

基础研究

日本血吸虫精氨酸酶编码基因 cDNA 的分离和序列分析

吴忠道, 郑亦男, 徐 劲, 余新炳

(中山医科大学寄生虫学教研室, 广东 广州 510089)

摘要: 【目的】分离和鉴定日本血吸虫(*Schistosoma japonicum*, *S. j*)新基因。【方法】应用免疫学方法筛选 *S. j* cDNA 文库获得阳性克隆; 通过 PCR 直接序列测定技术, 表达序列标签(EST)同源性检索对阳性克隆插入片段进行鉴定; 采用亚克隆技术及生物信息学等技术, 对获得的日本血吸虫精氨酸酶编码基因序列进行结构与功能的分析。【结果】获得了一个含 1 061 bp 外源插入片段的阳性克隆, 其 cDNA 序列含有一个 840 bp 的阅读框, 编码 279 个氨基酸, 属精氨酸酶家族, 与国际蛋白质库中的酵母、蛙、人和大鼠精氨酸酶序列的同源性分别为 44%, 50%, 51%, 53% 和 54%。【结论】获得了编码日本血吸虫精氨酸酶编码基因全长 cDNA, 为日本血吸虫精氨酸酶基因功能的研究奠定了基础。

关键词: 日本血吸虫; 精氨酸酶; 序列分析, cDNA

中图分类号: R383.2⁺ 文献标识码: A 文章编号: 1000-257X(2001)02-0085-04Isolation and Sequence Analysis of the cDNA Encoding *Schistosoma japonicum* Arginase

WU Zhong-dao, ZHENG Yi-nan, XU Jing, YU Xin-bing

(Department of Parasitology, Sun Yat-sen University of Medical Sciences, Guangzhou 510089, China)

Abstract: 【Objectives】To isolate and identify novel genes of *Schistosoma japonicum* (*S. j*) from cDNA library. 【Methods】cDNA clones were screened from *S. j* cDNA library by immunological method, the inserts of positive cDNA clones were identified by PCR directed sequencing and homologous searching of EST, the cDNA sequence encoding *S. j* arginase was analyzed by the subcloning and bioinformatics. 【Results】The positive cDNA clone with 1 061 bp insert was isolated from *S. j* cDNA library, which sequence of the insert included a ORF with 840 bp, and translated into 279 amino acid residues. The deduced protein was similar to the arginase family, the homology at amino acid sequence level was 44%, 50%, 51%, 53% and 54% with *schizosaccharomyces pombe* (fission yeast) arginase, *xenopus laevis* (african clawed frog) arginase, *homo sapiens* (human) arginase ii precursor and *rattus norvegicus* (rat) arginase ii precursor respectively. 【Conclusion】The full-length cDNA sequence encoding *Schistosoma japonicum* arginase was cloned, which gave the basis for further studying of *Schistosoma japonicum* arginase.

Key words: *Schistosoma japonicum*; arginase; sequence analysis, cDNA

血吸虫侵入宿主体内后迅速生长, 成虫期则大量排卵, 其蛋白质合成十分旺盛。因此, 研究血吸虫氨基酸的代谢途径和机制, 探索阻断其代谢途径的方法, 对于预防和治疗血吸虫病都具有重要的价值。精氨酸是血吸虫虫体的主要氨基酸组分之一, 尽管目前对血吸虫精氨酸的代谢还研究的不多, 但

已有的研究发现, 日本血吸虫具有很高的精氨酸酶活力, 提示精氨酸酶在血吸虫氨基酸代谢中起作重要的作用^[1]。我们采用血吸虫感染者血清, 对日本血吸虫成虫 cDNA 文库进行免疫学筛选, 获得了一个含 1 061 bp 外源插入片段的阳性克隆。经序列测定和序列分析, 初步确认该插入片段为日本血吸

收稿日期: 2000-07-04

基金项目: 中山医科大学“211”重点学科建设基金资助项目(98169)

作者简介: 吴忠道(1962-), 男, 江西南昌人, 博士, 副教授, 硕士生导师; 余新炳(1953-), 男, 安徽宿松人, 博士, 教授, 博士生导师

虫精氨酸酶编码序列。

1 材料与方 法

1.1 cDNA 文库及试剂

日本血吸虫(中国大陆株)成虫 cDNA 文库由南京医科大学陈淑贞教授等构建^[3]。大肠杆菌 *E. coli* Y1090 和噬菌体 λ gt11 均购自 Promega 公司, *Taq* 酶(5×10^6 U/L)和 2 mmol/L dNTP 购自加拿大真达公司, HRP-羊抗人 IgG 购自广州华美公司。

1.2 抗体探针的制备

在日本血吸虫重度流行区, 采用静脉抽血方法收集日本血吸虫感染者(粪检虫卵阳性者)血清, 分装后低温保存。血清经 λ gt11/Y1089 裂解物反复吸附后制备成筛库用抗体探针^[3]。

1.3 阳性克隆的筛选及 PCR 鉴定

按吴海玮等方法对 cDNA 文库进行免疫学筛选和阳性克隆噬菌体 DNA 的扩增^[3]。

选取 λ gt11 Sfi-Not 载体插入位点上下游邻近区核苷酸序列, 设计上下列引物一对并由上海生工公司合成。上游引物(P1)5'-GGT GGC GAC GAC TCC TGG AGC CCG-3'; 下游引物(P2)5'-TTG ACA CCA GGC CAA CTG GTA ATG-3'。

1.4 阳性克隆 EST 的获得及同源性检索

扩增 150~200 μ L 单个阳性重组克隆的 cDNA 插入片段, 采用 PCR 产物直接序列测定技术, 对插入片段的 5' 端进行部分序列测定^[4], 获得阳性克隆表达序列标签(EST)。通过国际互联网将测得的阳性克隆 EST 序列与 EMBL 血吸虫基因库中的核苷酸序列进行同源性搜寻, 该基因库网址为: <http://www2.ebi.ac.uk/blast2/parasites.html>, 所用程序为 Blastn。

1.5 阳性克隆插入片段的通测及序列分析

将阳性克隆插入片段 PCR 产物与质粒 pMD18-T 连接, 构建测序质粒, 并进行全序列的测定。本项工作由大连宝生物公司(TakaRa)协助完成。

利用国际互联网上的有关生物生物信息网站如 Webgen (<http://www.itba.mi.cnr.it/web-gene/>) 和 ExPASy (<http://www.expasy.ch/tools/>) 进行序列的结构分析, 分析内容包括核苷酸序列中开放读框(ORF)的寻找, 编码氨基酸的推导, 推导出的氨基酸与 Swiss-Prot 蛋白质库中序列

的同源性比较等。

2 结 果

2.1 阳性克隆的筛选及 EST 的产生

应用预处理过的混合血清, 对 *S. j* cDNA 文库大约 2.5×10^4 个噬菌体克隆进行了筛选。初筛获得克隆 12 个, 然后再挑取单个初筛阳性噬菌斑进行复筛, 最后确定了 5 个阳性克隆。5 个阳性克隆均扩增出单一片段, 大小从 400~1 100 bp 不等, 而野生型 λ gt11 未出现扩增片段, 表明经过复筛确定的阳性克隆均含有外源性 cDNA 插入片段。其中阳性克隆 JAYW0p22.GY 的 cDNA 插入片段约为 1.1 kb。

采用 PCR 产物直接测序法对筛选到阳性克隆的 cDNA 插入片段 PCR 产物进行了 5' 端部分序列测定(长度为 200~600 bp), 获得了相应的 EST(此部分内容另文发表)。其中阳性克隆 JAYW0p22.GY 的 EST 核苷酸序列长度为 578 bp, 其 GenBank 进入号为 AW160123。

2.2 EST 同源性检索

采用 WU-Blastn 程序进行同源性检索, 结果发现, AW160123 序列与 EMBL 中 AA294801 序列高度同源, 其一致性为 98%(277/280), AA294801 为国际基因库中已登录的一个日本血吸虫表达基因 EST, 序列长度 311 bp, 为精氨酸酶同源序列(GenBank 对 AA294801 的注释为: JA00A108. QA3 Adult SjC 7/94 Schistosoma japonicum cDNA clone SJADA108 5' similar to Arginase, mRNA sequence, Length 311 bp)。

2.3 插入片段的通测及序列分析

将阳性克隆 JAYW0p22.GY 插入片段 PCR 产物亚克隆入质粒 pMD18-T, 并测定其全长 cDNA 序列。该 cDNA 序列长 1 061 bp, 含有 6 个可能的开放性阅读框(ORF), 其中最大的一个 ORF 为 840 bp, 起始密码子为 ATG, 终止密码子为 TGA, 编码 279 个氨基酸, 3' 末端的非编码区长度为 164 bp, Poly(A)信号在 999~1 006 bp, Poly(A)尾出现在 1 052 bp 处(图 1)。

2.4 推导氨基酸序列的结构分析

开放读框内的 279 个氨基酸组成的多肽分子为 30.527 ku, 等电点 PI 为 6.36, 氨基酸平均亲水性值为 -0.191。采用 InterPro 程序对 SWISS-

```

GGA GAT GTC ATA CCT GTA GAA CTA AGT GAA TCA GAG GAT CCA GAA CTT TGT GGA 54
ATG AAA TGG TCA AGA AGT TTC ACA CAG ACC ACT TTG AAA ATA GCT GAC CGT GTA 108
M K W S R S F T Q T T L K I A D R V
GAA CAG TTG GTA AAA GGG TCA AAT AAA CAT AGT ATT GAA TCC AGT AAT TCG AAA 162
E Q L V K G S N K H S I E S S N S K
CCA TCA CCA TTA GTA ATT GTT GGC GGT GAT CAT AGT ATG GCG ACT GGA ACT ATA 216
P S P L V I V G G D H S M A T G T I
CTT GGA CAT GCT AGA GCC AAA CCA GAT GTG TGC ATT ATA TGG GTT GAT GCT CAT 270
L G H A R A K P D V C I J W V D A H
GGT GAT ATA AAT ACA CCA CCA AAC TCA ACT ACT GGA AAT ATA CAT GGA ATG CCA 324
G D I N T P P N S T T G N I H G M P
TTA AGT TTT CTA GTA AAA GAA CTA CAA GAT CAA ATT CCA TGG TTG GAT GAC TTT 378
L S F L V K E L Q D Q I P W L D D F
CAT AGT ATA AAA CCA TOT CTG GAT GCC AGC AAT CTT GTT TAC ATT GGT TTA CGG 432
H S I K P C L D A S N L L V Y I G L R
GAT TTA GAC GTT TAT GAA ACA CCG GAT ATA AGA AAG CAT GGT ATA GCT TAT TTT 486
D L D V Y E T R D I R K H G I A Y F
ACA AGT CTT GAC GTT GAT CGA ATG GGA ATG GAA GCC GTC ATT AAA GAA CCA TTA 540
T M L D V D R M G M E A V I K E A L
CAA GCT GTG AAT CCG AGA TTA GAG AAA CCT ATT CAT TTA AGT TTT GAT ATT GAT 594
Q A V N P R L E K P I H L S F D I D
GCA TTG GAT CCT TCA ATT GCT CCA AGT ACT GGT ACT GCT GYT CCA GGT GGT TTA 648
A I D P S I A P S T G T A V P G G L
ACA TTA CGT GAA GGT TTA AGA ATA TGT GAA GAA ATT TCA GCT ACA GGA AAA CTT 702
T L R E G L R I C E E I S A T G K L
TCT ATT GTT GAA TTG GCT GAA TTA AAT CCT TTG TTA GGA TCT AAA GAA GAT GTT 756
S I V E L A E L N P L L S K E D V
GAA AAA ACG CAA TCA TCT GCT GTG CAC ATT TTA AGG GCA TCG TTA GGA CAT TGT 810
E K T Q S S A V H I L R A S L G H C
CGT TCA GGT CAA TTA CCG ATG AAA GTT AAC AAT TCA ACC ACT AAT AGT ATT GTT 864
R S G Q L P M K V N N S T T N S I V
AGA CAA GCT GAA CGT ATA CAG ATA AAG TGA TAA TTATTCCTTCTCAATGCAATTA 921
R Q A E R I Q I K *
ATGATTAATTCCTTATAATAATAATAATCAATGATCAATATGATTAATAATAATGTTACTAAC 986
AAAAATAATGTAATAATACAAATGATTTAAAGTATTTTCTAAATATACTACTATTAATTAATTTG 1051
AAAA AAAAAA 1061

```

图 1 日本血吸虫表达基因 cDNA 克隆 JAYW0P22. GY 插入片段的核苷酸序列及其推导的氨基酸序列

Fig. 1 cDNA sequence and its deduced amino acid sequence of the clone JAYW0P22. GY from *Schistosoma japonicum* cDNA library

The amino acids were indicated under the nucleotide sequences; signature sequence of the amino acids was underlined by two lines; * shows stop codon. Poly(A) signal was underlined

PROT 蛋白质数据库进行检索, 鉴定出该氨基酸序列属精氨酸酶家族, 其序列中含有 3 段精氨酸家族特征性序列, 它们分别是“VGGDHSMATGTILG”、“IIWVDAHGD”和“SFDIDLPSIAPSTGTAVPGG”。将该氨基酸序列与 SWISS-PROT 中的已知蛋白质序列进行比较, 发现与大鼠 (*rattus norvegicus*)、人 (*homo sapiens*)、蛙 (*xenopus laevis*) 及酵母 (*schizosaccharomyces pombe*) 精氨酸酶氨基酸序列的同源性较高, 分别为 54% (118/216), 53% (115/216), 51% (111/216), 50% (110/216) 和 44% (99/224), 见图 2。

3 讨论

利用生物信息学 (bioinformatics) 技术研究新基因的功能和预测基因表达产物的性质已成为当基因组研究的主要技术策略^{5, 6}。如加拿大科学家

```

1 —MKWSRS—————FTQT TLKI ADRVEQLVKGSNKHSI E—SSNSKPSPL
2 MFLRSSYSRLRHQQI PCA—L TRSVSVAVVGFPSRGGKKKGVVEYGPAA I REAGLKLRLSMLGCHI KDFGDLFSF
3 MSLRGSLSRLLQTRVHSI—L KKS VHSVAVI GAFPSSGOKRKGVEHGPAA I REAGLMLRSLGCHIKDFGDLFSF
4 MS I RSNFVRLLLKQVSI I LK QKKCSHSVAVI GAFPSSGOKRKGVEHGPAA I RSAGL I ERLSNLGGKVCDFGDLHF
5 —MSPHK I PEVHRH I MSS—RYMEGNVSI I NMFPSSGGQPKDGAELAPEM I EAAGLPEDLERLGSYVNVQNPKF

*****1*****
1 —————V I VGGDHSMATGT I LGHARAKPDVCI I IWDADHD
2 TNVP—KDDPYNNLVVYPRVSG I ANOELAEVSVRAVSGYSCVTLGGDHSLA I GTI I SGHARHPDLCV I MVDAHAD
3 TPVP—KDDLYNNL I VNPFRVGLANOEAEVSVRAVSDGYSCVTLGGDHSLA I GTI I SGHARHPDLCV I MVDAHAD
4 SQVP—NDEL VNS I VKHPRITVGLACKVLAEEVSKAVGAGHT CVTLGGDHSLAFGS I TGHAAQQDPDLCV I MVDAHAD
5 KSRPLKEGPNQALMKNPLTVSNVTRQVRN I VQOELKQR I AVN I GGDHSLA I GTVEGVQVYDDAGLVN—DAHAD

*****2*****234*****
1 INTPPNNTGNI I HGMP L SFLYKELQQQ I PMLDDFHS I KPCLDASNLVY I GLRDLVYVETRD I RKHG I AYF TMLDV
2 INTPLTVSGNI I HGQPLSFL I RELQDQVPLPGFSWI I KPCLSPMLVY I GLRDPVEAEHF I LKSFDI QYFSMRDI
3 INTPLTSSGNLHGQPVYFLLRELQDQVPLPGFSWI I KPCLSSAS I VY I GLRDPVPEHF I LKNDY I SYFSMRDI
4 INTPLTTPSGNLHGQPVYFLLRELQDQVYFP I PGFSWAKPCLSKSD I VY I GLRDLPAEAF I LKNDY I SYFSMRDI
5 INTDPSSEKSNLHGQPLSFL—GYAEPLPE—EFANTRRV I EERRLAF I GLRDLDPMEARFLERS I ATAYMHV

*****3*****
1 DRMGAEV I KEALGAVNPREKPI I HLSFDI DALDPS I APSTGTAVPGLTLREGLR I GEE I SATGKLS I VELAEI
2 DRLGI QKVMQTFDRL I GKRRRP I HLSFDI I DAFDPKLPATGTPVVGGLTYREGLY I TEE I HSTGLLSALDLVEV
3 DRLGI QKVMQTFDRL I GKRRRP I HLSFDI I DAFDPLAPATGTPVVGGLTYREGLY I AEE I HNTGLLSALDLVEV
4 DCMG I KKVMEKTFDGLRRDRP I HLSFDI I DAFDPLAPATGTPV I GGLTYREGLY I TEE I HNTGLMSALDLVEV
5 DKYGI ARVVEALEHI I NPGRRRP I HLSFDVDAADCP I VAPATGTRVPGLTFREAMY I CESVAETGSLVAVDVMEV

*****23451 *****15324
1 NPLLQ—SKEDVEKTSQSAVHI I LRASLGHCRSGDL ————PW—KVNMTTNS I VROAER I Q I K
2 NPHLATSEEEAKATASLAVDVI I ASSFGOTREGGH ————JAYDHLPTPSSPHESEKEECVRI—
3 NPQLATSEEEAKTANLAVDVI I ASSFGOTREGGH ————IYDQLPTPSSPDESEMOARVRI—
4 NPVLAATSEEVKATANLAVDVI I ASOFGOTREGAHTRADTI I IVDLPTPSTSYESDNEQVRI—
5 NPLLQ—NKEEAKTIVDLARS I VRTCLGQ —————TLL—————

```

图 2 推导的日本血吸虫精氨酸序列与其它生物精氨酸序列的同源性比较

Fig. 2 Homology analysis of Deduced amino sequence of *Schistosoma japonicum* arginase and other arginase sequences

*: Conserved amino; First line: Deduced amino sequence of *Schistosoma japonicum* arginase; Second line: ARG2-RAT 008701 *rattus norvegicus* (rat), arginase ii precursor; Third line: ARG2-HUMAN P78540 *homo sapiens* (human), arginase ii precursor; Fourth line ARG2-XENLA Q91554 *xenopus laevis* (afican clawed frog), arginase; Fifth line: ARG1-SCHPO P37818 *schizosaccharomyces pombe* (fission yeast), arginase

Riordan 等人^[7]在得到蛋白产物之前, 通过对囊性纤维变性过膜传导调控因子 (cystic fibrosis transmembrane conductance regulator) 的核苷酸序列的分析, 准确地预测了该蛋白的结构及其作用模式。

对新获得的核苷酸序列, 通常按以下步骤来进行基因的识别和鉴定。①预测其中可能的编码区; ②搜寻信号序列, 如启动子、翻译起始位点、RNA 剪接信号、终止子等; ③将编码区连为一个完整的序列, 并翻译蛋白质。然后以蛋白质序列为模板序列, 对序列库作同源性检索; ④预测序列的二级和三级结构, 对其可能的功能进行理论预测, 如亲水性、抗原表位等 (如果为 cDNA 序列, 问题则大为简化, 可根据起始密码子直接预测/确定可能的编码区或开放读框)。在此分析基础上, 再采用生物化学实验手段进行基因的鉴定。为此, 利用生物信息学识别基因与应用生物学实验方法鉴定基因, 两

者可相辅相成, 互相促进, 从而加速新基因发现和鉴定。

目前, 生物信息学发展迅速, 国际核苷酸序列如 GenBank/EMBL 和氨基酸序列库如 SWISS-PROT 等定期更新和公布所收集的序列资料, 并可通过国际互联网方便地进行序列检索。同时, 序列分析软件如 GCG 具有强大的功能, 能基本满足对已获得的 DNA 序列结构与功能分析的需要。此外, 各类生物信息学网站为基因序列和氨基酸序列的分析提供了更加有效的技术途径。如著名的 Webgen 和 ExPasy 网站等, 这些网站提供了从核苷酸序列编码读框(ORF)的预测到推导氨基酸一、二、三级结构的理论分析等一系列相关程序。本文主要应用其中的“ORF and protein analysis”, “Hcpoly a”, “Compute PI/MW”, “InterPro Search”, “ScanPileScan”和“MSA”等程序, 并取得了较满意的分析结果。通过对获得的日本血吸虫 cDNA 阳性克隆 JAYW0p22. GY 的插入片段 cDNA 序列的结构分析, 结果显示, 该序列含有一个 840 bp 的阅读框, 编码 279 个氨基酸, 属精氨酸酶家族, 其氨基酸序列与国际蛋白质库中的酵母、蛙、人和大鼠精氨酸酶序列的同源性高达 44%~54%。表明, 我们已获得了编码日本血吸虫精氨酸酶编码基因全长 cDNA, 这为日本血吸虫精氨酸酶基因功能的研

究奠定了基础。目前, 我们正在对该 cDNA 序列的重组表达产物进行实验鉴定。

参考文献:

- [1] 毛守白. 血吸虫生物学与血吸虫病防治[M]. 北京: 人民卫生出版社, 1991. 214~215.
- [2] 陈淑贞, 吴海玮, 张兆松, 等. 一个定向克隆的日本血吸虫基因表达文库[J]. 中国人兽共患病杂志, 1997, 13(6): 23.
- [3] 吴海玮, 陈淑贞, 王荣芝, 等. 日本血吸虫中国大陆株成虫 cDNA 文库的免疫筛选及 PCR 初步鉴定[J]. 中国寄生虫病防治杂志, 1998, 10(1): 7.
- [4] 吴忠道, 余新炳, 郑亦男, 等. 日本血吸虫(中国大陆株)表达基因的分离和 EST 序列测定[J]. 中国人兽共患病杂志, 2000, 16(1): 3.
- [5] 杨焕明. “人类基因组计划”自启动至“后基因组计划”的转折[J]. 生物工程进展, 1996, 16(2): 2.
- [6] Foote S J, Speed T, Handman E. What can bioinformatics do for parasitology research[J]? Parasitology Today, 1998, 14(9): 346.
- [7] Riordan J R. From the gene to the physiopathology: the functions of CFTR protein. Arch Pediatr, 1996, 3 Suppl 1: 108s.

(编辑 张敏瑞)

·新成果·

治疗性角膜移植的系列研究

课题负责人 陈家祺 等

(中山医科大学眼科中心, 广东 广州 510080)

各种原因引起的角膜溃疡或穿孔是我国最严重及最常见的角膜病, 每年患病人数达数十万, 治疗性角膜移植是此类患者治疗的重要措施及复明的唯一方法。本研究自 1969 年开始历时 30 多年, 经过 1 万例以上的角膜移植手术病例观察, 对治疗性角膜移植的若干临床重点问题进行系列研究。通过全身及局部免疫学调查, 发现了蚕蚀性角膜溃疡是一种以体液免疫为主细胞免疫参与的自身免疫性疾病, 修正了国际上关于此病发病机制的传统观点。在国际上率先提出了完全切除角膜及结膜病灶组织后行板层角膜移植及术后应用免疫抑制剂治疗蚕蚀性角膜溃疡的综合治疗方案, 通过国际上最大样本的临床应用, 其治愈率位于国际首位。自行设计出分别用于有晶体眼及无晶体眼的人工角膜, 此人工角膜较国外所设计及生产的人工角膜视野宽、清晰, 在我国率先开展眼球前、后段联合手术治疗角膜白斑伴有视网膜脱离患者, 其手术效果明显优于国外报告。在国际上最系统地提出了根据角膜溃疡或穿孔的原因、位置、大小等综合因素考虑治疗性角膜移植手术的时机、方案、技巧及术后处理原则。这些理论已基本成为我国治疗性角膜移植的经典原则, 被眼科学教科书及各种角膜手术专著引用, 并被全国各大医院普遍采用。该成果于 2000 年获广东省科学技术进步一等奖。

(陈丽芳)