

口蹄疫病毒株 OA/58 L 蛋白酶的结构构建和功能分析

周建华, 丛国正, 高闪电, 常惠芸

(中国农业科学院兰州兽医研究所, 家畜疫病病原生物学国家重点实验室,
农业部畜禽病毒学重点开放实验室, 国家口蹄疫参考实验室, 甘肃 兰州 730046)

摘要:以口蹄疫病毒株 OA/58 RNA 为模板, 反转录并扩增目的 cDNA, 然后与 pGEM-T Easy 载体连接并转化 JM109 菌株, 提取的重组质粒用凝胶电泳、PCR 和 *EcoR* I 酶切法鉴定。该毒株与 A12, O1K, O1Campos 和 TW45/97 毒株序列通过对比分析发现, *L* 基因中有 2 个起始密码子, 第二个是首选; 并且确定氨基酸保守区主要位于第 35~39, 43~54, 65~67, 75~80, 90~111, 113~142, 144~146, 148~157, 159~172 和 176~187 位。L 蛋白酶含有球状区域, 碳端有一柔性杆状结构。合成的 L 蛋白酶可形成二聚体结构。第 144 位的 Lys、148 位的 His 和 163 位的 Asp 可能是 L 蛋白酶的活性位点。保守区氨基酸残基在维持蛋白的空间构像和功能方面具有重要作用。由拉马钱德兰图可证明本试验构建 L 蛋白酶的空间结构是合理的, 它对于进一步的研究具有指导性。

关键词: 口蹄疫病毒; L 蛋白酶; 活性位点

中图分类号: S852.66 文献标识码: A 文章编号: 1000-7091(2007)04-0172-04

Construction and Analysis of L Protease Structure and Function from a Foot-and-mouth Disease Virus Strain OA/58

ZHOU Jian-hua, CONG Guo-zheng, GAO Shan-dian, CHANG Hui-yun

(Chinese Academy of Agriculture Science, State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Animal Virology of Ministry of Agriculture, National FMD Reference Laboratory, Lanzhou 730046, China)

Abstract: Foot-and-mouth disease virus strain OA/58 RNAs were used as templates for RT-PCR. The amplified cDNA products were cloned into pGEM-T Easy vectors and transformed into JM109. The recombinant plasmids were identified by electrophoresis, PCR, and *EcoR* I cleavage. The nucleotide and amino acid sequences were compared with the *L* genes of the other 4 reference strains. The results show that between 2 initiation codons, there is a special function sequence which enables small subunit of ribosome to recognize and utilize the second AUG to translate L protease that is called for Lb; the regions of 35–39th, 43–54th, 65–67th, 75–80th, 90–111th, 113–142th, 144–146th, 148–157th, 159–172th and 176–187th in L protease most probably are conservative. By homology modeling the FMDV strain OA/58 L protease, the 3D mold of this protease was obtained. Resolution of the 3D structure of L protease showed a compact globular form with a flexible C terminal extension from 187th to 201th. Depending on the region of hydrophilicity residues forming hydrophilicity power, L proteases can form dimmers. Lys₁₄₄, His₁₄₈ and Asp₁₆₃ may be active amino acids which form the active site of L protease. By Ramachandran Plot showing, 3D mold of a FMDV strain OA/58 L protease is rather reasonable, this research should be used as an instruction in order to direct the work on FMDV L protease.

Key words: FMDV; L protease; Active site

口蹄疫是由口蹄疫病毒(Foot-and-mouth disease virus, FMDV)引起的一种急性、烈性、高度接触传染性和可快速远距离传播的动物疾病。口蹄疫的暴发给当地带来严重的经济损失, 疫区乃至整个国家的

畜产品出口受到封锁。这使得口蹄疫成为家畜疾病中的头号“经济病”^[1]。FMDV 属于微 RNA 病毒科, 口疮病毒属, 其基因组为正链单股 RNA, 它既是转录中的 mRNA, 又是负链合成的模板, 长度约为 8 500

收稿日期: 2007-04-18

基金项目: 国家“973”项目资助(2005CB523201)

作者简介: 周建华(1981-), 男, 内蒙古呼和浩特人, 硕士, 主要从事病毒基因工程研究

通讯作者: 常惠芸(1965-), 女, 河南许昌人, 研究员, 博士, 主要从事病毒基因工程的研究工作。

个核苷酸。L 蛋白酶是 FMDV 基因组编码的第一个蛋白质, L 蛋白具有自我剪切功能, 可从多聚蛋白中自动裂解下来, 形成具有活性的蛋白酶^[2]。在 7 个血清型 FMDV 基因组中对应的 L 蛋白编码区均存在 2 个起始密码子 (AUG), 它们起始编码的多聚蛋白裂解后形成 2 种 L 蛋白酶, 分别是 Lab 和 Lb。虽然 2 种 L 蛋白酶已经在体外感染的细胞中检测到, 但体内试验证实 Lb 是被首选合成的。L 蛋白酶是一种毒性蛋白, 它主要的功能是裂解宿主细胞中的真核转录起始因子 eIF4G 和 p220 蛋白。

同源模建方法是蛋白质三维结构构建的主要方法。对蛋白质数据库 PDB 分析可以得到这样的结论: 任何一对蛋白质, 如果两者的序列一致部分超过 30%, 则它们具有相似的三维结构, 即 2 个蛋白质的基本折叠相同, 只是在非螺旋和非折叠片层区域的一些细节部分有所不同。对于具有 60% 一致部分的序列, 用同源模建方法确定的三维模型非常准确。若序列的一致部分超过 60%, 则构建结果将接近于试验得到的测试结果。为了明确 FMDV L 蛋白酶的结构和功能, 本试验就 FMDV L 蛋白酶的氨基酸序列和结构功能分析进行研究。

1 材料和方法

1.1 材料

FMDV OA/58 株由国家口蹄疫参考实验室种毒保藏中心保存。主要试剂分别购自大连宝生物公司、博大泰克公司和 Promega 公司。上游引物和下游引物是参考其他 FMDV 毒株 L 基因两端保守区设计的。上游引物: 5'-GTGCCCAGGTTAAAAGCTTC-3'; 下游引物: 5'-CAGTCGTTGTTCTGGGTGTTG-3'。

1.2 方法

1.2.1 目的基因扩增 病变组织经研磨后, 按照提毒试剂盒说明书提取病毒总 RNA。参考《分子克隆》方法进行反转录和扩增。扩增产物经琼脂糖凝胶电泳纯化和琼脂糖凝胶 DNA 提取试剂盒回收。

1.2.2 克隆与测序 按照 pGEM-T easy 载体试剂盒说明书进行 PCR 产物与载体的连接, 并用 JM109 菌株进行连接产物的转化和克隆。然后提取质粒, 进行凝胶电泳、PCR 和 *EcoR* I 酶切鉴定, 样品送上海生工测序。

1.2.3 序列分析 用 DNASTAR 软件分析比较 OA/58 和 China/99 (本实验室测定), OIK, A12, TW45/97 (源自 GenBank) L 基因序列和氨基酸序列, 从而确定 OA/58 L 蛋白酶氨基酸序列的保守区域。

1.2.4 搜索 PDB 数据库 用 NCBI 的 BLAST 程序分

析 OA/58 L 蛋白在蛋白质数据库 (PDB) 中的同源蛋白, 查询结果显示 PDB 代号为 1QMY 的 FMDV L 蛋白酶三聚体中的亚基氨基酸序列与 OA/58 L 蛋白的一致性高达 92.9%, 以 1QMY 作为同源模建的模板。

1.2.5 同源模建 我们根据同源模建的思路^[3], 运用 Swiss-pdbViewer 蛋白质分析软件构建 L 蛋白酶的 3D 模型。最后利用 PROCHECK 程序检测构建的 L 蛋白酶 3D 模型是否合理。

2 结果与分析

2.1 FMDV OA/58 L 基因 TR-PCR 和转化结果

凝胶电泳显示 PCR 扩增产物约 600 bp, 这与预期的扩增长度一致。从培养 JM109 转化菌的平皿中挑取单克隆扩增培养后提取质粒后经琼脂糖凝胶电泳、PCR 和 *EcoR* I 酶切鉴定, 目的片段长约 600 bp, 证明所克隆的 cDNA 为 L 基因。

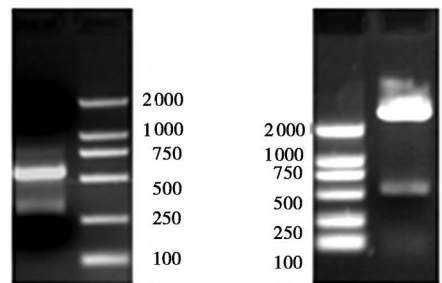


图 1 L 基因 PCR (左) 和 *EcoR* I 酶切鉴定 (右)

Fig. 1 (Left) electrophoresis of L gene; (Right) identification of *EcoR* I cleavage

核苷酸序列 (Nucleotide sequence):

ATGAACACAACCTGATTGTTTATCGCTTTGGTGACGCACTCAGAGAGATTAAGACACTGT
TTCCTTTCACGGACACAAGGAAAAATGGAATTCACACTGTACAATGGTGAGAAGAAGACAT
TTTACTCCAGACCAACAACACGACAACTGCTGGCTCAACACCATTCCTNAGTTGTTTCAG
GTACGTCGATGAACCTTTCTTCGACTGGGTCTACAACCTGCCTGAGAACCTTACTCTTGAG
GCTATTAGGCAGTTGGAGGAGGTAACCTGCTTTCGCTGACGAGGGTGACCGCCCGCT
CTCGTCATTGGAACATCAACACTTGTCTCACACCGGAATCGGTACTGCCTCGCGCCCA
GCGAGGTGTGTATGGTAGACGGGACGACATGTGCTTGGCTGACTTCCACGCTGGCATT
CTTGAAAGGACAAGAACACGCGCTGTTTGCCTGCGTTACCTCCAACGGGTGGTACGCGAT
CGACGACGAGGACTTTTACCCTGGACGCGGACCGCTCCGATGTCTTGGTGTGTTGCCCG
TACGATCAAGAACCGCTCAATGGGGAATGGAAGCAAGGTTTCAAGACGGCTCAAG

氨基酸序列 (Amino acid sequence):

MNTTDCFIALVHALREIKTLFLSRITQKMEFTLYNGEKKTFYSRPNHNDNLNTILXLFER
YVDEPFDDWYNSPENLTLEAIRQLEEVTTGLEHGGPPALVIWNKHLHTGIGTASRPS
EYCMVDGTDMLADFHAGIFLKGQEHAFACVTSNGWYAIIDDEDFYPTPDPSDVLVFPVY
DQEPLNGEWKAKVQKRLK

图 2 OA/58 L 基因核苷酸序列和推导的氨基酸序列

Fig. 2 The nucleotide and amino acid sequence of L gene in FMDV strain OA/58

2.2 FMDV OA/58 L 蛋白酶核苷酸序列和氨基酸推导序列

OA/58 L 基因序列全长 603 个核苷酸, 编码 201 个氨基酸组成的多肽链 (图 2)。

低。第二个 AUG 前后具有真核核糖体小亚基识别的特征性序列 AAAAAUGG, 并且起始甲硫氨酸乙酰化。此外, 2 个 AUG 之间有一段连续的嘧啶区, 与真核 18S 核糖体 RNA 的 3 端互补, 这些特性决定了第二个 AUG 使用频率高。由此可知, Lb 是 L 蛋白酶的主要形式(图 2 氨基酸序列下划线表示)。

从图 3 和 4 可知, L 蛋白酶 3D 模型可被划分为 2 个区域, 一个为球状区域, 另一个为具有柔性的杆状区域(187~201aa), 这与 Guarne 等^[5]提出的 L 蛋白酶在其碳端尾部(Asp₁₈₄至 Lys₂₁₀)研究结果基本一致。图 3 中的 1 号 α -螺旋由 11 个氨基酸残基组成并且亲水性氨基酸残基和疏水性氨基酸残基在螺旋中有一定的规律性, 即亲水性氨基酸残基多出现在亲水侧面(位于 L 蛋白表面), 而疏水性氨基酸残基多出现在疏水侧面(朝向 L 蛋白疏水核心), 这种现象在一级结构中表现为亲、疏水性残基间隔模式, 这种模式有利于 L 蛋白酶高级结构的稳定。2 号 α -螺旋由 6 个非极性氨基酸构成, 其中 Ala, Leu, Val 和 Ile 具有极强的疏水性。虽然 2 号 α -螺旋位于 L 蛋白酶表面, 但其周围存在有大量非极性氨基酸, 其数量多于同区域中极性氨基酸的数量, 由此确定此区域为 L 蛋白的疏水区域。由这一结论我们认为 L 蛋白酶将利用这一疏水区域彼此之间通过疏水作用力形成二聚体, 其结果使疏水区隐藏在二聚体内部, 确保 L 蛋白酶结构稳定。这一推测与 Guarne 等^[6]和 George 等^[7]提出的 L 蛋白酶是以二聚体形式发挥酶学活性的研究结果基本一致。图 3 中 3 号 α -螺旋位于 L 蛋白酶内部, 由 13 个氨基酸残基构成, 其中 9 个为非极性氨基酸残基并且位于螺旋疏水侧面(朝向 L 蛋白酶疏水核心区), 4 个极性氨基酸残基位于亲水侧面, 此 α -螺旋呈现两亲性有利于结构的稳定。图 3 中 4 号 α -螺旋由 6 个氨基酸残基构成, 虽然在序列上表现为极性、非极性氨基酸残基间隔出现, 但不具有两亲性。蛋白质所具有的功能在很大程度上取决于其空间结构, 而 OA/58 L 蛋白酶核苷酸序列中的保守区在维持 L 蛋白酶的空间构象并且为 L 蛋白酶的活性氨基酸有效发挥其功能等方面发挥着重要作用。因此对于蛋白质空间结构的研究在蛋白质工程中有其极其重要的意义。

从图 6 可知, Lys₁₄₄-His₁₄₈-Asp₁₆₃为 L 蛋白酶催化三联体。其中 Lys₁₄₄侧链上 -氨基基团带有剩余的电子对, 同时它的侧链是 4 个碳的直链, 柔性较大, 使侧链的 -氨基反应活性增大, 可以对底物缺电子基团发动亲核攻击, 从而拉近酶与底物的距离。

His₁₄₈中的咪唑基值得注意, 因为它既是一个很强的亲核集团又是一个有效的广义酸碱功能集团。在生理 pH 值条件下, 咪唑基既可以作为质子的供体又可以作为质子的受体, 而且供出或接受质子的速度几乎相等。因此, 咪唑基是催化中最有效最活泼的催化集团。Asp₁₆₃侧链中的羧基在生理条件下带负电, 有吸引质子的倾向。His₁₄₈中咪唑基团在 Asp₁₆₃侧链中的羧基影响下, 使咪唑基团去质子化, 这使咪唑基团具有亲核攻击的能力。这说明 L 蛋白酶对底物的酶切反应是亲核催化。

X 射线晶体学方法是至今为止研究蛋白质结构最有效的方法, 但是由于对样品的需要量大、纯度要求高, 被测定的蛋白质的分子量一般不超过 20 000 Da 等原因, 也受到很大限制。对蛋白质结构的研究是现代分子生物学的基础之一, 可以说分子生物学所进行的一切研究都离不开结构信息做指导。因而迫切需要一种不依赖晶体培养、迅速、简便易行的测定蛋白质结构的方法。利用蛋白质的一级结构所提供的氨基酸序列信息来进行高级结构预测的方法就是应这种需要而发展起来的。本试验就是以此为目的, 对 FMDV OA/58 L 蛋白酶进行了有意义的尝试, 并且分析的结果许多都与前人的研究基本吻合, 这说明此种方法运用到 FMDV 编码蛋白功能的研究是可行的, 对试验具有指导意义。

参考文献:

- [1] Grubman M J, Baxt B. Foot-and-mouth disease [J]. Clin Microbiol Rev, 2004, 17: 465-493.
- [2] Skern T, Fita I, Guarne A. A structural model of picornavirus leader proteinases based on papain and bleomycin [J]. J Gen Virol, 2001, 79: 301-307.
- [3] 丁达夫, 汤海旭, 张保红, 等. 同源蛋白质三级结构的预测[M]//理论物理与生命科学. 上海: 上海科学技术出版社, 1997: 119-148.
- [4] Gorbalenya A E. Putative papain-related thiol proteases of positive-strand RNA viruses [J]. FEBS Lett, 1991: 201-205.
- [5] Guarne A, Hampoelz B, Glaser W, et al. Structure and biochemical features distinguish the foot-and-mouth disease virus leader proteinase from other papain-like enzymes [J]. J Mol Biol, 2000, 302: 1227-1240.
- [6] Guarne A, Tomo J, Kirchweger R, et al. Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition [J]. EMBO, 1998, 17: 7469-7479.
- [7] George M. The non-structural leader protein gene of foot-and-mouth disease virus is highly variable between serotypes [J]. Virus Genes, 2001, 22: 271-278.