2010

Subcellular Localization of the Non-Structural Proteins 3C and 3CD of the Honeybee Virus Deformed Wing Virus

Cierra Nichole Danko
Bucknell University

Follow this and additional works at: https://digitalcommons.bucknell.edu/honors_theses

Part of the Medical Sciences Commons

Recommended Citation
Danko, Cierra Nichole, "Subcellular Localization of the Non-Structural Proteins 3C and 3CD of the Honeybee Virus Deformed Wing Virus" (2010). Honors Theses. 35.
https://digitalcommons.bucknell.edu/honors_theses/35

This Honors Thesis is brought to you for free and open access by the Student Theses at Bucknell Digital Commons. It has been accepted for inclusion in Honors Theses by an authorized administrator of Bucknell Digital Commons. For more information, please contact dcadmin@bucknell.edu.
Subcellular Localization of the Non-Structural Proteins 3C and 3CD of the
Honeybee Virus Deformed Wing Virus

by

Cierra N Danko

A Proposal Submitted to the Honors Council
For Honors in Biology

10 May 2010

Approved by:

Adviser: Dr. Marie Pizzorno

Department Chairperson: Dr. Marie Pizzorno
Acknowledgements

First, I would like to thank Dr. Marie Pizzorno for mentoring, support, guidance, and invaluable input (not to mention late afternoon chats about nothing that has to do with biology and long griping/crying sessions to keep my sanity). I would also like to thank my mother, Christine Casson, for putting up with her “budding scientist” for twenty-two long years. I could not have made it through this thesis without them, Sophia Drinis, 4SE, Karan Shah, Gary Deng, Chen Mao, and the ingenious invention of the “Turbo Shot” by Dunkin Donuts.

I need to thank the Bucknell University Biology Department, Dr. Matthew Heintzelman for antibodies, Dr. Kathleen Page and Joshua Ripple for SDS-PAGE supplies, Dr. Joe Moore and Dr. Julie Gates for use of the confocal microscope, John Tomtishen for cells, and Tom Vogler for support in cloning, analysis, and through all the silly mistakes I made in the lab.

A special thanks to my committee: Dr. Marie Pizzorno

Dr. Elizabeth Marin

Dr. Thomas Selby
# Table of Contents

Abstract...........................................................................................................................................1

Introduction........................................................................................................................................3

*Honeybee Viruses*..........................................................................................................................3

*Deformed Wing Virus (DWV) and Colony Collapse Disorder (CCD)*.................................5

*Transmission of Deformed Wing Virus (DWV)*...........................................................................7

*Effect of Deformed Wing Virus (DWV) on Infected Colonies*..................................................8

*Picornavirus Structure and Replication*.......................................................................................9

*Characterization of Deformed Wing Virus (DWV)*.................................................................11

*Poliovirus: A Human Analog*........................................................................................................15

*Green Fluorescent Protein (GFP) as a Localization Tag*.............................................................16

*Hypothesis*......................................................................................................................................18

Materials and Methods..................................................................................................................19

*Polymerase Chain Reaction (PCR) and Isolation of Constructs*................................................19

*Insertion of Constructs into TA-cloning Vectors*......................................................................21

*Transformation of One Shot Cells and Isolation of Colonies Containing Inserts*........21

*Miniprep Procedure and Restriction Digests*.............................................................................22

*Gel Isolation and Insertion of Constructs into EGFP Vectors*...............................................22

*Transformation of DH5α Cells and Isolation of Colonies Containing Inserts*..................23

*Transfection and Western Blotting*..............................................................................................24

*Transfection and Confocal Microscopy*......................................................................................25
Results..................................................................................................................26
Discussion..............................................................................................................39
Bibliography...........................................................................................................45
Appendix A: Sequences of TA clone inserts and NCBI amino acid blast analyses........52
Appendix B: Original GFP/DAPI merged confocal images.......................................59
List of Tables

Table 1  Sequence of primers used to isolate DWV non-structural coding regions........20
Table 2  Predicted size of insert fragments compared to observed size..........................34
Table 3  Predicted molecular weight of fusion proteins compared to observed weight....36
List of Figures

Figure 1  The Deformed Wing Virus (DWV) genome..................................................14
Figure 2  Orientations of eight total DWV/GFP fusion proteins..................................32
Figure 3  Agarose gel electrophoresis of digested DWV/GFP fusion proteins.............33
Figure 4  Western blot of transfected DWV/GFP fusion proteins...............................35
Figure 5  Subcellular localization of DWV/GFP fusion proteins.................................37
Figure 6  Subcellular localization of DWV/GFP fusion proteins (continued).................38
Abstract

*Apis mellifera* L., the European honeybee, is a crucial pollinator of many important agricultural crops in the United States. Recently, honeybee colonies have been affected by Colony Collapse Disorder (CCD), a disorder in which the colony fails due to the disappearance of a key functional group of worker bees. Though no direct causal relationship has been confirmed, hives that experience CCD have been shown to have a high incidence of Deformed Wing Virus (DWV), a common honeybee virus. While the genome sequence and gene-order of DWV has been analyzed fairly recently, few other studies have been performed to understand the molecular characterization of the virus. Since little is known about where DWV proteins localize in infected host cells, the objective of this project was to determine the subcellular localization of two of the important non-structural proteins that are encoded in the DWV genome. This project focused on the protein 3C, an autocatalytic protease which cleaves itself from a longer polyprotein and helps to cut all of the other proteins apart from one another so that they can become functional, and 3D, the RNA-dependent RNA polymerase (RdRp) which is critical for replication of the virus because it copies the viral genome. By tagging nested constructs containing these two proteins and tracking where they localized in living cells, this study aimed to better understand the replication of DWV and to elicit possible targets for further research on how to control the virus. Since DWV is a picorna-like virus, distantly related to human viruses such as polio, and picornavirus non-structural proteins aggregate at cellular membranes during viral replication, the major hypothesis was that
the 3C and 3CD proteins would localize at cellular organelle membranes as well. Using confocal microscopy, both proteins were found to localize in the cytoplasm, but the 3CD protein was found to be mostly diffuse cytoplasmic, and the 3C protein was found to localize more specifically on membranous structures just outside of the nucleus.
Introduction

Honeybee Viruses

In general, the study of the molecular biology of insect viruses is a relatively new and expanding field; though researchers have known about insect viruses since the early 1900’s, molecular characterization studies were not performed on insect viruses until the 1970’s (Friesen and Miller 2001). The study of insect viruses is important because it can elicit valuable information about both the mechanisms present in insect viruses that have analogs in human pathogens and about cellular processes themselves. For example, the presence of a 5’ cap (7-methyl-guanosine) on processed eukaryotic mRNAs was discovered while studying a member of the reovirus family that infects multiple insects (Friesen and Miller 2001; Furuichi and Miura 1975). The study of honeybee viruses is particularly important because honeybees are critical pollinators of many crops in the United States, and viral infection can compromise the critical role that honeybees play in agricultural crop pollination (McGregor 1976).

Of the eighteen viruses known to infect honeybees, sixteen of them are picorna-like viruses with positive-sense, single-stranded RNA genomes (Allen and Ball 1996). The six most common viral pathogens of *Apis mellifera*, the European honeybee, are Deformed Wing Virus (DWV), Acute Bee Paralysis Virus (ABPV), Chronic Bee Paralysis Virus (CBPV), Sacbrood Virus (SBV), Black Queen Cell Virus (BQCV), and Kashmir Bee Virus (KBV) (Chen and Siede 2007). DWV is the most common honeybee virus and is spread through either contact with infected food, eggs laid by an infected
queen, or parasitism by an infected Varroa mite. The virus gets its name from the crumpled wings that are observed on adult bees during symptomatic infection.

SBV is the most geographically wide-spread honeybee virus, and it is a pathogen that attacks mainly younger bees, especially larvae that are approximately two days old (Bailey and Ball 1991). The virus gets its name from the “sac-like” appearance of infected larvae (Chen and Siede 2007). BQCV attacks larvae in colonies that are in the midst of rearing a queen, and its prevalence has a strong association with infection by the protozoan Nosema apis. The virus gets its name from the black color that the queen bee cells in the hive turn after the larvae have been killed by the virus (Bailey and Ball 1991). KBV infects honeybees through direct cuticle contact rather than through feeding, and though symptomatic infection is much less common than with the other honeybee viruses, studies have shown that KBV is the most potent, killing some bees within three days (Bailey et al. 1979). The virus gets its name from the region of India in which it was first discovered. ABPV causes trembling and inability to fly in infected bees, and it kills honeybees relatively quickly. The virus gets its name from the sudden paralysis and death experienced by the infected bees. CBPV was one of the first viruses isolated from honeybees, and though it produces the same paralysis symptoms as ABPV, the virus takes longer to actually kill the honeybees because it is less virulent. The virus gets its name from the persistent symptoms of paralysis that the bees experience before death (Bailey et al. 1963). Though each virus is pathogenic on its own, many of the common honeybee viruses show an increase in prevalence in conjunction with the presence of another honeybee parasite. For example, the presence of the ecto-parasitic mite, Varroa
*destructor*, has been correlated with increased viral levels of DWV, SBV, KBV, and ABPV (Chen and Siede 2007).

**Deformed Wing Virus (DWV) and Colony Collapse Disorder (CCD)**

*Apis mellifera* is a crucial pollinator of many important agricultural crops in the United States (McGregor 1976; Sabara and Winston 2003). Recently, honeybee colonies have been affected by Colony Collapse Disorder (CCD), an ecological phenomenon that is caused by the disappearance of adult worker honeybees. CCD is characterized by honeybee colonies that fail due to the rapid loss of a critical functional group of bees—the adult foragers (vanEngelsdorp et al. 2009). When a colony experiences a decrease in the number of adult foragers, the remaining younger bees in the hive rapidly mature to compensate for the lack of adult workers (Robinson 1992). When young honeybees were reared in isolation without the presence of any adult foragers, Huang and Robinson (1992) found that the expression of juvenile hormone increased dramatically, and the expression of juvenile hormone induces the rapid and premature maturation of younger bees in hives which experience a loss of adult workers prior to colony collapse. This rapid maturation of young bees causes added stress on the entire colony because there is suddenly a decrease in the number of bees serving other roles inside the hive, such as nurse bees that typically feed the brood. This effect is further exaggerated during environmentally stressful conditions, such as over the winter months, when a lack of food
can combine with a decrease in the number of forager bees to cause the colony to collapse (Tentcheva et al. 2004).

Currently, CCD is speculated to be caused by viruses, and though no direct causal link has been established, a microarray analysis comparing transcripts found in CCD affected (CCD positive) and non-affected (CCD negative) colonies showed that many picorna-like viral RNAs were much more prevalent in CCD positive colonies as compared to CCD negative colonies (vanEngelsdorp et al. 2009, Johnson et al. 2009). The microarray analysis also screened for up-regulation of detoxification genes in the gut of the bees because pesticides were once suspected to contribute to CCD. However, no significant difference in expression of these genes was found between CCD positive and CCD negative colonies (Decourtye et al. 2004, Johnson et al. 2009). Two of the viral “front-runners” attributed to causing CCD are Israeli Acute Paralysis Virus (IAPV) and Deformed Wing Virus (DWV), but the microarray analysis by Johnson et al. (2009) that included a broad sampling of CCD colonies from different regions of the United States showed that DWV had the highest expression level in CCD positive colonies compared to CCD negative colonies (Cox-Foster et al. 2007, Palacios et al. 2008, Johnson et al. 2009). Also, of the sixteen picorna-like viruses known to infect honeybees, DWV is the most common to infect *Apis mellifera* (Allen and Ball 1996, Johnson et al. 2009).

Even though recent microarray studies have shown that DWV is the viral “front-runner” for inducing CCD, no direct causal relationship has been established because there are many other pathogens which can infect honeybees along with DWV (Johnson et al. 2009, de Miranda and Genersch 2010). For example, in screening for the infection of
Apis mellifera L. by six picorna-like honeybee viruses [Deformed Wing virus (DWV), Acute Bee Paralysis virus (ABPV), Chronic Bee Paralysis virus (CBPV), Sacbrood virus (SBV), Black Queen Cell virus (BQCV), and Kashmir bee virus (KBV)] in Devon, England using reverse-transcription PCR (RT-PCR), Baker and Schroeder (2008) found that 97% of the 69 analyzed colonies showed some level of DWV infection, and 32% of the colonies showed multiple viral infections. Therefore, DWV may be only one of multiple stressors on the honeybee colonies leading to their collapse.

Transmission of Deformed Wing Virus (DWV)

DWV can be transferred between honeybees through both horizontal and vertical transmission. The virus is transmitted both through the presence of Varroa mites, ectoparasites on the bees which act as vectors for carrying the virus, and through direct feeding contact between adults and young bees (Chen et al. 2006). The spread of DWV to different regions of the world has been linked to the presence of Varroa destructor mites (Bowen-Walker et al. 1999, Calderon et al. 2003). Varroa mites transmit the virus from an infected bee to one that has not yet been infected, and in fact, Varroa mites and their associated pathogens are responsible for the mortality of greater than one third of the managed honeybee population in the United States (Allen and Ball 1996). This transmission of the virus by mites is part of the horizontal transmission pathway of DWV, and it takes place as the mites infest an uninfected bee and feed on that bee after having previously fed on an infected honeybee (Bowen-Walker et al. 1999). Vertical
transmission, on the other hand, occurs when an infected worker bee feeds a younger, uninfected bee. In the same light, an infected queen can vertically transfer the virus to its progeny by producing infected royal jelly, which is used as a food source for larvae, or by physically laying DWV-infected eggs (Chen et al. 2006; De Miranda and Fries 2008).

Effect of Deformed Wing Virus (DWV) on Infected Colonies

The presence of a high density of DWV in a colony can suppress the normal functioning of honeybee immune systems. In addition, infection of a colony with DWV can lead to either symptomatic or asymptomatic infection, depending on the condition of the honeybee immune systems. The immune systems of the bees can be affected by both environmental conditions and the presence of ecto-parasitic mites, such as the Varroa mites, whose presence correlates with a suppression of immunity-related genes (Yang and Cox-Foster 2005). A typical hive with a symptomatic infection contains bees that show physical signs of infection such as crumpled wings, bloated abdomens, paralysis, learning deficits, and a shortened life span (Bailey and Ball 1991). Though studies have shown that asymptomatic bees infected with DWV have impaired associative learning and cannot form memories, they also tend to have lower virus concentrations (Iqbal and Mueller 2007; Bailey and Ball 1991). An asymptomatic colony can be transformed into a symptomatic colony during environmentally stressful conditions, such as winter, when the bees are restricted to the hive, have less food available, and are in close proximity to
one another. This leads to a decrease in performance among adult worker bees, which can lead to collapse of the colony (Tentcheva et al. 2004).

Though the presence of Varroa mites has been shown to increase the severity of symptomatic DWV infections, such as the increased incidence of physical wing deformities, DWV has also been implicated in the collapse of honeybee colonies over the winter months even with very low presence of the Varroa mites (Yang and Cox-Foster 2005, Highfield et al. 2009). This suggests that DWV alone could cause the colony to collapse under stressful conditions.

**Picornavirus Structure and Replication**

Based on their genome structure and replication cycles, the majority of honeybee viruses, including DWV, are most closely related to picornaviruses. The Picornaviridae family is a large family of viruses that infect mammals, including humans. Members of Picornaviridae include poliovirus and the human rhinovirus, the cause of the common cold. These are relatively small, non-enveloped viruses with a genome consisting of single-stranded, positive-sense RNA. The capsids of picornaviruses are usually composed of four structural proteins — VP1, VP2, VP3, and VP4, though some viruses only have three capsid proteins. These proteins form a twenty-sided icosahedral capsid shell. The capsid proteins interact with receptor proteins on the host-cell surface, which vary for each individual virus, and a conformational change causes the virus to either fuse with part of the membrane and release the genome into the cellular cytoplasm or to be
taken up into the cell by endocytosis where acidification of the endosome causes the genome to be released. Since the picornaviral genome is a single-stranded, positive sense RNA, it can be translated immediately by cellular ribosomes upon entry into the cytoplasm. The viral RNA must be translated before replication can occur because picornaviruses do not package any enzymatic proteins into their capsids to aid in replication (reviewed in Racaniello 2001).

Since the picornavirus genome does not have a 5’ cap like typical cellular messenger RNA (mRNA) to attract the ribosome for initiation of translation, the genome must have another element to outcompete the cellular mRNAs for attraction of the ribosomes. The 5’ UTR of the genome contains a region of complex secondary and tertiary structure with many stem loops that acts as an IRES (Larsen et al. 1981). The IRES has a high affinity for cellular translation factors, which in turn attract and position the ribosome to initiate translation at the correct AUG (methionine) start codon (reviewed in Racaniello 1991). The genome is translated as one long polyprotein that is subsequently cleaved into individual functional proteins by viral proteases, such as 3C. Since cellular RNA is never copied into complementary RNA during normal cellular processes, there are no cellular RdRp enzymes. Therefore, the picornavirus genome must encode its own polymerase (RdRp) for replication. After proteolytic processing releases the 3D RdRp protein, the viral polymerase copies a strand of the genome into a full-length negative-sense RNA template. One negative-strand template is then used by the same RdRp to produce many copies of the full-length positive-sense genome.

Picornavirus replication occurs on cytoplasmic cellular membranes and not within the
nucleus, though some viral proteins may enter the nucleus to prevent transcription of cellular mRNA (Caliguir and Tamm 1970; Sharma et al. 2004). After replication is complete, the capsid proteins assemble and package only the positive-sense RNA strands as the viral genome (reviewed in Racaniello 1991).

Characterization of Deformed Wing Virus (DWV)

Deformed Wing Virus (DWV) was first identified by scientists in Japan in the 1980’s, but the first thorough molecular characterization of the virus was not performed until 2006 (Allen and Ball 1996; Lanzi et al. 2006). DWV’s single-stranded RNA genome is enclosed within an icosahedral capsid that is about 30 nm in diameter, and its three capsid proteins, VP1, VP2, and VP3, closely resemble those of other picorna-like viruses. These proteins are structural proteins that form the capsid shell that surrounds and protects the viral genome. In addition to capsid proteins, the DWV genome also codes for non-structural proteins which are important for viral replication, such as an RNA helicase to unwind any transient double-stranded RNA, the 3C protease, and a 3D RNA-dependent RNA polymerase (RdRp) (Figure 1). The genome contains a long 5’ UTR, which most likely acts as an internal ribosome entry site (IRES) to help the viral genome bind to host-cell ribosomes. This is critical because viruses do not produce their own ribosomes for protein translation, and the DWV genome does not have a 5’ cap structure like normal cellular mRNAs to attract the cellular ribosome. The genome is approximately 10,000 nucleotides in length (10 knt), and DWV has been shown to be closely related to Varroa
destructor virus 1 (VDV-1), SBV, and *Ectropis oblique* picorna-like virus (EoPV) (Lanzi et al. 2006).

Before the genome of DWV was analyzed in 2006, a very similar RNA virus called Kakugo virus (KV) was identified by Fujiyuki et al. in 2004. When Fujiyuki et al. exposed *Apis mellifera* worker bees to a giant hornet (*Vespa mandarinia japonica*) hanging from a string, they noticed that many of the honeybees aggressively attacked the hornet. After collecting the attacking honeybees and comparing the RNA sequences found in their brains to RNA from non-aggressive honeybees collected as a control, the group identified a novel RNA present in the aggressive bees that resembled the genome of a picorna-like virus. The authors reported that the viral RNA was most closely related to SBV, and they mentioned that even though their novel RNA was shown to share 97-98% sequence homology with DWV, they still believed that KV was a completely separate virus from DWV because of almost 200 nucleotide differences between the two genomes and a slightly longer 5’ untranslated region (UTR) in KV (Fujiyuki et al. 2004). However, the differences in the lengths of the 5’ UTR may be due to technical difficulties in obtaining full-length cDNA clones at the 5’ end of the viral genome. In a subsequent paper, Fujiyuki et al. (2006) went on to report that multiple strains of KV that they surveyed were in fact the same virus because the RNA encoding for their RNA-dependent RNA polymerases (RdRp) only varied in sequence homology by less than 2%. As Fujiyuki et al. (2006) stated, it is very hard to consider RNA viruses with less than 2% variation in sequence homology to be separate viruses. Because of this discrepancy, even though Fujiyuki et al. consider KV to be completely different from DWV (even though
their RNA sequences are ~98% homologous), KV was considered to be indistinguishable from DWV for the purpose of this particular project. Because of the extreme sequence homology between the viruses, the primers designed in this study to produce the 3C and 3CD protein coding sequences of DWV would not selectively amplify DWV over Kakugo, and it is possible that Kakugo virus 3C and 3CD proteins could have been isolated as well.

Both DWV and KV are members of the *Iflavirus* genus, a genus that was established in 2006 by the International Committee on Taxonomy of Viruses (ICTV) to account for the monocistronic RNA insect viruses, which are more closely related to the animal picornaviruses than members of the insect *Dicistroviridae* family (Mayo and Ball 2006; Lanzi et al. 2006). Monocistronic viruses, such as DWV and poliovirus, have their genomes translated by host cell ribosomes into one large polyprotein from a single open reading frame (ORF), with the structural proteins at the 5’ end and the non-structural proteins at the 3’ end to generate multiple proteins from a single, small RNA genome. Utilizing proteolytic cleavage of a polyprotein maximizes the efficiency of encoding the proteins needed for viral replication and subsequent infection within such a small genome (reviewed in Flint et al. 2004). Dicistronic viruses, such as Cricket paralysis virus (CrPV), contain two ORFs in their genome and thus produce two separate polyproteins—the polyprotein translated from the 5’ end of the genome contains the non-structural proteins and the polyprotein translated from the 3’ end of the genome contains the structural proteins. The genomes of the dicistroviruses contain two IRES sequences, one at the beginning of each ORF, where the cellular ribosome binds to initiate translation of
the viral polyproteins (reviewed in Racaniello 2001). Thus, the members of the *Dicistroviridae* family posses more control over the amounts of structural and non-structural proteins that they produce because the two ORFs are translated independently.

Figure 1 The Deformed Wing Virus (DWV) genome [taken from Lanzi et al. 2006]. The important non-structural proteins, including 3C and 3D (RdRp), are located at the 3’ end of the genome. Solid vertical lines indicate predicted cleavage sites with their amino acid sequences shown above the genome. Viral 3C proteases typically cleave after glutamine (Q) or glutamic acid (E) (Lanzi et al. 2006).
Poliovirus: A Human Analog

The closest human viral analog to DWV is poliovirus, a member of the picornavirus family. Just like DWV, the genome of poliovirus is first translated into one long polyprotein, and that polyprotein is subsequently cleaved into smaller individual structural and enzymatic proteins (Jacobson and Baltimore 1968). Unlike DWV, the poliovirus genome encodes two viral proteases, 2A and 3C, which are both autocatalytic as well. The 2A protease cleaves between specific tyrosine (Tyr) and glycine (Gly) residues in the polyprotein, while the 3C protease cleaves between glutamine (Gln) and Gly (reviewed in Flint at al. 2004). This project focuses on the 3C protein from DWV because it performs critical cleavages of both viral and cellular proteins within the infected host cell. For example, poliovirus 3C has been shown to cleave human TATA-binding protein, a critical transcription factor required for the transcription of host cell mRNA, in addition to most of the viral polyprotein (Clark et al. 1993). Thus, 3C is responsible for maintaining viral persistence through both a positive effect on viral processes (polyprotein cleavage) and a negative effect on host processes (through suppression of cellular transcription). Because of this dual functionality and lack of information on the actual subcellular localization of DWV 3C, the 3C protein presented itself as an intriguing candidate for this particular molecular characterization study.

Since poliovirus is a model human picornavirus, the localization of its proteins has been studied. While determining the subcellular localization using Green Fluorescent Protein (GFP) as a tag, Sharma et al. (2004) found that the poliovirus 3CD and 3D
proteins remained cytoplasmic when they were tagged alone, but they found that both proteins localized to the nucleus of HeLa cells when the cells were super-infected with poliovirus after transfection with the GFP fusion-protein plasmids. A nuclear localization signal (NLS), which targets proteins for transport into the nucleus, was discovered within the 3D protein which contained three lysines (K) in a row (KKKRD). When the NLS was mutated to KKAAA, substituting neutral alanines for the charged lysines and arginine, neither 3CD nor 3D localized to the nucleus after super-infection (Sharma et al. 2004). Though 3C alone has not been shown to localize to the nucleus during early poliovirus infection, high concentrations of 3C have been shown to inhibit RNA Polymerase III transcription in mammalian cells, which occurs in the nucleus (Clark et al. 1991).

**Green Fluorescent Protein (GFP) as a Localization Tag**

Green Fluorescent Protein (GFP) has been successfully used as a tag to track the subcellular localization of many cellular and viral proteins (reviewed in Tsien 1998), including the picornavirus non-structural proteins (Sharma et al. 2004, Ghildyal et al. 2009). GFP has been utilized successfully in many organisms and tissue culture systems, ranging from the yeast *Saccharomyces cerevisiae* to mammalian cell lines such as HeLa and HEK 293 (Human Embryonic Kidney) (Cubitt et al. 1995). For example, when a mitochondrial localization signal was added to GFP expressed in HeLa cells, the
fluorescent proteins correctly localized to the mitochondria (Rizzuto et al. 1995; Cubitt et al. 1995).

What is now utilized as GFP was first discovered by Shimomura et al. (1962) in conjunction with the aequorin protein in the jellyfish *Aequorea*. When aequorin fluoresced blue (at around a wavelength of light near 470 nm), Shimomura et al. (1962) noticed that GFP responded with a bioluminescence that gave off green light around a wavelength of 508 nm (Tsien 1998). The GFP chosen for use in the experiments described here was enhanced GFP (EGFP), and according to Shaner et al. (2005), EGFP’s peak fluorescence occurs when it is illuminated with 488 nm blue light, and its peak emission gives off light of 507 nm wavelength. Many variants of GFP have been developed through mutation of the original protein because the wild-type protein has a complex expression spectrum. The variants are divided into seven classes based on unique components present within their chromophore, the specific region of GFP that is responsible for generating the observed color through fluorescence. Enhanced GFP, which is part of class 2, has a phenolate anion in its chromophore due to substitution of threonine (Thr) for serine (Ser) at the 65th amino acid position from wild-type GFP (Tsien 1998). This mutation suppresses a secondary excitation peak near 375 nm found in wild-type GFP, which simplifies the excitation and emission pattern of the protein and increases brightness by amplifying the peak near 507 nm nearly six-fold as well (Heim et al. 1995). According to Brejc et al. (1997), the extra methyl group attached to Thr in the mutated chromophore of EGFP adds steric hindrance, which prevents the normal ionization of glutamate (Glu), thus affecting the hydrogen bond network and allowing the
chromophore to become anionic instead of taking on the typically neutral state of the wild-type chromophore (Ormö et al. 1996, Tsien 1998).

**Hypothesis**

Currently, there is no available insect tissue culture system for growing or infecting cells with DWV. Therefore, it was necessary to used cloned genes of individual viral proteins in this study. In addition, there are no antibodies available commercially that recognize the non-structural proteins of DWV, so portions of the non-structural coding region of the virus were fused to EGFP to study their expression and localization in mammalian cells.

Since DWV is a picorna-like virus, distantly related to human viruses such as polio, and picornavirus non-structural proteins aggregate at cellular membranes during viral replication, the major hypothesis of this project was that the 3C and 3CD proteins would localize at cellular membranes in the cytoplasm (Moore and Eley 1991; Caliguiri and Tamm 1970). The subcellular localization of each viral protein was visualized using confocal microscopy.
Materials and Methods

*Polymerase Chain Reaction (PCR) and Isolation of Constructs*

One single forward primer and four reverse primers for PCR amplification of each of the four constructs were designed using the amino acid boundaries of 3C and 3D reported by Lanzi et al. (2006). The primers contained between 35 and 42 nucleotides each, including a Kozak sequence to help initiate translation (Kozak 1986) followed by a HindIII restriction enzyme cleavage site in the forward primer and a BamHI restriction enzyme cleavage site in each of the reverse primers, all in the correct reading frame with the pEGFP-N1 and pEGFP-C1 vectors (Table 1). Each primer was designed to contain ~50% guanine and cytosine nucleotides, and the primers were ordered from Eurofins MWG Operon.

Plasmid DNA containing the cDNA from the 3’ end of the DWV genome (Deng and Pizzorno, unpublished data) was diluted 1:99 in distilled water (dH₂O), and PCR mixes were prepared using 2x Master Mix D, 1µL of cDNA, and 10µL of each respective primer, following the manufacturer’s protocol (Epicentre). PCR was performed for 35 cycles of 94°C for 1 minute, 60°C for 1 minute, and then 72°C for 2 minutes. The PCR products were electrophoresed on a 1% agarose gel at 80 V for one hour and stained with ethidium bromide to verify that the isolated fragments were the expected size predicted from the DWV genome sequence (Lanzi et al. 2006).
Table 1  Sequence of primers used to isolate DWV non-structural coding regions (Kozak sequence in bold)

<table>
<thead>
<tr>
<th>Primer Type</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer</td>
<td>5’ GTCGTGCAAGCT<strong>TCCGCCAT</strong>GGAAACCAAGTACTATTTTAAG 3’</td>
</tr>
<tr>
<td>First Reverse Primer (to isolate 3C)</td>
<td>5’ GTAGTAGGATCCATTCCTTCAGTACCAGCAACATG 3’</td>
</tr>
<tr>
<td>Second Reverse Primer (to isolate 3C+)</td>
<td>5’ GTAGTAGGATCCCAGGATATCGAATCAAACCATC 3’</td>
</tr>
<tr>
<td>Third Reverse Primer (to isolate 3CD)</td>
<td>5’ GTAGTAGGATCCATAGCATGAGTCCAATTCGCTG 3’</td>
</tr>
<tr>
<td>Fourth Reverse Primer (to isolate 3CD+)</td>
<td>5’ TTAGTAGGATCCATCGCGCTAGCCTAGCTAACATCTAAC 3’</td>
</tr>
</tbody>
</table>
Insertion of Constructs into TA-cloning Vectors

After isolation of each PCR DNA band, a ligation was performed with commercially prepared pCRII vector from a TA cloning kit using T4 DNA ligase, following the manufacturer’s protocol (Invitrogen). TA-cloning was performed prior to EGFP-cloning because TA plasmids can utilize blue/white screening for the presence of inserts and transformation is more efficient. In addition, sequencing of the TA plasmids allows for confirmation of the sequence before subcloning into the EGFP vectors.

Transformation of One Shot® Cells and Isolation of Colonies Containing Inserts

The ligation products were each transformed into One Shot® competent INVαF’ cells following the manufacturer’s protocol (Invitrogen), except that SOB medium was used instead of SOC medium. To grow up the bacterial colonies, 80µL of 2% X-gal (20ng/µL in N-methylformamide) was added to each Luria broth (LB) agar plate containing ampicillin (100 µg/mL), since the TA-plasmids contain an ampicillin resistance gene, and 75µL of transformed bacteria was spread onto each plate. The plates were incubated at 37°C overnight. White colonies from each plate were picked and were grown in 4mL of LB containing 100 µg/mL of ampicillin at 37°C in a shaking incubator overnight.
Miniprep Procedure and Restriction Digests

Miniprep procedures were performed on each of the transformed bacterial cultures to isolate the TA-plasmid DNA, following the manufacturer’s protocol (QIAGEN). After isolation, 5µL of each miniprep DNA was digested with EcoRI in a total reaction volume of 20µL for 2.5 hours at 37°C. The digested miniprep DNA was electrophoresed on a 1% agarose gel at 80V with ethidium bromide to verify the presence of each insert in their respective TA plasmids. After verification that the plasmid contained an insert, the original miniprep DNA was digested sequentially with HindIII in Buffer 2 incubated at 37°C for 1.5 hours followed by BamHI (with 9µL Tris, 6µL NaCl, and 2µL BSA added) incubated at 37°C for an additional 1.5 hours in a total reaction volume of 200µL. In conjunction, pEGFP-N1 and pEGFP-C1 plasmids were sequentially digested with HindIII and BamHI as well.

Gel Isolation and Insertion of Constructs into EGFP Vectors

The total volume of sequentially digested miniprep DNA was run on a 1% agarose preparative gel at 80V. The gel was soaked briefly in an ethidium bromide/water bath to visualize the DNA bands. The fragment corresponding to each insert was cut out of the gel, and the DNA was extracted from the gel using a QIAEX II isolation kit following the manufacturer’s protocol (QIAGEN). The total volume of each sequentially digested EGFP plasmid DNA was phenol-chloroform extracted and then ethanol precipitated (Sambrook et al. 1989). The DNA pellet was resuspended in 20 µL of sterile dH₂O. The
concentration of each sample was determined using a NanoDrop1000, and the concentrations were verified visually by running a small volume of each insert and EGFP plasmid next to a DNA low-mass ladder on a 1% agarose gel containing ethidium bromide. Ligation of each sequentially digested product into each of the EGFP vectors was performed using T4 DNA Quick Ligase, following the manufacturer’s protocol (New England BioLabs).

*Transformation of DH5α Cells and Isolation of Colonies Containing Inserts*

The ligation products were each transformed into DH5α competent cells following the manufacturer’s protocol (Invitrogen). To select for bacterial colonies containing the pEGFP vector, 200µL of each transformed bacterial solution was added to an LB plate containing kanamycin (30 µg/mL) since the EGFP plasmids contain a kanamycin resistance gene, and the plates were incubated at 37°C overnight. Since there is no blue/white screening for EGFP plasmids, colonies from each plate were chosen at random and were grown up in 4mL LB containing 30 µg/mL kanamycin overnight. Minipreps were performed on each culture as above, and the miniprep DNA was sequentially digested with *HindIII* in Buffer 2 and then with *BamHI* in Buffer 3 at 37°C in a final total reaction volume of 20µL (New England Biolabs). The sequentially digested miniprep DNA was electrophoresed on a 1% agarose gel containing ethidium bromide at 80V for one hour to screen for the presence of each insert. Once a plasmid was found to contain an insert of the correct size, the original culture that produced that
miniprep DNA was grown up in 50mL of LB containing 30 μg/mL kanamycin at 37°C in a shaking incubator overnight. Midiprep procedures were performed on each of the bacterial cultures to isolate the EGFP plasmid DNA containing the inserts, following the manufacturer’s protocol (QIAGEN).

Transfection and Western Blotting

Twenty-four hours prior to transfection, ~2x10⁵ COS-7 cells were plated in 5ml of DMEM/10% FCS media without antibiotics in each well of two 6-well plates. The cells in each well were transfected with an experimental plasmid or pEGFP-N1 by itself as a control using Lipofectamine™2000 following the manufacturer’s protocol (Invitrogen). The media was changed the next day to DMEM/10% FCS with penicillin and streptomycin to prevent contamination. The cells from each well were harvested by scraping in 1x PBS at 48 hours post-transfection. After harvesting, the cells were centrifuged for one minute, the PBS was removed, and the cells were resuspended in 25μL of cold PBS. An equal volume (25μL) of 2x protein sample buffer (4% SDS) was added, the solutions were heated at ~95°C for 5 minutes, and the tubes were centrifuged for five minutes to pellet any cell debris. β-mercaptoethanol (4% final concentration) was not added to the cell lysates until just before loading the gel.

Each transfected cell lysate was loaded onto a 10% SDS-PAGE gel, which was run at 200 V for one hour. The gel was then blotted to nitrocellulose paper in Western Transfer buffer (25mM Tris, 192mM glycine, 0.006% SDS, 20% methanol) at 75V for
two hours. The blot was blocked in 10% dry milk/1xTBS for two hours at room temperature. The blot was then incubated in 1% dry milk/1x TBS containing a monoclonal antibody to GFP at a 1:1000 dilution (JL-8 antibody from Clontech) for one hour, washed with TBST (1x TBS, 0.5% Tween-20), and then incubated with a rabbit-anti-mouse antibody conjugated to alkaline phosphatase (Santa Cruz Biotechnology) for 45 minutes. After extensive washing, the blot was incubated with NBT/BCIP as substrates for color development by the alkaline phosphatase enzyme.

Transfection and Confocal Microscopy

Twenty-four hours prior to transfection, ~1x10⁵ HeLa cells were plated in 500µL of DMEM/10% FCS without antibiotics in each well of a 24-well plate containing coverslips. The cells in each well were transfected with an experimental plasmid or pEGFP-N1 by itself as a control using Lipofectamine™2000 following the manufacturer’s protocol (Invitrogen). The media was changed once over ~48 hours to DMEM/10% FCS with penicillin and streptomycin to prevent contamination.

Approximately 48 hours post-transfection, the media was removed from the wells and the cells were washed once with 1xPBS. They were then fixed in 4% paraformaldehyde/1x PBS for 10 minutes at room temperature. The cells were then washed three times with PBST (1x PBS, 0.1% Triton-X 100, 0.05% Tween 20) and once in 1x PBS. Coverslips were removed from the 1xPBS, quickly rinsed in dH₂O, and then mounted with DAPI containing mounting media (Vector Laboratories). The coverslips
were sealed with clear nail polish and allowed to set for 1 hour. Cells were then imaged on a Leica SP5 confocal laser scanning microscope. Images were processed using the LAS AF software.

Results

To determine where the 3C and 3CD proteins localize in cells, the part of the DWV genome that codes for each of the proteins was isolated from cDNA of the 3’ end of the DWV genome and was inserted into EGFP plasmids. In designing the plasmids, nested constructs were chosen to more closely mimic the structure of the functional protease generated by DWV during infection in the host cell. In most picornaviruses, 3C is autocatalytic and cleaves itself out of the polyprotein (a “cis-cleavage” event) to further process the rest of the same polyprotein. The 3C protease also processes other complete polyproteins (a “trans-cleavage” event), and generates a regulated amount of 3CD protein, which consists of 3C and 3D cleaved out of the polyprotein while still joined together (Flint et al. 2004). Though poliovirus 3C has been shown to have protease activity on its own, that activity increases when it remains attached to the 3D protein. Likewise, 3C must be attached to 3D (as 3CD) for viral replication to occur, and it is thought that 3CD mediates the binding of the polymerase (3D by itself) to the RNA template (reviewed in Semler and Wimmer 2002; Cornell and Semler 2002).

Isolation of the genes encoding the desired protein regions was accomplished by designing primers that were used in PCR amplification of specific genome segments and
which contained restriction enzyme sites that aided in inserting the DNA fragments into the pEGFP plasmids. The primers for this study were designed to amplify the 3C protein (3C), 3C plus the extra nucleotides before 3D (3C+), 3C through 3D (3CD), and 3C through the end of the genome (3CD+) (Figure 1), using the gene sequences reported by Lanzi et al. (2006), which are also available in GenBank.

Before inserting the isolated constructs into the EGFP vectors, the preliminary TA clones containing the amplified segments were sequenced. According to a nucleotide BLAST search using NCBI, the 3C coding region was 99% identical to the corresponding region of Deformed Wing Virus Isolate PA (accession number AY292384.1), the 3C+ coding region was 98% identical, the 3CD coding region was 98% identical, and the 3CD+ coding region was 98% identical. According to an amino acid BLAST search using NCBI, the 3C amino acid sequence was 99% identical to the corresponding region of Polyprotein [Deformed Wing Virus] (accession number AAP49283.1), the 3C+ amino acid sequence was 98% identical (accession number AAP49008.1), the 3CD amino acid sequence was 99% identical (accession number AAP49283.1), and the 3CD+ amino acid sequence was 99% identical (accession number AAP49283.1). The only major amino acid change present within the functional region of the polyprotein coding regions was a change from proline to serine in the 3C region of the 3C+ construct (Appendix A).

After isolation and sequencing, each protein coding region was cloned into an EGFP plasmid, which created a fusion protein consisting of two different proteins connected together, when the vector was translated into protein in a cell. EGFP vectors have two different available orientations; the viral gene can either be inserted so that the
translated viral protein is at the N-terminus of the EGFP protein (pEGFP-N1) or at the C-terminus of the EGFP protein (pEGFP-C1). All four isolated protein coding regions were inserted separately into both vectors, creating eight total plasmids that expressed a viral protein fused to EGFP (Figure 2). Both orientations of the EGFP vectors were utilized in case the EGFP interfered with folding of the viral protein when present at one end.

The success of isolating each construct in both the pEGFP-N1 and pEGFP-C1 vectors was determined by sequentially digesting the plasmids with HindIII and BamHI and electrophoresing the products on an agarose gel (Figure 3). The sizes of the digested fragments were compared to the predicted DNA bp lengths from Lanzi et al. (2006), and the fragments were very close to their predicted sizes (Table 2). Because of the faintness of the insert band, it should be noted that the 3CD+/N1 DNA was probably a mixed population with a high percentage of pEGFP-N1 lacking an insert due to lack of streaking for a single colony before midiprep procedures were performed.

The plasmids were introduced into COS-7 cells via transfection, and a Western blot was performed to verify that the fusion proteins were actually produced in mammalian cells and to determine if the 3C protease actively cleaved itself from the EGFP (Figure 4). Since antibodies against DWV proteins are not available commercially, EGFP antibodies were used to confirm the sizes of the fusion proteins. Although the use of honeybee cells would have been ideal for these experiments, mammalian cells were used because there is no reliable honeybee cell culture system, mammalian cells are easy to grow and to transfect, and they were readily available in the laboratory. The sizes of the fusion proteins were compared to the predicted molecular
weights using the sequence lengths from Lanzi et al. (2006) and the standard conversion that 1000 bp = 37 kDa (New England BioLabs) (Table 3). All of the fusion proteins were close to their expected sizes, though the bands on the Western blot, and thus the amount of protein produced, varied in intensity.

The 3C/N1 sample produced one high intensity band at the predicted molecular weight and a few slightly lower molecular weight bands tapering in intensity. The 3C/C1 sample produced a band with intermediate intensity at the predicted molecular weight, a higher intensity band at a slightly lower molecular weight, and a high intensity band corresponding to EGFP alone. The 3C+/N1 sample produced a band with intermediate intensity at the predicted molecular weight and a light band corresponding to EGFP alone. The 3C+/C1 sample produced a high intensity band at the predicted molecular weight and a few slightly lower molecular weight bands tapering in intensity. The 3CD/N1 sample only produced a very light band at the predicted molecular weight and a light band corresponding to EGFP alone, probably due to low transfection efficiency. The 3CD/C1 sample produced a single high intensity band at the predicted molecular weight. The 3CD+/N1 sample did not produce a band corresponding to the predicted fusion protein and produced a very high intensity band corresponding to EGFP alone, probably due to high contamination by pEGFP-N1 lacking an insert. The 3CD+/C1 sample produced a single band with intermediate intensity at the predicted molecular weight. The pEGFP-N1 control produced a high intensity band corresponding to EGFP alone.
To examine the sub-cellular localization of each of the fusion proteins, the plasmids were transfected into mammalian HeLa cells. EGFP fluoresces green when excited by blue light, so the cells containing the fusion proteins were excited using a confocal microscope with an argon laser at 488nm. The cells were also excited separately by an ultraviolet laser at 405nm to visualize the DAPI DNA stain. In the cells containing EGFP attached to viral proteins (verified through the Western blots), the subcellular localization of the 3C, 3C+, 3CD, and 3CD+ fusion proteins was determined by observing where in the cell the EGFP fluoresced (Figure 5 and Figure 6; Appendix B). All of the DWV/GFP fusion proteins localized in the cytoplasm, with either diffuse or aggregated localization or a combination of both.

The 3C fusion proteins showed more strongly aggregated localization than the 3CD proteins, especially with the EGFP-C1 orientation of the plasmids. The 3C/N1 fusion protein showed diffuse cytoplasmic localization with a few areas of aggregation just outside of the nucleus. The 3C/C1 fusion protein localized as aggregates in the cytoplasm, most likely on cytoplasmic membranes. The 3C+/N1 fusion protein showed diffuse cytoplasmic localization. The 3C+/C1 fusion protein localized strongly in aggregates just outside of the nucleus and throughout the cytoplasm.

The 3CD fusion proteins showed more diffuse cytoplasmic localization than the 3C proteins. The 3CD/N1 fusion protein did not demonstrate strong enough fluorescence for imaging. The 3CD/C1 fusion protein showed diffuse cytoplasmic localization. The 3CD+/N1 fusion protein was not produced in the transfected cells; the image from the 3CD+/N1 sample, which resembled the control, showed diffuse cytoplasmic and nuclear
localization because it consisted of EGFP-N1 alone due to contamination. The 3CD+/C1 fusion protein showed diffuse cytoplasmic localization with numerous small areas of strong aggregation throughout the cytoplasm. The EGFP-N1 control showed both diffuse cytoplasmic and nuclear localization.
Figure 2  Orientations of eight total DWV/GFP fusion proteins generated by inserting DWV constructs individually into both pEGFP-N1 (3CD+/N1, 3CD/N1, 3C+/N1, and 3C/N1) and pEGFP-C1 (3CD+/C1, 3CD/C1, 3C+/C1, and 3C/C1) vectors.
Figure 3  Agarose gel electrophoresis of DWV/GFP fusion proteins sequentially digested with *Hind*III and *Bam*HI. The digests were run on a 1% agarose gel with ethidium bromide at 75V.
Table 2  Predicted size of insert fragments from Lanzi et al. (2006) compared to observed size calculated by using the Band Analysis function of Kodak 1-D software.

<table>
<thead>
<tr>
<th>Lane on Agarose Gel</th>
<th>Insert Fragment</th>
<th>Predicted DNA size (bp)</th>
<th>Observed DNA size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>None (EGFP-N1 alone)</td>
<td>4700</td>
<td>4333</td>
</tr>
<tr>
<td>3</td>
<td>3C</td>
<td>463</td>
<td>519</td>
</tr>
<tr>
<td>4</td>
<td>3C</td>
<td>463</td>
<td>539</td>
</tr>
<tr>
<td>5</td>
<td>3C+</td>
<td>927</td>
<td>966</td>
</tr>
<tr>
<td>6</td>
<td>3C+</td>
<td>927</td>
<td>955</td>
</tr>
<tr>
<td>7</td>
<td>3CD</td>
<td>1965</td>
<td>1932</td>
</tr>
<tr>
<td>8</td>
<td>3CD</td>
<td>1965</td>
<td>1932</td>
</tr>
<tr>
<td>9</td>
<td>3CD+</td>
<td>2153</td>
<td>2147</td>
</tr>
<tr>
<td>10</td>
<td>3CD+</td>
<td>2153</td>
<td>2088</td>
</tr>
</tbody>
</table>
Figure 4  Western blot of DWV/GFP fusion proteins harvested from transfected mammalian COS-7 cells. Proteins were visualized using a monoclonal anti-GFP antibody (JL-8 from Clontech). Black arrow heads show the full length fusion proteins.
Table 3  Predicted molecular weight of fusion proteins from Lanzi et al. (2006) compared to observed molecular weight calculated by using the Band Analysis function of Kodak 1-D software.

<table>
<thead>
<tr>
<th>Lane on Western Blot</th>
<th>Fusion Protein</th>
<th>Predicted Molecular Weight (kDa)</th>
<th>Observed Molecular Weight (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3C/N1</td>
<td>44.0</td>
<td>48.6</td>
</tr>
<tr>
<td>3</td>
<td>3C/C1</td>
<td>44.0</td>
<td>46.9</td>
</tr>
<tr>
<td>4</td>
<td>3C+/N1</td>
<td>61.2</td>
<td>60.3</td>
</tr>
<tr>
<td>5</td>
<td>3C+/C1</td>
<td>61.2</td>
<td>62.1</td>
</tr>
<tr>
<td>6</td>
<td>3CD/N1</td>
<td>99.6</td>
<td>88.3</td>
</tr>
<tr>
<td>7</td>
<td>3CD/C1</td>
<td>99.6</td>
<td>88.3</td>
</tr>
<tr>
<td>8</td>
<td>3CD+/N1</td>
<td>106.6</td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>3CD+/C1</td>
<td>106.6</td>
<td>96.1</td>
</tr>
<tr>
<td>10</td>
<td>None (GFP alone)</td>
<td>27</td>
<td>25.2</td>
</tr>
</tbody>
</table>
Figure 5  Subcellular localization of DWV/GFP fusion proteins in HeLa cells. The green fluorescence in the first column is GFP, the blue false color in the second column is DAPI, and the third column contains a merged image of the previous two columns. A, B, and C are 3C/N1; D, E, and F are 3C/C1; G, H, and I are 3C+/N1; J, K, and L are 3C+/C1.
Figure 6  Subcellular localization of DWV/GFP fusion proteins in HeLa cells (continued). The green fluorescence in the first column is GFP, the blue false color in the second column is DAPI, and the third column contains a merged image of the previous two columns. A, B, and C are 3CD/C1; D, E, and F are GFP-N1 contamination in the 3CD+/N1 sample; G, H, and I are 3CD+/C1; J, K, and L are EGFP-N1 without any insert. The 3CD/N1 sample did not demonstrate strong enough fluorescence for imaging.
Discussion

The aim of this project was to further understand the molecular characterization of an important honeybee virus, DWV. In general, critical information can be learned from studying the molecular components of a virus which can lead to a better understanding of its infectivity and pathogenesis. Studying viruses at the molecular level builds a foundation for understanding the virus as a whole and the processes that they use to infect cells. Therefore, studying individual proteins produced by DWV will provide a better understanding of its infection processes.

The 3C and 3CD proteins were chosen for this study because they perform critical functions involved in viral replication (Lanzi et al. 2006). Each construct containing the nested coding regions was successfully inserted into both pEGFP-N1 and pEGFP-C1 vectors, though the 3CD+/N1 insert band was very faint compared to the other fragment bands when the plasmids were sequentially digested and electrophoresed (Figure 3). This faintness is most likely due to contamination by a large population of pEGFP-N1 vector that does not contain any insert. This type of mixed population contamination could be prevented in the future by streaking for a single bacterial colony before growing up the transformed bacteria for midiprep plasmid DNA isolation. There are also a few faint bands present on the agarose gel in the lanes with 3C/N1, 3C/C1, and 3C+/N1 which all appear to be ~600 bp in length (Figure 3). Since these bands are extremely faint and are present in both the 3C and 3C+ populations, they are most likely artifacts from the midiprep isolation, such as denatured plasmid DNA.
The large percentage of pEGFP-N1 in the 3CD+/N1 sample explains the lack of a high molecular weight band on the Western blot of 3CD+/N1 (Figure 4). Only GFP itself appeared in the 3CD+/N1 sample, probably due to a lack of transfection by the actual 3CD+/N1 plasmid because it composed such a small percentage of the total sample. The 3CD+/N1 sample was also the only experimental sample to have all of its cells resemble the EGFP-N1 control under the confocal microscope, with both cytoplasmic and nuclear diffuse GFP observed throughout the cell (Figure 6F). This discrepancy was probably again due to the extremely low concentration of 3CD+/N1 actually present in a sample highly contaminated with EGFP-N1 plasmids. Also, it should be noted that the cells transfected with 3CD/N1 did not produce enough protein to capture an image, probably due to the extremely low transfection efficiency of that particular plasmid as can be seen by its faint bands on the Western blot.

One potential issue that had to be addressed during this study was that since 3C is an autocatalytic protease, it could possibly cleave itself from its EGFP tag after it was translated into protein inside of the transfected cells. If 3C cleaved itself from the EGFP tag, then EGFP would show up on the Western blot by itself because it would no longer be attached to any other protein. This autocatalytic protease capability of 3C was observed on the Western blot, though the appearance of GFP alone in a given sample could be due to contamination by GFP plasmids lacking inserts as well. Since most of the proteins did not have a band of GFP in one orientation but had a GFP band in the other orientation, it is likely that proteolytic cleavage is the cause of this phenomenon. The bright bands on the agarose gel after double digestion in most samples support this
hypothesis as well. It is interesting to note that proteolytic activity was observed in all of the constructs when the viral protein was at the N terminus (EGFP-N1 fusion). In the larger fusion proteins that did not undergo cleavage (EGFP-C1 fusions), the 3C portion of the protein was in the middle of the construct, stuck between EGFP and 3D. One potential reason for this observed pattern could be that the enzymatic activity of 3C is physically blocked when it is constrained by 3D and EGFP instead of another viral protein on each side. It is also possible that the sequence connecting EGFP to 3C in the EGFP-C1 orientation does not contain a typical 3C protease cleavage site.

In both the 3C+/N1 and 3C/N1 samples, a few proteins slightly smaller than the experimental proteins were identified by the EGFP antibody as well. This could possibly be due to partial degradation of either the viral protein or GFP by cellular proteases. The smaller molecular weight band of EGFP present in most lanes (~18 kDa) is probably due to cellular protease activity as well (Dr. Matthew Heintzelman, personal communication). In the future, to prevent viral proteolysis, the cleavage sites of the 3CD protein could be mutated as was done by Sharma et al. (2004) for poliovirus. A comparison of the proteins produced by the plasmids with a wild-type or mutated cleavage site would give more definitive support for either viral or cellular cleavage of the proteins to produce multiple bands on the Western blot.

It was critical to use both orientations of the EGFP vector in generating clones for this project in case EGFP at one end of the construct would interfere with viral protein folding or localization. Since some of the fusion proteins had different cytoplasmic localizations depending on the orientation of the EGFP vector that was used, one
orientation of the vector most likely gives a more accurate representation of their localizations than the other. For example, the 3C+ protein was very diffuse cytoplasmic when fused to EGFP-N1, yet it was extremely localized in aggregates, potentially on cytoplasmic membranes, when fused to EGFP-C1 (Figure 5). Though none of the fusion proteins localized in the nucleus at all, the contradiction between diffuse cytoplasmic and aggregated cytoplasmic localization cannot currently be resolved. Generally, the 3C proteins (Figure 5) showed more intense aggregation than the 3CD proteins, which were mainly diffuse cytoplasmic (Figure 6). Once live virus is available in culture, and once antibodies against DWV proteins are produced, the 3C and 3CD proteins could be tracked using antibody staining during infection. This method, though not currently available, would determine which orientation of the GFP vector most accurately depicts the actual localization of 3C and 3CD during DWV infection. If aggregated cytoplasmic is the correct localization, antibodies against specific cellular membranes, such as Golgi apparatus and endoplasmic reticulum, could be utilized to determine at which particular membranes the DWV proteins are aggregating.

DWV is a particularly difficult virus to work on and has not been successfully grown in culture yet. Though this project determined the subcellular localization of DWV 3C and 3CD by themselves, which was never done before, an ideal experiment would include localization in combination with superinfection by live virus as well, just as Sharma et al. (2004) showed for poliovirus. It is likely that either 3C or 3CD (or both) would localize to the nucleus upon live infection because picornavirus 3C proteins have been shown to degrade cellular transcription factors, such as the TATA-binding protein,
which reside in the nucleus of host cells (Clark et al. 1993). Sharma et al. (2004) found a putative NLS in the poliovirus 3CD amino acid sequence that was consistent with a relatively well-conserved domain across all picornaviruses that has the potential to act as an NLS, “KKRDI.” When the potential NLS was mutated, the 3CD protein did not localize in the nucleus even with superinfection.

In scanning the amino acid sequence of the DWV 3CD protein, a potential NLS was discovered during this study. The 3D protein contains an amino acid motif consistent with the conserved picornavirus “KKRDI” domain: RKKGI (Lanzi et al. 2006). Since K and R are both positively-charged amino acids and D and G both contain amine groups in their side chains, this sequence is similar enough to potentially function as the DWV 3CD NLS. If transfected cells were superinfected with live DWV virus, an accessory protein could potentially assist in delivering 3CD to the nucleus for entry.

Though this project only examined two particular proteins created by the DWV genome, its broader implications are related to the devastating effect that DWV has on honeybees. To understand the functioning of any virus, it is critical to understand the molecular foundations of its component parts. Since DWV has been shown to be prevalent in honeybee colonies with CCD, the virus may have a devastatingly negative impact on honeybee populations in the United States (Johnson et al. 2009).

Agricultural crops have been shown to have increased quality and quantity due to pollination by honeybees (Sabara and Winston 2003, Sabbahi et al. 2005). For example, a forty-six percent increase in seed yield of canola plants was observed in the presence of honeybees compared to the absence of the bees (Sabbahi et al. 2005). Honeybees are
economically valuable for use as pollinators, and one honeybee colony can even be used to pollinate two or three specific crops at the same time (Morse and Calderone 2000). Since honeybees are crucial for pollination, the trickle-down effects a DWV infection could be devastating, both to the honeybee colony itself and to farmers and consumers as part of the whole agricultural ecosystem. Since little previous research has been done on the molecular characterization and localization of DWV proteins, this project is crucial to start building a foundation for understanding the molecular workings and infectivity of DWV.
Bibliography


Appendix A: Sequences of TA clone inserts and NCBI amino acid blast analyses

Sequence of 3C TA clone insert using the T7 primer (restriction digest sites in bold)

AAGCCTTCCGCCATGGGAACCAGTAGCTATATTTAAATATTTCCATAATCAAGAGACTAGAATGTTCTGCTCTTGATATTTCTGGTATTG
AAATTGATTTGTTGAATTTACCTAGATTGTATTATGGTGGTCTCGCGGGAGAGGAGTCGTTTGATAGTAATACCGTGCTTGTGACTATGCCTAATCGTATTCCTGAGTGTAAGAGCATTATTAAATTATAGCGTCACATAATGAACATATACGTGCTCAGAATGATGGAGTGTTAGCGTACTGGCGACCATACTCAGCTATTGGCTTTCGAGAATAATAATAAAACTCCGATAGTATTAACGCTGATGGTTTGTA
TGAGGTTATACTTCAAGGAGTATATACC
TATCCATACCACGGCGATGTTGTGGGATCTCAGAAATGGAGGTGGTATTTGGTTCGGAATTTCACAACGCGCAATTATAGGTATCCATGTTGCTGACTGCAAGGATCC
GGATCTTCTCTTCAGTACCAGCAACATGGATACCTATAATTGGCCGTTGTA
AATTCCGAGACAACATAATTCCGAAACCAACACACCACATCGCCTGGTATGGATA
GGTATATACTCCTTGAAGTATAACCTATACACAAACCCATACGCACGTTAATTACCTTA
TCGGAGTTTTATTATTATTCTCGAAGGCAAATAGCTGAGTATAGGTCGCCAGTT
ACTAACACCCATCATCATTCCAGAGCAGTATATGTCTTATTATGTGACGCTATAAA
TTAAATAATGCTTACACTCAGGAATACGATTAGGCATAGTCACAAGCAGG
GTATTACTATCAAAACAGACTCCTTCCGGCAGACCCCATAGTACATACATCAG
GTAAATTCACAAACTCAATTTTCAATACCAGAAATATTCACCAGACATTCTAGTCTCTGATTATGAAATATACCCATTTGCTGGCGGAGAAGCTT
**Sequence of 3C+ TA clone insert using the T7 primer (restriction digest sites in bold)**

GGATCCCCGAGATATCGAATCACAACCCCATCTAATCCAGGCACACCACATACA
GCATCTGGCAAACCTCCAATCTTGGCAAACTTTATTTGTTAATACTACTGAAAC
TAATTTTTCTTTTCAATAGGATTTTGTGCTAATCTCGAGATTTTCCTATTAAACGG
TGAACAAGGCATGCCATGCTTTTCACACCCATACTTAAAGGATCATGCGGC
GCTATTCTGGATAGCCATTGTGCATTGTTGTGCATGTTCTCTCATACATCAA
TGTTCCATGGAATAAAGGCTCTTTTAAATCCCAAGTGAAGGCTTTGAGCATGAG
CTAGCTTTGCATCCTCCTACTTACCAATCGGATATAAATCGGTATCTAATTTG
GCATTGTGTTAAAATTCGCGAACAACATACCAGAACCCACCATCGCC
TGTTTCACTCTAAATAGTATTTAAAGCAGTTATGATAGTAGTAT
GGTCGCCCACTGATCTGACATCGGGATGTGGCTCTCCTACATCAAA
GAGCTTTGGTAAATTTTTCGGAGACAACAAATATCGCAAAACCAACACATCGCC
TGATAGTTATATAACTTTCTTAAAGTATACATCAAACATCCATCGCC
GGATCC

**Sequence of 3C+ TA clone insert using the SP6 primer (restriction digest sites in bold)**

AAGCTTCCGCCATGGGAACCAAGTACTATTATATATCTATACATCTAATCAAGA
GCTAGAATGCTCTGGTATATTTCCTGGTAGATTTGGAATTTAC
CTAGATTGTATTATGGTTGCTCGCAGGAGAGGTCTTTGGAATTATATC
GCTGTTGTGACATGTCTACTCATGCATCGTCTAGTGAAGATATATC
TAGAGCTGCACTATAGCTGCTAGTTGCTAGATAGTAGTAT
GCTGCGGCCATCTGACTCGTCTAGTTGCTAAGATGAGATAT
ACTGCCGCCATCAGTTGCTAGTTGCTAGATAGTAGTAT
GCTAGATTAACGCTGATGTTGTGATAGATATTCTTAAAGATATACC
TATCCACACCGCGCAGTGTTTTTGTTGTTGATATTTGATGTCCGAATT
AACAAGGCCCAATTATCTATCCATCTGTTGCTAGTGATAGGAATTGCAGGC
TTGGAGATTGCTAAAACACTTGTATAGTAGATTTCCACTGATTAAGGACAT
GATATGAACAGAAGCAGTACGATTGTTGTATAGACATCGTCTGGAACATTA
GATGAAATCTGATATAAGGATTATATCCTAATCGGATTTGGAAGGTTAAG
TGCAGAAGTCAGCTACCAACAGCCCTTCCACTGGAATTAAAAAGACGGCT
ATCCATGGGAAAATTTGATGTTGAACTGAAACATCCATCGTATCGTACATG
ATCCAAAGATAAGCCCGCATGATCTTTTCAAGGTTAGGTTGGAAGGAACATG
CATGCTCTGGTCTCAGTTAAATAGGAAACATCTGGGATTAGCAAAAATTAC
TGAAGAAGAAAAATTGTCTAATAGGATATAAAAAACAAATATGTTGAAAGG
AGTTTGGCAAGATGCTGTATGTGCTGTGCTGATGAGATTTGGAATTCGA
TATCTCGGAGATCC
Sequence of 3CD TA clone insert using the T7 primer (restriction digest sites in bold)

AAGCTTCCGCCATGGGAACCAAGTACTATTTTGAATGATATTCATAATGCAAGA
GACTAGAATGTCTGGTGTATATTCTGTTGATGGAAATGGTTTGGTTTAC
CTAGATTGTATATTGTTGGTCTCAGGCGGGAGAGGAGTGCTTGGATAGTAATATC
ATGCTTGTGACATATGCTAATCGTATTTCTGAGTGTAAGAGCATTATAATAGTT
TATAGCGTACATAAATAGGATATACGTACGTCGACAGATGAGTGGATTAGTA
ACTGGGCAACATACTAGCTATTGGCTGCCAGGATAAAATAAACCTCGA
TAAGATATTAACGCTGATGGTTTGTATGAGGTATTACTTCAAGGAGGTATATACC
TACCCATAACCAGGCGATGTGTTTGCAGTATTTGTTGTTCGGAATT
ACAACGGCAGATTATAGGTATACCATGTGTTTGGTACTGAGATGTTCACTGGA
TTTGAGAATGGCTGATCACCATTGTCAGATGTTGCTAAAGACGTATTGAGG
AGATGAAGAGGGGCGTACGATGTTAAGCTGTTGCTAGATTCC
TGAGTGAAAGAGACGTGACGGTCAGGGTTGATGTTAGTAGTTAATT
TATAGCGTCACAGTCAAGTTTCAGAGGA

Sequence of 3CD TA clone insert using the SP6 primer (restriction digest sites in bold)

GGATCCATAGCAGATGCTCCAAATTCGTTCTCTTCTACTACCGGAGACCCTTGCCAG
GTTAGCTGAAACACAGGTCTAGTGGATGTATAAGAAACCCATGGTTTAAAGA
AGATGACGTGCTTAGAAGCTGTCGACCTTTCACTAGTATTCTCTGATTTTATCC
TACGTAAATTTATGTTTATGTACGTGACAGTGCTCAGGTTGAAAACATGGGA
GAAAATCAGATATGCTAGTTCATATTCTCGGCATGTTCCACATTACGTCG
TACGATCTGTCTCTTCTACGCTACGTCGACCTTTCACTTTCTTGCAGGC
GAGCAGACATGGTGTTAATCTGATATTCTGATTTCAATCAGGATATAC
TACCCAAAGCTAATATTTCAACAATTTGGAATATGATTCAATATATCC
GTTATGAGGAACCTCGAATTCCGCGAAGGTACTCGGTGCAACACAGATCC
GATACGAGACTAGCGCTCTAAATCTTCTTGCAGGAATGGTTACGTTGCA
TTATTTTGCCTTTTACTCTCTTCTTGGTATAATGTAATACCAATCCGAATA
ATTTCGGAAGGCTAGGAAAGCTCCAGAACATGCGAATTCCAAACGGACAGA
TTTATGCTCAGGATAGGATGCAGCTACTATTTAGCAACCTCGGTTGCA
TTTGTCCACCTACTAAGCTGTTAATGCTACAAATACATTTCTGCTAGCATTAG
TCCCGCAAGCTCAGATGCTAGACGCTGTTACGCTACGTTGGAACGGT
ATGGTAAAAGACTGTCGGGAGCTTATACTAAATATATCAGTTTGGGATTCACTT
ACATTTTTC
Sequence of 3CD+ TA clone insert using the T7 primer (restriction digest sites in bold)

```
AAGCTTCCGCCATGGGAACCAAGTACTATTTTTAAGTATATTCATAATCAAGA
GACTGAAATGTCTGGTGATATTCTCTGTATTGAAAATGTATTTGGTTGAATTAC
CTAGATTTGATATTAGTGATTCTCAGCGGAGAGGAGCTGTTTGGATAGATAATATC
GTGCTTTGTGACTATGCTAATCTGATATCTCTGAGTGCAAGAGCATTTATTTAAATT
TATAGCGTCATACATGAAATGACATATACGTGCTCAGAATGATGGAGTGTTAGTA
ACAAACGGCAATTATAGGTATACCATGGCTTGTCTCAGTACTTTGGAGATTGCATGGA
TTTGAGATGCTGAACTTCCGTTGCGTGAATTA
GATGAATCTGATATTGGTTTAGATACCGATTTATATCCGATTGGTAGGAGTA
TGCAAAGGCTGCTGATGCTGCTCAAGGACCTTCTACTGGGATATAAAAGAGCTG
ATCCATGGAACATTTGATGTAAGGACTGAACCAAATCCGATGTCATCACGTG
ATCCATGAAATAGCGCCGCATGATCCTTTGAAGTTAGGGTGTGAAAAGCATGG
CATGCCCTTGTCACCCGTTTAATAGGAACATCTGGGAATTAGCAACAAATACATT
TGAAAGAAAAATTAGTGGTCAGTTAATGGAATACCCCAATAAATGTTGCAAGATTGG
AG
```
NCBI amino acid BLAST of partial 3C insert

```
>gb|ACY38662.1| polyprotein [Deformed wing virus]
Length=2993
Score = 306 bits (785), Expect = 4e-82
Identities = 149/150 (99%), Positives = 149/150 (99%), Gaps = 0/150 (0%)
Frame = +3
Query  9  GTKYFKYIHQQTEFMGDSISIEIDLLNLPLYGLGLGEEESDFRNVTLMNPRIPEC
       188
       GGTKYFKYIHQQTEFMGDSISIEIDLLNLPLYGLGLGEEESDFRNVTLMNPRIPEC
Sbjct 2181 GGTKYFKYIHQQTEFMGDSISIEIDLLNLPLYGLGLGEEESDFRNVTLMNPRIPEC
       2240
Query 189 KSIKFIASHNEHRARQNQDVGVTGHDQTLLAFENHNTKPI3INADDGLYEYILQGVTYP
       368
       KSIKFIASHNEHRARQNQDVGVTGHDQTLLAFENHNTKPI3INADDGLYEYILQGVTYP
Sbjct 2241 KSIKFIASHNEHRARQNQDVGVTGHDQTLLAFENHNTKPI3INADDGLYEYILQGVTYP
       2300
Query 369 YHSDDGVCSSILLSLRLQRPFIIGHVAGTEG 458
       YHSDDGVCSSILLSLRLQRPFIIGHVAGTEG
Sbjct 2301 YHSDDGVCSSILLSLRLQRPFIIGHVAGTEG 2330
```

NCBI amino acid BLAST of partial 3C+ insert

```
>gb|AAP49008.1| polyprotein [Deformed wing virus]
Length=1566
Score = 632 bits (1629), Expect = 2e-179
Identities = 308/312 (98%), Positives = 309/312 (99%), Gaps = 0/312 (0%)
Frame = +3
Query  9  GIKYFKYIHQQTEFMGDSISIEIDLLNLPLYGLGLGEEESDFRNVTLMNPRIPEC
       188
       GIKYFKYIHQQTEFMGDSISIEIDLLNLPLYGLGLGEEESDFRNVTLMNPRIPEC
Sbjct  866 GIKYFKYIHQQTEFMGDSISIEIDLLNLPLYGLGLGEEESDFRNVTLMNPRIPEC
       925
Query 189 KSIKFIASHNEHRARQNQDVGVTGHDQTLLAFENHNTKPI3INADDGLYEYILQGVTYP
       368
       KSIKFIASHNEHRARQNQDVGVTGHDQTLLAFENHNTKPI3INADDGLYEYILQGVTYP
Sbjct  926 KSIKFIASHNEHRARQNQDVGVTGHDQTLLAFENHNTKPI3INADDGLYEYILQGVTYP
       985
Query 369 YHSDDGVCSSILLSLRLQRPFIIGHVAGTEG 548
       YHSDDGVCSSILLSLRLQRPFIIGHVAGTEG
Sbjct  986 YHSDDGVCSSILLSLRLQRPFIIGHVAGTEG 1045
Query 549 YELPLELDEALTIDLTDLPIGRVDKALHADQSPSTGIISKIKLHTLHGFVDRTEPHMSRS
       728
       YELPLELDEALTIDLTDLPIGRVDKALHADQSPSTGIISKIKLHTLHGFVDRTEPHMSRS
Sbjct 1046 YELPLELDEALTIDLTDLPIGRVDKALHADQSPSTGIISKIKLHTLHGFVDRTEPHMSRS
       1105
Query 729 DPRPAPHDPLGCEKHMPCSFERNKHELATNLHKEKLVSVVKPGCKGGLQDAC
       908
       DPRPAPHDPLGCEKHMPCSFERNKHELATNLHKEKLVSVVKPGCKGGLQDAC
Sbjct 1106 DPRPAPHDPLGCEKHMPCSFERNKHELATNLHKEKLVSVVKPGCKGGLQDAC 1165
Query 909 GVPGLDGFDSIS 944
       GVPGLDGFDSIS
Sbjct 1166 GVPGLDGFDSIS 1177
```
NCBI amino acid BLAST of partial 3CD insert
NCBI amino acid BLAST of partial 3CD+ insert
Appendix B: Original GFP/DAPI merged confocal images

*EGFP-N1 (no insert)*
Lipofectamine 2000 (no plasmid transfection)
$3C^+/N1$
$3C^+/C1$
3CD/C1
3CD+/N1
3CD+/C1