猪带绦虫成虫凝溶胶蛋白基因的生物信息学分析*

申萍香[1]，王 宇[2]，黄 江[2**]
(1. 贵阳医学院 多媒体形态学实验室，贵州 贵阳 550004；2. 贵阳医学院 寄生虫学教研室，贵州 贵阳 550004)

[摘 要] **目的：** 分析猪带绦虫(*Taenia solium*)成虫凝溶胶蛋白(gelsolin，GEL)基因及编码蛋白的结构和功能。**方法：** 利用生物信息网站美国国家生物技术信息中心(NCBI)和瑞士生物信息学研究所的蛋白分析专家系统(ExPASY)中生物信息学分析工具，并结合其它分析软件，从获得的猪带绦虫成虫全长 cDNA 质粒文库的表达序列标签(EST)中识别凝溶胶蛋白基因，分析该基因的结构并预测其编码蛋白质的结构和功能特征。**结果：** 猪带绦虫成虫凝溶胶蛋白基因与细粒棘球绦虫凝溶胶蛋白的一致性为 88%，相似性为 93%；全长 1 514 bp，编码区为 167～1 263 bp，编码 365 个氨基酸，无各种亚细胞定位序列，具有多个潜在的磷酸化位点，蛋白在溶液中性质稳定。**结论：** 应用生物信息方法从猪带绦虫成虫 cDNA 质粒文库中筛选出了猪带绦虫凝溶胶蛋白 cDNA 全长序列并预测其结构与功能。

[关键词] 有钩绦虫猪肉绦虫；凝溶胶蛋白；蛋白质序列分析
[中图分类号] R383.39；R532.39 [文献标识码] A [文章编号] 1000-2707(2011)02-0139-04

Bioinformatics Analysis on Gelsolin Gene of *Taenia Solium* Imago

SHEN Pingxiang[1], WANG Yu[2], HUANG Jiang[2]
(*1. The Multimedia Laboratory of Morphological Science, Guiyang Medical College, Guiyang 550004, Guizhou, China; 2. Department of Parasitology, Guiyang Medical College, Guiyang 550004, Guizhou, china*)

[**Abstract**] **Objective**: To analyze the structure of gelsolin gene from *Taenia solium* adults (Ts GEL) and the structure and function of the encoded protein. **Methods**: Gelsolin gene was identified from *T. solium* full-length cDNA plasmid libratory by analyzing tools. Its structure was analyzed, and the structure and function characteristics of the encoded protein were predicted. **Results**: Consistency and similarity of Ts GEL with *Acephalocystis granulosus* gelsolin gene were 88% and 93% respectively. The full length of Ts GEL was 1 514 bp. Its coding region was 167～1 263, encoding 365 amino acids. The encoded protein didn't contain any kinds of subcellular localization sequence, but contained several potential phosphorylation sites. It was stable in solution. **Conclusions**: The full cDNA sequence of gelsolin gene is screened from cDNA plasmid library of *T. solium* imago with bioinformatical method. The structure and function of the gene and encoded protein are predicted.

[**Key words**] *Taenia solium*; gelsolin; protein array analysis

猪带绦虫成虫寄生于人体肠道，引起猪带绦虫病，幼虫寄生于人体皮下、肌肉或内脏引起囊尾蚴病[1]。猪带绦虫在发展中国家分布很广，在发达国家的分布也有明显增加趋势，目前猪带绦虫病和囊虫病的发病仍然是不可忽视的公共卫生问题。本研究从猪带绦虫成虫全长 cDNA 文库中筛选了凝溶胶蛋白基因，对该基因及其编码的蛋白结构和特征进行全面的生物信息学分析，以指导后续目的基因的克隆表达、生物学功能及免疫学特性的研究，为其在猪带绦虫病的诊断、药物及疫苗研究中的应用提供依据。

1 材料和方法

1.1 材料 猪带绦虫成虫全长 cDNA 质粒文库的

*[基金项目]贵州省科技攻关项目[黔科合 NY 字(2008)3060]。
**通讯作者 E-mail：mmm_hj@gmc.edu.cn

构建和 EST 测序及 Unigene 分析，由本课题组与上海联合基因有限公司合作完成，Unigene 通过 Wu－blastx 方法进行识别[2]。编码猪带绦虫成虫凝溶胶蛋白（*Ts* GEL）基因文库质粒编号为 ZDTC 21－C7。

1.2 方法 通过 NCBI 网站（http://www.ncbi.blastX）对不同进化阶元的其它物种凝溶胶蛋白的同源蛋白序列进行比对，分析该基因的翻译序列与其他蛋白质氨基酸序列的一致性，判断其是否为全长基因；利用 Vector NTI Suite 8.0 软件包中的 AlignX 程序进行多序列联配分析，利用 ORF Finder（开放阅读框探寻器）确定其完整编码序列并显示其编码的氨基酸序列。利用蛋白分析专家系统服务器 ExPASy Proteomics Server（http://ca.expasy.org/）所提供的蛋白质在线分析工具预测氨基酸序列的分子量、等电点、稳定性指数等理化性质，InterPro Scan（http://www.ebi.ac.uk/InterProScan/）和 MotifScan（http://myhits.isb-sib.ch/cgi-bin/motif_scan）预测氨基酸序列中的基序（motif）和功能域；利用 PredictProtein（http://cubic.bioc.columbia.edu/predictprotein/）预测氨基酸序列的跨膜区和拓扑结构以及二级结构、通过 IEDB Analysis Recource（http://tools.immuneepitope.org/tools/bcell/iedb_input）分析抗原表位，利用 SWISS-MODEL（http://swissmodel.expasy.org/）进行三级结构的同源建模，预测蛋白质的空间构象，并通过 Vector NTI suite 8.0 软件包中的 3D molecule viewer 看图软件显示。利用 PCgene 软件分析包中抗原决定簇分析程序将凝溶胶蛋白基因的主要抗原表位标示在三维结构图上。

2 结果

2.1 BlASTX 的分析结果 从 BLASTX 的分析结果显示，该基因是凝溶胶蛋白的同源基因，与 GenBank 中细粒棘球绦虫的同源基因 AAK 15 753.1 的氨基酸序列的一致性达 93%，相似性达 96%；从比对结果来看，该克隆基因的 5′端和 3′端序列均长于细粒棘球绦虫完整凝溶胶蛋白编码序列，故推测为猪带绦虫成虫凝溶胶蛋白的全长基因序列，其最大的 ORF 就是其完整的编码区，该基因全长 1 514 bp，编码区为 167～1 263 bp，编码 365 个氨基酸，在 5′端和 3′端都有非翻译区，见图 1。

```
   1  GCTAGCTCGCCGCAG CCGAACGACCGAGCG CAGCGAGTCAGTGAG CGAGGAAGCGGCCGC ATAACTTCGTATAGC ATACATTATACGAAG
                                                                                 M  A  G  L  I
  91  TTATCAGTCGACGGT ACCGGACATATGCCC GGGAATTCGGCCATT ACGGCCGGGGATCTT CGAGTAAGCCTTGGC TATGGCAGGTCTGAT
       K  P  K  T  Y   D  W  K  D  S   N  V  A  S  V   G  S  A  M  D   R  E  V  K  K   E  S  A  M  T
 181  AAAGCCTAAGACCTA TGACTGGAAGGACTC CAACGTTGCCAGCGT TGGCTCGGCGATGGA CAGAGAGGTTAAGAA GGAGTCTGCGATGAC
       E  K  C  W  E   P  V  G  R  A   K  S  P  F  L   M  V  W  R  V   K  Q  F  K  L   E  P  V  P  S
 271  GGAGAAGTGTTGGGA GCCTGTTGGCAGAGC GAAATCGCCCTTCCT CATGGTCTGGCGGGT TAAGCAATTCAAGCT CGAGCCAGTGCCCAG
       D  E  V  G  N   F  Y  N  G  D   S  Y  V  I  C   R  A  T  R  S   P  G  G  D  K   L  L  Y  N  V
 361  TGATGAGGTTGGGAA TTTTTACAATGGCGA TTCCTACGTCATCTG CAGGGCAACGAGAAG CCCTGGTGGTGACAA GCTACTCTACAATGT
       H  F  W  I  G   K  H  S  T  A   D  E  Y  G  T   A  A  Y  K  T   V  E  L  D  T   F  L  D  D  A
 451  TCACTTCTGGATTGG CAAACACAGTACAGC TGATGAGTACGGTAC CGCTGCCTACAAGAC TGTCGAATTGGACAC CTTCCTCGATGATGC
       A  V  Q  H  R   E  V  E  G  Y   E  S  P  L  F   K  S  Y  F  D   K  L  V  I  L   K  G  G  Y  A
 541  TGCGGTCCAACATCG TGAGGTCGAAGGCTA CGAGTCGCCACTCTT CAAGAGCTACTTTGA CAAGCTTGTCATCCT AAAGGGTGGCTACGC
       S  G  F  R  H   V  K  P  E  E   Y  R  P  R  L   L  R  F  C  K   E  G  K  I  T   Y  M  R  Q  E
 631  CTCTGGTTTCCGTCA TGTCAAGCCGGAAGA GTATAGGCCACGGTT GCTACGCTTCTGCAA GGAAGGGAAAATCAC TTATATGCGCCAGGA
       T  F  S  K  Q   S  V  H  S  G   D  V  F  I  L   D  L  G  D  R   A  Y  Q  F  N   G  S  K  C  S
 721  GACCTTTAGCAAGCA ATCTGTCCACTCTGG TGACGTGTTCATCCT GGATCTGGGCGATCG GGCCTATCAGTTCAA CGGCTCCAAGTGCTC
       A  F  E  K  S   A  A  A  A  F   L  Q  D  L  E   S  K  R  N  G   R  C  N  T  S   V  L  D  E  A
 811  TGCCTTTGAGAAGAG CGCGGCCGCAGCCTT TCTGCAGGATTTGGA AAGCAAACGTAATGG ACGCTGCAACACTTC CGTCTTGGATGAAGC
       H  T  P  Q  D   H  E  F  W  T   T  L  P  D  T   P  V  E  E  P   E  P  P  K  E   V  V  K  S  L
 901  CCACACACCACAGGA TCACGAATTCTGGAC TACGCTACCCGACAC ACCTGTAGAAGAGCC AGAACCTCCTAAAGA AGTTGTCAAGTCCCT
       Y  K  I  S  D   A  S  G  K  M   E  L  T  I  V   R  E  G  S  A   P  K  R  D  I   K  P  D  D  V
 991  YTACAAGATATCTGA CGCTAGTGGTAAGAT GGAGTTGACCATTGT GAGGGAGGGCTCGGC CCCCAAGCGTGACAT TAAGCCCGACGATGT
       Y  M  I  L  A   K  E  G  P  F   V  Y  V  G  R   D  C  F  V  L   E  K  R  N  A   L  S  N  A  H
1081  TTACATGATCCTTGC CAAGGAGGGTCCCTT CGTTTATGTAGGCAG GGATTGTTTTGTTTT GGAGAAGAGAAACGC TCTTTCCAACGCCCA
       K  F  L  Q  S   C  P  N  P  F   L  P  V  P  V   V  T  D  E  Q   G  E  S  F  L   R  G  V  W  D
1171  CAAATTTTTGCAGTC CTGCCCCAACCCCTT CTTGCCCGTCCCCGT CGTGACTGATGAGCA GGGGGAATCATTCCT AAGGGGCGTTTGGGA
       E  *
1261  TGAGTAGACGGTGCA AGACTTGTTTCGACT TCGACCAAACAAAAC GCAAATTATGTCGCC AATGCTTTTTTTAAT TTAGCATATCAAGTG
1351  CCTTAAGTTTCATTT TTATTTTAATCTTGA TGTTCGTTCCTGTTT TAACTGATTTCATTA AAAGCATTGTTAGTG TAAAAAAAAAAAAAA
1441  AAAAAAAAAAAAAAA AAAAAACATGTCGGC CGCCTCGGCCCTCGA GAAGCTTTCTAGACC ATTCGTTTGGCGCG
```

图 1 Ts GEL 全长 cDNA 序列及其 ORF 编码的氨基酸序列

Fig. 1 Full-length cDNA sequence of *Ts* GEL and the amino acid sequence encoded by ORF

2.2　*Ts* GEL 蛋白质的理化性质　猪带绦虫 *Ts* GEL 蛋白质的理论分子量和等电点分别是41 364.8和5.99,在酵母和大肠杆菌中体内表达的半衰期分别大于20 h 和10 h;在溶液中的不稳定指数为41.98,高于阈值40,在溶液中性质不稳定。脂肪族指数为67.81,总亲水性 -0.541,蛋白质总体疏水性高。

2.3　*Ts* GEL 蛋白翻译后修饰与结构的特征性序列　模序(Motif)分析结果显示:*Ts* GEL 蛋白含有2个可能的糖基化位点:210~213,238~241 位氨基酸;9 个潜在的酪蛋白激酶 II(CK2)磷酸化位点:22~25,103~106,120~123,142~145,215~218,240~243,247~250,256~259,260~263 位氨基酸;3 个潜在的 N 肉豆蔻酰位点:69~74,152~157,211~216 位氨基酸; 4 个潜在的蛋白激酶 C(PKC)磷酸化位点:35~37,231~233,274~276,281~283 位氨基酸;在 29~32 位氨基酸存在一潜在的 cAMP 和 cGMP 依赖的蛋白激酶磷酸化位点。

2.4　*Ts* GEL 的亚细胞定位序列　ExPASY 中 TargetP 预测该基因无分泌型信号肽, 未发现线粒体、过氧化酶体、溶酶体和细胞核等亚细胞定位序列。

2.5　InterPro Scan 扫描一级结构中包含的结构和功能域特征序列 ExPASY 中 InterPro Scan 扫描结果显示,该氨基酸序列中含有2个凝溶胶蛋白保守功能结构域,结果见图2。

2.6　*Ts* GEL 蛋白质的拓扑结构、二级结构和亲水性特征　Htm 预测该蛋白没有跨膜序列,见图3。Sec 预测 α 螺旋(H)、β 折叠(E)和无规卷曲(L)的比例分别是 21.64∶28.77∶49.59, 二级结构以无规则卷曲为主。根据每个氨基酸残基的溶剂可及性(Acc)分析,预测在水溶性中,该蛋白质折叠成一个不甚致密的球状分子。

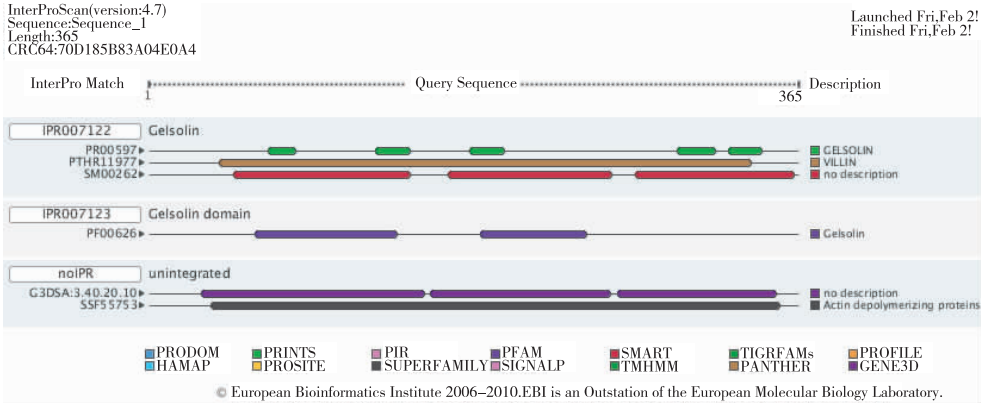


图 2　猪带绦虫 *Ts* GEL 的功能域与结构域

Fig. 2　The fuctional domain and structural domain of *Ts* GEL



图 3　*Ts* GEL 蛋白质的拓扑结构和跨膜区(Predictprotein 分析)

Fig. 3　The topology structure and transmembrane regions of *Ts* GEL analyzed by Predictprotein

2.7 *Ts* GEL 的抗原表位分析 通过 IEDB Analysis Recource 提供的蛋白质抗原表位在线分析工具,预测其主要线性抗原决定簇的位置分别是:9 ~ 35、61 ~ 73、102 ~ 113、127 ~ 138、242 ~ 253、256 ~ 272、291 ~ 302、349 ~ 357 位氨基酸,推测该蛋白是一个多功能的蛋白质,其 B 细胞抗原表位,见图 4。

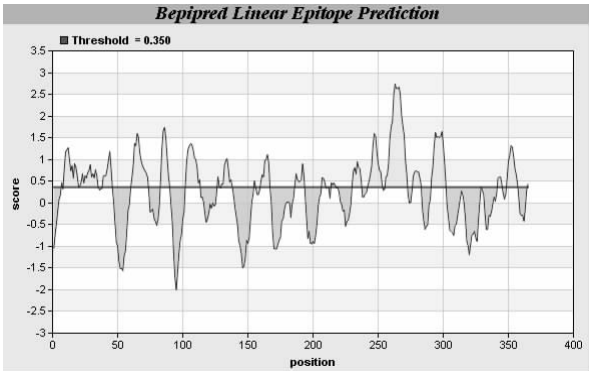


图 4 *Ts* GEL 的 B 细胞抗原表位图

Fig. 4 B cell epitope map of *Ts* GEL

2.8 *Ts* GEL 的三维结构图及线性表位的位置 通过 Swiss-model 将 *Ts* GEL 与蛋白结构数据库中蛋白质三维结构进行匹配,输回模拟的凝溶胶蛋白三维结构图文件,在 vector NTI suite 软件包中打开该蛋白质结构文件,将三个主要的线性表位标示在结构图上,见图 5。3 个棍状区域代表 3 个主要线性表位为 61 ~ 73、102 ~ 113、127 ~ 138 位氨基酸。

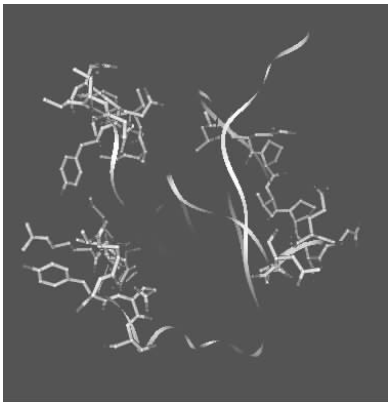


图 5 猪带绦虫 Ts GEL 的 3D 结构和线性表位的空间位置

Fig. 5 3D structure and spacial position of linear epitope of *Ts* GEL

3 讨论

猪带绦虫广泛存在于发展中国家,随着移民的增加也逐渐分布于发达国家和地区,全世界每年大约有 5 亿人感染猪带绦虫[3,4],并且每年因该寄生虫而死亡的人数可达 5 万[5],猪带绦虫病和囊虫病的发病在世界范围内仍然是不可忽视的公共卫生问题,因此在疾病的预防和诊治方面尚需更多的研究。为了筛选在猪带绦虫病的诊断、免疫或治疗方面的靶分子,本课题组近年构建了猪带绦虫成虫全长 cDNA 文库,从猪带绦虫基因组及蛋白质组学的研究入手,系统地开展猪带绦虫基因的鉴定和功能研究。

细胞骨架重排发生在多种细胞生命活动中,使细胞质在凝胶态和溶胶态之间转换。这一可逆性过程涉及多种蛋白质,其中凝溶胶蛋白(gelsolin) 超家族通过切断肌动蛋白丝、脱离肌丝终端 、促进肌动蛋白聚集成核等方式控制肌动蛋白的构建,使肌动蛋白在单体和聚合体两种形式间通过聚合解聚作用维持动态平衡[6]。凝溶胶蛋白作为凝溶胶蛋白家族的成员之一,能够与肌动蛋白结合并且调节肌动蛋白的聚合解聚作用,是一种重要的肌动蛋白结合蛋白,也是目前研究较为详尽和透彻的重要蛋白分子,其主要功能包括调节肌动蛋白活性、细胞运动、细胞凋亡、吞噬功能以及血小板的形成和活化等[7]。研究表明,凝溶胶蛋白与凋亡、凝血、免疫、肿瘤及炎症等病理生理过程密切相关[8]。绦虫有自动脱落节片的生理功能,根据凝溶胶蛋白的生物学功能显示,该蛋白在真核生物中扮演着重要的角色,是否在绦虫体内也起到重要的作用,尤其与绦虫在体内的运动和节片脱落的关系值得深入研究。同时根据该蛋白的功能显示,绦虫凝溶胶蛋白有可能是一个重要的药物靶标。

为了对其机制进行更深入的研究,本文通过 Blastx 从猪带绦虫 cDNA 文库中识别出了一个编码凝溶胶蛋白的全长基因,该基因该基因全长 1 514 bp,编码区为 167 ~ 1 263 bp,编码 365 个氨基酸,在 5′端和 3′端都有非翻译区。并通过生物信息学的方法对其所编码的蛋白的结构和功能进行预测,结果显示该蛋白为多功能蛋白,在溶液中不稳定,其 InterPro Scan 扫描结果显示具有两个保守功能结构域,二级结构预测该蛋白可能以胞浆型存在,无各种亚细胞定位序列。对凝溶胶蛋白的研究进展和生物信息学分析,有助于采取合理的克隆和表达策略,选用适合的表达载体,提高目的蛋白高效的可溶性表达的可能性及获得有活性的

(下转第 146 页)

海马 LTP 是目前公认的从突触水平上研究学习记忆的神经模型,海马 LTP 分为诱导和维持两个阶段,其诱导依赖 NMDA 受体激活,而维持则与 A-氨基酸-3 羟基-5 甲基-4 异恶唑丙基受体功能和电压依赖性 Ca^{2+} 通道开放有关。PS 幅值是 LTP 中的一项指标,反映了突触后细胞同步活动。在本研究中,海洛因戒断组在高频刺激前后的 PS 幅值变化率无明显变化,说明戒断组的突触传递增强受到了抑制,而海洛因依赖组的 PS 幅值变化率与对照组比较无显著差异,表明突触可能产生了适应性的变化。突触传递的适应性变化已经被认为是药物成瘾的一个重要因素[9],伴随着大鼠对海洛因的依赖形成,正常情况下依赖于海洛因的 CA3 区的突触功能在海洛因的作用下,产生了适应性的变化,当停用海洛因后,突触功能便受到影响。据文献报告,MK-801 能在 CA1 区抑制 LTP 的诱导,表明 CA1 区的突触传递可能有 NMDA 的介导,而本次研究中在 LPP-CA3 通路观察 PS 幅值的变化率,发现与对照组的变化率无显著差异,表明此通 路的突触传递功能持续增强可能是非 NMDA 受体所介导。但总的来说,LTP 的形成与药物的成瘾机制仍是十分复杂的问题,有待进一步研究。

4 参考文献

[1]邓祥发. 海洛因成瘾大鼠脑神经元死亡方式及其超微结构的变化[J]. 广西医科大学学报,2004(5): 660 – 662.

[2]潘贵书,徐国强,李淑芳,等. 海洛因成瘾大鼠模型的建立[J]. 贵阳医学院学报,1997(增):46.

[3]Maldonado R,Negus S,Koob GF. Precipitation of morphine withdrawal syndrome in rats by administration of mu delta and kappa selective opioid antagonists[J]. Neuropharmacology, 1992(12):1231.

[4]Wei E,Loh HH,Way EL. Quantitative aspects of precipitated abstinenle in morphine-dependent rats[J]. J Pharmacol Exp Ther,1973(2):398 – 403.

[5]张开镐. 药物依赖性的动物实验方法(一)[J]. 中国药物依赖性杂志,1999(1):23 – 26.

[6]叶晓明,王新华,李家乐,等. 吗啡依赖和戒断大鼠某些脑核团中 β-内啡肽含量的变化[J]. 中国疼痛医学杂志,1999(4):233 – 236.

[7]包新民. 大鼠立体定位图谱. 北京:人民卫生出版社,1991:175.

[8]Ali DW, Salter MW. NMDA receptor regulation by Src kinase signaling in excitatory transmission and plasticiy[J]. Curr opin Neuro-biol,2001(3):336 – 342.

[9]Milner TA, Drake CT Ultrastructural, evidence for presynaptic mu-o-piod receptor modulation of synaptic plasticity in NMDA-receptor containing dendrites in the dentate gyrus[J]. Brain Res Bull,2001(2):131.

(2011 – 03 – 08 收稿,2011 – 03 – 21 修回)

(上接第 142 页)

重组蛋白等,为进一步研究其生物功能及其该蛋白在诊断、药物及疫苗研究中的应用提供依据。

4 参考文献

[1] 吴观凌. 人体寄生虫学[M]. 3 版. 北京:人民卫生出版社,2005:515 – 530.

[2] 黄江,胡旭初,徐劲,等. 猪带绦虫成虫 cDNA 质粒文库的构建及 EST 测序[J]. 中国人兽共患病学报,2008(12):1126 – 1128.

[3] Aguilar-Diaz H,Bobes RJ,Carrero JC,et al. The genome project of Taenia solium [J]. Parasitol Int,2006(55):127 – 130.

[4] Sciutto E,Fragoso G,Fleury A,et al. Taenia solium disease in humans and pigs: an ancient parasitosis disease rooted in developing countries and emerging as a major health problem of global dimensions [J]. Microbes Infect, 2000(2):1875 – 1890.

[5] Mafojane NA ,Appleton CC,Krecek RC,et al. The current status of neurocysticercosis in Eastern and Southern Africa [J]. Acta Trop,2003(87):25 – 33.

[6] Yin HL. Gelsolin:calicium and polyphosophoinositide regulated actin modulating protein [J]. Bioessays,1987(4): 176 – 179.

[7] 陈琦,姚咏明. 凝溶胶蛋白结构和功能研究新进展[J]. 中国病理生理杂志,2010(7):1433 – 1439.

[8] 李明娟,闵锐. 凝溶胶蛋白及其生物医学作用[J]. 生命科学,2010(3):267 – 271.

(2011 – 03 – 09 收稿)