

百合无症病毒衣壳蛋白基因克隆和蛋白分析

徐品三*, 刘纪文, 高晓蓉, 尹雅蕾

(大连理工大学 生命科学与技术学院, 辽宁 大连 116024)

摘要: 根据已报道的 LSV CP 基因序列合成两条寡聚核苷酸引物, 模板为感染 LSV 的百合叶片的总 RNA, 通过反转录-聚合酶链式反应(RT-PCR)扩增出大小为 876 bp 的 LSV CP 基因, 经测序后, 对该基因编码区全长序列及相应的氨基酸序列用生物信息学软件系统进行序列分析及结构功能预测. 结果表明: 该基因由 876 个核苷酸组成, 编码 291 个氨基酸; 与 GeneBank 公布的其他 LSV 分离物的基因序列同源性为 93.4%~99.0%, 氨基酸同源性为 84.8%~99.5%; 它含有一个卷曲螺旋结构和多个磷酸化位点, 平均疏水值为 -0.432; 含有 *Carlaviruses* 完整的衣壳蛋白保守结构域, 二级结构以 α -螺旋和无规则卷曲为主.

关键词: 百合无症病毒; 衣壳蛋白; 序列分析; 功能预测

中图分类号: S682.12; Q785 **文献标志码:** A

0 引言

百合无症病毒(LSV)属于香石竹潜隐病毒属(*Carlavirus*), 是感染百合科植物的主要病毒之一. 百合无症病毒感染百合后植株无明显症状, 但导致种球退化, 给百合种植业造成极大的经济损失^[1,2]. 对于该病毒目前还没有十分有效的方法来加以防治, 关于百合病毒的研究还集中于病毒检测和利用传统的方法进行病毒脱除^[3,4], 因此培育抗病毒百合已经迫在眉睫.

近年来, RNA 干扰(RNAi)介导的病毒抗性逐渐发展为一种新型的基因工程抗病毒策略, 并且在一些植物抗病性育种上获得成功^[5,6]. 利用病毒衣壳蛋白(CP)基因培育抗病性品种的研究报道很少^[7], 在百合抗病性分子育种上的研究还未见报端. 利用衣壳蛋白基因的 dsRNA 来抗 LSV, 首先要对百合无症病毒衣壳蛋白(LSV CP)基因的遗传性和序列进行分析, 百合衣壳蛋白基因克隆和蛋白分析是构建抗 LSV 表达载体、转化百合、培育抗病毒新品种的前提.

本研究在利用 RT-PCR 获得的 LSV 衣壳蛋白基因组序列的基础上, 运用计算机技术和分子

生物学软件对 LSV 衣壳蛋白核苷酸序列的同源性、二级结构和其功能进行分析和预测, 明确 LSV 不同分离物之间存在的分子差异, 以期为深入研究 LSV CP 的结构, 在细胞内的复制、转运、表达机制和功能定位奠定基础, 为彻底根治百合病毒病提供理论参考.

1 材料与方法

1.1 材料来源与试剂

感染 LSV 的百合材料取自大连. 大肠杆菌 DH5 α 为本实验室所保存. 克隆载体 pMD18-T vector、RNA 提取试剂 TRIzol、RNA 反转录试剂、DNA 聚合酶以及引物等均购自大连 TaKaRa 公司.

1.2 CP 基因 RT-PCR 扩增及克隆

根据 GeneBank 已报道的 LSV 序列设计 (GeneBank 登录号为 NC 005138) CP 基因全序列引物, 其引物序列如下: 上游引物 LSV CP1, 5'-ATGCAATCAAGACCAGCACA-3'; 下游引物 LSV CP2, 5'-TCATCCATTATTTGCGTA-3'. 采用 TRIzol 法提取感病百合叶片总 RNA, 溶解于经 DEPC 处理过的灭菌超纯水中, 在 10 μ L 反

收稿日期: 2010-11-20; 修回日期: 2012-05-13.

基金项目: 国家自然科学基金资助项目(30771760, 31070621).

作者简介: 徐品三* (1958-), 男, 副教授, E-mail: xupinsan@sina.com.

应体系中加入 1 μg 总 RNA 作模板, 50 pmol LSV CP1 为引物, 在 AMV-逆转录酶作用下合成 cDNA 第 1 链, 其热循环程序为室温放置 10 min, 42 $^{\circ}\text{C}$ 60 min 后, 冰上放置 2 min, 反转录合成 cDNA。取 5 μL cDNA 产物, 补加引物 LSV CP1 后进行 PCR 扩增, 扩增条件: 94 $^{\circ}\text{C}$ 变性 1 min, 55 $^{\circ}\text{C}$ 复性 30 s, 72 $^{\circ}\text{C}$ 延伸 1 min, 30 个循环, 72 $^{\circ}\text{C}$ 后延伸 10 min, PCR 产物用 1.2% 琼脂糖凝胶进行电泳, 利用试剂盒将目标条带回收, 与克隆载体 pMD18-T vector 连接后转入宿主菌株 *E. coli*, 经过质粒抽提和 PCR 检测, 获得重组阳性克隆后, 送大连 TaKaRa 公司进行测序。

1.3 CP 基因核苷酸序列分析及蛋白质翻译

CP 基因测序后, 应用 DNAMAN 软件进行核苷酸序列分析和蛋白质翻译, 通过蛋白分析系统 ExPASy 预测 LSV 衣壳蛋白的理化性质。采用 ClustalX 1.81 软件与已报道的侵染百合 LSV 分离物进行病毒核苷酸序列、氨基酸序列同源性分析, 用 MEGA3.1 软件进行系统进化树绘制。

1.4 CP 基因功能及编码蛋白二级结构预测

利用 rpsblast 和 Kyte-Doolittle 分别对 CP 基因保守功能域和 LSV CP 的亲水性进行分析; 同时利用 COILS 算法和 ScanProsite 软件分别对 LSV CP 卷曲螺旋及 LSV CP 功能位点进行预测; 最后用 GOR 等蛋白分析软件进行 LSV CP 的二级结构预测分析。

2 结果讨论与分析

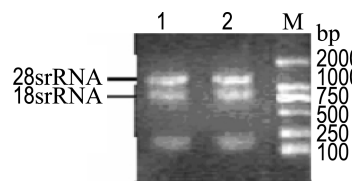
2.1 CP 基因 RT-PCR 扩增及克隆

用 TRIzol 法提取感病百合叶片总 RNA 作为模板(图 1), LSV CP1、LSV CP2 为特异引物, 通过 RT-PCR 扩增出大小为 900 bp 左右的目的基因片段(图 2), 与预期目的片段大小相符。

2.2 CP 基因核苷酸序列分析及蛋白质翻译

回收上述 RT-PCR 产物, 与克隆载体 pMD18-T vector 连接, 并转入宿主菌 *E. coli*, 获得重组阳性克隆后送去测序, 测序结果(GeneBank 登录号为 FJ 456352)表明 LSV CP 基因由 876 个核苷酸组成。利用 DNAMAN 软件进行核苷酸序列分析和蛋白质翻译(图 3), 结果表明, 克隆片段覆盖了 LSV CP 基因的全长序列, 含有一个开放阅读

框架, 编码 291 个氨基酸, 蛋白相对分子质量为 32 kD, 等电点 pI 为 7.02; 含有 3 个半胱氨酸, 预测这 3 个半胱氨酸之间形成二硫键的可能性较小, 280 nm 处的摩尔消光系数为 21 555 $\text{L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$,

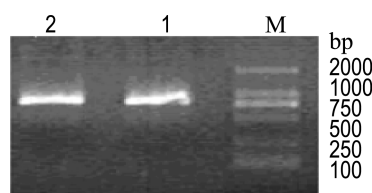


M DL 2000 DNA 标记物;

1、2 百合叶片 RNA

图 1 百合叶片 RNA 提取

Fig. 1 Extracted RNA from lily leave



M DL 2000 DNA 标记物;

1、2 RT-PCR 产物

图 2 琼脂糖电泳检测 RT-PCR 扩增产物

Fig. 2 The electrophoresis of RT-PCR product in agarose gel

```

1 atgcaatcaagaccagcacaagaatccggctccgcaagcgagacc
M Q S R P A Q E S G S A S E T
46 cctgcacgtggcaggcccccacaccagtgatgcaccaagagacgag
P A R G R P T P S D A P R D E
91 ccacccaattacaataataacgctgagtcacgtttagagcaacga
P T N Y N N N A E S L L E Q R
136 ctaccccgctgacgagagagtcacgtgcaaaagcgcaattcc
L T R L I E K L N A E K R N S
181 aatctgcgaatgtggtcttcgagatcgaggagccctcgctggaa
N L R N V A F E I G R P S L E
226 ccacagagtgccatgcgaagggaacccctgcgaacccctacgggaga
P T S A M R R N P A N P Y G R
271 ttctcaattgacgaactcttcaagatgaaggttggtgctgtatct
F S I D E L F K M K V G V V S
316 aacacatggcaacccactgaacaaatggccaaagatagcgctcagac
N N M A T T E Q M A K I A S D
361 atgcaggcgcttgggtaccacacggagcagctcgcatcagtaata
I A G L G V P T E H V A S V I
406 ttgcgaatggctcatcatgtgtgcttgctgagcagttcgccgttc
L Q M V I M C A C V S S A F
451 cttgacccctgaggcgagcatttgagtttgagaatggagcggtggcc
L D P E G S I E F E N G A V P
496 gtgattccatcactgcgacatgaagaagcagctggagctgccc
V D S I T A I M K K H A G L R
541 aaagtctgcggctctatgcccctatgctggaacagcagctgtt
K V C R L Y A P I V W N S M L
586 gtgcgaatcaaccaccagctgattggcaagctatggcttccaa
V R N Q P P A D W Q A M G F Q
631 tacaacacgcggttcgcagcttttgacaccttcgactacgtgacc
Y N T R F A A F D T F D Y V T
676 aaccaggcggtatcccaacctgtcgaggggacatcagggacccc
N Q A A I Q P V E G I I R R P
721 acttctgctgaggtcattgcccacacgcgcacaaagcaattggcc
T S A E V I A H N A H K Q L A
766 ctggataggtcaaacgcgaatgagcgtctgaggagcctggaaacc
L D R S N R N E R L G S L E T
811 gaataacacgggggggactaggagctgaattcgctcgaatcat
E Y T G G V L G A E I V R N H
856 cgatacgcaataatggatga 876
R Y A N N G *

```

图 3 LSV CP 基因核苷酸序列及推测编码氨基酸序列

Fig. 3 Nucleotide sequence of LSV CP gene and deduced amino acid sequence

0.1%(1 g/L)的吸光度为 0.673;假设二硫键全部打开时,280 nm 处的摩尔消光系数为 $21\,430\text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$,0.1%(1 g/L)的吸光度为 0.669,在哺乳动物网状红细胞体外表达的半衰期为 30 h,在酵母和大肠杆菌中体内表达的半衰期分别大于 20 h 和 10 h. 在溶液中的不稳定指数为 46.27,高于阈值 40,在溶液中性质较不稳定,脂肪族指数 77.49,平均疏水值 -0.432 ,蛋白质总体疏水性较高.

2.3 CP 基因核苷酸、氨基酸序列同源性分析

利用 ClustalX 1.81 软件分别对所获得的 CP 基因序列与 10 个来源于不同地区百合的 LSV 分离物对应序列进行核苷酸与氨基酸的亲缘性比较,结果显示,LSV 大连卷丹分离物与其他 10 个不同地区 LSV 分离物之间均具有较高的核苷酸和氨基酸亲缘性,分别为 93.4%~99.0% 和 84.8%~99.5%,而与日本分离物核苷酸亲缘性有较大差异,为 87.3%(图 4),但两种分离物的氨基酸同源性却很高,为 99.5%(图 5).

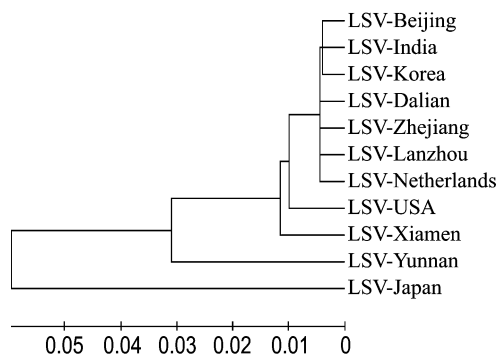


图 4 根据 LSV CP 基因核苷酸序列构建的亲缘关系树

Fig. 4 Phylogenetic tree based on the nucleotide sequences of LSV CP gene

2.4 CP 基因编码蛋白功能预测

疏水性分析表明平均疏水值(H)为 -0.432 ,该蛋白在 120~150 aa 有一个较强的疏水区,这可能与核酸结合和病毒粒子组装有关(图 6). 卷曲螺旋预测表明,如果卷曲度值(C)超过 0.5 就被认为存在卷曲螺旋结构,LSV 衣壳蛋白卷曲度值为 1.0,因此认为存在卷曲螺旋结构(图 7). 通过 rpsblast 分析其保守功能域发现在 45~96 aa 有 *Carlaviruses* 完整的衣壳蛋白保守结构域,在

105~244 aa 有 *Potexviruses* 和 *Carlaviruses* 完整的衣壳蛋白保守结构域(图 8).

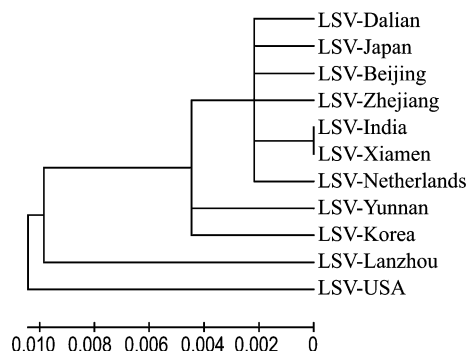


图 5 根据 LSV CP 基因氨基酸序列构建的亲缘关系树

Fig. 5 Phylogenetic tree based on the amino acid sequences of LSV CP gene

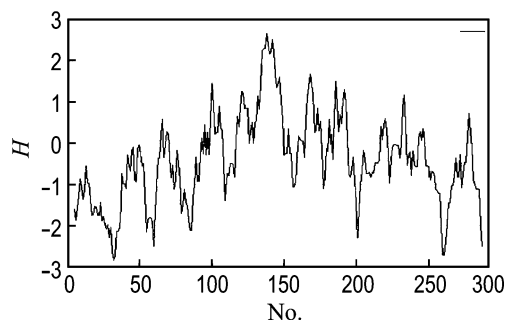


图 6 LSV CP 基因氨基酸序列的疏水性分析

Fig. 6 Hydrophobicity analysis of amino acid sequence encoded by LSV CP gene

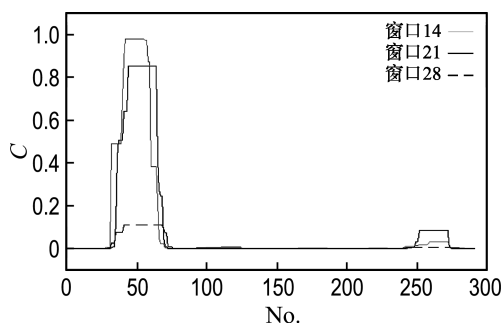


图 7 LSV CP 基因的氨基酸序列卷曲螺旋预测

Fig. 7 Coiled coil analysis of amino acid sequence encoded by LSV CP gene

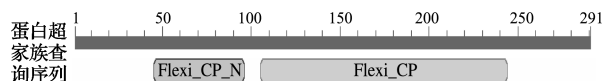


图 8 LSV 衣壳蛋白的保守功能域预测

Fig. 8 Prediction of conserved function domain of LSV CP

利用 ScanProsite 软件预测 LSV 衣壳蛋白的功能位点表明:有 1 个潜在的 cAMP 和 cGMP 依赖蛋白激酶磷酸化位点 238~241, PKA 调控代谢酶活性, PKG 主要调节胞内钙离子, 可能与 RNA 转移有关; 5 个潜在的酪蛋白激酶 II (CK2) 磷酸化位点: 11~14 aa, 22~25 aa, 40~43 aa, 92~95 aa, 241~244 aa, 酪蛋白激酶 II 是 Ser/thr 类蛋白激酶, 其活性不受环核苷和钙离子的调节, 能将靶蛋白的 Ser/thr 氨基酸磷酸化, 实现信号的级联放大和传递; 4 个潜在的 N-肉豆蔻酰位点: 102~107 aa, 266~271 aa, 274~279 aa, 275~280 aa, 豆蔻酸通过其脂肪链与一些水溶性蛋白共价结合, 使这些水溶性蛋白被锚定在质膜的细胞质面, 成为膜蛋白; 1 个潜在的蛋白激酶 C (PKC) 磷酸化位点: 259~261 aa, 蛋白激酶 C 是一种细胞质酶, PKC 将靶蛋白的 Ser/thr 氨基酸磷酸化, 激活细胞质中某些酶参与生化反应的调控; 1 个潜在在原核细胞膜脂蛋白质的附着点: 134~144 aa (图 9)。

cAMP- and cGMP-dependent protein kinase phosphorylation site
Site : 238 to 241 RRPT Identity.
Protein kinase C phosphorylation site
Site : 259 to 261 SNR Identity.
Casein kinase II phosphorylation site
Site : 11 to 14 SASE. Identity.
Site : 22 to 25 TPSD. Identity.
Site : 40 to 43 SLLE. Identity.
Site : 92 to 95 SIDE. Identity.
Site : 241 to 244 TSAE. Identity.
N-myristoylation site
Site : 102 to 107 GVVSNN. Identity.
Site : 266 to 271 GSLETE. Identity.
Site : 274 to 279 GGVQGA. Identity.
Site : 275 to 280 GVQGA. Identity.

图 9 LSV 衣壳蛋白的功能位点预测

Fig. 9 Prediction of functional site of LSV CP

2.5 CP 基因编码蛋白二级结构预测

预测结果显示, 组成 LSV 衣壳蛋白的 291 个氨基酸中, 编码蛋白含有 α -螺旋、 β -折叠、卷曲等二级结构, 其中, 101 个氨基酸可能形成 α -螺旋结构, 含量为 34.71%; 42 个氨基酸可能形成 β -折叠片, 含量为 14.43%; 148 个氨基酸可能形成无规则卷曲, 占 50.86%。从上述预测结果可以看出, 在第 40~60 aa、93~102 aa、110~120 aa、165~180 aa、243~255 aa 处可能存在 α -螺旋结构; 在第 180~195 aa、230~240 aa 和 279~285 aa 处可能存在 β -折叠片结构 (图 10)。

GOR4:

Alpha helix	(Hh) :	101 is	34.71%
3 ₁₀ helix	(Gg) :	0 is	0.00%
Pi helix	(Ii) :	0 is	0.00%
Beta bridge	(Bb) :	0 is	0.00%
Extended strand	(Ee) :	42 is	14.43%
Beta turn	(Tt) :	0 is	0.00%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	148 is	50.86%
Ambiguous states	(?) :	0 is	0.00%
Other states	:	0 is	0.00%

helix	10	20	30	40	50
sheet	MQSRPAQESCSA	SETFARCRFT	PSDAPRDEPT	NYNNNAESLLE	QRLTLRLI
turns	T	TTTT	T	T	TTTT
coil	CCCCCCCCCCCC	CCCC	CCC	CCCC	
helix	60	70	80	90	100
sheet	HCLNAEKHNENLRN	VAFETGRPSLEPT	SAMRRNFANPYGRF	SIDELFKMK	
turns	HHHHHHHH	EEEE	EE	EEEE	HHHHHHHH
coil	T	CCCC	T	T	TT
helix	110	120	130	140	150
sheet	VCVVENNMTTEGMCKI	ASDIAGLGVPTEHVA	SVILQMWIMC	ACVSSAF	
turns	HH	HHHHHHHHHH	HHHHHHHHHH	H	
coil	CCCC	CCCC	CCCC	CCCC	TTTT
helix	160	170	180	190	200
sheet	LDPEGSIEFENGAVFVD	SIAALMKKHVGLR	KVCRLYAFIV	WNSMLVRNQF	
turns	HHHH	HHHHHHHHHHHHHH	EEEEEEEEEEEE	EEEE	TTTT
coil	CCCC	CC	CCCC	CC	CC
helix	210	220	230	240	250
sheet	PADWQAMCFQYNTRFAA	FDTLDYVTNQAAI	QFVEGII	RRPTSIAEVIAHNA	
turns	H	H	H	EEEEEEEE	HHHHHHHH
coil	CC	TT	CCCC	CC	CC
helix	260	270	280	290	
sheet	HCGLALDRSNRNR	LGSLTEYTCGVQGA	EIVRNHRYANNG		
turns	E	EE	EEEEEEEE		
coil	TT	TT	TTTT	TTTTTT	

图 10 LSV 衣壳蛋白二级结构预测结果

Fig. 10 Secondary structure prediction result of LSV CP

3 结 论

通过对大连卷丹的 LSV CP 基因核苷酸序列和其对应的氨基酸的同源性分析可以看出, 不同地区百合无症分离株其衣壳蛋白基因的核苷酸存在差异, 相对应的氨基酸序列的同源性非常高; 而在对不同品种百合所分离的百合无症病毒的衣壳蛋白基因核苷酸与氨基酸的同源性对比中却发现, 其核苷酸序列的同源性较好, 而相对应氨基酸的同源性却较低 (另作报道), 从这个结果可以初步推断 LSV CP 基因序列的同源性与其寄主特异性关系较大, 但与地理起源关系较小。在对不同品种百合提取 RNA 时发现, 不同品种百合细胞内的蛋白质、多糖等物质有很大的差异。因此, 本文预测在 LSV 变异进化中不受地域差异的影响, 而出现了寄主适应性变异。

通过对百合无症病毒衣壳蛋白结构与功能的预测发现, 其在溶液中稳定性不高, 功能位点也比较少, 这可能是百合无症病毒无症的原因所在。另外, 研究发现百合无症病毒衣壳蛋白含有

*Carlaviruses*和*Potexviruses*完整的衣壳蛋白保守结构域,可以通过这些区域设计 siRNA,利用 RNAi 技术更广范围地进行百合抗病毒育种研究.

参考文献:

- [1] Asjes C J, Blom-Barnhoorn G J, van Schadewijk A R. Effect of seasonal detection of *lily symptomless virus* and *lily mottle virus* on aphid-borne virus spread in *Lilium* in the Netherlands [J]. *Acta Horticulturae*, 2002, **568**(1):201-207.
- [2] Singh A K, Mahinghara B K, Hallan V, *et al.* Recombination and phylogeographical analysis of *lily symptomless virus* [J]. *Virus Genes*, 2008, **36**(2): 421-427.
- [3] Nimi Y, HAN D S, Mori S, *et al.* Detection of cucumber mosaic virus, *lily symptomless virus* and *lily mottle virus* in *Lilium* species by RT-PCR technique [J]. *Scientia Hort*, 2003, **97**(1):57-63.
- [4] XU Pin-san, Niimi Y, Araki H. Production of virus-free bulblets from callus induced from scale culture of *Lilium longiflorum* 'Georgia' [J]. *Journal of the Japanese Society for Horticultural Science*, 2000, **69**(1):97-102.
- [5] Zamore P D, Tuschl T, Sharp P A, *et al.* RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21-to 23-nucleotide intervals [J]. *Cell*, 2000, **101**(1):25-33.
- [6] 陈银华,张俊红,欧阳波,等.花椰菜 ACC 氧化酶基因的克隆及其 RNAi 对内源基因表达的抑制作用[J]. 遗传学报, 2005, **32**(7):764-769.
CHEN Yin-hua, ZHANG Jun-hong, OUYANG Bo, *et al.* Cloning of ACC oxidase gene and inhibition of endogenous gene expression with RNAi in cauliflower [J]. *Acta Genetica Sinica*, 2005, **32**(7):764-769. (in Chinese)
- [7] 赵 丹,孔宝华,李 凡,等. 侵染百合的黄瓜花叶病毒亚组 II 分离物 cp 基因克隆和序列分析[J]. 植物病理学报, 2007, **37**(5):456-460.
ZHAO Dan, KONG Bao-hua, LI Fan, *et al.* Cloning and sequencing on cp gene of *Cucumber mosaic virus* subgroup II isolated from *Lily* (*Lilium* cv. Oriental hybrids) [J]. *Acta Phytopathologica Sinica*, 2007, **37**(5):456-460. (in Chinese)

LSV coat protein genes cloning and protein analyses

XU Pin-san*, LIU Ji-wen, GAO Xiao-rong, YIN Ya-lei

(School of Life Science and Technology, Dalian University of Technology, Dalian 116024, China)

Abstract: A pair of primers are designed and synthesized based on the oligosaccharides nucleotide sequences of coat protein (CP) genes of *lily symptomless virus* (LSV) reported, then, a fragment of 876 bp of LSV CP gene sequence is amplified by reverse transcription polymerase chain reaction (RT-PCR) from the total RNA infected lily leaves. The function and the secondary structure of LSV CP are predicted by the bioinformatics software system after analyzing the sequence. The results show that the gene is 876 bp in length coding 291 amino acids. This isolation has a homology of 93.4%-99.0% at nucleotide and 84.8%-99.5% at amino acid levels respectively compared with those published sequences of LSV according to GeneBank. The functional study prediction shows that the grand average of hydropathicity is -0.432. The protein has one coiled-coil domain, several phosphorylation sites and a conserved structure domain of *Carlaviruses* CP. The secondary structure is mainly composed of α -helix and random coil-based structures.

Key words: *lily symptomless virus*; coat protein; sequence analysis; functional prediction