Vol. 169, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 31, 1990 Pages 203-216

CARCINOEMBRYONIC ANTIGEN GENE FAMILY MEMBERS IN SUBMANDIBULAR SALIVARY GLAND: DEMONSTRATION OF PREGNANCY-SPECIFIC GLYCOPROTEINS BY cDNA CLONING

Fairouz Zoubir, Wasif Noor Khan, and Sten Hammarström

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

Received April 12, 1990

We have recently shown that human submandibular salivary gland and saliva contain a number of glycoproteins belonging to the carcinoembryonic antigen (CEA) gene family. The members of the CEA family can be divided into the CEA subgroup and the pregnancy specific β1 glycoprotein (PSG) subgroup. The latter glycoproteins are abundant in placenta and fetal liver. Here we report that PSG's are expressed in normal adult submandibular salivary gland. Thus, cDNA cloning and sequencing gave two clones (SG5 and SG9) which coded for glycoproteins with a domain arrangement of N-A1-A2-B2-C and a third clone (SG8) which coded for a glycoprotein with a domain arrangement of N-A1-B2-C. SG5 is identical to PSG3, and SG9 to PSG1d, while SG8 most probably corresponds to PSG2. The 3' untranslated regions of the different members of the PSG subgroup contain highly homologous segments, suggesting a common evolutionary origin.

Carcinoembryonic antigen (CEA) (1) of the human digestive tract is a widely used tumor marker. Determination of the circulating levels of CEA is of proven value in monitoring patients with colorectal cancer after surgical removal of tumor. CEA is a promising target antigen for radioimmunotherapy of colorectal carcinomas. However, CEA belongs to a family of closely related macromolecules with wide distribution in cancerous and normal tissues and immunological cross-reactivity of anti-CEA antibodies with related antigens reduces the clinical utility of these antibodies. Efforts have been focused therefore on the molecular

Abbreviations: PSG, pregnancy specific β1 glycoprotein; CEA, carcinoembryonic antigen; NCA, nonspecific cross-reactive antigen; FL-NCA, fetal liver NCA; BGP, biliary glycoprotein; TM-CEA, transmembrane CEA; hsCGM, homo sapiens CEA gene family member; UTR, untranslated region.

0006-291X/90 $1.50

definition of the antigen-family, and determination of the primary structure of CEA-related molecules should hopefully lead to the development of more specific antibodies and/or other probes.

The CEA family belongs to the immunoglobulin superfamily (2). All members have a characteristic N-terminal domain (N) corresponding to the variable region of immunoglobulin, and varying numbers of repeat domains (A) and (B). The (A) and (B) domains correspond to the constant region (C2) of immunoglobulin. A variety of C-terminal domains are present giving rise to either trans-membrane, cytoplasmic (3), and phosphatidyl inositol-linked (4,5) membrane-bound forms, or to secreted forms (6,7,8). The family can be divided into the CEA subgroup and the PSG subgroup (9,10). Glycoproteins of the latter subgroup are synthesized in placenta (6,11) and fetal liver (7,8) and increase in concentration in blood of pregnant women, reaching a peak concentration at partus (12). So far 17 members of the CEA gene family have been identified. Six of these belong to the CEA subgroup and 11 to the PSG subgroup (Table I).

Using immunochemical techniques and RNA blot analysis (13) we have recently demonstrated the co-expression of different CEA family members (NCA 55/95, BGP, PSG) in human submandibular salivary gland. Here we report the cDNA cloning and sequencing of three members of the PSG group from this gland.

MATERIALS AND METHODS

Probes

The following DNA-probes were used: 1) PSG1d [FL-NCA-1 (7)] as a universal CEA/PSG-family probe; 2) a BGP-specific 24-mer oligonucleotide (5' AGGTGGGTCATTGGAGTGGTCCTG 3') corresponding to position 1504-1527 in TM1-CEA (3); 3) NCA 55/95 specific 24-mer oligonucleotide (5' CACAGACTCCTCCCTGTTCAGCTG 3') corresponding to position 1678-1703 of NCA 55/95 cDNA (14).

The oligonucleotide probes are complementary to the coding sequence.

Screening of the cDNA library
A submaxillary salivary gland (=submandibular salivary gland) cDNA library in gt 11 (Clontech, Palo Alto, USA) was screened with the complete PSG1d cDNA probe. Prehybridization and hybridization of the recombinant phage DNAs on nylon membranes (Amersham, England) was carried out at 55°C, in 4 x SET (1 x SET = 0.15 M NaCl, 0.001 M EDTA, 0.02 M Tris-HCl pH 7.8), 0.1 % sodium pyrophosphate, 0.1 % SDS, 5 x Denhardts solution [1x = 0.02 % Ficoll (Pharmacia), 0.02 % polyvinylpyrrolidone and 0.02 % bovine serum albumin] and 250 μg/ml of denatured calf thymus DNA. The cDNA was labelled with [α-^{32}P]dCTP by hexanucleotide random priming. The probe (specific activity: 1.5 x 10^9 cpm/μg) was used at a concentration of 0.7 x 10^6/ml. Final post-hybridization washing was carried out in 0.1 x SET, 0.1 % sodium pyrophosphate and 0.1 % SDS at 68°C for two hours. For identification of positive plaques membranes were exposed to Kodak X-omat X-ray film.

Insert analysis by Southern blot hybridization
The recombinant phage DNA was digested with EcoR1, and size separated on a 1.2% agarose gel and transferred to a nylon membrane. The prehybridization and hybridization was carried out in 6 x SSC (1 x SSC = 0.15 M NaCl, 0.015 sodium citrate pH 7), 5 x Denhardt's solution, 0.1 % SDS and 100 μg/ml denatured calf thymus DNA at 65°C overnight. DNA inserts were identified by autoradiography.

DNA-sequencing
Nucleotide sequencing of whole or subcloned DNA fragments was performed by the dideoxy chain termination method (15) on double stranded DNA in pUC19 or Bluescript plasmid vectors.

RESULTS

We recently showed that human submandibular salivary gland and saliva contain several different CEA family members (13). To establish whether this tissue also contained the PSGs we screened a cDNA library from human submandibular salivary gland in phage vector gt11 using PSG1d cDNA as probe. About 1200 positive clones were found when 3.0 x 10^5 recombinant phage plaques were screened. Ten clones were plaque purified and analyzed further. The size of the cDNA inserts varied from 0.4 to 2.0 kb. Nine inserts hybridized with the PSG1d-probe (Figure 1). In contrast, none hybridized with the NCA 55/95- and/or the BGP-specific probes (data not shown).

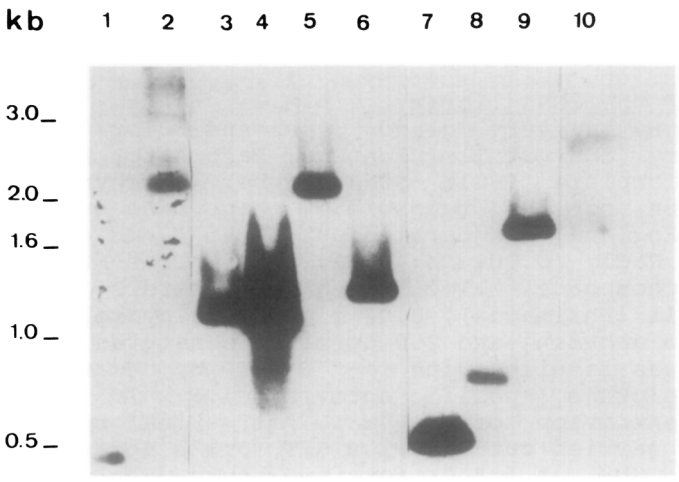

Figure 1. Southern blot analysis of cDNA clones from submandibular salivary gland. cDNA clones were digested with EcoR1 and fractionated on 1.0 % agarose gels. PSG1d positive cDNA clones (SG 1-10) are shown: Molecular weight markers are shown in kilobases.

Three cDNA clones of different sizes, SG5 (2.0 kb), SG8 (0.7 kb) and SG9 (1.7 kb) were chosen for sequence determination. The restriction enzyme maps and sequencing strategies are given in Figure 2.

SG5 and SG9 were found to be complete cDNA clones with a domain arrangement of N-A1-A2-B2-C while SG8 was a partial cDNA clone with B2 and C domains (Figure 3). The three clones were different from each other but encoded closely related genes that were found to belong to the PSG subgroup. when compared with known members of the CEA gene family. Accordingly, the A- and B domains were found to contain 2 cysteins at conserved positions, giving rise to Ig C2-like domains. SG9 had 7 Asn-Xaa-Ser/Thr glycosylation sites and SG5 had 4. The latter were found at identical positions in SG9.

The SG5 cDNA insert consisted of 1943 nucleotides of which 84 nucleotides belonged to the 5' UTR and 575 to the 3' UTR. The

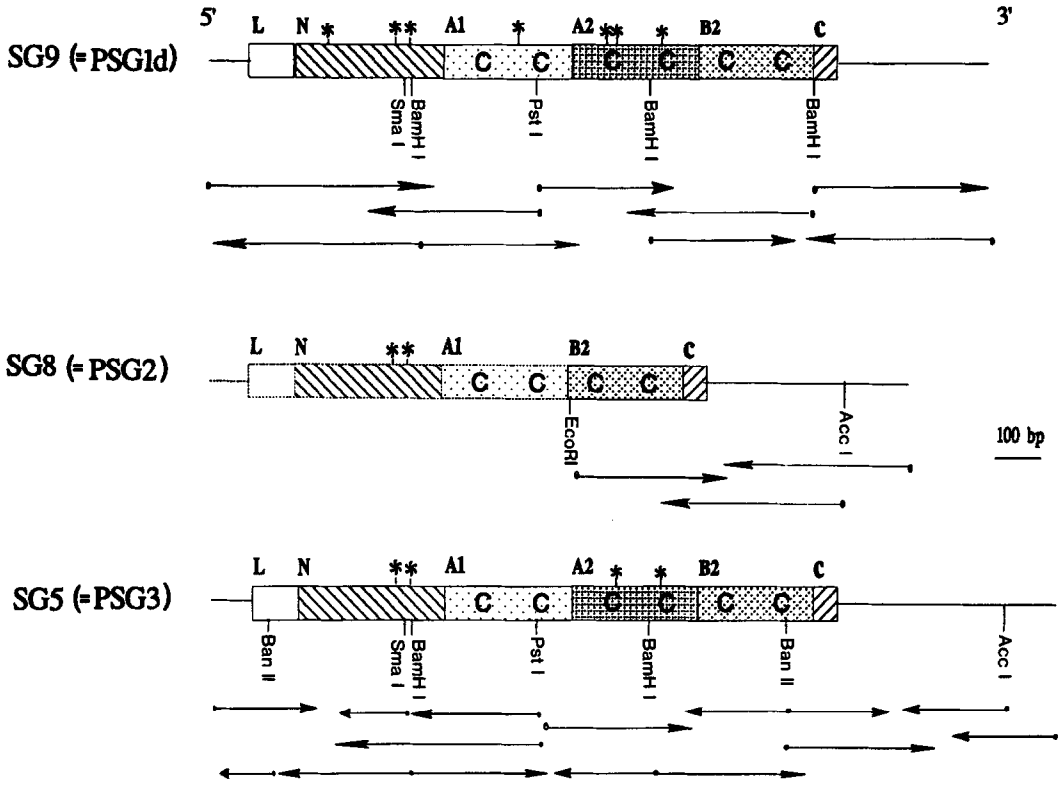

Figure 2. Schematic representation of domain arrangement, restriction sites and sequencing strategy of three PSG clones from human submandibular salivary gland. Arrows indicate sequencing strategy. L, leader peptide; N, N-terminal domain; A1 and A2, large internal domain; B, small internal domain; C, C-terminal domain. Solid lines represent 5' and 3' UTRs. Glycosylation sites are marked by stars. Hypothetical domains and UTR are indicated by broken lines. 100 bp are indicated by a bar.

latter included the polyadenylation signal sequence and 82 As at the 3' terminus. This clone coded for 428 amino acids of which 34 residues belonged to the leader peptide, giving rise to a mature 394 amino acids polypeptide with a molecular weight of 44,345 Daltons. The C-terminal domain contained 14 residues of amphipathic character. The N-domain of SG5 contained the Arg-Gly-Asp sequence (position 93-95) which is a cell recognition site on fibronectin and related matrix proteins (16).

SG9 had a 1721 bp cDNA insert, of which 97 nucleotides at the 5' end and 347 at the 3' end belonged to the untranslated region.

SG9 (=PSG1d)

```
                                                                                                   ->L
GGGAGGACAGCACAGCTGACAGCCGTGCTCAGAGAGTTTCTGGATCCTAGGCTTATCTCCACAGAGGAGAACACACAAGCAGCAGAGACCATGGGAACCCTCTCAGCCCCTCCCTGCACA  120
                                                                                          M  G  T  L  S  A  P  P  C  T    -10
                                                                                ->N
CAGCGCATCAAATGGAAGGGGCTCCTGCTCACAGCATCACTTTTAAACTTCTGGAACCTGCCCACCACTGCCCAAGTCACGATTGAAGCCGAGCCAACCAAAGTTTCCGAGGGGAAGGAT  240
Q  R  I  K  W  K  G  L  L  L  T  A  S  L  L  N  F  W  N  L  P  T  T  A  Q  V  T  I  E  A  E  P  T  K  V  S  E  G  K  D     16

GTTCTTCTACTTGTCCACAATTTGCCCCAGAATCTTACCGGCTACATCTGGTACAAAGGGCAAATGAGGGACCTCTACCATTACATTACATCATATGTAGTAGACGGTGAAATAATTATA  360
V  L  L  L  V  H  N  L  P  Q  N  L  T  G  Y  I  W  Y  K  G  Q  M  R  D  L  Y  H  Y  I  T  S  Y  V  V  D  G  E  I  I  I     56
                              *
TATGGGCCTGCATATAGTGGACGAGAAACAGCATATTCCAATGCATCCCTGCTGATCCAGAATGTCACCCGGGAGGACGCAGGATCCTACACCTTACACATCATAAAGGGAGATGATGGG  480
Y  G  P  A  Y  S  G  R  E  T  A  Y  S  N  A  S  L  L  I  Q  N  V  T  R  E  D  A  G  S  Y  T  L  H  I  I  K  G  D  D  G     96
                                       *->A1               *
ACTAGAGGAGTAACTGGACGTTTCACCTTCACCTTACACCTGGAGACTCCTAAGCCCTCCATCTCCAGCAGCAACTTAAATCCCAGGGAGACCATGGAGGCTGTGAGCTTAACCTGTGAC  600
T  R  G  V  T  G  R  F  T  F  T  L  H  L  E  T  P  K  P  S  I  S  S  S  N  L  N  P  R  E  T  M  E  A  V  S  L  T  C  D    136
                                                                                                                      •
CCTGAGACTCCAGACGCAAGCTACCTGTGGTGGATGAATGGTCAGAGCCTCCCTATGACTCACAGCTTGAAGCTGTCCGAAACCAACAGGACCCTCTTTCTATTGGGTGTCACAAAGTAT  720
P  E  T  P  D  A  S  Y  L  W  W  M  N  G  Q  S  L  P  M  T  H  S  L  K  L  S  E  T  N  R  T  L  F  L  L  G  V  T  K  Y    176
                                                                                ->A2*
ACTGCAGGACCCTATGAATGTGAAATACGGAACCCAGTGAGTGCCAGCCGCAGTGACCCAGTCACCCTGAATCTCCTCCCGAAGCTGCCCAAGCCCTACATCACCATCAACAACTTAAAC  840
T  A  G  P  Y  E  C  E  I  R  N  P  V  S  A  S  R  S  D  P  V  T  L  N  L  L  P  K  L  P  K  P  Y  I  T  I  N  N  L  N    216
                  •
CCCAGGGAGAATAAGGATGTCTTAAACTTCACCTGTGAACCTAAGAGTGAGAACTACACCTACATTTGGTGGCTAAATGGTCAGAGCCTCCCGGTCAGTCCCAGGGTAAAGCGACCCATT  960
P  R  E  N  K  D  V  L  N  F  T  C  E  P  K  S  E  N  Y  T  Y  I  W  W  L  N  G  Q  S  L  P  V  S  P  R  V  K  R  P  I    256
                           *        •                 *                                                                  ->B2
GAAAACAGGATCCTCATTCTACCCAGTGTCACGAGAAATGAAACAGGACCCTATCAATGTGAAATACGGGACCGATATGGTGGCATCCGCAGTGACCCAGTCACCCTGAATGTCCTCTAT 1080
E  N  R  I  L  I  L  P  S  V  T  R  N  E  T  G  P  Y  Q  C  E  I  R  D  R  Y  G  G  I  R  S  D  P  V  T  L  N  V  L  Y    296
                                    *                    •
GGTCCAGACCTCCCCAGAATTTACCCTTCATTCACCTATTACCGTTCAGGAGAAGTCCTCTACTTGTCCTGTTCTGCGGACTCTAACCCACCGGCACAGTATTCTTGGACAATTAATGAA 1200
G  P  D  L  P  R  I  Y  P  S  F  T  Y  Y  R  S  G  E  V  L  Y  L  S  C  S  A  D  S  N  P  P  A  Q  Y  S  W  T  I  N  E    336
                                                                     •
AAGTTTCAGCTACCAGGACAAAAGCTCTTTATCCGCCATATTACTACAAAGCATAGCGGGCTCTATGTTTGCTCTGTTCGTAACTCAGCCACTGGCAAGGAAAGCTCCAAATCCATGACA 1320
K  F  Q  L  P  G  Q  K  L  F  I  R  H  I  T  T  K  H  S  G  L  Y  V  C  S  V  R  N  S  A  T  G  K  E  S  S  K  S  M  T    376
          ->C                                                        •
GTCGAAGTCTCTGGTAAGTGGATCCCAGCATCGTTGGCAATAGGGTTTTAGGTGGAGTCTATCTGGCATTCAGAGAAGAGTCAGGAAAACAATTGTATTCCCAGCCTGTGTCCCATGGGC 1440
V  E  V  S  G  K  W  I  P  A  S  L  A  I  G  F  End                                                                         392
ACAAGCAAATCCCAAATTCTCCTCCTGAACCCTCCAAATTTGTCTAAGAACTTCGAAAACTTTAACAAACAGGCTGATATCTTCATAATATTCCCAGCCTAGACCAAGCAGGAAGCATTG 1560
ATTTCATTGAAATAATTGATAATAATGAAGATAATGTTTTTATGATTTTTATTTGAAAATTTGCTGATTCTTTAAATGGTTTGTTTTCTACATTGATGGAATTTTTCTCTTAATTAATCT 1680
ATCTACAGCTTATAGCAGTTCAATAAA                                                                                                1707
```

SG8 (=PSG2)

```
->B2
 GAATTCACCCTTCATACACCAATTACCGTTCAGGAGATAACCTCTACTTGTCTTGCTTCGCGAACTCTAACCCACCGGCACAGTATTCTTGGACAATTAATGGGAAGTTTCAGCAATCA 119
   I  H  P  S  Y  T  N  Y  R  S  G  D  N  L  Y  L  S  C  F  A  N  S  N  P  P  A  Q  Y  S  W  T  I  N  G  K  F  Q  Q  S     39
                                                     •                                                                 ->C
GGACAAAATCTGTTTATCCCCCAAATTACTACAAAGCATAGCGGGCTCTATGTTTGCTCTGTTCGTAACTCAGCCACTGGGCAGGAAAGCTCCACATCGTTGACAGTCAAAGTCTCTGCT 239
G  Q  N  L  F  I  P  Q  I  T  T  K  H  S  G  L  Y  V  C  S  V  R  N  S  A  T  G  Q  E  S  S  T  S  L  T  V  K  V  S  A     79
                                                     •
TCTACAAGAATAGGACTTCTTCCTCTCCTTAATCCAACATAGCAGCTGTGATGTCATTTCTGTATTTCAGGAAGACTGGCAGGAGATTTATGGAAAGGTCTCTTACAAGGACTCTTGAAT 359
S  T  R  I  G  L  L  P  L  L  N  P  T  End                                                                                   92
                              *
ACAAGCTCCTGATAACTTCAAGATCATACCACTGGACTAAGAACTTTCAAAATTTTAATGAACAGGCTGATACCTTCATGAAATTCAAGACAAAGAAGAAAAATACTCAATGTTATTGGA 479
CTAAATAATCAAAAGGATAATGATTTCATAATTTTCTATTTGAAAATGTGCTGATTCTTGGAATGTTTCATTCTCCAGATTTATGAACATTTTTTCTTGAGCAATTGGTAAAGTATACTT 599
TTGTAAACAAAAATTGAAACATTTCCTTTTGCTCTCTATCTGAGTGCCCCAGAATTGGGAATCTATTCATGAGTATTCATATGTTTATGGTAATAAAGCTATTTGCACAAGTTCAAAAAA 719
AAAAAAAAAAAA                                                                                                               731
```

SG5 (=PSG3)

```
                                                                                                  ->L
 GGACAGCACAGCTGAGAGCCATGCTCAGGAAGTTTCTGGATCCTAGGCTCAGCTCCACAGAGGAGAACACGCAGGCAGCAGAGACCATGGGGCCCCTCTCAGCCCCTCCCTGCACACAG  119
                                                                                         M  G  P  L  S  A  P  P  C  T  Q    -11
                                                                            ->N
CGCATCACCTGGAAGGGGCTCCTGCTCACAGCATCACTTTTAAACTTCTGGAACCCGCCTACCACTGCCCAAGTCACGATTGAAGCCGAGCCAACCAAAGTTTCCAAGGGGAAGGACGTT  239
R  I  T  W  K  G  L  L  L  T  A  S  L  L  N  F  W  N  P  P  T  T  A  Q  V  T  I  E  A  E  P  T  K  V  S  K  G  K  D  V     17

CTTCTACTTGTCCACAATTTGCCCCAGAATCTTGCTGGCTACATCTGGTACAAAGGGCAAATGAAGGACCTCTACCATTACATTACATCATACGTAGTAGATGGTCAAATAATTATATAT  259
L  L  L  V  H  N  L  P  Q  N  L  A  G  Y  I  W  Y  K  G  Q  M  K  D  L  Y  H  Y  I  T  S  Y  V  V  D  G  Q  I  I  I  Y     57

GGGCCTGCATACAGTGGACGAGAAACAGTATATTCCAATGCATCCCTGCTGATCCAGAATGTCACCCGGGAGGACGCAGGATCCTACACCTTACACATCGTAAAGCGAGGTGATGGGACT  479
G  P  A  Y  S  G  R  E  T  V  Y  S  N  A  S  L  L  I  Q  N  V  T  R  E  D  A  G  S  Y  T  L  H  I  V  K  R  G  D  G  T     97
                                    *->A1              *                                                   ‾‾‾‾‾‾‾
AGAGGAGAAACTGGACATTTCACCTTCACCTTATACCTGGAGACTCCCAAGCCCTCCATCTCCAGCAGCAACTTATACCCCAGGGAGGACATGGAGGCTGTGAGCTTAACCTGTGATCCT  599
R  G  E  T  G  H  F  T  F  T  L  Y  L  E  T  P  K  P  S  I  S  S  S  N  L  Y  P  R  E  D  M  E  A  V  S  L  T  C  D  P    137
                                                                                                                   •
GAGACTCCGGACGCAAGCTACCTGTGGTGGATGAATGGTCAGAGCCTCCCTATGACTCACAGCTTGCAGTTGTCCAAAAACAAAAGGACCCTCTTTCTATTTGGTGTCACAAAGTACACT  719
E  T  P  D  A  S  Y  L  W  W  M  N  G  Q  S  L  P  M  T  H  S  L  Q  L  S  K  N  K  R  T  L  F  L  F  G  V  T  K  Y  T    177
                                                                                         ->A2
CAGCCGCAGTGACCCAGTCACCCTGAATCTCCTCCCGAAGCAGGACCCTATGAATGTGAAATACGGAACCCAGTGAGTGCGCTGCCCAAGCCCTACATCACCATCAACAACTTAAACCCC  839
A  G  P  Y  E  C  E  I  R  N  P  V  S  A  S  R  S  D  P  V  T  L  N  L  L  P  K  L  P  K  P  Y  I  T  I  N  N  L  N  P    217
               •
AGGGAGAATAAGGATGTCTTAGCCTTCACCTGTGAACCTAAGAGTGAGAACTACACCTACATTTGGTGGCTAAATGGTCAGAGCCTCCCGGTCAGTCCCAGGGTAAAGCGACCCATTGAA  959
R  E  N  K  D  V  L  A  F  T  C  E  P  K  S  E  N  Y  T  Y  I  W  W  L  N  G  Q  S  L  P  V  S  P  R  V  K  R  P  I  E    257
                              •                 *                                                                      ->B2
AACAGGATCCTCATTCTACCCAGTGTCACGAGAAATGAAACAGGACCCTATCAATGTGAAATACAGGACCGATATGGTGGCATCCGCAGTTACCCAGTCACCCTGAATGTCCTCTATGGT 1079
N  R  I  L  I  L  P  S  V  T  R  N  E  T  G  P  Y  Q  C  E  I  Q  D  R  Y  G  G  I  R  S  Y  P  V  T  L  N  V  L  Y  G    297
                                 *                    •
CCAGACCTCCCCAGAATTTACCCTTCATTCACCTATTACCATTCAGGAGAAAACCTCTACTTGTCCTGCTTCGCGGACTCTAACCCACCAGCAGAATATTCTTGGACAATTAATGGGAAG 1199
P  D  L  P  R  I  Y  P  S  F  T  Y  Y  H  S  G  E  N  L  Y  L  S  C  F  A  D  S  N  P  P  A  E  Y  S  W  T  I  N  G  K    337
                                                                  •
TTTCAGCTATCAGGACAAAAGCTCTTTATCCCCCAGATTACTACAAAGCATAGCGGGCTCTATGCTTGCTCTGTTCGTAACTCAGCCACTGGCATGGAAAGCTCCAAATCCATGACAGTC 1319
F  Q  L  S  G  Q  K  L  F  I  P  Q  I  T  T  K  H  S  G  L  Y  A  C  S  V  R  N  S  A  T  G  M  E  S  S  K  S  M  T  V    377
       ->C                                                        •
AAAGTCTCTGCTCCTTCAGGAACAGGACATCTTCCTGGCCTTAATCCATTATAGCAGCCGTGATGTCATTTCTGTATTTCAGGAAGACTGGCAGACAGTTGCTTTCATTCTTCCTCAAAG 1439
K  V  S  A  P  S  G  T  G  H  L  P  G  L  N  P  L                                                                            394
TATTTACCATCAGCTACAGTCCAAAATTGCTTTTTGTTCAAGGAGATTTATGAAAAGACTCTGACAAGGACTCTTGAATACAAGTTCCTGATAACTTCAAGATCATACCACTGGACTAAG 1559
AACTTTCAAAATTTTAATGAACAGGCTGATACTTCATGAAATTCAAGACAAAGAAAAAAACCCAATTTTATTGGACTAAATAGTCAAAACAATGTTTTCATAATTTTCTATTTGAAAATG 1679
TGCTGATTCTTTGAATGTTTTATTCTCCAGATTTATGCACTTTTTTTCTTCAGCAATTGGTAAAGTATACTTTTGTAAACAAAAATTGAAACATTTGCTTTTGCTCCCTAAGTGCCCCAG 1799
AATTGGGAAACTATTCAGGAGTATTCATATGTTTATGGTAATAAAGTTATCTGCACAAGTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1919
AAAAAAAAAAAAAAAAAAAAAAAA                                                                                                   1943
```

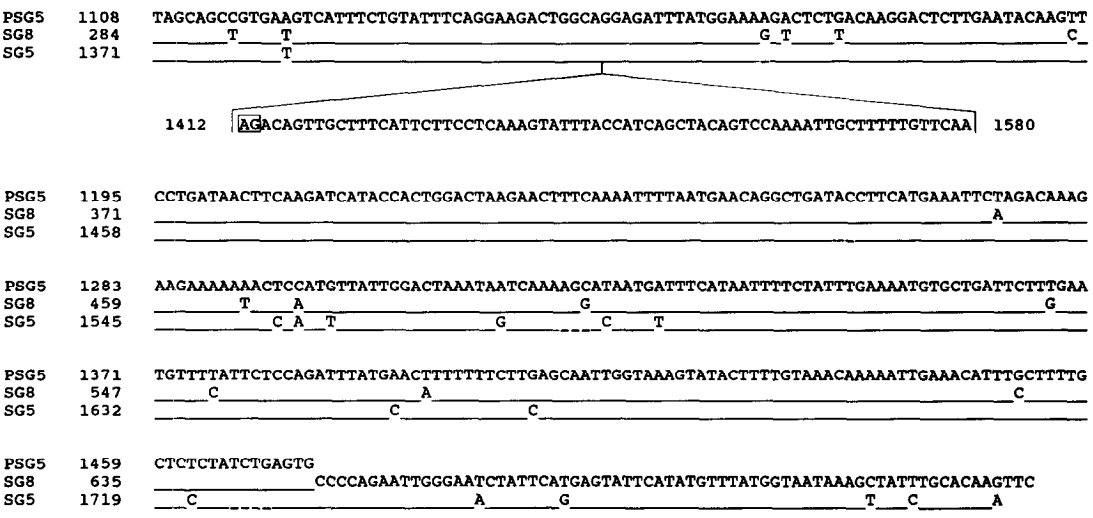

Figure 4. Nucleotide alignments of 3' untranslated regions of PSG5, SG8 and SG5. Identical nucleotides are represented with a line, only missmatches are written out. Broken lines show gaps. Nucleotide numbers are given to the left. The 67 nucleotide deletion (position 1412-1580) in SG5 is shown under the alignment. The consensus splice acceptor site in SG5 is boxed at the point of deletion.

The pre-protein of SG9 was 426 amino acids and the molecular weight of the polypeptide without the leader was 44,120 Daltons. The C-terminal domain consisted of 12 residues and was hydrophobic in character.

SG8, an EcoR1 fragment of 731 nucleotides, was a truncated cDNA lacking most of the coding sequence at the 5' end. The 5' EcoR1 site was an internal site and this clone probably has resulted from incomplete EcoR1 site protection during construction of the cDNA library. The coding sequence started at the twentysecond nucleotide of the B2 domain and contained 78 amino acids. This domain was followed by the C-terminal domain, which comprised 14 amino acids, and the 3' UTR (453 bp). There was a polyadenylation

Figure 3. Nucleotide and deduced amino acid sequence of SG9, SG8 and SG5. Sequences are marked with domain boundaries. Cysteines are marked with dots and Asn-Xaa-Ser/Thr glycosylation sites with stars. RGD sequence and poly (A) signal are underlined.

site at position 690-695 and the cDNA ended with an 18 oligo (A) stretch.

The 3' UTRs of SG5 and SG8 were highly homologous (95 %) (Figure 4). On the other hand, the 3' UTR of SG9 differed from that of SG5 and SG8, except for 80 % homology over a stretch of 92 bp (1553-1645) located 30 bp 5' to the poly A signal of SG9.

DISCUSSION

Data base comparisons reveal that the three submandibular salivary gland clones described here are identical to recently described cDNA clones from placenta and fetal liver. Thus, SG5 was identical to PSG3, originally described in placenta while SG9 was identical to PSG1d, expressed in fetal liver and placenta (Table I and II). As the SG8 partial cDNA clone is identical to the 3' end EcoR1 fragment of PSG2, which includes the B2 and C domains as well as the entire 3' UTR, we may reasonably conclude that SG8 and PSG2 code for an identical glycoprotein although it cannot be excluded that the SG8 and PSG2 mRNAs are two differentially spliced variants of the same gene, having a different domain arrangement.

This study shows clearly that PSGs are expressed in normal submandibular salivary gland. Thus PSGs are not confined to placenta and fetal liver, although it would appear as if they are only expressed at high levels in these tissues. Earlier studies, using reactivity with antisera or monoclonal antibodies and molecular size as the criteria to identify PSGs, indicated that normal skin fibroblast (17,18,19), and cells of the myelomonocytic lineage (20) express PSG. However, due to the difficulty in distinguishing different CEA family members from

Table I. The carcinoembryonic antigen gene family

Subgroup	Old nomenclature	New nomenclature	CHO sites	RGD	Domain structure	Ref
CEA	CEA	CEAa	28	-	N-A1-B1-A2-B2-A3-B3-M	28,29,30
	CEA	CEAb	28	-	N-A1-B1-A2-B2-A3-B3-M	9
	NCA 55/95	NCA 55/95	12	-	N-A1-B1-M	14,30,31
	hsCGM6, GN-1, M6	CGM6	11	-	N-A1-B1-M	32,33,34
	BGPI, TM1-CEA	BGPa**	20	-	N-A1-B1-A2-TM-Cyt	35,3
	TM2-CEA	BGPb**	16	-	N-A1-B1-TM-Cyt	3
	TM3-CEA	BGPc**	19	-	N-A1-B1-A2-TM-Cyt	3
	TM4-CEA	BGPd**	15	-	N-A1-B1-TM-Cyt	3
	hsCGM1*					9
	hsCGM2*					9
PSG	PSG93,PSβG D,FL-NCA-2, hPSP11, hPS3, PSG1a	PSG1a**	7	-	N-A1-A2-B2-C	6,36,7,23,37
	PSG16	PSG1b**	7	-	"	6
	PSβG C	PSG1c**	7	-	"	36
	FL-NCA-1,PSG1d,CO-NCA, SG9	PSG1d**	7	-	"	7,37,21
	PSβG E, SG8	PSG2	5	+	N-A1-B2-C	36
	pSP1-i,hc17,PS35,SG5,hTS6[&]	PSG3	4	+	N-A1-A2-B2-C	38,39,40,23
	hsCGM4*, PSG4,hHSP2	PSG4	6	-	"	9,37,22
	FL-NCA-3,PSβGHL	PSG5	4	+	N-A2-B2-C	8,25
	hsCGM3* PSG6	PSG6	7	+	N-A1-A2-B2-C	9,37
	PSG7	PSG7	?	-	"	41
	CGM35*,hTS1[&]	PSG8	?	-	(N)-A1-[B1]-A2-B2-Cdcab	26,23
	PSGkα	PSG9	?	?	?	24
	PSG10	PSG10	?	?	?	42
	FL-NCA-4,PS34,PSGB	PSG11	6	+	N-A1-A2-B2-C	43,40,44

Nomenclature according to ISOBM Workshop proposal (10).
*, genomic clones; **, a, b, c and d are alternative splice forms; (N) indicates that N-domain exon is absent in CGM35; [B1] indicates that exon is abortive, thus, not processed into mRNA. &, incompletely processed mRNA.

Table II. Distribution of PSGs in normal human tissues and tumor cell lines

PSG	Tissues and cells							
	Normal tissues					Tumor cell lines		
	Placenta	Fetal liver	Salivary gland	Colon	Testis	HL-60	KG-1	HeLa
1	+	+	+	+	+(t)			
2	+		+(t)					
3	+		+		+(t,i)			
4		+						+(t)
5		+				+(t)		
6		+						
7	+							
8					+(i)			
9							+	
10							+(?)	
11	+	+						

Data compiled from the following sources: placenta and fetal liver (see Table I); submandibular salivary gland (this paper); normal colon (21); testis (22,23), HL-60 (25); KG-1 (24); HeLa (22). (t), truncated cDNA; (i), cDNA containing intervening sequences.

each other by immunological reagents, these earlier studies do not conclusively prove that the tissues express PSGs. Table II gives tissues and tumor cell lines where the presence of PSGs has been established by molecular cloning and sequencing of the corresponding cDNA. As can be seen PSGs have so far been demonstrated in normal submandibular salivary gland (this paper), colon (21), and testis (22,23). Note, however, that all three PSG cDNA from testis either contained intervening non coding sequences or were truncated at the 5' end (Table II). Among tumor cell lines PSG cDNAs were demonstrated in libraries from the erythroleukemia cell line, KG-1 (24), the promyelocytic leukemia cell line, HL-60 (25) and the cervix carcinoma cell line, HeLa (22). It is interesting to note that different tissues or cells

tend to express different PSGs. PSG expression may therefore be regulated in a tissue specific - as well as in a developmentally related-manner. Experiments aiming at identifying regulatory elements in selected PSG genes are in progress.

There is a strikingly high homology (94 %) between the 3' UTR of SG5, SG8 and PSG5 (Figure 4). This homology extends to PSG1d (SG9) and to CEA since they contain segments which are homologous to SG5, SG8 and PSG5. Moreover, 3' UTR of NCA 55/95 was found to be highly homologous (430 bp) to PSG1a (26). These findings strongly suggest a common evolutionary origin of the genes.

As mentioned above, the homology between the 3' UTRs of SG5, SG8 and PSG5 was of the order of 94 %. However, a deletion of 67 nucleotides was found at position 1146 and 321 of PSG5 and SG8, respectively. The 67 bp segment begins with a consensus splice acceptor site. The results indicate that the 3' UTRs of these genes comprise more than one exon.

There are four differentially spliced forms of the PSG1 gene (a-d) (Table I) which give rise to PSGs with different C-terminal domains. In this tissue we have so far only identified PSG1d. The intron-exon organization is known only for one member of the PSG subgroup, i.e. PSG8 (ref 26, Table I). The predicted PSG1 gene is analogous to PSG8 in that all the domains of PSG1 have corresponding exons in the PSG8 gene, including the C-terminal domains. The four C-terminal domain exons are arranged in two sets of two overlapping exons where Cd overlaps with Cc and Ca overlaps with Cb. PSG1d is an unspliced mRNA, the C-terminal of which is a continuation of the open reading frame of the B2 domain. It utilizes the first poly A signal which lies in the

coding part of the 1c C-terminal domain. Recently, Streydio et al (27) reported a second variant of PSG1d cDNA (Ci), which gave the same 12 amino acid C-terminal as does the PSG1d studied here while utilizing the poly A signal present in the next exon, to form a longer 3' UTR. CEA mRNA also occurs in two different size fractions (i.e. 4.0 and 3.5 kb). They apparently code for an identical peptide but utilize two different polyadenylation signals.

In summary, we show through cDNA cloning the expression of three PSGs in human submandibular salivary gland. Protein and mRNA analysis of submandibular salivary gland revealed the presence of NCA 55/95 and BGP a/c in this normal adult tissue (13). Thus there are at least 5 CEA gene family members expressed in salivary gland although not necessary in the same cell type.

ACKNOWLEDGMENTS

We thank Ms Anne Israelsson, Ms Marie Svensson and Ms Elisabeth Granström for their technical assistance. Dr Teresa Ramos for critical reading of the manuscript and Ms Christine Boström for secretarial assistance. This work was supported by a grant from the Swedish Cancer Society.

REFERENCES

1. Gold, P., and Freedman, S.O. (1965) J. Exp. Med. 122, 467-481.
2. Paxton, R., Mooser, G., Pande, H., Lee, T., and Shively, J. (1987) Proc. Natl. Acad. Sci. USA 84, 920-924.
3. Barnett, T.R., Kretschmer, A., Austen, D.A., Scott, J.G., Hart, J.T., Elting, J.J., and Kamarck, M.E. (1989) J. Cell Biol. 108, 267-276.
4. Hefta, S.A., Hefta, L.J., Lee, T.D., Paxton, R.J., and Shively, J.E. (1988) Proc. Natl. Acad. Sci. USA 85, 4648-4652.
5. Kolbinger, F., Schwarz, K., Brombacher, F., von Kleist, F., and Grunert, F. (1989) Biochem. Biophys. Res. Commun. 161, 1126-1134.
6. Watanabe, S., and Chou, J. (1988) J. Biol. Chem. 263, 2049-2054.
7. Khan, W.N., Osterman, A., and Hammarström, S. (1989) Proc. Natl. Acad. Sci. USA 86, 3332-3336.
8. Khan. W.N., and Hammarström, S. (1989) Biochem. Biophys. Res. Commun 86, 525-535.

9. Thompson, J.A., Mauch, E.-M., Chen, F.-S., Hinoda, Y., Schrewa, H., Berting, B., Barnett, S., von Kleist, S., Shively, J.E., and Zimmermann, W. (1989) Biochem. Biophys. Res. Commun. 158, 996-1004.
10. Barnett, T., and Zimmermann, W. (1990) Tumor Biol. 11, 59-63.
11. Horne, C.H.W., Towler, C.M., Pugh-Humphreys, R.G.P., Thompson, A.W., and Bohn, H. (1976) Experientia, 32, 1197-1199
12. Grudzinskas, J.G., Gordon, Y.B., Jeffrey, D., and Chard, T. (1977) Lancet 1, 333-334.
13. Zoubir, F., Khan, W.N., Hammarström, M-L., Zeromski, J, Carenfelt, C., Szczyrzyca, H., and Hammarström, S. (1990) Molec. Immun. (submitted).
14. Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., Riggs, A., and Shively, J. (1988) J. Biol. Chem. 263, 3202-3207.
15. Sanger, F., Nickmen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Ruoslahti, E., and Pierschbacher, M. (1987) Science 238, 491-497.
17. Rosen, S., Kaminska, J., Calvert, I., and Aaronson, S. (1979). Am. J. Obstet. Gynecol. 134, 734-738.
18. Engvall, E., Miyashita, M., and Ruoslahti, E. (1982) Cancer Res. 42, 2028-2033.
19. Chou, J.Y. Oncodev. Biol. Med. (1983) 4, 319-326.
20. Heikinheimo, M., Gahmberg, C., Bohn, H., and Andersson, L. (1987) Blood 70, 1279-1283.
21. Khan, W.N., Osterman, A., Zoubir, F., and Hammarström, S. (1989) In "The carcinoembryonic antigen gene family" (A. Yachi and J.E. Shively, eds.) p. 87-96 (1989) Elsevier Science Publishers.
22. Chan, W.-Y., Borjigin, J., Zheng, Q.-X., and Shupert, W.L. (1988) DNA 7, 545-555.
23. Borjigin, J., Tease, L.A., Chan, W.Y. (1990) Biochem. Biophys. Res. Commun. 166, 1063-1071.
24. Barnett, T., Pickle II, W., Rae, P., Hart, J., Kamarch, M., Elting, J. (1989) Am. J. Hum. Genet. 44, 890-893.
25. Oikawa, S., Inuzuka, C., Kuroki, M., Matsuoka, Y., Kosaki, G., and Nakazato, H. (1989) Biochem. Biophys. Res. Commun. 163, 1021-1031.
26. Oikawa, S., Inuzuka, C., Kosaki, G., and Nakazato, H. (1988) Biochem. Biophys. Res. Commun. 156, 68-77.
27. Streydio, C., and Vassart, G. (1990) Biochem. Biophys. Res. Commun. 166, 3, 1265-1273.
28. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A., and Stanner, S. (1987) Mol. Cell. Biol. 7, 3221-3230.
29. Oikawa, S., Nakazato, H., and Kosaki, G. (1987) Biochem. Biophys. Res. Commun. 142, 511-518.
30. Barnett, T., Goebel, S.J., Nothdurft, M.A., and Elting, J.J. (1989) Genomics 3, 59-66
31. Tawargi, Y., Oikawa, S., and Matsuoka, Y. (1988) Biochem. Biophys. Res. Commun. 150, 89-96.
32. Berling, B., Zimmermann, W., Thompson, J. (1989) Abstract XVIIth meeting of International Society for Oncodevelopmental Biology and Medicine, Freiburg.
33. Fumiko, A., Kuroki, M., Misumi, Y., Oikawa, S., Nakazato, H., and Matsuoka, Y. (1990) Biochem. Biophys. Res. Commun. 166, 1063-1071.
34. Boucher, D., Cournoyer, D., Fuks, A., and Stanners, C.P. (1989) Abstract XVIIth Meeting of the International Society for Oncodevelopmental Biology and Medicine, Freiburg.
35. Hinoda, Y., Neumaier, M., Hefta, S., Drzeniek, Z., Wagener, C., Shively, J., and Paxton, R. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 6959-6963.

36. Streydio, C., Lacka, K., Swillens, S., and Vassart, G. (1988) Biochem. Biophys. Res. Commun. 154, 130-137.
37. Zimmermann, W., Wagner, K., Weiss, M., and Thompson, J.A. Biochem. Biophys. Res. Commun. 163, 1197-1209.
38. Rooney, B., Wilson Horn, C., and Hardman, N. (1988) Gene 71, 439-449.
39. McLenachan, T., and Manfield, B. (1989) Biochem. Biophys. Res. Commun. 162, 1486-1493.
40. Arakawa, F., Kuroki, Mo., Misumi, Y., and Matsuoka, Y. (1989) Abstract XVIIth Meeting of the International Society for Oncodevelopmental Biology and Medicine, Freiburg.
41. Thompson, J., Koumari, R., and Wagner, K. (1989) Abstract XVIIth Meeting of the International Society for Oncodevelopmental Biology and Medicine, Freiburg.
42. Barnett, T., Picklett, W., Austen, D., Hart, J., Nothdurft, M.A., Rae, P., Kretschmer, A., Kamarck, M., and Elting, J. (1989) Abstract XVIIth Meeting of the International Society for Oncodevelopmental Biology and Medicine, Freiburg.
43. Khan, W.N., and Hammarström, S. (1990) Biochem. Biophys. Res. Commun. (in press).
44. Streydio, C., Swillens, S., Georges, M., Szpirer, C., and Vassart, G. (1990) Genomics (in press).