1-1-2007

The characterization of a 21 kDa protein in the brine shrimp, Artemia franciscana, as a Group 1 late embryogenesis abundant (LEA) protein.

Michelle Ann Sharon
University of Windsor

THE CHARACTERIZATION OF A 21 kDa PROTEIN IN THE BRINE SHRIMP, *ARTEMIA FRANCISCANA*, AS A GROUP 1 LATE EMBRYOGENESIS ABUNDANT (LEA) PROTEIN

by

Michelle Ann Sharon

A Thesis
Submitted to the Faculty of Graduate Studies and Research
through Biological Sciences
in partial fulfillment of the requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

2007

Your file Votre référence
ISBN: 978-0-494-35060-7
Our file Notre référence
ISBN: 978-0-494-35060-7

ABSTRACT

Late embryogenesis abundant (LEA) proteins are hydrophilic molecules believed to provide desiccation tolerance to plant seeds, certain prokaryotes, and some animal species. The aquatic crustacean *Artemia franciscana* (brine shrimp) can exist as a dormant encysted embryo that is highly resistant to dehydration, potentiating molecular mechanisms of tolerance to dehydration, such as LEA protein synthesis. The experiments described herein characterize a 21kDa protein abundant in encysted *Artemia franciscana* embryos as a Group 1 LEA protein. A protocol for the purification of the protein has been established. Through trypsin digestion, MALDI mass spectrometry and de novo sequencing, as well as amplification of the encoding gene by PCR, the complete sequence of the Group 1 LEA gene has been identified. Eight tandem repeats homologous to the 20-amino acid sequence that characterizes Group 1 LEA proteins have been identified. This is the first identification of a Group 1 LEA protein in an animal species.

ACKNOWLEDGEMENTS

I wish to extend my sincere thanks to my advisor, Dr. Warner for the opportunity to complete a Master of Science degree in his laboratory, as well as for his continuous guidance and encouragement. I would also like to thank my committee members, Dr. Hubberstey and Dr. Vacratsis, for their interest in my research project and for providing direction. Thank you to my advisor, my committee members, and the chair of my defence, Dr. Corkum, for their review of my thesis. I would also like to thank faculty and graduate students in the Department of Biological Sciences for their support and assistance. Thank you to my family and friends for their endless support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 ANHYDROBIOSIS

Anhydrobiosis is the ability of an organism to survive in the dry state by protection of cellular structure and machinery (Hand *et al.*, 2006). Anhydrobiosis is characterized by nearly complete desiccation of the organism, metabolic arrest, and the accumulation of the disaccharide trehalose in anhydrobiotic animal species (Kikawada *et al.*, 2006). Organisms are also capable of surviving a desiccated state due to upregulation of selected hydrophilic proteins, such as late embryogenesis abundant (LEA) proteins (Goyal *et al.*, 2005). Anhydrobiosis is reversible upon rehydration (Goyal *et al.*, 2005). The importance of protective mechanisms for survival of desiccation is exemplified in studies of plant seeds subjected to this environmental stress. When desiccation-sensitive plant embryos are exposed to dehydration, the loss of water causes the collapse of the embryo's tissue systems. However, scanning electron microscopy of desiccation-tolerant plant embryos exposed to dehydrating conditions reveals the preservation of both external and internal tissue systems, facilitating the survival of the embryo (Wakui & Takahata, 2002). Similarly, the existence of desiccation tolerant life cycle stages in animal species, such as the encysted diapause embryo of the brine shrimp, *Artemia franciscana*, the desiccation-resistant dauer juvenile larval stage of the nematode, *Caenorhabditis elegans*, and anhydrobiotic larva of the chironomid *Polypedilum vanderplanki*, (Tunnacliffe & Wise, 2007) facilitate the survival of these species during dehydration.

1.2 LATE EMBRYOGENESIS ABUNDANT (LEA) PROTEINS

1.2.1 EXPRESSION OF LEA PROTEINS

First characterized by Galau, G.A. and Dure, L. III (1981) in the seeds of cotton, late embryogenesis abundant (LEA) proteins are hydrophilic, heat stable proteins postulated to have a considerable role in tolerance to desiccation (Gal *et al.*, 2004, Campos *et al.*, 2006). LEA protein expression has been identified primarily in plant seeds, but also in pollen, vegetative organs of plants during water deficit, animals such as

the nematodes *Caenorhabditis elegans* and *Aphelenchus avenae*, and prokaryotes such as *Haemophilus influenzae* and *Bacillus subtilis* (Gal *et al.*, 2004, Wise, 2003). LEA proteins continue to be identified in additional organisms, such as the recent identification of two mRNA molecules believed to encode two Group 3 LEA proteins in diapause and post-diapause stages of the brine shrimp, *Artemia franciscana* (Hand *et al.*, 2006).

LEA proteins are produced late in embryo development, comprising up to 4% of the cellular protein (Wise & Tunnacliffe, 2004, Hinniger *et al.*, 2006). The expression of LEA proteins can be developmentally programmed, as in maturing plant seeds, or induced by environmental stresses such as desiccation stress (Dure *et al.*, 1989), cold stress, and osmotic stress (Hinniger *et al.*, 2006), improving tolerance to these conditions, as well as by exogenous abscisic acid (Dure *et al.*, 1989, Soulages *et al.*, 2002). A small portion of the LEA proteins are expressed constitutively (Goyal *et al.*, 2005).

In plant species, LEA proteins and their mRNA levels peak when desiccation is experienced during seed development (Ohno *et al.*, 2003). However, LEA proteins are thought to provide other plant tissues with desiccation tolerance upon the imposition of water-related stresses or induction by abscisic acid (ABA) throughout the plant's life cycle (Dure *et al.*, 1989, Hinniger *et al.*, 2006, Espelund *et al.* 1992). LEA proteins and their encoding mRNAs are believed to be rapidly degraded upon germination of plant seeds or cessation of the stress and rehydration of the organism (Swire-Clark & Marcotte, 1999, Chourey *et al.*, 2003).

The expression patterns of LEA proteins in invertebrate species are sometimes difficult to unravel as messenger RNA (mRNA) levels may not correlate with LEA protein levels. For example, the gene encoding the Group 3 LEA protein Aav-lea-1 in the nematode *Aphelenchus avenae* is up-regulated during dehydration, although the protein is already present in the unstressed animal as well as during dehydration. Desiccation has been found to induce cleavage of this LEA protein (Goyal *et al.*, 2005). Disconcordant expression of mRNA and LEA proteins has not been reported for other invertebrate species in which LEA proteins are synthesized (Tunnacliffe & Wise, 2007).

1.2.2 MECHANISMS OF ACTION OF LEA PROTEINS

Although the mechanisms of action of LEA proteins have not been fully elucidated, these proteins are postulated to protect cellular and molecular structures from desiccation-induced damage (Goyal *et al.*, 2005, Campos *et al.*, 2006). LEA proteins may function by binding water molecules, thereby decreasing the rate of water loss from cells during dehydration (Hinniger *et al.*, 2006). When a cell dehydrates, there is an increase in the concentration of intracellular components, such as ions, that can damage macromolecules (Tunnacliffe & Wise, 2007). LEA proteins may function by sequestering and scavenging these ions (Soulages *et al.*, 2002), thwarting osmotic stress conditions. LEA proteins have been postulated to protect dehydration-sensitive proteins from denaturation or aggregation during water stress conditions by participating in protein-protein interactions (Hinniger *et al.*, 2006, Soulages *et al.*, 2002). However, LEA proteins may also act as molecular shields that occupy space in the cytoplasm of the dehydrating cell, decreasing interactions between partially denatured proteins and preventing protein aggregation (Tunnacliffe & Wise, 2007). LEA proteins may also play a role in the renaturation of unfolded proteins (Wise, 2003). Associations with membranes and the subsequent enhancement of membrane stability during dehydration is another proposed function of LEA proteins (Tunnacliffe & Wise, 2007). LEA proteins may play a role in the cytoskeleton that resists physical stresses imposed by desiccation (Kikawada *et al.*, 2006, Hand *et al.*, 2006). In this case, the LEA proteins would help to prevent the collapse of the cell as its water is depleted (Tunnacliffe & Wise, 2007). It has also been suggested that LEA proteins may integrate with sugar matrices, forming protein-sugar interactions within seeds that are thought to increase the mechanical stability of the cell, enhancing desiccation tolerance (Shih *et al.*, 2004). Studies of seedlings of a rice species (*Oryza sativa* L.) undergoing salinity stress revealed that LEA proteins may be partly responsible for the growth arrest of these seedlings during salt stress, an adaptation strategy for the conservation of energy and resources during adverse conditions. Therefore, the degradation of the LEA proteins is thought to aid in the recovery of the cells from the stress and the continuation of development (Chourey *et al.*, 2003).

The extreme hydrophilicity of LEA proteins is thought to prevent them from acting as transmembrane proteins or as globular proteins with hydrophobic cores

(Campos *et al.*, 2006). In fact, LEA proteins have been found to exhibit specific subcellular localization patterns to the nucleus, mitochondria, endoplasmic reticulum, cytoplasm, chloroplasts, peroxisomes, and vacuoles (Tunnacliffe & Wise, 2007). Since LEA proteins exhibit specific localization patterns within the cell and tissues, it is possible that different groups of LEA proteins or different group members have separately defined roles in the protection of the organism from desiccation stress (Soulages *et al.*, 2002). The overlapping functions of these multiple LEA proteins make it difficult to identify the mechanism of action of an individual LEA protein. Furthermore, the LEA proteins may be extremely multi-functional (Tunnacliffe & Wise, 2007). It is believed that the widespread expression of LEA proteins throughout subcellular locations and tissues is an indication of the importance of LEA proteins in all cellular compartments during desiccation (Tunnacliffe & Wise, 2007).

1.2.3 CATEGORIZATION, AMINO ACID COMPOSITION, AND SECONDARY STRUCTURE OF LEA PROTEINS

LEA proteins have been categorized into six different groups, based on differing expression pattern and sequence. Groups 1 – 3 are considered to be the major LEA protein groups. Proteins in the LEA groups are characterized by an abundance of hydrophilic amino acids that are charged (Wise, 2003, Campos *et al.*, 2006). The hydrophilicity and abundance of charged amino acid residues are thought to contribute to the high solubility of LEA proteins during boiling (Shih *et al.*, 2004). For example, a Group 2 LEA protein, also called a dehydrin, remained soluble even after 10 minutes of boiling (Ohno et *al.*, 2003). Members of the LEA protein groups are quite variable in size, ranging from an 8 kDa Group 3 LEA protein in the pollen of *Typha latifolia* to a 77 kDa Group 3 LEA protein in the nematode *Caenorhabditis elegans* (Tunnacliffe & Wise, 2007).

Repeated amino acid sequence motifs are frequent among LEA proteins and postulated to contribute to the desiccation tolerance function of these proteins (Dure *et al.*, 1989, Espelund *et al.*, 1992). It is these sequence motifs that facilitate the categorization of LEA proteins into their respective groups. For instance, Group 1 LEA proteins are characterized by a hydrophilic repeated motif of twenty amino acids

(Cuming, 1999). Group 2 LEA proteins possess at least two of three characterized sequence motifs, entitled Y, S, and K (Close, 1997), while Group 3 LEA proteins display a motif of eleven amino acids that is repeated multiple times (Cuming 1999). The number of repeats appearing in proteins of a particular LEA protein group varies widely (Dure *et al.*, 1989), often ranging from one to thirteen repeats of a particular sequence (Wakui & Takahata, 2002, Hong-Bo *et al.*, 2005). These repeated sequences may appear in tandem or with intervening tracts of amino acids divergent from the repeated sequence (Dure *et al.*, 1989). The degree of similarity of repeat regions also varies within the sequence of an individual protein and among sequences of different proteins (Dure *et al.*, 1989). As described for LEA proteins belonging to Group 3, the repeats may become more degenerate toward the C-terminus of the protein (Hand *et al.*, 2006). Although it is a general consensus that the repeated sequence motifs found in LEA proteins are important to protein function, the precise role of the repeat regions remains elusive. It has been postulated that the conserved sequence motifs are recognition sequences that allow for the interaction of the LEA protein with other cellular components (Mouillon *et al.*, 2006). The sequence motifs could also be important for the folding of the LEA protein, for subcellular localization, or for protein processing events (Tunnacliffe & Wise, 2007).

LEA proteins in solution are said to possess little or no ordered secondary structure (Campos *et al.*, 2006), remaining largely unfolded in solution and being classified as intrinsically unstructured proteins. LEA proteins from Groups 1 – 4 are predicted to be at least 50% unfolded (Tunnacliffe & Wise, 2007). However, desiccation may induce conformational changes that cause the proteins to undergo transition from unordered to ordered structures, such as α-helices, during drying, which may be related to the proteins' mechanisms of action (Wolkers *et al.*, 2001). Conformational changes in response to dehydration have been observed in at least two Group 3 LEA proteins from anhydrobiotic nematodes and *Typha latifolia* pollen (Shih *et al.*, 2004). However, these dehydration-induced conformational changes may not be a characteristic of all LEA proteins (Shih *et al.*, 2004).

Post-translational modifications, such as phosphorylation, have been proposed for LEA proteins (Tunnacliffe & Wise, 2007). Processing of LEA proteins into smaller cleavage products during dehydration has also been reported for Group 3 LEA proteins in

the nematode *Aphelenchus avenae*, the chironomid *Polypedilum vanderplanki*, and the bdelloid rotifer *Adineta ricciae*, and has been proposed to increase a protein's specific activity. Protein cleavage is not believed to be a characteristic of all LEA proteins expressed in invertebrate species (Tunnacliffe & Wise, 2007).

1.3 GROUP 1 FAMILY OF LEA PROTEINS

1.3.1 CHARACTERISTICS OF GROUP 1 LEA PROTEINS

In accordance with all six defined groups of LEA proteins, those belonging to Group 1 are hydrophilic, heat-stable molecules that counteract the damaging effects of water loss in their host organism (Stacy & Aalen, 1998; Soulages *et al.*, 2002), preventing total cellular desiccation (Swire-Clark & Marcotte, 1999). According to published work by Stacy and Aalen (1998) and a review of LEA protein literature by Tunnacliffe and Wise (2007), the expression of Group 1 LEA proteins has not been previously identified in an animal species. However, proteins belonging to this group have been frequently identified in plant seeds (Kamisuga & Cuming, 2005). It has been suggested that the Group 1 LEA proteins mediate their protective effects by binding and replacing water (Swire-Clark & Marcotte, 1999), thereby minimizing its loss from cells (Soulages *et al.*, 2002), by sequestering ions, by preventing protein aggregation induced by water stress (Tunnacliffe & Wise, 2007), and by preserving enzyme activity after cycles of dehydration and rehydration (Goyal *et al.*, 2005). Interestingly, the potential for DNA binding has also been postulated for the Group 1 LEA proteins (Wise & Tunnacliffe, 2004, Wise 2003).

The hydrophilic nature of these proteins is a product of numerous charged and polar amino acids (Wise & Tunnacliffe, 2004). Group 1 LEA proteins have an abundance of the amino acids arginine, glycine, and glutamate and a moderately strong representation of lysine. Proteins belonging to this group also have an under-representation of the amino acids cysteine, phenylalanine, tryptophan, isoleucine, leucine, and asparagine (Wise, 2003, Wise & Tunnacliffe, 2004).

Members of the Group 1 family of LEA proteins are believed to possess numerous unstructured loop regions (Wise & Tunnacliffe, 2004), forming natively

unfolded proteins. Studies of the p11 Group 1 LEA protein from pea (*Pisum sativum*) revealed the protein to be almost entirely unstructured, with only about 2% of the protein forming an α-helix in water (Tunnacliffe & Wise, 2007). Similar to the other LEA protein groups, it has been postulated that Group 1 LEA proteins may be induced to undergo folding into functional conformations upon exposure to dehydration. Studies of the Group 1 LEA protein rGmD-19 from soybean (*Glycine max*) illustrated stress-induced conformational changes to a more ordered state (Soulages *et al.*, 2002).

Like members of other LEA protein groups, Group 1 LEA proteins in plant seeds are thought to be proteolytically degraded after seed germination, when their protective effects are no longer needed (Kamisugi & Cuming, 2005). Studies of the wheat Em protein, a Group 1 LEA protein, have shown that both the Em protein and its mRNA are rapidly degraded within the first 24 hours of the grain contacting water molecules (Taylor & Cuming, 1993). The degradation of this protein is believed to be a selective process within the cytosol carried out by a cysteine proteinase during the first hours of germination (Taylor & Cuming, 1993). This is paralleled by the finding that germinating wheat embryos have high and increasing levels of proteolytic activity (Taylor & Cuming, 1993). Genes encoding proteins homologous to the Em protein in wheat have been identified in cotton, rice, maize, sunflower, barley, radish, *Arabidopsis*, carrot, and mung bean. These proteins may be expected to experience a similar specific, cysteine proteinase-mediated degradation (Swire-Clark & Marcotte, 1999).

1.3.2 REPEATED AMINO ACID SEQUENCE MOTIF IN GROUP 1 LEA PROTEINS

LEA proteins belonging to Group 1 possess a highly conserved, repeated twenty amino acid motif that is extremely hydrophilic in nature (Stacy & Aalen, 1998, Kamisugi & Cuming, 2005). Group 1 LEA proteins have repeated sequences similar to the following consensus twenty amino acid sequence: GGQTRREQLGEEGYSQMGRK (Cuming, 1999). The degree of repetition of this amino acid motif varies among the proteins belonging to this group and has been found one to five times in currently described members of Group 1 LEA proteins (Swire-Clark & Marcotte, 1999, Stacy & Aalen, 1998). For instance, in Group 1 LEA proteins of the plant species *Arabidopsis*

thaliana, the characteristic hydrophilic motif is repeated up to four times (Stacy & Aalen, 1998). These hydrophilic repeats are expected to be of utmost importance to the function of this group of proteins during stress conditions (Stacy & Aalen, 1998). Studies of Group 1 LEA proteins in barley and cotton revealed that the twenty amino acid motif characteristic of this protein group contains five conserved glycine residues, whereby the sequence motifs are usually separated by duplicate glycine residues. This creates a pattern throughout the amino acid sequence of the protein with repeated hydrophilic domains separated by double glycine residues. These glycine residues are the areas of maximum flexibility of the protein (Espelund *et al.*, 1992).

1.3.3 STRUCTURAL SIMILARITIES AMONG GENES ENCODING GROUP 1 LEA PROTEINS

In order to investigate the general structure of genes encoding Group 1 LEA proteins, a representative gene can be considered. *Physcomitrella patens* is a moss species that can tolerate the loss of 90% of its water content (Frank *et al.*, 2005). Studies of the *PpLEA-1* gene which encodes a Group 1 LEA protein in *P. patens* revealed that this gene is characterized by a 149 bp 5'-untranslated region, a 264 bp sequence encoding the Group 1 LEA protein containing the characteristic Group 1 amino acid sequence motif, and a 166 bp 3'-untranslated region. The *PpLEA-1* gene has a single intron of 192 bp located within the codon for the 36^{th} amino acid, a position conserved among the Group 1 LEA genes of higher plants. In fact, all previously identified genes encoding Group 1 LEA proteins possess an intron in this conserved location (Kamisugi & Cuming, 2005).

Comparisons of Group 1 LEA genes in plant species such as wheat, barley, *Arabidopsis*, carrot, soybean, and mung bean reveal significant similarities in gene structure (Kamisugi & Cuming, 2005). Furthermore, certain LEA proteins belonging to Group 1 display conserved elements in the untranslated regions, with minor deviations throughout the 3'- untranslated region (3'-UTR). Such similarities have allowed for the suggestion that noncoding portions (5'-UTR and 3'-UTR) of the mRNA molecules encoding LEA proteins may have regulatory functions (Espelund *et al.*, 1992).

Other similarities have also been observed throughout examinations of the amino acid sequence of Group 1 LEA proteins. For example, PpLEA-1 is characterized by a sequence of 39 amino acids at the N-terminus of the protein, prior to the start of the characteristic Group 1 LEA protein sequence motif. Group 1 LEA proteins in wheat, barley, *Arabidopsis*, carrot, soybean, and mung bean display a 42 – 43 amino acid domain N-terminal to the characteristic Group 1 amino acid sequence motif (Kamisugi & Cuming, 2005).

1.3.4 EXPRESSION OF GROUP 1 LEA PROTEINS

The expression of Group 1 LEA proteins has been previously identified in plant species and in only one non-plant species, the prokaryote *Bacillus subtilus* (Stacy & Aalen, 1998). Although it is known that Group 1 LEA proteins accumulate during late embryogenesis in response to desiccation, there is some discrepancy in the literature as to the induction of Group 1 LEA protein expression in non-embryonic tissues during later stages of development. Certain researchers attest to the induction of Group 1 LEA gene expression in non-embryonic tissues in response to dehydration or the application of abscisic acid (ABA) (Swire-Clark & Marcotte, 1998). However, others state that the Group 1 family of proteins differs from other LEA protein groups in that the proteins are not physiologically induced in mature plant vegetative tissues despite environmental or hormonal cues (Kamisugi & Cuming, 2005). These researchers believe that Group 1 LEA proteins are expressed in a strictly seed-specific manner (Kamisugi & Cuming, 2005).

In higher plants, Group 1 LEA genes are believed to be transcriptionally activated by the interaction of a transcription factor with a responsive element in the gene's promoter (Espelund *et al.*, 1992). Genes encoding Group 1 LEA proteins are thought to require transcriptional activation by a transcription factor encoded by the *VP1* (maize)/*ABI3* (*Arabidopsis*) gene, which is associated with the induction and breakage of embryonic dormancy. Since this transcription factor is expressed almost entirely in seeds, it is thought to regulate expression of Group 1 LEA proteins in seeds (Kamisugi & Cuming, 2005). However, transcriptional activation of LEA genes has not been fully explained as embryos and vegetative tissues with mutations in the gene encoding the VP1 transcription factor remain capable of a certain degree of LEA protein synthesis

(Espelund *et al.*, 1992). The induction of LEA gene expression may be further complicated by the suggestion that different signal transduction pathways may function in embryos and vegetative tissues for the expression of LEA proteins in response to abscisic acid, osmotic stress, and water stress. Furthermore, transcriptional and post-transcriptional events are expected to be involved in the regulation of LEA protein expression (Espelund *et al.*, 1992).

1.4 BRIEF REVIEW OF OTHER PROMINENT LEA PROTEIN GROUPS

Group 2 LEA proteins, also called dehydrins, are characterized by the appearance of at least two of three conserved sequence motifs. The K segment is a highly conserved, lysine-rich fifteen amino acid stretch with the consensus sequence: EKKGIMDKIKEKLPG (Close, 1997). This sequence may appear once in a single polypeptide or may appear in eleven repeated units and is thought to form amphipathic α-helices (Ohno *et al.*, 2003). This motif has also been proposed to be involved in the hydrophobic interactions of Group 2 LEA proteins and partially denatured proteins or membranes (Ohno *et al.*, 2003). A tract of serine residues called an S-segment is another amino acid sequence motif common to Group 2 LEA proteins (Ohno *et al.*, 2003). Another of the potential repeat sequences that characterizes Group 2 LEA proteins is the Y-segment, which has the following consensus sequence: DEYGNP (Close, 1997). The Group 2 LEA proteins are thought to have the most complex domain structure of all the LEA protein groups and have only been identified in plant species (Tunnacliffe & Wise, 2007).

Group 3 LEA proteins are characterized by a conserved sequence motif of eleven amino acids that frequently appears in tandemly repeated units. This repeated motif has the following consensus sequence: TAQAAKEKAXE (Cuming, 1999), where "X" denotes any amino acid. This sequence motif is repeated five times in Group 3 LEA proteins of cotton and thirteen times in the Group 3 LEA proteins of rapeseed (*Brassica napus*) (Wakui & Takahata, 2002). These repeats are thought to be responsible for the formation of amphipathic α-helical structures in these proteins. All previously described LEA proteins in animal species belong to the Group 3 LEA protein family (Kikawada *et al.*, 2006).

The general characteristics and proposed mechanisms of action of LEA proteins described in section 1.2 are applicable to proteins belonging to all of the LEA protein groups. In order to afford protection from dehydration, it is possible that LEA proteins function through interactions with members of differing LEA protein groups. It is also possible that LEA proteins function individually or may mediate their protective effects through individual actions in addition to interactions with other LEA proteins (Swire-Clark & Marcotte, 1999).

1.5 THE STUDY ORGANISM: *ARTEMIA FRANCISCANA*

Commonly known as the brine shrimp, the model organism *Artemia franciscana* is an aquatic crustacean that inhabits harsh and often hypersaline habitats ranging from deserts, to tropics, to mountains (Tanguay *et al.*, 2004). A chitinous exoskeleton covers the body and is shed periodically during molting. Paired, broad, leaf-like appendages radiate from the segmented body and assist in swimming, respiration, osmoregulation, and feeding. Adult males are 8 to 10 mm in length and 4 mm in width, while adult females are usually 10 to 12 mm in length with a 4 mm width (Abatzopoulos *et al.*, 2002).

Dual developmental pathways characterize this organism's complex life history and facilitate survival in fluctuating or adverse environmental conditions. As a bisexual species, mating and fertilization produce zygotes that are held within the female's lateral pouches. During ovoviviparous development, the female releases the offspring as free-swimming larvae that develop within the habitat. Progression through the larval stages to adulthood requires the offspring to experience numerous molts, often ranging from 16 to 20 molting cycles (MacRae, 2003). Oviparous development is characterized by the release of encysted embryos that develop within the habitat until the gastrula stage. At this time, the offspring enter diapause, a genetically programmed, reversible state of dormancy (Liang & MacRae, 1999). Diapause cysts become surrounded by chitinous shells, experience developmental arrest, and decrease their metabolism to extremely low levels. Under favourable water content, molecular oxygen, and temperature, diapause is terminated, allowing the embryo to emerge from the diapause shell (Clegg & Trotman, 2002). Development then resumes, allowing for the progression of the offspring through

the larval stages to adulthood, experiencing the necessary molting stages. The termination of diapause also allows metabolism to increase to physiological levels. The biophysical and biochemical pathways that signal diapause remain poorly understood. However, the developmental pathway of the offspring is determined prior to fertilization. Oocytes destined for oviparous development become yellow in colour and display a brown shell gland, in contrast to green oocytes with an absent shell gland that develop ovoviviparously (Liang & MacRae, 1999).

Encysted embryos of the brine shrimp, *Artemia franciscana* are "among the most stress-resistant of all multicellular eukaryotes to environmental insults" (Warner *et al.*, 2004). Dormant diapause embryos are capable of surviving conditions such as severe desiccation (Clegg & Trotman, 2002), anoxia for a reported duration of 5.6 years, changing salinities, extremes of temperature, variable food reserves, changing aeration, and radiation exposure (Tanguay *et al.*, 2004, Warner & Clegg, 2001). The encysted embryo is also characterized by the synthesis of proteins exclusive to the diapause stage of the life cycle and an absence of detectable protein turnover (Clegg *et al.*, 2000).

One of the proteins exclusive to the diapause stage is p26, a small heat shock/α-crystallin protein. This is a transcriptionally regulated molecular chaperone of proteins, a GTP-binding protein, and a GTPase, conferring stress tolerance to the cysts. p26 is abundant in the encysted embryo, comprising 10-15% of the total non-yolk protein (Warner & Clegg, 2001). Artemin is another protein that is exclusive to the diapause embryo. Believed to function as a molecular chaperone for mRNA, artemin binds mRNA during stressful conditions, providing protection from aggregation or denaturation. Artemin comprises another 10-15% of total non-yolk protein (Warner *et al.*, 2004).

1.6 *ARTEMIA FRANCISCANA* AND LEA PROTEINS

Throughout the duration of this project, two mRNA molecules believed to encode two Group 3 LEA proteins were identified in *Artemia franciscana* by Hand *et al.* (2006). These mRNA molecules were found to be most abundant in diapause and 4 hour post-diapause embryos and minimal in larvae that are incapable of tolerating desiccation (Hand *et al.*, 2006). The encoded proteins each have a predicted molecular weight of 39 kDa and are expected to be strongly hydrophilic. The protein denoted AfrLEA1 is

predicted to display a repeated thirty-nine amino acid sequence, while AfrLEA2 is thought to have a repeated fourteen amino acid sequence motif. Although these repeats appear in tandem and tend to degenerate toward the C-termini of the proteins, they do not match the conserved repeat motifs used to characterize Group 3 or Group 1 LEA proteins (Hand *et al.*, 2006).

1.7 RESEARCH OBJECTIVES

Given the ability of encysted embryos of the brine shrimp, *Artemia franciscana*, to survive severe desiccation, these organisms should possess a molecular mechanism that facilitates dehydration tolerance. The synthesis of proteins exclusive to diapause in *A. franciscana*, such as p26 and artemin, and the widespread expression of late embryogenesis abundant (LEA) proteins in desiccation-tolerant organisms are consistent with the identification of an LEA protein in this organism. The appearance of a previously uncharacterized, abundant protein upon the examination of the protein complement of the encysted embryo of *Artemia franciscana* led to the pursuit of the following objectives:

1. Characterize a 21 kDa protein abundant in encysted embryos of the brine shrimp, *Artemia franciscana*
2. Develop a protocol for the purification of the 21 kDa protein of interest
3. Obtain the DNA sequence of the gene encoding the 21 kDa protein of interest from an *A. franciscana* encysted embryo cDNA library
4. Examine potential interactions between the 21 kDa LEA protein in *Artemia franciscana* and an antibody specific for a Group 3 LEA protein in the nematode, *Aphelenchus avenae*

2. MATERIALS

Pure populations of encysted brine shrimp embryos of the species *Artemia franciscana* were obtained from Sanders Brine Shrimp Company (Ogden, Utah). Polyclonal antibodies to Group 3 LEA proteins in the nematode, *Aphelenchus avenae*, were a gift from Dr. Alan Tunnacliffe, Cambridge University, United Kingdom. The *A. franciscana* encysted embryo cDNA library in the bacteriophage Lambda ZAPII containing the Bluescript phagemid was prepared by Leandro Sastre, Madrid, Spain. Genomic DNA of *A. franciscana* was previously isolated from nauplii by Dr. A.H. Warner, University of Windsor.

3. METHODS

3.1 DEVELOPMENT OF A PROTOCOL FOR THE PURIFICATION OF A 21 kDa PROTEIN FROM ENCYSTED *ARTEMIA FRANCISCANA* EMBRYOS

The following sections describe the sequence of steps that was developed for the isolation and purification of the 21 kDa protein of interest from encysted embryos of the brine shrimp, *Artemia franciscana*.

3.1.1 HYDRATION AND HOMOGENIZATION OF ENCYSTED EMBRYOS OF *ARTEMIA FRANCISCANA*

A 10g sample of dry encysted embryos was combined with 0.25x seawater (26.25mM NaCl, 0.45mM KCl, 1.41mM $MgCl_2$, 1.57mM $MgSO_4$, 0.0844mM $CaCl_2$, 0.030mM $NaHCO_3$) and allowed to hydrate for a minimum of 3 hours at 0 - 4°C. After hydration, distilled water was added to the preparation to dilute the seawater and hydrated cysts were allowed to settle for approximately 5 minutes. Broken shells were removed from the surface of the preparation by aspiration. The preparation was filtered using a glass Pyrex filter and rinsed with distilled water for the removal of residual salt. The sample of hydrated cysts was weighed, obtaining a wet weight of approximately 2.5x the weight of the starting dry cysts.

The preparation of hydrated encysted embryos was homogenized in Buffer K (150mM sorbitol, 70mM K-gluconate, 5mM $MgCl_2$, 5mM KH_2PO_4, 35mM HEPES, pH

7.4) using a motorized mortar and pestle at 10°C. The total volume (mL) of Buffer K added to the homogenized preparation was 5x the mass of the hydrated cyst sample measured in grams. Homogenization was allowed to proceed until a thick slurry was obtained.

3.1.2 CENTRIFUGATION OF HOMOGENIZED CYSTS

The homogenized sample was centrifuged at 12,100 x g for 15 minutes (4°C) to obtain the post-mitochondrial supernatant. Lipids were removed from the surface of the centrifuged sample by suction and the supernatant was removed for subsequent steps. The pellet, including broken shells, nuclei, and yolk platelets, was discarded. The supernatant was centrifuged as described above to remove additional lipids from the surface of the sample. The post-mitochondrial supernatant was subjected to centrifugation at 123,000 x g for 16.5 hours at 4°C. Lipids were removed from the surface by suction, while the flocculent material just below the surface was avoided during the removal of the supernatant. The sediment, containing ribosomes, glycogen and certain dominant proteins, such as artemin, was discarded. This post-ribosomal supernatant was kept on ice for subsequent purification steps.

3.1.3 HEATING THE POST-RIBOSOMAL SUPERNATANT

The post-ribosomal supernatant was heated at 70°C for 8 minutes, with occasional agitation of the test tubes for uniform heat distribution. After heating, the sample was placed on ice for approximately 1 hour to allow for precipitation of heat insoluble materials. This sample was then centrifuged at 12,100 x g for 15 minutes (4°C) to remove heat insoluble materials. The supernatant, containing heat soluble proteins, was saved for subsequent steps.

3.1.4 AMMONIUM SULFATE PRECIPITATION OF PROTEINS

Ammonium sulfate was added to the heat soluble proteins to a concentration of 25% saturation. The preparation was placed on ice for approximately 1 hour, then centrifuged at 12,100 x g for 15 minutes (4°C) to remove any insoluble material. To the 25% ammonium sulfate supernatant, additional ammonium sulfate was added to 75%

saturation. After sitting on ice for approximately 1 hour, the preparation was centrifuged at 12 100 x g for 15 minutes (4°C) to collect the proteins insoluble between 25-75% ammonium sulfate saturation.

3.1.5 DIALYSIS OF 25-75% AMMONIUM SULFATE INSOLUBLE PROTEINS

The 25-75% ammonium sulfate insoluble fraction was suspended in buffer A containing 0.10M NaCl and 0.025M Tris-HCl, pH 7.45 and dialyzed in this suspension buffer for 2 hours in an ice bath using a Slide-a-Lyzer® Dialysis Cassette (Pierce) with a 10,000 molecular weight cutoff. This dialysis step allowed for the removal of excess ammonium sulfate from the re-suspended sample.

3.1.6 HEAT TREATMENT AFTER AMMONIUM SULFATE CONCENTRATION AND DIALYSIS

After dialysis, the 25-75% ammonium sulfate fraction was heated again, this time at 80°C for 10 minutes, with occasional agitation. The sample was chilled on ice for 1 hour, then centrifuged at 12 100 x g for 15 minutes (4°C). The supernatant, containing heat soluble material was retained for further purification.

3.1.7 SIZE EXCLUSION CHROMATOGRAPHY ON A SEPHADEX G-150 COLUMN

A column of Sephadex G-150 (SF) (72cm x 1.5cm) was equilibrated with buffer A (0.10M NaCl and 0.025M Tris-HCl, pH 7.45). To calibrate the column, several proteins of known molecular weight were applied to the column and eluted with the buffer. The standard proteins were bovine serum albumin (BSA) (67kDa), chymotrypsinogen (25kDa), and ribonuclease A (13.7kDa). The heat soluble proteins of interest were then fractionated on the G-150 column. Fractions of approximately 1.5 mL were collected throughout the elution, and the absorbance of each fraction was analyzed in a UV spectrophotometer at 280 nm. The cumulative volume eluted from the column was also recorded in order to determine the volume at which the protein standards and the protein of interest eluted from the column.

3.1.8 SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS

A 10% polyacrylamide gel was prepared according to the procedure of Laemmli (1970) and the column fractions containing high absorbance readings at 280 nm and therefore abundant protein, were analyzed compared to molecular weight standard proteins (Fermentas). The gel was stained with Coomassie Brilliant Blue (50% methanol, 10% acetic acid, 0.05% Coomassie Brilliant Blue R-250) for 1 hour with agitation, then de-stained using a solution containing 5% methanol and 7.5% glacial acetic acid. The band pattern displayed on the gel allowed for the identification of column fractions with the highest quantity of the 21 kDa protein of interest.

3.1.9 CONTROL OF ENDOGENOUS PROTEASES

The column fractions containing the 21 kDa protein were pooled, and to minimize degradation of the protein of interest by endogenous proteases, the serine protease inhibitor PMSF (dissolved in 100% DMSO) was added to a final concentration of 1mM.

3.1.10 CONCENTRATING SAMPLES CONTAINING THE 21 kDa PROTEIN

Centricon YM-10 Centrifugal Filter Devices (Amicon) with a molecular weight cutoff of 10,000 were used to concentrate the pooled column fractions (ie. 21 kDa protein) to less than 500 μL.

3.1.11 DIALYSIS OF CONCENTRATED 21 kDa PROTEIN

Prior to further purification, the crude 21 kDa protein sample was dialyzed in buffer B (0.025M NaCl and 0.01M Tris-HCl, pH 8.5) for 2 hours at 4°C in a Slide-a-Lyzer® Dialysis Cassette (Pierce). The sample was removed from the chamber and centrifuged at 14,000 x g for 3 minutes in a desktop centrifuge to sediment any insoluble material.

3.1.12 ANION EXCHANGE HIGH PERFORMANCE LIQUID CHROMATOGRAPHY ON MONO Q

Ion exchange columns separate proteins based on differences in the net ionic charge at the given buffer conditions. The Mono Q anion exchange column HR 5/5

(Pharmacia Biotech) has a positively charged matrix. The Mono Q column was equilibrated with buffer C (0.025M NaCl and 0.01M Tris-HCl, pH 8.5) and the elution of protein was carried out with a salt gradient to 1M NaCl buffered with 0.01M Tris-HCl, pH 8.5 at 0.5mL/min. over 60 minutes. As the protein eluted from the column, it was collected in tubes containing the serine protease inhibitor PMSF, for a final inhibitor concentration of 1mM. Each column fraction was analyzed by SDS-PAGE as previously described to identify column fractions containing the 21 kDa protein of interest. Mono Q column fractions with abundant 21 kDa protein were pooled and concentrated to less than 1mL using a freshly hydrated Centricon YM-10 concentrator (Amicon) according to the manufacturer's instructions.

3.1.13 HEATING OF MONO Q EFFLUENT AND CENTRIFUGATION

The concentrated sample from the Mono Q column was heated in a 80°C water bath for 10 minutes as previously described, then placed on ice for 1 hour to allow for the precipitation of any residual heat insoluble proteins, including potential proteases that could degrade the 21 kDa protein of interest. The sample was then centrifuged at 14,000 x g for 3 minutes in a desktop centrifuge to sediment any insoluble materials which were then discarded. The supernatant was saved for further purification steps.

3.1.14 CATION EXCHANGE HIGH PERFORMANCE LIQUID CHROMATOGRAPHY ON MONO S

Prior to chromatography on a Mono S column (HR 5/5) (Pharmacia Biotech), the sample from the Mono Q column was dialyzed as described above but in buffer D (0.025M NaOAc, pH 5.5). After equilibration of the Mono S column with buffer D, the protein was injected onto the column and eluted with a gradient of salt to 0.4M NaCl in the equilibration buffer. Fractions were collected into test tubes containing the serine protease inhibitor PMSF as described for the Mono Q fractions above. The Mono S fractions containing UV absorbing material at 280 nm were subjected to SDS-PAGE as previously described to identify column fractions with the 21 kDa protein and to assess the homogeneity of these fractions. After purification of the 21 kDa protein to near homogeneity on Mono S, and removal of aliquots for SDS-PAGE analysis, the fractions

containing the 21 kDa protein were pooled and concentrated using a Centricon YM-10 concentrator (Amicon), as described above (section 3.1.10). The concentrated sample was adjusted to contain 1mM PMSF then stored at –20°C.

3.2 EXAMINATION OF THE DNA SEQUENCE ENCODING THE POTENTIAL LEA PROTEIN

The following sections describe experimental protocols leading to the determination of the DNA sequence encoding the 21 kDa protein of interest from an *Artemia franciscana* embryonic cDNA library.

3.2.1 TRYPSIN DIGESTION, MATRIX-ASSISTED LASER DESORPTION / IONIZATION (MALDI) MASS SPECTROMETRY, AND DE NOVO SEQUENCING

A sample of the most pure preparation of the 21 kDa protein of interest was analyzed in the laboratory of Dr. Panayiotis Vacratsis, University of Windsor. The protein was digested into fragments using trypsin, which cleaves on the carbonyl side of lysine and arginine. The trypsin fragments were subjected to MALDI mass spectrometry. In this technique, protein fragments migrate within an electric and/or magnetic field as a function of their mass-to-charge ratio, allowing the mass of the protein to be deduced with high precision (Nelson & Cox, 2000). Amino acid sequences of various peptides were obtained from the MALDI mass spectrometry data. These sequences were placed into an NCBI BLAST search (http://www.ncbi.nlm.nih.gov/) for comparisons with other known sequences. An *Artemia* codon usage table was used to convert these amino acid sequences to predicted nucleotide sequences. Oligonucleotide primers in the forward and reverse directions were designed according to these predicted base pair sequences for use in polymerase chain reactions.

3.2.2 PRIMER SELECTION AND DESIGN

In order to perform the polymerase chain reaction (PCR), primers were designed based on the available amino acid sequences of selected tryptic peptides. The design of primers was made to ensure that primer pairs had an absence or near absence of

homodimer and heterodimer formation. Primers were also designed so that melting temperatures of primer pairs were as close to one another as possible. The length of each primer was also considered, in order to achieve specificity and product formation. Primers were synthesized by Sigma Genosys.

The T7 promoter primer and the M13reverse primer are complementary to sequences in the pBluescript vector containing the *A. franciscana* embryonic cDNA library. Primers LEA-1F and LEA-1R were deduced from the amino acid sequence provided by trypsin digestion, MALDI mass spectrometry, and de novo sequencing, according to an *Artemia* codon usage table. The LEA-6Ra primer was designed according to the amino acid sequence provided from trypsin digestion, MALDI mass spectrometry, and de novo sequencing data, using codons that appeared most frequently in previously obtained sequencing results. All other LEA primers were designed according to sequence data obtained during subsequent PCR and sequencing steps. The sequences and melting temperatures of the primers used in these experiments are shown in Table 1.

Table 1. Primers used in the polymerase chain reactions and sequencing reactions and their calculated melting temperatures:

Primers to sequences within the pBluescript vector:	***Tm(°C)***
T7 promoter primer: 5'-TTGTAATACGACTCACTATAGGGC-3'	60.0
M13 reverse primer: 5'-CAGGAAACAGCTATGACC-3'	55.5
Primers to LEA sequences within the A. franciscana cDNA library:	
Forward Primers:	
LEA-1F: 5'-GGTCATCAAGGTTATGTTGAAATG-3'	63.0
LEA-3F: 5'-TTTAAACGAAGTTCAAGCGTTCTC-3'	64.2
Reverse Primers:	
LEA-1R: 5'-CATTTCAACATAACCTTGATGACC-3'	63.0
LEA-3R: 5'-GAGAACGCTTGAACTTCGTTTAAA-3'	64.2
LEA-6Ra: 5'-TGCTGCATAGTCTGCTCTTGACATTT-3'	68.7

3.2.3 INITIAL AMPLIFICATION FROM *ARTEMIA FRANCISCANA* EMBRYOMIC cDNA LIBRARY USING THE POLYMERASE CHAIN REACTION

In order to amplify the DNA sequence encoding the potential LEA protein, polymerase chain reactions were performed using the *Artemia franciscana* embryonic cDNA library in the phagemid Bluescript as a template (see Appendix C for pBluescript map). For each of the polymerase chain reactions, a 50μL mixture was prepared containing 2mM each of dATP, dGTP, dTTP, dCTP (Promega), 5μL 10x Mg^{2+}-free buffer (Promega), 2.0mM $MgCl_2$ (Promega), approximately 200ng *A. franciscana* embryonic cDNA in the Bluescript library, 2.5 units *Taq* DNA polymerase (Promega), 20pmol of each of the two primers (Sigma Genosys), and the required volume of distilled water to adjust the reaction volume to 50μL. The reaction mixtures were placed into thin-walled tubes with a drop of mineral oil above the mixture to prevent evaporation. Polymerase chain reactions were performed using a desktop Thermal Cycler (Bio Rad). The primer combination T7 promoter/LEA-1R was used with the following PCR program: 5 minutes at 95.0°C for denaturation, 35 cycles with 45 seconds at 95.0°C for denaturation, 45 seconds at an annealing temperature 4°C below the lower melting temperature listed in Table 1, 2 minutes at 72°C for elongation. These cycles were followed by a final 10 minute elongation step at 72°C. When polymerase chain reactions did not amplify DNA fragments from the template, or if greater specificity of amplification was desired, adjustments were made to the chosen annealing temperature.

3.2.4 AMPLIFICATION OF REGIONS OF THE POTENTIAL LEA GENE FROM GENOMIC DNA OF *ARTEMIA FRANCISCANA* USING POLYMERASE CHAIN REACTIONS

In order to detect potential introns in the gene encoding the 21 kDa protein, the polymerase chain reaction was used to amplify the encoding gene sequence from *Artemia franciscana* genomic DNA and the sequences compared to those amplified from the *A. franciscana* embryonic cDNA library. Polymerase chain reaction mixtures were prepared as previously described (see section 3.2.3) using 200ng *A. franciscana* genomic DNA for the template.

3.2.5 ELECTROPHORESIS OF PCR PRODUCTS ON AGAROSE

To examine the DNA fragments amplified by PCR, the samples were subjected to electrophoresis. To prepare the samples, 10μL of each of the PCR reaction mixtures were combined with 2μL of 6x loading dye (Promega). A 1kb DNA ladder (Fermentas) was also prepared for each gel according to the suggested protocol. Samples were loaded into a 1% agarose gel prepared with 1x TAE (see Appendix A) and allowed to migrate down the gel with an applied voltage of 70-78V. The gel was then stained in 1x TAE containing ethidium bromide or Gel Red (Biotium, Inc.) for 30 minutes of agitation. DNA fragments on the gel were visualized under ultraviolet light and photographed using the ChemiGenius Bio Imaging System (SynGene) with GeneSnap Image Acquisition Software (SynGene).

3.2.6 EXCISION AND PURIFICATION OF DNA FRAGMENTS

When PCR products were visualized, a 1% agarose gel of increased thickness was prepared for electrophoresis of larger volumes of the PCR mixtures. The thicker 1% agarose gel allowed for 20μL to 35μL sample volumes to be electrophoresed under the conditions previously described (see section 3.2.5). After staining and visualization of the gel, bands representing the DNA fragments of interest were excised from the gel using a clean scalpel and transferred to 1.5mL tubes. The DNA fragments were purified using the QIAquick® Gel Extraction Kit (Qiagen) following the suggested protocol. The DNA fragments were eluted with 50μL distilled water adjusted to pH 7.8 with NaOH.

3.2.7 RE-AMPLIFICATION AND ELECTROPHORESIS OF DNA FRAGMENTS

Additional polymerase chain reactions were performed in order to re-amplify the DNA fragments that were generated, increasing the concentration of DNA within the samples. Excised and purified DNA fragments at approximately 200ng or serial dilutions of the initial polymerase chain reaction were used as the template for further polymerase chain reactions. The re-amplifications were prepared according to the conditions previously described (see section 3.2.3), using the same primer combinations and annealing temperatures as for the initial amplification of the DNA fragment. Samples of the PCR products were then subjected to electrophoresis as previously described. The

DNA fragments of interest were excised from the gel and purified using the QIAquick® Gel Extraction Kit (Qiagen), following the suggested protocol. When electrophoresis displayed faint bands for the DNA fragment of interest, the sample was again re-amplified by the polymerase chain reaction, following the same protocol used for initial polymerase chain reactions. Electrophoresis and purification of the DNA fragments could then be performed, in order to obtain a sufficient concentration of the DNA fragment of interest for cloning.

3.2.8 LIGATION OF DNA FRAGMENTS INTO PLASMID VECTORS

Ligation reactions were performed using the Invitrogen TA Cloning® Kit, allowing the DNA fragment of interest to be inserted into the plasmid vector pCR® 2.1 in a region flanked by *EcoRI* restriction enzyme sites (see Appendix D for pCR® 2.1 map). The ligation mixture was prepared by combining 1μL of 10x ligation buffer (Invitrogen), 2μL of pCR® 2.1 vector at 25ng/μL (Invitrogen), the required volume of the purified DNA fragment to achieve a 1:3 vector to insert ratio, 1μL T4 DNA ligase (Invitrogen), and distilled water to a reaction volume to 10μL. This procedure was performed according to the suggested protocol for the TA Cloning® Kit (Invitrogen). The ligation reaction was incubated for a minimum of 4 hours, but preferably for 16-18 hours at 14°C.

3.2.9 TRANSFORMATION OF THE PLASMID VECTOR INTO COMPETENT BACTERIAL CELLS

Transformation using the Invitrogen TA Cloning® Kit allowed the plasmid vector pCR® 2.1 containing the ligated DNA insert to be taken up by One Shot® chemically competent *E. coli* cells (INVαF'). The chemically competent cells were stored at -80°C and thawed on ice prior to use. According to the suggested protocol, the chemically competent *E. coli* cells (INVαF') were incubated and heat shocked in the presence of 2μL of the ligation reaction containing the plasmid vector with the ligated DNA insert, to encourage the cells to take up the plasmid. LB/agar plates (see Appendix A) were prepared containing ampicillin and X-gal at 50μg/mL and 70μg/mL, respectively. Two different volumes of the transformed bacterial cells were spread on each plate which were allowed to remain right-side-up for 30 minutes for absorption of the cells. Plates were

inverted and incubated for approximately 18 hours at 37°C or until colonies had a diameter near 1mm. The LB/agar plates were then transferred to 4°C to allow for proper colour development (blue versus white) in each of the colonies.

Since the pCR® 2.1 plasmid contains an ampicillin resistance gene, transformed *E. coli* cells were capable of growing in the presence of ampicillin. Differences in the colour of the colonies allowed for the determination of bacterial cells containing the plasmid vector alone and those containing the plasmid vector with the ligated DNA insert. Bacterial colonies that appeared blue in colour were avoided, as these cells lacked the DNA insert in the plasmid, having a functional β-galactosidase enzyme and hydrolyzing X-gal within the agar plate. White colonies were re-grown as these cells contained plasmid with a DNA insert that interrupted the Lac Z gene, producing non-functional β-galactosidase enzyme, and lacking the ability to hydrolyze X-gal within the LB/agar plate.

3.2.10 RE-GROWTH OF WHITE COLONIES IN LIQUID CULTURE AND ISOLATION OF PLASMID DNA FROM BACTERIAL CELLS

Adhering to sterile technique, white colonies were grown individually in 2.5mL of LB broth (see Appendix A) containing 150μg ampicillin in a 37°C incubator with agitation for 16 - 18 hours. Cloudiness of the broth indicated bacterial growth. Following the suggested protocol of the Wizard® Plus Minipreps DNA Purification System (Promega), pCR® 2.1 plasmid DNA was isolated from the bacterial cells using 40-50μL of sterile water at pH 7.8 for the elution.

3.2.11 MEASURMENT OF PLASMID DNA CONCENTRATION USING A UV SPECTROPHOTOMETER

To determine the concentration of plasmid DNA isolated from bacterial cells, a 1:200 dilution of each plasmid DNA sample was analyzed at 260 nm and 280 nm in a UV spectrophotometer. Plasmid DNA concentrations were calculated based on the assumption that an absorbance of 1 OD at 260 nm corresponds to a double stranded DNA concentration of 50ng/μL.

3.2.12 RESTRICTION ENZYME DIGESTION OF PLASMID DNA AND ELECTROPHORESIS OF DIGESTED SAMPLES

Due to the presence of *EcoRI* restriction sites flanking the insert sequence in the pCR 2.1® plasmid, each isolated plasmid was digested with the restriction enzyme *EcoRI* in order to confirm the presence of the DNA insert within the plasmid. For each restriction enzyme digestion, a reaction mixture was prepared containing 1μL of 10x Buffer H (Promega), 1μL of the *EcoRI* restriction enzyme (12u/μL) (Promega), 500-600ng of plasmid DNA, and distilled water to adjust the total reaction volume to 10μL. Each reaction was incubated at 37°C for a minimum of 2 hours. Each sample was combined with 2μL of 6x loading dye (Promega), electrophoresed on a 1% agarose gel, and visualized as previously described (see section 3.2.5).

3.2.13 SEQUENCING OF PLASMIDS CONTAINING THE DESIRED INSERT

Plasmid DNA samples containing the DNA insert of expected size according to initial polymerase chain reactions were sent to the Robarts Institute (London, ON) for sequencing. Sequencing results were analyzed by identifying the plasmid's *EcoRI* sites in order to discard the sequence belonging to the vector and obtain the sequence of the ligated DNA fragment.

3.2.14 FURTHER AMPLIFICATION OF THE GENE ENCODING THE 21 kDa PROTEIN OF INTEREST FROM THE EMBRYONIC cDNA LIBRARY USING THE POLYMERASE CHAIN REACTION

Additional polymerase chain reactions were performed as described above (see section 3.2.3) to further amplify DNA sequences of the gene encoding the 21 kDa protein of interest from the embryonic cDNA library of *A. franciscana*. These reactions used primers designed in accordance with sequence data obtained from initial PCR and sequencing steps and from trypsin digestion, MALDI mass spectrometry, de novo sequencing results. These steps followed the same amplification, purification, and sequencing protocols previously described (see sections 3.2.3 – 3.2.13) in order to obtain additional sequence data for the encoding gene (see Appendix E for depiction of PCR strategy).

3.2.15 PERMENENT STORAGE OF POSITIVE CLONES

Clones that contained a DNA insert of expected size were stored in a glycerol stock in order to preserve these cultures for potential future use. These samples were prepared by combining 150μL of sterile 100% glycerol and 850μL of LB media (see Appendix A) containing the cells of interest. The samples were mixed and stored at -80°C.

3.3 SDS-PAGE FOR EDMAN DEGRADATION / N-TERMINAL SEQUENCING ANALYSES

In order to confirm the amino acid sequence at the N-terminus of the protein that was deduced from DNA sequencing results obtained through PCR, a sample of the most pure preparation of the 21 kDa protein was prepared for Edman degradation / N-terminal sequencing. SDS-PAGE was performed as previously described (section 3.1.8), loading 3μg of purified 21 kDa protein per lane. Samples were run in triplicate, with a protein standard (Fermentas) on each side of the gel. SDS-PAGE was performed in 1x Tris-Glycine-SDS buffer with an applied voltage of 150V. After completion of electrophoresis, one protein standard and one sample lane were cut from the gel and stained in Coomassie Brilliant Blue as previously described (see section 3.1.8) for visualization of the bands. The remaining portion of the gel was soaked in transfer buffer (3.0g/L Tris, 14.4g/L glycine, 200mL/L methanol) for 30 minutes with gentle shaking. A PVDF membrane and two pieces of Whatman filter paper (3MM) cut to match the size of the gel were also soaked in transfer buffer for 30 minutes with gentle shaking, after first wetting the membrane in methanol. The transfer apparatus was assembled in the presence of transfer buffer. A current of 250mAmp was applied throughout the 2 hour transfer of protein from the gel to the PVDF membrane.

After completion of the transfer, the transfer apparatus was disassembled and the PVDF membrane was soaked in three changes of 10mM 3-cyclohexylamino-1-propanesulfonic acid, 10% methanol. The PVDF membrane was then soaked in distilled water for 5 minutes, after which the membrane was stained in Coomassie Brilliant Blue (50% methanol, 10% acetic acid, 0.05% Coomassie Brilliant Blue R-250) for 5 minutes, then de-stained in three changes of 50% methanol, 10% acetic acid for a total exposure

time of approximately 7 minutes. The membrane was rinsed of the de-staining agent with distilled water, then allowed to air-dry. The membrane was sent to the National Research Council of Canada's Biotechnology Research Institute, Montreal, Quebec for N-terminal sequencing analyses using the Edman degradation reaction.

3.4 PREPARATION OF A RADIOLABELLED LEA PROBE BY PCR

The polymerase chain reaction was used to prepare a radiolabelled LEA probe of the potential LEA gene. Similar to polymerase chain reactions described above (see section 3.2.3), a 50µL mixture was prepared containing 5µL 10x Mg^{2+}-free buffer (Promega), 2.0mM $MgCl_2$ (Promega), approximately 50ng of the plasmid pCR 2.1® (Invitrogen) containing a LEA-3F/LEA-1R insert, 2.5 units *Taq* DNA polymerase (Promega), and 20pmol each of the primers LEA-3F and LEA-1R (Sigma Genosys). The reaction mixture also contained 2mM each of dATP, dGTP, dTTP and 0.4mM dCTP (Promega). The concentration of dCTP was reduced because 2.5µL of dCTP radiolabelled with ^{32}P (Mandel/NEN) was also added to the reaction mixture, equivalent to 25µCi upon initial usage, for incorporation into the probe. Distilled water was added to a final volume of 50µL. The preparation was placed into a thin-walled tube and covered with a drop of mineral oil to prevent evaporation. The polymerase chain reaction was performed using a desktop Thermal Cycler (Bio Rad). PCR was performed using the following PCR program: 5 minutes at 95.0°C for denaturation, 35 cycles with 45 seconds at 95.0°C for denaturation, 45 seconds at the annealing temperature 59.0°C, 2 minutes at 72°C for elongation. These cycles were followed by a final 10 minute elongation step at 72°C. The radiolabelled probe was purified on a G-50 Sephadex column (1 x 10cm) equilibrated with column buffer containing 10mM Tris-HCl pH 8.0, 50mM NaCl, 1 mM EDTA pH 8.0. Column fractions containing the radiolabelled probe as determined by liquid scintillation counting were stored at -20°C until further use.

3.5 ISOLATION OF BLUESCRIPT PHAGEMID FROM XL1-BLUE MRF' CELLS AND SCREENING OF EMBRYONIC cDNA LIBRARY

The *A. franciscana* embryonic cDNA library in the bacteriophage λZAPII was obtained from the laboratory of L. Sastre, Madrid, Spain. The Bluescript phagemid with the library was previously excised from the λZAPII bacteriophage and used to transform competent XL1-Blue MRF' cells (Butler, A.M., 1999). These cells had been stored at -80°C until a freezer malfunction caused the loss of cell viability. Therefore, the Bluescript phagemid containing the *A. franciscana* embryonic cDNA library was isolated from these nonviable cells, as described below.

3.5.1 ISOLATION OF ADDITIONAL QUANTITIES OF THE BLUESCRIPT PHAGEMID CONTAINING THE EMBRYONIC cDNA LIBRARY USING THE ALKALINE LYSIS METHOD

Nonviable XL1-Blue MRF' cells containing the Bluescript phagemid with the *A. franciscana* embryonic cDNA library were removed from -80°C storage and allowed to thaw. The alkaline lysis method (Sambrook *et al.*, 1989) was used to isolate the Bluescript phagemid from these nonviable cells to obtain sufficient quantities of the library for screening procedures. Three vials of cells containing the phagemid were centrifuged at 12,100 x g for 10 minutes to pellet the cells. The cells were re-suspended in alkaline lysis solution I (50mM glucose, 10mM EDTA, 25mM Tris-HCl, pH 8.0), adding 100μL of the solution for every 1.5mL of the original cell suspension. The sample was vortexed to facilitate re-suspension and mixing. Alkaline lysis solution II (0.2N NaOH, 1% SDS) was added to the preparation in a ratio of 200μL of the solution for every 1.5mL of the original cell suspension. The samples were mixed by inverting the tubes and allowed to sit on ice for 5 minutes. Alkaline lysis solution III (3.0M potassium, 5.0M acetate) was added to the preparation in a ratio of 150 μL of the solution for every 1.5mL of the original cell suspension. The preparation was gently mixed and allowed to sit on ice for 10 minutes before centrifugation at 12 100 x g and 4°C for 5 minutes. The supernatant was removed without disturbing the pellet and transferred to a clean tube. An equal volume of phenol-chloroform-isoamyl alcohol was added to the collected supernatant and mixed well by vortexing. The preparation was centrifuged at 12 100 x g

and 4°C for approximately 5 minutes for phase separation. After the transfer of the upper aqueous layer of the preparation to a clean tube, two volumes of 95% ethanol were added. The preparation was mixed well, allowed to sit at room temperature for 5 minutes, then centrifuged at 12 100 x g and 4°C for 15 - 30 minutes to collect the nucleic acid precipitate. The insoluble precipitate was washed twice with 70% ethanol and residual ethanol was removed by air-drying the pellet. The precipitated DNA was re-suspended in 100 - 200μL of distilled water and analyzed at 260 nm and 280 nm, as described in section 3.2.11. The plasmid preparation was then treated with DNAse-free RNAse at 20 - 50 μg/mL in TE buffer (see Appendix A) for 1 hour, then inactivated with approximately 200 μL of chloroform. The sample was centrifuged for phase separation and the above steps repeated for DNA precipitation.

3.5.2 POLYMERASE CHAIN REACTION FOR CONFIRMATION OF THE PRESENCE OF THE GENE ENCODING THE 21 kDa PROTEIN OF INTEREST IN THE ISOLATED BLUESCRIPT PHAGEMID

In order to confirm the presence of the cDNA of the gene encoding the 21 kDa protein on interest in the newly isolated Bluescript phagemids, the polymerase chain reaction was used to amplify gene sequences from the phagemid. The polymerase chain reaction mixture was prepared as previously described (section 3.2.3), using approximately 260ng of newly isolated *A. franciscana* embryonic cDNA library in the phagemid Bluescript for the PCR template and 20pmol each of primers LEA-3F and LEA-1R. After completion of the polymerase chain reaction, samples were examined by electrophoresis as previously described (see section 3.2.5) for visualization of the DNA fragments amplified from the embryonic cDNA library in the newly isolated Bluescript phagemids.

3.5.3 PREPARATION OF VIABLE, COMPETENT XL1-BLUE MRF' CELLS

To grow XL1-Blue MRF' cells, a sterile wire loop was used to take a sample of viable cells from the surface of a XL1-Blue MRF' cell stock stored at -80°C. This was performed without allowing the cells to thaw from their frozen sate. The sample of cells was combined with 500μL of sterile LB broth and mixed thoroughly. A loopful of this

broth was then used to streak an LB/agar plate. The inoculated plate was allowed to sit right-side-up for approximately 30 minutes to allow for absorption. The plate was then inverted and incubated at 37°C for approximately 18 hours or until the diameter of each colony was near 1mm. Once sufficient colony growth was achieved, the plate was wrapped in parafilm and stored at 4°C until further use.

XL1-Blue MRF' cells were grown in broth culture by combining 2.5mL of sterile LB broth with one colony picked from the XL1-Blue MRF' culture plate. This broth was incubated at 37°C with agitation for approximately 18 hours to achieve sufficient cloudiness, indicating bacterial growth. Since bacterial cells are most susceptible to the uptake of plasmid DNA during the logarithmic stage of growth, the XL1-Blue MRF' cells were re-grown in broth culture to ensure log phase. In a sterile test tube, a 100μL sample of the broth culture was combined with 2.5mL of sterile LB broth and allowed to incubate at 37°C with agitation for approximately 2 hours.

The XL1-Blue MRF' cells were made competent using the procedure described by Tu *et al.* (2005). A 1mL sample of the XL1-Blue MRF' broth culture was centrifuged at 5200 x g for 1 minute for the collection of cells. The supernatant was removed by suction. The cell pellet was re-suspended in 500μL of cold TB solution (see Appendix A), gently mixed, and incubated on ice for 25 minutes. The cells were collected by centrifugation at 5200 x g for 1 minute and the supernatant removed by suction. The newly competent cells were re-suspended in 100μL of cold TB buffer with gentle mixing.

3.5.4 INFECTING COMPETENT XL1-BLUE MRF' CELLS WITH THE BLUESCRIPT PHAGEMID CONTAINING THE EMBRYONIC cDNA LIBRARY

To transform the competent XL1-Blue MRF' cells with the phagemid Bluescript containing the *A. franciscana* embryonic cDNA library, a procedure was followed similar to that described for the transformation of other cell types. A 50μL sample of competent XL1-Blue MRF' cells was combined with 2 - 3μL of the isolated Bluescript phagemid and gently mixed with a pipette tip. The cells were incubated on ice in the presence of the phagemid for 30 minutes, given a heat shock of 42°C for 30 seconds, and placed immediately on ice. S.O.C. medium (250μL) (Invitrogen, see Appendix A) at room temperature was added to the cells and incubated at 37°C for 1 hour with shaking to allow

for phenotypic delay. Samples of transformed cells (50 - 100μL) were spread on LB plates in the presence of 50μg/mL ampicillin. The inoculated plates were allowed to remain right-side-up for approximately 30 minutes to allow for cell absorption to the agar then inverted and incubated at 37°C for 18 hours. After sufficient colony growth, plates were wrapped in parafilm and stored at 4°C until further use.

3.5.5 TRANSFER OF TRANSFORMED BACTERIAL COLONIES TO NITROCELLULOSE MEMBRANES

For each of the culture plates, a nitrocellulose disk labeled with a ballpoint pen was placed atop the bacterial colonies and allowed to sit for 4 - 5 minutes. During this time, India ink was injected in three locations around the perimeter of the nitrocellulose membrane for orientation purposes. The nitrocellulose disks were lifted from the plates using blunt forceps, inverted so that the colony-containing side of the membranes faced up, and placed atop filter paper saturated with 10% SDS. The nitrocellulose disks were allowed to sit for approximately 3 minutes. The disks were then placed on top of filter paper saturated with denaturing solution (1.5M NaCl, 0.5N NaOH) for 2-5 minutes, keeping the same orientation with the colony-containing side of the disks facing up. Next the membranes were allowed to sit on filter paper soaked in neutralizing solution #1 (1M Tris-HCl pH 5.5, 1.5M NaCl) then on filter paper containing neutralizing solution #2 (1M Tris-HCl pH 7.2, 1.5M NaCl) for approximately 5 minutes each. Lastly, the membranes were placed atop filter paper saturated with 2x SSC for 1 - 2 minutes. The nitrocellulose disks were rinsed with 2x SSC and wiped with tissue to remove residual debris. The disks were allowed to air-dry on paper towels for approximately 30 minutes. Culture plates from which bacterial colonies were transferred were inverted and incubated at 37°C for approximately 4 hours to allow for re-growth of colonies.

3.5.6 BAKING NITOCELLULASE MEMBRANES CONTAINING TRANSFERRED BACTERIAL COLONIES

Nitrocellulose disks containing DNA transferred from bacterial colonies were placed between two pieces of Whatman 3MM filter paper, with the transfer side of the membranes facing up. The disks were baked under vacuum at 80°C temperature for

1 - 1.5 hours in a gel dryer (Bio Rad). The membranes were then removed from the filter paper and stored at room temperature in a dark, dry location until further use.

3.5.7 HYBRIDIZATION OF LEA RADIOACTIVE PROBE

Two nitrocellulose membranes were placed in each glass hybridization bottle with the transfer side of the disks facing the interior of the bottle. Into each bottle, 10mL of pre-hybridization solution (6x SSC, 0.3x Denhardts's Solution, 0.1% SDS) was added and the bottles rotated for 1 hour at 64°C in a hybridization oven (Fisher Scientific). The pre-hybridization solution was discarded and replaced with 6x SSC (0.9M NaCl, 0.09M sodium citrate, pH 7.0) for hybridization. The radiolabelled LEA probe was removed from -20°C storage, allowed to thaw, incubated in a boiling water bath for 5-6 minutes to denature the double-stranded DNA probe, then placed immediately on ice. A volume of heat-denatured probe equivalent to $1x10^6$ – $2x10^6$ cpm was added to the bottle. The membranes were allowed to rotate in the hybridization solution containing the radiolabelled LEA probe for approximately 18 hours at 64°C.

Upon completion of the hybridization reaction, the membranes were washed as follows. For the first washing, 20mL of 6x SSC + 1% SDS was added to the bottle and the membranes were allowed to rotate for approximately 30 minutes at 64°C in the hybridization oven. The membranes were then washed sequentially at 64°C with 20mL of 2x SSC + 1% SDS (30 minutes), then with 20mL of 0.2x SSC + 1% SDS (10 - 13 minutes). The nitrocellulose membranes were then covered with plastic wrap. The plastic wrap was folded around the membranes to form a sealed envelope containing the disks. The Geiger-Mueller counter was used to assess the quantity of radiolabelled probe bound to the nitrocellulose membranes and to determine the time of exposure to an X-ray placed with the membranes in an X-ray cassette.

3.5.8 EXPOSURE OF X-RAY FILM TO RADIOLABELLED NITROCELLULOSE MEMBRANES

The X-ray cassette containing the membranes and film was stored at -80°C for 1 – 5 days, depending on the quantity of radioactivity detected on the membrane with the Geiger-Mueller counter. Upon completion of the exposure, the X-ray film was processed

with developer for approximately 4 minutes, briefly washed with stop solution (5% acetic acid), placed in fixer for approximately 8 minutes, then submerged in water and rinsed for approximately 3 minutes. The X-ray was allowed to dry before examination and comparison to the nitrocellulose membranes and culture plates.

3.5.9 ANALYSIS OF EXPOSED X-RAYS

The X-ray was positioned on the bottom side of the culture plate, aligning the India ink markings on the plate to those on the X-ray to identify spots on the X-ray indicative of LEA probe hybridization. Bacterial colonies that correlated with the locations of these dark spots on the X-ray were selected for subsequent screening. Colonies that produced positive signals were re-grown by first combining the chosen colony with 150μL of sterile LB broth. This broth was then spread on a LB/agar plate in the presence of 50μg/mL ampicillin. Inoculated plates were allowed to sit right-side-up for 30 minutes to allow for absorption before inverting the plates and incubating at 37°C for approximately 18 hours.

3.5.10 SECOND SCREENING OF BACTERIAL COLONIES CONTAINING PHAGEMID BLUESCRIPT WITH EMBRYONIC cDNA LIBRARY

Bacterial colonies were transferred to nitrocellulose membranes and baked at 80°C in a gel dryer as previously described (see sections 3.5.5 – 3.5.6). Hybridization was allowed to proceed as previously described, followed by appropriate washing steps. The membranes were wrapped in plastic wrap and processed as described before (section 3.5.7 – 3.5.8). Colonies that produced positive reactions on the X-ray were re-grown in broth culture by combining the selected colony with 2.5mL of sterile LB broth and 50μg/mL ampicillin. The samples were allowed to incubate at 37°C for 18 hours with agitation of the broth.

After sufficient bacterial growth, $1:1x10^6$ dilutions of the broth cultures were made using sterile LB broth and 150 μL samples used to inoculate LB/agar plates in the presence of 50μg/mL ampicillin. The plates were grown as previously described (section 3.5.4). Individual colonies were then transferred to two LB/agar plates with 50μg/mL ampicillin, inoculating approximately $1cm^2$ areas, in order to examine numerous colonies

while reducing the quantity of plates to be screened. The inoculated plates were incubated as previously described (see section 3.5.4).

3.5.11 THIRD SCREENING OF BACTERIAL COLONIES CONTAINING PHAGEMID BLUESCRIPT WITH EMBRYONIC cDNA LIBRARY

After sufficient growth of positive colonies from the second screening, these colonies were transferred to nitrocellulose membranes using the appropriate transfer reagents, as previously described (see section 3.5.5). The membranes were baked at 80°C, prehybridized, and hybridized with $1x10^6$ - $2x10^6$ cpm of the LEA probe, as described for previous screenings (see sections 3.5.6 - 3.5.7). The membranes were washed and an X-ray film was exposed in order to identify positive colonies from the third screening.

3.5.12 ISOLATION OF BLUESCRIPT PHAGEMID CONTAINING EMBRYONIC cDNA LIBRARY FROM POSITIVE AND NEGATIVE CLONES

Three positive colonies and three colonies that did not produce a positive result on the X-ray film were all re-grown in broth culture as previously described (see section 3.5.10). Following the suggested protocol of the Wizard® Plus Minipreps DNA Purification System (Promega), the Bluescript phagemid containing the embryonic cDNA library was isolated from re-grown colonies showing a positive result after the third screening and from colonies that served as negative controls. The concentrations of isolated phagemids were calculated from absorbance measurements at a 260 nm wavelength in a UV spectrophotometer as described previously (see section 3.2.11).

3.5.13 RESTRICTION ENZYME DIGESTION OF PHAGEMID BLUESCRIPT

During the construction of the embryonic cDNA library, cDNA molecules were inserted into the Bluescript phagemid between *XhoI* and *EcoRI* restriction sites (Sastre, personal communication). In order to excise the cDNA of interest from the phagemid, the vector was digested with these restriction enzymes, which digested the phagemid at locations flanking the cDNA insert. These restriction digests were prepared by combining 500 – 600ng of purified phagemid DNA, 1µL of *EcoRI* (12u/µL) (Promega), 1µL of *XhoI* (10u/µL) (Promega), 1µL of 10x Buffer D (Promega), and dH_2O to a final volume of

10μL. This double digestion was allowed to proceed overnight at 37°C as the chosen buffer provided 100% activity for the restriction enzyme *XhoI* and 50-75% activity for the restriction enzyme *EcoRI*. These restriction digests were visualized by electrophoresis on a 1% agarose gel as previously described (see section 3.2.5).

3.5.14 SOUTHERN BLOT ANALYSIS

After electrophoresis in 1x TAE buffer (see Appendix A), the lane of the agarose gel containing the DNA ladder was excised and excess agarose trimmed away using a clean scalpel in order to perform subsequent steps with a gel containing only the restriction digests of the Bluescript phagemid. The gel was soaked in denaturing solution (1.5M NaCl, 0.5N NaOH) for 45 minutes, then rinsed briefly with water. The gel was soaked in neutralizing solution #1 (1M Tris-HCl pH 5.5, 1.5M NaCl) for 20 minutes, then in neutralizing solution #2 (1M Tris-HCl pH 7.2, 1.5M NaCl) for an additional 25 minutes.

The transfer apparatus was assembled by placing a glass plate atop a dish containing transfer buffer (10x SSC, see Appendix A). A long Whatman 3MM filter paper wick was placed across the glass plate, allowing the sides of the wick to be submerged in the transfer buffer. Atop the wick was placed a rectangular piece of filter paper, followed by the inverted agarose gel. The Biodyne B membrane (PALL Life Sciences) was cut to a size matching that of the agarose gel, notched at its lower left corner, moistened with water, and allowed to soak in transfer buffer for 2 minutes. The Biodyne B membrane hydrated in transfer buffer was placed atop the agarose gel, and covered with a piece of filter paper moistened in transfer buffer. A stack of paper towels, cut to match the size of the lower transfer layers, was placed on top of the filter paper to allow absorption of transfer buffer. A small weight was placed on top of the stack of paper towels to apply pressure, facilitating capillary transfer of DNA from the agarose gel to the Biodyne B membrane. The transfer was allowed to proceed for 18 - 20 hours, with saturated paper towels replaced throughout the process. Following the transfer, the apparatus was disassembled and the lanes of the gel were marked on the membrane using a ballpoint pen. The membrane was allowed to air-dry on paper towels for approximately 30 minutes, placed between two pieces of Whatman 3MM filter paper, and baked at 80°C

under vacuum for 1 hour in a gel dryer (Bio Rad). Upon completion of baking, the membrane was stored in a dry, dark place until further use.

3.5.15 HYBRIDIZATION OF BIODYNE B MEMBRANE WITH THE LEA PROBE

The Biodyne B membrane containing the transferred DNA from the restriction digestion was placed in a glass hybridization bottle and covered with 10mL of pre-hybridization solution (6x SSC, 0.3x Denhardt's Solution, 0.1% SDS). The bottle was rotated at 65°C for 30 minutes in the hybridization oven (Fisher Scientific). The solution was replaced with fresh pre-hybridization solution (6x SSC, 0.3x Denhardt's Solution, 0.1% SDS) containing $1x10^6$ – $2x10^6$ cpm of the radiolabelled probe prepared as previously described (see section 3.4). The bottle was allowed to rotate at 65°C in the hybridization oven. After 18 hours, the hybridization solution was discarded, and the membrane was washed first with 10mL of 6x SSC + 0.1% SDS at 65°C for 20 minutes, then with 2x SSC + 0.1% SDS for 20 minutes at 65°C, then with 10mL of 0.2x SSC + 0.1% SDS for 13-15 minutes at 65°C. The Biodyne B membrane was removed from the hybridization bottle and covered with plastic wrap.

3.5.16 EXPOSURE OF X-RAY FILM TO RADIOLABELLED BIODYNE B MEMBRANE

The radiolabelled membrane was placed in an X-ray cassette, covered with a sheet of X-ray film, and stored at -80°C for 1-4 days, depending on the quantity of radiolabelled probe bound to the membrane. After the desired exposure time, the film was developed and analyzed as described above (see section 3.5.8 – 3.5.9). Positive clones should yield a positive signal while negative clones should yield plasmid inserts that do not hybridize with the radiolabelled probe. Positive Bluescript phagemid samples were sent to the Robarts DNA Sequencing facility (London, ON) for analyses.

3.6 EXAMINING POTENTIAL ANTIBODY REACTIONS

The following sections describe the examination of potential interactions between the 21 kDa protein of interest from *A. franciscana* and antibodies specific for a Group 3 LEA protein in nematode (*Aphelenchus avenae*).

3.6.1 WESTERN BLOTTING

Samples of the protein preparation at each stage of the purification protocol were subjected to SDS-PAGE as previously described (section 3.1.8). Duplicate gels were prepared in order for one gel to be stained in Coomassie Brillant Blue and used for comparison. The un-stained gel was soaked in transfer buffer (3.0g/L Tris, 14.4g/L glycine, 200mL/L methanol) for 30 minutes, with agitation. A nitrocellulose membrane large enough to cover the gel, and two similarly sized pieces of filter paper, were also soaked in transfer buffer for 30 minutes. The transfer apparatus was assembled to allow for the movement of the negatively-charged proteins from the gel to the nitrocellulose membrane. Proteins were transferred to the nitrocellulose membrane in the transfer apparatus over 18 hours, with an applied current of 100mAmp.

The nitrocellulose membrane containing transferred proteins was submerged in blocking agent (9% skim milk) for 1 hour with gentle shaking. It was then washed with two 5 - 10 minute cycles of TTBS (see Appendix A) and one 5 - 10 minute cycle of TBS (see Appendix A), with agitation. Next, the membrane was incubated for 1.5 - 2 hours with an antibody specific for a Group 3 LEA protein in the nematode *A. avenae* in TTBS, using a 1:250 antibody dilution, as suggested by the supplier. Washing of the membrane was repeated as previously described. The nitrocellulose membrane was then incubated with agitation for 1.5 - 2 hours in the secondary antibody solution, prepared with TTBS and goat anti-rabbit IgG conjugated with horseradish peroxidase (1:3000 dilution of antibody; BioRad). The membrane was again washed as previously described.

The nitrocellulose membrane was covered with ECL detection reagent (Amersham Bioscienes) and allowed to incubate for 5 minutes. The membrane was placed with an X-ray film in a cassette and exposed for various times, to a maximum of 45 minutes. The X-ray films were developed to detect interactions between the potential

21 kDa LEA protein in *Artemia franciscana* and an antibody specific for a Group 3 LEA protein in the nematode, *Aphelenchus avenae*.

4. RESULTS AND DISCUSSION

4.1 DEVELOPMENT OF A PROTOCOL FOR THE PURIFICATION OF A 21 kDa PROTEIN FROM ENCYSTED *ARTEMIA FRANCISCANA* EMBRYOS

The following sections discuss the development of a protocol for the isolation and purification of the 21 kDa protein of interest, a potential Group 1 late embryogenesis abundant (LEA) protein, from encysted embryos of the brine shrimp *Artemia franciscana*, outlining the importance of the successive steps.

4.1.1 CENTRIFUGATION OF HOMOGENIZED CYSTS

The initial centrifugation of the hydrated and homogenized preparation of encysted embryos produced a visibly layered pellet that included broken shells, yolk platelets, and nuclei. The supernatant was opaque with peach to orange pigment and contained the 21 kDa protein of interest. The second centrifugation of this sample produced a peach to orange coloured pellet containing mitochondrial components and a mainly opaque supernatant bearing peach pigment. Ultracentrifugation of the post-mitochondrial supernatant for 16.5 hours proved to be a beneficial step in the purification of the protein of interest due to the removal of glycogen, ribosomal components, and other dominant proteins, such as artemin, from the sample. These components were collected in a pellet that displayed an orange pigment and a mucoid consistency. The post-ribosomal supernatant was transparent with an orange pigment. All of these centrifugation steps allowed for the removal of lipids, carotenoids, and lipoproteins from the surface of the supernatant.

4.1.2 HEATING STEPS

Heating the post-ribosomal supernatant, the 25-75% ammonium sulfate fraction, and the Mono Q effluent followed by cooling on ice allowed for the denaturation and precipitation of heat insoluble proteins from each preparation. The initial 70°C/8min. heating of the post-ribosomal supernatant allowed for the bulk removal of heat insoluble proteins and the collection of these proteins in a white to peach coloured pellet after centrifugation. The inclusion of the 80°C/10min. heat steps of the 25-75% ammonium

sulfate fraction and the Mono Q effluent were included in the purification protocol in an attempt to denature proteases that were suspected to be degrading the 21 kDa protein of interest in the preparations. The heat soluble 21 kDa protein that was later identified to be the LEA protein was recovered by centrifugation, removing protein that precipitated during the heat treatments.

4.1.3 AMMONIUM SULFATE PRECIPITATION

Precipitation of proteins with ammonium sulfate was a method of obtaining selected protein fractions, while others remained in solution. In order to determine the appropriate ammonium sulfate fraction to use in the 21 kDa protein purification protocol, different fractions were precipitated, re-suspended, and analyzed by SDS-PAGE. It was determined that using the 25-75% ammonium sulfate fraction maximized the quantity of the 21 kDa protein precipitated while excluding unwanted proteins from the sample.

4.1.4 SIZE EXCLUSION CHROMATOGRAPHY: SEPHADEX G-150 SUPERFINE COLUMN

To determine the molecular mass of the protein of interest, a Sephadex G-150 (SF) column (72cm x 1.5cm) was calibrated using proteins of known molecular weight: bovine serum albumin (BSA, 67 kDa), chymotrypsinogen (25 kDa), and ribonuclease A (13.7 kDa). The column fractions from this calibration run were measured in a UV spectrometer at 280 nm to determine their elution positions from the column (see Appendix B).

Size exclusion chromatography via the Sephadex G-150 (SF) column was then used to determine the molecular size of the LEA protein and to achieve initial crude chromatographic separation of the protein preparation. The heat soluble, ammonium sulfate precipitate was applied to the Sephadex G-150 (SF) column and fractions analyzed by SDS-PAGE to identify column fractions containing the 21 kDa protein of interest. These results are shown in Figure 1. The fractions containing the 21 kDa protein were pooled and concentrated for subsequent steps.

Figure 1. Analysis of Sephadex G-150 (SF) column fractions after elution of the sample containing the 21 kDa protein of interest

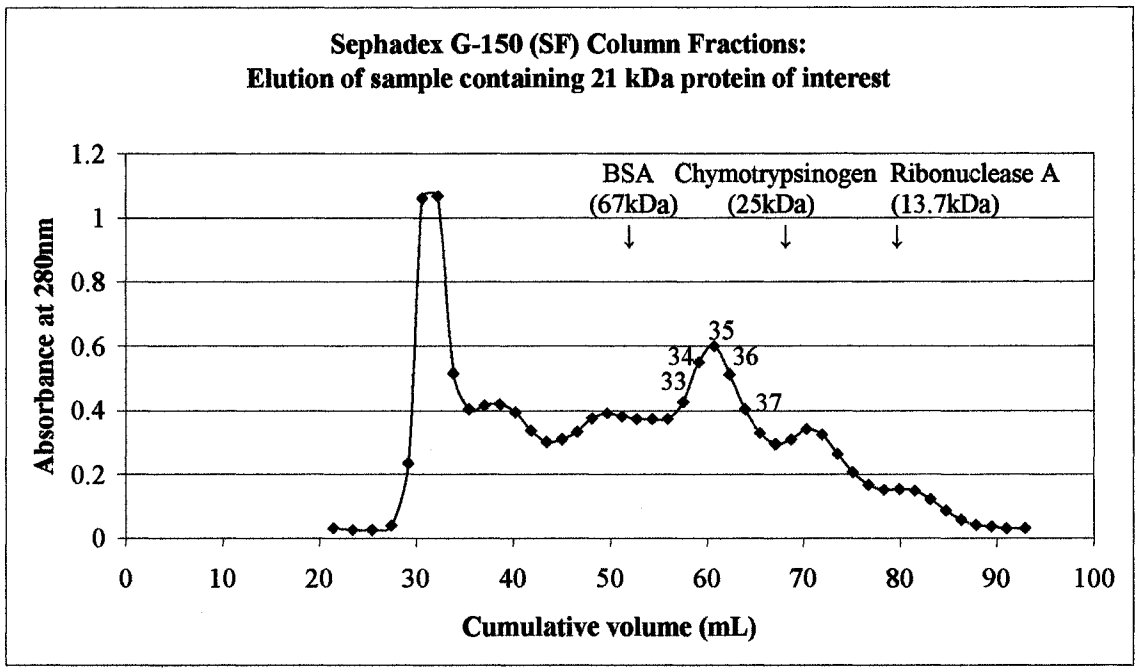
Column fractions eluted from the Sephadex G-150 (SF) column were analyzed by recording the volume of each fraction and the absorbance after measurement in a UV spectrophotometer at 280 nm. Column fractions were analyzed by SDS-PAGE.

Panel A. Graph of the absorbance of each column fraction at 280 nm versus the cumulative volume of each successive fraction. The elution positions of the protein standards bovine serum albumin (BSA, 67 kDa), chymotrypsinogen (25 kDa), and ribonuclease A (13.7 kDa) from the calibration of the column are shown with arrows at the top of the graph. Column fractions 33 - 37 are labelled on the graph as these are the fractions containing the 21 kDa protein of interest.

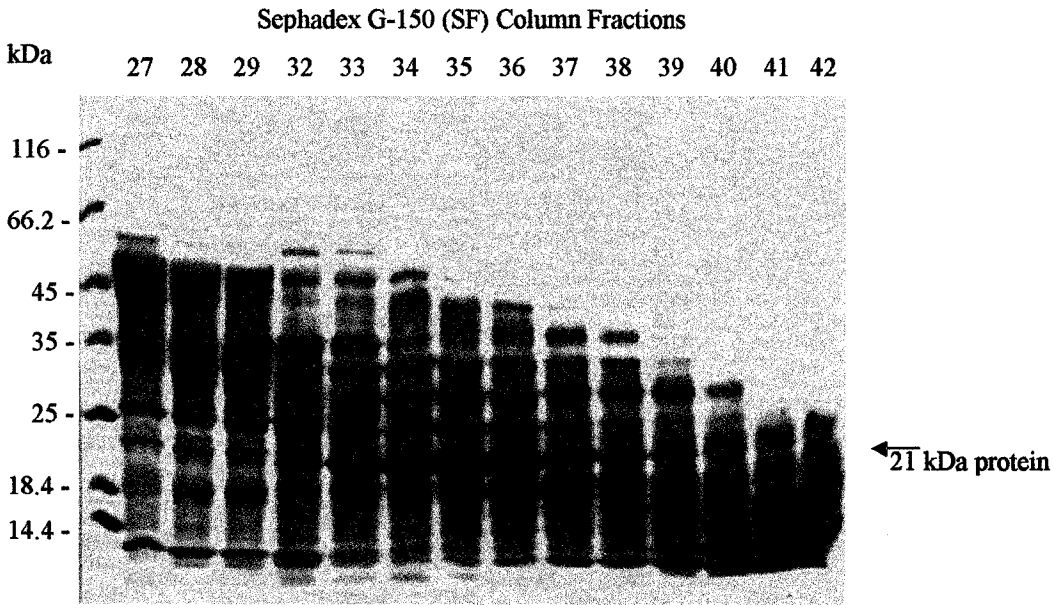
Panel B. SDS-PAGE analysis of column fractions, showing the elution of the 21 kDa protein in fractions 33 - 37. These column fractions were pooled for subsequent steps in the purification protocol. The protein of interest eluted from the Sephadex G-150 (SF) column with an average elution volume of 60.7 mL.

Figure 1.

A.



B.



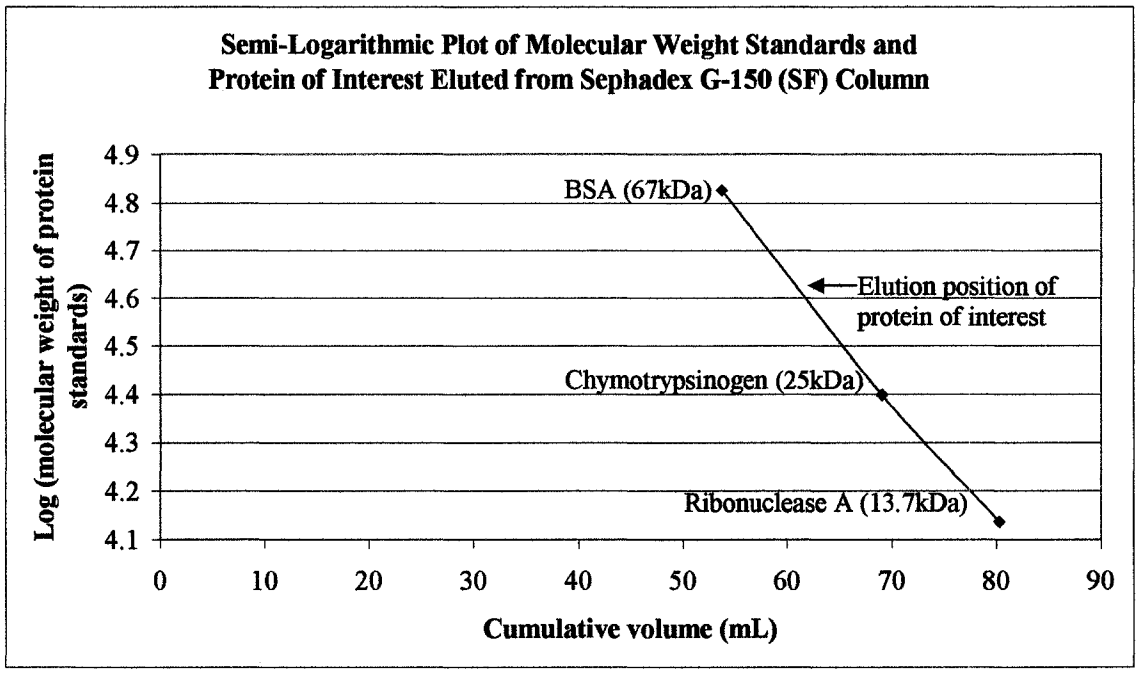
4.1.5 DETERMINING THE MOLECULAR MASS OF THE PROTEIN OF INTEREST BY CHROMATOGRAPHY ON A SEPHADEX G-150 SUPERFINE COLUMN

To determine the molecular mass of the protein of interest, data from the protein molecular weight calibration run were used to compose a standard curve, depicted in Figure 2. Based on the calibration of the column with the protein molecular weight standards, it was initially expected that the protein of interest, which displayed a mass of 21 kDa on SDS-PAGE would have an elution volume of approximately 73 mL on the Sephadex G-150 (SF) column. However, analysis of the column fractions by SDS-PAGE showed the protein of interest to have an elution volume of 60.7 mL. This elution volume (60.7 mL) corresponded to the $\log_{10}$ (molecular weight) of 4.635, which is calculated to be a molecular mass of ~43.0 kDa. This indicates that the protein, with a molecular mass of 21 kDa from SDS-PAGE, is actually a dimer, probably a homodimer, with an approximate molecular weight of 43 kDa.

Figure 2. Standard Curve – Semi-logarithmic plot of protein molecular weight standards eluted from Sephadex G-150 (SF) column

The logarithm_{10} of the molecular weight of each of the protein standards (BSA (67 kDa), chymotrypsinogen (25 kDa), ribonuclease A (13.7 kDa)) were plotted against the cumulative volume at which each of the standards were eluted from the Sephadex G-150 (SF) column, producing a standard curve with the following equation of the trendline: $y = -0.0261x + 6.2189$. When the protein-rich preparation was fractionated using the G-150 (SF) column, the protein of interest eluted from the column at 60.7 mL. According to the standard curve, this elution volume corresponds to a protein with a molecular weight of approximately 43 kDa. Therefore, the protein of interest appears to be a dimer.

Figure 2.



4.1.6 ANION EXCHANGE CHROMATOGRAPHY: MONO Q COLUMN

A Mono Q anion exchange column with cationic groups bound to the column beads was used for the separation of the protein of interest from other proteins in the preparation based on differences in net charge under the given conditions. The sample containing the protein of interest was dialyzed against 0.025M NaCl, 0.01M Tris-HCl pH 8.5, then applied to the Mono Q column equilibrated with the same buffer. The proteins were eluted using a gradient of NaCl in the starting buffer to a 1M final concentration. The proteins eluted from the column were analyzed by SDS-PAGE as shown in Figure 3B. The 21 kDa protein eluted from the column in fractions 4 - 8, having little interaction with the column beads. These fractions were pooled for subsequent steps. Column fractions eluted at later time points were also analyzed by SDS-PAGE and shown to contain other proteins that were not of interest in this study. These fractions were excluded from the pooled sample, effectively separating these proteins from the purified sample.

4.1.7 CATION EXCHANGE CHROMATOGRAPHY: MONO S COLUMN

The cation exchange Mono S column was used for further purification of the 21 kDa protein. The protein from the Mono Q column was dialyzed against buffer of 0.025M NaOAc, pH 5.5, then applied to the Mono S column and eluted with NaCl in the starting buffer to 0.4M. As shown in Figure 4A, the major proteins eluted from the column between 0.20 and 0.25M NaCl. Analysis of the column fractions by SDS-PAGE revealed that the 21 kDa protein eluted in the large peak of the chromatogram shown by the arrow in Figure 4A. SDS-PAGE analyses revealed that the 21 kDa protein of interest had been purified to near homogeneity (Figure 4B). A summary of steps throughout the purification protocol is depicted in Figure 5.

Throughout the development of the protocol, adjustments were made to the buffer system for the Mono S column in an attempt to optimize the isolation and purification of the protein of interest. After altering the concentration of NaCl in the high salt buffer, it was determined that a NaCl gradient from 0 to 0.4M NaCl produced optimal separation. Also, adjusting the pH of the buffer system facilitated the choice of a buffer system at pH 5.5 for optimal protein separation.

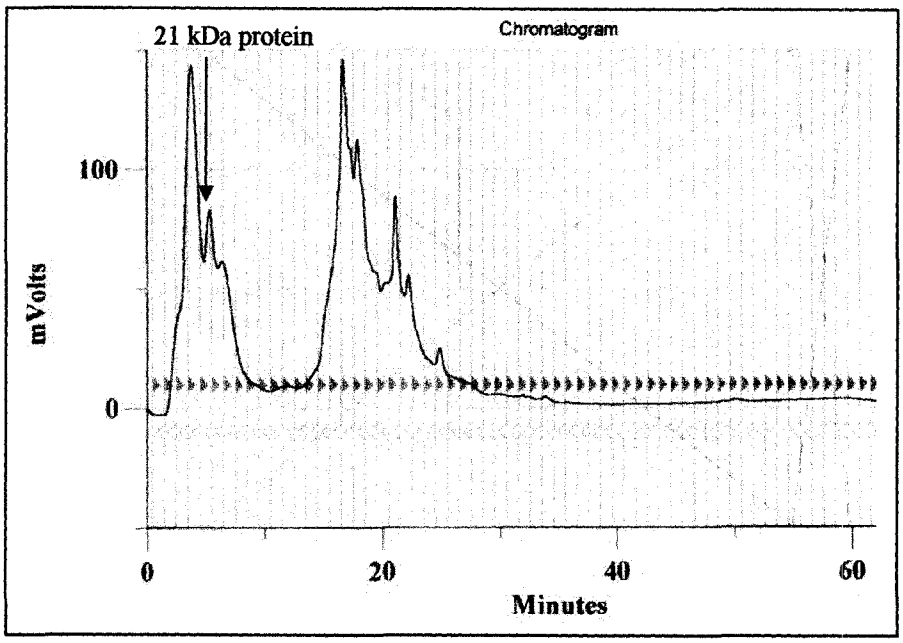
Figure 3. Analysis of Mono Q column effluent after the sample containing the 21 kDa protein from the Sephadex G-150 column was applied to the column

Panel A: Chromatogram representing the fractionation of the protein-containing sample via the Mono Q column. Peaks in the chromatogram represent areas of protein elution. The arrow at the second peak of the chromatogram denotes the elution position of the 21 kDa protein of interest. The arrowheads represent column fractions.

Panel B: SDS-PAGE analysis of Mono Q column fractions. The arrow at the right of the gel indicates the position of the 21 kDa protein which eluted in column fractions 4 - 8. These fractions were pooled for subsequent steps.

Figure 3.

A.



B.

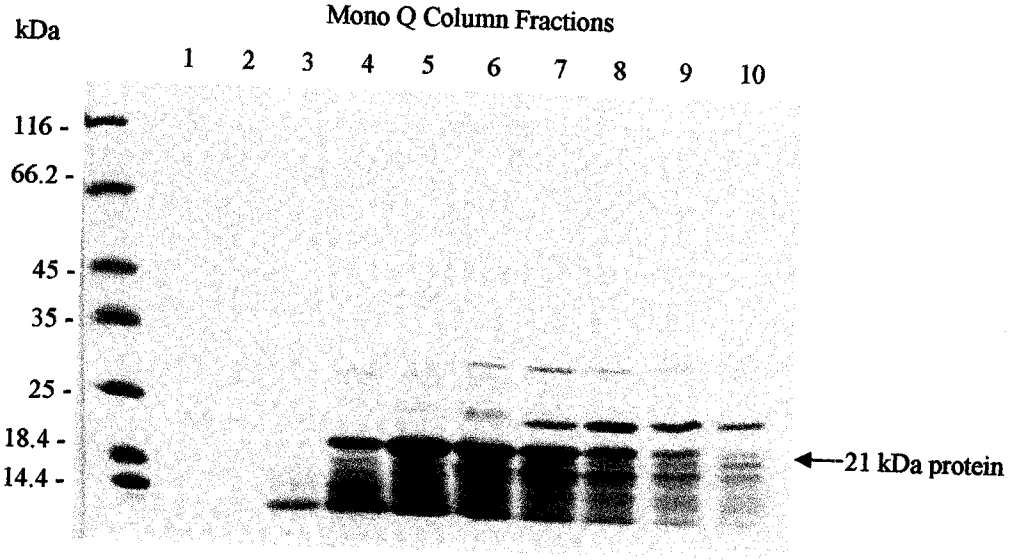


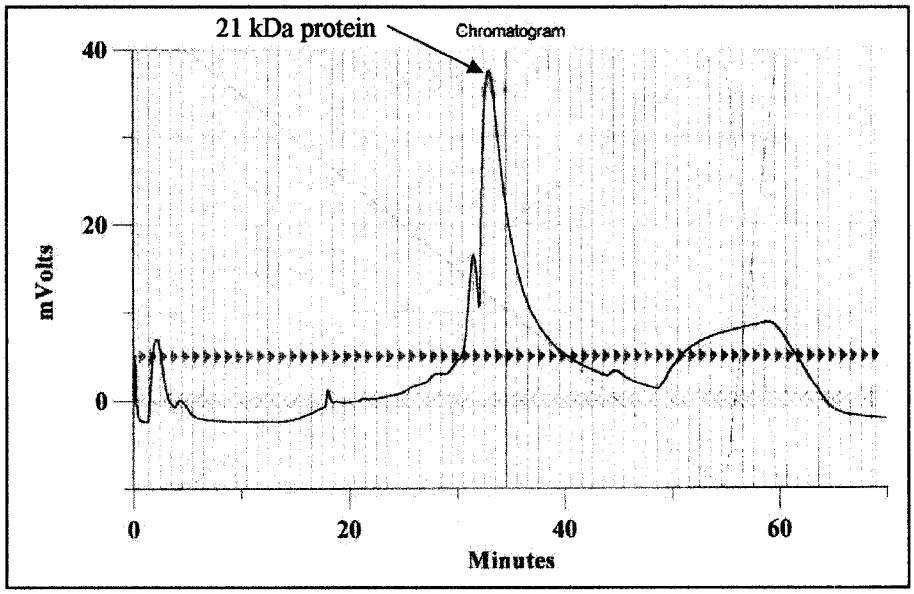
Figure 4. Analysis of Mono S column effluent after the sample containing the 21 kDa protein from the Mono Q column was applied to the column

Panel A: Chromatogram representing the fractionation of the protein-containing sample via the Mono S column. Peaks in the chromatogram represent areas of protein elution. The arrow at the large peak in the chromatogram denotes the elution position of the 21 kDa protein of interest. The arrowheads represent column fractions.

Panel B: SDS-PAGE analysis of Mono S column fractions. The arrow at the right of the gel represents the position of the 21 kDa protein which eluted in column fractions 32 - 34. These fractions were pooled for subsequent steps. As shown in this SDS-PAGE, the 21 kDa protein of interest has been purified to near homogeneity by the established purification protocol.

Figure 4.

A.



B.

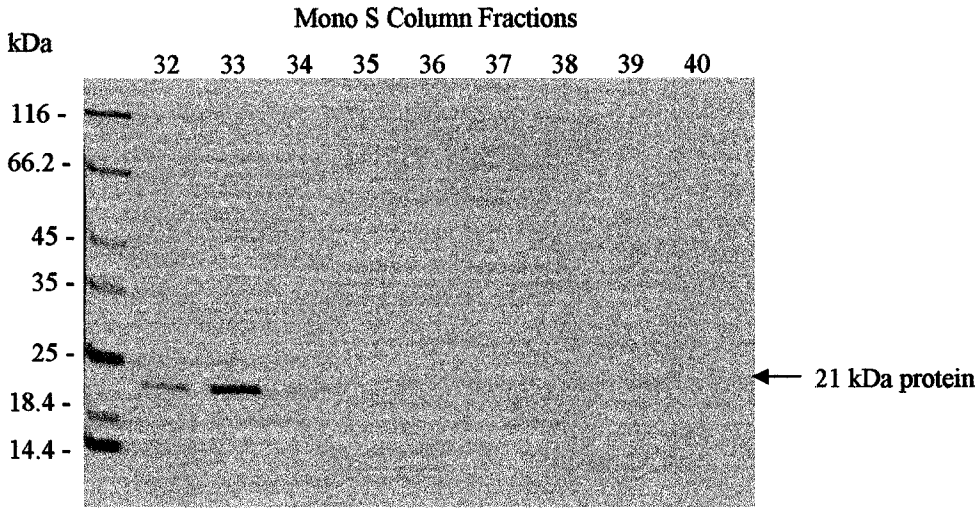


Figure 5. Summary of 21 kDa protein purification protocol

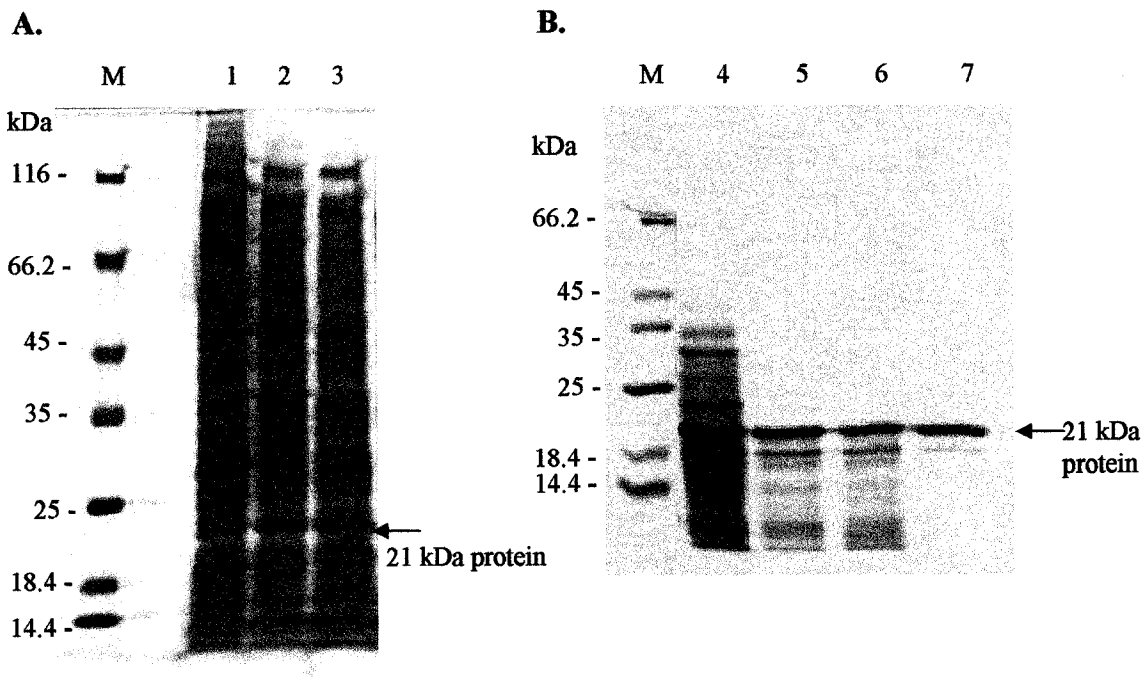
Samples from each stage of the purification protocol were analyzed by SDS-PAGE to assess the purity of the 21 kDa protein of interest with each step of the purification protocol.

Panel A: Protein samples from early stages of the purification protocol (Lane M: protein molecular weight standard, Lane 1: post-mitochondrial supernatant, Lane 2: post-ribosomal supernatant, Lane 3: 70°C/8min. heat soluble proteins).

Panel B: Protein samples from later stages of the purification protocol (Lane M: protein molecular weight standard, Lane 4: Sephadex G-150(SF) fraction, Lane 5: Mono Q fraction, Lane 6: 80°C/10min. heat soluble proteins, Lane 7: Mono S fraction).

With each successive step of the purification protocol shown in Panels A and B, the 21 kDa protein band increased in intensity and purity. At the end of the purification protocol, the 21 kDa protein of interest had been purified to near homogeneity. The less intense, smaller molecular weight band (~19kDa) depicted in the Mono S fraction of Panel B (Lane 7) was shown by MALDI mass spectrometry to represent degradation products of the 21 kDa protein of interest.

Figure 5.



4.2 CONTROL OF PROTEASES DURING 21 kDa PROTEIN PURIFICATION

The following sections discuss experiments to control the degradation of the 21 kDa protein of interest by a heat stable, endogenous protease in the protein preparation.

4.2.1 EXAMINING THE EFFECT OF A SERINE PROTEASE INHIBITOR, PMSF

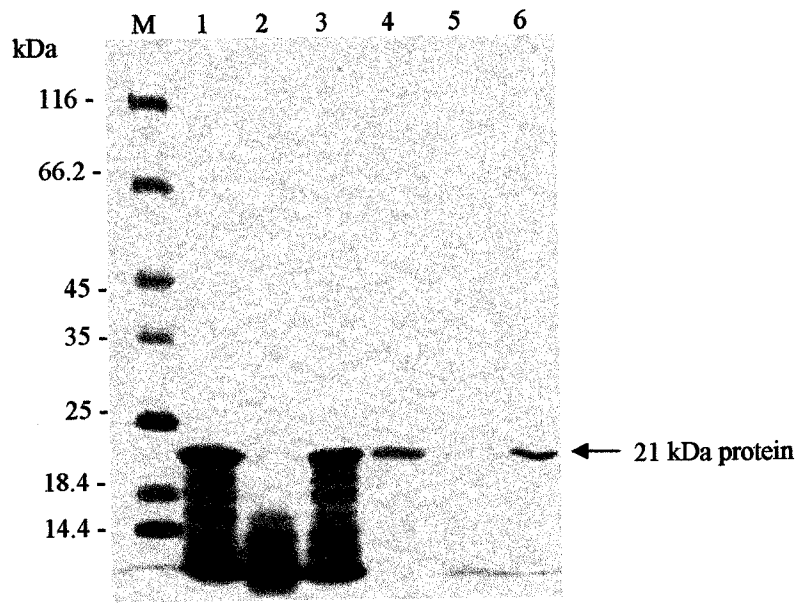
Samples of the Mono Q purified protein were examined after storage in a -20°C freezer, compared with storage overnight at room temperature, and in the presence of 1mM PMSF. When these samples were compared using SDS-PAGE analysis, it was found that all column fractions containing the 21 kDa protein were degraded after overnight incubation at room temperature. Degradation was most severe in the early eluting protein-rich fractions and so severe that the 21 kDa band became nearly absent from the gel. This degradation was not specific for the protein of interest, as other protein bands disappeared from the gel after the incubation period. These results are shown in Figure 6 (compare lanes 1 and 2). When the serine protease inhibitor PMSF was added at 1mM to the Mono Q purified samples and incubated overnight at room temperature, the 21 kDa protein of interest was greatly protected from degradation (see Figure 6, lane 3). Other proteins in the preparation were also protected from degradation. Similar results occurred when samples from the two Mono S trials were subjected to the same treatment conditions. Overnight incubation at room temperature showed significant degradation of the 21 kDa protein, while the addition of the serine protease inhibitor PMSF (1mM) protected the protein of interest from this degradation, producing a band at 21 kDa with an intensity similar to that of the samples stored in the -20°C freezer (compare lanes 4-6, Figure 6).

These assays indicated that the 21 kDa protein was being degraded by a heat-stable serine protease. The protease did not appear to be specific for the protein of interest, as numerous other proteins in the Mono Q preparation were degraded during overnight incubation at room temperature. The serine protease inhibitor PMSF provided significant protection against degradation of the 21 kDa protein by the endogenous protease.

Figure 6. Analysis of protease degradation of the 21 kDa protein and protease inhibition

This SDS-PAGE analysis shows samples containing the 21 kDa protein from the Mono Q and Mono S columns under different treatment conditions (Lane M: protein molecular weight standards, Lane 1: Sample from Mono Q column, storage at -20°C, Lane 2: Sample from Mono Q column, overnight incubation at room temperature, Lane 3: Sample from Mono Q column, overnight incubation at room temperature with 1mM PMSF, Lane 4: Sample from Mono S column, storage at -20°C, Lane 5: Sample from Mono S column, overnight incubation at room temperature, Lane 6: Sample from Mono S column, overnight incubation at room temperature with 1mM PMSF). When compared to the control samples stored at -20°C (lanes 1 and 4), samples incubated at room temperature overnight showed significant degradation of the 21 kDa protein (lanes 2 and 5). When the serine protease inhibitor PMSF was added to the samples to a final concentration of 1mM and the samples incubated at room temperature overnight, the 21 kDa protein was protected from degradative effects of the endogenous protease (lanes 3 and 6).

Figure 6.

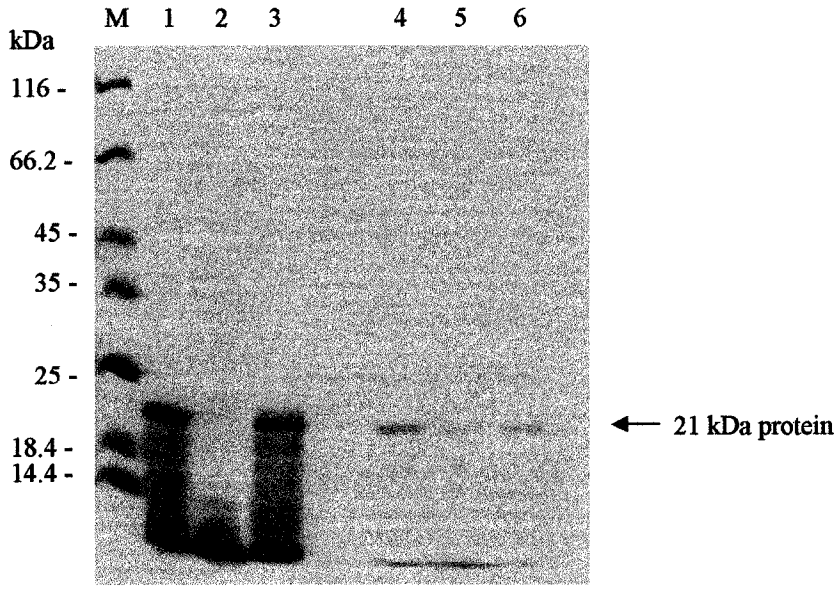


4.2.2 EXAMINING THE EFFECT OF UREA

Protein-rich samples from the Mono Q and Mono S columns were examined by SDS-PAGE after storage at -20°C, after ~48hr. incubation at room temperature, and after ~48hr. incubation at room temperature in the presence of 4M urea. For the Mono Q column preparation, incubation at room temperature showed significant degradation of the 21 kDa protein of interest. As shown in Figure 7 (lane 2), the 21 kDa protein band as well as many other protein bands in the preparation are degraded when incubated in the absence of 4M urea. The addition of 4M urea appeared to protect many proteins from degradation, including the 21 kDa protein (lane 3, Figure 7). Similar results occurred when the Mono S column preparation of the 21 kDa protein was examined. Although the 21 kDa protein band was still visible upon SDS-PAGE after incubation without urea, the band was much less intense when compared to the sample stored at -20°C, indicating protein degradation (compare lanes 4 and 5, Figure 7). Once again, the addition of 4M urea prior to incubation provided some protection of the 21 kDa protein from protease-induced degradation (lane 6, Figure 7).

Figure 7. Analysis of the degradation of the 21 kDa protein and the effect of urea
This SDS-PAGE displays protein samples after fractionation with the Mono Q and Mono S columns and storage under different conditions (Lane M: protein molecular weight standards, Lane 1: Sample from Mono Q column, storage at -20°C, Lane 2: Sample from Mono Q column, ~48hr. incubation at room temperature, Lane 3: Sample from Mono Q column, ~48hr. incubation at room temperature with 4M urea, Lane 4: Sample from Mono S column, storage at -20°C, Lane 5: Sample from Mono S column, ~48hr. incubation at room temperature, Lane 6: Sample from Mono S column, ~48hr. incubation at room temperature with 4M urea). When samples stored at -20°C (lanes 1 and 4) were compared to samples incubated at room temperature for approximately 48 hours (lanes 2 and 5), degradation of the 21 kDa protein was apparent. However, samples incubated at room temperature for approximately 48 hours in the presence of 4M urea (lanes 3 and 6) showed some protection of the 21 kDa protein of interest from endogenous protease degradation.

Figure 7.



4.3 FURTHER INVESTIGATION OF POTENTIAL PROTEIN-PROTEIN INTERACTIONS

The examination of the elution of the protein of interest from the Sephadex G-150 size exclusion column suggested that the protein may be a subunit of a dimeric complex. This possibility was further examined using the Sepharose 12 gel filtration column and high performance liquid chromatography (HPLC) of the purified protein. The following describes the findings from this set of experiments.

4.3.1 EXAMINATION OF THE SIZE OF THE PROTEIN OF INTEREST ELUTED FROM A HPLC GEL FILTRATION COLUMN

The protein standards bovine serum albumin (BSA, 67 kDa), ovalbumin (45 kDa), chymotrypsinogen (67 kDa), and ribonuclease A (13.7 kDa) were used to calibrate the Sepharose 12 gel filtration column. This experiment allowed for the formation of the standard curve shown in Figure 8. Using the Sepharose 12 column, the protein of interest eluted at 27.8 minutes, consistent with a molecular mass of approximately 42 kDa according to the standard curve. Since the protein of interest produced a band at approximately 21 kDa when analyzed by SDS-PAGE, it appears that the 21 kDa monomer is part of a dimeric protein of approximately 42 kDa, confirming results from the Sephadex G-150 (SF) column.

While it seems plausible that two identical 21 kDa proteins become bound to form a homodimer in the physiological state, the possibility of a protein heterodimer cannot be ignored. It is possible that the 21 kDa protein monomer under investigation is bound to a protein with a different function that happens to have a molecular weight near 21 kDa. Since these two proteins would have such a similar molecular weight, they would be difficult to differentiate with SDS-PAGE. The possibility of heterodimer formation is perhaps supported by the appearance of amino acid sequences from trypsin digestion, MALDI mass spectrometry, and de novo sequencing that do not match sequences deduced from polymerase chain reactions. The appearance of weak signals aside from the major amino acid signal from N-terminal sequencing may also indicate the presence of a protein of a different nature complexed with the 21 kDa protein of interest. Proteins with a molecular weight less than 21 kDa visible upon SDS-PAGE and confirmed by trypsin

digestion, MALDI mass spectrometry, and de novo sequencing to be degradation products of the 21 kDa protein may suggest that the protein of interest forms a dimeric complex with an endogenous protease. This protease could perhaps degrade the 21 kDa protein of interest upon rehydration of the encysted embryo and termination of diapause when the potential desiccation-tolerance function of the protein is no longer needed. This rehydration-induced degradation of the 21 kDa protein of interest in *A. franciscana* would mimic the protease-mediated degradation of the Em protein, a Group 1 LEA protein in wheat, that occurs upon rehydration of the grain (Taylor & Cuming, 1993).

Figure 8. Analysis of the molecular weight of the protein of interest eluted from a gel filtration column

Panel A: Standard Curve – Semi-logarithmic plot of protein molecular weight standards eluted from the Sepharose 12 gel filtration column

The logarithm_{10} of the molecular weight of each of the protein standards (BSA (67 kDa), ovalbumin (45 kDa), chymotrypsinogen (25 kDa), ribonuclease A (13.7 kDa)) were plotted against the time at which these proteins eluted from the Sepharose 12 column. This produced a standard curve with the following equation of the trendline:

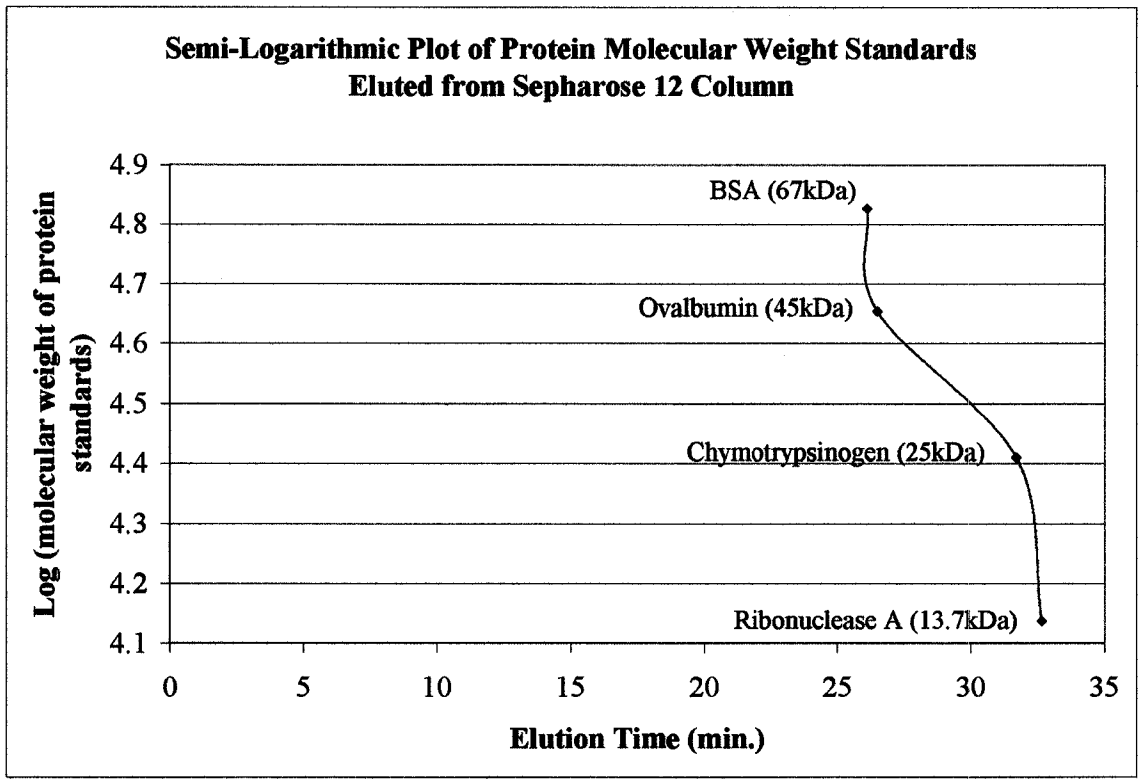
$y = -0.0831x + 6.9351$.

Panel B. Analysis of the elution position of the protein of interest from the Sepharose 12 gel filtration column

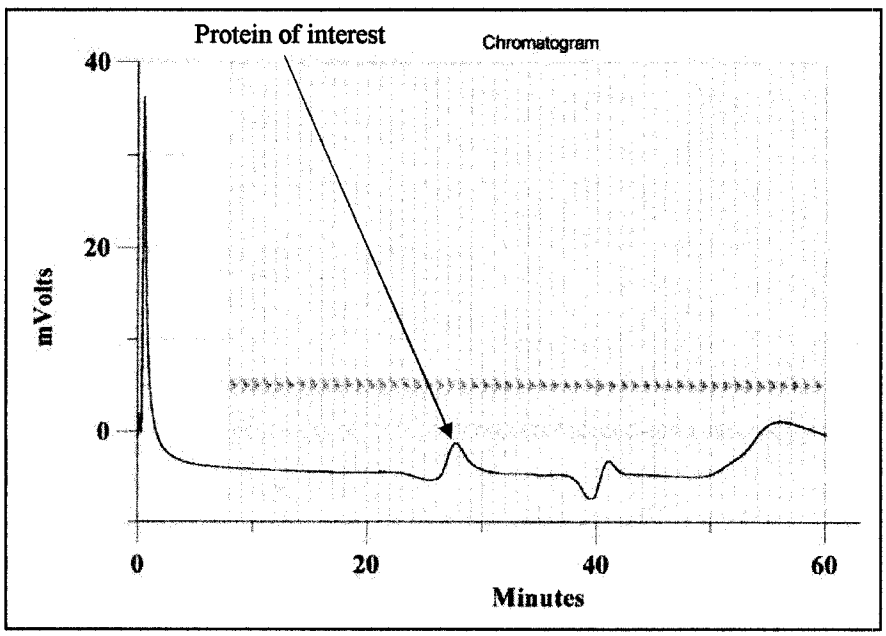
Upon completion of the steps of the established purification protocol, a pure preparation of the protein of interest was placed on the Sepharose 12 column. As shown by the arrow in the chromatogram, the protein under investigation eluted from the column at 27.8 minutes. When this value was placed into the equation of the trendline formed from the standard curve, the protein corresponded to a molecular weight of approximately 42 kDa. The protein of interest was eluted from the column as a homodimer of two identical 21 kDa protein subunits or as a heterodimer of one potential LEA protein of 21 kDa and another protein, potentially a protease, with a similar molecular weight. The smaller peak in the chromatogram that eluted near 41 minutes was confirmed to be devoid of the protein of interest.

Figure 8.

A.



B.



4.4 EXAMINATION OF THE DNA SEQUENCE ENCODING THE 21 kDa PROTEIN OF INTEREST

The following sections discuss the examination of the DNA sequence that encodes the 21 kDa protein on interest in encysted embryos of the brine shrimp, *Artemia franciscana*.

4.4.1 TRYPSIN DIGESTION, MATRIX-ASSISTED LASER DESORPTION / IONIZATION (MALDI) MASS SPECTROMETRY, AND DE NOVO SEQUENCING

Analysis of the 21 kDa protein of interest using trypsin digestion, MALDI mass spectrometry, and de novo sequencing in the lab of Dr. Panayiotis Vacratsis, University of Windsor, provided various amino acid sequences. When placed into an NCBI BLAST search (http://www.ncbi.nlm.nih.gov/BLAST), the fourteen amino acid sequence NH_2- AEQLGHEGYVEMGR -COOH revealed significant similarity to late embryogenesis abundant (LEA) proteins. In particular, this amino acid sequence displayed 86% identity to a Group 1 LEA protein in the moss species *Physcomitrella patens* and 86% identity to an LEA protein in the broomcorn, *Sorghum bicolor*. Furthermore, this fourteen amino acid sequence was 71% identical to the Group 1 LEA protein consensus sequence described by Cuming (1999): NH_2-GGQTRREQLGEEGYSQMGRK -COOH. The forward oligonucleotide primer LEA-1F and the reverse primer LEA-1R were designed according to the nucleotide sequence encoding the above amino acid stretch from analyses of the 21 kDa protein (see Table 1 for details).

Analysis of another trypsin-produced peptide yielded the following amino acid sequence, in which forward slashes indicate the potential presence of either of two amino acids at the given location: NH_2- Q/K Q/K Q/K M S R A D Y A A M R R -COOH. Due to the discovery of numerous repeated regions near the protein's N terminus and the unique nature of the above amino acid sequence, this amino acid sequence was thought to be near the C-terminus of the potential LEA protein. When analyzed via an NCBI BLAST search (http://www.ncbi.nlm.nih.gov/BLAST), this amino acid sequence did not show similarity to any of the LEA proteins. However, the repeated Group 1 LEA protein

sequence motifs may degenerate near the C-terminus of the protein, as described for LEA proteins belonging to Group 3 (Hand *et al.*, 2006). Since the above amino acid sequence is predicted to be near the C-terminus of the protein, it is plausible that the sequence would bear little similarity to the Group 1 LEA consensus motif. The reverse primer LEA-6Ra was designed according to this amino acid sequence, using codons that appeared most frequently in sequence data obtained from previous polymerase chain reactions.

4.4.2 INITIAL AMPLIFICATION AND SEQUENCING OF DNA FROM *ARTEMIA FRANCISCANA* EMBRYOMIC cDNA LIBRARY USING POLYMERASE CHAIN REACTION

The initial amplification of the gene encoding the potential LEA protein from the *A. franciscana* embryonic cDNA library used the T7 promoter for the forward primer and LEA-1R for the reverse primer. The T7 promoter primer was complementary to a sequence within the Bluescript phagemid, the vector containing the embryonic cDNA library. The LEA-1R primer was designed according to the fourteen amino acid sequence provided from trypsin digestion, MALDI mass spectrometry, and de novo sequencing. DNA fragments amplified by PCR using these primers are shown in Figure 9A, with the *EcoRI* restriction digest of isolated plasmids and the excised insert shown in Figure 9B. DNA sequencing revealed that the initial PCR amplification from the embryonic cDNA library produced a 280 base pair DNA fragment with approximately two repeated motif regions of the amino acid sequence: NH_2- GGQARAEQLGHEGYVEMGRK -COOH. These amplified fragments are shown in detail in Figure 9C and summarized in Table 2. The sequence of the LEA-1R primer was within this repeated region, indicating that the sequence provided by MALDI mass spectrometry was a repeated region in the potential LEA gene. The first full-length repeat revealed from these sequencing results was 75% identical to the Group 1 LEA protein consensus motif (NH_2-GGQTRREQLGEEGYSQMGRK -COOH) reported by Cuming (1999), classifying the 21 kDa protein of interest as a Group 1 LEA protein (see Appendix E for depiction of PCR strategy).

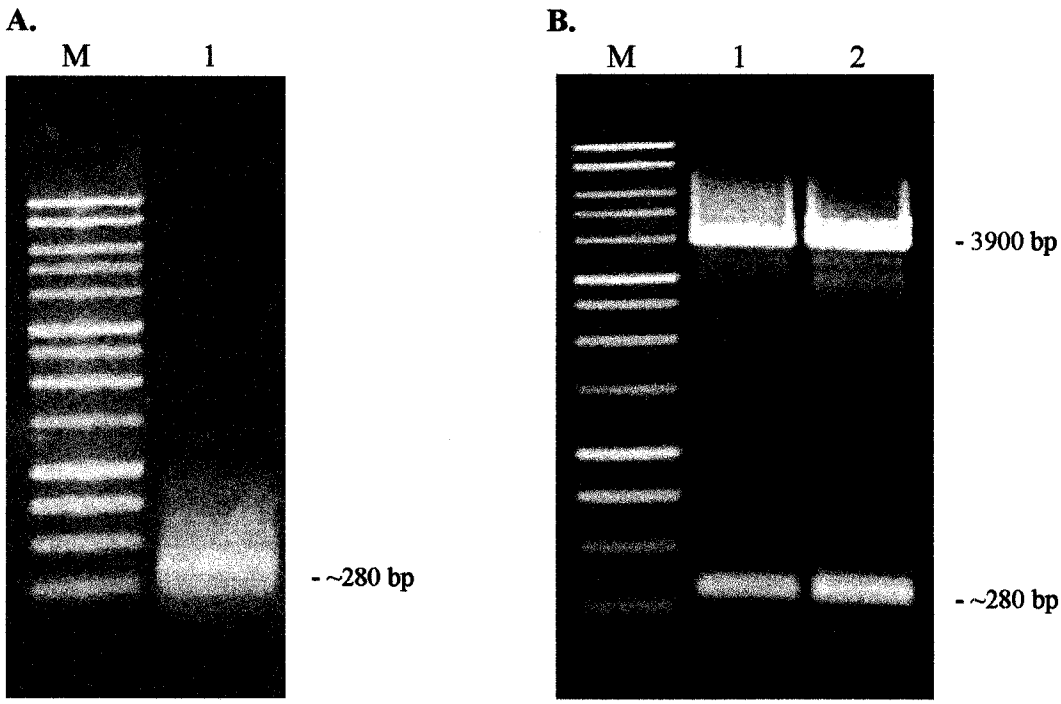
Figure 9. Analysis of PCR products amplified from the *A. franciscana* embryonic cDNA library using the primer combination T7 promoter and LEA-1R

Panel A. Agarose gel electrophoresis of PCR products amplified using the T7 promoter primer complementary to a region of the phagemid Bluescript containing the embryonic cDNA library and the LEA-1R primer deduced from trypsin digestion, MALDI mass spectrometry, and de novo sequencing data (Lane M: 1 kb DNA ladder, Lane 1: PCR products T7 promoter:LEA-1R). An intense band of approximately 280 base pairs was apparent.

Panel B. *EcoRI* restriction digest of the plasmid pCR® 2.1 (Invitrogen) containing the ligated T7 promoter:LEA-1R insert of approximately 280 bp (Lane M: 1 kb DNA ladder, Lanes 1-2: Plasmid DNA containing the T7:LEA-1R PCR product digested with *EcoRI* and separated on a 1% agarose gel, both lanes showing the insert of approximately 280 bp).

Panel C. Nucleotide sequence alignment (ClustalW; http://www.ebi.ac.uk/clustalw) and encoded amino acid sequence of DNA fragments amplified from the embryonic cDNA library with the T7 promoter primer and the LEA-1R primer. The LEA-1R primer sequence is shown in bold and the T7 promoter primer has been removed, as it was complementary to the plasmid vector sequence. Asterisks indicate identity between the two base pair sequences from two different clones. Pink and teal highlighting of the amino acid sequence indicates a repeated sequence motif. Bases shown in red indicate the sequence to which the LEA-3F forward primer was generated for subsequent PCR.

Figure 9.



C.

```
M4_T71R     TTTAAACGAAGTTCAAGCGTTCTCCATTGGCATTTTTATCGTTTATTTGGAAAGAAATAT   60
M5_T71R     TTTAAACGAAGTTCAAGCGTTCTCCATTGGCATTTTTATCGTTTATTTGGAAAGAAATAT   60
            ************************************************************
            F  K  R  S  S  S  V  L  H  W  H  F  Y  R  L  F  G  K  K  Y

M4_T71R     CCATCAAAAATGAGTGAACAGGGAAAGCTAAGCCGCCAAGAAGCTGGACAAAGAGGTGGT  120
M5_T71R     CCATCAAAAATGAGTGAACAGGGAAAGCTAAGCCGCCAAGAAGCTGGACAAAGAGGTGGT  120
            ************************************************************
            P  S  K  M  S  E  Q  G  K  L  S  R  Q  E  A  G  Q  R

M4_T71R     CAAGCAAGGGCTGAACAGCTTGGTCATGAAGGATATGTAGAGATGGGCCGGAAAGGTGGT  180
M5_T71R     CAAGCAAGGGCTGAACAGCTTGGTCATGAAGGATATGTAGAGATGGGCCGGAAAGGTGGT  180
            ************************************************************

M4_T71R     CAAGCTAGAGCAGAACAGTTAGGTCGTCRRGGTTATGTTGARRTG                 225
M5_T71R     CAAGCTAGAGCAGAACAGTTAGGTCGTCRRGGTTATGTTGARRTG                 225
            *********************************************
```

Table 2. DNA fragments amplified from the *A. franciscana* embryonic cDNA library by PCR using the T7 promoter primer and LEA-1R primer

Primer Combination	*Size of sequenced DNA fragment*	*Number of repeated sequence motifs*
T7 promoter / LEA-1R (Clone designation: M4_T71R)	225 bp	2
T7 promoter / LEA-1R (Clone designation: M5_T71R)	225 bp	2

4.4.3 ADDITIONAL AMPLIFICATIONS OF REGIONS OF THE LEA GENE FROM THE EMBRYONIC cDNA LIBRARY USING THE POLYMERASE CHAIN REACTION WITH PRIMERS LEA-3F AND LEA-1R

Based upon sequencing of the DNA fragments amplified with PCR using the primer combination T7 promoter primer:LEA-1R, an additional primer entitled LEA-3F was designed. This primer was complementary to a region near the 5'-end of the LEA gene and did not appear to be a repeated motif in the given sequence. Polymerase chain reactions with the embryonic cDNA library as the template and the primers LEA-3F near the 5'-end of the gene and LEA-1R, the repeat sequence within the LEA gene, yielded several bands of differently sized DNA fragments, as shown on the agarose gel in Figure 10A. Multiple purifications, re-amplifications, ligations, and transformations allowed for the sequencing of numerous plasmids, many with differently sized DNA inserts. The DNA sequences of these amplified products are shown in Figure 10C and listed in Table 3.

The amplification of DNA fragments of at least four different sizes using this LEA-3F:LEA-1R primer combination can be explained by the complementarity of the LEA-1R primer with the repeated sequence motif in the LEA gene that is characteristic of Group 1 LEA proteins. While the LEA-3F primer appears to bind a unique 24 base pair sequence near the 5'-end of the LEA gene, the LEA-1R primer can bind to any of the repeat sequences that appear throughout the LEA gene, amplifying DNA fragments of different sizes. Sequencing of the DNA fragments amplified using this primer combination revealed that the 20-amino acid sequence motif is repeated five times in these amplified fragments. However, it remained possible that the sequence motif appeared in additional repeats. The LEA-1R primer may not have bound to repeats closer to the 3'-end of the LEA gene, preventing the amplification of DNA fragments of increased length. Additional polymerase chain reactions assisted in the elucidation of the number of repeats of the conserved 20-amino acid motif throughout the LEA gene.

Figure 10. Analysis of PCR products amplified from the *A. franciscana* embryonic cDNA library using the primer combination LEA-3F and LEA-1R

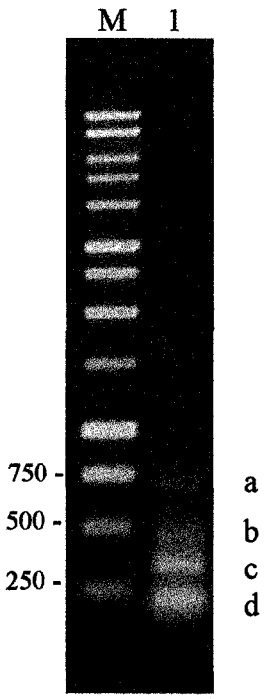
Panel A. Agarose gel electrophoresis of PCR products amplified using the LEA-3F primer complementary to upstream sequence data obtained from previous experiments and the LEA-1R primer deduced from trypsin digestion, MALDI mass spectrometry, and de novo sequencing data and complementary to a repeat region (Lane M: 1 kb DNA ladder, Lane 1: PCR products LEA3F:LEA-1R, DNA fragments a,b,c,d: differently sized DNA fragments amplified using the LEA-3F:LEA-1R primer combination).

Panel B. *EcoRI* restriction digest of plasmids (pCR® 2.1, Invitrogen) containing ligated LEA-3F:LEA-1R inserts of four different sizes (Lane M: 1 kb DNA ladder, Lane 1: uncut control, Lane 2 - 7: Plasmid DNA containing the LEA-3F:LEA-1R PCR products digested with *EcoRI* and separated on a 1% agarose gel. Lanes 2, 5, 6, and 7 contain plasmids with inserts of expected sizes, with labels to match the inserts to the differently sized PCR products shown in Panel A).

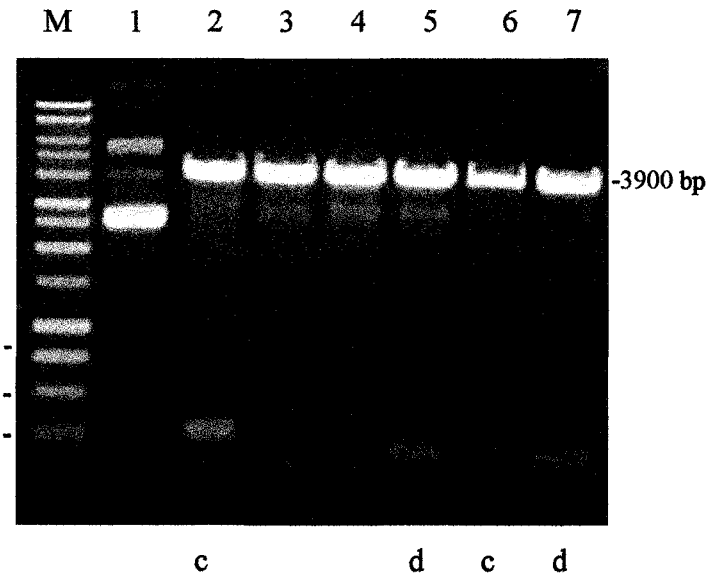
Panel C. *EcoRI* restriction digest of the plasmid pCR® 2.1 (Invitrogen) containing a ligated LEA-3F:LEA-1R insert (Lane M: 1 kb DNA ladder, Lane 1: Plasmid DNA containing a LEA-3F:LEA-1R PCR product digested with *EcoRI* and separated on a 1% agarose gel. This lane shows digested plasmid DNA with a DNA insert corresponding to the PCR product labelled 'b' in Panel A).

Figure 10.

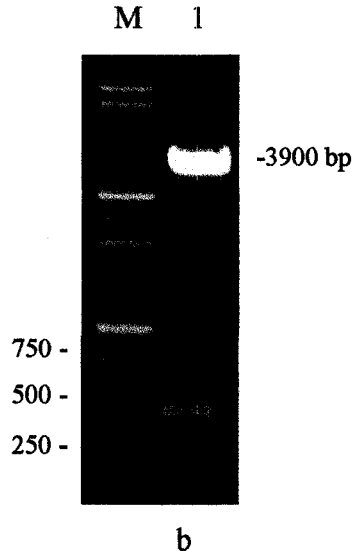
A.



B.



C.



Panel D. Nucleotide sequence alignment (ClustalW) and encoded amino acid sequence of DNA fragments amplified from the embryonic cDNA library with the LEA-3F primer and the LEA-1R primer. Sequences from six different clones are aligned. Primer sequences are shown in bold and asterisks indicate identity between the six base pair sequences from six different clones. Base pairs highlighted in green indicate a discrepancy in the identity of the base pair in one of the aligned sequences. Base pairs highlighted in yellow indicate polymorphic base pair regions, where the aligned sequences differ in the base pair at a specific location. Duplicate or triplicate amino acids listed at a given position indicate the amino acids encoded by polymorphic base pair regions, where the amino acids are listed in order of decreasing frequency. Highlighted amino acids show the 20-amino acid sequence motif, repeated up to 5 times in the longest sequence.

D.

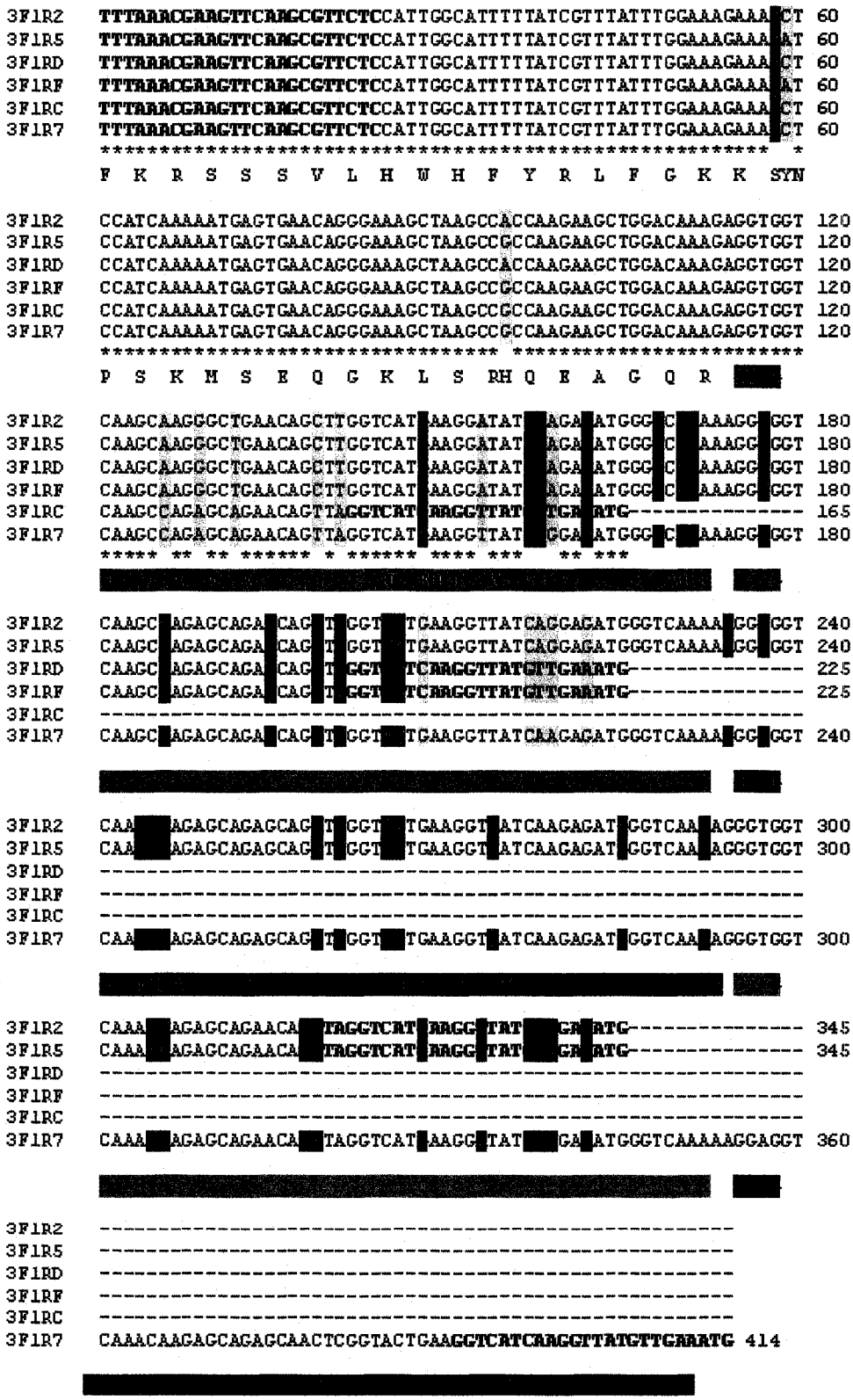


Table 3. DNA fragments amplified from the *A. franciscana* embryonic cDNA library by PCR using the primers LEA-3F and LEA-1R

Primer Combination	*Size of sequenced DNA fragment*	*Number of repeated sequence motifs*
LEA-3F / LEA-1R (Clone designation 3F1RC)	165 bp	1
LEA-3F / LEA-1R (Clone designation 3F1RD)	225 bp	2
LEA-3F / LEA-1R (Clone designation: 3F1RF)	225 bp	2
LEA-3F / LEA-1R (Clone designation: 3F1R2)	345 bp	4
LEA-3F / LEA-1R (Clone designation: 3F1R5)	345 bp	4
LEA-3F / LEA-1R (Clone designation: 3F1R7)	414 bp	5

4.4.4 AMPLIFICATIONS OF REGIONS OF THE LEA GENE USING THE PRIMER COMBINATION LEA-3F / LEA-6Ra

Polymerase chain reactions using the primer LEA-3F deduced from previous sequence data near the 5'-end of the LEA gene and the LEA-6Ra primer thought to be near the 3'-end of the LEA gene and deduced from trypsin digestion, MALDI mass spectrometry, and de novo sequencing results were attempts to obtain the entire sequence of the LEA gene from the embryonic cDNA library. This polymerase chain reaction generated a DNA fragment of 582 base pairs (Figure 11). Sequencing of this amplified fragment revealed that the Group 1 amino acid sequence motif identified in initial amplifications is repeated seven times, as shown in Figure 11C and listed in Table 4. This is not entirely surprising as Hong-Bo *et al.* (2005) have reported repetitions of amino acid sequence motifs in LEA proteins up to thirteen times. From these sequencing results, it appears that the tandem repeats degenerate near the 3'-end of the LEA gene (Figure 11C), as predicted for repeated sequence motifs in other LEA proteins (Hand *et al.*, 2006). Therefore, these amplified DNA fragments confirm the presence of seven repeats of the characteristic Group 1 LEA protein 20-amino acid sequence motif and may suggest a limit to the number of times the motif is repeated throughout the LEA gene. However, these amplified DNA fragments do not extend to the 3'-end of the LEA gene. Additional polymerase chain reactions were performed in attempt to elucidate the 3'-end of the LEA gene sequence and potentially a portion of the 3'-untranslated region.

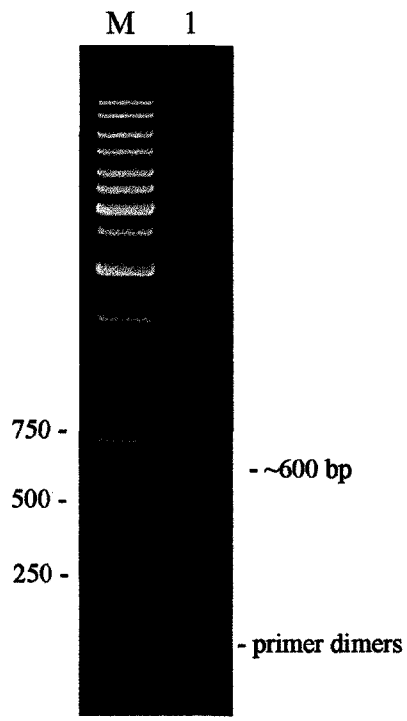
Figure 11. Analysis of PCR products amplified from the *A. franciscana* embryonic cDNA library using the primer combination LEA-3F and LEA-6Ra

Panel A. Agarose gel electrophoresis of PCR products amplified using the LEA-3F primer complementary to sequence data obtained from previous experiments and the LEA-6Ra primer deduced from trypsin digestion, MALDI mass spectrometry, and de novo sequencing data (Lane M: 1 kb DNA ladder, Lane 1: PCR products LEA3F:LEA-6Ra). A DNA fragment of approximately 600 bp was amplified.

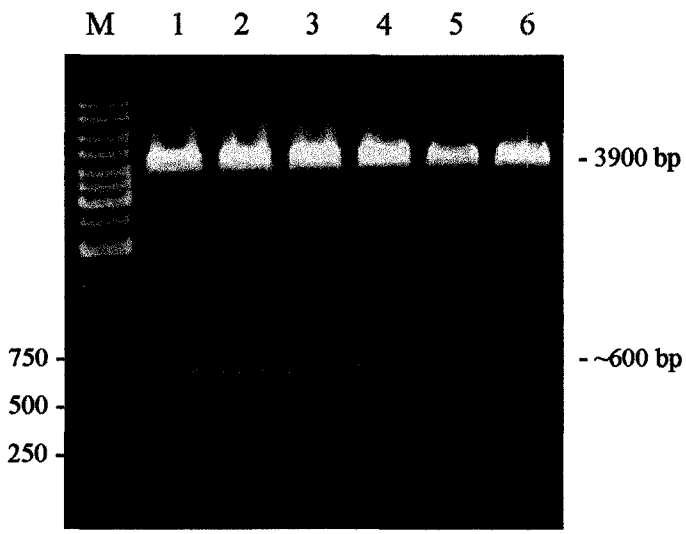
Panel B. *EcoRI* restriction digest of the plasmid pCR® 2.1 (Invitrogen) containing the ligated LEA-3F:LEA-6Ra insert of approximately 600 bp (Lane M: 1 kb DNA ladder, Lane 1 - 6: Plasmid DNA containing the LEA-3F:LEA-6Ra PCR product digested with *EcoRI* and separated on a 1% agarose gel, where all 6 lanes show the excision of the ~600 bp DNA insert).

Figure 11.

A.



B.



Panel C. Nucleotide sequence alignment (ClustalW) and encoded amino acid sequence of DNA fragments amplified from the embryonic cDNA library with the LEA-3F primer and the LEA-6Ra primer. Sequences from three different clones are aligned. Primer sequences are shown in bold and asterisks indicate identity between the three base pair sequences. Base pairs highlighted in green indicate a discrepancy in the identity of the base pair in one of the aligned sequences. Duplicate amino acids listed at a given position indicate the amino acids encoded by different potential codons due to base pair discrepancies, where the amino acids are listed in order of decreasing frequency. An "X" in the amino acid position denotes a stop codon. Highlighted amino acids show the 20-amino acid sequence motif, repeated 7 times in the three sequences.

C.

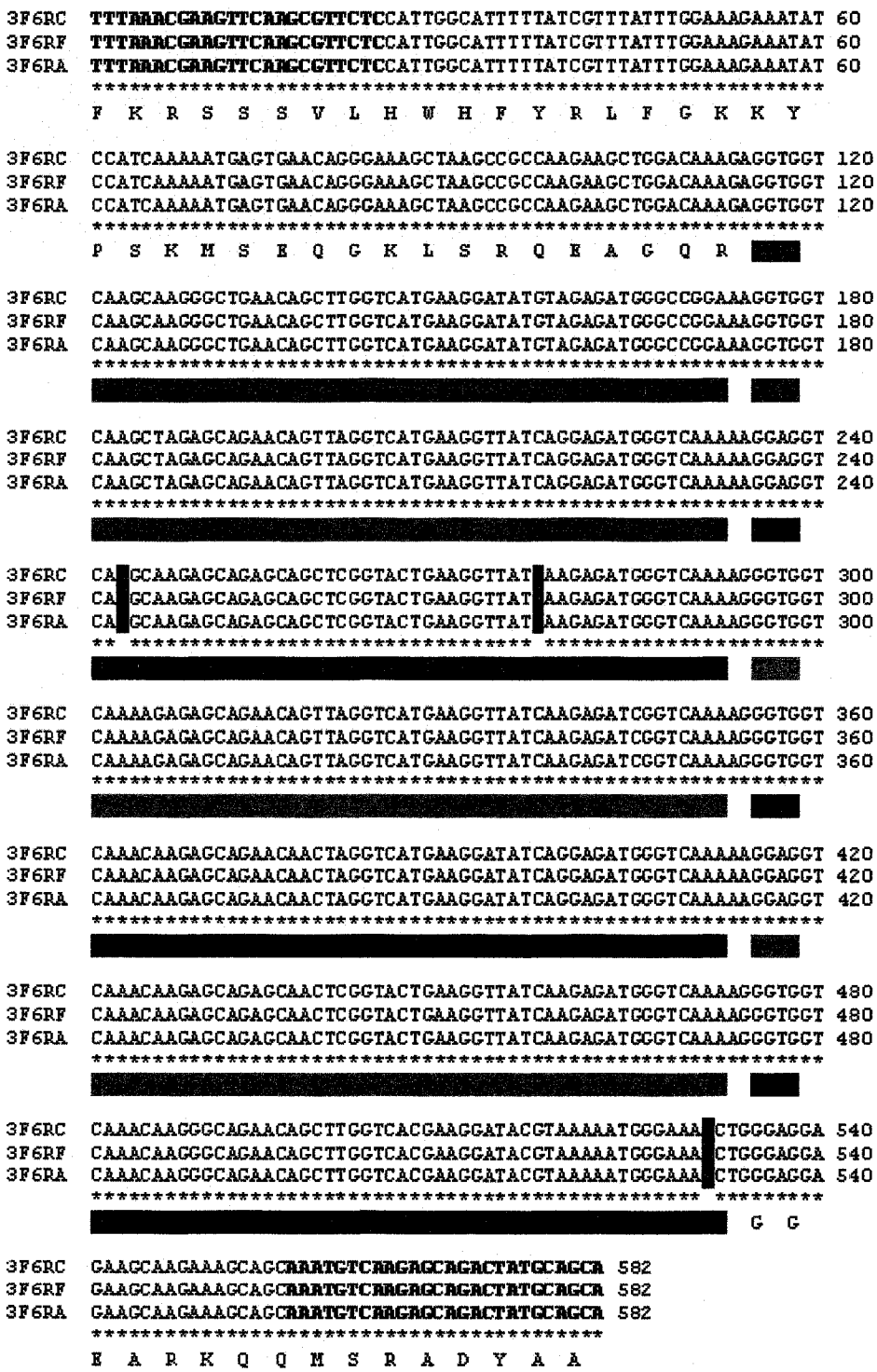


Table 4. DNA fragments amplified from the *A. franciscana* embryonic cDNA library by PCR using the primers LEA-3F and LEA-6Ra

Primer Combination	*Size of sequenced DNA fragment*	*Number of repeated sequence motifs*
LEA-3F / LEA-6Ra (Clone designation: 3F6RA)	582 bp	7
LEA-3F / LEA-6Ra (Clone designation: 3F6RC)	582 bp	7
LEA-3F / LEA-6Ra (Clone designation: 3F6RF)	582 bp	7

4.4.5 AMPLIFICATION OF THE LEA GENE USING THE PRIMER COMBINATION LEA-3F / M13R

Polymerase chain reactions using the LEA-3F primer within the known gene sequence in combination with the M13R primer within the Bluescript phagemid containing the embryonic cDNA library were attempts to amplify the complete LEA gene from the library. PCR using the LEA-3F primer, complementary to sequence data near the 5'-end of the gene, in combination with the M13R primer in the Bluescript phagemid containing the embryonic cDNA library amplified a DNA fragment of approximately 800 base pairs, shown in Figure 12A. This fragment was accompanied by smaller DNA fragments that were much less intense when analyzed by agarose gel electrophoresis (Figure 12A). Attempts to obtain the sequence of this largest, most intense DNA fragment confirmed the LEA gene sequence that had been obtained from previous PCR results and provided additional sequence data for the complete Group 1 LEA gene. Sequencing results from PCR using the primers LEA-3F and M13R revealed the nucleotide sequence within the 5'-untranslated region, the 546 base pair region encoding the LEA gene, the stop codon, the 3'-untranslated region, and the poly-A tail. These results also confirmed the presence of eight repeats of the characteristic Group 1 LEA protein sequence motif, shown in Figure 12C and listed in Table 5. The confirmation of eight repeats throughout the Group 1 LEA gene is reminiscent of the multiple repeat regions reported by Wakui and Takahata (2002) and Hong-Bo *et al.* (2005) for other LEA proteins.

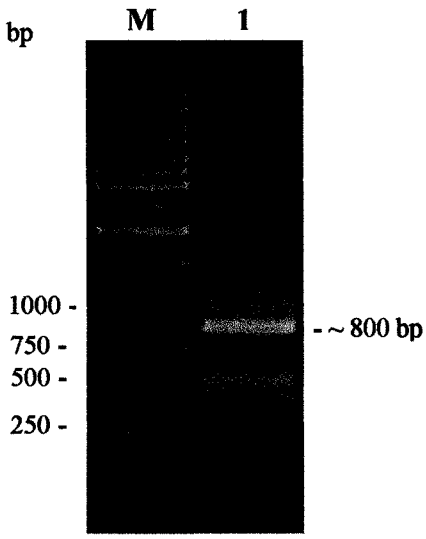
Figure 12. Analysis of PCR products amplified from the *A. franciscana* embryonic cDNA library using the primer combination LEA-3F and M13R

Panel A. Agarose gel electrophoresis of PCR products amplified using the LEA-3F primer complementary to sequence data obtained from previous experiments and the M13R primer complementary to a region of the Bluescript phagemid containing the embryonic cDNA library (Lane M: 1 kb DNA ladder, Lane 1: PCR products LEA3F:M13R). A DNA fragment of approximately 800 bp was amplified.

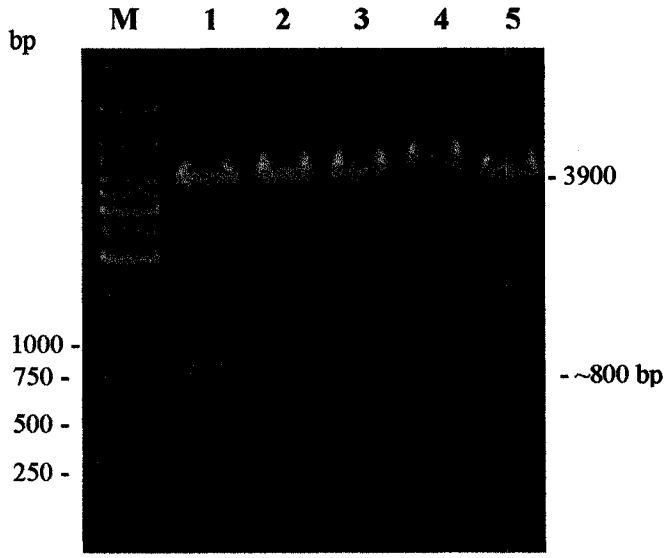
Panel B. *EcoRI* restriction digest of the plasmid pCR® 2.1 (Invitrogen) containing the ligated LEA-3F:M13R insert of approximately 800 bp (Lane M: 1 kb DNA ladder, Lane 1 - 5: Plasmid DNA containing the LEA-3F:M13R PCR product digested with *EcoRI* and separated on a 1% agarose gel, where Lane 1, 2 and 5 show the excision of the DNA insert).

Figure 12.

A.



B.



Panel C. Nucleotide sequence alignment (ClustalW) and encoded amino acid sequence of DNA fragments amplified from the embryonic cDNA library with the LEA-3F primer and the M13R primer. Sequences from two different clones are aligned. The LEA-3F primer sequence is shown in bold and the M13R primer has been removed, as it was complementary to the plasmid vector sequence. Asterisks indicate identity between the two base pair sequences from two different clones. Base pairs highlighted in green indicate a discrepancy in the identity of the base pair in one of the aligned sequences. Duplicate amino acids listed at a given position indicate the amino acids encoded by different potential codons due to base pair discrepancies. An asterisk in the amino acid position denotes a stop codon. Highlighted amino acids show the Group 1 LEA protein 20-amino acid sequence motif, with eight full length repeats shown throughout the gene sequence.

C.

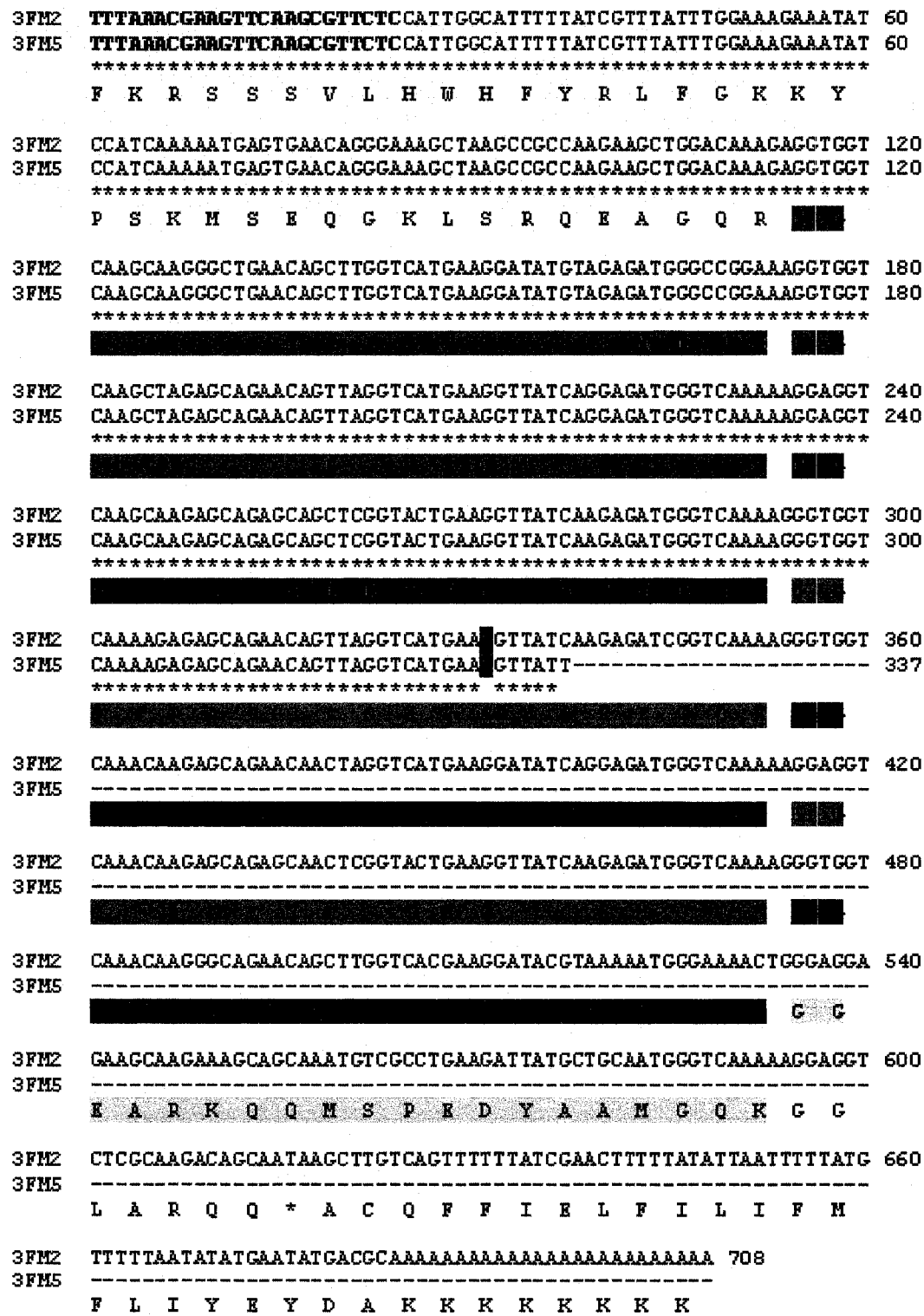


Table 5. DNA fragments amplified from the *A. franciscana* embryonic cDNA library by PCR using the primers LEA-3F and M13R

Primer Combination	*Size of sequenced DNA fragment*	*Number of repeated sequence motifs*
LEA-3F / M13R (Clone designation: 3FM2)	708 bp	8
LEA-3F / M13R (Clone designation: 3FM5)	337 bp	4

4.4.6 AMPLIFICATIONS OF REGIONS OF THE LEA GENE FROM GENOMIC DNA OF *ARTEMIA FRANCISCANA*

Amplification of regions of the LEA gene by PCR using *Artemia franciscana* genomic DNA for the template and the primer combination LEA-3F and LEA-1R generated DNA fragments of three visibly distinct sizes, as shown in Figure 13A. Re-amplification, purification, ligation, and transformation reactions allowed for the sequencing of the amplified DNA fragments. This produced sequence data for 345 and 225 base pair regions of the LEA gene, with four and two repeats of the conserved amino acid sequence motif, respectively (Figure 13C). When compared to the DNA fragments amplified from the embryonic cDNA library, the sequences were identical, except for polymorphic base pair regions. This indicated that throughout the 345 base pair region amplified from the *A. franciscana* genomic DNA, the LEA gene is devoid of introns. The alignment of DNA sequences amplified from the embryonic cDNA library and those amplified from genomic DNA is depicted in Figure 14.

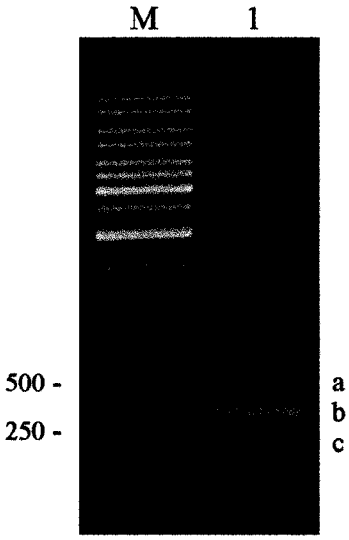
Figure 13. Analysis of PCR products amplified from the *A. franciscana* genomic DNA using the primer combination LEA-3F and LEA-1R

Panel A. Agarose gel electrophoresis of PCR products amplified using the LEA-3F primer complementary to sequence data obtained from previous experiments and the LEA-1R primer deduced from trypsin digestion, MALDI mass spectrometry, and de novo sequencing data and complementary to the repeat region (Lane M: 1 kb DNA ladder, Lane 1: PCR products LEA3F:LEA-1R, DNA fragments a,b,c: differently sized DNA fragments amplified from genomic DNA using the primer combination LEA-3F:LEA-1R).

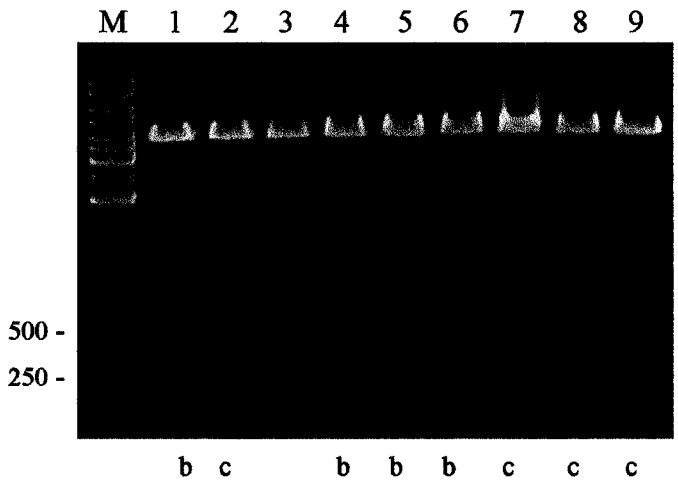
Panel B. *EcoRI* restriction digest of plasmids (pCR® 2.1, Invitrogen) containing ligated LEA-3F:LEA-1R inserts of three different sizes (Lane M: 1 kb DNA ladder, Lane 1-9: Plasmid DNA containing the LEA-3F:LEA-1R PCR products digested with *EcoRI* and separated on a 1% agarose gel. Lanes 1 - 2 and 4 - 9 contain plasmids with inserts of expected sizes, with labels to match the differently sized PCR products shown in Panel A).

Figure 13.

A.



B.



Panel C. Nucleotide sequence alignment (ClustalW) and encoded amino acid sequence of DNA fragments amplified from *A. franciscana* genomic DNA with the LEA-3F primer and the LEA-1R primer. Sequences from four different clones are aligned. Primer sequences are shown in bold and asterisks indicate identity between the four base pair sequences. Base pairs highlighted in green indicate a discrepancy in the identity of the base pair in one of the aligned sequences. Duplicate amino acids listed at a given position indicate the amino acids encoded by different potential codons due to base pair discrepancies, where the amino acids are listed in order of decreasing frequency. Highlighted amino acids show the 20-amino acid sequence motif, repeated four times in these amplified regions.

C.

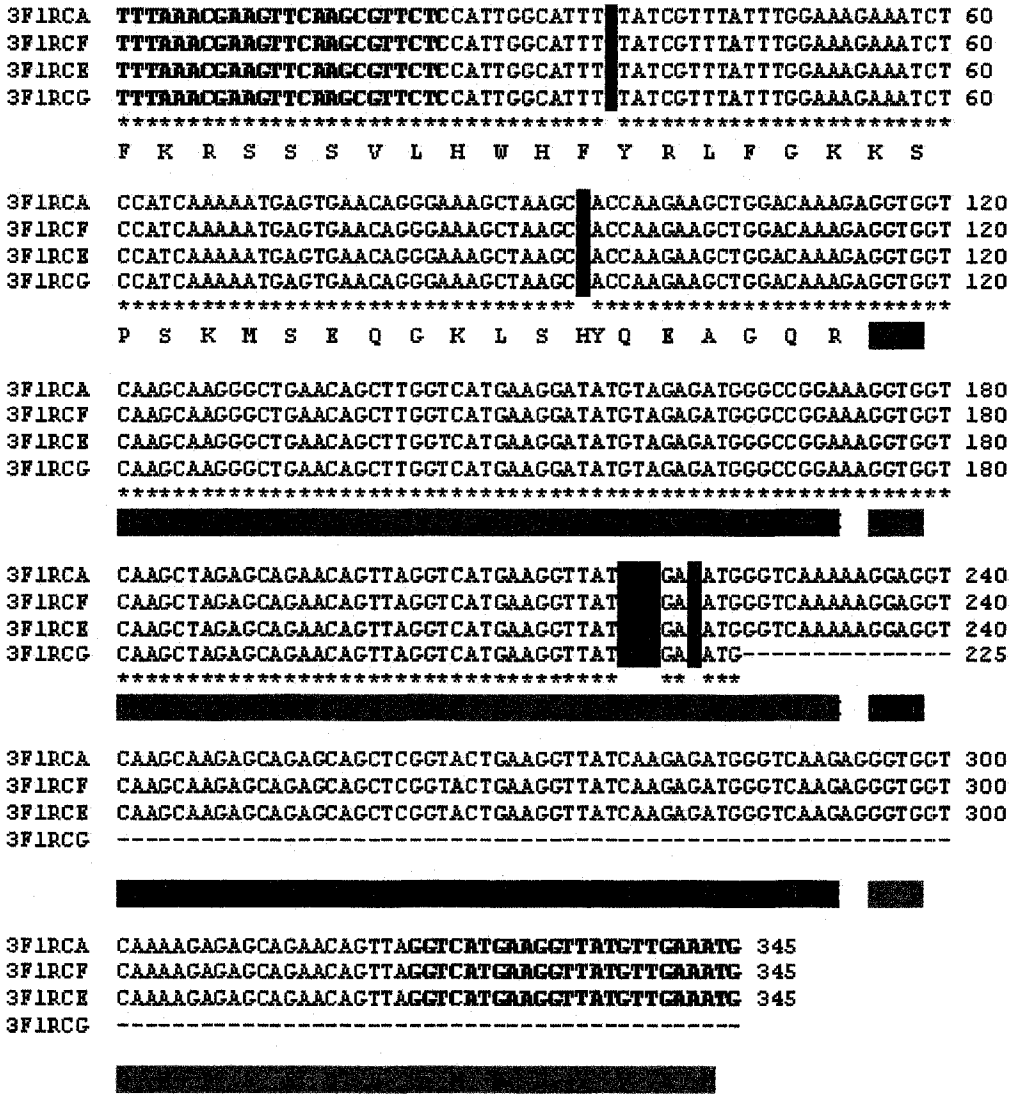


Table 6. DNA fragments amplified from *A. franciscana* genomic DNA by PCR using the primers LEA-3F and LEA-1R

Primer Combination	*Size of sequenced DNA fragment*	*Number of repeated sequence motifs*
LEA-3F / LEA-1R (Clone designation: 3F1RCA)	345 bp	4
LEA-3F / LEA-1R (Clone designation: 3F1RCE)	345 bp	4
LEA-3F / LEA-1R (Clone designation: 3F1RCF)	345 bp	4
LEA-3F / LEA-1R (Clone designation: 3F1RCG)	225 bp	2

4.4.7 COMPARISON OF NUCLEOTIDE SEQUENCES OF THE GROUP 1 LEA GENE OBTAINED FROM POLYMERASE CHAIN REACTIONS

Seventeen nucleotide sequences amplified from the embryonic cDNA library and genomic DNA by PCR were aligned using Clustal W (http://www.ebi.ac.uk/clustalw) in order to confirm the sequence of the Group 1 LEA gene and to establish a consensus sequence for the LEA gene. This also allowed for the identification and confirmation of the number of repeated sequence motifs in each of the selected clones and for the confirmation of sequence data outside of the tandem repeats. Sequencing results from PCR using the T7 promoter primer and the LEA-1R primer with the embryonic cDNA library as the template revealed one full-length and one partial repeat of the conserved Group 1 LEA sequence motif in addition to sequence within the 5'-untranslated region. Additional polymerase chain reactions using the LEA-3F primer near 5'-end of the sequence and the LEA-1R primer to the repeat region produced sequencing results that confirmed the 5'-untranslated region and a portion of the coding sequence with up to five repeats of the Group 1 LEA protein motif. Further amplifications using the upstream LEA-3F primer with the LEA-6Ra primer near the 3'-end of the sequence produced sequence data that once again confirmed the 5'-untranslated region and revealed seven tandem repeats of the conserved domain. DNA fragments amplified from genomic DNA using the primers LEA-3F and LEA-1R were also aligned throughout the comparison of the seventeen clones, as these sequences contained four repeat motifs and were devoid of introns. DNA fragments amplified using the primers LEA-3F and M13R produced the most complete gene sequence, revealing a portion of the 5'-untranslated region, the nucleotide sequence encoding the mature protein, and the 3'-untranslated region that included the poly-A tail. Since the LEA-3F:M13R fragment revealed the full gene sequence, the presence of eight repeats of the Group 1 LEA motif was confirmed. Since the other amplified DNA fragments had decreased length compared to the LEA-3F:M13R fragment and did not reveal the complete LEA gene sequence, these shorter fragments had fewer numbers of the repeated domain. The alignment of sequence data from these seventeen different clones is shown in Figure 14.

Figure 14. ClustalW alignment of seventeen DNA sequences of regions of the LEA gene amplified by PCR using the *Artemia franciscana* embryonic cDNA library and genomic DNA

Sequence data for seventeen separate clones were obtained from amplification of regions of the LEA gene by polymerase chain reactions using primers internal to the LEA gene sequence and primers to sequences within the Bluescript phagemid containing the embryonic cDNA library. Sequences amplified from genomic DNA are also aligned. These sequences have been individually presented in the description of PCR results. Primer sequences are shown in bold and primers to the Bluescript phagemid have been removed. Nucleotides highlighted in green indicate a discrepancy in the identity of the base pair in one of the aligned sequences. Nucleotides highlighted in yellow indicate polymorphic base pair regions, where the aligned sequences differ in the base pair at a specific location. Asterisks indicate nucleotide identity among the seventeen sequences.

Sequences are identified by their clone designations:

3F1R* = amplified from embryonic cDNA library using primers LEA-3F and LEA-1R
(3F1RF = clone F, 3F1RD = clone D, 3F1R2 = clone 2,
3F1R5 = clone 5, 3F1RC = clone C, 3F1R7 = clone 7)

3FM* = amplified from embryonic cDNA library using primers LEA-3F and M13R
(3FM5 = clone 5, 3FM2 = clone 2)

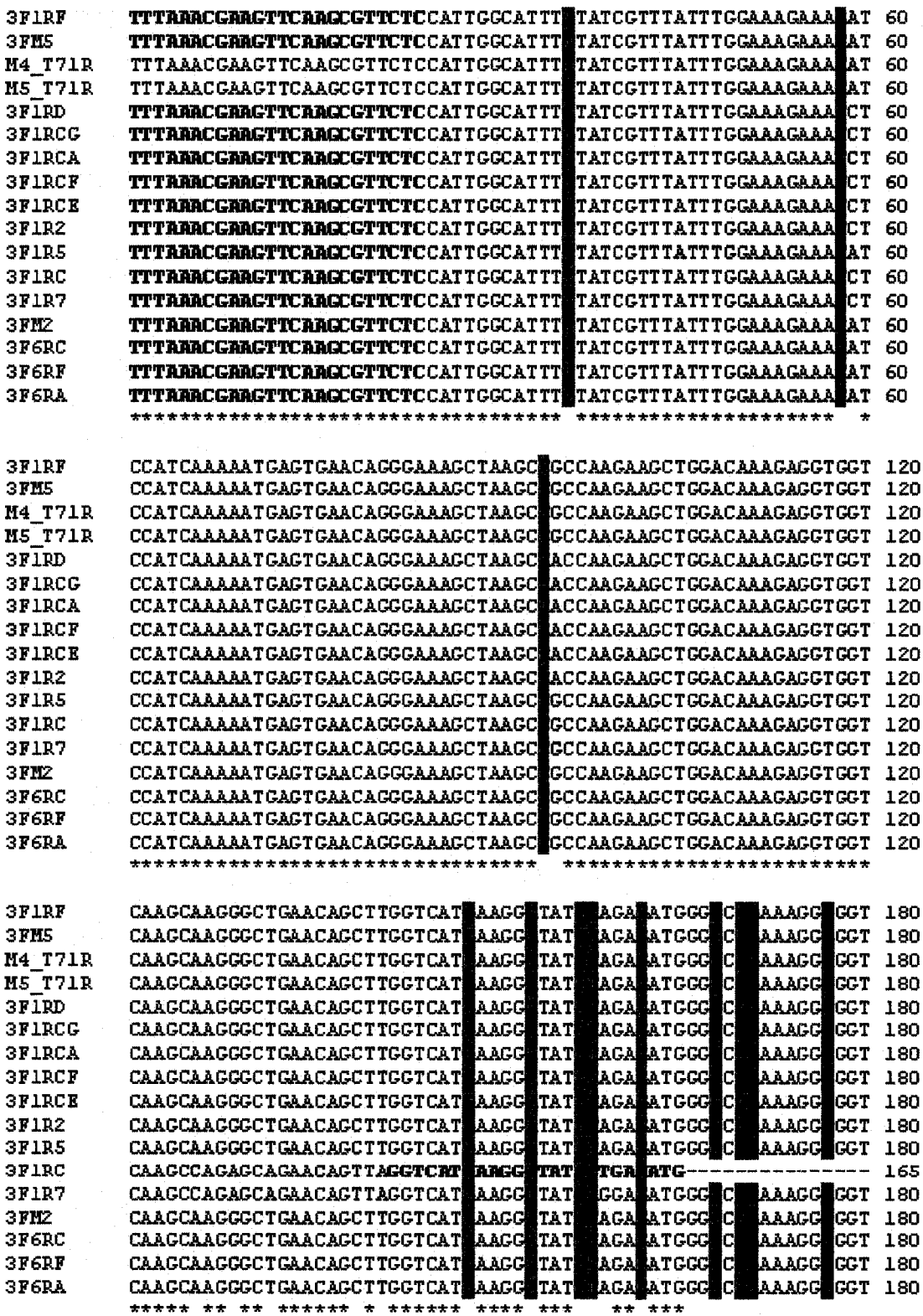
M4_T71R = amplified from embryonic cDNA library using the PCR primers
T7 promoter and LEA-1R, clone M4

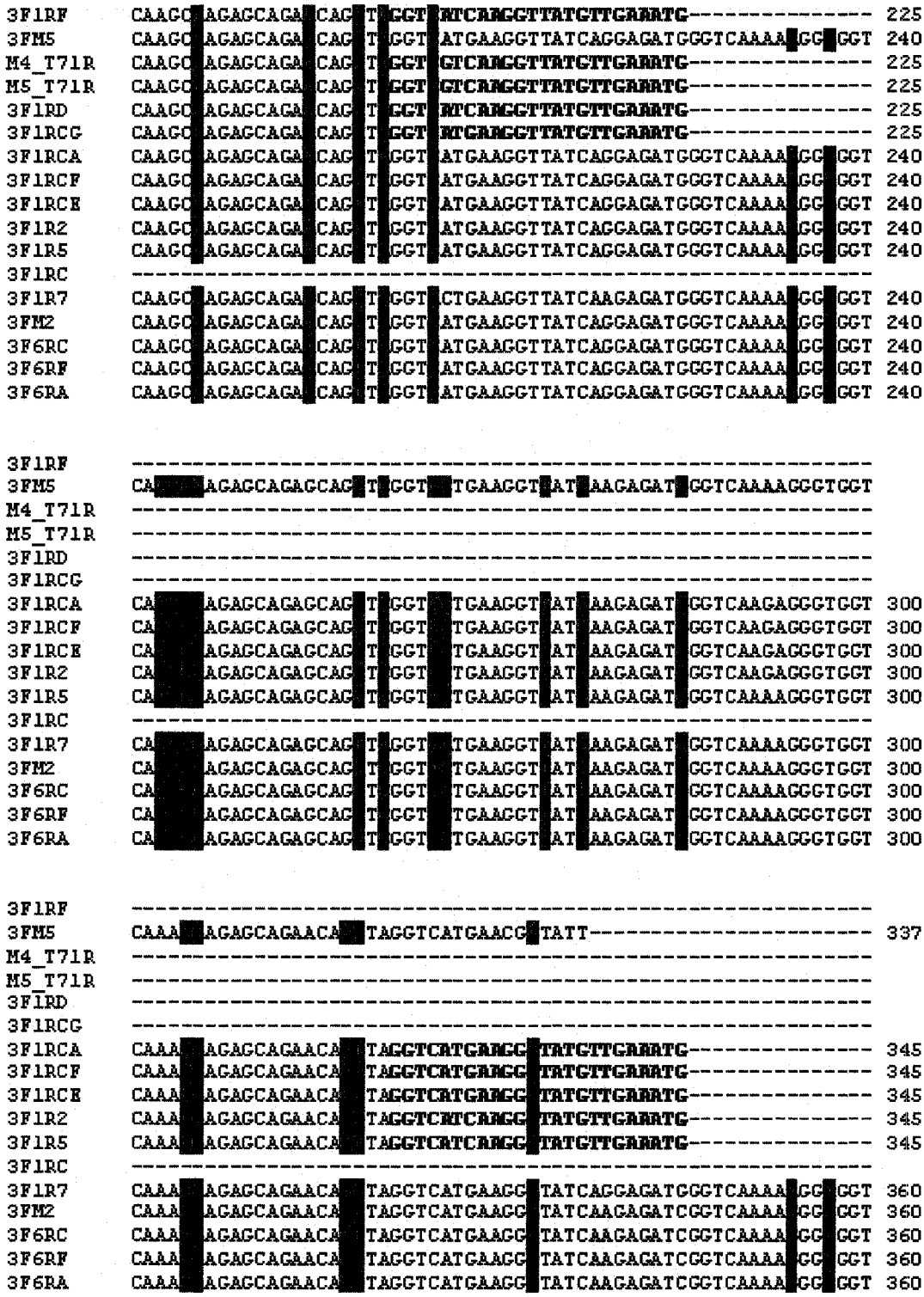
M5_T71R = amplified from embryonic cDNA library using the PCR primers
T7 promoter and LEA-1R, clone M5

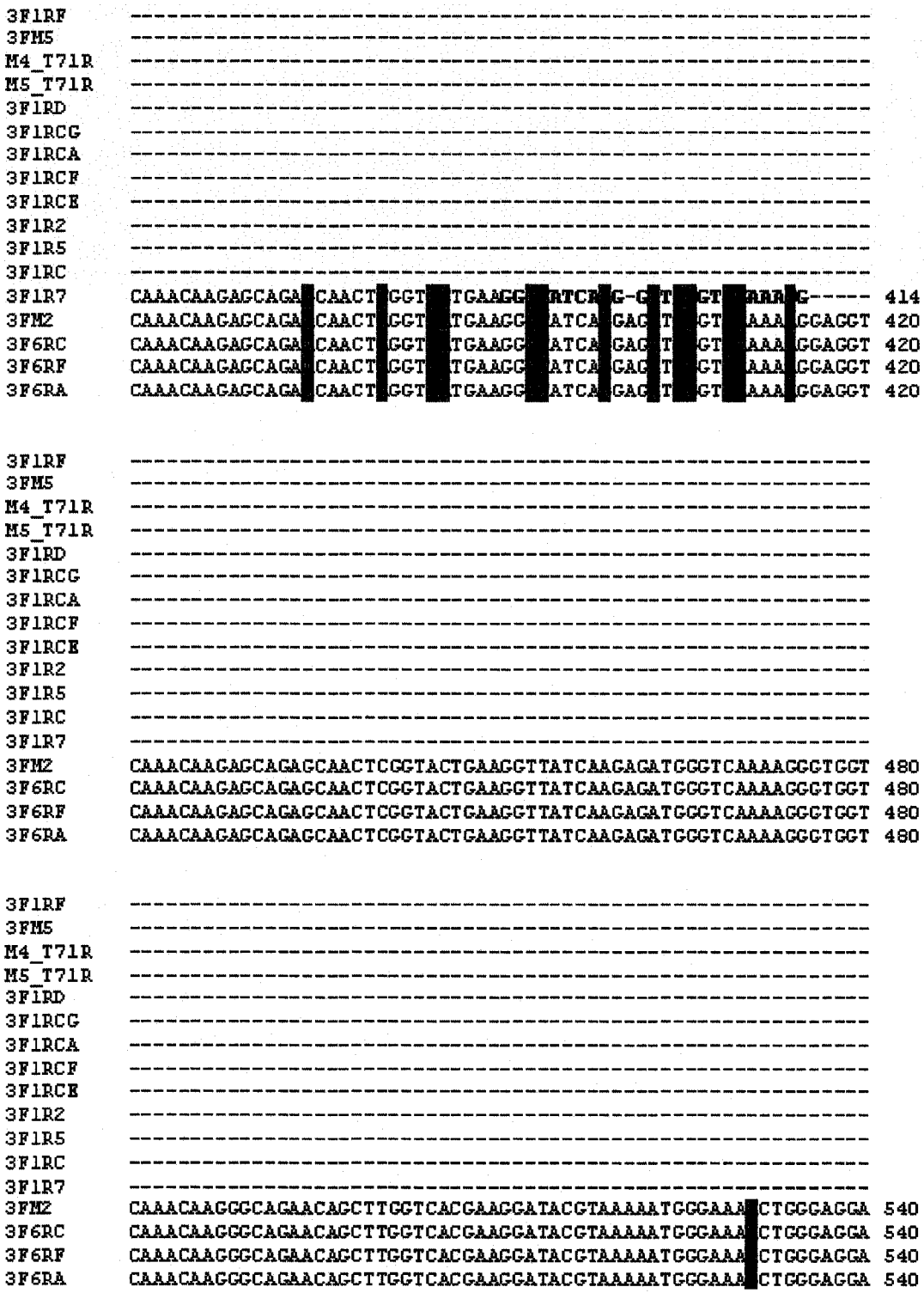
3F1RC* = amplified from genomic DNA using PCR primers LEA-3F and LEA-1R
(3F1RCG = clone G, 3F1RCA = clone A, 3F1RCF = clone F, 3F1RCE = clone E)

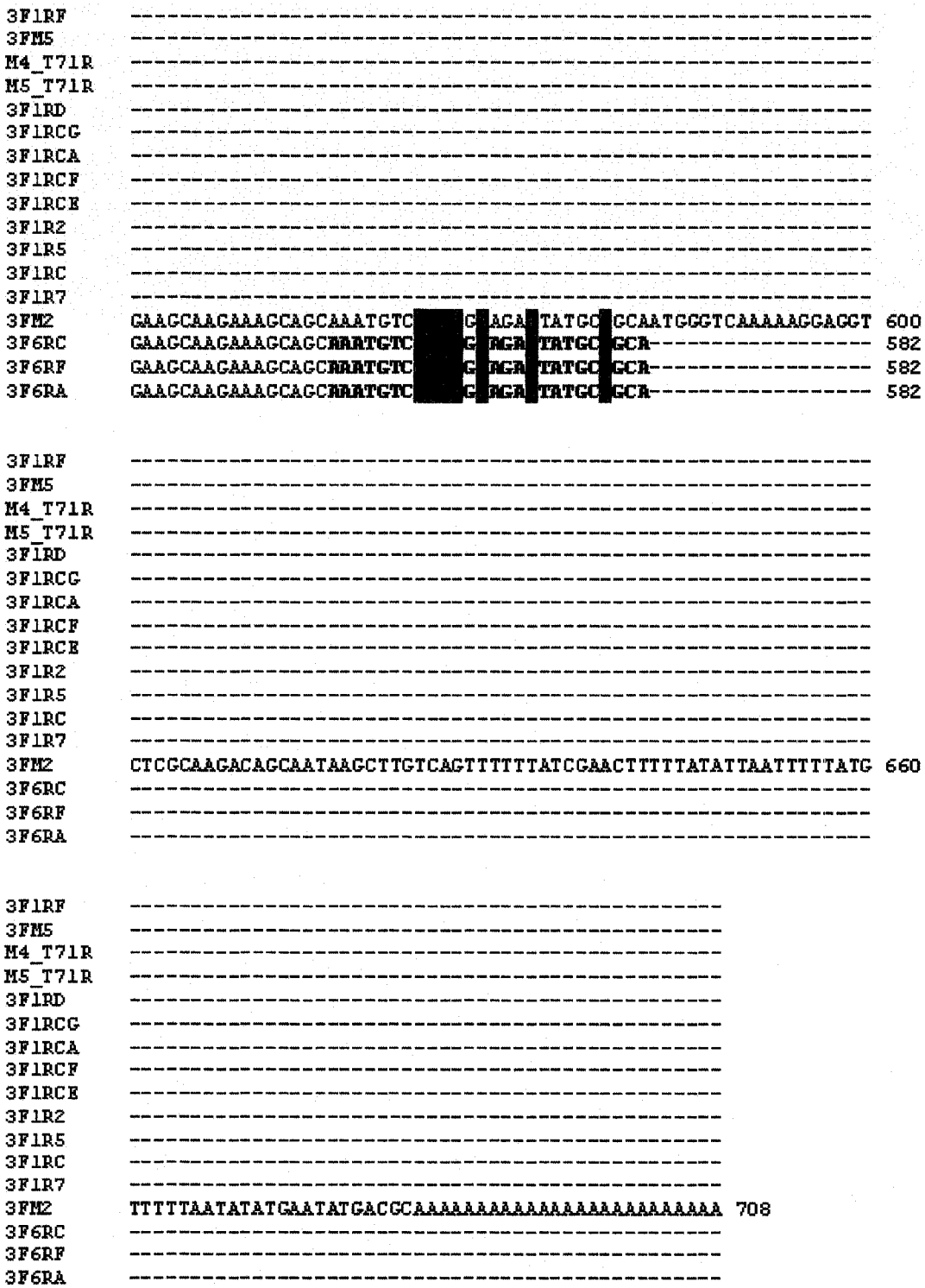
3F6R* = amplified from embryonic cDNA library using primers LEA-3F and LEA-6Ra
(3F6RC = clone C, 3F6RF = clone F, 3F6RA = clone A)

Figure 14.









4.5 N-TERMINAL SEQUENCING ANALYSES OF THE *ARTEMIA FRANCISCANA* 21 kDa LEA PROTEIN

N-terminal sequencing of a Mono-S purified sample of the 21 kDa LEA protein revealed the following amino acid sequence at the N-terminus of the protein, where brackets indicate ambiguity in the stated residue: NH_2 - (S) E Q Y K L S R Q E A G Q R (G) G Q A R A E (K) (L) (S) (G) (Y). Comparisons of the amino acid sequence deduced from PCR results and the sequence of 26 amino acids obtained from N-terminal sequencing reveal 81% identity between the two sequences (Figure 15). The amino acid sequence deduced from PCR results revealed an additional 23 amino acid stretch at the protein's N-terminus that did not appear in the N-terminal sequencing results. Since the N-terminal sequencing procedure revealed the amino acid sequence of the mature protein and was immediately preceded by a methionine residue, this procedure confirmed the initiation of the sequence encoding the LEA protein. It is possible that polymerase chain reactions amplified an upstream regulatory region of the gene, a pro-region that is transcribed but not translated into protein, within the 5'-untranslated region of the LEA gene. The N-terminal sequencing results also included a portion of the characteristic Group 1 LEA motif found to be repeated eight times in DNA fragments amplified by PCR, providing further confirmation of the sequence of this repeat region. Figure 15 depicts the alignment of sequence data deduced from PCR and the results of the N-terminal sequencing procedure.

Figure 15. Comparison of Group 1 LEA amino acid sequences deduced from PCR products and N-terminal sequencing procedures

The upper line, labelled "PCR seq.", shows the amino acids encoded by the consensus sequence that was formed from the sequencing of PCR products. The lower line, labelled "N-terminal seq.", shows the amino acid sequence produced from N-terminal sequencing of a pure preparation of the Group 1 LEA protein. Brackets in this sequence indicate ambiguity in the amino acid at a particular position. Asterisks indicate identity between the two sequences. This alignment allows for the comparison of amino acid sequences generated using two different methods, confirming the N-terminal sequence of the mature Group 1 LEA protein and identifying a portion of the 5'-untranlated region. The sequences are 81% identical.

```
PCR seq.          F K R S S S V L H W H F Y R L F G K K Y P S K M S E Q
N-terminal seq.                                                (S)E Q
                                                                * * *

PCR seq.          G K L S R Q E A G Q R G G Q A R A E Q L G H E …
N-terminal seq.   Y K L S R Q E A G Q R(G)G Q A R A E(K L S G Y)
                    * * * * * * * * * * * * * * * * *   *
```

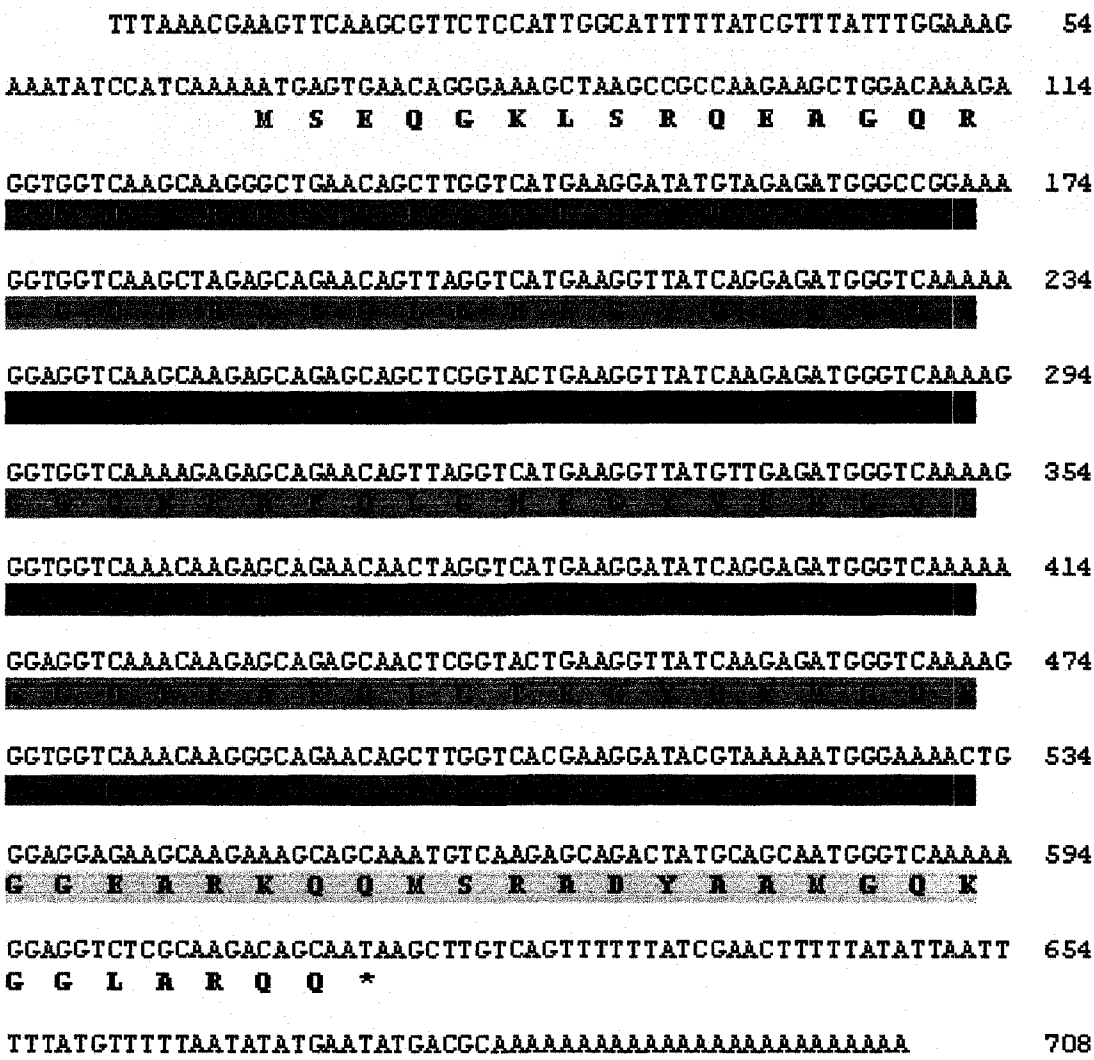
4.6 FORMATION OF A CONSENSUS SEQUENCE FOR THE GROUP 1 LEA GENE

The alignment of the seventeen different clones revealed individual locations where there were discrepancies in the base pair in one of the aligned sequences (nucleotides highlighted in green, Figure 14). The alignment also allowed for the identification of polymorphic base pair regions, where two or more different nucleotides were listed at a given location among the seventeen clones (nucleotides highlighted in yellow, Figure 14). In the formation of the consensus sequence for the Group 1 LEA gene, base pair discrepancies were addressed by placing the nucleotide that appeared in the greatest number of clones into the consensus sequence. Once the consensus nucleotide sequence of the Group 1 LEA gene was established, the sequence was translated into a consensus amino acid sequence, shown in Figure 16. The amino acids at the N-terminus of the protein were confirmed by N-terminal sequencing results (Figure 15). Within this consensus sequence, the eight repeats of the Group 1 LEA sequence motif were highlighted (Figure 16). The Group 1 LEA gene consensus sequence provided the best possible representation of the seventeen different clones, enabling comparisons to be made among the eight repeated motifs and other gene sequences, as described in section 4.7 (see Appendix F for the NCBI submission of the consensus sequence).

Figure 16. Formation of the consensus sequence of the LEA gene from the seventeen DNA sequences obtained from different clones

Through the alignment of the seventeen sets of sequence data from seventeen different clones obtained from amplification of regions of the LEA gene by PCR, a consensus sequence was created for the LEA gene. Where the seventeen clones showed differences in the nucleotides at a particular location, the base pair that appeared in the highest number of clones was selected for the consensus sequence. The encoded amino acids were then listed. At two locations, two nucleotides occurred with equal frequency. In these locations, the eight repeated sequence motifs were compared and the nucleotide that allowed for the insertion of the amino acid occurring with highest frequency was selected. The amino acids of the encoded protein are listed and an asterisk in the amino acid position denotes the stop codon. Highlighted amino acids indicate the eight repeated sequence motifs.

Figure 16.



4.7 COMPARISON OF EIGHT GROUP 1 LEA PROTEIN REPEATS

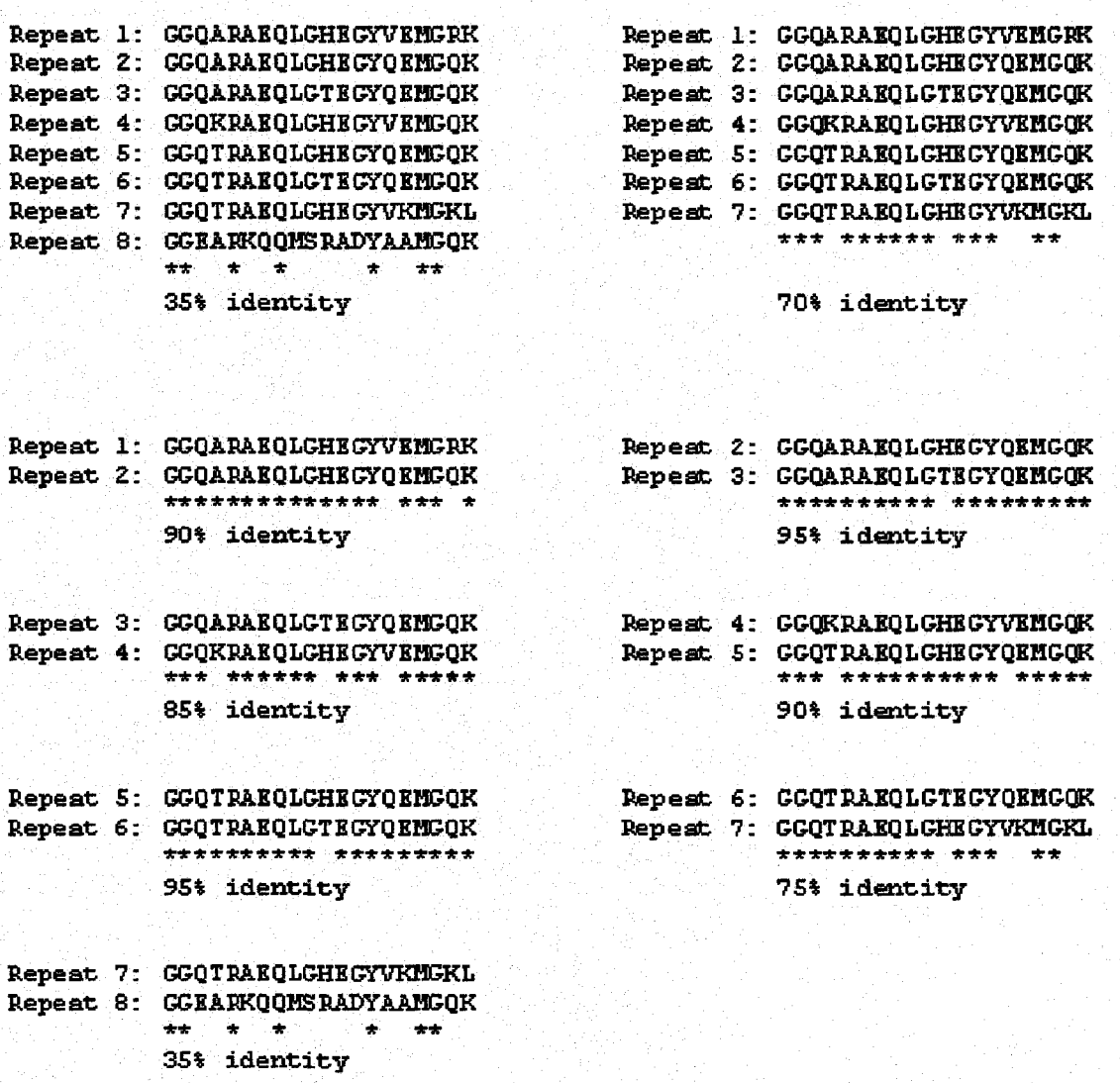
Comparisons of the eight repeats of the Group 1 LEA sequence motif composed of twenty amino acids per repeat revealed that the repeats are highly similar to one another. For example, the second and third repeat sequences, as well as the fifth and sixth repeat sequences are both 95% identical to one another. Similarly, comparisons of the first and second repeat sequences, as well as the fourth and fifth repeat sequences reveal that these repeat pairs are 90% identical. When all eight repeat sequences are compared to one another, the repeats are 35% identical. However, when repeat number eight, the repeat closest to the C-terminus of the protein is removed from the comparison, the first seven repeat sequences are 70% identical. Degeneracy of the repeat sequences near the C-terminus of the protein was also reported for LEA proteins belonging to Group 3 (Hand *et al.*, 2006). For some of the comparisons, such as the alignment of the first and second repeat sequences, as well and the comparison of the sixth and seventh repeats, the differences between the repeats occur near the C-terminus of the twenty amino acid sequence. Comparisons of the sequence repeats are shown in Figure 17.

High identity among the internal repeats of a Group 1 LEA protein is also observed in the GsiB protein of *Bacillus subtilis*, the only other Group 1 LEA protein currently identified in a non-plant species. In the GsiB protein, the five internal repeats of twenty amino acids per repeat are 85% identical to one another and the second and third repeats are 100% identical (Stacy & Aalen, 1998). There is also evidence of high repeat identity in Group 1 LEA proteins found in plant species, as the four repeats of the B19 protein in barley are 85% identical to one another, with the second and third repeats showing 100% identity to one another (Stacy & Aalen, 1998). This high identity among the Group 1 LEA sequence repeats contrasts the lower identity among the repeats of two proposed Group 3 LEA proteins in *Artemia franciscana* (Hand *et al.*, 2006). Based upon mRNA sequence data, the five published repeats in the AfrLEA1 sequence are only 19% identical to one another, with the most similar repeat pair having 66% identity. The three published repeats in the AfrLEA2 sequence are only 29% identical to one another. Thus, there appears to be a comparably high conservation of the twenty amino acid sequence repeated eight times within the Group 1 LEA protein of *Artemia franciscana*.

Figure 17. Comparison of the eight repeats of the 20-amino acid sequence motif from the Group 1 LEA protein consensus sequence

The sequences of the eight repeated motifs from the Group 1 LEA protein consensus sequence were aligned in order to compare the amino acid composition of each of the repeat regions. Asterisks indicate amino acid identity among the aligned sequences. The comparisons show the high identity among the aligned sequences and the degeneracy of the repeats near the C-terminus of the protein. Repeat 1 is nearest to the protein's N-terminus while repeat 8, showing greatest degeneracy, is nearest to the protein's C-terminus.

Figure 17.



4.8 COMPARISON OF THE GROUP 1 LEA GENE IN *ARTEMIA FRANCISCANA* TO LEA GENES IN OTHER ORGANISMS

Although all Group 1 LEA proteins have variations of the conserved sequence motif described by Cuming (1999) (GGQTRREQLGEEGYSQMGRK), it is interesting to investigate the similarity of entire Group 1 LEA gene sequences among different organisms. Since the only other Group 1 LEA protein identified in a non-plant species is the GsiB protein in the prokaryote *Bacillus subtilis*, this published gene sequence was aligned with the Group 1 LEA gene of *Artemia franciscana* (Figure 18A). The aligned sequences show identity in some localized regions, as shown in Figure 18A, with an overall 44% identity between the two sequences.

Recently, Hand *et al.* (2006) reported the existence of two mRNAs that they propose to encode two Group 3 LEA proteins in *Artemia franciscana.* Although the sequences published by Hand *et al.* (2006) belong to LEA protein Group 3 and the sequence identified herein is being classified as a Group 1 LEA protein, LEA protein classification is often based on factors such as expression pattern in addition to sequence similarity (Tunnacliffe & Wise, 2007). This variability in the classification criteria, in addition to the potential identification of multiple LEA proteins within the same species, prompted a comparison of the Group 1 LEA gene sequence with the two mRNA sequences AfrLEA1 and AfrLEA2 identified by Hand *et al.* (2006). These comparisons, shown in Figure 18B and C, reveal random amino acid similarities throughout the sequences, but no prominent domain identities.

When the consensus amino acid sequence of the Group 1 LEA protein in *Artemia franciscana* is entered into an NCBI BLAST search (http://www.ncbi.nlm.nih.gov/), the consensus sequence shows homology with an Em-like protein in the plant species *Arabidopsis thaliana.* The protein in *A. thaliana* is a homologue of the Em protein, a representative Group 1 late embryogenesis abundant protein in wheat. Figure 19A shows the alignment of the amino acid sequences of the Group 1 LEA protein in *A. franciscana* and the Em-like protein in *A. thaliana*, where the complete sequences are 55% identical. The NCBI BLAST search also revealed similarity with a Group 1 late embryogenesis abundant protein in the plant species, *Brassica napus.* The described protein was identified in the plant's leaf tissue and its inducibility upon the imposition of abiotic

stress was noted. The alignment of the amino acid sequences of these Group 1 LEA proteins in *A. franciscana* and *B. napus* is shown in Figure 19B where the sequences display 55% identity. The aligned sequences in Figure 19A and 19B show greatest amino acid identity in the central regions of the sequences, as opposed to near the N-terminal and C-terminal regions of the proteins. This can be explained by the observation that the characteristic Group 1 LEA repeat sequence is prominent in the central region of the sequence from *A. franciscana*. The first fifteen amino acids at the N-terminus of the protein are not included in the tandem repeats, and the seven amino acids at the C-terminus of the protein exhibit degeneracy of the repeat motif.

Figure 18. Amino acid sequence comparisons

Panel A. Clustal W was used to align the amino acid sequence of the Group 1 LEA consensus sequence from *Artemia franciscana* with the amino acid sequence of the GsiB Group 1 LEA protein from *Bacillus subtilis*. The complete sequences are 44% identical.

Panel B. Clustal W alignment of the consensus amino acid sequence from the Group 1 LEA protein in *Artemia franciscana* and the proposed amino acid of a Group 3 LEA protein, designated AfrLEA1, also in *Artemia franciscana* (Hand *et al.*, 2006). Random amino acid similarities were observed throughout the sequences.

Panel C. Clustal W alignment of the consensus amino acid sequence from the Group 1 LEA protein in *Artemia franciscana* and a second proposed amino acid of a Group 3 LEA protein, designated AfrLEA2, in *Artemia franciscana* (Hand *et al.*, 2006). Once again, only random similarities were observed throughout the two sequences.

Figure 18.

A.

```
A.fran_Grp1LEA   MSEQGKLSRQEAGQRGGQARAEQLGHEGYVEMGRKGGQARAEQLGHEGYQEMGQKGGQAR 60
B.subt_Grp1LEA   MADNNKMSREEAG------------------RKGGETTSKNHDKEFYQEIGQKGGEAT 40
                 *:::.*:**:***                  ****:: ::: .:* ***:*****:*

A.fran_Grp1LEA   AEQLGTEGYQEMGQKGGQKRAEQLGHEGYVEMGQKGGQTRAEQLGHEGYQEMGQKGGQTR 120
B.subt_Grp1LEA   SKNHDKEFYQEIGEKGGEATSKNHDKEFYQEIGEKGGEATSENHDKEFYQEIGRKGGEAT 100
                 ::: ..* ***:*:***:  ::: .:* * *:*:***:: :*: .:* ***:*:***::

A.fran_Grp1LEA   AEQLGTEGYQEMGQKGGQTRAEQLGHEGYVKMGKLGGEARKQQMSRADYAAMGQKGGLAR 180
B.subt_Grp1LEA   SKNHDKEFYQEIGSKGGNARNND---------------------------------- 123
                 ::: ..* ***:*.***::* ::

A.fran_Grp1LEA   QQ 182
B.subt_Grp1LEA   --
```

B.

```
Grp1_AfranLEA    ----------------------------------------MSEQGKLSRQEAGQRGG 17
Grp3_AfrLEA1     MAEPEEPPGIYEKVKSAFVSAPDRAQEAYNQAYESARSVFDDAVRSARKMKNTAAEQAQG 60
                                                        : .  *:..  * *  *

Grp1_AfranLEA    QAR-----------------AEQLGHEGYVEMG--------RKGGQARAEQLGHEGYQ 50
Grp3_AfrLEA1     AYEGLKESPENLQRVTRDIYHQAQDTGKGAYETVAGSADDAYRRAQETAQAAQEQSKGFL 120
                   .                  *:: *: .*  :.       *    *:* *   :*:

Grp1_AfranLEA    ---------------------EMGQKGGQARAEQLG----------------TEGYQE 71
Grp3_AfrLEA1     NRVKDTLTAPFSSSSDQAKETYDRTKDEAQYRAQQAADAGQFFGKVKDTITAPFTSGYDQ 180
                                 :  :. .* **:* .                  *.**::

Grp1_AfranLEA    MGQ------KGGQKRAEQLGHEG-------------YVEMGQKGGQTRAEQLGHEGYQEM 112
Grp3_AfrLEA1     TQEGYERARRSAEEAAQQAADQGQTLFERAKDTITSPFSSGSEQAQESFERAKRAAEEQV 240
                   :      :..:: *:* ..:*              .. *.: .*   *:  : . :::

Grp1_AfranLEA    GQ--------KGGQTRAEQLGTEGYQEMGQKGGQTRAEQLGHEG------------YVK 151
Grp3_AfrLEA1     EQSKGMFQNIKGTITSPFNSAADTAKEAGQRAKKQAEEAADQSQGFMQKVKDTFASPFLS 300
                  *        **  * . : .::  :* **:. :   *  .:.                ::.

Grp1_AfranLEA    MGKLGGEARKQQMSRADYAAMGQKGGLARQQ------------------------ 182
Grp3_AfrLEA1     AGEESQEAIERTKREAEEARHQGEGFLHRVADTIMHPFQSSSEQVGEAADRIKRGA 356
                  *: . ** ::   .*: *    :* * *
```

C.

```
Grp1_AfranLEA    ----------------------------------------------MSEQGKLSRQEA 12
Grp3_AfrLEA2     MPKAAAKGIGETVKADADVVEGMASTGYEKLKSAFGIASMKTKDAAENVAESARATKDYT 60
                                                             ::*..: ::: :

Grp1_AfranLEA    GQRGGQARAEQLG---------------HEGYVEMGKKGGQARAEQLG--HEGYQEMGQ 54
Grp3_AfrLEA2     VDSAKSAYDKTVDSTKSAYDKTTDSAKSVHDSTADTAKSAYNKATETLGSAYDKTKDTAQ 120
                  : . .*  : :.                *:. .: .:.. :  :* **  ::  :: .*

Grp1_AfranLEA    -KGGQARAEQLGTEGYQEMGQKGGQKRAEQLGHEGYVEMGQKGGQ--------TRAEQLG 105
Grp3_AfrLEA2     STYDQVTGAAHSAYDKTAEATKSAYDKTADAAHSVYNKTGDAGKQAYDSTKEAARSTGKS 180
                  . .*. .   .: .    . *.. .:: : .*. * : *: * *         :*:   .
```

Figure 19. Amino acid sequence comparisons

Panel A. Clustal W alignment of the consensus amino acid sequence of the Group 1 LEA protein from *Artemia franciscana* with the amino acid sequence of the Em-like protein from *Arabidopsis thaliana* (http://www.ncbi.nlm.nih.gov/). The complete sequences are 55% identical.

Panel B. Clustal W alignment of the consensus amino acid sequence from the Group 1 LEA protein in *Artemia franciscana* and the amino acid sequence of a Group 1 LEA protein from *Brassica napus* (http://www.ncbi.nlm.nih.gov/). The complete sequences displayed 55% identity.

Figure 19.

A.

```
A.fran_Grp1LEA   MSEQGKLSRQEAGQRGGQARAEQLGHEGYVEMGRKGGQARAEQLGHEGYQEMGQKGGQAR 60
A.thal_Em-like   MASK-QLSREELDEKAKQG-------ETVVPGGTGGHSLEAQEHLAEG----RSKGGQTR 48
                 *:.: :***:* .::. *.       *  *  *  * . .*::   **     .****:*

A.fran_Grp1LEA   AEQLGTEGYQEMGQKGGQKRAEQLGHEGYVEMGQKGGQTRAEQLGHEGYQEMGQKGGQTR 120
A.thal_Em-like   KEQLGHEGYQEIGHKGGEARKEQLGHEGYQEMGHKGGEARKEQLGHEGYQEMGHKGGEAR 108
                  **** *****:*:***: * ******** ***:***::* ***********:***::*

A.fran_Grp1LEA   AEQLGTEGYQEMGQKGGQTRAEQLGHEGYVKMGKLGGEARKQQMSRADYAAMGQKGGLAR 180
A.thal_Em-like   KEQLGHEGYKEMGRKGGLS----------TMEKSGGERAEEEGIEIDESKFTNK----- 152
                  **** ***:***:*** :           _* * ***  :::  . * : : :*

A.fran_Grp1LEA   QQ 182
A.thal_Em-like   --
```

B.

```
A.fran_Grp1LEA   MSEQGKLSRQEAGQRGGQARAEQLGHEGYVEMGRKGGQARAEQLGHEGYQEMGQKGGQAR 60
B.napu_Grp1LEA   MASK-QQSREELDEKARQG-------ETVVPGGTGGKSVEAQERLAEG----RSKGGQTR 48
                 *:.: : **:* .::. *.       *  *  *  * ...*::   **     .****:*

A.fran_Grp1LEA   AEQLGTEGYQEMGQKGGQKRAEQLGHEGYVEMGQKGGQTRAEQLGHEGYQEMGQKGGQTR 120
B.napu_Grp1LEA   KEQLGHEGYQEMGHKGGETRKEQLGHEGYQEMGHKGGETRKEQLGHEGYQEMGHKGGETR 108
                  **** *******:***:_* ******** ***:***:** ***********:***:**

A.fran_Grp1LEA   AEQLGTEGYQEMGQKGGQTRAEQLGHEGYVKMGKLGGEARKQQMSRADYAAMGQKGGLAR 180
B.napu_Grp1LEA   KEQLGHGGYKEMGRKGGLS----------TMDKSGGERAEEEGIEIDESKFTNK----- 152
                  ****  **:***:*** :           _*_* ***  :::  . * : : :*

A.fran_Grp1LEA   QQ 182
B.napu_Grp1LEA   --
```

4.9 ANALYSIS OF GROUP 1 LEA PROTEIN CHARACTERISTICS

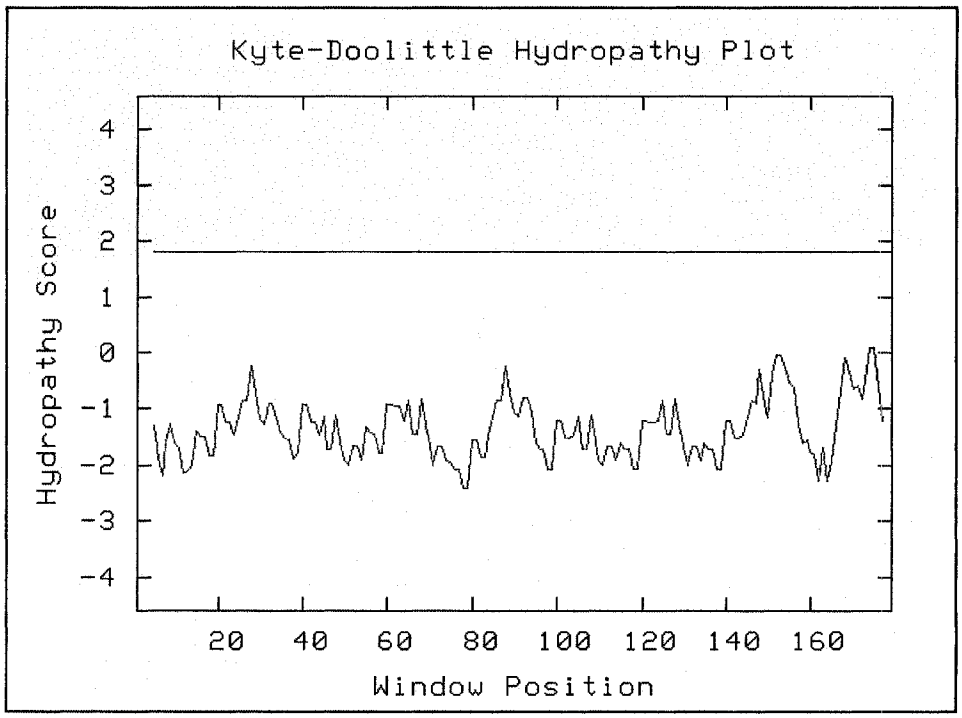
4.9.1 ASSESSMENT OF HYDROPHILICITY

When the sequence of a conventional globular protein is analyzed via a hydropathy plot, different portions of the protein sequence are distributed between the hydrophilic and hydrophobic regions of the graph, corresponding to the protein's internal hydrophobic patches and hydrophilic surface regions. However, hydropathy plots for LEA proteins reveal that a large portion of the protein or the entire protein sequence is distributed within the hydrophilic space of the graph, reflective of the high hydrophilicity of LEA proteins (Tunnacliffe & Wise, 2007). The 182 amino acid sequence of the Group 1 LEA protein of *Artemia franciscana* was analyzed via a Kyte-Doolittle hydropathy plot (Kyte & Doolittle, 1982), as shown in Figure 20. All of the peaks in the hydropathy plot are in the negative range, predicting high protein hydrophilicity. There is an absence of peaks in the graph that extend above a hydropathy score of 1.8, indicating that none of the sequence extends into the hydrophobic region of the graph. The graph even seems to reflect the periodicity of the amino acid sequence that is a result of the eight tandem sequence motifs. The hydropathy plot for the Group 1 LEA protein predicts a highly hydrophilic protein, one of the hallmarks of the LEA proteins.

The hydrophilicity of this LEA protein is related to other protein characteristics. For instance, the extreme hydrophilicity of LEA proteins is believed to be responsible for their lack of ordered secondary structure in the hydrated state. The hydrophilic, unstructured nature of LEA proteins is also thought to be responsible for the ability of LEA proteins to remain soluble in elevated temperatures. Aggregation of proteins due to heat is a result of partial protein denaturation and the association of exposed hydrophobic patches. Since LEA proteins remain natively unfolded and are highly hydrophilic, this method of protein aggregation due to heat is inapplicable to LEA proteins (Tunnacliffe & Wise, 2007).

Figure 20. Hydropathy plot of Group 1 LEA protein

The hydropathy plot for the Group 1 LEA protein with a window size of nine depicts the extreme hydrophilicity of the LEA proteins. The entire protein sequence remains below the threshold hydropathy score of 1.8, indicating high protein hydrophilicity (http://gcat.davidson.edu/rakarnik/kyte-doolittle.htm).



4.9.2 PREDICTED POST-TRANSLATIONAL MODIFICATIONS, SECONDARY STRUCTURE, MOLECULAR WEIGHT, AND SUBCELLULAR LOCALIZATION OF THE GROUP 1 LEA PROTEIN

Potential phosphorylation sites within the Group 1 LEA consensus amino acid sequence were examined using a phosphorylation site predictor program (Blom *et al.*, 1999). This analysis revealed two potential serine phosphorylation sites, eight potential tyrosine phosphorylation sites, and three potential threonine phosphorylation sites, although the threonine sites were predicted with minimal probability. The prediction of N-terminal acetylation sites (Kiemer *et al.*, 2005) showed the potential for amino-terminal acetylation of the Group 1 LEA protein. These predicted post-translational modifications have the potential to influence the function of this protein within the cell.

Similar to other characterized LEA proteins and specifically Group 1 LEA proteins (Wise & Tunnacliffe, 2004), it is likely that the Group 1 LEA protein in *Artemia franciscana* has little ordered secondary structure, forming a natively unfolded protein. However, as reported for a Group 1 LEA protein in soybean (Soulages, 2002), it is possible that dehydration induces conformational changes in the LEA protein of interest to a more ordered structure. Based upon a secondary structure prediction program (Garnier *et al.*, 1996), it can be hypothesized that the Group 1 LEA protein may form alpha-helical structures upon desiccation.

Analysis of the consensus amino acid sequence of the Group 1 LEA protein gave a predicted molecular weight of approximately 20 kDa (Bjellqvist *et al.*, 1993). However, this prediction is based solely on the amino acid sequence of the protein and does not take into account potential post-translational modifications that would increase the molecular weight of the protein. These modifications could account for the initial description of the LEA protein as a 21 kDa protein based upon SDS-PAGE analyses. Furthermore, elution of the Group 1 LEA protein from gel filtration columns revealed that the protein of interest is likely a 21 kDa subunit of a homodimer or heterodimer, discussed in sections 4.1.5 and 4.3.1.

The consensus amino acid sequence of the Group 1 LEA protein was also examined for potential subcellular localization signals. This protein is not predicted to be targeted to mitochondria, the endoplasmic reticulum, peroxisomes, or the nucleus

(Horton *et al.*, 2006, Claros & Vincens, 1996). Due to the currently elusive nature of this protein's mechanism of action, and the multitude of predictions as to the mechanisms of action of all the LEA proteins, it is difficult to make any further predictions concerning the subcellular localization of the protein of interest.

4.10 SCREENING THE EMBRYONIC cDNA LIBRARY FOR THE COMPLETE LEA GENE SEQUENCE

Screening the *A. franciscana* embryonic cDNA library using a radiolabelled LEA probe was unsuccessful. Throughout the procedure, bacterial growth was achieved when XL1-Blue MRF' cells transformed with the Bluescript phagemid containing the embryonic cDNA library were used to inoculate LB/agar plates containing 50μg/mL ampicillin. Transfer of these colonies to nitrocellulose disks, hybridization with the radiolabelled LEA probe, and exposure of an X-ray revealed both positive and negative colonies throughout the screening process, as expected. However, when phagemid DNA isolated from positive colonies was digested with the restriction enzymes *XhoI* and *EcoRI* for the excision of cDNA inserts, transferred to a Biodyne B membrane, and hybridized with the radiolabelled LEA probe, the excised cDNA did not produce a positive signal. Instead, the radiolabelled LEA probe hybridized with the Bluescript phagemid after restriction digestion and Southern blotting of phagemid DNA isolated from positive and negative colonies (data not shown). The LEA probe-Bluescript phagemid hybridization created great difficulty in the screening of the embryonic cDNA library in the Bluescript phagemid using the given probe.

The hybridization of the radiolabelled LEA probe with the Bluescript phagemid during Southern blotting was quite puzzling since the probe hybridized to only some DNA samples isolated from bacterial colonies during the three screening procedures. Since the XL1-Blue MRF' cells are grown in the presence of ampicillin, only those cells that have taken up the Bluescript phagemid containing the embryonic cDNA library would be capable of growth, due to the presence of an ampicillin resistance gene in the phagemid. Since all of the colonies capable of growth in the presence of ampicillin would have the Bluescript phagemid, all of the colonies would be expected to produce a positive signal if the probe was specific for the Bluescript phagemid sequence.

To further examine the specificity of the radiolabelled LEA probe, additional hybridization experiments were performed. It was found that the radiolabelled LEA probe displayed specificity for the LEA gene, as it hybridized with PCR products amplified from the embryonic cDNA library using the primer combinations LEA-3F/LEA-1R and LEA-3F/LEA-6Ra (data not shown). PCR products amplified from the embryonic cDNA library using the primer combinations LEA-3F/LEA-1R and LEA-3F/LEA-6Ra and ligated into the cloning vector pCR® 2.1 were digested with the restriction enzyme *EcoRI*. Since the multiple cloning site in this vector is flanked by *EcoRI* sites, this digestion excised the PCR products from the vector. The reactions were electrophoresed on a 1% agarose gel, Southern blotting was performed, and the membrane was probed with the radiolabelled LEA probe. Examination of the exposed X-ray revealed that the radiolabelled LEA probe hybridized to the excised PCR products, indicating specificity for the LEA gene, but also hybridized to the pCR® 2.1 vector, indicating specificity for the plasmid vector sequence (data not shown).

Successful screening of the embryonic cDNA library may have required a radiolabelled probe complementary to an alternate region of the LEA gene, reducing potential specificity for vector sequences. However, due to the large number of repeated motifs throughout the LEA gene, unique base pair regions were limited. Instead of continuing to screen the embryonic cDNA library to obtain the entire LEA gene sequence, additional polymerase chain reactions were performed using primers within the established LEA gene sequence (LEA-3F) in combination with a primer in the Bluescript phagemid that contains the embryonic cDNA library (M13R) (see section 4.4.5).

4.11 WESTERN BLOT: EXAMINING POTENTIAL ANTIBODY REACTIONS

The following sections discuss reactions between proteins at various stages of the purification protocol and antibodies specific for a Group 3 LEA protein in the nematode, *Aphelenchus avenae*.

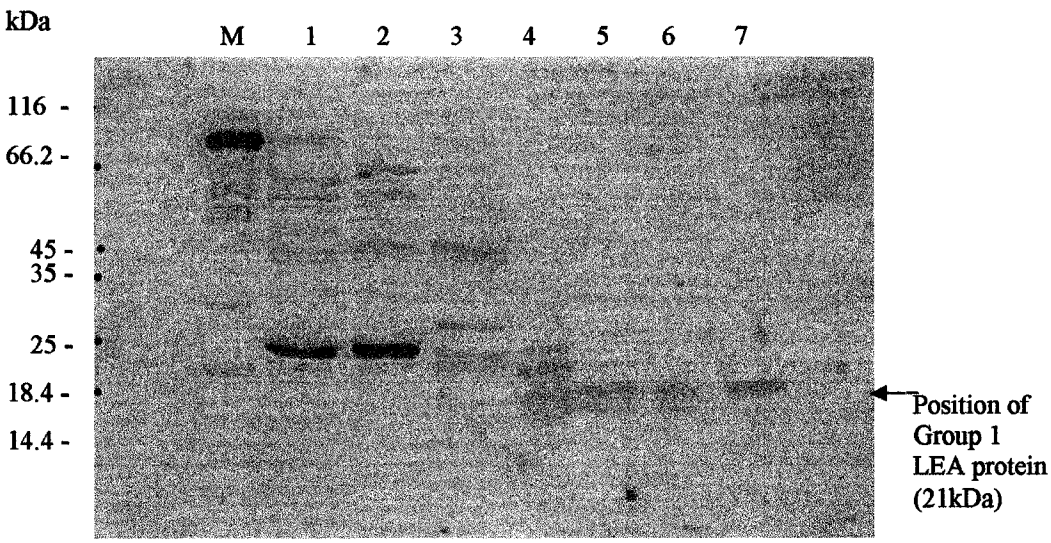
4.11.1 IDENTIFICATION OF A POTENTIAL GROUP 3 LEA PROTEIN IN ENCYSTED EMBRYOS OF *ARTEMIA FRANCISCANA* BY WESTERN BLOTTING

When samples from each stage of the purification protocol were probed with an antibody specific for a Group 3 LEA protein in nematode, a strong reaction occurred with a protein of approximately 25 kDa but no significant reaction occurred with the 21 kDa LEA protein (Figure 21). The reaction between the ~25 kDa LEA protein and the nematode antibody was present in the post-mitochondrial supernatant and the post-ribosomal supernatant preparations. Since the protein being investigated has a molecular mass of approximately 21 kDa upon SDS-PAGE and the two Group 3 LEA proteins predicted by Hand *et al.* (2006) are both expected to be 39 kDa, this reaction indicated the potential presence of another LEA protein in the encysted embryo of *A. franciscana*. However, contrary to the 21 kDa LEA protein, this ~25 kDa protein does not appear to be heat stable as the dark band on the X-ray disappeared in the heated (70°C/8 min.) preparation. These results suggest that additional LEA proteins are present in the encysted *A. franciscana* embryos, which is supported by the identification of numerous LEA proteins in a single organism (Tunnacliffe & Wise, 2007). Although the 21 kDa protein of interest in *A. franciscana* is thought to belong to the Group 1 LEA protein category, the results of this Western Blot suggest the presence of LEA proteins belonging to other protein groups, such as Group 3 LEA proteins. Since all LEA proteins identified in animal species belong to Group 3 (Stacy & Aalen, 1998, Tunnacliffe & Wise, 2007), with the exception of the LEA protein described herein, it seems plausible that the encysted embryos of *A. franciscana* express at least one LEA protein belonging to the Group 3 category in addition to its 21 kDa Group 1 LEA protein.

Figure 21. Western blot using antibodies specific for a Group 3 LEA protein in nematode and *A. franciscana* encysted embryo protein samples from stages of the purification protocol

This X-ray shows protein preparations from stages of the purification protocol probed with an antibody specific for a Group 3 LEA protein in the nematode *Aphelenchus avenae* (Lane M: protein molecular weight standards, Lane 1: post-mitochondrial supernatant, Lane 2: post-ribosomal supernatant, Lane 3: 70°C/8min. heat soluble proteins, Lane 4: Sephadex G-150 (SF) fraction, Lane 5: Mono Q fraction, Lane 6: 80°C/10min. heat soluble proteins, Lane 7: Mono S fraction). The antibody reaction in Lane M is insignificant, as this lane represents the protein molecular weight standards used during SDS-PAGE. An abundant reaction between the antibody and another potential LEA protein of approximately 25 kDa was apparent in the post-mitochondrial and post-ribosomal supernatant preparations (Lanes 1 and 2). After this 45 minute exposure time, weak interactions between the antibody and the 21 kDa LEA protein were visible in later stages of the purification protocol (Lanes 4 - 7), but are likely to be insignificant background exposure due to the prolonged exposure time with the ECL reagents and X-ray film.

Figure 21.



5. REVIEW AND SUMMARY

Encysted embryos of the brine shrimp, *Artemia franciscana*, define an early life cycle stage that withstands severe desiccation among other harsh environmental conditions (Warner *et al.*, 2004, Clegg & Trotman, 2002). The ability of this organism to survive dehydration potentiates a molecular mechanism of stress-tolerance, such as the synthesis of late embryogenesis abundant (LEA) proteins. Many parallels can be drawn between the life cycle stage during which LEA proteins are expressed in *Artemia franciscana* and the stages expressing LEA proteins in other animal species. Similar to adult brine shrimp, adults of the nematode *Caenorhabditis elegans* are sensitive to desiccation. However, when subjected to limited food reserves, *C. elegans* is capable of producing a dauer juvenile larval stage that survives the unfavourable environment. The expression of an LEA-like protein in this larval stage, entitled Ce-lea-1, provides the larvae with tolerance to desiccation (Gal *et al.*, 2004). This is reminiscent of the production of encysted brine shrimp embryos that enter diapause, surviving harsh conditions and, as shown in these experiments, synthesize an LEA protein. The anhydrobiotic nematode *Aphelenchus avenae*, anhydrobiotic larva of the chironomid *Polypedilum vanderplanki*, desiccation-tolerant bdelloid rotifers, and the nematode *Steinernema feltiae*, that is partially desiccation-tolerant, all express LEA proteins as a mechanism of surviving desiccation. While LEA proteins currently identified in animal species belong to Group 3 and mRNA believed to encode Group 3 LEA proteins have been identified in *Artemia franciscana* (Hand *et al.*, 2006), the LEA protein identified in these experiments is characterized as Group 1.

Results of experiments described herein provide immense support for the characterization of the identified protein as a Group 1 LEA protein. The identified protein is hydrophilic, heat-soluble, and contains eight repeats of a 20-amino acid sequence highly similar to the characteristic Group 1 LEA sequence motif described by Cuming (1999). The first six repeat sequences in the identified protein are 75% identical to the characteristic Group 1 LEA motif, and the seventh repeat is 70% identical to the published motif. Furthermore, each repeat is initiated by the characteristic double glycine thought to provide flexibility between the repeats of the Group 1 LEA proteins (Espelund *et al.*, 1992). The identification of the initial repeat sequence was made using a protein-

based approach, whereby a pure preparation of the protein was analyzed via trypsin digestion, MALDI mass spectrometry, and de novo sequencing (Dr. Panayiotis Vacratsis, Anna Kozarova), as well as by polymerase chain reactions that amplified the gene from a cDNA library. After obtaining a consensus amino acid sequence for the identified protein, searches of the NCBI databases revealed identity to other Group 1 LEA and LEA-like proteins.

Although the expression of the identified protein during various developmental stages of *Artemia franciscana* has not yet been investigated, it is possible that the Group 1 LEA protein is expressed exclusively within the encysted diapause embryos, since this is the desiccation-tolerant life cycle stage, and absent from swimming larvae and adults, as these stages are desiccation-sensitive. Given that the expression of the proteins artemin and p26 is exclusive to diapause in *Artemia franciscana* and important for stress-tolerance (Warner & Clegg, 2001, Warner *et al.*, 2004), a similar expression pattern could apply to the Group 1 LEA protein. The expression pattern of the Group 1 LEA protein at various life cycle stages of *Artemia franciscana* will be investigated in future experiments.

The identification of a Group 1 LEA protein in encysted embryos of *Artemia franciscana* is, to my knowledge, the first identification of a Group 1 LEA protein in an animal species. This identification, along with the recent characterization of two mRNA molecules believed to encode Group 3 LEA proteins in *Artemia franciscana* (Hand *et al.*, 2006), and the suggestion of another potential LEA protein of 25 kDa from Western blot procedures (see section 4.11.1) may increase the repertoire of identified hydrophilic proteins that provide desiccation-tolerance to the encysted embryos of this species. These findings, along with potential future discoveries of other LEA proteins in this species, increase our understanding of the tolerance mechanisms employed by this model organism, the wide-ranging expression of LEA proteins, and mechanisms of tolerating desiccation.

Appendix A

Description of commonly used reagents

LB solid medium (Luria-Bertani medium): 10.0 g/L bacto-tryptone, 5.0 g/L bacto-yeast, 10.0 g/L NaCl, 15.0 g/L agar, pH 7.5

LB liquid medium (Luria-Bertani medium): 10.0 g/L bacto-tryptone, 5.0 g/L bacto-yeast, 10.0 g/L NaCl, pH 7.5

TTBS (Tween 20/TBS): 100mM Tris-Cl, pH7.5, 0.9% NaCl, 0.1% Tween 20

TBS (Tris-buffered saline): 100mM Tris-Cl, pH 7.5, 0.9% NaCl

1x TAE (Tris-acetate/EDTA electrophoresis buffer): 0.04M Tris-acetate, 0.001M EDTA

TE buffer (Tris-EDTA), pH 7.4: 10mM Tris-HCl (pH 7.4), 1mM EDTA (pH 8.0)

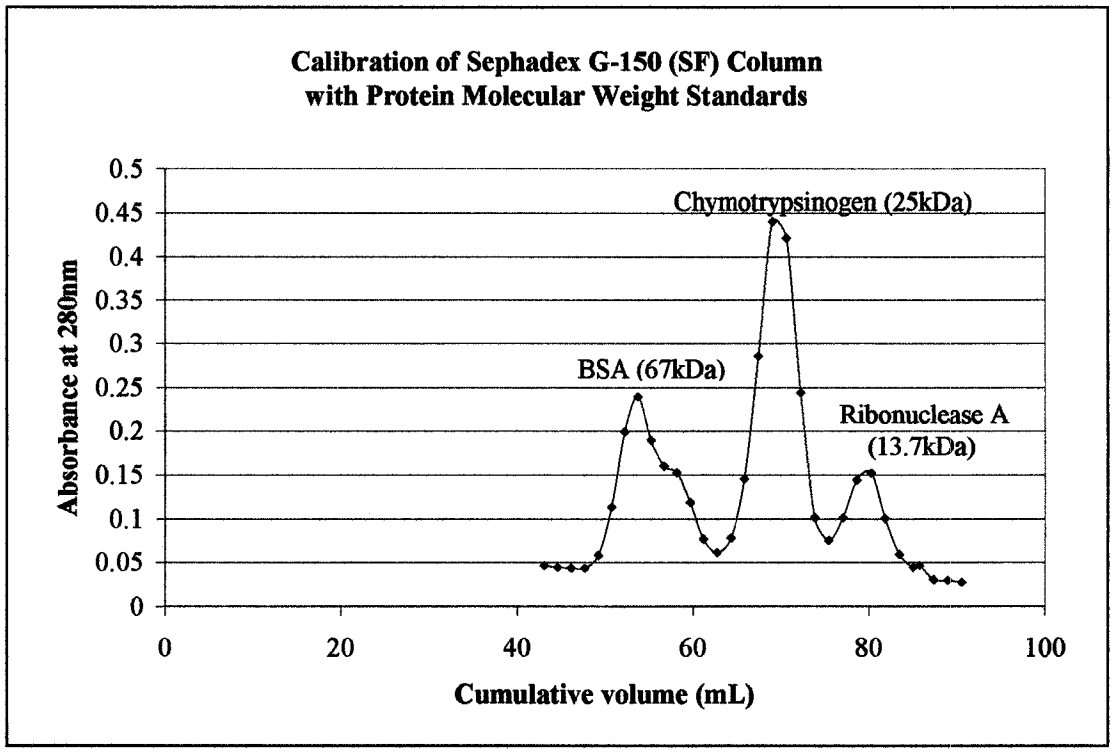
TB solution ($CaCl_2$ solution): 10mM Pipes, 55mM $MnCl_2$, 15mM $CaCl_2$, 250mM KCl

10x SSC: 1.5M NaCl, 0.15M sodium citrate, pH 7.0

Appendix B

Calibration of Sephadex G-150 (SF) Column

Protein molecular weight standards (bovine serum albumin (BSA, 67kDa), chymotrypsinogen (25kDa), ribonuclease A (13.7kDa)) eluted from the Sephadex G-150 (SF) column at the following elution positions, allowing for the formation of the standard curve shown in Figure 2.

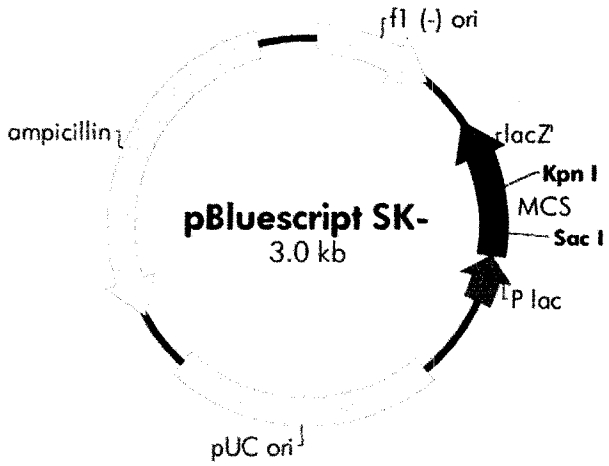


Appendix C

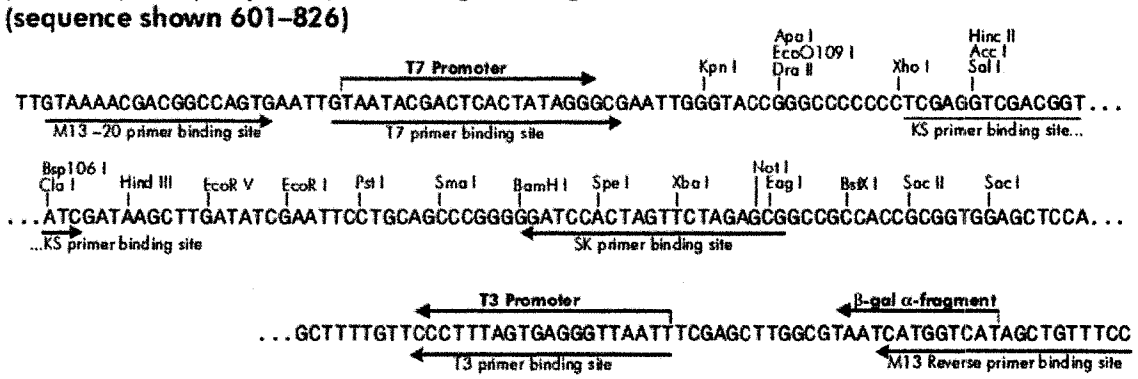
Map of the Bluescript phagemid (www.stratagene.com)

The embryonic cDNA library is contained within the Bluescript phagemid between *XhoI* and *EcoRI* restriction sites.

f1 (–) origin 24–330
β-galactosidase α-fragment 463–816
multiple cloning site 653–760
lac promoter 817–938
pUC origin 1158–1825
ampicillin resistance (*bla*) ORF 1976–2833



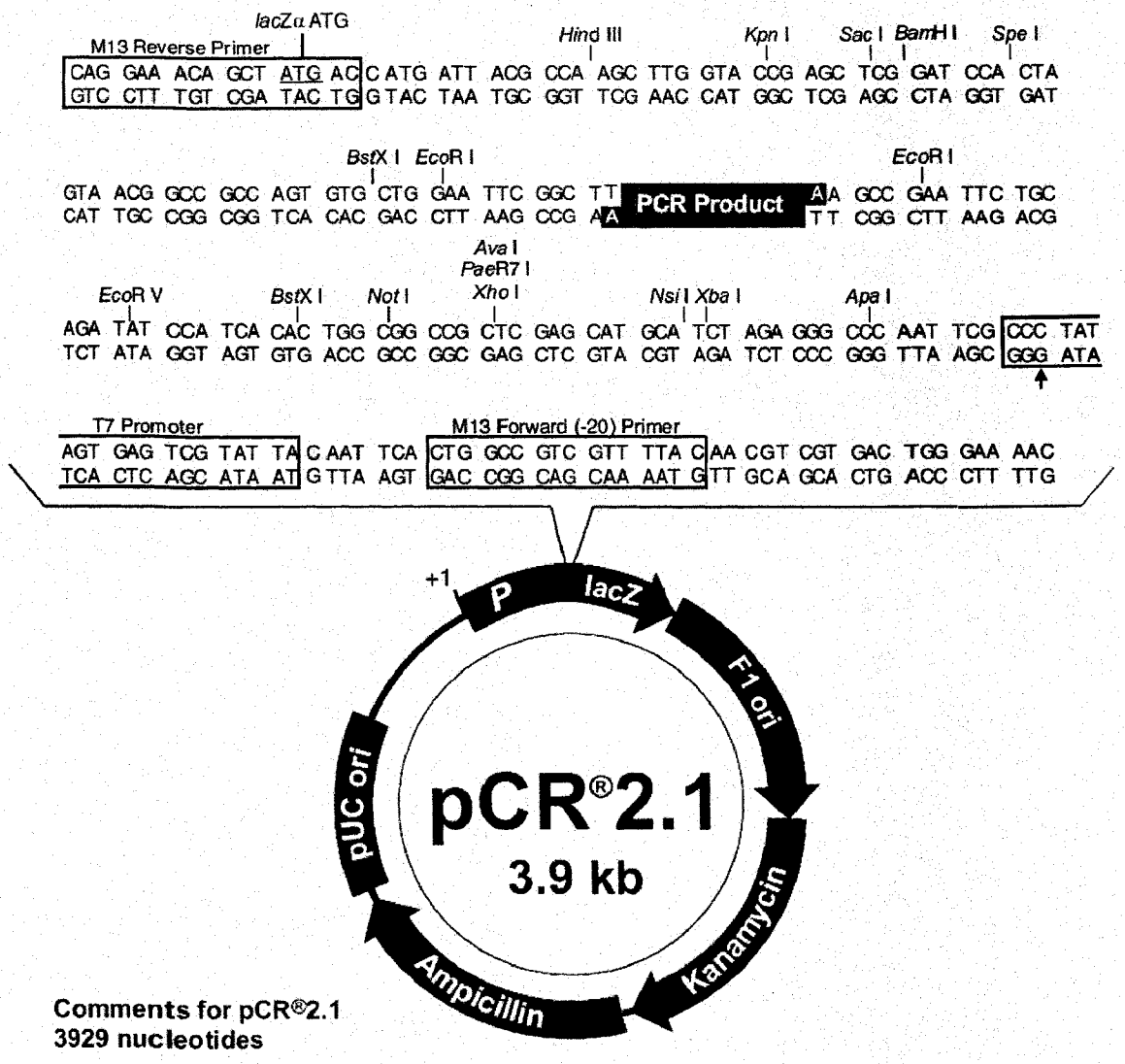
pBluescript SK (+/–) Multiple Cloning Site Region (sequence shown 601–826)



Appendix D

Map of the pCR® 2.1 vector (Invitrogen, www.invitrogen.com)

The plasmid vector pCR® 2.1 used for the ligation of purified PCR products



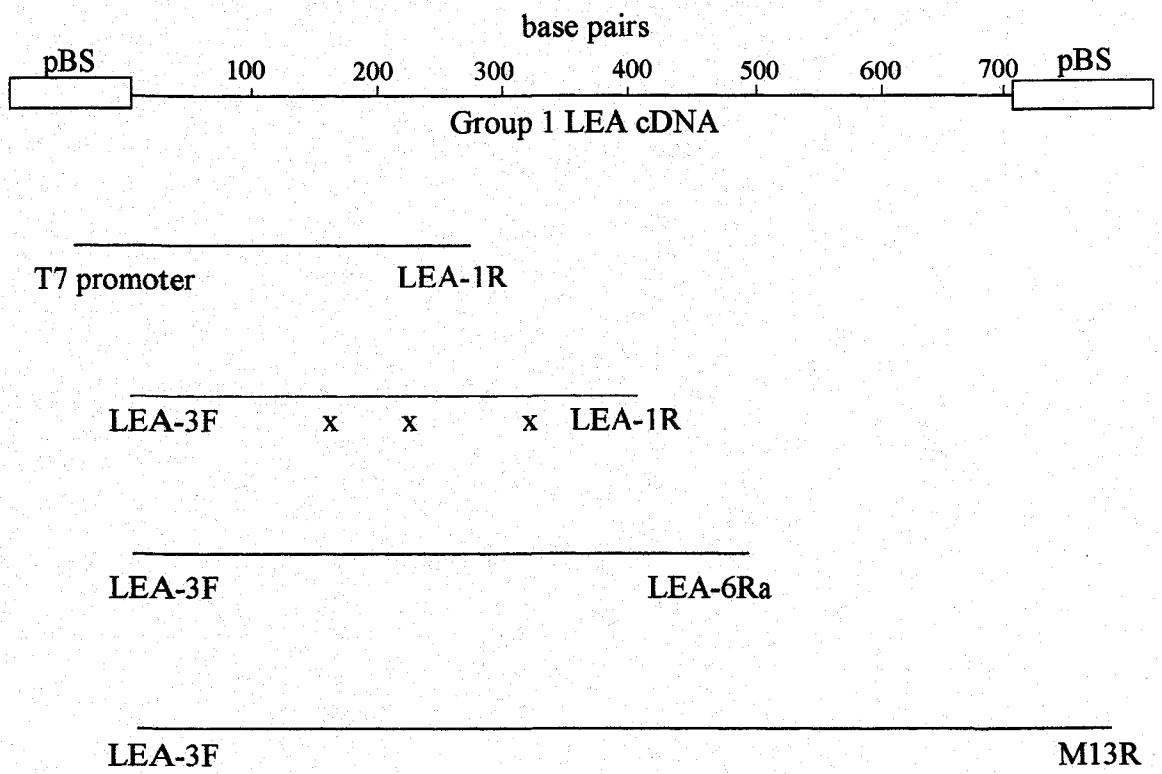
Comments for pCR®2.1
3929 nucleotides

LacZα gene: bases 1-545
M13 Reverse priming site: bases 205-221
T7 promoter: bases 362-381
M13 (-20) Forward priming site: bases 389-404
f1 origin: bases 546-983
Kanamycin resistance ORF: bases 1317-2111
Ampicillin resistance ORF: bases 2129-2989
pUC origin: bases 3134-3807

Appendix E

Summary of PCR Strategy

Depiction of the Group 1 LEA gene from the *Artemia franciscana* embryonic cDNA library in the phagemid Bluescript (pBS) and DNA fragments amplified from the cDNA library by PCR using each primer pair. Each horizontal line below the LEA gene represents the DNA fragment amplified by PCR using the listed primers. The symbol "x" along the line for the LEA-3F/LEA-1R primer combination represents the sequencing of differently sized DNA fragments amplified from this PCR due to binding of the LEA-1R primer to multiple repeats throughout the gene sequence (see Figure 11).



Appendix F

NCBI Submission: Group 1 LEA protein in *Artemia franciscana*

The following description of the Group 1 LEA gene was submitted to NCBI:

```
From gb-admin@ncbi.nlm.nih.gov.....................> bankit919706

Comment: The translated sequence contains eight 20-mer tandem repeats

--To be released immediately after processing--

LOCUS       bankit919706     708 bp    mRNA    linear   INV 07-JUN-2007
DEFINITION  Artemia franciscana late embryogenesis abundant protein
            (LEAP) mRNA, complete cds.
ACCESSION   919706
VERSION
KEYWORDS    .
SOURCE      Artemia franciscana
ORGANISM    Artemia franciscana
            Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda;
            Anostraca; Artemiidae; Artemia.
REFERENCE   1  (bases 1 to 708)
  AUTHORS   Sharon,M.A., Kozarova,A., Vacratsis,P.O. and Warner,A.H.
  TITLE     Characterization of a group 1 late embryogenesis abundant
            (LEA)protein in Artemia embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 708)
  AUTHORS   Sharon,M.A., Kozarova,A., Vacratsis,P.O. and Warner,A.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2007) Biological Sciences, University of
            Windsor, 402 Sunset Ave, Windsor, ON N9B3P4, Canada
COMMENT     Bankit Comment: The translated sequence contains eight 20-
            mer tandem repeats.
FEATURES             Location/Qualifiers
     source          1..708
                     /organism="Artemia franciscana"
                     /mol_type="mRNA"
                     /db_xref="taxon:6661"
     CDS             70..618
                     /note="hydrophilic protein"
                     /codon_start=1
                     /protein_id="PROT_1_bankit919706"

/translation="MSEQGKLSRQEAGQRGGQARAEQLGHEGYVEMGRKGGQARAEQL
GHEGYQEMGQKGGQARAEQLGTEGYQEMGQKGGQKRAEQLGHEGYVEMGQKGGQTRAE
QLGHEGYQEMGQKGGQTRAEQLGTEGYQEMGQKGGQTRAEQLGHEGYVKMGKLGGEAR
KQQMSRADYAAMGQKGGLARQQ"
```

BASE COUNT 272 a 94 c 198 g 144 t

ORIGIN

```
   1 tttaaacgaa gttcaagcgt tctccattgg cattttatc gtttatttgg aaagaaatat
  61 ccatcaaaaa tgagtgaaca gggaaagcta agccgccaag aagctggaca aagaggtggt
 121 caagcaaggg ctgaacagct tggtcatgaa ggatatgtag agatgggccg gaaaggtggt
 181 caagctagag cagaacagtt aggtcatgaa ggttatcagg agatgggtca aaaaggaggt
 241 caagcaagag cagagcagct cggtactgaa ggttatcaag agatgggtca aaagggtggt
 301 caaaagagag cagaacagtt aggtcatgaa ggttatgttg agatgggtca aaagggtggt
 361 caaacaagag cagaacaact aggtcatgaa ggatatcagg agatgggtca aaaaggaggt
 421 caaacaagag cagagcaact cggtactgaa ggttatcaag agatgggtca aaagggtggt
 481 caaacaaggg cagaacagct tggtcacgaa ggatacgtaa aaatgggaaa actgggagga
 541 gaagcaagaa agcagcaaat gtcaagagca gactatgcag caatgggtca aaaaggaggt
 601 ctcgcaagac agcaataagc ttgtcagttt tttatcgaac tttttatatt aatttttatg
 661 tttttaatat atgaatatga cgcaaaaaaa aaaaaaaaaa aaaaaaaa
```

BIBLIOGRAPHY

Abatzopoulos, T.J., Beardmore J.A, Clegg, J.S., and Sorgeloos P. *Artemia*: Basic and Applied Biology. Kluwer Academic Publishers, The Netherlands: 2002.

Altschul, S.F., L.M. Thomas, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Research 25, 3389-3402. (http://www.ncbi.nlm.nih.gov/)

Bjellqvist, B., G.J. Hughes, C. Pasquali, N. Paquet, F. Ravier, J.C. Sanchez, S. Frutiger, and D.F. Hochstrasser (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis 14, 1023-1031.

Blom, N., S. Gammeltoft, and S. Brunak (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology 294(5), 1351-1362.

Browne, J.A., K.M. Dolan, T. Tyson, K. Goyal, A. Tunnacliffe, and A.M. Burnell (2004) Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode *Aphelenchus avenae*. Eukaryotic Cell 3(4), 966-975.

Campos, F., F. Zamudio, and A.A. Covarrubias (2006) Two different late embryogenesis abundant proteins from *Arabidopsis thaliana* contain specific domains that inhibit *Escherichia coli* growth. Biochemical and Biophysical Research Communications 342, 406-413.

Chourey, K., S. Ramani, and A.K. Apte (2003) Accumulation of LEA proteins in salt (NaCl) stressed young seedlings of rice (*Oryza sativa* L.) cultivar Bura Rata and their degradation during recovery from salinity stress. Journal of Plant Physiology 160, 1165-1174.

Claros, M.G. and P. Vincens (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. European Journal of Biochemistry 241, 779-786.

Clegg J.S., Trotman, C.N.A. "Physiological and Biochemical Aspects of *Artemia* Ecology." *Artemia*: Basic and Applied Biology. Ed. TJ Abatzopoulos, JA Beardmore, JS Clegg, P Sorgeloos. The Netherlands: Kluwer Academic Publishers, 2002, 129-131.

Clegg, J. S., S.A. Jackson, and V.I. Popov (2000) Long-term anoxia in encysted embryos of the crustacean, *Artemia franciscana*: viability, ultrastructure, and stress proteins. Cell Tissue Research 301, 433-446.

Close, T.J. (1997) Dehydrins: a commonality in the response of plants to dehydration and low temperature. Physiologia Plantarum 100, 291-296.

Clustal W Multiple Sequence Alignment, European Bioinformatics Institute. (http://www.ebi.ac.uk/clustalw)

Cuming, A.C. "LEA proteins." Seed Proteins. Ed. PR Shewry, R Casey. The Netherlands: Kluwer Academic Publishers, 1999, 753-780.

Dure, L., M. Crouch, J. Harada, T.H.D. Ho, J. Mundy, R. Quatrano, T. Thomas, and K.R. Sung (1989) Common amino acid sequence domains among the LEA proteins of higher plants. Plant Molecular Biology 12, 475-486.

Espelund, M., S. Sæbøe-Larssen, D.W. Hughes, G.A. Galau, F. Larsen, and K.S. Jakobsen (1992) Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differently by abscisic acid and osmotic stress. The Plant Journal 2(2), 241-252.

Frank, W., D. Ratnadewi, and R. Reski. (2004) *Physcomitrella patens* is highly tolerant against drought, salt, and osmotic stress. Planta 220(3), 384-394.

Gal, T.Z., I. Glazer, and H. Koltai (2004) An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress. FEBS Letters 577, 21-26.

Garnier, J., J.F. Gibrat, and B. Robson. "GOR secondary structure prediction method version IV." Methods in Enzymology, Vol. 266. Ed. R.F. Doolittle. 1996. 540-553.

Goyal, K., C. Pinelli, S.L. Maslen, R.K. Rastogi, E. Stephens, and A. Tunnacliffe (2005) Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode. FEBS Letters 579, 4093-4098.

Hand, S.C., D. Jones., M.A. Menze, and T.L. Witt (2006) Life without water: expression of plant LEA genes by an anhydrobiotic arthropod. Journal of Experimental Zoology Part A 307(1), 62-66.

Hinniger, C., V. Caillet, F. Michoux, M.B. Amor, S. Tanksley, C.W. Lin, and J. McCarthy (2006) Isolation and characterization of cDNA encoding three dehydrins expressed during *Coffea canephora* (Robusta) grain development. Annals of Botany 97, 755-765.

Hong-Bo, S., L. Zong-Suo, and S. Ming An (2005) LEA proteins in higher plants: structure, function, gene expression and regulation. Colloids and Surfaces B: Biointerfaces 45, 131-135.

Horton, P., K.J. Park, T. Obayashi, and K. Nakai (2006) Protein subcellular localization prediction with WoLF PSORT. Proceedings of the 4th Annual Asia Pacific Bioinformatics Conference APBC06, 39-48.

Kamisugi, Y., and A.C. Cuming (2005) The evolution of the abscisic acid-response in land plants: comparative analysis of Group 1 LEA gene expression in moss and cereals. Plant Molecular Biology 59, 723-737.

Kiemer, L., J.D. Bendtsen, and N. Blom (2005) NetAcet: Prediction of N-terminal acetylation sites. Bioinformatics 21(7), 1269-1270.

Kikawada, T., Y. Nakahara, Y. Kanamori, K. Iwata, M. Watanabe, B. McGee, A. Tunnacliffe, and T. Okuda (2006) Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. Biochemical and Biophysical Research Communications 348, 56-61.

Kyte, J. and R. Doolittle (1982) A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157, 105-132.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-685.

Liang, P., and T.H. MacRae (1999) The synthesis of a small heat shock/α-crystallin protein in *Artemia* and its relationship to stress tolerance during development. Developmental Biology, 207, 445-456.

MacRae, T.H. (2003) Molecular chaperones, stress resistance and development in *Artemia franciscana*. Seminars in Cell & Developmental Biology 14, 251-258.

Mouillon, J.M., P. Gustafsson, and P. Harryson (2006) Structural investigation of disordered stress proteins: Comparison of full-length dehydrins with isolated peptides of their conserved segments. Plant Physiology 141, 638-650.

Nelson D.L., and Cox M.M. Lehninger Principles of Biochemistry, 3rd ed. Worth Publishers, New York: 2000.

Ohno, R., S. Takumi, and C. Nakamura (2003) Kinetics of transcript and protein accumulation of a low-molecular-weight wheat LEA D-11 dehydrin in response to low temperature. Journal of Plant Physiology 160, 193-200.

Sambrook, J., Fritsch, E.F., and Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbour Laboratory Press, New York: 1989.

Shih, M., S. Lin, J. Hsieh, C. Tsou, T. Chow, T. Lin, and Y.C. Hsing (2004) Gene cloning and characterization of a soybean (*Glycine max* L.) LEA protein, GmPM16. Plant Molecular Biology 56, 689-703.

Sorgeloos, P. "Life history of the brine shrimp *Artemia*." The Brine Shrimp *Artemia*, Vol. 3. Ed. G Persoone, P Sorgeloos, O Roels, E Jaspers. Belgium: Universa Press, 1980. xix-xxiii.

Soulages, J.L., K. Kim, C. Walters, and J.C. Cushman (2002) Temperature-induced extended helix/random coli transitions in a group 1 late embryogenesis-abundant protein from soybean. Plant Physiology 128, 822-832.

Stacy, R.A.P., and Aalen, R.B. (1998) Identification of sequence homology between the internal hydrophilic repeated motifs of Group 1 late-embryogenesis-abundant proteins in plants and hydrophilic repeats of the general stress protein GsiB of *Bacillus subtilis*. Planta, 206, 476-478.

Swire-Clark, G.A. and W.R. Marcotte, Jr. (1999) The wheat LEA protein Em functions as an osmoprotective molecule in *Saccharomyces cerevisiae*. Plant Molecular Biology 39, 117-128.

Taylor, R.M., and A.C. Cuming (1993) Selective proteolysis of the wheat Em polypeptide: Identification of an endopeptidase activity in germinating wheat embryos. FEBS Letters 331, 71-75.

Tanguay, J.A., R.C. Reyes, and J.S. Clegg (2004) Habitat diversity and adaptation to environmental stress in encysted embryos of the crustacean *Artemia*. Journal of Biosciences 29(4), 489-501.

Tu, Z., H. Guangyuan, K.X. Li, M.J. Chen, J. Chang, L. Chen, Q., Yao, and D.P. Liu (2005) An improved system for competent cell preparation and high efficiency plasmid transformation using different *Escherichia coli* strains. Electronic Journal of Biotechnology 8(1), 114-120.

Tunnacliffe, A., and M.J. Wise (2007) The continuing conundrum of the LEA proteins. Naturwissenschaften PMID: 17479232 [Epub ahead of print].

Wakui, K., and Y. Takahata (2002) Isolation and expression of *Lea* gene in desiccation-tolerant microspore-derived embryos in *Brassica* spp. Physiologia Plantarum 116, 223-230.

Warner, A.H., R.T. Brunet, T.H. MacRae, and J.S. Clegg (2004) Artemin is an RNA binding protein with high thermal stability and potential RNA chaperone activity. Archives of Biochemistry and Biophysics 424, 189-200.

Warner, A.H., and J.S. Clegg (2001) Diguanosine nucleotide metabolism and the survival of artemia embryos during years of continuous anoxia. European Journal of Biochemistry 268, 1568-1576.

Wise, M.J. (2003) LEAping to conclusions: A computational reanalysis of late embryogenesis abundant proteins and their possible roles. BMC Bioinformatics 4(52), 1471-2105.

Wise, M.J. and A. Tunnacliffe (2004) POPP the question: what *do* LEA proteins do? Trends in Plant Science 9 (1), 13-17.

Wolkers, W.F., S. McCready, W.F. Bandt, G.G. Lindsay, and F.A. Hoestra (2001) Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses *in vitro*. Biochimica et Biophysica Acta 1544, 196-206.

VITA AUCTORIS

NAME:	Michelle Ann Sharon
PLACE OF BIRTH:	Windsor, Ontario
YEAR OF BIRTH:	1982
EDUCATION:	St. Thomas of Villanova Secondary School, LaSalle 1996-2001
	University of Windsor, Windsor, Ontario 2001-2005, B.Sc.
	University of Windsor, Windsor, Ontario 2005-2007, M.Sc.