Glycoconjugate Journal 21, 267–273, 2004

Cloning and tissue distribution of the human *B3GALT7* gene, a member of the β1,3-Glycosyltransferase family*

Chaoqun Huang[1,2], Jialiang Zhou[1], Shiliang Wu[1], Yuxi Shan[2], Shaolei Teng[1] and Long Yu[2]

[1] *Department of Biochemistry and Molecular Biology, Soochow University, Suzhou 215007, People's Republic of China,* [2] *The Institute of Genetics, Fudan University, Shanghai 20043, People's Republic of China*

We report here the cloning and tissue distribution of the human *B3GALT7* gene, a member of the β1,3-Glycosyltransferase family, structurally related to the β1,3-Galactosyltransferase family and β1,3-*N*-acetylglucosaminyltransferase family, isolated from a human lung cDNA library. *B3GALT7* is mapped to chromosome 19q13.2 by browsing the UCSC genomic database. It contains an ORF with length of 1191bp, encoding a protein with a signal peptide sequence and galactosyl-T domain, and its molecular weight and isoelectric point is predicted to be 43.3 kDa and 8.67 respectively. The molecular weight of the protein when expressed in *E. coli* corresponded to that expected. Northern blotting showed that *B3GALT7* was highly expressed in lung, throat and ileum, whereas the expression level was low in tongue, breast, uteri, testis. In addition, it was also demonstrated that *B3GALT7* is differentially transcribed in human tumor cell lines.
Published in 2004.

Keywords: **β1,3-Galactosyltransferase, β1,3-*N*-acetylglucosaminyltransferase, clone, northern blot, tumor cell line**

Introduction

The glycobiology field has evolved dramatically in the last decade with the isolation, cloning and expression of recombinant forms of an increasing number of the enzymes, glycosyltransferase, which catalyze the synthesis of complex oligosaccharides and glycoconjugates. Several β3Gal-T activities that form Galβ1-3Hex(NAc) α/β linkages exist in animals. These include type 1 chain synthase activity(Galβ1-3GlcNAcβ1-R), mucin-type core 1 synthase activity(Galβ1-3GalNAcα1-R), several glycosphingolipid synthase activity that form G_{M1}, Gal-Gb4, the histo-blood group A associated Gal-A glycolipids(Galβ1-3GalNAcα/β1-RA). It was therefore expected that a large β3Gal-T gene family would exist. Indeed, a large family of β1,3-Glycosyltransferases(CAZy family 31) (http://afmb.cnrs-mrs.fr/CAZY/index.html) has been identified and includes human β1,3-Galactosyltransferases (β3Gal-T1, T2, T3, T4, T5, T6) [1–5], β1,3-*N*-acetylglucosaminyltransferases [6], fringe [7,8], core 1 beta 1,3-galactosyltransferase [9,10], etc. β 1,3-Glycosyltransferases(CAZy family 31) gene family show diverse enzymatic functions. Enzyme activities using different donor substrates(UDP-Gal and UDP-GlcNAc) and different acceptor sugars(GlcNAc, Gal, and GalNAc) have been characterized, but all form the β1-3 linkage. Furthermore, acceptor substrate specificity studies with natural glycoconjugate reveal marked differences in biological functions. It was reported that it is not a single β3Gal-T enzyme that controls type 1 chain synthesis in epithelial tissues such as lung, colon, stomach, so it is likely that additional genes encoding type 1 chain synthases exist [4]. In this report, We describe the identification of a novel gene *β3GALT7*(UDP-Gal:betaGal beta 1,3-galactosyltransferase polypeptide 7), and show that it may be the member of β1,3-Glycosyltransferases(CAZy family 31) family. The protein sequence displays close homology with the sequence of human β1,3-Galactosyltransferase and β1,3-*N*-acetylglucosaminyltransferase families.

To whom correspondence should be addressed: Chaoqun Huang, Department of Biochemistry and Molecular Biology, Soochow University, Suzhou 215007, People's Republic of China. Tel.: +86-0512-67164423; E-mail: 011023471@fudan.edu.cn

*Sequence data from this article have been deposited with the GenBank Data Library under Accession No. AY277592.

Materials and methods

Cloning of *B3GALT7*

TBLASTN GenBank searches with the reported coding sequence of human *B3GALT1* sequence(GenBank accession

number AF117222) allowed us to indentify two ESTs (CA431056, BM726913) which were shown to have no homology to known galactosyltransferases. This two ESTs were used to screen human dbEST and the thereby identified human ESTs(accession No. BM478343, BM147124, BM726913, CA431056, BM771654, CA418485) were assembled *in silico* into a contig, which contains an open reading frame of 1191 nucleotides encoding a protein of 397 amino acids. To verify the sequence of the contig, primers of *B3GALT7*-A-up (5′-CACCAGAAACCAGCCCTAACACG-3′ corresponding to nucleotides 142–164 nt) and *B3GALT7*-B-down(5′-AAGACCTGGAAGGAGGGGCTGAG-3′ corresponding to nucleotides 1708–1730 nt) were designed and synthesized (Shenggong BioTech, Com.). PCR was performed with *B3GALT7* primers, and several human cDNA libraries (lung, colon, Clontech Co. ltd.) were used as templates. The PCR conditions were: denaturing at 94°C for 45 s, annealing at 58°C for 45 s, elongation at 72°C for100 s. The PCR product was subject to T-A cloning and sequencing.

Northern blot analysis of *B3GALT7*

Premade Multiple Tissue Northern (MTN) Blots isolated from sixteen human tissues were obtained from Shen Zhen King Grace Biotechnologies INC in China. The blot was probed with (α -^{32}p) dCTP labeled form of the complete coding region of human *B3GALT7* cDNA. Hybridization was performed according to the manufacturer's recommendations. The blot was also probed with radiolabeled β-actin as an indicator of RNA loading.

RNA isolation and reverse-transcription and polymerase chain reaction (RT-PCR)

Six human tumor cell lines(H1299: lung adenocarcinoma in which P53 is not expressed; A549: lung adenocarcinoma; SPCA-1: lung adenocarcinoma; H446: small cell lung cancer; H460: large cell lung carcinoma; SHG-44: neuroglioma) were purchased from ATCC(the American Type Cluture Collection), and were cultured as previously reported [11]. Total RNA was extracted from six different tumor cell lines with single-step isolation method using TRIzol reagent according to the manufacturer's instructions. cDNA was synthesized using 2 mg of total RNA, SuperscriptII reverse transcriptase (Gibco BRL) plus Oligo(dT)15 (Promega) according to the manufacturer's protocols. First-strand cDNA was subjected to RT–PCR amplification on FS-918 DNA Amplifier (Shanghai Fusheng Institute of Biotechnology) using the primer pair of *B3GALT7*-RT-A(5′-CTATGTGCCCGAGTCCTTCTTCG-3′ corresponding to nucleotides 1324–1346 nt) and *B3GALT7*-RT-B(5′-GCAGTTGTTTCCAGAGCCGAATG-3′ corresponding to nucleotides 1603–1625 nt). PCR product size was 302 bp. The expression of the β-MG was analyzed as a control: sense prime (5′-CTCGTGCTACTCTCTCTTTC-3′), antisense prime (5′-CATGTCTCGATCCCACTTAAC-3′). The PCR conditions were the same as in cloning except for that the elongation time was changed to 30 s.

Construction of the expression plasmid to produce recombinant *B3GALT7* protein

The *B3GALT7* coding sequence was generated by PCR amplification using Pfu polymerase (Stratagene Cloning Systems, La Jolla, CA). The primers used to generate the sequence were: pGEX-6p-1-*B3GALT7*-A(5′-TCGAATTCATGCGCTGCCCCAAGTG-3′) containing a EcoR I restriction site and pGEX-6p-1-*B3GALT7*-B (5′-TTCTCGAGGCACTGGAGCCTTGGGTC-3′) containing a Xhol I restriction site. The PCR product was restriction enzyme digested with EcoR I and Xhol I. The resulting product was ligated into the EcoR I/Xhol I restriction site of the pGEX-6p-1 vector. The sequence of the construct obtained was verified by automated DNA sequencing. The bacterial overexpression was performed with *E. coli* strain BL21(DE3). After transformation of the plasmid construct obtained in *E. coli* and plating on LB-agar (100 mg/ml ampicillin), a single colony was used to inoculate a 5 ml culture in LB-medium (200 mg/ml ampicillin). The cells were grown at 37°C. At an OD595 between 0.6 and 0.8, overexpression of recombinant human *B3GALT7* was induced by adding IPTG to a final concentration of 0.3 mM and incubated at 30°C for an additional 5 h. The cell were harvested by centrifugation at 10000 g for 2 min and The cell pellet was used for SDS-PAGE analysis.

In silicon analysis mapping and chromosomal location and gene structure

To determine the mapping information, the sequence of *B3GALT7* was used for searching the genomic database at http://genome.ucsc.edu. Related tools used for included Sequence Utility (http://hgsc.bcm.tmc.edu/searchlauncher), SMART for domain searching (http://smart.embl-heidelberg.de) and Vector NTI package (Informax, Co. ltd.). Cellular localization of protein *B3GALT7* was predicted by PSORT (http://psort.nibb.ac.jp).

Results

Cloning and sequence analysis of human *B3GALT7* gene

The PCR product encoding B3GALT7 was amplified from a human lung cDNA library, Cloned into pMD18-T vector (TaKaRa) and subjected to sequencing. The determined sequence was consistant with that defined for the *in silico* contig. In accordance with the guidelines of Human Genome Organization (HUGO) Nomenclature Committee (http://www.gene.ucl.ac.uk/nomenclature), this gene was named *B3GALT7*(UDP-Gal:betaGal beta 1,3-galactosyltransferase polypeptide 7) because it encodes a protein with a galactosyl-T domain and the protein sequence displays close homology with human β1,3-Galactosyltransferase. The gene symbol is provisional,

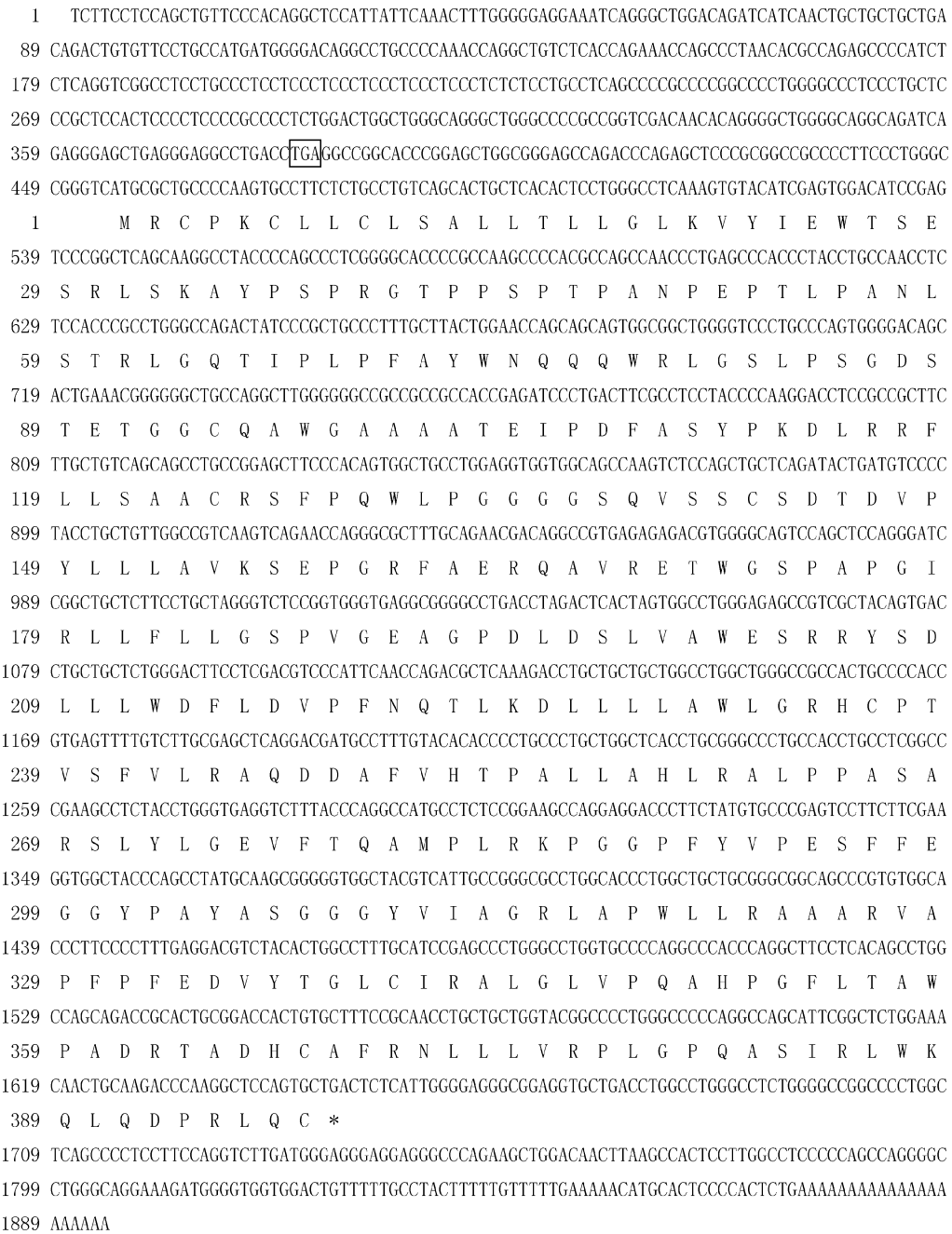

Figure 1. Nucleotide and amino acid sequences of *B3GALT7*. A stop codon upstream of the ORF at the same reading frame are boxed. Asterisk represents the stop codon.

dependent on the subsequent enzymological data. And the cDNA sequence submitted to GenBank with the GenBank Accession No AY277592. *B3GALT7* is 1894 bp in length, encoding a protein 397 amino acids (Figure 1). The molecular weight and isoelectric point is predicted to be 43.3 kDa and 8.67 respectively using OMEGIA software [12]. It is composed of three exons and two introns (Table 1), the open reading frame is from 455 nucleotides to 1645 nucleotides, The ATG start codon (nucleotides 455–457) is proceded by an in-frame stop codon TGA (nucleotides 383–385). Although there is not typical polyadenylation signal, the polyadenylation A tail starts at nucleotide 1873. *B3GALT7* is mapped to19q13.2 by browsing the UCSC genomic database (http://genome.ucsc.edu). Result of searching in SMART database showed that *B3GALT7* protein contains a signal peptide sequence and galactosyl-T domain (http://smart.embl-heidelbeg.de). B3GALT7 protein is predicted to be localized at extracellular by PSORT (http://psort.nibb.ac.jp).

Searches of homology in international protein databases of NCBI using the BLAST service

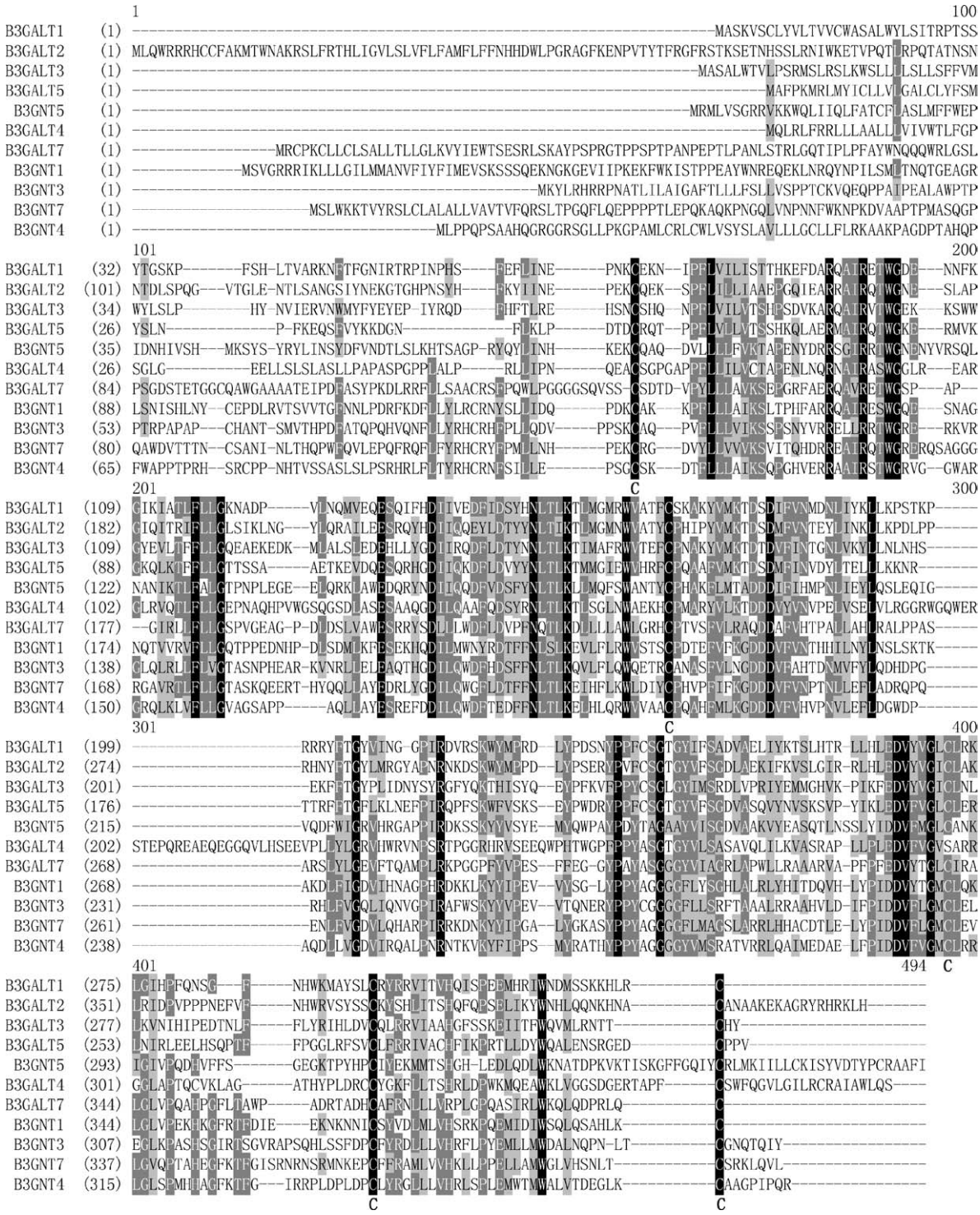

Figure 2. Alignment of the human B3GALT7 protein and human B3GALT1, B3GALT2, B3GALT3, B3GALT4, B3GALT5, B3GnT1, B3GnT3, B3GnT4, B3GnT5, B3GnT7 protein. Identical amino acids are shaded in black, similar amino acids are shaded in gray. Five conserved cysteines residues are showed.

(http://www.ncbi.nlm.nih.gov/blast) with the deduced protein sequence revealed that human B3GALT7 shares a high degree of homology with the hypothetical protein MGC32391 in mouse (60% identity and 65% similarity) and rat (60% identity and 65% similarity). By multiple sequence aligment, It also shows homology to the members in the B3GALT family and the B3GnT family (Figure 2). There are several conserved short sequence motifs and five

Table 1. Nucleotide sequence of exon-intron junctions of human *B3GALT7*

3′ Splice acceptor	*Exon*	*Size (bp)*	*5′ Splice donor*	*Intron*	*Size (bp)*
cDNAendTCTTCCTCCAGC	1	135	CCCCAAACCAGG***gt***ggagtgaagg	1	1025
caccctttccc***ag***CTGTCTCACCAG	2	287	CCAGACCCAGAG***gt***aaaaaagcat	2	473
gctcccttgt***ag***CTCCCGCGGCCG	3	1452	TCCCCACTCTGAaaaaaaaaaaaa	polyA	

Intron sequence is shown in lowercase and exon sequence in uppercase, respectively. Bold italics lettering indicators donor and acceptor splice sites.

cysteines residues highly conserved in these two families. The result of phylogenetic analysis of family 31 members including fringe, core 1 galactosyltransferase, β1,3-*N*-acetylglucosaminyltransferases, β1,3-galactosyltransferases and β1,3-*N*-acetylgalactosaminyltransferase indicate that it is associated together with some members of β1,3-*N*-acetylglucosaminyltransferases (Figure 3). These results suggested that B3GALT7 is the member of β1,3-Glycosyltransferase (CAZy family 31) family.

Tissue distribution of *B3GALT7* gene

The tissues in northern blot analysis include brain, heart, liver, spleen, stomach, testis, muscle skeleton, kidney, throat, ileum, tongue, uteri, ovary, breast, pancreas. The tissue distribution patterns of *B3GALT7* mRNA is shown in Figure 4. Northern blots revealed that *B3GALT7* was highly expressed in lung, throat and ileum, whereas the expression level was low in tongue, breast, uteri, testis. The size of the transcripts was about 1.9 kb. In other tissues, no significant signal was detected. The expression pattern of *B3GALT7* mRNA in six tumor cell lines was also determined by RT-PCR (Figure 5). The result showed that *B3GALT7* gene was differentially expressed in these tumor cell lines. The expression level was higher in A549 cells, but was low in H466 cells. There was no signal detected in SHG-44 cells.

Expression of recombinant B3GALT7 in prokaryote

Recombinant B3GALT7 protein was expressed in *E. coli* strain BL21(DE3). SDS-PAGE showed that the molecular weight of GST-tagged recombinant protein was about 69 KDa (Figure 6).

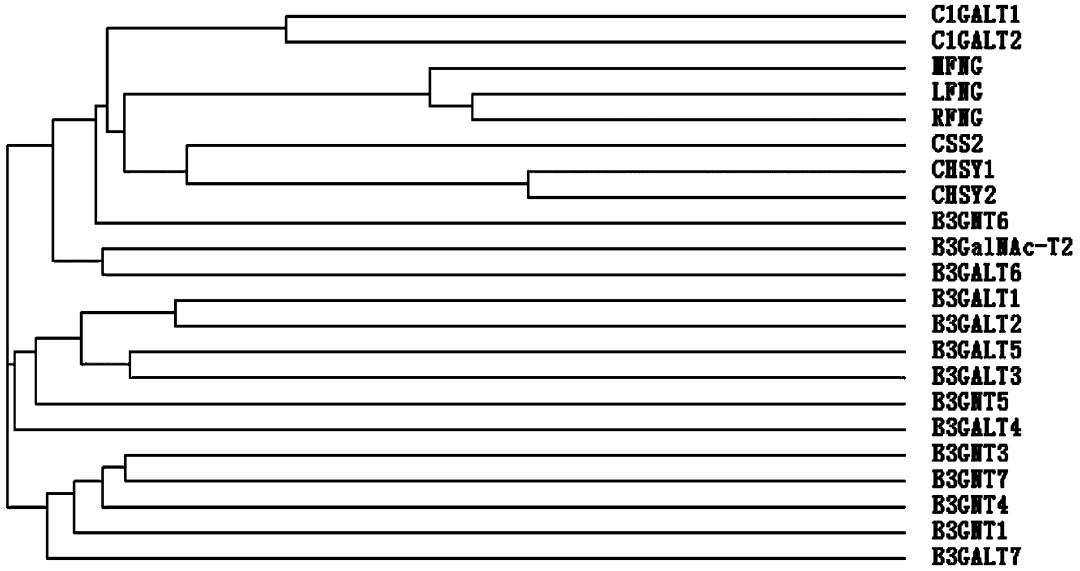

Figure 3. Phylogenetic tree of the CaZy family 31 members. The phylogenetic tree was plotted online (http://www.ebi.ac.uk/clustalw). The corresponding accession numbers are NP_064541, NP_689905, NP_002396, Q8NES3, Q9Y644, NP_078812, NP_055733, CAD43233, NP_006867, NP_689703, NP_542172, NP_066191, NP_003774, NP_149363, NP_003772, NP_114436, NP_003773, NP_055071, NP_660279, NP_110392, NP_150274, NP_940942, respectively (from top to bottom). C1GALT indicates core 1 UDP-galactose:*N*-acetylgalactosamine-alpha-R beta 1,3-galactosyltransferase, MFNG indicates Beta-1,3-*N*-acetylglucosaminyltransferase manic fringe, LFNG indicates Beta-1,3-*N*-acetylglucosaminyltransferase lunatic fringe, RFNG indicates Beta-1,3-*N*-acetylglucosaminyltransferase radical fringe, CSS2 indicates chondroitin sulfate synthase 2, CHSY indicates chondroitin *N*-acetylgalactosaminyltransferase, B3GalNAc-T2 indicates beta 1,3-*N*-acetylgalactosaminyltransferase 2, B3GALT indicates beta 1,3-galactosyltransferase, B3GNT indicates beta-1,3-*N*-acetylglucosaminyltransferase.

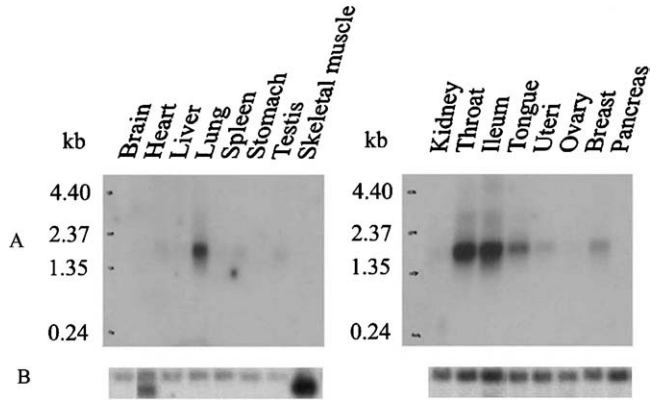

Figure 4. Northern blot analysis of *B3GALT7* mRNA in human sixteen tissues. (A) Northern blots containing mRNA from 16 human tissues were hybridized with a probe derived from *B3GALT7* cDNA. A 1.9-kb transcript was detected. (B) The same blots were hybridized with β-*actin* cDNA as a control.

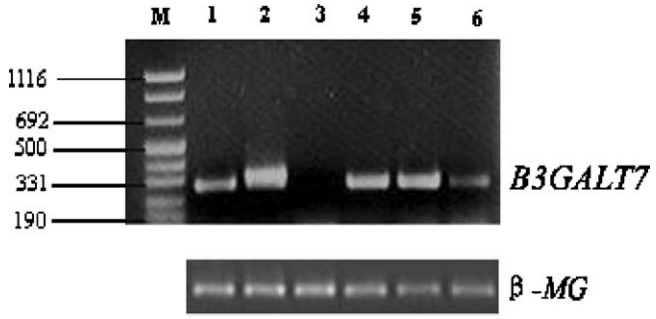

Figure 5. Expression of *B3GALT7* in six human tumor cell lines. M. DNA marker; Lane 1 lung adenocarcinoma H1299 in which P53 is not expressed; Lane 2 lung adenocarcinoma A549; Lane 3 neuroglioma SHG-44; Lane 4 large cell lung carcinoma H460; Lane 5 lung adenocarcinoma SPCA-1; Lane 6 small cell lung cancer H446.

Due to the GST-tag of around 26 KDa, the molecular weight of the recombinant B3GALT7 is around 43 KDa, which is similar to that predicted by OMIGA.

Discussion

We report the isolation and sequencing of the human *B3GALT7* cDNA, which encodes a protein structurally related to the β1,3-Galactosyltransferase family and β1,3-*N*-acetylglucosaminyltransferase family, which is mapped to chromosome 19q13.2. It contains an ORF with length of 1191 bp, encoding a protein with a signal peptide sequence and galactosyl-T domain. Analysis of sequence similarities between the six β3-Galactosyltransferase genes and five β1,3-*N*-acetylglucosaminyltransferase genes revealed that B3GALT7 share homology with β1,3-Galactosyltransferase and β1,3-*N*-acetylglucosaminyltransferase family members, and several conserved short sequence motifs were found, Interestingly, five cysteine residues are high conserved in these

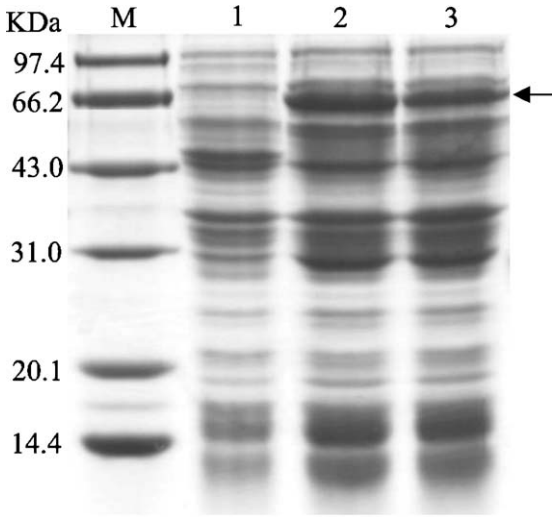

Figure 6. Expression of recombinant *B3GALT7* in *E.coli* strain BL21(DE3). Coomassie Blue-stained SDS-PAGE gel of the *B3GALT7* expression construct. M. molecular mass markers; Lane 1 not induced by IPTG; Lane2,3 induced by adding IPTG to a final concentration of 0.3 mM.

proteins. The result of the phylogenetic analysis indicated that B3GALT is associated together with some members of β1,3-*N*-acetylglucosaminyltransferases. This gene is more related to GlcNAc transferases than to Gal transferases of CaZy family 31, B3GALT7 may be a GlcNAc transferase on the basis of the phylogenetic analysis. Sequence analysis of*B3GALT7* showed that the coding region of this gene is located within a single exon, it is consistant with *B3GALT1*, *B3GALT2*, *B3GALT3*, *B3GALT4*, *B3GALT5* , the coding regions of which were also located within a single exon [13]. All members of the β1,3-Galactosyltransferase family have different chromosomal localizations; *B3GAL-T1* at 2q31.1, *T*2 at 1q31, *T*3 at 3q25, *T*4 at 6p21.3, *T*5 at 21q22.3, *T*7 at 19q13.2. The exsistence of multiple β1,3-Galactosyltransferases and β1,3-*N*-acetylglucosaminyltransferases is suggestive of a high degree of redundancy for genes of seemingly similar functions, which could represent a comprehensive genetic back-up. However, it is equally likely that this high number of enzymes have evolved as a result of specific requirements for enzymes with different functions. Northern analysis of *B3GALT7* revealed that it was highly expressed in lung, throat and ileum, and the expression level was low in tongue, breast, uteri, testis. Whereas Northern analysis indicates that *B3GALT1* is mainly expressed in brain [2,13]; *B3GALT2* was mainly expressed in brain, although heart also expressed [2,13,14]; *B3GALT3* was more widely expressed, and was found in brain, pancreas, kidney, and reproductive organs [2,14]; *B3GALT4* appears to be ubiquitously expressed, although higher levels were found in thymus and spleen [15]; *B3GALT5* was expressed in epithelial tissues including pancreas and intestine [4,16]; *B3GALT6* was widely expressed [5]. B3GnT2 was ubiquitously expressed, whereas B3GnT3

was expressed in colon, jejunum, stomach, esophages, placenta and trachea. B3GnT4 was mainly expressed in brain [6]. These data indicated that β1,3-Galactosyltransferase and β1,3-*N*-acetylglucosaminyltransferase genes were differentially expressed in human tissues, suggesting that different β1,3-Galactosyltransferase and β1,3-*N*-acetylglucosaminyltransferase play roles in different tissues as well as perform different functions even though they all form β1-3 linkage.

Acknowledgments

We thank for Dr. Songming Jiang (Institute of Genetics, Fudan University, China) for reviewing this paper.

References

1 Sasaki K, Sasaki E, Kawashima K, Hanai N, Nishi T, Hasegawa M, (July 5, 1994) Japanese Patent 0618759 A 940705.
2 Kolbinger F, Streiff MB, Katopodis AG, Cloning of a human UDP-galactose:2-Acetamido-2-deoxy-D-glucose 3-galactosyltransferase catalyzing the formation of Type 1 chains, *J Biol Chem* **273**, 433–40 (1998)
3 Amado K, Almeida R, Carneiro F, Leverly SB, Holmes EH, Nomoto M, Hollingsworth MA, Hassan H, Schwientek T, Nielsen PA, Bennett EP, Clausen H, A family of human β3-Galactosyltransferases, *J Biol Chem* **273**, 12770–8 (1998).
4 Isshiki S, Togayachi A, Kudo T, Nishihara S, Watanabe M, Kubota T, Kitajima M, Shiraishi N, Sasaki K, Andoh T, Narimatsu H, Cloning, Expression, and characterization of a novel UDP-galactose: β-*N*-Acetylglucosamine β1,3-Galactosyltransferase (β3Gal-T5). Responsible for Synthesis of Type 1 chain colorectal pancreatic epithelia and tumor cells derived thereform, *J Biol Chem* **274**, 12499–507 (1999).
5 Bai XM, Zhou DP, Brown JR, Crawford BE, Hennet T, Esko JD, Biosythesis of the linkage region of glycosaminoglycans, *J Biol Chem* **276**, 48189–95 (2001).
6 Norihiko S, Ayumi N, Akira T, Tetsuo E, Tomohiro A, Yoji Y, Nobuyuki I, Satoshi N, Satoshi K, Susumu S, Hisashi N, Katsutoshi S, Identification and characterization of three novel β1, 3-*N*-acetylglucosaminyltransferases structurally related to the β1,3-Galactosyltransferase family, *J Biol Chem* **276**, 3498–507(2001).
7 Moran JL, Johnston SH, Rauskolb C, Bhalerao J, Bowcock AM, Vogt TF, Genomic structure, mapping, and expression analysis of the mammalian Lunatic, Manic, and Radical fringe genes, *Mamm Genome* **10**, 535–41 (1999).
8 Johnston SH, Rauskolb C, Wilson R, Prabhakaran B, Irvine KD, Vogt TF, A family of mammalian Fringe genes implicated in boundary determination and the Notch pathway, *Development* **124**, 2245–54 (1997).
9 Ju T, Brewer K, D'Souza A, Cummings RD, Canfield WM, Cloning and expression of human core 1 beta1,3-galactosyltransferase, *J Biol Chem* **277**, 178–86 (2002).
10 Ju T, Cummings RD, A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase, *Proc Natl Acad Sci USA* **99**, 16613–8 (2002).
11 Anna B, Maurizia V, Marco T, Differential expression of β1,3-Galactosyltransferases in human colon cells derived from adenocarcinomas or normal mucosa, *FEBS Letters* **451**, 75–80 (1999).
12 Kramer JA, Omiga: A PC-based sequence analysis tool, *Molecular Biotechnology* **19**, 97–106 (2001).
13 Amado M, Almeida R, Carneiro F, Levery SB, Holmes EH, Nomoto M, Hollingsworth MA, Hassan H, Schwientek T, Nielsen PA, Bennett EP, Clausen H, A family of human β3-Galactosyltransferase: Characterisation of four members of a UDP-galactose: β-*N*-acetylglucosamine/β-*N*-acetylgalactosamineβ1,3-Galactosyltransferase family, *J Biol Chem* **273**, 12770–8 (1998).
14 Hennett T, Dinter A, Kuhnert P, Mattu TS, Rudd PM, Berger EG, Genomic cloning and expression of three murine UDP-galactose: β-*N*-acetylglucosamine β1,3-Galactosyltransferase gene, *J Biol Chem* **273**, 58–65 (1998).
15 Miyaki H, Fukumoto S, Okada M, Hasegawa T, Furukawa K, Expression cloning of rat cDNA encoding UDP-galactose G(D2) β1,3-Galactosyltransferase that determines the expression of G(D1b)/G(M1)/G(A1), *J Biol Chem* **272**, 24794–9 (1997).
16 Zhou D, Berger EG, Hennett T, Molecular cloning of a human UDP-galactose:GlcNAcβ1,3GalNAc β1,3 Galactosyltransferase gene encoding an O-linked core 3-elongation enzyme, *Eur J Biochem* **263**, 571–6 (1999).

Received 18 May 2004; revised 19 June 2004; accepted 22 June 2004