THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 263, No. 7, Issue of March 5, pp. 3202-3207, 1988
Printed in U.S.A.

Characterization of a cDNA Clone for the Nonspecific Cross-reacting Antigen (NCA) and a Comparison of NCA and Carcinoembryonic Antigen*

(Received for publication, August 28, 1987)

Michael Neumaier‡, Wolfgang Zimmermann§, Louise Shively, Yuji Hinoda, Arthur D. Riggs, and John E. Shively

From the Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, California 91010 and the §Institut für Immunbiologie der Universität, D7800 Freiburg, Federal Republic of Germany

NCA (nonspecific cross-reacting antigen), a glycoprotein found in normal lung and spleen, is immunologically related to carcinoembryonic antigen (CEA), which is found in over 95% of colon adenocarcinomas. From a human genomic library, we previously cloned part of an NCA gene and showed that the amino-terminal region has extensive sequence homology to CEA (Thompson, J. A., Pande, H., Paxton, R. J., Shively, L., Padma, A., Simmer, R. L., Todd, Ch. W., Riggs, A. D., and Shively, J. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 2965–2969). We now present the nucleotide sequence of a cDNA clone, containing the entire coding region of NCA (clone 9). The clone was obtained from a λgt10 library made from the colon carcinoma cell line SW 403; the clone contains a 34-amino acid leader sequence, 310 amino acids for the mature protein, and 1.4 kilobases of 3′-untranslated region of the NCA gene. A comparison of the NCA sequence to the CEA sequence (Oikawa, S., Nakazato, H., and Kosaki, G. (1987) *Biochem. Biophys. Res. Commun.* 142, 511–518; Zimmerman, W., Ortlieb, B., Friedrich, R., and von Kleist, S. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 2690–2694) shows that both proteins contain doublets of an immunoglobulin-like domain, of which there are one copy in NCA and three copies in CEA, a 108-amino acid amino-terminal domain with no cysteine residues, and a carboxyl-terminal hydrophobic domain of sufficient length to anchor the glycoproteins in the cell membrane. Overall, the corresponding coding regions possess 85% sequence homology at the amino acid level and 90% homology at the nucleotide level. Forty nucleotides 3′ of their stop codons, the CEA and NCA cDNAs become dissimilar. The 108-amino acid amino-terminal region together with part of the leader peptide sequence corresponds exactly to a single exon described in our previous work. The data presented here further demonstrate the likelihood that CEA recently evolved from NCA by gene duplication, including two duplications of the immunoglobulin-like domain doublet of NCA.

Carcinoembryonic antigen (CEA)[1] is a 180-kDa highly glycosylated glycoprotein. CEA was first described by Gold and Freedman in 1965 (1) as a colon tumor-associated antigen. Immunoassays for CEA have found wide acceptance as a diagnostic tool in primary clinical diagnosis of colon cancer as well as in the long term monitoring of patients following colorectal tumor resection. The original concept that CEA is a tumor-specific antigen was abandoned when small amounts of CEA were found in normal adult colon (2, 3). In addition, several CEA-related antigens expressed in normal and malignant tissues have been described. NCA, the nonspecific cross-reacting antigen, was first described in normal lung and spleen in 1972 (4, 5). Using monoclonal antibodies, Buchegger *et al.* (6) have differentiated between two molecular weight forms of NCA: NCA-55, a 55-kDa glycoprotein found in granulocytes and epithelial cells, and NCA-95, a 95-kDa glycoprotein found only in granulocytes. Other CEA-like antigens include a unique 128-kDa colon tumor-associated antigen (7), two meconium antigens of molecular weights 160 kDa (NCA-2) and 100 kDa (7, 8), and an 85-kDa biliary glycoprotein (BGP I) (9). CEA-like antigens in the serum of normal blood donors have been isolated and characterized with monoclonal antibodies to CEA (10). Four antigens with molecular weights of 200, 180, 114, and 85 kDa were found. The 180- and 85-kDa antigens correspond to CEA and BGP I, respectively. The identities of the remaining antigens are unknown, but these and the meconium antigens may arise from CEA, BGP I, and NCA by proteolytic cleavage or may be distinct gene products. All of the antigens are obviously related by sharing common antigenic determinants but, in most cases, each has also been shown to possess unique antigenic determinants. For CEA and NCA, amino acid sequence information has demonstrated extensive sequence homology between the two antigens and placed them within the immunoglobulin supergene family (11).

Earlier, we reported the sequences of a partial genomic clone for NCA (12) and cDNA clones for CEA (13). These data demonstrated that CEA and NCA each contain a 108-amino acid amino-terminal domain which has no cysteine residues. In addition, CEA contains multiple copies of a domain of 178 amino acid residues, each of which has four cysteine residues and presumably, two disulfide loops. Oikawa *et al.* (14) obtained a near full-length clone for CEA, extending our data and confirming the occurrence of three copies of this 178-amino acid immunoglobulin-like domain in CEA.

* This work was supported by National Cancer Institute Grant CA37808 and a grant from the Dr. Mildred Scheel Stiftung für Krebsforschung. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03550.

‡ Recipient of fellowship support from the Deutsche Forschungsgemeinschaft.

[1] The abbreviations used are: CEA, carcinoembryonic antigen; NCA, nonspecific cross-reacting antigen; kb, kilobase; BGP, biliary glycoprotein; IPTG, isopropyl-1-thio-β-D-galactopyranoside; bp, base pairs.

Since the biological role of CEA and its relationships to other members of its gene family is still not clear, we have continued our efforts to isolate and characterize CEA-like genes. In this report, we present data on an NCA cDNA clone isolated from a colon tumor cell line and a comparison of NCA with CEA.

MATERIALS AND METHODS

Chemicals—Nitrocellulose filters were purchased from Schleicher & Schuell. Restriction enzymes were from Boehringer Mannheim and Bethesda Research Laboratories. S_1 nuclease, T_4 polymerase, T_4 kinase, and T_4 ligase were from Bethesda Research Laboratories. Avian myeloblastosis virus reverse transcriptase was from Life Sciences and deoxy and dideoxy nucleotides were from Pharmacia LKB Biotechnology Inc. For sequencing with T_7 polymerase, the Sequenase kit from United States Biochemicals were used. Sequencing of single-stranded DNA was performed using the Amersham Corp. sequencing kit. [α-^{32}P]dATP, [γ-^{32}P]ATP, [α-^{32}P]dCTP, and [α-^{35}S]dATP were from Du Pont-New England Nuclear. All other reagents were of analytical grade.

Construction and Screening of the cDNA Library—Total cytoplasmic RNA was prepared from SW 403 cells according to Dashal *et al.* (15) and cDNA was synthesized from 5 μg of poly(A)$^+$ RNA essentially as described by Maniatis *et al.* (16). Double-strand cDNA was rendered blunt-ended with T_4 polymerase and then treated with T_4 kinase to ensure that all 5′ ends of the cDNA molecules were phosphorylated.

For convenient cloning into λgt10, asymmetrical *Eco*RI adapters, also containing a *Cla*I site, were added to ends of the cDNA by blunt end ligation (16). The sequence of the adaptor was

5′-AATTCCGTATCGATGTGC
GGCATAGCTACACG-5′

The *Cla*I site is underlined. Only the 5′ end of the shorter strand (14 nucleotides) was phosphorylated, and a 200-fold excess of adaptor over calculated cDNA ends was used during the ligation reaction (16). After size selection by gel electrophoresis and phosphorylation of the 5′ ends, fragments larger than 600 bp were cloned into dephosphorylated arms of λgt10 (Stratagene, San Diego, CA). For *in vitro* packaging, Gigapack packaging extracts (Stratagene) were used. A total of 5×10^5 independent recombinant clones were obtained. The library was amplified using CES 200 cells (rec BC$^-$), which reduces loss of clones containing repetitive sequences (17). Approximately 2×10^5 clones were screened according to the method of Benton and Davies (18) with a 1400-bp *Eco*RI fragment from the genomic NCA clone λ39.2, which codes for a portion of the Ig-like domain of NCA (12). Seven strong positives were obtained. Rescreening was done with a 561-bp *Sau*3A fragment of the same genomic clone that codes for a portion of the amino-terminal domain of NCA. One clone (clone 9) was found to be positive with this probe. Labeling of the probes was performed as described elsewhere (19), and hybridization conditions were as given in the protocol by Cami and Kourilsky (20).

Subcloning into a Phagemid Vector—Restriction analysis of clone 9 showed that the insert was 3.5 kb and contained an *Eco*RI site. The two *Eco*RI fragments of the insert were 2.1 and 1.4 kb. The 3.5-kb insert obtained by *Cla*I digestion and the *Eco*RI fragments were gel-purified and subcloned into the phagemid Bluescript (Stratagene), which can be used to produce either single-stranded DNA or supercoiled double-stranded DNA.

Because color discrimination was poor and only one orientation of the insert DNA could be obtained in recombinant clones, a procedure involving plating on nitrocellulose was developed. In some experiments the transformation mix was spread directly onto nitrocellulose filters on LB/amp plates lacking IPTG and X-gal. After overnight growth of the transformed bacteria, the nitrocellulose filters were transferred onto plates containing 5 mM IPTG and 40 μg/ml X-gal. Excellent color discrimination was obtained 2 to 4 h after transfer to the plates containing inducer and substrate. Plasmid DNA from recombinant clones was prepared by the method of Hattori and Sakaki (21) or by banding in cesium chloride gradients (16).

DNA Sequence Determination—The 1.4-kb fragment was sequenced by standard M13 single-stranded DNA methods. The 2.1- and 3.5-kb fragments were sequenced by using double-stranded plasmid DNA. Three different enzymes and procedures were used at various times. Sequencing with T_7 polymerase was performed with the Sequenase kit from United States Biochemicals adapted to double-stranded sequencing. Commercially available Bluescript sequencing primers were used (Stratagene) or, alternatively, 17- or 18-residue oligonucleotides were designed from known sequences of the insert DNA. Supercoiled plasmid DNA (2 μg) was alkaline-denatured in 100 mM NaOH in a total volume of 22 μl and heated to 65 °C for 5 min. After addition of 50 ng of the respective primer, the sample was neutralized by addition of 2 μl of 1 M sodium acetate, pH 4.5, quickly precipitated with 2.5 volumes of 100% ethanol and kept on solid dry ice for about 20 min. The DNA was then pelleted by centrifugation in a microfuge for 10 min at 4 °C, washed once with 70% ethanol, and dried. Annealed primer/template hybrids could be kept up to 2 weeks at −20 °C.

For sequencing with avian myeloblastosis virus reverse transcriptase and Klenow polymerase, nucleotide working solutions were as given by Zagurski *et al.* (22) and Strauss *et al.* (23). The dried primer/template hybrids were resuspended in deionized water and the appropriate reaction buffer to give a volume of 15 μl. Four μl of [α-^{35}S]dATP (500 Ci/mmol) were added and the sequencing reactions were carried out as described (22, 23). For sequencing with the Sequenase kit the primer/template hybrids were resuspended in a total volume of 15 μl consisting of the reagents supplied with the kit and 10 μCi of [α-^{35}S]dATP. The sequencing reactions were then performed according to the United States Biochemicals protocol. Samples were run on 4–8% polyacrylamide gels (0.4 mm), using standard size gels as well as 85-cm-long gels with up to two loadings per sample.

RESULTS

A cDNA library was constructed in λgt10 using mRNA isolated from the colon carcinoma cell line SW403 and screened with a 1400-bp *Eco*RI probe from a genomic NCA clone (12). Seven positives out of 200,000 clones were obtained, one of which (clone 9) also hybridized with a probe specific for the amino terminus of NCA. Restriction analysis of this clone with *Eco*RI revealed two fragments of size 2.1 and 1.4 kb. When the clone was cut with *Cla*I the full-length 3.5-kb insert was obtained. The 2.1-kb fragment hybridized to probes from the coding region of NCA or CEA. Both *Eco*RI fragments as well as the 3.5-kb *Cla*I insert were subcloned into the respective polylinker sites of the phagemid Bluescript. Restriction analysis and T-tracking of recombinant clones containing the 2.1-kb fragment showed that only one orientation of the insert was obtained when the colonies were grown in the presence of IPTG and X-gal. However, by first growing the cells on nitrocellulose filters without induction by IPTG, the opposite orientation of the insert was obtained. We also observed a considerable improvement of color development when fully grown colonies were induced by transferring the filters to plates containing IPTG and X-gal.

The 1.4-kb fragment was sequenced by standard dideoxy methodology using single-stranded DNA (24). However, the

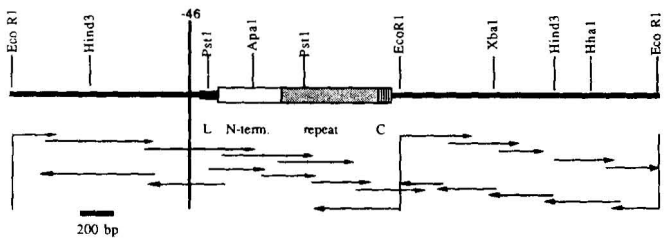

FIG. 1. **Restriction map and sequencing strategy for NCA cDNA clone 9.** The 3.5-kb insert contains 1.0 kb of 5′-flanking region, 1.5 kb of 3′-untranslated region, and 1032 bp of an open reading frame. The open reading frame, shown by *boxes*, contains a leader sequence (*L*, 102 bp), an amino-terminal domain (*N-term*, 324 bp), an Ig-like repeat (*repeat*, 528 bp), and a carboxyl-terminal domain (*C*, 78 bp). The internal *Eco*RI site begins at nucleotide 2108. *Horizontal arrows* indicate regions and orientation of sequence analysis. The *vertical bar* indicates the position at which a cloning artifact in clone 9 may have occurred resulting in the fusion of the NCA gene to an unrelated sequence of 974 bp 5′ of nucleotide *−46*. (See text for a discussion of the 5′-flanking region.)

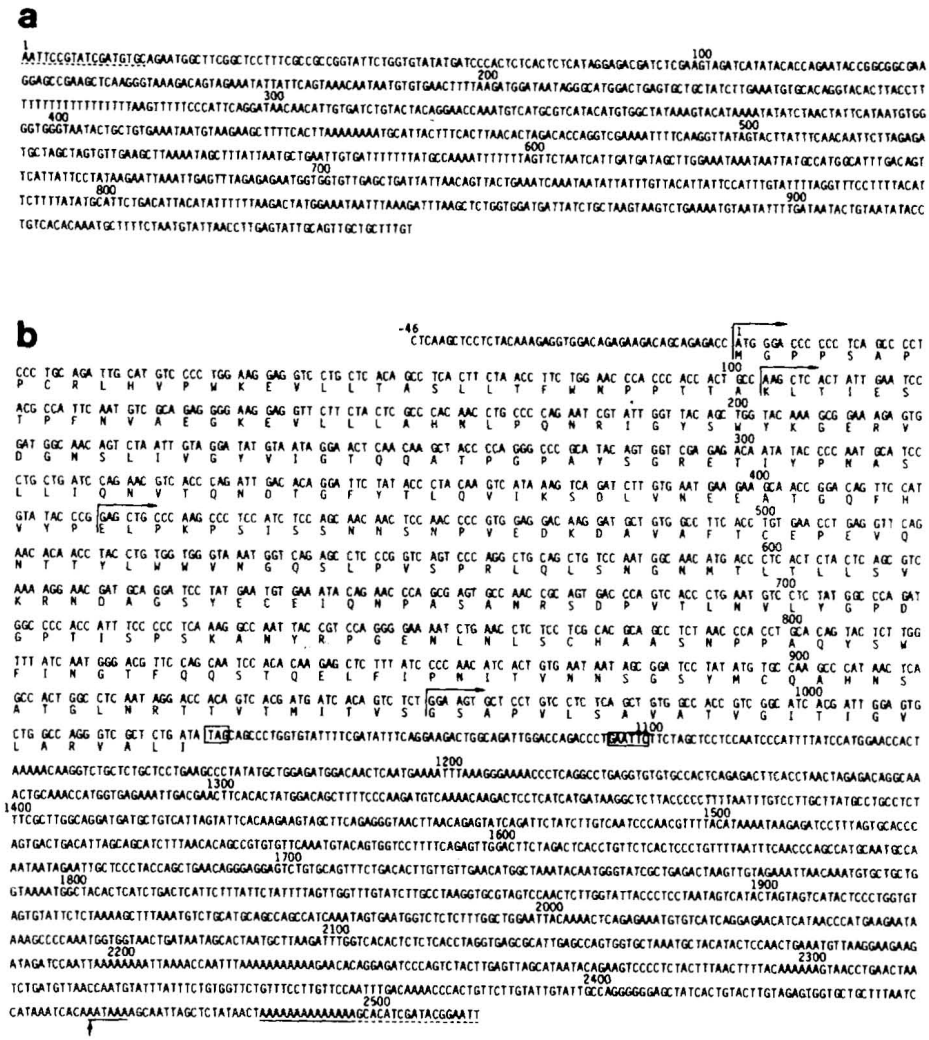

FIG. 2. *a*, nucleotide sequence of the 974 bp that flank the NCA gene at its 5′ end in clone 9. (For a discussion see text.) The cloning adaptor is shown in *broken lines*. *b*, nucleotide sequence of the NCA cDNA gene. The open reading frame begins at nucleotide 1, indicated by the *first horizontal arrow*. The *second horizontal arrow* indicates the amino terminus of the mature protein. The *third horizontal arrow* indicates the beginning of the Ig-like repeat, and the *fourth* the beginning of the hydrophobic, carboxyl-terminal domain. The stop codon and internal *Eco*RI site are *boxed*. The *vertical arrow* indicates the proposed polyadenylation site, followed by a 14-nucleotide poly(A) tail (*underlined*). The asymmetric cloning adaptor is shown by *broken lines*.

2.1-kb *Eco*RI fragment and the 3.5-kb full-length insert were directly sequenced using double-stranded supercoiled plasmids (22, 23). For sequencing of double-stranded DNA, best results were obtained with the T_7 DNA polymerase. For the 3.5-kb subclone, only the region around the internal *Eco*RI site was sequenced, demonstrating that no small fragment was lost during subcloning. The restriction map and sequencing strategy is shown in Fig. 1. Clone 9 has a long 5′-flanking region of 1020 bp, followed by an open reading frame encoding 34 amino acids of a leader peptide, 310 amino acids of the mature protein, and 1430 bp of a 3′-untranslated region ending in a poly(A) tail of 14 adenine residues. The entire sequence of clone 9 is shown in Fig. 2. Recently, a partial genomic clone for NCA has been reported by us (12) and Oikawa *et al.* (31). The genomic clones contain almost 600 bp of sequences upstream of the start of the NCA signal peptide. The nucleotide sequence of the 5′-flanking region of clone 9 (the cDNA containing the NCA gene) diverges from these genomic sequences at position −46 upstream of the translational start. The genomic sequences surrounding nucleotide −46 could be interpreted as an acceptor splice signal (32). To test the hypothesis, whether the divergence was due to an alternative splicing event in the precursor mRNA, we rescreened the cDNA library with a probe specific for the 5′-flanking region of clone 9. Two independent clones (cDNA clones 13 and 2-3) with insert DNAs of 1.9 and 1.5 kb were obtained and partially sequenced (Fig. 3). They match the sequence of the 5′-flanking region of clone 9 (containing the NCA gene) only upstream of nucleotide −46. While there is no homology to the NCA gene downstream of this position, clones 13 and 2-3 show identical sequences, which makes them likely to be transcripts of the same gene. Both clones have the polyadenylation signal AATAAA, and clone 13 possesses also 23 adenine residues. In addition, the sequences surrounding the position of divergence from clone 9 show no indication for a donor splice site. Furthermore, Northern blot analysis of total RNA obtained from SW 403 cells shows two

```
NCA clone 9   AATGCTTTTCTAATGTATTAACCTTGAGTATTGCAGTTGCTGCTTT
    clone 13  AATGCTTTTCTAATGTATTAACCTTGAGTATTGCAGTTGCTGCTTT
    clone 2-3 AATGCTTTTCTAATGTATTAACCTTGAGTATTGCAGTTGCTGCTTT

                    -46                                  1
NCA clone 9   GTCTCAAGCTCCTCTACAAAGAGGTGGACAGAGAAGACAGCAGAGACCATGG
    clone 13  GTACAGAGGTTACTGCAATAAAGGAAGTGGATTCATTAAACCTAAAAAAAAA
    clone 2-3 GTACAGAGGTTACTGCAATAAAGGAAGTGGATTCATTAAACCT

                    10
NCA clone 9   GACCCCCCTCAGCC
    clone 13  AAAAAAAAAAAAA
```

FIG. 3. **Comparison of the clones 2-3 and 13 with clone 9 containing the NCA gene.** Sequences upstream of nucleotide −46 are shared among the three clones (*boldface letters*). Sequences unique to the respective clones are in *lightface letters*. The putative polyadenylation signals are *underlined*. Numbering refers to clone 9. The translational start site for NCA is shown in *underlined boldface letters*.

different size messages, depending on which fragments of clone 9 are used as probes. A 795-bp *Eco*RI/*Bsm*I fragment from the 5′ end of clone 9 identifies a distinct 7-kb message; while using the coding region of NCA as a probe, only a 2.5-kb band is seen, consistent with the length of the actual NCA gene in clone 9 (data not shown). For these reasons we conclude that the first 974 nucleotides in clone 9 may be a cloning artifact rather than the product of a splicing event in the 5′-untranslated region of NCA and represent the 3′-untranslated sequence of an unrelated gene which has been fused to the 5′-untranslated region of the NCA gene.

The amino-terminal sequences of NCA-55 and NCA-95 and a number of sequences from internal peptides of NCA-95 (11) are identical to that predicted by the cDNA sequence. The directly determined amino acid composition of NCA sequence contains 12 potential *N*-glycosylation sites, some of which were previously sequenced by Paxton *et al.* (11).

A comparison of the NCA sequence to CEA (14) is shown in Fig. 4. CEA contains three copies of the immunoglobulin-like repeat domain, only one of which is shown here. Three domains are depicted: a 108-amino acid amino-terminal domain which contains no cysteine residues, a 176-amino acid immunoglobulin-like domain containing two disulfide loops (4 cysteine residues), and a 26-amino acid hydrophobic domain. The first domain is identical to the amino-terminal 107-amino acids of genomic NCA shown by Thompson *et al.* (12). In the domains shown, the homology is 85% at the amino acid level and 90% at the nucleotide level. A comparison of homologies at the nucleotide level of the three copies of the repeated domain in CEA to the single copy in NCA reveals the highest homology for the first copy (90%) compared to 86 and 83% for the second and third copies, respectively. Sequence homology at the nucleotide level continues to 40 bp beyond the stop codon, after which there is no apparent homology between both messages. The nucleotide sequence of approximately 800 bp of the 3′-untranslated region of CEA has been reported. CEA contains a truncated Alu sequence in its 3′-untranslated region which is lacking in NCA.

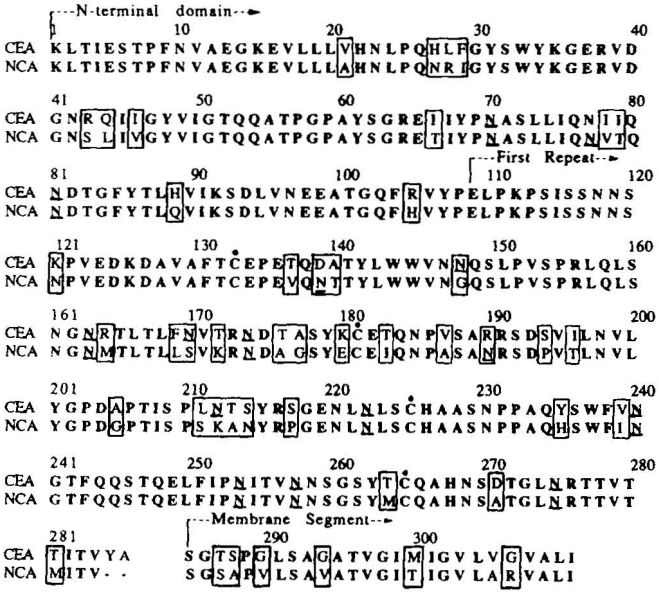

FIG. 4. **Comparison of CEA and NCA amino acid sequences.** *Numbers* refer to the NCA sequence. Only the first Ig-like repeat of CEA is shown. Amino acid differences are *boxed*. Cysteine residues are indicated by a *solid circle*. Asparagine-linked glycosylation sites are *underlined*. The three domains are indicated by *arrows*.

DISCUSSION

Characterization of a cDNA clone containing the entire coding region of NCA makes possible a detailed comparison of NCA to CEA, the cDNA of which was recently cloned by Zimmermann *et al.* (13) and Oikawa *et al.* (14). A high degree of sequence homology was previously noted by amino-terminal sequence analysis (11, 25) and by the extensive immunological cross-reactions between CEA and NCA. Although amino acid differences between CEA and NCA occur throughout their corresponding sequences (Fig. 4), there are regions of perfect homology (*e.g.* residues 45–77) and regions with multiple substitutions (*e.g.* residues 164–196). The 4 cysteine residues of the Ig-like domain are conserved, as are several residues that are invariant in many immunoglobulin supergene family sequences (11).

There are also differences in glycosylation sites. NCA contains 12 potential glycosylation sites, whereas CEA contains 13 potential glycosylation sites for the corresponding sequence. Of the 12 sites in NCA, 8 are conserved in CEA; the unique sites in NCA occur at residue 77 in the amino-terminal domain and residues 139 and 190 in the Ig-like domain. The five glycosylation sites unique to CEA occur at residues 148, 170, 174, 212, and 240, all within the repeated domain. From the number of potential glycosylation sites in NCA, it cannot be concluded whether this clone is equivalent to NCA-95 or NCA-55. However, the similar amino acid compositions (Table I) and amino-terminal sequences (11) for both suggest that the protein sequences are the same for both species of NCA. Limited studies on the glycosylation patterns of NCA indicate a mixture of bi-, tri-, and tetraantennary Asn-linked carbohydrate chains (26). Assuming 12 biantennary chains (M_r = 2,200/chain), NCA could have a minimum molecular weight of 60,000. Assuming 12 tetraantennary chains (M_r = 4,400/chain) NCA could have a maximum molecular weight of 86,000. Thus, it is possible that the two forms of NCA are different only in their glycosylation patterns.

The regions of amino acid sequence differences, especially those which produce different glycosylation patterns, are good candidates for antigenic differences between CEA and NCA. The length of the carboxyl-terminal domains are the same between CEA and NCA and both are of sufficient hydropho-

TABLE I

Amino acid compositions of NCA-95, NCA-55, and NCA cDNA

Data for NCA-95 and NCA-55 are taken from Paxton *et al.* (11). Data for Trp are taken from Engvall *et al.* (25). Data for NCA cDNa are calculated from the nucleotide sequence. The predicted molecular weight for the mature protein is 33,619, based on DNA sequence data.

Amino acid	NCA-95	NCA-55	NCA cDNA
		mol %	
½Cys	1.6	1.4	1.3
Asx	12.7	12.2	11.3
Thr	8.4	8.3	8.7
Ser	9.4	9.3	9.4
Glx	11.3	11.6	10.0
Pro	7.7	7.4	7.1
Gly	6.9	8.1	6.8
Ala	6.2	6.2	7.1
Val	7.0	7.0	8.1
Met	1.0	1.3	1.0
Ile	4.4	4.3	5.5
Leu	8.8	8.6	8.7
Tyr	4.8	4.8	3.9
Phe	2.7	2.5	2.3
His	1.9	1.3	1.6
Lys	2.8	3.2	2.6
Arg	2.6	2.8	3.2
Trp		2.0	1.6

bicity and length to function in membrane insertion. These hydrophobic sequences are reminiscent of the carboxyl-terminal domain found in the Thy-1 antigen (27).

Prior to amino acid sequence analysis of CEA (11), there was no indication that CEA or CEA-like antigens were members of the immunoglobulin gene superfamily. Indeed, the biological function of CEA is still unknown. The discovery of the Ig-like domains in CEA allows the prediction that these domains may fold like adjacent domains of immunoglobulin (*i.e.* CH2 and CH3). With respect to the observed homology, a similar prediction can be made for the Ig-like domain in NCA. Cleavage of the disulfide bonds which connect the two sets of antiparallel β-sheets may lead to the loss of antigenic activity in reduced and *S*-alkylated CEA or NCA (28). Thus, the identification of immunologically reactive epitope-containing domains within CEA and NCA should be assisted by the predicted structure of immunoglobulin. As noted above, each copy of the repeat domains in CEA and NCA actually contain two Ig-like domains (equivalent to CH2 and CH3), which would be interactive, and also likely to form dimeric two-chain structures. Thus, it should now be possible to express the amino-terminal and Ig-like domains of CEA or NCA independently and to test the potential for an immunoglobulin-like fold in these proteins directly. If the folding pattern is indeed like Ig, then the glycosylation sites should be on the surface of the molecule. We are currently using molecular graphic techniques to simulate the structure of CEA and NCA. Preliminary results are in concordance with this model. Evolutionary relationships between CEA and NCA can be deduced from a comparison of their structures. First, it is likely that CEA evolved from the simpler NCA gene by a process of gene duplication and subsequent divergence (Fig. 5). In this view, the Ig-like domain doublet in NCA would have been copied twice in CEA. It may also be expected that some other member of the CEA gene family would have two copies of this repeated domain. The molecular weight of the 128 kDa antigen (7) is of the right size for containing two but not three copies. We are currently sequencing this antigen to test this hypothesis. A comparison of the degree of homology for each copy of these repeats relative to each other and NCA would reveal the approximate time of gene duplication. Since NCA but not CEA is found in lower mammals by immunological tests (29), it is likely that the duplication and gene

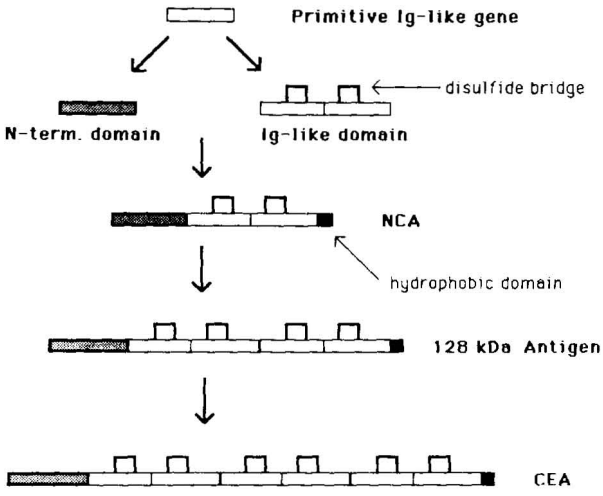

FIG. 5. **Postulated evolution of NCA and CEA genes from a primitive ancestor.** The ancestral gene may have arisen from a single Ig-like gene and duplicated several times to produce this gene family. CEA would be the latest member of the family.

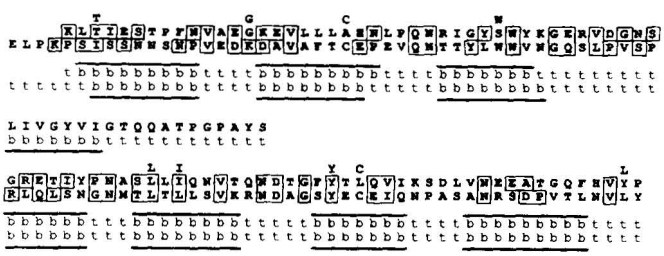

FIG. 6. **Alignment of the amino-terminal and Ig-like domains in NCA.** The first two lines compare the amino acid sequences of the two domains, starting the alignment 6 residues within the Ig-like domain. Homologies which were obtained by dot matrix alignments are indicated with *boxes*, allowing conservative replacements for the groups S,T,G,A,P; Q,N,D,E; H,R,K; I,V,L,M; and Y,F,W. Possible positional displacements are also shown. Maximal homology requires the separate consideration of residues 44–62 (see Oikawa *et al.* (30)). The second pair of lines compares β-strand predictions using the Chou-Fasman algorithm and aligning the Ig-like domain with the CH2 domain of immunoglobulin. Predicted β-strands are shown with a *b* and interconnecting loops or turns are shown with a *t*. Residues 44–62 may correspond to an extra β-strand not found in immunoglobulin. The size of the homologous domains roughly corresponds to one immunoglobulin domain. Invariant residues of the immunoglobulin supergene family members are shown above the amino acid comparisons.

expansion event occurred recently, along with the divergence of higher primates.

The amino-terminal domain has low but significant homology to the Ig-like domain (Fig. 6), suggesting that it also is related to immunoglobulin. The homology is low at the nucleotide level (30%), but higher at the amino acid level (40%), if one allows for conservative replacements and possible positional shifts. The positional shifts are reasonable, given the overall β-sheet structure prediction for this domain. The size of the amino-terminal domain may indicate that it corresponds to one-half the Ig-like domain doublet or the equivalent of a single CH2 or CH3 domain in immunoglobulins. Alignment of CH2 or CH3 to one-half of this NCA domain results in the assignment of seven β-strands. If the amino-terminal domain is aligned to the Ig-like domain in NCA, seven β-strands also are found. The low homology, lack of the disulfide loops, and the size difference suggest that extensive divergence would have had to occur to explain its origin from the Ig-like domains in this molecule. More importantly, the dissimilarity of the amino-terminal domain from the Ig-like domains calls attention to its probable different functional aspect. Since the biological function of these molecules is still unknown, we await the results of further experiments to assign functions to individual domains.

Acknowledgments—We would like to thank Drs. Raymond Paxton and John Thompson for many helpful suggestions and discussions in preparing this manuscript.

REFERENCES

1. Gold, P., and Freedman, S. O. (1965) *J. Exp. Med.* **121,** 439–462
2. Egan, M. L., Pritchard, D. G., Todd, C. W., and Go, V. L. W. (1977) *Cancer Res.* **37,** 2638–2643
3. Fritsche, R., and Mach, J.-P. (1977) *Immunochemistry* **14,** 119–127
4. von Kleist, S., Chavanel, G., and Burtin, P. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69,** 2492–2494
5. Mach, J.-P., and Pusztaszeri, G. (1972) *Immunochemistry* **9,** 1031–1034
6. Buchegger, F., Schreyer, M., Carrel, S., and Mach, J.-P. (1984) *Int. J. Cancer* **33,** 643–649
7. Neumaier, M., Fenger, U., and Wagener, C. (1985) *J. Immunol.* **135,** 3604–3609

8. Burtin, P., Chavanel, G., and Hirsch-Marie, H. (1973) *J. Immunol.* **111,** 1926–1928
9. Svenberg, T. (1976) *Int. J. Cancer* **17,** 588–596
10. Neumaier, M., Fenger, U., and Wagener, C. (1985) *Mol. Immunol.* **22,** 1273–1277
11. Paxton, R. J., Mooser, G., Pande, H., Lee, T. D., and Shively, J. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 920–924
12. Thompson, J. A., Pande, H., Paxton, R. J., Shively, L., Padma, A., Simmer, R. L., Todd, Ch. W., Riggs, A. D., and Shively, J. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 2965–2969
13. Zimmermann, W., Ortlieb, B., Friedrich, R., and von Kleist, S. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 2960–2964
14. Oikawa, S., Nakazato, H., and Kosaki, G. (1987) *Biochem. Biophys. Res. Commun.* **142,** 511–518
15. Dashal, I., Ramirez, S. A., Ballal, R. N., Spohn, W. H., Wu, B., and Busch, H. (1976) *Cancer Res.* **36,** 1026–1034
16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, a Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Nader, W. F., Edlind, T. D., Huettermann, A., and Sauer, H. W. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 2698–2702
18. Benton, W. D., and Davies, R. W. (1977) *Science* **196,** 180–182
19. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13
20. Cami, B., and Kourilsky, P. (1978) *Nucleic Acids Res.* **5,** 2381–2390
21. Hattori, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152,** 232–238
22. Zagursky, R. J., Baumeister, K., Lomax, N., Berman, M. L. (1985) *Gene Anal. Techn.* **2,** 89–94
23. Strauss, E. C., Kobori, J. A., Siu, G., and Hood, L. E. (1986) *Anal. Biochem.* **154,** 353–360
24. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
25. Engvall, E., Shively, J. E., and Wrann, M. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 1670–1674
26. Kessler, M. J., Shively, J. E., Pritchard, D. G., and Todd, C. W. (1978) *Cancer Res.* **38,** 1041–1048
27. Williams, A. F., and Gagnon, J. (1982) *Science* **216,** 696–703
28. Westwood, J. H., and Thomas, P. (1975) *Brit. J. Cancer* **32,** 708–715
29. Haagensen, D. E., Jr., Metzgar, R. S., Swenson, B., Dilley, W. G., Cox, C. E., Davis, S., Murdoch, J., Zamcheck, N., and Wells, S. A., Jr. (1982) *J. Natl. Cancer Inst.* **69,** 1073–1076
30. Oikawa, S., Imajo, S., Noguchi, T., Kosaki, G., and Nakazato, H. (1987) *Biochem. Biophys. Res. Commun.* **144,** 634–642
31. Oikawa, S., Kosaki, G., and Nakazato, H. (1987) *Biochem. Biophys. Res. Commun.* **146,** 464–469
32. Ohshima, Y., and Gotoh, Y. (1987) *J. Mol. Biol.* **195,** 247–259