

**THE HUMAN PREGNANCY-SPECIFIC GLYCOPROTEIN GENES ARE TIGHTLY LINKED ON
THE LONG ARM OF CHROMOSOME 19 AND ARE COORDINATELY EXPRESSED**

**John Thompson, Rosa Koumari, Klaus Wagner, Sabine Barnert, Cathrin Schleussner,
Heinrich Schrewe, Wolfgang Zimmermann,
Gaby Müller¹, Werner Schempp¹,
Daniela Zaninetta², Domenico Ammaturo², and Norman Hardman²**

Institute of Immunobiology, University of Freiburg, Stefan-Meier-Str. 8, D-7800 Freiburg, FRG

¹Institute of Human Genetics, University of Freiburg, Albertstr. 11, D-7800 Freiburg, FRG

²Department of Molecular Biology, Biotechnology Section, Ciba-Geigy AG, CH-4002 Basel, Switzerland

Received January 22, 1990

The pregnancy-specific glycoprotein (PSG) genes encode a group of proteins which are found in large amounts in placenta and maternal serum. In situ hybridization analyses of metaphase chromosomes reveal that all the human pregnancy-specific glycoprotein (PSG) genes are located on the long arm of chromosome 19 (19q13.2-13.3), overlapping the region containing the closely-related carcinoembryonic antigen (CEA) gene subgroup. Higher resolution analyses indicate that the PSG genes are closely linked within an 800kb SacII restriction endonuclease fragment. This has been confirmed through restriction endonuclease mapping and DNA sequence analyses of isolated genomic clones, which show that at least some of these genes are located in very close proximity. Further, these studies have helped to identify a new member of the PSG gene sub-family (PSG7). DNA/RNA hybridization analyses, using gene-specific oligonucleotide probes based on published sequences, showed that five from six PSG genes tested are coordinately transcribed in the placenta. Due to the close proximity of these genes and their coordinated expression pattern, common transcriptional regulatory elements may exist. © 1990 Academic Press, Inc.

The pregnancy-specific β 1 glycoproteins (PSG) consist of a heterogenous group of molecules (1) which are expressed in the placenta during pregnancy and in tumors of trophoblastic origin (2). At term pregnancy, the PSG molecules represent some of the major placental proteins (3). Their presence in choriocarcinomas has indicated their potential utility as tumor markers (2). Until recently, little was known about the structure and complexity of this protein family, and their function still remains unknown. At least three immunologically-crossreacting PSG species with differing molecular sizes have been isolated from human placenta (4). Through molecular analyses of genomic and cDNA clones, ten separate PSG genes have been identified so far, and their coding regions could be deduced (4-14). Differential splicing of at least one gene has been reported, whereby four mRNA variants have been found, encoding proteins which differ only in their short C-terminal domain regions (4,5,9,11,14). This increases the potential number of PSG proteins to thirteen, assuming that all the transcripts are translated. Sequence comparisons reveal that these genes are homologous to those encoding the carcinoembryonic antigen (CEA) and related

molecules (15). The PSG/CEA gene family belongs to the immunoglobulin (Ig) superfamily (15,16). Basically, as derived from DNA sequence data, the proteins consist of one N-terminal Ig variable (IgV)-like, and a differing number of immunoglobulin constant (IgC)-like domains (17). Differences can be found in the C-terminal domain regions, whereby the CEA-related proteins appear to be membrane-bound, either after post-translational modification through processing and the addition of a glycosyl phosphatidylinositol tail (18-22), or as an integral membrane protein (23,24). Most of the PSG molecules apparently lack a transmembrane region and their deduced primary structures indicate that they are secreted (4,5,8,9,11).

The CEA-related genes are clustered on the long arm of chromosome 19 (25-27). In this paper we have determined the fine chromosomal location of the PSG genes. DNA/RNA hybridization investigations using presumed gene-specific oligonucleotides have been applied to study the transcriptional activity of individual genes. These basic studies should help us to learn more about the regulation of PSG gene expression, and pave the way for functional analyses of this important group of pregnancy-specific glycoproteins.

MATERIALS AND METHODS

In situ hybridization: Preparation of chromosomes and in situ hybridizations were carried out as described previously (27) using a 1026bp EcoRI/SstI restriction endonuclease fragment from λ hscGM3-1 (10) which contains the N-terminal domain exon, as a probe. After 3 weeks exposure at 4°C, the chromosomes were RBG-banded by the method of Perry and Wolff (29).

Isolation of high molecular weight genomic DNA and pulsed field gel electrophoresis analyses: High molecular weight DNA was isolated from an Epstein Barr virus-immortalized, human lymphoblast cell line, L1164 (gift from G. Scherer, Institute of Human Genetics, Freiburg University, FRG) in agarose blocks according to the European Molecular Biology Laboratory (EMBL, Heidelberg, FRG) protocol. Pulsed field gel electrophoresis of large restriction endonuclease fragments was carried out on a CHEF-type model (EMBL, Heidelberg, FRG), at 4 V/cm and a pulse time of 90s in a 0.8% agarose gel for 43h at 15°C as described by the manufacturers, and the DNA was transferred to a charged nylon membrane (GeneScreen-Plus, Dupont, Bad Homburg, FRG) according to the manufacturer's protocol. The membrane was hybridized with a probe containing the total coding region minus the 3'-untranslated portion of the PSG1a cDNA clone described elsewhere (11). The hybridization was carried out overnight at 42°C in 40% deionized formamide, 1M NaCl, 10% dextran sulphate, 1% SDS, 100 μ g/ml denatured salmon sperm DNA. Final washes were performed at 62°C in 0.1x SSPE (1x SSPE: 180mM NaCl, 10mM NaH₂PO₄, 1mM EDTA pH 7.4), 0.1% SDS.

Bacterial strains, human genomic libraries and cloning vectors: All work with recombinant DNA was carried out according to the German and Swiss safety regulations. A human leucocyte genomic library in λ EMBL3 (HL1006; Clontech, Palo Alto, USA) was amplified in *E. coli* 803. A human cosmid genomic library (gift from W. Lindenmaier, Gesellschaft für Biotechnologische Forschung, Braunschweig, FRG), inserted in the vector pHC79-2cos/TK, was amplified in the rec A *E. coli* strain ED8767. For subcloning of the insert DNA fragments, either Bluescript phagemid (Stratagene, San Diego, USA) was used in *E. coli* strains JM109 and RR1 Δ M15, or M13mp18/19 phage, which was amplified in *E. coli* strain JM107.

Screening of the human genomic libraries: Various DNA fragments were radiolabelled (28), and used to screen the human genomic libraries (30). For the lambda library, recombinant bacteriophages were amplified and transferred to nitrocellulose filters (Schleicher and Schuell, Dassel, FRG) as described (31). Screening and plaque purification was carried out by using the radioactively-labelled cDNA insert of pSP₁-i (8). Hybridization was carried out at 65°C in a solution containing 5x Denhardt's, 50 μ g/ml denatured salmon sperm DNA, 0.05% SDS and 6x SSC (1x SSC = 0.15M NaCl, 0.017M Na-citrate pH 7.0). Final washes were performed at 65°C using 0.1x SSC. For the cosmid library, the original screening was carried out using a 1.4kb EcoRI restriction endonuclease fragment, covering the 3'-untranslated region of the NCA cDNA clone 9 (32). Colony hybridization on nitrocellulose filters (30) was used to identify the cosmid clones. Pre-hybridization was carried out at 42°C in 40% deionized formamide, 6x SSPE, 1x Denhardt's, 0.5% SDS, 100 μ g/ml denatured salmon sperm DNA for 2h. Hybridization was under the same conditions but in the absence of SDS. Final washes were performed at 65°C in 2x SSPE, 0.5% SDS.

Restriction endonuclease mapping: Lambda clones were mapped by restriction endonuclease digests aided by hybridization with the pSP₁-i cDNA insert (8). Detailed mapping of clone λ PSG6/7-1 was carried out using DNA fragments subcloned in Bluescript plasmids. The cosmid clones were mapped by linearizing with the restriction endonucleases, PvuI or NheI, which cut in the vector at a short distance to the left and right of the cloning site, respectively. The resulting fragments were then partially digested with different restriction endonucleases (EcoRI, SstI and BamHI), separated by gel electrophoresis in 0.5% agarose overnight at 1 V/cm (4°C) and transferred to charged nylon membranes prior to hybridization with oligonucleotide probes whose complementary sequences are located in the vector between the PvuI site (oligo cos1), or the NheI site (oligo cos2) and the cloning site. The oligonucleotides were as follows:

oligo cos1 5' TAGGCGTATCACGAGGCCCTTTCG 3'
 oligo cos2 5' GGCGATGCTGTCCGAATGGACG 3'

Hybridization was carried out in 30% deionized formamide, 1M NaCl, 1% SDS, 10% dextran sulphate, 100 μ g/ml denatured salmon sperm DNA at 10-15°C below the theoretical melting temperature (33). Final washes were carried out in 2x SSPE, 0.5% SDS at 3-5°C below the theoretical melting temperatures. Size determination of the fragments hybridizing allowed the linear mapping of restriction endonuclease sites. Uncertainties arising from this method could be resolved by isolating specific DNA fragments, followed by digestion with a second restriction endonuclease and by agarose gel electrophoresis separation. This was accompanied by hybridization analyses with various DNA probes from the coding region of a PSG1a cDNA clone (11), which were labelled by random priming (28) and hybridized at 42°C in 40% deionized formamide, 1M NaCl, 1% SDS, 10% dextran sulphate, 100 μ g/ml denatured salmon sperm DNA. Final washes were carried out in either 0.1x, or 2x SSPE, 1% SDS at 65°C. The exact positions of the exon/intron borders were determined by DNA sequencing.

Determination of DNA sequences: Subcloned restriction endonuclease DNA fragments in phage M13mp18/19 and Bluescript phagemid were sequenced as single or double stranded templates by the dideoxy-chain-termination method (34), using universal, or internal oligonucleotide primers and a sequencing kit (United States Biochemicals). For comparison of the nucleotide and deduced amino acid sequences, the computer program "Align" (M. Trippel, Freiburg, FRG and R. Friedrich, Giessen, FRG: unpublished) was used.

RNA analyses by blot hybridization: RNA was extracted from a placenta of the 31st week of pregnancy according to the method of Chomczynski and Sacchi (35), or from colon tumour according to Fiddes and Goodman (36), as described elsewhere (27). Poly(A) RNA was purified once over an oligo(dT) column and size separated on methylmercury hydroxide denaturing agarose gels as described (27). Gene-specific oligonucleotides which were complementary to corresponding mRNAs were synthesized, based on known DNA sequence data, from the N-terminal domains of different PSG genes (Table I). Hybridization was carried out in 20-30% deionized formamide, 1% SDS, 1M NaCl, 10% dextran sulphate, 100 μ g/ml denatured salmon sperm DNA at 10-15°C below the theoretical melting temperatures (33). The final washes were performed in 2-6x SSPE at a temperature where only hybrids with none or maximally one mismatch are stable.

Primer extension experiments: In order to determine the transcriptional start for two PSG genes, gene-specific oligonucleotides which are complementary to corresponding mRNA species were used to prime extension reactions. For PSG5, the oligonucleotide used was oligo PSG5 (Table I). For comparison, we used an oligonucleotide from the 5'-untranslated region of PSG1 (oligo PSG1-5' = 5'GATAAGCCTAGGATCCAGAACTCT 3'), based on sequence data from a corresponding genomic clone (λ PSG β -1: [10], and unpublished data). For the annealing reaction, 5-10ng, 5'-end labelled oligonucleotide was mixed with 2.5 μ g poly(A) RNA isolated from a term placenta, in 10 μ l 5mM sodium phosphate buffer pH 7.0, 5mM EDTA. After 5 min denaturation at 90°C, 0.4 μ l 2M NaCl was added and hybridization took place in separate reactions at 40, 45, 50, 55 or 60°C for 1-2h. After this time the mixtures were either allowed to cool slowly to room temperature, or were directly taken for the extension reaction. The hybridization mix was adjusted to 17.5mM Tris/Cl pH 8.3, 4.3mM MgCl₂, 1.75mM dithiothreitol, 3.5mM dNTP, 1ng/ μ l actinomycin D, 2U/ μ l RNasin (Atlanta, Heidelberg, FRG), 0.8U/ μ l reverse transcriptase (Pharmacia, Freiburg, FRG), in 25 μ l volume and incubated at 42°C for 1h. Following a phenol extraction and ethanol precipitation, the extension products were denatured and loaded onto an 8M urea/6% polyacrylamide DNA sequencing gel.

S1 nuclease analyses: The transcriptional start of PSG1 was independently determined using S1 nuclease mapping. A 1.7kb Sst I restriction endonuclease fragment from the PSG β -1 clone described elsewhere (10), containing the first exon from PSG1 (data not shown), was subcloned in M13mp19. A single-stranded DNA fragment, which was complementary to the mRNA was synthesized from this template using the same ³²P-labelled primer oligonucleotide as for the primer extension experiment (see above). For this, the oligonucleotide was 5' end-labelled with polynucleotide kinase and annealed to the single-stranded M13 subclone according to the Perkin Elmer Cetus GeneAmp™ DNA amplification protocol, followed by extension with Taq-polymerase (Amersham, Braunschweig, FRG), at 72°C for 30 min. The extension product was digested with Pst I, separated on a preparative 4% polyacrylamide/8.3M urea gel, and the resulting 362nc fragment was purified by electroelution. The hybridization with poly(A) RNA from a term placenta (see above), and the S1 nuclease mapping was essentially according to Maniatis et al. (30). For digestion of single strands, the

DNA/RNA complex was incubated with 1000 U/ml nuclease S1 for 30 min at 18°C. Analysis of the double-stranded DNA/RNA complex was carried out on a DNA sequencing gel.

RESULTS

Chromosomal localization of the PSG genes: The results of the in situ hybridization analyses with metaphase chromosomes are shown in Figure 1. Utilizing published sequence data from the coding regions of the N-terminal domains (4,5,8-10) in the equation presented by Lathe (33), it is possible to predict that all the members of the PSG gene subgroup but not from the CEA subgroup, should cross-hybridize with the λ hsCGM3-1 probe (10) used, which corresponds to the gene PSG6 (13), under the conditions applied. Therefore, it appears that all the genes belonging to the PSG gene-subgroup are clustered on the long arm of chromosome 19 (Figure 1A). Finer mapping of the grain distribution locates this gene cluster to 19q13.2-q13.3 (Figure 1B). Closer studies regarding their location were carried out by pulsed field gel electrophoresis analyses. In order to test the hybridization stringency conditions for PSG gene subgroup specificity, isolated cosmid clones, cosPSG4/5-1 and cosPSG4/5-2 (see below), as well as lambda clones λ hsPS β G-1 and λ hsCGM3-1 (10), which, according to the newly proposed nomenclature system (13), contain the genes encoding PSG4, PSG5, PSG1 (formerly named PS β G) and PSG6 (formerly named hsCGM3) respectively, were co-hybridized as positive controls. Cosmid clone cosCEA1 (data to be published) and lambda clones λ 39.2 (37) and λ hsCGM2-1 (10), which contain the CEA, NCA and hsCGM2 genes, respectively, were co-hybridized as negative controls. Under the stringency conditions shown, all the PSG genes were found to hybridize, whereas the negative controls did not (results not shown). A control for cross-hybridization with genes PSG2 (= PS β G-E [9]), PSG3 (SP1 [8]), and PSG8 (= CGM35 [7]) was not carried out, but the high sequence similarity of all the PSG genes would suggest that they too should be positive under the stringency conditions applied. It can, therefore, be concluded that the strong signal seen in the genomic analysis (Figure 2) probably represents all the PSG genes, which are localized within an 800kb SacII restriction endonuclease fragment. The weaker band at around 980kb represents the region of limited mobility for large DNA fragments. This could contain incompletely digested DNA fragments, although it cannot be ruled out that a small number of PSG genes are located on at least one other, larger DNA fragment.

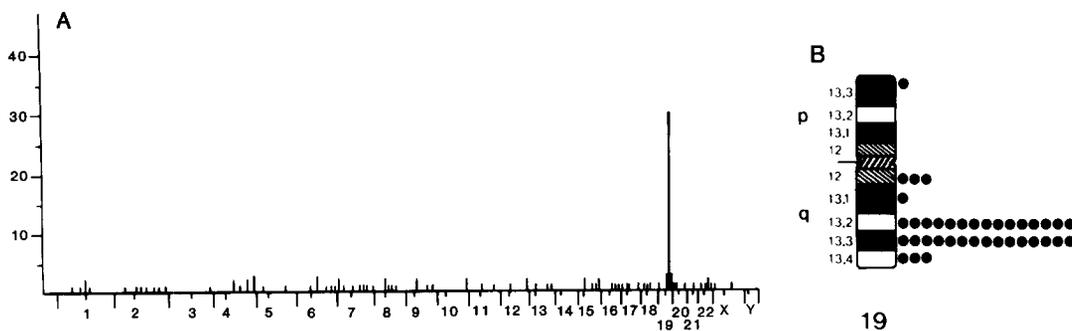


FIG. 1. (A) In Situ hybridization with a genomic PSG6 probe (see text). The histogram shows the grain distribution in metaphase chromosomes indicating that all PSG genes are located on chromosome 19. (B) Finer mapping of grain distribution over chromosome 19 localizes the PSG gene subfamily to the long arm at positions 19q13.2-q13.3.

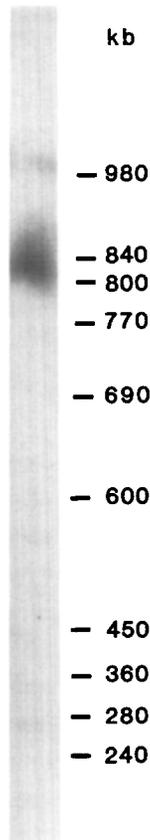


FIG. 2. Pulsed field gel electrophoresis analysis of human lymphoblast cell line L1164 DNA, digested with *SacII* and hybridized with the total coding region of a PSG1a cDNA clone (see text) under conditions whereby all PSG genes should hybridize. Calibration at the right was achieved by inclusion of *Saccharomyces cerevisiae* chromosomes as length markers.

Fine structural analyses of cloned PSG genes: In order to study the gene linkage more exactly, we started to screen a cosmid library by chromosome walking, so as to determine how close neighbouring PSG genes are. The two cosmid clones cosPSG4/5-1 and cosPSG4/5-2 were chosen for closer analyses because they revealed very similar restriction endonuclease fragment sizes for a given enzyme, as well as similar hybridization patterns with a number of different DNA probes (results not shown), indicating that they represent two overlapping clones, and thus, provide a better chance for finding linkage than analysing single clones. Furthermore, one of these clones (cosPSG4/5-1, adopting the new nomenclature system [13]) has already been shown to contain the N-terminal domain exon for a member of the PSG gene subgroup (10), where it was formerly named cosmid hsCGM4-1. Restriction endonuclease mapping, along with hybridization analyses and DNA sequencing was carried out for these different clones in order to determine the exon/intron structures and the identity of the genes contained therein (Figure 3B). It could be confirmed that the two cosmid clones overlap through DNA sequencing. Together these clones contain the entire coding regions of two PSG genes, PSG4 and PSG5.

In parallel to the cosmid analyses, a lambda library was screened adopting the following strategy. Since the PSG cDNA insert of pSP₁-i (8) contained internal *Bam*HI cleavage sites, it was supposed

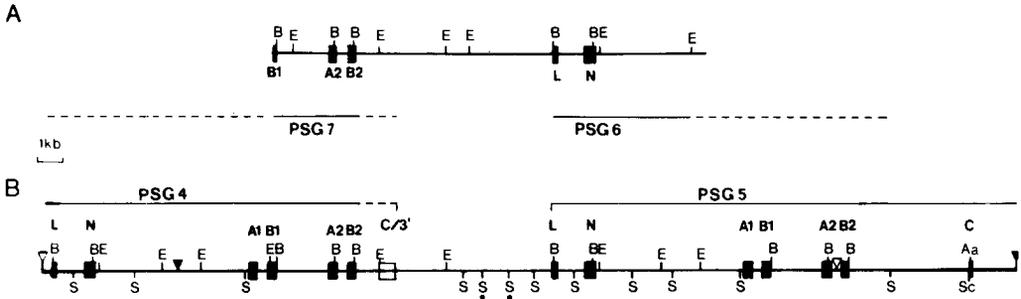


FIG. 3. Restriction map and sequencing strategy of (A) the λ PSG6/7 clone, and (B) the combined cosPSG4/5-1 and cosPSG4/5-2 clones, whereby the boundaries of the former clone are indicated with open, and the latter with filled-in triangles. The following restriction endonuclease sites have been mapped: E = EcoRI, B = BamHI, S = SstI, Sc = ScaI, Aa = AatII. For the latter two enzymes, only the sites used for mapping are shown. Filled in circles depict uncertain restriction sites. For Sau3a (Sa), only the sites used for sequencing are shown. The exons are indicated as filled-in boxes, apart from the C/3' exon of PSG4, which has only been provisionally characterized (see text), and is designated as an open box. L = Leader exon, N = N-terminal exon, A, B = IgC-like exons A and B, C = C-terminal exon in PSG5, C/3' = C-terminal and 3'-untranslated exon in PSG4. The borders of the transcription units are shown as far as they are known, and dashed lines indicate presumed regions, in analogy to known PSG gene borders.

that digestion of genomic DNA clones with BamHI, followed by hybridization with pSP₁-i might aid mapping. Of seven independent clones analysed, five contained two characteristic cDNA-hybridizing BamHI fragments of 0.7 and 2.3kb. Clone λ PSG6/7-1 also contained an additional cDNA-hybridizing band of 1.7kb, and was thus chosen for further analysis as containing potentially the most structural information. This lambda clone indeed contains sequences from two neighbouring PSG genes (Figure 3A). The N-terminal domain exon of the 5'-lying gene is not present within this clone and it cannot be identified with respect to known PSG genes. However, the high sequence similarity to other PSG genes in the corresponding repeat domain exons reveals its identity as a member of this gene subgroup (92-97% at the nucleotide level), rather than the CEA subgroup (66-78%). Furthermore, sequence comparisons of these exons with the corresponding regions for other members of the PSG subgroup reveal that they differ from all known sequences. For this reason, we have designated this gene PSG7, which has already been included in the list compiled by Barnett and Zimmermann (13). The second gene (PSG6) in this clone is identical in its N-terminal domain exon to the recently described hsCGM3 gene (10), and corresponds to the cDNA clone λ PSG6, described recently (11).

DNA sequencing of all three clones was restricted to the whole exons and parts of their surrounding introns. For PSG5, the putative promoter region was sequenced as far as 380 nucleotides upstream from the initiator codon. The sequence data are not presented here, but are being forwarded to GenBank, Los Alamos, USA.

Identification of transcripts for individual PSG genes: We have started to study the expression patterns of individual PSG genes, which may be expected to be coordinated, due to their tight linkage and strong sequence similarities. Gene-specific oligonucleotides were synthesized after comparing the DNA sequences from the N-terminal domains of known PSG genes or their transcripts (Table I). Each of these oligonucleotides should, at a high stringency, only hybridize with transcripts of one PSG gene. This was first calculated for each oligonucleotide according to Lathe (33) and confirmed for oligo PSG2 and oligo PSG4 by hybridizing them with isolated PSG

TABLE 1: PSG gene-specific oligonucleotide sequences.

Name	Sequence	Source of sequence and reference
oligo PSG1	5' CTAGTCCCATCATCTCCCTTTATG 3'	(PSG93 [4])
oligo PSG2	5' CTCGAAACTTTTGGTGGCTGGGC 3'	(PSβG-E [9])
oligo PSG3	5' TTATTTGACCATCTACTACGTATG 3'	(SP1-i [8])
oligo PSG4	5' GTGACTTGGGCAGTTGTGGGCGGA 3'	(λhsCGM4 [10])
oligo PSG5	5' GTGACTTGAGCAGTGATAGGCAGG 3'	(cosPSG4/5)
oligo PSG6	5' AGTATAAGGTGACAGTAAATATC 3'	(λhsCGM3 [10])

The corresponding cDNA/genomic clones from which they are derived are shown in brackets.

genomic clones (data not shown). On hybridizing all these oligonucleotides with poly(A) RNA from placenta, transcripts for five of the genes (PSG1-5) could be identified (Figure 4A, even numbered lanes). Although it cannot be conclusively ruled out that one or the other oligonucleotide could cross-hybridize with transcripts from more than one PSG gene, the different hybridization patterns using the various oligonucleotides indicate a high degree of specificity for the individual probes. Only oligonucleotides PSG2 and PSG5 show a similar hybridization signal. This is probably due to the fact that the corresponding transcripts are of similar sizes, as deduced from published cDNA data (6,9), rather than to cross-hybridization. For some of these oligonucleotides, more than one transcript could be visualized, e.g. for oligo PSG1, two mRNA species (Figure 4A, lane 2), and with oligo PSG4, three could be differentiated (Figure 4A, lane 8). Poly(A) RNA from a colon adenocarcinoma, which could be shown to contain mRNAs for CEA (Figure 4B, lane 1), was also hybridized, whereby no transcripts for any of the PSG genes could be identified (Figure 4A, odd numbered lanes). Furthermore, CEA transcripts could not be found in the placenta (Figure 4B, lane 2).

Due to differences in the specific activities of the radioactively-labelled oligonucleotides, it is not possible to quantitate variability in expression between the different gene transcripts. However,

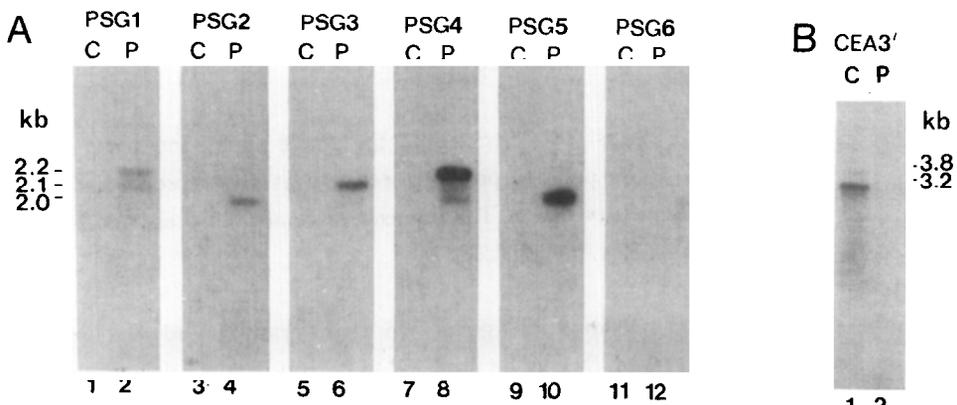


FIG. 4. DNA/RNA (Northern) hybridization analyses of poly(A) RNA from: (A) colon tumor (C) in the odd numbered lanes, and 31st week placenta (P) in the even numbered lanes, hybridized at high stringency (see text), with the following oligonucleotides (see Table 1): oligo PSG1 (lanes 1 and 2), oligo PSG2 (lanes 3 and 4), oligo PSG3 (lanes 5 and 6), oligo PSG4 (lanes 7 and 8), oligo PSG5 (lanes 9 and 10) and oligo PSG6 (lanes 11 and 12), which are specific for the genes indicated above. (B) colon tumor (lane 1) and term placenta (lane 2), hybridized with the *Rsa*I/*Pst*I fragment from the 3'-untranslated region of the pCEA3 cDNA clone described elsewhere (11) and washed at high stringency (0.1x SSPE, 0.1% SDS, 65°C).

no transcript could be found in the placenta using the PSG6 gene-specific oligonucleotide, which as a control, did hybridize with the corresponding genomic clone λ hsCGM3-1 (10) under the same hybridization conditions (data not shown). No gene-specific oligonucleotide was tested for PSG7.

Determination of the transcriptional starts for two PSG genes: As a basis for studying the promoter regions of PSG genes in a search for common regulatory elements, we have determined the transcriptional start sites for the PSG1 and PSG5 genes. The oligonucleotide used for PSG1 (oligo PSG1-5') exhibited an identical hybridization pattern with term placental poly(A) RNA (data not shown) to the gene-specific oligo PSG1 (cf. Figure 4A, lane 2), and was used because it is situated closer to the transcriptional start than oligo PSG1, thus allowing a more precise determination. Analysis of the primer extension experiments (Figure 5A) indicates multiple transcriptional

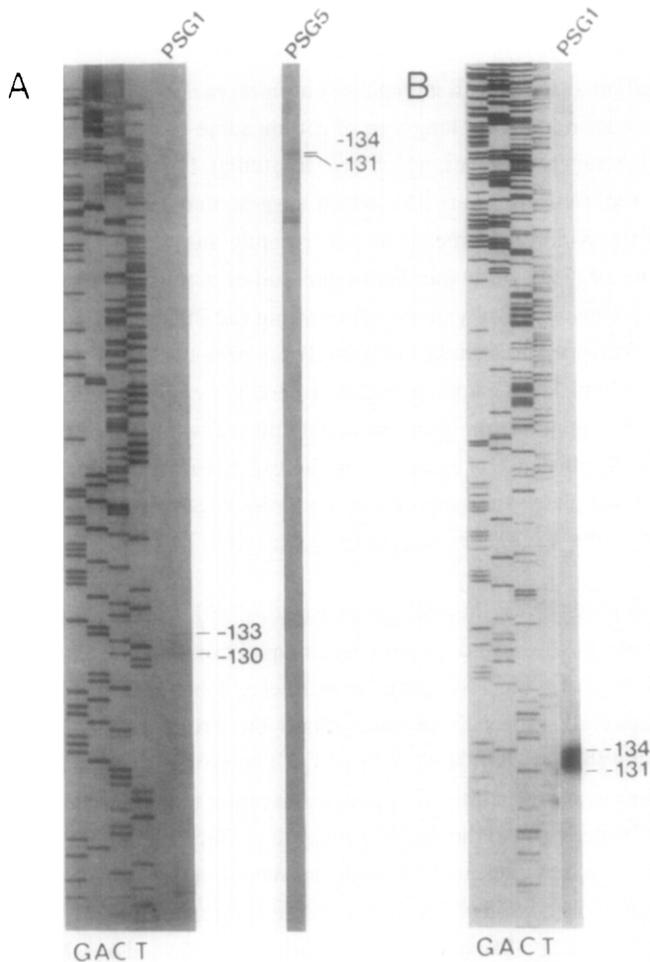


FIG. 5. Determination of the transcriptional start of PSG1 and PSG5 through: (A) primer extension analyses. As a length marker, a DNA sequence reaction for PSG1 was run, using the same 5'-oligonucleotide primer as for the PSG1 primer extension reaction (see text), directly to the right. The PSG5 primer extension product was run on the same gel and is longer, because the primer used there binds in the N-terminal region of that gene (see text). In both cases, the numbers shown indicate the position of the longest extension products upstream from the translational starts of each gene. (B) S1 nuclease analysis. As a length marker, an NCA genomic sequence from the 5' region of that gene (37) was used. The sizes indicate the positions of the 5' ends of the protected fragment with respect to the translational start of PSG1. The longer product visible as a band at the top of this lane corresponds to the 362nc DNA fragment (see text).

starts for both the PSG1 and PSG5 genes, where the longest extension products lie 130-134 nucleotides upstream of the translational starts, in both cases. Hybridization at different temperatures only revealed this pattern at 45°C with oligo PSG1-5', and above or below, these bands were not seen (results not shown). Using oligo PSG5, similar patterns were found up to 50°C, and above no signal was visible. The shorter, strong bands at approximately -100 may represent alternative start sites, because their presence or absence shows an identical annealing temperature dependence pattern to the -130 to -134 bands at the different annealing temperatures. In contrast, other, smaller extension products probably represent premature termination points, due to secondary, structural hindrance of the reverse transcriptase, because they reveal a different annealing temperature dependence. The S1 nuclease assay for PSG1 (Figure 5B) confirms the multiple starts at the same positions found with the primer extension analyses. No recognizable TATA box could be identified for either gene.

DISCUSSION

The *in situ* hybridization analyses with metaphase chromosomes demonstrate, unequivocally, that all the PSG genes are located on the long arm of chromosome 19 at position q13.2-q13.3 (Figure 1B), close to the CEA subgroup genes, which are located at 19q13.1-13.2 (25-27). These results conflict with those reported elsewhere (5), which suggest that the PSG genes are located on chromosome 6 and the X chromosome. However, recently another group also located the PSG genes to chromosome 19 (38), and in parallel to our studies, a third group located these genes to 19q13.1-13.2 (39). The close proximity of the genes within the PSG subgroup as suggested through the pulsed-field gel electrophoresis studies (Figure 2) has been confirmed for some of these genes through restriction endonuclease mapping studies, whereby it could be shown in two cases that a distance of only ca. 6kb separates the transcriptional units of two pairs of PSG genes (Figure 3A and B). The fact that this distance is conserved in the two cases analysed, along with the suggested localization of this complete gene subgroup within an 800kb SacII restriction endonuclease fragment, indicates that all the PSG genes are similarly clustered.

The existence of a gene cluster for the PSG genes, together with their strong sequence similarities suggests that the genes have evolved recently by unequal crossing over, which has led to gene multiplication. On comparing the PSG genes found here, it becomes obvious that certain restriction sites are conserved, especially in and around the exon regions. When comparing the restriction sites around the closely-related pairs of IgC-like exons (A, B), a stronger conservation between corresponding exons from different genes exists rather than between IgC-like exons within a gene, eg. the BamHI sites close to the A2/B2 exons are conserved in PSG7, PSG4 and PSG5 (see Figure 3). These sites are, however, only partially conserved in the A1/B1 exon regions of these genes, where intergene conservation of restriction endonuclease sites for SstI/BamHI can be found. Indeed this higher intergene IgC-like domain conservation is confirmed by DNA sequencing, where the derived amino acid sequence similarity is 90-95% for corresponding intergene repeat exons, but only 43-46% for internal repeat units (i.e. A1 vs. A2, B1 vs. B2). Oikawa et al. (7) find similar results in their PSG gene, CGM35 (80-95% and 44-49%, respectively). This suggests that during evolution, the IgC-like exon pairs within a primordial gene duplicated prior to the whole gene unit, whose own duplication gave rise to new PSG genes more recently. This supplements the evolutionary models which we have reported elsewhere (17,40).

The exon structures of all four genes characterized here are identical to those already found for other members of the CEA/PSG gene family (7,10,12,37,41), whereby a strong correlation exists between the exon and the domain borders. Interestingly, corresponding cDNA clones for PSG4, which have been reported elsewhere (hHSP2: [5], and λ PSG4: [11]), contain the A1, A2, B2 but lack the B1 repeat subdomains. We find only four nucleotide differences between the genomic exon sequences and the combined cDNA sequences, indicating polymorphism for this gene (data not shown). These exchanges lead to three amino acid exchanges. The cDNA clone FL-NCA3 (6) corresponds to PSG5 and shows eleven nucleotide differences, leading to five amino acid exchanges, contains only the A2 and B2, but lacks both the A1 and B1 subdomains. Recently, Oikawa et al. (12) have analysed the same gene. The genomic analyses for PSG4 and PSG5, show that the splice sites for these missing exons in PSG4 and PSG5 are apparently not degenerate, as has been reported for another PSG gene, CGM35 (7). It is interesting to note that the B1 domains from PSG4 and PSG5, which are lacking in their corresponding cDNA clones hHSP2 (5) and FL-NCA-3 (6), respectively, both contain stop codons. In the case of PSG4, this is due to a point mutation, whereas in PSG5, a deletion of two nucleotides, when compared with the B1 exon of other PSG genes, leads to a frame shift, which is followed by a stop codon. The A1 exon of PSG5, which is missing from the corresponding cDNA clone contains an open reading frame.

The nucleotide sequence determined downstream from the B2 exon of the PSG4 gene shows homology to the corresponding region region in the CGM35 gene (7). Furthermore, the conservation of the first EcoRI restriction endonuclease site, downstream of the B2 exon in the 3'-untranslated region of PSG4 (Figure 3B), with respect to the already sequenced CGM35 clone (7), indicates similarity to the latter further downstream. We also have provisional sequence data (not shown) downstream from this EcoRI site in PSG4 which confirm continued homology to the corresponding PSG4 cDNA (11), and the CGM35 clone (7). This whole region in CGM35 has been shown to contain a number of exons (7), which could be differentially spliced to give rise to four mRNA species, differing in their 3'-untranslated regions, as shown for PSG1 (11). The PSG5 3'-untranslated region, on the other hand, appears to differ with respect to the EcoRI site, and indeed the nucleotide sequence diverges from the PSG4 and CGM35 sequences (7), 152 nucleotides downstream from the 3'-end of the B2 exon (data not shown). This position coincides with the point of divergence between the CEA 3'-untranslated region and the corresponding region in CGM35, as reported elsewhere (7). It was recently suggested that this divergence may have been caused by a deletion in the CGM35 gene and correspondingly in the PSG4 gene (11). This hypothesis is strengthened by the present data. We could identify the presence of a separate exon, downstream from B2 in PSG5 through DNA sequencing, which contains coding information for the short C-terminal domain and the first 44 nucleotides of the 3'-untranslated region, which confirms the model which we recently proposed (11). The rest of the 3'-untranslated region is encoded by at least one separate exon, which is not present within this cosmid clone, as confirmed by hybridization with an oligonucleotide from that region (data not shown). The 3'-untranslated region of PSG7 has not been analysed, but the conserved distance between PSG7 and PSG6 as compared to PSG4 and PSG5 (Figure 3) would indicate similarity in the 3'-exon containing region to PSG4.

In order to see whether or not these closely linked PSG genes reveal similar expression patterns, we have looked for individual PSG transcripts in the placenta and a colon tumor (Figure 4A),

whereby signals for five of six genes tested could be found in the placenta only. In a previous study, we were unable to detect transcripts for PSG genes in a number of tumor tissues (10). Together, these results indicate a high degree of tissue-specific expression. With the gene-specific oligonucleotides for PSG1 and PSG4, multiple transcripts could be visualized (Figure 4A, lanes 2 and 8). For PSG1, four mRNA splice variants have been described through cDNA cloning (4,5,9,11,14). A similar set of splice products can be predicted to exist for PSG4, PSG7 and CGM35 (7), which appear to have similar 3'-untranslated regions to PSG1 (see above). For PSG2 (=PS β G-E [9]), PSG3 (=SP1-i [8]) and PSG5, only one transcript could be found in each case (Figure 4A, lanes 4,6 and 10). These three genes have previously been predicted, through cDNA analyses, to have a similar exon structure in their 3'-untranslated regions (11), showing differences to the PSG4-related group. The smaller sizes of the PSG2 and PSG5 mRNAs are probably due to a deletion, caused by alternative splicing, which has not been observed for PSG3. Although no transcript could be identified for PSG6, we have recently been able to confirm the transcriptional activity of this gene through the isolation of a corresponding cDNA clone (λ PSG6) from a fetal liver cDNA library (11). No gene-specific oligonucleotides have so far been tested for PSG7.

The coordinated expression for five of the PSG genes in a tissue-specific manner suggests that common regulatory elements may control their transcriptional activities. The structural organization of these genes within a closely linked gene cluster would facilitate such controls. Alternatively, there is room for individual promoters between each PSG gene (Figure 3), which may, in analogy to the PSG coding sequences, also be highly conserved, following the suggested recent duplication of complete PSG gene units (cf.40). This would in turn lead to highly similar regulation patterns. Closer analyses of the putative promoter regions should clarify this.

ACKNOWLEDGMENTS

We wish to thank Prof. S. von Kleist for laboratory use, and Nicole Brückel for excellent technical assistance. The oligonucleotides were synthesized by Dr. Gabor Igloi on an automated DNA synthesizer, which was a gift from the "Fonds der Chemischen Industrie" to Dr. Hans Kössel. A special thankyou is made to Dr. W. Lindenmaier, Braunschweig, FRG., for the human cosmid library. This work was supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Bohn, H. (1972) *Blut* 24, 292-302.
2. Tatarinov, Y.S. (1978) *Gynecol. Obstet. Invest.* 9, 65-97.
3. Lin, T.-M., Halpert, S.P., and Spellacy, W.N. (1974) *J. Clinic. Invest.* 54, 576-582.
4. Watanabe, S., and Chou, J.Y. (1988) *J. Biol. Chem.* 263, 2049-2054.
5. Chan, W.-Y., Borjigin, J., Zheng, Q.-X., and Shupert, W.L. (1988) *DNA* 7, 545-555.
6. Khan, W.N., and Hammarström, S. (1989) *Biochem. Biophys. Res. Commun.* 161, 525-535.
7. Oikawa, S., Inuzuka, C., Kosaki, G., and Nakazato, H. (1988) *Biochem. Biophys. Res. Commun.* 156, 68-77.
8. Rooney, B.C., Horne, C.H.W., and Hardman, N. (1988) *Gene* 71, 439-449.
9. Streydio, C., Lacka, K., Swillens, S., and Vassart, G. (1988) *Biochem. Biophys. Res. Commun.* 154, 130-137.
10. Thompson, J.A., Mauch, E.-M., Chen, F.-S., Hinoda, Y., Schrewe, H., Berling, B., Barnert, B., von Kleist, S., Shively, J.E., and Zimmermann, W. (1989) *Biochem. Biophys. Res. Commun.* 158, 996-1004.
11. Zimmermann, W., Weiss, M., and Thompson, J.A. (1989) *Biochem. Biophys. Res. Commun.* 163, 1197-1209.
12. Oikawa, S., Inuzuka, C., Kuroki, M., Matsuoka, Y., Kosaki, G., and Nakazato, H. (1989) *Biochem. Biophys. Res. Commun.* 163, 1021-1031.

13. Barnett, T., and Zimmermann, W. (1990) *Tumor Biol.* in press.
14. Khan, W., Osterman, A., and Hammarström, S. (1989) *Proc. Natl. Acad. Sci. USA.* 86, 3332-3336.
15. Watanabe, S., and Chou, J.Y. (1988) *Biochem. Biophys. Res. Commun.* 152, 762-768.
16. Paxton, R., Mooser, G., Pande, H., Lee, T.D., and Shively, J.E. (1987) *Proc. Natl. Acad. Sci. USA* 84, 920-924.
17. Thompson, J., and Zimmermann, W. (1988) *Tumor Biol.* 9, 63-83.
18. Grunert, F., Kolbinger, F., Schwarz, K., Schwaibold, H., and von Kleist, S. (1988) *Biochem. Biophys. Res. Commun.* 153, 1105-1115.
19. Hefta, S.A., Hefta, L.J.F., Lee, T.D., Paxton, R.J., and Shively, J.E. (1988) *Proc. Natl. Acad. Sci. USA* 85, 4648-4652.
20. Jean, F., Malapert, P., Rougon, G., and Barbet, J. (1988) *Biochem. Biophys. Res. Commun.* 155, 794-800.
21. Kolbinger, F., Schwarz, K., Brombacher, F., von Kleist, S., and Grunert, F. (1989) *Biochem. Biophys. Res. Commun.* 161, 1126-1134.
22. Takami, N., Misumi, Y., Kuroki, M., Matsuoka, Y., and Ikehara, Y. (1988) *J. Biol. Chem.* 263, 12716-12720.
23. Barnett, T.R., Kretschmer, A., Austen, D.A., Goebel, S.J., Hart, J.T., Elting, J.J., and Kamarck, M.E. (1989) *J. Cell Biol.* 108, 267-276.
24. Hinoda, Y., Neumaier, M., Hefta, S.A., Drzenick, Z., Wagener, C., Shively, L., Hefta, L.J., Shively, J.E., and Paxton, R.J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 6959-6963.
25. Kamarck, M.E., Elting, J.J., Hart, J.T., Goebel, S.J., Rae, P.M.M., Nothdurft, M.A., Nedwin, J.J., and Barnett, T.R. (1987) *Proc. Natl. Acad. Sci. USA* 84, 5350-5354.
26. Willcocks, T.C., Craig, S.P., and Craig, I.W. (1989) *Ann. Hum. Genet.* 53, 141-148.
27. Zimmermann, W., Weber, B., Ortlieb, B., Rudert, F., Schempp, W., Fiebig, H.-H., Shively, J.E., von Kleist, S., and Thompson, J.A. (1988) *Cancer Res.* 48, 2550-2555.
28. Feinberg, A.P., and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6-13.
29. Perry, P., and Wolff, S. (1974) *Nature* 251, 156-158.
30. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York.
31. Benton, W.D., and Davis, R.W. (1977) *Science* 196, 180-182.
32. Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., and Riggs, A.D. (1988) *J. Biol. Chem.* 263, 3202-3207.
33. Lathe, R. (1985) *J. Mol. Biol.* 183, 1-12.
34. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
35. Chomczynski, P., and Sacchi, N. (1987) *Analyt. Biochem.* 162, 156-159.
36. Fiddes, J.C., and Goodman, H.M. (1979) *Nature* 281, 351-356.
37. Thompson, J.A., Pande, H., Paxton, R.J., Shively, L., Padma, A., Simmer, R.L., Todd, C.T., Riggs, A.D., and Shively, J.E. (1987) *Proc. Natl. Acad. Sci. USA* 84, 2965-2969.
38. Barnett, T.R., Pickle, W., Rae, P.M.M., Hart, J., Kamarck, M., and Elting, J. (1989) *Am. J. Hum. Genet.* 44, 890-893.
39. Niemann, S.C., Schonk, D., van Dijk, P., Wieringa, B., Grzeschik, K.H., and Bartels, I. (1989) *Cytogenet. Cell. Genet.* 52, 95-97.
40. Rudert, F., Zimmermann, W., and Thompson, J. (1989) *J. Mol. Evol.* 29, 126-134.
41. Oikawa, S., Kosaki, G., and Nakazato, H. (1987) *Biochem. Biophys. Res. Commun.* 146, 464-469.