

文章编号 1006-8147(2014)03-0188-04

论 著

衣原体噬菌体衣壳蛋白 Vp1 及其高保守区二级结构及 B 细胞表位研究

姚卫锋¹, 卢桂玲¹, 谢艳秋¹, 于 旺¹, 宋蒙蒙¹, 李士颖²

(1. 天津市中医药研究院附属医院皮肤科, 天津 300120; 2. 天津公安医院内三科, 天津 300042)

摘要 目的: 分析衣原体噬菌体 Vp1 蛋白及其高保守区的二级结构并预测其 B 细胞表位。方法: 以 ClustalX 程序比对各株衣原体噬菌体衣壳蛋白 Vp1 序列获得高保守区序列。以 Vp1 氨基酸序列为基础, 采用 Gamier-Robson 法、Chou-Fasman 法、Eisenberg 法和 Karplus-Schulz 法分析蛋白二级结构; 按 Kyte-Doolittle 法、Emmini 法和 Jameson-Wolf 法预测蛋白的抗原表位。结果: 衣原体噬菌体 Vp1 蛋白的二级结构以 β 折叠为主, 有少量 α 螺旋; 其高保守区含多个抗原位点, 预测其 N 端 1-8, 103-110, 158-164, 189-196, 322-332, 427-434, 478-488 为优势表位。结论: 衣原体噬菌体 Vp1 蛋白及其高保守区存在复杂的蛋白结构, 可形成多个可选表位。

关键词 微病毒; 蛋白; 结构; 分析

中图分类号 Q7

文献标志码 A

Bioinformation study of capsid protein Vp1 high conservative region parts from Chlamydia phage

YAO Wei-feng¹, LU Gui-ling¹, XIE Yan-qiu¹, YU Wang¹, SONG Meng-meng¹, LI Shi-ying²

(1. Department of Dermatology, Tianjin Academy of Traditional Chinese Medicine Affiliated Hospital, Tianjin 300120, China; 2. Internal Medicine 3, Tianjin Gong'an Hospital, Tianjin 300042, China)

Abstract Objective: To predict the secondary structure and B cell epitope of capsid protein Vp 1 in high conservative region. **Methods:** The conservative region was acquired by comparing the Vp1 protein sequences. The secondary structure of these regions were predicted by the method of Gamier-Robson, Chou-Fasman, Eisenberg and Karplus-Schulz while its cell epitope was predicted by the method of Kyte-Doolittle, Emmini and Jameson-Wolf. **Results:** The main constructions of Chlamydia phage Vp1 were β regions, including several α regions. And the sections of 1-8, 103-110, 158-164, 189-196, 322-332, 427-434, 478-488 in the N-terminal of the conservative region could be the epitopes of B cell. **Conclusion:** The conservative region of Vp1 protein has the complicated structures and can form into multiple epitopes.

Key words microvirus; protein; structure; analysis

衣原体感染与人类的多种疾病关系密切^[1], 同时本身也是衣原体噬菌体的宿主。衣原体噬菌体与衣原体的关系逐渐成为研究的热点。近 10 年发现的衣原体噬菌体已达 5 株, 其 20 面对称衣壳均含有 3 种蛋白分子, 其中 Vp1 (virus protein 1) 分子量最大, 在衣壳中含量最高, 肽链长度最长, 有形成标志性抗原表位的良好物质基础^[2]。本文旨在分析各株噬菌体 Vp1 氨基酸序列的保守区及其结构信息, 为实证研究奠定基础。

1 材料与方法

1.1 材料 衣原体噬菌体 Vp1 蛋白氨基酸序列 Genbank 序列号分别为 Chp2 (NP_054647.1); Chp3 (YP_022479.1); Chp4 (YP_338238.1); phiCPAR39 (NP_510872.1); phiCPG1 (NP_510872.1)。

1.2 方法 以 ClustalX 程序比对各株衣原体噬菌体衣壳蛋白 Vp1 序列获得高保守区序列。蛋白二级结构预测应用 DNASTar 软件提供的 Protean 模块进行, 采用 Gamier-Robson 方法、Chou-Fasman 方法、Eisenberg 方法和 Karplus-Schulz 方法分析蛋白二级结构。B 细胞表位预测以 7 个氨基酸残基为一组, 分别按 Kyte-Doolittle、Emmini 和 Jameson-Wolf 方案预测其亲水性、表面可及性和抗原性指数。

2 结果

2.1 序列高保守区比对结果 衣原体噬菌体 Vp1 高保守蛋白序列: ClustalX 序列比对结果见图 1 (以 Chp2Vp1 氨基酸序列定位): N 端第 1-215, 300-565 区域比对后标为连续 "*", 一致性很强, 为高保守区; 各序列第 216-299 位及第 461-470 位比对后无连续 "*" ":", 其间甚至出现比对 "空沟 (以 "-" 表示)", 序列差异显著。

作者简介 姚卫锋 (1977-), 男, 主治医师, 博士, 研究方向: 皮肤病学;
通信作者: 李士颖, E-mail: shiying7803@163.com。

CLUSTAL 2.0.11 multiple sequence alignment

```

cpar39      MVRNRRLPSVMSHSFAQVPSARIQRSSFDRCGLKTTFDAGYLIPFCDEVLPDGTIFSLK 60
phiCPG1VP1  MVRNRRLPSVMSHSFAQVPSARIQRSSFDRCGLKTTFDAGYLIPFCDEVLPDGTIFSLK 60
chp4        MVRNRRLPSVMSHSFAQVPSARIQRSSFDRCGLKTTFDAGYLIPFCDEVLPDGTIFSLK 60
chp2        MVRNRRLPSVMSHSFAQVPSARIQRSSFDRCGLKTTFDAGYLIPFCDEVLPDGTIFSLK 60
chp3Vp1     MVRNRRLPSVMSHSFAQVPSARIQRSSFDRCGLKTTFDAGYLIPFCDEVLPDGTIFSLK 60
*****

cpar39      EAFLARMATPIFFPLMDNLRDLTQYFFVPLRLIWSNFQKFCGEQDNPGDSTDFLTPVLTAP 120
phiCPG1VP1  EAFLARMATPIFFPLMDNLRDLTQYFFVPLRLIWSNFQKFCGEQDNPGDSTDFLTPVLTAP 120
chp4        EAFLARMATPIFFPLMDNLRDLTQYFFVPLRLIWSNFQKFCGEQDDPGDSTDFLTPILTAP 120
chp2        EAFLARMATPIFFPLMDNLRDLTQYFFVPLRLIWSNFQKFCGEQDNPGDSTDFLTPILTAP 120
chp3Vp1     EAFLARMATPIFFPLMDNLRDLTQYFFVPLRLIWSNFQKFCGEQDNPGDSTDFLTPVLTAP 120
*****

cpar39      -SGGFTEGSIHDYLGSLPTKVAGIECVAFWHRAYNLIWNQYYRDENIQESVDVEMGDITSN 179
phiCPG1VP1  -SGGFTEGSIHDYLGSLPTKVAGIECVAFWHRAYNLIWNQYYRDENIQESVDVEMGDITSN 179
chp4        QNGGFAEGSIHDYLGSLPTKVAGVQCVAFWHRAYNLIWNQYYRDENIQDSVEVQMGDITAD 180
chp2        -AGGFTEGSIHDYLGSLPTKVAGVQCVAFWHRAYNLIWNQYYRDENIQESVEVQMGDITTD 179
chp3Vp1     -TGGFTEGSIHDYLGSLPTKVAGVQCAAFWHRAYNLIWNQYYRDENIQESVEVQMGDITTD 179
*** *****

cpar39      EVNNYKLLKRGKRYDYFTSCLFPWPQKGAFTIGVGGIVFVQGLGIQWGNSSAPNPFITASS 239
phiCPG1VP1  EVNNYKLLKRGKRYDYFTSCLFPWPQKGAFTIGVGGIVFVQGLGIQWGNSSAPNPFITASS 239
chp4        EVNNYKLLKRGKRYDYFTSCLFPWPQKGAFTIGVGGIVFVQGLGIQWGGSTGNPFITASD 240
chp2        EVNNYLLKRGKRYDYFTSCLFPWPQKGAFTIGVGGKAPIEGLYMNVNNSNPFVGKFLVDS 239
chp3Vp1     EVKNYELLKRGKRYDYFTSCLFPWPQKGAFTIGVGGKAPIEGLYMNVNNSNPFVGKFLVDS 239
**:* *****

cpar39      WINSVNPFIIN----SITPTPTGINQILNYGQAYYIKKF-----GEATIDPTFPA 285
phiCPG1VP1  WINSVNPFIIN----SITPTPTGINKILNYGQAYYIKKF-----GEATIDPTFPA 285
chp4        WRDSVNFTYVN----SATQPTPTGINKILSYGQAYYIKKF-----GEATIDFAPRA 286
chp2        QSTPRVLQDLQGNKLSGIAAYNQITGKHVYVNSAWYTVIPQSEPAATLENGNYTTTQKQFI 299
chp3Vp1     QSTPRVLQDLQGNKLSGIAAYNQITGKHVYVNSAWYTVIPQSEPAATLENGNYTTTQKQFI 299
.      : : * : . * : : ..*:* . *      * : * * :

cpar39      YVDLGSTSPVTINSLREAFQLQKLYERDARGGTRYIEIIRSHFNVSQSPDARLQRAEYLG 345
phiCPG1VP1  YVDLGSTSPVTINSLREAFQLQKLYERDARGGTRYIEIIRSHFNVSQSPDARLQRAEYLG 345
chp4        YVDLGSTSPVTINSLREAFQLQKLYERDARGGTRYIEIIRSHFNVSQSPDARLQRAEYLG 346
chp2        YADLGATSPVTINSLREAFQLQKLYERDARGGTRYIEIIRSHFNVSQSPDARLQRAEYLG 359
chp3Vp1     YADLGATSPVTINSLREAFQLQKLYERDARGGTRYIEIIRSHFNVSQSPDARLQRAEYLG 359
* , * * : *****

cpar39      SSTFVNISPIPQTSSTDSTSPQGNLAAYGTAIGSKRVFTKSFTEHGVILGLASVRADLNY 405
phiCPG1VP1  SSTFVNISPIPQTSSTDSTSPQGNLAAYGTAIGSKRVFTKSFTEHGVILGLASVRADLNY 405
chp4        SSTFVNISPIPQTSSTDSTSPQGNLAAYGTAIGSKRVFTKSFTEHGVILGLASVRADLNY 406
chp2        SSTFVNISPIPQTSSTDSTSPQGNLAAYGTAIGSKRVFTKSFTEHGVILGLASVRADLNY 419
chp3Vp1     SSTFVNISPIPQTSSTDSTSPQGNLAAYGTAIGSKRVFTKSFTEHGVILGLASVRADLNY 419
*****

cpar39      QQGLDRMWSRRIRWDFYWPALSHLGEQAVLNKEIYCQGPVAVKDAQNGNIVVDEQVFGYQE 465
phiCPG1VP1  QQGLDRMWSRRIRWDFYWPALSHLGEQAVLNKEIYCQGPVAVKDAQNGNIVVDEQVFGYQE 465
chp4        QQGLDRMWSRRIRWDFYWPALSHLGEQAVLNKEIYCQGPVAVKDAQNGNIVVDEQVFGYQE 466
chp2        QQGLDRMWSRRIRWDFYWPALSHLGEQAVLNKEIYCQGPSVKNS--GGEIVDDQVFGYQE 477
chp3Vp1     QQGLDRMWSRRIRWDFYWPALSHLGEQAVLNKEIYCQGPSVKNS--GGEIVDDQVFGYQE 477
*****

cpar39      RFAEYRYKTSKITGKFRSNATGSLDAWHLAQGFENLPTLSPEFIEENPFMDRVVAVDTEP 525
phiCPG1VP1  RFAEYRYKTSKITGKFRSNATGSLDAWHLAQGFENLPTLSPEFIEENPFMDRVVAVDTEP 525
chp4        RFAEYRYKTSKITGKFRSNATSSLDWHLAQGFENLPTLSPEFIEENPFMDRVVAVDTEP 526
chp2        RFAEYRYKTSKITGKFRSNATSSLDWHLAQGFENLPTLSPEFIEENPFMDRVVAVDTEP 537
chp3Vp1     RFAEYRYKTSKITGKFRSNATSSLDWHLAQGFENLPTLSPEFIEENPFMDRVVAVDTEP 537
*****

cpar39      DFLLDGWFSRLCARPMFVYSVPGLIDHF 553
phiCPG1VP1  DFLLDGWFSRLCARPMFVYSVPGLIDHF 553
chp4        DFLLDGWFSRLCARPMFVYSVPGLIDHF 554
chp2        DFLLDGWFSRLCARPMFVYSVPGLIDHF 565
chp3Vp1     HFLLDGWFSRLCARPMFVYSVPGLIDHF 565
*****

```

* 为一致;: 为性质相似; 为低度类似; 横线为比对空沟

图1 Vp1 蛋白序列比对

Fig 1 Vp1 protein sequence alignment

2.2 保守区二级结构分析结果 蛋白二级结构预测 Chou-Fasman 方案示:小长度的 α 螺旋(alpha regions)稀疏散布、长短不均的 β 折叠(beta regions)相对密集分布于序列全长;相对长度更短的转角结构(turn regions)在以上两者间的全序列范围频繁出现,且该结构与柔性区域(flexible regions, Karplus-Schulz 方法)的位置基本重叠。Gamier-Robson 方案示:全序列只有 3 个短长度 α 螺旋,散在分布; β 折叠长度比较均匀,小间隔密布于序列全长,其间隔长度多小于邻近 β 折叠长度;转角结构和无规卷曲结构(coil regions)位置也与柔性区域基本重叠,大部分富集于序列 N 端 1-300 位氨基酸残基范围内,位于 β 折叠间隔区域,长度整体上大于 Chou-Fasman 方案预测结果。Eisenberg 方案预测 α 螺旋、 β 折叠长度不均,全序列范围疏密相间,二者位置与 Gamier-Robson 方案类似(图 2)。综合 3 种方法: α 螺旋集中区域为 N 端第 57-101,145-201,231-264,311-356,418-454,507-534; β 折叠在全序列范围小段密集间断分布,富集范围为 N 端第 33-74,135-174, 211-249, 294-353,409-438, 473-499,528-565。转角结构在全序列范围以小片段形式

排布,富集范围为 N 端第 22-55,93-128,172-215,227-291,328-382,494-552。

2.3 B 细胞表位预测结果 以 Kyte-Doolittle 方案分析,蛋白亲水性较高的区域(hydrophilicity plot)为 N 端第 1-11,21-32,99-111,157-167,174-197,201-208, 226-233,240-254,261-269, 277-300,321-333,343-358,372-384,416-436,474-501,521-529。Jameson-Wolf 法预测抗原指数(antigenic index)较高的区域为 N 端第 1-11,21-37,52-59,99-114,123-130, 160-169, 173-183,189-196,203-208,214-221, 227-234,237-244,250-256,261-266, 278-292,323-335,345-353,373-384,392-399,427-435,459-472,478-504,520-531,536-541,548-555。大多数的高抗原指数区域与预测的亲水区重合。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 1-8,103-110,158-164, 175-183, 189-196,240-244,260-266,278-283, 289-299,322-332, 345-357, 372-381,417-423,427-434,478-488,523-528 区段很可能位于蛋白质的表面(图 2)。

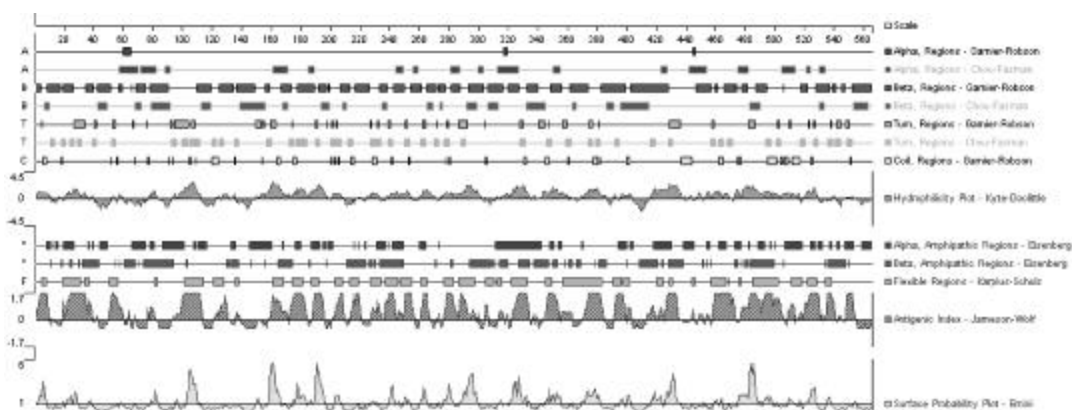


图 2 二级结构及 B 细胞抗原表位预测图

Fig 2 Predict the secondary structure and B cell epitope

3 讨论

迄今为止全世界已发现 6 种衣原体噬菌体,分别是 Chp1、Chp2、Chp3、Chp4、CPAR39 和 phiCPG1。第一株噬菌体 Chp1 发现后并未引起注意,随着衣原体感染病例及衣原体耐药性的增加,近十年衣原体的噬菌体研究逐渐引起重视,相继发现了 5 株新的噬菌体。研究显示衣原体噬菌体属微病毒家族,都有 20 面体衣壳和单链核苷酸序列^[3]。

衣原体噬菌体的衣壳是由 Vp1、Vp2、Vp3 3 种衣壳蛋白以符合热力学定律原则对称组装的^[2]。Vp1 是分子量最大的结构蛋白分子,其插入环 IN5 暴露

于衣壳表面,由多肽链的第 216 和 299 之间的氨基酸构成,能形成蘑菇样的突起^[4],推测 Vp1 蛋白参与噬菌体与宿主的识别过程。以各株噬菌体 Vp1 多肽序列进行比对并构建种系发生树发现 Chp1 与其他 5 株噬菌体的亲缘关系最远,其他 5 株噬菌体分为 2 个亚群:分别为 phiCPG1/phiCPAR39 和 Chp2/Chp3/Chp4。因此,我们对后 5 株噬菌体的 Vp1 蛋白序列展开分析。剔除 Chp1Vp1 序列后,发现 N 端第 1-215,300-565 区域保守度很高,未出现差异显著的比对空沟,说明这两个区域对 Vp1 蛋白分子的结构构成有重要价值。考虑到噬菌体的突变率很高,

上述两个区域在各种 Vp1 蛋白中能保有高度一致,说明其能稳定遗传、难以替代。因而对该区域的进一步结构分析将能为寻找通用标志物(表位)提供理论基础。

相对三级/四级结构预测的低可信度,二级结构预测的可信度较高且仍在不断改进。Gamier-Robson 法、Chou-Fasman 法、Eisenberg 法和 Karplus-Schulz^[5-6] 方案都预测到蛋白的 β 折叠比 α 螺旋的数量多,可以认为蛋白的二级结构以 β 折叠为主:该结构依靠氨基酸残基间的氢键形成“刚性”折叠,成为蛋白分子中心的“支架”,除特殊情况暴露于分子表面外,普遍难以被免疫系统识别^[7]。Vp1 蛋白基本的结构骨架是 β 折叠组成的, α 螺旋只是辅助的结构构成单位。转角和无规卷曲常常位于分子表面,因而有可能与抗体/受体作用,而且糖基化、磷酸化等蛋白加工过程也相应成为可能^[8]。转角这种“结构”在进化中保守性很强^[9],这种“保守性”提示转角对蛋白质的结构和功能进化都有重要意义。分析中发现 Vp1 蛋白除多变区外,其 N 端第 1-215,300-565 区域有频繁出现的转角和无规则卷曲结构,说明 Vp1 蛋白有较为复杂的空间结构,肽链不断折叠、卷曲,这样能形成大量的潜在突出于分子表面的蛋白结构,成为潜在的蛋白间相互作用区域。

研究表明暴露于蛋白表面的区域容易被识别,潜在表位的蛋白分子表面可及性分析必不可少^[10]。同时,蛋白质分子中疏水性氨基酸残基一般埋藏在蛋白质分子内部,而亲水性氨基酸残基位于蛋白质分子表面,亲水性部位分析也成为暴露区域分析的重要一环。相对肽链中氨基酸残基的性质,同时提出了抗原指数分析方案^[11]。抗原指数分析结果表明有 25 个集中的抗原表位区域且指数值较高,并与亲水性部位、表面可及性综合分析,认为蛋白 N 端第 1-8,103-110,158-164, 189-196, 252-259, 273-284,322-332, 427-434,478-488 是优势 B 细胞表位。除去 252-259, 273-284 两个表位,其它均位于序列保守区。上述区域均避开了分析的“刚性结构— α 螺旋及 β 折叠”区域。

目前已经从人类衣原体中发现了一株衣原体噬菌体—phiCPAR39,并观察到其与肺炎衣原体株

CPAR39 的相互作用。因而对噬菌体 Vp1 蛋白高保守区的分析将为在更广范围内寻找人类衣原体噬菌体奠定基础。同时高保守区分析也为重组和建立衣原体基因输送工具奠定了基础。此外,衣原体噬菌体具有不同的、重叠的细胞嗜性特点—即它可以感染原来的宿主衣原体也可感染其他数种衣原体,为今后构建多宿主敏感型重组衣原体噬菌体提供了理论支持。

参考文献:

- [1] Saka H A, Valdivia R H. Acquisition of nutrients by *Chlamydiae*: unique challenges of living in an intracellular compartment[J]. Curr Opin Microbiol, 2010, 13(5): 4
- [2] Sait M, Livingstone M, Graham R, et al. Identification, sequencing and molecular analysis of Chp4, a novel chlamydia phage of *Chlamydia abortus* belonging to the family Microviridae[J]. Gen Virol, 2011, 92(7): 1733
- [3] Sliwa-Dominiak J, Suszyńska E, Pawlikowska M, et al. Chlamydia bacteriophages[J]. Arch Microbiol, 2013, 195(10): 765
- [5] Kodaira K, Nakano K, Okada S, et al. Nucleotide sequence of the genome of the bacteriophage 3: interrelationship of the genome structure and the gene products with those of the phages, X174, G4 and K[J]. Biochim Biophys Acta, 1992, 30(11): 277
- [6] Ponomarenko J V, Bourne P E. Antibody-protein interactions: benchmark datasets and prediction tools evaluation[J]. Struct Biol, 2007, 7(63): 1
- [7] Su C H, Pal N R, Lin K L, 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): 617
- [8] Larsen J E, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes[J]. Immunome Res, 2006, 2(2): 1
- [9] Beaver J E, Bourne P E, Ponomarenko J V. EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the immune epitope database and analysis resource [J]. Immunome Res, 2007, 2(2): 3
- [10] Yoo S Y, Merzlyak A, Chung W J, et al. Facile growth factor immobilization platform based on engineered phage matrices[J]. Soft Matter, 2011, 14(7): 1660
- [11] 郑小莉, 张艺, 谢晓东, 等. 鼠 Minele 蛋白 B 细胞抗原表位预测及多克隆抗体制备[J]. 免疫学杂志, 2011, 27(8): 683
- [12] Saraswathi S, Fernández-Martínez J, Koliński A, et al. Distributions of amino acids suggest that certain residue types more effectively determine protein secondary structure[J]. Mol Model, 2013, 19(10): 4337

(2013-12-13 收稿)