Tumor Biology

Vol. 9, 1988

The Journal of the International Society for Oncodevelopmental Biology and Medicine (ISOBM)

Founded 1980 as 'Oncodevelopmental Biology and Medicine' by the ISOBM (editors W.H. Fishman and H. Hirai), continued as 'Tumour Biology'
(editors H. Hirai and M. Neville) 1984–1986

KARGER

Contents Vol. 9, 1988

Review

Tumor Biol 1988;9:63–83

1010–4283/88/0093–0063$2.75/0

The Carcinoembryonic Antigen Gene Family: Structure, Expression and Evolution

John Thompson, Wolfgang Zimmermann

Institute of Immunobiology, University of Freiburg, FRG

Key Words. Molecular cloning · Carcinoembryonic antigen (CEA) · Domain model · Exon structure · Immunoglobulin superfamily · Chromosomal localization · mRNA · Epitope mapping · Rat

Abstract. The molecular cloning of carcinoembryonic antigen (CEA) and several cross-reacting antigens reveals a basic domain structure for the whole family, which shows structural similarities to the immunoglobulin superfamily. The CEA family consists of approximately 10 genes which are localized in two clusters on chromosome 19. So far, mRNA species for five of these genes have been identified which show tissue variability in their transcriptional activity. Expression of some of these genes in heterologous systems has been achieved, allowing the localization of some epitopes. The characterization of a CEA gene family in the rat and a comparison with its human counterpart has been utilized in the development of an evolutionary model.

Introduction

Carcinoembryonic antigen (CEA), whose expression appears to be oncodevelopmentally regulated, is one of the most widely used human tumour markers. It was originally discovered by Gold and Freedman [1] in colonic tumours and fetal gut and was apparently absent from the normal intestine. However, later studies also revealed the presence of CEA in normal tissues [2–5]. Despite the lack of tumour specificity, the CEA concentration in sera is an important parameter in the post-operative surveillance of cancer patients [6, 7]. The presence of CEA on the surface of tumour cells has been exploited for the immunolocalization of primary tumours and their metastases using radiolabelled antibodies [8]. First results have also been presented in an animal system as well as in cell culture, for the potential use of such antibodies linked with cytotoxic substances for therapeutic purposes [9, 10]. A number of review articles summarize the clinical utility of CEA and the reader is referred to more recent articles for detailed information [11, 12].

CEA is a highly glycosylated molecule with a molecular weight of 180,000. Glycosylation inhibition studies show the protein

moiety to consist of a single polypeptide chain with an apparent molecular weight of approximately 80,000 [13]. Immunobiochemical studies have revealed a number of closely related molecules, suggesting the existence of a gene family (table I). However, the exact size of this family has been very difficult to ascertain for the following reasons:

(1) Various laboratories have described molecules which cross-react with CEA, and each molecule has received a different name. In some cases, these molecules could be identical; e.g., the classical NCA [31] is presumed to be identical with NGP [35], CEX [19], CCEA-2 [20] and CCA III [21].

(2) The molecular weights for what have been described as the same molecules vary in the literature; e.g., for NCA, values ranging from 40,000 to 120,000 can be found [29, 31, 33, 35, 41].

(3) The faecal antigens may or may not represent degradation products of CEA or other cross-reacting antigens, as already discussed [32].

(4) Variations in the molecular weight of the same molecule could exist, depending on the degree of glycosylation. Such variations have been found for CEA-180 extracted from different tumours (160,000–240,000) [42, 43]. Similarly, NCA-50 (MW 50,000) and TEX (MW 85,000), which have been shown to have the same protein backbone, apparently differ only in their carbohydrate moieties [44].

(5) Investigators using polyclonal antisera may earlier have falsely described contaminants in their CEA preparations as being cross-reactive; e.g., α1-anti-chymotrypsin, which is often co-purified with CEA [45].

With the advent of monoclonal antibodies (MAbs), an attempt has been made toward a more reliable definition of the various members of the CEA family. An exchange of monoclonal antibodies between a number of laboratories has been arranged under the auspices of the International Society for Oncodevelopmental Biology and Medicine, and each participating group agreed to test all the MAbs with one standard preparation for each of the various CEA-related molecules. Although the results of these tests remain outstanding, and much useful information regarding the specificity of the different MAbs should be forthcoming, a completely reliable definition of the different family members still cannot be expected. Only an exact description of the basic structure for each molecular species can lead to a precise definition of this family.

The existence of an antigen family is one of the reasons which have complicated the clinical application of CEA. As table I shows, CEA and many of the cross-reacting antigens occur in malignant and normal tissues. The classical NCA, for instance, is found in granulocytes and macrophages [30, 46–49], whereas BGP-1 is present in the epithelium of bile canaliculi [15]. This can lead to problems in the immunolocalization and immunotherapy of CEA-producing tumours, unless highly specific antibodies are employed [50]. A more exact description of the various members of this family, including an identification of the different epitopes, should help to improve the reliability of CEA as a tumour marker.

The existence of a gene family was already suggested through immunological studies, because most of the epitopes, including those present on more than one family member, are located on the protein moiety of these molecules [51]. Protein sequence

Table I. The CEA family

Antigen	MW	CHO	Source	Reference
βE-Protein	?		colon tumour	14
Biliary glycoprotein I (BGP-I)	90K	40	bile	15
Biliary glycoprotein II (BGP-II)	?		infected bile	15
Biliary glycoprotein III (BGP-III)	?		infected bile	15
Breast carcinoma glycoprotein (BCGP)	?		breast tumour	16
Carcinoembryonic antigen 200 (CEA-200)	200K		colon tumour	17
Carcinoembryonic antigen 180 (CEA-180)	180K	60	colon tumour	1
Carcinoembryonic antigen 160 (CEA-160)	160K		colon tumour	18
CEA-associated protein (CEX)	?		colon/plasma	19
Colonic carcinoembryonic antigen 2 (CCEA-2)	?		colon/plasma	20
Colonic carcinoma antigen III (CCA-III)	60K		serum	21
Fetal sulphoglycoprotein antigen (FSA)	?		stomach	22
Gastric CEA-like antigen (Celia)	?		stomach juices	23
Meconium antigen (MA)	185K		meconium	24
Meconium antigen-100 (MA-100)	105K	50	meconium	25
Melanoma/carcinoma cross-reacting oncofetal Ag	95–150K		melanoma/carcinoma	26
Non-specific cross-reacting antigen 2 (NCA-2)	160K	50	meconium	27
Non-specific cross-reacting antigen 160 (NCA-160)	160K		granulocytes	28
Non-specific cross-reacting antigen 110 (NCA-110)	110K		spleen	29
Non-specific cross-reacting antigen 97 (NCA-97)	97K		colon tumour	18
Non-specific cross-reacting antigen 95 (NCA-95)	95K		granulocytes	30
Non-specific cross-reacting antigen 90 (NCA-90)	90K		granulocytes	28
Non-specific cross-reacting antigen 75 (NCA-75)	75K		colon tumour	18
Non-specific cross-reacting antigen 50 (NCA-50)	50K	30	granulocytes	31
Normal colon washings antigen (NCW)	?		colon	5
Normal faecal antigen 1 (NFA-1)	20–30K	13	faeces	32
Normal faecal antigen 2 (NFA-2)	170K		faeces	33
Normal faecal cross-reacting antigen (NFCA)	80–90K		faeces	34
Normal glycoprotein (NGP)	60K		lung	35
Tumour-extracted antigen (TEX)	85K	60	colon tumour	36
165K antigen	165K		meconium	25
160K antigen	160K		colon tumour	37
128K antigen	128K	50	colon tumour	38
90K antigen	90K		colon tumour	39
50K antigen	50K	25	colon tumour	40
40K antigen	40K		colon tumour	37

The antigens are listed alphabetically, or where no name exists, according to their molecular weights. CHO = Carbohydrate moiety expressed as a percentage of the total molecular weight (MW).

data has already revealed close homology but also distinctive differences between CEA and NCA [29, 36]. However, the determination of the complete primary amino-acid structure has proven difficult due to the large size and high degree of glycosylation of CEA and related molecules (table I). Such information can be more easily provided at the DNA level. Determination of the primary structures of the various molecules belonging to the CEA family should give a better idea of their number as well as their structural interrelationships.

Through cloning the genes for CEA and its cross-reacting antigens a number of problems concerning the reactivity pattern of CEA can be dismissed. For example, studies on engineered mutants for each gene should help to characterize the different epitopes, and molecule-specific determinants may be resolved. Comparative expression studies of the various genes in different tumours and their corresponding normal tissues could reveal the possibility that other members of the CEA family may also be useful as tumour markers. From a biological point of view, it may be possible to gain information as regards the function and evolution of the CEA family. Also, studies on a tissue-specific or oncodevelopmentally regulated expression of the different CEA family members should lead to much useful information with respect to the regulation of gene expression in general. These various aspects will be discussed in more detail below.

The purpose of this review is to summarize the results gained so far through studies at the DNA/RNA level. We shall not, however, review all that is known about the immunochemistry of CEA and related molecules, as this has been well documented elsewhere [12, 52].

History of CEA Cloning

Despite the existence of monospecific, polyclonal CEA antisera and CEA-specific MAbs as well as knowledge of the first 26 amino-acids from the N-termini of both CEA and NCA [29, 36], attempts to clone CEA and related genes were unsuccessful for a long time. The first results on the characterization of CEA mRNA were gained through in vitro translation of total poly(A)-mRNA from a rectal carcinoma, followed by immunoprecipitation with antibodies to CEA [53]. A polypeptide (approximate MW 80,000) was identified as a non-glycosylated CEA precursor, with a relative abundance of 0.6% of total translated protein. This protein was apparently resistant to CNBr cleavage, a result in agreement with the lack of methionine in colonic CEA. The size of the CEA mRNA was estimated to be 3,100 nucleotides by agarose gel electrophoresis [54].

However, before cDNA or genomic clones could be isolated, additional protein sequence data was necessary. A major limitation in gaining this data, as mentioned, has been the high level of glycosylation for different members of the CEA family [12] (table I). This problem was eventually resolved [55] and the newly gained protein sequences paved the way for the molecular cloning and an unequivocal identification of different members of the CEA family. Thus, the isolation and characterization of CEA cDNA clones has recently been achieved in different laboratories at approximately the same time. Basically, two approaches were successful:

(1) Screening of a cDNA library derived from colonic tumours or from a colonic tumour cell line poly(A)-mRNA, using synthetic oligonucleotides based on known protein sequence data [56, 57];

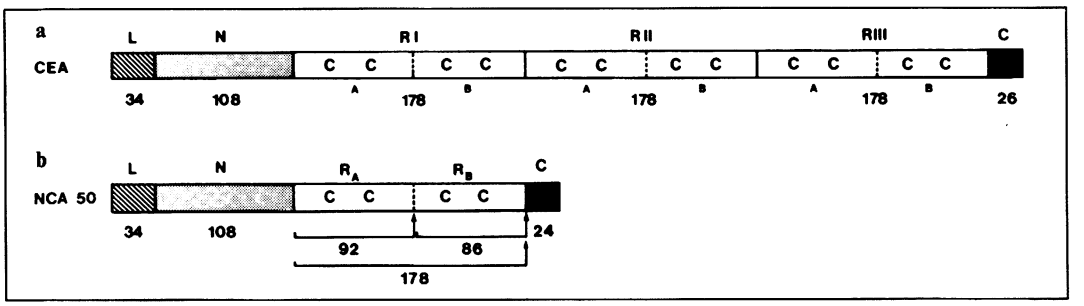

Fig. 1. Domain structure of CEA (**a**) and NCA-50 (**b**), based on deduced amino-acid sequences. Four major domains are indicated by differential shading and letters above the blocks: a leader peptide (L), a N-terminal domain (N), a repeating unit (R), of which three copies are present in CEA (**a**: RI–RIII), but only one in NCA-50 (**b**), and a hydrophobic, carboxyl-terminal region (C). The repeat regions can be further divided into two subdomains, R_A and R_B. Amino-acid numbers are presented below the blocks. The positions of cysteine residues are indicated in the blocks (C).

(2) screening of cDNA libraries derived from colonic tumours and normal colon poly(A)-mRNA as well as from a colonic tumour cell line (LoVo) in expression vectors with CEA antisera [58, 59].

Parallel to this, an NCA-50 gene was isolated from a human genomic library through screening with synthetic oligonucleotides [60, 61]. A second gene belonging to the CEA family was isolated after transfection of genomic DNA from a CEA-producing cell line (LoVo). In order to achieve this, primary and secondary mouse L-cell transfectants expressing a CEA-like molecule were selected by fluorescence-activated cell sorting, and the transfected DNA was identified through hybridization to a CEA-like cDNA clone [59].

Primary Structure of CEA and Its mRNA

The primary structure of CEA has been reported independently from four laboratories [56–58, 62]. The authenticity of these clones has been proven through a comparison with protein sequence data for CEA [55, 63, 64]. However, it has still not been determined unequivocally whether the gene corresponding to these cDNAs codes for all CEAs (table I) or individual CEA species. The amino-acid sequences, as deduced from the nucleotide sequences of the cDNAs, show that CEA is synthesized as a precursor of 702 amino-acids. The leader peptide (34 amino-acids) is followed by the mature CEA peptide (668 amino-acids). The molecular weight for the protein moiety of the mature CEA molecule (minus leader) can be calculated to be 72,893. Due to the presence of three internal repeats, this peptide can be divided into a number of structural domains, which are shown in figure 1a.

The three repeat domains of 178 amino-acids each reveal an exceptionally high degree of sequence similarity, having between 67 and 73% of their amino-acids in common (table II), and each contains four cysteine residues at precisely the same positions. If conservative exchanges are allowed, then the

degree of similarity is even higher. The degree of conservation at the nucleotide level is also very high (80–83% similarity). Other proteins with internal repeats have been reported in the literature, but the internal degree of similarity of the repeating domains in CEA is the highest reported so far [56]. It is also noticeable that the CEA repeats are relatively long [65]. As shown in figure 1, they can be further divided into two subdomains (R_A and R_B) of approximately equal size which reveal sequence similarity to each other as well as to the N-terminal domain (approximately 20% at the amino-acid level) [66]. The importance of this will be discussed in more detail below.

Another region of interest is the C-terminal domain which consists of 26 amino-acids and is strongly hydrophobic. This provides a potential insertion region for CEA into the plasma membrane and has just the right length to span the lipid bilayer [67]. Surface labelling experiments with intact cells indicate that the rest of the molecule projects outward from the cell [R. Bhardwaj: personal commun.]. However, this sequence may be the signal for attachment of a glycophospholipid tail after cleavage of the hydrophobic domain by analogy with LFA-3 [68], Thy-1 and presumably N-CAM [69]. The lack of hydrophilic amino-acids at the carboxyl terminus, which normally anchor membrane proteins, would support this mode of attachment. Determination of the presence or absence of this region in the mature protein should clarify this speculation.

Twenty-eight potential N-glycosylation sites (fig. 3) are found in the CEA peptide [57, 58], many of which have been identified as being glycosylated through protein analyses [55]. This value is considerably lower than the values of 70 [12] or 40 sites [70] predicted from chemical analyses. In order to achieve a molecular weight of 180,000 and assuming that all 28 potential sites are glycosylated, the carbohydrate chain length would have to average 20 residues. The possible existence of O-glycosylation sites as discussed by Shively and Beatty [12] should only lead to marginal changes in these values.

One of the criteria which has been utilized to distinguish CEA from NCA is the apparent lack of methionine in CEA [71]. However, as has been shown from the cDNA-derived primary structure, one methionine does exist in the C-terminal domain of CEA [56, 58]. As this amino-acid is located only 11 amino-acids from the carboxyl terminus, the larger peptide resulting from

Table II. Sequence similarity between corresponding domains of CEA and NCA-50

Domains compared	Degree of sequence similarity	
	nucleotide level, %	amino-acid level, %
L_{CEA}/L_{NCA}	83	74
N-terminus$_{CEA}$/ N-terminus$_{NCA}$	93	89
RI_{CEA}/R_{NCA}	90	84
RII_{CEA}/R_{NCA}	86	76
$RIII_{CEA}/R_{NCA}$	83	73
RI_{CEA}/RII_{CEA}	83	73
$RI_{CEA}/RIII_{CEA}$	82	72
$RII_{CEA}/RIII_{CEA}$	80	67
C-terminus$_{CEA}$/ C-terminus$_{NCA}$	85[a]	71[a]

[a] The C-terminal domain of NCA-50 is two amino-acids (or six nucleotides) shorter than the corresponding domain in CEA. Therefore, the calculations are based only on the corresponding 24 amino-acids and their codons.

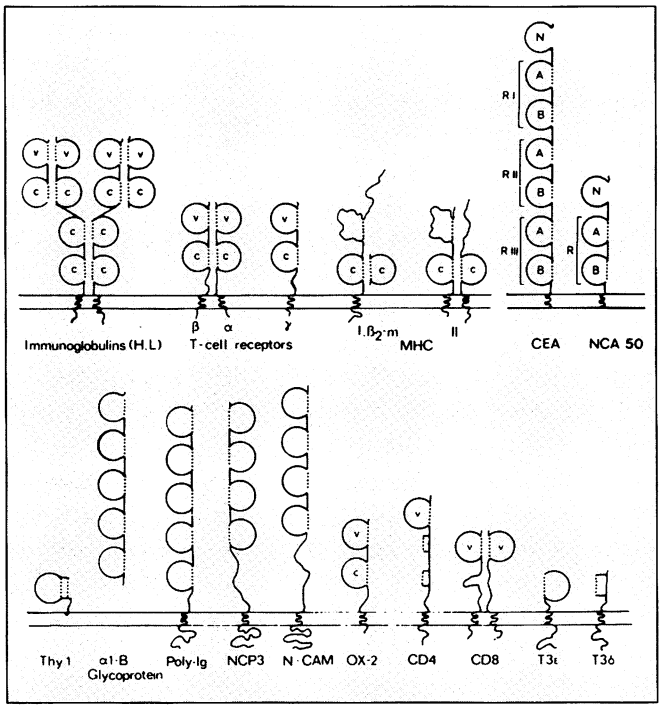

Fig. 2. Domain model for CEA and NCA-50 reveals a close resemblance to the immunoglobulin superfamily. The domains for CEA and NCA-50 are designated as in figure 1. Variable (V) and constant domains (C) are indicated for the immunoglobulin superfamily. The dotted lines represent disulphide bridges. This figure has been modified after Hunkapiller and Hood [100], where the individual members of the immunoglobulin superfamily are described in more detail.

the standard CNBr treatment would not be noticeably smaller in SDS-PAGE analyses than the native CEA molecule. The small peptide would not be visualized as this will migrate out of the gel. For this reason it is understandable that earlier investigations could not detect the presence of methionine in CEA [72]. On the other hand, as already discussed, this methionine-containing C-terminal region may be absent from the mature molecule, if CEA is attached to the membrane by a glycophospholipid tail. In this case, the mature CEA molecule would contain no methionine.

A domain model has been proposed [57, 60, 66], which assumes that neighbouring cysteine residues form disulphide bonds. A model for CEA is shown in figure 2. It is obvious that such a secondary structure shows a strong similarity to members of the immunoglobulin superfamily [73]. The relevance of this will be discussed in more detail below.

Analysis of the CEA cDNA also revealed a region of interest in the non-coding region of the mRNA. The 3′-untranslated region of the processed CEA mRNA contains a truncated, Alu-type, repetitive sequence flanked by a pair of direct repeats. Human Alu DNA is a head-to-tail dimer of two similar sequences with a length of approximately 130 bp. Both monomers can be distinguished by the presence of a 31-bp insert in the second (right) repeat [74]. In the CEA mRNA, approximately 70 nucleotides are deleted from the 5′-terminal region of the first (left) half of the dimer [56, 58]. The significance of the presence of this truncated Alu sequence is

unknown. The 5′-untranslated region has a length of 102 nucleotides as determined by primer extension experiments [57].

Primary Structure of NCA-50 and Its mRNA

The primary structure of NCA-50 has recently been determined [75, 101]. The authenticity of the NCA-50 cDNA clones was proven through a comparison with protein sequence data which were recently extensively investigated [55, 76]. NCA-50 is synthesized as a precursor of 344 amino-acids. The leader peptide (34 amino-acids) is followed by the mature NCA-50 peptide (310 amino-acids). A molecular weight for the protein moiety of the mature NCA-50 (minus leader) can be calculated to be 33,456. In contrast to CEA, NCA-50 does not contain an internal repeating structure. However, due to its strong sequence similarity to CEA, NCA-50 can also be structurally subdivided into a number of domains, which are presented in figure 1b.

It is obvious that the major difference between NCA-50 and CEA is that the former contains only one 'repeat' domain which accounts for the large difference in size between the two polypeptides. This is the major distinction between the two proteins; otherwise, an extremely high conservation between the corresponding domains is apparent at both the amino-acid as well as at the nucleotide levels. These values are presented in table II and will also be discussed in more detail below.

NCA-50 contains 11 potential N-glycosylation sites [75] (fig. 3). Amino-acid sequence analyses have revealed that some of these sites are always glycosylated whereas others are only occasionally or never glycosylated [76]. It has recently been shown that TEX (MW 85,000, see table I) and NCA-50 (MW 50,000), after deglycosylation, yield a polypeptide of the same size [44]. Protein sequence data for both molecules indicate identity, which suggests that the size difference is, in this case, due to variable glycosylation of just one polypeptide.

A few other interesting features about NCA-50 should be mentioned. First of all, in contrast to CEA, NCA-50 contains three methionine residues, which correlates with earlier observations [29]. Secondly, NCA-50 also contains the hydrophobic C-terminal domain, which, although being slightly shorter than the corresponding region in CEA, also provides a potential membrane insertion region. This suggests that NCA-50, like CEA, is membrane-bound (fig. 2). Thirdly, the regions of variability between CEA and NCA-50 at the amino-acid level are often clustered, suggesting the existence of potential molecule-specific epitopes (fig. 3). This important finding could possibly be utilized in future in the production of molecule-specific antibodies which would be of great clinical significance.

Figure 2 shows a domain model for NCA-50, which has obvious similarities to CEA and to molecules belonging to the immunoglobulin superfamily.

Analysis of the cDNA has revealed that the 3′-untranslated region of NCA-50 mRNA diverges completely from the corresponding region in the CEA mRNA [75]. Apart from the first 40 nucleotides directly downstream from the carboxyl terminus, no homology can be found beyond this point. Further, the NCA-50 mRNA does not contain any Alu-like repetitive sequences. As will be described below, these differences

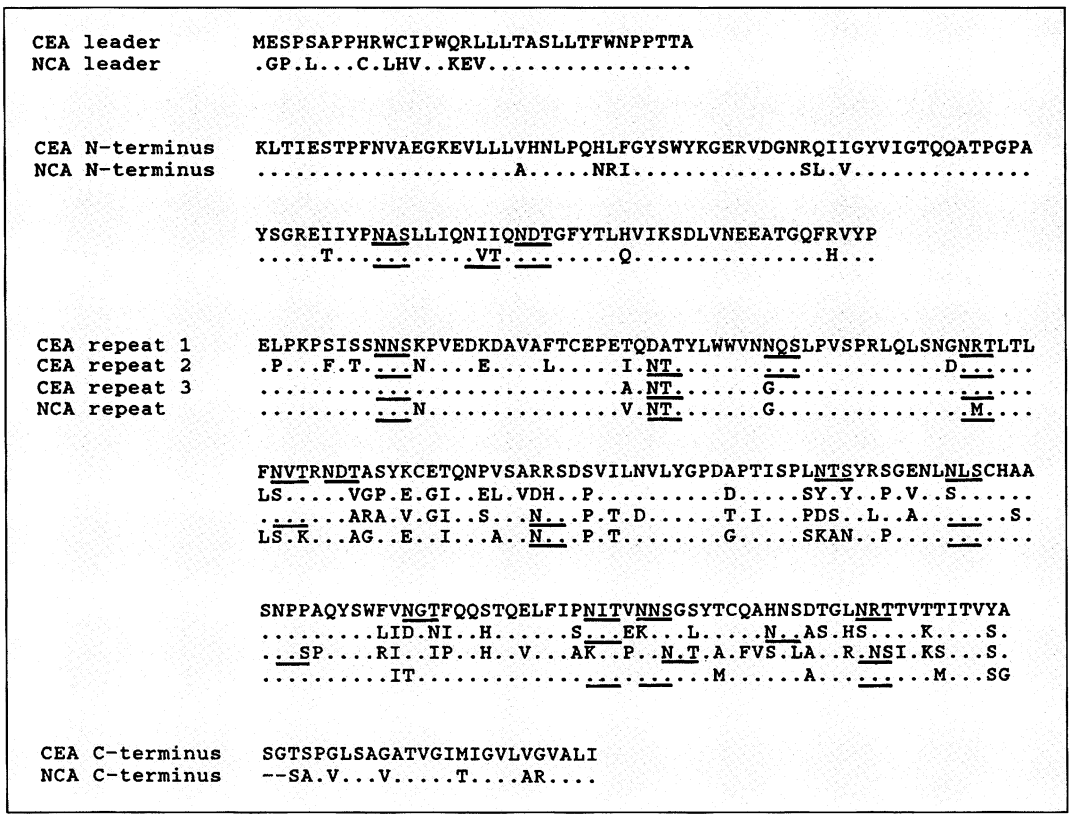

Fig. 3. Amino-acid sequence comparison between CEA and NCA-50. Corresponding domains are grouped together. In each case, dots indicate identity to the amino-acids of the CEA domains shown in the top line of each group. Dashes indicate amino-acid deletions in comparison with CEA. Potential N-glycosylation positions are underlined.

have been utilized in the identification of the CEA and NCA-50 mRNAs in RNA blot hybridization experiments. The NCA-50 3′-untranslated region is similar to corresponding regions in other members of the CEA family, whereas the 3′-non-coding region from the CEA mRNA is apparently unique to two genes, maximally [77]. Assuming a common origin for the CEA-gene family, one may speculate that a new exon could have been inserted into the NCA-50-like, 3′-untranslated region of a primordial CEA gene. A similar event has been suggested to explain the abrupt divergence in the 3′-untranslated region of preprocathepsin B mRNA in the rat and mouse [78]. The major differences between the CEA and NCA-50 mRNAs in this region may also have some functional importance. It has been discussed elsewhere that mRNA 3′-untranslated regions may have regulatory functions [79]. It will be of interest in the future to test for such possible functions in the regulation of expression of CEA and NCA-50 genes.

Relationship to the Immunoglobulin Superfamily

As already indicated in figure 2, the proposed secondary structures for CEA and NCA-50 reveal a strong similarity to members of the immunoglobulin superfamily. This structure is based on the supposition that neighbouring cysteine rests form disulphide bridges, although this still remains to be experimentally verified. However, a comparison of deduced amino-acids has shown conservation of critical amino-acids in comparison with other immunoglobulin superfamily members, indicating that the CEA gene family may be placed within this superfamily [55, 57, 66]. As has already been mentioned, the half repeats (R_A and R_B) as well as the N-terminal domains of CEA and NCA-50 (fig. 1) show sequence similarity to each other. Closer computer analyses of the deduced secondary structure of these regions indicate an anti-parallel arrangement of beta strands in two beta sheets, which is indeed the typical feature of the basic immunoglobulin fold [73]. So far, conservation of the disulphide bond has been adopted as being invariant for the immunoglobulin superfamily. Indeed, this characteristic may be applicable to the half repeat domains of CEA and NCA-50. However, the N-terminal domains lack cysteine residues, but their computer-deduced secondary structures indicate that they, too, can form the typical immunoglobulin fold. This has been discussed in a recent review by Williams [73], who suggests that the conserved disulphide bond should no longer be the last invariant characteristic of these immunoglobulin-related domains. Accordingly, the N-terminal regions of CEA and NCA-50 should, therefore, also be classified as immunoglobulin-like domains. This means that CEA is composed of seven such domains, and NCA-50 of three. These domains form the basic backbone of these two antigens, and as will be discussed below, presumably for the whole CEA family.

Evidence has been presented that CEA may exist in the cell as a non-covalently bound dimer, composed of identical or closely similar monomers [80]. Similar intermolecular interactions have been suggested for NCA [81]. Such structures are known to exist for other members of the immunoglobulin superfamily (fig. 2). However, before conclusive evidence is available, CEA and NCA-50 are depicted as monomers in figure 2.

It is of great interest to note that many of the other members of the immunoglobulin superfamily are often, but not invariably, membrane-bound (fig. 2). Furthermore, where known, all appear to have some kind of recognition or receptor function. The membrane localization, as well as the strong resemblance of CEA and NCA-50 to other immunoglobulin superfamily members is strongly indicative of a similar function for these two gene products. Although no experimental evidence is so far available, this is the first indication of a function for CEA and NCA-50, which may be used as a starting point for directed experiments along these lines.

The CEA-Gene Family

From the above-mentioned primary structures for CEA and NCA-50 a basic structure for the whole family may be predicted. In order to get a more exact idea of the size of this family, DNA fragments from different domains of the NCA-50 gene were

used to probe total human DNA, digested with different restriction endonucleases, in Southern blot hybridization analyses [58–60]. The mean number of fragments which hybridized with a given probe was taken as an indication of the number of gene loci. The results indicate the existence of approximately 9 separate genes. However, the exact size of the family can only be determined when the separate genes are isolated and characterized. Work is in progress in a number of laboratories at both the cDNA as well as at the genomic level in this direction. In our laboratory, we have isolated six independent clones from a human genomic library through differential hybridization with various probes. The results indicate that each clone represents a different member of the CEA family, and the basic domain structure seen for CEA and NCA-50 appears to be conserved. A model may be proposed which could in part explain the size variability of the CEA family, based on differences in the number of immunoglobulin-like domains, which is, indeed, applicable for CEA and NCA-50.

However, recent results indicate that this model alone cannot explain the total variability of the CEA family. Additional domains have also been found for at least one family member [62]. In this case, an additional cytoplasmic domain has been identified, the structure of which shows no homology to any of the other domains described so far. Furthermore, the degree of glycosylation may also have profound effects on the molecular size. As discussed above, recent results indicate that TEX (table I) and the classical NCA-50 probably have the same protein backbone, and the variations in molecular weight are due to differential glycosylation [44].

Genomic Organisation and Chromosomal Localization

Analyses at the genomic level for members of the CEA family indicate a strong correlation between the exon structure and the domains [60, 61, 82]. The structure for NCA-50, as far as it is known, is as follows:

The first exon apparently contains the 5′-untranslated region and approximately two thirds of the leader sequence. The second exon contains the rest of the leader and the complete N-terminal domain. The third exon correlates exactly with the R_A 'repeat' subdomain [82] (fig. 1b). Such half-repeats, corresponding to immunoglobulin-like domains, appear to represent a basic building unit for the CEA gene family, as will be discussed in more detail below. Unpublished data reveal a conservation of this exon/domain structure in other members of the CEA gene family.

We have recently analysed three other genes belonging to the CEA family in and around the N-terminal exon, and a number of interesting results have been obtained [results to be published]. So far, these genes have been unequivocally identified as belonging to the CEA gene family through strong sequence similarities to CEA and NCA-50. However, distinctive differences reveal that they represent separate genes. At present they cannot be assigned to a particular protein, assuming that they are active genes.

For this reason we propose a temporary nomenclature system for these as well as other genes, until their products have been identified. In other to avoid further confusion with the existing nomenclature (table I), we propose that each gene should be assigned numerically, according to the species.

In addition, their assignment to the CEA gene family should also be included. Thus, the following abbreviation is suggested for unidentified human CEA-like genes: hsCGM 1, hsCGM 2 etc. (*Homo sapiens* CEA gene-family member). As will be discussed below, four CEA-like genes have been found in the rat *(Rattus norvegicus)*, which we shall name rnCGM1–rnCGM4.

A poly(dC–dA) sequence has been found in the first intron of the NCA-50 gene [60, 61]. This potential regulatory region is located approximately 100 nucleotides upstream from the start of the second exon, which contains part of the leader sequence and the complete N-terminal domain. Poly(dC–dA)·poly(dG–dT) sequences have been reported in a number of genes [83], and they are believed to form Z-DNA [84]. Such sequences could play a role in the regulation of gene expression [85, 86], or in gene rearrangement [87], as well as recombination [88, 89]. Although the intron sequences for three other genes are very similar in comparison with NCA-50 in the region between the first and second exons, they vary with regard to this poly(dC–dA) sequence. In one gene (hsCGM 1), this sequence is very short [$(dC–dA)_3$], in a second gene (hsCGM 2), it is of medium length [$(dC–dA)_8$], and in a third gene (hsCGm 3), it has the same length as in the NCA-50 gene [$(dC–dA)_{18}$]. In addition, it is followed by a poly(dG–dT) region of approximately equal length. In this latter case, a secondary structure could be formed, which may also have some regulatory function. Future experiments should help us to understand the potential role played by this interesting sequence in the CEA gene family.

The chromosomal localization of the different members of the CEA gene family was independently determined using two different techniques. One approach was to analyse human/mouse somatic cell hybrids with a partial CEA cDNA probe [59]. The second approach utilized human metaphase chromosome spreads which were hybridized with probes from the 'repeat' regions of the NCA-50 gene and a CEA cDNA [77]. In both cases, all genes belonging to the CEA family were located on chromosome 19. A major gene cluster is found on the long arm of this chromosome, and a second locus is indicated at the end of the short arm. Such gene clusters have been found for other gene families; e.g., the immunoglobulins [90], the major histocompatibility complex [91], or the globin gene family [92]. In all cases, where known, the members of a given gene family show similar functions, and the existence of gene clusters may help coordinate expression of the individual genes. For example, cis-acting enhancer sequences could activate more than one gene within such a cluster.

Expression of the CEA-Gene Family

The expression of the CEA-gene family has been studied in a number of tumours and some normal tissues [77, 93]. For these studies, various regions from the coding and non-coding regions were used to probe RNAs in Northern blot hybridization analyses. It has been possible to identify at least six different RNA molecules belonging to this family with gene-specific DNA fragments. For CEA, two mRNAs have been found which appear to be encoded by one gene with two polyadenylation sites [58]. The mRNA for NCA-50 could be shown to be gene-specific through hybridization under stringent conditions, with the NCA-50 3′-untranslated region. A fourth mRNA species

has been found in RNA insolated from chronic myeloid leukaemia granulocytes. This mRNA hybridizes with probes from the coding regions of NCA-50 or CEA, but differs in its size and the 3'-untranslated region. At least two other mRNA species belonging to the CEA family have been reported to be expressed in a choriocarcinoma and in fetal liver [62, 94].

Preliminary studies on the expression of these mRNAs in human tumours derived from different tissues have already revealed some interesting results [77]. Firstly, CEA and NCA-50 mRNAs show great variability in their absolute as well as their relative amounts in a number of different tumours. A strong expression of CEA mRNA was found in adenocarcinomas of the colon, and to a lesser extent, in an adenocarcinoma of the pancreas. This correlates well with the high CEA-positivity of these tumours [12]. These two tumour types also expressed NCA-50 mRNA, whereas a squamous carcinoma of the lung only expressed the NCA-50 mRNA. It was recently reported [56, 93] that normal colon and colon tumours contain similar amounts of CEA mRNAs, although pronounced differences in the relative antigen amounts are obvious: approximately 10–100 × more CEA is found in colon tumours than in normal colon. This indicates that a post-transcriptional control may be important in regulating the expression of CEA in colonic tissues. The mRNA species found in granulocytes from patients with chronic myeloid leukaemia appears to be unique to this type of malignancy and may represent a member of the CEA family which is specifically expressed in the myeloid cell lineage. Indeed, recent immunological and biochemical evidence suggests that this mRNA may code for an NCA-97 in granulocytes [F. Grunert, personal commun.], which is presumed to be identical to NCA-95 [30] (table I). Such results indicate that other members of the CEA family may also be of potential clinical use as tumour markers. It will be of interest to make more extended screenings of human tumours as well as their corresponding normal tissues for the presence of mRNAs of members of the CEA gene family in order to attempt a correlation between their expression and the malignant phenotype.

Expression of CEA and NCA-50 in Heterologous Systems

Prokaryotic Systems

The subcloning of CEA, NCA-50, as well as fragments of these cDNA molecules into expression vectors, followed by transfection into *Escherichia coli,* has allowed their expression in prokaryotes. This system has so far been utilized either to isolate CEA family cDNA clones, by identification of expressed peptides with polyclonal anti-CEA sera [58, 59], or to map epitopes responsible for CEA/NCA MAb-binding [95, 96]. The expressed proteins have thus been immunologically confirmed as belonging to the CEA family.

Although some MAbs, which recognize epitopes common to different members of the CEA family [97], bind to a CEA fragment lacking the N-terminal domain [96], the available data support the notion that the N-terminal regions of CEA and NCA-50 contain some of the major antigenic sites. An NCA-specific, as well as a CEA/NCA cross-reacting MAb bind to the last third of the N-terminal domain of an NCA-50 deletion mutant [95], and a further cross-reacting

MAb recognizes the N-terminal region of CEA [96]. Finer mapping of these epitopes will eventually identify the amino-acids responsible for the specificity of MAb binding. A possible explanation for the high antigenicity of these N-terminal domains may be found in the predicted 'immunoglobulin fold' for these regions in CEA and NCA-50 [73]. The region on NCA-50 which binds the NCA-specific MAb [95] is analogous to the surface loop in immunoglobulins which connects the two antiparallel beta sheets [73]. Furthermore, the N-terminal domains of CEA and NCA-50 are not stabilized by a disulphide bridge. The increased flexibility and the surface accessibility may well account for the high antigenicity of this region.

However, the assessment of conformational epitopes, utilizing the prokaryotic expression of proteins, is hampered by the fact that bacteria exhibit a high reductive potential, which could possibly induce atypical foldings of heterologously expressed proteins. Furthermore, the lack of protein glycosylation plus the production of inclusion bodies containing over-produced, crystallized proteins cause additional difficulties in obtaining 'native' antigenic proteins. The use of eukaryotic expression systems circumvents most of these problems.

Eukaryotic Systems

Expression of CEA and CEA-like molecules has so far been achieved either by transfection of mouse L-cells with total human DNA [59] or by transfection of a eukaryotic expression vector containing the coding region of CEA into Chinese hamster cells [57].

Immunofluorescence staining with anti-CEA sera as well as cell surface labelling followed by immunoprecipitation localized CEA-like molecules onto the cell membrane. The molecule identified on L-cells had a lower molecular weight (150,000) than that isolated from the hamster cells (180,000). The latter has the exact weight of CEA purified from human tumours and indeed reacts with a MAb specific for CEA. Whether the smaller molecular weight of the CEA-reactive molecule in L-cells is due to different glycosylation, differential processing, or expression of a cross-reacting antigen remains to be clarified.

The advantages of eukaryotic expression systems for analysing the expression of members of the CEA gene family are manifold. For example, transfected L-cells provide an elegant method for raising highly specific antibodies against CEA and related proteins in syngeneic animals [59]. Secondly, analyses of transfectants should yield useful information regarding post-transcriptional and post-translational processing of individual gene products, which would be synthesized as a controlled variable in an otherwise constant system. Finally, functional studies may be carried out. As already mentioned, the structural similarity of the CEA family to members of the immunoglobulin superfamily could suggest a recognition or receptor function. Using this system, first investigations may be directed toward studying, for example, the possible role of CEA and related proteins in intercellular recognition.

Evolution of the CEA Gene Family

Recent investigations in our laboratory have revealed the existence of a CEA-like gene family as far down the evolutionary scale as rodents [results to be published].

Investigations below this animal order have not been performed. DNA hybridization analyses indicate a strong sequence divergence between rodents and primates, as human CEA/NCA-50 probes barely hybridize with restriction endonuclease digests of total rodent DNA. Despite this, a number of clones have been isolated from a rat genomic library which represent four different gene loci and which show sequence homology to the N-terminal domain of human NCA-50 as well as conservation of the exon structure [results to be published]. The sequence divergence with respect to the human CEA gene family, as indicated throught the DNA hybridization studies, has been confirmed at the nucleotide level for each rat gene. That genes of the CEA-like family are also expressed in the rat has been shown through hybridization of restriction endonuclease fragments containing part of the exon region of each clone, with rat poly(A)-RNA. Despite the description of molecules with similar expression patterns and biochemical properties to CEA in the rat [98], earlier investigations were unable to detect CEA or NCA below the higher primates using immunochemical methods [99]. This is understandable when we see the great sequence divergence below the primates.

On comparing different members of the CEA family in humans, the extremely high conservation of sequences for corresponding domains is obvious (table II). It may be of interest to note that the N-terminal domain as well as the first repeat from CEA show the highest conservation with respect to the corresponding NCA-50 domains. This could indicate that these domains are under a higher selective pressure than the second and third CEA repeats, possibly for functional reasons. In this manner, other, or modified functions could have evolved for the second and third CEA repeats. Such a mechanism of domain duplication, followed by functional divergence, has been discussed as a general phenomenon enabling proteins to perform new or more complex functions [65]. The general high level of sequence similarity between CEA gene family members is not confined to the coding regions alone, as conservation has also been found in the 3′- and 5′ untranslated regions (compare 57 and 75) as well as in intron regions [results to be published]. The analyses carried out so far in the rat system also indicate a relatively high degree of sequence conservation between the different CEA-like genes within this species [results to be published]. Taking into account the strong sequence divergence between the CEA families in the two species, this leads us to the idea that the families may have arisen independently through parallel gene amplification of a primordial gene, which was present before mammalian radiation took place. A model for the evolution of this gene family is proposed in figure 4. The existence of gene clusters on chromosome 19, as described above, suggests the duplication of a primordial gene through unequal crossing over, giving rise to the various members of the CEA gene family.

The basic building unit can be inferred by comparing internal sequence similarities within the different domains. The N-terminus and the half-repeats reveal sequence similarities, especially in one region (fig. 4, shaded area), indicating a common origin [66, 75]. It is of interest to note that the rat clones which have been isolated also show a higher sequence conservation in this same region in comparison with the N-terminus of CEA and of NCA-50. This indicates that this region may have functional importance. The

MAb recognizes the N-terminal region of CEA [96]. Finer mapping of these epitopes will eventually identify the amino-acids responsible for the specificity of MAb binding. A possible explanation for the high antigenicity of these N-terminal domains may be found in the predicted 'immunoglobulin fold' for these regions in CEA and NCA-50 [73]. The region on NCA-50 which binds the NCA-specific MAb [95] is analogous to the surface loop in immunoglobulins which connects the two antiparallel beta sheets [73]. Furthermore, the N-terminal domains of CEA and NCA-50 are not stabilized by a disulphide bridge. The increased flexibility and the surface accessibility may well account for the high antigenicity of this region.

However, the assessment of conformational epitopes, utilizing the prokaryotic expression of proteins, is hampered by the fact that bacteria exhibit a high reductive potential, which could possibly induce atypical foldings of heterologously expressed proteins. Furthermore, the lack of protein glycosylation plus the production of inclusion bodies containing over-produced, crystallized proteins cause additional difficulties in obtaining 'native' antigenic proteins. The use of eukaryotic expression systems circumvents most of these problems.

Eukaryotic Systems

Expression of CEA and CEA-like molecules has so far been achieved either by transfection of mouse L-cells with total human DNA [59] or by transfection of a eukaryotic expression vector containing the coding region of CEA into Chinese hamster cells [57].

Immunofluorescence staining with anti-CEA sera as well as cell surface labelling followed by immunoprecipitation localized CEA-like molecules onto the cell membrane. The molecule identified on L-cells had a lower molecular weight (150,000) than that isolated from the hamster cells (180,000). The latter has the exact weight of CEA purified from human tumours and indeed reacts with a MAb specific for CEA. Whether the smaller molecular weight of the CEA-reactive molecule in L-cells is due to different glycosylation, differential processing, or expression of a cross-reacting antigen remains to be clarified.

The advantages of eukaryotic expression systems for analysing the expression of members of the CEA gene family are manifold. For example, transfected L-cells provide an elegant method for raising highly specific antibodies against CEA and related proteins in syngeneic animals [59]. Secondly, analyses of transfectants should yield useful information regarding post-transcriptional and post-translational processing of individual gene products, which would be synthesized as a controlled variable in an otherwise constant system. Finally, functional studies may be carried out. As already mentioned, the structural similarity of the CEA family to members of the immunoglobulin superfamily could suggest a recognition or receptor function. Using this system, first investigations may be directed toward studying, for example, the possible role of CEA and related proteins in intercellular recognition.

Evolution of the CEA Gene Family

Recent investigations in our laboratory have revealed the existence of a CEA-like gene family as far down the evolutionary scale as rodents [results to be published].

Investigations below this animal order have not been performed. DNA hybridization analyses indicate a strong sequence divergence between rodents and primates, as human CEA/NCA-50 probes barely hybridize with restriction endonuclease digests of total rodent DNA. Despite this, a number of clones have been isolated from a rat genomic library which represent four different gene loci and which show sequence homology to the N-terminal domain of human NCA-50 as well as conservation of the exon structure [results to be published]. The sequence divergence with respect to the human CEA gene family, as indicated throught the DNA hybridization studies, has been confirmed at the nucleotide level for each rat gene. That genes of the CEA-like family are also expressed in the rat has been shown through hybridization of restriction endonuclease fragments containing part of the exon region of each clone, with rat poly(A)-RNA. Despite the description of molecules with similar expression patterns and biochemical properties to CEA in the rat [98], earlier investigations were unable to detect CEA or NCA below the higher primates using immunochemical methods [99]. This is understandable when we see the great sequence divergence below the primates.

On comparing different members of the CEA family in humans, the extremely high conservation of sequences for corresponding domains is obvious (table II). It may be of interest to note that the N-terminal domain as well as the first repeat from CEA show the highest conservation with respect to the corresponding NCA-50 domains. This could indicate that these domains are under a higher selective pressure than the second and third CEA repeats, possibly for functional reasons. In this manner, other, or modified functions could have evolved for the second and third CEA repeats. Such a mechanism of domain duplication, followed by functional divergence, has been discussed as a general phenomenon enabling proteins to perform new or more complex functions [65]. The general high level of sequence similarity between CEA gene family members is not confined to the coding regions alone, as conservation has also been found in the 3′- and 5′ untranslated regions (compare 57 and 75) as well as in intron regions [results to be published]. The analyses carried out so far in the rat system also indicate a relatively high degree of sequence conservation between the different CEA-like genes within this species [results to be published]. Taking into account the strong sequence divergence between the CEA families in the two species, this leads us to the idea that the families may have arisen independently through parallel gene amplification of a primordial gene, which was present before mammalian radiation took place. A model for the evolution of this gene family is proposed in figure 4. The existence of gene clusters on chromosome 19, as described above, suggests the duplication of a primordial gene through unequal crossing over, giving rise to the various members of the CEA gene family.

The basic building unit can be inferred by comparing internal sequence similarities within the different domains. The N-terminus and the half-repeats reveal sequence similarities, especially in one region (fig. 4, shaded area), indicating a common origin [66, 75]. It is of interest to note that the rat clones which have been isolated also show a higher sequence conservation in this same region in comparison with the N-terminus of CEA and of NCA-50. This indicates that this region may have functional importance. The

peptide of the carcinoembryonic antigen synthesized by human tumor cell lines in the presence of tunicamycin. Mol Immunol 1984;21:743–746.
14 Ørjasaeter A: Demonstration of carcinoembryonic antigens (CEA), non-specific crossreacting antigens (NCA) and an associated alpha protein in normal human tissues and fluids by immunodiffusion techniques. Acta Pathol. Microbiol Scand Section B 1974;82:387–395.
15 Svenberg T: Carcinoembryonic antigen-like substances of human bile. Isolation and partial characterization. Int J Cancer 1976;17:588–596.
16 Kuo T, Rosai J, Tillack TW: Immunological studies of membrane glycoproteins isolated from human breast carcinomas. Int J Cancer 1973;12: 532–542.
17 Imai K, Moriya Y, Fujita H, et al: Immunological characterization and molecular profile of CEA detected by monoclonal antibodies. J Immunol 1984;132:2992–2997.
18 Grunert F, AbuHarfeil N, Schwarz K, et al: Two CEA and three NCA species, although distinguishable by monoclonal antibodies, have nearly identical peptide pattern. Int J Cancer 1985;36: 357–362.
19 Darcy DA, Turberville C, James R: Immunological study of carcinoembryonic antigen (CEA) and a related glycoprotein. Br J Cancer 1973;28:147–160.
20 Turberville C, Darcy DA, Laurence DJR, et al: Studies on carcinoembryonic antigen and a related glycoprotein CCEA-2: preparation and chemical characterization. Immunochemistry 1973;10:841–843.
21 Newman ES, Petras SE, Georgiadis A, et al: Interrelationship of carcinoembryonic antigen and colon carcinoma antigen-III. Cancer Res 1974;34: 2125–2130.
22 Häkkinen I: Immunological relationship of CEA and the fetal sulfoglycoprotein antigen. Immunochemistry 1972;9:1115–1119.
23 Vuento M, Engvall E, Seppälä M, et al: Isolation from human gastric juice of an antigen closely related to the carcinoembryonic antigen. Int J Cancer 1976;18:156–160.
24 Primus FJ, Freeman JW, Goldenberg DM: Immunological heterogeneity of carcinoembryonic antigen: purification from meconium of an antigen related to carcinoembryonic antigen. Cancer Res 1983;43:679–685.
25 Siepen D, Paxton RJ, Neumaier M, et al: Carcinoembryonic antigen (CEA) and two crossreactive antigens of 165 kD and 105 kD isolated from meconium exhibit identical amino terminal sequences. Biochem Biophys Res Commun 1987; 147:212–218.
26 Liao S-K, Kwong PC, Clarke BJ, et al: Monoclonal antibody recognizing melanoma-carcinoma cross-reacting oncofetal antigen epitopically associated with carcinoembryonic antigen. JNCI 1985;74:1047–1058.
27 Burtin P, Chavannel G, Hirsch-Marie H: Characterization of a second normal antigen that crossreacts with CEA. J Immunol 1973;111:1926–1928.
28 Audette M, Buchegger F, Schreyer M, et al: Monoclonal antibody against carcinoembryonic antigen (CEA) identifies two new forms of crossreacting antigens of molecular weight 90,000 and 160,000 in normal granulocytes. Mol Immunol 1987;24: 1177–1186.
29 Engvall E, Shively JE, Wrann M: Isolation and characterization of the normal crossreacting antigen: homology of its NH2-terminal amino acid sequence with that of carcinoembryonic antigen. Proc Natl Acad Sci USA 1978;75:1670–1674.
30 Buchegger F, Schreyer M, Carrel S, et al: Monoclonal antibodies identify a CEA crossreacting antigen of 95 kD (NCA-95) distinct in antigenicity and tissue distribution from the previously described NCA of 55 kD. Int J Cancer 1984;33:643–649.
31 von Kleist S, Chavanel G, Burtin P: Identification of an antigen from normal human tissue that crossreacts with the carcinoembryonic antigen. Proc Natl Acad. Sci USA 1972;69:2492–2494.
32 Kuroki M, Shinoda T, Takayasu T, et al: Immunological characterization and structural studies of normal fecal antigen-1 related to carcinoembryonic antigen. Mol Immunol 1982;19:399–406.
33 Matsuoka Y, Kuroki M, Koga Y, et al: Immunochemical differences among carcinoembryonic antigen in tumor tissues and related antigens in meconuim and adult feces. Cancer Res 1982;42: 2012–2018.
34 Kuroki M, Koga Y, Matsuoka Y: Purification and characterization of CEA-related antigens in normal adult feces. Cancer Res 1981;41:713–720.
35 Mach JP, Pusztaszeri G: Demonstration of a par-

tial identity between CEA and normal glycoprotein. Immunochemistry 1972;9:1031–1034.
36 Kessler MJ, Shively JE, Pritchard DG, et al: Isolation, immunological characterization, and structural studies of a tumor antigen related to carcinoembryonic antigen. Cancer Res 1978;38:1041–1048.
37 Blaszczyk M, Pak KY, Herlyn M, et al: Characterization of gastrointestinal tumor-associated carcinoembryonic antigen-related antigens defined by monoclonal antibodies. Cancer Res 1984;44:245–253.
38 Neumaier M, Fenger U, Wagener Ch: Monoclonal antibodies for carcinoembryonic antigen (CEA) as a model system: Identification of two novel CEA-related antigens in meconium and colorectal carcinoma tissue by western blots and differential immunoaffinity chromatography. J Immunol 1985;135:3604–3609.
39 Herlyn M, Blaszczyk M, Sears HF, et al: Detection of a carcinoembryonic antigen and related antigens in sera of patients with gastrointestinal tumors using monoclonal antibodies in double determinant radioimmunoassays. Hybridoma 1983; 2:329–339.
40 Laferte S, Krantz MJ: Purification and characterization of a 50,000 mol. wt. glycoprotein antigen, and localization of determinants involved in cross-reactivity with carcinoembryonic antigen. Mol Immunol 1983;20:421–431.
41 Hammarström S, Svenberg T, Hedin A, et al: Antigens related to carcinoembryonic antigen. Scand J Immunol 1978;7 (suppl 6):33–46.
42 Banjo C, Shuster J, Gold P: Intermolecular heterogeneity of the carcinoembryonic antigen. Cancer Res 1974;34:2114–2121.
43 Grunert F, Hernando JJ, Schwarz K, et al: Do CEA molecules exhibit organ specific differences? Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 131.
44 Paxton RJ, Hefta S, Shively JE, et al: Protein sequence analysis of the carcinoembryonic antigen gene family. Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 139.
45 Grunert F, AbuHarfeil N, Luckenbach GA, et al: Isolation and characterization of two proteins copurifying with carcinoembryonic antigen. Tumour Biol 1984;5:221–232.
46 Burtin P, Quan PC, Sabine MC: Nonspecific cross reacting antigen as a marker for human polymorphs, macrophages and monocytes. Nature 1975;255:714–716.
47 Bordes M, Knobel S, Martin F: Carcinoembryonic antigen (CEA) and related antigens in blood cells and hematopoetic cells. Eur J Cancer 1975;11: 783–786.
48 Heumann D, Candarjis P, Carrel S, et al: Identification of the normal glycoprotein (NGP) cross-reacting with CEA as a differentiation antigen of myeloid cells and macrophages; in Lehmann FG (ed): Carcino-embryonic Proteins. Amsterdam, Elsevier/North Holland Biomedical Press, 1979, vol 11, pp 3–14.
49 Wahren B; Gahrton G, Hammarström S: Nonspecific crossreacting antigen in normal and leukemic myeloid cells and serum of leukemic patients. Cancer Res 1980;40:2039–2044.
50 Dillman RO, Beauregard JC, Sobol RE, et al: Lack of radioimmunodetection and complications associated with monoclonal anticarcinoembryonic antigen antibody cross-reactivity with an antigen on circulating cells. Cancer Res 1984;44:2213–2218.
51 Hammarström S, Engvall E, Johansson BG, et al: Nature of the tumor-associated determinant(s) of carcinoembryonic antigen. Proc Natl Acad Sci USA 1975;72:1528–1532.
52 Rogers GT: Carcinoembryonic antigens and related glycoproteins: molecular aspects and specificity. Biochim Biophys Acta 1983;695:227–249.
53 Zimmermann W, Friedrich R, Grunert F, et al: Characterization of messenger RNA specific for carcinoembryonic antigen. Ann NY Acad Sci 1983;417:21–30.
54 Zimmermann W, Thompson J, Grunert F, et al: Identification of messenger RNA coding for carcinoembryonic antigen. Contrib Oncol 1984;19:64–74.
55 Paxton RJ, Mooser G, Pande H, et al: Sequence analysis of carcinoembryonic antigen: Identification of glycosylation sites and homology with the immunoglobulin supergene family. Proc Natl Acad Sci USA 1987;84:920–924.
56 Zimmermann W, Ortlieb B, Friedrich R, et al: Isolation and characterization of cDNA clones encoding the human carcinoembryonic antigen reveal a highly conserved repeating structure. Proc Natl Acad Sci USA 1987;84:2960–2964.
57 Beauchemin N, Benchimol S, Cournoyer D, et al: Isolation and characterization of full length func-

tional cDNA clones for human carcinoembryonic antigen (CEA). Mol Cell Biol 1987;7:3221–3230.
58 Oikawa S, Nakazato H, Kosaki G: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence. Biochem Biophys Res Commun 1987;142:511–518.
59 Kamarck ME, Elting JJ, Hart JT, et al: Carcinoembryonic antigen family: expression in a mouse L-cell transfectant and characterization of a partial cDNA in bacteriophage λgt11. Proc Natl Acad Sci USA 1987;84:5350–5354.
60 Thompson JA, Pande H, Paxton RJ, et al: Molecular cloning of a gene belonging to the carcinoembryonic antigen gene family and discussion of a domain model. Proc Natl Acad Sci USA 1987;84: 2965–2969.
61 Oikawa S, Kosaki G, Nakazato H: Molecular cloning of a gene for a member of carcinoembryonic antigen (CEA) gene family; signal peptide and N-terminal domain sequences of nonspecific cross-reacting antigen (NCA). Biochem Biophys Res Commun 1987;146:464–469.
62 Barnett T, Goebel S, Elting J: Cloning and characterization of cDNAs for CEA and related family members. Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 150.
63 Terry WD, Henkart PA, Coligan JE, et al: Structural study of the major glycoprotein in preparations with CEA activity. J Exp Med 1972;136: 200–204.
64 Paxton RJ, Simmer RL, Shively JE: Protein structure and molecular cloning of carcinoembryonic antigen and a related tumor associated antigen (TEX) (abstract). Tumour Biol. 1985;6:327.
65 Li W-L: Evolution of duplicate genes and pseudogenes; in Nei M, Koehn RK (eds): Evolution of Genes and Proteins. Sunderland, Sinauer, 1983, pp 14–37.
66 Oikawa S, Imajo S, Noguchi T, et al: The carcinoembryonic antigen (CEA) contains multiple immunoglobulin-like domains. Biochem Biophys Res Commun 1987;144:634–642.
67 Yarden Y, Escobedo JA, Kuang W-J, et al: Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. Nature 1986;323:226–232.
68 Dustin ML, Selvarj P, Mattaliano RJ, et al: Anchoring mechanisms for LFA-3 cell adhesion glycoprotein at membrane surface. Nature 1987;329: 846–848.
69 Low MG: Biochemistry of the glycosyl-phosphatidylinositol membrane protein anchors. Biochem J 1987;244:1–13.
70 Chandrasekaran EV, Davila M, Nixon DW, et al: Isolation and structures of the oligosaccharide units of carcinoembryonic antigen. J Biol Chem 1983;258:7213–7222.
71 Plow EF, Edgington TS: Speciation of carcinoembryonic antigen and its immunodiagnostic implications: in Herberman RB, McIntire KR (eds): Immunodiagnosis of Cancer. New York, Dekker, 1979, vol 9 (part I), pp 181–239.
72 AbuHarfeil N, Grunert F, von Kleist S: Non-specific crossreacting antigen (NCA) does not contain methionine. Tumour Biol 1984;5:339–350.
73 Williams AF: A year in the life of the immunoglobulin superfamily. Immunol Today 1987;8: 298–303.
74 Jelinek WR, Schmid CW: Repetitive sequences in eukaryotic DNA and their expression. Annu Rev Biochem 1982;51:813–844.
75 Neumaier M, Zimmermann W, Shively L, et al: Characterization of a cDNA clone for the nonspecific crossreacting antigen (NCA) and a comparison of NCA and carcinoembryonic antigen (CEA). J Biol Chem, in press.
76 Grunert F, Schwarz K, Paxton R, et al: Comparison of amino acid sequences from deglycosylated NCA-50 and DNA-deduced sequences. Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 132.
77 Zimmermann W, Weber B, Ortlieb B, et al: Chromosomal localization of the carcinoembryonic antigen gene family and differential expression in various tumors. Cancer Res, in press
78 Chan SJ, San Secundo B, McCormick MB, et al: Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. Proc Natl Acad Sci USA 1986;83:7721–7725.
79 Miyata T, Yasumaga T, Nishida T: Nucleotide sequence divergence and functional constraints in mRNA evolution. Proc Natl Acad Sci USA 1980; 77:7328–7332.
80 Lisowska E, Krop-Watorek A, Sedlaczek P: The dimeric structure of carcinoembryonic antigen (CEA). Biochem Biophys Res Commun 1983;115: 206–211.
81 Lisowska E, Krop-Watorek A, Sedlaczek P: Intermolecular interactions in CEA and NCA. Int Soc

Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 146.
82 Thompson J, Rudert F, Mauch E-M, et al: Intra- and interspecies evolution of the CEA gene family. Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 133.
83 Hamada H, Petrino MG, Kagunaga T, et al: Characterization of genomic poly(dT-dG)·poly(dC-dA) sequences: Structure, organization, and conformation. Mol Cell Biol 1984;4:2610–2621.
84 Vorlickova M, Kypr J, Stokrova S, et al: A Z-like form of poly(dA-dC)·poly(dG-dT) in solution? Nucleic Acids Res 1982;10:1071–1080.
85 Quax W, Egberts WV, Hendriks W, et al: The structure of the vimentin gene. Cell 1983;35:215–223.
86 Hamada H, Seidman M, Howard BH, et al: Enhanced gene expression by the poly(dT–dG)·poly(dC–dA) sequence. Mol Cell Biol 1984; 4:2622–2630.
87 Haniford DB, Pulleyblank DE: The in vivo occurrence of Z-DNA. J Biomol Struct Dynam 1983;1: 593–609.
88 Proudfoot NJ, Maniatis T: The structure of a human α-globin pseudogene and its relationship to α-globin gene duplication. Cell 1980;21:537–544.
89 Slightom JL, Blechl AE, Smithies O: Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 1980;21:627–638.
90 Tonegawa S: Somatic generation of antibody diversity. Nature 1983;302:575–581.
91 Steinmetz M, Hood L: Genes of the major histocompatibility complex in mouse and man. Science 1983;222:727–733.
92 Maniatis T, Fritsch EF, Lauer J, et al: The molecular genetics of human hemoglobins. Annu Rev Genet 1980;14:145–178.
93 Cournoyer D, Boucher D, Beauchemin N, et al: Use of a cDNA probe for carcinoembryonic antigen for the analysis of CEA gene expression in malignant and normal human tissues. Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 148.
94 Khan W, Osterman A, Hammarström S: Molecular cloning of carcinoembryonic antigen and related macromolecules. Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987. abstr. 144.
95 Schwarz K, Brückel N, Kolbinger F, et al: Expression of a member of the carcinoembryonic antigen glycoprotein family in *Escherichia coli.* Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 137.
96 Matsuoka Y, Kuroki M, Oikawa S, et al: Antigenic reactivity of recombinant CEA peptides. Int Soc Oncodev Biol Med: XVth Annu Meet, Quebec, 1987, abstr. 143.
97 Kuroki M, Kuroki M, Koka Y, et al: Monoclonal antibodies to carcinoembryonic antigen: A systematic analysis of antibody specificities by using related normal antigens and evidence for allotypic determinants on carcinoembryonic antigen. J Immunol 1984;133:2090–2097.
98 Abeyounis CJ, Milgrom F: A thermostable antigen characteristic for carcinogen-induced rat intestinal tumors. J Immunol 1976;116:30–34.
99 Wahren B, Gadler F, Gahrten G, et al: NCA: A differentiation antigen of myelopoietic cells in humans and hominoid monkeys. Ann NY Acad Sci 1983;417:344–358.
100 Hunkapiller T, Hood L: The growing immunoglobulin gene superfamily. Nature 1986;323:15–16.
101 Tawaragi Y, Oikawa S, Matsuoka Y, et al: Primary structure of nonspecific crossreacting antigen (NCA), a member of carcinoembryonic antigen (CEA) gene family, deduced from cDNA sequence. Biochem Biophys Res Commun 1988; 150:89–96.

Accepted: November 19, 1987

John Thompson
Institut für Immunbiologie der
Universität Freiburg
Stefan-Meier-Strasse 8
D-7800 Freiburg (FRG)