J. Biochem. **133**, 799–810 (2003)
DOI: 10.1093/jb/mvg102

Molecular Cloning and Functional Characterization of Crustapain: A Distinct Cysteine Proteinase with Unique Substrate Specificity from Northern Shrimp *Pandalus borealis*

Hitoshi Aoki[1,2,*], Md. Nazmul Ahsan[1,*] and Shugo Watabe[1,†]

[1]Laboratory of Aquatic Molecular Biology and Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657; and [2]Research & Development Center, Nichirei Corporation, 9 Shinminato, Mihama, Chiba 261-8545

Received March 3, 2003; accepted April 11, 2003

A cDNA clone encoding a cysteine proteinase of the papain superfamily has been isolated from the hepatopancreas of northern shrimp *Pandalus borealis* (NsCys). NsCys shares the highest identity of 64% with a cathepsin L–like cysteine proteinase from lobster, and its identity to the well-characterized mammalian cathepsins S, L, and K falls within a narrow range of 54–59%. However, it differs from each of these cathepsins in certain key residues including, for example, the unique occurrence of tryptophan and cysteine residues at the structurally important S2 subsite. Consequently, NsCys produced in *Pichia pastoris* appears to be distinct in various physicokinetic properties. The recombinant enzyme is active and stable over a wide range of pH values, and its substrate specificity is unusual, as demonstrated by its poor affinity for phenylalanine residues. Instead, it shows the highest specificity for proline residues, a property similar to cathepsin K. Unlike cathepsin K, however, NsCys cleaves valine residues more efficiently than leucine. Similar results were obtained with the natural peptide substrate glucagon. The shrimp proteinase is further distinguished by its potent collagenolytic activity, resulting in a cleavage pattern reminiscent of bacterial collagenase. To distinguish such unique structural and enzymatic properties, we propose the trivial name "crustapain" for the shrimp proteinase, indicating that it is a papain-like cysteine proteinase from a crustacean species.

Key words: cathepsin, collagen, crustapain, cysteine proteinase, northern shrimp.

Abbreviations: Bz, benzoyl; E-64, L-3-carboxy-*trans*-2,3-epoxypropionyl-leucyl-amido-(4-guanidino)butane; MCA, 4-methyl-7-coumarylamide; NsCys, northern shrimp *Pandalus borealis* cysteine proteinase; PMSF, phenylmethylsulfonyl fluoride; RACE, rapid amplification of cDNA ends; Suc, 3-carboxy-propionyl; Z, benzyloxycarbonyl.

Cysteine proteinases [EC 3.4.22] of the papain superfamily, often referred to as cathepsins, comprise a large number of enzymes from both prokaryotes and eukaryotes, with representative members expressed in bacteria, fungi, protozoa, plants, and human (*1*). In mammals, they are well documented as ubiquitously expressed intracellular housekeeping enzymes involved in general lysosomal protein breakdown, which include, for example, cathepsins L, B, H, and O. In addition, extracellular cysteine proteinases such as cathepsins S, K, and W have been implicated in various pathophysiological processes correlated with their specific tissue distributions (*2*).

The primary structures of these enzymes share a common domain organization, with a signal peptide of 16–18 amino acids, followed by a propeptide of 62–100 residues and a catalytically active mature region of about 220–230 amino acids (*1*, *2*). The signal sequence, containing stretches of hydrophobic amino acids, facilitates the targeting of nascent protein into secretory pathways, while the propeptide is responsible for folding, temporary inhibition and transport of the proenzyme as an inactive precursor to the endosomal/lysosomal compartment (*2–4*). The catalytic moiety of the mature enzyme is composed of a triad consisting of cysteine, histidine and asparagine residues and is folded into a two-domain structure (*1*, *2*). However, these enzymes differ in tissue distribution and in some enzymatic properties, including substrate specificities, due to minor changes in the core structural regions.

Most of the knowledge about the structural features that distinguish an array of cathepsins from each other has been derived from mammalian sources because of their relevance to various pathological conditions in human (*1*, *2*, *4*). Recently, interest in these enzymes, particularly cathepsin L from parasitic invertebrates, has increased considerably because of their distinct tissue-specific roles in digestion, development, and immune progression (*5–8*). However, the subtleties of how these enzymes might have evolved in different organisms to serve different functions yet share the same structural topology have largely been overlooked, even though such understanding might assist in the search for enzymes with novel specificity.

Enzymes derived from organisms living in cold marine waters differ notably from those in terrestrial animals. Most striking is the fact that these marine enzymes are more active at low temperature and also differ from their warm-blooded counterparts in their pH optima, pH sta-

*These authors contributed equally to this work.
†To whom correspondence should be addressed. Tel: +81-3-5841-7520, Fax: +81-3-5841-8166, E-mail: awatabe@mail.ecc.u-tokyo.ac.jp

bilities, substrate affinities and catalytic efficiencies (*9*, *10*). In this regard, the northern shrimp *Pandalus borealis* is particularly interesting because it is adapted to live at temperatures as low as –1.6°C in the marine waters of northern hemisphere (*11*). The study of enzymatic molecules in such poikilothermic species is of great interest, especially in relation to the strategies adopted by these organisms to achieve a normal level of proteolysis at temperatures well below that of homoeothermic species.

In a preliminary experiment as part of our search for proteinases with unique substrate specificities from cold-adapted marine species, we observed a collagenolytic cysteine proteinase activity in northern shrimp hepatopancreas, although repeated attempts to purify the enzyme ended up with what appeared to be cathepsin L. Subsequently, by using degenerate oligonucleotide primers, we cloned two cDNAs, one encoding cathepsin L (unpublished data), and the other one encoding a similar but distinct cysteine proteinase (NsCys). To overcome the problem of obtaining pure enzyme in sufficient quantities, we then overexpressed the encoded protein in methylotrophic yeast *Pichia pastoris* as a fully functional enzyme in quantities amenable to its detailed enzymatic characterization. Sequence comparison both at nucleotide and predicted amino acid levels identified the shrimp protein as a cysteine proteinase member of the papain superfamily, which differed considerably in structurally important S2 subsite [nomenclature according to Schechter and Berger (*12*)] from other well-characterized family members. The expressed protein was highly specific for a proline residue at P2 position, a property similar to human cathepsin K. However, unlike most cathepsins, the shrimp proteinase showed a broad pH activity and stability profiles and was able to extensively degrade type I native collagen molecules. In this communication, we describe these unique properties of the shrimp proteinase and discuss them with regard to the structural differences between it and related members of the papain superfamily. Considering the present situation of nomenclature for cathepsin and taking note of the unique structural and enzymatic properties of the shrimp proteinase, we propose the trivial name of "crustapain" for this enzyme, indicating that it is a papain-like cysteine proteinase from a crustacean species.

MATERIALS AND METHODS

Chemicals—7-Amino-4-methylcoumarin (MCA) used as the standard was purchased from Peptide Institute (Osaka). Fluorogenic peptide substrates Z-Arg-Arg-MCA, Z-Phe-Arg-MCA, Z-Pro-Arg-MCA, Z-Val-Val-Arg-MCA, and Z-Leu-Leu-Arg-MCA and cysteine proteinase inhibitors E-64 were obtained from Bachem (Bubendorf, Switzerland). Porcine skin acid-soluble type I collagen and glucagon were obtained from Wako Pure Chemicals Ltd. (Tokyo). All other reagents used were of highest analytical grade available.

Shrimp—Live northern shrimps *Pandalus borealis* used in this study (16.5 ± 1.2 g body weight) were procured from a fishery cooperative society in Ishikawa Prefecture. Organs of interest were dissected, immediately frozen in liquid nitrogen, and stored at –80°C until use.

cDNA Cloning of NsCys—Standard molecular biology techniques were performed essentially following Sambrook *et al.* (*13*) to isolate the cDNA clone encoding northern shrimp cysteine proteinase (NsCys). Total RNAs from frozen hepatopancreas tissues of shrimp were extracted with an Isogen solution (Nippon Gene, Tokyo). First strand cDNA was synthesized from total RNAs using a 3′-RACE kit according to the manufacturer's protocol (Life Technologies GibcoBRL). The peptide sequence determined by N-terminal amino acid sequencing of the cathepsin L-like enzyme isolated from northern shrimp in an earlier study (unpublished data) was used to design two degenerate oligonucleotide sense primers (nsCatF1: 5′-GA(T/C)TGGCG(G/C)GA(T/C)AA(A/G)GG(G/C)GC-3′ and nsCatF2: 5′-CA(A/G)TG(C/T)GG(C/T)TC(C/T)TG(C/T)TGGGC-3′). The design of the oligonucleotide primers was guided by the codon usage of the corresponding residues found in *Homarus americanus* cysteine proteinase 1 (*14*) in order to minimize the degeneracy of the primers. Two rounds of PCR were performed with each of these primers and an antisense anchor primer, AUAP, contained in the RACE kit to amplify the 3′-region of NsCys using a DNA thermal cycler (GeneAmp PCR system 2400; Applied Biosystems). Typically, 100 μl of reaction mixtures contained 20 pmol of forward and reverse primers, about 500 μg of first-strand cDNA as template, 20 nmol of dNTP mixture, 10 μl of 10 X *Ex Taq*™ buffer and 2.5 units of *Ex Taq*™ DNA polymerase (TaKaRa). Subsequently, the 5′-site of the cDNA was amplified using three gene-specific primers (GSP1: 5′-GACTCATCTCCTGAAGCTGG-3′; GSP2: 5′-ACTGCTGACGGTGGCGCCGA-3′ and GSP3: 5′-GGCAATTGTCATCAATTGCC-3′) with a 5′-RACE system (Life Technologies GibcoBRL) according to the manufacturer's protocol. First-strand cDNA was synthesized with GSP 1, while PCR amplification was performed with primers AUAP and GSP2 followed by nested PCR with an internal primer GSP3. Amplified DNA fragments were subcloned into plasmid vector pGEM-T Easy (Promega), using *Escherichia coli* strain JM 109 (Promega) as a host bacterium. Sequencing was performed on both strands using a DNA sequencer model 373S (Applied Biosystems) after labeling the DNA with Dye-Deoxy™ Terminator Cycle Sequencing kit (Applied Biosystems).

Northern Blot Analysis—Total RNAs collected from skeletal muscle, hepatopancreas and intestine tissues were subjected to Northern blot analysis as described by Sambrook *et al.* (*13*). Hybridization was carried out at 65°C for 20 h with a probe derived from the 3′-site of NsCys cDNA (891–1196 bp; see Fig. 1). The probe was randomly labeled with [α-^{32}P]dCTP using a random primer DNA labeling kit ver. 2 (TaKaRa). Autoradiography was performed on an X-ray film (Kodak) for 24 h at –80°C using an intensifying screen (Fujifilm).

Overexpression of NsCys in Yeast—Northern shrimp cysteine proteinase (NsCys) was heterologously expressed in the methylotrophic yeast *Pichia pastoris* using EasySelect™ Echo-Adapted™ *Pichia* Expression Kit (Invitrogen). A 924-bp cDNA encoding the complete precursor form of NsCys excluding the signal peptide was amplified by PCR and subcloned into pUniD/V5-His-TOPO vector supplied with the kit, following the supplier's instructions. The resulting vector was then recom-

```
CACTTTAGCAAGATGAGGTCTCTGTTTCTTATCCTTCTCGGGCTGGCTGCGGTCTCCGCC   60
          M  R  S  L  F  L  I  L  L  G  L  A  A  V  S  A        16
ATTGGAGAATGGGAAAACTTCAAGACGAAGTTTGGCAAGAAGTATGCCAACTCAGAAGAG  120
 I  G  E  W  E  N  F  K  T  K  F  G  K  K  Y  A  N  S  E  E    36
GAGAGTCACAGAATGTCTGTTTTCATGGACAAACTGAAGTTCATTCAGGAGCACAATGAA  180
 E  S  H  R  M  S  V  F  M  D  K  L  K  F  I  Q  E  H  N  E    56
CGATACGATAAGGGAGAAGTCACTTATTGGCTGAAAATCAACAACTTCTCCGATTTGACC  240
 R  Y  D  K  G  E  V  T  Y  W  L  K  I  N  N  F  S  D  L  T    76
CACGAAGAGGTCTTGGCCACCAAGACTGGAATGACCAGGAGACGACACCCTCTTTCCGTA  300
 H  E  E  V  L  A  T  K  T  G  M  T  R  R  R  H  P  L  S  V    96
TTGCCCAAATCTGCCCCAACCACACCAATGGCCGCAGACGTTGACTGGAGGAATAAGGGG  360
 L  P  K  S  A  P  T  T  P  M  A  A  D  V  D  W  R  N  K  G   116
GCTGTCACCCCCGTCAAGGATCAGGGACAATGCGGATCATGCTGGGCTTTCTCAGCTGTC  420
 A  V  T  P  V  K  D  Q  G  Q  C  G  S  C  W  A  F  S  A  V   136
GCCGCCTTGGAAGGAGCGCACTTCCTGAAGACCGGAGATTTGGTCAGCCTGTCTGAACAG  480
 A  A  L  E  G  A  H  F  L  K  T  G  D  L  V  S  L  S  E  Q   156
AATTTGGTTGACTGCTCTTCGTCTTACGGTAACCAAGGATGTAATGGTGGATGGCCATAC  540
 N  L  V  D  C  S  S  S  Y  G  N  Q  G  C  N  G  G  W  P  Y   176
CAAGCTTATCAATACATCATTGCCAATCGTGGCATTGACACCGAATCGTCATACCCTTAC  600
 Q  A  Y  Q  Y  I  I  A  N  R  G  I  D  T  E  S  S  Y  P  Y   196
AAGGCAATTGATGACAATTGCCGATATGATGCCGGAAACATCGGCGCCACCGTCAGCAGT  660
 K  A  I  D  D  N  C  R  Y  D  A  G  N  I  G  A  T  V  S  S   216
TATGTCGAACCAGCTTCAGGAGATGAGTCCGCACTTCAGCATGCTGTCCAGAATGAAGGA  720
 Y  V  E  P  A  S  G  D  E  S  A  L  Q  H  A  V  Q  N  E  G   236
CCCGTCAGCGTCTGCATTGATGCTGGTCAATCATCTTTCGGTAGTTACGGAGGAGGTGTT  780
 P  V  S  V  C  I  D  A  G  Q  S  S  F  G  S  Y  G  G  G  V   256
TACTATGAACCAAACTGCGATTCCTGGTACGCCAACCATGCCGTGACAGCCGTCGGCTAC  840
 Y  Y  E  P  N  C  D  S  W  Y  A  N  H  A  V  T  A  V  G  Y   276
GGCACCGACGCCAACGGAGGAGATTACTGGATCGTCAAGAACTCGTGGGGTGCATGGTGG  900
 G  T  D  A  N  G  G  D  Y  W  I  V  K  N  S  W  G  A  W  W   296
GGAGAGAGTGGCTACATCAAGATGGCCAGAAACAGGGACAACAACTGTGCCATTGCTACC  960
 G  E  S  G  Y  I  K  M  A  R  N  R  D  N  N  C  A  I  A  T   316
TATAGTGTCTACCCTGTTGTTTAAGATCTTTTATTGACACTCACAATGATTTTCTTTCCA 1020
 Y  S  V  Y  P  V  V                                          323
TCATTTATCATTGGGGAACTTTTAATATTCATTTGGGGTTTTCATTTGATATTTTGTGTA 1080
AGTCTCAGTCAATCCCATTAGACATGTTTTGTTACGGTGGATTCTTAAGTCAACCTTTGA 1140
ATCAAACACTTTTGTCAAATTACAATGAACACATCCAACAGATGATGATACATATGAAAA  1200
TAAAGATACAACAGATAAAAAAAAAAAAAAAAAAAAAAAAA                     1242
```

Fig. 1. **DNA nucleotide and deduced amino acid sequences of the cDNA clone encoding northern shrimp cysteine proteinase.** Numbers with normal and italic numerals on the right margin indicate nucleotides from the 5′-end of cDNA and amino acids from the first methionine, respectively. The catalytic triad residues of cysteine proteinases are reversed. The initiation (ATG) and termination (TAA) codons are shown in boldface. A putative polyadenylation signal, AATAAA, and poly(A) tail are underlined. Open and shaded boxes in the amino acid sequence indicate the putative signal peptide and propeptide, respectively.

bined with *P. pastoris* shuttle vector pPICZα-E using a Cre recombinase-mediated plasmid fusion system (*15*) such that the NsCys cDNA was placed downstream of the yeast α-mating factor secretion signal. Subsequent procedures for transformation with *P. pastoris* host cells, screening for positive clones and induction of protein expression were carried out according to the supplier's protocol and further optimized empirically, details of which will be published elsewhere. Briefly, after linearizing the fusion plasmid construct with *Pme*I, the *P. pastoris* KM71H (*arg4 aox1::ARG4*) competent cells were transformed by electroporation (GenePulser; Bio-Rad), and the positive transformants with multiple copies of integrated NsCys gene were isolated by selection with increasing concentration of zeocin. One of the highly productive clones was selected for large-scale recombinant protein production to obtain a pure preparation of shrimp proteinase from the culture medium following a single size-exclusion chromatography step of the concentrated medium.

SDS-PAGE and Gelatin Zymography—Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to the method of Laemmli (*16*) using 4–20% gradient polyacrylamide slab gels (Tefco, Tokyo). Gelatin zymography was performed as described by Heussen and Dowdle (*17*) with slight modifications. Electrophoresis was performed at 4°C, using 15% polyacrylamide slab gels containing 0.1% gelatin (Tefco, Tokyo). After electrophoresis, SDS was removed from the gels by washing twice for 30 min in 2.5% Triton X-100. The gels were incubated at room temperature for 3 h in the incubation buffer (100 mM sodium acetate, pH 5.5, 100 mM NaCl, 2 mM DTT, 2 mM EDTA, and 0.01% Brij-35), stained with 0.05% Coomassie Brilliant Blue R 250, and destained with 10% acetic acid.

Determination of Protein Concentration—The concentrations of the purified recombinant NsCys were determined by the method of Bradford (*18*) using bovine serum albumin as a standard. For kinetic studies of the shrimp proteinase, the molarity of the enzyme was determined by active site titration with E-64 as described by Barrett and Kirschke (*19*).

Enzymatic Assays—Routine enzyme assays were performed at 25°C using intramolecularly quenched methylcoumarylamide (MCA) substrates in a buffer containing 100 mM sodium acetate, pH 6.0, 100 mM NaCl, 2 mM DTT, 2 mM EDTA, and 0.01% Brij-35. Stock substrate solutions were prepared at concentration of 20 mM in dimethyl sulfoxide. Hydrolysis reactions were initiated by the addition of the enzyme diluted with the same buffer, and the enzymatic activities were followed by monitoring the release of the fluorogenic leaving group, MCA, at an excitation wavelength of 380 nm and an emission wavelength of 460 nm using a microplate reader

(SpectraMax Gemini spectrofluorometer; Molecular Devices). The S2 subsite specificity was investigated using various fluorogenic di- and tripeptide MCA substrates under pseudo-first-order conditions, *i.e.*, using a substrate concentration far below the estimated K_m, where the initial rate, v_0, is directly proportional to the k_{cat}/K_m value. To obtain second-order rate constants, K_m and k_{cat}, assays were performed at different concentrations of substrates covering the range of 10–200 μM with the enzyme concentration being fixed at 5 nM. The kinetic constants were obtained by non-linear regression fitting of the initial velocity data to the Wilkinson transformation (*20*) of the Michaelis-Menten equation using the EnzPack program (Biosoft, U.K.). Under all the experimental conditions, the fluorescence change was linear over the assay time (5 min). All measurements were performed in triplicate in two or more separate experiments, and the data are reported as means ± SD.

pH and Temperature Activity and Stability Profiles—The pH activity profile of the recombinant NsCys was determined at 10 μM substrate concentration under pseudo-first-order conditions as described above. The following buffers were used: 100 mM sodium citrate (pH 3.0–6.0), 100 mM sodium phosphate (pH 6.0–8.0), and 100 mM sodium borate (pH 8.0–11.0). All buffers contained 2 mM DTT, 2 mM EDTA, and 300 mM NaCl. The enzyme was incubated with these buffers for 30 min at 25°C, and the residual activities were determined using the fluorogenic substrate as described above.

To determine the effect of temperature on the Z-Pro-Arg-MCA hydrolyzing activity of NsCys, the reaction buffers containing the substrate were preincubated for 10 min at various temperatures. After addition of the enzyme solutions, the reactions were allowed to proceed for 5 min, and the fluorescence changes were recorded as described above. For thermal stability determination, the enzyme solutions were incubated at 30 to 60°C, and aliquots were withdrawn at intervals and immediately cooled on ice. Residual activities towards Z-Pro-Arg-MCA were then measured at 25°C.

Digestion of Glucagon—A sample of glucagon (1 μM) was digested with 12.5 nM of recombinant NsCys in 1 ml of 100 mM sodium acetate buffer (pH 6.0) containing 100 mM NaCl, 2 mM DTT, 2 mM EDTA, and 0.01% Brij-35 at 25°C for 4 h. The sample was then acidified with acetic acid (15%), and the resulting peptide fragments were immediately separated by reversed phase HPLC on a Tosoh CCPM equipment fitted with an ODS-120A column (25 × 0.4 cm; Tosoh, Tokyo). The column was washed with 0.1% trifluoroacetic acid in water until the baseline in absorbance at 215 nm was reached, and elution was performed with a 0–60% linear gradient of 0.1% trifluoroacetic acid in 95% acetonitrile at a flow rate of 1.0 ml/min. The eluents corresponding to each peak of absorption at 215 nm were collected, dried under vacuum and subjected to N-terminal amino acid sequencing on an Applied Biosystems protein sequencer model 476A.

Digestion of Collagen—Porcine skin acid-soluble type I collagen was diluted into 100 mM sodium acetate buffer, pH 6.0 containing 150 mM NaCl, 2 mM DTT, and 2 mM EDTA to obtain a molar concentration of 2.5 μM and incubated with 125 nM shrimp proteinase with or without 10 μM E-64. Samples were removed at predetermined intervals, immediately added to SDS-PAGE sample buffer, and boiled for 5 min. Digestion of collagen was monitored by SDS-PAGE using 4–20% gradient gel (TEFCO, Tokyo) followed by Coomassie Blue staining.

RESULTS

Isolation and Characterization of NsCys cDNA—Rapid amplification of cDNA ends (RACE) was employed to isolate the cDNA clone encoding northern shrimp cysteine proteinase (NsCys). For 3′-RACE, an initial amplification of northern shrimp hepatopancreas first-strand cDNA by PCR with primers nsCatF1 and AUAP was followed by a round of PCR with the first PCR product as a template and primers AUAP and nsCatF2, which was designed as a nested primer immediate by downstream of nsCatF1. Subsequent nucleotide sequencing of the PCR product revealed that it contained a coding region of 597 bp, a 3′-untranslated region (UTR) of 261 bp with a putative polyadenylation signal, AATAAA, and a stop codon, TAA. A BLAST search for the amino acid sequence deduced from this partial nucleotide sequence revealed that its highest homology was with members of the cathepsin L subfamily of cysteine proteinase from various organisms.

To obtain the complete nucleotide sequence of NsCys, 5′-RACE was performed using three gene-specific primers (GSP) derived from the sequence information of this cloned fragment. PCR amplification was performed with primers AUAP and GSP2 followed by nested PCR with an internal primer GSP3. A 622-bp sequence of the amplified product was found to contain a putative ATG start codon flanked by a 12-bp 5′-UTR and an overlapping region of 235 bp with the sequence obtained by 3′-RACE. As there was no difference in the overlapping region of 5′-RACE and 3′-RACE sequences, this sequence was considered to be the 5′-site of NsCys. Finally, to avoid PCR artifacts associated with overlapping RACE strategy, a gene-specific primer designed from the 5′-end of the nucleotide sequence obtained by 5′-RACE was used in a single PCR with the 3′-anchor primer, AUAP, to isolate the full-length cDNA encoding the complete precursor form of NsCys.

The 1,242-bp nucleotide sequence and the deduced amino acid sequence of the protein encoded by the single

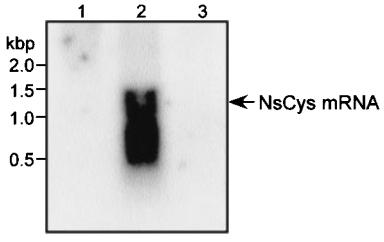

Fig. 2. **Tissue distribution of transcripts encoding northern shrimp cysteine proteinase as studied by Northern blot analysis.** Approximately 10 μg of total RNAs extracted from skeletal muscle (lane 1), hepatopancreas (lane 2), and intestine (lane 3) were subjected to electrophoresis on a 0.9% agarose gel. Subsequently, RNAs were blotted onto a nylon membrane and hybridized with a 306-bp DNA probe derived from the 3′-site of northern shrimp cysteine proteinase cDNA (891–1196 bp; see Fig. 1). The position of an RNA size marker (TaKaRa) is indicated on the left.

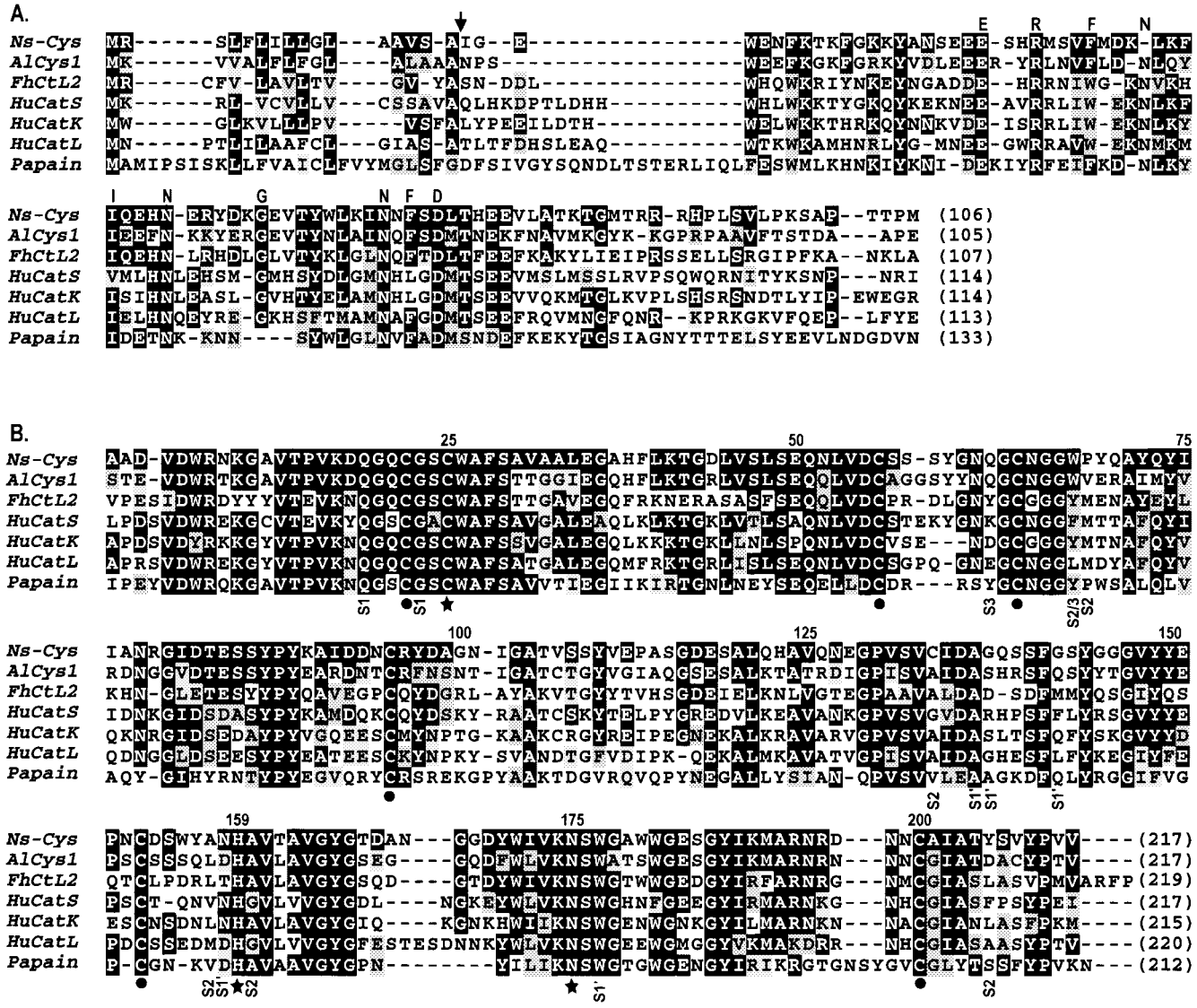

Fig. 3. **Comparison of the deduced amino acid sequence of northern shrimp cysteine proteinase with those of other cysteine proteinases.** Gaps were introduced to maximize the alignment, and proregions (A) are shown separately from mature enzymes (B) for simplicity. Conserved residues relative to northern shrimp cysteine proteinase (NsCys) in at least four sequences are highlighted: identical, similar and unrelated residues with dark, light, and white backgrounds, respectively. A: Numerals on the right margin indicate the total numbers of amino acid residues from the first methionine in the proregions. The conserved ER(W/F)NIN (*24*) and GNFD (*25*) motifs are indicated on top of the alignment. Note that the glycine residue of the GNFD motif in the papain sequence cannot be aligned because papain has fewer insertions than other sequences and thus left unaligned for simplicity. An arrowhead indicates a putative signal cleavage site in NsCys. B: The mature proteins are numbered according to the consensus papain numbering (*26*). Asterisks represent active-site residues, and filled circles mark cysteine residues of putative disulfide bridges. Amino acid residues shown to form different substrate binding pockets (*1*) are also indicated. Aligned amino acid sequences extracted from the GenBank/EMBL/DDBJ (accession numbers in parentheses) are papain (M15203); AlCys, American lobster *Homarus americanus* cysteine proteinase (X63567); FhCtL2, *Fasciola hepatica* cathepsin L2 (U62289); HuCatL, human cathepsin L (X12451); HuCatK, human cathepsin K (X82153); and HuCatS, human cathepsin S (M90696). The alignment was generated using Clustal W together with the BOXSHADE server (http://www.ch.embnet.org/software/BOX_form.html).

uninterrupted open reading frame (ORF) beginning at the first ATG are shown in Fig. 1. The ORF encodes a protein of 323 amino acids with a calculated molecular weight of 35500. The 5′-UTR prior to the ATG start codon is 12 bp long, whereas the 3′-UTR after the TGA stop codon consists of 261 bp and includes a consensus polyadenylation site, AATAAA, and run of poly(A) sequences presumably derived from the poly(A)-rich tail of mRNA. Database search for the deduced amino acid sequence revealed its homology to a number of cysteine proteinases, with higher scores for members of the cathepsin L subfamily from various vertebrate and invertebrate organisms.

Northern blot analyses of total RNAs extracted from muscle, intestine and hepatopancreas tissues with a 305-bp cDNA probe corresponding to the 3′-UTR of NsCys produced a single band at about 1.2 kb for hepatopancreas tissues, indicating that the cloned cDNA might represent the complete mRNA transcript (Fig. 2). Of note is

its exclusive expression in the hepatopancreas tissue, which suggests its possible extracellular digestive role in shrimp gut, as has been observed for other invertebrates (*5*, *6*, *21*, *22*).

Comparison of NsCys with Other Cysteine Proteinase—Multiple alignment of the shrimp proteinase with several cathepsin sequences appearing on the top of the BLAST scores allows the identification of the putative signal and activation peptides as expected for the prodomains found at the N-terminus of cysteine proteinases (*2*). Thus, NsCys contains a stretch of hydrophobic amino acids with a basic residue near the first methionine, which is typical of eukaryotic signal peptides (*23*) and predicted to be cleaved after alanine preceding a serine residue (Fig. 3). The cleavage site between the activation peptide, *i.e.*, the proregion, and the mature proteinase as shown in Fig. 3 was identified, when the peptide sequence obtained by N-terminal amino acid sequencing of the purified enzyme as mentioned above was aligned. Because these proregions show conservation in secondary structural features rather than at the primary sequence level, they show very little sequence similarity and therefore are aligned separately (Fig. 3A) from the mature proteins (Fig. 3B).

Although, as expected, the proregion of NsCys shows less homology with other cathepsins, it contains the ER(F/W)NIN motif identified by Karrer *et al.* (*24*) as a characteristic feature of members of cathepsin L subfamily (Fig. 3A). In addition, NsCys also contains the GNFD motif in the proregion, which is generally well conserved in most of the cysteine proteinases of the papain superfamily (*2*, *25*). The predicted amino acid sequence of NsCys also contains a series of residues proposed to be important in the catalytic mechanism of these enzymes, such as the active-site cysteine residue that is transiently acetylated during peptide hydrolysis, as well as the asparagine and histidine residues that form the catalytic triad of cysteine proteinases (*1–3*).

The presence of a well-conserved, so-called ERFNIN motif [$EX_3RX_2(V/I)FX_2NX_3IX_3N$ motif] distinguishes the cathepsin L subfamily, which includes cathepsins L, V, K, S, W, F, and H, from the rest of the papain family members including cathepsins B, C, O, and X (*2*, *24*). In NsCys the interspersion distance EX_3R is reduced to EX_2R and the first N of the ERFNIN motif is missing. The fact that NsCys falls within the cathepsin L subfamily is further reflected in the conserved tryptophan residue at position 6, which is invariably present in these family members (*24*). In addition, an aspartate or asparagine precedes the active site histidine in the ERFNIN group of cysteine proteinases, which is also conserved in NsCys. These observations clearly set NsCys apart from the cathepsin B subfamily, which differs distinctively at these positions. Apart from the putative catalytic triad, Cys25, His159, and Asn175 (papain numbering (*26*) is used throughout), the mature protein of NsCys contains all other residues conserved in cysteine proteinases, including Gln19 and Gly23, which comprise the oxyanion hole (*27*), as well as Trp177, which interacts closely with the P1 residue (*1*). Besides the catalytic triad residue Cys25, NsCys has six additional cysteine residues with the same structural

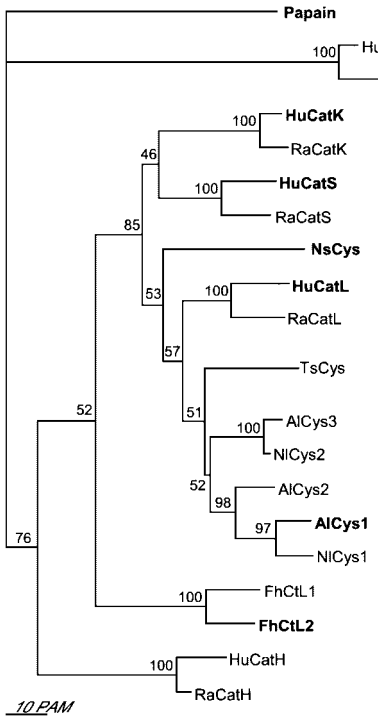

Fig. 4. **Phylogenetic analysis of northern shrimp cysteine proteinase with representative cathepsins of the papain superfamily.** Only mature proteins (*i.e.*, excluding the N-terminal preproregions) were taken for tree construction. The phylogram constructed by the neighbour-joining method was based on Clustal W-generated paired alignment of all amino acid sequences excluding gaps, with the papain sequence as an outgroup. The tree was drawn in scale with DRAWTREE, whereas distances were calculated with PROTDIST using the Dayhoff matrix and expressed in PAM, the number of amino acid substitutions per 100 positions. Bootstrap majority consensus values on 1,000 replicates were calculated with CONSENSE and are indicated in percent at each branch node. All these programs are parts of the PHYLIP package, version 3.5c and freely available at http://evolution.genetics.washington.edu/phylip.html. The shrimp proteinase (NsCys) compared with those obtained from the GenBank/EMBL/DDBJ databases (abbreviations followed by accession numbers in parenthesis) are human cathepsins B and H (HuCatB, M14221, and HuCatH, X16832, respectively), rat cathepsins B, H, K, L, and S (RaCatB, M11305; RaCatH, Y00708; RaCatK, AF010306; RaCatL, Y00697; and RaCatS, S51520, respectively), American lobster *Homarus americanus* cysteine proteinase 2 and 3 (AlCys2, X63568, and AlCys3, X63569, respectively), Norway lobster *Nephrops norvegicus* cysteine proteinase 1 and 2 (NlCys1, X80989 and NlCys2, X80990, respectively), tropical shrimp *Penaeus vannamei* cathepsins L-like cysteine proteinase (TsCys, X85127) and parasitic helminth *Fasciola hepatica* cathepsin L1 (FhCtL1, U62288). Sequences appearing in Fig. 3 and cited therein are in boldface.

topology shown to form three disulfide bridges in papain (*26*).

While the shrimp proteinase has the highest degree of identity of 64% at the amino acid level with a cysteine proteinase from American lobster *Homarus americanus*, its identity to the well-characterized mammalian cathepsins falls within a narrow range of 58, 54, and 53% respectively with human cathepsins S, L and K, for example. Thus, it appears difficult to assign the shrimp protein unequivocally to any of the types belonging to cathepsin L subfamily from sequence alignment alone

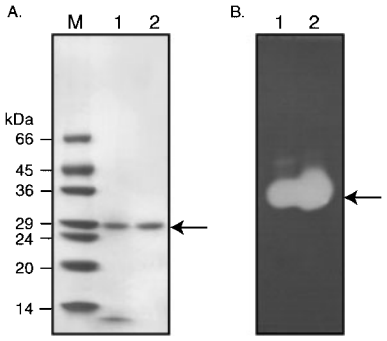

Fig. 5. **SDS-PAGE (A) and gelatin zymography (B) patterns of recombinant northern shrimp cysteine proteinase.** The cDNA encoding the complete precursor form of the shrimp enzyme was heterologously expressed in *Pichia pastoris*, and two active preparations of the recombinant enzyme were purified from the yeast culture medium. Lane 1 corresponds to a proform of NsCys, while lane 2 corresponds to the mature enzyme, which was used for enzymatic characterization in this study. M, low molecular weight marker proteins (Sigma). Each arrow indicates the mature enzyme.

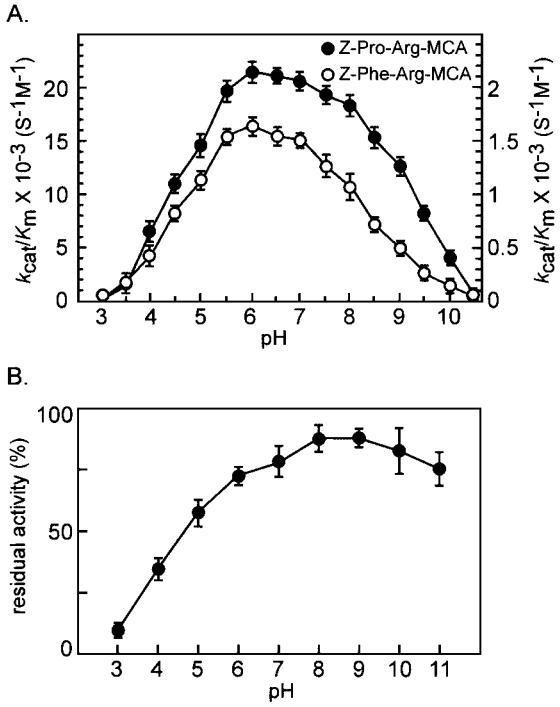

Fig. 6. **pH activity (A) and stability (B) profiles of northern shrimp cysteine proteinase.** A: The second-order rate constants (k_{cat}/K_m) for pH-dependant activities of the shrimp enzyme for the hydrolysis of Z-Pro-Arg-MCA and Z-Phe-Arg-MCA measured under pseudo-first-order conditions are shown on the left and right axis, respectively. B: pH stability was studied by monitoring the residual activities against Z-Pro-Arg-MCA substrate at pH 6.0 after incubating for 30 min in buffers of various pH values. The data for each pH treatment are expressed as percent activity relative to an untreated sample.

because of its apparent relatedness to or deviation from each member. To further address the question of its position in the cysteine proteinase superfamily, a number of different cathepsin sequences were extracted from the databases and used in phylogenetic analyses. The phylogram constructed by the neighbor-joining method shows an apparent clustering of NsCys with cathepsin L-like proteins from different invertebrate species with the same node bearing vertebrate cathepsin S (Fig. 4). However, it should be mentioned that, unlike mammalian cathepsins, the 'cathepsin L-like' designation of the invertebrate proteinases is not based on detailed enzymatic studies.

Overexpression and Purification of Recombinant NsCys—In order to generate sufficient quantities of biologically active protein for various enzymatic characterizations, a 924-bp cDNA excluding the signal peptide of NsCys was PCR-amplified and cloned into pPICZα-E downstream of the yeast α-factor secretion signal. The transfer vector was linearized with *Pme*I and electroporated into *Pichia pastoris* KM71H host cells. A *P. pastoris* clone that secreted high levels of recombinant enzyme into the culture medium was selected for large-scale expression and subsequent purification of NsCys. The recombinant shrimp enzyme was purified in two electrophoretically homogeneous fractions (Fig. 5A), both of which showed proteolytic activities by gelatin zymography (Fig. 5B). The N-terminal amino acid sequences of the 30-kDa protein in both fractions exactly matches that of the mature proteinase deduced from the gene sequence (see Fig. 1), while that of the lower band visible at the 10-kDa position in the major fraction (Fig. 5A; lane 1) corresponds to the prodomain of NsCys (data not shown). However, it must be noted that the apparent molecular mass of 30 kDa of the recombinant mature NsCys is larger than that expected from the gene sequence (~24 kDa), a difference that is commonly observed in these proteinases and reported to be independent of associated glycosylation (*28*). Although both preparations of the purified recombinant NsCys showed comparable enzymatic properties in preliminary experiments, the fraction corresponding to the mature protein (Fig. 5A; lane 2) was quantitatively and reproducibly inhibited by the prototypic cysteine proteinase inhibitor E-64 and therefore assigned for subsequent enzymatic characterization.

pH Activity Profile and pH Stability—The pH-dependent activity profile and stability are sensitive measures to understand an enzyme's functional and structural integrity. The pH dependence of the Z-Pro-Arg-MCA hydrolyzing activity, the most preferred substrate of NsCys (see below), in the range of pH 2.5 to 11.0 is shown in Fig. 6, together with that obtained against the most common substrate Z-Phe-Arg-MCA for comparison. Like typical cysteine proteinases, NsCys shows an acidic pH optimum of 6.0 for both synthetic substrates. However, it has a broader alkaline activity profile characterized by about 80% of its maximum activity at pH values as high as 8.5 against the best substrate Z-Pro-Arg-MCA (Fig. 6A). All mammalian cathepsins characterized so far are either completely inactive or show minimal activities at this alkaline pH value (*1*, *2*, *19*). To determine the pH stability of NsCys, we incubated the enzyme at various pH values at 30°C for 2 h and measured the residual activities at pH 6.0. NsCys is unstable at pH values less than 5.0 and remains fully active over a wide alkaline pH range (Fig. 6B), which is also in contrast to the stability profile of mammalian cathepsins (*2*, *19*).

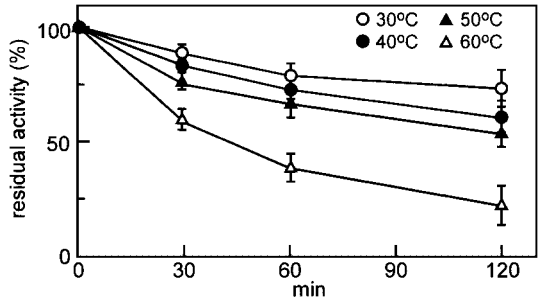

Fig. 7. **Thermal stability of northern shrimp cysteine proteinase.** The enzyme solutions were incubated at the indicated temperatures, and aliquots were withdrawn at intervals and immediately cooled on ice. Residual activities towards Z-Pro-Arg-MCA were measured at 25°C and expressed as percentage of the initial activity.

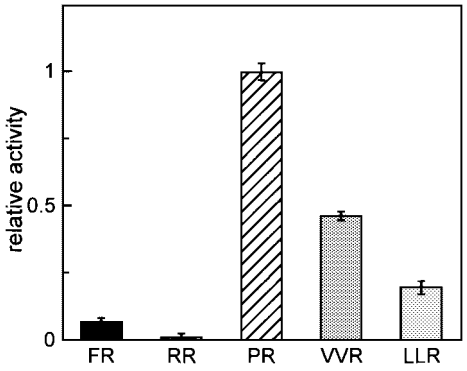

Fig. 8. **Investigation of the subsite specificity of northern shrimp cysteine proteinase using fluorogenic di- or tripeptide substrates.** Di- or tripeptide fluorogenic substrates of Z-Xaa-Xaa-Arg-MCA were used, where Xaa represents different amino acid residues as indicated with single letter codes. Specificity constants, k_{cat}/K_m, were determined under pseudo-first-order conditions as described under "MATERIALS AND METHODS." Values are normalized by assigning 1 to the best substrate.

The recombinant shrimp proteinase shows maximum activity at 40°C, with considerable activity at temperatures as low as 20°C, as would be expected for a cold-adapted enzyme (data not shown). However, unlike other cold-adapted enzymes reported so far (*9*, *10*), the shrimp enzyme is remarkably thermostable, retaining 60% of its initial activity after incubation at 50°C for 2 h (Fig. 7).

Substrate Specificity and Kinetic Parameters—The substrate specificity of NsCys was characterized with several synthetic di- or tripeptide MCA substrates varying at P2 position, since in all cysteine proteinases of the papain superfamily the so-called S2 subsites define the primary substrate specificity (*1–3*). In general, cathepsins L and S prefer bulky side-chains of hydrophobic residues phenylalanine and leucine, respectively, to the smaller β-branched valine, whereas a positively charged arginine residue results in poor substrate hydrolysis (*1*, *2*). Interestingly, the preference for leucine over valine in these proteinases is reversed in NsCys (Fig. 8) and, quite unexpectedly, NsCys exhibited more than 10-fold lower affinity for phenylalanine than proline, in sharp contrast to other known cathepsins. Although mammalian cathepsin K also prefers a proline residue at P2 position, it almost equally accepts leucine as a P2 residue (*29*) and shows considerable affinity towards phenylalanine (*30*, *31*).

The second-order rate constants (k_{cat}/K_m) of NsCys obtained for MCA substrates containing both proline and phenylalanine at different pH values are summarized in Table 1. Compared to Z-Pro-Arg-MCA, the lower specificity constants obtained for Z-Phe-Arg-MCA at all pH values examined can be attributed predominantly to about one order of magnitude lower k_{cat} values. These results suggest that the substrate containing a phenylalanine residue still interacts with NsCys but in an unproductive way resulting in possible substrate inhibition during deacylation of the enzyme-substrate complex.

Table 1. **Kinetic parameters of northern shrimp cysteine proteinase.** Kinetic constants were determined at different pH values using Z-Phe-Arg-MCA and Z-Pro-Arg-MCA as substrates at 25°C. Data are shown as means ± SD (n = 4).

Substrate	pH	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($s^{-1}M^{-1}$)
Z-Phe-Arg-MCA				
	pH 4.0	0.069 ± 0.005	118 ± 17	585
	pH 6.0	0.124 ± 0.008	78 ± 11	1,590
	pH 8.0	0.088 ± 0.007	86 ± 13	1,023
Z-Pro-Arg-MCA				
	pH 4.0	0.866 ± 0.05	98 ± 10	8,837
	pH 6.0	1.392 ± 0.11	65 ± 7	21,415
	pH 8.0	1.303 ± 0.09	70 ± 11	18,614

Hydrolysis of Glucagon and Collagen—The above results with the synthetic substrates suggest that NsCys might have very different substrate specificity from its mammalian counterparts (*29–31*). To substantiate this notion, we further investigated the substrate specificity of NsCys by studying the peptide fragments generated by digestion of the natural polypeptide substrate glucagon, which contains phenylalanine residues but lacks a proline residue. Consistent with the findings with the synthetic substrates, NsCys shows the highest preference for valine at P2 position, followed by threonine and alanine residues (Fig. 9). Furthermore, no peptide fragments derived from P2 leucine residue could be resolved under the experimental conditions, which is also in line with the previous observation that NsCys shows poor affinity towards substrate with a P2 leucine residue.

The extraordinarily high specificity of NsCys for proline residues suggests its possible involvement in collagen turnover *in vivo*. To examine this possibility, we investigated the ability of NsCys to cleave type I collagen, a triple helical protein molecule rich in proline residues. Surprisingly, incubation of type I collagen with NsCys at a substrate-to-enzyme molar ratio of 20:1 resulted in complete disappearance of collagen γ- and β-chains after 30 min, followed by extensive degradation of α-chains after 60 min, as evidenced by the increase in low molecular weight protein bands on the SDS-PAGE gel (Fig. 10). The stability of the triple helical collagen molecule at 25°C, a temperature well below its melting point, is obvious (*32*) and can also be judged from the apparent intactness of all three chains of collagens incubated as controls under the same assay conditions. Moreover, the fragmentation of collagen is not an artifact, but results from cysteine proteinase activity of NsCys during incubation,

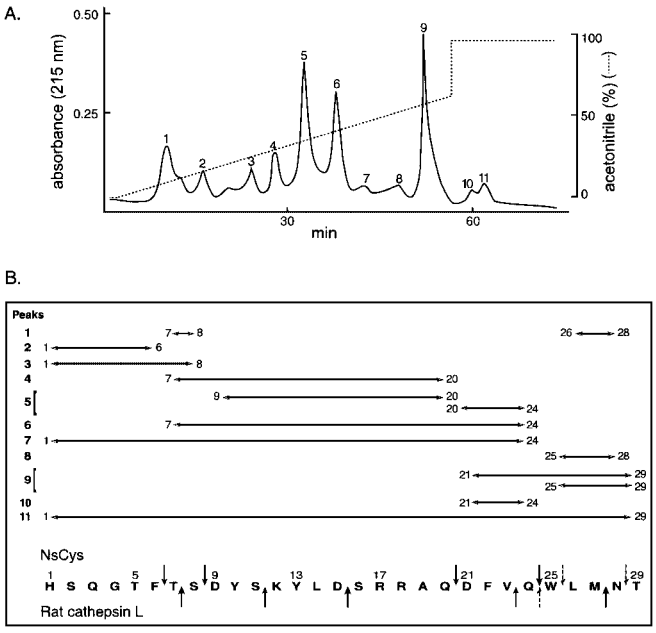

Fig. 9. **Cleavage specificity of northern shrimp cysteine proteinase towards natural polypeptide substrate glucagon.** Recombinant shrimp enzyme (12.5 nM) was incubated with 10 μM glucagon at pH 6.0 and 25°C for 2 h. The reaction was stopped with acetic acid (15%), and the resulting peptide fragments were separated by reverse phase HPLC as described under "MATERIALS AND METHODS." Glucagon fragments, as shown sequentially in the chromatogram (panel A), were subjected to N-terminal amino acid sequencing. The sequence of glucagon is shown by the single-letter amino acid code, while those determined for the HPLC peaks are represented by lines spanning the corresponding parts of glucagon (panel B). The degree of susceptibility of each cleavage site as assumed from the peak height in the chromatogram as well as from amino acid sequence analyses is indicated as major (thick arrow), moderate (thin arrow) and minor (broken arrow). For comparison, the cleavage sites reported for rat cathepsin L (*45*) are also shown.

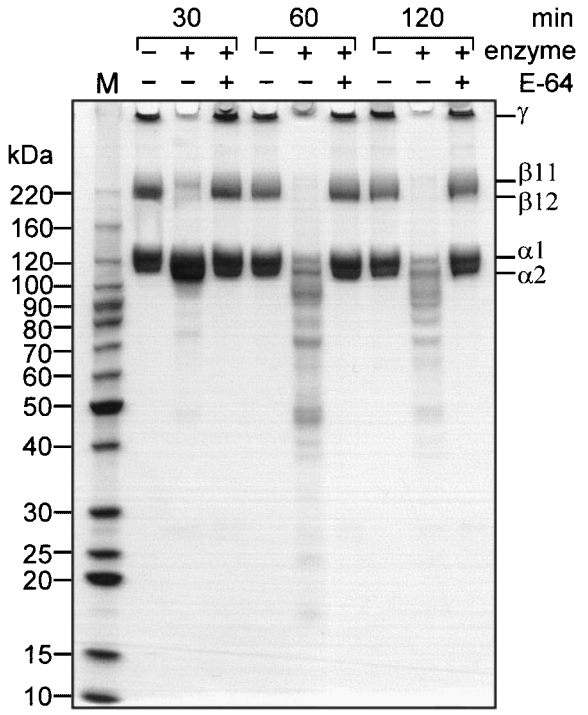

Fig. 10. **SDS-PAGE pattern of type I collagen after digestion with northern shrimp cysteine proteinase.** Porcine skin type I collagen was incubated with the shrimp enzyme at a collagen-to-enzyme molar ratio of 20:1 in the presence or absence of E-64 at pH 6.0 and 25°C. Other experimental conditions are as described under "MATERIALS AND METHODS." SDS-PAGE was performed by using 4–20% gradient concentration of acrylamide followed by Coomassie Blue staining. M, 10 kDa ladder protein size marker (Life Technologies GibcoBRL).

because no such fragments were generated in the presence of the cysteine proteinase inhibitor, E-64.

DISCUSSION

This is the first proteinase from an Arctic shrimp species, *Pandalus borealis*, which has been cloned, sequenced, produced in a heterolgous expression system, and characterized. The predicted primary structure of the shrimp proteinase revealed considerable homology to cathepsin L-like cysteine proteinases, which include, for example, cathepsins L, S, and K. Because of the extensive sequence similarity between these related cathepsins, it is not possible to unequivocally relate them with new orthologs unless this is supported by detailed biochemical characterization at the protein level. There are no extraordinary structural features that would immediately distinguish the cathepsin L-family members from each other. Rather, the particular enzymatic properties that define them are derived from subtle changes at or around the primary specificity pocket (*1–3*). In the cysteine proteinases, the S2 subsite comprises the only true substrate-binding pocket and therefore appears to be the major specificity determinant (*1*, *2*). Based on X-ray crystallographic studies of various cysteine proteinases, it has been claimed that this pocket is mostly hydrophobic and composed of residues 67, 68, 133, 157, 160, and 205 (*1*, *33*, *34*). The size and shape of the side-chains of these residues tend to determine the specificity of different cathepsins by modulating the topology of the binding pocket and fine-tuning the interaction with the incoming substrate residues. NsCys has a large aromatic residue at position 67 as in cathepsin K, but a polar cysteine residue at position 133, in contrast to alanine in cathepsins B, L, and K and glycine in cathepsin S (see Fig. 3). Hydrophobic residues of varying sizes are commonly found at positions 68, 157, and 160 (*1*, *29*), which are also conserved in NsCys. On the other hand, Tyr205 in NsCys seems interesting, as residue 205 is considered to be the single most important residue in determining the specificity preferences observed in different cathepsins (*1–3*, *29*, *35*).

These subtle yet structurally important changes in NsCys should therefore be reflected by its substrate preference. In fact, the recombinant NsCys appears to be unique among its well characterized mammalian equivalents in terms of both physical and kinetic properties. Amongst the cysteine proteinases of the papain superfamily characterized to date, only mammalian cathepsin K (*29–31*) and cathepsin L2 from parasite *Fasciola hepatica* (*36*) are known to accept a P2 proline residue. How-

ever, unlike NsCys, they also accept a phenylalanine residue with comparable affinity. Smooker et al. (*36*) showed that a single tyrosine residue at position 67 of *F. hepatica* cathepsin L2 is responsible for its unusual specificity for proline residues. The same line of experimental evidence has recently been reported for cathepsin K by Lecaille *et al*. (*29*), who showed that in addition to Tyr67, Leu205 also plays a key role in binding of proline residues. In this regard, it is worth mentioning that the residues proposed to constitute the S2 subsite are completely identical in these two cathepsins. Considering their similar specificity preferences, it is tempting to speculate that cathepsin L2 from *F. hepatica* is indeed a helminth homologue of mammalian cathepsin K (see Fig. 3). However, papain, the prototypic member of the superfamily, has the same tyrosine residue at position 67 yet shows no affinity for proline residues (*1*–*3*). Hou *et al*. (*37*) recently showed that mutation of Tyr212 (residue 97 according to papain numbering) associated with pycnodysostosis disease abolishes its collagenolytic activity. Interestingly, most cathepsins, including NsCys, also have this residue at the same position (Tyr97; see Fig. 3), while possessing different substrate specificities. It is therefore plausible that a complex interplay of other residues might also be involved in the specificity preference for proline residues. In NsCys residue 67 is tryptophan, and interestingly, the tyrosine residue makes a topological switch to occupy position 205 (Fig. 3). Thus, while the preservation of shape and nature of residue 67 in NsCys may cause its unusual specificity for P2 proline residues, the occurrence of tyrosine at the base of the S2 pocket would presumably further increase the specificity for proline. Such a structural arrangement in NsCys may also explain its unexpectedly lower preference for P2 phenylalanine and leucine residues with large side-chains, due presumably to the stearic hindrance caused by the aromatic rings of Trp67 and Tyr205.

The shrimp proteinase also contrasts with other cysteine proteinases in its physicokinetic behavior. It is less efficient in terms of k_{cat}/K_m ratio than other cysteine proteinases (*19*, *29*, *30*, *36*), although comparison is only possible for parameters measured under the same experimental conditions and for enzymes having similar specificity. Bearing in mind the limitation of interpretation, however, the kinetic parameters of other cysteine proteinases suggest that the observed lower catalytic efficiency of NsCys is due to its lower binding affinity (increased apparent K_m value). This inference is reasonable in view of the kinetic behavior of cold-adapted enzymes. At any given temperature, an enzyme from a cold-adapted species would exhibit lower affinity for a substrate than an ortholog from a warm-adapted species. However, if measured at their respective physiological temperatures, their affinities would be similar (*38*). The restricted expression of NsCys in the shrimp hepatopancreas together with its broad pH profile suggests a possible extracellular digestive function, as has been reported for similar proteinases from several other invertebrates (*5*, *6*, *21*, *22*). In this case, the second-order rate constant (k_{cat}/K_m) seems irrelevant, because at relatively high substrate concentration *in vivo*, zero-order reaction kinetics would prevail and thus the catalytic rate constant (k_{cat}) should be the best index for evaluating the enzyme's performance.

The broad pH activity and alkaline stability profiles of NsCys are quite unusual for cysteine proteinases, most of which show a bell-shaped activity profile with a sharp acidic pH optimum and are quite unstable at alkaline pH values (*1*, *29*, *30*). However, cathepsin L from several invertebrates such as *F. hepatica* (*39*) and ciliate *Paramecium tetraurelia* (*40*) tolerates a broad range of pH values. Interestingly, the residual activity of alkali-treated NsCys increased to some extent, suggesting a possible protonation of one or more reactive groups on the enzyme molecule at higher pH values. From the pH activity and stability profiles of NsCys, it is apparent that an ionizable group with a pK_a value greater than 9 shifts the activity towards alkalinity; and the precise identification and mechanism of this shift are currently under investigation. The presence of many ionizing groups has been reported for cathepsin B (*41*) and cruzipain (*42*) catalyzing the hydrolysis of different synthetic substrates, but the identity of those in the alkaline range remains unknown. In this regard, the presence of Tyr205 and Cys133 with a side-chain pK_a value of about 9 at the core of the S2 subsite in NsCys seems interesting. In particular, the side-chain of residue 133 lies closest to and points directly at the P2 side-chain (*1*, *24*). Thus the assumption that protonation of this residue at higher pH values may offset the possible deprotonation effect of catalytic His159, thereby modulating the specificity preference of the S2 pocket, would be an interesting topic for further investigation.

To the best of our knowledge, the potency of NsCys to cleave type I collagen so extensively is unparalleled by any cysteine proteinase reported so far. A 30-min incubation of collagen with NsCys at 25°C increases monomeric α chains at the expense of trimeric γ and dimeric β chains. The same period of incubation also results in weakening of the α chains, and further incubation up to 60 min results in complete disappearance of γ and β chains associated with extensive degradation of α chains (see Fig. 10). Cleavage of the collagen molecule by cathepsin K (then termed as cathepsin O2) was first demonstrated by Brömme *et al*. (*30*), and follow-up studies by Garnero *et al*. (*32*) showed that the cleavage pattern is similar to that of bacterial collagenase (*43*): *i.e*., unlike other cysteine proteinases (*44*), it can cleave within the helical region of intact collagen molecule at multiple sites. Comparison of experimental data of these reports with those of the present study, whenever possible, indicates that cathepsin K requires considerably longer incubation period (4–8 h) to degrade collagen molecules to an extent apparently comparable to that of NsCys.

The results of the present study do not allow the assignation of NsCys to any of the well-characterized members of cathepsin L group. Comparison with available data suggests that NsCys could be the shrimp homologue of mammalian cathepsin K. It differs, however, from the latter in some key structural features and notably in physicokinetic properties as discussed above. We therefore propose for the shrimp proteinase the trivial name "crustapain," to indicate that it is a papain-like cysteine proteinase from a crustacean species and to distinguish its unique structural and enzymatic properties.

This study was supported, in part, by a Grant-in-Loan from the Bio-oriented Technology Research Advancement Institution (BRAIN), an affiliation of the Ministry of Agriculture, Forestry and Fisheries, Japan, to Nichirei Corporation. M.N.A. was the recipient of a post-graduate research scholarship from the Ministry of Education, Culture, Sports, Science and Technology of Japan. Nucleotide sequence data of the northern shrimp cysteine proteinase (NsCys) are available in the DDBJ/EMBL/GenBank database under the accession no. AB091669.

REFERENCES

1. McGrath, M.E. (1999) The lysosomal cysteine proteinases. *Annu. Rev. Biophys Biomol. Struct.* **28**, 181–204
2. Turk, B., Turk, D., and Turk, V. (2000) Lysosomal cysteine proteases: more than scavengers. *Biochim. Biophys Acta* **1477**, 98–111
3. Tao, K., Streams, N.A., Dong, J, Wu, Q.L., and Sahagian, G.G. (1994) The pro region of cathepsin L is required for proper folding, stability, and ER exit. *Arch. Biochem. Biophys.* **311**, 19–27
4. Chapman, H.A., Riese, R.J., and Shi, G.P. (1997) Emerging roles for cysteine proteinases in human biology. *Annu. Rev. Physiol.* **59**, 63–88
5. Brady, C.P., Dowd, A.J., Brindley, P.J., Ryan, T., Day, S.R., and Dalton, J.P. (1999) Recombinant expression and localization of *Schistosoma mansoni* cathepsin L1 support its role in the degradation of host haemoglobin. *Infect. Immun.* **67**, 368–374
6. Lima, A.P., dos Reis, F.C., Serveau, C., Lalmanach, G., Juliano, L., Menard, R., Vernet, T., Thomas, D.Y., Storer, A.C., and Scharfstein, J. (2001) Cysteine proteinase isoforms from *Trypanosoma cruzi*, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors. *Mol Biochem. Parasitol.* **25**, 41–52
7. Lalmanach, G., Boulange, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F., and Authie, E. (2002) Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. *Biol. Chem.* **383**, 739–749
8. Li, R., Chen, X., Gong, B., Selzer, P.M., Li, Z., Davidson, E., Kurzban, G., Miller, R.E., Nuzum, E.O., McKerrow, J.H., Fletterick, R.J., Gillmor, S.A., Craik, C.S., Kuntz, I.D., Cohen, F.E., and Kenyon, G.L. (1996) Structure-based design of parasitic proteinase inhibitors. *Bioorg. Med. Chem.* **4**, 1421–1427
9. Smalås, A.O., Heimstad, E.S., Hordvik, A., Willassen, N.P., and Male, R. (1994) Cold adaption of enzymes: structural comparison between salmon and bovine trypsins. *Proteins* **20**, 149–166
10. Ahsan, M.N. and Watabe, S. (2001) Kinetic and structural properties of two isoforms of trypsin isolated from the viscera of Japanese anchovy *Engraulis japonicus*. *J. Prot. Chem.* **20**, 49–58
11. Shunway, S.E., Perkins, H.C., Schick, D.F., and Stickney, A.P. (1985) Synopsis of biological data of the pink shrimp, *Pandalus borealis* Krøyer, 1838. NOAA Technical Report NMFS 30, *FAO Fish. Synop.* **144**, 3–6
12. Schechter, I. and Berger, A. (1967) On the size of the active site in proteinases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162
13. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
14. Laycock, M.V., MacKay, R.M., Di Fruscio, M., and Gallant, J.W. (1991) Molecular cloning of three cDNAs that encode cysteine proteinases in the digestive gland of the American lobster (*Homarus americanus*). *FEBS Lett.* **292**, 115–20
15. Liu, Q., Li, M.Z., Leibham, D., Cortez, D., and Elledge, S. (1998). The univector plasmid fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* **8**, 1300–1309
16. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
17. Heussen, C. and Dowdle, E.B. (1980) Electrophoretic analysis of plasminogen activators in polyacrylamide gels containing sodium dodecyl sulfate and copolymerized substrates. *Anal. Biochem.* **102**, 196–202
18. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254
19. Barrett, A.J. and Kirschke, H. (1981) Cathepsin B, cathepsin H, and cathepsin L. *Methods Enzymol.* **80C**, 535–561
20. Wilkinson, G.N. (1961) Statistical estimation in enzyme kinetics. *Biochem. J.* **80**, 324–332
21. Johnson, K.S. and Rabosky, D. (2000) Phylogenetic distribution of cysteine proteinases in beetles: evidence for an evolutionary shift to an alkaline digestive strategy in Cerambycidae. *Comp. Biochem. Physiol.* **126B**, 609–619
22. Dowd, A.J., McGonigle, S., and Dalton, J.P. (1995) *Fasciola hepatica* cathepsin L cleaves fibrinogen and produces a novel fibrin clot. *Eur. J. Biochem.* **232**, 241–246
23. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijine, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6
24. Karrer, K.M., Peiffer, S.L., and DiTomas, M.E. (1993) Two distinct gene subfamilies within the family of cysteine proteinase genes. *Proc. Natl Acad. Sci. USA* **90**, 3063–3067
25. Vernet, T., Berti, P.J., de Montigny, C., Musil, R., Tessier, D.C., Menard, R., Magny, M.C., Storer, A.C., and Thomas, D.Y. (1995) Processing of the papain precursor. The ionization state of a conserved amino acid motif within the pro region participates in the regulation of intramolecular processing. *J. Biol. Chem.* **270**, 10838–10846
26. Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A., and Drenth, J. (1984) Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.* **179**, 233–256
27. Menard, R., Carriere, J., Laflamme, P., Plouffe, C., Khouri, H.E., Vernet, T., Tessier, D.C., and Thomas, D.Y. (1991) Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain. *Biochemistry* **30**, 8924–8928
28. Smith, S.M., Kane, S.E., Gal, S., Mason, R.W., and Gottesman, M.M. (1989) Glycosylation of procathepsin L does not account for species molecular-mass differences and is not required for proteolytic activity. *Biochem. J.* **262**, 931–938
29. Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C.S., and Brömme, D. (2002) Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity. *Biochemistry* **41**, 8447–8454
30. Brömme, D., Okamoto, K., Wang, B.B., and Biroc, S. (1996) Human cathepsin O2, a matrix protein-degrading cysteine proteinase expressed in osteoclasts. *J. Biol. Chem.* **271**, 2126–2132
31. Bossard, M.J., Tomaszek, T.A., Thompson, S.K., Amegadzie, B.Y., Hanning, C.R., Jones, C., Kurdyla, J.T., McNulty, D.E., Drake, F.H., Gowen, M., and Levy, M.A. (1996) Proteolytic activity of human osteoclast cathepsin K. *J. Biol. Chem.* **271**, 12517–12524
32. Garnero, P., Borel, O., Byrjalsen, I., Ferreras, M., Drake, F.H., McQueney, M.S., Foged, N.T., Delmas, P.D., and Delaisse, J.M. (1998) The collagenolytic activity of cathepsin K is unique among mammalian proteinases. *J. Biol. Chem.* **273**, 32347–32352
33. McGrath, M.E., Klaus, J.L., Barnes, M.G., and Bromme, D. (1997) Crystal structure of human cathepsin K complexed with a potent inhibitor. *Nat. Struct. Biol.* **4**, 105–109
34. Fujishima, A., Imai, Y., Nomura, T., Fujisawa, Y., Yamamoto, Y., and Sugawara T. (1997) The crystal structure of human cathepsin L complexed with E-64. *FEBS Lett.* **407**, 47–50
35. Gillmor, S.A., Craik, C.S., and Fletterick, R.J (1997) Structural determinants of specificity in the cysteine proteinase cruzian. *Protein Sci.* **6**, 1603–1611
36. Smooker, P.M., Whisstock, J.C., Irving, J.A., Siyaguna, S., Spithill, T.W., and Pike, R.N. (2000) A single amino acid substi-

tution affects substrate specificity in cysteine proteinase from *Fasciola hepatica*. *Protein Sci.* **9**, 2567–2572
37. Hou, W., Brömme, D., Zhao, Y., Mehler, E., Dushey, C., Weinstein, H., Miranda, C.H., Fraga, C., Greig, F., Carey, J., Rimoin, D.L., Desnick, R.J., and Gelb, B.D. (1999) Characterization of novel cathepsin K mutations in the pro and mature polypeptide regions causing pycnodysostosis. *J. Clin. Invest.* **103**, 731–738
38. Hochachka, P.W. and Somero, G.N. (2002) *Biochemical Adaptation: Mechanasims and Processes in Physiological Evolution*. Oxford University Press, Oxford, NY
39. Roche, L., Dowd, A.J., Tort, J., McGonigle, S., McSweeney, A., Curley, G.P., Ryan, T., and Dalton, J.P. (1997) Functional expression of *Fasciola hepatica* cathepsin L1 in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **245**, 373–80
40. Völkel, H., Kurz, U., Linder, J., Klumpp, S., Gnau, V., Jung, G., and Schultz, J.E. (1996) Cathepsin L is an intracellular and extracellular proteinase in *Paramecium tetraurelia*. Purification, cloning, sequencing and specific inhibition by its expressed propeptide. *Eur. J. Biochem.* **238**, 198–206
41. Khouri, H.E., Plouffe, C., Hasnain, S., Hirama, T., Storer, A.C., and Menard, R. (1991) A model to explain the pH-dependent specificity of cathepsin B-catalysed hydrolyses. *Biochem. J.* **275**, 751–757
42. Serveau, C., Lalmanach, G., Hirata, I., Scharfstein, J., Juliano, M.A., and Gauthier, F. (1999) Discrimination of cruzipain, the major cysteine proteinase of *Trypanosoma cruzi*, and mammalian cathepsins B and L, by a pH-inducible fluorogenic substrate of trypanosomal cysteine proteinases. *Eur. J. Biochem.* **259**, 275–280
43. French, M.F., Mookhtiar, K.A., and van Wart, H.E. (1987) Limited proteolysis of type I collagen at hyperreactive sites by class I and II *Clostridium histolyticum* collagenases: complementary digestion patterns. *Biochemistry* **26**, 681–687
44. Kirschke, H., Kembhavi, A.A., Bohley, P., and Barrett, A.J. (1982) Action of rat liver cathepsin L on collagen and other substrates. *Biochem. J.* **201**, 367–372
45. Kargel, H.J., Dettmer, R., Etzold, G., Kirschke, H., Bohley, P., and Langner, J. (1981) Action of rat liver cathepsin L on glucagon. *Acta Biol. Med. Ger.* **40**, 1139–1143