Biochemical and Structural Characterization of the Adenylyl Cyclase Activity of a Fusion Epsin N-terminal Homology Protein from *Arabidopsis thaliana*

DS KHUNOU

orcid.org 0000-0002-3897-8667

Dissertation submitted in fulfilment of the requirements for the degree Master of Science in Biology at the North West University

Supervisor: Prof O Ruzvidzo

Co-supervisor: Dr. DT Kawadza

Co-supervisor: Dr. BT Dikobe

Graduation ceremony: July 2020

Student number: 23783877

DECLARATION

I, Semakaleng Dalitah Khunou, solemnly declare that this dissertation entitled **"Biochemical and structural characterisation of the adenylyl cyclase activity of a fusion Epsin N-terminal homology protein from *Arabidopsis thaliana*"** is my work and has been submitted to the Department of Botany at the North-West University, Mafikeng Campus, for the Masters of Science in Biology. This work has never been submitted to any institution of learning elsewhere for examination or other purposes. This is my own work and all the sources used or quoted here have been properly indicated and sincerely acknowledged.

SIGNATURES

..... Date: .....

Semakaleng Dalitah Khunou

(Student)

..... Date: .....

Prof O Ruzvidzo

(Supervisor)

TABLE OF CONTENTS

DEDICATION

This research is dedicated to my parents Mrs. Khutsafalo Mita Khunou and The late Mr. Ephraim Khunou, my family Regomoditswe Yvonne Khunou, Rorisang Oarabile Khunou, Kaboentle Trinity Khunou, the late Omphile Gift Khunou and Ofentse John Khunou, my guardians Mr. Geelboy Tshipo, my late aunt Mrs. Gosiamemang Tonny Maria Khunou and her family. Thank you for all your extraordinary support and motivation. I would also like to dedicate this work to my supervisors Prof. Oziniel Ruzvidzo, Dr. Dave Kawadza, Dr. Tshegofatso Dikobe and all members of the 2017-2019 Plant Biotechnology Research Group in the Botany Department at the North-West University (Mafikeng Campus) for their mentorship and guidance, but most of all, their patience.

TO GOD BE THE GLORY

ACKNOWLEDGMENTS

I would like to extend my sincere gratitude to the following individuals and organizations for their contributions towards the completion of this work: My supervisor; Prof. O Ruzvidzo and co-supervisors; Dr. TD Kawadza, Dr. TB Dikobe for their patience, encouragement, excellent supervision and strategic mentorship throughout the period of this study. Most importantly, I want to thank them for their outstanding efforts towards the accomplishment of this research, by teaching me the new techniques and giving me the opportunity to work with them in the Laboratory. When there was a lack of hope you all kept me motivated.

During the course of this research, I went through some difficult moments. However, as I mourned, those around me reminded me that the race wasn't over yet. Thus, I would love to extend my appreciation to my family and supportive companions for their prayers, encouragement and all kinds of support given at all costs to see the achievement of this research. To my parents, thank you for funding my studies and basic needs. Mom, I wouldn't have done it without you. During the challenges and losses, your words kept me sane and the memories of those I lost kept me strong.

To my colleagues and my mentors, more especially Mrs. Katlego Sehlabane-Abotseng and Mrs. Enetia Bobo, thank you once again for the knowledge, laboratory technique skills, guidance, generosity and motivation in my research work. My sincere gratitude also goes to my friends Gerald Motsatsi and Lesego Molopyane for their endless motivation and support in all the situations during the course of the practical work. Honors and Masters Teams, you have all been like family, thank you. To the 2018 Department of Botany student assistants, I enjoyed working with you, thank you for the support and cooperation. May God bless and recognize each and everyone who played a role in this work.

Last but not least, I would also like to acknowledge my sponsor, the North-West University (Mafikeng Campus) for granting me the Postgraduate Masters Bursary and the opportunity to enrol for this course and using their resources.

DEFINITION OF TERMS

Adenosine triphosphate (ATP): A high energy molecule that stores and supplies cells with the energy they require to perform specific tasks.

Adenylate cyclases (ACs): Enzymes capable of converting adenine 5′-triphosphate (ATP) to cyclic 3′,5′-adenosine monophosphate (cAMP).

***Arabidopsis thaliana*:** A small flowering plant that is widely used as a model research organism in Plant Biology.

Coatomers: A large protein complex that forms part of the coat of specific Golgi intercisternal transport vesicles that are involved in constitutive vesicular transport between the endoplasmic reticulum, Golgi apparatus, and plasma membrane.

Cyclic adenosine monophosphate (cAMP): An important second messenger molecule in many biological processes, derived from ATP and used for intracellular signal transduction systems in many different organisms.

Cytokinesis: The cytoplasmic division of a cell at the end of mitosis or meiosis, bringing about the separation of the dividing cell into two daughter cells.

Endocytosis: The process of actively transporting a molecule into the cell by engulfing it with its membrane.

Enzyme immunoassay: An antibody-based diagnostic technique used in Molecular Biology for the qualitative and quantitative detection of specific biological molecules.

Epsin N-terminal homology: A structural domain that is found in proteins involved in endocytosis and cytoskeletal machinery.

Epsin: A protein that has a function in maintaining cell membrane curvature.

Glacial period: A period in the earth's history when polar and mountain ice sheets were unusually extensive across the earth's surface.

Greenhouse effect: The trapping of the sun’s warmth in a lower atmosphere, due to the greater transparency of the atmosphere to visible radiation from the sun than to infrared radiation emitted from the planet’s surface.

Guanylate cyclases (GCs): Enzymes capable of converting guanine 5′-triphosphate (GTP) to cyclic 3′,5′-guanosine monophosphate (cGMP).

Helicases: Enzymes that use the energy derived from the hydrolysis of nucleoside triphosphates to unwind the double-stranded helical structure of nucleic acids.

His-tagged: An epitope tag (histidine tag) based on a short stretch (~6) of histidine residues added to either the N- or C-terminus of a protein, sometimes with an added region susceptible to endopeptidase cleavage to allow stripping of the tag from the recombinant protein.

Interglacial period: A geological period of increased temperatures that occurs between major glacial phases, and which can last between 10 000 to 20 000 years.

Isopropyl-β,D-thiogalactopyranoside: A molecular biology reagent or a molecular mimic of allolactose, a lactose metabolite that triggers transcription of the *lac* operon, and it is therefore, used to induce protein expression where the gene is under the control of the *lac* operator.

Luria-Bertani broth: A nutrient growth medium typically used for the maintenance and propagation of *E. coli*.

Mass spectrometry: A biochemical method used to detect biological molecules according to their quantities and molecular weights.

Protein purification: The process of isolating a protein of interest from its natural expression environment.

Protein expression: A process by which proteins are synthesized, modified and regulated in living organisms also referring to the laboratory techniques required to manufacture proteins.

Proteome: A collection of cellular proteins, whose expression levels are co-regulated by a single and specific signaling molecule.

Reverse transcriptase-polymerase chain reaction (RT-PCR): A molecular method used to amplify and convert a short RNA segment into a DNA product termed copy DNA (cDNA) using an RNA-dependent DNA polymerase enzyme.

Second messenger: A biological molecule capable of transmitting external cellular signals within the cell for the development of appropriate cellular responses through regulated gene expressional and metabolic events.

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE): A technique used in Molecular Biology to separate different protein molecules according to their sizes and migration levels in a polyacrylamide gel system subjected to a strong electrical field.

LIST OF ABBREVIATIONS

2YT	**:**	Double concentrated yeast-tryptone medium
AC	**:**	Adenylate cyclase
ANOVA	**:**	Analysis of variance
ANTH	**:**	AP180 N-terminal homology
AP-2	**:**	Activator protein 2
ATP	**:**	Adenosine 5′-triphosphate
cAMP	**:**	Cyclic 3′,5′-adenosine monophosphate
CAP	**:**	Clathrin assembly protein
CCVs	**:**	Clathrin-coated vesicles
cGMP	**:**	Cyclic 3′,5′-guanosine monophosphate
CME	**:**	Clathrin-mediated endocytosis
ENTH	**:**	Epsin N-terminal homology
GC	**:**	Guanylate cyclase
GTP	**:**	Guanosine triphosphate
IBMX	**:**	3-Isobutyl-1-methylxanthine
IPTG	**:**	Isopropyl-β-D-1-thiogalactopyranoside
I-TASSER	**:**	Iterative Threading Assembly Refinement
LB	**:**	Luria-Bertani

NBS-LRR	**:**	Nucleotide-binding site-leucine-rich repeat
Ni-NTA	**:**	Nickel-nitrilotriacetic acid
OD	**:**	Optical density
PPMV	**:**	Parts per million by volume
PDE	**:**	Phosphodiesterase
PPM	**:**	Parts per million
RPM	**:**	Revolutions per minute
SDS-PAGE	**:**	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
TAIR	**:**	The Arabidopsis Information Resource
UIM	**:**	Ubiquitin-interacting motifs
UV	**:**	Ultraviolet
YT	**:**	Yeast-tryptone

LIST OF FIGURES

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 eukaryotic ACs and their product cAMP have been firmly established as important signalling molecules with critical roles in cellular signal transduction pathways. Therefore, and in an effort to augment information currently known about ACs in higher plants, this study targeted an Epsin N-terminus homology protein from *Arabidopsis thaliana* (AtENTH), whose gene was initially annotated as a probable AC bioinformatically and then recently confirmed as a *bona fide* AC practically. The study recombinantly expressed the AtENTH protein followed by a comprehensive characterization of its enzymatic AC activity biochemically using the enzyme immunoassaying technique and structurally through structural modelling, simulations and molecular docking techniques. Findings from the study, technically confirmed the AtENTH protein as a multi-domain multi-functional higher plant AC, whose catalytic activity has a very strong specificity for ATP as its sole substrate compared to the other known organic triphosphates, and binding to it using the E residue located at position 2 within its catalytic AC center. The AtENTH ultimately catalyzes the conversion of this preferred ATP substrate into cAMP using the K residue located at position 14 of its catalytic center.

Keywords: *Arabidopsis thaliana*, Epsin N-terminal homology protein, adenylate cyclase, cAMP, structural modelling, molecular docking.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Over the years, people around the world have been using a large number of fossil fuels (e.g. coal, oil, and natural gas) for energy needs in their homes, factories and vehicle industries. According to NASA(2017), fossil fuels release carbon dioxide (CO_2), a heat-trapping gas, into the atmosphere, which is the main reason why the climate is changing. Heat-trapping gases are also called greenhouse gases. They exist naturally in the atmosphere, where they help keep the Earth warm enough for plants and animals to live. People are adding extra greenhouse gases to the atmosphere every day; these extra gases are causing the Earth to get warmer, setting off all sorts of other changes around the world, and also affecting the atmosphere. Weather conditions and patterns that occur around an area in a specific period of time, are usually influenced by factors such as latitude, elevation, topography, ocean currents, nearby water, and vegetation, and all these are considered abiotic factors (UCS, 2017).

According to Terashima et al. (2014), the atmospheric CO_2 concentrations have always fluctuated between 180–200 and 250–280 parts per million by volume (ppmv) during the different glacial and interglacial periods respectively. At the beginning of the current interglacial period, CO_2 levels remained around 280 ppmv for more than 10 000 years. However, since the industrial revolution of the 1800s, CO_2 concentration levels have been increasing with annual fluctuations peaking during winters and the most recently, exceeding 400 ppmv. This then brings about the possibilities of the concentrations rates reaching twice the rates prior to the 1800s by the end of this century.

Climatologists have studied and reported climate changes across the globe, leading to global warming and greenhouse effect awareness. Some of the changes that have been observed include ocean acidification and global temperature rise in the world. The changes in climate directly and indirectly affect humans, plants, and animals. Bierwirth (2018) states that human beings and animals are able to deal with elevated levels of CO_2 in the short term due to their various compensation mechanisms in the body. For example, the increased CO_2 levels lead to acidity in the blood, which then triggers compensatory mechanisms, including pH buffering systems in the blood, increased breathing to reduce excess CO_2 in the bloodstream, increased excretion of acid by the kidneys to help restore the acid-base balance and nervous system stimulation to counteract the direct effects of pH changes on heart contractility and vasodilation. On the other hand, when living conditions are unfavourable due to the indirect effects of the rise in CO_2 concentration levels, human beings and animals can relocate to other places to find food and other means of protecting themselves against any kind of harm caused by the climate change (Bierwirth et al., 2018).

Unlike humans and animals, plants however, are sessile and thus cannot keep pace with the rapid changes in the temperature and increases in CO_2 by means of adaptation alone. This is because most plants are adapted to an atmospheric CO_2 of below 300 ppmv. Due to this reason, various studies have been conducted to date, on both the direct and indirect effects of the increase in CO_2 on plant growth and crop yield, and have shown that high CO_2 does not necessarily enhance plant growth as several plant cellular functions are down-regulated by high CO_2 . As a result, the precise direct effects of CO_2 on plant growth and performance are largely unknown and thus more research is required to unravel and understand the detrimental effects of high CO_2 (Terashima et al., 2014).

Ideally, the sessile nature of plants makes them vulnerable to both biotic and abiotic stress factors. Climate changes form part of the abiotic stress factors whilst living organisms such as

humans, animals, pathogens, etc. form part of the biotic stress factors. According to Pandey et al. (2017), due to issues such as global warming and the potential climate abnormalities associated with it, crops typically encounter an increased number of abiotic and biotic stress combinations, which severely affect their growth and yield. Plants are essential for life existence on earth; they supply food for nearly all terrestrial organisms, including human beings. Plants also supply us with industrial materials, raw goods and clothes but most importantly, they maintain the oxygen balance in the atmosphere (Pandey et al., 2017).

Climate change has also caused several concerns due to its impact on food production, agricultural livelihood and food security globally. Climate has been recorded as a major constraint of agricultural development as it causes various negative impacts on crop growth and production. Global warming leads to the concurrence of several abiotic and biotic stress factors causing destructive consequences on crops. Drought, high and low temperatures, and salinity influence the occurrence and spread of pathogens, insects, and weeds. These abiotic factors alter plant-pest interactions by enhancing host susceptibility to pathogenic organisms, insects and other organisms as well as reducing the host's competitive ability against weeds. In summary, the abiotic factors affect plant physiology and defense responses, making them highly susceptible (Pandey et al., 2017).

As plants become more susceptible to various stress factors, there is a reduced level of crop production, which somewhat has devastating effects onto the world food supplies, causing serious problems such as malnutrition even in developed countries (Pandey et al., 2017). With droughts and floods persistently occurring across the globe, people are forced to move to neighboring countries and abroad to find better and safer places for their families, causing a high demand for food due to the growing populations and ultimately causing threats of conflicts, wars, famine, food security, and economic declines. Basically, the continued rise in CO_2 and increase in temperatures will leave half the world population unable to grow its food, taking care

of its livestock and the overall human health and well-being being drastically affected. Since most organisms on earth depend on plants for several means, scientists of various disciplines like botany, agronomy, and plant biotechnology have started carrying out studies on ways to possibly improve plant evolution and production.

Scientists realized that with the ever-changing climatic conditions and plants being affected like already outlined, it meant that all life on our mother planet, the earth, may seize to exist and it was now time to broadly cross-examine and understand plant physiology and development (Pandey et al., 2017). By identifying the agriculturally important morpho-physiological traits that can be utilized to identify genotypes with combined stress tolerance, then stress tolerant plants may be developed and sustained. Scientists knew that a thorough investigation of plants, will have a positive impact on crop production. Instead of studying actual high plant species that are more complex, model plant species are ideally used, and those include *Zea mays L.* (the common maize), *Oryza sativa* (the usual rice), *Arabidopsis thaliana* (the mouse-ear cress), etc. Most model plants are angiosperms, which are closely related to all high plants; and therefore, complete sequencing of their genomes and discovering functions of their proteins will offer much needed information about the roles of proteins in higher plants, and particularly agricultural crops (Koornneef and Meinke, 2010).

The National Science Foundation (2017) stated that although *A. thaliana* is a non-commercial member of the mustard family, it is the most used model organism in plant research because of its affordability, it grows much faster and its response to stress and diseases in much the same way as crop plants but most importantly, it has a small genome. Its small genome simplifies and facilitates genetic analysis. These are some of the reasons why this research, reported herein this species. Specifically, this study focused on the adenylyl cyclase (AC) activity of an Epsin N-terminal homology protein (ENTH) or PICALM9A encoded by the *At1g25240* gene in Arabidopsis and annotated by the National Centre for Biotechnology Information (NCBI).

According to The Arabidopsis Information Resource (2018), ENTH protein is specifically localized in the Golgi apparatus, clathrin-coated pit, vesicles, and mitochondrion (TAIR, 2018). This protein is also known as the putative clathrin assembly protein that is involved in clathrin coat assembly and endocytosis. This protein plays a role in 1-phosphatidylinositol binding, clathrin binding, and phospholipid binding. ACs are enzymes that synthesize cyclic adenosine monophosphate (cAMP) from adenosine triphosphate (ATP), which increases the intracellular levels of cAMP and thus cell signalling and transduction systems. At the moment, the At1g25240 gene or its AtENTH protein has been confirmed as a functional AC molecule (unpublished data) but its specific function has not yet been fully elucidated. Therefore, this study was aimed at functionally characterizing the enzymatic AC activity of an AtENTH protein from Arabidopsis thaliana using biochemical techniques such as *in vitro* enzyme immunoassaying system and structural techniques for example structural modelling, simulations and molecular docking.

1.2 Literature Review

1.2.1 Adenylate Cyclase and Cyclic Adenosine Monophosphate

The cyclic nucleotide 3'-5'-cyclic adenosine monophosphate (cAMP) is the original member of the family of second messengers discovered by Dr. Earl W Sutherland around 1956 during his studies of the mechanism of hormone action, specifically glycogen phosphorylase. In 1971, Dr. Sutherland was awarded the Nobel Prize for this work, which ultimately proved to be the first of five Nobel Prizes recognizing research on cAMP. Ideally, cAMP has been firmly established as a universal regulator of cellular functions, an important signaling molecule and second messenger in all organisms, including animals and lower eukaryotes. Biological processes

mediated by this second messenger include memory, metabolism, gene regulation, and immune function. The cAMP also plays a role in polarized growth of pollen tubes (Serezani et al., 2008).

Naturally, the intracellular levels of cAMP are regulated by the balance between the activities of two enzymes, AC and cyclic nucleotide phosphodiesterase (PDE) as is shown in Figure 1.1(below).Different isoforms of these enzymes are encoded by a large number of genes, which differ in their expression patterns and mechanisms of regulation, generating cell-type and stimulus-specific responses. cAMP generated as a consequence of AC activation can activate several effectors, the most well-studied being the cAMP-dependent protein kinase (PKA), which is a central cascade that transmits extracellular stimuli and governs cell responses through the second messenger cAMP (Sassone-Corsi, 2012, Pierce et al., 2002). On the other hand, cAMP levels decrease in the presence of PDE. Surprisingly and even though cAMP has increasingly been recognized as an important signaling molecule in higher plants, its generating enzymes, ACs have largely remained somewhat elusive and a matter of huge controversy (Gehring, 2010).

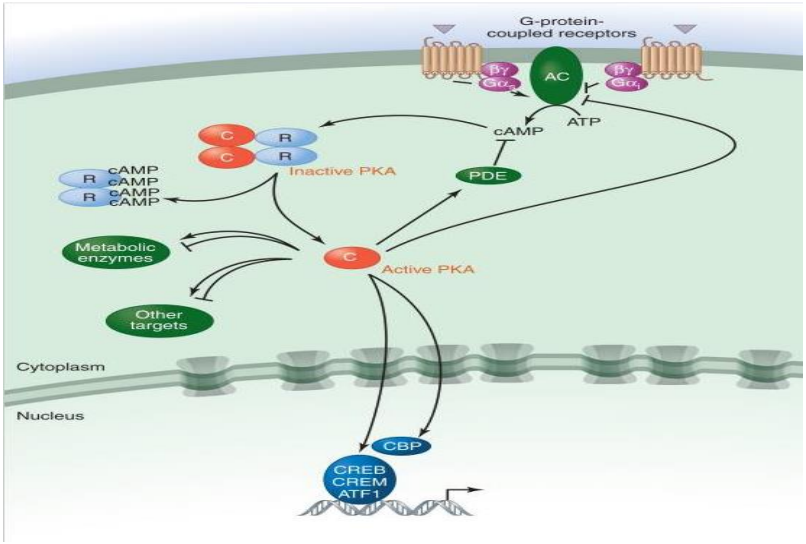

Figure 1.1: The cAMP-dependent protein kinase pathway. ACs are activated downstream from the G-protein-coupled receptors (GPCRs) such as the β-adrenoceptor by interactions with the α-subunit of the G_s protein (αs), which is released from the heterotrimeric αβγ G-protein complexes following binding of

agonist ligands to GPCRs (e.g., epinephrine in the case of β adrenoceptors). The generated cAMP then activates the cAMP-dependent protein kinase (PKA) (Sassone-Corsi, 2012, Pierce et al., 2002).

According to Gehring (2010), cAMP levels in plants are low compared to those found in animals. For instance, it was reported that cAMP levels in plants are typically <20 pmol/g fresh weight (e.g., ≤12 pmol/g fresh weight in the ryegrass endosperm cell cultures and <16 pmol/g fresh weight in the *Lilium longiflorum* pistils) whereas animal levels are typically >250 pmol/g wet weight. Incidentally, low levels of yet another cyclic nucleotide, cGMP, were also reported in plants, where for example, specific responses to a virulent pathogen increased the cytosolic cGMP from <0.4 pmol/g fresh weight to 1 pmol/g fresh weight. It is, however, noteworthy that 0.5 pmol/g fresh weight of a cyclic nucleotide corresponds to a cytosolic concentration of approximately 500 pM, and that signalling in the picomolar range is not uncommon in plants. Nonetheless and despite the low, seemingly un-physiological and certainly not animal-like levels of cAMP in plants, the notion that plants also have a functional cAMP-dependent signal system remained alive, mainly because both cell permeant 8-BrcAMP and stimulation of *albeit* unknown ACs with forskolin, could elicit concentration and time-dependent biological responses such as increases in Ca^{2+} influx across the plasma membrane (Gehring, 2010).

Based on the confirmation that cAMP plays important roles in signaling in higher plants, various scientists, including Professor Christoph Gehring's attempted to identify probable AC molecules, particularly in *A. thaliana*. From that work, a total of 14 putative protein candidates were identified including the ENTH (At1g25240) reported in this study. Notably, at that time, the only annotated and experimentally confirmed AC in higher plants was a *Zea mays* pollen protein with a role in polarized growth of pollen tubes (Moutinho et al., 2001). The Arabidopsis orthologue of this protein (At3g14460) is annotated as disease resistance protein belonging to the nucleotide-binding site-leucine-rich repeat (NBS-LRR) family, used for pathogen sensing. It has a role in defense responses and apoptosis downstream signaling, which is enabled by cAMP.

Considering that cyclic nucleotides have important and diverse roles in plant signaling via cyclic nucleotide-responsive protein kinases, binding proteins, and ion-gated channels, it raised suspicions that it was unlikely that a single AC or Guanylate Cyclase (GC) can account for all the known and reported cAMP- and cGMP-dependent processes in higher plants. In line with his hypothesis was the fact that a number of Arabidopsis molecules with different domain organizations and experimentally confirmed GC activity had, by that time, been reported (Gehring, 2010).

Previously, within the Arabidopsis genome, functionally tested GCs had been identified with a 14 amino acid long search term deduced from an alignment of conserved and functionally assigned amino acids (Figure 1.2) in the catalytic center of annotated GCs from lower and higher eukaryotes. A similar approach was then used for the discovery of novel ACs in *A. thaliana* (Gehring, 2010).

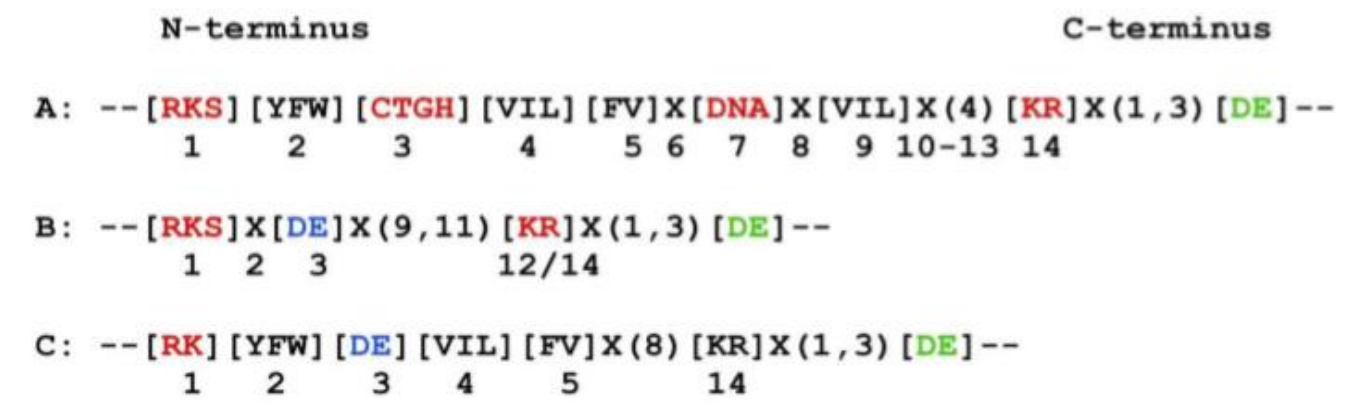

Figure 1.2: Catalytic center motifs of nucleotide cyclases. (A) Centre motif of experimentally tested GCs in plants. The residue (red) in position 1 does the hydrogen bonding with the guanine, the amino acid in position 3 confers substrate specificity and the residue in position 14 stabilizes the transition (GTP/cGMP). The Mg2+/Mn2+-binding site is C-terminal (green). In the derived motifs (B and C) specific for ACs, position 3 (blue) has been substituted to [DE] to allow for ATP binding (Gehring, 2010).

1.2.2 The Model Plant

A. thaliana is a small dicotyledonous species that is universally recognized as a model research organism in Plant Biology. This plant is a member of the mustard or *Brassicaceae* family, which includes cultivated species such as cabbage and radish (TAIR, 2017). According to Wixon (2001), *A. thaliana* was discovered by Johannes Thal in the Harz Mountains, in the sixteenth century. Arabidopsis was originally adopted as a model organism because of its usefulness in genetic experiments. Important features included its short generation time, small genome, small size that limited the requirement for growth facilities, and prolific seed production through self-pollination. According to Song et al. (2012), the Arabidopsis genome encodes 22 proteins with an ANTH/ENTH domain. Although mammalian Epsin1 is the best characterized protein of this class, the first isolated protein with this domain came from plants two years before the domain was biochemically described and termed the Epsin N-terminal homology (ENTH) (Zouhar and Sauer, 2014).

1.2.3 The ENTH Domain

Plants are eukaryotic organisms; one feature that defines them is the compartmentalization of their cytoplasm into different membrane-associated organelles. In order to maintain these compartments, cells have developed mechanisms to ensure that specific proteins are targeted to particular organelles. Sub-cellular compartmentalization became an essential feature in these organisms, allowing the correct interrelations of certain intracellular components and enabling reactions to occur efficiently and orderly. To connect all compartments, proteins and lipids are trapped into transport vesicles, which are made of “coatomers”, specific for a particular compartment. Clathrin-coated vesicles (CCVs) are transport vesicles that have been extensively

characterized in a wide variety of eukaryotic cells and mediate the variety of cargo molecules to the endosomal/ lysosomal compartments. These vesicles are covered by a layer that consists of scaffold proteins, clathrin, and several oligomeric and monomeric adapter proteins, of which these proteins bind clathrin and also recognize specific classification signals that are present in the cytosolic domains of transmembrane proteins, producing accumulation of these proteins in the CCVs (Feliziani and Touz, 2017). The accessory proteins that are involved in CCV formation include Epsin1 and AP180, which act as adaptors of clathrin.

Epsin1 has a unique domain at the N-terminus that is shared by a group of proteins termed the Epsin N-terminal homology domain-containing proteins. Whilst AP180 has a domain that displays a high amino acid sequence similarity to the ENTH domain, however, the domain is termed the AP180 N-terminal homology. These protein domains have the ability to bind inositol phospholipids in the membrane, most notably phosphatidylinositol 4, 5-biphosphate [PtdLns (4, 5)P2], and that is why they are often grouped together as the A/ENTH domain (Song et al., 2012). ENTH and ANTH phospholipid-binding proteins are supposed to play a crucial role in the initiation of clathrin-coated pits at the plasma membrane because they serve as bridges/adaptors between phosphatidylinositol 4, 5-bisphosphate (P1[4,5]P2) and several components of the clathrin endocytic machinery (CEM) (Holstein and Oliviusson, 2005). Although these two domains are often grouped as A/ENTH domain, they have different structural features and lipid-binding properties (Song et al., 2012).

Furthermore, ENTH has an N-terminal structured region in front of an α-helix1 that is converted to an α-helix upon binding to the PtdLns(4,5)P_2 in membranes and is thus referred to as H_0 (α0), which is essential for membrane-binding and de-forming activities of the ENTH domain. Basically, what happens is that the H_0 helix penetrates the outer leaflet of the bilayer and forms the α-helix in both the membrane vesicles and their preformed membrane tubes. The ANTH however, does not have the N-terminal H_0, it consists of an α-helix1, an α-helix2 and a loop

between them to bind the PtdLns(4,5)P_2. This ANTH domain also does not have the membrane-deforming activity. It is solely involved in endocytosis in both animal and plant cells.

Epsin related proteins are involved in two different trafficking pathways: endocytosis and lysosomal/vacuolar trafficking. The majority of Epsin-related proteins are involved in endocytosis whilst very few are involved in lysosomal/vacuolar trafficking. Many CCV mediated endocytosis involved proteins have been shown to accumulate in the cell plate during cell division (Song et al., 2012). During the last stage of mitosis known as cytokinesis, the cytoplasm of a dividing cell is portioned to daughter cells. The plant cell employs a more complicated mechanism for cytokinesis, where a phragmoplast is formed from the remains of a spindle microtubules and a new cell wall is generated at the mid-plane of the phragmoplast, thereby separating the cytoplasm. Secretory vesicles originating from the trans-Golgi network (TGN) are delivered to the division plane and fuse to each other via a homotypic fusion to form the cell plate. The fused vesicles at the growing cell plate are then further processed via intermediate structures, such as the tubule-vesicular networks to form some planar fenestrated sheets.

Many vesicle trafficking-related proteins are involved in cell plate formation. These proteins include the cytokinesis-specific t-SNARE KNOLLE, a syntaxin-binding protein, KEULE, AtSNAP33, dynamin-related proteins, ESCRT (for endosomal sorting complex required for transport) components, and proteins of the exocyst complex. These observations suggest that both exocytosis and endocytosis play critical roles in cell plate formation. The clathrin-mediated vesicle budding is not restricted to the plasma membrane only, where it serves to take up nutrients, signaling receptors and other proteins from the cell surface via endocytosis but it is also involved in the transport of different cargo proteins from TGN to endosomes and lysosomes. This proves that E/ANTH domains are indeed universal tethering components of the clathrin-mediated vesicle budding machinery.

Clathrin-mediated endocytosis (CME) is one of the essential cellular processes that require the coordinated action of multiple membrane proteins functioning together. The ENTH is a major player in clathrin-mediated endocytosis (Gleisner et al., 2016). The Epsin plays an important role in inducing membrane curvature and recruiting accessory proteins in the early stage of CME. The Epsin contains multiple conserved binding motifs that can interact with several accessory proteins associated with CME, for example, Adaptor Protein 2 (AP2), Epidermal Growth Factor Receptor Pathway Substrate 15 (EPS15), clathrin and ubiquitinated proteins (Lai et al., 2012). CME is a process that is driven by a chain of remodeling events and an elaborate set of proteins acting in an orchestrated manner, whereby an almost flat patch of the plasma membrane is transformed into a closed, cargo containing vesicle. It is known that the ENTH of Epsin1 binds specifically to the receptor lipid PtdLns (4,5)P2 and endocytosis is a fundamental and complex process supporting many essential cellular functions in eukaryotes, including nutrient uptake, receptor signaling, down-regulation and defense against pathogens. This process consists of membrane invagination, budding and formation of transport vesicles directed to an intracellular membrane-bound organelle specialized in receiving internalized materials, known as the early endosome (Law et al., 2012). Plant endocytosis impacts many critical events during the life cycle of a plant, including embryo patterning, lateral organ differentiation, root hair formation, hormone signal transduction, and defense responses. Several studies have clearly demonstrated the crucial role of endosomes in several plant processes, including cell fate specification, abscisic acid and auxin signaling, tropic responses, and pathogen defense.

ENTH proteins include proteins that contain an Epsin homolog (De Camilli et al., 2002). The ENTH domain forms a compact solenoid of eight alpha-helices, comprising roughly 130 to 150 amino acids. Epsins are proteins capable of fulfilling different roles at budding in endocytic sites. According to Sen et al. (2012), they bear UIMs (ubiquitin-interacting motif), that act as endocytic adaptors by directly binding to ubiquitinated cargo. Epsins also interact with other elements of the endocytic machinery, including the hubs of the endocytic network, i.e., AP2 and

clathrin. This means that they are accessory proteins by consolidating and regulating the endocytic network. Endocytic sites are composed of patches of actin-interacting with a complex series of endocytic regulatory proteins. Several of these endocytic proteins contain independently folded protein modules, such as the ENTH domain, that interacts with phospholipids, including inositol phospholipids, to induce and stabilize membrane curvature. The ENTH domain has a compact α-helical structure and binds to phosphatidylinositol 4,5-bisphosphate [PtdLns (4,5)P2], which is enriched on the plasma membrane. When the ENTH domain binds to PtdIns (4,5)P2-rich membranes, unstructured residues at the N-terminus form a new α –helix, called α0, that inserts into the inner leaflet of the bilayer, inducing curvature. The creation of membrane curvature has been considered a major function for the ENTH domain (Ritter and McPherson, 2006).

The gene of focus for this study is the At1g25240 gene that encodes an ENTH protein (Figure 1.3). This gene is of the ENTH/VHS/GAT family proteins that are involved in clathrin coat assembly and endocytosis. Clathrin coat assembly is the process that results in the assembly of clathrin triskelia into the ordered structure known as a clathrin cage, whilst endocytosis is a process by which extracellular materials are taken up into a cell by invagination of the plasma membrane to form vesicles enclosing these materials. The gene of interest is located in the following organelles of the living cell; the Golgi apparatus, the clathrin-coated pit, the clathrin-coated vesicle, and mitochondrion. According to the TAIR (2018) website, this gene is known to play a role in 1-phosphatidylinositol binding, clathrin binding and phospholipid binding besides its recent discovery as a functional AC. However, despite its established AC activity (as a result of it harbouring an annotated AC catalytic center (Figure 1.3)), such a function has not yet been fully characterized, which was the specific focus of this study.

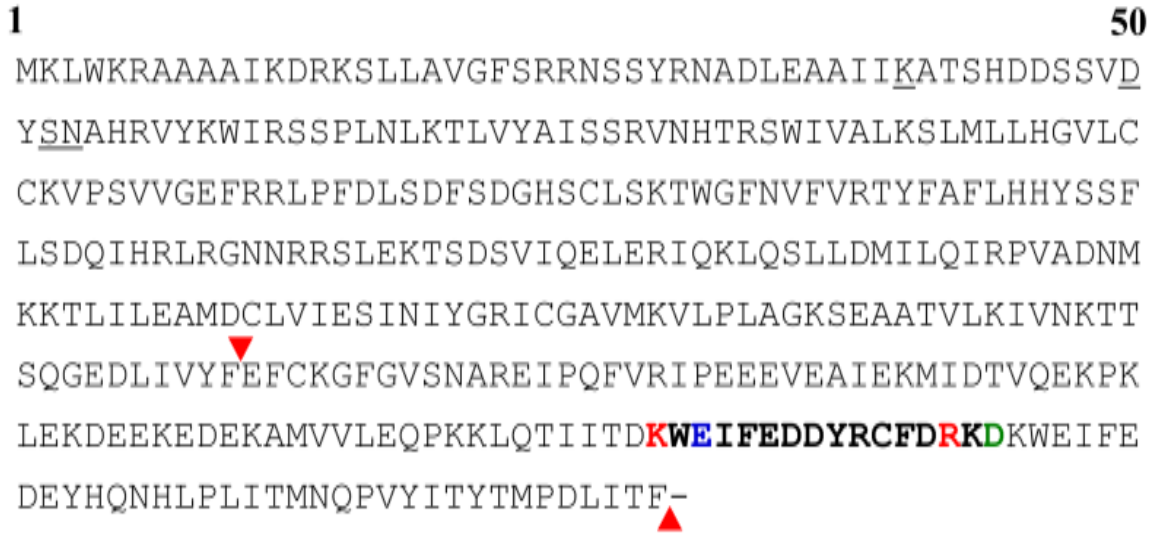

Figure 1.3: The complete amino acid sequence of Arabidopsis ENTH with the AC catalytic center highlighted in bold and the 119 amino acid fragment confirmed for AC activity indicated within the inverted red triangles. The underlined amino acids mark an N-terminal phosphatidylinositol 4,5-bisphosphate (also referred to as PtdIns(4,5)P2, PIP2 or PI(4,5)P2) binding site.

1.3 Problem Statement

Due to the extremely changing climatic conditions, plants have become more vulnerable to abiotic and biotic stress factors. Although plants have been studied for years, information on their adaptation and defense mechanisms to stress and particularly enzymes involved in such processes have not been fully elaborated. Adenylate cyclases (ACs) are one group of the enzymes that are centrally involved in the core cellular signaling of plants, however, both their structures and biochemical properties remain partially characterized. This study was thus set to elaborate on the structural and enzymatic properties of a fusion ENTH protein from Arabidopsis (AtENTH) that has recently been confirmed as an AC so that its potential roles in cell communication and signal transduction systems could be better understood.

1.4 Research Aim

The aim of this study was to characterize the enzymatic AC activity of an Epsin N-terminal homolog protein from *A. thaliana* using biochemical and structural techniques.

1.5 Research Objectives

1. To fully express the putative fusion AtENTH protein as a recombinant product.
2. To affinity purify the fully expressed recombinant fusion AtENTH protein under native non-denaturing conditions.
3. To validate the AC activity of the purified recombinant fusion AtENTH protein through complementation testing.
4. To determine the substrate specificity of the purified recombinant fusion AtENTH protein via enzyme-immunoassay.
5. To determine the functional and regulatory elements of the AtENTH protein via structural modelling.
6. To characterize the functional properties of the AtENTH protein through molecular docking.
7. To infer the probable mechanisms by which AtENTH participates in stress response and adaptation mechanisms.

1.6 Significance of the Study

Upon the successful completion of this study, there shall be more clarity on the AC activity of the annotated ENTH protein and its probable functional roles in the model plant *Arabidopsis*

thaliana and other closely related higher plants. Using such generated information, efforts can be made to genetically manipulate agricultural important crops for yield improvement and ultimately, food security in the country.

CHAPTER 2

RESEARCH METHODOLOGY

2.1 Recombinant Expression and Affinity Purification of the AtENTH Protein

2.1.1 Recombinant Expression of the AtENTH Protein

An *E. cloni* EXPRESS BL21 (DE3) pLysS cell colony harbouring the pTrcHis2-TOPO:AtENTH expression construct (Plant Biotechnology Lab, Department of Botany, North-West University) was used to inoculate 10 ml of the 2YT media (16 g tryptone, 10 g yeast extract, 5 g NaCl and 4 g glucose per L, pH 7.0) supplemented with 0.5% (w/v) glucose, 34 μg/ml chloramphenicol and 100 μg/ml ampicillin in a 15 ml falcon tube. The Falcon tube was incubated overnight at 37ºC, with moderate shaking in an orbital shaker at 200 rpm. A fraction (200 μl) of the overnight culture were then used to inoculate 25 ml of fresh 2YT media containing 34 μg/ml chloramphenicol, 100 μg/ml ampicillin and 0.5% (w/v) glucose. The culture was then incubated at 37ºC, with moderate shaking at 200 rpm until an OD_{600} of 0.5-0.6 was reached as measured by the Hekios Spectrophotometer (Merck, Gauteng, RSA). Immediately at that point, the culture was split into two sets of 5 ml (control) and 20 ml (experimental) volumes respectively. The bigger culture was induced to express the intended ENTH recombinant by adding 1 mM of isopropyl-β,D-thiogalactopyranoside (IPTG) (Sigma-Aldrich Corp., Missouri, USA) while the control culture was left un-induced. The split cultures as then agitated in an incubator (200 rpm) at 37ºC for 3 hours. After incubation, 10 μl portions of each culture were collected for analysis by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) while the rest of the other cultures were centrifuged at 8 000 x g for 5 minutes to pellet out the cells and stored for the other forthcoming downstream analysis.

2.1.2 Affinity Purification of the Recombinant AtENTH Protein

The pelleted out bacterial cells carrying the expressed recombinant ENTH protein was then purified under native non-denaturing conditions using a His-Select nickel-nitrilotriacetic acid (Ni-NTA) affinity matrix, according to the manufacturer's protocol (Catalog # P6611; Sigma-Aldrich Inc., Missouri, USA). The pelleted induced cells were re-suspended in 5 ml phosphate saline (PBS) buffer (140 mM NaCl, 3 mM KCl, 4 mM $Na_2HPO_4.2H_2O$ and 1.5 mM KH_2PO_4) supplemented with 10 mM imidazole and 0.5 mM phenylmethylsulfonyl fluoride (PMSF), and sonicated for 6 minutes (10 seconds pulsing and 10 seconds chilling cycles) to rapture and release the cell contents. After cell rupturing, the crude lysate was then clarified into a cleared lysate through centrifugation at 10 000 x g for 20 minutes and the cleared lysate was kept at 4˚C for the subsequent purification steps. A portion of the cleared lysate (about 20 µl) was collected for analysis by SDS-PAGE.

Furthermore about 1 ml of the 50% (in 70% ethanol) Ni-NTA slurry matrix (Lot # MG159557; Thermo Scientific Inc., Rockford, USA) were washed three times with 6 ml of sterile distilled water on a rotary mixer (Bio-Rad Laboratories., California, USA) for 5 minutes by adding 2 ml of water followed by a brief mixing on the mixer and then draining out the wash water. The washed Ni-NTA beads were then equilibrated for 5 minutes in 5 ml of the PBS buffer. After equilibration, the Ni-NTA beads were then mixed with the generated AtENTH cleared lysate was rotated for 1 hour on an adjustable Bench Revolver (Labnet International Inc., New Jersey, USA) at 30 rpm at 4˚C. This step was undertaken so that the AtENTH protein in the cleared lysate could specifically bind to the bead matrix using its Histidine-tagged segment. After binding, the mixture was then sedimented at 2 000 x *g* for 5 minutes before the supernatant (flow-through) was removed and a portion (about 20 µl) saved for SDS-PAGE analysis.

After successfully binding the AtENTH recombinant, the unbound proteins were all thoroughly washed off the Ni-NTA bead matrix three times with 15 ml of a wash buffer (100 mM NaH_2PO_4,

10 mM Tris-HCl; pH: 8.0, 500 mM NaCl, 20 mM β-mercaptoethanol, 7.5% (v/v) glycerol, and 10 mM imidazole) whereby on each wash, the beads were completely re-suspended into 5 ml wash buffer followed by a total removal of the buffer. Portions of each wash together with part of the washed bound beads (about 20 µl) were then all saved for SDS-PAGE analysis.

2.1.3 Elution of the Recombinant AtENTH Protein

The bound and fully purified AtENTH recombinant protein was then eluted off the Ni-NTA bead matrix through the addition of 2 ml elution buffer (200 mM NaCl, 50 mM Tris-Cl (pH: 8.0), 250 mM imidazole, 0.5 mM PMSF, and 20% (v/v) glycerol) by mixing the buffer and matrix together and allowing the mixture to settle for 10 minutes. The resultant supernatant containing the eluted AtENTH protein was then collected and stored at 4ºC for downstream use while a portion (about 20 µl) was also collected for SDS-PAGE analysis.

2.1.4 Concentration and Desalting of the Recombinant AtENTH Protein

The eluted and purified recombinant AtENTH protein was freed from the buffering salts and excess water by pouring the 2 ml eluent into the upper chamber of a Spin-X UF de-salting and concentrating device (Corning Corp., New York, USA). The device was then centrifuged at 2 540 x g at 4ºC for 4 hours or until the final volume had gone down to 100 µl. The concentrated and desalted protein fraction was then removed from the device and transferred to a new Eppendorf tube. Protein concentration was then determined using a 2000 Nanodrop Spectrophotometer (Thermo Scientific Inc., Califonia, USA) and the recovered protein sample stored at -20ºC for further use.

2.2 Activity Assaying

2.2.1 Validation of the AC Activity of the Recombinant AtENTH Protein

Before the AC activity of the eluted and purified recombinant ENTH protein was assessed and further elucidated, its exact inherent AC activity was firstly confirmed through a complementation testing. In this process, some chemically competent mutant *E. coli* host cells, the *cya*A SP850 strain (Coli Genetic Stock Center, Yale University, Connecticut, USA) were collected (Plant Biotechnology Lab) and divided into two portions. The first portion was transformed with the pTrcHis2-TOPO:AtENTH expression construct (section 2.1.1) while the other portion was left un-transformed (control). Alongside this, a MacConkey agar plate supplemented with 15 μg/ml kanamycin and 0.1 mM IPTG (Sigma-Aldrich Corp., Missouri, USA) was also prepared and then sub-divided into 3 segments using a permanent marker. The first segment was left un-streaked (no *cya*A cells), the second segment was streaked with the non-transformed *cya*A mutant cells while the last segment was streaked with the *cya*A mutant cells transformed with the pTrcHis2-TOPO:AtENTH expression construct. The plates were inverted and incubated at 37ºC for 40 hours. After the incubation period, all segments were then visually analyzed for the various phenotypic characteristics. In this case, a deep red or purplish color on the transformed mutant cells would indicate positive AC activity for the cloned and recombinantly expressed AtENTH protein while a colourless or yellowish colour on the same cells would indicate a non-AC activity on the cloned and expressed recombinant protein.

2.2.2 Determination of the Substrate Specificity of the Recombinant AtENTH Protein

The *in vitro* AC activity of the purified AtENTH recombinant protein was characterized by assessing its most preferred substrate among ATP, GTP, CTP, and TTP in a Tris-buffered reaction system. Here, 10 μg portions of the recombinant AtENTH protein were separately

mixed with either 1 mM ATP or GTP or CTP or TTP, plus 2 mM 3-isobutyl-1-methylxanthine (IBMX, Sigma-Aldrich Corp.,) to inhibit phosphodiesterases (PDEs) and 5 mM Mn^{2+} in a final volume of 200 µl of the Tris-HCl buffer at pH 8.0. Any residual cAMP contents resulting from the non-protein AC activity were also monitored in tubes containing the incubation medium with no protein. All prepared reaction mixtures were then incubated at 25ºC for 20 minutes and terminated through the addition of 10 mM of ethylene di-amine tetra-acetic acid (EDTA), with boiling for 3 minutes. The reaction mixture was transferred to ice for 2 minutes, and centrifuged at 9 200 x g for 3 minutes for clarification. The resulting supernatants were then assayed for cAMP content using the cAMP-linked enzyme immunoassaying kit (Catalogue # CA201) following the acetylation version of its protocol as described by the manufacturer's manual (Sigma-Aldrich Corp., Missouri, USA). All results and outcomes were then subjected to the one-way statistical analysis of variance (ANOVA) in triplicate sets.

2.2.3 Structural Analysis of the Recombinant AtENTH Protein

In order to further characterize the purified recombinant AtENTH protein, its functional and regulatory elements were assessed and determined via structural modelling. In this regard, the full-length amino acid sequence of the AtENTH protein was obtained from the PROSITE database located within the Expert Protein Analysis System (ExPASy) proteomics server (https://www.expasy.ch/) and submitted to the iterative threading assembly refinement (I-TASSER) server available on-line at: http://zhanglab.ccmb.med.umich.edu/I-TASSER/ and various ENTH models generated by the same I-TASSER method (Zhang, 2008). The model with the highest quality, based on its C-score, was then downloaded from the server and its structural features and properties visualized and analyzed (through simulations) by the UCSF Chimera (v.1.10.1.) supported by the NIGMS P41-GM103311 program. The final model image was then created, again, using the UCSF Chimera (v.1.10.1.) (Pettersen et al., 2004).

2.2.4 Characterization of the Recombinant AtENTH Protein through Molecular Docking

After obtaining the best structural model for the AtENTH, its AC center was then docked with the different substrates (ATP, GTP, CTP or TTP) using AutoDock Vina (v.1.1.2) (Trott and Olson, 2010). Using the same AutoDock Vina, the frequencies of positive binding pose for each of the 4 tested different substrates were generated across a total of 18 simulations per substrate and then evaluated and expressed as percentages. Finally, an image of the AtENTH with a substrate with the best docking frequency and positive binding pose was then created and presented with PyMOL (v.1.7.4.) (The PyMOL Molecular Graphics System, Schrödinger, LLC).

CHAPTER 3

RESULTS

3.1 Partial Expression of the Recombinant AtENTH Protein

For the AtENTH protein to be characterized in this study, *E. cloni* BL21 (DE3) pLysS DUOs cells harbouring the pTrcHis2-TOPO:AtENTH expression construct were chemically induced with 1 mM IPTG to partially express the desired AtENTH recombinant protein. As is shown in Figure 3.1, the resultant AtENTH recombinant protein was produced as a His-tagged fusion product (Crowe et al., 1994) with an approximate size of 17.630 kDa.

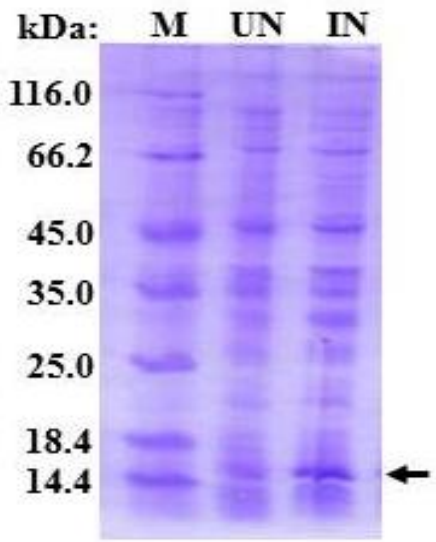

Figure 3.1: Partial expression of the recombinant AtENTH protein. An SDS-PAGE of protein fractions expressed in *E. cloni* BL21 (DE3) pLysS DUOs cells harbouring the pTrcHis2-TOPO:AtENTH expression construct, where M is representing the unstained low molecular weight marker (Thermo Scientific International Inc., Burlington, Canada), UN is representing the non-induced culture and IN is representing the culture induced with 1 mM IPTG. The arrow is marking the expressed and desired recombinant AtENTH protein.

3.2 Affinity Purification of the Recombinant AtENTH Protein

After partial expression of the recombinant AtENTH as a His-tagged product, its purification was then undertaken on a Ni-NTA affinity system (Lindwall et al., 2000, Stempfer et al., 1996) under native non-denaturing conditions (QIAGEN, 2003). As is shown below in Figure 3.2, the successful purification of the desired recombinant AtENTH was achieved.

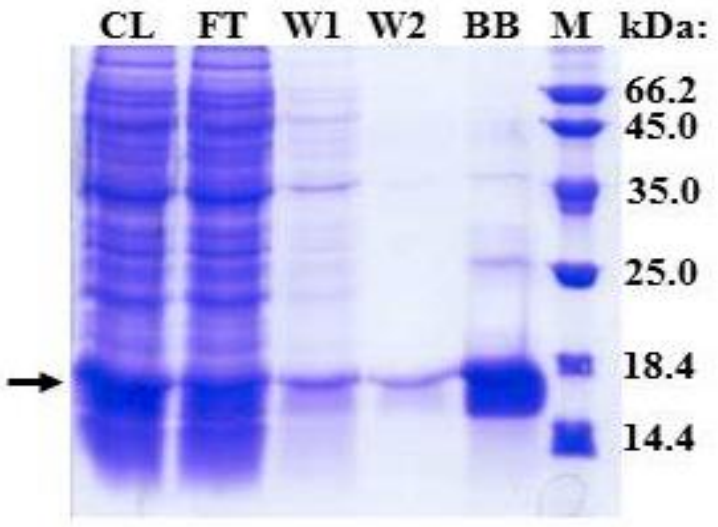

Figure 3.2: Affinity purification of the recombinant AtENTH protein under native non-denaturing conditions. An SDS PAGE of the recombinant AtENTH fractions collected at different steps of its purification procedure. CL is the cleared lysate generated after cell rupturing in PBS buffer and clarification through centrifugation, FT is the flow-through of the cleared lysate after it was passed through the Ni-NTA resin matrix, W1 is the first wash of the bound AtENTH onto the Ni-NTA matrix with wash buffer, W2 is the second wash, and BB is the purified and bound AtENTH. M represents the unstained low molecular weight marker (Fermentas International Inc., Burlington, Canada) while the arrow is marking the recombinant AtENTH protein.

3.3 Chemical Elution, Desalting, and Concentration of the Purified AtENTH Protein

After affinity purification, the purified recombinant AtENTH protein was then eluted off the Ni-NTA matrix, followed by desalting and concentration. The resultant purified protein product is shown in Figure 3.3.

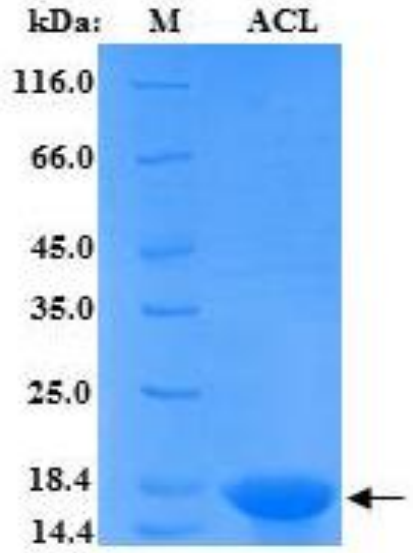

Figure 3.3: Chemical elution, desalting, and concentration of the purified recombinant AtENTH protein. An SDS-PAGE of the eluted, desalted and concentrated purified recombinant AtENTH protein, where M represents the low molecular weight marker (ThermoFisher Scientific Inc., Missouri, USA) while ACL represents the eluted, desalted and concentrated purified recombinant AtENTH protein. The arrow is marking the final purified recombinant AtENTH protein product.

3.4 Validation of the AC Activity of the Recombinant AtENTH Protein

The inherent AC activity of the expressed recombinant AtENTH protein was confirmed through a complementation testing before its purified version was functionally characterized. In this process, some chemically competent mutant *E. coli* host cells or the *cya*A SP850 strain (Coli Genetic Stock Center, Yale University, Connecticut, USA), which lacks the AC activity was transformed with the pTrcHis2-TOPO:AtENTH expression construct, followed by assessment of the ability of the recombinant AtENTH protein to rescue the mutant host and enabling it to metabolize lactose in MacConkey agar. As shown in Figure 3.4, the recombinant AtENTH protein successfully complement the mutant host to ferment lactose.

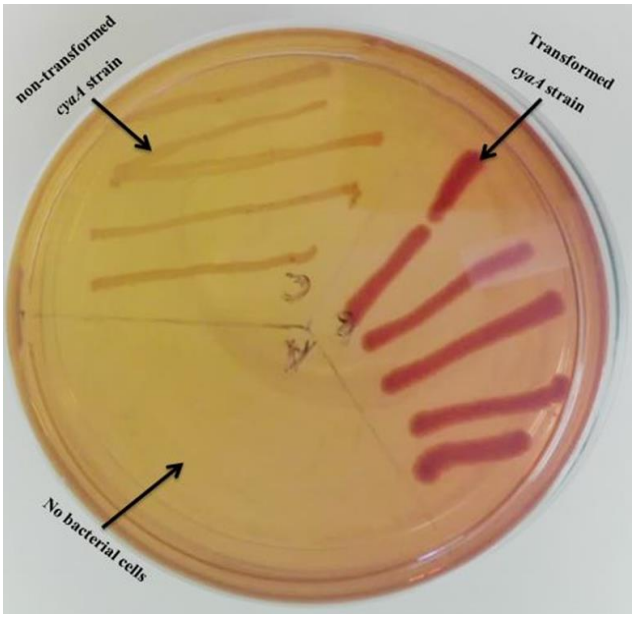

Figure 3. 4: Complementation testing of the recombinant ENTH protein. Segment A contains no bacterial cells, segment B contains the transformed *E. coli cya*A SP850 mutant cells harbouring the recombinant pTrcHis2-TOPO: AtENTH expression construct and showing a magenta/deep purple phenotype that signifies a lactose-fermenting phenotype and segment C contains the non-transformed *E. coli cya*A SP850 mutant cells, which do not ferment lactose and thus produced whitish/yellowish colonies.

3.5 Determination of the Substrate Specificity of the Recombinant AtENTH Protein

After expression and purification of the recombinant AtENTH protein, the purified protein was then tested for its substrate specificity among the commonly known four triphosphates (ATP, GTP, CTP, and TTP) using enzyme immunoassay (catalog # CA201; Sigma-Aldrich Inc., Missouri, USA). As shown in Figure 3.5, the recombinant AtENTH strongly preferred ATP as its sole substrate of activity than the other three ordinary triphosphates.

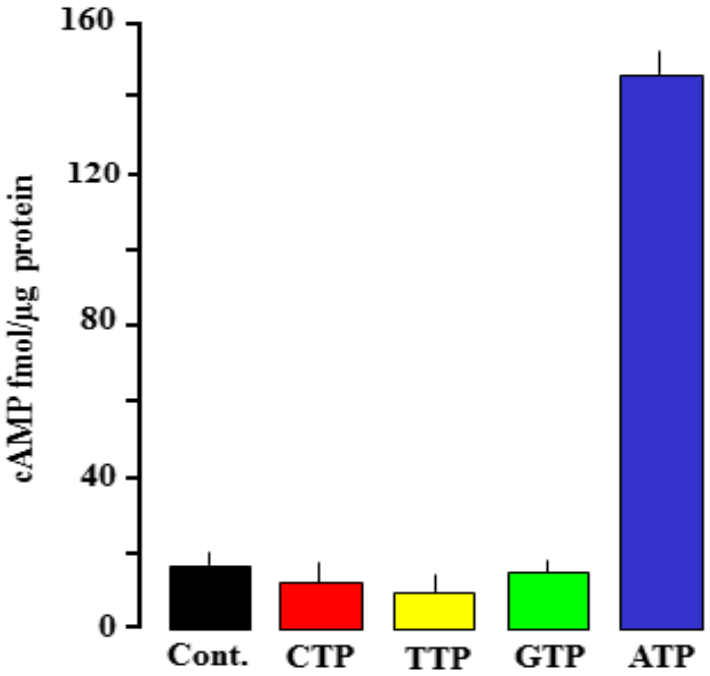

Figure 3.5: Determination of the substrate specificity of the recombinant AtENTH protein. Reaction mixtures containing 10 µg of the purified recombinant AtENTH protein; 50 mM Tris-HCl; pH: 8.0; 2 mM IBMX; 5 mM Mn^{2+}; and 1 mM of either CTP, TTP, GTP or ATP were prepared and incubated at room temperature for 20 minutes. The generated cAMP was measured with an cAMP-specific enzyme immunoassaying system (catalog # CA201; Sigma-Aldrich Inc., Missouri, USA) based on the acetylation version of its protocol. Cont is the cAMP content generated in the absence of the purified recombinant AtENTH while CTP, TTP, GTP or ATP represent the cAMP levels generated by the purified recombinant AtENTH using either CTP, TTP, GTP or ATP as a substrate. Error bars represent the standard errors (SEM) of the means of three independent and representative assays ($n = 3$; $p < 0.05$).

3.6 Structural Analysis of the Recombinant AtENTH Protein

After determining the substrate specificity of the AtENTH, its physical 3-dimensional (3-D) structure was then modeled and configured so that its associated functional and regulatory elements could be determined. The 3-D structure was modeled and configured via the iterative threading assembly refinement (I-TASSER) method (Zhang, 2008) and the process was carried out in such a way that both the ventral (front) and dorsal (back) sides of the AtENTH protein could be easily visualized and assessed (Figure 3.6).

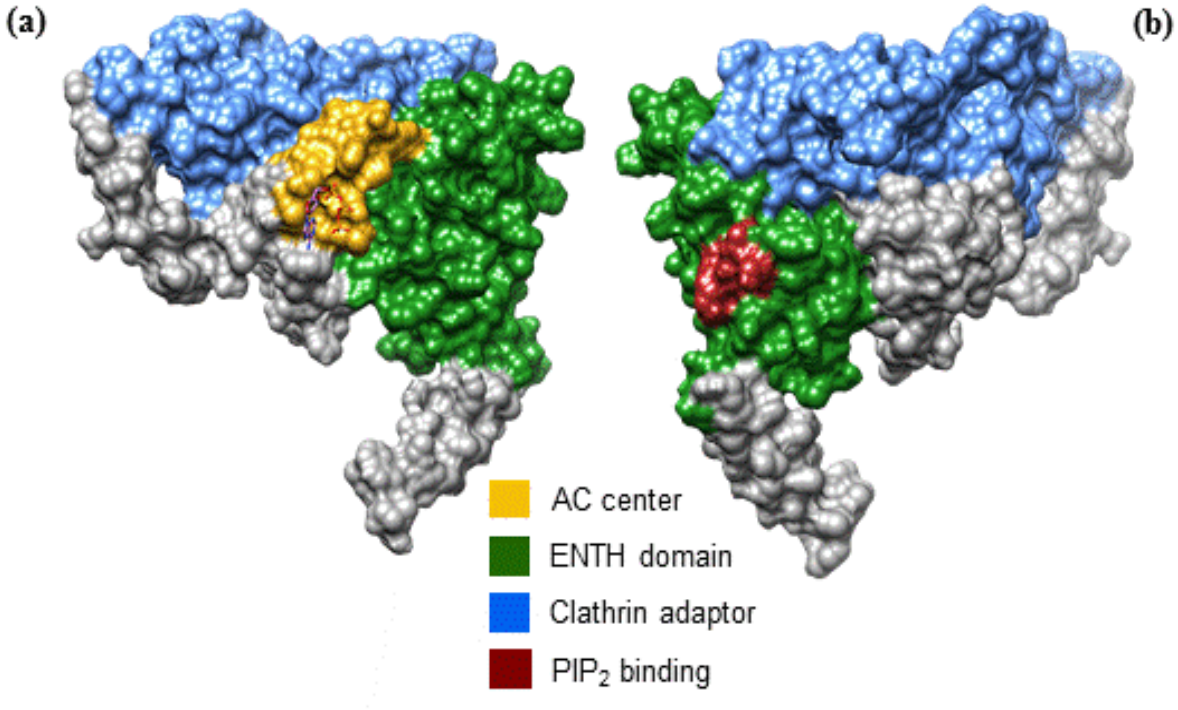

Figure 3.6: Structural features of the recombinant AtENTH protein. Full-length model of the AtENTH protein, showing (a) the AC catalytic center (gold) at its ventral side and (b) the PIP_2 binding site (brown) at the dorsal side. The ENTH domain (green) and clathrin adaptor (blue) make up the membrane-binding regions of the whole protein (grey) at either side. All these domains are key and essentially involved in the functional and regulatory processes of the AtENTH protein.

3.7 Functional Characterization of the Recombinant AtENTH Protein through Chemical Docking

After obtaining the model of the full-length AtENTH and identifying its various functional and regulatory elements, the AC catalytic center of the model was then chemically analyzed through docking with the various probable substrates (ATP, GTP. CTP and TTP) and noting the interaction (frequency of docking and positive binding pose) of each of these substrates with key residues at the center. The chemical docking was carried out on both the surface and ribbon models of the protein, using AutoDock Vina (ver. 1.1.2) [Trott and Olson, 2010] .

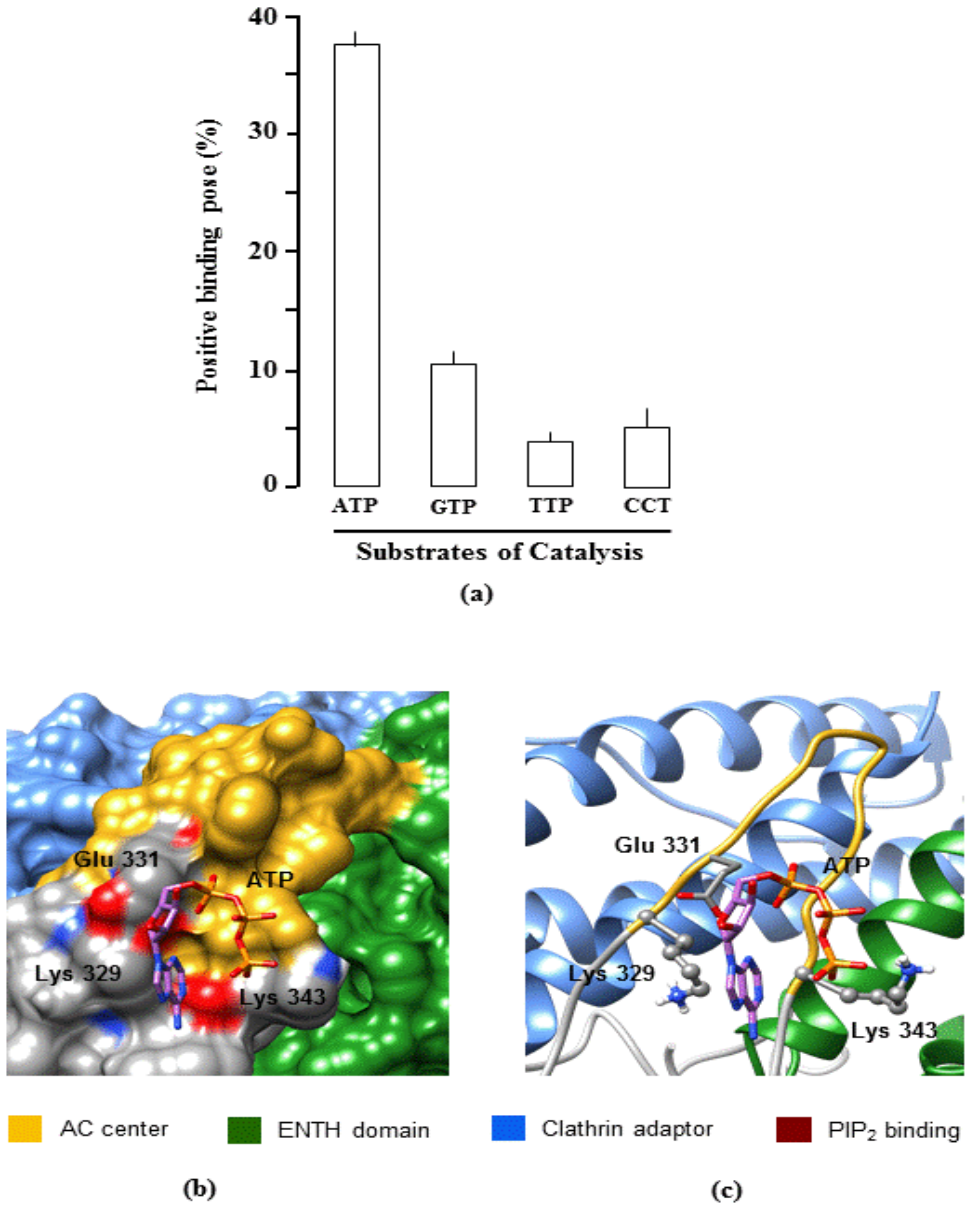

Figure 3.7: Docking of the AC center of the recombinant AtENTH protein. (a) Frequency of the positive binding pose for the various substrate molecules onto the AC center of AtENTH and interaction of ATP with key residues at the AC catalytic center (gold) of the (b) surface and (c) ribbon models of AtENTH. Residues implicated in interactions with ATP are colored according to their charges in the surface model while in the ribbon model, they are shown as individual atoms.

CHAPTER 4

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

4.1 Discussion

Plants are essential for life existence on earth; they supply nearly all terrestrial organisms, including human beings, with various food and resources such as shelter , however, and most importantly, they maintain the oxygen balance in the atmosphere. With the rise in CO_2 levels in the atmosphere influencing the ever-changing climatic condition, plant existence now appears to be at risk. Plants are sessile and therefore, cannot relocate or escape when conditions are extremely unfavourable. Thus, understanding plant morphology and physiology are important for the basis of possible improvement of their defense and adaptation mechanisms and for this, study model plants such as *Arabidopsis thaliana* or *Zea mays* are used. This is because model plants are angiosperms, which mean they are closely related to all high plants; and therefore, complete sequencing of their genes and discovering their functions will offer essential information about the roles of their proteins in higher plants (Koornneef and Meinke, 2010).

For this study, *A. thaliana* was used as a model plant. The Arabidopsis genome encodes 22 proteins with an ANTH/ENTH domain (Song et al., 2012) of which the gene of interest for this study (the At1g25240 gene) encodes an ENTH protein. This protein belongs to the ENTH/VHS/GAT family proteins that are involved in clathrin coat assembly and endocytosis. Previously, the protein was annotated as an adenylyl cyclase (AC) as already stated in Chapter 1 and was one of the nine AC genes Professor Gehring discovered using an AC catalytic motif (Gehring, 2010). More so, this protein was recently confirmed in our research lab as a functional AC (unpublished data). However, its functional, structural and biochemical properties remained

uncharacterized. Hence in this study, this protein was expressed and purified followed by characterization of its enzymatic AC activity using biochemical and structural techniques.

To achieve the set objectives for the study, the *E. cloni* BL21 (DE3) pLysS DUOs cells harbouring the pTrcHis2-TOPO:AtENTH expression construct obtained from the Plant Biotechnology Research Laboratory (North-West University) were grown in 2YT media followed by their induction with 1 mM IPTG to express the AtENTH protein. As is shown in Figure 3.1, the intended and desired AtENTH recombinant protein was successfully expressed and produced as a His-tagged fusion product (Crowe et al., 1994) of an approximate size of 17.630 kDa.

The His-tagged expressed protein was then purified on a Ni-NTA affinity system (Lindwall et al., 2000, Stempfer et al., 1996) under native non-denaturing conditions (QIAGEN, 2003). and according to the manufacturer's protocol (Catalog # P6611; Sigma-Aldrich Inc., Missouri, USA). The purification process is important because it actually isolates the protein of interest from various components within its environment. Ideally, since the AtENTH recombinant was expressed as a His-tagged product, its 6x negatively-charged histidine tail was taken advantage of, to affinity bind to the 6x positively-charged Ni-NTA platform and thus getting purified (QIAGEN, 2003). As is shown in Figure 3.2, the desired purification status of the recombinant AtENTH protein was successfully achieved. The purified recombinant protein was then eluted off the Ni-NTA matrix, followed by its desalting and concentration (Figure 3.3) before its AC activity was subsequently characterized.

After concentration of the eluted and purified recombinant AtENTH, validation of this protein's AC activity was nonetheless done using the complementation test prior to further assessment and elucidation. In this test, a chemically competent mutant *E. coli* host strain, the *cya*A SP850 (Coli Genetic Stock Centre, Yale University, Connecticut, USA) that is deficient in AC activity and incapable of fermenting lactose was transformed with the pTrcHis2-TOPO:AtENTH expression

construct, followed by assessment of its probable ability to ferment lactose on MacConkey agar. As shown in Figure 3.4, a deep reddish or purple color on the transformed mutant host cells was observed, which essentially indicated a positive AC activity for the cloned and recombinantly expressed AtENTH protein and ideally, signifying its ability to now enable the *cya*A mutant strain to metabolize lactose (Moutinho et al., 2010).

After successfully validating the AC activity of the recombinant AtENTH protein, its substrate specificity among the commonly known four organic triphosphates (ATP, GTP, CTP, and TTP) was assessed and determined using the cAMP-linked enzyme immunoassaying system (Sigma-Aldrich Corp., Missouri, USA). As shown in Figure 3.5, the recombinant AtENTH protein strongly preferred ATP other than the other three triphosphates for its activity and thus showing it (ATP) as its sole substrate. While other nucleotide cyclases such as AtWAKL-10 (Meier et al., 2010) and AtPSKRI1 (Kwezi et al., 2011) are capable of optionally utilizing either of the two purines (ATP and GTP), some cyclases such as AtPPR-AC (Ruzvidzo et al., 2013), AtClAP (Chatukuta et al., 2018) and AtLRRAC1 (Bianchert et al., 2018) strictly prefer ATP as their sole substrate of activity. Thus, the behavior of our AtENTH in the reported study was not unusual as it is technically consistent with findings of the other previous undertaken studies.

In addition to the establishment of ATP as the most preferred substrate for AtENTH, we also used computational methods to assess and determine the various functional and regulatory elements of the AtENTH protein. As shown in Figure 3.6, the protein has several functional and/or regulatory elements on its structure, which among others; include the AC catalytic center at its ventral side, the PIP2 binding site at the dorsal side and the ENTH and clathrin adaptor domains that make up its membrane-binding regions throughout its structure. This indicates that the protein is indeed a multi-domain multi-functional molecule related to the other previously established nucleotide cyclases such as AtHNOX (Mulaudzi et al., 2011) and AtClAP (Chatukuta et al., 2018). Additionally, we also showed in this model that the AC catalytic center

in the AtENTH structure is solvent-exposed, thus allowing for unimpeded substrate interactions and presumably substrate catalysis (Figure 3.6).

Further probing of the AC centre through molecular docking with ATP and three other co-triphosphates (GTP, CCT and TTP) showed that ATP has the highest frequency of positive binding pose of approximately 39% followed by GTP (~12%) then CTP (~7%) and lastly, TTP (~6%) (Figure 3.7a). Incidentally, such a finding was very consistent with our previous *in vitro* enzyme immunoassaying finding that established ATP as the most preferred substrate for the AtENTH protein (Figure 3.5) and thus the sole substrate for protein activity. In addition, further probing of the AC centre also suggested that the amino acid residue that stabilizes the transition state from ATP to cAMP in this protein is conferred by the residue at position 14 (Figure 3.6b & c) just like in AtKUP7 (Al-Younis et al., 2015) and centres of the other experimentally confirmed plant ACs (Moutinho et al., 2001; Ruzvidzo et al., 2013; Ito et al., 2014; Al-Younis et al., 2018; Chatukuta et al., 2018, Bianchert et al., 2018). The finding thus provided practical evidence that the annotated intracellular AC catalytic centre contained within the structural domain of AtENTH is indeed a functional motif in plants since a number of higher plant molecules harbouring this same motif have been practically confirmed as functional ACs (Figure 4.1) (Chatukuta et al., 2018, Bianchet et al., 2018, Moutinho et al., 2001, Ruzvidzo et al., 2013, Ito et al., 2014, Al-Younis et al., 2015, Al-Younis et al., 2018). In addition, this also further enhanced the level of confidence into the parameters used to design the motif for use in searching ACs in higher plants (Gehring, 2010).

Figure 4.1: Alignment of the AC centre of AtENTH (AT1G25240) with those of AtPPR-AC (AT1G62590), AtKUP7 (AT5G09400), AtLRRAC1 (AT3G14460), ZmPSiP (AJ307886), and NbAC (ACR77530). All these protein molecules have been experimentally confirmed as functional ACs in higher plants. (Chatukuta et al., 2018, Bianchet et al., 2018, Moutinho et al., 2001, Ruzvidzo et al., 2013, Ito et al., 2014, Al-Younis et al., 2015, Al-Younis et al., 2018)

4.2 Conclusion

Findings of this study technically confirmed AtENTH as a multi-domain multi-functional protein (higher plant AC). The protein has a very strong specificity for ATP as its sole substrate compared to the other known organic triphosphates, and it binds this substrate using the E residue located at position 2 within its catalytic AC center and ultimately catalyzing its conversion into cAMP using the K residue located at the position 14 of its catalytic center (Figure 1.2).

4.3 Recommendations

- Research in this field should be continued through large-scale cultural experiments as well as other molecular physiology and/or mutational studies, to clarify the respective and precise mechanisms behind the observed various downstream processes that occur under high temperatures, exposure to toxic metals and high CO_2 conditions.

- Information gleaned from such studies could be used to optimize plant performance under various environmental stress conditions and in a tailor-made manner that generates valuable materials for both crop breeding and crop production.

- In the laboratories, however, expression of the full-length AtENTH protein, followed by a concerted assaying of its *in vitro* AC activity in response to inositol and/or calcium probing is highly recommended.

REFERENCES

AL-YOUNIS, I., WONG, A. & GEHRING, C. 2015. The *Arabidopsis thaliana* K^+-uptake permease 7 (AtKUP7) contains a functional cytosolic adenylate cyclase catalytic centre. *FEBS Letters,* 589, 3848-3852.

AL-YOUNIS, I., WONG, A., LEMTIRI-CHLIEH, F., SCHMOCKEL, S., TESTER, M., GEHRING, C. & DONALDSON, L. 2018. The *Arabidopsis thaliana* K^+-uptake permease 5 (AtKUP5) contains a functional cytosolic adenylate cyclase essential for K^+ transport. *Front Plant Sci,* 9, 1645.

BIANCHET, C., WONG, A., QUAGLIA, M., ALQURASHI, M., GEHRING, C., NTOUKAKIS, V. & PASQUALINI, S. 2018. An *Arabidopsis thaliana* leucine-rich repeat protein harbors an adenylyl cyclase catalytic center and affects responses to pathogens. *J Plant Physiol,* 232, 12-22.

BIERWIRTH, P. 2018. Carbon dioxide toxicity and climate change: a major unapprehended risk for human health [online]. *Figshare.* doi:10.6084/m9.figshare.1256261.v2.

CHATUKUTA, P., DIKOBE, T., KAWADZA, D., SEHLABANE, K., TAKUNDWA, M., WONG, A., GEHRING, C. & RUZVIDZO, O. 2018. An Arabidopsis clathrin assembly protein with a predicted role in plant defense can function as an adenylate cyclase. *Biomolecules,* 8, 15.

CROWE, J., DOBELI, H., GENTZ, R., HOCHULI, E., STUBER, D. & HENCO, K. 1994. 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol Biol,* 31, 371-87.

DE CAMILLI, P., CHEN, H., HYMAN, J., PANEPUCCI, E., BATEMAN, A. & BRUNGER, A. T. 2002. The ENTH domain. *FEBS Letters,* 513, 11-18.

FELIZIANI, C. & TOUZ, M. C. 2017. To a better understanding of the giardial ENTH protein function. *Biosci Trends,* 11, 115-119.

GEHRING, C. 2010. Adenyl cyclases and cAMP in plant signaling - past and present. *Cell Commun Signal,* 8, 15.

GLEISNER, M., KROPPEN, B., FRICKE, C., TESKE, N., KLIESCH, T. T., JANSHOFF, A., MEINECKE, M. & STEINEM, C. 2016. Epsin N-terminal homology domain (ENTH) activity as a function of membrane tension. *J Biol Chem,* 291, 19953-61.

HOLSTEIN, S. E. & OLIVIUSSON, P. 2005. Sequence analysis of *Arabidopsis thaliana* E/ANTH-domain-containing proteins: membrane tethers of the clathrin-dependent vesicle budding machinery. *Protoplasma,* 226, 13-21.

ITO, M., TAKAHASHI, H., SAWASAKI, T., OHNISHI, K., HIKICHI, Y. & KIBA, A. 2014. Novel type of adenylyl cyclase participates in tabtoxinine-β-lactam-induced cell death and occurrence of wildfire disease in *Nicotiana benthamiana*. *Plant Signaling & Behavior,* 9, e27420.

KOORNNEEF, M. & MEINKE, D. 2010. The development of Arabidopsis as a model plant. *Plant J,* 61, 909-21.

LAI, C. L., JAO, C. C., LYMAN, E., GALLOP, J. L., PETER, B. J., MCMAHON, H. T., LANGEN, R. & VOTH, G. A. 2012. Membrane binding and self-association of the epsin N-terminal homology domain. *J Mol Biol,* 423, 800-17.

LAW, A. H. Y., SHEN, J. & JIANG, L. 2012. SCAMP, VSR, and plant pndocytosis. *Endocytosis in Plants.* Springer, 217-231.

LINDWALL, G., CHAU, M., GARDNER, S. R. & KOHLSTAEDT, L. A. 2000. A sparse matrix approach to the solubilization of overexpressed proteins. *Protein Eng,* 13, 67-71.

MOUTINHO, A., HUSSEY, P. J., TREWAVAS, A. J. & MALHO, R. 2001. cAMP acts as a second messenger in pollen tube growth and reorientation. *PNAS,* 98, 10481-10486.

MULAUDZI, T., LUDIDI, N., RUZVIDZO, O., MORSE, M., HENDRICKS, N., IWUOHA, E. & GEHRING, C. 2011. Identification of a novel *Arabidopsis thaliana* nitric oxide-binding molecule with guanylate cyclase activity *in vitro*. *FEBS Letters,* 585, 2693-2697.

NASA. 2017. *Carbon dioxide* [Online]. California Institute of Technology. Available at: https://climate.nasa.gov/vital-signs/carbon-dioxide [Accessed: 10 April 2018].

NSF. 2017. *The role of Arabidopsis in plant science research* [Online]. Available at: https://www.nsf.gov/bio/pubs/reports/arabid/chap1.htm [Accessed: 15 March 2018].

PANDEY, P., IRULAPPAN, V., BAGAVATHIANNAN, M. V. & SENTHIL-KUMAR, M. 2017. Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Front Plant Sci,* 8, 537.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem,* 25, 1605-12.

PIERCE, K. L., PREMONT, R. T. & LEFKOWITZ, R. J. 2002. Signalling: seven-transmembrane receptors. *Nature Rev Mol Cell Biol,* 3, 639.

QIAGEN 2003. QIA*Expressionist:* A handbook for high-level expression and purification of 6xHis-tagged proteins. Germany QIAGEN GmbH QIAGEN Strasse 1 • 40724 Hilden.

RITTER, B. & MCPHERSON, P. S. 2006. There's a GAP in the ENTH domain. *PNAS,* 103, 3953-3954.

RUZVIDZO, O., DIKOBE, B. T., KAWADZA, D. T., MABADAHANYE, G. H., CHATUKUTA, P. & KWEZI, L. 2013. Recombinant expression and functional testing of candidate adenylate cyclase domains. *Cyclic Nucleotide Signaling in Plants.* Springer.

SASSONE-CORSI, P. 2012. The cyclic AMP pathway. *Cold Spring Harbor Perspectives in Biology,* 4, a011148.

SEN, A., MADHIVANAN, K., MUKHERJEE, D. & AGUILAR, R. C. 2012. The Epsin protein family: coordinators of endocytosis and signaling. *Biomol Concepts,* 3, 117-126.

SEREZANI, C. H., BALLINGER, M. N., ARONOFF, D. M. & PETERS-GOLDEN, M. 2008. Cyclic AMP: master regulator of innate immune cell function. *American J Resp Cell and Mol Biol,* 39, 127-132.

SONG, K., JANG, M., KIM, S. Y., LEE, G., LEE, G. J., KIM, D. H., LEE, Y., CHO, W. & HWANG, I. 2012. An A/ENTH domain-containing protein functions as an adaptor for clathrin-coated vesicles on the growing cell plate in Arabidopsis root cells. *Plant Physiol,* 159, 1013-25.

STEMPFER, G., HOLL-NEUGEBAUER, B. & RUDOLPH, R. 1996. Improved refolding of an immobilized fusion protein. *Nat Biotechnol,* 14, 329-34.

TAIR. 2017. *About Arabidopsis* [Online]. Available at: https://www.arabidopsis.org/portals/education/aboutarabidopsis.jsp [Accessed: 2 April 2018].

TAIR. 2018. *About Arabidopsis : Locus: AT1G25240* [Online]. Available at: https://www.arabidopsis.org/servlets/TairObject?id=30371&type=locus [Accessed: 2 April 2019].

TERASHIMA, I., YANAGISAWA, S. & SAKAKIBARA, H. 2014. Plant responses to CO_2: background and perspectives. *Plant Cell Physiol,* 55, 237-40.

TROTT, O. & OLSON, A. J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem,* 31, 455-61.

UCS. 2017. *Global Warming Science* [Online]. Available at: https://www.ucsusa.org/our-work/global-warming/science-and-impacts/global-warming-science [Accessed:2 April 2018].

WIXON, J. 2001. Featured organism: *Arabidopsis thaliana*. *Comp Funct Genomics,* 2, 91-8.

ZHANG, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics,* 9, 40.

ZOUHAR, J. & SAUER, M. 2014. Helping hands for budding prospects: ENTH/ANTH/VHS accessory proteins in endocytosis, vacuolar transport, and secretion. *Plant Cell,* 26, 4232-44.