

Cloning and characterization of a novel nuclease from shrimp hepatopancreas, and prediction of its active site

Wen-Yi WANG*, Shwu-Huey LIAW† and Ta-Hsiu LIAO*¹

*Institute of Biochemistry, College of Medicine, Taiwan University, 1, Sec. 1, Jen-Ai Road, Taipei, Taiwan, and †Department of Life Science, National Yang-Ming University, 155, Sec. 2, Li-Nong Street, Taipei, Taiwan

Approximately 95% of the amino acid sequence of a shrimp (*Penaeus japonicus*) nuclease was derived from protease-digested peptides. A 1461-base cDNA for the nuclease was amplified and sequenced with degenerate primers based on the amino acid sequence and then specific primers by 3' and 5' RACE (rapid amplification of cDNA ends). It contains an open reading frame encoding a putative 21-residue signal peptide and a 381-residue mature protein. The N-terminus of the enzyme is pyroglutamate, deduced from composition and matrix-assisted laser desorption ionization–time-of-flight MS analyses, and confirmed by a glutamine residue in the cDNA sequence. The enzyme has 11 Cys residues, forming five intramolecular disulphides. The eleventh Cys residue was linked to a thiol compound with an estimated molecular mass of between 500 and 700 Da. A sequence similarity search revealed no homologous proteins but residues 205–255

shared a conserved active-site motif within a distinct group of nucleases. His²¹¹ in this conserved motif was shown to be very important in catalysis by site-specific modification with ¹⁴C-labelled iodoacetate. The shrimp nuclease, previously designated DNase I, does indeed possess a low level of hydrolytic activity towards RNA in the presence of Mg²⁺ and Ca²⁺. The conservation of functionally important residues during distant evolution might imply that the catalytic mechanisms are similar in these nucleases, which should be classified in one subfamily. Finally, an active-site structure for shrimp nuclease was proposed on the basis of published structural data and the results of mutational and biochemical analyses of *Serratia* nuclease.

Key words: cDNA cloning, mixed disulphide, N-terminal pyroglutamate, nuclease active site.

INTRODUCTION

Nucleases, in a broad sense, are enzymes capable of hydrolysing nucleic acids, including not only DNase and RNase but also the sugar-non-specific nucleases that have been named nucleases for short. DNase I, cleaving double-stranded DNA to yield 5'-phospho polynucleotides in the presence of bivalent metal ions under neutral or alkaline conditions [1], occurs in many tissues of animals and plants and also in micro-organisms. Bovine pancreatic DNase I (EC 3.1.21.1) is the DNase I that has been studied most thoroughly [2]. Interestingly, cyanobacterial [3], bacterial [4], fungal and mitochondrial nucleases [5] also contain a DNase activity similar to that of bovine DNase I. However, the sequences of these nucleases are distinct from those of DNases I of bovine and other vertebrates [6–11]. To understand the evolutionary relationships of the bovine DNase-I-like proteins, we investigated DNase I in species of greater evolutionary separation: shrimp DNase I was therefore purified and characterized [12]. Although its enzymic properties are typical of DNase I, there were several distinctive structural features, such as a lack of glycosylation, high Cys content, high molecular mass and protease resistance [12,13]. These features prompted us to study its structural and functional relationships.

In the present study, we determined the protein and cDNA sequences for shrimp DNase I. Surprisingly, the sequences exhibited no similarity to those of the bovine DNase-I-like proteins. However, in the motif of residues 205–255, a number of identical residues were found when aligned with the conserved active sites of several nucleases. This finding has prompted us to investigate the intrinsic RNase activity in shrimp DNase I. In the presence of Mg²⁺ and Ca²⁺, shrimp DNase I did indeed exhibit

RNase activity. Therefore shrimp DNase I, which was named incorrectly, is referred to as shrimp nuclease throughout the rest of this paper.

EXPERIMENTAL

Materials

Live shrimp (*Penaeus japonicus*) were obtained from the local market. Calf thymus DNA and dithiothreitol were purchased from Sigma. Shrimp nuclease was purified [12] and its DNase I activity was assayed as described previously [14,15]. Trypsin [treated with tosylphenylalanylchloromethane ('TLCK')] and chymotrypsin (TLCK-treated) were from Worthington. Iodoacetic acid and 2-mercaptoethanol were obtained from Wako. Iodo[2-¹⁴C]acetic acid (50 μ Ci/ μ mol) was from Amersham.

Protein sequencing

Determination of amino acid composition, peptide cleavage and separation were performed as described previously [7,16]. Except for disulphide pairing, the protein was reduced and S-carboxymethylated before digestion with protease [17]. Peptides were sequenced on an Applied Biosystems Sequencer, model 477A.

RNA isolation and reverse-transcriptase-mediated PCR

Total RNA was isolated from 0.4 g of hepatopancreas of live shrimp by the guanidinium/phenol/chloroform method [18]. A cDNA library was synthesized from total RNA with oligo(dT) primers and Moloney-murine-leukaemia virus reverse trans-

Abbreviations used: MALDI–TOF–MS, matrix-assisted laser desorption ionization–time-of-flight MS; RACE, rapid amplification of cDNA ends.

¹ To whom correspondence should be addressed (e-mail thliao@ccms.ntu.edu.tw).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AJ1334337.

criptase (Novagen). The PCR consisted of one initiation cycle (5 min at 95 °C), 35 amplification cycles (1 min at 95 °C, 1 min at 50 °C and 1 min at 72 °C) and one termination cycle (7 min at 72 °C).

Rapid amplification of cDNA ends (RACE)

By using the cDNA library with a 5'-degenerate primer 5'-CCNCAYGARCARTAYTAYTTYGC-3' (sense; shrimp nuclease residues 201–208) and a 3'-degenerate primer 5'-TCR-TTNACRTCRAARTCDATCCA-3' (anti-sense; shrimp nuclease residues 342–349), a fragment of shrimp nuclease cDNA was obtained. On the basis of the nucleotide sequence of this cDNA fragment, specific primers were then synthesized for 3' and 5' RACE (Life Technologies). For 3' RACE, the primer 5'-GGAATTCCTGACCTTGATGACATCAATG-3' (sense; shrimp nuclease residues 272–280) was used. For 5' RACE, the primer 5'-ATCTCCACCGGGTTGCCATTGA-3' (anti-sense; shrimp nuclease residues 278–285) was used for first-strand synthesis and the nested primer 5'-GGAATTCCTAGGTCGCGTGTGCAAACCTC-3' (anti-sense; shrimp nuclease residues 249–256; underline indicates *EcoRI* cloning site) was used in PCR amplification.

Gene cloning and DNA sequencing

The PCR products of 5' and 3' RACE were purified with GeneClean III (BIO 101). The recovered DNA fragments were treated with *SalI* and *EcoRI* (Boehringer Mannheim) and ligated into pBS(+) plasmid (Stratagene). This inserted plasmid was transformed into *Escherichia coli* strain TG1. Recombinant plasmids were isolated and sequenced on a DNA sequencer (Perkin-Elmer, model 373).

Disulphide pairing

Intact shrimp nuclease (1 mg) was treated with CNBr in 70% (v/v) formic acid for 24 h at 25 °C [6] and the products were separated by gel filtration on a Sephadex G-75 column (0.9 cm × 67 cm) eluted with 50% (v/v) acetic acid. Two fragments were obtained and each was digested further with chymotrypsin (2 µg). The digests were separated on HPLC; Cys-containing peptides were identified. They were reduced with 2 mM dithiothreitol in 0.05 M Tris/HCl, pH 8.0, and chromatographed again on HPLC. Each separated into two reduced Cys-containing peptides, which were identified on the basis of results of amino acid composition and sequencing. The two reduced Cys-containing peptides were thus assigned for each of the paired disulphides.

RNase assay

The assay method was essentially similar to that for RNase T1. In brief, each microcentrifuge tube contained 50 µl of 0.2 M Tris/HCl, pH 7.5, 20 µl of 25 mM CaCl₂, 5 µl of 0.25 M MgCl₂, 75 µl of water and 25 µl of enzyme. After the mixture had been prewarmed at 37 °C for 5 min, 62.5 µl of freshly prepared baker's yeast RNA (1.2%, w/v) was added and the incubation continued. After exactly 15 min, 0.1 vol. of 7.5 M ammonium acetate and 3 vol. of ethanol were added and the entire solution was cooled to -20 °C for at least 30 min, then centrifuged. The A₂₆₀ of the supernatant was measured. One unit of RNase activity was defined as an increase of one absorbance unit under these conditions.

RNase activity staining *in situ*

The zymogram method was used as described by Rosenthal and Lacks [19], with minor modifications. Baker's yeast RNA (Worthington) was added to a final concentration of 160 µg/ml in the separating gel solution. After electrophoresis, the gel was soaked in 100 ml of 10 mM Tris/HCl, pH 7.0, containing 5 mM MgCl₂ at 25 °C for 30 min with gentle shaking. This process was repeated twice. Finally, the gel was transferred to fresh buffer also containing 2 mM CaCl₂ at 37 °C for 30 min and stained with ethidium bromide. RNase activity was revealed as dark bands in the gel. SDS/PAGE was performed by the method of Laemmli [20].

RESULTS

Amino acid and cDNA sequences of shrimp nuclease

Approximately 95% of the amino acid sequence of shrimp nuclease was derived from protease-digested peptides and agreed well with the cDNA-translated sequence (Figures 1 and 2). The uncertain parts of the protein sequence were supplemented by

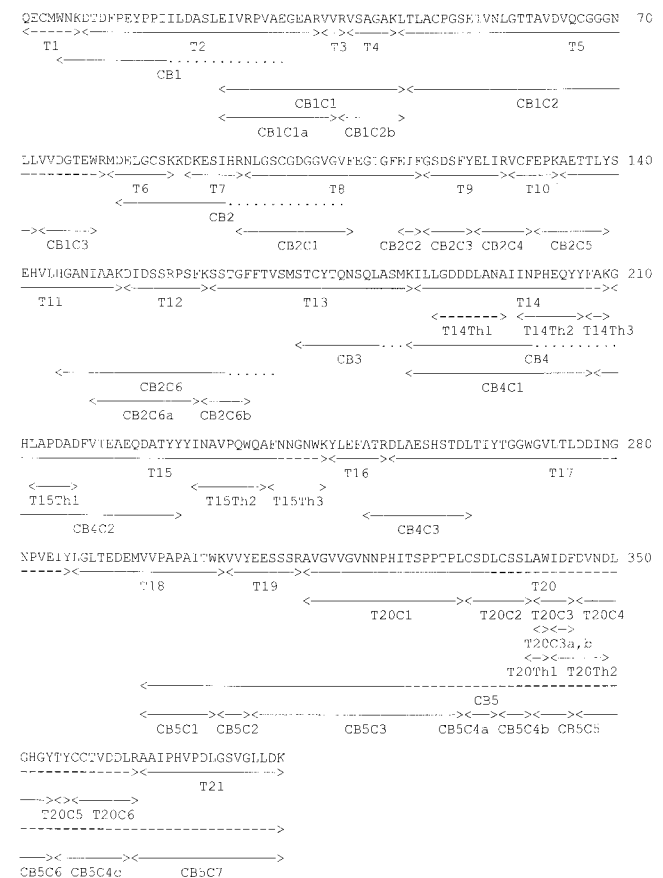


Figure 1 Complete amino acid sequences of shrimp nuclease

Reduced and S-carboxymethylated nuclease was digested by trypsin. The large tryptic peptides were digested further by thermolysin or chymotrypsin. The intact nuclease cleaved with CNBr and its products were digested by chymotrypsin. Designations: T, tryptic peptide; Th, thermolytic peptide; C, chymotryptic peptide; CB, CNBr peptide. Daughter peptides are designated with letters after the names of the parental peptides. Solid lines, sequences completed by Edman degradation; broken lines, sequences by composition only; dotted lines, incomplete sequencing by Edman degradation.

CTC AGT CAA CCA AGA ATG GCG GGA TTT GGC CTC CAG GAT TTC TTC ATT GTG ACT CTC CTT GGG GTC GGA GTT ACT 75
M A G F G L Q A F F I V T L L G V G V T 26
 GGC CAA GAC TAC ATC TGG AAT AAG GAC ACC GAC TTC CCC GAA TAC CCA CCC ATC ATT CTC GAT GCT TTC CTT GAG 150
G Q E C M W N K D T D F P E Y P P I I L D A S L E 45
 ATC GTC AGC CCC GTC GKH GAA GGA GAG GCG AGA GGG GTC GGA GFA TGA GGG GGA GTC AMG TTG ACC CTC GGG TGC 225
I V R P V A E G E A R V V R V S A G A K L T L A C 70
 CCC GCT AGC TAA ATA CTT AAT CTT GAA ACC ACA GTC GGG GAC GTG CAG TGC GGC GGA GGA AAT CTT CTT GTT GTC 300
P G S E I V N L G T T A V D V Q C G G G N L L V V 85
 GAT GGG ACT GAA TGG GGA ATG GAC GAA CTC GGG TCC ACC AAA AAG GAC AAG GAA TCG ATT CAC CTC AAC CTT GGG 375
D G T E W R M D E L G C S K K D K E S I H R N L G 120
 AGC TGC GAG GAC GGC GGT GTC GGG GTC GAA GGG ATC GGA TTC GAG ATC TTC GGT TCC GAT AGC TTT TAC GAA 450
S C G D G G V G V F E G I G F E I F G S D S F Y E 165
 CTG ATT CXC TCT TTT TFC GAA CCC AAA GCG GAG ACC AGC TTA TAT TCC GAG CAC CTC CTC GAC GGG GCC AAC ATC 525
L I R V C F E P K A E T T L Y S E H V L H G A N I 170
 GGG GAG AAG GAC ATC GAC TCC TCC AGG CCA TGG TTC AAG TGT TCC ACA GAA TTC TTC ACT GGG TCC ATG TCC ACC 600
A A K D I D S S R P S F K S S T G F F T V S M S T 295
 TGC TAC ACC GAG AAT TCG CAG CTC GCA TGG ATG AAG ATT CTT CTG GGA GAT GAT GAC CTC GGC AAC GCC ATC ATC 675
C Y T Q N S Q L A S M K I L L G D D D L A N A I I 220
 AAC CCC CAC GAA CAA TAT TAT TTC GCC AAA GGG CAC CTG GCT CXC GAC GCA GAC TTC GGG ACG GAG GCG GAG GAA 750
N P H E Q Y Y F A K G H L A P D A D F V T E A E Q 245
 GAT GCG ACG TAC TAC TAC AAT AAT GTC GGT GAT CAG TGG CAG GAC TTC AAT AAC GGC AAC TGG AAG TAC CTA GAG 825
D A T Y Y Y I N A V P Q W Q A F N N G N W K Y L E 270
 TTT GCA ACA GGC GAC CTA GTC GAA TCC CAC AGC ACT GAC CAG ACC ATC TAT ACT GGC GGC TGG GGC GTC CAG ACC 900
F A T R D L A E S H S T D L T I Y T G G W G V I T 285
 TGG GAT GAC ALC AAT GGC AAC CCG GTG GAG ATT TAC CTT GGG TTG ACT GAG GAC GAG ATG GPT GTC CCC GCA CCG 975
L D D I N G N P V E I Y L G L T E D E M V V P A P 320
 GCT ATC ACT TGG AAG GTC GTG TAC GAG GAA AGC AAT TCT CCA GCC GTG GGC GTC GTG GCG GGG AAC AAC CCA CAC 1050
A I T W K V V Y E E S S S R A V G V V G V N N P H 485
 ATC ACA GGC CCG CCC ACT CCA CTG TTT TCG GAC CTG TCC AGT TCT CTC CCG TGG ATT GAC TTC GAC GTG AAC GAC 1125
I T S P P T P L C S D L C S S L A W I D F D V N D 370
 CTC GAC CAC GGC TAC TTT TAC TGT TGT ACC GTC GAT GAC CTA AAG GCG GGT ATA CCA GAT GFA CTF GAC CTC GGG 1300
L G H G Y T Y C C T V D D L R A A I P H V P D L G 395
 TGG GPT GAC CTT GTG GAC AAG TGA ATA TGG GGC TAT TTC CCC AAA TTC AAC AAA TCA ACA ACA AAT TGG AAT GGG 1275
S V G L L D K * 402
 CAA AHA AGG AAT GAA GAA AAC GAA AGA AAT TAA ACA CAG CCA GAG TTG TGG TGG GGA TGT TTC AIT TAA TTT CTT 1350
 CAG TTC TCG GAA TTT AIG GAA AGT TCG TGT GAT TAC GAA GGC TCA GFA TTT TTC CTT AAT GAT GFA GCA ATA TGA 1425
 ATT ATA GAT GAT GAT GAA AAA AAA AAA AAA AAA 1461

Figure 2 Nucleotide and deduced amino acid sequences of shrimp nuclease

Numbers at the right relate to the last nucleotide on each line. The translated protein sequence is shown under the nucleotide sequence. The asterisk indicates the stop codon. The putative signal peptide of 21 amino acid residues is underlined.

that of the cDNA. The sequences of the 5' and 3' RACE fragments were connected by an overlap to deduce a 1461 bp polynucleotide for the cDNA of shrimp nuclease (Figure 2). It has a 5' untranslated region, an open reading frame and a 3' untranslated region followed by a poly(A) tail. The translated amino acid sequence contained 402 amino acid residues, including the mature enzyme of 381 residues and a putative 21-residue signal peptide.

Determination of the blocked N-terminus

It was not possible to determine the N-terminal sequence of either the native or the denatured shrimp nuclease by Edman degradation on the sequencer. During the sequencing of the tryptic peptides, one of the peptides (T1) also failed to be sequenced. The amino acid composition and the cDNA sequence suggested that peptide T1 should be assigned to the N-terminus and that its calculated molecular mass was 997.25 Da. The actual molecular mass of peptide T1 measured by matrix-assisted laser desorption ionization-time-of-flight MS (MALDI-TOF-MS) was 980.61 Da (Figure 3). On the basis of these results, cyclization of Gln to form pyroglutamate at the N-terminus of shrimp nuclease is a reasonable deduction.

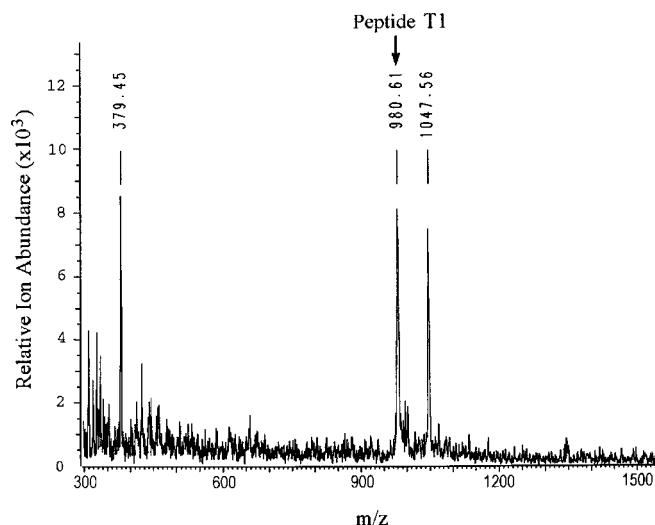


Figure 3 Mass spectrum of tryptic peptide T1 at the N-terminus

The molecular mass of peptide T1 is shown as 980.61 Da with two internal standards, α -cyano-4-hydroxycinnamic acid (379.45 Da) and angiotensin II (1047.56 Da) with a MALDI-TOF mass spectrometer (Bruker-Franzen Biflex).

Pairing of disulphide bonds

The shrimp nuclease sequence revealed eleven Cys residues. In 8 M urea without 2-mercaptoethanol, treatment of shrimp nuclease with iodoacetate showed no carboxymethyl Cys residues, but after reduction and S-carboxymethylation in 8 M urea, 11 carboxymethyl Cys residues were obtained (results not shown), indicating that all of the eleven Cys residues are involved in disulphide bonds. The first three disulphide bonds (Cys³-Cys¹²⁹, Cys⁴⁹-Cys⁶⁶ and Cys⁸⁶-Cys¹⁷⁵) were paired, on the basis of identification of the Cys-containing chymotryptic peptides obtained from the large CNBr fragment (Figure 4). Pairing for the last four Cys (Cys³³³, Cys³³⁷, Cys³⁵⁷ and Cys³⁵⁸) poses a rather difficult problem because no protease can digest the bond between Cys³⁵⁷ and Cys³⁵⁸. A peptide isolated from the chymotryptic digest of the CNBr fragment (CB5 in Figure 1) contained all these four Cys residues. The sequencing results for this chymotryptic peptide showed that the pairings would be either Cys³³³-Cys³⁵⁷; Cys³³⁷-Cys³⁵⁸ or Cys³³³-Cys³⁵⁸; Cys³³⁷-Cys³⁵⁷.

Except for five intramolecular disulphide bonds mentioned above, the eleventh Cys residue must form a disulphide with an unknown thiol compound. We treated the enzyme with 2-mercaptoethanol in a buffer without denaturing agent, followed by S-carboxymethylation with iodo[2-¹⁴C]acetate. It was still

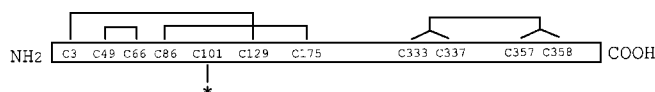


Figure 4 Disulphide bridges in shrimp nuclease

Five disulphide bridges are shown; the eleventh Cys connecting to a small thiol compound (asterisk) is Cys¹⁰¹. The pairings for the last four Cys residues (Cys³³³, Cys³³⁷, Cys³⁵⁷ and Cys³⁵⁸) at the C-terminus were uncertain.

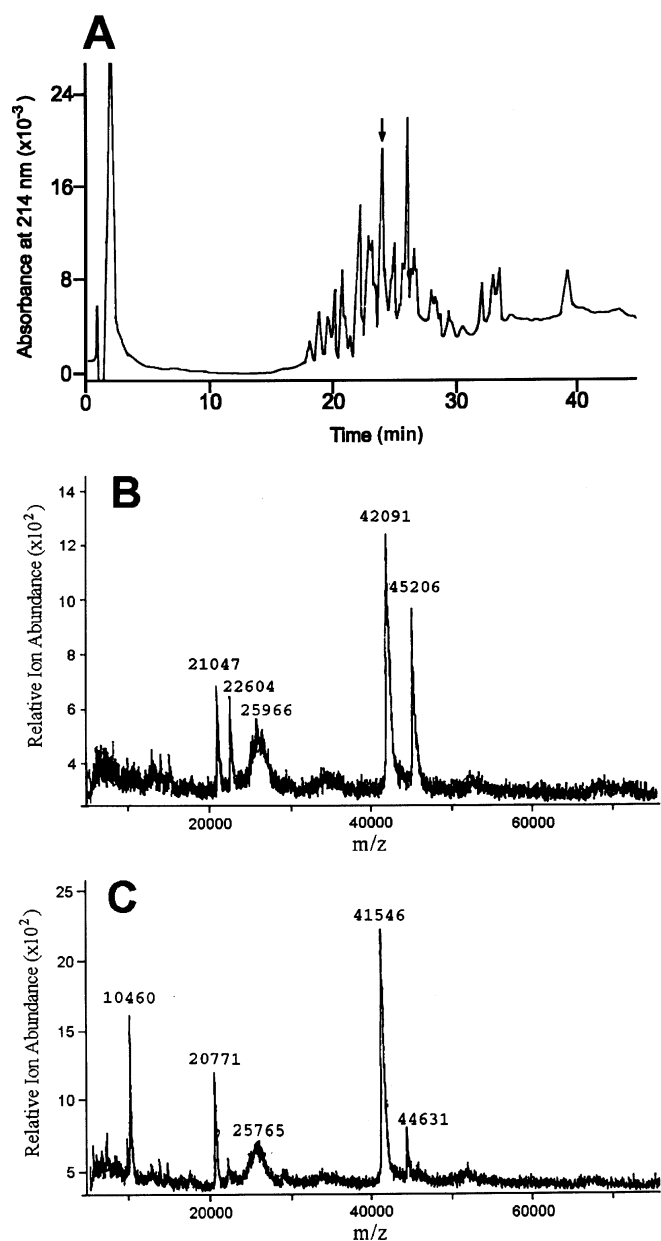


Figure 5 Identification of the 11th Cys residue and its cross-linked thiol compound

(A) HPLC profile of the radiolabelled peptide. Shrimp nuclease was reduced with 2-mercaptoethanol (9 mM) in 0.1 M Tris/HCl, pH 8.0, for 30 min and then alkylated with iodo[2- 14 C]acetate (10 mM). The labelled enzyme after digestion with trypsin for 2 h was subjected to gel filtration on a Sephadex G-25 column (0.4 cm \times 70 cm). The radioactive fractions collected were digested further with thermolysin and the products were run through the same Sephadex G-25 column. The 14 C-containing fractions were then analysed by HPLC. The arrow indicates the elution position for the 14 C-containing peptide. (B, C) MALDI-TOF-MS of native shrimp nuclease (B) and the reduced and alkylated shrimp nuclease (C). The reduction and alkylation procedure was the same as that in (A) except that unlabelled iodoacetate was used. The difference between the two molecular masses (42 091–41 546 Da) plus the molecular mass of acetate minus 1 (for a proton) is 604 Da. Because the error of MALDI-TOF-MS is approx. 0.1%, the estimated molecular mass for the thiol compound attached to the native shrimp nuclease is between 500 and 700 Da. The polypeptide with ten Cys residues cross-linked, one Cys residue free and a blocked pyroglutamate at the N-terminus has a calculated molecular mass of 41 425 Da.

fully active but with 1 mol of 14 C incorporated per mol of the enzyme. The labelled enzyme was then digested with proteases and subsequently analysed on HPLC (Figure 5A). The

```

Shp  FAKGH LAPDAD-FVTEAEQDATYYY-INAVPQWQA-FNNGNWKYIEFATRDLAESH 259
Sem  VDRGHQAPLAS-LAGVSDWESLNYL-SNITPQKSD-LNQGAWARLEDDQERKLI DRA 137
Hum  FDRGHLAAAANHRWSQKAMDPTFYL-SKVA PQVFI-LNQAMHNLKYSRSLTFSY 190
Cal  FDRGHAAAGNHRWSQKAMDPTFYL-SKVA PQVFI-LNQAMHNLKYSRSLTFSY 643
Cuc  FDRGHAAAGNHRWSQKAMDPTFYL-SKVA PQVFI-LNQAMHNLKYSRSLTFSY 137
Syr  VDRGHQAPADP-DISSQAMQDPTFYL-SNMA PQVGFNRHYWAYLEGGEMDRLTQNE 205
Scp  VDRGHQVPAADCKFSSQAMNPTFYL-SNMC PQVGDGFRHYWAYLEGGEMDRLTQNE 181
Bob  YDRGHIVSSADMSFSEAMKDTFYL-SNMS PQKSE-FNSGILKLEKLVREMAISK 94
Trb  LSRGHLAAACFHKSSSTVELAQTFHNNANTV PQ-DMTQNAVDFLRLNLRKLRYY 271
Ana  YDRGH TAPSADRTKTTEDNAATFFIM-FNMM PQTPDN-NRNTWGNLEDDYCRELVSQG 173

```

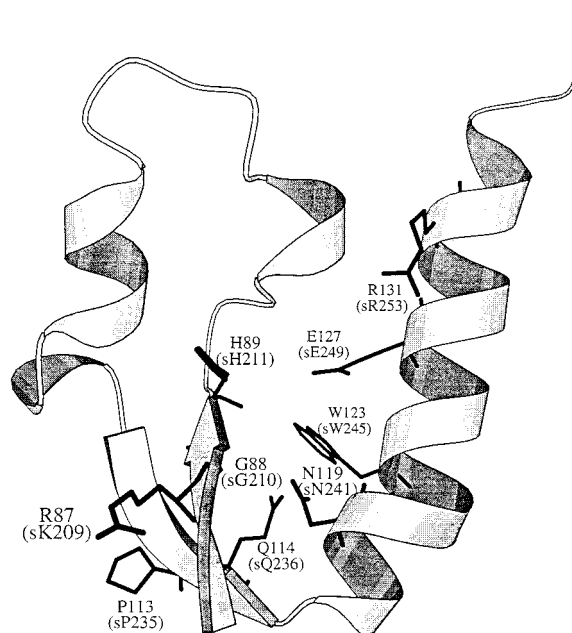


Figure 6 Alignment of the conserved motifs with selected nucleases from different organisms and proposal of the structure of the shrimp nuclease active site

Upper panel: amino acid sequence alignment showing a conserved motif within a few nucleases. Numbers at the right relate to the last residue on each line. Consensus residues are shaded. Abbreviations: Shp, shrimp; Sem, *Serratia marcescens*; Hum, *Homo sapiens*; Cal, *Caenorhabditis elegans*; Cuc, *Cunninghamella cuneae*; Syr, *Syncephalastrum racebaena*; Scp, *Schizosaccharomyces pombe*; Hep, *Helicobacter pylori*; Bob, *Borrelia burgdorferi*; Trb, *Trypanosoma brucei*; Ana, *Anabaena sp.* Lower panel: the active site of *Serratia* nuclease, drawn by the MOSCRIPT program. The side chains of the highly conserved residues are shown by sticks. The residues of *Serratia* nuclease are illustrated and the parentheses indicate the corresponding residues in shrimp nuclease. The potential functions of these residues were proposed on the basis of the *S. marcescens* studies (see the text).

sequencing results showed a peptide, NLGScmCGDGGV (in which cmC represents carboxymethyl Cys), which indicated that Cys¹⁰¹ of shrimp nuclease was radiolabelled. On the basis of MALDI-TOF-MS analysis (Figures 5B and 5C), the thiol compound linking to the 11th Cys at position 101 has an estimated molecular mass between 500 and 700 Da.

Conserved nuclease active site

Screening of the sequence databases at the NCBI (NIH) with the shrimp nuclease sequence by using the BLASTP program revealed no homologous genes. However, residues 205–255 in shrimp nuclease shared some similarity with residues 83–134 in *Serratia* nuclease. Subsequent screening of the *Serratia* nuclease revealed that these residues were a significantly conserved motif within a distinct group of nucleases occurring in species ranging from animals to micro-organisms (Figure 6, upper panel). Among these nucleases, *Serratia* nuclease was the most extensively studied. The crystal structure, site-directed mutagenesis and enzyme kinetic analysis of *Serratia* nuclease provided strong

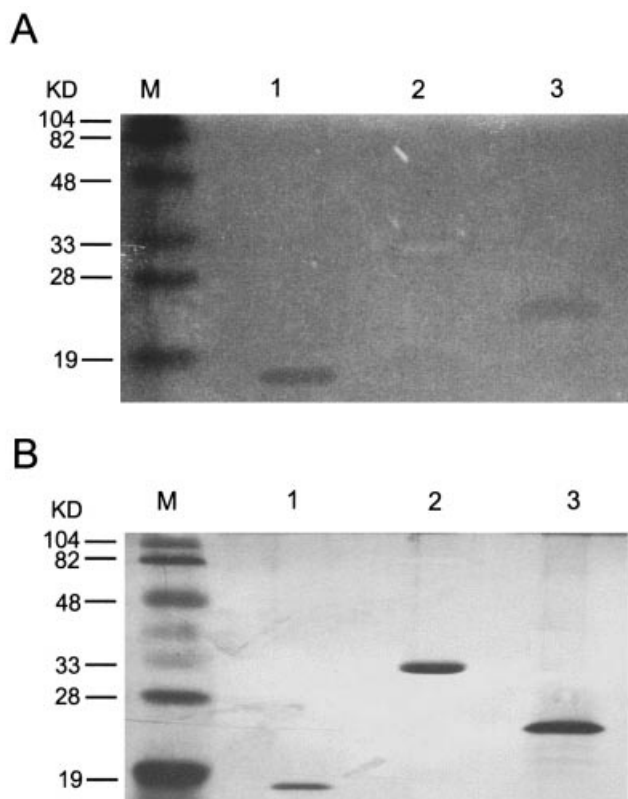


Figure 7 Intrinsic RNase activity by SDS/PAGE

(A) The RNase activity gel stained *in situ*. (B) The silver-stained gel. Lane M, molecular markers [their molecular masses are indicated (in kDa) at the left]; lane 1, micrococcal nuclease; lane 2, bovine pancreatic DNase I; lane 3, shrimp nuclease.

evidence that the conserved motif was the enzyme's active site [21–24] and possible functions of the conserved residues were suggested. These functionally important residues were all found in shrimp nuclease. We therefore proposed a model for the active site of shrimp nuclease on the basis of the crystal structure of *Serratia* nuclease (Figure 6, lower panel).

Intrinsic RNase activity

In view of sequence similarity with other nucleases as illustrated in Figure 6 (upper panel), one cannot eliminate the possibility that shrimp nuclease has the ability to degrade RNA. When the purified shrimp nuclease was assayed for RNase in solution, it showed a specific activity of 18 ± 1 units/mg, which was 1/20 that of micrococcal nuclease (360 ± 20 units/mg). The possible contamination of a polypeptide containing RNase activity in the purified shrimp nuclease preparation was excluded by RNase activity staining *in situ* after SDS/PAGE (Figure 7). For the same amount of protein applied (silver-stained gel, Figure 7B), shrimp nuclease did indeed possess RNase activity, whereas bovine DNase I did not (Figure 7A). Although a previous paper [12] reported that shrimp nuclease had no detectable RNase activity, the discrepancy was probably due to a difference in assay conditions. The previous RNase assays were performed at acidic pH and without metal ions, whereas those used in the present study were at alkaline pH and with $MgCl_2$ and $CaCl_2$.

Identification of the active-site His residue

Shrimp nuclease in a solution containing 0.1 M $MnCl_2$ and 0.1 M iodoacetate at 47 °C was rapidly inactivated. The reaction kinetics was similar to that for bovine DNase I [25]. The inactivation of bovine DNase I by iodoacetate was due to the alkylation of a single His residue [25] at position 134 [6]. His¹³⁴ was considered to be at the active site; a catalytic mechanism involving His¹³⁴ was proposed on the basis of the X-ray structure [26]. For the identification of this His residue, shrimp nuclease was first modified with iodo[2-¹⁴C]acetate. A major radioactive peptide was then isolated in the tryptic digest of the labelled enzyme and His²¹¹ was the radiolabelled residue. His²¹¹ in shrimp nuclease, corresponding to His⁸⁹ in the *Serratia* nuclease, was the highly conserved histidine residue (Figure 6, upper panel).

DISCUSSION

Although bovine DNase I and shrimp nuclease have similar enzymic specific activities [12], their protein sequences are different. In the present study, sequence alignment revealed that the DNase I activity-containing proteins can be grouped into two categories. The first group is bovine DNase-I-like proteins, which is highly conserved and is found only in vertebrates [27]. The second group consists of some nucleases distributed widely in organisms ranging from cyanobacterium to human [3–5,28]). The amino acid sequences for the second group are all different except in a 50-residue segment that is known to be the nuclease active site. Shrimp nuclease belongs to the second group. In evolution, shrimp has not yet quite evolved to differentiate its hepatopancreas. For digestive purposes it employs an ancient molecule, a sugar-non-specific nuclease, rather than the DNA-specific DNase I that is used in the more evolved vertebrates, which have differentiated liver from pancreas.

To the best of our knowledge, shrimp nuclease is the first nuclease to have been cloned and characterized among the Arthropoda. A sequence similarity search revealed no similar genes. Besides several distinctive properties in structure [12,13], the present study demonstrated two unique features: a blocked N-terminus and an odd number of Cys residues forming an unusual mixed disulphide. These two features have never previously been reported among all known nucleases or DNases I. From the results of this study, we established not only the exact number and locations of the Cys residues but also the pairing of disulphide bonds. Shrimp nuclease contained 11 Cys residues. They formed five intramolecular disulphide bonds; the eleventh Cys at position 101 was linked to a thiol compound with an estimated molecular mass of between 500 and 700 Da. The enzymes in which Cys¹⁰¹ has been reduced or alkylated or disulphide-linked are all active, indicating that Cys¹⁰¹ is not involved in catalysis. It has been reported [29–31] that a thiol group of a protease was cross-linked to thiomethanol. The enzyme activity could be generated only by reduction to remove the thiol compound. Recently an anti-tumour RNase from *Rana pipiens* oocytes, called onconase, was reported to have a pyroglutamate residue at the N-terminus [32]. A recombinant mutant onconase (N1M), with a Met residue replacing pyroglutamate, had decreased enzyme activity and cytotoxicity [33]. The biological significance of a blocked N-terminus and an unusual mixed disulphide bond in shrimp nuclease has yet to be elucidated.

The conserved active site found in the family of sugar non-specific nucleases (Figure 6) from organisms with distant evolutionary origins might imply that their catalytic mechanisms are similar. In particular, His²¹¹ and Asn²⁴¹ in shrimp nuclease are also highly conserved in two other different families of nucleases,

the DNA-entry nuclease of the *Streptococcus pneumoniae* family and the Cys-His box containing the nuclear homing endonuclease family. The crystal structure of I-PpoI homing endonuclease complexed with a homing-site DNA oligonucleotide revealed that the conserved Asn residue ligates the essential bivalent cation, and the conserved His interacts with the negative charge on the scissile phosphate oxygen; it was suggested that the His serves as a Lewis acid to stabilize the transition state and that a metal-bound water molecule is deprotonated by a general base and attacks the phosphate group [34]. However, on the basis of mutational and biochemical studies, Friedhoff and co-workers [35] preferred the notion that the conserved His residue functions as a general base in the activation of the water molecule and the bivalent cation as a Lewis acid in phosphodiester bond cleavage.

In view of available structural, mutational and biochemical data of *Serratia* nuclease and I-PpoI, we predict that the active site structure of shrimp nuclease is similar to that of *Serratia* nuclease (Figure 6, bottom panel). In this structure, His²¹¹ and Asn²⁴¹ are very important for catalytic activity, as mentioned above. Lys²¹⁰ and Arg²⁵³ might be involved in substrate binding. The *Serratia* nuclease R87K mutant maintained wild-type activity, showing the importance of a positive charge at this position (Lys²¹⁰ in shrimp nuclease). Other conserved residues, including small residues, bulky hydrophobic residues, Gly²¹⁰, Asn²³², Pro²³⁵, Gln²³⁶, and Trp²⁵⁴, may be involved in the structure stabilization of the active site. In particular, Asn²³² is involved in a hydrogen-bond network supporting the proper orientation of the imidazole ring of His²¹¹ to abstract a proton from the water molecule, as well as Gln²³⁶ to Asn²⁴¹, and Trp²⁴⁵ to Glu²⁴⁹. Moreover, because this nuclease's active site is distinguishable at the sequence level, it can be used as a fingerprint for putative new members belonging to this subfamily.

We have expressed the cloned shrimp nuclease gene in *E. coli* but the expressed protein failed to show nuclease activity. This was probably because shrimp nuclease has more than five disulphide bonds to cross-link. The previous study [13] showed that once the disulphide bonds were reduced, the polypeptide chain was unable to fold back properly to regain its enzymic activity. The difficulties encountered prevented us from a further study of the important role of His²¹¹ by site-directed mutagenesis. Nevertheless, the site-specific modification of His²¹¹ has clearly demonstrated its important catalytic role.

This work was supported in part by a grant from the National Science Council of the Republic of China (NSC87-2316-B-002-010).

REFERENCES

- Laskowski, Sr. M. (1971) in *The Enzymes*, 3rd edn, vol. 3 (Boyer, P. D., ed.), pp. 289–311, Academic Press, New York
- Moore, S. (1981) in *The Enzymes*, 3rd edn, vol. 14 (Boyer, P. D., ed.), pp. 281–296, Academic Press, New York
- Muro-Pastor, A. M., Flores, E., Herrero, A. and Wolk, C. P. (1992) *Mol. Microbiol.* **6**, 3021–3030
- Nestle, M. and Roberts, W. K. (1969) *J. Biol. Chem.* **244**, 5219–5225
- Fraser, M. J. and Low, R. L. (1993) in *Nucleases*, 2nd edn (Linn, S. M., Lloyd, R. S. and Roberts, R. J., eds), pp. 171–207, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Liao, T. H., Salnikow, J., Moore, S. and Stein, W. H. (1973) *J. Biol. Chem.* **248**, 1489–1495
- Paudel, H. K. and Liao, T. H. (1986) *J. Biol. Chem.* **261**, 16006–16011
- Paudel, H. K. and Liao, T. H. (1986) *J. Biol. Chem.* **261**, 16012–16017
- Shak, S., Capon, D. J., Hellmiss, R., Marsters, S. A. and Baker, C. L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9188–9192
- Polzar, B. and Mannherz, H. G. (1990) *Nucleic Acids Res.* **18**, 7151
- Hsiao, Y. M., Ho, H. C., Wang, W. Y., Tam, M. F. and Liao, T. H. (1997) *Eur. J. Biochem.* **249**, 786–791
- Chou, M. Y. and Liao, T. H. (1990) *Biochim. Biophys. Acta* **1036**, 95–100
- Lin, J. L., Wang, W. Y. and Liao, T. H. (1994) *Biochim. Biophys. Acta* **1209**, 209–214
- Kunitz, M. (1950) *J. Gen. Physiol.* **33**, 349–362
- Liao, T. H. (1974) *J. Biol. Chem.* **249**, 2354–2356
- Chang, Y. M., Lin, S. and Liao, T. H. (1994) *Biotechnol. Appl. Biochem.* **19**, 129–140
- Crestfield, A. M., Moore, S. and Stein, W. H. (1963) *J. Biol. Chem.* **238**, 622–627
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Rosenthal, A. L. and Lacks, S. A. (1977) *Anal. Biochem.* **80**, 76–90
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Friedhoff, P., Kolmes, B., Gimadutdinov, O., Wende, W., Krause, K. L. and Pingoud, A. (1996) *Nucleic Acids Res.* **24**, 2632–2639
- Miller, M. D., Tanner, J., Alpaugh, M., Benedik, M. J. and Krause, K. L. (1994) *Nat. Struct. Biol.* **1**, 461–468
- Friedhoff, P., Gimadutdinov, O. and Pingoud, A. (1994) *Nucleic Acids Res.* **22**, 3280–3287
- Miller, M. D. and Krause, K. L. (1996) *Protein Sci.* **5**, 24–33
- Price, P. A., Moore, S. and Stein, W. H. (1969) *J. Biol. Chem.* **244**, 924–928
- Weston, S. A., Lahm, A. and Suck, D. (1992) *J. Mol. Biol.* **226**, 1237–1256
- Liao, T. H. (1997) *J. Formos. Med. Assoc.* **96**, 481–487
- Ho, H. C. and Liao, T. H. (1999) *Biochem. J.* **339**, 261–267
- Ferdinand, W., Stein, W. H. and Moore, S. (1965) *J. Biol. Chem.* **240**, 1150–1155
- Bustin, M., Lin, M. C., Stein, W. H. and Moore, S. (1970) *J. Biol. Chem.* **245**, 846–849
- Lo, S. S., Fraser, B. A. and Liu, T. Y. (1984) *J. Biol. Chem.* **259**, 11041–11045
- Ardelt, W., Mikulski, S. M. and Shogen, K. (1991) *J. Biol. Chem.* **266**, 245–251
- Boix, E., Wu, Y. N., Vasandani, V. M., Saxena, S. K., Ardelt, W., Ladner, J. and Youle, R. J. (1996) *J. Mol. Biol.* **257**, 992–1007
- Flick, K. E., Jurica, M. S., Monnat, Jr, R. J. and Stoddard, B. L. (1998) *Nature (London)* **394**, 96–101
- Friedhoff, P., Franke, I., Krause, K. L. and Pingoud, A. (1999) *FEBS Lett.* **443**, 209–214

Received 18 October 1999/7 December 1999; accepted 7 January 2000