INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

DIFFERENTIAL GENE EXPRESSION IN RESPONSE TO FREEZING AND ANOXIA IN THE INTERTIDAL MARINE GASTROPOD, *LITTORINA LITTOREA*.

BY

TAMARA ERICA ENGLISH

B.SC. HONS. MOUNT ALLISON UNIVERSITY, 1993
M.SC. CARLETON UNIVERSITY, 1995

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COLLEGE OF NATURAL SCIENCES

CARLETON UNIVERSITY
OTTAWA ONTARIO, CANADA

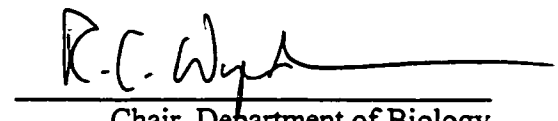
The undersigned hereby recommend acceptance of the thesis,

Differential gene expression in response to freezing and anoxia in the intertidal marine gastropod, *Littorina littorea.*

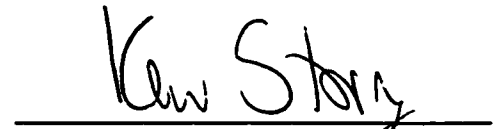
by

Tamara Erica English
B.Sc. hons. Mount Allison University, 1993
M.Sc. Carleton University, 1995

to the Faculty of Graduate Studies and Research,
in partial fulfillment of the requirements for the degree of Doctor of Philosophy



Chair, Department of Biology



Thesis Supervisor

External Examiner

Carleton University

2000

ABSTRACT

The intertidal zone is a highly variable environment where temperature, salinity and oxygen availability all fluctuate on a daily and seasonal basis. As a result, animals that inhabit this zone possess a high degree of metabolic plasticity. *Littorina littorea,* an intertidal marine gastropod, is tolerant of both freezing and anoxia. This study searched for changes in gene expression that may underlie the animal's ability to endure these stresses. Screening of cDNA libraries was employed to identify freezing- and anoxia-induced genes in foot muscle of *L. littorea*; two libraries were synthesized from mRNA isolated from foot muscle of animals exposed to either 1, 12 and 24 hours of freezing or 1, 12 and 24 hours of anoxia. Differential screening of the frozen or anoxic libraries with mRNA isolated from stressed versus control (5°C) snails, followed by northern blot analysis, resulted in six transcripts that were confirmed as stress-induced. DNA sequence analysis identified 3 clones as myosin heavy chain, beta actin and cytochrome oxidase subunit 2. These clones were isolated from the freezing library but showed greatest transcript accumulation during recovery after stress. The translated amino acid sequence of a fourth clone, LL_{MET}, elicited a putative identification of metallothionein, a heavy metal binding protein with a possible antioxidant role. The remaining two clones, LL_{GRP} and $LLAF_{W}$, were novel, showing little or no homology to DNA or amino acid sequences in various databases. Analysis of their translated amino acid sequences indicated that both proteins possessed a secretory signal at the N terminus, suggesting that they had either a membrane location or are excreted from the cell. Structural predictions based on previously analyzed proteins, suggested that LL_{GRP} is a membrane channel protein of the porin class whereas $LLAF_{W}$ fits the criteria of an anti-parallel bundle (apb) protein of the alpha-class. None of the six genes/proteins found by this study has a cellular function that integrates well with any of the previously-identified biochemical adaptations that support anoxia or freezing tolerance but each suggests that there is still much more to

learn about the types of molecular adjustments that are needed to support natural stress tolerance.

ACKNOWLEDGEMENTS

The following deserve thanks for their contributions to this work:

•Those who blazed trail: QY (Qin Yin Cai), Tolga Bilgen, Bill Willmore. For showing me the ropes and leaving me the frayed ends to clean up.

•Ken and Janet Storey, for advice and support FAR above and beyond the call of science, infinite patience, making me "do the right thing", and for caring.

•My parents, for laying the foundation. My sisters, in particular Trishy, my saner, younger twin, for long evenings of "talking me down" .

•Friends, for unconditional love, support, humour, brutal honesty and *shelter* (!).

•G and B for their confidence and encouragement.

Real knowledge is to know the extent of one's ignorance.
Confucius

Imagination is more important than knowledge.
Albert Einstein

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	amino acid
AFP	anti-freeze proteins
BLAST	Basic alignment search tool. Variations: blastp – input is an amino acid sequence blastn – input DNA blastx – input is a DNA sequence, which is then translated in all 3 ORFs, each of which is aligned with protein sequences from the database. Most useful choice with novel sequences.
cDNA	complementary DNA
cpm	counts per minute
DEPC	diethylpyrocarbonate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediamine tetraacetate
EtBr	eithidium bromide
GTE	glucose, tris, EDTA buffer
INP	Ice nucleating protein
Kb	kilobase
KDa	kilo Dalton
LLAF	*Littorina littorea* anoxic foot
LLFF	*Littorina littorea* frozen foot
MOPS	3-[N-morpholine] propanesulfonic acid
mRNA	messenger RNA
MT	metallothionein

CHAPTER 1

GENERAL INTRODUCTION

Stress can be defined as any change in the environment that causes a change in physiology or biochemistry. Environmental stressors such as temperature, oxygen tension, water and food availability are a fact of life for all organisms and can fluctuate on daily or seasonal time scales. For example, mammals and terrestrial ectotherms must deal with the seasonal cold of winter; diving animals and intertidal invertebrates with intermittent oxygen deprivation; and desert animals with heat and desiccation. Each of these scenarios has been shown to induce metabolic responses as illustrated in several reviews (Cheng and DeVries, 1991; Davies and Hew, 1990; Duman *et al.*, 1991; Storey and Storey, 1988, 1990, 1992). As the behavioral, physiological, and metabolic adaptations that help these organisms to survive have been elucidated, research has shifted to the genetic responses to stress.

Mechanisms of stress tolerance

Cellular metabolism is an intricate dance, coordinating biochemical pathways that integrate thousands of different enzymes, functional and structural proteins, as well as their substrates and cofactors and their interactions with numerous physical parameters (e.g. pH, temperature, ionic strength). Environmental factors such as low oxygen tension, low temperature and low water availability are all known to perturb metabolic homeostasis. Studies have shown that the adaptation of cellular metabolism to environmental stress can involve a variety of mechanisms including changes in metabolic

rate, amounts or activities of enzymes and functional proteins, intracellular pH, metabolite and ion concentrations, and membrane composition. For cells and organisms reacting to short term environmental stress, some fine-tuning of metabolism is often all that is required to maintain homeostasis. This can include regulatory mechanisms such as altering the availability of enzyme substrates and co-factors, the actions of allosteric effectors, changes in physical parameters, and covalent modification of proteins (e.g. reversible protein phosphorylation).

In the case of extreme or long-term stress, additional "coarse" controls are generally required. A common strategy is metabolic rate depression, an escape into a hypometabolic or dormant state, whereby organisms lower metabolic rate, often to <10 % of normal and sometimes to a virtually ametabolic state, to stretch their metabolic fuel reserves to last weeks, months or even years until environmental conditions favorable for renewed growth and development return (Hochachka and Guppy, 1987; Storey and Storey, 1990). One component of this strategy is to alter the amounts of selected proteins through regulation of the rates of protein synthesis or degradation. Since gene expression plays a vital role in all aspects of cellular activities including the responses to internal signals and external environmental factors and since the level and types of cellular proteins are partially controlled by the regulation of gene expression, it is clear that gene regulation must be a part of natural stress survival. Analysis of changes in gene expression using molecular biology techniques can reveal the range of systems employed by stress-tolerant animals to improve the chances of survival under extreme environmental conditions. In addition, these techniques may also help to identify the full range responses that may not be detectable with biochemical and physiological methods.

Studies of the contribution of gene expression in animal responses to environmental stress are the focus of my research project.

Regulation of gene expression

Gene expression involves several major steps and can be regulated at several points: processing, stability, transcription and translation. In eukaryotic cells mRNA stability is one of the key regulatory mechanisms involved (Beelman and Parker, 1995). mRNA decay is triggered by at least three types of initiating events: poly(A) shortening, premature translational termination, and endonucleolytic cleavage (Jacobson and Peltz, 1996). Translational control also modulates gene expression, primarily via control over the efficiency of mRNA translation, which is regulated at the translation initiation step via reversible phosphorylation of various initiator factors (Hershey, 1991; Mathews *et al.*, 1996; Pain, 1996).

In general, the strongest controls are at the level of transcription. In addition to the basic transcriptional machinery, transcription is modulated by various regulatory proteins, known as enhancers and silencers, that bind to DNA sequences. As their names indicate, enhancers stimulate transcription, whereas silencers repress it. These DNA regulatory elements are highly varied and gene specific, and can be located near to or at a great distance upstream or downstream from the core promoter. The interaction of a specific binding protein (activator or repressor) with its DNA regulatory element influences assembly of the initiation complex at the core promoter, which modulates the activity of the RNA polymerase II and hence the resulting rate of mRNA synthesis (Latchman, 1990; Mitchell and Tjian, 1989). Recent studies have further complicated the situation by

identifying regulatory proteins called co-activators and co-repressors that can function as signaling intermediates between the DNA binding proteins and the basal transcription machinery (Bjorklund and Kim, 1996; Verrjzer and Tjian, 1996).

There are a wide variety of response elements and their associated binding proteins (activators or repressors) which up- or down-regulate specific genes in response to chemical and physiological stimuli. A variety of extracellular stimuli, including growth factors, heat shock, osmotic stress, and ATP depletion can activate signal transduction pathways (Cohen, 1997; Kyriakis *et al.*, 1994; Kyriakis and Avruch, 1996). Signal transduction via protein kinase and phosphatase cascades is initiated by the binding of a ligand molecule (such as a growth factor or hormone) to a specific receptor on the plasma membrane or inside of a cell, which in turn alters the activities of certain protein kinases or phosphoprotein phosphatases and the levels of second messengers, such as cyclic AMP (cAMP), calcium ions, etc., leading to cellular responses. Many of the signal-mediated cascade pathways responding to stress lead to changes in gene expression (Hunter, 1995; Seger and Krebs, 1995). Mitogen-activated protein kinases (MAPKs), for example, are activated by osmotic stresses (Matsuda *et al.*, 1995). Hence, gene expression is one of the endpoints of signal transduction pathways and continually reshapes the cell in response to metabolic needs and environmental stimuli.

Stress-Induced Gene Expression

Environmental stress can trigger changes in gene expression. To date, the heat shock response has been the best characterized of the stress responses and is often used as an example. In response to elevated temperatures, cells produce a select group of heat-

shock proteins (hsps) in high levels. Some of these hsps are present constitutively in cells; others appear as de novo protein products under stress conditions only (Lindquist and Craig, 1988). Most hsps appear to function to protect cellular proteins from stress-induced denaturation. Some act as molecular chaperons to assist protein folding, higher order assembly, translocation and protein degradation (Lindquist, 1986; Parsell and Lindquist, 1993). The heat-shock response is widespread and highly conserved among organisms. Hsps are also induced by a variety of other stresses including heavy metals, amino acid analogues, alcohol and oxidants (Hightower, 1991) as well as in pathophysiological circumstances such as cerebral and cardiac ischemia (Morimoto et al., 1994). Regulation of the heat shock response involves a highly conserved regulatory region of the hsp genes called the heat shock element (HSE) which is activated by a heat shock transcription factor (HSF). HSF is present in a latent state under normal conditions and is activated upon heat stress to bind HSE for transcription activation. The stress signal is thought to be transduced to HSF by changes in the physical environment, by the activity of HSF-modifying enzymes, or by changes in intracellular levels of hsps (Morimoto, 1993).

Gene expression in response to an abrupt shift to lower temperatures (cold-shock) has also been investigated recently (Jones and Inouye, 1994; Thieringer *et al.*, 1998). Data from studies with *Escherichia coli* have shown that selected cold-shock proteins (csps) are induced when cells are abruptly shifted to a lower temperature, the magnitude of induction being generally dependent on the size of the temperature shift. Since an inhibition of translation by certain antibiotics (tetracycline, *etc.*) can also induce the cold-shock response, it has been suggested that the state of the ribosome may be the

physiological sensor for the induction. The initiation of translation and ribosome assembly are inhibited at low temperature and, thus, the cold shock response may function to overcome the partial block of protein synthesis (Thieringer *et al.*, 1998).

The heat shock and cold shock responses are transient responses that are intended to help minimize damage to cells from brief exposures to extreme conditions that are generally outside the normal experience of the organism. However, many organisms must deal with prolonged low temperature exposure during the winter months and the role of gene expression in adapting organisms to long term cold stress is of great interest. To date, the subject has been most extensively studied in plants due to interest in improving the cold- or frost-hardiness of agricultural crops (Thomashow, 1993; Jaglo-Ottosen *et al.*, 1998). Various cold-induced proteins have been identified that act as membrane protectors to enhance freezing survival of plant cells (Artus *et al.*, 1996). Tulips, which require cold exposure to flower, show cold-induced expression of both a water channel protein (gammaTIP) and the enzyme invertase, which together lead to an increase in osmotic potential and provide a driving force for the growing stalk (Balk and de Boer, 1999). Cold-induced gene expression has also been documented in some fish species (Tiku *et al.*, 1996) and in fish cell culture systems (Yamashita *et al.*, 1996).

Gene expression is also affected by oxidative stress, low oxygen tension, and cell volume changes. Oxidative stress arises from the harmful effects of reactive oxygen species (ROS) which damage nucleic acids, protein and lipids. Cells deal with oxidative stress by expressing proteins with antioxidant activities and enzymes to repair damaged macromolecules (Sen and Packer, 1996). Low oxygen tension is a feature of many physiologic and pathologic conditions, such as wound healing or ischemia, and induces

the transcription of many genes (Bunn and Poyton, 1996). Transcription of the erythropoietin gene (*Epo*) is a well-known example. Erythropoietin is a plasma glycoprotein that stimulates production of red blood cells, thus increasing the oxygen-carrying capacity of the blood and allowing the animal to tolerate limited oxygen availability. Hypoxic conditions also induce the transcription of HIF-1, the hypoxia inducible transcription factor (Semenza and Wang, 1992), which activates a range of genes, such as those encoding erythropoietin, growth factors, phosphoglycerate kinase 1, and glucose transporters (Bunn and Poyton, 1996). To date these responses have only been characterized in mammals. Finally, changes in cell volume associated with dehydration, osmotic and water stress also result in specific gene expression (Burg *et al.*, 1997; Matsuda *et al.*, 1995).

The paragraphs above describe a few examples of environmentally induced alterations in gene expression but an enormous roster of genes are affected when stress is imposed upon an organism. An illustration of the gene responses to oxygen deprivation (Michiel et al., 2000; Semenza, 2000; Wenger, 2000; Zhu and Bunn, 1999; Shih and Claffey, 1998; Gleadle and Ratcliffe, 1999) and cold exposure (Los and Murata, 1999; Yamanaka, 1999; Shinozaki and Yamaguchi-Shinozaki, 2000; Phadtare, et al., 1999; Thieringer et al., 1998) can be found in these reviews . Clearly, the majority of studies to date have been conducted using isolated cells or cell lines or using animals that are not tolerant of the given stress. Thus, to date, we have a limited outlook on the extent of metabolic functions affected when cells/animals respond to stress. In the present study I set out to examine the gene expression responses to anoxia and freezing by an organism that is tolerant of both of these environmental stresses.

An animal model system for stress-induced gene expression: *Littorina littorea*

The intertidal ecological niche

The intertidal zone consists of the landscape created when the ocean tide is at its lowest ebb (Figure 1.1). The rocky shores of the Atlantic provide an ideal scaffold for the creation of small, temporary tide pools, forming an entire ecological niche for a variety of animal species (Figure 1.2). These pools, and indeed the entire beach, are highly variable environments where factors such as temperature, salinity and oxygen availability fluctuate on daily and seasonal levels. As a consequence, inhabitants must possess a high degree of metabolic plasticity. Most are highly tolerant of oxygen deprivation which ensures survival of these gill-breathing species when the tide recedes. Species living at high latitudes are also frequently exposed to subzero air temperatures at low tide during the winter months and, therefore, many have developed the ability to survive bouts of freezing. The marine gastropod mollusc *Littorina littorea* (the periwinkle snail) is one such intertidal inhabitant on northern shores which displays remarkable abilities to endure wide variations in multiple environmental parameters.

The subject animal: *Littorina littorea*

L. littorea is both freeze-tolerant and anoxia-resistant. It can survive freezing at -8°C for at least 8 days although mortality increases at lower temperatures; LD_{50} values at -9, -11 and -13°C were 7 days, 28 hours and 4 hours, respectively (Murphy and Johnson, 1980; Murphy, 1983). Tolerance of freezing is greater in winter than summer and is enhanced by acclimation to higher salinity, which increases intracellular

osmolality, thereby reducing the amount of ice that can form at any given temperature. Unlike many insects and other invertebrates, *L. littorea* does not produce specific cryoprotectants such as polyhydric alcohols (Storey and Storey, 1988) but amino acids, anaerobic end-products and Ca^{2+} appear to help stabilize membrane structure during freezing (Murphy, 1983; Hayes and Loomis, 1985). Pre-exposure to anoxic conditions also enhances freezing survival by at least two mechanisms: (1) a build up of metabolic end products raises cellular osmolality, and (2) anoxia-induced metabolic rate depression lowers energy demands by ice-encased tissues (Storey and Storey, 1988).

Anoxia Survival

Gill breathers such as *L. littorea* can only extract oxygen from aqueous solution. Therefore, when these organisms are out of water at low tide, they soon deplete their "on-board" supplies of oxygen and subsequent survival depends on their ability to survive without oxygen for extended time periods. Anoxia tolerance for most intertidal invertebrates is accomplished via a combination of biochemical strategies. Firstly, the tissues of these animals maintain large stores of fermentable fuels (including glycogen and the amino acid, aspartate) that can be catabolized to produce ATP without the participation of oxygen. In addition, they catabolize these fuels via alternative pathways (ie. not just glycogen → lactate as is the standard in mammalian tissues) that are linked with additional sites for substrate-level phosphorylation of ADP and that produce end-products including alanine, succinate and volatile fatty acids (deZwaan, 1983; Storey and Storey, 1990; Storey 1992). Thus, whereas the fermentation of 1 mole of glucose to

lactate produces 2 moles of ATP, conversion of glucose to succinate or propionate yields 4 or 6 moles ATP/mole glucose, respectively (Hochachka and Guppy, 1984). Finally, and perhaps most importantly, anoxia-tolerant marine molluscs sharply lower their metabolic rate when oxygen is depleted so that anaerobic metabolic rate is less than 10% of the resting aerobic rate. Hypometabolism greatly extends the time that their fixed reserves of metabolic fuels can support survival.

Freezing tolerance

At higher latitudes, intertidal animals often encounter subzero temperatures during aerial exposures at low tide in the winter. Survival depends on mechanisms that allow the animals to either avoid or endure freezing, adaptive strategies that are termed freeze avoidance and freeze tolerance, respectively. Freeze-avoiding species supercool to maintain their body fluids in a liquid state and avoid the formation of ice in their tissues. Freeze tolerant species permit and even encourage ice to form in extracellular fluid spaces of their bodies but take steps to prevent ice formation within cells. These contrasting strategies have similar components including a seasonal reduction in the amount of freezable water in the body, synthesis of low molecular weight cryoprotectants (such as glycerol, glucose) that increase the osmolality of body fluids, and a shift to a hypometabolic state (Storey and Storey, 1988). Unique proteins are synthesized in both types of animals to enhance survival. Antifreeze proteins (AFPs) occur in most freeze-avoiding species to inhibit the growth of ice crystals whereas freeze tolerant animals often have specific ice-nucleating proteins (INPs) that initiate and regulate ice formation in appropriate extracellular spaces. AFPs have also been found in

some freeze tolerant species (despite the apparent contradiction in terms) where they appear to function to minimize recrystallization, the process by which small crystals regroup to form larger crystals which could do physical damage to tissues. Both types of proteins are synthesized seasonally in response to environmental cues. AFPs have been reported in one species of intertidal gastropod whereas INPs have been found in several intertidal bivalves (e.g. *Mytilus edulis*) and gastropods (e.g. *Melampus bidentatus*). Although Loomis *et al.* (1985) did not find AFPs in *L. littorea,* the seasonal expression of INPs or AFPs in intertidal species suggests a strong genetic component to the induction and maintenance of freezing survival strategies.

Biochemical observations indicate that responses to freezing and anoxia are different

While oxygen deprivation and freezing are by definition different stressors, they do share several characteristics. Anoxia is a highly specific single stress, affecting only the available cellular oxygen supplies, but it elicits a variety of metabolic responses to deal with the potential energy stress caused by blocking oxidative phosphorylation. Freezing is a more complex stress and contains several intertwining components. One of these is oxygen deprivation because the freezing of extracellular body fluids cuts off the delivery of oxygen to tissues. Plasma or hemolymph freezing also causes cessation of substrate delivery and waste removal via the blood. Extracellular freezing also withdraws a large percentage of total body water into crystals of pure ice (often up to 65%) and this has huge effects on intracellular volume, ionic strength and osmolality. The greater complexity of freezing stress may be the reason that a wider variety of metabolic end-

products accumulate during freezing versus anoxia; for example, in the marine bivalve, *Geukensia demissus*, only succinate accumulates during anoxia whereas lactate, succinate and alanine all accumulate during freezing to levels 3-4 fold higher than in anoxia (Storey and Churchill, 1995). In addition, all tissues build up glucose during freezing but not during anoxia. These differences suggest that the metabolic controls and responses that underlie anoxia tolerance and freezing survival have common and disparate elements.

Genetic observations in stress tolerant animals

Although the biochemical adaptations in response to stress have been elucidated in many stress-tolerant animals, including *L. littorea*, gene expression responses that enhance survival are largely unexplored. In recent years, however, several lines of supportive evidence for the involvement of stress-induced gene expression in natural adaptation to low oxygen or freezing stresses, have been delineated. For example, analysis of anoxia-induced changes in protein synthesis (via ^{35}S-methionine or ^{14}C-leucine labeling) in turtles has documented the production of new proteins in the anoxic state (Brooks and Storey, 1993; Hand and Hochachka, 1995). In addition, an assessment of changes in organ mRNA populations during anoxia and recovery via *in vitro* translation showed organ-specific changes in mRNA complement (presumably indicative of altered protein synthesis) during both anoxia exposure and aerobic recovery (Douglas et al., 1994). None of these studies, however, were able to identify or assign functions to any of the differentially-expressed radiolabeled proteins. In designing the current study, therefore, a genetic approach to the problem seemed to hold the best probability for finding and identifying new proteins that contribute to stress tolerance in *L. littorea*.

Experimental approach:

Studying gene expression in response to environmental stress in *L. littorea.*

Gene expression is a key factor that regulates cellular protein complement and changes in gene expression are the primary mechanism of cellular response to extracellular signaling. Combined with the fact that a plethora of environmental stresses induce gene expression, it is reasonable to predict that changes in gene expression also occur in stress tolerant animals when faced with harsh external conditions. However, the broad range of affected genes, combined with the novel nature of stress-tolerant animals, makes it impossible to hypothesize which genes will be up-regulated in *L. littorea.* Although many have a high probability of being induced by freezing or anoxia in *L. littorea*, to specifically hunt for them would be a prohibitively time-consuming "hit and miss" strategy. A more useful approach, when studying an animal whose genetic reponses have never been previously investigated, is to construct and differentially screen a cDNA library. Isolation and characterization of clones that show increased mRNA accumulation in a stressed state can highlight the participation of selected genes and the cellular functions that they support, as important in the tolerance of stress. Therefore, I propose that the frozen state or changes in oxygen levels respresent stressors that can induce *L. littorea* to respond with increased expression of specific genes. The main questions to be addressed in this thesis are:

1) Do freezing and anoxia induce gene expression in *L. littorea*? cDNA libraries were constructed from the foot muscle of freezing-exposed and anoxia-exposed snails. These were differentially screened to isolate stress-responsive genes.

2) What genes are up-regulated? Stress-induced genes were sequenced and submitted as queries to public databases to determine whether they represented known proteins or novel proteins specific to *L. littorea.*

3) What are the differences (if any) between the two stresses? Northern blot analysis was used to determine whether the genes that were up-regulated in response to either freezing or anoxia were also responsive to the other stress. In addition, by calculating the fold increase in transcript levels (relative to the control state) the relative level of gene induction was assessed for each stress.

4) Are the genes tissue-specific? Northern blot analysis was used to determine whether genes that were up-regulated by anoxia or freezing in foot muscle were also stress-induced in snail hepatopancreas.

Figure 1.1: **The intertidal zone along the Bay of Fundy, near Annapolis Royal, Nova Scotia.** The Bay of Fundy boasts the highest tides in the world. The photo shows the tide near its lowest ebb and illustrates the rocks and kelp beds that create a unique environmental niche and the native habitat of *Littorina littorea.*

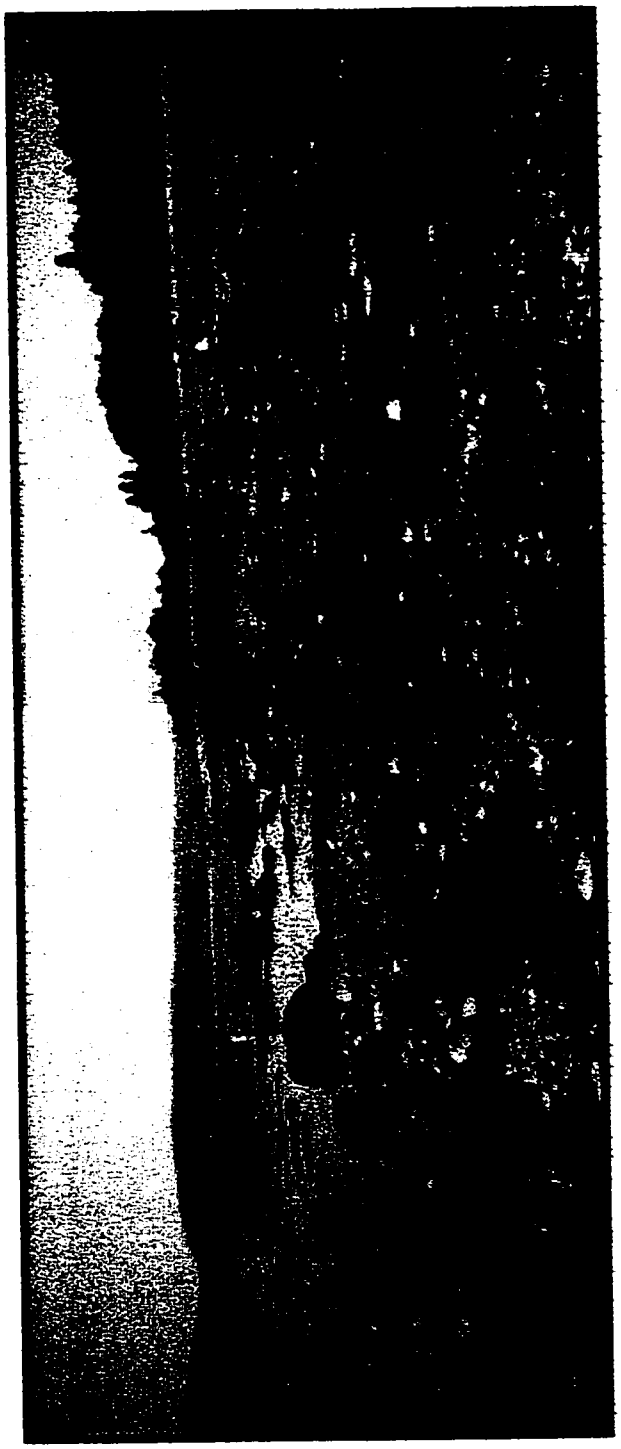
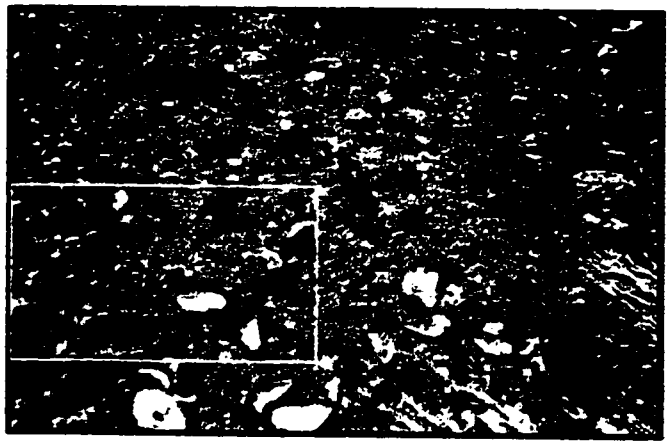
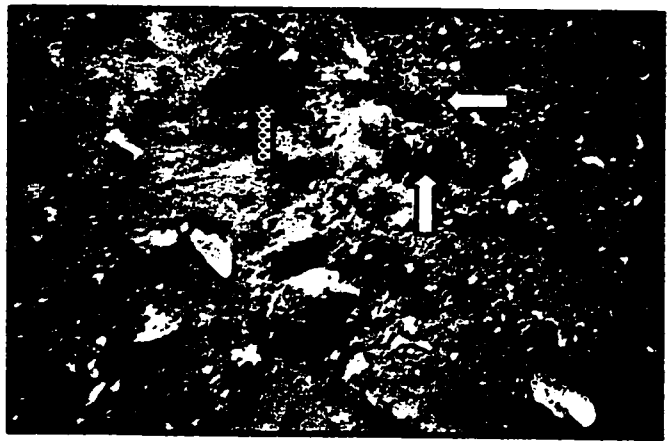


Figure 1.2: Intertidal pool created at low tide along the Bay of Fundy.

These photos illustrate a single tidal pool at several magnifications. Arrows in the bottom photo show *L. littorea* (white) and blue mussels, *Mytilus edulis* (dotted).





CHAPTER TWO

DIFFERENTIAL GENE EXPRESSION DURING FREEZING AND ANOXIA IN THE FOOT MUSCLE OF *LITTORINA LITTOREA*: SCREENING OF cDNA LIBRARIES

INTRODUCTION

The cells of all species, whether bacterial, animal or plant, are capable of cellular and molecular alterations in response to external stresses or environmental challenges. Differential gene expression is a common and essential response to stressors in an effort to survive the stress. Although the population of message RNAs present at any given time results from the balance between the rates of its synthesis and degradation, the composition of the mRNA pool generally provides a good indication of the proteins that are currently being synthesized. By comparing the pools of mRNA transcripts present in a tissue under two different states (e.g. control versus stressed), key stress-induced changes in mRNA transcript levels can be spotted. Subsequent identification of the corresponding genes and proteins indicates various specific cellular loci and types of cell functions that are modified for survival of extreme stress.

Although mRNA is the desired target for analysis of differential gene expression, messages are notoriously labile and difficult to amplify. Therefore, techniques have been developed that make use of reverse transcription (RT) to synthesize a complementary DNA (cDNA) that can be inserted into a phage vector. This increases the stability of any given transcript while the nature of the phage allows for self replication of the plasmid with the insert DNA. A representative cDNA library should contain full-length copies of the original population of gene transcripts from the source tissue (due to the size-selection process described in the methods) so screening of such a library with ^{32}P-labeled cDNA probes synthesized using mRNA isolated from stressed vs. unstressed tissues will reveal differentially expressed genes. The cDNA clones that are up- or down-regulated in response to the stress or change of state can then be selected for sequence analysis to identify the differentially expressed genes.

The lambda ZAP insertion vector system (Strategene, La Jolla, CA) is commonly used for the construction of cDNA libraries. The vector system combines the high

efficiency of lambda library construction and the convenience of a plasmid system with blue-white colour selection. The vector can accommodate inserts of up to 10 Kb in length, can be screened with DNA probes or antibody probes, and can be used for in vivo excision to form a phagemid. The polylinker of the phagemid has 21 unique cloning sites flanked by T_3 and T_7 promoters and a choice of 6 different primer sites for DNA sequencing. The vector can also be used for site-directed mutagenesis, unidirectional deletions, in vitro transcription to generate riboprobes used for Southern and northern blotting, and the expression of fusion proteins.

Several concurrent studies of environmentally induced gene expression have been carried out in our lab using differential screening of cDNA libraries. The subjects of these experiments have included a wide variety of animals (crustaceans, insects, reptiles, amphibians, mammals) and the isolated up-regulated genes fall into several categories. Proteins encoded on the mitochondrial genome have been the most commonly isolated, and include NADH–ubiquinone oxidoreductase (subunits 2, 4, and 5 from turtle heart), cytochrome b (*Cytb*) (turtle brain), cytochrome oxidase (subunit III in frog brain, subunit I in turtle heart), and the transfer-RNA gene series WANCY (turtle heart) (Cai and Storey, 1996, 1997; Wu 1999). Genes encoding proteins that are associated with ATP were up-regulated in response to freezing in the brain of *Rana sylvatica* (ATPase subunits 6 and 8) and anoxia in turtle heart (ADP/ATP translocase) (Cai and Storey, 1997; Wu 1999). Ribosomal proteins were found to be up-regulated by freezing in the frog *Rana sylvatica* and the insect *Eurosta solidaginis*, two unrelated animals that are both freeze-tolerant (Wu, 1999; Bilgen 1998). The gene for the acidic ribosomal protein was up-regulated in both animals, whereas ribosomal protein L7 was isolated using differential display PCR from frog skin (Wu, 1999). Finally, the gene for the ventricular isoform of myosin light chain 1 (*MLC1v*) was induced during hibernation in heart of Richardson's ground squirrels, a result that may indicate a restructuring of muscle proteins for function at the low body temperature (usually <5°C compared with 37°C in euthermia) of the

hibernating state (Fahlman and Storey, 2000). In addition to identifiable genes, differential screening of cDNA libraries from unique systems frequently results in the isolation of novel clones representing proteins that have never before been characterized. For example, a gene encoding a novel protein with a nuclear export sequence was shown to be up-regulated during freezing in wood frogs (Cai and Storey, 1997). Every other cDNA library screening that has been completed by our lab has also yielded at least one novel sequence. These stress-induced genes/proteins have been found in multiple species: anoxic crayfish (hepatopancreas and tail muscle), glucose-injected wood frogs (liver and heart), anoxic adult turtles (mixed organ library), and the estivating snail, *Otala lactea.*

Finally, initial genetic studies have indicated the involvement of metabolic functions that had never before been suspected as contributing to stress tolerance. For example, in the freeze tolerant frog, *Rana sylvatica,* transcripts for fibrinogen α and γ subunits were induced 3-fold over control levels in the liver, lung and gut early in the freezing process (Cai and Storey, 1997a). Expression of this plasma clotting protein during freezing suggests that it may be part of an enhanced damage-repair mechanism to deal with any bleeding injuries as the result of ice damage. In another example, expression of riboflavin binding protein (RfBP) is induced in liver during long-term estivation in the desert toad, *Scaphiopus couchii.* RfBP functions to bind plasma riboflavin and load the vitamin into the eggs or fetus of animals. Up-regulation of this gene in liver of estivating toads may be linked with maturation of eggs in preparation for the explosive breeding that occurs immediately upon emergence from estivation. Elevated levels of RfBP might also function to allow the adult to "cache" riboflavin and maintain an endogenous vitamin pool over the 9-10 months of each year that toads are dormant (Storey et al., 1999). Such results illustrate the most useful aspect of the cDNA library screening approach and that is that it allows us to identify previously unknown contributors to the stress-survival mechanism.

The present chapter describes the construction and differential screening of cDNA libraries made from mRNA isolated from the foot muscle of anoxia- and freeze-exposed *L. littorea* and the isolation of several genes that are up-regulated during anoxia or freezing exposure. As a methodological note, since the up-regulation of different genes may take place after different lengths of exposure to stress (in this case freezing or anoxia), the cDNA libraries were made using mRNA pooled from tissues sampled at multiple time points: 1, 12 and 24 hours of stress treatment.

MATERIALS AND METHODS

Collection and treatment of animals

Marine periwinkles, *Littorina littorea*, were purchased from a local seafood market and held for 3 weeks in the laboratory at 5°C in aerated full strength seawater (1000 mOsmol). Control snails were sampled from this condition. Shells were quickly cracked open, and foot and hepatopancreas were immediately removed, frozen in liquid nitrogen, and transferred to -80°C for storage until use. To impose anoxia, other snails were transferred to closed jars with about 1-2 cm of seawater in the bottom. The water had been previously bubbled with nitrogen gas (containing 0.03% CO_2) for at least 10 minutes and N_2 gassing was continued for 15-20 minutes after adding snails. Jars were then sealed and maintained at 5°C for 1, 5, 12, 24 , 48 or 144 hours (during this period, most snails climbed up the walls of the jar). Animals from one jar were sampled at each time point. For freezing exposure, animals were removed from aerated seawater, placed in closed plastic containers, and transferred to -8.0°C in an incubator. In initial tests, cooling and ice nucleation were monitored for a few snails by placing a thermistor inside the shell in contact with the mantle. Nucleation was detected by the sharp rise in body temperature when freezing began (latent heat release of crystallization) and typically occurred within 45 minutes under the experimental conditions used. In subsequent trials, therefore, a 45 minute cooling period was allowed prior to beginning to time the experimental freezing exposures of 1, 5, 12, 24 , 48 or 144 hours. After freezing or anoxia exposure, snails were rapidly removed from the containers and processed as for controls.

Total RNA extraction

Total RNA was extracted from frozen tissue using Trizol™ reagent according to the manufacturer's instructions (Gibco-BRL, Bethesda MD). In brief, tissues were homogenized in Trizol reagent (1-2 mL per 100 mg) and then incubated at room

temperature for 5 minutes. Chloroform (0.2 mL per mL of Trizol) was added to the homogenate which was then shaken and left at room temperature for a further 10 minutes. Samples were then centrifuged at 12,000 x g at 4°C for 15 minutes to separate the homogenate into 3 layers. The upper, aqueous layer contained the RNA and was removed to a new sterile tube. Addition of n-propanol (0.5 mL per mL of Trizol) precipitated the RNA. After 10 minutes at room temperature, the samples were then centrifuged as above and the RNA pellet was washed in 70% ethanol, air dried for 10 minutes and redissolved in DEPC-treated water (smallest amount required). Samples were stored at -20°C (short-term) or -70°C (long-term). Quality and purity were assessed spectrophotometrically as well as visually using denaturing formaldehyde gels (see methods in Appendix 2). Intact total RNA from invertebrates shows a single broad ribosomal band at roughly 1.34 Kb.

mRNA isolation

Where required, polyA mRNA was isolated from total RNA using oligo dT cellulose (New England Biolabs) chromatography according to the method of Sambrook *et al.* (1989). Dry oligo dT cellulose was combined with 0.1 M NaOH, poured into a sterile column, then equilibrated in loading buffer. Total RNA was heated to 65°C (10 minutes) and brought to a final concentration of 0.5 M NaCl. The sample was loaded onto the column and washed with an equal volume of 1X loading buffer. The eluant was collected, heated and reloaded twice more. The column was washed with 10 mL loading buffer followed by 10 mL middle wash until the A_{260} of the eluant was zero. The mRNA was then eluted in 6 aliquots (0.5 mL each) of elution buffer and precipitated with sodium acetate and ethanol overnight at -20°C. The mRNA was pelleted by centrifugation (20 minutes at 12,000 x g) and resuspended in 10 µL DEPC dH_2O. Quantity and purity were assessed spectrophotometrically (see Appendix 2) and the mRNA was stored at -70°C for future use.

cDNA library synthesis

Periwinkles tolerate both freezing and anoxia, thus cDNA libraries were constucted for each condition. I made the first library using mRNA extracted from foot muscle of frozen snails, whereas the second (from foot muscle of anoxia-treated animals) was made by Dr. Qin Yin Cai (now of Bio S&T, Lachine, PQ). In both cases, equal amounts of mRNA isolated from animals treated for 1, 12 and 24 hours were pooled (totaling 5 μg) and used to construct cDNA libraries using the Uni-ZAP-cDNA synthesis kit (#066004, Strategene, La Jolla, CA) according to manufacturer's instructions. The following general protocol was used for both libraries.

Preparation of host cells

Two strains of *E. coli* were used during these experiments: XL1 Blue and SOLR. XL1Blue cells were grown on LB-agar plates containing 12.5 μg of tetracycline per mL of medium, while SOLR cells were maintained on LB plates containing kanamycin (50 μg/mL). When liquid cultures were required, LB broth containing appropriate antibiotic was inoculated with a single colony and grown with shaking at 37°C to late log phase.

First and second strand synthesis

The first strand of cDNA was made from the substrate mRNA. A nucleotide mixture containing dNTP's (methylated CTP) was added to the mRNA with RNase inhibitor, linker primers, ^{32}P-dATP and reverse transcriptase (type MMLV). First strand synthesis was complete after incubation at 37°C for one hour. A 45 μL aliquot of this mixture was then combined with a fresh dNTP mixture (CTP not methylated), ^{32}P-dATP and the appropriate buffer. After mixing, RNase H and DNA polymerase I were added. RNase H nicks the RNA bound to the first strand DNA, producing multiple fragments that act as primers for DNA polymerase I. Incubation was carried out at a constant 16°C.

Modifications to the cDNA fragments: blunting the cDNA termini, ligating EcoR1 adapters, phosphorylation of the EcoR1 Ends

The termini of the cDNA were blunted by incubation with *Pfu* DNA polymerase, which fills in any missing nucleotides in the double stranded cDNA. Following removal

of protein and ethanol precipitation, cDNA was resuspended in EcoR 1 adapters which have the following sequences:

5′ AATTCGGCACGAG 3′

3′ GCCGTGCTC 5′

The adapters were ligated to the blunt ends of the cDNA with T_4 DNA ligase, thus ensuring the correct orientation of the clone when inserted into the vector. The adapter ends were then phosphorylated to enable insertion of clones into the dephosphorylated vector arms. Xho 1 digestion releases the EcoR 1 adapter and residual linker-primer from the 3' end of the cDNA. Appropriately sized fractions were ligated to the Uni-ZAP XR vector arms with T_4 ligase. Finally, the cDNA was packaged into lambda phage using the Gigapack II packaging extracts. The recombinant phage were then used to infect XL1-Blue *E. coli* cells.

The cDNA is inserted into the β-galactosidase gene of the vector DNA thereby disabling the gene. By growing infected XL1-blue cells on plates with X-gal and IPTG, it is possible to determine both the titer of the library, number of plaque-forming units (pfu) per mL, and the percentage of pfu that do not contain an insert. Intact phage produce blue plaques, while recombinant phage produce white plaques.

Titre determination

The titre of a phage stock is presented in terms of pfu per mL. Serial dilutions of the library were prepared in SM buffer and XL1-Blue host cells were cultured at 37°C in LB broth. A 100 μL aliquot of each library dilution was mixed with 400 μL cells in separate tubes and incubated at 37°C for 20 minutes. The cells of each dilution were added to 5 mL of molten NZY top agar containing IPTG (140μg/ml) and X-gal (350μg/mL), then thoroughly mixed and poured over an NZY bottom agar plate pre-warmed to 37°C. The plates were incubated overnight at 37°C. The titer was determined by counting the numbers of plaques on each plate, and using the following formula:

pfu/ml = (number of plaques × dilution factor) ÷ volume of extract plated.

Amplification of the library titre

To enhance the stability of the library, it is necessary to increase the titre. Approximately 50,000 plaques were grown on each of ten 24 cm x 24 cm NZY plates using the same procedure as for titre determination, but using 1 mL of XL-1 blue cells and 30 mL of top agar. After 10 hours of incubation at 37°C, phage was recovered from the surface of the plates by overlaying them with SM buffer and incubating overnight at 4°C with rocking to bring the phage into solution. The buffer was then collected, brought to 5% chloroform and centrifuged at 5000 x g for 15 minutes to remove cell debris. The supernatant was transferred to another sterile container. After adding dimethylsulfoxide (DMSO) to 7% v/v, the amplified library was aliquoted into sterile 1.5 ml tubes and stored at -80°C. The titre of the amplified library was determined as described above.

Screening Protocol

Differential screening was used to isolate stress-induced genes. The protocol used was adapted from Strategene kit instructions and Sambrook *et al.* (1989). Duplicate plaque lifts were prepared and hybridized with ^{32}P-labeled probe synthesized from mRNA. One lift was hybridized with probe constructed using mRNA from the stressed condition, the other with control. Autoradiography films of the corresponding lifts were then compared. Plaques present on the stress film and not the control were isolated as putative stress-induced genes.

Plaque lifts

A sufficient number of 24 x 24 cm NZY plates were prepared such that they contained 40,000 pfu so that each clone would be represented at least once. Duplicate plaque lifts were made with Hybond-N 0.45 μm nylon membranes (Amersham, now Pharacia Biotechnologies). The adsorption time was one minute for the first lift and 2 minutes for the second. Membranes were then soaked in denaturation (2 minutes), neutralization (4 minutes) and rinse (30 seconds) solutions. Adsorbed DNA was

permanently fixed to the membrane with UV exposure (120,000 μjoules/cm^2, 254 nm) followed by heating at 80°C for 1-2 hours.

Probe synthesis and hybridization

To prepare probe, 1 μg of denatured mRNA was incubated with dNTP's (no dCTP), oligo dT primers, RNasin and DTT for 10 minutes at room temperature to allow the primers to anneal. Then reverse transcriptase and ^{32}P-dCTP were added and the mixture was incubated for 1 hour at 37°C. Following cDNA synthesis, the RNA was degraded with NaOH at 65°C for 30 minutes. Once cooled to room temperature, the mixture was neutralized with 2 M HCl, and passed through a Sephadex G-50 spun-column to remove unincorporated nucleotides.

One of the duplicate plaque lifts was hybridized in a Labline Model 308 hybridization oven (VWR Canlab) with single-stranded cDNA probe made from control animals whereas the other was incubated with cDNA probe from stressed animals. Stress probes were prepared using a mixed pool of mRNA containing equal amounts of message from each of the three time points used for library synthesis. One million cpm of probe was added per mL of hybridization fluid. Hybridization was continued for 12-15 hours whereupon the plaque lifts were washed (15 minutes) in serial dilutions of SSC (6X, 2X, 0.5X, 0.2X) and 0.2% SDS until the counts on the lifts measured between 1000-5000 cpm (as determined with a Geiger counter). Lifts were then exposed to Kodak X-OMAT film for an appropriate length of time (equal time for control and stressed hybridizations in each comparison). Control and stress films were aligned and compared to identify clones present either solely during stress or at a much higher intensity than control.

Putative positive clones were excised from the original NZY plate using a flamed Pasteur pipette. For each agar plug containing a positive plaque, 0.5 mL of SM buffer and 100 μL of chloroform were added to kill bacterial cells. After removing cellular debris, the supernatant containing the phage was titred and plated. Depending on the number of putatively up-regulated clones obtained from the primary screen, a secondary screen was

also performed. Excised phage were pooled into one or more groups, titred, replated, and new duplicate plaque lifts were made and screened. When the number of positive clones was of a manageable level (<50 were usually identified after a secondary screen), they were isolated and plated onto separate 10 cm NZY plates at a density of 50-100 pfu/mL. A final round of plaque lifts were screened with control and stress probe (steps as above). In most cases the appropriate positive plaque could be isolated at this point. The clones were then transferred to the p-BLUESCRIPT vector (method below) to allow the cDNA to be excised and/or sequenced, as well as to maintain a stable stock in bacterial cells.

***In vivo* excision of potential λZAP clones to pBLUESCRIPT**

The λZAP vector can be converted from its bacteriophage, double stranded DNA form to a phagemid form with assistance from a helper phage. The phagmid can then be used to infect a different strain of *E. coli* called SOLR. In SOLR cells, the single stranded phagemid circularizes to form the double stranded phagemid, pBLUESCRIPT, which can then replicate autonomously in the host cells.

Broth cultures of XL1 Blue and SOLR cells (optical density at 600 nm (OD_{600}) = 0.2-0.5) were prepared from fresh streak plates. The cells were centrifuged at 4000 rpm for 5 minutes and resuspended in 10 mM $MgSO_4$ to an OD_{600} of 1.0. XL1BLUE cells (200 μL) were combined with 250 μL of phage stock (potential clone with a titer $>10^5$) and 1 μL of ExAssist helper phage (Strategene). After incubating at 37°C for 15 minutes, 3 mL of LB broth were added followed by incubation at 37°C for 2-2.5 hours. Following centrifugation at 1500 x g (15 minutes) the supernatant was transferred to a new sterile 15 mL polypropylene tube and then heated to 70°C for 15 minutes. Cell debris was removed by centrifugation at 4000 rpm (15 minutes) leaving the supernatant as a filamentous phagemid stock. This was transferred to sterile 1.5 mL tubes and stored at 4°C.

Each filamentous phage stock was then used to infect SOLR cells and grown on ampicillin plates (37°C, overnight). One of the resulting colonies was streaked on a new

LB-ampicillin plate to create a master plate of the SOLR cells containing the phagemid of interest. Plates were stored at 4°C for up to 2 weeks. One colony from each plate was used to inoculate LB broth and grow a saturated liquid culture to be used for a plasmid preparation (see below) and create a glycerol stock (see Appendix 1).

Preparation of plasmid (miniprep)

One mL of an overnight culture of the cells containing the clone of interest was added to 10 mL of fresh LB ampicillin broth and grown to an OD_{600} of 1.5-2.0. Cells were pelleted (2000 x g), resuspended in 1 mL GTE and placed on ice. Plasmid isolation was achieved in 3 steps: cell lysis with addition of 2 mL 0.2 M NaOH + 1% w/v SDS (incubated at room temperature for 5 minutes); neutralization with addition of 1.5 mL KAcF (10 minute incubation on ice) to precipitate chromosomal DNA and cellular debris which was then removed by centrifugation (12,000 x g at 4°C for 10 minutes); and precipitation, with addition of 4.5 mL of isopropanol. After a 15 minutes incubation at room temperature, the sample was centrifuged as above, and the DNA pellet was air-dried for 5 minutes and then resuspended in 1 mL of $T_{50}E_{10}$ buffer.

RNaseA (150 μg) was added to digest residual RNA (37°C for 30 minutes). Contaminating protein was removed by serial extraction with equal volumes of phenol, then phenol/chloroform (1:1 mix), then chloroform and the top, aqueous phase was transferred to a new tube after each extraction. DNA was precipitated with the addition of sodium acetate (0.1 volumes, 3.2 M) and 100% ethanol (2.5 volumes) followed by 1 hour incubation at -20°C. The precipitated plasmid was pelleted by centrifugation (10 min, 12000 x g), washed with 70% ethanol, and resuspended in sterile dH_2O.

DNA agarose gel electrophoresis

Purified plasmid DNA was treated with restriction enzymes and analyzed on an agarose gel to estimate the size of the insert and purity of the mini-prep. Enzymes and related buffers were obtained from Gibco-BRL. A 2-10 μg aliquot of plasmid DNA in 7 μ

L with sterile water was combined with 10 units each of EcoR1 and Xho1, and 1 μL of 10X reaction buffer. The mixture was incubated at 37°C for 90 minutes and then loaded onto a TAE 1% agarose gel with 2 μL of loading dye. DNA fragments were separated with electrophoresis at 10V/cm and the DNA was stained with ethidium bromide.

Reverse northern as a tertiary screen and DNA sequencing

To further eliminate false positives, reverse northern blots[Ψ(see below)] were performed on clones isolated from the library screen. Purified plasmid was digested and electrophoresed as above, but only 200 ng of each plasmid digest was loaded onto duplicate gels. When electrophoresis was complete, the gel was rinsed in a denaturing solution for 30 minutes, followed by several brief rinses in distilled water, 2 x 15 minute washes in a neutralization solution, and 15 minutes in 20X SSC (composition of solutions are listed in appendix 2). DNA was transferred to a nylon membrane (Hybond) by upward capillary transfer, with 20X SSC as the transfer buffer. After 5 hours, membranes were exposed to UV and then heated to 80°C for 2 hours to fix the DNA. Duplicate membranes were hybridized with ^{32}P-labeled probes from control and stressed tissue as described above. Positive clones were sequenced using an automated procedure by BioS&T (Lachine, PQ) or Canadian Molecular Research Services (CMRS, Ottawa, ON).

Ψ A standard northern blot analysis is a protocol whereby total RNA from control and stressed conditions are separated in a denaturing agarose gel, and then transferred to a nylon membrane which is hybridized with a gene-specific probe to elucidate the gene's expression pattern. A reverse northern blot contains the gene-specific insert on the membrane and is hybridized with total mRNA probe from a given state.

Formaldehyde agarose gel RNA electrophoresis and northern blot assembly

RNA molecules were separated on a denaturing agarose gel according to the method of Sambrook *et al.* (1989). The gel was prepared by melting 3 g of solid agarose (electrophoresis grade, GIBCO BRL) in 210 mL DEPC-treated water which was then cooled to ~60°C in a waterbath. Twenty-five mL of 10x running buffer (see appendix 2 for composition), 13 ml of 37% formaldehyde (v/v) and 2 µL EtBr (10mg/mL stock) were then added. The solution was poured into a 20 x 35 cm gel mold to set. (Note: higher levels of EtBr result in very high background staining and, furthermore, since EtBr migrates toward the cathode rather than the anode, a few µL were also added to the running buffer in the anode buffer tank to ensure even staining). To prepare RNA samples, 15 µg of total RNA (made up to 5 µL total volume with DEPC-treated water), were heated at 65°C for 3-5 min, and then chilled on ice. A 19 µl aliquot of sample buffer (composed of 12.5 µL formamide, 2.5 µL 10× MOPS buffer, 4 µl 37% formaldehyde) and 2.5 µl of 10× RNA loading buffer (0.9 ml 50% v/v glycerol, 2 µl 0.5 M EDTA pH 8.0, 50 µl 0.25% w/v bromophenol blue, 50 µl 0.25% w/v xylene cyanol FF) were then added. After few seconds of centrifugation, the RNA sample was loaded onto the agarose gel which had been pre-run for 10-15 min at 90 V (4.5 V/cm) with 1× MOPS as the electrophoresis buffer. A 7 µg aliquot of RNA ladder (1µg/µl, 0.24-9.5 kb, GIBCO BRL) was applied as a size marker in one lane. The gel was run at 90 V for 4-5 h, then removed, rinsed in DEPC-treated water for 10 min, 10×SSC for 20 min, and then photographed on a UV light transilluminator (Foto/UV 26, Fotodyne Inc.) at a wavelength of 300 nm with a DS-34 camera and Polaroid instant pack films (Type 667).

The RNA was then transferred onto a nylon membrane (pore size 0.2 μm, Hybond-N, Amersham) by upwards capillary transfer with 10× SSC (Sambrook *et al.*, 1989). Total RNA was fixed to the membrane as for plaque lifts. The blot was then stained with a 0.02% w/v methylene blue solution to visualize the standards and ribosomal RNA bands.

Probe synthesis

A gene-specific DNA probe was derived from the cloned insert DNA. The insert was isolated by restriction enzyme digestion of the plasmid and the restriction fragments were separated by agarose gel electrophoresis. The insert DNA was then purified from the gel with a silica gel purification kit as per manufacturer's instructions (Geneclean III kit, BIO 101) and ^{32}P-labeled using a random priming method adapted from Sambrook *et al.* (1989). Approximately 50-250 ng of insert DNA was adjusted to 10 μl with sterile ddH_2O, heated in boiling water for 3-4 min and chilled on ice. Then 1 μl of 1 mM dNTPs (without dCTP), 2 μl of 10× Reaction Buffer (ReACT 2, Gibco BRL), 1.5 μl of DNA random hexamers (1 μg/μL), 0.5 μL of Large Fragment of DNA Polymerase I (Klenow fragment, exonuclease minus, 5 U/μl, Gibco BRL), and 5 μL of [α-^{32}P]dCTP (3,000 Ci/mmol, Amersham) were added and the mixture was incubated at 37°C for 30 min. The reaction was stopped by adding 1 μl of 0.5 M EDTA (pH 8.0) and unincorporated nucleotides were removed by passing the mixture through a Sephadex G-50 spun-column. TE buffer was added to the DNA probe sample to 0.5 ml and radioactivity was measured

Northern blots were hybridized with probe (1×10^6 cpm/mL hybridization fluid) for approximately 16 hours, then washed as described for library screening to achieve a good

signal-to-noise ratio. Blots were exposed to Kodak X-OMAT film for an appropriate time period (explanation in Appendix I) and then analyzed as outlined below.

Northern blot analysis

Analysis of the developed autoradiograms was carried out using computer imaging software. A Scan Jet3C scanner (Hewlett-Packard, Palo Alto, CA) was used in conjunction with HP Deskscan II v2.2 software (Hewlett-Packard, Palo Alto, CA) to obtain a computer pixel image. If required, the scanned images were manipulated as a whole using Adobe Photoshop v. 2.0 to improve resolution. Imagequant v3.22 (Innovative Optical Systems Research) was then employed to measure relative band intensity across the lanes of the autoradiogram image. This analysis gave both final confirmation of increased message accumulation of the candidate clones, and a profile of relative transcript levels throughout the freezing or anoxia exposures in comparison with the untreated (control) state. Band intensities were normalized to their corresponding rRNA bands, and these values were compiled as ratios of transcript levels with respect to control values.

RESULTS

Construction of cDNA libraries

Message RNA from the foot muscle of freeze-treated *L. littorea* (1, 12, and 24 hours pooled) was used to successfully synthesize a cDNA library in the Uni-Zap XR vector. The initial titre of 5.5×10^5 pfu/mL was amplified to 1.5×10^9 pfu/mL with 98% recombination efficiency. The library prepared from foot muscle of anoxia-treated snails had similar characteristics, with an amplified titre of 4.7×10^{10} pfu/mL. For both libraries, the size range of cDNA inserts ranged from 0.4 to 8 Kb as determined by a mass excision and plasmid isolation procedure.

Screening of freezing and anoxic cDNA libraries

Stress inducible clones were isolated by screening the libraries with ^{32}P labeled total cDNA probe derived from polyA mRNA as described in the methods section. Probes for the stressed condition were synthesized from a mixture of equal amounts of mRNA isolated from foot muscle of 1, 12, and 24 hour anoxic or frozen tissue. The control probe was synthesized from the mRNA of animals acclimated to 5°C under aerobic conditions. The control probe used was the same for screening both libraries because the original experimental treatments involved three groups of animals (control, anoxic, frozen) that were derived from the same batch of snails.

Figure 2.1 shows a typical result from the first round of differential screening. The top plaque lift was hybridized with control probe, the lower with probe made using mRNA from freeze-exposed foot muscle. Screening of approximately 3.6×10^5 pfu from the frozen library revealed 336 potential plaques showing stronger binding of stress probe as opposed to control. These were isolated and divided into 4 pools, each of which was used for secondary screening. The second round of screening left 37 plaques with obviously stronger signals when hybridized with stress probe. These were isolated

individually and plated separately for the final round of screening, a typical result of which is shown in figure 2.2. At this point, twenty-one clones appeared to be up-regulated by the stress. These were designated $LLFF_1$-$LLFF_{21}$, where LLFF represents *Littorina littorea* frozen foot. The phage was isolated from the excised plaques and used as the substrate for *in vivo* excision. The resulting pBluescript plasmids were purified and used in a reverse northern blot analysis to further select up-regulated clones. The inserts were released from the plasmid with a double restriction enzyme digest (Xho1 and EcoR1), separated on duplicate 1% TAE agarose gels and transferred to Hybond membranes. Duplicate membranes were screened with total cDNA probes (control vs. freezing) to confirm stronger binding of probe in the stressed state. As illustrated in figure 2.3, 10 clones showed stronger binding in the frozen state, and these were further pursued with northern blot analysis, the results of which are further detailed in chapters 3 and 4. This eliminated several more clones because they either did not bind to a transcript on the membrane or were expressed at the same concentration throughout the freezing time course. The remaining clones were LLFF 3, 13, 15, 20, and 21 and these were deemed to be differentially regulated.

The same process was used to isolate differentially expressed clones from the anoxic library. After screening 3×10^5 plaques, over 200 produced stronger signals with the anoxic probe and were isolated for secondary screening. After secondary and tertiary screening, 78 clones remained. Many of these were multiple plaques on a single plate that showed the same level of up-regulation. After 3 rounds of screening, positive plaques on the same plate are frequently identical, so steps were taken to eliminate replicate clones. After *in vivo* excision, the plasmids were digested and the fragments separated on an agarose gel. Those found to contain the same size insert were considered to be identical clones. At this point, ~30 clones remained. These were also examined with a reverse northern, but all inserts bound the anoxic probe more strongly than the control so the procedure was unable to eliminate any false positives in this case. The clones were

then subjected to northern blot analysis (detailed in later chapters) to confirm stress-induced up-regulation. Finally, there were only 5 positive clones to pursue, and these were denoted as LLAF W, EE, N, 4.2.1, and 7.1.4, where LLAF represents *Littorina littorea* anoxic foot. The cDNA insert size was determined using a double restriction digest to release the insert with separation of the fragments through an agarose gel. The results of the restriction digests are shown in figure 2.4.

DNA sequences and characteristics of the foot muscle cDNA clones

All positive clones were sequenced and analyzed for key characteristics of messenger RNA. These include the presence of a polyadenylation tail (the factor that distinguishes mRNAs from rRNA and tRNA), an open reading frame (ORF), start and stop codons, and a polyadenylation signal. In addition, alignment analysis was conducted to determine if any of the clones were replicates, which were then grouped and compared to find the most complete ORF. The following paragraphs describe the characteristics of each clone or group of clones.

All positive clones isolated from the tertiary screens of both libraries were subjected to *in vivo* excision and plasmid isolations were carried out as described in the methods section. Restriction digests showed single inserts for each candidate and allowed size estimations on the basis of their migration with respect to a set of DNA standards. Using the digested inserts as templates to synthesize radiolabeled probes, northern blots containing equal amounts of total RNA from freezing and anoxia time courses, were probed to reveal the level of transcript induction. An overview of findings is presented here with a summary of clone characteristics in Table 2.1. Detailed information regarding the analyses of sequences and northern blots for individual clones are presented separately in the following chapters.

Novel clones

$LLFF_3$, $LLAF_N$, and $LLAF_{4.2.1}$

These three clones were all isolated individually and appeared to be up-regulated under both the anoxia and freezing stresses. They were isolated from independent tertiary screens and the sizes of their cDNA inserts varied greatly, which initially suggested that they might all be unique. However, after sequence analysis it became apparent that these three were all copies of the same message but truncated to varying degrees. The nucleotide sequences of the three clones are presented in figure 2.5 which also highlights the small variations in their sequences. Given the extremely high level of sequence identity among the 3 sequences, the few non-identical nucleotides that were found were attributed to errors in the sequencing protocol. A nucleotide search in the NCBI Genbank database elicited only low-scoring matches, hence the DNA sequence did not allow for identification.

All three clones showed a polyadenylation signal (AATAAA) in the 3' region, a polyA tail, and stop codons. $LLAF_N$ and $_{4.2.1}$ were lacking the 5' region, so only $LLFF_3$ contained a start codon and 5'UTR. These characteristics are shown for the $LLFF_3$ sequence in figure 2.6A. Figure 2.6B illustrates the three reading frames in the direction of translation, 5'-3'. In this diagram, the clone sequence is represented with a line, while the perpendicular tick marks above the line indicate a start codon, those below the line, a stop codon. Possible ORFs are indicated with an arrow. As illustrated, each reading frame shows an ORF, but only the first is of any significant size, representing the translation of 300 nucleotides.

$LLFF_{13}$ and $LLAF_{EE}$, $LLAF_{7.1.4}$

$LLFF_{13}$ was isolated from the freezing library, whereas $LLAF_{EE}$ and $LLAF_{7.1.4}$ were from the anoxic library. Although the sequences of the overlapping portions of $LLFF_{13}$ and $LLAF_{EE}$ were identical, $LLFF_{13}$ (at 324 bp in length) lacked the roughly 877

bases present in the 5'region of $LLAF_{EE}$ (figure 2.7). $LLAF_{7.1.4}$ was also truncated as compared with $LLAF_{EE}$ and, as such, both it and $LLFF_{13}$ were lacking a start codon, 5'UTR and a complete open reading frame (ORF). $LLAF_{EE}$ possessed a start codon at 38nt and an apparently complete ORF of 968 bases (figure 2.8). In addition, all three clones contained a polyadenylation signal near the terminus of the 3'UTR, a mere 19 bases from the start of the polyA tail.

A DNA homology search in Genbank produced no matches for this group of clones. Although these were not the only unidentifiable clones that were isolated from the *L. littorea* libraries, they were the most unique of the isolated genes. A quick glance at any of these three sequences shows a remarkable lack of thymidine residues. $LLAF_{EE}$, for example, contained only 11% T, whereas C, A, and G residues each accounted for roughly 30% of the total bases. In addition, the bases C, A and G were arranged in multiple repeating patterns (highlighted in figure 2.8). Finally, this group of clones, with such a unique sequence, showed the most dramatic increase in probe binding from the control to stressed state.

$LLAF_W$

At only 878 bases and with stop codons present throughout the last 500 nt, this gene also appeared to be truncated, lacking a start codon and 5'UTR (figure 2.9). $LLAF_W$ has two putative polyadenylation signals and a polyadenylated 3' end containing 22 adenosines. DNA sequence similarity search did not show a match with any known DNA sequences.

Identified clones

Clones $LLFF_{15}$ and $LLFF_{21}$ do not share similar sequences, but each was isolated solely from the frozen library. $LLFF_{15}$ was 1551 nt in length with a polyA signal at 1115 bp, followed by a 22 base polyA tail. Although this was one of the longest clones isolated, it lacked both a 5'UTR and a complete ORF. In addition, the first stop codon

(TGA) was located only 382 nt from the 5' end, after which the rest of the sequence was riddled with stop codons in all reading frames, suggesting that the sequence primarily composed of the 3'UTR of a much larger gene. A BLAST homology search through Genbank supported this observation because nucleotides 1-380 showed strong sequence identity with the last 300 bases of the gene for myosin. Figure 2.10 illustrates the complete cDNA sequence and characteristics common to message RNAs.

The truncated reading frame possesses multiple frameshifts which are due to sequencing errors. Figure 2.11 illustrates the 3 reading frames coded for by $LLFF_{15}$ and highlights those sections that align with a known myosin gene from another molluscan species, the scallop *Argopecten irradians.* As demonstrated in figure 2.11b, the reading frame of $LLFF_{15}$ encodes only a tiny portion of the C-terminal end of the myosin protein.

Finally, figure 2.12 illustrates the transcript level of $LLFF_{15}$ in foot muscle in response to freezing. Over the freezing time course, expression increased gradually to a maximum of 2-fold greater than control after 24 hours exposure. The largest increase however was seen during recovery when, after 24 hours, $LLFF_{15}$ expression increased to more than 4-fold that of control. Anoxia induced a more moderate increase of only 2-fold, with no further increase during recovery. The probe for this clone did not bind to hepatopancreas total RNA, a result that will be addressed in the discussion.

$LLFF_{21}$ was 2023 nt in length. It was sequenced from both the 5' and 3' ends towards the centre of the clone. The results from the two sequences did not overlap and could not be assembled into a single cDNA, indicating that the sequence was missing several nucleotides in the centre of the sequence. A DNA sequence homology search showed that the gene was definitely that of beta-actin, and in light of this, I did not consider it necessary to pursue further sequencing and obtain the complete DNA sequence.

The two sections of DNA sequence are presented in figure 2.13. Figure (A) shows the 5' section of the sequence, which, although lacking a start codon, contained a large uninterrupted coding region (from base 1 to 980 nt) where the stop codon was present. The remainder of this sequence and the entire region shown in figure 2.13 (B) represents the 3' section of the sequence. The polyadenylation signal was located at base 1981 of sequence (B), followed 17 bases later by the polyA tail.

The translated coding region of $LLFF_{21}$ is shown in figure 2.14(A), while figure (B) shows the alignment of the *L. littorea* actin with those of other species. The alignment of each translated reading frame with known actin revealed several frame-shifts. Where a shift occurred, the different reading frames were placed on different lines.

The change in transcript content in foot muscle for clone $LLFF_{21}$ is illustrated in figure 2.15 . Anoxia had essentially no effect on actin transcript, as illustrated in figure 2.15(D), where the bar graph for each time point hovered around control levels. Freezing however, induced a 3-fold increase after 12 hours of freezing, and remained at this level until 24 hours of recovery.

The expression pattern of $LLFF_{21}$ was very different in the hepatopancreas. $LLFF_{21}$ gradually increased over both the anoxia and freezing time courses to a maximum value at 12 hours of stress exposure (figure 2.16). The maximum increase of 4X that of the control level was attained after 12 hours of freezing, and remained stable until one hour of recovery, after which it dropped to control values. Anoxia induced a similar pattern, but the peak increase was attained after just 5 hours of anoxia and was only maintained through 12-24 hours. Transcript levels then gradually declined to control levels for the remainder of the stress and recovery period.

$LLFF_{20}$

The clone $LLFF_{20}$ was 688 bases long, with a polyadenylation signal at bases 667-673. The stop codon overlaps with the polyadenylation signal, and poly A tail in the

following fashion, T**AATTAAA**(18A) which is a characteristic of mitochondrial transcripts. Although lacking a portion of the 5'region (including a start codon) a Genbank BLAST search indicated that the clone had a strong similarity with the gene for cytochrome c oxidase, subunit II (COXII), a mitochondrially encoded protein. The sequence is illustrated in figure 2.17, along with indications of the stop codon, polyA signal and coding region. The translated reading frame, using the invertebrate mitochondrial code, is shown in figure 2.17(B) and aligned with COXII protein sequences from other sources. The high degree of similarity in shown in the matrix table in figure 2.17 (C).

Table 2.1. Characteristics of up-regulated clones isolated through differential screening of cDNA libraries synthesized from total message RNA from the foot muscle of freezing or anoxia treated *Littorina littorea*. Each of the clones contained a stop codon, 3'UTR, polyadenylation signal (AATAAA) and a polyA tail at the 3' end of the sequence. Some also showed a start codon and a 5'UTR whereas others were incomplete.

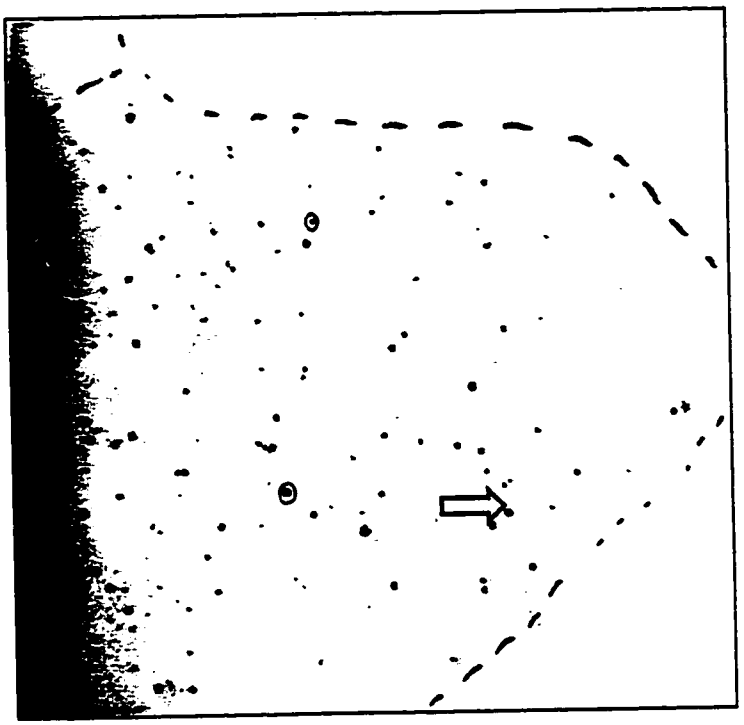
Clone	Isolated from: Frozen library	Anoxic library	Size nt	5'UTR	Start codon	Number of bases translated	3'UTR character for complete or near complete clones Size	% of total nt	Identity
$LLFF_3$	*	*	1205	Yes	Yes	300	829	62%	Putative metallothionein
$LLFF_{13}$	*	*	324	No	No		N/A		Novel gene
$LLFF_{15}$	*		1651	No	No	365 partial ORF	N/A		Myosin heavy chain
$LLFF_{20}$	*		688	No	No		N/A		Cytochrome oxidase
$LLFF_{21}$	*		2023	No	No		~1021	51.4%	Actin
$LLAF_W$	*		792	Yes	Yes		592	74.5%	Novel gene
$LLAF_{EE}$	*	*	1092	Yes	Yes		183	15.4%	Novel gene
$LLAF_{7.1.4}$	*		645	No	No	453	N/A		Novel gene
$LLAF_{4.2.1}$	*	*	825	No	No				Putative metallothionein
$LLAF_N$	*	*	572	No	No		N/A		Putative metallothionein

Figure 2.1 **Results of a typical primary differential screening of the cDNA library made from foot muscle of freeze-exposed *L. littorea*.**

(A) Plaque lift was hybridized with total cDNA probe synthesized from mRNA isolated from control foot muscle.

(B) Duplicate plaque lift hybridized with cDNA probe synthesized from equal amounts of mRNA isolated from 1, 12 and 24 hour freeze-treated animals.

(A)



(B)

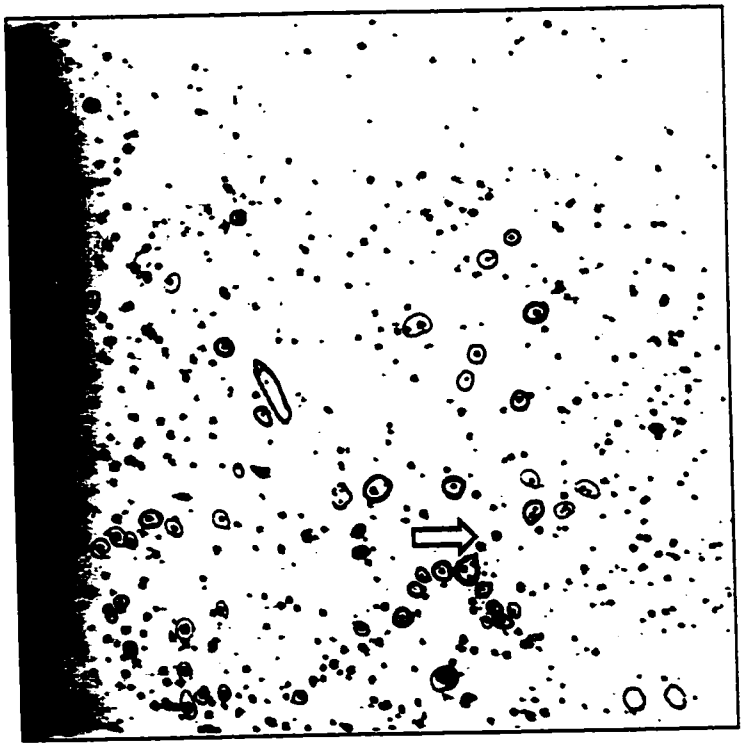


Figure 2.2: A typical tertiary screening result from the freezing library after two rounds of differential screening.

(A) and (B) show duplicate plaque lifts, created from the same plate of plaques. Figure (A) illustrates the lift hybridized with control probe, whereas (B) was hybridized with probe made from mRNA of freeze-treated animals (as described in figure 2.1).

(A)

(B)

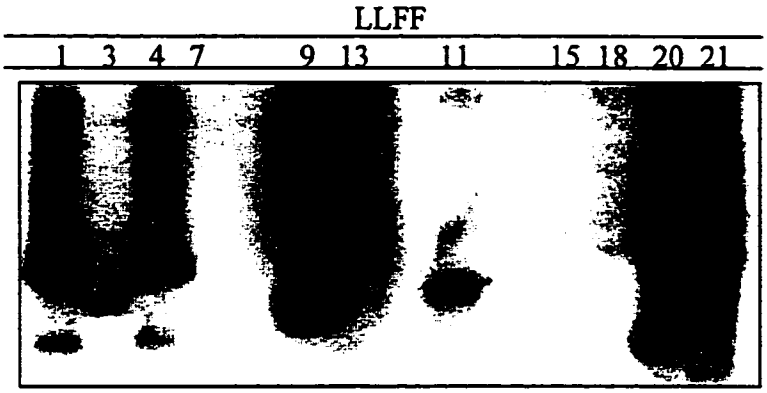
Figure 2.3 **Reverse northern analysis of clones isolated from the cDNA library created from freeze-treated *L. littorea* foot muscle.** Plasmids were digested with restriction enzymes (EcoR1 and Xho1) to isolate the inserts. Digestion products were seperated on duplicate 1% TAE agarose gels and transferred to a nylon membranes. One membrane was hybridized with ^{32}P-labelled total cDNA probe made from control animals, the other with stress probe made from mRNA isolated from animals treated with 1, 12 and 24 hours freezing.

(A) Membrane hybridized with control probe

(B) Duplicate membrane hybridized with freezing probe. Arrows indicate the clones that were deemed up-regulated.

(C) Restriction digests of up-regulated clones. The largest band in each lane (~3Kb) is the pBluescript plasmid, indicated with an arrow. Molecular weight standards (1Kb ladder from Gibco BRL) are indicated to the left of the gel photo. $LLFF_{15}$ was the only clone of this group that showed 3 bands after digestion. In light of the sequencing data, which shows the insert size to be ~1.5Kb, the band at ~2.2Kb results from single-stranded, uncut plasmid (which includes the $LLFF_{15}$ insert).

(A)



(B)

(C)

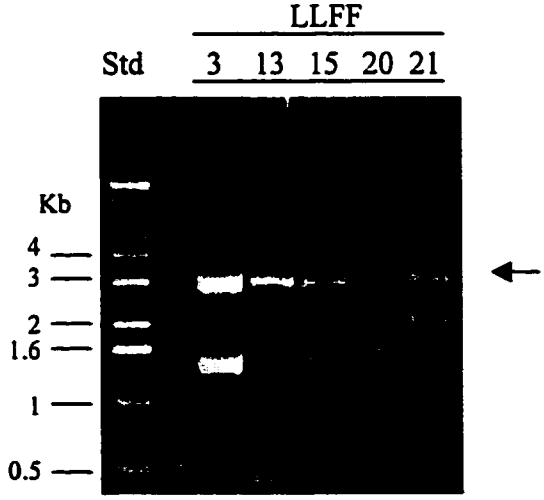


Figure 2.4: Results of restriction digests for clones isolated from the cDNA library made from mRNA of 1, 12, and 24 hour anoxia-treated animals.

Plasmids were digested with Xho1 and EcoR1 and the fragments separated on a 1% TAE agarose gel. Sizes of the DNA standards and plasmid fragments are indicated

Std N W
3
2
1.6
1
0.5
Std 4.2.1 7.1.4
3
2
1.6
1
0.5
Std EE
3
2
1.6
1

Figure 2.5: **Alignment of replicate cDNA clones $LLFF_3$, $LLAF_{4.2.1}$ and $LLAF_N$.** Clone $LLFF_3$ is the longest in this group, hence the first 5 lines of sequence belong to it alone. There are only minimal variations among the nucleotide sequences and these are highlighted in black. The residues of the sections that align are 96% identical.

LEGEND:

Base differences white text

Gaps in sequence ....

Identical bases ---

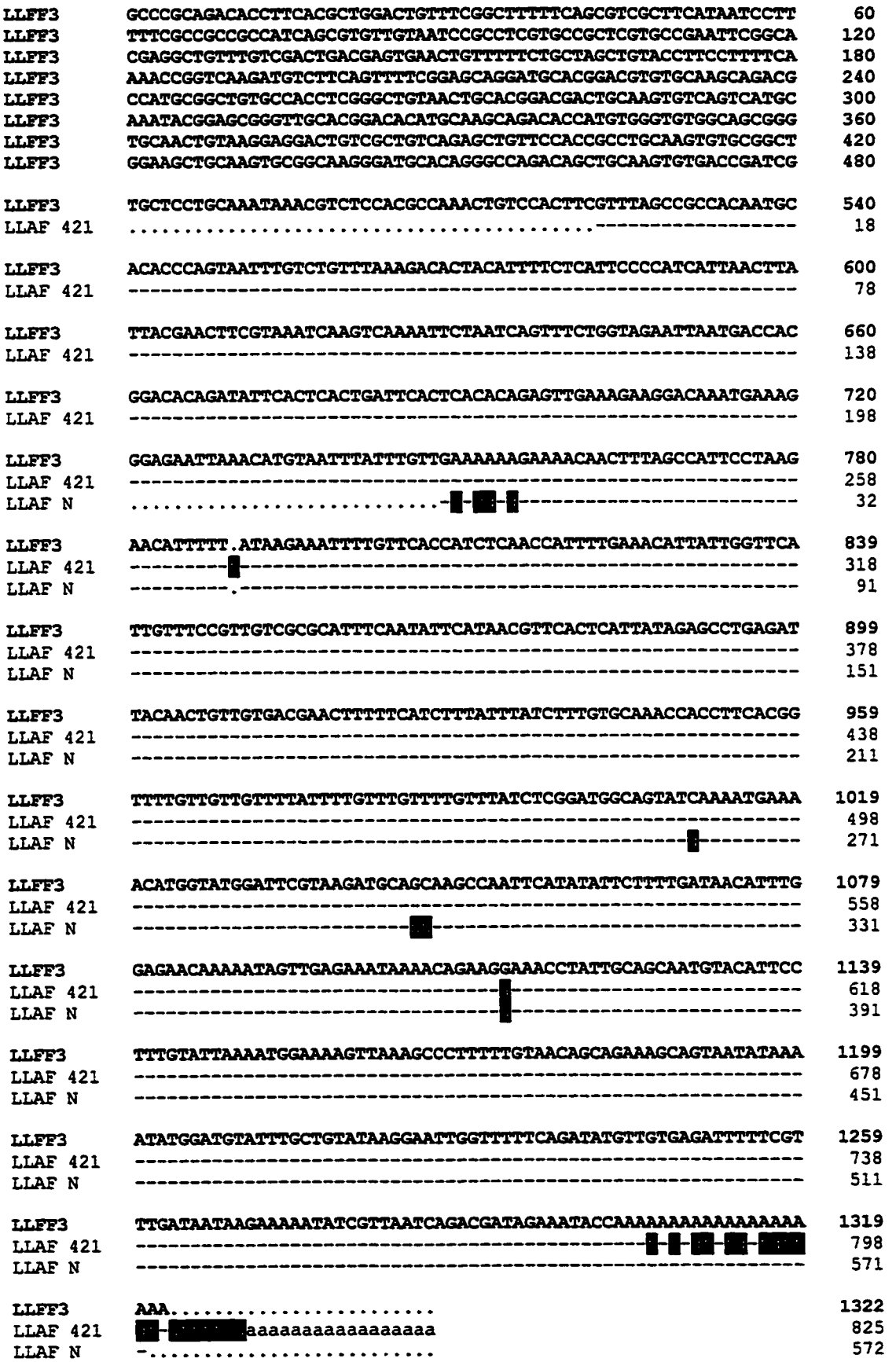

LLFF3 GCCCGCAGACACCTTCACGCTGGACTGTTTCGGCTTTTTCAGCGTCGCTTCATAATCCTT 60
LLFF3 TTTCGCCGCCGCCATCAGCGTGTTGTAATCCGCCTCGTGCCGCTCGTGCCGAATTCGGCA 120
LLFF3 CGAGGCTGTTTGTCGACTGACGAGTGAACTGTTTTTCTGCTAGCTGTACCTTCCTTTTCA 180
LLFF3 AAACCGGTCAAGATGTCTTCAGTTTTCGGAGCAGGATGCACGGACGTGTGCAAGCAGACG 240
LLFF3 CCATGCGGCTGTGCCACCTCGGGCTGTAACTGCACGGACGACTGCAAGTGTCAGTCATGC 300
LLFF3 AAATACGGAGCGGGTTGCACGGACACATGCAAGCAGACACCATGTGGGTGTGGCAGCGGG 360
LLFF3 TGCAACTGTAAGGAGGACTGTCGCTGTCAGAGCTGTTCCACCGCCTGCAAGTGTGCGGCT 420
LLFF3 GGAAGCTGCAAGTGCGGCAAGGGATGCACAGGGCCAGACAGCTGCAAGTGTGACCGATCG 480
LLFF3 TGCTCCTGCAAATAAACGTCTCCACGCCAAACTGTCCACTTCGTTTAGCCGCCACAATGC 540
LLAF 421 18
LLFF3 ACACCCAGTAATTTGTCTGTTTAAAGACACTACATTTTCTCATTCCCCATCATTAACTTA 600
LLAF 421 78
LLFF3 TTACGAACTTCGTAAATCAAGTCAAAATTCTAATCAGTTTCTGGTAGAATTAATGACCAC 660
LLAF 421 138
LLFF3 GGACACAGATATTCACTCACTGATTCACTCACACAGAGTTGAAAGAAGGACAAATGAAAG 720
LLAF 421 198
LLFF3 GGAGAATTAAACATGTAATTTATTTGTTGAAAAAAGAAAACAACTTTAGCCATTCCTAAG 780
LLAF 421 258
LLAF N 32
LLFF3 AACATTTTT.ATAAGAAATTTTGTTCACCATCTCAACCATTTTGAAACATTATTGGTTCA 839
LLAF 421 318
LLAF N 91
LLFF3 TTGTTTCCGTTGTCGCGCATTTCAATATTCATAACGTTCACTCATTATAGAGCCTGAGAT 899
LLAF 421 378
LLAF N 151
LLFF3 TACAACTGTTGTGACGAACTTTTTCATCTTTATTTATCTTTGTGCAAACCACCTTCACGG 959
LLAF 421 438
LLAF N 211
LLFF3 TTTTGTTGTTGTTTTATTTTGTTTGTTTTGTTTATCTCGGATGGCAGTATCAAAATGAAA 1019
LLAF 421 498
LLAF N 271
LLFF3 ACATGGTATGGATTCGTAAGATGCAGCAAGCCAATTCATATATTCTTTTGATAACATTTG 1079
LLAF 421 558
LLAF N 331
LLFF3 GAGAACAAAAATAGTTGAGAAATAAAACAGAAGGAAACCTATTGCAGCAATGTACATTCC 1139
LLAF 421 618
LLAF N 391
LLFF3 TTTGTATTAAAATGGAAAAGTTAAAGCCCTTTTTGTAACAGCAGAAAGCAGTAATATAAA 1199
LLAF 421 678
LLAF N 451
LLFF3 ATATGGATGTATTTGCTGTATAAGGAATTGGTTTTTCAGATATGTTGTGAGATTTTTCGT 1259
LLAF 421 738
LLAF N 511
LLFF3 TTGATAATAAGAAAAATATCGTTAATCAGACGATAGAAATACCAAAAAAAAAAAAAAAAA 1319
LLAF 421 798
LLAF N 571
LLFF3 AAA..... 1322
LLAF 421 aaaaaaaaaaaaaaaa 825
LLAF N 572

Figure 2.6: Nucleotide sequence of the cDNA clone $LLFF_3$.

(A) The clone appeared to be a complete message as it contains all the hallmarks of mRNAs:

LEGEND

start codon	ATG
putative ORF	plain black text
stop codon	box
polyA signal	**black, bold type**
3' and 5' UTR	*italics*

(B) This figure illustrates the 3 possible 5'-3' open reading frames. Each line represents the length of the clone; the tick marks above the line represent start codons (ATG), while those below show possible stop codons. The thick black arrows indicate a complete ORF with both start and stop codons. The longest ORF runs from bases 193 – 496.

(A)

```
1     GCCCGCAGACACCTTCACGCTGGACTGTTTCGGCTTTTTCAGCGTCGCTT     50
51    CATAATCCTTTTTCGCCGCCGCCATCAGCGTGTTGTAATCCGCCTCGTGC     100
101   CGCTCGTGCCGAATTCGGCACGAGGCTGTTTGTCGACTGACGAGTGAACT     150
151   GTTTTTCTGCTAGCTGTACCTTCCTTTTCAAAACCGGTCAAG[redacted]TCTTC     200
201   AGTTTTCGGAGCAGGATGCACGGACGTGTGCAAGCAGACGCCATGCGGCT     250
251   GTGCCACCTCGGGCTGTAACTGCACGGACGACTGCAAGTGTCAGTCATGC     300
301   AAATACGGAGCGGGTTGCACGGACACATGCAAGCAGACACCATGTGGGTG     350
351   TGGCAGCGGGTGCAACTGTAAGGAGGACTGTCGCTGTCAGAGCTGTTCCA     400
401   CCGCCTGCAAGTGTGCGGCTGGAAGCTGCAAGTGCGGCAAGGGATGCACA     450
451   GGGCCAGACAGCTGCAAGTGTGACCGATCGTGCTCCTGCAAATAAACGTC     500
501   TCCACGCCAAACTGTCCACTTCGTTTAGCCGCCACAATGCACACCCAGTA     550
551   ATTTGTCTGTTTAAAGACACTACATTTTCTCATTCCCCATCATTAACTTA     600
601   TTACGAACTTCGTAAATCAAGTCAAAATTCTAATCAGTTTCTGGTAGAAT     650
651   TAATGACCACGGACACAGATATTCACTCACTGATTCACTCACACAGAGTT     700
701   GAAAGAAGGACAAATGAAAGGGAGAATTAAACATGTAATTTATTTGTTGA     750
751   AAAAAGAAAACAACTTTAGCCATTCCTAAGAACATTTTTATAAGAAATTT     800
801   TGTTCACCATCTCAACCATTTTGAAACATTATTGGTTCATTGTTTCCGTT     850
851   GTCGCGCATTTCAATATTCATAACGTTCACTCATTATAGAGCCTGAGATT     900
901   ACAACTGTTGTGACGAACTTTTTCATCTTTATTTATCTTTGTGCAAACCA     950
951   CCTTCACGGTTTTGTTGTTGTTTTATTTTGTTTGTTTTGTTTATCTCGGA     1000
1001  TGGCAGTATCAAAATGAAAACATGGTATGGATTCGTAAGATGCAGCAAGC     1050
1051  CAATTCATATATTCTTTTGATAACATTTGGAGAACAAAAATAGTTGAGAA     1100
1101  ATAAAACAGAAGGAAACCTATTGCAGCAATGTACATTCCTTTGTATTAAA     1150
1151  ATGGAAAAGTTAAAGCCCTTTTTGTAACAGCAGAAAGCAGTAATATAAAA     1200
1201  TATGGATGTATTTGCTGTATAAGGAATTGGTTTTTCAGATATGTTGTGAG     1250
1251  ATTTTTCGTTTGATAATAAGAAAAATATCGTTAATCAGACGATAGAAATA     1300
1301  CCAAAAAAAAAAAAAAAAAA                                   1321
```

(B)

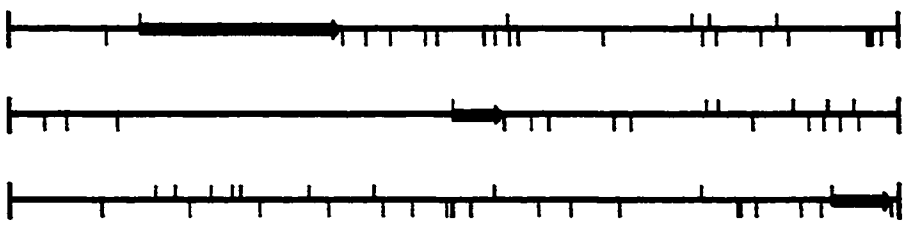


Figure 2.7: Alignment of DNA sequences for replicate clones $LLAF_{EE}$, $LLFF_{13}$, and $LLAF_{7.1.4}$.

Clone $LLAF_{EE}$ is the longer of the pair and hence composes the first 9 lines of sequence. There are only minimal variations among the nucleotide sequences and these are highlighted in black.

LEGEND:

Base differences white text

Identical bases -----

Gaps in sequence ....

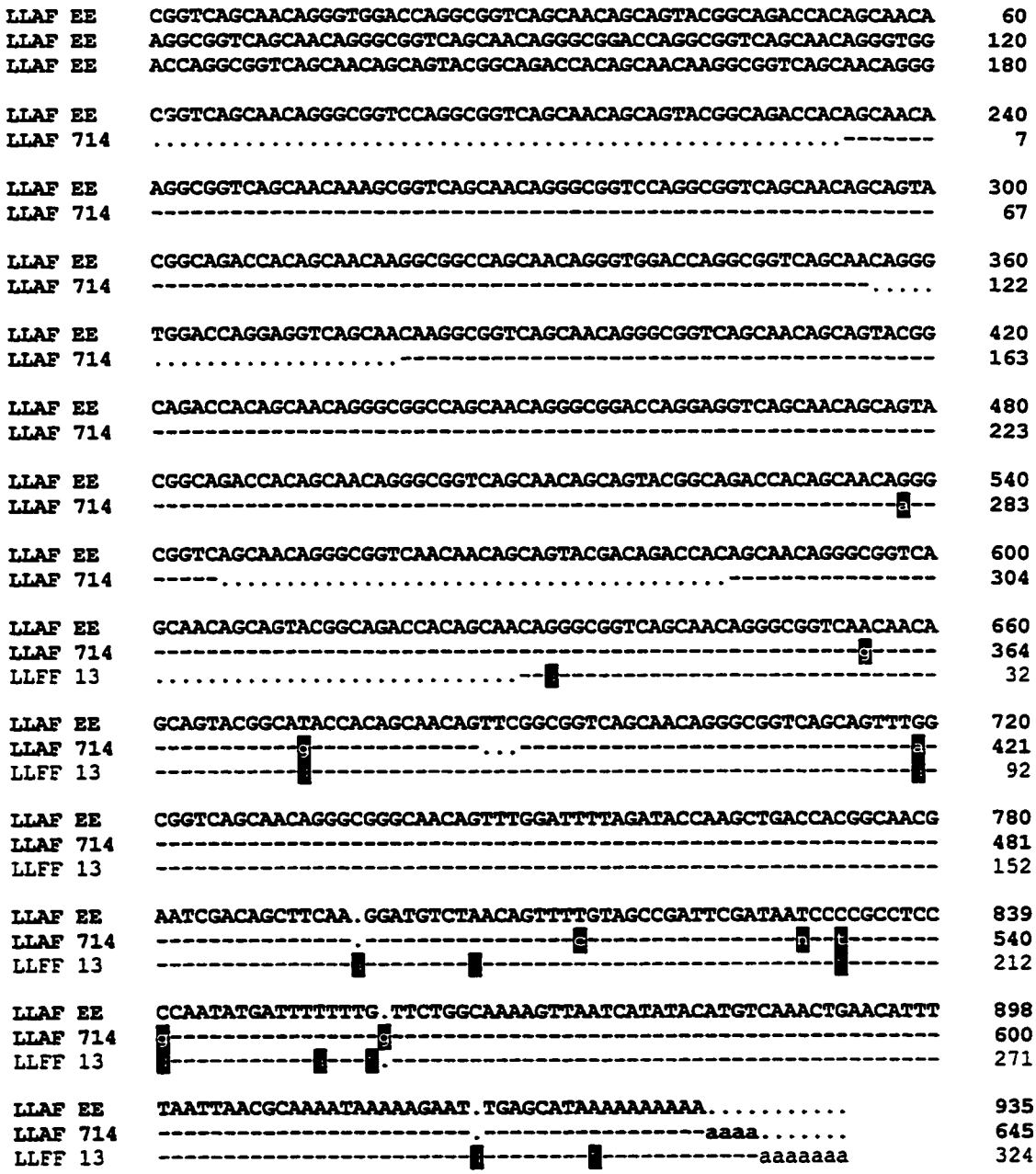

LLAF EE CGGTCAGCAACAGGGTGGACCAGGCGGTCAGCAACAGCAGTACGGCAGACCACAGCAACA 60
LLAF EE AGGCGGTCAGCAACAGGGCGGTCAGCAACAGGGCGGACCAGGCGGTCAGCAACAGGGTGG 120
LLAF EE ACCAGGCGGTCAGCAACAGCAGTACGGCAGACCACAGCAACAAGGCGGTCAGCAACAGGG 180
LLAF EE CGGTCAGCAACAGGGCGGTCCAGGCGGTCAGCAACAGCAGTACGGCAGACCACAGCAACA 240
LLAF 714 ....------- 7
LLAF EE AGGCGGTCAGCAACAAAGCGGTCAGCAACAGGGCGGTCCAGGCGGTCAGCAACAGCAGTA 300
LLAF 714 ---------- 67
LLAF EE CGGCAGACCACAGCAACAAGGCGGCCAGCAACAGGGTGGACCAGGCGGTCAGCAACAGGG 360
LLAF 714 ----------..... 122
LLAF EE TGGACCAGGAGGTCAGCAACAAGGCGGTCAGCAACAGGGCGGTCAGCAACAGCAGTACGG 420
LLAF 714 .....---------- 163
LLAF EE CAGACCACAGCAACAGGGCGGCCAGCAACAGGGCGGACCAGGAGGTCAGCAACAGCAGTA 480
LLAF 714 ---------- 223
LLAF EE CGGCAGACCACAGCAACAGGGCGGTCAGCAACAGCAGTACGGCAGACCACAGCAACAGGG 540
LLAF 714 ----------a-- 283
LLAF EE CGGTCAGCAACAGGGCGGTCAACAACAGCAGTACGACAGACCACAGCAACAGGGCGGTCA 600
LLAF 714 -----.....---------- 304
LLAF EE GCAACAGCAGTACGGCAGACCACAGCAACAGGGCGGTCAGCAACAGGGCGGTCAACAACA 660
LLAF 714 ----------g----- 364
LLFF 13 .....-- 32
LLAF EE GCAGTACGGCATACCACAGCAACAGTTCGGCGGTCAGCAACAGGGCGGTCAGCAGTTTGG 720
LLAF 714 ----------g----------...----------a- 421
LLFF 13 ----------.----------.- 92
LLAF EE CGGTCAGCAACAGGGCGGGCAACAGTTTGGATTTTAGATACCAAGCTGACCACGGCAACG 780
LLAF 714 ---------- 481
LLFF 13 ---------- 152
LLAF EE AATCGACAGCTTCAA.GGATGTCTAACAGTTTTGTAGCCGATTCGATAATCCCCGCCTCC 839
LLAF 714 ----------.----------c----------n--t------- 540
LLFF 13 ----------.-------.----------.------- 212
LLAF EE CCAATATGATTTTTTTG.TTCTGGCAAAAGTTAATCATATACATGTCAAACTGAACATTT 898
LLAF 714 g----------g---------- 600
LLFF 13 .----------.--.---------- 271
LLAF EE TAATTAACGCAAAATAAAAAGAAT.TGAGCATAAAAAAAAAA..... 935
LLAF 714 ----------.----------aaaa..... 645
LLFF 13 ----------.-------.----------aaaaaaa 324

Figure 2.8: Nucleotide sequence of the cDNA clone $LLAF_{EE}$.

(A) DNA sequence with highlighted mRNA characteristics of clone $LLAF_{EE}$ repeating nucleotide sequences (CAGCAACAG and **CAGCAACAA**) are highlighted.

LEGEND:

start codon	ATG
putative ORF	plain black text
stop codon	box
polyA signal	**black, bold type**
3' and 5' UTR	*italics*

(B) This figure illustrates the 3 possible 5'-3' open reading frames. Each line represents the length of the clone with the tick marks illustrating the presence of start codons (above the line) and stop codons (blow the line). The thick black arrows indicate a complete ORF with both start and stop codons. The sequence presents a single ORF, in reading frame 2, (38-1006 nt).

1 CAACATCGTCTTGCCTAGGTTAAACAAACCTGCCAACATGTGGAAGATCG

51 TGACCTTGACCCTTCTCGCAACTCTAGCAGCTGCTCAGTTCGGCGGTCAG

101 CAGGAAGGCGGTCAACAGCAGTACGGCAGACCA[redacted]GGCGGT**CA**

151 **GCAACAA**GGCGGT[redacted]GGCGGTCAACAGCAGTACGGCAGACCA**C**

201 **AGCAACAA**GGCGGT[redacted]GGCGGT[redacted]GGCGGACCAGGC

251 GGT[redacted]GGTGGACCAGGCGGT[redacted]CAGTACGGCAGACC

301 A**CAGCAACAA**GGCGGT[redacted]GGCGGT[redacted]GGCGGACCAG

351 GCGGT[redacted]GGTGGACCAGGCGGT[redacted]CAGTACGGCAGA

401 CCA**CAGCAACAA**GGCGGT[redacted]GGCGGT[redacted]GGCGGTCC

451 AGGCGGT[redacted]CAGTACGGCAGACCA**CAGCAACAA**GGCGGT**CAGC**

501 **AACAA**AGCGGT[redacted]GGCGGTCCAGGCGGT[redacted]CAGTAC

551 GGCAGACCA**CAGCAACAA**GGCGGC[redacted]GGTGGACCAGGCGGT[redacted]

601 [redacted]GGTGGACCAGGAGGT**CAGCAACAA**GGCGGT[redacted]GGCG

651 GT[redacted]CAGTACGGCAGACCA[redacted]GGCGGC[redacted]

701 GGCGGACCAGGAGGT[redacted]CAGTACGGCAGACCA[redacted]GG

751 CGGT[redacted]CAGTACGGCAGACCA[redacted]GGCGGT[redacted]

801 [redacted]GGCGGTCAACAACAGCAGTACGACAGACCA[redacted]GGCGGT[redacted]

851 [redacted]CAGTACGGCAGACCA[redacted]GGCGGT[redacted]GGCGG

901 TCAACAACAGCAGTACGGCATACCA[redacted]TTCGGCGGTCAGCAAC

951 AGGGCGGTCAGCAGTTTGGCGGT[redacted]GGCGGGCAACAGTTTGGA

1001 TTT**TAG***ATACCAAGCTGACCACGGCAACGAATCGACAGCTTCAAGGATGT*

1051 CTAACAGTTTTGTAGCCGATTCGATAATCCCCGCCTCCCCAATATGATTT

1101 TTTTGTTCTGGCAAAAGTTAATCATATACATGTCAAACTGAACATTTTAA

*1151 TTAACGCAA**AATAAA**AAGAATTGAGCAT**AAAAAA***

(B)

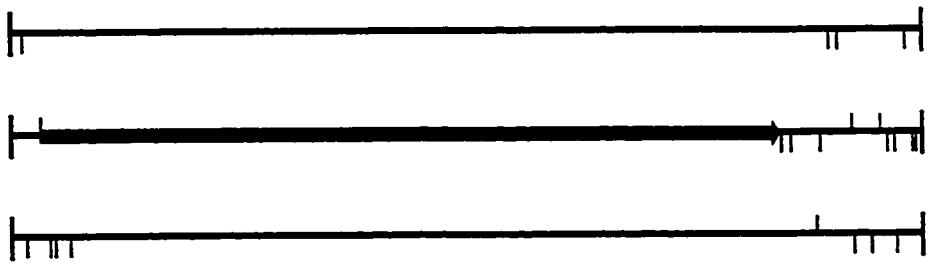


Figure 2.9: DNA sequence of clone $LLAF_W$.

LEGEND

putative coding region	plain black text
stop codon	box
polyA signal	**bold plain**
3' UTR	italics

1 GCGGCAGTGGGGGCAGTGGGAAAGCGTGACCTGTGTGCTGACCTGCTGACCCCTGACCAG

61 GTGCAACAACTAGCCACTGACCTACAGAACACCTGTGCCAACATCCCGATCAACCTGGCC

121 GCAGATGGTCTTGACTTTGATGAGGTCAAGGGCATTTTCACGGACAGCGATGTTGATGGA

181 GATGGAAAACTGTTGGCCACAGAGTTGGCAGACTTTTCTGCTGCAATCCAAACAACAGAA

241 GCCTGTGCGGAGGTGGAGGCCGAGATTCAGATATAAACCTGATATGATCATTTACTTGAC

301 GACACTGACAGAGAGCGAGATAGAATAGATCCATTCTTTTTTTCTTCTTCAAAATAACGT

361 AATTCAAATTAAGTAAATAGGTGCACGAACAAAAAATTCTTCCAATACAACTGGGCCCTC

421 ATGAAAAAAGATGATAAATACTAAGAAGAAAAATGAAAGACCGTGTCACATCTGTAAGAA

*481 **AAAATAAA**ACAAACAAAGGTATTTCGTTGAAATATATCAAAAGAAAAGGATTAAAATAAG*

541 AAGTGCATCATTTACTAAAAGAAGAACACAATAACAAAAATGTCCATTCAGTTAAAGCGC

601 AAGAGATCAACGTATGCACGCTTCTTTTGTTGTTGTTGCTTAAGACAACTTTTATTCAAT

661 CGTTTCCATTTATTTTTCAATTTATTAAAGAGACAAAAAGCAGCAATTGGTCATGGTGTG

721 AATTGTGTGTGATGAGAATCGCTGATGTTTTATTGTTGTTCTTCGCTACTTGATAATGTG

781 TCGCCCAAAGTCTATCAGCTTCTTTGAGAGTAATGGTTGTGTAAAATAAAAAGTGTCTAA

*841 **AAAATCAAAAAAAAAAAAAAAAAAAAA***

Figure 2.10: DNA sequence of clone $LLFF_{15}$.

The 5' region of this clone shows high similarity with known genes for myosin and consequently it was possible to determine that the clone is missing the 5'UTR and start codon. The remaining features of a mRNA are present, including:

LEGEND:	
stop codon	box
polyA signal	**bold type**
3' UTR	italics
coding region	plain black text

1 CGGTTATGGG TGATGAGTGC GCAGGRACAG GAACAGTNCA TGCAGGTGGA GAAACAGCGC

61 AAGGGCCTGG AACAGAAGGA CAAGGACATG CAGGACAGAC TGGACGAGGG CGAGGCTCAG

121 GCACTCAAGG GAGGCAAGAA GARCATCCAG AAACWCGAAC AGAGAGTGCG CGAGDCGAGA

181 TGGAGCTTGC CAGCGAGCAG CGCCGACACG GGGAGACCCA GAAGAACATG CGCAAGGCCG

241 ACAGACGCGT CAAGGAGCWG GCCTGACAGA GCGACGAGGA CCGCAAGAAC CAGGAGCGTC

301 TACAGGAGCT CGTGGACAAG CTGCAGGACA AGARCAAGAG CTACAAGAGA CAAGTMGAGG

361 AGGCCGAGGA GACCGCCGCC GTGAACCTGG CTAAGTACCG CAAGTGCAGA CCGAGCTGGA

421 AGATGCCGAG GAGCGCGCCG ACACCGCCGA AGGAACCCTG CAGAAACTCC GCGCCAAGAA

481 CCGAWCCTCC GTCTCTGTCC AGCGTACCTC CGCTAGCCCC GCGCCCGGAC TGTCGGCAGC

541 CCGAGCCATG CGTGGTTCCT CTGTGACTGA CTAATTGCTC ACCCGCCTAT TACAACCTGC

601 GCGGCTCGAT GACAGGACTG AACACTTCCT TCCGTTCAGC CACGCCGTCC AATGAGGGGG

661 ACTTTGACGA CTACTGATAC ACCCACTTCC GTCACTGACC AGCCAATAGC AATGCATGTT

721 ACAACCGCCA TTAGCAACTT CACCAAGGGG ACAGTACTCG AAGGACAGGA ACTTGTACAC

781 CATGAAAACC ACTGTGCTGA TTTTTTCCTC TCGATTTGTT TTTCTTTCTG CGTGTGGATG

841 GATGGACCGA ATTGAGGAAT GACGAAGACT TGTACGAGGA CTATGAWAAW ATAATTATAT

901 CTTATTACGG ACGTAATAGG ACGTGACATT GATTCTATTA GCTGTTGTTG GTGACCTTTC

961 AGTGTCCGTT TTACTTTCTG TATATTTTAT CGATGTGTAC TCTGCGGAGC AGGCCGTTGC

1021 TGGGAGCCGT ATTGTGAGAT GCAATGTATG TGATACAATT ATACTCTTGT CCDTTTTACA

1081 AGATTATTTA GTGCAATTTT CAAGGGCGGG TCGAAGAGGA GATGCAACGG AATGAAAAAG

1141 TTGATTTGCG GAATTCTCCG GAACGAGATT GAGATGGTTA ACGACTAGAC GGAATCCAGT

1201 TATATGGTGT AGCTGCTAAT GTAGACGAGT TTGACTTCTT CGWTTTTGCW AGAWAWWTAT

1261 TACGCACGTT TGACWTCTTT AATTGAATTR ATTWTCAATT TTTCGTGTGA TTTTTTACCC

1321 TATCGTGGGT GCTCATGTAC AGACCATCAT TAGCAGTTCC GCAGTGACGA GGAACAGCCC

1381 CTCTCATTGG ACGACACAGT TGAAACAGAC ATGGCTGTTT TGATAACAGA TACTTGCCTA

1441 TTTCTTCTGT GCTTCTATAA TTTCTGTCTT TTTTGCAAAA AAAATATTTA ATGATTTCAA

1501 TCTCATACCA GTGAAATAAA GAGCTGCCAA GAAAAAAAAA AAAAAAAAAA A

Figure 2.11 Alignment of $LLFF_{15}$ with the amino acid sequence of myosin heavy chain from the scallop, *Argopecten irradians*.

(A) Sequence anomalies induced multiple frameshifts throughout the sequence, therefore, $LLFF_{15}$ amino acids that were identical to the scallop myosin were highlighted with various colours.

(-) Indicates identical amino acid between scallop and $LLFF_{15}$.

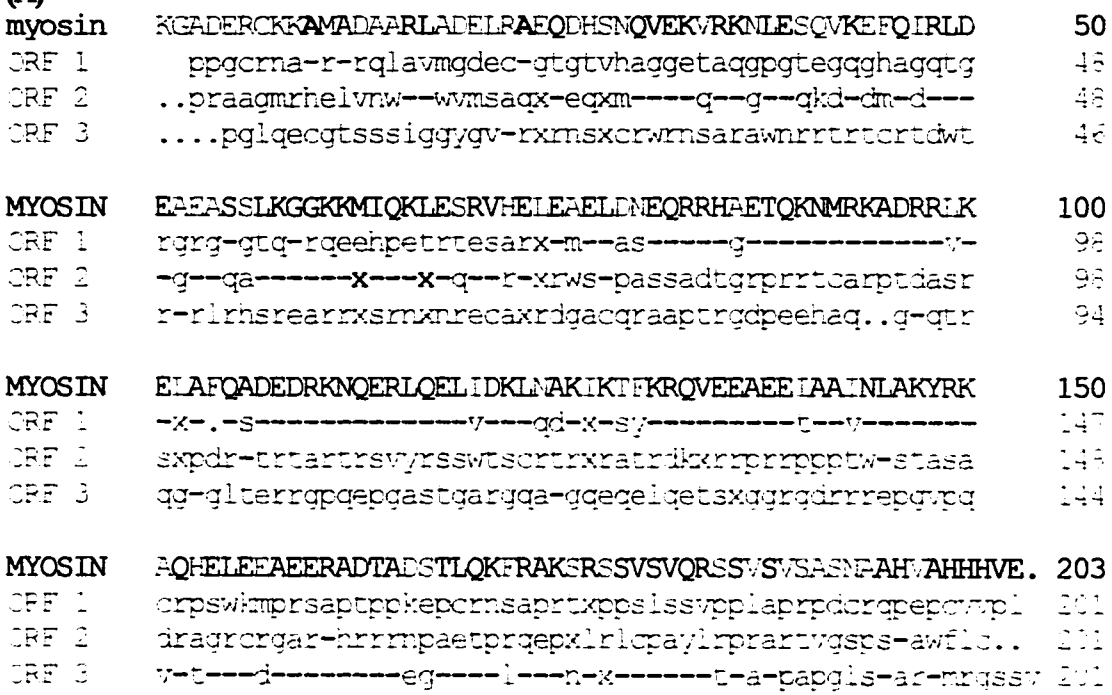
ORF 1 Blue text

ORF 2 Red text

ORF 3 Green text

(B) Complete myosin heavy chain amino acid sequence from *Argopecten irradians*. Of the 1951 amino acids, only the red section was used in the comparison with $LLFF_{15}$.

(A)



```
myosin   KGADERCKKAMADAARLADELRAEQDHSNQVEKVRKNLESQVKEFQIRLD      50
ORF 1      ppgcrna-r-rqlavmgdec-gtgtvhaggetaqgpgtegqghagqtg      48
ORF 2    ..praagmrhelvnw--wvmsaqx-eqxm----q--g--qkd-dm-d---     48
ORF 3    ....pglqecgtsssiggygv-rxrnsxcrwrnsarawnrrtrtcrtdwt     46

MYOSIN   EAEASSLKGGKKMIQKLESRVHELEAELDNEQRRHAETQKNMRKADRRLK     100
ORF 1    rgrg-gtq-rqeehpetrtesarx-m--as-----g----------v-      98
ORF 2    -g--qa------X---X-q--r-xrws-passadtgrprrtcarptdasr     98
ORF 3    r-rlrhsreaxrxsrnxnrecaxrdgacqraaptrgdpeehaq..g-qtr     94

MYOSIN   ELAFQADEDRKNQERLQELIDKLNAKIKTFKRQVEEAEEIAAINLAKYRK     150
ORF 1    -x-.-s----------v---qd-x-sy---------t--v-------      147
ORF 2    sxpdr-trtartrsvyrsswtscrtrxratrdkxrrprrppptw-stasa     146
ORF 3    qg-glterrqpqepgastgarqqa-gqeqelqetsxggrqdrrrepgvpq     144

MYOSIN   AQHELEEAEERADTADSTLQKFRAKSRSSVSVQRSSVSVSASNPAHVAHHHVE. 203
ORF 1    crpswhmprsaptppkepcrnsaprtxppslssvpplaprpdcrqpepcvrpl  201
ORF 2    dragrcrgar-hrrrmpaetprqepxlrlcpaylrprartvgsps-awflc..  201
ORF 3    v-t---d--------eg----l---n-x------t-a-papgls-ar-mrqssv 201
```

(B)

```
mnidfsdpdfqylavdrkklmkeqtaafdgkkncwvpdekegfasaeiqsskgdeitvkivadsstrtvkk
ddiqsmnppkfekledmanmtylneasvlynlrsrytsgliytysglfciavnpyrrlpiytdsviakyrg
krkteipphlfsvadnayqnmvtdrenqsclitgesgagktestkkvimyfarvaanlykqkqeepttttha
rasnledqiieanpvleafgnaktvrnnnssrfgkfirihfgptgkiagadietylleksrvtyqqsaern
yhifyqicsnaipelndvmlvtpdsglysfinqgcltvdniddveefklcdeafdilgftkeekqsmfkct
asilhmgemkfkqrpreeqaesdgtaeaekvaflcginagdllkallkpkvkvgtemvtkgqnmnqvvnsv
galakslydrmfnwlvrrvnktldtkakrnyyigvldiagfeifdfnsfeqlcinytnerlqqffnhhmfi
leqeeykkegiawefidfgmdlqmcidliekpmgilsileeecmfpkaddksfqdklyqnhmgknrmftkp
gkptrpnqgpahfelhhyagnvpysitgwleknkdpinenvvallgaskeplvaelfkapeepagggkkkk
gkssafqtisavhreslnklmknlysthphfvrciipnelkqpglvdaelvlhqlqcngvlegiricrkgf
psrliysefkqrysilapnaipqgfvdgktvsekilaglqmdpaeyrlgttkvffkagvlgnleemrderl
skiismfqahirgylirkaykklqdqriglsviqrnirkwlvlrnwqwwklyskvkpllsiarqeeemkeq
lkqmdkmkedlakterikkeleeqnvtlleqkndlflqlqtledsmgdqeerveklimqkadfesqikele
erlldeedaaadlegikkkmeadnanlkkdigdlentlqkaeqdkahkdnqistlqgeisqqdehigklnk
ekkaleeankktsdslqaeedkcnhlnklkakleqaldelednlerekkvrgdvekakrkveqdlkstqen
vedlervkreleenvrrkeaeisslnskledeqnlvsqlqrkikelqarieeleeeleaernarakvekqr
aelnreleelgerldeaggatsaqielnkkreaellkirrdleeaslqheaqisalrkkhqdaanemadqv
dqlqkvksksekekqqlrsevedlqaqiqhisknkgcsekvmkqfesqmsdlnarledsqrsinelqsqks
rlqaensdltrqledaehrvsvlskeksqlssqledarrsleeetrarsklqnevrnmhadmdaireqlee
eqesksdvqrqlskanneiqqwrskfeseganrteeledqkrkllgklseaeqtteaanakcsalekaksr
lqqeledmsievdranasvnqmekkqrafdkttaewqakvnslqselensqkesrgysaelyrikasieey
qdsigalrrenknladeihdltdqlseggrstheldkarrrlemekeelqaaleeaegaleqeeakvmraq
leiatvrneidkriqekeeefdntrrnhqralesmqasleaeakgkadamrikkkleqdinelevaldasn
rgkaemektvkryqqqiremqtsieeeqrqrdearesynmaerrctlmsgeveelraaleqaerarkasdn
eladandrvneltsqvssvqgqkrklegdinamqtdldemhgelkgaderckkamadaarladelraeqdh
snqvekvrknlesqvkefqirldeaeasslkggkkmiqklesrvheleaeldneqrrhaetqknmrkadrr
lkelafqadedrknqerlqelidklnakiktfkrqveeaeeiaainlakyrkaqheleeaeeradtadstl
qkfraksrssvsvqrssvsvsasnaahvahhhve
```

Figure 2.12: Hybridization of clone $LLFF_{15}$ against total RNA from the foot muscle of freezing and anoxia treated animals.

(A) Binding of $LLFF_{15}$ to its complementary message in the total RNA on a northern blot.

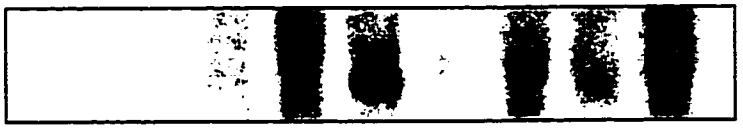
(B) Ethidium bromide-stained rRNA band from the formaldehyde gel.

(C) Illustration of the percentage of clone $LLFF_{15}$ pixel density with respect to the corresponding rRNA band.

(A)


C 1 12 24 24R 1 12 24 24R
Frozen
Anoxic

(B)



(C)

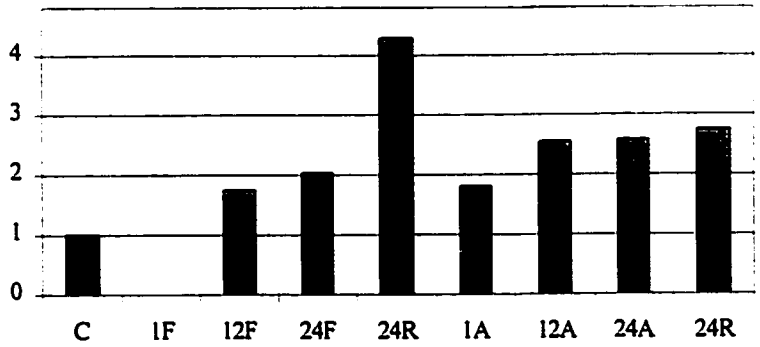

4
3
2
1
0
C 1F 12F 24F 24R 1A 12A 24A 24R

Figure 2.13: DNA sequence of clone $LLFF_{21}$

The 5' region of this clone shows high similarity with known genes for actin and consequently it was possible to determine that the clone is missing the 5'UTR and start codon. The remaining features of a mRNA are present, including:

(A) 5'-section of the clone sequence

LEGEND:

coding region	plain black text
stop codon	**box**
3' UTR	*italics*

(B) 3' section of clone sequence

3' UTR	*italics*
polyA signal	**black, bold type**
polyA tail	***bold italics***

(A) 5′-3′ sequence

1 CAGGAMTTCS GCACGAGGGG CGACGAGGCC AGCGCAAGCG TGGTATCNNC ACCCTCAAGT
61 ACCCCATCGA GCACGGCATC GTCACCAACT GGGACGATAT GGAGAAGATC TGGCACCACA
121 CCTYNTACAA CGAGCTGCGT GTGGCCCCCG AGGAGCACCC CGTCCTGCTG ACAGAGGCTC
181 CCCTCAACCC CAAGGCTAAC CGTGAGAAGA TGACCCAGAT CATGTTCGAG ACCYTCAACG
241 CCCCCGCCAT GTACGTCGCC ATCCAGGCCG TGCTCTCCCT GTACGGTTCC SGTCGTACCA
301 CCGGCATCGT GCTCGACTCC GGAGATGGTG TCACCCACAC CGTGCCCATC TACGAGGGTT
361 ACGCTCTCCC CCACGCCATC CTCCGTCTGG ACTTGGCTGG CCGTGACCTC ACTGACTACC
421 TGATGAAGAT CCTCACAGAG CGCGGTTACT CCTTCACCAC CACCGCCGAG CGTGAGATCG
481 TTCGCGACAT CAAGGAGAAG CTCTGCTACG TCGCCCTGGA CTTCGAGCAG GAGATGGCCA
541 CCGCCGCCTC CTCCTCMAAW RRGGAGAAGA GCTACGAGCT TCCCGACGGT CAGGTCATCA
601 CCATCGGAAA CGAGCGCTTC ASGTGCCCCG AGTCTCTCTT CCAGCCTTCC TTCCTGGGTA
661 TGGCATCTCC TGGTATCCAC GAGACCACCT ACAACTCCAT CATGAAGTGC GACGTGGACA
721 TCCGTAAGGC CCTGTACGCC AACACCGTGC TGTCCGGAGG CACCACCATG TTCCCCGGTA
781 TCGCCGACAG GATGCAGAAG GAGATCMCCG CCCTGGCTCC CAGCACCATG AAGATCAAGC
841 CCACCGCTCC CCCCGAGCGC AAATACTCCG TATGGCCCSG YGGCTCCATC CTCGCTTCCC
901 TGTCCACCTT CAGCAGATGT GGATCTCCAA GCAGGAGTAC GACGAGTCCG GCCCTTCCAT
961 CGTCCACCGA AAGTGCTTCT AAACAGCAGA TACTGGACA

(B)

1 AACCCCCCCC CGAAAAAAAA ACCCCGTAGG GACGGGGCCC CAACCCCCCG TCCCTGCCAC
61 CTTCCAGCAG ATGGGGTCCC CAAGGCGGAG TCGACGAGTC GGCCCTTCCA CCGCCCACCG
121 CAAGTCCTAC TAAACAGCAG ATACTGRAAA ATAACAAATT TTCTGATCCT CTTAGCATTT
181 TTCCAATCTG TCTTGGGGGG GGGRCTCCCG TCTTCTTCTC AAATCAAATT TCGGCCATTT
241 TTTTGTTAGC CCGCTTCAAA GTTCCAACAA GCGAGAAACA AACTTTTGTT CATGCCGCCT
301 GACACRAAAT AAACTTTTTC CAACCTTGRC CTTCACCGTC TTCTTCTTGA CATGRGKTCG
361 TGTGAATGTT TCTGAAGATA AAAAGACACC CGCGGCTGCT ACGTGTTTGA GAAGACAGAG
421 TGTCTGGVCC TTWAHTTTTA CATCCCCCGA CATCTTATCT TGGTCAAATG GTGCKGGAWA
481 AAAATACCAG GCTTTGTGAA TCTAACTGCC AGAAGTGGAC GGAGAAAAAT CCGGCCGGAG
541 ATGGTTATTA CCCCCCCCCS CCTGGGTCAG AAAGACTTAA CTCTTTGTAA ACGCGCATAG
601 AATGTTCATG TCTGATTGTA TATGTCGATT GAACGAAAGV TCTTGGSCTT CGACTCATTT
661 TATTCCTATA AATAAATGTG CGAGTCAGTC TTTCAACCAA TCATCAGCAA CCAAGTCGAA
721 CGVAGAGCCA ATCGAAAGAA AGAACACAGA GGGGGCGCGC GATATACAAA CTCAGGTGAA
781 CTTTCTGTSG CTTATACAGA TGTGTTCAAK GTTCGAACCK KGNRGTAGAA AAAATTAAAA
841 GAGAAAAACC TAATTGAAAT ACTGCCATGT TGANCGTGTV AGAATGTGAT ATCTTTTTGT
901 ATCACTTCAC CATGTCCACT TTATTCACTG TGVRNGATGC AATGTTGAAC GTAGAATTTC
961 CATCCGCCAC ATTAAATAGT TTTCAATAAA TGATTGTGTA CATTTGTAAA AAAAAAAAAA
1024 AAAA

Figure 2.14 Nucleotide and translated amino acid sequence of clone $LLFF_{21}$.

(A) Frame shifts in the coding reading frame are indicated by continuing the amino acid sequence on another line

(B) Alignment of $LLFF_{21}$ with published beta actin proteins from: the sea scallop *Placopecten magellanicus*, Artemia, and *Xenopus laevis*. Dashed lines indicate residues that are identical among the sequences.

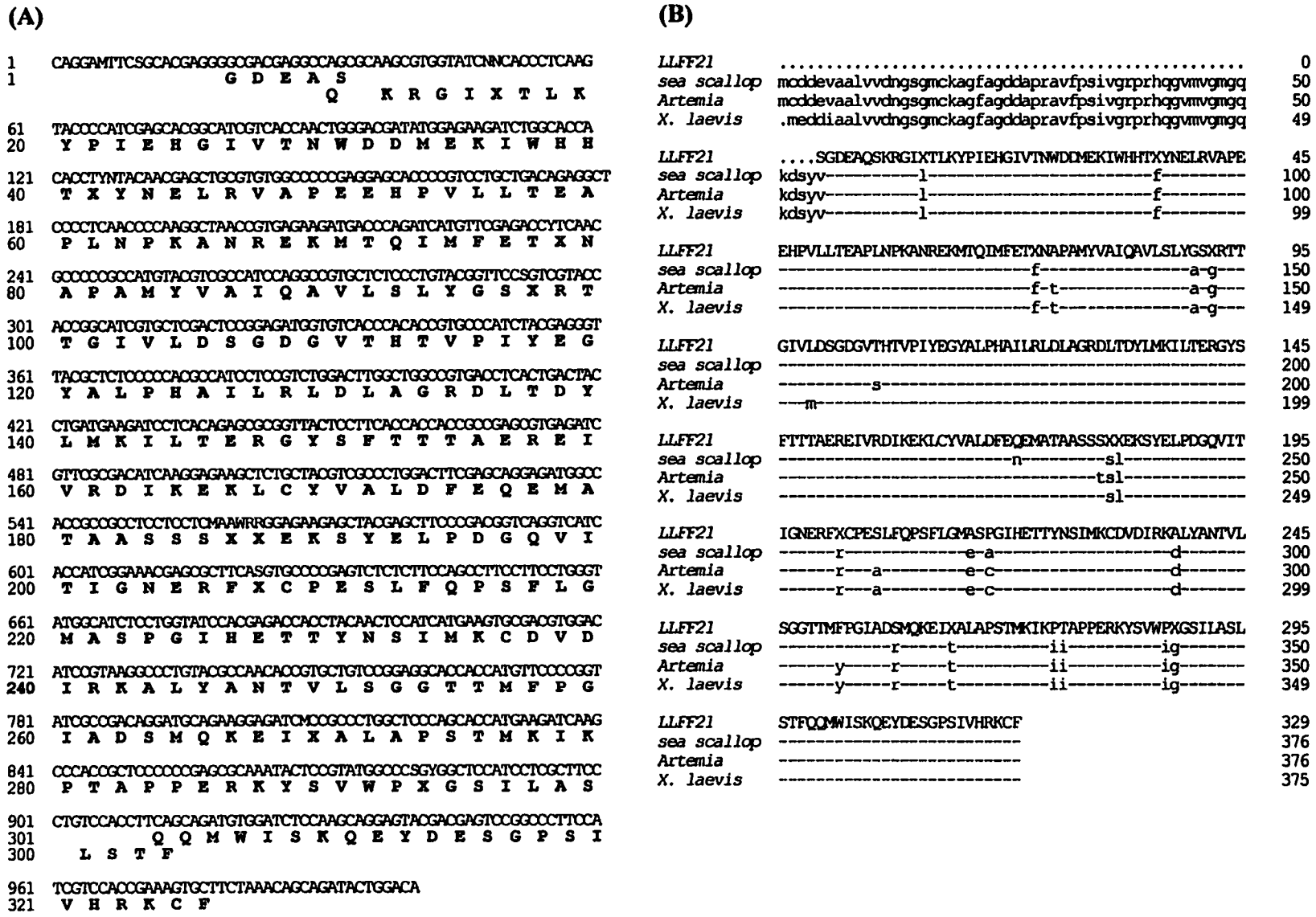

(A)
(B)
LLFF21
sea scallop
Artemia
X. laevis

Figure 2.15 Hybridization of clone $LLFF_{21}$ against total RNA from the foot muscle of freezing and anoxia treated animals.

(A) Time points

(B) Northern blot probed with $LLFF_{21}$

(C) Ethidium bromide staining of rRNA from blot (B)

(D) Graphical representation of the pixel density ratio of $LLFF_{21}$ with respect to the corresponding rRNA band.

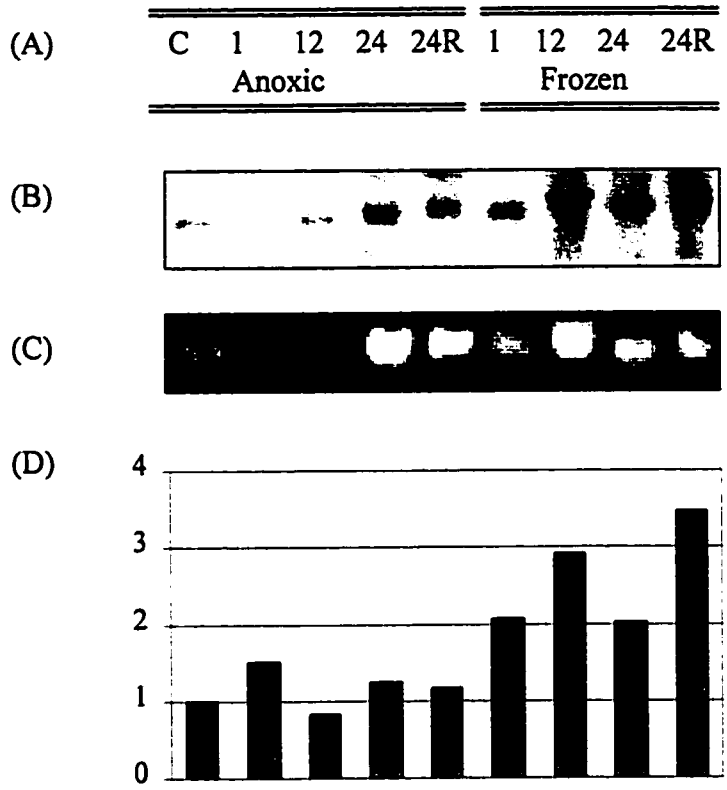
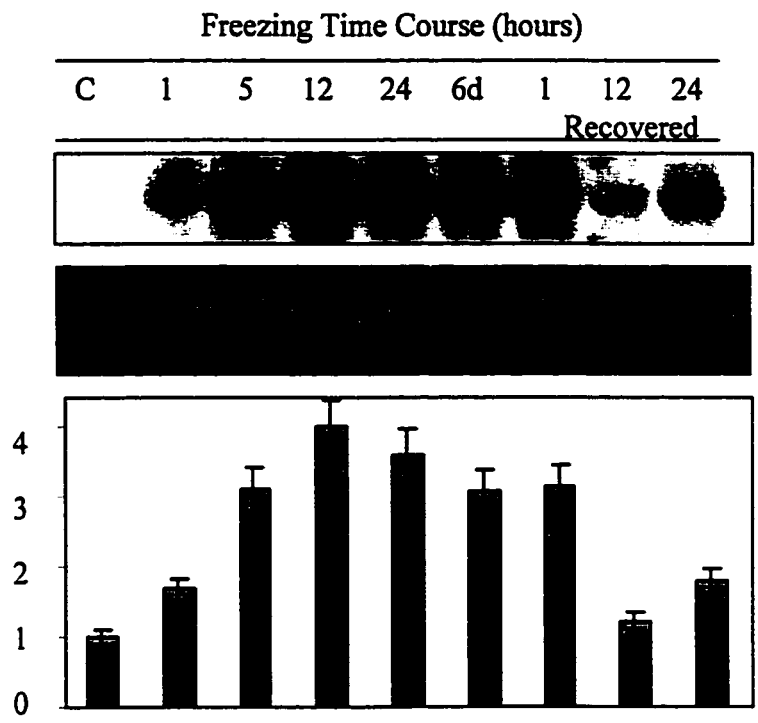

(A)
C 1 12 24 24R 1 12 24 24R
Anoxic
Frozen
(B)
(C)
(D)
4
3
2
1
0

Figure 2.16 Hybridization of clone $LLFF_{21}$ against total RNA from the hepatopancreas of freezing and anoxia treated animals.

(A) Freezing time course.

(B) Anoxia time course.

(A)


Freezing Time Course (hours)
C 1 5 12 24 6d 1 12 24
Recovered
4
3
2
1
0

(B)

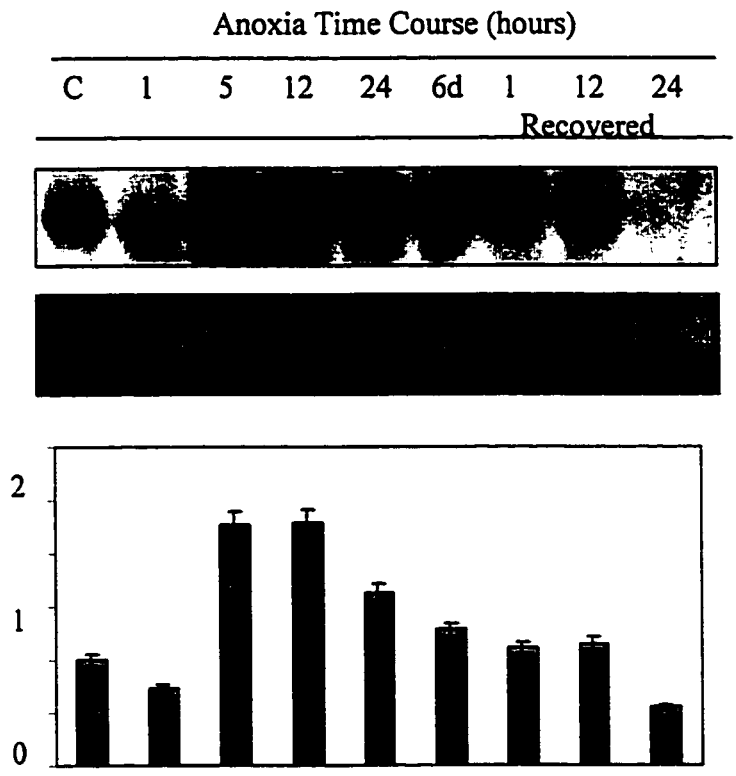

Anoxia Time Course (hours)
C 1 5 12 24 6d 1 12 24
Recovered
2
1
0

Figure 2.17: DNA sequence of clone $LLFF_{20}$.

(A) This clone was identified as that for cytochrome c oxidase subunit 2, which is encoded on the mitochondrial genome. These genes have the same characteristics as nuclear genes, but are arranged in a more compact manner. The stop codon, poly adenylation signal and polyA tail all overlap at the 3' end. The clone is incomplete, lacking a start codon.

LEGEND:

Coding region	plain black text
Stop codon	box
PolyA signal	**black, bold text**
PolyA tail	italics

(B) Alignment of $LLFF_{20}$ with other COXII amino acid sequences. Identical amino acids are indicated with a dashed line (---).

(C) Homology matrix of the 5 sequences

(A)

```
1     GAGGATTTCA AGACGCTGCT TCTCCTTTAA TAGAAGAACT AATTTTTTTC CATGACCACG
61    CAATAATAAT TTTAGTTATA ATTATTAGCC TAGTTGGCTA TGCTGCTTTA TCCTTAGTAT
121   TAAATAAATA TACTTGCCGA TCTCTTGTTG AGGGTCAAGA AATTGAGACC ATCTGAACTA
181   TTATCCCAGC TATCATTTTA GTATTCCTAG CTTTACCTTC TCTCCGACTT TTATACCTTC
241   TAGATGAAGT TGGTGACTGC AGCTTAACTG TGAAAAGTAT TGGTCACCAA TGATATTGAA
301   GCTACGAGTA CTCCGATTTT TTAAATATTG AGTTTGATTC CTACATAATT CCAACTAACG
361   AATTAGAAAG TGGGGACTTT CGCTTACTAG AAGTAGACCA CCGTGTTGTC CTTCCTACTC
421   AGACTGACAT CCGAGTCTTA GTAACGTCTG CTGACGTAAT CCACTCTTGG ACCGTACCTT
481   CTTTAGGGAT TAAAGCTGAC GCCGTTCCCG GACGTTTAAA CCAACTTAGA TTTTATATCA
541   AGTACCCAGG AGTATTTTAT GGCCAATGTT CTGAAATTTG CGGAGCAAAT CACTCTTTTA
601   TACCAATTGT GGTTGAGGCT GTACCCCTTA AGAACTTTAT ACAATGGGTC GTGCATGTTT
661   CTGAATAATT AAAAAAAAAA AAAAAAAA
```

(B)

```
L. littorea   .......GFQDAASPLMEELIFFHDHAMMILVMIISLVGYAALSLVLNKYTCRSLVEGQE  53
L. saxatilis  mslwgqw----------------------i----------m--------------  60
K. tunicata   mafwsqw----g---i--q---------l--i-----ls-g-v--mn-sfls--tl-s--  60
human         mahaaqv-l---t--i-----t-----l--ifl-cf--l--lfltltt-l-ntnisda--  60
Drosophila h. mstwanl-l--s------q--------ll-----tv----lmfm-ff--hvn-f-lh--l  60

L. littorea   IETIWTIIPAIILVFLALPSLRLLYLLDEVGDCSLTVKSIGHQWYWSYEYSDFLNIEFDS 113
L. saxatilis  ------------------------------------------------------------ 120
K. tunicata   --iv---l--vv-i---f---q-------leepa--i-vv---------------l---- 120
human         m--v---l------li------i--mt---n-p---i---------t---t-yggli-n- 120
Drosophila h. --m----l-----l-i-f----------inep-v-l----------------n------ 120

L. littorea   YMIPTNELESGDFRLLEVDHRVVLPTQTDIRVLVTSADVIHSWTVPSLGIKADAVPGRLN 173
L. saxatilis  ----------------------------v---------------------------- 180
K. tunicata   ---sled--e--y--------s-v-mk-kv-----a---l---------v---------- 180
human         --l-plf--p--l---d--n-----ieap--mmi--q--l---a--t--l-t--i----- 180
Drosophila h. --v-----svdg----d--n-i---mnsq--i---a--------i-a--v-v-gt----- 180

L. littorea   QLSFYIKYPGVFYGQCSEICGANHSFMPIVVEAVPLKNFMQWVVHVSE.            221
L. saxatilis  ----------------------------------------a-----.              228
K. tunicata   ----fan----------------------l-v-dsss-ik-imfng-a             229
human         -tt-tatr---y-----------------l-li---i-emgp-ftl..             227
Drosophila h. -tn-f-nr--l-----------------i-s--vny-ik-isnnmns              229
```

(C)

L. littorea	**100%**				
L. saxatilis	98.2%	100%			
K. tunicata	70.1%	71.1%	100%		
Human	67.9%	66.2%	62.4%	100%	
Drosophila	61.4%	59.9%	53.3%	55.5%	100%

DISCUSSION

In recent years, advances in the development of molecular biology techniques have made it faster, easier and perhaps, most importantly for those involved in basic research, cheaper, to determine the genetic changes induced by environmental stress or differing biological states. Investigators with widely varied interests have used identification of differentially expressed genes as an experimental approach to understand not only gene function, but also molecular mechanisms underlying a biological process. A huge variety of systems have been studied in this manner, as illustrated by the more than 11,000 articles returned after a search for "differential gene expression" in the Medline database.

Studies of differential gene expression have only recently been conducted on stress-tolerant species. The first studies with respect to cold tolerance were completed in plants, where cold exposure seems to induce the expression of embryogenesis abundant protein (Baker *et al.*, 1988), dehydrin (Close *et al.*, 1989), heat-shock proteins (Neven *et al.*, 1993), and antifreeze proteins (AFPs) implicating a role in creating cold hardiness in plants (Cherry, 1994). Studies with animals have revealed genes encoding for desaturase in fish (Tiku *et al.*, 1996), ferritin H subunit in rainbow trout cells (Yamashita *et al.*, 1996), and for AFPs in cold water fish species (Davies *et al.*, 1999) and cold-hardy insects (Tyshenko *et al.*, 1997). Recent studies in our lab employing differential screening of a wood frog liver cDNA library lead to the discovery of genes up-regulated under freezing exposure (fibrinogen, ADP/ATP translocase and a 10 kDa novel sequence) which may be involved in cold adaptation in this freeze-tolerant species (Cai and Storey, 1997a, 1997b; Cai et al., 1997)

Combined with the genes outlined in appendix I, the breadth of these differentially expressed genes makes it clear that when embarking on a new aspect of stress survival, the straight-ahead accumulation of raw genetic data is a valid and

productive approach. Given the accessibility of techniques used in the detection of differentially expressed genes, both in terms of cost and "user-friendliness", it seems absurd to *not* employ the diffuse approach of comparative screening. It is a crucial step that highlights areas of physiology or biochemistry that may not have previously been considered a part of stress tolerance.

Additional differential techniques

cDNA libraries have been employed to isolate differentially expressed genes for decades. However, the costs and technical challenges associated with the technique, made it a daunting project for comparative biochemists or physiologists. However, the commercial systems and pre-packaged reagents now available have made it a somewhat simpler endeavor. The drive to shorten the time span required to obtain data has lead to the development of additional techniques that can be used to obtain the same results, each of which has its own advantages and problems.

Perhaps the most accessible and popular of the techniques is differential display PCR. DD-PCR employs random primers and a radio-labeled nucleotide to amplify a subfraction of total mRNA from 2 different populations. The PCR products are run side by side on sequencing gels and the cDNAs that produce signals of differing intensity under the two conditions are isolated, reamplified and cloned into a vector. Since its first description in 1992 (Liang and Pardee), over 1000 articles in PubMed describe the use of this technique, but there are also nearly 100 articles that describe either a modification to the protocol or a call for improvement. In contrast to a cDNA library screen, obtaining the differentially expressed cDNAs is the easy part of DD-PCR. A multitude of problems arise after this step, including a high rate of false positives and a limited ability to identify both rare and abundant mRNAs. But the greatest annoyance with DD-PCR is the small size of the cDNA clones which are predominantly, if not completely, composed of 3'UTR. Given that UTR sequences can vary greatly even among genes coding for the

same functional protein, such a clone can make cDNA identification through BLAST seraches difficult. In addition, the lack of gene specificity makes DD-PCR products difficult to use as probes to screen a cDNA library for the complete gene, nor are they appropriate for 5'RACE (a PCR method to elucidate a longer fragment of gene sequence) because preparation of gene-specific primers is difficult. To solve such problems, a recent modification reported by Jurecic et al. (1998) shows that incorporation of long-distance PCR into DD-PCR enables highly reproducible amplification and comparative display of cDNAs ranging in size from 150 bp to 2 kb. Despite the traditional technical problems, when all circumstances are at their best, DD-PCR is one of the fastest of the expression-screening techniques.

Subtractive hybridization is a technique that removes the cDNAs which are expressed equally in both conditions. In general, the cDNAs from the stressed state are hybridized with a large excess of cDNA from the control condition, followed by gel chromatography to remove the double stranded nucleic acid hybrids from the single-stranded cDNAs (which correspond to differentially expressed mRNAs). The result is a pool of cDNAs that is greatly enhanced with respect to differentially expressed messages. At this point, the cDNA can be used for either library construction or probe synthesis. Unfortunately, the long polyA tracts of the mRNA can form hybrids with polyT sequences of entirely unrelated genes, resulting in the loss of template during the normalization step (Wang et al., 2000).

DNA microarrays consist of measured amounts of DNA sequences in an organized pattern on a nylon membrane or small glass plate that are screened as for cDNA library screens. Two populations of cDNAs are labeled and simultaneously hybridized to identical arrays, revealing genes that are present more intensely in one condition over the other. This technique can be used for quantitative monitoring and analysis of cell-, tissue- and process-specific gene expression profiles. However, despite the tremendously simple screening protocol, commercially available arrays exploit

genetic information obtained from common animal models, which are widely divergent from *L. littorea*. Therefore, discovery of novel periwinkle clones would be unlikely using a commercial array.

Oddly enough, while the microarrays evolved from generalized screening techniques, they can now be used with a specific goal in mind because they can be obtained with a series of related genes. For example, an investigator interested in the effect of a stress or a given biological process (such as cell cycle progression) on a particular gene can purchase an array, differentially screen it and have the answer, avoiding the time and technical skills required for DD-PCR or in the synthesis of a cDNA library. However, even microarrays cannot remove the task of confirming upregulation. Even when an upregulated gene has been identified, either northern blot hybridization or RT-PCR must be completed to illustrate that the transcript levels do indeed change when the stress is imposed.

Results of cDNA library screening

In this study I investigated the role of differential gene expression in response to environmental stress in the foot muscle of *Littorina littorea*. This gastropod species inhabits the coastal regions of the Atlantic ocean along north America and Europe. As an intertidal species, this animal faces decreased oxygen availability on a daily basis, which is complicated by freezing temperatures over the course of winter. Although unlikely to experience such an extended period of anoxia or freezing, this species can survive 6 days under a nitrogen environment or frozen at –4°C. To my knowledge, this is the first study of this stress-tolerant species to use differential screening of cDNA libraries as a tool to elucidate the contribution of gene expression to stress survival. I identified a net total of six upregulated clones that fall into 3 categories: upregulated in response to freezing, or anoxia or both.

Novel clones

$LLFF_3$, $LLAF_N$, and $LLAF_{4.2.1}$ were more than 99% identical where they overlapped. Such redundancy in the data indicates a very thorough screening of the cDNA libraries and, by extrapolation, that all genes that *could* be detected using this technique were indeed isolated. In addition, upregulation of the same message in response to both freezing and anoxia, indicates that a common mechanism induces expression. Unfortunately, the clone could not be identified by comparison to DNA sequences in public databases, making further analysis necessary. However, the DNA sequence of $LLFF_3$, the longest of these clones, shows 2 complete ORFs that can be translated to an amino acid sequence. The following chapter contains a detailed analysis of this clone and the use of a translated amino acid sequence to obtain identification.

Clones, $LLAF_{EE}$, $LLAF_{7.1.4}$, and $LLFF_{13}$ are ***also*** novel DNA sequences that do not show similarity to published sequences. Where they overlap, the DNA sequences of $LLAF_{EE}$ and $LLFF_{13}$ are identical, and represent a novel gene that was up-regulated in response to both freezing and anoxia. Interestingly, whereas $LLAF_{7.1.4}$ is very similar to $LLAF_{EE}$, its DNA sequence has several large gaps that prevent a completely identical alignment. By way of a possible explanation, these clones could either represent a family of related genes, each of which is transcribed individually, or a single gene that is alternatively spliced into a series of message RNAs that are translated separately.

Such a possibility can be investigated by probing a Southern blot. A series of blots containing digested *L. littorea* genomic DNA could be probed separately with ^{32}P-labeled $LLAF_{EE}$, and $LLAF_{7.1.4}$, whereas a third blot would be hybridized with a mixture of both probes. If the clones belong to a single gene that is differentially spliced at the transcription stage, all probes will bind to the same, solitary gene, hopefully represented by one identical band on the blot. However, if the clones represent a gene family, each probe will bind to a differently sized band, and the blot hybridized with the mixed probe

will show multiple bands. However, the high degree of similarity among these clones makes it unlikely that this would be a conclusive experiment.

$LLAF_W$, is also a novel DNA sequence. Unlike the other novel clones, $LLAF_W$ was isolated only from the anoxic library, indicating an anoxia-specific trigger that induces gene expression. The DNA sequence can be translated from the first nucleotide until a stop codon at 283 bp, perhaps indicating that the clone does not represent a complete message. Thus, further experiments were required to confirm a stress-specific response and transcript size, isolate a longer cDNA if necessary, and analyze the complete translated ORF for clues as to function. These points are addressed in detail in chapter 4.

Identified clones

Screening of the library created from the foot muscle of freeze-treated snails revealed genes that were conclusively identified. Myosin, actin, and subunit 2 of cytochrome c oxidase (COXII) all showed stronger signals on a reverse northern probed with total mRNA probe from the freezing state. However, they were not isolated from the anoxic library, indicating a freezing-specific response, which will be addressed below.

Actin and Myosin

Clones $LLFF_{15}$ and $_{21}$, isolated from the *L. littorea* foot muscle cDNA freezing library, contained nucleotide sequences encoding the C-terminal ends of myosin and actin, respectively, but the N terminus was absent in both cases. However, the 329 amino acids of $LLFF_{21}$, and the 201 residues of $LLFF_{15}$ shared a high percentage of residues with the respective proteins from other invertebrate sources. The amino acid sequence of $LLFF_{21}$ was 94.4, 93.1 and 93.4 % identical with that of actin from the sea scallop, *Atermia* and the toad, *Xenopus laevis*, respectively. $LLFF_{15}$ was more difficult to compare due to the number of frame shifts and possible sequencing errors but 50%

residues compared in figure 2.11 (A) were shared between $LLFF_{15}$ and the myosin heavy chain (MHC) of the scallop, *Argopecten irradians*.

A recent study involving Richardson's ground squirrels showed increased expression of myosin light chain (MLC) in the heart during hibernation (Fahlman et al., 2000). The authors suggested that this could occur in an effort to restructure the contractile apparatus and maintain appropriate heart function in the hypothermic state. While this is a possible explanation in *L. littorea*, it is not probable, because a snail should not have a tremendous biological need to maintain strong contractile properties in the foot muscle during or after stress, as they are likely immobile during the stress. In addition, the pattern of transcript accumulation is most obvious during exposure to and recovery from freezing. Such a delayed response does not support a function for increased actin and myosin transcripts during freezing itself, but since freezing can cause damage to structural components of the cell, the induction of genes for proteins of the contractile apparatus may have a function in preparing for freeze-induced tissue damage.

Cytochrome c Oxidase, subunit 2

Mitochondria contain their own DNA (mtDNA), a circular molecule that is strongly conserved. Completely sequenced mtDNA sequences for invertebrates range from 18,224 bp in *Crassostrea gigas* (accession #: NC_001276) to15,532 bp in *Katharina tunicata* (NC_001636). Mitochondrial genomes usually encode 2 rRNAs (12S and 16S rRNA), 22 tRNAs and 13 proteins. The genome is economically organized, with tRNA genes interspersed between the rRNA and protein-coding genes with few non-coding nucleotides between the coding sequences. The encoded proteins are subunits of respiratory chain enzymes, including cytochrome c oxidase subunit II (COXII).

Upregulation of other mitochondrial genome-encoded genes in response to stress has also been reported. Other studies in our lab showed that genes encoding NADH-ubiquinone oxidoreductase, subunit 5 and cytochrome oxidase subunit 1 (COI) were both

up-regulated by oxygen deprivation in the heart of anoxia-tolerant turtles (*Trachemys scripta elegans*) (Cai and Storey, 1996) and freezing in the freeze-tolerant hatchlings of the turtle *Chrysemys picta marginata* (Cai and Storey, 1996). In Richardson's ground squirrel, the transcript for the mitochondrial gene encoding subunit 2 of NADH-ubiquinone oxidoreductase (Nad2) increased 2-fold in the heart and liver, and 4-fold in muscle during hibernation.

Accumulation of selected mitochondrial gene transcripts has also been reported in mammals in acute myocardial stress in a cold environment. The upregulation of mtATPase 6 & 8, COI, II, III and cyt b genes under global ischemia and reperfusion was found in the rat heart (Das et al., 1995). The mitochondrial 16S rRNA, and ATPase 6 & 8 gene transcripts were accumulated in high levels in human myoblasts and increased progressively as myotubes matured (Webster *et al.*, 1990). The accumulation of mitochondrial COXII transcript and 16S rRNA was also reported in the liver of rats exposed to cold stress (at 4^{o}C) (Martin *et al.*, 1993). By contrast, oxidative stress leads to the down-regulation of mammalian mitochondrial RNAs (such as the 16S RNA) (Crawford *et al.*, 1997).

Reverse northerns using $LLFF_{20}$ consistently bound more message using total mRNA probe from the freezing state, indicating stress-induced expression. Unfortunately, northern blots did not yield the same result, showing consistent expression across the time courses of both freezing and anoxia. I believe this discrepancy may result from a technical detail. The number of transcripts in the sample is derived from the message RNA on a reverse northern and total RNA on a northern blot. In terms of sample size, upwards of 10-20 animals were used to isolate the mRNA, but only 2-3 were used for the total RNA on the northern blot. So, perhaps the increase on the reverse northern is due to a larger sample size, and the COXII transcript does not change in every animal, every time.

In summary, the differential screening of these cDNA libraries revealed known and novel genes that were upregulated at different stages of stress and recovery. Details regarding the pattern of upregulation of the identified genes were presented above, but the novel clones required further analysis. In chapter 3, the isolation of a putative metallothionein gene is reported, while chapter 4 presents the two novel clones whose translated amino acid sequences contain N-terminal secretory signals. In each case, detailed analysis of the amino acid sequences was required to identify motifs in both the nucleotide and corresponding amino acid sequences and then use such motifs to attempt identification or define a putative function.

CHAPTER 3

FREEZING AND ANOXIA STRESSES INDUCE EXPRESSION OF A PUTATIVE METALLOTHIONEIN GENE IN THE FOOT MUSCLE OF *LITTORINA LITTOREA*.

INTRODUCTION

Metallothioneins (MTs) are a family of low molecular weight, cysteine-rich proteins that characteristically bind metals. MTs are a common protein class that has been identified in microorganisms, invertebrates, vertebrates and plants. While their precise physiological relevance is not clear, the broad taxonomic range and tissue distribution indicate that this protein family makes an important contribution to the cellular machinery.

Four MT isoforms have been identified in mammals, each with distinct sequences and some with tissue-specific expression. MT-1 and MT-2 are coded by nonallelic genes and are expressed to varying degrees in all tissues. MT-3 was isolated from human brain by Palmiter et al. (1992) and seems to be specific to that tissue, whereas MT-4 is present only in stratified squamous epithelial tissue (Quaife et al., 1994). The original classification system for MTs was proposed by Nordberg & Kojima (1979) and was based on these mammalian isoforms and by 1985 was extended, by introducing a subdivision of all MTs into three classes (Fowler et al., 1987). Class I included all MTs with cysteine residues aligned closely with those of mammalian forms, while those that lacked this characteristic were assigned to class II. Class III was composed of metalloisopolypeptides that containing gammaglutamyl-cysteinyl subunits. However, the growing number and variety of identified sequences made such a system inadequate, and so, Binz and Kägi (1999) have recently proposed an alternative "family" system. A MT family accounts for MTs which share a particular set of sequence-specific characters and are thought to be evolutionary related, hence each family is identified by a number and its

taxonomic range, and then further seperated into subfamilies containing MTs that share a set of more stringent phylogenetic features. The molluscan MTs are denoted as family 2, and subdivided into the following groups: mo1 (mussel MT-1), mo2 (mussel MT-2), mog (gastropod MTs), and mo (other mollusc MTs).

Regardless of classification, all MTs have sequence and structural similarities. They consistently show few or no aromatic amino acids and an unusually large number of cysteines (20-25% of total residues). The cysteines also fall into a common motif, Cys-X-Cys, where X = any amino acid. While this motif is common in mammalian species, it has also been identified in most MTs, including the terrestrial snail *Helix pomatia*, where 18 cysteines of 66 total amino acid residues are arranged in seven Cys-X-Cys motifs (Dallinger et al., 1993). Although there is minimal sequence identity among MTs from different species, the presence of these motifs is a unique marker that makes it possible to propose the identity of MT for a novel sequence.

Spatial structures of mammalian, crustacean, and an echinoderm MT have been derived from 2D NMR spectroscopy and X-ray crystallography. Despite great variation among the primary amino acid sequences, the spatial structures are very similar. They all resemble a dumbell-like shape with two separate protein domains containing a core of complexed mineral-like clusters. All cysteines are involved in binding the metals, which are organized in several tetrahedral units, composed of metal($^{2+}$)-cys units. Finally, MTs contain almost no regular secondary structure elements, with the majority of residues presented as random coil.

The ability to bind metals allows metallothionein to participate in several functional roles in the cell. It can maintain homeostasis of essential trace metals while

also segregating toxic heavy metals, such as cadmium and mercury, from the cytosol. In addition, MT can behave as a reservoir for copper and zinc, which can then be donated to other metalloproteins such as transcription factors. However, despite these rather obvious roles, recent studies have demonstrated that the metal-binding characteristics may be neither the primary nor most important role for MTs. Growing in popularity is the idea that MTs can protect against oxidative damage.

Oxygen stress and damage by reactive oxygen species

An unavoidable consequence of aerobic metabolism is the production of reactive oxygen species (ROS), highly reactive molecules that are capable of damaging the integrity of biological macromolecules, including lipids, proteins, and DNA (reviewed in Halliwell and Gutteridge, 1989). All organisms that consume oxygen maintain basal levels of antioxidant defenses to scavenge and detoxify ROS. However, a sudden or massive change in oxygen content can overwhelm these defenses and lead to ROS-induced damage. For example, it has become clear that much of the damage resulting from an ischemic insult occurs during the reperfusion phase when the rush of oxygen re-entering a tissue leads to the rapid generation of ROS such as superoxide ($^{\bullet}O_2^-$) and hydroxyl ($^{\bullet}OH$) radicals as well as hydrogen peroxide (H_2O_2) (Floyd, 1990; Traystman et al., 1991).

Metallothionein an antioxidant

Evidence to support an antioxidant role for MTs has been summarized in several recent reviews (Viarengo et al., 2000, Palmiter, 1998). In general, tissues that show good

resistance to ROS, heavy metals or DNA damaging reagents react to such stresses by increasing the cellular complement of MT (Kling and Olsson, 2000; Campagne et al., 1999; Kondo et al., 1997; Kang, 1999). Circumstancial genetic evidence for the antioxidant role can be extrapolated from several studies. The superoxide radical is a major mediator of reperfusion injuries. It is normally detoxified by the enzyme, superoxide dismutase (SOD). Yeast that lack the gene for SOD are highly sensitive to oxidative stress, but when they are induced to synthesize the copper-binding MT protein, many of the deleterious effects of oxidative stress are lessened. This suggests that MT can act as a substitute for SOD in protecting yeast from oxidative stress (Tamai et al., 1993). Mammalian cells that over-express MT are also protected against the damaging effects of hydrogen peroxide (Wang et al., 1999), nitric oxide (Schwarz et al., 1995; Montoliu et al., 2000) and many other electrophilic agents. The inference of these studies being that MTs make a superb of scavenger of deleterious oxygen radical because their cysteins are readily oxidized. Although yet to be conclusively proven, it is very likely that MT could function in a similar manner *in vivo*. Finally, this protein very likely produces a secondary antioxidant effect by cutting oxidant production off at the source. By complexing transition metals displaying Fenton reactivity (Fe,Cu), MT can reduce the generation of ROS.

L. littorea antioxidant character

As mentioned, ROS damage macromolecules such as lipids and DNA. However, previous biochemical studies with *L. littorea* have shown that foot muscle does not show signs of lipid damage after extensive anoxia and recovery (Pannunzio and Storey, 1998)

indicating that the animal has defenses that leave it well prepared to deal with abrupt changes in tissue oxygen levels within the range encountered during the normal anoxic-aerobic transitions. In other words, ROS have limited deleterious effects on *L. littorea*.

The current chapter presents data supporting the induction of a novel metallothionein in the tissues of *L. littorea* in response to both freezing and anoxia treatments. The putative MT gene, $LLFF_3$ or $LLAF_N$ was isolated via differential screening of the cDNA libraries as outlined in Chapter 2. The present chapter describes experiments used to probe the expression of the gene including northern blot analysis to assess differential gene expression, the fold increase in expression during stress, and the pattern of message accumulation over a time course of stress treatment and recovery. In addition, the clone sequence was analysed using several publically available computer programs to determine the correct ORF, presence of known protein motifs, similarity with known genes and proteins, and possibile secondary structures of the translated amino acid sequence. These data are consistent with the role of the gene product as a metallothionein.

MATERIALS AND METHODS

cDNA library screening, northern blot preparation, probe synthesis and DNA sequencing were performed as in chapter 2. The northern blots were conducted in triplicate, each representing different RNA isolates. The pixel density value of each mRNA band was divided by that of the corresponding control, ribosomal RNA. These ratios were averaged for each time point, and the SEM was calculated for this average value. Then each average was normalized to the control condition (such that control value = 1.0) and the SEM represents the same percentage as for the first average.

Sequences were analyzed using a variety of computer programs in order to determine an identity for the clones, illustrate the open reading frame, translate the nucleotide sequence to amino acids and align the clone sequence with homologous proteins found in public databases. Finally, immunoblotting was attempted in an effort to illustrate corresponding changes in protein content across the stress time courses.

DNA sequencing and analysis

Complete gene sequences were obtained using an automated procedure by BioS&T (Lachine, PQ) or Canadian Molecular Research Services (CMRS, Ottawa ON). Several internet databases and sites were used to analyze the clone sequences. BLAST (Basic Alignment Search Tool, NIH, Bethesda Maryland) and FASTA (Neimes, France) alignment databases were used to attempt gene identification. Because the start codon of a transcript is bracketed by a conserved series of nucleotides called the Kozak sequence (AorGXXATGG) (Kozak, 1987), a program called ATG^{PR} (http://www.hri.co.jp/atgpr/)

was used to calculate the probability that any given ATG triplet in a DNA sequence is a start codon. The software DNAman v. 4.11 (Lynnon Biosoft, Vaudreuil, Quebec) was used to identify open reading frames (ORF), polyadenylation signals, translate the DNA sequence, and align multiple DNA and amino acid sequences.

Western Immunoblotting

Immunoblotting was performed in an attempt to determine if MT protein content increased along with gene expression. Electrophoresis equipment and reagents were mainly obtained from Bio-Rad Laboratories. Other chemicals were purchased from Sigma Chemical Co. (St. Louis, MO) or the equivalent, unless specified. Chemicals were of electrophoretic grade or the equivalent.

Total soluble protein was extracted from foot muscle and hepatopancreas of control, frozen and anoxic snails. Tissue samples were removed from -80°C storage and homogenized (1:10 w/v) in buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, and protease inhibitors (1 μg/ml aprotinin, 1μg/ml leupeptin, 1 μg/ml pepstatin, 0.1 mM phenylmethylsulfonyl fluoride [PMSF]) that were added immediately prior to homogenization. Homogenates were centrifuged at 14,000 rpm for 15 min at 4°C in a Biofuge 15 (Baxter Canlab). The supernatants were collected and quantified for protein concentration using the Bio-Rad protein assay (Bradford, 1976). SDS-polyacrylamide gel electrophoresis (PAGE) was used to separate proteins. Recipes for electrophoresis stock solutions are listed in appendix I. Aliquots of supernatant were mixed with 2x SDS-PAGE sample buffer (1:1, v/v) and boiled 3 min. Equal amounts of protein (20 μg per lane) were loaded into each sample well on a mini-gel. Proteins were separated using

15% gels (acrylamide:bis-acrylamide 29.2:0.8; w/w) using the discontinuous buffer system of Laemmli (1970) and with a Bio-Rad mini-gel apparatus. Separation generally required 1 h at 200 V.

Proteins were then transferred to polyvinylidene difluoride (PVDF) membranes (Immobilon-P Transfer Membrane, Millipore Corp. Bedford, MA) using a wet transfer apparatus (BioRad) at 4°C for 1 hour at constant voltage (100V). The transfer buffer, containing 25 mM Tris (pH 8.5), 192 mM glycine, 0.05% SDS and 20% v/v methanol, was cooled to 4C prior to the transfer.

A commercial antibody made against mouse metallothionein (type II) was purchased from Stressgen (Victoria, BC).. Western immunoblotting was performed as recommended by the supplier. In general, the PVDF membrane was first rinsed with TBST buffer (10 mM Tris, pH 7.5, 150 mM NaCl, 0.05% Tween-20) and then incubated in blocking buffer (TBST with 5% non-fat dry milk) for one hour. The blot was then added to blocking buffer containing the primary antibody at 1:750 dilution and incubated at room temperature for one hour or at 4°C overnight. Following 5 X 5 minute washes in TBST, the blot was then incubated with the secondary antibody (goat and mouse), diluted 1:5000 in blocking solution. The target protein was visualized by the enhanced chemiluminescence assay (ECL) (NEN Life Science Products, Inc. Boston, MA).

RESULTS

The putative snail metallothionein clones ($LLFF_3$, $LLAF_{4.2.1}$ and $LLAF_N$) were isolated independently through differential screening of both the freezing and anoxic libraries. As illustrated in Chapter 2, the sequences of these two clones are identical where they overlap. Therefore, for subsequent work they are referred to as $LLFF_{MET}$ and the longest clone, $LLFF_3$, was used for further characterization.

Northern blot analysis

Northern blots were hybridized with radio-labeled probe synthesized from the digested insert of the $LLFF_{MET}$ plasmid. Blots contained equal amounts of total RNA isolated foot muscle sampled from each of the experimental groups: control, 1, 12, and 24 h stress (either freezing or anoxia), and 24 h recovery after 6 days of freezing stress. Transcript size was estimated by comparison with the RNA standards. The $LLFF_{MET}$ probe hybridized with a band of roughly 1.2 Kb in size. The pattern of stress-induced transcript accumulation was similar for both freezing and anoxia. As illustrated in Figure 3.1a, transcript levels rose quickly in response to stress treatments, with a marked increase in transcript levels within 1 h of stress exposure. Although the ethidium bromide-staining of the corresponding rRNA bands on the same blot showed that more RNA is present for 1 hour of stress, the fold changes in LL_{MET} expression were calculated with respect to the rRNA (for each time point) so changes in $LLFF_{MET}$ transcript levels can be considered stress-specific and not due to variations in RNA loading onto gels (figure 3.1b).

Densitometric analysis (Figure 3.2c) graphically illustrated the changes in transcript levels. $LLFF_{MET}$ transcripts were present in low amounts in control foot muscle but increased 2-3 fold after only one hour of stress exposure. Transcript levels stayed high in muscle through 12 h of stress exposure and remained high in 24 h frozen snails. After 24 h anoxia exposure, however, transcript levels had dropped to near control values. After 6 d of stress (freezing or anoxia) followed by 24 h of recovery $LLFF_{MET}$ transcript levels were reduced from their peak values under either stress but were still markedly higher than control values.

$LLFF_{MET}$ gene transcripts also accumulated in *L. littorea* hepatopancreas during anoxia and freezing exposures. As apparent in figure 3.2, the transcript signal from this tissue was very weak, but changes were still detected. In freeze-treated heptopancreas, $LLFF_{MET}$ accumulation was slightly elevated after 1 hour, (roughly 3-fold) but the maximum increase of 7-fold was seen after 6 days of freezing, a level that was maintained until 12 hours into the 5°C recovery period at. By 24 hours of recovery, $LLFF_{MET}$ transcript declined to roughly 3-fold that of control values.

In contrast, transcript levels rose 2-fold after only one hour of anoxia exposure. This was maintained through 12 hours of anoxia, at which point it fell to control levels after 24 hours anoxia treatment. Transcript levels then remained relatively constant throughtout the rest of the anoxia treatment (up to 6 days) and the aerobic recovery period.

Analysis of DNA and amino acid sequences

$LLFF_{MET}$ is a complete message RNA. The 1304 nt insert contained a polyadenylation signal (AATAAA) at bases 1100-1105, within the 818 nt 3'UTR, and a poly (A) tail starting at base 1293. As described in chapter 2, this clone shows 2 possible open reading frames. The web-based analysis program, ATG^{PR}, was used to identify the most likely ATG codon that would initiate translation. The result of this Kozak analysis (presented in table 3.1) shows that the first ATG codon in the sequence (193-195 nt) represents the point of translation initiation. From this result, it was possible to determine the probable open reading frame and remaining mRNA characteristics. Bases 1-193 represent the 5'UTR, the start and stop codons are located at 193 and 493 nt respectively, creating an ORF of 300 nt or 100AA. These characteristics are highlighted in figure 3.3 which illustrates both the DNA and translated AA sequences.

The translated ORF of $LLFF_{MET}$ has 27 cysteine residues. All but 7 of these fell into Cysteine-X-Cysteine motifs throughout the sequence, where X is any amino acid. These patterns are illustrated in figure 3.3, and are compared to other known metallothioneins in figure 3.4. In addition, molluscan metallothioneins have a conserved C-terminal motif, CxCxxxCTGxxxCxCxxxCxCK, which is also present in the *L. littorea* amino acid sequence and is illustrated in Table 3.2.

Western Blot Analysis

Immunoblotting was attempted using a commercially available antibody made against the mouse metallothionein. Western blots containing total protein from each of

the stressed and recovered time points can be an effective illustration of changes in protein content. However, after multiple trials the results were inconclusive, showing binding with only large molecular weight proteins. As will be described in the discussion, this may be the result of limited homology among MT from different sources.

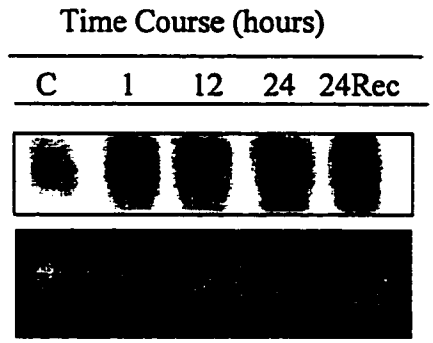
Figure 3.1: Results of northern blot hybridization for clone LL_{MET} against total RNA from the foot muscle of *L. littorea*.

(A) Total RNA from freeze-treated animals (top) and the corresponding ribosomal RNA bands.

(B) As for (A), using total RNA from anoxic animals.

(C) Graphical analysis of A and B, to illustrate transcript levels of LL_{MET} across the time course.

(A) Frozen RNA



(B) Anoxic RNA

(C)

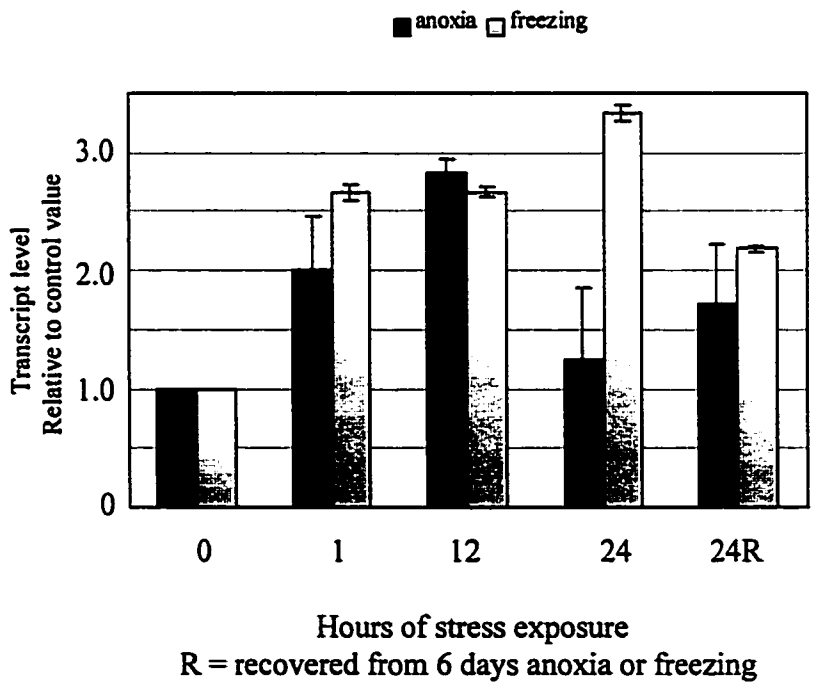
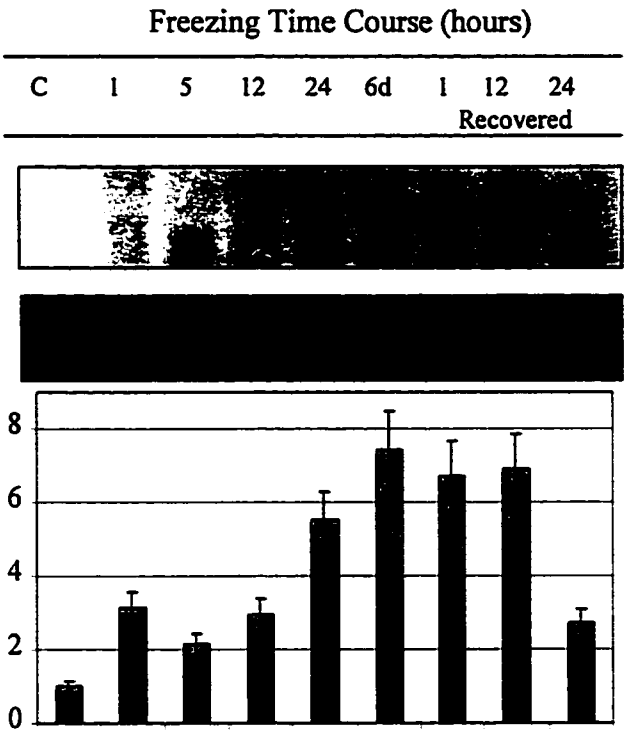


Figure 3.2: Results of northern blot hybridization for clone LL_{MET} against total RNA from the hepatopancreas of *L. littorea*.

(A) LL_{MET} probe (top), corresponding ribosomal RNA bands; graphical representation to illustrate transcript levels of LL_{MET} across the time course.

(B) As for (A), but the total RNA was isolated from anoxic animals.

(A)


Freezing Time Course (hours)
C 1 5 12 24 6d 1 12 24
Recovered
8
6
4
2
0

(B)

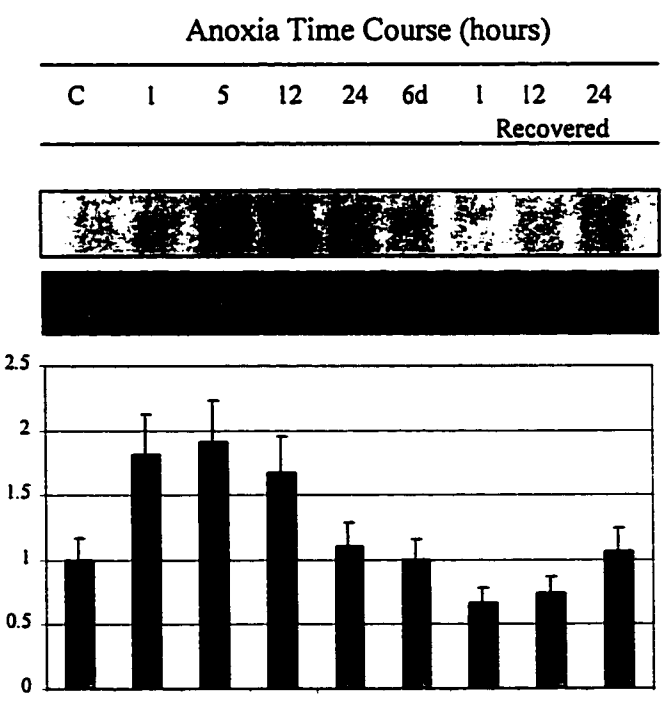

Anoxia Time Course (hours)
C 1 5 12 24 6d 1 12 24
Recovered
2.5
2
1.5
1
0.5
0

Figure 3.3: **LL_{MET} DNA sequence and translated open reading frame.** The amino acid sequence is presented in **bold**; metallothionein motifs are boxed, the polyadenylation signal is in ***bold italics***.

```
  1 GCCCGCAGACACCTTCACGCTGGACTGTTTCGGCTTTTTCAGCGTCGCTTCATAATCCTT   60
 61 TTTCGCCGCCGCCATCAGCGTGTTGTAATCCGCCTCGTGCCGCTCGTGCCGAATTCGGCA  120
121 CGAGGCTGTTTGTCGACTGACGAGTGAACTGTTTTTCTGCTAGCTGTACCTTCCTTTTCA  180

181 AAACCGGTCAAGATGTCTTCAGTTTTCGGAGCAGGATGCACGGACGTGTGCAAGCAGACG  240
  1             M  S  S  V  F  G  A  G  C  T  D  V  C  K  Q  T   16

241 CCATGCGGCTGTGCCACCTCGGGCTGTAACTGCACGGACGACTGCAAGTGTCAGTCATGC  300
 17   P  C  G  C  A  T  S  G  C  N  C  T  D  D  C  K  C  Q  S  C   36

301 AAATACGGAGCGGGTTGCACGGACACATGCAAGCAGACACCATGTGGGTGTGGCAGCGGG  360
 37  K  Y  G  A  G  C  T  D  T  C  K  Q  T  P  C  G  C  G  S  G   56

361 TGCAACTGTAAGGAGGACTGTCGCTGTCAGAGCTGTTCCACCGCCTGCAAGTGTGCGGCT  420
 57  C  N  C  K  E  D  C  R  C  Q  S  C  S  T  A  C  K  C  A  A   76

421 GGAAGCTGCAAGTGCGGCAAGGGATGCACAGGGCCAGACAGCTGCAAGTGTGACCGATCG  480
 77  G  S  C  K  C  G  K  G  C  T  G  P  D  S  C  K  C  D  R  S   96

481 TGCTCCTGCAAATAAACGTCTCCACGCCAAACTGTCCACTTCGTTTAGCCGCCACAATGC  540
 97  C  S  C  K  *                                               100

541 ACACCCAGTAATTTGTCTGTTTAAAGACACTACATTTTCTCATTCCCCATCATTAACTTA  600
601 TTACGAACTTCGTAAATCAAGTCAAAATTCTAATCAGTTTCTGGTAGAATTAATGACCAC  660
661 GGACACAGATATTCACTCACTGATTCACTCACACAGAGTTGAAAGAAGGACAAATGAAAG  720
721 GGAGAATTAAACATGTAATTTATTTGTTGAAAAAAGAAAACAACTTTAGCCATTCCTAAG  780
781 AACATTTTTATAAGAAATTTTGTTCACCATCTCAACCATTTTGAAACATTATTGGTTCAT  840
841 TGTTTCCGTTGTCGCGCATTTCAATATTCATAACGTTCACTCATTATAGAGCCTGAGATT  900
901 ACAACTGTTGTGACGAACTTTTTCATCTTTATTTATCTTTGTGCAAACCACCTTCACGGT  960
961 TTTGTTGTTGTTTTATTTTGTTTGTTTTGTTTATCTCGGATGGCAGTATCAAAATGAAAA 1020
1021 CATGGTATGGATTCGTAAGATGCAGCAAGCCAATTCATATATTCTTTTGATAACATTTGG 1080
1081 AGAACAAAAATAGTTGAGAAATAAAACAGAAGGAAACCTATTGCAGCAATGTACATTCCT 1140
1141 TTGTATTAAAATGGAAAAGTTAAAGCCCTTTTTGTAACAGCAGAAAGCAGTAATATAAAA 1200
1201 TATGGATGTATTTGCTGTATAAGGAATTGGTTTTTCAGATATGTTGTGAGATTTTTCGTT 1260
1261 TGATAATAAGAAAAATATCGTTAATCAGACGATAGAAATACCAAAAAAAAAAA          1304
```

Table 3.1: Identification of putative Kozak sequence in the nucleotide sequence of LL_{MET} using the web-based analysis program, ATG^{PR}.Start codon analysis.

No.of ATG from 5'end	Reliability	Frame	Identity to Kozak rule A/GXXATG G	Start (bp)	Finish (bp)	ORF Length (aa)	Stop codon found?	Sequence (refer to figure 3.2for the intervening AA)
1	0.64	2	AXXATGt	269	568	100	Yes	MSSV....CKCDRSCSCK
8	0.11	1	AXXATGc	613	624	4	Yes	MHTQ
9	0.05	3	GXXATGa	729	803	25	Yes	MTTDTDIHSLIHSHRVERR TNEREN
10	0.05	1	cXXATGa	790	852	21	Yes	MKGRIKHVIYLLKKENNF SHS
14	0.05	3	AXXATGG	1098	1166	23	Yes	MVWIRKMQQANSYILLIT FGEQK

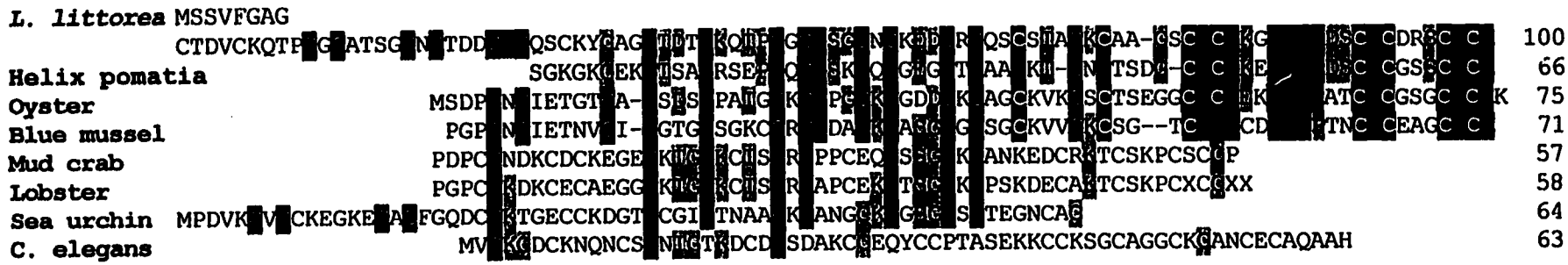
Figure 3.4: Amino acid sequence comparison among LL_{MET} and known metallothioneins from other invertebrate species.

(A) Amino acid sequence alignments

(B) Homology matrix of MT sequences; values present as %.

(C) Relationship among known metallothioneins.

(A)



(B) Homology matrix of MT sequence; values present as %.

	L. littorea	*H. pomatia*	Oyster	Blue mussel	Mud crab	Lobster	Sea urchin	*C. elegans*
L. littorea	100							
H. pomatia	59.1	100						
Oyster	44.7	57.1	100					
Blue mussel	44.3	49.2	57.7	100				
Mud crab	31.6	33.3	32.7	34.5	100			
Lobster	32.8	27.7	30.9	38.2	72.7	100		
Sea urchin	25.9	40.4	36.8	30.9	21.7	17.8	100	
C. elegans	22.3	22	27.1	23.4	29.5	22.7	27.7	100

(C)

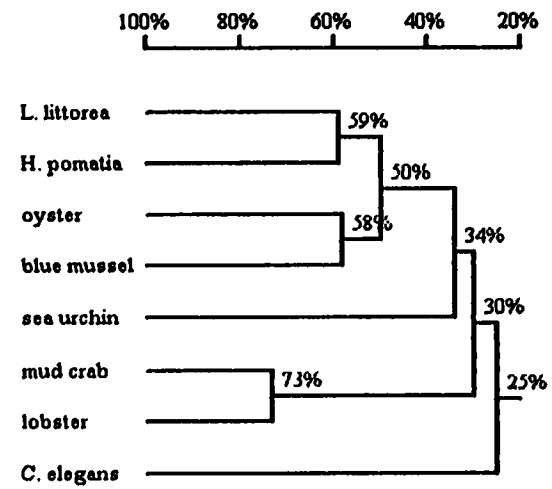


Table 3.2: Sequence homolgy with an molluscan metallothionein motif, CxCxxxCTGxxxCxCxxxCxCK. This sequence is located at the C-terminus of molluscan MTs and is also one of the few highly conserved regions of the protein sequence. Conserved residues are bolded.

Animal	MT motif CxCxxxCTGxxxCxCxxxCxCK
L. littorea	88 **CKCG**KG**CTGP**DS**CKC**DRS**CSCK** 100
Helix pomatia	44 **CKC**GKE**CTGP**DS**CKC**GSS**CSCK** 66
Oyster	53 **CKCG**EK**CTGP**AT**CKC**GSG**CSCK** 74
Blue mussel	49 **CKCG**CD**CTGP**TN**CKC**EAG**CSCK** 71

Table 3.3 **Comparrison of predicted secondary structures for each amino acid of known metallothioneins and LL_{MET}.**

	Predicted secondary structure % of total AA falling into each category		
Source	Helix	Strand	Coil
L. littorea	0	0	100
Helix pomatia	0	0	100
Oyster	0	5.3	94.7
Blue mussel	0	9.8	90.2
Mud crab	0	0	100
Lobster	0	0	100
Sea urchin	0	4.6	95.4
C. elegans	7.9	0	92.1

DISCUSSION

The clone denoted LL_{MET} was isolated from both the freezing and anoxic libraries. On the basis of structural characteristics of the gene and protein sequences shared with MTs from other sources, a possible identification as a metallothionein was established. However, in light of size and amino acid composition differences, care was taken to conduct several comparisons with known MTs to support this putative identity.

Structural comparisons

The nucleotide sequence of LL_{MET} contained two possible ORFs that were roughly equivalent in size. The mechanism of eukaryotic translation allows the prediction of the "real" reading frame. The function of the 5'cap of mRNAs is to enhance the binding of ribosomes to the message. Once bound, the ribosome searches for a start codon, and in 95% of the cases, translation is initiated at the first AUG. Only if the nucleotides around the start codon do not meet the very flexible criteria of the Kozak model does the ribosome ignore the start site. Therefore, the program ATG^{PR} was used to indicate the most likely start site. ATG^{PR} was written in such a way that it takes into account the location of the AUG site in the probability calculation. Hence, as in most cases, the first AUG codon was deemed the site of translation initiation. The translated amino acid sequence obtained from this ORF was submitted to a BLASTp search to attempt identification, since the nucleotide sequence of LL_{MET} was of little usefulness (described in chapter 2). The results indicated a limited identity with the MT family of proteins.

MTs show significant variations in size and amino acid composition across species. Vertebrate MTs are characterized by common features, including a MW of 6-7 KDa, and a total of 61 amino acids, 20 of which are cysteine residues and none are aromatic. Metallothioneins have also been identified in a long list of marine invertebrates, including bivalves, gastropods and echinoderms. However, invertebrate MTs are more variable in terms of MW and amino acid composition, and they often contain a few aromatic residues. In addition, molluscan MTs have a wide tissue distribution, being found in digestive gland (George and Pirie, 1979), mantle (Carpene et al., 1980), gill (Carpene and George, 1981) and adductor mussel (Carpene et al, 1983), whereas the corresponding MT isoforms in mammalian systems are primarily restricted to the liver and kidney.

The coding region of LL_{MET} is substantially longer than other characterized MTs. Coding for 100 amino acids, it is 25-38 residues longer than MT's from mussel (71), the terrestrial snail *Helix pomatia* (66), oyster (75), *C. elegans* (63), human (61), and pigeon (74). A comparison of the primary amino acid sequence of LL_{MET} with those previously reported reveals strongest similarity with MT from another gastropod, *H. pomatia,* at 59.1%. In addition, MTs from 2 bivalve species, oysters and *M. edulis*, share 40% of their residue composition with *L. littorea*, while the crustacean species, mud crab and lobster, are only 21-26% identical.

All molluscan MTs identified thus far contain a conserved C-terminal sequence that must be present in a putative MT protein for it to be classified as a member of the molluscan subclass of metallothioneins. As illustrated in Table 3.2, this sequence was also found in LL_{MET}, further supporting the similarity with molluscan MTs. In addition,

Dallinger et al. (1993) described a gastropod MT with a unique N-terminal region showing similarity with a segment of histones H2A and H4, which he suggested could indicate DNA stabilizing function. In light of this unique motif, a gastropod family was added to the MT classification system, with the histone motif as the required component. LL_{MET} does not have this sequence, however, in light of the fact that there are only five MT protein sequences from two gastropod species GenBank, it would appear that this generalization is not valid. Clearly, a classification based on only two species leaves room for modification.

As a family, MTs are structurally and functionally similar, but are notoriously variable in nucleotide sequence and amino acid composition. Therefore, comparison of protein structural elements can be a stronger supporter of identity. The most consistent characteristic of metallothioneins is the presence of a large number of cysteine residues that are found in the repeating motif, Cys-X-Cys. Figure 3.3 illustrates that, although the compared sequences varied in length, where they overlap, the Cys-X-Cys motifs align either exactly or very closely among the species. Such highly conserved, protein-specific areas strongly support the conclusion that LL_{MET} is a metallothionein. It is also interesting to note that LL_{MET} contains 27 cysteine residues, 4 to 7 more than the afore-mentioned species. Perhaps the additional cysteines allow for a stronger antioxidant activity, with the addition of extra thiol-containing cysteine residues.

In terms of higher-order structures, all MTs that have been studied using NMR and X-ray crystallography are similar. Homonuclear H1 NMR results from the sea urchin MT show that it is composed of 2 globular domains: an N-terminal 4-metal domain (residues 1 to 36) and a C-terminal 3-metal domain composed of residues 37-65

(Riek et al., 1999). The protein binds 7 equivalents of divalent metal ions, and all cysteines are in the reduced form, and co-ordinate with metals in a tetrahedral geometry (Kaji and Kojima, 1987). The secondary structure consists primarily of random coil, whereas it is the complexing of metals with cysteines that provides a scaffold and gives the tertiary structure its characteristic dumbbell shape. In keeping with this observation, the amino acids of LL_{MET} were predicted to form coiled-coils (Figure 3.5). However, every MT has a unique connectivity pattern of the Cys-metal coordination bonds, creating unique local folds in the polypeptide backbone. These differences are related to variations of the cysteine positions and again emphasize the special role of the cysteine residues in defining the structure of MTs.

METAL BINDING

A unique biological feature of MT genes is their inducibility by a variety of agents and conditions. Biosynthesis of the protein is enhanced by cytokines, tumour promoters, hormones, growth factors, as well as by environmental stressors such as hypoxia and psychological stress. Such a wide range of inducers leaves the specific cellular roles for MT unclear. Unanimous support is only found for an essential role in cellular metal homeostasis, because exposure to selected heavy metals (Cd, Cu, Hg, Zn) rapidly induces MT mRNA. This leads to increased biosynthesis of MT, which can then bind excess metals, storing essential heavy metals and detoxifying others. In line with this observation, Huang (1993) proposed that MT plays a rescue function, acting as a donor of essential metals to structures previously compromised by metal binding.

The rate of MT synthesis and degradation continually changes in response to the environmental and physiological state of the animal. Turnover is related to transcriptional control of MT mRNA synthesis as well as protein degradation in the lysosomal system. The half-life of the MT protein varies from hours to days in mammals, while in a cadmium-treated mussel, Bebiano and Langston (1993) calculated a half-life of 25 days. As older MTs degrade, the released cadmium induces further synthesis of the protein, to which the metal becomes re-sequestered. The slow MT turnover rates (compared with mammals) and the lack of significant cadmium excretion shows the relatively stable nature of the cadmium-metallothionein complex in these invertebrates and a continual synthesis of protein in response to a chronic stress.

Despite the large number of MTs identified in invertebrates, this chapter presents the only data I am aware of that shows enhanced gene expression of MT in response to natural environmental stressors in a tolerant animal. The northern blot analysis showed that there was a three-fold increase in MT mRNA levels in response to both anoxia and freezing. While this change is not particularly dramatic, it is of similar magnitude to other stress-induced genes that we have seen in the lab. In addition, the net amount of new protein synthesized from the transcript will also depend on the stability of the mRNA transcript, not just the total amount.

The induction response was essentially the same in both stressors (anoxia and freezing), indicating a common mechanism or trigger that is responsible for activating the transcription of the MT gene in response to either stress. One possibility for such a common trigger could be low oxygen. By definition, anoxia is a low oxygen stress, but freezing also results in limited oxygen availability to tissues. This is because extracellular

ice formation imposes ischemia on organs due to the freezing of hemolymph plasma. Although snails have an open circulatory system, which pulses hemolymph rather than blood to bathe the tissues, ice still prevents movement of fluids surrounding tissue.

Reasons for upregulating MT

Specific cellular functions for MTs are unclear in all species under normal circumstances, so conclusions with respect to its function in stress tolerant animals are speculative. However, considering the potential effects these putative functions could have on a hypoxic animal, the speculations are highly appealing. The most likely role is one of an antioxidant, as discussed below. While the evidence in the literature is largely circumstantial, the induction of MT expression in a stress tolerant animal that is known to require substantial antioxidant defenses and shows limited damage in situations where reactive oxygen species (ROS) generation should be high, may be seen as strong anecdotal support for such a role.

Antioxidant properties of metallothioneins

MT has been shown to scavenge oxygen radicals in cell-free systems (Kennedy et al., 1993) and while such an effect has not been shown in vivo, there is ample evidence to support an antioxidant role for MT. Using transgenic mice, it is possible to determine the effect of a single genetic change on the "normal" response to stress. While oxidative stress has been shown to contribute to ischemia-induced cell death in the mammalian brain (Chan, 1996), Campagne et al. (1999) demonstrated that transgenic mice over-expressing the MT-1 isoform are protected against focal cerebral ischemia and reperfusion. In their study, the mRNA levels for MT1 increased 7.5 fold over baseline

values upon application of ischemia, and while tissues were still damaged, the sizes of the affected regions were approximately 40% smaller. In a reverse study, Lazo et al. (1995) showed an enhanced sensitivity to oxidative stress in cultured embryonic cells that lack genes for MT-I and II.

A cold-induced MT response in the whole animal has been reported in the literature, though not in a stress tolerant species. Beattie et al. (1996) found MT was induced in response to cold in rats that were transferred from 26 to 6 degrees. Following an increase in mRNA content, the MT protein itself increased 1.4-3 fold in kidney and liver, while the thermogenic organ, brown adipose tissue, showed a 16-fold increase in MT protein. Thermogenesis is a process that requires high oxygen consumption, which then causes a rise in the amount of ROS during heat production. The researchers therefore concluded that MT was produced to supply antioxidant protection. In *L. littorea*, LL_{MET} was induced after 1 hour of anoxia and 12 hours of freezing, conditions where oxygen consumption is very low. This may suggest a "preemptive strike" against the inevitable oxygen influx and production of ROS during recovery.

Other supporting evidence comes from studies of the effects of iron and cadmium on intertidal mussels, *Mytilus galloprovincialis*. Iron stimulates ROS formation due to its role in Fenton reactions but cadmium does not. Thus, cadmium exposure increased MT content in mussel tissues without changing reduced glutathione levels or antioxidant enzyme activities. However, prior exposure to cadmium greatly increased the survival rates of mussels when subsequently exposed to iron in an anoxic environment. This protective effect was attributed to the cadmium induction of MT and its presumed oxyradical scavenging effect (Viarengo et al., 1999).

Hemocyanin

Hemocyanin, the copper-based, oxygen-carrying protein of most invertebrates shows expression in response to low oxygen conditions. Laboratory studies with lobster and shrimp have shown that environmental hypoxia induces higher levels of hemocyanin (Haegerman, 1968; Baden et al., 1990). The blue crab shows a similar response, both to hypoxia and hyposalinity (deFur et al., 1990; Mason et al., 1983). Given that MT is also induced under such low oxygen conditions, these results suggest a link between the biosynthesis of hemocyanin, copper metabolism and MT. Indeed, Brouwer and Brouwer (1998) described such a connection in the blue crab. In light of the expression of MT in *L. littorea* in response to hypoxia and freezing, it would be very interesting to determine if LL_{MET} can bind copper ions, thereby indicating a function as a reservoir or cache of copper for transfer to hemocyanin.

Heavy metals

Detoxification of heavy metals was the first proposed role for MTs. In this study, LL_{MET} mRNA levels were enhanced more in response to freezing and for a longer period of time. Because ice excludes all solutes from its crystal lattice, freezing causes an increase in concentrations of all solutes in both intra- and extracellular body fluids; conversion of 65% of body water to ice (which is survivable by *L. littorea*) must necessarily triple the concentration of all solutes in the remaining unfrozen fraction. This includes any metals present. Since elevated metal concentrations are a commonly accepted inducer of MT gene expression, freeze-induced increases in MT expression may come from two sources: metal-stimulated induction and ischemia-stimulated induction,

perhaps explaining the more pronounced induction of LL_{MET} mRNA during freezing versus anoxia.

FUTURE DIRECTIONS

After extensive work has been carried out to show differential expression of a gene, the ideal case is to also be able to show that the increased message RNA accumulation is followed by increased protein production and accumulation of the protein product. While there are many ways to accomplish this, one of the best techniques to use is immunoblotting, especially for a protein that lacks a measurable enzymatic activity. Unfortunately, attempts to detect MT in the tissues of *L. littorea* using a commercial antibody against human MT were not successful, despite repeated titrations of the primary and secondary antibodies in an effort to optimize the binding. In light of the minimal MT sequence identity across species, this result was disappointing but not surprising. Antibodies generally recognize an epitope (binding site) consisting of only a few amino acids. Although studies with the blue crab have shown that stress (usually heavy metal exposure) quickly induces mRNA accumulation which is followed by de novo MT synthesis, there needs to be illustrative proof of this in *L. littorea*, rather than merely extrapolating a conclusion. Therefore, several proposals for further work can be considered and are presented below.

Despite the failed western analysis, a specific antibody can elegantly illustrate changes in protein content. In light of the specificity problem outlined above, it is clear that a more specific antibody is required. With pure LL_{MET} protein, antibodies can be synthesized. The protein itself could be obtained by cloning the DNA sequence into an

expression vector and overexpressing the protein in bacteria or insect cells, or by purifying MT directly from *L. littorea* tissue. Purification of MTs from marine invertebrates was attempted by Dallinger et al. (2000) with limited success. However, their isolation protocols are very straightforward and protein purification from *L. littorea* tissue source may be the faster road to a pure protein. In either case, an N-terminal sequence or amino acid composition of the isolated protein should be completed to ensure it is consistent with that deduced from the LL_{MET} clone.

In addition to immunoblotting, the MT content could be measured directly in tissue homogenates. The original protocol dates back to 1958 but there have been many modifications outlined in Isani et al. (2000). In general, the procedure is based on ethanol/chloroform fractionation of the cytosol followed by spectophotometric measurement of SH content.

Finally, after demonstrating a stress-induced increase in protein content, the novel MT should be characterized in terms of its metal binding character. Many species express multiple MTs, each with different metal-binding characteristics. For example, the blue crab has two Cd inducible MT isoforms (CdMT-I and CdMTII) and a third isoform that is induced by copper, but not Cd. In general, copper-binding MTs do not bind Zn or Cd. Whether the induced *L. littorea* MT is a copper or a Zn/Cd isoform has important implications for the function of the MT and purpose for its induction during anoxia or freezing stresses.

CHAPTER 4

FREEZING AND ANOXIA INDUCE EXPRESSION OF NOVEL GENES CONTAINING SECRETORY SIGNALS IN THE FOOT MUSCLE OF *LITTORINA LITTOREA*

INTRODUCTION

With the advent of high throughput sequencing techniques in the 1990's, a great push developed towards sequencing entire genomes. Although the "human genome project" is the most famous of these ventures, there are currently 600 organisms listed in the NCBI database, *Entrez Genomes,* with the data representing both completely sequenced organisms and those for which sequencing and / or assembly are in progress. All three main domains of life - bacteria, archaea, and eukaryota - are represented, as well as viruses and mitochondria. With this explosion of DNA sequence information it is essential to have methods for reliable alignments and information retrieval. The previous chapters have illustrated the ease and usefulness of such databases since a quick search with a DNA alignment tool such as BLAST, can either assign a clone an identity (as in Chapter 2) or provide enough clues to give a weakly homologous gene a putative identity (Chapter 3).

Unfortunately, with respect to truly novel gene sequences, alignment programs lose their effectiveness. Even though efforts to sequence entire genomes have added greatly to the genetic knowledge base, they also demonstrate the limits of our functional interpretations. The shear number of genes (or ESTs, expressed sequence tags) orphaned without an identity illustrates how much we *do not* know. Hence, when searching BLAST with a novel cDNA the only result that can be provided is the lack of information (i.e. no identity).

Analysis of DNA itself presents an intrinsic problem. The genetic code is startlingly simple, composed of a 4 letter alphabet that can be combined and rearranged

to form genes. However, whereas the 4 letters (bases) are distinct molecules, they are chemically similar and in and of themselves provide no information about the gene product. By translating the DNA to the 20-letter amino acid alphabet, the resulting molecule has a great deal more diversity in terms of physical and chemical properties. Factors such as whether an amino acid is basic or acidic, hydrophobic or hydrophilic, and the structure of side chains all determine how closely and in what manner the amino acids interact.

All protein sequences, whether determined directly or via translation of an open reading frame in a nucleotide sequence, contain intrinsic information that is valuable in determining their structure or function. While direct structural evidence can be derived from experiments such as CD spectroscopy, X-ray crystallography and NMR, these are specialized techniques that demand a great deal of highly technical work by well-trained personnel. As a result, the rate of data accumulation with respect to protein structure cannot keep pace with the relatively simple elucidation of nucleotide sequences. The gap between protein sequence and protein structure data is evident by comparing the content of their respective databases. As of July 2000, there were 525,722 sequences entries in the non-redundant protein sequence database, but only 12,771 structures in PDB (Protein Data Bank). To help narrow this gap, a variety of predictive programs have been developed by pooling structural data from proteins of known structures to develop guidelines to predict amino acid behaviour in a given situation. This then allows the prediction of structural and functional information for a completely novel amino acid sequence in the absence of biochemical data.

This chapter presents two novel cDNA sequences that code for unique proteins. The clones differ greatly in their responses to anoxia and freezing, but due to their very unique sequences, they required extensive computer assisted analysis to attempt to supply a putative function. Such analysis was necessary in order to propose rational future experiments aimed at discovering the function of these novel proteins.

MATERIAL AND METHODS

cDNA library screening, northern blot hybridization and analysis, and DNA sequencing of the clones were performed as described in Chapters 2 and 3. Clone $LLAF_W$ showed a transcript size significantly larger than the clone itself, implying the clone was an incomplete message. 5'RACE, Rapid Amplification of cDNA Ends, a method for obtaining full-length cDNA by PCR (Zhang and Frohman, 1997) was carried out to isolate a longer message.

Isolation of a full-length cDNA using RACE

A commercial kit from Gibco BRL (*5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0*) was used to isolate the missing sequence of $LLAF_W$. The manufacturer's protocol was followed exactly, and is available on the world wide web, at:

http://www.lifetech.com/Content/Tech-Online/molecular_biology/manuals_pps/18374058.pdf

The primers used for this procedure were synthesized at BioS&T (Lachine, Quebec) and designed using Primer designer.

Computer-assisted analysis

Both clones discussed in this chapter are unique, showing no similarity with other DNA sequences currently contained in public databases. Fortunately, there are a variety of predictive programs available on the world wide web that allow analysis of amino acid

sequences to elucidate protein characteristics such as motifs, sites for post-translational modification, and signal peptides. While not absolutely conclusive, the suggested presence of such characters can lend insight into the function of novel proteins.

Protein characteristics

Common protein characteristics such as the isoelectric point (pI), molecular weight (MW), hydrophobicity, and amino acid composition, were determined using the commercial program, DNAman (Lynnon Biosoft, Vaudrieul, PQ). The translation initiation site was identified by the presence of a Kozak sequence, which was determined using ATG^{PR}, (http://www.hri.co.jp/atgpr) as described in Chapter 3.

Identification of secretory signal peptides in proteins

SignalP is a useful tool for the detection of N-terminal secretory signal peptides and their cleavage sites in an amino acid sequence. The program makes use of "artificial neural networks", a computational process that is particularly useful where the problem in question requires the involvement of patterns and trends (see Appendix V for a simple description). Such programs compare the input sequence to a "training set" of data, which is composed of known signal sequences. SignalP employs 2 neural networks: one to distinguish between signal peptides and non-signal peptides, the other to recognize the cleavage site.

The SignalP WWW server returns three scores between 0 and 1 for each amino acid in the sequence for each of three predictions. The raw cleavage site score (or C) recognizes cleavage sites vs. other sequence positions. The signal peptide score (or S)

recognizes signal peptides as opposed to non-signal-peptides. Finally, a combined cleavage site score (the Y-score) predicts the cleavage site location and is optimized by observing where the C-score is high and the S-score changes from a high to a low value. For a typical signal peptide, the C- and Y-scores will be high at position +1 to the cleavage site, while the S-score will be high before the cleavage site and low thereafter. For a typical non-secretory protein, all three scores are very low throughout the sequence.

Sub-cellular localization

PSORT II, developed by Kenta Nakai at the Human Genome Center, Tokyo, Japan, was used to identify the presence of a range of motifs that indicate where the newly translated protein may be sent. The program takes into account both the amino acid sequence as well as its source (bacterial or eukaryotic), then analyses it by applying "rules" for various sequence features of known protein sorting signals. A range of known signal sequences are incorporated into this analysis, including those that direct proteins to the nucleus, mitochondria, or Golgi apparatus for sorting (a complete list can be found in Appendix VI). Finally, the program collates the results and reports the probability for the protein to be localized at a given subcellular site.

Secondary structure predictions for the amino acid sequence

Protein structure was predicted from the amino acid sequence using the Protein Structure Analysis (PSA) web server. This system determines the probable placement of secondary structural elements along the sequence. The analysis algorithm is based on probabilistic Discrete State-space Models (DSMs) and optimal filtering and smoothing

algorithms as described in Stultz et al. (1993). Based on this analysis, patterns of alpha-helices, strands, tight turns and loops are assigned to the sequence. In addition, the results of the secondary structure prediction allows for the classification of the putative protein sequence into a tertiary structure class, which makes it possible to hypothesize the function of an unknown protein. Appendix VII lists the 4 "superclasses" that denote the four major folding classes of globular single-domain proteins, as well as the multiple sub-classes that further detail the predicted structure.

SAPS

Statistical analysis of protein sequences (SAPS) employs statistical data, gathered from identified protein sequences, to evaluate novel amino acid sequences for the presence of various functional and structural characteristics. The output is organized into the following sections:

- compositional analysis,
- charge distributional analysis -- charge clusters; high scoring (un)charged segments; charge runs and patterns
- distribution of other amino acid types -- high scoring hydrophobic and transmembrane segments, cysteine spacings
- repetitive structures -- in the amino acid alphabet and in an 11-letter reduced alphabet
- multiplets -- counts, spacings, and clusters in the amino acid and charge alphabets,

- periodicity analysis
- spacing analysis.

RESULTS

SECTION 4.1: CLONE LL_{GRP}

As described in Chapter 2, differential screening revealed two clones, one isolated from the anoxic library and the other from the freezing library, with nucleotide sequences that were identical where they overlapped. Compared to the 1092nt $LLAF_{EE}$, the $LLFF_{13}$ segment was considerably shorter at 324nt. Accordingly, these 2 clones were paired and denoted LL_{GRP}, *Littorina littorea* glutamine-rich protein, and only the longest gene sequence ($LLAF_{EE}$) was used for further analysis and discussion.

Northern blot analysis

The probe for LL_{GRP} consistently bound to a single transcript at approximately 1.2 Kb. Among all the isolated clones, LL_{GRP} offered the most dramatic alteration of gene expression in the foot muscle of *L. littorea*, as illustrated in Figure 4.1.1. In the control state the transcript was barely evident, making "fold increase" a dubious term. Whereas the transcript accumulated in response to both freezing and anoxia, the time scale of change was different between the two stresses. After 1 hour of freezing exposure there was little or no message present, but after 12 hours, the transcript showed a dramatic increase to roughly 4-fold the level in controls. Anoxia, on the other hand, induced a similar increase in mRNA transcripts, but required only one hour of treatment to do so. For both stresses, the message accumulation was sustained throughout the time course, up to at least 24 hours of recovery, the longest recovery period available.

The hepatopancreas also showed up-regulation of this transcript in response to anoxia and freezing, although to a more limited degree (Figure 4.1.2). Transcripts levels increased 2-fold over control values after one hour of freezing and remained relatively stable throughout the freezing treatment. After one hour of recovery, the transcript level increased somewhat further to a maximum of 2.5-fold that of control but after 24 hours of recovery levels had declined to roughly equal to control. In response to anoxia LL_{GRP} transcripts increased 2-2.5 fold in hepatopancreas after 1 hour anoxia; this was maintained through 12 hours, after which transcript levels decreased for the duration the anoxia exposure and throughout the aerobic recovery period.

Sequence analysis: DNA and amino acids

The DNA and translated amino acid sequences for LL_{GRP} are illustrated in Figure 4.1.3. Most of the characteristic mRNA motifs were discussed in Chapter 2, but in addition to these, it is interesting to note that the 3'UTR is very short (172 nt before the start of the polyA tail) and the 5'UTR is ambiguous. Unlike most complete genes this sequence has a fairly short 5'UTR and no stop codon was present before the putative ORF. Despite this, the DNA sequence was assumed to be near-complete, because the length of 1152 nt was very close to that of the transcript size, which was estimated to be 1.2 Kb after comparison with RNA standards on the northern blot. Additional support for this reading frame comes from analysis of the DNA sequence with ATG^{PR}, the results of which indicate that the first ATG codon is the most likely translation initiation site because it best fits the Kozak consensus sequence (Table 4.1.1).

Compositional analysis of clone LL_{GRP}.

Translation of the putative ORF produced a 322 amino acid sequence representing a 33.9 KDa protein. The amino acid composition itself, listed in Figure 4.1.4a, is quite unusual. Altogether, glutamine (Q), glycine (G) and proline (P), compose 85% of the amino acids. SAPS analysis provided additional information by indicating whether a given amino acid was over- or under-represented, and to what degree, in comparison with the training data. In accordance with the high percentage composition listed above, glutamine and glycine are in the positive 1% quantile (i.e. present as frequently as the highest X-containing proteins in the reference set) while 13 other AA are highly under-represented in the sequence.

The putative protein is predominantly neutral; 95% of its residues are uncharged, only 2 are negatively charged and 12 carry a positive charge (Figure 4.1.4b). As a consequence, the predicted pI is a very basic 10.4. SAPS also provides an illustration of charge patterns throughout the sequence because this can be an indication of amphipathic secondary structures. However, LL_{GRP} showed no positive, negative or mixed charge clusters. The distribution of hydrophobic amino acids also showed no significant segments greater than 15 residues but it did pick out a single transmembrane segment, from bases 4-18 (Fig. 4.1.4b).

The abundant G, P, and Q amino acids are present in a sequence, GGQQQQY(G or D)RPQQQGGQQQ(G or Q), that is repeated 11 times between residues 24-287. The residues involved in these repeats are highlighted with blue text in Figure 4.1.4b. Nine are immediately obvious but near the C-terminus, two repeats overlap with 2 others. These final 2 repeats (238-255 and 269-287) are each encapsulated in a

box. At the protein's C-terminus, a shorter but similar sequence, QQFGGQQQGG, is found twice at residues 298-307 and 308-317 (Figure 4.1.4b, red text).

In contrast to the repeating nature throughout the bulk of the sequence, the first 16 residues at the N terminus shows much more variability, entirely lacking the prominent amino acids Q, P, and G. Analysis with SignalP (figure 4.1.5) indicated that the first 16 amino acids, MWKIVTLTLLATLAAA, compose a probable N-terminal secretory signal peptide. All three predictive calculations used by the SignalP server reached this conclusion, thus conveying a high degree of confidence in the result. This conclusion is further supported by the protein's hydropathic profile, which shows a strongly hydrophobic N-terminus, whereas the remainder of the sequence is consistently hydrophilic. Finally, the only transmembrane segment identified in the sequence (described above) is also encompassed with this signal peptide.

Secondary structure prediction

The secondary structure of the putative protein encoded by LL_{GRP} was determined with the www-based analysis program PSA and the results are illustrated graphically in Figure 4.1.6. LL_{GRP} falls into the beta class of proteins (Fig. 4.1.6 a), a classification that is further subdivided into 12 groups, indicated on the X-axis of Fig. 4.1.6 b. PSA prediction placed 100% probability that LL_{GRP} is a porin, a group of non-hydrophobic membrane proteins that contain a series of beta strands arranged in a barrel motif. As illustrated in Figure 4.1.6 c, the sequence shows a repeating pattern, with alternating strand and loop motifs, interspersed with 4-amino acid tight turns, roughly corresponding to the location of proline residues in the primary sequence.

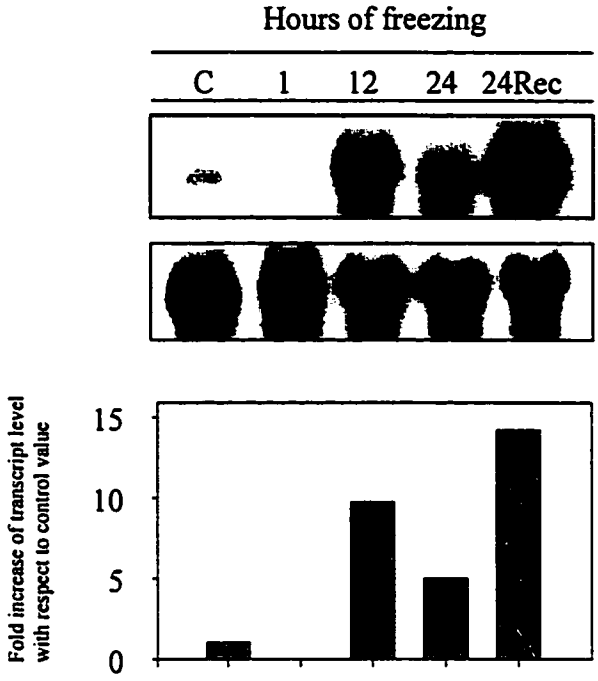
Figure 4.1.1: Northern hybridization of clone LL_{GRP} against total RNA from the foot muscle of freezing and anoxia treated animals.

Figures A and B illustrate (from top to bottom): northern blot hybridized with LL_{GRP} probe, control binding of the gene for ribosomal RNA (18S subunit), and an illustration of the percentage of clone LL_{GRP} pixel density with respect to the corresponding rRNA band.

(A) RNA from freeze treated foot muscle

(B) RNA from anoxia treated foot muscle

(A)


Hours of freezing
C
1
12
24
24Rec
Fold increase of transcript level with respect to control value
15
10
5
0

(B)

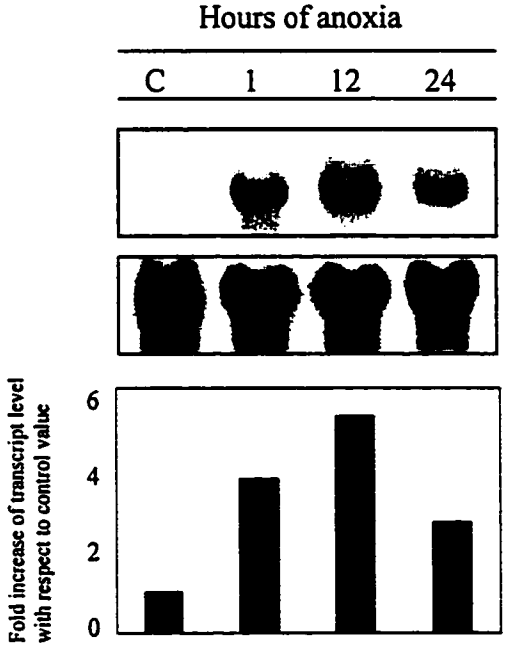

Hours of anoxia
C
1
12
24
Fold increase of transcript level with respect to control value
6
4
2
0

Figure 4.1.2: Northern hybridization of clone LL_{GRP} against total RNA from the hepatopancreas of freezing and anoxia treated animals.

(A) Freezing time course

(B) Anoxic time course

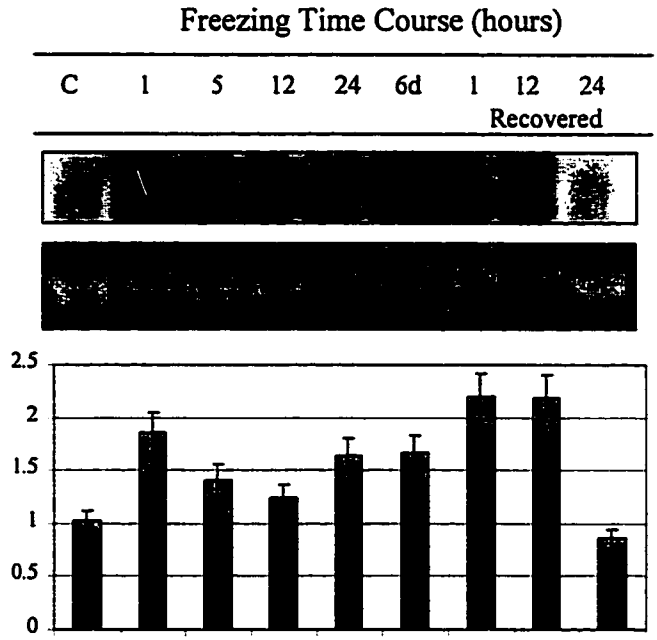
Both A and B illustrate:

(i) Binding of LL_{GRP} to its complementary message in the total RNA on a northern blot.

(ii) Ethidium bromide-stained rRNA band from the formaldehyde gel

(iii) Illustration of the percentage of clone LL_{GRP} pixel density with respect to the corresponding rRNA band.

(A)


Freezing Time Course (hours)
C
1
5
12
24
6d
1
12
24
Recovered
2.5
2
1.5
1
0.5
0

(B)

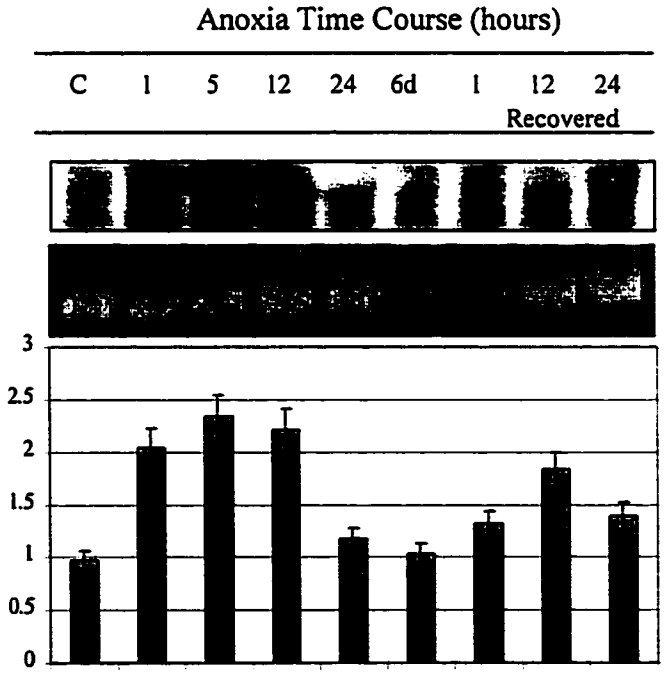

Anoxia Time Course (hours)
C
1
5
12
24
6d
1
12
24
Recovered
3
2.5
2
1.5
1
0.5
0

Table 4.1.1: Identification of putative Kozak sequences in nucleotide sequence of LL_{GRP} using the web-based analysis program, ATP^{PR} (see Appendix IV for a description of the program and reference).

The most likely site of translation initiation is the ATG codon of the longest ORF identified in the gene sequence.

No. of ATG from 5'end	Reliability	Frame	Identity to Kozak rule A/GXXATGG	Start (bp)	Finish (bp)	ORF Length (aa)	Stop codon found?	Sequence (refer to figure 4.1.3 for the intervening AA)
1	0.75	2	aXXATGt	38	1003	322	Yes	MWKIV...QQFGF
2	0.04	3	aXXATGt	1047	1094	16	Yes	MSNSFVADSIIPASPI
3	0.04	2	aXXATGa	1094	1138	15	Yes	MIFLFWQKLIIYMSN
4	0.04	2	tXXATGG	1130	1138	3	Yes	MSM

Figure 4.1.3: Nucleotide and amino acid sequences for clone LL_{GRP}, a novel cDNA clone isolated independently via differential screening of freezing and anoxic cDNA libraries.

LEGEND

Secretory signal peptide (box).	XXXXX
Nucleotides	*italics*
Amino acids	black text

```
-37   CAACATCGTCTTGCCTAGGTTAAACAAACCTGCCAAC

38    ATGTGGAAGATCGTGACCTTGACCCTTCTCGCAACTCTAGCAGCTGCTCAGTTCGGCGGT
1     M  W  K  I  V  T  L  T  L  L  A  T  L  A  A  A  Q  F  G  G

61    CAGCAGGAAGGCGGTCAACAGCAGTACGGCAGACCACAGCAACAGGGCGGTCAGCAACAA
21     Q  Q  E  G  G  Q  Q  Q  Y  G  R  P  Q  Q  Q  G  G  Q  Q  Q

121   GGCGGTCAGCAACAGGGCGGTCAACAGCAGTACGGCAGACCACAGCAACAAGGCGGTCAG
41     G  G  Q  Q  Q  G  G  Q  Q  Q  Y  G  R  P  Q  Q  Q  G  G  Q

181   CAACAGGGCGGTCAGCAACAGGGCGGACCAGGCGGTCAGCAACAGGGTGGACCAGGCGGT
61     Q  Q  G  G  Q  Q  Q  G  G  P  G  G  Q  Q  Q  G  G  P  G  G

241   CAGCAACAGCAGTACGGCAGACCACAGCAACAAGGCGGTCAGCAACAGGGCGGTCAGCAA
81     Q  Q  Q  Q  Y  G  R  P  Q  Q  Q  G  G  Q  Q  Q  G  G  Q  Q

301   CAGGGCGGACCAGGCGGTCAGCAACAGGGTGGACCAGGCGGTCAGCAACAGCAGTACGGC
101    Q  G  G  P  G  G  Q  Q  Q  G  G  P  G  G  Q  Q  Q  Q  Y  G

361   AGACCACAGCAACAAGGCGGTCAGCAACAGGGCGGTCAGCAACAGGGCGGTCCAGGCGGT
121    R  P  Q  Q  Q  G  G  Q  Q  Q  G  G  Q  Q  Q  G  G  P  G  G

421   CAGCAACAGCAGTACGGCAGACCACAGCAACAAGGCGGTCAGCAACAAAGCGGTCAGCAA
141    Q  Q  Q  Q  Y  G  R  P  Q  Q  Q  G  G  Q  Q  Q  S  G  Q  Q

481   CAGGGCGGTCCAGGCGGTCAGCAACAGCAGTACGGCAGACCACAGCAACAAGGCGGCCAG
161    Q  G  G  P  G  G  Q  Q  Q  Q  Y  G  R  P  Q  Q  Q  G  G  Q

541   CAACAGGGTGGACCAGGCGGTCAGCAACAGGGTGGACCAGGAGGTCAGCAACAAGGCGGT
181    Q  Q  G  G  P  G  G  Q  Q  Q  G  G  P  G  G  Q  Q  Q  G  G

601   CAGCAACAGGGCGGTCAGCAACAGCAGTACGGCAGACCACAGCAACAGGGCGGCCAGCAA
201    Q  Q  Q  G  G  Q  Q  Q  Q  Y  G  R  P  Q  Q  Q  G  G  Q  Q

661   CAGGGCGGACCAGGAGGTCAGCAACAGCAGTACGGCAGACCACAGCAACAGGGCGGTCAG
221    Q  G  G  P  G  G  Q  Q  Q  Q  Y  G  R  P  Q  Q  Q  G  G  Q

721   CAACAGCAGTACGGCAGACCACAGCAACAGGGCGGTCAGCAACAGGGCGGTCAACAACAG
241    Q  Q  Q  Y  G  R  P  Q  Q  Q  G  G  Q  Q  Q  G  G  Q  Q  Q

781   CAGTACGACAGACCACAGCAACAGGGCGGTCAGCAACAGCAGTACGGCAGACCACAGCAA
261    Q  Y  D  R  P  Q  Q  Q  G  G  Q  Q  Q  Q  Y  G  R  P  Q  Q

841   CAGGGCGGTCAGCAACAGGGCGGTCAACAACAGCAGTACGGCATACCACAGCAACAGTTC
281    Q  G  G  Q  Q  Q  G  G  Q  Q  Q  Q  Y  G  I  P  Q  Q  Q  F

901   GGCGGTCAGCAACAGGGCGGTCAGCAGTTTGGCGGTCAGCAACAGGGCGGGCAACAGTTT
301    G  G  Q  Q  Q  G  G  Q  Q  F  G  G  Q  Q  Q  G  G  Q  Q  F

961   GGATTTTAGATACCAAGCTGACCACGGCAACGAATCGACAGCTTCAAGGATGTCTAACAG
       G  F  *

1021  TTTTGTAGCCGATTCGATAATCCCCGCCTCCCCAATATGATTTTTTTGTTCTGGCAAAAG
1081  TTAATCATATACATGTCAAACTGAACATTTTAATTAACGCAAAATAAAAAGAATTGAGCA
1141  TAAAAAAAAAAA
```

Figure 4.1.4: LL_{GRP} amino acid character (calculated using SAPS).

(A) LL_{GRP} amino acids and the corresponding charge for each AA.

LEGEND

Bold text	Amino acid sequence
Grey text	AA charge
Grey box	transmembrane region
Blue text	18AA repeat – The clear boxes highlight 2 repeats that overlap with each other (shown in blue)
Red text	10AA repeat – there are 2 of these repeats that run together; the first is underlined to distinguish them.

(B) Compositional analysis of clone LL_{GRP}.

The symbols present after the amino acid indicate whether it is over- or under- represented in the sequence, and to what degree.

LEGEND

Low usage

-- Amino acid X is present as little as the 1% least X-containing proteins in the reference set

- Amino acid X is present as little as the 5% least X-containing proteins in the reference set

High usage

+ present in the 95% quantile

++ present in the 99% quantile

(A)

AA	#	%composition	AA	#	%composition	AA	#	%composition
A--	4	(1.2%)	C	0	(0.0%)	D--	1	(0.3%)
E--	1	(0.3%)	F-	5	(1.6%)	**G++**	**101**	**(31.4%)**
H--	0	(0.0%)	I--	2	(0.6%)	K--	1	(0.3%)
L--	4	(1.2%)	M-	1	(0.3%	N--	0	(0.0%)
P	**21**	**(6.5%)**	**Q++**	**152**	**(47.2%)**	R	11	(3.4%)
S--	1	(0.3%)	T--	3	(0.9%)	V--	1	(0.3%)
W	1	(0.3%)	Y	12	(3.7%)			

(B)

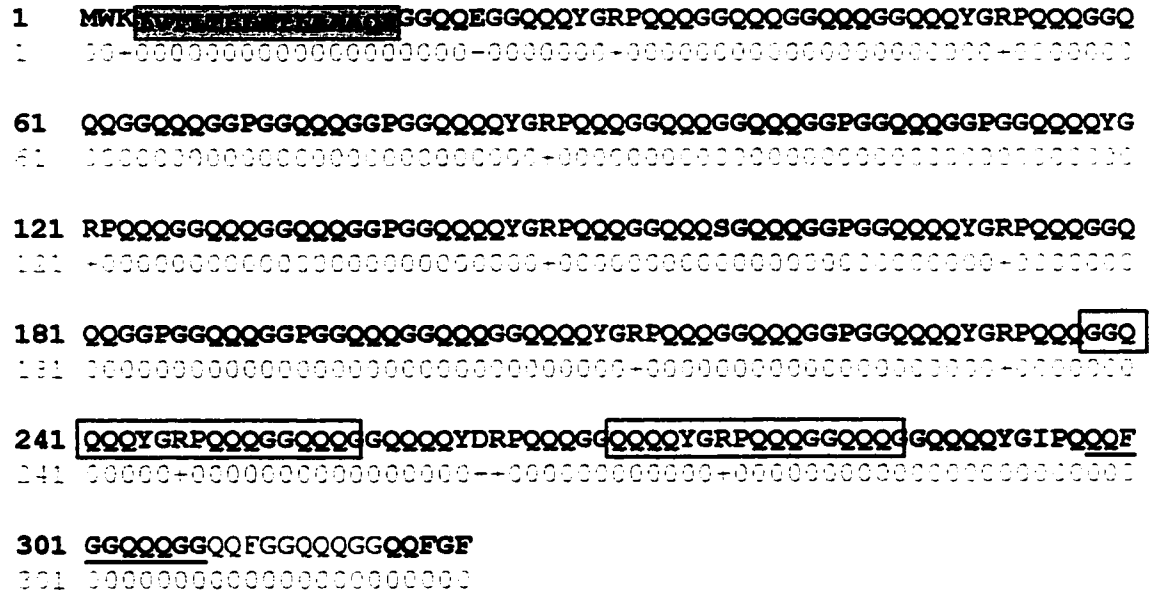


Figure 4.1.5: Statistical evidence for the presence of a secretory signal peptide in the translated amino acid sequence of clone LL_{GRP}. Data was calculated using the web-based analysis program, Signal-P, which returns three scores between 0 and 1 for each position in the amino acid sequence. (See Appendix V for an in depth explanation of these calculations)

C-score (raw cleavage site score)
Derived from networks trained to recognize cleavage sites vs. other sequence positions. Value is trained to be:

high at position +1 (immediately after the cleavage site)
low at all other positions

S-score (signal peptide score)
Derived from networks trained to recognize signal peptide vs. non-signal-peptide positions. Value is trained to be:

high at all positions before the cleavage site
low at 30 positions after the cleavage site and in the N-terminals of non-secretory proteins.

Y-score (combined cleavage site score)
Optimized by observing where the C-score is high and the S-score changes from a high to a low value. The Y-score formalizes this by combining the height of the C-score with the slope of the S-score.

All three scores are averages of five networks trained on different partitions of the data.

(A) The probability that any given amino acid is part of a secretory signal peptide is calculated for each of the first 24 AA. The results of these calculations indicate that a signal sequence is present at amino acids 1-16 (indicated in bold type).

(B) Graphical representation of the above data. The arrow indicates the site of cleavage.

(C) **Hydrophobicity profile.** The x-axis represents the linear arrangement of amino acids, whereas the y-axis indicates the hydrophobic character of the corresponding AA. Higher values indicate stronger hydrophobicity.

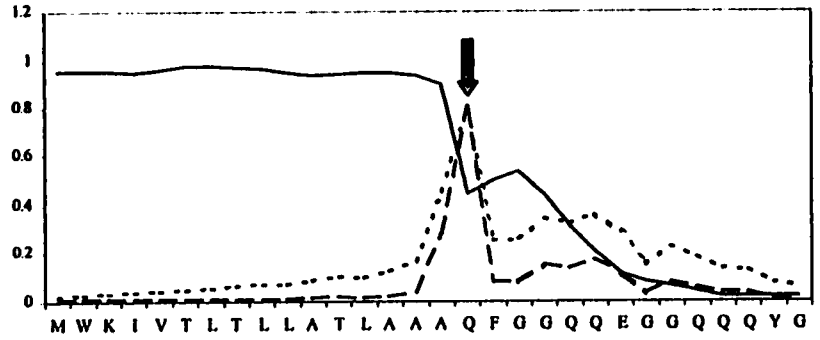
(A)

pos	aa	C	S	Y
1	**M**	**0.012**	**0.957**	**0.020**
2	**W**	**0.012**	**0.957**	**0.026**
3	**K**	**0.012**	**0.956**	**0.032**
4	**I**	**0.012**	**0.950**	**0.038**
5	**V**	**0.012**	**0.961**	**0.042**
6	**T**	**0.011**	**0.977**	**0.047**
7	**L**	**0.012**	**0.982**	**0.054**
8	**T**	**0.013**	**0.976**	**0.062**
9	**L**	**0.013**	**0.967**	**0.068**
10	**L**	**0.012**	**0.951**	**0.071**
11	**A**	**0.016**	**0.944**	**0.086**
12	**T**	**0.020**	**0.949**	**0.101**
13	**L**	**0.017**	**0.955**	**0.097**
14	**A**	**0.024**	**0.950**	**0.123**
15	**A**	**0.040**	**0.943**	**0.164**
16	**A**	**0.279**	**0.903**	**0.453**
17	Q	0.808	0.449	0.796
18	F	0.083	0.503	0.254
19	G	0.082	0.537	0.253
20	G	0.153	0.439	0.347
21	Q	0.132	0.314	0.321
22	Q	0.170	0.203	0.360
23	E	0.113	0.115	0.286
24	G	0.034	0.082	0.152

Measure	Pos.	Value	Cutoff	Conclusion
max.C	17	0.808	0.37	YES
max.Y	17	0.796	0.34	YES
max.S	7	0.982	0.88	YES
meanS	1-28	0.995	0.48	YES

#Most likely cleavage site between pos. 16 and 17: AAA-QF

(B) Black = S-score, Short dash = Y-score, Long dash = C-score



(C)

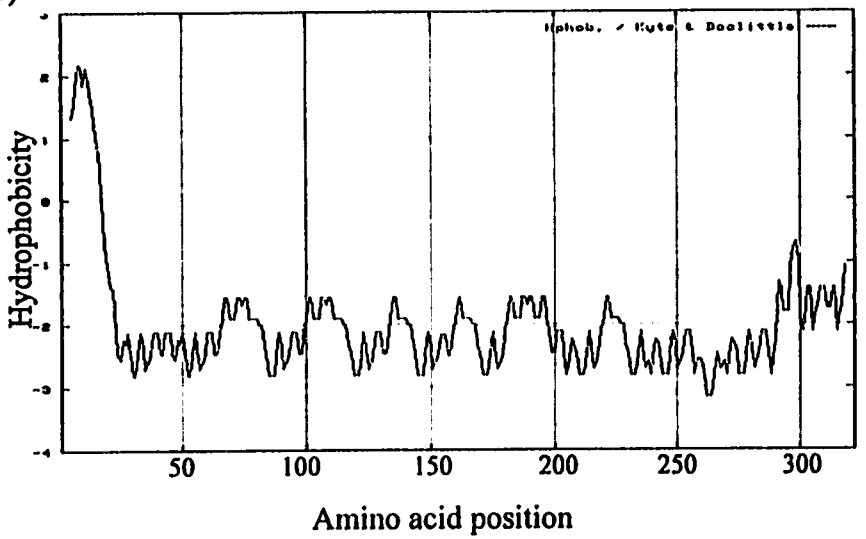


Figure 4.1.6: Secondary structure predictions for LL_{GRP}. Results were obtained using the WWW-based analysis program, PSA (The Protein Sequence Analysis. The server predicts protein secondary and tertiary structure based on the amino acid sequence as well as secondary structure elements, their lengths, and connectivity.

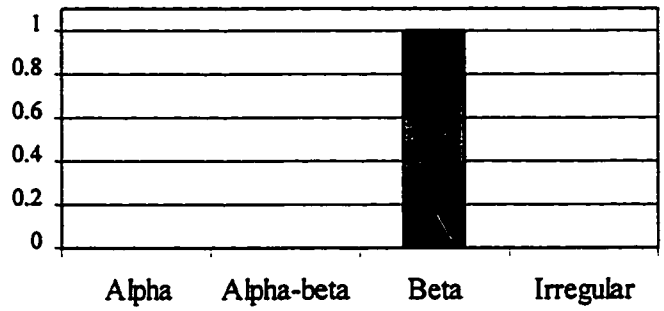
(A) Predicted protein superclass.

(B) Predicted Beta-subgroup classification for LL_{GRP}. The twelve subgroups are indicated on the x-axis.

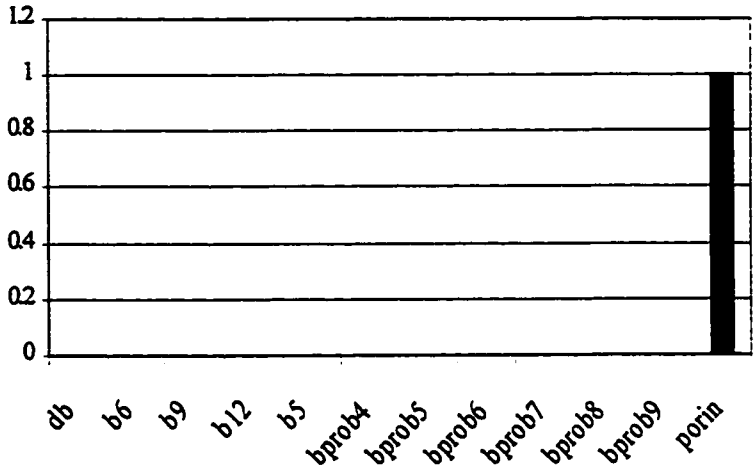
(C) Predicted secondary structure for each amino acid.
Each row corresponds to a different secondary structural state, and each column corresponds to a different residue position. The four rows labeled "TURN 1, 2, 3, 4" denote four-residue tight turns. The probabilities of each residue being in each of the structural states are depicted using contour lines of constant probability in increments of 0.1. Areas surrounded by many contour lines are regions of high probability, while areas outside of the contours have low probabilities of less than 0.1.

For example, residue 52, has a 20% probability of being in a loop because it is encircled by 2 concentric lines in the loop row.

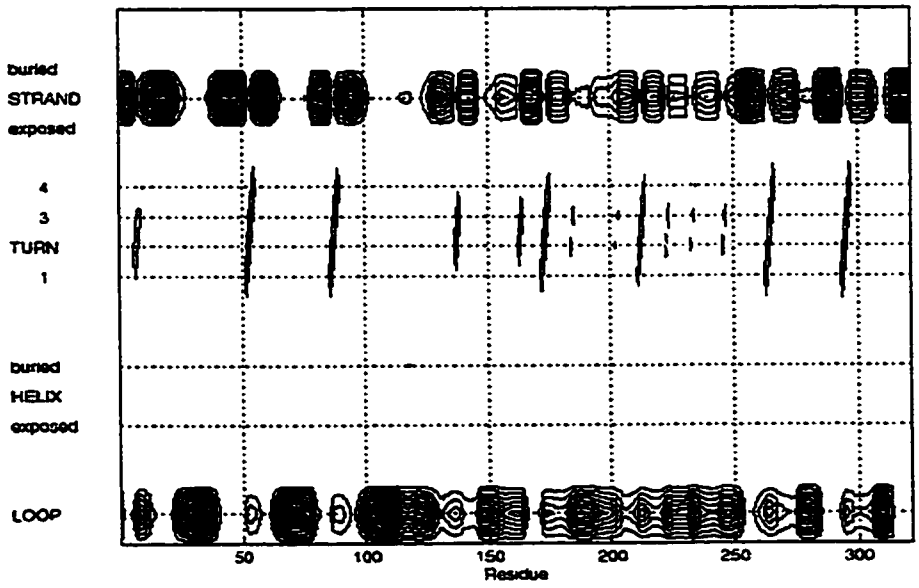
(A)


1
0.8
0.6
0.4
0.2
0
Alpha
Alpha-beta
Beta
Irregular

(B)


1.2
1
0.8
0.6
0.4
0.2
0
db
b6
b9
b12
b5
bprob4
bprob5
bprob6
bprob7
bprob8
bprob9
porin

(C)


buried
STRAND
exposed
4
3
TURN
1
buried
HELIX
exposed
LOOP
50
100
150
200
250
300
Residue

RESULTS

SECTION 4.2: CLONE $LLAF_W$

Northern blot analysis

Clone $LLAF_W$ was isolated via differential screening of the anoxic library alone. When used as a probe for northern blots, it hybridized to a single transcript with a size of approximately 1.1 Kb. Transcript accumulation was fairly modest and increased gradually over the course of anoxia (Figure 4.2.1). After 12 h of anoxia, the message levels had increased 5-fold over the control values, but then continued to climb during the recovery period to a maximum increase of 14-fold higher after 12 h aerobic recovery.

However, in keeping with the results of library screening, there was no change in binding intensity with freeze treated foot muscle. In fact, for several northern trials with RNA from frozen foot the probe either did not bind or presented a very weak signal. As a result, expression of $LLAF_W$ was difficult to quantify for the frozen state.

Environmental stress had no effect on the expression of clone $LLAF_W$ in the hepatopancreas (Figure 4.2.2). Transcript levels were strongest in the control state, and did not rise at any of the sampling times during either the freezing or anoxia timecourses.

Sequence analysis: DNA and amino acid

The clone isolated from the library was roughly 800 nt in length, yet it bound to a 1.1 Kb transcript on northern blots. Since the 3' polyA end was used for selection with the cDNA library, it was clear that the clone was truncated at the 5' end. Using primers designed from the $LLAF_W$ sequence, the full length cDNA of 1117 nt was elucidated.

This sequence is presented in Figure 4.2.3 along with the translated amino acid sequence. The mRNA motifs of the original clone were illustrated in Chapter 2, but in addition to these, the RACE product showed a 5'UTR, a start codon, and a complete ORF. As with clones LL_{MET} and LL_{GRP}, Kozak sequence analysis supported the first ATG codon as the translation initiation site (Table 4.2.1).

Amino acid composition and physical character

Translation of the putative ORF produced a 143 amino acid sequence representing a 15.4 KDa protein with a predicted isoelectric point of 4.60 (values calculated using DNAman). As with LL_{GRP}, the N-terminus of the sequence has a high Lys and Val content, which is characteristic of secretory signal peptides. Accordingly, results from SignalP analysis (Figure 4.2.4) supported the identification of the 18 amino acids of the N-terminal sequence, MMTKTVLVLLIVCAMLVA, as a cleaved secretory signal peptide. The amino acid composition, charge profile, hydrophobicty are shown in Figure 4.2.5. Unlike LL_{GRP}, the putative protein does not show over- or under- representation of any amino acid (with respect to the SAPS data set). In addition, with the exception of histidine, each amino acid is present at least once in this sequence. $LLAF_W$ has one positive charge cluster, at amino acids 24-49, but no negative or mixed charge clusters. Finally the SAPS analysis did not identify any significant hydrophobic or transmembrane regions in the AA sequence.

Secondary structure prediction

Analysis with PSA catalogued $LLAF_W$ as an alpha class protein. More specifically, an APB (anti-parallel bundle) structure, a protein with 4 amphipathic alpha-helices of nearly identical lengths. The probability distributions for these results are illustrated in Figures 4.2.6 a and b. Figure 4.2.6 c illustrates the pattern of secondary structures throughout the sequence. Tight turns and beta-strand motifs are practically non-existent but there is a very clear alternating pattern of loops and alpha helices along the sequence. Figure (D) also illustrates the loops and helices, but indicates whether an amino acid is buried or exposed to the solvent. The figure clearly indicates that the first 20 amino acids fit the helix pattern, but are also buried from the solvent, in stark contrast to the remaining helices which have alternating buried and exposed amino acids, indicating an amphipathic structure.

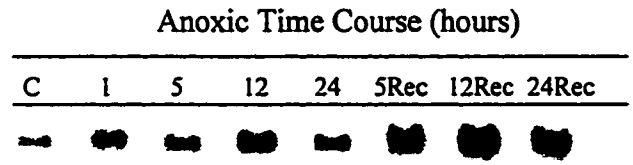
Figure 4.2.1: Northern hybridization of clone $LLAF_W$ against total RNA from the foot muscle of anoxia-treated animals.

(A) Binding of $LLAF_W$ to its complementary message in the total RNA.

(B) Loading controls. Binding of ^{32}P-labeled probe of the gene for ribosomal RNA to the above blot. Also presented is the ethidium bromide-stained ribosomal RNA bands of the same northern blot.

(C) Illustration of clone $LLAF_W$ pixel density as a percentage of the corresponding rRNA band.

A)


Anoxic Time Course (hours)
C
1
5
12
24
5Rec
12Rec
24Rec

B)


Loading controls
rRNA
Gene for rRNA
16s rRNA

C)

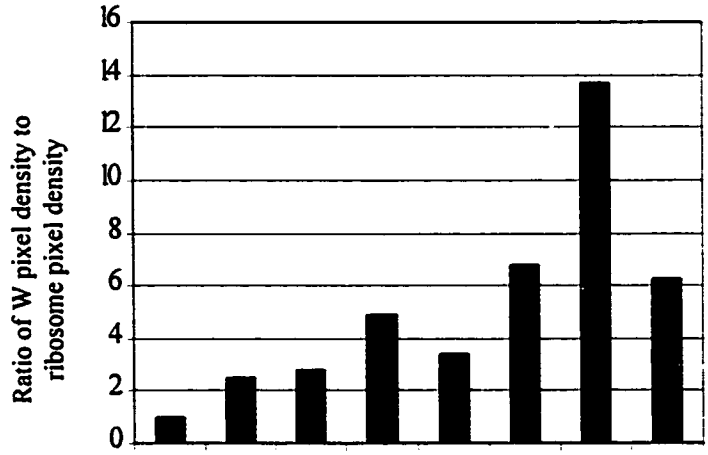

Ratio of W pixel density to ribosome pixel density
16
14
12
10
8
6
4
2
0

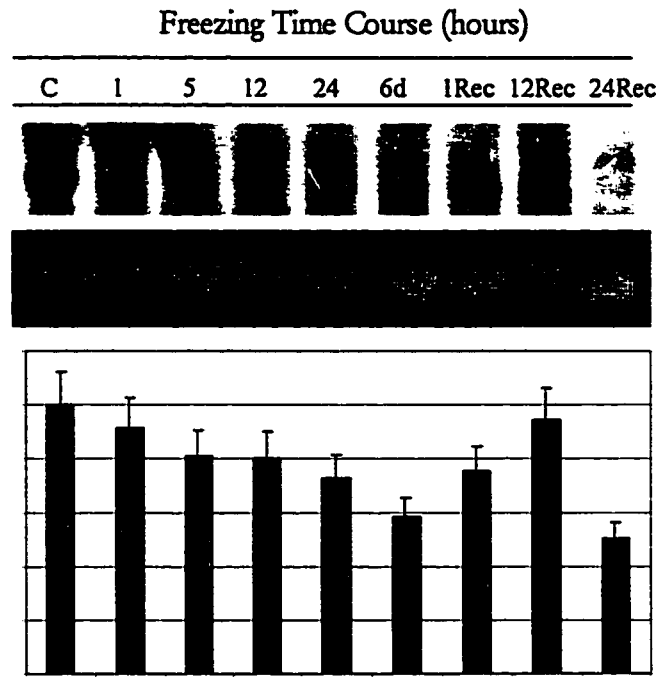
Figure 4.2.2: Northern hybridization of clone $LLAF_W$ against total RNA from the hepatopancreas of anoxia and freeze-treated animals.

Figures A and B indicate (from top to bottom) the time course of stress treatment; the northern blot hybridized with ^{32}P-labeled $LLAF_W$; the ethidium bromide-stained ribosomal RNA bands; and a graphical illustration of clone $LLAF_W$ pixel density as a percentage of the corresponding rRNA band.

(A) Freezing exposure

(B) Anoxia treatment

A)


Freezing Time Course (hours)
C
1
5
12
24
6d
1Rec
12Rec
24Rec

B)

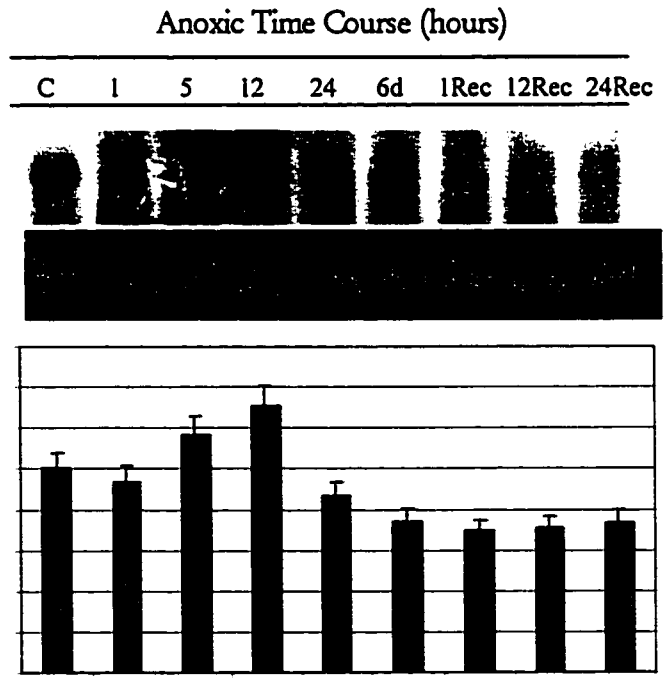

Anoxic Time Course (hours)
C
1
5
12
24
6d
1Rec
12Rec
24Rec

Table 4.2.1: Identification of putative Kozak sequences in nucleotide sequence of $LLAF_W$ using the web-based analysis program, ATP^{PR} (see Appendix IV for a description of the program and reference).

The most likely site of translation initiation is the ATG codon of the longest ORF identified in the gene sequnce.

# of ATG from 5'end	Reliability	Frame	Identity to Kozak rule A/GXXATGG	Start (bp)	Finish (bp)	ORF Length (aa)	Stop codon found?	Sequence (refer to figure 4.2.3 for the intervening AA)
1	0.69	1	AXXATGa	94	522	143	Yes	MMTKTV...EACAEVEAEIQI
2	0.27	1	AXXATGa	97	522	142	Yes	MTKTV...EACAEVEAEIQI
4	0.10	1	GXXATGc	136	522	129	Yes	MLVADS... EAÇAEVEAEIQI
6	0.08	2	cXXATGG	374	574	67	Yes	MVLTLMR.....YLTTLTESEIE

Figure 4.2.3: Nucleotide and translated amino acid sequence for clone $LLAF_W$, a novel cDNA isolated from the cDNA library synthesized from anoxia-treated foot muscle of *L. littorea*. The originally isolated clone (described in chapter 2) was truncated at the 5 end. This sequence represents the orignal clone and the sequence of the 5'RACE product.

LEGEND

Secretory signal peptide (box)	XXXXX
Nucleotides	black text
Amino acids	grey text

```
-93     GGCCACGCGTCGACTAGTACGGGGGGGGGGDGGDDGTGGAAATATCGTCATTTTCAAGGC
-33     ACAACACTGGTTCTAACGACAGACTTCGCCACGA

1       ATGATGACCAAGACCGTCCTGGTATTACTGATCGTATGTGCAATGCTTGTAGCAGACTCT
1       M  M  T  K  T  V  L  V  L  L  I  V  C  A  M  L  V  A  D  S

61      GACTGCTGGAGACGAAGACGACGAAGAAGAGGTTTCTTTCGAGGGGTTTTTCGGATTGTG
21       D  C  W  R  R  R  R  R  R  R  G  F  F  R  G  V  F  R  I  V

121     AAGACCGGTGTCTATGCGACCGGAAAAGCGGTGGGGGCGGCAGTGGGGGCAGTGGGAAAG
41       K  T  G  V  Y  A  T  G  K  A  V  G  A  A  V  G  A  V  G  K

181     CGTGACCTGTGTGCTGACCTGCTGACCCCTGACCAGGTGCAACAACTAGCCACTGACCTA
61       R  D  L  C  A  D  L  L  T  P  D  Q  V  Q  Q  L  A  T  D  L

241     CAGAACACCTGTGCCAACATCCCGATCAACCTGGCCGCAGATGGTCTTGACTTTGATGAG
81       Q  N  T  C  A  N  I  P  I  N  L  A  A  D  G  L  D  F  D  E

301     GTCAAGGGCATTTTCACGGACAGCGATGTTGATGGAGATGGAAAACTGTTGGCCACAGAG
101      V  K  G  I  F  T  D  S  D  V  D  G  D  G  K  L  L  A  T  E

361     TTGGCAGACTTTTCTGCTGCAATCCAAACAACAGAAGCCTGTGCGGAGGTGGAGGCCGAG
121      L  A  D  F  S  A  A  I  Q  T  T  E  A  C  A  E  V  E  A  E

421     ATTCAGATATAAACCTGATATGATCATTTACTTGACGACACTGACAGAGAGCGAGATAGA
141      I  Q  I  *

481     ATAGATCCATTCTTTTTTTCTTCTTCAAAATAACGTAATTCAAATTAAGTAAATAGGTGC
541     ACGAACAAAAAATTCTTCCAATACAACTGGGCCCTCATGAAAAAAGATGATAAATACTAA
601     GAAGAAAAATGAAAGACCGTGTCACATCTGTAAGAAAAAATAAAACAAACAAAGGTATTT
661     CGTTGAAATATATCAAAAGAAAAGGATTAAAATAAGAAGTGCATCATTTACTAAAAGAAG
721     AACACAATAACAAAAATGTCCATTCAGTTAAAGCGCAAGAGATCAACGTATGCACGCTTC
781     TTTTGTTGTTGTTGCTTAAGACAACTTTTATTCAATCGTTTCCATTTATTTTTCAATTTA
841     TTAAAGAGACAAAAAGCAGCAATTGGTCATGGTGTGAATTGTGTGTGATGAGAATCGCTG
901     ATGTTTTATTGTTGTTCTTCGCTACTTGATAATGTGTCGCCCAAAGTCTATCAGCTTCTT
961     TGAGAGTAATGGTTGTGTAAAATAAAAAGTGTCTAAAAAATCAAAAAAAAAAAAAAAAA
1021    AAAA
```

Figure 4.2.4: Statistical evidence for the presence of a secretory signal peptide in the translated amino acid sequence of clone $LLAF_W$. Data was calculated using the web-based analysis program, Signal-P. The position of the Y-score maximum indicates the cleavage site; if the S-score is greater than 0.5, the sequence is predicted to be a signal peptide.
See Appendix V for an in depth explanation of these calculations.

(A) The probability that any given amino acid is part of a secretory signal peptide is calculated for each of the first 20-30 AA. The results of these calculations indicate that a signal sequence is present at amino acids 1-18 (indicated in bold type).

(B) Graphical representation of the above data. The arrow indicates the site of cleavage.

(A)

Residue position	AA	C-score	S-score	Y-score
1	**M**	**0.012**	**0.944**	**0.000**
2	**M**	**0.012**	**0.949**	**0.000**
3	**T**	**0.012**	**0.966**	**0.014**
4	**K**	**0.014**	**0.971**	**0.024**
5	**T**	**0.012**	**0.974**	**0.031**
6	**V**	**0.012**	**0.962**	**0.038**
7	**L**	**0.011**	**0.971**	**0.044**
8	**V**	**0.012**	**0.960**	**0.050**
9	**L**	**0.013**	**0.971**	**0.059**
10	**L**	**0.012**	**0.974**	**0.061**
11	**I**	**0.014**	**0.970**	**0.069**
12	**V**	**0.012**	**0.963**	**0.068**
13	**C**	**0.017**	**0.968**	**0.085**
14	**A**	**0.027**	**0.952**	**0.114**
15	**M**	**0.111**	**0.909**	**0.238**
16	**L**	**0.019**	**0.909**	**0.101**
17	**V**	**0.072**	**0.891**	**0.201**
18	**A**	**0.048**	**0.861**	**0.167**
19	D	0.451	0.602	0.520
20	S	0.135	0.536	0.282

#Measure	Position	Value	Cutoff	Conclusion
max. C	23	0.493	0.37	YES
max. Y	19	0.520	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.948	0.48	YES

(B)

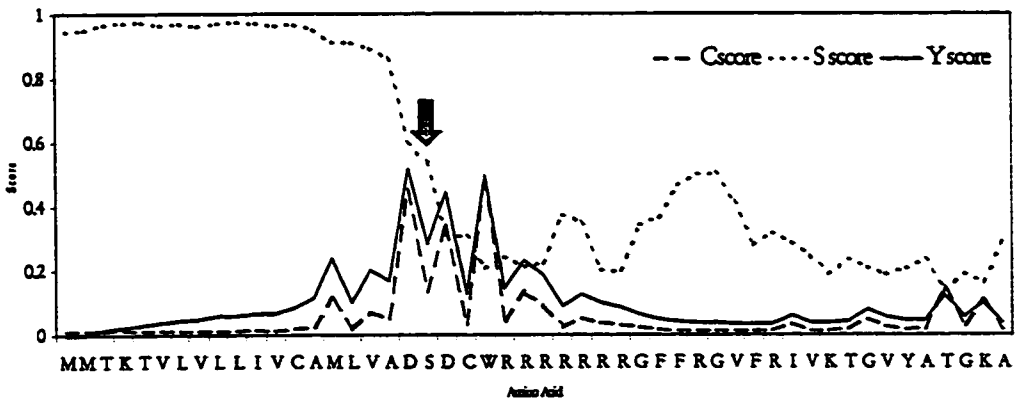


Figure 4.2.5: $LLAF_W$ amino acid character (calculated using SAPS).

(A) $LLAF_W$ amino acids, the corresponding charge and most likely secondary structure for each AA.

LEGEND

Bold text	Amino acid sequence
Grey text	AA charge

Secondary structure:

e	strand
h	helix
t	turn
c	coil

(B) Compositional analysis of clone $LLAF_W$.
In contrast to LL_{GRP}, no amino acid is over- or under- represented in this sequence.

(C) Hydrophobicity profile. The x-axis represents the linear arrangement of amino acids, while the y-axis indicates the hydrophobic character of the corresponding AA. Higher values indicate stronger hydrophobicity.

(A)

```
1    MMTKTVLVLLIVCAMLVADSDCWRRRRRRRGFFRGVFRIVKTGVYATGK
1    000+0000000000000000-0-00+++++++000+000+00+0000000+
1    hhhhhhhhhhhhhhhhhhttthhhhhhtccchheeeeeeeeeeeethh

51   AVGAAVGAVGKRDLCADLLTPDQVQQLATDLQNTCANIPINLAADGLDF
51   0000000000+--000-0000-0000000-00000000000000-00-0
51   hhhhhhhhhccccchhhhccccchhhhhhhhhhhhhccccccceeeecttcch

101  DEVKGIFTDSDVDGDGKLLATELADFSAAIQTTEACAEVEAEIQI
101  --0+0000-0-0-0-0+0000-00-00000000-000-0-0-000
101  hhhheeeeccccctthhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```

(B)

AA	#	%composition	AA	#	%composition	AA	#	%composition
A	19	(13.3%)	C	5	(3.5%)	D	14	(9.8%)
E	6	(4.2%)	F	6	(4.2%)	G	11	(7.7%)
H	0	(0.0%)	I	8	(5.6%)	K	6	(4.2%)
L	14	(9.8%)	M	3	(2.1%)	N	3	(2.1%)
P	2	(1.4%)	Q	6	(4.2%)	R	10	(7.0%)
S	3	(2.1%)	T	11	(7.7%)	V	14	(9.8%)
W	1	(0.7%)	Y	1	(0.7%)			

(C)

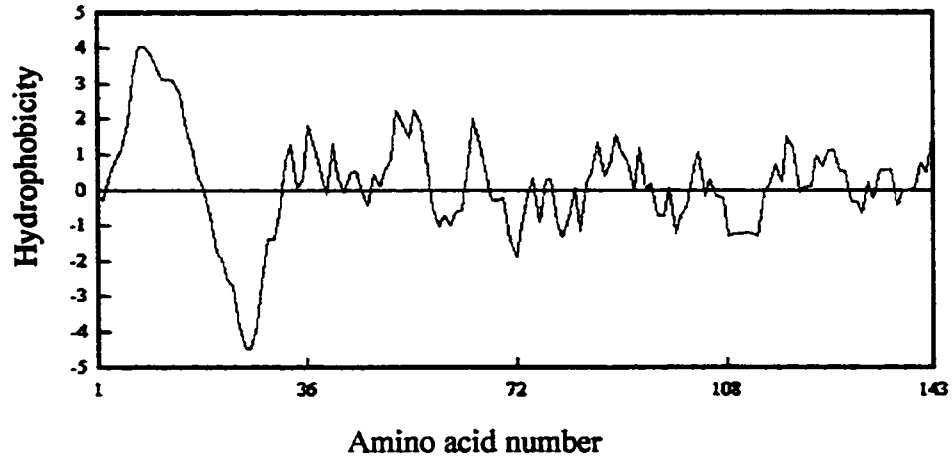


Figure 4.2.6: Secondary structure prediction for the $LLAF_W$ amino acid sequence using the WWW-based analysis program, PSA.

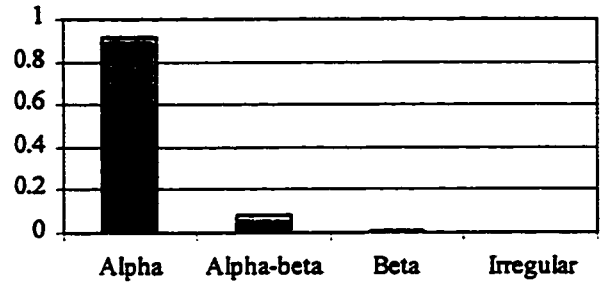
(A) Predicted protein superclass, which are based on the folding pattern of a protein. $LLAF_W$ falls into the alpha class with a 92.8% probability.

(B) Predicted alpha and alpha-beta subgroup classification, each of which has slight structural variations. $LLAF_W$ is most likely of the APB, antiparallel bundle, class.

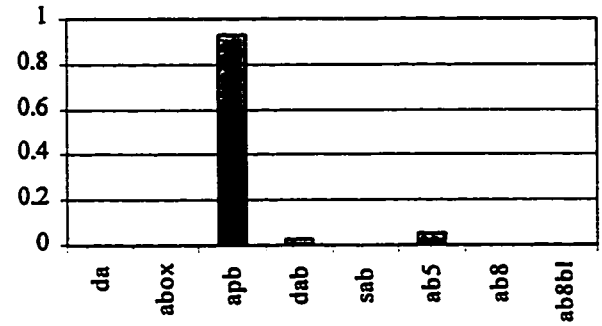
(C) Predicted secondary structure for each amino acid. The linear strand of amino acids is represented on the X-axis, while the y-axis indicates the probability that a given AA is in a loop, helix, turn, or strand.

(D) This figure also illustrates the secondary structure for each AA, as in Figure C, but with the additional information of whether the residue in a helix is likely to be exposed or buried. High probabilities are represented with more concentric lines. (Note: Figure C illustrates that essentially every amino acid is found in either a loop or helix, therefore only these two structure distributions are depicted here.)

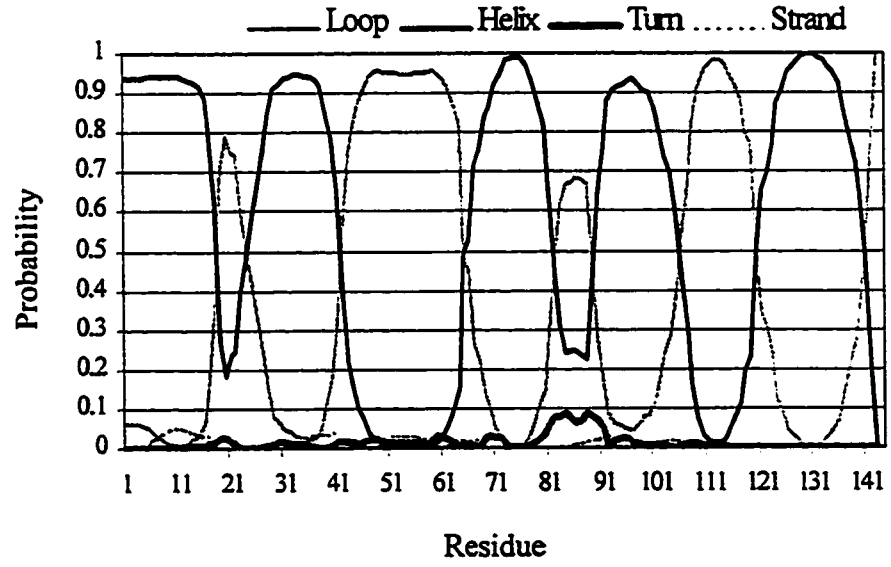
(A)


1
0.8
0.6
0.4
0.2
0
Alpha
Alpha-beta
Beta
Irregular

(B)


1
0.8
0.6
0.4
0.2
0
da
abox
apb
dab
sab
ab5
ab8
ab8bl

(C)


Loop
Helix
Turn
Strand
Probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
11
21
31
41
51
61
71
81
91
101
111
121
131
141
Residue

(D)

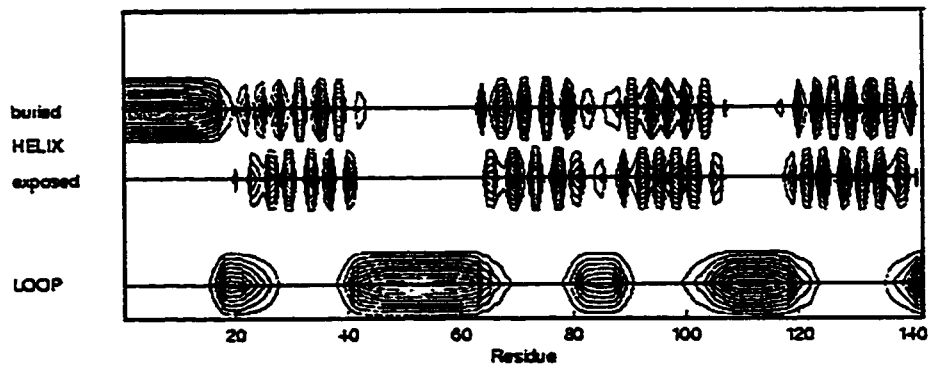

buried
HELIX
exposed
LOOP
20
40
60
80
100
120
140
Residue

Table 4.1 Results of PSORT analysis. This program is used to predict the subcellular location of a protein from its amino acid sequence.

	Clone LL_{GRP}	Clone $LLAF_W$
Presence of a signal peptide **PSG** calculation **GvH** calculation	5.06 Cleavable signal peptide 1-16	6.75 Cleavable signal peptide 1-22
ALOM Detect TM region	None	None
MTOP Membrane topology	N-terminal inside	C-terminal side inside **Caution:** Inconsistent mtop result with signal peptide
MITDISC Mitochondrial target sequence	none	None
NUDISC Nuclear localization signal	Pat4: none Pat7: none Bipartite: none Content of basic residues: 3.7% NLS Score: -0.47	pat4: RRRR (5) at 24 pat4: RRRR (5) at 25 pat4: RRRR (5) at 26 pat4: RRRR (5) at 27 pat7: none bipartite: none content of basic residues: 11.2% NLS Score: 0.78
KDEL – ER retention sig	None	None
SKL and **SKL2** Peroxisomal target signal in the C-terminus	None	None
VAC	None	None
RNA-binding motif	None	None
NMYR Myristoylation pattern	None	None
Prenylation motif	None	None
MemTQRL; transport motif from cell surface to golgi	None	None
Tyrosines in the tail	None	None
Dileucine motif in tail	None	None
NNCN Prediction of cytoplasmic vs. nuclear localization	Nuclear Reliability 94.1	Cytoplasmic Reliability 94.1
Coiled-coil regions	Zero residues	Zero residues
k-NN prediction **PREDICTION**	44.4% extracellular 33.3% nuclear 22.2% cytoplasmic **Extracellular**	44.4 %: extracellular, including cell wall 22.2 %: vacuolar 22.2 %: plasma membrane 11.1 %: cytoplasmic **Extracellular**

DISCUSSION

In an early text about sequence analysis, von Heijne stated that "molecular biology is all about sequences" (1987). This prophetic statement has only become more true over the following 13 years, as illustrated by the recently published rough draft of the human genome (a massive 10^9 bp) as well as the whole or partially complete genomes of 600 other organisms. This enormity of sequence data requires creative methods of sequence analysis (for DNA and proteins) a crucial area of molecular research, if we are ever to "decode" and characterize newly discovered genes. This chapter illustrates the usefulness of several predictive methods in the effort to obtain descriptive information.

Protein Structure

The approach to protein structure elucidation is generally one of "knowledge-based prediction". Programmers utilize data from the structures of known proteins to make analogies with unknown amino acid sequence at all levels of the structural hierarchy: secondary, motifs, domains, and quaternary interactions. The www-based analysis program PSA is one such program that is publicly available and is a very useful tool when dealing with novel protein sequences. Because the *L. littorea* sequences are novel, clues about their secondary structures that are derived from comparative analysis with known proteins can indicate possible functions or cellular locations. Identification of specific structural elements in the unknown proteins and knowledge of how those

structural elements function in other proteins can suggest possible roles of the *L. littorea* proteins.

LL_{GRP} was predicted to belong to the porins, a group of beta-barrel proteins found in the outer membranes of Gram-negative bacterial cells, and the outer-membranes of mitochondria and chloroplasts. Clearly, the only class of porin that is appropriate for this study is the mitochondrial porins. These proteins are referred to as VDACs (Voltage Dependent Anion Channels) and represent a regulated gate between the cytosol and intramitochondrial compartments for the exchange of metabolites. Believed to occur as dimers, these mostly beta sheet transmembrane proteins form a hydrophilic pores in the outer membrane that allow neutral molecules (up to 5 kDa) to diffuse in and out the mitochondrion. However, keeping in mind that all algorithms used to analyze unknown protein sequences should come with the warning *"Do not expect your computer to tell you the truth"*, this conclusion should not be blindly accepted. First, the sequences of known mitochondrial porins are quite well conserved, and do not resemble that of LL_{GRP}. Secondly, the secondary structures have been determined for several different porins when embedded in lipid membranes (Vogel and Jahnig, 1986) and the results indicate a structural composition of 50-60% beta sheet, 20% turns, and less than 15% helix. Such a report leaves no room for loop structures, which although non-repetitive, can still be an ordered structure. As illustrated in Figure 4.1.6, the residues of LL_{GRP} are split roughly 50/50 between beta sheets and loops, with no alpha helices whatsoever. While I am not convinced LL_{GRP} is a functional porin, the characteristics described here are such that the possibility remains that LL_{GRP} represents a new membrane protein.

$LLAF_W$ was deemed a member of the alpha class of proteins. These predominantly contain α-helices, and almost no beta sheets or turns, features that are clearly illustrated in Figure 4.2.6(c). Of the four structural classes, on average, alpha proteins have the shortest chain length at 129 residues (Chou, 1989). At 143 amino acids $LLAF_W$ certainly falls into this size range. Alpha proteins can also be further delineated into eight subclasses. When the signal sequence is cleaved from $LLAF_W$, the protein consists of 4 helices connected by loops, a structure that is characteristic of the anti-parallel bundle (apb) class. This group contains a range of proteins including the globins (hemoglobin and myoglobin), erythrins, cytochromes, and the cytokine interleukin-4. Such a list leaves an intriguing possibility that this protein could be responsive to oxygen and, as such, may have a role in the well developed anoxia tolerance of *L. littorea*.

Sequence databases continue to grow in size, and by virtue of evolutionary relationships, in redundancy. However, the number of protein families has leveled off. Therefore, identities derived from protein family relationships and known secondary structures allows a relatively high level of confidence in the predictions. Currently, secondary structures can be predicted to 70% accuracy (Chou 1989; Greene and Henikoff, 1999). Specialized structures such as trans-membrane regions and coiled coils can also be predicted, but there are far fewer sites devoted to this purpose simply because the rules for folding of these structures are far less well understood. However, with

advances in modeling and the continuing accumulation of direct protein structure data, this can only improve in the future.

Secretory Signal Peptides

The impetus for reporting clones $LLAF_W$ and LL_{GRP} together in this chapter, was the presence of a secretory signal peptide in the translated amino acid sequences. The concept of a such signals was first proposed by Blobel (1980), where he indicated that newly made proteins are targeted and imported into the various organelles by built-in signal sequences that are composed of short stretches of amino acids. Such a mechanism is crucial to the construction of a functional cell because a typical mammalian cell has 10-20,000 different proteins, each of which must be located to the correct aqueous space, membrane or organelle to fulfill its function.

Proteins that are membrane-bound or destined for excretion are synthesized by ribosomes associated with the membrane of the endoplasmic reticulum (ER). A secretory signal peptide is found at the N-terminus of the nascent protein and directs it to the ER and the vesicular sorting pathway. The signal peptide, emerging from the ribosome, binds to the signal-recognition particle (SRP), which then docks, as a complex, to the SRP-receptor and channel, called a translocon. The SRP dissociates from the receptor and the nascent polypeptide chain is translocated through the channel into the ER lumen. The signal peptide is finally cleaved off, while translation of the rest of the protein continues. Ultimately, the protein is secreted out of the cell or sorted to the appropriate organelle.

Secretatory signal peptides typically consist of 15-30 residues, which are conserved with respect to the character of the amino acids. They retain a positively charged N-region, followed by a hydrophobic h-region and a neutral, yet polar, c-region. In addition, because the signal sequences are cleaved when the protein reaches the ER, there are restrictions on the amino acids near the cleavage site. The (-3, -1) rule states that the residues at positions –3 and –1 (with respect to the cleavage site) must be small and neutral. The predictability allows for their identification in both novel and known proteins. SignalP, an algorithm created by Nielsen et al. (1997) is accepted as the gold standard in secretory signal sequence detection, because it is designed to use three predictive calculations on which the final conclusion is based. Using this program, secretory signals were detected in both the unknown proteins presented in this chapter.

The signal peptide is most obvious in LL_{GRP}, as 2 of the 3 calculations reached the same conclusion that a 16 amino acid signal sequence was present that was cleaved between residues 16 & 17. In addition to the statistical support, visual inspection shows that the residues found in the sequence are consistent with the character of known secretory proteins. The hydrophobicity profile in Figure 4.1.5 (c) shows that the residues in the secretory signal are indeed hydrophobic, in strong contrast to the rest of the protein, and the (-3, -1) residues are alanines, often termed the "default amino acid" due to its small size and neutral, nonpolar character. The demarcation of the signal in clone $LLAF_W$ is less obvious. While each calculation was again positive, the results of the predictions were very close to the statistical cut-off (Figure 4.2.4a). But, like LL_{GRP}, the amino acid hydrophobic character within the signal is consistent with those of previously described secretory proteins. However, the (-3, -1) rule was not explicitly followed; the -1

position is occupied by an alanine, but the -3 position is lysine, a residue with a long, charged side chain. This is likely an allowable exception to the rule because the lysine side chain is very flexible and may not hinder cleavage.

A variety of proteins, with wide-ranging functions, are classified as secretory, including cytokines, glucagon, mucin. Because this class of proteins is either excreted from the cell or membrane-bound, it is logical to presume that these novel proteins perform a function similar to one of the known secretory proteins. While not a perfect method of comparison, the number of secretory proteins is far smaller than the total protein complement of the cell.

PSORT analysis

While predictive data has great potential efficacy, the results are still an educated guess. Obtaining the information is easy - interpreting it is not. Various programs can produce conflicting results. Such conflicts are illustrated by the PSORT analysis described in Table 4.1. The function of this site is to predict the subcellular location of a protein. The site was designed such that it amalgamates predictions from several other algorithms, and calculates the potential subcellular location of the protein on the basis of this data. On the surface, the scores for NNCN (nuclear and cytoplasmic) and the k-NN prediction (extracellular) seem at odds. NNCN is designed to distinguish between the tendency of a protein to be in either the nucleus or the cytoplasm, whereas the k-NN score allows for the possibility of an extracellular location. However, despite these discrepancies, the final conclusion, that LL_{GRP} and $LLAF_W$ are extracellular proteins, is in accordance with the SignalP conclusion, which does not discriminate between signal

peptides that direct proteins to the cell surface fot excretion versus those directing proteins to organelles.

Discussion of Results

Northern blot hybridization with foot muscle total RNA clearly demonstrated that clone LL_{GRP} showed the most remarkable upregulation of all the differentially expressed genes isolated during this study. From amounts that were practically undetectable in the control state, transcript levels increased 10-20 times upon imposition of anoxia or freezing. Although a longer exposure to freezing was required to reach maximum transcript levels in the frozen state, expression seemed to be enhanced equally in response to anoxia and freezing. The similar response to both stressors implies a common induction mechanism and a function that is appropriate to both freeze-tolerance and anoxia. In addition, since expression was much lower in the hepatopancreas, it might also be implied that the function of the gene product might be more crucial to stress survival by muscle than by hepatopancreas. This type of organ-specific variations in protein content can be found in other animals. For example, the serum AFPs of fish are produced by liver but skin produces its own endogenous AFPs which stay and act locally and lack the export signal sequence of the liver AFPs. In addition, riboflavin binding protein is typically a liver product that is exported to the yolk of eggs, but in birds (not reptiles or amphibians) a second form of RBP is produced in the oviduct and located to the egg white. So, synthesis in 2 places in the snail could indicate that all tissues need this protein to some extent during freezing and/or anoxia, with the extent depending on the particular function of the protein and its importance for stress survival in each organ.

Alternatively, it could indicate that there are major and minor sites of synthesis, perphaps producing proteins with different fates. Currently such conclusions are conjecture, but northern and western blots of all tissues and blood could provide clarification.

Although the transcript size of approximately 1.3 Kb is close to that of the clone sequence (1092 nt) there is evidence that the LL_{GRP} sequence may be incomplete. Anecdotally, stop codons are usually present in the 5'UTR before the translation initiation site, but there were no such codons in the 5' region of this sequence. However, analysis with ATG^{PR} supported the first possible initiation site as the functional one. As described in Chapter 3, this analysis program takes into account the biological fact that eukaryotic ribosomes are not choosy, and barring any strong variations from the Kozak definition, will declare the first ATG the start codon.

As illustrated in Figure 4.2.2, accumulation of $LLAF_W$ transcripts was far more modest than for LL_{GRP}. In addition, $LLAF_W$ accumulation showed tissue and stress specific differences from the LL_{GRP} patterns. Although the transcript was present in the hepatopancreas, expression in this organ actually decreased in response to both anoxia and freezing. In the foot muscle, $LLAF_W$ was not detected during freezing, whereas transcript levels showed a gradual increase throughout the anoxia time course, but the maximum expression occurred after 12 hours of aerobic recovery. Conclusions are tentative at this point, but it seems likely that this expression pattern is reflective of a protein that provides a recuperative function.

Future experiments

Western immunoblotting is a fast, easy method of protein detection. According to the protocol outlined in Chapter 3, this is likely the fastest, method to confirm protein changes during stress or recovery, and to hopefully illustrate that the protein content does indeed increase with mRNA accumulation. Given the presence of secretory signals, it seems crucial to determine if the proteins are present within the cell or predominantly excreted. The immunoblotting procedure can be expanded to illustrate these details. Using fresh tissue, protein samples isolated from subcellular fractions (mitochondria, membranes and cytosol) could be compared, on the same blot, to blood proteins and localize the protein to one of these options.

While predictive algorithms are clearly useful in the effort to obtain descriptive information about a novel protein. However, they still cannot provide positive identity. A newly reported technique, UPA (universal protein array) (Ge, 2000) is a shotgun approach to detect protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. Intended as an efficient method to screen proteins as targets for new drug ligands, it could be a simple, fast method to illustrate the proteins or ligands with which a novel protein will interact. As published, the majority of proteins in the assay are transcription factors, activators and co-factors. Many of the proteins are commercially available at high purity, therefore, using a simple 96-well blotting apparatus, it would be simple to create a "homemade" blot and screen for potential interactions with key cellular pathways.

General Discussion

General Discussion

Strategies to deal with environmental stresses are as varied as the animals that experience them. Some terrestrial and bird species avoid temperature extremes by migrating large distances while others, not capable of traveling, find shelters that limit ambient temperature fluctuations. Some small mammals spend the winter tunneling through the snow, a very effective insulator. Other terrestrial and aquatic animals take a different tact, hibernating in dens or deep water where the temperature is maintained above freezing. Still others actually submit to freezing temperatures yet survive via one of two mechanisms: freeze tolerance or freeze avoidance. Freeze-tolerant species permit the formation of ice crystals when the temperature of body fluids drops below the freezing point, but they limit the size of the crystals and restrict growth to the extracellular fluids. Intracellular ice is always lethal, so tolerant animals must possess adaptations that enhance the rate of survival. Freeze-avoiding species supercool the body fluids so ice does not form, even when the body temperature has dropped well below zero. Natural survival of freezing body temperatures is found widely among invertebrates (Storey and Storey, 1992a) as well as some terrestrial amphibians and reptiles (Storey, 1990; Storey and Storey, 1992b).

The marine periwinkle, *Littorina littorea*, is an ideal model system for the study of natural tolerance to both freezing and anoxia. While this animal can tolerate long term freezing at -8C, extracellular ice still presents several potentially lethal complications, including: (1) potential physical damage due to ice crystal expansion; (2), ischemia, caused by arrested delivery of O_2 and nutrients to tissues; and (3), extracellular ice

effectively increases the solute concentration in the extracellular space resulting in dehydration and increased ionic strength in the cells. Intertidal snails also endure anoxia stress under multiple circumstances: during exposure to air at low tide, when trapped in stagnant, shallow tide pools, or when tissues freeze in winter. Therefore, *Littorina littorea* must have developed effective adaptations to survive these stresses, as they cannot physically escape them.

Past studies of freeze- and anoxia-tolerant species have been biochemical in nature. These studies revealed the existence of ice nucleating and antifreeze proteins, as well as alterations to intermediary metabolism that allow the production of cryoprotectants and modifications to regulatory enzymes. Recently, molecular biology techniques have been applied to these animals, with the goal of illustrating the role played by gene expression in environmental stress survival. Studies from our lab have elicited a range of genes expressed in evolutionarily unrelated animals, in response to cold and anoxia. Freezing induces the expression of alpha and gamma fibrinogen (Cai and Storey 1997a) and mitochondrial ATP/ADP translocase (Cai et al., 1997) in the freeze-tolerant frog, *Rana sylvatica*. Over-wintering insects, *Epiblemma scudderiana* and *Eurosta solidaginis*, showed cold-induced expression of a LIM protein, which signals muscle differentiation, and acidic ribosomal protein, respectively (Bilgen, 1999). Estivating spadefoot toads and terrestrial snails, anoxia-treated crayfish, hypoxic turtles and hibernating mammals all show stress-induced expression of selected genes (unpublished data) but the lists are never identical and each study revealed at least one novel clone that appears to be specific to a given species. Therefore, delving into the genetic responses of *L. littorea* required the broadest possible approach, first finding inducible genes, a time-

consuming process, and only then moving on to characterization of these genes and their corresponding proteins.

This thesis further demonstrates that regulation of gene expression is an important part of natural stress survival. Through extensive screening of *L. littorea* foot muscle cDNA libraries, I isolated five clones that were up-regulated in the frozen state: actin, myosin, cytochrome c oxidase subunit 2, metallothionein and a novel gene (denoted LL_{GRP}) that was predicted to encode a membrane protein. The library derived from anoxia-treated animals also revealed LL_{GRP} and LL_{MET}, and an anoxia-specific gene denoted $LLAF_{W}$.

The clones described in chapter 4, LL_{GRP} and $LLAF_{W}$, were classified as secretory proteins, a group that encompasses secreted and plasma membrane proteins, which play crucial roles in a variety of physiological and developmental processes. The characteristic N-terminal signal peptide either targets a protein to the appropriate membrane or dictates its secretion from the cell. Strong importance can be attached to the discovery of such proteins, because they often represent genes such as cytokines that play roles in intercellular communication. Systematic cloning of the genes encoding these proteins is of such great interest that a method to effectively trap signal sequences during cDNA library screening has been developed (Tashiro et al., 1993). Signal sequence trap (SST) methods have been used to screen cDNA libraries and isolate only clones encoding N-terminal signal sequences. Tsuruga et al. (2000) identified known and novel genes expressed during differentiation in adipocytes, whereas others have used the technique to isolate and profile the secreted or membrane proteins of *Arabodopsis* (Goo et al., 1999a) and *Drosophila* (Goo et al., 1999b). In each of these studies, over 50% of the isolated

genes encoded novel sequences, illustrating that the SST method can be used to effectively isolate secreted and membrane proteins without knowledge of their functions, and that the majority of these broadly-functioning proteins still remain undiscovered. Given the importance of this class they may make a crucial contribution to natural stress survival. More generally, isolating novel genes demonstrates the importance of differential screening techniques. For example, a clone such as LL_{GRP}, which revealed the most dramatic message accumulation in response to both anoxia and freezing, would remain undiscovered had all messages in the tissue not been examined.

Alteration to the complement of muscle proteins has been shown to occur under several environmental conditions. For example, myosin heavy chain (MHC) expression decreased in rat diaphragm and skeletal muscles under hyberbaric (hypoxic) conditions (Bigard et al., 2000; Mortola and Naso, 1995) while hypertrophic hearts show increased MHC expression (Moalic et al., 1993). These results, as well as data from the ground squirrel heart (Fahlman et al., 2000) illustrate the reorganization of muscle protein expression, most likely in an effort to maintain muscle function. Whereas freezing induced the expression of myosin and actin in the muscle of *L. littorea,* transcript levels were actually highest during the recovery period. This could indicate that unlike the previous examples of altered MHC expression to adjust contractile function, *L. littorea* may increase expression after the stress is removed in order to repair damage incurred to these proteins during freezing.

Injury caused by reactive oxygen species (ROS) is a major source of cellular damage in response to environmental stress. *L. littorea* can be viewed as a model for reperfusion injury because long bouts of anoxia are followed by a sudden availability of

oxygen. Such rapid changes can leave cells unable to cope with ROS, leading to damage. While *L. littorea* does show some alteration in the complement of antioxidant enzymes during and after stress, it shows little evidence of lipid damage (Pannunzio and Storey, 1998). Chapter 3 illustrated the isolation of a putative metallothionein. LL_{MET} was upregulated in response to freezing and anoxia, indicating that limited oxygen availability is a likely trigger to increase its expression. MTs are believed to scavenge ROS in the cell and since marine invertebrates, unlike other species, maintain basal levels of MT transcript, it is well-poised to play a role in the prevention of reperfusion injury.

The marine snail *Littorina littorea* has developed a remarkable tolerance for freezing and anoxia. The microenvironment of the intertidal pool can change several times a day, with wide variations in temperature, salinity, and oxygen exposure. The highly flexible responses to temperature and oxygen variability make this species an excellent model animal for studies of natural stress survival. Future experimental approaches (gene screening, protein expression and specific immunoblotting) combined with ever-advancing proteomics analysis will reveal the full range of changes in gene products that support stress adaptation.

APPENDICES

APPENDIX I

COMMON PROTOCOLS AND SOLUTIONS IN MOLECULAR BIOLOGY

Denaturation of double stranded plasmid for sequencing

A sample of the plasmid (3-5 µg per sequencing reaction) is denatured with 0.1 volume 2 M NaOH/ 20 mM EDTA at 37°C for 30 minutes, then neutralized with 0.1 volume 3 M sodium acetate, and precipitated with 2 volumes of ice cold ethanol. Plasmid precipitation is carried out for 15 minutes at -80°C. The plasmid DNA is pelleted, washed with 70% ethanol and resuspended in 7 µL dH_2O.

Film Exposure

When exposing a membrane hybridized with a ^{32}P-labeled probe, it's important to take into account what the image will be used for. For quantitative measurements (eg. with Imagequant) leave the film somewhat under-exposed, so the developed images will be well-within the linear range of the film. If the goal is a great picture/scan, then aim for a darker image with more contrast, i.e. expose the film for a longer period of time.

Formaldehyde Gel

1. Before running a total RNA formaldehyde gel, the beakers, gel tray, cell and combs are thoroughly washed with 3 % H_2O_2, rinsed in normal ddH_2O and then rinsed with 100 % methanol. Allow all components to air dry. If glassware cannot be treated with DEPC one day prior to use, then heat over a Bunsen burner, both inside and out.

2. Remove 16 µg of total RNA from frozen stock isolation and place in a 0.5 mL Eppendorf tube. This amount of RNA should be in approximately 5 µL of water. If the volume is too low, add DEPC-treated water; if the volume is too high, concentrate the sample down by vacuum desiccation.

3. Melt agarose in DEPC-treated water (3 g of agarose in 186.4 mL of water for large gels) in the microwave, and then leave the melted agarose solution in a 65°C water bath. After cooling to 65°C, add 5 X formaldehyde gel running buffer (50 mL for large gels) and 37 % formaldehyde stock solution (13.6 mL) and gently stir in. Try not to create bubbles while stirring.

4. Pour into the gel mold and let solidify.

5. While the gel is solidifying, premix the sample buffer:
   number of samples X 3 µL of 5 X formaldehyde gel running buffer per sample
   number of samples X 5.25 µL of 37 % formaldehyde per sample
   number of samples X 15 µL of formamide per sample

   Do not place the sample buffer on ice. Formaldehyde in an aqueous solution will precipitate at 4°C.

6. Add 17 µL of the above premixed sample buffer to the 5 µL sample of total RNA and the 0.24-0.95 Kb RNA ladder used as a standard (use 7 µL of the 1 µg/µl stock).

Incubate the samples and standards for 15 minutes at 55°C and then place them immediately on ice. Microfuge at maximum speed for 15 seconds.

7. Pre-run the gel for 10 min at 90 V for large gels (60 V for small gels). This will get rid of any contaminating RNases coming from the gel comb.

8. Add 4 µL of sterile DEPC-treated formaldehyde loading buffer to each sample and standards.

9. Immediately load the samples and standards into the gel.

10. Run the gel in 1 X formaldehyde gel running buffer at a constant voltage (90 V for large gels, 60 V for small gels) until the dye front has migrated 8 cm from the gel wells.

11. After the run is complete, cut off the bottom unused portion of the gel and just above the gel wells. Quickly rinse the gel 3 X in normal ddH_2O before staining. Stain in 400 mL of normal ddH_2O containing 30 µL of a 10 mg/mL stock solution of ethidium bromide. Stain for 30 minutes. Destain in normal ddH_2O for 30 minutes Wash in 400 mL of 10 X SSC for 20 minutes. Take a picture of the gel.

G-50 spun column

This technique is used to remove unincorporated nucleotides from newly synthesized probe.

Create a Sephadex G-50 spun column by pushing a bit of cotton to the bottom of a sterile 1 cc syringe using the plunger of the syringe. Fill the syringe to the top with a thick slurry of Sephadex G-50 (in $T_{10}E_1$ buffer). Place the syringe in a 15 mL screw cap tube and spin at a setting of 7 in a tabletop centrifuge for 1 minute. Dispose of the liquid in the bottom of the tube and replace the spun column. Layer the random primer reaction on top of the Sephadex G-50, and spin as above. Save the eluant containing the labeled probe and dispose of the spun column in the radioactive waste. If using the probe for cDNA library screening, dilute the radioactive probe to 500 µL with $T_{10}E_1$ buffer; if using the probe for a northern blot, do not dilute. Count 1 µL in scintilation counter.

Glycerol stock

Combine 700 µl of fresh bacterial broth culture with 300 µl of sterile 50% glycerol in a 1.5 mL Eppendorf tube. Close the tube and drop into liquid nitrogen to flash-freeze; store at -80°C.

Isolating DNA from an agarose gel slice

Cut band of interest fromo the gel (stained with ethidium bromide and visualized on UV transilluminator box. For the purpose of probe synthesis (labelling the DNA with 32P and random primers) the DNA can be spun from the gel and used directly. Create a sieve tube: punture the bottom of a 500uL ependorf tube with an 18 guage needle, then plug the bottom with a piece of cotton, glass wool or aquarium wool. Place the gel slice in this

tube, cap, and place inside a 1..5 mL tube. Centrifuge at 10000rpm in a microfuge for 3-5 minutes. Discard the sieve tube, and precipitate the DNA. Resuspend the pellet in water ot TE buffer.

Nucleotide quantification
RNA and DNA were quantified using spectrophotometric methods. One absorbance unit at 260 nm correlates to 40 μg/mL of RNA or 50 μg/mL of double stranded DNA (as for plasmids). The ratio of A_{260}/A_{280} demonstrates purity; generally, 1.5-2.0 is acceptable.

Precipitating nucleic acids
Add 0.1 volumes sodium acetate (3 M) and 2 volumes 100% ethanol. Place at -20°C for 1 hour; centrifuge at 10,000 rpm for 10 minutes. Wash pellet with 70% ethanol, spin as before, aspirate off residual alcohol, and dry at room temperature for 5-10 minutes. Resuspend in TE buffer or water.

Washing blots after hybridization with radioactive probes (northerns or plaque lifts).
Blots were washed of excess probe in this order, each for 10 minutes at room temperature:
6 X SSC + 0.2% SDS
2 X SSC + 0.2% SDS
0.5 X SSC + 0.2% SDS
0.2 X SSC + 0.2% SDS
After each wash, the blots were checked with a Geiger counter to determine the number of remaining counts. Washing was stopped when 750-2000 cpm remained. If the blot contained more than 5000cpm after the highest stringency wash, they were washed at 65° C 0.2 X SSC + 0.2% SDS.

IV. Transfer of DNA to hybond for a Southern Blot

Note: In a Southern Blot, the DNA molecules in the gel are double-stranded, so they must be made single stranded in order for the probe to hybridize to them. To do this, the DNA is transferred using a strongly alkaline buffer, which causes the DNA strands to separate - this process is called denaturation - and bind to the filter as single-stranded molecules. RNA an protein are run in the gels in a state that allows the probe to bind without this pre-treatment.

1. Depurination solution: 15 minutes
2. Denaturation solution: 30 minutes
3. Rinse briefly in sterile distilled water.
4. Neutralizing solution: 2 X 15 minutes
5. 20X SSC: 2 X 15 minutes

4. Prepare transfer stack as for a northern blot (MB09B). In brief, fill a buffer reservoir with 20XSSC. Place a support over this tray; place the wick on the support with ends in the SSC. Place the gel on the wick and place saran wrap around the perimeter of the gel to prevent the transfer solution from by-passing the gel. Equilibrate an appropriately-sized piece of nylon membrane in water (1 min) and 20XSSC (5 min) and then place it on

the gel. Remove any bubbles and then place 2 pieces of blotting paper and 4-6 inches of paper towel on top of the membrane. Top off with a weight to ensure contact between the gel and membrane. Transfer overnight.

5. Disassemble the stack. Fix the DNA to the membrane with the UV cross-linker and then allow the blot to air dry on the bench for about 1 hour. Finally, bake at 80° for 2hrs.

Recipes for Part IV

Depurination Solution: 20.8 mls of concentrated HCl in 1 L of H_20 (use ultrapure water). Be careful with the concentrated acid: wear lab coat, gloves, and safety glasses, pour and mix in the fume hood. Add the acid to the water carefully, it will steam a bit.

Denaturing Solution: 20 g of NaOH, 58.4 g of NaCl per liter of ultrapure water. NaOH is caustic, so be careful with it.

Neutralizing Solution: 78.8 g Tris base, 87.6 g NaCl in 800 mls of ultrapure water. Adjust the pH to 7.4. Add ultrapure water to 1 liter.

20X SSC: 175.3 g NaCl, 88.2 g of trisodium citrate; add ultrapure water to 1 liter.

COMMON SOLUTIONS USED IN MOLECULAR BIOLOGY

Ampicillin

Stock solution of 50 mg/mL in sterile dH_2O. Filter through of 0.2 μm filter and store at -20°C in the dark.

Formaldehyde Gel Loading Buffer:

0.72 mL formamide
0.16 mL 10 X MOPS buffer
0.26 mL formaldehyde
0.18 mL DEPC-treated water
0.1 mL 80 % glycerol in DEPC-treated water
0.08 mL bromophenol blue (from a 0.5 % w/v saturated solution)
Store at -20°C.

5 X Formaldehyde Gel Running Buffer:

0.1 M MOPS (pH 7.0)
40 mM sodium acetate
5 mM EDTA
Dissolve 20.6 g MOPS in 800 mL of DEPC-treated 50 mM sodium acetate. Check the pH with a pH meter; adjust to 7.0 with 2 N HCl. Add 10 mL 0.5 M EDTA (pH 8.0). Adjust the volume to 1 L with DEPC-treated water. Pass through a sterile 0.2-micron Millipore filter. Store at room temperature.

50 % Formamide Prehybridization/Hybridization Solution:

Make 50 mL of solution in a sterile 50 mL Falcon tube prewarmed to 42°C. Alternatively, you can make up large batches in sterilized glassware and freeze 45 mL aliquots in Falcon tubes at -20°C. Mix each component in well before adding the next.

12.5 mL of 1 M $NaPO_3$ (pH 7.2)
7 mL of sterile water
2.5 mL of 5 M NaCl
0.1 mL of 0.5 M ethylenediaminetetraacetic acid (pH 8.0)
25 mL of formamide

Then add: 1.0 g of sodium dodecyl sulphate (molecular grade) slowly into solution (42°C) mixing as you add.
Then add: 5 g polyethylene glycol 4000 (molecular grade) slowly into solution (42°C) mixing as you add. Replace at 42°C until dissolved. If a precipitate forms, filter through a hybridization screen. You may freeze solution at -20°C.

GTE (1 L):

Glucose	50 mM	9 g
TRIS-Cl (pH 8)	25 mM	25 mL of 1M solution
EDTA	10 mM	25 mL of 0.5 M solution

Sterilize by filtration (0.2 μm). Store at 4°C.

KAcF (1 L):

K-acetate	3M	294 g
Formic acid	1.8 M	68 mL of a 98% solution

Autoclave and store at 4°C.

LB Agar (1 L)

LB broth
20 g agar
Adjust pH to 7.0 with 5 N NaOH; add deionized water to 1 L; autoclave.

LB Broth (1 L)

10 g NaCl
10 g Select Peptone 140 (bacto-tryptone)
5 g Select yeast extract (bacto-yeast extract)
Adjust pH to 7.0 with 5 N NaOH.; add deionized water to 1 L, autoclave

6 X DNA Loading Buffer:

0.25 % (w/v) bromophenol blue
0.25 % (w/v) xylene cyanol FF
40 % (w/v) sucrose in sterile water
Store at 4°C.

10 X MOPS Buffer

0.2 M MOPS (3-[N-morpholino]propanesulfonic acid)
0.05 M sodium acetate
0.01 M EDTA
Bring to a final pH of 5.5-7.0 with NaOH; do not autoclave; store in the dark.

mRNA isolation solutions

loading buffer: 0.5 M NaCl, 10 mM Tris·HCl, pH 7.5, 0.1% w/v SDS
middle wash: 0.2 M NaCl, 10 mM Tris·HCl, pH 7.5, 0.1% w/v SDS
elution buffer: 2 mM EDTA and 0.1% SDS

NaOH/SDS

0.2 M NaOH and 1% SDS. Store at room temperature. Do **not** store on ice.

NZY Broth: (for 1 L)

5 g NaCl
2 g $MgSO4.7H_2O$
5 g yeast extract
10 g NZ amine (casein hydrolysate)
Adjust pH to 7.5 with 10 N NaOH. Autoclave.

NZY top agar

NZY Broth + 0.7 % w/v agar (top agar)Autoclave, let cool to 55°C then pour into petri dishes
Use 30 mL for a 24x24 cm plate, 3-5 mL for a 10 cm round plate

NZY bottom agar

NZY Broth + 1.5 % w/v agar (bottom agar)Autoclave, let cool to 55°C then pour into petri dishes
Use 200 for a 24x24 cm plate and 5-25 mL for a 10cm round plate

Plaque lift wash solutions:

Denaturation (for 100 ml): 30 ml 5 M NaCl (1.5 M), 2 g NaOH (0.5 M)
Autoclave. Store at room temperature.
Neutralization (for 100 ml): 30 ml 5 M NaCl (1.5 M), 50 ml 1 M Tris-HCl, pH 8 (0.5 M)
Rinse (for 100 ml):20 ml 1 M Tris-HCl, pH 7.5 (0.2 M), 10 ml 20 x SSC (2x)

10 X Random Primer Buffer:

0.5 M Tris-HCl (pH 7.6)
20 mM dithiothreitol (DTT)
50 mM $MgCl_2$
0.4 M KCl
Store at -20°C.

SM buffer:(for 1L)
5.8 g NaCl
2.0 $MgSO_4.7H_2O$
50 ml 1 M Tris-Cl, pH 7.5
5 ml 2 % gelatin
Autoclave. Store at room temperature.

20 X SSC Buffer:

3 M NaCl (175 g/L)
0.3 M Na3.citrate.2 H_2O (88 g/L)

Adjust pH to 7.0 with 1 M HCl. No need to autoclave. Store at room temperature

Stop solution (for sequencing reactions)

(90% formamide, 20 mM EDTA, bromophenol blue and xylene cyanol). temperature.

50 X TAE Buffer:

242 g Tris base
57.1 mL glacial acetic acid
37.2 g $Na_2EDTA.2\ H_2O$
Adjust the pH to approximately 8.5. Store at room temperature.

10x TBE buffer for 1 L:

108 g Tris base
55 g boric acid

9.3 g EDTA

Dissolve in 800 ml of ddH_2O. Check that the pH is approximately 8.3. Adjust as necessary, then bring up to volume. Store at 4°C.

Tetracycline

Stock solution of 25 mg/mL in 100% ethanol. Store at 4°C.

APPENDIX II
GENETIC CODES

The Universal genetic code

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	'CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Initiation Codon: AUG

Invertebrate variations from the universal code:

Codon	Invertebrate code		Standard code	
AGA	Ser	S	Arg	R
AGG	Ser	S	Arg	R
AUA	Met	M	Ile	I
UGA	Trp	W	Ter	*

Comments:

Wolstenholme, 1985; Gadaleta et al., 1988). AUU is not used as an initiator in Mytilus (Hoffmann et al., 1992).

APPENDIX III:

Symbols for amino acids

A	Ala	alanine
B	Asx	asparagine or aspartic acid
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	[henylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine
Z	Glx	glutamine or glutamic acid

APPENDIX IV

ATP^{PR}: program to identify Kozak sequences in DNA sequences;

Help file for kozak sequence analysis results table

No. of ATG from 5' end
A sequence may have many ATG triplets. This information tells you the number of an ATG's occurence relative to the 5' end

Reliability
Although the program works using the score from a linear discriminant function this is difficult for users to interpret. So, instead of showing the user the ldf score, we calculate a reliability score that states the number of times predictions are normally correct when a particular ldf score has been achieved. Specificallyi, the estimate of reliability is derived from our tests using a non-redundant set of 660 sequences.

For example, a reliability of 0.12 means that in our tests the reliability of the ldf score was only 12% and should therefore be treated with the upmost caution.

Frame
Tells the user in which of the three possible frames each ATG was located

Identity to Kozak rule
Kozak was the first to note that real initiation codons have a strong preference for the pattern A/GXXATGG. This section of the table shows the user how close a predicted ATG is to this pattern

Start (bp)
The number of the nucleotide at which the translation starts. ie the A of the initiating ATG.

Finish (bp)
The number of the nucleotide at which the translation finishes. This could be the end of the input sequence. See Stop codon found?

ORG length (aa)
The length of the predicted ORF in amino acids. This is (Finish-Start)/3.

Stop codon found?
Tells the user whether a termination codon was found or whether the program read through to the end of the input sequence

APPENDIX V

SIGNAL P, A PUBLICALLY AVAILABLE WEB-BASED PROGRAM FOR THE PREDICTION OF SECRETATORY SINAL PEPTIDES IN PROTEINS.

Henrik Nielsen, Jacob Engelbrecht, Søren Brunak and Gunnar von Heijne: Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10, 1-6 (1997).

ABSTRACT

We have developed a new method for identification of signal peptides and their cleavage sites based on neural networks trained on separate sets of prokaryotic and eukaryotic sequences. Discrimination between cleaved signal peptides and uncleaved N-terminal signal-anchor sequences is also possible, though with lower precision.

Background: Prediction of signal peptides

With the huge amount of unprocessed sequence data, automatic prediction of protein function and location becomes increasingly important. One important aspect of this is the prediction of secretory signal peptides and location of their cleavage sites. The most widely used method is a weight matrix published by von Heijne in 1986 (Nucleic Acids Res., 14, 4683-4690). The method is still used with the original matrix data today, even though the amount of signal peptide data available has increased by a factor of 5-10 since 1986. Here, we present a method based on a larger and more recent data set, and a more sophisticated computational tool: artificial neural networks. These have been used for many biological sequence analysis problems, including prediction of protein secondary structure and mRNA splice sites. In this approach, we use one network to distinguish between signal peptides and non-signal peptides and another network to recognize the cleavage site.

Signal peptide data

The data were taken from SWISS-PROT version 29. The data sets were divided into prokaryotic and eukaryotic entries, and the prokaryotic data sets were further divided into Gram-positive eubacteria (Firmicutes) and Gram-negative eubacteria (Gracilicutes), excluding Mycoplasma and Archaebacteria. Viral, phage, and organellar proteins were not included. Additionally, two single-species data sets were selected, a human subset of the eukaryotic data, and an *E. coli* subset of the Gram-negative data.

From secretory proteins, the sequence of the signal peptide and the first 30 amino acids of the mature protein were included in the data set. From cytoplasmic and (for the eukaryotes) nuclear proteins, the first 70 amino acids of each sequence were used. Additionally, a set of eukaryotic signal anchor sequences, i.e. N-terminal parts of type II membrane proteins, were extracted. Redundancy in the data sets was avoided by excluding pairs of sequences which were functionally homologous, i.e. where the cleavage site of one signal peptide could be located by simply aligning it to the other. The numbers of non-homologous sequences remaining in the data sets are shown below:

	Number of sequences		
Source	signal peptides	non-secretory proteins	signal anchors
Human	416	251	97
Eukaryotes	1011	820	28
E. coli	105	119	-
Gram-	266	186	-
Gram+	141	64	-

Results: Characteristics of signal peptides

The common structure of signal peptides from various proteins is commonly described as a **positively charged n-region**, followed by a **hydrophobic h-region** and a **neutral but polar c-region. The (-3,-1)-rule states that the residues at positions -3 and -1 (relative to the cleavage site) must be small and neutral for cleavage to occur correctly.**

We have analyzed the characteristics of the new signal peptide data set and displayed them in the form of sequence logos. The differences between signal peptides from different organisms are apparent from the sequence logos:

Results: Predictive performance

SignalP is the most powerful prediction method for signal peptides published. In order to compare the strength of the neural network approach to the weight matrix method, we recalculated new weight matrices from the new data and tested the performance of these. The weight matrix method was comparable to the neural networks when calculating C-score, but was practically unable to solve the S-score problem and therefore did not provide the possibility of calculating the combined Y-score. The ability to distinguish signal anchors from signal peptides has not been evaluated for any of the earlier published methods for signal peptide recognition.

The best prediction of cleavage site location is provided by the position of the Y-score maximum. The best prediction of sequence type (signal peptide or non-secretory protein) is given by the mean S-score (the average of the S-score in the region between position 1 and the position immediately before the Y-score maximum): if mean S-score is larger than 0.5, the sequence is predicted to be a signal peptide.

These prediction performances are minimal values. They are measured on the test sets (i.e. data which were not used to train the networks), and due to the redundancy reduction of the data, the sequence similarity between training and test sets is so low that the correct cleavage sites cannot be found by homology. Consequently, the prediction accuracy on sequences with some degree of homology to the sequences in the data sets will in general be higher. So, the predictions are not based on homology, so it can predict a signal peptide in a protein that is novel. BUT, if the protein does have homolgy to something in the "training" set, it will give a somewhat higher score.

APPENDIX VI

Protein Structure Analysis, program

http://huxley.bu.edu/psa/

Overview

The Protein Sequence Analysis (PSA) server predicts protein secondary and tertiary structure based on sequence, and is available for researchers who have amino acid sequences for proteins of unknown structure and for which no homologous sequences are known. To use PSA, one submits a single amino acid sequence to the server, which may be instructed to analyze the sequence in one of three ways: using Type-1, Type-2, or WD-repeat DSMs. DSMs are Discrete State-space Models for patterns of alpha-helices, strands, tight turns, and loops in specific structural classes.

Type-1 models are for complete sequences from monomeric, single-domain, globular, water-soluble proteins in several recognized structural classes. Now, however, there is an exception, because a set of Type-1 DSMs have been included for transmembrane proteins with a beta-barrel fold like porin. In any case, the Type-1 analysis may be applied to sequences having lengths in the range from 40 to 350 residues. Type-1 DSMs are also appropriate for those subsequences of membrane-spanning proteins that are believed to extend beyond the membrane (based on a hydropathy profile). See the "Description of Type-1 DSMs" section for more information on Type-1 models.

Type-2 models, in contrast, are for either partial or complete sequences from potentially large proteins that violate one or more of the modeling assumptions embodied in Type-1 models. For example, Type-2 models are appropriate for proteins that have one or more of the following properties: (1) they are multimeric; (2) they have more than one structural domain; or (3) they are not globular or soluble (e.g., membrane-spanning proteins). Type-2 models can be applied to sequences up to 1000 residues long. See the "Description of Type-2 DSMs" section for more information on Type-2 models.

References

Ahmed N., Siddiqui A.A., and Ahmed, A.. (2000). DDRT-PCR: use of agarose gels for detection of amplified products. *Mol. Vis.* **31**:144-7.

Alberts B., Bray D., Lewis J., Raff M., Roberts K. and Watson J. D. (1994) *Molecular Biology of the Cell.* Garland Publishing, Inc. New York, NY.

Almendral J.M., Sommer D., Macdonald-Bravo H., Burckhardt J., Perera J., Bravo R. (1988) Complexity of the early genetic response to growth factors in mouse fibroblasts. *Mol Cell Biol.* **8**:2140-8.

Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J.H., Staden R. and Young I.G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**:457-465

Artus N. N., Uemura M., Steponkus P. L., Gilmour S. J., Lin C. and Thomashow M. F. (1996) Constitutive expression of the cold-regulated *Arabidopsis thaliana COR15a* gene affects both chloroplast and protoplast freezing tolerance. *PNAS* **93**:13404-13409.

Baden S.P., Phil L. and Rosenburg, R. (1990) Effects of oxygen depletion on the ecology, blood physiology, and fishery of the norway lobster *Nephrops norvegicus. Mar. Ecol. Prog. Ser.* **67**:141-155.

Baker J., Steel C. and Dure L. (1988) Sequence and characterization of 6 *Lea* proteins and their genes from cotton. *Plant Mol. Biol.* **11**:277-291.

Balk P.A. and de Boer A.D. (1999) Rapid stalk elongation in tulip (*Tulipa gesneriana* L. cv. Apeldoorn) and the combined action of cold-induced invertase and the water-channel protein gamma TIP. *Planta.* **209**:346-54.

Beattie J.H., Black D.J., Wood A.M., Trayhurn P. (1996) Cold-induced expression of the metallothioein-1 gene in brown adipose tissue of rats. *Am. J. Phys.* **270**: R971-977.

Bebianno M.J. and Langston W.J. (1993) Turnover rate of metallothionein and cadmium in *Mytilus edulis. Biometals.* **6**:239-44.

Beelman C.A. and Parker R. (1995) Degradation of mRNA in eukaryotes. *Cell.* **81**:179-83.

Bigard AX, Sanchez H, Birot O, Serrurier B. (2000) Myosin heavy chain composition of skeletal muscles in young rats growing under hypobaric hypoxia conditions. *J Appl Physiol.* 88:479-486.

Bilgen T. (1998) Differential gene expression in two cold-hardy insects in response to low temperatures. Doctoral Thesis. Carleton University.

Binz P.A. and Kägi J.H.R. (1999) Metallothionein: molecular evolution and classification. In: *Metallothionein IV,* Klaassen C. (Ed.), Birkhauser Verlag, p7-13.

Bjorklund S. and Kim Y.J. (1996) Mediator of transcriptional regulation. *Trends Biochem. Sci.* **21**:335-337.

Blobel G. (1980) Intracellular protein topogenesis. *PNAS.* 77:1496-1500.

Boore J.L. and Brown W.M. (1994) Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata. Genetics.* **138**:423-443

Bradford M.. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein using the principle of protein-dye binding. *Anal. Biochem.* 72:248-254.

Brooks S.P.J. and Storey K.B. (1993) *De novo* protein synthesis and protein phosphorylation during anoxia and recovery in the red-eared turtle. *Am. J. Physiol.* **265**: R1380-R1386.

Brouwer M. and Brouwer T.H. (1998) Biochemical defense mechanisms against copper-induced oxidative damage in the blue crab, *Callinectes sapidus. Arch Biochem Biophys.* **351**:257-64.

Bunn H.F. and Poyton R.O. (1996) Oxygen sensing and molecular adaptation to hypoxia. *Physiol Rev.* **76**:839-85.

Burg M. B., Kwon E. D. and Kultz D. (1997) Regulation of gene-expression by hypertonicity. *Annu. Rev. Physiol.* **59**: 437-455.

Cai Q. and Storey K. B. (1996) Anoxia-induced gene expression in turtle heart: Upregulation of mitochondrial genes for NADH-ubiquinone oxidoreductase subunit 5 and cytochrome c oxidase subunit 1. *Eur. J. Biochem.* **241:** 83-92.

Cai Q. and Storey K. B. (1997a) Freezing-induced genes in wood frog (*Rana sylvatica*): fibrinogen upregulation by freezing and dehydration. *Am. J. Physiol.* **272**:R1480-R1492.

Cai Q. and Storey K. B. (1997b) Upregulation of a novel gene by freezing exposure in the freeze-tolerant wood frog (*Rana sylvatica*). *Gene* .**198:** 305-312.

Cai Q., Greenway S., and Storey K. B. (1997) Differential regulation of the mitochondrial ADP/ATP translocase gene in wood frogs under freezing stress. *Biochim Biophys Acta.* **1353**:69-78.

Campagne, M. vL., Thibodeaux, H., von Bruggen, N., Cairns, B., Gerlai., Palmer, J.T., Williams, S.P., and Lowe, D.G. (1999) Evidence for a protective role of metallothionein-1 in focal cerebral ischemia. *PNAS*, **96**: 12870-12875.

Carpene E., and George S.G. (1981) Absorption of cadmium by gills of *Mytilus edulis*(L.) *Mol. Physiol.* **1**:23-24.

Carpene E., Cattani O., Hakim G. and Serrazanetti G.P. (1983) Metallothionein from foot and posterior adductor muscle of *Mytilus galloprovincialis*. *Comp. Biochem. Physiol.* **74C**: 331-336.

Carpene E., Cortesi P., Crisetig G., and Serrazanetti G.P.,. (1980) Cadmium-binding proteins from the mantle of *Mytilus edulis* (L.) *Thal. Jug.* **16**: 317-323.

Chan P.H. (1996) Superoxide radical is a mediator of reperfusion injury. *Stroke.* **27**: 1124-1129.

Cheng C. C. and DeVries A. L. (1991) The role of antifreeze glycopeptides and peptides in the freezing avoidance of cold-water fish. In: *Life Under Extreme Conditions* (de Prisco, G., ed) pp. 1-14, Springer, Berlin.

Cherry J. H. (1994) *Biochemical and Cellular Mechanisms of Stress Tolerance in Plants.* Springer, Berlin.

Chou P.(1989) Prediction of protein structural classes from amino acid composition. In: *Prediction of protein structure and the principles of protein conformation.* (Fassman, G. D., ed.) pp 549-586. Plenum Press, New York.

Churchill T.A. and Storey K.B. (1995) Metabolic effects of dehydration on an aquatic frog, *Rana pipiens. J Exp Biol.* **198**:147-54.

Close T. J., Kortt A. A. and Chandler P. M. (1989) A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. *Plant Mol. Biol.* **13**:95-108.

Cohen, D. M. (1997) Mitogen-activated protein kinase cascades and the signaling of hyperosmotic stress to immediate early genes. *Comp. Biochem. Physiol. A* **117**:291-299.

Crawford, D. R., Wang, Y., Schools, G. P., Kochheiser, J. and Davies, K. J. A. (1997) Down-regulation of mammalian mitochondrial RNAs during oxidative stress. *Free Radic. Biol. Med.* **22**:551-559.

Dallinger, R., Berger, B., Gruber, C, Hunziker, P., and Sturtzenbaum, S. (2000) Metallothioneins in terrestrial invertebrates: structural aspects, biological significance, and implications for their use as biomarkers. *Cell. mol. biol.* **46**:331-346.

Dallinger, R., Berger, B., Hunziker, P.E., Birchler, N., Hauer, C.R., and Kagi, J.H.R. (1993) Purification and primary structure of snail metallothionein. Similarity of the with histones H4 and H2A. *Eur. J. Biochem.* **216**:739-746.

Das, D. K., Maulik, N. and Moraru, I. I. (1995) Gene expression in acute myocardial stress. Induction by hypoxia, ischemia, reperfusion, hyperthermia and oxidative stress. *J. Mol. Cell Cardiol.* **27**:181-193.

Davies PL, Hew CL. (1990) Biochemistry of fish antifreeze proteins. *FASEB J.* **4**:2460-2468.

Davies, P. L., Fletcher, G. L. and Hew, C. L.(1999) Freeze-resistance strategies based on antifreeze proteins. In: *Environmental Stress and Gene Regulation* (Storey, K. B., ed) pp. 61-80, BIOS Scientific Publishers Ltd, Oxford, UK.

deFur, P.L., Mangum, C.P. and Reese, J.E. (1990) Respiratory responses of the blue crab Callinectes sapidus to long term hypoxia. *Biol. Bull. Mar. Biol. Lab., Woods Hole*. **178**:46-54.

Douglas, D.N., Giband, M., Altosaar, I., and Storey, K.B. (1994) Anoxia induces changes in translatable mRNA populations in turtle organs: a possible adaptive strategy for anaerobiosis. *J Comp Physiol B.* **164**:405-14.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. and Burd, C. G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**:289-321.

Duman, J. G., Wu, D. W., Xu, L., Tursman, D. and Olsen, T. M. (1991) Adaptations of insects to subzero temperatures. *Quart. Rev. Biol.* **66**:387-410.

Fahlman, A., Storey, J.M. and Storey, K.B. (2000) Gene Up-regulation in heart during mammalian hibernation. *Cryobiology*. **40**:332-42.

Floyd RA. (1990) Role of oxygen free radicals in carcinogenesis and brain ischemia. *FASEB J.* **4**:2587-2597.

Fowler B.A., Hildebrand C.E., Kojima Y., and Webb M. (1987) Nomenclature of metallothionein. *Experientia Suppl.* **52**:21.

Ge, H. (2000) UPA a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucleic Acids Res.* **15**:28:e3.

George, S.G. and Pirie, B.J.. (1979) The occurrence of cadmium in sub-cellular particles in the kidney of the marine mussel, Mytilus edulis, exposed to calcium. The use of electron microprobe analysis. *Biochim. Biophys. Acta.* **580**: 234-244.

Gleadle, J.M. and Ratcliffe, P.J. (1998) Hypoxia and the regulation of gene expression. Mol. Med. Today. **4**:122-9.

Goo JH, Ahn Y, Park OK, Park WJ (1999a) Selection of Drosophila genes encoding secreted and membrane proteins. *Mol. Cells*. **9**:564-568

Goo JH, Park AR, Park WJ, Park OK. (1999b) Selection of Arabidopsis genes encoding secreted and plasma membrane proteins. *Plant Mol Biol.* **41**:415-23.

Greene, E.A. and Henikoff, S. (1998) Getting more from your sequence on the web. *Nat Genet.* **18**:209.

Haegerman L. (1968) Haemocyanin concentration in the shrimp *Cragnon cragnon* (L.) after exposure to moderate hypoxia. *Comp. Biochem. Physiol.* **85A**:721-724.

Halliwell B. and Gutteridge J.M. (1989) Iron toxicity and oxygen radicals. *Baillieres Clin Haematol.* **2**:195-256.

Hand, S.C., and Hochachka, P.W. (1995) A heme-protein-based oxygen-sensing mechanism controls the expression and suppression of multiple proteins in anoxia-tolerant turtle hepatocytes. *PNAS.* USA. **92**:7505-7509.

Hayes, D.R. and Loomis, S.H. (1985) Evidence for a proteinaceous ice nucleator in the hemolymph of the pulmonate gastropod *Melampus bidentatus. Cryo Lett.* **6**:418-421.

Hershey, L. W. B. (1991) Translational control in mammalian cells. *Annu. Rev. Biochem.* **60**:717-755.

Hightower, L. E. (1991) Heat shock, stress proteins, chaperones, and proteotoxicity: meeting review. *Cell* **66**:191-197.

Hochachka, P. W. and Somero, G, N. (1984) *Biochemical Adaptation.* Princeton University Press, Princeton, NJ.

Hochachka, P.W. and Guppy, M. (1987). Hochachka, P.W. & M. Guppy. Metabolic Arrest And The Control of Biological Time. Harvard Univ. Press, Cambridge, Mass. 237 pp.

Huang, P.C. (1993) Metallothionein structure/function interface. In: *Metallothionein III: Biological roles and medical implications.* Suzuki, K.T., Imura, N., and Kimura, M. (eds.) Birkhauser, Basel. pp. 407-426.

Hunter, T. (1995) Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* **80**:225-236.

Isani, G., Andreani, G., Kindt, M., and Carpene, E. (2000) Metallothioneins (MTs) in marine molluscs. *Cell. Mol. Biol.* **46**: 311-330.

Jacobson, A. and Peltz, S. W. (1996) Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu. Rev. Biochem.* **65:** 693-739.

Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O. and Thomashow, M. F. (1998) *Arabidopsis CBF1* overexpression induces *COR* genes and enhances freezing tolerance. *Science* **280**:104-106.

Jones, P. G. and Inouye, M. (1994) The cold-shock response – a hot topic. *Mol. Microbiol.* **11:** 811-818.

Jurecic R, Nachtman RG, Colicos SM, Belmont JW. (1998) Identification and cloning of differentially expressed genes by long-distance differential display. *Anal Biochem.* **259**: 235–244.

Kaji, and Kojima,. (1987). Chemistry and biochemistry of metallothionein. *Experientia.* **52**: 25-61.

Kang, Y.J.. (1999) The antioxidant function of metallothionein in the heart. *Proc Soc Exp Biol Med.* **222**:263-73.

Kennedy MC, Gan T, Antholine WE, Petering DH. (1993) Metallothionein reacts with Fe^{2+} and NO to form products with A g = 2.039 ESR signal. *Biochem Biophys Res Commun.* **196**:632-5.

Kling P.G. and Olsson, P.. (2000) Involvement of differential metallothionein expression in free radical sensitivity of RTG-2 and CHSE-214 cells. *Free Radic Biol Med.* **28**:1628-37.

Ko, C.M., Bae, S.W., Lee, N.S., Son,Y.H. and Jung, H.M. Direct Submission to NCBI GenBank for *Xenopus laevis* cytoplasmic beta actin. Accession AAC27796

Kondo Y, Rusnak JM, Hoyt DG, Settineri CE, Pitt BR, Lazo JS. (1997) Enhanced apoptosis in metallothionein null cells. *Mol Pharmacol.* **52**:195-201.

Kozak, M.. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**:8125-48

Kyriakis, J. M. and Avruch, J. (1996) Protein kinase cascades activated by stress and inflammatory cytokines. *BioEssays* **18**:567-577.

Kyriakis, J. M., Banerjee, P., Nikolakaki, E., Dai, T., Rubie, E.A., Avruch, J. and Woodgett, J.R. (1994) A MAP kinase subfamily activated by cellular stress and tumor necrosis factor. *Nature* **369**:156-160.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:681-685.

Latchman, D. (1990) *Gene Regulation: A Eukaryotic Perspective*. Unwin Hyman Ltd., London, UK.

Lazo, J.S.m Kondoy, Y., Dellapiazza, D., Michalska, A.E., Choo, K.H.A., and Pitt, B. R. (1995) Enhanced sensitvity to oxidative stress in culteures embryonic cells from transgenic mice deficient in MT I and II genes. *J.B.C.* **270**:5506-5510.

Liang P, and Pardee AB. (1995) Recent advances in differential display. *Curr Opin Immuno.* 7:274–280.

Liang, P., and Pardee, A.B.. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*. **57**:967-71.

Lindquist, S. (1986) The heat-shock response. *Annu. Rev. Biochem.* **55**:1151-1191.

Lindquist, S. and Graig, E. A. (1988) The heat-shock proteins. *Annu. Rev. Genet.* **22**:631-677.

Loomis, S.H. (1985) Seasonal changes in the freezing tolerance of the intertidal pulmonate gastropod *Melampus bidentatus*. *Can. J. Zool.* **63**:2021-2025.

Lorus, J.J., Denome, S.A., and Dunlap, J.C. (1989). Molecular cloning of genes under control of circadian clock in Neurospora. *Science*. **243**:385-388.

Los, D.A. and Murata, N. Responses to cold shock in cyanobacteria. *J. Mol. Microbiol. Biotechnol.* (1999) **1**:221-30.

Martin, I., Vinas, O., Mampel, T., Iglesias, R. and Villarroya, F. (1993) Effects of cold environment on mitochondrial genome expression in the rat: evidence for a tissue-specific increase in the liver, independent of changes in mitochondrial gene abundance. *Biochem. J.* **296**:231-234.

Mason, R.P., Mangum, C.P. and Godette, G. (1983) The influence of inorganic ions and acclimation salinity on hemocyanin oxygen binding in the blue crab, *Callinectes sapidus*. *Biol. Bull. Mar. Biol. Lab., Woods Hole*. **164**:104-123.

Mathews, M. B., Sonenberg, N. and Hershey, J. W. B. (1996) Origins and targets of translational control. In: *Translational Control* (Hershey, J. W. B., Mathews, M. B., and Sonenberg, N., eds) pp. 1-30, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Matsuda, S., Kawasaki, H., Moriguchi, T., Gotoh, Y. and Nishida, E. (1995) Activation of protein kinase cascades by osmotic shock. *J. Biol. Chem.* **270**:12781-12786.

Michiels, C., Arnould, T., and Remacle, J. (2000) Endothelial cell responses to hypoxia: initiation of a cascade of cellular interactions. *Biochim. Biophys. Acta.* . **1497**:1-10.

Mitchell, P. J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**:371-378.

Moalic JM, Charlemagne D, Mansier P, Chevalier B, Swynghedauw B. (1993) Cardiac hypertrophy and failure--a disease of adaptation. Modifications in membrane proteins provide a molecular basis for arrhythmogenicity. *Circulation*. 87:(suppl):IV21-6. Review.

Montoliu, C., Monfort, P., Carrasco, J., Palacios, O., Capdevila, M., Hidalgo, J., and Felipo, V.. (2000) Metallothionein-III prevents glutamate and nitric oxide neurotoxicity in primary cultures of cerebellar neurons. *J. Neurochem*. **75**:266-73.

Morimoto, R. I. (1993) Cell in stress: transcriptional activation of heat shock genes. *Science* **269**:1409-1410.

Morimoto, R. I., Tissiers, A. and Georgopoulos, C. (1994) *The Biology of Heat Shock Proteins and Molecular Chaperones*. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Mortola JP, Naso L. (1995) Electrophoretic analysis of contractile proteins of the diaphragm in chronically hypoxic rats. *Am J Physiol.* **269**:L371-376.

Murphy DJ. (1983) Freezing resistance in intertidal invertebrates. *Annu Rev Physiol.* **45**:289-99.

Neven, L. G., Haskell, D. W., Hofig, A., Li, Q. B. and Guy, C. L. (1993) Characterization of a spinach gene responsive to low temperature and water stress. *Plant Mol. Biol.* **21**:291-305.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G.. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering.* **10**:1-6.

Nordberg M. and Kojima Y. (1979) Experientia. Suppl. **34**:48-55.

Pain, V. M. (1996) Initiation of protein synthesis in eukaryotic cells. *Eur. J. Biochem.* **236**:747-771.

Palmiter, R.D., Findley, S.D., Whitmore, T.E., and Durnam, D.M.. (1992) MT-III, a Brain-Specific Member of the Metallothionein Gene Family. *PNAS* **89**:6333-7.

Palmiter, R.D.. (1998) The elusive function of metallothioneins. *PNAS* **95**:8428-8430.

Pannunzio, T.M. and Storey, K.B. 1998. Antioxidant defenses and lipid peroxidation during anoxia stress and aerobic recovery in the marine gastropod *Littorina littorea. J. Exp. Mar. Biol. Ecol.* **221**: 277-292.

Parsell, D. A. and Lindquist, S. (1993) The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annu. Rev. Genet.* **27**:437-496.

Patwary, M.U., Reith, M. and Kenchington, E.L. (1996) Isolation and characterization of a cDNA encoding an actin gene from sea scallop (*Placopecten magellanicus). J. Shellfish Res.* **15**:265-270

Phadtare, S., Alsina, J., and Inouye, M. (1999) Cold-shock response and cold-shock proteins. *Curr. Opin. Microbiol.* **2**:75-80.

Quaife, C.J., Findley, S.D., Erickson, J.C., Froelick, G.J., Kelly, E.J., Zambrowicz, B.P., and Palmiter, R.D. (1994). Induction of a new metallothionein isoform (MT-IV) occurs during differentitation of stratefied squamous epithelia. *Biochemistry.* **33**: 725-7259.

Riek, R., Precheur, B., Wang, Y., Mackay, E.A., Wider, G., Guntert, P., Liu, A., Kagi, J.H.R., and Wuthrich, K. (1999) NMR Structure of the Sea Urchin (*Strongylocentrotus purpuratus*) metallothionein MTA. *J. Mol. Biol.* **291**:417-428.

Roeder, R. G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**:327-337.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schwarz, M.A., Lazo, J.S., Yalowich, J.C., Allen, W.P., Whitmore, M., Bergonia, H.A., Tzeng, E., Billiar, T.R., Robbins, P.D., and Lancaster, J.R.. (1995) Metallothionein protects against the cytotoxic and DNA-damaging effects of nitric oxide. *PNAS.* **92**:4452-6.

Seger, R. and Krebs, E.G. (1995) The MAPK signaling cascade. *FASEB J.* **9**:726-35.

Semenza, G. L. and Wang, G. L. (1992) A nuclear factor induced by hypoxia via *de novo* protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. *Mol. Cell. Biol.* **12**:5447-5454.

Semenza, G.L.(2000) HIF-1: mediator of physiological and pathophysiological responses to hypoxia. *J. Appl. Physiol.* **88**:1474-80.

Sen, C. K. and Packer, L. (1996) Antioxidant and redox regulation of gene transcription. *FASEB J.* **10**:709-720.

Sharp, P. A. (1994) Split genes and RNA splicing. *Cell* **77**:805-815.

Shih, S.C. and Claffey, K.P. (1998) Hypoxia-mediated regulation of gene expression in mammalian cells. *Int. J. Exp. Pathol.* **79**:347-57.

Shinozaki K. and Yamaguchi-Shinozaki, K. (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Curr. Opin. Plant. Biol.* **3**:217-23.

Smith,J.J. and Bush,G.L. (1997) Phylogeny of the genus *Rhagoletis (Diptera: Tephritidae)* inferred from DNA sequences of mitochondrial cytochrome oxidase II. *Mol. Phylogenet. Evol.* **7**:33-43

Spicer,G.S. and Pitnick,S. (1996) Molecular systematics of the *Drosophila hydei* subgroup as inferred from mitochondrial DNA sequences. *J. Mol. Evol.* **43**:281-286

Storey, K. B. (1999) Stress-induced gene expression in freeze-tolerant and anoxia-tolerant vertebrates. In: *Environmental Stress and Gene Regulation* (Storey, K. B., ed) pp. 1-23, BIOS Scientific Publishers Ltd, Oxford, UK.

Storey, K. B. and Storey, J. M. (1988) Freeze tolerance in animals. *Physiol. Rev.* **68**:27-84.

Storey, K. B. and Storey, J. M. (1990) Facultative metabolic rate depression: molecular regulation and biochemical adaptation in anaerobiosis, hibernation, and estivation. *Quart. Rev. Biol.* **65**: 145-174.

Storey, K.B. (1992). The basis of enzymatic adaptation. In: Fundamentals of Medical Cell Biology, Vol. 3A (Bittar, E., ed.) JAI Press, Greenwich, Connecticut, 137-156.

Storey, K.B. and Storey, J.M. (1992) Natural freeze tolerance in ectothermic vertebrates. *Annu Rev Physiol.* **54**:619-37.

Storey, K.B., Dent, M.E. and Storey, J.M.. (1999) Gene expression during estivation in spadefoot toads, *Scaphiopus couchii*: upregulation of riboflavin binding protein in liver. *J Exp Zool.* **284**:325-33.

Syring RA, Brouwer TH, Brouwer M. (2000) Cloning and sequencing of cDNAs encoding for a novel copper-specific metallothionein and two cadmium-inducible metallothioneins from the blue crab *Callinectes sapidus*. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol.* **125**:325-32.

Tamai KT, Gralla EB, Ellerby LM, Valentine JS, Thiele DJ. (1993) Yeast and mammalian metallothioneins functionally substitute for yeast copper-zinc superoxide dismutase. *PNAS.* **90**:8013-8017.

Tanaka, T., Kondo, S., Iwasa, Y., Hiai, H. and Toyokuni, S. (2000). Expression of stress-response and cell proliferation genes in renal cell carcinoma induced by oxidative stress. *Am. J. Pathol.* **156**:2149-57

Tashiro K., Tada H., Heilker R., Shirozu M., Nakano T., and Honjo T.. (1993) Signal sequence trap: a cloning strategy for secreted proteins and type I membrane proteins. *Science.* **261**:600-603.

Thieringer, H.A., Jones, P.G. and Inouye, M. (1998) Cold shock and adaptation. *BioEssays* **20**:49-57.

Thieringer, H.A., Jones, P.G., Inouye, M. (1998) Cold shock and adaptation. *Bioessays.* **20**:49-57.

Thomashow, M.F. (1993) Genes induced during cold acclimation in higher plants. In: *Advances in Low Temperature Biology* (Steponkus, P. L., ed.) vol. 2, pp. 183-210, London: JAI Press.

Tiku, P.E., Gracey, A.Y., Macartney, A.I., Beynon, R.J., and Cossins, A.R. (1996) Cold-induced expression of ν^9-desaturase in carp by transcriptional and post-translational mechanisms. *Science* **271:** 815-818.

Traystman, R.J., Kirsch, J.R. and Koehler, R.C. (1991) Oxygen radical mechanisms of brain injury following ischemia and reperfusion. *J Appl Physiol.* 71:1185-95.

Tsuruga H, Kumagai H, Kojima T, Kitamura T. (2000) Identification of novel membrane and secreted proteins upregulated during adipocyte differentiation. *Biochem. Biophys. Res. Commun.* **272**:293-297.

Tyshenko, M. G., Doucet, D., Davies, P. L. and Walker, V. K. (1997) The antifreeze potential of the spruce budworm thermal hysteresis protein. *Nature Biotech.* **15**:887-890.

Uptain, S. M., Kane, C. M. and Chamberlin, M. J. (1997) Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **66**:117-172.

Verrijzer, C. P. and Tjian, R. (1996). TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.* **21:** 338-342.

Viarengo, A. Burlando, B., Ceratto, N., and Panfoli, I.. (2000) Antioxidant role of metallothioneins: a comparative overview. *Cell. Mol. Biol.* **46**:407-417.

Viarengo, A., Burlando, B., Cavaletto, M., Marchi, B., Ponzano, E., and Blasco, J. (1999) Role of metallothionein against oxidative stress in the mussel *Mytilus galloprovincialis. Am. J. Phys.*, **277**:R1612-R1619.

Vogel, H. and Jahnig, F. (1986) Models for the structure of outer-membrane proteins of Escherichia coli derived from Raman spectroscopy and prediction methods. *J. Mol. Biol.* **190**:191-199.

Von Heijne, G. (1987) *Sequence Analysis in Molecular biology: Treaseure Trove or Trivial Pursuit?* Academic Press Inc..

Wahle, E. and Keller, W. (1992) The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.* **61**:419-440.

Wang, G.W., Schuschke, D.A. and Kang, Y.J. (1999) Metallothionein-overexpressing neonatal mouse cardiomyocytes are resistant to H_2O_2 toxicity. *Am. J. Physiol.* **276**:H167-75.

Wang, S.M., Fears, S.C., Zhang, L., Chen, J., and Rowley, J.D.. (2000) Screening poly(dA/dT) cDNAs for gene identification. *PNAS* **97**:4162-4167.

Webster, K. A., Gunning, P., Hardeman, E., Wallace, D. C. and Kedes, L. (1990) Coordinate reciprocal trends in glycolytic and mitochondrial transcript accumulations during the in vitro differentiation of human myoblasts. *J. Cell. Physiol.* **142**:566-573.

Wenger, R.H. (2000) Mammalian oxygen sensing, signalling and gene regulation. *J. Exp. Biol.* **203**:1253-63..

Wilding, C.S., Mill,P.J. and Grahame,J. (1999) Partial sequence of the mitochondrial genome of *Littorina saxatilis*: relevance to gastropod phylogenetics. *J. Mol. Evol.* **48**:348-359

Wu, S.W. (1999) *Differential gene expression under environmental stresses in the freeze tolerant wood frog,* Rana sylvatica. Doctoral Thesis. Carleton University.

Yamanaka, K. (1999) Cold shock response in Escherichia coli. *J. Mol. Microbiol. Biotechnol.* 1:193-202.

Yamashita, M., Ojima, N. and Sakamota, T. (1996) Molecular cloning and cold-inducible gene expression of ferritin H subunit isoforms in rainbow trout cells. *J. Biochem. Chem.* **271:** 26908-26913.

Zhang Y. and Frohman M.A. (1997) Using rapid amplification of cDNA ends (RACE) to obtain full-length cDNAs. *Methods Mol Biol.* **69**:61-87.

Zhang, J. and Zhang, L. (1996) Rapid identification of differentially expressed RNA transcripts in apoptotic T lymphocytes. *Immunol Methods.* **195**:113-23.

Zhu, H. and Bunn, H.F. (1999) Oxygen sensing and signaling: impact on the regulation of physiologically important genes. *Respir. Physiol.* **11**:239-47.