Characterization of murine carcinoembryonic antigen gene family members

Fritz Rudert,[1,*] Ann M. Saunders,[2] Sabine Rebstock,[1] John A. Thompson,[1] and Wolfgang Zimmermann[1]

[1]Institut für Immunbiologie, Universität Freiburg, Stefan-Meier-Strasse 8, D-7800 Freiburg, FRG; [2]Division of Neurology, Duke University Medical Center, Durham, North Carolina 27710, USA

Abstract. The carcinoembryonic antigen (CEA) is a human tumor marker whose gene belongs to a family with more than 20 members. This gene family codes for a group of proteins with in vitro cell adhesion properties and for a group of abundantly expressed pregnancy-specific glycoproteins (PSG) with unknown functions. As a basis for in vivo functional studies, we have started to analyze the murine CEA gene family and have identified five new members (*Cea-2* to *Cea-6*). cDNA clones were isolated for *Cea-2*, *Cea-3*, and *Cea-6*. The deduced amino acid sequences of *Cea-2* and *Cea-6* indicate three IgV-like (N), followed by one IgC-like (A) domain (N1-N2-N3-A). We have also partially characterized the *Cea-2* gene and two additional ones, *Cea-4* and *Cea-5*. *Cea-2* and *Cea-4* are separated by only 16 kb, suggesting a close linkage of murine CEA-related genes, as found for the human CEA gene family. *Cea-5* was located to the proximal region of mouse Chromosome (Chr) 7, which is syntenic to part of human Chr 19, containing the human CEA gene family cluster. *Cea-2*, *Cea-3*, and a *Cea-4*-like gene are differentially transcribed in the placenta during pregnancy, but not in other organs tested. This expression pattern strongly suggests that they represent counterparts of the human PSG subgroup members, despite the presence of multiple IgV-like domains, a feature not found for human PSGs. The more distantly related *Cea-5* seems to be ubiquitously expressed. The putative promoter region of *Cea-2* lacks typical TATA- or CAAT-boxes, but contains other conserved motifs that could play a role in the initiation of transcription.

Introduction

Carcinoembryonic antigen (CEA), which is widely used as a tumor marker (Shively and Beatty 1985), the closely related nonspecific crossreacting antigen (NCA), and biliary glycoproteins (BGP), as well as their rodent counterparts (mmCGM2 and C-CAM 105), have recently been shown to function in vitro as homo- and heterophilic cell adhesion molecules (reviewed in Thompson et al. 1991). The mRNAs encoding these antigens show distinct and rather restricted expression patterns, being predominantly found in colonic mucosa and adenocarcinomas (CEA, NCA), liver (BGP), and cells of the myeloid lineage [NCA, CGM6 (Zimmermann et al. 1987, 1988; Hinoda et al. 1990)]. The human CEA-related antigens are encoded by a gene family with ~20 members (reviewed in Thompson et al. 1991), which represents a branch of the immunoglobulin gene superfamily (Williams and Barclay 1988). The CEA gene family, which is clustered on the long arm of Chr 19 (Thompson et al. 1992), can be subdivided into the CEA subgroup, containing the genes for the classical crossreacting antigens and the pregnancy-specific glycoprotein (PSG) subgroup (Watanabe and Chou 1988). The PSGs comprise a group of very closely related proteins of unknown function encoded by at least 11 genes (Khan et al. 1992). In contrast to the CEA-related proteins, which are membrane-bound, they appear to be secreted because most of them lack hydrophobic C-terminal domains suitable for membrane anchorage. PSGs are produced in large amounts by the syncytiotrophoblast and represent the major group of placental proteins in the serum of women at term pregnancy (Lin et al. 1974). As tumors of trophoblastic origin also produce PSGs, these proteins have been used as markers to monitor the treatment of choriocarcinoma patients (Sakuragi 1982).

Recently, putative PSG genes have also been identified in the rat (Kodelja et al. 1989; Rebstock et al. 1990). However, they show relatively low sequence

---

similarity to the human CEA gene family. Furthermore, the deduced domain organization of a rat PSG-related antigen (rnCGM1) has been found to be completely different from its human counterparts. Several immunoglobulin variable-like (IgV) domains are followed by one immunoglobulin constant-like (IgC) domain (Rebstock et al. 1990), whereas one IgV and two or three IgC domains are present in human PSGs (Thompson et al. 1991). These observations, together with detailed "silent site" mutation analyses, imply that the CEA gene families have evolved independently in rodents and primates (Rudert et al. 1989). For these reasons, a direct assignment of individual members or subgroups of the rodent CEA gene family to their human counterparts simply by sequence comparisons has not been possible. On the other hand, transcripts of the rnCGM1 gene have been found so far only in the placenta. Therefore, rnCGM1 and closely related genes (rnCGM3 and rnCGM4) most probably represent PSG homologs (Kodelja et al. 1989; Rebstock et al. 1990).

As a basis for evolutionary and in vivo functional studies, we have isolated PSG/CEA-related genes and cDNAs from the genetically well-characterized species mouse and determined the chromosomal localization of one gene.

Materials and methods

Isolation and sequencing of mouse genomic and cDNA clones

A cosmid library from BALB/c liver DNA (Steinmetz et al. 1985) and λgt11 library constructed from RNA of a day-13 mouse placenta (Jackson et al. 1986) were plated and screened according to standard procedures (Sambrook et al. 1989). For the cosmid library the following mixture of [^{32}P]labeled genomic and cDNA probes was used: The 289-bp HindIII/BamHI fragment from λrnCGM2-1 (Rudert et al. 1989), the 661-bp PvuII/PstI fragment from λrnCGM3-1, as well as the 748-bp SstI/BamHI fragment from λrnCGM4/5-1 (Kodelja et al. 1989) and the EcoRI cDNA inserts of λrnCGM1a and λrnCGM1b (Rebstock et al. 1990). The cDNA library was screened with the 296 bp XmnI/EcoRI fragment from λrnCGM1-1 (Rudert et al. 1989). The filters were washed at low stringency in 2 × SSPE, 0.1% SDS at 60°C (1 × SSPE: 180 mM NaCl, 10 mM sodium phosphate pH 7.4, 1 mM EDTA). For isolation of full-length cDNA clones the anchored polymerase chain reaction (PCR) was used. Two μg of total RNA from day-13 placenta was reverse transcribed with avian myeloblastosis virus reverse transcriptase (Promega) with an oligo(dT)-containing oligonucleotide (5'-GGAATTCGTCGACAAGC-TT(T)$_{12}$G/C/A-3') as a primer. This primer can hybridize at the ends of all polyadenylated mRNAs because of the redundancy at its 3'-end (Leibrock et al. 1989), and it contains EcoRI, SalI and HindIII recognition sites for subcloning. CEA-related cDNAs were amplified by 30 cycles (denaturation: 1 min, 94°C; annealing: 15 s, 50°C; extension: 3 min, 72°C) in a Thermocycler 60 (bio-med) after addition of either 5'-primer FR1 (pos: 73–89, Fig. 1C) from the 5'-untranslated region, or FR2 (pos: 556–572, Fig. 1C) from the N1-domain of Cea-2 cDNA and Taq polymerase (Promega) according to Sambrook and co-workers (1989). Sequencing was performed on single- or double-stranded templates with Sequenase (United States Biochemicals) or T7 DNA polymerase (Pharmacia) with internal or universal oligonucleotide primers. The program "Clustal" from PC-GENE (Genofit), which is based on average linkage cluster analysis (Higgins and Sharp 1988), was used to construct the dendrograms. Specific primer selection for PCR was aided by the program "Primer" (Lucas et al. 1991). Ks values were determined with the program LWL85 (Li et al. 1985).

Mapping of cosmid clones

Restriction endonuclease sites on lorist B cosmids were mapped by partial digestion after conversion into phosmids, as described previously (Hohn 1983; Little and Cross 1985). Exon positions were determined by sequencing and PCR (as above), using primers from different exons of Cea-2. The distances between and positions of the exons were inferred from the size of the amplified genomic DNA fragments in combination with known internal restriction endonuclease sites. Linearized cosmid DNA (1 μg) was used as template with 5'-oligonucleotides FR24 (N2-domain, pos: 791–807), FR28 (N3-domain, pos: 1050–1066), and 3'-oligonucleotides FR25 (3'-untranslated region, pos: 1932–1948), FR27 (3'-untranslated region, pos. 1746–1762), FR29 (A-domain, pos: 1603–1619) in the following combinations: FR24/FR29, FR28/FR25, FR28/FR27, and FR28/FR29.

RNA isolation and Northern blot analyses

Placentae, fetuses, and adult tissues from STU mice (Schäfer, 1979) were pulverized in liquid nitrogen, and total RNA was extracted according to Chomczynski and Sacchi (1987). RNA blot hybridizations were performed as described (Maniatis et al. 1982).

Determination of the transcriptional start sites of the Cea-2 gene by S1 nuclease analyses

The 1.45-kb HindIII/EcoRI fragment from cosE7, which contains the first exon of Cea-2, was subcloned into the M13mp19 vector and used as template for probe synthesis. Synthesis of a complementary strand by Taq polymerase was primed with the [^{32}P]oligonucleotide 10C-L (pos: 84–107 in Cea-3 sequence). The resulting double-stranded DNA was digested with HpaII, and the 435 nc single-stranded fragment was purified on a denaturing polyacrylamide gel (Schrewe et al. 1990). Hybridization to 50 μg of total RNA from day-13 placenta and S1 analysis were performed as described by Maniatis and colleagues (1982). The size of the protected single-stranded DNA fragments was determined on a denaturing polyacrylamide gel with sequencing reaction products of the above M13 template and the 10C-L oligonucleotide as a primer.

Chromosomal localization

The chromosomal localization of Cea-5 was determined by segregation analyses in a panel of 334 [(C3H/HeJ–gld/gld × Mus spretus)F$_1$ × C3H/HeJ–gld/gld] interspecific backcross mice. The breeding and isolation of genomic DNA from these mice has been described previously (Seldin et al. 1988). The segregation of restriction fragment length variants (RFLVs) was examined in 10 μg genomic DNA samples, which were digested with TaqI, electrophoresed in 0.9% agarose gels, transferred to Nytran membranes (Schleicher & Schuell), hybridized at 65°C (Seldin et al. 1988), and washed at 65°C in 0.2× SSC. Gene linkage was determined by segregation analysis (Green 1981). Gene order was established by minimizing the number of double recombinants between genes within a linkage group and confirmed by maximum likelihood analysis (Bishop 1985).

Results

Isolation and analysis of mouse PSG-related cDNA clones

Analysis of RNA from different adult and fetal mouse tissues with N-terminal domain exon-containing genomic fragments from rat PSG-related genes rnCGM1, rnCGM3, and rnCGM4 (Kodelja et al. 1989; Rebstock et al. 1990) revealed related transcripts in the placenta
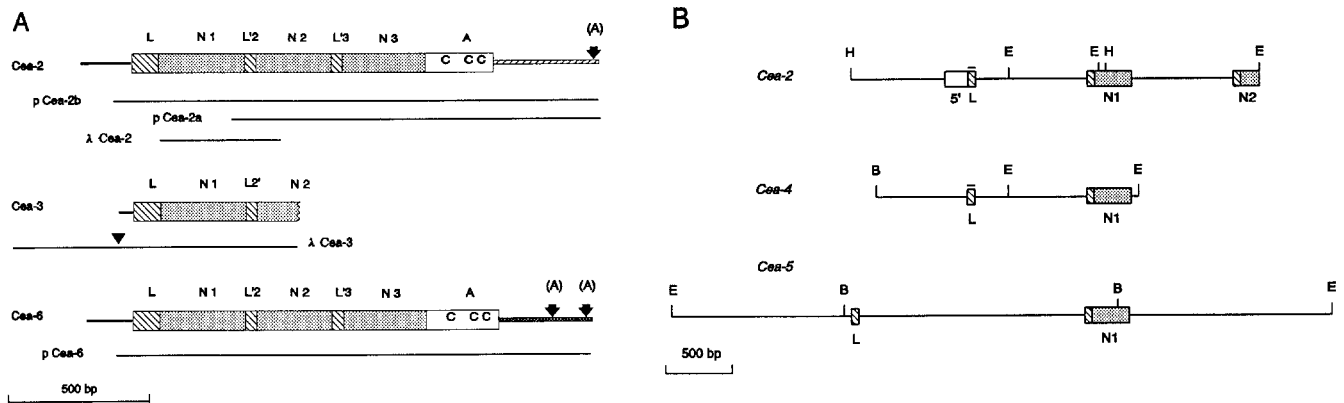
Fig. 1. Structure of cDNA (A) and genomic DNA fragments (B) of members of the murine CEA gene family and their nucleotide sequence (C). Deduced protein domains and exons are depicted as **blocks**, whereby homologous regions are indicated by the same *shading*. The sequence divergence in the 3'-untranslated regions of Cea-2 and Cea-6 mRNAs is indicated by a change of the fill-in pattern. The coding regions are composed of a leader peptide (L), N or IgV-like domains (N1–N3), interspersed leader fragments (L'2, L'3), and an IgC-like domain of the A-type (A). $5' = 5'$-noncoding region; C = cysteine in A domains. The **vertical arrows** indicate polyadenylation signal sequences. The **arrowhead** marks a possible cloning artefact in the Cea-3 cDNA clone, upstream of which no sequence homology to any of the so far analyzed CEA gene family members is found. Transcriptional start sites and domain borders are indicated by **arrows**. The 5'-untranslated region of Cea-2 cDNA has been completed with genomic sequences. The 5'-start of the longest Cea-2 cDNA is marked by an **arrowhead**. The oligonucleotide used as 5'-primer for synthesis of full length cDNAs by PCR is *underlined*. Since this sequence was derived from a genomic clone of *Cea-2*, it possibly does not reflect the authentic Cea-6 sequence in this region. The recognition sequences for potential N-glycosylation are shaded **gray**. Three **asterisks** mark the stop codons. Three **diamonds** have been inserted into the amino acid, and a **dash** into the nucleotide sequence of Cea-6, to indicate the one base pair deletion (see text). A nucleotide difference found in an independent Cea-6 cDNA clone is shown above the nucleotide sequence. Polyadenylation signal sequences are indicated by **broken lines**. The RGD-related motifs are **underlined by a bar**. A purine-rich sequence directly upstream of the transcriptional start sites in *Cea-2* is **underlined**. The multiple putative initiation codons in *Cea-5* are shown in **bold print**, the in frame stop codons are **underlined**. The approximate size of nonsequenced regions in the genomic DNA are shown in **brackets** and are taken into account for the numbering system. N, unknown nucleotide.

(data not shown). On the basis of these results, we screened a mouse day 13-placenta cDNA library (Jackson et al. 1986) with the rnCGM1 probe. Out of 10^6 recombinant phages, the DNA of 81 hybridized. The cDNA inserts from two of these clones (λ-Cea-2 and λ-Cea-3) were sequenced (Fig. 1A, Fig. 1C). When compared with the deduced domain organization of the rat PSG analog rnCGM1 (Rebstock et al. 1990), λCea-2 contains nearly the whole of the first IgV-like domain (N1) and part of the second (N2), whereas λCea-3 also covers sequences of the 5'-noncoding region and the leader (Fig. 1A). In order to gain the complete sequence for Cea-2 mRNA, the anchored PCR method was applied (see Materials and methods). The resulting DNA fragments of about 1.9 kb and 1.5 kb were subcloned and sequenced (Fig. 1A). In addition to the expected full-length and partial Cea-2 cDNA clones (pCea-2a,b), a full-length clone from a third, closely related gene (*Cea-6*) was obtained. In the latter cDNA a single nucleotide is missing in the leader domain. Since this deletion was found in two independently isolated Cea-6 cDNA clones, a PCR amplification or cloning artefact can be ruled out. Therefore, *Cea-6* is probably a pseudogene or a nonfunctional allele. In the following comparisons, the deletion was neglected. The Cea-2 cDNA has a length of 2044 bp and covers a 72-bp 5'-untranslated region, an open reading frame of 1425 bp, corresponding to 475 amino acids ($M_r = 52{,}912$; M_r minus leader $= 49{,}292$), and a 3'-noncoding sequence of 427 bp including a degenerate polyadenylation signal 15 nucleotides upstream of the poly(A) tail (Fig. 1C). The deduced protein con-

sists of a leader peptide, three related IgV-like domains (N1–N3) with 53–66% similarity at the nucleotide and 40–51% at the amino acid level and an IgC-related domain of the A-type. The N domains are separated by leader fragments (L'2, L'3) with one or two cysteine residues. Seven N-glycosylation sites are present, all of which are located in the IgV-like domains. On the assumption of an increase of the molecular weight of 4000 per site by addition of carbohydrate chains (Neumaier et al. 1988), the relative molecular mass of the mature proteins can be predicted to be about 77,000. *Cea-6* mRNA is very similar up to 40 nucleotides downstream of the stop codon; thereafter the two sequences diverge completely. When the *Cea-6* nucleotide sequence is compared with the 3'-untranslated region of *rnCGM1* mRNA, continuous homology is found with 69% of the nucleotides being identical. As the A domain exon in the *Cea-2* gene is contiguous with the 3'-noncoding region (see below) in *Cea-6* and *rnCGM1* mRNAs, splice events probably occurred in the vicinity of the stop codon. The lack of sequences matching the splice donor consensus sequence (Cech 1983) in the corresponding region of Cea-2 pre-mRNA could explain why splicing is not observed for the *Cea-2* gene in this region. The 3'-untranslated region of Cea-6 contains two polyadenylation consensus sequences. The putative Cea-6 protein shows an identical domain organization as found for Cea-2. Eight putative glycosylation sites exist, which would lead to a predicted M_r of 81,000 of the glycosylated protein. It has 60% of its amino acids in common with the Cea-2 protein.

C

Cea-2 cDNA

```
   1 GGTAGCACTGAGAGTGGAAGTGAGTGAGCAGTGCTGTGGTCACAGACCCCACCACCCACAGCCTATAGGATTCTGTGAAGTGTTCCTGCCTGAGAGAAAACTCAGCTCAGAAGGAGGAAG

 121 GACAGCAGAGGCCAGGTGCCTTGCAGGAGCTGAAGGTCTCCTCAGAGAAAGGCCAGTACAGTGAGAAGAGATATGGAGGTGTCCTCTGAGCTTCTCAGCAATGGGTGTACCCCCTGGCAA
                                                                           MetGluValSerSerGluLeuLeuSerAsnGlyCysThrProTrpGln
                                                                          ⌐► L

 241 AGGGTTCTGCTCACAGCCTCCCTCTTATCCTGCTGCCTCCTGCCCACCACTGCCAGAGTCACTGTGGAATTCTTACCTCCCCAAGTGGTTGAAGGAGAAAATGTTCTTCTACGCGTTGAT
     ArgValLeuLeuThrAlaSerLeuLeuSerCysCysLeuLeuProThrThrAlaArgValThrValGluPheLeuProProGlnValValGluGlyGluAsnValLeuLeuArgValAsp
                                                           ⌐► N1

 361 AATCTGCCAGAGAATCTTCTAGGTTTTGTGTGGTACAAAGGGGTGGCAAGTATGAAGCTTGGAATTGCACTGTATTCACTGCAATATAATGTAAGTGTGACAGGGCTTAAGCACAGCGGT
     AsnLeuProGluAsnLeuLeuGlyPheValTrpTyrLysGlyValAlaSerMetLysLeuGlyIleAlaLeuTyrSerLeuGlnTyrAsnValSerValThrGlyLeuLysHisSerGly

 481 AGAGAGACACTGCACAGAAATGGGTCCCTGTGGATCCAAAATGTCACCTCGGAGGACACAGGATATTACACCCTTCGAACCGTAAGTCAACGTGGAGAACTGGTATCAGATACATCCATA
     ArgGluThrLeuHisArgAsnGlySerLeuTrpIleGlnAsnValThrSerGluAspThrGlyTyrTyrThrLeuArgThrValSerGlnArgGlyGluLeuValSerAspThrSerIle
                                   ⌐► L2'                    ⌐► N2

 601 TTCCTTCAGGTGTACTCCTCTCTTTTCATCTGTGAGCGTCCTACTACCCTTGTACCTCCCACTATTGAATTAGTGCCGGCCAGCGTTGCTGAAGGAGGAAGTGTTCTTTTCCTTGTTCAC
     PheLeuGlnValTyrSerSerLeuPheIleCysGluArgProThrThrLeuValProProThrIleGluLeuValProAlaSerValAlaGluGlyGlySerValLeuPheLeuValHis

 721 AATCTACCAGAGTATCTTATATCACTTACCTGGTACAAAGGAGCAGTTGTGTTTAACAAGCTTGAGATTGCCCGATACAGAACAGCCAAGAATTCAAGTGTGCTTGGCCCTGCCCACAGC
     AsnLeuProGluTyrLeuIleSerLeuThrTrpTyrLysGlyAlaValValPheAsnLysLeuGluIleAlaArgTyrArgThrAlaLysAsnSerSerValLeuGlyProAlaHisSer

 841 GGTAGAGAAACGGTGTTCAGCAATGGATCCCTGCTGCTCCAGAATGTCACCTGGAAAGACACCGGATTCTACACCCTACGAACTCTGAATAGATATCCGAGAATAGAATTAGCACACATT
     GlyArgGluThrValPheSerAsnGlySerLeuLeuLeuGlnAsnValThrTrpLysAspThrGlyPheTyrThrLeuArgThrLeuAsnArgTyrProArgIleGluLeuAlaHisIle
                                  ⌐► L3'                      ⌐► N3

 961 TACCTTCAGGTGGACACCTCCCTTTCCTCGTGCTGTCACCCTCTTGACTCTCCCCAGCTCAGTATCGATCCATTGCCACCGCATGCTGCTGAAGGGGGAAGGGTTCTTCTCCAGGTCCAT
     TyrLeuGlnValAspThrSerLeuSerSerCysCysHisProLeuAspSerProGlnLeuSerIleAspProLeuProProHisAlaAlaGluGlyGlyArgValLeuLeuGlnValHis

1081 AATCTGCCAGAAGATGTGCAAACCTTTTCCTGGTACAAAGGCGTGTATAGCACTATATTATTTCAAATTGCAAAATATAGCATAGCAACGAAGTCCATCATCATGGGCTATGCACGGAGT
     AsnLeuProGluAspValGlnThrPheSerTrpTyrLysGlyValTyrSerThrIleLeuPheGlnIleAlaLysTyrSerIleAlaThrLysSerIleIleMetGlyTyrAlaArgSer

1201 AGAAGAGAGACAGTGTATACCAATGGATCCCTGCTGCTCCAGGATGTCACTGAGAAAGACTCTGGTGTGTACACACTAATAACAACAGACAGCAATATGGGAGTTGAAACAGCGCATGTG
     ArgArgGluThrValTyrThrAsnGlySerLeuLeuLeuGlnAspValThrGluLysAspSerGlyValTyrThrLeuIleThrThrAspSerAsnMetGlyValGluThrAlaHisVal
                                 ⌐► A

1321 CAAGTCAACGTCCACAAGCTTGCCACACAGCCCGTCATAAAAGCCACAGACAGCACAGTACGAGTACAGGGCTCAGTGATCTTCACTTGCTTCTCAGACAACACTGGGGTCTCCATCCGT
     GlnValAsnValHisLysLeuAlaThrGlnProValIleLysAlaThrAspSerThrValArgValGlnGlySerValIlePheThrCysPheSerAspAsnThrGlyValSerIleArg

1441 TGGCTCTTCAACAATCAGCGTCTGCAGCTCACAGAGAGGATGACCCTGTCCCCGTCAAAGTGCCAACTCTGGATACGTACTGTGAGGAAGGAGGATGCTGGAGAGTACCAATGTGAGGCC
     TrpLeuPheAsnAsnGlnArgLeuGlnLeuThrGluArgMetThrLeuSerProSerLysCysGlnLeuTrpIleArgThrValArgLysGluAspAlaGlyGluTyrGlnCysGluAla

1561 TTCAACCCAGTCAGCTCAAAGACCAGTCTCCCAGTCATCCTGGCTGTGATGATCGAGTGACCCATCCTCTCATGCTATAGCAGGGCGGGGCATTTCTGTATTGAGATGTCCACACCTACC
     PheAsnProValSerSerLysThrSerLeuProValIleLeuLeuAlaValMetIleGlu***

1681 CACCCCTGCCAGTTATAGATTTCCATTCATTCGCATGGCCATTTGGCCATCTCTCCCATCCCTCCCCACACCTGATCCTGAACCCCTCCATTTTCTCCCACATCCCCTCTCCTACCCAGT

1801 TCCATTCCTCCATCTGCCTTTTACTACTATTTTTCCCCTTCTACATGAGATTCAAGCATCCTCCCTCGGGCTTCCAACTTGTTTAACTTTTTTGTATCTGTGTAGTGTAGTGTGTGTATC

1921 TTGTATTTTATGGCTAATATCCAGTTATAAGTGAGTATATACCATGCATGCCATTTTGGAACCTGGTTACCTCACTCAGGATGATACTCTCAAGTTCTATCCAAATTAATGATGCCTTCA

2041 AATC
```

Cea-3 cDNA

```
   1 GCCTTGCAGGAGCTGAAGGTGTTCTCCTCAGAGAAAGGCCAATATAGTGAGAAGAGATATGGAGTTGTCATCTGAACTTTTCAGCAACGGGTGTACCTCCTGGCAGAGGGTTCTCCTCAC
                                                           MetGluLeuSerSerGluLeuPheSerAsnGlyCysThrSerTrpGlnArgValLeuLeuThr
                                                          ⌐► L

 121 AGCCTCCCTCTTAACCTGCTGGCTCCTGCCCACCACTGCCAGAGTCACCATTGAATCTTTACCACCCCAAGTGTATGAAGGAGAAAATGTTCTTCTACGTGTTGACAATATGCCAGAGAA
     AlaSerLeuLeuThrCysTrpLeuLeuProThrThrAlaArgValThrIleGluSerLeuProProGlnValTyrGluGlyGluAsnValLeuLeuArgValAspAsnMetProGluAsn
                   ⌐► N1

 241 TCTTCTAGTGTTTGGCTGGTACAGAGGAATGACAAATTTGTGGCAAGCAATTGCACAGCATTGGCTGTACTACTATAGTGTAATGGTGAAGGGGCTGAATCACAGCGGCAGAGAGATATT
     LeuLeuValPheGlyTrpTyrArgGlyMetThrAsnLeuTrpGlnAlaIleAlaGlnHisTrpLeuTyrTyrTyrSerValMetValLysGlyLeuAsnHisSerGlyArgGluIleLeu

 361 ATACATCAACGGGTCCCTGTGGATCCAAAATGTCACACAAGAGGACACAGGATATTACACTTTTCAAACCATAAGTAAACGAGGAGAAATAGTATCAAATACATCCCTGTACTTGCACGT
     TyrIleAsnGlySerLeuTrpIleGlnAsnValThrGlnGluAspThrGlyTyrTyrThrPheGlnThrIleSerLysArgGlyGluIleValSerAsnThrSerLeuTyrLeuHisVal
                  ⌐► L2'                   ⌐► N2

 481 GTACTCCTCTCTTTTCATCTGTGTGCGTCCTACTACCCTTATATCTCCGACTATTGAATTAGTGCCAGCCAGCGTTGCTGCAGGGGGAAGCATTCTTCTCCTTGTTCACAATATTCCAAA
     TyrSerSerLeuPheIleCysValArgProThrThrLeuIleSerProThrIleGluLeuValProAlaSerValAlaAlaGlyGlySerIleLeuLeuLeuValHisAsnIleProLys

 601 GTATCTTCAATCGCTTTTCTGGTACAAAGGGCTGATTGCATTTAACAAGGTTGAGATTGCTCGATACAGAACAGCCAAGAATTC
     TyrLeuGlnSerLeuPheTrpTyrLysGlyLeuIleAlaPheAsnLysValGluIleAlaArgTyrArgThrAlaLysAsn
```

Isolation and analysis of mouse CEA/PSG-related genes

In parallel to the isolation of PSG-related cDNAs, we screened a mouse cosmid library for CEA/PSG-related genes. First, a pool of candidate clones was identified by hybridization at low stringency with a mixture of probes from rat CEA/PSG-related genes (*rnCGM1*, 2, 3 and 4; Kodelja et al. 1989) and a cDNA (*rnCGM1*; Rebstock et al. 1990). Seventy-five out of 3×10^5 clones analyzed hybridized. Six of these clones gave a strong signal with the N1 exon probe from *rnCGM1*,

Cea-6 cDNA

```
   1 CTGTGAAGTGTTCCTGCCTGAGAGAACACTCAGATCAGAAGAAGGAAGGACAGCAAATGCTGGTGCTGTGTTGGACGGGAAGTTCTTCTTCTCAAAGAGAGGACAGTACAGTGAGAAGAG
       r+ L                       T.                                                                                       r+ N1
 121 AGATGGAGGTGTCCTTGGTGCTTCTCTACCA-GGGTGTACACCCTGGCAGGGTCTTCTGCTCACAGCTTTTCTTTTAACCTCTTGGTTCCTGCCCACCACTGTCCAAGTCACCATTGAAT
      MetGluValSerLeuValLeuLeuTyr+++GlyCysThrProTrpGlnGlyLeuLeuLeuThrAlaPheLeuLeuThrSerTrpPheLeuProThrThrValGlnValThrIleGluLeu

 241 TAGTGCCACCACAAGTGGCTGAAGGAGAAAATGTCCTTATTATTGTTTACAGTCTGCCAGAGGATCTTACAGCCATAGCCTGGTTCAAAGGAGTGACAAATATGAACCTCGGAATTGCAT
      ValProProGlnValAlaGluGlyGluAsnValLeuIleIleValTyrSerLeuProGluAspLeuThrAlaIleAlaTrpPheLysGlyValThrAsnMetAsnLeuGlyIleAlaLeu

 361 TGTATGCACTGGCCTCTAACATCAGTGTGAAAGGGCCCGAACACAGTGGTAGAGAGACAGTGTTCAGCAATGGATCCCTGTTGCTTCACAATGTCACCCAGAAGGACACAGGATTCTATA
      TyrAlaLeuAlaSerAsnIleSerValLysGlyProGluHisSerGlyArgGluThrValPheSerAsnGlySerLeuLeuLeuHisAsnValThrGlnLysAspThrGlyPheTyrThr

 481 CTATACGGACCTTAAATAGACATGGAAAAATTGTATCCACAACATCCATATACCTCCACGTGTACACCTTCTTCTGGACTTGTGGACCTCTTTCCCCCTCTGCCCATCTCACTATTGAAT
      IleArgThrLeuAsnArgHisGlyLysIleValSerThrThrSerIleTyrLeuHisValTyrThrPhePheTrpThrCysGlyProLeuSerProSerAlaHisLeuThrIleGluSer

 601 CAGTGCCACCCAAAGTTGCCGAAGGGGGAAGTGTTCTTCTAGTTGTTCACAATCTCCAGGGGAATCTTCGATCCCTTTTCTGGTACAAAGGGATGATTGTGTCCAGGAACCTTGAGGTTG
      ValProProLysValAlaGluGlyGlySerValLeuLeuValValHisAsnLeuGlnGlyAsnLeuArgSerLeuPheTrpTyrLysGlyMetIleValSerArgAsnLeuGluValAla

 721 CAAGCACACATAATAGACACAAATTTAAGTGTGCATGGGCCTTTACACAGTGGTAGAGAGACAATATACAGCAATGGATCCCTCATGTTCTACAATGTGACATGGAAGGACAGTGGATTAT
      ArgHisIleIleLeuAspThrAsnLeuSerValHisGlyProLeuHisSerGlyArgGluThrIleTyrSerAsnGlySerLeuMetPheTyrAsnValThrTrpLysAspSerGlyLeuTyr
                                                                                                r+ L3'                    r+ N3
 841 ACACCCTACGAACTCTAAGTACAGATATGAAAACAGAATTAGCACATGTACAACTCCAGGTGGACGCTTCCCTTTCTCCGTGCTGTAATCCCCTTACCTCTTCTCAGCTCATGATTGAAC
      ThrLeuArgThrLeuSerThrAspMetLysThrGluLeuAlaHisValGlnLeuGlnValAspAlaSerLeuSerProCysCysAsnProLeuThrSerSerGlnLeuMetIleGluPro

 961 CAGTGCCACGCTATGCTGTTGAAGGAGAAAGTGTTCTTTTCATGGTTCATAACCTGCCAAAAGATCTTCAAACCTTTTCCTGGTACAAATCAGTGTATGGGGCTGAGATATTAAAAATCA
      ValProArgTyrAlaValGluGlyGluSerValLeuPheMetValHisAsnLeuProLysAspLeuGlnThrPheSerTrpTyrLysSerValTyrGlyAlaGluIleLeuLysIleThr

1081 CAGAATACAGCAGAGCCATGAGTTCCACCACACGTGGAAGTGAACTCAAAAGAAGAGAGCGTGTGTACACCAATGGATTCCTCCTGCTCCAGAGTGCCACTGAGAAAGATGCAGGAATGT
      GluTyrSerArgAlaMetSerSerThrThrArgGlySerGluLeuLysArgArgGluArgValTyrThrAsnGlyPheLeuLeuLeuGlnSerAlaThrGluLysAspAlaGlyMetTyr
                                                              r+ A
1201 ACATACTAGAAACTTTAAGCAGAGATTTCAAAATTGAAAAAGCACAGGTGCAATTATATGTAAACAAGCCTATATCACAGCCCTTCATTCGAGTCACCAACACCACGGTCACAATACAGA
      IleLeuGluGluThrLeuSerArgAspPheLysIleGluLysAlaGlnValGlnLeuTyrValAsnLysProIleSerGlnProPheIleArgValThrAsnThrThrValThrIleGlnSer

1321 GCTCTGAGGTCCTCACCTGCCTCTCAGCTGACACTGGAATCTTCATCCGTTGGATCTTCAATAACCAGAGTCTGCAGCTTACAGAGAGGATGACACTATCCCCAACAAAGTGCCAACTCA
      SerGluValLeuThrCysLeuSerAlaAspThrGlyIlePheIleArgTrpIlePheAsnAsnGlnSerLeuGlnLeuThrGluArgMetThrLeuSerProThrLysCysGlnLeuSer

1441 GCATAGATCCTGTCAGGAAGGAGGATGCTGGAGAGTACCAATGTGAGGTCTCCAACCCAGTCGGTTCACAGACCAGTCTCCCAGTCAGCCTGGCTGTGACAAATGAGTGACACTTCCTTT
      IleAspProValArgLysGluAspAlaGlyValTyrGlnCysGluValSerAsnProValGlySerGlnThrSerLeuProValSerLeuAlaValThrAsnGlu***

1561 CATCCTATAAAGGAGTGGGGGCATTTCTTTATTGATAAGGTACAAAATGGAGTAAAGTTATGGTGAAAATTGTCAGTTACCACTCAGGTACAGCCACCATTGTGAGTCATGTCTGTAATG

1681 CTGGGATAATTGTGTATATAAGGACAGACCAGGACCTAAAATGAGACTCTGTCTCCAAAGGAAAACAGACATCAATGTAATAAACAGAAATACAAAGGCATCAAGGATTTAAGTTGTGGC

1801 TCAAATATATAGCCAGTCATCAGAATACATGGCAATCCAATTTTGAGAGCCCCCAAACTTTGAGTCTGCTGTTTAAATGTTGCAGCGTTTGAATAGAAATAAATGCTCCTGTCATATACC
```

whereas one clone (cosC3) hybridized with the N-terminal domain exon probe from *rnCGM2*. This gene is a more distantly related member of the rat CEA gene family. cosC3 and two of the clones, positive with the *rnCGM1* probe (cosE7, cosG11), were characterized in detail. The genomic DNA inserts of cosE7 and cosG11 were found to overlap extensively, covering a 51-kb genomic region (Fig. 2). DNA fragments containing exons 1 and 2 were identified by hybridization, then subcloned and partially sequenced (Fig. 1B,C). These analyses revealed the presence of two tandemly arranged genes in the overlapping cosmids cosE7 and cosG11. One gene coded for Cea-2; the second, located 16 kb downstream, was a new member of the CEA gene family (*Cea-4*) and closely related to *Cea-2* (Fig. 2). Their second exons have 84% of their nucleotide and 72% of their amino acid sequences in common. The close relationship of both genes is also observed in the region 5' of the translational start and the first intron, where similarities of 95% and 92% respectively are found. The complete exon organization of *Cea-2* was determined. The first exon encodes the 5'-untranslated region and two-thirds of the leader peptide. The following three exons each code for the last third of the leader or leader-like sequences (L'2, L'3), respectively, and for individual N domains. The IgC-like domain and the 3'-untranslated region found in the

Cea-2 cDNA clones are encoded by the last exon. The whole Cea-2 transcription unit covers a region of 7.5 kb. Cosmid cosC3 covers 34.5 kb of genomic DNA that contains another new member of the murine CEA gene family (*Cea-5*). Although *Cea-5* shows, as far as determined, the same overall genomic organization, it is more distantly related to *Cea-2* and *Cea-4*. This is exemplified by the larger first intron and the different arrangement of the restriction endonuclease sites. Furthermore, its exon 2 sequence reveals only 55% similarity of its nucleotides and 40% of its amino acids with the corresponding sequences of *Cea-2*. However, it shows a similarity of 93% and 92% at the nucleotide and amino acid level, respectively, with the rat CEA-related gene *rnCGM2*. The rates of synonymous substitutions in the N domain exons of *rnCGM2* and *Cea-5* (Ks = 0.211 ± 0.06) indicate that these genes diverged ~16 × 10^6 years ago, close to the time of rat/mouse speciation (Bulmer et al. 1991), so they probably represent orthologous genes. The translational start of the putative Cea-5 protein is ambiguous because of the presence of four potential in-frame ATG start codons (Fig. 1C). The first two are soon followed by in-frame stop codons and, therefore, are irrelevant for initiation of translation of the putative Cea-5 precursor. The fourth ATG codon (pos. 172–174) displays a sequence context closest to the trans-
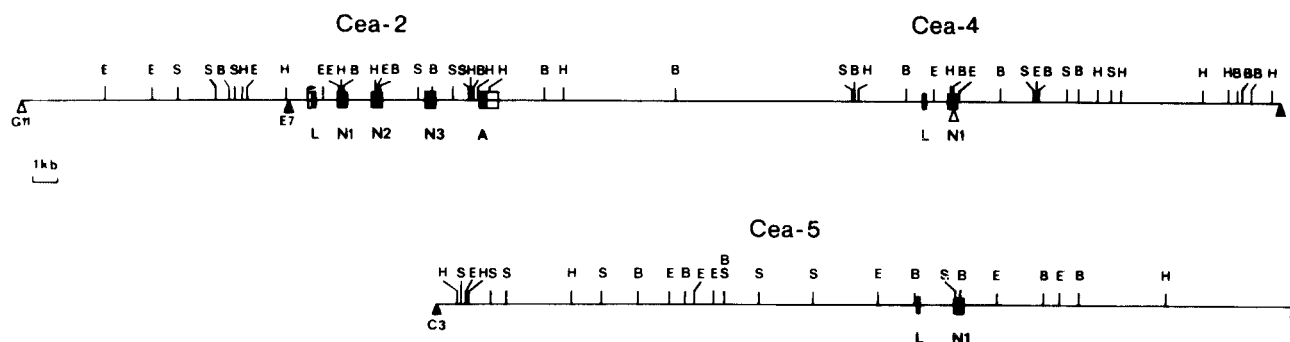
Cea-2

```
   1 GATCAGGGCAGGTCATATTTGTGGTGTTGAACNCCCCNNCTGTAGCCCTAGGCATGAGACTTCAGTTCTCCCTGAGAACTCCAGGAAGTTCAGCAACTCATGTTGCTCACTCAACACTAG

 121 GAACATTGTTTCCTTGAGCTAGGAGACCTTGTCTAGGAGTCTGCTTTCCATTCCTGTNCCCACTAACCTTGATTTCCCTTCCTAGACCTTGTTCTGTCTCTTGTGTCTGAGAAGACCCCT

 241 CTCTTTCTGGAGAACCTGAAAGTCTCCTCCCCACTGGCATCTCCCATAGTACAAGACAGGATTGCTGATGGGGAATCACCCAGAAAAGGGCAGAGCTGACCCAGGAAGACCAGCGAGAAC

 361 TGTGCTCAGCCAACACACAGCACAGCCAACTCATAAGGCTGCAAGGTCCATGGGCACACAGGTCACCTCTTCCCTAACACCAAGAACATGAAATTCCTCACCCAGCTGTTCCCGGGCCTC

 481 TGTCTTGGGTCAGGATCTTGACTGGTGACATTAGTGACAGAGGGTTTATCTTCATCCCCCAATCATCACATAACTGGTCCCTTCCAAACCTCACAGACATCTGTGCTACCCTCCTGCTGG

 601 GGAATCCACCTTCCCAGAGCACTGAGGACACAGGGCAGACTGGGTGCTCAGCCGGGTCTGTGTGTCACAGAGACACGTAGGAAAGTTGGAGGCTCCCTCTTTGATGCTGACAGCCTCATG

 721 TCCAGGAGTCAGCAGATGACAGCCCTTAGATGAGTTGTTCTTTGAGGGCATGTCGATGCTTATCAGCCTTGTTGCTCAAGGTCTTGTTGGAGAAGAAACATAAAGAGAGAAAATGTGAGA

 841 TACAGGGTAGCACTGAGAGTGGAAGTGAGTGAGCAGTGCTGTGGTCACAGACCCCACCACCCACAGCCTATAGGATTCTGTGAAGTGTTCCTGCCTGAGAGAAAACTCAGCTCAGAAGGA

 961 GGAAGGACAGCAGAGGCCAGGTGCCTTGCAGGAGCTGAAGGTCTCCTCAGAGAAAGGCCAGTACAGTGAGAAGAGATATGGAGGTGTCCTCTGAGCTTCTCAGCAATGGGTGTACCACCT
                                                                                           MetGluValSerSerGluLeuLeuSerAsnGlyCysThrThrTrp

1081 GGCAAAGGGTTCTACTCACAGGTAAGTGTTCCAGCTATCTGATTGTGTGTGGGAAGGGAAACTCAGAAGTCCAAAGTATTCTGTAGAGGAGATGTCTAGAGAAGCCTTTGATTTCTGTTT
     GlnArgValLeuLeuThrA

1201 GTAGTCATGGAGCAGAAGGTGCAGTGAGGGACAGAGGCTCAGCAGCAGGACCCTCAGAGCAGACAGGAGCTGAGGAGGAAAAGAAGAGCCTCAGCTGTCCCTGTTCAAAGTTAGTCACAT

1321 TGGCTGCCATGGTCTGAAGGCCAATGTTCCTAACCATGACAGGGTGACTCTCCAAGTAGAAATTAAACCTGGGAATGCCAGAGAGATGGCTCAGTGTGTGGAGACATGCCAGTCTGAATT

1441 CATTCCTCAGCTCTCACAGGGCAGATGGAGATACAGGCTCCTGCACGATCTTCTCTGCCCTCCACACAGGTGTTATGCACCCACAATGTGCATAGTGAGACAGGAGCCTTCCCAAACTTG

1561 TGCATCCAATACAGAATTATGGAATCACACTTGAACATAATGATTGCCATTTGTTCCAAACCTGGGACTATATAGATAAATCCATGGAGATATCCTTCCTTCCTTCCTTCCTTCCTTCCT

1681 TCCTTCCTTTCCTTTGATTTCTTCCCTTCCTTT---(0.32 kb)---ATTCCTCAGAGAGGAAATCATACCCACTGAAGCTCTGAGGGCGTGAGGGCACGATGCTCACCTCTATTATGG
                                                                                                                    r→ N1
2121 TCAAACCCTTACCCACAGGCCTAAGCTGAGGGATGCTCACACCTGCTCAGGATTGGTGGCTTCTCTCTTTTCCCTTCTAGCCTCCCTCTTATCCTGCTGCCTCCTGCCCACCACTGCCAG
                                                                                                      laSerLeuLeuSerCysCysLeuLeuProThrThrAlaArg

2241 AGTCACTGTGGAATTCTTACCTCCCCAAGTGGTTGAAGGAGAAAATGTTCTTCTACGCGTTGATAATCTGCCAGAGAATCTTCTAGGTTTTGTGTGGTACAAAGGGGTGGCAAGTATGAA
     ValThrValGluPheLeuProProGlnValValGluGlyGluAsnValLeuLeuArgValAspAsnLeuProGluAsnLeuLeuGlyPheValTrpTyrLysGlyValAlaSerMetLys

2361 GCTTGGAATTGCACTGTATTCACTGCAATATAATGTAAGTGTGACAGGGCTTAAGCACAGCGGTAGAGAGACACTGCACAGAAATGGGTCCCTGTGGATCCAAAATGTCACCTCGGAGGA
     LeuGlyIleAlaLeuTyrSerLeuGlnTyrAsnValSerValThrGlyLeuLysHisSerGlyArgGluThrLeuHisArgAsnGlySerLeuTrpIleGlnAsnValThrSerGluAsp

2481 CACAGGATATTACACCCTTCGAACCGTAAGTCAACGTGGAGAACTGGTATCAGATACATCCATATTCCTTCAGGTGTACTGTAAGTAATTCTTT---(0.5 kb)--CCTCAGCTGAGCT
     ThrGlyTyrTyrThrLeuArgThrValSerGlnArgGlyGluLeuValSerAspThrSerIlePheLeuGlnValTyrS
                                                                                      r→ L2'
2601 ATGACACAGCTGCTCAGACACAGGCGCCATCTTAGTCAAAAGAAGAAAAATCCTAATATTGATGGATCTCTCTTCCCTTTTAGCCTCTCTTTTCATCTGTGAGCGTCCTACTACCCTTGT
                                                                                      erSerLeuPheIleCysGluArgProThrThrLeuVal
     r→ N2
2901 ACCTCCCACTATTGAATTAGTGCCGGCCAGCGTTGCTGAAGGAGGAAGTGTTCTTTTCCTTGTTCACAATCTACCAGAGTATCTTATATCACTTACCTGGTACAAAGGAGCAGTTGTGTT
     ProProThrIleGluLeuValProAlaSerValAlaGluGlyGlySerValLeuPheLeuValHisAsnLeuProGluTyrLeuIleSerLeuThrTrpTyrLysGlyAlaValValPhe

3021 TAACAAGCTTGAGATTGCCCGATACAGAACAGCCAAGAATTC
     AsnLysLeuGluIleAlaArgTyrArgThrAlaLysAsn
```

lational start consensus (Kozak 1989). Usage of the third initiation codon (pos. 157–159), although obviously less favorable because of the lack of a purine at position −3, would lead to a leader peptide of 34 amino acids, a size that is predicted for all CEA-related precursor proteins characterized so far.

Classification of rodent CEA/PSG-related genes

The derived amino acid sequences from the second exons of the newly characterized and previously described members of the murine and rat CEA families were compared and displayed as a dendrogram (Fig. 3A). These exons were chosen because they encode domains of the same type (IgV-like) in all known rodent and human CEA family members. The rodent CEA-related genes can be subdivided into three groups. The first contains the murine genes mmCGM1 and mmCGM2 and the rat gene ectoATPase (subgroup 1). The tightly coupled mouse genes Cea-2 and Cea-4, as well as Cea-3 and Cea-6, can be grouped together with the rat genes rnCGM1, rnCGM3, and rnCGM4

(subgroup 2). Cea-5 and rnCGM2 represent the third group (subgroup 3). Nucleotide sequence comparisons show mean similarities of 78%, 78%, and 93% within subgroups 1, 2, and 3, respectively, and between subgroups, from 54% to 61%. Multi-alignment of the amino acid sequences encoded by exon 2 of human and rodent CEA-related proteins demonstrated that none of the rodent subgroups can be assigned to any of the human CEA family subgroup, which form separate branches in the dendrogram (Fig. 3A). Sequence analyses of all N domains of rodent subgroup 2 members revealed that they can be subdivided into two main groups: the first consists of all N1 domains, while the second contains the remaining ones (Fig. 3B). Within the latter group, N domains with the same position number are most similar. This is also true for the five N domains of rnCGM1 if one takes the recent duplication of N2 and N3 into consideration (Rebstock et al. 1990). In order to identify regions that are possibly important for the function of members of the rodent subgroup 2, their N domain amino acid sequences were compared for conserved motifs present in only

Cea-4

```
   1 GGATCCTGAAATTTCCCTCCCTCTTGGAATCTCCCAAGGGGAAGACAGGACTCTTGCTGAAGAATCACCCAGACAAAGGGCATAGGTGACCCAGGAGGACCAGAGAGCAGTTGGGCTCAG

 121 CAAACACACAGCCCAATTAACCCATAATGCTGGAAGGTTGCCTGGAAATAGTCCATAGGAGAGGAGCCCACCGTCCTTTCTAGGACACACAGGTCACCTATTCCCCAACACCAAGAACAT

 241 GAAACTCCTCACACAGCTGCTCCCGGGCCTCTGTCTTGGGTAGGATCTTGGCTGGTGGCATTAGTGATAAAAGGTTTATCTTCATATCCAGTCATCAGACAACTGGTCCCTTCCAACCTT

 361 CACAGACATCTGTCACCTTCCTGCTGGGAAATCCCACCTTCCCAGAGCACTGAGGACACAGGCCAGACTGGGTGCTTAGCTGGGTCTGTGTGTCTCAGGGACACATAGGAAAGATAGAGG

 481 CTCCCTTTTTGATGCTGACAGACTCATGTCCAGGAGTCAGCAGATGTCAGCCCTTGGATGAGTTGTTGTTCTCGGAGGGCATGTTGATGCTTATCAGCCTTGTTGCTCAAGGTCTTGTTG

 601 GAGAAGAAACATAAAAAGAGAAAATGTGAGAAACATGCAGGGCAACACTGAGAGTGGAAGTGACTGAGCAGTGCTGTGGACACAGACCCCACCACTCACAGCCCATAGGATTCTGGGAAG

                                                                                                              ┌─► L
 721 TGCTCCTGCCTGAGAGAACACTCAGCTCAGAAGGAGGAAGGACAGCAGAGGCCAGGTGCCTTGCAGGAGCTGAAGGTTTCCTCAGAGAAAGGCCAGTACAGTGAGAAGAGATATGGAGGT
                                                                                                                   MetGluVal

 841 GTCCTCTGAGCTTCTCAGCAATGGGTGGACCTCCTGGCAGAGGGTTCTGCTCACAGGTAAGTGTTCCAGCTATCTGATTGTGTGTGGGAAGGGAAACTCCGAAGCCCAAAGTCTTCTGAA
     SerSerGluLeuLeuSerAsnGlyTrpThrSerTrpGlnArgValLeuLeuThrA

 961 GAGGAGATGTCTAGAGAAGCCTTTGATTTCTGTTTGTAGTCATGGAGCAGAAGGTGCAGTGAGGGACAGAGGCTCAGCAGCAGGACCCTCAGAGCAGACAGGAGCTGAGGAGGGAAAGAA

1081 GAGCCTCAAGCTGTCCCTGTTCAAAGTTAGTCACATTGGCTGCCATGGTCAGAAGGCCAATGTTCCTAACCATGACAGGGTGATTCTCCAAGTAGAAATTAAACCTGGGAAAGCCAGGGA

1201 GACGGCTCAGTGTGTGGAGACATGCCAGTCTGAATTCCATTCCTCAGCTCTCACAGGGCAGATGGAGATACAGGCTCCTGCAGGATCTTCTCTGCCCTCCATACAGGTATGTGCTATGCA

1321 CCCACAATGTGCACAGTGAGACATGAATCTTCCCAAACTTGTGCATCCAATACAGAATTATGAAATCACACTTGAAAGTAATGATTACCATTTGTTCCAAACCTGGGACTATATAGATAA

1441 ATCCATGGAGATATCCTTCCTTCCTTCCTTCTTTCCTTCCTT--(0.45 kb)--GCGCTCAGATGTGCTTCCCTTCCCTCTAGCCTCCCTCTTAACCTGCTGGTTCCTGCCGATCACTG
                                                               laSerLeuLeuThrCysTrpPheLeuProIleThrAla

        ┌─► N1
2011 CCCGAGTCACCATCGAATCCGTACCACCCAAATTGGTCGAAGGAGAAAATGTTCTTCTACGAGTGGACAATCTGCCAGAGAATCTTCGAGTCTTTGCCTGGTACAGAGGGGTAATAAAAT
     GlyValThrIleGluSerValProProLysLeuValGluGlyGluAsnValLeuLeuArgValAspAsnLeuProGluAsnLeuArgValPheAlaTrpTyrArgGlyValIleLysPhe

2131 TTAAGCTTGGAATTGCACTGTATTCACTGGACTATAACACAAGTGTGACAGGACCTGAGCACAGTGGTAGAGAGACATTGCACAGCAACGGGTCCCTGTGGATCCAAAGTGCCACCCGGG
     LysLeuGlyIleAlaLeuTyrSerLeuAspTyrAsnThrSerValThrGlyProGluHisSerGlyArgGluThrLeuHisSerAsnGlySerLeuTrpIleGlnSerAlaThrArgGlu

2251 AAGACACAGGATATTACACGTTTCAAACCATAAGTAAAAATGGAAAAGTGGTATCAAATACATCCATGTTCCTTCAGGTGTACTGTAAGTAATTCTTTGTGA
     AspThrGlyTyrTyrThrPheGlnThrIleSerLysAsnGlyLysValValSerAsnThrSerMetPheLeuGlnValTyr
```

Cea-5

```
   1 ACAGACCTCCAGCAAGAAGAGTTGACGCAGATCACAGATCATAGCCTTGGGCAGATTCCCTGCACACAGCCCATGAGACTCTGGGAGATGCCCCTGGGGTAAGGACAGGAGAAATAAGAG

                                           ┌─► L
 121 GGAAGGATCCACAGGGCAGACACCAGGACAGCACACATGGGAGCAGAGACCATGGAGTCCCCTTCTCTGTTTCTCTGCAAAGGGCTCCTGCTCACAGGTGAGGCCAGCCCTGGAGTGGTG
                                           MetGlyAlaGluThrMetGluSerProSerLeuPheLeuCysLysGlyLeuLeuLeuThrA

 241 TGGGGAGGATGCTCTGACAGGAAGGGACTCTGGGAGGACCCTCTGCTTCAGGTTGAGAGGGGAGGACACGGAGGGACCCAGGCTGCTCAGGTGGATGCTCT---(1.5 kb)---GATAT

                                                         ┌─► N1
1861 TTATAGGATTTTCTCTCCTAGCCTTCCTCTTAACCTGCTGGAATGCACCCGCCGCTGCCGAGCTCACTATTGAATTAGTGCCACCCATGGTTGCCGAAGGCGGAAATCCGTTCTGTTTG
     laPheLeuLeuThrCysTrpAsnAlaProAlaAlaAlaGluLeuThrIleGluLeuValProProMetValAlaGluGlyGlyAsnSerValLeuPheVal

1981 TGCATGAAATGCCACTGAACGTCCAGGCGTTTTACTGGTACAAACAGAGAGATTCGACGAAGAGCTACGAAGTCGCACGGTACTTAACACCCACGAACCAAAGTTCGAAGATGCCTCAGC
     HisGluMetProLeuAsnValGlnAlaPheTyrTrpTyrLysGlnArgAspSerThrLysSerTyrGluValAlaArgTyrLeuThrProThrAsnGlnSerSerLysMetProGlnHis

2001 ACAGTGATAGGAAAACCGTATTCTACAGTGGATCCCTGCTGATCAGAAACGTCACCAAGGCTGACAGTGGAGTCTATACCTTACTAACATTTAACACGGAAATGGAAAGCGAATTAACAC
     SerAspArgLysThrValPheTyrSerGlySerLeuLeuIleArgAsnValThrLysAlaAspSerGlyValTyrThrLeuLeuThrPheAsnThrGluMetGluSerGluLeuThrHis

2121 ATGTGCATCTGGAAGTGCAGGGTAAGTGGTTGCGTGACCTCTGGGTGCTGGGGGGCGG
     ValHisLeuGluValGln
```

one of the two main N domain groups. The most prominent motif specific for the N1 domains is located near the carboxy-terminal end (Fig. 1C) within a region of low sequence conservation between subgroups. The motif consists of a positively charged amino acid (Arg, His), followed by a glycine and a positively or negatively charged amino acid (Lys, Glu). A similar motif is present at the same position in nearly all members of the PSG subgroup. The sequence motif in human PSGs is mostly ArgGlyAsp (LysGlyAsp in PSG1; Khan and Hammarström, 1990).

Expression of CEA/PSG-related genes

In order to achieve assignment of mouse CEA/PSG-related genes to their human counterparts, we com-

pared expression patterns, assuming that mRNAs of analogous genes have a similar tissue distribution. Northern blot analyses were carried out under stringent conditions with total RNA from placenta (days 11, 13, 16, 17, and 18 of gestation) and from colon, uterus, ovary, liver, kidney, and fibroblasts of adult mice, with the cDNA insert of λCea-2 as a probe (Fig. 1A). We found a 2.1-kb transcript only in placentae from day-13 of gestation onwards (Fig. 4, lanes 2 and 10, and data not shown). An mRNA of the same size was also detected with a Cea-3 cDNA probe in RNA of day-13, but not in day-11 mouse placenta (Fig. 4, lanes 3, 4, 11, and 12). At low stringency, weak cross-hybridization with a 3.3 kb mRNA species was detectable in day-11 and day-13 placental RNA with both probes (Fig. 4, lanes 1–4). A Cea-4 genomic probe hybridized under low stringency conditions with a 3.3-kb

Cea-2

Cea-4

Cea-5

Fig. 2. Restriction map of *Cea-2*, *Cea-4*, and *Cea-5*. The following restriction endonuclease sites were used to map the genomic DNA inserts of cosmid clones G11, E7, and C3: *Bam*HI (B), *Eco*RI (E), *Hind*III (H), *Sst*I (S). For G11 and E7 only the *Eco*RI sites confirmed by sequencing and subcloning have been included. Triangles of the same shading indicate the borders of genomic DNA fragments of individual cosmids. **Boxes** represent exons; coding regions are shown as **filled in**, untranslated regions as **open** boxes. For abbreviations see legend to Fig. 1.

mRNA species strongly in day-11 and weakly in day-13 placental RNA. Strong cross-hybridization with the 2.1-kb mRNA species in total RNA of day-13 placenta was observed (Fig. 4, lanes 5 and 6). Under stringent conditions, all hybridization signals with this probe disappeared (Fig. 4, lanes 13 and 14). Therefore, the 3.3-kb mRNA is probably derived from a gene closely related to *Cea-4*. As a control for RNA loading, identical samples were separated on the same gel and hybridized with a human β-actin probe (Fig. 4, lanes 7, 8, 15, and 16). The 2-kb *Bam*HI DNA fragment of the *Cea-5* gene covering the first and part of the second exon (Fig. 1B and 2) hybridized at low stringency (2× SSPE, 55°C) with a 3.4-kb mRNA in total RNA from NIH/3T3 mouse fibroblasts and whole day-11 and day-13 embryos (data not shown).

Sequence comparisons of putative promoter regions of the CEA gene family

S1 nuclease analyses revealed a cluster of transcription initiation sites at positions −187, −190, −191, and −192 relative to the start codon of *Cea-2* (data not shown). No classical TATA or CAAT promoter sequence motifs (Cordon et al. 1980) could be detected at the expected positions. However, a purine-rich region was noted directly upstream of the transcription initiation sites (Fig. 1C). To identify common promoter or regulatory elements, we compared the regions upstream of the translational start of all human, rat, and mouse CEA gene family members where the transcription start sites have been determined (Fig. 5). Despite low overall sequence conservation in the regions compared, a conserved 50-nucleotide stretch is visible surrounding the transcriptional starts of *CEA*, *NCA*, *PSG1*, and *PSG5*. In *Cea-2* and *rnCGM1*, however, the conserved region is located in the middle of the 5'-untranslated region. The 3'-half of this stretch is formed by an especially well-conserved, purine-rich sequence. In the vicinity of the transcriptional start sites, repeated sequence motifs ("transcriptional start motifs" TSM1 and TSM2) with two consensus sequences are found: the heptanucleotide sequence PyTGAGAPu (TSM1), which is present three times in the conserved region, and a related hexanucleotide PuCAGAPu (TSM2), where apparently preferential initiation of transcription of *CEA* and *NCA* is observed. Within or close to TSM1, initiation of transcription takes place in *PSG1*, *PSG5*, *Cea-2*, as well as in *rnCGM1*, which are all expressed in the placenta.
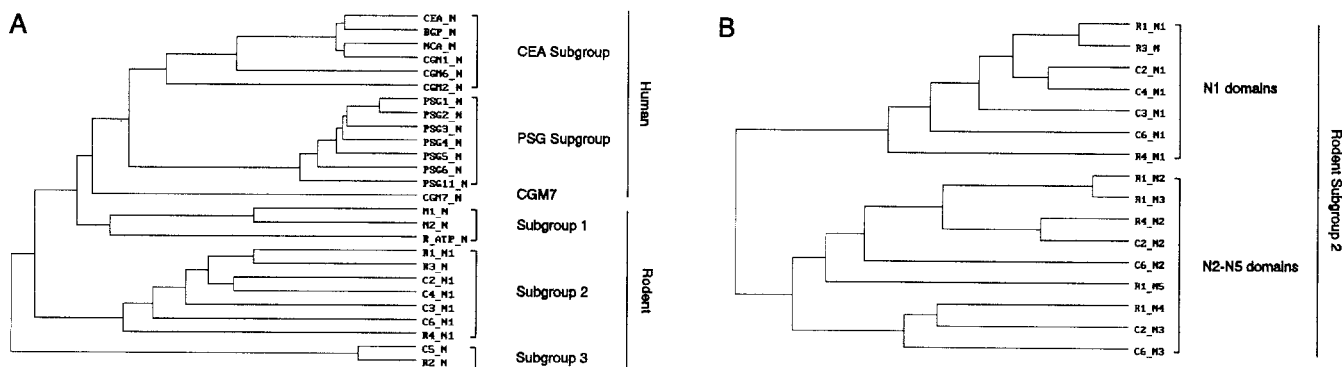
Fig. 3. Relationship of human and rodent CEA families. The amino acid sequences encoded by the indicated N domain exons were compared and displayed as dendrograms. The length of the branches reflect inversely the approximate degree of similarity between the compared sequences. C, *Cea*-; M, mmCGM; R, rnCGM (except for R_ATP; rat ectoATPase); N, amino acid sequence encoded by the N exons. Sequence sources are summarized in Thompson and colleagues (1991).
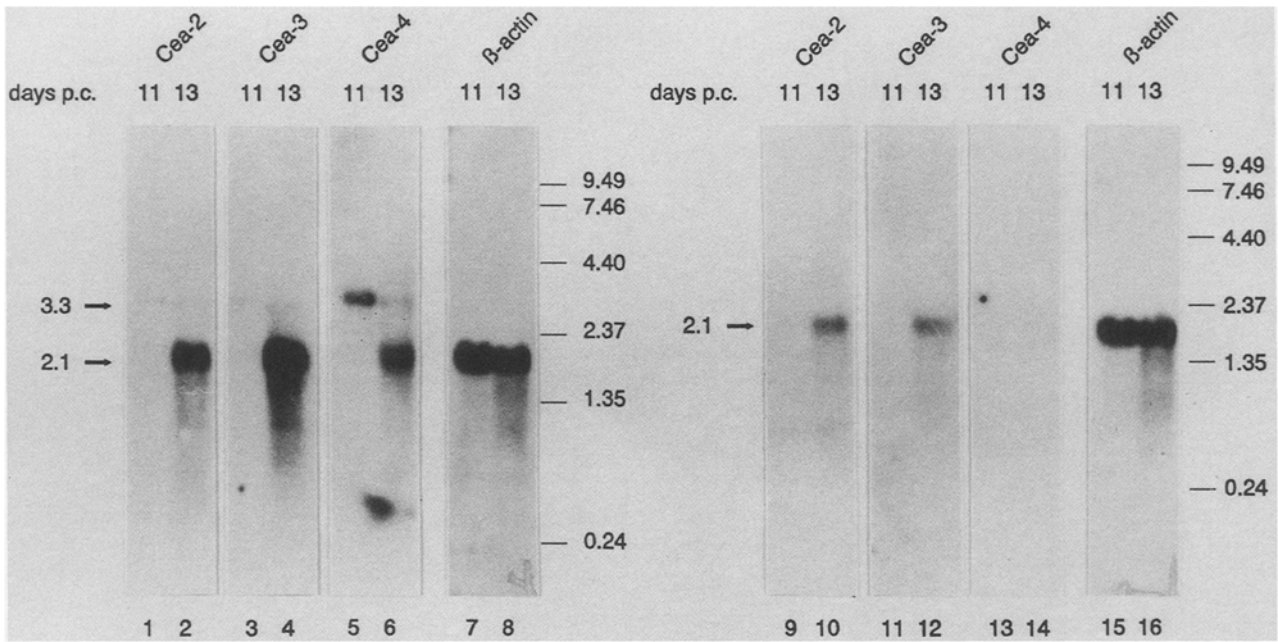
Fig. 4. Developmental expression of Cea-2 and Cea-3 mRNAs in mouse placenta. Fifteen μg of total RNA from day-11 and day-13 placenta were size separated in formaldehyde/agarose gels, transferred to charged nylon membranes, hybridized with the 1.1-kb genomic *Eco*RI DNA fragment of *Cea-4*, the 0.8 kb *Eco*RI/*Bam*HI fragment of λCea-3 (including 0.4 kb of the artefactually added DNA), and the cDNA insert of λCea-2, respectively (see Fig. 1A), and washed under the indicated conditions. The integrity and amount of RNA were checked by hybridization with a human β-actin probe of identical RNA samples, separated on the same gels. The sizes of marker RNAs (GIBCO-BRL) are indicated at the right margins; p.c., post coitum.
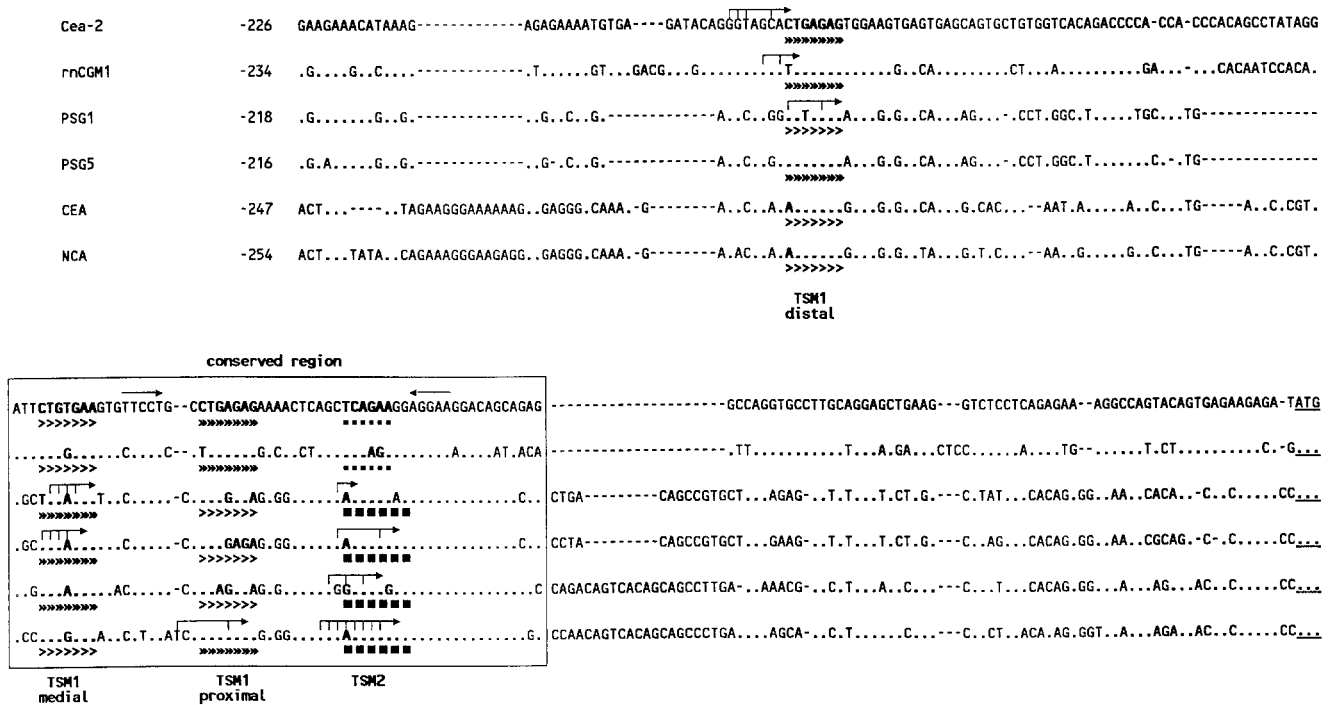
Fig. 5. Comparison of 5'-untranslated and upstream regions of members of the human and rodent CEA gene families. Dots indicate identical nucleotides; dashes, deletions that have been introduced for optimal alignment. Transcriptional start sites are marked by arrows, whereby main sites are shown in bold. The conserved sequence is boxed. Inverted arrows symbolize short inverted repeats. The conserved sequence motifs in the vicinity of the transcriptional start sites are shown in bold and are underlined with arrowheads (TSM1) and small boxes (TSM2), respectively. Sequences deviating from the consensus motifs are underlined with fewer arrowheads and larger boxes, respectively. The numbers indicate the size of the compared sequences upstream of the translational start (underlined). References for sequences and transcriptional start sites: *rnCGM1* (Rebstock et al. 1990), *PSG1* (Thompson et al. 1990; Watanabe and Chou 1988), *PSG5* (Thompson et al. 1990), *CEA* and *NCA* (Schrewe et al. 1990).

Chromosomal localization of Cea-5

The Cea-5 probe was hybridized to DNA from the homozygous parent generation, C3H/HeJ–gld/gld × Mus spretus, and the heterozygous F₁ generation, C3H/HeJ–gld/gld × Mus spretus. A unique Mus spretus 3.65 kb TaqI RFLV was identified (Fig. 6A). To examine the segregation of the RFLV, TaqI-digested DNA from 334 [(C3H/HeJ–gld/gld × Mus spretus)F₁ × C3H/HeJ–gld/gld] interspecific backcross progeny was analyzed. The Cea-5 RFLV segregated with genes previously mapped to the proximal region of mouse Chr 7 (Saunders and Seldin 1990a,b; Fig. 6B). Minimization of crossover frequency of these loci resulted

in the following gene order (±standard deviation): (centromere)–Pkcg-0.6 ± 0.42 cM–Cea-5–0.9 ± 0.52 cM–[Apoe, Atpa-2, Ckmm, D19S19H, Ercc-2].

Discussion

Here we report on five new members of the murine CEA gene family Cea-2, Cea-3, Cea-4, Cea-5, and Cea-6 (formerly named mmCGM5, mmCGM6, mmGMC7, mmCGM8, and mmCGM9; Thompson et al. 1991). Sequence comparisons with all known members of the murine and rat CEA gene families indicate that, as in humans, subgroups exist. None of these subgroups or individual genes, however, can be assigned to a human counterpart. This is probably owing to a rapid evolution of the CEA-related genes after separation of the primate and rodent orders (Rudert et al. 1989).

Rodent subgroup 1 contains the murine genes mmCGM1 and mmCGM2 (McCuaig et al. 1992; Turbide et al. 1991) and the rat gene coding for ecto-ATPase (Lin and Guidotti 1989), which was recently shown to be identical to C-CAM 105 (Aurivillius et al. 1990). This subgroup probably corresponds to the human CEA subgroup, since both encode mainly membrane-anchored proteins with a similar domain organization and expression pattern. Like BGP, rat ecto-ATPase/C-CAM 105 is found mainly in the epithelia of bile ducts (Svenberg 1976; Odin et al. 1988). Similarly, mmCGM1 and mmCGM2 transcripts have been detected in relatively large amounts in mouse liver (Turbide et al. 1991). Members of rodent subgroup 2 show a very different domain organization to their human counterparts and rodent subgroup 1 members. Instead of containing one IgV-like domain as found for all human CEA-related proteins, they are composed of three (Cea-2, Cea-6; Fig. 1A) or five (rnCGM1) such domains and one IgC domain (Rebstock et al. 1990). The existence of truncated leader sequences with cysteines between the IgV-like domains may have structural consequences (for example, for dimerization or formation of intramolecular disulfide bonds). Rodent subgroup 2 members lack any hydrophobic region and, therefore, seem to be secreted. Furthermore, as they are exclusively expressed in placenta, they most probably represent analogs of the human PSG subgroup. One may assume that human and rodent PSGs share a common function. As the N1 domain is the only domain of identical type (IgV-like) and position (first domain) common to rodent and human PSGs, it probably plays a major role in determining their function. This is supported by the finding that rodent PSG N1 domains form a separate group (Fig. 3B). The sequence motif present only in the rodent PSG N1 and human PSG N domains is indicative of a potential function. The tripeptide sequence ArgGlyAsp (RGD in one-letter code) found in human PSGs is known to be responsible for the interaction of some extracellular matrix proteins with cell surface receptors of the integrin family (Ruoslahti and Pierschbacher 1987). Although this sequence is not strictly conserved in mouse and rat N1
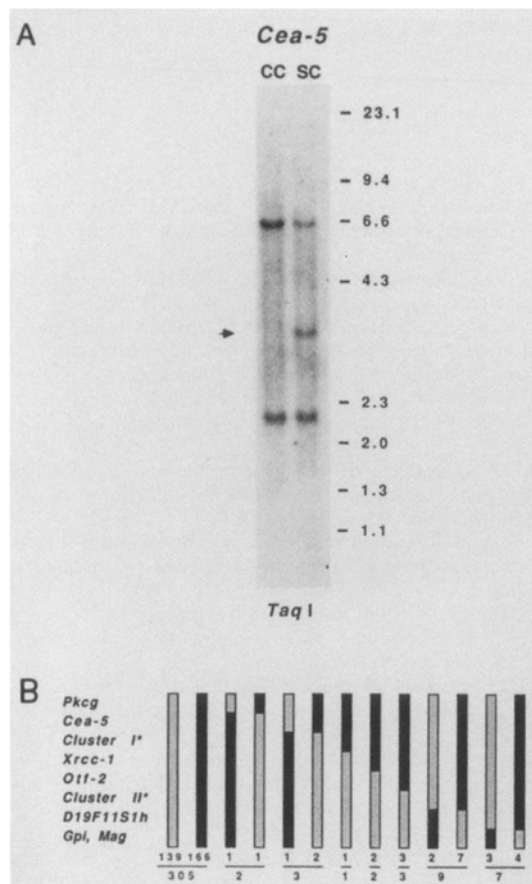
Fig. 6. Chromosomal localization of Cea-5. (A) Southern blot identification of a unique M. spretus RFLV for Cea-5. A [³²P]-labeled 2-kb BamHI DNA fragment from cosmid C3 covering the genomic region of Cea-5 with exon 1 and part of exon 2 was used as a probe. The arrowhead signifies the unique fragment present in DNA from (C3H/HeJ–gld/gld × M. spretus)F₁ (SC) but not present in homozygous C3H/HeJ–gld/gld (CC) mice. TaqI was the informative restriction endonuclease. The position and size (in kb) of marker fragments are shown on the right. (B) Analysis of interspecific backcross haplotypes. Haplotype distribution for mouse Chr 7 loci located within the syntenic human Chr 19q13.1–13.4 segment. For these loci, 334 backcross progeny were examined. Each column represents a chromosomal haplotype identified in the backcross progeny that was inherited from the (C3H/HeJ–gld/gld × M. spretus)F₁ parent. The number of backcross mice with each observed haplotype is indicated at the bottom of each column. Solid bars represent M. spretus-like RFLVs, and gray bars represent C3H/HeJ–gld/gld-like RFLVs. The cluster* designation includes (I) Apoe, Atpla3, Ckmm, D19S19h, and Ercc-2 and (II) Cyp1a, Cyp2b, Rip, and Tgfb in which no crossovers were observed.

domains, it shares a common spatial and charge pattern with the RGD motif. The fact that a variable number of Ig-like domains (3, 4, and 6, respectively) exist in different PSG members indicates that the length of these molecules is not necessarily relevant for function. It may even be that no murine PSG exists with five IgV-like domains as found in the rat, because computer-assisted analysis of the sequences (Li et al. 1985; Rebstock et al. 1990) reveals that the rnCGM1 N2/N3 domain duplication took place approximately 10×10^6 years ago, at a time after mouse/rat speciation (12×10^6 years ago; Bulmer et al. 1991). On the other hand, single PSGs may have specialized functions. Although the PSG genes of mice, rats, and humans seem in general to be coordinately expressed in placenta (Kodelja et al. 1989; Thompson et al. 1990), individual members, like the Cea-4-like gene, might function only at a certain time during development. In this context it is interesting to note that Cea-4 so far is the only rodent PSG member that lacks the RGD-related motif. The third subgroup consists of Cea-5 and rnCGM2, which show a remarkably high degree of similarity and are probably homologs. Sequence conservation is probably due to selective pressure since, although determination of Ks values reveals that separation of this gene pair (16×10^6 years) took place before the duplication of the N2 and N3 domains in the rat PSG rnCGM1 (10×10^6 years; Rebstock et al. 1990), the amino acid sequence conservation is more pronounced between the N domains of Cea-5 and rnCGM2 (93% vs. 76%). As rodent subgroups 1 and 2 are counterparts of the human CEA and PSG subgroups, respectively, members of rodent subgroup 3 may be analogous to human CGM7, which is the most distantly related member of the human CEA subgroup. Determination of their expression patterns in both species should help clarify this.

Like all human (Schrewe et al. 1990; Thompson et al. 1990) and rodent (Rebstock et al. 1990) CEA-related genes, Cea-2 lacks typical promoter consensus sequences. As the conserved motifs found often correlate with transcriptional starts, they may represent recognition sites for transactivating factors. Indeed, the transcriptional start sites of the placentally expressed PSG and rodent subgroup 2 genes often coincide with the motif TSM1, whereas for CEA and NCA TSM2 is found. Moreover, gel retardation experiments showed that an apparently ubiquitously expressed nuclear factor(s) can bind to oligonucleotides, the sequences of which have been derived from the conserved region of Cea-2 shown in Fig. 5 (results to be published elsewhere).

The localization of a murine CEA gene family member (Cea-5) to the proximal region of mouse Chr 7 is of interest because it was recently reported that the presumed CEA subgroup member mmCGM2 is also located on this chromosome (Robbins et al. 1991). Similarly, using a human CEA cDNA probe, Siracusa et al. (1991) colocalized a CEA-related gene with Otf-2 to the proximal region of mouse Chr 7. The human CEA gene family is clustered on the long arm of human Chr 19 (Thompson et al., 1992), in a region

(19q13.1–13.2) that is syntenic to the proximal part of mouse Chr 7 (Saunders and Seldin, 1990a). In consideration of this finding, together with the fact that Cea-2 and Cea-4 are closely linked, it may be expected that all members of the murine CEA gene family are clustered in this chromosome region.

References

Aurivillius, M., Hansen, O.C., Lazrek, M.B.S., Bock, E., and Öbrink, B.: The cell adhesion molecule Cell-CAM 105 is an ecto-ATPase and a member of the immunoglobulin superfamily. *FEBS Lett 264:* 267–269, 1990.

Bishop, D.T.: The information content of phase-known matings for ordering genetic loci. *Genet Epidemiol 2:* 349–361, 1985.

Bulmer, M., Wolfe, K.H., and Sharp, P.M.: Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. *Proc Natl Acad Sci USA 88:* 5974–5978, 1991.

Cech, T.R.: RNA splicing: three themes with variations. *Cell 34:* 713–716, 1983.

Chomczynski, P. and Sacchi, N.: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem 162:* 156–159, 1987.

Cordon, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P.: Promoter sequences of eukaryotic protein-coding genes. *Science 209:* 1406–1414, 1980.

Green, E.L.: Linkage, recombination and mapping. In E. Green (ed.); *Genetics and Probability in Animal Breeding Experiments,* pp. 77–113, Macmillan, New York, 1981.

Higgins, D.G. and Sharp, P.M.: CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene 73:* 237–244, 1988.

Hinoda, Y., Imai, K., Nakagawa, N., Ibayashi, Y., Nakano, T., Paxton, R.J., Shively, J.E., and Yachi, A.: Transcription of biliary glycoprotein I in malignant and non-malignant human liver tissues. *Int J Cancer 45:* 875–878, 1990.

Hohn B.: DNA sequences necessary for packaging of bacteriophage lambda DNA. *Proc Natl Acad Sci USA 80:* 7456–7460, 1983.

Jackson, L.L., Colosi, P., Talamantes, F., and Linzer, D.I.H.: Molecular cloning of mouse placental lactogen cDNA. *Proc Natl Acad Sci USA 83:* 8496–8500, 1986.

Khan, W.N. and Hammarström, S.: Identification of a new carcinoembryonic antigen (CEA) family member in human fetal liver—cloning and sequence determination of pregnancy-specific glycoprotein 7. *Biochem Biophys Res Commun 168:* 214–225, 1990.

Khan, W.N., Teglund, S., Bremer, K., and Hammarström, S.: The pregnancy-specific glycoprotein family of the immunoglobulin superfamily: identification of new members and estimation of family size. *Genomics, 12:*780–787, 1992.

Kodelja, V., Lucas, K., Barnert, S., von Kleist, S., Thompson, J.A., and Zimmermann, W.: Identification of a carcinoembryonic antigen gene family in the rat: analysis of the N-terminal domains reveals immunoglobulin-like, hypervariable regions. *J Biol Chem 264:* 6906–6912, 1989.

Kozak, M.: The scanning model for translation: an update. *J Cell Biol 108:* 229–241, 1989.

Leibrock, J., Lottspeich, F., Hohn, A., Hofer, M., Hengerer, B., Masiakowski, P., Thoenen, H., and Barde, Y.A.: Molecular clon-

ing and expression of brain-derived neurotrophic factor. *Nature* *341:* 149–152, 1989.

Li, W.-H., Wu, C.-I., and Luo, C.-C.: A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol 2:* 150–174, 1985.

Lin, S.-H. and Guidotti, G.: Cloning and expression of a cDNA coding for a rat liver plasma membrane ecto-ATPase. *J Biol Chem 264:* 14408–14414, 1989.

Lin, T.-M., Halpert, S.P., and Spellacy, W.N.: Measurement of pregnancy-associated plasma proteins during human gestation. *J Clin Invest 54:* 576–582, 1974.

Little, P.F.R. and Cross, S.H.: A cosmid vector that facilitates restriction mapping. *Proc Natl Acad Sci USA 82:* 3159–3163, 1985.

Lucas, K., Busch, M., Mössinger, S., and Thompson, J.A.: An improved microcomputer program for finding gene-, or gene family-specific oligonucleotides suitable as primers for polymerase chain reactions or as probes. *CABIOS 7:* 525–529, 1991.

Maniatis, T., Fritsch, E.F., and Sambrook, J.: *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1982.

McCuaig, K., Turbide, C., and Beauchemin, N.: mmCGM1a: A mouse carcinoembryonic antigen gene family member, generated by alternative splicing, functions as an adhesion molecule. *Cell Growth Differen,* in press, 1992.

Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., Riggs, A.D., and Shively, J.E.: Characterization of a cDNA clone for the nonspecific cross-reacting antigen (NCA) and a comparison of NCA and carcinoembryonic antigen (CEA). *J Biol Chem 263:* 3202–3207, 1988.

Odin, P., Asplund, M., Busch, C., and Öbrink, B.: Immunohistochemical localization of cellCAM 105 in rat tissues: appearance in epithelia, platelets and granulocytes. *J Histochem Cytochem 36:* 729–739, 1988.

Rebstock, S., Lucas, K., Thompson, J.A., and Zimmermann, W.: cDNA and gene analyses imply a novel structure for a rat carcinoembryonic antigen-related protein. *J Biol Chem 265:* 7872–7879, 1990.

Robbins, J., Robbins, P.F., Kozak, C.A., and Callahan, R.: The mouse biliary glycoprotein gene *(Bgp):* partial nucleotide sequence, expression, and chromosomal assignment. *Genomics 10:* 583–587, 1991.

Rudert, F., Zimmermann, W., and Thompson, J.A.: Intra- and interspecies analyses of the carcinoembryonic antigen (CEA) gene family reveal independent evolution in primates and rodents. *J Mol Evol 29:* 126–134, 1989.

Ruoslahti, E. and Pierschbacher, M.D.: New perspectives in cell adhesion: RGD and integrins. *Science 238:* 491–497, 1987.

Sakuragi, N.: Serum SP1 and hCG beta-subunit (hCG beta) levels in choriocarcinoma, invasive mole, and hydatidiform mole: clinical significance of SP1/hCG beta ratio. *Gynecol Oncol 13:* 393–398, 1982.

Sambrook, J.E., Fritsch, J., and Maniatis, T.: *Molecular Cloning: A Laboratory Manual,* 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

Saunders, A.M. and Seldin, M.F.: The syntenic relationship of proximal mouse chromosome 7 and the myotonic dystrophy gene region on human chromosome 19q. *Genomics 6:* 324–332, 1990a.

Saunders, A.M. and Seldin, M.F.: A molecular genetic linkage map of mouse chromosome 7. *Genomics 8:* 525–535, 1990b.

Schäfer, W.: The inbred mouse strain STU: development and properties. *Z Naturforsch 34c:* 306–309, 1979.

Schrewe, H., Thompson, J., Bona, M., Hefta, L.J.F., Maruya, A., Hassauer, M., Shively, J.E., von Kleist, S., and Zimmermann, W.: Cloning of the complete gene for the carcinoembryonic antigen: analysis of its promoter indicates a region conveying cell type-specific expression. *Mol Cell Biol 10:* 2738–2748, 1990.

Seldin, M.F., Morse, H.C., Reeves, J.P., Scribner, C.L., LeBoeuf, R.C., and Steinberg, A.D.: Genetic analysis of "autoimmune" *gld* mice. I. Identification of a restriction fragment length polymorphism closely linked to the *gld* mutation within a conserved linkage group. *J Exp Med 167:* 688–693, 1988.

Shively, J.E. and Beatty, J.D.: CEA-related antigens: molecular biology and clinical significance. *CRC Crit Rev Oncol Hematol 2:* 355–399, 1985.

Siracusa, L.D., Rosner, M.H., Vigano, M.A., Gilbert, D.J., Staudt, L.M., Copeland, N.G., and Jenkins, N.A.: Chromosomal location of the octamer transcription factors, *Otf-1, Otf-2,* and *Otf-3,* defines multiple *Otf-3*-related sequences dispersed in the mouse genome. *Genomics 10:* 313–326, 1991.

Steinmetz, M., Stephen, D., Dastoornikoo, G.R., Gibb, E., and Romaniuk, R.: Methods in molecular immunology: chromosomal walking in the major histocompatibility complex. *In Immunological Methods,* pp. 1–19, Academic Press, New York, 1985.

Svenberg, T.: Carcinoembryonic antigen-like substances of human bile: isolation and partial characterization. *Int J Cancer 17:* 588–596, 1976.

Thompson, J., Koumari, R., Wagner, K., Barnert, S., Schleussner, C., Schrewe, H., Zimmermann, W., Müller, G., Schempp, W., Zaninetta, D., Ammaturo, D., and Hardman, N.: The human pregnancy-specific glycoprotein genes are tightly linked on the long arm of chromosome 19 and are coordinately expressed. *Biochem Biophys Res Commun 167:* 848–859, 1990.

Thompson, J.A., Grunert, F., and Zimmermann, W.: The carcinoembryonic antigen gene family: molecular biology and clinical perspectives. *J Clin Lab Analysis 5:* 63–83, 1991.

Thompson, J., Zimmermann, W., Osthus-Bugat, P., Schleussner, C., Eades, A.-M., Barnert, S., von Kleist, S., Willcocks, T., Craig, I., Tynan, K., Olsen, A., and Mohrenweiser, H.: Long range chromosomal mapping of the carcinoembryonic antigen (CEA) gene family cluster. *Genomics, 12:*761–772, 1992.

Turbide, C., Rojas, M., Stanners, C., and Beauchemin, N.: A mouse carcinoembryonic antigen (CEA) gene family member is a calcium dependent cell adhesion molecule. *J Biol Chem 266:* 309–315, 1991.

Watanabe, S. and Chou, J.Y.: Isolation and characterization of complementary DNAs encoding human pregnancy-specific β1-glycoprotein. *J Biol Chem 263:* 2049–2054, 1988.

Williams, A.F. and Barclay, A.N.: The immunoglobulin superfamily-domains for cell surface recognition. *Annu Rev Immunol 6:* 381–405, 1988.

Zimmermann, W., Ortlieb, B., Friedrich, R., and von Kleist, S.: Isolation and characterization of cDNA clones encoding the human carcinoembryonic antigen reveal a highly conserved repeating structure. *Proc Natl Acad Sci USA 84:* 2960–2964, 1987.

Zimmermann, W., Weber, B., Ortlieb, B., Rudert, F., Schempp, W., Fiebig, H.-H., Shively, J.E., von Kleist, S., and Thompson, J.A.: Chromosomal localization of the carcinoembryonic antigen gene family and differential expression in various tumors. *Cancer Res 48:* 2550–2554, 1988.