

亚洲带绦虫烯醇酶基因及其蛋白质的结构与功能

黄 艳¹, 黄 江³, 胡旭初¹, 徐 劲¹, 曹开源², 余新炳¹, 包怀恩³, 郎书源³
(中山大学中山医学院 1. 寄生虫学教研室, 2. 免疫学教研室, 广东广州 510089;
3. 贵阳医学院寄生虫学教研室, 贵州贵阳 550004)

摘要:【目的】分析和预测亚洲带绦虫烯醇酶基因及其编码蛋白的结构和特性,用于指导其生物学功能的实验研究。【方法】利用美国国家生物技术信息中心和瑞士生物信息学研究所的蛋白分析专家系统中有关基因和蛋白的序列和结构信息分析的各种工具,结合其他生物信息学分析软件包,如 Pgene 和 Vector NTI suite,从亚洲带绦虫全长 cDNA 质粒文库中识别烯醇酶基因及其编码区,分析、预测该基因编码的蛋白质的理化特性、翻译后的修饰位点、功能域、亚细胞定位、拓扑结构、二级结构、三维空间构象等。【结果】该基因全长 1737 bp,编码区为第 205~1503 bp,编码 433 个氨基酸,为全长基因。GenBank 中与棘口吸虫烯醇酶氨基酸序列同源性最高,达 78%。相对分子量理论预测值为 46653.5 Ku。没有质体、线粒体定位序列。预测编码蛋白有 1 个跨膜区,3 个亲水性部位。与吸虫属的烯醇酶进化关系最近。【结论】应用生物信息方法从亚洲带绦虫成虫 cDNA 文库中筛选出了亚洲带绦虫烯醇酶 cDNA 全长序列并预测得到其结构与功能方面信息。

关键词: 亚洲带绦虫; 烯醇酶; 结构; 功能; 生物信息学

中图分类号: R383.3+2 文献标识码: A 文章编号: 1672-3554(2008)05-0575-06

Structure and Characteristics of Gene and Protein of Enolase from *Taenia asiatica*

HUANG Yan¹, HUANG Jiang³, HU Xu-chu¹, XU Jin¹, CAO Kai-yuan²,
YU Xin-bing¹, BAO Huai-en³, LANG Shu-yuan³

(1 Department of Parasitology, 2 Department of Immunology, Zhongshan School of Medicine, SUN Yat-sen University, Guangzhou 510089, China; 3 Department of Parasitology, Guiyang Medical College, Guiyang 550004, China)

Abstract: 【Objective】 To get the messages on the structures and characteristics of enolase from *Taenia asiatica* (T.a.ENO) by bioinformatics. 【Methods】 A full-length cDNA sequence encoding enolase from cDNA plasmid library of *Taenia asiatica* was identified by using tools of bioinformatics at webs sites of NCBI. The characteristics of the deduced protein including the physico-chemical characteristics, modification sites after translation, domains, subcellular location, topological structure, secondary structures, and 3D structure were predicted by employing bioinformatics software package supplied by the website of ExPaSy. 【Results】 The full cDNA sequence encoding T.a.ENO includes a complete open reading frame of 1299bp which encoded a putative protein of 433 amino acids. The coding region is 205 bp ~ 1503 bp. The amino acids sequence has a high identity with enolase from other species in GenBank. The protein has one transmembrane region and stable physico-chemical characteristics. The molecular weight of T.a.ENO is predicted to be 46653.5u. The protein has three hydrophilic regions. The relationship of phylogenesis between T.a.ENO and enolase of other trematodes is close. 【Conclusion】 The cDNA sequence encoding enolase was screened from cDNA library of adult *Taenia asiatica* by bioinformatics. The structure and characteristics of the gene and protein of T.a.ENO were obtained.

Key words: *Taenia asiatica*; Enolase; structure; function; bioinformatics

[J SUN Yat-sen Univ(Med Sci), 2008, 29(5): 575-580]

收稿日期: 2008-03-31

基金项目: 贵州省科技攻关项目 [黔科合 NY 字(2006)3037]

作者简介: 黄 艳(1977-), 女, 湖南衡阳人, 博士研究生; 黄 江, 通讯作者, 副教授, E-mail: mmm_hj@126.com

亚洲带绦虫(*Taenia asiatica*, *T.a.*)广泛分布于东南亚,包括我国西南地区及台湾,韩国、泰国、印尼和菲律宾等地^[1-3]。人们通过食生或半生含有亚洲带绦虫囊尾蚴的猪、或野猪的内脏,特别是肝脏而感染,对劳动生产力和畜产品破坏极大。亚洲带绦虫成虫形态与牛带绦虫成虫相似,但其幼虫却与猪带绦虫的囊尾蚴相似。亚洲带绦虫成虫与牛带绦虫成虫的形态极为相似,人们长期以来把亚洲带绦虫误认为是牛带绦虫。上世纪 80 年代以来人们对其形态学、流行分布、中间宿主及实验动物感染、遗传学进行了研究,但大部分工作仍局限在细胞水平^[4]。本课题组构建了亚洲带绦虫成虫的 cDNA 质粒文库,获得了大量的 Unigene,在这些工作的基础上开展了对亚洲带绦虫基因组及蛋白质组学的研究,以期从分子水平寻求 3 种带绦虫的起源、演化和彼此间的亲缘关系及宿主选择性的形成等问题的答案。本文分析的烯醇酶(enolase, ENO)是进行这方面研究中感兴趣的分子之一。

1 材料与方 法

1.1 材 料

亚洲带绦虫成虫全长 cDNA 质粒文库,由上海联合基因公司构建。大规模测序得到多个表达序列标签(EST),Washington University BLAST(WU-BLASTX)方法归并 EST 获得 UniGene^[5],由本课题组与该公司合作完成。编码亚洲带绦虫烯醇酶(*T.a.* ENO)基因的文库质粒编号为 HC1-G6。其他寄生虫及其他物种的 ENO 氨基酸序列源自 GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>): 肝片形吸虫烯醇酶基因(*Fasciola hepatica* ENO, 登录号 AAA57450),棘口吸虫烯醇酶基因(*Echinostoma caproni* ENO, 登录号 ABI26619),秀丽隐杆线虫烯醇酶基因(*Caenorhabditis elegans* ENO, 登录号 CAH10783),布氏锥虫烯醇酶基因(*Trypanosoma brucei* ENO, 登录号 EAN77714),人烯醇酶基因(*Homo sapiens* ENO1, 登录号 AAY43128; *Homo sapiens* ENO2, 登录号 AAH02745; *Homo sapiens* ENO3 登录号 AH17249),褐鼠(*Rattus norvegicus* ENO 登录号 AAH83566),牛烯醇酶基因(*Bos taurus* ENO 登录号 AAI02989),野猪烯醇酶基因(*Sus scrofa* ENO 登录号 ABC75829)。

1.2 方 法

1.2.1 *T.a.* ENO 基因的识别 通过美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI, <http://www.ncbi.nlm.nih.gov/>)网站的基本局部比对搜索工具(basic local alignment search tool, BlastX, <http://www.ncbi.nlm.nih.gov/BLAST/>)程序^[6],将文库质粒编号为 HC1-G6 的插入序列与 GenBank 中的序列进行比对,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 rpsblast 分析其保守功能域。

1.2.2 *T.a.* ENO 核酸和氨基酸序列分析 综合性蛋白核酸分析工具包(vector NTI suite)中的 ORF Finder 确定其完整的编码序列(complete coding sequence, cds),然后用 Translation 程序推导并输出氨基酸序列。AlignX 对 *T.a.* ENO 与 GenBank 中其他物种的同源蛋白氨基酸序列进行比对分析,构建分子进化树。

1.2.3 *T.a.* ENO 蛋白理化性质及结构分析 通过瑞士生物信息学研究所的蛋白分析专家系统(Expert Protein Analysis System, ExPASy, <http://ca.expasy.org/>)所提供的蛋白组学和序列分析工具,对目的基因及其产物进行生物信息学分析。预测 *T.a.* ENO 的理化性质,如分子量、等电点、氨基酸组成、摩尔消光系数、重组产物在细菌、酵母和哺乳动物细胞中的半衰期、在溶液中的稳定性等。预测 *T.a.* ENO 一级结构中糖基化、脂酰化、磷酸化、硫酸化等修饰位点、亚细胞定位。预测氨基酸序列的跨膜区和拓扑结构以及二级结构、分子的亲水性、溶液中的分子形态等,通过二级结构比对和折叠,对蛋白质的空间构象建模。

1.2.4 *T.a.* ENO 的亲水性分析 Pcgene 软件分析绘制氨基酸亲水性分布图,确定强亲水性的线性表位位置。

2 结 果

2.1 文库质粒编号为 HC1-G6 插入序列的 Blastx 分析

该基因是烯醇酶的同源基因,与 GenBank 中棘口吸虫(*Echinostomatidae caproni*)的烯醇酶同源性高达 78%。该克隆基因的 5' 端序列长于棘口吸

虫烯醇酶的完整编码序列, 所以该基因应该是亚洲带绦虫烯醇酶的全长基因序列, 其最大的 ORF

就是其完整的编码区(图 1)。用 rpsblast 分析发现有完整的烯醇酶的保守结构域(图 2)。

```

1      GGGTTCGTGACTCCTC GAGGTCGACGGTATC GATAAGCTTGATATC GAATTCGTATCGGCC ATTACGGCCGGGGGA CAGTGTGGTGGGTCT
91     TTTGTTTCATCTGGT GTCGCTGAGCTCGAA CTAAGCTCGAATTCAT GCTTCTACTTTTATC ACGCAGAGGCTGTGT ATCTAACTAGAAGTT
      M S I Q K I H A R Q I F D S R G N P T V E V
181    TTCATTCATTGAGGC TTCAACGGAATGTC ATCCAAAAGATTCAT GCTCGTCAGATCTTC GACAGTCGTGGAAAT CCCACTGTGAGGTT
      D L T T A K G M F R A A V P S G A S T G V H E A V E L R D G
271    GATTAACCACCGCC AAGGGGATGTTTCG GCTGCCGTCCCTCT GGTGCTCCACTGCT GTACACGAAGCGTT GAACTCCCGATGCT
      D K N A Y M G K G V L N A V K N V N E V I A P A L L K E K F
361    GACAAAAATGCCTAC ATGGCAAGGGTGT CTCAATGCCCTCAAG AACGTCAATGAAGTC ATCGCTCCTGCTTG CTAAGGAGAAATTC
      I V T D Q E K I D E F M I K L D G S P N K G K L G A N A I L
451    ATCGTACCACCGCAG GAGAAGATCGATGAG TTCATGATCAAGCTG GATGGATCACCAAC AAAGCAAGCTGGGA GCAAACGCCATCTTG
      G V S L A V C K A G A A E K G V P L Y R H I A D L A G N K D
541    GGTGTTTCTCTGGCG GTCTGCAAGGCTGGA GTCGCCGAGAAGGTT GTTCCCCTCTACCG CACATTGCTGACTTG GCTGTAACAAGGAC
      V V L P V P S F N V L N G G S H A G N K L A M Q E F M I L P
631    GTCGTCCTTCTGTG CCCTCATTCAACGTG CTCAATGGTGGCAGC CACGCTGGGAACAAG CTGGCTATGCAGGAG TTCATGATTCTCCC
      T G A K N F T E A M K M G T E V Y H H L K S V I K G K Y G L
721    ACCGGTCCCAAGAAC TTCACAGAAGCCATG AAGATGGCCACCGAG GTATACCATCACCTC AAGTCCGTTATCAAG GGCAAAATACGGTCTA
      D A C N V G D E G G F A P N I Q E H M E G L E L L K T A I D
811    GACGCCGTAAAGCTT GCGCATGAGGGTGGC TTTGCTCCTAACATC CAGGAACATATGGAG GGTCTTGAATTTCTC AAGACCCCAATTGAC
      R A G Y T G K V K I G M D V A A S E F Y Q N G K Y N L D F K
901    AGGCAAGGTACACC GGCAAGGTCAAGATC GGCATGGATGTGGCC GCTTCTGAATTCTAC CAAAACGGCAAGTAC AATCTGGACTTCAAG
      N P A A V A S S I V P G S K L A E I Y L E M L S K Y P I V S
991    AATCCCGCGCGGTT GCGTCTTCCATCGTT CCTGGATCGAAGCTG GCTGAGATCTACTTG GAGATGCTTTCAAA TACCCTATTGTCTCC
      I E D P F D Q D D W A A W T A F N A K A G I Q I V G D D L T
1081   ATTGAGGATCCCTTC GATCAAGATGACTGG GCCGCTGGACTGCC TTAACGCCAAGGCT GGGATTGAGATTGTT GGGGACGATTGACT
      V T N P E R V Q Q A I D K K A C N A L L L K V N Q I G S V T
1171   GTGCAAAACCCGAG CGTGTGCAGCAAGCT ATCGAACAAGAAGCT TGCAACGCCCTTCTG CTGAAGGTGAATCAG ATCCGGTCCGGTGACG
      E S I K A C K M S R A A G W G V M V S H R S G E T E D S T I
1261   GAGTCTATCAAGGCT TGCAAGATGCTCGC GCTGCTGTTGGGGC GTAATGGTGTGCGAC CGCTCAGGTGAAACG GAGGACTCCACAATT
      A D I V V G L R T G Q I K T G A P C R S E R L A K Y N Q L L
1351   GCTGACATCGTTGTC GGTCTCCGACCGGA CAGATCAAGACTGGT GCGCATGTGCGTCA GAGCGTCTAGCTAAG TACAACCAGCTCTCG
      R I E E E L G F K A V Y A G E H F R N P L *
1441   CGTATTGAGGAGAA CTTGGCTTTAAGGCC GTTACGCTGGCGAG GTTTCCGCAATCCT TTGTAATCACTGTTT ATGTACCTTGTGCTG
1531   CTGGCGGATTTGTC TAGATTTAACTGCGT TTGTTATTACAACCT CCTCGACAGCGGTAG TGATTGCGCAATTT TTTAAATTAATTTT
1621   TTGGAAGTTACAAA AAAAAAAAAAAAAA AAAAAAAAAAACATGT CGGCCGCTCGGCT ATGTGCGCGCCAC CGCGGTGGAGCTCCA
1711   GCTTTGTTCCCTTT AGTGAGGGTAAT

```

图 1 T.a.ENO 全长 cDNA 序列及其 ORF 编码的氨基酸序列图

Fig.1 Full-length cDNA sequence of T.a.ENO and the amino acid sequence encoded by ORF

The full length sequence of T.a.ENO contains 1 737 base pairs(bp). The complete coding sequence(CDS)is between 205-1 503 bp which encodes a putative protein of 433 amino acids. There are untranslatable sequences at the 5' and 3' termini.

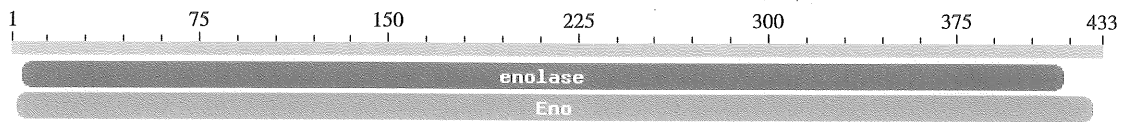


图 2 rpsblast 预测 T.a.ENO 的保守功能域图

Fig.2 Prediction of conserved domain of T.a.ENO by rpsblast

The number stands for the ordinal number of amino acid

2.2 T.a.ENO 蛋白质的理化性质

T.a.ENO 的相对分子量理论值和等电点分别为 46 653.5 和 6.77。含有 5 个半胱氨酸, 预测这 5 个半胱氨酸之间形成二硫键的可能性较小, 该蛋白在水溶液中 280 nm 处的摩尔消光系数为 33 140 mol·L⁻¹·cm⁻¹; 蛋白浓度为 1 g/L 时, 半胱氨酸未形成二硫键时吸光系数(Abs)为 0.708。若其成熟肽 N 端为蛋氨酸时, 在哺乳动物网状红细胞体外表达的半衰期为 30 h, 在酵母和大肠埃希菌中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 32.33, 在溶液中性质稳定。疏水

指数为 89.47, 疏水性较高。

2.3 T.a.ENO 翻译后修饰、亚细胞定位的预测

用 Motif scanning(Motifscan)分析 T.a.ENO 特定定位点结果显示, T.a.ENO 含有 6 个潜在的酪蛋白激酶 II (CK2)磷酸化位点, 5 个潜在的蛋白激酶 C(PKC)磷酸化位点, 2 个酪氨酸激酶磷酸化位点, 10 个潜在的 N-肉豆蔻酰位点, 1 个潜在的天冬氨酸糖基化位点。T.a.ENO 没有分泌信号肽序列和质体以及线粒体定位序列。

2.4 T.a.ENO 的拓扑结构、二级结构和亲水性特征

用 Predict protein 预测结果如图 3 所示。Htm

预测该蛋白是一个膜蛋白,有 1 个跨膜区(M),N 端位于膜内(i),C 端位于膜外(o)。Sec 预测 α 螺

旋(H)、β 折叠(E)和无规卷曲(空白部分)的比例分别是 40.42 : 21.15 : 38.43。

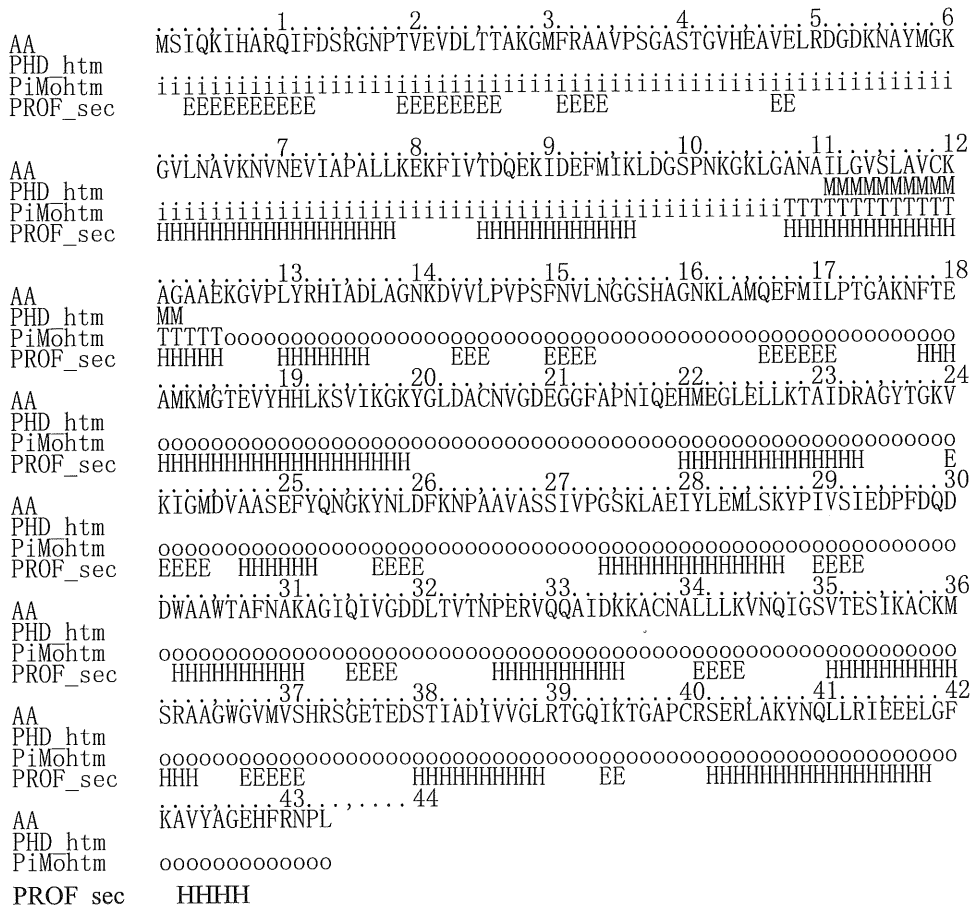


图 3 T.a.ENO 的拓扑结构、二级结构图

Fig.3 The topology, secondary structure of T.a.ENO by Predictprotein

AA: amino acid; PHD htm: neural network prediction of transmembrane helices; PROF sec: secondary structure; PiMohtm: prediction of membrane topology

2.5 T.a.ENO 的亲水性分析

利用 Pcgene 软件包预测 T.a.ENO 氨基酸的亲水性分布(图 4)。推导其线性抗原决定簇的位置分别是: ①Ah = 2.03 From 50 to 55: Arg-Asp-Gly-Asp-Lys-Asn; ②Ah = 1.73 From 86 to 92: Asp-Gln-Glu-Lys-Ile-Asp-Glu; ③Ah = 1.48 From 373 to 380: Arg-Ser-Gly-Glu-Thr-Glu-Asp-Ser (Ah, average hydrophilicity, 平均亲水性)。

根据拓扑图,其中①和②序列位于膜内,③序列位于膜外,是另一个高亲水性的线性表位。该序列位于膜外区域,而且③的序列是 ENO 蛋白质特征指纹序列。

2.6 T.a.ENO 的三维结构图和酶关键氨基酸的位置 利用同源建模法服务器 (SWISS-MODEL)将

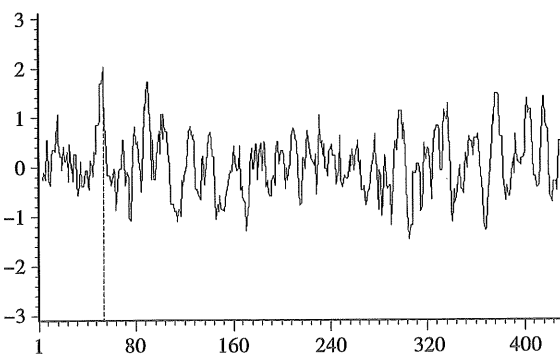


图 4 T.a.ENO 的亲水性分布图

Fig.4 Hydrophilicity profile of T.a.ENO

The ordinates and abscissa stand for the Ah and amino acids respectively. The vertical line stands for the amino acid site with highest Ah.

是一个膜蛋白,没有质体及线粒体的定位序列,很多研究发现烯醇酶定位于细菌、真菌、原虫^[9]、蠕虫^[10]的表面,也可通过免疫定位来确定 T.a.ENO 是否位于亚洲带绦虫的表膜,从而进一步研究其在致病及免疫方面的作用。

绦虫生理代谢所需的能量来自糖酵解,虽然烯醇酶不是糖酵解途径的关键酶,但它是一种多功能蛋白,是一个嗜神经因子。它能与细胞骨架蛋白和多聚核苷酸结合、还具有热休克蛋白的功能^[11-13]。此外,它还是纤溶酶原及层粘连蛋白的受体^[14],在寄生虫侵袭宿主组织过程中发挥作用^[9],它在感染和免疫中还作为抗体作用的靶分子^[15]。利用生物信息学方法分析得到的结果将有助于全面了解 T.a.ENO 的功能。

参考文献:

- [1] Fan PC, Lin CY, Kosman ML, et al. Experimental infection of Indonesia Taenia (Somosir strain) in domestic animals [J]. *Int J Parasitol*, 1989, 19(7):809-812.
- [2] Fan PC, Chung WC, Lin CY, et al. Experimental infection of Thailand Taenia (Chiengma strain) in domestic animals [J]. *Int J Parasitol*, 1990, 20(1):121-123.
- [3] Fan PC, Lin CY, Chung WC, et al. Experimental infection of philippine Taenia in domestic animals [J]. *Int J Parasitol*, 1992, 22(2):235-238.
- [4] 王正蓉. 亚洲牛带绦虫分类学研究进展[J]. *贵阳医学院学报*, 2001, 26(1):43-45.
- [5] 黄江,胡旭初,包怀恩,等. 亚洲带绦虫成虫全长 cDNA 质粒文库的构建及 EST 测序[J]. *热带医学杂志*, 2007, 7(2):116-118.
- [6] 吴忠道,余新炳,徐劲,等. 日本血吸虫(大陆株)成虫基因表达谱的研究[J]. *中山医科大学学报*, 2002, 23(6):401-404.
- [7] Zhang E, Brewer JM, Minor W, et al. Mechanism of enolase: the crystal structure of asymmetric dimer enolase-2-phospho-D-glycerate/enolase-phosphoenolpyruvate at 2.0 Å resolution [J]. *Biochemistry*, 1997, 36(41):12526-12534.
- [8] Babitt PC, Hasson MS, Wedekind JE, et al. The enolase superfamily: a general strategy for enzymecatalyzed abstraction of the α-protons of carboxylic acids [J]. *Biochemistry*, 1996, 35(51):16401-16489.
- [9] Pancholi V. Multifunctional alpha-enolase: its role in diseases [J]. *Cell Mol Life Sci*, 2001, 58(7):902-920.
- [10] Jolodar A, Fischer P, Bergmann S, et al. Molecular cloning of an alpha-enolase from the human filarial parasite *Onchocerca volvulus* that binds human plasminogen [J]. *Biochim Biophys Acta*, 2003, 1627(2-3):111-120.
- [11] Williams LA, Ding L, Horwitz J, et al. Tau-crystallin from the turtle lens: purification and partial characterization [J]. *Exp Eye Res*, 1985, 40(5):741-749.
- [12] Takei N, Kondo J, Nagaie K, et al. Neuronal survival factor from bovine brain is identical to neuron-specific enolase [J]. *J Neurochem*, 1991, 57(4):1178-1184.
- [13] al-Giery AG, Brewer JM. Characterization of the interaction of yeast enolase with polynucleotides [J]. *Biochim Biophys Acta*, 1992, 1159(2):134-140.
- [14] Carneiro CR, Postol E, Nomizo R, et al. Identification of enolase as a laminin-binding protein on the surface of *Staphylococcus aureus* [J]. *Microbes Infect*, 2004, 6(6):604-608.
- [15] Terrier B, Degand N, Guilpain P, et al. Alpha-enolase: a target of antibodies in infectious and autoimmune diseases [J]. *Autoimmun Rev*, 2007, 6(3):176-182.

(编辑 孙慧兰)

《中山大学学报(医学科学版)》征订启事

中山大学学报(医学科学版)(原名为中山医科大学学报)是国家教育部主管、中山大学主办的综合性医学学术期刊,1980年创刊,双月刊,全国公开发行。本刊主要反映国内外医学研究的最新成果,是科研人员交流医学学术思想的重要平台,主要发表基础医学研究、临床医学研究和公共卫生事业等方面的论文。中山大学学报(医学科学版)被引频次、影响因子连续多年位于国内同类刊物前列。依照清华大学2007年中国学术期刊计量指标均值分类统计报告的数据,本刊被引频次817,影响因子0.721,在同类刊物中排名第二。

本刊为双月刊,刊号:ISSN 1672-3554 CN 44-1575/R,每期售价10元,全年6期共60元,欢迎单位和个人向当地邮局订购(邮发代号:46-141)。电话:020-87331643, Fax:020-87330827。

网址: <http://xuebao.sysu.edu.cn/yx>

E-mail: XBmed@mail.sysu.edu.cn