



杨扇舟蛾颗粒体病毒晚期表达因子的分析*

梁振普** 赵晨光 张小霞*** 王彩平 吴慧 桑思瑶

(河南农业大学生命科学院, 郑州 450002)

摘要 【目的】本研究对杨扇舟蛾颗粒体病毒 (*Clostera anachoreta* granulovirus, ClanGV) 的晚期表达因子 (Late expression factors, LEFs) 进行生物信息学分析, 并对 ClanGV 的 *lefs* 与其它杆状病毒的 *lefs* 进行对比, 对鉴定 *lef* 基因和研究 ClanGV 的感染机制具有重要意义。【方法】使用 ORF finder 在线程序进行开放阅读框 (ORFs) 分析, BLASTP 程序用于蛋白序列的同源分析, DNASTar 和 DNAClub 生物学软件用于对序列进行处理和分析, 采用 BLAST 在公共数据库中对序列进行比对, Clustal W Version 1.8 和 GeneDoc 2.7 用于多序列的比对和分析, 用 MEGA 6.06 的 NJ 法进行系统发育分析。【结果】ClanGV 基因组包括 15 个 *lef* 基因, 分别为 *ie-1*、*lef-2*、*39k*、*lef-11*、*p35*、*p47*、*lef-1*、*lef-6*、*lef-5*、*lef-4*、*dnapol*、*lef-3*、*lef-9*、*lef-8* 和 *lef-10*。ClanGV 与云杉枞色卷蛾颗粒体病毒 (*Choristoneura occidentalis* granulovirus, CoGV)、黄地老虎颗粒体病毒 (*Agrotis segetum* granulovirus, AsGV)、苹果异形小卷蛾颗粒体病毒 (*Cryptophlebia leucotreta* granulovirus, CIGV)、马铃薯块茎蛾颗粒体病毒 (*Phthorimaea operculella* granulovirus, PoGV) 和苹果蠹蛾颗粒体病毒 (*Cydia pomonella* granulovirus, CpGV) 有着较近的亲缘关系。颗粒体病毒中只有 ClanGV 和分月扇舟蛾颗粒体病毒含有 *p35*。ClanGV 的 *lef-2* 的转录方向不同于其它 12 个颗粒体病毒。 γ 和 δ 杆状病毒分别只含有 7 个和 3 个 ClanGV 的 *lef* 同源基因。ClanGV 的 *lefs* 跟其它杆状病毒 *lefs* 的相似性较低, 其中 ClanGV 与其它颗粒体病毒间的相似性高于 ClanGV 与核型多角体病毒间的相似性。ClanGV 的 *lef-2*、*39k*、*p35*、*lef-5*、*dnapol* 和 *lef-9* 在启动子上游拥有 TATA box; *p47* 和 *lef-1* 在 TATA 下游 20~40 bp 处还有一个 CACT 基序; *lef-11*、*p35*、*lef-6*、*lef-5*、*lef-3*、*lef-9*、*lef-8* 和 *lef-10* 这 8 个基因有晚期启动子基序 TAAG; 而 *p35*、*lef-5*、*lef-9* 和 *lef-10* 上游既有 TATA 又有 TAAG 基序。【结论】结果为研究杆状病毒的基因功能和感染机制提供了数据基础。

关键词 杨扇舟蛾颗粒体病毒, *lefs*, 启动子, 复制, 转录

Analysis of the late expression factor genes in *Clostera anachoreta* granulovirus

LIANG Zhen-Pu** ZHAO Chen-Guang ZHANG Xiao-Xia***
WANG Cai-Ping WU Hui SANG Si-Yao

(College of Life Science, Henan Agricultural University, Zhengzhou 450002, China)

Abstract 【Objectives】To identify late expression factor (LEF) genes and investigate the ClanGV infection mechanism. 【Methods】We analyzed the ClanGV LEF family of genes of *Clostera anachoreta* granulovirus, including gene identification and organization, gene sequences, and structural characteristics at the DNA and protein levels, and compared the data obtained with those on other baculoviruses. Open reading frames (ORFs) were confirmed using the ORF finder online program. Homology searches were carried out using the BLASTP program. Sequence data assembly and analysis were performed using DNASTar and DNAClub software. Sequence comparison in public databases was done using the BLAST

*资助项目 Supported projects: 国家自然科学基金 (31570151 和 31272094); 河南省高校重点科研项目 (15A210039)

**第一作者 First author, E-mail: 1650950319@qq.com

***通讯作者 Corresponding author, E-mail: lzpzx@126.com

收稿日期 Received: 2016-01-21, 接受日期 Accepted: 2016-03-29

program. Multiple sequence analysis was performed using ClustalW Version 1.8 and GeneDoc 2.7. Phylogenetic analysis was carried out with MEGA 4.0 software using the neighbor joining (NJ) distance method (1 000 bootstrap replicates). **[Results]** Sequence analysis indicated that the *C. anastomosis* granulovirus ClanGV genome includes 15 LEF genes, annotated as *ie-1*, *lef-2*, *39k*, *lef-11*, *p35*, *p47*, *lef-1*, *lef-6*, *lef-5*, *lef-4*, *dnapol*, *lef-3*, *lef-9*, *lef-8* and *lef-10*, respectively. *C. anastomosis* granulovirus ClanGV shares higher homology with the *Choristoneura occidentalis* granulovirus, *Agrotis segetum* granulovirus, *Cryptophlebia leucotreta* granulovirus, *Phthorimaea operculella* granulovirus, and the *Cydia pomonella* granulovirus, than with other granuloviruses. With the exception of 9 nucleopolyhedroviruses, *P35* was only found in *C. anastomosis* granulovirus ClanGV and the *C. anastomosis* granulovirus. *C. anastomosis* granulovirus ClanGV *lef-2* differed from that of the other 12 granuloviruses in transcription orientation. The nucleopolyhedrovirus belonging to gammabaculovirus and deltabaculovirus include 7 and 3 ClanGV *lefs*, respectively. *C. anastomosis* granulovirus ClanGV *lefs* have very low homology with baculoviruses, although their homology with granulovirus *lefs* is higher than with those of nucleopolyhedroviruses. *C. anastomosis* granulovirus ClanGV *lef-2*, *39k*, *p35*, *lef-5*, *dnapol*, and *lef-9*, possess an early promoter motif (TATA). *p47* and *lef-1* have a CACT motif 20-40 bp downstream of TATA, and eight genes; *lef-11*, *p35*, *lef-6*, *lef-5*, *lef-3*, *lef-9*, *lef-8*, and *lef-10*, have a late promoter motif TAAG. However, only *p35*, *lef-5*, *lef-9*, and *lef-10*, possess both the TATA and TAAG. **[Conclusion]** The results of this study are of considerable importance to research on the ClanGV granulovirus gene function and infection mechanism.

Key words *Clostera anachoreta* granulovirus, *Lefs*, promoter, replication, transcription

杆状病毒是节肢动物的病原体,其宿主主要是鳞翅目、膜翅目和双翅目的昆虫。杆状病毒家族的成员有着双链、闭合、环状的 DNA 基因组,包括 4 个属: α 杆状病毒(鳞翅目昆虫的核型多角体病毒)、 β 杆状病毒(鳞翅目昆虫的颗粒体病毒)、 γ 杆状病毒(膜翅目昆虫的核型多角体病毒)和 δ 杆状病毒(双翅目昆虫的核型多角体病毒)(Liu *et al.*, 2014)。目前对于核型多角体病毒(Nucleopolyhedrovirus, NPVs)的研究已经非常深入和广泛,但是颗粒体病毒(Granulovirus, GV)研究相对较少(Liang *et al.*, 2013)。

在病毒感染期间杆状病毒基因的表达受到严格的调控,其感染过程可分为 3 个明显的阶段:早期、晚期和极晚期。早期基因通过宿主的 RNA 聚合酶 II 转录,而晚期和极晚期基因通过一个抗 α 鹅膏蕈碱的 RNA 聚合酶来转录(Berretta *et al.*, 2013)。早期阶段先于病毒 DNA 的复制,晚期和极晚期阶段表达病毒装配所需的基因,并且晚期的转录依赖于病毒 DNA 的复制(Wu *et al.*, 2010)。目前,关于杆状病毒感染过程中从早期到晚期阶段和从晚期到极晚期阶段转变的机制并不很清楚。

晚期表达因子(Late expression factors, LEFs)属于杆状病毒的早期基因,它们能够调

节晚期基因的表达,但自身却在早期阶段进行转录和表达(Hefferon *et al.*, 2004)。目前为止一共有 20 种 *lefs* 已经在苜蓿银纹夜蛾核型多角体病毒(*Autographa californica* multiple nucleopolyhedrovirus, AcMNPV)和家蚕核型多角体病毒(*Bombyx mori* nucleopolyhedrovirus, BmNPV)中被鉴定(Rapp *et al.*, 1998; Wu *et al.*, 2010)。*Lef-1*、*lef-2*、*lef-3*、*lef-7* 和 *lef-11* 既参与晚期基因的复制又参与晚期基因的转录(Evans *et al.*, 1997; Mikhailov *et al.*, 1998; Yu and Carstensen *et al.*, 2010; Zhang *et al.*, 2014),但 *lef-4*、*lef-5*、*lef-6*、*lef-8*、*lef-9*、*lef-10*、*lef-12*、*p47* 和 *39k* 只参与转录(Su *et al.*, 2011),其中 *lef-4*、*lef-8*、*lef-9* 和 *p47* 是病毒 RNA 聚合酶相关蛋白(Acharya and Gopinathan, 2002; Knebel-Morsdorf *et al.*, 2006; Rasheedi *et al.*, 2010)。*ie-1* 和 *ie-2* 能够作为反式激活因子参与复制(Olson *et al.*, 2002; Lin *et al.*, 2010)。*p35* 是抗细胞凋亡基因,它可以间接地保障病毒 DNA 的复制和晚期基因的表达(Zoog *et al.*, 1999; Sahdev *et al.*, 2010)。*Dnapol* 是 DNA 聚合酶基因(Wen *et al.*, 2007; Feng and Krell, 2014)。目前, NPVs 的 *lef* 基因已经进行了大量的鉴定与功能研究工作,但 GV 的 *lefs* 被研究

的很少 (樊磊等, 2012)。

杨树是一种重要的木材和绿化树种, 杨扇舟蛾 *Clostera anachoreta* 是杨树的一种重大害虫。从林间分离得到的杨扇舟蛾颗粒体病毒 (*Clostera anachoreta* granulovirus, ClanGV) 是一种高度特异和高效的杆状病毒病原体 (Liang *et al.*, 2013)。我们已经在多个地区用 ClanGV 对杨扇舟蛾进行了防治, 取得了较好的结果。杨扇舟蛾颗粒体病毒可以开发为一种具有重要应用价值的生物杀虫剂。杆状病毒一般需要 3~5 d 才能致死害虫。为了缩短致死时间, 目前大都通过分子生物学手段构建重组杆状病毒或添加生物增效剂来提高病毒杀虫剂的杀虫速度。*Lef* 基因在病毒感染复制过程中具有至关重要的作用, 本研究对 ClanGV 的 *lef* 家族进行了详尽的生物信息学分析, 并对 ClanGV 的 *lefs* 与其它杆状病毒的 *lefs* 进行了对比, 研究结果将为高效重组病毒的分子改良提供理论依据。

1 材料与方法

1.1 病毒基因组

本课题组已完成 ClanGV 基因组序列测定, 序列号为 HQ_116624 (Liang *et al.*, 2011)。

1.2 序列分析

开放阅读框 (ORFs) 通过 ORF finder 在线程序 (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) 来分析。超过 150 bp 的 ORFs 被作为假定的基因; 任意两个 ORFs 的重叠被设置为最多 25 个氨基酸, 否则选取较大的一个 (Wennmann *et al.*, 2015)。使用蛋白质 BLAST 程序进行了同源分析 (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)。

使用 DNASTar 和 DNAClub 生物学软件对序列进行处理和分析, 采用 BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) 在公共数据库中对序列进行比对。Clustal W Version 1.8 和 GeneDoc 2.7 用于多序列的比对和分析。用 MEGA 6.06 的 NJ 法 (1 000 bootstrap replicates) 进行了系统发育分析 (Zhu *et al.*, 2014)。

2 结果与分析

2.1 ClanGV 的晚期表达因子分析

序列分析发现, 在 ClanGV 基因组中有 15 个 *lef* 基因分别为 *ie-1*、*lef-1*、*lef-2*、*lef-3*、*lef-4*、*lef-5*、*lef-6*、*lef-8*、*lef-9*、*lef-10*、*lef-11*、*p35*、*p47*、*39k* 和 *dnapol*。将这些基因的转录方向、在基因组中的位置、序列一致性跟 12 个已完成测序的 GV 进行了比较, 结果表明几乎所有这些基因在 GV 基因组中位置和转录方向都是保守的。有两个 ClanGV 的 *lefs* 较为特殊。一个是 *p35*, 除了与 ClanGV 亲缘关系很近的分月扇舟蛾颗粒体病毒 (*Clostera anastomosis* granulovirus, CalGV) 也含 *p35* 外, 该基因仅在 ClanGV 中存在, 但 CoGV 中有一个 *p35/p49* 同源物, 至于该基因在 GV 中的功能还有待研究。另一个是 ClanGV 的 *lef-2*, 它具有跟其它 12 个 GV 不同的转录方向。

同源分析表明, *lefs* 在 ClanGV 和其它 GV 之间的相似性较低, 为 16%~73% (大多数小于 60%)。基因比对显示, 除了 *dnapol-lef-3* 和 *39k-lef-11* 外, 两个邻近的 *lef* 基因间的 ORFs 数在这些 GV 间都有不同程度的差异。从 *lefs* 在基因组中的位置和排列比较的结果看, ClanGV 在 *lefs* 的基因结构、转录方向和长度上与云杉枞色卷蛾颗粒体病毒 (*Choristoneura occidentalis* granulovirus, CoGV)、黄地老虎颗粒体病毒 (*Agrotis segetum* granulovirus, AsGV)、苹果异形小卷蛾颗粒体病毒 (*Cryptophlebia leucotreta* granulovirus, CIGV)、马铃薯块茎蛾颗粒体病毒 (*Phthorimaea operculella* granulovirus, PoGV) 和苹果蠹蛾颗粒体病毒 (*Cydia pomonella* granulovirus, CpGV) 有着更高的相似性 (图 1)。

除了上述提到的 GV 们, 本文还将 ClanGV 与 NPVs 进行了对比, 包括 α 杆状病毒的代表种: AcMNPV、BmNPV、棉铃虫核型多角体病毒 (*Helicoverpa armigera* nuclear polyhedrosis virus, HaNPV) 和草地贪夜蛾核型多角体病毒 (*Spodoptera frugiperda* multiple nucleopolyhedrovirus, SfMNPV)。结果表明, AcMNPV 和 BmNPV 基因组中不含

lef-11 基因, HaNPV 和 SfMNPV 不含 *p35* 基因。 δ 杆状病毒只含有库蚊核型多角体病毒 (*Culex nigripalpus* nucleopolyhedrovirus, CnNPV) 一个成员, 该病毒只含有 *lef-1*、*lef-8* 和 *dnapol* 3 种 *lef* 基因。已测序的 γ 杆状病毒包括冷杉锯角叶蜂核型多角体病毒 (*Neodiprion abietis* nucleopolyhedrovirus, NaNPV)、红头松叶蜂核型多角体病毒 (*Neodiprion lecontei* nucleopolyhedrovirus, NINPV) 和欧洲松锈锯角叶蜂核型多角体病毒 (*Neodiprion sertifer*

nucleopolyhedrovirus, NsNPV) 3 个成员, 该属病毒含有 7 个 *lef* 基因的同源物 (*p47*、*lef-1*、*lef-5*、*lef-4*、*dnapol*、*lef-9* 和 *lef-8*) , 并且除了 *lef-8* 外, 其余基因的转录方向都跟多角体蛋白基因一样。ClanGV 中的 15 个 *lef* 基因, 在 NPV 基因组中的排列顺序与在 ClanGV 及其它 GVs 中的顺序不一致。除 *lef-1* 外它们的转录方向也彼此不一致(表 1)。*lef* 基因排列及转录方向的差异是否决定了其在不同杆状病毒中会有不同的生物学功能, 还有待研究。

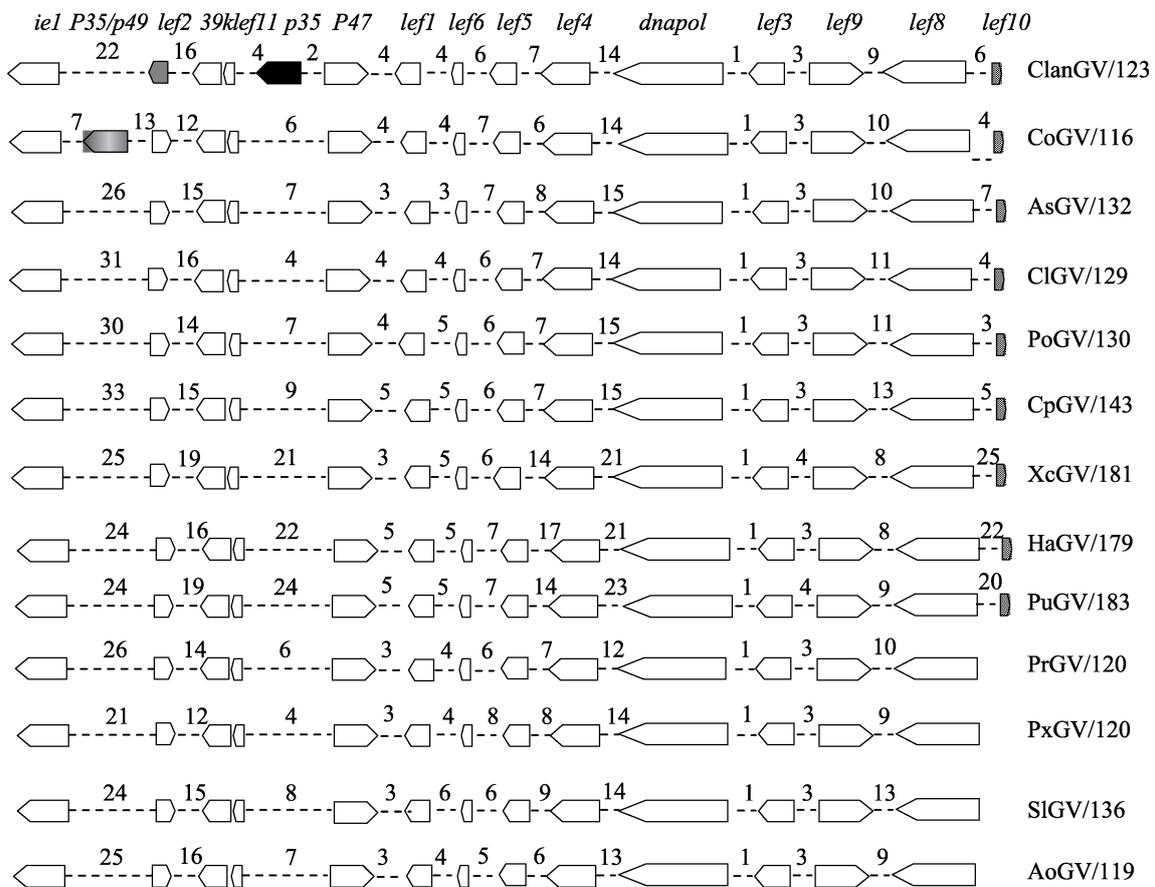


图 1 *lef* 基因在 ClanGV 和 12 个 GVs 基因组中的结构排列

Fig. 1 Schematic illustration of *lef* gene structure and orientation in the genome of ClanGV and other 12 sequenced GVs

箭头表示 ORFs 的相对长度和方向, 线上的阿拉伯数字表明两个基因间的 ORFs 数。GVs 名字后的阿拉伯数字表明相应

基因组的 ORFs 总数。用于比较基因组分析的序列 GenBank 检索号分别为 NC_008168 (CoGV)、NC_005839 (AsGV)、NC_005068 (CIGV)、NC_004062 (PoGV)、NC_002816 (CpGV)、NC_002331 (XcGV)、NC_010240 (HaGV)、NC_013772 (PuGV)、NC_013797 (PrGV)、NC_002593 (PxGV)、NC_009503 (SIGV) 和 NC_005038 (AoGV)。

The arrows represent the relative length and direction of the ORFs. The arabic numerals on lines indicate the ORF number between two genes, and those behind GV names show the total ORF number included in corresponding genome. GenBank accession numbers of genomes used in this comparison are NC_008168 (CoGV), NC_005839 (AsGV), NC_005068 (CIGV), NC_004062 (PoGV), NC_002816 (CpGV), NC_002331 (XcGV), NC_010240 (HaGV), NC_013772 (PuGV), NC_013797 (PrGV), NC_002593 (PxGV), NC_009503 (SIGV) and NC_005038 (AoGV), respectively.

表 1 ClanGV 的 *lef*s 与其它 NPVs 的 *lef*s 的比对*
Table 1 Comparison of ClanGV *lef*s with those of NPVs*

	Alpha					Gamma			Delta
	ClanGV	AcNPV	BmNPV	HaNPV	SfMNPV	NaNPV	NINPV	NsNPV	CnNPV
<i>ie-1</i>	471	582/10	584/10	661/11	683/8				
	6(-)	134 (+)	121 (+)	14 (+)	136 (-)				
<i>lef-2</i>	176	210/26	210/21	242/16	214/20				
	29(-)	153 (+)	135 (+)	117 (-)	16 (+)				
<i>39k</i>	262	275/12	277/12	312/13	310/10				
	46(-)	29 (-)	27 (-)	31 (-)	121 (+)				
<i>lef-11</i>	101			127/24	133/21				
	47(-)			32 (-)	120 (+)				
<i>p35</i>	307	299/39	299/39						
	52(-)	128 (+)	112 (+)						
<i>p47</i>	402	401/38	399/37	412/42	399/38	392/19	389/19	386/22	
	55(+)	33 (-)	31 (-)	35 (-)	116 (+)	49 (-)	46 (-)	49 (-)	
<i>lef-1</i>	233	266/27	270/27	245/35	239/29	211/27	211/28	190/23	235/21
	60(-)	7 (-)	6 (-)	124 (-)	18 (-)	68 (-)	65 (-)	68 (-)	45 (-)
<i>lef-6</i>	106	113/11	173/11	187/12	154/12	106			
	65(-)	21(+)	19 (+)	24 (-)	128 (+)	65 (-)			
<i>lef-5</i>	245	265/38	265/37	315/31	276/40	235/28	235/28	230/27	
	72(-)	92(+)	83 (+)	87 (+)	66 (-)	58 (-)	55 (-)	58 (-)	
<i>lef-4</i>	453	464/28	465/28	454/30	459/32	468/22	468/22	457/26	
	80(-)	83(+)	73 (+)	79 (+)	76 (-)	62 (-)	59 (-)	62 (-)	
<i>dnapol</i>	1037	984/29	986/29	1020/31	1041/31	919/21	923/20	913/20	1138/16
	95(-)	58(-)	53 (-)	67 (+)	92 (+)	12 (-)	20 (-)	28 (-)	91 (+)
<i>lef-3</i>	336	385/10	385/10	379/11	387/11				
	97(-)	60(-)	55 (-)	65 (+)	90 (+)				
<i>lef-9</i>	495	516/48	490/50	519/51	498/50	529/31	503/33	507/33	
	101(+)	55(+)	50 (+)	52 (+)	97 (-)	39 (-)	37 (-)	40 (-)	
<i>lef-8</i>	849	876/45	877/45	901/44	881/46	843/29	843/30	846/29	922/20
	111(-)	43(-)	39 (-)	38 (-)	113 (+)	83 (+)	78 (+)	81(+)	26 (-)
<i>lef-10</i>	69	78/23	78/21	71/24	75/27				
	118(+)	46(+)	42 (+)	46 (+)	106 (-)				

*比对所用到的 α 杆状病毒有 AcMNPV、BmNPV、HaNPV、SfMNPV, γ 杆状病毒有 NaNPV、NINPV、NsNPV, 杆状病毒有 CnNPV; 针对每个 *lef* 基因, 第一行前后两个阿拉伯数字分别表示氨基酸数以及与 ClanGV 的 *lef* 基因的相似性; 第二行阿拉伯数字表示在基因组中的 ORF 编号; “+”和“-”表示基因转录方向; 这个比对中用到的基因组的 GenBank 检索号分别为 NC_001623 (AcMNPV)、NC_001962 (BmNPV)、NC_003094 (HaNPV)、NC_009011 (SfMNPV)、NC_008252 (NaNPV)、NC_005906 (NINPV)、NC_005905 (NsNPV) 和 NC_003084 (CnNPV)。下表同。

* Alphabaculovirus used in this comparison are AcMNPV, BmNPV, HaNPV, SfMNPV; gammabaculovirus used in this comparison are NaNPV, NINPV, NsNPV; deltabaculovirus used in this comparison is CnNPV; for each gene, the arabic numerals in first lines indicate the gene length in amino acid sequence and its identity with that of ClanGV; the arabic numerals in second lines indicate the ORF number in genomes; “+”and“-” show the gene transcription direction; GenBank accession numbers of genomes used in this comparison are NC_001623 (AcMNPV), NC_001962(BmNPV), NC_003094 (HaNPV), NC_009011 (SfMNPV), NC_008252 (NaNPV), NC_005906(NINPV), NC_005905 (NsNPV) and NC_003084 (CnNPV), respectively. The same below.

2.2 启动子分析

杆状病毒早期基因可能拥有 TATA 启动子序列以及 TATA 下游 20~40 bp 处的 CAC/GT 元件。杆状病毒晚期基因的表达跟一个 ATAAG 基序有关, 这个基序是病毒编码的 RNA 聚合酶进行转录的起始位点。

对 ClanGV 的 *lef* 基因上游序列分析表明, *lef-2*、*39k*、*p35*、*lef-5*、*dnapol* 和 *lef-9* 有一个早期的启动子基序 (TATA), 它位于 *lef* 基因起始密码子上游 190 bp 以内。*P47* 和 *lef-1* 有一个位于 TATA 盒下游 20~40 bp 的 CACT 基序。*le-1*、*lef-11*、*lef-6*、*lef-4*、*lef-3* 和 *lef-8* 在 TATA 基序上游处有一个 CAC/GT 基序。*Lef-11*、*p35*、*lef-6*、*lef-5*、*lef-3*、*lef-9*、*lef-8* 和 *lef-10* 含有晚期启动子基序: TAAG。*le-1* 上游还有一个类似增强子的序列: CGTGC。只有 *p35*、*lef-5*、*lef-9* 和 *lef-10* 既含有 TATA 又含有 TAAG 基序 (表 2)。

2.3 与复制和转录相关的 ORFs

已有研究证明杆状病毒的 *lef-1* 的编码产物是一个 DNA 引发酶, 由于 *lef-2* 和 *lef-1* 之间存在相互作用, 因此它被称作引发酶相关因子。在杆状病毒感染过程中 *lef-1* 和 *lef-2* 至少与两个过程有关: DNA 复制和晚期/极晚期基因的表达。*lef-3* 编码与 DNA 复制有关的单链 DNA 结合蛋白 (SSB), *lef-3* 对 DNA 复制、出芽病毒生成和极晚期基因的表达来说是必须的。瞬时晚期表达实验表明, *lef-11* 在晚期启动子的高效转录和病

毒 DNA 的复制中是必须的。

ClanGV 的 *lef-1* 长 702 bp, 编码 233 个氨基酸, 预测分子量为 27.6 ku。在氨基酸水平上对 12 个 GV 和 40 个 NPVs 的 *lef-1* 进行比对发现, ClanGV 的 *lef-1* 与其它 GV 的 *lef-1* 相似性为 43%~73%, 高于 ClanGV 与 NPVs 的相似性 (21%~34%)。ClanGV 的 *lef-1* 的表达产物有一个引物酶相似基序 WVIDVD, 它存在于所有测过序的 GV 和部分 NPVs 中。

ClanGV 的 *lef-2* 长 531 bp, G+C 含量为 47.27%, 编码一个全长 176 个氨基酸的蛋白质, 预测分子量为 20.4 u。同源比对结果显示, ClanGV 的 *lef-2* 与 NPVs 的 *lef-2* 的相似性为 12%~21%, 与其它 GV 的 *lef-2* 的相似性为 26%~58%。值得注意的是, 在其 C 末端有一个长 37 个氨基酸的半胱氨酸富集基序。从整个 ORF 看, GV 的 *lef-2* 的长度都是相同的, 除了 PxGV 的 *lef-2* 有一个更长的 C 末端。

ClanGV 的 *lef-3* 长 1 011 bp, C+G 含量为 38.28%。这个基因在翻译的终止密码子下游 3 bp 处有一个典型的 PolyA 信号序列。尽管 *lef-3* 的编码产物对杆状病毒 DNA 的复制非常重要, 但在不同杆状病毒中它们的相似性却非常低。ClanGV 的 *lef-3* 与其它已完成测序的 12 个 GV 的 *lef-3* 的平均相似性为 26%。比对结果显示, 所有杆状病毒中 *lef-3* 的 N 末端的保守性更高。目前为止, 对 GV 的 *lef-3* 的研究还十分欠缺, 尤其是该基因的具体功能以及其 SSB 保守域的功能。

表 2 ClanGV 的 *lef* 基因的启动子
Table 2 Promoters of ClanGV *lef* genes

	<i>ie-1</i>	<i>lef-2</i>	<i>39k</i>	<i>lef-11</i>	<i>p35</i>	<i>p47</i>	<i>lef-1</i>	<i>lef-6</i>	<i>lef-5</i>	<i>lef-4</i>	<i>dnapol</i>	<i>lef-3</i>	<i>lef9</i>	<i>lef-8</i>	<i>lef-10</i>	
TATA	-	- 97	- 31	-	- 45	- 43	- 184	-	- 57	-	- 143	-	-	160	-	- 95
CACT	- 74	-	-	- 165	-	- 24	- 178	- 106	-	- 81	-	-	-	-	-	- 50
CAGT	-	-	-	- 134	-	-	-	- 119	-	-	-	- 17	-	-	142	-
TAAG	-	-	-	- 81	- 176	-	-	- 195	- 55/ - 7	-	-	-	-	-	- 26	- 17
CGTGC	- 139	-	-	-	-	-	-	-	-	-	-	-	- 87	-	-	-

- 表示没有对应的启动子。

- indicates there is no corresponding promoter.

lef-11 对病毒晚期基因的表达和病毒 DNA 的复制是必须的。氨基酸水平比对显示, ClanGV 的 *lef-11* 跟其它 GV 的 *lef-11* 的相似性为 40%~58%, 跟 NPVs 的 *lef-11* 的相似性为 14%~27%。ClanGV 的 *lef-11* 只有 306 bp, 编码一个长 101 个氨基酸, 分子量为 11.69 ku 的蛋白质, 其中第 9~26 位氨基酸在所有已知的杆状病毒的 *lef-11* 中高度保守, 但 ClanGV 的 *lef-11* 不含 AcMNPV 的 *lef-11* 中存在的锌指基序。

2.4 病毒 RNA 聚合酶 (转录) 相关 ORFs

杆状病毒代表种 AcMNPV 中, 核心的病毒 RNA 聚合酶由 4 种病毒蛋白组成: *lef-4*、*lef-8*、*lef-9* 和 *p47*。*lef-8* 和 *lef-9* 含有 β 和 β' RNA 聚合酶共有基序, 它们构成催化位点。*Lef-4* 是一个 mRNA 加帽酶, *p47* 的功能目前还不清楚。

Lef-4 是 AcMNPV 中鉴定的 4 个 RNA 聚合酶亚基中的一个, 在所有 GV 和绝大多数 NPVs 中有其同系物。ClanGV 的 *lef-4* 长 1 362 bp, 编码一个有 453 个氨基酸、分子量为 52.2 ku 的蛋白质。同源性分析表明 ClanGV 的 *lef-4* 跟其它杆状病毒的 *lef-4* 相似性较低, 其中跟 GV 的相似性为 39%~49%, 跟 NPVs 的相似性为 21%~30%。

通过 BlastP, 发现在 73 株杆状病毒基因组中存在 ClanGV 的 *lef-8* 的同源物。在氨基酸水平, ClanGV *lef-8* 跟其它 GV 的相似性为 56%~63%, 高于跟 NPVs 的相似性(19%~47%)。ClanGV 的 *lef-8* 基因全长 2 550 bp, 编码一个长 849 个氨基酸, 分子量为 98 ku 的蛋白质。在它的 C 端有一个含 13 个氨基酸的基序 (GLKICGIHGQKGV), 此基序在杆状病毒中高度保守, 并且跟原核生物 RNA 聚合酶的第二大亚基同源。

ClanGV 的 *lef-9* 基因全长 1 488 bp, 编码一个长 495 个氨基酸, 分子量为 56.42 ku 的蛋白质。同源比对结果显示, 与其它 *lefs* 相比, ClanGV 的 *lef-9* 在杆状病毒中有着更高的保守性, ClanGV *lef-9* 和其它 GV 的平均相似性为 59.25%, 和 NPVs 的平均相似性为 42.16%。ClanGV 的 *p47* 基因全长 1 209 bp, 编码一个分子量为 46.66 ku

的蛋白质。比对结果显示, 杆状病毒的 *p47* 的 N 末端大约 1/3 的氨基酸序列相似性很低, 具有较大的序列多样性。

2.5 与转录相关的其它 ORFs

对 AcMNPV 研究表明, *lef-5*、*Lef-6*、*39K*、*lef-10* 也与转录相关。已完成测序的 GV 和 NPVs 的 *lef-5* 有 3 个保守区域, 分别位于 67~156, 167~222 和 304~337 位氨基酸区域。BlastP 分析结果表明, 在 ClanGV *lef-5* 的 C 端有一个锌指区, 该锌指区能通过促进 RNA 聚合酶连续阅读进而调节 RNA 链的延伸过程。*lef-5* 通过与它自身的相互作用来影响晚期转录复合体的形成, 并且通过实验证明位于 C 末端的 71 个氨基酸跟该相互作用无关。ClanGV 的 *lef-6* 是一个比较短的基因, 整个 ORF 全长 321 bp, 编码一个 12.2 ku 的蛋白质, 在其终止密码子下游 55 bp 处有个 PolyA。Blastp 的结果表明, ClanGV 的 *lef-6* 与其它 GV 的相似性为 22%~41%, 与 NPVs 的 *lef-6* 的相似性很低, 仅为 5%~13%。尽管 GV 和 NPVs 的 *lef-6* 的相似性非常低, 但是在 N 端它们都有两个相似的保守区域, 并且所有杆状病毒的 *lef-6* 在 V³⁷, Y⁶⁵, G⁶⁸ 处都是保守的。

Blastp 结果显示, *39K* 存在于所有 GV 和 NPVs 中。ClanGV 的 *39K* 与其它 GV 的相似性较低, 只有 16%~43%, 尤其 C 末端比其余部位的差异性更大。从蛋白的大小看, 这些 39 K 彼此之间的长度也存在明显差异 (225~295 aa)。ClanGV 的 *39K* 基因全长 789 bp, 编码一个 30.46 ku 的蛋白质。ClanGV 的 *lef-10* 只有 210 bp, 编码一个 7.76 ku 的小分子量蛋白质。同源比对显示, ClanGV 的 *lef-10* 同其它 GV 的相似性为 30%~37%, 同 NPVs 的相似性为 19%~29%。另外, 该基因往往与相邻的 ORF 有大片段的重叠, 因此在进行杆状病毒基因组序列分析时需注意。尽管杆状病毒的 *lef-10* 间相似性非常低, 但是其含有 3 个保守区域, 分别位于 D⁹-V³¹, L³⁹-I⁴⁷ 和 M⁶⁰-S⁷¹ 处。

2.6 IE-1、P35 和 DNAPOL

Ie-1 是一个晚期基因表达所需的多功能调

控蛋白。杆状病毒的 *p35* 是一个凋亡蛋白酶 Caspase 的有效抑制剂。*dnapol* 能在螺旋失稳蛋白缺失时以单链 DNA 为模板聚合脱氧核苷酸, 并且具有 5'-3' 的核酸外切酶活性。瞬时复制实验中 AcMNPV 的 *dnapol* 被认为是质粒复制所需的 6 个基因中的一个。

ie-1 在绝大多数杆状病毒中都存在, 例如 AcMNPV、BmNPV 和 *Orgyia pseudotsugata* multiple nucleopolyhedrovirus (OpMNPV)。同源比对显示, ClanGV 的 *ie-1* 与其它 GV 的相似性为 20%~41%, 与 NPVs 的相似性非常低, 只有 5%~10%。ClanGV 的 *ie-1* 有 1 416 bp, 编码一个含 471 个氨基酸, 分子量为 54.77 ku 的蛋白质。跟其它 GV 与 NPVs 的 *ie-1* 进行对比时发现, ClanGV 的 *ie-1* 在 81 至 471 个氨基酸中有两个保守区, 但是 NPVs 的 *ie-1* 的 N 端存在的氨基酸在 GV 的 *ie-1* 中不超过 80 个。

目前为止, *p35* 只在 ClanGV 和 CalGV 跟 9 种 NPVs (AcMNPV、MvMNPV、PxMNPV、LsNPV、RoMNPV、SINPV、BmNPV、HcNPV 和 ToNPV) 中被发现。ClanGV 的 *p35* 编码一个有 307 个氨基酸, 分子量为 35 ku 的蛋白质。同源分析表明, 虽然 *p35* 在 NPVs 之间的相似性比较高 (58.4%~99.7%), 但 ClanGV 的 *p35* 与 NPVs 的 *p35* 的氨基酸序列相似性却非常低, 只有 12%~39%。已经知道的 *p35* 的裂解位点 D⁸⁷-G⁸⁸ 在以上所有杆状病毒中都是保守的。除此之外, ClanGV 的 *p35* 比 NPVs 的 *p35* 在 C 末端多出 16 个氨基酸, 核酸水平的分析显示, 这个 ORF 在翻译终止密码子下游 60 个核苷酸处有一个推测的 PolyA 基序, 因此, 需要进一步开展功能研究才能验证 ClanGV 的 *p35* 是否具有相关功能。

ClanGV 的 *dnapol* 长 3 114 bp, 明显比其它 *lef* 基因要长, 在终止密码子下游 98 bp 处有一个 PolyA 基序。同源性分析表明, ClanGV 的 *dnapol* 与其它 GV 的相似性为 44%~64%, 与 NPVs 的相似性为 20%~32%。通过 Blastp 分析, 在 ClanGV 的 *dnapol* 中发现了在真核和原核生物中共有的 DNA 聚合酶保守基序 (N 端-IV-II-VI-III-I-VII-V-C 端)。这表明它是 B 型 DNA 聚合酶家族的

一员, 另一方面, 真核和原核生物的 5'-3' 核酸外切酶区域也在 ClanGV 的 *dnapol* 中检测出来。同源比对结果表明, 所有 GV 的 *dnapol* 在 N 末端和 C 末端约 100 个氨基酸内的差异性比较明显。

3 讨论

对 *lef* 基因的序列、结构特点、组织方式进行了分析, 对 ClanGV 和其他杆状病毒的 *lefs* 进行了比较, 结果对鉴定新的 *lef* 基因和研究 ClanGV 的感染机制具有重要意义。

ClanGV 的 *lef-1* 的表达产物有一个引物酶相似基序 WVIDVD, 它存在于所有测过序的 GV 和部分 NPVs 中。有研究表明 AcMNPV 的 WVIDVD 基序能够引发瞬时 DNA 复制 (Evans *et al.*, 1997), 当这个基序突变为 WVVQAD 时, *lef-1* 就不再具有这个功能。颗粒体病毒的 *lef-1* 编码的 WVIDVD 基序是否如 NPVs 的 WVIDVD 基序一样对整个蛋白的功能有重要影响, 还需要通过基因功能研究才能得出结论。这也显示 NPVs 和 GV 经历了不同的进化历程后, 相应地拥有了不同的复制转录机制。

对 AcMNPV 研究表明, *lef-5* 对杆状病毒的晚期转录是必须的, 它可能是转录复合物的一个组件 (Hefferon, 2004; Su *et al.*, 2011)。 *lef-6* 对 AcMNPV 的 DNA 复制和晚期基因的转录都不是必须的, 但它的缺失会使晚期转录明显延迟, 因此推测 *lef-6* 可能执行着加速病毒感染周期的功能 (Lin and Blissard, 2002)。AcMNPV 的 39K 是一个晚期表达因子, 产物是一个核磷酸化蛋白, 存在于被感染细胞的细胞核中的病毒发生基质, 并且 AcMNPV 的 39K 具有 DNA 结合活性 (Broussard, 1996)。关于 *lef-10* 的功能目前还不清楚, 它跟其它功能基因没有明显的相似性, 推测其可能是转录过程中的一个辅助因子 (Hefferon, 2004)。

AcMNPV 的 *ie-1* 的 N 端 8~118 和 168~222 位氨基酸区域有启动子反式激活功能, C 端四分之三的氨基酸区域有 DNA 结合功能。跟其它 GV 与 NPVs 的 *ie-1* 进行对比时发现, ClanGV 的 *ie-1* 在 81 至 471 个氨基酸中有两个保守区,

但是 NPVs 的 *ie-1* 的 N 端存在的氨基酸在 GV 的 *ie-1* 中不超过 80 个。这些不同属的杆状病毒成员的 *ie-1* 在一级结构上的差异是否会导致其功能的改变目前还不清楚, 需要进一步开展 *ie-1* 尤其是 GV *ie-1* 的功能域的相关鉴定。

lef 基因对杆状病毒的复制至关重要, 并且它们彼此都具有不同的功能, 但是大多数杆状病毒只含有部分已鉴定的 20 个 *lef* 基因。例如 ClanGV 含有 15 种 *lef* 基因, 该病毒基因组没有 *lef-7*、*lef-12*、*p147* 和 *p39*。 γ 杆状病毒只含有 7 种 *lef* 基因 (*p47*、*lef-1*、*lef-5*、*lef-4*、*dnapol*、*lef-9* 和 *lef-8*)。研究这些 *lefs* 的功能和解析不同属杆状病毒的复制机制具有重要意义, 这将大大促进人们对杆状病毒的分子改良。例如, *lef-7* 与病毒的宿主范围有关, 研制相关重组病毒将可能扩大杆状病毒的宿主范围, 促进杆状病毒杀虫剂的应用。

尽管 ClanGV 的 *lefs* 与 GV 的 *lefs* 的相似性高于 ClanGV 与 NPVs 的相似性, 但 ClanGV *lefs* 与 GV *lefs* 间的相似性仍然较低 (大多数低于 60%)。因此我们推测, 在杆状病毒中普遍存在的 *lefs* 对杆状病毒的复制是必不可少的, 但舟蛾科的杨扇舟蛾与其它昆虫有较远的亲缘关系, 因此 ClanGV 在与宿主的共同进化过程中基因发生了较大的改变。除了 *dnapol-lef-3* 和 *39k-lef-1* 外, 在 12 个已测过序的 GV 及 ClanGV 中, 相邻的 *lef* 基因之间的 ORFs 数差别很大。需要进一步的实验来研究这些位于 *lefs* 间的基因与 *lefs* 功能之间的关系。

参考文献 (References)

- Acharya A, Gopinathan KP, 2002. Characterization of late gene expression factors *lef-9* and *lef-8* from *Bombyx mori* nucleopolyhedrovirus. *Journal of General Virology*, 83(Pt8): 2015–2023.
- Berretta MF, López MG, Taboga O, Sciocco-Cap A, Romanowski V, 2013. Functional analysis of *Spodoptera rugiperda* nucleopolyhedrovirus late expression factors in Sf9 cells. *Virus Genes*, 46(1): 152–161.
- Broussard DR, Guarino LA, Jarvis DL, 1996. Dynamic phosphorylation of *Autographa californica* nuclear polyhedrosis virus pp31. *Journal of Virology*, 70(10): 6767–6774.
- Evans JT, Leisy DJ, Rohrmann GF, 1997. Characterization of the interaction between the baculovirus replication factors LEF-1 and LEF-2. *Journal of Virology*, 71(4): 3114–3119.
- Feng G, Krell PJ, 2014. *Autographa californica* multiple nucleopolyhedrovirus DNA polymerase C terminus is required for nuclear localization and viral DNA replication. *Journal of Virology*, 88(18): 10918–10933.
- Hefferon KL, 2004. Baculovirus late expression factors. *Journal of Molecular Microbiology and Biotechnology*, 7(3): 89–101.
- Knebel-Morsdorf D, Quadt I, Li Y, Montier L, Guarino LA, 2006. Expression of baculovirus late and very late genes depends on LEF-4, a component of the viral RNA polymerase whose guanylttransferase function is essential. *Journal of Virology*, 80(8): 4168–4173.
- Liang ZP, Zhang XX, Yin XM, Cao SM, Xu F, 2011. Genomic sequencing and analysis of *Clostera anachoreta* granulovirus. *Archives of Virology*, 156(7): 1185–1198.
- Liang ZP, Zhang XX, Yin XM, Song XF, Shao XF, Wang LW, 2013. Comparative analysis of the genomes of *Clostera anastomosis* (L.) granulovirus and *Clostera anachoreta* granulovirus. *Archives of Virology*, 158(10): 2109–2114.
- Lin G, Blissard GW, 2002. Analysis of an *autographa californica* multicapsid nucleopolyhedrovirus *lef-6*-null virus: *lef-6* is not essential for viral replication but appears to accelerate late gene transcription. *Journal of Virology*, 76(11): 5503–5514.
- Lin XA, Chen Y, Yi YZ, Zhang ZF, 2010. Baculovirus immediately early 1, a mediator for homologous regions enhancer function in trans. *Virology Journal*, 7(1): 32.
- Liu XP, Yin FF, Zhu Z, Hou DH, Wang J, Zhang L, Wang ML, Wang HL, Hu ZH, Deng F, 2014. Genomic sequencing and analysis of *Sucra jujuba* nucleopolyhedrovirus. *PLoS ONE*, 9(10): e110023.
- Mikhailov VS, Mikhailova AL, Iwanaga M, Gomi S, Maeda S, 1998. *Bombyx mori* nucleopolyhedrovirus encodes a DNA-binding protein capable of destabilizing duplex DNA. *Journal of Virology*, 7(4): 3107–3116.
- Olson VA, Wetter JA, Friesen PD, 2002. Baculovirus transregulator IE1 requires a dimeric nuclear localization element for nuclear import and promoter activation. *Journal of Virology*, 76(18): 9505–9515.
- Pan Q, Hu Y, Li LL, 2012. Functional analysis of the late expression factor genes of *Plutella xylostella* granulovirus. *Chinese Journal of Virology*, (5): 560–566. [樊磊, 胡原, 黎路林, 2012. 小菜蛾颗粒体病毒晚期表达因子基因的功能分析. 病毒学报, (5):

- 560–566.]
- Rapp JC, Wilson JA, Miller LK, 1998. Nineteen baculovirus open reading frames, including lef-12, support late gene expression. *Journal of Virology*, 72(12): 10197–10206.
- Rasheedi S, Suragani M, Haq SK, Ghosh S, Ehtesham NZ, Hasnain SE, 2010. Characterization of LEF4 ligand binding property and its role as part of baculoviral transcription machinery. *Molecular and Cellular Biochemistry*, 333(1/2): 83–89.
- Sahdev S, Saini KS, Hasnain SE, 2010. Baculovirus P35 protein: an overview of its applications across multiple Therapeutic and biotechnological arenas. *Biotechnol Progress*, 26(2): 301–312.
- Su J, Lung O, Blissard GW, 2011. The *Autographa californica* multiple nucleopolyhedrovirus lef-5 gene is required for productive infection. *Virology*, 416(1/2): 54–64.
- Wen R, Ou R, Chen B, 2007. Identification, transcriptional and phylogenetic analysis of the DNA polymerase gene of *Pieris rapae* granulovirus. *Virus Genes*, 34(3): 351–358.
- Wennmann JT, Gueli Alletti G, Jehle JA, 2015. The genome sequence of *Agrotis segetum* nucleopolyhedrovirus B (AgseNPV-B) reveals a new baculovirus species within the *Agrotis baculovirus* complex. *Virus Genes*, 50(2): 260–276.
- Wu CP, Huang YJ, Wang JY, Wu YL, Lo HR, Wang JC, Chao YC, 2010. *Autographa californica* multiple nucleopolyhedrovirus lef-2 is a capsid protein required for amplification but not initiation of viral DNA replication. *Journal of Virology*, 84(10): 5015–5024.
- Yu M, Carstens EB, 2010. Identification of a domain of the baculovirus AcMNPV single-strand DNA binding protein LEF-3 essential for viral DNA replication. *Journal of Virology*, 84(12): 6153–6162.
- Zhang J, Dong ZQ, Zhang CD, He Q, Chen XM, Cao MY, Li HQ, Xiao WF, Lu C, Pan MH, 2014. Identification of a novel nuclear localization signal of baculovirus late expression factor 11. *Virus Research*, 184: 111–119.
- Zhu Z, Yin FF, Liu XP, Hou DH, Wang J, Zhang L, Arif B, Wang HL, Deng F, Hu ZH, 2014. Genome sequence and analysis of *Buzura suppressaria* nucleopolyhedrovirus: a group II Alphabaculovirus. *PLoS ONE*, 9(1): e86450.
- Zoog SJ, Bertin J, Friesen PD, 1999. Caspase inhibition by baculovirus p35 requires interaction between the reactive site loop and the β -sheet core. *The Journal of Biological Chemistry*, 274(37): 25995–26002.