· 实验研究 ·

细粒棘球绦虫延伸因子 1 生物信息学分析*

李超群[1], 樊海宁[2], 张耀刚[1], 曹得萍[1,2]

摘　要: 目的　利用生物信息学技术对细粒棘球绦虫的延伸因子 1(EF-1)进行分析,探讨其诊断价值。**方法**　通过 Expasy 系统预测 EF-1 的理化性质,使用 DNAStar 软件分析蛋白亲水性、柔性区域、抗原指数及表面可及性,利用 ABCpred 与 IEDB 软件综合预测 B 细胞线性抗原表位,通过 DNAStar 软件预测 T 细胞表位,使用 SOPMA 软件预测二级结构,SWISS-MODEL 网站构建 EF-1 的三级结构,使用 MEGA 软件选择 neighbor joining 法构建 EF-1 氨基酸序列的系统发育树。**结果**　EF-1 基因序列全长 1 412 bp,含有 2 个内含子;EF-1 亲水性得分较高的区域为 35 ~ 45、52 ~ 70、126 ~ 140、158 ~ 183、301 ~ 322、357 ~ 380、424 ~ 448;柔性区域为 26 ~ 45、54 ~ 71、162 ~ 176、300 ~ 318、357 ~ 381、432 ~ 448;抗原性指数得分较高的区域分布与柔性区域一致;表面可及性得分较高的区域较少,主要分布在 39 ~ 54、62 ~ 76、128 ~ 132、161 ~ 170、220 ~ 234、315 ~ 329、354 ~ 372、424 ~ 448;可能的 B 细胞线性表位的氨基酸序列为 120 ~ 130、286 ~ 295、306 ~ 320、365 ~ 378;可能的优势性 T 细胞表位区域为 34 ~ 48、164 ~ 177、222 ~ 244、280 ~ 290、321 ~ 330、396 ~ 407;EF-1 的二级结构中 α 螺旋占 32.81%、β 折叠占 22.54%、β 转角占 10.94%、无规则卷曲占 33.71%;系统发育树结果表明多房棘球绦虫为一枝,其余聚于一枝,且内部形成许多梳齿状分枝。**结论**　得到 4 个可能形成 B 细胞表位的区域、6 个优势性 T 细胞表位区域,EF-1 具有较高的保守性,可作为免疫诊断和药物治疗靶点 。

关键词: 细粒棘球绦虫; 延伸因子 1(EF-1); 生物信息学

中图分类号: R 383.3*3　**文献标志码**: A　**文章编号**: 1001-0580(2017)04-0607-04　DOI: 10.11847/zgggws2017-33-04-23

Property and structure of elongation factor 1 in *Echinococcus granulous*, a bioinformatics analysis

LI Chao-qun*, FAN Hai-ning, ZHANG Yao-gang, et al (* *Department of Pathogenic Biology, Medical College, Qinghai University, Xining, Qinghai Province* 810001, China)

Abstract: Objective　To analyze the property and structure of elongation factor 1 (EF-1) in *Echinococcus granulous* with bioinformatics technology and to explore its value in diagnosis of echinococcosis. **Methods**　Physicochemical property of EF-1 was predicted using Expasy system. Hydrophiloity plot, flexible regions, antigenic index, and surface accessibility were analyzed with DNAStar software. B cell linear antigen epitopes were forecasted with ABCpred and IEDB software. T cell antigen epitopes were forecasted with DNAStar software. The secondary and three-dimensional structure of EF-1 were determined using SOPMA software and SWISS-MODEL website separately. The phylogenetic tree for amino acid sequences of EF-1 was built with neighbor joining method of MEGA software. **Results**　The whole gene of EF-1 was 1 412 bp in length, containing two introns. The high score areas of hydrophiloity plot were 35 – 45, 52 – 70, 126 – 140, 158 – 183, 301 – 322, 357 – 380, and 424 – 448. The flexible regions were 26 – 45, 54 – 71, 162 – 176, 300 – 318, 357 – 381, and 432 – 448. The regions with high score of antigenic index were similar to the flexible regions. There were a few regions with the high score of surface accessibility, mainly locating at 39 – 54, 62 – 76, 128 – 132, 161 – 170, 220 – 234, 315 – 329, 354 – 372, and 424 – 448. The potential amino acids sequences of B cell linear epitopes were 120 – 130, 286 – 295, 306 – 320, and 365 – 378. The potential advantaged T cell epitope areas were 34 – 48, 164 – 177, 222 – 244, 280 – 290, 321 – 330, and 396 – 407. The proportions of secondary structure of EF-1 were 32.81% for α-helix, 22.54% for β-sheet, accounted 10.94% for β-corner, and 33.71% for random coil, respectively. The constructed phylogenetic tree of EF-1 demonstrates a branch of *Echinococcus* multilocularis had its own branch and another branch with many internal dentate branches. **Conclusion**　Four B cell epitopes and six advanted T cell epiopes of EF-1 were identified with the bioinformatics analysis and phylogenetic tree reveals that EF-1 is high conservative. These findings suggest that the epitopes of EF-1 in *Echinococcus granulous* could be adopted as targets for immunological diagnosis and drug treatment of echinococcosis.

Key words: *Echinococcus granulous*; elongation factor 1; bioinformatics analysis

棘球蚴病(echinococcosis)又称包虫病,是一种危害严重、呈世界性分布的人兽共患性寄生虫病,中国属高发地区之一。目前对于包虫病的免疫学诊断所用抗原是通过粗提原头节、囊液蛋白而获得,虽具

* **基金项目**: 国家自然科学基金(81360255)
作者单位: 1. 青海大学医学院病原生物学教研室,西宁 810001; 2. 青海大学包虫病重点实验室
作者简介: 李超群(1992 –),女,河南人,硕士在读,研究方向:病原生物学。
通讯作者: 曹得萍,E-mail: qhmccdp@163.com
数字出版日期: 2017 – 3 – 29 8:47
数字出版网址: http://kns.cnki.net/kcms/detail/21.1234.R.20170329.0847.016.html

有较高的敏感性和特异性[1]，但抗原成分复杂，易发生过敏反应，且抗原来源困难，不利于大规模生产，限制其在临床应用。从棘球蚴原头节、囊液中筛选出具有特异性的抗原，通过基因工程技术大规模合成有望成为今后的发展方向。而抗原表位是抗原分子抗原性的基础，因此预测抗原分子的抗原表位，是寻找特异性抗原分子有效的方法。确定抗原表位常用的技术有 X 射线衍射与核磁共振分析、实验技术、生物信息学技术[2-3]，前两者具有很高的准确性但成本昂贵，技术要求高，而通过生物信息学技术预测抗原表位，不仅可节省大量经费及时间，其预测准确性也随着生物信息学的发展而不断提高。有研究发现，延伸因子 1（elongation factor 1，EF-1）是一种具有潜在诊断价值的抗原[4-5]，本研究通过生物信息学技术对其进行分析，旨在了解 EF-1 在细粒棘球绦虫的免疫学诊断价值。

1　材料与方法

1.1　基因及氨基酸序列　查询 GenBank 并经 BLAST 比对确定 EF-1 基因序列，采用网站自动翻译的氨基酸序列用来后续分析。

1.2　理化性质推测　EF-1 的理化性质预测使用 Expasy 系统的 protparam 数据库（http://www.expasy.org/tools/protparam.html）。使用 DNAStar 软件 Kyte-Doolittle 法，预测 EF-1 亲水性；Karplus-Schulz 法预测 EF-1 柔性区域；Jameson-Wolf 法预测 EF-1 抗原性指数；Plot-Emini 法预测 EF-1 表面可及性。

1.3　B、T 细胞表位预测　利用 ABCpred 在线软件（http://www.imtech.res.in/raghava/abcpred/）及 IEDB 软件（http://tools.immuneepitope.org/main/html/tcell_tools.html）综合预测 EF-1 B 表位，选取 2 者重叠部分中可能形成 B 细胞表位的部分。为提高预测准确性，ABCpred 只保留得分≥0.8 的结果，IEDB 软件预测结果中只保留长度 >7 的结果。应用 DNAStar 软件 AMPHI 法预测 Th 细胞抗原表位、Rothbard-Taylor 预测含特定基序（motif）的 T 淋巴细胞抗原表位，综合 2 种预测结果推测优势性 T 细胞表位。

1.4　二、三级结构预测　二级结构预测使用 SOPMA 在线软件（https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl? page =/NPSA/npsa_sopma.html），三级结构预测运用 SWISS-MODEL 网站（http://swissmodel.expasy.org/）同源建模方法，进行自动建模。

1.5　EF-1 氨基酸序列同源性分析　从 GenBank 中下载部分绦虫 EF-1 的氨基酸序列、使用 MEGA 软件选择 neighbor joining 法构建系统发育树。

2　结　果

2.1　EF-1 基因及氨基酸序列　EF-1 基因序列（No. AB306934）全长 1 412 bp，含有 2 个内含子（1 ~ 144、745 ~ 776），网站自动翻译的氨基酸序列（No. BAF63675）由 448 个氨基酸残基组成。通过 protparam 数据库预测得到 EF-1 相对分子量为 48 856.7 kDa，等电点 pI 为 9.23，半衰期为 30 h（mammalian）、>20 h（yeast）、>10 h（E. coli），不稳定指数为 28.32（为稳定蛋白）。

2.2　EF-1 蛋白亲水性、柔性区域、抗原指数及表面可及性（图 1）　Kyte-Doolittle 法预测的 EF-1 亲水性得分较高的区域为 35 ~ 45、52 ~ 70、126 ~ 140、158 ~ 183、301 ~ 322、357 ~ 380、424 ~ 448；Karplus-Schulz 法预测的柔性区域散在分布，较大的几个区域为 26 ~ 45、54 ~ 71、162 ~ 176、300 ~ 318、357 ~ 381、432 ~ 448；Jameson-Wolf 法预测的抗原性指数得分较高的区域分布与 Karplus-Schulz 法预测的柔性区域一致；Plot-Emini 法预测的表面可及性得分较高的区域较少，主要分布在 39 ~ 54、62 ~ 76、128 ~ 132、161 ~ 170、220 ~ 234、315 ~ 329、354 ~ 372、424 ~ 448。

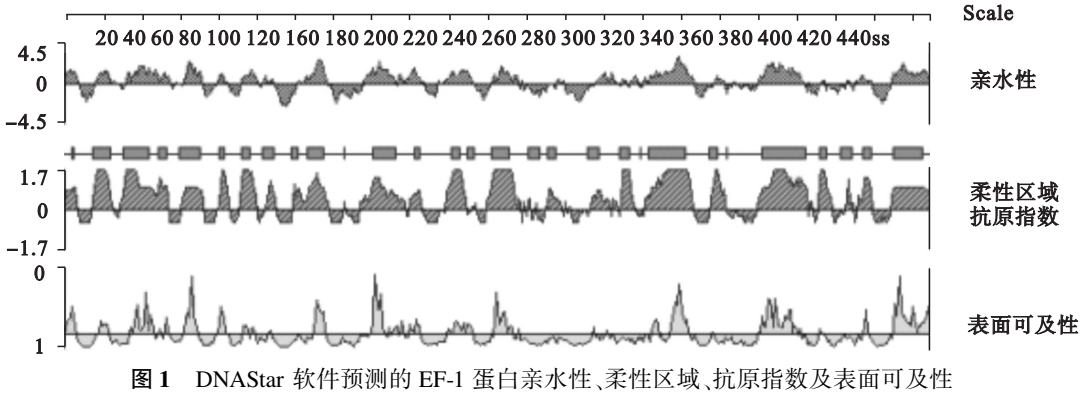

图 1　DNAStar 软件预测的 EF-1 蛋白亲水性、柔性区域、抗原指数及表面可及性

中国公共卫生 CPH

2.3　B 细胞表位预测(图 2)　通过 ABCpred 软件,得到的可能形成 B 细胞线性表位的氨基酸序列区域为 87 ~ 103、114 ~ 130、146 ~ 162、187 ~ 203、253 ~ 269、279 ~ 295、304 ~ 320、330 ~ 346、336 ~ 352、362 ~ 378、377 ~ 393,IEDB 软件预测的区域为 17 ~ 24、41 ~ 52、120 ~ 134、198 ~ 210、221 ~ 231、286 ~ 296、306 ~ 325、365 ~ 383,两者重叠区域中可能形成 B 细胞线性表位的氨基酸序列为:120 ~ 130、286 ~ 295、306 ~ 320、365 ~ 378。

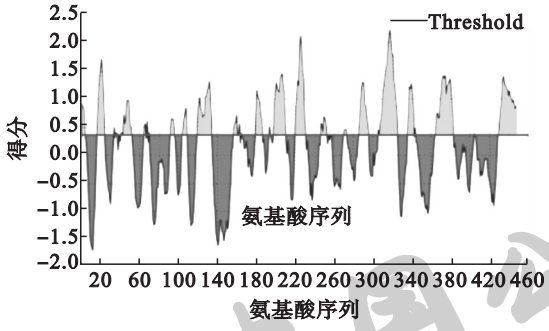

图 2　IEDB 软件预测的 EF-1 B 细胞表位

2.4　T 细胞表位预测　AMPHI 法预测可能形成 T 细胞表位的区域为 15 ~ 22、34 ~ 49、82 ~ 92、97 ~ 106、163 ~ 175、212 ~ 251、258 ~ 266、280 ~ 290、367 ~ 378、397 ~ 408;Rothbard - Taylor 法预测可能的 T 淋巴细胞抗原表位为 36 ~ 51、94 ~ 101、155 ~ 162、172 ~ 179、181 ~ 188、217 ~ 224、236 ~ 242、281 ~ 291、321 ~ 330、349 ~ 355、356 ~ 362、398 ~ 405、420 ~ 428,取两者重叠部分,得到可能的优势性 T 细胞表位区域为 34 ~ 48、164 ~ 177、222 ~ 244、280 ~ 290、321 ~ 330、396 ~ 407。

2.5　二、三级结构(图 3)　SOPMA 在线软件预测的 EF-1 二级结构中 α 螺旋占 32.81%、β 折叠占 22.54%、β 转角占 10.94%、无规则卷曲占 33.71%,SWISS-MODEL 网站预测的 EF-1 的三级结构见图 3。

图 3　SWISS-MODEL 网站构建的 EF-1 三级结构

2.6　EF-1 基因同源性分析(图 4)　从构建的系统发育树来看,EF-1 具有较高的保守性,除多房棘球绦虫外,其他聚于一枝,并且在这枝内部形成梳齿状分枝,人作为外类群单独处于一枝。

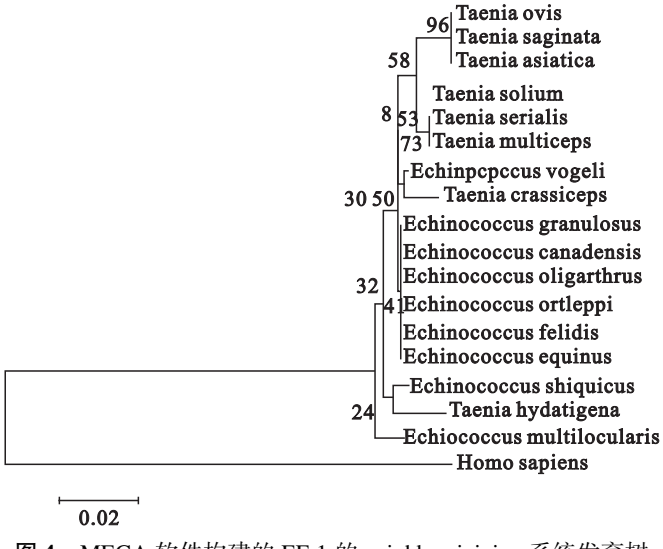

图 4　MEGA 软件构建的 EF-1 的 neighbor joining 系统发育树

3　讨　论

EF-1 是一种多聚体核糖体蛋白质,在细胞内普遍存在且大量表达,真核细胞可能有 α、β、γ、δ 4 个亚基。其中 α 亚基属 G 蛋白家族,与鸟苷酸结合调节蛋白(guanine nucleotide-binding regulatoryprotein,GTP)、氨基酰-tRNA 组成复合体,促进氨基酰-tRNA 进入氨基酰位,EF-1β 具有 GTP 酶活性,在 EF-1α 从核糖体脱落构象由 GDP 变为 GTP 准备结合新氨基酰 - tRNA 的过程中起催化作用,γ 亚基常与 β 亚基形成复合物,可提高 β 亚基的 GTP 酶活性[6-7]。有研究发现,在脊椎动物中还有 δ 亚基,其 C 端与 β 亚基具有同源性,但 N 端没有,提示,它们可能执行着不同功能[8]。

生物信息学是由生物学、计算机和信息论等融合而成的一门学科,组织和存储了海量由基因遗传学、分子生物学和生物技术推动的生物信息[9]。已在药物设计、蛋白结构预测,表位疫苗预测等方面得到广泛应用。随着理论和技术的积累,大量蛋白、抗原表位数据库得以建立,如 MHC 配体和多肽基序的数据库 SYFPEITHI、欧洲蛋白序列数据库(Swiss-Prot Protein Sequence Database, SWISS-PROT)、免疫表位数据库(immune epitope database, IEDB)、B 细胞表位预测数据库(prediction of linear B-cell epitopes, Bcepred)等,预测结果精确度不断提高,使得通过生物信息学技术预测蛋白结构、抗原表位成为及时高效的手段。

抗原表位是指免疫应答过程中被特异的效应分子或 B、T 淋巴细胞识别的抗原分子中具有特殊结构与免疫活性的化学基团,根据所结合细胞的不同,分为 B 细胞表位和 T 细胞表位[10-11]。B 细胞表位多为亲水性氨基酸或单糖残基,常位于抗原分子表面,而 T 细胞表位常为隐藏于抗原内部的疏水性肽段[2]。抗原表位决定着抗原分子的抗原性,研究抗原分子的抗原表位,对诊断试剂的制备、多肽疫苗的合成、单克隆抗体的筛选具有重要意义[12]。

对于 B 细胞线性表位的预测主要通过蛋白亲水性、柔性区域、抗原指数、表面可及性、二级结构综合来推测可能形成结合抗原的区域[13-16],亲水性区域位于抗原表面与 B 细胞表位关系密切。本研究结果显示,EF-1 蛋白亲水性区域占 50% 以上,柔性区域的分布与抗原指数的分布区域一致,占 60% 以上,这为 EF-1 形成 B 细胞抗原表位提供了有利条件;表面可及性分布的区域较少提示,EF-1 也可能形成构象表位。结合 ABCpred 和 IEDB 软件分析的结果得到可能形成 B 细胞抗原表位区域为:120 ~ 130、286 ~ 295、306 ~ 320、365 ~ 378,与抗原指数得分较高的区域分布有较大的重叠区域。研究表明二级结构中,α 螺旋、β 折叠常位于抗原内部,形态固定,结构规则,不利于抗体结合,而 β 转角和无规则卷曲位于抗原的表面,利于抗体结合,易于形成抗原表位[17]。SOPMA 在线软件预测的 EF-1 二级结构中 β 转角占 10.94%、无规则卷曲 33.71%,为形成 B 细胞抗原表位提供了有利条件;EF-1 三级结构中有较多的无规则卷曲,为形成 B 细胞线性表位提供了可能。

T 细胞表位常位于抗原内部的疏水性区域。本研究结合 AMPHI 与 Rothbard-Taylor 的预测结果,得到 EF-1 可能的优势性 T 细胞表位区域为 34 ~ 48、164 ~ 177、222 ~ 244、280 ~ 290、321 ~ 330、396 ~ 407,由亲水性预测结果来看,这些区域多分布在得分为负值的疏水性区域,结果较为准确。从构建的系统发育树来看,EF-1 具有较高的保守性,除多房棘球绦虫外,其他均聚在一枝,并在这枝内部呈现梳齿状分枝。提示,EF-1 在棘球绦虫与其他绦虫之间存在种属差异,利用这种差异,寻找细粒棘球绦虫 EF-1 特有表位,可为设计表位抗原疫苗提供新的选择[18-19]。

参考文献

[1] Zhang W, Wang S, McManus DP. *Echinococcus granulosus* genomics: a new dawn for improved diagnosis, treatment, and control of echinococcosis[J]. Parasite, 2014, 21: 66.

[2] 宋帅,李春玲,贾爱卿,等. 抗原表位研究方法进展[J]. 动物医学进展, 2010, 12: 87 - 91.

[3] Dall'Antonia F, Pavkov-Keller T, Zangger K, et al. Structure of allergens and structure based epitope predictions[J]. Methods, 2014, 66(1): 3 - 21.

[4] Matsubayashi M, Teramoto-Kimata I, Uni S, et al. Elongation factor-1α is a novel protein associated with host cell invasion and a potential protective antigen of *Cryptosporidium parvum*[J]. J Biol Chem, 2013, 288(47): 34111 - 34120.

[5] 刘玉江,戴佳琳,黄江,等. 猪带绦虫成虫 EF-1 基因生物信息分析及原核表达[J]. 中国公共卫生, 2011, 27(1): 83 - 84.

[6] 庞建新,黄江,胡旭初,等. 亚洲带绦虫成虫延伸因子 - 1 基因的克隆和分析[J]. 热带医学杂志, 2007, 06: 513 - 515.

[7] Sasikumar AN, Perez WB, Kinzy TG. The many roles of the eukaryotic elongation factor 1 complex[J]. Wiley Interdiscip Rev RNA, 2012, 3(4): 543 - 555.

[8] Cao Y, Portela M, Janikiewicz J, et al. Characterisation of translation elongation factor eEF1B subunit expression in mammalian cells and tissues and colocalisation with eEF1A2[J]. PLoS One, 2014, 9(12): e114117.

[9] Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, et al. An overview of bioinformatics tools for epitope prediction: implications on vaccine development[J]. J Biomed Inform, 2015, 53: 405 - 414.

[10] Zhang C, Li Y, Tang W, et al. The relationship between B-cell epitope and mimotope sequences[J]. Protein Pept Lett, 2016, 23(2): 132 - 141.

[11] Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine[J]. Genome Med, 2015, 7: 119.

[12] Saravanan V, Gautham N. Harnessing computational biology for exact linear B-cell epitope prediction: a novel amino acid composition-based feature descriptor[J]. OMICS, 2015, 19(10): 648 - 658.

[13] 梁瑾,王靖飞. B 细胞表位预测方法研究进展[J]. 生命科学, 2009, 2: 320 - 323.

[14] Lian Y, Ge M, Pan XM. EPMLR: sequence-based linear B-cell epitope prediction method using multiple linear regression[J]. BMC Bioinformatics, 2014, 15: 414.

[15] Hu YJ, Lin SC, Lin YL, et al. A meta-learning approach for B-cell conformational epitope prediction[J]. BMC Bioinformatics, 2014, 15: 378.

[16] Huang JH, Wen M, Tang LJ, et al. Using random forest to classify linear B-cell epitopes based on amino acid properties and molecularfeatures[J]. Biochimie, 2014, 103: 1 - 6.

[17] Nasrul Islam M, Iqbal S, Katebi AR, et al. A balanced secondary structure predictor[J]. J Theor Biol, 2016, 389: 60 - 71.

[18] 彭传林,魏川川,吴建伟,等. 家蝇溶菌酶 MDLZM 基因克隆、表达和序列分析[J]. 中国公共卫生, 2015, 31(6): 764 - 766.

[19] 赵学军,国果,修江帆,等. 家蝇伴侣蛋白 CCTη 基因序列分析与表达模式[J]. 中国公共卫生, 2015, 31(8): 1043 - 1046.

收稿日期: 2017-02-20　　(解学魁编校)