

DOI:10.13350/j.cjpb.220213

• 论著 •

细粒棘球绦虫抗原 Fis1 的生物信息学分析*

肖静^{1,2,3}, 曲晨⁴, 赵嘉庆^{1,2,3**}

(1. 宁夏医科大学基础医学院, 宁夏银川 750004; 2. 宁夏医科大学科学技术研究中心;
3. 宁夏常见传染病重点实验室; 4. 宁夏医科大学临床医学院)

【摘要】 目的 利用生物信息学的方法分析细粒棘球绦虫 Fis1 (EgFis1) 蛋白的抗原表位, 为分子肽疫苗的研发奠定理论基础。方法 在 NCBI 数据库中检索并下载 EgFis1 蛋白的氨基酸序列; 利用 EXpasy 软件预测蛋白的理化性质; 采用 SignalP5.0 和 TMHMM sever 2.0 软件预测 EgFis1 蛋白的信号肽和跨膜结构域; 利用 SOPMA 和 SWISS-MODLE 预测 EgFis1 蛋白的二级结构和三级结构; 采用 IEDB、ABCpred 和 SYFPEITHI 数据库预测 EgFis1 蛋白的 T、B 细胞表位。结果 细粒棘球绦虫 EgFis1 是由 157 个氨基酸组成的等电点为 5.86, 分子质量单位为 16.93 ku 的蛋白质; 不具有信号肽序列, 但具有一个跨膜结构域; 其二级结构中 α -螺旋占比例为 48.41%, 延伸连占 19.11%, β -转角占 7.01%, 无规则卷曲占 25.48%。亲水性较强的区域主要位于 12aa-24aa, 42aa-52aa, 64aa-69aa, 81aa-90aa, 97aa-105aa, 126aa-150aa。EgFis1 含有 3 个优势 B 细胞表位, 7 个 T 细胞表位以及 2 个 T、B 联合表位。结论 生物信息学方法预测 EgFis1 蛋白含有 4 个优势 B 细胞表位, 7 个 T 细胞表位以及 2 个 T、B 联合表位, 可作为免疫治疗和药物治疗的靶点。

【关键词】 细粒棘球绦虫; 生物信息学; 抗原表位; Fis1 蛋白

【中图分类号】 R383.33

【文献标识码】 A

【文章编号】 1673-5234(2022)02-0191-04

[Journal of Pathogen Biology. 2022 Feb;17(2):191-194.]

Bioinformatics analysis of the antigen FIS1 of *Echinococcus granulosus*

XIAO Jing^{1,2,3}, QU Chen⁴, ZHAO Jia-qing^{1,2,3} (1. Basic Medical College of Ningxia Medical University, Yinchuan, Ningxia 750004, China; 2. Science and technology research center of Ningxia Medical University Yinchuan; 3. Ningxia Key Laboratory of Prevention and Control of Common Infectious Diseases; 4. Clinical College of Ningxia Medical University)

【Abstract】 **Objective** The antigenic epitopes of Fis1 (EgFis1) protein of *Echinococcus granulosus* were analyzed by bioinformatics method, which laid a theoretical foundation for the development of molecular peptide vaccines. **Methods** The amino acid sequence of EgFis1 protein was examined and downloaded from the NCBI database, the physicochemical properties of the protein were predicted using EXpasy software, the signal peptide and transmembrane structural domains of EgFis1 protein were predicted using SignalP5.0 and TMHMM sever 2.0 software, the secondary and tertiary structures of EgFis1 protein were predicted using SOPMA and SWISS-MODLE, the T and B cell epitopes of EgFis1 protein were predicted using IEDB, ABCpred and SYFPEITHI databases. The T and B cell epitopes of EgFis1 protein were predicted using IEDB, ABCpred and SYFPEITHI databases. **Results** *E. granulosus* EgFis1 is a 157 amino acid protein with isoelectric point of 5.86 and molecular weight of 16.93 ku. It does not contain signal peptide sequence, but has a transmembrane domain. Its secondary structure consists of α -helices (48.41%), extended strands (19.11%), β -turns (7.01%) and random coils (25.48%). The hydrophilic regions were mainly 12aa-24aa, 42aa-52aa, 64aa-69aa, 81aa-90aa, 97aa-105aa and 126aa-150aa. EgFis1 contains three dominant B cell epitopes, seven T cell epitopes and two T, B combined epitopes. **Conclusion** EgFis1 protein contains four dominant B cell epitopes, seven T cell epitopes and two combined T and B epitopes, which can be used as a target for immunotherapy and drug therapy.

【Key words】 *Echinococcus granulosus*; bioinformatics; antigenic epitope; Fis1 protein

***细粒棘球绦虫病是由细粒棘球绦虫的幼虫感染人体所引起的一种危害严重的寄生虫病^[1-2]。宁夏回族自治区也是细粒棘球绦虫病主要流行的地区之一^[3-4]。该病原体进入人体, 主要寄生在肝脏、肺脏等器官, 引起一系列症状。目前, 该类疾病主要以手术治疗和药物治疗为主, 但手术治疗创伤性大, 费用高并且术后易复发。药物治疗极易出现耐药性而新药的研发缓慢^[5]。

疫苗是预防和控制传染病最经济最有效的手段^[6], 但

* **【基金项目】** 宁夏回族自治区重点研发项目 (No. 2020BEG03014)。

** **【通讯作者】** 赵嘉庆, E-mail: zhaojq@nxmu.edu.cn

【作者简介】 肖静 (1995-), 女, 湖北人, 硕士研究生在读。研究方向: 感染与免疫。E-mail: 13997817310@163.com

目前尚无预防人体感染的疫苗。随着分子生物学和生物信息学的快速发展,棘球绦虫的转录组数据逐渐公布,使得筛选具有保护力的抗原成为可能,但将全抗原作为疫苗使用同样也面临许多问题^[2]。因此,分子肽疫苗成为了当前科研的热点。

线粒体裂变蛋白 1(Fis1)介导着线粒体裂变,在线粒体膜崩解过程中发挥关键作用。线粒体融合和分裂过程对于维持细胞器的形态和分布,以及细胞的增殖和分化至关重要^[7]。Fis1 能够促进细胞凋亡^[8],但是 Fis1 在细粒棘球绦虫感染中的作用目前尚不明确。本研究利用生物信息学的方法对 EgFis1 蛋白进行分析,旨在了解 EgFis1 在细粒棘球绦虫免疫诊断方面的价值。

材料与方法

1 氨基酸序列获取

通过 NCBI(<https://www.ncbi.nlm.nih.gov/>)在线数据库检索并获取 EgFis1 蛋白的氨基酸序列,其登录号为 CDS20063.1。

2 蛋白质理化性质的预测

利用在线预测软件 EXpasy 里的 Protparam(<https://web.expasy.org/protparam/>)工具预测 EgFis1 蛋白质的理化性质;利用在线预测软件 ProtScale(<https://web.expasy.org/protscale/>)中的 Hphod./Hoop& Woods 算法预测蛋白质的亲/疏水性。

3 信号肽及跨膜结构域的预测

利用在线预测软件 SignalP5.0(<http://www.cbs.dtu.dk/services/SignalP/>)预测该蛋白是否含有信号肽序列;使用在线预测软件 TMHMM server 2.0(<http://www.cbs.dtu.dk/services/TMHMM/>)预测该蛋白是否含有跨膜结构域。

4 蛋白空间结构的预测

利用 Prabi 在线预测软件的 SOPMA(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)程序预测该蛋白的二级结构。使用 SWISS-MODEL(<https://swissmodel.expasy.org/interactive#sequence>)预测其三级结构。

5 抗原表位的预测

使用 ABCpred 和 IEDB 数据库预测蛋白的优势 B 细胞表位;使用 IEDB 和 SYFPEITHI 数据库预测蛋白的优势 T 细胞表位。

结果

1 EgFis1 的理化性质

Eg Fis1 蛋白由 157 个氨基酸组成,分子质量单位为 16.93 ku,等电点是 5.86,半衰期为 30h(mammali-

an)>20 h(yeast)>10 h(*Escherichia coli*),分子式为 C₇₄₂H₁₂₀₃N₂₀₅O₂₂₈S₉。

2 信号肽和跨膜结构域

该蛋白不含有信号肽序列,但含有一个跨膜结构域,位于 151-157aa 之间。

3 蛋白的亲/疏水性

利用在线预测软件 ProtScale 中的 Hphod./Hoop& Woods 算法预测 EgFis1 蛋白的亲/疏水性,结果如图 1。该蛋白在第 84 个氨基酸异亮氨酸处亲水性最强,亲水性最大值为-1.278。在第 118 位氨基酸赖氨酸处的疏水性最强,最大值为 2.078。EgFis1 蛋白亲水性较强的区域主要位于 12aa-24aa,42aa-52aa,64aa-69aa,81aa-90aa,97aa-105aa,126aa-150aa。在亲水性较强的区域可能有抗原表位存在。

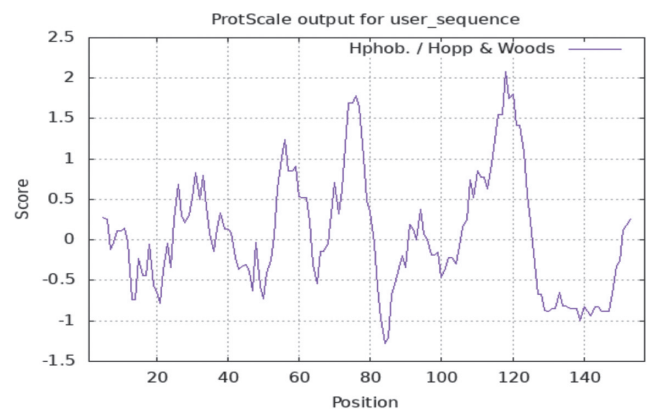


图 1 Eg Fis1 蛋白质亲/疏水性预测
Fig. 1 The hydrophilicity and hydrophobicity map of EgFis1 protein

4 蛋白的空间结构

4.1 二级结构 利用 Prabi 在线预测软件的 SOPMA 程序预测该蛋白的二级结构,结果如图 2,α-螺旋所占的比例为 48.41%,延伸连占 19.11%,β-转角占 7.01%,无规则卷曲占 25.48%。



图 2 EgFis1 蛋白的二级结构预测
Fig. 2 Prediction results of EgFis1 secondary structure

4.2 三级结构 利用 SWISS-MODLE 预测蛋白质的三级结构,结果如图 3。

5 抗原表位

5.1 优势 B 细胞表位 利用 ABCpred 软件预测 Eg Fis1 蛋白的 B 细胞表位,阈值>0.5 的 B 细胞表位肽有 5 条,分别位于 12aa-28aa,140aa-156aa,68aa-78aa,

102aa-118aa 和 30aa-46aa。经过 IEDB 数据库预测氨基酸长度 >6aa 的表位肽, 总共有 5 条, 分别位于 5aa-21aa, 33aa-43aa, 71aa-76aa, 104aa-110aa 和 118aa-124aa。由此可得优势 B 细胞表位肽有 4 条, 分别是 12aa-21aa, 33aa-43aa, 71aa-76aa 和 104aa-110aa (表 2)。

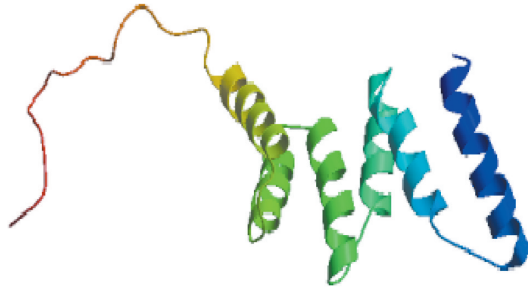


图 3 EgFis1 蛋白的三级结构预测
Fig. 3 The three-level structure of EgFis1

表 1 EgFis1 的理化性质
Table 1 Physicochemical properties of EgFis1

EgFis1 的理化性质 Physicochemical properties of EgFis1	预测结果 Predicted results
sequence	MELLDLNPEVTCARVMWGDYCDVYFQ KSREAYVAEGKAGFSNYDTQISYALDLL KTKDENALMTAESLLKIADDAKKRECLF LIALANTKMGAI ECDNLLAVNPQDHQS SDLKAEIKRRRRPGLTGLGAVGLAVGGV ALGLIGAAAAAYGIQKKGKQ
Number of amino acids	157
Molecular weight	16932.49
Theoretical pI	5.86
Formula	C ₇₄₂ H ₁₂₀₃ N ₂₀₅ O ₂₂₈ S ₉
The estimated half-life	30h>20h>10h
The instability index	43.07
Aliphatic index	95.22

表 2 优势 B 细胞表位序列
Table 2 The sequence of dominant B cell epitopes

序号 Order	氨基酸的位置 The position of amino acid	氨基酸的序列 The sequence of amino acid
1	12aa-21aa	CARVMWGDYC
2	33aa-43aa	VAEGKAGFSNY
3	71aa-76aa	IADDAK
4	104aa-110aa	NPQDHQS

5.2 T 细胞表位的预测

5.2.1 CD4⁺T 细胞表位 使用 SYFPEITHI 程序预测 HLA-DRB1 * 0401 的 Th 细胞表位, 筛选出评分大于 20 的表位肽总共有 7 条, 分别是 80aa-94aa, 40aa-54aa, 46aa-60aa, 79aa-93aa, 98aa-112aa, 128aa-142aa 和 137aa-151aa; 预测 HLA-DRB1 * 0701 的表位总共有 11 条, 分别是 44aa-58aa, 40aa-54aa, 50aa-64aa, 83aa-97aa, 14aa-28aa, 51aa-65aa, 58aa-72aa, 59aa-63aa, 80aa-94aa, 81aa-95aa 和 140aa-154aa。通过

IEDB 数据库预测 HLA-DRB1 * 0401 的 Th 细胞表位总共有 8 条, 分别是 85aa-99aa, 86aa-100aa, 72aa-86aa, 105aa-119aa, 87aa-101aa, 118aa-132aa, 104aa-118aa 和 106aa-120aa; 预测 HLA-DRB1 * 0701 的表位总共有 16 条, 分别是 106aa-120aa, 68aa-82aa, 69aa-83aa, 110aa-124aa, 108aa-122aa, 109aa-123aa, 93aa-107aa, 102aa-116aa, 107aa-121aa, 92aa-106aa, 90aa-104aa, 101aa-115aa, 91aa-105aa, 88aa-102aa, 89aa-103aa 和 33aa-47aa。综合上述两数据库预测结果, EgFis1 的 HLA-DRB1 * 0401 优势 CD4⁺T 表位肽是 85aa-94aa, 98aa-100aa 和 104aa-112aa; HLA-DRB1 * 0701 优势 CD4⁺T 细胞表位是 40aa-47aa, 88aa-97aa, 68aa-72aa 和 80aa-83aa (表 3, 表 4)。

表 3 EgFis1 蛋白的 HLA-DRB1 * 0401 限制性 CD4⁺T 细胞表位
Table 3 HLA-DRB1 * 0401 restricted CD4⁺T cell epitopes
of egfis1 protein

序号 Order	氨基酸的位置 The position of amino acid	氨基酸的序列 The sequence of amino acid
1	85aa-94aa	ALANTKMGAI
2	98aa-100aa	DNL
3	104aa-112aa	NPQDHQSSD

表 4 EgFis1 蛋白的 HLA-DRB1 * 0701 限制性 CD4⁺T 细胞表位
Table 4 HLA-DRB1 * 0701 restricted CD4⁺T cell epitopes
of egfis1 protein

序号 Order	氨基酸的位置 The position of amino acid	氨基酸的序列 The sequence of amino acid
1	40aa-47aa	FSNYDTQI
2	88aa-97aa	NTKMGAI EEC
3	68aa-72aa	LLKIA
4	80aa-83aa	CLFL

5.2.2 CD8⁺T 细胞表位 利用在线网站 SYF-PEITHI 预测 EgFis1 蛋白的 HLA-A0201 限制性 CD8⁺T 细胞表位(9aa), 筛选出评分大于 23 的表位总共有 6 条, 分别位于 132aa-141aa, 125aa-134aa, 61aa-70aa, 63aa-72aa, 127aa-136aa 和 15aa-34aa(表 5)。

表 5 EgFis1 蛋白的 HLA-A0201 限制性 CD8⁺T 细胞表位
Table 5 HLA-A0201 restricted CD8⁺T cell epitopes
of egfis1 protein

序号 Order	氨基酸的位置 The position of amino acid	氨基酸的序列 The sequence of amino acid
1	132aa-141aa	GLAVGGVALG
2	125aa-134aa	TLGLGAVGLA
3	61aa-70aa	ALMTAESLLK
4	63aa-72aa	MTAESLLKIA
5	127aa-136aa	GLGAVGLA
6	15aa-34aa	VMWGDYCDVY

5.2.3 T、B 联合细胞表位 结合上述 T、B 细胞表位的预测, 分析得出 T、B 联合细胞表位总共有 2 条, 位于 104aa-110aa 和 40aa-43aa 两处(表 6)。

表 6 EgFis1 蛋白的 T、B 联合细胞表位
Table 6 T and B combined cell epitopes of EgFis1 protein

序号 Order	氨基酸的位置 The position of amino acid	氨基酸的序列 The sequence of amino acid
1	104aa-110aa	NPQDHQS
2	40aa-43aa	FSNY

讨 论

Fis1 是一种介导线粒体裂变的蛋白质,在线粒体膜的崩解过程中发挥着主要作用,同时参与 Dnm1 复合物的正确组装、膜分布和功能^[7]。在哺乳动物中, Fis1 主要参与细胞增殖、分化以及细胞凋亡,但在寄生虫中的作用尚不清楚。

随着科学的发展,生物信息学也成为了科研中重要的一个环节。它是由生物学,计算机和信息学三门学科相互融合所形成的学科^[9]。截至目前,生物信息学已广泛用于药物的设计和开发,抗原表位的预测以及蛋白质结构的预测等方面。随着研究的深入,越来越多的数据库建立并供利用,使得利用生物信息学的方法筛选抗原表位成为了可能。

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,可以分为 T 细胞表位和 B 细胞表位^[10,11]。B 细胞表位一般位于蛋白表面,大多数都是亲水性较强的分子。T 细胞表位又可以分为线性表位和构象表位,一般位于蛋白质的内部^[12]。

本研究通过在线数据库预测 EgFis1 蛋白的优势 B 细胞表位肽有 4 条,位于 12aa-21aa, 33aa-43aa, 71aa-76aa 和 104aa-110aa 氨基酸之间。而这些 B 细胞表位与抗原指数较高、亲水性较强的区域相互对应。预测 CD4⁺ T 细胞表位有 7 条,位于 40aa-47aa, 88aa-97aa, 68aa-72aa 和 80aa-83aa, 以及 85aa-94aa, 98aa-100aa 和 104aa-112aa 氨基酸之间。预测 T、B 联合细胞表位肽有 2 条,位于 104aa-110aa 和 40aa-43aa 氨基酸之间。蛋白质空间结构预测结果显示,在 EgFis1 的二级结构中存在比例较高的 β -转角以及无规则卷曲等结构,而这些结构位于抗原的表面,容易形成抗原表位^[13-14],这也进一步提示 EgFis1 具有发展成疫苗的潜

在可能。

【参考文献】

[1] 王芬,陈林,赵明才,等. 细粒棘球绦虫 Calpain 蛋白的生物信息学分析[J]. 中国病原生物学杂志,2019,14(10):1171-1177,1193.

[2] 李超群,樊海宁. 细粒棘球绦虫延伸因子 1 生物信息学分析[J]. 中国公共卫生,2017,33(4):607-610.

[3] Wang W, Zhou X, Cui F, et al. Proteomic analysis on exosomes derived from patients' sera infected with *Echinococcus granulosus* [J]. Korean J Parasitol, 2019, 57(5): 489-497.

[4] Cadavid Restrepo AM, Yang YR, McManus DP, et al. Spatial prediction of the risk of exposure to *Echinococcus* spp. among school-children and dogs in Ningxia Hui Autonomous Region, People's Rep of China [J]. Geospat Health, 2018, 13(1): 644.

[5] Brunetti E, Kern P, Vuitton DA et al. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans [J]. Acta Trop, 2010, 114(1): 1-16.

[6] Pourseif MM, Yousefpour M, Aminianfar M, et al. A multi-method and structure-based in silico vaccine designing against *Echinococcus granulosus* through investigating enolase protein [J]. Bioimpacts, 2019, 9(3): 131-144.

[7] Wang N, Zhan J, Guo C, et al. Molecular Characterisation and Functions of Fis1 and PDCD6 Genes from *Echinococcus granulosus* [J]. Int J Mol Sci, 2018, 19(9): 2669.

[8] Stojanovski D, Koutsopoulos OS, Okamoto K, et al. Levels of human Fis1 at the mitochondrial outer membrane regulate mitochondrial morphology [J]. J Cell Sci, 2004, 117(Pt 7): 1201-1210.

[9] Chen C, Huang H, Wu CH. Protein bioinformatics databases and resources [J]. Methods Mol Biol, 2017(1558): 3-39.

[10] 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.

[11] 王芬,胡容,李龙,等. 细粒棘球绦虫胰岛素受体蛋白的生物信息学分析[J]. 中国病原生物学杂志,2020,15(11):1295-1298.

[12] Lian Y, Ge M, Pan XM. EPMLR: sequence-based linear B-cell epitope prediction method using multiple linear regression [J]. BMC Bioinformatics, 2014, 15(1): 414.

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

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

【收稿日期】 2021-09-11 【修回日期】 2021-11-21