

文章编号:2095-7386(2016)04-0026-05

DOI:10.3969/j.issn.2095-7386.2016.04.005

Rv3457c 基因及其编码蛋白基本特性 及抗原表位的生物信息学分析

赵隆麒,王心情,孙妍,占卫红,吴启航,李海波,余晓丽
(武汉轻工大学 生物与制药工程学院,湖北 武汉 430023)

摘要:运用生物信息学对 Rv3457c 基因编码蛋白的主要特征进行分析,同时预测并筛选其 T/B 细胞表位。运用 ProtParam、SignalP、TMHMM 来分析基因编码蛋白的理化性质、信号肽区、跨膜区;运用 NetMHC、Bimas、NetCTL 和 SYFPEITHI 对其 T、B 细胞进行细胞表位分析预测。得到该基因编码蛋白由 1 044 个氨基酸组成,原子数为 10 498,理论分子式为 $C_{2980}H_{4918}N_{1044}O_{1216}S_{340}$,分子质量为 85.7284 KD,等电点的理论值为 4.99,半衰期估计值 4.4 h,不稳定系数 52.31,平均亲水性 0.938,脂肪系数 20.11;含有潜在 T 细胞表位和 B 细胞表位。得出至少含有一个 T 细胞和一个 B 细胞表位,可作为疫苗的研制提供基础。

关键词:Rv3457c;生物信息学;细胞表位

中图分类号:Q 93

文献标识码:A

Bioinformatics analysis to characteristics and antigen epitopes of Rv3457c

ZHAO Long-qi, WANG Xin-qian, SUN Yan, ZHAN Wei-hong, WU Qi-hang, LI Hai-bo, YU Xiao-li

(School of Biology and Pharmaceutical Engineering, Wuhan Polytechnic University, Wuhan 430023, China)

Abstract: Analyzing main features of gene Rv3457c by using bioinformatics, predicting and screening its T/B cell epitope. Using ProtParam, SignalP, TMHMM to analyze physical and chemical properties, signal peptide and transmembrane region of the protein coded by the gene. NetMHC, Bimas, NetCTL and SYFPEITHI are used to predict its T and B cell epitope. The protein coded by the gene consists of 1044 amino acid, an atomic number of 10498, a theoretical molecular formula of $C_{2980}H_{4918}N_{1044}O_{1216}S_{340}$, a molecular mass of 85.7284KD, a theoretical isoelectric point of 4.99, an estimated half-life of 4.4h, an instability index of 52.31, an average hydrophilicity of 0.938, an aliphatic index of 20.11. The gene has potential T and B cell epitope. There is one B and T cell epitope at least, which can be the basics of vaccine development.

Key words: Rv3457c; bioinformatics; cell epitope

1 引言

结核病是一种由结核分枝杆菌 (*Mycobacterium tuberculosis*) 造成的危害性极大的慢性传染病,通常

感染肺部和淋巴系统,但对身体的其他器官及系统(如脑,循环系统,中枢神经系统)也会造成感染。结核分枝杆菌对肺部的感染通常造成肺结核,主要症状为发热、咳嗽、咳血、胸痛和呼吸困难。患上结

收稿日期:2016-10-18.

作者简介:赵隆麒(19 -),男,硕士研究生,E-mail:q304451487@vip.qq.com.

通信作者:余晓丽(1963-),女,教授,E-mail:yxl268@126.com.

核病会对人体的活动能力有很大的影响。

我国是结核病高负担国家,全国的结核菌感染者超过 5 亿,结核病确诊患者多达 500 万,每年感染结核病的人数约为 130 万,结核病感染人数居于全球第二^[1]。我国的结核疫情存在地区差异,西部地区结核病患病率为中部的 1.7 倍;农村地区为城市的 1.6 倍^[2]。从二十世纪四十年代起,随着链霉素的问世,开启了结核病的药物治疗时代。随后氨基水杨酸,乙胺丁醇,异烟肼,利福平相继出现,为结核病的治疗提供更强有力的保障^[2]。但是药物的出现也同样造成了耐药菌的出现,结核病也出现了耐药结核菌(MDR-TB)和广泛耐药结核菌(XDR-TB),这使得结核病疫情又变得严峻起来。21 世纪以来随着各项技术的发展成熟,人们开始寻找新的靶细胞位点,以寻求对耐药结核病的有效治疗。抗原表位即抗原决定簇,是指位于抗原表面可决定抗原特异性的特定化学基团。抗原表位的存在使抗原能和 T、B 淋巴细胞上的抗原受体进行特异性的结合,从而引起机体 T、B 淋巴细胞发生免疫应答反应。人们通过对结核分枝杆菌的研究,试图通过其抗原表位特异性的特点来找到合适的靶细胞药物,作为结核病一种新的诊断和治疗制剂。

H37RV 一直以来作为结核分枝杆菌研究的标准株,在 1998 年得到其全基因组后,人们对其基因进行了一系列的分析研究,以寻求对结核病的防治方法。Rv3457c 是 H37RV 标准株上的一段基因序列,笔者通过对其进行一系列的生物信息学分析,来预测该段基因编码的蛋白是否具有合适的抗原表位以及可否用于结核病的诊断。

2 实验方法

运用生物信息学方法对 Rv3457c 的理化性质、信号肽区、跨膜区和 T、B 细胞表位进行分析和预测。

2.1 Rv3457c 理化性质分析

运用 ExPASy-ProtParam (<http://web.expasy.org/protparam/>) 程序测定基因编码蛋白的氨基酸组成、分子质量和等电点等理化性质。

2.2 Rv3457c 信号肽结构预测

利用 SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) 在线程序分析该基因编码蛋白是否含有信号肽及切割位点。

2.3 Rv3457c 跨膜区结构预测

采用 TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) 在线分析软件对其是否存在跨膜区

进行分析。

2.4 Rv3457c T 细胞表位预测

2.4.1 Rv3457c CD4⁺T 细胞表位预测

主要组织相容性复合体 II(MHCII)分布于 B 细胞、单核/巨噬细胞、树突状细胞及其他抗原递呈细胞(APC)上,将外源性抗原递呈给 CD4⁺T 细胞。人类的 MHCII 分子主要包括 HLA-DP、HLA-DQ、HLA-DR、HLA-N、HLA-O、HLA-M 等基因位点编码的所有膜抗原。在这里我们选用 HLA-DR 来进行预测,打开 NetMHC (<http://www.cbs.dtu.dk/services/NetMHCIIpan/>) 网站,将 Rv3457c 的氨基酸序列输入,设定氨基酸长度为 15^[3],选择等位基因(Select Allele)设定为 DRB1_0101、DRB1_0301、DRB1_0401、DRB1_0701、DRBA_0802、DRB1_0901、DRB1_1101、DRB1_1302、DRB1_1501,将排列方式选择为按亲和值(Affinity)排列,开始分析预测。

2.4.2 Rv3457c CD8⁺T 细胞表位预测

打开网站 NetMHC (<http://www.cbs.dtu.dk/services/NetMHC/>), 打开 NetMHC 登陆界面,此软件可分析多种 HLA 类的 MHC 分子。与 HLA 强结合的肽段数通常为 9 个,因此将氨基酸长度设定为 9,同时选择 HLA-A 0201 和 HLA-A 0301,即 A2 和 A3 作为限制条件^[4]。选择以亲和值(Affinity)大小排序,然后输入氨基酸序列开始分析。

进入 Bimas 网站 (http://www-bimas.cit.nih.gov/molbio/hla_bind/), 选择同上的 MHC 分子类型,9 个氨基酸,输入序列开始分析。

打开 NetCTL 的网址 (<http://www.cbs.dtu.dk/services/NetCTL/>), 选择同上 MHC 分子类型,选择分类标准为 Combined score,输入氨基酸序列开始分析。

打开 SYFPEITHI 网站 (<http://www.syfpeithi.de/#userconsent#>), 同样选择 HLA-A 0201 和 HLA-A 0301 为 MHC 类型,氨基酸长度为 9 个,输入序列,开始分析。

2.5 Rv3457c B 细胞表位预测

IEDB 对 B 细胞表位进行预测提供了 5 种参数,选取 Chou&Fasman 法预测氨基酸编码蛋白质的 β 转角(β -Turn);选取 Emini 法预测其表面可及性(Accessibility);选取 Karplus&Schulz 法预测其柔韧性(Flexibility);运用 Kolaskar&Tongaonkar 法预测其抗原性(Antigenicity);Parker 用来预测其亲水性(Hydrophilicity)。打开网站 IEDB (<http://tools.immuneepitope.org/main/bcell/>), 输入氨基酸序列,并

分别选择这 5 种方法分别进行分析。

Bcepred 对 B 细胞表位预测提供了 7 种参数,分别为采用 Parker 法预测其亲水性;Karplus 法预测其柔韧性;Emini 法预测其表面可及性;Pellequer 法预测 β 转角;Janin 法预测其暴露表面 (Exposed Surface);Ponnuswamy 法预测其极性 (Polarity);Kolas-kar 法预测其抗原倾向性 (Antegenic Propensity)。打开网站 Bcepred (http://www.imtech.res.in/raghava/bcepred/bcepred_submission.html),输入氨基酸序列,选定这 7 种参数,然后运行。

3 结果

3.1 Rv3457c 理化性质分析

经 ExPASy-Protparam 在线软件分析, Rv3457c 编码蛋白 1 044 个氨基酸组成,原子数为 10 498,理论的分子式为 $C_{2980}H_{4918}N_{1044}O_{1216}S_{340}$,分子质量单位为 85.728 4 KD,等电点的理论值为 4.99,半衰期估计值 4.4 h,不稳定系数 52.31,平均亲水性 0.938,脂肪系数 20.11。

3.2 Rv3457c 信号肽结构预测

采用 SignalP 程序预测该基因的信号肽,结果显示该基因信号肽的切割位点在第 21 和第 22 之间。

3.3 Rv3457c 跨膜区结构预测

采用 TMHMM 程序预测该基因的跨膜区,结果显示该基因不存在跨膜区结构。蛋白质在细胞外。

3.4 Rv3457c T 细胞表位预测

3.4.1 Rv3457c CD4⁺T 细胞表位预测

将预测结果汇总入表 1。由表中数据可知,DRB1_0301 中肽段的结合数及强结合肽段数都是最多的,在此之中肽段的可选性比较强,故从 DRB1_0301 中来选其中强结合的 11 个肽段 72-86VTEIILNLKSLVVSS, 71-84DVTEIILNLKSLVVS, 50-64VTSIRIDGVLHEFTT, 49-63AVTSIRIDGVLHEFT, 73-87TEIILNLKSLVVSSE, 70-84EDVTEIILNLKSLVV, 51-65TSIRIDGVLHEFTTV, 48-62AAVTSIRIDGVLHEF, 74-88EIIILNLKSLVVSSEE, 52-66SIRIDGVLHEFTTVP, 69-83KEDVTEIILNLKSLV。

表 1 Rv3457c 与 HLA 分子结合总肽段数汇总

人类白细胞抗原 (HLA) 分子亚型	Rv3457c 的总肽段数为 339	
	强结合肽段数	弱结合肽段数
DRB1_0101	9	15
DRB1_0301	11	16
DRB1_0401	1	18
DRB1_0701	5	20
DRB1_0802	0	18
DRB1_0901	9	8
DRB1_1101	0	14
DRB1_1302	5	17
DRB1_1501	5	11

3.4.2 Rv3457c CD8⁺T 细胞表位预测

将四个软件预测的得分较高的序列进行整理汇总,如表 2 所示。

表 2 CD8⁺T 细胞表位预测结果汇总

序列序号	序列	序列得分			
		NetMHC (Affinity)	Bimas (Score)	NetCTL	SYFPEITHI (Score)
76 — 84	IILNLKSLV	306.63	60.154	0.8060	23
300 — 308	KLHQLGLSL	337.43	74.768	1.0471	25
166 — 174	SIYSPVLKV	106.07	70.387	1.0839	29
1 — 9	MLISQRPTL	62.62	36.316	1.0391	25

结合表 1 和表 2 可知重复的肽段为 76—84 IILNLKSLV,故 T 细胞的预测表位为 76—84 IILNLKSLV。

3.5 Rv3457c B 细胞表位预测

根据软件预测得出的数据选取其中综合得分较高的几项,分别标号为 1,2,3,4。1 表示序列 86—

92 SSEEDEP,2 表示序列 98—104 RKQGPGE,3 表示序列 309—315 KDSPPSF,4 表示序列 337—343 DEQDYAE。统计的各个序列得分情况如表 3 所示。由表 3 中数据可知,序列 86—92 SSEEDEP 在 IEDB 和 Bcepred 中的综合得分更加高一些,因此更加适合做 B 细胞表位。

表 3 B 细胞表位预测结果汇总

参考参数	IEDB 得分				Bcepred 得分			
	1	2	3	4	1	2	3	4
Hyd	6.929	5.314	3.386	5.971	2.041	1.360	1.290	2.240
Access	4.689	3.463	2.417	4.464	1.954	1.910	1.946	2.319
Fle	1.089	1.110	1.067	1.053	0.871	1.403	1.518	0.556
Anti	0.930	0.926	1.006	0.953	1.046	0.935	1.391	-0.276
Turn	1.151	1.189	1.281	1.026	-0.083	-1.423	1.290	-2.066
Polarity	-	-	-	-	1.733	1.163	0.955	1.756
Exp	-	-	-	-	1.563	1.730	1.804	2.013

注:IEDB 不含有参数 Polarity 和 Exposed Surface 的预测;Bcepred 得分取 7 个氨基酸的平均值四舍五入。

4 讨论

在多年的研究中,人们已经发现了关于结核的很多特异性抗原^[4],以这些结核抗原作为对象进行分析研究,为结核疫苗的研究提供了基础。同时还有很多的细胞位点没有被发现,这些潜在的细胞位点很可能成为结核病快速诊断的一种方法。机体产生免疫反应的因素不仅仅只在于其含有编码抗体的基因,更重要的是其有与抗体结合的空间,能够结合在相应的位置等因素。

人体免疫的特异性免疫过程主要为细胞免疫,即 T 细胞免疫和 B 细胞免疫。这两个免疫过程需要一套完整的免疫应答过程才能完成。从抗原入侵到细胞呈递最后到产生抗体,只有经过这一系列的步骤才是一套完整的免疫过程。已有的研究证明,CD4⁺T 细胞所产生的细胞因子 IFN-γ 可以对小鼠的结核模具产生保护作用^[5],而且当 CD4⁺T 细胞有损坏时,艾滋病患者更易患上结核^[6]。同时,对小鼠模具的研究表明 CD8⁺T 细胞对其有保护作用^[7]。此类实验证明 CD4⁺T 细胞和 CD8⁺T 细胞对于结核病有一定的特异性反应。笔者所研究的结核分枝杆菌基因 Rv3457c 尚未有人进行研究,生物信息学进行的分析在某些方面只作为理论参考,还需要进一步的实验来证明其是否适合作为抗原表位。

笔者采用一系列的生物信息学方法,对 Rv3457c 基因做了基础分析,并预测了其 T、B 细胞抗原表位。在使用 NetMHCIIpan3.1 Server 对其 CD4⁺T 细胞进行预测时,选取肽段结合数最多的 DBR1_0301 中的强结合肽段作为备选,然后使用 NetMHC、Bimas、NetCTL 和 SYFPEITHI 四个软件对其 CD8⁺T 细胞进行预测,综合选择四个软件中得分较高的肽段,再与 CD4⁺T 细胞的强结合肽段进行对

比,综合选取重复片段即为预测的可能性较高的抗原表位。蛋白质的理化性质决定其是否具有此类空间即表位供抗体结合,因此对于蛋白质理化性质的分析至关重要。据研究发现可以通过利用蛋白质的理化性质来预测其 B 细胞表位。现在已被公认的预测单参数方案有 6 种,亲水性方案(Hydrophilicity)、表面可及性方案(Surface Accessibility)、柔韧性方案(Flexibility)、抗原性方案(Antigenicity)、二级结构预测方案(Secondary structure)、抗原指数方案(Antigenivity Index)。这些方案有较好的预测效果。这里我们根据 IEDB 和 Bcepred 里面的数据来进行分析。

参考文献:

[1] 马麦卷. 结核免疫应答相关基因多态性与肺结核易感性的关联研究[D]. 北京:中国人民解放军军事医学科学院,2012.

[2] 唐神结,肖和平. 结核病流行趋势及治疗未来展望[J]. 中国实用内科杂志,2012(8):565-568.

[3] 王心情,孙妍,占卫红,等. 结核分枝杆菌 Rv3607c T 细胞表位分布情况预测及分析[J]. 武汉轻工大学学报,2016(2):36-39.

[4] 叶娟,张舒林,刘文第. 结核分枝杆菌 RD12 区 T 细胞表位分布情况预测及分析[J]. 上海交通大学学报(医学版),2014(1):7-12.

[5] Rosa Daniela Santoro, Ribeiro Susan Pereira, Cunha-Neto Edecio. CD4⁺T cell epitope discovery and rational vaccine design[J]. Archivum Immunologiae et Therapiae Experimentalis, 2010,58(2):121-130.

[6] Oldstone M B. The role of cytotoxic T lymphocytes in infectious disease: history, criteria, and

- state of the art[J]. *Curr Top Microbiol Immunol*, 1994, 189: 1-8.
- [7] Wang Xinjing, Cao Zhihong, Jiang Jing, et al. Elevated expression of Tim-3 on CD8 T cells correlates with disease severity of pulmonary tuberculosis[J]. *Journal of Infection*, 2011, 62(4): 292-300.
- [8] 南文龙, 金宁一, 鲁会军, 等. H5N1 亚型禽流感病毒血凝素 Th 和 B 细胞表位预测及抗原性分析[J]. *中国免疫学杂志*, 2009(7): 630-633, 637.
- [9] 刘海灿. 结核分枝杆菌八种蛋白抗原中人 T/B 细胞抗原表位多态性研究[D]. 北京: 中国疾病预防控制中心, 2014.
- [10] Nobuhiro Nakamoto, David E Kaplan, Jennifer Coleclough, et al. Functional Restoration of HCV -Specific CD8 T Cells by PD-1 Blockade Is Defined by PD-1 Expression and Compartmentalization[J]. *Gastroenterology*, 2008 (7): 927-1937.
- [11] Li Fan, Yang Di, Wang Yi-qin, et al. Identification and modification of an HLA-A * 0201-restricted cytotoxic T lymphocyte epitope from Ran antigen[J]. *Cancer Immunology, Immunotherapy*, 2009(12): 2039-2049.
- [12] Grotzke J E, Lewinsohn D M. Role of CD8⁺ T lymphocytes in control of *Mycobacterium tuberculosis* infection [J]. *Microbes Infect*, 2005, 7(4): 776-788.
- [13] Mommaas B, Kamp J, Drijfhout JW, et al. Identification of a novel HLA- B60- restricted T cell epitope of the minor histocompatibility antigen HA-1 locus [J]. *J Immunol*, 2002, 169(6): 3131-3136.

(上接第 25 页)

量改变 Logistic 动力学参数随之改变, 这种动力学研究方法可以为固态发酵条件优化提供理论技术支持。

参考文献:

- [1] 邱立友. 固态发酵工程原理及应用[M]. 北京: 中国轻工业出版社, 2008.
- [2] Santos M M, Rosa A S, Dal' Bait S, et al. Thermal denaturation: is solid-state fermentation really a good technology for the production of enzyme [J]. *Bioresour Toehnlol*, 2004 (93): 261-268.
- [3] 郑建丰. WMC-15 黑曲霉固态发酵生产糖化酶的研究[J]. *江苏食品与发酵*, 1998(4): 2-5.
- [4] 张良均, 谢佳标, 杨坦, 等著. R 语言与数据挖掘[M]. 北京: 机械工业出版社, 2016.
- [5] 陈陶声. 酶制剂生产技术[M]. 北京: 化学工业出版社, 1994.
- [6] 苏畅, 夏文水, 姚惠源. 氨基葡萄糖和乙酰氨基葡萄糖的测定方法 [J]. *食品工业科技*, 2003(6): 75-77.
- [7] 陈书峰, 赵亮, 刘德华. 绿色木霉在稻壳和麸皮混合基质上固态发酵生产纤维素酶的研究 [J]. *食品与发酵工业*, 2004, 30(1): 8-12.
- [8] 路秀玲, 赵树欣. 固态发酵中生物量的测定方法[J]. *天津轻工业学院学报*, 2000(4): 57-62.
- [9] Weiss R M, Ollis D F. Extracellular microbial polysaccharides. I. Substrate, biomass, and product kinetic equations for batch xanthan gum fermentation[J]. *Biotechnology and Bioengineering*, 1980, 22(4): 859-873.