

# 蚯蚓纤溶酶新基因 *EfP-0* 的电子克隆及其编码区序列 RT-PCR 验证 In Silico Cloning of *EfP-0* , a Novel Earthworm Fibrinolytic Enzyme Gene and Verification of its Coding Region by RT-PCR

赵晓瑜\* ,高 珊 ,崔大昶 ,耿风廷

ZHAO Xiao-Yu\* , GAO Shan , CUI Da-Ling and GENG Feng-Ting

河北大学生命科学学院 ,保定 071002

College of Life Sciences , Hebei University , Baoding 071002 , China

**摘 要** 利用生物信息学手段,以期获得蚯蚓纤溶酶 F-I-0 组分的基因。根据从粉正蚓(*Lumbricus rubellus*)中分离的 F-I-0 组分的 N 端氨基酸序列 VVGSDTTIGQYPHQL 利用 DNAMAN 软件通过电子克隆方法,从 *Lumbricidae* 的 dbEST 中获得该组分的核酸序列信息,设计特异引物,经过 RT-PCR,成功地从赤子爱胜蚓(*Eisenia foetida*)中克隆到一条蚯蚓纤溶酶新基因,命名为 *EfP-0*。*EfP-0* 基因全长 678bp,编码 225 个氨基酸的成熟肽,属丝氨酸蛋白酶,胰蛋白酶家族,与 F-I-0 组分的氨基酸组成非常接近。BLAST 证明,*EfP-0* 与已报道的蚯蚓纤溶酶基因之间的相似性均低于 40%,因此为蚯蚓纤溶酶中的一个新基因,GenBank 登录号为 DQ836917。构建的 pMAL-c2x-*EfP-0* 重组质粒,在大肠杆菌 TB1 中获得融合蛋白 MBP-*EfP-0* 的可溶性表达,表达产物有酪蛋白平板溶解活性。

**关键词** 蚯蚓纤溶酶,电子克隆,*EfP-0*

中图分类号 Q781.0811.4 文献标识码 A 文章编号 1000-3061(2006)06-0897-05

**Abstract** There are four different types of N-terminal amino acid sequences (F-I-0, F-I, F-II, F-III) in the multi-components of earthworm fibrinolytic enzymes (EFE). In GenBank 21 nucleic acid sequences of EFE have been reported. Among them, most of the N-terminal amino acid sequences belong to the F-III type, few belong to the F-II type. Only one is similar to the F-I type, but none to F-I-0. In this research we hoped to obtain the gene encoding component F-I-0 of EFE by the bioinformatics tools. Based on the N-terminal amino acid sequence VVGSDTTIGQYPHQL of the F-I-0 type from *Lumbricus rubellus*, a nucleic acid sequence was obtained by in silico cloning from dbEST of *Lumbricidae* using the software DNAMAN. A new gene of EFE from *Eisenia foetida* was successfully obtained by RT-PCR using specific primers designed according to this sequence. The new gene named *EfP-0* was cloned in pMAL-c2x and expressed as the fusion protein MBP-*EfP-0* in the supernatant of lysate. The fusion protein MBP-*EfP-0* purified by affinity chromatography had hydrolytic activity on casein plate. Sequencing result shows, *EfP-0* has 678bp and encodes a protein of 225 amino acids. The protein is a *serine protease belonging to trypsin family*. It has similar amino acid composition to F-I-0. BLAST in GenBank shows that the similarity is lower than 40% between *EfP-0* gene and other EFE genes. By this we conclude that *EfP-0* gene of EFE is a novel gene and it is the first time to be reported, its accession number for Genbank is DQ836917.

**Key words** earthworm fibrinolytic enzymes, in silico cloning, *EfP-0*

Received: July 26, 2006; Accepted: September 4, 2006.

This work was supported by the Key Subject of Biotechnology of Hebei Province.

\* Corresponding author. Tel: 86-312-5079623, E-mail: yuoaix@yahoo.com.cn

河北省生物工程重点学科资助。

© 中国科学院微生物研究所期刊联合编辑部 <http://journals.im.ac.cn>

电子克隆<sup>[1]</sup>(*in silico cloning*),是指利用计算机技术,依托现有的网络资源(EST数据库、核苷酸数据库、蛋白质数据库、基因组数据库等)采用生物信息学方法(包括同源性检索、聚类、序列拼接等)延伸EST序列,以期获得部分乃至全长cDNA序列的一种方法<sup>[2-3]</sup>。近几年来,EST数据库发展非常迅速,随着人类基因组测序及生物信息学的快速发展,产生了大量可供利用的基因组信息和分析软件。充分利用这些数据和工具来寻找新基因,了解基因的功能,是功能基因组学研究的主要任务,也是目前基因克隆的主要策略之一<sup>[1,2,4]</sup>。

蚯蚓纤溶酶(*earthworm fibrinolytic enzymes, EFE*)是一类在蚯蚓体内分离到的具有纤溶蛋白作用和激活纤溶酶原作用的蛋白质,具有抗凝、纤溶和溶栓作用<sup>[5]</sup>,可以作为溶栓药物治疗因血栓引起的各种疾病。在其分子生物学研究方面,国内外学者分别在GenBank中递交了21条核酸序列和28条蛋白质序列(21条通过核酸序列翻译所得,7条是直接报道的部分蛋白质序列)。根据N-末端氨基酸序列比对,这些序列与Mihara H报道<sup>[6]</sup>的6个组分中的5个组分一致或相似性极高,但与F-I-0组分序列一致的核酸序列至今尚未见报道。本研究利用现有的*Lumbricidae*的dbEST,借助生物信息学中的电子克隆技术,克隆得到了该基因。

## 1 材料与方法

### 1.1 材料

蚯蚓(*Eisenia foetida*);Trizol试剂购自赛百盛公司;cDNA合成试剂盒、TaqDNA聚合酶、DNA回收试剂盒、DL3000 DNA Marker、pMD19-T Vector购自TaKaRa公司;*E. coli* TB1、pMAL-c2x购自NEB公司。

### 1.2 方法

**1.2.1 验证 dbEST(*Lumbricidae*)的完整性:**利用dbEST(*Lumbricidae*, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=nucleotide>)对多条已知序列进行电子克隆来验证它的完整性。

**1.2.2 F-I-0基因的电子克隆:**由于F-I-0没有源EST作为种子序列,所以直接利用其N端氨基酸序列,在NCBI的BLAST工具中选择tblastn,取E值e-06以上的序列用DNAMAN拼接,将结果作为新的种子序列,采用blastn方式比对, DNAMAN拼接,如此重复至第四次,contigs不再延长,为结果序列,进行后续分析。

**1.2.3 RT-PCR的验证:**用Trizol试剂提取新鲜蚯蚓的总RNA,反转录合成cDNA第一链,根据已经电子克隆获得的序列,在序列CDS区域两端(678bp)设计引物,正向引物5'-GTCGTCGGTGGAAGTGATACC-3',反向引物5'-TTATTG(GC)GC(AG)ATGATCCAGGT-3',PCR扩增获得目的片段。回收PCR产物克隆pMD19-T Vector,由北京华大基因研究中心测序。

**1.2.4 表达质粒的构建和表达产物的鉴定:**将PCR产物克隆pMAL-c2x载体, IPTG诱导表达,亲和层析纯化,酪蛋白平板检测表达蛋白活性<sup>[7]</sup>。

## 2 结果

### 2.1 对 dbEST(*Lumbricidae*)完整性的验证

*Lumbricidae* dbEST(18330条核酸序列信息)中17222条是从*Lumbricus rubellus*来的,能否用于电子克隆需要用已知序列进行验证。分别取Genbank登录号为AY438624(*Eisenia foetida*)、AY684711(*Eisenia foetida*)和LRAJ3152(*Lumbricus rubellus*)三个蚯蚓纤溶酶基因的一部分作为种子序列进行电子克隆,克隆结果与原序列进行比对,相似性分别为87.27%、86.12%和93.53%。说明本研究利用的dbEST能够在软件拼接后得到与原完整序列相似性很高的序列,这一方面证明了dbEST中的数据是比较完整的,另一方面也说明蚯蚓纤溶酶在种属间的差异很小,同时也验证了本研究所采用的方法和软件是可信的。

### 2.2 F-I-0基因的电子克隆

根据F-I-0组分的蛋白质N-末端序列VVGGSDDTTIGQYPHQL作为拼接前的种子序列,用DNAMAN软件拼接获得序列(图1)。为了验证不同软件的拼接效果,还采用了seqmanII方法进行拼接,两种软件得到的结果相似性达到91.02%。

经由DNAMAN软件分析,最长的阅读框在38-778位之间,包含246个氨基酸残基,第38位碱基为起始密码子,推测38-100位碱基为信号肽,在成熟蛋白中被剪切,第101位碱基起始编码的蛋白,同F-I-0N端氨基酸残基一致,第778位碱基为终止密码子,915位碱基以后为Poly A尾巴,将阅读框架翻译成氨基酸序列,得到了长度为225个氨基酸的成熟肽。

### 2.3 *Efp-0*编码基因的获得

以蚯蚓(*Eisenia foetida*)总RNA为模板,根据2.2拼接的阅读框设计一对特异引物进行RT-PCR,结果

1 GCGGATTCCG CACGAGCGCA AAGGTTTCAA TCGTAACATG AGGATCCTGA TCTTGCTGTG  
61 TTTGGCTCTG CCCGCTCTGG GTTCCGGTGT GGATCAGCAT GTCGTCGGTG GAACTGATAC  
121 CACCATCGGC CAATACCGCG ATCAGCTGTC GTTGGCAGTG ACTGGATCTC ACTCATGCGG  
181 AGCATCTCTG ATTGGAACCA CTCGAGCTGT GACCGCGGCT CACTGCACAG GATCTGCCAT  
241 CGGCGTCTAC TCTATTCTTG GCGTAACCAC TGATCGTACT ACTACAACT GCGCCACTTG  
301 TGTACTCAGA GATCTCAACT TCCTAAAAAG GCATCCTAAT TACGATGGAA ACGGCGCGGG  
361 ATATCCGAAT GACGTAGCCG TTATTGGATT CGCTGCTGTG TCGACCAACA CCAACCTCCA  
421 GACAATTGCC CTGGCCACTC CATCGGACGG AAAC TTGCT GGTGACAGCT GTGTATCAC  
481 CGGATGGGGC CAAACTGGAT CCATCGGTGG TCTTCCAGAC ACCTTGCAGC TGGCTACAAT  
541 GAGTGTCTTG ACCAACACGG CATGTGCCG CATCTGGGGT GCAGTTCGCA TCAACGACGG  
601 ACACATCTGC GTCCATGCAG CAGGAAGAAG CGCCTGCAGT GCGGACAGCG GTGGCCCACT  
661 GGAGTGCAGT AACAACTGG CTGGAGCGAC ATCTTGGGGT GAGGCTCGT GTGACCCATC  
721 CTACCGCTCC GTCTACACAA GAGTCAGCTA CTTCTACACC TGGATCATCG CCCAATAAGC  
781 GACGCAATGA CTGGACCACG GATTCACTTC TCATGATTTC TGCATCCGAT TAAATAACAA  
841 GAAATAAAAA ACTGTTATCT TATATAAATT TCAGCTAAAC GATAATCGCT TAAATAAACA  
901 ATTTGCAACC TAAAAAAAAA AA

图 1 电子克隆( DANMAN )F- I -0 组分的序列

Fig.1 Sequence of F- I -0 in silico cloning by DNAMAN

得到了 680bp 左右的片段( 图 2 ),与预期的片段大小相符合。将此片段克隆至 pMD19-T 载体中筛选阳性克隆,进行测序,获得序列图谱,与电子克隆预测序列比对,相似性达 90.41%,将所得到的基因命名为 *EfP-0*。GenBank 登录号为 DQ836917。

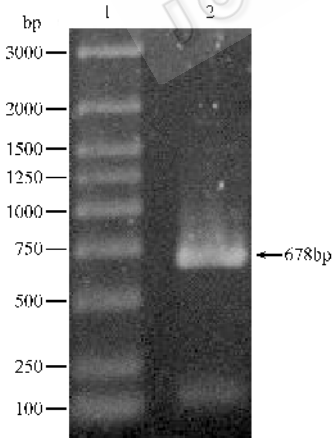


图 2 RT-PCR 扩增产物的琼脂糖凝胶电泳

Fig.2 Agarose gel electrophoresis of the product by RT-PCR

1 :DL3000 DNA Marker 2 :RT-PCR product.

2.4 *EfP-0* 的生物信息学预测分析

*EfP-0* 基因编码 225 个氨基酸的成熟肽,即 *EfP-0*。根据电子克隆推测的成熟肽序列,以及从 *Eisenia fetida* 中获得的 *EfP-0* 序列计算每个分子中各种氨基酸的含量,与 Mihara H 分离的天然 F- I -0

组分的氨基酸色谱数据<sup>[8]</sup>进行比较( 表 1 ),*EfP-0* 的氨基酸组成与 F- I -0 更为接近。

表 1 F- I -0 的电子克隆推测的成熟肽、*EfP-0* 与天然 F- I -0 的氨基酸组成的比较

Table 1 Comparison of amino acids composition of speculate peptide、*EfP-0* and F- I -0

Amino acid	Speculate peptide	<i>EfP-0</i>	F-I-0 <sup>[8]</sup>
Ala	25	27	31
Cys + halfcys	10	10	3
Asp + Asn	22	25	24
Glu + Gln	8	8	9
Phe	4	4	4
Gly	30	27	29
His	6	5	6
Ile	13	12	15
Lys	2	0	0
Leu	15	15	16
MET	1	1	1
Pro	7	9	11
Arg	8	9	9
Ser	21	21	20
Thr	26	24	25
Val	15	16	16
Trp	4	4	4
Tyr	8	8	8

用 DNAMAN 预测 *EfP-0* 的分子量为 22989D,等电点 pI 为 5.45,这与日本 Nakajima<sup>[8]</sup>报道的 F- I -0 质谱分子量 23013D,等电点 4.85 相近,也与本实验室生化分离纯化的 *EfP-0-1* 和 *EfP-0-2* 组分的质谱

分析结果接近 23008D 和 22967D。

经过蛋白质亚细胞定位分析( <http://www.bioinfo.tsinghua.edu.cn/SubLoc/eu-predict.htm> ),发现该蛋白质为细胞外蛋白,与蚯蚓纤溶酶的分泌特性一致。

*EjP-0* 属于丝氨酸蛋白酶中的胰蛋白酶家族 ( His-41, Asp-91, Ser-183 ) ,具有组氨酸活性位点( 37-42 )和丝氨酸活性位点( 177-188 )。除此之外, *EjP-0* 还具有蛋白激酶 C 磷酸化位点( 17-19、33-35、57-59、190-192 ) 酪蛋白激酶 II 磷酸化位点( 113-116、151-154、198-201 ) 肉豆蔻酰化位点( 3-8、4-9、22-27、32-37、44-49、83-88、117-122、130-135、133-138、137-142 )。

对 *EjP-0* 进行糖基化位点分析,序列中没有发

现 N-连接糖基化位点的特征氨基酸 Asn-Xaa-Ser/Thr ( N-x-S/T ) ,但是有潜在的 O-连接糖基化位点( 7、8、33、34、197 ) ,而本实验室生化分离纯化的 *EjP-0-1* 和 *EjP-0-2* 组分为含有少量糖链的糖蛋白。

2.5 数据库 BLAST 比对证明

将 *EjP-0* 基因的核酸序列在 NCBI 的 BLAST 数据库中进行比对,结果显示 *EjP-0* 基因与其他已报道的蚯蚓纤溶酶基因均无明显同源性( 图 3 )。 *EjP-0* 的氨基酸序列的比对结果显示, *EjP-0* 与其他蚯蚓纤溶酶相似性均低于 40% ,仅活性位点部位的序列有较高的同源性。因此证明 *EjP-0* 基因为蚯蚓纤溶酶的一个新基因。

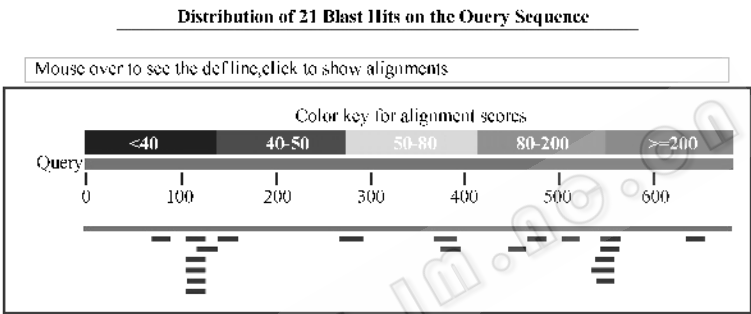


图 3 *EjP-0* 在 NCBI 中的 BLASTn 结果

Fig. 3 BLASTn of *EjP-0* in NCBI

2.6 *EjP-0* 基因在大肠杆菌中的表达及活性测定

将 *EjP-0* 基因克隆入 pMAL-c2x 载体构成 pMAL-c2x- *EjP-0* 重组质粒,其中 *EjP-0* 基因的上游为 MBP 基因,它与 *EjP-0* 基因以融合形式表达。

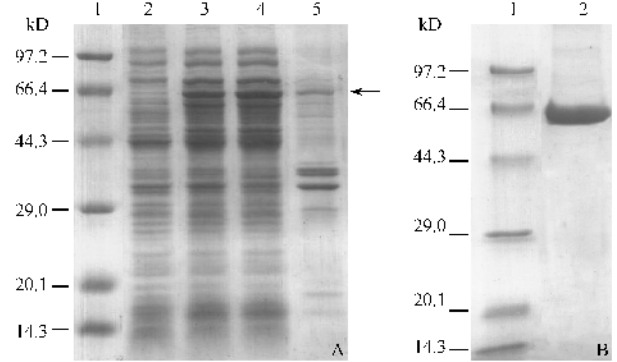


图 4 SDS-PAGE 检测 MBP-*EjP-0* 在大肠杆菌中的表达和纯化

Fig.4 SDS-PAGE analysis of MBP-*EjP-0* expression and purification in *E. coli*

A : 1 :protein molecular weight marker ;2, 3 :TB1/pMAL-c2x- *EjP-0* induced for 0h 4h ;4 :The supernatant of lysate of TB1/pMAL-c2x- *EjP-0* ;5 :The deposit of TB1/pMAL-c2x- *EjP-0*. B : 1 :protein molecular weight marker 2 :MBP-*EjP-0* purified by affinity chromatography.

MBP 大小为 42kD,如果融合 23kD 的 *EjP-0*,表达蛋白大小应为 65kD 左右。将 pMAL-c2x- *EjP-0* 重组质粒转化 *E. coli* TB1,随后进行 IPTG 诱导,表达产物经 SDS-PAGE 检测,显示有 65kD 左右的表达条带,破菌处理后发现该融合蛋白 MBP- *EjP-0* 为可溶性蛋白( 图 4 ,A )。经亲和层析纯化得到单一蛋白条带( 图 4 ,B ),用纯化后的蛋白进行酪蛋白平板活性检测有溶圈( 图 5 ),说明该蛋白具有纤溶酶活性。

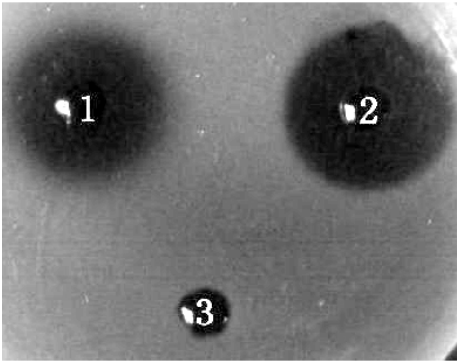


图 5 酪蛋白平板检测 MBP-*EjP-0* 活性

Fig.5 The analysis of MBP-*EjP-0* activity on casein plate

1 MBP-*EjP-0* 2 positive control( EFE );3 Negative control( MBP ).

### 3 讨论

Mihara H 首次从粉正蚓 (*Lumbricus rubellus*) 中发现蚯蚓纤溶酶<sup>[6]</sup>, 并将分离到的 6 个组分分为 3 个大组: F-I、F-II、F-III, F-I 又分为 F-I-0、1、2, F-III 分为 F-III-1 和 F-III-2<sup>[9]</sup>。到目前为止一直没有对 F-I-0 组分基因研究的报道。F-I-0 组分又与赵晓瑜等<sup>[10]</sup>分离的赤子爱胜蚓 (*Eisenia foetida*) 纤溶酶组分 I (*EfP-0-1*) 和 II (*EfP-0-2*) 具有相同的 N 端氨基酸序列和质谱分子量, 而且均属于胰凝乳蛋白酶(待发表)。本文根据从粉正蚓 (*Lumbricus rubellus*) 中分离的 F-I-0 组分的 N 端氨基酸序列, 通过电子克隆方法获得序列信息, 并成功的从赤子爱胜蚓 (*Eisenia foetida*) 中克隆到蚯蚓纤溶酶 *EfP-0* 基因( GenBank 登录号为 DQ836917 ), 说明蚯蚓纤溶酶在种间差异很小, 尤其是 5' 和 3' 序列具有很高的保守性。BLAST 对已知序列比对的结果, 相似性均低于 40%, 而且将该基因构建的表达载体在大肠杆菌中表达的产物具有酪蛋白水解活性, 充分说明该序列为首次报道蚯蚓纤溶酶的一个新基因。

选用生物信息学常用软件 DNAMAN 和 SeqmanII 能够得到相似度很高的序列, 该序列由 50 条左右 EST 序列支持, 包含 225 个氨基酸的开放性阅读框, 分子量 22989D, 经由氨基酸色谱分析各残基成分与预测基本一致, 只有 Cys 和 Ala 有差异(这个差异或许是由软件算法决定, 因为用 DNAMAN 分析其它已知序列, 都与色谱分析的相应组分的 Cys 有较大差异, 以后可以通过算法改进来解决该问题)。

原则上, 利用 N-末端序列设计简并引物和 oligo dT 进行 RT-PCR 也可以获得 cDNA, 但实际上可能由于简并度过高而影响了扩增的特异性。本文曾扩增到一个大小类似的序列, 然而测序的结果为序列两侧反向互补, 是一条未知序列。

电子克隆是一种获得新基因的方法, 随着 dbEST 数据库的不断增加, 为了解和获得更多基因提供了一条有效途径。本文利用该方法从已有的数据库中得到了蚯蚓纤溶酶组分的新基因序列的相关信息, 指导了新基因的顺利获得。

### REFERENCES(参考文献)

[ 1 ] Peng YD( 彭永德 ), Li YB( 李月彬 ), Song HD( 宋怀东 ). Novel full length cDNA cloning from normal adrenal gland and pheochromocytoma and functional prediction. *National Medical Journal of China*( 中华医学杂志 ) 2002, **82**( 1 ): 31 - 34

[ 2 ] Peterson DG, Schulze SR, Sciara EB *et al.* Integration of Cot analysis, DNA cloning, and high through put sequencing facilitates genome characterization and gene discovery. *Genome Res* 2002, **12**: 795 - 807

[ 3 ] Lander ES, Linton L M, Birren B *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001, **409**: 860 - 921

[ 4 ] Huminiecki L, Bicknell R. In silicon cloning of novel endothelial-specific genes. *Genome Res* 2000, **10**: 1796 - 1806

[ 5 ] Mihara H, Sumi H, Akazawa T *et al.* Fibrinolytic enzyme extracted from the earthworm. *Thromb Haemostas* 1983, **50**: 258 - 263

[ 6 ] Mihara H, Sumi H, Matsuura A *et al.* A thrombolytic agent. *European patent application*, EP0 105 092 A<sub>2</sub>, 1983-06-28

[ 7 ] Wu JX( 武金霞 ), Zhao XY( 赵晓瑜 ). Casein-Plate Method for Determination of Earthworm Fibrinolytic Enzyme. *Nature Magazine* ( 自然杂志 ) 2004, **26**( 3 ): 184 - 185

[ 8 ] Nakajima N, Mihara H, Sumi H. Characterization of Potent Fibrinolytic Enzymes in Earthworm, *Lumbricus rubellus*. *Biosci. Biotech. Biochem* 1993, **57**( 10 ): 1726 - 1730

[ 9 ] Mihara H, Sumi H, Yoneta T *et al.* A novel fibrinolytic enzyme extracted from the earthworm, *Lumbricus rubellus*. *Japanese J Physiol* 1991, **41**: 461 - 472

[ 10 ] Zhao XY( 赵晓瑜 ), Jing TY( 静天玉 ) *et al.* Antigenicity of Components Constituting Earthworm Fibrinolytic Enzyme. *Chinese Journal of Biochemistry and Molecular Biology*( 中国生物化学与分子生物学报 ) 2002, **18**( 2 ): 209 - 212