Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 2755–2759, April 1990
Cell Biology

Epithelial glycoprotein is a member of a family of epithelial cell surface antigens homologous to nidogen, a matrix adhesion protein

(tumor-associated antigens/cell-adhesion molecules/monoclonal antibodies/epidermal growth factor repeat/thyroglobulin domain)

BABETTE SIMON, DANIEL K. PODOLSKY, GERHARD MOLDENHAUER, KURT J. ISSELBACHER, SEBASTIANO GATTONI-CELLI, AND STEPHEN J. BRAND

Gastrointestinal Unit and Cancer Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114

Contributed by Kurt J. Isselbacher, December 29, 1989

ABSTRACT **The cell surface antigen, epithelial glycoprotein, defined by the monoclonal antibody HEA 125, is expressed on virtually all epithelial cell membranes but not on mesodermal or neural cell membranes. The cDNA encoding epithelial glycoprotein was isolated by HEA 125 antibody enrichment of colon tumor cDNA expressed transiently in COS cells. The sequence of the epithelial glycoprotein antigen is identical to the cell membrane protein recognized by the monoclonal antibody KS 1/4 and is homologous to the tumor-associated antigen GA733. These proteins share sequence homology to nidogen, an extracellular matrix component that appears to participate in cell–matrix adhesion. These proteins also share a homologous domain found in the B1 chain of laminin, a matrix adhesion protein, and placental protein 12, an insulin-like growth factor I binding protein secreted during pregnancy that has been implicated in regulation of fetal growth. This common domain is also repeated multiple times within the thyroglobulin precursor. These findings suggest epithelial glycoprotein is a cell surface molecule involved in cell–cell or cell–matrix interaction.**

The normal epithelial cell surface contains a complex and functionally diverse array of membrane glycoproteins that mediate many different cellular recognition functions. These include a wide variety of receptors and transporters as well as glycoproteins that mediate cell–cell or cell–matrix interactions (1, 2). Often these glycoproteins have been recognized through characterization of surface antigens on tumor cells by the development of monoclonal antibodies after immunization of mice with cells or cell-membrane preparations (3–6). Characterization of antigens first identified through their association with transformation has led to the recognition of functionally important constituents of the normal epithelial cell surface. Indeed, most tumor-associated cell surface antigens identified in this manner have proven to be present on nontransformed cells (3).

Although the functions of most tumor cell-associated surface determinants identified in this manner remain unknown, some of those membrane glycoprotein antigens may participate in cell–cell or cell–matrix interactions (7, 8). Thus the carcinoma embryonic antigen (CEA) was identified after immunization with tumor tissue (37). However, subsequent studies demonstrated that CEA is not a tumor-specific antigen (38). Although neoplastic transformation results in quantitative and qualitative abnormalities in CEA expression, it is also expressed in normal epithelial cells (7). The homogeneous cell surface expression of tumor cells resembles the pattern of surface expression seen in fetal epithelia (7). Most interestingly, studies have shown that CEA confers homophilic binding on CEA-expressing cells (7).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Among surface antigens defined by monoclonal antibodies (mAbs), the epithelial glycoprotein antigen (EGP) recognized by mAb HEA 125 may prove especially interesting (9, 10). Immunoprecipitation studies have indicated that the mAb HEA 125 recognizes a protein epitope on a 34-kDa membrane glycoprotein (10). Although mAb HEA 125 was first isolated in attempts to define a tumor-specific marker, immunohistochemical studies of human tissues demonstrated the expression of EGP in almost all normal epithelial tissues, with few exceptions (epidermal keratinocytes, gastric parietal cells, thymic cortical epithelium, myoepithelial cells, and hepatocytes). However, EGP expression has not been detected in either normal or malignant cells of mesodermal or neuroectodermal origin (9).

Although EGP is expressed in a polarized manner at the basolateral surface of normal epithelial cells, EGP in staining was more homogeneous on the membranes of tumor cells (9). Therefore, the changes in EGP expression seen with transformation resemble those of CEA, which has been shown to contribute to cell–cell adhesion of some colon tumor cells (7). These observations suggest that EGP may also play a role in fundamental cell–cell or cell–matrix interactions of epithelial cells.

To further elucidate the function of EGP, the structure of EGP* was determined through molecular cloning using mAb HEA 125 in the "cell-panning" technique (13, 14). The EGP cDNA encodes a transmembrane protein that is identical to the antigen recognized by the mAb KS 1/4 (11) and highly homologous to the GA 733-1 antigen (12). Most significantly EGP also shares significant sequence homology to nidogen, an extracellular matrix protein, as well as an insulin-like growth factor I (IGF-I) binding protein, placental protein 12. This suggests that EGP may be a member of a tumor-associated antigen family of adhesion molecules, growth factor receptors, or both.

MATERIALS AND METHODS

Cell Lines. The colon carcinoma cell lines HT-29, CaCo-2, SW837, and SW948 as well as the HPB lymphoblastic leukemia cell line were obtained from the American Tissue Culture Collection. The SY5Y neural cell lines were provided by Gerhard Heinrich (Massachusetts General Hospital) and the COS cells were provided by Brian Seed (Massachusetts General Hospital).

Isolation of EGP cDNA Clones. The cDNAs encoding the EGP surface antigen were isolated using the antibody-enrichment method of Seed and Aruffo (13, 14). A cDNA

Abbreviations: EGP, epithelial glycoprotein antigen; mAb, monoclonal antibody; CEA, carcinoma embryonic antigen; IGF-I, insulin-like growth factor I.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M32306).

library, prepared from the HT-29 (human colon carcinoma derived) cell line, was ligated into the cdm8 expression vector and then transfected into COS cells by the DEAE-dextran method. Transfected COS cells were resuspended 48 hr later, then incubated with mAb HEA 125 (1 μg/ml), and replated onto dishes coated with affinity-purified goat anti-mouse IgG antibody. A Hirt extract of episomal DNA prepared from adherent cells was used to transform competent *Escherichia coli*. The resulting bacterial colonies were pooled, amplified, and subjected to spheroblast fusion with COS cells. Transfected COS cells expressing EGP were subjected to a second round of selection by "antibody panning." After a third round of panning, DNA was isolated from bacterial colonies by the alkaline miniprep method and transfected into COS cell cultures. The presence of EGP was detected by indirect immunofluorescence (15).

Immunoprecipitation of Biosynthetically Labeled EGP. Cells from the colon carcinoma cell line HT-29, untransfected COS cells, and COS cells transfected with a cdm8 vector expressing the EGP cDNA were incubated with cysteine- and methionine-free Dulbecco's modified Eagle's medium containing 10% (vol/vol) dialyzed fetal calf serum and [^{35}S]cysteine and [^{35}S]methionine (40 mCi/ml; specific activity, 1000 Ci/mmol; 1 Ci = 37 GBq) for 18 hr at 37°C. Cells were then washed, detached with isotonic phosphate-buffered saline (PBS)/1 mM EDTA, and lysed in PBS/0.5% Nonidet P-40/0.05% deoxycholate/0.05% SDS/1 mM phenylmethylsulfonyl fluoride. Cell lysates were precleared with protein A-Sepharose CL-4B beads (Pharmacia) coated with goat anti-mouse immunoglobulin (Cappel Laboratories). For immunoprecipitation, precleared lysates were mixed with 15 μg of mAb HEA 125 for 4 hr at 4°C followed by incubation with goat anti-mouse immunoglobulin protein A-Sepharose. The precipitate was washed four times with lysis buffer, resuspended in reducing loading buffer [0.08 M Tris·HCl, pH 6.8/2% (wt/vol) SDS/0.1 M dithiothreitol/10% (vol/vol) glycerol/0.1% bromphenol blue/5% 2-mercaptoethanol], and electrophoresed on a 10% polyacrylamide gel containing SDS. The gel was fixed, soaked in EN^{3}HANCE (New England Nuclear), dried, and exposed to Kodak x-ray film at −70°C. ^{14}C-Methylated molecular weight markers (Amersham) calibrated the molecular weight of labeled proteins.

RNA Blot Hybridization. Total RNA was extracted from cells and tissues by using the method of Cathala *et al.* (16), and poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose chromatography (17) and then subjected to Northern blot analysis after electrophoresis through a 1% denaturing agarose gel (18). RNA blots were hybridized with a ^{32}P-labeled EGP cDNA insert labeled by nick-translation (19). After hybridization, filters were washed at 65°C in 0.1× SSC/0.1% SDS (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0) and subjected to autoradiography at −70°C.

cDNA Sequence Analyses. The EGP cDNA inserts were subcloned into M13mp18 (20) and the cDNA sequence was determined by the dideoxynucleotide method (21) using ^{35}S-labeled dATP and T7 DNA polymerase (Sequenase; United States Biochemical). Some sequence data were obtained by specific priming of M13 clones with synthetic oligonucleotides prepared by an automatic DNA synthesizer (Applied Biosystems). The sequences of clone 1 and 3 were determined in both strands; clone 5 was sequenced in only one direction.

RESULTS AND DISCUSSION

Isolation of a cDNA Encoding EGP. A cDNA library prepared from the HT-29 cell line was ligated into the cdm8 vector and transfected into COS cells by the DEAE-dextran method (13, 14). The transfected cells were pooled 48 hr after transfection, incubated with mAb HEA 125, and then panned on dishes coated with goat anti-mouse affinity-purified antibody. After three washes, the adherent cells were lysed;

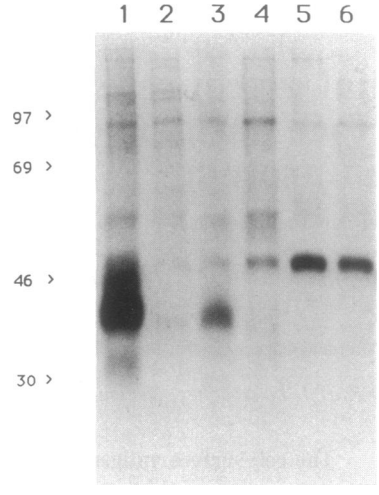

FIG. 1. Immunoprecipitation of EGP synthesized by HT-29 cells (lanes 1 and 2). EGP cDNA-transfected COS cells (lanes 3 and 4) and untransfected COS cells (lanes 5 and 6) are shown. Lysates of labeled cells were immunoprecipitated with EGP-specific antibody mAb HEA 125 (lanes 1, 3, and 5) or with control antibody mAb 45-2D9 (lanes 2, 4, and 6). Cell proteins were labeled with [^{35}S]methionine and [^{35}S]cysteine, immunoprecipitated, and electrophoresed in a 10% polyacrylamide gel under reducing conditions. Molecular masses of labeled standard proteins are shown in kDa.

episomal DNA was purified and used to transform competent *E. coli*. After repeated rounds of enrichment, plasmid DNA from three bacterial colonies expressed EGP when transfected into COS cells. Restriction analysis revealed a 1.4-kilobase (kb) insert in clone EGP1 and a 1.56-kb insert in clones EGP3 and EGP5. Clones EGP3 and EGP5 were identical by restriction analysis.

Expression of the EGP cDNA-Encoded Protein Is Identical to the HEA 125 Antigen Expressed by HT-29 Cells. Electrophoresis under reducing conditions showed a labeled protein of approximately 40 kDa in EGP3 cDNA-transfected COS cells immunoprecipitated by mAb HEA 125 but not by the control antibody mAb 45-2D9 (Fig. 1). Immunoprecipitation of HT-29 cells with mAb HEA 125 showed an identical 40-kDa protein as the major labeled protein plus a 34-kDa minor labeled protein. Prior studies have demonstrated a major 34-kDa protein and a minor 40-kDa protein immunoprecipitated by HEA 125 after labeling cell surface proteins on HT-29 cells by ^{125}I iodination. This difference in observed ratio of EGP-like proteins in HT-29 cells probably results from the different methods of radiolabeling cell proteins—the 34-kDa protein being on the cell surface whereas the 40-kDa form of EGP is the predominant intracellular form. However, this inference needs to be confirmed by pulse–chase analysis.

Expression of EGP mRNA in Various Cell Types. A ^{32}P-labeled EGP3 cDNA was hybridized to Northern blots of poly(A)$^+$ RNA isolated from various cell types (Fig. 2). A 1.6-kb EGP transcript was found in HT-29, SW948, and SW837, colon carcinoma-derived cell lines. Approximately equal levels of the 1.6-kb EGP mRNA were seen in undifferentiated (sucrase-minus) and differentiated (sucrase-positive) CaCo2 cells (Fig. 2). EGP mRNA levels in normal colon mucosa were equivalent to those seen in colon tumor cells consistent with the previous conclusion that EGP is expressed by normal colonic epithelial cells and is not a tumor-specific antigen. EGP mRNA was not expressed in the hemopoietic cell line HPB-ALL or the neural cell line SY5Y. These results are consistent with immunohistochemical data on the expression of EGP protein exclusively in epithelial cells (9, 10).

EGP cDNA Sequence. Three clones, EGP1, EGP3, and EGP5, were sequenced and shown to encode the same

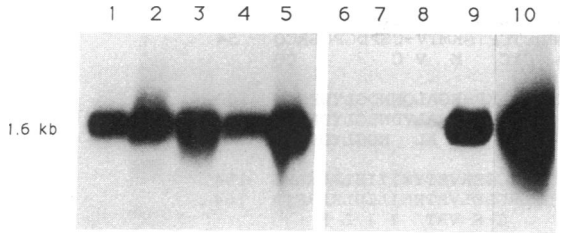

FIG. 2. Northern blot analysis of EGP mRNA expression in various tumor cell lines (lanes 1–9) and normal colon mucosa (lane 10). Poly(A)$^+$ RNA (15 μg) was electrophoresed on a 1% agarose gel, blotted, and hybridized with a ^{32}P-labeled nick-translated probe of the *Xba* I–*Xba* I cDNA insert of clone EGP3. Lanes: 1, CaCo-2 sucrase-positive colon adenocarcinoma cells; 2, CaCo-2 sucrase-negative colon adenocarcinoma; 3, HT-29 colon adenocarcinoma cell line; 4, SW837 colon adenocarcinoma cells; 5, SW948 colon adenocarcinoma cells; 6, SY5Y neural cell line; 7, HPB lymphoblastic leukemia cells; 8, untransfected COS cells; 9, HT-29 colon adenocarcinoma cell line; 10, normal colon mucosa.

protein sequence (Fig. 3). Clones EGP3 and -5 cDNA differed from EGP1 by having 160 base pairs of 5′ noncoding sequence compared to 16 base pairs in EGP1. All three clones had identical 3′ sequences terminating in an AATAAA polyadenylylation signal followed 15 nucleotides downstream by a poly(A) sequence. The EGP cDNA clones contained a 942-nucleotide open reading frame after the initiator AUG codon. This AUG codon is preceded by a CAGGC sequence, which conforms to the Kozak consensus for eukaryotic translational initiation (22).

The cDNA encodes a protein sequence of 314 amino acids that resembles a transmembrane protein. The first 21 amino acids comprise a characteristic hydrophobic signal sequence. By following the −3 −1 von Heijne rule (23), the predicted site of signal cleavage is after the Thr-Phe-Ala sequence, implying that an alanine is the N-terminal residue of EGP expressed on the cell membrane. The extracellular domain is comprised of the N-terminal 80% of mature EGP, which is anchored to the membrane by a hydrophobic 21-amino acid sequence followed by a short charged sequence of 4 basic amino acids lying in the cytoplasm that acts as a stop-transfer sequence. The predicted intracytoplasmic domain of EGP is very short, comprising only the C-terminal 26 amino acids. The predicted extracellular domain of the protein contains three potential N-linked glycosylation sites (Asn-Xaa-Ser/Thr). If these sites are glycosylated posttranslationally, the predicted molecular mass of the native protein encoded by the EGP cDNA would be approximately 40 kDa assuming

```
1    GCTCTAAAGGCGCGCGCAGGCATGGCGCCCCCGCAGGTCCTCGCGTTCGGGCTTCTGCTTGCCGCGGCGACGGCGACTTT
                          M  A  P  P  Q  V  L  A  F  G  L  L  L  A  A  A  T  A  T  F      20
                                                                SP

81   TGCCGCAGCTCAGGAAGAATGTGTCTGTGAAAACTACAAGCTGGCCGTAAACTGCTTTGTGAATAATAATCGTCAATGCC
      A  A  A  Q  E  E  C  V  C  E  N  Y  K  L  A  V  N  C  F  V  N  N  N  R  Q  C  Q     47
                        *     *                       *                 *
161  AGTGTACTTCAGTTGGTGCACAAAATACTGTCATTTGCTCAAAGCTGGCTGCCAAATGTTTGGTGATGAAGGCAGAAATG
       C  T  S  V  G  A  Q  N  T  V  I  C  S  K  L  A  A  K  C  L  V  M  K  A  E  M      73
       *                          *                       *
241  AATGGCTCAAAACTTGGGAGAAGAGCAAAACCTGAAGGGGCCCTCCAGAACAATGATGGGCTTTATGATCCTGACTGCGA
      N  G  S  K  L  G  R  R  A  K  P  E  G  A  L  Q  N  N  D  G  L  Y  D  P  D  C  D    100
                                                                                  *
321  TGAGAGCGGGCTCTTTAAGGCCAAGCAGTGCAACGGCACCTCCATGTGCTGGTGTGTGAACACTGCTGGGGTCAGAAGAA
       E  S  G  L  F  K  A  K  Q  C  N  G  T  S  M  C  W  C  V  N  T  A  G  V  R  R  T    127
                           *                    *     *
401  CAGACAAGGACACTGAAATAACCTGCTCTGAGCGAGTGAGAACCTACTGGATCATCATTGAACTAAAACACAAAGCAAGA
       D  K  D  T  E  I  T  C  S  E  R  V  R  T  Y  W  I  I  I  E  L  K  H  K  A  R     152
                              *
481  GAAAAACCTTATGATAGTAAAAGTTTGCGGACTGCACTTCAGAAGGAGATCACAACGCGTTATCAACTGGATCCAAAATT
       E  K  P  Y  D  S  K  S  L  R  T  A  L  Q  K  E  I  T  T  R  Y  Q  L  D  P  K  F    180

561  TATCACGAGTATTTTGTATGAGAATAATGTTATCACTATTGATCTGGTTCAAAATTCTTCTCAAAAAACTCAGAATGATG
       I  T  S  I  L  Y  E  N  N  V  I  T  I  D  L  V  Q  N  S  S  Q  K  T  Q  N  D  V    207

641  TGGACATAGCTGATGTGGCTTATTATTTTGAAAAAGATGTTAAAGGTGAATCCTTGTTTCATTCTAAGAAAATGGACCTG
       D  I  A  D  V  A  Y  Y  F  E  K  D  V  K  G  E  S  L  F  H  S  K  K  M  D  L     233

721  ACAGTAAATGGGGAACAACTGGATCTGGATCCTGGTCAAACTTTAATTTATTATGTTGATGAAAAAGCACCTGAATTCTC
       T  V  N  G  E  Q  L  D  L  D  P  G  Q  T  L  I  Y  Y  V  D  E  K  A  P  E  F  S    260

801  AATGCAGGGTCTAAAAGCTGGTGTTATTGCTGTTATTGTGGTTGTGGTGATAGCAGTTGTTGCTGGAATTGTTGTGCTGG
       M  Q  G  L  K  A  G  V  I  A  V  I  V  V  V  V  I  A  V  V  A  G  I  V  V  L  V    287
                                                          TM
881  TTATTTCCAGAAAGAAGAGAATGGCAAAGTATGAGAAGGCTGAGATAAAGGAGATGGGTGAGATGCATAGGGAACTCAAT
       I  S  R  K  K  R  M  A  K  Y  E  K  A  E  I  K  E  M  G  E  M  H  R  E  L  N     313

961  GCATAACTATATAATTTGAAGATTATAGAAGAAGGGAAATAGCAAATGGACACAAATTACAAATGTGTGTGCGTGGGACG
     A

1041 AAGACATCTTTGAAGGTCATGAGTTTGTTAGTTTAACATCATATATTTGTAATAGTAAAGCCTGTACTCAAAATATAAGC

1121 AGCTTGAAACTGGCTTTACCAATCTTGAAATTTGACCACAAGTGTCTTATATATGCAGATCTAATGTAAAATCCAGAACT

1201 TGGACTCCATCGTTAAAATTATTTATGTGTAACATTCAAATGTGTGCATTAAATATGCTTCCACAGTAAAATCTGAAAAA

1281 CTGATTTGTGATTGAAAGCTGCCTTTCTATTTACTTGAGTCTTGTACATACATACTTTTTTATGAGCTATGAAATAAAAC

1361 CATTTTAAACTGAAAAAAAAAAAA
```

FIG. 3. Nucleotide and derived amino acid sequences of EGP cDNA. The nucleotide numbering is at the left and amino acid numbering is at the right. The signal peptide (SP), potential N-linked glycosylation sites, and the presumed transmembrane domain (TM) are underlined; the location of cysteine residues is indicated by asterisks. The single-letter amino acid code is used.

```
EGP          MAPPQV-LAFGLLLAAATATFAAAQEECVCENYKLAVNCFVNN-NRQCQ   47
GA733     MARGPGLAPPPLRLPLLLLVLAAVTGHTAAQDNCTCPTNKMTV-CSPDGPGGRCQ   54
                 APP   L   LL  AA     AAQ  C C   K  V C        CQ

EGP       CTSVGAQNTVICSKLAAKCLVMKAEMNGSKLGRR-AKP-EGALQNNDGLYDPDCD  100
GA733     CRALGSGMAVDCSTLTSKCLLLKARMSAPKNARTLVRPSEHALVDNDGLYDPDCD  109
          C   G    V CS L  KCL  KA M   K  R    P E AL  NDGLYDPDCD

EGP       ESGLFKAKQCNGTSMCWCVNTAGVRRTDK-DTEITCSERVRTYWIIIELKHKARE  154
GA733     PEGRFKARQCNQTSVCWCVNSVGVRRTDKGDLSLRCDELVRTHHILIDLRHRPTA  164
            G FKA QCN TS CWCVN  GVRRTDK D    C E VRT  I I L H

EGP       KPYDSKSLRTALQKEITTRYQLDPKFITSILYENNVITIDLVQNSSQKTQNDVDI  209
GA733     GAFNHSDLDAELRRLFRERYRLHPKFVAAVHYEQPTIQIELRQNTSQKAAGDVDI  219
                 L   L      RY L PKF     YE   I I L QN SQK   DVDI

EGP       ADVAYYFEKDVKGESLFHSKK-MDLTVNGEQLDLDPGQTLIYYVDEKAPEFSMQG  263
GA733     GDAAYYFERDIKGESLFQGRGGLDLRVRGEPLQVE--RTLIYYLDEIPPKFSMKR  272
           D AYYFE D KGESLF      DL V GE L      TLIYY DE  P FSM

EGP       LKAGVIAVIVVVVIAVVAGIVVLVISRKKRMAKYEKAEIKEMGEMHRELNA      314
GA733     LTAGLIAVIVVVVVALVAGMAVLVITNRRKSGKYKKVEIKELGELRKEPSL      323
          L AG IAVIVVVV A VAG  VLVI       KY K EIKE GE   E
```

FIG. 4. Homology between EGP and GA733 (12) proteins. The entire sequences of EGP residues (1–314) and GA733 residues (1–323) were compared using the ALIGN program with the mutation data matrix (250-point-accepted mutations) and a gap penalty of 6. The alignment score was 39.85 standard deviations; a score greater than 8 standard deviations indicates highly significant homology. Identical amino acid residues are shown below. The single-letter amino acid code is used.

an average oligosaccharide chain length of 2.5 kDa. This predicted 40-kDa size agrees with the observed size of biosynthetically labeled EGP expressed by HT-29 cells and COS cells transfected with EGP cDNAs.

a

EGP repeat	I	(27- 59)	C X_1 C X_8 C X_7 C X_1 C X_{10} C
	II	(66-135)	C X_{32} C X_{10} C X_5 C X_1 C X_{16} C
GA733	I	(34- 66)	C X_1 C X_7 C X_8 C X_1 C X_{10} C
	II	(73-145)	C X_{34} C X_{10} C X_5 C X_1 C X_{17} C
Nidogen	IId	(776-809)	C X_4 C X_5 C X_8 C X_1 C X_{10} C
	IIe	(819-889)	C X_{28} C X_{10} C X_6 C X_1 C X_{20} C
Placental protein 12		(150-225)	C X_{29} C X_{10} C X_{10} C X_1 C X_{20} C
Thyroglobulin II		(96-160)	C X_{23} C X_{10} C X_6 C X_1 C X_{19} C

b

```
 1  PDCDESGLFKAKQCNGTSM     CWCVNTA GVRRTDKDTEIT     CSERV
 2  PDCDPEGRFKARQCNQTSV     CWCVNSV GVRRTDKGDLSLR    CDELV
 3  PQCDEYGHYVPTQCHHSTGY    CWCVDRD GRELEGSRTPP GMRPPCLSTV
 4  PNCNKNGFYHSRQCETSMDGEAGLCWCVYPWNGKRIPGSPEIR GDPN CQIYF
 5  PQCAEDGSFQTVQCQNDGRS    CWCVGAN GSEVLGSRQP  GRPVACLSFC
 6  PQCQDSGDYAPVQCDVQHVQ    CWCVDAE GMEVYGTRQL  GRPKRCPRSC
 7  PSCRRNGDYQAVQCQTEGP     CWCVDAQ GLEMHGTRQQ  GEPPSCAEGQ
 8  PSCTTEGSYEDVQCFSGE      CWCVNSW GKELPGSRVRD GQPR CPTDC
 9  PACTSEGHFLPVQCFNSE      CYCVDAE GQAIPGTRSAI GLPKKCPTPC
10  CDCDPRGIETP QCDQSTGQ    CVCVEGVEGPRCDKCTRGYSGVFPDCTPCH
    P C   G     QC          CWCV    G           G    C
```

c

```
EGP       24 QEECVCENYKLAVNCFVNNNR-QCQCTSVGAQNTVIC--SKLA-AKCLVMKA
NIDOGEN  773 VDECQHSRCHPDAFCYNTPGSFTCQCKPGYQGDGFRCMPGEVSKTRCQLERE
               EC          C        CQC          C         C

EGP       EMNGSKLGRRAKPEGALQNNDGLYDPDCDESGLFKAKQCN-GTSMCWCVNTAG 123
NIDOGEN   HILGAAGGADAQRP-TLQ---GMFVPQCDEYGHYVPTQCHHSTGYCWCVDRDG 873
             G   G  A     LQ   G   P CDE G     QC   T  CWCV   G
```

FIG. 5. (*a*) EGP, GA733 (12), nidogen (IId and IIe correspond to the EGF-like repeats d and e in domain II; ref. 27), placental protein 12 (28, 29), and human thyroglobulin (represented by repetition unit II; ref. 30) have similar repeats of cysteine (C) residues. Intervening amino acid residues are represented by X_n. Amino acid positions are shown in parentheses. (*b*) EGP, nidogen, placental protein 12, and laminin B1 chain (31) share common thyroglobulin-like domains. Sequences aligned are as follows. Sequences: 1, EGP residues 97–139; 2, GA733 residues 106–149; 3, nidogen residues 846–893; 4, placental protein 12 residues 179–230; 5, thyroglobulin residues 31–77; 6, thyroglobulin residues 99–145; 7, thyroglobulin residues 298–343; 8, thyroglobulin residues 599–643; 9, thyroglobulin residues 666–711; 10, laminin B1 chain residues 1111–1159. Identical amino acid residues are shown at the bottom. (*c*) EGP and nidogen share similar sequence beyond the thyroglobulin-like domain. The sequence homology was evaluated using the ALIGN program. The score was 8.67 standard deviations using the parameters in Fig. 4. Identical amino acid residues are shown below. The single-letter amino acid code is used.

The extracellular domain has two tandem sequences containing repeated cysteine residues (repeats 1 and 2, Fig. 3 and see Fig. 5*a*), which are also found within the epidermal growth factor precursor and in many transmembrane proteins that act as specific recognition molecules [low density lipoprotein receptor, cell surface immunoglobulins, T-cell receptors, and cell-adhesion molecules (24)]. It is interesting that CEA also contains cysteine repeats and, like EGP, also has a very short intracellular domain (25). These similarities suggest that EGP may function as a cell-adhesion molecule in a manner similar to CEA rather than as a receptor. The latter usually have larger intracellular domains that mediate signal transduction. However, the short intracellular domain does not preclude receptor function by EGP if it is coupled to other membrane-anchored proteins that mediate transmembrane signaling in a manner similar to the T-cell membrane proteins, CD4 and CD8. Both of these proteins have short intracellular domains but act as receptors through a tight association with the signaling lymphocyte-specific $p56_{lck}$ tyrosine kinase in the cell membrane (26).

The protein recognized by the mAb HEA 125 appears to be the same surface membrane protein recognized by mAb KS 1/4 reported by Perez and Walker (11). EGP is also highly homologous to another tumor antigen recognized by the mAb GA 733 (12) (Fig. 4). Although sequence identities extend through the entire length of the protein, highest sequence conservation occurs between residues in the second cysteine repeat (EGP amino acids 91–130) and in the hydrophobic transmembrane domain. Furthermore, EGP shows sequence similarity to nidogen (27), a cell matrix protein, and placental protein 12, an IGF-I-binding protein secreted by the placenta (28, 29). EGP shares with these proteins a similar pattern of cysteine repeats (Fig. 5*a*) and a homologous sequence of 50 amino acids that is also repeated at least five times in the thyroglobulin precursor (ref. 30 and Fig. 5*b*). This common 50-amino acid sequence is also similar to the sequence between residues 1111 and 1159 in the laminin B1 chain—another matrix adhesion protein (ref. 31 and Fig. 5*b*).

The homology between EGP and nidogen extends beyond the thyroglobulin-like domain to include the adjacent upstream amino acids (Fig. 5*c*). In nidogen, the EGP-homologous sequence is part of a much larger multifunctional 1217-amino acid protein that comprises three domains: an N-terminal 85-kDa globular domain, a central 27-kDa rod-like domain, and a 38-kDa globular C-terminal domain. The C-terminal domain binds laminin and collagen, anchoring nidogen in the basement membrane. The central rod-like domain contains five epidermal growth factor-like cysteine repeats and a single RGD (Arg-Gly-Asp) sequence, which mediates cell adhesion (27). Although the central rod-like domain of nidogen contains the EGP-like sequence, the RGD sequence mediating nidogen cell adhesion lies upstream of the EGP-like sequence and this common cell adhesion sequence is not found in EGP.

EGP is also homologous to placental protein 12, which is secreted by the endometrium into the amniotic fluid during the first trimester of pregnancy (28, 32, 33). Interestingly, many types of tumors express high levels of placental protein 12 (34, 35). While the biological actions of placental protein 12 are unknown, the IGF-I binding capability and its secretion into amniotic fluid suggests that placental protein 12 may regulate fetal growth (36).

Although the sequence homology of EGP to nidogen and placental protein 12 may provide insights into the function of EGP, the cellular localizations of these three proteins are very different. EGP is a transmembrane protein, whereas nidogen is a cell matrix protein and placental protein 12 is a protein secreted into tissue fluids. Additional observations noted above also indicate that nidogen and placental protein 12 have very different functions. Nidogen is a cell-adhesion molecule acting as a link between the extracellular matrix and the cell membrane. In contrast, placental protein 12 is a binding protein for IGF-I in tissue fluids and has been implicated in the regulation of fetal growth. Although further studies are required, EGP could function as a cell-adhesion molecule or as a receptor for a specific ligand since the homologous domain lies on the external surface of the cell membrane.

We thank Drs. Brian Seed and Ivan Stamencovic (Department of Molecular Biology, Massachusetts General Hospital) for advice and guidance in the antibody panning cDNA cloning techniques and for providing the HT-29 colon cell cDNA expression library. This work was supported by grants from the National Institutes of Health (DK 41557 and DK 36350). B.S. was supported by a fellowship from the German Research Foundation, Federal Republic of Germany.

1. Takeichi, M. (1988) *Development* **102,** 639–655.
2. Edelman, G. (1986) *Annu. Rev. Cell Biol.* **2,** 81–116.
3. Edwards, P. A. W. (1985) *Br. J. Cancer* **51,** 149–155.
4. Herlyn, M., Steplewiski, Z., Herlyn, D. & Koprowski, H. (1986) *Hybridoma* **5,** S3–S8.
5. Varki, N. M., Reisfeld, R. A. & Walker, L. E. (1984) *Cancer Res.* **44,** 681–688.
6. Girardet, C., Vacca, A., Schmidt-Kessen, A., Schreyer, M., Carrel, S. & Mach, J. (1986) *J. Immunol.* **136,** 1497–1505.
7. Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K. & Stanners, C. P. (1989) *Cell* **57,** 327–334.
8. Shimoyama, Y., Hirohashi, S., Hirano, S., Noguchi, M., Shimosto, Y., Takeichi, M. & Abe, O. (1989) *Cancer Res.* **49,** 2128–2133.
9. Momburg, F., Moldenhauer, G., Hammerling, G. J. & Moller, P. (1987) *Cancer Res.* **47,** 2883–2891.
10. Moldenhauer, G., Momburg, F., Moller, P., Schwartz, R. & Hammerling, G. J. (1987) *Br. J. Cancer* **56,** 712–721.
11. Perez, M. S. & Walker, L. E. (1989) *J. Immunol.* **142,** 3662–3667.
12. Linnenbach, A. J., Wojcierowski, J., Wu, S., Pyrc, J. J., Ross, A. H., Dietzschold, B., Speicher, D. & Koprowski, H. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 27–31.
13. Seed, B. & Aruffo, S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3365–3369.
14. Aruffo, A. & Seed, B. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8573–8577.
15. Harlow, E. & Lane, D. (1989) *Antibodies: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Cathala, G., Savoure, J.-F., Mendez, B., West, B. L., Karin, M., Martial, J. A. & Baxter, J. D. (1983) *DNA* **2,** 329–335.
17. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
18. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5201–5205.
19. Rigby, P. W. J., Dieckmann, M., Rhoads, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
20. Messing, J. (1983) *Methods Enzymol.* **101,** 20–77.
21. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
22. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
23. von Heijne, G. (1983) *Eur. J. Biochem.* **133,** 17–21.
24. Appella, E., Weber, I. T. & Blasi, F. (1988) *FEBS Lett.* **231,** 1–4.
25. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A. & Stanners, C. P. (1987) *Mol. Cell. Biol.* **7,** 3221–3230.
26. Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. (1988) *Cell* **55,** 301–308.
27. Mann, K., Deutzmann, R., Aumailley, M., Timpl, R., Raimondi, L., Yamada, Y., Pan, T., Conway, D. & Chui, M. L. (1989) *EMBO J.* **8,** 65–72.
28. Grundman, U., Nerlich, C., Bohn, H. & Rein, T. (1988) *Nucleic Acids Res.* **16,** 8711.
29. Julkunen, M., Koistinen, R., Aalto-Setala, K., Seppala, M., Janne, O. A. & Kontula, K. (1988) *FEBS Lett.* **236,** 295–302.
30. Malthiery, Y. & Lissitsky, S. (1987) *Eur. J. Biochem.* **165,** 491–498.
31. Pikkarainen, T., Eddy, R., Fukushima, Y., Byers, M., Shows, T., Pihlajaniemi, T., Saraste, M. & Tryggvason, K. (1987) *J. Biol. Chem.* **262,** 10454–10462.
32. Rutanen, E. M., Bohn, H. & Seppala, M. (1982) *Am. J. Obstet. Gynecol.* **144,** 460–463.
33. Tornehave, D., Fay, T. N., Teisner, B., Chemnitz, J., Westergaard, J. G. & Grunzinskas, J. G. (1989) *Eur. J. Obstet. Gynecol. Reprod. Biol.* **30,** 221–232.
34. Rutanen, E. M., Wahlstrom, T., Koistinen, R., Sipponen, P., Jalanko, H. & Seppala, M. (1984) *Tumour Biol.* **5,** 95–102.
35. Iino, K., Seppala, M., Heinonen, P. K., Sipponen, P. & Rutanen, E. M. (1986) *Cancer* **58,** 2294–2297.
36. Frauman, A. G., Tsuzaki, S. & Moses, A. C. (1989) *Endocrinology* **124,** 2289–2296.
37. Gold, P. G. & Freedman, S. O. (1965) *J. Exp. Med.* **122,** 467–481.
38. Martin, F. & Martin S. (1970) *Int. J. Cancer* **6,** 352–360.