Polymerase Chain Reaction Cloning of a Developmentally Regulated Member of the Sialyltransferase Gene Family*

(Received for publication, March 30, 1993, and in revised form, April 9, 1993)

Brian D. Livingston and James C. Paulson†
From Cytel Corporation, San Diego, California 92121

We have cloned a new member of the sialyltransferase gene family from newborn rat brain cDNA using the polymerase chain reaction and degenerate primers based on sequences found in a continuous region of homology present in three previously cloned sialyltransferases. The predicted amino acid sequence of the new gene contains 24 of the 31 amino acids that are invariably conserved in the previously cloned sialyltransferases and encodes a type II transmembrane protein characteristic of glycosyltransferases. In contrast to previously cloned sialyltransferase genes, the expression of the new gene is highly restricted, being abundantly expressed in newborn brain but not in adult brain or any other adult or newborn tissues.

The sialyltransferases represent a family of terminal glycosyltransferases that catalyze the attachment of sialic acids to carbohydrates of glycoproteins and glycolipids, many of which have been shown to function in biological processes ranging from intracellular targeting to cell adhesion (1-3). Sialyloligosaccharide structures vary considerably in both the linkage of the sialic acid as well as the acceptor sugar to which it is attached. Based on the fact that these enzymes are specific for the synthesis of a single linkage, a survey of the known sialic acid-containing oligosaccharides suggests that this gene family contains at least 10-12 members. To date only five sialyltransferases have been purified, and four of these have been cloned (4-11).

Comparison of the cloned sialyltransferases has revealed a conserved amino acid sequence motif located in the catalytic domain of these enzymes, which we refer to here as the sialylmotif. The most obvious homologous region of the motif consists of a highly conserved 55-amino acid sequence near the center of these enzymes (10). In this region, 42% of the amino acids are identical in each of the three cloned sialyltransferases. Recently, Drickamer (12) extended the motif by observing two additional regions of homology in the C-terminal end of these glycosyltransferases. None of the elements in the conserved motif are found in other glycosyltransferases such as fucosyltransferases (13-17), galactosyltransferases (18-21), or N-acetylgalactosaminyltransferases (22). Thus, in this report we have referred to this motif as the sialylmotif.

The conservation of the sialylmotif in all three cloned sialyltransferases predicted that other members of the sialyltransferase gene family might also be expected to contain the same motif. This suggested that it might be possible to use sequence information present in the sialylmotif to isolate new members of this gene family. Accordingly we attempted to use degenerate primers to either end of the highly conserved N-terminal segment of the sialylmotif in a PCR-based approach to clone new members of this gene family. Using newborn rat brain cDNA as a template a PCR fragment encoding a unique sialylmotif was identified. This PCR fragment was used as a probe to obtain a full-length copy of the gene. Like the other three cloned members of the sialyltransferase gene family, the new gene codes for a protein that contains the complete sialylmotif and is predicted to have a type II membrane topology characteristic of glycosyltransferases. The most striking characteristic of the new gene is its unusual pattern of regulation; whereas the other cloned members of the sialyltransferase gene family are differentially expressed in a variety of tissues (9, 10, 23), the new gene is only expressed in newborn brain.

EXPERIMENTAL PROCEDURES

PCR Cloning with Degenerate Oligonucleotides—Two degenerate oligonucleotides corresponding to the 5'- and 3' ends of the N-terminal segment of the sialylmotif (Fig. 1) were synthesized (Genosys). The sequences of the 5'- and 3'-primers were 5'GGAAGCTTTGSCRNMGSTGGYWRCTGCOT and 5'CGGATCCGGTRG"TTNSNSCCSACRCTC A+G+T+C;S=G+C;R=A+G;M=A+C;Y=C+T), respectively. PCR experiments were performed using 100 pmol of each primer and first strand cDNA synthesized from newborn rat brain as a template. Amplification was carried out by 30 cycles of 94 °C for 1 min, 37 °C for 1 min, and 72 °C for 2 min. The PCR products were digested with BamHI and HindIII and subcloned into these sites of Bluescript KS (Stratagene). Subclones were characterized by sequencing with a T3 primer.

Cloning of the STX Gene—Random primed newborn rat brain cDNA (24) was ligated with EcoRI-NorI linkers and then subsequently ligated into EcoRI-digested Agt10 (Promega). The resultant library was packaged using a Stratagene Gigapack II packaging extract and plated on Escherichia coli C600. Approximately 109 plaques were screened with the cloned PCR fragment. Four clones, STX1-4, were purified and subcloned into the NorI site of Bluescript (Stratagene) for further analysis.

Northern Analysis—Total RNA from rat tissues was prepared using an acid phenol procedure as described previously (25). newborn RNA samples were isolated from rat pups within 4 days of birth. RNA was electrophoresed in a 1% agarose gel containing formaldehyde, transferred to nitrocellulose, and hybridized following standard procedures (24). Northern blots were probed with a gel-purified, radiolabeled, 900-bp EcoRI fragment isolated from STX1.

Construction of a Soluble Form of STX—A truncated form of STX, lacking the first 31 amino acids of the open reading frame, was prepared by PCR amplification with a 5'-primer containing an in-frame BamHI site and a 3'-primer located 50 bp downstream of the stop codon. Amplification was carried out by 30 cycles of 94 °C for 1 min, 45 °C for 1 min, and 72 °C for 2 min. The fusion vector pGR120IprotA was constructed by inserting a BclIV/BamHI fragment, isolated from pRIT5 (Pharmacia LKB Biotechnology Inc.), encoding the protein A IgG binding domain into the BamHI site of pGR120I (a gift from Dr. K. Drickamer, Columbia University). The amplified fragment was subcloned into the BamHI site of pGR120IprotA resulting in fusion of STX to the protein A IgG binding domain.

*The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L13445.

Deoxyribonucleic acid samples were isolated from rat pups within 4 days of birth. RNA was electrophoresed in a 1% agarose gel containing formaldehyde, transferred to nitrocellulose, and hybridized following standard procedures (24). Northern blots were probed with a gel-purified, radiolabeled, 900-bp EcoRI fragment isolated from STX1. Amplification was carried out by 30 cycles of 94 °C for 1 min, 45 °C for 1 min, and 72 °C for 2 min. The fusion vector pGR120IprotA was constructed by inserting a BclIV/BamHI fragment, isolated from pRIT5 (Pharmacia LKB Biotechnology Inc.), encoding the protein A IgG binding domain into the BamHI site of pGR120I (a gift from Dr. K. Drickamer, Columbia University). The amplified fragment was subcloned into the BamHI site of pGR120IprotA resulting in fusion of STX to the protein A IgG binding domain.

The abbreviations used are: PCR, polymerase chain reaction; bp, base pair(s); kb, kilobase(s); G418, 418 ganciclovir; NeuAc2,3Galβ1,3GalNAcβ1,4(NeuAc2,3)Galβ1,4Gle-ceramide; G418, 418 ganciclovir; NeuAc2,3Galβ1,3GalNAcβ1,4(NeuAc2,3)Galβ1,4Gle-ceramide.
insulin signal sequence and the protein A present in the vector. A 100-bp fragment containing the fusion protein was subcloned into pSVL (Phar- 
imaceutical) resulting in the expression plasmid AX78.

Expression of the Soluble Form of STX—The expression plasmid AX78 (10 μg) was transfected into COS-1 cells in 10-cm plates using lipofectin as suggested by the manufacturer (Bethesda Research Laboratories). Two days after transfection 1 ml of the culture medium was collected and incubated with 10 μl of IgG-Sepharose (Pharmacia) for 1 h at room temperature. The beads were washed, resuspended in 25 μl of assay buffer, and assayed for sialyltransferase activity using oligosac- 
charides, glycoproteins, and gangliosides (26). Transfer of sialic acid to these acceptors was measured using ion-exchange and size exclusion chromatography (26).

Identical transfections were performed for pulse-chase labeling experiments. Following a 36-h expression period the plates were incubated at 37 °C with 2.5 ml of Dulbecco’s modified Eagle’s medium. After 1 h 250 μCi of Trans-P32-label (Amersham Corp.) was added to the medium, and the plates were incubated for an additional 3 h. At the end of this time the plates were washed with PBS and incubated overnight with 5 ml of complete Dulbecco’s modified Eagle’s medium. Labeled fusion protein was isolated by incubation of 1 ml of labeled medium with 10 μl of IgG-Sepharose (Pharmacia). Binding was carried out for 1 h at room temperature, the beads were washed with PBS and boiled in Laemmli sample buffer, and the released proteins were analyzed by SDS-polyacrylamide gel electrophoresis/fluorography.

RESULTS

PCR Amplification of a Sialylmotif Fragment Related to Those Found in Characterized Sialyltransferases—While the sialylmotif is composed of a number of amino acids dispersed throughout the sialyltransferase catalytic domain, the largest continuous regions of conservation are found in the sequences that border the highly conserved N-terminal segment of the motif. This fact suggested that primers corresponding to these amino acid sequences could be used in a PCR-based approach to amplify cDNAs that share this region of the sialylmotif. This approach would allow for the rapid characterization of the largest continuous portion of a new sialylmotif without first having to clone the entire gene. Accordingly oligonucleotide primers were made to the N-terminal and C-terminal ends of the highly conserved N-terminal segment of the sialylmotif (Fig. 1). The strong conservation of the C-terminal sequence of this region allowed for the design of an oligonucleotide primer with only a 256-fold degeneracy. The N-terminal sequence of the sialylmotif is more variable and is comprised of amino acids with higher codon redundancy; as a result, in order to account for all the nucleotide sequences found to encode for this region of the sialylmotif the primer was synthesized with a 1026-fold degeneracy. This degree of complexity is near the threshold of degeneracy allowable in PCR experiments although such primers have been successfully used in the past (27).

A 150-bp cDNA was chosen as a source from which to isolate cDNA to use in the initial PCR experiments since it is an abundant source of sialyloligosaccharides, such as polysialic acid (28) and the carbohydrate structures found in gangliosides (7), which are synthesized by sialyltransferases that have not yet been cloned. Using newborn rat brain cDNA as a template, PCR with the degenerate primers resulted in the amplification of a 150-bp product equal to the known size of the highly conserved region of the sialylmotif. Subcloning and sequencing revealed that the band was a mixture of two DNA fragments. Of 30 clones characterized 17 encoded the highly conserved N-terminal region of the c2,6-sialylmotif; the remaining clones encoded an incomplete sialylmotif fragment. This new sialylmotif fragment, STX (Fig. 1), encodes 17 of the amino acids found to be invariable in the three previously cloned sialyltransferases. Nine additional amino acids were found in two of the three previously characterized sialylmotifs. An additional 10 amino acids were found in at least one of the three sequences, suggesting that residues that were previously thought to be variable residues might also be conserved.

Primary Structure of the STX Gene—In order to isolate the complete coding sequence of the gene containing the new sialylmotif, the amplified fragment STX was used to screen a newborn rat brain cDNA library. Sequence analysis of the largest clone, STX1 (1.5 kb), revealed a continuous 375-amino acid open reading frame that contained the highly conserved N-terminal segment of the STX sialylmotif (Fig. 2). The predicted amino acid sequence also contains five of the additional eight invariant C-terminal conserved residues identified by Dricka- 
mere in the other cloned sialyltransferases (Fig. 3). The primary sequence characteristics of the STX protein indicates additional similarities to other members of the sialyltransferase gene family, such as the size and the apparent topology of the STX protein. The protein encoded by the STX gene contains a 17-amino acid N-terminal hydrophobic sequence bordered by charged residues that would be anticipated to serve as a signal- 
anchor domain, which would result in a type II transmembrane orientation characteristic of all other glycosyltransferases cloned to date (29).

Expression of the STX Protein—To facilitate functional anal- 
ysis of STX a soluble form of the protein was generated by replacing the first 31 amino acids including the transmembrane domain with the cleavable insulin signal sequence and the protein A IgG binding domain (20). The IgG binding domain was included in the construction to aid in the detection of the soluble STX protein. Similar fusions with the ST3N are readily secreted from expressing cells, bound by IgG-Sepharose, and are enzymatically active. When an expression plasmid containing the protein A/STX fusion (AX78) was expressed in COS-1 cells, an 85-kDa protein was detected (Fig. 4). The size of the fusion protein is approximately 15 kDa greater than the predicted molecular mass of the polypeptide suggesting that several of the potential N-linked glycosylation sites present in the amino acid sequence are being utilized.

The bound fusion protein was assayed for sialyltransferase activity using a variety of acceptor substrates. Activity was not detected using glycoprotein acceptors including asialo deriv- 
atives of mucin or antifreeze glycoprotein. Activity assays using oligosaccharide acceptors containing either Galβ1,3 or Galβ1,4GlcNAc linkages, Galβ1,3GalNAc sequences, or colo- 
minal acid and sialylated T antigen were also negative. STX fusion protein was also assayed for activity using a variety of glycolipids including lactosylceramide, GMI, GD3, and mixed gangliosides isolated from bovine brain; however, these also failed to serve as an acceptor substrates.

Developmental and Tissue-specific Expression of STX—In or- 
der to determine the pattern of expression and message size of the STX gene, total RNA isolated from newborn brain, adult brain, newborn kidney, adult kidney, liver, spleen, intestine, submaxillary gland, and lung was probed with a 500-bp EcoRI fragment isolated from STX1. Of the various tissues examined hybridization of a 5.5-kb message was only observed in new- 
born rat brain RNA (Fig. 5). No cross-hybridization to known

FIG. 1. Predicted amino acid sequence of the amplified fragment, STX, and comparison to the highly conserved N-terminal segment of the sialylmotif in the three previously cloned sialyltransferases. The three cloned sialyltransferases are the porcine Galβ1,3GlcNAc α2,3-sialyltransferase (ST3O, residues 142–189), the rat Galβ1,3/4GlcNAc α2,3-sialyltransferase (ST3N, residues 156–203), and the rat Galβ1,4GlcNAc α2,6-sialyltransferase (ST6N, residues 176–225). Amino acids conserved in any two sequences are boxed. Positions of the PCR primers are indicated by arrows.
sialyltransferase mRNAs was observed. The restricted expression of STX is a departure from the differential tissue-specific expression found with characterized sialyltransferases. While each of these genes is independently regulated resulting in different patterns of tissue-specific expression, in general each sialyltransferase is variably expressed in a number of diverse tissues (9, 10, 12). In contrast STX is only expressed in newborn brain; the expression does not appear to be a generalized embryonic phenomenon as the message was not detected in newborn kidney.

**DISCUSSION**

In recent years the carbohydrates of glycoproteins and glycolipids have been increasingly recognized for their information content in mediating biological processes (1–3). Consequently there has been greater interest in the diversity and regulation of the glycosyltransferase genes responsible for their synthesis. Of the 150–200 glycosyltransferase genes estimated to be required for their synthesis, approximately 12 have been cloned to date. Cloning these enzymes from protein sequence information has resulted in the isolation of several sialyltransferases (15, 16), galactosyltransferases (18–21), and a N-acetylgalactosaminyltransferase (22), but this approach is severely limited by the difficulty in purifying these low abundance proteins. An alternative approach using an expression cloning system developed by Lowe and co-workers (13–15) has achieved significant success in obtaining genes encoding fucosyltransferase genes. Additional members of the fucosyltransferase gene family have been isolated by a low stringency approach (16, 17). While high stringency, cross-hybridization using the cloned α2,6-sialyltransferase as a probe identified several alternatively spliced messages produced from this gene (30), low stringency hybridization was unsuccessful in isolating any new sialyltransferase genes.

**FIG. 4. Expression of a soluble protein A/STX fusion protein.** COS-1 cells transfected with the expression plasmid AX78 or mock transfected were pulse-chase labeled. Following the chase incubation the medium was collected and incubated with IgG-Sepharose. beads were washed and boiled in Laemmli sample buffer, and the eluted protein was subjected to SDS-polyacrylamide gel electrophoresis/fluorography. While there was no apparent binding of proteins in the medium of mock transfected cells, an 85-kDa protein was readily isolated from the medium of cells transfected with AX78. Taking into account the use of the STX N-linked glycosylation sites the molecular weight corresponds to the predicted size of the fusion protein.
Fig. 5. Developmental and tissue-specific expression of STX. RNA from various adult and newborn rat tissues were hybridized with a probe for STX. A 5.5-kb message was detected in newborn rat brain but not in adult brain or any of the other tissues examined. The filter was exposed for 8 h; longer exposures (12–72 h) did not indicate hybridization to any additional messages.

As described in this report, a PCR-based approach using sequence information in the sialylmotif seemed an attractive alternative for obtaining new members of the sialyltransferase gene family for two reasons. The relative position of the sialylmotif and the nature of the conserved residues suggest that the sialylmotif might be necessary for sialyltransferase activity. If this hypothesis was true, then the sialylmotif might be a structural characteristic common to many of the genes in the sialyltransferase gene family. Second, the most highly conserved groups of residues in the sialylmotif are at either end of a continuous 45-amino acid region. Thus, degenerate primers corresponding to the 5′- and 3′-ends were expected to amplify this region in related genes and resulted in the successful cloning of the STX gene.

The most strikingly disparate characteristic of the STX gene relative to the other members of the sialyltransferase gene family regards the regulation of the gene. Each of the other members of this gene family is independently regulated such that each is expressed at variable levels in a variety of tissues. Insights into the means by which sialyltransferase expression is regulated have come from analysis of the a2,6-sialyltransferase gene. While transcription can occur from multiple promoters, the high expression of a2,6-sialyltransferase observed in liver results from the activation of a liver-restricted promoter by the binding of liver-restricted transcription factors (31). The regulated expression of the STX gene may result from a similar mechanism mediated by stage-specific transcription factors.

The developmental regulation of the STX gene suggests a possible explanation for the difficulties in the detection of enzymatic activity. Preliminary analysis has only entailed known oligosaccharide acceptor structures. It is possible that the STX protein may be involved in the biosynthesis of a developmentally regulated, tissue-specific sialyoligosaccharide structure that has not yet been described. If this is the case, then the appropriate oligosaccharide substrates may not have been tested. Although the fusion protein was readily secreted from transfected cells, it is possible that the difficulties in the enzymatic characterization of STX stem from the expression of an incorrectly folded form of the protein. Alternatively, STX may be enzymatically unrelated to the sialyltransferases but have functional aspects similar to these genes, such as CMP-sialic acid binding or possibly a lectin-like activity.

Traditionally, cloning sialyltransferases required that the protein first be purified, a difficult process that has limited the extent to which these enzymes have been studied. The STX gene is the first member of the sialyltransferase gene family to be cloned without first being enzymatically defined and subsequently purified. The characterization of the sialylmotif and the demonstration that these conserved sequences can be used to clone new members of this gene family represent an advancement that may lead to a greater understanding of this family of glycosyltransferases.

Acknowledgments—We thank Drs. Hiroshi Kitagawa and Dawn Wen for advice and suggestions and Dr. Mark Williams and Sue Hymer for generously providing the fusion vector pGIR201protA.

REFERENCES

1. Kornfeld, S. (1987) FASEB J. 1, 462-468
2. Philips, M. L., Nudelman, E., Gaeta, F. C. A., Perez, M., Singhal, A. K., Harkomi, S., and Paulson, J. C. (1990) Science 250, 1130-1134
3. Paulson, J. C. (1989) Trends Biochem. Sci. 14, 272-276
4. Weinstein, J., de Souza-e-Silva, U., and Paulson, J. C. (1982) J. Biol. Chem. 257, 13835-13844
5. Weinstein, J., de Souza-e-Silva, U., and Paulson, J. C. (1982) J. Biol. Chem. 257, 13845-13853
6. Sadler, J. E., Rearick, J. J., Paulson, J. C., and Hill, R. L. (1970) J. Biol. Chem. 254, 5934-5941
7. Melkerson-Watson, L. J., and Sweeney, C. C. (1991) J. Biol. Chem. 266, 4448-4457
8. Weinstein, J., Lee, E., McEntee, K., Lai, P., and Paulson, J. C. (1987) J. Biol. Chem. 262, 17735-17743
9. Gillespie, W., Kelm, S., and Paulson, J. C. (1992) J. Biol. Chem. 267, 21004-21010
10. Wen, D., Livingston, B. D., Medzhihradszky, K., Kelm, S., Burlingame, A. L., and Paulson, J. C. (1992) J. Biol. Chem. 267, 21011-21019
11. Grundmann, U., Nerlich, C., and Zettlmeissl, G. (1990) Nucleic Acids Res., 18, 867
12. Dricknamer, K. (1993) Glycobiology 3, 2-3
13. Larsen, R. D., Ernst, L. K., Nair, R. P., and Lowe, J. B. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 6674-6678
14. Kukowska-Latallo, J. F., Larsen, R. D., Nair, R. P., and Lowe, J. B. (1990) Genes & Dev. 4, 1288-1303
15. Geolz, B. A., Hession, C., Goff, D., Griffiths, B., Tizard, R., Newman, B., Chi-Rosso, G., and Lobb, R. (1990) Cell 63, 1349-1356
16. Lowe, J. B., Kukowska-Latallo, J. F., Nair, R. P., Larsen, R. D., Marks, R. M., Machler, B. A., Kels, S., and Ernst, L. K. (1991) J. Biol. Chem. 266, 17467-17477
17. Weston, B. W., Nair, R. P., Larsen, R. D., and Lowe, J. B. (1992) J. Biol. Chem. 267, 4152-4160
18. Shaper, J. H., Hollis, G. F., Douglas, J. G., Kirsch, I. R., and Shaper, J. H. (1988) J. Biol. Chem. 263, 10420-10428
19. Joziassie, D. H., Shaper, J. H., Van den Enjenden, D. H., Van Tunen, A. J., and Shaper, N. L. (1989) J. Biol. Chem. 264, 14280-14287
20. Yamamoto, F., Marken, J., Tsui, T., White, F., Clausen, H., and Hakomori, S. (1990) J. Biol. Chem. 265, 1146-1151
21. Yamamoto, F., Clausen, H., White, T., Marken, J., and Hakomori, S. (1990) Nature 345, 229-233
22. Sarkar, M., Hull, E., Nishikawa, Y., Simpson, R. J., Moritz, R. L., Dunn, R., and Schachter, H. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 234-238
23. Paulson, J. C., Weinstein, J., and Schauer, A. (1988) J. Biol. Chem. 263, 10931-10934
24. Krieger, M. (1990) Gene Transfer and Expression, pp. 125-130, Stockton Press, New York
25. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156-159
26. Weinstein, J., de Souza-e-Silva, U., and Paulson, J. C. (1982) J. Biol. Chem. 257, 13835-13844
27. Innis, M., Gelfand, D. H., Sninsky, J. J., and White, T. (1990) PCR Protocols, pp. 39-46, Academic Press, San Diego, CA
28. Rutishauser, U., Ahlborn, A., Hall, A. K., Mann, D. M., and Sunshine, J. (1988) Science 240, 53-57
29. Paulson, J. C., and Colley, J. K. (1989) J. Biol. Chem. 264, 17615-17618
30. Wen, D. X., Svensson, E. C., and Paulson, J. C. (1992) J. Biol. Chem. 267, 25123-25128
31. Svensson, E. C., Soreghan, B., and Paulson, J. C. (1990) J. Biol. Chem. 265, 20863-20868