Glycoconjugate Journal 16, 337–350 (1999)

Combinatorial PCR approach to homology-based cloning: Cloning and expression of mouse and human GM3-synthase

Dmitri Kapitonov, Erhard Bieberich, and Robert K. Yu*

Department of Biochemistry and Molecular Biophysics, Medical College of Virginia, Virginia Commonwealth University Richmond, VA 23298-0614

GM3-synthase, also known as sialyltransferase I (ST-I), catalyzes the transfer of a sialic acid residue from CMP-sialic acid onto lactosylceramide to form ganglioside GM3. In order to clone this enzyme, as well as other sialyltransferases, we developed an approach that we termed combinatorial PCR. In this approach, degenerate primers were designed on the basis of conserved sequence motifs of the ST3 family of sialyltransferases (STs). The nucleotide sequence of the primers was varied to cover all amino acid variations occurring in each motif. In addition, in some primers the sequence was varied to cover possible homologous substitutions that are absent in the available motifs. A panel of cDNA from 12 mouse and 8 human tissues was used to enable cloning of tissue- and stage-specific sialyltransferases. Using this approach, the fragments of 11 new putative sialyltransferases were isolated and sequenced so far. Analysis of the expression pattern of a particular sialyltransferase across the panel of cDNA from the different tissues provided information about the tissue specificity of ST expression. We chose two new ubiquitously expressed human and mouse STs to clone full-length copies and to assay for GM3-synthase activity. One of the STs, which exhibited the highest homology to ST3 Gal III, showed activity toward lactosylceramide (LacCer) and was termed ST3 Gal V according to the suggested nomenclature [1]. The other ubiquitously expressed sialyltransferase was termed ST3Gal VI. All isolated sialyltransferases were screened for alternatively spliced forms (ASF). Such forms were found for both human ST3Gal V and ST3Gal VI in human fetal brain cDNA library. The detailed cloning strategy, functional assay, and full length cDNA and protein sequences of GM3 synthase (ST3Gal V, or ST-I) are presented.

Keywords: **PCR, GM3-synthase, sialyltransferase, molecular cloning of glycosyltransferases, cDNA**

Abbreviations: **ST, sialyltransferase; ST-I, CMP-NeuAc:lactosylceramide α2-3 sialyltransferase; ST3, a family of sialyltransferases that transfers a sialic acid residue from CMP-sialic acid to the third carbon of a sugar acceptor molecule, forming an α2-3 bond; LacCer, lactosylceramide or Galβ1-4Glcβ1-1′Cer; Gg3, GalNAcβ1-4Galβ1-4Glcβ1-1′Cer; GM3, NeuAcα2-3Galβ1-4Glcβ1-1′Cer; GM1, Galβ1-3GalNAcβ1-4Gal(3-2αNeuAc)β1-4Glcβ1-1′Cer; GD1a, NeuAcα2-3Galβ1-3GalNAcβ1-4Gal(3-2α NeuAc)β1-4Glcβ1-1′Cer; ASF, alternatively spliced form.**

Introduction

GM3 synthase (sialyltransferase I, or ST-I) is the enzyme involved in the last step of GM3 biosynthesis. It catalyzes the transfer of a sialic acid moiety from CMP-sialic acid onto lactosylceramide, forming an α2–3 linkage. GM3 is a common precursor for nearly all of the naturally occurring gangliosides. GM3 and the products of GM3 metabolism have important functions in normal organism development as well as in pathogenesis. They have been implicated in modulation of cellular growth [2,3], proliferation, differentiation [4–8], and apoptosis [9]. The modulatory role of GM3 and its metabolites, such as lyso-GM3 and deNAcGM3, in signal transduction [10] has been shown to be mediated through growth factor- or hormone receptor-associated cytoplasmic protein kinases [11–15], protein tyrosine phosphatase [16], protein kinase C (PKC) [14,17–19], phospholipase C (PLC) δ1 [20], and Ca^{2+} -ATPase [21]. In addition, a close association of transducer proteins (c-Src, Ras, FAK, Rho A, H-Ras) and neurotrophic factors such as prosaposin [22] with GM3 in the cell surface microdomains has been reported, suggesting a functional association [23]. GM3 has also been reported to

*To whom correspondence should be addressed. Robert K. Yu, Department of Biochemistry and Molecular Biophysics, Medical College of Virginia Campus, Virginia Commonwealth University, Box 980614, Richmond, VA 23298-0614. Tel. 804-828-9762; Fax 804-828-1473; E-mail: rkyu@vcu.edu

interact with other glycosphingolipids, such as LacCer and Gg3, to provide adhesion of melanoma cells onto endothelial cells [24,25] and function as a fusion co-factor for HIV-1 and HIV-2 [26,27]. For these reasons, interest is mounting in the cloning of cDNAs that code for glycosyltansferases and glycosidases involved in ganglioside metabolism. With the availability of these cDNAs, it becomes possible to investigate the mechanisms underlying the expression of gangliosides in normal and abnormal cellular development.

In general, three approaches are used for the cloning of the cDNA coding for a specific sialyltransferase: (a) protein purification followed by protein sequencing and cDNA library screening [28,29], (b) expression cloning [30,31], and (c) direct cDNA cloning with degenerate primers designed based on the most conserved area of homology between family members [32–34]. In this report, we improved the direct cDNA cloning approach by utilizing (1) degenerate primers designed to accommodate the variations in the entire conserved pattern (similar amino acid sequences present in different family members) as well as possible homologous substitutions and then utilizing different combinations of these primers in PCR; (2) a panel of cDNA libraries as a template, which assures complete representation of differentially expressed and tissue-specific sialyltransferases; (3) a pooled cDNA library as a template in the 96-well thermoplate, which assures amplification of cDNA for low-abundance sialyltransferases with less than perfect homology with the primers; (4) polyacrylamide gel with an alternative cross-linker to assure complete separation of PCR products that differ only a few bases in length; (5) a step-down PCR protocol and PCR with nested primers to increase the specificity of the reaction; (6) the primers for a new sialylmotif, which we termed STP (sialylmotif petit) and which is probably implicated in the recognition of the nucleotide portion of CMP- sialic acid (see below). With this improved approach for the molecular cloning of sialyltransferases, we have cloned fragments of 11 new putative ST3 sialyltransferases. In addition, 8 new full-length cDNAs for sialyltransferases have been isolated and sequenced [35] and one has been enzymatically characterized. Here we illustrate the isolation and characterization of mouse and human GM3-synthases (ST-I, ST3Gal V) and sialyltransferases ST3Gal VI by this approach. A preliminary account of this work has appeared [35].

Materials and Methods

Materials

The NG108-15 cell line, a hybrid of murine neuroblastoma and rat glioma cells, was kindly provided by Dr. Robert Ledeen (New Jersey School of Medicine, Newark, NJ). P19 embryonal carcinoma cell line was purchased from ATCC (ATCC #CRL 1825, ATCC, Manassas, VA). The human fetal glioma cell line, N-370 FG, was kindly provided by Dr. Toshio Ariga (Medical College of Virginia, Richmond, VA). Dulbecco's modified Eagle's medium (DMEM) was obtained from GibcoBRL (Grand Island, NY). Tissue culture dishes were from Falcon/Becton Dickinson Co. (Franklin Lakes, NJ, USA) and high-performance thin-layer chromatographic (HPTLC) plates were from Merck (Darmstadt, Germany). Mouse multiple tissue cDNA (MTC™) panel (#K1423-1) and human fetal MTC™ panel (#K1425-1) were purchased from Clontech (Clontech Laboratories, Inc., Palo Alto, CA). SuperScript™ human fetal brain (#10662-013) and mouse 15.5 day embryo (#10667-012) plasmid cDNA library were purchased from Life Technologies (Gaithersburg, MD). All other chemicals were of analytical grade or higher, and solvents were freshly redistilled before use.

Primers

In ST cloning, in addition to the well-known STL and STS motifs (Fig. 1), we took advantage of a newly characterized STP motif [35], located between STS and C-terminus, 36–51 amino acid residues from the C-terminus of the protein and 7–14 amino acid residues away from STS motif. The STP motif has the following sequence (H,C,R)(Y,H)(Y,W,F)(D,E,H,Y) (Fig. 1). The conserved motifs for ST3 sialyltransferases along with amino acid variations used in the primer design are shown in Fig. 1A and Table 1. Based on homology analysis, we subdivided the ST families [1] into 7 subfamilies [35] to facilitate primer design. In this classification, the ST3 family was divided into 2 subfamilies: 4 and 5. Forward primers 1F (a–d) were designed based on a portion of the STL motif, with variations to accommodate all of its amino acid substitutions, and possible homologous substitutions such as K↔R, known for other STs, V↔I, and V↔A [35]. In the amino acid sequences for the design of forward primers 2F(a, b), G was substituted for S, and the second G residue substituted for H according to the variations in the motif. Combinations of these substitutions were used in the design of the primers 2F(c, d). In the sequence for the 3F primers, S was substituted for N according to the pattern for family 4 of ST3. A broader range of substitutions, L↔M and V↔T, was used in other experiments. In the amino acid sequence for the primers 4F, F↔T and S↔G substitutions were used according to the motif, and their combinations were used in the design of the primers 4Fd and 4Fe. A↔Y substitutions were used in the design of 1R primers according to the motif and their combination was used in other experiments. N↔H and Y↔F substitutions were used according to the motif for ST3 family in the design of 2Ra and 2Rb primers. Additional Y↔L and A↔(Y/F) substitutions were used in other experiments. For the primers designed based on the STP motif, Y↔W↔F and D↔E substitutions were used in different combinations (Fig. 1, Table 1).

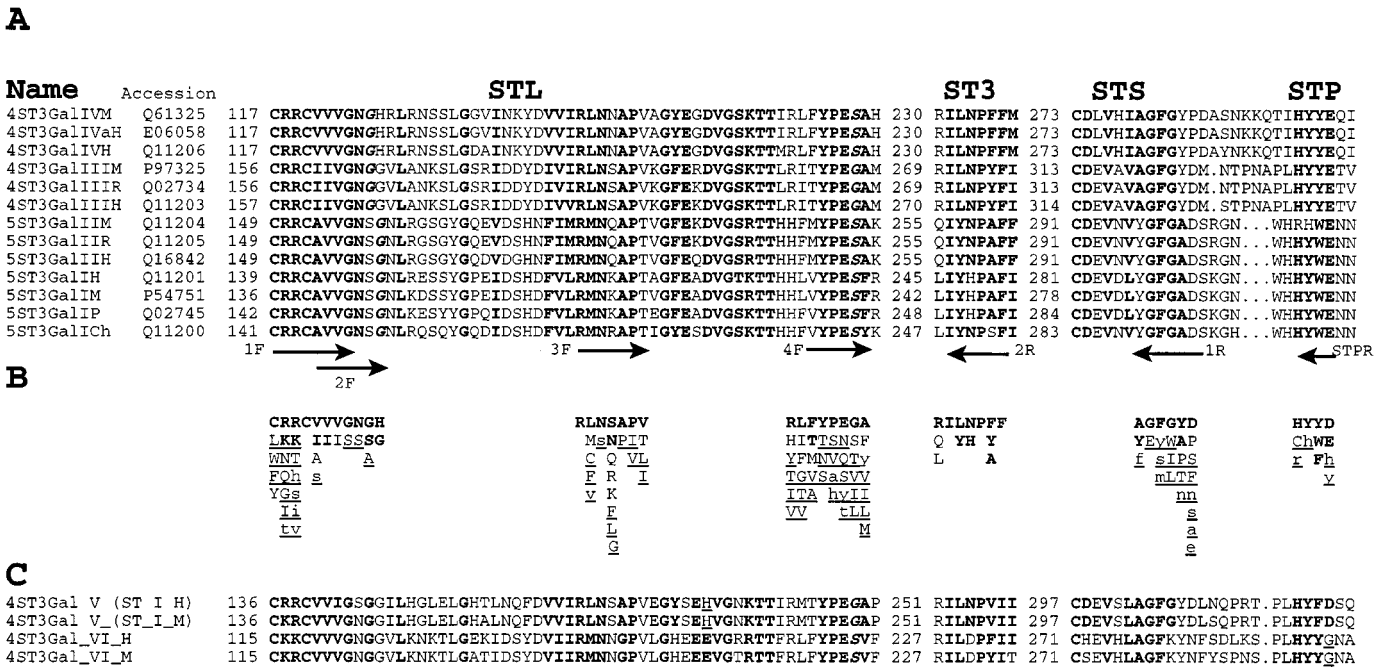

Figure 1. A: Position of the PCR primers used in this work along the amino acid sequence of the conserved ST3 domains. ST3Gal I to ST3Gal IV, all known ST3 sialyltransferases. M - Mouse, H - Human, a - very close isoforms, R - Rat, P - Pig, and Ch - Chicken sequences. The number in front of the ST indicates belonging to a subfamily of sialyltransferases. F - forward primer, R - reverse primer, STL - large sialylmotif, STS - small sialylmotif (only the last 13 amino acid residues are shown), STP - petit sialylmotif, number indicates the position of amino acid residue. **B:** Variations in the motifs considered in the primer design. Amino acid residues shown in bold were used in combinatorial PCR in this study. Underlined amino acid residues indicate variations in the motif outside the ST3 family. Amino acid residues shown in small letters indicate variations known for only one ST. The lowest variability was found in amino acid sequence for 1F, 2F, 1R and STP motif. C: motifs in the new, ubiquitously expressed members of ST3 family isolated in this study. The conserved amino acid residues are shown in bold, and the unique amino acid residues not found in any other STs are underlined.

To assure amplification of cDNA for all ST family members, different combinations of forward and reverse primers were used in the combinatorial PCR approach along with a panel of mouse and human cDNAs as templates. Using a panel of cDNAs as a template provided additional information about tissue- and stage-specific expression of the particular ST.

Library amplification

A set of 12 96-well flat-bottom tissue culture plates were inoculated at a density of 2,000 clones/well with mouse or human SuperScript cDNA plasmid libraries in 0.3 ml of Terrific Broth (TB) containing ampicillin (100 µg/ml) and carbenicillin (50 µg/ml). The library was amplified for 24 hr at 30 °C, and 20 µl aliquots from each well of a particular row were combined into one well of a master plate so that row A of the first plate became well 1A of a master plate, row B became 1B, and so forth. Aliquots of 0.75 ml of TB media with the antibiotics in 1.2-ml cluster tubes (Costar Corp., Cambridge, MA) were inoculated with 20 µl of the combined pools of the master plate, and cells were collected by centrifugation at 1,600 g for 5 min. Plasmid DNA was isolated following the standard alkaline lysis procedure with 70 µl of solubilization, denaturing, and neutralization buffer, and then precipitated with 150 µl of isopropanol in 96-well U-shape tissue culture plates (Coster Corp., Cambridge, MA). Following centrifugation at 1,600 g, the resulting DNA pellet was redissolved in 150 µl of water.

cDNA synthesis

Isolation of mRNA from human N-370 FG, HeLa, or differentiated mouse P19 cells was achieved by using PolyA-Tract® system III (Promega, Madison, WI). The sample was reverse-transcribed using 10 pmol of STP-R (a,b,c,d,e) primers (separately or as a mixture) and SuperScript™ reverse transcriptase.

PCR

For PCR with cDNA panels as a template (12 reactions for the mouse or 8 for the human panel, multiplied by the number of different primer combinations in the primary PCR reaction), 1 nmol of each primer, 100 µl of 10X Taq polymerase buffer [final concentration: 20 mM Tris sulfate, 3.5 mM $MgSO_4$, 16 mM $(NH_4)_2SO_4$, and 150 µg/ml of BSA, which was omitted if silver staining was used], 20 µl of 10 mM

Table 1. List of primers

Name	*Amino acid seq.*	*Primer*	*Degeneracy*
1Fa	CRRCVVVG	TG(C/T)CG(C/G)CG(C/G)TG(T/C)GT(G/C/T)GT(G/C/T)GT(G/C/T)GG	432
1Fb	CRRCIIVG	TG(C/T)CG(C/G)CG(C/G)TG(T/C)AT(C/T)AT(C/T)GT(G/C/T)GG	192
1Fc	CKRCVVVG	TG(C/T)AA(A/G)CG(C/G)TG(T/C)GT(G/C/T)GT(G/C/T)GT(G/C/T)GG	432
1Fd	CKKCVVVG	TG(C/T)AA(A/G)AA(A/G)TG(C/T)GT(G/C/T)GT(G/C/T)GT(G/C/T)GG	432
1Fe	CRRCAVVG	TG(C/T)CG(C/G)CG(C/G)TG(T/C)GCNGT(G/C/T)GT(G/C/T)GG	576
2Fa	VVVGNGG	GT(G/C/T)GT(G/C/T)GT(G/C/T)GGNAA(T/C)GGNGG	864
2Fb	VVVGNGH	GT(G/C/T)GT(G/C/T)GT(G/C/T)GGNAA(T/C)GGNCA	864
2Fc	VVVGNSG	GT(G/C/T)GT(G/C/T)GT(G/C/T)GGNAA(T/C)(T/A)(C/G)NGG	3456
2Fd	VVVGNSH	GT(G/C/T)GT(G/C/T)GT(G/C/T)GGNAA(T/C)(T/A)(C/G)NCA	3456
3Fa	RLNSAPV	AG(G/A)(T/C)TGAA(T/C)(T/A)(C/G)NGCNCCNGT	2048
3Fb	RLNNAPV	AG(G/A)(T/C)TGAA(T/C)A(T/C)GCNCCNGT	256
4Fa	TYPEGA	ACNTA(C/T)CCNGA(G/A)GGNGC	256
4Fc	FYPESA	TT(T/C)TA(C/T)CCNGA(G/A)(A/T)(G/C)NGC	512
4Fd	TYPESA	ACNTA(C/T)CCNGA(G/A)(A/T)(G/C)NGC	1024
4Fe	FYPEGA	TT(T/C)TA(C/T)CCNGA(G/A)GGNGC	128
4Fb	RLFYPES	(C/A)GNCTNTT(C/T)TA(C/T)CCNGA(G/A)(T/A)C	2048
1Ra	rev. AGFGYD	(A/G)TC(A/G)TANCC(A/G)AANCCNGC	512
1Rb	rev. YGFGAD	(G/A)TCNGCNCC(G/A)AANCC(A/G)TA	512
2Ra	rev. RILNP(F/Y)	(A/G)(A/T)ANGG(A/G)TTNA(A/G)NATNC	4096
2Rb	rev IY(H/N)PAF	(A/G)AANGCNGG(A/G)T(T/G)(A/G)TA(T/G/A)AT	768
STP-Ra	rev. HYYDx	*NNNNNN*(A/G)TC(A/G)TA(A/G)TA(A/G)TG	
STP-Rb	rev. HYYEx	*NNNNNN*(T/C)TC(A/G)TA(A/G)TA(A/G)TG	
STP-Rc	rev. HYWEx	*NNNNNN*(T/C)TCCCA(A/G)TA(A/G)TG	
STP-Rd	rev. HYWDx	*NNNNNN*(A/G)TCCCA(A/G)TA(A/G)TG	
STP-Re	rev. HYFDx	*NNNNNN*(A/G)TC(A/G)AA(A/G)TA(A/G)TG	
ST3-III	N/A	F: ATCATCGTGGGCAATGGAG	1
Human		R: CCTGGATGAAATATGGGTTGAG	Tm = 58
ST3-V	N/A	F 619: TATTGGAAGCGGAGGAATACTG	1
ASF		R 1253: CTTTCACCACTCCCTCTTTGAC	Tm = 58

F - forward, R - reverse, rev. - amino acid sequence was used to design reverse primers, seq. - sequence, Degeneracy - degree of degeneracy (number of possible combinations). *NNNNNN* - six random nucleotides were included in the STP primer sequence when it was used in primary PCR reaction, but were omitted when it was used in the cDNA synthesis. Tm - annealing temperature used in PCR. M - Mouse, H - Human.

dNTP mix, 10 U of Taq polymerase, and 0.5 µl of the template were combined and brought with water to the final volume of 1 ml. Aliquotes of 10 µl of this mix were transferred into each of the 96 wells of a Thermowell plate (07-200-248, Fisher Scientific, Pittsburgh, PA), combined with 0.5 µl of a template DNA, and overlaid with mineral oil. Following initial denaturation at 95° for 1 min, the reaction was allowed to proceed for 43 cycles at 95 °C for 20 sec, 62 °C for 30 sec, and 68 °C for 45 sec. During the first seven cycles, the annealing temperature was decreased 1 °C per cycle. When the annealing temperature reached 54 °C, the reaction was allowed to proceed for an additional 35 cycles. The final extension was performed at 72 °C for 10 min. An aliquot of 0.5 µl of the primary PCR reaction mixture was reamplified with the internal set of primers following the same PCR profile. The volume per reaction in the secondary PCR was increased to 20 µl, and all the components of the reaction mixture were increased correspondingly. For the PCR with a pooled cDNA library, the same reaction mixture was used with 1 µl of the template and the same PCR profile with a starting annealing temperature of 65 °C and a final annealing temperature of 59 °C for the last 35 cycles. The higher annealing temperature was necessary to decrease nonspecific priming on the vector itself. In primary PCR with STP primers the starting and final annealing temperatures were decreased by 4 °C. The PCR products were separated by agarose or, if reamplification was necessary, by polyacrylamide gel electrophoresis. In order to verify the specificity of PCR, Southern hybridization analysis was performed with the random primer-labeled fragments spanning the amplified areas of several known sialyltransferases.

Sequencing

The sequencing reaction was performed using a BigDye™ dye-terminators sequencing kit (Applied Biosystems, Inc.)

following the manufacturer's instructions with the following modifications: volumes of the reagents used were decreased so that 0.5 μg of plasmid DNA in 1 μl of water were combined with 1.5 μl of terminators mix and 5 pmol of the sequencing primer in 1 μl of water. Products of the sequencing reaction were separated on an ABI PRISM™377 analyzer.

Library screening

Gene-specific primers were synthesized using the sequencing information, and a second round of screening was then performed using the DNA sample from the primary plates as a template. Since each well of the master plate corresponds to a single row on a particular primary plate, an aliquot of bacterial suspension from each well of the positive rows (the row corresponding to the well from the master plate that produced a positive signal in the first round of PCR) was analyzed in the second round of PCR. Bacterial cells from the positive well of the primary plate were plated on a Nylon filter at the density of 10,000 colonies/150 mm plate. A replica filter was then prepared. Following denaturation at 70 °C in 0.5 M NaOH for 10 min with subsequent neutralization in 1 M Tris-HCl, pH 7.0, the replica filter was hybridized with 1×10^6 cpm/ml of hybridization buffer (50 mM Tris-HCl, pH 7.0 at 60 °C, 1 M NaCl, 1% SDS, 10% dextran sulfate MW 400,000, and 100 μg/ml denatured shredded salmon sperm DNA) for 4 h, washed four times in 1% SDS, 1X SSC buffer at 60 °C and exposed to an X-ray film for 10 to 30 min. Plasmid DNA from the positive clones was further sequenced. The primer design for the primer walking was assisted by the Prime program (Genetic Computer Group, Madison, WI) and EST database analysis.

Cell culture and transfection

NG108-15 cells were propagated in DMEM, supplemented with 10% fetal calf serum in a humidified atmosphere at 5% CO_2. Cells were transfected with pFLAG-CMV-5 ST-I vectors following a lipofectamine transfection protocol (Life Technologies, Gaithersburg, MD) and harvested following 48 hours of incubation. pCMV-Sport 2 plasmid containing ST-I cDNA with N-terminal deletion was used as a control.

Enzyme assay

Transfected cells corresponding to a cell mass equivalent to 10 mg of protein were solubilized with 100 μl of enzyme assay buffer (25 mM sodium cacodylate, pH 6.5, 10 mM $MgCl_2$, 0.3% Triton CF-54) for 30 min at 4 °C. Any insoluble material was removed by centrifugation at 14,000 g for 5 min, and aliquots of 20 μl corresponding to approximately 150 μg of solubilized protein were taken from the supernatant for determination of enzyme activity. The aliquots were added to 80 μl of enzyme assay buffer supplemented with 40 nmoles of CMP-N-[^{14}C]acetylneuraminic acid (0.2 μCi) and 20 nmoles of an acceptor glycolipid (GM1, GD1a, GM3, or LacCer). After incubation for 2 h at 37 °C, the incubation mixture was supplemented with 400 μl of water and 500 μl of $CHCl_3/CH_3OH/H_2O$ (8:4:3 by volume) for Folch-extraction of the assay product [^{14}C]-GM3 and then subjected to gel chromatography on Sephadex G-50 as described elsewhere [36]. The incorporated radioactivity was determined by liquid scintillation counting of the Sephadex G-50 eluate or by HPTLC of the Folch extract in $CHCl_3/CH_3OH$/0.2% aq. $CaCl_2$ (50:45:10 by volume), followed by autoradiography of the developed chromatogram. Blank assays were performed with the heat-inactivated enzyme.

Results and Discussion

In this paper we describe an approach that we developed for homology-based PCR cloning. In this approach, protein sequences of all homologous members of an enzyme family (ST in this paper) were aligned, and amino acid motifs were generated based on the conserved areas of homology. Forward and reverse degenerate primers were designed so that most variations of amino acid residues and potential conserved substitutions were included in the motif. All possible combinations of forward and reverse primers were then used in the PCR with a panel of cDNAs from different tissues and different organisms (mouse and human in this paper), thus providing amplification of tissue-specific and developmentally regulated enzymes efficiently.

For cloning of GM3-synthase, an aliquot of each cDNA panel was amplified by a primary PCR with the following combinations of primers (Table 1): 1Fa-1Ra, 1Fb-1Ra, 1Fc-1Ra, 1Fd-1Ra, 1Fe-1Ra; 1F(a,b,c,d,e)-1Rb; followed by the secondary PCR with 2Fa-2Ra, 2Fb-2Ra, 2Fa-2Rb, 2Fb-2Rb; 3F(a,b)-2R(a,b); 4F(a,b)-2R(a,b). The primary PCR with 2F(a,b)-1R(a,b) combinations of primers was followed by the secondary PCR with 3F(a,b)-2R(a,b) and 4F(a,b)-2R(a,b) combinations of primers. The primary PCR with 3F(a,b)-1R(a,b) combinations of primers was followed by the secondary PCR with 4F(a,b)-2R(a,b) combinations of primers. The position of amino acid residues used for the design of primers is shown in Figure 1. Results of the PCR with 1Fa-1Ra combinations of primers followed by 2Fa-2Ra are shown in Fig. 2A, (lanes 1a–12a), followed by 3Fa-2Ra (Fig. 2A, lanes 1b–12b), and by 4Fa - 2Ra (Fig. 3A, lanes 1a–12a). Results of the PCR with 2Fa - 1Ra combination of primers followed by 3Fa - 2Ra are shown in Fig. 2A (lanes 1c–12c) and followed by 4Fa - 2Ra (Fig. 3A, lanes 1b–12b). Results of PCR with 3Fa - 1Ra combinations of primers followed by 4Fa - 2Ra are shown in Fig. 3A (lanes 1c–12c). Southern hybridization analysis with the fragment of 4ST3 Gal IV cDNA was performed to verify the specificity of PCR, i.e., fragments of known sialyltransferases could be

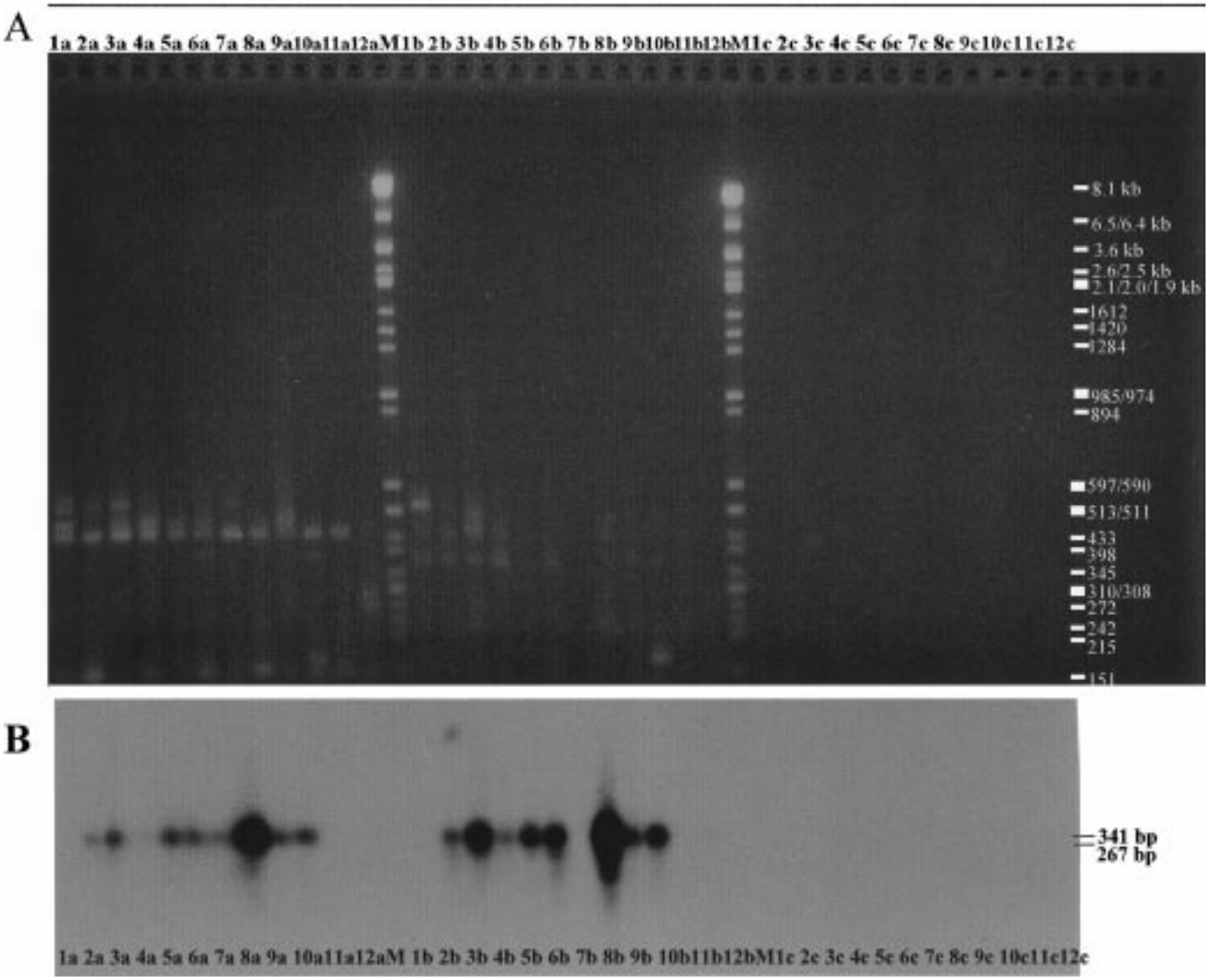

Figure 2. A: Agarose gel electrophoresis of the PCR products with mouse cDNA panel as a template and the following combinations of primers: 1Fa-1Ra followed by 2Fa-2Ra is shown in (lanes 1a–12a); followed by 3Fa-2Ra (lanes 1b–12b). Results of the PCR with the 2Fa-1Ra combination of primers followed by 3Fa-2Ra (lanes 1c–12c). Lane 1: heart; lane 2: brain; lane 3: spleen; lane 4: lung; lane 5: liver; lane 6: skeletal muscle; lane 7: kidney; lane 8: testis; lane 9: 7-day embryo; lane 10: 11-day embryo; lane 11: 15-day embryo; lane 12: 17-day embryo. Lane M: λ Ava II DNA size standard with corresponding size standards indicated on the right margin. **B:** Southern blot hybridization analysis of the same gel with a fragment of 4ST3 Gal IV cDNA. The corresponding size of the PCR products is indicated on the right margin.

amplified in combinatorial PCR as well (Figs. 2B and 3B). Gel-purified PCR products were cloned and sequenced. Since we employed mouse and human panels, it was possible to isolate both mouse and human copies of a particular ST. Mouse and human sialyltransferases were considered functionally identical if the percentage of homology in the coding region was higher then 80% on a cDNA level. If neither mouse nor human homolog was found, we designed degenerate primers based on a specific amino acid sequence in the STL or STS rather than on an entire motif, and we re-screened the pooled cDNA library by PCR. In addition, standard library screening was performed with either a mouse or a human counterpart as a probe. Based on the sequence of PCR products, gene-specific primers were designed, and the pooled cDNA library was screened again. Positive pools were plated and screened with PCR product as a probe to isolate full-length cDNA.

In order to compensate for possible underamplification of homologous STs due to the lower degree of homology with the designed primers, pooled human fetal brain and mouse total embryo cDNA plasmid libraries with a pool size of 2,000 independent colonies were screened. This pool

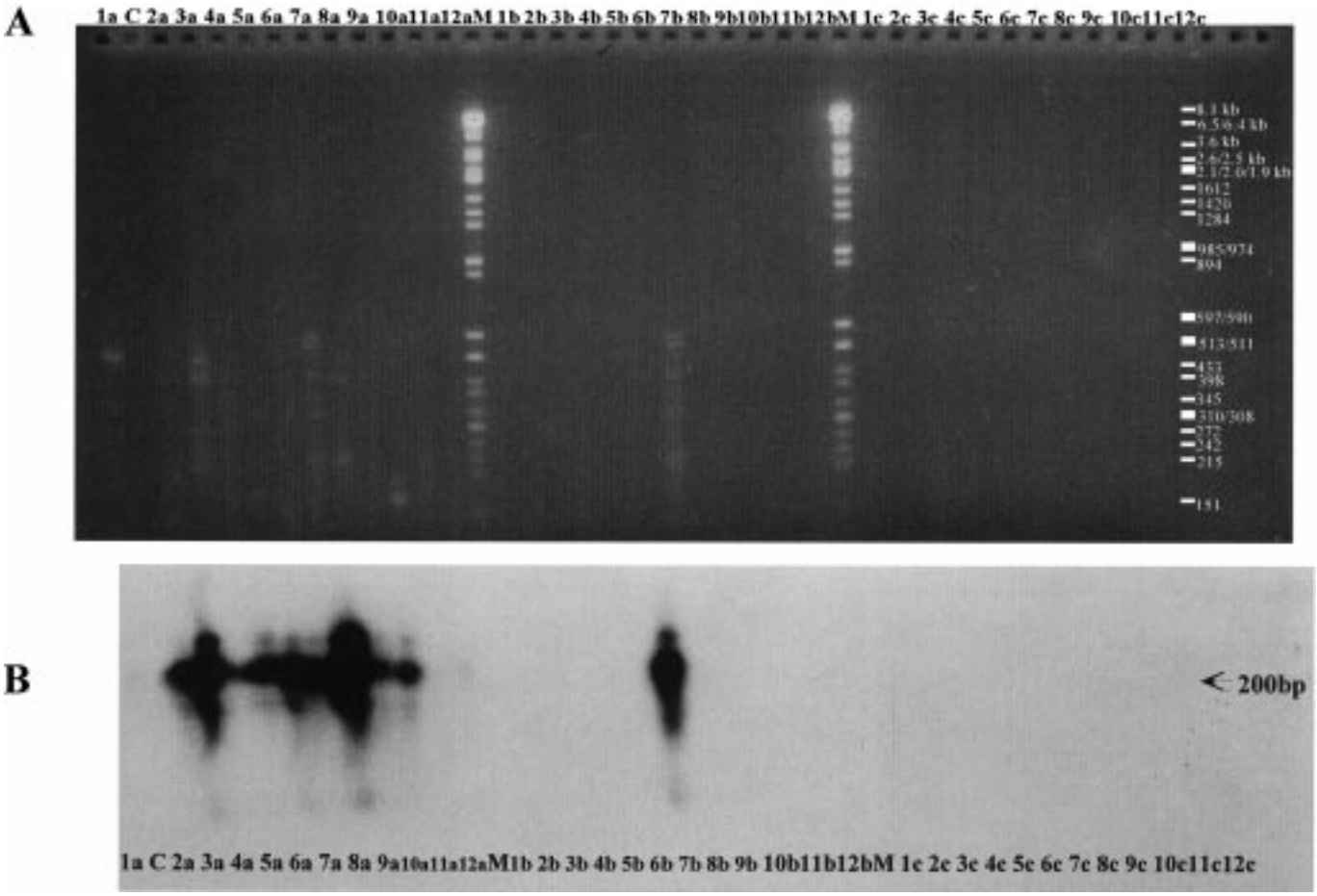

Figure 3. A: Agarose gel electrophoresis of the PCR products with mouse cDNA panel as a template and the following combinations of primers: 1Fa-1Ra followed by 4Fa-2Ra (lanes 1a–12a), 2Fa-1Ra combination of primers followed by 4Fa-2Ra (lanes 1b–12b) and 3Fa-1Ra combination of primers followed by 4Fa - 2Ra (lanes 1c–12c). Lane 1: heart; lane 2: brain; lane 3: spleen; lane 4: lung; lane 5: liver; lane 6: skeletal muscle; lane 7: kidney; lane 8: testis; lane 9: 7-day embryo; lane 10: 11-day embryo; lane 11: 15-day embryo; lane 12: 17-day embryo. Lane M: λ Ava II DNA size standard with the corresponding size standards indicated on the right margin. **B:** Southern blot hybridization analysis of the same gel with a fragment of 4ST3 Gal IV cDNA. The corresponding size of the PCR products is shown on the right margin.

size provided representation of no more than one sialyltransferase in most of the pools, thus assuring adequate and specific amplification when the template exhibited sub-optimal homology to the primers (compare lanes 2F and 2G in Fig. 4).

The expression pattern across the panel of tissues provided information about the tissue-specificity of the particular ST expression. Since GM3 is synthesized in all tissues, we cloned two different STs that are ubiquitously expressed in mouse and human tissues. Owing to their closer homology to ST3 Gal III sialyltransferase, one of these STs was termed ST3Gal V and the other one termed ST3Gal VI because it showed a higher homology to ST3Gal IV.

The human copy of ST3Gal V has three alternatively spliced forms (ASFs) that could be detected using primers F 619 and R 1253 (Table 1). According to the size of PCR products amplified using these primers, the ASFs were termed ASF 800, with a PCR product of 792 bp (total length, 3494 bp, GenBank # AF119417); ASF 600, with a PCR product of 634 bp (full length, 2288 bp, GenBank # AF119415); and ASF 400, with a PCR product of 454 bp (total length, 2415 bp, GenBank # AF119418). The mouse GM3-synthase was deposited with GenBank accession # AF119416. Two ASFs were found for the human ST3Gal VI (GenBank # AF119391). Mouse ST3Gal VI was deposited in the GenBank with accession number AF119390. To date, no ASFs were found for the mouse counterparts. Human ASFs of ST3Gal V and VI were used to assay for the GM3-synthase and GD1a-synthase activity using LacCer and GM1 as substrates, respectively (Fig. 5). Of all three alternatively spliced forms of ST3Gal V and two forms of ST3Gal VI, only ASF 600 of ST3Gal V revealed GM3-synthase activity. The complete nucleotide and the corresponding amino acid sequence of this ASF, together with the sites of alternative splicing, are shown in Fig. 6. Among

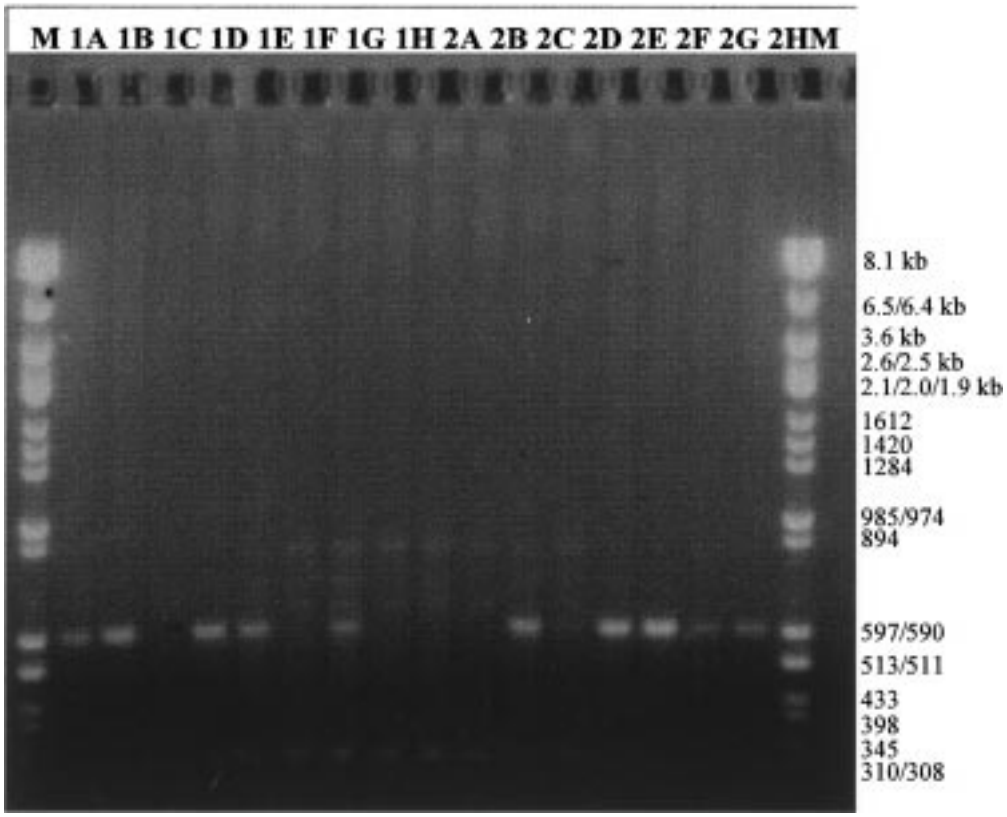

Figure 4. Agarose gel electrophoresis of the products of PCR with pooled human fetal brain cDNA library as a template. Each lane corresponds to the individual well of a master 96-well plate, where individual rows of primary plates of 2,000 cfu/well were combined, thus bringing the complexity of a library to 24,000 cfu/well. Lanes 1A–2H - PCR products of the wells from columns 1 and 2 of a master plate as a template. M - λ Ava II DNA size standard with the corresponding size standards indicated on the right margin.

all ASFs of ST3Gal V, ASF 600 had the highest homology with mouse ST3Gal V and with other STs. ASF 600 of ST3Gal V contained three conserved in-frame potential initiation codons. The amino acid sequence that starts from the third initiation codon had the highest homology with other STs and was sufficient for expression of functional activity. Therefore, numbering of amino acid residues along the protein sequence starts with M of the third reading frame (Fig. 6). Human GM3-synthase contains 362 amino acid residues and the mouse analogue has 359 amino acid residues. The degree of similarity between the two enzymes is 90.5% and the degree of identity is 86.4% at the amino acid level and 84.7% at the cDNA level in the coding region, starting with ORF 3. This degree of similarity corresponds to the average similarity of 85% expected for identical mouse and human genes. The similarity in the 5′ untranslated region drops to 81% and in the 3′ untranslated region drops to 54%. The hydrophobicity plot of human and mouse GM3-synthase is shown in Fig. 7. It reveals an N-terminal transmembrane domain common to all mammalian STs and a predicted type II transmembrane topology [37–39]. The human 4ST3Gal V has a calculated molecular mass of 41.74 KD and a pI of 8.74 and the mouse enzyme has a calculated molecular mass of 41.24 KD and a pI of 7.50. Human and mouse ST3Gal V have conserved STL, STS, and STP motifs common to other STs, as well as an ST3 motif common to all ST3 STs. Each ST3Gal V contains three potential N-glycosylation sites that are conserved for the mouse and human enzymes, five conserved potential casein kinase II phosphorylation sites, two conserved potential cAMP/cGMP-dependent protein kinase phosphorylation sites, as well as several non-conserved phosphorylation sites [40–42]. Mouse ST3Gal V has a $(GT)_{27}$ repetitive element in the 3′ untranslated region (Fig. 6).

Since GM3-synthase is the first entry sialyltransferase that gates the biosynthesis of all complex gangliosides, its cloning opens the possibility of investigating the regulatory mechanisms for ganglioside biosynthesis at the molecular biological level. After we reported cloning and sequencing of ST3Gal V and ST3Gal VI (GM3-synthase) from mouse and human [35], one group reported the sequence of human GM3-synthase, which is identical to ours at the amino acid level [43] and another group presented the sequence of mouse GM3-synthase [44], which shows several differences from the one reported in this paper. All the differ-

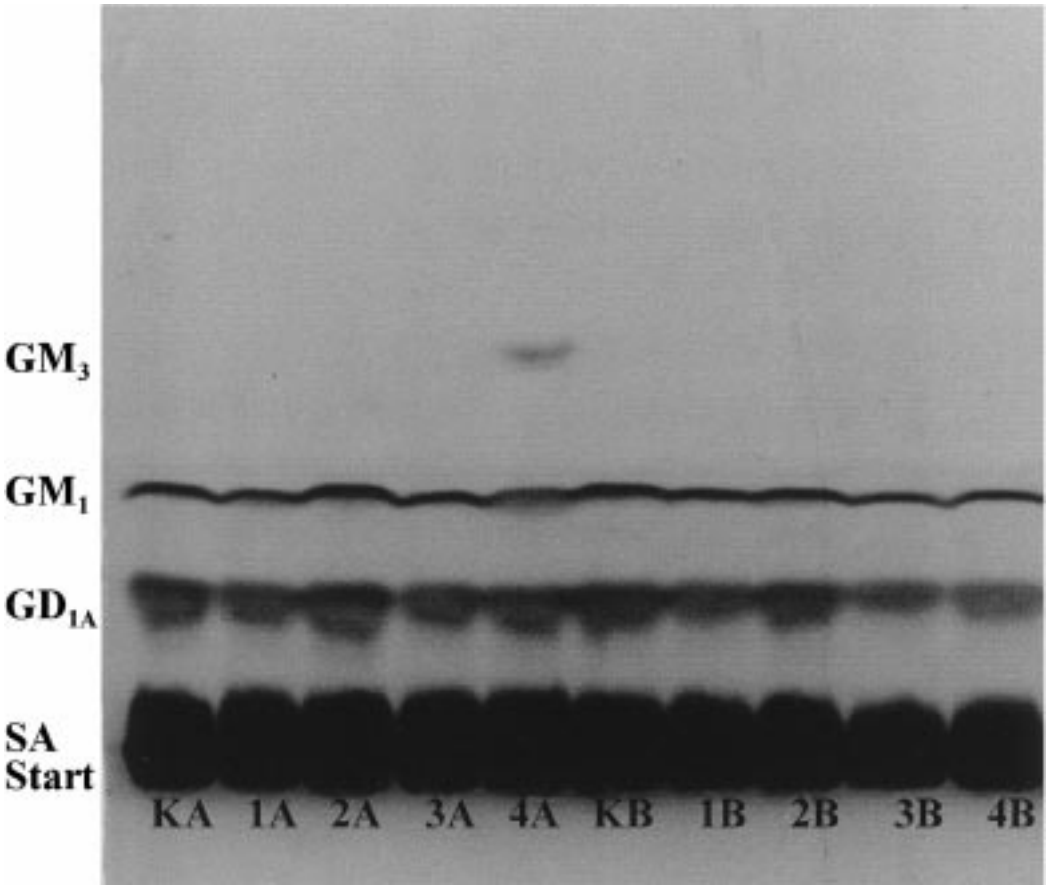

Figure 5. Autoradiogram of a TLC plate with separated products of enzyme assay. The corresponding markers are indicated on the left margin. Lanes KA and 1A–4A: products of the enzyme assay with LacCer as a substrate. Lanes KB and 1B–4B: products of the enzyme assay with GM1 as a substrate. Lane 1: human 4ST3Gal VI; lane 2: human 4ST3Gal V alternatively spliced form (ASF) 800; lane 3: human 4ST3Gal V ASF 400; lane 4: human 4ST3Gal V ASF 600. K: control for the endogenous enzyme activity of untransfected cells. The heavy GM1 and GD1a bands in all lanes arose from endogenous enzyme activities.

ences are shown in Figure 6. The reason for the discrepancy is not known.

The functional role of the sialylmotifs is still obscure. Most interestingly, we identified an STP motif which has the general structure (H,C,R)(Y,H)(Y,W.F)(D,E,H,Y) [35] (Fig. 1). This motif consists of four amino acid residues and is present in all known vertebrate sialyltransferases characterized so far. We speculate that it may play a role in interaction with the cytosine moiety which is common to the donor nucleotide, CMP-sialic acid. Computer modeling revealed interaction of aromatic ring of Y/F/W residues with cytosine, electrostatic interaction between amino acids H or C and oxygen from cytosine or phosphate group, and electrostatic interaction between D/E or Q residue and amino group of the cytosine moiety. The correctness of this model must await confirmation by site specific mutagenesis and/or by X-ray crystallization of the enzymes or NMR nuclear Overhauser experiments.

As a result of our combinatorial PCR approach, we have isolated 11 fragments of new sialyltransferases and cloned 8 full-length cDNAs. Work is in progress for the determination of substrate specificity of other ASFs of ST3Gal V, ST3Gal VI and the other sialyltransferases that we have cloned. Recently we have identified common domain structures of a number of glycosyltransferase families [45]. Using the conserved pattern of these domains and combinatorial PCR approach we were able to clone fragments of two new inverting glycosyltransferases and two new retaining glycosyltransferases, which proves that this approach can be used for cloning of the members of other homologous families (unpublished).

Acknowledgment

This work was supported by USPHS Grant NS11853.

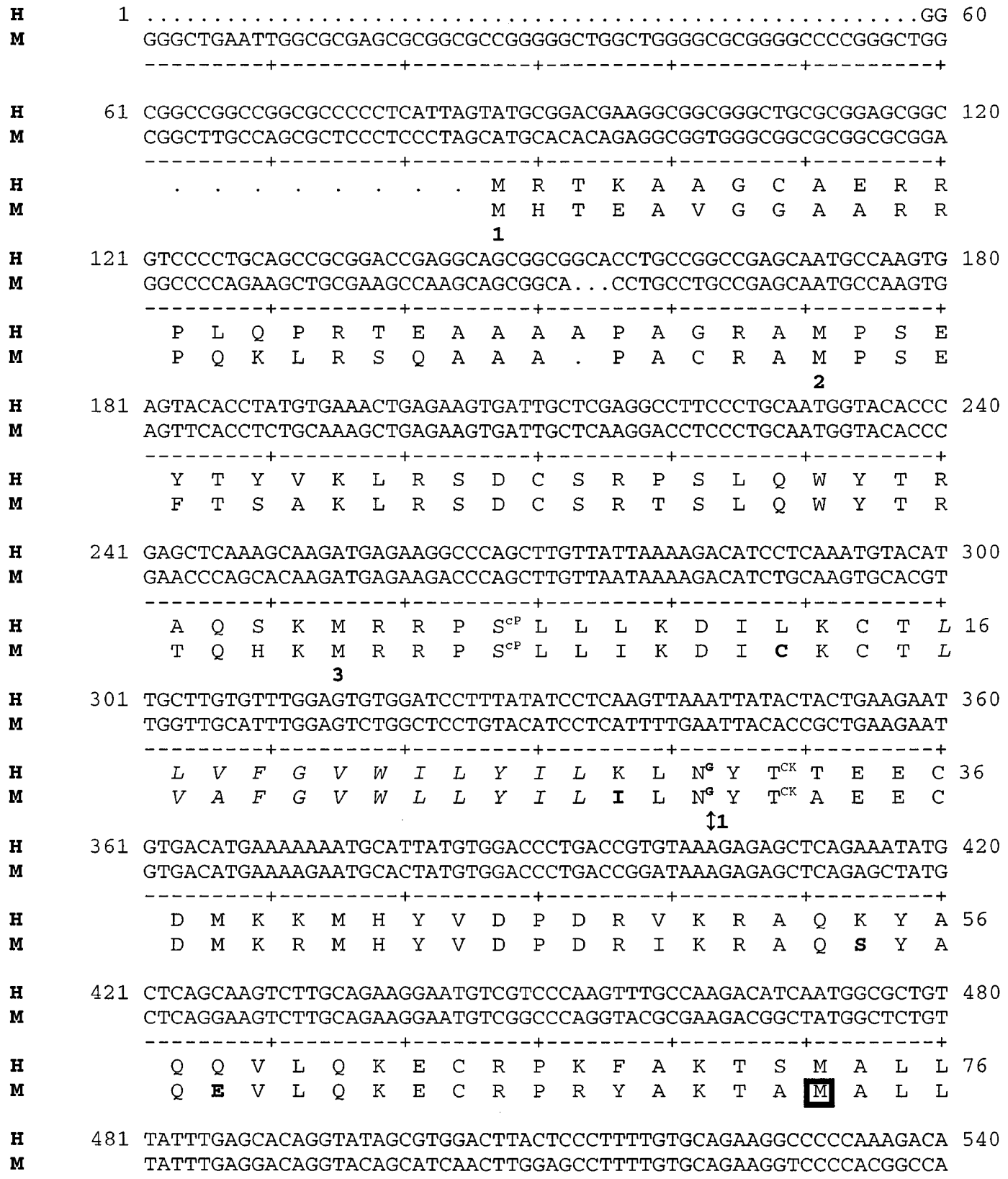
H 1 ..... GG 60
M GGGCTGAATTGGCGCGAGCGCGGCGCCGGGGGCTGGCTGGGGCGCGGGGCCCCGGGCTGG
H . . . . . M R T K A A G C A E R R
M M H T E A V G G A A R R
1
H 121 GTCCCCTGCAGCCGCGGACCGAGGCAGCGGCGGCACCTGCCGGCCGAGCAATGCCAAGTG 180
M GGCCCCAGAAGCTGCGAAGCCAAGCAGCGGCA...CCTGCCTGCCGAGCAATGCCAAGTG
H P L Q P R T E A A A A P A G R A M P S E
M P Q K L R S Q A A A . P A C R A M P S E
2
H 181 AGTACACCTATGTGAAACTGAGAAGTGATTGCTCGAGGCCTTCCCTGCAATGGTACACCC 240
M AGTTCACCTCTGCAAAGCTGAGAAGTGATTGCTCAAGGACCTCCCTGCAATGGTACACCC
H Y T Y V K L R S D C S R P S L Q W Y T R
M F T S A K L R S D C S R T S L Q W Y T R
H 241 GAGCTCAAAGCAAGATGAGAAGGCCCAGCTTGTTATTAAAAGACATCCTCAAATGTACAT 300
M GAACCCAGCACAAGATGAGAAGACCCAGCTTGTTAATAAAAGACATCTGCAAGTGCACGT
H A Q S K M R R P S^cP L L L K D I L K C T L 16
M T Q H K M R R P S^cP L L I K D I C K C T L
3
H 301 TGCTTGTGTTTGGAGTGTGGATCCTTTATATCCTCAAGTTAAATTATACTACTGAAGAAT 360
M TGGTTGCATTTGGAGTCTGGCTCCTGTACATCCTCATTTTGAATTACACCGCTGAAGAAT
H L V F G V W I L Y I L K L N^G Y T^CK T E E C 36
M V A F G V W L L Y I L I L N^G Y T^CK A E E C
↕1
H 361 GTGACATGAAAAAAATGCATTATGTGGACCCTGACCGTGTAAAGAGAGCTCAGAAATATG 420
M GTGACATGAAAAGAATGCACTATGTGGACCCTGACCGGATAAAGAGAGCTCAGAGCTATG
H D M K K M H Y V D P D R V K R A Q K Y A 56
M D M K R M H Y V D P D R I K R A Q S Y A
H 421 CTCAGCAAGTCTTGCAGAAGGAATGTCGTCCCAAGTTTGCCAAGACATCAATGGCGCTGT 480
M CTCAGGAAGTCTTGCAGAAGGAATGTCGGCCCAGGTACGCGAAGACGGCTATGGCTCTGT
H Q Q V L Q K E C R P K F A K T S M A L L 76
M Q E V L Q K E C R P R Y A K T A M A L L
H 481 TATTTGAGCACAGGTATAGCGTGGACTTACTCCCTTTTGTGCAGAAGGCCCCCAAAGACA 540
M TATTTGAGGACAGGTACAGCATCAACTTGGAGCCTTTTGTGCAGAAGGTCCCCACGGCCA

Figure 6.

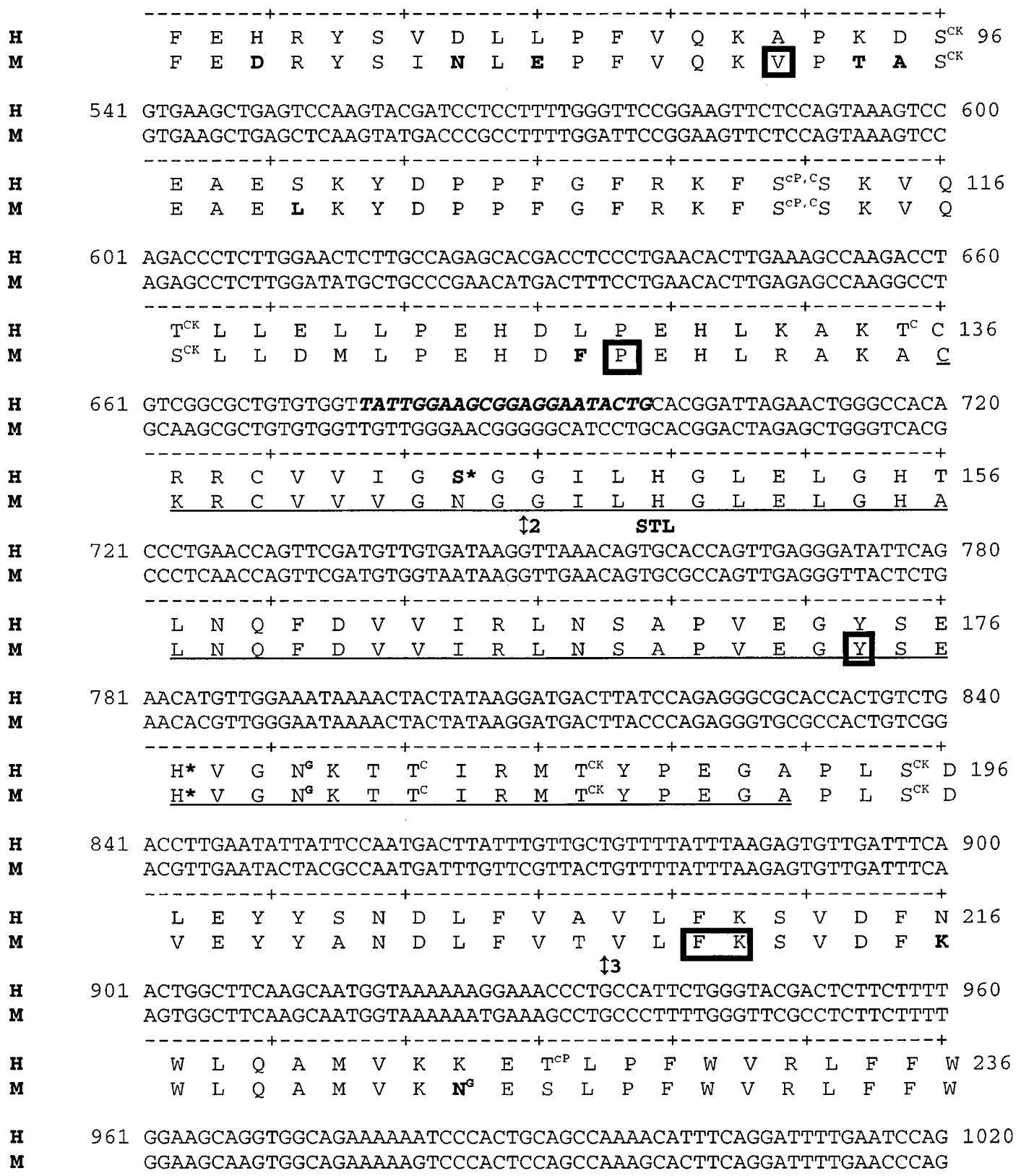

Figure 6. (*continued*)

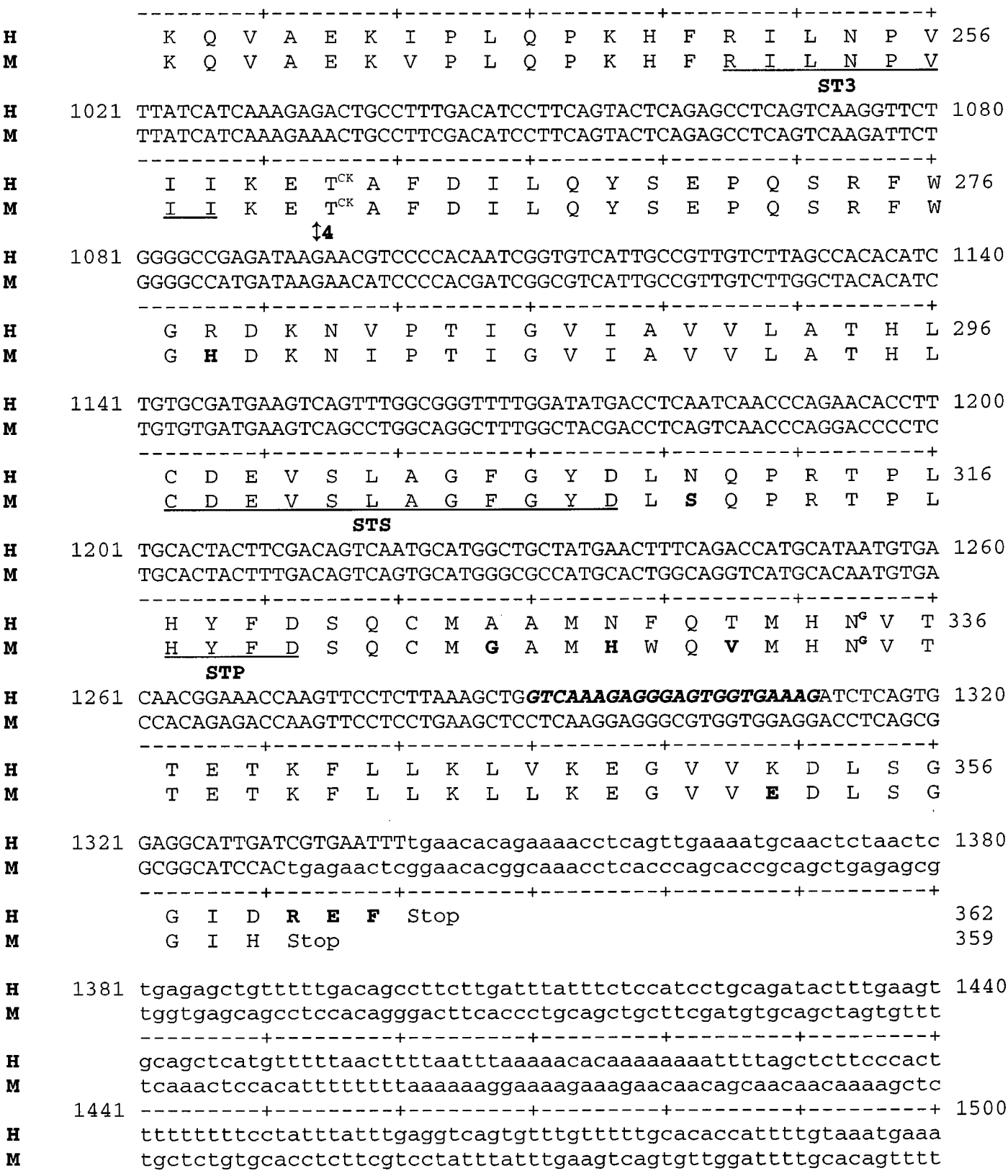

Figure 6. (*continued*)

```
      1501 ---------+---------+---------+---------+---------+---------+ 1560
H          cttaagaattgaattggaaagacttctcaaagagaattgtatgtaacgatgttgtattga
M          gtaagttaatcttaagaatgggattggaaggacttttcaaagagaattgtatagtttatt
      1561 ---------+---------+---------+---------+---------+---------+ 1620
H          tttttaagaaagtaatttaatttgtaaaacttctgctcgtttacactgcacattgaatac
M          gtttttaaggaagtaatttaatttgcagaaactgtacacacgtactctgctcaggtgttg
      1621 ---------+---------+---------+---------+---------+---------+ 1680
H          aggtaactaattggaaggagaggggaggtcactcttttgatggtggccctgaacctcatt
M          agtggaggagagggcttctggccctggatgatggctgtgatgcccgatactggggtctgc
      1681 ---------+---------+---------+---------+---------+---------+ 1740
H          ctggttccctgctgcgctgcttggtgtgacccacggaggatccactcccaggatgacgtg
M          tgctctgtttggtagaactgatggcagagaaacttcctgcctccaggataaagggcttac
      1741 ---------+---------+---------+---------+---------+---------+ 1800
H          ctccgtagctctgctgctgatactgggtctgcgatgcagcggcgtaggctggctggttga
M          tcatcacctctggcagctgctagacaagttcataaccccttttctgctagtccatctgcca
      1801 ---------+---------+---------+---------+---------+---------+ 1860
H          gaaggtcacaacccttctctgttggtctgccttctgctgaaagactcgagaaccaaccag
M          gctggctcgcaggactcaggcagggcagctgtcccggaggctgctggttggtgagccact
      1861 ---------+---------+---------+---------+---------+---------+ 1920
H          ggaagctgtcctgaaggtccctggtcggagagggacatagaatctgtgacctctgacaac
M          gtcagctgagcgccgtgatgttgccccagggtggaagaagccacacttcctacactgtca
      1921 ---------+---------+---------+---------+---------+---------+ 1980
H          tgtgaagccaccctgggctacagaaaccacagtcttcccagcaattattacaattcttga
M          gggcacttttaaacttctggaggg**gtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt**
      1981 ---------+---------+---------+---------+---------+---------+ 2040
H          attccttggggatttttactgccctttcaaagcacttaagtgttagatctaacgtgttc
M          **gtgtgtgtgtgtgtgtgt**tcattctgccccttccaaatcatctaagtgttatttaaggca
      2041 ---------+---------+---------+---------+---------+---------+ 2100
H          cagtgtctgtctgaggtgacttaaaaaatcagaacaaaacttctattatccagagtcatg
M          ctctgctgtttgtatgagatggttcatagaaattatgacaaagcctttgttatccaggcc
      2101 ---------+---------+---------+---------+---------+---------+ 2160
H          ggagagtacacccttccaggaataatgttttgggaaacactgaaatgaaatcttcccag
M          atgggaaaaggaaaaagaaaagaaagaaagaaaag*aataaa*agcttttgaggagcccctg
      2161 ---------+---------+---------+---------+---------+---------+ 2220
H          tattataaattgtgtatttaaaaaaaagaaacttttctgaatgcctacctggcggtgtat
M          ttaaaaaaaaaaaaaaa - 2235
H          accaggcagtgtcccagtttaaaaagatgaaaaag*aataaa*aactttgaggaaaaaaaa
H          aaaaaa - 2288
```

Figure 6. (*continued)* Nucleotide and amino acid sequence of human **(H)** and mouse **(M)** GM3 synthase. Position along nucleotide and amino acid sequence is indicated on the right and left margin, based on a position along mouse sequence. **1, 2,** and **3** - potential in frame initiation codons. The third M is selected as the first amino acid residue in the protein sequence based on homology with other sialytransferases. **STL** - large sialylmotif, **STS** - small sialylmotif, **ST3** - sialylmotif specific for ST3 family of sialyltransferases, **STP** - petit sialylmotif, probably involved in the recognition of cytidine portion of CMP-sialic acid. H* - The unique H residue in the place of D/E common to all known animal sialyltransferases. S* - The unique S residue in the place of N common to all other ST3 sialyltransferases. Non-conservely substituted amino acid residues are shown in capital bold. Stop - stop codon. The potential N-terminal transmembrane domain is shown in italic. The potential polyadenylation signal is shown in underlined italic. $(GT)_n$ family of repeats is shown in bold $\mathbf{(gt)_{27}}$. N^G indicates the potential N-glycosylation sites. S^{cP} or T^{cP} - indicates the potential cAMP and cGMP-dependent protein kinase phosphorylation sites. S^{CK} or T^{CK} - indicates the potential casein kinase II phosphorylation sites (only sites conserved between mouse and human are shown). S^C or T^C - potential sites of phosphorylation by protein kinase with the same substrate specificity as protein kinase C. $S^{cP,C}$ - serine residue that can be potentially phosphorylated by both cAMP or cGMP - dependent protein kinases and by protein kinases with the same substrate specificity as protein kinase C. ↕ followed by the number indicates places of alternative splicing sites. The nucleotide sequence of the primers used for the detection of the alternatively spliced forms (ASF) is shown in bold italic. Boxed amino acid residues indicate differences with GM3-synthases published by another group [44].

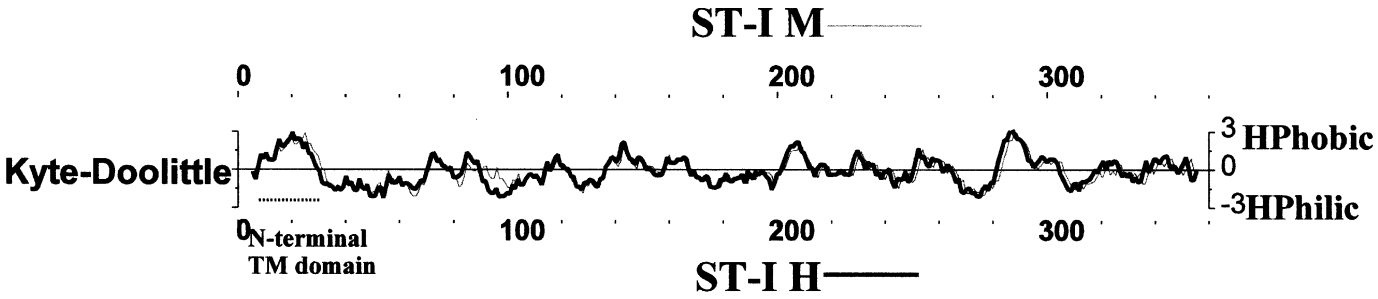

Figure 7. Hydrophobicity plot of mouse (grey line) and human (black line) GM3-synthase. The dotted line indicates N-terminal transmembrane domain common to all STs.

References

1 Tsuji S, Datta AK, Paulson JC (1996) [letter]. *Glycobiology* **6:** v–vii.
2 Bremer EG, Hakomori S, Bowen-Pope DF, Raines E, Ross R (1984) *J Biol Chem* **259:** 6818–25.
3 Bremer EG, Schlessinger J, Hakomori S (1986) *J Biol Chem* **261:** 2434–40.
4 Nojiri H, Takaku F, Ohta M, Miura Y, Saito M (1985) *Cancer Res* **45:** 6100–106.
5 Nojiri H, Takaku F, Terui Y, Miura Y, Saito M (1986) *Proc Natl Acad Sci USA* **83:** 782–86.
6 Nojiri H, Kitagawa S, Nakamura M, Kirito K, Enomoto Y, Saito M. (1988) *J Biol Chem* **263:** 7443–46.
7 Yada Y, Okano Y, Nozawa Y (1991) *Biochem J* **279:** 665–70.
8 Rahmann H, Rosner H, Kortje KH, Beitinger H, Seybold V (1994) *Brain Res* **101:** 127–45.
9 Ferrari G, Greene LA (1998) *Ann NY Acad Sci* **845:** 263–73.
10 Hakomori S (1990) *J Biol Chem* **265:** 18713–16.
11 Hakomori S (1997) *Sphingolipid-mediated Signal Transduction.* R.G.Landes Company and Chapman & Hall.
12 Hakomori S, Igarashi Y (1995) *J Biochem (Tokyo)* **118:** 1091–103.
13 Hynds DL, Burry RW, Yates AJ (1997) *Neurosci Res* **47:** 617–25.
14 Katoh N (1995) *Toxicology* **104:** 73–81.
15 Yates AJ, Rampersaud A (1998) *Ann NY Acad Sci* **845:** 57–71.
16 Suarez Pestana E, Greiser U, Sanchez B, Fernandez LE, Lage A, Perez R, Bohmer FD (1997) *Br J Cancer* **75:** 213–20.
17 Goldenring JR, Otis LC, Yu RK, DeLorenzo RJ (1985) *J Neurochem* **44:** 1229–34.
18 Kreutter D, Kim JY, Goldenring JR, Rasmussen H, Ukomadu C, DeLorenzo RJ, Yu RK (1987) *J Biol Chem* **262:** 1633–37.
19 Kim JY, Goldenring JR, DeLorenzo RJ, Yu RK (1986) *J Neurosci Res* **15:** 159–66.
20 Matecki A, Stopa M, Was A, Pawelczyk T (1997) *Int J Biochem Cell Biol* **29:** 815–28.
21 Yang FY, Wang LH, Yang XY, Tsui ZC, Tu YP (1997) *Biophys Chem* **68:** 137–46.
22 Misasi R, Sorice M, Garofalo T, Griggi T, Campana W, Giammatteo M, Pavan A, Hiraiwa M, Pontieri M, O'Brien J (1998) *J Neurochem* **71:** 2313–21.
23 Hakomori S, Yamamura S, Handa AK (1998) *Ann NY Acad Sci* **845:** 1–10.
24 Kojima N, Hakomori S (1991) *J Biol Chem* **266:** 17552–58.
25 Kojima N, Shiota M, Sadahira Y, Handa K, Hakomori S (1992) *J Biol Chem* **267:** 17264–70.
26 Hammache D, Yahi N, Pieroni G, Ariasi F, Tamalet C, Fantini J (1998) *Biochem Biophys Res Commun* **246:** 117–22.
27 Hammache D, Pieroni G, Yahi N, Delezay O, Koch N, Lafont H, Tamalet C, Fantini J (1998) *J Biol Chem* **273:** 7967–71.
28 Paulson JC, Beranek WE, Hill RL (1977) *J Biol Chem* **252:** 2356–62.
29 Gillespie W, Kelm S, Paulson JC (1992) *J Biol Chem* **267:** 21004–10.
30 Haraguchi M, Yamashiro S, Yamamoto A, Furukawa K, Takamiya K, Lloyd KO, Shiku H (1994) *Proc Natl Acad Sci USA* **91:** 10455–59.
31 Sasaki K, Watanabe E, Kawashima K, Sekine S, Dohi T, Oshima M, Hanai N, Nishi T, Hasegawa M (1993) *J Biol Chem* **268:** 22782–87.
32 Kim YJ, Kim KS, Do S, Kim CH, Kim SK, Lee YC (1997) *Biochem Biophys Res Commun* **235:** 327–30.
33 Nakayama J, Fukuda MN, Hirabayashi Y, Kanamori A, Sasaki K, Nishi T, Fukuda M (1996) *J Biol Chem* **271:** 3684–91.
34 Lee YC, Kurosawa N, Hamamoto T, Nakaoka T, Tsuji S (1993) *Eur J Biochem* **216:** 377–85.
35 Kapitonov D, Yu RK (1997) [dissertation] Medical College of Virginia of Virginia Commonwealth University, Richmond
36 Gu X, Preuss U, Gu T, Yu RK (1995) *J Neurochem* **64:** 2295–302.
37 Sipos L, von Heijne G (1993) *Eur J Biochem* **213:** 1333–40.
38 Nakashima H, Nishikawa K (1992) *FEBS Lett* **303:** 141–46.
39 Hartmann E, Rapoport TA, Lodish HF (1989) *Proc Natl Acad Sci USA* **86:** 5786–90.
40 Nigam SK, Blobel G (1989) *J Biol Chem* **264:** 16927–32.
41 Ou WJ, Thomas DY, Bell AW, Bergeron JJ. (1992) *J Biol Chem* **267:** 23789–96.
42 Sfeir C, Veis A (1995) *J Bone Miner Res* **10:** 607–15.
43 Ishii A, Ohta M, Watanabe Y, Matsuda K, Ishiyama K, Sakoe K, Nakamura M, Inokuchi J, Sanai Y, Saito M (1998) *J Biol Chem* **273:** 31652–55
44 Kono M, Takashima S, Liu H, Inoue M, Kojima N, Lee YC, Hamamoto T, Tsuji S (1998) *Biochem Biophys Res Commun* **253:** 170–75.
45 Kapitonov D, Yu RK (1999) *Glycobiology* (in press).

Received 15 April 1999, revised and accepted 25 June 1999