2003

Identification and Partial Characterization of a Family of Putative Palmitoyltransferases in Dictyostelium Discoideum

Brent Elliot Wells

IDENTIFICATION AND PARTIAL CHARACTERIZATION OF A FAMILY OF PUTATIVE PALMITOYLTRANSFERASES IN *DICTYOSTELIUM DISCOIDEUM*

By

Brent Elliott Wells

B.S. University of Utah, 1994

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Biochemistry)

The Graduate School

The University of Maine

December, 2003

Advisory Committee:

Robert E. Gundersen, Associate Professor of Biochemistry, Advisor

Keith W. Hutchison, Professor of Biochemistry

Mary Rumpho-Kennedy, Professor of Biochemistry

LIBRARY RIGHTS STATEMENT

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at The University of Maine, I agree that the Library shall make it freely available for inspection. I further agree that permission for "fair use" copying of this thesis for scholarly purposes may be granted by the Librarian. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Signature: Brent E Wells

Date: 12/13/03

IDENTIFICATION AND PARTIAL CHARACTERIZATION OF A FAMILY OF PUTATIVE PALMITOYLTRANSFERASES IN *DICTYOSTLIUM DISCOIDEUM*

By Brent Elliott Wells

Thesis Advisor: Dr. Robert Gundersen

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Biochemistry)
December, 2003

Heterotrimeric guanine nucleotide binding proteins (G-proteins) are essential components of a wide variety of eukaryotic cellular signaling pathways. Heterotrimeric G proteins consist of a 40 kDa α-subunit, a 36 kDa β-subunit and a small 8-10 kDa γ- subunit. Acting as molecular switches, G proteins relay molecular information from membrane bound receptors to downstream intracellular effectors. Most G-proteins require lipid modification by myristic acid and palmitic acid for proper localization and function. Protein palmitoylation is a post-translational, reversible thioester linkage of palmitic acid (C16:0) to an N-terminal cysteine residue of a substrate protein. Palmitoylation of G-proteins occurs specifically on the α subunit. In the slime mold *Dictyostelium discoideum*, the transition from vegetative growth to multicellular development during

starvation has been shown to be dependent upon the G-protein G2. The Gα2 subunit has been shown to be palmitoylated, *in vivo*; however, the mechanism by which this modification occurs has proven to be elusive. Recently, Erf2p, a 41 kDa membrane bound protein, has been identified as a Ras palmitoyltransferase in *Saccharomyces cerevisiae*. Erf2p contains a zinc finger DHHC Cysteine Rich Domain (DHHC-CRD) that has been suggested to be involved in protein-protein or protein-DNA interactions.

To determine whether Erf2p homologs exist in *Dictyostelium*, a BLAST search against the sequenced *Dictyostelium* genome was performed. A family of twelve Erf2p putative homologs was identified within the genome. To determine whether these putative homologs are expressed during the *Dictyostelium* life cycle, RT-PCR was performed using primers designed around the internal DHHC sequence found in each of the newly identified open reading frames. Amplified bands of the expected size in each reaction suggest that all twelve sequences are expressed during the life cycle. Real time PCR experiments showed that several of the open reading frames exhibit peak expression at 8 hours of starvation, making these genes candidates to study the palmitoylation of the Gα2 subunit.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Heterotrimeric G-Proteins

Since their discovery, heterotrimeric guanine nucleotide binding regulatory proteins (G-proteins) have been shown to be essential for a wide variety of cellular activities (reviewed in Gilman, 1987). Processes such as modulation of adenylyl cyclase activity in response to agonist stimulation (*e.g.* epinephrine, gonadotropins, and ACTH) or inhibition (*e.g.* α_2 adrenergic and muscarinic agonists) and control of intracellular cGMP concentrations during the visual phototransduction cascade are known to require heterotrimeric G-proteins to function properly in a wide variety of organisms.

Other cellular activities, in which control by G proteins is implicated but is not as well understood, include the regulation of neuronal Ca^{2+} channels (Hescheler *et al.*, 1987), exocytotic secretory events (Burgoyne, 1987), protein translocation (Connolly and Gilmore, 1986), and phospholipase A_2 activity (Jelsema and Axelrod, 1987). G-proteins provide a bridge between G-protein coupled receptors (GPCRs) and receptor specific downstream effectors. GPCRs are a superfamily of heptahelical receptor proteins that are dependent upon G-proteins to function and are characterized by seven transmembrane domains. These transmembrane domains and their associated intracellular and extracellular loop domains provide a considerable portion of the specificity of extracellular ligand binding and intracellular G-protein association (reviewed in Bourne, 1997). GPCR's are one of the largest superfamilies in the human genome (International Human Genome Sequencing Consortium, 2001), having more identified receptors than there are known ligands to bind to them.

Heterotrimeric G-proteins are comprised of three subunits (Fig. 1): a Gα subunit, a β subunit and a γ subunit. In mammals, there are a large number of isoforms of each of these subunits (23 α subunits, reviewed in Kehrl, 1998; 6β subunits and 12γ subunits, reviewed in Hamm, 1998), and each GPCR interacts with heterotrimers composed of distinct combinations of G-protein subunits.

Four main classes of G-proteins can be grouped together based on all of the known subunits: G_s, which activates adenylyl cyclase; G_i, which inhibits adenylyl cyclase; G_q, which activates phospholipase C; and G_{12} and G_{13}, whose function is currently unknown (reviewed in Hamm, 1998). The 23 known α subunit isoforms are encoded by 17 different genes that are divided into four subfamilies based on primary sequence homology and shared intracellular effector molecules: $G_{i\alpha}$ ($G_{i\alpha1}$, $G_{i\alpha2}$, $G_{i\alpha3}$, $G_{z\alpha}$, $G_{o\alpha1/2}$, $G_{t\alpha}$, and $G_{gust\alpha}$); $G_{q\alpha}$ ($G_{q\alpha}$, $G_{11\alpha}$, $G_{14\alpha}$, and $G_{15/16\alpha}$); $G_{s\alpha}$ ($G_{s\alpha}$ and $G_{olf\alpha}$); and $G_{12\alpha}$ ($G_{12\alpha}$ and $G_{13\alpha}$) (reviewed in Kehrl, 1998).

The β and γ subunits exist either as part of the entire heterotrimer or together as a dimer and function as one unit, despite being encoded by separate genes (Schmidt *et al.*, 1992). These two subunits do not appear to pair with each other indiscriminately, as shown by the fact that particular β subunits prefer specific γ subunits (reviewed in Kerhl, 1998).

Critical to the function of many signaling pathways, the three subunits of heterotrimeric G-proteins act as a molecular switch, relaying information from the activated ligand-bound receptor to the appropriate downstream effectors (*e.g.* adenylyl cyclase, phosphatidylinositol-specific phospholipase C, cGMP phosphodiesterase and ion

Figure 1. Crystal Structure of the Heterotrimeric G-protein Transducin. This figure shows the heterotrimeric G-protein transducin. The α-subunit is blue, the β-subunit is green and the γ-subunit is yellow. The cell membrane would be at the bottom of the figure, where lipid modifications of the N-terminal region of Gα2 and the C-terminal region of the γ subunit occur. It is these two regions where the G-protein is anchored to corresponding GPCRs. Figure was generated using Rasmol. PDB number-1GOT.

channels; Fig. 2). In the inactive state, the three subunits are bound together and form a complex with the inactive receptor. In this state, the α subunit has a GDP molecule tightly bound to its guanine nucleotide-binding site and the βγ subunits serve to stabilize this Gα-GDP conformation. Activity of the heterotrimer is prevented by binding of the GPCR, which obstructs exposure of G-protein active sites to downstream effectors.

G-protein signaling pathways are an integral part of the transmission of an immense selection of molecular information in a large number of eukaryotic systems. The tremendous variety of signaling molecules that GPCRs recognize convey their specific messages to a considerably smaller and less diverse group of G-proteins on the intracellular side of the plasma membrane. Through these relatively few G-protein subunits flows a complex system of intracellular molecular signaling mechanisms that integrate with each other to form a complete organism. Despite the complexity in signaling pathways that exists in eukaryotic systems, there is still significant conservation of key mechanisms within these pathways. It is these mechanisms that form the basis for signal transduction within most eukaryotic organisms.

1.2. Signaling Through Heterotrimeric G-proteins

Heterotrimeric G-proteins play a critical role in determining specificity and temporal characteristics of cellular responses to specific molecular signals. Ligand binding to the GPCR induces a conformational change in the receptor. This conformational change induces activity of the associated intracellular G-protein complex, by initiating a specific series of events. Activated receptor acts as a guanine nucleotide exchange factor (GEF) for the Gα subunit (reviewed in Bourne, 1997). GDP is released

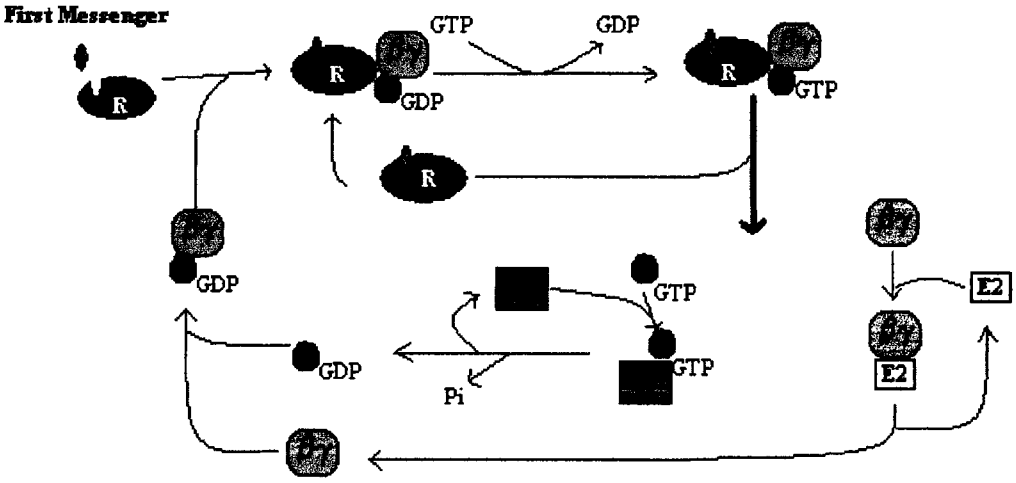


Figure 2. Signaling Mechanism Through Heterotrimeric G-proteins. Binding of extracellular ligand to the receptor stimulates heterotrimeric G-proteins to initiate a signaling cascade. Activation of Gα is dependent upon the presence of a bound GTP molecule. Upon completion of the cascade, Gα binds to the βγ subunit, and the heterotrimer returns to the membrane in the inactive state.

and replaced by GTP, the reaction probably being driven by a higher cytoplasmic concentration of GTP in the cell (Gilman, 1987). Receptor conformational changes and the exchange of GDP for GTP induce a conformation change in the Gα subunit, which lowers the stability of the heterotrimer, inducing βγ complex dissociation from the α subunit. At this moment, the receptor has a low affinity for the ligand. The Gα-GTP also has a low affinity for the GPCR and the complex dissociates. In the active state, both members of the G-protein complex have separate roles in activating downstream effectors and propagating the molecular signal (reviewed in Clapham and Neer, 1997). In addition to regulating specific signal transduction pathways, the Gα subunit also regulates GTP hydrolysis, which functions to control re-association of the Gα subunit with the βγ complex.

Deactivation of the heterotrimeric G-protein complex, resulting in termination of the corresponding signal cascade, appears to be regulated by the GTPase activity of the Gα-GTP subunit (reviewed in Casey and Gilman, 1988). The lifetime of various Gα-GTP complexes is in the range of tens of seconds, thereby stimulating prolonged amplification of downstream effector products.

1.3. Dictyostelium discoideum as a Model System

1.3.1. Dictyostelium discoideum Development

Dictyostelium discoideum is a primitive, haploid, eukaryotic organism that has become a useful and popular model system for study in the laboratory (reviewed in Kessin, 2001). It is a cellular slime mold that is found ubiquitously in the soil where the

primary food source is bacteria. Modern experimental work with this organism was first done with strains isolated from decaying leaves in forests near Asheville, North Carolina (Raper, 1935). During times of sufficient food availability, individual amoebae in the vegetative state are generally motile and chemotax toward sources of food by extending pseudopods in the direction of folate gradients that are by-products of bacterial metabolism (reviewed in Kay, 2002). Upon starvation, individual cells emit extracellular pulses of cAMP as a chemotactic stimulus to other cells in the vicinity (Van Haastert, 1995). Receiving cells relay this stimulus by emitting their own cAMP pulse into the environment, thus propagating the chemotactic signal. This signaling process initiates a transition from the unicellular vegetative stage to an aggregation stage, where cells migrate toward the source of the signal. Aggregation continues until a multicellular mound is formed. The mound continues to develop into a mobile slug and finally, into a fruiting body with the entire developmental cycle taking approximately twenty-four hours. A diagram of the *Dictyostelium* life cycle can be found at http://www.uni-kassel.de/fb19/genetics/projects/devcyc.gif (Kassel University Department of Genetics) During the developmental life cycle, gene expression within individual cells changes and the cells differentiate into functionally different pre-spore and pre-stalk cells. The fruiting body, which is the culmination of the life cycle, consists of a stalk of dead cells, which supports a sac of spores. Spores are heat and desiccant resistant and, in their natural habitat, the stalk extends out from the soil, enabling the spores to be passively moved to other sites by attaching to mobile organisms in the soil.

Watts and Ashworth (1970) developed useful strains of *Dictyostelium* that can survive on nutrient broth mixture (axenic strains). Starvation can easily be induced in

these strains by switching to a non-nutrient buffered medium and the resulting biochemical alterations can readily be investigated in the laboratory. *In vivo*, characteristic chemotaxis up a cAMP gradient during aggregation is completely dependent upon the extracellular signaling processes that require G-proteins (Coukell *et al.*, 1983; Kesbeke *et al.*, 1988).

1.3.2. Dictyostelium discoideum Gα2 subunit

At present, eight distinct Gα subunits have been cloned in *Dictyostelium* (Wu and Devreotes, 1991). Based on sequence comparison within the sequenced genome, it is thought that other Gα subunits probably exist (Parent and Devreotes, 1996). Among the eight Gα subunits, there is a 30-35% amino acid sequence identity and a comparable amount of identity exists between *Dictyostelium* and mammalian Gα subunits. The primary mammalian Gα subunits, $G\alpha_i$, $G\alpha_s$, and $G\alpha_q$ (Simon *et al.*, 1991) do not show any direct correlation to any of the *Dictyostelium* Gα subunits, but there are several amino acids in critical areas of the proteins that are identical and identify these proteins as Gα subunits.

One of the eight Gα subunits in *Dictyostelium discoideum* is Gα2, which is coupled to the GPCR cyclic AMP receptor (cAR)-1. Activation of downstream effectors during development is explicitly dependent upon expression of Gα2, as evidenced by the failure of Gα2-null cells to respond to cAMP mediated stimulation *in vivo*, resulting in cell arrest in the initial phase of development (Kumagai *et al.*, 1989). The relationship between cAR-1, Gα2, the βγ subunits and several of their downstream effectors

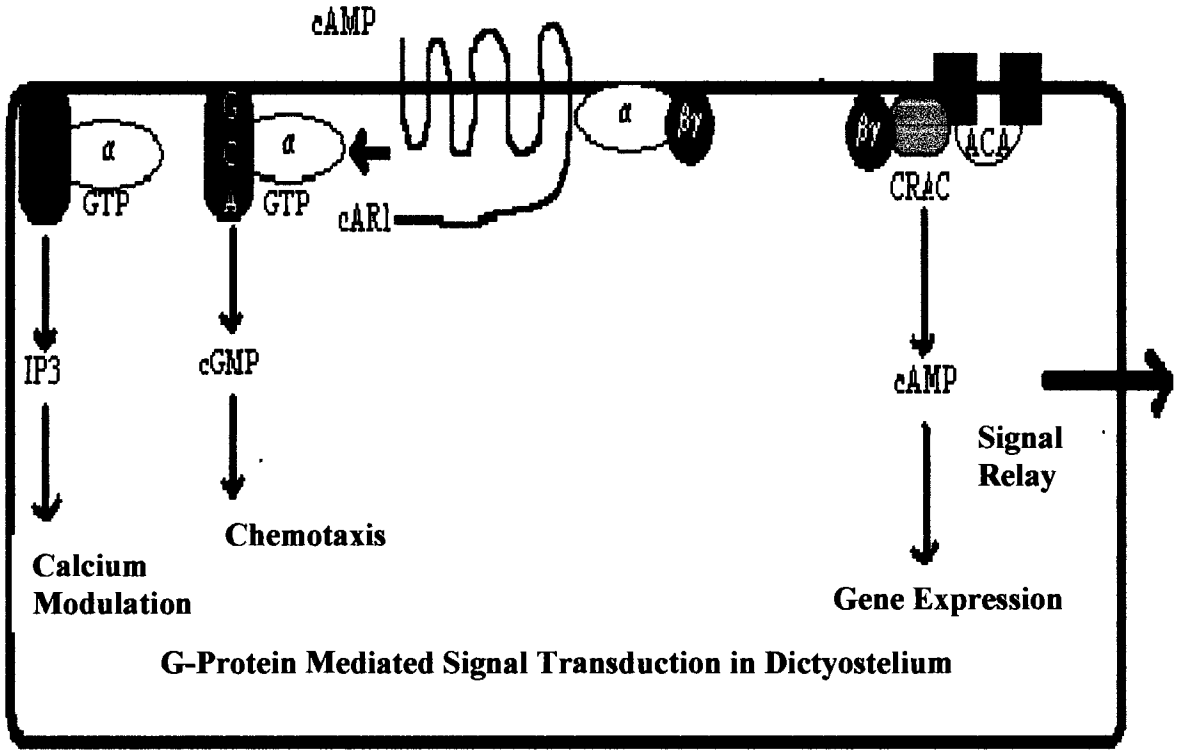


Figure 3. Gα2 Mediated Pathways in *Dictyostelium discoideum.* Both Gα2 and the βγ subunits have distinct and critical roles in the life cycle of *Dictyostelium discoideum.* Upon activation, Gα2 mediates chemotaxis and Ca^{+2} modulation. Simultaneously, the βγ subunits regulate gene expression and cAMP relay to the environment.

(reviewed in Kessin, 2001) is outlined in Fig. 3. Upon activation by cAMP, cAR-1 stimulates activation of G2. Through separation of the α subunit from the βγ complex, each of these subunits is able to initiate distinct signaling pathways. The βγ complex binds to cytosolic regulator of adenylyl cyclase (CRAC), which activates adenylyl cyclase (AC). This results in production of cAMP, which regulates developmental gene expression and initiates a signal relay to other cells in the vicinity. The activated α subunit, complexed to a GTP molecule, has a dual purpose once activated. Not only does this subunit activate Phospholipase C, thereby modulating intracellular Ca^{+2} concentrations, but it may also activate guanylyl cyclase (GCA), which catalyzes synthesis of cGMP. cGMP is critical to the chemotaxis response upon cAMP stimulation.

1.4. Lipid Modifications of Cellular Substrates

Although receptor-catalyzed guanine nucleotide exchange and α subunit GTP hydrolysis are the most well understood means of G-protein regulation, covalent modifications of heterotrimeric G-proteins have also been shown to provide additional signaling regulation levels. Folch and Lees (1951) were the first to describe covalent lipid attachment to proteins. The requirement for lipid modification of many essential components of signaling pathways is now known to be maintained across eukaryotic species. In the more than fifty years since Folch and Lees' discovery, advances in signal transduction research have led to the identification of several common lipid modifications, such as palmitoylation (Schmidt *et al.*, 1979), myristoylation (Carr *et al.*, 1982) and prenylation (Hancock *et al.*, 1989, Casey *et al.*, 1989). Many members of the G-protein superfamily of proteins are substrates for a number of covalent modifications.

Among these modifications are ADP-ribosylation, prenylation, acylation and phosphorylation (reviewed in Yamane and Fung, 1993; Fig. 4). Each of these different modifications have a critical function in mediating the interactions of the specific substrate G-protein with other proteins, as well as maintaining contact with the appropriate membrane compartments. A review of palmitoylation, myristoylation and prenylation, as they relate to G-proteins follows.

More than 200 proteins, including viral and cellular proteins (Casey, 1995; Milligan *et al.,* 1995; Resh, 1994), are now known to be fatty acylated in some fashion. All G-protein α subunits are modified in the N-terminal region by covalent attachment of myristate and/or palmitate (Fig. 4). Myristoylation is the irreversible addition of a saturated fourteen-carbon myristate molecule to a glycine residue at the N-terminal region of the protein. Although myristoylation is a co-translational event, it first requires that the start methionine be cleaved. A stable amide bond links the myristate molecule to the protein. Six Gα subunits are known to be myristoylated ($G\alpha_{i1}$, $G\alpha_{i2}$, $G\alpha_{i3}$, $G\alpha_{o}$, $G\alpha_{z}$ and $G\alpha_{t}$).

G-protein γ subunits are covalently modified by addition of the twenty-carbon isoprenoid geranylgeranyl (Fig. 4), although in the case of retinal specific γ_1, modification is carried out by the fifteen-carbon isoprenoid farnesyl (reviewed in Wedegaertner *et al.*, 1995). As is the case with all prenylated proteins, G-proteins are prenylated by a stable thioether linkage of either geranylgeranyl or farnesyl moieties to a cysteine residue located at, or near, the C-terminal "CAAX" box of the subunit. This modification is followed by proteolytic removal of the C-terminal three amino acids and carboxyl methylation at the new C-terminus.

Myristoylation and Palmitoylation

α_s	**M G C L G N S K T E D Q R N E**	**Palmitate**
α_{i1}	**M G C T L S A E D K A A V E R**	**Myristate, Palmitate**
α_t	**M G A G A S A E E K H S R E L**	**Myristate**
α_q	**M T L E S I M A C C L S E E A K A R R I N**	**Palmitate**

Prenylation

γ_1	**K G I P E D K N P F K E L K G G C V I S**	**Farnesyl**
γ_2	**T P V P A S E N P F R E K K F F C A I L**	**Geranylgeranyl**

Figure 4. Sites of Lipid Modifications in G Proteins. The N-terminal sequences of several G protein α subunits and the C-terminal sequences of two G protein γ subunits are shown. Myristoylated glycines are in blue, palmitoylated cysteines are in red, farnesylated cysteines are in green and geranylgeranylated cysteines are in orange.

Although farnesylation and geranylgeranylation of the γ subunit is not required for coupling with the β subunit, studies have indicated that a proteolytically truncated and isoprenylated γ subunit will not dimerize with the β subunit, indicating that there is a specific order to the modifications that must be followed for dimerization to occur (Higgins and Casey, 1994).

1.4.1. Palmitoylation of Cellular Substrates

Schmidt and co-workers (1979) found that palmitate was attached to cysteine residues in the glycoproteins of Sindbis virus via a thioester linkage. In the years since their work, research on the G-proteins has shown that all Gα subunits investigated, except $G\alpha_t$, are palmitoylated by way of a labile, reversible thioester linkage to a cysteine near the N-termini (reviewed in Wedegaertner *et al.*, 1995). As more types of lipid modifications were described, evidence indicated that the thioester linkage is a feature characteristic and distinctive to palmitoylation (Magee and Courtneidge, 1985). As a result, the terms "S-acylation" and "thioacylation" are not uncommon when referring to palmitoylation, although these terms could technically refer to isoprenylation. While a more appropriate name for this lipid modification would be acyl thioesterification or thioesteracylation, palmitoylation is most commonly used.

Palmitoylation is a post-translational and dynamic modification (McIlhinney *et al.*, 1985, Magee *et al.*, 1987). Dynamic palmitoylation has been demonstrated in a number of instances and for a number of GPCRs, G-proteins and other signaling proteins (Fig. 5). Because a protein can undergo multiple palmitoylation and depalmitoylation cycles during its lifetime, the hypothesis that palmitoylation may serve as a regulatory

Protein	Myristate	Palmitate	Isoprenoid	Modified Sequence
Non-receptor tyrosine kinases				
p60 src	+	-	-	**M G S...**
p56 lck	+	+	-	**M G C V C...**
p59 fyn	+	+	-	**M G C V Q C...**
p59 hck	+	+	-	**M G C...**
p61 hck	+	-	-	**M G G...**
Ras proteins				
Ha-Ras	-	+	C15	**...G C M S C K C V L S**
N-Ras	-	+	C15	**...G C M G L P C V V M**
K(B)-Ras	-	-	C15	**...K K K S K T K C V I M**
Other proteins				
PSD-95	-	+	-	**M D C L C I T T...**
AKAP18 (AKAP15)	+	+	-	**M G Q L C C...**
Vac8p	+	+	-	**M G S C C S C...**
SNAP-25	-	+	-	**...K F C G L C V C P C N K L**
GAP-43	-	+	-	**M L C C M...**

Figure 5. Selected Other Lipid Modified Proteins. Myristoylated glycines are blue; palmitoylated cysteines are red; isoprenylated cysteines are underlined.

element has been developed (reviewed in Mumby, 1997). This theory is given support by studies that have shown that a) certain stimuli can induce an increase in palmitate turnover (*e.g.* James and Olson, 1989), b) in some systems, palmitoylation is developmentally regulated (Melendez and Bizzozero, 1996; Patterson and Skene, 1999; Veit *et al.* 2000), c) palmitoylation of certain proteins requires the cell cycle (Mundy and Warren, 1992), and d) palmitoylation can occur in anucleated cells (erythrocytes; Staufenbiel, 1988) and platelets (Seehafer *et al.*, 1988; Huang, 1989).

Regardless of the type of fatty acid involved in a particular modification, the fatty acid must first be activated before is can be available for use by the cell. A common method of fatty acid activation involves the formation of a high-energy thioester linkage of the fatty acid to coenzyme A (CoA). This ATP driven reaction is performed by long chain fatty acyl ligase (or synthase).

Much of the evidence supporting the notion that palmitoyl CoA is the *in vivo* donor of palmitate comes from *in vitro* studies with cell free systems (Slomiany *et al.*, 1984; Berger and Schmidt, 1984). However, separate experiments have shown that palmitoylated proteins, themselves, can serve as the donor of palmitate for other substrate proteins (Quesnel and Silvius, 1994). This suggests that the underlying mechanism for some palmitoylation reactions may involve transacylation.

1.4.2. Functions of Palmitoylation

When discussing the possible functions of palmitoylation, a distinction must be made between constitutive and dynamic palmitoylation. In this thesis, constitutive palmitoylation will refer to the original palmitoylation that takes place prior to the

substrate protein reaching its position at the membrane, while dynamic palmitoylation will refer to the regulation of the acylated state in the substrate protein once it has begun functioning within a signaling pathway.

Initial palmitoylation of GPCRs has been shown to occur in or around the Golgi complex (Bradbury *et al.*, 1997; Petaja-Repo and Bouvier, unpublished observations). Several studies have suggested that constitutive palmitoylation may play a role in processing and targeting the newly made GPCRs from the Golgi to the plasma membrane (reviewed in Qanbar and Bouvier, 2003). In several instances, palmitoylation is suggested to play an important role in expression of receptors on the cell surface (Karnik *et al.*, 1993; Zhu *et al.*, 1995; Fukushima *et al.*, 2001 and Percherancier *et al.*, 2001).

Numerous studies have also suggested that dynamic palmitoylation of GPCRs may serve a role separate from constitutive palmitoylation. The reversibility of this modification may allow it to play a role in the modulation of receptor activity. A majority of the information that has been acquired concerning dynamic palmitoylation comes as a result of mutation studies, metabolic labeling, and chemical or enzymatic depalmitoylation studies (reviewed in Qanbar and Bouvier, 2003). Changes in palmitate half-life on GPCRs and G proteins under differing conditions further suggest a functional importance for reversible palmitoylation. In addition, the loss of a particular enzyme function as a result of depalmitoylation or mutation of a known palmitoylation site also lends credence to the hypothesis that dynamic palmitoylation has functional importance to the substrate protein. It must be acknowledged, though, that the functional loss due to mutation of a palmitoylation site may be due to the actual loss of the cysteine residue, itself, rather than the loss of the palmitate molecule, as the cysteine sulfhydryl itself can

be necessary for some protein functions (reviewed in Qanbar and Bouvier, 2003). Because of this important consideration, further study of dynamic palmitoylation is necessary to more fully understand its purpose in the cell.

1.5. Identification of Palmitoyltransferases

In the more than two decades since Schmidt and co-workers made the discovery of the palmitoylation of Sindbis virus glycoproteins, much work has been done to elucidate the mechanism by which this modification takes place. While this work has produced many advances, such as *in vitro* palmitoylation assays, partial characterization of palmitoylation activities and the purification to homogeneity of several candidate proteins, until recently, this effort had failed to characterize, definitively, any specific palmitoyltransferases (PAT). Discussion of recently identified PATs will take place later in this paper.

Despite all of the evidence discussed previously for the existence of distinct palmitoyltransferases, the fact that none had been identified led many to speculate that spontaneous palmitoylation occurred, *in vivo*. Lending credibility to this hypothesis is the observation that spontaneous transfer of palmitate can occur to myelin proteolipid protein (PLP) (Bizzozero *et al.*, 1987; Ross and Braun, 1988) and a splice variant of PLP, called DM20 (Yoshimura *et al.*, 1987). Since these observations, autoacylation has been observed in many experimental systems with many different substrate proteins.

Even though autoacylation has been shown to take place without a catalyzing enzyme, this reaction does not take place randomly on any cysteine in any known amino acid sequence. Some studies have shown that the presence of specific amino acids

bordering palmitoylated cysteines influence the rate and efficiency of the palmitoylation reaction (Bizzozero *et al.*, 2001; Belanger *et al.,* 2001) with some amino acids promoting palmitoylation and others inhibiting it. These observations, along with the fact that long chain fatty acid selectivity is maintained by enzyme-free reactions, suggest that autoacylation does occur, although not without influence from amino acids surrounding the reaction site. The physiological prevalence and importance of autoacylation is unclear at this time.

Despite indications of *in vivo* autoacylation and the lack of any specifically characterized palmitoyltransferases, much evidence has been acquired to suggest their existence. Work done over the decades indicates that the palmitoylation reaction is susceptible to inhibitors and that the putative activity is membrane enriched, heat sensitive and increases turnover of substrate palmitoylation (reviewed in Qanbar and Bouvier, 2003). While it was apparent that the reaction is indeed enzyme catalyzed, the identity of the responsible enzyme(s) remained unclear for a long time. Partly responsible for this is the fact that no consensus sequence has been found at palmitoylation sites, suggesting that more than one palmitoyltransferase exists. The N termini of 7 of the 9 members of the Src family of protein tyrosine kinases and several α subunit isoforms of heterotrimeric G-proteins, however, contain the sequence myrGlyCys, where the Cys-3 is palmitoylated (Milligan *et al.,* 1995; Resh, 1994).

The enzymes that catalyze myristoylation and prenylation (*i.e.* prenyl and myristoyl transferases) are well characterized and have been shown to be conserved in many species, from yeast to humans (Roth *et al.,* 2002). These enzymes have become appealing candidates as drug targets, such as anticancer agents. In theory, because many

significant signaling processes depend upon them, palmitoyltransferases should provide similar drug targeting opportunities that other, better understood, lipid modification enzymes now present. Because of this, it is critical to obtain a better understanding of the enzymes that are involved in this modification.

Several suspected PAT activities have been isolated, but a lack of further study of these proteins, thus far, has left identification of their specific roles in palmitoylation incomplete. Thiolase A, an enzyme required for fatty acid β-oxidation (Liu *et al.,* 1996; Liu *et al.,* 1999) was purified and identified as a PAT of mammalian H-Ras, but this enzyme is localized to peroxisomes, which makes it an unlikely candidate as a cytoplasmic acyltransferase. PAT activities using substrates such as viral glycoproteins, non-receptor tyrosine kinase $p59^{fyn}$ or G proteins have all been detergent solubilized, but instability of the isolated activity has made further study difficult (Chamoun *et al.,* 2001; Lee and Treisman, 2001; Bizzozero *et al.,* 1987).

Chamoun *et al.* (2001) have recently identified skinny hedgehog, in *Drosophila*, as a palmitoyltransferase of the Hedgehog family of proteins. While this discovery is interesting, the Hedgehog family of proteins is extracellular in nature and the mechanism of palmitoylation of this enzyme is fundamentally different from palmitoylation of most intracellular proteins, as evidenced by the fact that Hedgehog proteins are palmitoylated by an amide linkage of the palmitate to the substrate protein. Most palmitoylation of intracellular proteins, including heterotrimeric G-proteins and Ras, occurs via a thioester linkage, and therefore, the palmitoyltransferase(s) that function with these substrates must act via a mechanism that results in a thioester linkage.

During the latter part of 2002, two separate laboratories published results indicating that a potentially extensive family of putative palmitoyltransferases exists in a wide variety of eukaryotic species. Through the use of genetic screens, two different gene products that affect Ras palmitoylation in *Saccharomyces cerevisiae* were identified. Shr5 (Jung *et al.,* 1995), a hydrophilic 26.5 kD protein with no significant sequence homology and Erf2p (Bartels *et al.,* 1999), a 41kD protein with four predicted transmembrane domains, have been identified as necessary for palmitoylation in yeast. Both of these proteins form a complex of unknown stoichiometry. Lobo *et al.* (2002) showed that Erf2p, based on its ability to form an Asp-His-His-Cys-Cysteine Rich Domain (DHHC-CRD) dependent acyl-intermediate, is the enzyme that catalyzes the transfer of palmitate to Ras, while the function of Shr5 is much less understood at this time. To this point, the putative metazoan homologs of Erf2p have all been identified based upon the existence of the well-conserved DHHC-CRD motif. This has lead to speculation that other members of the DHHC-CRD family may also be palmitoyltransferases (Bartels *et al.,* 1999).

Shr5, also known as Erf4p, has been shown to be necessary for stable expression or solubilization of Erf2p from yeast cells, although the exact function of Erf4p remains elusive. It is possible that Erf4p acts as a chaperone for Erf2p, although other functions in mediating palmitoylation are being examined (Lobo *et al.,* 2002). Erf4p does not have any clearly defined domains that can be used to predict function, therefore making investigation of its function more difficult. Erf4p homologs have been found in some fungal genomes (Lobo *et al.,* 2002), however, it has not been possible to identify metazoan homologs based on sequence identity, alone.

In addition to the above proteins, another gene product required for palmitoylation in yeast has been identified. Akr1p, an 86 kD protein with six predicted transmembrane domains, has been demonstrated to be required for proper localization of type I casein kinase Yck2p to the plasma membrane *in vitro* (Feng and Davis, 2000), although it is not known whether additional proteins are required for palmitoylation of this substrate, *in vivo*. In addition, Akr1p, itself, becomes palmitoylated, suggesting a role for an Akr1p/palmitate intermediate in the reaction mechanism. Mutations in the DHHC-CRD region of this protein eliminate both trans- and auto-palmitoylation activities, indicating a critical role for this domain in the reaction. This is in agreement with the studies by Lobo *et al.* (2002), which also specify important roles for the DHHC-CRD region in Erf2p palmitoylation of Ras.

Both Akr1p and YOL003c can palmitoylate Ras2p (Zhou and Deschenes, unpublished observations), although at a lower efficiency than Erf2p (Lobo *et al.,* 2002). In all probability, both of these proteins can palmitoylate other substrates, *in vivo*. It will be interesting to determine whether the metazoan DHHC-CRD proteins thus far identified also possess palmitoylation capabilities. For the most part, little is known about the functional abilities of these other DHHC-CRD proteins (Putilina *et al.,* 1999), however, a c-Abl interacting protein (Aph2) containing a DHHC-CRD motif has been identified, using a two hybrid screening process (Li *et al.,* 2002). Co-expression of Aph-2 with c-Abl leads to induction of apoptosis in COS-7 cells (Lobo *et al.,* 2002).

Sequence homology between Erf2p and Akr1 is limited to a 50 base pair long internal region of both proteins called the DHHC-CRD. This highly conserved region is defined by a core Asp-His-His-Cys (DHHC) tetrapeptide sequence and multiple cysteine

Table 1. Putative Acyltransferase Sequence Homology. Numerous proteins with DHHC-zinc finger domains have been identified in a variety of species, although the function of many of these proteins is still unclear. While the sequence immediately surrounding the DHHC shows a high degree of homology to other zinc finger proteins, the rest of the sequence throughout the proteins is highly variable.

Q9D419	MOUSE	PPRTKHCSRCNFCVDDFDHHCKWLNNCIGGRNYRPFLMLV
PSLA	YEAST	PPRSKHCSICNRCVMVMDHHCIWINNCIGKGNYLQFYLFL
Q9H923	HUMAN	PPRCSHCSICNNCVERFDHHCPWVGNCIGRRNYRYFFMFI
Q9BLV0	LEIMA	PYRSKHCKYCNSCVRRYDHHCPWTGNCIGAKNMRYYLVFL
Q9NPG8	HUMAN	PARSKHCSICNRCVMRFDHHCGWMNNCIGKWNTRYFMIYL
Q9VBM8	DROME	PPRSWHCPLCKMCVLRKDHHCFFTGACVGMGNQRCFLLFL
Y041	CAEEL	PDRAMHCRVCGRCVLKMDHHCPWINNCVGYSNYKYFILFL
Consensus		P+R++HC++C++CV+++DHHC+Wh+NChG++N+++FhhFL

residues on either side of the region (Table 1). The DHHC-CRD, which is a variant of the C_2H_2 zinc finger domain (Putilina *et al.,* 1999), is found in Erf2p between transmembrane domains 2 and 3 and in the Akr1 sequence, with a single amino acid variation (DHYC), between transmembrane domains 4 and 5. The functional significance of the second histidine being changed to a tyrosine, in Akr1p, is not known at this time.

This highly conserved region, also called NEW1 or a zf-DHHC domain (PF01529) is found in a large family of membrane bound proteins in eukaryotes, ranging from single cell species to humans (Bohm *et al.,* 1997; Putilina *et al.,* 1999). Other genes that encode DHHC-CRD proteins in yeast include AKR2 (Dujon *et al.*, 1997), PSL10 (Murphy *et al.*, direct submission to Swiss-Prot., 1995), YOL003 (Dujon *et al.*, 1997), and YNL326c (Maftahi *et al.*, 1995).

In higher eukaryotes, DHHC-CRD proteins have been implicated in the pathology of Huntington's disease (Faber *et al.*, 1998), leukemia (Ruault *et al.*, 2002) and colorectal cancer with liver metastasis in humans (Oyama *et al.*, 2000). Additionally, this region has been shown to be involved in FSH activation of testicular Sertoli cells in rats (Chaudhary and Skinner, 2002).

While the function of this highly conserved domain within each of these systems is not completely understood at this time, it is thought that the zinc finger domain is involved in protein-protein or DNA-protein interactions (Putilina *et al.,* 1999). In addition, Lobo *et al.* (2002) found that mutations within conserved residues C189, H201 and C203 of the DHHC-CRD all abolished Ras PAT activity, suggesting that this region is involved in palmitoylation of substrates. Further site-directed mutagenesis studies are

needed to distinguish the specific roles of each of these conserved residues, as well as to determine other residues are necessary for palmitoylation activity.

At this time, the mechanisms by which Erf2p/Erf4p and Akr1 act to palmitoylate specific substrates are not well understood. It is apparent, however, that the Erf2p/Er4p complex does exhibit a preference for yeast Ras substrates, rather than multiple substrate proteins. Limitations in some of the assays used in recent experiments (sub-saturating concentrations of Ras substrate) and the difficulty of estimating the amount of active Er2p/Er4p in GST pull-down experiments have made resolving Erf2p/Erf4p mechanistic issues impossible, to date. Future analysis will require obtaining Erf2p/Erf4p complexes that can better be used in mechanistic studies (Lobo *et al.*, 2002).

1.6. Mechanism of Depalmitoylation

While discovery and characterization of specific palmitoyltransferases is only now beginning, the existence of acyl protein thioesterases (APTs) has been known for several years. Just as palmitoylation results in the formation of a thioester linkage to the substrate protein, depalmitoylation is the result of the cleavage of that thioester linkage by an APT. During activation of G proteins, it has been proposed that the α/GTP complex must be depalmitoylated for it to be released from the plasma membrane (Wedegaertner *et al.*, 1995; Wedegaertner and Bourne, 1994). This would allow for the α subunit to activate the next enzyme in the signal cascade. Evidence contradicting this hypothesis, though, has been shown by Huang *et al.*, (1999), who were unable to detect substantial release of G protein α subunit from rhesus monkey kidney cells upon depalmitoylation of the α subunit. They suggest that palmitoylation of the α subunit may

actually serve to regulate interactions with other proteins by potentially promoting interactions with the βγ subunit complex, by decreasing affinity for regulators of G protein signaling and/or by modulating other interactions that have not yet been characterized. These differing conclusions stand to emphasize the importance of further study of this critical signaling process.

The acyl protein thioesterase family of enzymes was first identified by Berger and Schmidt (1986), who observed that membranes from different sources were able to catalyze depalmitoylation of a variety of substrates. The first of these enzymes to be purified was protein palmitoyl thioesterase (PPT)-1 (Camp and Hofmann, 1993). Deficiency of PPT-1 leads to a lysosomal storage disease called infantile neuronal ceroid lipofuscinosis (Vesa *et al.*, 1995). PPT2 was cloned and characterized in 1997 (Soyombo and Hofmann, 1997), indicating that there may be an entire family of enzymes. Another of the proteins in this family, APT1, was shown to depalmitoylate several proteins, including $G_{\alpha s}$ (Duncan and Gilman, 1998), eNOS (Yeh *et al.*, 1999) and palmitoylated viral proteins (Veit and Schmidt, 2001). These enzymes are selective in the substrates that they will depalmitoylate as evidenced by the fact that APT1 will not depalmitoylate caveolin (Yeh *et al.*, 1999) or rhodopsin (Qanbar and Bouvier, 2003). Evidence that depalmitoylation is dynamically regulated was provided by the observation that eNOS depalmitoylation by APT1 was controlled by Ca^{+2}/calmodulin activity (Yeh *et al.*, 1999).

1.7. Rationale

The critical process of lipid modification of proteins has been well studied in many species. Four major forms of lipid modifications have been identified in a wide variety of eukaryotic and viral systems: myristoylation, prenylation, addition of glycophosphatidyl inositol and palmitoylation. Of these, palmitoylation appears to be the most versatile and the most common lipid modification. Although the importance of palmitoylation has been thoroughly investigated, the mechanisms by which this modification takes place, including the enzyme(s) responsible for the reaction, are poorly understood. Recently, two distinct palmitoyltransferases in yeast were characterized using a palmitoylation assay and by isolation of an intermediate bound to palmitate. These discoveries have opened up the possibility that homologs to these proteins can be found in other organisms, such as *Dictyostelium discoideum.*

To date, no palmitoyltransferases have been identified in *Dictyostelium discoideum. Dictyostelium* is a useful model system in which to study putative palmitoyltransferases because the genome of this haploid organism has been sequenced and is available on the Dictyostelium Genome Sequencing Project website (http://www.dictybase.org/). The life cycle is also clearly defined and relatively easy to manipulate. In addition, because *Dictyostelium* has only one β subunit gene (Lilly *et al.*, 1993), studies of lipid modifications of G proteins, especially of the α subunit and resulting effects of the removal of these modifications, can easily be performed.

With this information in mind, it will be useful to identify and study homologs of the yeast palmitoyltransferases in a variety of species, because that will allow for a greater understanding of signaling pathways in eukaryotic systems. Since *Dictyostelium*

is relatively easy to culture in the laboratory and there is a great deal of information known about the organism, it will be advantageous to identify palmitoyltransferases in this organism. In addition, because there are multiple substrate proteins in *Dictyostelium* that require palmitoylation, it will be useful to determine whether specific palmitoyltransferases are expressed at different times during the life cycle. This information will be helpful in many areas of future study, allowing researchers to examine regulation of signaling pathways with greater detail.

2. MATERIALS AND METHODS

2.1 Cell Culture

Ax-2, an axenic, wild type *Dictyostelium discoideum* cell line was used in this research. Wild type cells were grown axenically in HL-5 medium in shaking culture or on plastic culture dishes at 22°C (Watts and Ashworth, 1970).

2.2 Starvation of Dictyostelium Discoideum Cells

Because prolonged starvation stimulates activation of a number of signaling pathways leading to several phenotypic changes including aggregation, it was necessary to starve cells for use in further experiments. To starve *Dictyostelium*, vegetative state cells grown in suspension were centrifuged at 700 x g for 4 minutes at room temperature. The supernatant was decanted and the cell pellet resuspended in 15 ml of Development Buffer (DB, 10 mM 1X phosphate buffer, 2 mM $MgSO_4$, 0.2 mM $CaCl_2$). The cells were again centrifuged at 700 x g for 4 minutes and the supernatant decanted. The pellet was resuspended to 5 x 10^6 cells ml^{-1} in DB (60 ml) and 10 ml of the cell suspension was pipetted onto each of 6 culture dishes with agar medium. The cells were allowed to settle for 10-15 minutes and the excess DB was decanted. Each cell culture was incubated at 22°C and the cells were washed off of one plate every four hours for use in RNA extraction (protocol follows). Time points were taken at 4, 8, 12, 16, 20 and 24 hours.

2.3 RNA Extraction from Starved Dictyostelium Discoideum

To determine whether a specific ALZ gene was expressed at different time points during the *Dictyostelium* life cycle, RNA was extracted from each of the time point

samples acquired by starvation, as outlined above. A total of 10^7 cells were used for each RNA extraction. After the cells were starved for the appropriate amount of time (including cells that were not starved and remained in the vegetative state), they were washed from the plate and 10^7 cells were centrifuged at 700 x g for 4 minutes, at room temperature. The supernatant was decanted from the pellets and the cells were resuspended in 15 ml of DB. The suspensions were centrifuged at 700 x g for 4 minutes at room temperature and the supernatant was again removed. The pellets were resuspended in 175 µl of buffer RLN (50mM Tris-Cl, pH 8.0, 140 mM NaCl, 1.5 mM $MgCl_2$, 0.5 % (v/v) Nonidet P-40 (1.06 g/ml)) and the solutions incubated on ice for 5 minutes. They were then centrifuged at 700 x g for 2 minutes at 4°C. The supernatant was transferred to a 1.5 ml microcentrifuge tube and treated with DNase by adding 20 µl of 10 X DNase I buffer and 5 µl DNase I (Takara Biotechnology Inc., Japan) and incubated at 37°C for 20-30 minutes. After incubation, 600 µl of buffer RLT (Qiagen, Valencia, CA) was mixed into the samples, followed by 430 µl of 100% EtOH.

After mixing thoroughly, 700 µl was transferred to an RNeasy mini column placed into a 2 ml collection tube (RNeasy total RNA isolation kit, Qiagen, Valencia, CA) and the samples were centrifuged for 15 seconds at ≥ 8,000 x g. The flow-through was discarded and the remaining sample was subjected to an identical treatment. The RNeasy sample columns were saved, 700 µl of buffer RW1 (Qiagen, Valencia, CA) was added and the samples were centrifuged for 15 seconds at ≥ 8,000 x g. The flow-through was discarded and the RNeasy columns were transferred to a new 2 ml collection tube, after which, 500 µl of buffer RPE (Qiagen, Valencia, CA) was added to the columns. The samples were centrifuged for 15 seconds at ≥ 8,000 x g and the flow-through discarded.

Another 500 µl of buffer RPE was added to the column and the samples were centrifuged at ≥ 8,000 x g for 2 minutes to dry the columns. The RNeasy columns were transferred to new 1.5 ml microcentrifuge tubes and the RNA was eluted from the column by adding 30-50 µl of RNase free water to the column and centrifuging the samples for 1 minute at ≥ 8,000 x g. The RNeasy columns were discarded and the eluted RNA samples were stored at -20°C until they were quantified and used in further experiments. Quantification of RNA samples was completed on a spectrophotometer at 260 nm.

2.4 RT-PCR Analysis

To determine whether the identified ALZ genes are expressed in *D. discoideum*, reverse transcription followed by polymerase chain reaction (RT-PCR), was performed on the vegetative RNA sample isolated by the protocol outlined above. The reagents used for the reactions were from Promega (Madison, WI). In a sterile microcentrifuge tube the following reagents were added:

AMV/T*fl* 1X Reaction Buffer
dNTP Mix (0.2 mM each dNTP)
Gene specific primer (1 µM) (Table 2)
12.5mM $MgSO_4$
AMV Reverse Transcriptase (0.1U/µl)
6 ng/µl RNA
to 50 µl H_2O

After all of the above reagents were added, 20 µl of mineral oil was added to the top of the sample to prevent evaporation during PCR. Reverse transcription was performed at 48°C for 45 minutes. PCR was performed with a 30 second denaturation

step at 94°C, 1 minute of annealing at 55°C and 2 minutes of extension at 68°C. After 25 or 40 cycles of PCR were completed, the samples were stored at -20°C until analyzed.

2.5 Gene Expression by Real Time PCR

To further characterize the putative palmitoyltransferases identified in the BLAST search, real time PCR was performed using *Dictyostelium* cDNA samples from the vegetative stage and all six starvation time-points to qualitatively and semi-quantitatively compare when each of the genes is expressed during the *Dictyostelium* life cycle.

2.5.1 Random Primer Reverse Transcription

Before reverse transcription of the RNA samples could take place, 2.8 µl of random primer (Invitrogen, Carlsbad, CA) was added to 2 µg of RNA isolated from vegetative stage cells and from each of the six time point RNA samples from the above RNA isolation protocol. RNase free water (up to 11 µl total volume) was added to each of the samples, except the 8-hour starved sample, which already had 14.6 µl of total volume. All of the samples were incubated at 70°C for 5 minutes and then incubated on ice for 5 minutes.

After incubation, the following reagents were added to each sample:

AMV Reverse Transcriptase 1X buffer
dNTP (1 mM of each dNTP)
AMV Reverse Transcriptase (0.66U/µl)
to 25 µl RNase free water (except the 8 hour sample, which already had 25.4 µl total volume)

The samples were incubated for 1 hour at 37°C, after which they were diluted 1:1 with RNase free water and stored at -20°C.

Table 2. Internal Primers for RT-PCR. All twelve primer sets surround the conserved sequence that translates to the DHHC amino acid sequence. Sense and antisense primers are shown for each of the identified genes.

Gene	Sense Strand	Antisense Strand
ALZ 1	TGGTGTGGTATACCACCAAGAG	CCTATAACGTGCTGATTCTATAC
ALZ 2	AGATCAGAGACATATAGATTTTG	CAAACATAAACACATGATATCC
ALZ 3	ATTACATGTGGATTATATAGAG	GACATTGATAAAAGATATAGTGC
ALZ 4	CCAGAACGAGCTCATCATTG	AACTTGGTAATGTTTTTTCCATTG
ALZ 5	CAAATCCCTTTGATTTGTCC	AACCAGAAATTACCAAGATAA
ALZ 6	GTGAAAAAGAAGATAATGATATTG	AACAAAACCAGTGACAATGATTC
ALZ 7	GCCGTCGAAAAGGAATTGG	GAATATAATGGGGAATTCTCATC
ALZ 8	TGTGCTTTTAAACCACCAAGAAC	CCTGATATAGTTAAAGTTGTG
ALZ 9	CATCTCATTGTTCAACATGTAATAG	AGAATAACTAGTGGTTACTATTG
ALZ 10	GGAGTTATTGTAAGAAATGTTC	ACCACCAAATAATAAAGGTAAAG
ALZ 11	TCACCTGATGAAGATTTAAAAGA	AATTGGAATTGCTATTGCACTATAC
ALZ 12	CTTATCCATATGATAGATATC	AAATCCACATAAATAAGCACC

2.5.2 Real Time PCR

Real time PCR reactions were performed for each time point and for each ALZ gene using the iCycler iQ Detection System (Bio-Rad Laboratories, Hercules, CA). Sense and antisense primers are listed in Table 2. To each reaction was added the following:

sense primer (0.2μM)
antisense primer (0.2μM)
IQ Supermix with SYBR Green (1X)
2 μl diluted cDNA template (3 ng/μl)
to 25 μl RNase free water

All reagents were mixed by pipetting and 25 μl of each sample was aliquoted into separate wells in a 96 well plate. Each time point sample was run in triplicate. The PCR reaction was run with the following parameters:

1) Initial denaturation step – 94°C, 3 minutes
2) Cycle denaturation step – 94°C, 30 seconds
3) Annealing step – 55°C, 30 seconds
4) Elongation step – 72°C, 30 seconds
5) Repeat steps 2-4 for a total of 35 or 40 cycles
6) Final elongation step – 72°C, 10 minutes

2.6 DNA Sequence Analysis

Sequencing of the newly identified palmitoyltransferase gene ALZ 6 was carried out by the University of Maine DNA Sequencing Facility (Patricia Singer, Director) on an ABI 373 Stretch DNA Sequencer with XL Upgrade, ABI Sequence Navigator DNA and Protein Sequence Comparison software was used for analysis. Reagents included: ABI Prism BigDye Terminator V 3.0 Cycle Sequencing Ready Reaction Kit).

The ALZ 6 cDNA sequence was isolated by RT-PCR using primer sets designed from the ALZ 6 gene sequence generated by the Blast search described previously:

ALZ 6 Forward Full Length Primer

5'-AAAGAAAATTATGAAATTAAATA-3'

ALZ 6 Reverse Full Length Primer

5'-TTAACCACAACAAGAATAATATTTATTATTCC-3'

The RT-PCR product was inserted into the pCR II-TOPO plasmid (Invitrogen, Carlsbad, CA), following the product specific protocol.

All other cDNAs used for sequencing were isolated and amplified by rapid amplification of 5' and 3' ends using a GeneRacer™ Kit (Invitrogen, Carlsbad, CA). Protocols used in amplification of sequences are outlined later in this section. Internal primer sequences specific to each gene were used in sequencing (Table 2).

2.7 Rapid Amplification of cDNA ends (RACE) for Sequencing DNA

To sequence several of the ALZ genes, 5' and 3' cDNA ends were amplified according to the GeneRacer™ Kit (Invitrogen, Carlsbad, CA) protocol.

2.7.1 Rapid Amplification of 5' cDNA Ends

Rapid amplification of cDNA Ends (RACE) was performed on an RNA sample from *Dictyostelium* cells that had been starved for eight hours and on a control HeLa RNA sample from the manufacturer kit (Invitrogen, Carlsbad, CA). To begin the RACE procedure, the sample total RNA had to be dephosphorylated. This eliminates the possibility of amplifying truncated mRNA in further steps. The same protocol was

followed for both samples. The following reagents were mixed in two sterile 1.5 ml microcentrifuge tubes:

0.1 µg/µl 8 hour starved total RNA or control RNA
1X Calf Intestinal Phosphatase (CIP) buffer
RNaseOut™ (4 U/µl)
CIP (1 U/ul)
to 10 ul with DEPC treated water

All reagents were mixed by pipetting, vortexed and centrifuged, briefly, at full speed. The samples were incubated at 50°C for at least one hour and then centrifuged briefly at top speed and put on ice. All subsequent centrifugations were at full speed unless indicated otherwise. DEPC treated water (90 µl) and phenol: chloroform were added to each sample, which were then vortexed vigorously for 30 seconds and centrifuged for 5 minutes at room temperature. The top (aqueous) phase was transferred to a new microcentrifuge tube and 2 µl of 10mg ml^{-1} mussel glycogen and 10 µl of 3M sodium acetate (pH 5.2) were added. The samples were mixed and 220 µl of 95% EtOH was added. After vortexing the samples briefly they were incubated at -20°C for 10 minutes. After incubation, the samples were centrifuged at room temperature for 20 minutes. The supernatant was removed and the pellet was washed in 500 µl of 70% EtOH. The samples were centrifuged for 2 minutes and the supernatant (EtOH) was removed. The remaining EtOH was collected by briefly centrifuging again and pipetting off the supernatant. Pellets were air dried for 2 minutes and resuspended in 7 µl DEPC treated water. The RNA samples were then taken through the protocol to remove the mRNA cap structures. This ensures a 5' phosphate available for future ligation reactions. The following reagents were added on ice to the 7 µl dephosphorylated RNA samples:

tobacco acid pyrophosphatase (TAP) buffer
RNaseOut™ (4 U/μl)
TAP (0.05 U/μl)

All of the reagents were mixed and vortexed, briefly after which the samples were centrifuged briefly, and incubated at 37°C for 1 hour. After incubation, the samples were briefly centrifuged and the RNA precipitated by adding 90 μl DEPC treated water and 100 μl phenol: chloroform followed by vortexing for 30 seconds. After centrifugation for 5 minutes at room temperature, the aqueous phase from each sample was transferred to a new microcentrifuge tube. To this was added 2 μl 10 mg ml^{-1} mussel glycogen and 10 μl 3 M sodium acetate, pH 5.2. The solution was mixed and 220 μl 100% EtOH was added, followed by vortexing. The samples were then stored overnight at -20°C. The following morning the RNA was pelleted by centrifugation for 20 minutes at room temperature. The supernatant was removed and the pellets washed by adding 500 μl 70% EtOH and inverting the samples several times and vortexing, briefly. The samples were centrifuged at top speed for 2 minutes and the supernatant removed. The remaining EtOH was removed by centrifuging the pellets at top speed again, briefly, and pipetting the remaining EtOH from the pellets. The pellets were air dried for 2 minutes at room temperature and resuspended in 7 μl DEPC water.

With the mRNA samples de-capped, they were ligated to the 5' oligo that could be recognized by the kit 5' primer in future amplification steps. To do this, the 7 μl de-capped mRNA samples were added to separate tubes containing pre-aliquoted, lyophilized GeneRacer™ RNA Oligo (0.25μg). The oligo was resuspended and the samples were centrifuged briefly to collect the solution in the bottom of the tube. The samples were incubated at 68°C for 5 minutes to relax the RNA secondary structure.

Following incubation, the samples were chilled on ice for 2 minutes and centrifuged briefly. To each sample was added the following:

1X Ligase buffer
1mM ATP
RNaseOut™ (4 U/µl)
T4 RNA ligase (0.5 U/µl)

All reagents were mixed and the samples incubated at 37°C for 1 hour. After incubation, the samples were centrifuged briefly, and placed on ice. RNA in each sample was precipitated as described earlier and then resuspended in 10 µl DEPC treated water.

2.7.2 Rapid Amplification of 3' cDNA Ends

After the 8-hour *Dictyostelium* RNA and the control HeLa RNA samples were dephosphorylated, de-capped and ligated to the 5' oligo provided in the kit, the samples were prepared for RT-PCR for future sequencing.

To 10 µl of ligated RNA, 1 µl of the GeneRacer™ Oligo dT Primer and 1 µl of dNTP (10mM each) mix were added. The samples were incubated at 65°C for 5 minutes, to remove any secondary structure, and then chilled on ice for 2 minutes. After centrifuging the samples briefly the following reagents were added to the 12 µl ligated RNA and primer mixture:

1X First Strand Buffer
Cloned AMV Reverse Transcriptase (0.75 U/µl)
RNaseOut™ (2 U/µl)
to 20 µl sterile water

All of the reagents were mixed with the RNA samples and the samples were then incubated at 45°C for 1 hour. Following incubation, the reverse transcriptase was

inactivated by incubating at 80°C for approximately 15 minutes. The samples were centrifuged briefly and stored overnight at -20°C.

The following day, the newly synthesized second strand cDNA samples were amplified by PCR using kit primers for the ends of the genes and gene specific primers (GSP, Table 2) for the internal regions of the genes. Both 5' and 3' RACE procedures were performed on the cDNA samples. PCR reactions were set up for both the control HeLa cDNA and the 8-hour starvation cDNA samples as shown below:

5' RACE PCR

3 µl GeneRacer™ 5' Primer (10 µM)
0.5 µl reverse GSP (20 µM)
1 µl unknown cDNA template
10 µl 5X PCR buffer
1 µl dNTP (10mM each)
0.5 µl *Taq* Polymerase (5 U/µl)
4 µl $MgSo_4$ (25mM)
to 50 µl with sterile water

5' RACE Control PCR

3 µl GeneRacer™ 5' Primer (10 µM)
1 µl Control Primer (B.1, 10 µM)
1 µl HeLa cDNA template
10 µl 5X PCR buffer
1 µl dNTP (10mM each)
0.5 µl *Taq* Polymerase (5 U/µl)
4 µl $MgSo_4$ (25mM)
to 50 µl with sterile water

3' RACE PCR

3 µl GeneRacer™ 3' Primer (10 µM)
0.5 µl forward GSP (20 µM)
1 µl unknown cDNA template
10 µl 5X PCR buffer
1 µl dNTP (10mM each)
0.5 µl *Taq* Polymerase (5 U/µl)
4 µl $MgSo_4$ (25mM)
to 50 µl with sterile water

3' RACE Control PCR

3 µl GeneRacer™ 3' Primer (10 µM)
1 µl Control Primer (A, 10 µM)
1 µl HeLa cDNA template
10 µl 5X PCR buffer
1 µl dNTP (10mM each)
0.5 µl *Taq* Polymerase (5 U/µl)
4 µl $MgSo_4$ (25mM)
to 50 µl with sterile water

After all of the reagents were added as outlined above, the samples were overlaid with 20 µl of mineral oil. The PCR reactions were run according to the protocol given in the GeneRacer™ manual.

3. RESULTS

3.1 BLAST Search

Using the nucleotide sequence for the yeast palmitoyltransferase ERF2 (YLR246w, Bartels *et al.,* 1999) as the template for a BLAST search, twelve putative homologs were identified in the *D. discoideum* genome (Fig. 6). The BLAST search was performed using an Internet program designed to compare genes of interest to the entire *Dictyostelium* genome, on a *Dictyostelium* Internet portal called Dictybase (Fey *et al.,* website: http://dictybase.org/). The BLAST search was performed with an expected cutoff of 1 e-5, using the tblastx (translated DNA vs. translated DNA) program. The search utilized all Predicted Nucleotide open reading frames (orfs) in the database (updated 2/2003). All twelve open reading frames showed a high degree of homology within the previously identified DHHC-CRD amino acid sequence, although sequence similarity outside of this domain was minimal. The newly identified genes have been termed Acyltransferase-like gene with a Zinc Finger Domain (ALZ 1-12). Primers were designed for the full-length sequence and for the DHHC-CRD internal region for all twelve orfs and were used in further experiments.

The DictyBase website allows users to search for *Dictyostelium* orfs based on a terminology database search through the entire website (search website: http://dicty.sdsc.edu/searchannot2003.html). As further confirmation that all of the known *Dictyostelium* DHHC domain orfs were located in the BLAST search, a database search was performed on the DictyBase website using the term DHHC. The same twelve orfs were returned in this search as in the BLAST search (Table 3).

TBLASTX 2.0.9 [May-07-1999]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= (1080 letters)

Database: dictydb/Dicty013103_pred_cdna.fasta
12,734 sequences; 18,443,583 total letters

nt-ng2564	Contig_0999 pred cdna pORF 3540 5369...	97	1e-20
nt-ng207	Contig_0080 pred cdna pORF 177 1832 ...	87	3e-20
nt-ng395	Contig_0162 pred cdna pORF 529 1941 ...	88	6e-18
nt-ng7349	Contig_3025 pred cdna pORF 77 2558...	84	1e-16
nt-ng4163	Contig_1649 pred cdna pORF 3628 2415...	81	7e-16
nt-ng7517	Contig_3092 pred cdna pORF 4123 2403...	81	7e-16
nt-ng6574	Contig_2694 pred cdna pORF 4773 7027...	74	1e-14
nt-ng10639	Contig_4459 pred cdna pORF 695 216...	77	1e-14
nt-ng10268	Contig_4303 pred cdna pORF 304 257...	75	4e-14
nt-ng8770	Contig_3633 pred cdna pORF 5499 6845...	74	1e-13
nt-ng1677	Contig_0632 pred cdna pORF 893 2724...	45	6e-11
nt-ng2062	Contig_0789 pred cdna pORF 4827 5913...	56	4e-08

Figure 6. BLAST Search Results. Twelve genes in *Dictyostelium discoideum* were identified as having significant homology to ERF2 in *Saccharomyces cerevisiae*. Results taken from the DictyBase website.

Table 3. ALZ Open Reading Frames. Open reading frames identified in *Dictyostelium discoideum* by searching the DictyBase website for the identifier "DHHC". The author of this thesis arbitrarily assigned Gene IDs.

Gene ID	Orf ID	aa Count	Homology	Accession	Pfam Descript.	Contig	Start nt	End nt
ALZ 1	ng207	458	Hypothetical protein	gb|AAL92226.1|	zf-DHHC DH	Contig 0080	177	1832
ALZ 2	ng2062	330	Unknown protein	gb|AAF14029.1|	zf-DHHC DH	Contig 0789	4827	5913
ALZ 3	ng2564	502	Unknown protein	gb|AAL85057.1|	zf-DHHC DH	Contig 0999	3540	5369
ALZ 4	ng10639	446	AT3G09320/F3L24 19.9/101	gb|AAM33196.1|	zf-DHHC DH	Contig 4459	695	2160
ALZ 5	ng6574	730	Putative ankyrin repeat protein	gb|AAL86959.1|	ank Ankyri	Contig 2694	4773	7027
ALZ 6	ng8770	448	Hypothetical protein	gb|AAL92625.1|	zf-DHHC DH	Contig 3633	5499	6845
ALZ 7	ng10268	698	Ankyrin repeat-containing protein AKR1	gb|AAL92258.1|	ank Ankyri	Contig 4303	304	2577
ALZ 8	ng395	442	"cell cycle regulator with zn-finger domain, putative"	emb|CAD51645.1|	zf-DHHC DH	Contig 0162	529	1941
ALZ 9	ng1677	449	Hypothetical protein	emb|CAD50894.1|	zf-DHHC DH	Contig 0632	893	2724
ALZ 10	ng4163	348	ABL-PHILIN 2	gb|AAM45351.1|	zf-DHHC DH	Contig 1649	3628	2415
ALZ 11	ng7349	706	Ankyrin repeat-containing protein AKR1	gb|AAL92258.1|	"ank Ankyr	Contig 3025	77	2558
ALZ 12	ng7517	433	"unknown protein, contains TNFR/NGFR cysteine-rich"	gb|AAG51421.1|	zf-DHHC DH	Contig 3092	4123	2403

3.2 Predicted Temporal Characteristics of ALZ 1-12

3.2.1 Prediction of Intron Location in ALZ 1-12

Although the *Dictyostelium* genome has been sequenced and many orfs have been identified within the genome, it is possible for mistakes to be made in predicting introns and exons within identified open reading frames. Errors can be made both in sequencing and in interpretation of results from programs designed to predict intron location. Because of these possibilities, it is important to sequence orfs of interest so that the correct sequence can be confirmed and predictions of introns can be verified.

In addition to gene sequencing, predictions were made as to the location of introns in the available gene sequences by comparison of the known predicted coding sequences to the contigs that contain the gene of interest. Each of the ALZ genes identified in the BLAST search is already associated with a contig that is available at the DictyBase website (http://dicty.sdsc.edu/searchannot2003.html). By comparing each of the ALZ orfs identified from the BLAST search with the corresponding contig, a prediction was made as to where introns are located within each of the ALZ genomic DNA sequences. Comparisons were done visually, as well as with the aid of CLUSTALWPROF, found on the Biology Workbench website of the San Diego Supercomputer Center (http://workbench.sdsc.edu) at the University of California at San Diego. All of the gene sequences with the predicted intron locations and sequences are found in the addendum at the end of this thesis.

Two of the identified orfs, ALZ 2 and ALZ 9, required substantial editing to correct what appeared to be misidentified introns and exons within both of these open

reading frames. Corrections were made to the ALZ 2 sequence based on visual analysis of the sequence and, as a result, the orf became more similar to the Erf2p sequences, in that ALZ 2 now shows 4 predicted transmembrane domains (Table 4).

A previously unidentified intron was also found in the middle of the DHHC conserved domain of the ALZ 9 gene sequence. Removal of this intron did not change the number or location of the predicted transmembrane domains, however, it did cause the DHHC conserved sequence to re-emerge when the sequence was translated.

In addition to those major changes in ALZ 2 and ALZ 9, it was found that predicted introns in the ALZ 3 and ALZ 4 sequences were shifted downstream; in their respective contigs by a single base. This shift does not destroy the reading frame that contains the conserved DHHC domain in either gene, however, the introns located in the predicted open reading frames do not start or finish with the expected GT or AG sequences, respectively.

The ALZ 6 gene sequence has been confirmed by sequencing and will be discussed later. The cDNA for each of the remaining ALZ genes will be sequenced in the future to confirm whether intron predictions are correct.

3.2.2 Prediction of Transmembrane Domains in ALZ 1-12

As described by Bartels *et al.,* (1999), Erf2 is a 41 kDa protein with four predicted transmembrane domains (TMD). If the ALZ genes identified here in *Dictyostelium* are indeed true homologs to ERF2, or possibly to AKR1, then it should be expected that they will have the same number of transmembrane domains and the location of the DHHC-CRD should correspond to the yeast genes, as well.

Each of the ALZ amino acid sequences found at the DictyBase website was analyzed for predicted TMDs using the Internet program Simple Modular Architecture Research Tool (SMART; website: http://smart.embl-heidelberg.de/smart/show_motifs.pl, data not shown). For verification of the predicted topology, several of the ALZ amino acid sequences were also analyzed for predicted TMDs using the "DAS" Transmembrane Prediction server (website: http://www.sbc.su.se/~miklos/DAS/tmdas.cgi; data not shown). Both programs function by searching for hydrophobic regions within the amino acid sequence. The results of these analyses are found in Table 4, with a graphic representation of the predicted topology shown in Fig. 7.

This analysis shows that many of the ALZ proteins show similarity to Erf2p, in that they have four predicted transmembrane domains and the DHHC domain is expected to be inside the cell. These predictions are based on the assumption that the N-terminus of each of the proteins is located within the cell. A signal sequence at the N-terminal region of the genes may suggest that this region is outside the cell, although neither program used to predict transmembrane domains predicted the presence of a signal sequence in any of the sequences.

There are four notable exceptions to the previous observations. ALZ 4 has 5 predicted TMDs and ALZ 7, 11 and 12 all have 6 predicted TMDs (Table 4). ALZ 4 is unusual in that the 5 predicted TMDs do not fit any known patterns for DHHC proteins. ALZ 7, 11 and 12 are more characteristic of Akr1p, which has 6 predicted TMDs. Akr1p also has a series of characteristic ankyrin repeats near the N-terminus of the protein, as does ALZ 7 and ALZ 11. Curiously, ALZ 12 does not appear to have these ankyrin repeats. Just as interestingly, though, is the fact that ALZ 5 does show predicted ankyrin

Table 4. Predicted Transmembrane Domains of ALZ 1-12 Proteins. Each of the twelve ALZ genes was analyzed using the Simple Modular Architecture Research Tool (SMART) program and all but ALZ 4 has been found to a have characteristic number of transmembrane domains.

Protein Name	# of Transmembrane Domains	Location of DHHC Domain
ALZ 1	Four	Intracellular
ALZ 2	Four	Intracellular
ALZ 3	Four	Intracellular
ALZ 4	Five	Cytoplasmic‡
ALZ 5	Four	Intracellular
ALZ 6	Four	Intracellular
ALZ 7	Six	Intracellular
ALZ 8	Four	Intracellular
ALZ 9	Four	Intracellular
ALZ 10	Four	Intracellular
ALZ 11	Six	Intracellular
ALZ 12	Six	Intracellular

‡ Transmembrane Domains predicted with SMART indicate that the DHHC domain is outside the cell membrane, however, prediction made using DAS suggest only 4 TMD and that the DHHC domain is inside the cell membrane.

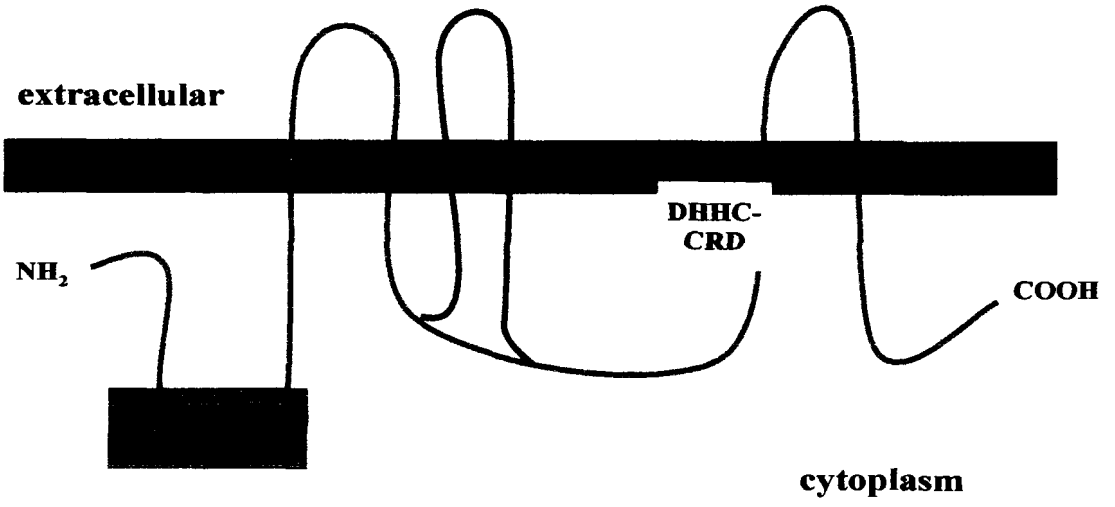


Figure 7. Predicted Transmembrane Topology of DHHC-CRD Proteins. Transmembrane domain analysis was performed, using the Simple Modular Architecture Research Tool (SMART) program. In addition to ALZ 5, 7, and 11, ALZ 12 may also have six transmembrane domains.

repeats near the N-terminus; however, it is only predicted to have 4 TMDs. This makes this protein similar to both Erf2p and to Akr1p. All of the other ALZ proteins show 4 predicted TMDs and have no obvious ankyrin repeats, making these proteins much more similar to Erf2p.

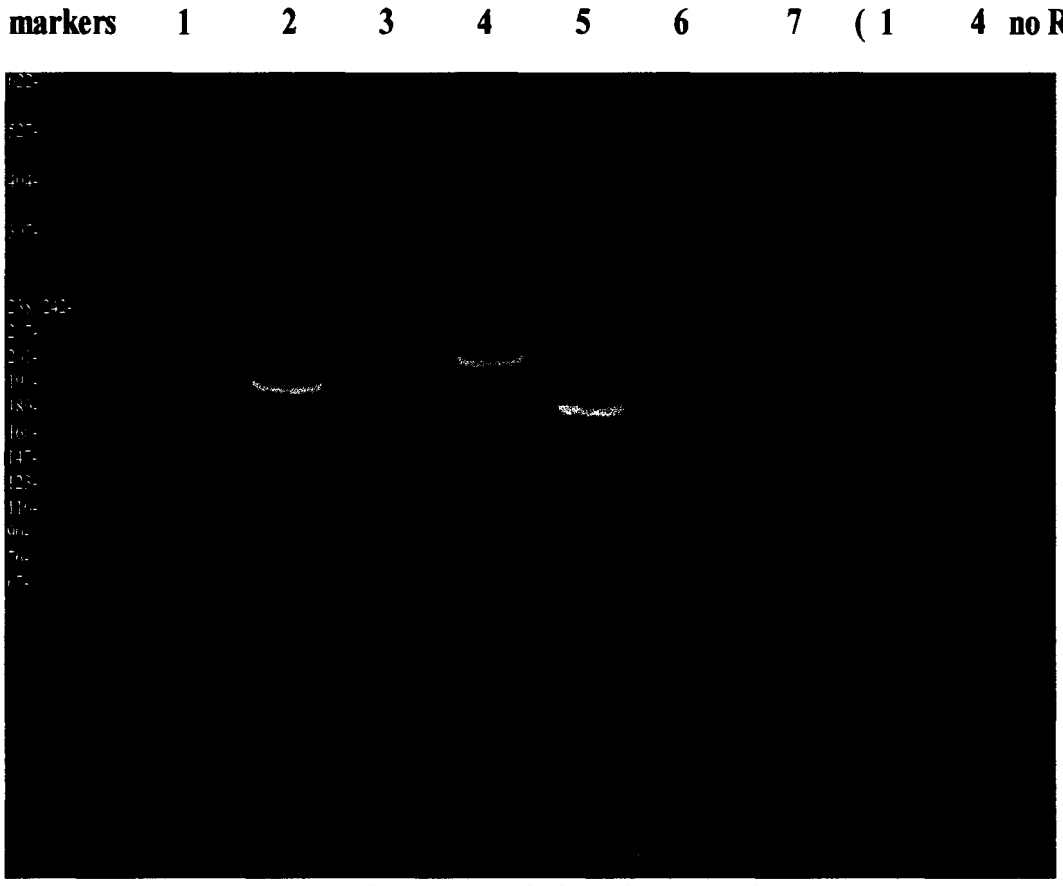
3.3 RT-PCR Analysis of Gene Expression

The twelve ALZ genes were examined for general expression in *Dictyostelium* using RT-PCR. RT-PCR for two of the ALZ genes, ALZ 1 and ALZ 4, was repeated without the reverse transcriptase, to determine whether any contaminating genomic DNA was present in the original sample. The products of the RT-PCR reactions were separated on acrylamide gels (Figs. 8 and 9).

General expression of all twelve genes was confirmed by the presence of bands in each sample. Based on the placement of each of the gene specific primer sets used to amplify the product in the RT-PCR experiments, a specific size can be predicted for each of the gene products (Table 5). Fig. 8 shows that a comparison of each of the ALZ 1-7 bands on the gel with the bands in the DNA marker lane indicates that each unknown band is approximately the right size. It is also apparent that there is no contaminating genomic DNA in the original samples, as no bands are seen in the lanes that correspond to a lack of reverse transcriptase in the original reactions (Fig. 8). Each of the bands shown for ALZ 8-12 is also the expected size (Fig. 9). Both of these experiments suggest that the twelve ALZ genes are expressed during the vegetative state of the *Dictyostelium* life cycle, although these results give no indication of the expression patterns characteristic of these genes. Unclear, though, is the specific identity of the multiple

Table 5. Size of ALZ 1-12 Products Using Internal Primers. Predicted size of PCR products when produced by the primers designed around the conserved DHHC domain, as described.

ALZ 1	ng207	210 nt
ALZ 2	ng2062	200 nt
ALZ 3	ng2564	180 nt
ALZ 4	ng10639	220 nt
ALZ 5	ng6574	190 nt
ALZ 6	ng8770	250 nt
ALZ 7	ng10268	260 nt
ALZ 8	ng395	185 nt
ALZ 9	ng1677	152 nt
ALZ 10	ng4163	215 nt
ALZ 11	ng7349	240 nt
ALZ 12	ng7517	230 nt



(15% acrylamide gel, stained with ethidium bromide)

Figure 8. RT-PCR of ALZ 1-7 Genes. RT-PCR was performed on vegetative state RNA, using primers designed against ALZ 1-7. The annealing temperature was set at 55°C. Each of the genes was amplified as described and the samples were run on a 15% acrylamide gel. Each of the genes shows a band of the expected size, when compared to the pBR322A DNA marker in the first lane.

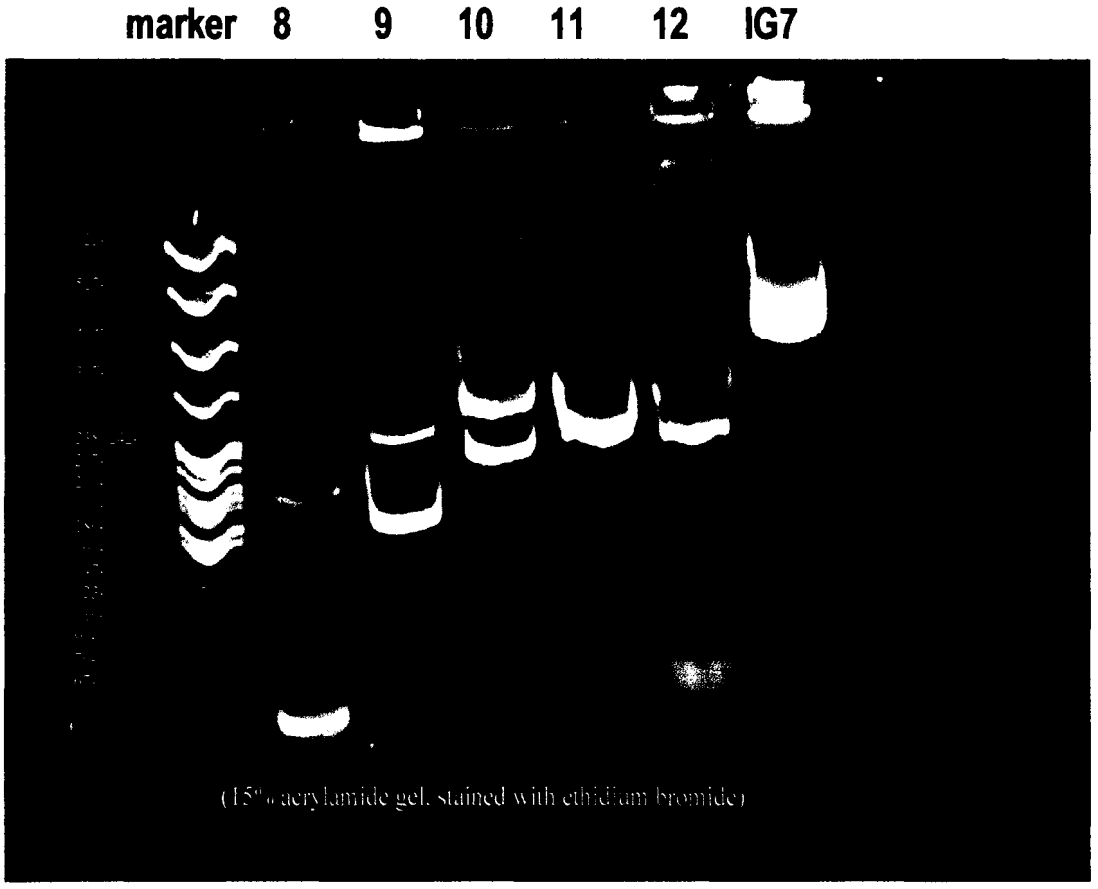


Figure 9. RT-PCR of ALZ 8-12 Genes and IG7. An RT-PCR experiment was done, using primers designed against ALZ 8-12 and IG7 (a constitutively expressed housekeeping gene). The annealing temperature was set at 55°C. Each of the samples produced an amplified band of the expected size when compared to the pBR322A marker..

bands in lanes corresponding to ALZ 2, 4, 8, 9 and 10, although, it is likely that these secondary bands in each lane are the result of mis-priming in the RT-PCR experiments.

3.4 Rapid Amplification of cDNA Ends (RACE) for DNA Sequencing

All attempts to amplify entire ALZ genes by the RACE method have thus far failed to produce complete sequences. In several attempts, measurable quantities of DNA were isolated from these reactions, but attempts to sequence the DNA have consistently led to unreadable results.

3.5 DNA Sequencing of ALZ Genes

The ALZ 6 gene was isolated from *D. discoideum* as described and subsequently sequenced. Two different clones were picked for sequencing and an alignment was created between the two cloned sequences and the published sequence using Clustalw. When compared with the ALZ 6 sequence found on the Dictybase website, the ALZ 6 plasmid sequence showed a minimum of 13 nucleotide differences depending upon which of the cloned sequences were compared to the published sequence. The alignment of one of the clones with the published sequence is shown in Figure 10. Nearly all of the observed nucleotide differences do not result in an overall frame change in the sequence, although, of particular interest is the deletion of an adenine from the cloned sequence which does create a premature stop codon in-frame. The differences in the sequences do result in a number of amino acid changes between the two sequences, but at this time it is impossible to discern which of the sequences is correct without picking new clones and re-sequencing.

```
ALZ 6 ----------------------ATGAAATTAAATAGTAATATTAATAATAATATAAATAA 38
6-6-2 GAATTCGCCCTTAAAGAAAATTATGAAATTAAATAGTAATATTAATAATAATATAAATAA 60
      **********************
ALZ 6 TAGTAGTAATAGTAATAATAACTTTGATGCAAAGAATATAATAGTTGATACAATAACACC 98
6-6-2 TAGTAGTAATAGTAATAATAACTTTGATGCAAAGAATATAATAGTTGATACAATAACACC 120

ALZ 6 ACCTGACCCATCGGTTGAATTTGAAAGGAAATTAGCGAAATCTATATTTTGTTTAGTTCA 158
6-6-2 ACCTGACCCATCGGTTGAATTTGAAAGGAAATTAGCGAAATCTATATTTTGTTTAGTTCA 180

ALZ 6 TTTTATAGTTTATTGCGTTATTATATTTAGAAAAGGTACAATTTTAGATCAAGCATTCAA 218
6-6-2 TTTTATAGTTTATTGCGTTATTATATTTAGAAAAGGTACAATTTTAGATCAAGCATTCAA 240

ALZ 6 AGATAAGGATTATTTTTATTTAATTTGGACACATTGTGTTTTCTTTTTTGCAATTGGAAC 278
6-6-2 AGATAAGGATTATTTTTATTTAATTTGGACACATTGTGTTTTCTTTTTTGCAATTGGAAC 300

ALZ 6 TTATTTCTTAATTTCAAGTAAAAGACCTGGTTTTGTTTCATTATCAAATCAAAATTTAAA 338
6-6-2 TTATTTCTTAATTTCAAGTAAAAGACCTGGTTTTGTTTCATTATCAAATCAAAATTTAAA 360

ALZ 6 TAATAATAATAATAATAATGGTAGTAGCAATAAATTTATATTAGAAGATTCAATGGGTTG 398
6-6-2 TAATAATAATAATAATAATGGTAGTAGCAATAAATTTATATTAGAAGATTCAATGGGTTG 420

ALZ 6 TATCCCTCAATTAAATATTAATCCAACACCAAATTATAGTAAAATAAGTAATATTAAGAG 458
6-6-2 TATCCCTCAATTAAATATTAATCCAACACCAAATTATAGTAAAATAAGTAATATTAAGAG 480

ALZ 6 AAAATTAAAAAATAGTAGTGGTGATATTACTAAAAATCAAGAAAATGAAGATTTAGTGCC 518
6-6-2 AAAATTAAAAAATAGTAGTGGTGATATTACTAAAAATCAAGAAAATGAAGATTTAGTGCC 540

ALZ 6 ATTGATGGAAATCTCAAAGAATATTGATGAAGATAGCATAAATGACGATACTATAACTAC 578
6-6-2 ATTGATGGAAATCTCAAAGAATATTGATGAAGATAGCATAAATGACGATACTATAACTAC 600

ALZ 6 AACCACTACCACTACCACCACCACCAGTACCTCAACAATACCAGAAATATCAAATGATGA 638
6-6-2 AACCACTACCACTACCACCACCACCAGTACCTCAACAATACCAGAAATATCAAATGATGA 660

ALZ 6 TGATGATAATAATAATGAAAACAATAACGATAATGTAAATAATAGAAATAATAATAATAG 698
6-6-2 TGATGATAATAATAATGAAAACAATAACGATA-TGTAATTA-TAGAAATAATAATAATAG 718
                                      *     *  *
ALZ 6 TAATGGTG-AAAAAGAAGATAATGATATTGATAAATTAAAGAATCATTATTTTTGTAAGA 757
6-6-2 GTATGGTGGAAAAAGAAGATAATGATATTGATAAATTAAAGAATCATTATTTTTGTAAGA 778
      **      *
ALZ 6 AATGTTTAGTTGATATACCATTAAGAACTAAACATTGTGTAAAATGTAATAGATGTGTTT 817
6-6-2 AATGTTTAGTTGATATACCATTAAGAACTAAACATTGTGTAAAATGTAATAGATGTGTTT 838

ALZ 6 TAAAGTATGATCACCATTGCGTATTTATAGGTGGTTGTGTTGGTTTAAATAATCATAAGA 877
6-6-2 TAAAGTATGATCACCATTGCGTATTTATAGGTGGTTGTGTTGGTTTAAATAATCATAAGA 898
```

Figure 10. Alignment of Predicted ALZ 6 Sequence and Experimental Sequence. Gene sequence of ALZ 6, as given by the University of Maine DNA Sequencing Facility, aligned with the ALZ 6 sequence found on Dictybase. Alignment was performed using Clustalw. The ALZ 6 sequence found on Dictybase is labeled ALZ 6 and the experimentally derived sequence is labeled 6-6-2. Additional sequence before and after the 6-6-2 sequence corresponds to experimental primers and to vector sequence. Divergent nucleotides are indicated with an asterisk. The A nucleotide missing from the 6-6-2 sequence is in bold red in the ALZ 6 sequence.

```
ALZ 6 ATTTTCTATTATTTTTATTGGCGGAATCATTATTACTACTTTTAGGTTTAAGAATCATTG 937
6-6-2 ATTTTCTATTATTTTTATTGGCGGAATCATTATTACTACTTTTAGGTTTAAGAATCATTG 958

ALZ 6 TCACTGGTTTTGTTAGAGAAAATTCAATTAAAGAATGGATCTTTTCAAATATTGCAATTA 997
6-6-2 TCACTGGTTTTGTTAGAGAAAATTCAATTAAAGAATGGATCTTTTCAAATATTGCAATTA 1018

ALZ 6 TTCCACCAACTTTATTAATTTTTGGTGGTTTATGTATGCCTTTTGCTTTATTCTGTTTTC 1057
6-6-2 TTCCACCAACTTTATTAATTTTTGGTGGTTTATGTATGCCTTTTGCTTTATTCTGTTTTC 1078

ALZ 6 ATAGTTTTTTAATTTTAACAAACCAATCTTCATGGTATGTGCTTATGAATAGATTTTTTT 1117
6-6-2 ATAGTTTTTTAATTTTAACAAACCAATCTTCATGGTATGTGCTTATGAATAGATTTTTTT 1138

ALZ 6 TTAAAATTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCTAATTTAATTAAATTTA 1177
6-6-2 TTAAAATTGGAAAAAAAAAAAAAAAAAAAAAAAAAN---ANCTAATTTAATTAAATTTA 1195
              *                                **** *
ALZ 6 ATAATAGGGAATTTAATAAGTATCAAAGAATTACATATTTAAAACCATTTTCTAAAAGAG 1237
6-6-2 ATAATAGGGAATTTAATAAGTATCAAAGAATTACATATTTAAAACCATTTTCTAAAAGAG 1255

ALZ 6 GTATTAATCCATTTAATAAAGGTCCATGGAATAATTTAAAAAAGTTTTTAAA-GGTGATG 1296
6-6-2 GTATTAATCCATTTAATAAAGGTCCATGGAATAATTTAAAAAAGTTTTTAAAAGGTGATG 1315
                                                        *
ALZ 6 AGAATCCTTCCGATTGGATATTATTATCAAAATATGAAGTTGACCAAATGA--------- 1347
6-6-2 AGAATCCTTCCGATTGGATATTATTATCAAAATATGAAGTTGACCAAATGAAGAAAAAAG 1375
                                                      *********
ALZ 6 -----
6-6-2 AAGAA 1380
      *****
```

Figure 10. Alignment of Predicted ALZ 6 Sequence and Experimental Sequence (cont.)

3.6 Real Time PCR Analysis of ALZ Gene Expression Patterns

In an effort to partially characterize the ALZ genes that have been identified, real time PCR experiments were performed as described. To normalize the results of each experiment, the housekeeping gene IG7 was run with each set of experiments. IG7 is a constitutively expressed ribosomal protein in *Dictyostelium discoideum*, and is therefore a useful housekeeping gene since the threshold cycle is expected to be similar at all time points tested.

Results of all of the real time PCR experiments are presented as average threshold cycle in Table 6. The data from all of the experiments were used to calculate the expression level of each of the genes at each time point, relative to IG7. To obtain a relative expression value for each gene, a modified version of the calculations from Bio-Rad Laboratories Real Time PCR calculation without a standard curve was used. An average threshold cycle number for IG7 was calculated from all time points and this value was used in future calculations. The average threshold cycle number for IG7 was subtracted from each of the threshold cycle values for each of the genes at each of the different time points. The threshold cycle number of IG7 was also subtracted from itself. The value obtained is the ΔCt for each of the genes at the respective time points. The expression value, E, was obtained for each data point by calculating 2 to the exponent ΔCt. This calculation sets the E value for IG7 at 1. Finally, the relative expression value for each data point was obtained by calculating (2log E)/E. This calculation sets the housekeeping gene value to 0.

This method was used to analyze results from all of the ALZ genes within their individual real time PCR experiments. The data from all twelve of the genes were

combined and a relative expression value was calculated for all of the data points (data not shown), and plotted on a bar graph (Fig. 13). Because ALZ 8 and ALZ 10 show expression dramatically higher than all of the other genes and do not allow evaluation of the other 10 genes, unless they are removed from the graph, separate graphs of the expression patterns of each individual gene were created (Figs. 14-25). All of these graphs were made using IG7 as the baseline value.

A standard curve was created for ALZ6 by starting with 10^8 copies of ALZ 6 plasmid DNA and diluting it by 1:10, down to a copy number of 10^1. Each dilution was run in a PCR reaction on the iCycler to determine the Ct, which was used to generate a standard curve and to calculate the copy number of ALZ6 in each sample (Table 7).

Finally, two graphs were created to show the expression of all of the ALZ genes in relation to ALZ 6 (Figs. 26 and 27). The first graph shows all of the ALZ genes (excluding ALZ 8) using ALZ 6 as the baseline value. This was done because ALZ 6 expression was quantified and, as a result, relative expression of the remaining genes in relation to a known quantity can be calculated. The second graph shows a comparison of ALZ 1, 9 and 12, in relation to ALZ 6. These three genes were chosen for the final graph, because each shows points in their expression where their respective expression levels drop below that of ALZ 6. The remaining genes were removed so that these three could be seen clearly.

The calculations for both of these final graphs are the same as those for the IG7 baseline graphs, with the exception that the ALZ 6 data were the baseline data used in the calculations.

Table 6. Real Time PCR Average Threshold Cycle Data. Data collected from real time PCR experiments for all twelve ALZ genes and the housekeeping gene, IG7. Values shown are averages of threshold cycles for each of the genes at the various sampling times. Each sample was analyzed in triplicate. Data for IG7 is an average of all of the housekeeping gene data from each experiment where it was used, eliminating outlying values.

	Vegetative	4 hour	8 hour	12 hour	16 hour	20 hour	24 hour
ALZ 1	35.6	35.7	35.5	35.3	36.3	35.4	36.1
ALZ 2	29.6	25.2	24.4	26.1	27.7	26.8	27.4
ALZ 3	28.9	23.9	23.7	25.3	27.1	26.6	29
ALZ 4	28.9	29.4	26.5	26.7	25.9	25	26.1
ALZ 5	27.9	25.4	23.8	25.5	28.2	26.4	31.9
ALZ 6	31.4	30.8	27.9	29.3	28.3	30.5	31.8
ALZ 7	26.3	25.7	26.1	25.7	25.4	25.9	24
ALZ 8	23.6	22.1	21.2	21.6	22.4	22.1	22.9
ALZ 9	29.6	31.2	30.8	30.6	30.9	32.3	30.3
ALZ 10	26.8	23.6	22.5	23.6	26.5	25.3	27.6
ALZ 11	27.5	29.3	27.4	25.5	27.4	27.9	28.4
ALZ 12	31	30.1	30.3	30.3	31.0	31.5	31.1
IG7	11.1	11.2	11.4	10.5	11.2	11.1	11.2

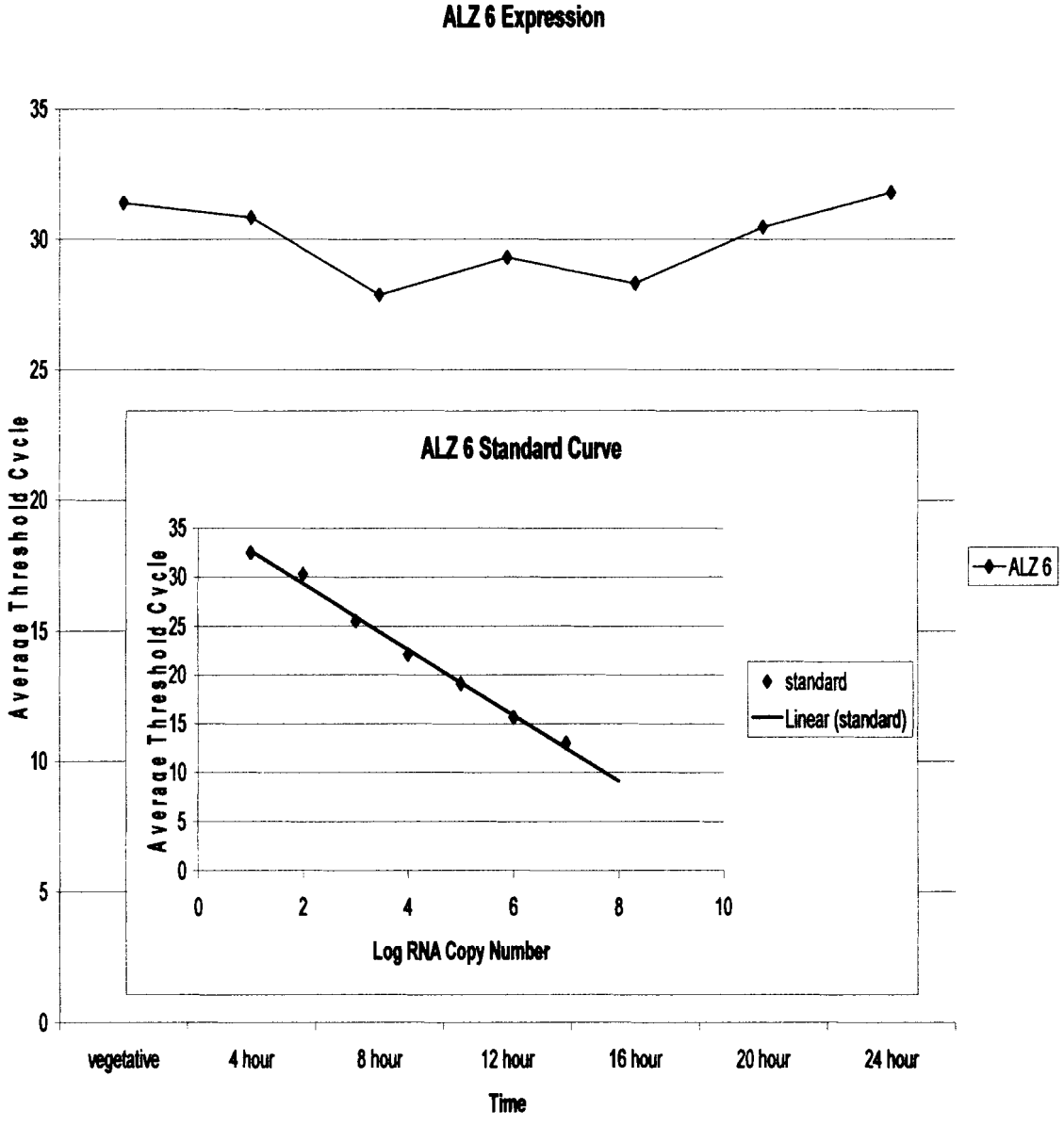


Figure 11. ALZ 6 Standard Curve and Expression. Inset graph is the standard curve created to quantify ALZ 6 expression. The X-axis of this graph shows the log to the base 10 of the RNA copy number used at each point on the graph. ALZ 6 average threshold cycle values are plotted as a separate line on the outer graph. The equation for the standard curve trend line is $y = -3.3643x + 36.05$.

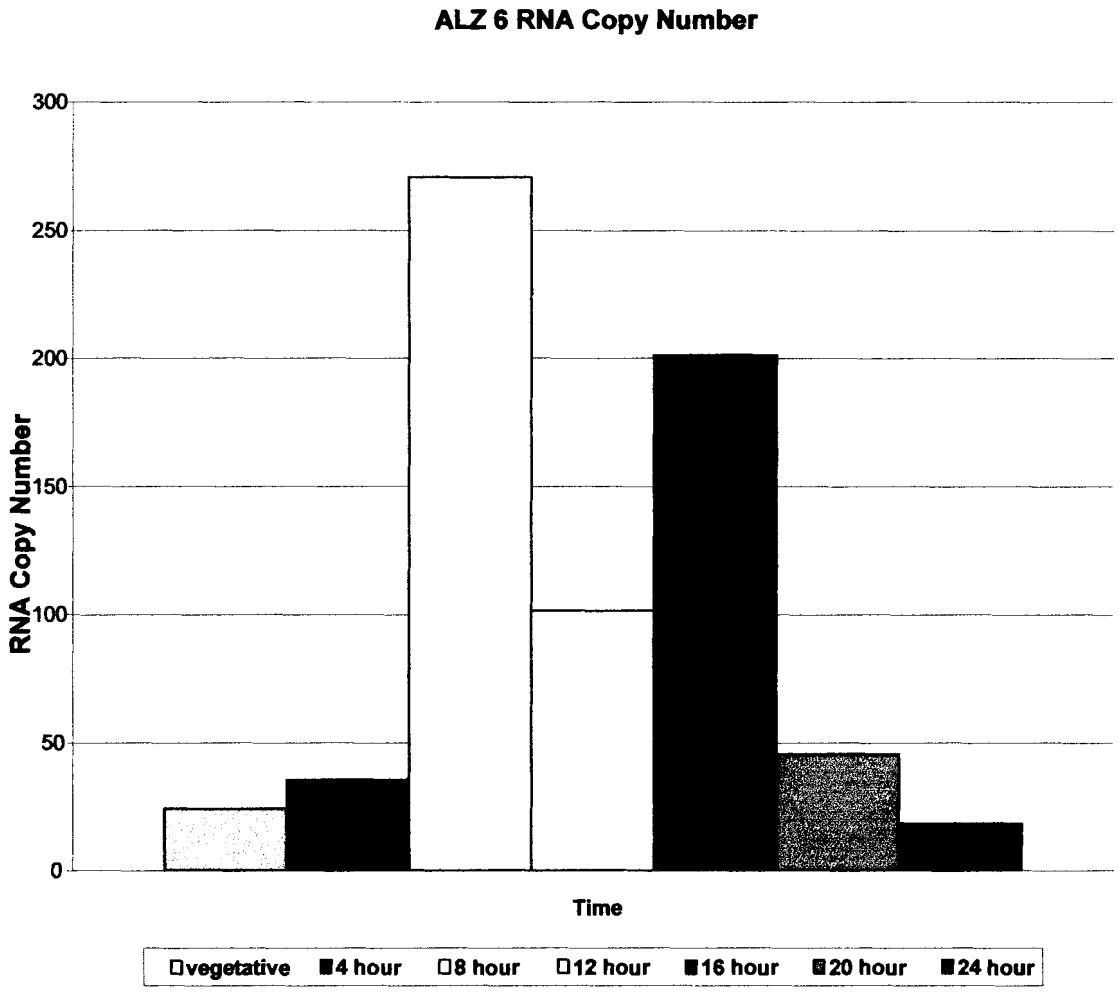


Figure 12. ALZ 6 RNA Copy Number. Data obtained by using the standard curve equation y = -3.3643x + 36.05 from figure 11. Copy number (x) was calculated by inserting the ALZ 6 threshold cycle value (y) for each time point into the equation and solving for x. The inverse log (x) was calculated and plotted onto the graph.

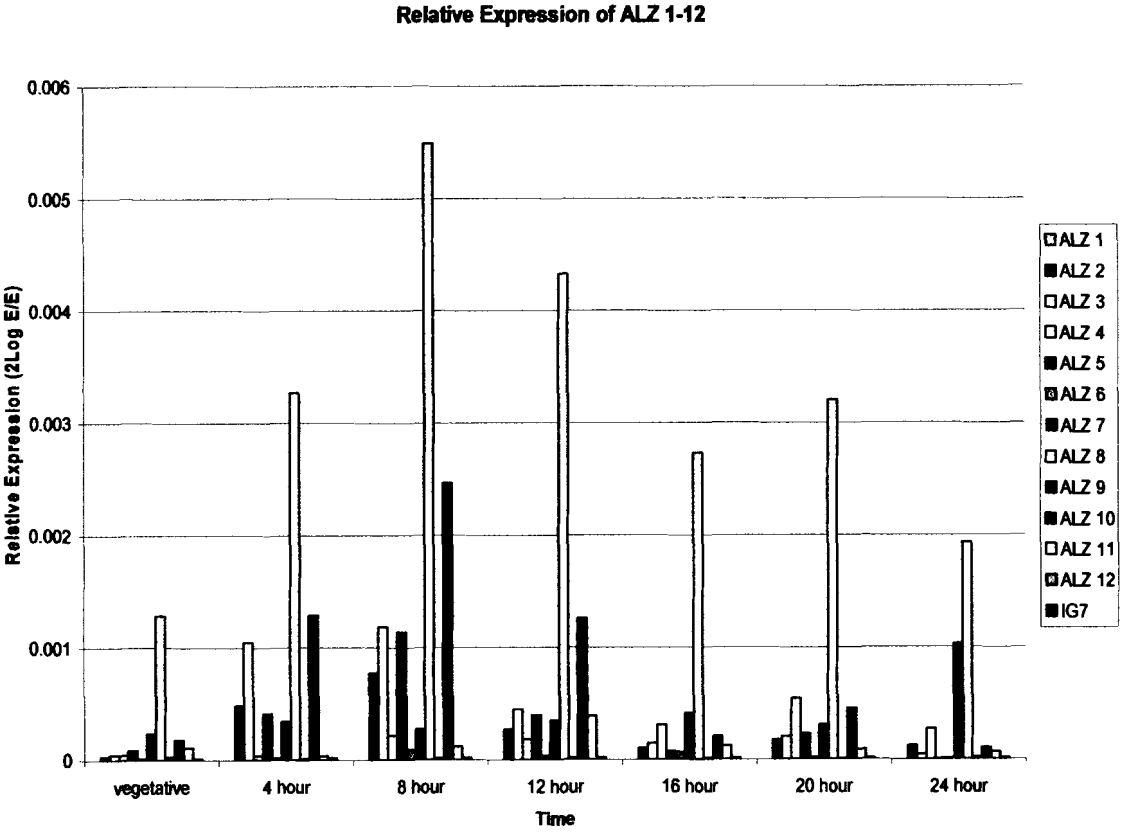


Figure 13. Semi-Quantitative Expression of ALZ Genes 1-12. The housekeeping gene, IG7, is used as the reference gene in this data set, which combines the results of all of the real time PCR experiments into one data set. IG7 expression is set to zero to allow normalized comparisons of the expression of each of the ALZ genes.

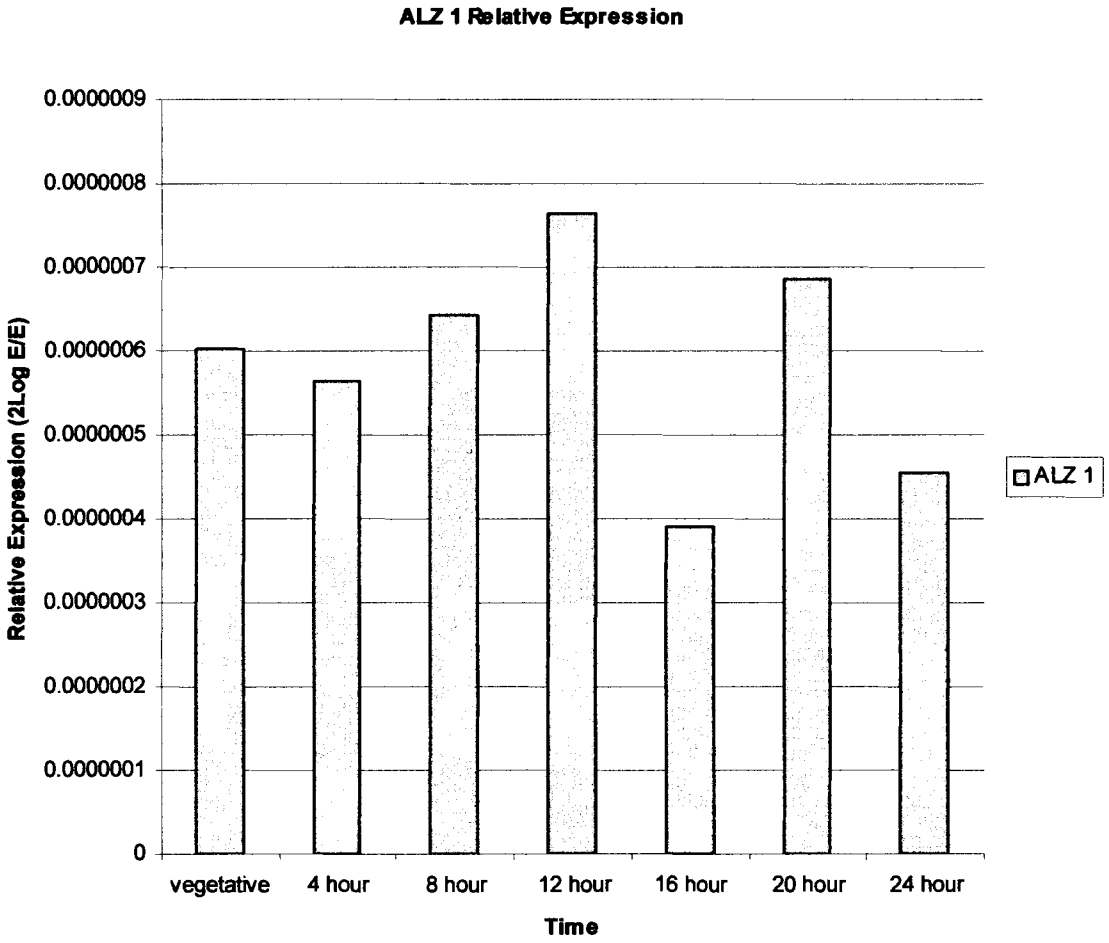


Figure 14. Semi-Quantitative Expression of ALZ 1. Graph shows the relative expression of the ALZ 1 gene, as calculated by normalizing the real time PCR threshold cycle data to the housekeeping gene IG7. The following graphs of ALZ 2-12 show the relative expression of the respective genes, calculated the same way.

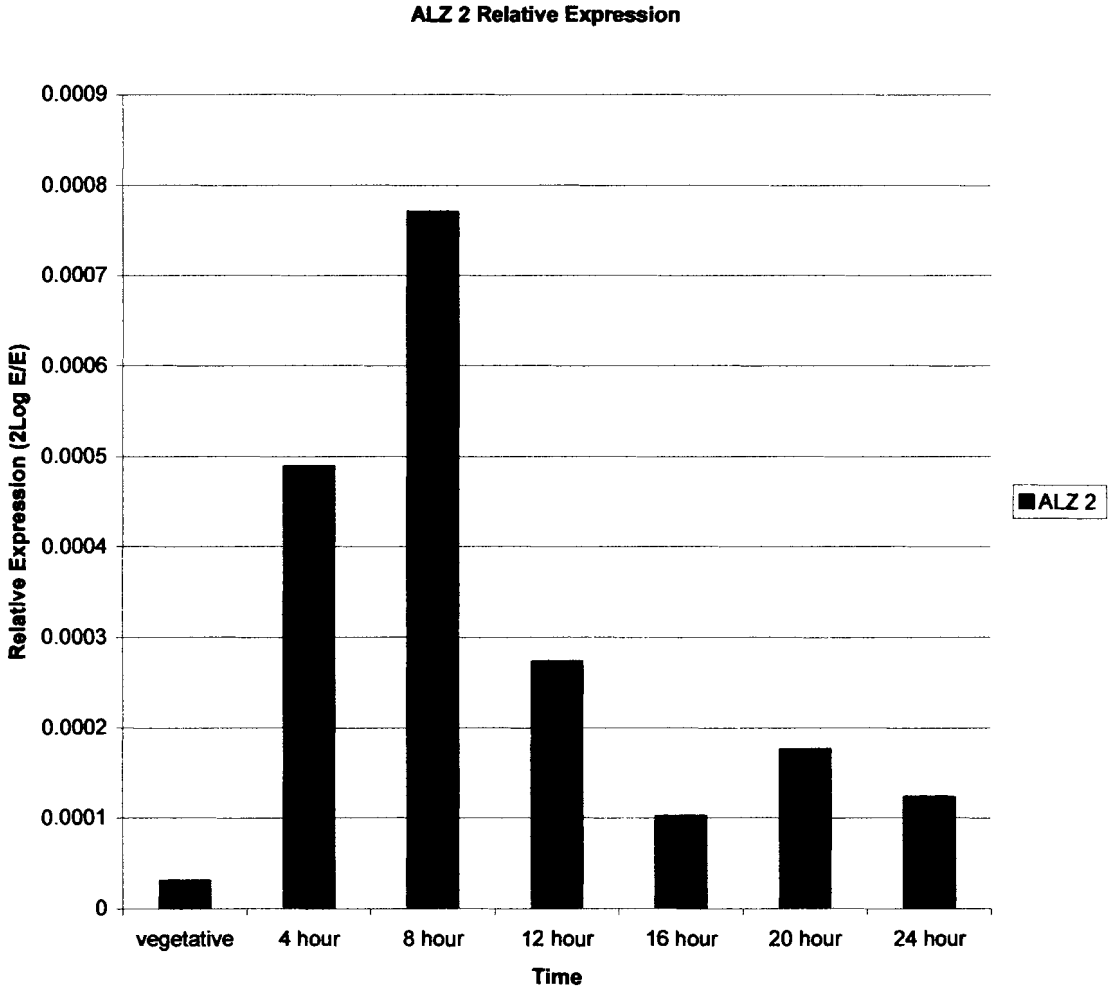


Figure 15. Semi-Quantitative Expression of ALZ 2.

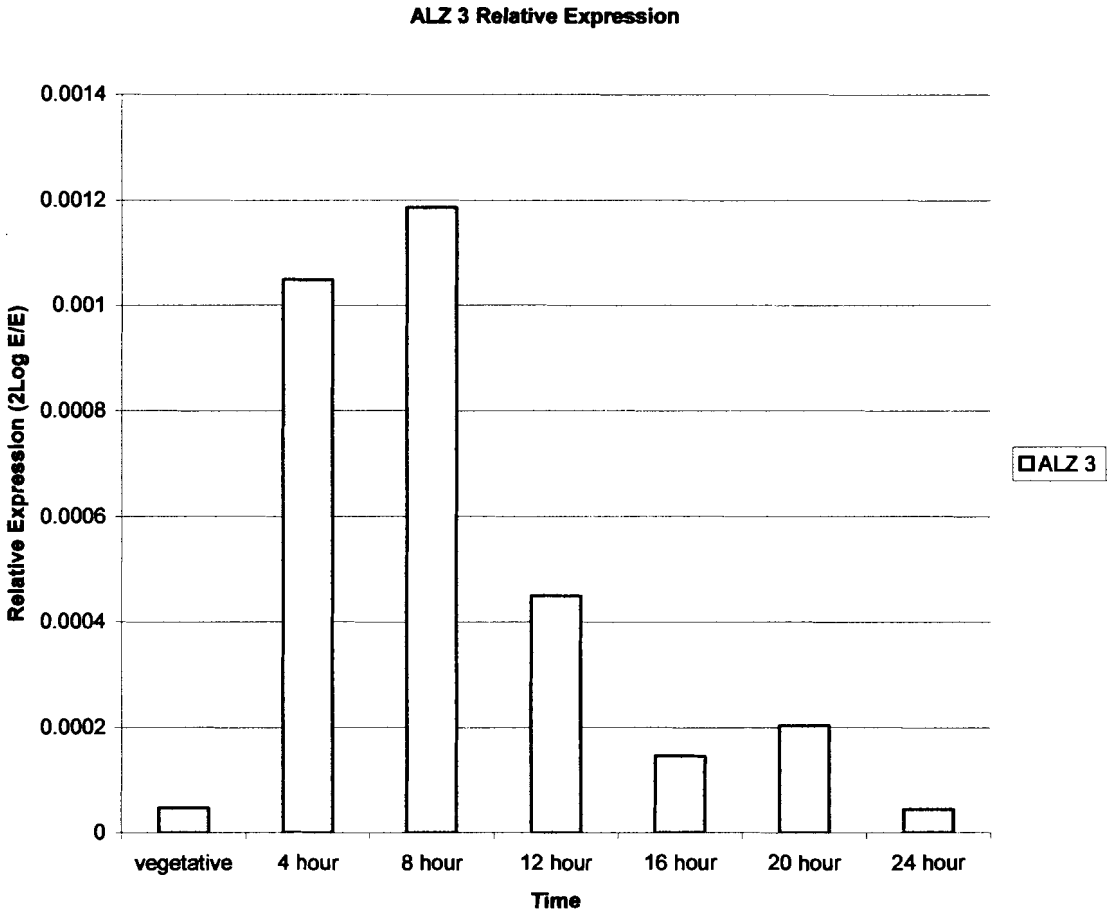


Figure 16. Semi-Quantitative Expression of ALZ 3.

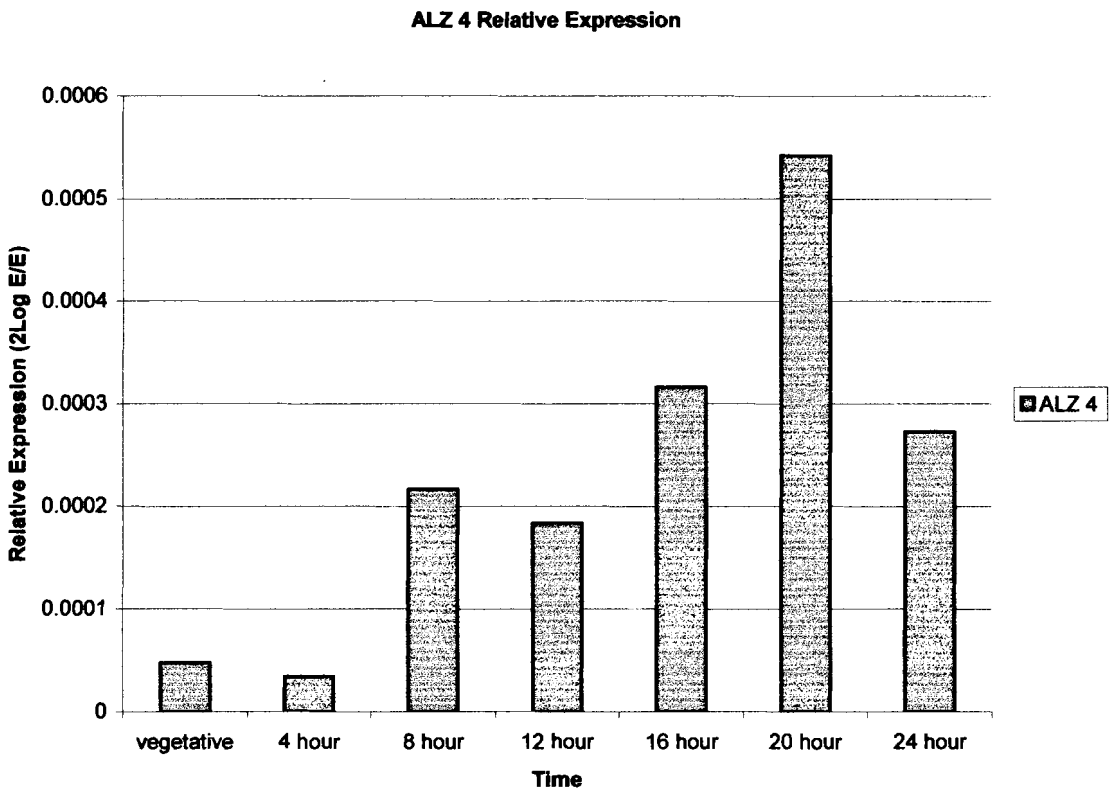


Figure 17. Semi-Quantitative Expression of ALZ 4.

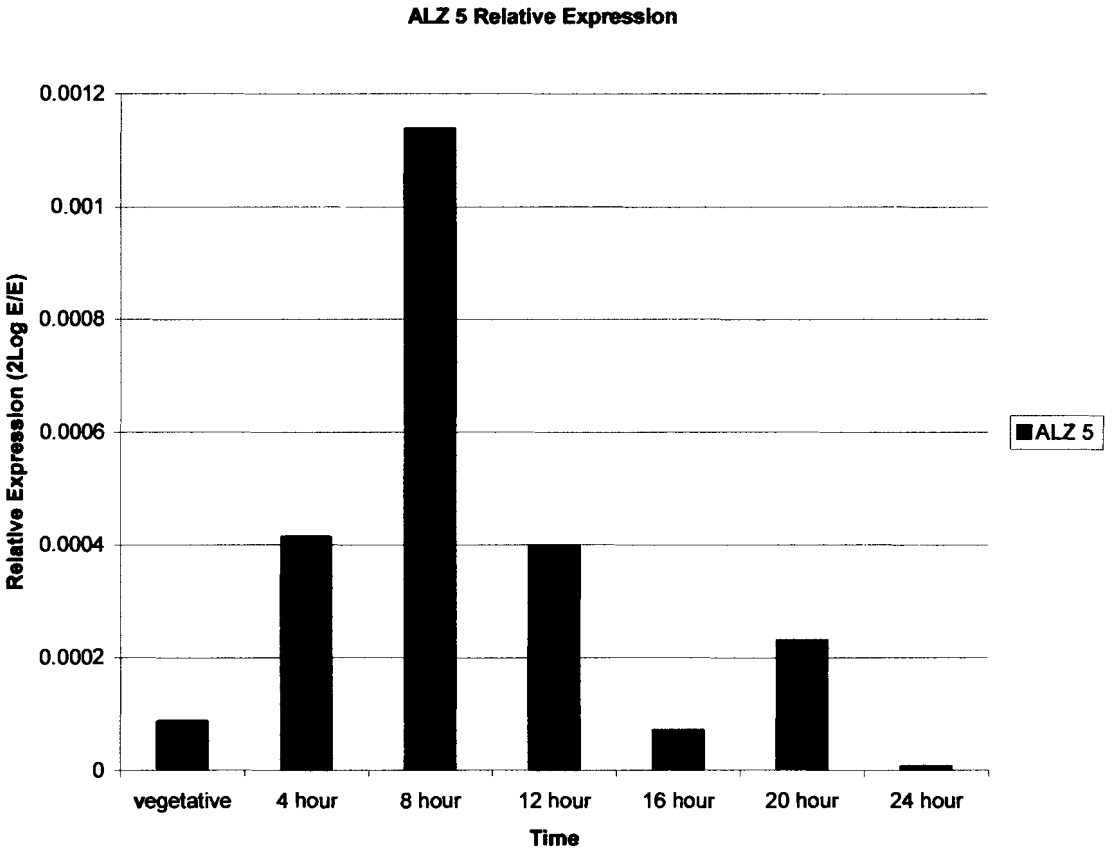


Figure 18. Semi-Quantitative Expression of ALZ 5.

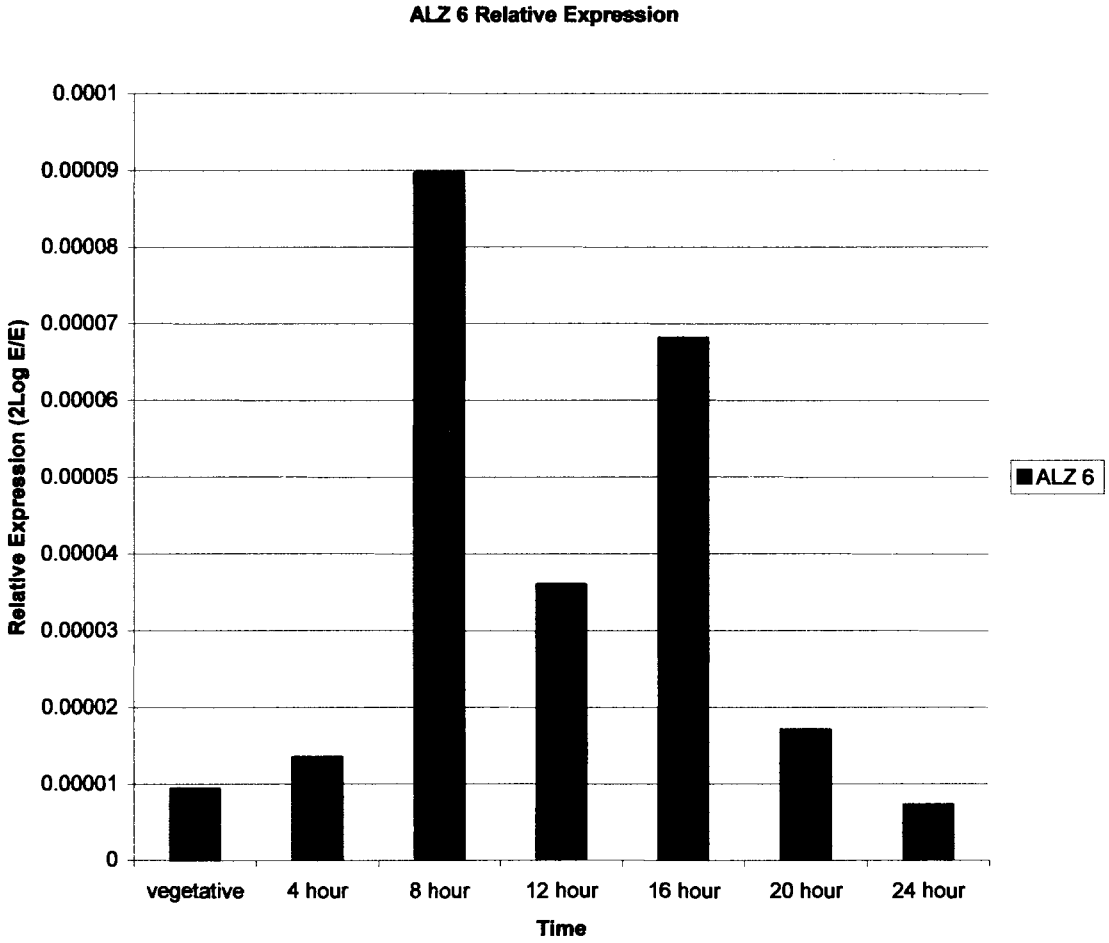


Figure 19. Semi-Quantitative Expression of ALZ 6.

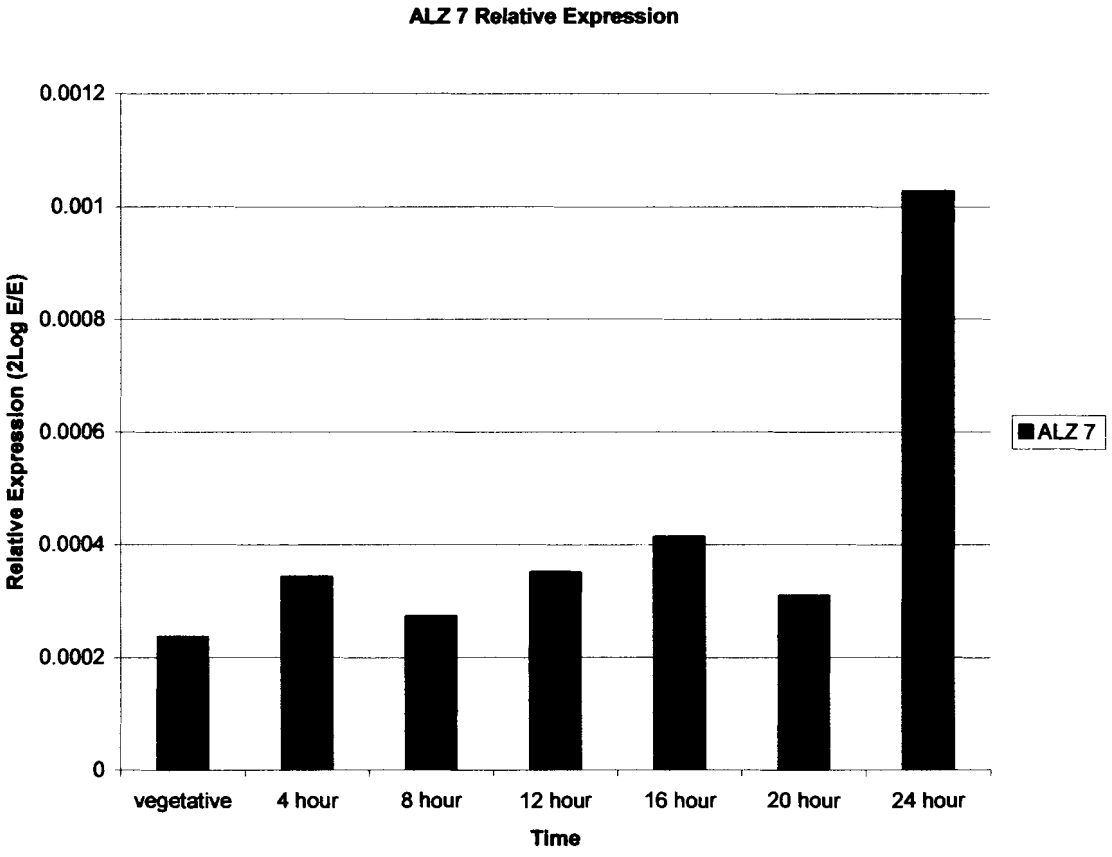


Figure 20. Semi-Quantitative Expression of ALZ 7.

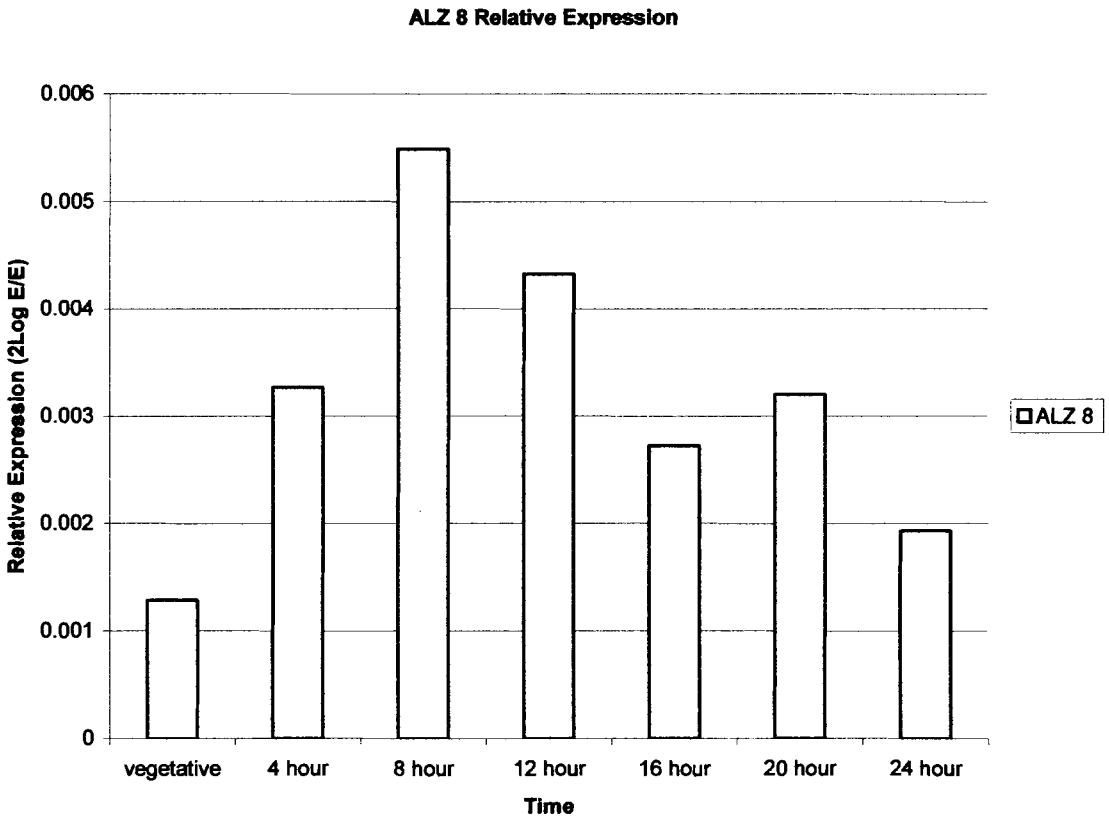


Figure 21. Semi-Quantitative Expression of ALZ 8.

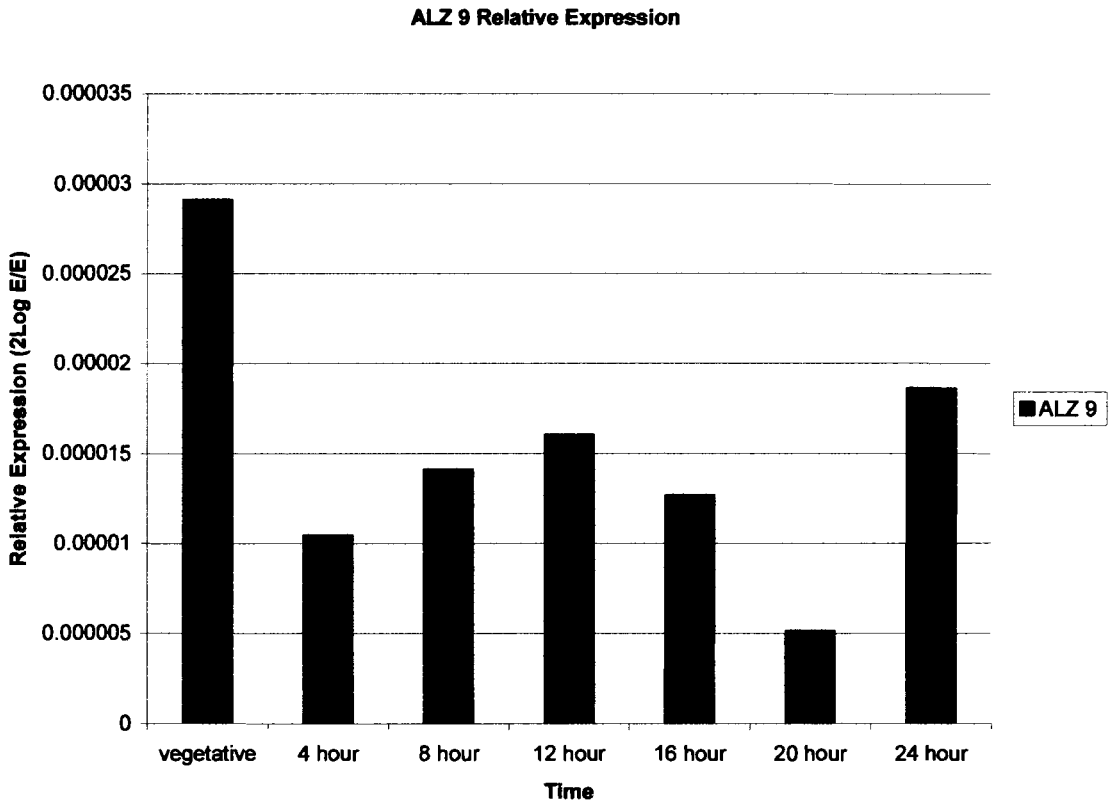


Figure 22. Semi-Quantitative Expression of ALZ 9.

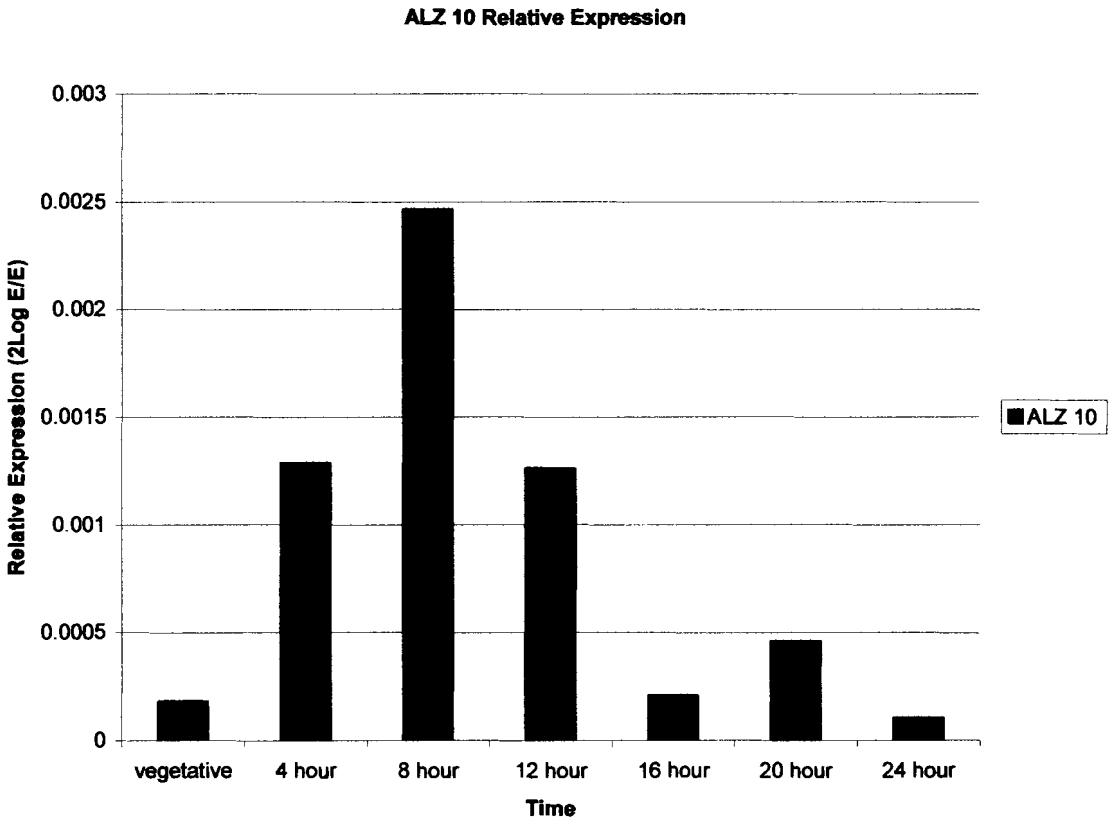


Figure 23. Semi-Quantitative Expression of ALZ 10.

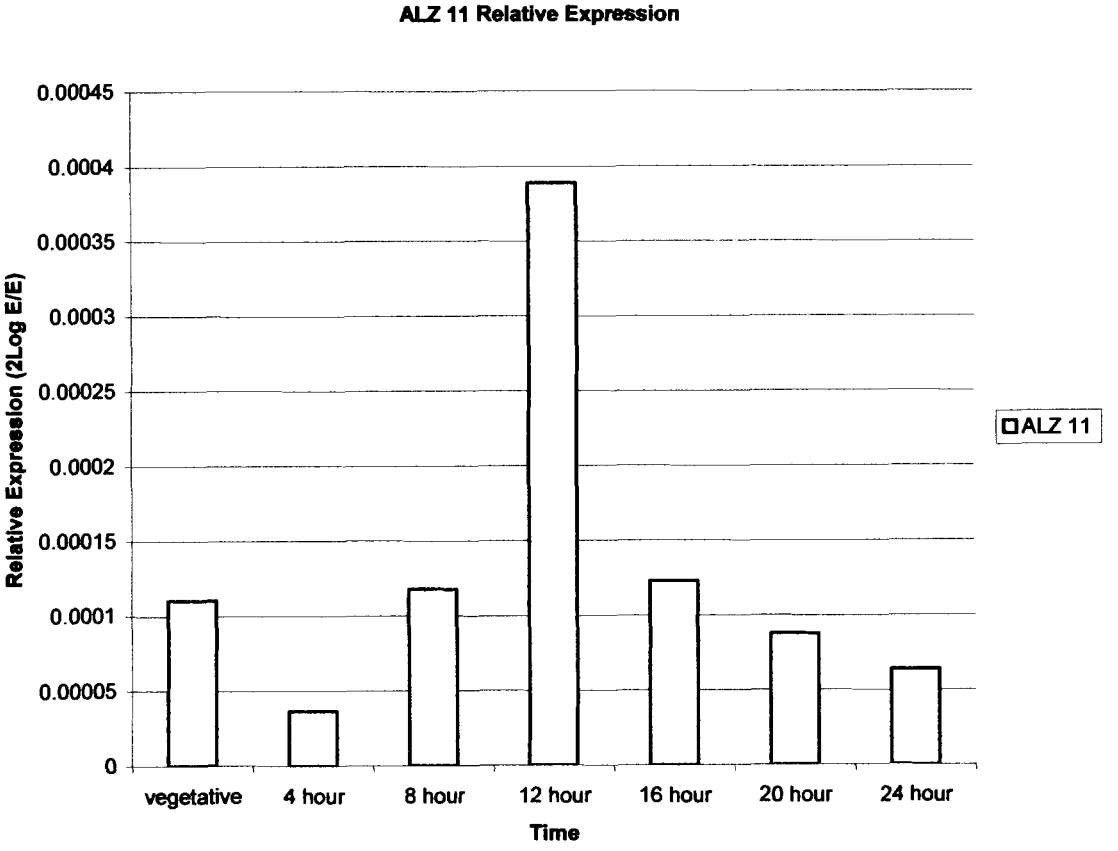


Figure 24. Semi-Quantitative Expression of ALZ 11.

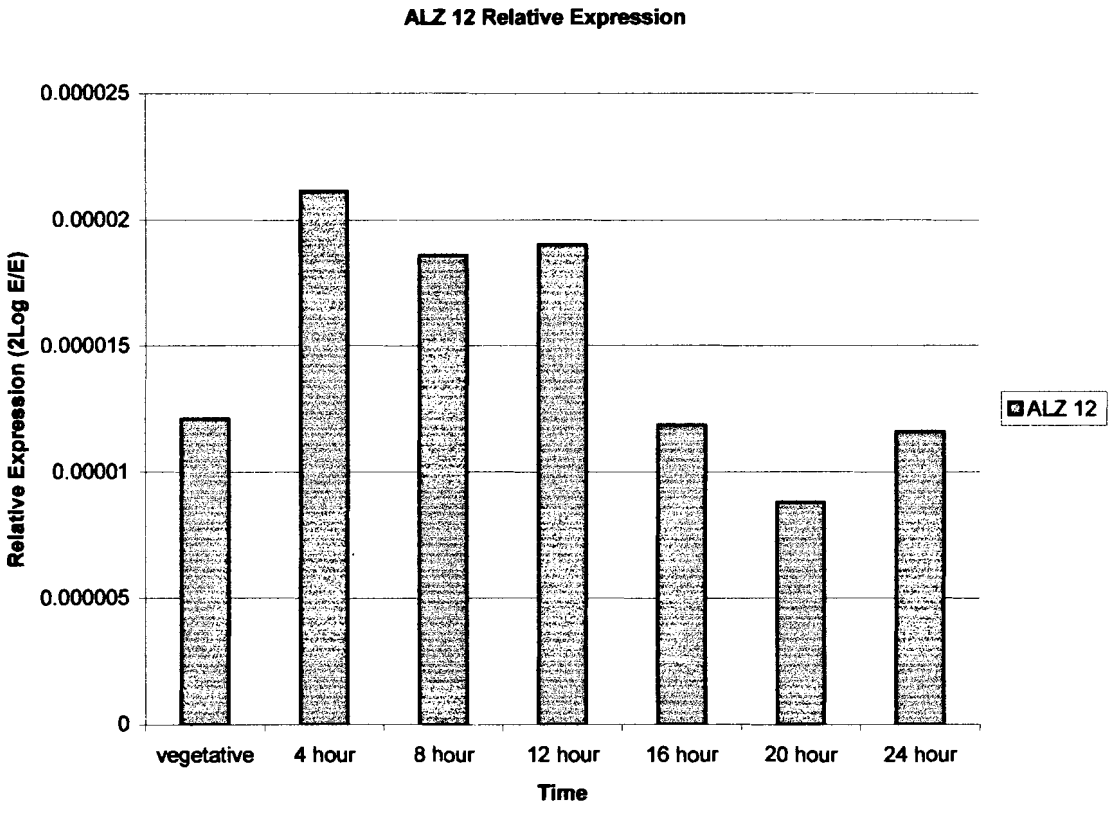


Figure 25. Semi-Quantitative Expression of ALZ 12.

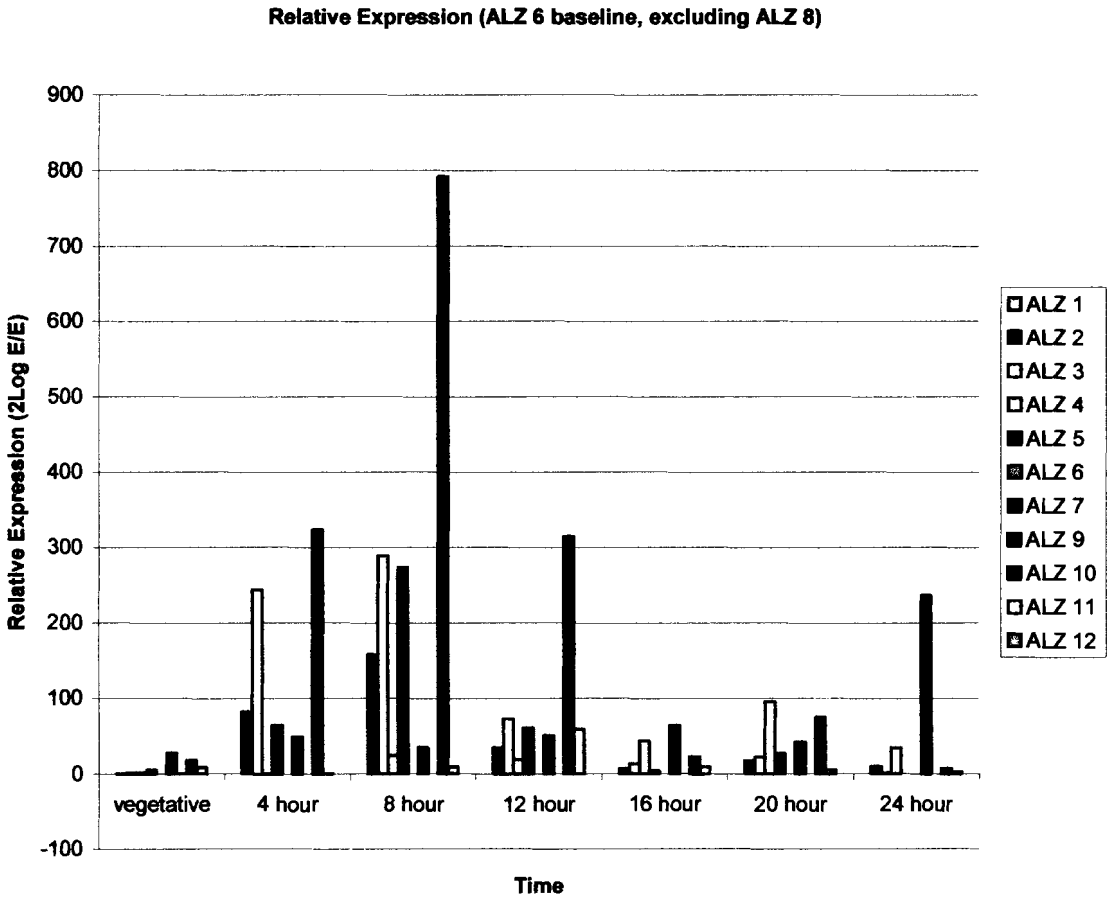


Figure 26. Relative Expression of ALZ 1-12 (Excluding 8) With ALZ 6 as Baseline. A different view of ALZ gene expression, in relation to ALZ 6. ALZ 8 has been removed from the graph to allow expression of the remaining genes to be more easily seen.

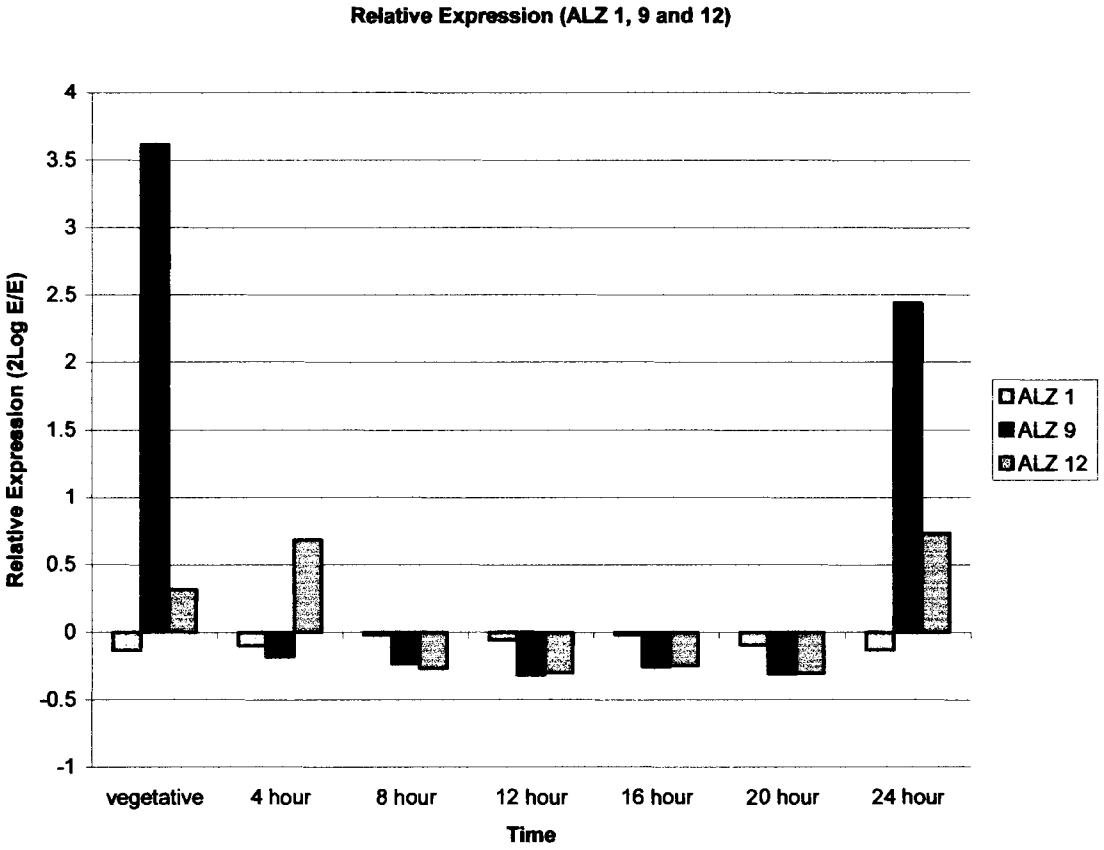


Figure. 27. Relative Expression of ALZ 1, 9 and 12 with ALZ 6 as Baseline. Data for all of the ALZ genes that show expression levels consistently higher than ALZ 6 have been removed, so that the genes that show expression levels lower than ALZ 6 can be viewed.

All of the graphs created with the data from the real time PCR experiments show some interesting and clearly defined expression patterns for ALZ 1-12. ALZ 1 expression (Fig. 14) is shown to be extremely low throughout the entire *D. discoideum* life cycle. In addition, expression of this gene does not fluctuate significantly at any time during the experiments, as shown by the highest ALZ 1 expression level (8 hour) being only twice the level of the lowest expression level (16 hour).

The results of these experiments show that several of the ALZ genes exhibit similar expression patterns, although overall expression levels vary from gene to gene. ALZ 2 expression (Fig. 15), although higher than the previously described ALZ 1, also is shown to be very weak during the vegetative state. By 4 hours of starvation, however, expression has begun to increase sharply and continues to do so through 8 hours starvation, at which time a peak in expression levels is experienced. The peak at 8 hours is approximately 30 times higher than the lowest peak, which is seen during the vegetative stage. By 12 hours starvation, ALZ 2 expression drops dramatically, to levels only 6 times greater than the vegetative state levels. Expression of this gene remains steady throughout the remainder of the life cycle.

ALZ 3 (Fig. 16), 5 (Fig. 18), 8 (Fig. 21) and 10 (Fig. 23) also exhibit the general pattern of gene expression shown by ALZ 2 although the level of expression of each of these genes varies. Of note, these experiments suggest that ALZ 8 expression is clearly the strongest of all of the ALZ genes investigated, during all of the time points tested. The peak of ALZ 10 expression (8 hours starvation) is second only to ALZ 8 in peak intensity.

ALZ 4 expression (Fig. 17) shows two separate peaks. Expression of this gene rises from its lowest levels between the vegetative stage and 4 hours starvation to a level 6-7 times higher at the 8 hours starvation stage. Expression drops by approximately one half at 12 hours starvation and then rises again to the second, higher peak at 20 hours starvation. This second peak is 14 times higher than the lowest point (4 hour) during the life cycle.

The results of ALZ 6 expression experiments are presented in two different forms (Figs. 11, 12 and 19). Expression of ALZ 6 was quantified at all of the time points taken throughout the *Dictyostelium* life cycle (Figs. 11 and 12). As a result of creating a standard curve for ALZ 6, consisting of a serial dilution of ALZ 6 contained within plasmid from 10^1 to 10^8 copies per sample, it is seen that ALZ 6 expression is relatively low during the vegetative state and through 4 hours of starvation, however, 8 hour expression levels are 5 times higher than that of the 4 hour sample. Although this dramatic increase results in the highest levels of this gene throughout the life cycle, similarly high levels of this gene are maintained through 16 hours of starvation. By 20 hours of starvation, ALZ 6 expression drops again to levels similar to the vegetative state and early starvation.

The equation for the slope of the ALZ 6 standard curve is $y = -3.3643x + 36.05$, where y is the threshold cycle number and x is the number of copies of DNA in the sample. When each of the ALZ 6 average threshold cycle numbers are substituted for y in the equation, a range of x values corresponding to the log of the DNA copy number for the particular sample can be obtained. By taking the inverse log of x, the DNA copy

Table 7. ALZ 6 Copy Number. Table shows the approximate number of copies of ALZ 6 RNA calculated to be present in *Dictyostelium discoideum* during each of the time points.

Time point	Log ALZ 6 copy number	ALZ 6 copy number
Vegetative	1.4	24.1
4 hour	1.6	35.5
8 hour	2.4	270.6
12 hour	2.0	101.5
16 hour	2.3	201.2
20 hour	1.7	45.7
24 hour	1.3	18.3

number for each of the ALZ 6 samples was calculated. The log of the copy number and the calculated copy number are shown in Table 7.

ALZ 7 (Fig. 20) exhibits an unusual expression pattern for this family of genes, in that expression of this particular gene is relatively stable from the vegetative state through 20 hours of starvation, before any significant change occurs. There are small fluctuations in expression during the 20-hour time frame, such as minor peaks at 4 and 16 hours starvation; however, neither of these fluctuations appears to be very significant. By 24 hours of starvation, however, a much larger peak in expression is seen, resulting in the highest point for this gene in the life cycle. This major peak is approximately 5 times greater than the lowest point (12 hours) in the cycle.

ALZ 9 (Fig. 22) expression is consistently low during the entire *Dictyostelium* life cycle, in comparison to a majority of the other ALZ genes. Expression of this gene appears to be highest during the vegetative state, after which it drops off at 4 hours starvation and remains consistently low throughout the rest of the 24-hour life cycle. Secondary peaks are seen, however, at 8 hours and 24 hours starvation. As with ALZ 8, the difference between the high point of expression during the vegetative state and the lowest point of expression at 20 hours starvation (ratio of 6:1) is not as great as that seen with several of the other ALZ genes.

ALZ 11 (Fig. 24) expression is unique among all of the genes detailed in this work. Expression of this gene starts out with a secondary peak during the vegetative state and quickly decreases at the onset of starvation. By four hours of starvation, the lowest levels of expression are reached in this gene and expression levels begin to climb again by 8 hours starvation. This increase continues to 12 hours starvation, at which time the

primary peak occurs. After this peak, expression decreases gradually to the end of the life cycle, although it does not decrease to the point seen at 4 hours starvation. The primary peak at 12 hours starvation is approximately 7 times greater than the 4-hour low point.

Finally, ALZ 12 (Fig. 25) expression is also unique, in that this gene maintains the primary expression peak through two time points. ALZ 12 expression through most of the time points is consistent and relatively low. At 4 hours starvation, though, expression increases to the primary peak and this peak is maintained through 8 hours starvation. By 12 hours, expression drops to nearly half of the peak value, and this level is maintained through 16 hours. By 20 hours, the low point of expression is reached; and it is at this time that the high point of expression is approximately 3 times greater than the low point.

An analysis of ALZ gene expression, in relation to ALZ 6, was also performed to obtain a better understanding of the quantities of ALZ 1-12 mRNA molecules that are found in the cell during each of the time points investigated. Although this analysis does not allow for quantification of any message copy numbers other than ALZ 6, it does still allow us to search for patterns that may be beneficial in future work. In doing this analysis, it is seen that all but three of the ALZ genes are consistently expressed at higher levels than ALZ 6 (Fig. 26). ALZ 1 (Fig. 27) is consistently expressed in lesser quantities than ALZ 6 throughout the entire life cycle. ALZ 9 (Fig. 27) is expressed at significantly higher levels than ALZ 6 during the vegetative state and 24 hours of starvation, however, between those times, expression of this gene drops below that of ALZ 6. Finally, ALZ 12 (Fig. 27) expression levels drop below that of ALZ 6 from 8 hours starvation through 20 hours, but are significantly higher during the vegetative state, 4 hours and 24 hours starvation.

4. DISCUSSION

We show by sequence homology that the single cell soil amoeba *Dictyostelium discoideum* possesses a family of at least twelve putative Erf2/Akr1 homologs that may function as palmitoyltransferases during the organism's life cycle. Of particular interest is the observation that the DHHC domain is completely conserved at the amino acid level in all twelve of these proteins. On either side of this domain, within a range of approximately 50 amino acids, there are a large number of amino acids that are highly conserved across all twelve ALZ proteins, as well as across a significant number of DHHC containing proteins in different species. The function of this domain is not yet understood, although, it is speculated that the DHHC domain may play a role in protein-protein or DNA-protein interactions (Putilina *et al.,* 1999). Recent work has also suggested that the domain also plays a critical role in palmitoylation of the substrate protein. Roth *et al.* (2002) present evidence that the Asp-His dipeptide and the Cys of the DHHC domain both are necessary for autopalmitoylation and substrate palmitoylation, as seen by the loss of palmitoylation function when these amino acids are mutated.

Lobo *et al.* (2002) specifically suggest that the cysteine rich domain may also play a vital role in substrate palmitoylation. Mutation of C189 to a Ser also abolished palmitoylation of Ras. Furthermore, they present evidence that the DHHC domain is involved in the direct transfer of palmitate to the substrate. Mutation of H201 to Ala abolishes palmitoylation of Ras in yeast; however, it does not abolish binding of palmitate to Erf2p.

The results presented here add reliability to the theory that the DHHC region is essential to the function of these proteins, especially when it is considered that outside the DHHC-CRD region there is little significant homology between the ALZ proteins and any other known proteins. Notable exceptions to this observation are the ALZ proteins exhibiting ankyrin repeats at the N-termini. It is possible that the DHHC region is the site where the palmitoyltransferase binds to the substrate protein, thus stabilizing the protein-protein interaction and allowing the palmitate molecule to be transferred to the substrate. While sequence homology does not necessarily indicate identical function, it does provide a starting point for work to be done to identify functional homologs to previously identified palmitoyltransferases.

Attempts to verify the gene sequences for all twelve of the ALZ genes have met with mixed results. As of this writing, significant experimental sequence data has been gathered only for ALZ 6; however, these results appear to contain several errors that put the complete sequence data in question. Most of these errors do not cause translational errors, but one of the errors does cause a stop codon to be inserted into the sequence prematurely. These differences in the available gene sequence and our sequencing results for the same gene are almost certainly due to errors during PCR or mis-reads during sequencing. A comparison of the sequence with the available sequence leads us to consider that we have isolated the ALZ 6 gene, but because of the errors we will have to sequence additional clones to determine the true ALZ 6 cDNA sequence.

In searching for predicted transmembrane domains in these genes, it has become apparent that, although the ALZ DHHC-CRD region is homologous to both Erf2p and Akr1p, the number of predicted transmembrane domains as well as the presence of

ankyrin repeats, makes some of these proteins more similar to Akr1p than to Erf2p. Further analysis of each of these proteins in the laboratory will allow for a more concrete identification of this family. This will also help in determining whether there are any more structural or functional differences between members of the Akr1 and Erf2 families of proteins.

Transmembrane domain analyses using computer models can only, by necessity, be speculative, since they are based on hydrophobicity of the amino acid sequence. A more precise determination of the number and location of transmembrane domains in the ALZ proteins will only come through experimentation in the laboratory. Such experimentation would require developing antibodies to selected sections of the amino acid sequences of the ALZ proteins. When the antibodies are added to an intact cell membrane, the antibodies will attach to the antigen if it is exposed to the environment. If the antigen is buried in the membrane, the antibodies will not bind and can easily be washed off the membrane.

In addition to confirming our predictions concerning the number of transmembrane domains in each of the proteins, it will be useful to compare the number and location of each of the transmembrane domains in the ALZ proteins to other DHHC proteins in other species. If it is shown that multiple DHHC proteins have transmembrane domains in the same location along the amino acid sequence, this will provide an indication that those domains may possess a specific importance to the function of the protein. We have shown, in this work, that the individual members of the family of ALZ proteins in *D. discoideum* show sequential and topological similarities to both Erf2 and to Akr1 in *E. coli*. Additional sequential and topological analysis of DHHC proteins in other

species will serve to further verify the ubiquitous nature of this family of proteins. With DHHC proteins known to exist in a large number of eukaryotic species, a determination of the sequential and structural similarities will add greatly to the ability to determine the function of this family.

The results of the real time PCR experiments presented here are useful in yielding only relative expression values. While not a quantitative value, relative values are still informative, in that they are able to show how much more each gene is expressed, relative to a the housekeeping gene and to each other, in these experiments. The results of the real time PCR experiments with ALZ 1 suggest that this gene is only expressed in very low numbers in relation to the other 11 genes, regardless of the time point in the life cycle. Another possibility is that the primers used are not optimally designed, causing them to mis-anneal often. Both hypotheses can only be tested by running the real time PCR experiments again, with a different set of ALZ 1 primers.

A less likely scenario is that ALZ 1 may actually be a pseudogene. Verification of this possibility can only come with *in vitro* and *in vivo* palmitoylation studies and gene knockouts. It is also possible that ALZ 1 may actually be expressed constitutively throughout the *Dictyostelium* life cycle, and that the low but constant level of expression at every time point may serve to keep palmitoylation of a specific substrate at consistent levels.

Another interesting observation is that ALZ 8 appears to be expressed much more strongly than any of the other genes, throughout the entire life cycle. This result may be misleading, however, when compared with the results from the RT-PCR experiment, which shows that there are two well-defined bands of product in the lane corresponding

to ALZ 8. Both of these DNA fragments will bind the SYBR Green dye during real time PCR and, therefore, will most likely produce a false threshold cycle. The results for ALZ 2, 4, 9 and 10 may also be deceptive because of the same problem, although, it appears that these genes may be affected to a lesser degree. As a result, caution must be used in interpreting the data from these experiments.

Also of significance in regards to the RT-PCR experiments is the observation that the secondary smaller band seen in the ALZ 8 lane is noticeably brighter than the band of interest. A possible explanation for this observation is that since this fragment is significantly smaller than ALZ 8, more copies of this fragment might be present in the aliquot used for the RT-PCR experiments. This would cause the primers to preferentially anneal to this fragment, and therefore, amplify it, instead of ALZ 8. A second possible explanation is that since ALZ 8 is considerably larger than the secondary fragment, the processivity of the polymerase may cause the enzyme to prematurely fall off the ALZ 8 template more frequently, therefore causing the secondary fragment to be amplified more frequently. To alleviate these potential problems in all the genes that show non-specific banding, a new set of more specific primers will need to be designed for all of the genes that show secondary RT-PCR bands.

Overall, some definitive expression patterns stand out from these experiments. The data suggest that ALZ 2, 3, 5, 8 and 10 may be interesting candidates for the enzyme that catalyzes the transfer of palmitate to the Gα2 subunit of *D. discoideum*. Each of these proteins experiences a peak in expression between 4 and 12 hours of starvation, in relation to the house keeping gene (IG7) and the other ALZ genes, which makes these genes particularly active during *Dictyostelium* aggregation. All of the other ALZ genes,

with the previously described exception of ALZ 1, show varying levels of expression during the life cycle; however, they do not appear to show selective expression during the aggregation stage. These proteins may serve as palmitoyltransferases for any or all of the seven other Gα subunits in *Dictyostelium*, as well as other proteins in the organism that require palmitoylation.

The work detailed in this thesis raises many different questions. One of the primary questions is whether these proteins actually function as palmitoyltransferases. Answering this question will first require *in vitro* palmitoylation studies using various proteins such as Gα2 as a substrate to which the palmitate can bind. Before this can take place, however, complete cDNA sequences for each of the ALZ genes will need to be acquired and then transformed into either bacterial expression plasmids or *Dictyostelium*, itself. Once this has been accomplished, radiolabeled palmitate can be added to the *in vitro* system to determine whether the lipid becomes covalently bound to the substrate protein. To further confirm the ALZ genes role in palmitoylation, mutations may be introduced into the gene sequence in an effort to determine whether palmitoylation can be abolished. In an *in vitro* system, if palmitoylation of the substrate no longer occurs when it previously had prior to introduction of mutations, this would further suggest that the ALZ gene in question is involved in palmitoylation.

If the ALZ gene products are shown to be palmitoyltransferases, it will be interesting to determine substrate specificity for each of the enzymes. Previous work has shown that palmitoyl-CoA is the substrate of choice for the palmitoylation process and it will be noteworthy to determine whether this is true for these proteins. Additionally, work will need to be done to determine whether these enzymes can use other lipids as

substrates, or whether they are restricted to palmitate. Finally, because at least three of the ALZ proteins appear to show similarity to Akr1p, rather than Erf2p, work will need to be done to determine whether these proteins can function as palmitoyltransferases on their own, or whether they require other proteins, such as the yeast Erf4p, to properly function.

Erf4p presents an unusual problem because it contains no distinguishable sequence homology to any other known proteins. This may make the study of Erf2 homologs difficult, as it has been shown to require Erf4 to function properly in *E. coli*. It is not clear, however, if Erf2 homologs in other species also require a chaperone-like companion protein such as Erf4. After determining that the ALZ proteins are actually palmitoyltransferases, it may be useful to conduct *in vitro* experiments using the ALZ proteins and a clone of Erf4. If there is a need for a companion protein in systems that are devoid of Erf4, one would expect to see a lack of substrate palmitoylation, while in experiments with Erf4 present, measurable substrate palmitoylation may be observed.

In an effort to determine the role of the ALZ genes in the overall life cycle of *Dictyostelium*, work will be done to determine the effects of knocking out each of the genes. It is not known whether any of the ALZ genes are required for *Dictyostelium* survival. We suggest that at least one of these genes is involved in aggregation of *Dictyostelium* during starvation and therefore, we would expect that when that gene is knocked out, aggregation may be altered, even though the cells would be going through starvation.

The DHHC domain of the ALZ proteins and their homologs is suggested to be a zinc finger domain. If this is true, then these proteins will have a specific zinc

requirement to function properly. At this time, no work has been done to detail the zinc requirement for any of this family of proteins. It will be interesting to determine what that requirement is in an *in vitro* system, as well as in an *in vivo* set of experiments. To do this, a chelator, such as EDTA, will be used to bind Zinc and each ALZ protein will be tested to determine whether zinc is required for palmitoylation to occur.

Finally, the ankyrin repeats found in ALZ 5, 7 and 11 need to be investigated to determine their importance to the function of these proteins. Little is known about ankyrin repeat segments, but it is thought that ankyrin is a structural protein, possibly anchoring cytoplasmic proteins to the cytoskeleton. It is possible that ankyrin repeat segments also perform a similar function, and we suggest that these segments may provide extra stability in anchoring ALZ 5, 7 and 11 to their locations within the cell.

As these putative palmitoyltransferases are further characterized and a greater understanding of their action is gained through *in vitro* and *in vivo* studies, it is anticipated that this information will help to further elucidate the action of a wide variety of signal transduction pathways in a vast number of organisms. In humans, a majority of hormones work through G-proteins and other palmitoylated proteins, and therefore, if we can gain a better understanding of the palmitoylation process, it will open a new area of potential therapeutic areas, as palmitoyltransferases could then be studied as drug and gene therapy targets.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

Bartels DJ, Mitchell DA, Dong X, and Deschenes RJ (1999). Erf2, a Novel Gene Product That Affects the Localization and Palmitoylation of Ras2 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 19:6775-87.

Belanger C, Ansanay J, Qanbar R, and Bouvier M (2001). Primary sequence requirements for S-acylation of β2-adrenergic receptor peptides. *FEBS Lett.* 499:59-64.

Berger M and Schmidt MF (1984). Identification of acyl donors and acceptor proteins for fatty acid acylation in BHK cells infected with Semliki Forest virus. *EMBO J.* 3:713-9.

Berger M and Schmidt MF (1986). Characterization of a protein fatty acylesterase present in microsomal membranes of diverse origin. *J. Biol. Chem.* 261:14912-8.

Bizzozero OA, Bixler HA, and Pastuszyn A (2001). Structural determinants influencing the reaction of cysteine-containing peptides with palmitoyl-coenzyme A and other thioesters. *Biochim. Biophys. Acta* 1545:278-88.

Bizzozero OA, McGarry JF, and Lees MB (1987). Autoacylation of myelin proteolipid protein with acyl coenzyme A. *J. Biol. Chem.* 262:13550-7.

Bohm S, Frishman D, and Mewes HW (1997). Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. *Nucleic Acids Res.* 25:2464-9.

Bourne HR (1997). How receptors talk to trimeric G proteins. *Curr. Opin. Cell Biol.* 9:134-42.

Bradbury FA, Kawate N, Foster CM, and Menon KM (1997). Post-translational processing in the Golgi plays a critical role in the trafficking of the luteinizing hormone/human chorionic gonadotropin receptor to the cell surface. *J. Biol. Chem.* 272:5921-6.

Burgoyne RD (1987). G proteins: control of exocytosis. *Nature* 328(6126):112-3.

Camp LA and Hofmann SL (1993). Purification and properties of a palmitoyl-protein thioesterase that cleaves palmitate from H-Ras. *J. Biol. Chem.* 268:22566-74.

Carr SA, Biemann K, Shoji S, Parmelee DC, and Titani K (1982). n-Tetradecanoyl is the NH2-terminal blocking group of the catalytic subunit of cyclic AMP-dependent protein kinase from bovine cardiac muscle. *Proc. Natl. Acad. Sci. USA* 79:6128-31.

Casey PJ (1995). Protein lipidation in cell signaling. *Science* 268(5208):221-5.

Casey PJ and Gilman AG (1988). G Protein Involvement in Receptor-Effector Coupling. *J. Biol. Chem.* 263:2577-80.

Casey PJ, Solski PA, Der CJ, and Buss JE (1989). p21ras is modified by a farnesyl isoprenoid. *Proc. Natl. Acad. Sci. USA* 86:8323-27.

Chamoun Z, Mann RK, Nellen D, von Kessler DP, Bellotto M, Beachy PA, and Basler K (2001). Skinny hedgehog, an acyltransferase required for palmitoylation and activity of the hedgehog signal. *Science* 293:2080-4.

Chaudhary J and Skinner MK (2002). Identification of a novel gene product, Sertoli cell gene with a zinc finger domain, that is important for FSH activation of testicular Sertoli cells. *Endocrinology* 143:426-35.

Clapham DE and Neer EJ (1997). G protein beta gamma subunits. *Annu. Rev. Pharmacol. Toxicol.* 37:167-203.

Connolly T, and Gilmore R (1986). Formation of a functional ribosome-membrane junction during translocation requires the participation of a GTP-binding protein. *J. Cell Biol.* 103:2253-61.

Coukell MB, Lappano S, and Cameron AM (1983). Isolation and characterization of cAMP unresponsive (frigid) aggregation deficient mutants of *Dictyostelium discoideum*. *Dev. Genet.* 3:283-97.

Dujon B, Albermann K, Aldea M, Alexandraki D, Ansorge W, Arino J, Benes V, Bohn C, Bolotin-Fukuhara M, Bordonne R, Boyer J, Camasses A, Casamayor A, Casas C, Cheret G, Cziepluch C, Daignan-Fornier B, Dang DV, de Haan M, Delius H, Durand P, Fairhead C, Feldmann H, Gaillon L, Kleine K, *et al.* (1997). The nucleotide sequence of Saccharomyces cerevisiae chromosome XV. *Nature* 387(6632 Supplement):98-102.

Duncan JA and Gilman AG (1998). A cytoplasmic acyl-protein thioesterase that removes palmitate from G protein α subunits and $p21^{RAS}$. *J. Biol. Chem.* 273:15830-7.

Faber PW, Barnes GT, Srinidhi J, Chen J, Gusella JF, and MacDonald ME (1998). Huntingtin interacts with a family of WW domain proteins. *Hum. Mol. Genet.* 9:1463-74.

Feng Y and Davis NG (2000). Akr1p and the type I casein kinases act prior to the ubiquitination step of yeast endocytosis: Akr1p is required for kinas localization to the plasma membrane. *Mol. Cell. Biol.* 20:5350-9.

Fey P, Gaudet P, Just EM, Dyck PA, Kibb, WA, and Chisholm RL "dictyBase" [online]. Available:http://www.dictybase.org/ (2003,October 31).

Folch J and Lees M (1951). Proteolipides, a new type of tissue lipoproteins. Their isolation from brain. *J. Biol. Chem.* 191:807-17.

Gilman AG (1987). G proteins: Transducers of receptor-generated signals. *Annu. Rev. Biochem.* 56:615-49.

Hamm HE (1998). The Many Faces of G Protein Signaling. *J. Biol. Chem.* 273:669-72.

Hancock JF, Magee AI, Childs JE, and Marshall CJ (1989). All ras proteins are polyisoprenylated but only some are palmitoylated. *Cell* 57:1167-77.

Hescheler J, Rosenthal W, Trautwein W, and Schultz G (1987). The GTP-binding protein, Go, regulates neuronal calcium channels. *Nature* 325:445-7.

Higgins JB and Casey PJ (1994). In vitro processing of recombinant G protein gamma subunits. Requirements for assembly of an active beta gamma complex. *J. Biol. Chem.* 269:9067-73.

Huang C, Duncan JA, Gilman AG, and Mumby SM (1999). Persistent membrane association of activated and depalmitoylated G protein α subunits. *Proc. Natl. Acad. Sci. USA* 96:412-17.

Huang EM (1989). Agonist-enhanced palmitoylation of platelet proteins. *Biochim. Biophys. Acta* 1011:134-9.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409:860-921.

James G and Olson EN (1989). Identification of a novel fatty acylated protein that partitions between the plasma membrane and cytosol and is deacylated in response to serum and growth factor stimulation. *J. Biol. Chem.* 264:20998-21006.

Jelsema C L and Axelrod J (1987). Stimulation of phospholipase A2 activity in bovine rod outer segments by the beta gamma subunits of transducin and its inhibition by the alpha subunit. *Proc. Natl. Acad. Sci. USA* 84:3623-7.

Jung V, Chen L, Hofmann SL, Wigler M, and Powers S (1995). Mutations in the SHR5 Gene of *Saccharomyces cerevisiae* Suppress Ras Function and Block Membrane Attachment and Palmitoylation of Ras Proteins. *Mol. Cell. Biol.* 15:1333-42.

Karnik SS, Ridge KD, Bhattacharya S, and Khorana HG (1993). Palmitoylation of bovine opsin and its cysteine mutants in COS cells. *Proc. Natl. Acad. Sci. USA* 90:40-4.

Kay RR (2002). Chemotaxis and cell differentiation in *Dictyostelium*. *Curr. Opin. Micro.* 5:575-79.

Kehrl JH (1998). Heterotrimeric G Protein Signaling: Roles in Immune Function and Fine-Tuning by RGS Proteins. *Immunity* 8:1-10.

Kesbeke F, Snaar-Jagalska BE, and Van Haastert PJM (1988). Signal transduction in *Discoideum fgdA* mutants with a defective interaction between surface cAMP receptors and a GTP-binding regulatory protein. *J. Cell Biol.* 107:521-8.

Kessin RH (2001). Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity. Cambridge University Press; Cambridge, England.

Kumagai A, Pupillo M, Gundersen R, Miake-Lye R, Devreotes PN, and Firtel RA (1989). Regulation and function of Gα protein subunits in *Dictyostelium*. *Cell* 57:265-75.

Lee JD and Treisman JE (2001). Sightless has homology to transmembrane acyltransferases and is required to generate active Hedgehog protein. *Curr. Biol.* 11:1147-52.

Li B, Cong F, Tan CP, Wang SX, and Goff SP (2002). Aph2, a protein with a zf-DHHC motif, interacts with c-Abl and has pro-apoptotic activity. *J. Biol. Chem.* 277:28870-6.

Lilly P, Wu LJ, Welker DL, and Devreotes PN (1993). A G-protein β-subunit is essential for *Dictyostelium* development. *Genes Dev.* 7:986-95.

Liu L, Dudler T, and Gelb MH (1996). Purification of a protein palmitoyltransferase that acts on H-Ras protein and on a C-terminal N-Ras peptide. *J. Biol. Chem.* 271:23269-76.

Liu L, Dudler T, and Gelb MH (1999). Additions and Corrections to Purification of a protein palmitoyltransferase that acts on H-Ras protein and on a C-terminal N-Ras peptide. *J. Biol. Chem.* 274:3252.

Lobo S, Greentree WK, Linder ME, and Deschenes RJ (2002). Identification of a Ras palmitoyltransferase in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 277:41268-73.

Maftahi M, Nicaud JM, Levesque H, and Gaillardin C (1995). Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV identifies the RPD3, PAS8 and KRE1 loci, five new open reading frames. *Yeast.* 11:567-72.

Magee AI and Courtneidge SA (1985). Two classes of fatty acid acylated proteins exist in eukaryotic cells. *EMBO J.* 4:1137-44.

Magee AI, Gutierrez L, McKay IA, Marshall CJ, and Hall A (1987). Dynamic fatty acylation of p21N-ras. *EMBO J.* 6:3353-57.

McIlhinney RA, Pelly SJ, Chadwick JK, and Cowley GP (1985). Studies on the attachment of myristic and palmitic acid to cell proteins in human squamous carcinoma cell lines: evidence for two pathways. *EMBO J.* 4:1145-52.

Melendez RF and Bizzozero OA (1996). Palmitoylation of myelin P0 protein is independent of its synthesis and parallels that of phospholipids. *J. Peripher. Nerv. Syst.* 1:34-41.

Milligan G, Parenti M, and Magee AI (1995). The dynamic role of palmitoylation in signal transduction. *Trends Biochem. Sci.* 5:181-7.

Mumby SM (1997). Reversible palmitoylation of signaling proteins. *Curr. Opin. Cell Biol.* 9:148-54.

Mundy DI and Warren G (1992). Mitosis and inhibition of intracellular transport stimulate palmitoylation of a 62-kD protein. *J. Cell Biol.* 116:135-46.

Murphy L, Shore L, Oliver K, Harris D, Barrell BG, Rajandream MA, and Walsh SV (1995). Direct Submission of PSL10 protein to Swiss-Prot.

Oyama T, Miyoshi Y, Koyama K, Nakagawa H, Yamori T, Ito T, Matsuda H, Arakawa H, and Nakamura Y (2000). Isolation of a novel gene on 8p21.3-22 whose expression is reduced significantly in human colorectal cancers with liver metastasis. *Genes Chromosomes Cancer.* 29:9-15.

Parent CA and Devreotes PN (1996). Molecular genetics of signal transduction in *Dictyostelium. Annu. Rev. Biochem.* 65:411-40.

Patterson SI and Skene JH (1999). A shift in protein S-palmitoylation, with persistence of growth-associated substrates, marks a critical period for synaptic plasticity in developing brain. *J. Neurobiol.* 39:423-37.

Putilina T, Wong P, and Gentleman S (1999). The DHHC domain: a new highly conserved cysteine-rich motif. *Mol. Cell. Biochem.* 195:219-26.

Qanbar R and Bouvier M (2003). Role of palmitoylation/depalmitoylation reactions in G-protein-coupled receptor function. *Pharmacol. Ther.* 97:1-33.

Quesnel S and Silvius JR (1994). Cysteine-containing peptide sequences exhibit facile uncatalyzed transacylation and acyl-CoA-dependent acylation at the lipid bilayer interface. *Biochemistry* 33:13340-8.

Raper KB (1935). *Dictyostelium discoideum,* a new species of slime mold from decaying forest leaves. *J. Agr. Res.* 50:135-47.

Resh MD (1994). Myristoylation and palmitoylation of Src family members: the fats of the matter. *Cell* 76:411-3.

Ross NW and Braun PE (1988). Acylation *in vitro* of the myelin proteolipid protein and comparison with acylation *in vivo*: acylation of a cysteine occurs nonenzymatically. *J. Neurosci. Res.* 21:35-44.

Roth AF, Feng Y, Chen L, and Davis NG (2002). The yeast DHHC cysteine-rich domain protein Akr1p is a palmitoyl transferase. *J. Cell Biol.* 159:23-8.

Ruault M, Brun ME, Ventura M, Roizes G, and De Sario A (2002). MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia. *Gene* 284:73-81.

Schmidt CJ, Thomas TC, Levine MA, and Neer EJ (1992). Specificity of G Protein β and γ Subunit Interactions. *J. Biol. Chem.* 267:13807-10.

Schmidt MF, Bracha M, and Schlesinger MJ (1979). Evidence for covalent attachment of fatty acids to Sindbis virus glycoproteins. *Proc. Natl. Acad. Sci. USA* 76:1687-91.

Seehafer JG, Slupsky JR, Tang SC, and Shaw AR (1988). The functional cell surface glycoprotein CD9 is distinguished by being the major fatty acid acylated and a major iodinated cell-surface component of the human platelet. *Biochim. Biophys. Acta* 952:92-100.

Simon MI, Stathmann MP, and Gautam N (1991). Diversity of G Proteins in Signal Transduction. *Science* 252:802-8.

Slomiany A, Liau YH, Takagi A, Laszewicz W, and Slomiany BL (1984). Characterization of mucus glycoprotein fatty acyltransferase from gastric mucosa. *J. Biol. Chem.* 259:13304-8.

Soyombo AA and Hofmann SL (1997). Molecular cloning and expression of palmitoyl-protein thioesterase 2 (PPT2), a homolog of lysosomal palmitoyl-protein thioesterase with a distinct substrate specificity. *J. Biol. Chem.* 272:27456-63.

Staufenbiel M (1988). Fatty acids covalently bound to erythrocyte proteins undergo a differential turnover *in vivo*. *J. Biol. Chem.* 263:13615-22.

Takida S and Wedegaertner PB (2003). Heterotrimer Formation, Together with Isoprenylation, Is Required for Plasma Membrane Targeting of Gbeta gamma. *J. Biol. Chem.* 278:17284-90.

Van Haastert PJM (1995). Transduction of the chemotactic cAMP signal across the plasma membrane of *Dictyostelium* cells. *Experientia* 51:1144-54.

Veit M, Becher A, and Ahnert-Hilger G (2000). Synaptobrevin 2 is palmitoylated in synaptic vesicles prepared from adult, but not from embryonic brain. *Mol. Cell Neurosci.* 15:408-16.

Veit M and Schmidt MF (2001). Enzymatic depalmitoylation of viral glycoproteins with acyl-protein thioesterase 1 *in vitro*. *Virology* 288:89-95.

Vesa J, Hellsten E, Verkruyse LA, Camp LA, Rapola J, Santavuori P, Hofmann SL, and Peltonen L (1995). Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature* 376:584-7.

Wall MA, Coleman DE, Lee E, Iniguez-Lluhi JA, Posner BA, Gilman AG, and Sprang SR (1995). The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell* 83:1047-58.

Watts DJ and Ashworth JM (1970). Growth of myxamoebae of the cellular slime mold *Dictyostelium discoideum* in axenic culture. *Biochem. J.* 119:171-4.

Wedegaertner PB and Bourne HR (1994). Activation and depalmitoylation of Gs alpha. *Cell* 77:1063-70.

Wedegaertner PB, Wilson PT, and Bourne HR (1995). Lipid Modifications of Trimeric G Proteins. *J. Biol. Chem.* 270:503-6.

Wu LJ and Devreotes PN (1991). *Dictyostelium* transiently expresses eight distinct G-protein α-subunits during its developmental program. *Biochem. Ciophys. Res. Commun.* 179:1141-7.

Yamane HK and Fung BK-K (1993). Covalent Modifications of G-Proteins. *Annu. Rev. Pharmacol. Toxicol.* 32:201-41.

Yeh DC, Duncan JA, Yamashita S, and Michel T (1999). Depalmitoylation of endothelial nitric-oxide synthase by acyl-protein thioesterase 1 is potentiated by Ca^{2+}-calmodulin. *J. Biol. Chem.* 274:33148-54.

Yoshimura T, Agrawal D, and Agrawal HC (1987). Cell-free acylation of rat brain myelin proteolipid protein and DM-20. *Biochem. J.* 246:611-7.

Zhu H, Wang H, and Ascoli M (1995). The lutropin/choriogonadotropin receptor is palmitoylated at intracellular cysteine residues. *Mol. Endocrinol.* 9:141-50.

APPENDIX
Predicted Intron Location in ALZ 1-12

Locations of predicted introns are indicated by a bold red carat (^).
Single nucleotide mismatches are indicated by a blue nucleotide (contig/gene).
Non intron-like stretches of nucleotides found in the contig, but not the published gene sequence, are in green.

Predicted ALZ 1 Sequence

ATGTCNCGACCGTCATATGCCTCAGCCACTAAAACATACTTTCATAATAGGCT
TGTTACAGGTCCTGATAGAGCATATTTTATTGTTGCGATGATATTAATGCTAA
TACCAGAGATACCATTTCTTATATTTGT^ATGCCCATTATTTGAAGAATGGATT
ACAGCAGCAATATACCCAGTATCAATTTATTTTTGGATTGCTTCATATATATT
TTTAATTCAAACAGCATATACAGATCCAGGTATTATACCTCGTGGAATTTATA
ATGATGATATTTTCGCACCAGATCACAGACAACCTCTATTTAAAAAAATTACA
GTTAAAGATACAAAACAAGAAATAAAATGGTGTGGTAT^ACCACCAAGAGCG
AATCATTGTGGTATTTGTAATAATTGTGTTGAAAGATTTGATCATCATTG^TCC
ATGGGTTGGAAATTGTATTGGAAGAAGAAATTATCAAACATTTTTATACTTTT
TATATTCATTAGGGTTTTTATGTATTTGGATAATGGGATTTTGTGTAGCACAT
ATATGTATAGAATCAGCACGTTATAGGGACAATCATCCATCTGCATCGAGTG
CAAAAGTATTTCAAGAAGGTATGAATAAGTCCCCCTATATATCGATTATAACT
ATAGTTTATGGGTTAGCAGGTTTAATGTTTGTTGGTTCATTAGGTGGATTTCA
CTTCTTTTTATTACTAACAAATCAAAGTACAAATGAAAAGATCAAAAAAACA
TATAAAAAAGTCTAATCCTTATCGTAAATCAGCATTTGCTAATTTTATTGAAGC
ATTTTGTCCACCACGTTATCCAAGTTTTTATAAATATACTTTAGATCATGAAA
AGGAATTAACAACAATACCAACTCCAAATAATATAAATGGAAATAATAATAA
TAGTATAAACAATAACAATAATAATAATAATAATAACAATAATAATAATAAT
AACAATAATAACAACAATAATAATAATAATAATATAAATAATGGTAATAGCG
GTGGTACAACTAATAATGGTTATACGCCACCAATTTCTCCACCTCAGATGTTA
CAACGTCAATCAAGTACAATAAGATATTCATTAGATAATTTAAGAACAAGTA
GTAATAGTAGTTTGGGTAATTTTAATAATTTGAAATCAAGTAGAGACCTTAAT
TTGTCAACTATAAGTGAAGATAAACCGAAAAATTTAAGTAATAGTAATAATA
ATAATAATACCAATAATAAGAACACTAGTGAAGATAATAATCATAGTAGTGG
AAGTGATTTTGGTGGTGATGAAGAAAATAATGAGGATGATTTTAAGAGTGAT
AACGATAAAGAAATTAATTCATCTTCCCTATCTTTAAATCATGAATTACAAGT
AAATGTTTAA

Predicted ALZ 2 Sequence

ATGACAACAACCTTTTTAGAAAGAGTTGAACATGTATTACAATTTTTATTAAA
ATTAGTAGGACCATTATTTGTAATATTTGCAACATTATTAATTGGTTCAATAG
CAGTAGTTCATTTT^GAATCAGATTATTCAGAAGAGAAAATAATAGAATTTA
AAAGTATTAAAACTATAAAGAGATCAGAGACATATAGATTTTGTATACATTG
TAGATTACCAAAAGAAGAGAGAACTCATCATTGTCAATTATGTGGTACATGT
GTTTTAAAAATGGATCATCATTGTCCATGGGTTAATAATTGTGTTGGAGCAAA
TAATCATCGATACTTTATGTTATTTTTAGTTTATCTTTGGATATCATGTGTTTA
TGTTTGTATATTATCCTATCCACATGTATTCAATAGTGAGAGTGGTTATATAC
CGTTTTCAATGTTAATGTCATTCGTTATCACACTAACGATAGCATTTGCATTG
GGTGGTTTATTAGGTTGGCAAATTTATTTAATACTTTCAAATCAAACTACAAT
CGAATTCCTTCATAATAGAACTCAATCTAAAAAAGCAAAAGCAAGAGGTGAA
ATTTATAAAAATCCTTATGATTTTGGTGTCTTACAAAATTTTAAACAATTCTTT
AAAGCTAATAATAATAATAATAATAATAATAATAATTT^TAATTGGTTTACAT
TTGCATTACCAACATTACAACATCCTGGAAAACAAAATAAGAAAATTGAACC
TTTATTGGAAATT^AAATCAAGATCTCAATTTAATTTTTATGATAATCCAATCC
TTAATATTTTTGCTAATATTTGGAGTAAAATGGAAAAAAAAATCAGAAAAGA
TTGGATGTCATTAATTGGAATAATAAAAAAATAA

Predicted ALZ 3 Sequence

ATGGTAAATAATAATAATAAAAATAATAAAATTAATGACAGAGAAAATGAA
GAAAATGAAAAAAATAAAAAAAAGGATAAAATATATGAAAATAAAATAGGA
ATTAATGAAAATAATAATGAAAATAATAATTACCAAAATGAAAATTTTATTT
ATACAAGCGGAGATAGTAATGATACACAAGAGGGTGATATTTCTGAAATACA
AGAAGAATCTACAAATGAAAATAATAAAAAATTAAAAAAAAGAAAAATAAA
AAAAACCAAAAGTCTTTATTTTTATAAATGTGTTACAGGGCCCGATAAATTAT
TTTTTATAGTTACTCAAATTTTTATATTAGTTCCATCATTATTTTTTGAAATCA
AATT^GGTACCAATAAGTTTAGAAATTTCAAAAAAATTAGAAATTTCAAATT
ATATCATTACTTTAACATTATTGGTTTTTATATTTTATAATTTATATAAATGCG
CATTTACAAATCCAGGTATAATTACAAGAAATAATATTGAAGAAAGTAAAAA
AGATAAAATAAAATTAACAAAATTACAAGCAACTAAAAAAGCTTTAAAGAA
ATTTATAACATCAGGTATGGATGAAATTGCAAATGATGATCCAATTTCATCAT
CATCTGATTTTAGTGATAGCGATGATGATGATCAAGATGAACAAGGTGGTAG
TAGTAGTGCTCGAAATAGTTTTGAAAATAGTGGTGATTTCATCGTAGAAATG
GAACAACCCACAAATAGTAACAACAACAACAGTAATAATAATAATAATAAT
AATAAAAATAGAAATAGAAATAATAACAATAATAATAATAATAATAATAAT
AATAAAAATCAATCAAAATATAAAGAAATTCAAATTGGTGATAGTGATTTTA
AATATAAATGTAAATTTTGTATTACATGTGGATTATATAGAGAACCAAGATCT
TTTCATTGCTCAACTTGTAATAATTGTGTAGAGAATTTTGATCATCATTGTGTT
TGGATTGGAAATTGTATTGGTCGTAGAAATTATAGAGAATTTTTTTATTTTAT
AACAACAACTTTAATTTATGCACTATATCTTTTATCAATGTCAATTGTATTTTT
AAATCAAATTGTAAATACAACTGAATCACCTGCAAATAAAATTAATAATAAT
AATATTAATAGTAATAGTAGCAATCACAATAGTAGTAATGATTTAAATGAAA
AGTTTGAAAAATCAATAAATAATATTTTATATGCTTTAAGAACTACAAGTGG
AGGTTTATG^TATTTTTATAATTATTTTTGGTTTTATAATGTCACTGTTGTTAG
GATTTTTAGTATCATATCATATAAGATTAACTCTTTCAAATAAATCAACAATC
GAGGATTTTAAAAAGATTTTCGAAAACCAAATAAACCCATCCGATAAAGGGT
GGTTGTTCAACTTAAAAATTAAACTTTTTAATTTCAATAATAGTAATAATAAT
ATTAATATTAATAATAATAATACTAATAATAATATTAATTCTTAA

Predicted ALZ 4 Sequence

ATGAACCTATACAACAATAGTAATAGTAGTGGTAGTAGTAATAGTAGTAGCA
GCAGTAACAATAAAACTAATATTGACTATAATGATATAAATAATAATGATAT
CTTAACTCCCAAAGGATATCACACAAGTATATATATAGACGATAATGATTTA
AATAATAATCAAATTATCAACAAAAATAATAACAACAACCATAATCGCAATA
ATAATAATAATAATAATAATACCAATAACCATAATAACCCGAAGAAAATGGT
AATAAATTTAAATCCATCTGGAATATTTATACCTTTAATGAATGATTTTGATG
ATGATGATGATGACGAACAAGAAAGTTTAATAAAAAATAATAAAATTACAAT
ACCAACAATAAATTCACAAATAGCGATACCAATATTAAATAATAACGATAAT
AATAATAATAGTAATAATAAAACTGAACAAACAACAACAACAACAACAATA
ATAAAGAATACAACGATTTTTAATAATAAAACTATTTCATTTGGTAGAATTGG
ATTTAGGTCAATTGTAATATTTTTAATATTAGTACCATATATTTATATATTGAA
TTTTGCAATATTTCCATGGACAGTTAATTATGAAACAGAAAGAAAGGGTAAA
ATTCATAGTTTCATATCTATGGCATTGGTGATTCAAATGTTATGTAATTATTAT
TTATGTTCAACTACTGATCCTGGTAGTTTCAAAGACACCATATCACCATCATA
CTATCTATTACATCCAATCTCTTCAACCGATTCAAATGATCATAAGAAATGGT
GTAACAAATGTAATCATCAAAAACCAGAACGAGCTCATCATTGTAGATATTG
TAATAGATGTGTACTAAGAATGGATCATCATTGTCAATGGTTACAAAATTGTA
TTGGTTTATTTAATCAAAAATATTTTGTACTCTTTTTATTCTATACTTCAATAT
CAATTATTTATTTCTTTACTTTATTAATTAAAAGATCAATTGAATTAGTAACA
AAATATACAATG^GAAAAAACATTACCAAGTTTTGATTTATTACATTTATTTT
TATTAGGAATTTTAATTATTATTTTAATTATAGCTGGTATTTCAATAATGGCAT
TGCTTTGGACTCAAATAGCATTAATTAGTAAAGGTTTAACAACGATTGAACAT
GAAGATAAAAAAAAGAAAATATCAACAACCAAATTATTTAAATTTATATAAAA
AGTATGATAAAGGTTCAATAATTTCAAATTTCTCAATTGTATTTGGAAATTTA
TCTTTTTTATGGTTATTACCAACCATACCAAATAATTTAAAAATAACTTCAAA
AAAAGGTGACATTTTTATAGTTTAA

Predicted ALZ 5 Sequence

ATGATACCAACACAAACACCCCCAAGACAACAAATGATGAATCATAATCATC
AACCACAACAAATACAACAACCACAACAAATACAACCACATCAACCACAAC
AACAACAACCACATCAATTTAGACCATTGGATGAAAATGCAGAAATTCCATT
TTTACAAATTGTAACACCAAATCAAATGATTAAATTACCACCAATGTTTCAAG
AACAATATAAACAATTATTAAATAGATGGTCAAATCAATGTATTAAAAAATC
TGGATTATTATATACTCAAGCAGTACCATATAATCAAGTTTTGGTATTATTAG
AAATTGCTAAAAATGGTGATATTCAAGAATATAGAGATTTCATTCAAGTAAT
TATTGAGTCAAATCAAAATAAAGATGGTGATGAAAATGAAGAATCAATTCAA
CAACAACAACAACAACAACAATTAACATTAAAATCAATAATAAATTTAATAG
ATTTAGAAGGTAATACATCAATTCATTATGCAATTATGAGAAAGAATAAAGC
AATGGTTAGTTATTTAATTGATTGTGGTGCAAATTTAGATATTCAAAATTTAG
AGGAAGGTCATACACCATTACATTGGGCATGTATTAAAGCAGATTCATCATTT
GTGTTACAATTGGTTGAAAATGGTGCAGATATTCATTTGACTGATAAGAGGG
GTTACAATGCATTACTTCATGCTGCACAATATAATGAAGTGAATTCAGTGCGT
TATTTAATTGAGAAAGGATTGGACCCAGTGCAATGCAAAGACATTCAATTAC
ATACTGCAGTGCATTGGACCTCGTTTCAAGGTCATGCAAATATGGCACGTTAT
TTCATAAGTTTAGGCGTTGACCCAAATGCACAAGACATTCAAGGCAGAACGG
CATTCCATTGGGGTT^ATAAAACCATTCGTGATAATGATGGTAAGACAGCCTA
CCAATTGGCAGAATCAAAGGAACATTATGAAATCATCGATTATTTAGACACA
AAGCTTAAAGATGATAAATTATTTGGCGGCAATGAGAGACTCTATCATAGAT
TTTGGACAGTGATGGGTGTGTTGACAGTATTGGTGCCAACCTGGATTCTATGC
TATGTGCCAGTGATATTTGGGTTACCATTGTTGGCTGTCGGTGGCTACTTTTTA
AAGAATTATTTACACTTAAATTATTGGGTACCAGAGAGAAACAATTGGTTATT
ACCCTCGATTTTATATTCTTCAGTATCTATTTGGTATTTGATTTATTTACTTAG
AATTGCACCATTGGTAATGGCAATTAACATTTTCCCAAATTTAATAATTAATG
CAACCTCGTGGTACTTTTTCTACTTTTTCATTCGTTTAACAAAGGAAGATCCA
GGCACAATTTCTAAACATATCTCTAAAGAGAAATCAAATGAAAACTTTATGA
ATGCTCTCTCAAGTGGTCGTCAAATCCCTTTGATTTGTCCAACCTGTTTAATTA
ATCGTCCAATTCGTTCAAAACATTGTCCATCTTGTAAAGGTTGTTTTGCACGT
TTTGATCATCATTGTGTTTGGATTAATAAATGTATTGGTATCAATAACCAAGC
TTTATTCATCACTGTACTCTTCTCATATGTTATCTTGGTAATTTCTGGTTTCATT
GTAACTTGGGATTACTTTAAATCTGATTCAAATGCTCCAATACTATCTGAAAG
TTATATTCAATCATTTTTATTCTACTATACCAATTACCCCTTTATTTTATTATTC
TCAATCTATGGTATTGGTATGGCTATTTGGATTGGTAAATTAGCAATCTCACA
AGTTTTAACAATTTTATTCAATAAAACTACTTATGAACAAATTCAACAAATTA
GAGAATTTGAAAGTAGACAAGGTCATGGTCATAGTCACGGTGGTGATCAACA
ATGTAATCATAGTCATGGTAGTGGTGGTGGTGATCATGGTCATAGCCATGGT
GGTAGTTCTGGTAGTGGTAGTGGTAGTGGTAGTGGTGATCATGGTCATAGTC
ATGGTGGTCAACCTTCACTTCCATCTTCAAGTGGTGGAAGTAAAAAAGATTCT
GATAAATCATTCCATAATCAACAAACTAATTTTGATATGTATCATCGTGGTGT
TGTAAATAATGTTAAAGAATTTTTATTCGATTCTCAAAAATTCTATTTCCAAA
CTGAAAATATTTATGCTGATTCAAGAGTTTAG

Predicted ALZ 6 Sequence (No predicted introns in sequence)

ATGAAATTAAATAGTAATATTAATAATAATATAAATAATAGTAGTAATAGTA
ATAATAACTTTGATGCAAAGAATATAATAGTTGATACAATAACACCACCTGA
CCCATCGGTTGAATTTGAAAGGAAATTAGCGAAATCTATATTTTGTTTAGTTC
ATTTTATAGTTTATTGCGTTATTATATTTAGAAAAGGTACAATTTTAGATCAA
GCATTCAAAGATAAGGATTATTTTTATTTAATTTGGACACATTGTGTTTTCTTT
TTTGCAATTGGAACTTATTTCTTAATTTCAAGTAAAAGACCTGGTTTTGTTTCA
TTATCAAATCAAAATTTAAATAATAATAATAATAATAATGGTAGTAGCAATA
AATTTATATTAGAAGATTCAATGGGTTGTATCCCTCAATTAAATATTAATCCA
ACACCAAATTATAGTAAAATAAGTAATATTAAGAGAAAATTAAAAAATAGTA
GTGGTGATATTACTAAAAATCAAGAAAATGAAGATTTAGTGCCATTGATGGA
AATCTCAAAGAATATTGATGAAGATAGCATAAATGACGATACTATAACTACA
ACCACTACCACTACCACCACCACCAGTACCTCAACAATACCAGAAATATCAA
ATGATGATGATGATAATAATAATGAAAACAATAACGATAATGTAAATAATAG
AAATAATAATAATAGTAATGGTGAAAAAGAAGATAATGATATTGATAAATTA
AAGAATCATTATTTTTGTAAGAAATGTTTAGTTGATATACCATTAAGAACTAA
ACATTGTGTAAAATGTAATAGATGTGTTTTAAAGTATGATCACCATTGCGTAT
TTATAGGTGGTTGTGTTGGTTTAAATAATCATAAGAATTTTCTATTATTTTTAT
TGGCGGAATCATTATTACTACTTTTAGGTTTAAGAATCATTGTCACTGGTTTT
GTTAGAGAAAATTCAATTAAAGAATGGATCTTTTCAAATATTGCAATTATTCC
ACCAACTTTATTAATTTTTGGTGGTTTATGTATGCCTTTTGCTTTATTCTGTTTT
CATAGTTTTTTAATTTTAACAAACCAATCTTCATGGTATGTGCTTATGAATAG
ATTTTTTTTTAAAATTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACC
TAATTTAATTAAATTTAATAATAGGGAATTTAATAAGTATCAAAGAATTACAT
ATTTAAAACCATTTTCTAAAAGAGGTATTAATCCATTTAATAAAGGTCCATGG
AATAATTTAAAAAAGTTTTTAAAGGTGATGAGAATCCTTCCGATTGGATATTA
TTATCAAAATATGAAGTTGACCAAATGA

Predicted ALZ 7 Sequence

ATGATAGATCATAATAATATTATAATTGGTAATTCAATTAATAGTAAAACATT
TTCATTAATTGAATCAGTAAAGAATGGTAAATTAAAAGAATGTATTGGATAT
TTAGAAAAGATTAGATTACAAAATCCGTCATTAGATTTAGGAGAAATAATAA
ATAGTGGTGATGACTGTGGCAATACAGCATTGCATTGGGCATGTTATAAGAA
ATGGTATGATATTGTAAAGTATTTACTATCGATGGGGGCCGACCCGAATATA
GCAAACACTGATGAATTACAAACACCATTTCAATGGGCATGCATCGGTGGTG
ATCTCCACATAGTGAAATACGTACTCAACAATGGCGGCGACCCACACCTCCA
AGACAAGCGTGGCTACAATTCACTAATTCACGCCACACAGTATAATGAAATC
AGCGTAGTTCGTTATCTGTTGGATAAAGGCGGTGTAAATGTAGACTCGCCCG
ATTTCCTACAAAAAACGTCTCTCCACTGGGCTGCCTACCAAGGTCATACCCAA
CTCCTACTCTTTCTAGTCAACAAAGGTGCCGACATCAACGCCTTGGACTCCCT
CGGCCGTTCTCCATTGCATTGGGCTGCCTTCAAAGGTAACTCTGACCCAATTA
AAGCACTCTGTGATTTTGGATCAAAAACTATGGAAAAAGATTCAAACAATCA
ATCTCCATCTGACATCTGTTCTTCTCAAAATCATAATTATCTCGCTCATTTCAT
TAAAACATTTAATTATCATCCATTTAGAAAAGTTGGTCCT^TTATTATATAAT
ATATTTTGGATTATTTTTGCAATTTTATTACAATTATATTTTGGATTTATATTTT
ATCATTTTACATTAATACCAGCATTGATATTATTTGGAGCAAGTTTAACATGT
TGCAAATTATTTATTGAACCAATTACATTGTCAAATTCACCAAATCCATTATT
ACCAACATGGATGATAACTTCATTTACGGTGTCATTCGTATACTATGTTAGAT
ATGTGGTACCAGCATTTCCAAATATAATTCTAACACATACGATCACATTGTTT
GTTTATTCAAGTTACTATTATTGTGCATTTAAATTGTTTTTCAGTGATCCTGGT
ACGGTTTCCTCTTCTACGACATCACAAGATTCAAAGGATTTCATTAATGCCGT
CGAAAAGGAATTGGAAATACCAGAGGTTTGTTCAACTTGTTTAATCAATAAA
CCAATTCGTGCAAAACATTGTAGAACTTGTAAAAGATGTGTAGCAAGATTCG
ATCATCATTGTGCTTGGATTAATAATTGTGTTGGTGTAAATAATAATTTATTA
TTCATAATTTTATTATGTTTATTCAGTTTAGCTTATATTATTAGTGTAACTTTT
AATTTTAAAT^TAATGTCAATTGATGAGAATTCGCCATTATATTCAGAAGGTA
AAATGGAATGGTGGACATATCACTATTCAACTTATAAAGGTTTAATTTTATTT
ACAATTTATAAATCATTCATTATGGCATGGTTAGCAAGATTATTATATGTACA
AATAACTGGTGTAATTAATAATGTCACAATGTTTGAATTAATGAAACCACCA
AAAGGTTCAAAGAAGAAATGTTGTAATCATGCTCCACAAGATCAAAATAATA
ATAATAATAATAATAATACTACTACTAATAATAGTAGTAGTAGCAGTAGTAG
CAATAATAATAGTACAGGTACAAATAATGATAATAATAATAATAATTTA
GGTACAAGTAGTAATGAAAGTGGTCAATCAAATTGTAATATTGATCAACATA
ATGATCATGATGGTCAATCAAGTGAACCATTATTAAATGAAGTTTCAATTCAA
ATTCGTAGTGAAAATGGAGTTGAGGAGGATGATGATGATGAAAATAATAAA
AATCCTTCAAATAAATCATTTAATAAAAACTCTAGAGATTTAGATCCAACTCA
AACTAATTTTAGAGATTCAATTTTAATTCCAAAAAGAACAAATAATAATGAA
AATCCATATGATAAAGGTTCAAGAGAAAATATTAGAGAATTCCTTTATGATA
CTTCAAAATGGTTTAGAACTACAACTTTTAGTAATAAAAACTTTTAA

Predicted ALZ 8 Sequence

ATGATAACAATTTCAGTTATTGGTGTATATATATTTATATTATTTGTTATATGT
AGTCAATGGATTTTAGTTTATGAACCATTAAATTTTTTTAATAATATATGGGG
ACAAACATATATAATAGTATATCATTTATTAGCATTATTATTATTAAGATGTT
ATAGTATGTCAGTATTAACAGATCCAGGTAGTCCACCATCAACATGGTTACC
AGAAGGTAAAACAAAACAAGATTTAACAATATTAATTGACCAATTCAAACAT
TTAAATCAAAGTAATAATAAATCTAAATCAACTACAGTATTGGTTGATGATG
ATTTAAGTACAATAACAATTGTAAATAGTAAAAACAGTAAAAACAACAACAA
CAACAACAACAAAGCTAGATTCTGTTCACAATGTTGTGCTTTTAAACCACCAA
GAACTCATCATTGTAAACAATGTAAAAGATGTATTCTAAAACATGATCATCA
TTGTCCTTGGATTGGTAATTGTACTGGTTTTAGAAATCAAAAGTTTTTCATTCA
ATTTTTATTTTATGTTGTAATATTAACTTCAATTACAATCACAACTTTAACTAT
ATCAGGTTTTTATATTTTAAATGTTAATTTAAATGTAGT^TGCTAAAATTAATA
ATAATAATAATAACAATTCAAATGAAGTTGATTTTTTAGTATCAAGTATAATA
GTAACAATAATGTATATTTTTAATTTCTCTGGTGTATTACCAGTATTGTTAGGT
GTTAGTGGTTTATTCTTCTTTCAAATGGAATTTTTGTTAGGAAATTATACACCC
GTTGAAAGATATGAAAGAAAGAAAGAAGGAAAATATGCAAGAAGGAATGGA
TTAAAATATAAATGGAAATTCGATAAAGGTTGGAAATTTAATTTTAGAGAAG
TTATGGGTGATACTTTAATTCAATGGTTTTTCCCAATAGGTTTTCCAAAAACT
GATGGTACTTATTGGAGAGAGAATGAACTTTATTTACAACAACAAAGATTAT
TAAATAATTCTTCTTCTCCTTCCTCCTCTTCCTCTTCAATAATAGCAACTACAA
ATAATCAACCAATCAATATTAATAATAATAATGATGATAATAATAATAATAA
TATTATTAATAAGAATGATAATGGCAATATAAAATTAGAAACTATACAACAA
CAGCCAATTAATGGCGCGAATATAGGTTATGATGATGATGACGACGATGATG
ATGATAGTGAATCAATTGAAATTAATGATTTTATAAGAAGTGACAGTAATCA
ACCTTTAATACCCTCAAATCAACAATCACAAAAACCATCATTTGGTTTTGTTG
AAAATGTTTAA

Predicted ALZ 9 Sequence

ATGGATGAAAATGAGAGTAAATATCTTTATTTTATTTCATTGGCTTTA^ATAA
TTATACCATCAACTGGGTTTTATGTGACGATTGCTCCATATTTATATGAAAAA
TTTGGAGTTATAATAATTATTTCACATTCATTAGTTTTAATATTAAGTGTTGGA
TCACTTAGTTTAGCGAAATTTACAGATCCAGGTTATATACCAAAACAAACAA
TTAAAGATAATAGTGATTATATTCTTCAACAAAATTTTAAATCTGAACAAAAA
CGATTACAAGAAAAAGAAGAGGAAGAAGAAGAAAAAGAAAGACAAGAACA
ACTTTCAAATAATAATAATAATGATAATAATAATAATAATTTAAATAGAGAA
CAACAAGAAGATGAAATTTTAATTAAAAAATATAATGAAAGACAAAAAAAG
AAACCTAAAAAGAAGGTTATACATTTATATAATGGAAATGTTATAAAATTAG
TTTATTGTA^AAACATGTAATTTTTTTAGACCACCAAGATCATCTCATTGTTCA
ACATGTAATAGATGTGTTTTAGAGTTTGATCATCATTG^TCCATGGGTTGGAA
ATTGTATAGGAAGAAATAATTATAAATATTTTGTATATTTTTTAATATGGACA
GTATTGTTATCAATAGTAACCACTAGTTATTCTTTATTACAATTGATTTCTTTA
TCAAAAGAGAAATATCCAGCATTTATAGATTTGGTTGCACATGCTCCATTTTC
AATTGTAATTGCAATTTATGCATTTCTTTTATTTTGGACATTGGTTGGTTTATG
TTTTTTTCATTTACATTTAATATCACGTGGTATCACAACAAGGGAAGAT^GCT
AAAGGTATTATTAATCCATATTTTTCAGGTGGATTTATTAGTGGATTTTTTAA
ATTGGCTTTTGGACCCAGATTTATAAAATCTTCAATAGATATTTATGAT^GTTT
CAACACAATATCAAAAATCTCAATCTTCCTTAAATGAATTTATAATCGATATT
CAAAATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
ATAATAATAATAATAATAATAATAATAAAAATAATAAATATAGACAATCGCC
AAATGTTTCACCATTTTTAGATAAAAAAGATAGTGTTAATATGAATGATTATG
GTAATAATAACACCAATGTAAATAGTAGCTCAGTTTTTTCGAATTTAAATAAT
AATAATATTAATAATAATAATAATATTAATAATAATAATAATAATAATAATA
ATAATAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATAA
TAATAATAATAATAATTCTACTCAATTTTCTTTATTATATAATGACGATACAA
ATATTGATGAAGGGGACGACAGAACAAGTTTACTCTCATATTATTAA

Predicted ALZ 10 Sequence

TTAAAAGATGTGATGATTATGGAAAATTAGTGAATAGATATTGTGGACCAGT
TTTTGTTGGATTTGCAACATCATTAATAACATTAATTGCAATCACCTATTTCAC
AATTATATTTCCAGCAACATCTGATTTTTCAAATTTATTTTTCATTTTCAATTT
CTTTTGTGACCTTAGTATTTCATTATTCCTTACCTATGGTATTTATTTCAATTAT
ATAAAAGCAATTATAACAAAACCTGGTTATCCAAATTTAAATTCAACTTTAAT
AAAT^AATATAAATAGCAATATAAATAATAGTAATAATAATAAAGATAAAGA
TAAAATATTGATAATTGATAATGTTAAATGGAGTTATTGTAAGAAATGTTCAA
AAGCTAAACCACCTAGATGTCATCATTGTAGTGTTTGCGATAAATGTGTTTTA
AAGATGGATCACCATTGCCCATGGATTGGTGGTTGTGTTGGTTTCTATAATTA
CAGATATTTCTTTTTATTTTTATCATATCTTTGGGTATCAGTTTGTTATGTATTA
GCTCATTCTTTACCTTTATTATTTGGTGGTTATTTA^TATAGTAAAAAATATAC
AGAAATTGATAGATTATTAGTTATTATATCATCAATTGGAAGTTTCATAACAT
TTGTTGCAGTTGGATCATTTGGTGGTTTTCATGCTTATTTAATTGGTAGTGGTC
AAACATCAATTGAAAATTTATATCCTCCAAAAAAGAGACCAAATTATAGTTT
AACTTCAATTAAAGATAATTTTCAAATAGTTTTAGGTAAAGGTGATTATTGGT
TCAGTGGTTTATTACCAATAAATTATACTCCAATTGGTAATGGCTGTGATTTT
AAACTAAATATTTCAAATGAAGATAATAATAAAAATAATGAAAATAATGAAA
ATAATGAAAATAATGAAATTGACAATCATAATAATAATAATAATAATAATAA
TAATAATAATAATAATGAAAAAGAAGATAATATTAATGAAAATGATAATTTA
ATAAGCTATGATACTGATGAATATAATAGGCATAAAAAATAA

Predicted ALZ 11 Sequence

ATGACACATTTAAAAAATAGGGAATCAAAACTAAATCATTTAAATAATAGTA
TTGAAAATGATGAAAAATATCAAATAAAATTTCAAAATAGTTGTAATAATAA
TAATAATAATAATAATAATAATAATAATAATAATAAAAAATTTAATTCC
ATAGAAAAAGTAGATGATAATTATAAAGAAGAGGAAGAAGAATACGATGAT
GATGATGATGGTGATGGTGATTATTCATCAGAAGAATATGATTCAGATAGTG
ATTATTCAAAATCAAATAAAAAATTAAATAAATTTACAATGTATGAAATGTT
AATTAGTGAATTAATACATTTTTCAGCAAAAGGGGATGTAGGAGATTGTGAG
GCAGTGATAGCAAAGATAAAGAGTCAACATAGTAATATCTCATTGAAGAGTT
TATTGAATTCGTTGGATTCAGAGGGTAATACAGCATTACATTGGGCATGTTAT
AGAAAACATTTTCAAGTTGTGAAATATTTAGTATCGATTGGAGCCAATGCAA
ATACGGCCAATTCATCAGAGAAGCAAACTCCATTTCATTGGGCATGTGTGGC
AGGTGACATTTATATCATACATCATTTGGTTCAGGTTGGAAATAGTGATATAT
TCCTTAGGGATCGCAGGGGTCTGAATTCACTGCTACTCTCGACCCACCACGAT
AGGTCATTGGACGTGGCGCGATATCTCCTCCACAAAGGCCTACCCGTGACCA
GTAAGGATGACGAGGGACACACCGCACTCCATTGGGCATCGTTCTCGGGAAA
TTTAAAACTGATCCGACTGTTGATAAATAGGGGTGCCGACATCAACTCAATC
GATAATTTACACAGAACTCCAATACATTGGTCGTCGTTCAAGGGCTACACCG
AATGTACAGTGGCACTCCATGAAGAGGGTGCCAATTTGAATCTAAAAGATGC
AGACAATAAAACTCCCTACGAATTGGCTTTAACTCGTTCAACTGATCCTTGTA
TTGCTTTCCTATGGAAAGCTCAAAAAATTGAGAAAAAGCAATTAACCAAGGG
TAAGTCATCTCTAGAGTTGCAAGATAACGCAAATGTATACAATTACTTTTGGA
CCTTTATGGCGTTACTTGGAAATTTCCTTTTCTTTTATATTCTCTACTCCTTTAA
ATGGTACATTTCAATTCCATCAATAATTATACTTGGTAATTTTTGTCGTATCTA
TCTAACTCATCTTTGGGTTGATTATTGTACAAATCCATTACCTGTAACTTGGTG
GATTGTTGGTTGTTTTATTTGTTATTGGACTTATTTCTTTCAAATTCTATGGT^
ATGTTCCAAATTATATATTTTGCCATACAGTTATATCAATTTTAAGTGCCATTT
TTTTTTATGGATTATGTTGTTTACCATTTTCAAATGCTGGTATTATAAAATCAT
CACCTGATGAAGATTTAAAAGATTTCATTTATAGAATTGAAAATAATCAACC
AATACCTGAAATTTGTCCAACTTGTGAT^ATTCATAAACCATTAAGATCAAAA
CATTGTAAATTTTGTAAACAATGTGTTGCTAGATTCGATCATCATTGTATTTG
GATTAATAATTGTGTTGGTACTGCAAATCATAGATTATTTTTATTAGTTTTGTC
ATTGTATAGTGCAATAGCAATTCCAATTTATTATGTTGCTTTTAAAT^TTTTAC
AATTAGATTTAAATGCACCATCATTTGATGATGGATATAAACAAGCATTTGA
ATATTATTATGATACACATAGAATGGTTTCAATATTTTTAGTTTATGGATTATT
AGCATGGATTTGGATATTAAAGTTGTTGAGCGCTCAAATTTTAGGTATCATTT
TCAATTGTACATTAAATGAATTATTAAACTTGGCAAGATATTCTTACCTAAGA
AAAGATGGTAAATGGAATGTATTTCATCGTGGTTTCTTTTCAAATATTACAGA
GTTCTTTATTGGTAGTAAAAAATGGTTTACATTTTTTAATTTTCCAATCGACTC
AAACTTAATTAATAATAATAATAATAATAATATTAATAATATTAATATTAATA
GTAACACTATAAATGTAAACACCAAAATCACTACAAATACTTCCAAAACTAT
TAAAGTTAATTAA

Predicted ALZ 12 Sequence

ATGGATATATTTTTAGTATTAATTGGTTGTTATGTTGTTTTTGCTGTAACGATT
TTATATACATTATTGTTAGGTCAATCAGAATTTCATAGAGATGGTTGTGTAGG
ATCTTTATATATATTTTTAACTTCTGGTTTACAAGATTGGTGTGT^TGGATTTT
TTGCAAGATGTTGTCCTAAAAAATTAAAAAATGGATCAGCTTCATGTTATAAT
TATTTTATGTATAAACCAAATAGAATATTACAAGGATTCTATTTAACTTTAGT
TTTATCAGGATTTTATTTCTTTTATTTTGATTGTTTCCCATATATTGGTGGACC
ATACATTTCATCAAATCATAAATATGGAGCATTTTTTGCAATCTCTTTCACTTT
ATTCACATTTGTTTTATCATCAAATTCAACACCAGGTTATATAAATGATTCAA
ATTATAAATTATTTAAAAATTCTTATCCATATGATAGATATCTTTATATTAAA
AAAAATTGTGAATCTTGTAATTTTATAAA^ACCAGCAAGATCCAAACATTGTA
GAATTTGTGATAGATGTGTTGGTAGATTTGATCATCATTG^TCCATGGATTAA
TAATTGTGTTGGAGAGAATAATTTAAGATATTTCTTATTATTTGTATTTTCAAC
ATCAATGTTATGTATGTATGGTGCTTATTTATGTGGATTTTCAATGTATAGTTT
TATGAAGATAAATGATGTTAAAAATTTAGGATATACAAAAGATGGTGTATGG
ACACCAATTCCAACAGCGATTTTATTGAAATATATAGCATTTGAATCACGTTC
AATCTTACCATTAGGTGCATTTTGTTTTGTAATTTCATTATTTTTATTTTATTTC
TTCTTTTATCATATTTGGTTAATTTCAAAGAATACAACTACAAATGAATCTTA
TAAATGGCAAGATATTAAAGATCAAATTAAAATTGAAAGATTAAAACAAAA
AATTATTGATCATGAAAAACTTGAAAAAGATAACGATAATAACAATAAAAAA
GATAACAATAAAAAAGATAATAATAAAAAAAATAATAATAAAAAAGATAAT
AATAAAAAAGATAATAAAAAAGAAAATAATAATGATAGTATTGTAGAAAAA
GTTAATGAAATTATAAAGAATGAAATTGGTGATTCAGATGAAGAAGGTACAA
ATTTAAATAAATTAAAGAAAAGAAATAAAAGACAACAATTATCGTCAAAGG
AATCAAAAGATAAAGATGAATTCTATACAACCTTACCATTACCAAAAACTTT
TAAAGAATTAAAAAATATTTATAATAAAGGTTTAATATC/TAA

BIOGRAPHY OF THE AUTHOR

Brent Elliott Wells was born in Long Beach, California on September 16, 1970. He was raised in Centerville, Utah and graduated from Viewmont High School in 1988. He attended the University of Utah and graduated in 1994 with a Bachelor's degree in Biology. After graduation, he worked as a Medical Technologist for Gull Laboratories and Laboratory Corporation of America, both in Salt Lake City, Utah, until entering the Biochemistry graduate program at the University of Maine in the fall of 1999.

As a result of receiving his degree, Brent has joined the Pediatrics Department of The Medical College of Wisconsin as a Research Technologist, to begin his career in the field of genetic research. Brent is a candidate for the Master of Science degree in Biochemistry from The University of Maine in December, 2003.