Vol. 150, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 15, 1988 Pages 89-96

PRIMARY STRUCTURE OF NONSPECIFIC CROSSREACTING ANTIGEN (NCA), A MEMBER OF CARCINOEMBRYONIC ANTIGEN (CEA) GENE FAMILY, DEDUCED FROM cDNA SEQUENCE

Yasunori Tawaragi[1], Shinzo Oikawa[1], Yuji Matsuoka[2], Goro Kosaki[3] & Hiroshi Nakazato[1]

[1]Suntory Institute for Biomedical Research, Shimamoto-cho, Mishima-gun, Osaka, 618, Japan
[2]Department of Biochemistry, School of Medicine, Fukuoka University, Nanakuma, Jonan-ku, Fukuoka 814-01, Japan
[3]Tokyo Metropolitan Komagome Hospital, Bunkyo-ku, Tokyo, 113, Japan

Received November 20, 1987

A cDNA containing the entire coding region for a member of carcinoembryonic antigen (CEA) gene family has been cloned from cDNA library of HLC-1 cells by immunochemical screening with the antibody specific to nonspecific crossreacting antigen (NCA). The cDNA encodes a precursor form of a polypeptide consisting of a 34-residue signal sequence, a 108-residue N-terminal (N-) domain, a 178-residue domain (NCA-I domain) and a 24-residue domain rich in hydrophobic amino acids (M-domain). Each domain has a distinct but homologous amino acid sequence to that of the corresponding domain of CEA. Unlike the coding sequences, the 3'-untranslated sequences differ markedly in the NCA and CEA cDNAs facilitating the preparation of probes that will discriminate between nucleotide sequences for CEA and NCA.

CEA (1,2) is a heterogeneous highly glycosylated protein with a molecular weight of ca. 180,000. Although it is one of the most widely used human tumor markers, it lacks absolute tumor specificity because of the presence of a number of immunologically closely related glycoprotein antigens, e.g., NCA or NCA-1 (3), NCA-2 (4), TEX (5), normal faecal antigens (NFA-1, NFA-2 and NFCA) (6) and BGP-1 (7). Therefore, it is of the utmost importance to elucidate the

Abbreviations: CEA, carcinoembryonic antigen; NCA, nonspecific crossreacting antigen; TEX, tumor-extracted CEA-related antigen; NFA, normal faecal antigen; NFCA, normal faecal cross-reacting antigen; BGP-I, biliary glycoprotein-I; kb, kilobase; bp, base pair.

0006-291X/88 $1.50

primary structure of the protein moieties of these antigens in order to understand their relationships to each other and the biological significance of each molecule.

Recently, we have cloned the cDNAs of mRNA which encode entire CEA peptide and shown that the CEA is a member of Ig supergene family and synthesized as a precursor polypeptide having a signal sequence of more than 30 residues. The 668-residue peptide which follows the signal is constructed as follows consecutively: a 108-residue N-terminal domain, three 178-residue repetitive domains, I, II and III, and a 26-residue putative membrane anchoring domain (8,9). We have also cloned a genomic sequence carrying exons encoding the signal sequence and the N-terminal domain of a CEA related antigen, probably NCA (10).

Recently, partial amino acid sequences of CEA and NCA have been obtained from direct amino acid sequence analysis on protease digests of deglycosylated antigens (11) and deduction from nucleotide sequences of cloned partial cDNA (12,13) and genomic sequence (14) by others.

We report here the cloning of cDNA and complete amino acid sequence of a precursor for an NCA deduced from the nucleotide sequence.

METHODS

cDNA cloning Poly(A)$^+$ RNA was prepared from total RNA extracted from 2 X 10 cells of a human lung carcinoma cell line, HLC-1 (15), as described (16) and was fractionated by sucrose density gradient centrifugation. RNA in each of the resulting fractions was analyzed by Northern blot hybridization (17) using CEA cDNA (N-terminal 201-bp Pst I-Pvu II fragment of pCEA 80-11 (8)) as a probe, and a cDNA library was constructed with RNA in the pooled positive fractions (about 2.5 to 4.0-kb in length) using a phage expression vector λgt11 (18). Double stranded DNA corresponding to the poly(A)$^+$ RNA was prepared according to the RNaseH method of Watson et al (19), ligated to the λgt11 DNA (Promega Biotec. U.S.A.) and packaged using Gigapack (Vector Cloning System, U.S.A.). Approximately 5 X 10^4 clones were plated and submitted to immunological screening using as the first antibody either rabbit anti-CEA antibody (DAKOPATTS, Denmark) or rabbit anti-NCA antibody completely absorbed with CEA (20). Four positive clones were identified and processed further for nucleotide sequencing as described in the legend to Fig.2.

Immunological Methods Immunological staining of plaques transfered to nitrocellulose membrane was performed according to the manufacturer's instruction (Express-Blot Assay Kit, Bio-Rad Lab., U.S.A.). The first antibodies for screening were described above. Peroxidase labelled goat anti-rabbit IgG antibody is from Bio-Rad Lab. (U.S.A.).

Hybridization Analysis One μg of DNA was digested with 30 units of EcoR I for 1 hr at 37 °C and electrophoresed in a 1 % agarose gel. The gel was dried, hybridized for 16 h at 37 °C with CEA or NCA specific oligodeoxyribonucleotide probe described in Fig.1 legend, and washed with Me_4NCl solution according to the method of William, et al (21).

RESULTS

To isolate NCA cDNA clones effectively, a cDNA library was constructed by using λgt11 expression vector and poly(A)$^+$ RNA isolated from HLC-1 cells which

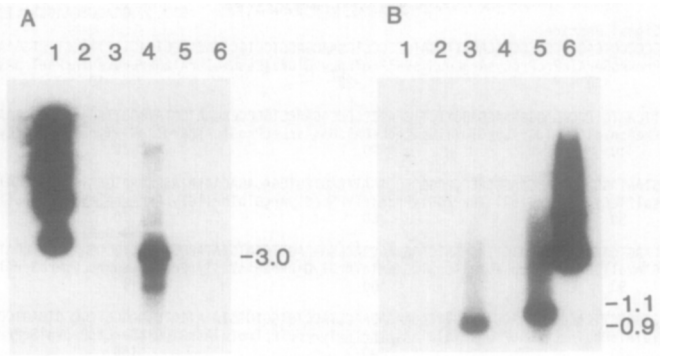


Fig.1 Autoradiogram of the direct hybridization analysis of EcoRI fragments of 4 immunopositive clones. One μg each of DNA digested with EcoRI was analysed as described in METHODS. Plasmid or phage DNA used was; lane 1, pCEA80-11 (CEA positive control, (8)); lane 2, λNCA2; lane 3, λNCA11; lane 4, λNCA14; lane 5, λNCA15; lane 6, pUCNCA4 (NCA positive control, (10)). As a probe, ^{32}P-labeled 17-mer oligodeoxyribonucleotide which corresponds to amino acid 26-31 of CEA (in A) or NCA (in B) was used. DNA fragments that exhibit strong hybridization are denoted by lines with the fragment size in kb.

are known to produce NCA much more than CEA (20). Immunochemical screening of the two 5 X 10^4 clones of the library using as the first antibody, anti-NCA antibody which is specific to NCA (20) or anti-CEA antibody which apparently recognizes NCA in addition to CEA, yielded 4 positive clones. Clone 11, 14 and 15 reacted with the latter and clone 2, 11 and 15 with the former suggesting that clone 2, 11 and 15 contain the NCA cDNA and clone 14 contains the CEA cDNA.

To distinguish the clones carrying NCA cDNA from those carrying CEA cDNA unequivocally, the EcoRI digests of cDNA clones were separated by electrophoresis through agarose gel, which was visually examined for the size of the cDNA fragments and then dried and hybridized with ^{32}P-labeled 17-mer oligodeoxyribonucleotide probes which correspond to amino acids 26-31 of CEA (8) and NCA (10,14). The size of the cDNA fragments released by EcoRI digestion of the clones is 0.9- and 1.4-kb for the clone 11 and 1.1- and 1.4-kb for the clone 15. The 0.9- and 1.1-kb fragments hybridized with the NCA probe but not with the CEA probe (Fig.1). The clone 14 released a single approximately 3-kb fragment which specifically hybridized with the CEA probe (Fig.1). The clone 2 did not give rise to a hybridizing fragment, for unknown reasons. From these results, we identified the clone 11 and clone 15 as those carrying NCA cDNA inserts, clone 15 carrying the longer insert.

Fig.2 depicts the nucleotide sequence derived from the cDNA inserts of the two overlapping clones 11 and 15, and the deduced amino acid sequence. The 2533-bp cDNA (excluding the poly(A) tract) has a long open reading frame ending at nucleotide 1032, the start of the frame being assigned at the nucleotide 1 which is the first nucleotide of the codon for the translation

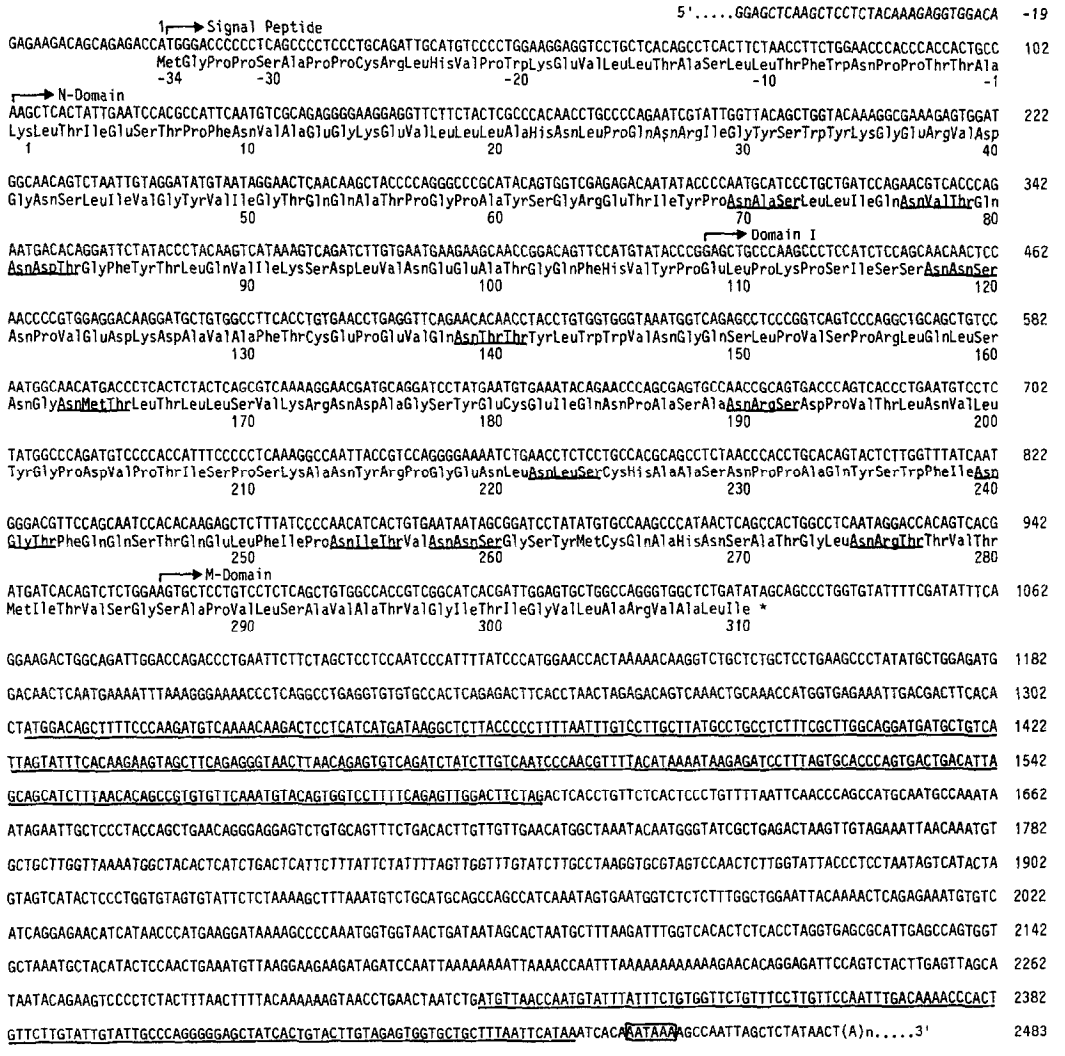


Fig.2 Nucleotide and amino acid sequence derived from NCA cDNA clones λNCA11 and λNCA15. Amino acids are numbered so that N-terminal sequence coincides with that reported for NCA (10,14); residues of the signal peptide are indicated by negative numbers. The starts of signal peptide and each domain are shown by ⌐→; possible N-glycosylation sites by underlines. The translational stop codon for NCA is shown by *. Long open reading frames other than that for NCA are underlined. AATAAA sequence is boxed. Nucleotide sequencing on λNCA15 insert was done by two method:1) EcoRI restriction fragments of the insert were recloned into pUC19 and sequenced by the chain termination method (22) using general and reverse primers; 2) appropriate restriction fragments were recloned into M13mp18 or M13mp19, and sequenced by the chain termination method. λNCA11 insert was analysed by chain termination method on EcoRI fragments recloned into pUC19 using general and reverse primers.

initiative methionine determined earlier on the sequence of the genomic clone (10,14). The 5'-untranslated sequence of 50 residues does not have any in-frame stop codon which, in the case of genomic sequence, is located at the position corresponding to -66 to -64 of this figure (10,14). The location of the cap site is not known at this stage and is left for future study. The

Table I. Comparison of the primary structure of CEA and NCA

Domain	No. of Amino Acid		No. of Asn-X-Thr/Ser		No. of Met		Homology(%)*	
	CEA	NCA	CEA	NCA	CEA	NCA	base	A.A.
S	34**	34	0	0	1	1	83.3	73.5
N	108	108	2	3	0	0	93.2	88.9
I	178	178	11	9	0	3	90.3	83.7
II	178	-	7	-	0	-	86.0	76.4
III	178	-	8	-	0	-	83.0	73.6
M	26	24	0	0	1	0	78.2	65.4
N to M	668	310	28	12	1	3		
(M.W.)	72,890	33,524						

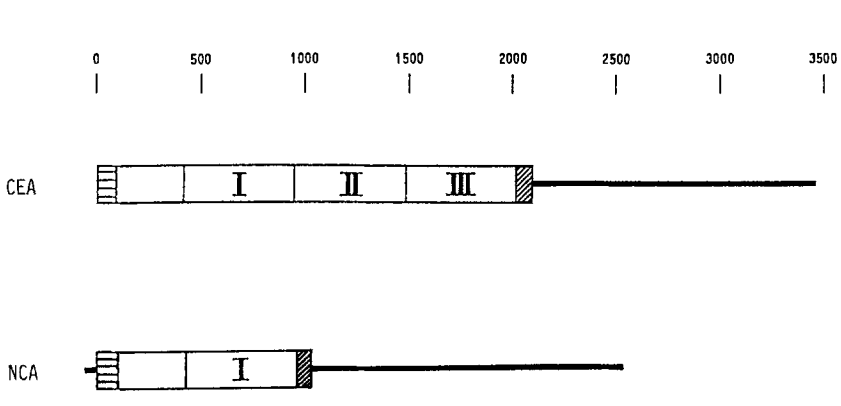
* NCA-I domain is compared with each of CEA-I, -II and -III domain.
** N-terminal four residues upstream the known 30 residues (8) have been deduced as MetGluSerPro (our unpublished results and (24)).

3'-untranslated region (excluding the poly(A) sequence) is 1451-bp long with the consensus poly(A) signal sequence, AATAAA, located 20 nucleotides upstream from the poly(A) addition site.

The peptide encoded consists of 344 amino acids, a 34-residue signal sequence and the following 108-residue N-terminal domain being exactly identical to those deduced from genomic sequence except for the 108th residue of the N-domain whose codon is divided by an intron and therefore could not be determined on the genomic sequence (10,14). The N-domain is followed by a 202-residue sequence, the first 178 residues being very homologous to the 178-residue repetitive domains of the CEA and C-terminal 24 residues being similar to the 26-residue putative M-domain of the CEA. The comparison of the features of primary structure of NCA deduced presently and CEA (8) is depicted in Table 1 and in Fig.3. The M-domains are compared as are aligned in Fig.3 rather than being aligned in other possible ways, for it allows the maximal homology for nucleotide sequences, without changing the length of the preceding domain of NCA.

It has been noted that the three repetitive domains of CEA are strikingly identical in their length and the location of four Cys residues (8). No less striking is the apparent strict conservation of the length of the N-terminal and 178-residue domains and the location of the Cys residues between NCA and CEA, although ambiguity surrounding the domain boundary does not allow one to determine the position of the boundary indisputably. For instance, N-terminal domain of NCA could be either 107- or 108-residue long because the codon encoding the 108th amino acid is interrupted by the insertion of an intron after the first nucleotide (10).

The sequence homology is the highest between the N-terminal domains (Table 1) and interestingly, the homology of the 178-residue domain of NCA



B

Repetitive Domain

NCA	E L P K P S I S S **N** N S N P V E D K D A V A F T C E P E V Q **N** T T Y L W W V N G	109-148
I	E L P K P S I S S **N** N S K P V E D K D A V A F T C E P E T Q D A T Y L W W V N **N**	109-148
II	E P P K P F I T S **N** N S N P V E D E D A V A L T C E P E I Q **N** T T Y L W W V N **N**	287-326
III	E L P K P S I S S **N** N S K P V E D K D A V A F T C E P E A Q **N** T T Y L W W V N G	465-504

NCA	Q S L P V S P R L Q L S N G **N** M T L T L L S V K R N D A G S Y E C E I Q N P A S	149-188
I	Q S L P V S P R L Q L S N G **N** R T L T L F **N** V T R **N** D T A S Y K C E T Q N P V S	149-188
II	Q S L P V S P R L Q L S N D **N** R T L T L L S V T R N D V G P Y E C G I Q N E L S	327-366
III	Q S L P V S P R L Q L S N G **N** R T L T L F **N** V T R N D A R A Y V C G I Q N S V S	505-544

NCA	A **N** R S D P V T L N V L Y G P D V P T I S P S K A N Y R P G E N L **N** L S C H A A	189-228
I	A R R S D S V T L N V L Y G P D A P T I S P L **N** T S Y R S G E N L **N** L S C H A A	189-228
II	V D H S D P V T L N V L Y G P D D P T I S P S Y T Y Y R P G V **N** L S L S C H A A	367-406
III	A **N** R S D P V T L D V L Y G P D T P I T S P P D S S Y L S G A N L **N** L S C H S A	545-584

NCA	S N P P A Q Y S W F I **N** G T F Q Q S T Q E L F I P **N** I T V **N** N S G S Y M C Q A H	229-268
I	S N P P A Q Y S W F V **N** G T F Q Q S T Q E L F I P **N** I T V **N** N S G S Y T C Q A H	229-268
II	S N P P A Q Y S W L T D G N T Q Q H T Q E L F I S **N** I T E K N S G L Y T C Q A **N**	407-446
III	S N P S P Q Y S W R I N G I P Q Q H T Q V L F I A K I T P N N **N** G T Y A C F V S	585-624

NCA	N S A T G L **N** R T T V T M I T V S G	269-286
I	N S D T G L **N** R T T V T T I T V V A	269-286
II	N S A S G H S R T T V K T I T V S A	447-464
III	N L A T G R **N** N S I V K S I T V S A	625-642

M-Domain

NCA	- - S A P V L S A V A T V G I T I G V L A R V A L I	287-310
CEA	S G T S P G L S A G A T V G I M I G V L V G V A L I	643-668

Fig.3 (A) Schematic structure of the cDNA of CEA and NCA. Open boxes and horizontal bars depict protein coding and non-coding regions, respectively. ▤, signal peptide; □, N-domain; [I], [II], [III], domain I, II, III; ▨, M-domain. (B) Comparison of amino acid sequences of the domains of NCA and CEA. Comparison of N-terminal domains has been reported (10) and is omitted here. The numbers of amino acid residues at the first and last ends of each line are given. Residues identical to those of NCA are boxed. * indicates the location of cysteine residue. Bold N is the asparagine residue at the possible N-glycosylation site.

(NCA-I domain) to each of the CEA repetitive domain I, II and III (CEA-I, -II and -III domain) is higher than that between any pair of the repetitive domains of CEA (8), the highest being to CEA-I domain (Table 1). The presence

of three Met residues in NCA-I domain in addition to the presence of Ala instead of Val at position 21 distinguish the peptide clearly from CEA. The number and location of the possible N-glycosylation sites are comparable to those of the CEA domains (Table 1 & Fig.3).

Another feature to be noted is that despite the high homology of the nucleotide sequences of the coding regions between the NCA and the CEA cDNAs, the 3'-untranslated regions are entirely different except for the first 42 residues after the stop codon, which are 85.7 % homologous. Unlike the CEA cDNA, the NCA cDNA lacks Alu family sequence but rather contains two long open reading frames in its 3'-untranslated region which could be translated into 100- and 43-residue polypeptides (Fig.2). Studies to determine whether the sequences are actually translated in cells or not are now in progress.

DISCUSSIONS

Here we have described the isolation and characterization of NCA cDNA clones. The most remarkable feature revealed by the deduced complete amino acid sequence of NCA is the resemblance and difference of the structure and construction of the domains between NCA and CEA (8). As is summarized schematically in Fig.3A, both proteins are apparently synthesized as a precursor having a signal sequence typical of secretory proteins, the mature form has a 108-residue N-terminal domain followed by one (NCA) or three (CEA) 178-residue domains consisting of subdomains A and B each of which, along with N-domain, appears to have Ig-related structures (9,23). A short C-terminal sequence of 24 to 26 mostly hydrophobic residues following the last 178-residue domain probably serves as a membrane anchoring site (8,12) directly or indirectly (23), therefore designated M-domain (9). Each corresponding domain shows extensive homology between the two proteins as is described in RESULTS and summerized in Table 1.

These results, together with the previous observations that the genomic sequences encoding a part of signal peptide and N-domain of NCA (10,14), homologues of subdomain A, and those of subdomain B comprise separate exons (9) and that N-domain (9,23) and subdomains A and B (9,11,14,23) have sequence similarity to the members of Ig supergene family, suggest that the ancestral genes for these three domains are derived from a primodial Ig gene and diversified through various events e.g. duplication, deletion, recombination, conversion, exon shuffling, point mutation, etc., to develop into present-day CEA gene family (11).

Contrary to the close similarity of the nucleotide sequences of the coding regions, those of 3'-noncoding regions are quite different between the CEA and NCA cDNAs, implying differential regulation of the expression of CEA

and NCA. From the practical point of view, this will enable one to prepare probes to study the expression of CEA and NCA in tissues at various developmental stages and under various physiological conditions.

REFERENCES

1. Gold, P. & Freedman, S. O. (1965) J. Exp. Med. 121, 439-462
2. Krupey, J., Gold, P. & Freedman, S. O. (1968) J. Exp. Med. 128, 387-398
3. von Kleist, S., Chavanel, G. & Burtin, P. (1972) Proc. Natl. Acad. Sci. USA. 69, 2492-2494
4. Burtin, P., Chavanel, G. & Hirsch-Marie, H. (1973) J. Immunol. 111, 1926-1928
5. Kessler, M. J., Shively, J. E., Pritchard, D. G. & Todd, C. W. (1978) Cancer Res. 38, 1041-1048
6. Kuroki, M., Koga, Y. & Matsuoka, Y. (1981) Cancer Res. 41, 713-720
7. Svenberg, T. (1976) Int. J. Cancer, 17, 588-596
8. Oikawa, S. Nakazato, H. & Kosaki, G. (1987) Biochem. Biophys. Res. Commun. 142, 511-518
9. Oikawa, S., Imajo, S., Noguchi, T., Kosaki, G. & Nakazato, H. (1987) Biochem. Biophys. Res. Commun. 144, 634-642
10. Oikawa, S., Kosaki, G. & Nakazato, H. (1987) Biochem. Biophys. Res. Commun. 146, 464-469
11. Paxton, R. J., Mooser, G., Pande, H., Lee, T. D., & Shively, J. E. (1987) Proc. Natl. Acad. Sci. USA. 84, 920-924
12. Zimmermann, W., Ortlieb, B., Friedrich, R. & von Kleist, S. (1987) Proc. Natl. Acad. Sci. USA. 84, 2960-2964
13. Kamarck, M. E., Elting, J. J., Hart, J. T., Goebel, S. J., Rae, P. M. M. Nothdurft, M. A., Nedwin, J. J. & Barnett, T. R. (1987) Proc. Natl. Acad. Sci. USA. 84, 5350-5354
14. Thompson, J. A., Pande, H., Paxton, R. J., Shively, L., Padma, A., Simmer R. L., Todd, C. W., Riggs, A. D. & Shively, J. E. (1987) Proc. Natl. Acad. Sci. USA. 84, 2965-2969
15. Akagi, T. & Kimoto, T. (1976) Jpn. J. Cancer Res. (Gann), 67, 483-492
16. Oikawa, S., Imai, M., Ueno, A., Tanaka, S., Noguchi, T., Nakazato, H., Kangawa, K., Fukuda, A., & Matsuo, H. (1984) Nature, 309, 724-726
17. Nambu, J. R., Taussig, R., Mahon, A. C. & Scheller, R. H. (1983) Cell, 35, 47-56
18. Young, R. A. & Davis, R. W. (1983) Proc. Natl. Acad. Sci. USA. 80, 1194-1198
19. Watson, C. J. & Jackson, J. F. (1985) In DNA cloning (D. M. Glover, ed.) Vol.I, pp79-88. IRL Press, Oxford, Washington, DC.
20. Ichiki, S., Kuroki, Ma., Kuroki, Mo., Koga, Y. & Matsuoka, Y. (1986) Jpn. J. Cancer Res. (Gann), 77, 462-472
21. William, I., Gitschier, W. J., Lasky, L. A., & Lawn, R. M. (1985) Proc. Natl. Acad. Sci. USA. 82, 1585-1588
22. Sanger, F., Nicklen, S. & Coulsen, A. R. (1975) Proc. Natl. Acad. Sci. USA. 72, 3961-3965
23. Williams, A. F. (1987) Immunol. Today, 8, 298-303
24. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuchs, A. & Stanners, C. P. (1987) Mol. Cell. Biol. 7, 3221-3230