Molecular Cloning and Functional Characterization of a Truncated Epsin N-terminus Homology (ENTH) Protein from *Arabidopsis thaliana*

T Moichwanetse

0000-0001-8875-076X

BSc.HONS Microbiology

Dissertation submitted in fulfilment of the requirements for the degree *Masters* in *Biology* at the Mafikeng Campus of the North West University

Supervisors : Prof. O Ruzvidzo

Graduation October 2017

Student number :23293470

http//www.nwu.ac.za/

It all starts here ™

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

Table of Contents

DECLARATION

I, Tsholofelo Moichwanetse declare that the mini-dissertation submitted to the Department of Biological Sciences at the North-West University, Mafikeng Campus, for the Masters in Biotechnology degree has never been submitted at this university or at any other institution elsewhere. This is my own work and all the sources used or quoted here have been properly indicated and sincerely acknowledged.

Student: Tsholofelo Moichwanetse

Signature: _____ Date: _____

Supervisor: Prof. O Ruzvidzo

Signature: _____ Date: _____

DEDICATION

I dedicate this work to my family and all of my friends.

ACKNOWLEDGEMENTS

I owe my gratitude to all those people who have made this dissertation possible including my family, for their constant love, concern, support and strength; the plant biotechnology research group; my friends who have helped me stay sane through this challenging journey.

I would like to express my sincere gratitude to my main supervisor Prof. O Ruzvidzo and co-supervisor Dr. TD Kawadza for their continuous support, patience, motivation, enthusiasm, and immense knowledge.

I am also indebted to the Biological Sciences Department of the North-West University Mafikeng Campus and the National Research Foundation for their financial assistance. Above all, I thank God mostly for giving me such an opportunity.

DEFINITIONS OF TERMS

Adenosine triphosphate: is a nucleoside triphosphate used in all cells as a coenzyme (ATP) to transports chemical energy for metabolism.

Adenylate cyclases: are enzymes that catalyze the formation of 3′,5′-cyclic AMP (cAMP) and pyrophosphate from adenosine triphosphate.

Cloning: is a process whereby copies of DNA fragments are obtained by allowing an inserted foreign DNA fragment to be replicated by a plasmid in a bacterial system.

Cyclic 3',5'-adenosine monophosphate: is a second messenger molecule derived from ATP used for most intracellular biological signal transduction systems in many different organisms.

Cyclic nucleotide: is a single phosphate nucleotide molecule with a cyclic bond arrangement between the sugar and phosphate group.

***Escherichia coli* BL21 pLysS cells:** are bacterial cell strains containing a T7 RNA polymerase gene that is controlled by the *lac* UV5 promoter in its chromosomal DNA1 and the T7 lysozyme gene in the pLysS plasmid.

G proteins: are a family of proteins involved in the transmission of signals from a variety of different stimuli outside a cell into the internal environment of the cell.

Phosphorylation: is the addition of a phosphate (PO_4^{3-}) group to a protein or other organic molecules, turning them on or off and thereby altering their functions and/or activities.

Primers: are short synthetic nucleic acid sequences capable of forming base pairs with a complementary template RNA or DNA strand and facilitating its specific amplification.

Protein kinase: is an enzyme that modifies other proteins by chemically adding a phosphate group to them (phosphorylation).

Signal transduction: is a cell communication system that occurs when an extracellular signalling molecule activates a cell surface receptor.

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE): is a technique used in molecular biology to separate different protein molecules according to their sizes and migration capacities in a polyacrylamide gel system subjected to a very strong electrical field.

LIST OF ABBREVIATIONS

AC	**:**	Adenylate cyclase
ATP	**:**	Adenosine triphosphate
BLAST	**:**	Basic Local Alignment Search Tool
cAMP	**:**	3′-5′-cyclic adenosine monophosphate
cGMP	**:**	3′-5′-cyclic guanosine monophosphate
ENTH	**:**	Espin N-terminal homology
GC	**:**	Guanylate cyclase
GTP	**:**	Guanosine triphosphate
IPTG	**:**	Isopropyl-β-D-thiogalactopyranoside
LB	**:**	Luria-Bertani
MS	**:**	Murashige and Skoog
OD	**:**	Optical density
PCR	**:**	Polymerase chain reaction
Rpm	**:**	Revolutions per minute
RT-PCR	**:**	Reverse transcriptase-polymerase chain reaction
SDS-PAGE	**:**	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TAIR	**:**	The Arabidopsis Information Resource
TBE	**:**	Tris-borate EDTA
TFB	**:**	Transformation buffer
UV	**:**	Ultraviolet
YT	**:**	Yeast-tryptone

LIST OF FIGURES

LIST TABLES

ABSTRACT

Adenylate cyclases (ACs) are enzymes that are known to produce 3′,5′-cyclic adenosine monophosphate (cAMP) from 5′-adenosine triphosphate (ATP) as a result of some associated extracellular stimulations. However, the question of whether or not cAMP does exist in plants has been an issue of debate for a while, mainly due to the less efficient methods employed to isolate this molecule and also because of its very low levels in plants. In contrast to plants, animals and lower eukaryotes ACs and their product cAMP have been firmly established as important signalling molecules with critical roles in cellular signal transduction pathways. Therefore, this study was carried out to determine if higher plants, particularly the *Arabidopsis thaliana*, have any other additional and functional ACs besides the few and currently known ones and if so, to test for their functional activities. Among the various proteins suspected to harbour an AC catalytic centre in *A. thaliana* is the epsin N-terminus homology (ENTH), whose truncated form harbouring the AC catalytic center (ENTH-AC) was investigated in this study. In this regard, some *A. thaliana* plants were initially generated followed by extraction of their total RNA. This total RNA together with some manually designed sequence-specific primers were then used to isolate and amplify the targeted ENTH-AC gene fragment via a specialized RT-PCR system. The isolated ENTH-AC gene fragment was then cloned into a pTrcHis2-TOPO expression vector to yield a pTrcHis2-TOPO:ENTH-AC fusion construct. The produced construct was then used to transform some chemically competent *cya*A SP850 mutant *E. coli* cells, through which its *in vivo* functional AC activity was then determined and confirmed. After this, more work was then undertaken bioinformatically on the ENTH protein, whereby a wide range of web-based softwares and computer-based programs were used to determine the exact functional role(s) of this protein in Arabidopsis and possibly, other related higher plants. All in all, findings from this study unequivocally indicated that the ENTH protein is a *bona fide* functional higher plant AC with a central role in cAMP-mediated fungal stress responses.

CHAPTER ONE

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Global warming refers to the overheating of the earth's surface and is one of the major leading problems faced by mankind world-wide. It is actually predicted to have a very negative effect on plant growth since high temperatures negatively affect plant development and thus crop productivity. This would then result in significant yield reductions, which create serious imbalances between the 'demand and supply' systems of the world causing global food insecurity (Bita, 2013). Apparently, as food supply continues to decrease, the population size continues to rapidly increase annually despite the nearly constant arable land size. Thus the increasing incidences of environmental stress are making crop farming somewhat unstable even though people solely depend on crops as their main source of food for survival (Stuhmer *et al.*, 1989). Thus, since the world population will always expand, crop productivity needs to be improved so as to sustain food security. To that end, it has already been suggested that world food production will have to be increased by 70% in order to meet the demand of an expected population size of 9 billion in 2050 (Bita, 2013).

There is no doubt that given the changes in plant growth environment associated with climate change, improving plant tolerance to stress will be a major priority. Cost-effective and efficient biotechnological strategies are required to develop crop plants with better stress detection and tolerance mechanisms (Abdallah, 2014). Unlike humans, plants are sessile organisms and, therefore, cannot move away when exposed to stress factors. However, plants have actually evolved an effective mechanism to escape or reduce damage from stress by using multiple signalling pathways that mediate their responses. Moreover, different forms of stress may activate or make use of multiple components such as genes, proteins and other signalling molecules, to mediate specific needed cellular responses.

Of late, biotechnologists have targeted protein molecules that systemically affect homeostasis in plants to possibly deliver effective plant stress response solutions (Braam, 1997). Therefore, an identification and characterization of such plant molecules will possibly contribute towards a better understanding of how plants exactly respond and adapt to stressful situations. One such potential molecule is the epsin N-terminal homology (ENTH) protein from *Arabidopsis thaliana* that putatively functions in regulating transcription, phospholipid, clathrin and phosphatidylinositol binding, and is also involved in endocytosis (TAIR). Additionally, the

ENTH gene has recently been bioinformatically annotated to harbour an adenylate cyclase (AC) catalytic centre, which in other organisms like animals and bacteria, plays a crucial role of generating the second messenger cyclic adenosine monophosphate (cAMP) that is centrally involved in cell signal transduction and responses to the various environmental stress factors. Apparently and even though the ENTH protein has recently been shown to harbour a putative AC catalytic centre, no studies to date have yet been undertaken to directly or indirectly establish this putative protein as a possible functional and/or *bona fide* plant AC (Gehring, 2010). Therefore, this study was designed to practically check if the ENTH is a functional plant AC and also to further elucidate its probable functions in plants, particularly with respect to cell stress responses and adaptation mechanisms.

1.2 Literature Review

1.2.1 Adenylate Cyclases and the Cyclic Adenosine 3′,5′-cyclic Monophosphate

Adenylate cyclases (also referred to as adenylyl or adenyl cyclases) (ACs) are enzymes that catalyze the conversion of ATP to cyclic AMP (adenosine 3′,5′-cyclic monophosphate) and a pyrophosphate (Gehring, 2010). For many decades, cyclic AMPs have been extensively studied as second messengers in animals, lower eukaryotes and bacteria, where they perform critical functions such as signalling, communication, and mediation of different physiological and biochemical processes including the activity of kinases (Ichikawa *et al.*, 1997). In bacteria, cAMP is involved in the positive regulation of the *lac* operon, whereby in low-glucose environments, it increases its levels and then binds to the allosteric site of the cAMP receptor protein (CRP), which is a transcription activator protein. Once the CRP is in an active configuration, it binds to a *cis*-element upstream of the *lac* promoter, activating transcription. In high-glucose environments, cAMP concentration decreases and the CRP detaches itself from the *lac* operon promoter, hence inhibiting transcription (Meiklejohn & Gralla, 1985).

In animals, cAMP was originally discovered as an intracellular mediator of many peptides and hormones, whereby an external stimulus will switch on the G protein-coupled receptors (GPCRs), activating ACs, which in turn, trigger a signal system that modulates cellular processes in response to the external stimulus. Since then, cAMP has been implicated in many signalling systems that intercede cellular responses to the imbalances of the external biosphere (Robison & Sutherland, 1971). Cyclic AMP was also described as a regulator of glycogen breakdown in the livers of animals (Assmann, 1995) and also for many features of the

development of the slime mold, *Dictyostelium discoideum*, a unicellular that then later develops as a multicellular organism (McMains *et al.*, 2008). An illumination of the structure and molecular conformation of cAMP (Fig. 1), which was first established in animals, helped in establishing the chemical properties and biological functions of this important cellular signalling secondary messenger molecule (Sutherland & Rall, 1960).

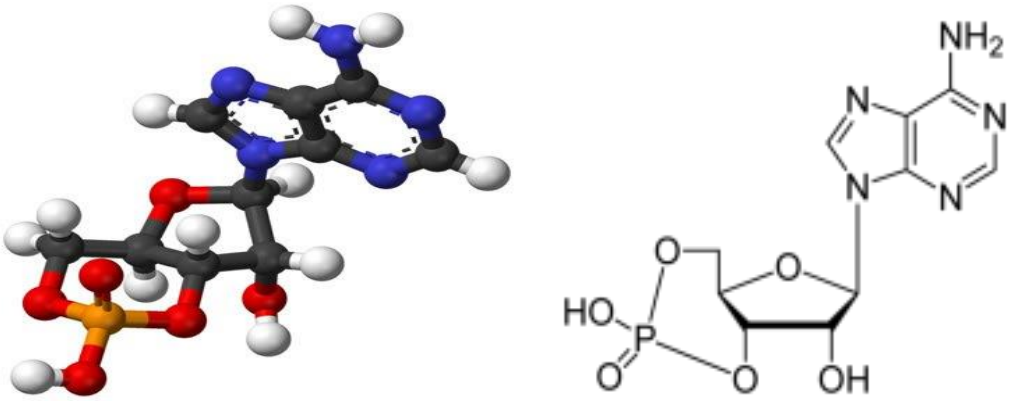

Figure 1.1: The chemical and three dimensional structures of cAMP (Sutherland & Rall, 1960).

1.2.2 The Long Term Dispute of cAMP in Higher Plants

Despite all this compelling evidence, the history of cAMP in higher plants has not been free of controversy and specifically, the question of whether or not it exists and/or function in plants has been fiercely debated for almost three decades starting in the late 1960s when the molecule was first accepted as a second messenger in animals, lower organisms and prokaryotes (Lemtiri-Chlieh *et al.*, 1995). The occurrence of cAMP in tissues of higher plants has been doubted by a number of reviewers due to the low yields of this product in plants as compared to animals (Gehring, 2010). Many experimental data and varying claims from several laboratories have also been published, suggesting some comparable roles for the molecule in higher plants. However, these were still highly criticized because some were lacking specificity in the effects elicited by cyclic nucleotides while others had uncertainties about cyclic nucleotide-generating enzymes (ACs), and others giving ambiguous identifications of the endogenous putative cyclic nucleotides, mainly due to the inadequate clarity of the then used chromatographic techniques (Newton & Smith, 2004).

By that time, many people were disputing the existence of cAMPs in higher plants to an extent that reviewers were even producing titles such as "*Evidence against the occurrence of adenosine 3',5'-cyclic monophosphate in higher plants*" (Amrhein, 1974). The main reason for such disputes regarding the existence of cAMP in higher plants was due to the low sensitivity of methods employed to detect it. Indeed, Pollar & Venere, (1970) reported that a particular radiolabeled product derived from barley seedlings could not be chromatographically separated from cAMP due to the inadequate resolving power of this system, which could not distinguish between cAMP and the RNA catabolic intermediate - 2′,3′-cAMP. Other publications arguing the presence of cAMP in higher plants included declarations such as "*Cyclic AMP does not mediate the action of gibberellic acid*" (Keates, 1973). However, and since the mid-1980s, several researchers embarked on a vicious journey attempting to demonstrate the occurrence of this molecule in tissues of higher plants (Assmann, 1995; Ludidi & Gehring, 2003; Wachten, 2006; Gehring 2010; Ruzvidzo *et al*., 2013), which then firmly confirmed the existence of cAMPs in higher plants. The potential roles of cAMP in higher plants include the regulation of ion channels and ion transport, cAMP-dependent signal transduction pathways, stimulation of the protein kinase activity, and the induction of stress responses and plant defense mechanisms. Overally, this shows that despite the low levels of cAMP in plants as compared to animals, the perception that plants have a functional cAMP still remains strongly alive (Lemtiri-Chlieh, 1995).

1.2.3 Search for Adenylate Cyclases in Higher Plants

Given the growing realization of the existence of ACs and cAMP in most organisms, it was not surprising to see plant scientists investigating whether such signalling systems were universal and also similarly operating in plants (Ruzvidzo *et al*., 2013), particularly in *Arabidopsis thaliana*, which is the subject of this study. Considering that cyclic nucleotides have important and diverse roles in plant signalling systems, it is unlikely that a single AC can account for all the currently reported and/or known cAMP-dependent processes in higher plants. In line with this hypothesis, is the fact that a number of Arabidopsis molecules with different domain organizations and experimentally confirmed ACs have recently been reported (Moutinho *et al*., 2001; Ruzvidzo *et al*., 2013; Ito *et al*., 2014; Swiezawska *et al*., 2014; Al-Younis *et al*., 2015). Briefly, some functionally tested guanylate cyclases (GCs) (the only nucleotide cyclases closely related to ACs) were previously identified in *Arabidopsis thaliana* using a 14 amino acid long search term deduced from an alignment of conserved and functionally assigned amino

acids in the catalytic center of annotated GCs from other organisms (Fig 1.2A) (Ludidi and Gehring, 2003). Based on this same concept, Gehring (2010) then made use of the same search term but now modified for specificity to ATP binding than GTP binding and with an additional C-terminal metal binding stretch (Fig 1.2B & C). From this approach, a sizable number of putative and novel AC molecules were then discovered (Table 1.1).

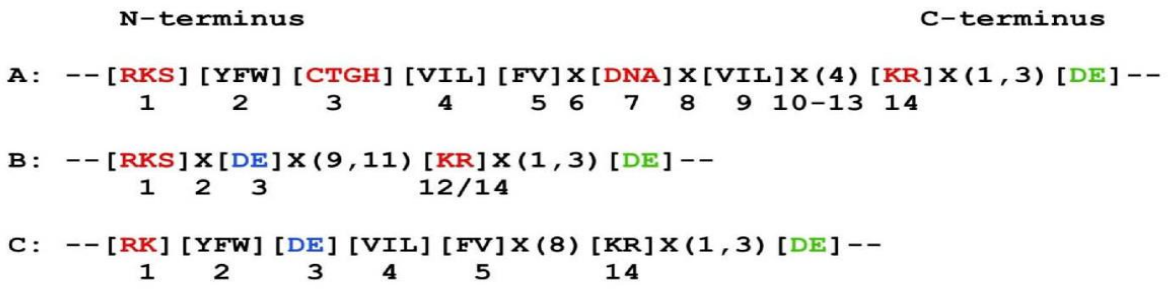

Figure 1.2: (A) Centre motif of practically tested and confirmed GCs in plants. Residue in position 1 (red), is for hydrogen bonding with guanine, residue in position 3 (red), confers substrate (GTP) specificity and residue in position 14 (red), stabilizes the transition (GTP/cGMP). The C-terminal metal binding segment is represented in green. In the derived motifs (B and C) specifically modified for ACs, position 3 (blue) has been substituted to [DE], to allow for ATP binding (Gehring, 2010).

Table 1.1: *Arabidopsis thaliana* proteins harbouring the putative AC search or catalytic motif: [RK][YFW][DE][VIL][FV]X(8)[KR]X(1,3)[DE] (Gehring, 2010).

ATG No.	Sequence	Annotation
At1g25240*	KWEIFEDDFCFTCKDIK	Epsin N-terminal homology
At1g62590	KFDVVISLGEKMQR--LE	Pentatricopeptide (PPR) prot.
At1g68110	KWEIFEDDYRCFDR--KD	Epsin N-terminal homology
At2g34780	KFEIVRARNEELKK-EME	Maternal effect embryo arrest
At3g02930	KFEVVEAGIEAVQR--KE	Chloroplast protein
At3g04220	KYDVFPSFRGEDVR--KD	TIR-NBS-LRR class
At3g18035	KFDIFQEKVKEIVKVLKD	Linker histone-like prot. - HNO4
At3g28223	KWEIVSEISPACIKSGLD	F-box protein
At4g39756	KWDVVASSFMIERK--CE	F-box protein

ATG represents the assigned *A. thaliana* gene bank numbers, followed by a nucleotide sequence suspected to be the annotated AC catalytic center, and the names to which each gene was bioinformatically inferred (annotations). ATG numbers follows a format for chromosomal based nomenclature, where AT stand for *Arabidopsis thaliana*, then the following number and letter representing one of the five *A. thaliana* chromosomes and then the last five digit code representing the location of a gene on the chromosome, and numbered from top/north to bottom/south. *The gene that was studied in this work.

These bioinformatically annotated Arabidopsis genes suspected to harbour the functional AC domain then left plant scientists with an open task to experimentally confirm the existence of ACs in higher plants as well as to further analyse their probable functional roles in plants. To this day and based on this phenomenon, a total of five functional ACs have since been identified in higher plants, which are the *Zea mays* pollen protein (Moutinho *et al*., 2001), the *Arabidopsis thaliana* pentatricopeptide repeat protein (Ruzvidzo *et al*., 2013), the *Nicotiana benthamiana* adenylyl cyclase protein (Ito *et al*., 2014), the *Hippeastrum hybridum* adenylyl cyclase protein (Swiezawska *et al*., 2014) and the AtKUP7 protein from *Arabidopsis thaliana* (Al-Younis *et al*., 2015).

1.2.4 The Epsin N-terminal Homology Protein

Amongst the bioinformatically identified candidates in *Arabidopsis thaliana* is the At1g25240 gene, which encodes for an epsin N-terminal homology (ENTH) protein (Gehring, 2010). Physiologically, ENTH is an evolutionarily conserved module of approximately 150 amino

acids found at the N-terminal of various proteins of different organism such as yeasts, plants, nematodes, frogs and mammals, even though in plants its exact role still remains elusive (Hussian, 2003; Gehring 2010) Ideally, the ENTH protein is also further described as an ENTH/ANTH/VHS superfamily protein that is located in the clathrin coat and annotated to function in phospholipid binding, clathrin binding and phosphatidylinositol binding, all involved in clathrin coat assembly and endocytosis (www.arabidopsis.org). Structural analyses and ligand-binding studies have also indicated that a set of proteins previously designated as harbouring the ENTH domain possess a highly similar, yet unique module referred to as an AP180 N-terminal homology (ANTH) domain (Legendre-Guillemiv, 2004).

Generally the ENTH domain is composed of multiple α-helices folded upon one another to form a compact globular structure of different lengths, which then interacts with lipids and proteins (Koshiba *et al*., 2001). The C-termini of the ENTH domain have the peptide motifs, which participate in clathrin-mediated endocytosis. Endocytosis is an essential mechanism for internalizing extracellular material and controlling the composition of the plasma membrane. It also controls cellular homeostasis, including the down-regulation of signalling receptors, which may be modulated to carry out the various effects of signalling molecules such as ACs, and the regulation of nutrients uptake as well as signal transduction (Aguilar *et al*., 2006). Therefore, understanding the relationship between ENTHs and the unique idea of ACs (signal transduction and maintenance of cell homeostasis), will potentially bring about a better understanding of the probable physiological and biological roles of ENTH proteins in higher plants.

1.2.5 Cell Signalling Pathway

Naturally, cells use a large number of signalling pathways to regulate their activities, however, there are several mechanisms involved in transmitting information in the cell. Most cell surface receptors stimulate intracellular target enzymes, which may either be directly linked or indirectly coupled to receptors by G proteins (Kholodenko, 2006). These intracellular enzymes serve as downstream signalling elements that propagate and amplify the signal initiated by ligand binding. A chain of reactions transmits signals from the cell surface to a variety of intracellular targets such as transcription factors that function to regulate gene expression. Thus intracellular signalling pathways do connect the cell surface to the nucleus, leading to changes in gene expression in response to extracellular stimuli. The signals are received by

the receptors, which are immobilized onto the cell membrane that function as molecular antennae (Kholodenko, 2006). These receptors then transfer the information across the membrane using signal transducers (triggers of signal cascades) and secondary messengers (amplifiers of signal, e.g., cAMP) that engage the specific signalling pathway - mediating specific cellular responses (Cooper, 2003).

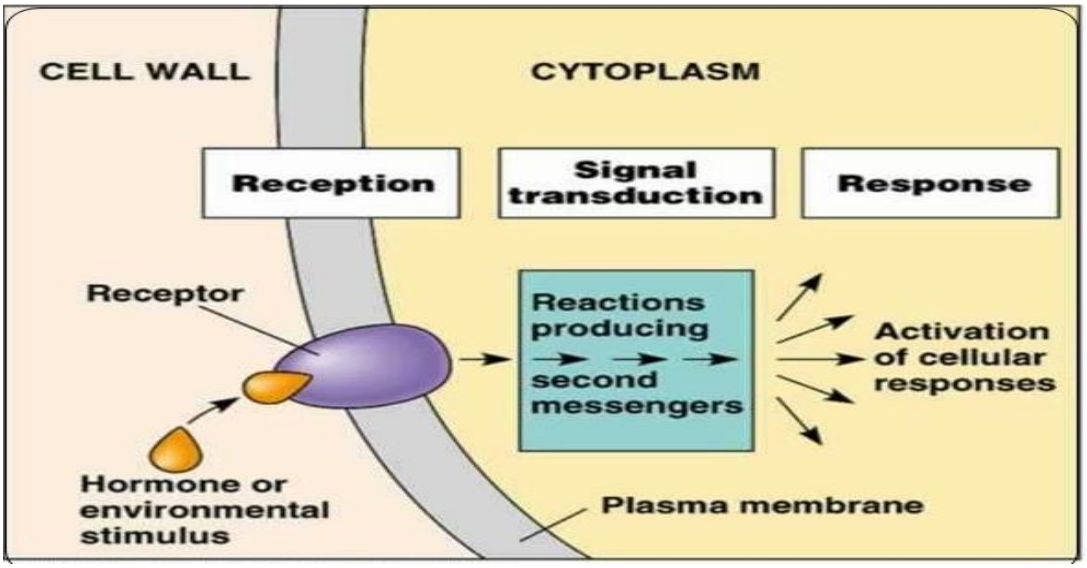

Figure 1.3: The basic principle of cell signalling pathway. An external stimulus binds to the receptor, (reception) then this result in the production of secondary messengers that transduce certain signals (signal transduction), which in turn mediate specific responses (Cooper, 2003).

In plant biotechnology, signalling molecules, which plant biochemists and plant physiologists mainly target, are those affecting homeostasis and cell communication, of which the ENTH protein from *A. thaliana* is one such molecule. In this regard, the ENTH protein was herein comprehensively studied so as to elucidate its probable functional roles in higher plants and, specifically, in important cellular processes such as growth, development and response to various environmental stress factors. Therefore, identifying and characterizing such critical components involved in stress response and adaptation mechanisms could immensely contribute towards a better understanding of how plants exactly function in their natural environments.

1.3 Problem Statement

Naturally, the biological molecules termed epsins play very important roles in cell signalling and also in the different steps of protein trafficking of various organisms. More so, many epsin

homologs have been identified in various plant cells through genome sequence analysis yet their roles have not been fully elucidated (Song, 2006). In 2010, Gehring bioinformatically annotated nine candidate proteins from the *Arabidopsis thaliana* plant as possible AC molecules amongst which the ENTH was one of them. However, no studies to date have yet been undertaken to experimentally demonstrate whether this ENTH is indeed a functional AC, and if so, whether it has any biological role in stress response and adaptation mechanisms. Hence this study was set to experimentally check and determine these specific and outlined aspects.

1.4 Research Aim

The main question of this study was to determine whether higher plants have any other additional and functional ACs besides the few and currently known ones and if so, to test for their functional activities. This question was, therefore, answered by experimentally testing for AC activity in the ENTH protein, which was recently identified in *Arabidopsis thaliana* as a probable AC molecule.

1.5 Objectives

The following specific objectives were set in order to attempt and address the proposed aim and the main research question of this study:

1. To isolate the annotated ENTH-AC gene fragment from *Arabidopsis thaliana*.
2. To clone the annotated ENTH-AC gene fragment from *Arabidopsis thaliana* into a stable and viable heterologous prokaryotic expression system.
3. To determine the enzymatic AC activity of the recombinant ENTH-AC protein.
4. To bioinformatically determine the functional roles of the recombinant ENTH-AC in *Arabidopsis thaliana* and other plants.

1.6 Significance of the Research Study

The undertaking of this project was significant in that:

1. If successful, more literature was going be added to the study of ACs and thus providing a clearer and better understanding of the mechanisms by which ACs function in plants

2. The study was also going to lay down a very good groundwork for further work on this important protein, leading to a better future understanding of its direct and/or indirect roles in plant stress responses and adaptation mechanisms.
3. More so, the study would add yet another functional higher plant AC to the group of the currently known ACs.

CHAPTER TWO

RESEARCH METHODOLOGIES

2.1 Generation of *Arabidopsis thaliana* Plants and their Growth Conditions

2.1.1 Seed Surface Sterilization and Stratification`

The *Arabidopsis thaliana*, ecotype Colombia (Col-O) plant was chosen as it grows fast and vigorously. About 0.1 g of the *A. thaliana* seeds were placed into a micro centrifuge tube. Using 70% of absolute ethanol, the seeds were surface sterilized by thoroughly vortexing them in 500 µl of the solution and subsequently removing the washing solution. The seeds were further washed with 500 µl of the seed sterilization buffer (0.1% (v/v) sodium dodecyl and 5% (v/v) sodium hypochlorite) before subjecting them to several washes (5 times) of the filter-sterile deionized water (2 500 µl). After these thorough washes, the seeds were then submerged into 500 µl of sterile water for 3 days and under cold conditions (at 4^{o}C) for 3 days (in order to stratify and preparing them for a synchronized or unified germination).

2.1.2 Plant Growth and Maintenance Conditions

Following stratification, the *Arabidopsis thaliana* seeds were germinated on Murashige and Skoog media (0.3% (w/v) sucrose, 0.43% Murashige nd Skoog basal salt mixture and 0.8% (w/v) tissue culture grade agar) supplemented with 1 ml of Gamborgs vitamins to nourish the seeds at pH 5.7. The seeds were germinated in petri dishes and allowed to grow for two weeks in a GC-300TL growth chamber (Lab Companion, Pretoria, RSA), under greenhouse conditions adjusted to 23/16 day/night for periods of 8/16 hours night/day at 10 000 light lux. The germinated seedlings were then transplanted into potting soil that contained 1:1:1 (w/w) vermiculite, humus, potting mix soil, respectively, which was watered with sterile distilled water. The seedlings were then covered with plastic cling wrap to retain moisture, and placed in the growth chamber to grow for a further 2-4 weeks. After two weeks the matured plant was used to extract total RNA.

2.2 Isolation and Cloning of the ENTH Gene Fragment into a Stable Heterologous Prokaryotic Expression System

2.2.1 Designing and Acquisition of Sequence-specific Primers

The genomic sequence of the At1g25240 gene encoding the putative ENTH protein was retrieved from The Arabidopsis Information Resource (TAIR) (http://www.arabidopsis.org). Two sequence-specific primers targeting the AC containing region of this protein (ENTH-AC) (Figure 2.1) were then manually designed based on this gene sequence and then sent for chemical synthesis (and subsequent supply) to the Inqaba Biotechnological Sciences (Pretoria, South Africa).

```
GAA TTT GGG GTC TCA AAC GCG CAC GAC ATT CCT AAA TTC GTG ACA ATC CCG GAA GAA GAC
 E   F   G   V   S   N   A   H   D   I   P   K   F   V   T   I   P   E   E   D
ATT AAA GCA ATC GAG AAA GTC ATC AAT GGA GTA GAA GAA GAA GAG GTG AAA AAA AAG GAA
 I   K   A   I   E   K   V   I   N   G   V   E   E   E   E   V   K   K   K   E
GAT GAA GTA GAG GAA GAG AAG AGT ATC ATA TTA GTG GAG AGA CCG GAG TTG CAG ACA ATC
 D   E   V   E   E   E   K   S   I   I   L   V   E   R   P   E   L   Q   T   I
ATA ACC GAT AAA TGG GAA ATT TTC GAA GAC GAC TTT TGT TTC ACA TGT AAG GAT ATT AAA
 I   T   D   K   W   E   I   F   E   D   D   F   C   F   T   C   K   D   I   K
GAA ACT GAT CAG CAT AGA AAG TTT AAC ATG GAT CCG AGT CCG CTG CCT CTA ATA GTT ATC
 E   T   D   Q   H   R   K   F   N   M   D   P   S   P   L   P   L   I   V   I
GAC GAG CCA GTT TAC TTC ACT CAC ACG TTA CCA GAT TTG ATT ACC TTC TAA
 D   E   P   V   Y   F   T   H   T   L   P   D   L   I   T   F   -
```

Forward primer: 5′- GAA TTT GGG GTC TCA AAC GCG CAC GAC ATT -3′
Reverse primer: 5′- GAA GGT AAT CAA ATC TGG TAA CGT GTG AGT-3′

Figure 2.1: The genomic sequence of the At1g25240-AC gene fragment and amino acid sequences of its associated truncated ENTH-AC protein. The blue highlight represents the forward priming site, the red highlight represents the annotated AC catalytic center while the green highlight represents the reverse priming site. The two manually designed primers are shown in yellow highlight. (Source: http://www.arabidopsis.org & http://www.expasy.org).

2.2.2 Extraction of the Arabidopsis Total RNA

Total RNA was extracted from 100 mg of plant leaf material harvested from the 6-week old *A. thaliana* ecotype Columbia (Col-0) plants generated in section 2.1.2 above. Extraction of the total RNA was undertaken using a Gene Jet Plant RNA Purification Mini Kit and according to the manufacturer's instructions (Thermo Scientific, Massachusetts, USA). Briefly, the leaves were placed in a cooled sterile mortar and then snap-frozen with liquid nitrogen (Afrox Industrial Gases, Klerksdorp, South Africa). The leaves were then ground to a fine powder with a chilled pestle. Using a chilled sterile metal spatula, the fine powder was quickly transferred into an RNase-free 1.5 ml micro centrifuge tube containing 500 µl of the Plant RNA Lysis Solution. The 1.5 ml tube was then vortexed for about 20 seconds and mixed thoroughly.

The tube was centrifuged for 5 minutes at 18,000x*g*. The supernatant was transferred to a clean 1.5 ml micro centrifuge tube, where a volume of 250 µl of 96% ethanol was added. The mixture was then transferred into purification column fitted within collection tube and centrifuged for a minute at 8 000x*g*. The flow-through was discarded while the column was washed with 700 µl of the Plant RNA Wash Buffer 1 through centrifugation for a minute at 8 000x*g*. This washing step was repeated with 700 µl of the Plant RNA Wash Buffer 2. The centrifugation step was repeated one more time for 2 minutes and in order to remove any excess wash buffer. Finally, the washed purification column was then placed in a new collection tube and 50 µl of the nuclease free water was then placed at the center of the column membrane followed by a centrifugation at 8 000x*g* and in order to elute RNA. The eluted RNA was then stored at-80^{o}C for further use.

2.2.3 Amplification of the ENTH-AC Gene Fragment

The targeted ENTH gene fragment harbouring the AC catalytic center (ENTH-AC) was synthesized from the total RNA generated above (section 2.2.2), whereby the total RNA was first used as a template to generate copy DNA (cDNA). The generated cDNA, together with the acquired sequence-specific primers, was then used to amplify the targeted ENTH-AC gene fragment in specialized a reverse transcriptase-polymerase chain reaction (RT-PCR) system using a Verso 1-Step RT-PCR Ready Mix RT-PCR kit, and as instructed by the manufacturer (Thermo Scientific Inc., Burlington, Canada). The proposed reaction mixtures and cycling conditions for this process are respectively shown in Tables 2.1 and 2.2 below.

Table 2.1: Components of the RT-PCR reaction mix for targeted amplification of the ENTH-AC gene fragment in a final reaction volume of 50 µl.

Composition	**Volume (µl)**	**Final Concentration**
Verso Enzyme Mix	1	
1-Step PCR Master Mix	25	1X
Forward Primer (10 µM)	1	200 Nm
Reverse Primer(10 µM)	1	200 Nm
RT Enhancer	2.5	
Water (PCR Grade)	19.5	
Template (RNA)	1	1 ng
Total Volume	50	

Table 2.2: The 1-step RT-PCR thermal cycling conditions used to amplify the ENTH-AC gene fragment.

Step	Temperature (°C)	Time	Cycles
cDNA Synthesis	50	15 min	1
Thermo-Start Activation	95	15 min	1
Denaturing	95	20 sec	45
Annealing	65	30 sec	
Extension	72	1 min	
Final Extension	72	5 min	1

2.2.3.1 Agarose Gel Electrophoresis of the Amplified ENTH Gene Fragment

The amplified RT-PCR product (Table 2.2 above) was resolved on a 1% agarose gel supplemented with 0.5 µg/ml ethidium bromide. The samples was resolved against a 100 bp Gene-Ruler DNA ladder (Fermentas International Inc., Burlington, Canada) and the gel immersed in a 1X TBE buffer at 80 volts and 250 mA current for 50 minutes. The gel was then visualized under UV light using a 2000 UV Trans-illuminator System (Bio-Rad Laboratories Inc., California, USA). Images of the resolved gel were then captured with a Chemi Doc Imaging System (Bio-Rad Laboratories Inc., California, and USA).

2.2.4 Cloning of the Amplified ENTH-AC Gene Fragment

2.2.4.1 Addition of 3′-adenines Overhangs

A total volume of 1 µl *Taq* polymerase was added to 40 µl of the RT-PCR reaction product (section 2.2.3) and the mixture was then incubated at 72°C for 10 minutes on a C1000 Thermo-cycler System (Bio-Rad Laboratories Inc., California, and USA). The resultant reaction mixture was kept on ice for further use.

2.2.4.2 Ligation of the ENTH-AC Insert into the pTrcHis2-TOPO Vector

An aliquot of 4 µl was collected from the adenylated ENTH gene fragment reaction mixture and then transferred into a fresh PCR tube followed by the addition of 1 µl of the pTrcHis2-TOPO expression vector (Invitrogen, Carlsbad, USA). The mixture was then incubated at

room temperature for 5 minutes before its subsequent use for the transformation process of competent *E. coli* expression cells.

2.2.4.3 Transformation of *E. coli* One Shot TOPO 10 Competent Cells with the pTrcHis2-TOPO:ENTH-AC Fusion Construct

Immediately after the ligation process, about 2 µl of the ligation mixture (pTrcHis2-TOPO: ENTH-AC fusion construct) was transferred into an ice-cold Eppendorf tube containing about 40 µl of the chemically competent *E. coli* One Shot TOPO 10 cells. The reaction mixture was gently mixed through pipette tip swirling and incubated on ice for 30 minutes. The mixture was then heat-shocked on a dry-bath heating block at 42°C for 30 seconds before being immediately incubated on ice for 5 minutes. Thereafter, the reaction mixture was supplemented with 250 µl of SOC medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM Mg_2SO_4 and 20 mM glucose) and incubated in a shaker at 37°C at 225 rpm for 30 minutes (to allow cells to produce the β-lactamase, which is useful for selection process of detoxifying ampicillin). The mixture was then plated (80 µl and 20 µl) onto two Luria Bertani (LB) agar plates (1% (w/v) agar, 1% (w/v) tryptone powder, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) supplemented with 50 µg/ml ampicillin and 0.5% glucose. The plates were then incubated at 37°C overnight.

2.2.4.4 Colony PCR

In order to check whether the colonies produced from section 2.2.4.3 above were carrying the generated pTrcHis2-TOPO: ENTH-AC fusion construct or not, single colonies were picked followed by performance of colony PCR, and in accordance with the standard MyTaq Mix protocol (Bioline, London, UK). In order to easily manipulate the selected colonies, each one of them was resuspended into 5 µl sterile distilled water followed by the use of such a suspension solution in the PCR reactions. The used reaction mixtures and cycling conditions for this process are respectively shown in Tables 2.3 and 2.4 below.

Table 2.3: Reaction components of a colony PCR to check for presence or absence of the pCRT7/NT-TOPO-ENTH-AC fusion construct in the transformed competent *E. coli* One Shot TOPO 10 cells.

Component	Volume (µl)
Template (DNA)	1
Insert Primers (20 µM)	1
Insert Primers (20 µM)	1
MyTaq Reddy Mix, (2x)	25
Water (sdH2O)	Up to 50

Table 2.4: Themocycling conditions for the 1 step-by-step colony PCR amplification of the ENTH-AC gene fragment.

Step	Temperature (ºC)	Time	Cycles
Initial Denaturation	95	1 min	1
Denaturing	95	15 sec	35
Annealing	65	15 sec	
Extension	72	10 sec	

The amplified PCR products were then resolved by agarose gel electrophoresis as has already been outlined in section 2.2.3.1 above. In all cases, a re-amplification of the ENTH-AC gene fragment would then mean a positive result.

2.2.4.5 Extraction of the pTrcHis2-TOPO:ENTH-AC Fusion Construct from the Transformed *E. coli* One Shot TOPO 10 Cells

After colony PCR, all colonies with the positive outcomes were then used for the subsequent extraction of the contained pTrcHis2-TOPO: ENTH-AC fusion construct. In this regard, each remaining 4 µl of the resuspension culture was individually used to inoculate 5 ml fresh double strength yeast-tryptone (2YT) media (0.8% (w/v) tryptone powder, 0.5% (w/v) yeast extract and 0.25% (w/v) NaCl) supplemented with 50 µg/ml ampicillin. The inoculated cultures were incubated overnight at 37ºC with shaking at 225 rpm. The following day, cells were harvested by centrifuging at 6 800x*g* for 5 minutes at room temperature and the supernatant discarded. Plasmid extraction was then performed at room temperature using the Gene Jet Plasmid Miniprep Kit (Thermo Fisher Scientific Inc., California, and USA).

Briefly, the pelleted cells were resuspended into 250 µl of Resuspension solution (supplemented with RNase) by pipetting up and down or until no cell clumps had remained. To this culture, a total of 250 µl of Lysis solution was added followed by thorough mixing through inversion of the tube 4-6 times or until the solution had become viscous and slightly clear. Thereafter, 350 µl of Neutralization solution was added and the solution immediately mixed thoroughly through inversion of the tube 4-6 times. Cell debris was then pelleted by centrifuging at 16 300x*g* for 5 minutes and the resultant supernatant then transferred into a supplied GeneJET spin column fitted into a collection tube. The spin column collection was then centrifuged at 16 300x*g* for 1 minute and its flow-through discarded while the column was re-placed into the same collection tube. About 500 µl of Wash solution (diluted with ethanol prior to first use) was added to the spin column and the collection further centrifuged at 16 300x*g* for 30-60 seconds. The flow-through was discarded while the column was placed back into the same collection tube. The washing procedure was then repeated twice more before the spin column was further centrifuged for an extra 1 minute, to remove any residual wash solution. The washed GeneJET spin column was then transferred into a sterile Eppendorf tube where the plasmid DNA was then be eluted by adding 50 µl of Elution buffer right to the center of the GeneJET spin column membrane followed by an incubation of 2 minutes at room temperature and then a centrifugation at 16 300x*g* for 2 minutes. The used spin column was then discarded while the purified plasmid DNA was stored at -20ºC for further use.

2.2.4.6 Analysis of Positive Clones

This step was carried out so as to check and confirm if the ENTH-AC gene insert had been successfully cloned (in the correct orientation) into the pTrcHis2-TOPO vector. Reaction mixtures and cycling conditions for this process were set as already been outlined in section 2.2.4.4 and presented in Tables 2.3 and 2.4 above but with some minor changes. Firstly, the template DNA in Table 2.3 was this time around changed from a cell colony culture to a purified pTrcHis2-TOPO: ENTH-AC plasmid DNA. Secondly, a second set of a reaction mixture like the one in Table 2.3 was prepared but with the ENTH-AC forward primer now replaced by the pTrcHis2-TOPO vector forward primer. Lastly, the annealing temperature of 65°C in Table 2.4 was this time around changed to 45°C for the second set of the reaction mixture. Both reactions were then run and the generated PCR products then resolved by agarose gel electrophoresis and as has already been outlined in section 2.2.3.1 above. In this case, a re-amplification of the ENTH-AC gene fragment in both reaction sets would mean its

positive and successful cloning into the pTrcHis2-TOPO vector. For further verification and confirmation, the correctly identified pTrcHis2-TOPO: ENTH-AC fusion construct was also send to Inqaba Biotechnological Sciences (Pretoria, RSA) for nucleotide sequencing.

2.3 *In Vivo* Assaying of the AC Activity of the Recombinant ENTH-AC Protein

2.3.1 Preparation of Competent *cya*A *E. coli* Cells

Some SP850 *E. coli cya*A cells were first obtained from the Coli Genetic Stock Centre (Yale University, Connecticut, USA) and then prepared to be chemically competent before they were transformed with the pTrcHis2-TOPO:ENTH-AC fusion construct. Briefly, a fresh 2YT media supplemented with 15 g/ml of kanamycin to select against non-*cya*A cells, was prepared whereby about 200 µl of the SP850 *E. coli cya*A glycerol stock was used to inoculate 10 ml of 2YT media. The culture was then incubated at 37°C overnight, shaking at 200 rpm in an SI-600 orbital shaker (Lab Companion, Boksburg, South Africa). The following morning, 1 ml of the overnight cell culture was then used to inoculate 100 ml of the fresh 2YT media supplemented with 15 g/ml of kanamycin. Thereafter, the culture was incubated on a shaker set at 200 rpm and 37°C, and up until an OD_{600} of 0.6 was reached. The cell culture was then placed on ice to cool down for 5 minutes and to arrest growth. The culture was then transferred into a sterile ice-cold micro centrifuge tube followed by centrifugation at a low speed of 4 000 rpm and a low temperature of 4°C for 5 minutes and in order to harvest the cells. The supernatant was discarded while the pelleted cells were gently resuspended into an ice-cold TFB1 buffer (Transformation buffer 1) containing 30 mM KAc, 50 mM $MnCl_2$, 75 mM $CaCl_2$, 15% glycerol, 85 ml SD-H_2O, and a pH (KOH) of 5.8). The resuspended solution was then incubated on ice for 90 minutes and the cells then harvested by centrifugation at 4000 rmp at a 4°C for 5 minutes. The harvested cells were then resuspended into 2 ml of the ice-cold TFB2 buffer (Transformation buffer 2) containing 10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% glycerol, 85 ml SD-H_2O, and a pH (KOH) of 6.8). Several aliquots of 200 µl of the cells were then prepared in sterile micro-centrifuge tubes that were then kept on ice for subsequent uses.

2.3.2 Transformation of the *cya*A *E. coli* Cells with the pTrcHis2-TOPO: ENTH-AC Expression Construct

In this section, the chemically competent *cya*A *E. coli* cells prepared in section 2.3.1 above were transformed with the pTrcHis2-TOPO: ENTH-AC fusion construct generated in section 2.2.4.5 above. The transformation process was as exact as the one previously described in section 2.2.4.3 above for the chemically competent *E. coli* One Shot TOPO 10 cells when transformed with the same plasmid fusion construct.

2.3.3 Activity Assaying

A MacConkey agar plate supplemented with 15 µg/ml kanamycin and 0.1 mM IPTG (Sigma-Aldrich Corp., Missouri, USA) was prepared and then sub-divided into 3 quadrants using a permanent marker. The first quadrant was left un-streaked (no *cya*A cells), the second quadrant was streaked with the non-transformed *cya*A mutant cells while the last quadrant was streaked with the c*ya*A mutant cells transformed with the pTrcHis2-TOPO:ENTH-AC fusion construct (from section 2.3.2). The plate was inverted and incubated at 37°C for 40 hours. After the incubation period, all quadrants were then visually inspected for various phenotypic characteristics. A deep red or purple colour observed on the transformed *cya*A mutant cells would mean a positive AC activity for the cloned and partially expressed ENTH-AC recombinant protein.

2.4 Bioinformatic Expression Analysis of the *ENTH* Gene in *Arabidopsis thaliana*

In order to augment our biochemical work on the recombinant ENTH-AC protein above, we resorted to the bioinformatic analysis of its gene (the At1g24250 or *ENTH*). In this approach, a wide array of web-based softwares and programs were employed so as to elucidate the probable and/or exact physiological roles of the ENTH protein in *Arabidopsis thaliana* and other related higher plants.

2.4.1 Anatomical Expression Analysis of the *ENTH* Gene

The expression levels of the *ENTH* gene in various tissues of the *A. thaliana* were obtained using a microarray database and expression-data analysis tool, GENEVESTIGATOR Version

V3 (www.genevestigator.com/gv/) (Zimmermann *et al*., 2004; Grennan, 2006). This tool uses the Affymetrix Arabidopsis genome array platform of the 260011_At probe to provide information about the genomic transcriptome information of a specific selected gene. In this case, the At1g24250 gene was used as the query term before the arbitrary values of the expression intensities of *ENTH* in various Arabidopsis tissues were retrieved.

2.4.2 Developmental Expression Analysis of the *ENTH* Gene

In order to determine the expressional levels of the *ENTH* gene in the Arabidopsis plant at the different stages of its development, the AtGeneExpress visualization and developmental tool (www.arabidopsis.org) were used, whereby the At1g24250 gene was queried against 10 different developmental stages of the plant (Zimmermann *et al*., 2004; Grennan, 2006). Such developmental stages were set from the radicle emerging from the seed coat up to the release of mature seeds from the pod.

2.4.3 Co-expressional Analysis of the *ENTH* Gene

To determine the co-expressional profile of the *ENTH* gene with the other related genes in the Arabidopsis, the Arabidopsis co-expression tool (ACT) (http://www.genevestigator.com) was used. This analysis was carried out across all microarray experiments using At1g24250 (*ENTH*) as the search gene, and leaving the gene list limit blank to obtain a full correlation list. The tool makes use of signal intensities from microarray experiments to obtain the Pearson correlation co-efficient (r-value), which indicate the linear associations of various expressions between a reference gene (At1g24250, *ENTH*) and all other Arabidopsis genes represented on the selection chip. The tool then calculates both the positive and negative correlations (ranges from -1 to +1), which are measures of statistical significance, expressed as probability (P) and expectation (E) values.

2.4.4 Stimuli Specific Analysis of the *ENTH* Gene

Using the stimulus tool, the expression profiles of the *ENTH* and its top most 50 co-expressed genes (*ECGG-50*) (the *ENTH*:*ECGG-50*) were screened over the ATH1:22K array affymetrix public microarray data in the GEVESTIGATORVersion V3 (http://www.genevestigator.com). The normalized microarray data were then downloaded and anaylzed the by GEO (NCBI)

(www.ncbi.nlm.gov/geo), the TAIR GeneExpress (www.ebi.ac.uk/microarrays-as/ac) and the NASC Arrays (www.affymetrix.arabidopsis.info/narrays/experimentbrowse.pl) for experiments that induced differential expression of the co-expressed genes. The fold change (log2) value was measured for each experiment that induced expression, and consequently providing expression values through the Multiple Array Viewer program of the Multiple Viewer (MeV) software package (version 4.2.01) (The institute for Genomic Research (TIGR)).

CHAPTER THREE

RESULTS AND DESCRIPTIONS

3.1 Isolation and amplification of the ENTH-AC Gene Fragment

In order to isolate the desired ENTH-AC gene fragment, total RNA was first isolated from the 6-week old Arabidopsis plants followed by the generation of copy DNA (cDNA). Using the manually designed sequence-specific primers and the extracted total RNA in a 1-Step RT-PCR system, the targeted ENTH-AC gene fragment was then successfully isolated. The isolated and amplified 348 bp ENTH-AC gene fragment was then resolved on a 1% agarose gel as is shown in Fig 3.1 below.

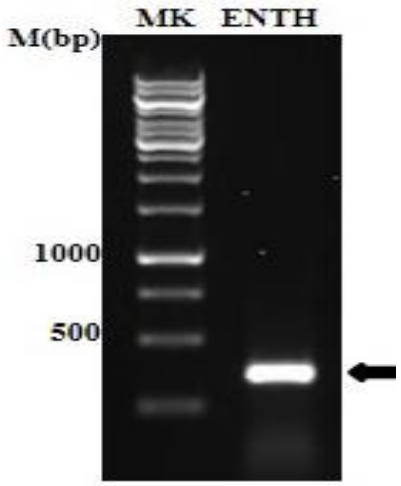

Figure 3.1: Agarose gel resolution of the ENTH-AC gene fragment isolated from the 6-week Arabidopsis thaliana plants using a Verso™ 1-Step RT-PCR Reddy Mix system. The code MK represents the 100 base pair molecular weight maker, the code ENTH represents the ENTH-AC gene fragment while the arrow is marking the amplified and resolved 348 bp ENTH-AC gene fragment.

3.2 Screening for the Successful Transformants

After isolating the ENTH-AC gene fragment from *Arabidopsis thaliana* by RT-PCR, it was then ligated into the pTrcHis2-TOPO expression vector to yield a pTrcHis2-TOPO:ENTH-AC fusion construct. This fusion construct was then used to transform some chemically competent *E. coli* One Shot TOPO 10 cells in preparation for the subsequent activity assaying of the anticipated ENTH-AC recombinant protein. The transformed cells were then checked by

colony PCR for successful transformation. As is shown in Fig 3.2 below, the assessment in deed showed that the transformed cells were harbouring the cloned pTrcHis2-TOPO:ENTH-AC fusion construct.

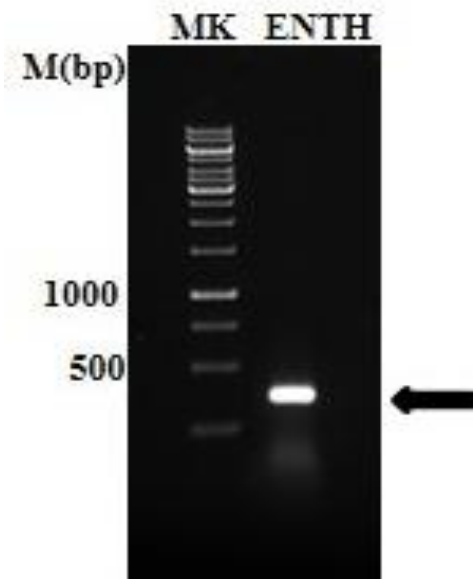

Figure 3.2: A 1% agarose gel resolution of the ENTH-AC gene fragment after its re-amplification by colony PCR of the *E. coli* One Shot TOPO 10 cells harbouring the pTrcHis2-TOPO:ENTH-AC fusion construct. MK represents the 100 base pair molecular weight maker while ENTH represents the ENTH-AC gene fragment and the arrow is marking the re-amplified 348 bp ENTH-AC gene fragment.

3.3 Screening for the Correct Recombinants

After obtaining the successful transformants by colony PCR, all the harboured pTrcHis2-TOPO:ENTH-AC fusion constructs in these transformants were then isolated followed by their screening by standard PCR for the correct in-frame orientation of the ENTH-AC gene fragment. Fig 3.3 below shows a gel resolution of a pTrcHis2-TOPO: ENTH-AC fusion construct carrying the ENTH-AC gene fragment in the correct in-frame orientation.

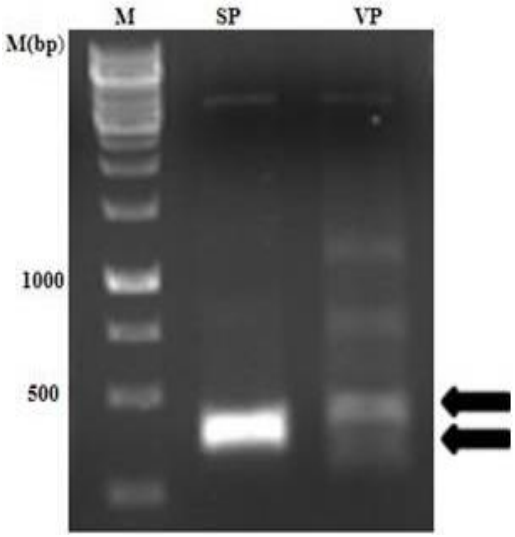

Figure 3.3 A 1% agarose gel resolution confirming the correct orientation of the ENTH-AC gene fragment in the pTrcHis2-TOPO expression vector, where M is the molecular weight marker, while SP is representing the ENTH-AC gene fragment amplified by its own primers and VP representing the ENTH-AC gene fragment amplified by one of its own primers and the vector primer. The two arrows are marking the amplified ENTH-AC gene fragments.

3.4 Determination of the *In Vivo* AC Activity of the Recombinant ENTH-AC Protein

After identifying the successfully cloned pTrcHis2-TOPO: ENTH-AC fusion construct, its possible recombinant expression and AC functional activity were then assessed by using it to transform some chemically competent mutant *cya*A SP850 *E. coli* cells followed by analysis on the MacConkey agar. This strain of *E. coli* is naturally non-lactose fermenting as a result of its lack of the AC activity. However, when complemented with a functional exogenous AC molecule, it can then ferment the lactose. Fig 3.4 below shows such a successful complementation scenario of this mutant strain by the recombinant ENTH-AC protein, and therefore signifying that the ENTH-AC gene fragment was both expressible and catalytically functional.

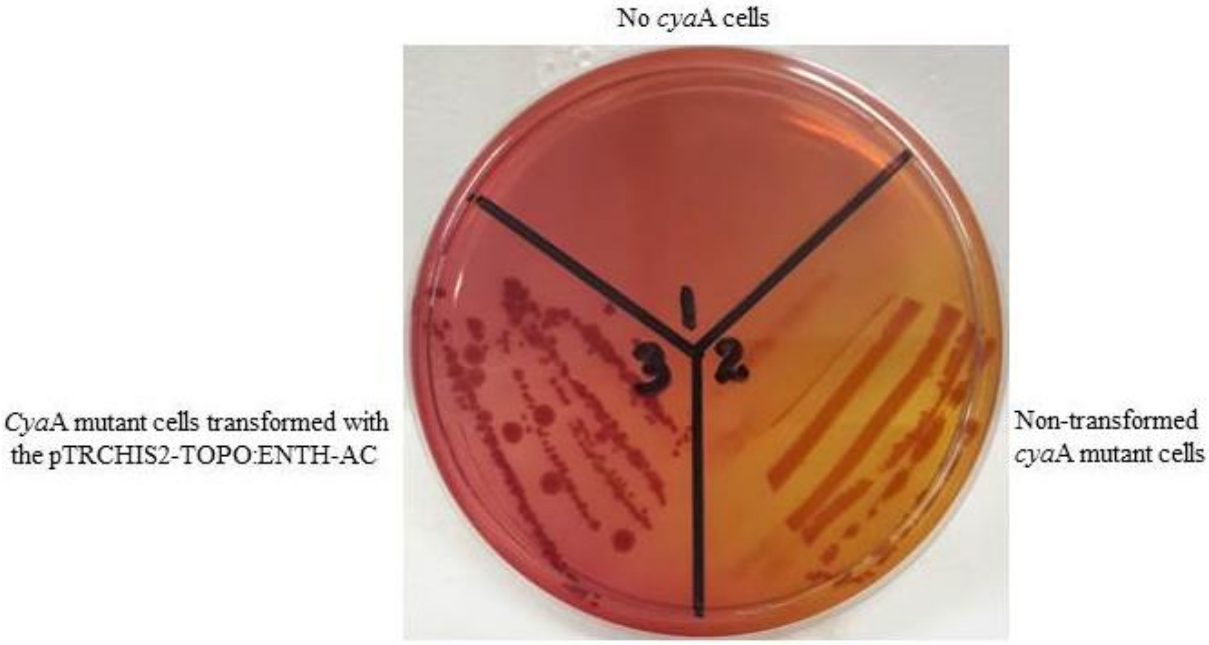

Figure 3.4: Complementation testing of the recombinant ENTH-AC protein using the mutant *cya*A cells. The 1st quadrant has no cells and hence no growth observed, the 2nd quadrant has the non-transformed *cya*A mutant cells which are non-lactose fermenters and producing yellow or white colonies, and the 3rd quadrant has the *cya*A mutant cells transformed with the pTrcHis2-TOPO:ENTH-AC fusion construct now with the ability to ferment lactose and producing deep red or purple colonies.

3.5 Determination of the Anatomical Expression of ENTH

A Genevestigator analysis of the At1g24250 gene showed that the ENTH protein is exclusively expressed in various tissues (over 105) of the *Arabidopsis thaliana* plant including the root, the collective flower carpel, the cotyledon, the leaf, the meristem, the stamen, the stem, the cell embryo and the seedling (www.arabidopsis.org). Furthermore, the analysis also indicated that the ENTH protein is highly expressed in the root protoplast, the primary root and the giant cell (Fig 3.5 below).

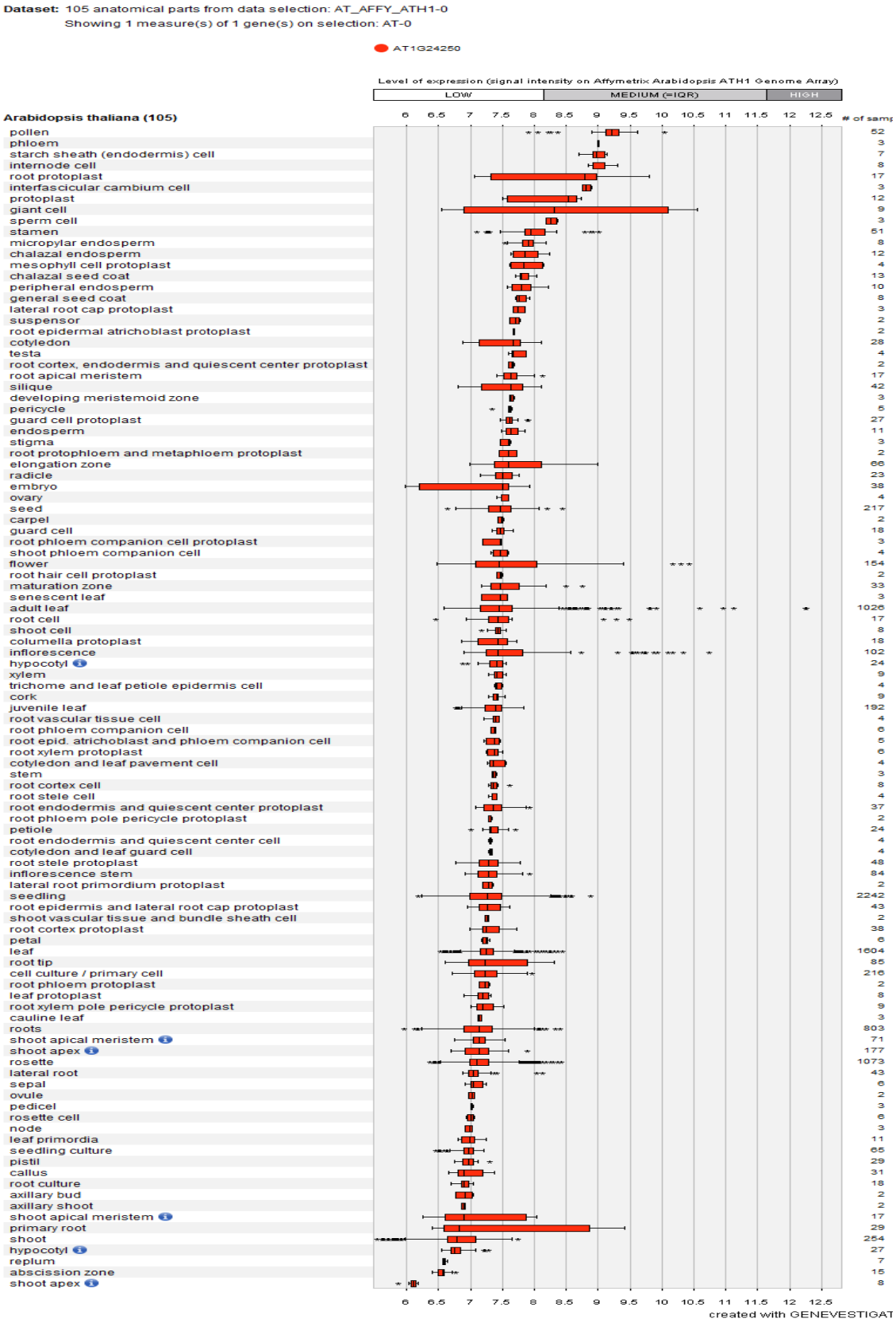

Figure 3.5: Expression intensity levels of the ENTH-AC protein in various tissues of the *Arabidopsis thaliana*. The ENTH protein shows that it is exclusively expressed across the various parts of the plant but most expressed in the root protoplasts, primary roots and the giant cells (Retrieved from the Genevestigator anatomy tool).

3.6 Determination of the Developmental Expression of ENTH

In order to obtain a comprehensive expressional profile of the ENTH protein during the various stages of development in the Arabidopsis plant, the AtGeneExpress visualization and development tool (Zimmermann *et al*., 2004) was used to screen the microarray data using At1g24250 as the search query. The results obtained, yielded 10 developmental stages during which the ENTH protein is expressed and indicating that the protein is highly expressed when the plant begins to flower and least expressed when the radicle emerges from the seed coat (Fig 3.6).

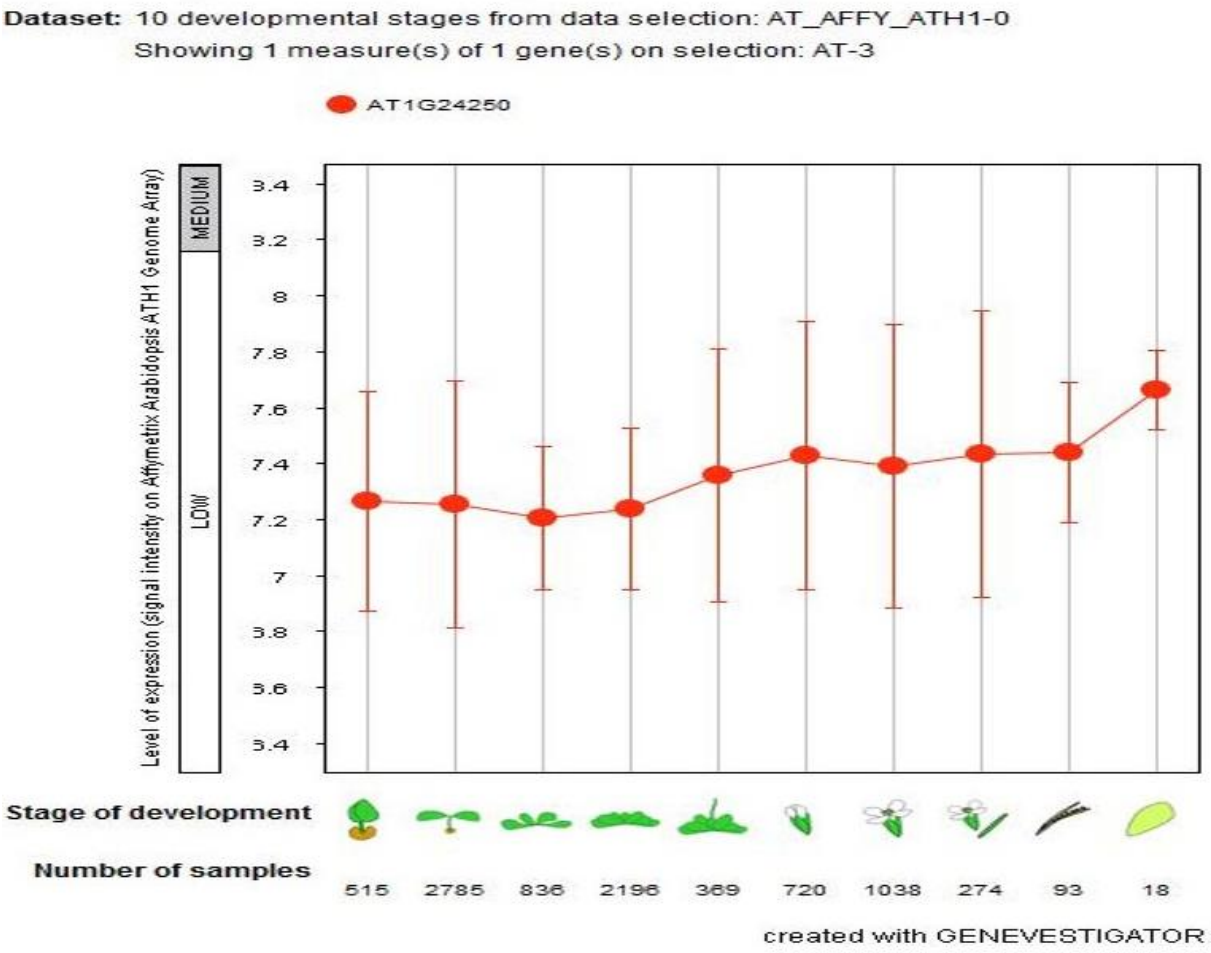

Figure 3.6: Expression profiles of the ENTH protein during the different developmental stages of the *Arabidopsis thaliana* plant. The above data shows measures of expression levels in relation to the various stages of development beginning from the time when the radicle emerges from the seed coat up to the time when mature seeds are released from the pods. The profile indicates that the ENTH protein is highly expressed during the flowering stage and least expressed during the seed germination stage.

3.7 Determination of the Co-expression Profile of ENTH

In order to obtain a list of other proteins that are co-expressed with the ENTH protein in the *Arabidopsis thaliana* plant, the expression profile values of the At1g24250 gene were measured using the Arabidopsis co-expression tool (ACT) (http://www.genevestigator.com) on the available microarray data with about 359 diverse transcriptome experiments. From these diverse chip experiments, the ENTH protein was found to be co-expressed with a number of other proteins and the 50 top most co-expressed proteins having Pearson correlation values of 0.87 to 0.98 (Table 3.1).

Table 3.1: A list showing the top most 50 proteins that are co-expressed with the ENTH protein in the *Arabidopsis thaliana*.

Rank	*r* value	Locus and GO terms	Description or Annotation
1	0.98	AT5G46050	Protein NT1/PTR Family
2	0.96	AT5G61110	Zinc ion binding
3	0.94	AT1G74870	RING/U-box superfamily protein
4	0.94	AT3G21000	Gag-Pol- related retrotransposon family protein
5	0.94	AT3G18610	Nucleolin like 2 (NUC-L2)
6	0.93	AT5G25820	Exostosin family protein
7	0.92	AT1G15320	Unknown protein
8	0.92	AT5G44330	Tetratricopeptide repeat (TPR) –like super family
9	0.92	AT5G24820	Eukaryotic aspartyl protease family protein
10	0.92	AT5G60090	Protein kinase superfamily
11	0.92	AT3G57370	Cyclin family protein
12	0.91	AT1G68200	Zinc finger (CCCH-TYPE) family protein
13	0.91	AT5G08250	Cytochrome P450 superfamily protein
14	0.91	AT5G51270	Putative U-box domain-containing protein kinase
15	0.91	AT3G58540	Unknown protein
16	0.91	AT5G35600	Histone deacytylase 7
17	0.90	AT1G60240	Transcriptional regulator protein
18	0.90	AT3g60020	Ubiquitin-protein ligase activity
19	0.90	AT5g23190	Cytochrome P450
20	0.90	AT1G32460	Unknown protein
21	0.90	AT3G52130	Bifunctional inhibitor/lipid-transfer protein
22	0.90	AT4G35420[PEF,PW.]	Tetraketide alpha-pyrone reductase 1
23	0.89	AT1G75110	Nucleotide-diphospho-sugar transferase protein
24	0.89	AT2G35736	Unknown protein

25	0.89	AT4G30040	Eukaryotic aspartyl protease family protein
26	0.89	AT4G34270	TIP41-like family protein
27	0.89	AT4G33930	Cupredoxin super family
28	0.89	AT4G03320	Translocon at inner envelope membrane protein
29	0.89	AT2G27630	Ubiquitin carboxyl-terminal hydrolase-related
30	0.89	AT5G41890	GDSL-like Lipase/Acylhydrokase superfamily
31	0.89	AT5G53680	RNA-binding (RRM/RBD/RND motifs) family
32	0.89	AT3G57390	Agamous-like 18
33	0.89	AT2G46480	Putative galacturonosyltranferase 2
34	0.89	AT4G10950	SGNH hydrolase-type esterase superfamily
35	0.88	AT5G40940	Putative fasciclin-likearabinogalactan protein 20
36	0.88	AT1G80970	XH domain-containing protein
37	0.88	AT1G65930	Cytosolic NADP+-dependent isocitrate
38	0.88	AT1G74150	Galactose oxidase/kelch repeat superfamily
39	0.88	AT5G52160	Bifunctional inhibitor/lipid-transfer protein
40	0.88	AT5G20240	K-box region and MADS domain transcrip-factor
41	0.88	AT3G61340	F-box and interaction domains-containing
42	0.88	AT5G62080 PEF ,PW	Bifunctional inhibitor/lipid-transfer protein
43	0.88	AT5G60080	Protein kinase superfamily protein
44	0.88	AT3G59530 CCA,PEF,PD	Calcium dependent phosphotriesterase
45	0.88	AT1G71230	COP9-signalosome 5B
46	0.88	AT4G39740	Thioiredoxin superfamily protein
47	0.88	AT1G69500 PEF,SP,CCA	Cytochrome P450,designated CYP704B1
48	0.88	AT1G05020	Clathrin coat assembly
49	0.87	AT2G31035	Oxysterol binding protein
50	0.87	AT2G18260	Syntaxin of plants 112

Abbreviations for the indicated GO terms:
PEF = pollen exine formation; **SP** = sporopollenin biosynthetic process; **PW** = pollen wall assembly, **PD** = pollen development; **CCA** = cellular component assembly involved in morphogenesis.

3.8 Determination of the Stimulus-specific Expression of ENTH

After establishing the co-expressional profile of the ENTH protein and its indicative functional roles together with its 50 top most co-expressed genes (ECGG-50), the whole set (ENTH:ECGG-50) was then subjected to an *in-silico* global expression analysis in which specific experimental conditions that could induce the differential expression of all genes in the set were identified. In accordance with the co-expression and GO analysis, the heat maps generated from the microarray expression analysis revealed that the transcriptional processes

of the ENTH protein and its related ECGG-50 proteins (the ENTH:ECGG-50 pool) are generally and collectively induced in response to a wide range of biotic stress factors (Fig 3.8).

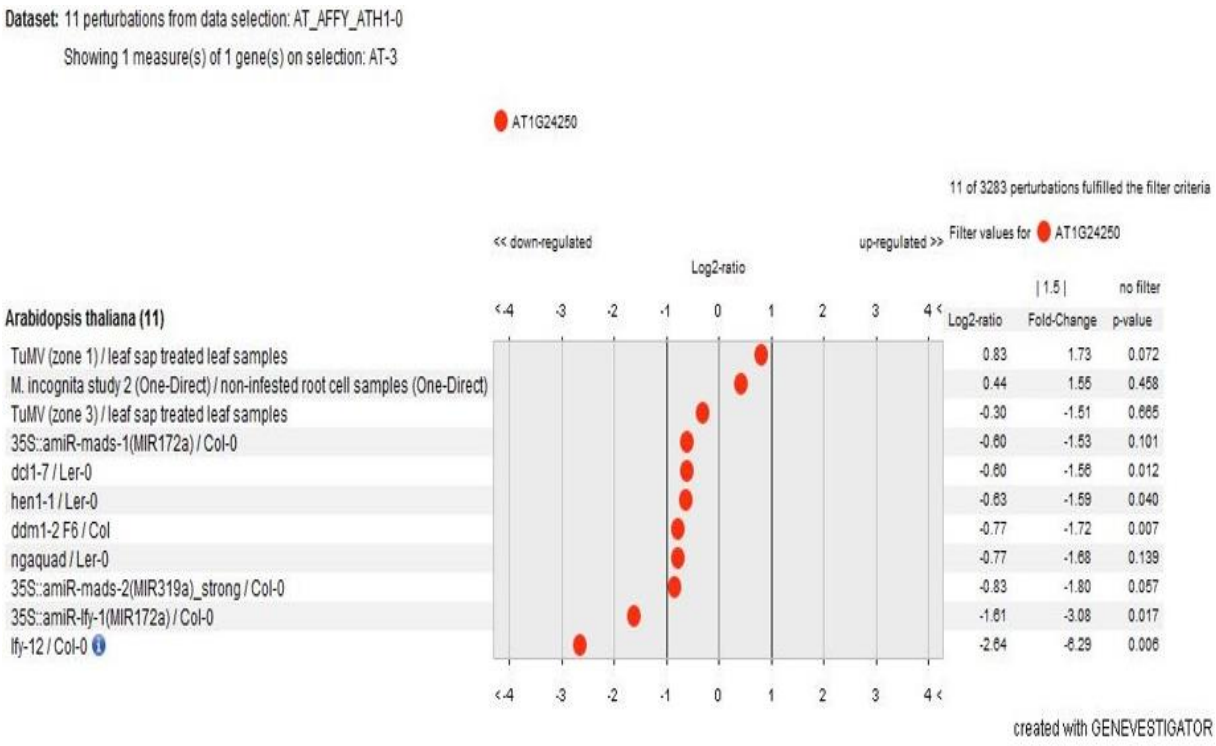

Figure 3.7: Expression profiles of the ENTH protein in *Arabidopsis thaliana* in response to the various stimulus-specific conditions. The profile indicates a differential induced expression of the protein in response to a wide array of biotic factors.

CHAPTER 4

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

4.1 Discussion

Adenylate cyclases (ACs) are enzymes which catalyse the conversion of adenosine 5′-triphosphate (ATP) to 3′,5′-cyclic adenosine monophosphate (cAMP) (Gehring, 2010). The generated cAMP then functions as a secondary messenger molecule with roles in various cell signalling pathways triggered by external stimuli (Hussain *et al*., 2003). Apparently, while the presence and functions of ACs and cAMP in the rest of other living organisms has been well documented and established, their presence and/or functions in higher plants has been doubted and strongly debated for quite some time (Robinson *et al*., 2011). This was mainly due to the less efficient methods employed to find these molecules as well as the low levels of cAMP (<20 pmol/g fresh weight) in plants compared to eukaryotes (>250 pmol/g) (Lemtiri-Chlieh, 2011), hence the necessity to study these molecules in a more detail in plants. Incidentally, Gehring (2010) bioinformatically identified nine candidate genes from *Arabidopsis thaliana* bearing the AC catalytic centre (Table 1.1) and as a model to demonstrate the presence of ACs in higher plants (Gehring, 2010). Amongst them was the At1g24250 gene coding for an epsin N-terminal homology (ENTH) protein and also suspected to harbour a functional AC catalytic center as already been discussed in Chapter 1 above. However, and up to the present day, no study has ever attempted to determine whether or not this protein is able to synthesize cAMP from ATP or having any functional role(s) in plants. This study was, therefore, designed to experimentally test and determine if the annotated ENTH protein is a functional AC or not. To do that, the study had to recombinantly clone and then functionally characterize its truncated version harbouring the annotated AC catalytic center (the ENTH-AC protein) at the laboratory level.

To begin with, some *A. thaliana* plants were generated and raised to fully grown plants followed by isolation of their total RNA. During that same time, some sequence-specific primers for the targeted ENTH-AC gene fragment were also designed and acquired (Fig 2.1). Using both the isolated total RNA and the manually designed sequence-specific primers, a specialized RT-PCR system was then used to isolate and amplify the desired 348 bp ENTH-AC gene fragment (Fig 3.1). The amplified ENTH-AC gene fragment was then cloned into a pTrcHis2-TOPO expression vector to generate a pTrcHis2-TOPO:ENTH fusion construct, which was then maintained and propagated in chemically competent *E. coli* One Shot TOPO

10 cells. Apparently, since the pTrcHis2-TOPO is a linear and non-directional expression vector with 5′-thymine (T) cloning overhangs, it was therefore, very essential to check and verify if the cloned ENTH-AC gene fragment was properly ligated in the correct expression orientation. Firstly and after transformation of the *E. coli* One Shot TOPO 10 cells with the pTrcHis2-TOPO:ENTH fusion construct, colony PCR was undertaken to validate presence of the ENTH-AC gene fragment (Fig 3.2). Secondly, a confirmatory PCR, whereby the pTrcHis2-TOPO:ENTH fusion construct was used as a template for a new set of primers (vector forward and insert reverse), was also undertaken. As is shown in Fig 3.3, this PCR setting succeeded to re-amplify the ENTH-AC gene fragment and thereby confirming its correct in-frame orientation in the pTrcHis2-TOPO vector. Otherwise if the orientation was not correct, then no amplification was going to be achieved. More so, the yielded fragment was slightly larger than the original ENTH-AC band (the 348 bp with its own self-primers) because part of the pTrcHis2-TOPO vector was in this case, included. Lastly, the confirmed pTrcHis2-TOPO:ENTH-AC fusion construct was also sent over to the Inqaba Biotechnological Sciences (Pretoria, RSA) for chemical sequencing, and the correct orientation was further validated (results not shown).

After confirming the correctness of the designed pTrcHis2-TOPO:ENTH-AC fusion construct, the probable AC functional activity of the ENTH-AC recombinant protein was then tested *in vivo* through a complementation system of the *cya*A SP850 mutant *E. coli* cells. This strain of *E. coli* totally lacks the AC activity in its system as a result of the unavailability of the only AC gene (*cya*A gene) in its cellular profile and therefore, cannot metabolize carbon sources such as lactose, maltose, and galactose (Moutinho *et al*., 2001). When this mutant strain is grown on MacConkey medium, there is a very distinct phenotypical difference between itself and its wild type, whereby it appears white or yellowish while the wild type appears deep red or purplish (Cotta *et al*., 1998). Technically, this mutant strain provides a very suitable biochemical platform for testing the probable AC activity of foreign recombinant proteins like the ENTH-AC via a complementation phenomenon of its *cya*A mutation (Cotta *et al*., 1998). Previously, it has already been demonstrated that when a *cya*A mutant is transformed with a heterologous system, which codes for a functional putative AC gene, the AC gene will rescue the *cya*A mutation (complementation) (Cotta *et al*., 1998) and enabling the mutant strain now to ferment lactose and producing deep red or purplish colonies on MacConkey agar (Perlman and Pastan, 1969).

In this case, when the pTrcHis2-TOPO:ENTH-AC fusion construct was used to transform the chemically competent *cya*A SP850 mutant *E. coli* cells, the transformed cells turned deep red (Fig 3.4) and therefore signifying the *in vivo* AC functionality of the recombinant ENTH-AC protein. This outcome is very similar to other previous related findings whereby the *E. coli cya*A SP850 mutant strain was transformed and tested against the various unknown foreign plant proteins. Such proteins include the PSiP protein from *Zea mays* (Moutinho *et al*., 2001); the AtPPR-AC protein from *Arabidopsis thaliana* (Ruzvidzo *et al*., 2013); the HpAC1 protein from *Hippeastrum hybridum* (Swiezawska *et al*., 2014); and the AtKUP7 protein from *Arabidopsis thaliana* (Al-Younis *et al*., 2015). In this regard therefore, the recombinant ENTH-AC protein was thus proved to be a functional AC molecule capable of generating cAMP from ATP and thereby becoming the fifth ever higher plant AC to be confirmed this way.

After confirming the AC functional activity of the ENTH-AC recombinant protein, a further functional analysis of ENTH protein in *A. thaliana* and possibly other related higher plants was then undertaken using bioinformatics, whereby a wide range of web-based softwares and computer-based programs were employed (Zimmermann *et al*., 2004). When assessed for expression in the various tissues of the Arabidopsis plant, the ENTH protein was affirmed that it is expressed the most in root protoplasm, the giant cells and primary roots but least expressed in other tissues such as the shoot apex and the abscission zones (Fig 3.5). Using the development tool to screen a microarray data with the At1g24250 gene as the search query, the obtained results indicated that the ENTH protein is expressed during 10 different developmental stages of the Arabidopsis plant, and is most expressed when the plant begins to flower but least expressed when the radicle emerges from the seed coat (Fig 3.6).

When assessed for its possible co-expression with the other related proteins in the *Arabidopsis thaliana,* the ENTH protein was found to be co-expressed with a wide range of other proteins of which the 50 top most of such proteins had the Pearson correlation values ranging between 0.87 and 0.98 (Table 3.1). Most of such co-related proteins have specific functional roles in pollen development and cellular component assembly, both processes of which are essentially mediated by cAMP. Therefore, by virtue of being co-expressed with proteins of known functions in the Arabidopsis plant, the ENTH protein is thus also similarly involved in the same or related functions since it is commonly accepted that in plants, genes which are functionally related may be transcriptionally co-expressed (Mutwil *et al*., 2008). A stimulus-specific microarray expression profiling of the ENTH protein essentially showed that this putative

protein is generally and specifically induced by a wide variety of biotic stress factors and most specifically, the fungal pathogens (Fig 3.6) The ENTH protein therefore, becomes the third ever higher plant AC to be directly implicated in pathogen related (biotic) stress responses after the HpAC1 protein from *Hippeastrum hybridum* has been shown to be involved in cAMP-mediated stress signalling in response to the *Phoma narcissi* infections (Swiezawska *et al*., 2014) and the NBAC from *Nicotiana benthamiana* protein responsible for the *Pseudomonas syringae* tabtoxinine-β-lactam induced cell deaths during wildfire diseases (Ito *et al*., 2014). Hence by combining the findings of Fig 3.4, whereby the recombinant ENTH-AC protein showed a functional *in vivo* AC activity and the findings of Fig 3.6, where the ENTH protein has been implicated in fungal pathogenic stress responses, it is therefore, conceivable to conclude that this protein is a real molecular candidate for the various cAMP-mediated stress response processes in higher plants.

4.2 Conclusion

This work has practically demonstrated that the ENTH-AC protein, which was previously annotated (bioinformatically) to harbour an AC catalytic center in the *Arabidopsis thaliana*, is indeed a *bona fide* higher plant AC with a role in cAMP-mediated fungal stress response processes.

4.3 Recommendations

1. Since the ENTH protein was previously annotated as a probable AC candidate together with the other eight Arabidopsis candidate genes, it is therefore, very essential that the other outstanding candidates are also experimentally tested and functionally confirmed as possible ACs in higher plants.
2. Since the ENTH protein has now been confirmed to have a functional role in cAMP-mediated fungal stress responses, its exact molecular mechanisms should, therefore, be further elucidated. This approach may involve the use of a fully purified ENTH-AC protein molecule and in either *in vitro* or *in planta* experimental studies.

REFERENCES

Abdallah NA, Moses V, Prakash C: **The impact of possible climate changes on developing countries: The needs for plants tolerant to abiotic stresses.** *GM Crops Food* 2014, **5:** 77-80.

Amrhein N: **Evidence against the occurrence of adenosine 3':5'-cyclic monophosphate in higher plants.** *Planta* 1974, **118:** 241-258.

Assmann SM: **Cyclic AMP as a second messenger in higher plants**. *Plant Physiol* 1995, **108:** 885-889.

Azhar S, Krishna Murti C.R: **Effect of indole-3-acetic acid on synthesis of cyclic 3'- 5' adenosine phosphate by Bengal gram seeds.** *Biochem Biophys Res Comm.* 1971, **4:** 58-64.

Bita C.E., Gerats T., **Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops**. *Front Plant Sci* 2013, **4:** 273.

Braam J, Sistrunk M.L, Polisensky DH, Xu W, Purugganan MM., Antosiewicz DM, Campbell P, Johnson KA: **Plant responses to environmental stress: regulation and functions of the Arabidopsis TCH genes.** *Planta* 1997, **203:** 35-41.

Chen H, Fre S, Slepnev VI, Capua MR, Takei K, Butler MH, Di Fiore PP, De Camilli P: **Epsin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis**. *Nature* 1998, **394:** 793-797.

Cotta MA, Whitehead TR, Wheeler MB: **Identification of a novel adenylate cyclase in the ruminal anaerobe.** *PubMed.* 1998, **164(2):** 257-260.

Gehring C**: Adenyl cyclases and cAMP in plant signalling – past and present.** *Cell Comm. Sign* 2010, **8:** 15.

Hussain NK, Natasha K, Yamabhai M, Bhakar AL, Metzler M, Ferguson SSG, Hayden MR, McPherson P.S, Kay BK: **A role for epsin N-terminal homology/Ap180 N-terminal homology (ENTH/ANTH) domains in tubulin binding.** *Biol Chem* 2003, **278:** 28823-28830.

Ichikawa T, Suzuki Y, Czaja I, Schommer C, LeBnick A, Schell J, Walden R: **Identification and role of adenylyl cyclase in auxin signalling in higher plants.** *Nature* 1997, **390:** 698-701.

Inas Al-Younis, Aloysius Wong, Chris Gehring: **The *Arabidopsis thaliana* K^+-uptake contains a functional cytosolic adenylate cyclase center.** *FEBS Letters.* 2015, (589): 3848-3852.

Ito M, Takahashi H, Sawasaki T, Ohnishi K, Hikichi Y: **Novel type of adenylyl cyclase participates in tabtoxinine-β-lactam-induced cell death and occurrence of wildfire disease in *Nicotiana benthamiana*.** *Plant Signal Behav.* 2014, **9:** e27420.

Kholodenko BN: **Cell signalling dynamics in time and space.** *Mol Cell Biol.* 2006, **7:** 165-176.

Koshiba S, Kigawa T, Kikuchi A, Yokoyama S: **Solution structure of the epsin N-terminal homology (ENTH) domain of human epsin.** *J. Struct. Funct Genom* 2002, **2:** 1-8.

Keates RAB: **Evidence that cyclic-AMP does not mediate action of gibberellic acid.** *Nature* 1973, **2:** 355-357.

Legendre-Guillemin V, Wasiak S, Hussain NK, Angers A, McPherson PS: **ENTH/ANTH proteins and clathrin-mediated membrane budding.** *J. Cell Sci* 2004, **117:** 9-18.

Lemtiri-Chlieh F, Thomas L, Marondedze C, Irving H, Gehring C: **Cyclic nucleotides and nucleotide cyclases in plant stress responses.** *InTech* 2011, **6:** 137-182.

Ludidi N, Gehring C: **Identification of a novel protein with guanylyl cyclase activity in *Arabidopsis thaliana*.** *J. Biol Chem* 2003, **278:** 6490-6494.

McMains, V.C, Liao, X.H. & Kimmel, A.R: **Oscillatory signalling and network responses during the development of *Dictyostelium discoideum***. *Ageing Res. Rev.* 2008, **7:** 234-248.

Messa M, Fernández-Busnadiego R, Sun EW, Chen H, Czapla H, Wrasman K, Wu Y, Ko G, Ross T, Wendland B, Camilli P: **Epsin deficiency impairs endocytosis by stalling the actin-dependent invagination of endocytic clathrin-coated pits.** *eLife* 2014, 3.

Meiklejohn AL, Gralla JD: **Entry of RNA polymerase at the *lac* promoter.** *Cell* 1985, **43:** 769-776.

Moutinho A, Hussey PJ, Trewavas AJ, Malho R: **cAMP acts as a second messenger in pollen tube growth and reorientation**. *Proc Natl Acad Sci USA* 2001, **98:** 10481-10486.

Mutwil M, Obro J, Willats WGT, and Persson S: **Web-tools that combine BLAST and co-expression analyses.** *Plant Physiol* 2008, **36** (Web Server Issue)**:** 320-326.

Perlman RL, Varmus HE., and Pastan I: **Regulation of inducible enzyme synthesis in *Escherichia coli* by cyclic adenosine 3′,5′-monophosphate.** *J. Biol. Chem* 1969, **244:** 5828-5835.

Polo S, Sigismund S, Faretta M, Guidi M, Capua M.R, Bossi G, Chen H, De Camilli Di Fiore PP: **A single motif responsible for ubiquitin recognition and mono-ubiquitination in endocytic proteins.** *Nature* 2002, **416:** 451-455.

Robison GA, Sutherland EW: **Cyclic AMP and function of eukaryotic cells: An introduction.** *Annals of the New York Acad Scie* 1971, **185:** 5-9.

Ruzvidzo O, Dikobe,BT, Kawadza DT, Mabadahanye GH, Chatukuta P, Kwezi L: **Recombinant expression and functional testing of candidate adenylate cyclase domains.** *Method Mol Biol* 2013, **1016:** 13-25.

Song J, Hui Lee M, Lee G, Yoo MC, Hwang I: **Arabidopsis EPSIN1 plays an important role in vacuolar trafficking of soluble cargo proteins in plant cells via interactions with clathrin, AP-1, VTI11, and VSR1.** *The Plant Cell* 2006, **18:** 2258-2274.

Stühmer W, Conti F, Suzuki H, Wang XD, Noda M, Yahagi N, Kubo H, Numa S: **Structural parts involved in activation and inactivation of the sodium channel.** *Nature* 1989, **339:** 597-603.

Sutherland, EW, Rall, TW: **The relation of adenosine 3',5'-phosphate and phosphorylase to the actions of catecholamines and other hormones.** *Pharmacol Rev* 1960, **12:** 265-299.

Swieżawska B, Jaworski K, Pawełek A, Grzegorzewska W, Szewczuk P, Szmidt-Jaworska: A**. Molecular cloning and characterization of a novel adenylyl cyclase gene, HpAC1, involved in stress signalling in *Hippeastrum hybridum*.** *Plant Physiol Biochem* 2014, **80:** 41-52.

The Arabidopsis Information Resource (TAIR) (www.arabidopsis.org).

Wachten S, Schlenstedt J, Gauss R, Baumann A: **Molecular identification and functional characterization of an adenylyl cyclase from the honeybee.** *J. Neurochem* 2006, **96:** 1580-90.

Wendland B, Steece KE, Emr SD: **Yeast epsins contain an essential N-terminal ENTH domain, bind clathrin and are required for endocytosis**. *EMBO J* 1999, **18:** 4383-4393.