Chinese Science Bulletin 2003 Vol. 48 No.1 68—71

cDNA cloning and expression of earthworm fibrinolytic enzyme component A

LIU Junfeng, WANG Xinquan, XU Lei, ZHANG Jiping,
LIANG Dongcai & CHANG Wenrui

National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
Correspondence should be addressed to Chang Wenrui (e-mail: wrchang@sun5.ibp.ac.cn)

Abstract **Earthworm fibrinolytic enzyme component A (EFEa), a protein with dual fibrinolytic activity, is one of the major therapeutically important earthworm fibrinolytic enzyme components. The cDNA fragment encoded the mature protein was cloned from earthworm (*Eisenia fetida*) by the RT-PCR technique. The deduced amino acid sequence of the EFE component A shows high homology with some members of serine proteases trypsin family, and the amino acid residues constituting the active sites are conserved in the EFEa as compared with the other proteins of the trypsin family. The cDNA fragment was subcloned into the expression vector pQE31 and pMAL-c2X of *E. coli*. The resulting expression plasmids, pQE-*efea* and pMAL-*efea*, were used to transform the *E. coli* strain M15. Recombinant protein bands corresponding with calculated molecular weights were induced. The induced His_6-EFEa fusion protein with pQE-*efea* was accumulated into inclusion body, while the induced MBP-EFEa fusion protein with pMAL-*efea* was soluble and showed fibrinolytic activities.**

Keywords: earthworm fibrinolytic enzyme, prokaryotic expression, cDNA.

Thrombosis is one of the most widely occurring diseases in modern life, which often causes disability and death. Medications using fibrinolytic enzymes are the most effective methods used in the treatment of thrombosis. Earthworm fibrinolytic enzyme (EFE) from earthworm was characterized and became commercially available in Korea and China as novel oral-administered fibrinolytic agents for prevention or treatment of cardiac and cerebrovascular diseases.

Further characterization revealed that EFE was composed of several components[1—8]. The amino acid sequences of these components are the basis for the analysis of their mechanisms of the fibrinolytic activities. Five cDNA sequences that encoded the components in EFE have been isolated from the earthworm (GenBank accession numbers: AB045719, AB045720, U25648, U25643 and AF109648)[8]. Four of them were isolated from *Lumbricus rubellus*, which encoded components F-Ⅱ-1(GenBank accession numbers: AB045720 and U25648) and F-Ⅱ-2 (GenBank accession numbers: AB045719 and U25643) respectively. The cDNA AF109648 was isolated from *Lumbricus bimastus*. Cloning of the genes that encoded the components in EFE will help to get more information about the molecular mechanisms of them to degrade fibrin.

Recently we have purified and crystallized a novel component in EFE from *Eisenia fetida* and named it earthworm fibrinolytic enzyme component A (EFEa)[9]. It is one of the major therapeutically important EFE components, and functions both as a direct fibrinolytic enzyme and a plasminogen activator. Some characteristics of EFEa, such as the molecular weight and the N-terminal 24 residues, have been obtained. The determination of the structure of EFEa is being carried out in our lab. In this note, we described the cloning of the cDNA fragment encoding the mature protein and expression of it in *E. coli*.

1 Materials and methods

(ⅰ) Strains and plasmids. Vector pMD18-T was from Takara Co. *E. coli* strain M15, pQE-31 were purchased from Qiagen Co. Vector pGEM-T was from Promega Co. Vector pMAL-c2X was purchased from New England Biorads Co.

(ⅱ) RT-PCR amplification and gene cloning. The total RNA was isolated from fresh earthworm using Trizol methods. The reaction of first-strand cDNA synthesis was catalyzed by Superscript Ⅱ RNaseH⁻ reverse transcriptase using random primers. A reaction product of 2 μL was added in the 50 μL PCR amplification system. The primers were as follows: primer 1: 5′-GTNATHGGNGGNA-CNAAYGC-3′ (N = A/G/T/C, H = A/C/T, Y=C/T), primer 2: 5′-CTAACGAGAGTTGTCTCCGA-3′. Primer 1 was a degenerate primer that was designed based on the N-terminal sequence of the EFEa (VIGGTNA), and primer 2 was designed based on the cDNA sequence of a component in EFE from *Lumbricus bimastus* (GenBank accession number: AF109648). PCR reactions were catalyzed by Ex Taq polymerase (Takara Co.) and the product was ligated into vector pMD18-T. The resulting construct, pMD-*efea*, was sequenced by the dideoxy chain termination method of Sanger. The inserted DNA fragment digested from pMD-*efea* with *Bam*HⅠ and *Hind*Ⅲ was cloned into pQE31 to get pQE-*efea*. The DNA fragment amplified from pMD-*efea* with primers gaattcCGTGATAGGGGGCACTA and aagcttCTAACGAGAGTTGTCTCC was ligated into vector pGEM-T. The resulting con-

struct, pGEM-*efea*, was digested with *Eco*R Ⅰ and *Hind* Ⅲ, and the DNA fragment including the ORF was cloned into pMAL-c2X to get pMAL-*efea*.

(ⅲ) Expression of the recombinant His_6-EFEa. Overnight culture of cells bearing the construct for expression was diluted 1∶100 into fresh LB broth in the presence of 100 μg/mL ampicillin and 30 μg/mL kanamycin. When the cell suspension reached an absorbance of 0.6 at 600 nm, the expression was induced by the addition of 1 mmol/L IPTG for 4 h. Cells were harvested by centrifugation at 6000 *g* for 10 min and resuspended in isolation buffer (0.5 mol/L NaCl, 50 mmol/L Tris-HCl, pH 8.0). The cells were sonicated on ice bath and the lysate was centrifuged at 16000 *g* for 30 min. Expression of the fusion protein was analyzed by SDS -PAGE.

(ⅳ) Expression and purification of the recombinant protein MBP-EFEa. Overnight culture of cells bearing the construct for expression were diluted 1∶100 into fresh LB broth in the presence of 100 μg/mL ampicillin and 30 μg/mL kanamycin. When the cell suspension reached an absorbance of 0.6 at 600 nm, the expression was induced by the addition of 0.3 mmol/L IPTG for 4 h. Cells were harvested by centrifugation at 6000 *g* for 10 min and resuspended in isolation buffer (20 mmol/L Tris–HCl, pH 7.4, 0.2 mol/L NaCl, 1 mmol/ L EDTA). The cells were sonicated on ice bath and the lysate was centrifuged at 16000 *g* for 30 min. The purification of the recombinant protein was performed according to the manufacture's instructions (New England Biorads Co).

(ⅴ) SDS-PAGE. The preparing of SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and dying with the Coomassie blue were performed according to Sambrook and Rusell[10].

(ⅵ) Analysis of the sequences of DNA and protein. The analysis, translation of DNA sequences were performed with DNAMAN (Ver 4.0). A homology search and alignment of the protein sequences were done with the BLAST (http://www.ncbi.nlm.nih.org) and the Interpro database was used to analyze the active sites of proteins (http://www.ebi.ac.uk/interpro).

(ⅶ) Analysis of the fibrinolytic activities. The fibrinolytic activities were measured using standard fibrin plates[4]. The enzyme solution of 10 μL (0.4 mg/mL in 0.1 mol/L sodium phosphate buffer, pH 7.6) was spotted onto the fibrin plates and incubated at 37℃ for 8 h.

2 Results

(ⅰ) Cloning of the EFEa cDNA fragment. Utilizing the designed primers, a cDNA fragment was amplified from the total RNA of *Eisenia fetida* by the RT-PCR technique (Fig. 1).

The length of this fragment from PCR reaction was in accordance with the expected value. The fragment was excised from agarose gel and ligated into vector pMD18-T to get pMD-*efea*. The nucleotide sequence and the deduced amino acid sequence of the insert are shown in Fig. 2 (GenBank accession number: AF393512).

The 726 bp DNA fragment was predicted to encode a polypeptide of 242 amino acid residues, and the N-terminal 24 residues were the same as those of the native

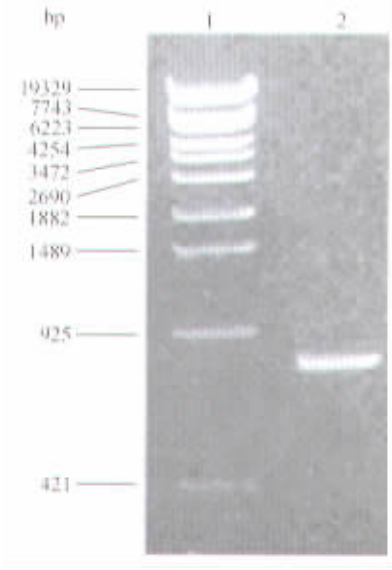

Fig. 1. The agarose gel electrophoresis of PCR products. 1, λ-EcoT14I DNA marker; 2, PCR products.

```
1    GTG ATA GGG GGC ACT AAC GCT AGC CCA GGA GAG TTC CCA TGG CAA   45
1     V   I   G   G   T   N   A   S   P   G   E   F   P   W   Q    15

46   CTG TCT CAG CAA CGC CAA AGC GGT TCT TGG TCG CAC AGC TGC GGA   90
16    L   S   Q   Q   R   Q   S   G   S   W   S   H   S   C   G    30

91   GCA TCG CTT CTG AGC TCA ACT TCC GCT CTC AGC GCA TCG CAC TGC  135
31    A   S   L   L   S   S   T   S   A   L   S   A   S   H   C    45

136  GTC GAT GGA GTG CTA CCC AAC AAC ATC CGA GTC ATC GCT GGT CTT  180
46    V   D   G   V   L   P   N   N   I   R   V   I   A   G   L    60

181  TGG CAA CAG TCG GAC ACA AGC GGC ACT CAG ACC GCT AAC GTC GAT  225
61    W   Q   Q   S   D   T   S   G   T   Q   T   A   N   V   D    75

226  AGC TAC ACC ATG CAT GAG AAC TAT GGT GCT GGT ACT GCT TCG TAC  270
76    S   Y   T   M   H   E   N   Y   G   A   G   T   A   S   Y    90

271  TCC AAT GAC ATT GCC ATT CTG CAC CTC GCA ACG TCC ATC AGC CTC  315
91    S   N   D   I   A   I   L   H   L   A   T   S   I   S   L   105

316  GGA GGA AAC ATC CAG GCA GCT GTC CTT CCC GCC AAC AAC AAC AAC  360
106   G   G   N   I   Q   A   A   V   L   P   A   N   N   N   N   120

361  GAC TAC GCC GGA ACC ACA TGC GTC ATC TCC GGA TGG GGT CGC ACA  405
121   D   Y   A   G   T   T   C   V   I   S   G   W   G   R   T   135

406  GAT GGA ACG AAC AAT CTG CCG GAC ATC CTT CAG AAG TCG TCA ATT  450
136   D   G   T   N   N   L   P   D   I   L   Q   K   S   S   I   150

451  CCG GTC ATA ACG ACC GCC CAG TGC ACC GCC GCC ATG GTT GGA GTC  495
151   P   V   I   T   T   A   Q   C   T   A   A   M   V   G   V   165

496  GGT GGA GCC AAC ATC TGG GAT AAT CAC ATC TGC GTC CAG GAT CCT  540
166   G   G   A   N   I   W   D   N   H   I   C   V   Q   D   P   180

541  GCT GGC AAC ACC GGA GCC TGC AAT GGC GAT AGC GGT GGC CCA CTG  585
181   A   G   N   T   G   A   C   N   G   D   S   G   G   P   L   195

586  AAC TGC CCA GAC GGC GGA ACT CGA GTG GTT GGT GTT ACT TCG TGG  630
196   N   C   P   D   G   G   T   R   V   V   G   V   T   S   W   210

631  GTT GTA TCC AGT GGC CTT GGT GCA TGT CTT CCG GAC TAC CCT TCC  675
211   V   V   S   S   G   L   G   A   C   L   P   D   Y   P   S   225

676  GTC TAC ACC CGC GTC AGC GCC TAC TTG GGT TGG ATC GGA GAC AAC  720
226   V   Y   T   R   V   S   A   Y   L   G   W   I   G   D   N   240

721  TCT CGT TAG                                               729
241   S   R   *                                                 242
```

Fig. 2. Nucleotide and deduced amino acid sequences of cDNA encoding EFEa.

EFEa. The molecular weight of the polypeptide is similar with that of the native EFEa (24667 D/24762 D). The coding sequence of EFEa shared 99.6%, 32.5% and 37.8% identities with genes AF109648 from *Lumbricus bimastus*, AB045719 and AB045720 from *Lumbricus rubellus*. At the amino acid level the identities were 99.2%, 27.2% and 27.8%.

A search of GenBank using the program BLAST revealed that the deduced primary sequence of EFEa shows high homology with some serine proteases. The sequence identities between EFEa and lugworm chymotrypsinogen (X95078), Pacific white shrimp trypsin (X86369), cow preproelastase II (X97635), Norway rat elastase II precursor (L00124) and African clawed frog pancreatic elastase 1 (AF468646) were 36%, 34%, 33%, 32% and 32%, respectively. EFEa also showed similar active sites with members of Serine proteases, trypsin family (IPR001254) based on the search of the Interpro database (Table 1). The catalytic triad composed of residues His57, Asp102 and Ser195 in trypsin correspond to His49, Asp93, and Ser191 in EFEa respectively. But the primary substrate specificity determinant, Asp189 in trypsin was changed to Gly185 in EFEa. These results show that EFEa may be a serine protease of trypsin family, and the substrate specificity is different between EFEa and trypsin.

Table 1 Some characteristic of EFEa function sites

	Catalytic triad			PSSD[a)]
	His	Asp	Ser	
Trypsin	57	102	195	Asp189
F-III-1&2	43	91	188	Asp182
EFEa	44	93	191	Gly185

a) PSSD: Primary substrate specificity determinant.

(ii) Expression of recombinant His_6-EFEa. The fragment encoding EFEa was cloned into the pQE31 vector, and the resulting plasmid pQE-*efea* will generate an expression product with the expected size of 28385 D (275 aa). The plasmid was transformed into the *E. coli* M15. The expression was induced with 1 mmol/L IPTG. In the SDS-PAGE gel, a protein band about 30 kD appeared after 4 h induction, but the induced protein existed in the inclusion bodies (Fig. 3).

(iii) Expression and purification of recombinant MBP-EFEa. The fragment encoding EFEa was cloned into the pMAL-c2X vector, and the resulting plasmid pMAL-*efea* will generate an expression product with the expected size of 67672 D (633 aa). The plasmid was transformed into the *E. coli* M15. The expression was induced with 0.3 mmol/L IPTG. In the SDS-PAGE gel, a protein band about 67 kD appeared and the induced protein existed in the supernatant (Fig. 4). The recombinant protein was purified using affinity chromatography. The fibrinolytic activities of the recombinant protein were analyzed using a fibrin plate. The recombinant protein dissolved fibrin plate and showed fibrinolytic activities, which was lower than that of the native EFEa (Fig. 5).

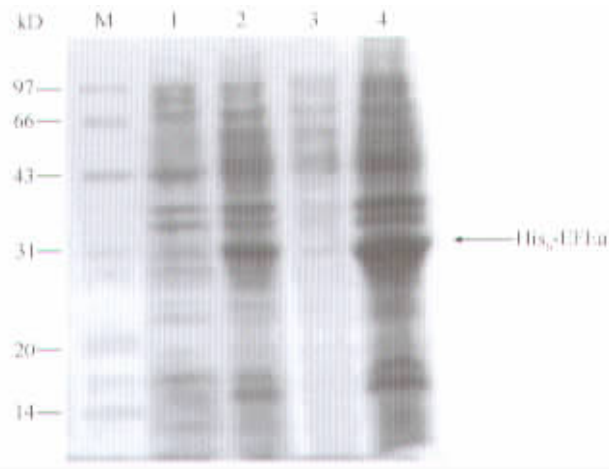

Fig. 3. Analysis of the expression of recombinant His_6-EFEa by SDS-PAGE. 1, Noninduced cells; 2, cells induced with IPTG; 3, cleared lysate; 4, pellet. M, low molecular mass protein standards.

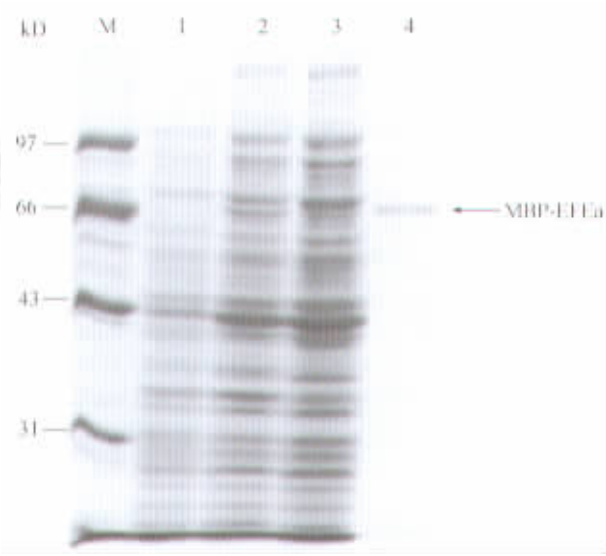

Fig. 4. Analysis of the expression of recombinant MBP-EFEa by SDS-PAGE. 1, Noninduced cells; 2, cells induced with IPTG; 3, cleared lysate; 4, recombinant MBP-EFEa. M, low molecular mass protein standards.

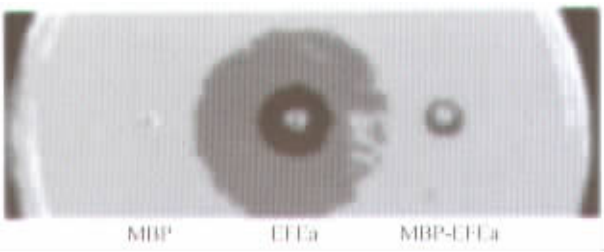

Fig. 5. Analysis of the fibrinolytic activities. MBP, Maltose-binding protein; EFEa; earthworm fibrinolytic enzyme component A; MBP-EFEa, recombinant MBP-EFEa.

3 Discussion

In this note, we described the cloning of the cDNA encoding EFEa from earthworm *Eisenia fetida* by RT-PCR. The sequence of the cDNA showed highly identities with that encoding an EFE component from *Lumbricus bimastu* (GenBank accession number: AF109648). The component F-Ⅱ isolated from *Lumbricus rubellus* showed the same N-terminal 24 amino acid residues with EFEa from *Eisenia fetida*. This shows that the three sequences should belong to the same gene, which is conserved in different families of earthworm.

The genes encoding F-Ⅰ-1 and F-Ⅰ-2 were cloned, and the deduced primary sequences of them showed high similarities with some serine proteases. The active sites (the catalytic triad and primary substrate specificity determinant) were conserved in them. These results showed that F-Ⅰ-1 and F-Ⅰ-2 were trypsin proteases, which agreed with their biochemical characteristics[4,5,8]. Although the deduced amino acid sequence of EFEa also showed high similarities with serine protease and the catalytic triad was conserved, the difference in primary substrate specificity determinant between EFEa and trypsin implied that the substrate specificity of them might be different. The primary crystallographic results of EFEa from *Eisenia fetida* showed that it adopted the fold of chymotrypsin-like serine proteases, and had essential S1 specificity determinants characteristic of elastase[11]. It proved our deductions from the analysis of the amino acid sequences mentioned above.

The expression of heterologous genes in *E. coli* is by far the simpliest and most inexpensive method available for obtaining large amounts of desired proteins. Unfortunately, the recombinant polypeptides were found accumulated in the inclusion bodies in many cases. The expression of the recombinant protein as a fusion with a highly soluble polypeptide can facilitate the solubility of the target protein in the cell lysate[12]. The soluble recombinant EFEa was expressed as a fusion with a highly soluble polypeptide maltose binding protein. More and more heterologous genes will be expressed using this method to obtain soluble recombinant proteins.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 39970174) and the Foundation of the National Key Research Development Project of China (Grant No. G1999075601).

References

1. Xiong, Y., Yang, S., Liu, X. et al., Separation, purification and partial sequence analysis of the earthworm fibrinolytic enzymes, Chinese Biochemical Journal, 1997, 13(3): 292.
2. Yang, J., Ru, B., Purification and characterization of an SDS-ac- tivated fibrinolytic enzyme from *Eisenia fetida*, Comp. Biochem. Physiol. B1, 1997, 18(3): 623.
3. Mihara, H., Sumi, H., Yoneta, T. et al., A novel fibrinolytic enzyme extracted from the earthworm *Lumbricus rubellus*, Japn. J. Physiol., 1991, 41: 461.
4. Nakajima, N., Mihara, H., Sumi, H., Characterization of potent fibrinolytic enzymes in earthworm, *Lumbricus rubellus*, Biosci. Biotech. Biochem., 1993, 57: 1726.
5. Nakajima, N., Ishihara, K., Sugimoto, M. et al., Chemical modification of earthworm fibrinolytic enzyme with human serum albumin fragment and characterization of the protease as a therapeutic enzyme, Biosci. Biotechnol. Biochem., 1996, 60(2): 293.
6. Nakajima, N., Sugimoto, M., Ishihara, K. et al., Further characterization of earthworm serine proteases: Cleavage specificity against peptide substrates and on autolysis. Biosci. Biotechnol. Biochem., 1999, 63(11): 2031.
7. Park, Y., Kim, J., Min, B. et al., Rapid purification and biochemical characteristics of lumbrokinase F-III from earthworm for use as a fibrinolytic agent, Biotechnol. Lett., 1998, 20(2): 169.
8. Sugimoto, M., Nakajima, N., Molecular cloning, sequencing, and characterization of cDNAs encoding fibrinolytic enzymes from earthworm, *Lumbricus rubellus*, Biosci. Biotechnol. Biochem., 2001, 65: 1575.
9. Tang, Y., Zhang, J., Gui, L. et al., Crystallization and preliminary X-ray analysis of earthworm fibrinolytic enzyme component A from *Eisenia fetida*, Acta Cryst D, 2000, 56: 1659.
10. Sambrook, J., Rusell, D., Molecular Cloning, 3rd ed., New York: Cold Spring Harbor Laboratory Press, 2001.
11. Tang, Y., Liang, D. C., Jiang, T. et al., Crystal structure of earthworm fibrinolytic enzyme component A: Revealing the structural determinants of its dual fibrinolytic activity, J. Mol. Bio., 2002, 321: 57.
12. Georgiou, G., Valax, P., Expression of correctly folded proteins in *Escherichia coli*, Cur. Opin. Biotech., 1996, 7: 190.

(Received September 20, 2002)