

DOI:10.13350/j.cjpb.230213

• 论著 •

细粒棘球绦虫 EGR_09314 蛋白的生物信息学分析*

殷广琦¹,董丹¹,于学婷¹,陈雪玲¹,吴向未²,王仙^{1**}

(1.石河子大学医学院免疫学教研室,新疆石河子 832002;2.石河子大学医学院第一附属医院普外科)

【摘要】 目的 利用生物信息学的方法预测、分析细粒棘球绦虫 EGR_09314 蛋白的结构、功能和生物学特性等,为包虫病分子肽疫苗的筛选奠定基础。方法 从 NCBI 数据库中下载 EGR_09314 基因的核苷酸序列及其所编码的氨基酸序列,利用 ORF Finder 在线工具分析该基因的开放阅读框;利用 Prot-Param 预测蛋白的理化性质,ProtScale 预测其亲水性和疏水性,NetPhos 3.1 Server 预测其磷酸化位点,NetNGlyc 预测其糖基化位点,GPS-SUMO 2.0 预测其棕榈酰化修饰位点;利用 SOPMA 预测蛋白的二级结构,SWISS-MODEL 预测蛋白的三级结构;利用 Euk-mPLoc 2.0 分析其亚细胞定位,SignalP 4.1Server 分析其信号肽位置,TMHMM 分析跨膜区域,SMART 预测其结构域位置;应用 ABCpred 和 IEDB 预测 EGR_09314 蛋白的 B 细胞抗原表位,应用 SYFPEITHI 和 IEDB 预测其 T 细胞抗原表位。结果 细粒棘球绦虫 EGR_09314 蛋白是由 491 个氨基酸组成的亲水性蛋白,其分子式为 $C_{2446}H_{3895}N_{719}O_{733}S_{33}$ 。该蛋白含有信号肽,且含有 1 个跨膜区域,位于第 3-25 位氨基酸,该蛋白定位于细胞膜及细胞外。二级结构中 α 螺旋结构占 35.03%,延伸链结构占 13.85%, β 折叠结构占 5.30%,无规则卷曲结构占 45.82%。EGR_09314 蛋白存在 45 个磷酸化位点,3 个糖基化位点,6 个棕榈酰化修饰位点。EGR_09314 蛋白有 8 个 B 细胞优势表位,8 个细胞毒性 T 细胞(CTL)优势表位和 8 个辅助性 T 细胞(Th)优势表位。结论 生物信息学预测 EGR_09314 蛋白可能是一种跨膜蛋白,也具有分泌性蛋白的特点,存在多个 B、T 细胞表位,为该蛋白的基因克隆表达及细粒棘球绦病的诊断或疫苗候选抗原筛选提供了理论基础。

【关键词】 细粒棘球绦虫;EGR_09314;生物信息学;抗原表位

【中图分类号】 R383.33

【文献标识码】 A

【文章编号】 1673-5234(2023)02-0190-07

[Journal of Pathogen Biology. 2023 Feb;18(2):190-196.]

Bioinformatic analysis of EGR_09314 of *Echinococcus granulosus*

YIN Guang-qi¹, DONG Dan¹, YU Xue-ting¹, CHEN Xue-ling¹, WU Xiang-wei², WANG Xian¹ (1. Department of Immunology, School of Medicine, Shihezi University, Shihezi 832000, Xinjiang, China; 2. Department of Hepatobiliary Surgery, the First Affiliated Hospital, School of Medicine, Shihezi University)***

【Abstract】 **Objective** To bioinformatically analyze the the structure, function and biological characteristics of *E. granulosus* EGR_09314 in order to lay the foundation for screening of molecular peptide vaccine against echinococcosis.

Methods The amino acid sequence EGR_09314 was downloaded from NCBI database. The physical and chemical properties of protein were predicted by Prot-Param, its hydrophobicity was predicted by PotScale, its phosphorylation sites were predicted by NetPhos 3.1 Server, its glycosylation sites were predicted by NetNGlyc, its palmitoylation modification sites were predicted by GPS-SUMO 2.0. Its secondary structure was predicted by SOPMA, its tertiary structure was predicted by SWISS-MODEL, its subcellular localization was predicted by SignalP 4.1Server, its transmembrane regions were predicted by TMHMM, its domains were predicted by SMART. Its B-cell epitopes were predicted by ABCpred and IEDB, and its T-cell epitopes were predicted by SYFPEITHI and IEDB. **Results** EGR_09314 consists of 491 aminoacides, and it is a hydrophilic protein with a signal peptide and a transmembrane region. Its molecular formula is $C_{2446}H_{3895}N_{719}O_{733}S_{33}$. It is located in cell membrane and extracellular. Its secondary structure consists of α -helices (35.03%), extended strands (13.85%), β -sheets (5.3%), and random coils (45.82%). There are 45 phosphorylation sites and 3 glycosylation sites in the secondary structure of this protein. It was predicted that the protein EGR_09314 had 8 dominant epitopes of B cells, 8 dominant epitopes of CTL cells and 8 Th cell epitopes. **Conclusion** Bioinformatics analysis showed that EGR_09314 protein has multiple B and T cell epitopes. It can provide a theoretical basis for gene cloning, expression of the protein, diagnosis of *E. granulosus* infection or vaccine candidate antigen.

【Key words】 *Echinococcus granulosus*; EGR_09314; bioinformatics; antigenic epitope

* **【基金项目】** 国家自然科学基金项目(No. 82060297, 81460246); 区域创新引导计划(No. 2021BB006); 青年创新培育人才项目(No. CXPY202003)。

** **【通讯作者】** 王 仙, E-mail: 307835271@qq.com

【作者简介】 殷广琦(1997-), 女, 山东菏泽人, 硕士研究生, 主要从事感染免疫研究。E-mail: 1453604081@qq.com

细粒棘球蚴病, 又称囊型包虫病 (cystic echinococcosis, CE), 是由细粒棘球绦虫 (*Echinococcus granulosus*, *E. g.*) 的中绦期幼虫寄生于人体的肝脏、肺脏等其他器官而引起的一种寄生虫病。该病高发于新疆、西藏等以畜牧业为主的地区, 带来极其严重的疾病负担和畜牧业损失。目前, 临床上以手术治疗为主^[1], 但术后常会引发再次感染, 而使用阿苯达唑药物治疗, 机体常伴随神经系统、心血管系统等发生不良反应^[2]。因此, 亟需寻找灵敏的早期诊断方法或有效的疫苗用于包虫病的早期防治。

抑制素 β A (INHBA) 是组成激活素和抑制素的亚基, 属于转化生长因子 (TGF- β) 超家族成员^[3], 与细胞生长、增殖、分化以及凋亡等生理功能密切相关^[4]。细粒棘球绦虫中存在的 INHBA 基因, 在细粒棘球蚴长期的慢性感染中的作用尚不清楚。本研究采用生物信息学的方法对 EGR_09314 基因编码的细粒棘球蚴 INHBA 蛋白的结构、功能及生物学特性进行分析, 为囊型包虫病的早期诊断和疫苗研制提供理论依据。

材料与方法

1 EGR_09314 氨基酸序列的获取及基因开放阅读框架分析

从美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) 的 Gene 检索目录中获取 EGR_09314 基因信息, 并下载核苷酸及其编码的氨基酸序列, 其氨基酸序列登录号为 XP_024347036。采用 NCBI 网站中的 ORF Finder (<http://www.ncbi.nlm.nih.gov/orffinder>) 分析 EGR_09314 基因的开放阅读框架。

2 EGR_09314 蛋白的基本理化性质分析

通过理化性质预测网站 Prot-Param (<http://web.expasy.org/ProtParam/>) 对 EGR_09314 基因编码氨基酸序列的理化性质, 包括氨基酸数目、分子式、分子质量、等电点、氨基酸组成等进行预测。

3 EGR_09314 蛋白的信号肽位置和跨膜区域分析及其亚细胞定位预测

利用 SignalP 4.1 Server 在线系统 (<http://www.cbs.dtu.dk/services/SignalP-4.1/>) 对 EGR_09314 蛋白的信号肽位置进行预测; 利用 TMHMM 2.0 在线系统 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>) 预测 EGR_09314 蛋白的跨膜区域; 利用 Euk-mPloc 2.0 在线系统 (<http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/>) 进行 EGR_09314 蛋白的亚细胞定位预测。

4 EGR_09314 蛋白的亲(疏)水性分析

利用亲水性及疏水性分析网站 PotScale

(<https://web.expasy.org/protscale/>) 分析。

5 EGR_09314 蛋白的糖基化位点、磷酸化位点及蛋白棕榈酰化修饰位点分析

利用 NetNGlyc-1.0 在线系统 (<http://www.cbs.dtu.dk/services/NetNGlyc/>) 对 EGR_09314 蛋白的糖基化位点进行预测; 利用 NetPhos 3.1 Server 在线系统 (<http://www.cbs.dtu.dk/services/NetPhos/>) 对 EGR_09314 蛋白的磷酸化位点进行预测; 利用 GPS-SUMO 2.0 在线系统 (<http://sumosp.biocuckoo.org/online.php>) 对 EGR_09314 蛋白的蛋白棕榈酰化修饰位点进行预测。

6 EGR_09314 蛋白空间结构的预测

利用 PRABI 数据库中的 SOPMA 在线系统 (http://npsa-prabi.ibcp.fr/cgi-bin/secpred_sopma.pl) 预测 EGR_09314 蛋白的二级结构; 利用 SWISS-MODEL 在线系统 (<http://swissmodel.expasy.org/>) 预测 EGR_09314 蛋白的三级结构。

7 EGR_09314 蛋白抗原表位的预测

利用 ABCpred 和 IEDB 数据库预测 EGR_09314 蛋白的优势 B 细胞表位; 利用 SYFPEITHI 和 IEDB 在线数据库预测 EGR_09314 蛋白的 T 细胞表位。

8 EGR_09314 蛋白的相互作用蛋白分析

将 EGR_09314 蛋白质的全长氨基酸序列上传至 STRING (<http://string-db.org/cgi/input.pl>) 在线软件, 预测 EGR_09314 的相互作用蛋白。

结果

1 EGR_09314 蛋白的基本信息及基因开放阅读框架分析

EGR_09314 蛋白由 491 个氨基酸组成, 其序列如下: MKCGIVLVALALLMLGSCPLIHALFRQPAIM DGTLLLEVPDQEEERMWVDPTPKIEDNDGNDV DDVGTKRTLEEAEEREKEAKRKADEEEEFERL IHIEKFKRTLLQRLHLSPPDFSHHGGMANRTH GRRVLRSLPLALQGRLNQMRAEDGMAEPPPD RTDERETLILLKHLHWRLPKVASATFGIEMADD IDPSRIQSAFLRFEHVEVWEIFMTPSEEEGKMTN AAVDQPSLEQYNNLTWMFERQPTKYTSLPAPT VIRRLGSPMERRRSGSIRIRPGRLAETFPVSPCG LVQVTFEISGSFAQWMSHRRRMLMRKLVRSI IVVCPDCSSHVDPVDVNGILEIHRNIVRRTRR SLDTNSSQHVPIGNPCSPKGHKFCCTQPFSLNL EDVGNWNNWILHPKTVEPNYCHGSCQADGIQKT PHSDLMHLYRSQNYDRLSEVQREAMLSCCHPV KMASTSVLYVDPDNELHMDTLHNIIVLECGCS。

EGR_09314 基因含有 3 个开放阅读框 (ORF), 其

中 ORF1 为最长的开放阅读框,起始密码子为 ATG,终止密码子为 TGA,编码 491 个氨基酸,与目的蛋白全长一致,表明 EGR_09314 蛋白全长可完全表达(图 1)。

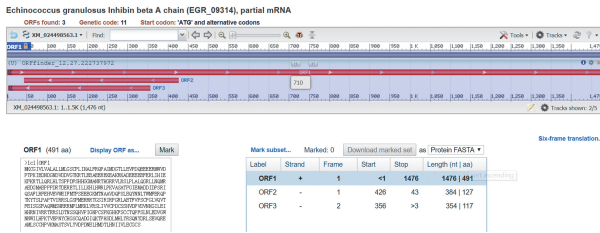


图 1 EGR_09314 基因开放阅读框分析

Fig. 1 Open reading frame analysis of EGR_09314 gene

2 EGR_09314 蛋白的基本理化性质

EGR_09314 蛋白的理论等电点 pI 为 6.13,相对分子质量为 56.16119×10^3 。该蛋白由 491 个氨基酸构成,包含 20 种氨基酸,其中以亮氨酸(10%)、谷氨酸(8.6%)、精氨酸(8.4%)、脯氨酸(6.9%)、丝氨酸(6.9%)5 种氨基酸的组成比例较高,不包括吡咯赖氨酸和硒半胱氨酸。带负电荷的残基(Asp+Glu)总数为 70,带正电荷的残基(Arg+Lys)总数为 60。EGR_09314 蛋白分子式为 $C_{2446}H_{3895}N_{719}O_{733}S_{33}$,原子总数 7826,EGR_09314 蛋白的消光系数为 48315,280 nm 处的吸光度为 0.860。当蛋白的 N 端残基为 Met 时,该蛋白在哺乳动物网状细胞中的半衰期为 30 h,在大肠埃希菌体内的半衰期 > 10 h。

3 EGR_09314 蛋白的信号肽、跨膜区及亚细胞定位

经 SignalP 4.1 Server 预测,EGR_09314 蛋白存在信号肽,位于第 23、24 位氨基酸之间(图 2)。Euk-mPLoc 2.0 预测 EGR_09314 蛋白位于细胞质、细胞外,所以推测该蛋白是分泌蛋白。TMHMM 分析 EGR_09314 蛋白在第 3-25 位氨基酸间可能存在跨膜结构,为跨膜蛋白,这与 SignalP 4.1 Server 预测结果一致(图 3)。

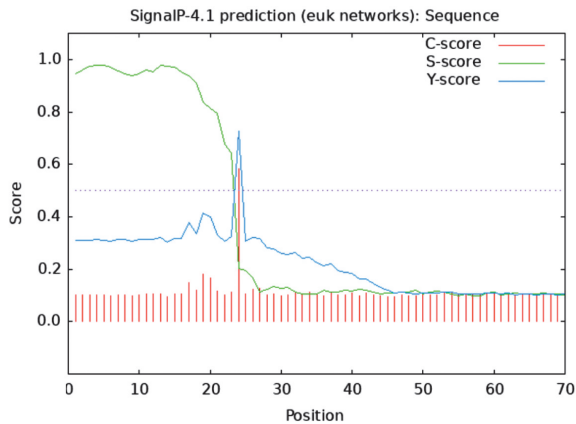


图 2 EGR_09314 蛋白信号肽分析

Fig. 2 Signal peptide analysis of EGR_09314 proteins

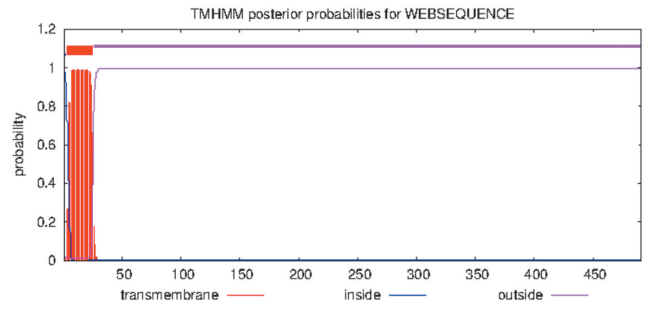


图 3 EGR_09314 蛋白跨膜区预测

Fig. 3 Prediction of EGR_09314 protein transmembrane structure

4 EGR_09314 蛋白的亲(疏)水性

PotScale 预测 EGR_09314 基因编码蛋白为亲水性蛋白,第 81 位氨基酸亲水性得分最高为 -3.267;第 9 位氨基酸疏水性得分最高为 3.522。该蛋白含有较多亲水性结构域,分别位于 39-50、53-94、123-132、146-166、235-253、267、285、311-318、404-416、424-447 位氨基酸,这些区域暴露在蛋白的表面,易形成抗原表位(图 4)。

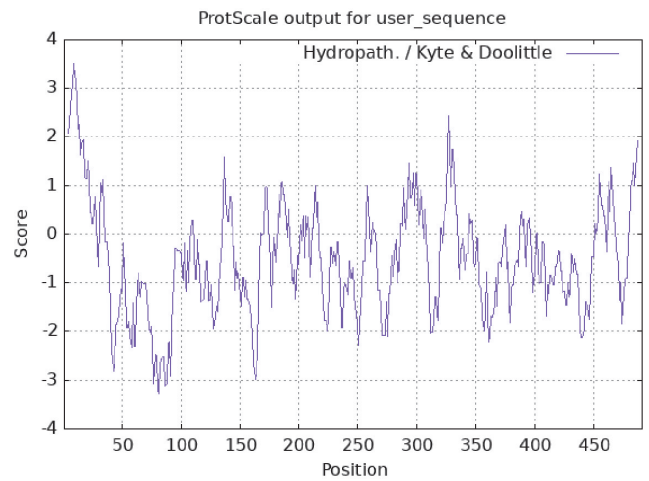


图 4 EGR_09314 蛋白亲/疏水性

Fig. 4 Hydrophilicity/ hydrophobicity of EGR_09314 protein

5 EGR_09314 蛋白的糖基化、磷酸化位点及棕榈酰化修饰位点

采用在线软件 NetNGlyc 分析 EGR_09314 的糖基化位点^[5],存在于 126、241 和 366 位氨基酸(图 5);经 NetPhos 3.1 预测分析,EGR_09314 蛋白存在 45 个磷酸化位点,包含 24 个丝氨酸(Ser)磷酸化位点(114、119、136、198、202、220、235、255、265、268、277、304、306、312、326、336、362、368、378、385、430、438、445、463),17 个苏氨酸(Thr)磷酸化位点(34、52、68、71、105、128、163、168、186、218、251、260、275、359、365、388、427)和 4 个酪氨酸(Tyr)磷酸化位点(239、254、414、468)(图 6)。这些位点可能介导体内相关信号通路中关键分子的去磷酸化过程。经 GPS-SUMO

2.0 预测分析^[6], EGR_09314 蛋白的棕榈酰修饰位点位于 54、169-173、296-300、327-331、347-351、483-487 残基。

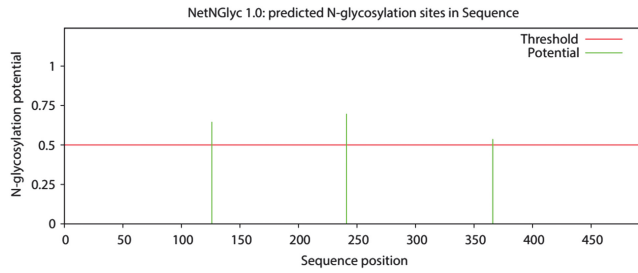


图 5 EGR_09314 蛋白的糖基化位点预测

Fig. 5 Analysis of glycosylation sites of EGR_09314 protein

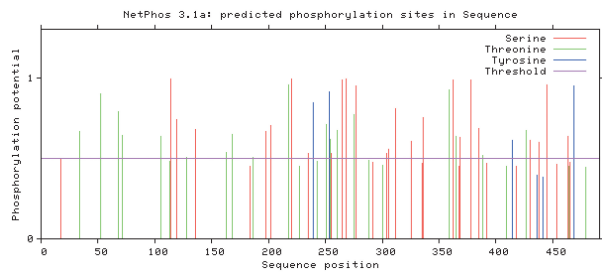


图 6 EGR_09314 蛋白的磷酸化位点预测

Fig. 6 Analysis of phosphorylation sites of EGR_09314 protein

6 EGR_09314 蛋白的二、三级结构

使用在线软件 SOPMA 预测 EGR_09314 蛋白的二级结构^[7], 其中 α -螺旋占 35.03%, 延伸链 13.8%, β -折叠 5.3%, 无规则卷曲 45.82% (图 7)。同时使用 SWISS-MODEL 软件对 EGR_09314 基因编码蛋白的三级结构进行预测, EGR_09314 蛋白主要结构为 α -螺旋及无规则卷曲 (图 8)。将其与蛋白的二级结构预测结果对比, 二者几乎完全一致。

7 柔性区域、抗原指数及表面可及性

利用 DNASTar 软件中的 Protean 程序对 EGR_09314 的抗原指数、柔性区域及表面可及性等代表性参数预测, 结果见图 9。以 Hopp 和 Woods 方案预测 EGR_09314 的亲水性区域位于 37-51, 52-98, 99-110, 110-138, 147-167, 217-256, 267-285, 309-322, 348-373, 376-386, 408-450, 469-479 位氨基酸; 用 Emini 蛋白可及性方案分析分值较高的区域位于 38-47, 50-62, 66-95, 155-167, 218-227, 232-245, 245-254, 256-269, 435-450 位氨基酸。经过 Karplus-Schulz 柔性区域的分析, 得出 EGR_09314 可塑性较强的区域位于 39-48, 51-93, 158-169, 219-228, 248-260, 265-279, 334-343, 358-371, 374-383, 423-432, 436-451 位氨基酸; 采用 Jameson-Wolf 方法预测 EGR_09314 的抗原性指数较高的表位区域位于 29-48, 51-98, 113-133, 150-168, 191-201, 218-240, 247-255, 261-286, 312-319, 332-

348, 351-369, 373-391, 405-432, 436-452, 469-477 位氨基酸。抗原指数得分较高的区域与柔性较高的区域分布一致。

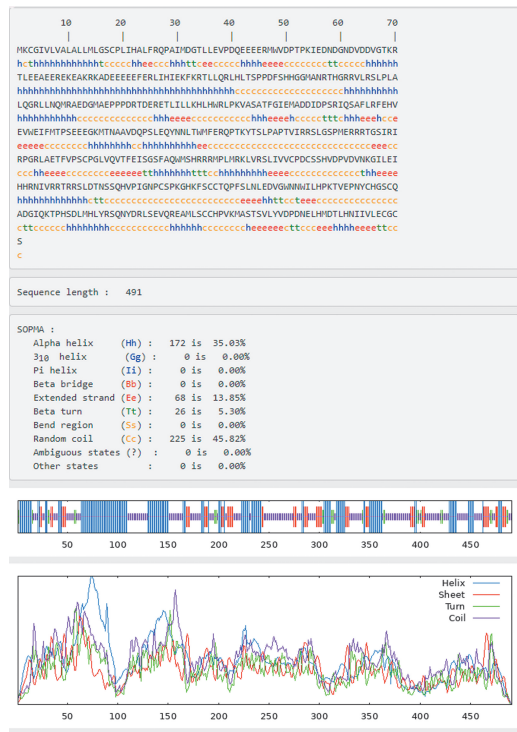


图 7 EGR_09314 蛋白的二级结构预测

Fig. 7 The secondary structure of EGR_09314 protein

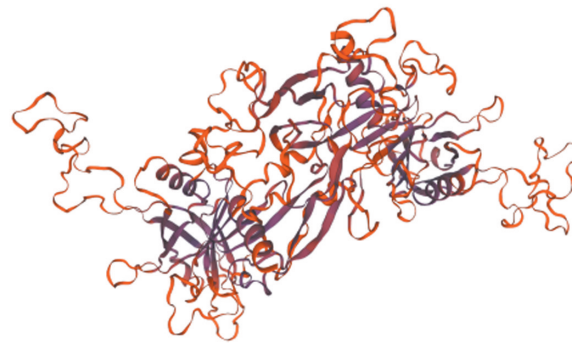


图 8 EGR_09314 蛋白的三级结构预测

Fig. 8 The three-level structure of EGR_09314 protein

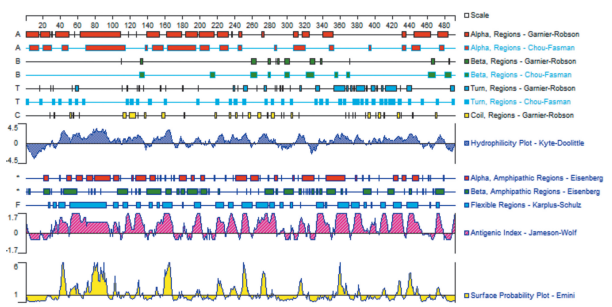


图 9 EGR_09314 蛋白的亲水性、柔性区域、抗原指数和表面可及性预测

Fig. 9 Hydrophilicity and flexible regions of EGR_09314 protein and prediction of antigen index and surface accessibility

8 EGR_09314 蛋白的 B 细胞表位

利用 IEDB 对 EGR_09314 进行 B 细胞抗原表位预测,根据 EGR_09314 蛋白的可塑性、表面可及性、线性表位、β-转角、表面可及性、亲疏水性、抗原 6 个方面筛选的 B 细胞表位肽有 9 条,分别位于第 39-45、51-56、157-166、219-225、234-240、247-254、269-274、357-367、437-443 位氨基酸。亲水性指数比较高的区域暴露于表面的几率较大,作为抗原表位的可能性也最大。利用 ABCpred 在线网站预测阈值 >0.6 的 B 细胞表位有 40 条,分别位于 456471、468-483、275-290、120-135、370-395、219-234、186-201、361-376、263-278、380-395、351-366、325-340、283-298、334-349、407-413、235-250、153-168、69-84、212-227、311-326、449-464、44-59、421-436、77-93、340-355、387-402、109-124、13-28、52-67、228-243、253-268、198-213、174-189、143-148、442-457、397-402、433-448、28-43、84-99、427-442。联合 DNASTar 及以上两数据库预测结果取交集得到优势 B 细胞表位有 8 条,分别位于 39-43、51-56、157-166、219-225、235-240、269-274、357-367、437-443 位氨基酸(表 1)。

表 1 EGR_09314 蛋白 B 细胞优势表位预测
Table 1 Prediction of dominant B cell epitope of EGR_09314 protein

序号 Order	氨基酸位置 The position of amino acids	表位序列 The position of amino acids
1	39-43	PDQEE
2	51-56	PTPKIE
3	157-166	EPPDRTDER
4	219-225	PSEEEGK
5	235-240	SLEQYN
6	269-274	PMERRR
7	357-367	RRTRRSLDTNS
8	437-443	RSQNYDR

9 EGR_09314 蛋白的 CD8⁺T 细胞表位

利用 SYFPEITHI 在线分析软件的表位预测功能分析蛋白的 MHC I 类 HLA-A0201 限制性(9aa)细胞毒性 T 细胞(CTL)表位,筛选出预测分值 ≥23 的表位肽,总共 12 条,其序列见表 2。利用 IEDB 在线网站预测 EGR_09314 的 HLA-A0201 限制性 CTL 表位,筛选出评分较高的细胞表位。与表 2 取交集,得到优势 HLA-A0201 限制性 CTL 表位位置为 30-38、174-181、13-20、322-330、133-141、204-212、133-141、451-459 位氨基酸。

10 EGR_09314 蛋白的 CD4⁺T 细胞表位

利用 SYFPEITHI 在线分析软件的表位预测功能分析蛋白的 MHC II 类 HLA-DRB1 * 0401 和 HLA-DRB1 * 0701 限制性(15aa)辅助性 T 细胞(Th)表位,筛选出预测分值 ≥23 的表位肽,分别有 16 条和 13 条,其序列见表 3、表 4。利用 IEDB 在线网站预测

EGR_09314 的 HLA-DRB1 * 0401 和 HLA-DRB1 * 0701 限制性 Th 表位,筛选出评分较高的细胞表位,与表 3 取交集,得到优势 HLA-DRB1 * 0401 限制性 Th 表位位置为 134-144、250-255 位氨基酸;与表 4 取交集,得到优势 HLA-DRB1 * 0701 限制性 Th 表位位置为 327-333、267-281、286-299、99-112、272-281、458-469 位氨基酸。

表 2 EGR_09314 蛋白的 HLA-A0201 限制性 CD8⁺T 细胞表位
Table 2 HLA-A0201 restricted CD8⁺T cell epitopes of EGR_09314 protein

序号 Order	氨基酸位置 The position of amino acids	表位序列 The position of amino acids	SYFPEITHI 得分 SYFPEITHI score (≥23)
1	30-38	IMDGTLLLEV	29
2	174-181	HLHWRLPKV	27
3	319-327	LMRKLVRSL	27
4	12-20	LLMLGSCPL	25
5	322-330	KLVRSLIVV	25
6	5-13	IVLVALALL	24
7	133-141	VLRSLPLAL	24
8	204-212	FLRFEHVEV	24
9	7-15	LVALALLML	23
10	13-21	LMLGSCPLI	23
11	295-303	GLVQVTFEI	23
12	451-459	AMLSCCHPV	23

表 3 EGR_09314 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 EGR_09314 protein

序号 Order	氨基酸位置 The position of amino acids	表位序列 The position of amino acids	SYFPEITHI 得分 SYFPEITHI score (≥23)
1	90-104	EEFERLIHIEKFKR	28
2	201-215	QSAFLRFEHVEVWEI	28
3	241-255	NLTWMFERQPTKYTS	28
4	9-23	ALALLMLGSCPLIHA	26
5	18-32	CPLIHALFRQPAIMD	26
6	44-58	EERMWVDPPTKIEDN	26
7	60-74	GNDVDDVGTKRRTLEE	26
8	96-110	LIHIEKFKRLLQRL	26
9	134-148	LRSPLALQGRLLNQ	26
10	317-331	MPLMRKLVRSLIVVC	26
11	327-341	LIVVCPDCSSHVDPV	26
12	344-358	NKGILEIHRNIVRR	26
13	429-443	HSDLMHLRYSQNYDR	26
14	441-455	YDRLSEVQREAMLSC	26
15	456-460	CHPVKMASTSVLYVD	26
16	474-488	ELHMDTLHNIIVLEC	26

11 EGR_09314 相互作用蛋白

STRING 在线分析软件预测的与 EGR_09314 (W6U417) 相互作用的蛋白有 Fibulin-2、W6UI50、W6UPM9 (Protein kinase C)、W6UDN2 (TKL Ser/Thr protein)、W6UEG8、W6USA6、W6UDU1、W6TYX3 和 W6UGG8(图 10)。

表 4 EGR_09314 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 EGR_09314 protein

序号 Order	氨基酸位置 The position of amino acids	表位序列 The position of amino acids	SYFPEITHI 得分 SYFPEITH score (≥23)
1	456-470	CHPVKMASTSVLYVD	30
2	99-113	IEKFKRTLQLRLHLT	26
3	9-23	ALALLMLGSCPLIHA	24
4	44-58	EERMWVDPTPKIEDN	24
5	212-226	VWEIFMTPSEEGKM	24
6	267-281	GSPMERRRTGSIRIR	24
7	269-283	PMERRRTGSIRIRPG	24
8	286-300	AETFPVSPGPLVQVT	24
9	298-312	QVTFEISGSFAQWMS	24
10	327-341	LIVVCPDCSSHVDPV	24
11	360-374	RRSLDTNSSQHVPIG	24
12	421-435	ADGIQKTPHSDLMHL	24
13	458-472	PVKMASTSVLYVDPD	24

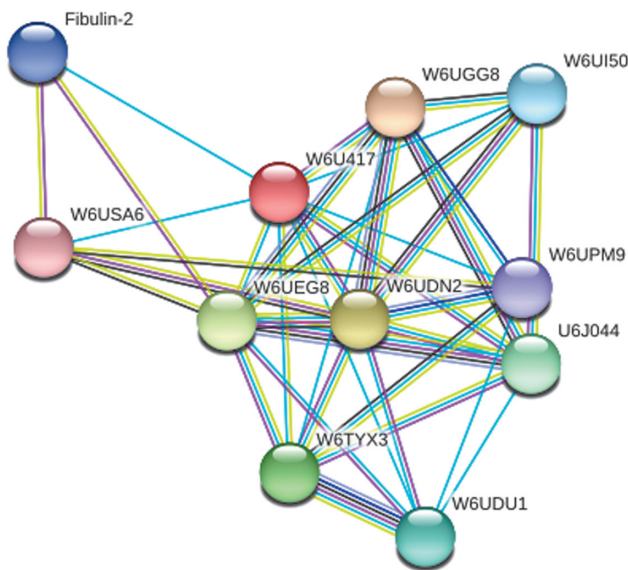


图 10 EGR_09314 蛋白的相互作用蛋白预测
Fig. 10 Interactive proteins of EGR_09314 protein

讨论

EGR-09314 基因编码的抑制素 β A 链 (INHBA) 属于转化生长因子超家族成员激活素的组成单体之一,其同源二聚体可构成激活素 A^[8],作为一种多功能细胞因子可通过旁分泌或自分泌的形式参与多种生物学过程。INHBA 可通过激活 TGF- β 信号通路促进胃癌的发展^[9],而且在肝损伤后激活素 A 通过诱导 Kupffer 细胞表达更高水平的肿瘤坏死因子- α 和转化生长因子- β ,从而引起肝星状细胞活化,导致肝纤维化的发生^[10]。除此之外,INHBA 基因的异常表达与多种癌症,如食管癌^[11]、结直肠癌^[12]及乳腺癌^[13]等的发生发展密切相关,但在细粒棘球蚴感染中的作用尚不清楚。

伴随着生物信息科学技术的发展,科研人员利用其对病原体的相关蛋白进行分析、处理和预测,越来越多的基因功能和关联蛋白进而被预测和发现^[14]。生物信息学的应用可减少开发疾病新抗原时的工作量和预算,同时还可提高工作效率,缩短发现新抗原表位的时间。细粒棘球蚴及其排泄分泌物 (E/S) 引发 Th1/Th2 相关细胞因子免疫失衡,逃避宿主的免疫攻击,以利于棘球蚴的生长、存活^[15],TGF- β /Smad 信号通路也在此发挥关键作用^[16]。而采用生物信息学方法可初步分析预测以上各成分的结构和生物学功能、筛选抗原表位,为疾病的临床治疗与防治提供新思路。

蛋白质经过磷酸化、糖基化、泛素化和脂质化等一系列复杂的修饰过程后才具有相应的生物活性或功能^[17]。本研究采用生物信息学方法对棘球蚴成虫 (细粒棘球蚴) INHBA 进行分析预测,结果显示 EGR_09314 基因含有 45 个磷酸化位点、3 个糖基化位点、6 个棕榈酰化修饰位点。这些分子的磷酸化、糖基化等修饰方式可辅助相关基因的转录、翻译,进而响应了寄生虫的需要。

多肽疫苗作为一种新型疫苗,是通过化学合成技术,按照病原体抗原基因中已知或预测的抗原表位序列制备的疫苗,较传统疫苗具有高效性、高特异性等优势^[18]。本研究通过 DNASTar 等软件预测 EGR_09314 蛋白优势 B 细胞抗原表位有 8 个,分别位于 39-43、51-56、157-166、219-225、235-240、269-274、357-367、437-443 位氨基酸。SYFPEITHI 及 IEDB 在线分析软件筛选出 EGR-09314 蛋白 Th 细胞优势表位有 8 个,其中,HLA-HLA-DRB1 * 0401 限制性 Th 表位有 2 个,分别位于 134-144、250-255 位氨基酸,HLA-DRB1 * 0701 限制性 Th 表位有 6 个,分别位于 327-333、267-281、286-299、99-112、272-281、458-469 位氨基酸。预测含有 1 个 T、B 细胞联合表位,位于第 269-274 位氨基酸。该蛋白有一个跨膜区,且有信号肽。该蛋白主要定位在细胞质、细胞外,其二级结构主要为 α -螺旋和无规则卷曲 (无规则卷曲占 45.82%),表明其容易与抗体嵌合,这对棘球蚴病疫苗的设计有重要意义。STRING 数据库分析显示 Eg-INH β (W6U417) 蛋白与 Fibulin-2、W6UI50、W6UPM9 (Protein kinase C)、W6UDN2 (TKL Ser/Thr protein)、W6UEG8、W6USA6、W6UDU1、W6TYX3 和 W6UGG8 有相互作用,提示 EGR_09314 蛋白可能诱导细胞膜上相应受体复合物并激活其丝氨酸/苏氨酸激酶 (TKL Ser/Thr protein) 活性,然后招募并活化下游的 Smad 蛋白信号通路,在各种人类疾病的纤维化发展中发挥重要作用^[19]。开发该蛋白相应受体靶向药物或抑制剂,可能为改善或治疗包虫病纤维化提

供可能性。

综上所述,EGR-09314 基因所编码的蛋白可能对细粒棘球蚴的生长、发育、寄生具有重要的调节作用,且含有多个 B 细胞及 T 细胞抗原表位,抗原性良好,为抗棘球绦虫感染药物的开发、疫苗的研制提供了信息基础。

【参考文献】

[1] Stojkovic M, Weber TF, Junghans T, et al. Clinical management of cystic echinococcosis: state of the art and perspectives[J]. Curr Opin Infect Dis, 2018, 31(5): 383-392.

[2] Mortezaei S, Afgar A, Mohammadi MA, et al. The effect of albendazole sulfoxide on the expression of miR-61 and let-7 in different in vitro developmental stages of *Echinococcus granulosus* [J]. Acta Trop, 2019(195): 97-102.

[3] Abdel MM, Pauklin S. TGFβ1/INHBA homodimer/nodal-SMAD2/3 signaling network: a pivotal molecular target in PDAC Treatment[J]. Mol Ther, 2021, 29(3): 920-936.

[4] Lyu S, Jiang C, Xu R, et al. INHBA upregulation correlates with poorer prognosis in patients with esophageal squamous cell carcinoma[J]. Cancer Manag Res, 2018(10): 1585-1596.

[5] 张萌萌, 代海兵, 王菡, 等. 原发性高血压与糖基化修饰相关研究进展[J]. 中华高血压杂志, 2022, 30(2): 133-137.

[6] 易李, 郑春福. 蛋白质棕榈酰化修饰在病毒感染过程中的作用[J]. 中国科学: 生命科学, 2020, 50(8): 836-848.

[7] 宋丹丹, 邓志东. RNA 二级结构预测的模糊模型[J]. 中国科学 (E 辑: 信息科学), 2007(10): 1285-1303.

[8] Bloise E, Ciarmela P, Dela Cruz C, et al. Activin A in mammalian physiology[J]. Physiol Rev, 2019, 99(1): 739-780.

[9] Chen ZL, Qin L, Peng XB, et al. INHBA gene silencing inhibits gastric cancer cell migration and invasion by impeding activation of the TGF-β signaling pathway[J]. J Cell Physiol, 2019, 234(10): 18065-18074.

[10] Kiagiadaki F, Kampa M, Voumvouraki A, et al. Activin-a causes hepatic stellate cell activation via the induction of TNFα and

TGFβ in Kupffer cells[J]. Biochim Biophys Acta Mol Basis Dis, 2018, 1864(3): 891-899.

[11] Kaneda H, Arai T, Matsumoto K, et al. Activin A inhibits vascular endothelial cell growth and suppresses tumour angiogenesis in gastric cancer[J]. Br J Cancer, 2011, 105(8): 1210-1217.

[12] He Z, Liang J, Wang B, et al. Inhibin, beta A regulates the transforming growth factor-beta pathway to promote malignant biological behaviour in colorectal cancer [J]. Cell Biochem Funct, 2021, 39(2): 258-266.

[13] Yu Y, Wang W, Lu W, et al. Inhibin β-A (INHBA) induces epithelial-mesenchymal transition and accelerates the motility of breast cancer cells by activating the TGF-β signaling pathway [J]. Bioengineered, 2021, 12(1): 4681-4696.

[14] 刘银凤, 张雷. 生物信息学数据库在医学研究中的应用[J]. 中国病原生物学杂志, 2014, 9(10): 961-962, 965.

[15] Pan W, Hao WT, Shen YJ, et al. The excretory-secretory products of *Echinococcus granulosus* protoscolices directly regulate the differentiation of B10, B17 and Th17 cells [J]. Parasit Vectors, 2017, 10(1): 348.

[16] Pang N, Zhang F, Li S, et al. TGF-β/Smad signaling pathway positively up-regulates the differentiation of Interleukin-9-producing CD4⁺ T cells in human *Echinococcus granulosus* infection[J]. J Infect, 2018, 76(4): 406-416.

[17] Linder ME, Deschenes RJ. Palmitoylation: policing protein stability and traffic[J]. Nat Rev Mol Cell Biol, 2007, 8(1): 74-84.

[18] Bhattacharya M, Sharma AR, Patra P, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach[J]. J Med Virol, 2020, 92(6): 618-631.

[19] Mendoza FA, Jimenez SA. Serine/threonine kinase inhibition as antifibrotic therapy: transforming growth factor-β and Rho kinase inhibitors [J]. Rheumatology (Oxford), 2022, 61(4): 1354-1365.

【收稿日期】 2022-09-22 【修回日期】 2022-12-16

(上接 189 页)

[14] Craig PS, Hegglin D, Lightowers MW, et al. Echinococcosis: Control and prevention[J]. Adv Parasitol, 2017(96): 55-58.

[15] Flisser A. Eliminating cystic echinococcosis in the 21st century [J]. Lancet Infect Dis, 2018, 18(7): 703-704.

[16] Norouzi R, Ataei A, Hejazy M, et al. Scolicidal effects of nanoparticles against hydatid cyst protoscolices *in vitro* [J]. Int J Nanomedicine, 2020(15): 1095-1100.

[17] Zheng H, Zhang W, Zhang L, et al. The genome of the hydatid tapeworm *Echinococcus granulosus* [J]. Nat Genet, 2013, 45(10): 1168-1175.

[18] Walker LS, Sansom DM. Confusing signals: recent progress in

CTLA-4 biology [J]. Trends Immunol, 2015, 36(2): 63-70.

[19] Zhang F, Li YH, Fan MW, et al. Enhanced efficacy of CTLA-4 fusion anti-carries DNA vaccines in gnotobiotic hamsters [J]. Acta Pharmacol Sin, 2007(28): 1236-1242.

[20] 赵商岐, 孔慧芳, 周彦霞, 等. 细粒棘球蚴 EgG1Y162-2 与 CTLA-4IgV-EgG1Y162-2 蛋白结构预测及表位分析比较[J]. 中国病原生物学杂志, 2021, 16(5): 546-551.

[21] Li SW, Yang KG, Zhao BF, et al. Epitope imprinting enhanced IMAC (EI-IMAC) for highly selective purification of His-tagged protein[J]. J Mater Chem B, 2016(4): 1960-1967.

【收稿日期】 2022-08-17 【修回日期】 2022-11-01