Molecular Cloning of cDNA That Encodes Chymotrypsin Inhibitor ECI from *Erythrina variegata* Seeds and Its Expression in *Escherichia coli*

Jun-ichi KURAMITSU, Shiroh IWANAGA, Nobuyuki YAMASAKI, and Makoto KIMURA

Laboratory of Biochemistry, Faculty of Agriculture, Kyushu University, Fukuoka 812, Japan

Received March 11, 1996

Synthetic oligonucleotides representing all possible sequences of the N-terminal and internal amino acid sequences of the chymotrypsin inhibitor ECI from *Erythrina variegata* seeds were used to generate a probe specific for ECI-related sequences by the polymerase chain reaction on the *E. variegata* genomic DNA. A lambda phage cDNA library constructed from poly(A) RNA from maturing seeds was screened with the ECI gene thus obtained as a probe and characterized by DNA sequencing. The cloned ECI cDNA comprised 737 nucleotides and one open reading frame that encoded a polypeptide chain of 203 amino acids including a signal peptide composed of 24 amino acids. An expression plasmid was designed for export of the recombinant inhibitor into the periplasm. For this purpose, the cDNA fragment encoding matured ECI was ligated into the *NcoI* and *BamHI* sites following the *pelB* signal sequence in the expression vector pET-22b and expressed in *Escherichia coli* BL21 (DE3). However, this attempt failed as the recombinant inhibitor caused the formation of inclusion bodies in *E. coli* cells as a heterologous preprotein (SR-ECI), with the *pelB* upstream leader. SR-ECI was made soluble and renatured by refolding and reoxidation, and subsequently processed with pronase to give rise to recombinant ECI (R-ECI) that had an extra methionine residue attached to the N-terminal amino acid of ECI. Purified R-ECI inhibited chymotrypsin almost as strongly as authentic ECI.

Key words: chymotrypsin inhibitor; *Erythrina variegata*; cDNA cloning; expression

Many serine proteinase inhibitors in the Kunitz family have been isolated from leguminous plants, and their structures and functions have been investigated extensively. The tertiary structure of complex of trypsin and soybean trypsin inhibitor (STI) has been analyzed with resolution of 2.6 Å, so the molecular mechanism of the interaction of STI and trypsin is understood, and STI is used as a model inhibitor in studies of substrate-trypsin interaction. It seems, however, that STI can bind two molecules of chymotrypsin and that a preformed binary complex of STI-trypsin still binds one molecule of chymotrypsin. These results suggested that the STI molecule has a second reactive site for chymotrypsin.

We have purified one chymotrypsin inhibitor, ECI, from the seeds of *Erythrina variegata*; The genus *Erythrina* is of deciduous leguminous trees and shrubs found widely in the tropics and subtropics. We also sequenced the amino acids of ECI by protein chemical methods. ECI consists of 179 amino acid residues in a single polypeptide chain and it forms a stoichiometric 1:2 complex with chymotrypsin, judging from the titration pattern of inhibitory activity. Limited proteolysis with chymotrypsin under acidic conditions unambiguously identified one of the reactive sites to be Leu**Ser** but no information was obtained about a second reactive site for chymotrypsin. Similar observations were previously reported for chymotrypsin inhibitors isolated from winged beans and *Erythrina latisima* seeds.

Kunitz family inhibitors that inhibit cysteine proteinase have been isolated from potato tubers; the aspartic acid proteinase inhibitor inhibited trypsin as well. A subtilisin-2-amylase inhibitor with sequence similarity to Kunitz family inhibitors was isolated from barley, but its reactive site for 2-amylase has not been reported.

These observations, together with those described above for Kunitz family inhibitors, led us to speculate that Kunitz family inhibitors might be “double heads”, one reactive site is for a serine proteinase and the other might be for serine proteinase, cysteine proteinase, aspartic acid proteinase, or 2-amylase. If so, the gene encoding a Kunitz family inhibitor might be suitable for use in the design of new multifunctional proteinase inhibitor by protein engineering. Study of the structural basis of interaction of these inhibitors with proteinases could lead to a more successful application of the Kunitz family inhibitor in drugs or transgenic plants, where it might confer pathogen resistance.

As an initial step, we decided to clone cDNA encoding ECI and express it at a high level in *Escherichia coli* cells. In this paper, we describe the molecular cloning of cDNA encoding ECI and its expression in *E. coli* with a T7 polymerase expression system.

**Materials and Methods**

*Materials.* Developing *E. variegata* seeds were harvested in May 1994 in Okinawa, Japan, and stored at –80°C until use. The oligonucleotides used in this study were purchased from Greiner Japan. A kit with predigested pET11a vector and a Time Saver cDNA synthesis kit

---

1. To whom correspondence should be addressed.

*Abbreviations:* ECI, *Erythrina* chymotrypsin inhibitor; IPTG, isopropyl-β-D-thiogalactopyranoside; ORF, open reading frame; R-ECI, recombinant ECI; RP-HPLC, reverse-phase high-performance liquid chromatography; STI, soybean trypsin inhibitor; SR-ECI, recombinant ECI with a signal sequence; TFA, trifluoroacetic acid; UTR, untranslated region.
were obtained from Pharmacia. A GigaPack II Plus phage packaging kit was purchased from Stratagene. A GeneAmp R PCR reagent kit with Ampli Taq R DNA polymerase, random primer DNA labeling kit, DNA ligase, and RedTaq DNA sequencing kit were obtained from Takara. Restriction enzymes and T4 DNA ligase were from Gibco BRL, and used as recommended by the supplier. Nitrocellulose membranes were from Schleicher & Schuell. x-32P dCTP was obtained from Amersham. x- 
Chymotrypsin and Pronase (a protease from Streptomyces griseus) were obtained from Sigma Chemical Co. All other reagents were of analytical grade for biochemical use.

Isolation of plant DNA. Genomic DNA was isolated from the germinating seeds of E. variegata by the method of Taylor and Powell.15 Approximately 3.8 mg high-molecular-mass DNA was obtained from 8 g of ground seed tissue. Southern blotting was done as described elsewhere,16 with the randomly primed ECI gene fragment labelled with 32P as the probe and then, the blotting was washed thoroughly.

RNA isolation and cDNA library construction. Total RNA was isolated from developing E. variegata seeds by the method of Manning.14 Poly(A+)
RNA was isolated with Oligo(dT)30 Super (Takara). Subsequently, a cDNA library was constructed with the cDNA synthesis kit and zgl11 as the cloning vector. For Northern blotting, total RNA was electrophoresed on a 1.5% agarose gel containing 2.2M formaldehyde, transferred to nitrocellulose membranes, and hybridized as in Southern blotting.

Preparation of ECI gene fragments. DNA fragments of ECI were amplified by PCR with E. variegata genomic DNA as the template. On the basis of the amino acid sequence of the ECI protein sequence found (EPLVYDELG at positions 18 and PWWTYVE at positions 88-94),17 two mixed oligonucleotide primers, 5′-GAAGCCGCCTTGTGAGCTTCTTGATGAAGG-3′ and 5′-TCACACGCTCCACCAAGG-3′ were designed. The PCR was done as recommended by the manufacturer on an automated thermal sequencer (ASTEC, Program Temp Control System PC-700). A 300-bp fragment was amplified, purified on a gel, treated with T4 DNA polymerase, and then cloned into the pUC18 vector digested by Smal. After confirmation of its sequence, the fragment was used as a hybridization probe to screen the cDNA library.

Screening of the E. variegata seed cDNA library. About 1.5 × 108 recombinant phages were plated on a lawn of E. coli Y1088 cells and the phages were screened by plaque hybridization with the ECI gene fragment as the probe as described before.11 Clones that hybridized with the PCR fragment in the second and third screenings were isolated and characterized.

DNA sequencing. Phage zgl11 DNA were prepared according to the standard procedure and the cDNA insert was obtained by EcoRI digestion, ligated to the vector pUC18, and then used to transform competent E. coli J1115. Restriction endonucleases were isolated from alkaline lysates and the DNAs were sequenced by the dyeoxy chain termination method19 with BoaBest sequencing kit, as described previously.

Expression of ECI in E. coli cells. For ligation of only the DNA sequence that codes matured ECI into the expression vector pET-22b (Novagen),18 the ECI cDNA fragment from nucleotide 1 to 540 of the sequence of the cloned ECI and flanked by artificial sites for Ncol and BamHI, was amplified by PCR with primers, 5′-CCATGGACCATCTTGCTGGATGTGAAGG-3′ and 5′-GGCGGATCCCTTGATGTGTTGAGTGGT- AGC-3′. After digestion of the PCR product with Ncol and BamHI, the DNA fragment was inserted between the Ncol and BamHI sites of pET-22b to yield the plasmid pETECI. The plasmid constructed was introduced into E. coli strain BL21(DE3) cells grown on LB plates containing 50 μg/ml ampicillin. BL21(DE3) cells containing pETECI were grown to the mid-log-phase at 37°C and then treated with IPTG at the final concentration of 1 mM. After this addition, the cells were cultured for 5h and then collected by centrifugation. The pellet was washed twice in 10 mM phosphate buffer, pH 7.2, containing 0.1 M NaCl and then lysed with 1% SDS. Preparation of the periplasmic fraction by osmotic shock and of the inclusion bodies after sonication of the cells were done by the methods of Nossal and Heppel20 and Marston,21 respectively. Samples were analyzed by SDS PAGE on a 17.5% acrylamide gel.22 Western blotting was performed as described in the legend to Fig. 2. Anti-ECI antibodies, developed using anti-mouse IgG conjugated with peroxidase, and then immune complexes were stained with diaminobenzidine as described by Towbin et al.23 Purification of the recombinant ECI (R-ECI). E. coli transformants harbouring the expression vector pETECI were cultured in 1 liter of L-broth containing 50 μg/ml ampicillin. Cells expressing recombinant ECI were harvested and washed twice in 30 mM Tris HCl, pH 7.5, containing 30 mM NaCl. Cells were sonicated on ice at a setting of 7 for 2 min periods followed by 5 min for cooling for a total of 70 min. Cell sonicates were spun at 12,000 × g for 20 min at 4°C to separate the pellet from the supernatant. The pellet thus obtained was suspended in 1 M sucrose solution and again centrifuged at 12,000 × g for 20 min for separation of inclusion bodies. The pellet was suspended in 2% Triton X-100 containing 10 mM EDTA and collected again by centrifugation at 12,000 × g for 20 min. The recombinant protein was made soluble in 30 mM Tris HCl, pH 7.5, containing 30 mM NaCl, 1 mM dithiothreitol, and 8 M urea, and dialyzed against the same buffer containing 4 M urea instead of 8 M urea and then against water, and finally lyophilized. The recombinant protein was purified by reverse-phase HPLC (RP-HPLC) on a YMC gel C4 column (4.6 × 250 mm) equilibrated with 0.1% trifluoroacetic acid (TFA). The protein was eluted with a linear gradient of 0 to 56% acetonitrile in 0.1% TFA for 30 min. The recombinant protein thus obtained was digested with pronase in the same buffer, pH 7.2, at an enzyme-substrate ratio of 1:100 for 5 h at 37°C. The purified protein was analyzed by SDS PAGE as described above. The N-terminal amino acid sequence of the protein was analyzed by a gas phase sequencer PSQ-1 (Shimadzu). Chymotrypsin inhibition was assayed as described previously,24 with casein (Merck) as the substrate. The concentration of chymotrypsin was estimated spectrophotometrically with a coefficient of 20.4 for a 1% solution and a 1-cm light-path. The concentrations of ECI and recombinant proteins were measured by amino acid analysis.

Results and Discussion

Construction and screening of the E. variegata cDNA library.

Northern blotting of total RNA with the ECI gene fragment obtained by PCR showed that the ECI transcript was some 800-bp in length. This RNA preparation was used to construct the cDNA library. Six micrograms of poly(A) RNA was purified from 500 μg of total RNA by affinity chromatography on Oligotex-dT30. Double-stranded cDNA was synthesized and inserted into the vector zgl11 at the EcoRI site and then the recombinant phages were used to transfect E. coli Y1088. From the library (1.5 × 108), six clones that gave strong hybridization signals were detected. Digestion of DNA isolated from these clones with EcoRI released 800-bp inserts that hybridized on Southern blotting to the PCR fragment used in the screening (data not shown). We cloned the largest insert into the EcoRI site of pUC18 and determined its nucleotide sequence.

Sequence of ECI cDNA

The nucleotide sequence of the clone that contained the largest insert and its deduced amino acid sequence are shown in Fig. 1. The cDNA had a single open reading frame (ORF) coding for 203 amino acids. It was preceded by a short untranslated region (UTR) of seven nucleotides (excluding the EcoRI linker sequence) and followed by a long UTR of 107 nucleotides excluding the polyA tail. The clone was close in length to that of ECI mRNA (800 bp) estimated by Northern blotting.

We compared the amino acid sequence deduced from the nucleotide sequence of ECI cDNA with that reported previously.24 Comparison showed that the ORF encoded a 203-residue prehormone with a signal peptide. The N-terminal 24 amino acids resemble a classical signal sequence, with a charged residue near the N-terminus followed by a long stretch of hydrophobic amino acids and
Fig. 2. Southern Blotting of E. variegata Genomic DNA.

Ten micrograms of genomic DNA from E. variegata was digested with the restriction enzymes EcoRI (lane 1) and BamHI (lane 2), and electrophoresed on a 0.8% agarose gel. ECI cDNA was used as the probe and hybridized as described in ref. 17. HindIII-digested lambda phage DNA was used as the size marker.

forms of the inhibitor. During purification of ECI by ion-exchange chromatography, in addition to the ECI protein used for the protein sequence study, another component with almost the same inhibitory activity was isolated (Kimura et al., unpublished).

In this connection, we examined the gene organization of the ECI structural gene in a preliminary way by Southern blotting. High-molecular-mass DNA from E. variegata was digested to completion with either EcoRI or HindIII and analyzed by Southern hybridization with the ECI cDNA clone obtained in this study. The ECI cDNA hybridized to three EcoRI fragments (two major fragments of 6.0 kb and 2.2 kb, and one minor fragment of 2.0 kb) and two HindIII fragments (one major fragment of 3.8 kb and one minor fragment of 5.6 kb, Fig. 2). ECI cDNA had neither EcoRI nor HindIII recognition sites, so the result suggest the presence of at least two structural genes in the E. variegata genome, although it is possible that there are EcoRI and HindIII recognition sites in an intron(s).

Expression of ECI in E. coli cells

Disulfide-bonded proteins, such as bovine pancreatic trypsin inhibitor and human growth hormone, when secreted into the E. coli periplasmic space, can be expressed, not forming aggregates and retaining full activity. Hence, we tried to express ECI, which contains two disulfide bonds, as a heterologous preprotein with the pelB signal sequence as its N-terminus. For this purpose, the cDNA fragment (positions 1–540) encoding mature ECI was specifically amplified by PCR and fused to the signal sequence as shown in Fig. 3, in the expression plasmid vector pET-22b. In this construction, we expected the recombinant protein to be secreted into the periplasm of E. coli cells. The expression plasmid PET-21 was transformed into E. coli strain BL21 (DE3) cells, which were treated as described in Materials and Methods. Synthesis of ECI was first looked for by SDS-PAGE followed by Western blotting of a cell
Fig. 3. Construction of the Expression Vector pETEC1.
The cDNA fragment encoding mature ECI was amplified by PCR and ligated with the expression vector pET-22b digested with the restriction enzymes NdeI and BamHI. The resulting plasmid, pETEC1, was used to transform E. coli BL21 (DE3).

SDS-PAGE
Western blotting

Fig. 4. Overproduction of Recombinant ECI Examined by SDS-PAGE and Immunoblotting.
Lanes 1 and 7 contain control ECI. Lanes 2 and 8 contain E. coli BL21 (DE3) cells carrying pETEC1 and not treated with IPTG; lanes 3 and 9 contain the same kind of cells 5 h after IPTG was added to 1 mM. Lanes 4 and 10 contain proteins obtained by osmotic shock. Lanes 5 and 11 contain soluble proteins obtained by sonication. Lanes 6 and 12 contain proteins obtained from inclusion bodies.

lysate. A protein of 23 kd was detected on SDS-PAGE (Fig. 4, lanes 3 and 9) and it reacted with anti-ECI polyclonal antibody after IPTG treatment, showing that the recombinant ECI protein was some 2 kd larger than the authentic protein (lanes 1 and 7). Then the bacterial pellet was suspended in 30 mM Tris–HCl, pH 8.0, containing 5 mM EDTA and 20% sucrose, and the mixture was incubated for 30 min. The periplasmic fraction in the supernatant thus obtained and the sonic extracts were examined by

Fig. 5. Pattern of RP-HPLC of Recombinant ECI Obtained from Inclusion Bodies.
Proteins from inclusion bodies were purified by RP-HPLC on a YMC gel C4 column (4.6 × 250 mm) equilibrated with 0.1% TFA at the flow rate of 1.0 ml min. The proteins were eluted with a linear gradient of 0 to 56% acetonitrile in 0.1% TFA. The recombinant protein obtained in the first peak was designated SR-ECI.

SDS-PAGE. The results showed no recombinant protein in the periplasmic fraction (lanes 4 and 10), where we had expected to find it; rather, most of the protein was in the inclusion bodies (lanes 6 and 12). Our attempt to cause production of recombinant ECI in the E. coli periplasmic space had failed. Instead, recombinant ECI with a signal peptide of 22 amino acids seemed to be produced in the
inclusion bodies.

Purification and characterization of the recombinant protein

The protein fraction was refolded by dialyzing against 4 M urea and then water, in turn. The recombinant protein was purified by RP-HPLC and evaluated in terms of purity and inhibition of chymotrypsin. RP-HPLC of the protein produced two peaks (Fig. 5). The protein in the first peak inhibited chymotrypsin, but the protein in the second peak did not (Fig. 6). Sequence analysis of the proteins, however, gave a single N-terminal sequence, Met-Lys-Tyr-Leu-Leu-, identical to that of the pelB reader sequence, so a signal peptide preceded the ECI protein. These observations suggested that the protein eluted at the first peak may have been correctly folded with correct disulfide bonds, but that the protein in the second peak might be incorrectly folded or folded with free cysteine residues. The recombinant protein obtained in the first peak was designated SR-ECI and digested enzymatically.

Several attempts to obtain the recombinant ECI (R-ECI) without the pelB signal sequence were made. A preliminary experiment showed that pronase digestion of SR-ECI gave a single protein band almost identical to that of authentic ECI in SDS-PAGE, so SR-ECI was digested by pronase to cleave the pelB signal peptide, and the resultant protein was purified by RP-HPLC. Figure 7 shows the results of SDS-PAGE of the protein R-ECI, together with those of authentic ECI and SR-ECI. Sequence analysis of the recombinant protein thus obtained gave the N-terminal sequence Met-Glu-Pro-Leu-. This finding indicated specific cleavage by pronase at the peptide bond between Ala-Met in the pelB signal sequence, giving rise to R-ECI which has an extra methionine residue attached to the N-terminal amino acid residue of ECI.

The ability of R-ECI to inhibit chymotrypsin was examined with casein as a substrate. R-ECI inhibited chymotrypsin (K_m 181 nm) to almost the same extent as authentic ECI (K_m 55 nm, Fig. 6), so the R-ECI molecule was correctly refolded into the native conformation and retained its inhibitory activity.

References