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 (PS β G) but differs from other PS β G 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 PS β G. 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 radio-immunotherapy (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 $[\gamma^{-32}P]ATP$ and used to screen a human fetal liver cDNA library in phage vector λ gt11. 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 *Hind*III 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-PS β G 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; PS β G, 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).

Medical Sciences: Khan et al.

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, $5 \times SSPE (1 \times SSPE =$ 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 $2 \times SSPE/1\%$ SDS at 65°C for 1 hr and in $0.1 \times$ SSPE/1% SDS at 68°C for 30 min. DNA fragments containing nucleotides 400-692 (Sma I-Pst I) and 1310-1694 (BamHI-*Eco*RI) of the cDNA sequence were labeled with $\left[\alpha^{-32}P\right]dCTP$ 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 Nterminal 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 CEArelated molecules. One cDNA species of 1.7 kb was by far the



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.

most frequent and was chosen for complete sequence determination. The glycoprotein it encodes will be referred to as

AGAGAGTTTCTGGATCCTAGGCTTATCTCCACAGAGGAGAACACACAAGCAGCAGAGACC

-34	Met ATG	Gly GGA	Thr ACC	Leu CTC	Ser TCA	Ala GCC	Pro CCT	Pro CCC	Cys TGC	Thr ACA	Gln CAG	Arg CGC	Ile ATC	Lys AAA	Trp TGG	Lys AAG	Gly GGG	Leu CTC	Leu CTG	Leu CTC	Thr ACA	Ala GCA	Ser TCA	Leu CTT	132
-10	Leu TTA	Asn AAC	Phe TTC	Trp TGG	Asn AAC	Leu CTG	Pro CCC	Thr ACC	Thr ACT	Ala GCC	GÎn CAA	Val GTC	Thr ACG	Ile ATT	Glu GAA	Ala GCC	Glu GAG	Pro CCA	Thr ACC	Lys AAA	Val GTT	Ser TCC	Glu GAG	Gly GGG	204
15	Lys AAG	Asp Gat	Val GTT	Leu CTT	Leu CTA	Leu CTT	Val GTC	His CAC	Asn AAT	Leu TTG	Pro CCC	Gln CAG	Asn AAT	Leu CTT	Thr ACC	Gly GGC	Tyr TAC	İle ATC	Trp TGG	Tyr TAC	Lys AAA	Gly GGG	Gln CAA	Met ATG	276
39	λrg AGG	Asp GAC	Leu CTC	Tyr TAC	His Cat	Tyr Tac	Ile ATT	Thr ACA	Ser TCA	Tyr TAT	Val GTA	Val GTA	Asp GAC	Gly GGT	Glu GAA	Ile ATA	Ile ATT	Ile ATA	Tyr TAT	Gly GGG	Pro CCT	Ala GCA	Tyr Tat	Ser AGT	348
63	Gly GGA	Arg CGA	Glu GAA	Thr ACA	Ala GCA	Tyr TAT	Ser TCC	Asn AAT	Ala GCA	Ser TCC	Leu CTG	Leu CTG	Ile ATC	Gln CAG	Asn <u>AAT</u>	Val GTC	Thr ACC	Arg CGG	Glu GAG	λsp GAC	Ala GCA	Gly GGA	Ser TCC	Tyr TAC	420
87	Thr ACC	Leu TTA	His CAC	Ile ATC	Ile ATA	Lys AAG	Gly GGA	Asp GAT	Asp GAT	Gly GGG	Thr ACT	λrg Aga	Gly GGA	Val GTA	Ťhr ACT	Gly GGA	Arg CGT	Phe TTC	Thr ACC	Phe TTC	Thr ACC	Leu TTA	His CAC	Leu CTG	492
111	Glu G A G	Thr ACT	Pro CCT	Lys AAG	Pro CCC	Ser TCC	Ile ATC	Ser TCC	Ser AGC	Ser AGC	Asn AAC	Leu TTA	Asn AAT	Pro CCC	Arg AGG	Glu GAG	Thr ACC	Met ATG	Glu GAG	Ala GCT	Val GTG	Ser AGC	Leu TTA	Thr ACC	564
135	Cys TGT	Asp GAC	Pro CCT	Glu GAG	Thr ACT	Pro CCA	Asp GAC	Ala GCA	Ser AGC	Tyr TAC	Leu CTG	Trp TGG	Trp TGG	Met ATG	λsn λλΤ	Gly GGT	Gln CAG	Ser AGC	Leu CTC	Pro CCT	Met ATG	Thr ACT	His CAC	Ser AGC	636
159	Leu TTG	Lys AAG	Leu CTG	Ser TCC	Glu GAA	Thr ACC	Asn AAC	Arg AGG	Thr ACC	Leu CTC	Phe TTT	Leu CTA	Leu TTG	Gly GGT	Val GTC	Thr ACA	Lys AAG	Tyr TAT	Thr ACT	Ala GCA	Gly GGA	Pro	Tyr TAT	Glu G AA	708
183	Cys TGT	Glu GAA	Ile ATA	Arg CGG	Asn AAC	Pro CCA	Val GTG	Ser AGT	Ala GCC	Ser AGC	Arg CGC	Ser AGT	Asp GAC	Pro CCA	Val GTC	Thr ACC	Leu CTG	Asn AAT	Leu CTC	Leu CTC	Pro CCG	Lys AAG	Leu CTG	Pro CCC	780
207	Lys AAG	Pro CCC	Tyr TAC	Ile ATC	Thr ACC	Ile ATC	Asn AAC	Asn AAC	Leu TTA	Asn AAC	Pro CCC	Arg AGG	Glu GAG	Asn AAT	Lys AAG	Asp Gat	Val GTC	Leu TTA	Asn AAC	Phe TTC	Thr ACC	Cys TGT	Glu G AA	Pro CCT	852
231	Lys AAG	Ser AGT	Glu GAG	Asn AAC	Tyr TAC	Thr ACC	Tyr TAC	Ile ATT	Trp TGG	Trp TGG	Leu CTA	Asn AAT	Gly GGT	Gln CAG	Ser AGC	Leu CTC	Pro CCG	Val GTC	Ser AGT	Pro CCC	Arg AGG	Val GTA	Lys Aag	Arb Cga	924
255	Pro CCC	Ile ATT	Glu GAA	Asn AAC	Arg AGG	Ile ATC	Leu CTC	Ile ATT	Leu CTA	Pro CCC	Ser AGT	Val GTC	Thr ACG	λrg λGλ	λsn <u>λλΤ</u>	Glu GAA	Thr ACA	Gly GGA	Pro CCC	Tyr TAT	Gln CAA	Cys TGT	Glu GAA	Ile ATA	996
279	Arg CGG	Asp GAC	Arg CGA	Tyr TAT	Gly GGT	Gly GGC	Ile ATC	Arg CGC	Ser AGT	Asp GAC	Pro CCA	Val GTC	Thr ACC	Leu CTG	λsn λλT	Val GTC	Leu CTC	Tyr Tat	Gly GGT	Pro CCA	λsp GλC	Leu CTC	Pro CCC	Arg Aga	1068
303	Ile ATT	Tyr TAC	Pro CCT	Ser TCA	Phe TTC	Thr ACC	Tyr TAT	Tyr TAC	Arg CGT	Ser TCA	Gly GGA	Glu G AA	Val GTC	Leu CTC	Tyr TAC	Leu TTG	Ser TCC	Cys TGT	Ser TCT	Ala GCG	λsp GAC	Ser TCT	Asn AAC	Pro CCA	1140
327	Pro CCG	Ala GCA	Gln C A G	Tyr TAT	Ser TCT	Trp TGG	Thr ACA	Ile ATT	Asn AAT	Glu GAA	Lys Aag	Phe TTT	Gln CAG	Leu CTA	Pro CCA	Gly GGA	Gln C AA	Lys Aag	Leu CTC	Phe TTT	Ile ATC	Arg CGC	His Cat	Ile ATT	1212
351	Thr ACT	Ťhr ACA	Lys AAG	His Cat	Ser AGC	Gly GGG	Leu CTC	Tyr TAT	Val GTT	Cys TGC	Ser TCT	Val GTT	Arg CGT	λsn λλC	Ser TCA	Ala GCC	Thr ACT	Gly GGC	Lys AAG	Glu GAA	Ser AGC	Ser TCC	Lys AAA	Ser TCC	1284
375	Met ATG	Thr ACA	Val GTC	Glu G AA	Val GTC	Ser TCT	Gly GGT	Lys AAG	Trp TGG	Ile ATC	Pro CCA	Ala GCA	Ser TCG	Leu TTG	Ala GCA	Ile ATA	Gly GGG	Phe TTT	End TAGG	TGGA	GTCT	атст	GGCA	TTC	1361
	AGAG	AAGA	GTCA	.GGAA	ааса	ATTG	TATT	CCCA	GCCT	GTGT	CCCA	TGGG	CACA	AGCA	латс	CCAA	ATTC	тсст	CCTG	AACC	стсс	AAAT	TTGT	CTA	1456
	AGAA	сттс	GAAA	ACTT	таас	AAAC	AGGC	TGAT	атст	TCAT	АЛТА	TTCC	CAGC	CTAG	лссл	AGCA	GGAA	GAAC	ATTG.	ATTT	CATT	GAAA	TAAT	TGA	1551
	таат	aatg.	AAGA	таат	GTTT	TTAT	GATT	TTTA	TTTG	****	TTTG	CTGA	ттст	ттаа	ATGG	TTTG	TTTT	CTAC	ATTG.	ATGG.	AATT	TTTC	тстт	тта	1646
	ATCT	ATCT.	ACAG	CTTA	TAGC	AGTT	с <u>аат</u>	<u>AAA</u> C	ТАТА	CTTC	TGGG.	****	X												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. 1G). 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. 1E). 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. 1F). 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 (PS β G), CEA, NCA-55, BGPI, and a PS β G/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 PS β G. Three closely related PS β Gs can be distinguished: group I, represented by PSBG clones 16 and 93 (21, 22) and PS β G clones C and D (23); group II, represented by the genomic clone CGM 35 (24); and group III, represented by PS β G 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 $PS\beta Gs.$ Moreover, the nucleotide sequence of FL-NCA differs completely from other PS β Gs from position 1303, where the C-terminal tail begins. (ii) FL-NCA, PS β G, 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

LANSE PERTUR (J) FL-CK,				FL-NCA	
PrK.A. -14 MCLEARPCTORINGULLINGUETALINNINGETA PEGO 14/071-2 MCLEARPCTORINGUELLINNINGETA	LEADER PEPTIDE	(L)			
<pre>pig0 fig0 fig0 results of the second se</pre>	FL-NCA -3		CTT SADDCTODI KUKGI I I TASI I NEWNI DTTA	PS8G E	
PS00 E -14 M055APETHEITKKGLUTYASLLTMPNFTAN MCA-30	PSBG 16/93* -3	4 1	IGTLSAPPCTORIKWKGLLLTASLLNFWNLPTTA		
CRA 55	PSBG E -3	4 P	IGPLSAPPCTEHIKWKGLLVTASLLNFWNLPTTA	BUP	
DGP 37 -14 HIGHEADLERVERVENGOLLERVERVENGOLLERVERVENGEN HIGHEADLERVERVENGOLLERVERVENGOLLERVERVENGEN D D HIGHEADLERVERVENGOLLERVERVENGEN D D HIGHEADLERVERVENGEND VLLAUBLERVERVENGEND VLLAUBLERVERVENGEND HIGHEADLERVERVENGEND VLLAUBLERVERVENGEND VLLAUBLERVERVENGEND HIGHEADLERVERVENGEND VLLAUBLERVERVENGEND VLLAUBLERVERVENGEND PERCEA OVELBALBERVENGEND VLLAUBLERVERVENGEND VLLAUBLERVERVENGEND PERCEA OVELBALBERVENGEND VLLAUBLERVENTSTERVENGENDERVENGENGENGENGEN VLLAUBLERVENGENGENGENGEN VLLAUBLERVENGENGENGENGENGENGENGENGENGENGENGENGENGE	CEA -3-	4 P	(ESPSAPPHRWCIPWORLLLTASLLTFWNPPTTA	NCA-55	
CA PINCA 0VTI BARFINYSERED-91_UNVERNITET ITEGORIED. YNVERNITET ITEGORIED. YNVERNITET YNODEI II YNRYSWETA - SERBEL ONTE AND STRUTT I MANDER YN YNTE YNDDEI II YNRYSWETA - SERBEL ONTE AND STRUTT I MANDER YN YNTE YNDDEI II YNRYSWETA - SERBEL ONTE AND STRUTT I MANDER YN YNTE YNDDEI II YNRYSWETA - SERBEL ONTE AND STRUTT I MANDER YN YNTE YNDDEI II YNRYSWETA - SERBEL ONTE AND STRUTT I MANDER YN YNTE YNDDEI II YNRYSWETA - SERBEL ONTE AND STRUTT I MANDER YN YNTE YNDDEI II YNRYSWETA - SERBEL ONTE AND STRUTT I MANDER YN YNTE YNDDEI II YNRYSWETA - SERBEL ONTE AND STRUTT - YNRYFTI I MANDER YN YNTE YNDDEI II YNRYSWETA - SERBEL ONTE YN YNDE YN YNTE YN YNDE YN YN YNTE YN YNDE I I YNRYSWETA - SERBEL ONTE YN YNDE YN YNTE YN YNDE YN YN YN YN YN YNTE YN YNDE YN	BGP -3	4 1	AGPPSAPPCREAVPWREVELIASEEITWAPPITA AGHLSAPLHRVRVPWOGLLLTASLLTFWNPPTTA		
HECOMULE (H) FL-NCA PSIO 1 OVTIEAGEPTINGSERUD-91LUVENLEGNLTETITEGGMERUTHYTTSTYVDOETII YEBAYSERTA-SSEREFLGYVTTEGAGSTELII ISKODOTGOVTURFTTILE PSIO 5 OVTIEAGEPTINGSERUD-91LUVENLEGNLTETITEGGMERUTHYTTSTYVDOETII YEBAYSERTA-SSEREJGYVTTEGAGSTELII ISKODOTGOVTURFTTILE PSIO 5 OVTIEAGEPTINGSERUD-91LUVENLEGNLTETITEGGMERUTHYTTSTYVDOETII YEBAYSERTA-SSEREJGYVTTEGASSERUTYTEGASSERUT		• •	••••••	CEA	
Image 2004/16 (17) PF-NCA 1 OVEL BARFYRWSERED - VLLWENLEQUITGET GEGORGELYHYTEST VDGET IL YRAYSGETA - YSHREL GENYTERASSET, HIT LAGDOTTE VTREFT TILL PEGA 16 / 3 1 OVEL BARFYRWSERED - VLLWENLEQUITGET GEGORGELYHYTEST VDGET IL YRAYSGETA - YSHREL GENYTERASSET, HIT LAGDOTTE VTREFT TILL PEGA 16 / 3 1 OVEL BARFYRWSERED - VLLWENLEQUITGET GEGORGELYHYTEST VDGET IL YRAYSGETA - YSHREL GENYTERASSET, HIT LAGDOTTE VTREFT TILL CEA 1 OVEL BARFYRWSERED - VLLWENLEQUITGET GEGORGELYHYTEST VDGET IL YRAYSGETA - YSHREL GENYTERASSET, HIT LAGDOTTE VTREFT FLWENCENCON CONTROL AND TARKEN AND THE PERSISSEN AND THE P					
FL-NCA PBC 16/93 OVIE BARFTWSERED-VLLWBILKONLTGET FREGORDEL YNYTESTWJOET II YN PAYSMERTA - YLBASLE GWYNEBASTE HI HAGDOTHE VTHWTFT LM BYNG 1 PSC 1 OVIE BARFTWSERED-VLLWBILKONLTGET FREGORDEL YNYTESTWJOET II YN PAYSMERTA - YLBASLE GWYNEBASTE HI KGODTHE VTHWTFT LM BYNG 1 NCA-55 I WYTE BARFTWSERED-VLLWBILKONLTGET FREGORDEL YNYTESTWJOET II YN PAYSMERTA - YLBASLE GWYNDITWEBASTE HI KGODTHE VTHWTFT LM BYNG 1 NCA-55 I KLEITSPERYKNEERE-VLLWRLKONT BYNG FREGORDEL YNYTESTWJOET II YN PAYSMERT - FRANKLE GWYNDITWEBASTE HI FRODOTHE VTHWTTT KLEITSPERYMAREE-VLLWRLKONT BYNG FREGORDEN HINT FREGORDEN - YN PASSEL GWYNDITWEBASTE I FRANKLE GWYNDITWEFT FRANKLE GWYNDITWEFT FRANKLE GWYNDITWEFT FRANKLE GWYNDITWEFT FRANKLE GWYNDITWEFT FRANKLE FRANKLE - ANDOPNY LAREE DORAIN (A) PEFL-NCA A1 110 LETERFERSISSINLBRETHEAVSL KUPPETP- ASLWMMO - GE WYTHOL KUPPETFERSISSINLBRETHEAVSL KUPPETFD- ASLWMMO - GE PERLENE A1 100 LETERFERSISSINLBRETHEAVSL KUPPETP- ASLWMMO - GE WYTHOL KUPPETFERSISSINLBREW KLEWPETFUL FRANKLEW FRANKLEWERTHEAVSL KUPPETFD- ASLWMMO - GE WYTHOL KUPPETFERSISSINLBREW KLEWPETFUL FRANKLEWERTHEAVSL KUPPETFD- ASLWMMO - GE WYTHOL KUPPETFERSISSINLBREW KLEWPETFUL FRANKLEWERTHEAVSL KUPPETFUL FRANKLEWERTHEAVSL KUPPETFUL FRANKLEWERTHEAVSL KUPPETFD- ASLWMMO - GE WYTHOL KUPPETFERSISSINLBREW KLEWPETFUL FRANKLEWERTHEAVSL KUPPETFUL FRANKLEWERTHEAVSL KUPPETFUL FRANKLEWERT	RH2-DOMAIN (N)				
FL-RCA 1 OVIERANT RUSSING-VALUNG UNDER CONDUCTIVE TRY VOOREIT UNDER VERSUURS AUTOR TRANSMERT - SEASEL SEGNT VERSUURS AUTOR TRY VERSUURS AUTOR TRANSMERT - SEASEL SEGNT VERSUURS AUTOR TRY VERSUURS AUTOR AU					
PSBE 10793 1 0071 BADEWASERLD FALL WELEVELT HER LEVEL THE LEVEL WELEVELY WORK OF THE STORM AND STATE IN RELAX DEFINITION OF THE STORM AND STATE IN RELAX D	FL-NCA	1	QVTI KAEPTKVSEGKD-VLLLVHNLPQNLTGY I WYKGQMR	DLYHYITS	YVVDGEIIIYOPAYSORETA-YSHASILIQOVTREDAGSITLHIIRGDDGTRGVTORFTFTLH
CEA 1 I.T.T.B.STEPHYNABEL-VILLANGELE-BYSKYRGENVENCEJ GWYNORDJONG WARDELL-FYNABELDUNIELNAWERWYN NCA-55 NCA-55 1 NLTTESTEPHYNABEL-VILLANELEGOL FORSKYRGENVONNOUT GYDATTERWYNBURGEL - FYNABELDUNIELANDERWYN BCP LARGE DORLIN (A)	PSBG 16/93 PSBG F	1	OVET EAGEPKYSKGKD-VLLLVHILEPULTGY IN XAGONR OVET EAGEPKYSKGKD-VLLLVHILEPULTGY IN XAGONR	DLYHYITS M.VHVTTS	YVVDGEIIIYAPAYSARTA-YSNASILLONVTREDAGSYTLHIIKGDDGTRGVZSRTTTLH WVVDGOIIIYAHAYSAPETA-YSNASILLONVTREDAGSYTLHIYEBGDGTBGVZSRTTTLH
NCA-55 1 CLITESTETYNARGE-VILLARKLEGARIGSSATALEXGARGUNGNSLÄVGY LOCGATEGA KERKET - GENERALGANGUTGETTE OVIKSDUWEE - ARGOPTIV LARCE DONAIN (A) FL-NCA AI 110 LETTREPEISSENLAPRETHEAVSLUTDETTP - ASTLANKULGSVELAEKGEV ON SLÄVET - GENERALGANG KERT -	CEA	î	KLTI STPFNVAEGRE-VILLVNNLPOHLFGYSWYRGERV	DGNROLIG	YVIGTOOATPGFAYEGREII-YPHANLLICHIIONDTGFYTLHVIKSDLVNEEATGGFRVY
BGP 1 OLDTESHEPTMARGED-VALUMELEQCLEGENVERGENVERANGTVORALGTOCALTOGOLAT	NCA-55	1	KLTIESTFFNVAEGRE-VLLLARKLECNRIGYSVYRAERV	DGNSLIVG	WIGTQQATPOPAYSGRETI-YPRASLLIGHVNQIDTGFYTLQVIRSDLVNEEATGQFHVY
LARGE DOMAIN (A) FL-NCA A1 110 LITTREESISSINLAPRETMEANS LECOPETIDO-ASTLMMINGO-GSL	BGP	1	QLITTESMPFNVABGKE-VLLLVINEPQQLFGYSWIRGERV	DGNRQIVG	YAIGTQQATPGPANSGRETI-YPMASLLIQMVTQNDTGF7TLQVIESDLVNEEASGQPHVY
LARGE DOMAIN (A) FL-NCA A1 110 LETTEREST SSSNLAPRETMEANS IS CODPETED - AST LMMMG - GSL					
FL-NCA A1 110 LETPERFERSIONLARGETHELYSLEDPETPD-ASTLWANGG-OSL PMTHSLKLEDTERTELGYTNYTAGTECTINEVASUEDPETPLATULL PSG E A1 110 LETPERFERSIONLARGETHELYSLEDPETPD-ASTLWANGG-OSL PMTHSLKLEDTERTELGYTNYTAGTECTINEVASUEDPETPLATULL PSG E A1 100 LETPERFERSIONLARGETHELYSLEDPETPD-ASTLWANGG-OSL PMTHSLKLEDTERTELGYTNYTAGTECTINEVASUESENTUTILL CGH 35 A1 TETERTESTISSENTUPREAMENTYL ICDPETPD-ASTLWANGG-OSL PMTHSLKLEDTERTELGYTNYTAGTECTINEVASUESENTUPLICUTILL CGH 35 A1 TETERTESTISSENTUPREAMENTYL ICDPETPD-ASTLWANGG-OSL PMSPLCLSNEETERGETTELGYTNYTAGTECTINEVASUESENTUPLICUTILL CGH 35 A1 TETERTESTISSENTUPLICUTILLENGTUPLICUTILLUSYTIMEGANGETTESTISSENTUPLICUTILLUSYTIMEGANGETTESTISSENTUPLICUTILLUSYTIMEGANGETTESTISSENTUPLICUTILLUSYTIMEGANGETTESTISSENTUPLICUTINUAL NEODENAL TESTISSENTUPLICUTILLUSYTIMEGANGETTESTISSENTUPLICUTINUAL NEODENAL TESTISSENTUPLICUTILLUSYTIMEGANGETTESTISSENTUPCOCTUPUCATION TESTISSENTUPLICUTINUAL NEODENAL TESTISSENTUPLICUTINU	LARGE DOMAIN (A)			
PL-NCA A1 110 LETERERSISSINUS RETERED AS LEMANDE - 93:					
PSBC 10/37 A1 110 LETROSSESSION CONTRACTOR AND	FL-NCA A1	110 1	LETPEPSISSISNLNPRETMEAVSLTCOPETPD-ASELWMMNG	OSE	PMTHSLKLSETNRTLFLLGYTKYTAGPYECHIRNPV\$ASBSDPVTLNLL
CGR 35 A1 108 PELEVERSISKLAD PERMEAVELISCOPETIPD-ASTLANDIGSIL	PSPG 10/93 AL	110 1	LETERPSISSISSICHERRETHERKSLILDERTPD-ASELWHMUG	051	
CEA A1 108 PELFEREISENNERVEDAAAFTEEPETOD-ATLIANVEN-GELPSEPELGENGEREITENT MUTASKEGTUGEVEALEBESVIEWU BGP A1 108 PELFEREISENNERVEDAAAFTEEPETOD-ATLIANVEN-GELPSEPELGENGEREITENT MUTASKEGTUGEVEALEBESVIEWU BGP A1 108 PELFEREISENNERVEDAAAFTEEPETOD-TTELMVIEN-GELPSEPELGENGEREITENT MUTASKEGTUGEVEALEBESVIEWU FL-NCA 2203 PELFEREISENNERVEDAAAFTEEPETOD-TTELMVIEN-GEL	CGM 35 A1		ETPEPSISSSKLNPREAMEAVSLTCDPETPD-ASYLWMMRG	051	PMSHRLOLSETWRTEFLLGYTKYTAGPYECEIR#PVSASESDPFTLELL
NCA-55 A 108 PELEKSSISSINSRYEDKOLANAPTCEERVON-TITLMINNG-QSL	CEA A1	108 1	PELPEPSIESNNSKPVEDKDAVAFTCEPETQD-ATTLWVNN	QSL	PVSPRLQLSNGBRTLTLFNYTRNDTASYKCTTORPVSARREDSVILBVL
PGP A1 100 PPL/PACEADSSINSRY/EDUDAVITACEPETUD-TICLW.INFQBL	NCA-55 A	108 1	PELFKFSIBSNNSNPVEDKDAVAFTCEPEVQN-TTTLWVNG	OSL	PVSPRLQLSNGMPTLTLLSVKRNDAGSTECHIGNPASANREDPVTLNVL
FL-RCA A2 203 PKLPERYSTINULNPERENKULNPTCEPKSEN-YTTUWLNG05L	BGP AI	108 1	PELYRPSIBSNNSNYVEDRDAVAFICEPETQD-TTELWWIWN		FVSPRLQESNG#RTETELSVTRNDIGPTECETQ#PVSAN#SDPVTG#VT
PSBG 16/93 A2 203 PKLPREYTINNLAPRENKDULNTGEPKSEN-TYTIMUNG-051	FL-NCA A2	203 1	PKLPKPYITINNLNPRENKOVLNFTCEPKSEN-YTYIWWLNG	051	PVSPRVKRPIENRILILPSVTNNETGPYQCEIRDRYGGIRSDPVTLNYL
CGM 35 A2 PRLPERPTITINULERPENKUVLAPTEEPKSEN-TTTIWULAG-OSS	PSBG 16/93 A2	203 1	PKLPKPYTTINNLNPRENKOVLNFTCEPKSEN-YTYIWWLNG	QSL	PVSPRVKRPIEMRILILPSVTRNETGPYQCEIRDRYGGVRSLPVTLNVL
CLA ALLEYESTSSINNSKPEDEDKONVAFTCEPERON-TYTLWYNG - SET	CGM 35 A2	286	PKLPKFYITINNLKPRENKOVLNFTCEPKSEN-YTYIWWLNG AFDDKNFITSNNSNDVFDFDAVAIWYFDFTON-TTVIWWLNG-	OSL	
BGP A2 289 PVVAKBOIKASKTTVTGDKÖSVNLTKSTNDTG-ISIRNFFKNÖSEPSSEINKLSQGNTTISINPVKREDAGTTVGEVFNPISKNOKDEINKNV SNALL DONAIN (B) SSEINKLSQGNTTISINPVKREDAGTTVGEVFNPISKNOKDE FL-NCA B 296 YGPDLFKITPSFTYTRSGEV-LYLSCSADENNPAQYSMTINEKKO	" A3	464	AELPKPSISSNNSKPVEDKDAVAFTCEPEAON-TTYLWVNG	OSL	
SHALL DOMAIN (B) LPGKLEX RHITTKHSGLVCSVRNSATKESSKSMTVEVS PSGG 16/93 B 296 YGEDLERI YESFTYRSGEV-EVISCSAD5MPPACKSWTINEKRO	BGP A2	289	PVVAKPOIKASKTTVTGDKDSVNLTCSTNDTG-ISIRWFFKN	Çsi	PSSERMKLSQGNTTESINFYKREDAGTYWCEVFNPISKNQ&DRIMENYN
SHALL DOMAIN (B)					
FL-NCA B 296 YGEDLEPETYDSFTYTRSGEV-LYLSCSADSNEPAGTSWTINEKKO LPGKLSTRHITTKHSGLYCSVRNSATKESSKSMTVEVS PSBG 16/93 B 201 YGEDLEPETYDSFTYTRSGEV-LYLSCSADSNEPAGTSWTINEKKO LPGKLSTRHITTKHSGLYCSVRNSATKESSKSMTVEVS PSBG 2 B 202 HGPDLEPETYDSFTYTRSGEV-LYLSCSADSNEPAGTSWTINEKKO LPGKLSTRHITTKHSGLYCSVRNSATKESSKSMTVEVS PSBG 2 B 201 YGEDLEPETYDSFTYTRSGEV-LYLSCSADSNEPAGTSWTINGKYO LSGKLSTPOTTKHSGLYCSVRNSATKESSKSMTVKVS CCM 35 B 201 YGEDLEPETYDSFTYTRSGEV-LYLSCSADSNEPAGTSWTINGKYO SGKLSTPOTTKHSGLYCSVRNSATKESSKSMTVKVS CEA B1 201 YGEDLEPETYDSFTYTRSGEV-LYLSCSADSNEPAGTSWTINGKYO SGKLSTPOTTKHSGLYCSVRNSATKESSKSMTVKVS CEA B1 201 YGEDLEPETYDSFTYTRSGEN-LASCHAASNPPAGTSWTINGKYO SGKLSTPOTTKHSGULYCSVRNSATKESSKSMTVKVS W B2 375 YGEDDEPETSSKYTTRGEN-LASCHAASNPPAGTSWTINGKYO SGKLSTPOTTKKSGUNGTKKESSKSMTVKVS W B3 557 YGEDDEPETSSKYTRYNSGEN-LASCHAASNPPAGTSWLINGTPO STELEXTPNKTVNNSGSTCANHSOTNULNTTVTNTTVKS WCA-55 B 201 YGEDDEPETSSKATKAASNPAGTSWEFFINGTPO STELEXTPNTTVNNSGSTCANHSOTNULNTTVKTHTVKTHTVKT FL-NCA 381 GKWIPASLAIGF SSGVTCANSNENTPAGTSSKENTPAGTSWEFFINGTPO STELEXTPNTTVNNSGSTCANHSOTNULNTTVKTHTVKTHTVKTHTVKTHTVKTHTVKTHTVKTHTVK	SMALL DOMAIN ((B)			
FL-NCA B 296 YGEPLERLYPSFTYTRSGEV-SYLGCSADSMPPAGYSMTINEKRO					
PSBG E B 201 IGPUERTINGSUP VIESTINGSUP VIESTINGKO Commentation of the second viesting vie	FL-NCA B 2	296	YGPDLPRIYPSFTYYRSGEV-LYLSCSADSNPPAQYSWTINEKF	ğ	L DODY I WE DU TOWN LOAD TWO COVENES ATTAKES SKSHTVEVS
CGM 35 B VGPDLPH3YPSPTYTRSGEV_SVISESADSNPPACPSMTINGKF0	PSBG E B 2	203 1	IGPOLPRIIFSFTIIRSGEV~LIGGESADSADPAGYSWTINEKF HGPOLPRTHPSYTNYRSGDN-LYGSEFANSKPPAGYSWTINGKF		OSCONET POITTKHSGLYVCSVRISATORESSTSLTVKVS
CEA B1 201 YGPDGPTISELNTSTRSGEN-ENLSCHAASNPPAGYSWIJDGNGTO	CGM 35 B		YOPDLPRIYPSFTYYRSGEV-LYLSCSADSNPPAGYSWTINGKF	ğ	LSGCKLFIPOITTKHSGLYACSVRNSATGKESSKSNTVKVS
B 379 YGPDDFTISRSYTYTRRVN-ISLBCHAASNPPAGYSMLINGTO B 357 YGPDDFTISRSYTYTRRVN-ISLBCHAASNPPAGYSMLINGTO NCA-55 B 201 YGPDCFTISRSKANTRFGEN-ENISCHASNPPAGYSMLINGTO BGP B 201 YGPDCFTISRSKANTRFGEN-ENISCHASNPPAGYSMLINGTO STORE IT PNITYNNSGSYTCHANNS	CEA B1 2	201	YGPDAPTI SPLNTSYRSGEN-LNLSCHAASNPPAQYSWFVNGTF	ÿ	OSTQELFIPNITVNNSQSTTCQAHUSDTQLNRTTVTTITVY
NCA-55 B 201 YGPDGPTISESKANTRPGEN-LALSCHAASNPPAGTSHFINGTPO	" B2 3	579	YGPDDPTISPSYTYYRPUVN-LSUSCHAASKPPAQYSWLIDGN1	<u>ğ</u>	
BGP B 201 YGPDTETESESDTYTRPGAN-ESESCYAASMPPAÖYSMLINGTPGGSTGELFIPNETVNNSGSTTCHAMMSVTCONRTTVKTIIVTELS TAIL (T) FL-NCA 381 GKWIPASLAIGF PSBG 61 381 DWTVP PSBG 73 381 DWTVP PSBG 7 381 ASTRIGLELINPT CGM 35 C3 ASTSSINTTAVY CEA 642 ASGTSPGLSAGATVGINIGVL/GVALI NCA-55 286	NCA-55 B 2	201	YOPDOPTISESKANYRPOEN-LNLSCHAASNPPACYSWFINGTF	ŏ	OSTOELFIPNITVNNSØSYMCOAHNSATGLNRTTVTHITVS
TAIL (T) FL-NCA 381 GKWIPASLAIGF PSSG 16 381 DMTVP PSSG 23 381 DMTVP PSSG C 381 ASTRIGLPLINPT CGM 35 C3 ASSSINTTAVY C1 DMTUP C2 EAL CEA 642 GS3FVLSAVATVGITIGVLSQUALI NCA-55 286 GSAFVLSAVATVGITIGVLARVALI	BGP B 2	201	YGPDTPTTSPSDTYTRPGAN-LSLSCYAASNPPAQYSWLINGTF	·õ	QSTQELFIPNITVNNSQSYTCHANNSVTOCNRTTVKTIIVTELS
TAIL (T) FL-NCA 381 GKWIPASLAIGF PSBG 16 381 EAL PSBG 73 381 DMTVP PSBG 6 381 AXSSSINTSGNRN PSBG 7 288 ASTRIGLIPLINPT CGM 35 C3 AXSSSINTTAVY C1 DMTLP C2 EAL CCA 642 NCA-55 286 C38 GSAPVLSAVATVGITIGVLARVALI					
FL-NCA 381 GKWIPASLAIGF PSG0 16 381 EAL PSG0 93 381 DMTVP PSG0 C 381 AXSSSINTTSONRN PSG0 E 288 ASTRIGLPLLNPT CCM 35 C3 AXSSSINTTAVY C1 DMTLP C2 EAL CCA 642ASGTSPGLSAGATVGIMIGVLVGVALI NCA-55 286GSAPVLSAVATVGITIGVLARVALI	TATI. (T)				
FL-NCA 381 GKWIPASLAIGF PSBG 6 381 EAL PSBG 7 381 DWTVP PSBG C 381 AVTSSINTSGNRN PSBG C 381 ASTRIGLEPLLNPT CGM 35 C3 AVSSSINTAVY C1 DWTLP DWTLP C2 EAL CEA 642 SGSPQLSAGATVGINIGVLVGVALI NCA-55 286 GSAPVLSAVATVGITIGVLARVALI					
PSBG 16 381 EAL PSBG 3 BAI DMTVP PSBG C 381 ASTRIGLPLINPT CGM 35 C3 ASTRIGLPLINPT CGM 35 C3 AVSSSINTAVY C1 DMTLP C2 EAL CEA 642 GS3FVLSAVATVGITIGVLAVALI	FL-NCA	381	GKWIPASLAIGF		
PSBG C 301 AYSSSINYTSGNRN PSBG E 208 ASTRIGLIPLINPT CGM 35 C3 AYSSSINYTAVY C1 DWTLP C2 EAL CEA 642ASGTSPGLSAGATVGIMIGVLVGVALI NCA-55 286GSAPVLSAVATVGITIGVLARVALI	PS8G 16	381 381			
PSBG E 288 ASTRIGLIPLINPT CGM 35 C3 AVSSSINTAVY C1 DWTLP DWTLP C2 EAL CEA 642 GSAPVLSAVATVGITIGVLAUGVALI NCA-55 286 GSAPVLSAVATVGITIGVLAUVALI	PSBG C	381	AYSSSINYTSGNRN		
CGM 35 C3 AXSSSINTTAVY C1 DWTLP DWTLP C2 EAL CEA 642 ASGTSPGLSAGATVGINIGVLVGVALI NCA-55 286 GSAPVLSAVATVGITIGVLARVALI	PSBG E	288	ASTRIGLLPLLNPT		
C2 EAL CEA 642ASGTSPGLSAGATVGIMIGVLVGVALI NCA-55 286GSAPVLSAVATVGITIGVLARVALI	CGM 35 C3		AYSSSINYTAVY		
CEA 642ASGTSPGLSAGATVGIMIGVLVGVALI NCA-55 286GSAPVLSAVATVGITIGVLARVALI	C1 C2		EAL		
NCA-55 286GSAPVLSAVATVGITIGVLARVALİ	CEA CEA	642	ASGTSPGLSAGATVGIMIGVLVGVALI		
	NCA-55	286	GSAPVLSAVATVGITIGVLARVALI		

Sold ---- GSAPVLSAVATVGITIGVLARVALI
 YNALPQENGLSPGAIAGIVIGVVALVALIAVALACFLHFGKTGRASDQRDLTSTNPQSPTTLRTTP#THLTR

FIG. 3. Amino acid sequence (standard one-letter symbols) comparison of FL-NCA, PSBG, CEA, NCA-55, BGPI; and CGM 35 (PS&G/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.

Medical Sciences: Khan et al.



FIG. 4. SDS/PAGE-autoradiographic analysis of immunoprecipitates. (A) Solubilized CHO cells after labeling for 2 hr with [³⁵S]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-PS β G (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-PS β G. The cells were labeled with [³H]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-PS β G 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/PS β G 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 *Hind*III fragments. In contrast, when the N-domain probe was used only one *Hind*III 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 (*BamHI-EcoRI* 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 μ g 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 PS β G 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 PS β G 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 PS β G 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 PS β G clones C and D, on the one hand, from PS β G clones 16 and 93 on the other.

PS β G groups I and II have the domain structure N-A1-A2-B2-T. PS β G 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 PS β G I that differ in the C-terminal tail are known. Alternative splicing can give rise to three different forms of PS β G II (24) with tails identical with or similar to three of the tails found in PS β G I: However, a tail corresponding to that of FL-NCA was not found in the sequenced portion of the PS β G II gene (24). If present, it must be 3' to the other C-terminal sequences. No cDNA clones corresponding to PS β G II has been sequenced. Only one form of PS β G 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 PS β G species (21). In addition, human placenta was found to contain a 54-kDa and a 64-kDa PS β G species (21). FL-NCA was immunoprecipitated by polyclonal anti-CEA and anti-BGPI sera as well as by an antiserum to PS β G. The finding that the former sera precipitated FL-NCA is consistent with the high degree of sequence similarity between FL-NCA/PS β G, 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/PSBG 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 subgroupspecific 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.
- 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.
- Paxton, R., Mooser, G., Pande, H., Lee, T. & Shively, J. (1987) Proc. Natl. Acad. Sci. USA 84, 920-924.
- Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A. & Stanner, S. (1987) Mol. Cell. Biol. 7, 3221-3230.
- Neumaier, M., Zimmerman, W., Shively, L., Hinoda, Y., Riggs, A. & Shively, J. (1988) J. Biol. Chem. 263, 3202–3207.
- 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.
- 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.
- 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.
- Tabor, S. & Richardsson, C. C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767–4771.
- Ellis, L., Clauser, E., Morgan, D. O., Ederly, M., Roth, R. A., & Rutter, W. J. (1986) Cell 45, 721–732.
- 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.
- 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.
- Streydio, C., Lacka, K., Swillens, S. & Vassart, G. (1988) Biochem. Biophys. Res. Commun. 154, 130-137.
- Oikawa, S., Inuzuka, C., Kosaki, G. & Nakazato, H. (1988) Biochem. Biophys. Res. Commun. 156, 68-77.
- 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.
- Williams, A. & Barclay, N. (1988) Annu. Rev. Immunol. 6, 381– 405.
- Björkman, P., Saper, M., Samraoui, B., Bennett, W., Strominger, J. & Wiley, D. (1987) Nature (London) 329, 506-511.