Detection of the SIRE1 Env-Like Protein, a Potential Retroviral Protein in Soybean

Garen Gaston

Loyola University Chicago

Recommended Citation
https://ecommons.luc.edu/luc_theses/564

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.
Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 2011 Garen Gaston
LOYOLA UNIVERSITY CHICAGO

DETECTION OF THE SIRE1 ENV-LIKE PROTEIN, A POTENTIAL RETROVIRAL PROTEIN IN SOYBEAN

A THESIS SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
MASTER OF SCIENCE

PROGRAM IN BIOLOGY

BY
GAREN GASTON
CHICAGO, ILLINOIS
AUGUST 2011
ACKNOWLEDGEMENTS

I would like to thank Eftychia Gouvas for helping initially orient me in the lab and being a constant source of helpful information, Dr. Kim Williamson for introducing me to Western blotting technique and suggesting the solution to one of my first major obstacles, Dr. Rochlin for lending me the necessary start-up materials for the project, and Dr. Castignetti, for opening his lab as the eternal chemical supply room. I would also like to thank Thomas Carter at the USDA and the University of Illinois/USDA germplasm collection for seed supplies necessary for the research.

I would like to thank Lauren Mogil and the other undergrads of our lab and the Doering lab who lent moral support through diversion, to the office staff for helping shield me from much of the university bureaucracy, and to Dr. Doering, personally, for letting us move into a lab with sunshine and a view.

In addition, I would like to thank my committee, Drs. Eric Schroeter, Stefan Kanzok, and F. Brian Pickett for their advice on my presentation and writing style and for helping widen my perspective in the myopic state of writing the thesis. And of course, Dr. Laten who has constantly steered me from blind paths, given me constant encouragement and has had to suffer not only through the initial drafts of this thesis, but also two extra unplanned years of me in the lab. I hope the work proves worth it.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER ONE: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER TWO: MATERIALS AND METHODS</td>
<td>25</td>
</tr>
<tr>
<td>CHAPTER THREE: RESULTS</td>
<td>46</td>
</tr>
<tr>
<td>CHAPTER FOUR: DISCUSSION</td>
<td>88</td>
</tr>
<tr>
<td>REFERENCE LIST</td>
<td>110</td>
</tr>
<tr>
<td>VITA</td>
<td>119</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Fig. 1 Genomic organization of retroelements
Fig. 2 Mechanisms of retroelement influence on gene transcription
Fig. 3 Genomic organization of retroelements and expression strategies
Fig. 4 Lifecycle of retroelements with and without env proteins
Fig. 5 Phylogenetic organization of the Sireviruses and Ty1/copia elements
Fig. 6 Comparison of Retroviral and SIRE1 envelope and env-like proteins
Fig. 7 Location of amplified regions on the env-like protein
Fig. 8 IBA Stargate two-vector system
Fig. 9 Vector map and partial sequence for pEntry-IBA51 (donor vector)
Fig. 10 Vector map and partial sequence for pASG-IBA53 (destination vector)
Fig. 11 Size verification of pre-insertion amplicons and plasmid inserts
Fig. 12 Initial induction for the expression of fusion peptides
Fig. 13 Characterization of peptides after His-tag Purification
Fig. 14 Characterization of peptides after tandem His-tag/Strep-tag Purifications
Fig. 15 Purification of peptide 21 by imidazole Gradient
Fig. 16 Re-purification of peptide 19 by imidazole Gradient
Fig. 17  Estimates of peptide purity preceding antibody production

Fig. 18  Verification of specificity of crude antisera

Fig. 19  Pre-immune antisera tests on plant material

Fig. 20  Soybean stress treatments probed with crude antisera

Fig. 21  Crude antisera tests: Day 58 vs. Day 35, Rabbit 1 vs. 2

Fig. 22  Test of purified anti-\textbf{p19} Ab without extensive washing

Fig. 23  Soybean treatments probed with purified anti-\textbf{p19} Ab

Fig. 24  Embryo and young plant tissues probed with purified anti-\textbf{p19} Ab
LIST OF TABLES

Table 1. Examples of Stressed induced LTR-retrotransposons

Table 2. Comparison of 60-residue proline-rich regions from *SIRE-1* and mammalian retroviruses, FeLV, feline leukemia virus; MLV, murine leukemia virus

Table 3. Molecular Weights of Proteins Detected in Tissue Culture and Embryo tissues with purified anti-19 Ab
ABSTRACT

Long terminal repeat (LTR) retroelements are mobile DNAs that include retroviruses, endogenous retroviruses and retrotransposons. Though retrotransposons, which are noninfectious, can make up as much as 80% of plant genomes, retroviral elements have never been confirmed in plants. However, the recently active, highly conserved soybean retrotransposon SIRE1 encodes a putative protein that is predicted to have features of a retroviral envelope protein, which could sponsor infectivity. Thus plants might host endogenous retroviruses and possibly still-active retroviruses. This project demonstrates that SIRE1 retrotransposons may be active and producing the env-like protein. Polyclonal antibodies generated against cloned translation products of the SIRE1 env-like gene were used to directly detect the protein in soybean tissue under conditions known to induce the expression of related plant elements. At least one 65 kDa protein was identified as a likely candidate for the SIRE1 env-like protein. Though the identity must be confirmed, the results provide preliminary evidence of SIRE1 env-like protein production, suggesting that SIRE1 may be an active endogenous retrovirus and possibly an infectious plant retrovirus.
CHAPTER ONE
INTRODUCTION

Retrotransposons, retroviruses and endogenous retroviruses are three subdivisions of a class of mobile DNA elements called retroelements. These viral and virus-like agents are flanked by identical copies of a long terminal repeat (LTR) [Fig. 1] and have the unique property of integrating into the chromosomes of the cells that host them. Retroelement genomes encode all of the necessary proteins required for their transposition and replication. Among other proteins, these include a functional reverse transcriptase (rt), which sponsors replication through an RNA intermediate, an integrase (int), which splices the newly formed cDNA into the host genome.

Fig. 1 Genomic organization of retroelements

Comparison of retroelements shows similar genomic organization. However both SIRE1 and retroviruses encode an extra ORF, which is not found in traditional retrotransposons.
Retroelements are ubiquitous across eukaryotic genomes, comprising as much as 80% of the entire genome of some plant genomes (Kumar and Bennetzen 1999). Most eukaryotic species host many different retroelement families, each present in dozens to thousands of copies (Eickbush and Malik, 2002). Because there are thousands of copies, many of which are peppered with disabling mutations, they are considered to be part of the repetitive DNA of the host genome, and have historically been referred to as “junk” DNA. However, active roles in many facets of genome evolution and function are becoming clearer.

No verified traces of retroviral elements have been found in plants though hundreds of thousands of retrotransposon copies can be found in a single plant genome and retroviruses might have risen multiple times in animal genomes (Malik et al. 2000). However, research over the last 20 years has identified putative proteins encoded by some plant retrotransposons that might sponsor infectivity. Evidence of their function and detection of active expression will expand the known range of potentially infectious retroviruses to plants.

**Importance of Retroelements In Genomes**

Retrotransposons and transposable elements in general were often characterized twenty to thirty years ago as “junk” DNA since much of this DNA is fragmented. Even functional copies do not encode proteins required by the host. In the more recent past however, focusing on functional copies, researchers coined the term “parasitic DNA” since the retroelements utilize host resources but have no apparent benefit. However,
more recent research has begun to recognize functionality in host genomes and their significance in speciation and the evolution of the genome.

Restructuring of the Chromosomal Landscape

LTR retroelements constitute the majority of the genetic material of many plants as well as 8-10% of mouse and human genomes (Lander et al., 2001; Waterston et al., 2002). In the case of maize, as much as 80% of the genome is composed of retroelement copies (Kumar and Bennetzen 1999). Rapid expansion of the retroelements can be responsible for gross genomic changes and have played a role in the evolution of many plant species (Ungerer, 2006; Vitte & Panaud, 2005; reviewed by Bennetzen, 2005). Retrotransposons in maize, for example, are hypothesized to be primarily responsible for the doubling in genome size to 2400 Mb in the last 6my (San Miguel, 1998). Retroelements have also been implicated in chromosomal instability and rearrangements (reviewed by Mieczkowski et al., 2006) which might promote speciation through post-zygotic reproductive isolation (Ladizinsky, 1998; Grieg, 2009; Baird, 2009).

Role in Disease

In animals, retroelements are commonly associated with disease (reviewed by Jern and Coffin, 2008). Retroviruses cause a range of diseases, such as immunodeficiencies and cancers in humans and other animals. Endogenous retroviruses have been implicated in animal cancers, and insertions of endogenous retroviruses into genes have also been shown to be the cause of widely studied mouse phenotypes such as *hairless*, *dilute*
(coloration), and *agouti*, which result among other things in, obesity, diabetes, and increased susceptibility to tumors.

Positive Genomic Change and Gene Regulation

Despite their deleterious effects, retroelements can effect positive change in a genome and might even be required as regulators of heterochromatic structural organization and as promoters of alternate temporal and spatial gene expression at the level of transcription [Fig. 2] (Reviewed in Grandbastien, 2010; Shapiro, 2005; Gogvadze and Buzdin, 2009). Retroelement LTRs contain regulatory elements. Much of the effect of the alternate regulatory patterns caused by retroelements is caused by the placement of their LTRs in and around genes. A retrotransposon insertion in *Drosophila* has been shown to promote insecticide resistance by increasing expression of a nearby gene.

Fig. 2 Mechanisms of retroelement influence on gene transcription

Red boxes retroelements, green boxes gene exons, green arrow gene transcriptional start site, purple oval enhancer element (adapted from Gogvadze and Buzdin, 2009)
involved in detoxification (Chung, 2007). On a larger scale, regulatory sequences in LTRs from related elements can act as enhancers and alternate promoters for a network of genes. These regulatory sequences from different families can be simultaneously triggered at different developmental stages for coordinated expression of neighboring genes (Peaston, 2004; reviewed in Shapiro, 2005). LTRs inserted in the inverted orientation can repress expression of neighboring genes by producing antisense RNAs. Over 48,000 antisense start sites described in the human genome fall within transposable elements (Conley, 2008).

**Activation of Retroelements**

Expression and mobilization of retroelements are kept under tight control by host regulatory systems. Both transcriptional and post-transcriptional mechanisms have been described that repress retrotransposition (Hirochika, 2000; Garfinkel, 2003; Matsuda and Garfinkel, 2009; Lim, 2009). Thus elements can be transcribed without being transpositionally active. In plants, DNA methylation of retroelements inhibits transcription and seems to be the predominant mechanism of silencing (Hirochika, 2000; Tsukahara, 2009), though there is some evidence that, in some instances, plants might employ post-transcriptional controls as well (see below under Methodology).

Transcriptional repression can be loosened, however. Tissue-specific transcripts have been routinely detected during different key developmental stages (Vicient, 2010; Jiao and Deng, 2007; Peaston, 2004). Also stress response in plants and mammals is tightly linked to the activation of retrotransposons (reviewed by Grandbastien, 2010;
Mansour, 2007). Transcriptional activation and/or transposition have been seen across the plant kingdom in response to a wide variety of biotic and abiotic stimuli that elicit a stress response [Table 1] including wounding, pathogen attack and chemical treatments (Mhiri et al, 1997; Takeda, 1998; Ansari, 2007). A particular effective activator of retrotransposons is tissue culture induction and maintenance (Hirochika, 1993; Hirochika et al, 1996; Rakocevic et al, 2009; Hirochika, 2000). Retrotransposition has also been seen in response to allopolyploidization and “genomic shock” caused by interspecific hybridization in plants (Kashkush et al, 2002, Madlung et al, 2005).

Table 1. Examples of Stressed induced LTR-retrotransposons

<table>
<thead>
<tr>
<th>STRESS TYPE</th>
<th>Elements</th>
<th>Element Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue culture</td>
<td>Tto1, Tos17</td>
<td>Copia</td>
</tr>
<tr>
<td>Cell-wall hydrolases</td>
<td>Tnt1</td>
<td>Copia</td>
</tr>
<tr>
<td>Chilling</td>
<td>Tos17</td>
<td>Copia</td>
</tr>
<tr>
<td>Adenine starvation</td>
<td>Ty1</td>
<td>Copia</td>
</tr>
<tr>
<td>Wounding</td>
<td>TLC1</td>
<td>Copia</td>
</tr>
<tr>
<td>Protoplast preparation</td>
<td>TLC1</td>
<td>Copia</td>
</tr>
<tr>
<td>High salt concentrations</td>
<td>TLC1</td>
<td>Copia</td>
</tr>
<tr>
<td>Interspecific hybridization</td>
<td>Wis 2-1A</td>
<td>Copia</td>
</tr>
<tr>
<td>Adaptation to a moisture gradient</td>
<td>Bare-1</td>
<td>Copia</td>
</tr>
<tr>
<td>Microbial Factors</td>
<td>Tnt1</td>
<td>Copia</td>
</tr>
<tr>
<td>Mechanical damage</td>
<td>Tnt1</td>
<td>Copia</td>
</tr>
<tr>
<td>In vitro regeneration</td>
<td>Tos17</td>
<td>Copia</td>
</tr>
<tr>
<td>Viral infection</td>
<td>Tos17</td>
<td>Copia</td>
</tr>
<tr>
<td>Cytosine demethylation</td>
<td>Tos17</td>
<td>Copia</td>
</tr>
<tr>
<td>UV light</td>
<td>Reme1</td>
<td>Copia</td>
</tr>
<tr>
<td>Resistance to bacterial blight and plant development</td>
<td>Tos17</td>
<td>Copia</td>
</tr>
<tr>
<td>Fungal infection</td>
<td>Romani PP and Erika LTR</td>
<td>Gypsy</td>
</tr>
<tr>
<td>Heat shock</td>
<td>MAGGY</td>
<td>Gypsy-like</td>
</tr>
</tbody>
</table>

(Adapted from Mansour, 2007)

Retrotransposons show diverse activation patterns. A recent global transcriptional study in rice showed the differential and overlapping responses of retrotransposon families in response to different stress stimuli and in key developmental stages and tissues (Jiao and Deng, 2007). A similar study that focused on maize cell culture and...
organs (Vicient, 2010) showed similar interfamily variation in transcription of retroelements. A study in tobacco by Beguiristain et al. (2001) showed that the differential response even extended to elements within the same family. The tobacco retrotransposon family Tnt1 has three quasispecies that only differ by a single segment of the promoter region, yet each subfamily has a distinct response to various stimuli.

There is a strong correlation between adaptation and element transposition. A comparison of a hardy variety and a cold-sensitive variety of alfalfa showed transcripts of a non-autonomous retrotransposon family were found in the hardy variety after cold-stressed conditions but not in the sensitive variety (Ivashuta et al. 2002). Changes in the genome can be seen as well. An increase in copy-numbers of the retrotransposon BARE1 in barley cultivars correlates with adaptation to higher degrees of aridity (Kalendar, 2000). It is not known whether retrotransposon activity is an evolutionary strategy of the host to adapt to environmental challenges or if it is a by-product of loosening of epigenetic or other control mechanisms. A recent study showed that, at least one family of retrotransposons has a direct role. Retrotransposons in the “resurrection plant”, Craterostigma plantagineum, are involved in generating siRNA that enables metabolic pathways for desiccation tolerance (Hilbricht, 2008).

**LTR Retroelements and the Relationships Among Retrotransposons, Retroviruses and Endogenous Retroviruses**

**The Retroelement Genome**

LTR retroelements are 6,000 to 15,000 base pairs in length (Eickbush and Malik,
The simplest LTR retroelement that is transpositionally competent is the retrotransposon. It encodes all of the required genes and regulatory sequences needed for replication and transposition. The small number of proteins made by retrotransposons is common to retroviruses and endogenous retroviruses. These proteins are cleaved by an encoded protease from precursor polyproteins [Fig 1]. The Group antigen (gag) protein is cleaved into small structural proteins - capsid, matrix, and nucleocapsid - while polyprotein (pol) is cleaved into the enzymes integrase, reverse transcriptase, and RNase H. gag and pol are often encoded by a single ORF, but some retroelements possess ORFs separated by a stop codon or a frameshift (Gao, 2003) [FIG. 3]. The protease gene is found between gag and pol with variations in reading frames. Retroviruses contain other genes, such as envelope (env) [FIG.1], which encode proteins necessary for infection (Jern and Coffin, 2008).

Based on the order of the integrase and reverse transcriptase genes, which are transcribed as parts of pol, LTR-retroelements can be broadly divided into two major families, Ty1/copia and Ty3/gypsy [Fig. 1]. Both families are found in virtually all eukaryotes, from fungi to higher plants and animals. (Eickbush and Malik, 2002).

Retroelement proteins are translated from mRNA that is transcribed by the host RNA polymerase II from a single promoter within the LTR (reviewed in Freed, 2001). Because of the differences in arrangements, retroelements must use many strategies for expression (Gao, 2003). Some elements produce a single RNA to express all of their genes while others use a multigene transcript along with spliced transcripts (Jern and Coffin, 2008). In addition, translation often requires ribosomal frame shift or stop-codon
readthrough.

Fig. 3  Genomic organization of retroelements and expression strategies

Different arrangements of gag-pol found in retrotransposons. Expression strategies can require frameshift and stop codon readthrough. Most express a single transcript but splicing is known for some genes (Adapted from Havecker and Voytas, 2005).

Retroelement populations are made up of elements with different degrees of functionality. Those that encode all of the proteins necessary for transposition are known as autonomous elements. However, possessing functional copies is not required for transposition. Non-autonomous elements that do not produce functional copies of any or all of the proteins but have retained key nucleic acid motifs can transpose using the functional proteins encoded by other retroelements in the host genome (reviewed Feschotte, 2002).

Life Cycles of LTR Retroelements:

**Retrotransposons and Retroviruses**

The major functional and evolutionary feature distinguishing non-infectious retrotransposons from infectious retroviruses is the presence of a functional env gene (which codes for a functional envelope protein). Like the other retroelement proteins, the
envelope protein is cleaved into smaller functional units. One domain of this protein binds to a surface receptor on the target cell membrane allowing another domain to mediate the fusion of the membrane with the lipid envelope that surrounds the virus. This loads the virion into the cell cytoplasm (reviewed in Jern and Coffin, 2008). The same protein can mediate the release of some viruses from the cell (Shaw et al, 2003) [Fig. 4].

Fig. 4  Lifecycle of retroelements with and without env proteins

![Diagram of Lifecycle of Retroelements](image-url)

The lifecycles of infectious retroelements with functional env genes and non-infectious retroelements are essentially identical. However as retroviruses assemble they bud from the cell. Without the ability to infect, retrotransposons must remain in the cell to carry out their lifecycle.

Highlighting the similarities of the retrovirus and retrotransposon lifecycles, Grandbastien titled a section of her 2010 review, “LTR-Retrotransposons lead an Intra Cellular Retroviral Life”. The virion from the infectious retrovirus that enters the cell is made up of the structural proteins encoded by the gag gene. The virion contains necessary transpositional enzymes and the RNA genome of the retrovirus. Similarly
virus-like particles of retrotransposons assemble inside the cell using a nearly identical protein, the retrotransposon gag. These virus-like particles, also contain necessary transpositional enzymes and an RNA genome. From this point, the intracellular life of the retroviruses and the retrotransposons are almost identical [Fig. 4] For both retroviruses and retrotransposons, the RNA genomes are then reverse transcribed into double-stranded DNA and integrated into the host genome by element-encoded enzymes. Transcription by the host RNA polymerase II starts from a promoter in the LTR. A single mRNA is transcribed for translation of a gag-pol polyprotein, and the RNA is also transcribed into a single genomic RNA for assembly into the particle.

At this point, the life cycles diverge [Fig. 4]. In the retrovirus, the envelope protein is transcribed from a spliced RNA. The retroviral envelope along with the gag and pol polyproteins are transported to the cell membrane where they assemble into a virion and bud from the cell, cloaked in cell membrane. The retrotransposon virus-like particle, however, must assemble within the cell. With no envelope protein, it cannot theoretically infect new host cells. Successful replication must be carried out within the initial cell (Grandbastien, 2010; Jern and Coffin, 2008).

**Endogenous Retroviruses**

The ability of retroviruses to insert into the genome of a host cell allows them to be passed vertically in that cell's lineage and horizontally via infectious transfer of virions. Retroviral infection of germline cells can result in a retrovirus being passed vertically from the parental generation to the offspring generation (reviewed by Jern and Coffin,
2008). The integrated retroviral genome will be present in the germline of the offspring and may become fixed in the genome of the host species. These retroviruses that are passed vertically between host generations are called endogenous retroviruses. Though the endogenous retrovirus might initially retain the potential for infection, the genome-bound copy does not excise from the DNA and is presumably not maintained by selection, but rather exists as a result of genetic drift. Therefore the DNA copy frequently becomes a “fossil” remnant due to the accumulation of mutations that render it nonfunctional.

Disruption of only the env gene removes the potential for infection. Nevertheless, the endogenous retrovirus may still remain transpositionally competent and, like a retrotransposon, may still be able to replicate within the confines of a given cell [Fig. 4] while mutations continue to accumulate in its env gene. Because of this relationship between retroviruses and their endogenous counterparts, they share the characteristic of encoding all or part of an ancestral envelope protein, which is not found in retrotransposons. Thus evidence of their origin as an offshoot of retroviruses is retained in the genomes of endogenous retroviruses, even though their infective nature is lost.

**SIRE1 – A Plant Retrovirus?**

The Host Range of Plant Retrotransposons with Env-like ORFs

Despite the fact that LTR retroelements are ubiquitous across eukaryotes and retrotransposons make up the majority of many plants DNA, retroviruses and endogenous retroviruses had only been found in vertebrates, nematodes and insects (Malik et al,
2000). But the discovery and characterization of the 2000-member SIRE1 retroelement family in soybean and related elements in other plants has led to the suggestion that plants also possess endogenous retroviruses and therefore have previously had or still might have infectious retroviruses of their own (Laten, 1998; Wright and Voytas, 1998; Peterson-Burch et al., 2000; Vicent et al. 2001a; Vicent et al. 2001b; Wright and Voytas, 2002). In addition to genes that are functionally conserved and unambiguously homologous to the structural and catalytic proteins found in hundreds of more fully characterized retroelements, SIRE1 possesses an extra ORF [Fig. 1]. The theoretical translation of the product structurally resembles mammalian retroviral envelope proteins.

SIRE1 is the canonical member of the sirevirus clade, which have proliferated across a wide variety of monocots and dicots ranging from maize to citrus trees and of course to soybean [Fig. 5]. Identifiable homologous segments of the env-like ORF have even been found in Arabidopsis, tomato (Laten, 1999) and several other plant species (Springer Index of Viruses in press).

Many families of retrotransposons that cluster with sireviruses have truncated env-like proteins or have lost the protein altogether (Vicent et al. 2001a; Vicent et al. 2001b; Havecker and Voytas, 2005). It is not known if the truncated proteins retain any functionality, but it appears that the env-like protein is dispensable to the sirevirus life cycle in plants. Despite this, the env-like gene ORF has been preserved without frameshift or interruption in both Lotus japonicus and soybean, which diverged over 50mya (Holligan, 2006; Vitte 2006; Laten, unpublished). This suggests that the gene might play a key role in the functioning of the retroelement.
As stated above, LTR-retroelements can be subdivided into two major families, Ty1/copia and Ty3/gypsy, based on their gene order [Fig. 1]. SIRE1 and the sireviruses are a part of the Ty1/copia family. All other known retroviruses and endogenous retroviruses are related to Ty3/gypsy. Some Ty3/gypsy elements in plants have also been suggested to have remnants of env-like proteins and thus to be endogenous retroviruses (Wright and Voytas, 2002), but like the sireviruses, their infectious ancestry remains unconfirmed. Confirmation of the SIRE1 env-like protein as a true envelope protein would not only extend the range of endogenous retroviruses to plants, but also to Ty1/copia elements as well.

(Adapted from Havecker and Voytas, 2005)
Env-like Versus Envelope

It should be asked why one would look for retroviruses or their endogenous counterpart in plants. As stated above, envelope proteins, which are found in the lipid envelopes of viruses, serve two major functions in viral infectivity: they sponsor infection into new animal cells by allowing fusion of the cell membrane with the viral envelope, and they can mediate viral budding from infected animal cells. The cell walls of plants create an obstacle for entry to and exit from the cell and would prohibit the systemic spread of a virus by cell-to-cell infection by means of membrane fusion.

Many viruses known to infect plants have adapted to this situation by employing movement proteins (reviewed in Lucas, 2006). These proteins carry the viral RNAs and DNAs through the plasmodesmata to infect neighboring cells. As Laten (1998) noted, if the env-like protein is involved in infection, it is unlikely to be a membrane protein. The env-like protein has no predicted structural features that would support movement protein function and is predicted to be over twice the size of known movement proteins, which are approximately 30 kDa.

Enveloped plant viruses are not uncommon, but the envelope proteins appear dispensable in plants. Tospoviruses and plant rhabdoviruses both have envelope proteins that they use to infect insect cells and thereby alternate between plant and insect hosts (Lazarowitz, 2007). Resende et al (1991) have shown using one family of tospoviruses that after generations of bypassing the insect host and serially inoculating plants by mechanical means, some of the viruses lost the function of their envelope protein due to mutations. Nevertheless they retained the ability to infect and multiply in the plant host.
The env-like ORF does not appear to be necessary for retrotransposon proliferation since it is absent from the vast majority of plant retrotransposons, but the env-like reading frame has nevertheless been retained across tens of millions of years of sirevirus evolution (from *Lotus japonicus* to soybean). This suggests two possibilities: Either the protein is an actual envelope protein that has recently been transmitted to and from an animal vector, or this protein has a novel function in sirevirus retrotransposons with no known homologs.

In support of the envelope hypothesis, the SIRE1 env-like protein has many shared characters with retroviral envelope proteins. The env-like protein was predicted to have transmembrane regions, glycosylation sites, \( \alpha \)-helices, and a coiled-coil that sponsors oligomerization - all features of mammalian retroviruses (Laten, 1998; Freed, 2001) [Fig. 6]. Further comparison shows that these elements’ \( \alpha \)-helices are similarly distributed in the SIRE1 and mammalian retroviral proteins, and that two of the transmembrane regions share similar positions in the two proteins. They are located near the N-terminus and just

![Fig. 6](image.png) Comparison of Retroviral and SIRE1 envelope and env-like proteins

Similarities can be seen in structural features and positions between the composite retroviral envelope protein (adapted from Laten, 1998) and predicted secondary structure features of the SIRE1 env-like protein.
upstream of the coiled-coil (Laten, 1998). On the other hand, the SIRE1 protein lacks a typical furin cleavage site common to retroviral envelope proteins and an anchor peptide (Laten, 1998). Other proteases besides furin are known to cleave envelope proteins though (Shulla, unpublished; see Discussion). In addition to the structural similarities, the amino acid composition of a proline-rich region of the predicted translation product is also similar to other retroviral envelope proteins (Laten, 1998) [Table 2].

Table 2. Comparison of 60-residue proline-rich regions from SIRE-1 and mammalian retroviruses. FeLV, feline leukemia virus; MLV, murine leukemia virus.

<table>
<thead>
<tr>
<th></th>
<th>FeLV</th>
<th>MLV</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>S+T</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>F+W+Y</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>I+L+V</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>N</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Q</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>A+G</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>D+E</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>K+R</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>C+M</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Similarities in amino acid composition can be seen in the Sire-1 env-like protein and retroviral proteins.

Despite the similarities between the SIRE1 env-like gene and mammalian retroviral envelope proteins, there is no evidence for sequence homology. The envelope of the insect endogenous retrovirus Gypsy is known to sponsor infection (Kim et al., 1994; Song et al., 1994; Teysset et al., 1998), and there is also no discernible evidence of homology between it and mammalian retroviral envelope proteins (Song et al., 1994). Furthermore, evolutionary kinship is unclear even between retroviral and influenza envelope proteins (Harrison, 2005), which are known to sponsor infection by nearly identical structural and functional mechanisms (Harrison, 2005). Though the hereditary connection between env
and the env-like gene is not clear, envelope proteins that sponsor infection have likely been captured many times by retroelements (Malik, 2000).

Possibility of Active and Infectious SIRE1

If indeed SIRE1 is an endogenous retrovirus, one key question is whether or not it has retained its infectivity. In general, the envelope protein of a new endogenous retrovirus should be intact, and the element may retain its ability to infect other cells. But an endogenous retrovirus is likely to be repressed quickly by gene silencing mechanisms of the host where the env-like gene is subject to mutation (Jern and Coffin, 2008). Though endogenous retroviruses in vertebrates are known with seemingly intact elements and potentially functional envelope proteins, none have has been shown to cause infection in its natural state (Jern and Coffin, 2008). Recently though a “resuscitated”, consensus human endogenous retrovirus was shown to sponsor infection in human cells (Dewannieux, 2006). The consensus was made because of significant polymorphisms in the element, which was thought to have had its last amplification about 5 mya.

SIRE1 on the other hand has undergone a recent expansion having elements less than 30,000 years old (Laten, 2003). Dozens of intact SIRE1 elements exist in which the env-like gene has no discernible mutations that would make the protein non-functional (Laten, 2003; Du, 2010; Laten, unpublished).

In Drosophila, the endogenous retrovirus Gypsy has an env-like protein that does sponsor infectious transfer (Kim et al, 1994; Song et al, 1994; Teysset et al, 1998). Interestingly, when it was first found, the range of retroviruses was limited to vertebrates, and Gypsy was presumed to be a retrotransposon. However, because Gypsy possesses an
extra ORF in the same position and with similar predicted structural features as the retroviral envelope, it was argued that Gypsy was an endogenous retrovirus (reviewed in Song et al, 1994). It was then shown that active Gypsy elements could be transferred horizontally and that the env-like protein could sponsor infectious transfer (Kim et al, 1994; Song et al, 1994; Teyssset et al, 1998). This not only reclassified Gypsy from a retrotransposon to an endogenous retrovirus, thereby expanding the range of retrovirally related elements to insects, but it also showed that endogenous retroviruses can also retain their infectious ability.

Though endogenous retroviruses are not yet known in plants, the case for SIRE1 has been argued with similar points (Laten, 1998). The fact that dozens of copies of intact elements exist (Laten, unpublished; Du 2010) bolsters the possibility that active SIRE1 elements can be found. Showing that SIRE1 elements are active and produce an env-like protein that sponsors infection will extend the range of endogenous retroviruses and potentially retroviruses to both plants and to Ty1/copia elements.

**Goals and Methods of This Study**

It has been the long-term goals of the lab to test the hypotheses that 1) SIRE1 is an active retroelement family, 2) that it is a specifically an endogenous retrovirus family and 3) that it is also infectious. This study aims to begin the characterization of the env-like protein in an effort to generate experimental data that would support the above goals. The first step in characterizing this protein is to show that the protein is being expressed in soybean plants.
Preliminary Evidence of Activity

Some experimental data show transcription of one or more SIRE1 copies. Twenty ESTs corresponding to SIRE1 sequences have been found in the Genbank EST database. (Laten, unpublished). Some of these cDNAs were derived from stressed tissues. However, approximately two thirds of the ESTs are antisense and thus unusable as a substrate for translation. It cannot be determined if sense the transcript represents transcription from a promoter within the LTR or if it is a product of readthrough from a cryptic or other promoter adjacent to one of the many copies of the family. cDNAs have also been recovered for SIRE1 in water stressed root tissue (Winfrey, unpublished). SIRE1 RNA has been detected in soybean tissues (Lin, 2001). Though no RNA was detected in Northern blots, RT-PCR showed evidence of SIRE1 transcripts in leaf and root tissues using primers that amplify gag and reverse transcriptase regions. Using primers for the env-like region, transcripts were detected in leaf tissues. However, detecting fragments of transcripts from a repetitive DNA family is not considered to be evidence for biologically relevant expression of protein products (see below). Despite this, it is encouraging evidence that the SIRE1 family is still active and that the env-like protein is being expressed.

Methodology

I have chosen to study production of the env-like protein by directly detecting it using antibodies generated against it. Though nucleic acid studies are faster and much cheaper, they do not provide direct evidence of the production a protein product and in
the case of repetitive DNAs making such an inference might be unreliable, as described below.

**Limits of Nucleic Acid Studies**

**RNA studies**

Mutations in methylation mechanisms, which lift transcriptional controls, have been shown to allow transposition of a variety of plant retrotransposons (Hirochika, 2000; Tsukahara, 2009). In these cases retroelement proteins have been produced. As stated above, transcriptional controls are thought to be the predominant mechanism for suppressing retroelement activity in plants (Hirochika, 2000; Feschotte, 2002).

However in a survey of transposable element ESTs in maize, the author, after seeing large amounts of transcription from some families of transposable elements suggested that either “transposition is more frequent than previously expected, or cells can control transposition at a post-transcriptional level.” (Vicient, 2010) Post-transcriptional gene silencing (PTGS) mechanisms are known to help control retrotransposons in organisms such yeast and *Neurospora crassa* (Garfinkel, 2003; Nolan, 2005). In addition PTSG is a common method of silencing plant viruses and is linked with TGS in concurrent initiation pathways (Vaucheret, 2001).

For the non-LTR retroelement Karma in rice, transcripts were detected in both cell cultures and the leaves of plants regenerated from the culture, but transposition was only found in the latter (Komatsu, 2003). The authors suggest this as evidence for post-transcriptional controls.
In support of post-transcriptional controls in LTR retroelements, some studies have shown discrepancies between RNA levels and transposition. In a heterologous Arabidopsis system, RNA levels from the tobacco retrotransposon Tto1 were compared between DNA methylation-deficient mutants and cultured tissue (Hirochika, 2000). Though the retrotransposon produced higher amounts of RNA in the mutant plants, transposition was only demonstrated in cultured cells. Likewise, in tobacco, Takeda (2001) found that Tto1 transposition intermediates, such as double-stranded DNA, were only found in tissue culture though retrotransposon RNA was expressed at higher levels in flowers.

A closer look at transcription in tissue-cultured tobacco found that the vast majority of the detected Tto1 RNA was missing 28 bases (Böhmdorfer, 2005). The RNA was poorly translated and was not a substrate for reverse transcription. This RNA was however able to form a double stranded segment, which the author suggested might promote RNA silencing.

Furthermore, in Arabidopsis methylation mutants, inhibition of transposition of one retroelement remained despite detectable transcripts. The element was shown to transpose only with a combination of mutations, which included mutations in Pol IV and Pol V (Mirouze, 2009). Pol IV and Pol V has recently been shown to play a key role in producing siRNA for post-transcriptional defense mechanisms (Pikaard, 2008; Pontes, 2009). Thus, relying on the presence of transcripts to infer production of a protein product should be made with caution. So while the detection of RNA suggests that SIRE1
elements may be still functional and active, transcription does not equate with functional activity. Therefore neither Northern blots nor RT-PCR studies are sufficient for its demonstration.

DNA studies

Transposition can be detected by several DNA methods. Proof of transposition to new sites would imply functional activity of some SIRE1 proteins. However the env-like protein is likely a product of readthrough of a stop codon (Havecker and Voytas, 2003), and its role in the proliferation of the element within the cell is not known. As there is a substantial hindrance to env-like protein production and since the env-like protein might not even be necessary for transposition, production of the env-like protein cannot be inferred.

Immunological Methods

In light of these arguments, evidence of production of the env-like protein was sought directly by immunoblotting with antibodies specific to the env-like protein.

Our lab has created separate DNA constructs that encode segments of the gene, including the coiled-coil region, for production in E. coli. The constructs encoded the gene segments as fusion peptides with an N-terminal His-tag and a C-terminal Strep II-tag (Strep-tag), which were used for purification of the fusion peptides. Custom polyclonal antibodies were then made against the purified peptides and used to probe
proteins from key developmental stages and distressed soybean tissues for the production of the env-like gene.

**Conclusion**

Our results are suggestive that the env-like protein can be induced under stress conditions and might also be present in embryo and seedling tissues. Additional verification steps are needed to confirm that the proteins found in this study correspond to the SIRE1 env-like protein, but these methods, are well established. If the protein(s) found in this study can be eventually be verified as an envelope protein that sponsors SIRE1 infection, it would be the first characterized plant retrovirus and the first retrovirus related to the Ty1/copia retroelement superfamily. In this study, I have laid the groundwork to begin to directly examine these questions.
CHAPTER TWO

MATERIALS AND METHODS

Polyclonal Antibody Production

Antigen Source

PCR and Recombinatorial Cloning

PCR Primers were designed to amplify various regions of the SIRE1-4 env-like gene (NCBI accession: HYPERLINK "http://www.ncbi.nlm.nih.gov/nuccore/AY205608" AY205608, REGION: 5978..8020). Regions chosen excluded putative transmembrane domains but included a predicted coiled-coil domain Fig. 7. These primers preserved the theoretical reading frames of the sequences and included a 4 bp extension to add recombination sites for use with the IBA Stargate (Germany) expression system. (Primer pairs synthesized at Eurofins Mfg Operon. Added non-soybean sequence is underlined.

**primer19** Fwd: 5’-**AATGACCTCACC**AAAAGAAACTGCA, Rev: 5’-**TCCCCATT**GG**GGAAGATGTT**; **primer20** Fwd: 5’- **AATGGCAG**AAAG**GTTAC**AA**GAG**ACA**, Rev: 5’- **TCCCTTTCCACT**TTTCAAC**ATTGC** ;

Fig. 7  Location of amplified regions on the env-like protein

![Location of amplified regions on the env-like protein](https://example.com/fig7)

Map of the predicted features of a conceptual translation of the env-like gene and approximate coverage of the cloned segments (marked by dumbbells).
**primer21 Fwd:** 5' -AA**T**GGAATCTCCTCTATCCCTGCAT, **Rev:** 5' -TCCC**T**TCTTCA**T**CATCAATGCCAC). 3’-phosphorothioate protected primers were used to prevent degradation of the primers when used with Pfu polymerase.

The PCR template was a purified *E. coli* plasmid, e006-1 (Laten, unpublished) that contains a copy of the env-like gene. A single *E. coli* colony harboring the plasmid was inoculated in 5 ml of LB and grown overnight at 37°C in 100 mg/L ampicillin. OD_{600} was between 1.5 and 5.0. Plasmid was purified with the Roche HighPure miniprep using the provided reagents and protocols. Plasmid was then eluted with 50 µl of water and centrifuged 13,000 x g for 1min.

The sequences were amplified using native Pfu DNA polymerase (Stratagene).

The PCR protocol used is as follows:

- 10x Pfu buffer: 5 µl/50 µl
- dNTPs: 250 µM
- Each primer: 0.5 µM
- Template: 15pg/50 µl
- Pfu pol.: 1.25U/50 µl

45s at 95°C, (45s at 95°C, 30s at 58°C, 4:30min at 72°C) x30, 10min at 72°C

Amplicon lengths from each primer pair were verified by gel electrophoresis on a 2% agarose gel with 0.5x Tris-Borate EDTA Buffer (44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA). Amplicons were run through Wizard PCR cleanup (Promega) using manufacturer provided reagents and protocols and then sent for sequencing (University of Chicago Cancer Research Center DNA Sequencing Facility) using the forward primers used in the amplification. Amplicons were then used in the IBA (Germany) Stargate...
expression system using reagents and protocols provided by the manufacturer. The Stargate system is a two-vector system [Fig. 8]. 4.6 µl of each of the three PCR products (14 µl total) was mixed with 10 µl of IBA reagents to be recombinatorialy inserted into the first vector or “donor vector” (pENTRY-IBA51, Fig. 9a and 9b). 10 µl from the 25 µl of reaction mixture was added to 100 µl of chemically competent *E. coli* (strain: TOP10; genotype: F- mcrA (mrr-hsdRMS-mcrBC) 80lacZ M15 lacX74 recA1 ara 139 (ara-

Fig. 8    IBA Stargate two-vector system

Step 1: In the first step PCR is used to amplify and extend the gene of interest (GOI) using primers that add recombinatorial sites. The PCR product is then recombined into the Donor Vector. Step 2: The GOI is recombined from the Donor Vector into one of an array of Destination (Expression) Vectors. The Destination Vector used in this project codes for a recombinant protein with an N-terminus 6xHis tag and a C-Terminus StrepII tag.

leu)7697 galU galK rpsL (StrR) endA1 nupG; IBA, Germany). Heat shock was performed according to IBA protocol. Cell/vector mixture was incubated on ice for 30 min. The mixture was then gently mixed and incubated at 37°C for 5 min. The mixture was then gently mixed and again incubated on ice for 5 min. 900 µl of LB was added and the mixture was shaken for 45 min at 37°C. 100 µl of the mixture was then plated on LB
agar containing 50 mg/L kanamycin and 50 mg/L X-gal. The residual 900 µl of cell mixture was centrifuged for 30 sec and the supernatant was removed. Cells were then resuspended in 100 µl of LB medium and plated as above. Colonies expressing kanamycin resistance were picked and inserts of their plasmids were identified by colony PCR (described below). Plasmids containing single inserts were isolated (HighPure miniprep, Roche) and sequenced using a vector specific primer provided by IBA (Germany; ENTRY-Primer-rev, 5’-CCCCTGATTCTGTGGATAACCG -3’, Fig. 9b). Only one primer was necessary since the read was over 500 bp and easily reached the beginning of the insertion.

Fig. 9  Vector map and partial sequence for pEntry-l

(a) Vector contains a kanamycin resistance gene(KanR).  (b) Sequence spanning the cloning site. Manufacturer-provided primers are marked. The 4bp on either side of the site are the recombinatorial sites and are complimentary to the 4bp extensions attached to the amplicons during the PCR step.
2 µl of each donor vector (2 ng/µl) with one of the three insertions (6 µl total) was mixed with IBA reagents to be recombinatorialy inserted into the second vector or “destination vector” (pASG-IBA43, Fig. 10a and 10b). This is the expression vector. 10 µl from the 12.5 µl of reaction mixture was added to 100 µl of chemically competent *E. coli* (TOP10). Cells were transformed by heat shock as above. 100 µl of the mixture was then plated on LB agar containing 100 mg/L ampicillin. The residual 900 µl of cell mixture was centrifuged for 30 sec and the supernatant was removed. Cells were then resuspended in 100 µl of LB medium and plated on LB with 100 mg/L ampicillin as above. Colonies expressing ampicillin resistance were picked and inserts of their plasmids were identified by colony PCR (described below). Plasmids containing single inserts were isolated (HighPure miniprep, Roche) and sequenced using a vector specific primer provided by IBA (Germany; ASG-Primer-rev, 5’- CGCAGTAGCGGTAAACG - 3’, Fig. 10b).

**Expression**

Pilot studies were done as follows: freshly plated *E. coli* colonies were grown overnight in 4 ml of LB with 100 mg/L ampicillin at 30°C. The culture was then inoculated into 200 ml of fresh LB/amp media for expression. When OD$_{550}$ reached 0.5-0.7, cells were induced with anhydrotetracycline (IBA, Germany) at a concentration of 200 ng/ml. After 3 hours at 37°C, cells were harvested by centrifugation, and pellets were stored at -20°C. For scaled-up expression, overnight culture and expression culture volumes were increased by a factor of 10.
A cleared lysate of the cells was then produced by native or denaturing protocols (IBA, Germany, purification protocols). For denaturing conditions, pellets were thawed on ice and resuspended in 5 ml/g of wet weight in Buffer B (100 mM NaH$_2$PO$_4$, 10 mM Tris-Cl, 8M urea; pH 8.0). Cell mixture was vortexed until translucent. Under native conditions, pellets were thawed on ice and resuspended in 5 ml/g of wet weight in Ni-NTA Lysis Buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM imidazole; pH 8.0).
1 mg/ml of lysozyme (Sigma) was added, and the mixture was incubated for 30 min on ice. The cell mixture was sonicated in an ice bath using a 1.5 cm tip on an Artek Sonic 300 Dismembrator set to 60% (~180 W) until solution was translucent. 20 sec bursts were alternated with 20 sec cooling periods. Debris was cleared from lysates by centrifugation at 10000 x g for 25 minutes. Expression was then verified by SDS-PAGE using an anti-6xHis monoclonal antibody (IBA, Germany).

**Purification of Peptides**

**His-tag purification:**

**Native purification:** a column was packed with 0.75 ml of settled HisPur Cobalt Resin (ThermoFisher), and storage buffer was allowed to drain. Pilot studies used only 0.5 ml of settled resin. All samples and buffers were drained by gravity flow. pH of the sample was adjusted to 7.4. The column was equilibrated by adding two resin-bed volumes (RV) of wash buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole; pH 7.4). Cleared *E. coli* lysate from the protein expression was added onto the resin and collected as flowthrough. Flowthrough was reapplied to column to maximize binding. Resin was washed 6-8 times with two RV of wash buffer. Peptides were eluted with two RV of elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 150 mM imidazole; pH 7.4). Elution was repeated two more times. Elutions were collected separately.

**Denatured His-tag purification:** All steps were carried out as in the native His-tag procedure above, except that all buffers contained 8 M urea in addition to the listed
chemicals. In preparation for Strep purification denatured eluants were either diluted 13x with (native) wash buffer or dialyzed as described below.

**Dialysis of samples:** Samples with small volumes were transferred to 2 ml Spectra/Por Ready-To-Use dialysis tubes with a 500 Da molecular-weight cut-off. Samples with larger volumes were transferred to SnakeSkin dialysis tubing (ThermoFisher) with 7 kDa molecular-weight cut-off. Samples were dialyzed against wash buffer without imidazole or urea at 4°C overnight. Buffer was then changed, and samples were dialyzed at 4°C for 2-4 more hrs.

**Gradient purification (native):** All steps leading up to washes were done as in the native His-tag purification above. For purification of peptide p21, wash volumes and concentrations of imidazole in the wash buffer are as follows: 4 RV containing 10 mM imidazole, 8 RV containing 30 mM imidazole, and finally 1 RV containing 50 mM imidazole. Elutions were all done with 2 RV of elution buffer at each of the following concentrations of imidazole: 75 mM, 100 mM, 130 mM, 160 mM, 200 mM, and two separate 250 mM aliquots. Each fraction was collected separately. For purification of peptide p19, wash volumes and concentrations of imidazole in the wash buffer are as follows: 12 RV containing 10 mM imidazole, 6 RV containing 10 mM imidazole at 37°C, 9 RV containing 30 mM imidazole, and finally 1 RV containing 50 mM imidazole. Elutions were all done with 4 RV of elution buffer at each of the following concentrations of imidazole: 50 mM, 80 mM, 110 mM, 140 mM, 170 mM, and 250 mM. Each fraction was collected separately.
Strep-tag purifications:

400 µl of settled Strep-Tactin Sepharose resin (IBA, Germany) was packed in a column. Pilot studies only used 100 µl of settled resin. All samples and buffers were drained by gravity flow. The resin was equilibrated with 2 RV of buffer W (100 mM Tris-Cl, 150 mM NaCl, 1 mM EDTA; pH 8.0). Eluants from an individual His-purification were pooled together, and pH was adjusted to 8.0. The eluants were then added to the Strep-tactin column. The column was washed 8 times with 1 RV of buffer W. The peptides were then eluted with 1 RV of buffer E (100 mM Tris-Cl, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin (IBA, Germany); pH 8.0). The Elution was repeated two more times. Elutions were collected separately.

Antibody Production

His and Strep purifications of peptides 19 and 20 were combined to get 1.89 mg of protein at 210 µg/ml. Approximately equal amounts of each 19 and 20 were used. The peptide mix was sent to OpenBiosystems for antibody production following their 70 day (10 week) 2-rabbit protocol. Total sera was received. Titration of the total sera was performed by OpenBiosystems using the supplied antigen and gave extremely high titer counts.

Titer protocol (Open Biosystems): Titers are reported as the reciprocal of serum dilution, and values were calculated by measuring the dilution point where the absorbance drops below 0.2 at OD405 (4 times background). Post-inject antibodies were detected by indirect EIA against the passively coated peptide or protein with anti-species IgG-HRP conjugate and ABTS substrate. Peptides were coated directly onto the plate,
and conjugates were not used in ELISA. This is to avoid any inflation of the titers due to the presence of linker or conjugate specific antibodies.

**Polyclonal Antibody Purification**

This Antibody purification protocol is adapted from the Molecular Cloning protocol (Sambrook, *et al* 1989; pp. 18.16-18.18). 355 µg of His- and Strep-purified 19, and 537.5 µg of His- and Strep-purified 21 were set aside for antibody purification. Typical nitrocellulose membranes have an average binding capacity 90 µg/cm² (Millipore). To match the binding capacity, two membranes were cut to 1.3x3.0 cm² and 2.0x3.0 cm², for 19 and 21 respectively. The gasket support plate and the vacuum manifold of a BioDot (BioRad) apparatus were used to blot the purified peptide onto the membranes. The gasket support plate was placed inside the manifold, and two sheets of thick filter paper were placed on top of the gasket support. Two holes roughly the same size as the membranes were cut into a piece of parafilm that otherwise covered the apparatus. The parafilm was taped to the apparatus so that all suction came through the two membrane-sized holes. The filter paper was primed by wetting it with Tris-Buffered Saline (TBS). A vacuum was pulled until the buffer disappeared. The membranes were then placed on top of the filter paper and wet with TBS. Again a light vacuum was pulled just until the buffer disappeared. Purified peptides were then applied to their respective membranes until each was entirely covered. A vacuum was pulled just until the peptide solution disappeared. Application of peptides followed by vacuum was repeated until all peptide had been added to the membrane.
Protein binding was checked by Ponceau stain. The membrane was then allowed to sit for one hour in blocking buffer (5% milk powder in TBS) at room temperature. The membranes were then placed around the inside of a 20 ml beaker so that there was no overlap. 9 ml of crude sera was added. The height of the sera was increased to fully cover the membranes by adding small vial to the center of the beaker. The beaker was gently shaken for 6hr at room temperature. The membranes were removed from the beaker and rinsed quickly in 0.15 M NaCl. The membranes were then washed first in 0.15 M NaCl with gentle shaking for 20min and then with PBS with shaking for 40min.

The membranes were then placed on parafilm that was pressed into the bottom of a Petri dish. Elution buffer (0.2 M Glycine (pH 2.8), 1 mM EGTA) was added to the membrane at a volume of 50 µl/cm² of membrane. The membranes were gently shaken in a 25°C humidified incubator for 20 min. Antibodies were pipetted off of the membranes and transferred to a microfuge tube. 0.1 volume of 1 M Tris base (pH 8.0) was added to the antibodies. The antibody mixture was spotted on pH paper to verify near-neutral pH. Then 0.1 volumes of 10x Phosphate-Buffered Saline and sodium azide to 0.02% were added. Membranes were re-blocked in blocking buffer for one hour, and the purification was done a second time.

**Sample Materials**

**Plant Materials**

Unless otherwise stated seeds from Williams 82 cultivars were germinated with light in a tall covered dish containing approximately 3 cm of moist, commercially available seed starter mix (sphagnum moss, vermiculite, perlite). Once seedlings grew to
approximately 10 cm, they were transferred to small pots with soil and grown with natural light. Soybean leaves taken for treatment were young leaves from the top two nodes of the plant. Whenever possible a selection of different sizes and ages (from the top nodes) were pooled together for individual treatments. Tobacco leaves were taken without regard to age. Root samples were taken from plants > 4wks old. Putative hybrid material (gift from Thomas Carter, USDA-ARS) comes from a putative mating of *G. max* maternal x *G. soja* paternal.

**Embryos**

Hot water was poured over dry seeds and left to sit for upwards of three minutes until soft enough to split. Seeds were split and embryos were excised. The number and cultivars of embryos harvested are as follows: 25 TCAxBO9 N7103 (gift from Thomas Carter, USDA-ARS), 10 Clark, and 10 Forrest.

**Seedlings**

Equal numbers of seeds from Williams 82, Er-hej-jan, Ark Soy, and PI96983 cultivars were allowed to germinate for three days on the bench top in a Petri dish between two moistened paper towels. Seedlings were split into two categories depending on total length: large seedlings >1.5 cm and small seedlings <1.5 cm. Small seedlings were ground whole while older seedlings were split into shoots and radicals.
**Tissue Culture Induction**

Williams 82 seeds were sterilized in a 10% bleach solution for 10 minutes. Seeds were then rinsed with 1% bleach then washed several times in sterile water. The seeds were then placed in sterile jars containing autoclaved germinating media (4.4 g/L MS salts (Sigma), 3% sucrose, 0.8% phytoagar; pH 5.8.;1x antibiotic/antimycotic (Sigma) added after autoclaving). After seedlings had reached about 10 cm, the plants were cut with a sterile razor blade into pieces that would expose as much of the undifferentiated tissue in the middle as possible. Pieces were then transferred to culture media (germinating media with 0.5 mg/L 6-Benzylaminopurine (BAP) and 5.0 mg/L 1-Napthylene acetic acid (NAA) both added after autoclaving). Root and stem explants were mixed together and incubated in darkness at room temperature and subcultured every 3-4 weeks. Explants from one week and three weeks after the third subculture were used for extraction.

**Water-Stressed Tissues**

The tray holding the potted plants was tilted to a slight pitch and Plants were not watered for 5 days. This gave a range of stress to the plants as detected by their leaves. Leaves were considered as lightly stressed if they showed only mild rippling. Severely stressed leaves felt papery to the touch. Root tissue was combined from plants showing lightly and severely stressed leaves. Moderately and mildly stressed leaves from the first and second nodes were used for leaf tissue.
Leaf Treatments

After each leaf treatment the leaves were allowed to rest for 24 hours in darkness between paper towels damp with .05% 2-(N-morpholine)-ethanesulfonic acid (MES) buffer, pH 5.7, or the treatment buffer. In cases of vacuum infiltration (VI), leaves were placed in the treatment solution in a 50 ml centrifuge tube in a vacuum bell jar. Air was evacuated by a vacuum pump pulling 27mmHg of pressure. Leaves were left under vacuum for >2 min. Release valve was turned, and air was allowed to return rapidly. This was repeated until >50% of the leaf tissue of each leaf tissue appeared translucent and saturated with treatment buffer. All material was ground by mortar and pestle in liquid nitrogen.

**Methyl Jasmonate (MJ)** – VI with 100 mM or 300 mM MJ in 0.05% MES soln.

**CuCl₂** – VI with 0.5 mM or 10 mM CuCl₂ in 0.05% MES at pH 5.7.

**Salicylic acid** - VI with 2 mM salicylic acid in 0.05% MES

**Cellulase** – either the petioles of leaves were bunched together and dipped into the cellulase solution (2U/ml) overnight or VI to 30-70% saturation of the leaves.

**Freeze** – Leaves were placed in a single layer between the backsides of two Petri dishes. Liquid nitrogen was poured into the top dish freezing the leaf. Leaves were left for 30s and then removed and allowed to rest.

**UV** – Leaves were exposed to direct UV light at 312nm using a Spectroline, UV Transilluminator (model: TVR-312R, (6) 15W UV-B) for 2-3 minutes on each side, then allowed to rest.
**Wounding** – Leaves were punctured using a 200 µl pipette tip at a density of approximately 5-10 punctures per cm².

* Though the youngest leaves on the plants were taken, plants in these treatments were just beginning to flower.

** Initial treatments were done with leaves from young plants. Later treatments were performed on leaves from older plants that were just beginning to flower.

---

**Protein Extraction from Plant Tissues**

A variation of a Trichloroacetic acid (TCA)-Acetone extraction method was used (Méchin, *et al.*, 2007). All samples were kept at -80°C until extraction. All centrifugations were at 12000 x g for 12-15 minutes at 4°C. Approximately 0.17-0.2 g of ground sample material was placed in a 15 ml plastic centrifuge tube and was mixed with 5 ml of precipitation solution (Acetone, 10% TCA, 0.07% BME) at -20°C. Samples were then incubated at -20°C for 1 hour. The samples were transferred to three 2 ml microcentrifuge tubes and centrifuged. Precipitation solution was pipetted off, and 1.8 ml of -20°C wash solution (Acetone, 0.07%BME) was added. Samples were incubated for 1 hour at -20°C. The samples were centrifuged, and the buffer was removed by pipetting. Wash solution was added, samples were centrifuged and wash solution was removed two more times without incubation. After the final removal of wash solution, samples were placed under vacuum to dry. At approximately 40 minutes, tubes were closed and tapped on the bench top to help drying by loosening the pellet. Samples were then placed back under vacuum for approximately 20 more minutes. Dried samples were recollected into one tube. Samples were then mixed with solubilizing buffer (urea 8 M, SDS 5% w/v,
Tris-HCl [pH 6.8] 40 mM, EDTA 0.1 mM). Generally two times the volume of the loosened pellet was used. Samples were then vortexed at low speeds for an hour and again centrifuged. Protein solution was removed and stored at -20°C.

**General Procedures**

**Colony PCR**

For this PCR reaction lysed colonies were used as template. This allowed numerous reactions and quick results by bypassing the plasmid isolation step for each colony. A 25 µl reaction was set up as below. For the template, a barely visible amount of a fresh isolated colony was picked with a toothpick. The colony was deposited into the reaction tube by twirling the toothpick. Lysis took place in the first step at 95°C, which is run for 4:30min. Primer pairs primer19, primer20 and primer21 were mixed into each reaction. The PCR reaction was set up as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume/Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2x GoTaq Green Master mix</td>
<td>12.5 µl/25 µl</td>
</tr>
<tr>
<td>Each primer</td>
<td>0.5 µM</td>
</tr>
<tr>
<td>Template</td>
<td>barely visible amount of single colony</td>
</tr>
</tbody>
</table>

4:30min at 95°C, (45s at 95°C, 30s at 58°C, 2:15min at 72°C) x30, 7min at 72°C.

**Protein Concentrations**

Protein concentrations were found by the BCA method (Smith, 1985) or by fluorometry using a Quantit spectrophotometer (Invitrogen). Samples were diluted as necessary.
SDS-PAGE and Western Blot:

Proteins were separated by SDS-PAGE in a Mini-Protean electrophoresis cell (BioRad) using 18% or 7.5% precast polyacrylamide gels (Ready Gel, BioRad) in a Tris-HCl buffer system. Samples were mixed 1:1 with Laemelli’s 2x Sample buffer (62.5 mM Tris-HCl (pH 6.8), 25% Glycerol, 2% SDS, 0.01% Bromophenol Blue, 5% BME added before use) and heated for 15 minutes to 70°C. Samples were mixed without dilution unless otherwise stated. Running buffer (49.5 mM Tris-HCl, 383.6 Glycine, 0.1% SDS) was added to both inner and outer chambers of the cell. The comb was removed from the gel and the wells were washed with running buffer by pipette. 17 µl of the sample was then loaded into the wells. 9 µl of Prestained Molecular Weight Markers, 26 kDa-180 kDa (Sigma) or Low-Range Rainbow Molecular Weight Markers, 3.5 kDa-38 kDa (GE Healthcare) were used as standards in 7.5% and 18% gels, respectively.

Gels were run at 110V for 1 hour. The gel was then removed from the cast, and the wells and foot were removed with a razor. A nitrocellulose membrane, filter paper and blotting pads were cut to the size of the gel. The gel was equilibrated with gentle shaking in cold Transfer Buffer (48 mM Tris, 39 mM Glycine, 20% methanol, pH ~ 9.2 without adjustment; Bjerrum and Schaffer-Nielsen, 1986) for 12-15 min. The membrane, blotting pads and filter paper were allowed to soak in transfer buffer for 15 minutes. The membrane was slid into the transfer buffer at a 45° angle to avoid introduction of bubbles into the nitrocellulose matrix.

The proteins in the gel were then blotted onto the membrane using a Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (BioRad). The unit was assembled as follows from the platinum anode upwards: blotting pad, membrane, gel, filter paper, blotting pad.
After each layer was added, air bubbles were rolled out with a pipette. The stainless steel cathode was then placed on top. The cell was run at 20V for 1 hour.

Transfer efficiency was checked using Ponceau and Coomassie stains. The membrane was allowed to soak in Ponceau Stain (1% (w/v) Ponceau S, 5% (v/v) acetic acid) for 20 minutes and then rinsed with water to remove any unbound stain. The gel was allowed to sit in Coomassie stain (0.01% Coomassie Brilliant Blue (R-250), 20% methanol, 0.5% acetic acid) overnight at 4°C and then placed in Coomassie Destain (10% acetic acid, 20% methanol). The gel was heated in a microwave for 30s and allowed to sit for one hour. Destain was changed and the gel reheated as necessary until unbound stain had diffused out of the gel.

After Ponceau staining, the membrane was soaked overnight at 4°C in Blocking Buffer (5% milk powder in TBS plus 0.1% Tween-20). The membrane was then transferred to a plastic bag with fresh Blocking Buffer and primary antibody and incubated for 1-2 hours. The membrane was removed and washed four times in TBST (TBS plus 0.1% Tween-20): once quickly and three times for 7 minutes with shaking. The membrane was then put in a new bag with fresh blocking buffer and AP (alkaline phosphatase)-conjugated secondary antibody and incubated at room temperature for 1 hour. The membrane was removed, washed five times in TBST: once quickly and four times for 10 minutes with shaking. Secondary antibody detection was performed by a colorimetric method using a premixed BCIP/NBT solution (5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium; Sigma) or by a chemiluminescent method using PhosphaGLO Reserve (KPL) following manufacturer’s protocols.

Antibody concentrations were dependent on the particular antibody and on the
final detection method. The standard concentrations used are as follows: Mouse Anti-6x HIS mAb (IBA, Germany), 1:1000; Goat Anti-Mouse Ab (Jackson ImmunoResearch) used with BCP/NBT 1:2500; Rabbit Anti-Env pAb, 1:1000; Goat Anti-Rabbit Ab (Jackson ImmunoResearch) used with BCIP/NBT 1:2500; Goat Anti-Rabbit Ab used with PhosphaGLO Reserve (KPL) 1:40000.

Fixing proteins in gel: Gels were allowed to sit in fixing solution (50% H₂O, 40%meOH, 10% Acetic Acid) for 1 hour. The gel was then drained and Coomassie staining was carried out as above.

**Yeast Positive and Negative Controls**

Denatured lysates were obtained from eft05, eft06, eft14, and eft16, which are yeast, cultures expressing fusion proteins (Gouvas, unpublished). These cultures were previously used for yeast 2-hybrid screening. Eft05 and eft06 express proteins fused to the gal-4 active domain. Eft16 and eft14 express proteins fused to the gal-4 binding domain. The construct in eft16 expresses the gal-4 binding domain fused to a full-length copy of the sire1-4 env protein. Once activity of the polyclonal antibodies against the env protein was established (see results), eft16 was used as a positive control in subsequent studies. Eft14 is identical to eft16 cultures except the gal-4 binding domain is fused to a different protein. Eft05 and eft06 are identical to eft16 except for nutritional requirements and plasmid constructs that contain neither the binding domain nor the env gene.
**His/Strep Dual-Tagged Positive Control**

The Stargate manufacturer (IBA) provided the positive control as a destination vector (pASG-IBA44) with an insert. This plasmid coded for a dual-tagged fusion peptide with an N-terminus Strep tag and a C-terminus His tag. The peptide had a molecular weight of 15.5 kDa. The vector was identical to plasmid (pASG-IBA43, Fig. 10a and 10b) used for experimental constructs except that the locations of the tags on the expressed peptide were reversed. Heat-shock transformation into *E. coli* and induction of cultures were identical to those used above for experimental peptides. Lysate from the induction of *E. coli* containing this plasmid was used as a positive control in Western Blots using the anti-6xHis mAb for detection.

**Comparison of Protein Amounts on Western Blots**

A 1ng sample of imidazole-purified **p19** (100mM elution) was loaded onto the gel and run with soybean samples for comparison. After Western blotting using purified anti-p19 Ab and detection with PhosphaGLO Reserve and x-ray film, images of blots were scanned. Bands of interest on the scanned images were analyzed with “Image J” following included instructions. (available by ftp at zippy.nimh.nih.gov/ or http://rsb.info.nih.gov/NIH-ImageJ; developed by Wayne Rasband, National Institutes of Health, Bethesda, MD). Calculations were made with the assumption that equal band intensities approximate equal molarity.
**BLAST of Fusion Peptides to Find Epiopes:**

Amino acid sequences for fusion peptides **p19** and **p21** were BLASTed against the EST and Protein databank using the blastn search (search of a translated nucleotide database using a protein query) and the protein search respectively. In separate searches, the “Organism” limits were set to “tobacco” or “all plants” minus “legumes”. The expect threshold was raised to 100 to allow shorter fragments. Potential cross-reacting epitopes were identified in ESTs as follows: Regardless of the + or – orientation of the read, if an EST translation matched at least 5 contiguous amino acids, any associated protein sequences were evaluated to see if the molecular weight corresponded to either of the tobacco bands. If no associated sequences were available, the EST’s DNA sequence was BLASTed against the nucleotide database to trace the EST back to its genomic source. Approximately 1500 bp of genomic DNA from either side of the sourced EST sequence was taken and translated to find a likely ORF. The potential epitope was checked to make sure that it was in the potential ORF and the translated ORF was evaluated for molecular weight.
CHAPTER THREE

RESULTS

Database Searches

Since ESTs and cDNAs have been found for SIRE1 (R. Winfrey, unpublished; Laten, unpublished) and transcript has even been detected for the env-like gene specifically, the NCBI EST database was probed to see if env-like protein could be identified. Using the SIRE1-4 env-like nucleotide and amino acid sequences as queries (nucleotide accession: AY205608; protein accession: AAO73521), EST databases were searched (tBLASTn, BLASTn and tBLASTx) and revealed six ESTs with an E value below $10^{-3}$. Three of these ESTs are in the opposite orientation of the env-like gene in plants under stressed conditions – a known activator of retrotransposon transcription. The fact that ESTs are detected and that cDNA has been found previously lends some support that the env-like protein might be produced in soybean.

Antigen Production

Vector Generation

For immunological detection methods, antibodies against the env-like protein need to be generated. In order to do this, regions of the env-like gene were cloned into IBA Stargate (Germany) plasmids that express fusion proteins with two terminal
purification tags (see Materials and Methods). Attempts to express the full-length protein in this system were not successful, so smaller regions (80-120 aa) were used. Three different regions (19, 20, and 21) of the env-like gene were chosen for expression as dual-tagged fusion peptides [Fig. 7]. The regions were chosen for wide coverage across the env-like protein. This helps ensure the detection of potential cleavage products in soybean. As a putative envelope protein, the endogenously expressed env-like protein is potentially cleaved into two pieces (see Discussion). Region 21 covers the coiled-coil region, but this is the only structural feature intentionally covered. Transmembrane regions were left out of the coverage to increase the likelihood that the fusion proteins would be soluble, thus making their isolation and purification straightforward.

The IBA Stargate system uses sequential recombination reactions to insert genes of interest into two vectors (see Materials and Methods). For insertion into the first vector the three different regions of the SIRE1-4 env-like gene were amplified using one of three sets of primers (primer19, primer20, primer21). The primer sets were designed to produce three products of different sizes: 350, 300, and 250 bp, respectively. The template for the reactions was a cloned copy of the SIRE1-4 env-like gene. Electrophoresis of the PCR reactions showed three amplicons of the expected sizes [Fig. 11]. After PCR-cleanup, equal masses of each amplicon were combined for recombination into the IBA Stargate “donor” vector as described in Materials and Methods. Plasmids from this reaction were then transformed into E. coli.

After plating and an overnight incubation, colony PCR (see Materials and Methods) was performed on 16 colonies using the primer sets mentioned above to
identify colonies that carried plasmids with only single insertions of 19, 20 or 21. Eight of the colonies are shown in [Fig. 11a]. Lane 2 shows possible multiple insertions as detected by the triple bands. However lanes 3-9 show only a single band at the expected sizes of 250, 300, or 350 bp. Plasmids from five of these colonies were isolated and sequenced using a plasmid-specific primer (see Materials and Methods) to confirm a single insertion without mutations and in the correct orientation. This was confirmed for each of the plasmids. The plasmids from three of the colonies - each containing one of the inserts - were then purified and used for recombination into the pASG-IBA43 “destination” vector (see Materials and Methods), which is an expression vector. Equal amounts of the three purified plasmids were mixed together for the reaction. The reaction products were then transformed into E. coli.

After plating and overnight incubation, colony PCR was performed on 31 colonies to identify colonies containing the destination vector with a single insertion. Fig. 11b shows ten of the colonies. Each of the three expected insertions was detected. Plasmids from six candidate colonies (two colonies for each construct) were isolated and sequenced using another plasmid-specific primer (see Materials and Methods). All amplicons from sequenced colonies contained in-frame fusions and no mutations.

Expression and Purification of Fusion Peptide

Insertions in the pASG-IBA43 vector were designed to be expressed as recombinant fusion proteins with molecular weights of 15 kDa, 13kDa and 11kDa, respectively, with N-terminus tags containing six histidine amino acids (His-tag) and C-
terminus Strep-Tactin affinity tags (Strep-tag). For evaluation of the expression of the constructs, three colonies, each containing the destination vector with one of the three amplicons (19, 20, or 21), were initially chosen. The expressed, dual-tagged fusion-peptides are hereafter referred to in bold as \textbf{p19}, \textbf{p20}, and \textbf{p21}. They were named after the inserted amplicons that coded for them. Expression was induced in \textit{E. coli} (see Materials and Methods) for 2-4 hours unless otherwise stated, and cleared cell lysates were produced under denaturing conditions (see Materials and Methods).

SDS-PAGE of lysates was analyzed by Western blot probed with monoclonal anti-six-histidine antibodies (anti-His mAb) [Fig. 12]. Proteins with relatively strong signals in the size range expected were detected in lanes with lysates from \textit{E. coli} that contained constructs of \textbf{p19} or \textbf{p21}. The band for \textit{E. coli} expressing peptide \textbf{p20} was much weaker. This initial SDS-PAGE run was on a 7.5\% gel. All constructs were expected to make peptides of <20 kDa. Resolution of peptides in this range on a 7.5\% gel was poor, and precise molecular weight characterization of these peptides was not possible. Nevertheless this gel was used as a qualitative assessment for production and showed that affinity purification was an appropriate next step.

\textbf{Initial Purifications and Peptide Characterization}

Custom polyclonal antibody production by Open Biosystems required a total of approximately 2mg of the antigen peptides. To achieve maximum purity at reasonable cost, lysates were enriched using a tandem purification strategy. The N-terminal His-tag allowed an initial enrichment of the lysates with HisPur cobalt ion resin (ThermoFisher).
This resin is relatively cheap and has a large binding capacity (Litchy 2005) but is far less efficient at removing contaminating proteins than more expensive resins such as the Strep-Tactin resin which binds the Strep tag (IBA, Germany) (Litchy 2005). After the initial enrichment, the eluants from the HisPur resin were then further enriched using the Strep-Tactin resin, as described below.

Pilot runs were performed to assess the efficacy of enriching each of the proteins using the tandem His-tag/Strep-tag strategy. Since protein sizes were not determined in the initial analysis of expression, this experiment was also used to verify and characterize the protein products of the constructs.

SDS-PAGE of eluant fractions (E1, E2, E3 etc.), washes, and flowthrough (FT) of the pilot purifications were run using 18% gels. The proteins were blotted to a membrane as described, and results were analyzed by Ponceau Staining to show total transferred protein and by Western blotting using anti-His mAb [Figs. 13 and 14]. Results are described below.

**Pilot His-Tag Purification**

Fusion proteins in these pilot purifications were enriched from lysates using HisPur cobalt ion resin (ThermoFisher) under denaturing conditions, as described. As mentioned above, the expected sizes for p19, p20, and p21 were 15 kDa, 13 kDa, and 11 kDa, respectively.

The Ponceau stained membrane and the Western blot of the initial elution of the p19 purification [Fig. 13, lanes 19E1-19E3] showed a strong band at 19 kDa that grows
weaker with subsequent elutions eventually falling below the detection limits of staining protocol. The first elution of the p20 showed a very weak band around 16 kDa, close to the expected 13 kDa, in lane 20E1 of the Western blot [Fig. 13b], but it was undetectable in the subsequent elution [lane 20E3] or in either lane on the Ponceau stain [Fig. 13a], suggesting poor expression. Bands were detected at 9 kDa, close to the expected 11 kDa, in eluants of the p21 peptide [lanes 21E1 and 21E2] on both the Western blot [Fig. 13b] and Ponceau stained membrane [Fig. 13a], though the stained band in 21E2 was extremely weak. The positive control supplied by IBA was detected at about 15 kDa [Fig. 13b] as expected.

A number of weaker, secondary bands were also detected on the Western blots. These were below the detection limits of the Ponceau staining. In eluant lanes of p21 [Fig. 13b, lanes 21E1 and 21E2] a secondary band was visible at 22 kDa, but is not detected in the flowthrough. A secondary band also appears at 31 kDa in the lane with the positive control, a non-purified cell lysate [Fig. 13b]. In both of these instances the protein is about two times the size of the protein of the primary band. Two secondary bands also appear in the first p19 elution [Fig. 13a and 13b, lane 19E1]. One faint band runs approximately 1 kDa above the primary band and is in large enough quantities to be detected by both Ponceau Staining [Fig. 13a] and Western blotting [Fig. 13b; for a stronger signal see Fig. 14a and 14b]. The other protein runs below the primary band at approximately 10 kDa [Fig. 13a (weak) and 13b]. Neither secondary band is seen in subsequent eluants of p19. The secondary bands might have been E. coli proteins that for
unknown reasons co-eluted with tagged proteins, but both are detectable by the anti-His mAb.

No His-tagged protein was detected in any of the flowthrough lanes for \textbf{p19}, \textbf{p20}, or \textbf{p21} on the Western blots [Fig. 13b, lanes 19FT, 20FT, and 21FT]. With the exception of the secondary bands of \textbf{p19}, which were detectable by anti-His mAb, no contaminating bands were observed in the eluant lanes in amounts large enough to be detected by Ponceau staining [Fig. 13a]. Primary and secondary protein sizes are evaluated below.

**Strep-Tag Purification**

As stated, the initial His-purifications were done under denaturing conditions, which are unsuitable for enrichment on the Strep-Tactin resin. For Strep-tag purifications, eluants from the His-purifications were diluted greater than 10-fold so that the concentration of urea was within the tolerances recommended by the supplier of the Strep-Tactin resin (IBA, Germany). Nevertheless, the pilot Strep-tag purifications resulted in no detectable protein in eluants or even in the diluted flowthrough [data not shown]. The dilution might have brought protein concentration in the flowthrough below detection limits, but the lack of enriched protein in eluant fractions suggests that the diluted eluants from denatured His-purifications were incompatible with the Strep purification or that all proteins irreversibly bound to the resin.

In order to address possible compatibility issues with the Strep-Tactin resin, fusion peptides were re-expressed and lysates were prepared under native conditions, as described. These lysates were then purified on HisPur (ThermoFisher) cobalt-resin
columns under native conditions, as described. The eluants were purified on a column containing Strep-tactin resin, as described. In an attempt to increase the amount of peptide p20 to be purified, the expression protocol was scaled-up 5-fold.

Due to the specificity of the Strep-tactin resin, which is derived from a biotin-streptavidin interaction (IBA), only the Strep-tagged peptides were expected to remain bound prior to the elution step. The fusion peptides were expected to be the 19 kDa, 16 kDa and 9 kDa products seen in the His-tag purifications. The proteins that generated secondary bands were expected to elute in the flowthrough.

After SDS-PAGE of the Strep-tag purifications, the proteins were blotted to a membrane and analyzed by Ponceau staining and by Western blotting [Fig. 14a and 14b]. Note that the standards used on this gel were added 5 minutes late, due to user error. As a result, bands appeared to be 1-2 kDa smaller than the corresponding bands in Fig. 13a and b. All but one of the primary and secondary proteins detected and described above for the His-tag purification were maintained through the Strep-tag purification. The exception was the primary band for p20, which was not observed in Western blot or Ponceau staining for the Strep-tag purification (data not shown). The Ponceau stain showed four detectable bands between 7-12 kDa in the flowthrough of p19 [Fig. 14a]. Of the four proteins, only the larger two, which ran near 12 kDa, were detected on the Western blot [Fig. 14b]. All four of these smaller proteins may be degradation products. Protein degradation is a common problem and should not affect their effectiveness as an antigen source (see Discussion). Interestingly two new secondary bands were detected by
Western blot in the 19E2-3 lane at and just above 38 kDa [Fig. 14b]. This is approximately twice the size of the two bands detected near 19 kDa.

In general, after the tandem enrichments, the strongest band in each eluant lane was found at approximately the expected size. Based on their amino acid sequence, the predicted sizes for p19, p20, and p21 should have been 15 kDa, 13 kDa, and 11 kDa, respectively. However, in the eluant lanes the bands showed apparent molecular weights of 19 kDa, 16 kDa, and 9.0 kDa, respectively, as measured on the SDS-PAGE run of the His-tag purification [Figs. 13a, 13b]. Minor differences of the predicted weight of a protein versus its behavior in SDS-PAGE are common (Rath et al, 2009). Interestingly, all secondary bands detected in eluants from the His-tag purification still remained after purifications with the Strep-tactin resin [Figs. 14a and 14b], and most can be detected by anti-His mAb [Fig. 14b]. It should also be noted that for three of the dual-tagged proteins - p19, p21, and the IBA-supplied positive control - there was a secondary protein that ran at approximately double the molecular weight of the primary band [Fig. 14b]. These bands might correspond to homodimer (see Discussion).

Scaled-up Purification Protocols and Imidazole Gradient His-tag Purification

Having verified that appropriate sized proteins were produced and could be at least partially purified using non-denaturing conditions, the protocols were scaled up another two fold (ten fold over the initial pilot purifications). Expression was induced in cultures from two separate clones carrying the p20 construct, and the proteins were enriched using non-denaturing His-purification protocols. Nevertheless no bands were
seen in Western blots (data not shown). To determine if expression of this construct required longer induction or if enrichment of p20 was more compatible with denatured extractions, an overnight induction was coupled with denatured extraction and a denatured protocol for His-tag purification (see Materials and Methods). Since the denatured eluant was not compatible with the Strep-tactin purification, the eluants from the purification were dialyzed to remove the urea and purified using Strep-tactin resin. However, no peptide was detected in the eluants or the flowthrough (data not shown) suggesting that peptide p20 was very unstable, remained irreversibly bound to the resin, or perhaps might have precipitated out in the dialysis. Though no precipitate was noticed upon casual inspection of the dialyzed proteins, precipitate would have been expected in very small amounts, which might have gotten lost in the dialysis tubing or been overlooked. One last attempt was made to purify peptide p20 using only the HisPur resin (ThermoFisher) under denaturing conditions. Using an imidazole gradient, as described in Materials and Methods, contaminants should elute from a cobalt column before any His-tagged peptide. On Western blots, only the positive control protein was detected and no bands were detected in the gradient-purified lysates (data not shown). These inconsistent results for p20 indicate that production of peptide p20 may have been highly variable, that in some cases p20 may not have bound to the cobalt column, or that p20 may have eluted at extremely low imidazole concentrations (below 40 mM which is the equivalent of a very stringent wash). Due to these results, purification of peptide p20 was not further pursued.
For peptides p19 and p21, results of the scaled-up protocol used with the Strep-tag purification suggested that much of the His-tagged protein could have been lost in the flowthrough fraction and/or in washes of the Strep-tag purification (see Fig. 14b, lanes 19FT, 19W1). To prevent this loss but retain purity, large-scale His-purification of p21 was attempted using an imidazole gradient under native conditions, as described in Materials and Methods. Stringent washes were done at three and five times the recommended concentration of imidazole preceding elution. For the His-tag gradient purification of p21, eluted proteins separated by SDS-PAGE were blotted to a membrane [Fig. 15a]. Ponceau staining of the membrane showed that the majority of the peptide eluted off of the column at imidazole concentrations between 75 mM-130 mM. The peptide appeared to be very pure except for a smaller band under 8.5 kDa [Fig. 15a]. This lower band was detected in separate Strep-tagged purifications of p21 [Figs. 15b and 17a, lanes 8 and 9]. Western blot analysis using anti-His mAb could faintly detect the smaller protein in elution fractions (data not shown). To compare relative purity of the gradient-purified samples, an elution fraction of the gradient-purified p21 was compared to a Strep-tag purified elution by Coomassie staining of the SDS-PAGE-separated proteins [Fig. 15b]. The Coomassie stain of the gradient-purified p21 showed a second faint band below the primary band not detected by the Ponceau staining. However this protein was also detected in the Strep-tag purification [Fig. 15b]. Overall, the gradient-purified protein seemed to be of a comparable purity to a tandem-purified protein.

Surprisingly, in one His-purified sample (not gradient purified) of p19, three new bands were detected with Ponceau staining [Fig. 16a]. No size standards were run on the
gel, but based on comparisons with other runs the three bands were estimated to be between 24 kDa and 31 kDa. To remove these contaminating proteins, eluants were re-purified using a gradient purification. Eluants were first dialyzed to remove imidazole and then added again to HisPur (ThermoFisher) resin. The resin was extensively washed and proteins were eluted by a gradient purification as above. Proteins were then separated by SDS-PAGE, fixed in the gel and Coomassie stained [Fig. 16b]. After gradient purification a weak band was still detected at 24 kDa in the 50 mM elution, but the other proteins were not present. None of the contaminating proteins was detected in subsequent elutions.

**Preparation of Antigen for Antibody Production**

To prepare the fusion peptides for antibody production, concentrations of elutions of all purifications were determined, and purified protein was pooled to amass 1.89 mg of total protein at a concentration of 210 µg/mL. This was within the amount and concentration of peptide needed for the protocol. The sample contained approximately equal masses of p19 and p21. Though the pooled samples mostly contained Strep-tag purifications, His-tag gradient purifications (without further Strep-tag purification) were also used (GradPur19 50 mM and 80 mM; GradPur21-2 100 mM and 130 mM).

Purities of the samples pooled together were estimated visually by Coomassie or Ponceau staining. The Coomassie stained gel of gradient-purified p19 is discussed above [Fig. 16b]. The Coomassie stain of gradient purified p21 and Strep-tag purified p21 show no unexpected bands [Fig. 15b]. In addition, all of the eluants that were pooled together
were run on SDS-PAGE and blotted to a membrane. The purity of was checked by Ponceau staining [Fig. 17a]. Primary bands of expected sizes were detected in all lanes. No other bands besides those of suspected breakdown products were detected. In lane 2 of the Ponceau-stained membrane [Fig. 17a] a breakdown product of a different size was detected when compared to neighboring lanes with purified peptide p19. This was unusual since this lane is part of a serial dilution [Fig. 17a, lanes 2-4]. A band of similar size was detected in lanes 3 and 4. Oddly, the band in these lanes was much weaker though the samples were more concentrated. This band was also similar in size to peptide p21 [seen in Fig. 17, lanes 8-12] and could represent a contamination of the lane by peptide p21. Overall, the Ponceau and Coomassie stains affirm the purity of the samples.

The purified and pooled peptides were then sent to Open Biosystems for a 10-week, 2-rabbit protocol. Antisera for both rabbits reached the maximum titer of 204800 on the Manufacturer’s scale. (See Materials and Methods.)

Efficiency and Purity of the Resin Purification Protocols

Two resins (Cobalt HisPur resin and StrepTactin resin) and two separate strategies (Tandem His-tag/Strep-tag purifications and Gradient His-tag purification) were used in this study to enrich peptides. In order to streamline the protocol for future experiments and to assess cost/benefit considerations, it was of interest to assess the efficiencies of the resins and the strategies in terms of peptide recovery and purity.

After expression and His-tag purifications, there was no visible western signal in the flowthrough lane (e.g. Fig. 13), which implies that the HisPur column is efficient at
binding the peptide given the amount of protein added to the column. Analysis of protein concentration of the gradient purifications of p19 and p21 showed that for a 1L starting culture 2-4 mg of peptide was recovered on 0.75 ml of HisPur resin. The column has a manufacturer-reported a capacity of 10 mg/ml of resin (7.5 mg/0.75 ml). Since only $\frac{1}{4}$-\(\frac{1}{2}\) of the capacity of the resin was recovered, the protocol could probably be adjusted to a longer induction time, a larger (possibly 2-3L) starting culture, or some combination of the two to take full advantage of the binding capacity of the resin. Longer induction times might be a better place to start since larger cultures also mean larger amounts of possible contaminating proteins.

Purification by StrepTactin resin resulted in noticeably less peptide. Though concentration often remained high in the first eluant fractions, the size of each fraction was diminished by at least four-fold (500 µl versus 2 ml). The manufacturer reported the resin capacity at 50 to 100 nmol/ml of resin or up to 3 mg/ml for a 30 kDa protein. Peptide p19 was about 15 kDa (theoretical weight), so the maximum expected capacity was between 750 µg-1.5 mg. However the average amount of protein eluted was closer to 250-300 µg for each of the peptides p19 and p21. Since native purifications were used, the cause of this discrepancy was unknown. It is possible that it merely represents differences in binding from protein to protein. It was clear though that a fraction of the peptide p19 was eluting in the initial washes. Later washes were not tested. Still it might be possible that less extensive washing could retain some of the p19 on the resin. Other possible remedies for fusion-protein loss would be to add the flowthrough back to the column one extra time, or perhaps incubating the samples with the resin for an hour (also
known as the batch method) before packing the resin into the column for washes and elution. Both might increase yields.

Except for the secondary bands (discussed below), there was very little evidence of contaminating proteins in any of the final elution fractions, which included both His-tag gradient purifications and Strep-tag purifications. After analysis using both Ponceau staining and Coomassie staining techniques [Figs. 15, 16b, and 17], it would appear that most of the His-gradient and Strep-tag purifications are equally pure. It might be possible to use only elution fractions from His-tag gradient purifications in the future. This is important since Strep-tactin resin is relatively expensive.

Though most of the scaled-up protocols and purifications of p19 and p21 gave no evidence of contamination as assessed by Ponceau staining, the one case of visible contamination was found in a single His-tag purification of p19. [Fig. 16a] The resin was washed extensively. The cause of the contamination was not clear. A second His-tag purification of the eluants using an imidazole gradient removed most of the contamination. Coomassie staining of the SDS-Page of elution fractions showed only one faintly visible contaminating protein left in the 50mM (imidazole) elution fraction [Fig. 16b]. This protein does not show in a blotted Ponceau stain of the same elution fraction [Fig. 17]. Since the average detection limit of Coomassie Staining is around 50ng of proteins (Winkler et al, 2007) and for Ponceau Staining is 250ng (Salinovich and Montelaro, 1986), a high-end estimate of the contaminating protein might be around 200ng of protein seen in the gel. Taken into consideration the amount of this sample used
in the pooled peptides, this would translate to <0.09% of the 1.89 mg of protein sent for antibody production.

Other elution fractions from imidazole-gradient His-tag purifications were also pooled for antibody production. This included the subsequent elution fraction (80mM) of the gradient-purification of p19, but the Coomassie stained gel showed no sign of contamination [Fig. 16b, lane 5]. Gradient His-tag purifications for p21 were also used. The purity of the fraction eluted with the lowest concentration of imidazole (100mM) showed comparable purity to a Strep-tag purified elution [Fig. 15b]. The balance of the pooled elutions was Strep-purified and contaminants were below the level of detection by Ponceau staining [Fig. 17].

**Experiments with Polyclonal Antibodies to p19 and p21**

Specificity of Crude Antisera for Purified Peptides and Env-like Proteins Produced in Eukaryotes

Assays to verify that the antibodies would bind both to the fusion peptides used as the antigen and to an env-like protein without His or Strep tags were run. Crude antisera from the day 58 blood draws were tested on the E. coli-produced, purified, pooled peptide and on a yeast protein prep from a strain expressing a full length SIRE1-4 env-like fusion protein (E. Gouvas, unpublished) [Fig. 18]. Equal volumes of the antisera from both rabbits were combined. SDS-PAGE on an 18% gel was used to achieve resolution of the smaller peptides.
Overall, binding of the crude antisera was strong. The crude antisera from each rabbit reached the maximum titer as detected by OpenBiosystems showing that the antigen produced a strong immunogenic reaction in the rabbits. As expected the binding of antibodies to the pooled-fusion peptide gave a strong signal at the recommended 1:500 starting dilution for the antibody [Fig. 18b, lane 11]. The ratio of the Ponceau stained signal [Fig. 18a] to the Western blot signal [Fig. 18b] was far more intense in blots that were probed with the antisera compared to those probed with commercial anti-His mAb [e.g. Fig. 14a and 14b].

Lane 11 of Fig. 18b shows a Western blot of the pooled peptides. Many proteins were detected, but most bands correspond to primary and secondary bands of p19 and p21 [Compare to Fig. 14b, Fig. 17a], as would be expected. The four most intense bands on the Western blot, corresponding to black arrows, were also visible on the Ponceau stained membrane. The two proteins with the most intense bands on the Ponceau stained membrane corresponded to the primary bands of p19 and p21. Other proteins were detected on the Western blot [Fig. 18b] just above and below the marked bands from p21, but they corresponded to the likely degradation products of p19 and p21 already discussed above [p19: Fig. 14, p21: Fig. 17a]. The dimer-sized proteins, indicated by red arrows, were also detected [Fig. 14b]. The protein detected just under 24 kDa corresponded to the dimer-sized protein detected after the Strep-tag purifications of p21 [Fig. 14b]. The two proteins detected around 38 kDa corresponded to the dimer-sized proteins of the 19 kDa and 20 kDa peptides in purifications of p19. There was also very faint band about 5 kDa below the 19 kDa band of p19. Interestingly, no protein had been
detected at this level by any method in any of the purifications, but the band however was extremely faint and was not further considered. The general background that was seen in the lane on the Western blot was probably due to overloading of the sample. Overall, this lane shows that the crude antisera reacts strongly to the peptide antigen, that few proteins were present in the antigen other than those already known.

A double band was detected on the Western blot in decreasing amounts across the membrane [Fig. 18, lanes 3-11]. It was likely due to a loading error since the bands were at the same level as bands for \textbf{p19} and were strongest in the blank and ladder lanes adjacent to the \textbf{p19/p21} peptide sample.

Lanes 6-7 [Fig. 18] contained a protein extract from yeast strain Eft16, which contains a plasmid expressing a full-length SIRE1-4 env-like fusion protein. The protein is referred to as \textbf{eft16}. The env-like portion was fused to a Gal4 protein for use in a yeast two-hybrid screen. (E. Gouvas, unpublished) Since there were no Strep or HIS tags on \textbf{eft16}, any antibody binding should have been to env-like epitopes. The yeast negative controls were protein extracts from strains Eft05, Eft06 and Eft14, containing plasmids that expressed other fusion proteins used in the same study.

For the yeast protein samples, crude antisera bound proteins only in the lysate expressing \textbf{eft16} [Fig. 18b, lane 6-7]. Three to four bands were detected at about 50 kDa, 59 kDa, and 67 kDa with a weak band between the latter two. This was more bands than expected, but they occurred only in yeast expressing \textbf{eft16}, and were most probably degradation or cleavage products of the env-like protein. The two most intense bands were detected at 59 kDa and 67 kDa. Both bands are smaller than other size estimations
from previous studies in yeast (Gouvas, personal communication) and in subsequent experiments in this study. Either the yeast proteins or the High MW standards seem to have run aberrantly for unknown reasons. Typically the two intense bands of \textbf{eft16} ran at approximately 94 kDa and 116 kDa [e.g. Fig. 23]. The 116 kDa band matches previous runs using this protein (Gouvas, personal communication). Despite the questionable molecular weights of the yeast proteins, it was clear that the crude antisera binds strongly to both the \textit{E. coli}-produced antigen source and to proteins in Eft16 lysates containing, a yeast-produced full-length env-like protein from SIRE1-4. Based on these results, it appeared likely that the antisera would be able to detect SIRE1 env-like protein and that experiments for its detection in soybean could begin.

Tests with Crude Antisera on Soybean

\textbf{Test of Crude Pre-immune Antisera on Soybean and Tobacco}

Antisera recovered from immunized rabbits contains not only polyclonal antibodies against injected antigens, but endogenous antibodies present in the pre-immunized animals. Binding of non-target plant proteins is a known problem with antibodies raised in rabbits for use in detection of proteins from higher plants (see Discussion). To evaluate the extent to which such antibodies might bind to the proteins in this study, a Western blot was run using pre-immune antisera from both rabbits. Fig. 19 shows that antisera from both pre-immune rabbits detected a 55 kDa protein in soybean and tobacco leaves. The 55 kDa band is consistent with the size of the rubisco large subunits of both soybean and tobacco (Parry, 1987). Rubisco is an enzyme used for
photosynthesis and would be expected in leaf tissues, and can comprise up to 50% of the protein in a soybean leaf (Krishnan, 2009). Antibodies to this protein appear to be present in the pre-bleed antisera [Fig. 19].

The crude pre-immune antisera also bound several other proteins in soybean, soybean tissue culture and tobacco including a ~57 kDa and a weak ~98 kDa in the tissue culture. Notably there were no sharp bands in Eft16 lysates, which contained a full-length env-like protein, nor were there sharp bands the 65 kDa position in the tissue culture lane. This corresponded to one of the env-like candidate bands seen with day58 antisera (below). These results strongly suggest that these rabbits possess antibodies against plant proteins found in their food. The impact that these detected antigens had on the interpretation of the experimental data are discussed below.

Tests of Stressed Samples

Despite the repression of retrotransposon activity in plants, treatments of plant tissues with various chemical and physical stresses have been shown to raise retrotransposon activity (reviewed in Grandbastien, 2010 and Mansour, 2007). It was not known what conditions might activate SIRE1, so soybean plants and tissues were subjected to various stresses and other treatments that have been reported to activate a variety retrotransposons in other plants (see Materials and Methods). Protein extracts from treated soybean plants and tissues were separated by SDS-PAGE, and analyzed by Ponceau Staining and Western blotting using the combined crude antisera [Fig. 20].
Despite the binding of the pre-immune antisera to plant proteins, it was hoped that env-like proteins could still be detected.

The Western blot [Fig. 20b] showed a few intense bands, some of which might be products of the env-like gene, but it also revealed several other bands of different sizes and intensities. As shown above [Fig. 19] at least some of these bands were due to contaminating pre-immune antibodies binding to proteins that were not targeted env-like proteins. In particular, the corresponding Ponceau stain [Fig. 20a] showed one heavy band in particular at 55 kDa in all of the extractions of leaf tissues. In the Western blot moderately intense bands that correspond to Ponceau bands were detected in lanes 5 and 9 with weaker bands in lanes 1, 3, 4 and 7 (lanes 5 and 9: untreated soybean and 1:5 dilution of wound treatment; lanes 1,3,4 and 7: undiluted wound, untreated tobacco, cellulase treated, and water-stressed, respectively). After longer exposures of the Western blot to film (over two minutes), a 55 kDa band was clearly present in all lanes except lanes 2, 6, and 11. Lanes without the 55 kDa band corresponded to samples from root and tissue-cultured cells, both of which had minimal if any light exposure. Again, this band was most likely the rubisco protein.

Moderately strong bands were detected however in proteins from tissue culture [Fig. 20, lane 2] at 51 kDa, 57 kDa, and a double band at 98 kDa. Only very weak bands were visible in the corresponding lane of the Ponceau stain. This implies that the proteins in these bands had strong antibody binding. Weak bands were also detected on the Western blot at 88 kDa, 60 kDa, 45 kDa, 41 kDa. Upon longer exposure, an apparent bubble appeared between 51 kDa and 57 kDa that might have blocked the transfer of
other proteins. There was also a band at 65 kDa roots from water-stressed plants (lane 6). A protein of the same size was visible in protein samples from the wound treatment (lane 1) and, upon longer exposure, in the diluted wound sample (lane 9).

The paired bands in the tissue culture lane (lane 2) at 98 kDa ran higher than the expected size of the SIRE1-4 full-length env-like protein (74 kDa), but post-translational modifications were predicted (Laten, 1998). Also with >1100 SIRE elements in the soybean genome, it was very possible that these proteins were products of elements that encode larger env-like proteins than SIRE1-4. However, one of the paired 98 kDa bands might correspond to a weak, approximately 98 kDa band in the pre-immune sera [Fig. 19; Rabbit 2, lane 8]. The 57 kDa protein in tissue culture, likewise, might correspond to the band at 57 kDa seen on the pre-immune blots [Fig. 19; Rabbit 2, lane 8]. Other smaller proteins in the tissue culture sample (lane 2) might have been cleavage or degradation products (see Discussion). The apparent difference in binding of the antibody towards the Eft16 lysates between Figs. 18 and 20 can be explained by the difference in sensitivity of the two detection methods, NBT/BCIP and PhosphaGLO Reserve (KPL).

Overall, though some variation in the size of the env-like protein would be expected, it seemed apparent that the antibodies were binding proteins other than the env-like protein. However, many if not all of these were most likely due to anti-plant protein antibodies present in the rabbits prior to inoculation.
Assessing Antisera Binding to Plant Proteins in Soybean (Troubleshooting)

The antibody production protocol (Open Biosystems) used two rabbits, and crude antisera was recovered before antigen injection and on post-injection days 35 and 58. Because the pooled antisera bound to numerous plant proteins and binding to a env-like protein could not be determined, Western blots using different antisera were compared to determine the extent of antisera binding to non-target proteins and to determine if any of the antisera could be used without further enrichment steps. Antisera from the two rabbits (Day 58: Rabbit 1 and Rabbit 2) and from two different time-points (Rabbit 2: Day 58 and Day 35) were used in separate blots loaded with the same set of samples [Fig. 21]. As expected, Day 58 bleeds from Rabbit 2 gave stronger signals than bleeds from Day 35 from the same rabbit. However, the two Day 58 bleeds gave markedly different results. Rabbit 2 had an overall stronger blot implying a stronger immune reaction - both to eft16, and to the plant proteins. No molecular weight standards were run with this blot, but bands were estimated by comparing them to bands in Eft16 lysates (the largest of which are at ~116 kDa, and ~95 kDa) and the presumed rubisco bands seen in the tobacco and untreated soybean lanes (~55 kDa). Many of the sharper bands do not agree between the rabbits. However one band that was prominent and present in both rabbits was the band at 98 kDa in the tissue culture lanes, although only a single band was visible in Rabbit 1 and a double band in Rabbit 2. This shows an added band at ~98 kDa for Rabbit 2 when compared to the pre-immune [Fig. 19]. Other bands were detected in the tissue culture lane of Rabbit 1 at 51 kDa and possibly a double band at ~57-60 kDa. These were present in Rabbit 2, but they appeared much weaker.
Antibody Purifications and Further Tests in Soybean

Though candidate bands had been identified, it was obvious that binding to non-targeted proteins would hinder interpretation of the results. In an attempt to minimize the detection of the non-targeted proteins, the antibodies were affinity-purified. Though stronger, non-target bands were detected with Rabbit 2 day58 antisera, this antisera also gave the strongest signals against proteins believed to be env-like. In hope that the purification would eliminate pre-immune antibodies and possibly cross-reacting species, Rabbit 2 day58 antisera was used for antibody purification in an attempt to optimize for the strongest signal possible.

Peptide antigen samples, \textbf{p19} and \textbf{p21}, were separately immobilized on a nitrocellulose membrane and incubated with crude antisera. The membranes were then washed, and the bound antibody was eluted as described. Westerns using these antibodies as probes showed that the purified anti-p21 Ab did not bind to \textbf{eft16} (data not shown), so it was not used for final Western blots. In Western blots with purified anti-p19 Ab, prominent bands in tobacco were detected at ~75 kDa and ~150 kDa, (the 75 kDa band can be seen in Fig. 22), but the presence of one or the other was variable and inconsistent from one Western blot to another.

To see if there were any possible short peptide sequences common to both the fusion peptides (\textbf{p19} and \textbf{p21}) and known tobacco proteins, both fusion peptide sequences were BLASTed against protein and translated EST sequence databases (see Materials and Methods and Discussion). These searches were extended to all other plants excluding plants in the fabaceae (legume) family. The Expect threshold (E value) was
increased to 100 to allow hits of smaller lengths, albeit of questionable significant value. Though epitope-sized hits (6-8 amino acids) were found, they were in proteins or ORFs with substantially different predicted molecular weights than the ones detected on the Western blots. No likely proteins candidates were found. With more extensive washes using 0.3% Tween-20, neither tobacco band was detected.

With extensive and stringent wash conditions and purified anti-p19 Ab, three tissue culture proteins at 98 kDa, 65 kDa, and 57 kDa were still easily detected [Fig. 23b, lane 5]. Faint bands were visible at 52 kDa and 42 kDa. Close comparison with the Ponceau stain [Fig. 23c] showed that the 65 kDa band was barely visible on the Ponceau stain which implies that the binding was robust and not solely due to high amounts of protein coupled with either weak, non-specific interactions or specific binding of low amounts of contaminating antibodies. Comparison of the Ponceau stain and the Western blot for the 98 kDa and the 57 kDa bands was complicated since there were more stained protein bands in those regions. However the stained bands close to 98 kDa are fairly weak.

Though three strong bands and some minor bands were consistently detected by Western blots in tissue culture samples, interestingly, the intensities of the bands changed with respect to each other while using the purified antibodies. In the initial Western blot with crude antisera, relatively strong bands were observed at 51 kDa and 57 kDa, a weak band at 65 kDa and a double band at ~95-98 kDa. The 51 kDa band was the most prominent [Fig. 20b, lane 2]. When tissue culture samples were probed with combined purified antibody, anti-p19 Ab, the band patterns adjusted to 57 kDa, 65 kDa and a single
band at ~98 kDa with the 65 kDa band being the most prominent [Fig. 22]. It is possible that the weakened and lost bands were false positives and were weakened due to loss of antibodies that were targeting proteins other than env-like. However with the loss anti-p21 antibodies the possibility that these are SIRE1 env-like proteins can not be ruled out (see discussion).

To estimate the amount of putative env-like protein in the 65 kDa band, the intensity of the band from a 1 ng load of gradient-purified peptide p19 was compared to that of the 65 kDa protein of the tissue culture sample (data not shown; see Materials and Methods). Both bands were from the same Western blot probed with the purified anti-p19 Ab. The band from the 65 kDa protein was estimated to have been ~30% of the intensity and therefore to have contained ~1 ng of protein per load. This constitutes >280 ng of the 65 kDa protein per 500 mg of tissue.

With the extensive washes, the bands visible in Fig. 20 in wound treatment samples and water-stressed root tissues were not observed [Fig. 23b, lanes 2 and 13]. Western blots from extractions of stress-treated leaves, including the dipped cellulase, salicylic acid, CuCl₂, and UV treatments, showed no bands under these assay conditions (data not shown).

Transcription of retrotransposons has also been seen in key developmental stages and tissues. To test the developmental activation of SIRE1 and the production of env-like, proteins from embryos and from seedling roots and shoots were examined. In lanes with embryo and seedling preparations [Fig. 24, lanes 2, 3, 6, and 7], bands that match the size of the tissue culture bands were visible at about 57 kDa, 65 kDa and 98 kDa.
Lanes 3 of Fig. 24 (long seedling shoot) also share the very weak ~47 kDa band with the tissue culture lane. New bands were detected at ~130 kDa and 80 kDa in seedling tissues [Fig. 24, lanes 3, 6, and 7 - corresponding to long seedling shoot, short seedling shoot and short seedling, respectively; summary of bands in Table 3.] None of the proteins detected in the embryonic tissue were detected in any of the leaf tissue samples [Fig. 24 lane 7, Fig. 23b lanes 1-4 and 9-11].

The rubisco band was still visible in many lanes of Fig. 23, though only faintly. In the tobacco lane it is more distinct due to an aberration in the run. Though minor contamination might have still been present from pre-immune antibodies, the band at 65 kDa seems to be a clear indication that the SIRE1-4 env-like protein was likely induced under tissue culture conditions.
Fig. 11  Size verification of pre-insertion amplicons and plasmid inserts

(a) Gel electrophoresis of pre-insertion PCR products used for recombination into the Donor vector. Lane 1 contains ladder. Lanes 3-4, 5-6, and 7-8 contain amplicons. using primer sets 19,20, and 21 respectively. Amplicons for each set run at their expected sizes of 350, 300, and 250.

(b and c) Gels of “colony PCR” products of isolated colonies. Template used in the reactions were the indicated vectors isolated from E. Coli. The reactions used primer sets 19, 20, and 21 combined together which. (b) Donor vector verification: lanes 1 and 10 contain ladder. Lanes 6 and 11 likely have at least two insertions as indicated by their bands at 250, 300 and a combined band at 550. In all other lanes, only one insertion is detected at 250bp or 300bp except 17 in which a single insertion at 350 can be detected. (c) Destination Vector insertion: lanes 1 and 10 contain ladder. Only a single band at 250, 300, or 350 can be detected in all lanes.
Fig. 12  Initial induction for the expression of fusion peptides

Ponceau stained membrane (a) and anti-His probed Western blot (b) of the blotted SDS-PAGE of cleared lysates from uninduced and induced *E. coli* cultures containing plasmids that encode the indicated peptides. A “smile” occurred on the gel causing the dye front to be higher in induced lanes. The proteins were run on a 7.5% gel causing proteins below 31 kDa to run at or near the dye front. Induced peptides are all expected to be below 20 kDa. Expression can be seen in induced cultures though the signal for 20 is significantly weaker.
Fig. 13  Characterization of peptides after His-tag Purification

a. Ponceau

![Ponceau stained membrane](image)

b. Western (anti-His)

![Western blot (anti-His)](image)

Ponceau stained membrane (a) and anti-His probed Western blot (b) of the blotted SDS-PAGE of cleared lysates from induced *E. coli* cultures containing plasmids that code for indicated peptides. Lysates were added to HisPur (ThermoFisher) cobalt columns for purification of His-tagged proteins. The Western blots was probed with anti-His mAb and detected by NBT/BCIP. Lanes are labeled as follows: Eluant fractions - first, second, etc. (E1, E2, etc.), Flowthrough (unbound protein) fraction (FT), First Wash (W1). The positive control was a cleared lysate from *E.coli* expressing a His/Strep dual-tagged peptide (see materials and methods). Standards are in Daltons. The standards used in SDS-PAGE of the Strep-tag purifications were added 5 minutes late, so peptides appeared 1-2kDa smaller than in the His-tag purification. Black arrows point to secondary bands of approximately dimer size on the Western blots. The red arrow point to secondary bands below the primary band on Western blots.
Fig. 14  Characterization of peptides after tandem His-tag/Strep-tag Purifications

Ponceau stained membrane (a) and anti-His probed Western blot (b) of the blotted SDS-PAGE of proteins from induced *E. coli* cultures that express the indicated peptide. Proteins were sequentially purified by their His-tag followed by their Strep-tag. Western Blots were probed with anti-6xHis mAb and detected by NBT/BCIP. Strep-tag purification of 20 showed no bands and is not included. Lanes are labeled as follows: Eluant fractions - first, second, etc. (E1, E2, etc.), Flowthrough (unbound protein) fraction (FT), First Wash (W1). The positive control was cleared lysate from E.coli expressing a His/Strep dual-tagged peptide (see materials and methods). Standards are in Daltons. The standards used in SDS-PAGE were added 5 minutes late, so peptides appeared 1-2kDa smaller. Black arrows point to secondary bands of approximately dimer size on the Western blots. The red arrows point to secondary bands below the primary band on Western blots. (c) Detail of Western blot in b. showing double banding at dimer level. (d) Close-up of a Western blot clearly showing two bands below the primary band in FT of Strep purifications of 19.
Fig. 15  Purification of peptide 21 by imidazole Gradient

(a) Ponceau stained membrane of the blotted SDS-PAGE of products of a His-tagged gradient purification. Fractions were eluted with increasing concentrations of imidazole between 50mM and 250mM. Imidazole concentrations (mM) for each eluted fraction are beneath each lane. Two different purifications were performed with lysates of induced *E. coli* containing plasmid coding for a His/Strep tagged-21. They are labeled GradPur21-1 and GradPur21-2. Standards are in Daltons. FT stands for flowthrough. Wash1 is the first wash after lysates have drained from the resin. Lanes with fractions pooled for analysis are labeled with concentrations separated by a slash (e.g. 130/170). Lane 250/250 represents two separate 250mM elutions. (b) Coomassie stained gel of an SDS-PAGE comparing eluants from a His-tag gradient purification with a Strep-tag purification.
Fig. 16  Re-purification of peptide 19 by imidazole Gradient

(a) Ponceau stained membrane of eluants from a His-Purified peptide 19 showing high contamination. Black arrows point to suspected contaminating bands. (b) Coomassie stained gel of contaminated eluants re-purified on HisPur by gradient purification of imidazole. Fractions were eluted with increasing concentrations of imidazole between 50mM and 250mM. Elutions were fixed in the gel and stained by Coomassie stain. Standards are in Daltons. Elution fractions are labeled by the concentration of imidazole (mM) in the buffer. Lanes with fractions pooled for analysis have two concentrations separated by a slash (e.g. 130/170). 250-1 and 250-2 (lanes 2-3) represent two separate 250mM elutions. All eluants were run by SDS-PAGE on 18% gels.
Fig. 17  Estimates of peptide purity preceding antibody production

Ponceau stained membrane of purified peptide that was to be pooled as the antigen for antibody production. Eluants were run using SDS-PAGE on 18% gels. The peptide expected in each sample is identified by the large 19 or 21 at bottom of the figure. Dilutions of the same sample across multiple lanes are indicated by a wedge. The point indicates the lowest concentration, and the mouth indicates the highest. His samples are single elutions from gradient purifications. The number following its label is its elution fraction and correspond to samples shown in Figs. 11 and 12. Strep samples are comprised of pools of Strep-purified product. The number following the Strep label indicates that they are different pools and has no other significance. Lanes 2-4 are serial dilutions of a mixture of products from a single fraction of a His gradient purification and a pool Strep purifications.
Fig. 18  Verification of specificity of crude antisera

(a) Ponceau stained membrane and (b) Western blot of an SDS-PAGE used to assess the polyclonal antibodies made against the purified peptide. The Western blot was probed with day58 crude antisera produced against 19 and 21. Lanes 1 and 9 contain molecular weight markers (labeled in Daltons). The High m.w. Standards ran slower than anticipated (see Results). Therefore, molecular weights inferred by lane 1 should be used cautiously. Lane 11 contains E. coli derived pooled 19 and 21 peptides used for antibody production. Lanes 6-7 contain protein extracted from yeast Eft16 that over-express a full-length SIRE1-4 env-like fusion protein. Lanes 2-5 and 8 contain other yeast lysates that express other fusion proteins. Samples that have a range of concentrations across multiple lanes are indicated by a wedge with the point indicating the lowest sample concentration. The black arrows show peptide bands commonly seen in previous purifications. Red arrows point to “dimer” bands seen in previous purifications. We see that the antisera detects only the pooled peptides and the yeast lysates containing the full-length env-like protein.
Fig. 19  Pre-immune antisera tests on plant material

Crude Antisera Rabbit 1

Crude Antisera Rabbit 2

Ponceau stained membranes and Western blots probed with pre-immune crude antisera from Rabbit 1 and Rabbit 2. Sizes are in Daltons. Lanes 3,4,7, and 8 contain soybean derived proteins. The tobacco and untreated lanes are leaf materials. The pre-immune antisera of both rabbits detects various proteins from the soybean derived materials. The arrows point to the putative Rubisco protein at about 55 kDa. Though aside from the putative Rubisco band only Rabbit 2 antisera shows banding in tobacco. Neither rabbits’ anti-sera detects proteins from Eft16 (lanes 1 and 5) which expresses a full-length SIRE1 env-like protein.
Fig. 20  Soybean stress treatments probed with crude antisera

(a) Ponceau stained membrane and (b) Western blot of SDS-PAGE separated protein extractions of soybean tissues subjected to the indicated treatments. Western blot was probed with day58 crude antisera and detected using Phosphaglo Reserve (KPL). Tissues in lanes 6, 7 and 11 as Rt (root) or Lf (leaf). Except for the tissue culture sample in lane 2 all others are leaf tissues. Marker weights are in Daltons. Lanes 9 is a 1:5 dilutions of the wound treatment. The persistent band at 55kDa was detected intensely on the Ponceau stain and to a lesser extent the Western blot. It was suspected to be RUBISCO. Tissue culture showed the most prominent bands at ~ 51 kDa, 57 kDa and a double bands at about 98 kDa. A barely detectable bubble was present between the 51 kDa and 57 kDa. Other notable bands were detected in lane 6 and lane1 at 64 kDa. On longer exposures the same band could also be detected in Lane 9.
Western blots of SDS-PAGE separated proteins from the indicated sources were probed with different antisera for comparison and detected with Phosphaglo Reserve (KPL). Lanes 3-4, 7-8, 11-12 are derived from soybean plant material. The putative Rubisco band was clearly detected by all antisera. The 51 kDa tissue culture band is marked by a red arrow and can be seen much more clearly in Rabbit 1 (lane 4) than in Rabbit 2 (lane 8). One of the few bands obviously detected by both Rabbit 1 and Rabbit two is a ~98 kDa band is circled in blue. In Rabbit 2, two separate bands were detected at this level, though the second band is weak.
Western blot of SDS-PAGE separated proteins from tissue culture, tobacco and treatments. Western blots were probed with purified Anti-\textbf{p19} Ab and washed lightly. Rubisco bands were barely detectable compared with Westerns done with crude antisera. However an unexpected band in tobacco was detected. Other bands were also detected in tissue culture, but with extensive washing both the tobacco bands and lighter tissue culture bands could not be detected (See Fig. 17 and 18). Tissue culture bands present in Fig. 17 and 18 are marked by arrows. The mark observed in the lane 1 (untreated) seems to be an artifact.
Fig. 23  Soybean treatments probed with purified anti-p19 Ab

(a) Ponceau stained membrane and (b) Western blots of SDS-PAGE separated proteins from Eft16, and tobacco leaves, tissue culture, and soybean leaves after the indicated treatments. Westerns were probed with purified anti-p19 Ab, washed extensively and detected with Phosphaglo Reserve (KPL). Standards are in Daltons. Lane 6 is a 1:3 dilution of the tissue culture protein. Lane 13 of the Ponceau stained membrane was enhanced to show protein. Three strong bands were detected in lane 5, and to some extent in lane 6, of the Western blot at 98 kDa, 65 kDa and 57 kDa. In leaf treatment lanes, No bands other than Rubisco background were seen. (c) Detail of Lane 5 comparing Ponceau stain (P) and Western blot (W). The 65 kDa band is especially dark on the Western blot, but relatively light on the Ponceau stained membrane compared to the other bands.
Fig. 24  Embryo and young plant tissues probed with purified anti-p19 Ab

(a) Ponceau stained membrane and (b) Western blot of SDS-PAGE separated proteins from embryos and seedlings (sdl). Lane 7 is a leaf treatment added for comparison. Westerns were probed with purified anti-p19 Ab, washed extensively and detected with Phosphaglo Reserve (KPL). Root (rt) and Shoot(sht) tissues are indicated where appropriate. Two strong bands were detected in lane 4 of the Western blot at 98 kDa, 65 kDa, and two faint bands were detected at 57 kDa, and 47 kDa. Similar sized bands could be detected in embryo and seedling tissues with the addition of new bands including 80 kDa and ~130 kDa. The 55 kDa band in lane 5 was probably to be rubisco. (c) Detail of lanes 2,3,4 and 7. Black arrows indicate bands detected in the tissue culture lanes as well embryo and young plant tissue lanes.
Comparison of the molecular weights of proteins from indicated samples. Weights for tissue culture were averaged from figures 17 and 18. Weights for embryo and long seedling shoot were taken from Fig. 17. Tissue culture bands in bold were consistently the strongest bands.

Table 3. Molecular Weights of Proteins Detected in Tissue Culture and Embryo tissues with purified anti-19 Ab

<table>
<thead>
<tr>
<th>Embryo Shoot</th>
<th>Tissue Culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>66</td>
<td>68</td>
</tr>
<tr>
<td>56</td>
<td>57</td>
</tr>
<tr>
<td>44</td>
<td>47</td>
</tr>
<tr>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>

Comparison of the molecular weights of proteins from indicated samples. Weights for tissue culture were averaged from figures 17 and 18. Weights for embryo and long seedling shoot were taken from Fig. 17. Tissue culture bands in bold were consistently the strongest bands.
CHAPTER FOUR

DISCUSSION

Introduction

Retroelements are ubiquitous features of the genomic landscapes of plants and animals. One class of retroelements, infectious retroviruses, possesses a third ORF that codes for an envelope protein that sponsors infectivity, which has been experimentally verified in vertebrates and insects. The envelope protein mediates the retroviral interaction with cell membranes allowing it to bud from the host cell and fuse with the membrane of the target cell. Sireviruses, which are found widely across families of plants, represent a class of endogenous retroelements, most of which possess a third ORF (env-like) whose conceptual translation product has predicted structural characteristics similar to those of a retroviral envelope protein (Laten, 1998; Havecker and Voytas, 2005). Infectivity, though, has yet to be verified.

The env-like gene of sireviruses presents an interesting question of function. Though enveloped plant viruses are known in plants (Lazarowitz, 2007) and though retroelements have repeatedly captured envelope proteins (Malik et al, 2000), no retroviruses, endogenous or otherwise, have been verified in plants. In addition, judging from the ubiquity of non-infectious retroelements in plants (Vitte and Panaud, 2005), envelope proteins are unnecessary for proliferation, yet the SIRE1 ORF has been preserved intact for 100 million years of evolution (Laten, unpublished). Determining
that SIRE1 possesses an envelope gene would expand the range endogenous retroviruses and possibly active retroviruses to plants as well as animals. If the ORF does not encode an envelope protein, the ORF, then, having no homologs, it would seem to be a gene of novel function among the small number of known retroelement genes.

Though recent transposition has been postulated (Laten, 2003), it is not known if any of the sireviruses are still actively retrotransposing. One hypothesis of this study is that SIRE1 elements are still active in soybean. SIRE1 cDNAs have been detected in water stressed roots (R. Winfrey, unpublished), but the BLAST searches done in the present study using LTR, gag, pol, and env sequences returned about 26 ESTs, two thirds of which are antisense. Even though sense ESTs were found, there are an estimated 2000 full or partial SIRE1 element copies (Laten, 1993). This makes it almost impossible to determine whether the transcripts were initiated in the element itself or were products of transcriptional readthrough from neighboring promoters, some of which may have been cryptic. Though transcription might be an unreliable indicator of SIRE1 activation, both sense and antisense strands might have biological function. Since sense-strand SIRE1 transcripts were found, copies of the retroelement might still be active and producing the env-like protein. Antisense transcripts could have regulatory function of the element as micro or siRNAs.

Since transcription alone does not indicate activation, downstream signs of biologically significant activity was sought by looking for translation of SIRE1 proteins in soybean. Retrotransposition mediated by Pol and Gag is well characterized, but the env-like gene is a unique feature of the sirevirus class. Therefore, the env-like protein was
chosen as the focus of this study. Evidence of this protein would not only suggest biologically significant activity of SIRE1 elements, but would also provide evidence that the unique sirevirus ORF actually encodes a translated product.

**Antibody Production**

Protocol Analysis

**Cloning**

With a limited supply of recombinatorial enzymes, a strategy was adopted to combine all reactions in each of the two recombinatorial steps (recombination of amplicons into the Entry vector and recombination of the Entry vector with inserts into the Donor vector). In the second recombination, the reaction volume was also reduced by half since this reduced volume was all that was required by the transformation protocol. At the end of the protocols for both reactions colonies with plasmid carrying a single insertion of 19 or 20 or 21 were identified using PCR. Reducing the reaction size by half and combining the reactions caused no obvious problems in the recombination reactions or in the identification of colonies containing each of the inserts. Since the recombinatorial enzymes are both the limiting factor and the most expensive component of the Stargate kit, this proved to be an excellent way to save on time and resources when cloning multiple, relatively small pieces into this system.

**Resin Purifications of Peptide**

As shown in Results, the imidazole gradient technique for His-tag purification seemed
to enrich the peptides to purities that were comparative to Strep-tag purifications. In
addition, the HisPur resin (ThermoFisher) might have been far below binding capacity.
For future peptide production, longer induction times for peptide expression coupled with
the a His-tag purification by an imidazole gradient might be more cost effective course in
terms of time, money, and production levels. However, cloning into a two-tagged system
might still be desirable. Having antibodies for each tag might make it easier to discern the
nature of any proteins at unexpected sizes.

Interpretation of Secondary Bands in Purifications

The secondary bands that were observed in resin-purified samples of peptides
p19, p21, or the manufacturer-supplied control [e.g. Fig. 14] can be divided into three
categories: the 20 kDa protein, “double-weight” proteins, and low molecular weight
proteins. They can be given two main interpretations: they were either natural E. coli
proteins that have co-purified, or they were tagged proteins related to the constructs.
Since each of the proteins was detected consistently and uniquely to only one of the
purified-peptide samples, it is unlikely that the detected proteins were simply E. coli
proteins that specifically or non-specifically bound to the column. Despite this E. coli
proteins could have interacted with resin-bound proteins. However, most of the
secondary bands were detectable by the anti-His mAb and were seen after the Strep-
purification, so this explanation likely accounts for a very small number of the secondary
proteins since it would require not only that the protein was bound non-specifically by the
antibody but also that it attached to a protein that remained after both purifications, e.g. the tagged-peptide.

Interpretations of construct-related secondary proteins can be addressed separately by category. In peptide **p19** purifications, a 20 kDa band [Fig. 14], about 1 kDa above the primary band was detected. One possibility was that the 20 kDa protein was derived from a population of *E. coli* cells in the culture with a mutation in the stop codon of the pASG-IBA43 vector. This would add a seven amino acid sequence that would increase the theoretical molecular weight as much as 760 Da, but would otherwise leave the affinity tags unchanged.

Many of the low molecular weight proteins - those found beneath the primary bands of **p19** and **p21** - were likely degradation products. The doublet pattern of the 19 kDa and 20 kDa proteins of **p19** seemed to recur in the low molecular weight bands suggesting degradation of the larger two proteins. Proteins of both **p19** and **p21** that were not detected after the Strep-purification, or are not detectable by the anti-His mAb might have lost one or the other of their tags. Since the degradation products should not introduce new epitopes to antibody production, antibodies binding to them should also bind to the full-length protein.

The third category of bands was detected by anti-His mAb at approximately double the molecular weight of the primary protein, and they were also retained by both purifications suggesting both tags were present. Single bands were detected above **p21** and the dual-tagged control [Fig. 14b ], and two bands were detected above the 19 kDa and 20 kDa bands **p19**. One interpretation could be that the proteins in the double-sized
bands were dimers of the primary bands formed by disulfide bridges, coiled coils, or by other non-covalent means. Peptide **p21** does theoretically have a coiled-coil, and the control protein, Azurin, is known to form disulfide bridges (Chen *et al.*, 1998). However, **p19** does not possess a predicted coiled-coil region and has no cysteines necessary for disulfide bond formation. In addition, the samples were boiled with SDS and BME for 7 minutes before being loaded onto the gel, and the dimer-sized protein in **p21** was initially detected in lysates denatured with 8M urea. Despite this, explanations for the bands can range from protein aggregation (Schagger, 2006; Sagne, 1996) to formation by SDS- and urea-resistant, coiled-coil oligomers (e.g. Monier, 1995; Kubista, 2004; Brueggemann, 2004; Rausell, 2004) to disulfide bond formation from weakened reducing agents (*e.g.* β-mercaptoethanol). Nevertheless, the proteins were very weakly detected by Western blot, and were not detectable by staining methods [Figs. 13a, 14a and 15-17] indicating that their concentrations in the elutions were very low.

If the secondary bands represent *E. coli* contaminants, they should be inconsequential since they were from a bacterial source and would not be expected to generate antibodies that detect proteins in eukaryotic systems. Thus antibodies that are specific to the contaminants would not be expected to have downstream consequences when used with soybean. Nonetheless, many of the secondary bands were likely to represent products related to the fusion peptides themselves and were therefore not likely to interfere with polyclonal antibody production.
Polyclonal Antibodies

Specificity of Crude Antisera for Purified Peptides and Env-like Proteins Produced in Eukaryotes

The experimental results in Fig. 18 showed strong binding to the antigen source, strong binding to proteins in yeast lysates (Eft16 lysates) which expressed a full-length env-like fusion protein (fused to a Gal4 binding domain), and zero binding to negative control yeast lysates which expressed fusion proteins other than the env-like protein. The lack of signal in negative controls indicated that despite the inclusion of possible contaminating proteins in the antigen used to raise the env-like antibodies, there were no cross-reacting antibodies that bound yeast proteins.

Though the proteins that were detected in Eft16 ran substantially lower than expected, these results were not reproduced in subsequent experiments. Subsequent Western blots of the Eft16 lysates probed with the crude or purified antisera showed that the highest protein runs at 116 kDa [e.g. Fig. 23b]. This was consistent with Western blots of Eft16 probed with a monoclonal antibody against the Gal4 domain (data not shown). This indicates that for unknown reasons either the proteins or the large molecular weight standards [Fig. 18b, lane 1] ran aberrantly.

While the monoclonal anti-Gal4 antibody only detected one protein, the anti-env antibodies detected as many as four. The smaller proteins detected in the Eft16 lysates were likely to have been breakdown products of the env-like protein since they were not detectable in any of the yeast negative controls. The second largest protein was known in other experiments to run at ~94 kDa [e.g. Fig. 23b]. This was consistent with the removal
of the entire Gal4 domain, which has a molecular weight less than 15 kDa. This would explain why the anti-Gal4 antibody did not detect them.

The experiment shown in Fig. 18 was important. It showed that no antibodies were generated that bind non-specifically in at least one eukaryotic system, yeast. It also confirmed that the antibodies that were generated were not only detecting proteins from the pooled, purification product, which were presumed to be tagged-fragments of anti-env, but also that the antibodies were detecting a yeast-generated copy of the protein. This indicated that antibodies were specifically detecting the env-like protein and that they were not binding only to the tagged regions or other E. coli proteins. Lastly we do not know what post-translational modifications occur in soybean, but the experiment did show that detection of the full-length protein is not precluded by post-translational modifications that might have occurred in the yeast. Taken together these results suggested a good possibility that the antisera could detect a naturally made env-like protein in soybean.

Soybean Treatments

Pre-immune, Post-injection Crude Antisera

Testing plant material with the pre-immune sera showed that both rabbits had many endogenous antibodies against plant proteins, including the putative rubisco protein [Fig. 19]. According to OpenBiosystems, antibodies raised in rabbits can bind plant proteins other than the supplied antigen and can be a problem in antisera used for detection in plant systems. Presumably this is because the rabbits are fed plant-rich diets.
One service OpenBiosystems offers is sending antisera from multiple, pre-immunized rabbits. The rabbits antisera can then be tested in the specific plant system, and individual rabbits with the lowest pre-existing antibody signals can then be chosen for immunization. For future antibody generation, taking advantage of this service or generating monoclonal antibodies might be the best options.

Despite the detection of obvious confounding proteins, candidate bands were identified when the crude antisera (post-immune) was used to probe stressed soybean plants and tissue-cultured soybean. In particular, three bands in the tissue culture lane [Fig. 20b, lane 2] were particularly strong and are discussed below.

**Purification of Antibodies**

Attempts were made to clarify results of the Western blots probed with crude antisera and to see if weaker bands might correspond to env-like proteins. In one experiment, crude antisera was incubated with soybean or tobacco lysates in an effort to bind all antibodies that were not specific to the fusion-peptides. In a converse approach, the antisera was incubated with the fusion peptides (antigen competition) to bind only fusion-peptide specific antibodies. With this strategy, blots probed with competed and non-competed antisera would be compared. Env-like bands should be noticeably diminished in the Western blots probed with competed antisera. However the results from these attempts revealed no new information as to which of the bands might correspond to env-like proteins (data not shown).
A third strategy was the purification of antibodies using immobilized peptides. This produced a significant reduction in the putative binding of rubisco and other proteins that were presumably not derived from env-like proteins [Fig. 22]. There were still traces of rubisco. However, it seems logical that if rubisco is so prevalent, antibodies specific to it are also probably in fair abundance. The antibody purification protocol (Sambrook et al, 1989) only calls for two washes. In retrospect, this seems inadequate. The Phosphaglo Reserve (KPL) system however is geared towards ultra sensitive detection for protein amounts as low as hundreds of femtograms. Trace amounts of rubisco-specific antibody were probably still present after purification. The weak 55 kDa bands were probably caused by the shear abundance of the putative rubisco protein binding to trace amounts of rubisco-specific antibody that was left in the purified antibodies.

Unexpectedly after purification, two tobacco bands were detected more intensely than in previous Western blots [Fig. 22]. Detection alternated between a 75 kDa (data not shown) band and a ~150 kDa band. In the initial Western blots using crude antisera, these bands were not detected in tobacco [Fig. 20b, lane 3]. Since only one or the other of the bands was often observed, this result could have been due to dimerization. In the experiments testing the polyclonal antibodies, boiling of the samples was substituted by incubation of the samples for 10 minutes at 65°. This was designed to avoid aggregate formation (Schagger, 2006). However this lower temperature might not been sufficiently high to eliminate some of the interactions leading to dimers.

Sireviruses have been recently found in tobacco (Grandbastien, unpublished) but no amino acid sequence similarity to soybean env-like peptides was present. BLAST
searches against the NCBI databases were done on protein sequences, translated genome sequences and translated EST sequences in tobacco using the tagged peptide sequences as queries. Though no proteins were found as potential cross-reacting targets, the tobacco genome is far from being fully sequenced. It is unclear why these protein bands were detected, but extensive washes removed all traces of the bands [Fig. 23], which suggests that they might simply have been due to strong yet non-specific binding, or weak binding to an abundant protein.

**Analysis of Detectable Bands in Soybean Samples**

**Tissue culture**

As mentioned in Results, the intensities of the bands detected in tissue culture were altered in Western blots probed with crude antisera versus the purified anti-p19 Ab. [Fig. 20b, lane 2 and Fig. 23]. Though this could have been merely experiment-to-experiment variation, several other things could account for these differences. First, the weakened bands might not have been env-like proteins and thus the purification process would have excluded antibodies targeting them. The possibility that candidate proteins are not env-like might account for the loss of one of the 98 kDa bands and the weakening of the 57 kDa band [Fig. 23]. Evidence from the Rabbit 2 pre-immune Westerns [Fig. 19] corroborated this. As mentioned, a weak, single 98 kDa band and a strong 57 kDa band was detected. The 57 kDa band also corresponds to an intense Ponceau band. The residual trace of the 57 kDa band might be due to contaminating antibodies similar to the
mechanism that was postulated for the continued detection of the rubisco band in Western blots probed with purified antibodies.

However, proteins with weakened signals such as the 57 kDa protein might still be SIRE1 env-like proteins. As reported in Results, after purification, only the anti-p19 Ab was effective. The initial mix of crude antisera from both rabbits presumably contained both anti-p19 Ab and anti-p21 Ab. Proteins with weakened signals might have been cleaved proteins that had a few p19 epitopes, but many p21 epitopes. Proteins with signals that disappeared entirely might have had p21 epitopes exclusively.

Similarly, the difference in immune response between the rabbits was a probable factor in the variation. For example the 51 kDa protein is not seen in the pre-immune antisera for either rabbit [Fig. 19], but sponsors a much stronger signal in the Day 58 sera from Rabbit 1 than from Rabbit 2 [Fig. 21]. The antibody purifications were done with Rabbit 2, Day 58 sera. If, by chance, only Rabbit 1 had a significant immune reaction to p21, these antibodies would have been lost, and diminished signals would have resulted in proteins with significant proportions of p21 epitopes.

As for the signals that were present after the switch to anti-p19 Ab, the 65 kDa protein gave a strong signal in all Westerns that used the purified antibody. Unlike the rubisco band in Fig. 23, the 65 kDa Western blot signal was disproportionately strong when compared to the Ponceau stained protein on the membrane [Fig. 23c versus rubisco in 23a and 23b]. It was not as clear that this was the case for the 57 kDa or the 98 kDa proteins since exact alignment between the images was difficult and there were other Ponceau-stained bands in these regions [Fig. 23c]. A longer SDS-PAGE run might
improve the resolution. However, as discussed, the 57 kDa band is possibly due to residual contaminating, pre-immune antibodies binding to an abundant protein [Fig. 19]. Though a 98 kDa band was also seen in pre-immune, two bands were seen in the day 58 antisera [Figs. 20 and 21]. After purification only one is again seen. One interpretation of this result is that the Rabbit 2 Day 58 antisera was detecting both an env-like protein and a non-env-like protein and the band that remains after purification corresponded to the env-like protein.

Overall the 65 kDa protein followed by the 98 kDa protein seem to be the most likely candidates for potential env-like proteins.

Stress treatments

None of the biotic or abiotic stress conditions resulted in detectable and or reproducible env-like protein detection using the purified antibodies. Whether this was due to the absence of protein or to tissue source and age or to technical limitations is not known, although these issues can be addressed in future studies. Vacuum infiltration of older plants on the verge of flowering was noticeably more difficult. Perhaps in future studies only younger plants should be used.

Detection of proteins with the crude antisera in water-stressed roots seemed to corroborate a previous study that found SIRE1 cDNAs in water-stressed roots (R. Winfrey, unpublished). However the detected bands were very weak. Furthermore, in the blots seen in Fig. 23, the concentrations of protein in each lane of both the water-stressed roots and the methyl jasmonate treatments were particularly low as evidenced by the
Ponceau staining [Fig. 23a]. Low protein levels, coupled with the extensive washing used in these Western blots, may have been factors in the lack of detectable bands, although the same amount of protein was used in the initial Western blot where weak bands were detected. Nevertheless, a higher amount of protein would have made it more likely that rare proteins could be found, but proteins were loaded onto the gel without dilution.

Like the case for the 51 kDa band in tissue culture, another explanation is the possibility that only antibodies specific to p21 might bind the initially detected proteins. Though initial detection of proteins in methyl jasmonate and water-stress treatments was not replicated with purified anti-p19 Ab, these treatments cannot be ruled out as potential inducers of env-like protein.

One sample does show bands in the stress treated samples. Upon very long exposure times, bands at 100 kDa and 65 kDa do appear in the wound treatment lane (data not shown). This corresponds to the bands seen in the tissue culture sample, but at this exposure time the likelihood of bands appearing from trace amounts of possible contaminating antibody is higher.

It should also be noted that, except for the water-stress treatment and the culturing of tissue, tissues subjected to stress treatments were collected after 24 hours. Tissue culture however is a long-term “condition” as cultures were grown for several weeks. Though previous studies have shown retrotransposon activity to occur within 24 hours and even within 12 hours of the initiation of bioitic and abiotic stress treatments (e.g. Takeda, 1998; Tapia, 2005; Mhiri, 1997), these studies focused on detecting transcripts,
not proteins. Since RNA has been an indicator in other studies, perhaps stress treatments should be looked at with RNA studies or by other means in tandem with or even preceding protein work, and perhaps stressed leaves should not be collected at 24 hours but should be pushed out to 48 or 72 hours, which might allow an accumulation of the protein.

An unconventional stress, hybrid stress, produced no detectable bands [Fig. 23 Lanes 8 and 9]. Transcription and amplification of plant retrotransposons has been correlated with interspecific hybridization (Reviewed by Grandbastien, 2010). One proposed theory on why transposable elements (Class I and Class II) are able to mobilize in hybrids suggests that maternal silencing mechanisms in one species do not recognize paternal transposable elements from another (Reviewed in Madlung and Comai, 2004). This theory proposes that micro RNAs used to program the epigenetic silencing of the embryo are supplied maternally to the egg cell but are species specific. The maternal RNAs do not recognize the paternal transposable elements after fertilization and therefore the failure to silence results in the activation of the paternal transposable elements.

The putative hybrids used in this study were crosses between a maternal *G. max* and a paternal *G. soja*. Since *G. soja* was the wild ancestral progenitor of cultivated *G. max*, and the two species have only been isolated for 10,000 years, it was perhaps unlikely that the elements in the paternal genome of *G. soja* would go unrecognized by the silencing molecules of the maternal *G. max* genome. Despite the recent divergence of the two species, it is still possible that elements in *G. max* elements could go unrecognized by the *G. soja* silencing pathway due to the cultivation of *G. max*, but
according to the theory above this would be best tested with a maternal *G. soja* and a paternal *G. max*. Unfortunately for the experiment the hybrid used above was the only one available. Still better hybrids might be or more distant crosses entirely.

Tissues from seedlings and embryos

Retrotransposon transcription has been recorded at varying stages of maize and rice development (Vicient, 2010; Jiao and Deng, 2007), so it is not surprising that SIRE1 env-like candidate proteins were detected in these tissues [Fig. 24a and 24b]. Due to the multi-copy nature of SIRE1 and the theoretical nature of env-like proteins (discussed below), it is also not surprising that proteins were detected at new molecular weights compared to those detected in tissue culture [Table 3].

Though the induction of tissue culture is considered a stress and stress-related genes are expressed during culture (Rakocevic et al, 2009), it might not be the stress response that is promoting activity of the SIRE1. Of the samples used, the embryo and seedling tissue and the tissue culture shared the characteristic that they are likely to have cells that they have actively growing undifferentiated tissues. A similar suggestion has been made to account for retroelement transcription in maize (Vicient, 2010). In that particular study however, transcription of retroelements was very low in embryos. In this experiment most of the other stress treatments were done on leaves, which might simply not have been the ideal tissue in which to find SIRE1 expression. It might be informative to see the response of apical meristems and root tips to the various stresses.
Interpretations of the Multiple Bands

Several proteins were identified as candidate env-like proteins. Interpretation of the bands is no easy task as the primary structures of the actual protein itself are not well understood, and the complex nature of expression of the env-like protein makes it difficult to tell what protein sizes might be precursors, mature proteins, cleavage products or degradation products. The multiple bands may have several explanations that are equally valid.

The 98 kDa protein might correspond to a full length env-like protein though the molecular weight of the SIRE1-4 env-like protein is theoretically only 74.35 kDa. The **Eft16**, a cloned version of the SIRE1 env-like protein fused to a Gal4 domain, ran at ~116 kDa yet has a theoretical molecular weight of 94 kDa. The shared difference in weight (~22-24 kDa) might be explained by expected post-translational modifications such as glycosylation (Laten, 1998).

It is also possible that some of the detected bands are precursors. As stated above in the Introduction, the SIRE1 env-like protein is thought to be a read-through product, therefore Gag-Pol-Env-like precursors of about 250 kDa are predicted (Havecker and Voytas, 2003). Nothing was detected in this range, but the proteins that were found were detected in tissue culture where the env-like protein theoretically constantly being produced and therefore has a chance to collect in the tissue. The precursor on the other hand might only be present in minute amounts at any given time. However other precursors would also be predicted as the retrotransposon proteins are proteolytically cleaved. Two possible precursors might be RNase+Env-like and reverse
transcriptase+RNase+Env-like. Using predicted sizes of proteins from the SIRE1-4 element, these two precursors should be 101 kDa (27 kDa + 74 kDa) and 128 kDa (28 kDa + 27 kDa + 74 kDa), respectively. The former corresponds closely to the 100 kDa protein found across many of the samples in Fig. 24 and the latter to the 130 kDa protein found in embryos and long seedlings [Fig. 24b]. As this is the first study to look directly at the presence of the envelope protein, it cannot be said what, if any, other precursors might exist nor the location of the cleavage sites between RNase H and env-like to form the mature protein. Therefore we cannot know what size the “full-length”, functional protein actually is.

The smaller proteins could be cleavage products of the “full length” env-like, which might be expected if env-like was an envelope protein. Both HIV and drosophila gypsy envelopes are cleaved at furin-like sites [RxxR] (Binley, 2002; Misseri, 2004). The HIV envelope protein, GP160, in particular is cleaved into a Surface protein (SU), GP120, and a transmembrane protein (TM), GP41, by a cellular protease. After cleavage, SU binds non-covalently to TM, which is anchored in the lipid envelope of the virus. SU binds cellular receptors and allows the TM to rearrange and mediate fusion with the cell membrane. These proteins also sponsor budding of new viruses from infected cells (Freed, 2001).

Though cleavage is necessary for a functional protein, the [RxxR] protease cleavage motif was not found in any of the copies of the SIRE1 env-like protein that are presumed functionally competent. Cleavage products cannot be ruled out however. A coronavirus variant that lacks the furin-like cleavage site necessary for its functionality
was found to be cleaved, though it is not known if it is by a furin-like protease using another site or an alternate protease entirely (e.g. serine protease) (Ana Shulla, unpublished). SIRE1 env-like might also be cleaved by other proteases.

The smaller bands could also be degradation products. The bands of presumed degradation products in Eft16 lanes are of similar size to the bands detected in embryo and tissue culture lanes [Fig. 24b], but there is another possibility. There are over 2000 copies of the env-like gene in the soybean genome and many of them have the potential to produce truncated copies of differing sizes. This also could explain the alternate bands seen in the embryonic tissue. As seen with the differential expression of elements within the TNT1 family (Beguiristain et al., 2001), different SIRE1 elements with varying sizes of env-like proteins might have been active in response to different development and stress triggers.

Assigning a justification for the presence of each band is obviously complicated and each may have been the result of more than one of these factors.

**Conclusions**

EST and cDNA data have been collected in support of active SIRE1 viruses and transcription of the env-like gene. Though this data is an excellent place to start looking for protein expression, retroelement transcriptional data by itself is fraught with dangers of interpretation as is evidenced by a recent paper that characterizes the over-interpretation of transcriptional data from human endogenous retroviruses (HERVs) as detecting “rumor” viruses (a pun on “tumor viruses”) – or false positives associations
with diseases (e.g. Voisset, 2008). Though there is much information to be mined from transcriptional studies, the critique above is a reminder of the assumptions that we make and the care that needs to be taken when analyzing any data - especially data from retroelements.

To my knowledge, the study undertaken here is not only the first study that has hinted at the activity of sireviruses by looking directly at the translated gene product, but also the first study that looks directly at env-like expression at the level of protein products in envelope-class retrotransposons in plants. This is significant since the env-like protein is the key element that separates the sireviruses and ATHILA-like viruses from retrotransposons that are ubiquitously found across eukaryotes.

In this study antibodies were successfully raised against purified fusion peptides that covered two regions of the env-like gene from SIRE1-4. Though sequencing will be required to positively identify the proteins, this study strongly suggests that I have detected up to three proteins with putative env-like regions in cultured soybean tissue. Though the main source of the protein identified here was cultured tissue, this study suggested that SIRE1 might be active in early development as well. SIRE1 and the env-like protein appear to still be produced in soybean, which suggests that they might have biological relevance.

Future and Corollary Studies

Even with the failure of the anti-p21 Ab in this study, my results further support the hypothesis that SIRE1 may be functionally activated in soybean. With newly purified
anti-p21 Ab, more proteins will likely be found in tissue culture extracts. Since the anti-p19 and anti-p21 Abs detect different regions of the env-like protein, Western blots using anti-p21 Abs will help answer questions as to the nature of the multiple bands by independent detection of env-like proteins containing both regions. In addition new, truncated proteins or cleavage products might also be detected.

There are reported polyclonal antibodies to the SIRE1 gag protein (anti-gag Ab; Voytas, unpublished). Our lab intends to use these antibodies to detect the gag protein. Read-through translation into the env-like region only occurs 5% of the time in an in vitro system (Havecker and Voytas, 2003). Thus gag protein might be produced in vivo at 20 fold the concentration of the env-like protein, which means there might be more than 20 ng of gag protein in tissue cultured preparations. Since proteins in embryo and seedling tissues gave weak signals, it would be informative to assay these tissues using the anti-gag Ab.

With evidence of functional expression in soybean, a next logical step would be to transform Arabidopsis with a SIRE1 element. Arabidopsis would be an ideal host since it is less likely to have any SIRE1-specific epigenetic silencing mechanisms. Other retrotransposons have successfully been introduced into Arabidopsis and have shown transposition (Hirochika, 2000; Perez-Hormaeche, 2008). Also BLAST searches against the SIRE1 env-like region reveal homologous fragments from distant and heavily mutated SIRE1 relatives in Arabidopsis. This implies that Arabidopsis has possessed sireviruses proteins in the past and might be able to make functional SIRE1 with its env-like protein.
Now that potential native sources for the SIRE1 env-like protein have been identified, the protein’s biochemical and structural nature can be further investigated. Studies can be conducted on the nature of translatable RNA, and using the purified antibodies, it might be feasible to visualize the protein and virus-like particles in vivo.

There are still yet many questions to answer about sireviruses and their unique protein. Hopefully this work will be a starting point from which questions about the nature and mysterious function of the env-like protein will be answered, and a first glance into the role of envelope-class retrotransposons in the evolutionary history of plants.
REFERENCE LIST


Retrotransposon Tnt1 Is Controlled by Reversible Transcriptional Gene Silencing.


VITA

Garen Gaston was born and raised in Woodward, Oklahoma. Before attending Loyola University Chicago, he attended the University of Wisconsin, Madison, where he earned a Bachelor of Natural Sciences with a focus on Genetics in 1998. While at Madison and directly after, Garen worked in a variety of biology labs that focused on plant genetics, plant cytogenetics and cancer research.

Garen came to Loyola to earn a Masters in Biology where he was awarded full tuition and stipend. At Loyola, he spent his time studying putative retroviruses in soybean. During his stay, he was a teaching assistant for Genetics and Genomics classes and was an instructor for Biology labs. He is a member of the American Association for the Advancement of Science (AAAS) and the American Society for Microbiology and presented part of his thesis work at the American Society for Microbiology Mobile DNA Conference in Montreal, Canada in 2010.

Currently, Garen has plans to continue molecular research though he will change location to a marine biology lab in Haifa, Israel.
THESIS APPROVAL SHEET

The thesis submitted by Garen Gaston has been read and approved by the following committee:

F. Bryan Pickett, Ph.D., Director
Professor of Biology
Loyola University Chicago

Eric Schroeter, Ph.D.
Professor of Biology
Loyola University Chicago

Stefan Kanzok, Ph.D.
Professor of Biology
Loyola University Chicago

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

________________________  ______________________
Date  Director’s Signature