Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4863–4867, July 1989
Biochemistry

Serine proteases from nematode and protozoan parasites: Isolation of sequence homologs using generic molecular probes

(molecular evolution/trypsin/cysteine/active sites/polymerase chain reaction)

JUDY A. SAKANARI*†, CATHERINE E. STAUNTON*‡, ANN E. EAKIN§, CHARLES S. CRAIK§, AND JAMES H. MCKERROW*

*Department of Pathology and §Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143; and ‡Cornell University Medical College, New York, NY 10021

Communicated by Russell F. Doolittle, March 27, 1989

ABSTRACT **Serine proteases are one of the biologically most important and widely distributed families of enzymes. Isolation of serine protease genes from organisms of widely diverged phylogenetic groups would provide a basis for studying their biological function, the relationship between structure and function, and the molecular evolution of these enzymes. Serine proteases for which little structural information is known are those that are important in the pathogenesis of parasitic nematode and protozoan diseases. Identification and isolation of protease genes from these organisms is a critical first step in understanding their function for the parasite and possibly suggesting innovative approaches to arresting parasitic diseases. Serine protease gene fragments were isolated from genomic DNA of the parasitic nematode *Anisakis simplex* by using degenerate oligonucleotide primers and the polymerase chain reaction. Primers were designed based upon the consensus sequence of amino acids flanking the active site serine and histidine residues of eukaryotic serine proteases. Four serine protease gene fragments from this parasite were sequenced and one is 67% identical to the rat trypsin II gene. Alignment of these two genes revealed that the intron–exon junctions are conserved between nematode and rat suggesting that this *Anisakis* serine protease is structurally and functionally similar to rat trypsin. The generality of this approach to identify serine protease genes from genomic DNA of two very divergent species, a parasitic protozoan and a mammal, was also confirmed. Genes for other enzymes or any protein with conserved structural motifs can be identified and isolated using this technology. Using a similar strategy, a cathepsin B-like cysteine (thiol) protease gene fragment was isolated from *Caenorhabditis elegans* DNA.**

Serine proteases are one of the most important families of enzymes found in nature. Members of this ubiquitous class of proteases hydrolyze peptide bonds and are involved in a broad range of biological processes including intra- and extracellular protein metabolism, digestion, blood coagulation, clot dissolution, immunological response, developmental regulation, and fertilization (1–4). The three-dimensional structure of serine proteases, in addition to biophysical, molecular biological, and enzymological studies, provides very useful models to understand the mechanism of enzyme action, the basis of substrate specificity, and the molecular evolution of the enzymes themselves (5–15).

Aside from their role in the physiology and metabolism of organisms, serine proteases have also been implicated in the pathogenesis of a number of infectious diseases. Among the most prevalent of these are parasitic diseases, such as schistosomiasis and onchocerciasis (African river blindness), which represent some of the world's greatest health problems (16–18). Parasite proteases may facilitate invasion of host tissue, metabolism of host proteins, and evasion of the host immune response. A generic molecular technology for isolation of serine protease genes from diverse sources would be of immense value in expanding the data base for serine proteases and to further our understanding of the function of these enzymes in a variety of organisms.

In the first step of developing such a generic molecular strategy, we report the isolation of four serine protease genes from a parasitic nematode using degenerate oligonucleotide primers and genomic DNA. These primers were designed based upon mammalian serine protease consensus sequences encoding the active site amino acids and were used to initiate the polymerase chain reaction (PCR) on genomic DNA from the nematode. The applicability of this technique to a wider spectrum of the eukaryotic serine protease family was confirmed when these primers and the PCR were used also to identify serine proteases from two widely diverged phylogenetic groups, mammals and protozoa. Analysis of amplified fragments revealed that the catalytic triad of the serine protease active site was conserved among these organisms.

Anisakiasis is a human disease caused by the ingestion of the larval nematode *Anisakis* found in raw seafood dishes such as sushi and sashimi (19). This parasite can be invasive and penetrate the wall of the stomach or intestine. We have found that secretions of tissue-invasive larvae contain a trypsin-like serine protease that may facilitate the invasion of host tissue (20). Based upon the enzymological characterization of this enzyme by substrate gel electrophoresis, peptide substrate assays, and inhibition studies, we hypothesized that it was structurally related to the trypsin family of serine proteases and not the subtilisin family. Western blot analysis of *Anisakis* extract with rat trypsin antisera identified a protein (M_r, 25,000) that shared similar epitopes with a mammalian trypsin (unpublished data).

All known members of the eukaryotic serine protease family have serine, histidine, and aspartic acid at the active site of the enzyme at amino acid positions 195, 57, and 102, respectively (21). The serine and histidine residues are known to be required for enzymatic activity based on chemical modification experiments using organophosphates for Ser-195 (22) and chloromethyl ketone affinity reagents for His-57 (23). X-ray crystallography and site-directed mutagenesis were used to verify the role of Asp-102 in catalysis (24, 25). Alignment of representative serine protease sequences reveals that amino acids flanking these active site residues are also conserved (26). Use of oligonucleotide probes based on these consensus sequences to screen cDNA or genomic libraries has not been a successful approach to isolation of serine protease cDNA or genes, because the corresponding

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: PCR, polymerase chain reaction.

†To whom reprint requests should be addressed.

nucleic acid codons for these conserved amino acids are very degenerate (2048-fold for the histidine region, 3072-fold for the aspartic acid region, and 8192-fold for the serine region), and false signals are common. The PCR has proved to be an extremely sensitive and remarkably selective means of identifying small amounts of specific genetic material (27, 28). We reasoned, therefore, that under low-stringency conditions, this technique might overcome the drawback of the degeneracy of the oligonucleotide primers because both primers flanking the region of interest in the gene must hybridize for the PCR to occur. DNA sequences between the active site serine- and histidine-coding sites could then be amplified and cloned by this technique. Even broader applicability of this strategy was suggested by the isolation of a cathepsin B-like cysteine (thiol) protease from the free-living nematode *Caenorhabditis elegans* using consensus primers based on mammalian cysteine protease sequences.

MATERIALS AND METHODS

Primers. Serine protease oligonucleotide primers were designed from the consensus sequences flanking the histidine, aspartic acid, and serine residues from eukaryotic organisms (26). The histidine primer or sense primer and serine primer or antisense primer were used to amplify the gene fragments in the PCR. The internal probe (consensus sequence flanking the aspartic acid residue) was used for hybridization to amplified products on Southern blots. Cysteine protease oligonucleotide primers were designed based upon the consensus sequences flanking the active site cysteine (primer = 4096-fold degenerate) and a conserved region located 40 amino acids downstream from this active site residue (primer = 1024-fold degenerate) (29). Restriction sites *Eco*RI and *Hin*dIII were added to the 5′ ends of each primer to allow cloning in a known orientation and subsequent sequencing. Three additional bases (ACA) were added to the 5′ ends to ensure polymerization through the restriction sites. Inosine was used to minimize degeneracy and to maximize base-pairing promiscuity. Oligonucleotides were synthesized and purified by using a β-cyanoethyl phosphoramidite protocol and purified by anion exchange on a FPLC column (Operon, San Pablo, CA).

Amplification of Protease Gene Fragments and Sequence Analysis. Genomic DNA from *Anisakis* larvae, rat liver, and *Trypanosoma cruzi* epimastigotes and plasmid pTN, which contains a cDNA insert of rat trypsin (30), were used in the PCR for amplification of serine protease genes. pTN (50 ng per 50 μl of reaction volume) was used as a positive control since the size of the amplified product could be predicted from the known cDNA sequence [434 base pairs (bp)]. *Anisakis* genomic DNA was isolated using the guanidine isothiocyanate/CsCl gradient technique for extraction of RNA and DNA (31); rat DNA was extracted as described by Craik *et al.* (32); *T. cruzi* epimastigote DNA was prepared as described by Castro *et al.* (33). Genomic DNA from *C. elegans* was used to amplify cysteine protease genes and was extracted using standard methods.

Genomic DNA (200 ng) was used in a 50-μl reaction volume and amplified in the PCR described by Rappolee *et al.* (28). DNA was amplified for 60 cycles of PCR with primers annealed at 25°C for 2 min. Fragments were visualized on 4% agarose gels [3% (wt/vol) NuSieve GTG/1% SeaKem GTG; FMC]. Gels of amplified serine protease gene fragments were Southern blotted, blots were hybridized overnight with the ^{32}P-labeled probes (internal aspartic acid probe, cDNA of rat trypsin and chymotrypsin) at 37°C and washed three times with 4× SSC/0.1% SDS at 37°C for 30 min (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0). After amplification of the serine protease gene fragments, reaction products were extracted once with phenol/chloroform [1:1 (vol/vol)], ethanol-precipitated, and digested with *Eco*RI and *Hin*dIII. Gel fragments were cut out of a 1% agarose gel, DNA was extracted with "glassmilk" (Geneclean), ligated into M13-mp18 and M13mp19 to obtain the sequences of both coding and anticoding strands, and sequenced by the Sanger dideoxynucleotide method using a modified T7 polymerase (Sequenase) and the M13 universal primer. The 150-bp gene fragment amplified using the cysteine primers was subcloned into Bluescript (Stratagene). DNA was prepared using the mini-prep alkaline lysis method (34) and used directly for sequencing with Sequenase and the M13 universal primer.

Amino acid sequences of both serine and cysteine proteases were derived from the nucleic acid sequences and analyzed using the sequence analysis package designed and provided by the Biomathematics Computation Lab, Department of Biochemistry and Biophysics, University of California, San Francisco.

RESULTS AND DISCUSSION

After amplification, products from the reactions with serine protease primers were visualized on ethidium bromide-stained gels (Fig. 1*A*) and probed with an oligonucleotide to an internally conserved amino acid sequence encoding the third member (aspartic acid) of the catalytic triad. The amplified DNA fragments were confirmed as serine protease gene fragments by (*i*) hybridization with a known serine protease gene (Fig. 1*B*) or (*ii*) DNA sequencing (Fig. 2*B*).

Sequence analysis of each of the amplified gene fragments from *Anisakis*, which were recognized by the internal oligonucleotide probe, suggested that they were serine protease genes. Each contained regions of DNA that encode the three amino acid sequences that represent the active site consensus sequences of serine proteases. To determine which of the four serine protease gene fragments may represent the trypsin-like serine protease identified in parasite secretions by enzyme assays, amplified products were hybridized to rat pancreatic trypsin II cDNA. The 728-bp fragment hybridized to this probe under moderate-stringency conditions (data not shown). The DNA sequence of this fragment is, in fact, 67% identical to the rat trypsin II gene (Fig. 3) and shows 57% sequence identity at the amino acid level. Alignment of the two gene sequences revealed that the intron–exon junctions in both organisms are conserved but that the sizes of the introns are smaller in the nematode. These data and the results of our biochemical studies and Western blot analysis with rat trypsin antisera suggest that this gene fragment from *Anisakis* encodes a serine protease that is structurally and functionally very similar to rat trypsin and may be involved in the degradation or digestion of host tissues.

Brenner (37) proposed that there are two subclasses of serine proteases: one contains a TCN (N = A, T, G, or C) codon and the other contains an AGY (Y = C or T) codon for the active-site serine residue. Sequence data showed that two of the gene fragments from *Anisakis* use the TCN codon (728 bp and 482 bp) and the other two use the AGY codon (470 bp and 195 bp) to code for the active-site serine residue. Both subclasses of serine protease are, therefore, present in this primitive metazoan. In addition, the 482-bp fragment appears to be a pseudogene with stop codons after the histidine and serine active-site residues (Fig. 2).

We further tested the universality of the active-site primers and PCR to identify serine proteases from organisms of diverse phyla. Genomic DNA from the rat and the parasitic protozoan *T. cruzi* (causative agent of Chagas disease) was amplified and analyzed by ethidium bromide staining. Seventeen bands were amplified from rat genomic DNA and 8 bands were amplified from *T. cruzi* DNA (Fig. 1*A*). The many gene fragments amplified from rat DNA are consistent with the multiple serine proteases present in vertebrates. Al-

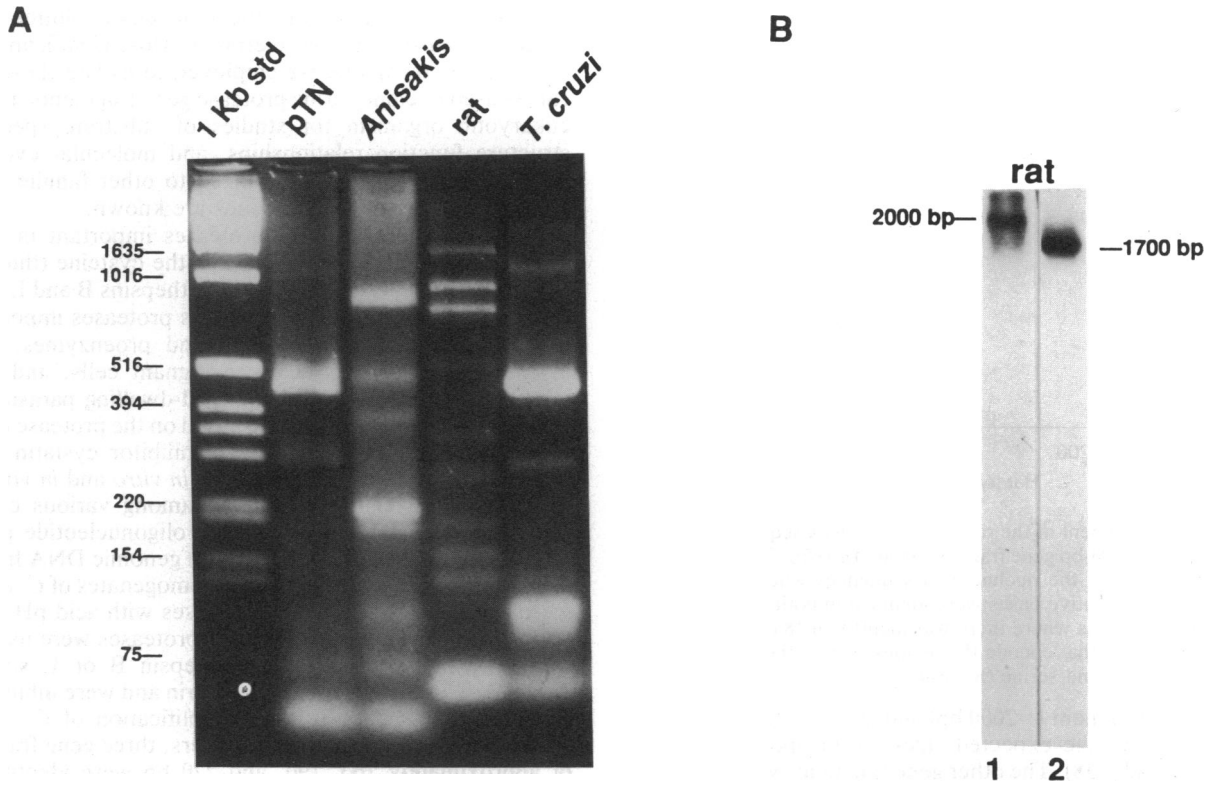

FIG. 1. Gene fragments visualized in agarose gels and identified by Southern hybridization. (*A*) Ethidium bromide-stained gel of serine protease gene fragments generated from the PCR using pTN, *Anisakis*, rat, and *T. cruzi* DNA. An *Ava* I digest of a recombinant pBR322 vector containing repeats of a 1014-kilobase insert (1 Kb DNA ladder; BRL) is in lane 1 Kb std. (*B*) Southern blot of rat genomic DNA fragments hybridized with ^{32}P-labeled rat trypsin II (lane 1) and ^{32}P-labeled rat chymotrypsin B (lane 2).

though every fragment has not yet been characterized, we identified two fragments by using available probes. Amplified DNA fragments from the rat were hybridized with rat trypsin and chymotrypsin gene probes for identification (Fig. 1*B*).

A

```
                          57              102                 195
     EcoRI TrpValValThrAlaAlaHisCys   GlyAsnAspIleAlaLeuLeu   GlyAspSerGlyGlyProLeuHindIII
5'-ACAGAATTCTGGGTIGTIACIGCIGCICAYTG-3' 3'-CCITTICTITAICGIXAIXA-5' 3'-CCICTX(TC/AG)ICCICCIGGIXATTCGAAACA-5'
```

B

Nematode (*Anisakis*)	- 728 bp	VTAA**H**CYKS	INN**D**IMLIQLS	DSCQGD**S**GGP
	- 482	VTAA**H**	SLN**D**LTVLSAT	GD**S**GGP
	- 470	VTAA**H**CVYD	RCS**D**IALMQLT	DACQGD**S**GGP
	- 195	VTAA**H**CDLA	PLA**D**LTVGSQT	VSVYGD**S**GGP
Protozoa (*T. cruzi*)	- 438 bp	VTAA**H**CWRI	GIS**D**AQWAKGL	DIVIGD**S**GGP
Mouse - H Factor		LTAA**H**CNVG	REG**D**LQLVRLK	DSCNGD**S**GGP
Rat - trypsin II		VSAA**H**CYKS	LNN**D**IMLIKLK	DSCQGD**S**GGP
Rat - chymotrypsin B		VTAA**H**CGVK	INN**D**ITLLKLS	SSCMGD**S**GGP
Rat - group specific protease		LTAA**H**CKGR	NLH**D**IMLLKLE	AAFMGD**S**GGP
Rat - elastase I		MTAA**H**CVSS	AGY**D**TALLRLA	SSCNGD**S**GGP
Rat - elastase II		LTAA**H**CISN	NGN**D**IALVKLA	SGCQGD**S**GGP
Human - prothrombin		LTAA**H**CLLY	LDR**D**IALMKLK	DACEGD**S**GGP
Human - TPA		LSAA**H**CFQE	YDN**D**IALLQLK	DACQGD**S**GGP
Human - plasminogen		LTAA**H**CLEK	TRK**D**IALLKLS	DSCQGD**S**GGP
Human - Factor IX		VTAA**H**CVET	YNH**D**IALLELD	DSCQGD**S**GGP
Trematode - (*Schistosoma mansoni*) elastase		LTAG**H**CVCS	SGP**D**IAIVMLA	LPAPGD**S**GGP

FIG. 2. (*A*) DNA primers and probe sequences used for amplification and isolation of serine protease homologs. The active site amino acid residues are numbered; X is A or G; Y is C or T; I is inosine. The amino acid sequences shown for the aspartic probe and serine primer were designed from the reverse complement of the DNA sequence shown. (*B*) Amino acid sequences surrounding the active site catalytic triad of *Anisakis* and *T. cruzi* serine proteases compared to other species. Alignment of these fragments from *Anisakis* and *T. cruzi* shows that they contain the conserved residues of the catalytic triad of serine proteases (shown in bold type). Conserved regions flanking the active site residues of mouse, rat, human, and trematode serine proteases are also shown for comparison (35, 36). The single-letter amino acid code is used. TPA, tissue plasminogen activator.

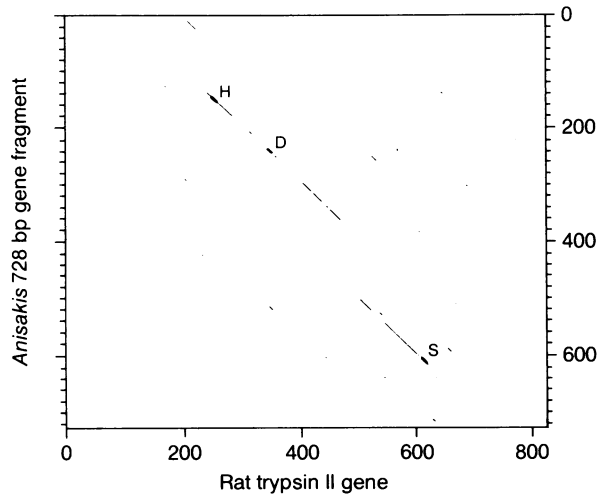

FIG. 3. Alignment of the rat trypsin II cDNA sequence with the sequence of the 728-bp gene fragment from *Anisakis*. This dot matrix represents areas of the nucleic acid sequences where 14 or more bases of 21 consecutive bases were identical in both species. Thick lines represent areas where there was identity in 18 of 20 bases and are the sequences that encode the regions around the histidine (H), aspartic acid (D), and serine (S) residues.

The trypsin fragment (≈2000 bp) and chymotrypsin fragment (≈1700 bp) are the expected sizes based upon known sequence data (32, 38). The other gene fragments will be further characterized by DNA sequencing and may include pseudogenes or serine protease genes not yet described. The major band (438 bp) of the amplified products from *T. cruzi* was sequenced and contains regions that encode the conserved amino acid sequences of the active sites of serine proteases (Fig. 2*B*). Although no serine protease activity has been identified from *T. cruzi*, the sequence data suggest that a serine protease gene may be present in this organism. Using the degenerate oligonucleotide primers in conjunction with the PCR, we could identify gene fragments from species belonging to widely diverged animal groups.

The importance of isolating these serine protease genes from parasitic organisms extends beyond the identification of a primitive trypsin gene. Parasite proteases play critical roles in parasitic infections and may be targets of anti-parasitic vaccines or chemotherapy (16–18, 36). Furthermore, Hill and Hastie (39) have suggested that parasite proteases may be one of the critical factors driving the molecular evolution of host serum protease inhibitors (serpins). Most significant is the fact that the technique we employed to isolate these genes can be used to clone serine protease gene fragments from any eukaryotic organism for studies of substrate specificity, structure–function relationships, and molecular evolution. This technique can be generalized to other families of enzymes whose active site residues are known.

Another major family of proteases important in a wide spectrum of biological processes is the cysteine (thiol) protease family. This family includes cathepsins B and L, which are lysosomal hydrolases, as well as proteases important in the processing of prohormones and proenzymes, in the invasion of human tissue by malignant cells, and in the metabolism of hemoglobin by blood-dwelling parasites (17, 18, 40, 41). A synthetic peptide based on the protease binding domain of the cysteine protease inhibitor cystatin C was found to have antibacterial activity *in vitro* and *in vivo* (42). Based on two regions conserved among various cysteine proteases (Fig. 4), we constructed oligonucleotide primers for amplification and tested these on genomic DNA from the free-living nematode *C. elegans*. Homogenates of *C. elegans* were shown to have several proteases with acid pH optima (44). At least two distinct cysteine proteases were identified that were active against the cathepsin B or L substrate Z-Phe-Arg-7-amino-4-methylcoumarin and were inhibited by leupeptin at pH 5 (44). After amplification of *C. elegans* genomic DNA by using these primers, three gene fragments of approximately 160, 190, and 220 bp were identified on ethidium bromide-stained gels (data not shown). The most prominent band (160 bp) was subcloned and sequenced. Analysis of this sequence showed 55% identity (over 53 amino acids) with rat cathepsin B (Fig. 4). This result provides a molecular probe for use in studying the functional role and regulation of expression of a protease in *C. elegans*, an organism widely used to define some of the basic mechanisms of cellular differentiation. The primers can also provide a means of isolating genomic fragments from a variety of other organisms for which cysteine proteases are known to play a critical role.

The serine and cysteine protease gene fragments isolated by this method provide probes to obtain the full-length cDNAs and genes. The laborious and time-consuming efforts of enzyme purification from natural sources can be circumvented by cloning, sequencing, and expressing the gene encoding the enzyme of interest. Characterization of the

A

```
                          25                                  65
     EcoRI GlnGlyGlnCysGlySerCysTrp          GluGlyCysAsnGlyGlyTyrHindIII
5'-ACAGAATTCCAXGGICAXTGYGGITC/AGITGYTGG-3'  3'-CTYCCIACXTTXCCICCIATXTTCGAAACA-5'
```

B

```
Nematode (C. elegans)  QGQCGSCWAFSTAEVISDGTCMASNGTQQPIICPTDLLTC.....CWNVCGEGCNGGY
Alignment              :: ::::::::::..:.:: .:. .:: . ....:::::        :   ::.:::::::
rat-cathepsin B        QGSCGSCWAFGAVEAMSDRICIHTNGRVNVEVSAEDLLTC.....CGIQCGDGCNGGY
rat-cathepsin H        QGACGSCWTFSTTGALESAVAIAS.GKMM.TLAEQQLVDC.....AQNFNNHGCQGGL
mouse Cys proteinase   QGQCGSCWAFSASGCLEGQMFLKT.GKLI.SLSEQNLVDC.....SHAQGNQGCNGGL
human-cathepsin B      QGSCGSCWAFGAVEAISDRICIHTN..VSVEVSAEDLLTC.....CGSMCGDGCNGGY
Dictyostelium Cys proteinase  QGQCGSCWSFSTTGNVEGQHFISQN.KLV.SLSEQNLVDCDHECMEYEGEEACDEGCNGGL
Chinese gooseberry-actinidin  QGECGGCWAFSAIATVEGINKITS.GSLI.SLSEQELIDC.....GRTQNTRGCDGGY
barley-aleurain        QAHCGSCWTFSTTGALEAAYTQAT.GKNI.SLSEQQLVDC.....AGGFNNFGCNGGL
papaya-papain          QGSCGSCWAFSAVVTIEGIIKIRT.GNLN.QYSEQELLDC.....DRRSYGCNGGY
                                25                                  65
```

FIG. 4. (*A*) DNA primers used for amplification and isolation of cysteine protease homologs. The active site cysteine (residue 25) is shown in bold type. Design of these primers was similar to that for the serine protease primers (see Fig. 2). (*B*) Comparison of the amino acid sequence of the cysteine protease homolog isolated from *C. elegans* with other cysteine proteases (29). Gaps (. . .) were added to maximize similarity. Alignment of the *C. elegans* protease with rat cathepsin B is indicated with a colon (:) denoting amino acid identity and a single dot (.) denoting conservative amino acid changes. Residue numbers follow the papain numbering system (43); the single-letter amino acid code is used.

overexpressed gene products will provide biochemical information on the function of these enzymes and the role they may play in the physiology of these organisms or the diseases they produce. Such analyses can also provide a data base for comparison of sequences that define amino acids critical for structure or function of the enzyme. Direct postulates can be made and tested with the genetic system using site-directed mutagenesis. This approach is not limited to the active site of enzymes but can be used for a variety of conserved structural motifs encoded by signature sequences (EF-hand for Ca^{2+} binding, nucleotide fold, immunoglobulin fold, GTPase binding domain, etc.) in protein families. Application of this strategy should greatly increase our understanding of the molecular evolution of enzymes and perhaps other protein families as well.

We thank S. Craig for the *T. cruzi* DNA; M. J. Banda, R. J. Fletterick, W. J. Rutter, and D. V. Santi for their suggestions; Z. Werb for allowing us use of her Perkin–Elmer/Cetus thermocycler; and E. M. Shimazu for preparing the manuscript. We are grateful to D. A. Rappolee for his advice, helpful discussions, and reviewing the manuscript. This work was supported in part by grants from the National Institute of Allergy and Infectious Diseases (AI20452), the Edna McConnell Clark Foundation, and the World Health Organization to J.H.McK., the National Science Foundation (DMB8608086 and EET8807179) to C.S.C., and an American Heart Association Medical Research Fellowship provided through Cornell University Medical College to C.E.S.

1. Stroud, R. M. (1974) *Sci. Am.* **231,** 74–88.
2. Kraut, J. (1977) *Annu. Rev. Biochem.* **46,** 331–358.
3. Neurath, H. (1984) *Science* **224,** 350–357.
4. Neurath, H. (1986) *J. Cell Biochem.* **32,** 35–49.
5. Birktoft, J. J. & Blow, D. M. (1972) *J. Mol. Biol.* **68,** 187–240.
6. Tulinsky, A., Vandlen, R. L., Morimoto, C. N., Mani, N. V. & Wright, L. H. (1973) *Biochemistry* **12,** 4185–4192.
7. Stroud, R. M., Kay, L. M. & Dickerson, R. E. (1974) *J. Mol. Biol.* **83,** 198–208.
8. Huber, R., Dietmar, K., Wolfram, B., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974) *J. Mol. Biol.* **89,** 73–101.
9. Sawyer, L., Shotton, D. M., Campbell, J. W., Wendell, P. L., Muirhead, H. & Watson, H. C. (1978) *J. Mol. Biol.* **118,** 137–208.
10. Bode, W., Chen, Z., Bartels, K. & Bartunik, H. (1983) *J. Mol. Biol.* **164,** 237–282.
11. Wright, C. S., Alden, R. A. & Kraut, J. (1969) *Nature (London)* **221,** 235–242.
12. Brayer, G. D., Delbaere, L. T. J. & James, M. N. G. (1978) *J. Mol. Biol.* **124,** 261–283.
13. Codding, P. W., Delbaere, L. T. & Hayakawa, K. (1974) *Can. J. Biochim.* **52,** 208–220.
14. Brayer, G. D., Delbaere, L. T. J. & James, M. N. G. (1979) *J. Mol. Biol.* **131,** 743–775.
15. Doolittle, R. F. & Feng, D. F. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 869–874.
16. McKerrow, J. H. (1988) in *Parasitic Infections: Contemporary Issues in Infectious Disease*, eds. Leech, J. H., Sande, M. A. & Root, R. K. (Churchill Livingstone, New York), Vol. 7, pp. 51–59.
17. McKerrow, J. H. & Doenhoff, M. F. (1988) *Parasitol. Today* **4,** 334–339.
18. McKerrow, J. H. *Exp. Parasitol.* **68,** 111–115.
19. Sakanari, J. A. & McKerrow, J. H. *Clin. Microbiol. Rev.*, in press.
20. Sakanari, J. A. & McKerrow, J. H. (1988) *J. Cell. Biochem.*, Suppl. 12B, 299 (abstr.).
21. Hartley, B. S. (1964) *Nature (London)* **201,** 1284–1287.
22. Dixon, G. H., Go, S. & Neurath, H. (1956) *Biochim. Biophys. Acta* **19,** 193–195.
23. Shaw, E., Mares-Guia, M. & Cohen, W. (1965) *Biochemistry* **4,** 2219–2224.
24. Craik, C. S., Roczniak, S., Largman, C. & Rutter, W. J. (1987) *Science* **237,** 909–913.
25. Blow, D. M., Birktoft, J. J. & Hartley, B. S. (1969) *Nature (London)* **221,** 337–340.
26. Craik, C. S., Rutter, W. J. & Fletterick, R. (1983) *Science* **220,** 1125–1129.
27. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230,** 1350–1354.
28. Rappolee, D. A., Mark, D., Banda, M. J. & Werb, Z. (1988) *Science* **241,** 708–712.
29. Portnoy, D. A., Erickson, A. H., Kochan, J., Ravetch, J. V. & Unkeless, J. C. (1986) *J. Biol. Chem.* **261,** 14697–14703.
30. Vasquez, J., Evnin, L., Higaki, J. & Craik, C. S. *J. Cell. Biochem.* **39,** 265–276.
31. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) *Basic Methods in Molecular Biology* (Elsevier, New York), pp. 130–135.
32. Craik, C. S., Choo, Q. L., Swift, G. H., Quinto, C., MacDonald, R. J. & Rutter, W. J. (1984) *J. Biol. Chem.* **259,** 14255–14264.
33. Castro, C., Craig, S. P. & Castañeda, M. (1981) *Mol. Biochem. Parasitol.* **4,** 273–282.
34. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 368–369.
35. Gershenfeld, H. K. & Weissman, I. L. (1986) *Science* **232,** 854–861.
36. Newport, G. R., McKerrow, J. H., Hedstrom, R., Petitt, M., McGarrigle, L., Barr, P. J. & Agabian, N. (1988) *J. Biol. Chem.* **263,** 13179–13184.
37. Brenner, S. (1988) *Nature (London)* **334,** 528–529.
38. Bell, G. I., Quinto, C., Quiroga, M., Valenzuela, P., Craik, C. S. & Rutter, W. J. (1984) *J. Biol. Chem.* **259,** 14265–14270.
39. Hill, R. E. & Hastie, N. D. (1987) *Nature (London)* **326,** 96–99.
40. Barrett, A. J. (1977) *Proteinases in Mammalian Cells and Tissues* (Elsevier/North-Holland, New York), pp. 181–208.
41. Sloane, B. F., Rozhin, J., Hatfield, J. S., Crissman, J. D. & Honn, K. V. (1987) *Exp. Cell Biol.* **55,** 209–224.
42. Björck, L., Åkesson, P., Bohus, M., Trojnar, J., Abrahamson, M., Olafsson, I. & Grubb, A. (1989) *Nature (London)* **337,** 385–386.
43. Light, A., Frater, R., Kimmel, J. R. & Smith, E. L. (1964) *Proc. Natl. Acad. Sci. USA* **52,** 1276–1283.
44. Sarkis, G. J., Kurpiewski, M. R., Ashcom, J. B., Jen-Jacobson, L. & Jacobson, L. A. (1988) *Arch. Biochem. Biophys.* **261,** 80–90.