Carcinoembryonic Antigens: Alternative Splicing Accounts for the Multiple mRNAs that Code for Novel Members of the Carcinoembryonic Antigen Family

Thomas R. Barnett,* Axel Kretschmer,‡ Douglas A. Austen,‡ Scott J. Goebel,* John T. Hart,‡ James J. Elting,* and Michael E. Kamarck‡

*Molecular Diagnostics, Inc. and ‡Molecular Therapeutics, Inc., Miles Research Center, West Haven, Connecticut 06516

Abstract. The recent cloning of complete cDNAs encoding carcinoembryonic antigen (CEA) and nonspecific cross-reacting antigen has revealed the existence of a new gene family belonging to the immunoglobulin gene superfamily. We have reported the isolation of a partial CEA cDNA and of L-cell transfectant cell lines that express human antigens cross-reactive with commercial antibodies directed to native CEA (Kamarck, M., J. Elting, J. Hart, S. Goebel, P. M. M. Rae, J. Nedwin, and T. Barnett. 1987. *Proc. Natl. Acad. Sci. USA.* 84:5350–5354). In this study, we describe the identification and cloning of 3.9-, 3.7-, 2.2-, and 1.8-kb cDNAs and a 23-kb genomic transcription unit, which code for new members of the CEA gene family. DNA sequence analysis of these cloned DNAs establishes the existence of a set of four alternatively spliced mRNAs which are expressed in several tumor cell lines, in human fetal liver, and in L-cell transfectants. Deduced amino acid sequences of the encoded isoantigens show extensive similarity to CEA and nonspecific cross-reacting antigens, but in addition demonstrate transmembrane and cytoplasmic domains. We designate members of this antigen family transmembrane CEAs. The transmembrane CEA isoantigens share general structural characteristics with members of the immunoglobulin gene superfamily and can be specifically compared to the cell adhesion molecules, N-CAM (neural cell adhesion molecule) and MAG (myelin-associated glycoprotein).

CARCINOEMBRYONIC antigen (CEA)[1] is a 180-kD glycoprotein that has been used extensively as a serum marker in some cancers, particularly those of colorectal, breast, and lung origin. Despite its initial promise as a tumor-specific marker, CEA has been shown to be a member of a family of 8–10 cross-reactive isoantigens which can be detected in a variety of normal and tumor tissue types (Shively and Beatty, 1985). CEA and nonspecific cross-reacting antigen (NCA) have recently been shown to have considerable structural and sequence similarity (Engvall et al., 1978; Paxton et al., 1987; Thompson et al., 1987), thus explaining their observed immunological cross-reactivity among anti-CEA antibodies (von Kleist et al., 1972; Darcy et al., 1973). This cross-reactivity extends to all members of the CEA isoantigen family and has made interpretation of diagnostic immunoassays for CEA both difficult and inaccurate.

Recently, the complete cDNAs for both CEA and NCA have been isolated and sequenced (Beauchemin et al., 1987; Oikawa et al., 1987; Barnett et al., 1988; Neumaier et al., 1988; Tawaragi et al., 1988). These antigens display a high degree of sequence similarity and also maintain a repeating series of disulfide loops with features of immunoglobulin superfamily members (Hunkapiller and Hood, 1986; Williams and Barclay, 1988). Both CEA and NCA have short hydrophobic peptide segments at their carboxyl termini, but appear to be anchored in the cell membrane by a glycolipid moiety (Takami et al., 1988).

We have previously isolated an L-cell transfectant line, 23.4 11+, and a bacteriophage λgt11 partial cDNA clone (λcLV7) that express epitopes of CEA (Kamarck et al., 1987). The partial cDNA clone encompasses the second and part of the third loop domains of CEA (nucleotides 1,084–1,916 in Barnett et al., 1988) and identifies several major Eco RI and Bam HI segments in genomic DNA. All of these genomic segments map to human chromosome 19, suggesting the existence of a CEA gene locus consisting of an extensive family of related members. In contrast, cDNA from λcLV7 hybridized to a single genomic segment in L-cells that had been transfected with human genomic DNA and selected for CEA expression using commercial antibod-

Axel Kretschmer's present address is BAYER AG, ZF-F Biotechnologie Q18, D-5090 Leverkusen, FRG.

1. *Abbreviations used in this paper*: CEA, carcinoembryonic antigen; MAG, myelin-associated glycoprotein; NCA, nonspecific cross-reacting antigen; N-CAM, neural cell adhesion molecule; TM-CEA, transmembrane carcinoembryonic antigen.

ies and the fluorescence-activated cell sorter. These transfectants expressed CEA family members distinguishable from both CEA and NCA based on the molecular sizes of their antigens and their messenger RNAs (Kamarck et al., 1987; Barnett et al., 1988).

In this study, we describe the cloning of a genomic transcription unit and of cDNAs for alternatively spliced mRNAs that code for four previously undescribed CEA isoantigens expressed by a number of tumor cell lines and by normal fetal liver. Despite their extensive amino acid sequence similarities, the four protein species differ from CEA and NCA in that they contain transmembrane and cytoplasmic domains. For this reason we have designated these molecules transmembrane CEAs (TM-CEAs). With the identification of this new class of CEA glycoproteins, we can now account for six of the 8–10 members of the CEA isoantigen family.

Materials and Methods

Cells and Cell Lines

Human tumor cell lines were obtained from the American Type Culture Collection (Rockville, MD) and grown according to specifications. Line designations are LoVo (CCL 229, colon adenocarcinoma), HT-29 (HTB 38, colon adenocarcinoma), BeWo (CCL 98, choriocarcinoma), SW403 (CCL 230, colon adenocarcinoma), KG-1 (CCL 246, acute myelogenous leukemia), SCaBER (HTB 3, squamous bladder carcinoma), and MIA PaCa-2 (CRL 1420, human pancreatic carcinoma). The L-cell transfectant line, 23.4 11+, has been previously described (Kamarck et al., 1987).

DNA Preparation and Analysis

DNA was extracted from cells essentially as described by DiLella and Woo (1987). DNA (10 μg) was digested at 37°C for 6–8 h with restriction endonucleases (Amersham Corp., Arlington Heights, IL) and subjected to electrophoresis through horizontal 0.8% agarose gels before transfer to nylon membranes. For the preparation of genomic DNA for library construction or subcloning, 100–200 μg of DNA was digested either with excess Eco RI or with a predetermined amount of Mbo I (mean DNA size produced = 12–15 kb) and subjected to electrophoresis overnight through horizontal 0.5% low melting temperature agarose gels for optimal separation. Appropriate size classes were excised with a razor blade, gel slices were melted in 0.3 M NaOAc, pH 5, and 0.1 M Tris-HCl, pH 7.8, DNA was purified by phenol extraction, precipitated with 95% ethanol, and resuspended in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA.

RNA Preparation and Analysis

Cell pellets stored at −80°C were thawed in the presence of 0.14 M NaCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.5, 0.5% NP-40 (Sigma Chemical Co., St. Louis, MO), 4 mM DTT, and 10 u/ml of human placental ribonuclease inhibitor (Stratagene Inc., San Diego, CA). Cells were fractionated by 20 strokes of a low clearance Dounce-type homogenizer and incubated in 0.1% Na deoxycholate for 10 min on ice. Cytoplasmic and nuclear fractions were separated by centrifugation for 20 min at 12,000 *g*, and the supernatant was added to an equal volume of 200 mM Tris-HCl, pH 7.5, 25 mM EDTA, 300 mM NaCl, 2% SDS, and 400 μg/ml of proteinase K. After incubation at 45°C, nucleic acids were extracted with phenol/chloroform (1:1 [vol/vol]), and ethanol precipitated in the presence of added NaOAc to 0.15 M. Poly A+ RNA was prepared by oligo dT column chromatography as described by Aviv and Leder (1972). For Northern analysis, nucleic acids were separated on 2.2 M formaldehyde–0.8% agarose gels as described by Lehrach et al. (1977), then transferred to nylon membranes.

Probe Preparation and Nucleic Acid Blots

Mini-preparations of plasmid DNA (Maniatis et al., 1982) were digested with restriction enzymes and fractionated by size on 1% agarose gels. Appropriate segments were excised from the gel and used directly for random priming (Feinberg and Vogelstein, 1984) in the presence of α-^{32}P-dCTP (6,000 Ci/mmol; Amersham Corp). Unincorporated nucleotide was removed by spun column chromatography over Sephadex G-50 (Pharmacia Fine Chemicals, Piscataway, NJ). Probes were used at 10^6 cpm/ml in 2× SSPE, 5× Denhardt's, 6% SDS at 68°C for 24–48 h. Posthybridization washes were in 0.2× SSPE, 0.25% SDS at 68°C.

cDNA Library Construction

Poly A+ RNAs from 23.411+ transfectant cells and from HT-29 cells were converted to double-stranded DNA essentially by the method of Gubler and Hoffman (1983). After Eco RI methylase treatment, DNA was fractionated on 0.8% LGT agarose gels, size selected based on identification of RNA from Northern blots, and eluted by melting (see above). Eco RI linkers were added and cDNAs were inserted into Eco RI–cleaved arms of ligated λgt10 DNA. Phage with inserts were selected on the Hfl^+ host, NM514. Screening of phage was by the method of Benton and Davis (1977) using radiolabeled DNA probes. Purified positive phage DNAs were cleaved with Eco RI, then inserts were fractionated on 1% LGT agarose gels and ligated to 5 ng of Eco RI–cleaved plasmid Bluescript KS+ (Stratagene). Plasmid minipreps were used as source of DNA for deriving nested deletion sets of cDNA inserts by exonuclease III/mung bean nuclease treatment (Henikoff, 1984) and/or for primer or oligonucleotide directed dideoxy sequencing on double-stranded DNA by the method of Sanger et al. (1977). Computer-aided analysis of DNA sequence information was performed using the Pustell program (IBI) or the series of programs published by the Genetics Computer Group (Madison, WI).

Genomic Library Construction

DNA segments of 12–15 kb from Mbo I partial digests were inserted into left and right arms of Bam HI–cleaved λ*Jam* phage DNA. This λ*am Bam* bacteriophage vector was constructed by substituting the 5.0-kb cosL–Ava I segment of λCharon 4A phage DNA for the equivalent segment of λJ1 DNA (Mullins et al., 1984) and selecting phage on su^-/su^+ hosts. A library of $>10^6$ independent recombinants was obtained after packaging the ligated λ*Jam* and transfectant DNAs in vitro and plating on *Escherichia coli* host LE392. Positive phage were purified, DNAs were prepared, and restriction segments were separated on 1% agarose gels before blotting on nitrocellulose membranes according to Southern (1975). Appropriate segments were subcloned into Bluescribe (−) or Bluescript KS+ or SK+ plasmids for double-stranded DNA sequencing.

Western Blot Analysis

Peptide/N-glycanase was purified from culture supernatants of *Flavobacterium meningosepticum* by the method of Tarentino et al. (1985). No endoglycosidase F or endogenous protease activity could be demonstrated in this preparation. Transfectant 23.4 11+ or normal mouse L cells were lysed in phosphate buffer containing Triton X-100 (PBS-T; 10 mM $NaPO_4$, pH 7.3, 140 mM NaCl, 10 mM EDTA, 10 mM benzamidine, 1 mM difluorophosphate, and 1% Triton X-100) by sonication at 4°C. Aliquots of lysate containing up to 100 μg of total protein were brought to 1% SDS and 1 mM β-mercaptoethanol and heated to 100°C for 3 min. After cooling, the denatured lysates were mixed with an equal volume of 2× endo F digestion buffer (1× = 0.2 M $NaPO_4$, pH 8.7, 0.75% NP-40, 5 mM EDTA, 5 mM *o*-phenanthrolene, and 0.5 mM PMSF). Peptide/*N*-glycanase (0.4 μg) was added and digestion was performed for 3 h at 37°C. An additional 0.1 μg of enzyme was added and digestion was continued for an additional hour. Controls for enzyme reactivity included the complete deglycosylation of an α_1-acid glycoprotein. Reactions were terminated by the addition of 1/3 vol of 4× SDS-PAGE sample buffer, then heated for 7 min at 100°C. Digested samples were analyzed by SDS-PAGE 10–20% gradient gels (Laemmli, 1970). Proteins were transferred to nitrocellulose by the method of Towbin et al. (1979). Membranes were blocked for 1 h in PBS-T containing 5% BSA. Rabbit anti-CEA Ig (DAKOPATTS, Copenhagen, Denmark) was used at 2 μg/ml and goat anti-rabbit Ig alkaline phosphatase conjugate (Promega Biotec, Madison, WI) was used at a dilution of 1:7,500. Detection was by incubation of the membrane in 100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$ containing 66 mg/ml of nitro blue tetrazolium and 33 mg/ml of 5-bromo-4-chloro-3-indolyl phosphate.

Results

Isolation of TM-CEA cDNAs

Northern blot analysis was performed on poly A+ RNA from the L-cell transfectant line, 23.4 11+ that had been selected for the expression of CEA isoantigens (Kamarck et al., 1987). Use of the CEA partial cDNA insert, LV7, as probe identified multiple hybridizing RNA bands with approximate molecular sizes of 3.9, 3.7, 2.2, and 1.8 kb (Fig. 1, lane *a*). This pattern also recurs in the colon adenocarcinoma cell line HT-29 (Fig. 1, lane *b*), which in addition expresses mRNAs for CEA (4.0 and 3.6 kb) and NCA (3.0 kb) (Barnett et al., 1988). To clone the unidentified mRNAs, 3–4-kb double-stranded cDNA from 23.4 11+ poly A+ RNA was size selected and inserted into λgt10. A library of 8×10^5 pfu was screened by LV7 hybridization. 13 phage were isolated, and the subcloned inserts from three of these, tc16–19 (3.6 kb), tc19–22 (3.5 kb), and tcE22 (3.3 kb), were analyzed by restriction endonuclease mapping. Insert tcE22 differed from the others by an apparent internal deletion of ~300 bp. Additional cDNAs were prepared from an HT-29 cDNA library that was size selected in the 1–3-kb region.

Based on the observation by Gold and Freedman (1965) that CEA is expressed during fetal development, we also screened a cDNA library (2×10^5 pfu; Clontech Laboratories Inc., Palo Alto, CA) prepared from normal human fetal liver poly A^+ RNA with the CEA probe, LV7. Five positive recombinant phage were isolated and the inserts were subcloned into plasmid vectors. Two overlapping inserts, cFL4 and cFL5, together comprised a 3.6-kb fetal liver cDNA identical in restriction map to tc16–19 and tc19–22.

Fig. 2 shows the compilation of DNA sequences and deduced amino acid sequences from the inserts of plasmids ptc16–19, ptc19–22, ptcE22, pcFL4, and pcFL5. These inserts together provided the longest open reading frame (nucleotides 73 to 1656) and the longest cDNA (3464 nucleotides). The compiled sequence codes for an apoprotein of 526 amino acids (M_r of 58 kD) with 19 potential extracellular N-linked glycosylation sites (Asn-X-Thr/Ser).

The coding sequence of the proposed apoprotein (designated TM1-CEA) demonstrates marked similarity to sequences found in CEA and NCA cDNA clones. This is indicated by the proposed domain structure illustrated in Fig. 3 and by the sequence relationships detailed in Table I. As is typical for other CEA isoantigens, the TM1-CEA polypeptide initiates with a 34-amino acid leader sequence that ends at a short side-chain amino acid, alanine. The proposed NH_2-terminus of TM1-CEA contains glutamine as the first residue that initiates at 108-amino acid NH_2-terminal domain. This domain is identical in length and highly similar in sequence to the comparable domain of CEA and NCA (Fig. 3 and Table I). This is followed by a disulfide-loop domain (LD) that is a characteristic feature of CEA and is ~178 amino acids in length (Oikawa et al., 1987).

This domain differs from the single disulfide loop domains of immunoglobulins in that each one contains four cysteine residues capable of forming 47- and 39-amino acid disulfide-linked loops. Loop domains are repeated three times in CEA and once in NCA (Fig. 3). The disulfide-loop domain size for TM1-CEA is similar to those of CEA and NCA, and the position of its cysteines is strictly conserved. In contrast to the domains of CEA and NCA, LD I of TM1-CEA has a fifth cysteine (residue 308; Fig. 2) that may not be involved in the formation of an intrachain disulfide loop.

An interesting structural region follows the conserved 177 amino acid loop-domain of TM1-CEA. This region, extending from nucleotides 1,030 to 1,340, codes for ~100 amino acids and has the appearance of a "half" disulfide-loop domain (*LD IIa* of Fig. 3). This is indicated by comparison of amino acids that initiate loop domain regions in CEA, NCA, and TM1-CEA and by the characteristic positioning of its only two cysteine residues to form a 47–amino acid disulfide loop. While each of the disulfide-loop domains of the three CEA classes is highly conserved (~70% amino acid similarity overall), LD IIa of TM1-CEA retains ~44% amino acid similarity compared to TM1-CEA LD I (Table I).

A 32–amino acid hydrophobic region comes after LD IIa and bears sequence similarity to the short hydrophobic tail that terminates CEA and NCA (Table I). Hydropathy plot analysis supports the inference that this is a membrane-spanning region (data not shown). By contrast, the carboxyl-terminal region of TM1-CEA is composed of a hydrophilic stretch of 71 amino acids that terminates at residue 526 (nucleotide 1667).

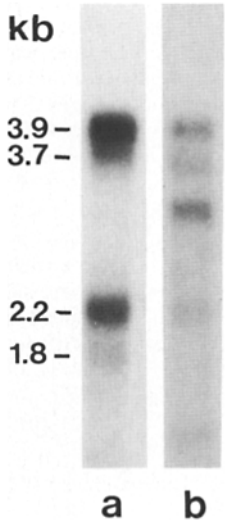

Figure 1. Expression of CEA-related poly A+ RNAs. One-tenth microgram of CEA-transfectant 23.4 11+ poly A+ RNA (lane *a*) and two micrograms of colon carcinoma cell line HT-29 poly A+ RNAs (lane *b*) were subjected to electrophoresis on a 2.2 M formaldehyde–0.8% agarose gel, transferred to nylon membrane, and hybridized with ^{32}P-labeled LV7 DNA.

Identification of TM-CEA cDNAs Derived from Alternatively Spliced mRNAs

Four different molecular sizes of CEA-related mRNA are detected in transfectant cell line 23.4 11+ and are observed in tumor cell line HT-29 (Fig. 1). Additional cDNAs isolated from 23.4 11+ and HT-29 cDNA libraries established that the observed mRNA species are derived by alternative RNA splicing.

DNA Sequence of TM2-CEA. Insert tcE22, isolated from the 23.4 11+ cDNA library, presented a restriction map indistinguishable from TM1-CEA clones when a 300-bp gap was positioned within the molecule. DNA sequence analysis showed that tcE22 codes for a second TM-CEA species (TM2-CEA), which is identical to the TM1-CEA apoprotein with the exception of a 100–amino acid gap that corresponds exactly to LD IIa (Figs. 2 and 3). We conclude that this cDNA accounts for the 3.7-kb mRNA observed in Fig. 1 and that it is derived by alternate splicing of the transcript that

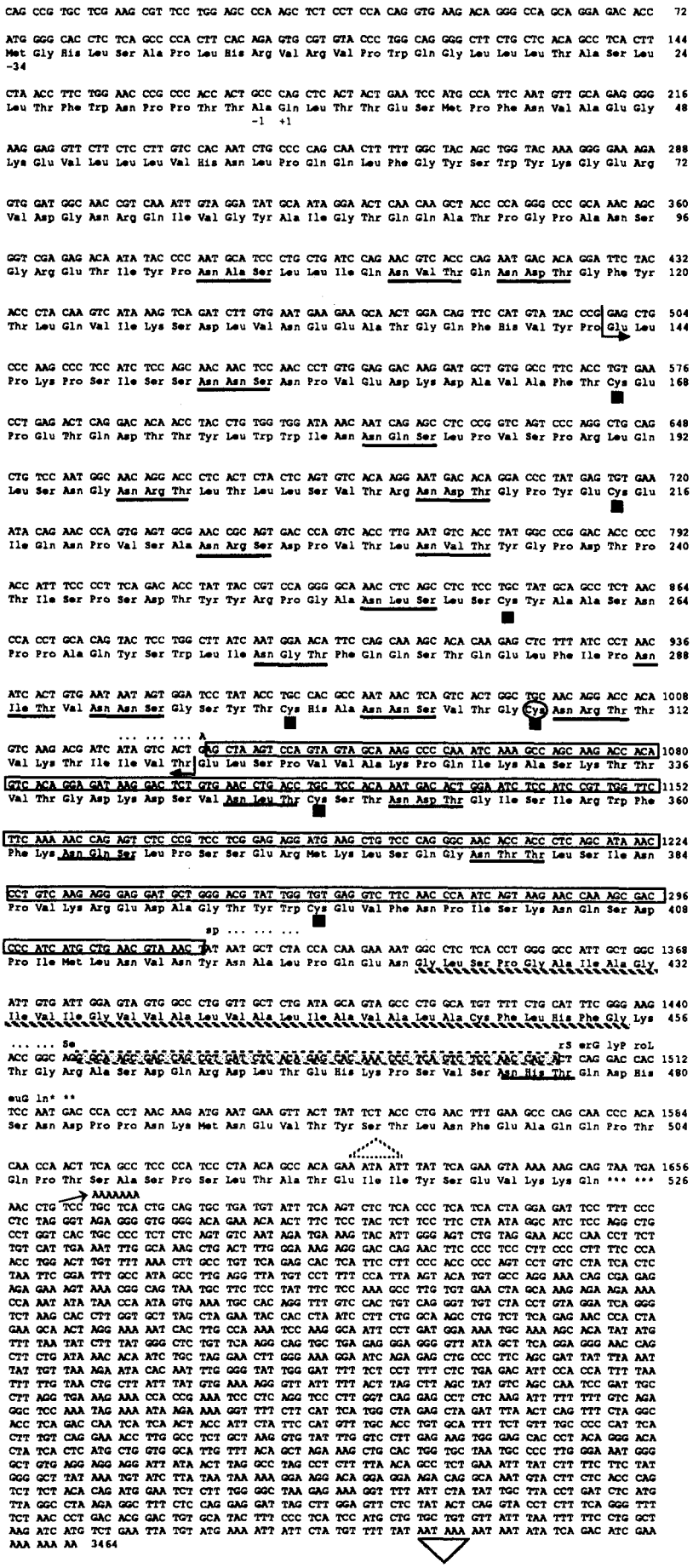

Figure 2. DNA and deduced amino acid sequences of TM-CEA cDNAs. The DNA sequences were compiled from double-strand dideoxy termination reactions on both strands of all clones. The full sequence from nucleotide 1 to 3,464 corresponds to TM1-CEA. The initiating methionine residue of the apoprotein is at position −34 relative to the proposed amino terminus of the mature protein at +1. Symbols ↳ ↲ demarcate the likely beginning and end of the first loop-domain based on similarity with CEA and NCA; solid bars note consensus sequences for N-linked glycosylation sites; ■ are cysteine residues in the extracellular portion of the molecule; (Cys) is the only unpaired cysteine residue in the extracellular region; the thick barberpole line indicates a hydrophobic transmembrane region. The open box around nucleic acid sequence corresponds to the "half" loop-domain (LD IIa) region deleted in TM2-CEA; the single amino acid that is altered as a result of "half" loop-domain deletion, i.e., aspartic acid (*Asp*), is shown as a split residue above the corresponding split codon in the TM1-CEA sequence; dots on either side of the Asp residue indicate that all adjacent sequences are as shown for TM1-CEA. The 53 nucleotides that are deleted in the cytoplasmic region of TM3-CEA are indicated by the shaded portion of the broken box. The new amino acid, serine (*Ser*), that is generated at the splice junction, and subsequent new amino acids that are translated by the induced frameshift, are shown above their corresponding new codons; dots amino-terminal to Ser indicate that all sequence to that point is identical to TM1-CEA. Termination of translation is indicated by ***. Solid and broken triangles show potential polyadenylation signals for TM1/TM2-CEA and for TM3-CEA, respectively; the up arrow after nucleotide 1,662 indicates the presence of a short oligo A stretch detected in the TM3-CEA cDNA.

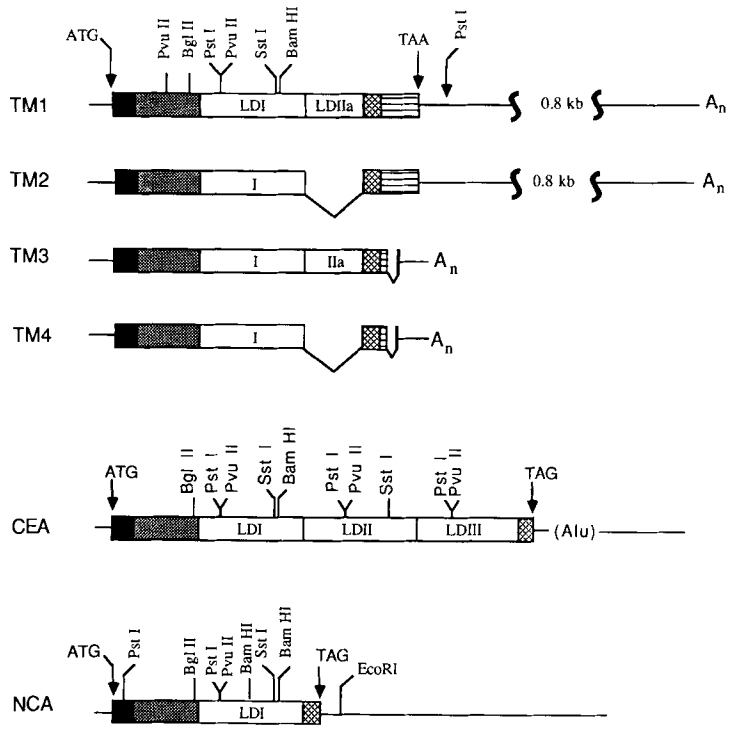

Figure 3. Schematic representation of proposed structural domains of CEA-isoantigens TM1-CEA, TM2-CEA, TM3-CEA, TM4-CEA, and their comparison with CEA (180 kD) and NCA. Thin lines correspond to the 5′ and 3′ untranslated regions. The coding regions of the apoproteins (rectangle) initiate with ATG and are divided into leader sequences (■), NH_2-terminal sequences (▩), loop-domains (□), hydrophobic regions (▩), and a cytoplasmic region for TM-CEA (▤). TAG and TAA are translation stop codons. Lines connecting structural domains indicate where the TM-CEA apoproteins differ from TM1-CEA. The symbol | indicates the segment resulting from the translational frame shift in TM3- and TM4-CEA (see text for details). A_n is poly A.

also generates TM1-CEA. The predicted apoprotein size of TM2-CEA is 47 kD.

DNA Sequence of TM3-CEA. Northern analysis of poly A+ RNAs indicated a prominent 2.2-kb RNA species when hybridized with LV7 probe DNA (Fig. 1). cDNA from HT-29 poly A+ RNA yielded two partial clones that could be distinguished from TM1- and TM2-CEA cDNAs by restriction site analysis. Efforts to isolate full-length versions of TM3-CEA cDNA from the 23.4 11+ library were not successful. The DNA sequence of these partial cDNAs corresponds exactly to nucleotides 345–1,448 and 1,502–1,662 of TM1-CEA cDNA. The deletion of 53 bp between nucleotides 1448 and 1502 occurs in the DNA coding for the cytoplasmic region (*shaded box* of Fig. 2). As 53 bp is not a multiple of a codon triplet, nucleotides after 1,448 yield a new translation reading frame and generate a new cytoplasmic component for TM3-CEA (schematically drawn in Fig. 3). The sequence indicates a translation of six new amino acids, followed by a TGA stop codon (nucleotide 1518). This results in a 9-amino acid cytoplasmic region for TM3-CEA compared to a 71-amino acid region for TM1- and TM2-CEA producing an apoprotein of 51 kD. A potential polyadenylation consensus sequence AATAAT (Wickens and Stephenson, 1984) in the 3′ untranslated region (UTR) of TM3-CEA mRNA is found at a position that precedes a short run of oligo A nucleotides where the sequence diverges from that of TM1- and TM2-CEA. The likely polyadenylation site occurs in sequence that codes for the TM1- and TM2-CEA cytoplasmic domain.

Table I. Amino Acid Sequence Identities among CEA Family Members

Members compared	Regions compared (in percent)			
	Leader	NH_2-terminal	Loop-domain	HR
NCA/CEA	73	89	85 (CEA I) 77 (CEA II) 76 (CEA III)	76
CEA/TM1-CEA	76	87	83 (CEA I) 79 (CEA II) 73 (CEA III)	67
NCA/TM1-CEA	76	89	86	62
TM1-CEA: domain I vs. IIa	–	–	44	–

For each comparison, the different regions correspond to signal sequences (amino acids −34–−1), NH_2-terminal segments (amino acids 1–108), loop-domains (amino acids 109–286, 287–464, and 465–642 for CEA, and 109–286 for NCA [Barnett et al., 1988], 109–285 and 286–389 for loop-domains I and IIa, respectively, of TM-CEA), and hydrophobic regions (*HR*) (amino acids 643–668 for CEA, 287–310 for NCA, and 390–421 for TM-CEA). Sequence comparisons are strictly on the basis of amino acid identity; conservative substitutions have not been permitted. Where one region is slightly different in length from its counterpart, percent identity is calculated on the shorter sequence.

DNA Sequence of TM4-CEA. Although we have yet to obtain a representative cloned cDNA sequence, we speculate that there exists an additional TM-CEA molecule, TM4-CEA (calculated M_r of 40 kD), in which both LD IIa and the cytoplasmic region are alternatively spliced as in TM2-CEA and TM3-CEA, respectively (illustrated in Fig. 3). Experimental support for a fourth TM-CEA molecule comes from our observation of a 1.8 kb mRNA that is observed in Northern blots of 23.4 11+ cell mRNA (Fig. 1, lane *a*). The estimated size of TM4-CEA mRNA without a poly A^+ tail is 1.4 kb. As our size estimates for each of the TM-CEA RNAs by Northern blot analysis is consistently 300–400 bases longer than that of the cDNA, it is likely that the 1.8-kb RNA codes for TM4-CEA.

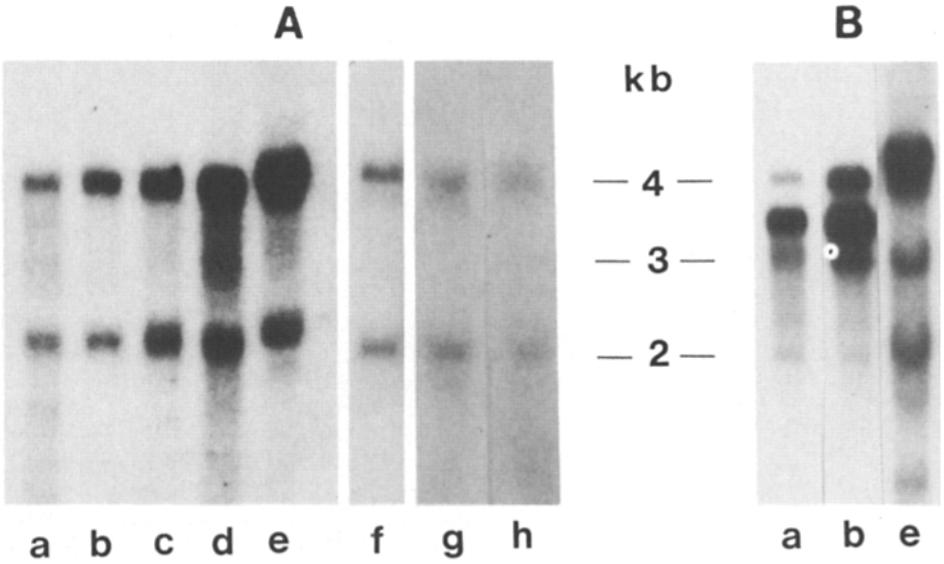

Figure 4. Tumor cell distribution of CEA-related poly A+ RNA. Two micrograms of poly A+ RNAs from various tumor cell lines were subjected to electrophoresis on a 2.2 M formaldehyde–0.8% agarose gel, transferred to nylon membrane, and hybridized with ^{32}P-labeled CYTO probe DNA ([*A*] lane *a*, SW403; lane *b*, LoVo; lane *c*, DLD-1; lane *d*, HT-29; lane *e*, KG-1; lane *f*, BeWo; lane *g*, SCaBER; and lane *h*, MIA PaCa-2) or LV7 DNA ([*B*]; for lanes *c*, *d*, *f*, *g*, and *h*; CYTO and LV7 patterns are identical). Autoradiographic exposures have been adjusted to accommodate the differences in TM-CEA mRNA concentration displayed by the cell lines SCaBER and MIA PaCa-2.

Expression of TM-CEA cDNAs

To investigate the specific expression of TM-CEA mRNAs in tumor cells of different origins, we probed cellular RNAs with a cDNA probe derived from the cytoplasmic region of the TM1-CEA cDNA (CYTO probe; nucleotides 1,443–1,676). Fig. 4 *A* shows hybridization of the CYTO probe to poly A+ RNAs from a variety of tumor cell lines (lanes *a–h*). All of the tumor cell lines shown, representing colorectal adenocarcinomas (SW403, lane *a*; LoVo, lane *b*; DLD-1, lane *c*; HT-29, lane *d*), acute myelogenous leukemia (KG-1, lane *e*), choriocarcinoma (BeWo, lane *f*), squamous bladder adenocarcinoma (SCaBER, lane *g*), and pancreatic adenocarcinoma (MIA PaCa-2, lane *h*) express the TM-CEA mRNAs. While some of these tumors express other CEA-isoantigen mRNAs based on hybridization to LV7 cDNA (Fig. 4 *B*, lanes *a*, *b*, and *e*), BeWo, SCaBER, and MIA PaCa-2 express only the transmembrane CEA isoantigen mRNAs (data not shown).

The Complete Transcription Unit of TM-CEA

We have reported the transfection of the TM-CEA transcription unit into primary and secondary L-cell transfectants (Kamarck et al., 1987). To clone genomic segments corresponding to this transcription unit, a λ*Jam* recombinant DNA library of 2 × 10^6 independently derived recombinants were obtained from an Mbo I partial digest of transfectant cell 23.4 11+ DNA. 96 phage that hybridized with radiolabeled human DNA were plaque purified and a number of these phage isolates (λ11, λ22, and λ39) also hybridized with the LV7 probe. DNA inserts of these three phage ranged in size from 12 to 20 kb and contained the 5.3-kb Bam HI restriction segment shared by all of the primary and secondary L-cell transfectants (Kamarck et al., 1987).

Defined restriction fragments of TM1-CEA cDNA were subcloned for analysis of the genomic DNA. A 5′ untranslated region probe (cDNA nucleotides 0–72) hybridized to genomic phage λ11, λ22, and λ39. A 3′ UTR probe (3′UTR3;

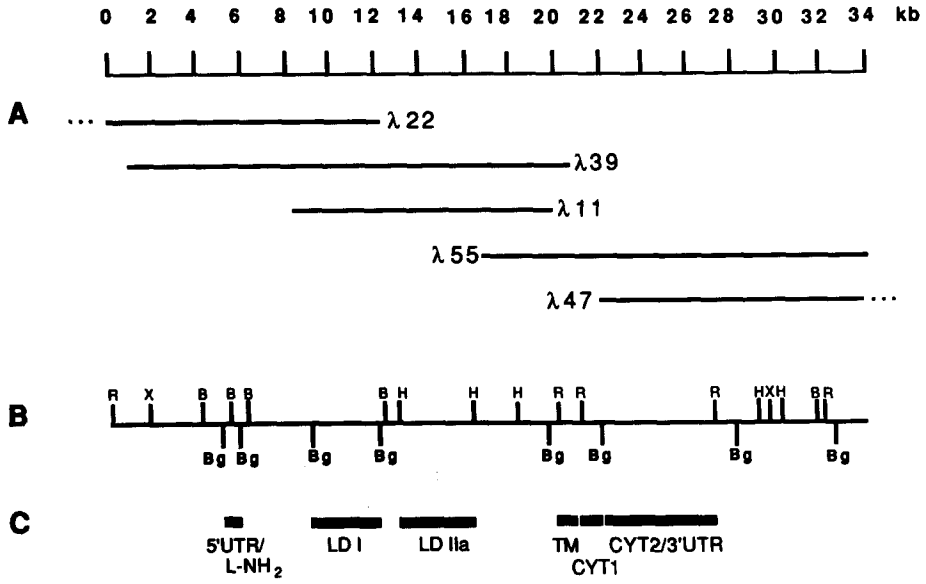

Figure 5. Map of TM-CEA genomic DNA. (*A*) Extent of human genomic DNA contained in each λ phage selected from the transfectant cell library. Dots indicate that the phage insert extends beyond that shown. (*B*) Composite restriction map contained in the five λ phage shown above. Symbols are: *R*, Eco RI; *B*, Bam HI; *Bg*, Bgl II; *H*, Hind III; and *X*, Xho I. (*C*) Restriction segments of genomic DNA homologous to subcloned portions of cDNA representing structural domains in the apoprotein. These segments are described in the text.

cDNA nucleotides 2,917-3,461) did not hybridize to these phage, but did identify two additional human recombinant phage (λ47 and λ55) derived from the transfectant genomic library. Restriction site mapping demonstrated that all of the identified phage overlapped (Fig. 5, *A* and *B*). Localization of the 5′ and 3′ UTR to genomic restriction segments indicated that the cDNA coding sequence is contained within ~23 kb of genomic DNA (Fig. 5, *B* and *C*).

cDNA restriction segments defining structural domains within TM-CEA apoprotein were used to determine their location within genomic DNA (Fig. 5 *C*). For example, cDNA sequences coding for the LD IIa structural domain (cDNA nucleotides 1017-1343; Fig. 2) were localized to a 3.8-kb Hind III genomic segment (Fig. 5 *C*). In a similar manner, coding sequences for the leader and NH_2-terminal domains (L-NH_2), the first disulfide-loop domain (LD I), the transmembrane domain (TM), and cytoplasmic domains (CYT1 and CYT2) were localized to genomic restriction segments (Fig. 5 *C*).

To look for other genomic sequences related to the TM-CEA cytoplasmic domain, the CYTO probe (nucleotides 1,443-1,676) was hybridized to a Southern blot of human genomic DNA cleaved with several restriction endonucleases. In contrast to the complex pattern obtained at high stringency with the domain probe LV7 (Kamarck et al., 1987), hybridization performed even at reduced stringency (60°C; 6× SSPE) produced one band per lane indicating that there is a single cytoplasmic region in the genome (data not shown).

Delineation of Exon-Intron Boundaries

To support the evidence for alternative RNA splicing, we sequenced selected exon boundaries of the TM-CEA genomic DNA. For example, the 3′ end of LD I was sequenced from a subclone of the 2.5-kb Bgl II genomic segment (Fig. 5 *B*) that hybridizes with the loop-domain probe, LV7. Double-stranded DNA sequencing using the dideoxy chain termination method was performed after extending an oligonucleotide primer from the 3′ end of LD I, in this example, into adjacent intron sequence. A summary of selected exon–intron junction sequences is presented in Table II. Our data demonstrate that the structural domains defined as LD IIa, transmembrane, and cytoplasmic regions are coded for by discrete exons flanked by splice donor–acceptor sites. In addition, the nucleotide sequences of the exon–intron boundaries confirm the proposed location of junctions for alternative splicing based on the cDNA sequences of TM1- through TM4-CEA.

Biochemical Analysis of TM-CEA Isoantigens

Western immunoblot analyses, using a broadly reactive anti-CEA polyclonal antibody on lysates of 23.4 11+ cells, showed a complex array of antigens with molecular masses ranging from 50 to 150 kD (Fig. 6, lane *A*). This result is consistent with the existence of multiple TM-CEA isoantigens displaying extensive N-linked glycosylation. To assess the number of different TM-CEA apoprotein species actually present, we subjected 23.4 11+ polypeptides to enzymatic deglycosylation with peptide/N-glycanase isolated from *F. meningosepticum* (Tarentino et al., 1985). As shown in Fig. 6, lane *B*, western analysis of deglycosylated transfectant cell proteins demonstrates the presence of four to five CEA isoantigen apoproteins. In contrast a lysate of control L cells shows no reactive species (Fig. 6, lane *C*). The mobilities of the major bands in lane *B* are consistent with the molecular sizes of apoproteins calculated from translation of cDNA sequences: TM1, 58 kD; TM2, 47 kD; TM3, 51 kD; TM4, 40 kD. The presence of an apparent doublet for the smallest species is presently not explained, but may be due to inaccessibility of some N-linked oligosaccharides to enzyme.

Discussion

In the past 20 yr, an array of 8–10 CEA-related glycoproteins has been identified by biochemical methods in a variety of tumor and nontumor cells (for review see Shively and Beatty, 1985). Recently, the genes for two members of this isoantigen family, CEA and NCA, have been cloned (Beauchemin et al., 1987; Oikawa et al., 1987; Barnett et al., 1988; Neumaier et al., 1988; Tawaragi et al., 1988). In this study, we identify mRNAs, cDNAs, and a genomic transcription unit that encode novel CEA isoantigen members with transmembrane and cytoplasmic domains (TM-CEAs). Immunological data confirm that TM-CEAs are expressed at the cell surface. Although we cannot assign TM-CEA glycoproteins to previously identified CEA isoantigens (Shively and Beatty, 1985), on the basis of molecular size we suggest that TM1-CEA may be related to the so-called "128K antigen" (Neumaier et al., 1985).

We have shown that an L-cell secondary transfectant, 23.4 11+, and a number of tumor cell lines express a set of four RNA transcripts (3.9, 3.7, 2.2, and 1.8 kb) that are unrelated to CEA (4.0 and 3.6 kb) or NCA (3.0 kb) poly A+ RNAs. As the secondary transfectants likely contain a single genomic transcription unit, we proposed that these mRNAs represented the products of alternative RNA splicing (Kamarck

Table II. DNA Sequence at Exon–Intron Junctions of TM-CEA Genomic DNA

Junction	Donor	Acceptor
LD I-LD IIa	. . . GTC ACT G gtaag . . .	. . . gacag AG CTA AGT . . .
	1030	1031
LD IIa-TM	. . . GTA AAC T gtaag . . .	. . . gacag AT AAT GCT . . .
	1318	1319
TM-CYT1	. . . ACC GGC AG gtatg . . .	. . . tttag G GCA AGC . . .
	1448	1449
CYT1-CYT2	. . . AAC CAC A gtaag . . .	. . . ag CT CAG GAC . . .
	1501	1502

Appropriate genomic restriction segments were subjected to DNA sequencing near the regions that are deleted or altered by splicing in TM2- and TM3-CEA cDNAs. Capitalized letters represent sequences contained in cDNA and genomic DNA, and adjacent small type represents intron sequence. Numbers refer to their positions in TM1-CEA cDNA. Abbreviations describing the exons are as defined in the text.

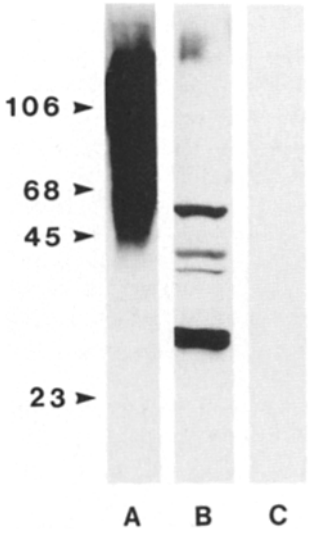

Figure 6. TM-CEA protein isoantigens expressed by 23.4 11+ transfectants. After transfer from gels, CEA isoantigens were detected using DAKO anti-CEA Ig followed by goat anti-rabbit alkaline phosphatase. (Lane *A*) Lysate from 23.4 11+ cells; (lane *B*) enzymatically deglycosylated lysate from 23.4 11+ cells; and (lane *C*) lysate from mouse L cells. Numbers on the left indicate the apparent molecular masses in kilodaltons of prestained molecular mass markers phosphorylase B (*106*), BSA (*68*), ovalbumin (*45*), and α-chymotrypsin (*23*).

et al., 1987). The cDNA sequences of the TM-CEAs reported here, coupled with genomic DNA sequences at exon–intron junctions, demonstrate that the mRNAs are indeed derived by alternative splicing from a single transcript. While TM1-CEA represents the largest transmembrane molecule, TM2-CEA lacks LD IIa, and TM3-CEA is defined by an alternate splice within the cytoplasmic domain, resulting in a unique intracellular peptide sequence. The existence of an mRNA that contains both alternate splices, TM4-CEA, is supported by expression in the transfectants of a 1.8-kb poly A+ RNA (Fig. 1, lane *a*) and a 40-kD apoprotein (Fig. 6).

We also have cloned, as overlapping recombinant phage, the complete TM-CEA transcription unit which is encoded in ~23 kb of genomic DNA. Hybridization with cDNA probes indicates that structural domains defined by comparison with other CEA-related genes correspond to discrete exons (Fig. 5). This was confirmed by the demonstration that regions adjacent to predicted exons are not found in cDNA and have consensus splice donor–acceptor sequences (Table II). These exon–intron junctions correspond exactly to the points of difference detected in the sequences of the alternatively spliced TM-CEAs.

Cell lines derived from colorectal tumors primarily produce CEA poly A+ RNA (Barnett et al., 1988), consistent with their secretion of 180-kD CEA. Using specific cDNA probes we have demonstrated that TM-CEAs are also expressed by a number of these colorectal lines, in addition to other tumor types. In some cell lines, only the set of TM-CEA mRNAs are expressed (e.g., BeWo; Fig. 4, lane *f*), while in others, numerous CEA-related RNAs are additionally produced (e.g., HT-29; Fig. 1, lane *b*). The expression of TM-CEAs is evidently not restricted to neoplastic tissue as TM1-CEA cDNAs have been cloned from an apparently normal fetal liver cDNA library. We cannot as yet correlate the expression of one or more CEA isoantigens with biological function, cellular morphology, substrate adhesion, or metastatic potential.

Members of the CEA isoantigen family display considerable nucleotide and amino acid sequence conservation (Table I). With the exception of the "half" disulfide loop-domain of TM1-CEA, all of the proposed extracellular regions are ≥70% similar at the amino acid level and ≥80% similar at the nucleotide level. This explains the ability of the TM-CEAs to be detected by several anti-CEA monoclonal antibodies and also illustrates the general cross-reactivity of anti-CEA immunoreagents.

Extracellular loop domains in CEA isoantigens demonstrate conserved amino acids surrounding the cysteine residues that are reminiscent of immunoglobulins and immunoglobulin superfamily proteins (Hunkapiller and Hood, 1986). We have noted previously (Barnett et al., 1988), and also observe here, the overall structural and amino acid similarity of all CEA glycoproteins to members of the immunoglobulin gene superfamily (Williams and Barclay, 1988). This family extends to cell surface recognition proteins (Thy-1 [Williams and Gagnon, 1982]), cell adhesion proteins (neural cell adhesion molecule [N-CAM; Cunningham et al., 1987], ICAM [Simmons et al., 1988], T cell erythrocyte receptor CD2 [Peterson and Seed, 1987], and S- and L-MAG [Salzer et al., 1987; Arquint et al., 1987]), receptor proteins (T4 receptor [Maddon et al., 1985] and PDGF receptor [Yarden et al., 1986]), and oncogenes (CSF-1 receptor/v-fms [Hampe et al., 1984]). Although their ligands have not yet been identified, we propose by analogy that members of the CEA isoantigen family act as cell surface recognition proteins.

This analogy can be extended to the alternative structural forms of neural cell adhesion molecule (N-CAM) and myelin-associated glycoprotein (MAG). Three different N-CAM isoantigens are generated from a single genomic segment (Cunningham et al., 1987): the *ssd* form, with a short hydrophobic COOH-terminus (like CEA and NCA), the *sd* form, with a short cytoplasmic domain (like TM3/4-CEA), and the *ld* form, with a longer cytoplasmic domain (TM1/2-CEA). Two cytoplasmic forms of MAG (S-MAG and L-MAG) have also been described (Salzer et al., 1987). For MAG, N-CAM, and TM-CEA, the alternative cytoplasmic regions are derived by differential splicing of those exons that code for intracellular peptide segments. A novel N-CAM has been described recently in human skeletal muscle where the *ssd* form contains a previously unidentified 37-amino acid extracellular domain (Dickson et al., 1987). This may compare with the TM1- and TM3-specific "half" disulfide-loop in these forms of CEA.

Alternative forms of the cytoplasmic region of N-CAM and L-MAG are derived by appropriate splice selection of in-frame exons (Cunningham et al., 1987; Salzer et al., 1987). In contrast, TM3-CEA and the proposed TM4-CEA contain a short but novel cytoplasmic peptide generated by a splice that results in out-of-frame translation. With this splice, the coding sequence for TM1/2-CEA cytoplasmic region becomes the 3′ UTR and also contains the consensus polyadenylation sequence (Figs. 2 and 3). To our knowledge, this result is unique among alternatively spliced RNAs. Unusual alternate COOH-terminal splices accompany the synthesis of the glial and endothelial cell forms of platelet-derived growth factor A chain (Tong et al., 1987; Bonthron et al., 1987) and of the hydrophobic and hydrophilic termini of decay accelerating factor (Caras et al., 1987), but in each case, the 3′ untranslated sequence and polyadenylation site remains constant.

The cytoplasmic domain of TM-CEA is unique, as it is represented as a single band in Southern blot genomic analysis and bears no significant resemblance to other CEA- or non-CEA-related sequences based on computer-aided searches

```
EPPELDLSYSHSDKGKRPTKDSYTLTEELAEYAEIRVK     L-MAG
*||  ***||       ||  | *||   *  | |**
DPPMNEVTYSEAQQPTQPTSASPSLTATEIIYSEVKKQ     TM1/2-CEA
     |     |
     NK    TLNF
```

Figure 7. Amino acid similarity in the cytoplasmic region of immunoglobulin gene superfamily members, L-MAG and TM-CEA. The carboxyl-terminal residues of L-MAG and TM1-/TM2-CEA are shown; the absolute COOH-terminus for each is on the right. Regions of identity or similarity are indicated, respectively, by a solid line (|) or an asterisk (*) extending between compared residues. For best physical alignment, deletion of some amino acids in TM-CEA relative to L-MAG is required; this is indicated by ^.

of the GenBank or EMBL data bases, or by direct comparison of 20 different immunoglobulin superfamily members. However, a segment of limited similarity has been detected between the COOH-terminal 44 amino acids of the TM1/TM2-CEA cytoplasmic domain and the COOH-terminal 38 amino acids of the L-MAG cytoplasmic domain (19/38 amino acids; Fig. 7). This cytoplasmic segment is relatively high in amino acids that are potential substrates for phosphorylation i.e., serine, threonine, and tyrosine (13/48 residues), and it is largely these amino acids that are shared between the molecules. Consensus signals for associated enzymatic activities like tyrosine kinase (Bairoch and Claviere, 1988) are not found.

In Fig. 8, we illustrate our interpretation of the extensive structural similarity among CEA isoantigens using immunoglobulin molecules as a model for their organization (Hunkapiller and Hood, 1986). CEA isoantigens are depicted as membrane bound: for CEA and NCA, the membrane attachment may occur via a phosphoinositol glycan moiety (Takami et al., 1988), while the TM-CEAs have structural domains that make them integral membrane glycoproteins. While we depict all CEA isoantigens as single polypeptide chains, sequence analysis of the TM-CEAs indicates that these proteins have the potential to exist in vivo as cell surface homo- or heterodimers. LD I of TM-CEAs contains a single cysteine residue (amino acid position 308; Fig. 2) that may not be involved in intrachain disulfide linkage and therefore by analogy, TM-CEA homodimers and heterodimers are depicted as immunoglobulin-like in surface character.

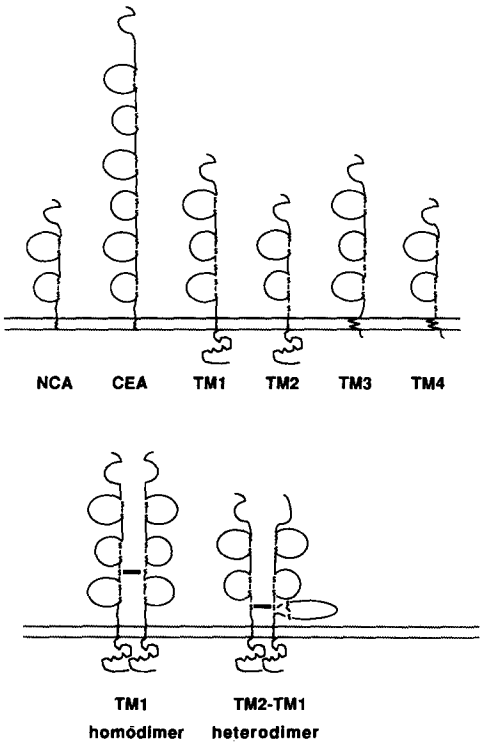

Figure 8. Hypothetical single chain and dimer forms of CEA isoantigens as immunoglobulin-like molecules. Each CEA family member is depicted with extracellular disulfide loops and is embedded in the lipid bilayer via phosphoinositol linkage (for CEA and NCA) or is an integral cell surface molecule (for TM-CEAs). Some TM-CEA members are depicted as homo- or heterodimers.

We wish to thank Mary Ann Nothdurft, Judy Dziuba, and Karen Wallberg for excellent technical assistance; Drs. Frank Ruddle and Vincent Marchesi for support; and gratefully acknowledge the many astute and helpful suggestions of Peter M. M. Rae. Quality graphics were by Suzy Pafka.

Received for publication 11 August 1988, and in revised form 10 October 1988.

Note Added in Proof: Hinoda et al. (1988. Proc. Natl. Acad. Sci. USA. 84:5350-5354) recently described the sequence of biliary glycoprotein I (BGP-I) which appears virtually identical to TM1-CEA. There are several significant differences between BGP-I mRNA and TM1-CEA mRNA described here. These include a shortened BGP-I cytoplasmic coding region which is attributable to a single nucleotide deletion at their position 1,401, a completely unrelated 3' UTR when compared to TM1-CEA, and drastically different Northern blot patterns when using unique cDNA segments.

References

Arquint, M., J. Roder, L.-S. Chia, J. Down, D. Wilkinson, H. Bayley, P. Braun, and R. Dunn. 1987. Molecular cloning and primary structure of myelin-associated glycoprotein. *Proc. Natl. Acad. Sci. USA.* 84:600-604.

Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408-1412.

Bairoch, A., and J.-M Claviere. 1988. Sequence patterns in protein kinases. *Nature (Lond.).* 331:22.

Barnett, T., S. Goebel, M. A. Nothdurft, and J. Elting. 1988. Carcinoembryonic antigen family: characterization of cDNAs coding for NCA and CEA, and suggestion of non-random sequence variation in their conserved loop domains. *Genomics.* 3:59-66.

Beauchemin, N., S. Benchimol, D. Cournoyer, A. Fuks, and C. Stanners. 1987. Isolation and characterization of full-length functional cDNA clones for human carcinoembryonic antigen. *Mol. Cell. Biol.* 7:3221-3230.

Benton, W., and R.W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. *Science (Wash. DC).* 196:180-182.

Bonthron, D., C. Morton, S. Orkin, and T. Collins. 1988. Platelet-derived growth factor A chain: gene structure, chromosomal location, and basis for alternative mRNA splicing. *Proc. Natl. Acad. Sci. USA.* 85:1492-1496.

Caras, I., M. Davitz, L. Rhee, G. Weddell, D. Martin, and V. Nussenzweig. 1987. Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature (Lond.).* 325:545-549.

Cunningham, B., J. Hemperley, B. Murray, E. Prediger, R. Brackenbury, and G. Edelman. 1987. Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science (Wash. DC).* 236:799-806.

Darcy, D., C. Turberville, and R. James. 1973. Immunological study of carcinoembryonic antigen (CEA) and a related glycoprotein. *Br. J. Cancer.* 28:147-154.

Dickson, G., H. Gower, C. Barton, H. Prentice, V. Elsom, S. Moore, R. Cox, C. Quinn, W. Putt, and F. Walsh. 1987. Human muscle neural cell adhesion molecule (N-CAM): identification of a muscle-specific sequence in the extracellular domain. *Cell.* 50:1119-1130.

DiLella, A., and S. Woo. 1987. Cloning large segments of genomic DNA using cosmid vectors. *Methods Enzymol.* 152:199-212.

Engvall, E., J. Shively, and M. Wrann. 1978. Isolation and characterization of the normal crossreacting antigen: homology of its NH_2-terminal amino acid sequence with that of carcinoembryonic antigen. *Proc. Natl. Acad. Sci. USA.* 75:1670-1674.

Feinberg, A., and B. Vogelstein. 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 137:266-267.

Gold, P., and S. Freedman. 1965. Specific carcinoembryonic antigens of the human digestive system. *J. Exp. Med.* 122:467-481.

Gubler, U., and B. Hoffman. 1983. A simple and efficient method for generat-

ing cDNA libraries. *Gene (Amst.)*. 25:263–269.
Hampe, A., M. Gobet, C. Sherr, and F. Galibert. 1984. Nucleotide sequence of the feline retroviral oncogene v-fms shows unexpected homology with oncogenes encoding tyrosine-specific protein kinases. *Proc. Natl. Acad. Sci. USA*. 81:85–89.
Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene (Amst.)*. 28:351–359.
Hunkapiller, T., and L. Hood. 1986. The growing immunoglobulin gene superfamily. *Nature (Lond.)*. 323:15–16.
Kamarck, M., J. Elting, J. Hart, S. Goebel, P. M. M. Rae, M. A. Nothdurft, J. Nedwin, and T. Barnett. 1987. Carcinoembryonic antigen family: expression in a mouse L-cell transfectant and characterization of a partial cDNA in bacteriophage λgt11. *Proc. Natl. Acad. Sci. USA*. 84:5350–5354.
Laemmli, U. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.)*. 227:680–685.
Lehrach, H., D. Diamond, J. Wozney, and H. Boedtker. 1977. RNA molecular weight determinations by electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry*. 16:4743–4751.
Maddon, P., D. Littman, M. Godfrey, D. Maddon, L. Chess, and R. Axel. 1985. The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family. *Cell* 42:93–104.
Maniatis, T., E. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 368–369.
Mullins, J., D. Brody, R. Binari, and S. Cotter. 1984. Viral transduction of c-myc gene in naturally occurring feline leukaemias. *Nature (Lond.)*. 308: 856–859.
Neumaier, M., U. Fenger, and C. Wagener. 1985. Monoclonal antibodies for carcinoembryonic antigen (CEA) as a model system: identification of two novel CEA-related antigens in meconium and colorectal carcinoma tissue by Western blots and differential immunoaffinity chromatography. *J. Immunol.* 135:3604–3609.
Neumaier, M., W. Zimmermann, L. Shively, Y. Hinoda, A. Riggs, and J. Shively. 1988. Characterization of a cDNA clone for the nonspecific cross-reacting antigen (NCA) and a comparison of NCA and carcinoembryonic antigen. *J. Biol. Chem.* 263:3202–3207.
Oikawa, S., H. Nakazoto, and G. Kosaki. 1987. Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence. *Biochem. Biophys. Res. Commun.* 142:511–513.
Paxton, R., G. Moser, H. Pande, T. Lee, and J. Shively. 1987. Sequence analysis of carcinoembryonic antigen: identification of glycosylation sites and homology with immunoglobulin supergene family. *Proc. Natl. Acad. Sci. USA*. 84:920–924.
Peterson, A., and B. Seed. 1987. Monoclonal antibody and ligand binding sites of the T cell erythrocyte receptor (CD2). *Nature (Lond.)*. 329:842–846.
Salzer, J., W. P. Holmes, and D. R. Colman. 1987. The amino acid sequences of the myelin-associated glycoproteins: homology to the immunoglobulin gene superfamily. *J. Cell Biol.* 104:957–965.
Sanger, F., S. Nicklen, and A. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–5467.
Shively, J., and J. Beatty. 1985. CEA-related antigens: molecular biology and clinical significance. *CRC Crit. Rev. Oncol. Hematol.* 2:355–399.
Simmons, D., M. Makgoba, and B. Seed. 1988. ICAM, an adhesion ligand of LFA-1, is homologous to the neural cell adhesion molecule NCAM. *Nature (Lond.)*. 331:624–627.
Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.
Takami, N., Y. Misumi, M. Kuroki, Y. Matsuoka, and Y. Ikehara. 1988. Evidence for carboyl-terminal processing and glycolipid-anchoring of human carcinoembryonic antigen. *J. Biol. Chem.* 263:12716–12720.
Tarentino, A., C. Gomez, and T. Plummer. 1985. Deglycosylation of asparagine-linked glycans by peptide:N-glycanase F. *Biochemistry*. 24: 4665–4671.
Tawargi, Y., S. Oikawa, Y. Matsuoka, G. Kosaki, and H. Nakazoto. 1988. Primary structure of nonspecific crossreacting antigen (NCA), a member of the carcinoembryonic antigen (CEA) gene family, deduced from cDNA sequence. *Biochem. Biophys. Res. Commun.* 150:89–96.
Thompson, J., H. Pande, R. Paxton, L. Shively, A. Padma, R. Simmer, C. Todd, A. Riggs, and J. Shively. 1987. Molecular cloning of a gene belonging to the carcinoembryonic antigen gene family and discussion of a domain model. *Proc. Natl. Acad. Sci. USA*. 84:2965–2969.
Tong, B., D. Auer, M. Jaye, J. Kaplow, G. Ricca, E. McConathy, W. Drohan, and T. Deuel. 1987. cDNA clones reveal differences between human glial and endothelial cell platelet-derived growth factor A-chains. *Nature (Lond.)*. 328:619–621.
Towbin, H., T. Stahelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA*. 76:4350–4354.
von Kleist, S., G. Chavanel, and P. Burtin. 1972. Identification of an antigen from normal human tissue that crossreacts with the carcinoembryonic antigen. *Proc. Natl. Acad. Sci. USA*. 69:2492–2496.
Wickens, M., and P. Stephenson. 1984. Role of conserved AAUAAA sequence: four AAUAAA point mutations prevent messenger RNA 3′ end formation. *Science (Wash. DC)*. 226:1045–1051.
Williams, A., and J. Gagnon. 1982. Neuronal cell Thy-1 glycoprotein: homology with immunoglobulin. *Science (Wash. DC)* 216:696–703.
Williams, A., and A. Barclay. 1988. The immunoglobulin superfamily-domains for cell surface recognition. *Annu. Rev. Immunol.* 6:381–405.
Yarden, Y., J. Escobedo, W.-J. Kuang, T. Yang-Feng, T. Daniel, P. Tremble, E. Chen, M. Ando, R. Harkins, U. Francke, V. Fried, A. Ullrich, and L. Williams. 1986. Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. *Nature (Lond.)* 323:226–232.