cDNA CLONING DEMONSTRATES THE EXPRESSION OF PREGNANCY-SPECIFIC GLYCOPROTEIN GENES, A SUBGROUP OF THE CARCINOEMBRYONIC ANTIGEN GENE FAMILY, IN FETAL LIVER

Wolfgang Zimmermann, Martina Weiss and John A. Thompson

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

Summary: The pregnancy-specific glycoprotein (PSG) genes constitute a subgroup of the carcinoembryonic antigen (CEA) gene family. Here we report the cloning of four cDNAs coding for different members of the PSG family from a human fetal liver cDNA library. They are derived from three closely related genes (PSG1, PSG4 and PSG6). Two of the cDNA clones represent splice variants of PSG1 ($PSG1_a$, $PSG1_d$) differing in their C-terminal domain and 3'-untranslated regions. All encoded proteins show the same domain arrangement ($N-R_{A1}-R_{A2}-R_{B2}-C$). Transcripts of the genes PSG1 and PSG4 could be detected in placenta by hybridization with gene-specific oligonucleotides. Expression of cDNA in a mouse and monkey cell line shows that the glycosylated $PSG1_a$ protein has a M_r of 65-66 kD and is released from the transfected cells. Sequence comparisons in the C-terminal domain and the 3'-untranslated regions of CEA/PSG-like genes suggests a complex splicing pattern to exist for various gene family members and a common evolutionary origin of these regions.
© 1989 Academic Press, Inc.

The carcinoembryonic antigen (CEA) is a human tumor marker of great significance in clinical oncology. Although CEA lacks absolute tumor specificity (1), its serum levels are used to monitor the success of therapy mainly for adenocarcinomas of the colon, breast and lung (2). The CEA gene is the prototype of a larger gene family which can be subdivided into the CEA and the pregnancy specific glycoprotein (PSG) subgroups (3). So far 14 CEA- and PSG- like genes have been identified, some of them only by partial genomic sequences (3-19). All members of the CEA family belong to the immunoglobulin superfamily and are composed of one variable (V) and a varying number (2, 3 or 6) of constant (C) domains (3,20-22). The in vivo function of CEA- and PSG-like antigens is unknown, although recent experiments indicate, that CEA might play a role in cell adhesion (23).

CEA- and PSG-like antigens comprise a heterogenous group of glycoproteins (for a review see 24; 25,26). PSG-like antigens are synthesized in the placenta and accumulate in large amounts in maternal serum during pregnancy (27,28). Furthermore, these antigens have been discussed as potential markers for tumors of trophoblastic origin (29). Expression of CEA-like genes has been shown for a number of normal and cancerous tissues. Transcripts of the CEA and/or of the non-specific crossreacting antigen (NCA) gene have been identified in tumors of the colon, breast, pancreas, gall bladder and lung and in normal colonic mucosa (30,31). The biliary glycoprotein I (BGP I) gene is transcribed in normal liver and in various tumor cell lines (13,14). In

contrast, not much is known about the expression of CEA- and PSG-like genes during fetal development, except that CEA is found in a variety of fetal tissues (32-34).

In order to identify other members of the CEA gene family which are expressed during human development, we decided to screen cDNA libraries derived from human fetal tissues for the presence of CEA/PSG-like cDNA clones. Such clones would allow delineation of the complete primary structures for members of the CEA family as well as helping to study their function during embryogenesis and tumor development.

MATERIALS AND METHODS

Cell lines and tissues: The human choriocarcinoma cell line BeWo (American Type culture collection ATCC CCL 98) and the SV40 transformed African green monkey kidney cells COS-7 (ATCC CRL 1651) were propagated in Dulbecco's modified Eagle's medium (DMEM) supplemented with penicillin, streptomycin and 15% or 10% fetal calf serum, respectively. The human colonic tumor was propagated in nude mice (35). The human tissues were immediately frozen in liquid nitrogen and stored at -140 °C until use.

Screening of the cDNA library: The PSG cDNA clones were isolated from a first trimester fetal liver cDNA library (Clontech, Palo Alto, USA) by screening the lambda gt11 recombinant phages with a mixture of CEA and NCA cDNA probes from pCEA1 (4), pCEA5 and pNCA1 (36). The cDNA inserts of these plasmids comprise part of the sequences coding for the N-terminal domains of CEA and NCA and repeated domains of CEA. Hybridization of the phage DNAs on nitrocellulose filters (Schleicher and Schuell, Dassel, FRG) was performed in the presence of 40% formamide, 5x Denhardt's solution (1x = 0.02% each of Ficoll (Pharmacia, Freiburg, FRG), polyvinyl pyrrolidon, bovine serum albumin (fraction V, Serva, Heidelberg, FRG), 6x SSPE (1x = 0.18 M NaCl, 10 mM sodium phosphate pH 7.4, 1 mM EDTA), 0.5% sodium dodecylsulfate (SDS) and 100 μg/ml heat-denatured salmon sperm DNA at 37 °C over night. After a final wash at 65 °C in 1x SSPE, 0.1% SDS for 30 minutes the filters were exposed to X-ray films. Positive plaques were isolated and some of them plaque-purified twice (37).

DNA sequencing and sequence analysis: The cDNAs were sequenced after subcloning into Bluescript (Stratagene, La Jolla, USA) or M13 vectors as double or single strands according to Sanger (38) using universal or internal oligonucleotide primers. The 17 mer oligonucleotide primers were synthesized and purified as described before (36). For comparison of the nucleotide and amino acid sequences the computer programs "Align" (M. Trippel and R. Friedrich, unpublished) and "Dot Matrix" (Rechenzentrum der Universität Freiburg) were used.

Expression of PSG1$_a$ cDNA in mouse 3T3 cells
Construction of pBEH/PSG1$_a$: The full length cDNA of PSG1$_a$ was excised from the vector by digestion with EcoRI and cloned into the expression vector pBEHpUC9 (39) downstream of the SV40 early promoter. First, the 1.3 kb EcoRI cDNA fragment containing the coding region of PSG1$_a$ was cloned into the vector. The correct orientation of the insert was demonstrated by digestion with SmaI (Fig. 4A). The plasmid with the correctly oriented EcoRI fragment was then partially digested with EcoRI, the linearized recombinant plasmid was recovered from the low melting point agarose gel after separation of the DNA fragments by electrophoresis and ligated to the 0.6 kb EcoRI fragment of PSG1$_a$. Among the clones obtained pBEH/PSG1$_a$ was identified as the expression vector construct with the correct 3'-location and orientation of the 0.6 kb EcoRI cDNA fragment by yielding fragments of expected size after double digestion with SmaI and SpeI (Fig. 4A).

Stable transfection of pBEH/PSG1$_a$ into 3T3 cells: Stable transfectants were obtained using the recently described lipofection technique (40). 0.5-1.0 x 10⁶ NIH 3T3 cells were seeded in a 75 cm plastic culture dish in DMEM supplemented with 10% fetal calf serum, penicillin and streptomycin. 24 hours later the cells were washed three times with DMEM without serum and treated with 20 μg of DOTMA (a gift from Phil Felgner, Syntex Research, Palo Alto, USA), 1 μg of the plasmid pSV2neo (41), which confers resistance toward the neomycin-derivative G418 and a tenfold molar excess (10 μg) of the PSG1$_a$ expression construct pBEH/PSG1$_a$ in 3 ml of DMEM without serum. Control cultures were only transfected with 1 μg of pSV2neo DNA. After 24 hours 10 ml of DMEM/10% fetal calf serum were added and 72 hours later selection for cells expressing the bacterial neo gene was started by addition of 1 mg/ml of the antibiotic G418 (GIBCO-BRL, Karlsruhe, FRG). After three weeks 100-300 large colonies were visible. They were harvested by trypsinization and used as a mixture for metabolic labeling experiments.

Metabolic labeling and immunoprecipitation of PSG1$_a$ protein:
Half confluent cultures of the transfectants in 75 cm culture dishes were incubated with [³⁵S]methionine (specific radioactivity >1000 Ci/mmole) at 170 μCi/ml in 1.5 ml of methionine-free DMEM/10% dialyzed fetal calf serum at 37 °C. After incubation for 7 hours the medium was chilled on ice, adjusted to 1 mM dithiothreitol, 1 mM phenymethylsulfonyl fluoride, 0.01% aprotinin (Sigma Chemie, Deisenhofen, FRG) and a tenfold excess of methionine to that normally present in DMEM, centrifuged for 1 minute at 12,000xg and stored at -70 °C. One third of the supernatant was incubated over night at 4 °C with 5 mg protein A-Sepharose (Pharmacia, Freiburg, FRG) saturated with the immunoglobulin fraction of a rabbit anti-human pregnancy-specific β1 glycoprotein (PSβ1G) or synonymously, "Schwangerschaftsprotein 1" (SP1) antiserum (DAKO, Copenhagen, Denmark) either in the presence or absence of 10 μg of purified SP1 (a gift from H. Bohn, Behringwerke, Marburg, FRG). The immunoprecipitates were washed three times with 0.5 ml of buffer J (10 mM Tris-Cl pH 7.5, 140 mM NaCl, 1 mM dithiothritol, 0.01% aprotinin, 0.1% SDS, 1% NP-40, 1% deoxycholate), once with 10 mM Tris-Cl pH 7.5 and subjected to electrophoresis on a 10% polyacrylamide gel containing 0.1% SDS (42). For fluorography, the gel was soaked with Enlightning (NEN Research Products, Dreieich, FRG) for 30 minutes, dried and exposed to X-ray film for 40 hrs.

RNA blot hybridization: Total RNA was extracted from fresh BeWo cells or in liquid nitrogen pulverized tissues by the guanidinium thiocyanate method (43) or alternatively, by a method according to Chomczynski and Sacchi (44). Poly(A)$^+$ RNA was isolated by one or in the case of colon tumors by two rounds of chromatography on oligo(dT) cellulose (Sigma Chemie, Deisenhofen, FRG) according to Aviv and Leder (45). Size fractionation of the RNAs on agarose gels in the presence of methylmercury hydroxide, transfer to charged nylon membranes and hybridization with [^{32}P]labeled cDNA fragments (46) was performed as described recently (30). When gene-specific [^{32}P]labeled oligonucleotides (specific radioactivity: 0.4-1.0x10^9 dpm/μg) were used as probes (Table I) the formamide concentration was lowered from 50% to 0-20%, depending on the calculated melting point of the hybrid (47), which was 10-15 °C above the hybridization temperature. After hybridization, the RNA blots were washed twice in 2xSSPE at room temperature for 10 minutes and once in 6xSSPE, 1% SDS for 60 minutes at a temperature where only hybrids with a maximum of one mismatch are stable.
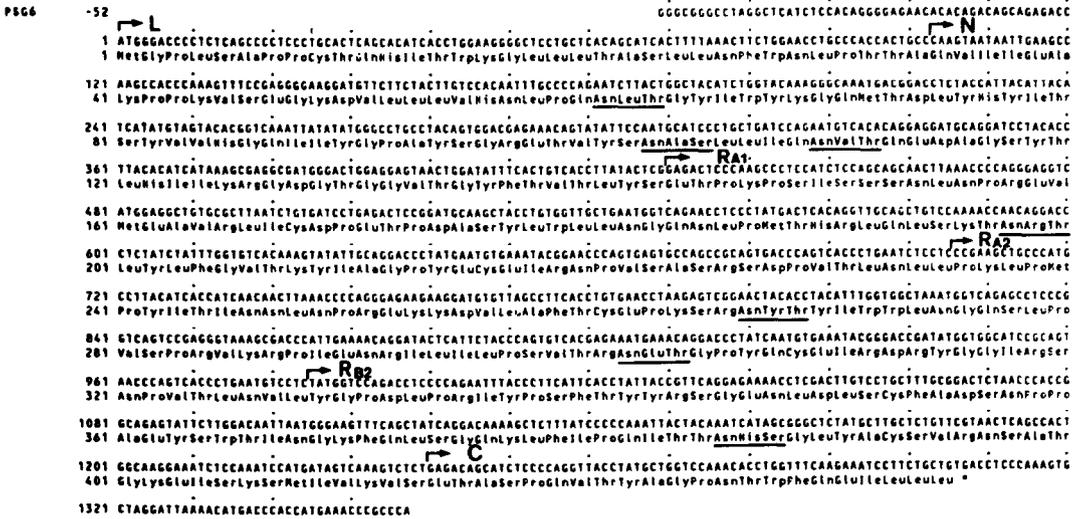
RESULTS

Isolation and analysis of pregnancy-specific glycoprotein cDNA clones. For the isolation of CEA/PSG-like cDNA clones from a fetal liver cDNA library, 10^5 plaques were hybridized with probes comprising coding regions of CEA and NCA cDNAs (for details see Materials and Methods). The applied hybridization conditions allowed cross-hybridization with members of all known subgroups of the CEA gene family, whose N-terminal domain exons reveal between 67 and 95% similarity at the nucleotide level (3). From the 150 positive clones seven were selected for further analysis. Digestion with restriction endonucleases and partial sequence analysis
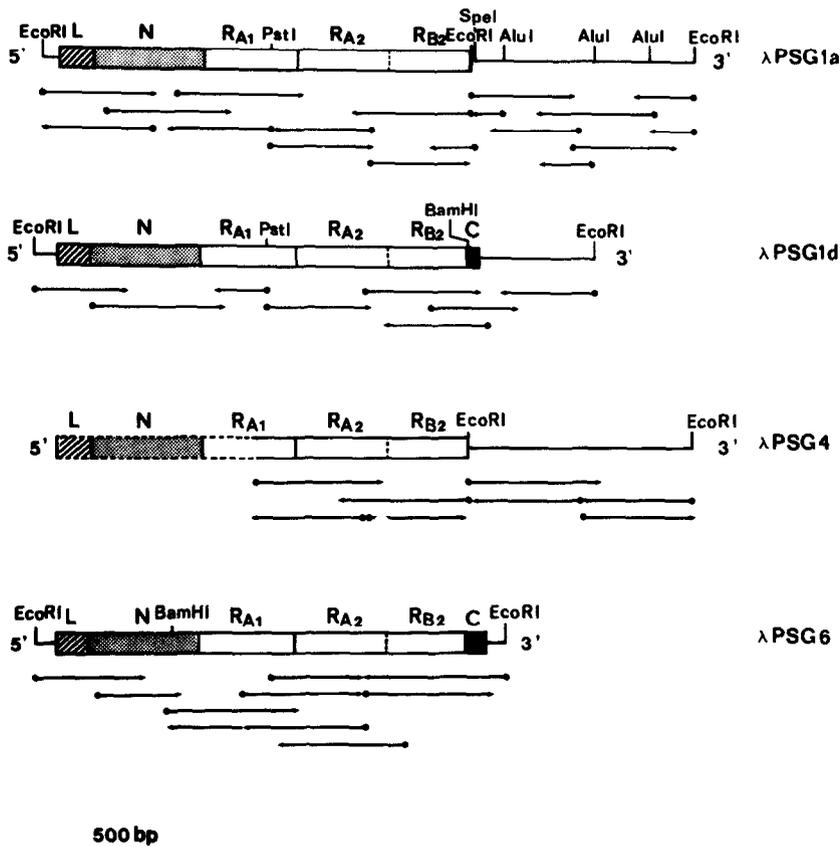
Figure 1: Domain structure and sequencing strategy of PSG1$_a$, PSG1$_d$, PSG4 and PSG6. The cDNAs were sequenced according to the strategy indicated by arrows. Only the AluI recognition sites relevant for sequencing are shown. The flanking EcoRI sites have been created by the cloning procedure. The coding regions can be subdivided into a leader [L], N-terminal or variable domain [N], halfrepeat or constant domains [R$_{A1,A2,B2}$] and C-terminal domain [C]. Lambda PSG4 is an incomplete clone and the dashed lines indicate the complete structure as derived from an overlapping cDNA clone published elsewhere (16).

indicated that these clones are derived from four different mRNAs. Four of these clones, lambda PSG1$_a$, lambda PSG1$_d$, lambda PSG4 and lambda PSG6 were completely sequenced according to the strategy shown in Fig. 1. The cDNA inserts of the clones lambda PSG1$_a$ and lambda PSG1$_d$ share a common sequence from their 5'-end downstream to position 1243 but differ completely beyond this point (Fig. 2). This indicates, that the corresponding mRNAs are derived from the same gene by differential splicing. On the other hand, the sequences of the cDNA inserts of lambda PSG4 and lambda PSG6, although very similar, differ from each other and from the PSG1 clones. They are, therefore, derived from two other closely related genes. All

PSG1$_a$

Figure 2: Nucleotide and deduced amino acid sequence of PSG1$_a$, PSG1$_d$, PSG4 and PSG6. The arrows demarcate the domain borders, the arrow head the start of the PSG1$_a$ cDNA. The recognition sequences for potential N-glycosylation sites are underlined, those for polyadenylation are boxed. The ArgGlyAsp (RGD) sequence in PSG6 starts at amino acid position 126.

Figure 2: Continued

clones except lambda PSG6 cover the 3'-ends of the corresponding mRNAs because of the presence of polyadenylation signals and an oligo(A) stretch in the case of lambda $PSG1_d$. The mRNAs represented by lambda $PSG1_a$ and lambda $PSG1_d$ encode preproteins of 419 (M_r 47,217; minus leader: 43,473) and 426 amino acids (M_r 47,860; minus leader: 44,116), respectively. They differ only in their C-terminal domains consisting of 5 and 12 amino acids, respectively (Fig. 2). Lambda PSG6 cDNA contains one long open reading frame of 435 amino acids (M_r 48,813; minus leader: 45,137), whereas the cDNA insert of lambda PSG4 is truncated at the 5'-end. In comparison to the domain structure of CEA and NCA (24) the coding regions of lambda $PSG1_{a,d}$ and lambda PSG6 are found to be composed of the following sequence of domains: a leader (34 amino acids), an N-terminal or immunoglobulin V-like domain of 110 and 109 amino acids, respectively, and three half repeats or immunoglobulin C-like domains (22), two of which belong to the A-type ($R_{A1,A2}$, 92 amino acids) and one to the B-type (R_{B2}, 85 amino acids) (16). The open reading frame of lambda PSG4 cDNA shows the same domain organization, however, the N-terminal domain and most of the R_{A1} domain is missing (Fig. 1). In lambda PSG4 and lambda PSG6 R_{B2} is followed by a short C-terminal domain with lengths of 5 and 22 amino acids, respectively. Except for lambda $PSG1_a$ and lambda PSG4 the C-terminal domains and 3'-untranslated regions seem to be unrelated (Fig. 2). The open reading frames of all four clones contain multiple, potential N-glycosylation sites (PSG1: 7; PSG4: 6; PSG6: 7), most at conserved positions, suggesting that the encoded proteins are highly glycosylated as is known for other members of the CEA family (for review see 24). In addition, the N-terminal domain of PSG6 contains the amino acid sequence RGD (Fig. 2), which is found in many extracellular matrix proteins, where it is known to convey cell adhesion (48). Comparison of the nucleotide sequences of the various domains of $PSG1_{a,d}$, PSG4 and PSG6 with the corresponding sequences of members of the CEA and PSG subgroups reveal degrees of similarity which clearly indicates that members of the CEA family belong to the PSG subgroup. Furthermore, $PSG1_{a,d}$, PSG4 and PSG6 correspond to genomic clones which have recently been partially sequenced (PSβG, hsCGM4 and hsCGM3, respectively [3]).

Table I: Gene- and splice variant-specific oligonucleotides. Oligonucleotides complementary to the corresponding mRNAs have been synthesized and purified as described before (2)

oligonucleotide	sequence 5'-3'	position	reference
PSG1	CTAGTCCCATCATCTCCCTTTATG	372	
[a]PSG1$_a$	GTCTTGCTCTTAGTGATTCCATGG	1292	
PSG16	CTTATAGGGCTTCAGAGACTTCGA	1298	26
PSG-C	CAATGGACCATGTAGGCTGTCTTC	1605	15
PSG4	GTGACTTGGGCAGTTGTGGGCGGA	490	[b]3
PSG6	AGTATAAGGTGACAGTGAAATATC	454	

[a] In the 3'-untranslated region of PSG4 the homologous region contains the same sequence except for one mismatch. Therefore, cross-hybridization of oligo PSG1$_a$ with PSG4 mRNA cannot be ruled out.
[b] Sequence was derived from clone hsCGM4.

Northern blot analyses. To characterize the sizes of the transcripts derived from the three PSG genes and to investigate whether these genes are active in tissues other than fetal liver, Northern blot experiments were performed. To be able to discriminate the mRNAs of these very closely related genes from each other as well as from the CEA subgroup, gene-specific oligonucleotides were used (Table I). Their sequences were selected from the N-terminal domains, where sequence information for most of the CEA gene family members is available. Hybridization with a [^{32}P]labeled oligonucleotide specific for PSG1 detected two transcripts of 2.0 and 2.2 kb in RNA from a 31 week human placenta (Fig. 3A,B; lane 1) whereas three transcripts of 1.8, 2.0 and 2.2 kb were found with a PSG4-specific oligonucleotide (Fig. 3B, lane 7). No transcripts could be identified for PSG6 in placenta, although the oligonucleotide used for this analysis hybridized in a control experiment with the corresponding genomic clone lambda hsCGM3 (3; data not shown). Although the presence of PSG in choriocarcinomas has been reported (for a review see [56]) none of the PSG subgroup members seems to be transcriptionally active in a
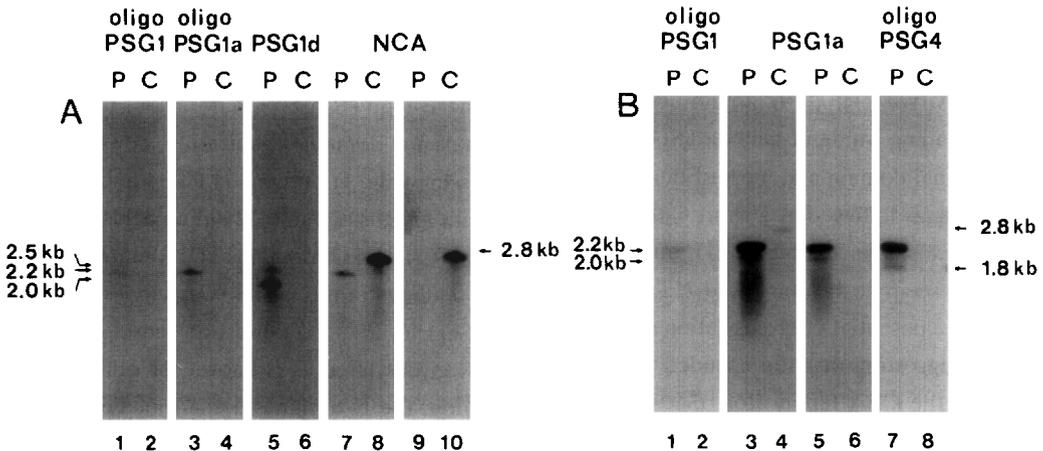
Figure 3: Analysis of PSG gene transcripts. 2 µg of poly(A)-containing RNA from a 31 week placenta (P) or a colonic tumor (C) were size-fractionated on a 1% agarose gel in the presence of methylmercury hydroxide, transfered to a nylon-membrane and hybridized with [^{32}P]labeled oligonucleotides (see Table I) specific for PSG1 transcripts (A,B lanes 1,2) or for the PSG1$_a$ splice variant (A lanes 3,4), for PSG4 mRNA (B lane 7,8) or with the BamHI-EcoRI cDNA fragment covering the 3'-noncoding region of PSG1$_d$ under low stringency conditions (wash in 2xSSPE, 60 °C; A lane 5,6). As a control a NCA-specific probe (the 1.4 kb EcoRI fragment of clone 9 containing the 3'-untranslated region [10]) was used to prove the intactness of the colon tumor RNA, which contains NCA mRNA. Cross-hybridization of the NCA and the PSG1$_a$ probe (the 670 bp EcoRI cDNA fragment covering the 3'-untranslated region) is observed with the 2.2 kb PSG1 and the 2.8 kb NCA transcript, respectively, under low stringency conditions (A lane 7; B lane 4), which is abolished after stringent washing (A lane 9; B lane 6).

choriocarcinoma cell line (BeWo), as demonstrated by hybridization under nonstringent conditions with the coding region of PSG1$_a$ (the 1.3 kb EcoRI fragment of lambda PSG1$_a$). In addition, neither of these PSG genes seems to be expressed in colon tumors (Fig. 3B, lanes 2, 8), where, however, transcripts of members of the CEA gene subgroup, namely NCA (Fig. 3A, lanes 8, 10) and CEA (30) are found. Inversely, the 2.8 kb NCA mRNA could not be detected in placental RNA Fig. 3A, lane 9). However, under nonstringent conditions with the probe covering the 3'-untranslated region of the NCA mRNA (10) a 2.2 kb mRNA species was visualized (Fig. 3A, lane 7). This cross-hybridization is probably due to the sequence similarity within the 3'-nontranslated region of the NCA and some of the PSG mRNAs (Fig. 5).

Together with PSG1$_a$ and PSG1$_d$, described here, four different splice variants of the PSG1 gene have been identified by cDNA cloning (15,16,19,26). In order to assign the 2.0 kb and 2.2 kb PSG1 mRNA species to individual splice variants, cDNA fragments from the 3'-untranslated regions of PSG1$_a$ and PSG1$_d$ and oligonucleotides specific for the splice variants PSG1$_a$ and the two others decribed before (PSG16 [26]; PSG-C [15]) were used as hybridization probes (Table I). The cDNA fragments from the 3'-untranslated region of PSG1$_a$ and the PSG1$_a$-specific oligonucleotide both hybridized most strongly with the 2.2 kb and weakly with a 2.0 kb mRNA species in placenta (Fig 3A, lane 3 and Fig. 3B, lanes 3, 5). The cDNA probe cross-hybridized only under nonstringent conditions with the 2.8 kb NCA mRNA in colon tumor RNA (compare in Fig. 3B, lanes 4 and 6). On the other hand, the cDNA probe covering the 3'-untranslated region of PSG1$_d$ hybridized strongly with a 2.0 kb and more weakly with a 2.5 kb mRNA species (Fig. 3A lane 5). The latter indicates that an additional mRNA species exists, which is possibly transcribed from another PSG gene containing a PSG1$_d$-like 3'-untranslated region. Therefore, the 2.2
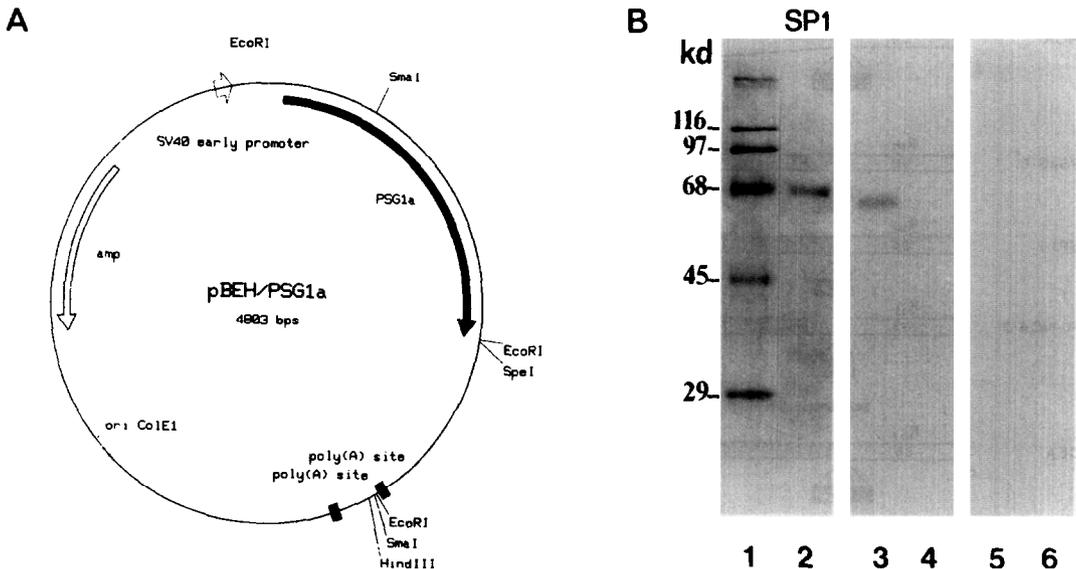
Figure 4: Expression of PSG1$_a$ cDNA in mouse 3T3 cells. (A) The complete cDNA of clone PSG1$_a$ has been inserted downstream of the SV40 early promoter into the EcoRI site of the eukaryotic expression vector pBEH. The coding region of PSG1$_a$ is indicated by a filled-in arrow. (B) The resulting plasmid pBEH/PSG1$_a$ was cotransfected with pSV2neo into 3T3 cells by lipofection. Stable G418-resistant transfectants were labeled with [^{35}S]methionine. The media from transfectants (lanes 3,4) or control transfectants without pBEH/PSG1$_a$ (lanes 5,6) were immunoprecipitated with rabbit anti-human SP1 in the absence (lanes 3,5) or presence of an excess of purified SP1 (=synonymous for PSβ1G). The precipitates were analysed by electrophoresis on a 10% SDS/polyacrylamide gel and autoradiographed. Proteins with the indicated sizes and SP1 (2 µg) were used as markers after staining with coomassie blue.

kb mRNA corresponds to the PSG1$_a$, the 2.0 mRNA to the PSG1$_d$ splice variant. No hybridization signal was obtained with an oligonucleotide specific for PSG16 and PSG-C, respectively. The former oligonucleotide spans the site, where PSG16 differs from PSG1$_a$ by insertion of a 86 bp sequence at the end of R$_{B2}$ in PSG1$_a$ (26).

Expression of PSG1$_a$ cDNA in eukaryotic cells. As an initial step toward studying the biochemical properties and function of individual mature members of the complex PSG family, we have started to express PSG1$_a$ cDNA in mouse NIH3T3 cells. This was achieved by cloning the complete cDNA into the expression vector pBEH (39) under the control of the Simian Virus 40 early promotor (Fig. 4A). The resulting plasmid pBEH/PSG1$_a$ was cotransfected into 3T3 cells
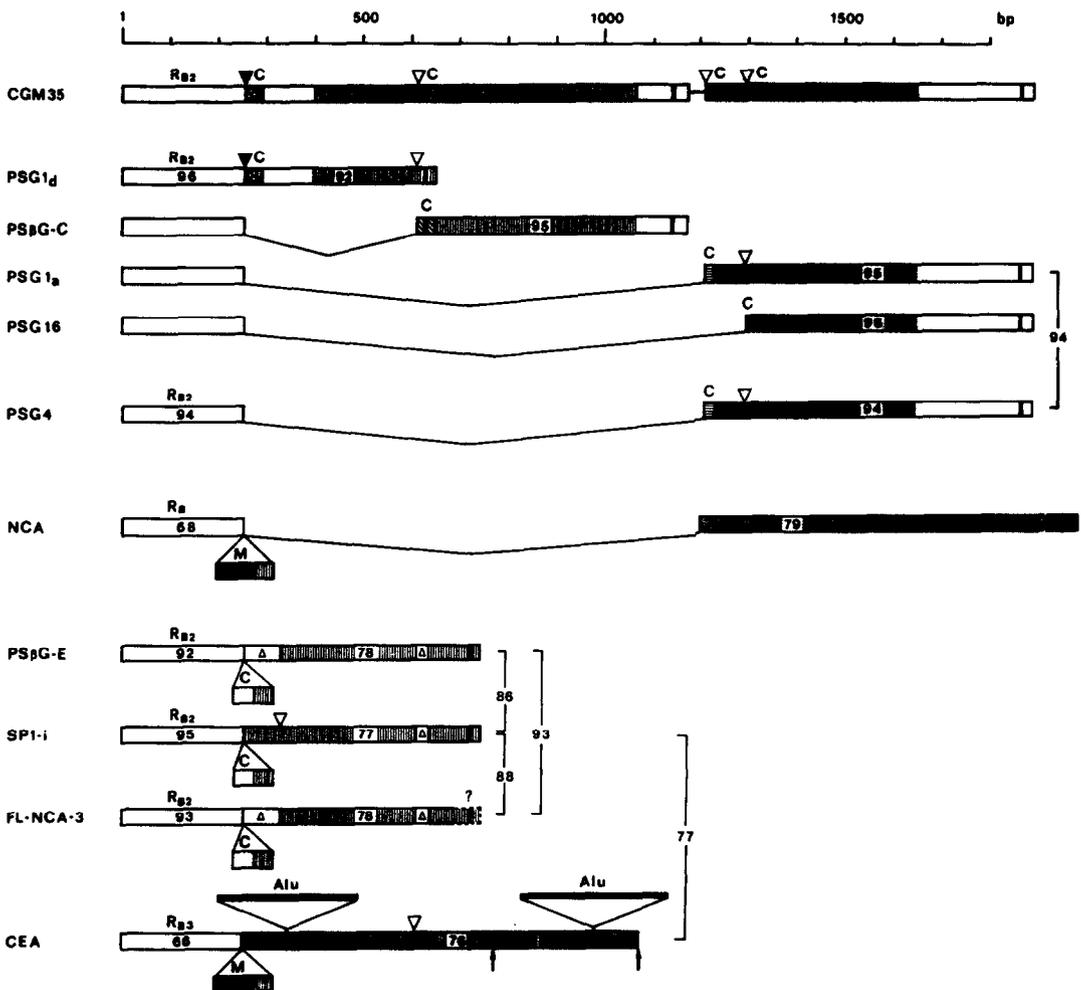
Figure 5: Comparison of the 3'-untranslated regions of PSGs, CEA and NCA cDNAs with the CGM35 gene. The nucleotide sequences compared begin with the last R$_B$ domain and continue to the 3'-end of each gene or cDNA. Putative splice donor (filled arrowheads) and splice acceptor sites (open arrowheads) are shown for the CGM35 gene (18). The single line adjoining the two blocks represents intron sequence. Homologous regions from mRNAs of other PSG/CEA-like genes, indicated by identical shading, are aligned as blocks below, whereby putative exon borders are taken into account. PSG1$_d$, PSβG-C, PSG1$_a$ and PSG16 represent splice variants of one gene. C = C-terminal domain, M = membrane domain, Alu = Alu-type repetitive element. Polyadenylation sites are indicated by black vertical bars (due to high sequence homology to PSβG-E, an identical polyadenylation site is assumed for FL-NCA-3), deletions larger than 20 nucleotides by triangles. Arrows demarcate the ends of the observed CEA mRNAs species formed by alternative polyadenylation. Numbers in blocks show degrees of nucleotide sequence similarity to corresponding regions of CGM35, numbers in brackets between individual cDNA species. The nucleotide sequences used for this comparison were taken from the following references: PSβG, PSβG-E (15), PSG16 (26), NCA (10), SP1-i (17), FL-NCA-3 (57), CEA (6,18).

together with pSV2neo, a plasmid which confers resistance toward the neomycin derivative G418. Pools of about 100 stable G418-resistant transfectants were labeled with [^{35}S]methionine. From the spent media of the cells transfected with pBEH/PSG1$_a$ a protein with an apparent M$_r$ of 66,000 could be precipitated with rabbit anti-human SP1 (Fig. 4B, lane 3), being slightly smaller than the major component of purified SP1 (lane 2). Addition of an excess of purified SP1 abolished the precipitation of this radioactively labeled protein (Fig. 4B, lane 4). No protein could be precipitated from media of control 3T3 cells transfected with pSV2neo alone (Fig. 4B, lanes 5, 6). Similar transfection experiments with African green monkey cos cells yielded a slightly smaller protein with an M$_r$ of 65,000 (data not shown).

DISCUSSION

Our findings obtained by cDNA cloning indicate that at least three different genes of the PSG subgroup are transcribed in liver during fetal development (PSG1, PSG4 and PSG6), whereby for PSG1 two splice variants are observed (PSG1$_a$ and PSG1$_d$). These results together with observations by Barnett et al. (13) who identified BGP I transcripts in human fetal liver suggest that also other members of the CEA gene family apart from the CEA gene play a role during fetal development. However, for confirmation and quantitation of transcription of the PSG genes Northern analyses with RNA from fetal liver have to be performed.

So far eight closely related PSG genes have been identified by isolation of genomic and cDNA clones (3,15-18,26,57 and results to be published elsewhere). cDNA clones with nearly identical nucleotide sequences as found in lambda PSG1$_a$ have been isolated recently from human placenta (PSG93 [26]; PSβG-D [15]; hPSP11 [16]), fetal liver (FL-NCA-2 [57]) or testis libraries (lambda hPS3 [49]). Lambda PSG1$_a$ differs from FL-NCA-2 whose sequence is identical with PSG93 and hPSP11 in 7 positions three of which are in the coding region. The latter differences all lead to amino acid substitutions. Comparison with the sequence of PSG93 reveals three different nucleotide exchanges which lead in one case to an amino acid change. These differences probably reflect genetic polymorphisms. A cDNA clone corresponding to our lambda PSG1$_d$ has also been recently described by Khan et al. (19). Furthermore, the sequence of a partial cDNA clone from a HeLa cDNA library has been reported which is obviously derived from the same PSG gene as lambda PSG4 (hHSP2 [16]). The sequences of hHSP2 and lambda PSG4 overlap for 138 bp. In this region they are identical except for the first 5 nucleotides of lambda PSG4 which is probably a cloning artefact. Together, they cover almost the complete coding sequence, lacking the first three amino acids. The deduced domain structures of the PSG antigens, where known, are very similar. The PSG1, PSG4, PSG6 and SP1 (17) proteins are composed of one variable or N-terminal domain (N) and three constant or half repeat domains (R$_{A1}$, R$_{A2}$, R$_{B2}$), whereas other PSGs (FL-NCA-3 [57], PSG-E [15]) lack the R$_{A1}$ or R$_{A2}$ domain, respectively. The R$_A$ and R$_B$ domains are encoded by separate exons as has been shown before for the NCA gene (50) and the PSG gene CGM35 (18). Two pairs of exons coding for half repeat domains (R$_{A1}$, R$_{B1}$ and R$_{A2}$, R$_{B2}$) are found in CGM35, whereby the splice acceptor signal at the R$_{B1}$ exon is mutated (18). This could explain the lack of R$_{B1}$ in the PSGs so far analysed or, in analogy, the absence of both the R$_{B1}$ and R$_{A2}$ domain in PSβG-E (15) or R$_{A1}$ and R$_{B1}$ in FL-NCA-3 (57). However, sequence analysis of a cosmid clone comprising the PSG4 gene reveals an intact splice acceptor consensus sequence for the R$_{B1}$ exon (results to be published elsewhere). Therefore,

inclusion or exclusion of $R_{A,B}$ domains by differential splicing, as is found for the BGP I gene (13), has to be considered.

The complexity of the protein family encoded by the PSG genes is increased by differential splicing. This is infered from the observation that the sequences of all PSG1 mRNAs diverge at the end of R_{B2} (Fig. 2; [15,16]). Obviously, the same R_{B2} exon is spliced to various exons which encode unrelated, 3-12 amino acid long C-terminal domains and 3'-untranslated regions (Fig. 5). These PSG1 C-terminal domains as well as the ones found in PSG4 (5 amino acids) and PSG6 (22 amino acids; Fig. 2) are very short and not particularly hydrophobic suggesting that these proteins are directly secreted and not membrane bound. This agrees well with the finding of large amounts of PSGs in maternal sera (28). However, presently it cannot be ruled out that the longer PSG C-terminal domains, especially that in PSG6, might serve as signals for glycosyl phosphatidylinositol membrane attachment as found for CEA (51-53) and NCA (54,55). In this case, the PSG6 protein which also contains the RGD sequence within its N-terminal domain could have a function in cell adhesion, as suggested for CEA (23).

The two splice variants of PSG1 which we have isolated from a fetal liver cDNA library, are also present in placenta, where two additional splice products have been found (15,26). PSG1$_a$ mRNA is readily detectable by hybridization with [^{32}P]labeled oligonucleotide (Fig. 3A), whereas PSG16 (26) and PSG-C (15) were not found under identical conditions. The PSG1$_a$ splice variant, therefore, seems to be one of the major transcripts of the PSG1 gene in placenta. A similar complex splicing pattern can be expected to exist for other genes. From the nucleotide sequences of the CGM35 gene, transcription products similar to PSG1$_d$, PSβG-C, PSG1$_a$ and PSG16 can be predicted ([18]; Fig. 5). In this gene, exons covering the various C-terminal domains and 3'-untranslated regions are clustered 3' of the R_{B2} exon. In fact, the region in CGM35 which encodes sequences homologous to the 3'-region of PSG1$_d$ is contiguous with the R_{B2} exon and partially overlaps the exon encoding the C-terminal domain and the 3'-untranslated region of the PSβG-C-like mRNA species (Fig. 5). Sequences homologous to the 3'-noncoding region of PSG1$_d$ are also present in transcripts of other CEA-like genes and might be expected to be present in CGM35 and PSG mRNAs (Fig. 5 and see below). This could explain the cross-hybridization of the PSG1$_d$ 3'-probe with a 2.5 kb mRNA (Fig. 3A). The observation of several mRNA species which hybridize with the PSG4-specific oligonucleotide also suggests differential splicing for the PSG4 gene (Fig. 3B), whereby the lambda PSG4 clone corresponds to the PSG1$_a$ splice variant. Due to the high degree of similarity between CGM35, PSG1$_a$ and PSG4 and the conservation of the splice acceptor signal for the generation of the PSG16-like splice variant (Fig. 5), the existence of mRNA species homologous to PSG1$_d$, PSβG-C and PSG16 can be predicted. No sequence similarity between the C-terminal domain and 3'-untranslated region of PSG6 and the corresponding region of the CGM35 gene was found. However, the homologous exon may be located further downstream of the known CGM35 genomic sequence.

It is interesting to note that sequences homologous to the 3'-regions of the PSG genes are also found in more distantly related genes of the CEA subgroup. For example the first 430 bp of the presumed 3'-region-exons of PSG1$_a$ and PSG4 show about 80% similarity to the first 430 bp of the 3'-untranslated region of NCA mRNA downstream of the point of complete divergence between CEA and NCA mRNA (10; Fig. 5). This explains the observed cross-hybridization under low stringency conditions between PSG1$_a$ and NCA mRNA (Fig. 3A lane 7 and Fig 3B

lane 4). Furthermore, a rather high similarity is found for the 3'-untranslated region of CEA and those of $PSG1_d$, PSβG-C, PSβG-E, SP1-i and FL-NCA-3 mRNA (Fig. 5). Closer sequence analyses revealed that the similarity with the CEA 3'-untranslated region stops abruptly in $PSG1_d$ and correspondingly in CGM35 upstream of position 1391. In the closely related 3'-noncoding regions of PSβG-E, SP1-i and FL-NCA-3 mRNAs, however the similarity continues beyond this point up to the stop codon of these PSG mRNAs and CEA mRNA (Fig. 5). In the CEA gene the hydrophobic membrane (M) domain together with 39 nucleotides of the 3'-untranslated region, are encoded in a separate exon (Fig. 5 and results to be published elsewhere). Based on these data, it can be hypothesized that the C-terminal domain and a small part of the 3'-untranslated region of PSβG-E, SP1-i and FL-NCA-3 are also encoded by separate exons. This idea is supported by the observation that a deletion of 67 nucleotides is found in PSβG-E and FL-NCA-3 mRNA ending exactly at the hypothetical C-terminal domain exon border (Fig. 5). This deletion is possibly caused by alternative splicing within the 3'-untranslated region exon. This explanation is supported by the presence of a splice acceptor consensus sequence in SP1-i mRNA at a position where the deletion ends in PSβG-E and FL-NCA-3 mRNA (Fig. 5). The putative C-terminal domain exon together with the surrounding intron sequences are obviously deleted in the CGM35 gene and in analogy in the PSG1 and PSG4 genes. This notion is stressed by the finding that in the gene corresponding to FL-NCA-3 which we have recently isolated, the sequences downstream of the R_{B2} exon share similarity with $PSG1_d$ and CGM35 exactly up to the position where the similarity with the CEA 3'-untranslated region begins (see Fig. 5; results to be published elsewhere). In conclusion, the similarities in the 3'-untranslated regions among the various members of the CEA gene family imply, that during evolution the exons containing the 3'-untranslated regions have been co-amplified together with the coding exons. Relatively late in evolution, i.e. after separation of the CEA and PSG subgroups, but before the appearance of the closely related PSG1, PSG4 and CGM35 genes, the deletion of the putative C-terminal exon and finally the insertion of the two Alu-like repetitive sequences into the CEA gene must have occurred.

Six of the eight PSG genes have been tested and shown to be transcriptionally active by Northern blot analyses and by cDNA cloning. Five of these genes are transcribed in placenta (Fig. 3 and unpublished results). Together with the proteins encoded by the PSG1 splice variants, at least eight very similar PSGs are expected to be synthesized in placenta. This complexity of PSG antigens is higher than anticipated from protein analyses, which identified three PSGs of 72, 64 and 54 kD (26). Because of the high degree of glycosylation of these antigens a direct assignment on the basis of their deduced primary sequence cannot be made. We, therefore, started to individually express PSG cDNAs in cells which do not express immunologically cross-reacting antigens. With a M_r of 66 kD, the $PSG1_a$ antigen was found to have a similar size to the major product of the placental PSG preparation (Fig. 4B). In similar experiments using the human HeLa cell line for transfection the size of the nearly identical FL-NCA-1 protein which corresponds to $PSG1_d$ was found to be 72 kD (19). As we expect many of the closely related PSG genes to code for antigens of similar size, these separately expressed, glycosylated antigens have to be further characterized. This might be achieved with monoclonal antibodies or, alternatively, with polyclonal antisera directed against peptide epitopes, specific for the various antigens. In conclusion, determination of the structure of all PSG genes and their splice variants,

together with assignment of the various genes to their respective proteins, will be a prerequisite to studying the function of individual PSGs.

REFERENCES

1. Kuroki, M., Arakawa, F., Yamamoto, H., Ikehara, Y., and Matsuoka, Y. (1988) Cancer Letters 43, 151-157.
2. Shively, J.E., and Beatty, J.D. (1985) CRC Critical Reviews in Oncology/Hematology 2, 355-399.
3. Thompson, J.A., Mauch, E.M., Chen, F.-S., Hinoda, Y., Schrewe, H., Berling, B., Barnert, S., von Kleist , S., Shively, J.E., and Zimmermann, W. (1989) Biochem. Biophys. Res. Commun. 158, 996-1004.
4. Zimmermann, W., Ortlieb, B., Friedrich, R., and von Kleist, S. (1987) Proc. Natl. Acad. Sci. USA 84, 2960-2964.
5. Oikawa, S., Nakazato, H., and Kosaki, G. (1987) Biochem. Biophys. Res. Commun. 142, 511-518.
6. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A., and Stanners, C.F. (1987) Mol. Cell. Biol. 7, 3221-3230.
7. 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.
8. 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.
9. Oikawa, S., Kosaki, G., and Nakazato, H. (1987) Biochem. Biophys. Res. Commun. 146, 464-469.
10. Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., and Riggs, A.D. (1987) J. Biol. Chem. 263, 3202-3207.
11. Tawaragi, Y., Oikawa, S., Matsuoka, Y., Kosaki, G., and Nakazato, H. (1988) Biochem. Biophys. Res. Commun. 150, 89-96.
12. Barnett, T., Goebel, S.J., Nothdurft, M.A., and Elting, J.J. (1988) Genomics 3, 59-66.
13. 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.
14. Hinoda, Y., Neumaier, M., Hefta, S.A., Drzeniek, Z., Wagener, C., Shively, L., Hefta, L.J.F., Shively, J.E., and Paxton, R.J. (1988) Proc. Natl. Acad. Sci. USA 85, 6959-6963. Correction: Proc. Natl. Acad. Sci. USA 86, 1668 (1989).
15. Streydio, C., Lacka, K., Swillens, S., and Vassart, G. (1988) Biochem. Biophys. Res. Commun. 154, 130-137.
16. Chan, W.-Y., Borjigin, J., Zheng, Q.-X., and Shupert, W.L. (1988) DNA 7, 545-555.
17. Rooney, B.C., Horne, C.H.W., and Hardman, N. (1988) Gene 71, 439-449.
18. Oikawa, S., Inuzuka, C., Kosaki, G., and Nakazato, H. (1988) Biochem. Biophys. Res. Commun. 156, 68-77.
19. Khan, W.N., Osterman, A., and Hammarström, S. (1989) Proc. Natl. Acad. Sci. USA 86, 3332-3336.
20. Paxton, Rj., Mooser, G., Pande, H., Lee, T.D., and Shively, J.E. (1987) Proc. Natl. Acad. Sci. USA 84, 920-924.
21. Oikawa, S., Imajo, S., Noguchi, T., Kosaki, G., and Nakazato, H. (1987) Biochem. Biophys. Res. Commun. 144, 634-642.
22. Williams, A.F., and Barclay, A.N. (1988) Ann. Rev. Immunol. 6, 381-405.
23. Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K., and Stanners, C.P. (1989) Cell 57, 327-334.
24. Thompson, J.A., and Zimmermann, W. (1988) Tumor Biol. 9, 63-83.
25. Bohn, H. (1972) Blut 24, 292-302.
26. Watanabe, S., and Chou, J.Y. (1988) J. Biol. Chem. 263, 2049-2054.
27. Horne, C.H.W., Towler, C.M., Pugh-Humphreys, R.G.P., Thompson, A.W., and Bohn, H. (1976) Experientia 32, 1197-1199.
28. Lin, T.-M., Halpert, S.P., and Spellacy, W.N. (1974) J. Clinic. Invest. 54, 576-582.
29. Tatarinov, Y.S. (1978) Gynecol. Obstet. Invest. 9, 65-97.
30. 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.
31. Cournoyer, D., Beauchemin, N., Boucher, D., Benchimol, S., Fuks, A., and Stanners, C.P. (1988) Cancer Res. 48, 3153-3157.
32. von Kleist, S., Winkler, J., Migule, I., and Böhm, N. (1986) Anticancer Res. 6, 1265-1272.
33. Wagener, C., Hain, F., Födisch, H.J., and Breuer, H. (1983) Histochemistry 78, 1-9.
34. Nap, M., Mollgard, K., Burtin, P., and Fleuren, G.J. (1988) Tumor Biol. 9, 145-153.
35. Fiebig, H.H., and Löhr, G.W. (1984) Med. Welt 35, 52-58.
36. Kodelja, V., Lucas, K., Barnert, S., von Kleist, S., Thompson, J.A., and Zimmermann, W. (1989) J. Biol. Chem. 264, 6906-6912.
37. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) In Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. .
38. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
39. Artelt, P., Morelle, C., Ausmeier, M., Fritzek, M., and Hauser, H. (1988) Gene 68, 213-219.
40. Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., Wenz, M., Northrop, J.P., Ringold, G.M., and Danielsen, M. (1987) Proc. Natl. Acad. Sci. USA 84, 7413-7417.
41. Southern, P.J., and Berg, P. (1982) J. Mol. Appl. Genet. 1, 327-341.
42. Laemmli, U.K. (1970) Nature 227, 680-685.
43. Fiddes, J.C., and Goodman, H.M. (1979) Nature 281, 351-356.
44. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156-159.

45. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
46. Feinberg, A.P., and and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.
47. Lathe, R. (1985) J. Mol. Biol. 183, 1-12.
48. Ruoslathi, E., and Pierschbacher, M.D. (1987) Science 238, 491-497.
49. Chan, W.Y., Tease, L.A., Borjigin, J., Chan, P.K., Rennert, O.M., Srinivasan, B., Shupert, W.L., and Cook, R.G. (1989) Human Reproduction 3, 677-685.
50. Thompson, J., Barnet, S., Berling, B., von Kleist, S., Kodelja, V., Lucas, K., Mauch, E.-M., Rudert, F., Schrewe, H., Weiss, M., and Zimmermann, W. (1989) In The CEA Gene Family (A. Yachi and K. Imai, eds.). Elsevier, Amsterdam (in press).
51. 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.
52. Takami, N., Misumi, Y., Kuroki, M., Matsuoka, Y., and Ikehara, Y. (1988) J. Biol. Chem. 263, 12716-12720.
53. Jean, F., Malapert, P., Rougon, G., and Barbet, J. (1988) Biochem. Biophys. Res. Commun. 155, 794-800.
54. Grunert, F., Kolbinger, F., Schwarz, K., Schweibold, H., and von Kleist, S. (1988) Biochem. Biophys. Res. Commun. 153, 1105-1115.
55. Kolbinger, F., Schwarz, K., Brombacher, F., von Kleist, S., and Grunert, F. (1989) Biochem. Biophys. Res. Commun. 161, 1126-1134.
56. Sorensen, S. (1984) Tumor Biol. 5, 275-302.
57. Khan, W.N., and Hammarström, S. (1989) Biochem. Biophys. Res. Commun. 161, 525-535.