Solution Structure of the Carboxyl-Terminal Cysteine-Rich Domain of the VHv1.1 Polydnaviral Gene Product: Comparison with Other Cystine Knot Structural Folds†

Jerrod Einerwold,‡,§ Mahesh Jaseja,‡,§ Kenneth Hapner,‡ Bruce Webb,|| and Valérie Copié*,‡

Department of Chemistry and Biochemistry, Montana State University, 108 Gaines Hall, Bozeman, Montana 59717, and Department of Entomology, University of Kentucky, Lexington, Kentucky 40546

Received July 18, 2001; Revised Manuscript Received September 28, 2001

ABSTRACT: Polydnaviruses are an unusual group of insect viruses that have an obligate symbiotic association with certain parasitic wasps. These viruses are transmitted with the wasp egg during oviposition into lepidopteran insects, enabling the survival and development of the egg inside the host larvae. We report the three-dimensional structure of a novel polydnaviral cysteine-rich motif (cys-motif), identified as the carboxyl-terminal domain of a two cys-motif containing polydnaviral VHv1.1 gene product, abbreviated "C-term VHv1.1". This 65-residue domain was identified experimentally by limited proteolysis of the full-length protein and was subsequently cloned in a bacterial expression system for NMR studies. The C-term VHv1.1 3D structure was determined in solution by two-dimensional ^{1}H NMR spectroscopy. Calculation of the structure was based on a total of 300 upper distance restraints and 20 dihedral angle constraints, and resulted in an ensemble of 25 representative conformers with an average rmsd of 0.47 Å from the mean structure for core backbone atoms. The protein core is made of a four β-strand scaffold held together in a compact structure by three disulfide bonds, which form a cystine knot. The four β-strands are arranged in an unusual configuration to form a triple-stranded β-sheet and double-stranded β-sheet. Comparison with other classes of cystine knots provides indication that C-term VHv1.1 represents a new and distinct cytine knot motif. This analysis provides a structural basis for interpretation of the genetic and amino acid sequence data classifying polydnavirus gene products as members of cysteine-rich protein families.

Polydnaviruses have developed a symbiotic relationship with parasitic Hymenoptera wasps and are essential for the successful parasitization of lepidopteran insects following wasp oviposition (*1*−*5*). Successful parasitization is accompanied by suppression of the insect host's multicellular immune response, encapsulation of foreign objects by hemocytes and developmental alteration of the host (*6*−*8*). Factors contributing to such effects include wasp venom, ovarian proteins, and the symbiotic group of polydnaviruses (*9*, *10*). The endoparasitic *Campoletis sonorensis* wasp injects polydnaviruses (CsIV)[1] of the ichnovirus genera, along with its egg, venom, and ovarian proteins, during parasitization of *Heliothis virescens* larvae. Species-specific expression of CsIV genes protects the wasp egg by selectively altering host physiology (*11*−*19*).

The most extensively characterized CsIV genes are members of the *cys* (cysteine-rich) gene family. Four genes belonging to this family, VHv1.1, VHv1.4, WHv1.0, and WHv1.6, have been reported (*13*, *20*−*23*). Sequence analysis reveals that these genes share a common gene structure, and code for two distinct protein subfamilies, labeled W- and V- (*23*). The CsIV *cys* gene family encodes for proteins whose primary sequences contain six invariable cysteines flanked by highly variable amino acid residues (Figure 1) (*23*). These characteristics are reminiscent of the structural arrangement of conotoxin genes, which code for absolutely conserved cysteine residues separated by highly variable intercysteine residues (Figure 1) (*24*, *25*).

The viral proteins encoded by WHv1.0 and WHv1.6 contain a single cys-motif, and possess a highly conserved 26-residue precysteine domain that is lacking in VHv1.1 and VHv1.4 (*26*, *27*). In contrast, the V-subfamily possesses proteins with two cys-motifs, consisting of ∼41 amino acids

† This work was funded in part by a Petroleum Research Fund, ACS Grant 32853-AC4, and the United States Department of Agriculture, USDA Grant 9803843 to V.C, and an NSF MCB 0094403 to B.A.W. The NMR experiments were recorded at Montana State University on a DRX600 spectrometer, purchased in part with funds from the NIH shared instrumentation grant program (SIG Grant 1-S10RR13878-01), and the NSF-EPSCOR program for the State of Montana.

* To whom correspondence should be addressed. E-mail: vcopie@chemistry.montana.edu. Phone: (406) 994-7244. Fax: (406) 994-5407.

‡ Department of Chemistry and Biochemistry.

§ The first two authors have contributed equally to the work presented here.

|| Department of Entomology.

[1] Abbreviations: NMR, nuclear magnetic resonance; cys-motif, cysteine-motif; CsIV, *Campoletis sonorensis* polydnavirus; C-term VHv1.1, carboxyl terminal cysteine-rich domain of the VHv1.1 polydnaviral gene product; MOI, multiplicity of infection; PBS, phosphate-buffered saline; Amp, ampicillin; Kan, kanamycin; IPTG, isopropyl β-thiogalactoside; MALDI-TOF, matrix-assisted laser desorption/ionization mass spectrometry; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; DQ-COSY, double-quantum filtered COSY; CNS, Crystallography & NMR System.

10.1021/bi011499s CCC: $20.00
Published on Web 11/07/2001

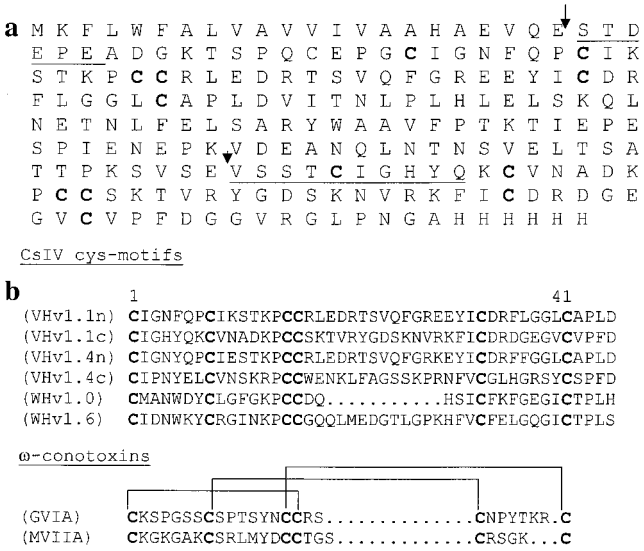

FIGURE 1: (a) Amino acid sequence of the full-length VHv1.1 polydnaviral protein. Arrows indicate cleavage sites where endoproteinase Glu-C (V8) enzyme cleave the protein into two major fragments. Underlined residues indicate stretches of amino acids identified by N-terminal sequencing data as the N-termini of the two fragments produced by V8 limited proteolysis. The N-terminal sequencing data were corroborated with molecular weight determination of the fragments by MALDI-TOF mass spectrometry. (b) Amino acid sequence alignment of cys-motifs identified in VHv1.1, VHv1.4, WHv1.0, and WHv1.6, polydnaviral proteins of the CsIV *cys* gene family. VHv1.1n/VHv1.1c and VHv1.4n/VHv1.4c refer to the N- and C-terminal cys-motifs of the two cys-motif containing VHv1.1 and VHv1.4 proteins, respectively. The last two amino acid sequences originate from two ω-conotoxin polypeptides, which are members of the ICK cystine knot superfamily. The sequence alignment exemplifies the analogous spacing of conserved cysteines among the CsIV cys-motif proteins and the smaller ω-conotoxin polypeptides.

each (*13*, *26*). The fact that VHv1.1 and VHv1.4 contain two cys-motifs and lack the precysteine domain may reflect functional divergence of the two subfamilies within the CsIV cysteine-rich family (*13*, *26*, *28*). By analogy with the evolutionary pressures observed on conopeptide structures and functions, evolution of cys-motif subfamilies could be the result of natural selection and the need for polydnavirus genes to evolve quickly as their hosts develop resistance genes (*29*). In an effort to elucidate selective forces driving the evolution of the cys-motif polydnaviral genes, patterns of synonymous and nonsynonymous nucleotide substitution have been recently analyzed (*30*). Such a study has identified nucleotide positions within the CsIV cys-motif coding regions that are highly conserved, under neutral selection, or under diversifying selection (*30*). This analysis has raised the prospect that nucleotide positions that are highly conserved or under neutral natural selection may correlate with amino acids within the polydnaviral protein sequences that are important for protein folding and structural stability (*30*). In contrast, nucleotide positions that are under diversifying selection pressure could correspond to amino acid residues that are important for insect host parasitism function (*30*).

Despite extensive analyses of the polydnaviral proteins at the genetic and primary amino acid sequence levels, the three-dimensional structure of these proteins is unknown. In an attempt to provide molecular insights into polydnaviral cys-motif protein function, we have undertaken the three-dimensional structural elucidation of the CsIV VHv1.1 gene product. Using limited proteolysis (*31*), we have identified two independently folded domains in the protein. In this manuscript, we report the high-resolution, three-dimensional structure of the VHv1.1 C-terminal cysteine-rich domain, abbreviated "C-term VHv1.1", in solution. The protein's 3D structure was solved by two-dimensional ^{1}H NMR spectroscopy. We describe similarities and differences between C-term VHv.1.1's 3D fold and the structures of known cystine knot families. We close by presenting evidence for a relationship between the structural features observed in C-term VHv1.1 and the genetic evidence linking nucleotide positions within the CsIV cys-motifs to regions under different selective pressures.

MATERIALS AND METHODS

Production of VHv1.1 Protein in High Five Insect Cells. Full-length VHv1.1 protein was produced initially from a baculovirus-insect cell expression construct developed in Dr. Webb's laboratory (*12*). This construct was identical to previously reported VHv1.1 vector (*12*) except that an asparagine residue, N-glycosylated in vivo, was replaced by a glutamine residue, to express a deglycosylated mutant VHv1.1 protein. For small scale protein expression, eight 75 cm^2 tissue culture flasks, containing 10 mL of fresh Ultimate Insect serum-free media (Invitrogen), were each seeded with a 5 mL culture of High Five cells (Invitrogen) that had been grown to confluency in a 25 cm^2 tissue culture flask. The 15 mL cultures were grown to 50% confluency and infected with a multiplicity of infection (MOI) of 10 (Invitrogen). Incubation media was collected 64 h post infection and centrifuged. In preparation for nickel affinity protein purification, the supernatant containing secreted VHv1.1 protein was concentrated to 10 mL, followed by dialysis with 1×-binding buffer (20 mM Tris-HCl pH 7.9, 500 mM NaCL, 5 mM imidazole). The sample was applied to a 1 mL of NTA-Ni column material (Qiagen), and washed with 1×-binding buffer containing 40 mM imidazole. The VHv1.1 protein was eluted from the column using a 20 mM Tris-HCl pH 7.9, 500 mM NaCL, 600 mM imidazole solution, followed by dialysis in phosphate-buffered saline (PBS: 25 mM phosphate pH 7.0, 150 mM NaCl).

Limited Proteolysis and Protein Fragment Identification. In preparation for limited proteolysis experiments, 25 μg of sequencing grade Endoproteinase Glu-C (V8) (Sigma Co.) was reconstituted in 20 μL of water, aliquoted into 5 μL samples, frozen in liquid nitrogen, and stored at −80 °C. Limited proteolysis was carried out on ice, as described (*31*). For the range-finding stage of the experiment, six 25 μL aliquots of full-length VHv1.1 protein were prepared. A 5 μL aliquot of the Endoproteinase Glu-C (V8) frozen stock was thawed on ice, and diluted in a series of 10-fold dilutions ranging from 10^1 to 10^5. In tube 1, 25 μL of PBS were added to 25 μL of substrate, no protease, as a control. In tubes 2−6, 25 μL of an enzyme dilution were added to 25 μL of substrate. After a 30 min incubation period, reactions were stopped by adding 30 μL of SDS-reducing sample buffer and boiling for 5 min. The samples were analyzed on SDS−PAGE, using a 15% reducing gel. Results from the range-finding experiment established that the 100-fold V8 enzyme dilution was optimal for the time-course experiment (*31*). In this second stage, an equal volume of a 100× dilution of

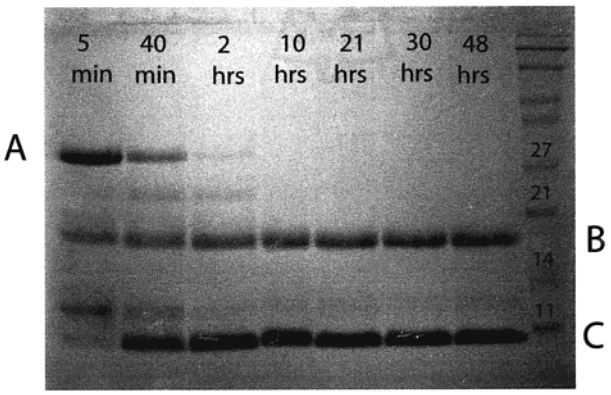

FIGURE 2: SDS−PAGE gel electrophoresis results of limited proteolysis experiments on full-length VHv1.1 using endoproteinase Glu-C (V8) protease. The A band corresponds to full length, uncleaved protein. The B and C bands correspond to two resulting protein fragments, which appear shortly after protease cleavage is initiated, and persist for 48 h following the initiation of limited protease digestion (*31*). The identity of the B and C bands was confirmed both by mass spectrometry and N-terminal sequencing. B and C were identified as the N-terminal and C-terminal cys-motif domains of VHv1.1, respectively.

V8 enzyme was mixed with VHv1.1 protein stock, and 20 μL aliquots were collected at 5 min, 40 min, 2 h, 5 h, 10 h, 21 h, 30 h, and 48 h (Figure 2). Analogous limited proteolysis experiments were carried out with trypsin (data not shown). Protein fragments were separated by HPLC reverse-phase chromatography and analyzed by MALDI-TOF mass spectrometry. For N-sequencing data analyses, the C-term, His-tag-containing VHv1.1 cys-motif was purified from the N-terminal protein fragment by nickel affinity chromatography. The exact amino acid sequences of the fragments were identified using a combination of molecular mass information from mass spectrometry analyses and N-terminal sequencing data.

Expression of VHv1.1 in a baculovirus-insect cell expression system permitted the unambiguous identification of the amino acid sequence corresponding to the C-term cys-motif of VHv1.1 However, the eukaryotic expression system could not produce sufficient amounts of protein for 2D ^{1}H NMR structural studies. For the latter, the cDNA-encoding C-term VHv1.1 was thus cloned into a prokaryotic expression system.

Vector Construction for Bacterial Expression of the C-Terminal Motif. A VHv1.1 cDNA construct, pVX900 clone (*12, 23*), was used as template in standard PCR reactions with cloned Pfu polymerase (Stratagene). *Nco*1 and *Xho*1 restriction enzyme sites were engineered into the forward and reverse primers respectively, to permit insertion of the C-term VHv1.1 cDNA into a pET-32b(+) vector (Novagen Inc.). A stop codon was engineered upstream from the *Xho*1 site to exclude the vector's C-terminal His-Tag. DNA sequencing was performed at the Iowa State DNA sequencing facility to confirm proper in-frame insertion. The pET-32b(+) vector allowed for expression of an N-terminal fusion protein, thioredoxin, to C-term VHv1.1 (*32*).

Bacterial C-Terminal VHv1.1 Protein Expression, Purification, and NMR Sample Preparation. One-liter cell cultures were initiated by inoculating 500 mL (×2) of LB-Amp (100 μg/mL)/Kan (30 μg/mL) with 5 mL of overnight culture. The cells were grown at 37 °C for ~3 h, then moved to a 16−17 °C growth chamber. The cells were harvested 20 h post-IPTG induction, and the pellets resuspended in 40 mL of 1×-binding buffer. C-term VHv1.1 was isolated and purified using nickel affinity and gel filtration chromatography. The fusion partner + His-tag was cleaved by enterokinase and separated from C-term VHv1.1 by reapplication of the protein solution through a nickel affinity column. Purified C-term VHv1.1 was dialyzed extensively against a 25 mM phosphate-buffered saline solution at pH 5.7, and concentrated to 750 μL using a 10 mL stir cell and a YM3 ultrafiltration membrane (Amicon, Inc.). Final NMR sample conditions consisted of a ~1 mM protein solution, in 25 mM sodium phosphate pH 5.7, 100 mM NaCl, 1 mM EDTA, and 0.01% sodium azide. For H_2O NMR samples, 5% D_2O was added for deuterium lock signal. For D_2O samples, the protein was lyophilized and redissolved in 99.996% D_2O (Cambridge Isotopes Laboratories, Cambridge, MA).

NMR Spectroscopy. ^{1}H NMR spectra were recorded on a Bruker DRX 600 spectrometer at Montana State University. To assign overlapping resonances in 2D ^{1}H NMR spectra, experiments were recorded at 315 and 310 K for a ~1 mM sample solution at pH 5.7, and at a temperature of 305 K for a sample solution at pH 5.0. Two-dimensional (2D) NMR data were recorded in phase-sensitive mode using the States-TPPI method for quadrature detection in the indirect t_1 dimension (*33*). Experiments typically made use of pulsed field gradients to suppress water signals and artifacts. Experiments on protein samples in H_2O made use of the WATERGATE pulse sequence for water suppression (*34*). 2D homonuclear NOESY spectra (*35*) were recorded with a mixing time of 200 ms. TOCSY spectra (*36*) were recorded using MLEV-17 spin-lock sequence with a 10 kHz RF field and a mixing time of 62 ms. Typically, spectra were acquired with 448−512 t_1 increments, 2048 data acquisition points, and a relaxation recovery delay of 1.5 s, with spectral widths of 6613.7 Hz. For DQF-COSY (*37*) spectra, 1024 t_1 increments were acquired. Spectra were recorded with 96 scans/t_1 increment for NOESY, 32 scans/t_1 increment for DQF-COSY, and 16 scans/t_1 increment for the TOCSY experiments. Spectra were processed using XWIN NMR (Bruker Inc.), Version 2.6. All spectra were zero-filled once in both dimensions, and sine-bell apodization functions, phase shifted by 45−90°, were applied in both dimensions prior to Fourier transformation. The three-bond $^3J_{HNH\alpha}$ coupling constants were measured from DQF-COSY spectra. Slowly exchanging amide protons were identified by dissolving the lyophilized protein in D_2O at pH 5.0 and 305 K, and recording 1D-^{1}H and 2D-TOCSY spectra after dissolution. ^{1}H chemical shift assignments were obtained for residue segments ranging from Thr 6 to Arg 55. No chemical shift information was obtained for the N- and C-termini segments of C-term VHv1.1, due to resonance overlaps or lack of NOE signals in these flexible regions. Practical details about the standard 2D ^{1}H NMR experiments used to assign a small protein like C-term VHv1.1 have been reviewed (*38*).

Structure Calculations. Distance restraints for structure calculations were based on 280 NOE cross-peaks assigned in 2D ^{1}H-NOESY spectra recorded at 315 K, with a 200 ms NOE mixing time. NOE restraints were classified into three categories: strong (1.8−2.5 Å), medium (1.8−3.7 Å), and weak (1.8−4.7 Å). Twenty four additional distance restraints

were included for 12 hydrogen bonds identified through both characteristic β-sheet NOE patterns and $^1H/^2H$ amide exchange experiments, and were given bounds of 1.8−2.3 Å (H−O) and 2.8−3.3 Å (N−O). Dihedral angle ϕ constraints were obtained from *J*-coupling constant measurements in DQ-COSY spectra. Structure calculations were performed using simulated annealing protocols in torsion-angle and Cartesian space of the Crystallography & NMR (CNS) program (*39*, *40*). The molecular dynamic scheme used in CNS consisted of the following stages: (a). Heating in torsion angle space at 50 000 K for 45 ps with the energy constant for the van der Waals parameters scaled by 0.1; (b) cooling in torsion angle space to 2000 K for 300 ps with ramping of the van der Waals parameters to full scale; (c) cooling in Cartesian space to 300 K for 75 ps using conventional molecular dynamics; and (d) 2000 steps of conjugate-gradient Powell minimization. The weights for the NOE restraints were set to 150 kcal/mol for stages a−c, and 75 kcal/mol for stage d. The functional form of the NOE distance restraints was a flat-bottomed parabolic function with a soft asymptote. A sum averaging function was used for both NOE and H-bond restraints (*39*, *40*). The weights of the dihedral angle restraints were set to 100 kcal/mol for stage a, 200 kcal/mol for stages b and c, and 400 kcal/mol for the final stage d. The quartic van der Waals repulsion energies were calculated using a force constant of 4 kcal mol^{-1} $Å^{-4}$, with van der Waals radii set to 0.8 times the values used in the CHARMM empirical energy function (*41*).

Final structure computations involved the calculation of 40 structures, from which 25 low energy structures were selected that had no NOE violations >0.3 Å and no dihedral angle violations >5 Å. Structures were analyzed using Quanta (Molecular Simulations), MOLMOL (*42*), and Procheck_NMR (*43*).

RESULTS

Limited Proteolysis. The amino acid sequence of the VHv1.1 protein, including the six histidines from the baculovirus expression system, is shown in Figure 1a. The N-glycosylated full-length protein comprises 223 amino acids and forms a molecular weight complex of ∼30 kDa. In our initial studies on VHv1.1, our goal has been to determine the three-dimensional structure of smaller domains of the full-length protein. Inspection of the protein's amino acid sequence indicates two possible cysteine-rich regions (Figure 1, panels a and b), thus leading us to suspect that the protein may be comprised of at least two independently folded domains. However, identifying the precise ends and beginnings of protein domains by inspection of amino acid sequences remains at best an intuitive art. Thus, our plan has been to use the more empirical and experimental approach of limited proteolysis (*31*) to identify folded domains within VHv1.1, which could be amenable to NMR structural investigations. The limited proteolysis approach is based on subjecting a protein to partial degradation with protease enzymes and to identify protein segments that are protected from protease cleavage due in part to the inaccessibility of cleavage sites within the folded core of the protein (*31*). To simplify the limited proteolysis experiments, a mutant recombinant baculovirus (Asn101Gln−VHv1.1) was used to generate a nonglycosylated recombinant protein. SDS−PAGE results of full-length VHv1.1 digested with V8 enzyme are shown in Figure 2. In lane 1, the major band at ∼30 kDa represents full-length, uncleaved VHv1.1 protein, (Figure 2, label A). As proteolysis proceeds, two major bands appear, one slightly below the 11 kDa marker and the other around the ∼20 kDa mark (Figure 2, labels C and B, respectively). In the 40 min and 2 h limited proteolysis time points, full-length protein is still detectable on the gel (lanes 2 and 3 of Figure 2). Yet in the subsequent time points, it is clear that the full-length protein has been cleaved into two major products. N-terminal sequencing results demonstrated that the heavy protein fragment begins with the 6-residue stretch, Ser-Thr-Asp-Glu-Pro-Glu (1st stretch of underlined residues in Figure 1a). Its molecular mass was determined to be 15 000 Da by MALDI-TOF mass spectrometry (data not shown), compared to a predicted molecular weight of 15 034 Da. The second smaller fragment was identified to begin at the 10-residue stretch Val-Ser-Ser-Thr-Cys-Ile-Gly-His-Tyr-Gln (2nd stretch of underlined residues in Figure 1a), with a molecular mass of 7136 Da as determined by mass spectrometry (data not shown). Altogether these data established unambiguously that the full-length VHv1.1 protein is cleaved by V8 into two major fragments, an N-terminal ∼15 kDa domain spanning residues Ser 23 to Glu 158, and a ∼7 kDa C-terminal domain spanning residues Val 159 to Ala 217 + His tag. These data are summarized in Figure 1a. Comparable cleaving patterns were found with trypsin enzyme (data not shown).

NMR Spectroscopy. Sequence-specific proton resonance assignments for C-term VHv1.1 were made using standard procedures (*38*, *44*) from 2D-NMR spectra recorded at different temperatures and pH. A ^1H chemical shift table is available as Supporting Information. Two of the three disulfide bonds were identified by inspecting the αH to βH and βH to βH NOE connectivities between cysteine pairs involved in disulfide bond formation. Using this approach, Cys7 and Cys 22 were identified as forming one of the three disulfide bonds found in C-term VHv1.1. A second disulfide bond was linked to Cys 14 and Cys 39. This left the third disulfide bond to involve cysteine residues Cys 21 and Cys 47. Regular secondary structure elements in C-term VHv1.1 were identified from patterns of sequential and short-range NOE cross-peaks, three-bond $^3J_{HNH\alpha}$ coupling constants, and slowly exchanging amide protons. Such data are summarized in Figure 3. Four β-strands emerge from the patterns of short-range NOE [αN(*i*,*i*+1) and βN(*i*,*i*+1) especially], and are further supported by observation of strong $^3J_{HNH\alpha}$ coupling constants, and slow $^1H/^2H$ exchanging amide protons in these regions: β-strand 1 (Tyr 11−Cys 14) and β-strand 2 (Lys 24−Val 26) are short, whereas β-strand 3 (Val 33−Asp 40) stretches over eight residues and is the longest of the four strands; lastly, β-strand 4 stretches from Val 46 to Phe 50 (Figures 3 and 4).

Protein Structure Calculations. The full set of constraints obtained from the nuclear magnetic resonance (NMR) experiments is summarized in Table 1. These constraints were used as input data for simulated annealing in CNS (*39*, *40*) to calculate the three-dimensional structure of C-term VHv1.1 in solution. An ensemble of 40 structures was calculated. These structures displayed no distance restraint violations above 0.3 Å and no dihedral restraint violations above 5 Å. The structural statistics reported for the ensemble of the 25 best, low conformational energy, C-term VHv1.1.

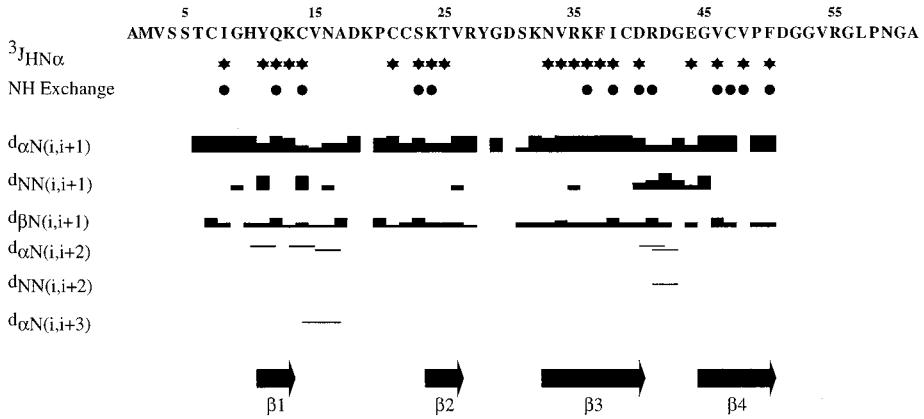

FIGURE 3: Summary of $^3J_{HNH\alpha}$, amide proton exchange, and patterns of sequential and short-range NOEs for C-term VHv1.1. The protein sequence is shown in the top line, using the one-letter abbreviation for the amino acids. Filled stars indicate three-bond *J*-coupling constants >8.5 Hz; filled circles denote residues for which the exchange time constant is >20 min. Sequential NOE cross-peaks are diagrammed in the rows $d_{NN}(i,i+1)$, $d_{\alpha N}(i,i+1)$, and $d_{\beta N}(i,i+1)$; the height of the box reflects the intensity of the NOE cross-peaks, classified as strong (tallest), medium, weak, and very weak (shortest). Short-range NOE cross-peaks are diagrammed in the rows $d_{NN}(i,i+2)$, $d_{\alpha N}(i,i+2)$, $d_{\alpha N}(i,i+3)$ as lines from the first residue involved in the cross-peak extending to the second residue. Elements of secondary structure are diagrammed in the last row, with arrows representing strands of β-sheet. Backbone regions between strands are referred to as loops I, II, and III.

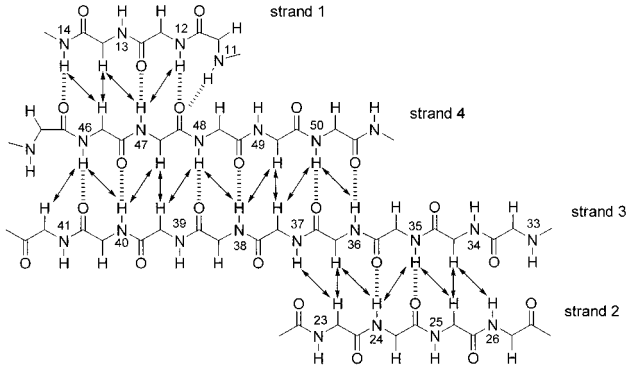

FIGURE 4: Schematic representation of the β-sheet structure of C-term VHv1.1. β-Strand interactions are depicted by hydrogen bonds (broken lines) and NOE connectivities (arrows) across strands. Amino acids are numbered at the NH position according to their location in the primary sequence.

structures indicate that the structures are well-defined and have good stereochemistry (Table 1). The root-mean-squared deviation from the mean (average rmsd) for the family of 25 structures is ~0.47 Å, with a corresponding pairwise root-mean-square atomic displacement of 0.67 ± 0.16 Å, for all β-strand backbone atoms. When the complete molecule is considered (residues 6−51) with the exception of the disordered N- and C-termini residues (residues 1−5, and residues 55−61), three flexible loop regions are included (Figures 3 and 6) and result in an overall pairwise rmsd of ~2.0 ± 0.4 Å for the family of 25 accepted structures (Table 1). The structures were also analyzed with the Procheck_nmr program (*43*) to examine the validity of the ϕ and φ dihedral angles for the polypeptide chain: 98% of the residues was found in the allowed regions of the Ramachandran plot.

C-Term VHv1.1 Tertiary Structure. The overall structure of the C-term VHv1.1 fragment is depicted in Figure 6, where the backbone heavy atoms of the best 14 structures have been superimposed over the whole molecule. The structure consists of a cysteine knot motif: the two [Cys 7(I)−Cys 22(IV)]and [Cys 14(II)−Cys 39(V)] disulfide bonds, connecting the polypeptide backbone, create a ring through which is threaded the third (Cys 21(III)-Cys 47(IV)) disulfide bond (Figure 5). The spacing between Cys I and Cys II consists of 6 residues and 16 between Cys (IV) and Cys (V),

Table 1: Structural Statistics for the Final Simulated Annealing Structures of C-Term VHv1.1

(A) experimental restraints	
total NOE restraints	280
intraresidue	67
sequential ($\|i-j\| = 1$)	112
short range ($1 < \|i-j\| < 4$)	12
long range ($\|i-j\| > 4$)	89
hydrogen bond restraints[a]	24
experimental dihedral angle ϕ restraints	20
restraint violations[b]	
distances	
number > 0.3 Å	0
dihedral	
number > 5°	0

(B) statistics for the calculated structures	{SA}[b]
deviation from idealized geometry	
bonds (Å)	0.0010 ± 0.0001
angles (deg)	0.30 ± 0.004
impropers (deg)	0.10 ± 0.01
final energies (kcal/mol)[c]	
distance restraints	1.40 ± 0.25
dihedral angles	0.01 ± 0.01
nonbonded (REPEL)	6.97 ± 0.86
root-mean-square deviations	
average atomic rmsd (Å)	
backbone core atoms[d]	
four-stranded β-sheet region	0.47
all heavy atoms from residues 6−51	1.36
pairwise atomic rmsd (Å)	
backbone core atoms	
four-stranded β-sheet region	0.67 ± 0.16
all heavy atoms from residues 6−51	1.95 ± 0.44

[a] Two distance restraints per hydrogen bond were used. [b] Values are for the ensemble 25 accepted structures, {SA}, obtained from simulated annealing in torsion angle and Cartesian space using CNS (*39*). [c] See text for details on structure calculations with CNS (*39*). [d] N, Cα, C, and O atoms were used for the superpositions of backbone coordinates. The four-stranded β-sheet core region includes residues: 23−25, 34−40, 46−50.

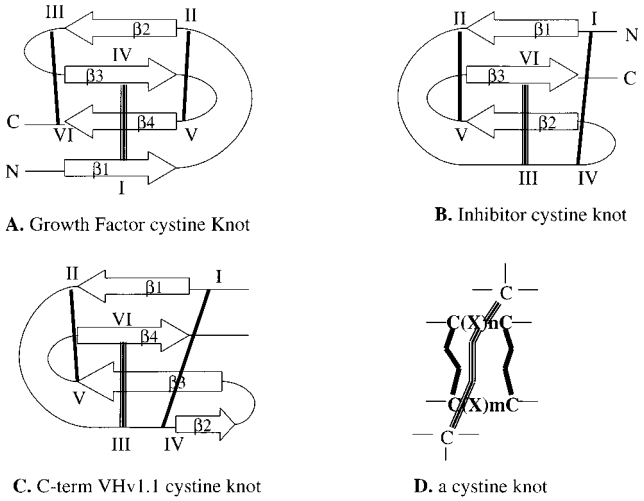

FIGURE 5: Topological arrangements of disulfide bonds in the cystine knot superfamilies. Comparison of the disulfide bond topology observed in the GFCK (panel A) and ICK (panel B) cystine knot superfamilies with the spatial arrangement of disulfide bonds identified in C-term VHv1.1 (panel C). The ring motif characteristic of cystine knots is shown in (panel D). The six cysteines forming the cystine knot disulfide bonds are referred to by roman numerals I through VI. The threaded disulfide bond in each motif is indicated by three vertical bars. β strands, labeled as β1 through β4, are not drawn to scale. In panel D, n represents the number of residue spacings between Cys (I) and Cys(II), and m is the number of residues between Cys (IV) and Cys (V). For the ICK superfamily, $n = 3-7$, $m = 1-4$ (46). In C-term VHv1.1, $n = 6$, $m = 16$. For the GFCK superfamily, the ring is delineated by disulfide bonds Cys(II)−Cys(V) and Cys(III)−Cys(VI). For the GFCK cystine knots, $n = 3-15$ and represents the number of residues between Cys(II) and Cys(III); m is equal to the number of residues between Cys(V) and Cys(VI) and is always equal to 1 (*47*).

forming a ring within the knot which is 22 residues long and involves all four β strands. This cystine knot stabilizes the unusual arrangement of the four β-strands forming the core of the structure. The four β-strands are arranged in an unusual configuration forming an antiparallel triple-stranded β-sheet plus a short antiparallel double-stranded β-sheet (Figures 4 and 8a). The C-terminal end of strand β3 hydrogen bonds with residues in the strands β1 and β4 to form an antiparallel triple-stranded β-sheet, while the N-terminal section of β3 forms a short antiparallel double-stranded β-sheet with β2. Residues involved in β-sheet formation are as follows: residue segment Val 34 to Lys 36 on strand β3 interacts with residues Lys 24 to Val 26 on strand β2 in the double-stranded β-sheet; residues Lys 36 to Asp 40 on strand β3 interact with residues Val 46 to Phe 50 on strand β4 in the triple-stranded β-sheet; finally, residues Tyr 11 to Lys 13 on strand β1 interact with residues Val 46 to Val 48 on strand β4 to complete the triple stranded β-sheet structures (Figures 4).

The pattern of hydrogen bonding in the β-sheets is regular except for Tyr 11 located at the edge of the β1 strand of the triple-stranded β-sheet. In the vicinity of Tyr 11, the amide protons of Tyr 11 and Gln 12 are hydrogen bonded to the same carbonyl moiety of residue 47 (Figure 4). The β4 strand and a large section of β3 exhibit a right-handed twist and cross over each other (Figures 6 and 8a). As mentioned, this β-sheet organization delineates three loops. The turn in loop I was identified as a type II′ turn involving residues Cys 14−Ala 17. Loop II, spanning residues Arg 27 to Asn 33, is the least defined part of the molecule and suggests increased flexibility in this region (Figures 3 and 6). Loop III spanning residues Arg 41 to Gly 45 exhibit overlapping turns that cannot be unambiguously characterized by the NMR data (Figure 3). The hydrophobic core of C-term VHv1.1 consists of side-chain interactions between Tyr 11 (strand 1), Phe 37 (strand 3) and Val 46 and Val 48 (strand 4) on one side of the molecule, referred to as the convex side (Figure 7, LHS), whereas the disulfide knot cystine residues, Phe 50 (strand 4) and Ile 38 are on the other side referred to as concave side (Figure 7, RHS). Most of the charged residues lie on the convex side of the molecule, with His 10, Lys 24, and Arg 35 forming one positively charged patch, and Lys 13 and Lys19 forming another. The lone negative charged patch on the convex side is formed by Asp 40, Asp 42, and Glu 44 (Figure 7).

DISCUSSION

The structure of C-termVHv1.1 presented herein is the first of its kind among the polydnaviral cysteine-rich protein family. The NMR structural data presented here clearly demonstrate the presence of a cystine knot structural motif. The patterns of disulfide bonding (Cys I−IV, Cys II−V, and Cys III−VI) identified in C-term VHv1.1 are identical to the ones found in all the cystine knot superfamilies (*45*−*49*). The cystine knot appears to impart a high degree of stability to the structural folds of proteins that contain such a motif. Currently, cystine knots are divided into three families: the growth factor cystine knots (GFCKs); the ion channels inhibitor cystine knots (ICKs); and the cyclic cystine knots (CCKs) (*46*, *47*). Comparing the 3D structure of C-term VHv1.1 to these three families, it is clear that C-term VHv1.1 represents a new and distinct cystine knot (Figure 8). In C-term VHv1.1, the four β-strands are integral components of the cystine knot, which consists of Cys I−IV and Cys II−V, i.e., Cys (*7*−*22*) and Cys (*14*−*39*), respectively, along with their connected polypeptide backbone chain, forming a large ring through which is threaded the third Cys III−VI disulfide bond, i.e., Cys (*21*−*47*). The four strands are arranged in an unusual configuration to form a triple-stranded β-sheet plus a double-stranded β-sheet, where strand β3 is involved in formation of both β-sheets (Figure 8a). By comparison, the growth factor cystine knots (GFCKs) include a group of diverse protein molecules that regulate cell growth, differentiation, and cell−cell communications (*48*, *49*). The 3D structures of NGF, TGF-β, PDGF, and glycoprotein hormones are prototypes of the GFCK protein fold (*48*). Like C-term VHv1.1, the GFCKs possess a cystine knot structure involving four β-strands. However, the structure of the cystine knot and the arrangement of the four strands are different. In GFCK cystine knot, the Cys II−V and Cys III−VI disulfide bonds along with their connected polypeptide chain form a ring through which Cys I−IV penetrates (*49*) (Figure 5). The four strands in GFCKs form a four-stranded antiparallel β-sheet, not a triple plus double β-sheet. Figure 8b illustrates these differences. In contrast, the inhibitory (ICK) and cyclic (CCK) cystine knots are very similar to each other topologically, except that the CCKs are cyclic in nature. These two cystine knot families include a group of small toxic and ion channels inhibitory polypeptides (*46*). The structure of the cystine knot in the ICKs and CCKs is

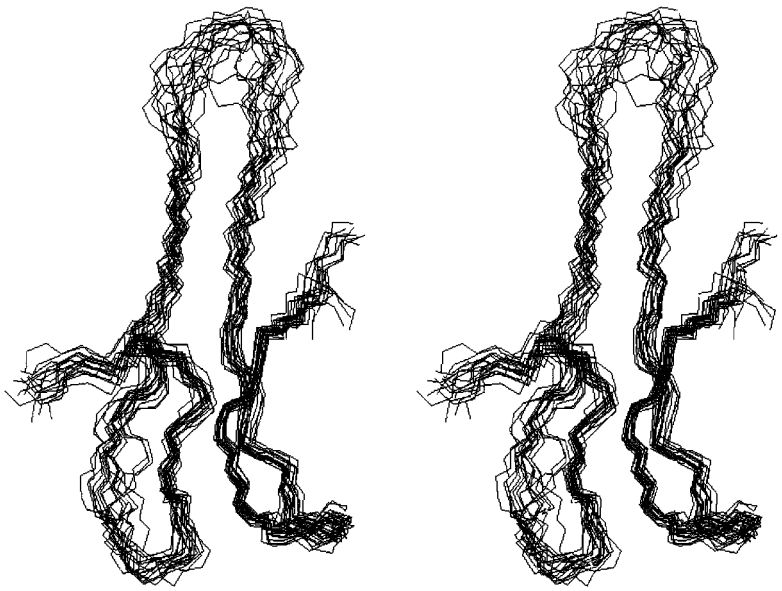

FIGURE 6: Schematic drawings of the structure of C-term VHv1.1. Stereoview of the backbone (C, Cα, N) atoms of 14 of the ensemble of 25 accepted NMR structures. The stereoview was generated with the program MOLMOL (42). Disulfide bonds have been omitted for clarity.

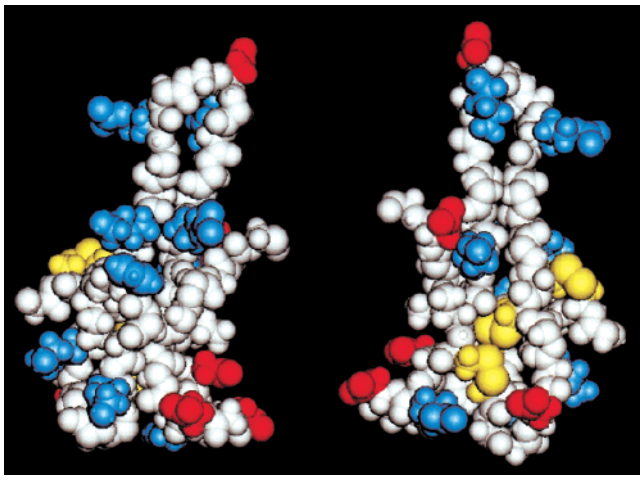

FIGURE 7: Solid-model representations of C-term VHv1.1 convex (LHS) and concave (RHS) surfaces. The blue and red side-chains represent positively and negatively charged residues, respectively. Cysteines are represented in yellow. On the convex surface, the three-disulfide bonds depicted in yellow are masked. The concave structure represents a 180° rotation about the vertical axis relative to the convex structure shown on the left and shows that the protein's three disulfide bonds are located predominantly on the concave side of the molecule.

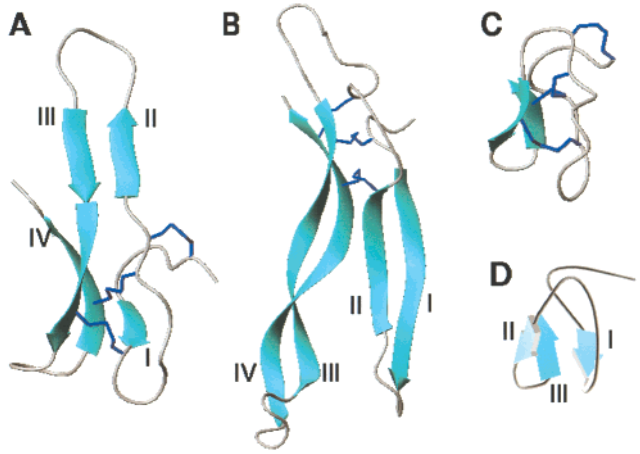

FIGURE 8: MOLMOL (*42*) representation of C-term VHv1.1 (panel A); the cystine knot structure of neurotrophin-3 (*49*), representing the structural motif of the GFCK cystine knot superfamily (panel B); and the ω-conotoxin MVIIA structure (*47*) representing the ICK cystine knots (panel C). Panel D is the Molscript (*50*) representation of panel C, and depicts the three-stranded β-sheet structure of ω-conotoxins (*47*). Figure 8 illustrates similarities and differences between C-term VHv1.1 and the 3D structures of representative members of two major cystine knot superfamilies. The β strands are labeled as roman numerals I, II, III, and IV from the amino terminal. The dark blue lines across the strands (seen in panels A−C) depict the proteins' disulfide bonds.

similar to that observed in C-term VHv1.1, i.e., the ring is formed by disulfide bonds Cys (I−IV) and Cys (II−V), with disulfide bond Cys (III, IV) threaded through the ring (*46*) (Figure 5). However, the number of β-strands involved in formation of the knot is different: in the ICK and CCK structures, three β-strands participate in knot formation (Figure 8d), whereas in C-term VHv1.1, four β strands are an integral part of the cystine knot motif (Figure 8a). All three strands involved in formation of the triple-stranded β-sheet in C-term VHv1.1 are longer than the three forming the triple-stranded β-sheet of ICKs and CCKs (Figure 8a). Though the sizes of the GFCK proteins are significantly larger than the ICK and CCK peptides, the ring size within the cystine knot motif is remarkably similar (*46*, *47*). For these three families, the ring size within the knot varies from 4 to 16 residues, not including cysteines. In the case of C-term VHv1.1, the ring is much longer and involves 22 noncysteine residues. Thus, the 3D structure of C-term VHv1.1 exhibits both similarities and differences when compared to the 3D structures of other cystine knot superfamilies. From this comparative analysis, we conclude that the 3D structure of C-term VHv1.1 represents a new and distinct example of a cystine knot. The protein scaffold consisting of four β-strands stabilized by three disulfide bonds appears robust enough to accommodate variations in the sequences of related *cys* family polydnaviral proteins.

Assessment of the Structural Information about C-Term VHv1.1 within the Context of the Genetic Data Available about Polydnaviral Cys-Motif Gene Sequences. In an effort to provide additional insights, we have compared structural regions of C-term VHv1.1 to nucleotide positions identified in genetic analyses to be under different selection pressures (*30*). Nucleotide positions that are highly conserved have been recognized as coding for invariant amino acids in analyses of CsIV cys-motif genes (*30*). In the case of C-term VHv1.1, these highly conserved nucleotide positions encode the six cysteines and two prolines. Recent frame-changing deletions and insertions were observed in nucleotide regions encoding amino acids located between the fourth and fifth cysteine residues and point to this gene region as possibly undergoing diversifying selection (*30*). However due to sequence gaps, this coding region cannot be aligned with other cys-motif genes, and nucleotide positions corresponding to amino acids in the loop II region of C-term VHv1.1 could not be analyzed (*30*). Interestingly, certain nucleotide positions found to be under neutral pressure correspond, in our structure, to amino acids that are hydrogen bonded, protected from $^1H/^2H$ exchange, and form the core region of C-term VHv1.1. These residues are Ile 8, Gln 12, Ile 38, Asp 40, Val 46, Val 48, and Phe 50 (Figures 3 and 4). In contrast other nucleotide positions, some corresponding to residues located within loops I and II, and the β2 strand of C-term VHv1.1 (Lys 24, Thr 25), correlate to gene regions that are under considerable diversifying pressure. The significance of diversifying positions in terms of structure and function of the cys-motif polydnaviral proteins is unclear at present. Diversifying pressures may have implications regarding modulation of these protein functions (*30*). However, information about the 3D structures of other members of the *cys* protein family are needed in order to fully comprehend the meaning, and the structural/functional implications, of the diversifying gene selection data.

We can conclude that the global fold of C-term VHv1.1 is governed by disulfide bonding patterns and β-strands. We would expect the 3D fold of other cys-motif containing polydnaviral proteins to be comparable to the 3D fold of C-term VHv1.1. However, subtle structural differences that may modulate variation in biological activity among the different proteins would also be anticipated. To provide further knowledge about this family of proteins, we have undertaken the structural determination of VHv1.1 N-terminal cys-motif and of the full-length VHv1.1. We hope that by solving the 3D structures of other polydnaviral cys-motifs, we will be able to provide further insights as to the relationship between structure and function governing CsIV *cys* proteins.

SUPPORTING INFORMATION AVAILABLE

A table of ^1H chemical shifts, as well as two figures representing the 2D COSY and NOESY spectra of the fingerprint region (NH to αH) of C-term VHv1.1. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Krell, P. J., Summers, M. D., and Vinson, S. B. (1982) *J. Virol. 43*, 859−870.
2. Brown, F. (1986) *Intervirology 25*, 141−143.
3. Stoltz, D. B., Krell, P. J., Summers, M. D., and Vinson, S. B. (1984) *Intervirology 21*, 1−4.
4. Stoltz, D. B., Beckage, N. E., Blissard, G. W., Fleming, J. G. W., Krell, P. J., Theilmann, D. A., Summers, M. D., and Webb, B. A. (1995) Polydnavirae. In *Virus taxonomy, Sixth Report of the International Committee on Taxonomy of Viruses* (Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A., and Summers, M. D., Eds.) pp 143−147, Springer-Verlag, Vienna, Austria.
5. Stoltz, D. B., and Vinson, S. B. (1979) Viruses and parasitism in insects. *Adv. Virus Res. 24*, 125−171.
6. Edson, K. M., Vinson, S. B., Stoltz, D. B., and Summers, M. D. (1981) *Science 211*, 582−583.
7. Strand, M. R., and Pech, L. L. (1995) *Annu. Rev. Entomol. 40*, 31−56.
8. Luckhart, S., and Webb, B. A. (1996) *Dev. Comput. Immunol. 20*, 1−21.
9. Webb, B. A., and Summers, M. D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4961−4965.
10. Webb, B. A., and Luckhart, S. L. (1996) *J. Insect Physiol. 42*, 33−41.
11. Blissard, G. W., Fleming, J. G. W., Vinson, S. B., and Summers, M. D. (1986) *J. Insect Phys. 32*, 351−359.
12. Summers, M. D., and Dib-Hajj, S. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 29−36.
13. Cui, L., and Webb, B. A. (1996) *J. Gen. Virol. 77*, 797−809.
14. Shelby K. S., and Webb, B. A. (1994) *J. Gen. Virol. 75*, 2285−2292.
15. Shelby, K. S., and Webb, B. A. (1997) *Insect Biochem. Mol. Biol. 27*, 263−270.
16. Webb, B. A., and Summers, M. D. (1992) *Experientia 48*, 1018−1022.
17. Li, X., and Webb, B. A. (1994) *J. Virol. 68*, 7482−7489.
18. Fleming, J. G. W., and Summers, M. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9770−9774.
19. Soldevila, A. I., Heuston, S., and Webb, B. A. (1997) *Insect Biochem. Mol. Biol. 27*, 201−211.
20. Blissard G. W., Vinson S. B., and Summers, M. D. (1986) *J. Virol. 57*, 318−327.
21. Blissard, G. W., Smith, O. P., and Summers, M. D. (1987) *Virology 160*, 120−134.
22. Blissard, G. W., Theilmann, D. A., and Summers, M. D. (1989) *Virology 169*, 78−79.
23. Dib-Hajj, S. D., Webb, B. A., and Summers, M. D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3765−3769.
24. 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., and Cruz, L. J. (1990) *Science 249*, 257−263.
25. Woodward, S. R., Cruz, L. J., Olivera, B. M., and Hillyard, D. R. (1990) *EMBO J. 9*, 1015−1020.
26. Cui, L., and Webb, B. A. (1997) *J. Virol. 71*, 8504−8513.
27. Cui, L., and Webb, B. A. (1997) *J. Gen. Virol. 78*, 1807−1817.
28. Cui, L., and Webb, B. A. (1997) *Arch. Insect Physiol. Biochem. 36*, 251−271.
29. Conticello, S., Gilad Y., Avidan, N., Ben-Asher, E., Levy, Z., and Fainzilber, M. (2001) *Mol. Biol. Evol. 18*, 120−131.
30. Dupas, S., Turnbull, M. W., and Webb, B. A. (2001) *Mol. Biol. Evol.* (submitted for publication).
31. Carey, J. (2000) *Methods Enzymol. 328*, 499−514.
32. Stewart, E. J., Aslund, F., and Beckwith, J. (1998) *EMBO J. 17*, 5543−5550.
33. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
34. Piotto M., Saudek V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
35. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.

36. Shaka, A. J., Lee, L. J., and Pines, A. (1988) *J. Magn. Res*. *77*, 274−293.
37. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun*. *117*, 479−485.
38. Redfield, C. (1993) Resonance assignment strategies for small proteins. In *NMR of macromolecules, a practical approach* (Roberts, G. C. K., Ed.) pp 71−99, IRL Press, Oxford, U.K.
39. Brunger, A. T., Adams, P. D., Clore, M. G., DeLano, W. T, Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S, Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
40. Stein, E. G., Rice, L. M., and Brunger, A. T. (1997) *J. Magn. Res*. *124*, 154−164.
41. Brooks, B. R., Bruccoleri, R. E., Olfason, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem*. *4*, 187−217.
42. Koradi R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
43. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
44. Wuthrich K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience Publication, New York.
45. Norton, R. S., and Pallaghy, P. K. (1998) *Toxicon 36* (11), 1573−1583.
46. Craik, D. J., Daly, N. L., and Waine C. (2001) *Toxicon 39*, 43−60.
47. Goldenberg, et al. (2001) *Protein Sci 10*, 538−550.
48. Sun, P. D., and Davis, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct*. *24*, 269−291.
49. Radziewski, et al. (1993) *Biochemistry 32*, 13350.
50. Kraulis (1991) *J. of App. Crystallogr*. *14*, 946−950.

BI011499S