Vol. 168, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 16, 1990 Pages 214-225

IDENTIFICATION OF A NEW CARCINOEMBRYONIC ANTIGEN (CEA) FAMILY MEMBER IN HUMAN FETAL LIVER - CLONING AND SEQUENCE DETERMINATION OF PREGNANCY- SPECIFIC GLYCOPROTEIN 7[1]

Wasif Noor Khan and Sten Hammarström

Department of Immunology, University of Umeå, S-901 85 Umeå, Sweden

Received March 5, 1990

The carcinoembryonic antigen gene family consists of the CEA- and the Pregnancy-Specific Glycoprotein- (PSG) subfamilies. Human fetal liver express several PSGs. Here we report cloning and sequencing of a new PSG subfamily member, PSG7. It is the fifth type of PSG found in fetal liver. PSG7 has the N-A1-A2-B2-C domain arrangement. Unlike other PSGs the N-terminal of PSG7 is unblocked. PSG7 has a cysteine in the C-terminal domain, which may allow dimerization. Variability analysis according to Wu and Kabat reveals that the region in the N-domain corresponding to complementarity determining region 3 of immunoglobulin is different between PSG subfamily members. Many members, including PSG7, contain the RGD sequence in this region. The CD2 region as well as two other short sequences (in N and A1 domains respectively) also show some variability. The function of PSGs is probably linked to the N-domain and the CDR2- and CD3-like regions are most likely responsible for ligand binding.

Pregnancy-specific β1 glycoprotein (PSG) [synonym: schwangerschafts-spezifisches β1 glykoprotein (SP1)] (1,2) is a group of human placental glycoproteins synthesized by the syncytiotrophoblast (3). PSG is detected in maternal serum during pregnancy and it increases in concentration as pregnancy progresses (2,4). Low levels of PSG (or PSG like) proteins are also found in serum from non-pregnant women (5). Elevated levels of PSG are seen in sera from patients with trophoblastic and some other types of tumors (for references see 6). PSG is not placenta

[1] In this communication we use the nomenclature for CEA-gene family members adopted at the XVIIth Meeting of the International Society for Oncodevelopmental Biology and Medicine in Freiburg, Western Germany 1989.

Abbreviations: PSG, pregnancy specific β1 glycoprotein; CEA, carcinoembryonic antigen; NCA, nonspecific cross-reactive antigen; FL-NCA, fetal liver NCA, BGP, biliary glycoprotein; CDR, complementarity determining region.

0006-291X/90 $1.50

specific. Thus, cultured human skin fibroblasts (7,8) and normal granulocytes (9) produce PSG. Moreover, cDNAs coding for PSG is not only found in placenta (10,12,13) but also in fetal liver (14-16), HeLa cells (17), testis (17) and submandibular salivary gland (18).

Sequence analysis of PSGs led to the discovery that these proteins are related to carcinoembryonic antigen (CEA) (11) and that they together with CEA (19,20), nonspecific cross-reacting antigen of 55 and 95 kd molecular weight (NCA-55/95) (21-23) and biliary glycoprotein (BGP) (24-26) comprise a subfamily within the immunoglobulin gene superfamily (27). Members of this gene family all contain a characteristic N-terminal immunoglobulin V region-like domain that lacks the two cysteines forming the intrachain disulphide bond. The variable region is followed by between 2 and 6 immunoglobulin C2 region-like domains (14). The C-terminus of the glycoproteins within the family differs between members. CEA and NCA 55/95 are anchored to the plasma membrane via phosphatidylinositol (28-31) while BGP has a regular transmembrane region followed by a cytoplasmic domain (26). In contrast, most, if not all PSG's appear to be "direct" secretion products (see below).

Recently CEA and NCA 55/95 have been shown to function as homotypic intercellular adhesion molecules (32,33). The function(s) of the pregnancy associated glycoproteins is(are) unknown.

We have recently cloned and sequenced cDNAs for three different PSGs from a human fetal liver library (14,15). Here we report the identification and sequencing of a new CEA gene family member belonging to the PSG group.

MATERIALS AND METHODS

Isolation and characterization of cDNAs

A first trimester human fetal liver cDNA library in phage vector λgt11 (Clontech) was screened with two oligodeoxynucleotide probes (14) which were derived from the N-terminal amino acid sequence of CEA. Positive cDNA clones and their restriction endonuclease fragments were subcloned into plasmid vector pUC19 or in the M13 based phage vector, phagescript (Stratagene). cDNA in phagescript was used for generating nested deletions. The nucleotide sequence of double stranded or single stranded DNA was determined by the dideoxy chain termination method (34).

RESULTS AND DISCUSSION

Identification and sequencing of cDNA clones from a human fetal liver library coding for PSG7. Approximately 200 positive clones were identified when 10^5 recombinant phage plaques were screened with two oligonucleotide probes corresponding to the N-terminal region of CEA (14). The sequence of three different cDNA clones, PSG1d (FL-NCA-1), PSG1a (FL-NCA-2) and PSG5 (FL-NCA-3) have been reported (14,15). Here we report the sequence of a fourth cDNA (FL-NCA-4) from fetal liver. It is named PSG7.

The complete sequence was obtained from two overlapping clones, A and B (Figure 1). The nucleotide sequence and the deduced amino acid sequence is shown in Figure 2. PSG7 consists of 1657 base-pairs of which 1278 nucleotides code for a polypeptide containing 426 amino acids. The 3' untranslated region contains the poly(A) signal. PSG7 has the L-N-A1-A2-B2-C domain arrangement found in most members of the PSG-subfamily (eg. PSG1, 3, 4 and 6). The N-

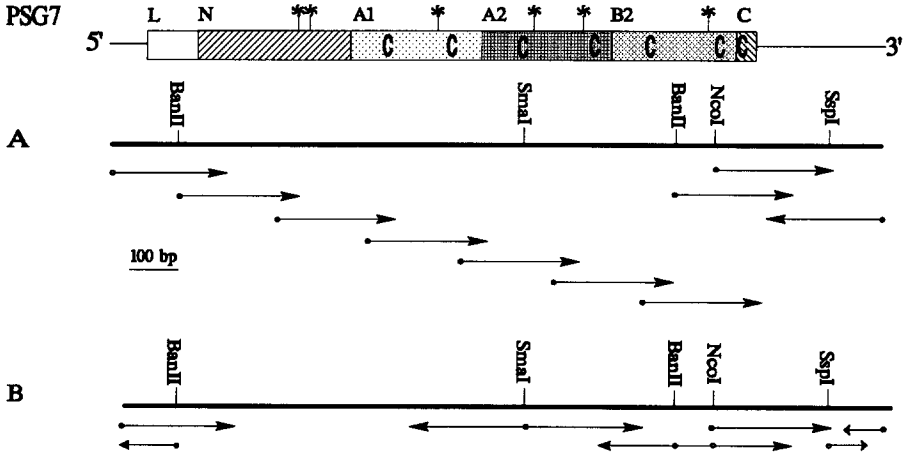

Figure 1. Alignment of two cDNA clones (A and B) and structural features of the PSG7 protein encoded by these cDNAs. Block diagram shows the domain structure, "C" indicates the positions of cysteines and stars the potential N-linked glycosylation sites. Clone A is 12 nucleotides longer than clone B at the 5' end and 10 nuclotides shorter than B at the 3' end. Arrows show the extent of DNA sequence determined.

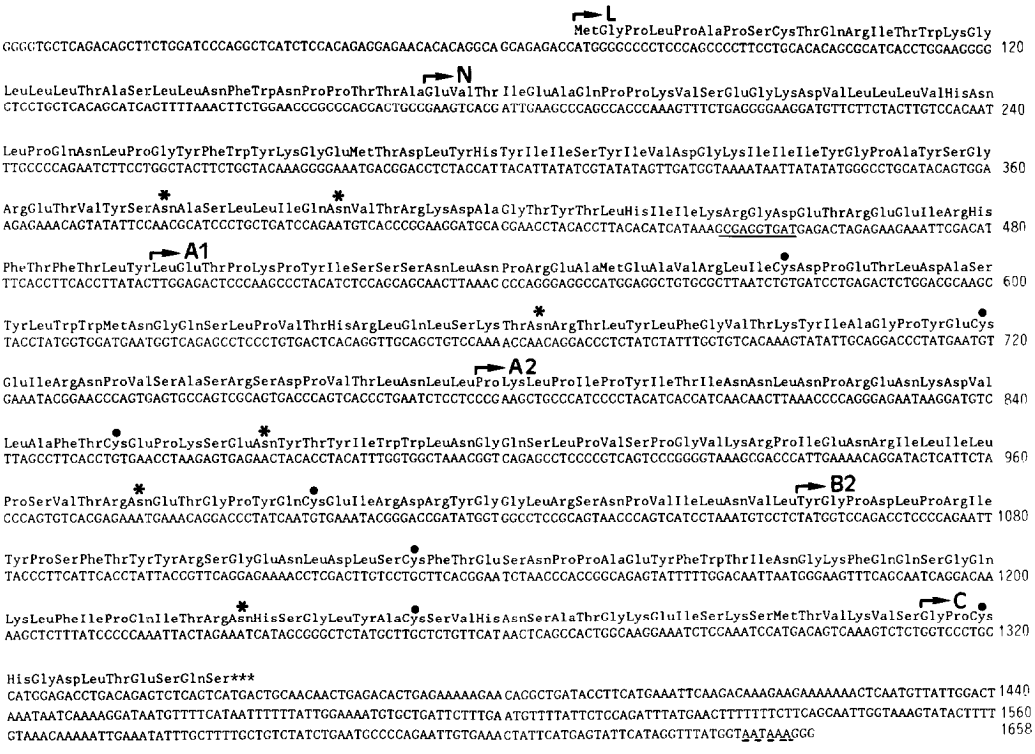

Figure 2. Nucleotide sequence and deduced amino acid sequence of PSG7. Domain boundaries are marked with arrows. Putative glycosylation sites are marked with stars. Cysteines are marked with dots, the RGD sequence is underlined and the poly (A) signal is underlined with a broken line.

domain lacks cysteines, while the three internal domains (A1, A2 and B2) each have two cysteines most probably forming an intrachain disulphide bridge in each of the three immunoglobulin C2-like domains. Interestingly, PSG7 contains an extra cysteine located in the C-terminal tail region (Figure 1, 2 and 5). This cysteine may allow PSG7 to form a dimer. The calculated molecular mass of the peptide moiety of PSG7 is 44,592 daltons.

PSG7 contains six putative N-glycosylation sites (Figure 1 and 2), five of which are located at the same positions as in the other PSG subfamily members. The sixth putative N-glycosylation site is found in the B2 domain close to the C-terminus, the only other PSG which has a glycosylation site at this position is

PSG6. The sequence identity at the amino acid and nucleotide levels between PSG7 and other members of the PSG-subfamily is 82-89 % and 93-95 % respectively, when corresponding domains are compared. PSG7 has an unblocked N-terminus in contrast to all other members of the PSG subfamily.

The amino acid sequence of PSG7 differs markedly from the sequence of other PSGs at positions 93-103 in the N-domain. This region, which corresponds to the complementarity-determining region 3 of immunoglobulin, contains the arginine-glycine-aspartic acid (R-G-D) sequence followed by a sequence of charged amino acids (6 out of 8 amino acids are charged). Structure predictions according to Chou and Fasman (35) indicate that this highly hydrophilic region forms an alpha-helix (Figure 4).

The PSG-subfamily. Southern blot analysis with a PSG subgroup specific N-domain probe indicate that the PSG subfamily contains 13 members or a multiple thereof (15). Presently, 8 different PSG subfamily members have been identified and sequenced (Figure 3). PSG1 occurs in four different alternatively spliced forms giving rise to different C-terminal domains (a-d) (Figure; references 5,10-12,14-16). PSG8, for which only the genomic sequence is known, may also occur in four forms with different C-terminal domains in analogy to PSG1, since it contains the corresponding exons (36). Three different domain arrangements are seen within the PSG-subfamily: N-A1-A2-B2-C (PSG1,3,4,6,7 and presumably 8); N-A1-B2-C (PSG2) and N-A2-B2-C (PSG5).

It is likely that most, if not all, PSGs are direct secretion products since they are found in serum and have short C-terminal domains. However, the possibility that some of them (eg. PSG6, PSG5, PSG1d, and PSG8d) are bound to the membrane via a

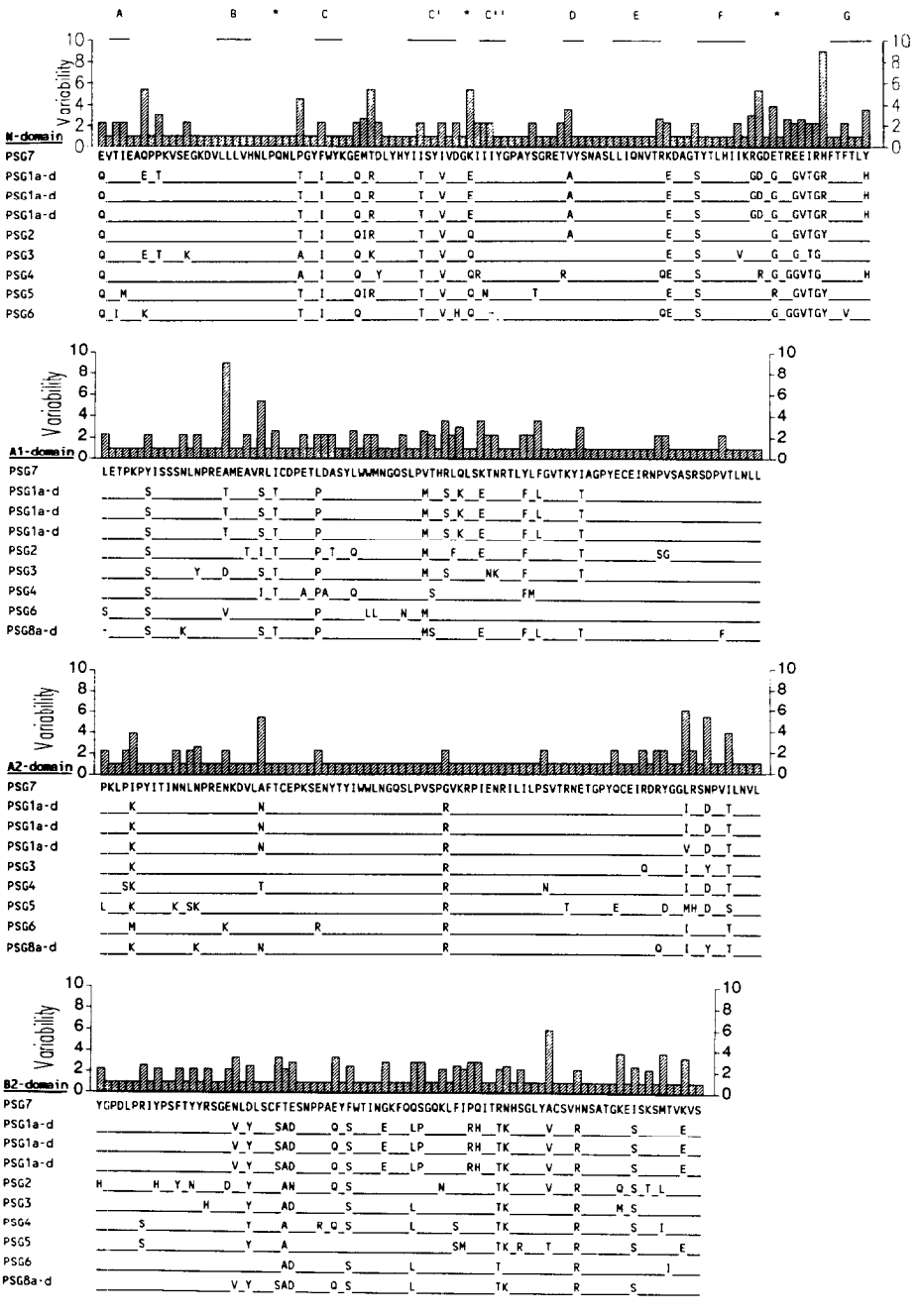

Figure 3. Amino acid sequence (standard one letter symbols) comparison and schematic representation of variability of PSG1 to 8. β strands (A-G) in N-domain are indicated by lines above the sequence. The positions of the hypothetical complementarity determining regions (CDR1-3) are indicated by stars. Block diagram shows variability = number of different amino acids occuring at a given position divided by frequency of the most common amino acid at that position (ref 37,38). PSG1 [PSG 16/93 (10,11); PSβG C/D (12); FL-NCA-1/2 (14,15); λPSG1 a/d (16); hPSP 11 (17], PSG2 [PSβG E (12)], PSG3 [pSP1-i (13); SG5 (18)], PSG4 [hHSP2 (17); hsCGM4 (43); λPSG4 (16)], PSG5 [FL-NCA-3 (15)], PSG6 [λPSG6 (16)], PSG7 [this communication], PSG8 [CGM35 (36)].

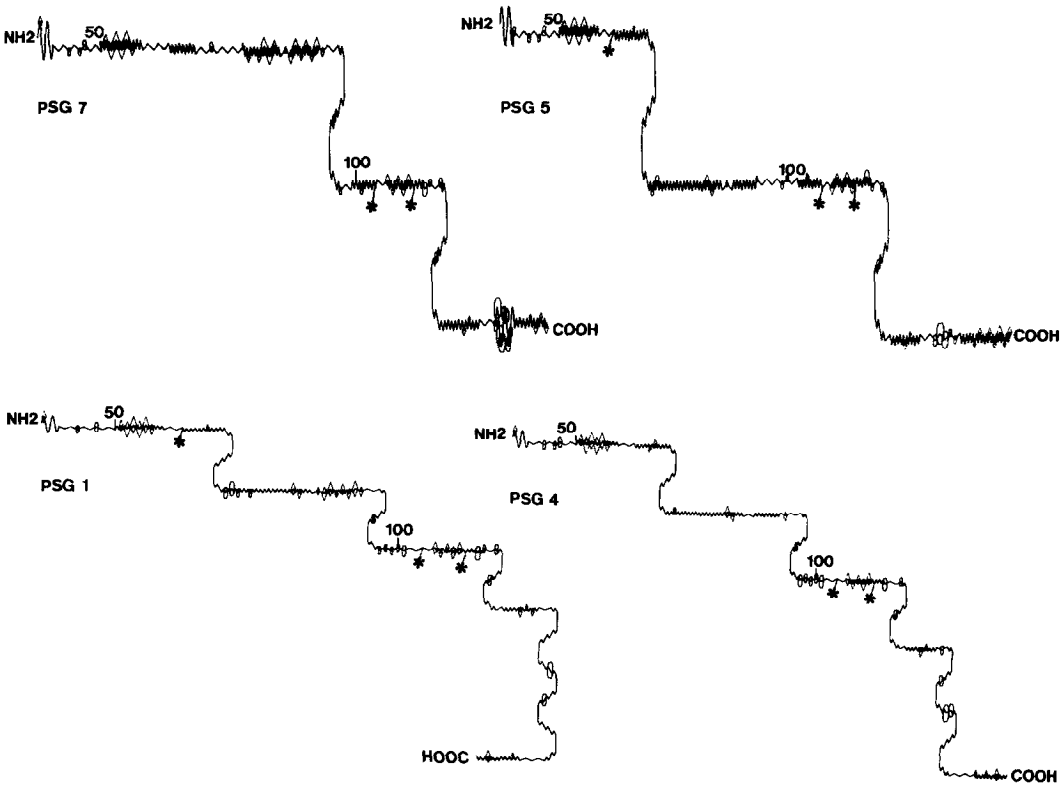

Figure 4. Secondary structure predictions and hydropathy of PSG N-domains. PSG2, 3 and 6 gave a similar plot to PSG1. The secondary structure predictions and hydropathy were obtained by computer assisted analysis according to Chou-Fasman (35) and Kyte-Doolittle (44) respectively. Numbering of amino acids starts at leader. Helices are shown as a sinus wave, beta sheets as a sharp saw tooth-wave, turns with 180 degree turns and coils with dull saw-tooth wave. Hydropathy is superimposed on secondary structure plot where (▯) represents hydrophilicity and (◊) hydrophobicity.

phosphatidylinositol linkage cannot be completely excluded since they have slightly more hydrophobic and/or longer C-terminals than the other (Figure 5). However, transfection experiments demonstrated that all PSG1a (16) and PSG1d (15) were secreted into the culture medium. In contrast, about 25 % of PSG5 (15) was retained on the cells.

The amino acid sequences of different PSG subfamily members are very similar to each other (Figure 3). Variability analysis

		C-Terminal domain
A	PSG1c	AYSSSINYTSGNRN
	PSG2	ASTRIGLLPLLNPT
	PSG3	APSGTGHLPGLNPL
	PSG5	APSGIGRLPLLNPI
	PSG8b	AYSSSINYTAVY
B	PSG1a	DWTVP
	PSG4	DWILP
	PSG8c	DWTLP
C	PSG1b	EAL
	PSG8a	EAL
D	PSG1d	GKWIPASLAIGF
	PSG8d	GKRIPVSLAIGI
E	PSG6	ETASPQVTYAGPNTWFQEILLL
F	PSG7	GPCHGDLTESQS

Figure 5. Carboxyl-terminal domain of PSG1 to 8. A to G show grouped C-terminal domains according to homology. Glycosylation sites are underlined and cystein is marked with a dot.

according to Wu and Kabat (37, 38) revealed, however, that there is one region in the N-domain, position 93-103, which shows considerable sequence variability. A lower degree of sequence variability is also found in three other regions. They are located at positions 37-40, 51-56 and 127-132 (A1 domain). Much of the other variability is non clustered and of a conservative nature (Figure 3). If the PSG domains have the immunoglobulin fold, as predicted from sequence comparison, the regions with high variability corresponds to the loop connecting β-strands C and C' (residues 37-40), CDR2 (residues 51-56) and CDR3 (residues 93-103) in the N-domain and the loop connecting β-strands A and B in the A1 domain. Interestingly, the CEA, NCA-55/95, BGP subfamily have entirely different sequences in these regions in comparison to PSGs, while remaining highly similar between themselves (Figure 6). Since most of the variability between the PSG subfamily members was found in the N-domain, we applied the Chou-Fasman algorithm (35) to predict the structure of the N-domain (Figure 4). The N-domains could be classified into four groups which differ from each other in terms of predicted structure. The N-domain structures of PSG2, 3 and 6 were similar to that of PSG1. The CD3 region is generally hydrophilic.

N-domain

PSG-group .V.αEA.P.KVS.GKDVLLLVHNLPQNL.GY.WYKG.α..LYHYI.SYαV.G....YGPAYδGRE..YSNASLLIQNVT..DAGδYTLHIαK..D.T.....FT.TL.
Consensus .α..E..P..V..GKεVLLL.H.LPQ...GY.WYK.....I..Y.α.....GPA.δGRE.αY.NASLLIQNα...D.G.YTL.ααK.....F.....
CEA-group .LT.ES.PFNVAEGKEVLLL.H.LPQ...GYSWYK.ERVDGN..IVGY.IGTQQATPGPA.SGRE.IYPNASLLIQNα.Q.DTGFYTL.VIKSDLVNEEATGQF.VY

A1-domain

PSG-group .ETPKP.ISSS.L.PβE.ME.V.L.CDP.T...SY.W.αNGQ.LPαδH...LS...RTL.α.GVTKY.AGPYECEIRN..SASRSDP.TLNLL
Consensus .E.PKP.ISS....P.E..ε.V...CεP.....δY.W.αN.Q.LPαδ..L.LS....TL.α..V.β.....Y.CE..N..SA.RSD...LNα.
CEA-group PELPKPSISSNNS.PVEDKDAVAFTCEPE.Q..TYLWWαN.QSLPVSPRLQLSNGN.TLTLL.V.RND...Y.CE.QNP.SA.RSD.V.LNV.

Figure 6. Amino acid comparison of consensus sequences of PSG-group (top line) and CEA-group (bottom line) of the CEA-gene family. The middle line shows the consensus sequence of the CEA-gene family. Conserved substitutions are shown by greek letters where α = A,I,L,V,M; β = H,K,R; δ = S,T; ϵ=D,E.

Three different PSG1a-d sequences have been identified: FL-NCA-1 and 2 (14,15) = PSβG-C/D (12) = hPSP11 (17); PSG 16/93 (10); PSG1 a/d (16). They differ from each other at a few positions in the nucleotide sequence, three of which lead to amino acid substitutions, eg. positions 7, 9 and 319 (Figure 3). Although it cannot be excluded that separate genes encode the three forms, it is perhaps more likely that they represent genetic polymorphism. The biological function(s) of the PSGs is (are) unknown. The findings that PSGs are synthesized in large amounts by syncytiotrophoblasts (1-4, 10) and fetal liver (14-16) and that they are found in the maternal circulation (1, 2, 4, 10) during normal pregnancy indicate that they have a function during pregnancy. Earlier studies by Bohn and associates moreover indicate that antibodies against human PSG induce abortion if injected into pregnant monkeys (39). How PSG exerts this effect(s) is, however, not known. It is possible that PSG interfers with the binding of cells to each other or with the attachment of cells to the extracellular matrix. The finding that the RGD sequence (40) is present in 5 out of 7 N-domain sequences suggests such a function. It is interesting to note that the sequence immediately following the RGD (=CDR3 region) differs substantially between members of the PSG subfamily. The different

family members may therefore have different binding specificities. The finding that most PSGs are secretion products does not mitigate such a function since the molecules may either act as extracellular matrixes or as inhibitors of cell attachment (scatter factors, compare reference 41). Presently it can, however, not be excluded that the PSGs is a family of immunomodulatory molecules. Recently it was shown that rat BGP contains the consensus sequence of the substrate binding site of ecto ATP-ase (42). All CEA family members have half of that site [amino acid residues 92-100; G-(PAYS)-GR]. The significance of this finding is unclear.

Taken together the data indicate that the function of the PSG subfamily is linked to the common N-domain. It seems likely that the CDR2- and CD3-like regions in this domain is directly involved in binding to the hypothetical ligand.

ACKNOWLEDGMENTS

We thank Ms Anne Israelsson, Ms Elisabeth Granström and Marie Svensson for excellent technical assistance and Ms Christine Boström for excellent secreterial assistance. This work was supported by a grant from the Swedish Cancer Society.

REFERENCES

1. Tatarinov, Y.S., and Masyukevich, V.N. (1970) Byull. Eksp. Biol. Med. USSR 69, 66-68.
2. Bohn, H. (1971) Arch. Gynaekol. 210, 440-457.
3. Horne, C.H.W., Towler, C.M., Pugh-Humphreys, R.G.P., Thompson, A.W., and Bohn, H. (1976) Experientia 32, 1197-1199.
4. Lin, T.M., Halbert, S.P., and Kiefer, D. (1976) J. Clin. Invest. 57, 466-472.
5. Mueller, U.W., and Jones, W.R., (1985) J. Reprod. Immunol. 8, 111-120.
6. Sorensen, S. (1984) Tumor Biol. 5, 275-302.
7. Rosen, S.W., Kaminska, J., Calvert, I.S., and Aaronson, S.A. (1979) Am. J. Obstet. Gynecol. 134, 734-738.
8. Chou, J.Y. (1983) Oncodevelop. Biol. Med. 4, 319-326.
9. Heikinheimo, M., Gahmberg, C.G., Bohn, H., and Andersson, L.C. (1987) Blood 70, 1279, 1283.
10. Watanabe, S., and Chou, J. (1988) J. Biol. Chem. 263, 2049-2054.

11. Watanabe, S., and Chou, J. (1988) Biochem. Biophys. Res. Commun. 152, 762-768.
12. Streydio, C., Lacka, K., Swillens, S. and Vassart, G. (1988) Biochem. Biophys. Res. Commun. 154, 130-137.
13. Rooney, B., Wilson Horn, C., and Hardman, N. (1988) Gene 71, 439-449.
14. Khan, W.N., Osterman, A., and Hammarström, S. (1989) Proc. Natl. Acad. Sci. USA 86, 3332-3336.
15. Khan, W.N., and Hammarström, S. (1989) Biochem. Biophys. Res. Commun. 86, 525-535.
16. Zimmermann, W., Weiss, M., and Thompson, J.A. (1989) Biochem. Biophys. Res. Commun. 163, 1197-1209.
17. Chan, W.-Y., Borjigin, J., Zheng, Q.X. and Shupert, W.L. (1988) DNA 7, 545-555.
18. Zoubir, F., Khan, W.N., and Hammarström, S. (1990) (submitted)
19. Gold, P., and Freedman, S.O. (1965) J. Exp. Med. 121, 439-462.
20. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A., and Stanner, S. (1987) Mol. Cell. Biol. 7, 3221-3230.
21. von Kleist, S., Chavanel, G., and Burtin, P. (1972) Proc. Natl. Acad. Sci. USA 69, 2492-2494.
22. Mach, J.-P., and Pusztaszeri, G. (1972) Immunochemistry 9, 1031-1033.
23. Neumaier, M., Zimmerman, W., Shively, L., Hinoda, Y., Riggs, A., and Shively, J.E. (1988) J. Biol. Chem. 263, 3202-3207.
24. Svenberg, T., Hammarström, S., and Hedin, A. (1979) Mol. Immunol. 16, 245-252.
25. Hinoda, Y., Neumaier, M., Hefta, S., Drzeniek, Z., Wagener, C., Shively, L., Hefta, L., Shively, J., and Paxton, R. (1988) Proc. Natl. Acad. Sci. USA 85, 6959-6963.
26. 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.
27. Paxton, R., Mooser, G., Pande, H., Lee, T., and Shively, J. (1987) Proc. Natl. Acad. Sci. USA 84, 920-924.
28. Hefta, S., Hefta, L., Lee, T., Paxton, R., and Shively, J. (1988) Proc. Natl. Acad. Sci. USA 85, 4648-4652.
29. Takami, N., Misumi, Y., Kuroki, M., Matsuoka, Y., and Ihehara, Y. (1988) J. Biol. Chem. 263, 12716-12720.
30. Grunert, F., Kolbinger, F., Schwarz, K., Schwaibold, H., and von Kleist, S. (1988) Biochem. Biophys. Res. Commun. 153, 1105-1115.
31. Kolbinger, F., Schwarz, K., Brombacher, F., von Kleist, S., and Grunert, F. (1989) Biochem. Biophys. Res. Commun. 161, 1126-1134.
32. Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K., and Stanner, C. (1989) Cell 57, 327-334.
33. Oikawa, S., Inuzuka, C., Kuroki, M., Matsuoka, Y., Kosaki, G., and Nakazoto, H. (1989) Biochem. Biophys. Res. Commun. 164, 39-45.
34. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
35. Chou, P.Y., and Fasman, G.D. (1978) Adv. Enz. 47, 45-147.
36. Oikawa, S., Inuzuka, C., Kosaki, G., and Nakazato, H. (1988) Biochem. Biophys. Res. Commun. 156, 68-77.
37. Wu, T.T., and Kabat, E.A. (1970) J. Exp. Med. 132, 211-250.
38. Kabat, E.A., Wu, T., Reid-Miller, M., Perry, H., and Gottesman, K. (1987) Sequences of Proteins of Immunological Interest (US Dept och Health and Human Services, Washington, DC), 4th Ed.
39. Bohn, H., and Weinmann, E. (1974) Arch. Gynäck. 217, 209-218.

40. Ruoslahti, E., and Pierschbacher, M. (1987) Science 238, 491-497.
41. Gherardi, E., Gray, J., Stoker, M., Perryman, M., and Furlong, R. (1989) Proc. Natl. Acad. Sci. USA 86, 5844-5848.
42. Lin, S.-H., and Guidotti, G. (1989) J. Biol. chem. 264, 14408-14414.
43. Thompson, J., Mauch, E.-M., Chen, F.-S., Hinoda, Y., Schrewa, H., Berting, B., Barnert, S., von Kleist, S., Shively, J.E., and Zimmermann, W. (1989) Biochem. Biophys, Res. Commun. 158, 996-1004.
44. Kyte, J., and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.