Molecular Cloning and Sequence Analysis of the Gene Encoding the Collagenase from *Cytophaga* sp. L43-1 Strain

Yoshikyo SASAGAWA,** Kazuo IZAKI,*** Yuko MITSUBARA,* Koki SUZUKI,* Hisao KOJIMA,* and Yoshinori KAMIO**

Department of Applied Biological Chemistry, Faculty of Agriculture, Tohoku University, 1–1 Amaniyama-machi, Tsutsuimidori, Aoba-ku, Sendai 981, Japan

*Research Institute of Nippi, Incorporated, Adachi-ku, Tokyo 120, Japan

Received May 15, 1995

*Cytophaga* sp. strain L43-1 secretes a collagenase [Y. Sasagawa et al., *Biosci. Biotech. Biochem.*, 57, 1894–1898 (1993)]. A *cog* gene encoding the collagenase from this strain was cloned, and the nucleotides were sequenced. The structural gene of *cog* consisted of 3846 base pairs, which encoded a polypeptide consisting of 1282 amino acid residues with a predicted molecular mass of 130 kDa which was synthesized as a pre-matured enzyme. The deduced N-terminal 14 amino acids sequence, molecular mass of 120 kDa, and pI of 4.96 of the predicted matured enzyme were consistent with those previously found for the collagenase purified from the strain. The *cog* gene was expressed in *Escherichia coli* using the lac promoter and ribosomal binding sequence in plasmid vector pUC119 or pKK223-3, but not its own putative promoter and Shine-Dalgarno sequence. The consensus amino acid sequence (His-Glu-Xaa-Xaa-His) of an active site of the metal proteases including the collagenase from *Vibrio arginolyticus* and a series of human MMPs was found in the Cog protein of the strain.

Up to date, various properties of the collagenases have been clarified by enzymatic studies on the purified collagenase from a variety of microorganisms and animals. In addition, many genetic studies on collagenases have been done from the cloning of the cDNA from animals including human. However, there have been published only a few reports with respect to bacterial collagenase genes. In our previous paper, a strain of bacterium that secretes collagenase out of the cell was isolated and identified as *Cytophaga* sp. L43-1. We purified the collagenase from the culture supernatant of this strain. The purified enzyme hydrolyzes native collagen. The enzyme recognizes the Xaa-Gly-Pro sequence in the collagen and cleaves the Xaa-Gly bond. The enzyme also hydrolyzes β-casein at the position of Pro-Gly and Met-Ala. These cleavage sites are the same sites cleaved by the collagenase from *Achromobacter isographus*. However, our enzyme has no activity toward three synthetic oligopeptides, *p*-phenylazobenzoxycarbonyl-Pro-Leu-Gly-Pro-Arg, carboxybenzoxyl-Gly-Pro-Gly-Pro-Ala, and carboxybenzoxyl-Gly-Pro-Leu-Gly-Pro, which are hydrolyzed by the usual collagenases including the *Achromobacter* collagenase at the bond of Xaa-Gly in the peptides, suggest that a longer amino acid sequence containing Xaa-Gly-Pro is required to be recognized by our enzyme. These characteristics of the collagenase from *Cytophaga* sp. L43-1 led us to clone the collagenase gene (*cog*) from this strain to examine the chemical structure and function of the enzyme. In this paper, we describe the molecular cloning and sequence analysis of the *cog* gene from *Cytophaga* sp. L43-1. We then discuss the structural relationship between this enzyme and the reported ones on the basis of the deduced amino acid sequences.

**Materials and Methods**

Isolation of genomic DNA from *Cytophaga* sp. L43-1. The cells of *Cytophaga* sp. L43-1 were grown aerobically in the medium described previously until stationary phase. The cells were harvested and lysed with 1% SDS in TE at room temperature, followed by a treatment with proteinase K (200 μg/ml) at 37°C for 15 h. The lysate was then treated at 37°C with Rnase (0.01 mg/ml) for 1 h. The cell lysate was then centrifuged at 10,000 × g for 20 min and the supernatant was extracted 3 times with phenol-chloroform–isoamyl alcohol (25:24:1, v/v/v). The chromosomal DNA in the aqueous phase was extensively dialyzed against TE and stored at 4°C.

Preparation of oligonucleotide probe for cloning of *cog* gene. The N-terminal amino acids sequence of the purified collagenase preparation from *Cytophaga* sp. L43-1 strain was identified by Edman degradation. A DNA probe corresponding to the N-terminal amino acid sequence (from 4th to 17th) of the enzyme (Tyr-Val-Ile-Pro-Val-Tyr-Phe-Ala-Ile-Tyr-Xaa-Asp-Xaa-Gln) was designed and synthesized as 5'-TAT GTI ATI CCI GTI TAT TTT GCI ATI ATI TAT III GAT I CAG-3'. The synthetic oligonucleotide probe was labeled with γ-32P-ATP (110 TBq/mmol) by polynucleotide kinase.

Construction of a genomic library from *Cytophaga* sp. L43-1. The purified chromosomal DNA was digested with restriction enzymes and fractionated using 10–40% (w/v) sucrose density gradient ultracentrifugation. The fractions containing DNA fragments hybridized with 32P-oligonucleotide

---

† All of the data in this paper were presented at the Annual Meeting of Japan Society for Bioscience, Biotechnology, and Agrochemistry on April 3, 1994 at Tokyo, Japan, and in the first author's thesis.

The nucleotide sequence data reported in this paper appear in the GSDB, DDBJ, EMBL, and NCBI nucleotide sequence databases with the accession number D20660.

** Present address: Biochemical Laboratory, Nitto Boseki Co., Ltd., Koriyama 963, Japan.

*** Present address: Department of Civil Engineering, Tohoku Institute of Technology, Taihaku-ku, Sendai 982, Japan.

**** Corresponding author. Fax: (81) 22 272 1870.

E-mail: ykamio@biochem.tohoku.ac.jp

Abbreviations: TE, 10 mM Tris–HCl buffer (pH 8.0) containing 1 mM EDTA; MMP, matrix metal protease; Xaa, unidentified amino acid residue.
probe were pooled. The DNA fragments were ligated to the appropriate restriction site of the plasmid vector pUC119 or Charomid vector 9-28 (Wako Chemicals, Tokyo). After transformation or transduction into E. coli DH5z, the recombinant E. coli that contained cog gene were selected by colony hybridization using the 32P-oligonucleotide probe.

Southern blot hybridization. Southern blot hybridization was done using 32PATGATGACCGTATTGCAGTATTTGACAGCACG as a DNA probe as described by Maniatis et al.16

DNA sequence. A series of deletion derivatives were obtained from the plasmids, pCCE101, pCCH111, pCE101, and pCCX111 (Fig. 1) by exonuclease III and mung bean nuclease digestion, and self-ligation. DNA was sequenced by cycle sequencing with reverse and forward primers (Applied Biosystem) using a DNA sequencer (Model 373A, DNA Sequencing System, Applied Biosystems).

Plasmid construction for cog gene expression in E. coli. The cog gene was sub-cloned into the pUC119 and pKK223-3 to achieve a high level of lac inducible expression in E. coli. A summary of the strategy for construction of the plasmids is shown in Fig. 3.

i) Construction of plasmid pCPL1.2. Plasmid pCXX111 was digested with EcoRI and XbaI. The 4-kbp (EcoRI-XbaI) fragment was extracted and 5' and 3' sticky ends of the fragment were blunted with dNTP by the Klenow fragment. The fragment was then ligated into the HinflII site of pUC119.

ii) Construction of plasmid pCPL2-1KS. The 4-kbp blunt-ended (EcoRI-XbaI) fragment was ligated with the Ncol linker (5'-GGCGATCCGGCG). The ligated fragment was then digested with Ncol. After both 5' and 3' sticky ends of the fragment were blunted with dNTP, the fragment was ligated into the Smal site of the plasmid vector pKK223-3.

Gene expression. The cells of the plasmid-containing E. coli were grown at 37°C for 6 h in LB medium containing ampicillin (100 µg/ml) and 1 mM IPTG with rigorous shaking. The cells were harvested, resuspended in 10 mM Tris-HCl (pH 7.5) containing 4 mM CaCl2, and disrupted by sonication on ice. The supernatant, freed from the cell debris by centrifugation at 15,000 × g for 30 min, was analyzed on SDS-PAGE by the method of Laemmli.17 For detection of the enzyme activity by zymography, 10% polyacrylamide containing 0.1% gelatin was used as a separating gel on SDS-PAGE.18 After electrophoresis, the gel was rinsed with 2.5% Triton X-100 followed with 50 mM Tris-HCl (pH 7.5) containing 4 mM CaCl2 at room temperature. The rinsed gel was then incubated at 30°C for 5 h, and stained with 0.1% Amido black in methanol-water-acetic acid (3:6:1, v/v/v) for 1 h, and washed with the methanol-water-acetic acid (3:6:1, v/v/v). The bands of proteins having gelatlinolytic activity were observed as clear zones.

Results and Discussion

Molecular cloning of cog gene of Cytophaga sp. L43-1 strain

A genomic library of Cytophaga sp. L43-1 was constructed using pUC119 or Charomid 9–28 vectors, with E. coli DH5z as the host strain. The 1.2 Kbp EcoRI-, 2.0 Kbp HindIII-, 2.5 Kbp PstI-, or 5.0 Kbp XbaI-digested fragment of the chromosomal DNA from the strain was strongly hybridized with the synthesized oligonucleotide probe (data not shown). Therefore, we subcloned the EcoRI and HindIII fragments into the corresponding sites of pUC119, followed by transformation into E. coli DH5z. The resulting plasmids containing the EcoRI fragment in different orientations, and those containing the HindIII one were designated pCCE101 and pCCE201, and pCCH111 and pCCH211, respectively (Fig. 1). Analysis of the nucleotide sequence of the fragments showed that one open reading frame existed in the fragments but it was truncated at their 3' ends. Therefore, we attempted to clone the cog gene from another gene library that was constructed using Charomid 9–28 and partial digests of chromosomal DNA with Sau3A. A 0.4-kbp HindII fragment was labeled with α-32P-ATP by the random palimer method and used as a DNA probe for Southern hybridization. One positive clone containing an approximately 14-kbp insert of DNA was obtained. Physical mapping and Southern hybridization analysis showed that a 5-kbp XbaI fragment contained the full length of the cog gene (Fig. 1). Thus, this fragment was subcloned into the XbaI site of pUC119. The resulting plasmid, designated pCCX111, was introduced into E. coli DH5z.

Nucleotide sequence of the cog gene

The nucleotide sequence analyzed comprised 5653 bp from the XbaI to the EcoRI site (Fig. 2). Within this sequence, we can identify an open reading frame (ORF), comprising 3846 bp, which begins with an ATG codon and ends with a TAA codon. The amino acid sequence of collagenase deduced from the DNA sequence corresponds to a polypeptide with an apparent molecular mass of 130 kDa, comprising 1282 amino acid residues. The N-terminal 14 amino acid residues from the 4th to 17th ones deduced from the nucleotide sequence correspond with that of the purified collagenase of the strain with the exception of difference of the 9th and 11th amino acid residues between the predicted and the analyzed ones. The difference in two amino acid residues might be due to an error of the analysis of the amino acid sequence of the enzyme preparation.

On the basis of the N-terminal amino acid sequence of the predicted collagenase, presence of a putative 19 amino acid signal sequence that showed several characteristic features of signal peptide was indicated. A cluster of positively charged amino acids adjacent to the methionine residue at the N-terminus (Met-Lys-Thr-Lys) was followed by a region rich in hydrophobic amino acids (Leu-Ph-Leu-Phe-Val-Ser-Leu-Leu-Met-Met-Leu). The putative cleavage site, Ala⁴⁸-Gln⁴⁹, of the signal peptide was identified by the weight-matrix approach by Von Heijne.21 The position of Ala⁴⁹-Glu⁵⁰ of the predicted pre-matured collagenase might be self-processed to be secreted as a matured form of collagenase. These results indicate that the ORF encodes the pre-collagenase, and that the N-terminus of the pre-form is processed in its conversion to the matured form. The predicted matured protein, comprising 1216 amino acid residues, has a molecular mass of 120 kDa which

![Fig. 1. Restriction Map and Sequencing Strategy for Collagenase Gene (cog).](image-url)
A palindromic sequence, which may form a transcription termination signal, was found between the 22 bp from 3689 to 3705 bp and the 22 bp from 3711 to 3733 bp. This sequence was homologous to the $\rho$-independent terminator.

Expression of cog gene in E. coli

The cell lysate and the culture supernatant were prepared from cultures of E. coli JM109 harboring either pCCX111 or pCCX211, and analysed by SDS–PAGE and gelatin.
zymography. However, no enzyme activity was detected in either pCCX110 or pCCX211, suggesting failure of the cog gene expression in E. coli due to the incompatibility of the SD sequence or signal peptide of the signal coincides with that found by SDS-PAGE. The polarity and hydrophy of the collagenase according to Capaldi and Vanderkooi and Kyte and Doolittle, respectively, were 15.1% and no significant hydrophobic segment in the matured collagenase. A calculated isoelectric point value of 4.96 for the deduced sequence of the matured enzyme correlated closely with a value of 4.8, which was calculated experimentally for the enzyme.

From the DNA sequence analysis of the flanking region of the structural cog gene, a presumed but not typical ribosomal binding sequence (AAACAAA) was found 15 bp upstream of the ATG translation initiation codon, which may correspond to the Shine–Dalgarno (SD) sequence, peptidease in E. coli. Therefore, we constructed the plasmids,
pCCOL3-2, in which the region encoding the first 12 amino acid residues of β-galactosidase was substituted for the region encoding the N-terminal 66 amino acid residues of the predicted pre-matured collagenase in pCCX111. In this plasmid, the cog gene expression is under the control of the lac promoter and SD sequence in pUC119 (Fig. 3, see the Material and Methods). In the same manner, plasmid pCCOL3-1KS was constructed. In this plasmid, three amino acid residues added (Met-Ala-Ala) was inserted between the tac promoter region including an SD sequence that is compatible in E. coli and the first amino acid residue of the predicted matured collagenase (Fig. 3). After E. coli harboring either pCCOL3-2 or pCCOL3-1KS was grown in LB medium containing 1 mM IPTG, the cells were centrifuged. The culture supernatant and the cell lysate were prepared and analyzed by SDS-PAGE and by zymography. As shown in Fig. 4, the enzyme activity was detected on the zymogram as a clear band with molecular masses of 130 kDa, 120 kDa, 70 kDa, 60 kDa, or 50 kDa, in the cell lysate fractions from both E. coli (pCCOL3-1KS) and E. coli (pCCOL3-2). The molecular mass of 130 kDa and 120 kDa of the active bands coincide with that of predicted matured recombinant collagenase and purified matured collagenase preparation from Cytophaga sp., respectively. The bands of 70 kDa, 60 kDa, and 50 kDa are likely to be formed due to proteolysis of the enzyme. On the stained gel after SDS-PAGE, the extra protein band of 130 kDa was also observed in the cell extract from only recombinant E. coli (Fig. 4, B, Arrow). These results indicate that the cog gene in recombinants was expressed from the lac or tac promoter on the vectors. The amount of collagenase expressed in either E. coli (pCCOL3-2) or E. coli (pCCOL3-1KS) gave one-fifth of Cytophaga sp. one under the calculation on the basis of the intensity of the clear zone on zymogram per unit weight of the cells. The low expression of the gene in E. coli might be due to the gross dissimilarity in the codon usage for amino acids between E. coli and Cytophaga sp. strain L43-1. The GC content of the entire sequence of cog gene was low (33.1%), which is reflected in the preferential use of codons with A or T in the third codon. Overall, the codon usage of cog gene in Cytophaga sp. was dissimilar to that in E. coli, except that preferential use of GAA for Glu, AAA for Lys, and GTT and GTA for Val are in common both in Cytophaga sp. and in E. coli. On SDS-PAGE, the protein band of 130 kDa with enzyme activity that is about 10 kDa higher than that of the intact one from Cytophaga sp., was observed in the cell lysate of E. coli (pCCOL3-1KS). (Fig. 4, A and B). As mentioned above, the molecular mass of 130 kDa of the recombinant collagenase coincides with that of the predicted matured protein from the DNA sequence. It should be noted that the prematured collagenase with molecular mass of 130 kDa was expressed in E. coli and that the prematured collagenase was processed to the matured form with molecular mass of 120 kDa in E. coli.

**Amino acid homology among the collagenase and the reported metalloproteases**

The deduced amino acid sequence of collagenase from Cytophaga was compared for homology, to those of the collagenases of bacterial strains and metalloproteases from human and microorganisms. As shown in Fig. 5, the sequence of the putative active site motif of the collagenase, which was contained as His-Glu-Phe-Gly-His- from the 166th amino acid residue to the 169th one, completely

![Fig. 3. Construction of Plasmid, pCCOL3-1KS and pCCOL3-2.](image)

![Fig. 4. Analysis of Culture Supernatant and Sonic Extract of E. coli with Plasmid by Gelatin Zymography (A) and SDS-PAGE (B).](image)
The entire amino acid sequence of the *Cytophaga* enzyme showed no homology to these other enzymes described above except the 5-residue sequence in the active site shown in Fig. 5.

References


