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

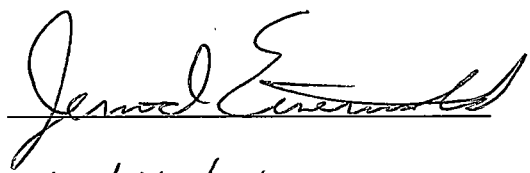
Approved for the College of Graduate Studies

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

1. INTRODUCTION.....1

2. MATERIALS AND METHODS.....9

Production of VHv1.1 protein in High Five™ insect cells.....9
Limited proteolysis and protein fragment identification.....10
Vector construction for bacterial expression of the C-terminal motif.....11
Bacterial C-terminal VHv1.1 protein expression,purification, and NMR sample preparation.....12
NMR Spectroscopy.....12
Structure calculations.....14

3. RESULTS.....16

Limited Proteolysis.....16
NMR spectroscopy.....20
Protein structure calculations.....20
C-term VHv1.1 tertiary structure.....24

4. DISCUSSION.....29

Assessment of the structural information about C-term VHv1.1 within the context of the genetic data available about polydnaviral *cys* gene sequences.....32

LITERATURE CITED.....36

LIST OF TABLES

Table Page

1. Structural statistics for the final simulated annealing structures of C-term VHv1.1.....22

LIST OF FIGURES

Figure Page

1. The life cycle of an endoparasitic Hymenoptera....2
2. The life cycle of a polydnavirus in relation to that of an endoparasitic wasp....3
3. Amino acid sequence of the full-length VHv1.1 polydnaviral protein....5
4. 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....7
5. SDS-PAGE gel electrophoresis results of limited proteolysis experiments on full-length VHv1.1 using endoproteinase Glu-C (V8) protease....18
6. MALDI TOF mass spectrum collected on large fragment produced by V8 digestion and RP-HPLC purification....19
7. Summary of $^{3}J_{HNH\alpha}$, amide proton exchange, and patterns of sequential and short range NOEs for C-term VHv1.1....21
8. Schematic representation of the β-sheet structure of C-term VHv1.1....23
9. Schematic drawings of the structure of C-term VHv1.1....25
10. Topological arrangements of disulfide bonds in the cystine knot superfamilies....26
11. MOLMOL (41) representation of C-term VHv1.1 (10a); the cystine knot structure of neurotrophin-3 (48), representing the structural motif of the GFCK cystine knot superfamily (10b); and the ω-conotoxin MVIIA structure (46) representing the ICK cystine knots (10c)....27
12. Solid-model representations of C-term VHv1.1 convex (LHS) and concave (RHS) surfaces....28
13. Molscript (49) representation of C-term VHv1.1 color-coded according to selection pressure found in particular regions....34

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

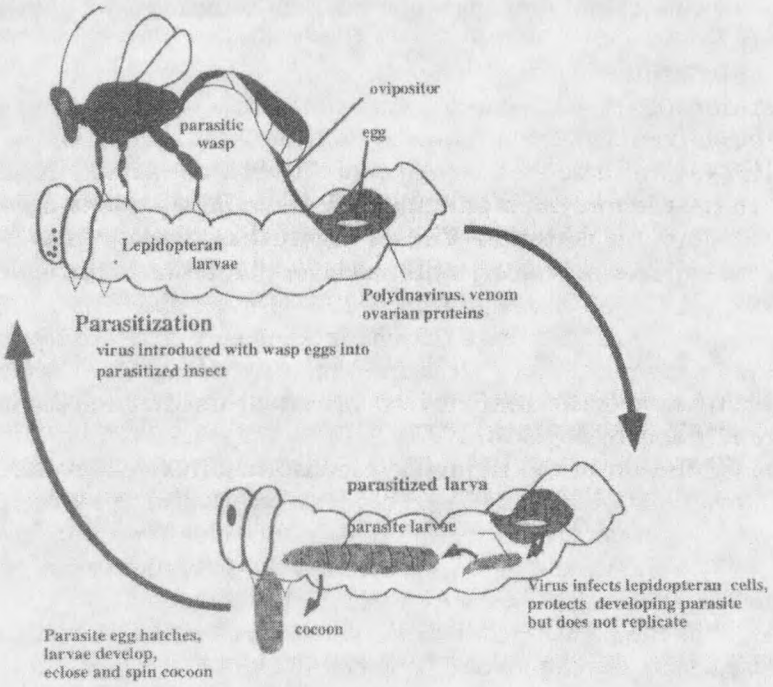


FIGURE 1. The life cycle of an endoparasitic Hymenoptera. Female wasps oviposit into host insects and at the same time inject polydnaviruses, venoms, and ovarian proteins that allow survival and development of the egg. The host dies shortly after or during parasite emergence.

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).

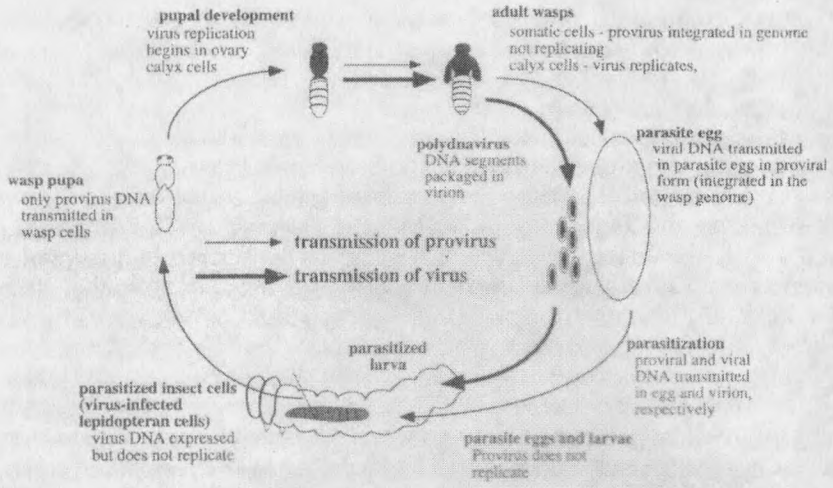


FIGURE 2. The life cycle of a polydnavirus in relation to that of an endoparasitic wasp. Viral DNA is transmitted to the host as proviral DNA in wasp chromosomes (thin arrows) and as circular DNA molecules within virions (thick arrows). Proviral DNA is the viral progeny and makes a complete cycle, whereas 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).

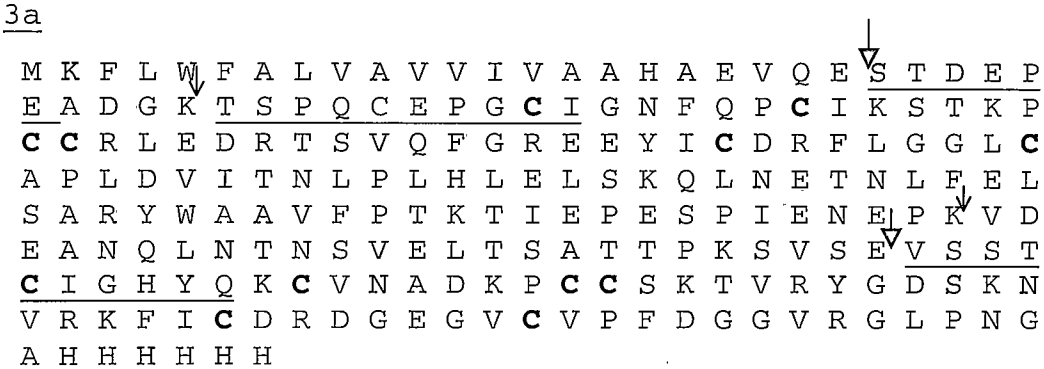
The endoparasitic *Campoletis sonorensis* wasp carries polydnaviruses of the ichnovirus genera. Among the estimated ten thousand ichnovirus species, the *Campoletis sonorensis* ichnovirus (CsIV) is the most studied (11). Injection of purified CsIV into

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