



ANALYSIS OF THE SIZE OF THE CARCINOEMBRYONIC ANTIGEN (CEA) GENE FAMILY: ISOLATION AND SEQUENCING OF N-TERMINAL DOMAIN EXONS

John A. Thompson, Eva-Maria Mauch, Fun-Shan Chen<sup>1</sup>, Yuji Hinoda<sup>1</sup>, Heinrich Schrewe, Barbara Berling, Sabine Barnert, Sabine von Kleist, John E. Shively<sup>1</sup> and Wolfgang Zimmermann

Institute of Immunobiology of the University of Freiburg, Stefan-Meier-Str. 8, D-7800 Freiburg, FRG

<sup>1</sup> Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, CA 91010

Received December 20, 1988

Five members of the human CEA gene family [human pregnancy-specific  $\beta$ 1-glycoprotein (PSBG), hscGM1, 2, 3 and 4] have been isolated and identified through sequencing the exons containing their N-terminal domains. Sequence comparisons with published data for CEA and related molecules reveal the existence of highly-conserved gene subgroups within the CEA family. Together with published data eleven CEA family members have so far been determined. Apart from the highly conserved coding sequences, these genes also show strong sequence conservation in their introns, indicating a duplication of whole gene units during the evolution of the CEA gene family. © 1989 Academic Press, Inc.

Carcinoembryonic antigen (CEA) is one of the most widely used human tumour markers, especially in the post-operative surveillance of patients with colonic tumours (1,2). A number of closely related molecules have been described which crossreact with CEA antisera, indicating a CEA-family. The exact size of the CEA family has been difficult to determine for a number of reasons which were recently discussed (3). However, protein sequencing (4,5), as well as DNA sequencing has lead to the identification of the complete primary structures of CEA (6,7,8), a non-specific crossreacting antigen (NCA: 9,10,11,12), biliary glycoprotein 1 (BGPI: 13), pregnancy-specific  $\beta$ 1-glycoprotein (PSBG-C/D, PSG16, PSG93: 14,15), PSBG-E (14), and "Schwangerschaftsprotein 1" (SP1: 16). These molecules exhibit sequence similarities, and can be subdivided into several domains: a leader peptide, an N-terminal domain, a variable number of repeating units, and apart from BGPI, a short C-terminal domain. In the case of BGPI, a transmembrane region and a cytoplasmic domain follow the repeat domains. The deduced primary and secondary structures of all these CEA-related molecules reveal homology to the immunoglobulin superfamily (4,17,18).

**Abbreviations:** CEA = carcinoembryonic antigen, NCA = non-specific crossreacting antigen, BGPI = biliary glycoprotein 1, PSBG = pregnancy-specific  $\beta$ 1-glycoprotein, SP1 = "Schwangerschaftsprotein 1", hscGM = Homo sapiens CEA gene-family member.

DNA/DNA hybridization studies indicate that approximately 10 genes exist in the CEA family (7,9,19). We have started to screen human genomic libraries in an attempt to isolate and identify all these genes. For this purpose, we have decided to concentrate our investigations on the N-terminal regions of the genes for two main reasons. Firstly, this domain appears to be present only once per antigen (3), whereas multiple copies of the repeat domain are known to exist in CEA (6,7,8), and in other genes belonging to this family (20). Secondly, in order to identify these genes with respect to their products, it is apparently easier to gain partial protein sequence data from the N-terminal domains than from other regions of CEA-like antigens (4,5). In this paper, we shall present sequence data from the N-terminal domains of five genes belonging to the CEA family.

### Materials and Methods

**Chemicals and oligonucleotides:** T4 DNA ligase, polynucleotide kinase, the Klenow fragment of DNA-polymerase I from *Escherichia coli*, and restriction endonucleases were from Boehringer Mannheim Biochemica, Bethesda Research Laboratories and Pharmacia/LKB. Low-gelling temperature agarose was obtained from International Biotechnologies Inc. and Sigma. Nitrocellulose filters were purchased from Schleicher and Schuell. Radiochemicals were supplied by Amersham and New England Nuclear. The sequencing kits were from United States Biochemicals and Amersham. RNasin was obtained from Amersham. All other chemicals were of analytical grade. Oligonucleotides were synthesized by the phosphoramidite method on an Applied Biosystems 308A DNA synthesizer, and purified according to the manufacturers protocol.

**Bacterial strains, human genomic library and cloning vectors:** All work with recombinant DNA was carried out according to the German and US safety regulations. A Charon 4A, human, genomic library (21) was amplified in *Escherichia coli* strains LE392 and BHB2600. A human cosmid genomic library (gift from W. Lindenmaier) was amplified in *E. coli* strain ED8767. For subcloning the insert DNA fragments, either Bluescript phagemid (Stratagene) was used in *E. coli* strains JM109 and RR1 M15, or M13mp18/19 phage, which was amplified in *E. coli* strain JM107.

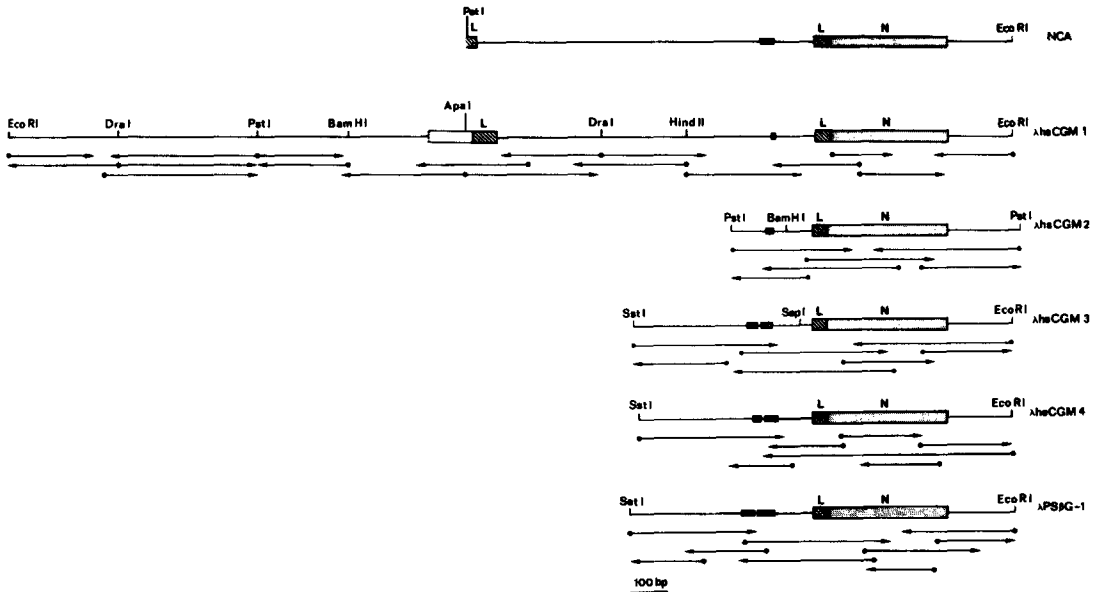
**Screening of the human genomic libraries:** Various DNA fragments and two oligonucleotides (see Results), were radiolabelled (22), and used to screen the human genomic libraries (23). Recombinant phages were amplified and transferred to nitrocellulose filters (24). Colony hybridization (23) was used to identify the cosmid clones. Hybridization with DNA fragments was carried out overnight at 44°C in 6x SSPE (1x SSPE = 0.18M NaCl, 10mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, 1mM EDTA), 0.5% (\*v/v) SDS, 5x Denhardt's solution, 30% formamide, 100µg/ml denatured calf thymus DNA, 10mM EDTA and 1 x 10<sup>6</sup> dpm/ml [ $\alpha$ -<sup>32</sup>P] dATP- labelled DNA fragment. The filters were washed twice (10-15 min at room temperature) in 2x SSPE, 0.5% SDS, followed by two low stringency washes (1h each at 65°C) in the same buffer. The oligonucleotide probes were 5'-end-labelled and hybridized as described (25). Filters were washed three times in 6x SSPE, 0.1% SDS (15 min at room temperature), followed by 15 min washes up to high stringency (60-65°C in the same buffer). Positive plaques were isolated and plaque purified (23).

**Determination of DNA sequences:** Subcloned restriction fragments in phage M13mp18/19 and Bluescript phagemid were sequenced as single or double stranded templates by the dideoxy-chain-termination method (26), using universal, reverse, T3, T7, SK, KS (Stratagene), or internal oligonucleotide primers. For comparison of the nucleotide and deduced amino acid sequences, the computer program "Align" (M. Trippel, Freiburg and R. Friedrich, Giessen, unpublished) was used. Secondary protein structure calculations were performed with the computer program "Novotny" (PCgene, Genofit, Switzerland).

**RNA blot hybridization:** Isolation and analysis of RNA by Northern blot hybridization was performed as described (27).

### Results

For the first screening, a 534bp Pst I fragment from the repeat domains of the CEA cDNA clone pCEA1 (6) was used to probe a human genomic lambda library. After applying low stringency hybridization conditions, 78 clones were picked. For further analyses, a Pst I fragment from an NCA cDNA clone, pNCA1, which covers approximately the last two thirds of the leader and half of the N-terminal domain



**Fig. 1.** Structure of subfragments of six genomic clones containing the N-terminal domain exons of CEA-related genes. Leader (L) and N-terminal domains are depicted by hatched and stippled boxes, respectively. The Pst I/Eco RI NCA fragment whose sequence has been published elsewhere (9) is included for reference. The 5'-untranslated region is depicted as an open box. The transcriptional start which is assumed in analogy to primer extension data for CEA (8) and unpublished data from us for NCA is indicated by a broken line. The smaller solid and horizontally-lined boxes symbolize simple repeated sequences of d(CA) and d(GT) units, respectively. Restriction endonuclease sites relevant for subcloning, sequencing and identification of probes for Northern analyses are indicated. The sequencing strategy is shown by arrows.

was used as a probe (unpublished results). Using the same low stringency conditions as for the first probe, 28 clones also hybridized with this DNA fragment. Three approaches were then adopted in an attempt to identify clones representing genes belonging to the CEA family which have so far not been analysed. Firstly, a direct isolation of the gene for CEA was attempted using a presumably CEA-specific (CEA 3+) oligonucleotide, which has been described elsewhere (9). At a high stringency wash, one clone ( $\Delta$ hsCGM1-1), was identified which preferentially hybridized with this oligonucleotide. Sequence analyses revealed that it does not represent the CEA gene, but a closely related one (see Figures 1, 2 and 3). In a second approach, a 1kb, Eco RI, DNA fragment from the intron between the N-terminal and first repeat exons of an NCA gene (9), was used as a probe. Of the 28 clones which had hybridized with the repeat and N-terminal probes, 4 were identified which did not hybridize with this DNA fragment and could not, therefore, represent the NCA gene. Digestion of the recombinant phage DNAs with different restriction endonucleases and hybridization with the various probes described above, revealed that three of these clones were apparently from the same chromosomal locus (data not shown). One of these three ( $\Delta$ hsCGM2-1), together with the fourth clone ( $\Delta$ hsCGM3-1), was selected for closer analyses of the N-terminal domain exon regions. In a third approach, a number of other clones were isolated which showed very similar restriction patterns to  $\Delta$ hsCGM3-1, although these clones did hybridize with the 1kb Eco RI intron fragment from the NCA gene. Prior to closer analyses, an oligonucleotide (see Figure 2: oligo PSBG 1), based on sequence data from the 3' end of the N-terminal domain of a cDNA clone which we recently isolated from a foetal liver library (results to be published), was used to probe these latter clones. This cDNA clone is identical to










Amino Acid Level							
	CEA	NCA	hsCGM1	hsCGM2	PSβG	hsCGM3	hsCGM4
CEA		90	89	66	57	58	58
NCA	94		92	66	57	58	57
hsCGM1	95	94		64	58	59	58
hsCGM2	81	81	82		52	52	51
PSβG	73	71	73	69		89	89
hsCGM3	73	72	73	68	93		91
hsCGM4	72	70	72	67	93	93	
Nucleotide Level							

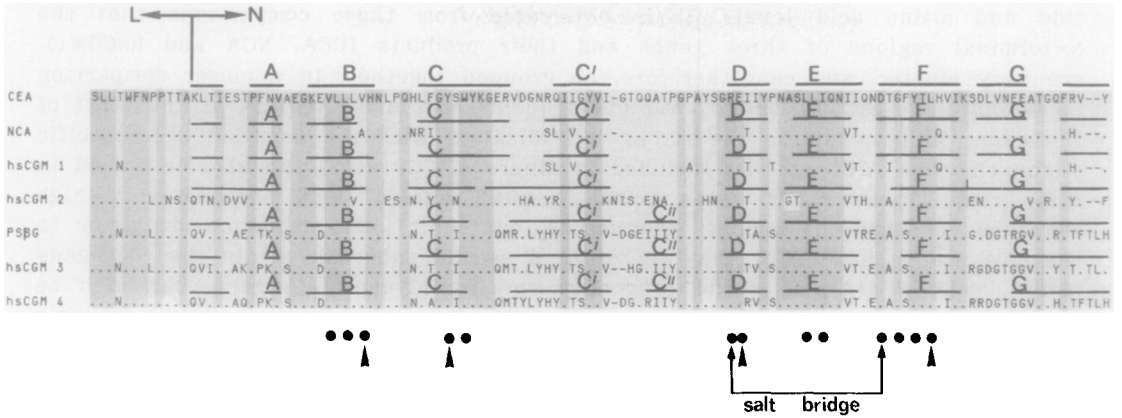
Fig. 3. Comparison of the N-terminal domain exons of human CEA-like genes at the nucleotide and amino acid levels. The percentage similarity values were calculated after optimal alignment. The CEA sequence was derived from cDNA (8) and the NCA from genomic DNA sequences (9).

one of the recently described PSβG cDNA clones (PSβG-D: 15). Oligo PSβG 1 hybridized under stringent conditions with one of these genomic clones (λPSβG-1), which was also selected for further analyses.

In parallel, we also screened a human cosmid genomic library with various probes, and in one approach we isolated a clone (cosmid hsCGM4-1), using a 1.5kb EcoRI fragment probe from the 3' non-translated region of the recently described NCA cDNA clone (11). This cosmid clone hybridized with oligo PSβG 1 under medium stringency conditions (50°C in 6x SSPE), but not at higher stringency (60°C in 6x SSPE), being, therefore, similar, but different to λPSβG-1.

Suitably sized restriction endonuclease fragments, which hybridized with the probe containing the N-terminal region of the NCA cDNA clone, pNCA1 (see above), were subcloned for each of these five genes. The sequencing strategies for these gene clones and their nucleotide and derived amino acid sequences are shown in Figures 1 and 2. The exon/intron borders can be defined by the presence of donor/acceptor consensus sequences (28), as well as a sequence comparison to known data for the NCA gene (9) and its corresponding cDNA (11). Potential glycosylation positions are also shown, whereby some variations can be noted (Figure 2). The deduced amino acid sequences are compared with those for CEA and NCA in Figures 3 and 4.

DNA/RNA hybridization analyses using specific probes for λhsCGM1-1, 2-1 and 3-1 were carried out to determine whether transcripts for these genes could be identified. The probe used for λhsCGM1-1 was an oligonucleotide synthesized from the 5' non-translated region, being complementary to a putative mRNA, which exhibits sequence differences to CEA and NCA (see Figure 2: oligo CGM1). For λhsCGM2-1, a Bam HI/Pst I genomic fragment, and for λhsCGM3-1, an Ssp I/Eco RI genomic fragment was used (see Figure 1), both of which contain the N-terminal domain exons. These probes only hybridized with one genomic restriction endonuclease fragment under high stringency conditions (results not shown). No probes were tested for cosmid hsCGM4-1 and λPSβG-1, as their sequences are so closely related to λhsCGM3-1 (see Figure 3), that any related mRNA species should also have hybridized with the latter probe. None of the probes used could detect corresponding mRNAs under higher stringency conditions in any of the following human tumours which had been propagated in nude mice: carcinomas of the gall bladder



**Fig. 4.** Comparison of the predicted amino acid sequences encoded by the N-terminal domain exons of human CEA-like genes with the corresponding sequences of CEA (8) and NCA (9). Amino acids are identified by the single letter code. Dots indicate amino acids which are identical to CEA. Dashes denote missing amino acids and were inserted to achieve optimal alignment of the compared sequences. Positions which are conserved throughout are shadowed grey. The border between the leader (L) and the N-terminal domain (N) is indicated. The position and extent of the  $\beta$  strands (A-G) are indicated by lines above the sequences having been calculated using the "Novotny" computer program. Residues which are highly conserved in the variable domain of the immunoglobulin superfamily are labelled by solid circles (18). Those not conserved in the CEA family are marked with arrow heads.

and pancreas, squamous carcinomas of the lung and cervix, two adenocarcinomas of the colon, a hepatoma and a bone metastasis of an oesophagus carcinoma. Also negative were the following primary tumours: four adenocarcinomas of the colon, an ovarian sarcoma, one serous and three mucinous ovarian carcinomas, one serous/mucinous ovarian carcinoma and leukocytes from a patient suffering from chronic myeloid leukaemia, as well as metastases from carcinomas of the stomach, colon and breast and normal colonic tissues. However, we were recently able to identify transcripts for PS&G and for hsCGM4 in a 31 week placenta using gene-specific oligonucleotide probes (results to be published).

**Discussion**

The data presented in Figures 2, 3 and 4 reveal that each of the five, newly isolated clones ( $\lambda$ hsCGM1-1,  $\lambda$ hsCGM2-1,  $\lambda$ hsCGM3-1,  $\lambda$ PS&G-1 and cosmid hsCGM4-1), represents an individual member of the CEA gene family. At present, only one of these genes (PS&G), can be identified with respect to a known cDNA (15). The N-terminal domain exon for this gene, reveals only two nucleotide differences to the published cDNA data for PS&G (15) at positions 553 according to our numbering (G instead of C) and 559 (A instead of C) (Figure 2). These two changes both lead to exchanges in the deduced amino acid sequences of the  $\lambda$ PS&G-1 genomic clone; Glu for Gln, and Thr for Pro, respectively. We assume that this gene represents an allele of the published PS&G cDNA because of the otherwise identical coding sequences. For the other newly isolated genes, we have adopted a temporary nomenclature system, as recently proposed (3), until the gene products, when existent, have been identified. We have designated each gene numerically, according to its species (Homo sapiens = hs), as a CEA gene-family member (CGM).

A comparison of the degrees of similarity for each gene with respect to the others, as well as to CEA and an NCA, is shown in Figure 3 at both the nucleo-

tide and amino acid levels. It is noticeable from these comparisons that the N-terminal regions of three genes and their products (CEA, NCA and hsCGM1), are very similar, and can, therefore, be grouped together. In a closer comparison (see Figure 4), hsCGM1 shows complete sequence similarity to CEA at the start of the exon, which explains the strong hybridization with the assumed CEA-specific oligonucleotide. Although not included in Figure 4, BGP1 (13) can also be placed in this group. Three other genes (PS&G, hsCGM3 and 4) also show a very high sequence conservation amongst themselves, but their lower sequence similarity to the CEA group shows that they represent a second subgroup within the CEA gene family. PS&G-E (14), a second pregnancy-specific glycoprotein gene, can also be placed within this subgroup, as can SP1 (16). The other gene (hsCGM2), reveals homology, but a lower sequence similarity to the CEA group, as well as to the PS&G group, indicating that it represents a separate entity or a third subgroup member within the CEA family.

The seven sequences compared in Figure 4 reveal homology to the immunoglobulin superfamily through the conservation of 9 out of 13 critical amino acids (18). This evolutionary relationship becomes more obvious when the deduced secondary structures are compared. In each case, a specific number of  $\beta$ -strands can be deduced, which are characteristic for the immunoglobulin-like fold. Williams (18), recently showed that the N-terminal domain of CEA reveals a close structural similarity to the variable domains of the immunoglobulins, whereas the CEA "half-repeats" are more similar to the constant domains. The extra  $\beta$ -strand (C'), which is a typical feature of immunoglobulin variable domains, appears to be present in the deduced secondary structures of the N-terminal regions for all seven members of the CEA family. In addition, the C'/C'' connecting loop, which represents a region of hypervariability in immunoglobulin variable domains, also reveals higher variability between the different subgroups of the CEA gene family. These structural similarities to members of the immunoglobulin superfamily indicate possible receptor or recognition functions for members of the CEA family. Specificity of function for each individual molecule could be determined by the hypervariable regions. It is noticeable that none of the seven human genes contain cysteine residues in their N-terminal domains. However, in all cases, the arginine and aspartic acid residues which build a salt bridge in the immunoglobulin variable domains (18), are conserved (Figure 4). Therefore, it may be assumed that this salt bridge is sufficient for stabilization of the two hypothetical  $\beta$ -sheets to form the immunoglobulin-like fold of this N-terminal domain in the CEA family.

The high degree of sequence similarity between these CEA-related genes is not restricted to the exon regions. Indeed, a high sequence conservation is also apparent in the surrounding introns (Table 1). However, one interesting region of variability between NCA and the five newly isolated genes is apparent at a location

**Table 1:** Comparison of the intron regions flanking the N-terminal domain exons

NCA	:	hsCGM1	87% <sup>a</sup>	hsCGM1	:	PS&G	80%
NCA	:	hsCGM2	83%	hsCGM2	:	hsCGM3	82%
NCA	:	hsCGM3	82%	hsCGM2	:	hsCGM4	80%
NCA	:	hsCGM4	79%	hsCGM2	:	PS&G	82%
NCA	:	PS&G	82%	hsCGM3	:	hsCGM4	94%
hsCGM1	:	hsCGM2	79%	hsCGM3	:	PS&G	93%
hsCGM1	:	hsCGM3	81%	hsCGM4	:	PS&G	93%
hsCGM1	:	hsCGM4	79%				

<sup>a</sup> The percentage similarity values were calculated after optimal alignment, and only the regions covered by all gene fragments were considered. The NCA sequence has been published elsewhere (9).

approximately 100 bp upstream from the start of the N-terminal exon. All six genes contain a region of simple-sequence DNA, [(dA-dC)<sub>n</sub>], but of varying length (see Figure 2). This upstream region has already been described for the NCA gene (9,10). In PS&G, hsCGM3, and 4 it is followed directly by a poly (dT-dG) sequence. In this latter subgroup a stable secondary structure in the form of a double-hair-pin, or cruciform could be formed. Such poly(dA-dC) · poly (dT-dG) sequences are highly repetitive in the human genome (29). Possible functions of these "TG-elements" have been discussed in the regulation of gene expression (30), or as recombination sites (31).

With regard to the expression of these genes, the PS&G gene has been shown, through Northern blot analyses (own unpublished results), as well as through isolation of corresponding cDNA clones to be transcribed in the placenta (15), and in foetal liver (own unpublished results). PS&G represents a member of a heterogeneous group of pregnancy-specific glycoproteins (32,33). At least three of these immunologically crossreacting glycoproteins were recently isolated from human placenta, with apparent molecular masses of 72, 64 and 54 kDa (15). The close sequence similarity of the PS&G gene subgroup (PS&G, hsCGM3 and 4) presented here (compare also PS&G-E: 14 and SP1: 16), indicates that these pregnancy proteins are encoded by distinct, but highly conserved genes. Investigations to determine this are underway. In order to analyse the potential of these gene products as tumour markers, as well as an attempt to identify potential transcripts for the non-identified genes, we have begun to screen different tumours in Northern analyses. So far, none of the tumours investigated revealed expression of any of these newly-isolated, CEA-related genes but as mentioned, expression of PS&G and hsCGM4 has been found in the placenta. This indicates a high tissue-specificity of expression. At present, it cannot be ruled out that hsCGM1, 2 and 3 are pseudogenes.

If we compare our results with published data (4-16), and assuming that the N-terminal domain occurs only once per gene, then a total of eleven genes have so far been identified which belong to the CEA gene family. These are; CEA (4-8), NCA (9-12), BGP1 (13), hsCGM1, hsCGM6 (a clone recently isolated in our laboratory from a cDNA library constructed from the spleen of a chronic myeloid leukaemia patient: results to be published), PS&G (cf. 14,15), PS&G-E (14), SP1 (16), hsCGM3, hsCGM4 and hsCGM2.

**Acknowledgments:** The oligonucleotides were synthesized by Gabor Igloi on an automated DNA synthesizer which was a gift from the "Fonds der Chemischen Industrie" to Hans Kössel. Assistance with the oligonucleotide purification by Klaus Schwarz is acknowledged. A special thankyou is made to W. Lindenmaier, Braunschweig, FRG, for the human cosmid genomic library. This work was supported by the Deutsche Forschungsgemeinschaft.

## References

1. Gold, P., and Freedman, S.O. (1965) *J. Exp. Med.* 121, 439-462.
2. Shively, J.E., and Beatty, J.D. (1985) *CRC Critical Reviews in Oncology/-Hematology* 2, 355-399.
3. Thompson, J.A., and Zimmermann, W. (1988) *Tumor Biol.* 9, 63-83.
4. Paxton, R., Mooser, G., Pande, H., Lee, T.D., and Shively, J.E. (1987) *Proc. Natl. Acad. Sci. USA* 84, 920-924.
5. Grunert, F., Kolbinger, F., Schwarz, K., Schwaibold, H., and von Kleist, S. (1988) *Biochem. Biophys. Res. Commun.* 163, 1105-1115.
6. Zimmermann, W., Ortlieb, B., Friedrich, R., and von Kleist, S. (1987) *Proc. Natl. Acad. Sci. USA* 84, 2960-2964.
7. Oikawa, S., Nakazato, H., and Kosaki, G. (1987a) *Biochem. Biophys. Res. Commun.* 142, 511-518.



8. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A., and Stanners, C.P. (1987) *Mol. Cell. Biol.* 7, 3221-3230.
9. Thompson, J.A., Pande, H., Paxton, R.J., Shively, L., Padma, A., Simmer, R.L., Todd, C.T., Riggs, A.D., and Shively, J.E. (1987) *Proc. Natl. Acad. Sci. USA* 84, 2965-2969.
10. Oikawa, S., Kosaki, G., and Nakazato, H. (1987c) *Biochem. Biophys. Res. Commun.* 146, 464-469.
11. Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., Riggs, A.D., and Shively, J.E. (1987) *J Biol. Chem.* 263, 3202-3207.
12. Tawaragi, Y., Oikawa, S., Matsuoka, Y., Kosaki, G., and Nakazato, H. (1988) *Biochem. Biophys. Res. Commun.* 150, 89-96.
13. Hinoda, Y., Neumaier, M., Hefta, S.A., Drzeniek, Z., Wagener, C., Shively, J.E., and Paxton, R.J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 6959-6963.
14. Streydio, C., Lacka, K., Swillens, S., and Vassart, G. (1988) *Biochem. Biophys. Res. Commun.* 154, 130-137.
15. Watanabe, S., and Chou, J.Y. (1988a) *J. Biol. Chem.* 263, 2049-2054.
16. Rooney, B.C., Wilson Horne, C.H., and Hardman, N. Gene, in press.
17. Oikawa, S., Imajo, S., Noguchi, T., Kosaki, G., and Nakazato, H. (1987b) *Biochem. Biophys. Res. Commun.* 144, 634-642.
18. Williams, A.F. (1987) *Immunol. Today* 8, 298-303.
19. Kamarck, M.E., Elting, J.J., Hart, J.T., Goebel, S.J., Rae, P.M.M., Nothdurft, M.A., Nedwin, J.J., and Barnett, T.R. (1987) *Proc. Natl. Acad. Sci. USA* 84, 5350-5354.
20. Oikawa, S., Inuzuka, C., Kosaki, G., and Nakazato, H. (1988) *Biochem. Biophys. Res. Commun.* 156, 68-77.
21. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G., and Maniatis, T. (1978) *Cell* 15, 1157-1174.
22. Feinberg, A.P., and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6-13.
23. Maniatis, T., Fritsch, E.F., and Sambrook, J. Eds., (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, pp.90-94.
24. Benton, W.D., and Davis, R.W. (1977) *Science* 196, 180-182.
25. Ullrich, A., Berman, C.M., Dull, T.J., Gray, A., and Lee, J.M. (1984) *EMBO J.* 3, 361-364.
26. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
27. Zimmermann, W., Weber, B., Ortlieb, B., Rudert, F., Schempp, W., Fiebig, H.-H., Shively, J.E., von Kleist, S., and Thompson, J.A. (1988) *Cancer Res.* 48, 2550-2554.
28. Mount, S.M. (1982) *Nucleic Acids Res.* 10, 459-472.
29. Hamada, H., Petrino, M.G., and Kakunaga, T. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6465-6469.
30. Hamada, H., Seidman, M., Howard, B.H., and Gorman, C.M. (1984) *Mol. Cell. Biol.* 4, 2622- 2630.
31. Slightom, J.L., Blechl, A.E., and Smithies, O. (1980) *Cell* 21, 627-638.
32. Bohn, H. (1972) *Blut* 24, 292-302.
33. Tatarinov, Y.S. (1978) *Gynecol. Obstet. Invest.* 9, 65-97.