Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 3765–3769, April 1993
Microbiology

Structure and evolutionary implications of a "cysteine-rich" *Campoletis sonorensis* polydnavirus gene family

(endoparasite/insect immunity/insect virus/gene duplication/overdominant selection)

S. D. DIB-HAJJ*†, B. A. WEBB‡, AND M. D. SUMMERS*†§

*Department of Entomology, †Center for Advanced Invertebrate Molecular Sciences, Institute of Biosciences and Technology, and §Texas Agricultural Experiment Station, Texas A&M University, College Station, TX 77843; and ‡Department of Entomology, Rutgers University, New Brunswick, NJ 08903

Contributed by M. D. Summers, January 26, 1993

ABSTRACT **For successful parasitization, the female *Campoletis sonorensis* endoparasitic wasp injects a polydnavirus into its host, *Heliothis virescens*, during oviposition. Viral gene expression induces immunosuppression and alters development of the host. We report here that three abundantly expressed genes, V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6, describe a polydnavirus "cysteine-rich" gene family which may be important in inducing these host manifestations. These genes have a similar primary gene structure and their proteins contain cysteine motifs characteristic of snail ion-channel ligands, the ω-conotoxins. Like the ω-conotoxins, the intercysteine amino acid residues are hypervariable with only three identical amino acids in all motifs. The conservation of this domain in the three viral genes may reflect an important functional role for these viral proteins in the parasitization of *H. virescens*. The three genes also contain introns similar in sequence at comparable positions in their 5′ untranslated leaders and coding sequences. V*Hv*1.1 contains two cysteine motifs, and each motif is interrupted by an intron at the same position as in the cysteine motifs of W*Hv*1.0 and W*Hv*1.6. Intron 2 sequences of W*Hv*1.0 and W*Hv*1.6 are 92% identical, while the immediately flanking exon sequences encoding the cysteine motifs are only 76% identical. This provides an example of nuclear pre-mRNA introns which are more conserved than flanking exons among members of a gene family.**

Successful parasitization by certain ichneumonid and braconid hymenopteran endoparasites is accompanied by suppression of the cellular immune response, the encapsulation of foreign objects by hemocytes, and alteration of development of the insect host (1, 2). Induction of these phenomena may require injection at oviposition of venom gland secretions alone (3, 4), both venom gland secretions and a polydnavirus (5, 6), or only a polydnavirus (7–10). The parasite *Leptoplina heterotoma* destroys *Drosophila melanogaster* hemocytes by lamellolysin, a secretion product of the venom gland of the parasite (11). Factor(s) induced by polydnaviruses from the parasites *Campoletis sonorensis, Hyposoter fugitivus*, and *Microplitis mediator* inhibit encapsulation in their respective hosts (10, 12, 13), which correlates with alterations in the host hemocytes (10). It is not known whether these factors are viral or virally induced cellular products. Polydnavirus and venom gland secretions from different species apparently perform analogous functions. This may suggest an evolutionary continuity between parasite and polydnavirus gene products that are essential for parasite survival (7). Webb and Summers (14) reported that *C. sonorensis* venom gland proteins and polydnavirus (*Cs*PDV) envelope proteins share conserved epitopes.

Successful parasitization of *Heliothis virescens* by the ichneumonid endoparasitic wasp *C. sonorensis* requires only a polydnavirus, *Cs*PDV. Injection of purified *Cs*PDV mimics natural parasitization by inducing immunosuppression and developmental arrest (7–10). The *Cs*PDV genome consists of ≈28 covalently closed superhelical DNA segments that range in size from 5 to 21 kbp and are assigned alphabetical letters in an increasing order of size (15). Different viral genes are expressed from specific genomic segments in the parasite, in the host larva, or in both (16–19). In *H. virescens*, viral gene expression is detected 2 hr post-parasitization and mRNAs persist even at 9 days post-parasitization (17, 18). Based on nucleotide sequence similarity and immunological crossreactivity, a significant fraction of the viral genes studied to date describe gene families (14, 16, 18). Two genes reported previously and reevaluated in this study map to genomic segment W (W*Hv*1.0 and W*Hv*1.6) (19). They encode 1.0- and 1.6-kb mRNAs, respectively, and share regions of significant similarity in their nucleotide and amino acid sequences (19). W*Hv*1.0 and W*Hv*1.6 open reading frames (ORFs) were expressed in Sf9 insect cells (20), and proteins of the expected molecular weight were processed and secreted into the medium (21).

We now describe another viral gene, V*Hv*1.1, that is abundantly expressed in *H. virescens* and maps to segment V and we also present new data on W*Hv*1.0 and W*Hv*1.6.¶ Sequence analysis of V*Hv*1.1 and a reevaluation of W*Hv*1.0 and W*Hv*1.6 revealed that the three genes share a common gene structure, including conserved introns interrupting the coding sequences of a cysteine motif characteristic of the snail ion-channel ligands known as ω-conotoxins (22). These data describe a "cysteine-rich" viral gene family that is abundantly expressed in *H. virescens*.

MATERIALS AND METHODS

C. sonorensis rearing conditions, *Cs*PDV purification from female oviducts, and DNA isolation have been described (15).

A cDNA clone (pcHv900) (14) containing an incomplete V*Hv*1.1 ORF was isolated from a parasitized *H. virescens* cDNA library in λgt10 (18) and cloned into the pBS(+/−) vector (Stratagene). PCR and hybridization screening were done using exon-specific oligodeoxynucleotides, derived from the pcHv900 sequence, and λgt10-specific oligodeoxynucleotides: (*i*) exon-specific primer A (5′-AGATCTGGCTCATCAGTTGATT-3′), (*ii*) exon-specific primer B (5′-CTATAACCACTGCGACCAGT-3′), (*iii*) λgt10-specific primer C (5′-AGTTCAGCCTGGTTAAGTCC-3′), and (*iv*) λgt10-specific primer D (5′-GGCTTATGAGTATTTCTTCCAGGG-3′). PCR was done on supernatants of plaque lysates, using oligonucleotide pairs A/C and A/D and Vent DNA polymer-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *Cs*PDV, *Campoletis sonorensis* polydnavirus; UTR, untranslated region; ORF, open reading frame.

¶The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L08243, L08244, and L08245).

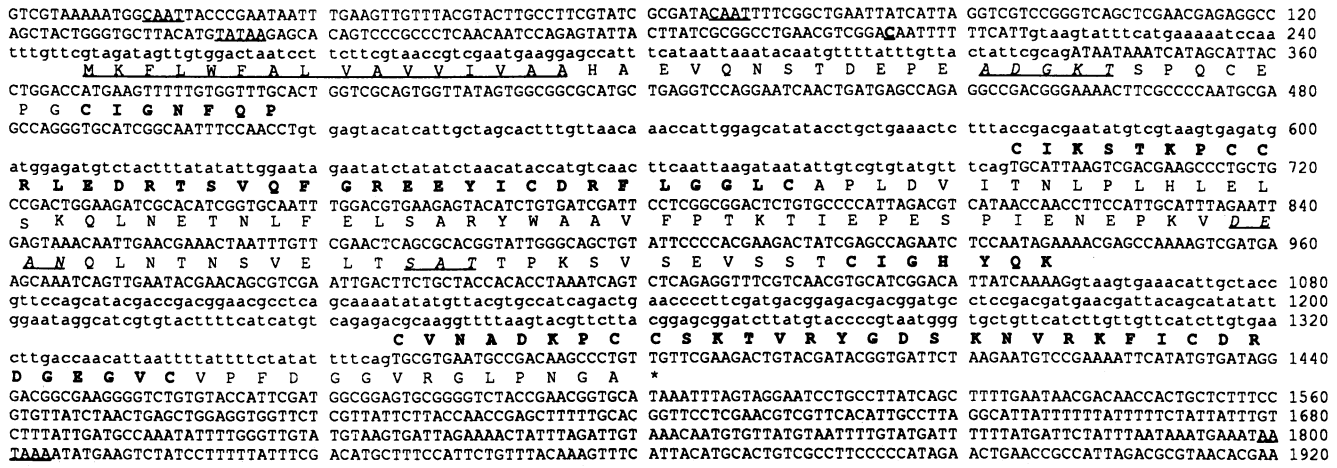

FIG. 1. Nucleotide and predicted amino acid sequences of the viral clone V*Hv*1.1. Two putative CAAT boxes and a TATA box are underlined. The 5′ nucleotide of the cDNA clone pcV*Hv*1.1 is in boldface type and underlined (position 204). A single ORF of 651 nucleotides encodes a protein of 217 amino acids. The underlined N-terminal 16 amino acid residues may function as a signal peptide according to the rules of Von Heijne (27). The N-terminal (L motif) and the C-terminal (R motif) cysteine motifs are shown in boldface type. Amino acid motifs shared with sequences of eukaryotic translation initiation factor eIF-4A are underlined and in italic type. These motifs are characteristic of the "DEAD" protein family (28). V*Hv*1.1 is interrupted by three introns shown here in lowercase type. Intron 1 interrupts the 5′ UTR; introns 2 and 3 immediately precede a cysteine codon. Based on the cDNA sequence, the polyadenylylation signal starts at position 1799.

ase (New England Biolabs). Amplification was done by the following protocol: 50 sec at 94°C, 2 min at 50°C, 2 min at 72°C for 30 cycles. Isolates containing a longer 5′ end than pcHv900 were further screened under high-stringency conditions (14) with 5′-end-labeled primer B. cDNA inserts from selected isolates were amplified by using the λgt10-specific primers C and D, as previously described, and ligated to pBS(+/−) cut with the restriction enzyme *Sma* I (Promega).

Genomic segment V was band isolated (19) and a 12-kbp *Eco*RI clone was selected which hybridized to the cDNA clone pcHv900 (14). A 3.8-kbp *Xho* I fragment (pVX3800) was subcloned into the pBluescript vector (Stratagene) by conventional methods (23). W*Hv*1.0 was cloned into pBS(+/−) vector (Stratagene) as a 4.3-kbp *Sal* I fragment (pWS4300) from segment W (19). The nucleotide sequence of the genomic subclones and the cDNA insert was determined by double-strand sequencing using the dideoxy chain-termination method (24) and Sequenase 2.0 (United States Biochemical).

Release 7.0 of the University of Wisconsin Genetics Computer Group software (25) was used for sequence analysis. In calculating percent similarity between sequences, the concept of maximum homology was applied (26), and gaps in the aligned sequence, irrespective of size, were counted as single mutational events.

RESULTS

Gene Structure of V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6. Previous reports (14, 17–19) described several *Cs*PDV genes, including the W*Hv*1.0/W*Hv*1.6 family, that are abundantly expressed after natural parasitization or injection of purified *Cs*PDV. Viral genomic (V*Hv*1.1) and cDNA (pcV*Hv*1.1) clones that encode a 1.1-kb mRNA detected from 2 hr post-paratization were mapped to segment V (ref. 14; unpublished data). The sequence of V*Hv*1.1 is presented here (Fig. 1); the sequence of the cDNA clone pcV*Hv*1.1 (accession number L08243) is in agreement with the genomic sequence. A putative TATA box and two CAAT boxes are located 63 bp, 137 bp, and 190 bp upstream of the 5′ end of the cloned cDNA (Fig. 1). V*Hv*1.1 is composed of four exons and three introns. A single reading frame encodes 217 amino acids. W*Hv*1.0 contains a putative TATA box and two introns at positions comparable to those in W*Hv*1.6 (Fig. 2; see also Fig. 5*a*). W*Hv*1.0 introns 1 and 2 are located at nucleotides 35 and 482, respectively, relative to the cDNA sequence (19).

The three genes are similar in structure (Fig. 3) and show varying degrees of similarity in different domains. (Sequences of different domains are aligned in Figs. 2, 4, and 5 and the results summarized in Table 1.) The 5′ untranslated region (UTR) and the signal peptide of V*Hv*1.1 show signif-

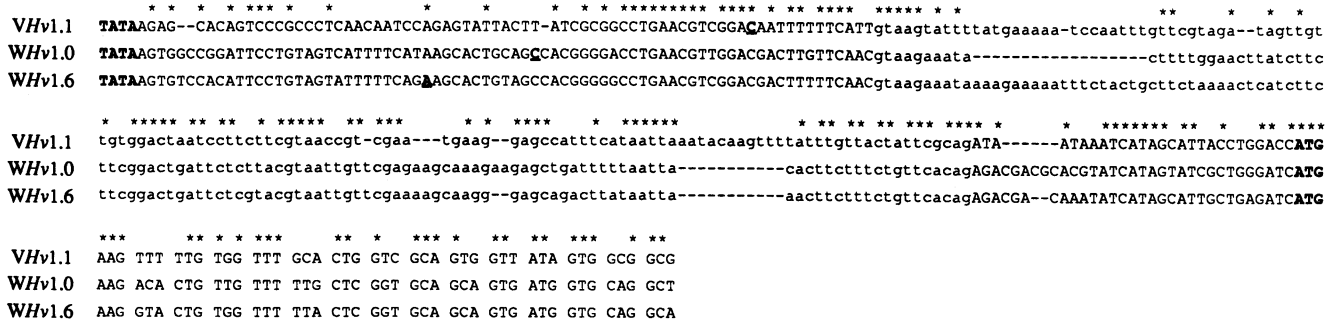

FIG. 2. Alignment of nucleotide sequences between the putative TATA box and the signal peptide of the genes V*Hv*1.1, W*Hv*1.0 (this study), and W*Hv*1.6 (19). Intron sequence is in lowercase type, stars indicate nucleotides conserved in the three genes, and dashes indicate deletions compared with the other sequences. For alignment purposes and based on the cDNA sequence of W*Hv*1.6 (19), the putative transcription initiation site (+1) in V*Hv*1.1 and W*Hv*1.0 is located 32 and 34 bp downstream of the putative TATA box of these genes, respectively. The first ATG in the transcribed sequence is in boldface type. Gaps were introduced to optimize the alignment. The 5′ nucleotide of the cloned cDNAs is in boldface type and underlined.

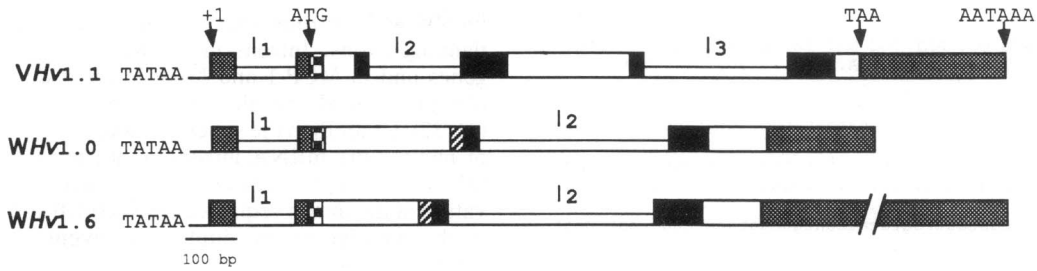

FIG. 3. Schematic comparison of gene structures of V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6 shown to scale. Checkered boxes indicate the sequence encoding the signal peptide. Black boxes indicate the sequence for the cysteine motif that is interrupted by an intron. Hatched boxes in W*Hv*1.0 and W*Hv*1.6 represent the precysteine, nearly identical, 27 amino acid residues (19). Cross-hatched boxes represent 5′ and 3′ UTRs. The interruption indicates an additional 300 bp in the 3′ UTR of W*Hv*1.6. The schematic of each gene ends at the polyadenylylation signal AATAAA.

icant levels of sequence identity to those of W*Hv*1.0 and W*Hv*1.6 (Fig. 2 and Table 1). The 3′ UTR of V*Hv*1.1 could not be confidently aligned with those of W*Hv*1.0 and W*Hv*1.6. The 5′ UTR of each gene is interrupted by intron 1 at a comparable position relative to the TATA box (Figs. 2 and 3). The 5′ UTR of V*Hv*1.1 is more similar to that of W*Hv*1.6 than to that of W*Hv*1.0 (Table 1). Introns 1 in W*Hv*1.0 and W*Hv*1.6 are 87.5% identical, whereas intron 1 of V*Hv*1.1 is only 60% identical to those of W*Hv*1.0 and W*Hv*1.6. The sequence identity of introns 1 of W*Hv*1.0 and W*Hv*1.6 is close to the identity of their 5′ UTRs (Fig. 2 and Table 1). Introns 2 and 3 of V*Hv*1.1 and introns 2 of W*Hv*1.0 and W*Hv*1.6 interrupt the ORF and immediately precede a cysteine codon (Figs. 1, 3, and 5*a*) (19). Introns 2 of W*Hv*1.0 and W*Hv*1.6 are 92% identical in sequence (Fig. 5*a* and Table 1). Introns 2 and 3 of V*Hv*1.1, when compared with each other or with introns 2 of W*Hv*1.0 and W*Hv*1.6, show significant similarity only at the 5′ and 3′ ends of the introns (Fig. 5*b* and Table 1). These similar sequences are not shared with intron 1 sequences of the three genes. The similarity in gene structure and nucleotide sequence of the three genes is best explained by common ancestry. The high similarity in the intron sequences between W*Hv*1.0 and W*Hv*1.6 suggests that these two genes may have evolved by a relatively recent gene duplication event.

Analysis of the Predicted ORFs of V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6. The protein encoded by V*Hv*1.1 has a putative 16-amino acid signal peptide and contains 6% cysteine residues arranged in a pattern (C . . . C . . . CC . . . C . . . C) characteristic of small peptide neurotoxins (ω-conotoxins) of carnivorous snails (22) and scorpions (30). V*Hv*1.1 has two such cysteine motifs, each consisting of 41 amino acids (Figs. 1 and 4). Introns 2 and 3 of V*Hv*1.1 occur immediately 5′ to the codon of the second cysteine of the motifs (Fig. 1). The V*Hv*1.1 ORF also has three clusters of amino acids (Fig. 1) shared with a wasp gene encoding a putative translation initiation factor eIF-4A (B.A.W. and M.D.S., unpublished data). These clusters (Fig. 1) are characteristic of the "DEAD" protein family of ATP-dependent RNA helicases with eIF-4A as their prototype (28). The order of occurrence

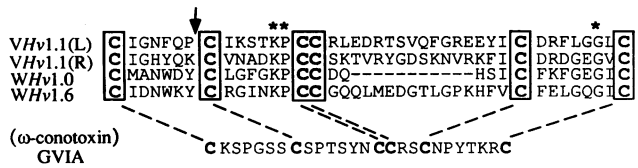

FIG. 4. Alignment of the cysteine motifs from the three viral genes V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6. The cysteine residues are in boldface type and boxed. Stars indicate the other conserved amino acid residues. Dashes indicate the 11-amino acid deletion from the motif in W*Hv*1.0. GVIA is the prototype ω-conotoxin (22) included here to highlight the cysteine residue pattern in common with the cysteine motifs of the three *Cs*PDV proteins.

of these clusters within the viral ORF is the same, but the inter-motif spacing is reduced when compared with the other members of the DEAD proteins. The functional significance of these putative DEAD motifs is not known.

Proteins encoded by W*Hv*1.0 and W*Hv*1.6 are secreted and also contain a single cysteine motif (Fig. 4); the coding sequence of each motif is interrupted by the second intron at the same position as in V*Hv*1.1 (Figs. 3 and 5*a*) (19). The signal sequences of the three proteins are similar (Table 1). The four cysteine motifs share only three identical intercysteine amino acid residues, but the cysteine motif in W*Hv*1.0 has a deletion of 11 amino acids (Fig. 4). The nucleotide sequences encoding the two cysteine motifs in V*Hv*1.1 are as similar to each other as they are to the sequences encoding the cysteine motif in either W*Hv*1.0 or W*Hv*1.6 (Table 1). However, the two V*Hv*1.1 cysteine motifs are more similar to each other at the amino acid level than to the cysteine motif in either of the other two genes (Table 1). The cysteine motifs of W*Hv*1.0 and W*Hv*1.6 are significantly more similar to each other at the nucleotide and amino acid levels than they are to the two V*Hv*1.1 cysteine motifs (Table 1). This is consistent with the hypothesis that W*Hv*1.6 and W*Hv*1.0 may have evolved by a gene duplication event and the two cysteine motifs of V*Hv*1.1 may have evolved by an intragenic duplication event.

The sequences encoding the cysteine motifs in W*Hv*1.0 and W*Hv*1.6 are 76% identical, but the two motifs are only 58% identical in amino acids (Table 1). The 76% identity in the sequences encoding the motif reflects 22 mutations, including 21 nucleotide substitutions and a single deletion of 33 bp in W*Hv*1.0 (Fig. 5*a*) (19). Only 6 of the 21 substitutions are third-position nucleotide changes; 5 of these changes are silent. Two of the silent substitutions are in the codons of the conserved cysteines. The remaining 15 substitutions consist of 5 first-, 2 second-, and 4 multi-position substitutions, all of which result in amino acid changes.

DISCUSSION

The three viral genes V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6 describe a cysteine-rich gene family with the following major characteristics. First, they have a common gene structure with introns of similar sequences at conserved positions. Second, their 5′ UTRs and signal peptides show significant levels of identity. Third, their predicted ORFs contain cysteine motifs characteristic of the ω-conotoxins. The cysteine motifs consist of hypervariable intercysteine amino acids and invariant cysteine residues. Fourth, transcripts of these genes are abundant in *H. virescens* after natural parasitization or injection of purified virus. Proteins encoded by W*Hv*1.0 and W*Hv*1.6 contain a precysteine domain of 26 identical amino acids followed by a serine-to-threonine change immediately N-terminal to the cysteine motifs (19). The protein encoded by V*Hv*1.1 lacks this precysteine do-

a

```
WHv1.0  TGC ATG GCT AAT TGG GAT TAT gtgagtacatt---gctgtc
WHv1.6  ... ..C .A. ... ... A.A ... ...........att......

        acttctccgttaagtgt------------------------------
        ................aatatcatatgcccctgtacctcacgctaaa

        ----------ttcattcttgtatttattgtcattcaatatatcaccat
        cattagttgt.............c.c.....................

        catcgtcgataaatagtacattaatgacaataagtcaaataattgagt
        ..................g...........t...............

        gtccac-tattacaataatatgcaatctgtaataacaa-aacatgatt
        ......c................t.........t..t........a

        aagg-gcaataccatcatccttgaacacaaacaaatcgtttaatatat
        g...ac...a..t....-......g........g......g.......

        tatttttccacactgcatttgtagacagtattcaacggccctttccg
        ....................................t....c...

        atccaatgaatggttacattccgc---aatagtcattgttaattttgt
        g..................t..tgc.........c..a.....a..

        tacatacgagggatacaaacagttctgttttgtctaccacggtaaaaa
        a.at-------------------------------------------

        tctaattccctaattacatggcgcatagtgacttatttaatgaacgtc
        ------------------------------------------....

        aacaaaattaatgttatattgtatgctttcag TGC CTG GGT TTC
        .....................t.at...... ..T .G. ... A..

        GGC AAG CCC TGC TGT GAT CAG --- --- --- --- ---
        AA. ... ... ... ... .GA ... CAA CTA ATG GAA GAT

        --- --- --- --- --- --- CAT TCA ATT TGC TTC AAA
        GGT ACG CTC GGA CCG AAG ..C .TC G.. ..T ..T G..

        TTC GGC GAG GGG ATC TGT
        C.. ..A C.. ... ... ...
```

b

```
                ****   ****  *****  *******
VHv1.1/I2  gtgagtacatcattgctagcactt---tgttaa //113 bp//
VHv1.1/I3  gtgaa-acat---tgcta--------ccgttcc //230 bp//
WHv1.0/I2  gtgagtacat---tgctgtcacttctccgttaa //361 bp//
WHv1.6/I2  gtgagtacattattgctgtcacttctccgttaa //337 bp//

        **  ***    ***** ******** ****
        tgtcaact---tcaattaagataatattgtcgtgtatg-tttcag
        tg--aacttgatcaac---attaattttattttctatattttcag
        tg--aacg---tcaacaaaattaatgttattgtatgctttcag
        ------cg---tcaacaaaattaatgttattgtttatttttcag
```

FIG. 5. (*a*) Alignment of intron 2 and the immediately flanking exons encoding the cysteine motifs in W*Hv*1.0 (this study) and W*Hv*1.6 (19). Dashes indicate deletions; dots indicate identity. Exon sequences are uppercase and intron sequences are lowercase. The cysteine codons are in boldface type. The nucleotide substitutions in the codons of the cysteine motifs are shown. (*b*) Alignment of the 5′ and 3′ termini of introns 2 and 3 of V*Hv*1.1 (Fig. 1) and introns 2 of W*Hv*1.0 (*a*) and W*Hv*1.6 (*a*) (19). Sequences corresponding to the consensus splicing signals (29) are in boldface type. Stars indicate nucleotides conserved in at least three introns; dashes indicate deletions compared with the other sequences. Gaps were introduced to optimize alignment.

main. The fact that V*Hv*1.1 contains two cysteine motifs and lacks this highly conserved domain may reflect the divergence of two subfamilies within the *Cs*PDV cysteine-rich gene family with related, but distinct, functions.

W*Hv*1.0 and W*Hv*1.6 appear to be the result of a relatively recent gene duplication event, despite the presence of highly divergent N and C termini of the mature proteins that are only 16% and 38% identical, respectively (19). Introns 1 and the flanking exon sequences, which encode the 5′ UTR and the signal peptides, are over 85% identical. Introns 2 are 92% identical, but the immediately flanking exon sequences encoding the cysteine motifs are only 76% identical. To our knowledge, this is the first example of members of a gene family with introns that are more conserved than their flanking exons. Introns in the α1 and α2 human α-globin genes and in MCP-1 and MCP-2 rabbit defensin genes are almost identical, but the exons are also equally conserved (31, 32). Other than the splicing signals, the primary sequence of nuclear pre-mRNA introns is not important for splicing (33); therefore, it is reasonable to conclude that the conservation of the intron sequences in W*Hv*1.0 and W*Hv*1.6 is due to the putative recent duplication event, rather than to a functional role.

The cysteine motifs of V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6 are characteristic of small peptide neurotoxins (ω-conotoxins) of carnivorous snails (22) and scorpions (30). The conservation of the cysteine motif was a clear indication that the three viral genes may be related despite the fact that only three intercysteine amino acid residues are identical in all four motifs (Fig. 4). Compared with ω-conotoxins, the proteins encoded by W*Hv*1.0 and W*Hv*1.6 are larger and do not contain a cleaved propeptide (21). Also, the cysteine motif is not as proximal to the C terminus of the proteins. A baculovirus gene more typical of ω-conotoxins has been reported (34). By analogy to conotoxins, the cysteine motifs of V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6 may function as a ligand for surface proteins of target cells.

The pattern of codon substitutions in the sequence encoding the polydnavirus cysteine motifs deviates from the predictions of the neutral theory of mutation (35). Also, it was unexpected to find that intron 2 sequences in W*Hv*1.0 and W*Hv*1.6 are more conserved than the immediate flanking exon sequences encoding the cysteine motifs. This may suggest selection for amino acid substitutions that do not alter the conserved structure of the cysteine motif. By analogy to the ω-conotoxins (36), whose members target different receptor subtypes, variability of the intercysteine amino acid residues may alter the target specificity of the motif without disrupting the conserved disulfide bridges.

A pattern of codon substitutions similar to that of the W*Hv*1.0/W*Hv*1.6 cysteine motifs exists in the antigen binding sites in the class I and II proteins encoded by the major histocompatibility complex; this pattern was explained by overdominant selection or positive Darwinian selection (37, 38). In these cases, the number of synonymous substitutions per synonymous site (d_S) is less than the number of nonsynonymous substitution per nonsynonymous site (d_N). In contrast, d_S is significantly greater than d_N in the remaining sequences not thought to be under positive selective pressure (37, 38). When the cysteine motifs of W*Hv*1.0 and W*Hv*1.6 or the L and R motifs of V*Hv*1.1 are compared with each other, d_S is only about 2-fold greater than d_N (data not shown). However, when the codons invariant in these four motifs are excluded from the analysis, d_S and d_N become comparable. This may indicate overdominant selection that could have been masked by inclusion in the analysis codons of amino acids not involved in binding to the putative target. Alternatively, if no positive selection pressure underlies the intercysteine amino acid hypervariability, this may suggest that these changes are under little or no negative selective pressure. It is conceivable that both of these mechanisms are at work on different amino acids in this domain.

In conotoxins, hypervariable regions are superimposed over a constant backbone of disulfide bridges, which results in the generation of high-affinity ligands with different specificity (22, 36). Recently, conserved cysteine-rich regions were proposed to be the surface receptor binding domain of a family of malaria erythrocyte-binding proteins (39). Also, a soluble "transformation" factor(s) has been proposed to explain altered behavior of plasmatocytes of parasitized insect larvae (10) that is correlated with suppression of encapsulation, the cellular immune response of insects. Because the *Cs*PDV cysteine-rich gene family encodes secreted

Table 1. Percentage nucleotide (nt) and amino acid (aa) identity of select regions of the cysteine-rich genes

Comparison	5′ UTR* (nt)	Signal peptide† nt	Signal peptide† aa	Cysteine motif‡	Cysteine motif‡ nt 90	Cysteine motif‡ nt 123	Cysteine motif‡ aa 30	Cysteine motif‡ aa 41	3′ UTR§ (nt)	Intron 1¶ (nt)	Intron 2‖ (nt)	Intron 3 (nt)
V*Hv*1.1/V*Hv*1.1	—	—	—	(L/R)	60	48	50	39	—	—	40**	—
V*Hv*1.1/W*Hv*1.0	56	63	56	(L)	54	—	40	—	45	60	40**	40**
				(R)	60	—	40	—				
V*Hv*1.1/W*Hv*1.6	63	69	63	(L)	56	49	37	29	46	60	40**	40**
				(R)	57	48	40	29				
W*Hv*1.0/W*Hv*1.6	87	90	94		76	—	58	—	83	88	92	—

*5′ UTR length in V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6 is 72, 79, and 77 bp, respectively.
†The putative signal peptide length is 16 amino acids.
‡The length of the cysteine motif in W*Hv*1.0 is 30 amino acids lacking 11 amino acids in loop 3 (Fig. 4). This deletion is considered a single mutation in calculating percentage similarity with the other motifs. Also presented are the percentage identities between the cysteine motifs of V*Hv*1.1 and W*Hv*1.6, including and excluding the corresponding 11 amino acids.
§3′ UTR length in V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6 is 300, 226, and 803 bp, respectively. The difference in length between W*Hv*1.0 and W*Hv*1.6 can be accommodated by two insertions (19).
¶Length of intron 1 of V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6 is 122, 107, and 117 bp, respectively.
‖Length of intron 2 of V*Hv*1.1, W*Hv*1.0, and W*Hv*1.6 is 193, 429, and 409 bp, respectively.
**Comparison of introns 2 and 3 of V*Hv*1.1. Intron 3 is 295 bp long. Also, intron 3 of V*Hv*1.1 was compared with intron 2 of W*Hv*1.0 and W*Hv*1.6 (last column). These introns show significant similarity at their 5′ and 3′ termini beyond the predicted splicing signals (Fig. 5*b*).

proteins with potentially different binding specificity that are apparently produced throughout the parasitization period, their possible roles as key factors for inducing host immunosuppression and developmental arrest should be investigated.

We thank Dr. Masatoshi Nei, Dr. Paul Feurst, and Dr. Rodney Honeycut for valuable discussions on evolutionary analysis. The computer program used to calculate d_S and d_N was a gift from Dr. Masatoshi Nei. We acknowledge the excellent technical help of Keith Barnhart, Debbie Ellis, and Valery Zackery. We thank Mr. Tatsuya Ota for performing the statistical analysis on the sequences encoding the cysteine motifs. This work was supported in part by National Science Foundation Grant IBN-9119827 and funds from Texas Experimental Station Project TEXO 8087.

1. Salt, G. (1970) *The Cellular Defense Reactions of Insects* (Cambridge Univ. Press, London), pp. 1–24.
2. Vinson, S. B. & Iwantsch, G. F. (1980) *Q. Rev. Biol.* **55**, 143–165.
3. Ables, J. R. & Vinson, S. B. (1982) *Entomophaga* **26**, 453–458.
4. Stoltz, D. B., Guzo, B., Belland, E. R., Lucaroti, C. J. & MacKinnon, E. A. (1988) *J. Gen. Virol.* **69**, 903–907.
5. Kitano, H. (1986) *J. Insect Physiol.* **32**, 369–375.
6. Tanaka, T. (1987) *J. Insect Physiol.* **33**, 413–420.
7. Edson, K. M., Vinson, S. B., Stoltz, D. B. & Summers, M. D. (1981) *Science* **211**, 582–583.
8. Theilmann, D. A. & Summers, M. D. (1986) *J. Gen. Virol.* **67**, 1961–1969.
9. Dover, B. A., Davies, D. H., Strand, M. R., Gray, R. S., Keeley, L. L. & Vinson, S. B. (1987) *J. Insect Physiol.* **33**, 333–338.
10. Davies, D. H., Strand, M. R. & Vinson, S. B. (1987) *J. Insect Physiol.* **33**, 143–153.
11. Rizki, R. M. & Rizki, T. M. (1991) *J. Exp. Zool.* **257**, 236–244.
12. Tanaka, T. (1986) *Mem. Fac. Sci. Kyoto Univ. Ser. Biol.* **11**, 1–27.
13. Stoltz, D. B. & Guzo, D. (1986) *J. Insect Physiol.* **32**, 377–388.
14. Webb, B. A. & Summers, M. D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4961–4965.
15. Krell, P. J., Summers, M. D. & Vinson, S. B. (1982) *J. Virol.* **43**, 859–870.
16. Theilmann, D. A. & Summers, M. D. (1988) *Virology* **167**, 329–341.
17. Fleming, J. G. W., Blissard, G. W., Summers, M. D. & Vinson, S. B. (1983) *J. Virol.* **48**, 74–78.
18. Blissard, G. W., Vinson, S. B. & Summers, M. D. (1986) *J. Virol.* **169**, 78–89.
19. Blissard, G. W., Smith, O. P. & Summers, M. D. (1987) *Virology* **160**, 120–134.
20. Summers, M. D. & Smith, G. E. (1987) *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Agricultural Experiment Station Bulletin 1555 (Tex. Agric. Exp. Stn., College Station, TX).
21. Blissard, G. W., Theilmann, D. A. & Summers, M. D. (1989) *Virology* **169**, 78–89.
22. Olivera, B. M., Rivier, J., Clark, C., Ramilo, C. A., Corpuz, G. P., Abogadie, F. C., Mena, E. E., Woodward, S. R., Hillyard, D. R. & Cruz, L. J. (1990) *Science* **249**, 257–263.
23. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Plainview, NY).
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
25. Devereux, J., Huebberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–397.
26. Czelusniak, J., Goodman, M., Moncriff, N. & Kehoe, S. M. (1990) *Methods Enzymol.* **183**, 601–615.
27. Von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
28. Schmid, S. R. & Linder, P. (1991) *Mol. Microbiol.* **6**, 283–292.
29. Senapathy, P., Shapiro, M. B. & Harris, N. L. (1990) *Methods Enzymol.* **183**, 252–278.
30. Fontecilla-Camps, J. C. (1989) *J. Mol. Evol.* **29**, 63–67.
31. Liebhaber, S. A., Goossens, M. & Kan, Y. W. (1981) *Nature (London)* **290**, 26–29.
32. Ganz, T., Rayner, J. R., Valore, E. V., Tumolo, L., Talmadge, K. & Fuller, F. H. (1989) *J. Immunol.* **143**, 1358–1365.
33. Padget, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150.
34. Eldridge, R., Li, Y. & Miller, L. K. (1992) *J. Virol.* **66**, 6563–6571.
35. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
36. Woodward, S. R., Cruz, L. J., Olivera, B. M. & Hillyard, D. R. (1990) *EMBO J.* **9**, 1015–1020.
37. Hughes, A. L. & Nei, M. (1988) *Nature (London)* **335**, 167–170.
38. Hughes, A. L. & Nei, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 958–962.
39. Adams, J. H., Sim, B. K. L., Dolan, S. A., Fang, X., Kaslow, D. C. & Miller, L. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7085–7089.