

衣原体噬菌体 Vp2 蛋白的生物临床价值分析

姚卫锋¹ 卢桂玲¹ 谢艳秋¹ 于旺¹ 宋蒙蒙¹ 李士颖^{2△}

【摘要】 目的 明确衣原体噬菌体 Vp2 蛋白在病毒重组、筛查研究中的作用,评估 Vp2 的临床价值。**方法** 以 COBALT 程序在线比对各株衣原体噬菌体衣壳蛋白 Vp2 蛋白序列;并以 ProteinBlast 的 Distance tree 功能构建种系发生树。以高保守区氨基酸序列为基础,采用 Gamier-Robson 法、Chou-Fasman 法分析蛋白二级结构;以 Karplus-Schulz 法分析柔性区域;Kyte-Doolittle 法、Hopp-Woods 法分析亲水性;Emini 法分析表面可及性;Jameson-Wolf 法分析抗原指数。**结果** 6 株衣原体噬菌体 Vp2 蛋白序列高度保守,差异主要在 Chp1 与其他 5 株噬菌体的 Vp2 蛋白之间亲缘关系较远。各株噬菌体的 Vp2 蛋白均有 α 螺旋为主的二级结构,蛋白序列高保守区存在多个细胞表位。**结论** Vp2 蛋白性质保守,是衣原体噬菌体衣壳的重要组分。其蛋白分子结构复杂,高保守区有较强的免疫原性,在病毒重组、沙眼衣原体噬菌体野生株筛查研究中有实际研究价值。

【关键词】 微病毒属;衣原体科;细菌噬菌体;蛋白质类;计算生物学

【中图分类号】 R373.9 **【文献标志码】** A **【DOI】** 10.3969/j.issn.0253-9896.2014.07.003

The Clinical Value of Chlamydia Phage Vp2 Protein

YAO Weifeng¹, LU Guiling¹, XIE Yanqiu¹, YU Wang¹, SONG Mengmeng¹, LI Shiyong²

1 Tianjin Academy of Traditional Chinese Medicine Affiliated Hospital, Tianjin 300120, China; 2 Tianjin Gonggan Hospital

LI Shiyong, E-mail:lishiyongmed@sohu.com

【Abstract】 Objective To evaluate the effect of chlamydia phage virus protein 2(Vp2) on the recombinant virus and virus screening research, and its clinical value thereof. **Methods** To compare the Vp2 protein sequences to get the conservative region with COBALT. A phylogenetic tree was built with ProteinBlast of Distance tree. The amino acid sequence in the high conservative region was predicted by the methods of Gamier-Robson and Chou-Fasman, and its flexible regions were predicted by Karplus method. The hydrophilicity plot was predicted by Kyte-Doolittle and Hopp-Woods method. The surface probability was analysed by Emini, and the antigenic index was analysed by Jameson-Wolf method. **Results** The six Chlamydia phage Vp2 proteins were the highly conserved sequences. There were obvious differences between Chp1Vp2 and other 5 Vp2 proteins. There were the main structure- α helix and some cell epitopes in the high conserved region. **Conclusion** Vp2 protein is the important component of chlamydia phage capsid with the conservative nature. Vp2 protein has complicated structures and high conservative region with strong immunogenicity, playing a practical value of research in virus recombinant and screening the wild strains of chlamydia trachomatis phage.

【Key words】 microvirus; chlamydiaceae; bacteriophages; proteins; computational biology

衣原体为胞内寄生菌,可经眼、呼吸道、泌尿生殖道黏膜感染,易感难治,多年来感染病例仍不断增加。随着细菌噬菌体研究的复兴^[1-2],近年发现的衣原体噬菌体为衣原体研究开辟了新领域^[3]。衣原体噬菌体是一类以衣原体为宿主的病毒^[4]。病毒蛋白 2(virus protein2, Vp2)是构成衣原体噬菌体衣壳的三种结构蛋白之一,核衣壳成熟过程中 Vp2 参与病毒装配^[5]。本文以衣原体的各株噬菌体 Vp2 蛋白序

列为目标,分析 Vp2 的共同特点,利用获得的生物信息评估 Vp2 的潜在研究价值。

1 材料与方法

1.1 材料 各株衣原体噬菌体 Vp2 蛋白氨基酸序列 GenBank 序列号分别为 Chlamydia phage1 Vp2 (Chp1Vp2, NP_044314.1)、Chlamydia phage2 Vp2 (Chp2Vp2, NP_054650.1)、Chlamydia phage3 Vp2 (Chp3Vp2, YP_022482.1)、Chlamydia phage4 Vp2 (Chp4Vp2, YP_338241.1)、Chlamydia pneumoniae phage CPAR39 Vp2 (phiCPAR39, NP_063896.1)、Guinea pig Chlamydia phage Vp2 (phiCPG1, NP_510874.1)。

1.2 方法 以 COBALT 程序在线比对 (<http://www.ncbi.nlm>).

基金项目:天津市中医药管理局科技基金资助项目(13017)

作者单位:1 天津市中医药研究院附属医院(邮编 300120);

2 天津市公安医院

[△]通讯作者 E-mail:lishiyongmed@sohu.com

nih.gov/tools/cobalt/cobalt.cgi) 各株衣原体噬菌体衣壳蛋白 Vp2 序列; 并以 ProteinBlast 的 Distance tree 功能构建种系发生树。以高保守区氨基酸序列为基础, 采用 Gamier-Robson 法^[6]、Chou-Fasman 法^[7]分析蛋白二级结构; 以 Karplus-Schulz 法分析柔性区域; Kyte-Doolittle 法^[8]、Hopp-Woods 法^[9]分析亲水性; Emini 法分析表面可及性; Jameson-Wolf 法^[9]分析抗原指数。

2 结果

2.1 序列高保守区比对结果 6 株衣原体噬菌体 Vp2 蛋白序列除衣原体噬菌体 Chp1 的 Vp2 蛋白由 263 个氨基酸残基(a.a)组成外, 其他 5 株的 Vp2 蛋白都包含 186 个 a.a。比对发现: Chp1Vp2 在 N 端 a.a 第 1~37(37 a.a), 193~213(20 a.a), 245~263(18 a.a) 节段与其他 5 种 Vp2 蛋白存在长距离“空沟”, 见图 1 (以短横线“-”表示), 其长度为上述 3 节段全长; 同时第 125、156、163、239 位出现单个 a.a 长度的比对空沟。分析各株 Vp2 的种系发生关系显示: Chp1Vp2 与其他 5 株亲缘关系较远, 为单独一组; 其余 5 株关系较近, 为一组, 同时又分为两个亚组: phiCPG1 和另 4 株噬菌体, 见图 2。以种系发生关系为指导, 去除 Chp1Vp2 序列后行序列比对分析, 发现 5 株噬菌体的 Vp2 蛋白无比对空沟, 序列构成及残基排列高度一致: 以 COBALT 程序网上比对发现一致度最低为 98%, 见图 1。

2.2 保守区二级结构分析结果 鉴于不同蛋白质的氨基酸序列有 30% 以上的相似性即可认为其结构相似^[10], 本文以衣原体噬菌体 Chp4 的 Vp2 蛋白为代表, 分析 5 株高度相似的衣原体噬菌体(即除去 Chp1 的其他 5 株衣原体噬菌体)的 Vp2 蛋白的二级结构。Chou-Fasman 方案和 Gamier-Robson 方案均示: 序列全长范围频繁出现长片段 α 螺旋(Alpha Regions), 而 β 折叠(Beta Regions)结构只有 3 个, 散在分布。转角结构(Turn Regions)分散排布于上两者之间, 数量不多。两种方案显示的 α 螺旋位置大部分重合; 但 β 折叠位置有冲突, 一致度高的位置为氨基酸残基 N 端第 25~37 区间。两方案分析转角结构与柔性区域(Flexible Regions, Karplus-Schulz 方法)的位置基本重叠, 柔性区域为 N 端第 5~13、36~45、56~63、69~74、82~91、95~102、105~108、115~120、125~132、135~138、140~143、145~150、162~175, 占全部氨基酸残基的 43.4%。说明有较多位置存在潜在的细胞表位。Gamier-Robson 方案示: 无规卷曲结构(Coil Regions)数量少而分散, 位置上同样与柔性区域基本重叠, 见图 3。综合上述: α 螺旋位置为 N 端第 3~7、34~57、62~73、77~81、101~104、115~135、144~154、158~165、175~181, 为全部氨基酸残基的 47.3%; β 折叠位置为 N 端第 26~36, 占全部氨基酸残基的 5.3%。说明 Vp2 蛋白二级

✓ YP_022482	1	MNPEQLTNTLGSVSGVAQGLSFLPGIASGVLGYLGAQKQKQNTAKQIAREQMAFQERMSNTAYQRAMEDMKKAGLN	76
✓ NP_510874	1	MNPEQLTNTLGSVSGVAQGLSFLPGIASGVLGYLGAQKQKQNTAKQIAREQMAFQERMSNTAYQRAMEDMKKAGLN	76
✓ NP_054650	1	MNPEQLTNTLGSVSGVAQGLSFLPGIASGVLGYLGAQKQKQNTAKQIAREQMAFQERMSNTAYQRAMEDMKKAGLN	76
✓ YP_338241	1	MNPEQLTNTLGSVSGVAQGLSFLPGIASGVLGYLGAQKQKQNTAKQIAREQMAFQERMSNTAYQRAMEDMKKAGLN	76
✓ NP_063896	1	MNPEQLTNTLGSVSGVAQGLSFLPGIASGVLGYLGAQKQKQNTAKQIAREQMAFQERMSNTAYQRAMEDMKKAGLN	76
✓ NP_044314	1	[37]AKQANLLNNLPLPSLDNVLNIGMFGLASGLLSYRAAKKQNKVMQDIANRQMAFQERMSSTAVRRHVEDLKKAGLN	113
✓ YP_022482	77	PMLAFSKGGASSPAGASWSPNNPVESAMNSGLAVQLTYERKKM-QAELQN-LREQNRLIRNQAIREGYLAERDKYMRVA	154
✓ NP_510874	77	PMLAFSKGGASSPAGASWSPNNPVESAMNSGLAVQLTYERKKM-QAELQN-LREQNRLIRNQAIREGYLAERDKYMRVA	154
✓ NP_054650	77	PMLAFSKGGASSPAGASWSPNNPVESAMNSGLAVQLTYERKKM-QAELQN-LREQNRLIRNQAIREGYLAERDKYMRVA	154
✓ YP_338241	77	PMLAFSKGGASSPAGASWSPNNPVENAMNSGLAVQLTYERKKM-QAELQN-LREQNRLIRNQAIREGYLAERDKYMRVA	154
✓ NP_063896	77	PMLAFSKGGASSPAGASWSPNNPVESAMNSGLAVQLTYERKKM-QAELQN-LREQNRLIRNQAIREGYLAERDKYMRVA	154
✓ NP_044314	114	PLLALG-GSASTPQGAFFYPVNPVPMESGLNSAISVADKVFQYQRLAHADFGGLNSAMSVVQLASAVQDYKRNKYGKFGVA	192
✓ YP_022482	155	GVPVATEMLDRTSGL-ISSSARAFKQLFSRKGR	186
✓ NP_510874	155	GVPVATEMLDRTSGL-LSSSAKAFKQLFSRKGR	186
✓ NP_054650	155	GVPVATEMLDRTSGL-LSSSAKAFKQLFSRKGR	186
✓ YP_338241	155	GVPVATEMLDRTSGL-ISSSAKAFKQLFSRKGR	186
✓ NP_063896	155	GVPVATEMLDRTSGL-LSSSAKAFKQLFSRKGR	186
✓ NP_044314	193	[20]KHPVGRAVSAANSGLAVKAGKGVNFKFSNMSS[18]	263

横线“-”为比对空沟

Fig. 1 The sequences alignment of six Vp2 proteins

图 1 6 株衣原体噬菌体的 Vp2 蛋白序列比对

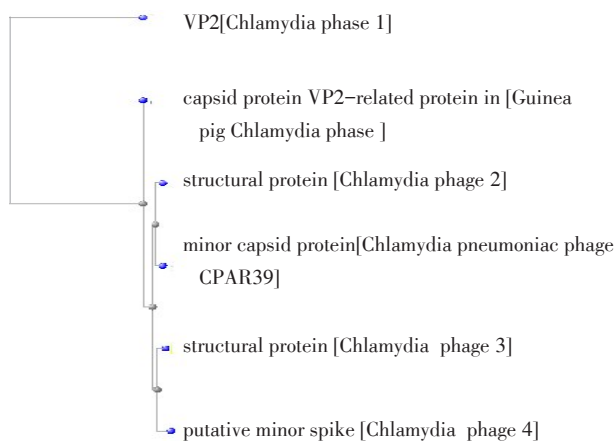


Fig. 2 The phylogenetic tree of six Vp2 protein sequences

图2 6种Vp2种系发生树

结构以 α 螺旋为主。转角结构位置为N端第58~62、84~86、89~94、137~142。

2.3 B细胞表位预测结果 Kyte-Doolittle 和 Hopp-Woods 方案分析蛋白亲水性显示亲水区域基本重叠,蛋白亲水性较高的区域(Hydrophilicity Plot)为N端第3~8、35~43、46~50、57~62、84~90、93~100、112~121、125~135、140~153、178~186,见图3。Jameson-Wolf法预测抗原指数(Antigenic Index)较高的区域为N端第1~8、3~74、83~92、95~105、110~52、157~168、170~177、180~186,见图3。高抗原指数区域与预测的亲水区重合度良好。Emini 方案预测氨基酸残基蛋白质表面可及性(Surface Probability Plot),显示N端第5~13、35~46、56~62、68~73、82~91、95~101、105~108、115~119、123~132、134~137、139~142、146~

150、160~174、180~183区段表面可及性良好,见图3。

3 讨论

3.1 衣原体噬菌体 Vp2 蛋白的生物多态性 Vp2 蛋白是衣原体噬菌体衣壳蛋白的三种成分之一,与其他两种蛋白以固定比例构成噬菌体的正20面壳,是噬菌体包装成熟过程中不可或缺蛋白成分。鉴于没有Vp2相关蛋白的晶体结构衍射分析,其三维结构不明。因而本文通过分析Vp2蛋白的一级结构,从理论上反复论证以获得Vp2蛋白更多的结构及相应的功能信息。对目前发现的全部6株噬菌体Vp2蛋白序列的分析,揭示了Vp2蛋白的氨基酸残基构成及排列均有明显的一致性。虽然出现了3段长距离空沟及4个单残基长度空沟,比对差异颇为明显,而这种差异是因为Chp1Vp2序列与其他5株Vp2序列的差异引起的。理论上,蛋白的生物多态性应该更为复杂,虽然都是衣原体噬菌体的Vp2蛋白,但不同衣原体噬菌体的蛋白必然有其自身特点。而6株Vp2蛋白的种系发生关系进一步清晰了衣原体噬菌体Chp1株的Vp2蛋白与其他5种Vp2蛋白差异的原因——Chp1Vp2与其他5株亲缘关系较远。实际上,上世纪80年代发现的第一株衣原体噬菌体Chp1并没引起关注^[10]。比对其余5种Vp2序列发现有98%~100%的一致性。比对结果表明:Vp2蛋白在进化过程中高度保守,尤其相对Chp1Vp2的263个氨基酸残基,其余五种来自不同衣原体噬菌体的Vp2蛋白的186个氨基酸残基从组成种类、数

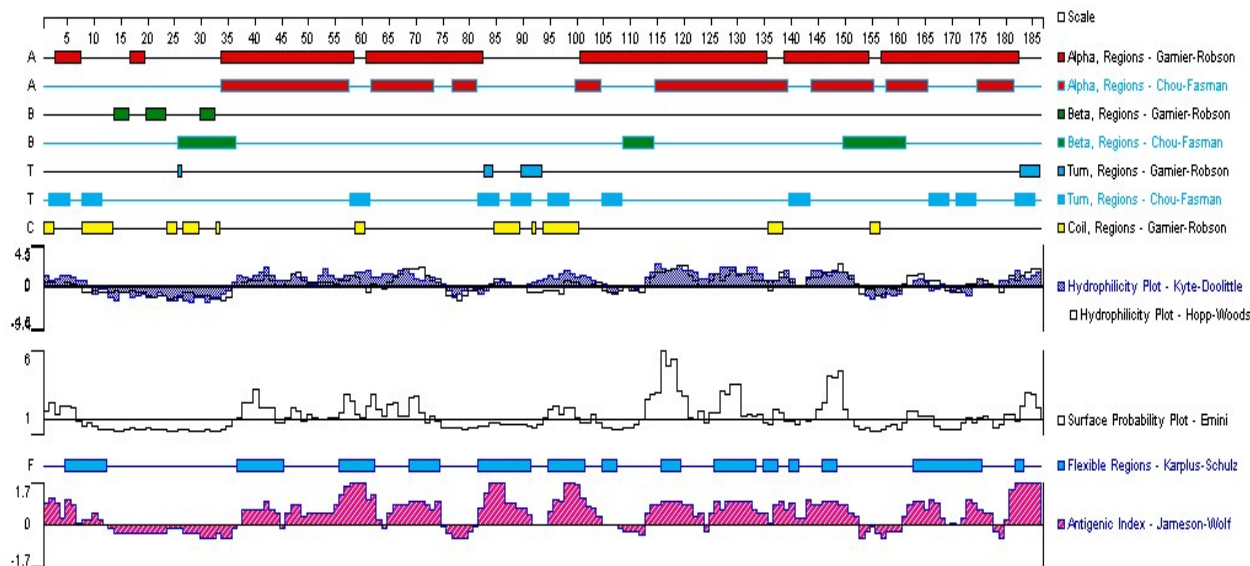


Fig. 3 The prediction map of secondary structure and B cell epitope

图3 二级结构及B细胞抗原表位预测图

量及排列次序上保守性很高。进一步,即使相对亲缘关系较远的 Chp1Vp2 也在序列中包含了这 186 a.a (即使相应位置的氨基酸残基不同,但性质相似)。这在频繁发生突变的简单微生物种类——病毒中足以引起重视^[1],Vp2 蛋白保守性极高,并没表现出明显的生物多态性。因而可以认为 Vp2 在噬菌体组装成熟过程中有不可替代的功能,以至于其可能的突变(理论上应该有较高突变率)将影响噬菌体的生活过程——核衣壳的成熟,这样可能的突变株不易存活。因此,Vp2 蛋白就具备了通用标志物的特点,以其为筛查目标,将极大提高衣原体噬菌体筛查的灵敏度,这对于在衣原体中,尤其是在沙眼衣原体中寻找噬菌体提供了合理的理论支持。

3.2 衣原体噬菌体 Vp2 蛋白的二级结构及细胞表位 通过序列比对,发现的长度为 186 a.a 蛋白序列仍有必要进一步挖掘其生物信息,并从这些信息中获得研究指导和提示。在晶体结构不明的情况下,应用生物信息工具分析其二级结构及细胞表位的研究方法已经为研究者所接受^[7-8]:随着分析工具的不断完善,多方案分析汇总后的结果与实证结构的符合率不断提高。总原则为:两种或两种以上不同原理的预测方案共同认可的预测结果将被采信。本文通过不同预测方案的交叉应用,分析了代表高保守 186 a.a 蛋白序列的 Chp4Vp2 蛋白序列。研究中普遍认为 α 螺旋、 β 折叠这两种结构很少能形成细胞表位^[7-9],因而在最终结果分析中剔除上述区间的预测表位。反之,对位于无规卷曲和转角结构这两种极可能形成细胞表位的结构区的预测表位,则在预测中保留,并进入下一步评估。进一步通过对亲水性、表面可及性、抗原指数指标的逐次筛选——满足高亲水性、高表面可及性和高抗原指数要求后,本研究认为氨基酸残基序列 N 端第 1~6、57~61、83~89、93~99、112~117、181~186 区间为 Vp2 蛋白的优势 B 细胞表位。以这些预测表位为基础,将大大简化后续核酸、蛋白合成步骤,节省研究时间及经济成本。

相对种系发生晚的流产衣原体、豚鼠结膜炎衣原体等较新的衣原体都已发现了自身的噬菌体,从进化角度推测,种系发生较早的沙眼衣原体也是临床关注度最高的衣原体,很可能存在自己的噬菌体。目前发现的总计 6 株衣原体噬菌体各以一种或几种衣原体为宿主,并且噬菌体宿主谱间有交叉重叠^[10]。这表明可以从沙眼衣原体之外的其他衣原体中寻找能感染沙眼衣原体的噬菌体。以上两方

面都需要以目前发现的衣原体噬菌体为基础,通过其某种共性来寻找新的衣原体噬菌体。Vp2 蛋白的高保守性、多优势表位的可选择性顺应了这个要求,因而是很好的标志蛋白。

本文分析了全部 6 株衣原体噬菌体的 Vp2 蛋白氨基酸序列,证明各株噬菌体的 Vp2 蛋白一致度很高,性质保守。序列的高保守区中有 6 个优势 B 细胞表位并获得各个表位的具体位置及氨基酸构成信息。明确了该蛋白在病毒重组、沙眼衣原体噬菌体野生株筛查研究中有重要价值。6 株衣原体噬菌体中的 5 株是近 10 余年发现的,该领域研究亟待拓展,并由此推进衣原体疾病的基础研究。

参考文献

- [1] Kutter E, De Vos D, Gvasalia G, et al. Phage therapy in clinical practice: treatment of human infections[J]. *Curr Pharm Biotechnol*, 2010,11(1):69-86.
- [2] Yoo SY, Kobayashi M, Lee PP, et al. Early osteogenic differentiation of mouse preosteoblasts induced by collagen-derived DGEA-peptide on nanofibrous phage tissue matrices[J]. *Biomacromolecules*, 2011,12(4):987-996. doi: 10.1021/bm1013475.
- [3] Yoo SY, Oh JW, Lee SW. Phage-chips for novel optically readable tissue engineering assays[J]. *Langmuir*, 2012,28(4):2166-2172. doi: 10.1021/la203840n.
- [4] Abedon ST, Kuhn SJ, Blasdel BG, et al. Phage Treatment of human infections [J]. *Bacteriophage*, 2011,1(2):66-85.
- [5] Sait M, Livingstone M, Graham R, et al. Identification, sequencing and molecular analysis of Chp4, a novel chlamydiaphage of *Chlamydia abortus* belonging to the family Microviridae[J]. *J Gen Virol*, 2011,92(Pt 7):1733-1737. doi: 10.1099/vir.0.031583-0.
- [6] Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes[J]. *Immunome Res*, 2006,2(2):1-7.
- [7] Ponomarenko JV, Bourne PE. Antibody-protein interactions: benchmark datasets and prediction tools evaluation [J]. *BMC Struct Biol*, 2007,7:64.
- [8] Saraswathi S, Fernández-Martínez JL, Koliński A, et al. Distributions of amino acids suggest that certain residue types more effectively determine protein secondary structure[J]. *J Mol Model*, 2013, 19(10):4337-4348. doi: 10.1007/s00894-013-1911-z.
- [9] Su CH, Pal NR, Lin KL, et al. Identification of amino acid propensities that are strong determinants of linear B-cell epitope using neural networks[J]. *PLoS One*, 2012,7(2):e30617. doi: 10.1371/journal.pone.0030617.
- [10] Sliwa-Dominiak J, Suszyńska E, Pawlikowska M, et al. Chlamydia bacteriophages[J]. *Arch Microbiol*, 2013,195(10-11):765-771. doi: 10.1007/s00203-013-0912-8.
- [11] Krupović M, Bamford DH. Order to the viral universe[J]. *J Virol*, 2010,84(24):12476-12479. doi: 10.1128/JVI.01489-10.

(2013-11-25 收稿 2013-12-30 修回)

(本文编辑 李鹏)