

猪带绦虫 CDC37 基因及其蛋白结构特征的生物信息学分析*

刘玉江¹, 陈晓睿², 黄 江^{3**}, 戴佳琳³, 廖兴江⁴

(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; 3. 贵阳医学院 法医学教研室, 贵州 贵阳 550004; 4. 贵阳医学院 物理学教研室, 贵州 贵阳 550004)

[摘 要] 目的: 分析和预测猪带绦虫细胞分裂周期蛋白 37(Ts CDC37) 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: 利用生物信息学网站所提供的有关基因和蛋白序列和结构功能分析工具, 以及专业的生物信息学软件包, 从猪带绦虫成虫全长 cDNA 质粒文库中识别 CDC37 基因, 对猪带绦虫 CDC37 基因编码的蛋白质进行生物信息学分析。结果: 该基因全长 1 203 bp, 编码 400 个氨基酸, 为全长基因。Genebank 中与日本血吸虫的 cell division side 37 homolog 蛋白的一致性最高, 达 60%。相对分子量理论预测值为 46 802.5 ku, 预测其有 6 个主要的 B 细胞抗原表位, 无亚细胞定位序列。结论: 应用生物信息方法从猪带绦虫成虫 cDNA 文库中筛选出了猪带绦虫 CDC37 基因全长序列并预测得到其结构与功能方面信息。

[关键词] 猪带绦虫; 基因; 蛋白质结构; 序列分析

[中图分类号] R383.3 **[文献标识码]** A **[文章编号]** 1000-2707(2012)04-0344-04

Bioinformatics Analysis on Cell Division Cycle 37 Gene of *Taenia solnium* and Structural Characteristics of Its Coding Protein

LIU Yujiang¹, CHEN Xiaorui², HUANG Jiang³, DAI Jialin³, LIAO Xingjiang⁴

(1. Laboratory of Multimedia Morphology, Guiyang Medical College, Guiyang 550004, Guizhou, China; 2. Department of Parasitology, Guiyang Medical College, Guiyang 550004, Guizhou, China; 3. Department of Forensic, Guiyang Medical College, Guiyang 550004, Guizhou, China; 4. Department of Physics, Guiyang Medical College, Guiyang 550004, Guizhou, China)

[Abstract] **Objective:** To analyze and predict the structural characteristics of cell division cycle 37 (CDC37) gene of *Taenia solnium* (Ts) and its encoding protein, and so as to guide the study on its biological function. **Methods:** Tools provided by bioinformatics webs for analysis on the structure, function and sequences of related gene and protein and bioinformatics software packages were employed to identify TsCDC37 gene from cDNA plasmid library, and bioinformatics analysis on the encoding protein of CDC37 gene was carried out. **Results:** This gene was 1 203 bp in length and was a full length gene coding 400 amino acids. The coding protein had highest coherence (60%) with CDC37 homolog protein of *Schistosoma Japonicum* in GeneBank. Relative molecular weight was 46 802.5 ku. The protein was predicted to have 6 B cell epitopes and have no subcellular location sequence. **Conclusions:** The full length sequence of Ts CDC37 gene is screened from cDNA library of adult *Taenia solnium* by bioinformatics method. The information about the structure and function of CDC37 gene and protein is obtained.

[Key words] *Taenia solnium*; gene; protein structure; sequence analysis

* [基金项目] 国家自然科学基金资助项目 (30760227)

** 通讯作者 E-mail: mmm_hj@gmc.edu.cn

猪带绦虫病和囊尾蚴病分布广泛,其流行广泛,危害严重,是我国重点防治的寄生虫病之一^[1~3]。本课题组构建了猪带绦虫、牛带绦虫以及亚洲带绦虫成虫全长 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猪带绦虫基因组及蛋白质组学的研究^[4]。在分析猪带绦虫细胞分裂周期蛋白 37 基因(Ts CDC37)全长 cDNA 序列的基础上,通过生物信息学方法预测该基因编码的蛋白的结构和功能特性,为实验研究和应用前景分析提供有价值的信息。

1 材料与方法

1.1 材料

猪带绦虫成虫全长 cDNA 质粒文库的构建、EST 大规模测序和 Unigene 分析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。编码猪带绦虫成虫 CDC37 基因文库质粒编号为 tscaa0_000445。其他寄生虫及其他物种 CDC37 氨基酸序列源自 GenBank。

1.2 方法

1.2.1 Ts CDC37 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(blastic local alignment search tool, BlastX, <http://www.ncbi.nlm.nih.gov/BLAST/>)程序,与 GenBank 中的其他物种 CDC37 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 resblast 分析其保守功能域。

1.2.2 Ts CDC37 核酸和氨基酸序列分析 综合性蛋白核酸分析工具包(vector NTIsuite)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。

1.2.3 Ts CDC37 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 Ts CDC37 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 Ts CDC37 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及二级结构、分子的疏水性、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。

2 结果

2.1 文库质粒编号为 tscaa0_000445 插入序列的分析

从 BLASTX 的分析结果来看,文库质粒 tscaa0_000445 中的插入 cDNA 序列全长 1 550 bp,编码区为 175 ~ 1 378 bp,编码 400 个氨基酸。其最大的 ORF 含 1 203 bp,5'端和 3'端均有非编码序列,起始密码子为 ATG,终止密码子都为 TGA,编码 400 个氨基酸,为猪带绦虫成虫 CDC37 的全长编码序列(图 1)。

2.2 Ts CDC37 蛋白质的理化性质

ExPASy 中 ProtParam 预测 Ts CDC37 的理论分子量和等电点分别是 46 802.5 Da 和 5.08。含有 8 个半胱氨酸,彼此之间形成二硫键可能性较小。假设形成 4 对二硫键时,280 nm 处的摩尔消光系数为 $45\,420\text{ M}^{-1}\text{ cm}^{-1}$, 0.1% 浓度的 Abs 为 0.970;假设二硫键全部打开时,该蛋白在 280 nm 处的摩尔消光系数为 $44\,920\text{ M}^{-1}\text{ cm}^{-1}$, 0.1% 浓度的 Abs 为 0.960。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,在酵母和大肠杆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 47.71,高于域值 40,在溶液中性质较不稳定。脂肪族指数 67.08,亲水性指数为 -0.997,蛋白质疏水性较高。

2.3 Ts CDC37 翻译后修饰 亚细胞定位的预测

ExPASy 中 TargetP 预测 Ts CDC37 无信号肽序列,也没有线粒体、过氧化酶体、溶酶体和细胞核等定位序列。

2.4 Ts CDC37 一级结构及二级结构预测分析

ExPASy 中 InterPro Scan 扫描结果显示该氨基酸序列中含有 2 段 Ts CDC37 保守结构域和功能域,分别位于 N 端和 C 端。扫描结果如图 2 所示。

ExPASy 中 PredictProtein 分析蛋白的拓扑结构和二级结构,该蛋白无跨膜区。Sec 预测 α 螺旋(H)、 β 折叠(E)和无规卷曲(L)的比例分别为 60.5%、2.00% 和 37.50%。该蛋白的二级结构以 α 螺旋为主。

1 GCTAGCTCGCCG CAG CCGAACACCCGAGCG CAGCGAGTCAGTGAG CGAGGAAGCGGCCG C ATAACCTCGTATAGC
76 ATACATTATACGAAG TTATCAGTCGACGGT ACCGGACATATGCCC GGGAAATTCGGCCATT ACGGCCGGGCCAC
M D Y S K W D H I E V S D D E D D
151 TTTTAGCCTTTT CACA TAATTAAGTTATGGA TTACAGTAAATGGGA TCATATCGAGGTTTC AGATGATGAGGACGA
T H P N I D T A S L F R W R H E A R L T R D R E W
226 TACCCATCCCAATAT AGATACTGCTAGTCT ATTCCGTTGGCGTCA CGAAGCTCGTCTTAC TCGTGACCGGAGTG
K E E K E K F E K E K K E H I E A L Q K A R S L Y
301 GAAAGAAGAAAAGGA GAAGTTTGAAAAAGA GAAGAAGGAGCATAT TGAAGCCCTACAGAA GGCTCGGTCACTTTA
E E G V K S N A P N V K D L E E N L R K L E V A D
376 CGAGGAGGGTGTCAA GAGTAATGCCCCAAA TGTGAAAGACTTGGA AGAAAATTTAAGGAA GCTGGAGGTG GCTGA
K E W Q E K E K E M N K K E R L R P L N V D T I S
451 CAAGGAATGGCAGGA GAAGGAAAAAGAAAT GAACAAAAAGAAAG GCTACGTCCTCTTAA TGTGATACATAAG
H E G R S R T F I N K E A L K E K P K V D E D N D
526 CCATGAAGGAAGATC TCGAACGTTTATCAA CAAGGAGGCTTTGAA AGAGAAGCCGAAGT AGACGAGACAATGA
D D V H E E A A D R L R D F V K K H E A E I R K F
601 TGATGATGTTACAGA GGAAGCCGCGGACCG ACTCAGAGATTTTGT GAAGAAACACGAAGC CGAAATTAGGAAGTT
G M F S R P V D S Q Q Y L L E K P F L V C D E T A
676 TGGCATGTTTAGCCG TCCGGTTGACTCACA ACAGTATCTTCTTGA AAAACCTTCCTTGT TGTGACGAAACCGC
N H L V L W C L D L A M E E K Y S L L D H V S H Q
751 CAATCATCTCGTCTT GTGGTGCTCGATCT TGCCATGGAAGAGAA GTACAGTCTTCTGGA TCAGGTCTCCCATCA
C I V M Q F M L E L A K S L K C D P R S C I R P F
826 ATGCATCGTGATGCA GTTCATGCTAGAAAT GGCCAAGTCCCTCAA ATGCGATCCACGATC ATGCAATCGACCCCTT
F N K F I N P E P E Y Q R M F E E E L T A F K A R
901 CTTCAACAAATTCAT CAACCCAGAGCCGGA GTACCAGCGCATGTT CGAGGAAGAGCTGAC GGCCTTCAAAGCAG
I R D R A R V K L T E A M Q Q Y E E E E R Q K R L
976 TATTCGTGATCGTGC TAGAGTAAACTAAC CGAGGCGATGCAGCA GTACGAGGAAGAGGA GCGCCAGAAACGCT
G P G G L D P Y E V I E T L P E P L R K C F E T R
1051 TGGCCCCGGTGGATT GGATCCCTACGAAGT CATAGAAACCTTCC TGAGCCCCTGCGAAA GTGTTTCGAGACTCG
D V E L L K K T L T E M D P K E A E E C M K R C I
1126 GGACGTTGAGTTGTT GAAGAAGACGCTGAC AGAGATGGATCCAAA AGAGGCGGAAGAGTG TATGAAGCGATGCAT
A S G L W V E N A A A A D A E G G E Q V Y D T T N
1201 TGCTCTGCGCTCTG GGTGAAAAATGCCGC TGC TGCTGATGCTGA AGGTGGAGAGCAGGT ATATGACACGACGAA
G V S V E E V K A S A G K E N V K K E E A S A N S
1276 CGGTGTGAGTGTTGA AGAGGTGAAGGCTTC GCGGGAAAGAGAG TGTGAAGAAGGAGGA GCGCTCGCGCAATTC
G D N A P N K S *
1351 CGGTGACAATGCTCC TAATAAGTCGTTT CTGTTTTCACCTCCT TGTGCTGGGTCTTA TCGGTCGTGCAAT
1426 ATTAATGTGATGCA CTAAATTCGCTTTCC TTCGAAAAAAGAGAGG AAAAAAAGAGAGAGG AAAAAAAGAGAGG
1501 TCGGCCGCTCGGCC CTCGAGAAGCTTCT AGACCATTCTTTGG CGCGC

图 1 文库质粒 tscaa0_000445 的插入 cDNA 序列及其最长 ORF 编码的氨基酸序列

Fig. 1 Inserted cDNA sequence in plasmid library tscaa0_000445 and the coding amino acid sequence of its longest ORF

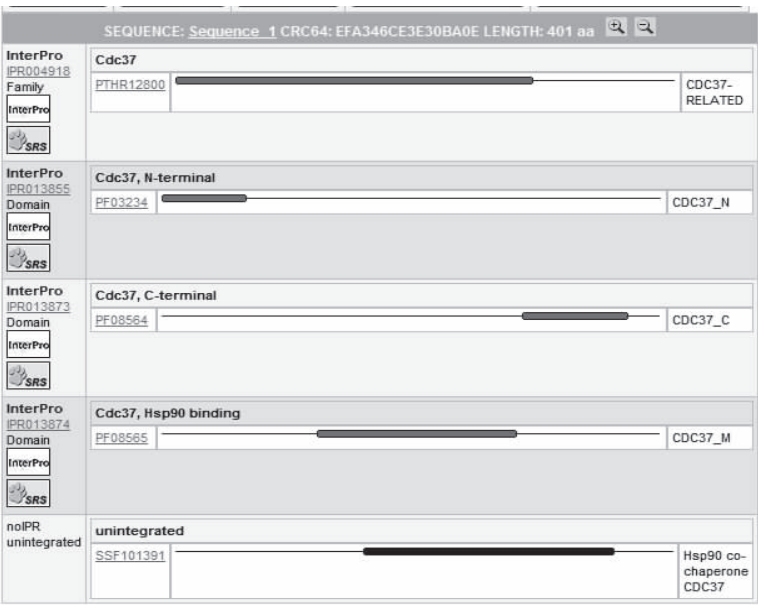


图 2 InterPro Scan 分析 Ts CDC37 结构域和功能域

Fig. 2 The domain and functional region of Ts CDC37 analyzed by InterPro Scan

2.5 Ts CDC37 的抗原表位分析

IEDB Analysis Resource 中 B Cell Epitope Prediction Tools 分析预测 Ts CDC37 的 B 细胞抗原表位结果显示,该氨基酸序列中有 12 个潜在抗原表位,其中主要的表位有六个:10 ~ 23 aa(EVSDDDED-

DTHPNID) ,68 ~ 81 aa(EEGVKS NAPNVKDL) ,92 ~ 105 aa (DKEWQEKEKEMNKK) 130 ~ 151 aa (ALKEKPKVDEDNDDDVHEEAAD) ,281 ~ 301 aa (QQYEEEEERQKRLGPGGLDPYE) , 327 ~ 336 aa (TEMDPKEAEE) 。结果如图 3 所示。

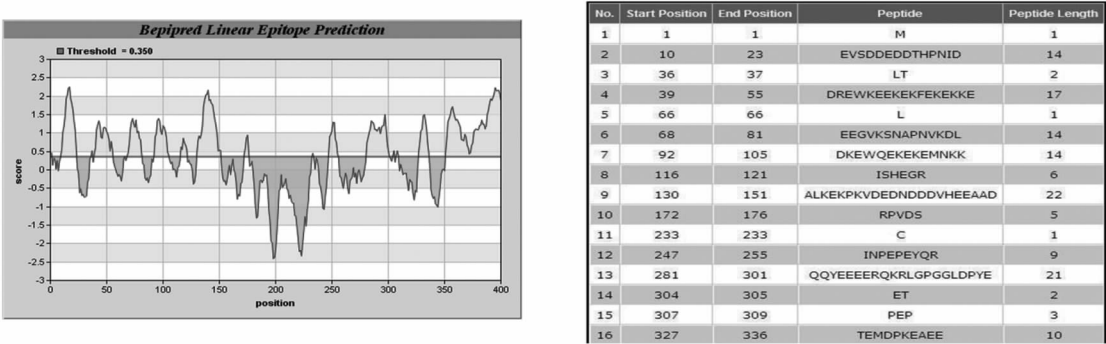


图 3 Ts CDC37 的 B 细胞抗原表位分布
Fig.3 Distribution picture of B cell epitopes of Ts CDC37

2.6 Ts CDC37 与其它物种 CDC37 分子的进化树的构建

应用 Vector NTI 软件对 8 个物种 CDC37 的氨基酸序列进行分子进化树的构建(图 4),结果显示在这 8 个物种中 Ts CDC37 与亚洲带绦虫 CDC37 的进化关系最近,其次是日本血吸虫和曼森血吸虫的 CDC37,而与非洲爪蛙、人类、短尾猴和斑鳢的 CDC37 进化关系较远。这与不同进化阶元的物种间的进化距离与种系演化进程相符,揭示 Ts CDC37 基因可以作为分子进化的指标。

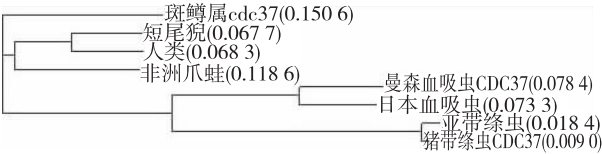


图 4 猪带绦虫 CDC37 与其他物种
CDC37 蛋白的分子进化分析
Fig.4 Molecular evolution analysis on CDC37
protein of *Taenia solium* and other species

3 讨论

生物信息学可以通过对已有的核酸和蛋白质序列数据库进行扫描和比对搜索目标序列特殊的结构特征(如各种亚细胞的定位信号、翻译后的修饰位点、功能域等)对基因的功能进行初步的理论

预测,为其功能研究寻找线索^[4]。
从猪带绦虫成虫全长 cDNA 质粒文库中筛选到 1 个与日本血吸虫同源性较高的 CDC37 基因,它是细胞分裂周期蛋白,最初是在芽殖酵母中被发现的,并在酵母的细胞分裂过程中起着举足轻重的作用^[5]。作为一种辅助伴侣分子,在 Hsp90 的众多辅助伴侣分子中,它可能是导致细胞转化的一个关键因素,以特异性地方式将蛋白激酶募集到 Hsp90 上,以促进它们进行正确的折叠。由于这些蛋白激酶参与了信号转导、细胞周期、细胞生存和增殖等生命活动,而且在肿瘤组织中特异性地高表达,因此,CDC37 的异常表达与肿瘤的发生、进展相关。作为癌基因的驱动者,它还能够通过多种蛋白激酶底物来实现其促进细胞转化的功能,在细胞生存过程中起着重要的作用^[6]。也有研究显示,如果抑制 CDC37 的功能,将诱导细胞凋亡。在寄生虫的研究领域,CDC37 在果蝇和线虫的生存过程中也起着重要的作用。综上所述,CDC37 基因的功能很强大,但对它的了解又甚少,特别是在带绦虫研究领域。该基因是否位于虫体的表面,是否与虫体损伤皮肤的修复和免疫逃避有关,是否缺失该基因虫体就会死亡以及它在带绦虫的生命活动中是否处于关键的地位或者与其他关键基因之间有什么相互作用,这都需要更进一步的实验去探究。

(下转第 351 页)

灵敏度高、重现性好,可用于表征注射用复方荭草冻干粉针生物活性信息,以达到综合控制产品质量的目的。建议采用该法测定时结果为阳性即可判断为合格,若初试结果为阴性允许复试后合并计算。

4 参考文献

- [1] 李仪奎. 中药药理实验方法学[M]. 2 版. 上海: 上海科学技术出版社, 2006: 311 - 312.
- [2] 肖贵南, 李瑾翡, 陈浩桢. 生物活性测定在中药质量控制中应用的可行性及研究思路[J]. 中药材, 2008 (4):

473 - 475.

- [3] 唐元泰, 芮菁. 关于中药标准采用生物活性测定项目的建议[J]. 中国药品标准, 2007 (6): 39 - 45.
- [4] 国家中医药管理局《中华本草》编委会. 中华本草[M]. 2 册. 上海: 科学技术出版社, 1999: 681 - 682.
- [5] 杨仓良. 毒药本草[M]. 北京: 中国中医药出版社, 1993: 435 - 436.
- [6] 国家药典委员会编. 中国药典[M]. 一部附录 XVIII C. 北京: 中国医药科技出版社, 2010: 132 - 133.

(2012 - 04 - 01 收稿, 2012 - 04 - 25 修回)

编辑: 周 凌

(上接第 347 页)

为研究此基因编码的蛋白质 CDC37, 首先利用生物信息学分析 CDC37 的理化性质, 如等电点、分子量、稳定性、氨基酸组成, 这有助于采取合理的克隆和表达策略, 提高目的蛋白的高效可溶性表达, 以期获得有活性的重组蛋白。将该序列推导出的氨基酸序列与 GenBank 中搜索到的同源蛋白的氨基酸序列比较发现, 它与日本血吸虫的同源性最高, 一致性达到 60%, 相似性达到 76%。分析发现 CDC37 在溶液中较稳定, 所以可以采用 HIS 标签的原核表达策略。在此基础上, 还选用其它一些有用的在线分析软件对功能域、亚细胞定位、二级结构、拓扑结构及抗原表位等全面的分析, 这些分析都会对该基因的研究起到指导作用。比如: 二级结构基本上以 α 螺旋为主, 提示该蛋白结构紧密而稳定, 未发现跨膜区 (M), 超过一半的氨基酸残基暴露在蛋白质外部, 推测不是膜蛋白; CDC37 的分子中没有信号肽, 提示它一般不会分泌到细胞外, 还不具有线粒体、过氧化酶体、质体、溶酶体等亚细胞定位序列, 提示可能是胞浆蛋白。生物信息学分析可以避免在后续实验工作中的盲目性, 以帮

助更快更全面的了解蛋白质, 尤其是新基因编码的蛋白质的结构功能。

4 参考文献

- [1] 贾万忠, 王玉潮, 闫鸿斌, 等. 人体带绦虫病诊断方法的研究进展[J]. 中国兽医科学, 2010 (2): 210 - 214.
- [2] 韩向新. 猪囊虫病的危害与防治措施[J]. 2008 (12): 73 - 74.
- [3] 食物源性寄生虫病的流行形势与控制对策[J]. 实用防医学, 2005 (4): 985 - 988.
- [4] 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.
- [5] 赵名, 于晓妣. 辅助伴侣分子 Cdc37 蛋白的研究进展[J]. 生物技术通讯, 2009 (3): 404 - 407.
- [6] Reed S I. The selection of *S. cerevisiae* mutants defective in the start event of cell division[J]. Genetics, 1980 (3): 561 - 577.

(2012 - 05 - 30 收稿, 2012 - 06 - 20 修回)

编辑: 张丽君

《贵阳医学院学报》加入期刊数据库启事

《贵阳医学院学报》已入编中文科技期刊数据库、中国期刊全文数据库、万方数据 - 数字化期刊群、中国核心期刊(遴选)数据库、解放军医学图书馆等数据库, 刊登在本刊的论文著作权使用费与本刊稿酬一次性给付, 不再另行发放。作者如不同意将文章入编相关数据库, 投稿时敬请说明。

《贵阳医学院学报》编辑部