Comprehensive physical map of the *Cydia pomonella* granulovirus genome and sequence analysis of the granulin gene region

Norman E. Crook,† Jackie D. James, Ian R. L. Smith‡ and Doreen Winstanley

Horticulture Research International, Wellesbourne, Warwick CV35 9EF, UK

A cloned strain of *Cydia pomonella* granulovirus, CpGV-M1, was obtained using successive rounds of an *in vivo* limiting dilution method. A detailed physical map of the genome was constructed using 11 restriction enzymes. The region containing the granulin gene and an open reading frame immediately upstream of the granulin gene was sequenced. This region showed a high degree of homology to the equivalent region from *Cryptophlebia leucotreta* granulovirus with 98% amino acid identity for the granulins and 68% identity for the putative polypeptides encoded by the upstream ORFs. These latter polypeptides contained two zinc finger-like motifs and showed a low degree of homology to ME53 from *Autographa californica* nucleopolyhedrovirus (AcMNPV). Evidence is presented for a similar upstream ORF in *Artogeia rapae* GV also. Hybridization studies showed that the CpGV genome had a similar overall organization to the *Artogeia rapae* GV genome. Hybridization between CpGV and AcMNPV was limited to fragments spanning about 15% of each genome suggesting that very few genes are highly conserved between GVs and NPVs.

Introduction

*Cydia pomonella* granulovirus (CpGV) is one of the most highly pathogenic baculoviruses known (Crook, 1991) and is a very effective control agent of the codling moth, *C. pomonella*, a worldwide pest on apples, pears and walnuts. It is also the only GV for which there currently exist susceptible cell lines (Winstanley & Crook, 1993). Very little is known about the molecular biology of GVs and how closely related they are to the more thoroughly studied nucleopolyhedroviruses (NPVs). These two genera, GVs and NPVs, within the *Baculoviridae* are distinguished primarily on the basis of their morphology but other differences may be just as fundamental. For example, NPVs replicate entirely within the intact nuclei of infected cells whereas GVs either cause early disruption of the nuclear membrane and replicate in the mixed nuclear and cytoplasmic region or, more rarely, replicate in the cytoplasm of cells which retain intact and apparently normal nuclei. We have recently demonstrated *in vitro* that CpGV uses both of these modes of replication (D. Winstanley & N. E. Crook, unpublished data).

The practical value of CpGV, the ability to be able to study its replication in detail and the potential for production of recombinant strains of the virus, make CpGV an excellent choice for a more detailed investigation of the molecular biology of a granulosis virus. A detailed physical map is a prerequisite for such work. We previously produced a physical map for the CpGV-M strain of the virus (Crook *et al*., 1985) but this contained only 36 restriction sites. Also, although restriction profiles of the CpGV-M isolate contained no visible submolar bands, most baculovirus isolates contain a mixture of genotypes and it would be highly desirable to eliminate the possibility of genotypic variability by working with a cloned strain of the virus.

In this paper, we present a much more detailed physical map constructed using a cosmid library of the genome from a cloned strain of CpGV-M. To gain an initial overview of areas of the genome containing genes closely related to those in *Autographa californica* NPV (AcMNPV), the best characterized NPV, a series of hybridizations was carried out. One of the

Author for correspondence: Doreen Winstanley.
Fax +44 1789 470552. e-mail Doreen.Winstanley@HRI.AC.UK
† Deceased.
‡ Current address: Central Science Laboratory, Sand Hutton, York YO4 1LZ, UK.

The nucleotide sequence reported here will appear in the EMBL, GenBank and DDBJ databases under accession no. Y09478.
most highly conserved baculovirus genes codes for the occlusion body protein, granulin or polyhedrin, and the sequence of the CpGV granulin gene together with an open reading frame immediately upstream of the granulin gene is reported.

Methods

- **Virus**. The CpGV-M isolate was that used previously (Crook et al., 1985). Cloned strains of this virus were obtained using three successive rounds of the in vivo limiting dilution method described by Smith & Crook (1988a).

- **Production of cosmid library**. Virus DNA, purified as described previously (Smith & Crook, 1988b), was partially digested with SalI and ligated into SalI-cut pVK102 (Knauf & Nester, 1982). Ligated DNA was added to a packaging mix (Promega) and used to transform E. coli HB101 cells. Colonies were screened using standard methods (Sambrook et al., 1989).

- **Restriction enzyme analysis and hybridization studies**. DNA was digested with restriction enzymes and electrophoresed in agarose gels using standard methods (Sambrook et al., 1989). Fragment sizes were computed using a BASIC version of the program given by Schaffer & Sederoff (1981). For hybridization, DNA was vacuum blotted onto membranes; Immobilon N (Millipore) and nitrocellulose (BA85; Schleicher and Schuell) were used for AcNPV and ArGV DNA, respectively. Probes were labelled with ³²P using an oligolabelling kit (Pharmacia). Hybridizations and washings were carried out under low stringency conditions at 55 °C. The hybridization buffer contained 6 x SSC, 0.1 M EDTA, 5 x Denhardt's solution, 0.5 % SDS and denatured salmon sperm DNA (5 μg/ml for Immobilon N and 100 μg/ml for nitrocellulose); the two post-hybridization washings for Immobilon N were 3 x SSC, 10 mM sodium phosphate pH 7.0, 10 x Denhardt's solution, 5 % SDS followed by 1 x SSC, 1 % SDS and for nitrocellulose two washes with 6 x SSC.

- **Sequencing**. Dideoxy sequencing was carried out on nested deletions generated with ExoIII and mung bean nuclease together with additional readings using synthesized primers. Sequencing reactions were done with Sequenase (Amersham) and 7-deaza nucleotides. Sequencing reactions were also run on 4 % NuSieve (Marine Colloids) gels to look for small fragments (down to about 200 bp), the only additional fragment detected was the previously mapped 299 bp Apal-C fragment (data not shown). The sizes of the fragments, shown in Table 1, were a mean of at least four determinations and in many cases were also based on the sizes of double digest fragments. The size of the entire genome was calculated to be 125.6 kbp, in close agreement with the earlier estimate of 126 kbp.

**Cosmid library of CpGV-M1**

Fragments produced by partial digestion of CpGV-M1 with SalI were ligated into pVK102. Analysis of the resulting clones showed that all the CpGV SalI fragments were represented. For most of the genome, clones were found containing segments differing by a single SalI fragment which allowed the order of SalI fragments along the genome to be ascertained. The area most poorly represented was that containing the SalI-H, -F, -O, -P, -Q, -W fragments with only a single clone found containing SalI-F, -O and -P. This was presumably due to the large number of small fragments (SalI-O, -P, -Q, -W and -V) adjacent to the large SalI-A fragment which reduced the likelihood of suitably sized partial digest fragments spanning this region.

**Physical map of the CpGV genome**

We had previously mapped the CpGV genome using EcoRI, BamHI, SmaI, Apal and HindIII, but this provided only 36 sites in total and left areas up to 15 kbp without any mapped sites. Digests of CpGV with SalI, Xhol, PstI, KpnI, SacI and BgIII gave a total of 76 fragments and thus mapping all these fragments would more than treble the number of sites compared to the earlier map. The location of most of these sites was determined from extensive double digest data on cosmids 64, 65, 17, 69, 2 and 73 although other cosmids were sometimes used, particularly to confirm the identity of fragments which crossed from one cosmid to another. In a few cases where some ambiguity remained, individual restriction fragments were cloned and mapped to provide additional data. This allowed all 112 sites to be mapped unambiguously (Fig. 2). The zero point for this map was the start of the granulin gene rather than the previously used BamHI restriction site which is now at 0.2 kbp. It seems most appropriate to include the entire granulin gene, including promoter region, at the left hand end of the map. Although the promoter region has not been analysed by any functional studies, the highly conserved nature of granulin promoter sequences shown in Fig. 4(b) suggests that the first T of this conserved sequence, i.e. nt 1084 on Fig. 3(b) would be the most appropriate starting base.

**DNA sequence of the granulin gene region**

Previous hybridization studies with a *Trichoplusia ni* GV (TnGV) granulin gene probe had shown the granulin gene to be located at the adjacent ends of EcoRI-I & -L and BamHI-K & -I
Fig. 1. Restriction enzyme profiles of CpGV-M1 DNA for SalI, EcoRI, BamHI, XhoI, PstI, KpnI, SacI, BglII, Smal, Apol and HindIII. λ DNA digested with HindIII is included for molecular size standards.

Table 1. CpGV-M1 DNA restriction fragment sizes (kbp)

<table>
<thead>
<tr>
<th>Fragment</th>
<th>SalI</th>
<th>EcoRI</th>
<th>BamHI</th>
<th>XhoI</th>
<th>PstI</th>
<th>KpnI</th>
<th>SacI</th>
<th>BglII</th>
<th>Smal</th>
<th>Apol</th>
<th>HindIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>23·1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>9·4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>6·6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td>23·1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td>2·3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>2·0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td>0·56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td>2·3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td>2·0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
<td></td>
<td></td>
<td>0·56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0·56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0·56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0·56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2·0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0·56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 2. Map of the CpGV-M1 genome for SalI, EcoRI, BamHI, XhoI, PstI, KpnI, SacI, BglII, Smal, Apal and HindIII. The position in kbp of each restriction site is shown. The zero point for the map is the start of the granulin gene. Fragments contained within selected cosmids are shown by hatched boxes at the top of the map.
Although the granulin promoter region is highly conserved, the rest of the intergenic region between the ORF124R homologues and granulin for CpGV, CIGV and ArGV shows great variability in both length (11–55 nt) and composition (19–64% GC) (Fig. 4 b).

Like CIGV ORF09, ORF124R showed low homology to AcMNPV ME53 (Knebel-Mörsdorf et al., 1993) and gave 22% aa identity and 44% similarity after the introduction of 11 gaps. These CpGV and CIGV polypeptides with 303 aa are also nearly one-third shorter than ME53 with 449 aa. This level of homology is too low to indicate whether these genes have a conserved function. However, the two GV ORFs and ME53 all have two putative zinc finger motifs which suggests some level of functional similarity.

The promoter region of ORF124R possessed early promoter elements but otherwise showed little similarity to the promoter region in CIGV apart from a short conserved sequence from $-5$ to $-12$. ORF124R has a potential early transcription start site (CAGT) at $-25$ but the only normal TATA box sequence (TATAA) is at $-103$, which is unusually distant from the CAGT sequence. There is, however, an AT region (TATTATAA) at the location ($-60$ to $-50$) where the TATA box might be expected to occur. It is possible, therefore, that this sequence acts as the basal promoter rather than the more normal TATAA sequence further upstream. Unlike ORF0909 and ME53, there is no baculovirus late promoter sequence (TAAG) in this region.

Hybridization studies

The relatedness of CpGV to ArGV and to AcMNPV was determined by hybridization to Southern blots of viral DNA. ArGV was chosen because, like CpGV, it is a highly virulent GV but with a very different host range. AcMNPV was compared because it is the best characterized baculovirus and areas of homology might thus be used to help locate genes in CpGV.

Hybridization to ArGV. In an initial hybridization to a Southern blot of EcoRI-digested CpGV DNA, using fragments representing the whole ArGV genome as a probe, all CpGV fragments gave a strong signal. To analyse this in more detail and confirm that the hybridization was specifically between individual segments of the two genomes, five cosmid clones spanning the ArGV genome were individually used as probes
Fig. 3. For legend see page 969.
CpGV map and granulin gene sequence

Fig. 4. (a) Comparison of amino acid sequences from CpGV ORF124R, ClGV ORF909 and a short region from ArGV just upstream of granulin. Identical residues are boxed. (b) Comparison of granulin promoters and upstream regions for CpGV, ClGV, ArGV, LoGV and TnGV. For CpGV, ClGV and ArGV the complete intergenic sequences between the stop codons (marked with asterisks) of the sequences shown in (a) and the start codons (also marked with asterisks) of granulin are shown.

with blots of EcoRI and BamHI digests of CpGV. Each cosmid hybridized to a subset of fragments from the two digests. Regions of CpGV DNA which hybridized to each ArGV cosmid are represented in Fig. 5. This showed that each segment of ArGV DNA hybridized to a contiguous set of CpGV DNA fragments located at a relatively similar position along the genome.

Hybridization to AcMNPV. Hybridizations to Southern blots of AcMNPV DNA digested with EcoRI, BamHI, PstI, HindIII and KpnI were performed using the CpGV genome as a probe. Almost 50% of the AcMNPV bands gave no hybridization signal and the remainder showed considerable variation in signal intensity. An analysis of this hybridization data with reference to the AcMNPV restriction map indicated that at least five distinct regions showed some hybridization to the CpGV genome. One of these regions, BamHI-F, contained the polyhedrin gene and since the CpGV granulin gene had previously been located, this was not used as a probe. Fragments from the other four regions of AcMNPV were then individually used as probes with Southern blots of CpGV digested with SalI, EcoRI, BamHI, PstI and XhoI, to identify the location of the homologous region on the CpGV genome. In each case, fragments which hybridized all overlapped a single region of the CpGV genome. Table 2 shows the minimum region identified in this way with each probe. The Southern blots used for these hybridizations also contained ArGV DNA digested with HindIII and BamHI, and TnGV DNA digested with EcoRI and HindIII. Two of the probes, AcMNPV EcoRI-M and PstI-M/EcoRI-E, hybridized to single bands in each of the ArGV digests which overlapped map units 88.8–84.1 respectively. These are very similar to the map unit locations with these probes for CpGV which were 89.0–94.7 and 79.6–85.5 respectively. EcoRI-M also hybridized to single...
bands in each of the TnGV digests, although the genomic location of these is not known. AcMNPV fragments HindIII-L and -P did not hybridize to either ArGV or TnGV, nor did the PstI-M/EcoRI-E probe hybridize to TnGV.

Discussion

A detailed physical map of a cloned strain of CpGV was produced as a basis for more detailed studies on the molecular biology of this virus. Mapping was greatly aided by construction of a cosmid library of the CpGV genome. For most purposes it was useful to have a selected set of cosmid clones which spanned the whole genome. The minimum number of clones which together contained all the Sall fragments was five, i.e. 83, 17, 69, 2 and 73. However, it was preferable for at least two reasons to have a selection in which each clone overlapped adjacent clones. This avoided the possibility that apparently adjacent clones were separated by very small Sall fragments (< 200 bp) which had gone undetected. Also, when using cosmids for purposes such as marker rescue, as was done to locate the CpGV iap gene (Crook et al., 1993), it is desirable to use a selection in which all viral genes are intact. A minimal overlap of 2–3 kbp would therefore be required to provide a high likelihood that any gene only partially located on one cosmid would be complete on the overlapping region of the adjacent cosmid. A selection of six or seven cosmids (Fig. 2) was required to achieve this. Cosmids 64, 65, 17, 69, 2 and 73 covered the whole genome with a minimum overlap of 3.5 kbp except at the site between Sall-Q and -W where there was no overlap. Cosmid 61 spanned this site although the overlap across Sall-Q of 1.9 kbp was slightly less than ideal. Many polyhedrin and a few granulin sequences have previously been determined and the data have been used to compile phylogenetic trees (Rohrmann, 1986; Zanotto et al., 1993). Although the number of granulin sequences was very limited, these trees showed that the granulins fell into a distinct group which were more closely related to each other than to the polyhedrins. The CpGV granulin gene sequence fits into this pattern. The CpGV granulin gene sequence is particularly closely related to ClGV granulin with 86% nt identity and 98% aa identity. GV granulin sequence is especially conserved 14 nt region which includes the ATAAG baculovirus late promoter element. In four of these five sequences, an extended region of 23 nt is almost completely conserved. The

![Image](https://example.com/image.png)
ORFs upstream of the granulin genes in CpGV and CIGV are also conserved though with a lower homology, showing 68% identity and 82% similarity at the aa level and 67% nt identity. The encoded polypeptides are not closely related to any other sequences currently on the protein databases but do have a low level of homology and some functional similarity with AcMNPV ME53. It is likely that ArGV also has an ORF similar to ORF124R immediately upstream of granulin.

More general information on the relatedness of CpGV to other baculoviruses was obtained by Southern blot hybridizations with ArGV and AcMNPV under low stringency conditions. This showed that large areas of the two GV genomes had significant homology to each other and that homologous areas were in relatively similar locations on the two genomes. This similarity of genome organization is supported by the finding of related ORFs upstream of the two granulins.

The very limited amount of hybridization with AcMNPV could be because there are many different genes in GVs or because there are related genes but with a level of homology too low to permit hybridization. A recent analysis of the genome of the DNA from three variants. The complete DNA sequence of the DNA from three variants. Journal of General Virology 67, 45–54.

References


Received 5 August 1996; Accepted 13 December 1996