

登革病毒 GD01/98 结构蛋白基因序列及 蛋白二级结构初步分析^{*}

张勤奋¹, 陈森雄¹, 蒋廉华², 卢耀英¹, 张景强¹

(1. 中山大学生物防治国家重点实验室, 广东 广州 510275;

2. 广州军区军事医学研究所, 广东 广州 510507)

摘要: 对登革病毒 GD01/98 结构蛋白基因采用 RT-PCR、分子克隆等方法进行测序, 获得了分离株的结构蛋白基因核苷酸序列 2 419 bp; 并通过 Internet、用 Blast 等软件进行同源性分析, 发现它同泰国分离株最近, 可能来源于泰国; 根据核苷酸序列, 推导蛋白一级结构, 并用 GOR、神经网络 (Hierarchical Neural Network) 等方法进行二级结构预测, 发现有 46.74% 氨基酸参与无规卷曲, 有 31.16% 氨基酸参与 α -螺旋, 没有 β 折叠。

关键词: 登革病毒; 结构蛋白; 核苷酸序列; 蛋白二级结构

中图分类号: Q71 **文献标识码:** A **文章编号:** 0529-6579 (2002) 02-0072-04

登革病毒引起的登革热 (Dengue Fever, DF)、登革出血热 (Dengue Hemorrhagic Fever, DHF) 和登革休克综合症 (Dengue Shock Syndrome, DSS) 是危害较大的急性传染病^[1,2]。但由于抗体依赖性感染增强作用 (ADE), 到目前为止还没有有效的疫苗可以防治登革病毒所引起的这些疾病。登革病毒的结构蛋白是组成病毒颗粒的主要成分, 也是毒粒的形态结构的主要物质基础。结构蛋白存在重要的抗原决定簇、受体结合位点等功能决定簇, 也与 ADE 有关系^[3,4]。因此研究登革病毒的结构蛋白对揭示这些活性位点的结构与功能, 控制其引起的疾病有重要的作用。此外, 已有研究表明, 毒力的变化除与登革病毒的血清型有关外, 也与毒株的疫源有很大的关系, 如: 登革-2 型东南亚株的毒力比其它地方株的强很多^[5], 因此, 搞清病毒的来源对控制其流行及治疗有重要作用。

本文提取了 1998 年在南海市分离得到的登革病毒 2 型分离株 (GD01/98) 的基因组 RNA, 通过设计的特异引物, 应用 RT-PCR、分子克隆等技术对该基因组 RNA 进行了研究, 获得了结构蛋白基因的核苷酸序列。并通过 Internet、用 Blast 等软件进行同源性分析; 同时根据核苷酸序列推导的结构蛋白氨基酸序列, 应用计算机及 Internet 等技术进行了结构蛋白二级结构的初步分析, 这些结果将为登革热等疾病的防治提供依据。

1 材料与方法

1.1 病毒、质粒和菌种

登革病毒 2 型 GD 01/98 株由广州军区军事医

学研究所分离、鉴定和保存^[6], 并按冯建勋方法培养、提纯^[7]。载体 pGEM-Teasy 购自 Promega 公司; 宿主菌 *E. coli* DH5 α 为中山大学海洋功能基因组国家开放实验室提供。

1.2 引物的设计和合成

用 Oligo 软件, 参照登革病毒-2 型 NGC 株的核苷酸序列设计 2 对特异引物:

引物 1

上游 5' GAC AGA TTC TTT GAG GGA CCT AAG 3'

下游 5' CTC TTG TCT ACT GAC GAT CAT GTG 3'

引物 2

上游 5' TCC AGC TCA CAA CGC AAC CAC TAT C 3'

下游 5' CCACAC GTA ACG GAG AAC CAC ACAT 3'

并由上海生工生物工程公司合成。

1.3 cDNA 合成和 PCR 扩增

将纯化的病毒用 GIBCO-BRL 公司的 TRIzol 细胞总 RNA 试剂盒, 参照 TRI Reagent 一步法分离病毒 RNA^[8], 用下游引物, 参照 cDNA Synthesis Kit 试剂盒说明方法进行病毒 cDNA 的合成, 以 cDNA 为模板, 按常规反应条件, 进行 32 个循环的 PCR 扩增, 反应体积为 50 μ L, 循环参数为: 94 $^{\circ}$ C 变性 3 min 后, 94 $^{\circ}$ C 35 s, 58 $^{\circ}$ C 50 s, 72 $^{\circ}$ C 2 min。

1.4 PCR 扩增产物的纯化

将 PCR 产物进行琼脂糖凝胶电泳, 切取与目的基因大小相符的核酸带, 参照 QIAGEN 的 QIA-quick Gel Extraction Kit 的说明进行提纯。

1.5 DNA 片段的克隆与测序

将 PCR 扩增产物与 Promega 公司的 pGEM-T

^{*} 收稿日期: 2001-08-12

基金项目: 国家自然科学基金资助项目 (39770161); 广东省自然科学基金资助项目 (970149); 全军医药卫生科研基金资助项目 (01Z014)

作者简介: 张勤奋 (1966-), 女, 讲师, 通讯联系人, 张景强; Email: js28@zsu.edu.cn

Easy Vector 连接, 转化到 *E. coli* DH5 α 菌株中, 筛选阳性克隆子, 培养至对数生长期, 用 Omega Bio-tek, Inc. 的 E. Z. N. A. Plasmid Miniprep Kit I, II 提纯质粒 DNA, 用 *Eco*R I 酶切, 琼脂糖凝胶电泳检测。用该 DNA 为模板, 由上海生工生物工程公司测序。

1.6 序列同源分析

摘录国际基因库 (GenBank) 中的几株登革 2 型病毒的核苷酸序列: 中国分离株 china04 (AF119661), Den2 NGC (D00346), Thai (AF022434), Strain 16681 (M84727.1), DV2 Jamaica (M0558.1), DV2 strain IQT1797 (AF100467), ThNH-793 (U31949), Puo 218 strain (D00345) 等, 用 BLAST, DNA STAR, Clustal W 等软件进行核苷酸序列分析和比较。

1.7 结构蛋白的二级结构初步分析

由所获得的 cDNA 序列, 推导氨基酸序列, 再通过 Internet, 用 GOR 二级结构预测、神经网络 (hierarchical neural network) 等方法对蛋白的二级结构进行预测分析^[9,10]。

2 结 果

2.1 GD 01/98 分离株结构蛋白基因克隆和测序

通过 2 对特异引物经 RT-PCR 对 GD 01/98 分离株的核酸进行扩增, 可分别获得产物约为 460 bp 和

2 000 bp 的片段 (图 1), 经克隆至 PGEM T Easy 载体后, 提取质粒, 重组子可得到这 2 片段。测序结果表明, 第 1 片段共有 465 bp, 包含了登革病毒全部的 C 蛋白基因和部分 PrM 基因, 第 2 个片段有 2 086 bp, 包含了全部的 PrM, M, E 蛋白基因和部分 C, NS1 的基因。将 2 片段连接起来获得 2 419 bp, 包含了全部的结构蛋白 C (70—411)、prM (411—680)、M (681—905)、E (905—2390) 的基因和部分 5' 端非编码区 (1—69)、部分的 NS1 (2391—2419) 基因, 并在所测得的第 70 个碱基上存在起始密码子, 没有终止密码子 (图 2)。

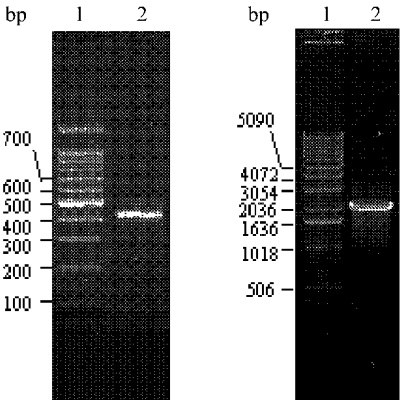


图 1 结构蛋白基因片段扩增结果
Fig. 1 PCR amplification of structural protein gene fragment
1 DNA 相对分子量标准; 2 PCR 产物

1 GAC AGATTCT TTAGGGGAGC TAGCTCAAC GTAGTCTCT AAC AGTTTTCTTA ATT AGAGAGC AG ATCTCTGA TGAATAACC A ACGGAAAAAG GC GAAAAACCA
101 CGCCTTTC AA TATGC TGAAA CCGC AG AGAA ACC GC GTGTC G ACTGTGTC AA CAGCTG AC AA AG AGATTTC ACTTGGAATG CTGCAGGG AC GAGG ACC ATT
201 AAAACTGTTC ATGGCCCTTG TGCC GTTCTT TCGTTTCTCA TCAATCCAC CAACAGCAGG GATATTTGAAG AGATGGGGAA CAATT AAAAA ATCAAAAGCC
301 ATC AATGTTT TG AGAGGGTT C AGG AAAGAG ATTGG AAGGA TGCTGAACAT CTTGAATAGG AGACGCAG AT CTGC AGGCAT GATTAATTATG CTGATTC CAA
401 C AGTAATGGC GTTCC ATC TA ACCACACGT A ACGGAGAACCC AC AC ATG ATC GTC AGT ATAC AAGAG AAAGG G AAAAGTCTT CTGTTTAAAGA CTGAGGATGG
501 CGTGAACATG TGC ACCCTCA TGGC CATGG A CCTTGGTG AA TTGTGTG AAG ACACAACTAC GTACAAGTGT CCCCCTTCTAA GGC AGAATGA GCCAGAAGAC
601 ATAGACTGTT GGTGC AATTC C ACGTCCAC A TGGGT AACCT ATGGGACTTG TACCACACG GGAGAACATA GAAGAGAAAA AAGATCAGTG GCATTGTTC
701 CAC ATGTGGG AATGGGACTG GAGACGCGAT CCGAAACATG GATGTCATCA GAAGGGCTT GGAAACATGC CAGAGAATT GAAACTTGA TCTGTGAGACA
801 TCCAGGCTTC ACCAT AATGG C ACG AATCC T GGC AT ACACC ATAGGGACGA CAC ATTTCC A AAGAGCCC TG ATTTTCATCC TACTGACAGC TGTCGTCCT
901 TCAATGACAA TGCCTGTC AT AGGAATATC A AATAGAGACT TTGTAGAAG GGTTCAGG A GGAAGCTGGG TTGACATAGT CTTAGAAC AT GGAAGCTGTG
1001 TGACGACGAT GGC AAAAAC AAACCAACAT TGGATTTTGA ACTGATAAAA ACGGAAGCC A AACAGCCTGC CACCCTAAGG AAGTACTGCA TAGAAGC AAA
1101 GCTAACCAAC ACACACAG AATCTCTTTG TCC AACAC AA GGG AACCCA GCCTAAAG A AGAG AGG AT AAGAGGTTTG TCTGC AAACATCC ATAGTA
1201 GAC AGAGGGT GGGAATGG ATGTGGAT A TTTGG AAAGG GAGGCATGT GACTGTGCT ATGTTTAC AT GC AAAAAGAA CATGG AAGGG AAAATCGTGC
1301 AACCAGAAAACTTGG AATAC ACCATTGTGG TAAACACCTCA CTCAGGGGAA GAGCATGCG A TC GGAATGA C ACAGAAAA CATGGC AAGG AAATCAAGT
1401 AAC ACCACAG AGTTTCGTCA C AGAAGCAG A ATTGACAGT TATGGCACCG TCACGATAG A GTGCTCTCCG AGAACAGGC TCGACTTC AA TGAGATGGTG
1501 TTGCTGCGAG ATGGAAAT AA AGCTTGCTG GTGCATAGC AATGGTTCTT AGACCTGCC A TTACCATGGC TGCCCGGAGC GGACACAC AA GGATCAAT
1601 GGATACAGAA AGAAGATGT GTCACTTTC A AAAATCCCCA TGCGAAGAAA CAGGATGTTG TTGTTTATAG ATCCCAAGAA GGGGCCATGC ATACAGC ACT
1701 CACAGGAGCC ACAGAAATCC AATGTGCTGC AGGAACTTA CTCTTACCCG GACATCTCAA GTGCAGGC TGAGAATGGACG AGCTACAGCT CAAAGGAATG
1801 TCATATCTCC ATGTGCACAGG AAGTTTCAAA GTTGTGAAGG AATAGCAGA AACACACAT GGAAAGATAG TTATTCAGAGT GCATATAG AA GGGGACGGCT
1901 CTCCATGTAA AATCCCTTTT GAGATAATGG ATTTGGAAAA GAGATATGTC TTAGGCGCC TGATCACAGT CAACCAATTT GTGACAGAAA AAGATAGCCC
2001 AGTCAACAT A GAGC AGAAC CTCC ATTCGG AGACAGCT AC ATCATCATAG GAGTAGAGCC GGGAC AACTG AAGCTCAC TGGTTC AAGAA AGGAAGTTCT
2101 ATCGGC AAAA TGTTTGAGAC AACATGAGG GGGC GAAGA GAATGGCCAT TTTGGGTGAC ACAGCCTGGG ATTTCCGGATC CCCGGGAGGA GTGTTTACAT
2201 CTATAGGAAA AGCTCTCC AC CAGTCTTTG GAGCAATTA TGGAGCTGCC TTCAGTGGG TTTCATGG AC TATGAAAATC CTCATAGGAG TCATATATAC
2301 ATGGATAGG AATGAACCTC ACAGCACCTC ACTGTCTGTG TCATAGTAT TGGTAGGAAT TGTGACATG TACTTGGGAG TCATGTTGCA GGGCAGATAGT
2401 GGTTCGCTTG TGACCTGA

图 2 GD 01/98 登革病毒结构蛋白基因 cDNA 核苷酸序列
Fig. 2 The cDNA nucleotide sequence of the structural proteins of Dengue virus II GD 01/98

3 讨 论

从结构蛋白基因系统树看, GD 01/98 与毒力强的东南亚疫源的泰国分离株同源, 而与中国 04 分离株不同源, 这一点同其非结构蛋白 NS1 部分序列分析的结构相似^[9]。蛋白二级预测表明登革病毒 2 型 GD 01/98 株的结构蛋白二级结构中最多的结构类型是无规卷曲, α 螺旋只占约 1/3, 但衣壳蛋白(C)中的 α 螺旋却占 72%。 α 螺旋是相当稳定的结构状态^[10], 这结果可以显示登革病毒结构蛋白中的衣壳蛋白结构相对紧密而稳定, 对病毒的形态起稳定作用。而前膜蛋白(PrM)中的 α 螺旋却只占 10% 左右, 说明其形态极不稳定。囊膜蛋白 E 的 α 螺旋约占 1/3, 这也说明其形态也是较为多变的。这些蛋白的结构特征与登革病毒的形态特征是相符的。

登革病毒的 E 蛋白在病毒的最外表, 其上有重要的抗原决定簇、受体结合位点等功能决定簇^[1-4]。已有实验证明 E 蛋白刺激机体产生的抗体能引起抗体依赖性感染增强作用 (ADE), 与登革出血热和登革休克综合征密切相关。在 E 蛋白的结构模型中, 有 3 个非重叠的抗原决定簇: A, B, C^[10]。GD 01/98 株同中国 04 株在 N 端 82—96 个残基的变化正好在 A 区, E 蛋白结构的这些变化, 尤其在 A 区的结构变化对其抗原决定簇的结构将会产

生影响, 进而有可能对其功能, 如毒力的变化产生不可忽视的作用。

参考文献:

[1] THOMAS P M. Dengue; the risk to developed and developing countries[J] . Proc Natl Acad Sci USA, 1994, 91: 2395—2400.

[2] HALSTEAD B S. Pathogenesis of Dengue; challenges to molecular biology[J] . Science, 1988, 239: 476—491.

[3] KAUFMAN M B. Monoclonal antibodies for dengue virus PrM glycoprotein protect mice against lethal Dengue infection[J] . Am J Trop Med Hyg. 1989, 41(9): 576—580.

[4] GOLLINS S W, PORTERFIELD J S. A new mechanism for the neutralization of enveloped viruses by antiviral antibody [J] . Nature, 1986, 321: 244—246.

[5] NICHOLAS J W. Variation in virulence of Dengue virus[J] . The Lancet, 1999, 354: 1401.

[6] 方美玉, 赵文忠, 刘建伟, 等. 登革病毒 2 型广东省分离株的鉴定及 NS1 基因序列分析[J] . 中华传染病杂志, 2001, 19(1): 18—21.

[7] 冯建勋, 卢 英, 赵春生, 等. 用于冷冻电镜的登革病毒 II 样品制备[J] . 中国学术期刊文摘, 1999, 5(6): 765—766.

[8] 李永明, 赵玉琪. 实用分子生物学方法手册[M] . 北京: 科学出版社, 1998: 205—251.

[9] GIBRAT J F, GARNIER J, ROBSON B. Further developments of protein secondary structure prediction using information theory[J] . J Mol Biol, 1987, 198: 425—443.

[10] 陈惠黎. 生物大分子的结构和功能[M] . 上海: 上海医科大学出版社, 1999: 16—42.

The Nucleotide Sequence and Secondary Structure
of the Structural Protein of Dengue Virus GD01/98 Strains

ZHANG Qin fen¹, CHEN Sen xiong¹, JIANG Lian hua², LU Xin ying¹, ZHANG Jing qiang¹

(1. State Key Laboratory for Biocontrol, Sun Yat-sen (Zhongshan) University, Guangzhou, 510275;
2. Military Medical Institute of Guangzhou Military District, Guangzhou 510507, China)

Abstract: The sequence about 2 419 bp of the structural protein gene of Dengue Virus GD01/98 strain was obtained by using RT-PCR, Clone technology etc. It could be concluded that the strain might be Thailand resources for the nucleotide sequence has highest homologous with ThNH 7/93. The preliminary secondary structure was studied by Gibrat (GOR3) methods and Hierarchical Neural Network methods together with the internet. The results demonstrated that there were 46. 74% random coils and 31. 16% α helix.

Key words: Dengue virus; structural protein; nucleotide sequence; secondary structure of protein