View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Virology 311 (2003) 350–365

VIROLOGY

www.elsevier.com/locate/yviro

The complete sequence of the *Adoxophyes orana* granulovirus genome☆

Sally Wormleaton,[a] John Kuzio,[b] and Doreen Winstanley[a,]*

[a] *Horticulture Research International, Wellesbourne, Warwick CV35 9EF, UK*
[b] *University of Victoria, Department of Biochemistry and Microbiology, P.O. Box 3055 STN CSC Victoria, B.C., Canada V8W 3P6*

Received 19 November 2002; returned to author for revision 18 December 2002; accepted 3 February 2003

Abstract

The nucleotide sequence of the *Adoxophyes orana* granulovirus (AdorGV) DNA genome was determined and analysed. The genome contains 99,657 bp and has an A + T content of 65.5%. The analysis predicted 119 ORFs of 150 nucleotides or larger that showed minimal overlap. Of these putative genes, 104 (87%) were homologous to genes identified previously in other baculoviruses. The mean overall amino acid identity of AdorGV ORFs was highest with CpGV ORFs at 48%. Sixty-three ORFs were conserved among all lepidopteran baculoviruses and are considered to be common baculoviral genes. Several genes reported to have major roles in baculovirus biology were not found in the AdorGV genome. These included chitinase and cathepsin, which are involved in the liquefaction of the host, which explains why AdorGV-infected insects do not degrade in a typical manner. The AdorGV genome encoded two inhibitor of apoptosis (*iap*) genes *iap-3* and *iap-5*. Among all of the granuloviruses genomes there was a very high level of gene collinearity. The genes shared by AdorGV and CpGV had exactly the same order along the genome with the exception of one gene, *iap-3*. The AdorGV genome did not contain typical homologous region (*hr*) sequences. However, it contained nine repetitive regions in the genome.
© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Baculovirus; Granulovirus; *Adoxophyes orana*; Summer fruit tortrix; Genome sequence

Introduction

The *Baculoviridae* are a family of invertebrate viruses with large circular, covalently closed, double-stranded DNA genomes which are packaged into nucleocapsids. The nucleocapsids are enveloped and embedded in proteinaceous occlusion bodies (OBs). There are two genera of baculovirus, the nucleopolyhedrovirus (NPV) and the granulovirus (GV) (van Regenmortel et al., 2000). The NPVs have large polyhedral OBs with diameters of between 1 and 5 μm with single or multiple virions embedded (Federici, 1997). The GVs have small ovoid OBs which average about 150 × 300–600 nm and normally contain a single virion (Federici, 1997). As a group, NPVs have a wider host range than GVs, being isolated mainly from the order Lepidoptera but also from the Diptera, Hymenoptera, and others (van Regenmortel et al., 2000). To date, GVs have been isolated only from lepidopteran larvae (van Regenmortel et al., 2000). The NPVs and GVs differ in their cytopathology (Crook, 1991; Federici, 1997). Early in GV infection, the nucleus enlarges and "clearing" occurs, where the nucleoli and chromatin move to the periphery of the nucleus (Winstanley and Crook, 1993). As the nucleus enlarges, the membrane disintegrates and the contents of the nucleus and cytoplasm become mixed. In NPV infections the nuclear membrane does not break down until late in infection. Little is known about the molecular causes of these differences.

Ten lepidopteran NPVs have been completely sequenced, namely *Autographa californica* (Ac) MNPV (Ayres et al., 1994), *Orgyia pseudotsugata* (Op) MNPV (Ahrens et al., 1997), *Lymantria dispar* (Ld) MNPV (Kuzio et al., 1999), *Bombyx mori* (Bm) NPV (Gomi et al., 1999), *Spodoptera exigua* (Se) MNPV (IJkel et al., 1999), *Helicoverpa armigera* (Ha) SNPV (Chen et al., 2001), *Spodoptera litura* (Splt) MNPV (Pang et al., 2001), *Mamestra configurata* (Maco) NPV (Li et al., 2002), *Epiphyas postvittana* (Eppo) MNPV (Hyink et al., 2002), and *Helicoverpa*

☆ The GenBank Accession Number of the AdorGV genome sequence reported in this paper is AF547984.

* Corresponding author. Fax. +44-0-1789-470552.
E-mail address: doreen.winstanley@hri.ac.uk (D. Winstanley).

0042-6822/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0042-6822(03)00149-1

zea (Hz) SNPV (Chen et al., 2002). In contrast, only three GVs, *Xestia c-nigrum* (Xecn) GV (Hayakawa et al., 1999), *Plutella xylostella* (Plxy) GV (Hashimoto et al., 2000), and *Cydia pomonella* (Cp) GV (Luque et al., 2001), have been sequenced so far. The only nonlepidopteran baculovirus completely sequenced is the mosquito-infecting virus *Culex nigripalpus* (Cuni) NPV (Afonso et al., 2001).

Adoxophyes orana GV (AdorGV, previously AoGV) is a "slow" GV (Winstanley and O'Reilly, 1999), i.e. typically the host dies in the final instar regardless of when it was infected. AdorGV is pathogenic to the summer fruit tortrix moth, *A. orana,* a pest on apples and pears in most of Europe and Japan. Several strains of the virus have been isolated, including one (AdhoGV) from *Adoxophyes honmai,* or the smaller tea tortrix, which is a pest of tea trees and is particularly prevalent in Japan (Nishi and Nonaka, 1996). No differences in restriction endonuclease profiles have been found between any of these isolates. The ecdysteroid UDP-glucosyltransferase (*egt*) gene of AdhoGV is identical to the *egt* gene of AdorGV (Wormleaton and Winstanley, 2001; Nakai et al., 2002).

The genome of AdorGV has previously been mapped and was estimated to contain approximately 100.9 kbp, and a small number of AdorGV genes have been identified (Wormleaton and Winstanley, 2001). Here, we present the complete sequence and gene organization of the AdorGV genome and compare it to the sequence and genetic organization from other baculoviruses.

Results and discussion

Sequence analysis of the AdorGV genome

The AdorGV genome is composed of 99,657 bp. This compares to 123,500, 178,733, and 100,999 bp for CpGV, XecnGV, and PlxyGV, respectively (Luque et al., 2001; Hayakawa et al., 1999; Hashimoto et al., 2000). It is the smallest lepidopteran baculovirus sequenced to date with the highest A + T (AT) content of 65.5%. The adenine residue of the *granulin* start codon was designated nt 1 and the sequence numbered in the direction of *granulin* transcription. The initial criteria used for selecting ORFs for further study were that they should be methionine-initiated ORFs of at least 50 amino acids which have minimal overlap with other ORFs. However, once the comparisons with other baculoviruses were completed, one exception was allowed. This was Ador44, which is 48 amino acids long and has homologues in all sequenced baculoviruses to date.

Early in the analysis of the AdorGV ORFs it was noted that the coding strands of many of the genes with homologues in other baculoviruses had a purine (AG) bias (>50%) and we suspect this is related to the codon, third-position, GC bias (GC3s) (Sharp and Lloyd, 1993). This observation was confirmed in other annotated, AT-rich, viral genomes (data not shown). We used coding strand AG composition as one criterion in assessing the validity of potential AdorGV ORFs.

It has been pointed out that genomes which are AT- or GC-rich have a depressed coding potential: fewer codon types are available to be used to encode a given peptide (Wan and Wootton, 2000). In an organism with an AT-rich genome, AT-rich codon types are preferred and GC-rich codon types are eschewed. In GC-rich genomes this is inverted. A consequence of this is that the overall amino acid composition of the peptides in extreme genome composition organisms is skewed. Peptides of AT-rich organisms have a higher proportion of Phe, Leu, Ile, Met, Asn, Lys, and Tyr and those of GC-rich organisms have more Pro, Arg, Ala, Trp, and Gly. The end result of this is that organisms with an extreme genome composition encode peptides of lower complexity (reviewed in Mount, 2001), as measured by a global complexity value G1 (Wan and Wootton, 2000). We calculated the G1 values for AdorGV ORFs and compared them to their respective coding strand AG composition (Figs. 1A and C). The average G1 for AdorGV peptides was 0.644, with a median distribution of 0.665. AT-rich (ca. 65%) genes from a variety of cellular organisms had median G1 values in the range of 0.72 to 0.78 (Wan and Wootton, 2000). While direct comparisons are not possible, due to differences in assembling data sets, it appears that the average complexity of the AdorGV protein complement is significantly lower than expected. Granulin (G1, 0.790) and p6.9 (G1, 0.249) stood out because they anchored the extreme ends of the distribution of G1 scores. Whereas most AdorGV ORFs had an AT composition (average 65%) close to the average AT composition of the AdorGV genome (66%), granulin and p6.9 had an AT composition that was significantly lower at 52 and 47%, respectively (Fig. 1B). In general, no overt correlation was evident between coding complexity and nucleotide composition for the extreme G1 valued ORFs (Table 1) or for AdorGV ORFs in general (data not shown). We note that for some proteins, such as granulin, it might be impossible for the virus to maintain its preferred nucleotide composition and codon usage and still encode a particular peptide.

Other measures of ORF validation were tried including dinucleotide bias and codon bias (Karlin, 1998). A summary of these analyses is shown in Table 1. Correspondence analysis (Peden, 1997) also failed to cluster any AdorGV genes in a manner that could be related to biological function such as temporal expression or level of expression on any of the axes examined (data not shown).

Our experience with various ORF prediction methods, particularly for genomes that are compositionally extreme (AT- or GC-rich), suggests that no one method will accurately define all possible ORFs. The AdorGV granulin gene is a striking example of this. It does not have the same nucleotide composition, as measured by AT-content, coding strand AG content, or dinucleotide frequencies, as the vast majority of other ORFs in the AdorGV genome. The codon usage of granulin does not conform to the overall codon

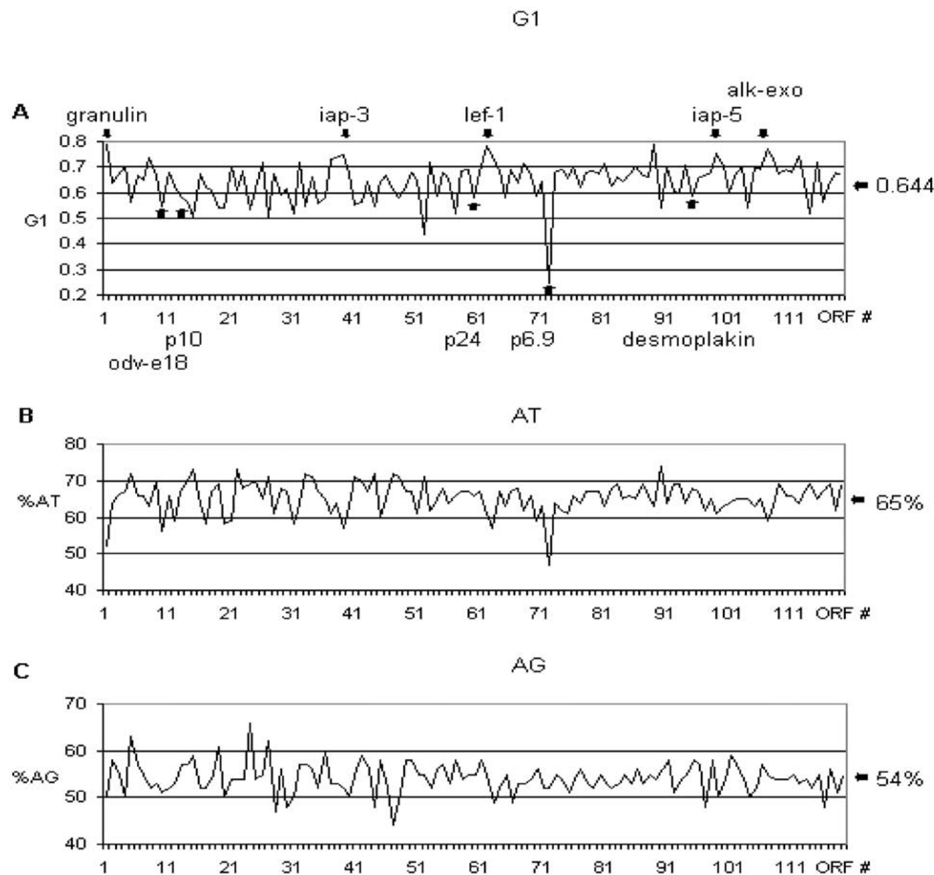

Fig. 1. Complexity of encoded peptides and nucleotide composition of AdorGV ORFs. (A) The G1 complexity values for the 119 AdorGV ORFs. The average G1 value for the selected ORFs is 0.644. The 5 highest and 5 lowest G1 values for known (homologous) AdorGV peptides are indicated with arrows. (B) The AT composition of AdorGV ORFs. The average ORF AT composition is 65%. Granulin (ORF 1, 52%) and p6.9 (ORF 72, 47%) significantly deviate from the average. (C) The AG composition of AdorGV coding strands. The purine content of the coding strand for the AdorGV ORFs is plotted. The average AG content is 54% and most of the annotated AdorGV ORFs are above 50%. The lowest AG content for an AdorGV ORF is ORF 47 at 44%.

usage of AdorGV ORFs. It had the appearance of an "alien gene" (Karlin, 1998), suggesting that it was recently acquired. However, we do not believe this is the case. We think it represents a relatively specific class of highly expressed, complex peptide that the virus encodes by sacrificing the constraints it maintains on other genes.

Twenty-five ORFs overlapped an adjacent ORF by an average of 13 codons per overlapping pair. There was only one significant overlap between ORFs (>45 codons) which was Ador29 and Ador30. Some overlapping ORFs, such as Ador104 and its opposite strand counterpart, were resolved through our screening process. A specific overlapping ORF encoding *lef*-10 was searched for and not found in AdorGV. The ORF that overlaps with the *lef*-10 ORF in NPVs, CpGV, and XecnGV (vp1054) is absent in AdorGV and PlxyGV. The AdorGV genome contained 9880 nt (9.9%) intergenic region with on average 107 nt between ORFs which did not overlap. One hundred and nineteen ORFs were selected for further study (Fig. 2, Table 2) and they were numbered from *granulin* in a clockwise direction.

The AdorGV ORFs had minimal intergenic distances and no preferred orientation (56.3% clockwise, 43.7% anticlockwise) or clustering according to expression or function. Predicted ORFs had a mean ORF length of 780 nucleotides.

Promoter analysis

Regions upstream of putative ORFs were screened for promoter elements. Two potential TATA-box element weight matrices were derived by Gibbs sampling (Ficket, 1996; Lawrence et al., 1993). One had a common TATA-box motif (derived core motif: TATAW) and the other was derived from TATA-like elements initially found upstream of some *lef* ORFs (derived core motif: TAATWAA). TATA-box motifs were found within 120 nt upstream of 73 ORFs and lef-TATA-like motifs were found upstream of 25 ORFs. Since 16 ORFs contained both types of TATA-box elements, we have identified a total of 82 potential AdorGV ORFs with early TATA-like transcription elements.

The CAKT element is an initiator element that is found at the transcription start site of many genes (Friesen, 1997). This element is approximately 30 nt downstream away from the TATA-box (RNA-polymerase II-binding site). A total

Table 1
A summary of the coding potential for the extreme G1 value encoded peptides of AdorGV.

ORF	Gene	aa	G1	AT	AG	diNuc	Bias	Axis 1	Axis 2	Axis 3	Axis 4
1	granulin	248	0.79	0.52	0.5	0.06	84.324	−0.11308	−0.02036	−0.12143	0.401842
89	iap-3	254	0.787	0.63	0.54	0.028	64.293	0.144089	−0.12912	0.03432	−0.02157
62	lef-1	233	0.781	0.61	0.55	0.027	54.597	−0.00051	0.12478	0.085414	0.29306
107	alk-exo	413	0.769	0.59	0.55	0.037	40.02	−0.06989	0.027183	−0.02688	0.225691
99	iap-5	279	0.755	0.61	0.5	0.028	58.328	0.085258	0.111869	−0.04124	0.070867
95	desmoplakin	539	0.589	0.68	0.58	0.025	54.544	0.042513	0.025207	0.171352	0.082995
13	p10	148	0.588	0.67	0.57	0.016	69.67	0.096699	−0.13917	−0.31609	0.051952
60	p24-capsid	156	0.583	0.66	0.55	0.018	74.333	−0.16811	0.255646	−0.11388	−0.07975
10	odv-e18	87	0.549	0.56	0.51	0.06	64.606	0.200902	0.251202	−0.10227	0.034373
72	p6.9	55	0.249	0.47	0.52	0.085	100.219	−0.222	−0.11694	−0.1541	−0.44388
ORF average		260	0.644	0.65	0.54	0.033					
Genome average				0.66							

The G1 values for AdorGV peptides were calculated and the five highest and five lowest, for peptides with known homologues in other baculoviruses, are shown. Short peptides generally produced lower G1 scores. The AT, coding strand AG, dinucleotide composition, and codon bias of AdorGV ORFs were calculated. The average AT composition of the ORFs reflects the overall AT composition of the AdorGV genome. No overt correlation was evident between coding complexity and nucleotide composition for the extreme G1 valued ORFs (as listed) or for AdorGV ORFs in general (data not shown).

of 13 AdorGV ORFs had a CAKT motif at 24–35 nt downstream of a TATA element. TATA elements operate independently of CAKT motifs; however, the presence of CAKT motifs, in a proper orientation, serves to confirm the likelihood of a true early promoter.

A non-TATA early motif has been reported for NPV (Bischoff and Slavicek, 1995). This motif, which has a CAKT-like core (consensus: C/G TCAGT C/T), was converted to a weight matrix and Gibbs sampled, and 120 nucleotides upstream of AdorGV ORFs were scanned. Significant scoring matches were found upstream of 15 genes. Three of these ORFs (p47, vp39, ORF 92) did not have a TATA-like element so the total number of putative early genes was elevated to 85.

A baculovirus late promoter element (DTAAG) is located, on average, at −60 nt upstream of the ORF initiation codon. This element was found within the nominal position of −15 to −120 nt upstream in 52 ORFs. In addition, 19 ORFs had this element immediately proximal to the initiation codon (within 15 nucleotides) and not elsewhere within 120 nt. A short leader sequence or initiation at an internal, downstream ATG would make such proximal DTAAG elements functional. It has been shown that the late promoter of the chitinase gene in AcMNPV is functional at 14 nucleotides upstream from the putative translation initiation codon (Hawtins et al., 1995). Therefore a total of 71 ORFs had potential late promoter elements.

In total, 103 ORFs were identified as having some type of early or late promoter element. Of these 103 ORFs, 53 had both early and late elements. Only 16 ORFs did not yield a match to the elements for which we searched. Of these 16 genes, a few (*iap-3, iap-5, fgf, fgf-2, lef-11, mp-nase, me53*) were homologues of previously characterised, expressed baculovirus proteins. In addition, AdorGV *ie-1,* a known immediate-early gene in other baculoviruses, had no recognisable TATA box in its immediate upstream region but did have a DTAAG late promoter element.

The AdorGV *ie-1* and *fgf-2* genes, which are poorly conserved and lack any type of early promoter element, may be nonfunctional in this virus. In considering this, we point out that the AcMNPV *ie-1* gene is the only baculovirus gene that is known to be spliced and, as such, is transcribed in AcMNPV from both its own early promoter and a distal, downstream element. It is possible that the AdorGV *ie-1* gene, if it is a functional early gene, is transcribed as a spliced gene only from a distal promoter element.

Given the minimum physical criteria we used to assess them (AG-content, G1), the allowable overlap of ORFs, and the fact that all AdorGV ORFs did not have some identifiable promoter element, it might be suggested that our standards were not sufficiently stringent to cull nonexpressed ORFs. However, we point out that the AcMNPV *lef*-10 ORF, which is half contained within a large expressed ORF on the opposite strand and half contained in the remaining intergenic region of two ORFs, was not identified in the original publication of the AcMNPV sequence (Ayres, 1994). We feel that baculoviruses have sufficient precedent to consider the possibility that overlapping ORFs are valid. This also supports our strategy that the most stringent criteria for culling ORFs in new genomes should not immediately be applied.

Comparison of AdorGV gene content with that of other baculoviruses

Fifteen ORFs are so far unique to AdorGV, which accounts for 6.2% of the genome. The remaining ORFs have homologues in other baculoviruses. Sixty-three of the ORFs were conserved among all 14 of the lepidopteran baculoviruses sequenced so far. Sixty-eight AdorGV ORFs had homologues in AcMNPV, while 101, 97, and 97 AdorGV ORFs had homologues in CpGV, XecnGV, and PlxyGV, respectively. Twenty-three of the ORFs are so far unique to GVs. Only one AdorGV ORF had a homologue in an NPV

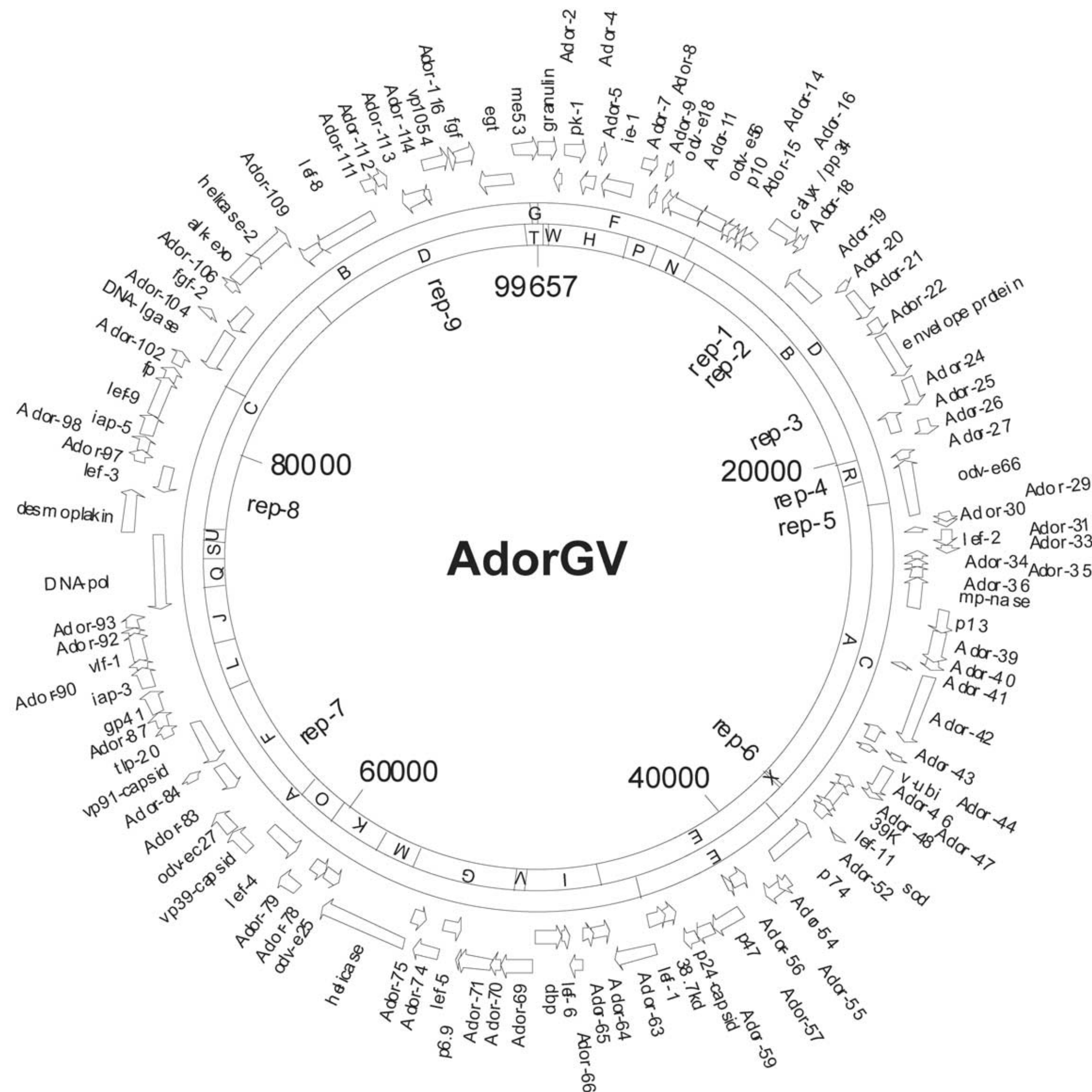

Fig. 2. Circular map of the AdorGV genome. The inner circles indicate the positions of cleavage sites for the following enzymes: inner circle, *Eco*RI; and outer circle, *Bam*HI. AdorGV ORFs are indicated outside these circles, with the arrow indicating the direction of transcription. The locations of repeat sequences are also shown (rep). A scale in bp is provided in the centre of the figure.

but not in a GV, which was Ador112 containing 20% aa identity to OpMNPV ORF 8.

The average amino acid sequence identity between AdorGV and CpGV, PlxyGV, XecnGV, and AcMNPV homologues is 48, 42, 41, and 34%, respectively. The most conserved ORFs are *polyhedrin/granulin* (Ador1), *ubiquitin* (Ador45), *p6.9* (Ador72), *lef-9* (Ador100), and *lef-8* (Ador110) (Table 2).

Genes specific to GVs

There are 23 ORFs present in all of the four sequenced GV genomes, which are absent in NPVs (Table 2). The most conserved GV-specific ORFs were Ador16 and Ador18, which are related to the previously described CpGV ORF 16L family (Kang et al., 1997). Ador99 encodes an IAP-5 that seems to be GV specific (Luque et al., 2001). Ador37 (homologous to Cp46, Xecn40, and Plxy35) is likely to be a member of the stromelysin family within the matrix metalloproteinase (MMP) superfamily. It is thought that this peptide is retained within infected cells until death, when it is released into the body of the insect, causing proteolysis of the tissues (Hashimoto et al., 2000; Hayakawa et al., 1999; Ko, et al., 2000).

AdorGV structural genes

The most conserved baculovirus structural protein is polyhedrin/granulin (89% average amino identity to GV homologues in Table 2), the major component of occlusion

Table 2
Potentially expressed ORFs in the genome of AdorGV

Description		Location			Size		Promoters					Composition			Identity						AcMNPV		Other	
ORF	Name	Left		Right	Len (aa)	Theor M_r	Early				Late	AT	AG (cs)	G1	CpGV		PxGV		XcGV		ORF	ID	ORF	OD
							T	L	I	N	np				ORF	ID	ORF	ID	ORF	ID				
1	granulin	1	>	744	248	29310	+				+	0.52	0.50	0.79	1	96	4	85	1	87	8	56		
2		747	<	1064	106	12059	+				+	0.64	0.58	0.64	2	34	5	28	2	31	—	—		
3	pk-1	1048	>	1872	275	32096					+	0.66	0.55	0.67	3	53	6	47	3	45	10	40		
4		1875	<	2447	191	22680	+		+			0.67	0.50	0.70	4	38	8	25	7	31	—	—		
5		2434	>	2682	83	9829						0.72	0.63	0.57	5	40	9	36	8	25	—	—		
6	ie-1	2757	<	4022	422	50055					+	0.66	0.57	0.67	7	44	10	30	9	30	147	19		
7		4111	>	4683	191	21662					+	0.66	0.55	0.65	8	40	11	36	10	32	146	27		
8		4753	<	5058	102	12203	+				+	0.63	0.52	0.74	9	50	12	49	11	55	145	24		
9		5069	>	5359	97	11730		+				0.70	0.53	0.67	—	—	—	—	—	—	—	—		
10	odv-e18	5419	<	5679	87	9210	+		+		++	0.56	0.51	0.55	14	83	13	61	12	72	143	42		
11		5682	>	7058	459	53897	+				++	0.66	0.52	0.68	15	50	14	41	13	40	142	33		
12	odv-e56	7048	<	8124	359	39442	+				+	0.59	0.53	0.63	18	65	16	63	15	62	148	44		
13	p10	8159	>	8602	148	16668	+	+			+	0.67	0.57	0.59	—	—	2	37	5	42	—	—	Sl p10	33
14		8470	<	8817	116	13382		+			+	0.70	0.57	0.56	—	—	3	27	—	—	—	—		
15		8846	>	9025	60	7157	+				+	0.73	0.59	0.51	19	22	17	28	16	17	29	22		
16		9066	<	9674	203	22763	+				+	0.64	0.52	0.67	20	54	20	45	17	50	—	—		
17	calyx pp34	9725	>	10732	336	36007					+	0.58	0.52	0.62	22	56	21	56	19	56	133	26*		
18		10835	>	11218	128	14515	+				+	0.67	0.55	0.61	23	56	23	47	18	49	—	—		
19		11597	<	13105	503	58549						0.69	0.61	0.54	—	—	—	—	—	—	—	—		
20		13326	>	13511	62	7180	+	+				0.58	0.50	0.54	—	—	—	—	—	—	—	—		
21		13751	>	14869	373	42814						0.59	0.54	0.70	29	25	24	23	25	21	—	—		
22		15158	>	15685	176	21064	+		+			0.73	0.54	0.61	30	28	25	30*	26	23	—	—		
23	F protein	15774	>	17495	574	67461	+	+				0.68	0.54	0.69	31	45	26	38	27	31	23	22		
24		17747	>	18724	326	38249	+		+			0.69	0.66	0.54	32	23	94	22	133	23	—	—		
25		18727	<	19578	284	33489						0.70	0.54	0.62	33	26	28	29*	29	23*	—	—		
26		19424	>	19984	187	21276					+	0.65	0.55	0.72	35	39	29	42	32	37	115	37		
27		20382	<	20783	134	16026	+				+	0.71	0.62	0.50	—	—	—	—	—	—	—	—		
28	odv-e66	20865	<	23105	747	83937	+				++	0.61	0.47	0.67	37	51	30	49	149	41	46	53		
29		23132	>	23443	104	12739	+				+	0.68	0.56	0.59	39	53	31	38	34	43	—	—		
30		23272	>	23610	113	13580						0.67	0.48	0.62	—	—	—	—	—	—	—	—		
31		23647	<	23820	58	7043	+				+	0.58	0.50	0.52	—	—	—	—	—	—	—	—		
32	lef-2	23834	>	24343	170	20422	+		+	+		0.64	0.57	0.72	41	50	32	40	35	41	6	26		
33		24392	>	24652	87	10263	+					0.72	0.57	0.55	42	35	—	—	36	32	—	—		
34		24645	<	24977	111	13515	+				+	0.71	0.56	0.66	—	—	—	—	—	—	—	—		
35		25020	<	25244	75	8657	+		+	+	+	0.67	0.52	0.56	—	—	—	—	—	—	—	—		
36		25283	<	25741	153	17956					+	0.65	0.60	0.58	45	36	34	30	39	24	—	—		
37	mp-nase	25824	<	27008	395	45699						0.61	0.53	0.73	46	34	35	29	40	34	—	—		
38	p13	26987	>	27784	266	31208					+	0.64	0.53	0.74	47	61	36	52	43	49	—	—	Mc76	48
39		27791	>	28915	375	42858	+				+	0.57	0.52	0.75	48	41	37	54	45	50	22	51		
40		28905	>	29318	138	16994						0.63	0.50	0.67	—	—	—	—	—	—	—	—		
41		29334	<	29561	76	8856					+	0.71	0.56	0.55	49	35	—	—	46	34	88	34		
42		29584	>	32208	875	210253					+	0.70	0.59	0.56	50	23	38	32	47	27	—	—		
43		32211	<	32816	202	23476	+	+			+	0.67	0.56	0.64	52	65	40	56	50	57	106/107	33/22		
44		32828	>	32971	48	5664	+			+		0.72	0.48	0.54	53	49	41	42	51	29	110	21		
45	v-ubi	33088	<	33369	94	10679	+		+			0.60	0.58	0.64	54	73	42	71	52	75	35	68		
46		33484	>	34530	349	40474	+				+	0.65	0.52	0.67	55	52	43	49	53	45	109	35		
47		34540	>	34764	75	9055	+				+	0.72	0.44	0.61	56	33	—	—	54	22	—	—		
48		34735	<	35091	119	14040		+			+	0.71	0.51	0.58	—	—	—	—	—	—	—	—		
49	39K	35120	<	35953	278	32158		+		+		0.67	0.58	0.63	57	36	45	25	55	25*	36	22		
50	lef-11	35943	<	36224	94	10802						0.67	0.58	0.68	58	49	46	43	56	46	37	22		
51	sod	36221	<	36676	152	16490					+	0.61	0.55	0.65	59	58	47	52	68	53	31	53		
52		36745	>	36912	56	6606		+		+		0.71	0.55	0.44	—	—	—	—	—	—	—	—		
53	p74	37370	<	39295	642	73504	+			+	++	0.62	0.52	0.72	60	56	49	50	77	45	138	43		
54		39340	>	39702	121	14053	+				+	0.65	0.56	0.59	—	—	—	—	—	—	—	—		
55		39686	>	40342	219	25084						0.68	0.57	0.68	—	—	—	—	—	—	—	—		
56		40678	<	41280	201	23224	+					0.64	0.53	0.66	—	—	—	—	—	—	—	—		
57		41338	<	41535	66	7559	+				+	0.66	0.58	0.52	62	32	50	24	83	31	—	—		
58	p47	41567	>	42730	388	45776				+	+	0.67	0.54	0.69	68	60	51	54	78	57	40	42		
59		42795	>	43469	225	26957	+				+	0.67	0.55	0.69	69	74	52	65	79	63	38	42		
60	p24 capsid	43505	>	43972	156	17580		+			+	0.66	0.55	0.58	71	52	53	53	80	44	129	31		
61	38.7kd	44004	<	44567	188	21491					+	0.67	0.58	0.68	73	25	54	29	81	27	13	18		
62	lef-1	44464	<	45162	233	28091	+	+				0.61	0.55	0.78	74	60	55	49	82	48	14	39		
63		45175	>	46773	533	60341					+	0.57	0.49	0.74	75	63	7	53	84	43	119	39		
64		46800	<	47492	231	27214	+					0.67	0.52	0.68	76	40	56	35	85	29	—	—	Ha110	25
65		47549	<	47917	123	14339						0.63	0.55	0.58	77	27	57	17	86	21*	—	—		
66		48052	>	48504	151	17966	+				+	0.67	0.49	0.69	79	41	59	37	87	36	150	25		
67	lef-6	48525	<	48830	102	12613	+	+			+	0.68	0.53	0.64	80	39	60	40	88	43	28	25		
68	dbp	48870	<	49877	336	38531	+					0.62	0.53	0.72	81	46	61	27	89	26	25	22		
69		49993	>	51180	396	47035	+	+			+	0.66	0.54	0.67	83	66	63	54	91	53	103	35		
70		51229	>	51582	118	13224					+	0.59	0.56	0.59	84	57	64	41	92	41	102	17		
71		51625	>	52812	396	45162	+				+	0.63	0.52	0.65	85	55	66	44	93	44	101	23		
72	p6.9	52845	>	53009	55	6974	+				+	0.47	0.52	0.25	86	67	67	65	94	73	100	56		
73	lef-5	53052	<	53792	247	28922	+					0.64	0.55	0.68	87	62	69	55	95	54	99	45		

(continued on next page)

Table 2 (*continued*)

Description		Location			Size		Promoters					Composition			Identity						AcMNPV		Other	
ORF	Name	Left		Right	Len (aa)	Theor M_r	Early				Late	AT	AG (cs)	G1	CpGV		PxGV		XcGV		ORF	ID	ORF	OD
							T	L	I	N	np				ORF	ID	ORF	ID	ORF	ID				
74		53739	>	54632	298	35490	+		+	+	+	0.62	0.54	0.69	88	61	70	51	96	54	98	41		
75		54633	<	55130	166	18555					+	0.61	0.51	0.66	89	50	71	37	97	45	96	30		
76	helicase	55114	>	58527	1138	134110	+	+	+	+	+	0.66	0.54	0.70	90	50	72	40	98	39	95	27		
77	odv-e25	58567	<	59217	217	24125	+				++	0.64	0.56	0.62	91	76	74	66	99	62	94	40		
78		59284	<	59775	164	19319	+	+	+		+	0.67	0.53	0.68	92	42	75	39	100	36	93	27		
79		59749	>	60549	267	31674						0.67	0.52	0.69	93	57	76	53	101	48	92	38		
80	lef-4	60775	<	62110	452	52468		+				0.67	0.55	0.67	95	51	78	44	110	45	90	32		
81	vp39capsid	62173	>	63045	291	33637				+	+	0.63	0.53	0.71	96	59	79	37	111	41	89	35		
82	odv-ec27	63099	>	63944	282	33309	+				+	0.68	0.52	0.63	97	62	80	44	112	48	144	27		
83		64471	<	65535	355	42103	+	+			+	0.69	0.53	0.66	99	31	82	35	113	30	—	—		
84		65591	>	65866	92	10893					+	0.65	0.53	0.65	100	31	83	32	116	23	—	—		
85	vp91capsid	65869	<	67641	591	67301	+				+	0.66	0.53	0.67	101	42	84	38	118	39	83	28		
86	tlp20	67613	>	67990	126	14354	+				+	0.65	0.56	0.70	102	40	85	29	119	19	82	22		
87		67968	>	68537	190	22069					+	0.69	0.53	0.67	103	67	86	59	120	54	81	44		
88	gp41	68554	>	69420	289	33570	+					0.66	0.55	0.66	104	49	87	46	121	45	80	31		
89	iap-3	69607	>	70368	254	29507						0.63	0.54	0.79	17	43	—	—	—	—	—	—	Op35	43
90		70416	>	70685	90	10590					+	0.74	0.56	0.54	105	38	88	35	122	32	78	21		
91	vlf-1	70672	>	71745	358	41920	+				+	0.64	0.58	0.70	106	65	89	51	123	53	77	32		
92		71753	>	72004	84	9628				+	+	0.69	0.51	0.61	107	58	91	48	125	46	76	28		
93		72022	>	72483	154	17835	+				+	0.69	0.54	0.60	108	57	92	41	126	29	75	25		
94	dnapol	72621	<	75713	1031	119695	+			+		0.64	0.55	0.71	111	59	93	50	132	51	65	34		
95	desmoplakin	75712	>	77328	539	63195					+	0.68	0.58	0.59	112	32	94	25	133	24	66	24		
96	lef-3	77526	<	78563	346	40497		+		+		0.67	0.57	0.65	113	43	95	29	134	24	67	29		
97		78532	>	78906	125	14726	+		+	+		0.62	0.48	0.67	114	58	96	50	135	44	68	31		
98		78963	>	79478	172	20129	+	+				0.65	0.58	0.68	115	27	97	28	136	27	—	—		
99	iap-5	79552	>	80388	279	32171						0.61	0.50	0.76	116	49	98	35	137	38	—	—		
100	lef-9	80366	>	81859	498	57662	+					0.63	0.54	0.70	117	69	99	61	139	62	62	53		
101	fp	81896	>	82336	147	17491	+			+		0.64	0.59	0.61	118	60	100	52	140	44	61	37		
102		82554	>	83084	177	20968	+			+		0.65	0.57	0.68	119	22	65	19	169	26	—	—		
103	dnaligase	83087	<	84727	547	65276	+					0.65	0.54	0.70	120	53	101	47	141	43	—	—	Ld22	24
104		84809	>	84997	63	6910	+	+	+			0.65	0.50	0.54	122	43	103	31	143	42	—	—		
105	fgf-2	84978	<	86045	356	41046						0.63	0.52	0.70	123	26	104	28	144	26	32	23		
106		86119	>	86439	107	12638	+				+	0.65	0.57	0.69	124	40	—	—	—	—	—	—		
107	alk-exo	86509	>	87747	413	47648	+	+			+	0.59	0.55	0.77	125	42	106	45	145	40	133	33		
108	helicase_2	87686	>	89170	495	57055	+			+		0.63	0.54	0.73	126	42	107	41	146	39	—	—	Ld50	43
109		89182	<	90024	281	32706	+	+				0.69	0.54	0.67	130	33	108	28	—	—	—	—		
110	lef-8	90020	<	92590	857	100556	+					0.66	0.54	0.69	131	68	109	66	148	64	50	51		
111		92684	>	93220	179	20867	+		+			0.66	0.55	0.68	134	54	112	38	171	40	53	22		
112		93225	>	93620	132	15616	+				+	0.64	0.53	0.75	—	—	—	—	—	—	—	—	Op8	20
113		93994	<	94899	302	35349	+	+			+	0.67	0.54	0.64	135	26	113	26	172	26	—	—		
114		94908	<	95102	65	7732		+			++	0.69	0.52	0.52	136	27	114	28	173	37	—	—		
115	vp1054	95174	>	96145	324	38242	+	+				0.65	0.55	0.72	138	53	115	42	175	42	54	36		
116		96209	>	96385	59	6752	+					0.67	0.48	0.56	—	—	—	—	—	—	—	—		
117	fgf	96385	>	97206	274	31990						0.69	0.56	0.64	140	32	117	25	178	25	32	25		
118	egt	97310	<	98644	445	50863	+				+	0.62	0.51	0.68	141	42	118	39	—	—	15	35		
119	me53	98717	>	99625	303	36065						0.69	0.55	0.67	143	48	120	41	180	35	139	22		

The positions of 119 putative ORFs in the AdorGV genome are shown and compared to homologues in other sequenced GVs and in AcMNPV. Common names, location, and direction of transcription of the identified ORFs are given. The length and the molecular weight (M_r) of the encoded peptide are calculated. Early elements identified in the −120 nt upstream regions through profile searches include TATA boxes (T, derived core: TATAW), lef-TATA-boxes (L, derived core: TAATWA), CAKT initiator elements within 25–35 nt downstream of a TATA-box (I), and CAGT-like nonconsensus TATA elements (N, core CAKT). Pattern searches for baculovirus late promoter elements (DTAAG motif) identified nominally (n, −15 to −120 nt upstream) and proximally (p, −5 to −15 nt upstream) positioned motifs. The AT and coding strand (cs) AG compositions of the ORFs are given as are the G1 values of their encoded peptides. The names and percentage identities (ID) to ORFs in other granuloviruses (CpGV, PlxyGV, XecnGV) as well as to the nucleopolyhedrovirus AcMNPV are listed. Identities in the "other" column are shown only when there was no AcMNPV homologue, but another NPV homologue, with the relevant virus. Abbreviations and GenBank Accession Numbers are as follows; S1 p10, SpliNPV (SLNPVP10); Mc76, MacoNPV (MBU59461); Ha110, HaSNPV (AF303045); Op35 and Op8, OpMNPV (OPU75930); Ld22 and Ld50, LdMNPV (AF081810). For amino acid identities a gap creation penalty of eight was used unless a correct alignment was not produced. This was in the case of some protein sequences with low similarity, in these cases a gap creation penalty of four was used (*).

bodies. Other conserved AdorGV structural genes are *p6.9* (Ador72) and *odv-e25* (Ador77) (both 68% average amino identity to GV homologues in Table 2). In contrast, *p24capsid* (Ador60), which encodes a protein associated with both ODV and BV (Wolgamot et al., 1993), is poorly conserved (50% average amino identity to GV homologues in Table 2). In CpGV, XecnGV and PlxyGV this gene also shows low similarity to its NPV homologues (Luque et al., 2001; Hashimoto et al., 2000; Hayakawa et al., 1999). AdorGV lacked homologues of the structural genes, *p80/p87-capsid* and ORF1629 (p78/83). The *p80/p87-capsid* gene is also absent in the other sequenced GVs. A putative ORF1629 (p78/83) has been identified in XecnGV (Xecn2), although it is less than half the size of the NPV ORFs and

has similarity concentrated only around a conserved proline-rich region (Hayakawa et al., 1999). All the GVs have a homologue of Xecn2 with similarity in the first 65 aa of the amino terminal region of the protein but show no similarity to ORF 1629 and do not contain a proline-rich region.

AdorGV contained homologues of the three putative p10 genes of XecnGV and PlxyGV. In NPV-infected cells, p10 forms fibrillar structures in the nucleus and cytoplasm. The protein is implicated in occlusion body morphogenesis and disintegration of the nuclear matrix, resulting in the dissemination of OBs (van Oers and Vlak, 1997). Three XecnGV ORFs (Xecn5, Xecn19, and Xecn83) present similarities to *p10.* Homologues of these three ORFs are present in PlxyGV (Plxy2, Plxy21, and Plxy50) and Hashimoto et al. (2000) suggested they are all *p10* homologues. AdorGV had a homologue to Xecn83/Plxy50/Cp62, which is Ador57. It was only 66 aa compared to 135–189 in the other GVs and showed sequence similarity to the N terminus of the other proteins. It did not contain any of the structural features common to p10s. Therefore it is unlikely that Ador57 is a *p10* homologue. There is a Xecn5/Plxy2 homologue present in AdorGV, called Ador13. If assumed to start from its second methionine it is a similar size to Xecn5/Plxy2. It showed moderate homology to Xecn5 (42%) and Plxy2 (40%) and also the p10 of *Spodoptera littoralis* NPV (SpliNPV) (33%). It contained a heptad repeat sequence and a basic C-terminus but did not contain a proline-rich domain. This has been named *p10* in the AdorGV genome and is the only ORF in the AdorGV genome not to have a CpGV homologue but to have other GV homologues. Another putative p10 was Ador17 which was 56% identical to Xecn19, Plxy21, and Cp22. Ador17 shared a number of motifs with p10, including a proline-rich domain and a heptad repeat sequence. It was 30% identical to AcMNPV p10 though it is significantly larger (336 vs 137 amino acids) and much of the sequence identity is between sequences of low complexity. Ador17 also showed similarity to the *calyx/pep* protein of NPVs. This ORF has not been identified in GVs before. However, electron microscopic evidence for a granule calyx exists (Pinnock and Hess, 1978). In NPVs, this structure is composed of a laminar polymer comprised of polysaccharides and the pp34 protein. The calyx is thought to act as a structural, stabilising factor for polyhedra in NPVs (Williams et al., 1989) and would likely play a similar role in GV.

A more difficult issue arises when assembly of the granule calyx is considered. The addition of electron dense spacers (precursor sheets of calyx) is thought to be mediated, in NPVs, by the fibrous bodies, though the experimental data supporting this belief contain conflicting observations (van Oers and Vlak, 1997; Funk et al., 1997; O'Reilly, 1997). The fibrous bodies are mainly composed of p10 protein. While the calyx and electron dense spacers are clearly evident in micrographs of GV infections, the fibrous body, p10-containing structures have yet to be identified. Furthermore, in review articles, discussion of fibrous bodies is exclusively limited to NPV. It is possible that the calyx of GV granules is attached via a p10/fibrous body-independent manner. This could indicate that there are no p10 proteins in GVs and that any such proteins identified as p10 in these viruses are solely due to fortuitous similarity via coiled-coil domains. Alternately, it might mean that GV fibrous bodies are smaller and more granular structures, similar to those in p10-fusion protein mutants created in NPVs (van Oers and Vlak, 1997). If this were the case, this would be a reflection of the expanded size of GV p10 proteins. Investigators could have overlooked such small structures.

When considering a possible lack of p10/fibrous bodies, it is interesting to note that baculoviral PTP has been found in close association with the fibrous body structures and that this peptide is absent in AdorGV (see below). However, a formal role for PTP in calyx formation or fibrous body structure has not been established.

As in all sequenced GVs and group II NPVs, AdorGV does not encode the envelope glycoprotein *gp64,* the major envelope fusion protein of AcMNPV, BmNPV, OpMNPV, and EppoMNPV (Monsma et al., 1996; Hyink et al., 2002). This protein appears to be unique to group I NPVs (Pearson et al., 2000; Ijkel et al., 2000). In LdMNPV, the envelope fusion protein is the product of the *Ld130* gene. AdorGV encodes an *Ld130* homologue, Ador23, which shows 28.5% amino acid identity to *Ld130. Ld130* homologues are present in all baculoviruses that have been completely sequenced, including those that contain *gp64.* The role of the *Ld130* homologue in the latter species is unclear (Pearson et al., 2000).

Genes involved in DNA replication and transcription

There are 19 *lef* genes in AcMNPV that have been implicated in DNA replication and transcription (Rapp et al., 1998). Early baculovirus genes are transcribed by the host cell RNA polymerase II, but these are often transactivated by genes such as *ie-0, ie-1, ie-2,* and *pe38* (Friesen, 1997). Of these genes, only *ie-1* is present in AdorGV, but it is poorly conserved. These genes seem to be poorly conserved among baculoviruses in general. Both *ie-2* and *pe38* are also absent from all group II NPVs and GVs with the exception of CpGV, which has a *pe38* gene. Six genes are reported to be essential for baculovirus DNA replication: *lef-1, lef-2, lef-3, dnapol, helicase,* and *ie-1* (Lu et al., 1997). Homologues of all are present in AdorGV. They are moderately well conserved, with the exception of *lef-3* and *ie-1* (Table 2). AdorGV does not have a *lef-7* homologue. This appears to be a group I NPV-specific gene. It stimulates transient DNA replication in AcMNPV and BmNPV (Gomi et al., 1997; Morris et al., 1994). AdorGV encodes a DNA ligase (Ador103) and a second helicase (Ador108), in common with LdMNPV, CpGV, XecnGV, and PlxyGV. The LdMNPV DNA ligase displays catalytic properties of a

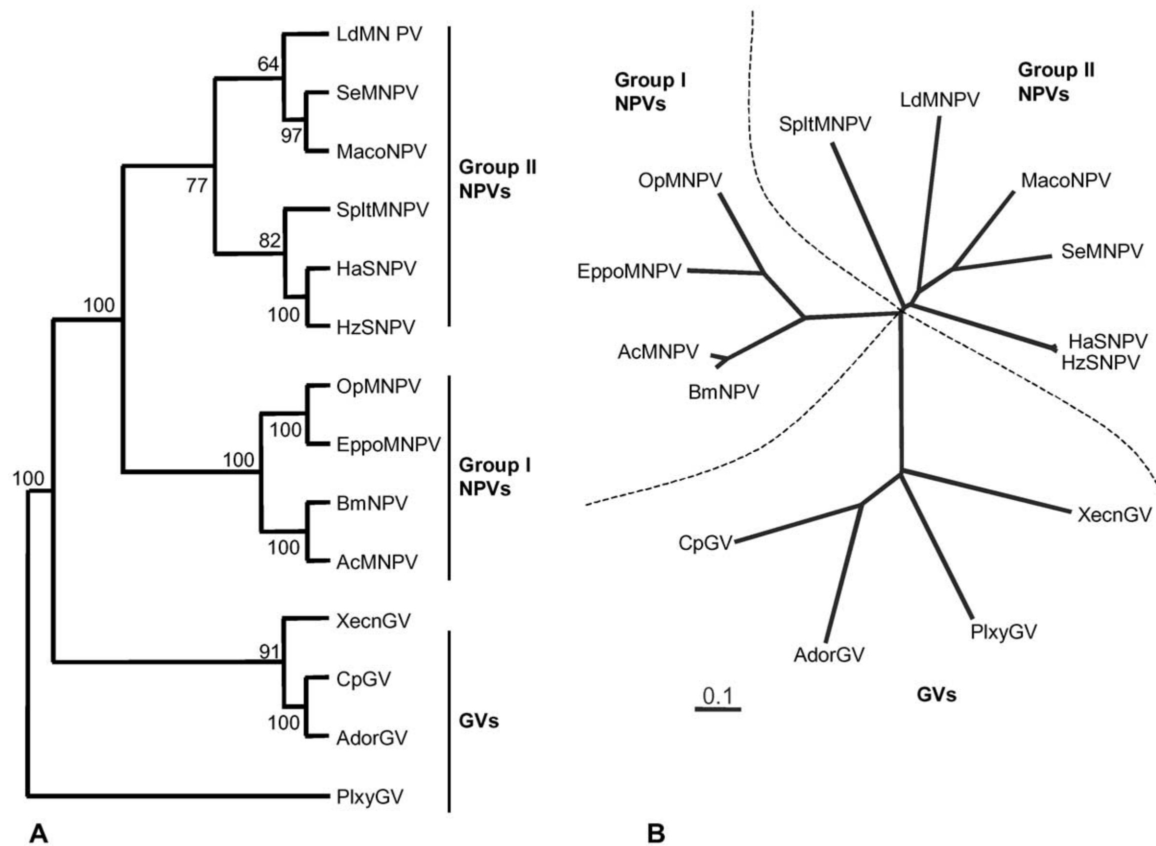

Fig. 3. Phylogenetic analysis of concatenated LEF protein sequences (LEFs 1–6, 8, 9, and 11). Unrooted trees were generated using maximum parsimony (A) and distance (B). Bootstrap percentage support values (1000 replications) are indicated for each internal branch. Full virus names are included in the text.

type III DNA ligase (Pearson and Rohrmann, 1998). The *helicase-2* gene shows similarities to a yeast mitochondrial helicase, called *pif-1* (Kuzio et al., 1999). Neither the *helicase-2* or *DNA ligase* gene stimulates transient DNA replication in LdMNPV. As their homologues are involved in DNA repair and recombination, they could be involved in DNA repair (Kuzio et al., 1999).

AdorGV lacks genes for enzymic functions in nucleotide metabolism, such as the large (*rr1*) and small (*rr2*) subunits of ribonucleotide reductase and deoxyuridyltriphosphate (*dUTPase*) as do AcMNPV, BmNPV, PlxyGV, XecnGV, HaSNPV, HzSNPV, and EppoMNPV (Ayres et al., 1994; Gomi et al., 1999; Hashimoto et al., 2000; Hayakawa et al., 1999; Chen et al., 2001, 2002; Hyink et al., 2002). These enzymes may facilitate virus replication in nondividing cells where dNTP pathways are inactive. Baculoviruses that do encode ribonucleotide reductase subunits are CpGV, OpMNPV, LdMNPV, SeMNPV, SpltMNPV, and MacoNPV (Luque et al., 2000; Ahrens et al., 1997; Ijkel et al., 1999; Kuzio et al., 1999; Pang et al., 2001; Li et al., 2002). These viruses also encode a *dUTPase,* except for CpGV.

Many genes required for late gene transcription have been described, including *lef 4–6, 8–11, 39K, p47,* and *vlf-1* (Lu and Miller, 1997). All of these are found in AdorGV, except for *lef-10.* Generally, these genes are more conserved than the early transcription activators (IJkel et al., 1999). The most conserved AdorGV LEF homologue was *lef-8* (Ador110, 66% average identity to other sequenced GVs) followed by *lef-9* (Ador100, 64%) and *p47* (Ador58, 57%), all of which are components of the virus-specific RNA polymerase (Jin et al., 1998). The most poorly conserved LEF is *lef-6* (40%). Hayakawa et al. (1999) reported it absent from XecnGV. The GV *lef-6* genes are smaller than the NPV *lef-6* genes (86–102 amino acids vs 138–187 amino acids). Functional studies will be necessary to determine whether the GV genes are true *lef-6* homologues.

Phylogenetic analysis based on gene sequences on the 14 baculovirus sequences available to date was performed using a concatenated alignment of all the common *lef* genes (*lefs* 1–6, 8, 9, and 11) (Fig. 3). The trees produced clearly divide the baculoviruses into three groups, the group I NPVs, the group II NPVs, and the GVs. Except for subtle differences in minor branch arrangement the two trees strongly support each other. Thus, the common set of *lef* genes contains minimal and adequate information to reconstruct a phylogenetic tree that is in agreement with the conclusions drawn using complete genomes in previous studies of baculovirus evolutionary relationships (Herniou et al., 2001; Hyink et al., 2002).

Genes with auxiliary functions

Auxiliary genes are genes that are not essential for viral replication, but provide it with some selective advantage (O'Reilly, 1997). The AdorGV genome does not contain either a chitinase or cathepsin gene. Chitinase and cathepsin genes have been identified in almost all other baculoviruses that have been completely sequenced to date with the exception of *Plutella xylostella* GV (Hashimoto et al., 1999). It appears that baculoviruses encode these enzymes to aid in the breakdown of insect tissues at the end of infection to release OBs into the environment and aid their horizontal spread. Deletion of either cathepsin or chitinase genes resulted in the failure of AcMNPV to cause liquefaction of the host (Hawtin et al., 1997). This indicated that these proteins function together to promote degradation of the host tissues at the end of the infection process (Hawtin et al., 1997). AdorGV infected larvae do not lyse at the end of infection. The infected larvae discharge large amounts of virus from their posterior end which is likely an alternative mechanism for disseminating virus to infect further larvae before they die.

Superoxide dismutase (*sod*) is a well-conserved gene of baculoviruses. The only genomes not to contain a SOD homologue are SpltMNPV and EppoMNPV (Pang et al., 2001; Hyink et al., 2002). SOD catalyses the dismutation of the superoxide radical O_{2-} into H_2O_2 and O_2 (Pardini, 1995). The superoxide radical and its metabolites cause damage to most cellular constituents. Therefore SODs provide some protection against oxygen toxicity (Pardini, 1995). However, the function of SODs in the baculovirus genome is unknown. The gene has been deleted in AcMNPV with no deleterious effect (Tomalski et al., 1991). Also host SODs are abundant in infected insect cells and therefore a baculovirus copy would not be required. Some studies suggest that SOD may serve to protect OBs from superoxide radicals generated by exposure to sunlight in the environment (Tomalski et al., 1991).

Ubiquitins are small proteins, which are present in abundance in eukaryotic cells. They are thought to be involved in an array of basic cellular processes and have been strongly implicated in protein degradation (Doherty and Mayer, 1992; Finley and Chau, 1991). Ubiquitin has been found in all of the baculoviruses sequenced to date. It has been found fused to *gp37* as a single ORF in SpltMNPV (Pang et al., 2001). It is a highly conserved protein and, with the exception of polyhedrin and granulin, it is the most highly conserved gene in the baculovirus genome (73% average amino identity to GV homologues in Table 2). Ubiquitin appears to be nonessential for viral replication (Reilly and Guarino, 1996) and its role in the virus life cycle is unclear.

AdorGV contains two putative fibroblast growth factors (*fgf*). The first is Ador117 which has a *fgf* domain and shows homology to GV ORFs Cp140, Plxy117, and Xecn178, all of which contain a *fgf* domain. However, these have not been annotated as *fgfs*. These are all homologues of the *fgf* of AcMNPV (Ac32). The second *fgf* is Ador105. This ORF shows homology to Cp123, Plxy104, and Xecn144 and also to Ac32. It does not contain an *fgf* domain and so is likely to be a dysfunctional *fgf*.

Enhancin is a metalloproteinase that disrupts the insect peritrophic membrane, facilitating the initiation of infection (Derksen and Granados, 1988; Wang and Granados, 1998). No *enhancin* homologue is present in AdorGV, CpGV or PlxyGV. In contrast, four *enhancin* homologues are present in XecnGV, two in LdMNPV, and one in MacoNPV. AdorGV also does not encode a *gp37* homologue. GP37 (spindlin) is related to the fusolins of entomopoxviruses, which also act as enhancing factors (Yuen et al., 1990). It remains to be seen whether AdorGV has alternative mechanisms of enhancing infection.

AdorGV, CpGV, and PlxyGV lack a *conotoxin-like* (*ctl*) homologue (Eldridge et al., 1992). Such a gene is present in XecnGV but lacking from BmNPV, SeMNPV, HaSNPV, and HzSNPV. Its biological role is unknown. Ador8 and Ador66 are homologues of Ac145/Ac150. Members of this gene family are also found in entomopoxviruses, and these ORFs contain a six-cysteine motif similar to chitin-binding proteins (Dall et al., 2001).

AdorGV possesses a gene encoding a protein kinase (PK) (Ador3) but does not contain any protein tyrosine phosphatases (PTPs). PTPs are a diverse group of proteins that can be divided into two subfamilies, the receptor and the nonreceptor PTPs, depending on whether they span the cell membrane. This gene is thought to be involved in the regulation of the phosphorylation status of the viral and host proteins during infection. A copy of *ptp* is present in all other sequenced lepidopteran baculoviruses to date with the exception of XecnGV, PlxyGV, LdMNPV, HaSNPV, and HzSNPV.

Inhibitors of apoptosis

Baculoviruses contain two types of genes with antiapoptotic activity; p35 and IAP (inhibitor of apoptosis). These genes block apoptosis induced by virus infection. P35 genes have only been identified in AcMNPV, BmNPV, and SpltMNPV. IAP homologues generally contain two baculovirus IAP repeats (BIR) (Birnbaum et al., 1994), which are associated with binding to apoptosis-inducing proteins (Vucic et al., 1997), and a C-terminal zinc finger-like (RING) Cys/His motif (Crook et al., 1993). Two members of the IAP gene family were observed in AdorGV, *iap-3* (Ador89) and *iap-5* (Ador99), using the nomenclature of Luque et al., 2001. The AdorGV *iap-3* gene has a similar amount of homology to all *iap-3* genes and does not appear to be more related to CpGV, unlike the majority of AdorGV genes. This gene is in a different region of the genome than the CpGV *iap-3* (70 and 10 m.u., respectively), and in fact is the only common gene which has a different position in the genome. This suggests that these may have arisen as independent acquisitions. The *iap-5* gene appears to be a GV-specific *iap*

which is in a similar position on the genomes. All GVs sequenced to date have *iap-5*.

Desmoplakin family

Ador95 is a homologue of Plxy94, which has been named desmoplakin as it shows similarity to an internal region of a human desmoplakin, an essential constituent of intracellular junctions (Hashimoto et al., 2000). Although a blast search using Ador95 did not come up with any desmoplakin hits, it is obviously a homologue of Plxy94, which is a gene common to all sequenced baculoviruses. The ORF Ador24 also has low similarity to Plxy94 and the homologues of Plxy94, and also has similarity to Cp32 and Xecn28. Ador24 again did not match any members of a desmoplakin family, although Cp32 did. The relationships between baculovirus desmoplakins were explored by phylogenetic analyses. AdorGV desmoplakins did not group together, suggesting they do not derive from duplications in AdorGV (data not shown).

Amplified ORFs

Baculovirus-repeated ORFs (*bro* genes) are present in a number of baculoviruses with between 1 and 16 copies. However, AdorGV did not contain any *bro* genes or possible truncated versions as in CpGV (Luque et al., 2001). The function of these genes is unclear but they have been shown to bind to DNA (Zemskov et al., 2000).

Unique AdorGV ORFs

There are 15 previously uncharacterised ORFs in the AdorGV genome. Only one of these shows significant similarity to sequences in GenBank and this is Ador56, which shares 27% aa identity to *Amsacta moorei* entomopoxvirus AMV225 which has an unknown function. Seven of these ORFs are under 100 aa. Four are larger than 100 aa with a baculovirus late promoter motif (Ador27, Ador34, Ador48, Ador54).

Repeated sequences

Many baculovirus genomes have several homologous regions (*hrs*) dispersed over the genome. Archetypal *hrs* are polydispersed. They occur as single copies or clusters of several copies at multiple locations along the genome. The *hrs* may serve as origins of replication (Kool et al., 1995) and as enhancers of transcription (Guarino and Summers, 1986; Guarino et al., 1986). An individual NPV *hr* typically comprises direct repeats usually centered around a palindrome. NPVs have several *hrs* dispersed around the genome, with variable numbers of the repeat unit in each *hr*. A second type of replication origin, a non-*hr* ori, has been identified in some NPVs (Habib and Hasnain, 2000; Heldens et al., 1997; Kool et al., 1994). These complex structures comprise multiple direct and inverted repeats within a region spanning 800–4000 bp and appear to be present only once per genome.

To date, no two GVs have had similar *hr* regions. Nine putative *hrs* have been identified in XecnGV (Hayakawa et al., 1999). These are different from most NPV *hrs* as they do not have a palindromic core. Four large *hrs* are present in PlxyGV and more like NPV *hrs,* in that the repeat unit is centered around a palindrome, although this is shorter than that found in NPVs. CpGV contains one 1.13 kbp non-*hr* like region and 13 imperfect palindromes of approximately 75 bp which show similarity to each other. These imperfect palindromes are dispersed throughout the genome with two pairs in close proximity. In the AdorGV genome there are nine regions that contain repeated regions which are unlike typical *hr* regions.

Ador19 is a large (503 aa) ORF that has weak similarity to other baculovirus low complexity ORFs such as HOAR (Le et al., 1997). It does not contain any recognisable baculovirus transcription initiation sites for early or late gene expression. It contains a repeated region (repeat 2) of 537 bp, which is very AT rich. Analysis of the coding region of Ador19 shows that the repeated sequence has a high coding strand AG content (61%) but poor dinucleotide bias (0.094) and the encoded peptide has an expected low G1 value (0.538). The HOAR ORFs have been characterised as being susceptible to rearrangement due to the repetitive sequences (Le et al., 1997). It is also of interest that Ador19 shows weak similarity to four XecnGV ORFs (ORFs 26, 42, 48, and 62). All the XecnGV ORFs are located at sites where XecnGV and AdorGV synteny breaks down. The repetitive sequence of Ador19 lacks stop codons and the presence of an ORF may have arisen through the fortuitous placement of flanking in-frame start and stop codons flanking a genetically plastic site. Similarly, AdorGV *odv-e66* may have acquired such a repeat near its C-terminal coding region.

The generation of *bros* is likely to be an artefact of genome replication and the association with *hrs.* The absence of canonical polydispersed *hrs* in AdorGV may explain the lack of repeated genes in this virus.

Repeat 1 is a 484 bp tandem repeat spanning from 10984 to 11467. It is an imperfect tandem repeat comprising two sequences of 260 and 207 bp in the C-terminal coding end of Ador18.

Repeat 2 is the largest region of repeated sequences, spanning 573 bp from 12261 to 12833 bp. This region is within Ador19, which may not be transcribed. This region contains many direct repeats from 23 to 143 bp long and is 76.2% AT rich. It is rich in the base T (51.0%) and very low in the base G (1.7%) with respect to the coding strand of granulin. This repeat region is in the same relative position as the CpGV non-*hr* like region and is closest to resembling a baculovirus non-*hr*.

Repeat 3 is 212 bp and spans from 17835 to 18077. It is located close to the N-terminus of Ador24 and consists of

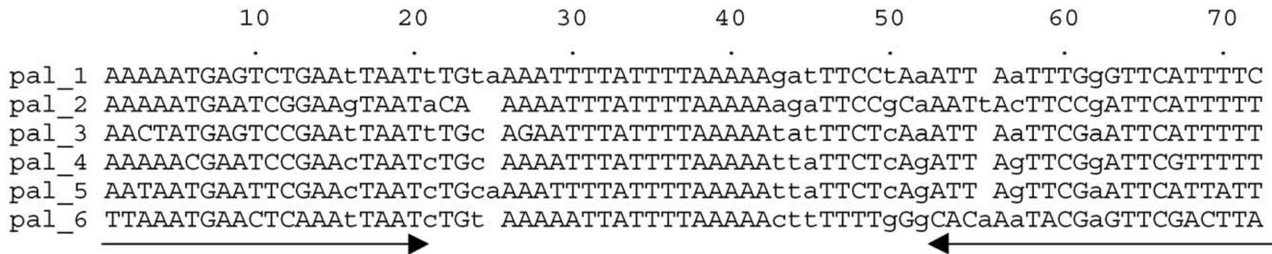

Fig. 4. The alignment of the residues comprising the imperfect palindromes of AdorGV repeat regions. Positions containing residues of at least 4 out of 6 identity are in uppercase. Palindromic regions that potentially contribute to a stem-and-loop structure (positions 1–20 and 53–72) are indicated with arrows. Alignments were performed interactively using Macaw (Schuler et al, 1991).

short direct repeats. Repeats are present in a similar position of the CpGV homologue of this gene, Cp32.

Repeat 4 is a 343 bp region from 19986 to 20328 bp which contains five small direct repeats of up to 15 bp and six inverted repeats. Within these repeats are three 72 bp imperfect palindromes (pal-1, 2 and 3), which are on average 77.5% identical to each other. This repeat region is not within an ORF and is 74.9% AT rich.

Repeat 5 is a 211 bp region of direct repeats within the N-terminus of *odv-e66* at 22754 to 22965 bp. The repeat consists of direct repeats of 16–69 bp. The sequence is G-rich (56.4%) and contains no C bases with respect to the coding strand of granulin.

Repeat 6 is a 329 bp region from 36731 to 37059 bp which contains two large inverted repeats and one large direct repeat. The direct repeat is an imperfect palindrome with the two 72 bp palindromes (pal-4 and 5) showing 91.5% identity to each other and 74.7–83.1% identity to the palindromes from repeat 4 (Fig. 4). Repeat 6 is not within an ORF and is 71.7% AT rich.

Repeat 7 is a 344 bp region from 64062 to 64405 of direct repeats which is not within an ORF. The repeat consists of direct repeats of 12–134 bp and is 64.5% AT rich.

Repeat 8 is a 72 bp imperfect palindrome (pal-6) which is in the intergenic region between the genes *desmoplakin* and *lef-3*. This palindrome is similar to the other five palindromes with on average 58.3% identity. Interestingly, one of the CpGV imperfect palindromes (rep-11) and a PlxyGV *hr* region (hr-4) are also located between their *desmoplakin* and *lef-3* genes.

Repeat 9 is a 199 bp inverted repeat at 93788–93986 bp. It has 69 bp stems and a 61 bp loop. It is not within an ORF and is 69.9% AT rich.

The largest repeat region is repeat 2 which is 537 bp and is most like non-*hr* regions, although the smallest found to date (Jehle, 2002). Repeat 4, repeat 6, and repeat 8 all contain similar imperfect 72 bp palindromes. These six imperfect palindromes come closest to representing poly-dispersed NPV *hrs.* The AdorGV palindromes are a similar length to those of CpGV and have a similar AT-rich core. Luque et al., have suggested that all CpGV *hrs* are singletons; i.e., they all only contain a single copy of the repeat unit and are not present as multiple tandem repeats like a typical *hr* element. This could also be the case for AdorGV. It has been shown that a single repeat element is sufficient for ori function (Leisy et al., 1995). However, only 53 and 57 bp separate the three palindromes in repeat 4 (pal-1, 2, and 3) and only 101 bp separates pal-4 and pal-5. It remains to be seen whether these elements actually function as replication origins or as transcription enhancers.

An interesting feature was observed in the intergenic region of some of the AdorGV ORFs with abutting C-termini (transcribed from complementary strands). Short, tandem repeats were found in the intergenic regions between Ador55/Ador56, Ador79/Ador80 (*lef-4*), and Ador82 (*odv-ec27*)/Ador83. It is possible that these are transcriptional regulatory regions.

Organization of the AdorGV genome

The genomic organization of AdorGV has been studied in a comparative manner using GeneParityPlot analysis (Hu et al., 1998) (Fig. 5). The gene order among the sequenced GVs is virtually identical with between only one and four common genes in different positions along the genome. CpGV appears to be the most collinear with only one gene, *iap-3,* in a different position. In AdorGV, 99 of 100 CpGV homologues, 93 out of 97 PlxyGV homologues, and 94 of 97 XecnGV homologues are in a conserved position. It can be seen from the gene arrangements of AdorGV compared with the other GVs that the first half of the AdorGV genome appears to contain more unique genes than the second half. This could be due to the fact that most of the repeated regions are in the first half of the genome which may have resulted in rearrangements, insertions, and deletions. The AdorGV has one main region of collinearity with the NPVs, although this is inverted with respect to the granulin/polyhedrin gene. This is from ORFs Ador69–101 (Ac103–61). A smaller region from Ador6–11 (Ac147–142) is also present.

The complete sequence of AdorGV has shown that there is a great deal of similarity in gene arrangement among the sequenced GVs. AdorGV is the second baculovirus not to have any obvious *hr* regions. It is the smallest lepidopteran baculovirus sequenced to date and it is a slow-killing virus.

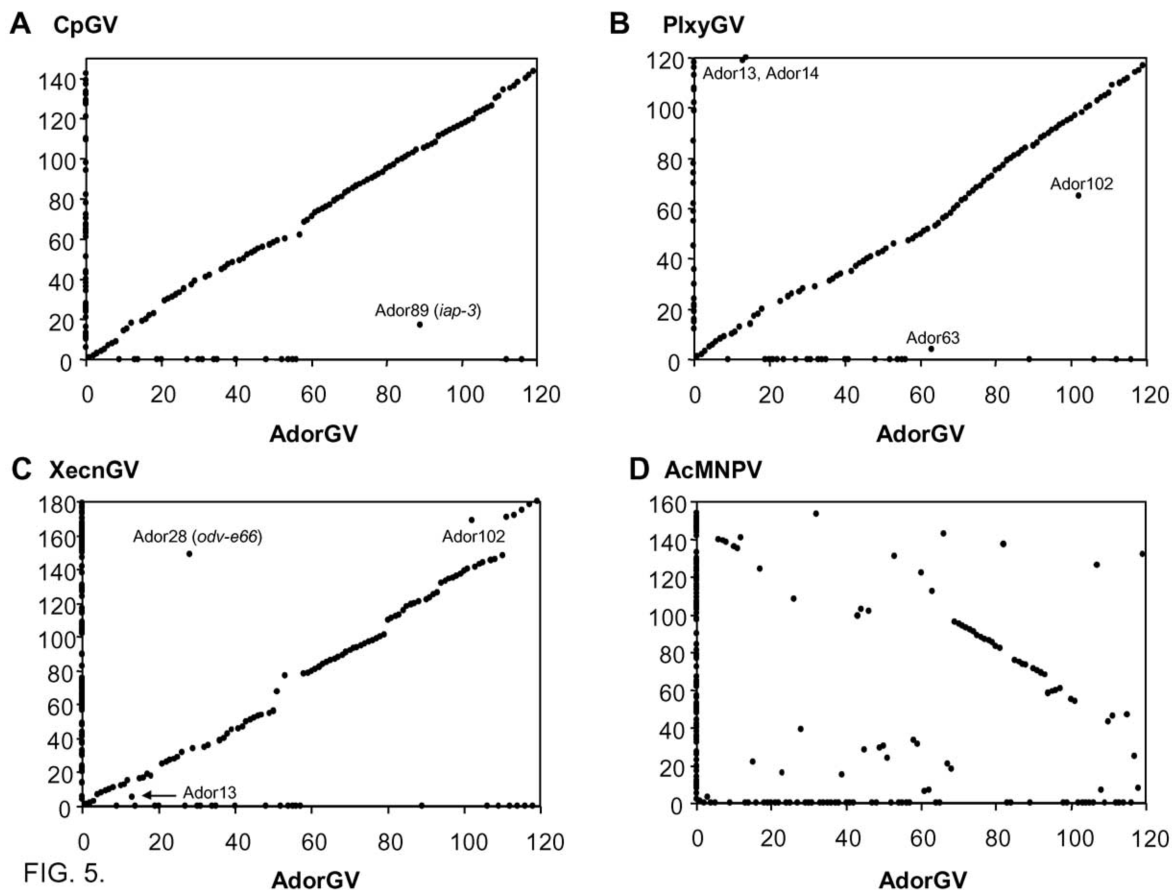

Fig. 5. Comparison of the AdorGV gene organization versus CpGV (A), PlxyGV (B), XecnGV (C), and AcMNPV (D). Homologues are plotted based on their relative location in the genome. ORFs with no homologues are aligned on the vertical and horizontal axes. In A, B, and C, homologues that have been separated from their expected neighbours are named.

Previously, slow speed of kill has been associated with a large genome and lack of an *egt* gene such as XecnGV, but AdorGV does have a functional *egt* gene (Wormleaton and Winstanley, 2001). AdorGV does not have any genes in common with slow-killing GVs, such as XecnGV, that were not present in CpGV, in fact CpGV seems to be the closest relation sequenced to date. AdorGV is a similar size to PlxyGV and lacks auxiliary genes such as chitinase and cathepsin that PlxyGV also lacks. AdorGV has 24 fewer genes than CpGV and is therefore useful for deciphering the common set of baculovirus or granulovirus genes.

Materials and methods

Virus

The English isolate of AdorGV was recovered from overwintering *A. orana* larvae collected in Kent, England, in 1993 and was propagated in laboratory stocks of *A. orana* larvae maintained on an artificial diet (Guennelon et al., 1981). A cloned genotype was obtained through three successive rounds of in vivo cloning, using the limiting dilution method described by Smith and Crook (1988a). The virus occlusion bodies were purified using glycerol and sucrose gradients following methods described previously (Crook and Payne, 1980) and the DNA was extracted and purified following methods described previously (Smith and Crook, 1988b).

Construction of genomic DNA libraries

Libraries of AdorGV DNA fragments were constructed by ligating restricted viral DNA fragments into pBluescript II SK+ (Stratagene) using T4 DNA ligase (Life Technologies). Recombinant plasmids were cloned and propagated in *Escherichia coli* DH5α and purified by alkaline lysis.

Sequencing

The nucleotide sequence of double-stranded DNA fragments of the AdorGV genome was determined utilising the dideoxynucleotide chain termination method (Sanger et al., 1977). Sequencing reactions, using ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems), were set up according to the

manufacturer's recommendations. The sequencing reactions were carried out using the Gene Amp PCR Systems 9600 and analysed on an ABI 377 automated DNA Sequencer (Applied Biosystems, Inc.). Universal pUC forward and reverse primers and custom primers were used for sequencing. Double-stranded DNA sequences were assembled using the SeqMan II sequence analysis package (Lasergene software Version 4.03; DNAstar, Inc., Madison, WI). The coding regions were predicted using the package GeneQuest II (DNAstar) by locating translation start and stop codons of open reading frames (ORFs) of 50 or more amino acids (aa). Database searches using the program PSI-BLAST were used to identify proteins sharing similarity. Percentage pairwise identities were calculated using the GAP program of Wisconsin Package Version 10.0, Genetics Computer Group (GCG) (Madison, WI) (Devereux et al., 1984), with default settings. Multiple alignments and phylogenetic trees were produced as in Wormleaton and Winstanley (2001). GeneParityPlot analysis was performed on the AdorGV genome versus other baculovirus genomes, as described previously (Hu et al., 1998).

Promoter regions within 120 nt upstream of ORFs were identified through simple pattern matching, or for less characterised elements, through a Gibbs sampling procedure (Ficket, 1996; Lawrence et al., 1993). In the Gibbs sampling, initial scan profiles were not derived at random, but were produced by seeding with an initial pattern match or profile search. After this, the Gibbs sampling procedure was performed as described. Since not all upstream regions were expected to contain any given element, regions lacking a significant hit to the generated weight matrix were removed from the pool and the Gibbs sampling procedure was repeated on the remaining upstream regions. This process was carried out until no further upstream regions could be culled. This resulted in a tightened scope for the final weight matrix and limited the reporting of false positives, as measured by the hits that deviated significantly from the original input pattern or profile.

Encoded peptides were measured for G1 global complexity (Wan and Wootton, 2000). AdorGV ORF dinucleotide frequencies and codon biases were calculated as described (Karlin, 1998). Correspondence analysis of codon usage was performed using CodonW (Peden, 1997).

Acknowledgments

We are grateful to Sarah Breeds for her contributions to the sequencing project. This work was supported by BBSRC core funding.

References

Afonso, C.L., Tulman, E.R., Lu, Z., Balinsky, C.A., Moser, B.A., Becnel, J.J., Rock, D.L., Kutish, G.F., 2001. Genome sequence of a baculovirus pathogenic for *Culex nigripalpus*. J. Virol. 75, 11157–11165.

Ahrens, C.H., Russell, R.L.Q., Funk, C.J., Evans, J.T., Harwood, S.H., Rohrmann, G.F., 1997. The sequence of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus genome. Virology 229, 381–399.

Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M., Possee, R.D., 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology 202, 586–605.

Birnbaum, M.J., Clem, R.J., Miller, L.K., 1994. An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. J. Virol. 68, 2521–2528.

Bischoff, D.S., Slavicek, J.M., 1995. Identification and characterization of an early gene in the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus. J. Gen. Virol. 76, 2933–2940.

Chen, X., WF, I.J., Tarchini, R., Sun, X., Sandbrink, H., Wang, H., Peters, S., Zuidema, D., Lankhorst, R.K., Vlak, J.M., Hu, Z., 2001. The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. J. Gen. Virol. 82, 241–257.

Chen, X.W., Zhang, W.J., Wong, J., Chun, G., Lu, A., McCutchen, B.F., Presnail, J.K., Herrmann, R., Dolan, M., Tingey, S., Hu, Z.H., Vlak, J.M., 2002. Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. J. Gen. Virol. 83, 673–684.

Crook, N.E., 1991. Baculoviridae: subgroup B: comparative aspects of granulosis viruses, in: Kurstak, E. (Ed.), Viruses of Invertebrates, Dekker, New York, pp. 73–110.

Crook, N.E., Clem, R.J., Miller, L.K., 1993. An apoptosis-inhibiting baculovirus gene with a zinc-finger like motif. J. Virol. 67, 2168–2174.

Crook, N.E., Payne, C.C., 1980. Comparison of three methods of ELISA for baculoviruses. J. Gen. Virol. 46, 26–37.

Dall, D., Luque, T., O'Reilly, D., 2001. Insect-virus relationships: sifting by informatics. BioEssays 23, 184–193.

Derksen, A.C.G., Granados, R.R., 1988. Alteration of a lepidopteran peritrophic membrane by baculoviruses and enhancement of viral infectivity. Virology 167, 242–250.

Devereux, J., Haeberli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.

Doherty, F.J., Mayer, R.J., 1992. Intracellular Proteins. IRL Press, Oxford.

Eldridge, R., Li, Y., Miller, L.K., 1992. Characterization of a baculovirus gene encoding a small conotoxin-like polypeptide. J. Virol. 66, 6563–6571.

Federici, B., 1997. Baculovirus pathogenesis, in: Miller, L.K. (Ed.), The Baculoviruses, Plenum, New York, pp. 33–60.

Ficket, J.W., 1996. Quantitative discrimination of MEF2 sites. Mol. Cell. Biol. 16, 437–441.

Finley, D., Chau, V., 1991. Ubiquitination. Annu. Rev. Cell Biol. 7, 25–69.

Friesen, P.D., 1997. Regulation of baculovirus early gene expression, in: Miller, L.K. (Ed.), The Baculoviruses, Plenum, New York, pp. 141–170.

Funk, J., Braunagel, S.C., Rohrmann, G.F., 1997. Baculovirus structure, in: Miller, L.K. (Ed.), The Baculoviruses, Plenum, New York, pp. 7–32.

Gomi, S., Majima, K., Maeda, S., 1999. Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. J. Gen. Virol. 80, 1323–1337.

Gomi, S., Zhou, C.E., Yih, W., Majima, K., Maeda, S., 1997. Deletion analysis of four of eighteen late gene expression factor gene homologues of the baculovirus, BmNPV. Virology 230, 35–47.

Guarino, L.A., Gonzalez, M.A., Summers, M.D., 1986. Complete sequence and enhancer function of the homologous DNA regions of *Autographa californica* nuclear polyhedrosis virus. J. Virol. 60, 224–229.

Guarino, L.A., Summers, M.D., 1986. Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. J. Virol. 60, 215–223.

Guennelon, G., Audemard, H., Fremond, J.C., El Idrissi Ammari, M.A., 1981. Progrès réalisés dans l'elevage permanent du Carpocapse (*Laspeyresia pomonella* L.) sur milieu artificiel. Agronomie 1, 59–64.

Habib, S., Hasnain, S.E., 2000. Differential activity of two non-hr origins during replication of the baculovirus *Autographa californica* nuclear polyhedrosis virus genome. J. Virol. 74, 5182–9.

Hashimoto, Y., Hayakawa, T., Ueno, Y., Fujita, T., Sano, Y., Matsumoto, T., 2000. Sequence analysis of the *Plutella xylostella* granulovirus genome. Virology 275, 358–372.

Hawtin, R.E., Arnold, K., Ayres, M.D., Zanotto, P.M.D., Howard, S.C., Gooday, G.W., Chappell, L.H., Kitts, P.A., King, L.A., Possee, R.D., 1995. *Autographa californica* nuclear polyhedrosis virus genome. Virology 212, 673–685.

Hawtin, R.E., Zarkowska, T., Arnold, K., Thomas, C.J., Gooday, G.W., King, L.A., Kuzio, J.A., Possee, R.D., 1997. Liquefaction of *Autographa californica* nucleopolyhedrovirus-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. Virology 238, 243–253.

Hayakawa, T., Ko, R., Okano, K.S, eong, S.-I., Goto, C., Maeda, S., 1999. Sequence analysis of the *Xestia c-nigrum* granulovirus genome. Virology 262, 277–297.

Heldens, J.G., Broer, R., Zuidema, D., Goldbach, R.W., Vlak, J.M., 1997. Identification and functional analysis of a non-hr origin of DNA replication in the genome of *Spodoptera exigua* multicapsid nucleopolyhedrovirus. J. Gen. Virol. 78, 1497–506.

Herniou, E.A., Luque, T., Chen, X., Vlak, J.M., Winstanley, D., Cory, J.S., O'Reilly, D., 2001. Use of whole genome sequence data to infer Baculovirus phylogeny. J. Virol. 75, 8117–8126.

Hu, Z.H., Arif, B.M., Jin, F., Martens, J.W.M., Chen, X.W., Sun, J.S., Zuidema, D., Goldbach, R.W., Vlak, J.M., 1998. Distinct gene arrangement in the *Buzura suppressaria* single- nucleocapsid nucleopolyhedrovirus genome. J. Gen. Virol. 79, 2841–2851.

Hyink, O., Dellow, R.A., Olsen, M.J., Caradoc-Davies, K.M.B., Drake, K., Herniou, E.A., Cory, J.S., O'Reilly, D.R., Ward, V.K., 2002. Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus. J. Gen. Virol. 83, 957–971.

Ijkel, W.F.J., van Strien, E.A., Heldens, J.G.M., Broer, R., Zuidema, D., Goldbach, R.W., Vlak, J.M., 1999. Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. J. Gen. Virol. 80, 3289–3304.

Ijkel, W.F.J., Westenberg, M., Goldbach, R.W., Blissard, G.W., Vlak, J.M., 2000. A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. Virology 275, 30–41.

Jehle, J.A., 2002. The expansion of a hypervariable, non-*hr* ori-like region in the genome of *Cryptophlebia leucotreta* granulovirus provides *in vivo* evidence for the utilization of baculovirus non-*hr* oris during replication. J. Gen. Virol. 83, 2025–2034.

Jin, J.P., Dong, W., Guarino, L.A., 1998. The LEF-4 subunit of baculovirus RNA polymerase has RNA 5′-triphosphatase and ATPase activities. J. Virol. 72, 10011–10019.

Kang, W., Crook, N.E., Winstanley, D., O'Reilly, D.R., 1997. Complete sequence and transposon mutagenesis of the *Bam*HI J fragment of *Cydia pomonella* granulosis virus. Virus Genes 14, 131–6.

Karlin, S., 1998. Global dinucleotide signatures and analysis of genomic heterogeneity. Curr. Opin. Microbiol. 1, 598–610.

Ko, R., Okano, K., Maeda, S., 2000. Structural and functional analysis of the *Xestia c-nigrum* granulovirus matrix metalloproteinase. J. Virol. 74, 11240–11246.

Kool, M., Ahrens, C.H., Vlak, J.M., Rohrmann, G.F., 1995. Replication of baculovirus DNA. J. Gen. Virol. 76, 2103–2118.

Kool, M., Goldbach, R.W., Vlak, J.M., 1994. A putative non-hr origin of DNA replication in the *Hin*dIII-K fragment of *Autographa californica* multiple nucleocapsid nuclear polyhedrosis virus. J. Gen. Virol. 75, 3345–52.

Kuzio, J., Pearson, M.N., Harwood, S.H., Funk, C.J., Evans, J.T., Slavicek, J.M., Rohrmann, G.F., 1999. Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. Virology 253, 17–34.

Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., Wootton, J.C., 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science 262, 208–214.

Le, T.H., Wu, T., Robertson, A., Bulach, D., Cowan, P., Goodge, K., Tribe, D., 1997. Virus Res 49, 67–77.

Leisy, D. J., Rasmussen, C., H. -T., K. Rohrmann, G. F., (1995). The *Autographa californica* nuclear polyhedrosis virus homologous region 1a: identical sequences are essential for DNA replication activity and transcriptional enhancer function. Virology 208, 742–752.

Li, Q.J., Donly, C., Li, L.L., Willis, L.G., Theilmann, D.A., Erlandson, M., 2002. Sequence and organization of the *Mamestra configurata* nucleopolyhedrovirus genome. Virology 294, 106–121.

Lu, A., Krell, P., Vlak, J., Rohrmann, G., 1997. Baculovirus DNA replication. In Miller, L.K. (Ed.), The Baculoviruses. Plenum, New York, pp. 171–192 Edited by.

Lu, A., Miller, L.K., 1997. Regulation of baculovirus late and very late gene expression. In Miller, L.K. (Ed.), The Baculoviruses. Plenum, New York, pp. 193–216 Edited by.

Luque, T., Finch, R., Crook, N., O'Reilly, D.R., Winstanley, D., 2001. The complete sequence of *Cydia pomonella* granulovirus genome. J. Gen. Virol. 82, 2531–2547.

Monsma, S.A., Oomens, A.G.P., Blissard, G.W., 1996. The GP64 envelope fusion protein is an essential baculovirus protein required for cell to cell transmission of infection. J. Virol. 70, 4607–4616.

Morris, T.D., Todd, J.W., Fisher, B., Miller, L.K., 1994. Identification of lef-7: a baculovirus gene affecting late gene expression. Virology 200, 360–9.

Mount, D.W., 2001. Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 308.

Nakai, M., Goto, G., Shiotsuki, T., Kunimi, Y., 2002. Granulovirus prevents pupation and retards development of *Adoxophyes honmai* larvae. Physiol. Entomol. 27, 157–164.

Nishi, Y., Nonaka, T., 1996. Biological control of the tea tortrix. Agrochem. Jpn. 69, 7–10.

O'Reilly, D., 1997. Auxiliary genes of baculoviruses, in: Miller, L.K. (Ed.), The Baculoviruses, Plenum, New York, pp. 267–300.

Page, R.D.M., 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

Pang, Y., Yu, J.X., Wang, L.H., Hu, X.H., Bao, W.D., Li, G., Chen, C., Han, H., Hu, S.N., Yang, H.M., 2001. Sequence analysis of the *Spodoptera litura* multicapsid nucleopolyhedrovirus genome. Virology 287, 391–404.

Pardini, R.S., 1995. Toxicity of oxygen from naturally-occurring redox-active pro-oxidants. Arch. Insect Biochem. Physiol. 29, 101–118.

Pearson, M.N., Groten, C., Rohrmann, G.F., 2000. Identification of the *Lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the Baculoviridae. J. Virol. 74, 6126–6131.

Pearson, M.N., Rohrmann, G.F., 1998. Characterisation of a baculovirus-encoded ATP-dependent DNA ligase. J. Virol. 72, 9142–9149.

Peden, J., (1997). CodonW. http://www.molbiol.ox.ac.uk/cu/culong.html.

Pinnock, D.E., Hess, R.T., 1978. Morphological variations in the cytopathology associated with granulosis virus in the fruit-tree leaf roller, *Archips argyrospila*. J. Ultrastruct. Res. 63, 252–260.

Rapp, J.C., Wilson, J.A., Miller, L.K., 1998. Nineteen baculovirus open reading frames, including LEF-12, support late gene expression. J. Virol. 72, 10197–10206.

Reilly, L.M., Guarino, L.A., 1996. The viral ubiquitin gene of *Autographa californica* nuclear polyhedrosis virus is not essential for viral replication. Virology 218, 243–247.

Sanger, F., Nicklen and Coulsen, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. USA 74, 5463–7.

Schuler, G.D., Altschul, S.F., Lipman, D.J., 1991. A workbench for multiple alignment construction and analysis. Proteins-Struct. Funct. Genet. 9, 180–190.

Sharp, P.M., Lloyd, A.T., 1993. Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. Nucleic Acids Res. 21, 179–83.

Smith, I.R.L., Crook, N.E., 1988a. *In vivo* isolation of baculovirus genotypes. Virology 166, 240–244.

Smith, I.R.L., Crook, N.E., 1988b. Physical maps of the genomes of four variants of *Artogeia rapae* granulosis virus. J. Gen. Virol. 69, 1741–1747.

Tomalski, M.D., Eldridge, R., Miller, L.K., 1991. A baculovirus homologue of a Cu/Zn superoxide dismutase gene. Virology 184, 149–161.

van Oers, M.M., Vlak, J.M., 1997. The baculovirus 10-kDa protein. J. Inverteb. Pathol. 70, 1–17.

van Regenmortel, M.H.V., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B., 2000. Virus Taxonomy—Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego.

Vucic, D., Kaiser, W.J., Harvey, A.J., Miller, L.K., 1997. Inhibition of reaper-induced apoptosis by interaction with inhibitor of apoptosis proteins (IAPs). Proc. Nat. Acad. Sci. USA 94, 10183–10188.

Wan, H., Wootton, J.C., 2000. A global compositional complexity measure for biological sequences: AT-rich and GC-rich genomes encode less complex proteins. Comput. Chem. 24, 71–94.

Wang, P., Granados, R.R., 1998. Observations on the presence of the peritrophic membrane in larval *Trichoplusia ni* and its role in limiting baculovirus infection. J. Invertebr. Pathol. 72, 57–62.

Williams, G.V., Rohel, D.Z., Kuzio, J., Faulkner, P., 1989. A cytopathological investigation of *Autographa californica* nuclear polyhedrosis virus p10 gene function using insertion/deletion mutants. J. Gen. Virol. 70, 187–202.

Winstanley, D., Crook, N.E., 1993. Replication of *Cydia pomonella* granulosis virus in cell cultures. J. Gen. Virol. 74, 1599–1609.

Winstanley, D., O'Reilly, D.R., 1999. Granuloviruses, in: Webster, R., Granoff, A. (Eds.), The Encyclopedia of Virology, second ed., Vol 1. Academic Press, London, pp. 140–146.

Wolgamot, G.M., Gross, C.H., Russell, R.L.Q., Rohrmann, G.F., 1993. Immunocytochemical characterization of p24, a baculovirus capsid-associated protein. J. Gen. Virol. 74, 103–107.

Wormleaton, S.L., Winstanley, D., 2001. Phylogenetic analysis of conserved genes within the ecdysteroid UDP-glucosyltransferase gene region of the slow-killing *Adoxophyes orana* granulovirus. J. Gen. Virol. 82, 2295–2305.

Yuen, L., Dionne, J., Arif, B., Richardson, C., 1990. Identification and sequencing of the spheroidin gene of *Choristoneura biennis* entomopoxvirus. Virology 175, 427–433.

Zemskov, E.A., Kang, W., Maeda, S., 2000. Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. J. Virol. 74, 6784–9.