Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 3332–3336, May 1989
Medical Sciences

Molecular cloning and expression of cDNA for a carcinoembryonic antigen-related fetal liver glycoprotein

(first-trimester human fetal liver cDNA library/pregnancy-specific glycoprotein/immunoglobulin gene superfamily)

WASIF NOOR KHAN, ANNIKA OSTERMAN, AND STEN HAMMARSTRÖM*

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

Communicated by Elvin A. Kabat, December 12, 1988

ABSTRACT **Carcinoembryonic antigen (CEA) is considered to be an embryonic antigen that is reexpressed in carcinomas. However, at the molecular level little is known about fetal forms of CEA. We have studied fetal liver, which was originally considered to contain CEA. A first-trimester cDNA library from fetal liver was screened with CEA-specific probes, and a dominant cDNA clone was identified and sequenced. This 1.7-kilobase cDNA codes for a complete protein of 426 amino acids, of which 34 constitute a leader peptide. Structurally, it can be divided into four immunoglobulin-like domains homologous to CEA (N–A1–A2–B2) and a hydrophobic tail (12 residues). The A and B domains each contain two cysteines; the N domain has none. The protein has seven potential sites for asparagine-linked glycosylation. It is a form of pregnancy-specific β_1-glycoprotein (PSβG) but differs from other PSβG species at the C terminus. The N and A1 domains show 45% and 51% amino acid sequence identity with the corresponding domains of the three CEA family members whose sequences have been determined. Expression studies showed that the cDNA codes for a 72-kDa glycoprotein that reacts immunologically with antisera to CEA, biliary glycoprotein I, and PSβG. The 72-kDa glycoprotein was released from the transfected cells. At least six mRNA species were identified in human tissues by using this cDNA as a probe. Genomic DNA analysis with an N-domain-specific probe indicated that the number of genes is relatively small.**

Carcinoembryonic antigen of the human digestive tract (CEA) is considered to be a fetal antigen that is reexpressed in adenocarcinomas, particularly in the gastrointestinal tract (1). However, the presence in normal adult colon of CEA, biochemically and immunologically indistinguishable from tumor CEA (2), as well as of CEA mRNA (3), is in apparent conflict with this concept. The function of CEA is unknown, but CEA is a clinically important tumor marker of proven value in the follow-up of patients with colorectal carcinomas and has considerable promise as a target antigen for radioimmunotherapy (4).

The amino acid sequence of CEA and two CEA-related normal adult tissue antigens, nonspecific crossreactive antigen of 55 kDa (NCA-55) (5) and biliary glycoprotein I (BGPI) (6), have been determined by cloning and sequencing of the corresponding cDNAs (7–10). Thirteen, most likely different, CEA-related macromolecules have been identified (for review, see refs. 11 and 12). Southern blot analyses have indicated that there are in the order of 10 genes in the CEA family (13).

CEA and related antigens in fetal tissues have been studied only to a limited extent. Fetal intestine and meconium contain an antigen closely similar to, but not identical with CEA, termed NCA 160, NCA-2, or meconium antigen (14).

In their original work on CEA, Gold and Freedman (1) identified a molecule in human fetal liver which, in immunodiffusion, gave a reaction of identity with colon carcinoma CEA. It was expressed between months 2 and 6 of gestation and disappeared during the last trimester. In order to understand the fundamental question of the oncofetal nature of CEA, we have initiated a study of fetal forms of CEA. Here we report the cloning, sequencing, and expression of the cDNA for the major CEA-related antigen in human fetal liver.†

MATERIALS AND METHODS

Screening of Human Fetal Liver cDNA Library, and Subcloning and Sequencing of Insert cDNA. Using the amino acid sequence information from the N-terminal portion of CEA (7), we synthesized two oligodeoxynucleotide probes, a 33-mer (3′-CCG-TTC-CTC-CAC-GAC-GAC-GAC-CAC-GTG-TTG-GAC-5′) and a 45-mer (3′-CGG-GTG-TTG-GAC-GGG-GTC-TTG-GCC-TAA-CCG-ATG-AGG-ACC-ATG-TTC-5′). These probes were end-labeled with bacteriophage T4 polynucleotide kinase and [γ-^{32}P]ATP and used to screen a human fetal liver cDNA library in phage vector λgt11. The mRNA source was first-trimester male fetal liver (Clontech). cDNA inserts from positive clones were subcloned into plasmid vector pUC19. Nucleotide sequence analysis was done as described by Sanger *et al.* (15) in pUC19 with Klenow fragment of DNA polymerase I or modified bacteriophage T7 DNA polymerase (16) (United States Biochemicals).

Transfection. The fetal liver cDNA was inserted into the expression vectors pECE (17) or pCMU-5 (18) at the *Eco*RI and *Hin*dIII sites, respectively. The vectors were kindly provided by Olof Karlsson (University of Umeå) and Lars Rask (Uppsala University). Transfection into HeLa or CHO cells was carried out by the calcium phosphate method (19).

Labeling and Immunoprecipitation. Transfected cells were radioactively labeled 48 hr after the addition of DNA. CHO cells were labeled with [^{35}S]methionine for 2 hr in methionine-free growth medium. HeLa cells were labeled with [^{3}H]mannose for 12 hr in medium lacking glucose. Both cell lysates and medium were subjected to immunoprecipitation using polyclonal anti-CEA, anti-BGPI, and anti-PSβG antibodies (Dako, Santa Barbara, CA) and protein A-Sepharose (Pharmacia). The eluates from immunoprecipitates were analyzed by electrophoresis in a SDS/5–15% polyacrylamide gradient gel in a discontinuous buffer system followed by autoradiography.

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.

Abbreviations: CEA, carcinoembryonic antigen; NCA, nonspecific crossreactive antigen; FL-NCA, fetal liver NCA; PSβG, pregnancy-specific β_1-glycoprotein; BGPI, biliary glycoprotein I.

*To whom reprint requests should be addressed.

†The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04539).

Northern Blot Analysis. Poly(A)$^+$ RNA from human term placenta and adult human liver was size-fractionated in a 1.2% agarose gel containing 2.2 M formaldehyde and transferred to a nylon membrane (Hybond-N, Amersham). The blots were prehybridized and hybridized at 45°C in a buffer containing 50% (vol/vol) formamide, 5× SSPE (1× SSPE = 0.18 M NaCl/0.01 M sodium phosphate, pH 7.7/1 mM EDTA), 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, and 0.5% (wt/vol) SDS. Filters were washed once in 2× SSPE/1% SDS at 65°C for 1 hr and in 0.1× SSPE/1% SDS at 68°C for 30 min. DNA fragments containing nucleotides 400–692 (*Sma* I–*Pst* I) and 1310–1694 (*Bam*HI–*Eco*RI) of the cDNA sequence were labeled with [α-^{32}P]dCTP by random priming and used as probes. RNA bands that hybridized with the probes were detected by autoradiography.

RESULTS

Identification and Sequencing of a cDNA Clone from a Human Fetal Liver Library Coding for a CEA-Related Macromolecule. To identify the CEA-related antigen(s) in human fetal liver, a cDNA library from this source was screened with two oligonucleotide probes corresponding to the N-terminal region of CEA. Approximately 200 positive clones were identified when 10^5 recombinant phage plaques were screened. Partial sequence analysis and restriction enzyme mapping of 20 inserts ranging in length from 1.5 to 2 kilobases (kb) demonstrated that they coded for three different CEA-related molecules. One cDNA species of 1.7 kb was by far the most frequent and was chosen for complete sequence determination. The glycoprotein it encodes will be referred to as

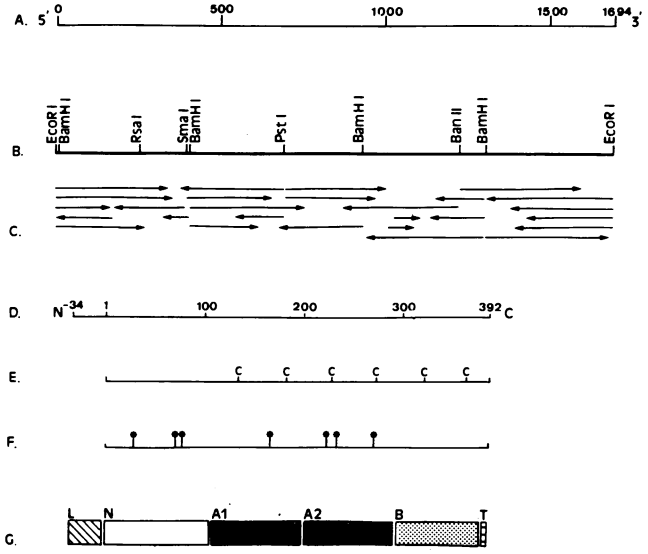

FIG. 1. Line diagrams of human FL-NCA cDNA and protein. (*A*) Nucleotide number line. (*B*) Partial restriction enzyme map. (*C*) Sequencing strategy. (*D*) Amino acid number line. (*E*) Cysteine residues. (*F*) Potential asparagine-linked glycosylation sites. (*G*) Domain structure based on internal homologies and homologies with CEA.

```
                                                                    AGAGAGTTTCTGGATCCTAGGCTTATCTCCACAGAGGAGAACACACAAGCAGCAGAGACC
-34  Met Gly Thr Leu Ser Ala Pro Pro Cys Thr Gln Arg Ile Lys Trp Lys Gly Leu Leu Leu Thr Ala Ser Leu
     ATG GGA ACC CTC TCA GCC CCT CCC TGC ACA CAG CGC ATC AAA TGG AAG GGG CTC CTG CTC ACA GCA TCA CTT  132
                                                  1
-10  Leu Asn Phe Trp Asn Leu Pro Thr Thr Ala Gln Val Thr Ile Glu Ala Glu Pro Thr Lys Val Ser Glu Gly
     TTA AAC TTC TGG AAC CTG CCC ACC ACT GCC CAA GTC ACG ATT GAA GCC GAG CCA ACC AAA GTT TCC GAG GGG  204

 15  Lys Asp Val Leu Leu Leu Val His Asn Leu Pro Gln Asn Leu Thr Gly Tyr Ile Trp Tyr Lys Gly Gln Met
     AAG GAT GTT CTT CTA CTT GTC CAC AAT TTG CCC CAG AAT CTT ACC GGC TAC ATC TGG TAC AAA GGG CAA ATG  276

 39  Arg Asp Leu Tyr His Tyr Ile Thr Ser Tyr Val Val Asp Gly Glu Ile Ile Ile Tyr Gly Pro Ala Tyr Ser
     AGG GAC CTC TAC CAT TAC ATT ACA TCA TAT GTA GTA GAC GGT GAA ATA ATT ATA TAT GGG CCT GCA TAT AGT  348

 63  Gly Arg Glu Thr Ala Tyr Ser Asn Ala Ser Leu Leu Ile Gln Asn Val Thr Arg Glu Asp Ala Gly Ser Tyr
     GGA CGA GAA ACA GCA TAT TCC AAT GCA TCC CTG CTG ATC CAG AAT GTC ACC CGG GAG GAC GCA GGA TCC TAC  420

 87  Thr Leu His Ile Ile Lys Gly Asp Asp Gly Thr Arg Gly Val Thr Gly Arg Phe Thr Phe Thr Leu His Leu
     ACC TTA CAC ATC ATA AAG GGA GAT GAT GGG ACT AGA GGA GTA ACT GGA CGT TTC ACC TTC ACC TTA CAC CTG  492

111  Glu Thr Pro Lys Pro Ser Ile Ser Ser Ser Asn Leu Asn Pro Arg Glu Thr Met Glu Ala Val Ser Leu Thr
     GAG ACT CCT AAG CCC TCC ATC TCC AGC AGC AAC TTA AAT CCC AGG GAG ACC ATG GAG GCT GTG AGC TTA ACC  564

135  Cys Asp Pro Glu Thr Pro Asp Ala Ser Tyr Leu Trp Trp Met Asn Gly Gln Ser Leu Pro Met Thr His Ser
     TGT GAC CCT GAG ACT CCA GAC GCA AGC TAC CTG TGG TGG ATG AAT GGT CAG AGC CTC CCT ATG ACT CAC AGC  636

159  Leu Lys Leu Ser Glu Thr Asn Arg Thr Leu Phe Leu Leu Gly Val Thr Lys Tyr Thr Ala Gly Pro Tyr Glu
     TTG AAG CTG TCC GAA ACC AAC AGG ACC CTC TTT CTA TTG GGT GTC ACA AAG TAT ACT GCA GGA CCC TAT GAA  708

183  Cys Glu Ile Arg Asn Pro Val Ser Ala Ser Arg Ser Asp Pro Val Thr Leu Asn Leu Leu Pro Lys Leu Pro
     TGT GAA ATA CGG AAC CCA GTG AGT GCC AGC CGC AGT GAC CCA GTC ACC CTG AAT CTC CTC CCG AAG CTG CCC  780

207  Lys Pro Tyr Ile Thr Ile Asn Asn Leu Asn Pro Arg Glu Asn Lys Asp Val Leu Asn Phe Thr Cys Glu Pro
     AAG CCC TAC ATC ACC ATC AAC AAC TTA AAC CCC AGG GAG AAT AAG GAT GTC TTA AAC TTC ACC TGT GAA CCT  852

231  Lys Ser Glu Asn Tyr Thr Tyr Ile Trp Trp Leu Asn Gly Gln Ser Leu Pro Val Ser Pro Arg Val Lys Arb
     AAG AGT GAG AAC TAC ACC TAC ATT TGG TGG CTA AAT GGT CAG AGC CTC CCG AGT CCC AGG GTA AAG CGA  924

255  Pro Ile Glu Asn Arg Ile Leu Ile Leu Pro Ser Val Thr Arg Asn Glu Thr Gly Pro Tyr Gln Cys Glu Ile
     CCC ATT GAA AAC AGG ATC CTC ATT CTA CCC AGT GTC ACG AGA AAT GAA ACA GGA CCC TAT CAA TGT GAA ATA  996

279  Arg Asp Arg Tyr Gly Gly Ile Arg Ser Asp Pro Val Thr Leu Asn Val Leu Tyr Gly Pro Asp Leu Pro Arg
     CGG GAC CGA TAT GGT GGC ATC CGC AGT GAC CCA GTC ACC CTG AAT GTC CTC TAT GGT CCA GAC CTC CCC AGA 1068

303  Ile Tyr Pro Ser Phe Thr Tyr Tyr Arg Ser Gly Glu Val Leu Tyr Leu Ser Cys Ser Ala Asp Ser Asn Pro
     ATT TAC CCT TCA TTC ACC TAT TAC CGT TCA GGA GAA GTC CTC TAC TTG TCC TGT TCT GCG GAC TCT AAC CCA 1140

327  Pro Ala Gln Tyr Ser Trp Thr Ile Asn Glu Lys Phe Gln Leu Pro Gly Gln Lys Leu Phe Ile Arg His Ile
     CCG GCA CAG TAT TCT TGG ACA ATT AAT GAA AAG TTT CAG CTA CCA GGA CAA AAG CTC TTT ATC CGC CAT ATT 1212

351  Thr Thr Lys His Ser Gly Leu Tyr Val Cys Ser Val Arg Asn Ser Ala Thr Gly Lys Glu Ser Ser Lys Ser
     ACT ACA AAG CAT AGC GGG CTC TAT GTT TGC TCT GTT CGT AAC TCA GCC ACT GGC AAG GAA AGC TCC AAA TCC 1284

375  Met Thr Val Glu Val Ser Gly Lys Trp Ile Pro Ala Ser Leu Ala Ile Gly Phe End
     ATG ACA GTC GAA GTC TCT GGT AAG TGG ATC CCA GCA TCG TTG GCA ATA GGG TTT TAGGTGGAGTCTATCTGGCATTC 1361

     AGAGAAGAGTCAGGAAAACAATTGTATTCCCAGCCTGTGTCCCATGGGCACAAGCAAATCCCAAATTCTCCTCCTGAACCCTCCAAATTTGTCTA 1456

     AGAACTTCGAAAACTTTAACAAACAGGCTGATATCTTCATAATATTCCCAGCCTAGACCAAGCAGGAAGAACATTGATTTCATTGAAATAATTGA 1551

     TAATAATGAAGATAATGTTTTTATGATTTTTATTTGAAAATTTGCTGATTCTTTAAATGGTTTGTTTTCTACATTGATGGAATTTTTCTCTTTTA 1646

     ATCTATCTACAGCTTATAGCAGTTCAATAAACTATACTTCTGGGAAAAA                                              1695
```

FIG. 2. Nucleotide sequence and deduced amino acid sequence of FL-NCA. Nucleotide numbering is at right; amino acid numbering is at left. The first 34 amino acids (−34 to −1) constitute the leader peptide. Coding sequence starts at nucleotide 61 and ends at the stop codon (nucleotides 1339–1341). Putative glycosylation sites and poly(A) signal are underlined. Cysteines are marked with dots.

fetal liver nonspecific crossreactive antigen (FL-NCA) to indicate its tissue source and crossreactivity with CEA (see below). The restriction enzyme map and sequencing strategy are shown in Fig. 1 *B* and *C*. Fig. 2 shows cDNA and the deduced amino acid sequence. The cDNA sequence includes a 1278-base-pair (bp) coding region, a 60-bp 5′ untranslated region, and a 357-bp 3′ untranslated region with a poly(A) tail.

Deduced Amino Acid Sequence of FL-NCA and Description of Structure. The FL-NCA precursor consists of 426 amino acids including a 34-amino acid leader sequence. The leader peptide contains an alanine at position −1 and a cluster of hydrophobic amino acids, which is consistent with the structure of most leader peptides (20). FL-NCA has a domain structure similar to that of CEA (8). Five domains can be distinguished (Fig. 1*G*). An N-terminal region (N) spanning residues 1–109, two A domains (A1 and A2) spanning residues 110–202 and 203–295, a B domain spanning residues 296–380, and a short hydrophobic C-terminal region (T) spanning residues 381–392. A1, A2, and B2 each contain two cysteines, whereas N has none (Fig. 1*E*). FL-NCA contains seven Asn-Xaa-Ser/Thr glycosylation sites—three each in the N and A2 domains and one in the A1 domain (Fig. 1*F*). The calculated molecular mass of the peptide moiety of FL-NCA is 44,120 daltons.

Relationship Between FL-NCA and Other Members of the CEA Gene Family. The deduced amino acid sequences of FL-NCA, pregnancy-specific β_1-glycoprotein (PSβG), CEA, NCA-55, BGPI, and a PSβG/FL-NCA gene, CGM 35, are shown in Fig. 3. The sequences have been divided into domains and aligned for maximum homology. The results of the sequence comparison can be summarized as follows. (*i*) The CEA-related fetal liver antigen studied here, FL-NCA, is a variant of PSβG. Three closely related PSβGs can be distinguished: group I, represented by PSβG clones 16 and 93 (21, 22) and PSβG clones C and D (23); group II, represented by the genomic clone CGM 35 (24); and group III, represented by PSβG clone E (23). FL-NCA belongs to group I. Its sequence is identical with clones C and D up to the C-terminal tail and differs from clones 16 and 93 only at positions 7, 9, and 285. However, its tail is different from those of all other PSβGs. Moreover, the nucleotide sequence of FL-NCA differs completely from other PSβGs from position 1303, where the C-terminal tail begins. (*ii*) FL-NCA, PSβG, CEA, NCA-55, and BGPI show a high degree of sequence identity. Forty-five percent of the amino acids in the N domain occur at identical positions in all six proteins. The corresponding figure for the A domains is 18%. However, when only the A1 domains are compared, the figure increases to 44%. In the B

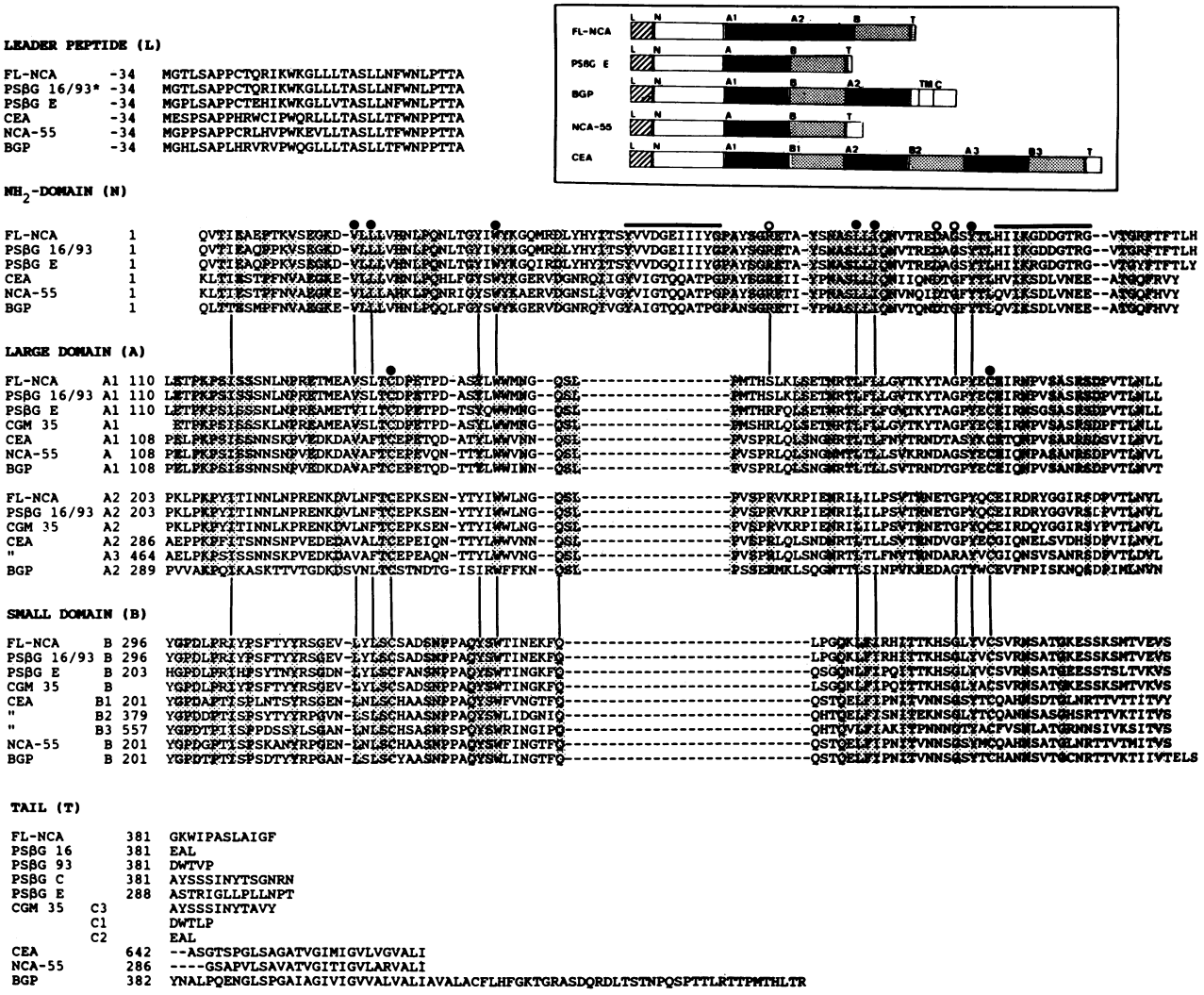

FIG. 3. Amino acid sequence (standard one-letter symbols) comparison of FL-NCA, PSβG, CEA, NCA-55, BGPI; and CGM 35 (PSβG/FL-NCA genomic clone). Identical residues at a given position in the proteins are shaded. Filled circles denote amino acids that are invariant in the immunoglobulin gene family. Open circles denote amino acids characteristic of variable-region (V-set) sequences. Positions in the sequence corresponding to complementarity-determining regions 2 and 3 of immunoglobulins are overlined. (*Inset*) Domain structures. TM, transmembrane; C, cytoplasmic.

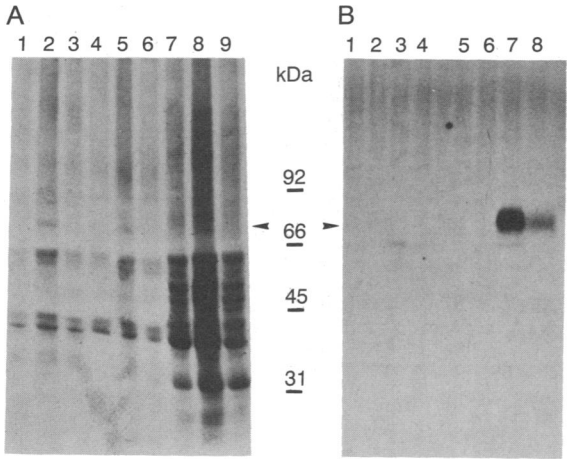

FIG. 4. SDS/PAGE–autoradiographic analysis of immunoprecipitates. (*A*) Solubilized CHO cells after labeling for 2 hr with [^{35}S]methionine. Lanes: 1, 4, and 7, untransfected cells; 2, 5, and 8, cells transfected with pECE with FL-NCA cDNA in correct orientation; 3, 6, and 9, cells transfected with pECE with cDNA in reversed orientation. Antibodies used for immunoprecipitation were anti-PSβG (lanes 1–3), anti-BGPI (lanes 4–6), and anti-CEA (lanes 7–9). (*B*) Solubilized HeLa cells (lanes 1–4) and spent medium (lanes 5–8), immunoprecipitated with anti-PSβG. The cells were labeled with [^{3}H]mannose for 12 hr. Lanes: 1 and 5, untransfected cells; 2 and 6, cells transfected with vector only; 3 and 7, cells transfected with pECE with cDNA; 4 and 8, cells transfected with pCMU.5 with cDNA. Positions and sizes of standard proteins run in parallel are indicated. Arrowheads point to 72-kDa glycoprotein.

domain 38% of the amino acids occur at identical positions. (*iii*) Each domain in all six proteins contains most of the invariant amino acids (filled circles in Fig. 3) conserved across the immunoglobulin gene superfamily (25, 26). (*iv*) The N domain is related to the variable region of immunoglobulin and T-cell antigen receptor (V-set sequences), and the A and B domains are related to the constant region (C2-set sequences). (*v*) Characteristic features of the subfamily seem to be that all members contain a V-set-related N-terminal domain lacking cysteines and one or several C2-set-related A and B domains (Fig. 3).

Expression of FL-NCA. HeLa and CHO cells were transfected with pCMU5 or pECE vector alone or with the vector containing FL-NCA cDNA in the correct or in the reversed orientation, respectively, and grown in the presence of [^{35}S]methionine or [^{3}H]mannose. The cell lysate and spent medium were then immunoprecipitated with anti-CEA, anti-BGPI, and anti-PSβG and analyzed by SDS/PAGE and autoradiography. Fig. 4 shows a representative experiment. A 72-kDa glycoprotein was specifically precipitated by all three antisera in cells transfected by FL-NCA cDNA in the right orientation. After a short labeling period the glycoprotein was found in the cell lysate, whereas after a longer labeling period the glycoprotein appeared in the medium.

Number of FL-NCA/PSβG Genes and Expression of mRNAs. Two probes corresponding to part of the N domain (*Rsa* I–*Sma* I fragment, nucleotides 261–400) and the N and A1 domains (*Sma* I–*Pst* I, nucleotides 400–691) of FL-NCA were used in Southern blot analysis of human genomic DNA. Under stringent conditions the N+A1 probe hybridized with 8–10 *Eco*RI, *Bam*HI, or *Hin*dIII fragments. In contrast, when the N-domain probe was used only one *Hin*dIII fragment (2.8 kb) was found to hybridize (data not shown).

Poly(A)$^+$ RNA from adult liver and placenta was prepared and analyzed by Northern blotting using probes corresponding to the coding region (*Sma* I–*Pst* I fragment) and 3′ noncoding region (*Bam*HI–*Eco*RI fragment) of FL-NCA cDNA. A typical experiment is shown in Fig. 5. Placenta

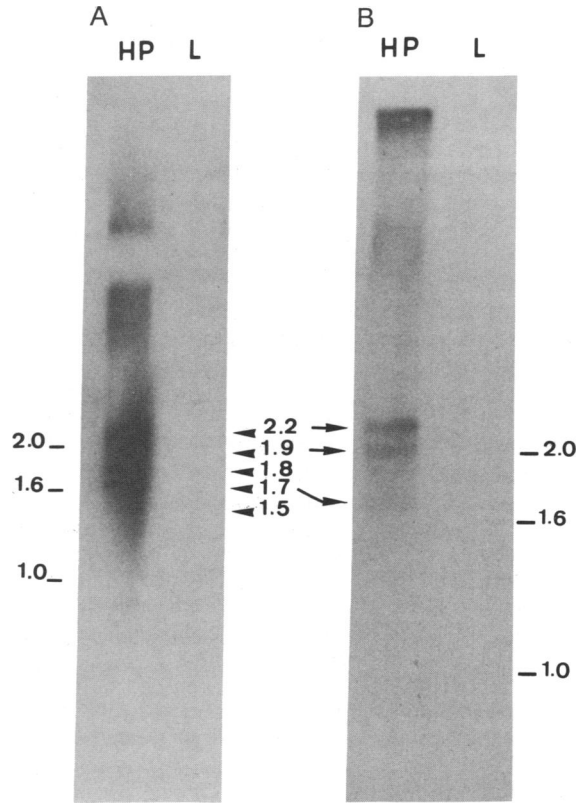

FIG. 5. Northern blot analysis of poly(A)$^+$ RNA from full-term human placenta (HP) and human adult liver (L). Poly(A)$^+$ RNA (2 μg per lane) was electrophoresed in formaldehyde/agarose gels. The blots were hybridized to radiolabeled *Sma* I–*Pst* I fragment (*A*) or to *Bam*HI–*Eco*RI fragment (3′ noncoding region) (*B*). Sizes are indicated in kilobases.

contained five species of mRNA that hybridized with the coding-region probe, ranging in size from 1.5 to 2.2 kb. When the 3′ probe was used, three mRNA species, of 1.7, 1.9, and 2.2 kb, were seen. The 1.7-kb mRNA corresponds to FL-NCA. No mRNA that hybridized with these probes was detected in adult liver.

DISCUSSION

We have identified FL-NCA as a member of the CEA family. It belongs to the PSβG subfamily, which appears to comprise at least three different groups of proteins (I–III) encoded by closely related genes. FL-NCA has a unique 12-amino acid hydrophobic tail not found in other members of the subfamily. It is likely that FL-NCA and PSβG clones C and D (23) are encoded by the same gene and that they are generated by alternative splicing of the primary transcript. Most likely PSβG clones 16 and 93 (21) also are encoded by this gene. Genetic polymorphism may be the explanation for the six nucleotide differences (three of which lead to amino acid changes) in the sequence 5′ to the tail that distinguish FL-NCA and PSβG clones C and D, on the one hand, from PSβG clones 16 and 93 on the other.

PSβG groups I and II have the domain structure N–A1–A2–B2–T. PSβG group III has the structure N–A1–B2–T. They show 80–92% sequence identity when the corresponding domains are compared. At present, four forms of PSβG I that differ in the C-terminal tail are known. Alternative splicing can give rise to three different forms of PSβG II (24) with tails identical with or similar to three of the tails found in PSβG I: However, a tail corresponding to that of FL-NCA was not found in the sequenced portion of the PSβG II gene

(24). If present, it must be 3′ to the other C-terminal sequences. No cDNA clones corresponding to PSβG II has been sequenced. Only one form of PSβG III is known. Its C-terminal tail differs from all others mentioned above (23).

In the fetal liver cDNA library (first trimester) the 1695-bp FL-NCA cDNA was most abundant, accounting for at least half of the CEA-crosshybridizing cDNA clones. Moreover, this clone represented 0.2% of the total clones screened. A 1.7-kb mRNA species hybridizing with a FL-NCA 3′-end probe was seen in human term placenta but not in adult human liver. Thus FL-NCA appears to be produced by both the fetal liver and the syncytiotrophoblast. Recently, we also detected a clone coding for FL-NCA in a cDNA library from apparently normal adult colon (unpublished results), but there the clone was very rare. A 3.4-kb mRNA species from submandibular salivary gland was found to hybridize with both coding and noncoding sequences of FL-NCA (W.N.K. and S.H., unpublished results).

The 72-kDa FL-NCA corresponds exactly in size to the major human placental PSβG species (21). In addition, human placenta was found to contain a 54-kDa and a 64-kDa PSβG species (21). FL-NCA was immunoprecipitated by polyclonal anti-CEA and anti-BGPI sera as well as by an antiserum to PSβG. The finding that the former sera precipitated FL-NCA is consistent with the high degree of sequence similarity between FL-NCA/PSβG, on the one hand, and CEA, NCA-55, and BGPI on the other. The results demonstrate that the two groups of proteins crossreact immunologically and that this crossreactivity must be considered in the design of specific CEA assays for clinical use and in the selection of monoclonal anti-CEA antibodies for radioimmunodetection and radioimmunotherapy.

A comparison of the amino acid sequences of the FL-NCA/PSβG group with those of the CEA/NCA-55/BGPI group of proteins reveals several interesting similarities and differences. (*i*) There are two regions in the N domain, positions 46–57 and 95–100, where the sequences are markedly different between the two groups while they are conserved within the groups (Fig. 3). With the possible exception of the sequence 174–179 in FL-NCA such differences are not found when other domains are compared. (*ii*) The subgroup-specific regions have the same positions in the sequence as complementarity-determining regions 2 and 3 in immunoglobulin variable-region sequence (25). (*iii*) Both regions end with an invariant glycine, a residue that may provide flexibility and permit movement of the rest of the sequence to allow better contact with a putative ligand (25). (*iv*) The region corresponding to complementarity-determining region 2 in the CEA group shows significant sequence homology with major histocompatibility complex-encoded class I and II antigens [corresponding to a sequence of β-strand in the bottom of the binding pocket of class I antigen (27)]. (*v*) The members in the family appear to form dimers. BGPI even contains an extra cysteine in the third domain (10), which may participate in the formation of an interchain disulfide bond. It is tempting to speculate that the proteins contain a ligand binding site made up of two N domains or by an N + A1 domain in which the regions mentioned above form part of the binding pocket. FL-NCA would then bind to one type of ligand, whereas CEA, NCA-55, and BGPI would bind to another. We do not know the nature of this hypothetical ligand. It should perhaps be searched for in the external environment, since at least one form of each protein is secreted into or facing a body lumen.

The skillful technical assistance of Ms. Anne Israelsson is gratefully acknowledged. This work was supported by a grant from the Swedish Cancer Society.

1. Gold, P. & Freedman, S. O. (1965) *J. Exp. Med.* **122,** 467–481.
2. Fritsche, R. & Mach, J.-P. (1977) *Immunochemistry* **14,** 119–127.
3. Oikawa, S., Nakazato, H. & Kosaki, G. (1987) *Biochem. Biophys. Res. Commun.* **142,** 511–518.
4. Buchegger, F., Vacca, A., Carrel, S., Schreyer, M. & Mach, J.-P. (1988) *Int. J. Cancer* **41,** 127–134.
5. Mach, J.-P. & Pusztaszeri, G. (1972) *Immunochemistry* **9,** 1031–1033.
6. Svenberg, T., Hammarström, S. & Hedin, A. (1979) *Mol. Immunol.* **16,** 245–252.
7. Paxton, R., Mooser, G., Pande, H., Lee, T. & Shively, J. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 920–924.
8. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A. & Stanner, S. (1987) *Mol. Cell. Biol.* **7,** 3221–3230.
9. Neumaier, M., Zimmerman, W., Shively, L., Hinoda, Y., Riggs, A. & Shively, J. (1988) *J. Biol. Chem.* **263,** 3202–3207.
10. Hinoda, Y., Neumaier, M., Hefta, S., Drzeniek, Z., Wagener, C., Shively, L., Hefta, L., Shively, J. & Paxton, R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6959–6963.
11. Thompson, J. & Zimmerman, W. (1988) *Tumor Biol.* **9,** 63–83.
12. Hammarström, S., Shively, J., Paxton, R., Beatty, B., Larsson, Å., Ghosh, R., Börmer, O., Buchegger, F., Mach, J.-P., Burtin, P., Seguin, P., Darbouret, B., Degorce, F., Sertour, J., Jolu, J.-P., Fuks, A., Kalthoff, H., Schmiegel, W., Arndt, R., Klöppel, G., von Kleist, S., Grunert, F., Schwartz, K., Matsuoka, Y., Kuroki, M., Wagener, C., Weber, T., Yachi, A., Imai, K., Hishikawa, N. & Tsujisaki, M. (1989) *Cancer Res*, in press.
13. Zimmerman, W., Weber, B., Ortlieb, B., Rudert, F., Schempp, W., Fiebig, H.-H., Shively, J., von Kleist, S. & Thompson, J. (1988) *Cancer Res.* **48,** 2550–2554.
14. Matsuoka, V., Kuroki, M., Koga, V., Kuriyama, H., Mori, T. & Kosaki, G. (1982) *Cancer Res.* **42,** 2012–2018.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
16. Tabor, S. & Richardsson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.
17. Ellis, L., Clauser, E., Morgan, D. O., Ederly, M., Roth, R. A., & Rutter, W. J. (1986) *Cell* **45,** 721–732.
18. Pääbo, S., Weber, F., Kämpe, O., Schaffner, S. & Petterson, P. A. (1983) *Cell* **35,** 445–453.
19. Chen, C. & Okayama, H. (1987) *Mol. Cell. Biol.* **7,** 2745–2752.
20. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91,** 524–545.
21. Watanabe, S. & Chou, J. (1988) *J. Biol. Chem.* **263,** 2049–2054.
22. Watanabe, S. & Chou, J. (1988) *Biochem. Biophys. Res. Commun.* **152,** 762–768.
23. Streydio, C., Lacka, K., Swillens, S. & Vassart, G. (1988) *Biochem. Biophys. Res. Commun.* **154,** 130–137.
24. Oikawa, S., Inuzuka, C., Kosaki, G. & Nakazato, H. (1988) *Biochem. Biophys. Res. Commun.* **156,** 68–77.
25. Kabat, E. A., Wu, T., Reid-Miller, M., Perry, H. & Gottesman, K. (1987) *Sequences of Proteins of Immunological Interest* (U.S. Dept. of Health and Human Services, Washington, DC), 4th Ed.
26. Williams, A. & Barclay, N. (1988) *Annu. Rev. Immunol.* **6,** 381–405.
27. Björkman, P., Saper, M., Samraoui, B., Bennett, W., Strominger, J. & Wiley, D. (1987) *Nature (London)* **329,** 506–511.