Isolation and characterization of a member of the cysteine-rich gene family from *Campoletis sonorensis* polydnavirus

Li Wang Cui and Bruce A. Webb*

Department of Entomology, University of Kentucky, Agricultural Science Building North, Lexington, KY 40546-9100, USA

The endoparasitic wasp *Campoletis sonorensis* injects a symbiotic polydnavirus into its host *Heliothis virescens*. Viral gene expression protects the wasp egg and larva from encapsulation by host haemocytes. Three related *C. sonorensis* polydnavirus (CsPDV) genes, which are expressed in parasitized *H. virescens*, have been previously isolated and grouped into a cysteine-rich gene family. In this report, a CsPDV gene encoding an abundant 1.4 kb mRNA expressed in parasitized insects was isolated and mapped to viral segment V (15.2 kb) by Southern blotting and PCR. The VHvl.4 eDNA is 1338 bp long and has an ORF that encodes 322 amino acids with two complete and one partial cysteine motifs. Similar to other characterized CsPDV cysteine motifs, the VHvl.4 motifs are also characterized by six cysteines at conserved positions and variable inter-cysteine amino acids. DNA sequence analyses show that the VHvl.4 gene shares regions of significant identity (73–97%) with the VHvl.1 gene, a member of the cysteine-rich gene family. The VHvl.4 and the VHvl.1 proteins are 62% identical overall; at the N termini including the signal peptide and the N-terminal cysteine motif the identity is greater (90%). Unlike other CsPDV cysteine-rich genes, the VHvl.4 cDNA has two conserved domains (77% identical in nucleotides, 55% identical in amino acids) that presumably result from the duplication of a portion of the gene. The VHvl.4 gene has four introns with splicing sites located at positions similar to VHvl.1 introns. Introns 2 and 3, located in the first and second domains respectively, have greater identity (97%) than the flanking exon sequences (77%). We propose, based on the evidence presented in this paper, that the VHvl.4 gene is a new member of the cysteine-rich polydnavirus gene family.

Introduction

For successful parasitization, endoparasitic wasps from the families Ichneumonidae and Braconidae employ factor(s) derived from the female reproductive tract to suppress encapsulation by the insect cellular immune response. Factors suppressing encapsulation include venom, ovarian proteins, and a group of symbiotic viruses, the polydnaviruses. Polydnaviruses are characterized by polydisperse, double-stranded, superhelical (SH) DNA genomes (Stoltz et al., 1984, 1995). Two morphologically distinctive groups of polydnaviruses are recognized: the bracoviruses found only in braconid wasps, and the ichnoviruses associated only with the ichneumonid wasps (Stoltz & Whitfield, 1992; Stoltz et al., 1995). To date, *Campoletis sonorensis* polydnavirus (CsPDV), the prototype of ichnoviruses, has been characterized to the greatest extent at the molecular level. CsPDV consists of 28 or more SH DNA segments ranging from 6 to 21 kb (Krell et al., 1982; Blissard et al., 1986a). CsPDV DNA is integrated in the wasp genome and vertically transmitted through the germ line (Fleming & Summers, 1991). Shortly after pupation CsPDV begins to replicate in specialized cells of the wasp's oviduct epithelium (calyx cells) and then buds into the oviduct lumen (Norton & Vinson, 1983; Webb & Summers, 1992). During oviposition, polydnavirus, venom, and ovarian proteins are injected together with the wasp eggs into *H. virescens* larvae. It has been shown that introduction of CsPDV alone into *H. virescens* larvae causes dramatic physiological changes that mimic natural parasitization (Edson et al., 1981; Shelby & Webb, 1994). However, recent studies demonstrate that an ovarian protein complex rapidly but transiently inactivates the *H. virescens* encapsulation response early in parasitization (Webb & Luckhart, 1994, 1996). Later, the persistent, long-term immunosuppression is provided by CsPDV gene expression (Luckhart & Webb, 1996).

Although CsPDV replication has not been detected in parasitized *H. virescens* (Theilmann & Summers, 1986),

---

* Author for correspondence. Fax +1 606 323 1120.
e-mail BAWEBB@pop.uky.edu

The sequence data reported in this paper have been deposited in GenBank under accession numbers U41655 and U41656.
Fig. 1. (a) For legend see facing page
**Fig. 1.** (a) Nucleotide and predicted amino acid sequences of the VHvl.4 gene. The 5' nucleotide of the cDNA clone is in bold and underlined. A putative TATA box 42 bp upstream of the 5' end of the cDNA and the polyadenylation signal (AATAAA) are underlined. The longest ORF of 966 nucleotides encodes a protein of 322 amino acids. N-Glycosylation sites (NXT) are in italic and underlined. Internal EcoRI and SalI restriction sites are underlined and indicated. Sequences used as VHvl.4-specific primers and the Y RACE primer are indicated by arrows. Cysteine residues of the cysteine motifs are in bold. Intron sequences are in lower-case letters.

(b) Schematic representation of VHvl.4 cDNA and genomic DNA. The 0.8 kb cDNA selected from the 2gtl 1 cDNA library and the 628 bp 3' RACE product are also shown. The numbering is relative to the 5' end of the cDNA. The 5' UTR, ORF and 3' UTR are distinguished by different boxes. The positions of splicing junctions of introns (In1, In2, In3 and In4) are indicated by dashed lines between the genomic and cDNA sequences. The ORF spans nucleotides 57-1025 of the cDNA. The 3' RACE fragment was used as the VHvl.4-specific probe for Southern and Northern hybridizations.
at least 12 viral transcripts are detected as early as 2 h post-parasitization (p.p.) (Fleming et al., 1983; Blissard et al., 1986a, b). Four viral genes, WHv1.6, WHv1.0, VHv1.1 and BHv0.9, which encode abundantly expressed at least 12 viral transcripts are detected as early as 2 h and grouped into gene families. BHv0.9 is a member of the 'repeat element' (rep) polydnavirus gene family that is characterized by the presence of a 540 bp repeat sequence (Theilmann & Summers, 1987, 1988). The other three genes belong to the cysteine-rich polydnavirus gene family (Blissard et al., 1987; Dib-Hajj et al., 1993). Genes within this family have related structures including conserved splicing sites in coding and noncoding regions, conserved cysteine motifs in their predicted amino acid sequences, and are expressed only in parasitized insects (Dib-Hajj et al., 1993). The conserved CsPDV cysteine motifs contain six invariable cysteine residues flanking highly variable inter-cysteine amino acids. Two cysteine-rich genes, WHv1.6 and WHv1.0, are located on viral DNA segment W and show regions of significant similarity in both nucleotide and predicted amino acid sequences (Blissard et al., 1987). The VHv1.1 gene is a cysteine-rich gene on viral segment V. The VHv1.1 cDNA cross-hybridizes to two mRNAs of 1-4 and 2.7 kb expressed in parasitized insects, suggesting that these transcripts may represent other members of the cysteine-rich gene family (Webb & Summers, 1990; Dib-Hajj et al., 1993).

Recently, Li & Webb (1994) showed that the VHv1.1 gene product, a 30 kDa protein, plays an important role in disrupting the immune system of H. virescens (Blissard et al., 1989). The VHv1.1 protein binds to granulocytes and plasmatocytes, the major immunocytes in insects. Using a baculovirus expression vector, this gene was introduced into H. virescens larvae, where its expression reduces the encapsulation response to wasp eggs (Li & Webb, 1994). Given the significance of the VHv1.1 gene, analysis of other related cysteine-rich genes is necessary. Here we describe the isolation and characterization of the 1-4 kb gene which also maps to viral segment V. Analysis of nucleotide and predicted amino acid sequences revealed high levels of identity between the VHv1.1 and VHv1.4 genes. Like the VHv1.1 gene, the VHv1.4 gene contains two complete cysteine motifs. In addition, the VHv1.4 gene has an incomplete cysteine motif, presumably resulting from the partial duplication of the N-terminal cysteine motif.

Methods

Insects and virus. Insect rearing, and virus and viral DNA purification from calyx fluid, were done as described by Krell et al. (1982). For RNA analyses requiring parasitized H. virescens, 15-20 third-instar larvae were parasitized by 8–10 mated female wasps within 30 min. At the end of this period, larvae were designated as 0 h p.p.

Cloning of the 1.4 kb cDNA. A 2gtl 1 cDNA library constructed from mRNA of parasitized H. virescens was screened by colony hybridization using the VHv1.1 cDNA as a probe (Sambrook et al., 1989). Hybridizing phage plaques were amplified by PCR using 2gtl 1 forward (5' GGGTGCGCAGACCTCCTGGACG 3') and reverse (5' GACACAAGCACAACTGTAATG 3') primers (Tung et al., 1989). Amplification reactions included 1 x PCR buffer (10 mm-Tris–HCl, pH 9.0, 50 mm-KCl), 50 µM of each dNTP, 1-25 mm-MgCl2, 0.5 µg of each of the 2gtl 1 primers, 5 µl of the phage suspension and 2-5 units of Taq polymerase (Promega). Phage DNA was denatured at 94°C for 2 min and 35 amplification cycles were performed (94°C, 2 min; 55°C, 2 min; 72°C, 2 min) in a model 480 DNA thermocycler (Perkin Elmer Cetus). The amplified DNA fragments were digested with EcoRI and cloned in Bluescript II KS(−) (Stratagene) for sequencing.

Rapid cloning of the 3' cDNA end (3' RACE). To clone the 3' end of the 1-4 kb cDNA, 0.5 µg of total RNA from parasitized H. virescens larvae at 24 h p.p. was reverse transcribed using an oligo(dT) primer (ACGCTTAAGCTG) (Frohman, 1990). Briefly, reverse transcription was performed at 42°C for 20 min in 20 µl of reaction mix containing 1 x DNA synthesis buffer (50 mm-Tris–HCl, pH 8.9, 10 mm-KCl, 2.5 mm-MgCl2), 200 µM of each dNTP, 10 µM-DTT, 0.2 µg of oligo(dT) primer and 200 units of MMLV reverse transcriptase (Promega). The reaction mixture was digested with 2 units of RNase H at 42°C for 10 min. One µl of the reverse transcription mixture was removed for PCR with oligo(dT) and 1-4 kb cDNA-specific primers (Fig. 1a) using 35 cycles of 94°C, 1 min; 48°C, 1 min; 72°C, 2 min. The PCR product was cloned in the pCR-TRAP vector (GeneHunter) for sequence analysis.

Cloning and mapping of viral segment V. A 129 kb EcoRI fragment of SH V, pVE12.9, has been cloned in PBS (Dib-Hajj et al., 1993). This clone was used to screen a CsPDV Pal library in Bluescript II KS(−) to select an overlapping 7.9 kb clone (pVP7.9) that hybridized to both termini of pVE12.9. Another CsPDV SacAI library made from partially digested viral DNA was probed with the 0.8 kb fragment of the 1-4 kb cDNA. Clones hybridizing to the probe were rescanned by PCR with 1-4 kb cDNA-specific primers 1, 2 and 3 (Fig. 1a). Amplification was performed with the GeneAmp 9600 system (Perkin Elmer Cetus) using the following protocol: 2 min at 94°C; 30 s at 94°C; 30 s at 55°C; 1 min at 72°C; 35 cycles. Overlapping SH V genomic clones were mapped with restriction enzymes.

Southern blots and mapping of the 1-4 kb gene. In order to map the 1-4 kb cDNA and VHv1.1 gene to viral genomic segments, Southern hybridization was performed. Restriction digestion, electrophoresis and Southern blot hybridization were done as described (Sambrook et al., 1989). Gene-specific fragments of the VHv1.1 (~600 bp of the 3' cDNA region; Dib-Hajj et al., 1993; Fig. 1, nucleotides 894-1800, excluding intron 3) and 1-4 kb cDNA (3' RACE product; see Fig. 1a, nucleotide 1351-2310, excluding intron 4) were used in the mapping experiments. These probes hybridized strongly to the corresponding mRNA (VHv1.1, 600 bp to 1.1 kb mRNA, and 3' RACE product to 1.4 kb mRNA) but weak cross-hybridization was also noted. A 200 bp fragment from the 5' end of the 1-4 kb cDNA (extending from the 5' of the cDNA to the first SfiI; Fig. 1a) was also used for mapping the 5' region. DNAs used for probes were extracted from agarose gels using GeneClean (Bio101). Random primer labelling with [32P]dATP was done using the Prime-a-Gene labelling system (Promega). Nylon blots were prehybridized for 1-5 h at 42°C in hybridization solution (50% formamide, 5 × Denhardt's solution, 6x SSC, 0.5% SDS). Hybridizations were performed at 42°C overnight with approximately 1-5 × 106 c.p.m./ml probe. Blots were washed under conditions of high stringency for 30 min in 0.1 × SSC, 0.1% SDS at 65°C and autoradiographed at −80°C.
Fig. 2. Comparison of the VHv1.1 and VHv1.4 genes. (a) Dot matrix plot of VHv1.1/VHv1.4 cDNAs. The nucleotide sequences are numbered on each axis. ORFs of VHv1.1 and VHv1.4 are shown as open or filled boxes, respectively. A ~280 bp region shown as a hatched box on VHv1.1 has a high degree of identity with two regions on VHv1.4. (b) Alignment of nucleotide (nt) and amino acid (aa) sequences of VHv1.4 cDNA domain 1 (D1, 182-463) and domain 2 (D2, 581-865). Dots indicate identical nucleotides and asterisks indicate conserved amino acids. Gaps (-) are introduced to optimize the alignment. Conserved cysteine residues are underlined. The highly variable regions on the two domains are underlined. These two domains show 77% nucleotide identity and 55% amino acid identity.
RNA analysis. Total RNA was extracted from *H. virescens* at 0, 2, 8, 24 and 48 h p.p. by the method of Chomczynski & Sacchi (1987). For Northern blots 10 μg of total RNA was separated on a 1% agarose gel with formaldehyde (Webb & Summers, 1990). RNA was transferred to nylon membrane in 10 x SSC and UV cross-linked in a Stratalinker.

Fig. 3 (a, b, c). For legend see facing page.
Fig. 3. Mapping the VHv1.4 gene to the SH V segment. (a) Restriction maps of the three overlapping clones of SH V: pVE12.9, pVP7.9 and pVS11.2. Hatched boxes show regions that hybridized to the VHv1.1-specific probe and filled boxes indicate hybridization to the VHv1.4-specific probe. The unique XhoI site (boxed) was arbitrarily designated the ‘0’ position for the composite SH V. Enzyme abbreviations: EcoRI, E; HindIII, H; PstI, P; Sall, Sa; SalI, S; XhoI, X. (b) Circular map of SH V (15.2 kb). The location and orientation of the VHv1.1 and VHv1.4 genes are shown as hatched and filled regions and by arrows. (c) PCR amplification of the three SH V clones using the VHv1.4 cDNA-specific primer pairs 2 + 3 and 1 + 3 shown in Fig. 1 (a). Each PCR reaction (50 μl) included 5 ng of plasmid DNA (pVE12.9, pVP7.9 or pVS11.2) as the template. Ten μl of the reactions was loaded in each lane. Lane 1, no DNA; lanes 2 and 6, 0.8 kb cDNA clone; lanes 3 and 7, pVS11.2; lanes 4 and 8, pVP7.9; lanes 5 and 9, pVE12.9. Note that the sizes of DNA produced when using genomic DNAs as the templates are larger than when using cDNA as the template due to the presence of intron(s). (d) Southern blots of SH V clones, pVS11.2 was digested by Sall (S) and HindIII + EcoRI (H + E). pVE12.9 and pVP7.9 were digested with Sall (S) and XhoI (X). Ethidium bromide stained gels (EtBr) of digestions are shown. The blots were probed with the 600 bp 3’ region of VHv1.1 cDNA (1.1) or with 3’ RACE product from the VHv1.4 cDNA (1.4).

DNA sequence determination and analysis. Nucleotide sequences were determined by the dideoxy termination method of Sanger et al. (1977) using the Sequenase 2.0 kit (United States Biochemical). Sequence data were analysed using the University of Wisconsin Genetics Computer Group DNA analysis software for the VAX computer (release 7.2).

Results

Nucleotide sequence of the 1.4 kb cDNA and predicted protein sequence

To isolate the 1.4 kb CsPDV gene expressed in parasitized H. virescens larvae, an existing λgt11 cDNA library was screened (B. Graham & M. D. Summers, unpublished). Twenty-four cDNA clones hybridizing to the VHv1.1 gene were identified by colony hybridization using the VHv1.1 cDNA as a probe. When amplified by PCR using λgt11 forward and reverse primers, ten clones produced amplimers of 0.8 kb (seven clones), 0.9 kb (two clones), and 2.3 kb (one clone). Sequencing of the 0.9 kb clones determined that these encoded the VHv1.1 cDNA. Sequencing of four 0.8 kb clones revealed that they were identical and highly homologous to, but different from the VHv1.1 cDNA. Nucleotide sequence analysis of the 3’ end of the 0.8 kb cDNA clones indicated an internal EcoRI site was present in the cDNA (i.e. the EcoRI site was not flanked by nucleotides on the EcoRI adaptor that was used for library construction), suggesting that the 3’ terminus may have been lost during library construction. To generate a full-length cDNA, 3’ RACE was performed with an oligo(dT) primer and the 1.4 kb cDNA-specific RACE primer (Fig. 1a). A ~600 bp 3’ RACE DNA fragment was the only detectable product and was cloned in the pCR-TRAP vector and sequenced. Sequence data showed partial overlapping of the 0.8 kb cDNA clones with the 3’ RACE clone up to the EcoRI site (Fig. 1b). Sequence and hybridization analyses of the 2.3 kb clone indicated
Table 1. Nucleotide and amino acid comparison (percentage identities) of members of the cysteine-rich polydnavirus gene family

Abbreviations: N, N-terminal cysteine motif; C, C-terminal cysteine motif; nt, nucleotides; aa, amino acids.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Signal peptide</th>
<th>Cysteine motif*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5' UTR† nt</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nt aa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nt aa</td>
<td></td>
</tr>
<tr>
<td>VHvl.1/VHvl.1</td>
<td>97 97 94</td>
<td>(N/C) 60 48 50</td>
</tr>
<tr>
<td>VHvl.1/VHvl.4</td>
<td>60 96 90 50</td>
<td>(N/N) 60 96 90</td>
</tr>
<tr>
<td>VHvl.1/VHvl.4</td>
<td>59 63 33 24</td>
<td>(N/C) 59 51</td>
</tr>
<tr>
<td>VHvl.1/VHvl.4</td>
<td>68 63 43 37</td>
<td>(C/C) 68 63</td>
</tr>
<tr>
<td>VHvl.1/VHvl.4</td>
<td>59 47 53 41</td>
<td>(C/N) 59 47</td>
</tr>
<tr>
<td>VHvl.1/VHvl.4</td>
<td>56 56 56</td>
<td>(N) 54 40</td>
</tr>
<tr>
<td>VHvl.1/VHvl.6</td>
<td>55 58 56</td>
<td>(C) 55 40</td>
</tr>
<tr>
<td>VHvl.1/VHvl.6</td>
<td>63 69 63</td>
<td>(N) 56 49 37</td>
</tr>
<tr>
<td>VHvl.1/VHvl.6</td>
<td>57 48 40 29</td>
<td>(C) 57 48</td>
</tr>
<tr>
<td>VHvl.1/VHvl.6</td>
<td>55 58 56</td>
<td>(N/C) 55 58 47</td>
</tr>
<tr>
<td>VHvl.1/VHvl.6</td>
<td>54 40 40 29</td>
<td>(N) 54 40</td>
</tr>
<tr>
<td>VHvl.1/VHvl.6</td>
<td>51 30 30 29</td>
<td>(C) 51 30</td>
</tr>
<tr>
<td>WHvl.0/WHvl.6</td>
<td>64 65 63</td>
<td>(N) 56 50 37</td>
</tr>
<tr>
<td>WHvl.0/WHvl.6</td>
<td>42 40 40 34</td>
<td>(C) 42 40</td>
</tr>
<tr>
<td>WHvl.0/WHvl.6</td>
<td>87 90 94</td>
<td>(N) 76 65</td>
</tr>
<tr>
<td>WHvl.0/WHvl.6</td>
<td>76 85 85 83</td>
<td>(C) 76 85</td>
</tr>
</tbody>
</table>

* The length of the cysteine motifs is typically 41 amino acids. The cysteine motif of WHvl.0 is only 30 amino acids in loop 3 (Fig. 4b). The deletion is considered a single mutation in calculating percentage similarity with the other motifs. To compare with the WHvl.0 cysteine motif (column 'nt 90' and 'aa 30'), sequences in other CsPDV cysteine motifs (nt 58-77, aa 19-29) were deleted.
† The 5' UTR lengths in VHvl.1, VHvl.4, WHvl.0 and WHvl.6 are 72, 70, 75 and 77 bp, respectively.
‡ The 3' UTR lengths in VHvl.1, VHvl.4, WHvl.0 and WHvl.6 are 300, 313, 226 and 803 bp, respectively.

that this clone was an artifact produced during library construction and was not derived from a 2.7 kb mRNA that cross-hybridized to the VHvl.1 gene (Webb & Summers, 1990). As the 2.7 kb mRNA only shows weak hybridization to VHvl.1 cDNA under low stringency and we used stringent hybridization conditions to screen the library, it is unlikely that clones representing the 2.7 kb gene would have been selected in this screen.

The composite sequence of the 1.4 kb cDNA which was generated by library screening and 3' RACE is 1338 bp long (Fig. 1a). The longest ORF of 966 nucleotides from nucleotide 57 (relative to the 5' end of the cDNA clone) to 1022 was identified. The sequence surrounding the first methionine codon of the 1.4 kb cDNA (ACCATGA) is consistent with the translation initiation consensus sequence (PuNNATGpu) (Kozak, 1983). A possible polyadenylation signal (AATAAA) is located 15 bp upstream of the poly(A) tail (Fig. 1a).

As revealed by dot matrix comparison of the cDNAs of the 1.4 kb and VHvl.1 genes, two regions of ~280 bp on the 1.4 kb cDNA (domain 1 and domain 2 are from nucleotides 182-463 and 581-865 from the 5' end of the cDNA, respectively) are homologous to the same region of the VHvl.1 gene (Fig. 2a). These two cDNA domains lie within the 1.4 kb gene coding region and are 77% identical in nucleotide sequence (Fig. 2b).

The longest ORF of the 1.4 kb cDNA encodes a protein of 322 amino acids with a predicted molecular mass of 42 kDa (Fig. 1a). The N-terminal sequence is very hydrophobic and encodes a signal peptide according to the rules of von Heijne (1986), indicating this protein is destined for insertion into the membrane or secretion. There are six potential N-glycosylation sites N-X-T within the VHvl.1 protein. Similar to the VHvl.1 protein, the predicted protein of the 1.4 kb cDNA has two complete cysteine motifs (Fig. 1a, amino acids 40-80 and 277-317, respectively). The two homologous domains (Fig. 2b) encode amino acid sequences with 55% identity and 72% similarity to each other. The first domain contains the N-terminal cysteine motif, while the second domain has only two cysteine residues residing at comparable positions to the N-terminal cysteine motif. A highly variable region of 10-11 amino acids is observed in comparisons of the two domains (Fig. 2b).

Mapping of the 1.4 kb cDNA to SH V and sequencing of the genomic DNA

In order to isolate the genomic DNA encoding the 1.4 kb gene, a Bluescript plasmid library was made from viral DNA partially digested with Sau3AI. Four clones were
Fig. 4. Alignments of predicted protein sequences and cysteine motifs. (a) Alignment of the predicted amino acid sequences encoded by the VHvl.4 and VHvl.1 genes. The full-length polypeptides are 62% identical and 72% similar. Sixteen N-terminal amino acids encoding a signal peptide are underlined. Two homologous domains D1 and D2 are underlined. Dots indicate identical amino acids and asterisks indicate conserved amino acid substitutions. Gaps (-) are introduced to maximize the alignment, (b) Alignment of the CsPDV cysteine motifs. CsPDV cysteine motifs have 41 amino acids except for the WHvl.0 cysteine motif which has 30 amino acids. Gaps (--) are introduced to optimize alignment of the cysteine residues. The only common inter-cysteine amino acid P (in bold) for the six cysteine motifs immediately precedes the third cysteine residue. N, N-terminal; C, C-terminal.

selected by colony hybridization with the [32P]dATP-labelled 0.8 kb fragment of the 1-4 cDNA. These four clones were chosen as templates for PCR amplification using 1-4 kb cDNA-specific primers (Fig. 1a). Only one clone of 11.2 kb had PCR products and was used for mapping with restriction enzymes.

Two subclones of band-isolated segment V, a 12.9 kb EcoRI clone (pVE12.9) and a 7.9 kb PsI clone (pVP7.9), were also mapped and contained 5.6 kb of overlapping sequence (Fig. 3a). The VHvl.1 gene mapped to a 3.8 kb XhoI fragment of the pVE12.9 clone (Dib-Hajj et al., 1993). Restriction enzyme mapping of the 11.2 kb clone, which contains the 1.4 kb gene, indicated an overlap with the pVE12.9 and pVP7.9 clones. Therefore, these overlapping clones are fragments of SH V, and were used to generate the complete physical map of SH V (Fig. 3a, b). Accordingly, the 11.2 kb Sau3AI clone is designated pVS11.2, and the 1.4 kb gene VHvl.4.

PCR and Southern hybridization analyses were done to map VHvl.4 on the cloned genomic fragments. PCR analyses of the three genomic clones showed 0.6 and 1.0 kb DNA amplimers when using 1.4 kb cDNA-specific primer pairs 2 + 3 and 1 + 3 (Fig. 3c), indicating that the VHvl.4 gene is located in the overlapping region of these three genomic clones. Due to the presence of intron(s) in genomic DNA, amplimers from the cDNA clone were smaller than from the genomic templates. For Southern blots, plasmid DNAs were digested by SalI, XhoI, and HindIII+EcoRI and the blots were hybridized to labelled cDNA probes. When probed with the 600 bp 3'
The lengths of introns 1 of VHvl.1, VHvl.4, WHvl.0 and WHvl.6 are 124, 124, 107 and 117 bp, respectively. The lengths of introns 2 of VHvl.1, VHvl.4, WHvl.10 and WHvl.6 are 186, 186, 429 and 409 bp, respectively. Introns 3 of VHvl.1 and VHvl.4 are 296 and 197 bp long, respectively. Intron 4 of VHvl.1 is 342 bp long.

<table>
<thead>
<tr>
<th>Genes</th>
<th>In 1</th>
<th>In 2</th>
<th>In 3</th>
<th>In 2/In 3</th>
<th>Inc2/In 4</th>
<th>In 3/In 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>VHvl.1/VHvl.1</td>
<td>-</td>
<td>-</td>
<td>38</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>VHvl.1/VHvl.4</td>
<td>99</td>
<td>96</td>
<td>47</td>
<td>96</td>
<td>56</td>
<td>72</td>
</tr>
<tr>
<td>VHvl.1/WHvl.10</td>
<td>50</td>
<td>41</td>
<td>46</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>VHvl.1/WHvl.0</td>
<td>49</td>
<td>39</td>
<td>46</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>VHvl.4/VHvl.1</td>
<td>-</td>
<td>-</td>
<td>97</td>
<td>51</td>
<td>48</td>
<td>-</td>
</tr>
<tr>
<td>VHvl.4/VHvl.6</td>
<td>49</td>
<td>42</td>
<td>42</td>
<td>44</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>VHvl.4/WHvl.1</td>
<td>50</td>
<td>43</td>
<td>43</td>
<td>46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WHvl.6/WHvl.0</td>
<td>88</td>
<td>92</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Abbreviation: In, intron.

Comparison of VHvl.14 with other cysteine-rich CsPDV genes

Database searches did not reveal homology with genes outside the CsPDV cysteine-rich gene family. Within the cysteine-rich gene family, VHvl.14 is very homologous to the VHvl.1 gene (cDNAs have 62% identity). Both VHvl.14 and VHvl.1 genes are highly conserved in the 5' regions and show lesser but varying degrees of identity in 3' sequences (Table 1). The 5' untranslated regions (5' UTR) of both genes have 97% identity and are interrupted by intron 1 at similar positions relative to the TATA box. Coding regions for the predicted signal peptides and N-terminal cysteine motifs are also highly conserved with 96% identity (Fig. 4a). The C-terminal cysteine motifs of the VHvl.14 and VHvl.1 genes show less identity (63%). The 3' UTRs of the two genes have only 52% identity (Table 1). Introns 1 of the VHvl.14 and VHvl.11 genes are both 124 bp in length and differ by only one nucleotide (Table 2). The other three introns of the VHvl.14 gene are located at conserved positions relative to the VHvl.1 introns in that all immediately precede the second cysteine codon in the cysteine motif. Introns 2 of the VHvl.11 and VHvl.14 genes are 96% identical to each other, and are 96% and 97% identical, respectively, to intron 3 of the VHvl.14 gene. VHvl.14 intron 2 and intron 3, located at conserved positions of the two conserved domains, have greater identity (97%) than their flanking exon sequences (77% identity). Introns more conserved than flanking exons were also observed for 3' coding regions of the two genes where VHvl.11 intron 3 and VHvl.14 intron 4 have greater identity (72%) than their immediate flanking exon sequences (64%). Clearly, VHvl.11 intron 2 is closely related to VHvl.14 introns 2 and 3, while VHvl.11 intron 3 is closely related to VHvl.14 intron 4.

Comparison of the nucleotide sequences of the four genes in the cysteine-rich gene family from 5' to 3', we observed that the regions encoding the signal peptides are more than 56% identical (Table 1). The cysteine motifs are 51-68% identical, except that the N-terminal cysteine motifs of VHvl.11 and VHvl.14 are 96% identical, and the cysteine motifs of WHvl.10 and WHvl.16 are 76% identical. The 3' UTRs have less identity (<52%) with the exception of the WHvl.0 and WHvl.16 3' UTRs (83% identical). Analysis of homologies between introns showed similar results (Table 2). Little identity was detected when VHvl.1 and VHvl.14 introns were compared to WHvl.10 and WHvl.16 introns.

Comparison of the VHvl.14 predicted protein sequence with the VHvl.11, WHvl.10 and WHvl.16 predicted protein sequences gave results similar to those at the nucleotide level. The two SH V proteins are more homologous to each other than to the SH W proteins (Fig. 4a). Within the cysteine-rich CsPDV gene family, the putative signal peptides are relatively more conserved (at least 56% identity) than other regions (Table 1). All the CsPDV cysteine-rich genes encode the characteristic cysteine motif (C---C---CC---C---C) (Fig. 4b). In addition, the inter-cysteine amino acids are highly variable with only one common amino acid for all the CsPDV cysteine motifs. The two SH V genes have very conserved N-terminal cysteine motifs (90% identity) (Table 1). Amino acid sequences of the two repetitive domains on VHvl.14 are 71% and 58% identical to their VHvl.1 homologue.
**Temporal expression of VHvl.4 and VHvl.1 genes**

Using the VHvl.1 3’ RACE product as the probe, we could specifically detect the 1.4 kb mRNA (Fig. 5a). Similarly, the VHvl.1 cDNA probe (600 bp 3’ region) is also relatively specific with only minor cross-reaction with the VHvl.4 mRNA (Fig. 5b). Northern analyses were performed to study the relative expression of the VHvl.1 and VHvl.4 genes. After a 30 min parasitization period, VHvl.4 and 1.1 transcripts were detected at 2 h p.p. At 8 h p.p. these two transcripts were very abundant, and remained throughout the following 2 days. The coordinate expression of the VHvl.1 and 1.4, and WHvl.0 and 1.6 genes suggests that their expression is coordinately regulated.

**Discussion**

To date, all of the sequenced CsPDV genes expressed in *H. virescens* are members of either the rep gene family or the cysteine-rich gene family. Here we report that another CsPDV cysteine-rich gene, VHvl.4, maps to SH V and is closely related to the VHvl.1 gene. VHvl.4 has a gene structure similar to that of other CsPDV cysteine-rich genes including conserved splicing sites in coding and noncoding regions, and homologous 5’ UTRs and putative signal peptides (> 55% identical). The protein encoded by VHvl.4 contains two complete copies of a conserved cysteine motif present in the CsPDV cysteine-rich gene family and several other biologically active proteins and polypeptides (Dib-Hajj et al., 1993; Olivera et al., 1990, 1991). Interestingly, the CsPDV cysteine-rich genes on the same DNA segments have greater identity (> 90% in some regions) than do members of the gene family located on different DNA segments. The ORF of the VHvl.4 gene is larger than the VHvl.1 ORF because it contains two domains homologous to the VHvl.1 N-terminal cysteine motif. The second homologous domain appears to be the result of an internal duplication of the N-terminal cysteine motif followed by the loss of most of the cysteine residues in the duplicated domain. The VHvl.4 C-terminal cysteine motif has greatest identity to the VHvl.1 C-terminal motif. Northern blot analyses revealed that VHvl.1 and 1.4 transcripts appear early and in abundance in parasitized insects and are likely to be coordinately regulated.

Although all the CsPDV cysteine-rich genes have similar gene structures, conserved cysteine motifs, and appear to have arisen from a common ancestral gene, there are significant variations in gene structure and sequence that likely reflect selection for variation in this gene family. The two SH W genes, WHvl.0 and WHvl.6 encode proteins that contain a single cysteine motif and a precysteine domain of 26 amino acids. In contrast, the SH V genes, VHvl.4 and VHvl.1 encode proteins with two cysteine motifs that lack a precysteine domain. The SH W cysteine-rich genes have two introns (Blissard et al., 1987), whereas the VHvl.1 and VHvl.4 genes have three or four introns, respectively (Dib-Hajj et al., 1993). Therefore, the WHvl.0, WHvl.6, VHvl.1 and VHvl.4 genes appear to form two distinct subfamilies, the W and V subfamilies. The SH V and SH W genes are more divergent than are genes residing on the same CsPDV genomic segment, indicating the result of relatively recent gene duplication events within the same viral segment. Interestingly, the VHvl.4 repetitive domain of
\[\sim 280\text{ nucleotides in length also appears to be the result of a duplication event but in this case the duplication has occurred within the \(VHvl.4\) gene. The first \(VHvl.1\) domain encodes the N-terminal cysteine motif, while the second encodes a cysteine motif that is very similar in amino acid sequence (55\% identity) but retains only two of the cysteine residues in the motif. Conservation of inter-cysteine residues in the two domains of the \(VHvl.4\) gene and the loss of cysteine residues in the second domain stand in marked contrast to the conservation of cysteine residues and high variability of the inter-cysteine residues in the six complete cysteine motifs.

Introns within the cysteine motif that are more conserved than flanking exons have been observed in the \(WHvl.1\) and \(WHvl.6\) genes (Dib-Hajj et al., 1993). This unusual phenomenon is seen not only between the 3' coding regions of the two SHV genes, but also between different domains of the \(VHvl.4\) gene. Within the two repetitive domains of \(VHvl.1\), intron 2 and intron 3 have greater identity (97\%) than their flanking exons (77\%). The conserved introns probably indicate recent duplication events rather than a functional role for introns which has led to their relative conservation and selection for variation in the coding sequences.

The four cysteine-rich CsPDV proteins have signal peptides that may direct the proteins for secretion. Using polyclonal antisera against \(VHvl.1\) protein, Li & Webb (1994) confirmed that this protein is secreted and present in abundance in the plasma of parasitized insects. \(WHvl.6\) and \(WHvl.1\) proteins are also efficiently secreted from recombinant baculovirus-infected cells (Blissard et al., 1989). All the characterized CsPDV cysteine-rich proteins are potentially glycosylated as N-glycosylation sites are located on the predicted proteins. Although \(WHvl.1\) and \(WHvl.6\) proteins both have one potential glycosylation site, [\(^{35}S\)mannose labelling and tunicamycin treatments of recombinant baculovirus-infected Sf9 cells indicate that \(WHvl.1\) protein is glycosylated, whereas \(WHvl.6\) protein is not or is only slightly glycosylated (Blissard et al., 1989). The \(VHvl.1\) protein, with one potential N-glycosylation site, is also glycosylated as determined by concanavalin A labelling and enzymatic deglycosylation (A. I. Soldevila & B. A. Webb, unpublished). Sequence analysis suggests that the \(VHvl.1\) protein may also be glycosylated since six typical N-glycosylation sites were identified. Using polyclonal antiserum against the \(VHvl.1\) protein, another protein of approximately 50 kDa was detected only in parasitized insects (Li & Webb, 1994). The appearance of this 50 kDa protein as a diffuse band may indicate that it is glycosylated. While the \(VHvl.1\) gene encodes a potential protein of 322 amino acids with a calculated molecular mass of 42 kDa, its glycosylated forms will be larger. The homology between the N termini of the \(VHvl.1\) and \(VHvl.1\) predicted proteins suggests that the \(VHvl.1\) protein could cross-react with \(VHvl.1\) antisera. Based on the size of the predicted protein, the putative glycosylation sites and the homology with the \(VHvl.1\) gene, our working hypothesis is that the \(VHvl.1\) gene encodes this 50 kDa protein.

Cysteine-rich proteins and peptides have been isolated from animal venoms, such as those of carnivorous snails, scorpions, spiders and snakes. In many of these venoms, the cysteine residues confer conformational stability on the venom toxin proteins or peptides (Olivera et al., 1990, 1991). The cysteine distribution of C---C---CC---C---C occurs in many other naturally occurring polypeptides such as the \(\alpha\)-conotoxins (Olivera et al., 1990, 1991), the \(cth\) gene product of \(Autographa\) \(californica\) nucleopolyhedrovirus (Eldridge et al., 1992) and neurophysins (Nojiri et al., 1987). The CsPDV cysteine-rich proteins are much larger (more than 198 amino acids long) than the proteins listed above, and their functions may be quite different. However, it is likely that the conserved cysteine motifs in the CsPDV cysteine rich genes have similar structural roles that confer and stabilize specific protein conformations that support their functional activities. In conotoxins, the combination of hypervariable regions and a constant backbone of disulphide bridges generates high-affinity ligands with various specificities (Olivera et al., 1990). The cysteine motifs of CsPDV cysteine-rich gene family also have highly variable inter-cysteine residues. On the basis of this similarity, we propose that the cysteine residues form a rigid structural backbone for the proteins that allows the CsPDV proteins to function as ligands for surface proteins of target cells, particularly, the haemocytes. Direct targeting of plasmocytes and granulocytes of \(H.\) \(virescens\) larvae by \(VHvl.1\) protein has recently been demonstrated (Li & Webb, 1994). Suppression of the host cellular encapsulation response by CsPDV may be the result of direct infection of the immunocytes and/or targeting by CsPDV gene products. In vivo functional analyses using the baculovirus expression system to introduce the \(VHvl.1\) gene into \(H.\) \(virescens\) larvae showed partial immunosuppressive activity of the \(VHvl.1\) protein (Li & Webb, 1994). Other CsPDV cysteine-rich genes may have similar functions as they have a common gene structure and are co-expressed in parasitized insects. The presence of many cysteine-rich genes in the CsPDV genome may allow targeting of many types of immunocyte receptors to effectively compromise the immunity of a range of host species. Whether other cysteine-rich CsPDV genes have additive or complementary functions to the \(VHvl.1\) gene in immunosuppression remains to be tested.

We thank B. Graham for providing the cDNA library and X. Li for the pVP7.9 clone. We also thank Kent Shelby, Ana Soldevila and Nancy Webb for critical review of the manuscript. This research is...
supported by the NIH (AI 3311402) and is publication 95-08-108 of the University of Kentucky Agricultural Experiment Station.

References


(Received 25 September 1995; Accepted 15 December 1995)