Structural characterization of a polydnaviral protein involved in wasp parasitism
by Jerrod Grant Einerwold

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Biochemistry
Montana State University

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. The endoparasitic Campoletis sonorensis wasp carries polydnaviruses of the ichnovirus genera (CsIV). Survival and development of the parisitoid wasp egg is made possible in part by the host specific expression of a family of cys genes that encode for cysteine-rich proteins. Studying the structures and functions of this family of proteins will help in understanding how polydnaviruses alter insect host physiology. The VHv1.1 viral gene encodes for a 217 residue protein containing two cysteine-rich domains. A 65-residue C-terminal cysteine-rich domain (C-term VHvl.l) was identified experimentally by limited proteolysis of the VHvl.l gene product, and was subsequently cloned in a bacterial expression system for NMR studies. The C-term VHvl.l three-dimensional structure was determined in solution by two-dimensional 1H-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 root-mean-square deviation (rmsd) of 0.47 A from the mean structure for core backbone atoms. The protein core is made of a four 13-strand scaffold held together in a compact structure by three disulfide bonds, which form a cystine knot. The four (3-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 an indication that C-term VHvl.l represents a new and distinct cystine 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. The genetic data representing the evolutionary pressures on the cys gene family seems to correlate with the C-term VHvl.l protein structure data. This allows for the extrapolation of possible structural similarities and differences among other CsIV cys-motif proteins.

STRUCTURAL CHARACTERIZATION OF A POLYDNAVIRAL PROTEIN INVOLVED IN WASP PARASITISM

by

Jerrod Grant Einerwold

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

October 2001

APPROVAL

of a thesis submitted by

Jerrod Grant Einerwold

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies,.

Dr. Valérie Copié Valérie Copié' 10/11/01
(signature) Date

Approved for the Department of Chemistry and Biochemistry

Dr. Paul Grieco Paul Grieco 10/11/01
(signature) Date

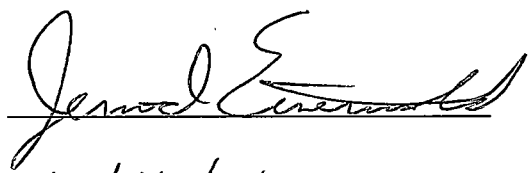
Approved for the College of Graduate Studies

Dr. Bruce Mcleod Bruce K. McLeod 10-11-01
(signature) Date

SATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University , I agree that the Library shall make it available to borrowes under rules of the Library.

Copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Signature 

Date 10/11/01

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

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. The endoparasitic *Campoletis sonorensis* wasp carries polydnaviruses of the ichnovirus genera (CsIV). Survival and development of the parisitoid wasp egg is made possible in part by the host specific expression of a family of *cys* genes that encode for cysteine-rich proteins. Studying the structures and functions of this family of proteins will help in understanding how polydnaviruses alter insect host physiology. The VHv1.1 viral gene encodes for a 217 residue protein containing two cysteine-rich domains. A 65-residue C-terminal cysteine-rich domain (C-term VHv1.1) was identified experimentally by limited proteolysis of the VHv1.1 gene product, and was subsequently cloned in a bacterial expression system for NMR studies. The C-term VHv1.1 three-dimensional 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 root-mean-square deviation (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 an indication that C-term VHv1.1 represents a new and distinct cystine 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. The genetic data representing the evolutionary pressures on the *cys* gene family seems to correlate with the C-term VHv1.1 protein structure data. This allows for the extrapolation of possible structural similarities and differences among other CsIV cys-motif proteins.

CHAPTER 1

INTRODUCTION

Polydnaviruses have developed a symbiotic relationship with endoparasitic Hymenoptera and are essential for the successful parasitization of Lepidopteran insects following wasp oviposition (1-5). Endoparasitic wasps must alter the immune and developmental responses of its larval Lepidopteran hosts to produce an environment that allows parasite survival and supports its development. Successful parasitization is accompanied by suppression of the multicellular immune response (encapsulation by hemocytes) and developmental alterations of the host (6-8). Polydnaviruses, wasp venoms, and ovarian proteins are injected into the host during oviposition and are responsible for the physiological effects of parasitization on the host.(9-10). A life cycle of an endoparasitic wasp is illustrated in Figure 1 (taken from 11).

Polydnaviruses are unique in terms of their polydisperse (segmented) DNA genomes, their association with Hymenoptera, and their distinctive life cycle (11). The segmented viral genome is dispersed and integrated throughout the wasp genome (proviral DNA). Although proviral DNA is present in all tissues of both male and female wasp, replication of the virus occurs only in the calyx cells of the female oviduct (11). Polydnavirus replication begins with the preexisting proviral DNA segments being excised from their integrated positions within the wasp genome (11). Virus replication is first detected during the late pupal stage. Virion particles accumulate in the oviduct to such a high density that they are detectable by a blueish tint (11). The dense population of

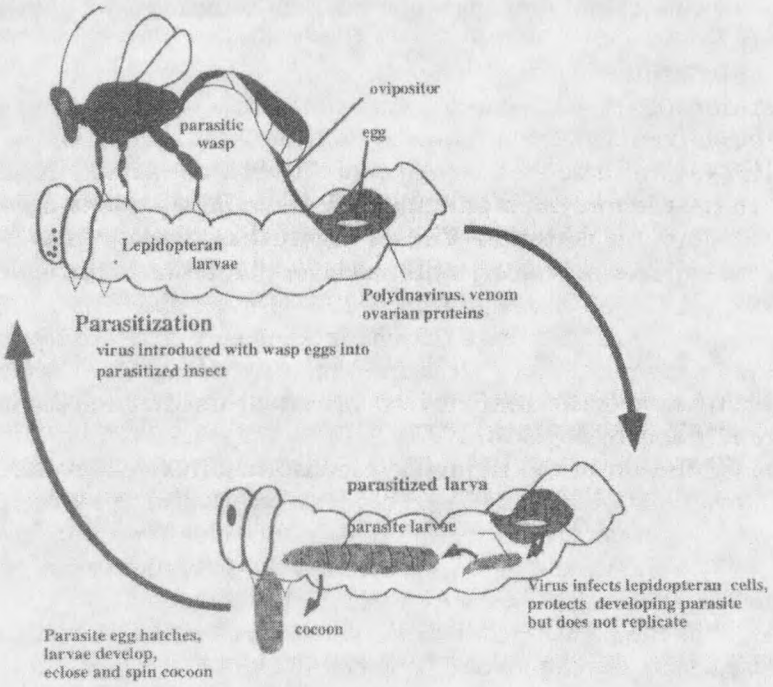


FIGURE 1. The life cycle of an endoparasitic Hymenoptera. Female wasps oviposit into host insects and at the same time inject polydnaviruses, venoms, and ovarian proteins that allow survival and development of the egg. The host dies shortly after or during parasite emergence.

virion particles, along with an ovarian protein solution, make up the calyx fluid which surrounds the egg in the oviduct. During oviposition, the egg, calyx fluid, and venom are introduced into the body of the host insect (11). This results in the introduction of two forms of viral DNA into the parasitized insect, one form being the virions that will infect Lepidopteran cells and express a subset of host specific genes, and the second form being the proviral DNA present within the chromosomes of the wasp egg (11). Polydnavirus life cycles are not typical in that virion infection takes place in the Lepidopteran host and the processes of replication along with transmission of viral progeny take place within the Hymenopteran wasp. As a result the polydnavirus has two hosts with which it distributes

its statges of infection, replication, and transmission of viral progeny (11). The life cycle of a polydnavirus in relation to the life cycle of an endoparasitic wasp is shown in Figure 2 (taken from 11).

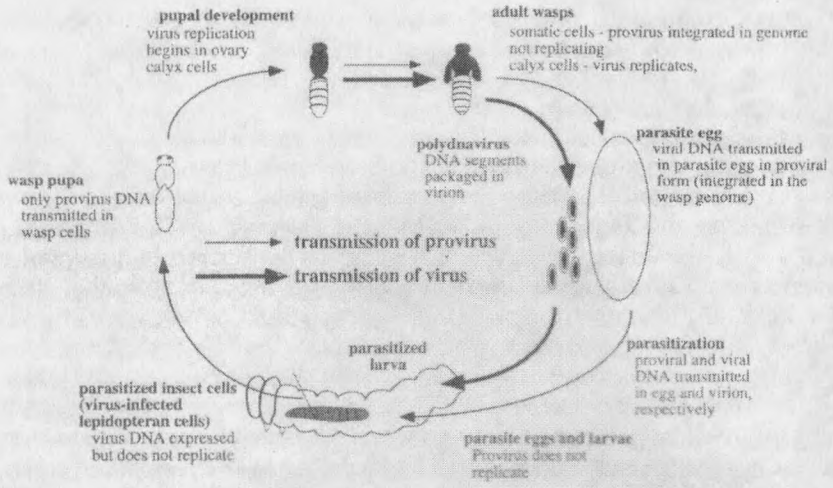


FIGURE 2. The life cycle of a polydnavirus in relation to that of an endoparasitic wasp. Viral DNA is transmitted to the host as proviral DNA in wasp chromosomes (thin arrows) and as circular DNA molecules within virions (thick arrows). Proviral DNA is the viral progeny and makes a complete cycle, whereas virions do not replicate or continue through the life cycle once in the lepidopteran host.

Independent evolutionary lineages gave rise to two genera of polydnaviruses: the bracoviruses which are associated with braconid wasps and the ichnoviruses which are associated with ichneumonid wasps (11). Since polydnaviral progeny are transmitted as integrated proviral DNA in the wasp genome, every wasp species carries a genetically isolated virus that represents a unique viral species (11).

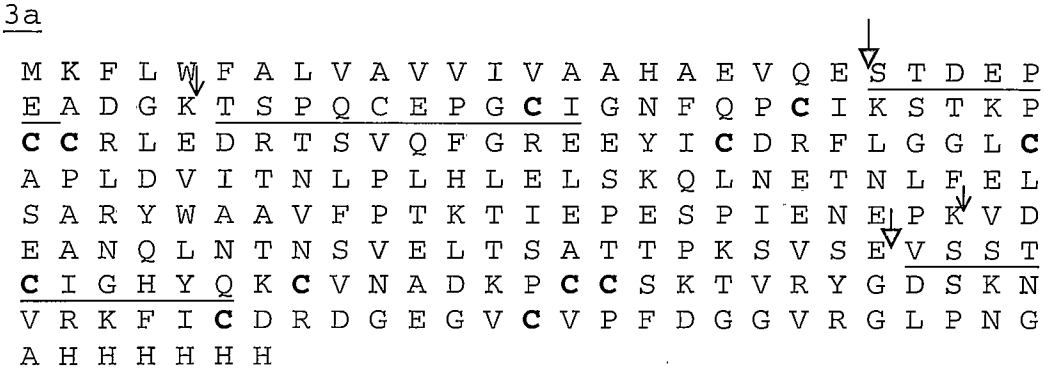
The endoparasitic *Campoletis sonorensis* wasp carries polydnaviruses of the ichnovirus genera. Among the estimated ten thousand ichnovirus species, the *Campoletis sonorensis* ichnovirus (CsIV) is the most studied (11). Injection of purified CsIV into

Lepidopteran host larvae mimics natural parasitization by suppressing host cellular immune responses and altering its growth (12).

The CsIV genome consists of approximately twenty-eight DNA segments that range in size from 5 to 21 kbp and are assigned alphabetical letters in an increasing order of size (11, 13). Two major gene families, the *rep* and *cys* gene families, have been identified in the CsIV genome (14). *Rep* genes are characterized by hybridization to a conserved 540 base-pair element, but only the BHv0.9 *rep* gene has been sequenced (14). 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 (15, 16-19). Sequence analysis reveals that these genes share a common gene structure, including conserved introns interrupting the coding sequences of cysteine-rich ("cys-motif") domains, and code for two distinct protein subfamilies, labeled W- and V- (19). Within the CsIV cysteine-rich gene family, introns are more conserved than flanking coding exons, and the protein primary sequences are characterized by six invariable cysteines flanking highly variable amino acid residues (Figure 3) (19). These characteristics are reminiscent of the primary structure arrangement of conotoxins, where conserved cysteine residues are separated by highly variable intercysteine residues (20, 21).

The CsIV cys-motif proteins have a putative ~16-amino acid signal peptide, allowing for protein secretion. The six-cysteine residues are arranged in a (C ...C ... CC ... C ... C) pattern characteristic of small peptide neurotoxins of carnivorous snails (ω-conotoxins) and scorpions, which are members of the ion-channels inhibitory toxin

3a



3b

CsIV cys-motifs

```
                I       II      III IV                      V         VI
(VHv1.1n)  CIGNFQPCIKSTKPCCRLEDRTSVQFGREEYICDRFLGGLCAPLD
(VHv1.1c)  CIGHYQKCVNADKPCCSKTVRYGDSKNVRKFICDRDGEGVCVPFD
(VHv1.4n)  CIGNYQPCIESTKPCCRLEDRTSVQFGRKEYICDRFFGGLCAPLD
(VHv1.4c)  CIPNYELCVNSKRPCCWENKLFAGSSKPRNFVCGLHGRSYCSPFD
(WHv1.0)   CMANWDYCLGFGKPCCDQ...........HSICFKFGEGICTPLH
(WHv1.6)   CIDNWKYCRGINKPCCGQQLMEDGTLGPKHFVCFELGQGICTPLS
```

ω-conotoxins

```
(GVIA)     CKSPGSSCSPTSYNCCRS.............CNPYTKR.C
(MVIIA)    CKGKGAKCSRLMYDCCTGS............CRSGK...C
```

FIGURE 3. 3a) Amino acid sequence of the full-length VHv1.1 polydnaviral protein. Thick arrows indicate cleavage sites where endoproteinase Glu-C (V8) enzyme cleave the protein into two major fragments. Thin arrows indicate sites were trypsin cleaved the protein. Underlined residues indicate stretches of amino acids identified, by N-terminal sequencing data, as the N-termini of the three fragments produced by V8 and trypsin limited proteolysis. The N-terminal sequencing data were corroborated with molecular weight determination of the fragments by MALDI-TOF mass spectrometry. 3b) Amino acid sequence alignment of cys-motifs identified in VHv1.1, VHv1.4, WHv1.0, and WHv1.6 CsIV *cys* genes. 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 superfamily. The sequence alignment exemplifies the analogous spacing of conserved cysteines among the CsIV cys-motif proteins and the smaller ω-conotoxin polypeptides. Roman numerals are used to indicate individual cysteine residues involved in disulfide bond formation.

cystine knot family (22). However, compared with the ω-conotoxins, the polydnaviral proteins are larger and do not contain a propeptide region. 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 (23, 24). In contrast, the V-subfamily possesses proteins with two cys-motifs, consisting of ~ 41 amino acids each (15, 19). For both the W- and V- protein subfamilies, the cysteine-rich motifs consist of hypervariable intercysteine amino acids and invariant cysteine residues (Figure 3). 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 (15, 19, 25). 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 (26). 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 (27). In this study twelve cysteine-motif coding regions from two species of wasps, one being *Campoletis sonorensis*, were analyzed. 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 (Figure 4) (27). 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 (27). In contrast, nucleotide positions under

diversifying selection pressure could correspond to amino acid residues that are important for allowing divergence in function between the CsIV cys-motif protein family members (27).

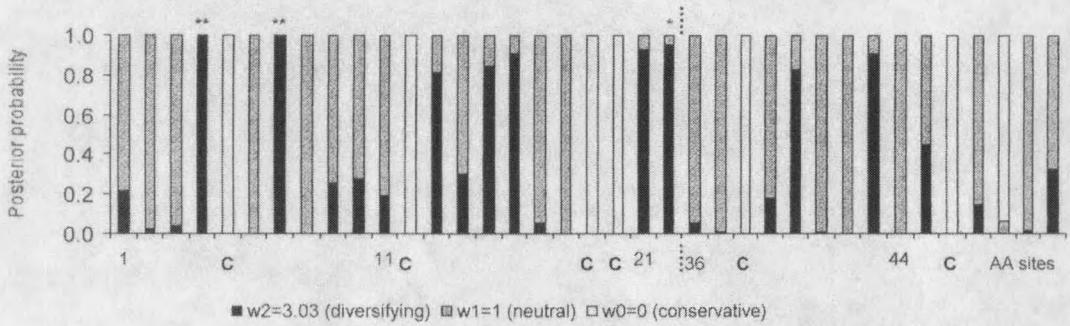


FIGURE 4. Taken from (27). Amino acid positions within the CsIV cysteine motif and their correlation to a posterior probability of their respective codon belonging to the conserved, neutral, or diversifying class of codon sites. The dashed line corresponds to the amino acid positions 23 to 35, a region partially or totally deleted in some motifs.

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. It has been shown that injection of a recombinant baculovirus, encoding for wild type VHv1.1, into Lepidopteran hosts reduces encapsulation of wasp eggs injected 24 hours post infection (28). Egg encapsulation is reduced by 50% in the presence of VHv1.1 compared to that of eggs injected into uninfected larvae (28). It is thought that VHv1.1 protein is taken up by hemocytes and disrupts the cytoskeloton making the cells nonfunctional, thereby inhibiting encapsulation of the egg (11). Recently, the cys-motif protein encoded by VHv1.4 has been implicated in the selective inhibition of protein synthesis in parasitized larvae (Webb et al., submitted) and there is also evidence

that the VHv1.1 gene product may have a similar activity (Kim and Webb, unpublished data).

In an attempt to provide molecular insights into polydnaviral cys-motif protein function, my research has involved determining the 3D structure determination of the C-terminal cysteine-rich domain found in the CsIV VHv1.1 gene product. Using limited proteolysis (29), two independently folded domains were identified in the protein. The domain's 3D structure was solved by two-dimensional ^{1}H-NMR spectroscopy. Similarities and differences between C-term VHv.1.1's 3D fold and the structures of known cystine knot families are described in this thesis. The discussion closes by presenting evidence supporting a possible relationship between the structural features observed in C-term VHv1.1 and the genetic evidence linking codon positions within the CsIV cys-motifs to regions under neutral or diversifying selective pressure.

CHAPTER 2

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 (30). This construct was identical to previously reported VHv1.1 recombinant baculovirus (30) 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 hours post-infection and centrifuged. In preparation for nickel affinity protein purification, the supernatant (containing secreted VHv1.1 protein) was concentrated to 10mL and dialyzed against 1X-binding buffer (20 mM Tris-HCl pH 7.9, 500 mM NaCL, 5 mM imidazole, 4L). The sample was applied to a 1 mL NTA-Ni column (Qiagen™), and washed with 1X-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. Fractions containing product were combined (two 1mL fractions) and dialyzed against phosphate buffered saline (PBS: 25 mM phosphate pH 7.0, 150 mM NaCl, 4L).

Limited proteolysis and protein fragment identification.

In preparation for limited proteolysis experiments, 25 μg of sequencing grade Endoproteinase Glu-C (V8) (Sigma Co.) were reconstituted in 20 μL of water, aliquoted into 5 μL samples, frozen in liquid nitrogen, and stored at -80°C. A 3mg/mL stock of trypsin was made with 1mM HCL, aliquoted into 20 μL samples, frozen in liquid nitrogen, and stored at -80°C. Limited proteolysis was carried out on ice, as described in (29). For the range-finding stage of the experiment, six 25 μl aliquots of full-length VHv1.1 protein were prepared. An aliquot of the Endoproteinase Glu-C (V8) frozen stock was thawed on ice, and diluted in a series of ten-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-minute incubation period, reactions were stopped by adding 30 μL of SDS-reducing sample buffer and boiling for 5 min. The range-finding experiment with trypsin was carried out exactly as the V8 range finding experiment above. The samples were analyzed on SDS-PAGE, using a 15% reducing gel. Results from the range-finding experiments established that the 100-fold enzyme dilution for both V8 and trypsin was optimal for the time-course experiments (29). In this second stage, an equal volume of a 100x dilution of V8 enzyme was mixed with VHv1.1 protein stock, and 20 μL aliquots were collected at 5 min, 40 min, 2 hrs, 5 hrs, 10 hrs, 21 hrs, 30 hrs, and 48 hrs. A trypsin time-course experiment was carried out exactly as the V8 time-course experiment described above. Protein fragments were separated by HPLC reverse-phase chromatography and analyzed by SDS-PAGE, MALDI-TOF mass spectrometry, and N-terminal sequencing.

The exact amino acid sequences of the fragments were identified using the 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 subcloned into a prokaryotic expression system.

Vector construction for bacterial expression of the C-terminal motif.

A VHv1.1 cDNA construct, pVX900 clone (19, 30), was used as template in standard PCR reactions with cloned Pfu polymerase (Stratagene®). Nco1 and Xho1 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 Xho1 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 and sequence. The pET-32b(+) vector allowed for expression of an N-terminal His-Tag® expressed C-terminal to an N-terminal fusion protein, thioredoxin, to C-term VHv1.1 (31).

Bacterial C-terminal VHv1.1 protein expression,

purification, and NMR sample preparation.

One-liter cell cultures were initiated by inoculating 500 mL(X2) 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 hours, then moved to a 16-17°C growth chamber. The cells were harvested 20 hours post IPTG induction, and the pellets resuspended in 40 mL of 1X-binding buffer. C-term VHv1.1 was isolated and purified using nickel affinity and gel filtration chromatography. The fusion partner + His-Tag® were 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 ~ 1mM protein solution, in 25 mM sodium phosphate pH5.7, 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.

With the assistance of Dr. Mahesh Jaseja, an NMR postdoctoral fellow in the laboratory, ^1H NMR spectra were recorded on a Bruker DRX 600 spectrometer at Montana State University. In order to assign overlapping resonances in 2D ^1H NMR spectra , experiments were recorded at 315 °K and 310 °K for a ~ 1mM sample solution

at pH 5.7, and at a temperature of 305 ^{o}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 (32). 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 (33). 2D homonuclear NOESY spectra (34) were recorded with a mixing time of 200 ms. TOCSY spectra (35) 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 to 512 t_1 increments, 2048 data acquisition points, and a relaxation recovery delay of 1.5 sec, with spectral widths of 6613.7 Hz. For DQF-COSY (36) spectra, 1024 t_1 increments were acquired. Spectra were recorded with 96 scans per t_1 increment for NOESY, 32 scans per t_1 increment for DQF-COSY, and 16 scans per 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 degrees, were applied in both dimensions prior to Fourier transformation. The three-bond $^{3}J_{HNH\alpha}$ 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 305K, 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 in (37).

Structure calculations.

Distance restraints for structure calculations were based on 280 NOE crosspeaks assigned in 2D ^{1}H-NOESY spectra recorded at 315° K, with a 200 msec 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). 24 additional distance restraints were included for 12 hydrogen bonds identified through both characteristic β-sheet NOE patterns and ^{1}H / ^{2}H 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 (38, 39). The molecular dynamic scheme used in CNS consisted of the following stages: a) Heating in torsion angle space at 50,000 °K for 45 psec with the energy constant for the van der Waals parameters scaled by 0.1; b) Cooling in torsion angle space to 2,000 °K for 300 psec with ramping of the van der Waals parameters to full scale; c) Cooling in Cartesian space to 300 °K for 75 psec using conventional molecular dynamics; 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 for the NOE distance restraints was a flat-bottomed parabolic function with soft asymptote. A sum averaging function was used for both NOE

and H-bond restraints (38, 39). The weights of the dihedral angle restraints were set to 100 kcal/mol for stage a, 200 kcal/mol for stage b-c, and 400 kcal/mol for 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 (40).

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 (41), and Procheck_NMR (42).

CHAPTER 3

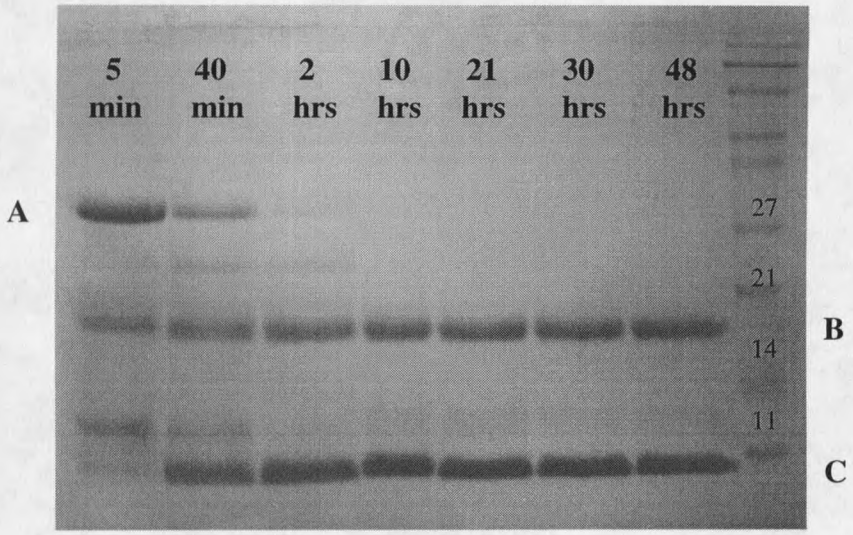
RESULTS

Limited Proteolysis.

The amino acid sequence of the VHv1.1 protein, including the six histidines from the baculovirus expression system, is shown in Figure 3a. 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 3a and 3b), 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 (29) 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 by 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 (29). In order 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 5a. In lane 1, the major band at ≈ 30 kDa represents full-length, uncleaved VHv1.1 protein, (Figure 5a, 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 5a, labels C and B, respectively). In the 40 min and 2 hours limited proteolysis time points, full-length protein is still detectable on the gel (lanes 2 and 3 of Figure 5a). 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, (1^{st} stretch of underlined residues in Figure 3a). Its molecular mass was determined to be 15,000 daltons by MALDI-TOF mass spectrometry (Figure 6a), compared to a predicted molecular weight of 15,034 daltons. The second smaller fragment, was identified to begin at the 10-residue stretch Val-Ser-Ser-Thr-Cys-Ile-Gly-His-Tyr-Gln (3^{rd} stretch of underlined residues in Figure 3a), with a molecular mass of 7,136 daltons as determined by mass spectrometry (Figure 6b). 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 3a. Unambiguous data was gathered for only one of the trypsin fragments. It was identified to begin with the 10-residue stretch Thr-Ser-Pro-Gln-Cys-Glu-Pro-Gly-Cys-Ile (2^{nd} stretch of underlined residues in Figure 3a) and mass spectrometry data suggests its C-terminal residue is Lys 133 (data not shown).

5a



5b

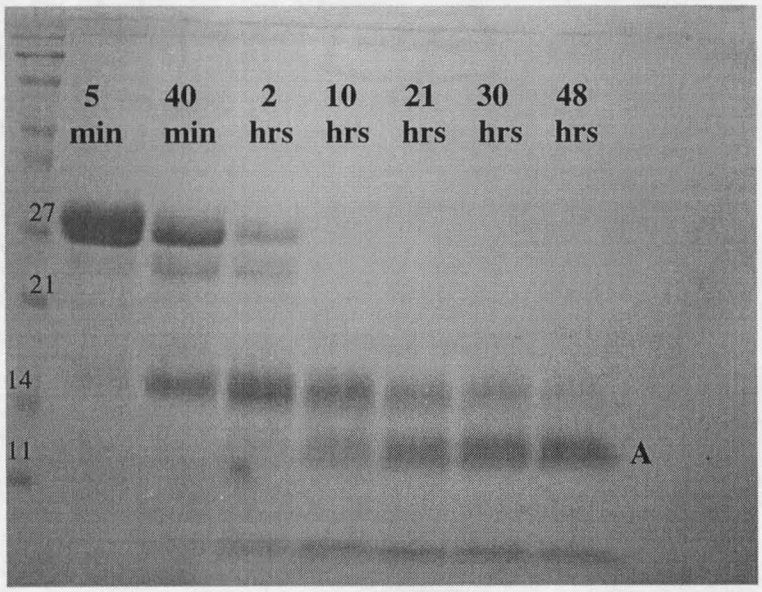
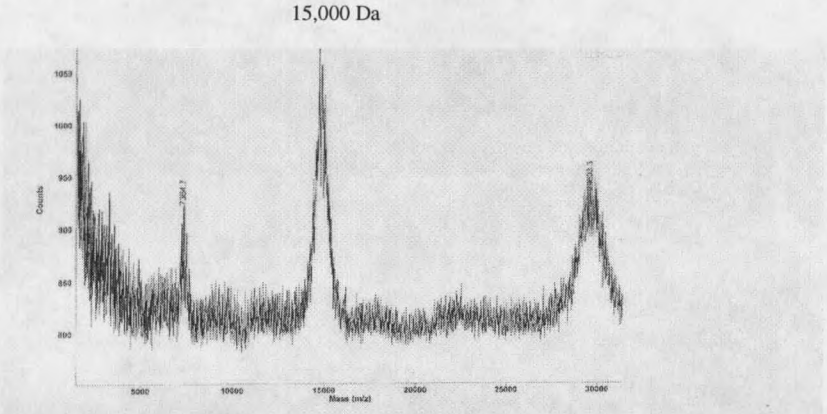


FIGURE 5. 5a) SDS-PAGE gel electrophoresis results of limited proteolysis experiments on full-length VHv1.1 using endoproteinase Glu-C (V8) protease. The A band corresponds to full length, uncleaved protein. The B and C bands correspond to two resulting protein fragments, which appear shortly after protease cleavage is initiated, and persist for 48 hours following the initiation of limited protease digestion. 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 (Figure 3a). 5b) SDS-PAGE gel electrophoresis results of limited proteolysis experiments on full-length VHv1.1 using trypsin. The A band was identified as the N-terminal cys-motif of VHv1.1 (Figure 3a).

6a



6b

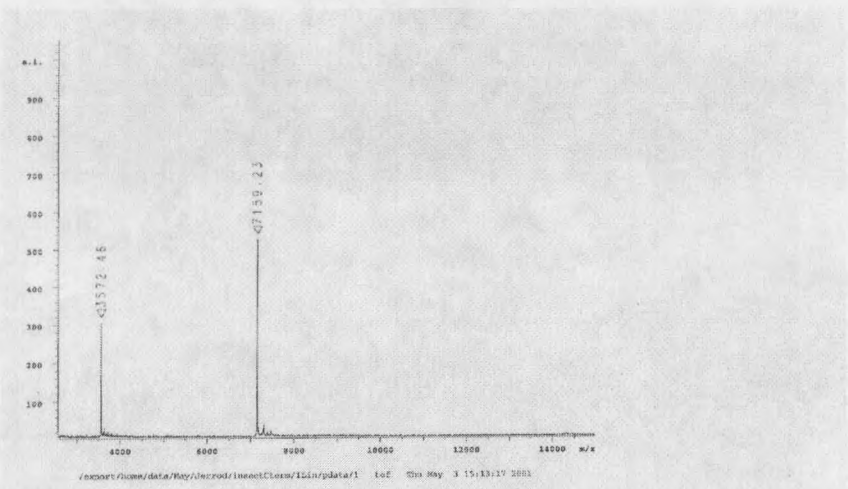


FIGURE 6. 6a) MALDI TOF mass spectrum collected on large fragment produced by V8 digestion and RP-HPLC purification. 6b) MALDI TOF mass spectrum collected on small fragment produced by V8 digestion and nickel affinity column purification, 23 mass units should be subtracted from the experimental mass to correct for sodium ion.

NMR spectroscopy.

Sequence specific proton resonance assignments for C-term VHv1.1 were made using standard procedures (37, 43) from 2D-NMR spectra recorded at different temperatures and pH.

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 crosspeaks, three-bond $^{3}J_{HNH\alpha}$ coupling constants, and slowly exchanging amide protons. Such data are summarized in Figure 7. 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 $^{3}J_{HNH\alpha}$ coupling constants, and slow $^{1}H/^{2}H$ exchanging amide protons in these regions: β-strand 1 (Tyr11-Cys13) and β-strand 2 (Lys 24-Val 26) are short, whereas β-strand 3 (Val 33-Asp 40) stretches over 8 residues, and is the longest of the four strands; lastly, β-strand 4 stretches from Val 46 to Phe 50 (Figures 7 and 8).

Protein structure calculations.

The full set of constraints obtained from the nuclear magnetic resonance (NMR) experiments are summarized in Table 1. These constraints were used as input data for

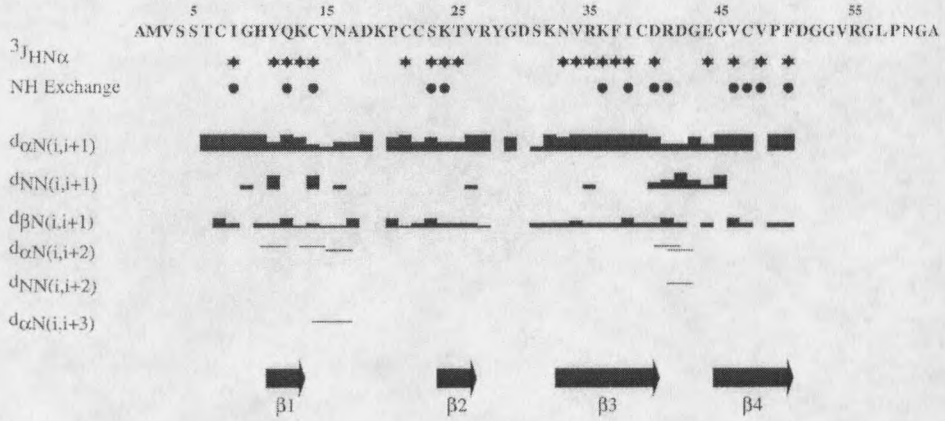


FIGURE 7. Summary of $^3J_{HN\alpha}$, amide proton exchange, and patterns of sequential and short range NOEs for C-term VHv1.1. The protein sequence is shown in the top line, using the one-letter abbreviation for the amino acids. Filled stars indicate three-bond J-coupling constants > 8.5 Hz; filled circles denote residues for which the exchange time constant is > 20 minutes. Sequential NOE crosspeaks 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 crosspeaks, classified as strong (tallest), medium, weak, and very weak (shortest). Short range NOE crosspeaks 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 crosspeak 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.

simulated annealing protocols in CNS (38, 39) 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 5Å. The structural statistics reported for the ensemble of the 25 best, low conformational energy, C-term VHv1.1. 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

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 constraints[a]	24
Experimental dihedral angle ϕ constrain	20
Restraint violations[b]	
Distances > 0.3 Å	0
Dihedral > 5 Å	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 Å
All heavy atoms (residues 6 to 51)	1.36 Å
Pairwise atomic rmsd Å	
Backbone core atoms[d]	
Four-stranded β-sheet region	0.67 ± 0.16 Å
All heavy atoms (residues 6 to 51)	1.95 ± 0.44 Å

[a]Two distance restraints per hydrogen bond were used
[b]Values are for the ensemble of 25 accepted structures, {SA}, obtained from simulated annealing in torsion angle and cartesian space using CNS (34)
[c]See materials and methods for details on structure calculations with CNS (34)
[d]N, C_α, 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.

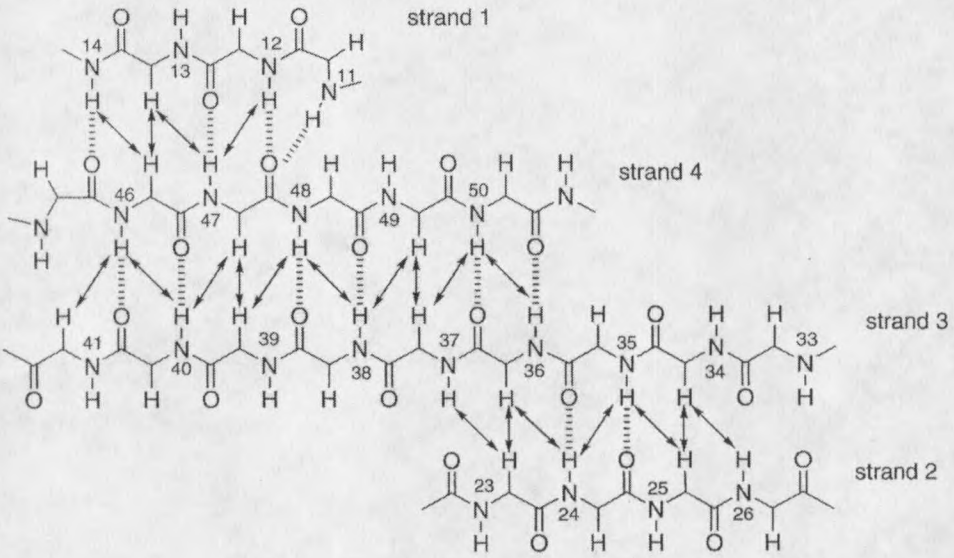


FIGURE 8. Schematic representation of the β-sheet structure of C-term VHv1.1. β-strand interactions are depicted by hydrogen bonds (broken lines) and NOE connectivities (arrows) across strands. Amino acids are numbered at the NH position according to their location in the primary sequence.

β-strand backbone atoms. When the complete molecule is considered with the exception of the disordered N- and C-termini residues (residues 1-5, and residues 52-61), three flexible loop regions are included (Figures 9 and 11), and result in an overall pairwise rmsd of ~ 2.0 ± 0.40 Å for the family of 25 accepted structures (Table 1). The structures were also analyzed with the Procheck_nmr program (42) to examine the validity of the ϕ and φ dihedral angles for the polypeptide chain: 98 % of the residues were 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 9, 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 (Cys7(I)-Cys22(IV)) and (Cys14(II)-Cys39(V)) disulfide bonds, connecting the polypeptide backbone, create a ring through which is threaded the third (Cys21(III)-Cys47(IV)) disulfide bond (Figure 10). The spacing between Cys I and Cys II consists of 6 residues, and 16 between Cys (IV) and Cys (V), 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 8 and 11a). 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 Val34 to Lys36 on strand β3 interacts with residues Lys24 to Val26 on strand β2 in the double-stranded β-sheet; residues Lys36 to Asp40 on strand β3 interact with residues Val46 to Phe50 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 8).

The pattern of hydrogen bonding in the β-sheets is regular except for a Tyr 11

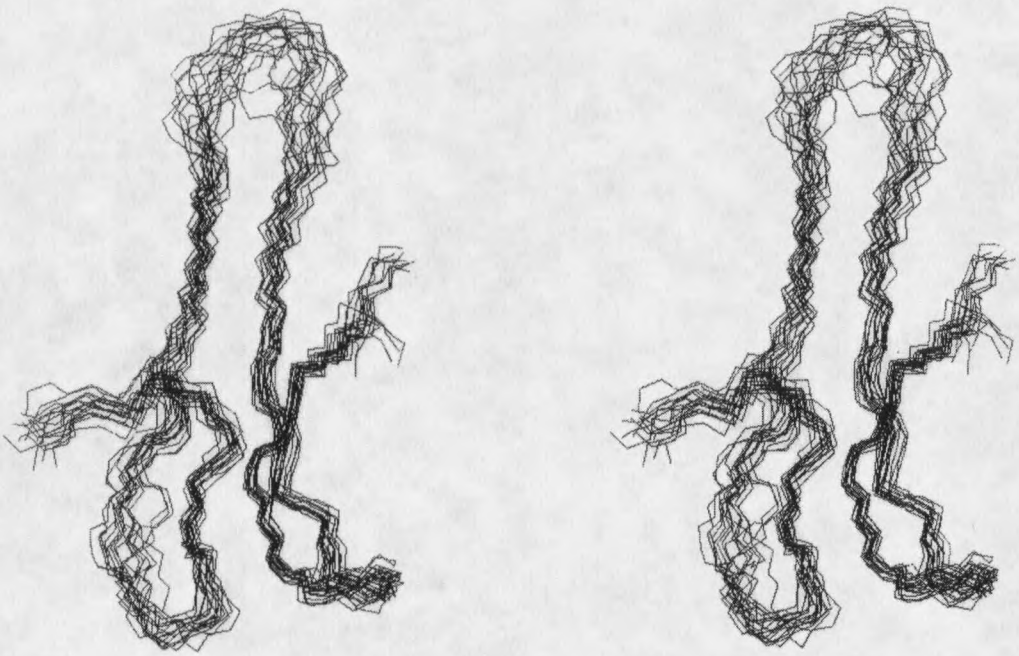
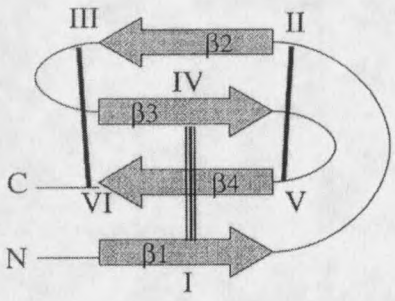
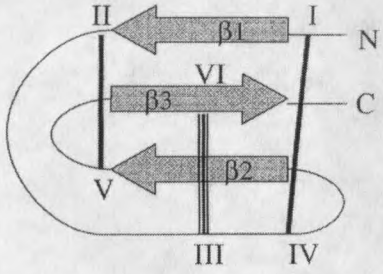


FIGURE 9. Schematic drawings of the structure of C-term VHv1.1. Stereoview of the backbone (C, Cα, N) atoms of 14 of the ensemble of 25 accepted NMR structures. The stereoview was generated with the program MOLMOL (41). Disulfide bonds have been omitted for clarity.

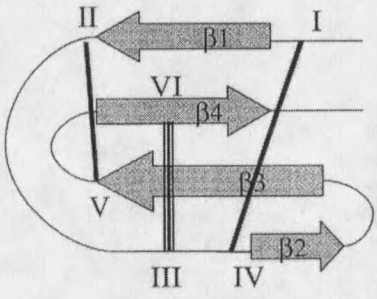
located at the edge of the β1 strand of the triple-stranded β-sheet. In the vicinity of Tyr 11, the amide protons of Tyr 11 and Gln 12 are hydrogen bonded to the same carbonyl moiety of residue 47 (Figure 8). The β4 strand and a large section of β3 exhibit a right-handed twist and cross over each other (Figures 9, 11a). 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 7, 9). Loop III spanning residues Arg 41 to Gly 45 exhibits overlapping turns that cannot be unambiguously characterized by the NMR data (Figure 7). The



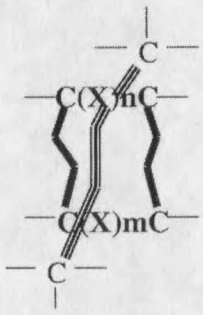
A. Growth Factor cystine Knot



B. Inhibitor cystine knot



C. C-term VHv1.1 cystine knot



D. a cystine knot

FIGURE 10. Topological arrangements of disulfide bonds in the cystine knot superfamilies. Comparison of the disulfide bond topology observed in the GFCK (10A) and ICK (10B) cystine knot superfamilies with the spatial arrangement of disulfide bonds identified in C-term VHv1.1 (10C). The ring motif characteristic of cystine knots is shown in (10D). 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 Figure (10D), 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 (43). 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 (44).

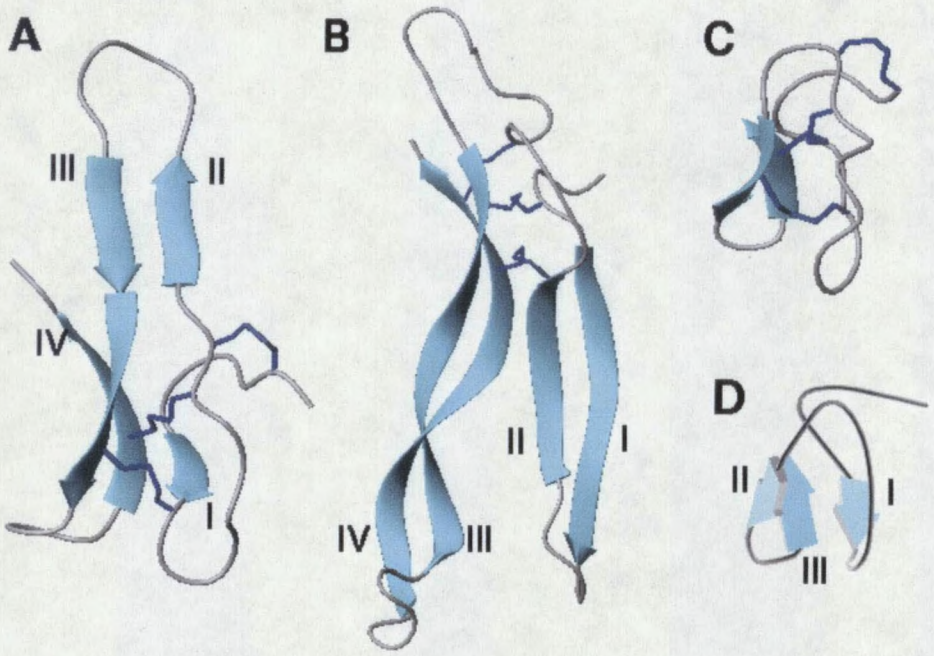


FIGURE 11. MOLMOL (41) representation of C-term VHv1.1 (11a); the cystine knot structure of neurotrophin-3 (48), representing the structural motif of the GFCK cystine knot superfamily (11b); and the ω-conotoxin MVIIA structure (46) representing the ICK cystine knots (11c). Molscript (49) representation of the ω-conotoxin MVIIA structure (11d), depicting the three-stranded β-sheet structure of ω-conotoxins (46). The figure 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 I, II, III, and IV from the amino terminal. Dark blue lines (11a-11c) represent disulfide bonds.

hydrophobic core of C-term VHv1.1 consists of side-chain interactions between Tyr 11 (strand 1), Phe 37 (strand 3), Val 46 and Val 48 (strand 4) on one side of the molecule, referred to as the convex side (Figure 12, 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 12, 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 12).

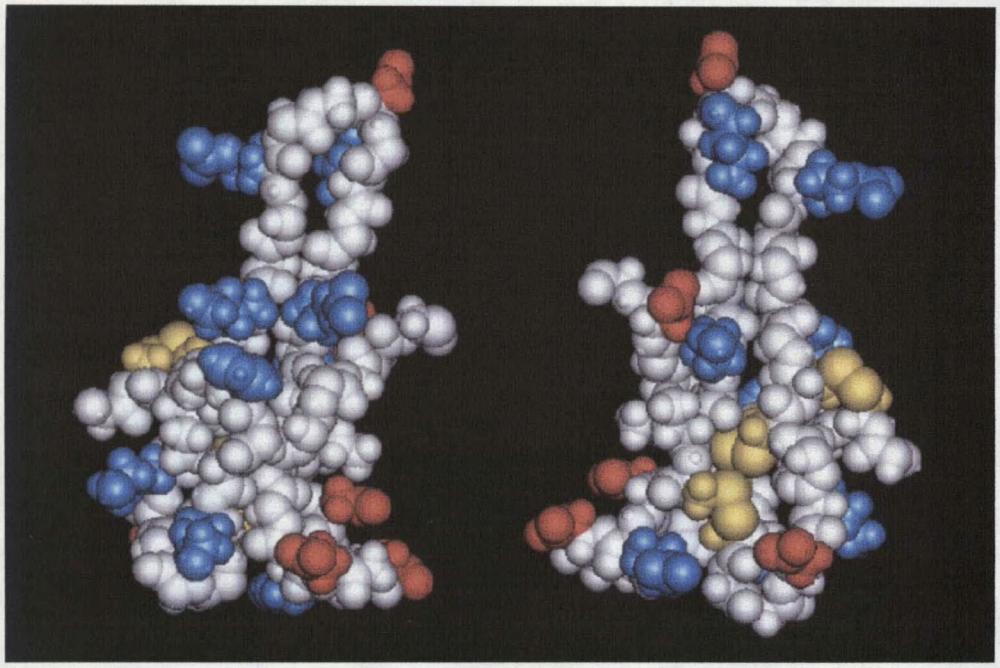


FIGURE 12. Solid-model representations of C-term VHv1.1 convex (LHS) and concave (RHS) surfaces. The blue and red sidechains represent positively and negatively charged residues, respectively. Cysteines are represented in yellow. On the convex surface, the three-disulfide bonds depicted in yellow are 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.

CHAPTER 4

DISCUSSION

Limited proteolysis experimentally identified the N- and C-terminal cys-motif domains of VHv1.1. Full-length recombinant VHv1.1 protein was subjected to partial protease degradation by V8 and trypsin. There are 26 potential V8- and 19 potential trypsin- cleavage sites uniformly dispersed within the amino acid sequence of full-length VHv1.1, and many of these sites are located within regions suspected to be part of VHv1.1's disulfide bonded core domains (Figure 3). What is striking about the limited proteolysis results is that neither V8 nor trypsin produced VHv1.1 fragments from cleavage sites located within the cysteine-rich motifs (Figures 3a). With such an approach, we were able to empirically identify the exact polypeptide sequence corresponding to C-term VHv1.1. The limited proteolysis data were then used to design a prokaryotic expression vector that allowed for production of larger amounts of C-term VHv1.1, an absolute prerequisite for preparation of NMR samples needed to solve the 3D structure of C-term VHv1.1 in solution.

The structure of C-termVHv1.1 presented herein is the first of its kind among the CsIV 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, & Cys III-VI, Figure 3b) identified in C-term VHv1.1 is

identical to the ones found in all the cystine knot superfamilies of its kind (figures 3b and 11) (44-48). 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 channel inhibitor cystine knots (ICKs); and the cyclic cystine knots (CCKs) (45-46). 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 11). 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 11a). By comparison, the growth factor cystine knots (GFCKs) include a group of diverse protein molecules that regulate cell growth, differentiation, and cell-cell communications (47, 48). The 3D structures of NGF, TGF-β, PDGF and glycoprotein hormones are prototypes of the GFCK protein fold (47). 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 (48) (Figure 10). The four strands in GFCKs form a four-stranded antiparallel β-sheet, not a triple plus double β-sheet. Figure 11a,b 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 channel inhibitory polypeptides (45). The structure of the cystine knot in the ICKs and CCKs is 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 (45) (Figure 10). 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 11c), whereas in C-term VHv1.1, four β strands are an integral part of the cystine knot motif (Figure 11a). 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 11a). 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 (45, 46). 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 non-cysteine 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, it can be concluded 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* gene sequences.

By analogy with the ω-conotoxins, the proteins belonging to the CsIV *cys* gene family most likely have similar cysteine motif structures yet function through differing molecular mechanisms. Zeroing in on the functional regions of these types of families could be done by comparing the three dimensional structures within a family and attributing differences to possible functional regions. This would require having the three dimensional structures and knowing the function of all of the family members. This is not the case for the *cys* family of proteins. Many years of exciting research will be done before this project can offer up structures and functions for all of the *cys* gene products.

What can be done at this stage is to correlate regions of the C-term structure to the cysteine motif coding regions, of the *cys* genes, that are highly conserved, under neutral pressure, or under diversifying pressure. Positions within the cysteine motif gene sequences that are conserved or under neutral pressure may be regions were nature does not want to commit to a great deal of experimenting because these positions give rise to protein interactions that are important in stabilizing the core structure. Positions that are under diversifying pressure may be regions were nature is doing a great deal of experimenting in an effort to adapt the functional mechanisms of this family of proteins to the evolving host defenses. Positions that are highly conserved are easily recognized in the amino acid sequences, these positions include the cysteine residues and proline 49 in the C-term motif (Figures 4 and 3b). Positions that are under neutral and diversifying pressures can not be recognized by the inspection of the amino acid sequences of the

cyseine motifs. They are found by the analysis of the genes using techniques described in (27). Recent frame changing deletions and insertions observed between the fourth and fifth cysteine residues points to this region as possibly undergoing diversifying selection, but did not allow for alignment with all of the genes and therefore positions correlating to T25 through K36 (loop II) in C-term VHv1.1 were not included in the analysis.

Residues exhibiting slowly exchanging amide protons were identified in the C-term motif via deuterium/proton experiments (Figure 7). These amide protons are thought to be under slow exchange because they are participating in stable and robust H-bonds that are crucial to the core stability of the C-term motif. Nine out of the thirteen residues that were found to have slowly exchanging amide protons, correlated with positions in the CsIV *cys* genes that are under neutral pressure (I8, Q12, I38, D40, V46, V48, F50, Figures 4 and 7) or positions that were highly conserved (C14, C47, Figures 4 and 7). The other four residues in the C-term motif that were found to be participating in stable H-bonds were either not included in the genetic analysis (K36, Figures 4 and 7) or found to correlate to positions within the CsIV *cys* genes that were under diversifying pressure (S23, K24, and R41, Figures 4 and 7). Interestingly, the H-bonding partners for all of the residues mentioned above shared analogous codon classifications, i.e. residues that correlated to neutral or conserved codon positions are H-bonded to residues which correlate to codon positions that are under neutral pressure; and residues that correlate to codon positions that are under diversifying pressure are H-bonded to residues that also correlate to codon positions under diversifying pressure (Figures 4, 7, and 8).

Secondary structure elements of the C-term motif can also be correlated to the

genetic analysis. β1, 1/2β3 (only ½ appeared in analysis due to deletion and insertion property of this region), and β4 strands were found to be in regions of the *cys* genes under neutral pressure. The β2 strand (S23, K24, T25) correlates to a region in the *cys* genes that is under considerable diversifying pressure. Finally, the V15 through D18 segment, a loop region of the C-term motif, correlates to an area of the *cys* genes under diversifying pressure (Figures 4 and 7). Figure 13 brings together the genetic analysis and the C-term structure by color-coding the Molscript (49) representation of C-term VHv1.1 according to the evolutionary selection pressure found in a particular region.

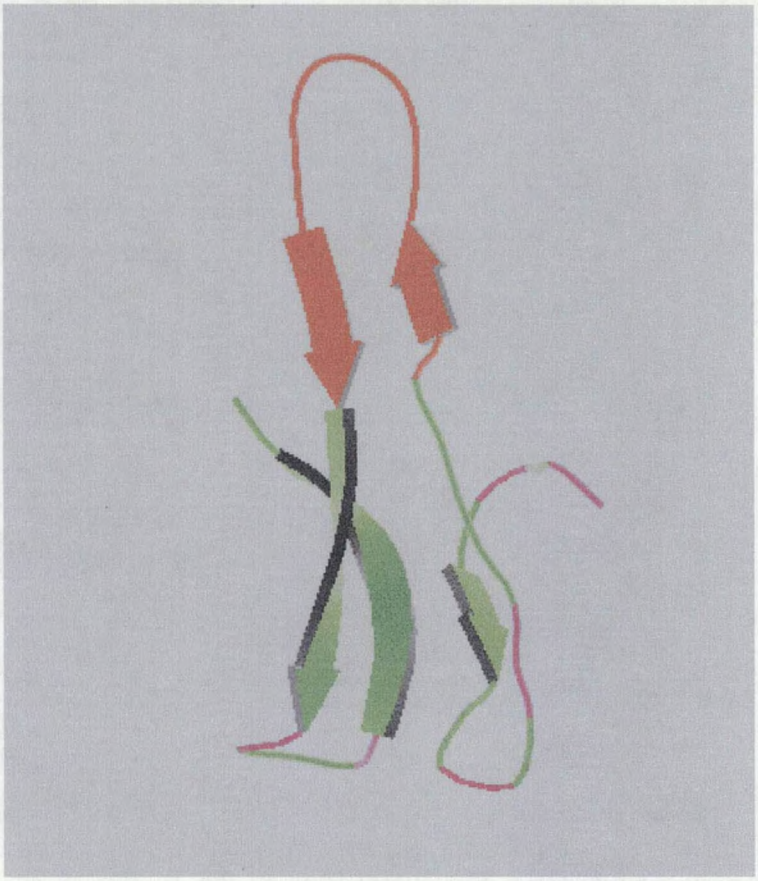


FIGURE 13. Molscript (49) representation of C-term VHv1.1 color-coded according to selection pressure found in particular regions. Green-neutral, pink-diversifying, red-not included in genetic analysis but thought to be under diversifying pressure.

This information may lead to the finding that the codon positions found to be under neutral pressure and correlating to H-bonded residues in the C-term motif will give rise to residues protected against exchange in all of the CsIV cysteine motifs and may be just as important for the stability of the core as the disulfide bonds. In addition, the codon positions found to be under diversifying pressure and correlating to H-bonded residues in the C-term motif may give rise to residues that are not always protected against exchange in all of the CsIV cysteine motifs thus, perhaps not as important for core stability in other motifs as compared to the C-term motif. It is tempting to speculate that the lengths of the β2 and β3 strands will differ among the CsIV cys motifs more so than the other secondary structure elements when known structures are compared. Due to the limited amount of structural data on CsIV cys-motifs, 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 (27). However, information about the 3D structures of other members of the cys-motif protein family are needed in order to fully comprehend the meaning, and the structural/functional implications, of the diversifying gene selection data.

In conclusion, the global fold of C-term VHv1.1 is governed by disulfide bonding patterns and β-strands. It may be expected that 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.

LITERATURE CITED.

1. Krell, P.J., Summers, M.D., and Vinson, S.B. (1982) *J. Virology* 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, pp. 143-147. In F.A. Murphy, C.M. Fauquet, D.H.L. Bishop, S.A. Ghabrial, A.W. Jarvis, G.P. Martelli, M.A. Mayo, and M.D. Summers (ed), *Virus taxonomy, Sixth Report of the International Committee on Taxonomy of Viruses*, 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. Comp. Immunol.* 20, 1-21.

9. Webb, B.A., and Summers, M.D. (1990) *Proc. Natl. Acad. Sci. USA* 87, 4961-4965.

10. Webb, B.A., and Luckhart, S.L., (1996) *J. Insect Physiol.* 42, 33-41.

11. Webb, B.A. (1998) Polydnavirus biology, genome structure, and evolution. In Miller, L.K., and Ball, A.L. (ed) *The Insect Virus,* Plenum Publishing, New York.

12. Soldevila, A.I., Heuston, S., and Webb, B.A. (1997) *Insect Biochem. Mol. Biol.* 27, 201-211.

13. Fleming, J.G.W. (1992) Annu. Rev. Entomology 37, 401-425.

14. Theilmann, D.A., and Summers, M.D. (1988) *Virology* 167, 329-341.

15. Cui, L., and Webb, B.A. (1996) *J. Gen. Virology* 77, 797-809.

16. Blissard G.W., Vinson S.B., and Summers, M.D. (1986) *J. Virology* 57, 318-327.

17. Blissard, G.W., Smith, O.P., and Summers, M.D. (1987) *Virology* 160, 120-134.

18. Blissard, G.W., Theilmann, D.A., and Summers, M.D. (1989) *Virology* 169, 78-79.

19. Dib-Hajj, S.D., Webb, B.A., and Summers, M.D. (1993) *Proc. Natl. Acad. Sci. USA* 90, 3765-3769.

20. 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.

21. Woodward, S.R., Cruz, L.J., Olivera, B.M., and Hillyard, D.R. (1990) *EMBO J.* 9, 1015-1020.

22. Olivera, B.M., Miljanich, G.P., Ramachandran, J., and Adams, M.E. (1994) *annu. Rev. Biochem.* 63, 823-867.

23. Cui, L., and Webb, B.A. (1997) *J. Virology* 71, 8504 –8513.

24. Cui, L., and Webb, B.A. (1997) *J. Gen. Virology* 78, 1807 –1817.

25. Cui, L. and Webb, B.A. (1997) *Arch. Insect Physiol. Biochem.* 36, 251-271.

26. Conticello, S., Gilad Y., Avidan, N., Ben-Asher, E., Levy, Z., and Fainzilber, M. (2001) *Mol. Biol. Evol.* 18, 120-131.

27. Dupas, S., Turnbull, M.W., and Webb, B.A. (2001) *Molecular Biology and Evolution*, submitted.

28. Li, X., and Webb, B.A. (1994) *J. Virology* 68, 7482-7489.

29. Carey, J. (2000) *Methods in Enzymology* 328, 499-514.

30. Summers, M.D., and Dib-Hajj, S.D. (1995) *Proc. Natl. Acad. Sci. USA* 92, 29-36.

31. Stewart, E.J., Aslund, F., and Beckwith, J. (1998) *Embo J.* 17, 5543-5550.

32. States, D.J., Haberkorn, R.A., and Ruben, D. J. (1982) *J. Mag. Res.* 48, 286-292.

33. Piotto M., Saudek V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661-665.

34. Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. (1979) *J. Chem. Phys.* 71, 4546-53.

35. Shaka, A.J., Lee, L.J., and Pines, A. (1988) *J. Mag. Res.* 77, 274-293.

36. Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wuthrich, K. (1983) *Biochem. Biophys. Res. Comm.* 117, 479-485.

37. Redfield, C. (1993) Resonance assignment strategies for small proteins, pp 71-99. In Roberts, G.C.K. (ed) *NMR of macromolecules, a practical approach*, IRL Press, Oxford, U.K.

38. 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 Cryst.* D54, 905-921.

39. Stein, E.G., Rice, L.M., and Brunger, A.T. (1997) *J. Mag. Res.* 124, 154-164.

40. Brooks, B.R., Bruccoleri, R.E., Olfason, B.D., States, D.J., swaminathan, S., and Karplus, M. (1983) *J. Comp. Chem.* 4, 187-217.

41. Koradi R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graphics* 14, 51-55

42. Laskowski, R.A., Rullmann, J.A.C, MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996) J. Biomol. NMR 8, 477-486.

43. Wuthrich K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience Publication, New York, N.Y

44. Norton, R.S., and Pallaghy, P.K. (1998) *Toxicon* 36(11), 1573-1583.

45. Craik, D.J., Daly, N.L., and Waine C. (2001) *Toxicon* 39, 43-60.

46. Goldenberg et al. (2001) *Protein Sci* 10, 538-550.

47. Sun, P.D., and Davis, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269-291.

48. Radziewski et al. (1993) *Biochemistry* 32, 13350.

49. Kraulis, (1991) *J. of Applied Crystallography* 14, 946-950.