Biosci. Biotechnol. Biochem., **73** (1), 21–28, 2009

Purification and Characterization of a Major Collagenase from *Streptomyces parvulus**

Yasuko SAKURAI,[1] Hideshi INOUE,[1,2] Wataru NISHII,[2] Takayuki TAKAHASHI,[1] Yuichi IINO,[1] Masayuki YAMAMOTO,[1] and Kenji TAKAHASHI[1,2,†]

[1]*Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[2]*School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan*

Received May 27, 2008; Accepted August 21, 2008; Online Publication, January 7, 2009
[doi:10.1271/bbb.80357]

A major collagenase was purified about 96-fold from a crude enzyme sample of *Streptomyces parvulus* by chromatography on Q-Sepharose, Sephacryl S-200, and butyl-Toyopearl. The purified enzyme showed a relative molecular mass of approximately 52,000 on SDS–PAGE and a pH optimum at about 9.0, and was strongly inhibited by metal-chelating agents. It also cleaved 4-phenylazobenzyloxycarbonyl-Pro-Leu-Gly-Pro-D-Arg specifically at the Leu-Gly bond, with a K_m value of 0.60 mM at pH 9.0 at 37 °C. Based on the amino acid sequences of the N-terminal region and internal tryptic peptides, the corresponding gene was cloned. The DNA sequence of the cloned gene indicated that the enzyme is produced as an 864-residue precursor protein with a 408-residue prepro sequence followed by a 456-residue mature enzyme moiety. The enzyme is most homologous with the collagenase from *S. coelicolor*, the identity being 73%, and it is thought to be a member of the *Vibrio* collagenase subfamily.

Key words: amino acid sequence; characterization; collagenase; purification; *Streptomyces parvulus*

Bacterial collagenases (*IUBMB*: EC 3.4.24.3) are metallo-endopeptidases capable of digesting native, triple-helical collagens, and are classified into two major subfamilies, M9A including *Vibrio* collagenases[1)] and M9B including *Clostridium* collagenases[2)] in the MEROPS classification. They are useful in tissue dissociation experiments, including the isolation of hepatocytes, adipocytes, and other cells for research purposes, as well as in medical procedures such as the preparation of vascular endothelial cells for seeding vascular prostheses and the isolation of pancreatic islets for transplantation.[2)] To date bacterial collagenases have been purified from various species, such as *C. histolyticum*,[3–6)] *C. perfringens*,[7)] *Pseudomonas marinoglutinosa*,[8)] *V. alginolyticus*,[9–11)] and *Streptomyces* sp.,[12–16)] and partially characterized. The *V. parahaemolticus* collagenase was obtained by expression of the gene in *Escherichia coli* and partially characterized.[17,18)] Moreover many more collagenase genes have been cloned and sequenced and the amino acid sequences of the corresponding enzyme precursors predicted, but collagenases that have been both enzymatically and structurally well characterized are rather scarce.

Crude collagenase preparation from *Streptomyces parvulus* subsp. *citrinus* (collagenase N-2, Nitta Zeratin, Osaka, Japan) is commercially available and has been used in various biochemical and biomedical studies, but it has not yet been purified completely and fully characterized. Hence it is important to purify the enzyme to homogeneity and to clarify fully its enzymatic and structural characteristics. In the present study we purified a major collagenase from *S. parvulus* and investigated various characteristics of it including the amino acid sequences of the precursor form (864 residues) and the mature enzyme (456 residues) by gene cloning and nucleotide/protein sequencing.

Materials and Methods

Materials. A bacterial strain of *S. parvulus* subsp. *citrinus* (IFO 14435) was obtained from the Fermentation Institute (Tsukuba, Japan). A crude enzyme sample (collagenase N-2) prepared from the same bacterial strain was from Nitta Zeratin. It had been prepared from a culture filtrate of the above bacteria by ammonium sulfate fractionation, followed by chromatographies on DEAE-Toyopearl and Sephadex G-100.[19)] Azocasein, benzamidine, bovine pancreatic trypsin inhibitor, diisopropylfluorophosphate, and N^{α}-tosyl-phenylalanine chloromethyl ketone were from Sigma (St. Louis, MO), and bestatin and leupeptin from Peptide Institute (Osaka, Japan). 4-Phenylazobenzyloxycarbonyl(PZ)-Pro-Leu-Gly-Pro-D-Arg-OH·$2H_2O$ (PZ-peptide) was from Nova Biochem (Lauferfingen, Switzerland). Q-Sepharose and Sephacryl S-200 were from Pharmacia Biotech (Piscataway, NJ), and butyl-Toyopearl and a TSKgel ODS-120T column were from Tosoh Co. (Tokyo). Reagents for protein and peptide sequencing were from Applied Biosystems (Tokyo). Other reagents were of analytical grade, largely obtained from Wako Pure Chemical (Tokyo).

Determination of enzyme activity. Azocoll was routinely used as a substrate unless otherwise specified, essentially as described elsewhere.[20)] Azocoll (4 mg) suspended in 890 µl of 50 mM Tris–HCl buffer, pH 7.5, was mixed with 10 µl of an enzyme solution in the same buffer, and the mixture was kept at 37 °C for 10–30 min under shaking. The reaction was stopped by the addition of 100 µl of 30% trichloroacetic acid, and the mixture was centrifuged at 40,000 × *g* for 10 min. The absorbance of the supernatant was measured at 540 nm. The blank sample was treated in the same manner without the enzyme. One unit

* The nucleotide sequence data for the collagenase are available from the DDBJ data bank under accession no. AB429498.
† To whom correspondence should be addressed. Fax: +81-49-297-8168; E-mail: kenjitak@ls.toyaku.ac.jp
Abbreviations: PZ, 4-phenylazobenzyloxycarbonyl; PCR, polymerase chain reaction

of enzyme activity was defined as the amount causing an increase of 1.0 in the absorbance at 540 nm per min under the standard assay conditions. When the volume of enzyme solution exceeded 10 µl, the volume of the buffer used to suspend azocall was reduced accordingly to adjust the final volume of the assay mixture to 900 µl.

Purification procedures. All purification steps were performed at 4 °C. A crude enzyme sample (collagenase N-2) (50 mg) was dissolved in 5 ml of 5 mM Tris–HCl buffer, pH 8.0, and 4 mM $CaCl_2$, and dialyzed against the same buffer (4,000 ml × 2). The dialyzed sample was loaded on a column of Q-Sepharose (1.6 × 10 cm) equilibrated with the same buffer, and then eluted with a linear gradient of 5 mM (500 ml) to 1 M (500 ml) of Tris–HCl buffer, pH 8.0, and 4 mM $CaCl_2$. The active fractions were pooled and concentrated using Centricell (Polyscience, Warrington, PA), and applied to a column (3.0 × 100 cm) of Sephacryl S-200 equilibrated and eluted with 50 mM Tris–HCl buffer, pH 8.0, 4 mM $CaCl_2$, and 0.2 M NaCl. The pooled active fraction was concentrated using Centricell and applied to a column (2.1 × 19.5 cm) of butyl-Toyopearl equilibrated with 50 mM Tris–HCl buffer, pH 8.0, 4 mM $CaCl_2$, and 20% ammonium sulfate. Elution was performed by a linear gradient from 50 mM Tris–HCl buffer, pH 8.0, 4 mM $CaCl_2$, and 10% ammonium sulfate (1,000 ml) to 5 mM Tris–HCl buffer, pH 8.0, and 4 mM $CaCl_2$ (1,000 ml). The active fractions were pooled and stored frozen at −20 °C. When necessary, part of the pooled fraction was concentrated using Centricell, desalted through a PD-10 column, and stored frozen at −20 °C. The protein concentration was determined from the absorbance at 280 nm assuming A_{280} (0.1% solution, 1-cm light path) = 1.0 unless otherwise specified.

SDS–PAGE. To check the homogeneity of the purified enzyme and to estimate its molecular mass, SDS–PAGE was performed under non-reducing conditions in 10% polyacrylamide gel by the method of Laemmli,[21] followed by Coomasie Brilliant Blue staining.

Determination of pH/activity profile. The activity toward azocoll was determined at 37 °C at pH 4.0–6.0 in 50 mM sodium acetate buffers, at pH 7.0–9.0 in 50 mM Tris–HCl buffers, at pH 9.5–11.0 in 0.1 M Na_2CO_3–$NaHCO_3$ buffers, and at pH 11.5–12.0 in 0.1 M Na_2CO_3–NaOH buffers. The experiments were performed 4 times.

Inhibition studies. The enzyme (0.008 µg in 10 µl of 7.5 mM Tris–HCl buffer, pH 8.0, and 4 mM $CaCl_2$) was mixed with the various reagents dissolved in 2–20 µl of 50 mM Tris–HCl buffer, pH 9.0 (containing a small volume of ethanol where necessary), and kept at 37 °C for 5 min, and then the remaining activity toward azocoll was determined under standard assay conditions.

Stability studies. The enzyme (0.12 µg in 150 µl of 7.5 mM Tris–HCl buffer, pH 8.0, and 4 mM $CaCl_2$) was kept at different temperatures for 30 min or 1 h, and then the remaining activity toward azocoll was determined under standard assay conditions. The experiments were performed 4 times.

Kinetic studies. The reaction mixtures contained various volumes of PZ-peptide[21] solution (0.34 nmol/µl in 50 mM Tris–HCl buffer, pH 9.0), enzyme solution (5 µl, 0.004 µg), 50 mM Tris–HCl buffer, pH 9.0 (80 µl), and distilled water in a total volume of 190 µl. The mixtures were incubated at 37 °C for 5 min unless otherwise specified. The reaction was stopped by heating at 100 °C for 2 min, and the digest was analyzed by HPLC using a Hitachi (Tokyo) 655A-11 system on a column (0.46 × 25 cm) of TSKgel ODS-120T (Tosoh, Tokyo). The peptides were eluted with linear gradients of acetonitrile (0–40% over 0–20 min and 40–100% over 20–25 min) in 0.1% trifluoroacetic acid at a flow rate of 0.8 ml/min. Under these conditions, PZ-peptide and its hydrolysis products, PZ-Pro-Leu and Gly-Pro-D-Arg, were eluted at 22.1, 22.8, and 1.9 min respectively (see Fig. 5). The extent of cleavage was estimated by amino acid analysis after acid hydrolysis (6 N HCl, 110 °C, 24 h) of the Gly-Pro-D-Arg fraction using an Applied Biosystems automated derivatizer-analyzer (420A/130A) (Applied Biosystems, Foster City, CA). V_{max} and K_m values were estimated from a Lineweaver-Burk plot using the reaction rates at 10 different substrate concentrations.

Determination of N-terminal and partial internal amino acid sequences. The N-terminal amino acid sequence of the purified protein was determined using an Applied Biosystems pulse-liquid protein sequencer model 477A. To determine the partial internal sequences of the enzyme, the protein was heat-denatured by incubation at 100 °C for 1 h. The denatured protein (about 1 nmol) was then digested at 37 °C with trypsin (Worthington, Lakewood, NJ) for 24 h in 83 µl of 70 mM Tris–HCl buffer, pH 8.0, and 10 mM $CaCl_2$. Trypsin (1.2 µg) was added at the start, and after 3 h and 20 h. The reaction was stopped by the addition of 10 µl of 1 N HCl. The resulting peptides were separated by HPLC using a Hitachi (Tokyo) 655A-11 system on a column (0.46 × 25 cm) of TSKgel ODS-120T (Tosoh). The peptides were eluted with linear gradients of acetonitrile (0–50% over 30 min, then to 100% over the next 5 min) in 0.1% trifluoroacetic acid at a flow rate of 0.8 ml/min. The effluent was monitored by measuring the absorbance at 215 and 280 nm, and the peptide peak fractions were collected and lyophilized. The peptide fractions were numbered with a suffix, "T," in the order of elution. An aliquot of each peptide fraction dissolved in water was submitted to automated amino acid sequencing using an Applied Biosystems pulse-liquid protein sequencer model 477A.

Cloning and nucleotide sequencing of the collagenase gene. Cloning of the collagenase gene was performed essentially as described previously.[22] Based on the amino acid sequences Leu-Arg-Ile-Arg-Ala-Gln-Glu-Met-Thr, Val-Leu-Val-Ile-Asn-His-Thr-Cys, and Phe-Gly-Asp-Gly-Thr-Thr-Ser-Ala-Ala-Ala-Asn-Pro-Ala, the degenerated primers were designed as follows: Fwd-1, CTI (C/A)GI ATI (C/A)GI GCI CA(G/A) GA(G/A) ATG AC; Fwd-2, GTI CTI GT(G/C) ATC AA(C/T) CA(C/T) AC(G/C/A/T) T; and Rev, GC IGG (G/A)TT IGC IGC IGC I(G/C)(A/T) IGT IGT ICC (G/A)TC ICC (G/A)AA. The first polymerase chain reaction (PCR) was performed using a DNA extract from *S. parvulus* as the template and a set of primers, Fwd-2 and Rev. The PCR conditions were as follows: an initial denaturation of 94 °C for 5 min; 25 cycles of 94 °C for 30 s, 55 °C for 2 min, and 72 °C for 2 min; and a final extension of 72 °C for 7 min. Using the 0.9-kb first PCR product as the template, the second PCR was performed with a set of primers, Fwd-1 and Rev. The PCR product was extracted from agarose gel after electrophoresis, labeled using a DIG Labeling Kit (Boehringer Mannheim, Mannheim, Germany) according to the manufacturer's instructions, and used as the specific probe for Southern and colony hybridizations. Since a 5-kb *Bam*HI fragment was detected as a major band by Southern hybridization (Fig. 10), the library for colony hybridization was prepared as follows: the *Bam*HI fragments around 5-kb were extracted from agarose gel after electrophoresis, ligated into the *Bam*HI site of plasmid BlueScript SK+, and introduced into *E. coli* DH5 α. Detection in Southern and colony hybridizations was performed using a DIG Luminescent Detection Kit for Nucleic Acid (Boehringer Mannheim). Plasmid DNA was prepared from a positive clone by the boiling lysis method. DNA sequencing was performed using a DNA sequencer model 4000L (LI-COR).

Sequence comparison. Sequence alignment and the construction of a phylogenetic tree were performed using the program Clustal W.[23]

Prediction of the signal sequence cleavage site, secondary structures, and the hydrophobicity profile. The signal sequence cleavage site was predicted following Heijne.[24] The secondary structures were predicted by the GOR4 program[25,26] and following Chou and Fasman.[27] A hydrophobicity plot was obtained by the method of Kyte and Doolittle.[28,29]

Results and Discussion

A major collagenase was purified from the crude enzyme sample (collagenase N-2) from *S. parvulus* by successive steps of chromatography on Q-Sepharose, Sephacryl S200, and butyl-Toyopearl. The results are shown in Fig. 1 and Table 1. The enzyme activity was largely eluted from the Q-Sepharose column with the equilibration buffer at the flow-through position, and most of the non-enzyme protein was bound to the

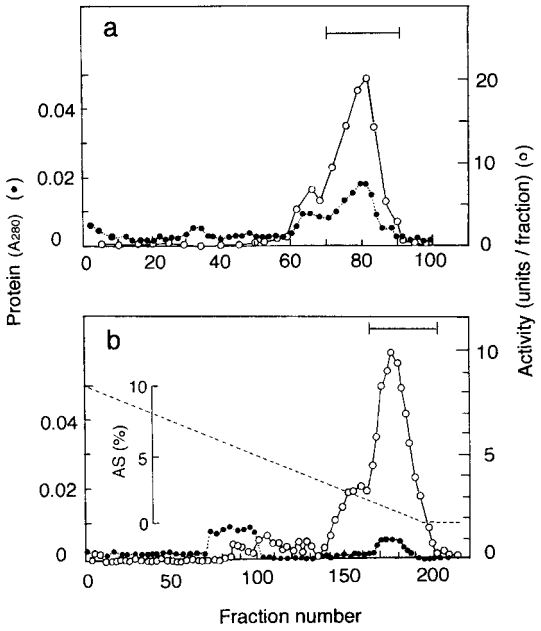

Fig. 1. Purification of a Collagenase from *S. parvulus*.

a, Chromatography on a Sepharose S-200 column. The active fraction, which had passed through the Q-Sepharose column, was chromatographed on the column (3.0 × 100 cm) equilibrated and eluted with 50 mM Tris–HCl buffer, pH 8.0, 4 mM $CaCl_2$, and 0.2 M NaCl at a flow rate of 28 ml/h. Fractions of 8.3 ml were collected, and the active fractions under the bar were pooled. b, Chromatography on a butyl-Toyopearl column. The active fraction from the Sepharose S-200 column was applied to a butyl-Toyopearl column (2.1 × 19.5 cm) equilibrated with 50 mM Tris–HCl buffer, pH 8.0, 4 mM $CaCl_2$, and 20% ammonium sulfate. Elution was performed by a linear gradient from 50 mM Tris–HCl buffer, pH 8.0, 4 mM $CaCl_2$, and 10% ammonium sulfate (1,000 ml) to 5 mM Tris–HCl buffer, pH 8.0, and 4 mM $CaCl_2$ (1,000 ml) at a flow rate of 50 ml/h. Fractions of 10 ml were collected, and the active fractions under the bar were pooled. The broken line in (b) shows the concentration of ammonium sulfate (AS) in the eluting buffer.

Table 1. Purification of a Collagenase from *S. parvulus*

Step	Protein[a] (mg)	Total activity (units)	Specific activity (units/mg)	Purification (-fold)	Yield (%)
Crude enzyme	50	488	9.8	1	100
Q-Sepharose	3.6	349	97	9.9	72
Sephacryl S-200	2.1	221	105	10.7	45
Butyl-Toyopearl	0.29[b]	270	938	95.7	55

A crude enzyme sample (collagenase N-2) (50 mg) was used as the starting material.

[a]The protein was estimated assuming A_{280} (0.1%, 1 cm) = 1.0, except for the crude enzyme.

[b]This value was corrected to 0.12 mg using a calculated A_{280} (0.1%, 1 cm) = 2.363.

column and eluted by the gradient elution (data not shown). The enzyme in the flow-through fraction from Q-Sepharose was further purified by chromatography on Sepharose S-200 and butyl-Toyopearl. The total activity was somewhat increased after butyl-Toyopearl chromatography (Table 1), presumably due to the removal of an inhibitory contaminant. Through these steps, the enzyme was purified from the crude sample about 96-fold at a yield of 55%. The purified enzyme gave a single peak on SDS–PAGE, and its relative molecular mass was estimated to be approximately 52,000 (Fig. 2). Upon chromatography on Sepharose S-200 and butyl-Toyopearl, a minor activity peak was observed in front

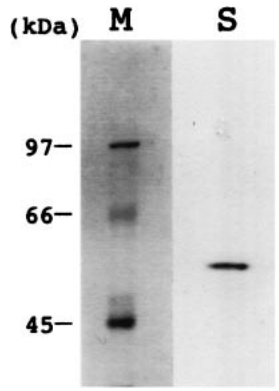

Fig. 2. SDS–PAGE of Purified *S. parvulus* Collagenase.

The *S. parvulus* collagenase finally purified by chromatography on butyl-Toyopearl was submitted to SDS–PAGE. Electrophoresis was performed under non-reducing conditions in a 10% gel, followed by Coomasie Brilliant Blue staining. M, molecular mass markers; S, sample.

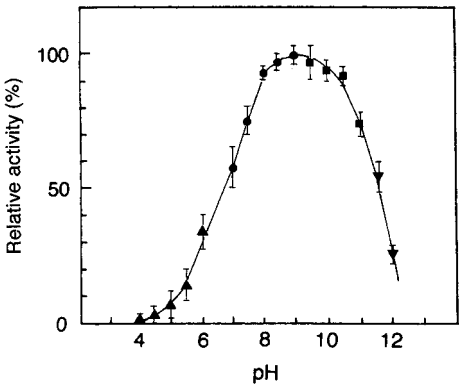

Fig. 3. pH-Dependence of Activity of *S. parvulus* Collagenase toward Azocoll.

Activity toward azocoll was measured at various pH values at 37 °C. ▲, sodium acetate buffers; ●, Tris–HCl buffers; ■, Na_2CO_3–$NaHCO_3$ buffers; ▼, Na_2CO_3–NaOH buffers. Each point represents an average of the values obtained in four independent experiments. The vertical bars represent standard deviations.

of the major peak. These results might indicate the presence of a minor collagenase component (see Fig. 10). The possibility cannot be excluded that a small amount of the minor enzyme contaminated the purified enzyme fraction. However, since no additional band was observed on SDS–PAGE, the amount might have been extremely small or its molecular mass might have been indistinguishable from that of the major enzyme.

The purified major enzyme was used in the subsequent characterization studies. To date, collagenases from *Streptomyces* species have been purified from various sources, but have been characterized only partially.[12–16] Figure 3 shows the pH-dependence of the activity toward azocoll; the enzyme is highly active in a pH range of 8–10, with a maximum around pH 9. This pH optimum is similar to that of 8–9 reported for *Streptomyces* sp. C-51,[14] but higher than the pH optima in a range of 6.0–8.0, reported for collagenases from *Streptomyces* sp. 1382[15] and 3B[16] and other collagenases.[4,9,10,18] As shown in Fig. 4, the enzyme was fairly stable up to 50 °C. Above 50 °C, it became remarkably unstable, and it was almost completely inactivated by incubation at 60 °C for 30 min (Fig. 4). The results obtained by incubation for 1 h were nearly the same, except that 45% of the original activity was lost at 55 °C.

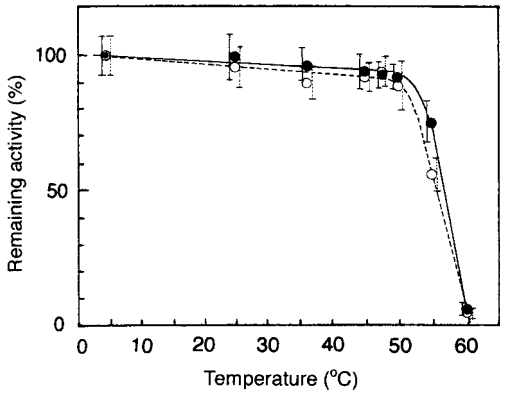

Fig. 4. Effects of Temperature on the Stability of *S. parvulus* Collagenase.

The enzyme was kept at various temperatures for 30 min (—●—) or 1 h (– –○– –) at pH 8.0, and then the remaining activity toward azocoll was determined under standard assay conditions. Each point represents an average of the values obtained in four independent experiments. The vertical bars represent standard deviations (solid line, 30 min; dotted line, 1 h).

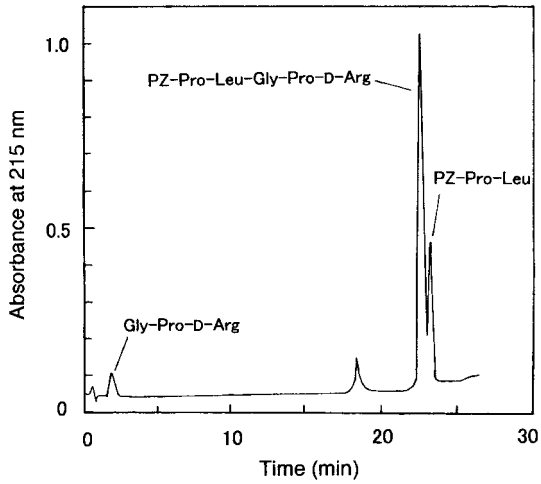

Fig. 5. HPLC Pattern of a Digest of PZ-Peptide by *S. parvulus* Collagenase.

PZ-peptide (24 nmol) was hydrolyzed by the enzyme (0.004 μg) for 5 min under standard assay conditions, and the digest was submitted to HPLC. The small peak (retention time, 18.2 min) eluted before the PZ-peptide is a non-peptide impurity.

Table 2. Effects of Various Reagents on the Activity of *S. parvulus* Collagenase

The purified enzyme was mixed with each reagent in 50 mM Tris–HCl buffer, pH 9.0, and kept at 37 °C for 5 min, and then the remaining activity toward azocoll was determined under standard assay conditions.

Inhibitor	Concentration[a] (mM)	Inhibition (%)
EDTA	1	89
8-Quinolinol	1	73
o-Phenanthroline	1	63
2,3-Dimercapto-1-propanol	1	53
L-Cysteine	10	26
Bestatin	1	16
Leupeptin	1	10
Iodoacetic acid	1	10
Benzamidine	1	10
Bovine pancreatic trypsin inhibitor	0.1	7
Diisopropylfluorophosphate	10	0
Tosyl-L-phenylalanine chloromethyl ketone	1	0

[a]Final concentration in the assay mixture.

The collagenase from *Streptomyces* sp. 3B has been reported to be less stable, with an optimal temperature of 37 °C.[16] Other collagenases have been reported to be stable up to 40 °C,[10,15] or near 50 °C.[8] Therefore the temperature stability of the present enzyme appears to be higher than those of most other collagenases. The enzyme was strongly inhibited by metal-chelating agents, such as EDTA, 8-quinolinol, *o*-phenanthroline, and 2,3-dimercapto-1-propanol, but was rather insensitive to inhibitors of serine and cysteine endopeptidases and aminopeptidases, as shown in Table 2. This indicates that the enzyme is a typical metalloendopeptidase, like many other bacterial collagenases containing an essential Zn^{2+}. Strong inhibition by EDTA and *o*-phenanthroline has also been reported for collagenases from *Streptomyces* sp. A8 and 3B.[13,16]

As shown in Fig. 5, PZ-peptide, a typical synthetic peptide substrate for collagenases, was cleaved specifically at the Leu-Gly bond. Under similar conditions, approximately 70% hydrolysis occurred in 30 min, and complete hydrolysis in several h at this peptide bond, and no hydrolysis was observed at other peptide bonds (data not shown). This is consistent with the fact that the crude enzyme (collagenase N-2) used in this study has been reported to cleave bovine tendon collagen (Sigma, St. Louis, MO) specifically at X-Gly bonds (X, predominantly Hyp and Ala), but scarcely to hydrolyze bovine serum albumin or casein.[19] These results indicate that the purified enzyme is a member of the bacterial collagenase family. The kinetic parameters, K_m and V_{max} values, were determined to be 0.60 ± 0.03 mM and 37 ± 2 μmol/l · min respectively. The *V. parahaemolyticus* and *C. histolyticum* collagenases have been reported to have K_m values of 1.06 mM and 14.82 mM respectively toward 2-furanacryloyl-Leu-Gly-Pro-Ala at pH 8.0 at 25 °C.[17] Thus the K_m value of the present enzyme toward PZ-peptide was apparently similar to that of the *V. parahaemolyticus* collagenase, although the substrates used were different.

N-terminal sequencing of the purified enzyme yielded a single sequence of 33 residues. In addition, internal amino acid sequences of 135 residues in total were also determined at the protein level by analyzing a tryptic digest of the purified enzyme. These peptide sequences were found to agree completely with the protein sequence deduced from DNA sequencing (Fig. 6), indicating the homogeneity of the purified enzyme. The complete amino acid sequence of the prepro form of the collagenase was deduced by cloning and nucleotide sequencing of the gene clone obtained from *S. parvulus*, as shown in Fig. 6. The total number of residues was 864, and the calculated relative molecular mass was 93,560. The N-terminal and some internal amino acid sequences determined at the protein level are also included in Fig. 6. Thus the prepro enzyme was composed of 864 amino acid residues, including a 408-residue prepro peptide and a 456-residue mature enzyme. In the prepro peptide, the N-terminal 33 residues and the remaining 375 residues were predicted to be the signal peptide and the propeptide respectively.

The mature enzyme is produced by cleavage at the Ser-Ser bond. *S. parvulus* therefore should produce an endopeptidase capable of cleaving this peptide bond. Purification and characterization of this processing enzyme would be interesting, especially since no

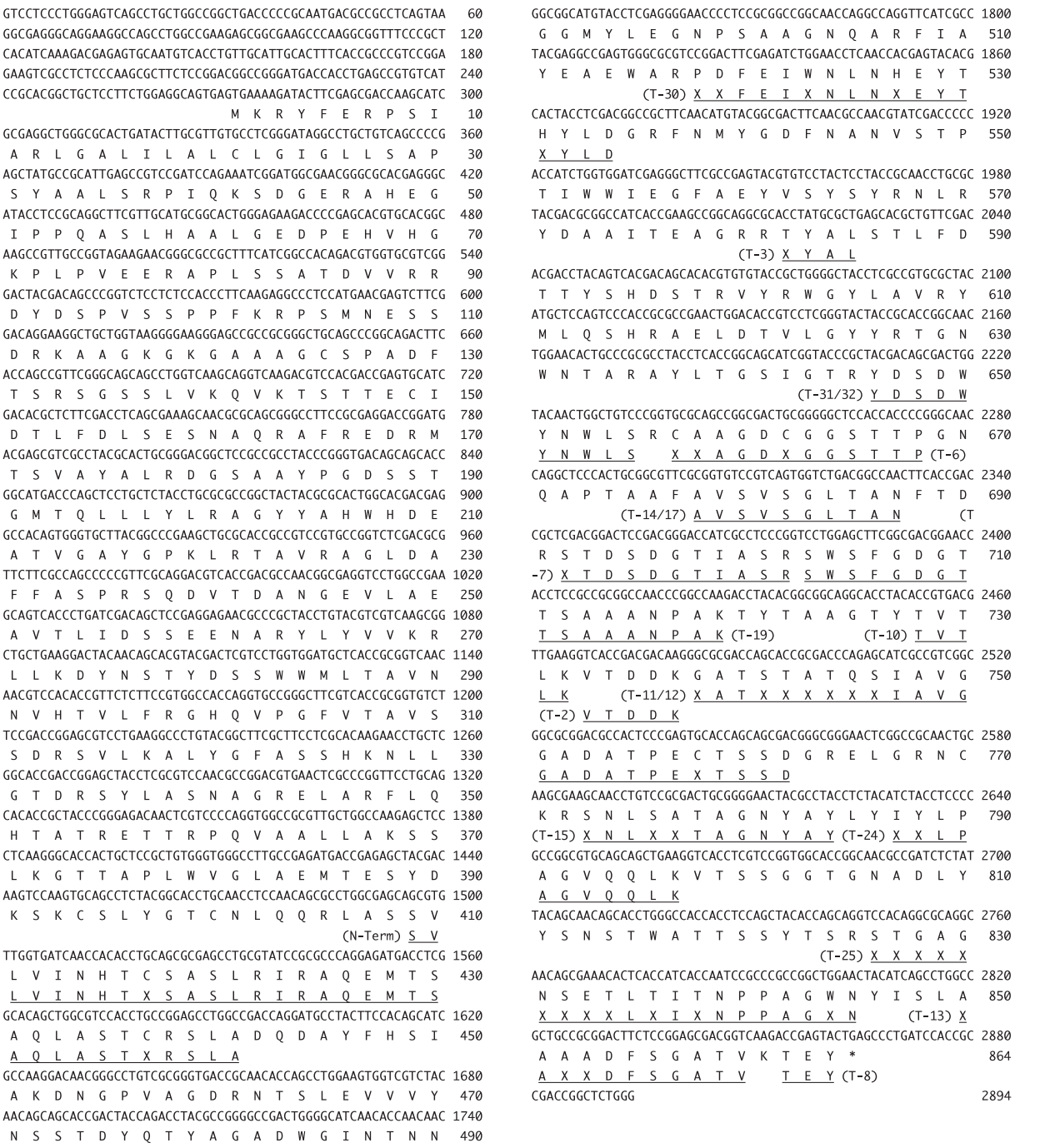

Fig. 6. Nucleotide and the Deduced Amino Acid Sequences of *S. parvulus* Collagenase.

The initiation codon was predicted to be GTG at position 271, as in the case of the *S. coelicolor* ortholog. The deduced N-terminal amino acid sequence and the amino acid sequences of the tryptic peptides determined are shown in one-letter code below the nucleotide sequence, and are underlined. N-Term, the N-terminal sequence of the enzyme; T, tryptic peptide. Numbers stand for peptide peak numbers on HPLC. The peptide number is shown before each peptide, except for T-6, T-8, and T-19 which are shown after the peptide. X, residue not unambiguously identified. An asterisk indicates a stop codon.

endopeptidase with such a specificity is known. The calculated relative molecular mass and pI value of the mature enzyme were 49,368 and 5.21 respectively. This relative molecular mass is roughly consistent with that (approximately 52,000) for the purified enzyme as determined by SDS–PAGE. The relative molecular masses reported for other collagenases from *Streptomyces* sp. are highly variable: 35,000,[12] 30,000–40,000,[15] 75,000,[13] 100,000 (type I) and 90,000–110,000 (type II),[14] and 116,000 (type I) and 97,000 (type II).[16] Thus the molecular mass of the present enzyme was different from any of these collagenases. These results indicate that there are marked variations in structural and enzymatic properties even among the *Streptomyces* collagenases.

The enzyme had a peptidase domain with a metallopeptidase active-site motif, His-Glu-X-X-His, at positions 527–531 (prepro protein numbering is used throughout), including the putative catalytic Glu528 and two metal ligands, His527 and His531, and a third putative metal ligand, Glu556. The enzyme is thought to have a PKD (polycystic kidney disease) domain[30] (residues 678–737) and a PPC (bacterial pre-peptidase C-terminal) domain[31] (residues 781–851) in the C-terminal region as judged by the domain structure of the *S. coelicolor* collagenase (CAA16449).

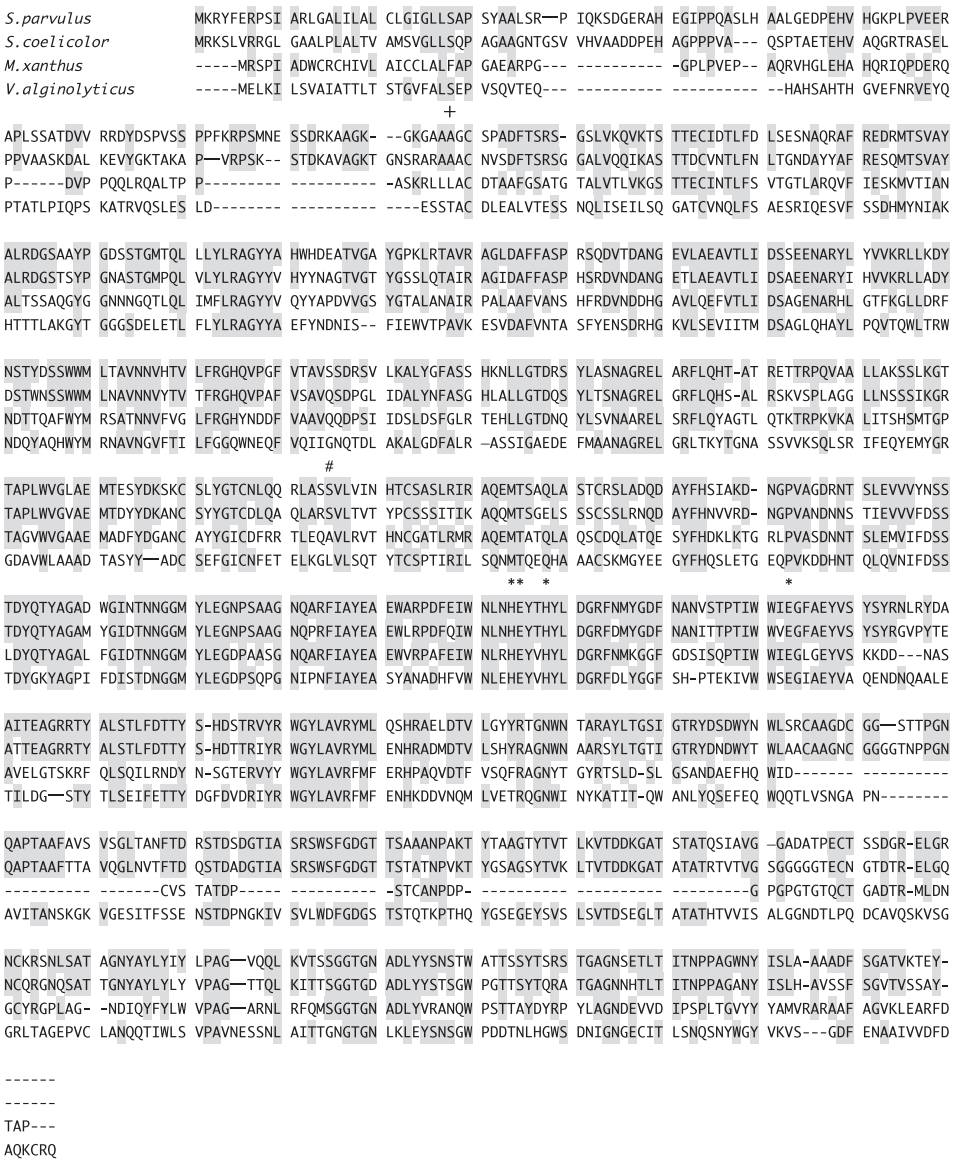

Fig. 7. Amino Acid Sequence Comparison of *S. parvulus* Collagenase with Homologous Collagenases.

The amino acid sequences of the prepro forms of the collagenases from *S. parvulus* (this study), *S. coelicolor* (CAA16449), *M. xanthus* (YP_634336), and *V. alginolyticus* (EAS77672) are aligned. The active site residues are indicated by *, and the N-terminal residues of the *S. parvulus* and *V. alginolyticus* enzymes are indicated by # and + respectively. The residues common to the *S. parvulus* collagenase are shaded. Dashes indicate deletions.

The amino acid sequences of a number of bacterial collagenases have been deduced from the nucleotide sequences of their genes, but only the sequence of the *S. coelicolor* collagenase has been determined among *Streptomyces* sp. Among these, the collagenases from *S. coelicolor*, *Micrococcus xanthus* (YP_634336), and *V. arginolyticus*[11] are among the most homologous to the present enzyme, and their sequences are compared in Fig. 7. Among the bacterial collagenases, the N-terminal sequences of the mature forms have been determined for the present enzyme (this study) and for the collagenases from *V. alginolyticus*,[11] *V. anguillarum*,[32] *C. perfringens*,[7] and *C. histolyticum*.[33] There is a marked difference in the number of residues from the N-terminus to the first His residue of the active-site motif. These are 118, 147, 412, and 416 residues for the present enzyme, *V. anguillarum*, *V. alginolyticus*, and *C. perfringens* enzymes respectively, and 414 and 415 residues for the two enzymes from *C. histolyticum*. Since the 456-residue enzyme is the only active form identified so far for the present enzyme, we assume that this form is the mature enzyme. The cleavage sites were found to be Ser-Thr, His-Ala, and Arg-Ala bonds in the collagenases from *V. alginolyticus*, *V. anguillarum*, and *C. perfringens* respectively, and Ser-Ile, Arg-Ala, and Arg-Val bonds in the collagenases G, A, and H from *C. histolyticum* respectively. Thus, the cleavage sites of Ser-Thr and Ser-Ile in the collagenases of *V. alginolyticus* and *C. histolyticum* (collagenase G) respectively are similar to that of the present enzyme (*viz.*, Ser-Ser). On the other hand, the cleavage sites in the other collagenases, His-Ala, Arg-Ala, Arg-Ala, and Arg-Val, have a basic residue at the P1 position and a small aliphatic residue at the P1′ position. No marked sequence homology was to be seen, however, beyond the P1 and P1′ positions. Hence a few types of processing enzymes of different specificities might be present. The differences in processing site, and hence in the size of the pro part, are thought to be due to differences in the mode of processing of proenzymes by

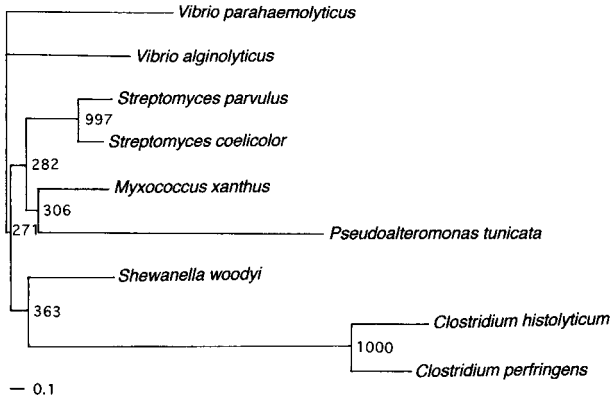

Fig. 8. Phylogenetic Tree Based on the Amino Acid Sequences of the Prepro Forms of Typical Bacterial Collagenases.

The accession numbers of the amino acid sequences of the collagenases used are as follows: *S. parvulus*, AB429498; *S. coelicolor*, CAA16449; *M. xanthus*, YP_634336; *P. tunicata*, EAR30787; *S. woodyi*, EAV38041; *V. alginolyticus*, EAS77672; *V. parahaemolyticus*, Q9AMB9; *C. histolyticum*, BAA86030; and *C. perfringens*, BAB79879. The lengths of branches indicate the evolutionary distances (substitutions/site) between the proteins. The bootstrap value of each node is shown at that node.

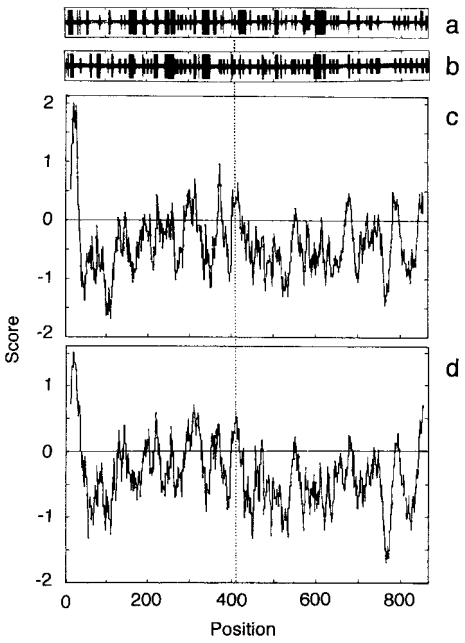

Fig. 9. Secondary Structure and Hydropathy Profile Comparison between the Prepro Forms of *S. parvulus* and *S. coelicolor* Collagenases.

The secondary structures of the *S. parvulus* collagenase (a) and *S. coelicolor* collagenase (b) were predicted by the GOR 4 program. Long vertical bar, α-helix; short vertical bar, β-strand. The hydropathy plots of the *S. parvulus* collagenase (c) and *S. coelicolor* collagenase (d) were obtained by the method of Kyte and Doolittle. The vertical dotted line indicates the N-terminal position (residue 409) of the mature *S. parvulus* collagenase and the corresponding position in the *S. coelicolor* collagenase.

specific processing enzymes in the respective bacteria. To our knowledge, the mechanism of activation of the proenzymes of bacterial collagenases has not been studied so far. It would be interesting to elucidate the activation mechanism of the present enzyme, including characterization of the processing enzyme.

When the sequence of the present enzyme in mature form was compared with the corresponding parts of other enzymes, the sequence identities were calculated

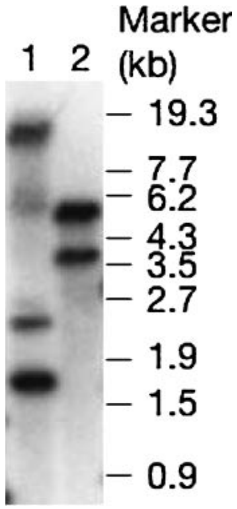

Fig. 10. Southern-Blot Hybridization Analysis of Genomic DNA from *S. parvulus*.

The genomic DNA was digested with restriction enzymes, as follows: lane 1, *Sac*I; lane 2, *Bam*HI.

to be 73%, 49%, and 38% with the collagenases from *S. coelicolor*, *M. xanthus*, and *V. alginolyticus* respectively. On the other hand, the identity with the *C. histolyticum* collagenase was 26%. Thus the present collagenase was most closely related to the collagenase from *S. ceolicolor*. Moreover, the *S. coelicolor* prepro enzyme (865 residues) shows significant similarity to the present enzyme in the size and sequence of the prepro moiety as well, as shown in Fig. 7. The N-terminal sequences of both prepro proteins start at the same position. These results indicate that the present enzyme is a member of the *Vibrio* collagenase subfamily (*MEROPS* M9A) rather than the *Clostridium* collagenase subfamily (*MEROPS* M9B). This was further confirmed by construction of a phylogenetic tree based on the amino acid sequences of the prepro forms of some typical bacterial collagenases, as shown in Fig. 8. The active-site motif and the third metal ligand Glu are all conserved in these enzymes.

The secondary structures of the prepro forms of the present enzyme and the *S. coelicolor* enzyme as predicted by the GOR4 program are compared in Fig. 9a and b. They are very much alike, but show some differences, especially in the N-terminal regions (positions 420–450) of the mature enzyme forms. The secondary structure contents of the mature forms of both enzymes were estimated to be as follows: present enzyme, α-helix, 20.2%, and β-strand, 24.3%; the *S. coelicolor* enzyme, α-helix, 12.4%, and β-strand, 34.9%. Similar results were obtained by the Chou-Fasman method (data not shown), and the secondary structure contents were estimated to be as follows: present enzyme, α-helix, 17.8%, and β-strand, 35.3%; the *S. coelicolor* enzyme, α-helix, 11.3%, and β-strand, 35.9%. Thus the present enzyme appears to be richer in α-helix than the *S. coelicolor* enzyme. Figure 9c and d shows a comparison of the hydropathy plots of the two enzymes. Again they are very similar, but show some differences. In the mature forms, significant differences are observed at positions 420–480 and 650–670. Hence both enzymes presumably possess very similar tertiary structures, but with some local differences. So far, however, no crystal structure has been solved for bacterial collagenases. Thus elucidation of the crystal

structures of the present enzyme and others is highly desirable for further understanding of the structure/function relationships and design of useful inhibitors of bacterial collagenases.

The results of Southern-blot hybridization analysis of the genomic DNA from *S. parvulus* are shown in Fig. 10. Although the DNA region used in Southern hybridizaion, nucleotides 1531–2423 in Fig. 6, involves no restriction site of the enzymes used for digestion of the genomic DNA, two or three hybridized bands are present in every lane of Fig. 10. This suggests the existence of another gene homologous to the cloned one. This gene might correspond to the minor enzyme that was not characterized, partly due to the paucity of the enzyme sample. Although the action of the minor enzyme on PZ-peptide was not analyzed, it is assumed to have the same specificity as the major enzyme, as judged by the specificity of the crude enzyme sample toward collagen.[19] Further studies are necessary to identify the minor enzyme.

Acknowledgments

We are grateful to Nitta Zeratin KK (Osaka, Japan) for the kind gift of crude collagenase (collagenase N-2). We also thank to Dr. S.B.P. Athauda for helpful technical advice. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1) Fukushima, J., and Okuda, K., *Vibrio* collagenase. In "Handbook of Proteolytic Enzymes" 2nd ed., Vol. 1, eds. Barrett, A. J., Rawlings, N. D., and Woessner, J. F., Elsevier Academic Press, London, pp. 414–415 (2004).
2) Van Wart, H. E., *Clostridium* collagenases. In "Handbook of Proteolytic Enzymes" 2nd ed., Vol. 1, eds. Barrett, A. J., Rawlings, N. D., and Woessner, J. F., Elsevier Academic Press, London, pp. 416–419 (2004).
3) Yoshida, E., and Noda, H., Isolation and characterization of collagenases I and II from *Clostridium histolyticum*. *Biochim. Biophys. Acta*, **105**, 562–574 (1965).
4) Kono, T., Purification and partial characterization of collagenolytic enzymes from *Clostridium histolyticum*. *Biochemistry*, **7**, 1106–1114 (1968).
5) Lwebuga-Mukasa, J. S., Harper, E., and Taylor, P., Collagenase enzymes from *Clostridium*: characterization of individual enzymes. *Biochemistry*, **15**, 4736–4741 (1976).
6) Bond, M. D., and Van Wart, H. E., Characterization of the individual collagenases from *Clostridium histolyticum*. *Biochemistry*, **23**, 3085–3091 (1984).
7) Matsushita, O., Yoshihara, K., Katayama, S., Minami, J., and Okabe, A., Purification and characterization of *Clostridium perfringens* 120-kilodalton collagenase and nucleotide sequence of the corresponding gene. *J. Bacteriol.*, **176**, 149–156 (1994).
8) Hanada, K., Mizutani, T., Yamagishi, M., Tsuji, H., Misaki, T., and Sawada, J., The isolation of collagenase and its enzymological and physico-chemical properties. *Agric. Biol. Chem.*, **37**, 1771–1781 (1973).
9) Lecroisey, A., Keil-Dlouha, V., Woods, D. R., Perrin, D., and Keil, B., Purification, stability and inhibition of the collagenase from *Achromobacter iophagus*. *FEBS Lett.*, **59**, 167–172 (1975).
10) Tong, N. T., Tsugita, A., and Keil-Dlouha, V., Purification and characterization of two high-molecular-mass forms of *Achromobacter collagenase*. *Biochim. Biophys. Acta*, **874**, 296–304 (1986).
11) Takeuchi, H., Shibano, Y., Morihara, K., Fukushima, J., Inami, S., Keil, B., Gilles, A. M., Kawamoto, S., and Okuda, K., Structural gene and complete amino acid sequence of *Vibrio alginolyticus* collagenase. *Biochem. J.*, **281**, 703–708 (1992).
12) Rippon, J. W., Extracellular collagenase produced by *Streptomyces madurae*. *Biochim. Biophys. Acta*, **159**, 147–152 (1968).
13) Chakraborty, R., and Chandra, A. L., Purification and characterization of a streptomycete collagenase. *J. Appl. Bacteriol.*, **61**, 331–337 (1986).
14) Endo, A., Murakawa, S., Shimizu, H., and Shiraishi, Y., Purification and properties of collagenase from a *Streptomyces* species. *J. Biochem.*, **102**, 163–170 (1987).
15) Ivanko, O. V., and Varbanets', L. D., Purification and physicochemical properties of *Streptomyces* sp. 1349 collagenase and *Streptomyces* sp. 1382 keratinase. *Mikrobiol. Z.*, **66**, 11–24 (2004).
16) Petrova, D., Derekova, A., and Vlahov, S., Purification and properties of individual collagenases from *Streptomyces* sp. strain 3B. *Folia Microbiol.* (Prague), **51**, 93–98 (2006).
17) Yu, M. S., and Lee, C. Y., Expression and characterization of the prtV gene encoding a collagenase from *Vibrio parahaemolyticus* in *Escherichia coli*. *Microbiology*, **145**, 143–150 (1999).
18) Luan, X., Chen, J., Zhang, X. H., Li, Y., and Hu, G., Expression and characterization of a metalloprotease from a *Vibrio parahaemolyticus* isolate. *Can. J. Microbiol.*, **53**, 1168–1173 (2007).
19) Wada, K., Nishio, T., and Yasugi, S., Japan Kokai Tokkyo Koho, 61-289885 (Dec. 19, 1986).
20) Wüensch, E., and Heidrich, H.-G., Zur quantitativen Bestimmung der Kollagenase. *Hoppe Seylers Z. Physiol. Chem.*, **333**, 149–151 (1963).
21) Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685 (1970).
22) Sambrook, J., and Russell, D. W., "Molecular Cloning, a Laboratory Manual" 3rd ed., Vol. 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001).
23) Thompson, J. D., Higgins, D. G., and Gibson, T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680 (1994).
24) von Heijne, G., A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.*, **14**, 4683–4690 (1986).
25) Garnier, J., Gibrat, J. F., and Robson, B., GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.*, **266**, 540–553 (1996).
26) Combet, C., Blanchet, C., Geourjon, C., and Deleage, G., NPS@: Network protein sequence analysis. *Trends Biochem. Sci.*, **25**, 147–150 (2000).
27) Chou, P. Y., and Fasman, G. D., Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **47**, 45–148 (1978).
28) Kyte, J., and Doolittle, R. F., A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132 (1982).
29) Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A., Protein identification and analysis tools on the ExPASy server. In "The Proteomics Protocols Handbook," ed. Walker, J. M., Humana Press, New York, pp. 571–607 (2005).
30) Hughes, J., Ward, C. J., Peral, B., Aspinwall, R., Clark, K., San Millán, J. L., Gamble, V., and Harris, P. C., The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat. Genet.*, **10**, 151–160 (1995).
31) Yeats, C., Bentley, S., and Bateman, A., New knowledge from old: *in silico* discovery of novel protein domains in *Streptomyces coelicolor*. *BMC Microbiol.*, **3**, 3–23 (2003).
32) Milton, D. L., Norqvist, A., and Wolf-Watz, H., Cloning of a metalloprotease gene involved in the virulence mechanism of *Vibrio anguillarum*. *J. Bacteriol.*, **174**, 7235–7244 (1992).
33) Matsushita, O., Jung, C. M., Katayama, S., Minami, J., Takahashi, Y., and Okabe, A., Gene duplication and multiplicity of collagenases in *Clostridium histolyticum*. *J. Bacteriol.*, **181**, 923–933 (1999).