

DOI:10.13350/j.cjpb.230506

• 论著 •

肺炎链球菌 StkP 蛋白 B 细胞及 T 细胞抗原表位的生物信息学分析*

李莎莎¹, 王华东², 杨丹¹, 马红¹, 伏慧¹, 李刚^{1,3**}

(1. 宁夏医科大学总医院医学实验中心, 宁夏银川 750004; 2. 中国人民解放军联勤保障部队军第九四二医院放射诊断科; 3. 宁夏医科大学总医院医学科学研究院)

【摘要】 目的 运用生物信息学方法对肺炎链球菌 StkP 基因编码蛋白的理化性质、结构等进行分析, 并预测其潜在的 B 细胞以及 T 细胞抗原表位。方法 通过 NCBI 数据库获取 StkP 蛋白氨基酸序列的相关信息; 运用 ProtParam 软件分析 StkP 蛋白的理化性质; 采用 ProScale 软件预测 StkP 蛋白的亲水性、柔韧性、表面可及性、 β -转角; 使用 SOMPA 和 SWISS-MODEL 在线服务器分析 StkP 蛋白的二级结构并构建其三级结构; 分别采用 ABCpred 及 IEDB 软件综合预测其 B 细胞表位, 采用 SYFPEITHI 软件分析确定其 T 细胞表位。结果 StkP 蛋白由 659 个氨基酸组成, 理论 pI 值 8.61, 原子组成为 C₃₁₉₇ H₅₂₃₄ N₈₇₂ O₉₉₇ S₁₄, 不稳定指数为 40.13, 归类为不稳定蛋白, 平均亲水系数 (GRAVY): -0.294, 属于亲水性蛋白; StkP 蛋白的二级结构中 α 螺旋占 33.08%, 延伸链占 19.58%, β -转角占 7.74%, 无规卷曲占 39.61%。ABCpred 及 IEDB 软件预测 StkP 蛋白潜在的 B 细胞表位分别为 9 个和 7 个; SYFPEITHI 在线程序预测该蛋白含有 8 个 CTL 细胞表位和 13 个 Th 细胞优势表位。结论 生物信息学方法预测肺炎链球菌 StkP 蛋白为不稳定亲水性蛋白, 含有多个潜在的 B、T 细胞抗原表位, 为以 StkP 蛋白为基础的新型亚单位疫苗的研究提供了理论基础。

【关键词】 肺炎链球菌; StkP 蛋白; B 细胞表位; T 细胞表位

【中图分类号】 R378.1

【文献标识码】 A

【文章编号】 1673-5234(2023)05-0525-04

[Journal of Pathogen Biology. 2023 May;18(5):525-528,535.]

Prediction and analysis of B cell and T cell epitopes in the StkP protein of *Streptococcus pneumoniae* by bioinformatics

LI Shasha¹, WANG Huadong², YANG Dan¹, MA Hong¹, FU Hui¹, LI Gang^{1,3} (1. General Hospital of Ningxia Medical University, Yinchuan 750004, China; 2. The 942 th Hospital of the PLA Joint Logistic Support Force; 3. Institute of Medical Sciences, General Hospital of Ningxia Medical University)

【Abstract】 **Objective** The physicochemical properties and protein structure of *Streptococcus pneumoniae* StkP protein were analyzed by bioinformatics method, and B cell and T cell epitopes were predicted. **Methods** The amino acid sequence of StkP protein was obtained from the NCBI database. The physical and chemical properties of StkP protein were analyzed by using ProtParam online software. The secondary structure of the protein was analyzed by online software SOMPA. ProScale online software was used to predict the surface accessibility, β -turn, hydrophilic, flexibility. The tertiary structure of StkP protein was constructed using SWISS-Model online software. The online software ABCpred and IEDB were used to predict B cell epitopes. T cell epitopes were analyzed by online prediction software SYFPEITHI.

Results StkP protein is composed of 659 amino acids. It's a hydrophilic protein. The secondary structure of StkP protein was analyzed by SOMPA online software. α -helix for 33.08%, β -turns for 7.74%, random coils for 39.61%. ABCpred and IEDB software were used to predict the B-cell epitopes of StkP protein, and the number of potential epitopes were 9 and 7, respectively. SYFPEITHI online software predicted CTL and Th cell epitopes, and there were 8 CTL and 13 Th cell epitopes. **Conclusion** The B cell and T cell epitopes of *S. pneumoniae* StkP protein can be predicted by biological information method, which provides a theoretical basis for the study of novel subunit vaccine based on StkP protein.

【Key words】 *Streptococcus pneumoniae*; StkP protein; B cell epitope; T cell epitope

***肺炎链球菌 (*Streptococcus pneumoniae*, Sp) 是一种 α 溶血的革兰阳性菌, 是造成儿童和老年人发病和死亡的主要原因^[1]。该菌通常经过呼吸道进入人体, 可在鼻咽部、支气管或肺内定植。当人体免疫力低下时, 可引起大叶性肺炎、脑膜炎、中耳炎等多种感染,

* **【基金项目】** 国家科技基础资源调查专项 (No. 2019FY101200); 宁夏回族自治区重点研发计划项目 (No. 2021BEG03090)。

** **【通讯作者】** 李刚, E-mail: gone.lee@163.com

【作者简介】 李莎莎 (1984-), 女 (满族), 硕士研究生, 主管技师。研究方向: 病原微生物检验。E-mail: 18309679166@163.com

严重的甚至导致死亡^[2-4]。据 WHO 统计数据,每年全球约有 160 万人死于肺炎链球菌引起的疾病^[5]。随着抗生素的广泛使用以及肺炎链球菌的遗传变异等问题,越来越多的研究报道肺炎链球菌对 β-内酰胺类、大环内酯类等抗生素出现耐药^[6-9]。因此,研制和应用有效的肺炎链球菌疫苗成为预防肺炎链球菌感染的有效方法。

StkP, 即丝氨酸-苏氨酸蛋白激酶 (Serine/Threonine protein Kinase), 是由肺炎链球菌 *stkP* 基因编码的一种蛋白激酶。多项研究表明, StkP 基因能够调控肺炎链球菌的细胞分裂过程^[10-15]。同时, *stkP* 基因序列在多种血清型的肺炎链球菌中均显示高度保守, 满足了 StkP 蛋白作为候选疫苗的研究条件。本研究采用生物信息学方法对 *stkP* 基因编码的蛋白进行分析, 预测其结构和功能, 并进一步预测其可能的 B 细胞及 T 细胞抗原表位, 为疫苗研发奠定理论基础。

材料与与方法

1 StkP 蛋白的氨基酸序列检索

从美国国家生物技术信息中心数据库 (National Center for Biotechnology Information, NCBI) 中检索肺炎链球菌 (*Streptococcus pneumoniae*) StkP 蛋白的氨基酸序列。

2 StkP 蛋白的理化性质预测

使用在线软件 ProtParam (<http://web.expasy.org/protparam/>) 预测 StkP 蛋白的氨基酸组成、分子量、理论等电点、原子组成、不稳定系数及亲疏水性等理化性质。

3 StkP 蛋白的空间结构预测

StkP 蛋白的二级结构利用线服务工具多参数评价软件 SOMPA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl/page=npsa_sopma.html) 预测 StkP 蛋白的 α 螺旋、β 转角、无规卷曲等二级结构。使用 SWISS-MODEL (<https://www.swissmodel.expasy.org/interactive>) 在线服务器预测蛋白的氨基酸序列并组合模板, 构建 StkP 蛋白的三级结构, 采用 Rasmol 软件进行分析, 明确蛋白的稳定性。

4 StkP 蛋白的 B、T 细胞表位预测

使用在线预测软件 ABCpred (<http://crdd.osdd.net/raghava/abcpred/index.html>) 及 IEDB (<http://www.iedb.org/>) 综合预测 StkP 蛋白的 B 细胞表位。应用在线预测软件 SYFPEITHI (<http://www.syfpeithi.de/>) 分析 StkP 蛋白的 T 细胞表位。主要组织相容性复合体 MHC I 类分子选择 HLA-A * 0201, MHC II 类分子选择 HLA-DRB1 * 0401, 通过比较分析最终确定 T 细胞表位。

结果

1 肺炎链球菌 StkP 蛋白的氨基酸序列

StkP 蛋白 GenBank 登录号为 SCA94634.1, 全长为 659 个氨基酸。将 StkP 蛋白的登录号 (SCA94634.1) 输入 NCBI 获得蛋白的具体氨基酸组成 (图 1)。

```
putative serine/threonine protein kinase StkP [Streptococcus pneumoniae]
GenBank: AAM47530.1
GenPept Identical Proteins Graphics
>AAM47530.1 putative serine/threonine protein kinase StkP [Streptococcus pneumoniae]
MIQIKIFAGRYRIVKQIGRGMADVILAKDLIDGEEVAVKVLRTNYQTDPIAVARFQREARAMADLDH
PHIVRIIDIGEEEDGQQVLAEMIVAGLDLKKRYIKEHYPLSNEEAVRIMRQILLAMLAHTRGIVHRDLKPKQ
NILLTPDGTAKVDFGIAVAFATSLTQTNSMLGSHVHLSPEQARGSKATVQSDIYAMGIIIFYEMLTGHI
PYDGSVAVTIALQHFQPLPSVIAENSSVPAQLENVLIKATAKLLINRYSVSEMVDLSSLSYNNRNE
SKLIFDTSKADIKTLPKVQSQTLSIPKVKQAQTEHKSINPQAVTEETYQPAPKPKRFRMRYLILLA
SLVLAASLIVLRSRPAIIAIPDVAQGVAAEAKATLKKANFIEGEEKTEASEKVEEGRIIRTPDAGAGT
RKEGKINLVSSGKQSFQISNYGRKSSDVI AELKKEKVPDNLKIEEESSESEAGTVLQKSLPEGIT
YDLKATQIVLTVAKKATTIQLGNVIGRNSTEVISELKKQKVPENLKIIEEESSESEPGTIMKQSPGAG
TIDVSKPTQIVLTVAKKTSVAMPSTYIGSSLEFKNLQIVTIGKEANIEVVEVTTAPAGSEVGMVQVE
SPRAGEKVDLNRTRKIVKISLYKPKTTSATP
```

图 1 StkP 蛋白的氨基酸序列
Fig. 1 Amino acid sequence of StkP

2 StkP 蛋白的理化性质

通过蛋白质分析软件 ExPASy 的 ProtParam 分析, StkP 蛋白由 659 个氨基酸组成, 相对分子质量为 72.28881×10^3 , 理论 pI 值 8.61。其中含有 79 个强酸性 (-) 氨基酸 (Asp + Glu), 83 个强碱性 (+) 氨基酸 (Arg + Lys)。该蛋白由 10314 个原子组成, 原子组成为 $C_{3197}H_{5234}N_{872}O_{997}S_{14}$ 。当该蛋白 N-末端是 M (Met) 时, 于体外哺乳动物网织红细胞中的半衰期为 30 h。StkP 蛋白的不稳定指数为 40.13, 归类为不稳定蛋白; 脂肪指数 93.23, 平均亲水系数 (GRAVY) 为 -0.294, 属于亲水性蛋白。

3 StkP 蛋白的空间结构

3.1 二级结构 利用 SOMPA 在线软件对 StkP 蛋白的二级结构进行预测, 结果如图 2。其中 α 螺旋占 33.08%, 延伸链占 19.58%, β 转角占 7.74%, 无规卷曲占 39.61%。

3.2 亲水性、柔韧性、表面可及性、β-折叠 分别使用 ExPASy 网站 ProScale 软件中的 Hydropath. / Kyte & Doolittle、Average flexibility、% accessible residues、beta-turn / Chou & Fasman 预测工具对 StkP 蛋白的亲水性、柔韧性、表面可及性和 β-折叠进行预测, 结果如图 3。图中横坐标代表蛋白氨基酸的定位, 纵坐标代表预测的分数。以柔韧性分析为例 (图 3B), 柔韧性指数最小为 3.589, 最大为 7.967。

3.3 三级结构 采用 SWISS-Model 在线软件, 通过与数据库中已有的蛋白进行序列比对, 选择同源性最高的模板 4eqm.1.A 进行建模, 预测 StkP 蛋白的三级结构, 结果如图 4。StkP 蛋白的预测结构与模板覆盖率为 46.91%。该模型 GMQE 评分为 0.31, QMEAN 评分为 -2.12, 模型结果可靠。

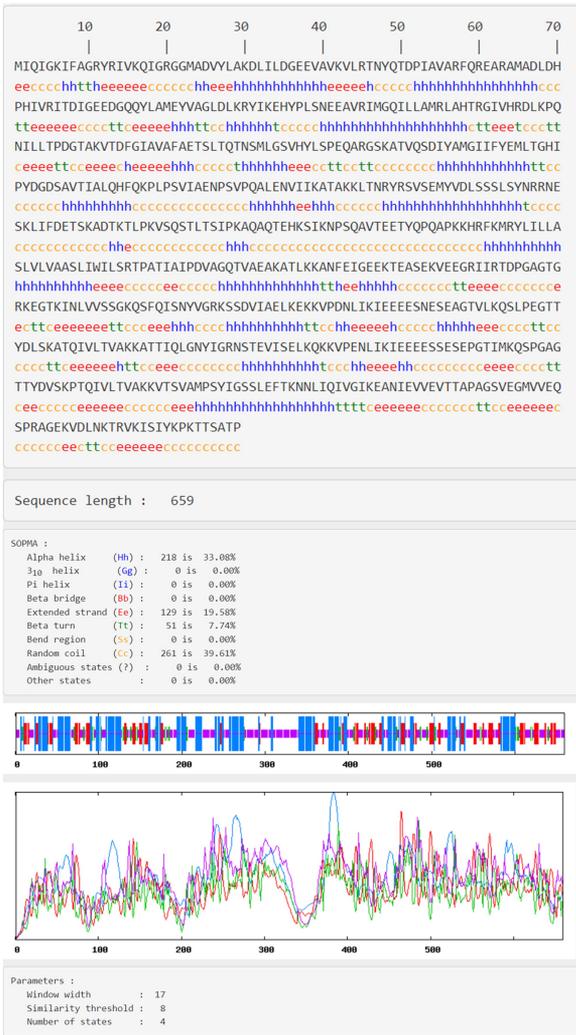


图 2 StkP 蛋白的二级结构预测

Fig. 2 Secondary structure prediction of StkP

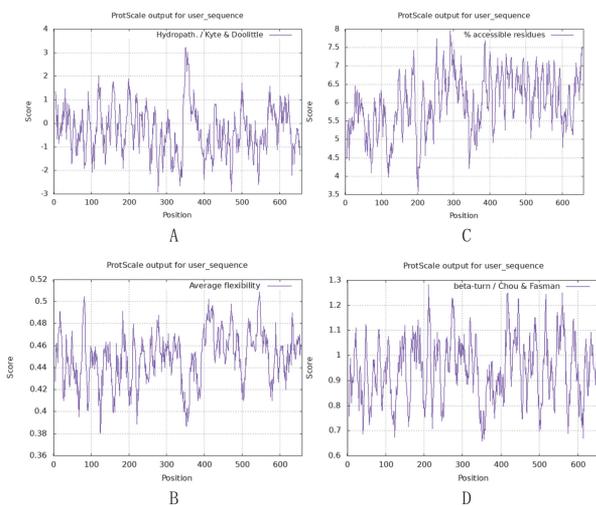


图 3 StkP 蛋白的亲水性(A)、柔韧性(B)、表面可及性(C)及 β -折叠(D)预测

Fig. 3 Prediction of hydrophilic flexible surface accessibility β -folding for StkP

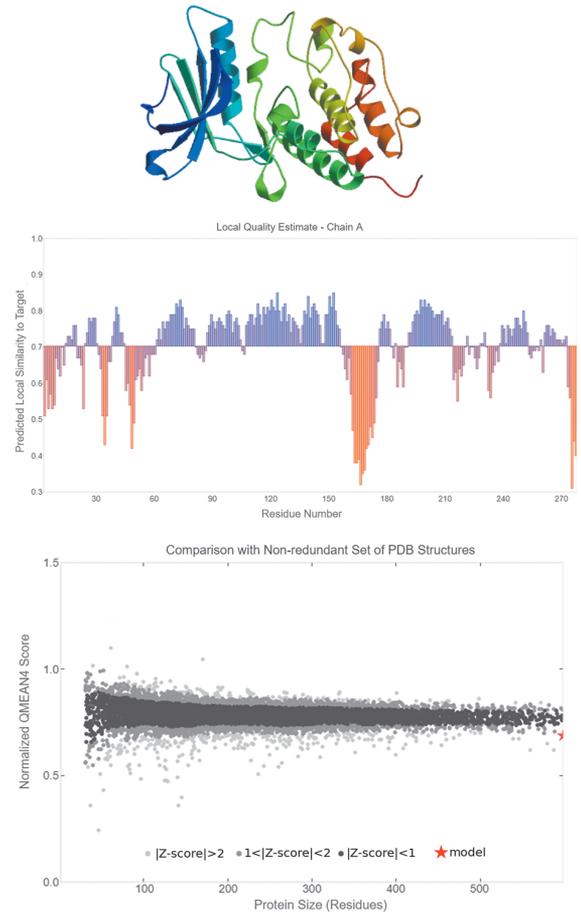


图 4 StkP 蛋白的三级结构预测

Fig. 4 Tertiary structure prediction of StkP

4 StkP 蛋白的 B、T 细胞抗原表位

4.1 B 细胞抗原表位 应用 ABCpred 软件对 StkP 蛋白的 B 细胞抗原表位进行预测,设置阈值为 0.51,共得到 33 条 B 细胞表位,选取其中评分大于 0.85 的 14 条表位进行后续分析。利用在线软件 IEDB 的 Bepipred Linear Epitope Prediction 2.0 对 StkP 蛋白的 B 细胞抗原表位进行分析,共预测出 25 条 B 细胞表位,去除长度过长及过短的表位后,选用 14 条。将上述两种方法选出的表位进一步通过在线软件 VaxiJen2.0 评估抗原性,最终确定的表位数量分别为 9 个和 7 个(表 1)。

4.2 T 细胞抗原表位 通过 SYFPEITHI 在线程序对 StkP 蛋白的 T 细胞抗原表位进行预测分析。其中,选择等位基因 HLA-A * 0201 预测 CTL 细胞抗原表位,选用长度为 9 个氨基酸进行分析(阈值设定为 23);以 HLA-DRB1 * 0401 预测 Th 细胞抗原表位,选用长度为 15 个氨基酸分析(设定阈值为 26)。将预测的 T 细胞表位应用 VaxiJen2.0 软件评估抗原性,选择预测表位中分数较高的表位,最终确定了 8 个 CTL 和 13 个 Th 细胞优势表位(表 2)。

表 1 StkP 蛋白 B 细胞表位预测

方法	起始位置	序列	
ABCpred	18	IGRGGMADVYLAKDLI	
	83	DGQQYLAMEYVAGLDL	
	139	PQNILLTPDGTAKVTD	
	196	YAMGIIFYEMLTGHIP	
	288	TSKADTKTLPKVSQST	
	307	IPKAQAQTEHKSINKP	
	408	GRIIRTPDGAGTGRKE	
	414	DPGAGTGRKEGTKINL	
	554	KQSPGAGTTYDVSKPT	
	IEDB	160	AFAETSLTQTNSM
		361	WILSRTPATIAIPDVAGQTVAE
		434	GKQSFQISNYVGRKSSD
		458	KKVPDNLKIEEEESNESEAG
577		KKVTSVAMPSYIGSSLEF	
600		IQIVGIKEANIEVVEVTTAPAGSV	
	630	QSPRAGEKVDLN	

表 2 StkP 蛋白 T 细胞表位预测

表位类型	起始点	序列	分数
CTL 表位	347	ILLASLVLV	30
	33	ILDGEEVAV	27
	88	LAMEYVAGL	27
	144	LTPDGTAKV	25
	345	YLILLASLV	25
	121	LLAMRLAHT	24
	346	LILLASLVL	23
	574	TVAKKVTSV	23
Th 表位	38	EVAVKVLRRTNYQTDP	26
	121	LLAMRLAHTRGIVHR	26
	141	NILLTPDGTAKVTDF	26
	156	GIAVAFAETSLTQTN	26
	215	DSAVTIALQHFQKPL	26
	262	VSEMYVDLSSLSYN	26
	343	MRYLILLASLVLVAA	26
	391	NFEIGEKEKTEASEKV	26
	424	GTKINLVVSSGKQSF	26
	464	LIKIEEEESNESEAG	26
	536	LIKIEEESESEPG	26
	609	NIEVVEVTTAPAGSV	26
	623	VEGMVVEQSPRAGEK	26

讨 论

疫苗是传染性疾病预防的重要手段之一,其具有安全、有效、无残留等优点。随着基因工程、单克隆抗体、短肽合成等技术的发展,分子疫苗的研制出现了新的技术途径和解决方法。抗原分子上存在的表位具有决定抗原的特异性、诱导特异的体液及细胞免疫应答的作用,抗原能激活机体的免疫反应,主要是其中特异性表位发挥作用。如果能够筛选出优势抗原表位,诱导出更加特异的抗体,则能够安全有效地预防疾病。表位肽是抗原的最小免疫活性区域,与 B 细胞和 T 细胞结合激活宿主免疫,并精确控制免疫反应的方向。

从理论上讲,它是疫苗开发中最基本和最有效的成分^[16]。表位肽疫苗完全由人工合成,不含核酸,能够克服常规疫苗的一些缺陷,被认为是一种更加高效、简便、稳定、经济的疫苗,目前肽疫苗是预防和控制感染性疾病和恶性肿瘤的主要新型疫苗研制方向之一^[17]。与全蛋白相比,表位疫苗能使机体产生高水平的抗体,细胞免疫也起着重要的辅助作用^[18-19]。

在信息时代,人类基因组计划不断深入,生物信息学发挥出其重要的作用,所涉及的内容包括生命科学等多个领域,为生命科学的研究与发展提供了技术支持和研究手段。随着计算机技术的发展和生物数据库的深入挖掘,从现有数据库中根据抗原表位的序列及结构特征预测可能的表位进行验证,为优势表位的筛选提供了一个更便捷、快速和经济实用的方法^[16]。

本研究利用生物信息学方法预测了 StkP 蛋白的二级结构,其中 α 螺旋占 33.08%,延伸链占 19.58%, β 转角占 7.74%,无规卷曲占 39.61%。通过二级结构预测及三级结构建模,发现蛋白存在多个无规则卷曲结构,其具有抗原潜力。利用在线软件 ABCpred 和 IEDB 等数据库分别对 B 细胞线性表位进行预测分析,排除位于 α 螺旋和 β 折叠区域的表位后,两种方法均预测到潜在的 B 细胞表位,起始点分别位于 18、83、139、160、196、288、361、577、600 等位置。T 细胞包括 CD8⁺ T 细胞和 CD4⁺ T 细胞,其中 CD8⁺ T 细胞也称细胞毒性 T 淋巴细胞 (cytotoxicity T lymphocyte, CTL),CD4⁺ T 细胞又称辅助性 T 细胞 (T helper, Th)。T 细胞抗原表位分别由 CTL 和 Th 细胞识别的 MHC I、II 类分子呈递。其中与 MHC I 类分子结合的内源性抗原肽由 CTL 识别,与 MHC II 类分子结合的外源性抗原肽则由 Th 细胞识别。本研究利用 SYFPEITHI 软件预测分析了该蛋白的 T 细胞抗原表位,HLA 等位基因选择 HLA-A * 0201 和 HLA-DRB1 * 0401 进行预测,最终筛选出 CTL 和 Th 细胞优势抗原表位,起始点分别为:347、33、88、144、331、536、609、623 等。

本研究预测了肺炎链球菌 StkP 蛋白作为候选分子可能包含的 B 细胞线性表位、CD4⁺ T 细胞及 CD8⁺ T 细胞表位,完成了表位疫苗研制的第一步,在此基础上可通过体内及体外实验最终获得保护性抗原表位,为全面评价以 StkP 蛋白为基础的新型亚单位疫苗的合理设计与开发提供了一种新的策略,为开展表位肽搭配串联后的免疫保护性研究奠定了理论基础。

【参考文献】

[1] 易中恒. 系统回顾全球 19 型肺炎链球菌的循环克隆复杂性和序列类型:多重耐药的重要性[J]. 微生物学免疫学进展,2021,49(5):6.

