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
<table>
<thead>
<tr>
<th>No. 1</th>
<th>Editorial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Time for Reflection</td>
</tr>
<tr>
<td></td>
<td>Neville AM, Hirai H</td>
</tr>
<tr>
<td></td>
<td>Review</td>
</tr>
<tr>
<td></td>
<td>Microheterogeneity of Human Alphafetoprotein</td>
</tr>
<tr>
<td></td>
<td>Breborowicz J</td>
</tr>
<tr>
<td></td>
<td>Original Papers</td>
</tr>
<tr>
<td></td>
<td>Biological Changes in Serum Antiproteases as Related to Tumor Progression</td>
</tr>
<tr>
<td></td>
<td>Sugar Receptors of Different Types in Human Metastases to Lung and Liver</td>
</tr>
<tr>
<td></td>
<td>Gabius H-J, Engelhardt R</td>
</tr>
<tr>
<td></td>
<td>Radioimmunoassay of Laminin P1 in Body Fluids of Pregnant Women, Patients with Gynaecological Cancer and Controls</td>
</tr>
<tr>
<td></td>
<td>Würz H, Crombach G</td>
</tr>
<tr>
<td></td>
<td>Elevated Serum CA 125 Levels in Patients with Benign Ascitic or Pleural Effusions</td>
</tr>
<tr>
<td></td>
<td>Mezger J, Wilmanns W, Lamerz R</td>
</tr>
<tr>
<td></td>
<td>Immunohistochemical Localization of Yolk Sac Antigen 1 on Rat Spermatogenic Cells</td>
</tr>
<tr>
<td></td>
<td>Sobis H, Van Hove L, Vandeputte M</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. 2-3</th>
<th>Editorial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Has the Carcinoembryonic Antigen Been a Valuable Discovery?</td>
</tr>
<tr>
<td></td>
<td>Neville AM</td>
</tr>
<tr>
<td></td>
<td>Review</td>
</tr>
<tr>
<td></td>
<td>The Carcinoembryonic Antigen Gene Family: Structure, Expression and Evolution</td>
</tr>
<tr>
<td></td>
<td>Thompson J, Zimmermann W</td>
</tr>
<tr>
<td></td>
<td>Original Papers</td>
</tr>
<tr>
<td></td>
<td>Attempts to Suppress Experimental Allergic Encephalomyelitis in Rats with Alpha Fetoprotein</td>
</tr>
<tr>
<td></td>
<td>Kikukawa A, Peterson BS, Deutsch HF</td>
</tr>
<tr>
<td></td>
<td>Evaluation of Fetal Hemoglobin in Various Malignancies with Reference to the Patients’ Age</td>
</tr>
<tr>
<td></td>
<td>Wolk M, Kieselstein M, Brufman G</td>
</tr>
<tr>
<td></td>
<td>Reactivity of Tumor Cells in Malignant Effusions with a Panel of Monoclonal and Polyclonal Antibodies</td>
</tr>
<tr>
<td></td>
<td>Shaw P, Buckman R, Law J, Baumal R, Marks A</td>
</tr>
</tbody>
</table>
Early Diagnosis of Hepatocellular Carcinoma with Lectin Electrophoresis of Serum Alpha-Fetoprotein
Taga H, Hirai H, Ishizuka H, Kaneda H ............................... 110

Expression of the Antigenic Determinant Recognized by the Monoclonal Antibody 44-3A6
on Select Human Adenocarcinomas and Normal Human Tissues

Evolution of Alpha-Fetoprotein: Sequence Comparisons among AFP Species and with Albumin Species
Baker ME .......................................................... 123

Monoclonal Antibody B72.3 Reacts with a Core Region Structure of O-Linked Carbohydrates
Gold DV, Mattes MJ .................................................. 137

Immunohistochemistry of Carcino-Embryonic Antigen in the Embryo, Fetus and Adult
Nap M, Mollgard K, Burtin P, Fleuren GJ .................................. 145

The Ovarian-Carcinoma-Associated Sebaceous Gland Antigen Is a Member of the HMFG2-Positive Family of Epithelial Mucins
de Kretser T .......................................................... 154

Erratum .............................................................. 164

No. 4
Perspectives
Initiators, Promoters and the Uncertainty Principle
Farber E .............................................................. 165

Original Papers
An Antigen Specific to Hyperplastic Liver Nodules Defined with Monoclonal Antibody: A New Marker for Preneoplastic Cells in Rat Chemical Hepatocarcinogenesis
Okita K, Esaki T, Kurokawa F, Takemoto T, Fujikura Y, Fukumoto T .................. 170

Heterogeneity of Glycolytic Enzyme Activity and Isozyme Composition of Pyruvate Kinase in Breast Cancer
Hennipman A, van Oirschot BA, Smits J, Rijksen G, Staal GEJ ......................... 178

Increased Expression of Thomsen-Friedenreich Antigens during Tumor Progression in Breast Cancer Patients (with 1 color plate)
Wolf MF, Ludwig A, Fritz P, Schumacher K ............................................. 190

Measurements of M-Subunit Lactate Dehydrogenase before Therapy as an Aid to Tumour Typing in Serous Effusions
Nevin GB, Mulholland P .................................................. 195

Simultaneous Estimation of Epidermal Growth Factor Receptors and Steroid Receptors in a Series of 136 Resectable Primary Breast Tumors

Purification and Characterization of Carcinoembryonic Antigen from GW-39, a Xenografted Human Colonic Tumor System
Tu YY, Primus FJ, Shochat D, Goldenberg DM ....................................... 212

Short Communication
Spontaneous Neoplasms in Amphibia
Tsonis PA, Del Rio-Tsonis K ................................................. 221
No. 5 Perspectives
Nunez, E.A. ......................................................... 225

Original Papers
Effect of an Ethanol Extract of Cancer Patients' Serum on the Vulnerability of Lymphocytes to Cortisol
Klein, A.; Bruser, B.; Malkin, A. ........................................... 233
Glycolytic Enzyme Activities in Breast Cancer Metastases
Hennipman, A.; van Oirschot, B.A.; Smits, J.; Rijksen, G.; Staal, G.E.J. ................................................ 241
Trophoblast-Specific β₁-Glycoprotein: Its Spatial Structure and Localization of Antigenic Determinants
Pavlenko, A.F.; Moroz, S.V.; Glazunov, V.P.; Vakorina, T.I.; Belogortseva, N.I.; Odinokov, S.E.; Ovodov, Y.S. ................................................ 249
Human Lung Cancer Nodules in Organotypic Culture: No Evidence of Correlation between the Antiproliferative Effects of Interferons and the Induction of 2',5'-Oligoadenylate Synthetase
Martyre, M.-C.; Beaupain, R.; Falcoff, E. ........................................... 263
An Enzyme Immunoassay for Carcinoembryonic Antigen Which Employs Simultaneous Incubation of Specimen with Solid Phase and Enzyme-Conjugated Antibodies
Lynch, D.M.; Briggman, J.; Borden, K.; Fritsche, H.; Malkin, A. ....................................................... 270

No. 6 Original Papers
Activities of Phosphohexose Isomerase and Other Glycolytic Enzymes in Normal and Tumor Tissue of Patients with Neoplastic Diseases: Comparison with Serum Activities and Correlation to Tumor Staging and Grading
Baumann, M.; Jezussek, D.; Lang, T.; Richter, R.-T.; Brand, K. ......................................................... 281
Correlation between Serum Tumor Marker Levels and Tumor Proliferation in Small Cell Lung Cancer
Establishment of Highly Metastatic Clones without Tumorigenicity Derived from Dunn Osteosarcoma
Chigira, M.; Watanabe, H. ....................................................... 293
Tissue Plasminogen Activator and Inhibitors of Fibrinolysis in Malignant Melanoma (with 1 color plate)
Kwaan, H.C.; Radosevich, J.A.; Xu, C.G.; Lastre, C. ......................................................... 301
Pancreatic Oncofetal Antigen and Carcinoembryonic Antigen in Breast and Colon Carcinoma
Kithier, K.; Samal, B.; Cejka, J.; Whitcomb, M.P.; Mood, D.W. ......................................................... 307
Comparison of the Tyrosine Kinase Activity with the Proliferation Rate in Human Colon Solid Tumors and Tumor Cell Lines
Kéry, G.; Balogh, A.; Teplán, I.; Csuka, O. ......................................................... 315
Acknowledgement to the Reviewers ................................................................. 323
Author Index .............................................................................. 324
Subject Index ............................................................................. 325
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

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

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, RA and RB. 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 (RA and RB) 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 glycolipid 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**

<table>
<thead>
<tr>
<th>Domains compared</th>
<th>Degree of sequence similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nucleotide level, %</td>
</tr>
<tr>
<td>LCEA/LNCA</td>
<td>83</td>
</tr>
<tr>
<td>N-terminusCEA/N-terminusNCA</td>
<td>93</td>
</tr>
<tr>
<td>RI_1CEA/RNCA</td>
<td>90</td>
</tr>
<tr>
<td>RI_2CEA/RNCA</td>
<td>86</td>
</tr>
<tr>
<td>RI_3CEA/RNCA</td>
<td>83</td>
</tr>
<tr>
<td>RI_4CEA/R_IICEA</td>
<td>83</td>
</tr>
<tr>
<td>RI_4CEA/R_IIIIEA</td>
<td>82</td>
</tr>
<tr>
<td>RI_5CEA/R_IIIIEA</td>
<td>80</td>
</tr>
<tr>
<td>C-terminusCEA/C-terminusNCA</td>
<td>85*</td>
</tr>
</tbody>
</table>

* 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.
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 glycoprophospholipid 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
<table>
<thead>
<tr>
<th>CEA leader</th>
<th>MESPSAPPFRWCIPWQRLLLTASLLLTTFWNPTTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCA leader</td>
<td>.GP.I....C.LHV...KEV..................</td>
</tr>
<tr>
<td>CEA N-terminus</td>
<td>KLTIESTFPNVAEGKEVLLLHVNLPLQHGFYGWSYKGERVDGRNQIEGTVGIGTQQATPGPA</td>
</tr>
<tr>
<td>NCA N-terminus</td>
<td>.GP.L...C.LHV...KEV..................</td>
</tr>
<tr>
<td>CEA repeat 1</td>
<td>ELPKPSISSSKNKFDVEKDVAFTCEPETQDATYLWLVVNNQLPVSQRLESLGNGRTLTL</td>
</tr>
<tr>
<td>CEA repeat 2</td>
<td>P......E.L...I.NT...G...{...D...N...E......P.T.N...G...M...</td>
</tr>
<tr>
<td>CEA repeat 3</td>
<td>A.N.T...</td>
</tr>
<tr>
<td>NCA repeat</td>
<td>FNVTRNTASYKTECTQNPVSARRSDVLNVLYGPDAPTISPLNTSYRSGENLNLSCAA</td>
</tr>
<tr>
<td>CEA C-terminus</td>
<td>SGTSPLSAGATVGYIMGVLVGVALI</td>
</tr>
<tr>
<td>NCA C-terminus</td>
<td>--SA.V...V......T...A.R.</td>
</tr>
</tbody>
</table>

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 inter-molecular 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 the 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 isolated 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 through 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 through 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


35 Mach JP, Pusztaszeri G: Demonstration of a par-
57 Beauchemin N, Benchimol S, Cournoyer D, et al: Isolation and characterization of full length func-
81 Lisowska E, Krop-Watorek A, Sedlaczek P: Intramolecular interactions in CEA and NCA. Int Soc


Accepted: November 19, 1987

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