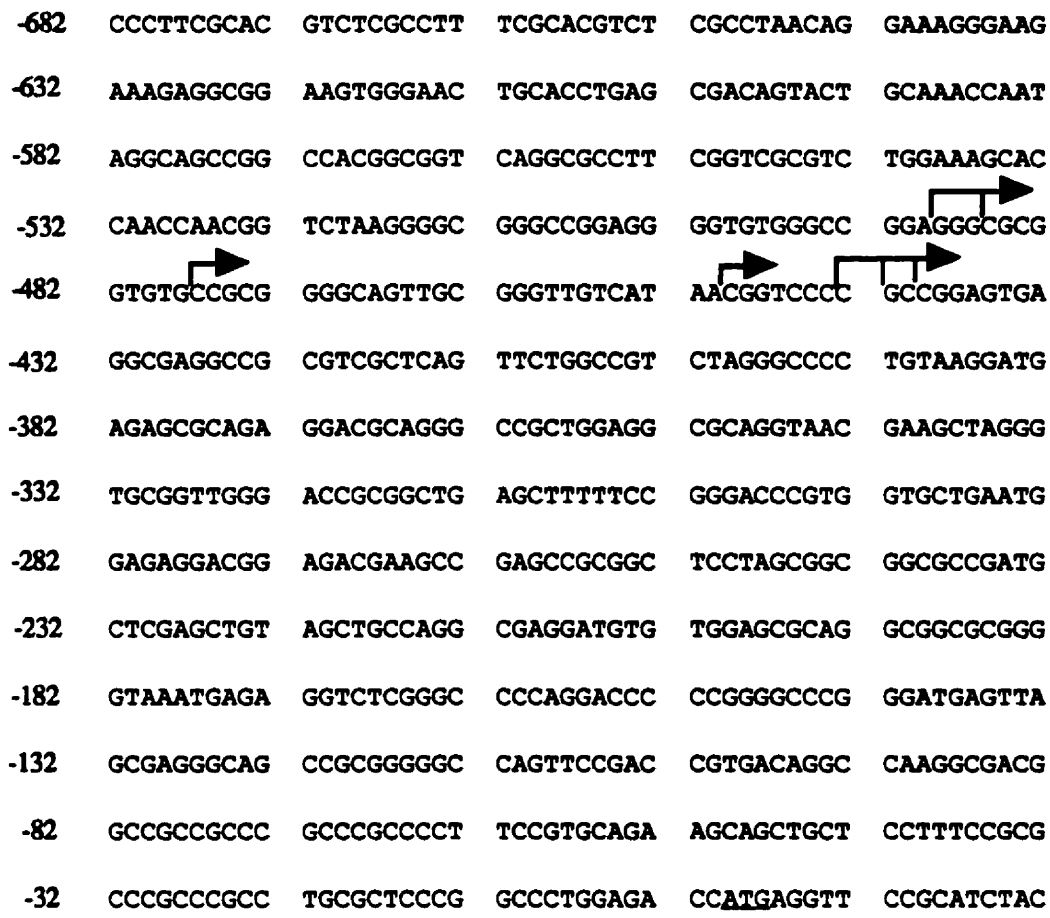


Figure 3.12 Summary of 5' RACE

This figure summarizes the multiple putative transcription initiation sites (shown with arrows) of the human GnT II gene as determined by 5' RACE. These putative initiation sites are spread over 50 bp within the 5' UTR. This feature is very common in housekeeping genes that lack a TATA-box. The translation start codon is underlined.

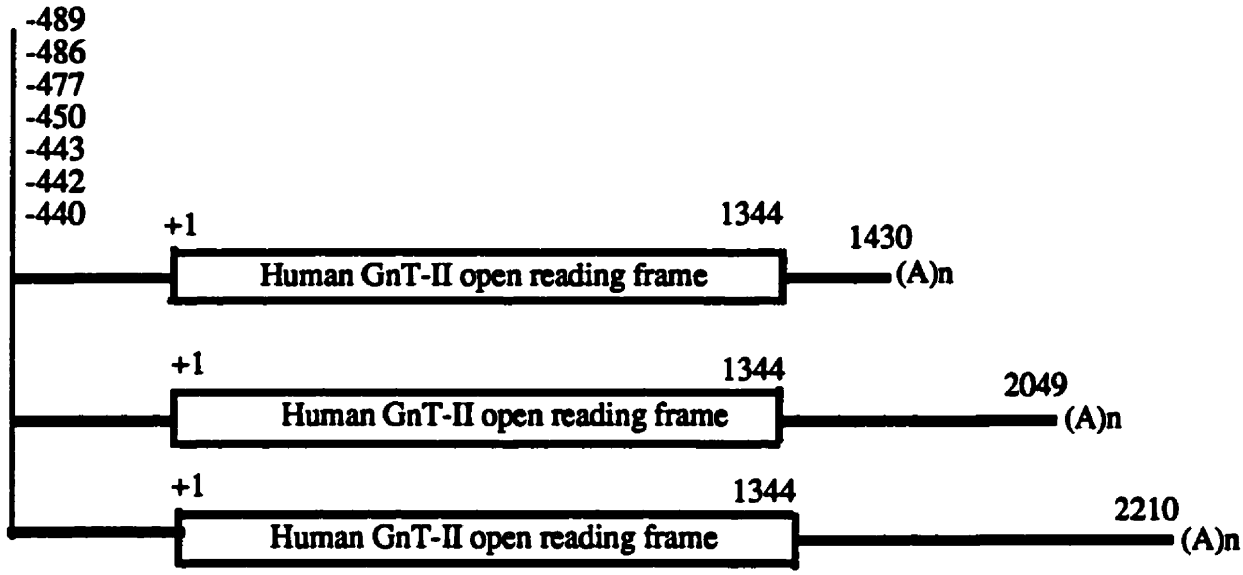


Figure 3.13 Summary of 5' and 3' RACE

Based on the 5' RACE results, the human GnT II gene is transcribed at multiple initiation sites covering ~50 bp. Transcription terminates at three distinct polyadenylation signals as determined by the 3' RACE experiments. Therefore, three transcripts are generated with the shortest one terminating at the first polyadenylation site, and the medium and longest transcripts terminating at the second and third polyadenylation elements. These results are consistent with the Northern blot analysis, in which multiple transcripts are detectable. The sizes of transcripts determined by 5' and 3' RACE are about 2.0 kb, 2.7 kb, and 2.9 kb in length.

reasoned that the promoter region might reside within this putative CpG island [Bohm, 1995].

A number of potential *cis*-elements for transcription factor binding within the GnT II gene 5'-flanking region have been identified (fig 3.14). An initiator (Inr)-like element (YYCANTYYY, see "Introduction") was identified at -417. It does not overlap with any of the determined putative transcription initiation sites. Although such overlap often occurs, this is not functionally necessary. An imperfect conserved Multiple start site Element Downstream (MED-1, GCTCCC/G, see "Introduction") is detected at -418 as GCTCAG. A conserved CCAAT box is found at -587. Members of the CCAAT/enhancer binding protein family (C/EBP) have been shown to participate in basal and activated transcription regulation. There is a Ying-Yang 1 or YY1 element [core conserved sequence 5' CAT 3', Lopez-Bayghen *et al*, 1996] at -458. YY1, a zinc finger protein related to the Kruppel family of transcriptional regulators of *Drosophila melanogaster*, is able to activate or repress transcription initiation depending on the promoter and the intracellular environment. Numerous Sp1 binding sites [(G/T)(G/A)GGC(G/T)(G/A)(G/A)(G/T), Faisst & Meyer, 1992] are found at -671, -655, -629, -570, -562, -517, -506, -498, -490, -446, -434, -356, and -319. Both YY1 and Sp1 are known to guide transcription initiation in some TATA-less promoters [Shiffman *et al*, 1996], and furthermore, the direct protein-protein interaction between Sp1 and TFIID complex has been shown to be essential for the assembly of the preinitiation complex [Pugh & Tjian, 1991]. Consistent with this are the YY1 and multiple Sp1 recognition sequences located upstream of the putative transcription initiation sites in the GnT II gene. Multiple binding sites for the Ets family of transcription factors [(C/A)GGA(A/T), Wasylyk *et al*, 1993] are

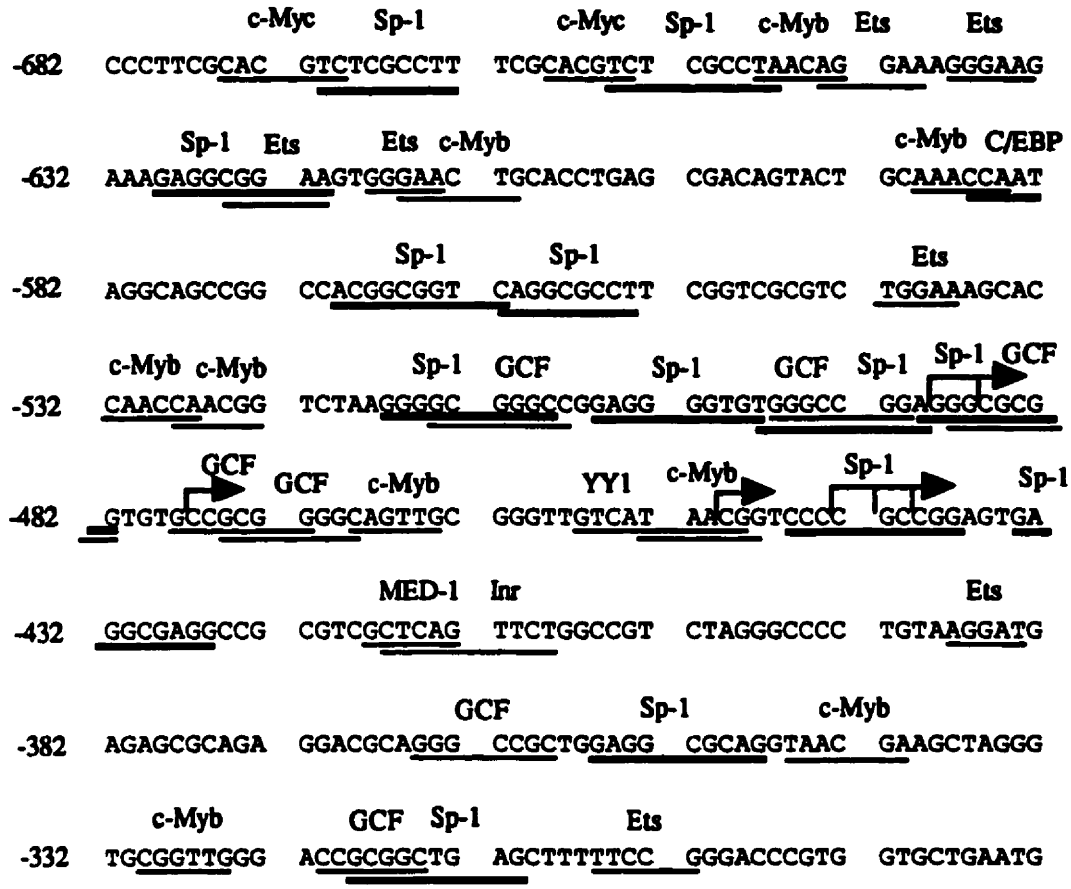


Figure 3.14 Nucleotide sequence of the 5'-end region of the human GnT II gene

Part of the DNA sequence upstream of the translation initiation codon ATG is shown in this figure. The putative transcription initiation sites determined by 5' RACE are indicated by arrows. Putative transcription factor binding sites are underlined and named above the sequence.

detected at -644, -638, -625, -618, -542, -388, and -306. Two c-Myc binding sites [Eilers *et al*, 1991] at -675 and -659, and several c-Myb binding sites [(T/C)AAC(T/C)(G/A), Nakagoshi *et al*, 1990] at -648, -616, -590, -532, -528, -469, -453, -346, and -330 are also identified. These latter nuclear oncoproteins have been implicated in regulation of gene expression during a variety of biological processes including growth control, differentiation and transformation, etc. There are several GCF recognition sites [(G/C)CG(G/C)(G/C)(G/C)C, Kageyama & Pastan, 1989; Angotti *et al*, 1994] at -514, -497, -488, -478, -475, -365, and -321. GCF recognizes a sequence similar to Sp1, but it acts as a repressor to downregulate gene expression. Other putative transcription factor binding sites include those for NF-1, leader binding protein 1 (LBP-1), E2F, NF-IL6, LF-A1 and TCF-1, etc. In addition to its well established function in cell cycle regulation, E2F may interact with the GC-rich promoter, by recruiting Sp1 to the GC-rich elements or *vice versa* [Shin *et al*, 1996].

3.4 Restriction map of human GnT II 5' flanking region

As the limited sequence information available for human GnT II 5' flanking region could have impeded our construction of plasmids for promoter activity assays, it was therefore necessary to obtain a restriction map of this region. The choices of restriction endonucleases used in this experiment were based on the map of the pCAT-Basic vector multiple cloning site so that appropriate enzymes that could produce cohesive ends for ligation could be tested. Digestion of pHG36 plasmid (fig 2.1) with either Sph I or Sal I gave a disperse EtBr staining picture indicative of supercoiled DNA structure, suggesting that there were no sites for Sph I or Sal I (fig. 3.15). Linearization of pHG 36 was obtained with either EcoR I,

Sac I or Xba I enzymes. Digestion with Xba I and either Sac I or EcoR I produced two corresponding sets of bands at ~1.8 kb / ~4.2 kb and ~1.4 kb / ~4.6 kb, respectively. Therefore, the cutting site of Xba I was mapped to ~1.2 kb upstream of the ATG codon. Digestion of pHG 36 with Pst I introduced two fragments at ~3.3 kb and ~2.7 kb. Since there is a Pst I digestion site at position +521 in the GnT II open reading frame, the second Pst I cutting site must be located at ~2.2 kb upstream of the ATG codon. The restriction map of pHG 36 plasmid is summarized in fig 3.15.

3.5 Functional analysis of the presumptive human GnT II promoter

In order to characterize the regions regulating the transcription activity of the GnT II gene, a series of chimeric constructs were made containing different lengths of the 5'-flanking region fused to the CAT reporter gene, and tested in transient transfection experiments as specified in "Materials and Methods". The 3' ends of all inserts in these constructs were identical and extended to -276bp upstream of the ATG codon. The human HeLa cell line was used as a recipient for the plasmids in these studies. The CAT activities of cell extracts were determined from a standard curve using different amounts of CAT enzyme standard (fig 3.16), and further normalized for transfection efficiency based on the β-galactosidase assays. The following CAT activity is therefore expressed as units CAT per unit β-galactosidase. Negative controls were performed by using cells without DNA transfection and cells transfected with pCAT-Basic plasmid. The CAT activities were 0.03±0.003 (n=4) and 0.04±0.006 (n=5), respectively. T-test between these two groups revealed no significant difference (p=0.080, >0.05). The positive control (pCAT-Promoter) had a

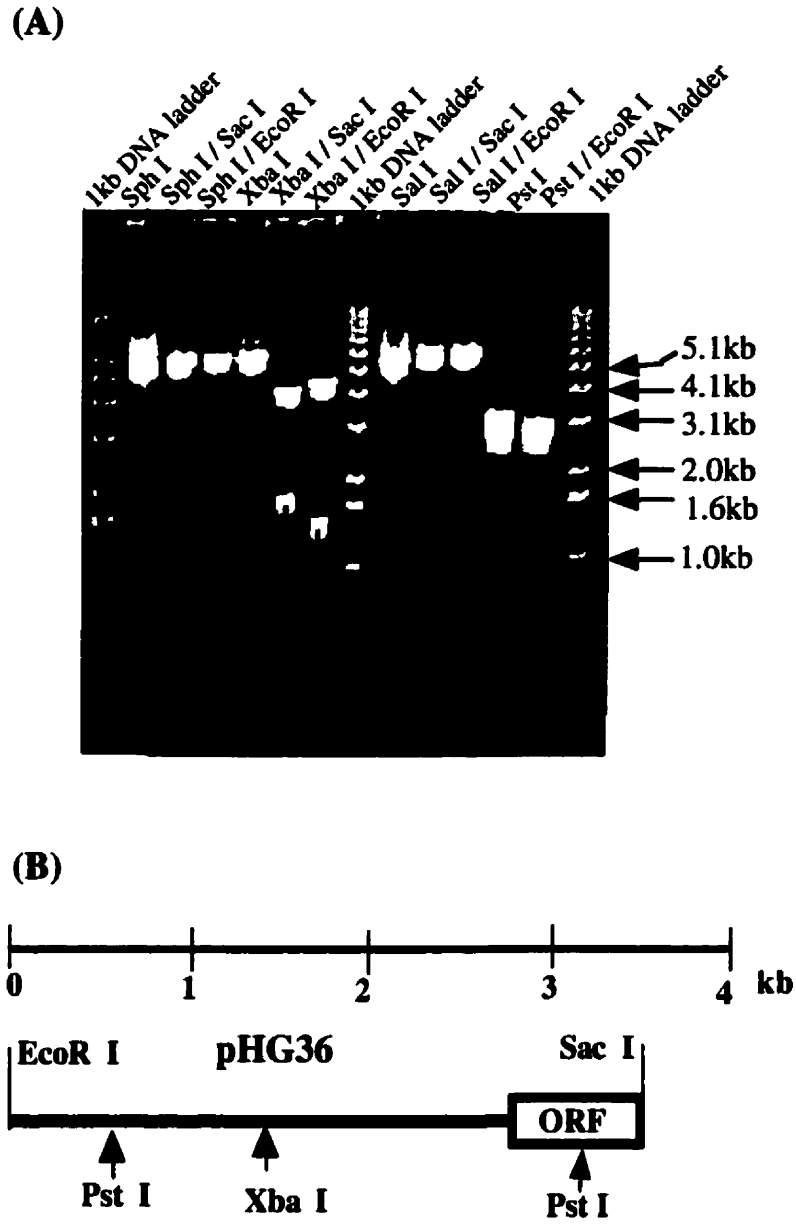


Figure 3.15 Restriction map of pHG36

(A) 1.54μg of plasmid pHG36 were digested with different combinations of EcoR I, Pst I, Sac I, Sph I and Xba I. The choice of these restriction enzymes was based on the multiple cloning site of pCAT-Basic vector so that the enzymes would produce cohesive ends in both the test insert and expression vector. The DNA fragments, digested to completion, were analyzed in 0.8% agarose gel stained by EtBr. (B) The simple map represents the restriction sites within plasmid pHG36 5' end. The diagram is drawn to scale.

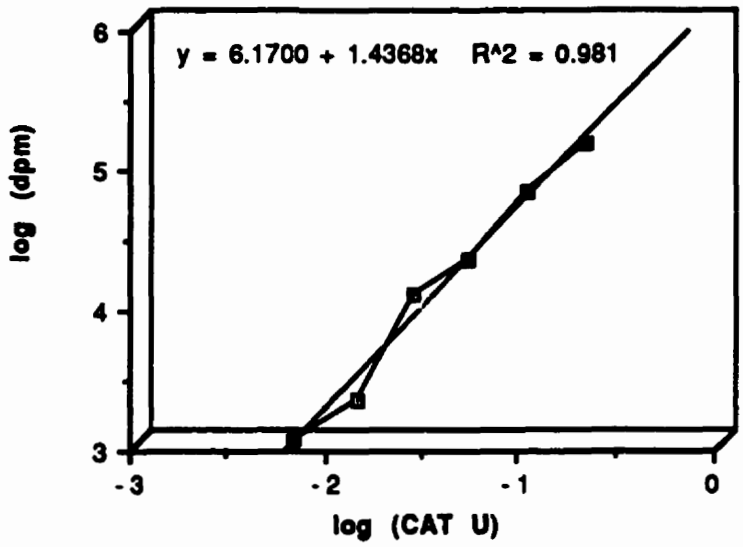


Figure 3.16 Standard curve of CAT assay

Various amounts of CAT enzyme (expressed as enzyme unit) were used in this experiment and the corresponding dpm were determined by the liquid scintillation counting assay. The correlation of CAT activities and dpm was plotted on logarithmic scales.

CAT activity 9.3-times higher than that of the pCAT-Basic control. The CAT activity of the construct containing the longest insert (p-1,200CAT) was 1.38±0.18 (n=3), a ~38-fold increase as compared to that of the pCAT-Basic control. The deletion of the 5'-flanking region from ~-1,200 to -680 (p-680CAT) resulted in a slight reduction in CAT activity (1.03±0.08, n=4). When the -680 to -276 DNA fragment was manipulated into the antisense direction in the pCAT-Basic vector, an unexpectedly high CAT activity was detected (data not shown), suggesting that some, if not all, *cis* -elements can function regardless of their directions. There were no differences in CAT activities between p-680CAT and p-636CAT plasmids. A further deletion of 83bp from p-636CAT plasmid gave rise to p-553CAT, which possessed an activity of 0.15±0.02 (n=4), a seven-fold reduction. Further deletions showed activities similar to p-553CAT plasmid (fig 3.17). It is thus concluded that the *cis*-acting regulatory elements located between -636 and -553 relative to the translation start site play central roles in regulating the expression of the human GnT II gene.

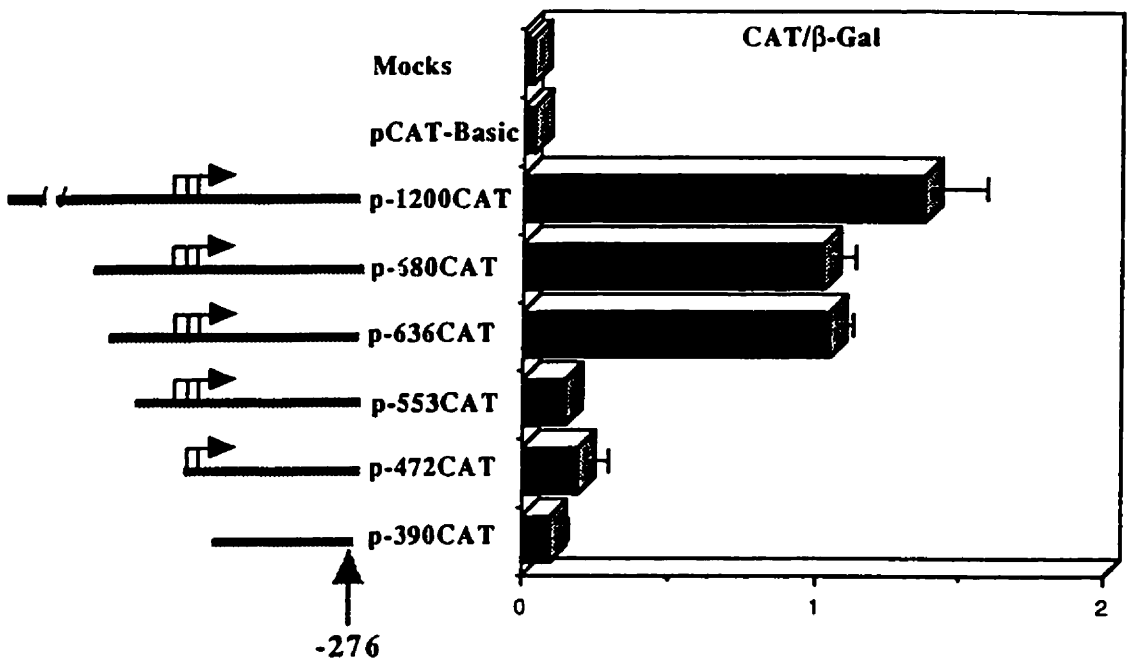


Figure 3.17 Promoter activities of the human GnT II 5' untranslated region

Deletion constructs of the GnT II 5'-flanking region linked to the chloramphenicol acetyltransferase (CAT) reporter gene were prepared as described in "Materials and Methods". Shown at the left are the constructs (not drawn to scale). The arrows indicate the putative transcription initiation sites determined by 5' RACE. All chimeras end at -276bp relative to the translational start site. The 5' end of each construct is indicated in its name. At the right are the CAT activities normalized to β-galactosidase (units CAT/unit β-galactosidase). At least three independent transfection experiments were performed with each construct. Values are expressed as means ±SD.

Chapter Four

Discussion

In this study, the putative transcription initiation and termination of the human GnT II gene were examined by the RACE method. The promoter region was also identified by examining the expression of a reporter gene. It is concluded that human GnT II initiates its transcription at multiple sites covering ~50 bp, and its mature mRNAs terminate at three distinct polyadenylation AAUAAA signals, leading to the production of multiple transcripts.

4.1 Selection of polyadenylation site

Studies have shown that the expression of the mouse Ig γ-H chains during B-cell development is controlled by distinct poly(A) site selection [Flaspohler & Milcarek, 1990]. In the early or memory B-cells, immunoglobin (Ig) heavy chain proteins exist as membrane-bound antigen receptors on the cell surface, while the secreted forms are produced by plasma cells. Transcription termination of the mouse Ig γ2a and 2b genes occurs downstream of the membrane-specific (mb) poly(A) site in both early and late-stage B cells, resulting in primary transcription units which contain different poly(A) sites [Flaspohler & Milcarek, 1990]. It is now believed that the balance of splicing of the last constant-region exon to exon M1 (mb mRNA) and polyadenylation at the 5' secretion-specific poly(A) site (sec mRNA) specifies the final sec-to-mb mRNA ratio, which is indicative of a B cell's developmental stage [Edwalds-Gilbert & Milcarek, 1995]. Since splicing has been shown to be constitutive and non-regulated [Peterson *et al*, 1991], the ratio of sec-to-mb mRNA is therefore controlled by polyadenylation. This result is further supported by the findings that the expression of mouse Ig γ2b gene is influenced by poly(A) site order and strength [Lassman & Milcarek, 1992], and that spacing between sec and mb

poly(A) sites, rather than splicing, governs the regulated production of the IgM heavy-chain transcription unit [Galli *et al*, 1988]. However, that differential transcriptional termination serves as another possible mechanism for the switch in poly(A) site usage during B-cell development can not be ruled out, as suggested by the expression of Ig μ transcription unit [Galli *et al*, 1987; Guise *et al*, 1988].

Another well-characterized example of the role of polyadenylation in the complex transcription unit in dictating the nature of mRNA output is the adenovirus major late transcription unit, which contains five poly(A) sites (L1 through L5) [Perscott & Falck-Pedersen, 1992]. In the early stage of infection, the 5' proximal L1 site is used almost exclusively, while in late stage all five sites are functional with L3 being slightly favored over the others. Using *in vitro* processing reactions, the L3 site has been demonstrated to be 5 to 10-fold more efficient than L1 as a substrate for RNA cleavage, 2-fold faster than L1 in the rate of cleavage, and 20-fold more effective than L1 in competing for processing factors [Perscott & Falck-Pedersen, 1992]. Similar to the findings that the sec poly(A) site is utilized only when located upstream of the mb poly(A) site [Peterson & Perry, 1989], the L1 site is used very inefficiently both in early and late infection when it is placed downstream of L3 [Proudfoot, 1991]. These results indicate that the position and strength of poly(A) site is important in determining mRNA biosynthesis. When L1 and L3 were placed in tandem in a minigene construct, the early-to-late poly(A) switch could be reproduced, indicating that the poly(A) sites themselves contain all of the *cis*-acting elements required for this shift [Falck-Pedersen & Logan, 1989]. These *cis*-acting elements may be (i) the variable downstream motifs that determine the stability of the processing complex, and thus the strength of

poly(A) site [Weiss *et al*, 1991], (ii) the upstream element as identified in the L3 site [Prescott & Falck-Pedersen, 1994], the SV40 late [Carswell & Alwine, 1989] and HIV poly(A) sites [Valsamakis *et al*, 1991], or (iii) the sequences immediately surrounding the AAUAAA hexanucleotide [Prescott & Falck-Pedersen, 1994; Ashe *et al*, 1995; Gilmartin *et al*, 1995]. It is surprising that the upstream enhancing elements are all U-rich, suggesting that they may be functionally similar to the downstream GU/U-rich elements. In the case of human GnT II, several UU-, GU-, and AU- rich clusters are identified surrounding the polyadenylation sites (fig 4.1). In addition to these *cis*-acting elements, several *trans*-acting factors have also been described, such as the 64-kDa subunit of CstF in L1/L3 poly(A) site selection [Mann *et al*, 1993] and in regulation of B-cell developmental Ig heavy chain expression [Edwalds-Gilbert & Milcarek, 1995], the 160-kDa subunit of CPSF in HIV-1 mRNA poly(A) site definition, and probably the heterogeneous nuclear ribonucleoprotein C proteins (RNP C) [Wilusz & Shenk, 1990]. However, the role of RNP C has yet to be elucidated.

Two models have been proposed to explain the role of multiple poly(A) sites in complex transcription units. Polyadenylation may be coupled to transcription involving a transcription-coupled scanning mechanism [Niwa *et al*, 1992; Wilson-Gunn *et al*, 1992]. In this model, the promoter-proximal poly(A) site has a processing advantage over more distal sites, simply because of the 5' to 3' directionality of the transcription complex. This property could also be explained by the inherent 5' to 3' directionality of the poly(A) site recognition machinery. The above model requires that the 5' poly(A) site must be less efficient than the 3' ones, otherwise the latter sites would not be utilized. In fact, this is compatible

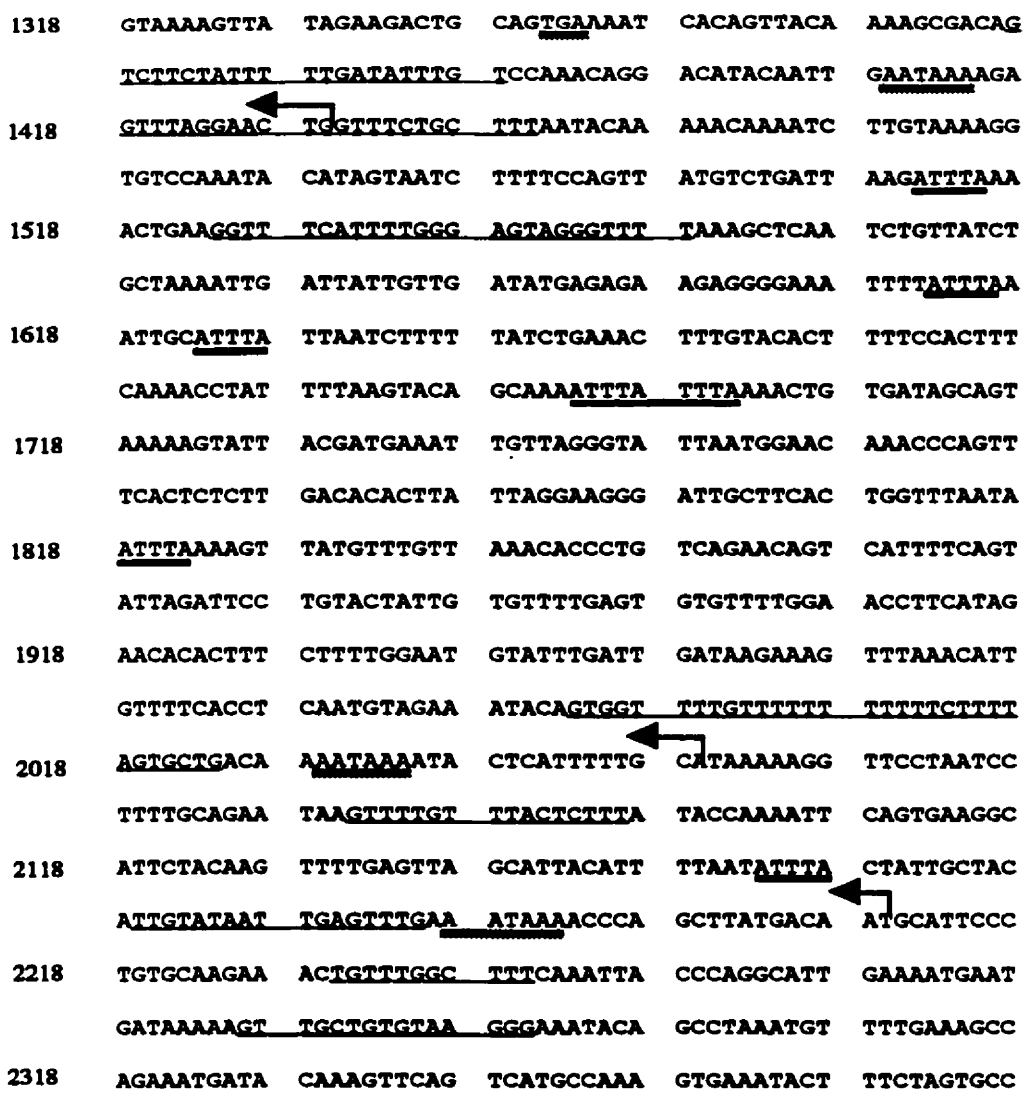


Figure 4.1 Genomic sequence of the human GnT II 3' UTR

Motifs shown in this picture are the G/T- or T-rich clusters (underlined). The translation stop codon and three conserved polyadenylation signals (AATAAA) are shown in hatch underlined. The three cleavage/polyadenylation sites determined by 3' RACE are indicated by arrows. The putative mRNA instability ATTTA motifs are shown in bold underlined.

with the strengths of the L1/L3 and sec/mb poly(A) sites. A second proposed model does not invoke 5' to 3' directionality. It depends on the amount of substrate pre-mRNA, the amount and activity of cleavage/ polyadenylation factors (e.g. CPSF and CstF), site-specific factors that selectively enhance or depress the use of particular poly(A) sites, and the specific sequences surrounding AAUAAA [Mann *et al*, 1993; Edwalds-Gilbert & Milcarek, 1995]. Changes in any of these factors, such as during B-cell development or viral infection, could therefore influence utilization of certain poly(A) sites. In other words, poly(A) site utilization, as suggested during adenovirus infection and IgM production, may be dictated by the stage of development and differentiation. Correlated with the role of complex N-glycans during development as demonstrated by CDGS II and other examples, GnT II expression may be regulated to some extent via the differential utilization of poly(A) sites. However, there is no direct evidence and this hypothesis needs to be further examined.

4.2 Functions of the 3' untranslated region

It has long been thought that the 3' untranslated region (UTR) of mRNA did not have any functional significance. However, an increasing body of evidence has implicated the 3' UTR in controlling polyadenylation, regulating mRNA stability [for review see Jackson, 1993], determining mRNA localization [Gavis & Lehmann, 1992; Mowry & Melton, 1992; Gavis & Lehmann, 1994], modulating mRNA translation [Ostareck-Lederer *et al*, 1994; Tanguay & Gallie, 1996], controlling oocyte maturation and early embryogenesis [Sheets *et al*, 1995; Walker *et al*, 1996], regulating growth and differentiation of myoblasts [Rastinejad & Blau, 1993a], and suppressing tumorogenesis [Rastinejad *et al*, 1993b].

Mutations in 3' UTR have shown to be responsible for some diseases such as myotonic dystrophy [Brook *et al*, 1992; Mahadevan *et al*, 1992].

4.2.1 3'UTR in regulation of mRNA stability

mRNA turnover plays an important role in determining the levels of gene expression. In view of the inability to detect mRNA stabilizers other than secondary structure, the 5' cap and the poly(A) tail, it has been proposed that mRNA is inherently stable in cells, and that the degradation rate of mRNA is controlled by the strength of its destabilizing sequences [for review see Sachs, 1993]. Receiving the greatest attention are the sequences within the 3' UTR. 3' UTR stimulates the poly(A) shortening rate by switching the PAB-dependent poly(A) nuclease from a distributive to a more processive enzyme [Lowell *et al*, 1992]. Shortening of poly(A) can in turn stimulate mRNA degradation via the polyadenylation-dependent decay pathway [Beelman & Parker, 1995]. On the other hand, a sequence-specific binding protein has been reported to interact with its 3' UTR target to stabilize the mRNA by controlling the length of the poly(A) tail [Geneste *et al*, 1996]. Interestingly, an mRNA instability determinant has been localized to the 3' UTR of a plant gene [Gil & Green, 1996].

Extensive searches have identified in the 3' UTR two elements that are involved in the determination of mRNA degradation. One of them is the iron-responsive element (IRE) found on the transferrin receptor mRNA [Klausner *et al*, 1993; Brinder *et al*, 1994]. The binding of IRE-binding protein (IRE-BP) to the stem-loop IRE structures, in response to low intracellular iron concentrations, prevents a sequence-specific endonuclease from targeting an uncharacterized destabilizing sequence in the vicinity of the IRE.

The second 3' UTR destabilizing sequence is the AU-rich element (ARE), originally identified in granulocyte-macrophage colony-stimulating factor (GM-CSF) mRNA [Shaw & Kamen, 1986]. AREs were subsequently found on other short-lived mRNAs, such as c-fos and c-Myc proto-oncogenes, human E-selectin gene and human insulin receptor gene, etc [Sachs, 1993; Chu *et al*, 1994; Levy *et al*, 1995]. The conserved sequence for the AREs is typically defined as one or multiple AUUUA repeats within a U-rich region of mRNA [Shaw & Kamen, 1986]. Although AREs have been demonstrated to be important for mRNA instability for almost a decade, little is understood about the mechanism(s) by which they selectively mediate mRNA turnover. Several AU-rich sequence-binding proteins (AUBPs), originally identified by *in vitro* UV-cross-linking, are thought to be involved in the regulation of mRNA degradation [Bohjanen *et al*, 1991; Bohjanen *et al*, 1992; Gillis & Malter, 1991; Zhang *et al*, 1993]. However, their function is still obscure. The binding affinities of ARE-binding proteins for AREs correlate with the potency of AREs to direct the degradation of mRNAs [DeMaria & Brewer, 1996]. Evidence that rapid mRNA turnover mediated by GM-CSF or c-fos AREs takes place predominantly in the cytoplasm and can be blocked by protein synthesis inhibitors implies that labile mRNA degradation and protein synthesis might be coupled [Shaw, G. & Kamen, R., 1986; Shyu *et al*, 1989]. This hypothesis is further supported by the findings that ribosome binding and ongoing or complete translation of the mRNA coding region are essential for activation of ARE-controlled mRNA turnover [Aharon & Schneider, 1993; Curatola *et al*, 1995; Winstall *et al*, 1995]. However, the detailed linkage between translation and ARE activation requires further investigation. In contrast, evidence has been obtained that some specific

proteins upon binding to AREs stabilize the mRNA [Rajagopalan & Malter, 1994]. This is not common.

In striking contrast to the sequence of the GC-rich promoter, the GnT II 3' UTR is relatively AU-rich. There are 7 conserved AUUUA repeats in the 3' UTR. Six of them reside between the first and second polyadenylation sites, and one lies between the second and third polyadenylation sites (fig 4.1). As predicted by the current consensus of the properties of these sequences, the shortest GnT II transcript should be most stable. This is consistent with our result that 2.0 kb (the shortest message) was the major transcript in human cell lines.

4.2.2 3' UTR in developmental regulation

The regulation of gene expression during development by 3' UTR can be realized through the control of polyadenylation of mRNA [Jackson, 1993; Sheets *et al*, 1995; Stebbins-Boaz *et al*, 1996]. This is particularly evident in *Xenopus laevis*, where there is no detectable transcription from the time of oocyte maturation to the 4,000 cell mid-blastula stage [Stebbins-Boaz *et al*, 1996, and references therein]. Following export from nucleus to cytoplasm, those mRNAs destined to be stored in a translationally dormant state undergo specific poly(A) shortening, usually from ~250 nucleotides to less than 20 bases. At specific stages such as maturation or following fertilization, these mRNAs are translationally activated through an event known as cytoplasmic polyadenylation which requires two *cis* elements in the 3' UTR of the responding mRNAs, the polyadenylation signal AAUAAA and the U-rich cytoplasmic polyadenylation element (CPE) residing within 50 bases upstream of the signal [Jackson, 1993; Stebbins-Boaz *et al*, 1996]. CPE binding protein (CPEB) has recently been shown to

bind several mRNAs essential for oocyte meiosis and maturation [Stebbins-Boaz *et al*, 1996]. The binding of CPEB to CPEs temporally drives the elongation of mRNA poly(A) tails which, through an uncharacterized mechanism(s), stimulate the initiation of translation of the messages.

The function of 3' UTR in the regulation of oocyte maturation or fertilization can also be mediated by modification of specific 3' UTR masking element binding proteins [Jackson, 1993]. It has been demonstrated that phosphorylation of an oocyte 82 kDa protein (p82) by the cdc 2-type kinase at fertilization blocks its binding for the cognate 3' UTRs of some translationally regulated maternal mRNAs, and thereby activates the translation of these genes. This event, together with others, virtually leads to a release from cell cycle arrest and maturation of the clam oocyte [Walker *et al*, 1996].

Another alternative mechanism by which 3' UTR functions in development is through an RNA intermediate that is antisense to the targeting RNA. This was demonstrated by the base pairing of lin-4 RNA with the lin-14 mRNA 3' UTR. The RNA-RNA interaction down-regulates lin-14 translation, and controls temporal pattern formation in *C. elegans* [Lee *et al*, 1993; Wightman *et al*, 1993].

4.3 Promoter region of the human GnT II gene

5' deletion analyses of the human GnT II gene specified that a major promoter element was within the 83 bp region from -636 to -553 relative to ATG as +1 (figs 3.14 and 3.17). A marked drop in CAT activity was observed between plasmids p-636CAT and p-553CAT. Several important transcription factor consensus elements are noted to be present within this 83bp region. The clustering of four Sp1 binding sites suggests that multiple

binding sites may be involved for synergistic activation of the promoter [Rajput *et al*, 1996]. A typical Sp1-like nuclear factor (Sp1, Sp2, Sp3, or Sp4) has been shown to contain a highly conserved zinc finger domain for DNA binding and a glutamine-rich domain for transactivation [Berg, 1992]. The transactivation domain allows these factors to interact with a class of co-activators and other transcription factors [Pugh & Tjian, 1990; Shin *et al*, 1996]. The control of human GnT II , probably a housekeeping gene, by Sp1 is consistent with the concept that Sp1 is not involved in tissue-specific transcription [Rajput *et al*, 1996]. Surprisingly, Sp1 has recently been shown to play an essential role in cell-type-specific gene regulation or in hormone/growth factor-mediated transcription regulation, and its expression level varies several fold in different cells and tissues, especially during development [Vihinen *et al*, 1996]. Furthermore, Sp1 appears to be involved in regulation of gene expression in the nervous system [Cibelli *et al*, 1996]. The differential regulation of human GnT II by Sp1 during development can therefore not be ruled out, since several lines of evidence have suggested the involvement of GnT II in nervous system development [Jaeken *et al*, 1994; Metzler *et al*, 1994].

A CCAAT element, which is targeted by the C/EBP-related family of nuclear transcription factors, is present immediately upstream of the Sp1 binding sites. Previous experiments have implicated the activities of the CCAAT element and inverted CCAAT box in the regulation of both basal and activated gene expression [Mcknight & Tjian, 1986; Dorn *et al*, 1987; Boularand *et al*, 1995; Pittman *et al*, 1995]. For example, the conservation of inverted CCAAT sequence in the cystic fibrosis transmembrane conductance regulator (CFTR) promoter was shown to be required for basal transcription by associating with the transcription initiation site

selection [Pittman *et al*, 1995]. Studies of heterotrimeric G-protein α_{i-2} subunit gene transcription and human fibronectin gene promoter reveal that the CCAAT box is involved in cAMP induced gene expression [Muro *et al*, 1992; Kinane *et al*, 1993]. In addition, the CCAAT box has recently been shown to control, at least in part, γ-globin gene silencing during normal development or the hereditary persistence of fetal haemoglobin (HPFH), a genetically inherited condition in which the fetal γ-globin genes are continually expressed in adult life [Ronchi *et al*, 1996].

An unusual feature of this important GnT II promoter region is the presence of proto-oncogene transcription factor targeting sites, two for c-Myb and another two for Ets proteins. The Ets family consists of about 30 related proteins characterized thus far, all of which contain a DNA-binding 'ets domain' [Wasylyk *et al*, 1993]. Evidence shows that the activity of Ets protein requires cooperation with other transcription factors, such as AP-1, c-Myb, Fos, and Jun, etc [Wasylyk *et al*, 1993; Sieweke *et al*, 1996]. Like many other oncoproteins, Ets family members have been implicated with specific development and differentiation and viral infectious processes [Wasylyk *et al*, 1993]. c-Myb protein is localized in the nucleus, and binds to DNA directly. Using co-transfection experiments, c-Myb has been demonstrated to activate simian virus (SV) 40 transcription in a c-Myb-dependent manner in which c-Myb binds sequence-specifically to a SV 40 enhancer [Nakagoshi *et al*, 1990]. A recent experiment shows that the c-Myb-specific transactivation is mediated in cooperation with the activity of the C/EBP family member, and is further co-activated by the CREB (cAMP response element binding protein) binding protein (CBP) [Oelgeschlåger *et al*, 1996]. The partial overlapping of c-Myb and C/EBP binding sites in the GnT II promoter suggests that these two factors could

interact directly or through a third protein (e.g. CBP). Consistent with the close relationship between Ets and c-Myb proteins, the fusion of which has a dramatic effect on cell transformation, the partial overlap of their binding sites is highly suggestive. It is of interest to note that the binding sites for c-Myb and Ets family have also been reported in the promoter region of GnT V, whose product is related to malignant transformation [Saito *et al*, 1995]. Since the action of GnT II is required for GnT V action, it is possible that the proto-oncogene proteins regulate both GnT II and V during transformation.

Taken together, the properties of transcription factors identified within this short GnT II promoter region suggest that regulation of GnT II gene expression may be involved in a number of aspects of normal development and differentiation. However, there is as yet no direct evidence.

Chapter Five

Summary and future directions

Based on the data presented in this thesis, it can be concluded that human GnT II is encoded by a single exon with multiple putative transcription initiation sites and three distinct polyadenylation sites. The multiple transcripts detected in human tissues and cell lines likely differ only in their 3' untranslated region.

Future work could be directed toward the understanding of the roles of 3' UTR in the GnT II gene. Different GnT II 3' UTRs can be subcloned immediately downstream of the open reading frame of a reporter gene (e.g. CAT), giving rise to hybrid constructs which differ only in their 3' UTRs. These constructs can then be expressed either *in vivo* or *in vitro*. The relative steady-state reporter gene specific mRNA level would indicate the effect of the GnT II 3' UTR on the mRNA stability. S1 nuclease protection assays can be carried out to determine quantitatively the relative strengths of distinct polyadenylation sites. Alternatively, monitoring of the reporter enzymatic activities could be used to reflect the changes of translation efficiency [Prescott & Falck-Pedersen, 1994]. Further experiments may also include the identification of specific proteins that bind to this region by UV-cross-linking. The localization of the binding site may subsequently be determined by deletion or site-directed mutagenesis.

Since a poly(A) site is dissected into an AATAAA hexanucleotide and its surrounding upstream and downstream sequences, the elements responsible for the different 3'-processing efficiencies between these three GnT II poly(A) sites can be identified by way of constructing series of chimeric poly(A) sites with one or two substitutions derived from other poly(A) sites. For instance, considering UnAnDn as indicating the compositions of the upstream sequence, AATAAA and downstream sequence of the n th Poly(A) site in the GnT II gene (n=1, 2 or 3), several

chimeric poly(A) sites such as $U_1A_1D_2$, $U_1A_2D_1$, $U_2A_1D_1$, etc., can be constructed. When these chimeras are placed at the 3'-end of the open reading frame of a reporter gene, the activities of the reporter gene determined following transient transfection could be used to indicate the influence of an individual motif on the 3' processing efficiency. Therefore, *cis*-acting elements that regulate the selection of polyadenylation site can be determined.

To study the promoter region further, other deletion series starting from -636, and additional site-directed mutagenesis experiments are essential for the final discrimination of *cis*-acting elements which regulate GnT II expression. DNase I footprinting, UV-cross-linking and gel mobility shift analyses are also required to define the *trans*-acting factors involved in GnT II regulation.

Finally, since GnT II seems to be regulated during development and differentiation, investigations of its expression in different developmental models might be fascinating.

REFERENCES

Aharon, T. and Schneider, R. J. (1993) Selective destabilization of short-lived mRNAs with the granulocyte-macrophage colony-stimulating factor AU-rich 3' noncoding region is mediated by a cotranslational mechanism. *Mol. Cell. Biol.* **13**, 1971-1980.

Angotti, E., Mele, E., Costanzo, F. and Avvedimento, E. V. (1994) A polymorphism (G -> A transition) in the -78 position of the apolipoprotein A-I promoter increases transcription efficiency. *J. Biol. Chem.* **269**, 17371-17373.

Ashe, M. P., Griffin, P., James, W. and Proudfoot, N. J. (1995) Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev.* **9**, 3008-3025.

Ast, G. and Weiner, A. M. (1996) A U1/U4/U5 snRNP complex induced by a 2'-O-methyl-oligoribonucleotide complementary to U5 snRNA. *Science* **272**, 881-884.

Beelman, C. A. and Parker, R. (1995) Degradation of mRNA in eukaryotes. *Cell* **81**, 179-183.

Berg, J. M. (1992) Sp1 and the subfamily of zinc finger proteins with guanine-rich binding sites. *Proc. Natl Acad. Sci. USA* **89**, 11109-11110.

Bierhuizen, M. F. A., Mattei, M. G. and Fukuda, M. (1993) Expression of the developmental I antigen by a cloned human cDNA encoding a member of a β-1,6-N-acetylglucosaminyltransferase gene family. *Genes Dev.* **7**, 468-478.

Bierhuizen, M. F. A., Maemura, K., Kudo, S. and Fukuda, M. (1995) Genomic organization of core 2 and I branching β-1,6-N-acetyl-

glucosaminyltransferases. Implication for evolution of the β-1,6-N-acetylglucosaminyltransferase gene family. *Glycobiology* **5**, 417-425.

Blair, W. S., Fridell, R. A. and Cullen, B. R. (1996) Synergistic enhancement of both initiation and elongation by acidic transcription activation domains. *EMBO J.* **15**, 1658-1665.

Bohjanen, P. R., Petryniak, B., June, C. H., Thompson, C. B. and Lindsten, T. (1991) An inducible cytoplasmic factor (AU-B) binds selectively to AUUUA multimers in the 3' untranslated region of lymphokine mRNA. *Mol. Cell. Biol.* **11**, 3288-3295.

Bohjanen, P. R., Petryniak, B., June, C. H., Thompson, C. B. and Lindsten, T. (1992) AU RNA binding factors differ in their binding specificities and affinities. *J. Biol. Chem.* **267**, 6302-6309.

Bohm, S., Gum, J. R. Jr., Erickson, R. H., Hicks, J. W. and Kim, Y. S. (1995) Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a TATA-less GC-rich sequence characteristic of a housekeeping gene promoter. *Biochem. J.* **311**, 835-843.

Boularand, S., Darmon, M. C., Ravassard, P. and Mallet, J. (1995) Characterization of the human tryptophan hydroxylase gene promoter (transcriptional regulation by cAMP requires a new motif distinct from the cAMP-responsive element). *J. Biol. Chem.* **270**, 3757-3764.

Brinder, R., Horowitz, J. A., Basilion, J. P., Koeller, D. M., Klausner, R. D. and Harford, J. B. (1994) Evidence that the pathway of transferrin receptor mRNA degradation involves an endonucleolytic cleavage within the 3' UTR and does not involve poly(A) tail shortening. *EMBO J.* **13**, 1969-6980.

Brockhausen, I. (1995) Biosynthesis, 3. Biosynthesis of O-glycans of the N-acetylgalactosamine-α-Ser/Thr linkage type. In *Glycoproteins*

(Montreuil, L., Schachter, H. and Vliegenthart, J. F. G., Eds), New Comprehensive Biochemistry, Vol. 29a, pp 201-259, Elsevier, Amsterdam.

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S.A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J. and Housman, D. E. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**, 799-808.

Buratowski, S. (1994) The basics of basal transcription by RNA polymerase II. *Cell* **77**, 1-3.

Buratowski, S. (1995) Mechanism of gene activation. *Science* **270**, 1773-1774.

Burley, S. K. (1996) Picking up the TAB. *Nature* **381**, 112-113.

Carcamo, J., Buckbinder, L. and Reinberg, D. (1991) The initiator directs the assembly of a transcription initiation factor IID-dependent transcription complex. *Proc. Natl. Acad. Sci. USA* **88**, 8052-8056.

Carswell, S. and Alwine, J. C (1989) Efficiency of utilization of the simian virus 40 late polyadenylation site: Effects of upstream sequences. *Mol. Cell. Biol.* **9**, 4248-4258

Charuk, J. H. M., Tan, J., Bernardini, M., Haddad, S., Reithmeier, R. A. F., Jaeken, J. and Schachter, H. (1995) Carbohydrate-deficient glycoprotein syndrome type II: An autosomal recessive N-acetylglucosaminyltransferase II deficiency different from typical hereditary erythroblastic multinuclearity, with a positive acidified-serum lysis test (HEMPAS). *Eur. J. Biochem.* **230**, 797-805.

Chu, W., Presky, D. H., Swerlick, R. A. and Burns, D. K. (1994) Alternatively processed human E-selectin transcripts linked to chronic expression of E-selectin *in vivo*. *J. Immunol.* **153**, 4179-4189.

Cibelli, G., Schoch, S., Pajunk, H., Brand, I. A. and Thiel, G. (1996) A (G+C)-rich motif in the aldolase C promoter functions as a constitutive transcriptional enhancer element. *Eur. J. Biochem.* **237**, 311-317.

Cohen, S. N. (1995) Surprises at the 3' end of prokaryotic RNA. *Cell* **80**, 829-832.

Curatola, A. M., Nadal, M. S. and Schneider, R. J. (1995) Rapid degradation of AU-rich element (ARE) mRNAs is activated by ribosome transit and blocked by secondary structure at any position 5' to the ARE. *Mol. Cell. Biol.* **15**, 6331-6340.

D'Agostaro. G. A. F., Zingoni, A., Moritz, R. L., Simpson, R. J., Schachter, H. and Bendiak, B. (1995) Molecular cloning and expression of cDNA encoding the rat UDP-N-acetylglucosamine:α-6-D-mannoside β-1,2-N-acetylglucosaminyltransferase II. *J. Biol. Chem.* **270**, 15211-15221.

Dehlin, E., von Gabain, A., Alm, G., Dingelmaier, R. and Resnekov, O. (1996) Repression of beta interferon gene expression in virus-infected cells is correlated with a poly(A) tail elongation. *Mol. Cell. Biol.* **16**, 468-474.

DeMaria, C. T. and Brewer, G. (1996) AUF1 binding affinity to A+U-rich elements correlates with rapid mRNA degradation. *J. Biol. Chem.* **271**, 12179-12184.

Doering, T. L. (1993) Glycosyl phosphatidylinositol-linked membrane proteins: structure, biosynthesis and function. In *Cell surface and extracellular glycoconjugates*, pp 83-117, Academic press.

Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) A multiplicity of CCAAT box-binding proteins. *Cell* **50**, 863-872.

Drapkin, R. and Reinberg, D. (1994a) The essential twist. *Nature* **369**, 523-524.

Drapkin, R., Sancar, A. and Reinberg, D. (1994b) Where transcription meets repair. *Cell* **77**, 8-12.

Dwek, R.A. (1995a) Glycobiology: "Towards understanding the function of sugars". *Biochem. Soc. Trans.* **23**, 1-25.

Dwek, R. A. (1995b) Glycobiology: More functions for oligosaccharides. *Science* **269**, 1324-1325.

Eckner, R., Ellmeier, W. and Birnstiel, M. (1991) Mature mRNA 3' end formation stimulates RNA export from the nucleus. *EMBO J.* **10**, 3513-3522.

Edge, C. J., Joao, H. C., Woods, R. J. and Wormald, M. R. (1993) The conformational effects of N-linked glycosylation. *Biochem. Soc. Trans.* **21**, 452-455.

Edwalds-Gilbert, G. and Milcarek, C. (1995) Regulation of poly(A) site use during mouse B-cell development involves a change in the binding of a general polyadenylation factor in a B-cell stage-specific manner. *Mol. Cell. Biol.* **15**, 6420-6429.

Eilers, M., Schirm, S. and Bishop, J. M. (1991) The MYC protein activates transcription of the α-prothymosin. *EMBO J.* **10**, 133-141.

Englund P. T. (1993) The structure and biosynthesis of glycosyl phosphatidylinositol protein anchors. *Annu. Rev. Biochem.* **62**, 121-138.

Faisst, S. and Meyer, S. (1992) Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res.* **20**, 3-26.

Falck-Pedersen, E. and Logan, J. (1989) Regulation of poly(A) site selection in adenovirus. *J. Virol.* **63**, 532-541.

Fiedler, K. and Simons K. (1995) The role of N-glycans in the secretory pathway. *Cell* **81**, 309-312.

Field, M. C. and Wainwright L. J. (1995) Molecular cloning of eukaryotic glycoprotein and glycolipid glycosyltransferases: a survey. *Glycobiology* **5**, 463-472.

Finger, E. B., Puri, K. D., Alon, R., Lawrence, M. B., von Andrian, U. H. and Springer, T. A. (1996) Adhesion through L-selectin requires a threshold hydrodynamic shear. *Nature* **379**, 266-269.

Flaspohler, J. A. and Milcarek, C. (1990) Myelomas and lymphomas expressing the IgG2a H chain gene have similar transcription termination regions. *J. Immunol.* **144**, 2802-2810.

Fukuda, M. N., Dell, A. and Scartezzini, P. (1987) Primary defect of congenital dyserythropoietic anemia type II: Failure in glycosylation of erythrocyte lactosaminoglycan proteins caused by lowered N-acetylglucosaminyltransferase II. *J. Biol. Chem.* **262**, 7195-7206.

Fukuda, M. N., Masri, K. A., Dell, A., Luzzatto, L. and Moremen, K. W. (1990a) Incomplete synthesis of N-glycans in congenital dyserythropoietic anemia type II caused by a defect in the gene encoding α-mannosidase II. *Proc. Natl Acad. Sci. USA* **87**, 7443-7447.

Fukuda, M. N. (1990b) HEMPAS disease: genetic defect of glycosylation. *Glycobiology* **1**, 9-15.

Galli, G., Guise, J. W., McDevitt, M. A., Tucker, P. W. and Nevins, J. R. (1987) Relative position and strengths of poly(A) sites as well as transcription termination are critical to membrane versus secreted μ-chain expression during B-cell development. *Genes Dev.* **1**, 471-481.

Galli, G., Guise, J., Tucker, P. W. and Nevins, J. R. (1988) Poly(A) site choice rather than splice site choice governs the regulated production of IgM heavy-chain RNAs. *Proc. Natl. Acad. Sci USA* **85**, 2439-2443.

Gavis, E. R. and Lehmann, R. (1992) Localization of *nanos* RNA controls embryonic polarity. *Cell* **71**, 301-313.

Gavis, E. R. and Lehmann, R. (1994) Translational regulation of *nanos* by RNA localization. *Nature* **369**, 315-318.

Geiger, J. H., Hahn, S., Lee, S. and Sigler, P. B. (1996) Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**, 830-836.

Geng, Y. and Johnson, L. F. (1993) Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA-less mouse thymidylate synthase promoter. *Mol. Cell. Biol.* **13**, 4894-4903.

Geneste, O., Raffalli, F. and Lang, M. A. (1996) Identification and characterization of a 44 kDa protein that binds specifically to the 3'-untranslated region of CYP2a5 mRNA: inducibility, subcellular distribution and possible role in mRNA stabilization. *Biochem. J.* **313**, 1029-1037.

Gil, P. and Green, P. (1996) Multiple regions of the *Arabidopsis SAUR-AC1* gene control transcript abundance: the 3' untranslated region functions as an mRNA instability determinant. *EMBO J.* **15**, 1678-1686.

Gillis, P. and Malter, J. S. (1991) The adenosine-uridine binding factor recognizes the AU-rich elements of cytokine, lymphokine and oncogene mRNAs. *J. Biol. Chem.* **266**, 3172-3177.

Gilmartin, G. M., Fleming, E. S., Oetjen, J. and Graveley, B. R. (1995) CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev.* **9**, 72-83.

Goodrich, J. A., Cutler, G. and Tjian, R. (1996) Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell* **84**, 825-830.

Guise, J. W., Lim, P. L., Yuan, D. and Tucker, P. W. (1988) Alternative expression of secreted and membrane forms of immunoglobulin μ chain is regulated by transcriptional termination in stable plasmocytoma transfectants. *J. Immunol.* **140**, 3988-3994.

Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C. and Mattaj, I. W. (1994) The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* **76**, 531-541.

Guo, Z. and Sherman, F. (1995) 3'-end-forming signals of yeast mRNA. *Mol. Cell. Biol.* **15**, 5983-5990.

Guzder, S. N., Giu, H., Sommers, C. H., Sung, P., Prakash, L. and Prakash, S. (1994a) DNA repair gene RAD3 of *S. cerevisiae* is essential for transcription by RNA polymerase II. *Nature* **367**, 91-94.

Guzder, S. N., H., Sung, P., Bailly, V., Prakash, L. and Prakash, S. (1994b) RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* **369**, 578-581.

Holstege, F. C. P., van der Vliet, P. C. and Timmers, H. Th. M. (1996) Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15**, 1666-1677.

Ihara, Y., Nishikawa, A., Toma, T., Soejima, H., Niikawa, N. and Taniguchi, N. (1993) cDNA cloning, expression and chromosomal localisation of human N-acetylglucosaminyltransferase III (GnT III). *J. Biochem.* **113**, 692-698.

Ince, T. A. and Scotto, K. W. (1995) A conserved downstream element defines a new class of RNA polymerase II promoters. *J. Biol. Chem.* **270**, 30249-30252.

Jackson, R. J. and Standart, N. (1990) Do the poly(A) tail and 3' untranslated region control mRNA translation? *Cell* **62**, 15-24.

Jackson, R. J. (1993) Cytoplasmic regulation of mRNA function: the importance of the 3' untranslated region. *Cell* **74**, 9-14.

Jacobson, R. H. and Tjian, R. (1996) Transcription factor IIA: A structure with multiple functions. *Science* **272**, 827-828.

Jaeken, J., Schachter, H., Carchon, H., De Cock, P., Coddeville, B. and Spik, G. (1994) Carbohydrate deficient glycoprotein syndrome type II: a deficiency in Golgi localised N-acetylglucosaminyltransferase II. *Arch. Dis. Child.* **71**, 123-127.

Kageyama, R. and Pastan, I. (1989) Molecular cloning and characterization of a human DNA binding factor that represses transcription. *Cell* **59**, 815-825.

Keller, W. (1995) No and yet to messenger RNA 3' processing! *Cell* **81**, 829-832.

Kinane, T. B., Shang, C., Finder, J. D. and Ercolani, L. (1993) cAMP regulates G-protein α_{i-2} subunit gene transcription in polarized LLC-PK1 cells by induction of a CCAAT box nuclear binding factor. *J. Biol. Chem.* **268**, 24669-24676.

Klausner, R. D., Rouault, T. A. and Harford, J. B. (1993) Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell* **72**, 19-28.

Kleene, R. and Berger, E. G. (1993) The molecular and cell biology of glycosyltransferases. *Biochim. Biophys. Acta* **1154**, 283-325.

Koleske, A. J. and Young, R. A. (1995) The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* **20**, 113-116.

Kornfeld, R and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631-64.

Kumar, R., Yang, J., Eddy, R. L., Byers, M. G., Shows, T. B. and Stanley, P. (1992) Cloning and expression of the murine gene and chromosomal location of the human gene encoding N-acetylglucosaminyltransferase I. *Glycobiology* **2**, 383-393.

Lal, A., Schutzbach, J. S., Forsee, W. T., Neame, P. J. and Moremen, K. W. (1994) Isolation and expression of murine and rabbit cDNAs encoding an α1,2-mannosidase involved in the processing of asparagine-linked oligosaccharides. *J. Biol. Chem.* **269**, 9872-9881.

Lassman, C. R. and Milcarek, C. (1992) Regulated expression of the mouse γ2b Ig H chain gene is influenced by poly(A) site order and strength. J. *Immunol.* **148**, 2578-2585.

Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-854.

Leuther, K. K., Bushnell, D. A. and Kornberg, R. D. (1996) Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: Implications for start site selection and initiation complex formation. *Cell* **85**, 773-779.

Levy, J. R., Hannah, S., Mooney, R. L., Hug, V. and Stevens, W. (1995) Sequence and functional characterization of the terminal exon of the human insulin receptor gene. *Biochim. Biophys. Acta* **1263**, 253-257.

Lo, N. and Lau, J. T. Y. (1996) Transcription of the β-galactoside α 2,6-sialyltransferase gene in B lymphocytes is directed by a separate and distinct promoter. *Glycobiology* **6**, 271-279.

Lopez-Bayghen, E., Vega, A., Cadena, A., Granados, S. E., Jave, L. F., Gariglio, P. and Alvarez-Salas, L. M. (1996) Transcriptional analysis of the 5'-noncoding region of the human involucrin gene. *J. Biol. Chem.* **271**, 512-520.

Lowell, J. E., Rudner, D. Z. and Sachs, A. B. (1992) 3'-UTR-dependent deadenylation by the yeast poly(A) nuclease. *Genes Dev.* **6**, 2088-2099.

Lutz, C. S., Murthy, K. G. K., Schek, N., O'Connor, J. P., Manley, J. L. and Alwine, J. C. (1996) Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. *Genes Dev.* **10**, 325-337.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-Macdonald, J., De Jong, P. J., Wieringa, B. and Korneluk, R. G. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**, 1253-1255.

Mann, K. P., Weiss, E. A. and Nevins, J. R. (1993) Alternative poly(A) site utilization during adenovirus infection coincides with a decrease in the activity of a poly(A) site processing factor. *Mol. Cell. Biol.* **13**, 2411-2419.

McEver, R. P., Moore, K. L. and Cummings, R. D. (1995) Leukocyte trafficking mediated by selectin-carbohydrate interactions. *J. Biol. Chem.* **270**, 11025-11028.

Mcknight S. and Tjian, R. (1986) Transcriptional selectivity of viral genes in mammalian cells. *Cell* **46**, 795-805.

McLauchlan, J., Gaffney, D., Whitton, J. L. and Clements, J. B. (1985) The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* **13**, 1347-1368.

Metzler, M., Gertz, A., Sarkar, M., Schachter, H., Schrader, J. W. and Marth, J. D. (1994) Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantation development. *EMBO, J.* **13**, 2056-2065.

Moremen, K. W. & Robbins, P. W. (1991) Isolation, characterization, and expression of cDNAs encoding murine α-mannosidase II, a Golgi enzyme that controls conversion of high mannose to complex N-glycans. *J. Cell Biol.* **115**, 1521-1534.

Mount, S. M. (1996) AT-AC introns: an ATtACk on dogma. *Science* **271**, 1690-1692.

Mowry, K. L. and Melton, D. A. (1992) Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in *Xenopus* oocytes. *Science* **255**, 991-994.

Murthy, K. G. K. and Manley, J. L. (1995) The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev.* **9**, 2672-2683.

Muro, A. F., Bernath, V. A. and Kornblihtt, A. R. (1992) Interaction of the -170 cyclic AMP response element with the adjacent CCAAT box in the human fibronectin gene promoter. *J. Biol. Chem.* **267**, 12767-12774.

Nakagoshi, H., Nagase, T., Kanei-Ishii, C., Ueno, Y. and Ishii, S. (1990) Binding of the *c-myb* proto-oncogene product to the simian virus 40 enhancer stimulates transcription. *J. Biol. Chem.* **265**, 3479-3483.

Nakatani, Y., Horikoshi, M., Brenner, M., Yamamoto, T., Besnard, F., Roeder, R. G. and Freese, E. (1990) A downstream initiation element required for efficient TATA box binding and *in vitro* function of TFIID. *Nature* **348**, 86-88.

Narimatsu, H. (1994) Recent progress in molecular cloning of glycosyltransferase genes of eukaryotes. *Microbiol. Immunol.* **38**, 489-504.

Niwa, M., MacDonald, C. C. and Berget, S. M. (1992) Are vertebrate exons scanned during splice-site selection? *Nature* **360**, 277-280.

Norman, K. E., Moore, K. L., McEver, R. P. and Ley, K. (1995) Leukocyte rolling *in vivo* is mediated by P-selectin glycoprotein ligand-1. *Blood* **86**, 4417-4421.

Oelgeschlåger, M., Janknecht, R., Krieg, J., Schreek, S. and Lüscher, B. (1996) Interaction of the co-activator CBP with Myb proteins: effects on Myb-specific transactivation and on the cooperativity with NF-M. *EMBO J.* **15**, 2771-2780.

Orkin, S. H. (1995) Transcription factors and hematopoietic development. *J. Biol. Chem.* **270**, 4955-4958.

Ossipow, V., Tassan, J., Nigg, E. A. and Schibler, U. (1995) A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell* **83**, 137-146.

Ostareck-Lederer, A., Ostareck, D. H., Standart, N. and Thiele, B. J. (1994) Translation of 15-lipoxygenase mRNA is inhibited by a protein that binds to a repeated sequence in the 3' untranslated region. *EMBO J.* **13**, 1476-1481.

Pang, J. S., Hung, R., Wu, C., Fang, Y. and Chau, L. (1996) Functional characterization of the promoter region of the platelet-activating factor receptor gene (Identification of an initiator element essential for gene expression in myeloid cells). *J. Biol. Chem.* **270**, 14123-14129.

Peterson, M. L. and Perry, R. P. (1989) The regulated production of μ_m and μ_s mRNA is dependent on the relative efficiencies of μ_s poly(A) site usage and the Cμ4-to-M1 splice. *Mol. Cell. Biol.* **9**, 726-738.

Peterson, M. L., Gimmi, E. R. and Perry, R. P. (1991) The developmentally regulated shift from membrane to secreted μ mRNA production is accompanied by an increase in cleavage-polyadenylation efficiency but no measurable change in splicing efficiency. *Mol. Cell. Biol.* **11**, 2324-2327.

Pittman, N., Shue, G., Leleiko, N. S. and Walsh, M. J. (1995) Transcription of cystic fibrosis transmembrane conductance regulator requires a CCAAT-like element for both basal and cAMP-mediated regulation. *J. Biol. Chem.* **270**, 28848-28857.

Prescott, J. C. and Falck-Pedersen, E. (1992) Varied poly(A) site efficiency in the adenovirus major late transcription unit. *J. Biol. Chem.* **267**, 8175-8181.

Prescott, J. and Falck-Pedersen, E. (1994) Sequence elements upstream of the 3' cleavage site confer substrate strength to the adenovirus L1 and L3 polyadenylation sites. *Mol. Cell. Biol.* **14**, 4682-4693.

Proudfoot, N. (1991) Poly(A) signals. *Cell* **64**, 671-674.

Pugh, B. F. and Tjian, R. (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* **61**, 1187-1197.

Pugh, B. F. and Tjian, R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* **5**, 1935-1945.

Rajagopalan, L. E. and Malter, J. S. (1994) Modulation of granulocyte-macrophage colony-stimulating factor mRNA stability *in vitro* by the adenosine-uridine binding factor. *J. Biol. Chem.* **269**, 23882-23888.

Rajput, B., Shaper, N. L. and Shaper, J. H. (1996) Transcriptional regulation of murine β1,4-galactosyltransferase in somatic cells. *J. Biol. Chem.* **271**, 5131-5142.

Rastinejad, F. and Blau, H. M. (1993a) Genetic complementation reveals a novel regulatory role for 3' untranslated region in growth and differentiation. *Cell* **72**, 903-917.

Rastinejad, F., Conboy, M. J., Rando, T. A. and Blau, H. M. (1993b) Tumor suppression by RNA from the 3' untranslated region of α-tropomyosin. *Cell* **75**, 1107-1117.

Roeder, R. G. (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex. *Trends Biochem. Sci.* **16**, 402-408.

Ronchi, A., Berry, M., Raguz, S., Imam, A., Yannoutsos, N., Ottolenghi, S., Grosveld, F. and Dillon, N. (1996) Role of the duplicated

CCAAT box region in γ-globin gene regulation and hereditary persistence of fetal haemoglobin. *EMBO J.* **15**, 143-149.

Roy, A. L., Meisterernst, M., Pognonec, P. and Roeder, R. G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature* **354**, 245-248.

Rüegsegger, U., Beyer, K. and Keller, W. (1996) Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J. Biol. Chem.* **271**, 6107-6113.

Ruppert, S. and Tjian, R. (1995) Human $TAF_{II}250$ interacts with RAP74: implications for RNA polymerase II initiation. *Genes Dev.* **9**, 2747-2755.

Sachs, A. B. (1993) Messenger RNA degradation in eukaryotes. *Cell* **74**, 413-421.

Sachs, A. & Wahle, E. (1993) Poly(A) tail metabolism and function in eukaryotes. *J. Biol. Chem.* **268**, 22955-22958.

Saito, H., Nishikawa, A., Gu, J., Ihara, Y., Soejima, H., Wada, Y., Sekiya, C., Niikawa, N. and Taniguchi, N. (1994) cDNA cloning and chromosomal mapping of human N-acetylglucosaminyltransferase V. *Biochem. Biophys. Res. Commun.* **198**, 318-327.

Saito, H., Gu, J., Nishikawa, A., Ihara, Y., Fujii, J., Kohgo, Y. and Taniguchi, N. (1995) Organization of the human N-acetyl-glucosaminyltransferase V gene. *Eur. J. Biochem.* **233**, 18-26.

Sambrook, J, Fritsch, E. F. and Maniatis, T. (1989) Construction and analysis of cDNA libraries. In *Molecular cloning* (2nd edition), pp 8.1-8.86, Cold Spring Harbor Laboratory Press.

Sauer, F., Hansen, S. K. and Tjian, R. (1995a) Multiple TAF$_{II}$s directing synergistic activation of transcription. *Science* **270**, 1783-1788.

Sauer, F., Hansen, S. K. and Tjian, R. (1995b) DNA template and activator-coactivator requirements for transcriptional synergism by *Drosophila* bicoid. *Science* **270**, 1825-1828.

Schachter, H. (1992a) Branching of N- and O-glycans: Biosynthetic controls and functions. *Trends Glycosci. Glycotechnol.* **4**, 241-250.

Schachter, H. and Brockhausen, I. (1992b) The biosynthesis of serine (threonine)-N-acetylgalactosamine-linked carbohydrate moieties. In *Glycoconjugates: Composition, structure and function* (Allen, H. J. and Kisailus, E. C., eds), pp 263-332, New York, Marcel Dekker.

Schachter, H. (1994) Molecular cloning of glycosyltransferase genes. In *Molecular Glycobiology* (Fukuda, M. and Hindsgaul, O., eds), pp 88-126, IRL Press.

Schachter, H. (1995a) Biosynthesis, 4b. Substrate level controls for N-glycan assembly. In *Glycoproteins* (Montreuil, L., Schachter, H. and Vliegenthart, J. F. G., Eds), New Comprehensive Biochemistry, Vol. 29a, pp 281-286, Elsevier, Amsterdam.

Schachter, H. (1995b) Biosynthesis, 2c. Glycosyltransferases involved in the synthesis of N-glycan antennae. In *Glycoproteins* (Montreuil, L., Schachter, H. and Vliegenthart, J. F. G., Eds), New Comprehensive Biochemistry, Vol. 29a, pp 153-199, Elsevier, Amsterdam.

Shaw, G. and Kamen, R. (1986) A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659-667.

Sheets, M. D., Wu, M. and Wickens, M. (1995) Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. *Nature* **374**, 511-516.

Shiffman, D., Brooks, E. E., Brooks, A. R., Chan, C. S. and Milner, P. G. (1996) Characterization of the human cyclin-dependent kinase 2 gene (Promoter analysis and gene structure). *J. Biol. Chem.* **271**, 12199-12204.

Shin, E. K., Tevosian, S. G. and Yee, A. S. (1996) The N-terminal region of E2F-1 is required for transcriptional activation of a new class of target promoter. *J. Biol. Chem.* **271**, 12261-12268.

Shyu, A., Greenberg, M. E. and Belasco, J. G. (1989) The *c-fos* transcript is targeted for rapid decay by two distinct mRNA degradation pathways. *Genes Dev.* **3**, 60-72.

Sieweke, M. H., Tekotte, H., Frampton, J. and Graf, T. (1996) MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. *Cell* **85**, 49-60.

Smale, S. T. and Baltimore, D. (1989) The "Initiator" as a transcription control element. *Cell* **57**, 103-113.

Stanley, P., Narasimhan, S., Siminovitch, L. and Schachter, H. (1975) Chinese hamster ovary cells selected for resistance to the cytotoxicity of phytohemagglutinin are deficient in a UDP-N-acetylglucosamine-glycoprotein N-acetylglucosaminyltransferase activity. *Proc. Natl. Acad. Sci. USA* **72**, 3323-3327.

Stebbins-Boaz, B., Hake, L. E. and Richter, J. D. (1996) CPEB controls the cytoplasmic polyadenylation of cyclin, Cdk2 and c-mos mRNAs and is necessary for oocyte maturation in *Xenopus*. *EMBO J.* **15**, 2582-2592.

Struhl, K. (1996) Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell* **84**, 179-182.

Tan, J., D'Agostaro, G. A. F., Bendiak, B., Reck, F., Sarkar, M., Squire, J. A., Leong, P. and Schachter, H. (1995) The human UDP-N-

acetylglucosamine:α-6-D-mannoside-β-1,2-N-acetylglucosaminyl-transferase II gene (*MGAT2*): Cloning of genomic DNA, localization to chromosome 14q21, expression in insect cells and purification of the recombinant protein. *Eur. J. Biochem.* **231**, 317-328.

Tan, J., Dunn, J., Jaeken, J. and Schachter, H. (1996) Mutations in the MGAT2 gene controlling complex N-glycan synthesis cause carbohydrate-deficient glycoprotein syndrome type II, an autosomal recessive disease with defective brain. *Am. J. Hum Genet.* in press.

Tanguay, R. L. and Gallie, D. R. (1996) Translational efficiency is regulated by the length of the 3' untranslated region. *Mol. Cell. Biol.* **16**, 146-156.

Tifft, C. J., Proia, R. L. and Camerinioteró, R. D. (1992) The folding and cell surface expression of CD4 requires glycosylation. *J. Biol. Chem.* **267**, 3268-3273.

Tjian, R. and Maniatis, T. (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**, 5-8.

Valsamakis, A., Zeichner, S., Carswell, S. and Alwine, J. C (1991) The human immunodeficiency virus type 1 polyadenylation signal: A 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation. *Proc. Natl. Acad. Sci. USA* **88**, 2108-2112.

Verrijzer, C. P., Yokomori, K., Chen, J. and Tjian, R. (1994) *Drosophila* $TAF_{II}150$: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science* **264**, 933-941.

Verrijzer, C. P., Chen, J., Yokomori, K. and Tjian, R. (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**, 1115-1125.

Vihinen, T., Määttä, A., Jaakkola, P., Auvinen, P. and Jalkanen, M. (1996) Functional characterization of mouse syndecan-1 promoter. J. *Biol. Chem.* **271**, 12532-12541.

Vliegenthart, J F. G. and Montreuil, J. (1995) Primary structure of glycoprotein glycans. In *Glycoproteins* (Montreuil, L., Schachter, H. and Vliegenthart, J. F. G., Eds), New Comprehensive Biochemistry, Vol. 29a, pp 13-28, Elsevier, Amsterdam.

Voet, D. & Voet, J. G. (1990) Other pathways of carbohydrate metabolism. In *Biochemistry*, Chapter 21, pp 561-585, John Wiley & Sons, New York.

Wahle, E. and Keller, W. (1992) The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.* **61**, 419-440.

Wahle, E. (1995a) 3'-end cleavage and polyadenylation of mRNA precursors. *Biochim. Biophys. Acta* **1261**, 183-194.

Wahle, E. (1995b) Poly(A) tail length control is caused by termination of processive synthesis. *J. Biol. Chem.* **270**, 2800-2808.

Walker, J., Dale, M. and Standart, N. (1996) Unmasking mRNA in clam oocytes: role of phosphorylation of a 3' UTR masking element-binding protein at fertilization. *Developmental Biol.* **173**, 292-305.

Wasylyk, B., Hahn, S. L. and Giovane, A. (1993) The Ets family of transcription factors. *Eur. J. Biochem.* **211**, 7-18.

Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J.* **6**, 3300-3309.

Weiss, E. A., Gilmartin, G. M. and Nevins, J. R. (1991) Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. *EMBO J.* **10**, 215-219.

Werner, M. H., Gronenborn, A. M. and Clore, G. M. (1996) Intercalation, DNA kinking, and the control of transcription. *Science,* **271**, 778-784.

Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal patter formation in *C. elegans*. *Cell* **75**, 855-862.

Wilson, C. J., Chao, D. M., Imbalzano, A. N., Schnitzler, G. R., Kingston, R. E. and Young, R. A. (1996) RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* **84**, 235-244.

Wilson-Gunn, S. I. J. E., Kilpatrick, J. E. and Imperiale, M. J. (1992) Regulated adenovirus mRNA 3'-end formation in a coupled *in vitro* transcription-processing system. *J. Virol.* **66**, 5418-5424.

Wilusz, J. T. and Shenk, T. (1990) A uridylate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substitutes for the downstream element of the polyadenylation signal. *Mol. Cell. Biol.* **10**, 6397-6407.

Winstall, E., Gamache, M. and Raymond, V. (1995) Rapid mRNA degradation mediated by the *c-fos* 3' AU-rich element and that mediated by the granulocyte-macrophage colony-stimulating factor 3' AU-rich element occur through similar polysome-associated mechanisms. *Mol. Cell. Biol.* **15**, 3796-3804.

Wolffe, A. P. and Pruss, D. (1996) Targeting chromatin disruption: Transcription regulators that acetylate histones. *Cell* **84**, 817-819.

Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R. N., Smolyar, A., Reinherz, E. L., and Wagner, G. (1995) Conformation and function of N-linked glycan in the adhesion domain of human CD2. *Science* **269**, 1273-1278.

Yoshimura, M., Nishikawa, A., Ihara, Y., Taniguchi, S. and Taniguchi, N. (1995) Suppression of lung metastasis of B16 mouse melanoma by N-acetylglucosaminyltransferase III gene transfection. *Proc. Natl. Acad. Sci. USA* **92**, 8754-8758.

Zawel, L. and Reinberg, D. (1995) Common themes in assembly and function of eukaryotic transcription complexes. *Annu. Rev. Biochem.* **64**, 533-561.

Zenzie-Gregory B., Sheridan, P., Jones, K. A. and Smale, S. T. (1993) HIV-1 core promoter lacks a simple initiator element but contains a bipartite activator at the transcription start site. *J Biol. Chem.* **268**, 15823-15832.

Zhang, W., Wagner, B. J., Ehrenman, K., Schaefer, A. W., DeMaria, C. T., Crater, D., DeHaven, K., Long, L. and Brewer, G. (1993) Purification, characterization, and cDNA cloning of an AU-rich element RNA-binding protein, AUF1. *Mol. Cell. Biol.* **13**, 7652-7665.