

日本血吸虫真核生物翻译起始因子 2 琢亚基全长 cDNA 的克隆与功能分析

卢晓昭 袁鸿娟 袁陈晓光 第一军医大学寄生虫学教研室 广东 广州 510515 袁

摘要 目的 对用表达序列标签 (EST) 策略从日本血吸虫 *Schistosoma japonicum* 尾蚴 cDNA 文库中筛选出的新基因进行全长 cDNA 克隆及功能预测。方法 将插入于 pTriplEx2 质粒上的 cDNA 进行测序。测序结果经 BLASTn 程序搜索。发现该插入 cDNA 序列与曼氏血吸虫真核生物翻译起始因子 2 琢亚基 (eukaryotic translation initiation factor 2 alpha subunit, eIF2 琢) 基因高度同源。根据该 EST 序列设计与 pTriplEx2 质粒上的 5' 端测序引物相匹配的 PCR 下游引物。从该 cDNA 文库中扩出包含 eIF2 琢全长 ORF 的 cDNA 片段。将 PCR 产物纯化后克隆到 pGEM-T 载体上, 并对此克隆进行测序得到其全长 cDNA 序列。用 NCBI 站点的 BLASTx 及 BLASTn 程序对所获得的新基因序列进行同源性搜索。用 blast two sequence 程序对同源性高的基因进行核苷酸及氨基酸水平的同源性比较。同时用网上分析软件进行基序和保守区域的搜索。结果 发现一个与曼氏血吸虫 eIF2 琢 mRNA 高度同源的日本血吸虫新基因。核苷酸与氨基酸水平的同一性分别为 87% 和 79%。编码由 327 个氨基酸组成的蛋白序列。结论 用 EST 策略筛选到一个日本血吸虫新基因。编码 eIF2 琢亚基全长编码序列与曼氏血吸虫 eIF2 琢 mRNA 高度同源。

关键词 血吸虫 日本血吸虫 曼氏血吸虫 表达序列标签 基因克隆 真核类起始因子 2

中图分类号 R785;R383.24 文献标识码 文章编号 000-2588(2003)04-0296-04

Cloning and function analysis of full-length cDNA sequence of *Schistosoma japonicum* eukaryotic translation initiation factor 2 alpha subunit

LUXiao-zhao, PENG Hong-juan, CHEN Xiao-guang

Department of Parasitology, First Military Medical University, Guangzhou 510515, China

Abstract: Objective To subclone the novel genes screened from the *Schistosoma japonicum* cercariae cDNA library through expressed sequence tag (EST) strategy and analyze its functions. Method The cDNA fragment inserted in pTriplEx2 vector was sequenced and the result retrieved with BLASTn program. It was found that this cDNA was highly homologous to *Schistosoma mansoni* eukaryotic translation initiation factor 2 alpha subunit (eIF2 琢) mRNA. According to the known EST sequence, the 3-terminal primer that matched the sequencing primer for the 5-terminal in pTriplEx2 plasmid was designed and used to amplify the full-length open reading frame (ORF) sequence of the eIF2 琢 from the cDNA library. After proper purification, the PCR product was linked to pGEM-T vector and the recombinant T-vector was sequenced to obtain the full length ORF, which was retrieved for homologous identification using NCBI blast program. These sequences that were highly homologous underwent comparison at the level of amino acids and nucleotides using BLAST2 Sequence program on NCBI BLAST site. The motif and conserved domain were also retrieved with the software available online. Result A novel cDNA sequence coding for a eIF2 琢 was found from the cDNA library of *Schistosoma japonicum* cercariae, which was highly homologous to the known *Schistosoma mansoni* eIF2 琢 mRNA, with the homology of 87% at the nucleotide level and 79% at the amino acid level. Conclusion The novel gene found by EST strategy may encode a eIF2 琢 which is highly homologous to *Schistosoma mansoni* eukaryotic eIF2 琢 mRNA.

Key words: *Schistosoma japonicum*; *Schistosoma mansoni*; expression sequence tag; gene clone; eukaryotic initiation factor 2

日本血吸虫病仍然在亚洲一些国家流行。特别是在中国的一些地区及菲律宾群岛。截至 2001 年底, 中国尚有 12 个省的 418 个血吸虫病流行县。流行区总

收稿日期 2003-01-30

基金项目 院联合国发展计划署 / 世界银行 / 世界卫生组织热带病研究和培训特别规划署基金 00690 袁 00191 袁

Sponsored by United Nations Development Programme/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases (A00690, A00191)

作者简介 卢晓昭 1981 袁 男 袁 云南陆良人 袁 第一军医大学临床医学本科学历

通讯作者 陈晓光 电话: 020-61648308

人口约 99 万 袁 患病人数居高不下。随着人类基因组计划的带动下, 1992 年美国基因组研究所 (Institute for Genomic Research, TIGR, USA) 发起了全球性的血吸虫基因组计划 (Schistosoma Genome Project, SGP)。袁 袁 该计划从 1994 年起受到 野 NDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) 基金的支持。袁 袁 血吸虫基因发现计划 是 SGP 的一部分, 该计划的主要目标之一是发现与鉴定日本血吸虫与曼氏血吸虫的新基因。袁 袁 而寻找新的治疗药物与疫苗。袁 袁 表达序列标签 (expression sequence tag, EST) 是文库中随机挑选的

cDNA末端的短序列约300碱基。通过EST可以利用数据库搜索进行DNA或蛋白质的同源性分析。而鉴定这些基因的起源。我们在国内外首次构建了日本血吸虫尾蚴cDNA文库。在用EST策略从文库中筛选到的201个EST序列中。有136个合乎要求的序列进行了同源性分析。在这136个EST中。有61.8%的序列与数据库中的蛋白或DNA序列无同源性。有38.2%的EST序列与已知基因或已知蛋白具有同源性。其中有5.9%及6.6%的EST分别与日本血吸虫及曼氏血吸虫已知基因高度同源。而有25.7%的EST与其它生物的相应基因或蛋白高度同源。我们对一个与曼氏血吸虫真核生物翻译起始因子2 α 亚基ukaryotic translation initiation factor 2 α subunit, eIF2 α mRNA序列高度同源的cDNA片段进行了研究。克隆了其全长开放读码框序列。

1 材料

1.1 文库

日本血吸虫中国大陆株尾蚴cDNA文库由本室用CLONTECH的SMARTTMcDNA文库构建试剂盒构建。一级文库库容量为1.8伊0⁷ pfu。

1.2 测序引物及测序

选出文库载体插入位点上游的序列设计测序引物。5'-CTCCGAGATCTGGACGAGC-3'。测序由上海基康公司完成。ERKINELMER公司全自动核酸分析仪BIPRISMTM310 Genetic Analysis测序。

1.3 质粒

pGEM-Tvector 购自PROMEGA公司。

1.4 工具酶和试剂

Taq E, NTP, coR, ho, 玉购自广州基因公司。DNA marker: GO2S100bp ladder 购自上海博彩科技有限公司。

2 方法

2.1 噬菌体转变成质粒

按Clontech SMARTTMcDNA文库构建试剂盒操作说明进行。

2.2 质粒的测序和EST的获取

碱裂解法提取质粒。用5'端测序引物测出一个反应的序列。将所得到的原始序列进行分析。编辑。除去质粒序列。留下插入的DNA片段。经过编辑后的序列称为一个EST。5'端不具有接头序列及长度小于150bp的EST都被舍弃。

2.3 EST分析

通过NCBI数据库。http://www.ncbi.nlm.nih.gov/blast。BLASTn, BLASTx程序分别在核苷酸及蛋

白质水平上对每一个EST进行同源性搜索。

2.4 全长cDNA序列的获取及功能分析

2.4.1 PCR扩增 以测序引物为上游引物: 5'-CTCCGAGATCTGGACGAGC-3'。并根据EST序列设计与之相匹配的下游引物: 5'-GCCTCGAGTCATATAATATATTTACTT-3'。以尾蚴cDNA文库为模板进行PCR扩增。琼脂糖凝胶电泳分析PCR结果。

2.4.2 目的基因的T载体连接及测序 PCR产物回收纯化后接入pGEM-T载体, 转化大肠杆菌DH5 α 。并筛选出阳性克隆。质粒送往基康公司进行测序。

2.4.3 对全长cDNA序列进行结构分析及功能预测。通过NCBI数据库。http://www.ncbi.nlm.nih.gov/blast。BLASTn, BLASTx程序分别在核苷酸及蛋白质水平上对每一个EST进行同源性搜索。利用NCBI BLAST站点的blast two sequence程序。http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html。对同源性高的基因进行核苷酸及氨基酸水平的同源性比较。利用Motif Scan in a Protein Sequence。http://hits.isb-sib.ch/cgi-bin/hits_motifscan。对cDNA序列所编码的蛋白质进行结构域搜索。利用CD-Search。http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi。程序对目标蛋白进行保守区域的搜索。

3 结果

3.1 EST Blast 结果

在所得到的EST序列中。CXP0006.FXS克隆的cDNA序列。片段长543 bp。EST登录号为gi|1147534。与曼氏血吸虫eIF2 α 的mRNA高度同源。

3.2 日本血吸虫eIF2 α 全长cDNA序列

日本血吸虫eIF2 α 全长cDNA序列与曼氏血吸虫eIF2 α 的mRNA。GenBank登录号为AF376135_1。高度同源。该序列包含一个完整的编码区。由起始密码子ATG至终止密码子TAA止。共有981 bp。编码由327个氨基酸组成的蛋白。

3.3 日本血吸虫eIF2 α 与曼氏血吸虫eIF2 α 基因的同源性比较

二者同源性为87%。具体见图2。

3.4 日本血吸虫与曼氏血吸虫eIF2 α 蛋白水平的同源性比较

用NCBI BLAST站点的BLASTx程序对Sj-eIF2 α cDNA序列进行搜索。发现其与Sm-eIF2 α 氨基酸水平的同源性达79%。

3.5 日本血吸虫eIF2 α 蛋白结构域搜索结果

利用Motif Scan in a Protein Sequence对cDNA序列所编码的蛋白质进行结构域搜索。结果表明。在

上述 sj-eIF2 琢氨基酸序列中从第 17 位 Asp 到第 86 位 Lys 是由 70 个氨基酸组成的 S1 结构域渊 domain 冤遥 利用 CD-Search 程序对目标蛋白进行保守区域的搜索袁发现保守区域遥

4 讨论

eIF2 在蛋白质生物合成的起始中扮演着重要角色袁是蛋白质合成不可缺少的起始因子遥 eIF2 是一个异构三聚体袁由 琢 茁 酌 三个亚基构成遥 Nonato 等^[10]报道了人类 eIF2 琢的部分结构袁该部分由 eIF2 琢的 3~182 位氨基酸残基组成袁占人类 eIF2 琢全长 N-末端的 2/3 遥其结构主要由两个部分组成袁一部分为寡聚核

Score=1131bits(588),Expect=0.0, Identities=850/976(87%), Gaps=6/976(0%)
 Sm:1 atgccatacagtcagatattctacgaggacttatccctgaggctggggatgtgctc60
 Sj:1 atgccatacagtcagatattctacgaggacttatccctgaggctggggatgtgtagt60
 Sm:61 gtgacggttaagtcacccatcaggtagctacggtggaacttctggagtacaagaac120
 Sj:61 gtaacaggttaagtcacccatcaggtagctacggtggaacttctggagtacaagaat120
 Sm:121atggaggatgatattacatagtgagctatccgctcgacgcaataagatctatcagcaag180
 Sj:121atggaggatgatattgctcagtgagctctcgcgacgataataagatcaatcagcaag180
 Sm:181cttgttgcgaatcgggtcaaaactgaggtactggtgtagcagtcagagtgccaagg240
 Sj:181cttgttgcgaatcgggtcaaaactgaggtactggtgtagcagtcagagtgccaagg240
 Sm:241tatattgatttgtaaaagaagagcctctgcagaagaaatagcgaagtgaagaaaga300
 Sj:241tatattgatttgtaaaagaagagcctctgcagaagaaatagcgaagtgaagaaaga300

```

1 ATGCCATACAGTCAGATATTATGAGGACTTATCCCTGAAGTTGGGGATGTTGTAGTT
1 M P I Q C R Y Y E D L F P E V G D V V V
61 GTAACAGTTAAGGTCATCCAACCTATGGGTAGCTATGTGGAACCTCTAGAGTACAAGAAT
21 V T V K V I Q P M G S Y V E L L E Y K N
121 ATTGAGGTATGATATTGCTCAGTGAGCTGTCTCGTACGATATAAGATCAATCAGCAAG
41 I G G M I L L S E L S R R R I R S I S K
181 CTTGTTCAATCGGGTCAAACTAGTGAGTACTGTTGACGAGTCGATAGTCCAAAAGGT
61 L V R I G S N T E V T V V R V D S A K G
241 TATATTGATTTATCAAAAAGAAGAGCTTCTGCAGAGGAGATCGAAAAGTCAAGGAAAGG
81 Y I D L S K R R A S A E E I A K C K E R
301 TTTGCAGAGGCTAAAGCGGTCAATCAAAATTTGAGGAATGTAGCCGAGAAGTTAGAGTAT
101 F A E A K A V N Q I L R N V A E K L E Y
361 GAGACAGATGTACAACCTCAAGAGCTGTGCGAGAAAAGTCTGTTGATTTTACAAAAAG
121 E T D V Q L E E L C R K T A W Y F D K K
421 ACTGGTCAAAGGCGAGGTTCTACGATATTTCAAGAAGTTGTAATTCACCAGAAAT
141 T G R K A G F Y D I F K K V V N S P E I
481 CTCGATGAGTGTGACATAGCAGCAGCAACAAAAGAAATGCTTCTTACGATATCAGGCAT
161 L D E C D I D Q P T K E M L L T D I R H
541 CGATTAACACCAAAGGCGAGTAAAATTCGCGCTGACTTTGAAGTTTCATGTTTCACTTAT
181 R L T P K A V K I R A D F E V S C F T Y
601 GATGGTATTGATGCCGTGAGAAGCGCACTTCGATCTGACTGAACTCAATTCAGATGCT
201 D G I D A V R S A L R S G L K L N S D A
661 CTTCAAATCCGCATCAATTTGATAGCGCCACCGCTTATGACTAACACGCAAAACAATG
221 L P I R I N L I A P P L Y V L T T Q T M
721 GATCGAGCTGCCGGCTTAGAACAGCTGAATGAAGTCTAAACGCTACCCAGACATCCATT
241 D R A A G L E Q L N E V L N V I Q T S I
781 GAAAGTCAAGTGGTTCTTTCAAAATTCGACAGGCTCCTCGTGTGTTTCAGATACAGAT
261 E S Q G G S F K I R Q A P R V V S D T D
841 GACGCAGAATTACAGCGTCAATGGATGAAGTAAAGGCGAATCGTGAAGTTTCTGGT
281 D A E L Q R Q M D E L E K A N R E V S G
901 GATGAGGATGATGAAGACGATGACGAGGATGAAGATGATGATGAGGAATCCAATGATGGA
301 D E D D E D D D E D E D D D E E S N D G
961 GATCAGAATGAACATAAGTAA
321 D Q N E H K *

```

图 1 日本血吸虫尾蚴 eIF2 琢 DNA 序列及所编码的氨基酸序列

Fig.1 cDNA and amino acid sequences of theof alpha subunit mRNA eukaryotic translation initiation factor 2 from Schistosoma japonicum cercariae

```

Sm:301tttgcgaagcaaaagcggttaccagatcttgaggaaatgtagctgagaagtagattat360
Sj:301tttgcgaagcctaaagcggctcaatcaaatcttgaggaaatgtagccgagaagtagagtat360
Sm:361 aagactgtagaacaacttgaagaactatgtagaaaacccgcttggtattttgacaggaag420
Sj:361 gagacagatgtacaactcgaagagcgtgtaggaagaaactgcttggtattttgacaggaag420
Sm:421actggtcgtagagctggttcgtacgacattttcaagaagttgtaaaatccaccagaaat480
Sj:421actggtcgaaagcagggcttctacgataatttcaagaagttgtaaaatccaccagaaat480
Sm:481cttgatgagtgtagacatagatcaggcaacaaaagaagtgcttctcaccgatatcagacat540
Sj:481ctcgatgagtgtagacatagaccagcaacaaaagaagtgcttctcaccgatatcagacat540
Sm:541cgattaacacacaaagcagtagaaatctgctgacttctgagggttctggtttcacttat600
Sj:541cgattaacacacaaagcagtagaaatctgctgacttctgagggttctggtttcacttat600
Sm:601gattggtatttgatgctgtttaaagtgactcggctcgactggaactcaatcagattct660
Sj:601gattggtatttgatgctgtcagaagcgactcggctcgactggaactcaatcagattct660
Sm:661ctcccaatccggtcaatctgatgaccaccacctttatgtactgaccacacaaacaat720
Sj:661cttccaatccggtcaatctgatgaccaccacctttatgtactgaccacacaaacaat720
Sm:721gactgagctgctggcttagaacagctaaatgaagtgtagatgtagatcaagaatcgaatt780
Sj:721gactgagctgctggcttagaacagctaaatgaagtgtagatgtagatcaagaatcgaatt780
Sm:781gaaagtcagggtggttcattcaaaatcaacaggtcctcgagttggtttcagatacagat840
Sj:781gaaagtcagggtggttcattcaaaatcaacaggtcctcgagttggtttcagatacagat840
Sm:841gacgaggacttacaacgcacaaatggatgaattagaaaaagcgaatcgtgaagttctgg900
Sj:841gacgcagaattacagcgtcaaatggatgaactagaaaaagcgaatcgtgaagttctgg900
Sm:901gattgaggagatgaagacgaagatgacgggtaggagaagatgtagtaggactcgaat960
Sj:901gattgaggatgtaggaagac---gattgacaggat---gaagatgattgaggactccaat960
Sm:961gattggagaccgaatg976
Sj:955gattggagaccgaatg970

```

图 2 日本血吸虫 eIF2 琢 DNA 与曼氏血吸虫 eIF2 琢 mRNA 序列同源性比较

Fig.2 Homology comparison between from Schistosoma mansoni(Sm) alpha subunit of translation initiation factor 2 mRNA and alpha subunit cDNA of Schistosoma japonicum(Sj) translation initiation factor 2
 -:ShowheterogenousgenebetweenSmandSj

Score= 490bits(1261), Identities=257/325 (79%)

Sj:1 MPIQCRYEDLFPEXXXXXXXXXXIQPMGSYVELLEYKNIIGMXXXXXXXXXXXXXXXXX 60
MPIQCR+YEDLFPE IQMGSYVELLEYKNIIGM

Sm:1 MPIQCRFYEDLFPEVGDVVLVTVKVIQSMGSYVELLEYKNIIGMILHSELRRIRRSISK 60

Sj:61 XXXXGSNTEVTVRVVDSAKGYIDLKRRASAEIAKCKERFAEAKAVNQILRNVAEKLEY 120
GSNTEVTVRVVDSAKGYIDLKRRASAEIAKCKERFA+AKAVNQILRNVAEKLY

Sm:61 LVRI GSNTEVTVRVVDSAKGYIDLKRRASAEIAKCKERFAKAKAVNQILRNVAEKLDY 120

Sj:121 ETDVQLEELCRKTAWYFDKKTGRKAGFYDIFKKVNSPEILDECDIQPTKEMLLTDIRH 180
+TDQLEELCRKTAWYFD+KTGR+AG YDIFKKVNSPEILDECDIQ TKEMLLTDIRH

Sm:121 KTDEQLEELCRKTAWYFDRKTRRAGSYDIFKKVNSPEILDECDIQATKEMLLTDIRH 180

Sj:181 RLTPKAVKIRADFEVSCFTYDGI DAVRSALRSGKLNSDALPIRINLIAPPLYVLTQTMT 240
RLTPKAVKIRADFEVSCFTYDGI DAV+SALRSGL+LNSD+LPRIINLIAPPLYVLTQTMT

Sm:181 RLTPKAVKIRADFEVSCFTYDGI DAVKSALRSGLELNSDSLPIRINLIAPPLYVLTQTMT 240

Sj:241 DRAAGLEQLNEVLNVIQTSIESQGGSFKIRQAPRVSDTDDAELQRQMDLEKANREVS 300
DRAAGLEQLNEVL+VI+ SIESQGGSFKI+QARVVSDTDDA+LQRQMDLEKANREVS

Sm:241 DRAAGLEQLNEVLVDVIKKSIESQGGSFKIQQAARVVSDTDDADLQRQMDLEKANREVS 300

Sj:301 --XXXXXXXXXXXXXXXXXNDGDQN 325
SNDGDQN

Sm:301 DEDEDEDDGDEEDDEASNDGDQN 325

图 3 日本血吸虫与曼氏血吸虫 eIF2 α 亚基蛋白序列同源性比较

Fig.3 Homology comparison of the protein sequences of alpha subunit of the translation initiation factor 2 from *Schistosoma mansoni* and *Schistosoma japonicum*

苷酸结合域 (ligonucleotide-binding domain) 是由五条肽链相互缠绕形成一个封闭的桶状结构。另一部分为螺旋状结构域 (helical domains)。这两部分结构共同形成了一个 S1 结构域。该结构域中包含 RNA 结合区 (RNA-binding domain)。但目前尚无实验证明 eIF2 α 与 RNA 结合。eIF2 α 具有重要的生物学功能即参与蛋白合成起始的调控。许多实验证明在病毒感染、热休克、重金属中毒、氨基酸极度缺乏的状态下的细胞以及一些血清的细胞中 eIF2 α 第 51 位丝氨酸 (Ser-51) 可被一些蛋白激酶如 RIK1、RNAPKR、CN2、ERK、KR-like ER kinase 等磷酸化。而翻译的起始被中断、蛋白合成被抑制、细胞的生长受到抑制。这一机制与病毒感染的细胞和肿瘤细胞中的蛋白质合成密切相关。因此许多有关 eIF2 α 磷酸化的研究正在进行之中。

我们从日本血吸虫尾蚴 cDNA 文库中筛到并克隆的与曼氏血吸虫 eIF2 α 亚基基因同一性高达 87% 的 cDNA 全编码区序列。经过几种蛋白分析软件的分析后发现它具有已知的 eIF2 α 亚基的特殊结构即 S1 结构域。从以往的资料可知 S1 结构域是首先从核糖体蛋白 S1 中发现的 RNA 结合区。S1 蛋白被证明通过其 S1 结构域连接 RNA。S1 结构域也被证明存在于其它一些物种的 eIF2 α 序列中。

综上所述所发现的 cDNA 编码日本血吸虫的 eIF2 α 亚基全长 cDNA 序列已经准确克隆到 pGEM-T 载体上。为进一步的功能研究奠定了基础。

参考文献

- 陈贤义, 姜庆五, 王立英等. 2001 年全国血吸虫病疫情通报. 中国血吸虫病防治杂志 (Chin J Schisto Control), 2002, 14(4): 241-3.
- Pena S. Third world participation in genome projects. Trends Biotechnol, 1996, 14(1): 74-7.
- Johnston AD. The WHO/UNDP/World Bank Schistosoma genome initiative: current status. Parasitol Today, 1997, 13(1): 45-6.
- Unnasch T. The filarial genome project. Parasitol Today, 1994, 10(3): 415-6.
- Gloria R, Mark D, Bento S, et al. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. Gene, 1995, 152(2): 141-7.
- Adams M, Dubrick M, Keriavage A, et al. Complementary cDNA sequencing: expressed sequence tags and human genome project. Science, 1991, 252(5013): 1651-6.
- 陈晓光, 李华, 彭鸿娟等. 日本血吸虫尾蚴 cDNA 文库的构建及分析. 中国寄生虫学与寄生虫病杂志, 2002, 20(1): 39-41.
- Chen XG, Li H, Peng HJ, et al. Construction and characterization of the cDNA library from *Schistosoma japonicum* cercariae. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi, 2002, 20(1): 39-41.
- Nonato MC, Widom J, Clardy J. Crystal structure of the N-terminal segment of human eukaryotic translation initiation factor 2. J Biol Chem, 2002, 277(19): 17057-61.
- Pain VM. Initiation of protein synthesis in eukaryotic cells. Eur J Biochem, 1996, 15: 236(3): 747-71.
- Kaufman RJ. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. Genes Dev, 1999, 15: 13(10): 1211-33.