Recent Progress in Molecular Cloning of Glycosyltransferase Genes of Eukaryotes

Hisashi Narimatsu

Division of Cell Biology, Institute of Life Science, Soka University, 1-236 Tangicho, Hachioji, Tokyo 192, Japan

Received April 1, 1994

Key words: Glycosyltransferase genes, Histo-blood type genes, Galactosyltransferase, Sialyltransferase, Fucosyltransferase, N-Acetylglucosaminyltransferase

1. Introduction

The enormous diversity of carbohydrate structures of glycoconjugates is biosynthesized in order by a series of glycosyltransferases. These enzymes have strict substrate specificities. No transferase can utilize more than one type of sugar donor. They can clearly recognize even subtle differences among acceptor molecules and different intersugar-linkages. Hence, it is considered that there are numerous different glycosyltransferases.

They are classified into families according to what kind of donor substrates is utilized for sugar transfer. The nine kinds of sugar nucleotides as donor substrates for eukaryote enzymes are listed in Table 1. The family of the enzymes which transfer the galactose from UDP-galactose (UDP-Gal) to the non-reducing terminus of the sugar acceptor substrate is called galactosyltransferases (GalTs). A UDP-Gal: N-acetylglucosamine \( \beta 1,4 \) galactosyltransferase (\( \beta 1,4 \text{GalT} \)), a member of the GalTs, catalyzes the transfer of galactose from UDP-Gal to the nonreducing end of N-acetylglucosamine (GlcNAc) with the \( \beta 1,4 \) intersugar-linkage. It is a definitively different enzyme from a UDP-Gal: N-acetylglucosamine \( \beta 1,3 \) galactosyltransferase (\( \beta 1,3 \text{GalT} \)) which transfers the galactose to GlcNAc with the \( \beta 1,3 \) -linkage.

It has been difficult to purify these enzymes because they occur in small quantities as membrane-bound forms in Golgi bodies and their activities are quite unstable. The \( \beta 1,4 \text{GalT} \) had been known to be present either as a membrane-bound form in the trans-Golgi cisternae or as soluble forms in body fluids, such as serum (16), colostrum, milk (4, 61) and ascitic fluids. Among many glycosyltransferases, the soluble form of bovine \( \beta 1,4 \text{GalT} \) was purified from milk and its gene was first cloned by us (42) from a cDNA library of bovine mammary glands, based on partial amino acid sequences of the purified enzyme. Almost at the same time, the other group succeeded in cloning the same bovine cDNA by an immuno-screening method (56). The reported sequences of these two clones were completely homologous even though they were cloned independently and by different methods. It was a dawn of molecular biological approach to new fields, named "Glycobiology and Glycotechnology," which were created during these years. The success in \( \beta 1,4 \text{GalT} \) cloning prompted the following researches on the gene cloning for the other glycosyltransferases.

Before the dawn of the molecular biological study, there had been many attractive hypotheses raising biological or biochemical questions on glycosyltransferases themselves and glycoconjugates, that would be solved by molecular biological experiments. The study subjects of interest eight years ago were as follows: 1) how strict the substrate specificities of glycosyltransferases are, 2) what are the genetic mechanisms which produce...
diversity of substrate specificities (50), 3) how different the membrane-bound enzyme and the soluble enzyme in body fluids are, 4) how homologous the primary sequences of the enzymes which share donor substrates are, 5) whether a hypothesis proposing that glycosyltransferases exist on the cell surface as intercellular recognition molecules (19, 49, 58) is correct or not, 6) whether a hypothesis (50) predicting that genes encoding a family of glycosyltransferases are located in mouse T/t complex loci on chromosome 17 is correct or not. None of the above questions could have been solved by any methods except for molecular cloning.

The evidences accumulated during these eight years have answered many questions mentioned above. cDNAs of more than 20 enzymes have been cloned and the genomic structures of some genes have been analyzed. The results of these studies have seemingly cleared many important points at issue. Hypotheses which are too attractive are often false ideas. Idea 6) was one of such wrong hypotheses. As regards 2) and 4), the sequences were not as homologous as expected among the enzymes that share the same donor substrate. There was a daring idea that the diversity of substrate specificities of the enzymes might be generated through genetic rearrangement like immunoglobulin genes, but, it was eventually disproved. Although many proven facts were contrary to expectation, the progress during the last eight years revealed the existence of glycosyltransferases and other findings based on the molecular biological evidence. These events ushered in the new phase of glycobiology.

2. Molecular Cloning of Yeast Glycosyltransferase Genes

The pathway of the dolichol (Dol)-oligosaccharide synthesis for N-linked sugar chain formation in the endoplasmic reticulum (ER) is shown in Fig. 1 (1, 15). Fourteen sugars are transferred to dolichol phosphate (Dol-P) by stepwise reaction of each glycosyltransferase, and Glc$_3$Man$_9$GlcNAc$_2$-PP-Dol is formed. Finally, the completed sugar chain is transferred en bloc from the Dol-oligosaccharide to an asparagine residue in proteins by an enzyme N-oligosaccharyltransferase (OTase) (17). These reactions in the ER are in common with all eukaryotes, from yeasts to animal cells (15, 24). The studies on glycosyltransferases synthesizing this core portion of N-linked sugar chain are promoted by developing temperature sensitive (ts) mutants of Saccharomyces cerevisiae. It is very difficult to obtain animal cell mutants on these enzymes because these mutations must be lethal. The S. cerevisiae mutants defective in this pathway were named alg (asparagine-linked glycosylation) mutants and numbered as shown in Fig. 1.

The ALG7 gene encodes N-acetylglucosamine-1-phosphate transferase (GPT) which catalyzes the first step of Dol-oligosaccharide synthesis. This enzyme catalyzes the transfer of GlcNAc-1-P from UDP-GlcNAc to Dol-P to form GlcNAc-PP-Dol. This activity is inhibited by the antibiotic tunicamycin, leading to cell death. The yeast gene has been cloned by a complementation method to rescue cells when exposed to tunicamycin. The deduced amino acid sequence of this enzyme revealed its topology in the ER as a type IV integral membrane protein, as seen in Fig. 2. The mammalian counterpart of this gene has also been cloned using yeast ALG7 gene as a probe (54, 75). Surprisingly, high sequence homology between yeast ALG7 and the mammalian counterpart gene was observed as the yeast probe revealed a hybridized band in Southern blotting analysis of mammalian DNA. Although the only mammalian glycosyltransferase genes involved in the synthesis pathway of Dol-oligosaccharide which have been cloned is the ALG7 gene counterpart, we might expect that these glycosyltransferase gene sequences would be highly homologous to the yeast counterparts, since they are the common enzymes through-

| Nucleotide-sugar donor substrates for eukaryote glycosyltransferases |
|-----------------------------|-----------------------------|
| UDP-Glc                     | Uridine 5'-diphosphate-glucose |
| UDP-Gal                     | Uridine 5'-diphosphate-galactose |
| UDP-GlcNAc                  | Uridine 5'-diphosphate-N-acetylglucosamine |
| UDP-GalNAc                  | Uridine 5'-diphosphate-N-acetylgalactosamine |
| UDP-GA                      | Uridine 5'-diphosphate-glucuronic acid |
| UDP-Xyl                     | Uridine 5'-diphosphate-xylene |
| CMP-SA                      | Cytidine 5'-monophosphate-sialic acid |
| GDP-Man                     | Guanosine-5'-diphosphate-mannose |
| GDP-fucose                  | Guanosine-5'-diphosphate-fucose |
out all species of eukaryotes. The other alg yeast mutants were isolated by a mannose suicide method (5, 24), in which the mutants defective in incorporation of mannose can survive because of less incorporation of radioactive mannose than wild type cells. Among a series of alg mutants, the ALG1 gene has been cloned and identified as a $\beta_1, 4$ mannosyltransferase ($\beta_1, 4$ ManT) catalyzing the transfer of mannose from GDP-Man to Dol-PP-GlcNAc$_2$ with the $\beta_1, 4$ linkage (2). It is assumed from the primary sequence that the $\beta_1, 4$ ManT has a topology as the type II integral membrane protein in the ER. As shown in Fig. 2, integral membrane proteins are classified into four types depending on the topology, based on their mode of insertion and their orientation within the membrane. The type II membrane-bound proteins have a C$_{\text{out}}$-N$_{\text{in}}$ topology, that is, the carboxyl terminal end of peptide faces the extracellular lumen whereas the amino terminal end faces the cytoplasm. As described later, more than 20 cDNAs of mammalian glycosyltransferases have been cloned, all of which are involved in the synthesis of terminal sequences of glycoproteins and glycolipids in the Golgi. They all have type II topology.

After completion of the assembly of lipid-linked

Fig. 1. Topology of $N$-glycosylation reaction in the endoplasmic reticulum. $\bullet$, dolichol; $\bigcirc$, mannose; ■, $N$-acetylglucosamine (GlcNAc); ▲, glucose. Reference (1).

Fig. 2. Four types of topology of integral membrane proteins. Reference (13).
precursor (Glc3Man9GlcNAc2-PP-Dol) in the ER, its core oligosaccharide is transferred en bloc to the asparagine residue within the possible Asn-X-Ser/Thr consensus sequences in nascent polypeptide chains, catalyzed by OTase (12, 17). The success in cloning of the yeast OTase gene (WBP1) was recently reported using the ts mutant strain (60). It was verified that the enzyme encoded by the WBP1 gene had the catalytic activity of OTase both in vivo and in vitro. As its sequence has not been reported, we do not know the conformation of its protein structure. It would be a useful probe for isolating a mammalian OTase gene by hybridization, as predicted above. The carbohydrate-deficient glycoprotein (CDG) syndrome, which is a human hereditary disease involving defective N-linked glycosylation, has been characterized biochemically (64). The most probable metabolic defect in the CDG patients would be suspected to exist in the early steps of N-linked protein glycosylation, and the low level of OTase would be the cause of this disease. The cloning of the mammalian OTase gene is anticipated for genetic analyses of this syndrome.

After the assembly of Dol-oligosaccharide in the ER which are similar in yeast and mammalian cells, the later stage of oligosaccharide chain processing in the Golgi is, however, quite different between yeast and mammalian cells. In yeast, the large mannose outer chains (Man15-100) are formed in the Golgi. There are several mannosyltransferases (ManT), α1, 2 ManT, α1, 3 ManT and α1, 6 ManT, catalyzing outer chain elongation through a step-wise addition of mannose residue in the Golgi, as shown in Fig. 3. Recently, one of the ManT genes involved in the outer chain synthesis was cloned and named OCH1 gene (38). The deduced amino acid sequence showed the typical type II topology which is common in a group of mammalian glycosyltransferases existing in the Golgi membranes. The OCH1 protein was proved to be one of the ManTs catalyzing the initiation step of outer chain elongation; it transferred a mannose with the α1, 6-linkage to the acceptor substrate (ManαGlcNAc2). For the first step of outer chain elongation, another α1, 6 ManT seems to be required for the transfer of α1, 6 mannose to the inner core (ManαGlcNAc2), to form ManαGlcNAc2 that is the substrate for the OCH1 enzyme. It was also suggested that the OCH1 differed from the α1, 6 ManTs which are responsible for further elongation of the outer chain. The other α1, 6 ManTs should be required for the following outer chain elongation after the OCH1 reaction.

The structures of O-linked sugar chains (O-glycan) of yeast are little known. Only one of the glycosyltransferase genes involved in the synthesis of yeast O-glycans has been cloned (20).

Very few of yeast glycosyltransferase genes have been cloned to date.

3. Molecular Cloning of Mammalian Glycosyltransferase Genes

(a) Cloning Methods for Mammalian Glycosyltransferase Genes

Studies on molecular cloning for mammalian glycosyltransferases started in 1986, when two groups reported success in cloning the bovine α1, 4 GalT gene almost at the same time. Afterwards, there was some confusion since the third group reported a novel cDNA sequence of α1, 4 GalT which had no homology with the former sequences. It was identified later on that the cDNA reported by the third group encoded a kind of CDC kinases, which might exert influence on α1, 4 GalT activity (8), but not the α1, 4 GalT. There are some other reports of cloning of false positives due to contamination with unrelated proteins in the glycosyltransferase preparations during purification.

Since 1986, many investigators rushed into this field and more than 20 mammalian glycosyltransferase genes have been cloned so far. The cDNAs of the enzymes, somehow possibly purified, such as β1, 4 GalT, UDP-Gal: Gal β1, 4 GlcNAc α1, 3 galactosyltransferase (α1, 3 GalT), CMP-sialic acid: β-galactoside α2, 6 sialyltransferase (α2, 6 ST), α1, 3 N-acetylgalactosaminyltransferase (α1, 3...
GalNAcT; a enzyme of ABO blood type) and β1, 2
N-acetylglycosaminyltransferase (β1, 2 GnT;
GnTI) were cloned by the methods of oligo-
nucleotides-probing based on partial amino acid
sequences or antibody screening. Once a cDNA was
cloned, it was very easy to clone cDNAs encoding
the same enzyme in other mammalian species
because the sequences are highly conserved among
species. The full-length β1, 4 GalT cDNAs of
human and mouse have been cloned by probing
with the bovine cDNA (31, 39, 43, 55).

For cloning of the genes encoding enzymes which
were too difficult to be purified, a method of
expression cloning was effectively employed (3). J.
Lowe et al first reported success in cloning two
fucosyltransferase (FucT) genes, α1, 2 FucT (14,
25) and α1, 3/4 FucT (23), by this method. α1, 2
FucT and α1, 3/4 FucT determine the carbo-
hydrate structures known as histo-blood types, H-
substance (O-type) and Lewis type, respectively. A
cDNA library constructed in an expression vector,
CDM8 (3), is first prepared from mRNA from cells
expressing the desired carbohydrate structure on the
cell surface. The plasmid DNAs are transiently
transfected to COS-1 cells in which they are strongly
expressed by the activity of the cytomegalovirus
promotor. In this combination, it is required that
the host cells, COS-1 cells, should have the pre-
cursor structure of carbohydrates and be negative in
the desired structure. In order to enrich the cells
which express the desired structure, sorting methods
by a cell sorter or panning can be employed using
an antibody or a lectin against the carbohydrate
epitope. Plasmid DNAs are rescued by Hirt’s
method from the enriched positive cells and
introduced into bacterial hosts by transformation.
Repeating this cycle, a single plasmid is finally
obtained as a cloned cDNA. Most of the recent
successes in glycosyltransferase gene cloning have
been done by this expression method.

(b) Topology of Glycosyltransferases in the Golgi
Apparatus

From comparison of the amino acid sequences
deduced from cDNA sequences, a consensus on the
domain structure of glycosyltransferases was
reached, as shown in Fig. 4. The domain structure
of β1, 4 GalT is depicted as representative among
them. They are all typical Type II membrane-bind-
ing proteins which have a membrane anchoring
domain at the N-terminal end (13). They all have a
short NH2-terminal cytoplasmic tail consisting of
4-24 amino acids, a transmembrane domain of about
20 amino acids and a stem region which contains a
proline-rich sequence. The stem region does not
appear to form secondary structures like the α-helix
and β-sheet. The length of the stem region varied
among enzymes and was followed by a large
COOH-terminal catalytic domain. The catalytic
domain oriented in the Golgi lumen, as illustrated
in Fig. 4, is sometimes released in soluble form into
body fluids, such as milk and serum, and the amount
of enzyme released depends on disease status and
inflammation. The stem region is a target site for
proteolytic cleavage, and the cleaved catalytic
domains are secreted from the cells. This was
proved by determination of the N-terminal amino
acid sequences of the purified soluble forms of
β1, 4 GalT and α2, 6 ST. Two cleavage sites for
proteases have been demonstrated for β1, 4 GalT (7).
The N-terminal sequences of two soluble forms of
β1, 4 GalT secreted into milk were the same as
those starting at the 79th and 96th amino acids
from the N-terminal end of the membrane-bound
enzyme. Although the kinds of proteases responsi-
bile for the cleavage have not been identified, there
have been some suggestions that cathepsin D-like
proteases within the acidic trans Golgi might be
involved in these reactions.

The soluble forms of β1, 4 GalT have been
detected as a useful prognostic marker in various
malignant tumors including ovarian cancers, but
clinical application of the measurement of β1, 4
GalT activity to diagnosis of malignant tumor had
been restricted because of false positives in
benign diseases. Physicochemical heterogeneity of
this enzyme determined by isoelectric focusing
and nondenaturing polyacrylamide gel electro-
phoresis has been observed by several investigators.
Some reported cancer-associated isoforms of β1, 4
GalT which were electrophoretically detected in
the sera of patients with malignant tumor. More recently, monoclonal antibodies (mAbs) recognizing specifically the cancer-associated isoform, that was named GAT (galactosyltransferase associated with tumor), have been prepared (63). By immunological screening using these mAbs, a cDNA encoding GAT has been successfully cloned (62). Experiments determining antigenic epitopes recognized by GAT-specific mAbs using deletion mutants of the cDNA, revealed that GAT is encoded by the same gene for the normal β1, 4 GalT, but GAT is produced by proteolytic cleavage at the site different from the normal β1, 4 GalT (62). These GAT-specific mAbs will possibly become a useful tool for cancer diagnosis.

(c) Signals Required for Retaining Glycosyltransferases in the Golgi Apparatus

It has recently been demonstrated that some proteins remaining in the ER as permanent residents share a consensus amino acid sequence, KDEL, which is required for retention in the ER (35, 44). Once this tetrapeptide sequence was deleted, the ER resident protein would be secreted. Consensus sequences in glycosyltransferases required for retention in the Golgi were searched for, but there were no homologous sequences predicting the Golgi retention signal. While both type I and II proteins can be found among ER resident proteins, all Golgi resident proteins identified so far have the type II topology, including glycosyltransferases.

To determine where the Golgi retention signals are located in the primary sequences of glycosyltransferases, experiments have been done by constructing chimeric cDNAs to generate hybrid proteins between some glycosyltransferases and the other type II cell surface membrane-bound proteins. In all of these experiments, it was reported that the minimal sequence required for Golgi retention exists in the transmembrane domain consisting of 16-20 amino acids (9-11, 32, 36, 45, 51, 69). Hybrid proteins possessing this domain could remain in the Golgi, whereas any other constructions in which this domain was either lacking or had been replaced by the transmembrane domain of cell surface-bound proteins were sorted to the secretion pathway or to the cell surface. Although molecular mechanisms of the Golgi retention of glycosyltransferases remain unclear, there are hypotheses. 1) There might be specific receptors for glycosyltransferases to retain them in the Golgi, based on the findings of the receptors for KDEL consensus sequence for ER retention, but this is unlikely since we could not find such consensus sequences among glycosyltransferases cloned to date. If receptors for anchoring glycosyltransferases were present in the Golgi, each enzyme would have different receptors. 2) Hydrophobic profiles and lengths of amino acid residues in the transmembrane domain might determine their localization, whether they are retained in the Golgi or sorted to the cell surface. This possibility was recently demonstrated by Qasba’s group (32). They constructed several mutant cDNAs encoding proteins changed in hydrophobicity of the transmembrane domain. According to their results, the increase in the length of hydrophobic amino acid residues overcame the Golgi retention signal and directed the enzyme to the plasma membrane. They concluded that the length of the hydrophobic region of the transmembrane domain is one of the parameters determining Golgi retention, but is not sufficient by itself.

(d) Galactosyltransferases

Glycosyltransferases utilizing UDP-Gal as the donor substrate are collectively called galactosyltransferases (GalTs). Three kinds of GalTs have been cloned, as listed in Fig. 6. The cDNAs of β1, 4 GalT were first cloned as described previously. Then, α1, 3 GalT cDNAs have been cloned by two different groups independently (22, 26). The third GalT is a B-type enzyme (B transferase) which determines B-type antigens on human erythrocytes in the human histo-blood type ABO system.

A large amount of lactose (Gal β1, 4 Glc) is secreted into the milk after it is synthesized by lactose synthetases in epithelial cells of lactating mammary glands. Two proteins are known to synthesize lactose, lactose synthetase A and B. The milk lactose synthetase A protein is capable of recognizing glucose as an acceptor in the presence of the B protein (α-lactalbumin; α-LA) and forms lactose, whereas, in the absence of α-LA, it transfers galactose to GlcNAc and synthesizes lactosamine (Gal β1, 4 GlcNAc). The first cloning of cDNA encoding β1, 4 GalT was done according to the partial amino acid sequences of the bovine lactose synthetase A (42). The Gal β1, 4 GlcNAc inter-sugar-linkage group occurs not only in the lactosamine but also in sugar chains of many mammalian glycoproteins, particularly asparagine-linked glycoproteins. In addition, this structure has been found in sugar chains of glycolipids, milk oligosaccharides, keratan sulfates and so on.
Glycoconjugates possessing the Gal β1, 4 GlcNAc intersugar-linkage are widely distributed throughout the human body. There had been a question whether the lactose synthetase A in mammary glands would be the same enzyme as those forming the Gal β1, 4 GlcNAc intersugar-linkage in the other tissues.

As Northern analyses revealed that β1, 4 GalT transcripts were ubiquitously expressed in a variety of tissues, the following isolation of β1, 4 GalT cDNA had been done by probing cDNA libraries of mouse F9 teratocarcinoma cells (39) and human placental cells (31, 43) with the bovine β1, 4 GalT cDNA. To determine the substrate specificities of β1, 4 GalT, cDNAs were expressed both in mamalian cells, COS cells (33, 40), and in Escherichia coli (41). Recombinant β1, 4 GalT expressed in E. coli had the same apparent K m values toward sugar nucleotide and sugar acceptors as those of the enzyme expressed in COS-1 cells (40, 41). It could synthesize a series of glycoconjugates, such as lactosamine, lactose in the presence of αLA, N-glycans, poly-N-acetyllactosamine and paragloboside (41). These results suggested that β1, 4 GalT, which had been cloned as the lactose synthetase A, functions in synthesizing lactose mainly in mammary gland, and is also involved in other tissues in the syntheses of Gal β1, 4 GlcNAc linkage of Asn-linked sugar chains and poly-N-acetyllactosamine chains of glycoprotein, in addition to the same linkage seen in glycolipids.

The α1, 3 GalT transfers galactose from UDP-Gal to the nonreducing terminus of a galactose residue of N-glycans, recognizing a minimal structure of Gal β1, 4 GlcNAc on the terminus. Once a terminal structure such as Gal α1, 3 Gal β1, 4 GlcNAc-R is formed by this enzyme, no further elongation of the N-glycan chain occurs. It had been known that catarrhines, which are Old World monkeys, apes and humans lack the α1, 3 GalT activity resulting in nonexpression of the carbohydrate epitope Gal α1, 3 Gal β1, 4 GlcNAc-R, whereas lower mammalian species except for catarrhines abundantly express both the α1, 3 GalT activity and this epitope. In this sense, this carbohydrate epitope is a species-specific antigen. It has been noted that significant amounts of naturally occurring antibody directed against this epitope are contained in the sera of catarrhines, but not of nonprimate mammals. The genes encoding this enzyme were first cloned from the mouse and cow in which the gene actively functions. Then, two of human genes homologous to the murine and bovine sequences were cloned and found to be inactivated by frameshift and nonsense mutations (27). This was the first evidence that clarified the molecular basis of species-specific carbohydrate epitopes. Sequence analyses of this gene in various primates suggested that it was inactivated in catarrhines after the evolutionary divergence of apes from monkeys.

Sequence comparison between two bovine GalTs, β1, 4 GalT and α1, 3 GalT, revealed that they do not share highly homologous sequences even though both enzymes utilize the same donor substrate, UDP-Gal. This was contrary to the expectation that all GalTs might have a highly conserved sequence required for UDP-Gal binding. However, looking carefully at the aligned sequences of the catalytic domains of the two enzymes, some conserved peptides could be found, as shown in Fig. 5 (70). Chemical labeling experiments on the purified β1, 4GalT conducted to identify the UDP-Gal binding region within the catalytic domain, suggested that two lysine residues, Lys 341 and Lys 351, seemed to be involved in binding to UDP-Gal. The region between Lys 341 and Lys 351 was thought to be associated with UDP-Gal binding (70). As seen in Fig. 5, some of the amino acid residues in this region are conserved in both the GalTs, β1, 4 GalT and α1, 3 GalT. It is also known that a disulfide linkage between two cysteine residues, C 134 and C 247, is required for β1, 4 GalT activity (71). More detailed analyses are required to determine its tertiary structure, and finally success in crystallization would conclusively determine the structure of the enzyme.

Even though the histo-blood type ABO system was discovered a long time ago, almost a century passed before the genes determining the ABO system were cloned. A cDNA encoding an A-type enzyme (A transferase) was first cloned in 1990 by oligonucleotide probing after obtaining peptide sequences of the purified enzyme (73). The third GalT, B transferase, was easily cloned by probing with the A transferase cDNA since the A and B transferase genes differ by only a few single-base substitutions within the 1,059-bp ORF, changing four amino acid residues. B transferase catalyzes galactose transfer to the galactose residue of Fuc α1, 2 Gal-R with the Gal α1, 3 Gal linkage. This is a quite similar reaction to that of α1, 3 GalT except that the acceptor substrate for B transferase has an additional fucose residue. The two enzymes, α1, 3 GalT and B transferase, form the Gal α1, 3 Gal intersugar-linkage in addition to sharing the donor substrate. Sequence comparison between α
496 H. NARIMATSU

1, 3 GalT and B type GalT genes revealed that they are phylogenically related genes, sharing highly homologous sequences within their catalytic domains. They have about 50% homology at the amino acid sequence level in the catalytic region but no homology outside this region.

All three GalT genes were reported to be localized to human chromosome 9, on which \( \beta 1, 4 \text{GalT} \), \( \alpha 1, 3 \text{GalT} \) and B type GaIT genes were mapped to 9p13, 9q33-q34 and 9q34, respectively. In addition, a gene encoding galactose-l-phosphate uridyltransferase, an enzyme involved in the synthesis of UDP-Gal, has also been mapped closely to the positions of the GaIT genes. This suggested that these genes involved in the galactose transfer are evolutionarily related. Considering the similarity in substrate specificities and sequence homology among these GaITs, they seemingly originated from the same ancestral gene and then diverged by gene duplication on chromosome 9. A hypothetical phylogenic tree indicating the divergence of GaITs is illustrated in Fig. 6. The \( \beta 1, 4 \text{GalT} \) gene first separated from the \( \alpha 1, 3 \text{GalT} \) gene, and then divergence of the B gene followed quite recently. As described later, A and O genes which determine A and O antigens on human erythrocytes respectively were generated from the B gene by point mutations.

\((e)\) Sialytransferases

Sialic acids on the terminal ends of carbohydrate chains of glycoconjugates play important biological roles as determinants involved in recognition events, such as adhesion of leukocytes and tumor cells to endothelial cells through recognition by selectins, binding of influenza virus to host cells and cell to cell interaction during embryogenesis the intensity of which is dependent on the length of polysialic acids in neural cell adhesion molecule (N-CAM).

Sialytransferases (ST) are a family of enzymes that catalyze the transfer of sialic acid (SA) from CMP-sialic acid (CMP-SA) to terminal positions on sugar chains of glycoproteins and glycolipids. The addition of SA on sugar chains terminates the chain elongation, except for polysialic acid chains as found in N-CAM and gangliosides. \( \alpha 2, 6 \text{ST} \) was the first ST cDNA cloned according to partial amino acid sequences of the purified enzyme (65); it catalyzes the transfer of SA to the terminal positions of many \( N \)-linked carbohydrate groups and forms the terminal structure, SA \( \alpha 2, 6 \text{Gal} \beta 1, 4 \text{GlcNAc-R} \). The gene encoding this enzyme spans 40 kb of genomic DNA and contains 6 exons (59). Compared with the genomic organization of \( \alpha 2, 6 \text{ST} \) with that of \( \beta 1, 4 \text{GalT} \) (21, 34), both genes were found to be very similar in size and exon...
organization, even though they have no sequence homology. This similarity may suggest correlations between exon organization and protein domains.

cDNAs for two different \( \alpha_2, 3 \) STs which transfer SA to acceptors with different intersugar-linkage from that of \( \alpha_2, 6 \) ST have been cloned after purification of the enzymes (18, 66). Although both of the \( \alpha_2, 3 \) STs transfer SA to the acceptor with the \( \alpha_2, 3 \) intersugar-linkage, they exhibit different specificities for the sequence of the acceptor sugar chains. Gal \( \beta_1, 3 \) GalNAc \( \alpha_2, 3 \) ST (ST3O) is an enzyme synthesizing the structure of \( \alpha_2, 3 \) Gal \( \beta_1, 3 \) GalNAc-R which is found in \( O \)-linked sugar chains of glycoproteins. Northern blot analysis revealed that ST3O is abundant in the salivary gland, liver, lung, and colon mucosa in which mucin-type glycoproteins possessing \( O \)-linked sugar chains are abundantly produced. The other \( \alpha_2, 3 \) ST, Gal \( \beta_1, 3 \) (4) GlcNAc \( \alpha_2, 3 \) ST (ST3N), which forms the structures of SA \( \alpha_2, 3 \) Gal \( \beta_1, 3 \) GlcNAc-R or SA \( \alpha_2, 3 \) Gal \( \beta_1, 3 \) GlcNAc-R has been cloned. These sugar sequences are mainly found on the termini of \( N \)-linked sugar chains, although they are also contained in \( O \)-linked sugar chains and in glycolipids.

A striking finding appeared from sequence comparison among the three STs, \( \alpha_2, 6 \) ST, ST3O and ST3N. They had a homologous region in a stretch of 55 amino acids from residue 156 to 210 within 374 amino acids in total located in their catalytic domains. The significant homologies among the three STs are shown in Fig. 7 (18, 66), although the remaining sequences outside this region lacked homology. The highly conserved sequences named as a sialyl motif enabled a polymerase chain reaction (PCR) strategy for cloning of cDNAs encoding the other STs (28). Primers designed for annealing to the 5' and 3' ends of the sialyl motif were used for PCR amplification on a mouse brain cDNA library from which PCR products (150 bp) were subcloned. Using one of the PCR products as a probe for screening the library, a full-length cDNA encoding a new \( \alpha_2, 3 \) ST has been isolated, and its product expressed in mammalian cells has been analyzed. The amino acid sequence deduced from the newly isolated cDNA clone, named STMB, showed 80% identity with that of ST3O. The STMB exhibited \( \alpha_2, 3 \) ST activity toward only the disaccharide moiety of Gal \( \beta_1, 3 \) GalNAc of glycoproteins and glycolipids. In \textit{in vitro} experiments for determining the substrate specificities of STMB, the STMB expressed in the mammalian cell could synthesize gangliosides, such as GM1b, GD1b and GT1b from asialo-GM1, GM1a and GD1b, respectively. It is not clear whether the STs for ganglioside synthesis are different from the STs for glycoproteins. STMB may be a candidate of \( \alpha_2, 3 \) STs responsible for synthesizing gangliosides, whereas STO3 is an \( \alpha_2, 3 \) ST responsible for mucin type glycoproteins. Some of the cDNA clones for novel STs are now being cloned by PCR based on the sialyl motif and are being characterized (personal communication). In a few years, more than ten of cDNAs for STs will be identified by this method.

Another success in cDNA cloning for ST using the expression cloning method has been reported (53). A cDNA library constructed in an Epstein-Barr virus-based cloning vector was transfected into a B-cell line. The cytotoxic lectin \textit{Ricinus communis} agglutinin 120, specific to galactose residues (especially to the Gal \( \beta_1, 4 \) GlcNAc structure), was used for positive selection of the cells in which galactose residues were sialylated and unexposed. Such cells could escape from killing by the lectin and grow in the culture. A cDNA encoding for a ST was finally isolated from the lectin resistant cells. The cDNA sequence revealed that the sialyl motif was shared by other STs, but there were no other regions with homology. The expression experiments confirmed that it is a new \( \alpha_2, 3 \) ST capable of synthesizing sialyl Lewis x (sLex) \textit{in vivo}. The sLex epitope is reported as a biologically active determinant as a ligand for selectins, E-, P- and L-selectin, which are cell adhesion molecules involved in leukocyte recruitment and the attachment of malignant cells to endothelial cells.

Both the PCR and expression cloning methods will be useful techniques for cloning other STs.
(f) *Fucosyltransferases*

Fucosyltransferases (FucTs) are the enzymes transferring fucose (Fuc) from GDP-Fuc to Gal, glucose (Glc) and GlcNAc via their 2, 3, 4 or 6 hydroxyls to produce a variety of structures. Expression cloning methods were employed for isolation of genes encoding FucTs.

A structure, Fuc α1, 2 Gal-R, is present on type I, Gal β1, 3 GlcNAc, and type II, Gal β1, 4 GlcNAc, precursors and the H antigen whose expression is normally restricted to the surfaces of human O-type erythrocytes and a variety of epithelial cells, urinary and respiratory tracts. To clone a genomic DNA encoding the α1, 2 FucT responsible for the H antigen expression, mouse L cells were chosen as host cells to be transfected with genomic DNAs isolated from a human cell line (A431) that has α1, 2 FucT activity since they have no H antigen on the cell surface. As L cells have H antigen precursors, the cells which received the α1, 2 FucT gene became positively stained with an anti-H antibody. The positive cells were enriched by the panning method and fluorescence-activated cell sorting, and then the human DNA fragment containing the α1, 2 FucT gene in the mouse cells was separated from mouse DNA by probing with a human-specific Alu-sequence probe (14). Finally, a cDNA encoding α1, 2 FucT was isolated from a human cDNA library by hybridization with the above genomic DNA probe (25). There are two phenotypes in individuals regarding the H antigen, secretors and non-secretors. The former expresses the ABH blood group antigens not only on the erythrocytes but also in bodily fluids such as the saliva and milk, whereas the latter has them only on the erythrocytes. It is generally believed that the secretor gene, Se gene, is a regulatory gene controlling expression of the α1, 2 FucT gene in epithelial cells which secrete H substances. In contrast, it has also been speculated that the H and Se genes are both structural genes encoding two different α1, 2 FucTs, one of which is mainly expressed in hematopoietic cells and the other in epithelial cells. Recent evidence supports the latter hypothesis whereby the Se gene is a structural α1, 2 FucT gene which might have a homologous sequence to the H gene. Successful cloning of the Se gene will be soon done by hybridization probed with the H gene.

The human Lewis histo-blood type system comprises two major antigens, Leα and Leβ. They were first discovered on erythrocytes, and later recognized in the plasma, saliva and other secretions. The structures of Leα and Leβ determinants are synthesized by the transfer of Fuc to the GlcNAc of type I precursor and to the H type I structure, respectively, with the α1, 4-linkage, as shown in Fig. 8. The gene synthesizing Lewis histo-blood type antigens, α1, 4 FucT gene, was cloned by the expression cloning method (23). The enzymatic characteristics determined by the recombinant product directed by this cDNA revealed a unique feature in that it can use both type I and type II oligosaccharide chains as acceptor substrates to produce the Fuc α1, 4- and α1, 3-linkages, respectively. This is an exceptional glycosyltransferase which has two linkage specificities, and for this reason it is called α1, 3/4 FucT. Southern blotting analysis proved with the Lewis gene (α1, 3/4 FucT gene) showed multiple bands suggesting the existence of highly homologous genes which may form a gene family of α1, 3 FucTs. This coincided with the results of enzyme purification experiment that had predicted the existence of several kinds of α1, 3 FucTs, of which acceptor specificities had revealed subtle differences (30). So far, three α1, 3 FucT genes in addition to the Lewis gene have been cloned by a hybridization method using the Lewis gene as a probe. They are numbered in order of date of isolation, as FucT III (Lewis type enzyme), IV (29), V (67) and VI (46, 68). Three α1, 3 FucT genes (FucT III, V and VI) were mapped on human chromosome 19, two of which (III and VI) were in close proximity to each other (46). The three genes on chromosome 19 had quite high sequence similarities, particularly in the carboxyterminal catalytic domain (Table 2), in which sequence identity at an amino acid level is more than 91%. The FucT IV gene was mapped on human chromosome 11 and had less sequence similarity to FucT III than to the others, although the amino acid sequence was 70%
homologous, hybridizable enough for cloning by the FucT III probe. It has been suggested that a gene duplication event first occurred and separated the α1, 3 FucT genes into two chromosomes, 11 and 19, followed by further gene duplications on each chromosome. Southern blot analysis of the FucT IV gene identified several bands besides FucT III, V and VI when carried out under conditions of decreased stringency, suggesting the existence of another multi-gene family, probably on chromosome 11, in addition to the family on chromosome 19. More members of α1, 3 FucT genes with sequences highly homologous to FucT IV are anticipated.

The expressed enzymes in mammalian cells directed by four FucT cDNAs (III, IV, V and VI) were characterized to determine their acceptor specificities. As summarized in Table 2, the enzymes had only the α1,3-linkage specificity by transferring Fuc to the type II chain generating Lex structures Gal β1, 4 [Fuc(α1, 3)] GlcNAc-R, except for FucT III, which has both α1,3- and α1,4-linkage specificities. FucT III, V, and VI could transfer Fuc to the sialylated type II chain and synthesized sLeα, SA α2, 3 Gal β1, 4 [Fuc(α1, 3)] GlcNAc-R, whereas FucT IV could not. The VIM2 epitope SA α2, 3 Gal β1, 4 GlcNAc β1, 3 Gal β1, 4 [Fuc(α1, 3)] GlcNAc-R (Fig. 9), was synthesized by the transfer of Fuc to the internal GlcNAc residue, not to the terminal GlcNAc. Three FucTs (III, IV and V) other than FucT VI, had the ability to produce the VIM2 epitope. Recently, it was orally reported that another α1, 3 FucT, named FucT VII, *was cloned by hybridization with a FucT IV probe. It differed from the other α1, 3 FucTs in that it was able to synthesize the sLeα determinant, but not the nonsialylated determinant, Leα. The genomic organization of FucT VII was analyzed and shown to contain multiple exons, whereas the previously cloned FucT III, IV, V and VI genes had no intron in their ORFs which is unusual among mammalian genes.

The biochemical information of purified enzymes from tissues had permitted the subdivision of human α1,3 FucTs into four types based on their physicochemical differences: Lewis type, myeloid type, plasma type and lung type. By comparison of fine specificities between purified enzymes and recombinant enzymes produced by their cDNA expression, it was revealed that the Lewis-type enzyme and the myeloid-type enzyme correspond to the products of FucT III and FucT IV genes, respectively, although the genes producing the plasma-type and lung-type enzymes remain uncertain.**

Table 2. Human α1,3 fucosyltransferase family

<table>
<thead>
<tr>
<th>FucT</th>
<th>Homology to FucT III(%)*</th>
<th>Chromosomal localization</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>III; (α1, 3/4); Lewis type</td>
<td>100</td>
<td>19</td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>IV; (α1, 3); Myeloid type</td>
<td>70</td>
<td>11</td>
<td>+ - - - - +</td>
</tr>
<tr>
<td>V; (α1, 3); ? type</td>
<td>91</td>
<td>19</td>
<td>+ + - - - +</td>
</tr>
<tr>
<td>VI; (α1, 3); plasma-type</td>
<td>91</td>
<td>19</td>
<td>+ + - - - +</td>
</tr>
<tr>
<td>VII; (α1, 3); ? type</td>
<td>42</td>
<td>9</td>
<td>- - - - - ?</td>
</tr>
</tbody>
</table>

* Amino acid level homology.

\[ \text{NecAc} \alpha 2,3 \text{Gal} \beta 1,4 \text{GlcNAc} \beta 1,3 \text{Gal} \beta 1,4 \text{GlcNAc} \beta - R + \text{GDP-Fuc} \]

Footnote: The paper regarding the FucT VII gene cloning* and the paper identifying the plasma-type FucT** were published during preparation of this manuscript.


It is of interest to know which FucT is responsible for synthesis of sLex determinants functioning as ligands for E-selectin on leukocytes and leukaemic cells. Although the transcripts of FucT IV (myeloid-type) were expressed at a significant level in all of the leukaemic cell lines examined, their role is still controversial since the recombinant FucT IV product showed an apparent inability to act on sialylated type II chain precursors (Table 2). It was suggested that the newly cloned FucT VII gene might code for sLex in the myeloid cell lineage. The sLex and sLea determinants are also expressed on malignant tumor cells and are thought to function as ligands for selectins, and these ligands may be involved in the mechanism of cell adhesion associated with hematogenous metastasis of cancers.

To examine which of α1, 3 FucTs is most responsible for expression of the sLex and sLea in malignancy, the levels of transcripts of various α1, 3 FucT genes were determined in a large number of tumor cell lines (72). It is noteworthy that the FucT III and VI transcripts were consistently observed at significant levels in all of the tumor cell lines. Activation of the FucT III and VI genes could be involved in malignant transformation by generating the ligands for selectins.

### (g) Glycosyltransferases for Determination of Histo-Blood Types

The epitopes of histo-blood type ABO antigens are determined by carbohydrate structures, as shown in Fig. 8. A individuals express α1, 3 GalNAcT (A transferase) activity whereas B individuals express α1, 3 GalT (B transferase). AB individuals express both activities and O individuals express neither. The difference between A and B genes was found to be a few base-substitutions resulting in the four amino acid substitutions within 353 amino acids encoded by their ORF (73). It was also proved that the FucT III gene has been considered to control the expression of Lewis antigens on erythrocytes and in secretions. To clarify the molecular genetic basis of Lewis blood group system, FucT III genes from Lewis negative, Le(−), individuals have been analyzed (47). In comparison of FucT III gene sequences between Le (+) and Le (−) individuals, two single base substitutions resulting in amino acid substitutions were found in Le (−) FucT III genes, one located in the transmembrane domain and the other in the catalytic domain. Furthermore, it was demonstrated that the single amino acid substitution in the catalytic domain is a cause of abolition of the FucT III activity, both α1, 3 and α1, 4 activities. This is the first case demonstrating that the single amino acid substitution inactivates enzyme activities to determine histo-blood type phenotype. It is now able to identify the ABO and Lewis genotypes by PCR amplification of genomic DNA followed by restriction enzyme digestions specifically recognizing point-mutated differences.

### (h) N-Acetylgalactosaminyltransferase

N-acetylgalactosaminyltransferases (GnTs) involved in the synthesis of N-glycans are named as shown in Fig. 10, of which, the genes for GnT I (52), III (48) and V (57) have been cloned. The GnT I is the enzyme that catalyzes an essential first step in the conversion of high-mannose N-glycans to hybrid and complex N-glycans. Recent success in targeted disruption of the GnT I gene clarified a crucial function of this gene during embryogenesis. Mouse embryos homozygous for the disrupted GnT I gene exhibited lethality at day 9.5 of gestation, although loss of GnT I gene function did not interfere with pre-implantation or early post-implantation development.

GnT III catalyzes the transfer of a GlcNAc residue to β1, 4 mannose in the core region of N-glycans and forms a bisecting GlcNAc. This enzyme

![Fig. 10. Numbering N-acetylgalactosaminyltransferase (Gn T) involved in syntheses of N-glycan. Gn, N-acetylgalactosamine; Man, mannose.](image-url)
is thought to be important in the regulation of the 
N-glycan biosynthesis since the bisecting GlcNAc 
formed by GnT III strongly inhibits subsequent 
enzyme reactions, such as β1, 6 FucT, GnT IV and 
GnT V.

GnT V is of particular interest because many cell 
lines transformed by tumor viruses and oncogenes 
exhibited a significantly increased level of this 
enzyme activity. Structural changes of N-glycans 
produced by the action of this enzyme are consid 
ered to correlate with changes in the metastatic 
potential of malignant cells. Comparison of the 
GnT V sequence with those of GnT I and III 
revealed no significantly similar region as observed 
in the sialyltransferase family.

One of the GnTs involved in the synthesis of O-
glycans has been cloned by the expression cloning 
method (6). This GnT forms a core� (C2) structure, 
Gal β1, 3 (GlcNAc β1, 6) GalNAc-R, which is 
often observed in O-glycans. The formation of 
the C2 branching structure is the first step critical 
for the following extension of poly-N-acetyllactos 
amine chains. Increased activity of this enzyme, 
C2GnT, has been observed in a variety of biologi 
cal processes, such as T-cell activation and immuno-
deficiency due to Wiskott-Aldrich syndrome 
and AIDS. The recombinant enzymes directed by 
the C2GnT cDNA had a fine specificity to the Gal 
β1, 3 GalNAc acceptor, but not to GlcNAc β1, 3 
GalNAc, GlcNAc β1, 2 Man and GlcNAc β1, 3 
Gal.

Another GnT which also forms a branching 
structure by transferring GlcNAc to the GlcNAc 
β1, 3 Gal acceptor with the β1, 6 linkage has been 
cloned. The branched form with this structure is 
called the I antigen, which is known to be an 
alloantigen that displays dramatic change during 
human development. The I antigen, one of the 
blood type antigens, is distinguished by its branch-
ed form of poly-N-acetyllactosamine from the i 
antigen with a linear form. The cDNA encoding 
for this enzyme (IGnT) has been cloned by the expres 
sion cloning method using anti-I antibody. Com 
parison of the sequences between the IGnT and the 
C2GnT revealed a highly conserved region within 
their catalytic domains. This is conceivable, since 
both the enzymes share not only the same donor 
substrate, UDP-GlcNAc, but also the same inter-
sugar β1, 6 linkage with only a subtle difference in 
the acceptor substrates, Gal or GalNAc. They have 
no sequence homology to the other GnTs 
mentioned above, which are entirely different 
enzymes in terms of the acceptor specificities and 
intersugar-linkage specificities. Southern blotting 
analysis using C2GnT and IGnT probes under low 
stringency conditions displayed multiple bands, 
suggesting a gene family of β1, 6 GnTs. It is 
anticipated that more genes included in the family 
will soon be cloned by cross hybridization probing 
with the C2GnT and IGnT genes.

(i) N-Acetylgalactosaminyltransferase

Beside the A transferase (A GalNAcT), a cDNA 
encoding β1, 4 GalNAcT involved in the synthesis 
of gangliosides has already been cloned (37) which 
catalyzes the transfer of GalNAc with the β1, 4 
linkage to GM2 and GD3, resulting in the synthesis 
of GM2 and GD3, respectively. It is of interest that the 
p40αx protein in the malignant cells infected 
with human T-cell lymphotropic virus type I 
(HTLV-I) can activate the β1, 4 GalNAcT gene, 
resulting in the expression of GM2 and GD3 on the 
cell surface as tumor markers.

References

glycosylation reactions in the endoplasmic reticulum. 
and transcript heterogeneity of the yeast gene ALG1, an 
essential mannansyltransferase involved in N-glycosyla-
CD28 cDNA by a high-efficiency COS cell expression 
diphosphate D-galactose D-glucose β-4-D-galacto-
syltransferase from bovine milk. J. Biol. Chem. 241: 
2672–2678.
5) Ballou, C.E. 1982. The molecular biology of the yeast 
Saccharomyces. p. 335–360. In Stralhem, J.N., Jones, 
E.W., and Broach, J.R. (eds), Metabolism and gene 
cloning of cDNA encoding UDP-GlcNAc; Galβ1-3 
GalNAc-R (GlcNAc to GalNAc) β1-6 N-acetyl-
glucosaminyltransferase by gene transfer into CHO cells 
7) Boegeman, E.E., Balaji, P.V., Sethi, N., Masibay, A.S., 
and Qasba, P.K. 1993. Expression of deletion mutants 
of bovine β1, 4-galactosyltransferase in Escherichia 
coli: importance of Cys134 for its activity. Protein 
8) Bunnell, B.A., Heath, L.S., Adams, D.E., Lahti, J.M., 
and Kidd, V.J. 1990. Increased expression of a 58-kDa 
protein kinase leads to changes in the CHO cell cycle. 


36) Munro, S. 1991. Sequence within and adjacent to the transmembrane segment of α2,6-sialyltransferase spec-


