The expansion of a hypervariable, non-\( hr \) ori-like region in the genome of Cryptophlebia leucotreta granulovirus provides \textit{in vivo} evidence for the utilization of baculovirus non-\( hr \) oris during replication

Johannes A. Jehle

State Education and Research Center for Agriculture, Viticulture and Horticulture (SLFA), Section Biotechnological Crop Protection, Breitenweg 71, 67435 Neustadt an der Weinstraße, Germany

In this report a naturally occurring hypervariable region within the genome of different Cryptophlebia leucotreta granulovirus genotypes is characterized. The region consists of a stretch of direct repeats, short palindromes and an unusual AT-rich region. Although the organization of these repeat sequences is unique to baculoviruses, it has the structural features of a ‘non-\( hr \)’ origin of DNA replication (ori). Restriction analysis and Southern hybridization revealed that this region is expanded during virus replication. Sequence comparison of different isolated genotypes indicated that the expansion is caused by concatenation of short repeats within the region or by concatenation of the complete region. These findings indicate that the expansion of non-\( hr \)-origin-like regions is not restricted to defective-interfering particles, as was found previously for baculoviruses propagated in cell culture. Moreover, it appears that non-\( hr \) complexity contributes to the natural heterogeneity and genetic plasticity of baculovirus genomes. Also, circumstantial evidence is discussed that \( hr \) oris might have developed from internal rearrangement and multiplication of a non-\( hr \) ori during baculovirus evolution.

Introduction

Cryptophlebia leucotreta granulovirus (CrleGV, previously ClGV) is a highly specific virus pathogen of the false codling moth, \textit{C. leucotreta}. CrleGV belongs to the family Baculoviridae (genus Granulovirus) and has a genome of double-stranded, circular DNA of about 112–4 kb. So far, three different virus isolates of CrleGV, CV, IC and SA, have been identified in infected larvae from the Cape Verde islands, the Ivory Coast and South Africa, respectively (Fritsch, 1989). All these isolates can be distinguished by small differences in their restriction patterns. The isolate CrleGV CV was biologically and molecularly characterized previously. Different genotypes of this strain were purified and propagated using an \textit{in vivo}-cloning procedure and a detailed physical map was constructed for genotype CV3 (Jehle \textit{et al.}, 1992). Cross hybridization experiments and sequence analyses indicated a close relationship of CrleGV to the \textit{Cydia pomonella} granulovirus (CpGV), which is pathogenic for the codling moth but also infectious to larvae of \textit{C. leucotreta} (Fritsch, 1989; Jehle \textit{et al.}, 1992). Both viruses share overall genome similarity and the highest degree of homology between both viruses was identified in the region of the granulin gene and in a highly conserved area of genes that is homologous to \textit{Autographa californica} multiple nucleopolyhedrovirus (AcMNPV) p6.9-lef5-p38 (Jehle & Backhaus, 1994a, b). These genes are arranged identically in all baculovirus genomes that have been sequenced so far (IJkel \textit{et al.}, 1999).

On the other hand, extensive intragenomic hybridization experiments and partial sequencing of the genome of CrleGV did not provide any evidence for the presence of homologous regions (hrs), which were supposed to be characteristic for many baculovirus genomes. hrs consist of a varying number of palindrome-containing repeats and are dispersed throughout the baculovirus genome. They were shown to be \textit{cis}-acting transcriptional enhancers of early gene expression in AcMNPV and \textit{Orgyia pseudotsugata} NPV (OpMNPV) (reviewed by Lu & Miller, 1997). Secondly, hrs may function as origins of
replication (ori), as suggested by an infection- or transfection-dependent replication of hr-containing plasmids in different cell/virus systems, including AcMNPV, OpMNPV and Spodoptera exigua MNPV (SeMNPV) (Kool et al., 1993; Ahrens et al., 1995; Broer et al., 1998).

Most of the baculovirus genomes sequenced so far contain hrs. The presence of hrs has been described for all lepidopteran and dipteran nucleopolyhedroviruses. However, they were not found in the completely sequenced genomes of CpGV, Adoxophyes orana GV or Neodiprion lecontei NPV (Luque et al., 2001; Wormleaton & Winstanley, 2001; Lauzon et al., 2001).

Another type of putative baculovirus ori is the so-called non-hr ori, which consists of palindromes and direct repeats neighboured by an AT-rich region. Non-hr oris do not contain hr-like sequences but have similar structural characteristics to eukaryotic oris (Kool et al., 1994). Non-hr oris have been characterized in AcMNPV, OpMNPV, SeMNPV and Spodoptera littoralis MNPV (Kool et al., 1994; Pearson et al., 1993; Heldens et al., 1997; Huang & Levin, 1999). Besides their activity in template-dependent replication assays, further evidence that non-hr oris may be involved in genome replication is derived from the finding that serial passage of AcMNPV results in the appearance of defective interfering particles (DIs) (Lee & Krell, 1992; Kool et al., 1991). Dls possess pseudoviral genomes harbouring major deletions and insertions. Dls do not replicate autonomously but need full-length helper virus to be propagated. The DNA of AcMNPV Dls isolated after 81 passages in Sf21 cells contained large reiterated segments of 2-8 kb fragment encompassing the non-hr ori. It was suggested that these Dls have a replication advantage due to their reduced size and increased number of putative oris (Krell, 1996). A similar finding was reported recently by Pijlman et al. (2002) for SeMNPV. In their study, multimers of fragments 2-6 and 3-0 kb containing the non-hr ori of SeMNPV were identified in SeMNPV Dls passed through Se-301 cells.

In this study, a naturally occurring hypervariable region within the genome of CrleGV is described. This region contains head-to-tail concatemers of a 900 bp sequence region that contains a variety of different palindromes and repeats. The structural organization of this area is similar to that of a baculovirus non-hr ori. These findings suggest that the expansion of putative non-hr oris is not restricted to passage in cell culture where artificial selection constraints might favour the generation of Dls. Comparison of the sequences of different CrleGV genotypes suggests an extensive variability of repeated sequences that contributes to the genetic diversity of baculoviruses.

Methods

**Viruses and viral DNA.** Viral DNA used in this study was derived from different genotypes of CrleGV. Genotypes CrleGV CV3, CV3.1 and CV4 were isolated by an in vivo-cloning procedure and viral DNA was purified according to techniques described elsewhere (Smith & Crook, 1988; Jehle et al., 1992).

**Restriction analysis and mapping.** Viral DNAs were digested with restriction enzymes and electrophoresed on 0.7% agarose gels in TAE buffer using standard techniques (Sambrook et al., 1989). Restriction mapping of genotypes CV4 and CV3.1 and of cloned fragments was performed on the basis of the restriction map of CrleGV CV3 by hybridization techniques and/or double digestion. For Southern hybridization, DNA fragments were blotted onto Hybond-N membranes (Amersham). Hybridization was carried out at 42 °C in 50% formamide according to standard methods, followed by chemiluminescent detection using DIG-labelled probes (Roche).

**Cloning and sequencing.** Viral DNAs of the genotypes CV3, CV3.1 and CV4 were digested with EcoRI and electrophoresed. The EcoRI-G fragments of genotypes CV3, CV4 and CV3.1 were isolated from agarose gels and cloned into pGem-7Zf (+). These plasmids were used to generate unidirectional exonuclease deletion clones by the method of Henikoff (1987). These were sequenced from double-stranded templates by the dideoxynucleotide chain-termination method using commercial automated sequencing (Centerprise) on an ABI 377 sequencer. Both strands of the templates were sequenced. Sequence data were compiled using the University of Wisconsin Genetics Computer Group (GCG) software package (Devereux et al., 1984) and Lasergene (DNASTAR).

Results

**Localization of the hypervariable region.**

CrleGV CV3 and CV4 are genotypes isolated previously that were purified by in vivo cloning of the Cape Verde isolate of CrleGV. Comparative restriction analyses between both genotypes indicated only minor differences consisting of small insertion or deletion mutations of not more than 50-100 bp accompanied by the loss of very few restriction sites in one of the genotypes (Fig. 1a, b). A major heterogeneity between both variants was observed when genomic DNA digests where hybridized to the labelled EcoRI-H fragment of CV3. Based on the restriction map of the genome of CrleGV, it was expected that this fragment would hybridize only with itself or with overlapping restriction fragments generated by other restriction enzymes. Instead, this hybridization experiment revealed a series of additional submolar hybridization signals that differed in size by multiples of about 900 bp from the EcoRI-H and Ndel-H fragments (Fig. 2a, lanes 1 and 3). More notably, these bands appeared to be present in decreasing molar ratios with increasing size. No additional submolar bands were observed at similar regular distances from other overlapping Ndel fragments E, M and Q, indicating that the observed variability was restricted to the overlapping region of EcoRI-H and Ndel-H, which was mapped to position 19-9-22-1 kb of the CrleGV genome. When labelled restriction fragments of a plasmid library covering the whole CrleGV genome were used to screen the CrleGV genome for further submolar variation, this region was the only area within the genome where variability was observed (data not shown). In contrast to CV3, hybridization to CV4 DNA digested with EcoRI/Ndel did not show any submolar bands but indicated
Fig. 1. (a, b) Agarose gels of digested DNA of CrleGV CV3 (a) and CV4 (b) using BamHI (lanes 1), EcoRI (lanes 2) and KpnI (lanes 3). (c) Agarose gel of EcoRI digests of isolated DNA of CrleGV CV3 (lane 1) and CV3.1 (lane 2). Single and double asterisks to the right indicate the 9.4 kb EcoRI-C/S fusion fragment and the expanded 6.1 kb EcoRI-H(+1) fragment of CV3.1, respectively. Numbers to the left indicate size in kb.

Fig. 2. (a) Southern blot of DNA digests of CrleGV CV3 and CV4 using labelled fragment CV3 EcoRI-H as probe. Lanes 1, CV3 DNA digested with EcoRI; 2, CV4 DNA digested with EcoRI; 3, CV3 DNA digested with NdeI; 4, CV4 DNA digested with NdeI. Letters and numbers to the right and to the left indicate the hybridizing EcoRI and NdeI restriction fragments and the ladder of submolar bands. (b) Southern blot of EcoRI digest of CV3 (lane 1) and CV3.1 (lane 2) using labelled EcoRI-H of CV3 as probe. Letters and numbers to the right indicate the hybridizing submolar restriction fragments. Size standards of λ DNA digested with HindIII (kb) are given to the left of (a) and (b).

Fig. 3. Restriction map and genetic organization of the EcoRI-H fragments of CV3 EcoRI-H, CV3.1 EcoRI-H(+1) and CV4 EcoRI-H and the corresponding genome area of CpGV. (a) Restriction maps of CV3, CV3.1 and CV4 for NdeI (N), CiaI (C) and BamHI (B). Solid lines indicate the sequenced parts of each fragment. Numbers to the right indicate the length of the sequenced fragments (compare with Fig. 4). Note the amplification of the 900 bp region (CiaI fragment) in CV3.1 EcoRI-H(+1), which is not present in CV3 and CV4 (dotted line). (b) Genetic organization of CV3 EcoRI-H, including ORFs, repeat units and AT-rich repeats identified by sequence analysis. ORFs are drawn as open arrows, repeats are indicated by rectangles. (c) Organization of the corresponding area in the CpGV genome according to Luque et al. (2001). ORFs and repeats are indicated by open arrows and rectangles, respectively. Identical hatching patterns with CrleGV ORFs indicate homologous ORFs.
small size differences of NdeI-H and NdeI-M (Fig. 2a, lanes 2 and 4).

The distinct pattern of submolar signals can be accounted for by the loss or reiterated amplification of an internal 900 bp region, which is indicated by the numbering −1, +1, + 2 in Fig. 2. This observation, and the size differences of the overlapping NdeI-H and NdeI-M fragments between CV3 and CV4, strongly suggested that EcoRI-H contains a hotspot of genomic variability within the CrleGV genome.

In order to analyse the nature of the heterogeneity in CV3, a spontaneous mutant named CV3.1 was identified by restriction analysis of the DNA of a number of different in vitro cloned CV3 genotypes. In the ethidium bromide-stained agarose gel, this mutant lacked the CV3 EcoRI-H fragment of 5.2 kb but contained instead a fragment of 6.1 kb, i.e. about 900 bp larger than the CV3 EcoRI-H fragment (Fig. 1c). The new genotype CV3.1 also contained a fragment of 9.4 kb, which was mapped to be the fusion product of the adjacent fragments EcoRI-C and EcoRI-S, as it is also found in CV4. This variation occurs about 35 kb from EcoRI-H and is not related to the variability within EcoRI-H (data not shown). When the labelled CV3 fragment EcoRI-H was hybridized to the EcoRI digest of CV3.1 DNA, the 6.1 kb fragment was the predominant hybridization signal (Fig. 2b). Again, a ladder of additional bands of increasing molecular mass and decreasing intensity could be observed, suggesting that CV3.1 is a variant in which EcoRI-H is mostly replaced by a fragment termed EcoRI-H(+1), which contains one additional reiteration of about 900 bp. In overexposed Southern blots, up to eight of these additional submolar bands were detectable.

Cloning, fine mapping and sequencing the hypervariable region

The conclusion that fragments EcoRI-H of CV3 and CV3.1 contain submolar multiple reiterations of an internal region, which was based on the distinct pattern of submolar hybridization bands, was proved further by fine mapping of the genotypes. EcoRI-H of CV3 and CV4, along with the collinear 6.1 kb fragment of CV3.1, were isolated from agarose gels and cloned into pGem-7Zf(+). Comparative restriction analyses showed that the major heterogeneity between these fragments was located between an NdeI restriction site at map position 19.9 kb and the right-hand EcoRI site at 22.1 kb. Within this area, CV3, CV3.1 and CV4 showed variation in the number and position of Clal sites. The EcoRI-H fragment of CV3.1 differed from that of CV3 by an additional Clal fragment of about 900 bp. When viral DNA of CV3 and CV3.1 was digested with Clal and hybridized to labelled CV3 EcoRI-H, the ladder-like submolar hybridization signals were replaced by a predominant supemolar signal of 900 bp, suggesting that all reiterations felled together into the proposed 900 bp repeat (data not shown).

For a detailed analysis of the hypervariable region, CV3 EcoRI-H and portions of CV3.1 EcoRI-H and CV4 EcoRI-H, including the hypervariable region, were sequenced (Figs 3 and 4). CV3 EcoRI-H had a length of 5161 bp and contained two open reading frames (ORFs) longer than 50 amino acids (ORF 348 and ORF 427) within the conserved part and a stretch of unusual short direct repeats, short palindromes and AT-rich repeats encompassing more than 1770 bp within the hypervariable part of the fragment. ORF 348 is oppositely oriented to the granulin gene of CrleGV and has a length of 954 bp. It appears to be a homologue of pe38, identified previously in AcMNPV, OpMNPV (ORF pe38) and CpGV (ORF 24) (Krappa & Knebel-Moersdorf, 1991; Theilmann & Stewart, 1992; Luque et al., 2001). Amino acid domains characteristic for PE38 proteins, such as a putative Q domain, C4HC4 RING finger domain and a leucine zipper, as well as an acidic and a basic domain, were found within the predicted amino acid sequence of ORF 348. AcMNPV pe38 is an immediate early-transcribed gene and might be involved in transcriptional regulation of late AcMNPV genes.

Upstream of ORF 348, ORF 427 also contains a C4HC4 RING finger domain and a putative leucine zipper, indicating that this gene might also be involved in DNA binding and, possibly, transcriptional regulation. The only known baculovirus homologue of ORF 427 occurs in the closely related CpGV, where it was identified as ORF 27, in close neighbourhood to the pe38 homologue of CpGV (ORF 24). In contrast to the CpGV genome, no homologues of CpGV ORF 25 and ORF 26 were present in the sequenced region of CrleGV (Fig. 3). The predicted translational start region of ORF 427 overlaps with a copy of a short palindromic repeat (r1p, nt 3325–3350 of CV3), which is found again about 105 nt upstream (r1p, nt 3457–486 of CV3).

These two repeats mark the left border of a stretch of short direct repeat and palindromic sequences that are present in all analysed genotypes. The extensive repeat area has a length of 1796 bp for CV3, 2729 bp for CV3.1 and 1707 bp for CV4. The different direct repeats (r2, r3, r4, r5, r6, r9 and r10), unique palindromes (p1 and p2) and repeated palindromes (r1p, r7p and r8p) are indicated in Fig. 4. As shown in Table 1, the repeats r1p, r2, r3 and r4, as well as r5, r6 and r7, contain common consensus motifs. The repeat r5 is composed of two similar sequences, r5I and r5II. Similarly, r6 is composed of two homologous repeats, r6I and r6II. The repeats r9 and r10 are AT-rich repeats of 26 and 21–24 bp lacking any C or G residues. In the case of CV4, the reiteration of these unusual AT repeats results in a sequence stretch of 419 bp without any C or G residues (Fig. 4). A common feature of all three genotypes is the invariable presence of at least two motifs composed of r6–r7p–r4, which are always found to the left and right of a variable number of r2 repeats. In contrast to CV3 and CV3.1, CV4 contains two additional palindromic sequences p1 and p2. Most intriguingly, CV4 lacks a stretch of 11 r2 repeats, which are interspersed in
Non-hr ori-like region of CrleGV

Fig. 4. For legend see page 2030.
Fig. 4. Sequence of the hypervariable area of EcoRI-H of CV3 and CV4 and EcoRI-H(+1) of CV3.1. The putative start codon of ORF 427 is indicated. The repeated 900 bp region of CV3.1 is aligned with its parental sequence. The different repeat types are given at the bottom of the alignment; those marked with an asterisk indicate truncated repeats.

Table 1. Nucleotide sequences of direct repeats and palindromic repeats within the hypervariable region of CrleGV

Left and right parts of palindromic sequences are single and double underlined, respectively. The consensus sequence is indicated in bold.

<table>
<thead>
<tr>
<th>Repeat name</th>
<th>Sequence</th>
<th>Length (bp)</th>
<th>Number of repeats in different genotypes (CV3/CV3.1/CV4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1p</td>
<td>AATCTAAACATTGCTCGGTTACATTTAGAAGAATTTAAAAC</td>
<td>36</td>
<td>3/3/2</td>
</tr>
<tr>
<td>r2</td>
<td>GGGGACTTAGAATAATTCACCCTAACC</td>
<td>28</td>
<td>16/31/5</td>
</tr>
<tr>
<td>r3</td>
<td>AGACACTGGAAAAATTCACAGTC</td>
<td>24</td>
<td>5/7/5</td>
</tr>
<tr>
<td>r4</td>
<td>CAGGACTTAGAATAATTCGCAG</td>
<td>25</td>
<td>5/8/4</td>
</tr>
<tr>
<td>Consensus</td>
<td>. RR . AYTRGA . AATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r51</td>
<td>TGACTGATAC . GATAG</td>
<td>16</td>
<td>2/3/2</td>
</tr>
<tr>
<td>r5II</td>
<td>GTGATAT . AGATAG</td>
<td>13</td>
<td>2/3/2</td>
</tr>
<tr>
<td>r6I</td>
<td>CGATACG . . AGTCAGTGGGGGCCC</td>
<td>21</td>
<td>3/4/3</td>
</tr>
<tr>
<td>r6II</td>
<td>ATGCAGA . AGTCAGTGGGGGCCC</td>
<td>21</td>
<td>1/1/1</td>
</tr>
<tr>
<td>Consensus</td>
<td>. YGAYTRC . ATAGTCAGTGGGCCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r7p</td>
<td>CGACTGACGTCAGA . . AGTGA</td>
<td>23</td>
<td>6/10/5</td>
</tr>
<tr>
<td>Consensus</td>
<td>. . RR . AYTRGA . AATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r8p</td>
<td>GTGTCACTTTGACAGGGTCTG</td>
<td>20</td>
<td>3/4/3</td>
</tr>
<tr>
<td>r9</td>
<td>TTTTTATTTATATTTATTTTTTTTTAA</td>
<td>26</td>
<td>9/9/10</td>
</tr>
<tr>
<td>r10</td>
<td>TTTATTTATTTTTATA (3) 6-10A</td>
<td>21–24</td>
<td>5/5/5</td>
</tr>
<tr>
<td>p1</td>
<td>GTAGGCCTGACGCGGCGCGGTACG</td>
<td>23</td>
<td>0/0/1</td>
</tr>
<tr>
<td>p2</td>
<td>HATCAGGCTGCTGGCTGCC</td>
<td>20</td>
<td>0/0/1</td>
</tr>
</tbody>
</table>

Fig. 5. Illustration of the hypervariable repeat region of CrleGV CV3, CV4 and CV3.1. Note that the expansion of EcoRI-H(+1) of CV3.1 is caused by a head-to-tail duplication of the 900 bp region.
CV3 and CV3.1 between two r7p–r4–r3–r7p–(r4) motifs. It appears that r2 repeats are expanding in CV3 or that the r2 repeat stretch is deleted in CV4 (Figs 4 and 5).

Expansion model of the hypervariable region of CrleGV

The arrangements of the repeats in CV3 and CV3.1 are identical. They differ only in a head-to-tail tandem of exactly 909 bp (called the 900 bp region), which encompasses the central part of the CV3 repeat structure (Figs 4 and 5). All other repeats, number of repeats and nucleotide positions are identical. This shows that the size difference between CV3 EcoRI-H and CV3.1 EcoRI-H(+1) is generated by duplication of an internal region, as was suggested by the hybridization data. From the sequence comparison of CV3 and CV3.1, it was concluded that the ladder of submolar bands in CV3 and in CV3.1 is the result of a concatenation of the 900 bp region that was found to be duplicated in CV3.1. This observation strongly suggests that the 900 bp region is amplified in submolar populations of CV3 and CV3.1, which finally resulted in the expansion of this region, as indicated by the hybridization data.

Discussion

In this study, a hypervariable region within the EcoRI-H fragment of three genotypes of CrleGV, CV3, CV3.1 and CV4, was identified. This hypervariable region is characterized by restriction site heterogeneity as well as the presence of a family of submolar bands, each differing in size by about 900 bp, for the two genotypes CV3 and CV3.1 (Figs 1 and 2). The observed regularity in size difference and the decreasing molar ratio with increasing size of the submolar bands indicated that a specific molecular mechanism, rather than random insertions, is involved in the generation of these bands.

Analyses of the cloned EcoRI-H fragments showed a more or less invariant region with similar restriction sites and a very variable region indicated by the presence of different Clal sites (Fig. 3). The invariant region of CV3 and the variable regions of CV3, CV3.1 and CV4 were sequenced and compared. The invariant region of EcoRI-H encompasses the nucleotide positions 1–3478 of CV3 EcoRI-H, it covers two ORFs, ORF 348 and ORF 427. GenBank analyses suggested that ORF 348 is a homologue of pe38 of AcMNPV, which is possibly involved in transcriptional regulation (Krapp & Knebel-Mörsdorf, 1991; Theilmann & Stewart, 1992). Interestingly, sequencing of different baculovirus genomes revealed that pe38 homologues are not present in group II NPVs and many GV [e.g. Lymantria dispar NPV, Helicoverpa armigera SNPV (HaSNPV), H. zea SNPV, SeMNPV, Xestia c-nigrum GV or Plutella xylostella GV], suggesting a phylogenetically scattered and polyphyletic distribution of this gene within baculoviruses (Herniou et al., 2001; Chen et al., 2001). ORF 427 also contains putative DNA-binding domains, such as a RING finger and a leucine zipper, suggesting some role in transcriptional regulation. So far, this ORF appears to be unique to CrleGV and CpGV, in which a homologue was identified as ORF 27 (Luque et al., 2001).

Within EcoRI-H of CrleGV CV3, some similarity to the gene arrangement of CpGV (genome position 18 500–22 000) was noted (Fig. 3). In both viruses, the CrleGV ORF 348 and ORF 427 and the CpGV ORF 24 and ORF 27 are in close vicinity. However, CpGV ORF 24 and ORF 27 are separated by two very short overlapping ORFs, ORF 25 and ORF 26, which were not identified in CrleGV. For CrleGV, these two ORFs may be absent or located at a different genomic region. Since these two ORFs are rather short and overlap with repeat structures, it is also possible that these ORFs are not expressed in CpGV.

The variable region of CrleGV consists of a stretch of short direct repeats, palindromes and AT-rich repeats. As shown in Table 1, some of these repeats can be considered as a modulation of a common sequence motif. These features resemble the non-hr oris found in AcMNPV, OpMNPV and SeMNPV (Kool et al., 1994; Pearson et al., 1993; Heldens et al., 1997). However, the complexity and the distinct arrangement of repeats and palindromes in CrleGV are very distinct from the corresponding region of the closely related CpGV and from any other baculovirus. It is noteworthy that no hr regions were identified for CpGV upon sequencing of the whole genome (Luque et al., 2001). This is similar to the situation in CrleGV, where genome mapping and preliminary genome sequencing did not indicate any evidence for the presence of hr-like regions interspersed throughout the genome (Jehle et al., 1992; unpublished data). The only genome region of CpGV that was suggested to be a potential non-hr orire comprises ORF 25, ORF 26 and the 3′ end of ORF 27 (Fig. 3) (Luque et al., 2001). This is in very close neighbourhood to the corresponding genome location of the repeat area in CrleGV identified in this study.

The three genotypes CV3, CV3.1 and CV4 contain the same repeat motifs in slightly different compositions. The major difference between CV3 and CV4 is the iterative insertion or deletion of a stretch of 11 copies of r2 repeats, which are present in CV3 but not in CV4. The genotype CV4 appeared to be the predominant genotype in the field isolate of CrleGV CV (Cape Verde islands) (Fritsch, 1989). Hence, CV3 and from that CV3.1 might harbour more deviated forms of the genome. For CV3.1, it appeared that a unit of exactly 909 bp (900 bp region) within the variable region became duplicated, resulting in an increase in the size of CV3 EcoRI-H to CV3.1 EcoRI-H(+1). On the basis of the hybridization, mapping and sequencing data, it is concluded that each step of the observed ladder of submolar bands contains an additional 900 bp region as it is found to be duplicated in CV3.1 EcoRI-H(+1). Since up to eight ladder steps were identified on overexposed Southern blots, the genome of the largest of these submolar genotypes may be increased by more than 7200 bp. It appears that this repeat region has the ability to
undergo rapid and extensive expansion under natural selection during the replication of CrleGV in larvae of *C. leucotreta*. It is noteworthy that no submolar bands were found for CV4, suggesting that the sequence expansion within EcoRI-H of CV3 and CV3.1 might need some minimum essential sequences that are not present in CV4.

Submolar bands forming a ladder of a decreasing molar ratio with increasing size have been identified previously in AcMNPV DIs. Lee & Krell (1992) reported the generation of AcMNPV DIs after 81 serial passages in Sf21 cells, which mainly consisted of concatenations of a 2.8 kb region identified later to harbour the non-*hr* ori of AcMNPV. Similarly, specific amplification of the non-*hr* ori of SeMNPV was observed when SeMNPV was propagated in different cell lines of *S. exigua* (Pijman et al., 2002). However, submolar concatamers of DNA are not always related to non-*hr* oris of replication. A temperature-sensitive mutant of AcMNPV, termed *tsb8*, contained concatamers of a 1.35 kb EcoRI fragment within EcoRI-P (map unit 0.89–0.90). Although this fragment does not contain a *hr* or a non-*hr* ori of AcMNPV, it is worth mentioning that it is adjacent to *hr5* and the non-*hr* ori located at map units 0.87–0.88 and 0.85–0.86, respectively. On the other hand, variation of the repeat number within *hr* regions has been observed for several viruses passed in cell culture or isolated from wild-type virus populations, e.g. *Choristoneura fumiferana* MNPV (Arif & Doerfler, 1984), *Anticarsia gemmatalis* MNPV (Garcia-Maruniak et al., 1996) and SeMNPV (Munoz et al., 1999).

The formation of DIs as well as plasmid-dependent replication assays in cell culture provided different lines of evidence that the non-*hr* oris of AcMNPV is most probably involved in DNA replication (reviewed by Lu et al., 1997). In an alternative approach, the utilization of the non-*hr* ori for AcMNPV replication was also suggested by competitive PCR studies, which demonstrated the increased amount of non-*hr* oris compared to a non-ori control sequence during AcMNPV replication in cultured insect cells (Habib & Hasnain, 2000). The observation that a non-*hr* ori-like region is expanding in a DI-like manner during replication of CrleGV in host larvae strongly supports the hypothesis that non-*hr* oris are also utilized during in vivo replication.

Sequence comparisons of the different CrleGV genotypes provide a complex picture of repeat sequence diversity that is unique to baculoviruses. Whereas the complexity and different kinds of repeats and palindromes in CV4 most resemble the non-*hr* oris of replication of other baculoviruses, the loss of repeat diversity (*p1* and *p2*) and the repetitive multiplication of a single repeat (*r2*) is observed in CV3. This slightly more homogeneous repeat structure, predominantly consisting of the repetition of a specific repeat unit, is more reminiscent of the structure of a baculovirus *hr* origin. In CV3.1, finally, it can be observed that one complete CV3 repeat unit is concatenated to form multiple tandem copies. This observation reveals the question of whether non-*hr* oris and *hr* oris have the same origin or whether *hr* oris derived from non-*hr* oris during baculovirus evolution. Indeed, there are some circumstantial evidences that might support this view. First, it was noted earlier that AcMNPV *hr* oris and non-*hr* oris have at least some common motifs: e.g. the non-*hr* repeat 5′ GATGATGTCAT 3′ can also be found in many *hr* repeats (Kool et al., 1994). In comparison to AcMNPV, the *hr* oris and non-*hr* oris of SeMNPV consist of a completely different nucleotide sequence and repeat arrangements. However, some repeats within the SeMNPV non-*hr* oris are very similar to the consensus sequence of the SeMNPV *hr* repeats (Fig. 6). These similarities of *hr* and non-*hr* sequences might reflect similar functional sequence requirements, such as binding of proteins involved in DNA replication or might be footprints of a shared phylogenetic history for *hr* and non-*hr* oris.

The variability in number of homologous repeats and internal rearrangement seem to be characteristic features of *hr* regions (Arif & Doerfler, 1984; Garcia-Maruniak et al., 1996; Munoz et al., 1999). The formation of concatamers might also favour the spread of *hr* oris within a genome at different loci. This spread could be triggered by translocation or transposition of repeat units. Chen et al. (2001) suggested that *hr2* and *hr3* of HaSNPV are the result of the insertion of two baculovirus-repeated ORF (*bro*) genes into what originally may have been a single *hr* region. A further property of some repeat structures, e.g. mini-satellites, is their ability to transpose provided that specific trans-acting factors are present (Hankeln et al., 1994). Hence, expansion of repeat regions or concatenation of repeat sequences might initiate their translocation or transposition to other genome regions where they could establish additional oris. It is not unlikely that this expansion of origins results finally in multi-ori genomes. The presence of more than one ori might provide some replicative advantage for large DNA genomes compared to genomes with a single ori, in terms that multi-ori genomes are less sensitive to single origin mutations and that these genomes can initiate replication from more than one genome loci, which might result in an accelerated speed of replication.

**Fig. 6.** Alignment of the consensus sequence of *hr* and repeats of the non-*hr* oris of replication of SeMNPV according to Heldens et al. (1997) and Munoz et al. (1999). Shaded nucleotides indicate identity. The consensus of *hr* oris is most probably established additional oris. It is not unlikely that this expansion of origins results finally in multi-ori genomes. The presence of more than one ori might provide some replicative advantage for large DNA genomes compared to genomes with a single ori, in terms that multi-ori genomes are less sensitive to single origin mutations and that these genomes can initiate replication from more than one genome loci, which might result in an accelerated speed of replication.
This research was initiated during my research stay at the Federal Institute for Biological Research for Agriculture and Forestry (BBA, Braunschweig, Germany). The initiative contributions of Horst Backhaus, Jürgen Samunskus and Antje Nickle (BBA Braunschweig, Germany) are gratefully acknowledged. I thank Claudia Vogel (SLFA Neustadt) for excellent technical assistance. I am indebted to Eva Fritsch (BBA Darmstadt, Germany) for providing viruses and viral DNA. The excellent comments by the journal’s reviewers, which helped to improve this manuscript, are gratefully acknowledged. This research was supported by a grant of the BMBF (0319306A) and by a grant of the DFG (JE245-4).

References


Received 18 December 2001; Accepted 28 March 2002