Journal of Neurochemistry
Raven Press, Ltd., New York

A Protease Inhibitor of the Serpin Family Is a Major Protein in Carp Perimeningeal Fluid: II. cDNA Cloning, Sequence Analysis, and *Escherichia coli* Expression

*Chang-Jen Huang, †Ming-Shyue Lee, *‡Fore-Lien Huang, and †Geen-Dong Chang

**Institute of Biological Chemistry, Academia Sinica; and †Graduate Institute of Biochemical Sciences and ‡Department of Zoology, National Taiwan University, Taipei, Taiwan*

Abstract: A cDNA clone, *pCP9*, has been isolated from a common carp liver cDNA library by immunoscreening with polyclonal antiserum raised against purified bighead carp α_1-antitrypsin. This clone is 1,396 bp in length and has an open reading frame encoding a protein of 410 amino acid residues. The deduced amino acid sequence shows moderate homology to human α_1-antitrypsin (38%), guinea pig contrapsin (35%), human α_1-antichymotrypsin (34%), and human proteinase C inhibitor (31%), all members of the serine protease inhibitor (serpin) family. To confirm further that the cDNA clone was derived from the authentic carp α_1-antitrypsin gene, the presumptive mature protein of *pCP9* was expressed in *Escherichia coli*. The molecular mass of the recombinant protein matched that predicted from the nucleotide sequence. This recombinant protein, which was recognized by antiserum against native α_1-antitrypsin, was capable of formation of serpin–enzyme complexes with trypsin, chymotrypsin, and elastase. Therefore, we conclude that the protein encoded by the *pCP9* clone is indeed carp α_1-antitrypsin. Expression of α_1-antitrypsin in brain was confirmed by reverse transcription and polymerase chain reaction performed on mRNA derived from both common carp and bighead carp brain. **Key Words:** Carp—Protease inhibitor—Serpin—Perimeningeal fluid.
J. Neurochem. **64,** 1721–1727 (1995).

The serpins (serine proteinase inhibitors) are a family of glycoproteins that include members involved in the control of blood coagulation, fibrinolysis, complement activation, and inflammation processes (Carrell and Boswell, 1986). Many of them are plasma protease inhibitors synthesized and secreted by the liver, such as α_1-antitrypsin (Long et al., 1984), contrapsin (Suzuki et al., 1991), α_1-antichymotrypsin (Chandra et al., 1983), complement C_1 inhibitor (Bock et al., 1986), and protein C inhibitor (Suzuki et al., 1987). However, a few of them are synthesized in other tissues, such as elastase inhibitor (Remold-O'Donnell et al., 1992), α_1-antitrypsin (Perlmutter et al., 1985; Perlmutter and Punsal, 1988), and complement C_1 inhibitor (Bensa et al., 1983) from monocytes, protease nexin from glial cells (Sommer et al., 1987; Van Nostrand et al., 1989), and plasminogen activator inhibitor from endothelial cells (Ny et al., 1986).

We have purified a 62-kDa protease inhibitor (p62) from bighead carp perimeningeal fluid (PMF) (Huang et al., 1995). p62 resembles mammalian α_1-antitrypsin in many respects. The protein forms complexes with and inhibits bovine trypsin, bovine chymotrypsin, and porcine pancreatic elastase. In addition, the protease inhibitor is a glycoprotein whose carbohydrate moiety can be removed by endoglycosidase F. Because α_1-antitrypsin in mammals is synthesized and secreted from the liver, attempts were made to isolate cDNA clones encoding this protease inhibitor from a carp liver cDNA library. We obtained two cDNAs whose expression products were recognized by a polyclonal antiserum against p62. Evidence is provided to support the proposal that one of these clones encodes functional p62.

MATERIALS AND METHODS

Enzymes and chemicals

Restriction enzymes, the Klenow fragment of DNA polymerase, and T4 DNA ligase were purchased from either Promega (Madison, WI, U.S.A.) or Boehringer Mannheim (Mannheim, Germany). Radiolabeled compounds were obtained from Amersham (Amersham, U.K.). Sources of other chemicals were described in our preceding article (Huang et al., 1995).

Received April 15, 1994; revised manuscript received August 17, 1994; accepted August 31, 1994.

Address correspondence and reprint requests to Dr. G.-D. Chang at Graduate Institute of Biochemical Sciences, National Taiwan University, P.O. Box 23-106, Taipei 10098, Taiwan.

The sequence data in this article have been submitted to the EMBL/Genbank Data Libraries under the accession numbers of L08689 and L27172.

Abbreviations used: PBS, phosphate-buffered saline; PCR, polymerase chain reaction; PMF, perimeningeal fluid; RT, reverse transcription; SDS, sodium dodecyl sulfate.

General methods in molecular biology

Standard procedures in molecular biology (Sambrook et al., 1989) were used for preparation of plasmid DNA, restriction enzyme digestion, DNA agarose gel electrophoresis, DNA ligation, and bacterial transformation.

Immunoscreening of a carp liver cDNA library

A carp liver (*Cyprinus carpio*) cDNA library prepared from poly(A)-enriched RNA by unidirectional insertion of cDNA into λ-ZAP II (Short et al., 1988) was purchased from Stratagene (La Jolla, CA, U.S.A.). To screen the library, λ-ZAP II phages were plated at a density of 5×10^4 plaques per agar plate (150 mm i.d.). A total of 10 plates was initially screened. After incubation for 3 h at 42°C, the plates were overlaid with nitrocellulose filters (pore size, 0.45 μm; Micron Separations, Westboro, MA, U.S.A.) that had been impregnated with 10 m*M* isopropyl-β-D-thiogalactopyranoside. Incubation was continued overnight at 37°C. The filters were then removed, washed with phosphate-buffered saline (PBS) at room temperature, and blocked in 3% skim milk in PBS for 1 h at room temperature. Following blocking, filters were probed with a polyclonal antiserum specific for carp α_1-antitrypsin (p62) at a 1:500 dilution in PBS containing 3 mg/ml of bovine serum albumin, 1 m*M* EDTA, and 0.4% Triton X-100 at 4°C for 16 h. The filters were then washed three times at room temperature in PBS and incubated with horseradish peroxidase-conjugated anti-guinea pig IgG (Sigma, St. Louis, MO, U.S.A.) for 1 h at room temperature. After three washes with PBS, the immune complexes were incubated in PBS containing 0.2 mg/ml of diaminobenzamidine and revealed by adding H_2O_2 (10 μl in 10 ml of PBS). Phages demonstrating stronger signals were isolated for secondary screening. Five positive clones were isolated after screening 5×10^5 plaque-forming units, and pBluescript plasmids containing cDNA inserts were obtained by in vivo excision according to the protocols provided by Stratagene.

DNA sequencing and sequence analysis

Plasmids carrying cDNA inserts were sequenced in both directions using T7 polymerase (U.S. Biochemical Corp., Cleveland, OH, U.S.A.) and the dideoxy chain termination method (Sanger et al., 1977). Programs from IntelliGenetics (Mountain View, CA, U.S.A.) were used to analyze the nucleotide sequences, and the SEQ program was used to deduce the amino acid sequence. A search for related sequences in GenBank, EMBL, SWISS-PROT, and Protein Identification Resource was carried out with an IFIND program using the FASTA algorithm of Pearson and Lipman (1988). Alignment of the deduced amino acid sequences with those of mammalian serpins was accomplished with the GENALIGN program using the FASTP algorithm of Lipman and Pearson (1985).

Expression and purification of recombinant α_1-antitrypsin

To express carp α_1-antitrypsin in *Escherichia coli*, the polymerase chain reaction (PCR) was carried out to amplify the DNA fragment containing the putative mature protein using two primers: F1, (5′-GGATCCGATCACTACCACCATCTCCAC-3′) and R1 (5′-TCTGTTTGGAGTTGGACCTCA-3′). PCR was performed on the plasmid DNA (pCP9) with the two primers and *Taq* polymerase by standard procedures (Mullis and Faloona, 1987). An amplified DNA fragment of 1.2 kb was isolated from an agarose gel, digested with *Bam*HI, and cloned into the *Bam*HI–*Sma*I sites of the expression vector pQE30, which was purchased from QIAGEN (Chatsworth, CA, U.S.A.). The sequence of the PCR product was verified by DNA sequencing. This construction introduced 12 amino acids (MRGSHHHHHHGS) into the amino-terminus of the mature protein. The resulting plasmid was transformed into *E. coli* strain JM109, and the resulting transformants were grown at 37°C overnight. The overnight culture was diluted 50-fold and regrown at 37°C for 1 h. Isopropyl-β-D-thiogalactopyranoside was then added to a final concentration of 1 m*M*, and the culture was continued for 5 h longer. The culture medium was centrifuged at 4,000*g* for 10 min, and the pellet was stored at −70°C for subsequent purification. Purification of His-tagged recombinant α_1-antitrypsin by Ni^{2+}-nitrilotriacetate resin (QIAGEN) was accomplished according to the procedures provided by the supplier.

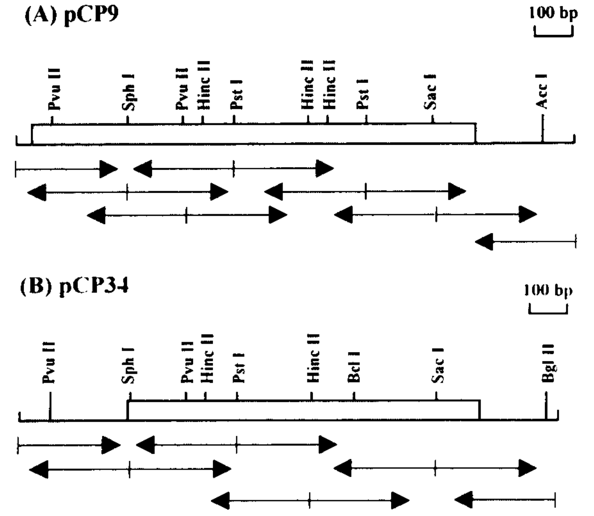


FIG. 1. Restriction enzyme maps of cDNA clones *pCP9* and *pCP34* and DNA sequencing strategy. The coding regions are indicated by the open boxes. Arrows indicate the direction and lengths of determined DNA sequences.

Cyanogen bromide cleavage

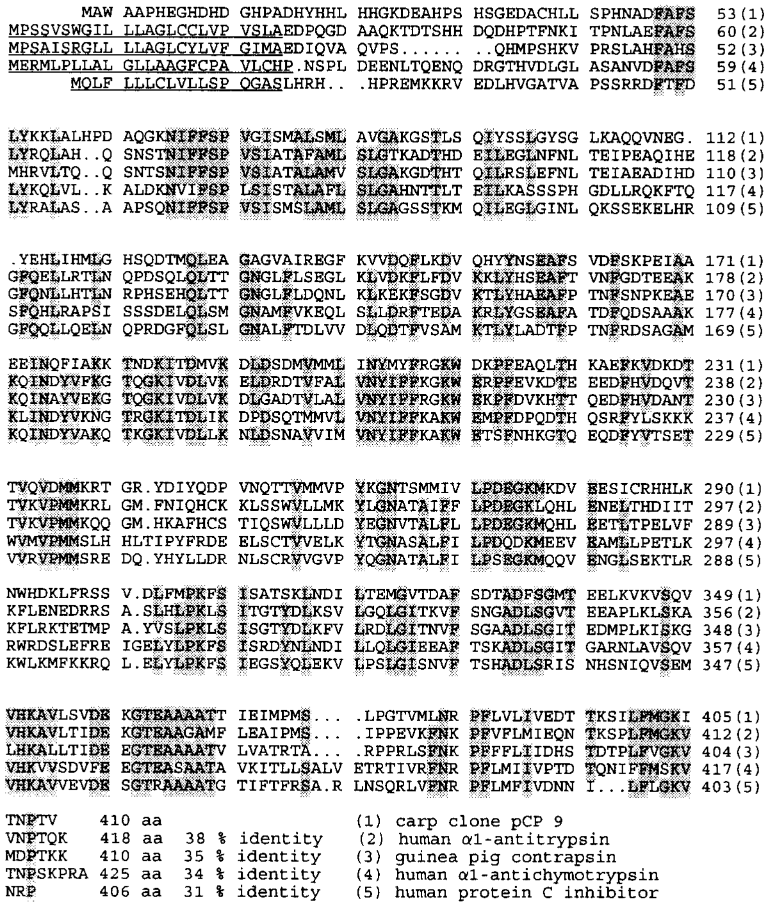
Purified bighead carp α_1-antitrypsin (100 μg) was reduced by 10 m*M* dithiothreitol in 8 *M* urea/50 m*M* Tris-HCl (pH 8.0) for 60 min at 37°C. Neutralized iodoacetate was added to a final concentration of 15 m*M*, and the sample was incubated for an additional 20 min. The sample was then dialyzed against distilled water, concentrated, and treated with cyanogen bromide (10 mg/ml) in 100 μl of 0.1 *M* HCl for 24 h at 25°C. Residual cyanogen bromide was evaporated in a Speed-Vac concentrator. The peptides were then dissolved in 0.1% trifluoroacetic acid and separated on a C_{18} HPLC column with a linear gradient of CH_3CN. Three peaks were chosen for amino acid sequencing. Automated amino acid sequencing was performed in a gas vapor sequenator (model 471A; Applied Biosystems, Forest City, CA, U.S.A.) equipped with an on-line phenylthiohydantoin analyzer.

Reverse transcription (RT) and PCR

Total RNA was isolated from the whole brain of common carp and bighead carp (*Aristichthys nobilis*) using a rapid acid-guanidinium procedure (Chomczynski and Sacchi,

FIG. 2. Nucleotide and deduced amino acid sequences of cDNA clones *pCP9* and *pCP34*. The nucleotide sequences of both clones are shown in the upper two lines, whereas deduced amino acid sequences are shown in the one-letter code below the nucleotide sequences. For *pCP34*, only nucleotides (in lowercase letters) and amino acid residues different from those of *pCP9* are indicated. The initiation and termination codons are underlined. Two potential *N*-glycosylation sites are double underlined. Amino acid residues that match peptide sequences derived from purified p62 are underlined with thick lines.

```
34 --------------------c-------------c---------tcaaa-c--------c--c---   66
 9 GGCACGAGGTGATTGCTGCCTTGCTGGTAGCAAATGGCCTGGGCCGCGCCTCACGAAGGTCATGAC   66
 9                                  M  A  W  A  A  P  H  E  G  H  D     11

34 --c-g--a-a--a-----------c-----------------------------c-----------  132
 9 CATGACGGCCACCCAGCTGATCACTACCACCATCTCCACCACGGGAAGGACGAAGCCCACCCCAGC  132
 9 H  D  G  H  P  A  D  H  Y  H  H  L  H  H  G  K  D  E  A  H  P  S     33

34 ---g-------------------ct-----g----------t------------------------  198
 9 CACAGTGGGGAGGATGCCTGCCATCTGCTTTCTCCACACAACGCTGACTTTGCCTTCTCCCTCTAC  198
 9 H  S  G  E  D  A  C  H  L  L  S  P  H  N  A  D  F  A  F  S  L  Y     55

34 ------------t-----t-a------a-------------------------a---------g--  264
 9 AAGAAACTTGCGCTCCATCCTGATGCCCAGGGCAAGAACATTTTCTTCTCCCCGGTCGGTATCTCA  264
 9 K  K  L  A  L  H  P  D  A  Q  G  K  N  I  F  F  S  P  V  G  I  S     77

34 -----------------------------------c-----a--t-----g---------------  330
 9 ATGGCTTTGAGCATGCTGGCTGTAGGTGCCAAGGGTAGCACTCTATCACAAATATACAGCAGTCTG  330
 9 M  A  L  S  M  L  A  V  G  A  K  G  S  T  L  S  Q  I  Y  S  S  L     99
34 -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -     22

34 --------t------c-----tca--------------t-----------------------a---  396
 9 GGTTACAGCGGGTTGAAGGCTCAGCAGGTCAATGAGGGCTATGAGCACTTGATCCACATGCTTGGC  396
 9 G  Y  S  G  L  K  A  Q  Q  V  N  E  G  Y  E  H  L  I  H  M  L  G    121
34 -  -  -  -  -  Q  -  S  -  -  -  -  -  -  -  -  -  -  -  -  -  -     44

34 -------g---ag-------------------c---------------------------------  462
 9 CACAGTCAGGACACCATGCAGCTGGAGGCAGGTGCTGGTGTGGCCATCAGAGAAGGCTTCAAAGTG  462
 9 H  S  Q  D  T  M  Q  L  E  A  G  A  G  V  A  I  R  E  G  F  K  V    143
34 -  -  R  E  A  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -     66

34 --------------------t--------------t--------g---------------------  528
 9 GTTGACCAGTTCCTGAAGGACGTTCAGCACTACTACAACAGCGAAGCCTTCAGCGTTGACTTCTCC  528
 9 V  D  Q  F  L  K  D  V  Q  H  Y  Y  N  S  E  A  F  S  V  D  F  S    165
34 -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -     88

34 -----------t-----------------t--------t--------g-----------g------  594
 9 AAGCCTGAAATCGCTGCAGAAGAGATTAACCAGTTCATCGCCAAGAAAACCAATGACAAAATAACC  594
 9 K  P  E  I  A  A  E  E  I  N  Q  F  I  A  K  K  T  N  D  K  I  T    187
34 -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -    110

34 a----------------------c--c-c-------------------------------------  660
 9 GACATGGTGAAGGACCTGGACTCTGATATGGTGATGATGCTGATTAACTATATGTACTTCAGAGGG  660
 9 D  M  V  K  D  L  D  S  D  M  V  M  M  L  I  N  Y  M  Y  F  R  G    209
34 I  -  -  -  -  -  -  -  -  T  -  -  -  -  -  -  -  -  -  -  -  -    132

34 --------------------t-----------c-----------c------------g--------  726
 9 AAGTGGGATAAGCCATTTGAAGCACAACTGACTCACAAAGCTGAGTTCAAAGTGGACAAGGACACC  726
 9 K  W  D  K  P  F  E  A  Q  L  T  H  K  A  E  F  K  V  D  K  D  T    231
34 -  -  -  -  -  -  D  -  -  -  -  -  -  -  D  -  -  -  -  E  -  -    154


34 -----------------------------------t------------------------------  792
 9 ACCGTGCAAGTTGACATGATGAAAAGAACCGGCCGCTATGACATCTATCAAGACCCTGTCAACCAA  792
 9 T  V  Q  V  D  M  M  K  R  T  G  R  Y  D  I  Y  Q  D  P  V  N  Q    253
34 -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -    176

34 --------------------c---------------------------a--t----------t---  858
 9 ACTACGGTCATGATGGTGCCATACAAAGGCAATACTTCCATGATGATCGTTCTTCCTGATGAAGGG  858
 9 T  T  V  M  M  V  P  Y  K  G  N  T  S  M  M  I  V  L  P  D  E  G    275
34 -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  I  F  -  -  D  -    198

34 -----------gc--------------a----a-----c---------------------------  924
 9 AAGATGAAGGATGTTGAAGAATCCATCTGCAGGCACCATCTTAAGAACTGGCATGATAAACTCTTC  924
 9 K  M  K  D  V  E  E  S  I  C  R  H  H  L  K  N  W  H  D  K  L  F    297
34 -  -  -  E  L  -  -  -  -  S  -  -  -  -  -  -  -  -  -  -  -  -    220

34 --------------------t--------------a---a-------t-----------a-g----  990
 9 AGAAGCTCTGTGGACCTGTTCATGCCCAAGTTCTCCATCTCTGCAACGTCCAAACTGAATGACATT  990
 9 R  S  S  V  D  L  F  M  P  K  F  S  I  S  A  T  S  K  L  N  D  I    319
34 -  -  -  -  -  -  -  -  -  -  -  -  -  T  -  -  -  -  -  K  G  -    242

34 ---gag--c---------------------g----------------a------t----------- 1056
 9 CTGACTGAAATGGGAGTGACTGATGCATTCAGTGACACAGCAGATTTCTCTGGGATGACAGAAGAG 1056
 9 L  T  E  M  G  V  T  D  A  F  S  D  T  A  D  F  S  G  M  T  E  E    341
34 -  E  D  -  -  -  -  -  -  -  G  -  -  -  -  L  -  -  L  -  -  -    264

34 g-------c-----------------c-----------------c--t------------------ 1122
 9 CTCAAAGTGAAGGTGTCACAGGTTGTGCATAAGGCAGTCCTCAGTGTGGATGAGAAGGGCACAGAG 1122
 9 L  K  V  K  V  S  Q  V  V  H  K  A  V  L  S  V  D  E  K  G  T  E    363
34 V  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -    286

34 --------t--------g-----a----------------------a---------a--t------ 1188
 9 GCAGCGGCCGCAACCACAATAGAGATCATGCCCATGTCCCTGCCAGGCACTGTGATGCTCAACCGA 1188
 9 A  A  A  A  T  T  I  E  I  M  P  M  S  L  P  G  T  V  M  L  N  R    385
34 -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  D  -  -  I  -  -  -    308

34 -----------------------------------------------------------c--t--- 1254
 9 CCTTTCTTGGTACTGATTGTAGAGGACACCACCAAGAGCATCCTCTTCATGGGCAAGATTACCAAT 1254
 9 P  F  L  V  L  I  V  E  D  T  T  K  S  I  L  F  M  G  K  I  T  N    407
34 -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -    330

34 -------a---------------------------c--gcac-ctgc--aa-gctgct----t-at 1320
 9 CCTACAGTGTGAGGTCCAACTCCAAACAGAACTTGACGAATTTTGAAATTTGAACTTGAAATCTTG 1320
 9 P  T  V  ***                                                         410
34 -  -  E                                                              333

34 -----g----c------a--------a-----gg------taactctt                   1367
 9 TGTATATCTTATTGTATACTGCAATGAATTGATTTGATTCTAATTAAATGAGTCATCATTAAAACA 1386

 9 ATTACACAAC                                                         1396
```



```
      MAW AAPHEGHDHD GHPADHYHHL HHGKDEAHPS HSGEDACHLL SPHNADFAFS  53(1)
MPSSVSWGIL LLAGLCCLVP VSLAEDPQGD AAQKTDTSHH DQDHPTFNKI TPNLAEFAFS  60(2)
MPSAISRGLL LLAGLCYLVF GIMAEDIQVA QVPS...... ..QHMPSHKV PRSLAHFAHS  52(3)
MERMLPLLAL GLLAAGFCPA VLCHP.NSPL DEENLTQENQ DRGTHVDLGL ASANVDFAFS  59(4)
     MQLF LLLCLVLLSP QGASLHRH.. .HPREMKKRV EDLHVGATVA PSSRRDFTFD  51(5)

LYKKLALHPD AQGKNIFFSP VGISMALSML AVGAKGSTLS QIYSSLGYSG LKAQQVNEG. 112(1)
LYRQLAH..Q SNSTNIFFSP VSIATAFAML SLGTKADTHD EILEGLNFNL TEIPEAQIHE 118(2)
MHRVLTQ..Q SNTSNIFFSP VSIATALAMV SLGAKGDTHT QILRSLEFNL TEIAEADIHD 110(3)
LYKQLVL..K ALDKNVIFSP LSISTALAFL SLGAHNTTLT EILKASSSPH GDLLRQKFTQ 117(4)
LYRALAS..A APSQNIFFSP VSISMSLAML SLGAGSSTKM QILEGLGINL QKSSEKELHR 109(5)

.YEHLIHMLG HSQDTMQLEA GAGVAIREGF KVVDQFLKDV QHYYNSEAFS VDFSKPEIAA 171(1)
GFQELLRTLN QPDSQLQLTT GNGLFLSEGL KLVDKFLFDV KKLYHSEAFT VNFGDTEEAK 178(2)
GFQNLLHTLN RPHSEHQLTT GNGLFLDQNL KLKEKFSGDV KTLYHAEAFP TNFSNPKEAE 170(3)
SFQHLRAPSI SSSDELQLSM GNAMFVKEQL SLLDRFTEDA KRLYGSEAFA TDFQDSAAAK 177(4)
GFQQLLQELN QPRDGFQLSL GNALFTDLVV DLQDTFVSAM KTLYLADTFP TNFRDSAGAM 169(5)

EEINQFIAKK TNDKITDMVK DLDSDMVMML INYMYFRGKW DKPFEAQLTH KAEFKVDKDT 231(1)
KQINDYVFKG TQGKIVDLVK ELDRDTVFAL VNYIFFKGKW ERPFEVKDTE EEDFHVDQVT 238(2)
KQINAYVEKG TQGKIVDLVK DLGADTVLAL VNYIFFRGKW EKPFDVKHTT QEDFHVDANT 230(3)
KLINDYVKNG TRGKITDLIK DPDSQTMMVL VNYIFFKAKW EMPFDPQDTH QSRFYLSKKK 237(4)
KQINDYVAKQ TKGKIVDLLK NLDSNAVVIM VNYIFFKAKW ETSFNHKGTQ EQDFYVTSET 229(5)

TVQVDMMKRT GR.YDIYQDP VNQTTVMMVP YKGNTSMMIV LPDEGKMKDV EESICRHHLK 290(1)
TVKVPMMKRL GM.FNIQHCK KLSSWVLLMK YLGNATAIFF LPDEGKLQHL ENELTHDIIT 297(2)
TVKVPMMKQQ GM.HKAFHCS TIQSWVLLLD YEGNVTALFL LPDEGKMQHL EETLTPELVF 289(3)
WVMVPMMSLH HLTIPYFRDE ELSCTVVELK YTGNASALFI LPDQDKMEEV EAMLLPETLK 297(4)
VVRVPMMSRE DQ.YHYLLDR NLSCRVVGVP YQGNATALFI LPSEGKMQQV ENGLSEKTLR 288(5)

NWHDKLFRSS V.DLFMPKFS ISATSKLNDI LTEMGVTDAF SDTADFSGMT EELKVKVSQV 349(1)
KFLENEDRRS A.SLHLPKLS ITGTYDLKSV LGQLGITKVF SNGADLSGVT EEAPLKLSKA 356(2)
KFLRKTETMP A.YVSLPKLS ISGTYDLKFV LRDLGITNVF SGAADLSGIT EDMPLKISKG 348(3)
RWRDSLEFRE IGELYLPKFS ISRDYNLNDI LLQLGIEEAF TSKADLSGIT GARNLAVSQV 357(4)
KWLKMFKKRQ L.ELYLPKFS IEGSYQLEKV LPSLGISNVF TSHADLSRIS NHSNIQVSEM 347(5)

VHKAVLSVDE KGTEAAAATT IEIMPMS... .LPGTVMLNR PFLVLIVEDT TKSILPMGKI 405(1)
VHKAVLTIDE KGTEAAGAMF LEAIPMS... .IPPEVKFNK PFVFLMIEQN TKSPLFMGKV 412(2)
LHKALLTIDE EGTEAAAATV LVATRTA... .RPPRLSFNK PFFFLIIDHS TDTPLFVGKV 404(3)
VHKVVSDVFE EGTEASAATA VKITLLSALV ETRTIVRFNR PFLMIIVPTD TQNIFFMSKV 417(4)
VHKAVVEVDE SGTRAAAATG TIFTFRSA.R LNSQRLVFNR PFLMFIVDNN I...LFLGKV 403(5)

TNPTV    410 aa                 (1) carp clone pCP 9
VNPTQK   418 aa  38 % identity  (2) human α1-antitrypsin
MDPTKK   410 aa  35 % identity  (3) guinea pig contrapsin
TNPSKPRA 425 aa  34 % identity  (4) human α1-antichymotrypsin
NRP      406 aa  31 % identity  (5) human protein C inhibitor
```

FIG. 3. Comparison of the deduced amino acid sequence of clone *pCP9* with other members of the serpin family. Known signal peptides are underlined. Gaps introduced for maximal alignment are indicated with dots. Amino acid positions are indicated at the end of each line. Shaded residues indicate amino acids that are identical in at least four proteins. The numbers at the end of each sequence show the percentage of identity between a given serpin and carp α_1-antitrypsin.

1987). Total cellular RNA (50 μg) was incubated at 65°C for 5 min in a buffer containing 50 m*M* Tris-HCl, 75 m*M* KCl, 3 m*M* $MgCl_2$, 10 m*M* dithiothreitol, 2 units of RNasin, and 1.25 m*M* deoxynucleotide triphosphates (dGTP, dATP, dTTP, and dCTP) and then kept on ice. Two hundred units of Superscript Moloney murine leukemia virus reverse transcriptase (GIBCO BRL, Gaithersburg, MD, U.S.A.) and 1 μg of oligo-dT_{12-18} primer were added and incubated at 37°C for 1 h. The reaction was then stopped by incubation at 95°C for 5 min.

Two sets of degenerated primers were designed according to the highly conserved amino acid sequence of the serpin family (see Fig. 3). They were 5′CA(T/C)AA(T/C)GCNGA(T/C)TT(T/C)GCNTT (F1), 5′GGNAA(A/G)TGGGANAA(A/G)CCNTT (F2), 5′AANGG(T/C)TTNTCCCA(T/C)TTNCC (R1), and 5′GCNGC(T/C)TCNGTNCC(T/C)TT(T/C)TC. PCR was carried out by addition of 20 μl of the heat-treated RT mixture and 80 μl of PCR buffer containing 25 m*M* Tris-HCl, 37.5 m*M* KCl, 1.5 m*M* $MgCl_2$, 5 m*M* dithiothreitol, 0.25 m*M* deoxynucleotide triphosphates, 0.25 μ*M* primers F1/R1 or F2/R2, and 2.5 units of *Taq* polymerase (Promega). Forty cycles of reaction was performed in a thermocylcer (Hybaid Ltd., Middlesex, U.K.) by the following program: 94°C for 1 min, 50°C for 1 min, and 72°C for 3 min. The PCR products of 510 (for primers F1/R1) and 475 bp (for primers F2/R2) were isolated from an agarose gel and subcloned into the *Sma*I site of the pUC19 vector (Yanisch-Perron et al., 1985).

Other analytical methods were essentially the same as described in our preceding article (Huang et al., 1995).

RESULTS

cDNA cloning

After screening 5 × 10^5 plaques with polyclonal antiserum raised against purified carp α_1-antitrypsin, five positive clones were isolated. On the basis of their restriction enzyme maps and partial sequences, two clones, *pCP9* and *pCP34*, were chosen for further characterization. The restriction map and sequencing strategy of the two clones are shown in Fig. 1.

The nucleotide sequences and the deduced amino acid sequences of both cDNA clones, *pCP9* and *pCP34*, are shown in Fig. 2. *pCP9* and *pCP34* have inserts of 1,396 and 1,367 bp, respectively. The *pCP9* clone contains 33 bp of the 5′-untranslated region, an open reading frame of 1,233 bp, and 130 bp of the 3′-untranslated region. A putative initiating ATG codon, which agrees with Kozak's rule (Kozak, 1987), is located at nucleotide 34. The open reading frame is predicted to encode a protein of 410 amino acids with a calculated molecular mass of 46,330 Da.

The nucleotide sequence of *pCP34* is similar to that of *pCP9* except for several point mutations. One point

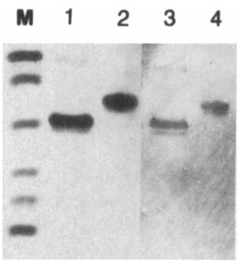


FIG. 4. SDS-polyacrylamide gel electrophoresis and western blot of recombinant and native α_1-antitrypsin. Coomassie Brilliant Blue R-250 staining of the gel is shown in lanes 1 and 2 (0.5 μg), whereas the western blot using an antiserum against α_1-antitrypsin is shown in lanes 3 and 4 (0.1 μg). M, markers of 94, 75, 45, 28, 22, and 18 kDa; lanes 1 and 3, purified recombinant protein; lanes 2 and 4, purified native p62.

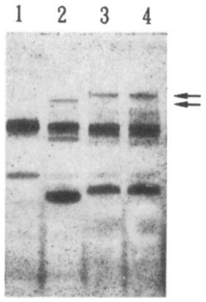


FIG. 5. Complex formation between recombinant α_1-antitrypsin and proteases. Recombinant His-p46 (0.3 μg) and protease (0.2 μg) in 5 μl of PBS were mixed and incubated at 4°C for 1 min. The mixtures were then analyzed by SDS-gel electrophoresis and visualized by silver staining. Lane 1, His-p46; lane 2, His-p46/trypsin; lane 3, His-p46/chymotrypsin; lane 4, His-p46/elastase. The positions of complexes are indicated by arrows.

mutation, T to C, occurs at the presumptive initiation ATG codon. As a result, the *pCP34* clone contains 264 bp of the 5′-untranslated region, an open reading frame of 1,002 bp, and 101 bp of the 3′-untranslated region. The open reading frame encodes a protein of only 333 amino acids with a predicted molecular mass of 37,629 Da. The deduced sequence of *pCP34* corresponds to amino acid residues 78–410 of *pCP9* with 93% identity at the amino acid level and 90% identity at the nucleotide level. Two potential *N*-glycosylation sites were identified in the deduced amino acid sequences of both cDNA clones.

Sequence homology

The deduced amino acid sequence of *pCP9* shows moderate sequence homology (31–38%) with other members in the serpin family such as human α_1-antitrypsin (Long et al., 1984), guinea pig contrapsin (Suzuki et al., 1991), human α_1-antichymotrypsin (Chandra et al., 1983), and human proteinase C inhibitor (Suzuki et al., 1987) (Fig. 3). The amino-terminal regions, including the signal peptide, are most variable. Stretches of highly conserved sequences were found, especially in the carboxy-terminal regions. A putative junction for the signal peptide and mature protein is located at a similar position as that in human α_1-antitrypsin and guinea pig contrapsin.

Characterization of recombinant protein

Most of the His-tagged recombinant protein was found in inclusion bodies and was thus solubilized with 8 *M* urea. The presence of 8 *M* urea did not affect the binding of recombinant protein to the metal column. Figure 4 shows sodium dodecyl sulfate (SDS)–gel electrophoresis and western blot analysis of the purified recombinant α_1-antitrypsin. A major protein band of ~46 kDa was detected (His-p46). The major band represented ~5% of the total cellular protein (data not shown). A band of lower molecular mass was also observed. This protein appears to be the degradation product of His-p46 because its amount increased during storage (data not shown). To demonstrate that the recombinant α_1-antitrypsin was active, the ability of recombinant protein to form complexes with serine proteases was examined. Figure 5 shows that this recombinant protein formed SDS-resistant complexes not only with trypsin but also with chymotrypsin and elastase. Polyclonal antiserum was also raised against this recombinant α_1-antitrypsin. This antiserum recognized native α_1-antitrypsin in PMF and in other body fluids with the exception of ovarian fluid (Fig. 6).

RT-PCR of brain mRNA

Two sets of primers were designed to perform RT-PCR on mRNA derived from common carp and bighead carp brain. PCR products from both species were identical in length, and only the products from bighead carp were sequenced. The derived amino acid sequence was 84% identical to that of the *pCP9* clone (Fig. 7).

DISCUSSION

We have isolated two cDNA clones, *pCP9* and *pCP34*, encoding common carp α_1-antitrypsin. A methionine codon with a flanking Kozak sequence (Kozak, 1987) is located at position 34 in *pCP9*. Because this methionine precedes the first two consensus sequences found in all serpins, FAFSLY in an α-helix hA domain and NIFFSPV in a β-sheet s6B domain, it is likely that this codon represents the translation initiation site. Therefore, the cDNA clone *pCP9* has an open reading frame encoding a protein of 410 amino acids with a calculated molecular mass of 46,330 Da, which is close to the deduced molecular mass of most members of the serpin family. Stretches of sequences

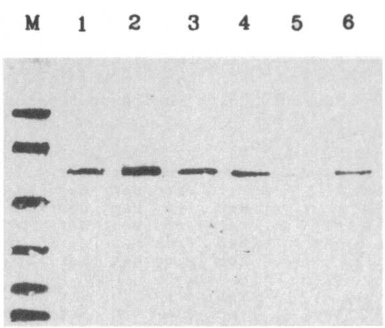
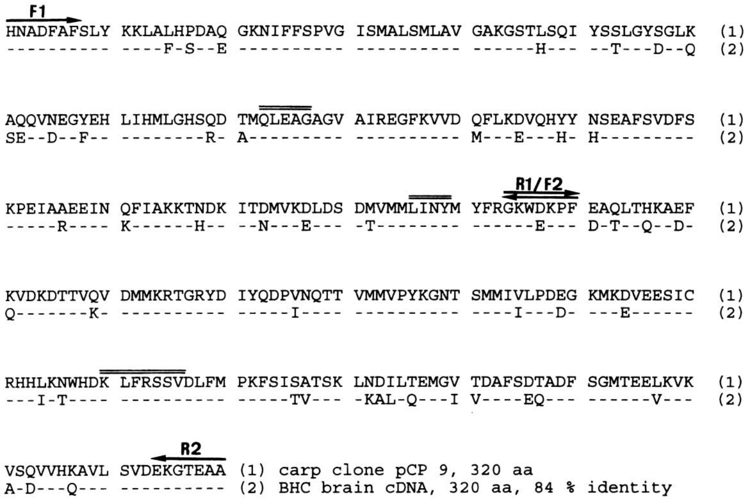


FIG. 6. Body fluid distribution of carp α_1-antitrypsin. Proteins from various body fluids as well as purified α_1-antitrypsin (0.1 μg) were electrophoresed in a SDS-polyacrylamide gel and subjected to western blotting using an antiserum against purified recombinant His-p46. Lanes 1 and 2, two isoforms of carp α_1-antitrypsin (Mono-P-a and Mono-P-b); lanes 3–6, 0.1 μl of serum, 0.1 μl of PMF, 1 μl of ovarian fluid, and 1 μl of milt.



```
F1
HNADFAFSLY KKLALHPDAQ GKNIFFSPVG ISMALSMLAV GAKGSTLSQI YSSLGYSGLK  (1)
---------- ----F-S--E ---------- ---------- ------H--- --T---D--Q  (2)

AQQVNEGYEH LIHMLGHSQD TMQLEAGAGV AIREGFKVVD QFLKDVQHYY NSEAFSVDFS  (1)
SE--D--F-- --------R- A--------- ---------- M---E---H- H---------  (2)

                                                R1/F2
KPEIAAEEIN QFIAKKTNDK ITDMVKDLDS DMVMMLINYM YFRGKWDKPF EAQLTHKAEF  (1)
-----R---- K------H-- --N---E--- -T-------- ------E--- D-T--Q--D-  (2)

KVDKDTTVQV DMMKRTGRYD IYQDPVNQTT VMMVPYKGNT SMMIVLPDEG KMKDVEESIC  (1)
Q-------K- ---------- -----I---- ---------- ----I---D- ---E------  (2)

RHHLKNWHDK LFRSSVDLFM PKFSISATSK LNDILTEMGV TDAFSDTADF SGMTEELKVK  (1)
---I-T---- ---------- -----TV--- -KAL-Q---I V----EQ--- ------V---  (2)

               R2
VSQVVHKAVL SVDEKGTEAA (1) carp clone pCP 9, 320 aa
A-D---Q--- ---------- (2) BHC brain cDNA, 320 aa, 84 % identity
```

FIG. 7. Comparison of deduced partial amino acid sequence of *pCP9* and PCR products from bighead carp brain. Identical amino acid residues are indicated by dashes. Amino acids that are identical to the peptide sequence derived from purified p62 are marked with double lines above the sequence. The positions of PCR primers are indicated by arrows.

that are conserved in all serpins, including pCP9, correspond to the structural features analyzed in detail in α_1-antitrypsin (Huber and Carrell, 1989).

The reactive center of the serpins is located near their carboxy-terminus with well-defined motifs (Carrell and Travis, 1985). Comparison of the sequence of this region also suggests that *pCP9* belongs to the α_1-antitrypsin subfamily (Fig. 8). The amino acid of p62 is homologous with α_1-antitrypsin from other species within this six-amino-acid region (P3–P′3). The above structural similarities strongly suggest that *pCP9* encodes a carp α_1-antitrypsin.

The cDNA clone *pCP9* was originally isolated by immunoscreening a carp liver cDNA library with antiserum against purified p62. We suggest that the gene product encoded by the *pCP9* is identical or similar to p62 based on the following observations. First, expressed protein encoded by the *pCP9* clone can be recognized by the same antiserum used for immunoscreening. Second, this recombinant protein is capable of forming complexes not only with trypsin but also with chymotrypsin and elastase. In addition, polyclonal antiserum raised against this recombinant protein recognized native p62. Finally, stretches of amino acid residues in positions 128–132, 201–204, and 295–301 in *pCP9* were identical to the sequence of peptides derived from purified p62 by cyanogen bromide cleavage. However, two discrepancies arose during comparison of the structures of purified p62 and the *pCP9* clone. First, one peptide fragment, KLFRSSV, was not preceded with a methionine residue. Second, the deglycosylated form of p62 has an apparent molecular mass of 53 kDa, which is 7 kDa larger than the value predicted from the nucleotide sequence.

Reactive site region of Serpins

	P3	P2	P1	P′1	P′2	P′3
Human α1-AT	Ile	**Pro**	**Met**	**Ser**	Ile	**Pro**
Rat α1-AT	Val	**Pro**	**Met**	**Ser**	**Leu**	**Pro**
GP α1-AT	**Met**	**Pro**	**Met**	**Ser**	**Leu**	**Pro**
Mouse α1-AT	Val	**Pro**	**Met**	**Ser**	Met	**Pro**
Rabbit α1-AT	**Met**	**Pro**	Ser	**Ser**	**Leu**	**Pro**
Carp p62	**Met**	**Pro**	**Met**	**Ser**	**Leu**	**Pro**
Human C1I	Val	Ala	His	Thr	**Leu**	Leu
Human tPAI	Val	Ala	Arg	Met	Ala	**Pro**
Human AT III	Ala	Gly	Arg	**Ser**	**Leu**	Asn
Human α1-ACT	Thr	Leu	Leu	**Ser**	Ala	Leu

FIG. 8. Comparison of the amino acid sequence of the reactive site region of carp α_1-antitrypsin (p62) and other serpins. α1-AT, α_1-antitrypsin; C1I, complement C_1 inhibitor; tPAI, tissue plasminogen activator inhibitor; AT III, antithrombin III; α1-ACT, α_1-antichymotrypsin. Sequence data of other serpins were taken from Huber and Carrell (1989) and from Suzuki et al. (1991).

Gene duplication occurs frequently in teleost fish. As a result, some species are tetraploid, including goldfish and common carp (Risinger and Larhammar, 1993). Goldfish and common carp have ~100 chromosomes, twice the number of chromosomes in other cyprinidae fish (Ohno et al., 1968). Should gene duplication occur, the two resulting genes may accommodate different degrees of mutations. One gene may retain the original function, whereas the other is free to accumulate mutations, resulting in gene products with null or different functions. In this study, we have cloned two genes in carp liver encoding α_1-antitrypsin. The open reading frame of *pCP34* encodes a protein of only 333 amino acids, corresponding to the amino acid residues 78–410 of *pCP9*. The gene product of *pCP34* lacks a signal peptide, the helix domain hA, the sheet domain s6B, and half of the helix domain hB, structural elements associated with α_1-antitrypsin. Some natural α_1-antitrypsin variants such as antitrypsin I and antitrypsin M Procida (Huber and Carrell, 1989) have single substitution in the hA domain, leading to predisposition for emphysema. Another variant, antitrypsin I Malton, has one phenylalanine deletion in the s6B domain and is also nonfunctional. By analogy with these human α_1-antitrypsin variants, the gene product of *pCP34* is expected to be a null mutant. In addition, the product of the *pCP34* gene was not detected in body fluid. Therefore, it appears that only one α_1-antitrypsin gene (*pCP9*) is functional. We have previously cloned two types of cDNA encoding the α (α1 and α2) subunit of carp gonadotropin (Chang et al., 1988). These two cDNAs encode proteins that differ by seven amino acids (three in the signal peptide and four in the mature polypeptide). Both recombinant α1 and α2 subunits, expressed in insect Sf21 cells, are able to associate with the β subunit, but only the $\alpha 1/\beta$ heterodimer displays biological activity (Huang et al., 1991). Therefore, tetraploid fish provide rich sources for the isolation of gene variants and provide valuable clues for the study of the structure–function relationship.

Acknowledgment: We thank Drs. L. S. Kao and S. R. Roffler for critically reading the manuscript. This study was supported by grants from the National Science Council and from Academia Sinica, Taiwan, R.O.C.

REFERENCES

Bensa J. C., Reboul A., and Colomb M. G. (1983) Biosynthesis in vitro of complement subcomponents C1q, C1s and C1̄ inhibitor by resting and stimulated human monocytes. *Biochem. J.* **216,** 385–392.

Bock S. C., Skriver K., Nielsen E., Thógersen H.-C., Wiman B., Donaldson V. H., Eddy R. L., Marrinan J., Radziejewska E., Huber R., Shows T. B., and Magnusson S. (1986) Human C1 inhibitor: primary structure, cDNA cloning, and chromosomal localization. *Biochemistry* **25,** 4292–4301.

Carrell R. W. and Boswell D. R. (1986) Serpins: the superfamily of plasma serine proteinase inhibitors, in *Proteinase Inhibitors* (Barrett A. J. and Salvesen G., eds), pp. 405–419. Elsevier Biomedical Press, Amsterdam.

Carrell R. W. and Travis J. (1985) α_1-Antitrypsin and the serpins: variation and countervariation. *Trends Biochem. Sci.* **10,** 20–24.

Chandra T., Stackhouse R., Kidd V. J., Robson K. J. H., and Woo S. L. C. (1983) Sequence homology between human α_1-antichymotrypsin, α_1-antitrypsin, and antithrombin III. *Biochemistry* **22,** 5055–5060.

Chang Y.-S., Huang C.-J., Huang F.-L., and Lo T.-B. (1988) Primary structure of carp gonadotropin subunits deduced from cDNA nucleotide sequences. *Int. J. Pept. Protein Res.* **32,** 556–564.

Chomczynski P. and Sacchi N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156–159.

Huang C.-J., Huang F.-L., Chang G.-D., Chang Y.-S., Lo C.-F., Fraser M. J., and Lo T.-B. (1991) Expression of two forms of carp gonadotropin α subunits in insect cells by recombinant baculovirus. *Proc. Natl. Acad. Sci. USA* **88,** 7486–7490.

Huang C.-J., Chen C.-C., Chen H.-J., Huang F.-L., and Chang G.-D. (1995) A protease inhibitor of the serpin family is a major protein in carp perimeningeal fluid: I. Protein purification and characterization. *J. Neurochem.* **64,** 1715–1720.

Huber R. and Carrell R. W. (1989) Implications of the three-dimensional structure of α_1-antitrypsin for structure and function of serpins. *Biochemistry* **28,** 8951–8966.

Kozak M. (1987) Compilation and analysis of sequences upstream from the translation start site in eukaryotic mRNAs. *Nucleic Acids Res.* **15,** 8125–8148.

Lipman D. J. and Pearson W. R. (1985) Rapid and sensitive protein similarity searches. *Science* **227,** 1435–1441.

Long G. L., Chandra T., Woo S. L. C., Davie E. W., and Kurachi K. (1984) Complete sequence of the cDNA for human α_1-antitrypsin and the gene for the S variant. *Biochemistry* **23,** 4828–4837.

Mullis K. B. and Faloona F. A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155,** 335–350.

Ny T., Sawdey M., Lawrence D., Millan J. L., and Loskutoff D. J. (1986) Cloning and sequence of a cDNA coding for the human β-migrating endothelial-cell-type plasminogen activator inhibitor. *Proc. Natl. Acad. Sci. USA* **83,** 6776–6780.

Ohno S., Wolf U., and Atkin N. B. (1968) Evolution from fish to mammals by gene duplication. *Hereditas* **59,** 169–187.

Pearson W. R. and Lipman D. J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.

Perlmutter D. H. and Punsal P. I. (1988) Distinct and additive effects of elastase and endotoxin on expression of α_1 proteinase inhibitor in mononuclear phagocytes. *J. Biol. Chem.* **263,** 16499–16503.

Perlmutter D. H., Cole F. S., Kilbridge P., Rossing T. H., and Colten H. R. (1985) Expression of the α_1-proteinase inhibitor gene in human monocytes and macrophages. *Proc. Natl. Acad. Sci. USA* **82,** 795–799.

Remold-O'Donnell E., Chin J., and Alberts M. (1992) Sequence and molecular characterization of human monocyte/neutrophil elastase inhibitor. *Proc. Natl. Acad. Sci. USA* **89,** 5635–5639.

Risinger C. and Larhammar D. (1993) Multiple loci for synapse protein SNAP-25 in the tetraploid goldfish. *Proc. Natl. Acad. Sci. USA* **90,** 10598–10602.

Sambrook J., Fritsch E. F., and Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sanger F., Nicklen S., and Coulson A. R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

Short J. M., Fernandez J. H., Sorge J. A., and Huse W. D. (1988) Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res.* **16,** 7583–7600.

Sommer J. S., Gloor S. M., Rovelli G. F., Hofsteenge J., Nick H., Meier R., and Monard D. (1987) cDNA sequence coding for rat glia-derived nexin and its homology to members of the serpin superfamily. *Biochemistry* **26,** 6407–6410.

Suzuki K., Deyashiki Y., Nishioka J., Kurachi K., Akira M., Yamamoto S., and Hashimoto S. (1987) Characterization of a cDNA for human protein C inhibitor. *J. Biol. Chem.* **262,** 611–616.

Suzuki Y., Yoshida K., Honda E., and Sinohara H. (1991) Molecular cloning and sequence analysis of cDNAs coding for guinea pig α_1-antiproteinase S and F and contrapsin. *J. Biol. Chem.* **266,** 928–932.

Van Nostrand W. E., Wagner S. L., Suzuki M., Choi B. H., Farrow J. S., Geddes J. W., Cotman C. W., and Cunningham D. D. (1989) Protease nexin-II, a potent antichymotrypsin, shows identity to amyloid β-protein precursor. *Nature* **341,** 546–549.

Yanisch-Perron C., Vieira J., and Messing J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33,** 103–119.