Analysis of the C-polyhedrin genes from different geographical isolates of a type 5 cytoplasmic polyhedrosis virus

Mark S. Galinski, Yang Yu, Beverly R. Heminway and George S. Beaudreau

The C-polyhedrin genes of two different geographic isolates of a type 5 cytoplasmic polyhedrosis virus (CPV) were cloned. A CPV infecting Orgyia pseudotsugata (OpCPV), isolated in the Pacific Northwest of the U.S.A., and a CPV infecting Heliothis armigera, isolated in South Africa, were studied. Both genes were found to be 883 nucleotides in length and encoded a predicted protein of 246 residues (M_r of 28890). Comparison of the nucleotide sequences of these two viruses with another type 5 geographic isolate, infecting Euxoa scandens (EsCPV; isolated in Eastern Canada), showed that there were only 17 nucleotide differences among the three genes. The only nucleotide variation that had an effect on the encoded protein was a deletion of nucleotide 774 in the gene of EsCPV. The deletion introduces a frameshift mutation resulting in the alteration of the carboxyl-terminal amino acid sequence. Sequence alignment of the OpCPV C-polyhedrin showed little homology to a type 1 CPV (infecting Bombyx mori) or with analogous proteins (N-polyhedrins) from two baculoviruses infecting O. pseudotsugata. Interestingly, most of the conserved residues between the N- and C-polyhedrins were either basic or aromatic amino acids.

Introduction

The cytoplasmic polyhedrosis virus (CPV) group, one of the genera classified within the Reoviridae family, is associated with inclusion-type diseases of insects (Payne & Mertens, 1983; Brown, 1989; Belloncik, 1989). These viruses are characterized by the production of large proteinaceous crystals in the cytoplasm of infected cells. The crystals contain a single protein, C-polyhedrin (25K to 31K), within which the mature virions are embedded. The viral genomes are composed of 10 discrete equimolar dsRNA segments. The electrophoretic separation of the genomic RNA has provided a means for classifying members of this group into 12 distinct classes (Payne & Rivers, 1976; Payne et al., 1977; Payne & Mertens, 1983; Mertens et al., 1989). Although the electrophoretic patterns of the genomic dsRNAs and virion-associated proteins are evidence for the diversity of this group of viruses, there are few reports on the sequence homologies of the various CPV types. Payne et al. (1978) examined the sequence homologies among three type 1 CPVs including Dendrolimus spectabilis CPV (DsCPV), Lymantria dispar CPV (LdCPV) and Bombyx mori CPV (BmCPV) and found that DsCPV and LdCPV were the same and each of these viruses was 52 to 76% identical with BmCPV. Galinski et al. (1983) reported no detectable homologies among type 1 (BmCPV), type 5 (Orgyia pseudotsugata CPV; OpCPV) or type 8 (Manduca sexta CPV) viruses or the human reovirus type 1. In contrast, there was significant sequence homology among three different geographical isolates of the type 5 virus, OpCPV, Euxoa scandens CPV (EsCPV) and Heliothis armigera CPV (HaCPV). OpCPV and EsCPV shared 43 to 44% of the sequence, and each was 6 to 13% identical with HaCPV. More recently, the nucleotide sequences of the C-polyhedrin genes from BmCPV (Arella et al., 1988) and EsCPV (Fossiez et al., 1989) have been reported. Analysis of these genes confirmed the previously reported differences in sequence homology between the type 1 and type 5 CPVs at the molecular level for this specific gene.

The work described in this paper was undertaken to determine the sequence similarities of the C-polyhedrins from three type 5 CPVs. These viruses were isolated from different hosts and in different geographical locations. The OpCPV was isolated in the Pacific Northwest of the U.S.A., EsCPV was isolated in Eastern Canada and HaCPV was isolated in South Africa. The data presented here demonstrate the high degree of sequence conservation of the polyhedrin gene for three type 5 CPV isolates.

Methods

Virus sources. EsCPV, a gift from S. Belloncik (Universite du Quebec a Trois-Rivieres, Canada) was originally described by Quiot &
Belloncik (1977) and further characterized by Grancher-Baray et al. (1981). HaCPV, a gift from R. Rubinstein (Medical School, Observatory, South Africa) was originally described by Rubinstein et al. (1976). OpCPV was originally described by Martignoni et al. (1969). BmCPV was a gift from C. C. Payne (AFRC Institute of Horticultural Research, Littlehampton, U.K.).

EsCPV, HaCPV and OpCPV were propagated in O. pseudotsugata larvae as previously described by Galinski et al. (1983). The genomic RNA was purified as described by Payne & Rivers (1976). Electrophoresis of the dsRNA was performed in 4.5% polyacrylamide-urea gels as described by Maxam & Gilbert (1980). Purification of the OpCPV polyhedrin gene (segment 10) was performed following electrophoresis of the genomic dsRNA on a low melting point agarose gel as described previously (Galinski et al., 1982).

Cloning and sequence analysis. Two synthetic oligodeoxyribonucleotides were used to determine the flanking (non-coding) sequences at the 5' and 3' ends of the OpCPV polyhedrin gene by direct dideoxynucleotide sequence analysis of the genomic RNA (Galinski et al., 1988). The sequences used for designing these primers were taken from the protein coding sequences reported for the EsCPV C-polyhedrin gene (Fossiez et al., 1989). Confirmation that the flanking ends of the gene were conserved provided the information for the design of two additional primers used for the cloning of the C-polyhedrin genes.

The OpCPV and HaCPV C-polyhedrin genes were cloned following cDNA synthesis and PCR amplification of the first-strand cDNA. Purified genomic dsRNA was used as a template for cDNA synthesis using two synthetic oligodeoxyribonucleotide primers, Op + (5' AGGGATCCAGTGTAAAATCTTACGAGGA 3') and Op− (5' AGGGATCCGCACTCAACTCTTGAACA 3'), which were identical to the ends of the plus and minus strands, respectively. These primers contained the exact 5' and 3' ends of the gene were highly conserved. Using this information, full length clones of the C-polyhedrin genes from OpCPV and HaCPV were generated. The sequence analysis of the three type 5 genes is summarized in Fig. 2. The OpCPV and HaCPV C-polyhedrin gene is 883 nucleotides in length and encodes a protein of 253 amino acids. These sequences are in agreement with the previously reported sequence for EsCPV (Fossiez et al., 1989). There are 17 nucleotide differences between the three genes. Of these, 16 are base transitions and one is a deletion. The single nucleotide deletion occurs in the

Results

Electrophoretic separation of the genomic RNAs from BmCPV, EsCPV, HaCPV and OpCPV using a 4.5% polyacrylamide gel is shown in Fig. 1. This shows that the three type 5 CPVs have remarkably similar electrophoretic profiles and that these profiles are clearly distinguishable from that of BmCPV. In addition, the HaCPV contains a small RNA that is not present in the other type 5 CPVs. Although the exact nature of this small RNA is not currently known, it is probably not a deletion mutant of the C-polyhedrin gene as has been reported in BmCPV (Arella et al., 1988) since it was not amplified during PCR (data not shown).

Direct dideoxynucleotide sequence analysis of the OpCPV C-polyhedrin gene confirmed that the flanking 5' and 3' ends of the gene were highly conserved. Using this information, full length clones of the C-polyhedrin genes from OpCPV and HaCPV were generated. The sequence analysis of the three type 5 genes is summarized in Fig. 2. The OpCPV and HaCPV C-polyhedrin gene is 883 nucleotides in length and encodes a protein of 253 amino acids. These sequences are in agreement with the previously reported sequence for EsCPV (Fossiez et al., 1989). There are 17 nucleotide differences between the three genes. Of these, 16 are base transitions and one is a deletion. The single nucleotide deletion occurs in the
limited using gel-purified segment 10 as a source of RNA. The analysis are shown in Fig. 3 and indicate that the clones contained authentic C-polyhedrin genes, codon and none of the encoded amino acids are affected. Variant nucleotides are found in the third position of the protein-coding sequences. Interestingly, all of the alterations, all are base transitions and 15 are located in the last seven amino acids and append 16 additional residues onto the carboxyl terminus of the EsCPV gene (OpCPV nucleotide 774) and results in a coding sequence is shown below the OpCPV C-polyhedrin sequence.

EsCPV gene (OpCPV nucleotide 774) and results in a frameshift. The alteration of the reading frame changes the last seven amino acids and appends 16 additional residues onto the carboxyl terminus of the EsCPV C-polyhedrin protein. The authenticity of the additional A residue in both the OpCPV and HaCPV was determined by dideoxynucleotide sequencing of the genomic dsRNA using the Op- primer. Of the remaining 16 alterations, all are base transitions and 15 are located in the protein-coding sequences. Interestingly, all of the variant nucleotides are found in the third position of the codon and none of the encoded amino acids are affected.

To confirm that the recombinant HaCPV and OpCPV clones contained authentic C-polyhedrin genes, in vitro transcription and translation reactions were performed to prepare [35S]methionine-labelled proteins. Authentic radiolabelled OpCPV C-polyhedrin protein was prepared using gel-purified segment 10 as a source of RNA. The radiolabelled proteins were gel-purified and subjected to limited S. aureus V8 proteolysis. The results of this analysis are shown in Fig. 3 and indicate that the recombinant clones encode a protein that is indistinguishable from the authentic OpCPV C-polyhedrin protein.

Pairwise sequence alignment between the OpCPV and BmCPV polyhedrins was performed to determine whether there were any identifiable similarities between the two proteins. Fig. 4(a) shows that there is little detectable sequence identity (17%, 45 out of 269 amino acids) between the two C-polyhedrins, and that any sequence similarity (29%, 78 out of 269) was evenly distributed throughout the proteins. These results are in agreement with the absence of nucleotide sequence homology previously reported for these viruses (Galinski et al., 1983).

Further analyses were performed to examine whether two baculovirus N-polyhedrins shared any similarities with the OpCPV C-polyhedrin. The proteins of these viruses were selected because they have been reported to infect O. pseudotsugata larvae naturally. The N-polyhedrin sequences were from a single nucleocapsid...
nuclear polyhedrosis virus (OpSNPV) and a multiple nucleocapsid nuclear polyhedrosis virus (OpMNPV). Alignment of these proteins (Fig. 4b) shows that there is little sequence identity (12%, 34 out of 275 amino acids). However, sequence similarities (24%, 66 out of 275) across the entire protein lengths are quite striking. Careful inspection of the conserved amino acids revealed that the positions of nearly half (14 out of 33) of the aromatic amino acids are maintained in all proteins. Since aromatic amino acids constitute 13% of the total, this conservation suggests a biological significance. Although not as striking, the conservation of basic residues is notable.

Comparison of the N-polyhedrin sequences has shown that these proteins are well conserved. Although these proteins may vary in amino acid sequence by up to 50% of their residues, their hydrophilicity plots are remarkably similar (Rohrmann, 1986). It would appear that much of the variation in these proteins is attributable to substitution for similar amino acids. Fig. 5 shows the hydrophilicity plots for BmCPV, OpCPV, OpSNPV and OpMNPV polyhedrin proteins. This shows that although the two N-polyhedrins have similar hydrophilicity plots, there is no apparent similarities between the C- and N-polyhedrins. Further, the C-polyhedrins appear remarkably different from one another.

### Discussion

The occluded viruses that are pathogenic for insects include the CPV group, the nuclear polyhedrosis viruses (Baculoviridae) and entomopoxviruses (Arif, 1984). The genomes of the latter two virus groups consist of large dsDNA molecules. The N-polyhedrins are similar in size to the C-polyhedrins, but the entomopoxviruses are occluded within a protein with an $M_r$ of about 120000-termed spheroidin (Banville et al., 1992). The maintenance of occlusion proteins in such a diverse group of viruses indicates the importance of polyhedrins in the life cycles of these viruses.

Previously we have shown that the three type 5 viruses examined in this study had no detectable sequence homologies with a type 1 or 8 CPV, using SI nuclease protection assays (Galinski et al., 1983). The results presented in this study confirm this analysis and demonstrate that the C-polyhedrins are significantly different not only at the nucleotide level but also at the amino acid level. The divergence of these proteins can also be seen in the hydrophilicity plots of the proteins (Fig. 5). Unlike the N-polyhedrins of baculoviruses, the relative distribution of hydrophilic and hydrophobic residues are remarkably different in the C-polyhedrins, suggesting that these proteins have changed significantly during the course of evolution. Whether other CPV types are as dissimilar in sequence is an area that needs further investigation.

In contrast to the intertypic differences, the type 5 C-polyhedrins appear to be remarkably similar. With the exception of the deletion of nucleotide 774 in the EsCPV gene, there are no differences among the three encoded proteins. These results confirm the high degree of sequence homology previously reported for these viruses (Galinski et al., 1983). In addition, since these viruses have been shown not to be identical, sequence analysis of other genomic segments will undoubtedly reveal differences in amino acid sequences among other proteins. Previous studies have shown that the OpCPV and EsCPV were 43 to 44% identical to each other, whereas each was 6 to 13% identical to the HaCPV, using liquid hybridization techniques and SI nuclease analysis (Galinski et al., 1983). Although the reported degrees of homology contrast with the data shown in this report, the minimum sequence identity necessary for hybridization was estimated at 70% under non-stringent conditions. Furthermore, analysis was performed using total genomic dsRNA, of which the C-polyhedrin gene accounts for less than 10% by mass. Thus, the apparent
C-polyhedrin genes of type 5 CPVs

relatedness of the C-polyhedrin genes would have been underestimated. The results reported here for the intertypic differences, together with the intratypic conservation, support the current classification of CPV groups based upon the electrophoretic migration of the genomic dsRNA (Payne & Rivers, 1976; Payne et al., 1977; Payne & Mertens 1983; Mertens et al., 1989).

Comparison among the three viruses infecting O. pseudotsugata shows that the three proteins share little amino acid homology except for the relative placement of aromatic amino acids (phenylalanine, tryptophan and tyrosine). The aromatic side chains have $pK_a$ values that are quite basic (9.11 to 10.78). Additional residues that are conserved are the basic lysine and arginine ($pK_a$ values of 8.95 to 12.48). It is intriguing to suggest that the distribution of these residues is related to the alkaline solubility properties of polyhedrins. Since release of infectious virus from the occlusion body, within the alkaline milieu of the insect midgut, is a critical step in virus replication, conservation of amino acids with alkaline $pK_a$ values might be expected.

Polyhedrins have received considerable attention because of the high level of expression that is normally attained in infected cells. Because of the robust expression of this protein, the regulatory elements of baculovirus N-polyhedrin genes have provided an invaluable tool for the construction of recombinant genes for expression of heterologous proteins. Unfortunately, similar strategies have not been applied to C-polyhedrins. Whether CPVs can be harnessed to perform similarly is an area that needs continued study.

References


(Received 13 December 1993; Accepted 14 March 1994)