Characterization of Two Promoters of a Rat Pregnancy-Specific Glycoprotein Gene†

Hungwen Chen,‡,§ Chao-Ling Chen,§ and Janice Yang Chou*,‡

Human Genetics Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, and Department of Large Animal Clinical Sciences, College of Veterinary Medicine, University of Florida, Gainesville, Florida 32610

Received March 29, 1994; Revised Manuscript Received May 19, 1994®

ABSTRACT: Pregnancy-specific glycoproteins (PSGs) are the major placental glycoproteins, that together with the carcinoembryonic antigens comprise a subfamily within the immunoglobulin superfamily. In order to develop an animal model for understanding the molecular mechanisms underlying the control of PSG expression, we isolated and characterized cDNA and genomic clones encoding a rodent PSG, rnCGM3. The rnCGM3 cDNA is 2761 bp in length and contains an open reading frame that encodes a 475 amino acid polypeptide with a domain arrangement of L_1N_1-L_2N_2-L_3N_3-A. The sequence in the 5′-untranslated and L_1 regions of rnCGM3 is identical to a previously identified cDNA, rnCGM6. The transcription initiation sites of both genes are located at nucleotide −197 upstream of the translation start site. In transient transfection assays using a chloramphenicol acetyltransferase (CAT) reporter gene, we demonstrated that DNA elements at nucleotides −326 to −185 (PI) and −147 and −86 (PII) relative to the translation start site of rnCGM3 could both function as promoters. The downstream promoter, PII, which is located within the first exon, shares high sequence identity with the minimal promoters of human PSG genes. Electrophoretic mobility shift assays (EMSAs) showed that protein factors in placental cell extracts formed three complexes (PIICI, PIICII, and PIICIII) with the PII promoter element. The PIICIII complex was also observed by DNase I footprinting analysis. Unlike PII, the upstream promoter, PI, contains a TATA box. DNase I footprinting analysis revealed two nuclear protein binding sites at nucleotides −311 to −290 (PISI) and −257 to −239 (PISII) in PI. EMSAs showed that protein factors in placental cell extracts bound to both sites and deletion of either site markedly reduced CAT expression. PISII contains a palindromic motif, *TGTTGCTCAACA*, and protein cross-linking and Southwestern hybridization analyses demonstrated that the protein factor binding to PISII had an apparent molecular mass of 40 kDa.

Pregnancy-specific glycoprotein (PSG) genes and the related carcinoembryonic antigen (CEA) genes comprise a family within the immunoglobulin (Ig) superfamily of genes [for reviews, see Thompson et al. (1991) and Chou and Plouzek (1992)]. Human PSG and CEA both play vital roles in clinical diagnosis: PSG is used to diagnose pregnancy, to predict pregnancy-related complications, and to monitor the treatment of choriocarcinoma (Tatarinov, 1978; Wurz et al., 1981; MacDonald et al., 1983; Masson et al., 1983; Tamsen et al., 1983; Sorensen, 1984), and CEA is employed to monitor postoperative tumor recurrences (Shively & Beatty, 1985). CEA proteins have been shown to function as adhesion molecules *in vitro* (Benchimol et al., 1989; Turbide et al., 1991), and PSGs have been implicated in the maintenance of pregnancy (Bohn & Weinmann, 1974; Wurz et al., 1981; Masson et al., 1983). However, the exact *in vivo* functions of PSG and CEA are currently unknown. In order to study the regulation of PSG/CEA expression and the in *vivo* roles of these proteins using an animal model, several rodent PSG/CEA cDNAs and genes have been isolated and characterized (Kodelja et al., 1989; Rebstock et al., 1990; Turbide et al., 1991; Chen et al., 1992; Rudert et al., 1992). Since understanding of the control of PSG/CEA expression may yield insight into possible function, we initiated study by examining the control of expression of a rodent PSG gene.

Both human and rodent PSG genes are expressed in high levels only in the placenta (Thompson et al., 1991; Chen et al., 1992). Studies have indicated that tissue-specific expression of a variety of genes is determined at the level of transcription (Maniatis et al., 1987). Therefore, the regulation of PSG expression offers an excellent model to study trophoblast-specific gene regulation. Recently, Nachtigal et al. (1993) identified highly conserved sequences (PSF-A and PSF-B) located in the 5′-flanking region of chorionic somatomammotropin and placental growth hormone variant genes that repress the expression of these genes in the pituitary but not in placental cells. A putative protein factor (designated PSF-1) which recognizes both PSF-A and PSF-B was shown to mediate pituitary-specific repression of placental genes by interacting with Pit 1, a pituitary-specific transcription factor (Nachtigal et al., 1993). This suggests that repression of placenta-specific gene expression in nontrophoblastic cells may be regulated by an array of cis-acting DNA elements and transcription factors.

In the present report, we have characterized both cDNA and genomic clones encoding rnGCM3, a rodent PSG, and defined sequences important in PSG expression. rnCGM3 contains three IgV-like (N) domains, each preceded by a leader peptide (L), and one IgC-like carboxyl (A) domain in a L_1N_1-L_2N_2-L_3N_3-A domain arrangement. Our study also shows that the 5′-flanking region of rnCGM3 contains two promoter

† The nucleotide sequences reported in this paper have been submitted to the GenBank/EMBL Data Bank under Accession Numbers U09814 and U09815.

* Correspondence should be addressed to this author at Building 10, Room 9S242, NIH, Bethesda, MD 20892. Telephone: 301-496-1094. Fax: 301-402-0234.

‡ National institutes of Health.

§ University of Florida.

® Abstract published in *Advance ACS Abstracts*, July 15, 1994.

elements (PI and PII) located upstream and downstream of the transcription start site, respectively. Two protein binding sites, PISI and PISII, were observed in the upstream promoter (PI), and deletion of either binding site decreased rnCGM3 promoter activity, indicating that each site contains activator elements essential for rnCGM3 expression. The PISII site contains a consensus half-site (TGTTG) for the putative transcription factor, PSF-1, found in chorionic somatomammotropin and placental growth hormone variant genes (Nachtigal et al., 1993). Whereas PSF-1 is shown to be absent in the placenta, we demonstrated that a 40-kDa protein in placental nuclear extracts bound to the PISII site in the rnCGM3 promoter. We also showed that the downstream promoter, PII, which is located in the first exon of the rnCGM3 gene, forms at least three protein–DNA complexes with factors present in placental cell extracts.

MATERIALS AND METHODS

Library Screening and Characterization of Genomic Clones. The rnCGM3 genomic clone was obtained by screening 4 × 10^5 plaques from a rat liver genomic library in EMBL3 (Promega Biotech, Madison, WI) with a rnCGM6 rat PSG cDNA probe (Chen et al., 1992). Two strongly positive overlapping clones were isolated and extensively characterized. The genomic inserts were subcloned into pGEM vectors for further characterization. Sequencing was accomplished by the Sanger dideoxy chain-termination method (Sanger et al., 1977). Both strands of the genomic clones were sequenced, and the sequences of the genomic clones were compared to cDNA sequences as a means to identify intron–exon junctions.

rnCGM3 cDNA Isolation. Reverse transcriptase–polymerase chain reaction (RT–PCR) and rapid amplification of cDNA ends (RACE) (Frohman et al., 1988) were utilized to isolate the cDNA encoding rnCGM3. The first-strand rat placental cDNA was synthesized using 5 μg of rat placental poly(A)$^+$ RNA as template and a first-strand cDNA synthesis kit obtained from Pharmacia LKB Biotechnologies Inc. The rnCGM3-specific primers used to PCR-amplify the rnCGM3 transcript were the following: P1s, 5′-*Bam*HI-CATTATAGCTTTCGCCTGGTA-3′ (nucleotides 183–203, Figure 1B); and P1as, 5′-*Xba*I-GTGTACAAGCCAGTATCTCTTGC-3′ (nucleotides 1082–1060). The amplified fragment was subcloned into a pGEM vector for further characterization.

To locate the 5′ end sequence and transcription start site of rnCGM3, we amplified rat placental poly(A)$^+$ RNA by RACE using a kit obtained from Boehringer Mannheim Biochemicals. The 5RT primer contains nucleotides 471–452 (5′-AGTGGGCTCGTCAATGTTAG-3′) in rnCGM3, and the 5AMP primer contains nucleotides 203–183 (5′-TACCAGGCGAAAGCTATAATG-3′). The rnCGM3-specific primer used for the 3′-RACE reaction contains nucleotides 1052–1074 (5′-ACCTCACTGCAAGAGATACTGGC-3′). The RACE products were directly cloned in the pCRII vector (Invitrogen, San Diego). The rnCGM3 sequence was verified by sequencing three or more PCR and RACE products.

Ribonuclease Protection Assays. Ribonuclease protection assays were performed essentially as described (Ausubel et al., 1992). Briefly, 25 μg of total rat placental RNA or yeast tRNA was hybridized at 45 °C for 18 h in a reaction mixture (30 μL) containing 80% formamide, 0.4 M NaCl, 0.04 M Pipes, pH 6.4, 1 mM EDTA, and an antisense rnCGM3 or rnCGM6 riboprobe (5 × 10^5 cpm). The hybrids were digested with RNase A (40 μg/mL) and RNase T1 (2 μg/mL) for 60 min at 30 °C and then electrophoresed through 8% polyacrylamide–urea sequencing gels. Single-base mismatches were normally not digested under the assay conditions used.

Construction of Promoter–CAT Fusion Genes, Transfection, and CAT Assays. rnCGM3 promoter fragments were synthesized by PCR using the rnCGM3 gene as a template. The 3′-end primers (5′ to 3′) are nucleotides −33 to −50, −86 to −104, and −185 to −202, and the 5′-end primers (5′ to 3′) are nucleotides −1465 to −1448, −1063 to −1047, −610 to −593, −513 to −496, −412 to −395, −326 to −310, −283 to −266, −236 to −219, −192 to −176, −147 to −130, and −96 to −79 in the rnCGM3 gene (see Figure 3A). Each oligo contains either an additional *Hin*dIII or an *Xba*I linker. The amplified fragments were digested with *Hin*dIII and *Xba*I and ligated into a modified promoterless and enhancerless pCAT-Basic-N plasmid (Lei et al., 1993). All constructs were confirmed by DNA sequencing. At least two preparations of each construct were tested by transfection.

rnCGM3(−412/−185ΔSI), containing an internal deletion of nucleotides −311 to −284, was generated as follows: a fragment containing nucleotides −412 to −312 of the rnCGM3 5′-flanking region was synthesized by PCR using rnCGM3 as a template and a pair of primers containing nucleotides −412 to −395 and −312 to −328 in rnCGM3. The amplified fragment was then ligated to the 5′ end of the rnCGM3 insert in the rnCGM3(−283/−185)CAT plasmid. rnCGM3(−412/−185ΔSII) and rnCGM3(−326/−185ΔSII), containing an internal deletion of nucleotides −283 to −237, were constructed similarly except the primers for rnCGM3(−412/−185ΔSII) were nucleotides −412 to −395 and −284 to −300 and for rnCGM3(−326/−185ΔSII), nucleotides −326 to −310 and −284 to −300. The amplified fragments were ligated to the 5′ end of the rnCGM3 insert in the rnCGM3(−236/−185)-CAT plasmid.

HP-A1 human placental cells (Lei et al., 1992a) were grown at 33 °C in α-modified minimal essential medium supplemented with 4% fetal bovine serum, streptomycin (100 μg/mL), and penicillin (100 units/mL). Cells in 150-cm^2 flasks were transfected in suspension by incubating with 2 mL of a calcium phosphate–DNA coprecipitate containing 50 μg of plasmid DNA and 2 μg of RSVluc as an internal control for transfection efficiency (de Wet el al., 1987). The CAT activity was assayed by incubating total cellular protein in a buffer containing 250 mM Tris, pH 7.8, 4 mM acetyl coenzyme A, and 0.1 μCi of [^{14}C]chloramphenicol (Fordis & Howard, 1987). Routinely, the assay was run for 1 h with the amount of extract required to convert 0.5–50% of the substrate to the acetylated forms. Assays outside this range were repeated using the appropriate amount of extract. The acetylated compounds were separated from chloramphenicol by thin-layer chromatography (95% chloroform–5% methanol v/v) on silica gel IB2 (Gilman Sciences) and quantitated by an AMBIS Radioanalytic Imaging System (San Diego, CA). The luciferase assays were performed as previously described (Lei et al., 1993).

EMSA and DNase I Footprinting Analysis. For EMSA, whole cell or nuclear extracts of HP-A1 cells were prepared as described by Snape et al. (1991) and Dignam et al. (1983), respectively. For DNase I footprinting analysis, nuclear extracts were prepared by the method of Ohlsson and Edlund (1986). Oligonucleotide probe was labeled with [α-^{32}P]dCTP using the Klenow fragment. In the EMSA, 10 μg of whole cell or nuclear extract was incubated with 0.5 ng of probe [(0.5–1) × 10^5 cpm] in a 20 μL reaction mixture containing 10 mM Tris, pH 7.5, 100 mM NaCl, 0.5 mM DTT, 10% glycerol, and 0.05% NP-40. After incubation at room

Table 1: Nucleotide Sequence of Exon/Intron Borders of the Rodent PSG Gene, rnCGM3[a]

Junction	Exon	Donor Intron Acceptor	Exon
$L-N_1$	...CTG CTC ACA	G gtaagtgtt.....ttcccttctag	CC TCA CTT TTA...
	Leu Leu Thr	A	la Ser Leu Leu
N_1-N_2	...CAC GTG AAA	T gtaagtaac.....ttcccttctag	CC TCT CTT TTC...
	His Val Lys	S	er Ser Leu Phe
N_2-N_3	...CAG GTG AAC	A gtaagtgat.....ttcccttccag	CC TCC CTT TCC...
	Gln Val Asn	T	hr Ser Leu Ser
N_3-	...ACC GTG CAC	A gtaagtgac.....	
	Thr Val His		

[a] Exon sequences are in upper-case letters. Consensus splice donors and acceptors are underlined.

temperature for 20 min, the reaction mixtures were analyzed by electrophoresis on 5% nondenaturing polyacrylamide gels. For competition, a nonlabeled fragment was preincubated with cell extract for 15 min on ice before the addition of probe. An Oct1 oligo (5′-GATCCATGCAAATGGATC-3′) was used as a nonspecific competitor.

DNase I footprinting analysis was performed essentially as described by Wu et al. (1987). Briefly, after incubation of a 50 μL EMSA reaction mixture at room temperature for 20 min, 10 units of DNase I was added, and the incubation was continued for an additional 70 s at room temperature. DNase I digestion was terminated by the addition of Na_2EDTA to a final concentration of 4 mM and BSA to 10 μg. The reaction mixture was separated in a 1% agarose gel and then blotted onto DEAE-nitrocellulose paper. The protected and free probes were eluted and analyzed on 8% urea–polyacrylamide gels.

Protein Cross-Linking and Southwestern Hybridization. For ultraviolet (UV) protein cross-linking analysis, an oligonucleotide containing nucleotides −260 to −237 (5′-GCCTTGTTGCTCAACATGTTGATT-3′) within the PISII site was annealed with a 10-mer containing nucleotides −237 to −246 (5′-AATCAACATG-3′) and extended by the Klenow fragment in a reaction mixture containing 50 μCi of [α-^{32}P]-dCTP, 0.025 mM dCTP, and 0.25 mM each of dATP, dGTP, and bromo-dUTP (BrdUTP). The BrdU-substituted probe was then incubated with HP-A1 nuclear extracts as described for the EMSA, and cross-linked by UV light essentially as described by McCormick et al. (1991). Briefly, samples in an ice bath were irradiated for 30 min in a Stratalinker (Stratagene) at a distance of 12 cm. The reaction mixture was then separated by 12% SDS–PAGE, dried, and autoradiographed.

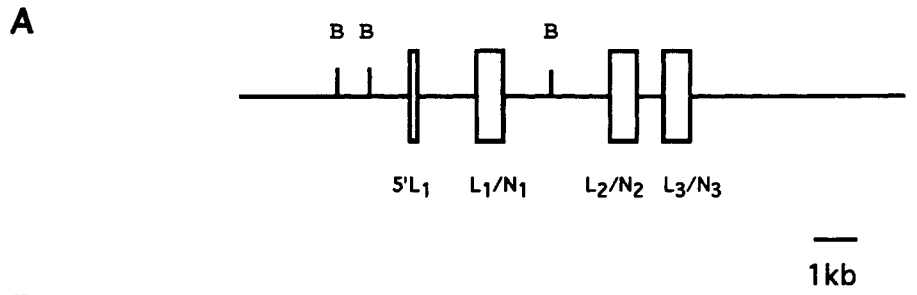
For Southwestern analysis, concatenated double-stranded oligonucleotides containing nucleotides −261 to −236 (5′-AGCCTTGTTGCTCAACATGTTGATTG-3′) in the rnCGM3 gene and an additional *Xba*I linker were labeled by the Klenow fragment. Fifty micrograms of HP-A1 nuclear extracts was separated by 12% SDS–PAGE and blotted onto a nitrocellulose membrane. Proteins were denatured by incubating the filter in binding buffer 10 (BF10) (25 mM Hepes, pH 7.9, 3 mM $MgCl_2$, 10 mM NaCl, and 1 mM DTT) containing 6 M guanidine hydrochloride at 4 °C for 5 min. Proteins were then renatured by a stepwise dilution of guanidine hydrochloride with an equal volume of BF10 until the final concentration of guanidine hydrochloride reached 0.187 M. The filter was blocked with 5% (w/v) nonfat dry milk in BF10 for 30 min at 4 °C and hybridized with 1×10^6 cpm/mL probe for 3 h at 4 °C in binding buffer 100 (BF100) (25 mM Hepes, pH 7.9, 3 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, and 10% glycerol) containing 0.25% (w/v) nonfat dry milk and 10 μg/mL sonicated salmon sperm DNA. The filter was washed in three 100 mL changes of BF100 containing 0.25% nonfat dry milk at 4 °C for up to 1 h, and autoradiographed.

RESULTS

Characterization of cDNA and Genomic Clones Encoding mCGM3. In earlier studies using a rat PSG cDNA, rnCGM6, as a probe, we demonstrated that rat PSG was expressed primarily in the placenta with low levels in the uterus (Chen et al., 1992). To study the regulation of rodent PSG expression, we screened a rat genomic library with the rnCGM6 probe and characterized two overlapping clones, L13 and L103. Restriction endonuclease mapping, Southern hybridization, and DNA sequencing analysis showed that the two clones contained the 5′-flanking region and the first four exons (5′L_1, L_1N_1, L_2N_2, L_3N_3) of a rat PSG gene, rnCGM3, but not the 3′-exon encoding the carboxyl-terminal A domain (Figure 1A). The exon–intron junctions were assigned by sequence comparison and alignment of cDNA and genomic sequences (Table 1).

The existence of the rat rnCGM3 gene was identified only through N-domain sequence analysis of rat PSG genomic clones (Kodelja et al., 1989); thus, the 5′ and 3′ regions of this cDNA have not been characterized. To demonstrate that rnCGM3 is actively transcribed in the rat placenta and to characterize the structure of this cDNA, we performed RT–PCR and RACE analyses. A pair of primers, nucleotides 183–203 and 1082–1060 (Figure 1B), containing rnCGM3 exon sequences were used to amplify DNA fragments from rat placental poly(A)$^+$ RNA by RT–PCR, and a fragment of expected size (900 bp) was obtained. Sequence analysis indicated that this fragment contained the predicted sequence of the rnCGM3 cDNA. To identify the 5′ and 3′ regions of rnCGM3, we amplified terminal fragments from rat placental RNA by the RACE method. The end of the 5′-untranslated region was shown to be located at nucleotide −197 with respect to the translation start site, and the carboxyl A domain was shown to follow the N_3 domain of rnCGM3 (Figure 1B).

The rnCGM3 cDNA is 2761 bp in length and has an open reading frame that encodes a polypeptide of 475 amino acids with a domain arrangement of L_1N_1-L_2N_2-L_3N_3-A (Figure 1B). The deduced rnCGM3 protein contains six potential



B

```
> Exon 5'L1
   GGAGCAGCATTGACAGTGGTAGGGACAGAGCAGTGCCTTGGACACTGACCCCACTATCCACAGCTCATAGGATTT     -123
TGAGAAGTGCTCCTGCCTGGGAAAAAACTCAGCTCTGAAAAAGGAAGGACAGCAGAGGCCAGGAGCCTTGTTGGAGCT     -45
GAAGGTCTTCTCCTCAGAGCAAGGGCAGTATAGTGAGGAGAGAC ATGGAGGTGTCCTCTGTGCTTCTCTGTTCAGGG      33
                                              M  E  V  S  S  V  L  L  C  S  G       11
                                  > Exon L1N1                          * N1
TGTAGCCCTTGGCAGGGGTTCCTGCTCACAGCCTCACTTTTAACCTGCTGGCTCCTGCCCACCACTGCCCAAGTCACC     111
C  S  P  W  Q  G  F  L  L  T  A  S  L  L  T  C  W  L  L  P  T  T  A  Q  V  T       37
CTCAAGTCCTCACCGCCCCAGGTGGTTGAAGGAGAAAACGTTCTTCTAAGTGCTGACAATCTGCCAGAGAACATTATA     189
L  K  S  S  P  P  Q  V  V  E  G  E  N  V  L  L  S  A  D  N  L  P  E  N  I  I       63
GCTTTCGCCTGGTACAAAGGGGAGACCGACATGAACCGTGGAATTGCACTGTATTCACTGAGGTATACTGTAAGTTTG     267
G  F  A  W  Y  K  G  E  T  D  M  N  R  G  I  A  L  Y  S  L  R  Y  T  V  S  L       89
ACGGGGCCTGTGCACAGTGGTCGAGAGACATTGTACAGCGACGGGTCCCTGTGGATCAAAAATGTCACCCAGGAGGAC     345
T  G  P  V  H  S  G  R  E  T  L  Y  S  D  G  S  L  W  I  K  N  V  T  Q  E  D      115
ACAGGATTTTATACCTTTCGAATCATAAATAAACATGGAAAAATTCAATCAAATACAACCCTGTTCCTTCACGTGAAA     423
T  G  F  Y  T  F  R  I  I  N  K  H  G  K  I  Q  S  N  T  T  L  F  L  H  V  K      141
 > Exon L2N2                          * N2
TCCTCTCTTTTCGTCTGTGGGCGCCCTTCTAACATTGACGAGCCCACTATTGAATCAGTGCCGGTCAGCGTTAGCGAA     501
S  S  L  F  V  C  G  R  P  S  N  I  D  E  P  T  I  E  S  V  P  V  S  V  S  E      167
TGGGGAAGCGTTCTTCTCCGTGTTCACAATCTTCCAGAGAATCTACGAACGCTTTTCTGGTACAAAGGGGTGCTTGTG     579
W  G  S  V  L  L  R  V  H  N  L  P  E  N  L  R  T  L  F  W  Y  K  G  V  L  V      193
TTGAAGAAGCTTGAGATTGCCCGATACAGAACAACCAAGAACACATATGAGCTGGGCCCTGTGCATAGTGGTAGGGAA     657
L  K  K  L  E  I  A  R  Y  R  T  T  K  N  T  Y  E  L  G  P  V  H  S  G  R  E      219
ACCGTGTACAGTAATGGGTCCCTGCTGCTCCAGAATGTCACCTGGAAAGACACCGGATTCTACACCCTAAGAACTCTG     735
T  V  Y  S  N  G  S  L  L  L  Q  N  V  T  W  K  D  T  G  F  Y  T  L  R  T  L      245
                                                  > Exon L3N3
AGTACAGATTTGAAAGTAGAAATAACACACATTTACCTTCAGGTGAACACCTCCCTTTCCTCGTGCTGTGACCCTGTC     813
S  T  D  L  K  V  E  I  T  H  I  Y  L  Q  V  N  T  S  L  S  S  C  C  D  P  V      271
         * N3
ACCTCTGTCCCACTCATGATAGAGCCAGTGCCACGGCATGCAGTTGAAGGGGAAAGTGTTCTTCTCTATGTCCACAAT     891
T  S  V  P  L  M  I  E  P  V  P  R  H  A  V  E  G  E  S  V  L  L  Y  V  H  N      297
CTGCCTGAAGCTCTGCAAACCTTTTCCTGGTACAAAGGAGTGTACAGCCTTAAGGAATTTAAAATTGCAGAATACAGC     969
L  P  E  A  L  Q  T  F  S  W  Y  K  G  V  Y  S  L  K  E  F  K  I  A  E  Y  S      323
ATAGCGACAAAGTCTGTCTTCCCGGGCCCGGCACACAGAGGAAGAGCGACAGGTTACACCAACGGATCTCTGCTCCTC    1047
I  A  T  K  S  V  F  P  G  P  A  H  R  G  R  A  T  G  Y  T  N  G  S  L  L  L      349
CAGGACCTCACTGCAAGAGATACTGGCTTGTACACACTAGTAACGTTAGACAGTAATTCCAAAATTAAATCAGCCCCT    1125
Q  D  L  T  A  R  D  T  G  L  Y  T  L  V  T  L  D  S  N  S  K  I  K  S  A  P      375
                  > Exon A
GTGCAAGTCACCGTGCACAAGCCTGTAACACAGCCCTTCCTGAGAGTCACAGAGAGCACAGTGACAGTACAGAGCTCA    1203
V  Q  V  T  V  H  K  P  V  T  Q  P  F  L  R  V  T  E  S  T  V  T  V  Q  S  S      401
GTGGTCTTCACTTGCTTGTCGGACAACACTGGAGTCTCCATCCGTTGGCTCTTCAAGAATCAGAATCTGCAGGTCACA    1281
V  V  F  T  C  L  S  D  N  T  G  V  S  I  R  W  L  F  K  N  Q  N  L  Q  V  T      427
GAGAGGATGACCCTATCCCCATCAAATTGCCAACTCAGGATACATGATGTCAGAAGGGAGGATGCTGGACAGTATCGA    1359
E  R  M  T  L  S  P  S  N  C  Q  L  R  I  H  D  V  R  R  E  D  A  G  Q  Y  R      453
TGTGAGGCCTTCAACCCAATCAGCTCCAAGACCAGTCGCCCAGTCAGCCTGGCTGTGATTAATGAGTGA CCCCTCCT    1436
C  E  A  F  N  P  I  S  S  K  T  S  R  P  V  S  L  A  V  I  N  E  END            475
CTCATCCTACAGCAGAGTGGGGACCTTTCTGTATTGAGATGTCCACACTTATGCAACCCTGCCAGTTACAGATTTCCA    1514
TTCATGCTCATGGCCAACTGTCCTTCTGTCCCATCCCTCCTCACACCTGATCCTGAACCCCCCCCATTATCTTCTACAT    1592
CCCTTCTTCCACCCAGTTCCTTTCCTTCTTCTTTCCCCTCCAACTCTATTTCCCCCTTCTAAGCTGAGATTCAAGCAA    1670
CCTCACGGGGCCTTTCTTCTTGTTTACCTTCTTTGTGTCAGTGGATTGTAGTGTGTGTATCTTGTATTTCATGGCTAA    1748
TATGCACTTATAAGTGCGACATGTCCTGGATGCCGGCTTTGGGGTTGGGCTACCTCAGCAAGTTCCATCCATGTGCCT    1826
TCAAAATTAATGATGTCTGTCTCTTCAATTGCTGAGTAGTAGCCCATGGTGTATATGAACCACATTGTCTGTATCCAT    1904
TCTTTCTAGTTACTGGTGCCCTCCTGGCAGTGCTCAGCCCACTATTGATGAGCAATGTAGCCCAAAGGTAAATGTTCA    1982
ATTTGTTCTTAGTGGACAGAGAGTTTCAAGCCTATATTGACATAAAGACGGGAACTGAGGTTGACAGCTGTCTTAGGG    2060
TTTTACTGCTGTGAACAGACTCCATGACCTATGCAAGTCTTATAAAGGACAACATTTATTGTGGCTGGCTTACAGCTC    2138
AGAGATTCAGTCCATTTTCATCAAGGAGGGAGCATGGAAGAGTCCATGCAAGCACAGTACAGCAGATTTGATTGTTCT    2216
ACAGCTTCATCTGAAGGATGCTAGCAGAATACCCATTTCCAGGCAGCTAGGGTTAGGGTATTAAAGCCCACACCCACA    2294
GTGACACACCTCCTCCAACAAGACTACACCTCCAACAGTGCAATTGGCCCAAGAATATACAAACCATCACAACAGCCA    2372
TGGCCTCACAGTTTGTAATAGCCACCAATCAAAACTTGCTGTTGTCTGAACATAGTGTCCAAGGCTCAGAACATGCCA    2450
ATGGATCACTGCTCCTTCCGAAAACCATCCAGATGAAGACAATGTTTTGACCCTTCAAATCTGAAATGCTGAAGTAGA    2528
ATGTGAATAAACATATGTGTATTACCACATAAACAG (A) n                                        2564
```

FIGURE 1: (A) Structure organization of rnCGM3. Only the first four exons of rnCGM3 are shown here. B, *Bam*HI. (B) Nucleotide and deduced amino acid sequences of rnCGM3 cDNA. Arrows indicate the locations of exons, and asterisks indicate the N domains. The potential N-linked glycosylation sites are underlined. The 3'-UTR destabilizing sequence is boxed, and the signal for polyadenylation is in boldface type.

N-linked glycosylation sites, suggesting that it is a glycoprotein. The sequences contained in the 5'-end exons (5'L_1) of rnCGM3 and rnCGM6 (Chen et al., 1992) are identical. In the coding region, rnCGM3 and rnCGM6 share 88% amino acid sequence identity except in the N_2 domain where they share 79% identity, suggesting that these two rat PSGs are closely related. On the other hand, amino acid sequences of rnCGM3 and rnCGM6 share only 52–69% sequence identity (depending upon the domains compared) with another rat PSG, rnCGM1 (Rebstock et al., 1990; Rudert et al., 1992).

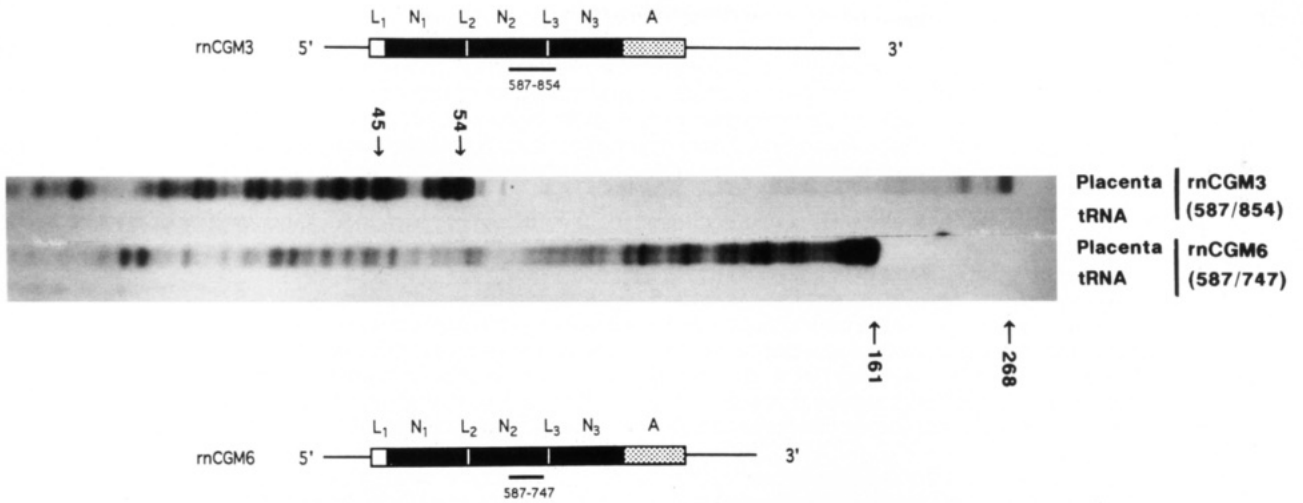


FIGURE 2: Ribonuclease protection analysis of rnCGM3 and rnCGM6 mRNAs in the rat placenta. The rnCGM3 probe contains nucleotides 587–854, and the rnCGM6 probe contains nucleotides 587–747, each indicated by a bar under the respective cDNA. Total RNA (25 μg) from rat placenta was annealed with the respective antisense riboprobe, and the protected fragments were separated on an 8% polyacrylamide–urea gel. Arrows indicate the fully protected fragments and the clusters of PSG mRNAs sharing sequence homology to rnCGM3. *ϕX174* digested with *Hin*fI was used as a marker (data not shown).

rnCGM3 Is Expressed in the Rat Placenta. PSGs are encoded by a group of closely related genes sharing high sequence homology. To demonstrate that the rnCGM3 transcript is expressed in the rat placenta, we examined rnCGM3 as well as rnCGM6 (Chen et al., 1992) mRNA expression by ribonuclease protection assays. An antisense riboprobe specific for rnCGM3 (nucleotides 587–854 in the N_2 domain) or rnCGM6 (nucleotides 587–747 in the N_2 domain) was hybridized with the rat placental RNA, and ribonuclease-protected fragments were analyzed on a sequencing gel (Figure 2). Fully protected bands of 268 and 161 bases predicted for the rnCGM3 and rnCGM6 mRNAs, respectively, were obtained. The clusters of protected fragments ranging from 40 to 55 bases using the rnCGM3 probe may represent PSG transcripts highly related to rnCGM3. Although it is difficult to quantitate the relative abundance of the mRNAs using different probes, the greater intensity of the rnCGM6-protected bands compared with the rnCGM3-protected bands suggests that the predominantly expressed PSG mRNA in rat placenta is rnCGM6, in agreement with previous observations (Chen et al., 1992). The ATTTA motif known to destabilize mRNAs (Sachs, 1993) is located at nucleotides 2114–2118 in the 3′-untranslated region of rnCGM3 (Figure 1B). Such a motif is not detected in the rnCGM6 mRNA.

Identification of Cis-Acting Elements in the rnCGM3 5′-Flanking Region Essential for PSG Expression. To delineate DNA sequences essential for expression of the rnCGM3 gene, we sequenced 1465 bp of the 5′-flanking region (Figure 3A). A putative TATA box, CATAAA, is located −226 to −221 bp upstream of the translation start size (Figure 3A). A TATA box-like sequence has not been observed in human PSG genes identified to date (Lei et al., 1992b, 1993). Nucleotides −1188 to −777 in the rnCGM3 5′-flanking region contain a portion of a long interspersed repeated DNA element (LINE) (Figure 3A), found in other rodent genes including oxytocin (Schmitz et al., 1991), vasopressin (Schmitz et al., 1991), hepatic steroid hydroxylase (Matsunaga et al., 1990), and salivary proline-rich protein (Lin & Ann, 1991). Nucleotides −176 to −29 in the rnCGM3 5′-flanking region share high sequence identity with nucleotides −176 to −42 in the human PSG gene 5′-flanking regions (Lei et al., 1993) (Figure 3B). To define DNA sequences that direct rnCGM3 gene expression in placental cells, we created rnCGM3 promoter–CAT fusion genes in a modified pCAT-Basic vector (pCAT-Basic-N) (Lei et al., 1993) and analyzed the promoter activity after transient expression in HP-A1 cells. HP-A1 cells are human placental cells immortalized with a recombinant SV40 virus which support expression of human PSG promoter–CAT fusion genes (Lei et al., 1992a). pCAT-Basic-N (containing no promoter or enhancer) and pSVCAT (containing the SV40 promoter and enhancer) were used as negative and positive controls, respectively.

Low levels of CAT activity (0.9–2% of pSVCAT) were observed in HP-A1 cells transfected with rnCGM3 promoter–CAT fusion genes containing nucleotides −1465/−33 to −412/−33 (Figure 4A). An additional 5′ deletion of nucleotides −411 to −284 abolished CAT expression, and further deletion of nucleotides −283 to −193 failed to restore CAT activity (Figure 4A). However, CAT activity rose to 8.9% [rnCGM3-(−147/−33)CAT] with a further 5′ deletion of nucleotides −192 to −148. The minimal downstream promoter (PII) was located at nucleotides −147 to −86 (Figure 3A) with respect to the translation start site because deletion of nucleotides −147 to −97 abolished CAT expression and deletion of nucleotides −85 to −33 reduced CAT expression to 1.2% (Figure 4A). Our study also suggests that nucleotides −85 to −33 contain activator elements and nucleotides −192 to −148 contain negative regulatory elements. The downstream promoter, PII, is located in the first exon which starts at nucleotide −197 with respect to the translation start site.

The presence of a TATA box at nucleotides −226 to −221 and the low levels of CAT expression directed by the rnCGM3-(−412/−33)CAT plasmid suggest that an additional promoter may exist upstream of the transcription initiation site. We therefore created rnCGM3 promoter–CAT fusion genes containing sequences upstream of nucleotide −185 (Figure 4A). The downstream promoter, at nucleotides −147 to −33, was deleted in these constructs. Transient expression in HP-A1 cells showed that deletions of nucleotides −1465 to −413 increased CAT expression to 2% [rnCGM3(−412/−185)CAT] of pSVCAT and nucleotides −610 to −413 constitute a negative regulatory element of the upstream promoter, PI (Figure 4A). Deletion of nucleotides −412 to −327 [rnCGM3(−326/−185)-CAT] decreased CAT expression to 27% of rnCGM3(−412/−185)CAT, and a further deletion of nucleotides −326 to −284 abolished CAT expression, locating PI at nucleotides −326 to −185 (Figure 4B).

A

```
GGATCCAGAGAAGGTCACAGTTGACTTGTTGATCTCCTCTGCTGTAGCCCTAGACATGAGATTGCAGTTCTC    -1394
CCAGAGAACTCCAGGGGGCGATACCAGGATGCTCACTCAACACTCAGACAATTGTTTCCCTGAATCATGACC    -1322
TTGACTAGGAGACAGCTTTCCTTCCTCTTCCTACTAACCTGGAGGTCCCTCACTAGACTCTGTTTCTTGTGT    -1250
CTGAGAATACCCCTCTCTGTGGGAACCGGAAAGTCCCCTCCCTCCTGGCATCTCCCATAGGGAAGAGAGGAT    -1178
TTCTTTTTTTTTTTTTTTCCGTTTTTTCTTTATTAACTTGAGTATTTCTTATTTACATTTCAAATGTTATTC    -1106
CCTTCCGGTTTCCGGGCCAACATCCCCCTAACCCCTCCAGCTCCCCTTCTATATGGGTGTTCCCCTCCCCAT    -1034
CCTCCCCCCCATTACTGCCCTCCCCGCAACAATCGCGTTCACTGGGGTTCAGTCTTAGCAGGACCAAGGGCTT     -962
CCCCTTCCACTGGTGCTCTTACTAGGCTATTCATTGCTACCTATGAGGTTGGAGTCCAGGGTCAGTCCATGT     -890
TTAGGCTTTGGGTAGTGGCTTAGTCCCTGGAAGCTCTGGTTGGTTGGCATTGTTGTTCATATGGGGTCTCGA     -818
GCCCCTTCAAGCTCTTCCAGTCCTTTGAAGAGAGGATTTCTGATGCGGACCCAGAGCAAGGGCAGAGCTGAC     -746
TCAGGAAGATGCCAGAGAACTTGGGCTCAGCCAACAAACAGGGAAGCCATCCCATACTGCTGCAAGAATGCC     -674
ATGAAAGTGTCCGTAGGAAGAAGAGCCCACAGTCCTTTTTAGGACAAACAGGTCCCCTCCTTTCCAAGACCA     -602
AGAACATAAAACTCCTCACACAGCTGCTCTTCAGCCTCTCTTGGGTCAGGAACTGGACTGGTGACATTAGTG     -530
ATAAGGATTAATCTTCATTCACAGGCATCACTCGGTTTGTCCTTTTCAGCTCTCAGAGAAATCTGTGTCACC     -458
TTCTGCTGGGGAATGCCACCTTCCCGGTTCAGTGAGGACATAGGGCAGACTGGGTGCTCATCTGGCTCTCTG     -386
                                                          PI (-326/-185)
TGTCACAGGTCCACATAGGAAAGATGGTGGCTCCCTCTTTGATGCTGGCAGAGTCACGTCCGGGACACAGCC     -314
GATGTCATCCCTGGGATGAGTGTTGTTCTCTGAGGGCATGTTGATGCTTATCAGCCTTGTTGCTCAACATGT     -242
                                                  ┌─> Exon 5'L1
TGATTGCAGGAGAAACATAAAGAGAGAAAAGTTGAAAGGGATGTGGAGCAGCATTGACAGTGGTAGGGACAG     -170
                                                  PII (-147/-86)
AGCAGTGCCTTGGACACTGACCCCACTATCCACAGCTCATAGGATTTTGAGAAGTGCTCCTGCCTGGGAAAA      -98
AACTCAGCTCTGAAAAAGGAAGGACAGCAGAGGCCAGGAGCCTTGTTGGAGCTGAAGGTCTTCTCCTCAGAG      -26
CAAGGGCAGTATAGTGAGGAGAGACATG                                                    3
```

B

```
rnCGM3   -176   GGGACAGAGCAG-TG-CCTTGGACACTGACCCCACTATCCACAGCTCATAGGATTTT
PSG12    -174   GGGACAGAG-AGGTGTCCTGGG-C-CTAACCCCAC---CCA----T--GAGCC---T

rnCGM3          GAGAAGTGCTCCTGCCTGGGAAAAAA-CTCAGCTCTGAAAAAGGAAGGACAGCAGAG
PSG12           GAGAAGTGCTCCTGCCCCAGGAAGAGGCTCAGCGCAGAAGGAGGAAGGACAGCACAG

rnCGM3          GCC-AGGAGCCTTGTTGGAG-CTGAAGGTCTTCT----CCTCA     -29
PSG12           -CCTA-CAGCCGTGCTC-AGG---AAG--TTTCTGGAACCT-A     -42
```

FIGURE 3: (A) Nucleotide sequence of the 5′-flanking region of rnCGM3. The partial sequence of a long interspersed repeated DNA element (LINE) is underlined. Direct repeated sequences flanking the LINE are bolded, and the putative TATA box is bolded and underlined. The PI (–326/–185) and PII (–147/–86) promoter elements are shaded. (B) Sequence comparison between the rnCGM3 PII promoter and the minimal promoter of the human PSG12 gene (Lei et al., 1993). Nucleotides that are identical between rnCGM3 and PSG12 are highlighted. Gaps are introduced for optimal alignment.

Effects of Sodium Butyrate on the Expression of rnCGM3 Promoter Constructs. In human PSG genes, we have shown that sodium butyrate greatly stimulated expression of PSG promoter–CAT fusion genes (Lei et al., 1993). Therefore, we examined the effects of sodium butyrate on rodent PSG (rnCGM3) expression. Sodium butyrate stimulated CAT expression in constructs containing primarily the upstream promoter, PI, and abolished the apparent negative regulation by the putative repressor at nucleotides –1465 to –413 (Table 2). In contrast, butyrate had no effect on expression of the downstream promoter, PII, and could not antagonize the negative regulation of nucleotides –192 to –148 on the rnCGM3 PII promoter (Table 2).

Identification of Protein Binding Sites in the Downstream Promoter, PII. Transient expression assays have located two promoter elements on each site of the transcription start site of the rat rnCGM3 gene. The downstream promoter, PII, is stronger than the upstream promoter, PI, *in vitro* (Figure 4A). To characterize protein binding sites within the PII promoter, EMSAs were performed using HP-A1 whole cell extracts and four double-stranded oligos corresponding to nucleotides –147 to –33, –147 to –86, –96 to –33, and –108 to –88 (Figure 5). Specific protein–DNA interactions were identified by the ability to block complex formation by the homologous oligos but not by unrelated ones. Three protein–DNA complexes (PIICI, PIICII, and PIICIII) were detected with rnCGM3(–147/–33) (Figure 5A) and rnCGM3(–147/–86) (Figure 5B), two (PIICI and PIICII) with rnCGM3(–96/–33) (Figure 5C), and one (PICIII) with rnCGM3(–108/–88) (Figure 5D).

When rnCGM3(–147/–33) (Figure 5A) or rnCGM3(–147/–86) (Figure 5B) was the target DNA, the formation of PIICI and PIICII was efficiently blocked by an excess of unlabeled target DNA, rnCGM3(–147/–86), or rnCGM3(–96/–33) DNA, indicating that transcription factor(s) that bind to nucleotides –147 to –86 or –96 to –33 in the PII promoter are closely related. On the other hand, rnCGM3(–108/–88), rnCGM3(–91/–72), or an unrelated Oct1 fragment failed to affect the formation of complexes PIICI and PIICII, locating the two putative binding sites at nucleotides –147 to –109 and –71 to –33. Similarly, the two protein–DNA complexes (PIICI and PIICII) formed between rnCGM3(–96/–33) and placental cell extracts could be efficiently blocked by an excess of unlabeled target DNA or rnCGM3(–147/–86), but not by rnCGM3(–91/–72) or Oct1 (Figure 5C). Sequence analysis indicates that multiple AGNAN motifs, at nucleotides –129 to –125, –120 to –116, –72 to –68, and –62 to –58 in the coding strand and at nucleotides –109 to –113 and –38 to –34 in the noncoding strand, are present within nucleotides –147 to –109 and –71 to –33 regions.

The PIICIII formed between rnCGM3(–147/–33) or rnCGM3(–147/–86) and placental cell extracts was efficiently competed only by the rnCGM3(–108/–88) DNA (Figure 5A,B, lanes 4). We do not know why formation of the PIICIII complex was not competed by the respective target DNA

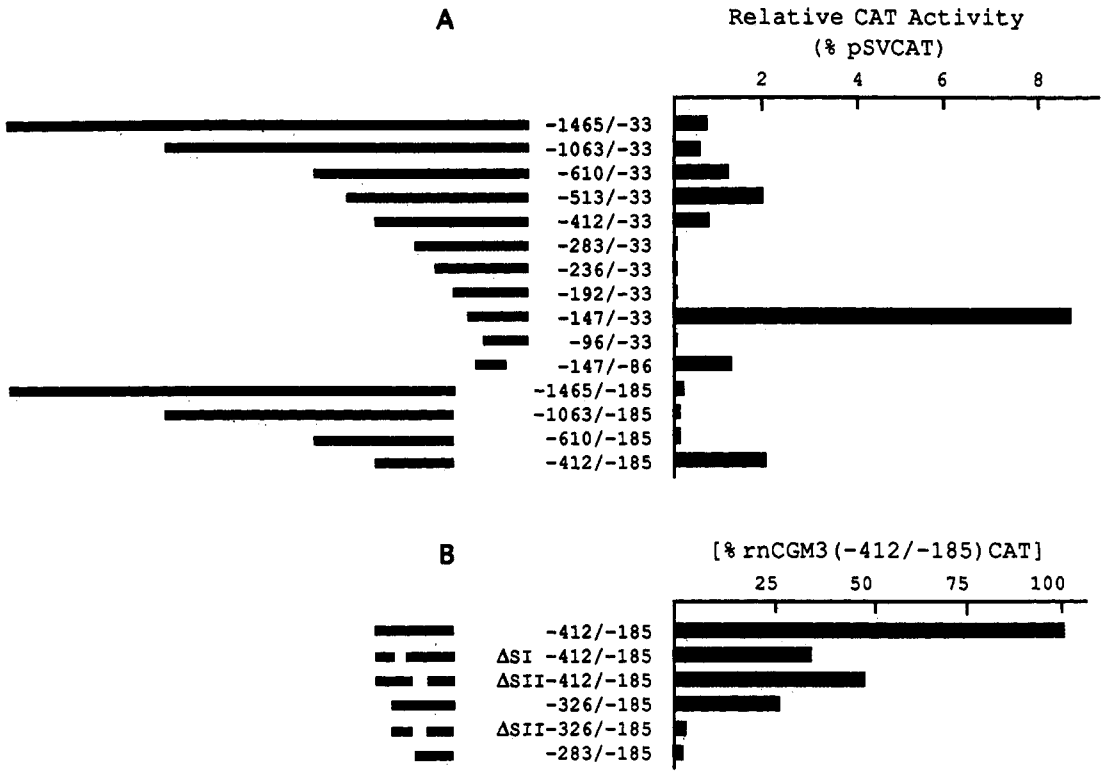


FIGURE 4: Bipartite promoter elements in the rnCGM3 5′-flanking region. (A and B) Characterization of promoter elements in rnCGM3. Fragments encompassing the respective rnCGM3 5′-flanking regions were inserted upstream of the modified promoterless and enhancerless pCAT-Basic-N plasmid (Lei et al., 1993). The pSVCAT (pCAT-Control) plasmid which contains both SV40 enhancer and promoter and pCAT-Basic-N were used as positive and negative controls, respectively. The CAT activity was determined for each construct after transient transfection into HP-A1 placental cells. The amount of CAT activity expressed after transfecting an equivalent amount of pSVCAT was 5.5 and 5.8 nmol min^{-1} (mg of protein)$^{-1}$ in (A) and (B), respectively. At least three independent experiments were conducted with two preparations of each construct. Two promoter elements at nucleotides −326/−185 (PI) and at nucleotides−147/−86 (PII) were identified. ΔSI−412/−185 stands for an internal deletion of nucleotides −311 to −284, and ΔSII−412/−185 ΔSII−326/−185 stand for an internal deletion of nucleotides −283 to −237.

Table 2: Effects of Sodium Butyrate on the Expression of rnCGM3 Promoter–CAT Fusion Genes

rnCGM3 constructs	rel CAT act.[a] (% of pSVCAT)	
	control	butyrate
−1465/−33	1.1	6.1 (5.5)
−1063/−33	1.2	4.3 (3.7)
−610/−33	1.4	4.8 (3.5)
−513/−33	2.1	4.9 (2.4)
−412/−33	0.9	3.7 (4.3)
−283/−33	ND[b]	ND
−192/−33	ND	ND
−147/−33	14.3	15.8 (1.1)
−1465/−185	ND	1.2
−1063/−185	ND	1.0
−412/−185	1.7	3.4 (2.0)
−283/−185	ND	ND

[a] The CAT activity was determined for each construct after transient transfection into HP-A1 placental cells in the absence or presence of 1 mM sodium butyrate. The amount of CAT activity expressed after transfecting an equivalent amount of pSVCAT into control or butyrate-treated cells was 2.3 and 4.5 nmol min^{-1} (mg of protein)$^{-1}$, respectively. Numbers in parentheses are butyrate to control ratios. [b] ND, nondetectable.

(Figure 5A,B, lanes 3). The specificity of PIICIII was further demonstrated when rnCGM3(−108/−88) was the target DNA (Figure 5D). The formation of PIICIII was blocked by an excess of rnCGM3(−108/−88), but not by rnCGM3(−91/−72) or Oct1 (Figure 5D).

To dissect further protein binding sites within the PII promoter of the rnCGM3 gene, we performed DNase I footprinting analysis using HP-A1 nuclear extracts and labeled rnCGM3(−147/−33) (Figure 6A) and rnCGM3(−283/−33) (Figure 6B) fragments. A discrete region corresponding to nucleotides −108 to −93 (PIISIII) was reproducibly protected from DNase I digestion both in the coding and in the noncoding strand by HP-A1 nuclear extracts using either rnCGM3(−147/−33) or rnCGM3(−283/−33) probe.

Identification of Protein Binding Sites in the Upstream Promoter, PI. To demonstrate that HP-A1 cells contain transcription factors that bind to sequences within the rnCGM3 upstream promoter, PI, we performed DNase I footprinting assays using two labeled fragments, rnCGM3(−283/−33) (Figure 6B) and rnCGM3(−412/−185) (Figure 6C). The rnCGM3(−283/−33) probe will also detect protein binding sites within PII at nucleotides −147 to −33. Two protected regions, PISII (−257/−239) and PIISIII (−108/−93), were found within rnCGM3(−283/−33), and two protected regions, PISI (−311/−290) and PISII (−257/−239), were found within rnCGM3(−412/−185). PIISIII has also been detected using the rnCGM3(−147/−33) DNA (Figure 6A).

The PISI site contains a palindromic sequence, *TCATCCCTGGGATGA*, located at nucleotides −309 to −295, and the PISII site contains two overlapping palindromic elements, *TGTTGCTCAACA*TGTTGA, located at nucleotides −256 to −239 (Figure 6D). To characterize individual protein binding sites within the rnCGM3 PI promoter, EMSAs were performed using HP-A1 nuclear extracts and double-stranded oligonucleotides corresponding to binding sites PISI (nucleotides −326 to −284) and PISII (nucleotides −283 to −237 and −260 to −237). A protein–DNA complex, PICI, was demonstrated with rnCGM3(−326/−284) (Figure 7A) and a complex, PICII, with rnCGM3(−283/−237) (Figure 7B) or rnCGM3(−260/−237) (Figure 7C).

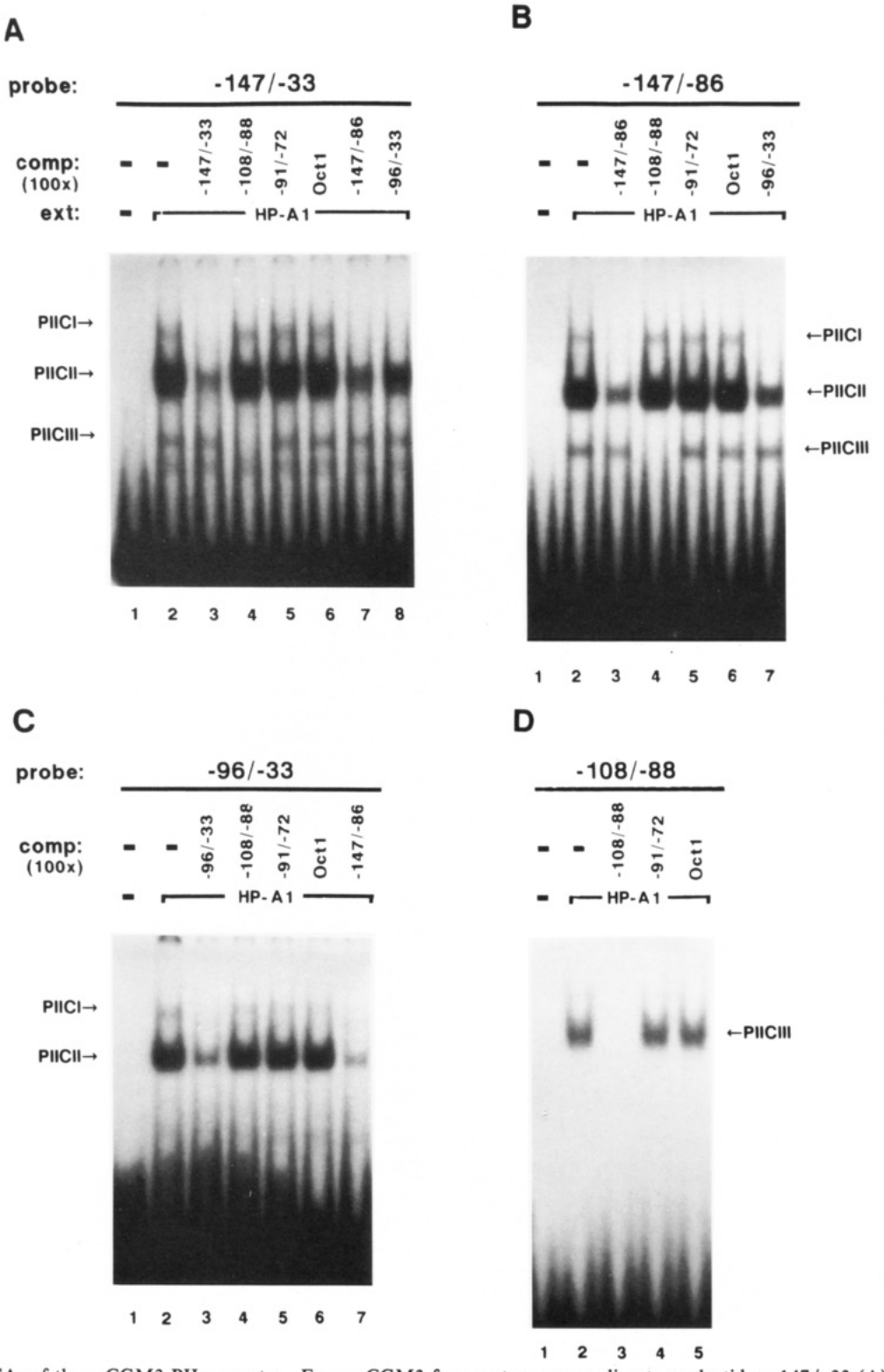
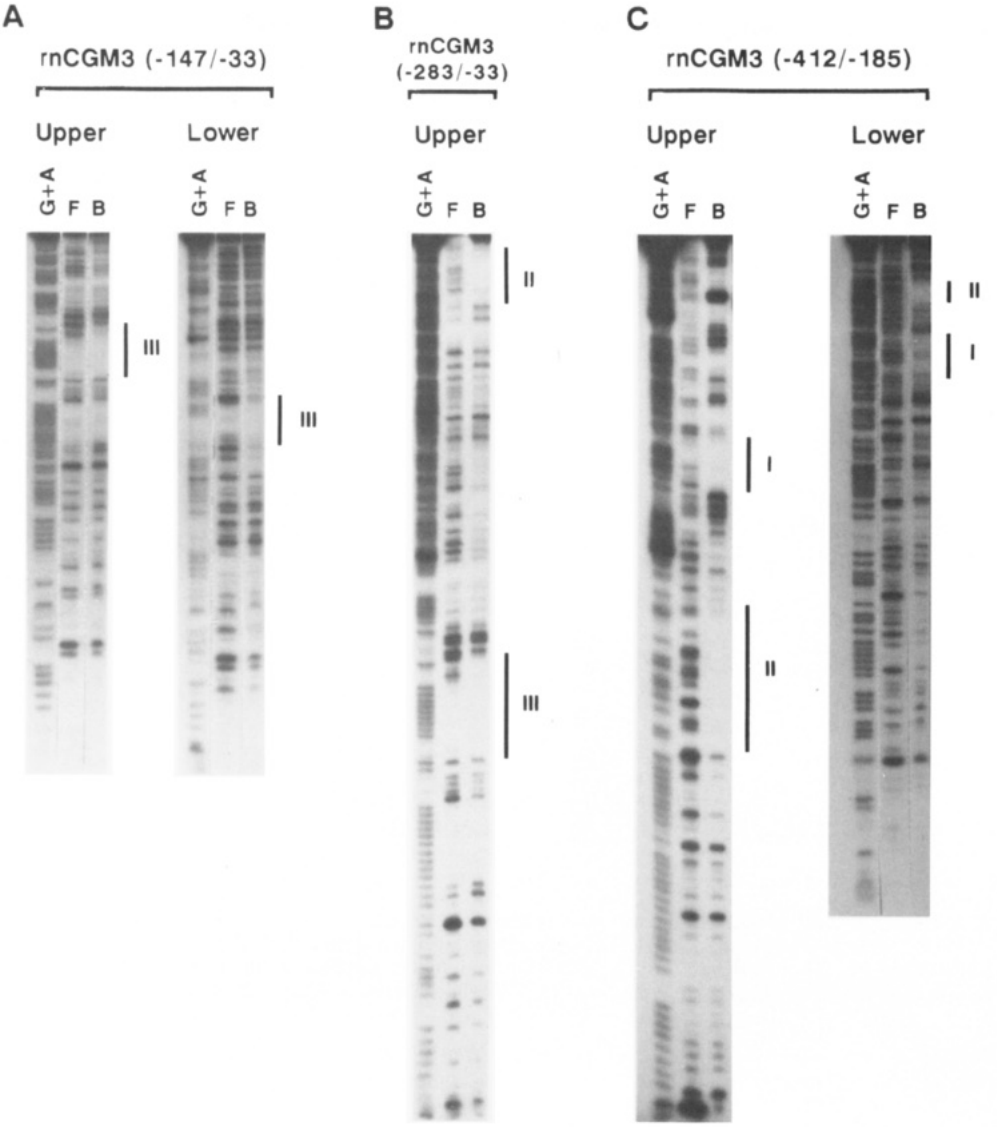


FIGURE 5: EMSAs of the rnCGM3 PII promoter. Four rnCGM3 fragments corresponding to nucleotides –147/–33 (A), –147/–86 (B), –96/–33 (C), and –108/–88 (D) were end-labeled and used in EMSAs with 10 μg of HP-A1 whole cell extracts. Specific complexes (PIICI, PIICII, and PIICIII) are indicated with arrows. Competitor oligonucleotides are present at 100-fold excess over the amount of target DNA. The reaction mixtures were analyzed on 5% nondenaturing polyacrylamide gels.

Functional Analysis of Protein Binding Sites within the rnCGM3 Upstream Promoter. To determine the role of protein binding sites PISI and PISII in transcription enhancement, we constructed mutants that carry internal deletions of either PISI or PISII. rnCGM3(–412/–185ΔSI)CAT contains an internal deletion of nucleotides –311 to –284, and rnCGM3-(–412/–185ΔSII)CAT and rnCGM3(–326/–185ΔSII)CAT contain an internal deletion of –283 to –237. The effect of deletion in these regions on enhancement of CAT expression directed by the rnCGM3 upstream promoter, PI, was then examined in HP-A1 cells (Figure 4B). Deletion of either PISI [rnCGM3(–412/–185ΔSI)CAT] or PISII [rnCGM3-(–412/–185ΔSII)CAT and rnCGM3(–326/–185ΔSII)CAT] resulted in a marked reduction in CAT expression (27–49%) as compared with the rnCGM3(–412/–185)CAT plasmid (100%). The near-complete of CAT expression in the



D

```
-326  CCGGGACACAGCCGATGTCATCCCTGGGATGAGTGTTGTTCTCTGAGGGCATGTTGATGCTTATCAGC
                                  I
-258  CTTGTTGCTCAACATGTTGATTGCAGGAGAAACATAAAGAGAGAAAAGTTGAAAGGGATGTGGAGCAG
           II
-190  CATTGACAGTGGTAGGGACAGAGCAGTGCCTTGGACACTGACCCCACTATCCACAGCTCATAGGATTT
                                                      AP-3
-121  TGAGAAGTGCTCCTGCCTGGGAAAAAACTCAGCTCTGAAAAAGGAAGGACAGCAGAGGCCAGGAGCCT
                H-APF-1    III              PEA3
 -54  TGTTGGAGCTGAAGGTCTTCTCCTCAGAGCAAGGGCAGTATAGTGAGGAGAGACATG
```

FIGURE 6: DNase I footprinting analysis of nuclear protein binding sites in the rnCGM3 5′-flanking region. (A–C) Three overlapping fragments of rnCGM3 5′-flanking region (flanked by 5′-*Hin*dIII and 3′-*Xba*I sites) were asymmetrically labeled at the *Xba*I site for analysis of the upper strand and at the *Hin*dIII site for analysis of the lower strand. Labeled fragments were incubated with HP-A1 nuclear extract, digested with DNase I, and analyzed on 8% polyacrylamide–urea gels. G + A, G plus A sequencing reaction of the probe; F, free probe; B, bound or protected probe. (D) Nucleotides –326 to +3 of the rnCGM3 5′-flanking region. The DNase I protected regions and the putative transcription factor binding sites are underlined. Both regions I and II contain palindromic sequences. Region III is a GA-rich element located downstream of the transcription initiation site at nucleotide –197.

rnCGM3(–326/–185ΔSII)CAT plasmid as compared with the rnCGM3(–326/–185)CAT suggests that the functional activation of rnCGM3 expression by PISI requires the simultaneous presence of PISII.

Characterization of the Protein Factors Binding to PISII. The *TGTTG*CT*CAACA* motif in PISII is also present in the 5′-flanking regions of the prolactin gene (*TGTTGCAACA*) (Peers et al., 1990) and chorionic somatomammotropin and placental growth hormone variant genes (*TGTTG*GTTGC-*CAACA*, termed PSF-A/PSF-B) (Nachtigal et al., 1993). Whereas the motif in the prolactin gene has not been characterized, the PSF-A/PSF-B motif has been shown to function as a repressor in pituitary but not placental cells. Our study indicates that PISII functions as an activator in placental cells. To characterize the protein factor that binds to the PISII motif, we performed protein UV cross-linking experiments using a BrdU-substituted probe, corresponding to nucleotides –260 to –237 in the rnCGM3 PI promoter. Two protein–DNA complexes of apparent molecular masses of 48 and 60 kDa were observed (Figure 8A). Complex

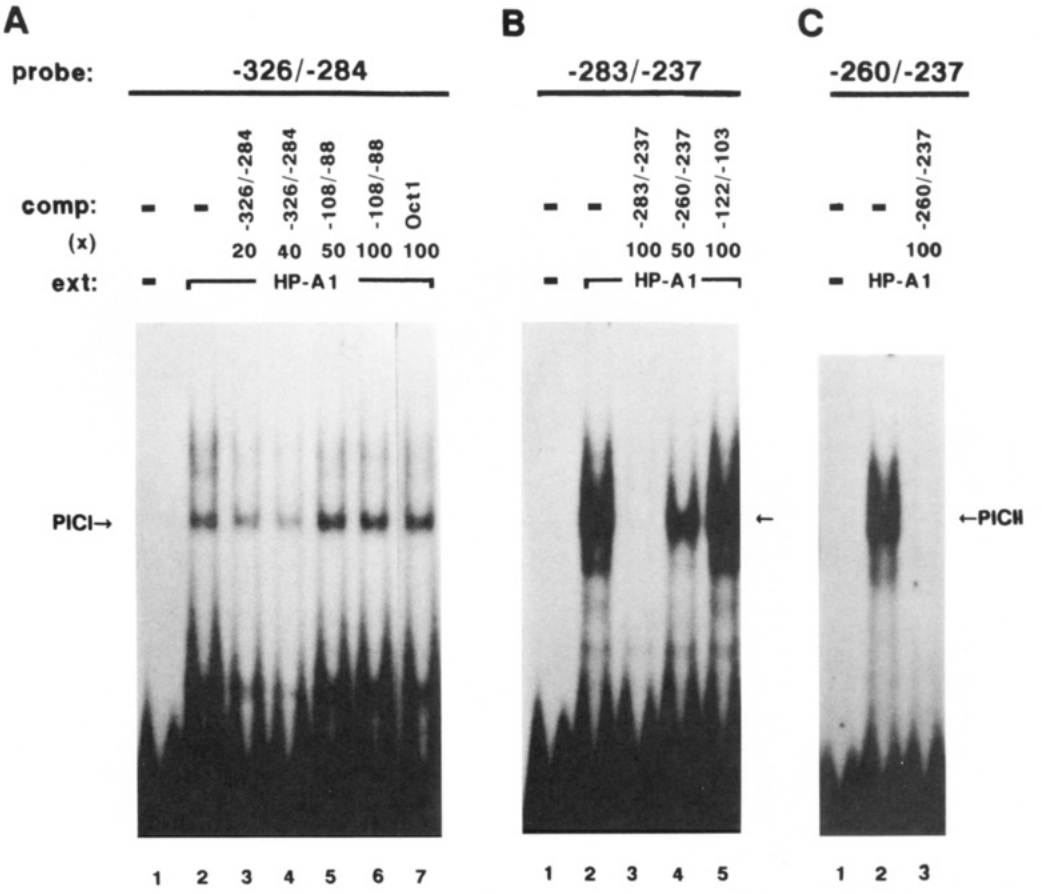


FIGURE 7: Binding of HP-A1 nuclear proteins to distinct sites in the rnCGM3 PI promoter. Three rnCGM3 fragments corresponding to nucleotides –326/–284 (A), –283/–237 (B), and –260/–237 (C) were end-labeled and used in EMSAs with 10 μg of HP-A1 nuclear extracts. Specific complexes (PICI and PICII) are indicated with arrows. The amounts of competitor oligonucleotides are indicated as fold over the amount of target DNA. The reaction mixtures were analyzed on 5% nondenaturing polyacrylamide gels.

formation could be blocked by an excess of unlabeled rnCGM3-(–260/–237) DNA, indicating the specificity of these complexes.

The nature of the protein factor that binds to PISII was further characterized by Southwestern analysis using a concatenated PISII probe. A placental protein factor of apparent molecular mass of 40 kDa was shown to bind to the PISII probe (Figure 8B). The difference in apparent sizes between the two assays may be due to the association of the PISII probe to the protein factor in the cross-linking analysis.

DISCUSSION

PSG genes are a group of closely related genes expressed in high levels only in the placenta. To study trophoblast-specific gene expression in an animal model, we isolated and characterized cDNA and genomic clones encoding a rat PSG, rnCGM3. rnCGM3 and a previously identified rat PSG, rnCGM6 (Chen et al., 1992), contain identical 5′-untranslated regions and L_1 domains. Both exhibit a L_1N_1-L_2N_2-L_3N_3-A domain arrangement. Moreover, the coding regions of rnCGM3 and rnCGM6 share 88% amino acid sequence identity. In contrast, rnCGM3 and rnCGM6 share only 52–69% amino acid sequence identity (depending upon the domains compared) with another rat PSG cDNA, rnCGM1 (Kodelja et al., 1989; Rebstock et al., 1990). Therefore, the rat PSG family appears to be more complicated than was originally anticipated. Ribonuclease protection analysis suggests that rat placenta expresses higher levels of rnCGM6 than rnCGM3. Sequence analysis of both cDNAs indicated that a mRNA destabilizing motif, AUUUA (Sachs, 1993), was present only in the 3′-untranslated region of the rnCGM3 transcript. It is possible that rodent PSG expression is regulated at both transcriptional and posttranscriptional levels.

Transient expression assays using rnCGM3 promoter–CAT fusion genes in a human placental cell line have identified two promoter elements, PI and PII, in the rnCGM3 5′-flanking region. The downstream promoter, PII, is located at nucleotides –147 to –86 (Figure 3A) within the first exon (nucleotides –197 to +64). Sequence analysis indicates that PII shares extensive sequence identity with the minimal promoters identified in the human PSG genes at nucleotides –172 to –34 upstream of the translation start site (Figure 3B) (Lei et al., 1993). The transcription initiation sites in human PSG genes are clustered at nucleotides –106 to –104 (Lei et al., 1993). Studies have shown that the promoter activity of a variety of genes is controlled by sequences located in the first exon. These include the skeletal troponin I gene, the *Drosophila melanogaster* heat shock genes, c-*myc*, and the gastrin gene (Hulmark et al., 1986; Yang et al., 1986; Theill et al., 1987; Nikovits et al., 1990). In addition, sequence comparison revealed potential transcription factor recognition sites in the PII promoter, including AP-3 (Mercurio & Karin, 1989), PEA3 (Rørth et al., 1990), and H-APF-1 (Majello et al., 1990) (Figure 6B). Whether these transcription factors interact with the downstream promoter, PII, and affect rnCGM3 expression *in vivo* is currently under investigation.

The upstream promoter, PI, at nucleotides –326 to –185 (Figure 3A), is located 5′ of the transcription initiation site at nucleotide –197. Several lines of evidence suggest that the upstream promoter, PI, may be the *in vivo* promoter. First,

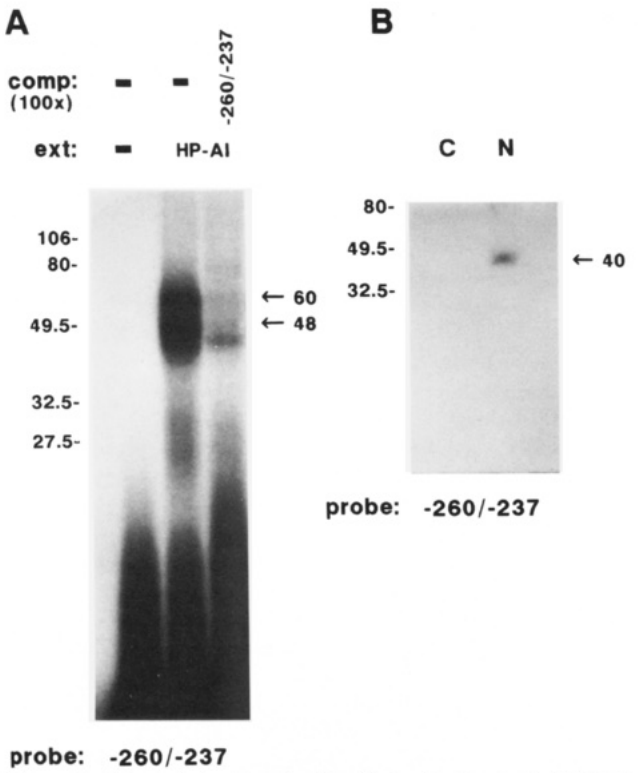


FIGURE 8: Characterization of nuclear factor binding to the TGTTG element in the PICII palindrome. (A) Protein UV cross-linking analysis. A BrdU-substituted, ^{32}P-labeled rnCGM3(–260/–237) DNA was incubated with 10 μg of HP-A1 nuclear extracts in the absence or presence of unlabeled rnCGM3(–260/–237) DNA. The reaction mixtures were irradiated with 254-nm UV light and separated by 12% SDS–PAGE. DNA–protein complexes are indicated by arrows. (B) Southwestern hybridization analysis of protein factor binding to rnCGM3(–260/–237). 50 μg of cytoplasmic (C) or nuclear (N) extract was subjected to electrophoresis on a 12% SDS–polyacrylamide gel and was transferred to a nitrocellulose membrane. The filter was incubated with ^{32}P-labeled concatenated rnCGM3-(–260/–237) as described under Materials and Methods.

a TATA box is found at nucleotides –226 to –221 upstream of the transcription initiation site. Second, expression of the PI promoter (nucleotides –326/–185) is stimulated by sodium butyrate, in contrast to the downstream promoter, PII (nucleotides –147 to –33) which is not responsive to this fatty acid (Table 2). Moreover, butyrate could not antagonize the negative regulation of PII by the putative repressor at nucleotides –192 to –148. It has been demonstrated that butyrate stimulates the expression of human PSG genes (Chou et al., 1990) and antagonizes negative regulation of DNA elements in the human PSG promoters (Lei et al., 1993). Third, the rnCGM3(–283/–33)CAT and rnCGM3(–192/–33)CAT constructs expressed undetectable CAT activity both in the absence and in the prsence of butyrate, suggesting that PII may not be functioning *in vivo*.

Two motifs, PISI and PISII, which bound protein factors in placental cell extracts, have been shown to be important for expression of the rnCGM3 PI promoter. The PISII site contains two overlapping palindromes, *TGTTG*-CT*CAACATGTTG*A, and functions as an activator in placental cells. A similar motif has been identified in the human prolactin gene promoter (*TGTTGCAACA*), although the function of this motif has not been elucidated (Peers et al., 1990). However, a similar palindrome, *TGTTG-TGTTG*GTTGC*CAACA* (termed PSF-A/PSF-B), found in chorionic somatomammotropin A, B, L, and growth hormone variant genes, has been shown to bind to a putative factor PSF-1 (Nachtigal et al., 1993). It is postulated that the putative PSF-1 factor binds to the PSF-A (or PSF-B) motif and mediates repression of the expression of these placental genes in nontrophoblastic tissues. The spacing between the palindromic half-sites varied among rnCGM3, prolactin, and chorionic somatomammotropin/growth hormone variant genes. It has been demonstrated that the spacing between the half-sites of steroid hormone/retinoic acid response elements specifies ligand specificity (Umesono et al., 1991). Therefore, the spacing between the "TGTTG" half-sites may dictate the specificity of the binding factors. The presence of a PISII-like motif in several trophoblast genes suggests that it may play an important role in regulating the expression of rat PSG genes. To this end, we have further characterized the PISII binding factor and shown that it is a polypeptide with an apparent molecular mass of 40 kDa. We are currently screening a placental expression library in order to further characterize this putative transcription factor.

Another palindrome, *TCATCCC*TGGGATGA, was also detected within the PISI site of the upstream promoter, PI. Recently, an Olf-1 motif, TCCCNNGGGA/G, shown to bind the olfactory neuronal transcriptional activator, Olf-1, has been identified in several olfactory-specific genes (Wang & Reed, 1993). Moreover, the early B-cell factor (EBF), an alternatively-spliced form of Olf-1 and a regulator of B lymphocyte-specific gene expression, has also been shown to bind a similar Olf-1 motif (Hagman et al., 1993). These findings raise the possibility that a similar factor in the placenta may recognize the PISI site and the transactivate PSG expression.

ACKNOWLEDGMENT

We thank Drs. M. Chamberlin, L. L. Shelly, K.-J. Lei, and C. Park for critical reading of the manuscript.

REFERENCES

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1992) pp 9.2.1–9.2.6, Greene Publishing and Wiley-Interscience, New York.

Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K., & Stanners, C. P. (1989) *Cell 57*, 327–334.

Bohn, H., & Weinmann, E. (1974) *Arch. Gynaekol. 217*, 209–218.

Chen, H., Plouzek, C. A., Liu, J.-L., Chen, C.-L., & Chou, J. Y. (1992) *DNA Cell Biol. 11*, 139–148.

Chou, J. Y., & Plouzek, C. A. (1992) *Semin. Reprod. Endocrinol. 10*, 116–126.

Chou, J. Y., Sartwell, A. D., Lei, K.-J., & Plouzek, C. A. (1990) *J. Biol. Chem. 265*, 8788–8794.

de Wet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R., & Subramani, S. (1987) *Mol. Cell Biol. 7*, 725–737.

Dignam, J. D., Lebovitz, R. M., Roeder, R. G. (1983) *Nucleic Acids Res. 11*, 1475–1489.

Fordis, M., & Howard, B. H. (1987) *Methods Enzymol. 151*, 382–397.

Frohman, M. A., Dush, M. K., & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8998–9002.

Hagman, J., Belanger, C., Travis, A., Turck, C. W., & Grosschedl, R. (1993) *Genes Dev. 7*, 760–773.

Hultmark, D., Klemenz, R., & Gehring, W. J. (1986) *Cell 44*, 429–438.

Kodelja, V., Lucas, K., Barnert, S., von Kleist, S., Thompson, J. A., & Zimmermann, W. (1989) *J. Biol. Chem. 264*, 6906–6912.

Lei, K.-J., Gluzman, Y., Pan, C.-J., & Chou, J. Y. (1992a) Mol. Endocrinol. *6*, 703–712.

Lei, K.-J., Sartwell, A. D., Pan, C.-J., & Chou, J. Y. (1992b) *J. Biol. Chem. 267*, 16371–16378.

Lei, K.-J., Wang, C., Chamberlin, M. E., Liu, J.-L., Pan, C.-J., & Chou, J. Y. (1993) *J. Biol. Chem. 268*, 17528–17538.
Lin, H. H., & Ann, D. K. (1991) *Genomics 10*, 102–113.
MacDonald, D. J., Scott, J. M., Gemmell, R. S., & Mack, D. S. (1983) *Am. J. Obstet, Gynecol. 147*, 430–436.
Majello, B., Arcone, R., Toniatti, C., & Ciliberto, G. (1990) *EMBO J. 9*, 457–465.
Maniatis, T., Goodbourn, S., & Fischer, A. (1987) *Science 236*, 1237–1245.
Masson, G. M., Anthony, F., & Wilson, M. S. (1983) *Br. J. Obstet. Gynaecol. 90*, 146–149.
Matsunaga, T., Nomoto, M., Kozak, C. A., & Gonzalez, F. J. (1990) *Biochemistry 29*, 1329–1341.
McCormick, A., Brady, H., Fukushima, J., & Karin, M. (1991) *Genes Dev. 5*, 1490–1503.
Mercurio, F., & Karin, M. (1989) *EMBO J. 8*, 1455–1460.
Nachtigal, M. W., Nickel, B. E., & Cattini, P. A. (1993) *J. Biol. Chem. 268*, 8473–8479.
Nikovits, W. J., Mar, J. H., & Ordahl, C. P. (1990) *Mol. Cell. Biol. 10*, 3468–3482.
Ohlsson, H., & Edlund, T. (1986) *Cell 45*, 35–44.
Peers, B., Voz, M. L., Monget, P., Mathy-Hartert, M., Berwaer, M., Belayew, A., & Martial, J. A. (1990) Mol. Cell. Biol. *10*, 4690–4700.
Rebstock, S., Lucas, K., Thompson, J. A., & Zimmermann, W. (1990) *J. Biol. Chem. 265*, 7872–7879.
Rørth, P., Nerlov, C., Blasi, F., & Johnsen, M. (1990) *Nucleic Acids Res. 18*, 5009–5017.
Rudert, F., Saunders, A. M., Rebstock, S., Thompson, J. A., & Zimmermann, W. (1992) *Mamm. Genome 3*, 262–273.
Sachs, A. B. (1993) *Cell 74*, 413–421.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A 74*, 5463–5467.
Schmitz, E., Mohr, E., & Richter, D. (1991) *DNA Cell Biol. 10*, 81–91.
Shively, J. E., & Beatty, J. D. (1985) *CRC Crit. Rev. Oncol. Hematol. 2*, 355–399.
Snape, A. M., Winning, R. S., & Sargent, T. D. (1991) *Development 113*, 283–293.
Sorensen, S. (1984) Tumor Biol. *5*, 275–302.
Tamsen, L., Johansson, S. G. O., & Axelsson, O. (1983) *J. Perinat. Med. 11*, 19–25.
Tatarinov, Y. S. (1978) *Gynecol. Obstet. Invest. 9*, 65–97.
Theill, L. E., Wiborg, O., & Vuust, J. (1987) *Mol. Cell. Biol. 7*, 4329–4336.
Thompson, J. A., Grunert, F., & Zimmermann, W. (1991) *J. Clin. Lab. Anal. 5*, 344–366.
Turbide, C., Rojas, M., Stanners, C. P., & Beauchemin, N. (1991) *J. Biol. Chem. 266*, 309–315.
Umesono, K., Murakami, K. K., Thompson, C. C., & Evans, R. M. (1991) *Cell 65*, 1255–1266.
Wang, M. M., & Reed, R. R. (1993) *Nature 364*, 121–126.
Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V., & Ueda, H. (1987) *Science 238*, 1247–1253.
Wurz, H., Geiger, W., Kunzig, H. J., Jabs-Lehmann, A., Bohn, H., & Luben, G. (1981) *J. Perinat. Med. 2*, 67–78.
Yang, J.-Q., Remmers, E. F., & Marcu, K. B. (1986) *EMBO J. 5*, 3553–3562.