Molecular identification and functional characterization of a novel adenylyl cyclase from *Glycine max*

ED Bobo

orcid.org 0000-0002-1558-1704

Thesis accepted in fulfilment of the requirements for the degree

Doctor of Philosophy in Biology

at the North-West University

Promoter: Prof O Ruzvidzo

Co-supervisor: Dr SS Mlambo

Co-supervisor: Dr TD Kawadza

Graduation ceremony: July 2020

Student number: 27537730

PREFACE AND ACKNOWLEDMENTS

This research work is a ground-breaking discovery in the study of secondary messengers in plants commonly known as cyclic adenosine monophosphate (cAMP) which are important signalling molecules. cAMP synthesis is made possible through the action of adenylyl cyclase (AC) enzymes that are responsible for increasing their concentration in cells. This ground-breaking research focused on cloning and expression of the first ever AC in *Glycine max;* XP_003529590, gene ID Glyma.07G251000 against a background of only 8 cloned and expressed ACs in higher plants. Five of these are from *Arabidopsis thaliana*, the other three from *Hippaestrum hybridium*, *Nicotiana tabacum* and *Zea mays*. The goal of the research was to elucidate the functional roles of the novel AC in soybean through molecular and bioinformatic characterisation. A lot of experimental work was covered in the Plant Biotechnology laboratory at the North West University in Mafikeng, South Africa to make this project a success.

First and fore-most I would want to express my sincere gratitude to my mentor and promoter Professor Oziniel Ruzvidzo for trusting and believing in me that I had it in me to handle a project of such great magnitude as I was coming from a purely ecological background. I also wish to thank Dr D Kawadza for ever being available to assist, particularly through the heavy hurdles of understanding the use and application of bioinformatics tools. This project would not have been a success without the assistance of the Plant Biotechnology Research Group, my greatest appreciation is to Selaelo Katlego Sehlabane for the support and assistance she provided in the laboratory. I wish also to thank the North West University for the NWU Doctoral bursary and the institutionary bursary that provided the much-needed financial assistance to facilitate my studies. My employer; Bindura University of Science Education and members of the Biological Sciences department for granting me study leave to focus on my studies. Last but not least, my children Sasha and Lisa for believing in me. Finally, all Glory be to God for his everlasting faithfulness.

PRELIMINARY SUMMARY

The overall aim of this research was to identify and characterise a predicted adenylyl cyclase (AC) enzyme in *Glycine max*; accession number XP_003529590; gene ID Glyma.07G251000. To start with, a preliminary bioinformatic analysis of the XP_003529590 gene was performed prior to the practical experimental work so as to gain a better understanding of the gene annotation, gene expression profile and its secondary structure. After that, total mRNA was then isolated from the soybean plant followed by amplification of the targeted XP_003529590 gene via RT-PCR and its subsequent cloning into the pTRcHis2-TOPO TA cloning vector. The successfully cloned XP_003529590 was then used to transform some chemically competent *E. coli* BL21 (DE3) pLysS expression cells followed by recombinant protein expression through induction with 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG). The expressed recombinant protein was herein referred to as GmAC1. After the expression, the ability of the expressed recombinant GmAC1 protein to generate cyclic adenosine monophosphate (cAMP) within the transformed cells was then assessed and determined endogenously using the enzyme immunoassaying system. An establishment of the actual AC activity of the recombinant GmAC1 protein was then undertaken via a complementation system using the SP850 *E. coli* mutant strain. After confirmation of the AC activity, expression of the GmAC1 protein was upscaled, followed by its affinity purification on a HisPur Ni-NTA resin matrix. After purification, an *in vitro* characterisation of the GmAC1's enzymatic activity was then undertaken using the enzyme immunoassaying system. Finally, the probable physiological roles of the XP_003529590 gene in soybean were then assessed and established through bioinformatic analysis. Consequently, the undertaken preliminary bioinformatic analysis showed that the gene ID for the XP_003529590 is Glyma.07G251000 (Glyma_07G251000), which is primarily expressed during the primary root development and in the primary meristems, and its protein product being a nucleic acid and/or compound binding alpha-helical pentatricopeptide protein. In addition, the undertaken endogenous assaying of the

expressed recombinant GmAC1 protein showed that this protein could enhance cAMP production in the transformed bacterial cells to about ≥ 3.0 folds. Eventually, the complementation testing then practically confirmed that the expressed recombinant GmAC1 protein is indeed a bona fide AC molecule as it could physiologically rescue the mutant SP850 *E. coli* host from being a non-lactose fermenter to a lactose fermenter. Subsequently, the *in vitro* characterisation of the GmAC1 showed that the recombinant protein was indeed a soluble AC (sAC) as its activity could be positively enhanced by the Mn^{2+}, Ca^{2+}, HCO_3^- molecular ions and not the F^- ion. Finally, the physiological evaluation of the XP_003529590 through bioinformatics strongly predicted its primary role in abiotic and biotic stress tolerence particularly during the juvenile developmental stages of the soybean plant. Therefore, the researched XP_003529590 or GmAC1 protein can be a very useful molecular component in possible further research to produce transgenic plants/crops that are tolerant to abiotic stresses such as drought, cold, flooding and salinity that affect crop plants during their early developmental stages.

Key terms: *Glycine max*; soybean; adenylyl cyclase (AC); cyclic adenosine monophosphate (cAMP); abiotic stress.

CHAPTER 1……5

INTRODUCTION AND LITERATURE REVIEW 5

1.1 Introduction 5

1.1.1 Background 5

1.1.2 Problem statement 8

1.1.3 Research aim 9

1.1.4 Research objectives 9

1.1.5 Significance of research 10

1.2 Literature review 10

1.2.1 Cyclic nucleotides 10

1.2.2 Plant adenylyl cyclases 14

1.2.3 Adenylyl cyclase activity in legumes 15

1.2.4 Adenylyl cyclases and the soybean plant 17

CHAPTER 2 20

PRELIMINARY BIOINFORMATIC ANALYSIS OF XP_003529590 PROTEIN 20

2.1 Introduction 21

2.2 Materials and methods 23

2.2.1 Gene annotation of the XP_003529590 24

2.2.2 Expression profile of the XP_003529590 protein coding gene in soybean tissues 24

2.2.3 Protein modelling and structure prediction 25

2.2.4 Protein-protein interaction of the XP_003529530 gene 26

2.3 Results 26

2.3.1 Gene annotations of the XP_003529590 AC gene 26

2.3.2 Expression profile of the XP_003529590 in soybean tissues 27

2.3.3 Protein modelling, prediction and analysis 28

2.3.4 Protein-protein interaction of the XP_003529590 in *G. max* 33

2.4 Discussion 33

2.5 Conclusion 36

CHAPTER 3 37

PARTIAL EXPRESSION OF THE RECOMBINANT GmAC1 PROTEIN AND MOLECULAR DETERMINATION OF ITS ENDOGENOUS ADENYLYL CYCLASE ACTIVITY 37

3.1 Introduction 38

3.2 Materials and methods 41

3.2.1 Preparation of plant material 41

3.2.2 Designing and acquisition of sequence-specific primers 41

3.2.3 Isolation of the targeted GmAC1 gene fragment 42

3.2.3.1 Extraction of the total RNA 42

3.2.3.2 Isolation and amplification of the targeted GmAC1 gene fragment 43

3.2.4 Agarose gel electrophoresis of the amplified GmAC1 gene fragment 45

3.2.5 Cloning of the amplified GmAC1 gene fragment 45

3.2.5.1 Addition of the 3´-adenine overhangs 45

3.2.6 Transformation of the chemically competent BL21 (DE3) pLysS *E. coli* cells with the pTRcHis2-TOPO:GmAC1 fusion expression construct 50

3.3 Partial expression of the recombinant GmAC1 protein 50

3.4. Determination of the endogenous AC activity of the recombinant GmAC1 protein 51

3.5 Results 52

3.5.1 Isolation of the GmAC1 gene fragment from *G. max* 52

3.5.3 Partial expression of the recombinant GmAC1 protein 54

3.5.4 Molecular determination of the endogenous AC activity of the GmAC1 protein 55

3.6 Discussion 56

3.7 Conclusion 60

CHAPTER 4 61

DETERMINATION OF THE *IN-VIVO* ADENYLYL CYCLASE ACTIVITY OF THE RECOMBINANT GmAC1 PROTEIN 61

4.1 Introduction 62

4.2 Materials and methods 64

4.2.1 Preparation of competent *E. coli* SP850 *cya*A mutant cells 64

4.2.2 Transformation of the competent *E. coli* SP850 *cya*A cells with the pTRcHis2-TOPO:GmAC1 expression construct 65

4.2.3 Complementation testing of the recombinant GmAC1 protein 66

4.3 Results 66

4.4 Discussion 67

4.5 Conclusion 69

CHAPTER 5 70

AFFINITY PURIFICATION OF THE RECOMBINANT GmAC1 PROTEIN AND *IN VITRO* CHARACTERISATION OF ITS ENZYMATIC ACTIVITY 70

5.1 Introduction 71

5.2 Materials and methods 75

5.2.1 Over-expression of the recombinant GmAC1 protein 75

5.2.2 Determination of the soluble or insoluble status of the expressed recombinant GmAC1 protein 76

5.2.3 Purification of the recombinant GmAC1 protein 76

5.2.4 Protein renaturation 78

5.2.4.1 Renaturation of the recombinant GmAC1 protein 78

5.2.5 Elution of the recombinant GmAC1 protein 79

5.2.6 Concentration and desalting of the recombinant GmAC1 protein 80

5.2.7 Functional characterisation of the recombinant GmAC1 protein 80

5.2.7.1 Sample preparation and enzyme immunoassaying 80

5.3 Results 82

5.3.1 Determination of the solubility/insolubility status of the recombinant GmAC1 protein 82

5.3.2 Affinity purification of the recombinant GmAC1 protein 83

5.3.3 Renaturation and elution of the recombinant GmAC1 protein 84

5.3.4 Characterisation of the enzymatic activity of the recombinant GmAC1 protein 85

5.4 Discussion 86

5.5 Conclusion..... 92

CHAPTER 6 93

EXTENSIVE BIOINFORMATIC ANALYSIS OF THE NOVEL XP_003529590 SOYBEAN GENE..... 93

6.1 Introduction 94

6.2 Materials and methods 100

6.2.1 Functional prediction of the XP_003529590 gene through multiple sequence alignments 100

6.2.2 Functional prediction of the XP_003529590 gene through co-expression analysis 100

6.2.3 Functional prediction of the XP_003529590 gene through stimuli expression responses 101

6.2.4 Functional prediction of the XP_003529590 gene through Gene Ontology..... 102

6.3 Results 102

6.3.1 Functional prediction of the XP_003529590 gene through multiple sequence alignments 102

6.3.2 Functional prediction of the XP_003529590 through co-expression analysis..... 105

6.3.3 Functional prediction of the XP_003529590 gene through stimuli expression responses 109

6.3.4 Gene Ontology FFPred analysis..... 113

6.4 Discussion 114

6.5 Conclusion..... 127

CHAPTER 7 128

GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS 128

REFERENCES 134

APPENDICES 186

Appendix A: cAMP enzyme immunoassay (EIA) procedure 186

Appendix B: PSI-BLAST DOMAINATION multiple sequence alignments 187

LIST OF TABLES

Table 1.1: The fourteen bioinformatically identified *Arabidopsis thaliana* proteins containing the AC catalytic search motifs (adapted from Gehring, 2010)..... 13

Table 2.1: BLAST result of binding surface sites on the XP_003529590 protein..... 31

Table 2.2: Values for localisation prediction of the XP_003529590 protein..... 32

Table 3.1: Components of an RT-PCR reaction mixture for amplification of the targeted GmAC1 gene fragment. 44

Table 3.2: The 1-step RT-PCR thermal cycling reaction conditions for amplification of the targeted GmAC1 gene fragment. 44

Table 3.3: Reaction components of a PCR reaction mixture to confirm successful ligation of the GmAC1 gene fragment (insert) into the pTrcHis2-TOPO expression vector. 49

Table 3.4: Reaction components of a PCR reaction mixture to confirm correct orientation of the GmAC1 gene fragment (insert) into the pTrcHis2-TOPO expression vector. 49

Table 3.5: Thermal cycle conditions for confirmation of the successful ligation and correct orientation of the GmAC1 gene fragment (insert) into the pTrcHis2-TOPO expression vector..... 49

Table 5.1: Renaturation conditions of the recombinant GmAC1 protein using the BioLogic DuoFlow Chromatographic system..... 79

Table 5.2: Molecular characterisation of the recombinant GmAC1 protein..... 82

Table 6.1: Gene IDs for the 7 domains with significant sequence scores to the XP_003529590 gene.. 103

Table 6.2: A Phytozome 12 inference of the HO4D005110 gene family with close similarity to the XP_003529590 (https://phytozome.jgi.doe.gov/pz/portal.html)..... 105

Table 6.3: The 25 top most genes positively co-expressed with the XP_003529590 gene in soybean during development..... 107

Table 6.4: The 25 top most genes positively co-expressed with the XP_003529590 gene in response to various perturbations..... 108

Table 6.5: FFPred Gene Ontology Analysis of the XP_003529590 gene..... 114

LIST OF FIGURES

Figure 1.1: Catalytic centre motifs of nucleotide cyclases. 13

Figure 1.2: Amino acid sequences of the XP_003529590 protein from *G. max*.. 19

Figure 2.1: Gene structure of Glyma.07G251000 from EnsemblGenomes**.** 28

Figure 2.2: Expression log2 scale of the XP_003529590 across the various tissues of *G. max* as is tested by Genevestigator..... 289

Figure 2.3: Predicted Alpha helix structure of the XP_003529590 protein product**.** 30

Figure 2.4: Secondary structure map from DISOPRED3 indicating the amino acid residues making up the secondary structure the soybean XP_003529590 protein..... 31

Figure 2.5: Intrinsic disorder profile of the XP_003529590 *G. max* protein**.**..... 31

Figure 2.6: 3-D structural prediction of the XP_003529590 protein binding sites..... 33

Figure 3.1: The nucleotide sequence of the XP_003529590 gene showing the GmAC1 gene fragment that was cloned and characterised in this study 43

Figure 3.2: Isolation of the GmAC1 gene fragment from *G. max* 55

Figure 3.3: Determination of the sucsessful cloning of the GmAC1 gene frgament in the pTRcHis-TOPO expression..... 56

Figure 3.4: Partial expression of the GmAC1 recombinant protein..... 57

Figure 3.5: Determination of the endogenous AC activity of the recombinant GmAC1 protein by enzyme immunoassay 58

Figure 4.1: Determination of the AC activity of the recombinant GmAC1 protein via complementation test**..** 69

Figure 5.1: Determination of the insolubility /solubility status of the recombinant GmAC1 protein..... 85

Figure 5.2: Purification of the recombinant GmAC1protein under non-native denaturing conditions**..** 86

Figure 5.3: Refolding and elution of the purified of the recombinant GmAC1 protein**.**..... 87

Figure 5.4: Characterisation of the AC activity of the recombinant GmAC1 protein. 88

Figure 6.1: Protein domains from BLASTP with sequence similarities to the XP_003529590 gene.....105

Figure 6.2: Similarity heat map based on the gene family HO04D005110 InterPro Tetratricopeptide-like helical protein. 106

Figure 6.3: Expression profile of the XP_003529590 gene in response to biotic stress..... 113

Figure 6.4: Expression profile of the XP_003529590 in response to nutrient supply 114

CHAPTER OUTLINE

CHAPTER 1: Introduction and literature review

The background to the research is highlighted in this Chapter, in which the problem statement, aim, specific objectives, significance and literature review of the research are covered.

CHAPTER 2: Preliminary bioinformatic analysis of XP_003529590 protein

A preliminary bioinformatic analysis of the putative XP_003529590 gene is covered in this Chapter. The analysis highlights include gene annotation, expression profile of the gene in soybean plant and the protein secondary structure prediction.

CHAPTER 3: Partial expression of the recombinant GmAC1 protein and molecular determination of its endogenous adenylyl cyclase activity

The extraction and amplification of the XP_003529590 gene fragment, its subsequent cloning, induction of protein expression and the ability of the expressed protein to generate cAMP endogenously are highlighted in Chapter 3.

CHAPTER 4: Determination of the *in vivo* adenylyl cyclase activity of the recombinant GmAC1 protein

Complementation testing of the recombinant GmAC1 protein using SP850 *E. coli* mutant cells was carried out in this Chapter.

CHAPTER 5: Affinity purification of the recombinant GmAC1 protein and *in vitro* characterisation of its enzymatic activity

The highlights of this Chapter include upscaling production of the recombinant GmAC1 protein, its purification through non-native denaturing conditions, refolding of the denatured recombinant protein, washing and its eventual elution. The ability of the purified and renatured recombinant protein to produce cAMP *in vitro* was also assessed.

CHAPTER 6: Extensive bioinformatic analysis of the novel XP_003529590 soybean gene

In this Chapter web-based bioinformatic tools were utilised to predict, confirm and understand the physiological functions of the Glyma.07G251000 gene in soybean.

CHAPTER 7: General discussion, conclusions and recommendations

This Chapter provides a general conclusion to the thesis, summarising the research findings and also providing the thesis conclusion and recommendation on future prospects and research.

LIST OF ABBREVIATIONS

AC	**:**	Adenylyl cyclase
ATP	**:**	3′,5′-Adenosine 5′-triphosphate
cAMP	**:**	Cyclic 3′,5′-adenosine monophosphate
cDNA	**:**	Copy DNA or DNA complementary to RNA
cGMP	**:**	Cyclic 3′,5′-guanine monophosphate
CNC	**:**	Cyclic nucleotide cyclase
DNA	**:**	Deoxyribonucleic acid
EIA	**:**	Enzyme immunoassay
GmAC1	**:**	*Glycine max* adenylyl cyclase 1
GTP	**:**	3′,5′-Guanosine 5′-triphosphate
HSP	**:**	Heat shock protein
IDP	**:**	Intrinsically disordered protein
IDRs	**:**	Intrinsically disordered regions
IPTG	**:**	Isopropyl-β-D-thiogalactopyranoside
LB	**:**	Luria Bertani
mRNA	**:**	Messenger ribonucleic acid
Ni-NTA	**:**	Nickel-nitrilotriacetic acid
OD	**:**	Optical density

PDE	**:**	Phosphodiestarase
PKA	**:**	Protein kinase A
PPR	**:**	Pentatricopeptide repeat
QTL	**:**	Quantitative trait loci
RNA	**:**	Ribonucleic acid
ROS	**:**	Reactive oxygen species
RT-PCR	**:**	Reverse transcriptase-polymerase chain reaction
sAC	**:**	Soluble adenylyl cyclase
SDS-PAGE	**:**	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
tmAC	**:**	Transmembrane adenylyl cyclase
TPR	**:**	Tetratricopeptide repeat
tRNA	**:**	Transfer ribonucleic acid

DEFINITION OF TERMS

Adenylyl cyclases: Enzymes which catalyse the cyclisation of adenosine 5′-triphosphate (ATP) to cyclic adenosine 3′,5′-monophosphate (cAMP) which requires the removal of a pyrophosphate (PPi). They are also known as adenylate cyclases.

Copy DNA: A DNA strand that is complementary to the RNA and synthesised by an RNA-dependent DNA polymerase.

Cyclic AMP: A cyclic nucleoside that acts as a second messenger, activating other enzymes within the cell. It is formed when the enzyme adenylyl cyclase is activated by the alpha subunit of a G protein.

***Glycine max*:** A legume that is commonly known as soybean and is an annual herbaceous plant in the Fabaceae family cultivated for its edible and highly proteinaceous seed.

mRNA: A sub-type of RNA that is generated during transcription, where a single strand DNA is decoded by a DNA-dependent RNA polymerase.

Plasmid DNA: A small circular double-stranded DNA molecule that is distinct from the commonly known chromosomal or nuclear DNA.

Primers: Short nucleic acid sequences that provide a starting point for DNA or RNA synthesis in a conventional PCR system.

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

Co-expression partner: A gene whose expression shows a similar pattern across different samples to that of a gene of interest.

Co-expression network: A system that describes genes that tend to show a coordinated expression pattern across a group of samples exposed to similar experimental conditions. In these networks, each node represents a gene and each edge represents the presence and/or strength of the co-expression relationship.

Disordered proteins: Regions on a protein that are unfolded and therefore lacking the 3-dimensional structure.

Gene: A DNA sequence that can be transcribed into a transcript. In the case of protein coding genes, this transcript can be translated into a protein that is fully functional.

Microarray: A platform for quantifying gene expression that assays mRNA molecules based on their hybridisation to probes present on an array, typically a glass slide.

Mitochondrion: A membrane bound organelle present in the cytoplasm of eukaryotic cells primarily responsible for the generation of ATP.

Recombinant protein: A protein molecule that is encoded by a gene in which the gene is cloned in a system that supports expression of the gene.

Regulatory gene: A gene that controls the expression of other genes.

Soluble adenylyl cyclase: An intracellular enzyme that is a source of the ubiquitous second messenger cAMP in response to bicarbonate and calcium or other external ligands.

Soybean: An annual legume that is cultivated for its proteinaceous seeds, forage and soil improvement. Its scientifically known as *Glycine max*.

Transcript: A single-stranded RNA molecule resulting from the transcription of a functional gene.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

1.1.1 Background

Soybean or *Glycine max* is one of the most important legume crops that provide sources of oil and protein for livestock and humankind. Apart from being consumable, soybean products have been gaining attention for their other additional attributes such as the anti-cancerous properties (Ko *et al*., 2013) in pharmaceuticals and the protein-based bio-degradable properties for possible consideration as alternatives in the plastic industry (Song *et al*., 2011). These diverse attributes of soybean make this legume a more widely desired crop plant, whose demand is rapidly increasing. Nevertheless, soybean production may be hindered by extreme weather conditions such as droughts, floods and heat, and also as a result of various diseases and pathogens (Deshmukh *et al*., 2014). It has been reported that growth and grain yield of this important cash crop are highly affected by water stress (Brevedan and Egli, 2003; Liu *et al*., 2003). However, plants respond and adapt to abiotic and biotic stresses or stress factors with an array of biochemical, physiological, and molecular modifications of which the soybean plant is no exception.

Since the sequencing of the entire soybean genome in 2010 (Schmutz *et al*., 2010), the majority of its protein coding genes remains experimentally unconfirmed (Chai *et al*., 2015). The soybean genome is almost 1.1 gigabases with approximately 46 430 protein coding genes (Turner *et al*., 2012). As such, genes responsible for encoding adenylyl cyclases (ACs) or enzymes responsible for generating the second messenger molecule adenosine 3′,5′-cyclic monophosphate (cAMP)

from adenosine 5′-triphosphate (ATP) in soybean have not been reported anyway. This entails that most genes involved in development, cell division, growth and responses to environmental stimuli are still experimentally unverified. In this regard, it therefore means that there is a real need to study signalling pathways in crop plants such as the soybean so as to gain a better understanding on the exact mechanisms by which plants grow, develop, respond and adapt to the various environmental stress factors. A comprehensive functional characterisation of genes encoding for second messengers involved in signal transduction is therefore, important. Studies in the sequencing of the soybean genome have provided a potential and powerful platform for the study and analysis of expressional and functional processes in this legume.

All along, plant studies have identified cAMP and guanosine 3′,5′-cyclic monophosphate (cGMP) as universal second messengers that play key roles in many physiological responses in higher plants. These cyclic nucleotides are key signalling molecules in most of the processes in plants, which include growth and differentiation, photosynthesis, and biotic and abiotic defenses (Gehring and Turek, 2017). The molecules play important roles in relaying external signals and modifying gene expression in cells of all phyla, where they transfer extracellular signals to appropriate molecules inside the cell (Wheeler, 2013). As stated earlier on, cAMP is generated from ATP by the action of ACs (Lemtiri-Chlieh *et al*., 2011) (often referred to as adenylate or adenyl cyclases) and both the ACs and cAMP play important roles in many signal transduction pathways (Frezza *et al*., 2018), whereby they carry the responsibility of amplifying stimuli received by eukaryotic cells (Neves-Zaph and Song, 2015). Thus, their involvement in signal transduction means they can be involved in regulating plant developmental programs and biotic and abiotic stress responses.

To date, various reports have been made, implicating cAMP in stress response (Choi and Xu, 2010; Thomas *et al*., 2013). The molecule has been reported in most plant specific processes such as stomatal closure since guard cell channels of *Vicia faba* could be modified by cAMP-dependent

phosphorylation (Jin and Wu, 1999) and the growth of pollen tubes of *Agapanthus umbellatus* and *Lilium longiflorum* have also been reported to be regulated by cAMP (Rato *et al*., 2004). Although the AC enzymes perform the same function in all prokaryotes and eukaryotes using the same substrate (ATP), they are different, and they vary in their expression, structure, activity and regulation (Cooper, 2005). Therefore, the importance of cAMP in regulating plant physiological processes requires a close scrutiny and a subsequent clear understanding based on the enzymes responsible for its synthesis and/or generation.

In plants and currently, there are only nine practically and experimentally confirmed ACs that have been reported. These include the *Zea mays* pollen signalling protein (PSiP; AJ307886.1) that participates in polarised pollen tube growth (Moutinho *et al*., 2001); the *Arabidopsis thaliana* pentatricopeptide repeat-containing protein (AtPPR-AC; At1g62590) responsible for chloroplast biogenesis and restoration of male sterility (Ruzvidzo *et al*., 2013), the *Nicotiana benthamiana* protein (NbAC; ACR77530) involved in tabtoxinine-β-lactum-induced cell death during wildfire disease (Ito *et al*., 2014), the *Hippaestrum hybridium* protein (HpAC1; ADM83595) involved in stress signalling (Świeżawska *et al*., 2014), two *A. thaliana* K^+ uptake permease proteins (AtKUP7; At5g09400 and AtKUP5; At4g33530) responsible for K^+ transport (Al-Younis *et al*., 2015; 2018), the *A. thaliana* clathrin assembly protein (AtCIAP; At1g68110) predicted to be involved in actin cytoskeletal remodelling during endocytic internalisation (Chatukuta *et al*., 2018) and the *A. thaliana* leucine-rich repeat protein (AtLRRAC1; At3g14460) involved in defense response against hemibiotrophic and biotrophics pathogens (Bianchet *et al*., 2019). In lower plants, there is the *Marchantia polymorpha* antheridium based reproductive protein (MpCAPE; Mapoly0068s0004) involved in male organ and cell development (Kasahara *et al*., 2016). Of these nine plant ACs, the AtPPR-AC, NbAC, HpAC1, AtKUP7, AtKUP5, AtCIAP and AtLRRAC1 possess the putative AC catalytic motif (Figure 1.1) previously annotated by Gehring in 2010 - this being a very strong sign, which indicates that this motif is indeed essentially functional and

thus increasing our confidence in continuously utilising it to search for more novel ACs in plants, and more importantly, in crop plants such as soybean.

Therefore, this implies that there is still a lot of work to be done on these enzymes, particularly in higher plants. In line with this, some recent and independent phylogenetic analysis of ACs in higher plants pointed to the existence of such molecules in *G. max*; XP_003529590 (Ito *et al*., 2014) and XP_003547191 (Świeżawska *et al*., 2014) and interestingly, none of these proteins has yet been functionally characterised. This research therefore, was set to try and fill this gap of knowledge through a specific analysis of the XP_003529590 protein, and against a luscious background, where past research had vehemently denied the existence of ACs and/or their activity in *G. max* (Yunghans and Moore, 1977). Of the two proteins, the XP_003529590 was chosen specifically because it possesses the annotated AC catalytic motif proposed by Gehring (2010) and later functionally confirmed in various studies (Ruzvidzo *et al*., 2013; Ito *et al*., 2014; Al-Younis *et al*., 2015; Al-Younis *et al*., 2018; Chatukuta *et al*., 2018; Bianchet *et al*., 2019). For the purpose of this study, this targeted protein was referred to as GmAC1 protein.

1.1.2 Problem statement

Although the role of cAMP has been recognised in most biological processes of animal cells, very little is presently known about this molecule and its associated signalling components (ACs) in higher plants. This is also despite the fact that both the cAMP and ACs have been shown to have very close links with essential transduction processes and/or physiological responses ranging from protein phosphorylation to the transcriptional activation of specific genes. To date, several efforts have been made in attempting to identify these molecules in plants and very few plants have been covered, leaving out important agronomic plants like *G. max*. Notably and in two recent related studies, two probable AC candidates; XP_003529590 and XP_003547191 were reported even

though none of them has been fully discussed anywhere. It is from this backdrop that this present study was then premised targeting the first candidate herein referred to as GmAC1. The XP_003529590 was selected over its other counterpart because within its genome, the AC catalytic motif as proposed by Gehring (2010) (shown in Figure 1.1) and functionally confirmed in various studies (Ruzvidzo *et al*., 2013; Ito *et al*., 2014; Al-Younis *et al*., 2015; Al-Younis *et al*., 2018; Chatukuta *et al*., 2018; Bianchet *et al*., 2019) is present.

1.1.3 Research aim

The major research aim of this research project was to isolate the XP_003529590 gene, test its AC function and then further characterise it in relation to stress and adaptation mechanisms.

1.1.4 Research objectives

The following key objectives were set out in order to properly address the proposed research aim:

1. To isolate and clone the AC-containing gene fragment (GmAC1) of the annotated XP_003529590 into a stable and viable heterologous prokaryotic expression system.
2. To optimise the partial expression strategies of the cloned XP_003529590 gene fragment into a recombinant GmAC1 protein.
3. To determine the endogenous and *in vivo* AC activity of the partially expressed and cloned recombinant GmAC1 protein.
4. To affinity purify the partially expressed recombinant GmAC1 protein and further characterise its *in vitro* AC activity

5. To bioinformatically determine and establish the correlation expressional and functional profiles of the XP_003529590 protein in *G. max*.

1.1.5 Significance of research

The molecular establishment of the GmAC1 protein as a *bona fide* higher plant AC and its subsequent functional characterisation as a signal molecule in soybean (particularly in important cellular processes such as growth, development, stress response and nitrogen fixation) would be of paramount and ground-breaking impact. This is because currently, there is no documentation on the existence and/or molecular function of this novel molecule in *G. max* and therefore, this research will immensely contribute towards the main body of science through new literature and new scholarship. Practical-wise, the elucidation of the biological/functional roles of the GmAC1 protein in soybean may be used in transgenics or cisgenics to increase growth and productivity of this very important agronomic crop, thus helping in the address of food security issues in the region (sub-Sahara) and beyond. Soybean is a plant of great economic value as it is a very good source of both protein and oil for human and animals.

1.2 Literature review

1.2.1 Cyclic nucleotides

The most commonly known natural cyclic nucleotide monophosphates include adenosine 3′,5′-cyclic monophosphate (cAMP) and guanosine 3′,5′-cyclic monophosphate (cGMP), which are catalytic products of adenosine 5′-triphosphate (ATP) and guanosine 5′-triphosphate (GTP) respectively. Adenylyl cyclases (AC) are protein enzymes that hydrolyse ATP into cAMP while guanylyl cyclases (GC) are protein enzymes that catalyse the formation of cGMP from GTP.

Several studies have previously reported both cAMP and cGMP as universal second messenger molecules in higher plants (Lemtiri-Chlieh *et al*., 2011; Mathieu-Demaziere *et al*., 2013), that play essential roles in many biological and physiological processes of plants. The two molecules similarly play essential roles in many physiological and developmental processes of all living organisms from prokaryotes (e.g. *E. coli*) to complex multicellular organisms such as *Homo sapiens* (Al-Younis *et al*., 2015). In higher plants, cAMP was reported to have a key role in the activation of protein kinases in the leaf of rice (Komatsu and Hirano, 1993) and in tobacco BY-2 cells, where it promotes cell division (Ehsan *et al*., 1998). cAMP has also been implicated in plant stress responses and defense mechanisms (Thomas *et al*., 2013).

In all organisms, signalling pathways essentially involve specific effector proteins, where in plants, these have been identified as the cyclic nucleotides (cAMP and cGMP), phosphodiesterases (PDEs), cAMP binding proteins known as protein kinase As (PKAs) and ACs (Jager *et al.,* 2012). Of these proteins, PDEs and ACs are responsible for the regulation of the proper cellular levels of cAMP. While ACs increase cAMP levels through the hydrolysis of ATP and in response to extracellular responses, PDE are responsible for the hydrolysis of cAMP to 5′-AMP thereby, lowering cAMP cellular levels (Hanoune and Defer, 2001; Kamenetsky *et al*., 2006; Omori and Kotera, 2007). PKAs are responsible for propagating cAMP responsive cell signalling events through the transfer of a γ-phosphate group of ATP to a downstream protein substrate (Turnham and Scott, 2016).

Apparently, given the point that a lot of research has dealt in detail with the roles of cAMP in plants, it is thus important to also study in detail the AC enzymatic systems responsible for the generation of cAMP. However, it has been so difficult to identify plant molecules with cyclic nucleotide cyclase (CNC) activity, generally because these molecules were reported to be outside the detection limit of an ordinary BLAST search (Wong and Gehring, 2013). A solution to this problem had however,been presented by Gehring in 2010, who proposed and tested a search

strategy that uses motifs deduced from conserved amino acids in the catalytic centre of experimentally annotated and functionally tested CNCs (Figure 1.1). From this approach, at least 14 plant AC molecules were annotated in the Arabidopsis genome (Table 1.1) (Gehring, 2010), while at least six higher plant ACs have since been identified. The identified proteins include the AtPPR-AC protein responsible for chloroplast biogenesis and the restoration of male sterility (Ruzvidzo *et al.,* 2013), the NbAC protein responsible for the tabtoxinine-β-lactam-induced cell deaths during wildfire diseases (Ito *et al*., 2014), the AtKUP7 protein responsible for vacuolar K^+ conductance (Al-Younis *et al*., 2015), the AtKUP5 protein responsible for cAMP-dependant K^+ flux (Al-Younis *et al*., 2018) the AtCIAP protein responsible for actin cytoskeletal remodelling during endocytotic internalisation (Chatukuta *et al*., 2018), and the AtLRRAC1 protein responsible for conferring tolerance to the biotrophic fungus, *Botrytis cinerea* (Bianchet *et al*., 2019). As such this same approach is currently being used by many plant biologists as it has already proved to be a very useful criterion for the successful identification of AC candidates in higher plants. In this regard, the same approach was similarly used in this study for the identification of the first ever AC molecule in soybean *G. max*.

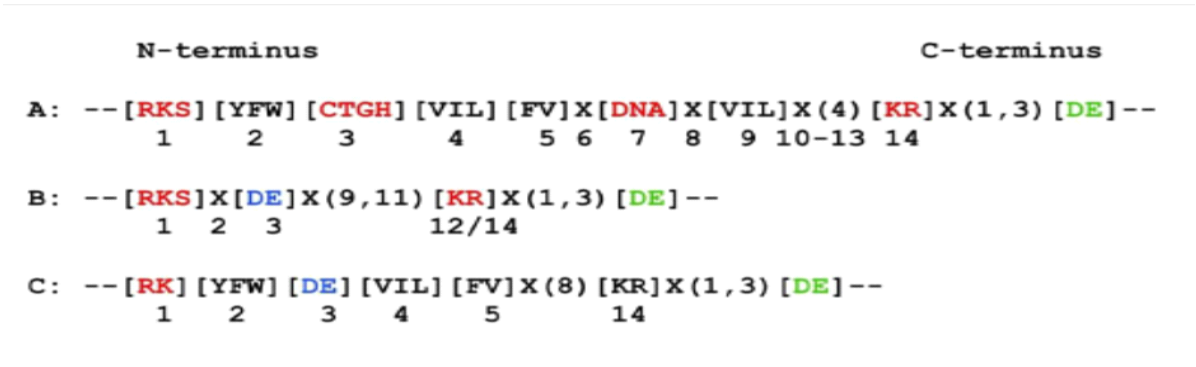


Figure 1.1: Catalytic centre motifs of nucleotide cyclases. (**A**) Centre motif of experimentally tested GCs in plants. The residue (red) in position 1 does the hydrogen bonding with the guanine, the amino acid (red) in position 3 confers substrate specificity and the residue (red) in position 14 stabilises the transition (GTP/cGMP). The $Mg^{2+/}Mn^{2+}$-binding site is C-terminal (green). In the derived motifs (**B** (relaxed) and **C**

(stringent)) specific for ACs, position 3 (blue) has been substituted to [DE] to allow for ATP binding (Gehring, 2010).

Table 1.1: The fourteen bioinformatically identified *Arabidopsis thaliana* proteins containing the AC catalytic search motifs (adapted from Gehring, 2010).

ATG NUM	SEQUENCE	ANNOTATION
At3g14460	-KYDVFPSFRGEDVR-KD-	Disease Resistance Protein
At1g26190*	-SADRVAMRNKNLKR-	Phosphoribulokinase/uridine kinase family protein
At1g73980*	-SVDSRMKYLHGGVSK-	AX4 AC domain containing protein
At2g11890*	-RVEEDEEEIEYWIGK-	G3 AC family protein
At3g21465*	-SSEAKHVENPTEAVK-	Unknown function
At1g25240	-KWEIFEDDFCFTCKDIKE-	Epsin N-terminal homology
At1g62590	-KFDVVISLGEKMQR--LE-	Pentatricopeptide (PPR) protein
At1g68110#	-KWEIFEDDYRCFDR—KD	Clathrin assembly protein
At2g34780	-KFEIVRARNEELKK-EME-	Maternal effect embryo arrest 22
At3g02930	-KFEVVEAGIEAVQR--KE-	Chloroplast protein
At3g04220	-KYDVFPSFRGEDVR--KD-	TIR-NBS-LRR class
At3g18035	-KFDIFQEKVKEIVKVLKD-	Linker histone-like protein – HNO4
At3g28223	-KWEIVSEISPACIKSGLD-	F-box protein
At4g39756	-KWDVVASSFMIERK--CE-	F-box protein

ATG represents the assigned *Arabidopsis thaliana* gene bank numbers for the fourteen putative AC proteins, followed by their amino acid sequences suspected to be their AC catalytic centres, and the names to which each protein was bioinformatically inferred (annotations).

*Proteins that contain the relaxed AC search motif present in the *G. max* while the rest contain the stringent motif.

1.2.2 Plant adenylyl cyclases

In all cellular systems including plants, the AC system is generally represented in two main forms; the transmembrane (tmAC) form (Kamenetsky *et al*., 2006) and the soluble (sAC) form (Lomovatskaya *et al*., 2007). While all tmACs are strictly activated by forskolin and the fluoride ion (Rail and Sutherland, 1958; Robison *et al*., 1968; Wuttke *et al*., 2001), all sACs are on the other hand, specifically activated by the calcium and bicarbonate ions (Garty and Salomon, 1987; Carricarte *et al*., 1988; Visconti *et al*., 1990; Chen *et al*., 2000). Furthermore, while the activity of all tmACs can flexibly depend on either the Mg^{2+} or Mn^{2+} metal ion as a co-factor (Robison *et al*., 1968), the activity of sACs is strictly dependent on the Mn^{2+} metal ion only (Braun, 1974; Braun, 1975). Principally, the fluoride ion non-specifically influences the activity of tmACs by targeting and modulating the nucleotide-binding site on the α–subunit of their G-protein (Howlett *et al*., 1979; Northup *et al*., 1983) and the sensitivity of tmACs to the fluoride ion is known to be relatively low and therefore, millimolar concentrations of this ion are usually required for enzymatic activation (Bigay *et al*., 1987). Furthermore, the activation of tmACs by the fluoride ion critically requires the presence of trace amounts of aluminium, a very important requirement that has long been overlooked because at the concentration of fluoride ion commonly used in laboratories, the solutions generally etch aluminium from the used glassware, which is relatively adequate for successful experimentations (Sternweis and Gilman, 1982; Bigay *et al*., 1987). On the other hand, the activities of all sACs are specifically mediated via the calcium-modulating protein (Kamenetsky *et al*., 2006), suggesting that their cAMP-dependent biological functions may specifically be mediated by calmodulin.

1.2.3 Adenylyl cyclase activity in legumes

In legumes, cAMP has been reported to have an important role in the sequence of biological events that regulate nodule formation and functioning (Gehring and Turek, 2017). Upchurch and Elkan (1978) proposed the involvement of cAMP in the regulation of ammonia assimilation in the rhizobia, *Bradyrhizobium japonicum.* Their proposal led to the isolation of genes encoding ACs from the *Rhizobium meliloti* (Beuve *et al*., 1990) and *B. japonicum* (Guerinot and Chelm, 1984). Further work led to the characterisation of cyclic PDE, which degrade cAMP to 5' AMP in *B. japonicum* (Catanese *et al*., 1989). This group of researchers were able to detect AC and cAMP-PDE in *B. japonicum* bacteroids (Catanese *et al*., 1989). The authors also reported that AC activity increased with nodule age whereas membrane bound cAMP-PDE activity decreased (Catanese *et al*., 1989). This then led to the realisation that the possible presence of cAMP in nodules implied its crucial role in symbiosis since AC activity increased with nodule age. Work by Terakado *et al.* (1997) reported the presence of significant levels of cAMP in cultured rhizobia strains and in symbiotic nodules of certain legumes. The nodules of soybean that were inoculated with *B. japonicum* contained 7 pmol g^{-1} f.wt of cAMP and therefore, this implying that cAMP can regulate nodule formation and function in legumes. Apart from being present in nodules of leguminous plants, the results also indicated occurrence of cAMP at 5-10 pmol g^{-} f.wt in leaves of *Phaseolus vulgaris* and *Vigna radiate* (Terakado *et al*., 1997). However, in soybean organs such as stems and roots, cAMP was not detected. These last findings were also very consistent with the results of Yunghans and Morre (1977), who reported inability to detect AC activity in the soybean hypocotyl. Notably and apart from the aspect of non-detection of cAMP in soybean stems, roots and hypocotyls, the fact that cAMP was systematically detected in nodules and leaves somewhat strongly points out to the existence of a cAMP-dependent signalling system in this important legume.

Apart from its reported role in symbiosis in legumes, cAMP has also been implicated in the induction of defense related genes in *P. vulgaris* (Bolwell, 1992). Most plants produce reactive oxygen species (ROS) as a defense response during pathogen attack and in *P. vulgaris*, elicitation with a pathogen lead to an oxidative burst (Wojtaszek and Bolwell, 1997). Forskolin, a potent activator of AC, has also been shown to elevate cAMP levels in *P. vulgaris* (Bindschedler *et al*., 2001). In that study, it was shown that addition of forskolin to *P. vulgaris* cells after elicitation with *Colletotrichum* resulted in the production of hydrogen peroxide (H_2O_2) and in the absence of the elicitor, no H_2O_2 production was stimulated. This clearly indicated the role forskolin (and certainly cAMP) plays in modulating signalling mechanism in legumes. This study by Bindschedler *et al.* (2001) certainly provided evidence that the activation of ACs in legumes is part of a signalling pathway that modulates the production of H_2O_2. The production of H_2O_2 on the other hand, is compatible with the induced production of ROS by forskolin in *P. vulgaris* and the increased level of cAMP in the same plant (Bolwell, 1992). The evidence provided above supports the possibility that cAMP is a component of signalling pathway in legumes that leads to the production of ROS in response to pathogen attack. However, nothing is known about the elicitor recognition through putative receptors or whether such receptors are directly coupled to a G-protein that activates ACs (Bindschedler *et al*., 2001). Nonetheless, it has been proven that in *P. vulgaris*, ACs are activated after pathogen attack, since cAMP increased after elicitation. Therefore, it is possible to assume that cAMP may directly act on cyclic nucleotide gated channels (CNGCs) and induce an increase in concentration of the cytosolic Ca^{2+}. This accumulation of the cytosolic Ca^{2+} is believed to be primary signal that is very important for subsequent downstream events. Some of the downstream events that follow include the production of ROS (Rajasekhar *et al*., 1999; Blume *et al*., 2000; Grant *et al*., 2000a) and the induction of defense-related genes (Blume *et al*., 2000).

1.2.4 Adenylyl cyclases and the soybean plant

Initial work in *G. max* by Yunghans and Moore (1977) had totally excluded any possibility for the existence of ACs and/or AC regulated systems in this plant. These researchers were attempting to search for AC activity in this legume using the hypocotyls of its etiolated stems. They employed ion exchange chromatography and radioactive scintillating techniques to assess for the intended AC activity, whereby 2 mM of ATP that contained the radioactive ATα^{32}P compound was used (Yunghans and Moore, 1977). In addition, they also performed an isotope dilution technique (a cAMP binding assay) in which a non-radioactive cAMP was used to compete for binding sites on a specific binding protein with a radioactive cAMP (competitive displacement). Finally, the anticipated cAMP was then measured using an assaying kit from Amersham Corporation (Arlington Heights, USA). Lastly, AC cytochemistry tissue localisation was also carried out in which the soybean hypocotyl tissues were placed in an incubation medium, fixed and then examined under electron microscopy. However, all these three different techniques nonetheless failed to detect any purported AC activity and/or cAMP in the soybean plant.

These unsuccessful findings from this work can be firmly attributed to the very low sensitivity limits of the techniques used to detect the AC and/or cAMP activities in plants during the seventies (Lemtiri-Chlieh *et al*., 2011). There are reports that the competitive displacement techniques and cAMP binding assays are normally hindered by sensitivity limits (Lemtiri-Chlieh *et al*., 2011), and hence the cAMP and ACs levels that could have been produced in this study were possibly too low for the detection limits. Nevertheless, modern technological advancement has now afforded techniques with increased sensitivity detections, which have afforded for both the accurate and precise *in vivo* and *in vitro* measurements of cAMP and AC activities in plants. Such techniques include the enzyme immunoassay (EIA) (Lomovatskaya *et al*., 2011; Ito *et al*., 2014; Al-Younis *et al*., 2018), mass spectrometry (Al-Younis *et al*., 2015; Bianchet *et al*., 2019) and radioimmunoassay, where radioactivity is measured in scintillation counters (Świeżawska *et al*.,

2014). Another contributory factor to the negative findings for Yunghans and Moore (1977)'s could be that in the seventies, scientists tended to strongly rely on the usage of similar analytical techniques for plants and animals (Lemtiri-Chlieh *et al*., 2011) yet these two organisms have relative very distinct and contrasting biochemical characteristics. More to this, the authors used an etiolated soybean plant, which implies that the plant had been grown in partial or a complete absence of light, and hence one can be forced to conclude that this could also have had affected the plant's developmental and physiological settings and thus resulting in such unapparent outcomes.

In line with the continued effort to search for ACs in soybean, subsequent work by Ito *et al*. (2014) and Świeżawska *et al*. (2014), provided some practical insights that there could be ACs and/or AC signalling systems in this plant. These authors independently performed some phylogenetic analysis of ACs in higher plants. Ito and his team studied the role played by a novel AC enzyme in *N. benthiana* (NbAC; ACR77530) and specifically, during the onset and subsequent establishment of the tabtoxinine-β-lactum (TβL) mediated cell death of the wild fire disease. In their efforts, they performed a phylogenetic analysis of ACs in higher plants using the Clustal W alignment (http://clustalw.ddbj.nig.ac.jp/) and their results revealed a similarity in protein/amino acid sequence between their own AC (NbAC) and an uncharacterised *G. max* AC (GmUCP), accession number XP_003529590. Parallel but equally important to the effort, Świeżawska and his team and while experimenting on molecular cloning and functional characterisation of yet another novel AC in *H. hybridum*, performed a protein/amino acid sequence alignment analysis of ACs in higher plants using the PRALINE software (http://www.ibi.vu.nl/programs/paralinewww) and their findings showed that these putative AC amino acid sequences are indeed strongly conserved in higher plants. A subsequent phylogenetic analysis using BLAST and based on the deduced protein sequence of HpAC1 showed a relatively high degree of similarity to other putative higher plant ACs, with another *G. max* uncharacterised protein, accession number XP_003547191,

also included (56%). These results from two independent studies (Ito *et al*., (2014) and Świeżawska *et al*., (2014)) thus therefore, strongly affirm the likely presence of ACs and or their activity in soybean, whose exact biological roles in the legume are yet to be elucidated.

In line with the continued search for ACs in soybean, our team herein used Cluster O alignment to carry out a preliminary search in the *G. max* genome with an AC search motif (Figure 1.1) that was previously used by Gehring (2010) to identify 14 putative AC candidates in *A. thaliana* (Gehring, 2010) and later confirmed to be catalytically functional by several authors (Ruzvidzo *et al*., 2013; Ito *et al*., 2014; Al-Younis *et al*., 2015, 2018; Chatukuta *et al*., 2018; Bianchet *et al*., 2019). The search indicated presence of this putative AC catalytic motif in the *G. max* genome and specifically, within its XP_003529590 protein, from the 325 to the 337 amino acid sequence of the protein (Figure 1.2).

```
MQVFSNARQA SRLLLSPHLR SSEAPHSTAL SLFSGLTQRD SRPVNTDPIQ    50
CFLSKAFYSS GVGTVEATPS EDVKELYDKM LDSVKVKRSM PPNAWLWSMI   100
ANCKHQPDIR LLFDILQNLR RFRLSNLRIH DNFNCNLCRE VAKACVHAGA   150
LDFGMKALWK HNVYGLTPNI ASAHHLLTNA KNHNDTKLLV EVMKLLKKND   200
LPLQPGTADI VFSICYNTDD WELINKYAKR FVMAGVKLRQ TSLETWMEFA   250
AKRGDIHSLW KIEKLRSNSM KQHTLITGFS CVKGLLLERK PSDAVAVIQV   300
LNQTLSDTKK SGIKGELQKL VSEWSLEVIK HQKEEDRKAL AASLKSDILV   350
MVSELLSMGL EANVSLEDLD RKEDIPQ                            377
```

Figure 1.2: Amino acid sequences of the XP_003529590 protein from *G. max*. The annotated AC catalytic motif is shown in bold and green highlight while its priming sites are shown in bold and yellow highlight.

In this regard therefore, this work was then set to recombinantly clone the AC-containing region of the XP_003529590 gene followed by its expression and extensive functional characterisation. This study was anticipated to be of paramount importance since this protein was going to be the first ever AC molecule to be identified in soybean. Additionally, the work would also assist in better understanding the possible biological roles of this novel protein in soybeans, particularly in critical cellular processes of growth, development and tolerance to the various environmental stress factors.

CHAPTER 2

PRELIMINARY BIOINFORMATIC ANALYSIS OF XP_003529590 PROTEIN

Abstract

A preliminary bioinformatic investigation of the novel adenylyl cyclase (AC) protein encoded by the gene XP_003529590 in *Glycine max* was performed. Work of this chapter aimed at understanding the annotation of the *G. max* XP_003529590, its expression profile in soybean and its protein secondary structure through various web-based computational tools. The preliminary investigation was vital as it provided baseline information required to continue with the molecular and functional characterisation of the XP_003529590 gene. Gene annotation, expression profile, protein modelling and protein secondary structure, and localisation predictions were performed. The results of the gene annotation revealed that the XP_003529590 gene name or identifier is Glyma.07G251000 (Glyma_07G251000) and is primarily located in the mitochondrion of the soybean plant cells. Expression profile analysis using Genevisible showed that this soybean gene is primarily expressed during the primary growth stages in the root apical meristems, root tips, shoot apical meristem and root hairs. Protein secondary structure prediction revealed that the XP_003529590 expression product is an alpha helical pentatricopeptide repeat (PPR) protein that is primarily involved in RNA binding. The DISOPRED3 webserver analysis tool also showed the helical structure is disordered, a characteristic that aids in the nucleic acid and compound binding ability of the protein. Overally, the computational analyses have indicated that this novel XP_003529590 product is a PPR protein involved in RNA and calcium binding.

2.1 Introduction

Arabidopsis thaliana is a model system that is used for identifying genes and their function, thus the creation of The Arabidopsis Information Resource (TAIR), an online database for genetic and molecular biology data for *A. thaliana* (Lamesch *et al*., 2012). The TAIR database serves as a reference for comprehending plant gene functions, therefore, unravelling mechanisms of plant development, physiology and biochemistry. The complete genome sequence of this eudicot plant provides the foundation for more comprehensive comparisons of conserved processes in all plants and in the identification of a wide range of plant-specific gene functions, therefore, establishing rapid systematic ways to identify genes for crop improvement. An understanding of the Arabidopsis genome is the basis for understanding the biology of all plants through bioinformatics tools (Clare *et al*., 2006), and to develop a direct and efficient access to understanding plant development and environmental responses. This also permits the assessment and understanding of the structure and dynamics of plant genomes.

In public databases, there are over six million protein sequences that have been deposited and the number continues to grow, however, the number of experimentally determined protein structures does not match the deposits (Kelly *et al*., 2015). As a result, this has led to the development of computational methods that are able to use protein sequences to predict protein structure (Kelly *et al*., 2015) and function (Makrodimitris *et al*., 2018). Furthermore, the complete sequence of the *A. thaliana* revealed thousands of unsuspected genes many of which were not ascribed putative function (Lurin *et al*., 2004) but through bioinformatics analyses, most of the genes have been studied. Making use of bioinformatics or computational gene function predictions is now a well-recognised field (Attwood, 2000; Hvidsten 2001; Syed and Yona, 2003). Therefore, access to primary DNA sequence is a fundamental resource in plant biology. So far, more than twenty plant genome studies have been completed and there are more than two hundred ongoing plant genomic studies (Martinez, 2013). As such, data mining has been equally and successfully used to predict

functions of genes in *Saccharomyces cerevisiae, Escherichia coli* and *Mycobacterium tuberculosis* (King *et al*., 2001; Clare and King, 2003), and the majority of these predictions being confirmed (King *et al*., 2004).

The success of bioinformatic tools in predicting gene/protein functions involved in AC and/or cAMP activities in plants have been demonstrated in predicting some of the Arabidopsis essential molecules such as the K^+-uptake permease 7 gene (*AtKUP7*) (Al-Younis *et al*., 2015), the clathrin assembly gene (*AtCIAP*) gene (Chatukuta *et al*., 2018) and the leucine rich repeat gene (*AtLRRAC1*) (Bianchet *et al*., 2019). It is also through bioinformatics analysis while screening *A. thaliana* proteins in mitochondria and chloroplasts that the pentatricopeptide repeat (PPR) gene family was discovered, which is involved in gene expression and restoration of male fertility (Ruzvidzo *et al*., 2013; Tan *et al*., 2014). Bioinformatic analyses have been used to understand genomic and genetic data on the expression, localisation and general function of most of the PPR proteins (Lurin *et al*., 2004) and have revealed that PPR proteins play essential roles in mitochondria and chloroplasts, through binding to organellar transcripts (Lurin *et al*., 2004).

Bioinformatics together with proteomic analysis, have been used to understand the Arabidopsis mitochondrial proteome (Heazlewood *et al*., 2003), further emphasising the role of computational tools in the identification of PPR motifs in plants (Gattiker *et al*., 2002; Finn *et al*., 2011). In protein structure predication, most of the bioinformatics tools currently in use today, rely on a method to compare a protein sequence of interest with large database of sequences. This helps in the construction of an evolutionary or statistical profile of the query sequence and to subsequently scan its profile against a database of profiles of known structures (Kelly *et al*., 2015). Therefore, an alignment between the two sequences, the unknown and known structures can be made and be used to construct a model of the unknown on the basis of the known structure. Alpha-helices, beta-strands and coils can be predicted using bioinformatics tools and the protein structure can

then be constructed from protein sequences. Knowledge of a protein structure ultimately sheds light into understanding the characteristics of novel proteins.

Changing climatic conditions have been predicted to reduce soybean yield such that the production yield forecasts usually consider challenges of extreme weather such as heat, cold, drought, floods and UV stress (Deshmukh *et al*., 2014). This means that it is crucial to extensively study soybean genome in order to identify and understand genes that can be used in producing cultivars that are stress tolerant. It has been reported that the soybean genome has undergone at least two whole genome duplication events (Shoemaker *et al*., 1996), approximately 59 and 13 million years ago respectively, resulting in nearly 75% of the genes in multiple copies (Walling *et al*., 2006; Gill *et al*., 2009). The genome duplication resulted in the generation of many duplicated genes that ultimately gave rise to a large number of new novel unique genes (Lynch and Conery, 2000) within the legume, hence the need to wholly sequence the *G. max* genome arose. The whole genome sequence for *G. max* Williams 82 Glyma1.01 was completed and published in 2010 (Schmutz *et al*., 2010) and the genome sequence was used in the study of gene structure (Upchurch and Ramirez, 2011) and identification of genes (Xia *et al*., 2012; Cook *et al*., 2014) among other uses. The genome sequence of soybean therefore, has provided a platform for further analysis and study of *G. max* genes, considering the genome duplication events that might have led to the emergence of novel genes, possibly ACs. It is through bioinformatic analyses that led to the discovery of a 2.55-fold duplication in the legume (Turner *et al*., 2012) compared to a 1.55-fold duplication in *A. thaliana* (Grant *et al*., 2000b).

2.2 Materials and methods

Preliminary bioinformatics analyses of the XP_003529590 gene were conducted to understand the protein coding gene, and to predict the structure and expression profiles of the soybean AC. This

was particularly important as it provided a baseline understanding of the characteristics of this AC gene prior to molecular and biochemical analyses. To achieve this various web-based servers and computer programs were utilised.

2.2.1 Gene annotation of the XP_003529590

Several web platforms were used to gain an understanding of the gene annotations of the *G. max* XP_003529590. These included the National Centre for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov), UniProt Knowledgebase (UniProtKB) (www.uniprot.org) (Chen *et al.*, 2017), EnSemblGenome (www.ensemblgenome.org) (Kersey *et al*., 2018) and PLAZA 4.0 (https://bioinformatics.psb.ugent.be/plaza) (van Bel *et al.*, 2018).

2.2.2 Expression profile of the XP_003529590 protein coding gene in soybean tissues

To predict the expression profile of the XP_003529590 gene in *G. max*, the Genevisible Affymetrix Soybean Genome Array platform from Genevestigator (https://genevestigator.com) (Zimmermann *et al*., 2004) was used. The platform tested 54 tissues from the soybean plant and 10 top tissues in which the XP_003529590 gene is expressed. The data was expressed as expression level on log2 scale. The determination of the expression profiles was an important aspect in the ultimate isolation of the targeted AC-containing fragment of the XP_003529590 gene in soybean.

2.2.3 Protein modelling and structure prediction

The Protein Homology/analoY Recognition Engine V 2.0 (Phyre2) (http://www.sbg.bio.ic.uk) (Kelly *et al*., 2015) was used to predict the secondary structure of the protein product coded for by the XP_003529590 gene. The webserver used advanced remote homology detection methods to build the sought XP_003529590 protein 3D model. XtalPred (http://ffas.burnham.org/XtalPred) (Slabinski *et al*., 2007) was used to provide additional information on the expressed XP_003529590 protein secondary prediction. An intrinsic disordered region prediction of the XP_003529590 protein sequence was then performed using DISOPRED3 (http://bioinf.cs.ucl.ac.uk/disopred) (Jones and Cozzetto, 2015). This platform was used to identify intrinsically disordered regions (IDRs) and the protein binding sites present within them. A further analysis to predict protein-protein, protein-nucleic acid, protein-compound and protein-metal ion binding sites was further performed using GenPRoBis (http://genprobis.insilab.org) (Konc *et al.*, 2017).

This web server had the potential to map sequence variants to protein structures from the Protein Data Bank (PDB) and to protein-binding sites. The protein-compound binding sites were understood to include the concept of glycosylation and other post translational modification sites. In this scenario, binding sites were defined through local structural comparisons of the whole protein structures by use of the Protein Binding Sites (ProBIS) algorithm and transposition of the ligands from similar binding sites found on the XP_003529590 query protein using ProBiS-ligands. The binding sites were generated as three-dimensional grids covering the space occupied by predicted ligands. TargetP 1.1 (http://www.cbs.dk) (Emanuelson *et al*., 2000), server was used to predict the subcellular location of the XP_003529590 protein. The location task was based on the predicted presence of the N-terminal amino acid sequences. If the sequence contained cTP, it implied a chloroplast transit peptide, mTP when the sequence possessed a mitochondrial targeting peptide, and when located in any other organ, it would indicate SP, thus implying that the protein

sequence is a signal peptide. The measure of prediction was based on reliability class and presented by RC, which is a measure of the size of difference between the highest and the second highest output scores (Emanuelson *et al*., 2000).

2.2.4 Protein-protein interaction of the XP_003529530 gene

To assess the interaction/associations of the XP_003529530 protein with other proteins in *G. max*, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (http://string-db.org) (Szklarczyk *et al*., 2017) was used. In this case, the use of STRING database was ideal as interactions between proteins help to describe and narrow down a protein's function (Szklarczyk *et al*., 2017).

2.3 Results

2.3.1 Gene annotations of the XP_003529590 AC gene

Results from the web platforms managed to reveal the following information on the *G. max* XP_003529590 gene. Its entry name in UniProtKB is I1KN29_SOYBN and its taxonomic identifier being 3847. The gene name often referred to as gene symbol in NCBI is 100792566 and is described as an uncharacterised LOC100792566. The locus tag (sometimes referred to as gene name or gene identifier) of the XP_003529590 gene is Glyma.07G251000 often written as Glyma_07G251000 and its mRNA ID is XM_003529542.3 as provided for in UniProtKB, EnsemblGenome, NCBI, Plaza 4.0 and other platforms not included here. However, the gene's alias is Glyma07G38080 as is in Plaza 4.0 and STRING databases. However, the alias Glyma07G38080 produces a 406 amino acid protein and therefore, not the XP_003529590 protein

that was intended for in this study. Plaza 4.0 has rather identified At4g15640 from *A. thaliana* as the best ortholog for Glyma.07G251000.

The XP_003529590 gene is located on chromosome 7 as follows; 7:42892578-42897371 Forward strand (EnsemblGenome) Chr07: 42893167-42897348: positive (Plaza 4.0) and chromosome 7 NC_038243.1 (NCBI). Information obtained from EnsemblGenome showed that the transcript for the gene of interest (GOI) is KRH50918, which correctly corresponds to 377 amino acids and protein size of 24.79 kDa (Figure 2.1). However, the transcript KRH50917 corresponds to the 406 amino acids protein (Glyma07G38080). The protein coding gene Glyma.07G251000 contains 11 exons, the transcript length is 1 746 against a translation of 377 residues as shown in Figure 2.1.

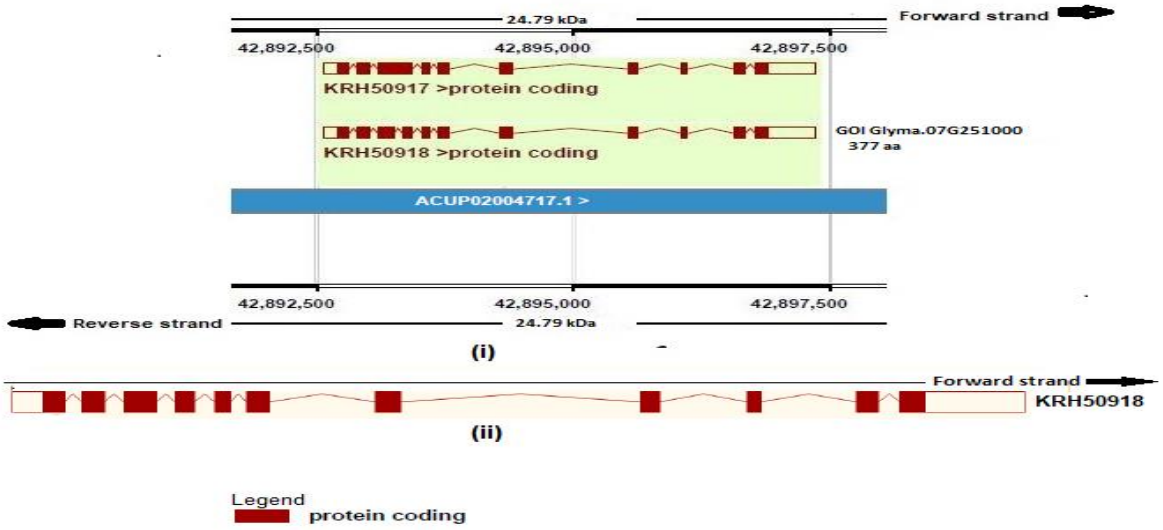


Figure 2.1: Gene structure of Glyma.07G251000 from EnsemblGenomes (Kersey *et al*., 2018). Transcript KRH50917 in (i) is an isoform of transcript KRH50918, which represents the AC Glyma.07G251000. (ii) A more detailed representation of the gene architecture with an exon count of 11, however, the isoform has 10 exons and 406 residues.

2.3.2 Expression profile of the XP_003529590 in soybean tissues

Across the tissues tested by Genevestigator, it was noted that the XP_003529590 gene is mostly expressed in the early primary growth stages of the soybean plant, primarily in the root apical

meristems, root tips, shoot apical meristem and root hairs as is shown in Figure 2.2. However, there is moderate expression of the gene in the inflorescence, seed, nodule and the pod.

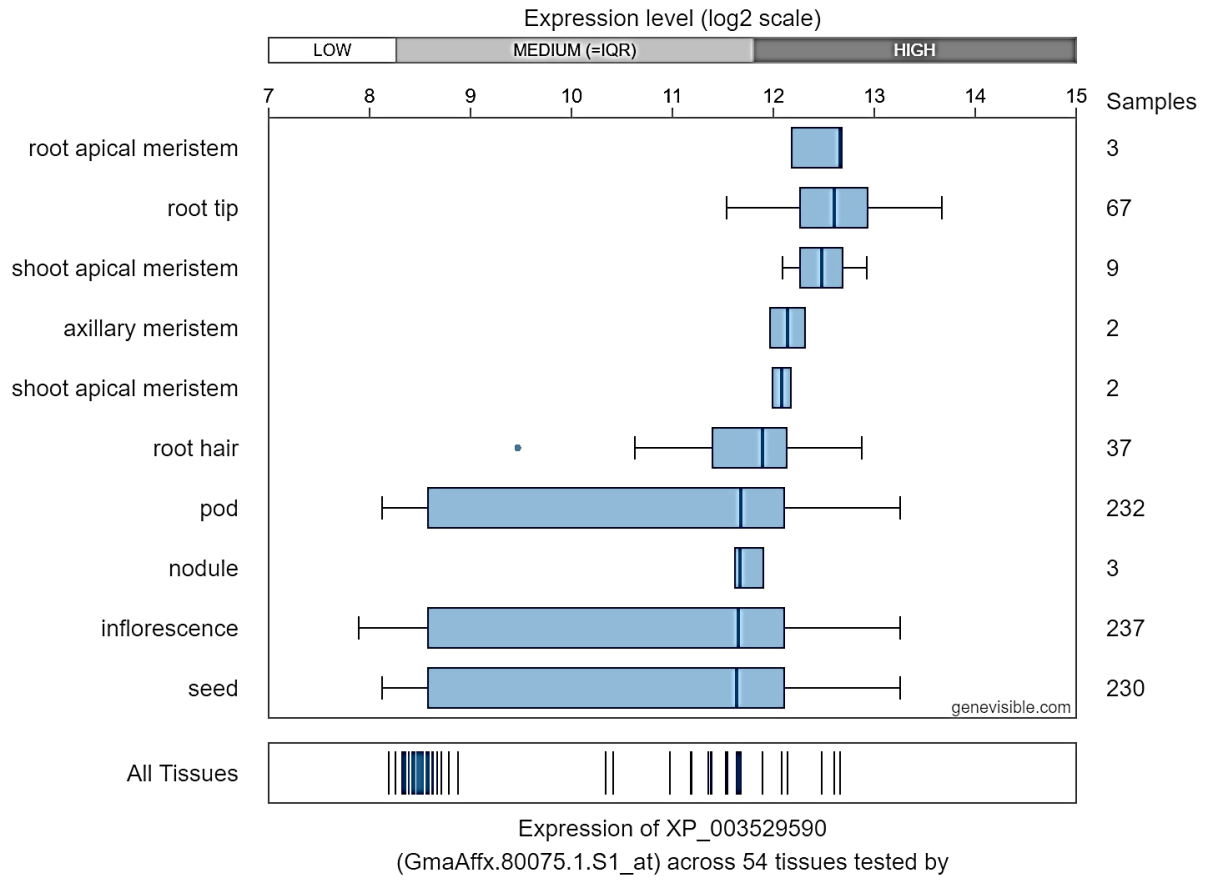


Figure 2.2: Expression log2 scale of the XP_003529590 gene across the various tissues of *G. max* as is tested by Genevestigator. The gene is mostly expressed in the root apical meristem, root tip and shoot apical meristem, the young developing organs of the legume and the early stages of development of the soybean plant as well in its reproductive phase as seen in the gene expression in the pod, inflorescence and the seed.

2.3.3 Protein modelling, prediction and analysis

Secondary structure prediction of the protein product generated by the XP_003529590 gene using the Phyre 2 Psi-pred 2.5 secondary protein structure prediction showed that the expressed product is largely an alpha helical protein as shown in Figure 2.3. The expressed product was predicted to

be a PPR protein. Analysis results indicated a 99.1% confidence similarity to a proteinaceous RNase PPR protein from *A. thaliana* At2g32230 (PRORP1) with 171 residues (45% of the sequence) and being identical to the At2g32230 PPR protein involved in RNA binding.

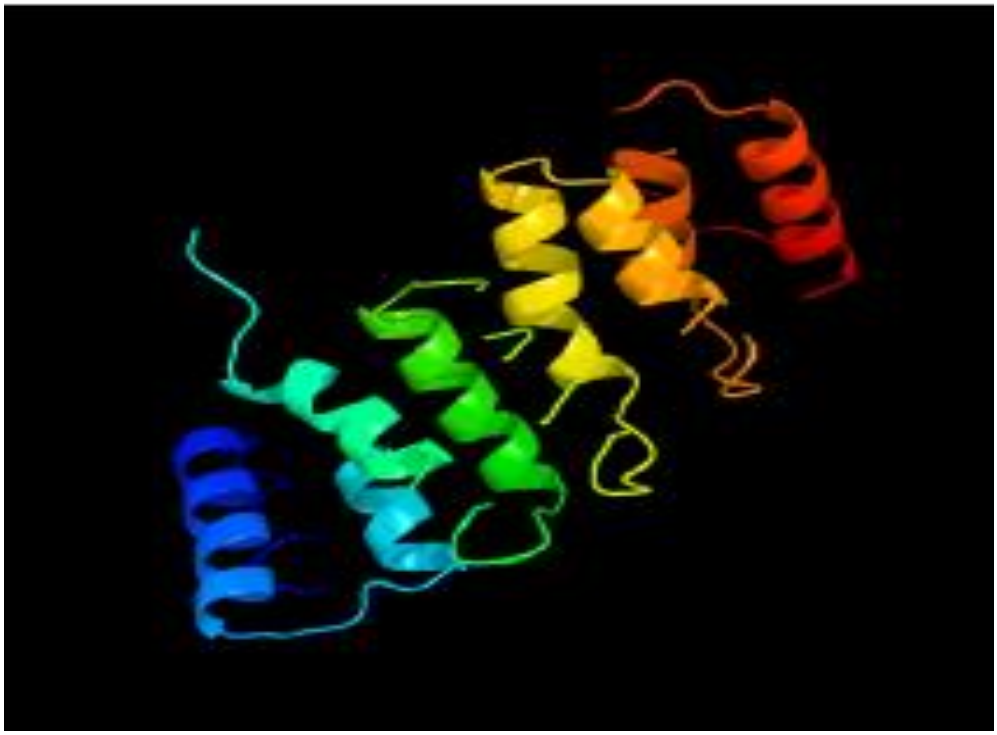


Figure 2.3: Predicted alpha helix structure of the XP_003529590 protein product. 72% of the amino acid sequences depict the α- helix, 0% depict the β-strand and 19% depict the disordered structures. Image coloured by rainbow means the N→ C terminus direction. (Model dimensions (Å): X:38.503 Y:60.189 Z:49.746.

The general protein features from XtalPred showed that the XP_003529590 protein has a gravity index of -0.27, an isoelectric point of 9.10 and a stability index of 41.99. The predictions further characterised the novel AC protein as possessing no transmembrane helices and thus indicating that the XP_003529590 protein is not a signal peptide. The amino acid residues making the α-helical structure of the XP_003529590 protein are shown in Figure 2.4 below.

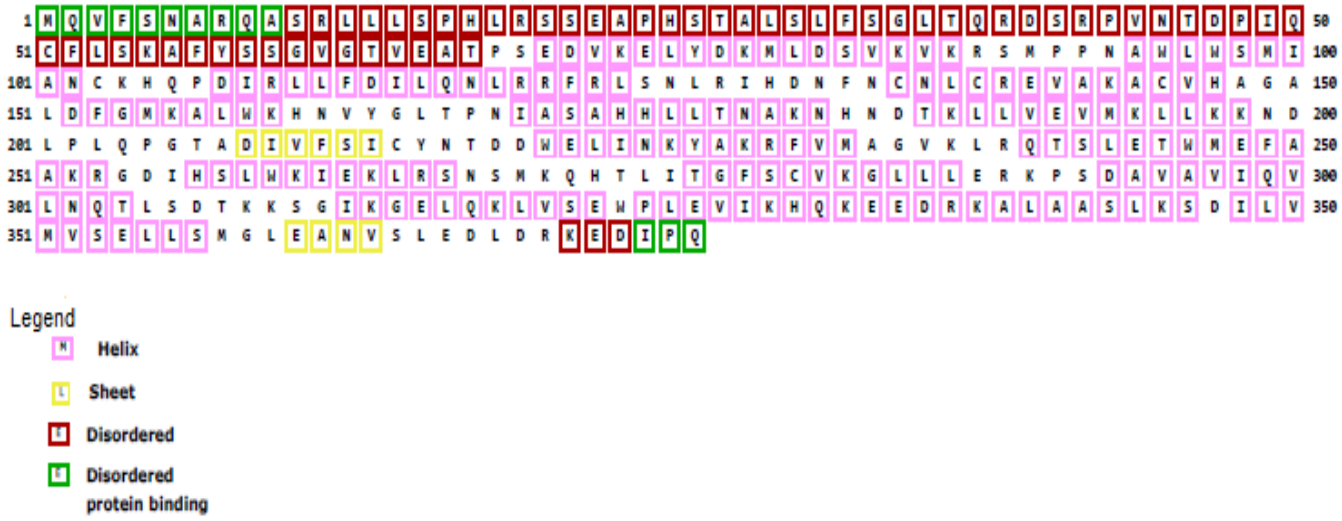


Figure 2.4: Secondary structure map from DISOPRED3, indicating the amino acid residues making up the secondary structure of the soybean XP_003529590 protein. The helical amino acid residues are in pink, disordered regions in red and the disordered amino acids involved in binding are in green.

Results from DISOPRED3 have predicted that the longest disordered region in the protein structure is made up of 36 amino acids (Leu 13: Disordered to Pro 48: Disordered) as is shown in Figure 2.4. However, there are some amino acids that are described as Helix: Disordered such as Ser 11, Arg 12 and Ile 49 just to mention but a few. Slabinski *et al*. (2007) described such amino acids as low complexity amino acids. It is interesting to note that the intrinsic disordered protein binding regions of the XP_003529590 protein start with the first amino acids Met 1, Gln 2, Val 3, Phe 4 and Ser 5, though low complexity amino acid (Asn 6 to Ala 10) are also described as Helix: Disordered: Binding. The intrinsic disordered binding regions on the XP_003529590 protein are also found on the last three amino acids Ile 375, Pro 376 and Gln 377. The instrinsic disorder profile of the XP_003529590 protein is shown in Figure 2.5 below.

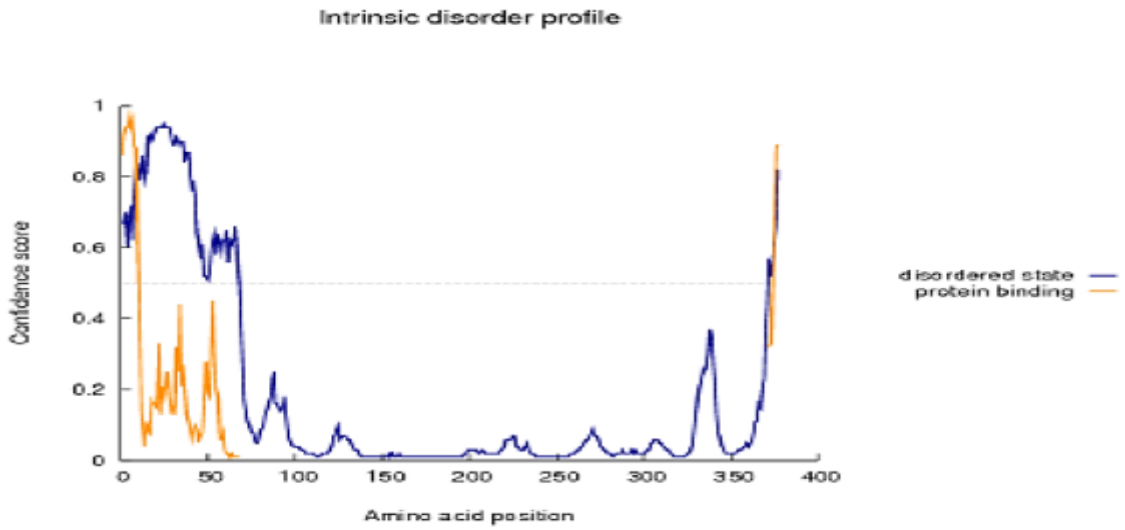


Figure 2.5: The intrinsic disorder profile of the XP_003529590 protein.

Binding sites of the XP_003529590 protein probe from GenPRoB and through BLAST identified seven major possible binding sites, which are shown in Table 2.1 below. Generally, the XP_003529590 protein is predicted to have nucleic acid- and compound-binding sites as is shown in Figure 2.6 respectively. The analog ligand (3X1L) for the nucleic acid-binding sites has been classified as an RNA-binding site (Osawa *et al.*, 2015), and the K21 ligand (1HAK) as a calcium-binding site (Kaneko *et al.*, 1997).

Table 2.1: BLAST result of binding surface sites on the XP_003529590 protein.

PDB ID	Ligand Analog	Binding Site
2oebA (6.6)	3X1L	Nucleic acid
4hreD (7.3)	K21	Compound
3x1lG (6.6)	3X1L	Nucleic acid
4hreC (7.)	K21	Compound
4hreB (7.3)	K21	Compound
3x1lF (6.6)	3X1L	Nucleic acid

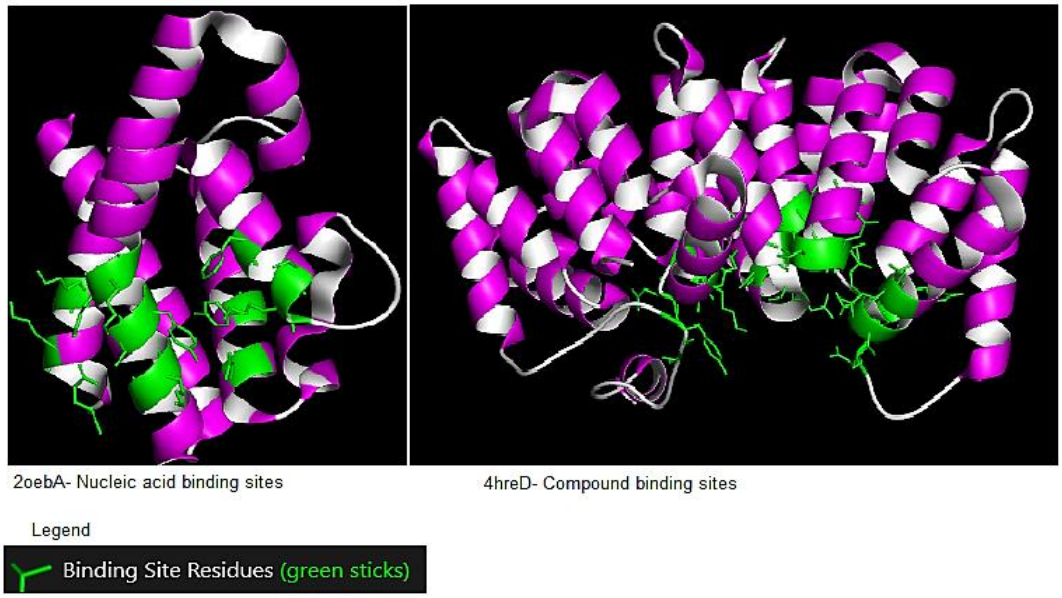


Figure 2.6: 3-D structural prediction of the XP_003529590 protein binding sites. 2oebA is showing binding residues for nucleic acids with the ligand analog 3X1L, which is classified as an RNA binding protein (Osawa *et al*., 2015). 4hreD is a typical compound binding site, whose analog K21 (4-[3-{1-(4-benzyl)piperodinyl}propionyl]-7-methoxy-2,3,4,5-tertrahydro-1,4-benzothiazepine) is 1HAK, which is classified as a calcium binding site (Kaneko *et al*., 1997).

2.3.3.1 Localisation prediction

The targetP v1.1 prediction result for the XP_003529590 protein showed an RC value of 5 and loc M as is shown in Table 2.2. This implies that the protein encoding AC XP_003529590 gene is located primarily in the mitochondrion.

Table 2.2: Values for localisation prediction of the XP_003529590 protein. Len indicates the length of the protein in amino acid residues, loc shows the primary location and RC is the reliability class.

Name	Len	cTP	mTP	sP	other	Loc	RC
I1KN29_SOY	377	0.520	0.614	0.024	0.012	M	5

2.3.4 Protein-protein interaction of the XP_003529590 in *G. max*

No results could be obtained for the 377 amino acids XP_003529590 protein in *G. max*. Only data available in STRING network was for the 406 amino acids alias Glyma07G38080 *G. max* protein, hence the search was abandoned.

2.4 Discussion

Gene annotation search for the protein encoding XP_003529590 gene has indicated that its gene identifier/gene name is Glyma.07G251000. Therefore, the XP_003529590 has sometimes been referred by its gene identifier Glyma.07G251000 (often written Glyma_07G251000) during the development of this thesis. Secondary structure prediction of the XP_003529590 indicated that its expressed product is an α-helical RNA binding PPR protein and that the gene is a structural homolog of the At2g32230 (PROteinaceous RNase P 1, AtPRORP1) from *Arabidopsis thaliana*. This Arabidopsis At2g32230 is a gene that primarily encodes a protein-only RNase P, involved in 5′ cleavage of tRNAs and RNA-like structures. The gene is involved in the maturation of plant mitochondrial mRNAs (TAIR: https://www.arabidopsis.org), therefore, is important in tRNA 5′ leader removal and tRNA processing. Mutants of this *A. thaliana* gene AtPRORP1, resulted in the drastic reduction in the levels of mature plastid tRNA-Phe (GAA) and tRNA-Arg (ACG), thereby limiting plastid gene expression (Howard *et al*., 2016).

Most *A. thaliana* PPR proteins have been reported to participate in a wide range of developmental processes through their involvement in RNA metabolism and posttranscriptional regulation of plant organelles (Laluk *et al*., 2011). This implies that the XP_003529590 gene plays crucial roles in regulating most of the soybean developmental process and maturation of the legume's mitochondrial mRNAs, thus promoting gene expression. An Arabidopsis AtPRORP1 gene has

two domains; the pentatricopeptide repeat (IPR002885) and the multi-antimicrobial extrusion protein: IPR002528 (Fujii *et al*., 2016). Considering that this gene has a 99.1% confidence structural similarity to XP_003529590, it is most likely that the XP_003529590 gene also possesses these domains. In *A. thaliana*, mitochondrial PPR domains have been reported to function in plant defense and abiotic stress responses (Laluk *et al*., 2011) and the multi-antimicrobial extrusion (MATE) proteins have been associated with plant defense responses via pathogen induced salicylic acid (SA) (Nawrath *et al*., 2002; Hammes *et al*., 2005; Zeng *et al*., 2011;). In *G. max*, the MATE family proteins have been associated with extrusion of xenobiotic compounds, responses to abiotic stresses and regulation of disease resistance (Liu *et al*., 2016). Therefore, the implication is that the PPR-like XP_003529590 gene could be involved in plant defense response against pathogens and abiotic stresses.

The α-helical structure of the XP_003529590 gene is consistent with reports elsewhere of the secondary structure of PPR proteins. They are described as a family of α-helical repeat proteins (Small and Peeters, 2000; Schmitz-Linneweber and Small, 2008) and their motifs are involved in direct RNA binding activity (Pfalz *et al*., 2009; Cai *et al*., 2011; Prikryl *et al*., 2011; Kobayashi *et al*., 2012; Okuda and Shikanai, 2012). A further analysis of the XP_003529590 potential binding sites using GenPRoBis, confirmed earlier findings that the gene is indeed involved in nucleic acid binding. GenPRoBis data has revealed seven binding surfaces on the *G. max* gene, which are primarily involved in nucleic acid and compound binding. ProBiS-ligands, a platform available in the GenPRoBis server revealed two types of ligand homologies that are able to bind on the identified binding surfaces, which are 3X1L and 1HAK (an example of K21). These ligands are classified as RNA binding and calcium binding respectively, confirming our earlier prediction that the alpha-helical XP_003529590 gene is primarily involved in RNA binding,

Conformational information to further strengthen the notion that the XP_003529590 protein is not a transmembrane signal peptide were from XtalPred and DISOPRED3, which further established the helical nature of the XP_003529590 protein. However, the regions of the XP_003529590 protein that lack the α- helical structure (known as the disordered region) is at 19% (Phyre2 server). The disordered regions of a protein have been reported to be involved in key cellular processes (Pazos *et al*., 2013). According to Dyson and Wright (2005) and Tompa (2005), disordered regions are characterised by stretches of charged and polar residues that almost lack hydrophobic cores, which are vital in the initiation of folding. As this maybe the case, research has indicated that these regions are associated with crucial cellular processes such as transcription, regulation and signalling cascades (Iakoucheva *et al*., 2002; Uversky *et al*., 2005). In eukaryotic organisms, it has been noted that most functions that correlate with the presence of long disordered regions, are associated with regulation via transcription and translation while functions that are related to lack of disorder, are dominated by enzymatic catalysis (Tompa, 2012). In plants, there are reports that have linked the presence of IDP or intrinsically disordered regions to a broader impact on many areas of plant biology such as abiotic stress, transcriptional regulation, light perception and development (Sun *et al*., 2013). The disordered amino acid residues from the XP_003529590 protein; Met 1, Gln 2, Val 3, Phe 4, Ser Ile 375, Pro 376 and Gln 377 have been implicated in binding and it therefore remains to assign the biological roles of the disordered amino acids in the genome of the XP_003529590 protein.

Results on the localisation prediction of the XP_003529590 protein has indicated that though found in the chloroplast and mitochondrion (0.520 and 0.614) respectively, this PPR-binding protein is primarily located in the mitochondrion (loc). Literature does support that the PPR proteins are primarily targeted to the mitochondria (Lurin *et al*., 2004) and that their roles in RNA binding include post-transcriptional processes. These processes include editing, splicing, and translation of transcripts (Schmitz-Linneweber *et al*., 2006). Therefore, the interaction of PPR

proteins with their target RNAs, undoubtedly points out to the involvement of the XP_003529590 gene in regulation of mitochondrial gene expression.

Finally, the presence of two *G. max* transcripts KRH50918 and KRH50917 on the same chromosome, 7:42892578-42897371 Forward strand (EnsemblGenome) Chr07: 42893167-42897348: positive (Plaza 4.0), could be bioinformatically providing evidence of gene duplication that possibly gave rise to the XP_003529590 gene and its annotated putative AC catalytic centre.

2.5 Conclusion

From the preliminary bioinformatic analysis of the gene of interest, it has been established that the XP_003529590 gene with a probable AC function in *G. max* is known as a Glyma.07G251000 or alternatively Glyma_07G251000, located on chromosome 7, primarily in the mitochondrion and is actually a binding protein with an alpha-helical structure. Therefore, the XP_003529590 gene will also be referred to by its gene identifier Gylma.07G251000. The protein is basically expressed in the root tip, root apical meristem, shoot apical meristem and the unifoliate leaf, and such an expression occurs during the early primary growth stages of the soybean plant, particularly during germination, the root and shoot developmental stages. After the discovery that this novel XP_003529590 protein does contain some intrinsically disordered regions, the remaining challenge is now to link these IDRs to their biological roles from the molecular to systems level in soybean plant and to establish the binding roles of the disordered and non-disordered amino acids. Lastly, it can be concluded the preliminary bioinformatic analyses has provided insights into some of the probable roles of the XP_003529590 gene in soybean that can be summarised as regulation of gene expression to events linked to root and shoot primary growth and development. The gene is also most likely to be involved in the legume's responses to biotic and abiotic stress factors during the early developmental stages of the *G. max* plant.

CHAPTER 3

PARTIAL EXPRESSION OF THE RECOMBINANT GmAC1 PROTEIN AND MOLECULAR DETERMINATION OF ITS ENDOGENOUS ADENYLYL CYCLASE ACTIVITY

Abstract

The sequencing of the soybean (*Glycine max*) genome that was completed in 2010 has provided a very good platform for plant functional genomic studies through gene cloning, protein expression and functional characterisations. However, even though the genome has been fully sequenced, a number of genes in this legume still remain experimentally unverified and among them being the XP_003529590 that harbours an annotated putative AC catalytic centre, which to date has since been functionally confirmed in a number of *Arabidopsis thaliana* proteins; AtPPR-AC, AtKUP7, AtClAP, AtKUP5 and AtLRRAC1. Thus with specific interest on this XP_003529590 gene, this study then targeted its AC-containing fragment (the GmAC1), isolated it from the *G. max* plant before cloning into a pTRcHis2-TOPO expression vector to generate a pTRcHis2-TOPO:GmAC1 expression construct. The resultant expression construct was then used to transform some chemically competent *E. coli* BL21 (DE3) pLysS cells followed by partial expression of the targeted recombinant GmAC1 protein. Finally, the ability of this truncated version of the XP_003529590 (GmAC1) protein to generate cAMP from ATP endogenously was then assessed via enzyme immunoassaying, where it emerged that the recombinant protein product could actually enhance the cellular cAMP levels of those prokaryotic expression hosts by a significant factor of at least 3.0 folds.

3.1 Introduction

The major challenge that has been brought about by genome sequencing is to assign a function to the large number of predicted genes. Expression of recombinant proteins is vital for the study of the biological functions of uncharacterised genes. The need to study and understand biological function of genes has resulted in the development of several protein expression systems aimed at producing target proteins. Some of the expression systems include bacteria, yeast, insect cells, mammalian cells and cell-free systems (Yamaguchi and Miyazaki, 2014). However, the most convenient and commonly used cell-free expression system to produce recombinant proteins is *Escherichia coli* (Baneyx, 1999; Swartz, 2001; Jana and Deb, 2005; Shrivastava *et al*., 2013). This is basically because of its well understood cell biology, simple fermentation process, easy handling and its ability to produce cost-effective and inexpensive large quantities of recombinant proteins (Gupta and Shukla, 2016). Recently, industrial enzymes and therapeutic proteins have been produced using *E. coli* as a host (Tripathi *et al*., 2009; Fakruddin *et al*., 2013; Spadiut *et al*., 2014). In this expression system, protein induction has been improved over the decades and until recently, isopropyl-β,D-1-thiogalactopyranoside (IPTG) is the most preferred inducer to promote protein expression (Briand *et al*., 2016).

The expression of most eukaryotic genes in *E. coli* has greatly increased over the years and as early as the 1980s, most genes had been expressed at high levels in the prokaryote (Schoner *et al*., 1986). Reports have implicated the success of protein expression on the use of multicopy cloning vectors that possess a strong promoter and an efficient ribosome binding site that would allow for transcription and translation of the cloned gene in *E. coli* (Crowl *et al*., 1985). In this expression system, an inducible T7 RNA polymerase is reported to be responsible for the production of high yields of recombinant proteins (Davanloo *et al*., 1984; Studier and Moffat, 1986). Basically,

during the design of the expression system, the coding sequence of the T7 RNA polymerase must be introduced into the bacterial chromosome under the control of the inducible *lacUV5* operon for transcription by the endogenous *E. coli* polymerase. Access to the T7 RNA polymerase is regulated by the *lac* repressor protein (*lacI*), which binds to the *lacUV5* operon. Therefore, addition of IPTG triggers the induction of protein expression and the T7 RNA polymerase that is produced after induction then transcribes the coding sequence of the protein inserted inside an expression vector under the control of the T7 promoter (Studier and Moffat 1986; Studier *et al*., 1990).

Whole genome sequencing has provided a comprehensive collection of plants genetic information and more than 40 plants of agronomical importance have so far been completely sequenced (Li *et al*., 2013). This therefore, provides a platform for plant functional genomic studies through gene annotation, cloning, expression and biochemical analyses (Chai *et al*., 2015). Determination of protein function is largely dependent on the knowledge of gene sequence and protein sequences. However, since the sequencing of the soybean genome in 2010 (Schmutz *et al*., 2010), the majority of its protein coding genes have remained experimentally unverified (Chai *et al*., 2015). There are reports that soybean studies are lagging compared to other common plants like rice and Arabidopsi*s* and this has been attributed to lower transformation efficiency and its genome complexity (Xia *et al*., 2013). Studies in soybean genome are vital in order to comprehend how the legume responds and adapts to several environmental stresses since the legume is considered as one of the most drought sensitive crops in the world (Tan *et al*., 2015). Apart from being an important crop, genome studies of this legume are crucial as the crop plant is a possible model plant to studying genome duplication (Wang *et al*., 2014; Tan *et al*., 2015), gene evolution and functional diversification (Tan *et al*., 2015). However, with the sequencing of the genome, a new era of gene cloning and functional analysis in soybean has thus begun.

To this day, most genome studies in *G. max* have focused mainly on genetic mapping of important quantitative trait loci (QTLs) for the improvement of agronomic traits but most of these genes have not yet been cloned, however, the cloning of genes is always very crucial in all genome studies as this would provide an insight into gene regulation network and specific features of the *G. max* genome. Most genes that have been cloned and functionally characterised in soybean include the *E* serials (*E1, E2, E3, E4*) that control flowering times and reproductive phases (Kumudini *et al*., 2007), the *Rhg4* gene responsible for resistance to the pathogen *Heterodera glycines*, which is a major constraint to soybean production (Liu *et al*., 2012). Tan *et al*. (2015) were able to isolate and clone the *DREB2* gene from *G. max* and expressed the recombinant protein in transgenic tobacco plants, thus inducing accumulation of proline, which in turn improved drought tolerance. The results indicated that the *G. max DREB2* gene could be used to enhance drought tolerance in soybean through overexpression of the gene. It has been reported that gene annotation in soybean is still incomplete and most of the predicted genes in this legume have not been supported by expression information (Wang *et al*., 2014).

To date, there exists no published literature on the molecular cloning, recombinant protein expression and functional characterisation of any AC enzyme in *G. max*. The molecular cloning and recombinant protein expression of AC enzymes from plants of economic significance such as soybean is vital as this will allow for identification and confirmation of the existence and perhaps roles of such enzymes in crop plants. Confirmation of the biological roles of the enzymatic products of ACs, i.e., the signalling molecule, cAMP can be used to improve major traits in the production and utilisation of this legume of great economic importance. Based on this notion, this study therefore, was undertaken to clone the AC-containing fragment of the XP_003529590 gene (GmAC1) into the pTrHis2-TOPO to generate the pTrHis2-TOPO:GmAC1 expression construct. The expression construct was then used to transform some chemically competent *E. coli* BL21 (DE3) pLysS expression cells followed by partial expression of the intended recombinant GmAC1

protein. After expression, the GmAC1 protein was then tested for its ability to generate cAMP from ATP within these prokaryotic expression hosts.

3.2 Materials and methods

3.2.1 Preparation of plant material

A total of eight *G. max* (cultivar number: NS 5909R) seeds were surface-sterilised by transferring them into a sterile 50 ml falcon tube with 15 ml 70% ethanol and vortexed for 30 seconds. The seeds were washed three times with a sterilisation buffer containing 0.1% (w/v) sodium dodecyl sulphate (SDS) and 5% (v/v) sodium hypochlorite, to remove traces of ethanol and other biological contaminants. The sterilised seeds were then washed and fully rinsed (5 times) with 500 ml sterile distilled water. Finally, the seeds were sown in suitable plastic pots, each containing sterile potting soil and vermiculite in a 3:2 quantity ratio. The pots were placed in a growth chamber at 28°C/20°C day/night temperatures and a photoperiod of 14h/10h day/night settings for a period of 7 days before germination. The soybean leaves, roots, and stems were collected when the first trifoliate leaves were well expanded (approximately 15 days after sowing) and the targeted XP_003529590 gene fragment then subsequently isolated from these 2 weeks old soybean plants.

3.2.2 Designing and acquisition of sequence-specific primers

The genomic sequence of the XP_003529590 gene was retrieved from UniprotKB (https://www.uniprot.org/) in FASTA (canonical) format. A set of sequence-specific primers for the targeted GmAC1 gene fragment were then designed specifically to flank the annotated AC catalytic motif of the XP_003529590 at about 50 base pairs either side (Figure 3.1). During the forward and reverse primer designing process, special care was taken to ensure that both primers

contained a 5′ G instead of and/or also not a 5′ T at all, since the subsequent downstream addition of 3′ A-overhangs using *Taq* polymerase and during the T-A cloning process is most efficient at adding a non-template 3′ A next to a C and not at all when the neighbouring base is an A (Brownstein *et al*., 1996). The designed primers were then sent to Inqaba Biotechnological Sciences (Pretoria, Republic of South Africa) for chemical synthesis and their subsequent supply.

atgcaagttttctctaacgctcggcaagcctctcggcttctcctctcgcctcatctccgaagctcggaagctcctca
ctccaccgcgctctctctcttctcagggctcacgcaacgtgattcccggccagtgaacactgatccaatacaatgct
tcttgtcgaaggctttttactcttccggagtgggtaccgttgaagcaactccatcagaggatgtgaaggaattatac
gacaaaatgcttgactctgtaaaagttaaacgatcaatgcctccaaatgcttggttgtggtcaatgattgcaaattg
caaacaccagcccgatattagacttctcttcgacattttgcagaacctccgcagatttagattgtcgaatcttcgaa
ttcatgacaatttttaattgcaatctctgtcgagaagttgctaaggcatgtgttcatgcaggagctcttgattttgga
atgaaggctttgtggaagcataatgtctatggactgactccaaatattgcatctgctcaccatttactgacgaatgc
taagaaccacaatgatactaaactgttggtggaagtaatgaagcttctgaagaagaatgatttaccattgcaaccag
gcacagcagatatagttttcagcatttgttataatacagatgattgggagttgattaataagtatgcaaaaagattt
gtcatggctggcgtaaaactacgacaaacctcattggaaacatggatggaatttgctgccaaaagaggagacattca
ctcattgtgg**aaaatagaaaagttgagatccaattcaatg**aagcagcacactttgataactgggttttcttgtgtca
agggtcttctgttggaacgtaaacccagtgatgcagttgctgtcattcaagttctaaatcagaccttgtctgataca
aagaagtcaggcatcaagggtgaacttcagaagcttgtatctgagtggcctttggaagttattaagcaccaaaaaga
a**gaggacagaaaggcattagcagcctctttgaaatctgat**atccttgtcatggttagtgagctactgagcatgggtc
ttgaggcaaatgtaagtttggaa**gacctggacagaaaggaagatattccacaa**tag

Forward: 5′-gAA ATA gAA AAg TTg AgA TCC AAT TCA ATg-3′
Reverse: 5′-gTg Tgg AAT ATC TTC CTT TCT gTC CAg gTC-3′

Figure 3.1: Nucleotide sequence of the XP_003529590 gene showing the GmAC1 gene fragment that was cloned and characterised in this study. The bolded green highlighted and underlined sequence represents the annotated AC catalytic motif of the XP_003529590 gene, while the bolded yellow highlighted sequences show the priming sites of the targeted and intended GmAC1. The two manually designed and used sequence-specific primers are highlighted in blue.

3.2.3 Isolation of the targeted GmAC1 gene fragment

3.2.3.1 Extraction of the total RNA

The targeted GmAC1 gene fragment was synthesised from total mRNA obtained from the 2-week soybean plants after harvesting 1-2 g of the plant material. The total mRNA was extracted using the Gene Jet Plant RNA Purification Mini Kit and according to the manufacturer's protocol (Catalogue #: K0801; Thermo Scientific Inc., Massachusetts, USA). Briefly, about 100 mg of

fresh soybean tissue, flash-frozen in liquid nitrogen, were ground thoroughly using deeply chilled sterile pestle and mortar into a fine powder. The soybean tissue powder was then immediately transferred to a 1.5 ml microcentrifuge tube containing 500 µl of Plant RNA Lysis Solution (supplemented with 10 µl of 2 M DTT) in order to release as much RNA as possible from the plant cells and prevent RNA degradation. The mixture was then vortexed for 20 seconds to thoroughly mix the contents. The mixed contents were then incubated at 56°C for 3 minutes and thereafter, centrifuged for 5 minutes at 20 000x*g* in an LSE High Speed Microcentrifuge (Corning Inc., Amsterdam, Netherlands). About 550 µl of the supernatant were collected and transferred into a clean 1.5 ml microcentrifuge tube containing 250 µl of 96% ethanol. The contents were mixed gently through pipetting. The prepared mixture was then transferred to a purification column inserted in a collection tube and centrifuged for 1 minute at 20 000x*g*. The flow through solution was discarded while the column and the collection tube were re-assembled. About 700 µl of Wash Buffer 1 were added to the purification column and centrifuged for 1 minute at 12 000x*g*. The flow-through and collection tube were discarded while the purification tube was placed into a clean 2 ml collection tube. A total of 500 µl of Wash Buffer 2 were added to the purification column and the contents were centrifuged for 1 minute at 12 000x*g*. The flow-through was discarded, while the column and collection tube were re-assembled. The wash procedure with Wash Buffer 2 was once again repeated. After centrifugation, the collection tube containing the flow-through was discarded while the purification column was transferred into a RNAase-free 1.5 ml Eppendorf tube. The isolated RNA was then eluted by adding 50 µl of nuclease-free water to the centre of the purification column membrane and centrifuging for 1 minute at 12 000x*g*. The total RNA was quantified using the nanometer and a value of 943.5 ng/µl was obtained.

3.2.3.2 Isolation and amplification of the targeted GmAC1 gene fragment

The extracted total RNA from section 3.2.3.1 above was then used as a template to generate copy DNA (cDNA) of the XP_003529590 gene. Together with the acquired sequence-specific primers, the generated cDNA was then used to amplify the targeted GmAC1 gene fragment in a reverse transcriptase-polymerase chain reaction (RT-PCR) system using the Verso 1-Step RT-PCR Reddy Mix RT-PCR kit and following the manufacturer's instruction (Catalogue # AB-1454/LD/A; Thermo Scientific Inc., California, USA). Components of the used reaction mixtures for the RT-PCR are shown in Table 3.1 below while the associated thermocycling conditions for the amplification of the targeted GmAC1 gene fragment are shown in Table 3.2.

Table 3.1: Components of an RT-PCR reaction mixture for amplification of the targeted GmAC1 gene fragment.

Reaction Component	Volume (µl)	Final Concentration
Verso Enzyme Mix	1.0	
1-Step PCR Master Mix (10X)	25.0	1X
Forward Primer (10 µM)	1.0	200 nM
Reverse Primer (10 µM)	1.0	200 nM
RT Enhancer	2.5	
Water (PCR Grade)	17.5	
Template (RNA)	2.0	1 ng
Total Volume	50.0	

Table 3.2: The 1-step RT-PCR thermal cycling reaction conditions for amplification of the targeted GmAC1 gene fragment.

Reaction Step	Time (mins)	Temperature (°C)	Cycles
cDNA Synthesis	15.00	50	1
RT Enzyme Inactivation	2.00	95	1
Denaturation	0.583	95	
Annealing	0.83	50	35

Extension	1.00	72	
Final Extension	5.00	72	1

3.2.4 Agarose gel electrophoresis of the amplified GmAC1 gene fragment

The amplified RT-PCR products were resolved on 1% agarose gel-electrophoresis stained with 0.5 µg/ml ethidium bromide using 1 kb Gene-Ruler DNA ladder (Thermo Scientific Inc., California, USA). The samples were run in 1X TBE buffer at 90 volts and 400 mA current for 50 minutes. The resolved gel was then visualised under UV light using the UV 2000 Trans-illuminator System (Bio-Rad Laboratories Inc., California, USA) and images finally captured with a ChemiDoc Imaging System (Bio-Rad Laboratories Inc., California, USA).

3.2.5 Cloning of the amplified GmAC1 gene fragment

3.2.5.1 Addition of the 3´-adenine overhangs

About 1 µl of *Taq* polymerase was added to 40 µl of the RT-PCR product reaction on ice and then incubated at 72ºC for 10 minutes on a C1000 Thermo-cycler System (Bio-Rad Laboratories Inc., California, USA). This process was crucial since it facilitated the cloning efficiency of the GmAC1 gene fragment into the pTRcHis2-TOPO expression vector by adding single A-overhangs to the 3′ blunt ends of the GmAC1 gene fragment. The reaction mixture was then placed on ice for further use in the pTRcHis2-TOPO TA cloning system.

3.2.5.2 Ligation of the adenylated GmAC1 gene insert into the pTrcHis2-TOPO vector

About 4 µl of the adenylated GmAC1 gene fragment were added to 1 µl of the pTRcHis2-TOPO expression vector. The mixture was mixed gently by swirling with a pipette tip prior to its incubation at room temperature for 5 minutes. The ligation mixture was then immediately used in the transformation of chemically competent One Shot TOPO 10 *E. coli* cells.

3.2.5.3 Transformation of the chemically competent One Shot TOPO 10 *E. coli* cells with the pTRcHis2-TOPO:GmAC1 fusion expression construct

The designed pTRcHis2-TOPO:GmAC1 construct from above (section 3.2.5.2) was used to transform some chemically competent One Shot TOPO 10 propagation cells, whereby 1 µl of the ligation mixture was aseptically added to 40 µl of the ice-cold chemically competent cells. The mixture was incubated on ice for 5 minutes before being heat-shocked for 30 seconds at 42°C and immediately placed on ice for 2 minutes. The incubation on ice and heat shocking steps facilitate the transformation of bacteria with recombinant plasmids. After the 2-minute ice incubation, 250 µl of the 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) was added to the transformed cells. The culture was then incubated in a shaking incubator at 200 rpm at 37°C for 60 minutes. The 60-minute incubation period and shaking of tubes at 200 rpm was an important step during the whole transformation process as it facilitated the chemically competent One Shot TOPO 10 cells to express β-galactamase, which in turn was essential in the detoxification of the antibiotic (ampicillin) during and later on in the selection process. 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 incubated at 37°C overnight in an Incubat 2001651 Bench-top Incubator (JP

Selecta SA., Barcelona, Spain). The next morning, the plates were visually inspected for colony growth. Successful growth of colonies indicated that only cells harbouring the gene construct were able to survive the antibiotic selection pressure on the LB plates.

3.2.5.4 Extraction of the pTRcHis2-TOPO:GmAC1 expression construct from the transformed One Shot TOPO 10 *E. coli* cells

Single colonies were selected from the overnight LB plates and incubated individually in 15 ml falcon tubes containing fresh LB broth media (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1.0% (w/v)) sodium chloride supplemented with 0.5% (w/v) glucose, 50 µg/ml ampicillin. The falcon tubes were incubated overnight at 37º shaking at 200 rpm. The cells were harvested the following morning by centrifuging at 6 800x*g* for 5 minutes at room temperature. The supernatant was discarded, and plasmid was extracted from the pelleted cells at room temperature using the GeneJet Plasmid Miniprep Kit and according to the manufacturer's instructions (Thermo Fisher Scientific Inc., California, USA).

In short, the pelleted cells were briefly resuspended in 250 µl of Resuspension solution so as to provide an optimal pH (pH 8.0), which was ideal for the subsequent lysis. The resuspended solution was transferred into a sterile 1.5 Eppendorf tube and 250 µl of lysis solution was added. The mixture was gently mixed by inverting the Eppendorf tube until the mixture became viscous and clear. Soon after, 350 µl of the Neutralisation solution was added and immediately mixed gently by inverting the tube 6 times. The cell suspension was then centrifuged for 5 minutes at 16 300x*g* to pellet the cell debris and the supernatant was carefully transferred into a supplied GeneJET spin column. The spin column was fitted into a collection tube and centrifuged for 1 minute at 16 300x*g* with the flow-through discarded while the column was replaced into the same collection tube. About 500 µl of Wash solution (diluted with 96% (v/v) ethanol prior to first use)

was added to the spin column after which the column was later centrifuged for 1 minute. The flow-through was discarded and the column was placed back into the same collection tube. The wash step was repeated before the spin column was centrifuged again for 1 minute to remove any residual wash solution. The washed and semi-dried GeneJET spin column was then transferred into a new and sterile 1.5 ml Eppendorf tube, where the plasmid DNA was then eluted with 50 µl sterile distilled water (pre-warmed to 56ºC) by adding the water directly onto the membrane of the column and allowing the column to sit and incubate for 2 minutes. The assembled column was then centrifuged at 16 300x*g* for 1 minute to elute the plasmid. The plasmid DNA (pTRcHis2-TOPO:GmAC1 fragment construct) was stored at -20ºC for further use while the used column was discarded.

3.2.5.5 Analysis of the positive clones

Convectional PCR procedure was used in the confirmation of positive clones in accordance with the standard MyTaq Mix protocol (Bioline, London, UK). This was performed to check if the GmAC1 insert was successfully ligated into the pTrHis2-TOPO cloning vector and in the correct orientation. The reaction mixtures for the two mentioned processes are presented in Tables 3.3 and 3.4 respectively and the thermal cycling conditions for both processes are given in Table 3.5.

Table 3.3: Reaction components of a PCR reaction mixture to confirm successful ligation of the GmAC1 gene fragment (insert) into the pTrcHis2-TOPO expression vector.

Component	Volume (µl)	Final Concentration
Template (10 ng DNA)	1	2 ng
Insert Primers (Each)	1	200 nM
MyTaq Reddy Mix (2X)	25	200 nM
Water sdH_20	Up to 25	-

Table 3.4: Reaction components of a PCR reaction mixture to confirm correct orientation of the GmAC1 gene fragment (insert) into the pTrcHis2-TOPO expression vector.

Component	Volume (µl)	Final Concentration
Template (10 ng DNA)	1	2 ng
Insert Forward and Vector Reverse Primers (20 µM each)	1	200 nM
MyTaq Reddy Mix (2X)	25	200 nM
Water sdH_2O	Up to 25	-

Table 3.5: Thermal cycle conditions for confirmation of the successful ligation and correct orientation of the GmAC1 gene fragment (insert) into the pTrcHis2-TOPO expression vector.

Step	Temperature (°C)	Time (seconds)	Cycles
Initial Denaturation	95	60	1
Denaturation	95	15	35
Annealing	50	15	
Extension	72		

3.2.5.6 Agarose gel electrophoresis

The PCR products from the analysis for successful ligation and correct orientation of the insert into the vector reactions were resolved on a 1% (w/v) agarose gel supplemented with 0.5 µg/ml ethidium bromide and images analysed as has already been outlined in section 3.2.4 of this chapter.

Double amplification of the targeted GmAC1 gene insert in both reaction samples of Tables 3.3 and 3.4 would therefore, correspondingly confirm its successful cloning (ligation and correct orientation) into the pTrcHis2-TOPO expression vector.

3.2.6 Transformation of the chemically competent BL21 (DE3) pLysS *E. coli* cells with the pTRcHis2-TOPO:GmAC1 fusion expression construct

The successfully designed pTRcHis2-TOPO:GmAC1 expression construct from section 3.2.5 above was used to transform some chemically competent *E. coli* BL21 (DE3) pLysS expression cells, whereby 5 µl of the ligation mixture were aseptically added to 40 µl of the ice-cold chemically competent BL21 (DE3) pLysS expression *E. coli* cells. The mixture was incubated on ice for 30 minutes before being heat-shocked for 30 seconds at 42°C and immediately placed on ice for 2 minutes. Soon after, 250 µl of the SOC medium were added to the transformed cells. The culture was then incubated in a shaking incubator at 200 rpm at 37°C for 60 minutes to allow expression of β-galactamase enzyme by the transformed cells. This enzyme later enables the transformed chemically competent BL21 (DE3) pLysS *E. coli* cells to detoxify ampicillin during the culture selection process. The incubation period was followed by plating onto 3 LB agar plates supplemented with 100 µg/ml ampicillin, 10 µg/ml chloramphenicol and 0.5% glucose. The following volumes were then plated; 20 µl, 50 µl and 80 µl. Finally, the plates were incubated at 37°C overnight to allow for growth of colonies.

3.3 Partial expression of the recombinant GmAC1 protein

A transformed BL21 (DE3) pLysS *E. coli* cell colony harbouring the correctly designed pTRcHis2-TOPO:GmAC1 fusion construct was used to inoculate 10 ml of fresh LB broth

supplemented with 0.5% glucose, 10 µg/ml chloramphenicol and 100 µg/ml ampicillin in a 50 ml falcon tube. The culture was incubated overnight at 37ºC and shaking at 200 rpm. On the subsequent day, 200 µl of the overnight culture were used to inoculate 20 ml of fresh LB broth containing 10 µg/ml chloramphenicol, 100 µg/ml ampicillin and 0.5% glucose. The culture was then incubated at 37°C, shaking at 200 rpm and until an OD_{600} of 0.5 was reached as measured by the Hekios spectrophotometer (Merck, Gauteng, RSA). Immediately, the culture was split into two 50 ml falcon tubes, each containing 10 ml of the split culture. One culture was induced to express the intended recombinant GmAC1 protein by adding 1 mM of IPTG (Sigma-Aldrich Corp, Missouri, USA) while the other culture was left un-induced (control). The split cultures were shaken in an incubator (200 rpm) at 37°C for 3 hours. After the 3 hours, 500 µl samples were transferred into Eppendorf tubes and together with the rest of the other cells, they were then separately centrifuged at 8 000x*g* for 5 minutes to pellet the cells. The pelleted 500 µl samples were then analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for protein expression while the rest of the bulk cells were kept at -20°C for further use.

3.4. Determination of the endogenous AC activity of the recombinant GmAC1 protein

The harvested cells (both the induced and the non-induced) from section 3.3 above were then lysed in 1 ml lysis buffer 1 (Amersham Healthcare, California, USA) supplemented with 2 mM 3-isobutyl-1-methylxanthine (IBMX) (Sigma-Aldrich Corp., Missouri, USA) to inhibit phosphodiesterases. The samples were shaken for 60 minutes at 200 rpm at room temperature on an orbital shaker to intensify the cell lysis process. The samples were then centrifuged at 9 200x*g* for 5 minutes before 200 µl of the lysates were transferred into fresh Eppendorf tubes. 200 µl of the lysis buffer 2 (Amersham Healthcare, California, USA) was added to the lysates and mixed thoroughly. After that, 200 µl portions of the reaction mixtures were transferred into fresh

Eppendorf tubes, where 11 µl of the acetylating reagent (Sigma-Aldrich Corp., Missouri) were added before the mixtures are gently mixed by pulsing. The endogenous cAMP contents from the prepared lysates were then measured by a cAMP-linked enzyme immunoassay kit (Catalogue # CA201; Sigma-Aldrich Corp., Missouri, USA) following the acetylation version of its protocol and as described by the manufacturer's manual. The measurements were taken using a Microplate Reader (Labtech International Limited, East Sussex, UK) at 405 nm and the obtained results then subjected to a one-way ANOVA.

3.5 Results

3.5.1 Isolation of the GmAC1 gene fragment from *G. max*

The targeted GmAC1 gene fragment was isolated from 15-day old *G. max* plants that were generated and maintained in a growth chamber at 28°C/20°C day/night temperature and a photoperiod of 14h/10h day/night. The isolation of the GmAC1 gene fragment was then done by first extracting total RNA from the plants followed by use of the extracted total RNA to generate cDNA of the XP_003529590 gene. The generated cDNA together with the manually designed sequence-specific primers were then used in a specialised Verso 1-Step RT-PCR protocol to amplify and isolate the targeted GmAC1 gene fragment at an approximate size range of 350 - 355 bp (Figure 3.2). The isolated GmAC1 gene fragment was then later cloned into a viable prokaryotic system for subsequent expression and ultimate functional characterisation of the intended recombinant GmAC1 protein.

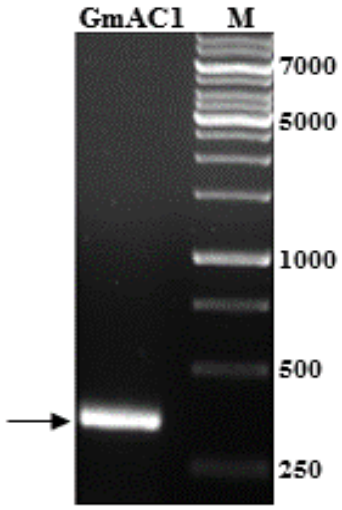


Figure 3.2: Isolation of the GmAC1 gene fragment from *G. max*. A 1% agarose gel resolution of the GmAC1 gene fragment isolated from *G. max* via a specialised RT-PCR system, whereby GmAC1 represents the amplified GmAC1 gene fragment, M is representing a 1 kb GeneRuler ladder (Thermo Scientific Inc., Burlington, Cananda) while the arrow is marking the expected 352 bp band of the amplified GmAC1 gene fragment.

3.5.2 Cloning of the GmAC1 Gene Fragment

The isolated targeted GmAC1 gene fragment from *G. max* was ligated into a pTRcHis2-TOPO expression vector to produce a pTRcHis2-TOPO:GmAC1 fusion expression construct. After that, the generated construct was then assessed for proper designing as is shown in Figure 3.3 below whereby both the successful ligation of the gene fragment and its correct orientation in the expression were checked and ascertained.

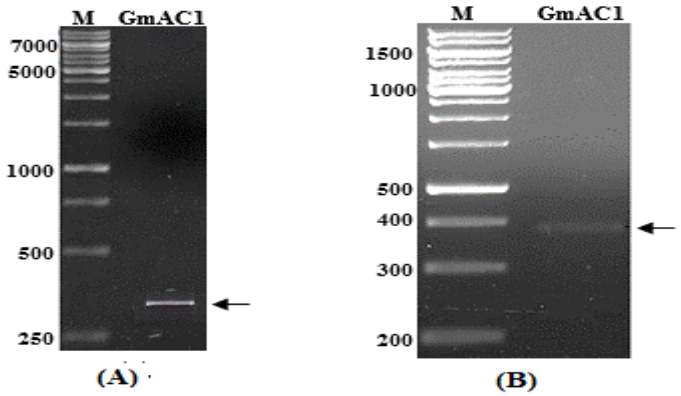


Figure 3.3: Determination of the successful cloning of the GmAC1 gene fragment in the pTRcHis2-TOPO expression vector. (**a**) Confirmation of the successful ligation of the GmAC1 gene fragment into the pTRcHis2-TOPO expression vector after its successful re-amplification with its own primers using the GmAC1 fusion expression construct as a template. (**b**) Confirmation of the correct orientation of the GmAC1 gene fragment in the pTRcHis2-TOPO:GmAC1 fusion expression construct after its successful re-amplification with its own forward primer and the pTRcHis2-TOPO vector reverse primer using the pTRcHis2-TOPO:GmAC1 fusion expression construct as a template. In both gels, GmAC1 represents the amplified GmAC1 gene fragment, M represents the GeneRuler ladders (1 kb in (**a**) and 100 bp in (**b**)) while the arrow is marking the respective amplified GmAC1 gene fragments.

3.5.3 Partial expression of the recombinant GmAC1 protein

The correctly generated and confirmed pTRcHis2-TOPO:GmAC1 fusion expression construct was used to transform some chemically competent *E. coli* BL21 (DE3) pLysS expression cells followed by induction of protein expression with 1 mM IPTG. As is shown in Figure 3.4 below, the expected recombinant GmAC1 protein was successfully expressed as a C-terminal His-tagged fusion product of approximately 24.79 kDa.

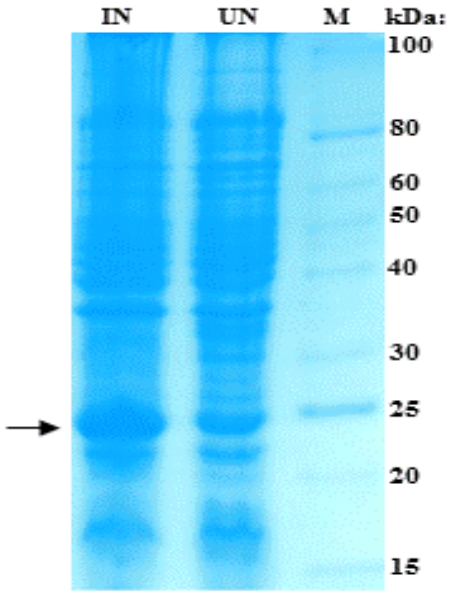


Figure 3.4: Partial expression of the GmAC1 recombinant protein. Induced (IN) and non-induced (UN) *E. coli* BL21 (DE3) pLysS cells harbouring the pTRcHis2-TOPO:GmAC1 fusion expression construct were resolved by SDS-PAGE on a 12% gel against a standard molecular weight marker (M). The arrow marks the expressed recombinant GmAC1 protein at its targeted size of approximately 24.79 kDa.

3.5.4 Molecular determination of the endogenous AC activity of the GmAC1 protein

After expression, the endogenous AC activity of the expressed recombinant GmAC1 protein or its ability to generate cAMP in the transformed competent *E. coli* BL21 (DE3) pLysS expression cells was assessed and determined by enzyme immunoassay. Cells induced by the 1 mM IPTG produced ±175 fmol/g of cAMP compared to the non-induced cells that produced only an insignificant amount of cAMP (±55 fmol/g) as shown in Figure 3.5 below.

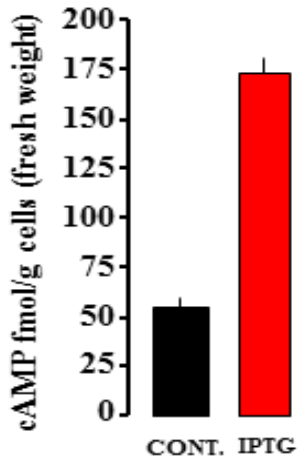


Figure 3.5: Determination of the endogenous AC activity of the recombinant GmAC1 protein by enzyme immunoassay. CONT is the control, which shows the level of cAMP generated within the un-induced competent *E. coli* BL21 (DE3) pLysS expression cells harbouring the pTRcHis2-TOPO:GmAC1 fusion expression construct while IPTG represents the level of cAMP generated within the induced competent *E. coli* BL21 (DE3) pLysS expression cells harbouring the pTRcHis2-TOPO:GmAC1 fusion expression construct.

3.6 Discussion

In this research, soybean plants were generated followed by extraction of their total RNA. Using this extracted total RNA, cDNA of the XP_003529590 gene was then generated via a specialised RT-PCR system. Since the main focus of this whole study was on the AC-containing fragment of the XP_003529590 gene (GmAC1), the generated XP_003529590 cDNA together with some GmAC1-sequence-specific primers were then used in the same specialised RT-PCR system to amplify and isolate the targeted and intended 352 bp GmAC1 gene fragment, which was visualised on 1% agarose as is shown in Figure 3.2. The amplified and isolated GmAC1 gene fragment is devoid of introns and therefore, consists only of the uninterrupted coding region (exons). In cloning systems, the possible presence of introns in cloned genes is generally believed to impose substantial and unnecessary energetic burden on protein expressing cells (Jo and Choi, 2015), and

hence it was therefore very necessary to generate and isolate a GmAC1 gene fragment devoid of the introns in our case.

The desired amplified and isolated GmAC1 gene fragment was then recombinantly cloned into the pTRcHis2-TOPO expression vector using the topoisomerase mediated (TOPO-TA) cloning system to produce pTRcHis2-TOPO:GmAC1 fusion expression construct, which would generate a C-terminus His-tagged protein product. The TOPO-TA cloning system uses the topoisomerase-based DNA cloning method that does not use restriction enzymes or DNA ligases, but *Vaccinia* virus- isolated topoisomerase 1 (Shuman and Moss, 1987), which was able to perform both the end modification and end joining (Shuman and Moss, 1987; Shuman, 1994) of the amplified GmAC1 gene fragment into the pTRcHis2-TOPO expression vector. The *Vaccinia* topoisomerase 1 is able to recognise, bind and cleave a precise sequence of 5′-(C/T) CCTT↓-3′ thus hydrolysing the phosphodiester bond at the 3′-end (Shuman, 1991). Once cleavage has occurred, a covalent bond is formed between the DNA fragment insert and the enzyme, forming a phosphor-tyrosyl bond, in which the bond energy is then used to relegate the 5′-hydroxyl end of the cleaved strand. Therefore, the *Vaccinia* topoisomerase 1 thus can relegate the same bond that was originally cleaved and/or religate it to a heterologous acceptor DNA possessing a 5′-hydroxyl tail that is complementary to that of the adduct, hence creating a recombinant molecule (Shuman, 1992).

In our case, the absence of the use of restriction digestion on both the cloning vector and its associated amplified GmAC1 gene fragment undoubtably, made the production of the intended recombinant DNA (pTRcHis2-TOPO:GmAC1 fusion expression construct) far much easier. In most researches, various researches have exploited the *Vaccinia* topoisomerase 1 to develop several high-throughput cloning systems. These include; Udo (2015), who concluded that the TOPO cloning was an excellent method to clone PCR amplified DNA though the study was based on cDNA cloning for expression in mammalian cells. Heyman *et al.* (1999) also exploited the strength of this technology in cloning and expressing human cDNA. A report on the successful

cloning and molecular characterisation of an AC gene using this same system was also made in *Hippeastrum hybridum* namely, the HpAC1 gene (Świeżawska *et al*., 2014). Another AC gene, At5g09400 that expresses ATKUP7, was also cloned and reported by Al-Younis *et al*. (2015) in which the gene was cloned into a Gateway compatible pCR8 vector (Invitrogen, Carlsbad, USA) by the TA cloning. Furthermore, Chatukuta *et al*. (2018) also recently cloned an AC gene fragment of the At1g68110 gene that expresses AtCIAP from *A. thaliana*, using cDNA generated through RT-PCR into a pTrcHis2-TOPO expression vector via the same TA cloning system.

Before initiating protein expression, the generated pTrcHis2-TOPO:GmAC1 fusion expression construct was first assessed for its correct designing i.e., presence of the insert and its correct orientation in the pTRcHis2-TOPO expression vector as is presented in Figure 3.3. Ideally, it was very important to check and validate that the cloned GmAC1 gene fragment was correctly ligated into expression vector in the 5′ to 3′ directional orientation for facilitation of effective expression of the intended recombinant GmAC1 protein. This verification procedure was then done; for checking and ascertaining the presence of the GmAC1 gene fragment in the vector, using the GmAC1 gene primers (Figure 3.3a) and for checking and ascertaining its correct orientation in the vector, using one GmAC1 gene fragment forward primer and the pTRcHis2-TOPO expression vector reverse primer (Figure 3.3b). The obtained results thus provided valid evidence of the correct cloning of the GmAC1 gene fragment into the pTRcHis2-TOPO expression vector and proper designing of the desired pTrcHis2-TOPO:GmAC1 fusion expression construct.

The successful cloning of the GmAC1 gene fragment into the pTRcHis2-TOPO expression vector was then followed by transformation of some chemically *E. coli* BL21 (DE3) pLysS expression cells using the pTRcHis2-TOPO:GmAC1 fusion expression construct. Protein expression was then induced using 1 mM IPTG and a recombinant GmAC1 protein product of an expected approximate size of 24.79 kDa was expressed (Figure 3.4). This size corresponds to a size prediction from EnsemblGenome web server as presented in Chapter 2, section 2.3.1.

Recombinant protein expression in *E. coli* BL21 (DE3) pLysS was successful since the pTrcHis2-TOPO cloning vector contains a hybrid promoter known as the *trc* promoter region that contain the -35 region from the *trpB* promoter and the -10 region from *lacUV5* promoter, thereby promoting high level expression (Egon *et al*., 1983) of the recombinant GmAC1 protein in the *E. coli* expression cells. The presence of the *lac* operator (*lacO*) allowed for the binding of the *lac* repressor encoded by the *lac1* gene, thereby allowing induction of protein expression of the recombinant GmAC1 protein in the presence of IPTG (Jacob and Monod, 1961; Müller-Hill *et al*., 1968). Premature transcription termination of the recombinant protein was reduced by the *rrnB* antitermination sequences also present in the pTrcHis2-TOPO cloning vector (Li *et al*., 1984), thus enabling for successful protein induction in the *E. coli* BL21 (DE3) pLysS expression cells. Within the cloning vector, there exists a T7 gene 10 translational enhancer sequence that promoted a more efficient translational initiation (Olins *et al*., 1988).

After expression, an assessment of the ability of the expressed GmAC1 recombinant protein to endogenously produce cAMP from ATP was then tested in the *E. coli* BL21 (DE3) pLysS transformed with the pTRcHis2-TOPO:GmAC1 fusion expression construct. The amount of cAMP produced in the *E. coli* BL21 (DE3) pLysS expression hosts after induction with 1 mM IPTG as measured by the cAMP-linked enzyme immunoassay kit was ±175 fmol/g compared to ±55 fmol/g cAMP produced in the un-induced *E. coli* BL21 (DE3) pLysS expression hosts (Figure 3.5). These results thus showed that the recombinant GmAC1 harboured by the expression host cells is a possible AC enzyme as exhibited by its inherent ability to produce cAMP in host cells. The findings are also closely linked to that of Moutinho *et al*. (2001) and Ruzvidzo *et al*. (2013), where the pRSET:PSiP and pCRT/7NT-TOPO:AtPPR-AC fusion expression constructs respectively were used to transform some chemically competent BL21 (DE3) and BL21 (DE3) Star pLysS *E. coli* cells and in both cases, generated almost 3.0-fold levels of the endogenous cAMP. The PSiP and AtPPR-AC proteins are novel plant molecules that recently, have been

confirmed to be cAMP-generating molecules in *Z. mays* (Moutinho *et al*., 2001) and *A. thaliana* (Ruzvidzo *et al*., 2013) respectively.

3.7 Conclusion

The isolated, cloned and partially expressed recombinant GmAC1 protein is capable of generating endogenous cAMP from ATP in living *E. coli* cells. Whether this is as a result of its inherent activity as a functional AC enzyme or simply by inducing the resident *E. coli* ACs, still needs to be further investigated and established.

CHAPTER 4

DETERMINATION OF THE *IN-VIVO* ADENYLYL CYCLASE ACTIVITY OF THE RECOMBINANT GmAC1 PROTEIN

Abstract

In prokaryotic and eukaryotic organisms, adenylate cyclases (AC) are enzymes responsible for increasing the intracellular levels of cAMP through the conversion of ATP into cAMP. In *E. coli*, cAMP has been reported to be an important regulator of energy source utilisation, whereby the metabolism of carbon compounds in this bacterial system solely depends on the levels of cAMP within the cell. Therefore, *E. coli* has been widely studied to understand inducible pathways that determine lactose metabolism. As an extension of this, plant biologists have also begun to exploit the ability of this bacterial system to ferment lactose by removing the AC catalytic domain from the *cya* gene thus creating *cya*A mutants (SP850 hosts) that are devoid of cAMP formation and consequently no lactose fermentation. Ideally, the system has been manipulated and extensively utilised to perform what is technically termed complementation test on suspected AC candidate molecules from various organisms, whereby such candidates are used as possible rescuing molecules of the SP850 hosts in lactose metabolism. In this chapter and using the same conceptual approach to test a potential AC candidate from *Glycine max* termed the GmAC1 protein, we used this protein candidate in form of its cloned expression construct (the pTRcHis2-TOPO:GmAC1) to transform *E. coli* SP850 *cya*A mutant cells and demonstrated its enzymatic ability to rescue the mutant hosts.

4.1 Introduction

In bacterial systems, cAMP has been reported as an important regulator of energy source utilisation (Botsford, 1981). cAMP has been implicated in modulating the activity of certain specific receptor proteins, which then regulate the transcription and ultimately expression of a large set of gene systems (Botsford, 1981). cAMP is a product of ATP conversion by adenylyl cyclases (ACs). In *Bordetella pertussis*, the AC toxin (*cya*A) has been reported as a calmodulin-activated AC that is able to by-pass the receptor-mediated endocytosis pathway (Wolf *et al*., 1980; Gordon *et al*., 1989). The toxin therefore, can penetrate directly across cytoplasmic membranes of many epithelial and immune effector cells, thereby causing activation of the intracellular calmodulin, leading to the *cya*A catalysing and the uncontrollable formation of cAMP. This cascade of events results in the incapacitation of host cells because of intoxication (Confer and Eaton, 1982), and as a result of this, the nature of the *cya*A system has made it the subject of most studies. Researchers have been able to abolish the invasive activity and severity hemeolytic activity of this *cya*A (Bellalou *et al*., 1990), which is involved in the catastrophic activity of its expressed AC. *Escherichia coli* also, has been reported to produce *cya*A cells and a study by Benz *et al.* (1994), revealed that both *E. coli* and *B. pertussis* express the toxin from the same genes. While the *cya*A toxins from *E. coli* cells have been classified into class 1 ACs, the calmodulin-activated toxins from *B. pertussis* have been classified into class II (Tellez-Sosa *et al*., 2002).

Naturally, in *E. coli*, the metabolism of carbon compounds has been established to involve inducible pathways, which strictly depend on the levels of cAMP in the cell (Perlman and Pastan, 1973). This system is one of the best studied inducible cellular pathways that involve genes that determine lactose (*lac*) metabolism (Brickman, 1973). Studies have revealed that the initiation of transcription of the *lac* operon requires cAMP and a receptor protein (Emmer *et al*., 1970; Zubay *et al*., 1970; deCrombrugghe *et al*., 1971; Eron and Block, 1971). In this bacterial system, cAMP receptor proteins (CRP) reportedly control gene expression in response to changes in the

intracellular levels of cAMP. In order to achieve greater carbon usage, the carbon catabolite repression is a major control system in bacteria (Kovarova-Kovar and Egli, 1998) through which the depletion of glucose promotes activation of the AC. This activation of the AC leads to an elevation of cAMP levels, which eventually binds to the CRP (Saier and Reizer, 1994), thus inducing a conformational change that enables it to bind to specific DNA sequences (Lauren *et al*., 1990). Consequently, this binding of genes stimulates transcriptional events that are necessary for utilisation of carbon sources other than glucose (Lauren *et al*., 1990). It was long confirmed in *E. coli* that uptake of glucose inhibited the synthesis of cAMP and thus metabolism of lactose (Perlman *et al*., 1969).

In *E. coli*, the deletion of the *cya* has resulted in formation of the SP850 *cya*A strain that lacks both the AC activity and cAMP production (Shah and Peterkofsky 1991). Consequent to this, successful complementation of the *E. coli* SP850 *cya*A mutant cells has been performed as early as 1998 using the *Prevotell ruminicola* (D31d) AC gene from ruminal anaerobic bacterium (Cotta *et al*., 1998). In that regard, the AC gene was cloned into a pUC18 plasmid and introduced into the *E. coli* SP850 *cya*A mutant cells by electroporation. The transformed cells were then screened on MacConkey lactose agar and produced one colony that had a deep red colour - an indication of lactose fermentation and ultimately complementation of AC deletion (Cotta *et al*., 1998). In plants and in studies involving characterisation of novel ACs, researchers have employed complementation analysis to confirm function of such novel ACs via the fermentation of lactose. The following authors were able to characterise AC activity from their respective expressed AC proteins: Moutinho *et al*. (2001) on the *Zea mays* PSiP protein, Świeżawska *et al.* (2014) on the *Hippeastrum hybridum* HpAC1 protein, Al-Younis *et al.* (2015, 2018) on the Arabidopsis AtKUP7 and AtKUP5 proteins respectively, Kasahara *et al.* (2016) on the *Marchantia polymorpha* MpCAPE protein, Chatukuta *et al.* (2018) on the Arabidopsis AtCIAP protein and Bianchet *et al*. (2019) on the Arabidopsis AtLRRAC1 protein.

In all of these studies, the authors utilised the *E. coli* SP850 strain, which is deficient in AC function, to confirm the cyclase activity of their novel AC genes/proteins, whereby the respective AC candidates were used to transform the mutant SP850 strain followed by growth on MacConkey agar and visual analysis of the resultant phenotype expression. All in all, their findings demonstrated AC functionality for their respective tested proteins since such proteins literally managed to convert the non-lactose fermenting mutant *E. coli* SP850 strain into a lactose fermenting wild type. Following these successful reported studies, this present study also, then embarked on assessing the previously cloned GmAC1 protein (Chapter 3) in a similar way. This is also on the backdrop that this recombinant protein (GmAC1) could not be conclusively declared as a *bona fide* an AC and/or accessory signalling molecule in the same chapter (Chapter 3).

4.2 Materials and methods

The pTRcHis2-TOPO:GmAC1 expression construct from generated in section 3.2.5.2 of Chapter 3 was herein used to test for the possible AC activity of the cloned GmAC1 gene fragment via complementation test.

4.2.1 Preparation of competent *E. coli* SP850 *cya*A mutant cells

E. coli SP850 *cya*A mutant cells (Coli Genetic Stock, Yale University, Connecticut, USA) were prepared to become chemically competent before they could be transformed with the pTRcHis2-TOPO:GmAC1 fusion expression construct and used for the intended work in this chapter. Briefly, 10 ml of LB broth (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1.0% (w/v) sodium chloride, 0.5% (w/v) glucose, 15 µg/ml kanamycin, pH: 7.0) were inoculated with the SP850 *cya*A mutant cells and incubated overnight on shaker at 200 rpm at 37°C. The following morning, 1 ml

of the overnight culture was used to inoculate 20 ml of fresh pre-warmed LB broth supplemented with 15 µg/ml kanamycin. The culture was then incubated at 37°C at 200 rpm and up until an OD_{600} of 0.5 was reached. The cells were harvested by centrifuging at 4 000x*g* for 5 minutes at 4°C. The supernatant was discarded while the harvested cells were resuspended in 15 ml of ice-cold Transformation Buffer 1 (30 mM potassium acetate, 50 mM manganese chloride, 100 mM rubidium chloride, 10 mM calcium chloride, 15% (v/v) sterile glycerol, pH: 5.8) and incubated on ice for 90 minutes. The resuspended cells were then harvested by centrifuging at 4 000x*g* for 5 minutes at 4ºC. The supernatant was discarded while the harvested cells were then resuspended in 4 ml of an ice-cold Transformation Buffer 2 (10 mM MOPS [3-(N-morpholino) propanesulfonic acid], 75 mM calcium chloride, 10 mM rubidium chloride, 15% (v/v) glycerol, pH: 6.8). The resuspended SP850 *cya*A mutant cells were then aliquoted into 100 µl portions and kept on ice for use in the subsequent transformation step.

4.2.2 Transformation of the competent *E. coli* SP850 *cya*A cells with the pTRcHis2-TOPO:GmAC1 expression construct

The prepared competent *E coli* SP850 *cya*A mutant cells from 4.2.1 were divided into two portions. The first portion was transformed (refer to section 3.2.6 for the transformation process) with the pTRcHis2-TOPO:GmAC1 expression construct (prepared in section 3.2.5.2), while the second portion was left un-transformed (control). The only variation in the transformation process herein from the previous one in section 3.2.6 was that the selection antibiotic was 35 µg/ml kanamycin instead of the 100 µg/ml ampicillin.

4.2.3 Complementation testing of the recombinant GmAC1 protein

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 segments, marked A, B and C respectively with a permanent marker. Segment A was left unstreaked (no *E. coli* SP850 *cya*A mutant cells), segment B was streaked with the chemically competent *E. coli* SP850 *cya*A mutant cells transformed with the pTRcHis2-TOPO:GmAC1 expression construct and segment C was streaked with the non-transformed *E. coli* SP850 *cya*A mutant cells. The plate was incubated at 37°C for 40 hours. After the incubation, all segments were then visually inspected for phenotypic characteristics. A deep red colour on the transformed *E. coli* SP850 *cya*A mutant cells would mean positive AC activity for the cloned and recombinantly expressed GmAC1 protein.

4.3 Results

The pTRcHis2-TOPO:GmAC1 expression construct transformed into chemically competent *E. coli* SP850 *cya*A mutant cells, which lack the ability to ferment lactose produced deep red cells much like the wild type. However, the mutant cells not transformed with the pTRcHis2-TOPO:GmAC1 expression construct were not able to ferment lactose hence produced pale yellow cells as is shown in Figure 4.1 below.

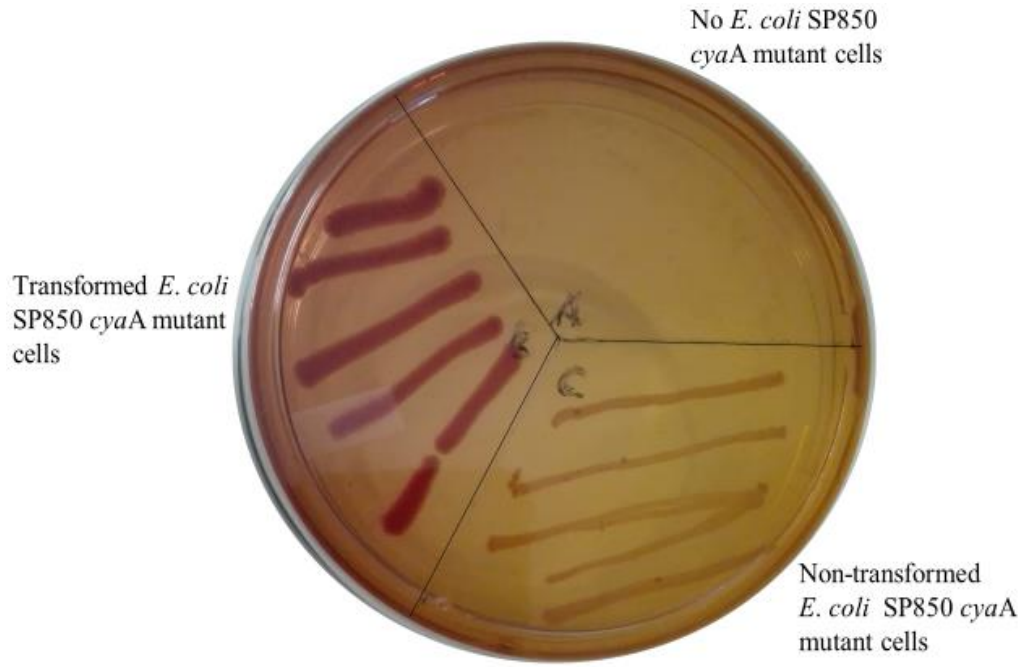


Figure 4.1: Determination of the AC activity of the recombinant GmAC1 protein via complementation test. The representation shows a Maconkey agar plate inoculated with various *E. coli* cell cultures. Segment A was left unstreaked (no bacterial cells), segment B was streaked with the mutant *E. coli* SP850 *cya*A mutant cells harbouring the pTRcHis2-TOPO:GmAC1 expression construct, which then showed a deep red phenotype, signifying a lactose-fermenting phenotype, and segment C was streaked with the non-transformed *E. coli* SP850 *cya*A mutant cells that produced pale yellowish colonies, signifying non-lactose fermentation.

4.4 Discussion

The SP850 is a *cya*A mutant strain of *E. coli* that is devoid of the AC activity and thus it is expected that this strain cannot use lactose and should produce pale/colourless colonies on MacConkey agar (Moutinho *et al*., 2001). The results of the complementation test of this study in Figure 4.1 showed that the non-transformed *E. coli* SP850 *cya*A mutant cells streaked on MacConkey agar were not able to ferment lactose hence producing the pale-yellow colonies (segment C). However, the *E. coli* SP850 *cya*A mutant cells that were transformed with the pTRcHis2-TOPO:GmAC1 expression construct were able to ferment lactose, hence the deep red colour observed in segment B. This implies that the pTRcHis2-TOPO:GmAC1 construct indeed carried a functional AC

enzyme (GmAC1) that in turn produced the deep red colour, an indication of lactose fermentation and complementation of the AC deletion in mutant *E. coli* SP850 *cya*A cells.

These obtained results are consistent with findings from a number of researchers who previously performed complementation analyses of their respective recombinants using the *E. coli* SP850 *cya*A mutant cells grown on MacConkey lactose agar. Cotta *et al*. (1998) performed a complementation analysis in an attempt to identify a novel AC in the ruminal anaerobe, *Prevotella ruminicola* D31d. Moutinho *et al*. (2001) performed complementation of the *cya*A mutation in *E. coli* (SP850) by a PSiP gene fragment from *Z. mays* that is involved in pollen tube growth and re-orientation for fertilisation and Świeżawska *et al*. (2014) used the *H. hybridum* HpAC1 protein to rescue the *E. coli* SP850 *cya*A mutant cells. Recently, several successful complementation of the *cya*A gene were performed by Ruzvidzo *et al*. (2013), Al-Younis *et al*. (2015), Kasahara *et al*. (2016), Al-Younis *et al.* (2018), Chatukuta *et al*. (2018) and Bianchet *et al*. (2019), who used the AtPPR-AC, AtKUP7, MpCAPE, AtKUP5, AtCIAP and AtLRRAC1 proteins respectively to rescue the SP850 host mutant.

Principally, in the *E. coli* SP850 *cya*A mutant cells transformed with the GmAC1 gene fragment, the generated cAMP was able to bind to the CRP of the bacterial cell to form the cAMP-CRP complex, which bound to the CRP binding region that is closely located to the *lac* promoter, thus enhancing transcription (Kuo *et al*., 2003). The lac promoter regulated the expression of genes that were involved in the catabolism of lactose (Ullmann and Danchin, 1983). Therefore, the non-transformed *E. coli* SP850 *cya*A mutant cells that failed to produce cAMP were unable to ferment the lactose, a condition that was overcome by the *E. coli* SP850 *cya*A mutant strain carrying the (functionally active) pTRcHis2-TOPO:GmAC1 expression construct.

4.5 Conclusion

Findings from this chapter provided solid evidence that the recombinant GmAC1 protein is indeed a *bona fide* functional AC capable of converting ATP into cAMP. The protein becomes the first ever such protein to be confirmed in soybean.

CHAPTER 5

AFFINITY PURIFICATION OF THE RECOMBINANT GmAC1 PROTEIN AND *IN VITRO* CHARACTERISATION OF ITS ENZYMATIC ACTIVITY

Abstract

In molecular biology, the cloning and expression of cAMP-producing proteins in *Escherichia coli* has made it possible to study their biological roles in various prokaryotic and eukaryotic organisms. However, expression of cloned genes in *E. coli* usually results in the production of biologically inactive recombinants, which must first be converted back into their active biological states before they can be studied and/or characterised. Nonetheless, an assessment of the quantity of endogenous cAMP in plants has greatly improved over the years with several technological advancements. To date, there exists various techniques with greater detection limits such as mass spectrometry, capillary spectrometry and enzyme immunoassay (EIA) that are being used to measure the levels of cAMP production and signalling. In this chapter, we report the over-expression of a recombinant GmAC1 protein from *Glycine max* in chemically competent *E. coli* BL21 (DE3) pLysS expression cells and its production as an insoluble heterologous product. The GmAC1 protein is the only soybean protein annotated to harbour a putative adenylate cyclase (AC) centre that may presumably generate cAMP. The expressed insoluble recombinant GmAC1 protein was then purified under non-native denaturing conditions using the Ni-NTA affinity purification system followed by its renaturation via a gradient-controlled BioLogic DuoFlow Chromatographic System. The resultant pure recombinant GmAC1 protein was then characterised enzymatically using EIA, wherein it was shown to be positively modulated by the Mn^{2+}, Ca^{2+}, and

HCO_3^{2-} molecular ions - meaning that it is actually a soluble AC (sAC), whose physiological roles may involve the sAC-cAMP-protein kinase A signalling system.

5.1 Introduction

Cyclic AMP has been described as an important second messenger in higher plants (Lemtiri-Chlieh *et al*., 2011; Mathieu-Demaziere *et al*., 2013) involved in signal transduction as plants respond to various environmental stresses and defense mechanisms (Thomas *et al*., 2013). The possibility availed by recombinant protein expression in *E. coli* has made it possible to study the biological roles of protein encoding genes involved in cAMP formation. cAMP is primarily produced by adenylyl cyclases (ACs) and the concentration of this signalling molecule in plant cells vary widely. The functional activity of ACs in plants have been studied through assessment of the quantity of endogenous cAMP (Lomovatskaya *et al*., 2005) produced in recombinant proteins. Generally, acceptance of cAMP as second messengers in plants was delayed (Gehring, 2010) because their levels in plants has been reported to be very much lower than in animals.

The progression of more accurate techniques with higher detection capabilities has provided a proper key on the occurrence of cAMP in plants, thus removing scepticism over its existence. Such scepticism was mainly overcome by the use of mass spectrometric analysis, which was able to provide unequivocal evidence on the existence of cAMP and other cyclic nucleotides in higher plants (Richards *et al*., 2002). Tandem mass spectrometry was also useful in demonstrating the activities of plant ACs (Pacini *et al*., 1993; Witters *et al*., 1996) and capillary HPLC electrospray mass spectrometry also proved to be a powerful analytical technique for the detection limits of low plant cAMP as much as 25 femtomoles (Witters *et al*., 1996). Thus, technological advancement has been very crucial in allowing the detecting of cyclic nucleotides in higher plants and their ultimate acceptance.

However, the most widely used methods to study endogenous cAMP are currently based on the methods of Gilman (1970), which are based on competitive interaction of labelled cAMP and sample cAMP being bound to a specific protein. For instance, the enzyme immunoassay (EIA) (Egorov *et al*., 1991) method that is currently used in endogenous plant cAMP studies to detect cAMP levels through the estimation of the rate of cAMP formation. This method is based on the measurement of cAMP levels in the plant sample through the use of antibodies (Lomovatskaya *et al*., 2011). The EIA methods are based on the principle of competitive displacement of a bound cAMP-enzyme conjugate with the cAMP present in the sample under analysis and the activity of the ACs then becomes directly proportional to the concentration of cAMP in the sample under analysis (Tijseen, 1985). Recent studies have demonstrated the essential role of EIA in cAMP quantification in plants (Ito *et al*., 2014; Świeżawska *et al*., 2014; Al-Younis *et al*., 2015; Chatukuta *et al*., 2018).

In soybean, previous efforts to determine AC activity were performed using the isotope dilution method, alternatively known as the cAMP protein-binding assay. The technique failed to detect any meaningful cAMP levels in this plant's tissues (Yunghans and Morre, 1977). However, progressive studies have shown that cAMP activity is indeed present in legumes since the nucleotide has been implicated in the regulation of nodule formation and function (Terakado *et al*., 1997). Therefore, it is important to gain an insight into the biological roles of the effector molecule AC behind the cAMP signalling pathway in the soybean plant since the concentration of cAMP is a vital indicator of the functional activity of the AC signalling system in this important legume (Lomovatskaya *et al*., 2011).

Reports from several laboratories have indicated that cAMP levels are correlated with a number of physiological processes in plants. In *Nicotiana plumbaginofolia*, cAMP has been implicated in the up-regulation of the cytosolic Ca^{2+} concentration (Volotovski *et al*., 1998), and in Arabidopsis roots, cAMP was reported to cause the down-regulation of Na^{+} influx and accumulation (Isner and

Maathuis, 2011). In the mesophyll cells of *Vicia faba*, intracellular application of cAMP showed a concentration-dependent increase in the outward K^+ current during whole-cell patch-clamp current recordings (Li *et al*., 1994). This implies that cation fluxes are regulated by cAMP, thus regulating ion and water homoeostasis in plants. Cell cycle progression in tobacco BY-2 cells was also reported to be influenced by the levels of cAMP with the highest levels observed in the S phase (Ehsan *et al.*, 1998) and the lowest levels noted at the beginning of the G_1 phase in arrested cells (Ehsan *et al*., 1999). Hence the progression of the cell cycle is largely influenced by stringent levels of cAMP. In *Orobanche minor*, seed germination was observed to require elevated levels of cAMP (Uematsu *et al*., 2007) and in nodules of *G. max* and *Vigna angularis*, high levels of cAMP have been implicated in the regulation of symbiotic root nodule formation (Terakado *et al*., 1997). These studies, therefore, indicate the vital role of cAMP as a modulator of plant metabolism and development.

The increasing recognition of cyclic nucleotides has shed more light in their involvement in the perception of extracellular biotic and abiotic stimuli through amplification and transduction of specific signals to corresponding responses (Gehring and Turek 2017). A lot of research has shed insights into the role of the endogenous cAMP in plant defense responses, resulting in transcript increases of defense related genes. Several studies have reported production of elevated cAMP concentrations at infection sites. These were demonstrated in *Phaseolus vulgaris* (Bolwell, 1992) and *Medicago sativa* treated with the glycoprotein of *Verticillium alboatrum* (Cooke *et al*., 1994), and in Arabidopsis exposed to the *Verticullium dahlia* toxins (Jiang *et al*., 2005). In all these situations, the elevated cAMP concentrations have resulted in improved disease resistance in the plants. There is no doubt to the important role played by cAMP in promoting plant defense against pathogen attack. Furthermore, in *A. thaliana* leaves, wounding stress was shown to result in almost a five-fold increase in the concentration of cAMP (van Damme *et al*., 2014). The evidence

discussed here undoubtedly establish the role of cAMP in the regulation of both the biotic and abiotic stress tolerance, thus aiding in plant development and survival.

The cloning of AC genes, their expression as recombinant proteins in suitable vectors and their subsequent purification has offered researchers in plant biology an excellent opportunity to study and determine the AC activities of novel ACs in plants. However, most expression and purification methods if heterologous and chemical, result in the production of biologically inactive recombinant proteins. During protein expression in *E. coli*, large quantities of the recombinant proteins are made, but the proteins are usually deposited as insoluble inactive inclusion bodies in the cytoplasm of *E. coli* (Clark, 1998,) due to mis-folding of the expressed proteins (Mihic and Harris, 1996). In some cases, expressed recombinant proteins have also been reported to contain endotoxins, which disrupt biological assays (Franken *et al*., 2000; Campos *et al*., 2008) hence require purification of the recombinant protein. Usually and during cloning, the recombinant protein is tagged to improve its solubility and purification processes. The tags are fused on either the N- or C-terminal end of the protein of interest (Gupta and Shukla, 2016). Currently there are several protein tags that can be used as fusion tags in *E. coli*, and these include the His6, GST, MBP, NusA, Trx and SUMO.

The His-tag is the most extensively used tag because of its efficient affinity purification properties on the fused protein as it contains 6 or more consecutive histidine residues (Gupta and Shukla, 2016). The His-tag allows for a one-step purification process of recombinant proteins even in the presence of denaturants like urea and guanidine hydrochloride (Kneusel *et al*., 1998). Proteins that have been tagged with the His-tag can be purified using a Ni-NTA affinity purification system (Campos *et al*., 2008). The Ni 6xHis act as a metal binding site for the recombinant protein thus facilitating the purification process. Furthermore, to effectively study the biological functions of genes, it is important to recover the biologically active recombinant protein from the inclusion bodies (Yamaguchi and Miyazaki, 2014). Protein refolding is a crucial process that is practised to

recover active recombinant proteins (Franken *et al*., 2000) and protein refolding can be defined as the change of protein confirmation from an unfolded to folded forms and is influenced by denaturant concentrations. The mis-folding of denatured proteins is usually as a result of a very rapid decrease in the concentration of the denaturant (Tsumoto *et al*., 2003; Ho *et al*., 2003). Otherwise a gradual decrease of the denaturant concentration usually leads to efficient protein refolding (Yamaguchi and Miyazaki, 2014).

Here we report the affinity purification of a recombinant GmAC1 protein from *G. max* using the Ni-NTA affinity purification system, followed by refolding of its denatured form using a BioLogic DuoFlow Chromatographic System. This recombinant protein, which is part of the larger protein complex XP_003529590 in soybean, was herein targeted because it is annotated to contain the putative AC catalytic centre (Gehring, 2010; Ito *et al*., 2014) presumably responsible for the generation of cAMP in the *G. max*. Ideally, whilst the general roles of ACs and cAMP in most plants have been extensively studied, not much is currently known in *G. max*.

5.2 Materials and methods

5.2.1 Over-expression of the recombinant GmAC1 protein

The general yield and levels of the recombinant GmAC1 protein, partially expressed in section 3.3.1 of Chapter 3, were enhanced through over-expression of the recombinant protein. Instead of the smaller volume of 20 ml and a lower IPTG concentration of 1 mM (as was in section 3.3 of Chapter 3) to express the protein, 100 ml of the LB broth supplemented with 35 µg/ml chloramphenicol, 100 µ/ml ampicillin and 0.5% glucose were induced with 4 mM ITPG to over-express the recombinant GmAC1 protein. After expression, the induced culture was then centrifuged at 9 200x*g* for 10 minutes to pellet the cells. The supernatant was discarded while the pelleted cells were then used for the intended protein purification.

5.2.2 Determination of the soluble or insoluble status of the expressed recombinant GmAC1 protein

The soluble/insoluble status of the expressed recombinant GmAC1 protein was determined through native non-denaturing conditions by lysing the pellet followed by an analysis of the cellular protein component through SDS-PAGE. The whole pellet of the cells harvested above (section 5.2.1) was thoroughly resuspended in 2.5 ml of sterile Tris-buffered saline (TBS) buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) and 0.5 µg/ml lysozyme (Catalogue number 62970; Sigma-Aldrich Inc., Missouri, USA). The resultant mixture was then incubated on ice for 60 minutes while periodically, being vortexed at medium speed for 10 minutes with some 1-minute intervals of ice incubation. This process was key to slowly rapture cells and release their contents. The resultant mixture was then centrifuged for 2 minutes at 9 200x*g* to separate the pellet from the supernatant. The supernatant was collected and kept as the soluble fraction of the protein whereas the pellet was kept as the insoluble fraction of the protein. 20 µl of each of the two protein fractions were then analysed by SDS-PAGE to determine the presence and relative quantities of the expressed recombinant GmAC1 protein. The presence of the recombinant protein in the supernatant fraction would then mean its soluble status while its presence in the pellet fraction would, on the other hand, would indicate its insoluble nature.

5.2.3 Purification of the recombinant GmAC1 protein

5.2.3.1 Preparation of the cleared lysate

After realising that the recombinant GmAC1 protein was wholly expressed in form of inclusion bodies, the whole insoluble fraction of the GmAC1 from section 5.2.2 above was then resuspended

in 5 ml of Lysis buffer (100 mM NaH_2PO_4, 10 mM Tris-Cl, 8 M urea, 10 mM β-mercaptoethanol, 500 mM NaCl, 15% glycerol, pH 8) and subjected to vigorous vortexing for 60 minutes. This was done to ensure that the pellet harbouring the insoluble recombinant GmAC1 protein was wholly solubilised by the chaotropic urea salt into a crude lysate. The produced crude lysate was then centrifuged at 10 000x*g* for 10 minutes at room temperature to pellet out the cell debris and generate a cleared lysate, which was collected and kept for further processing on the HisPur Ni-NTA purification matrix (Catalogue # 88221, ThermoFisher Scientific) following the QIAexpressionist protocol (QIAexpressionist, 2006).

5.2.3.2 Equilibration of the HisPur Ni-NTA resin matrix

An aliquot of about 4 ml of the HisPur Ni-NTA resins supplied as 50% slurry in 20% ethanol were transferred into a 15 ml falcon tube at room temperature. The resin was then washed three times with filter sterilised distilled water to remove the ethanol. The resin was then equilibrated by mixing it with 4 ml of the Lysis buffer through rotation of the mixture for 10 minutes at room temperature on a Rotator Revolver (Labnet International Inc., New Jersey, USA). The resin-buffer mixture was allowed to settle, and the buffer was removed and discarded while the equilibrated resins were kept at 4°C for further use.

5.2.3.3 Affinity purification of the recombinant GmAC1 protein

The cleared lysate produced in section 5.2.3.1 was transferred to the equilibrated HisPur Ni-NTA resins from section 5.2.3.2. The mixture was mixed gently for 60 minutes at room temperature on a Rotator Revolver at 30 rpm. This stage was crucial as it allowed for the recombinant GmAC1 protein to bind to the HisPur Ni-NTA resins. The binding was facilitated by the presence of three

divalent metal-binding sites on the resin, which allowed specific and high binding capacity with the 6xHis-tagged recombinant GmAC1 protein. Thereafter, the mixture was allowed to settle, and the supernatant collected as flow-through. About 30 µl of the flow-through was collected in an Eppendorf tube for analysis by SDS-PAGE. The bound HisPur Ni-NTA resins were then washed with the Wash buffer (100 mM NaH_2PO_4, 10 mM Tris-Cl, 8 M urea, 10 mM β-mercaptoethanol, 500 mM NaCl, 15% glycerol, pH 8) three times. During each wash, a 30 µl portion was collected and stored for analysis by SDS-PAGE. Finally, an aliquot (30 µl) of the bound HisPur Ni-NTA resins was also collected for analysis by SDS-PAGE.

5.2.4 Protein renaturation

The pellet solubilisation performed in section 5.2.3.1 above, resulted in the denaturation of the recombinant GmAC1 protein. Therefore, in order to restore the functional state of this targeted protein, a refolding or renaturation process had to be performed.

5.2.4.1 Renaturation of the recombinant GmAC1 protein

The washed HisPur Ni-NTA resins carrying the recombinant GmAC1 protein were first resuspended in 10 column volumes (cv) of the Refolding Buffer A (8 M Urea, 200 mM NaCl, 50 mM Tris-HCl, 500 mM glucose, 20 mM β-mercaptoethanol, pH: 8.0). The resuspended slurry was then loaded into a clean XK16 column (Bio-Rad Laboratories Inc, California, USA) before the column was manually connected to a Biologic DouFlow Chromatography System (Bio-Rad Laboritories Inc., California, USA). A linearised gradient system was then created based on the parameters listed in Table 5.1. During this refolding process, the Gradient Buffer A was slowly and linearly diluted to 0 M urea concentration with a Refolding Buffer B (200 mM NaCl, 50 mM

Tris-HCl, 500 mM glucose, 0.05% (w/v) polyethlygylcol, 4 mM reduced glutathione, 0.04 mM oxidised glutathione, and 0.5 mM phenylmethylsulphonyl floride). Both buffer systems were connected to a BioLogic DuoFlow Chromatographic System to facilitate a controlled renaturation process of the denatured recombinant GmAC1 protein. Therefore, during this process the purified recombinant GmAC1 protein was converted into its native and biologically active status.

Table 5.2: Renaturation conditions of the recombinant GmAC1 protein using the BioLogic DuoFlow Chromatographic system.

Variable	Value
Column volume	1.00 ml
Flow rate	0.50 ml/min
Column pressure limit	2.80 MPa
Average time for UV	1.00 sec
System pump for automatic pressure and flow regulation	Normal
Starting concentration of Buffer B	0.00%
Target concentration of Buffer A	100%
Length of gradient	900 min

5.2.5 Elution of the recombinant GmAC1 protein

The refolded, bound and fully purified recombinant GmAC1 protein was ultimately eluted off the HisPur Ni-NTA resins through the addition of 2 ml Elution buffer (200 mM NaCl, 50 mM Tris-HCl (pH 8.0), 250 mM imidazole, 0.5 mM PMSF, and 20% (v/v) glycerol) using the elution process directly linked to the DuoFlow system. The resultant supernatant containing the eluted protein was then collected and stored at 4ºC for further use.

5.2.6 Concentration and desalting of the recombinant GmAC1 protein

The eluted refolded recombinant GmAC1 protein was freed from its buffering salts and excess water by pouring the whole 2 ml eluent into the upper chamber of the Spin-X UF de-salting and concentrating device (Corning Life Sciences Inc., New York, USA) followed by centrifugation at 2 540x*g* at 4^{o}C for 4 hours or until the final volume had gone down to 100 µl. The concentrated and de-salted protein fraction was then removed from the device and transferred into a new Eppendorf tube. Protein concentration was subsequently determined by a 2000 Nanodrop spectrophotometer (Thermo Scientific, California, USA). About 30 µl of the eluted recombinant protein were collected for analysis by SDS-PAGE while the rest of protein was stored at 4ºC for further use.

5.2.7 Functional characterisation of the recombinant GmAC1 protein

The *in vitro* characterisation of the AC activity of the purified and refolded recombinant GmAC1 protein was performed via a cAMP-linked enzyme immunoassaying (EIA) system, following its acetylation protocol and in accordance with the manufacturer's instructions (Catalogue #CA201: Sigma-Aldrich Inc., Missouri, USA). The process involved determination of the functional effects/influences of various ionic and/or molecular components such as GTP, Mg^{2+}, Mn^{2+}, Ca^{2+}, HCO_3^- and F^- onto the enzymatic activity of the recombinant GmAC1 protein.

5.2.7.1 Sample preparation and enzyme immunoassaying

In nine 1.5 ml Eppendorf tubes, 200 µl reaction mixes were prepared as is detailed in Table 5.2 below. All the reaction mixes were then incubated for 20 minutes at room temperature before reactions were stopped by the addition of 1 mM EDTA, which chelated out all the divalent metal

ions, thus removing the co-factors necessary for the enzymatic action of the recombinant GmAC1 protein. The samples were then boiled for 5 minutes to inactivate the recombinant protein. The samples were then centrifuged at 16 300x*g* for 5 minutes to clarify them followed by collection and use of their clarified fractions. The clarified fractions in all the 9 tubes were then acetylated by adding an acetylating reagent (2:1, triethylamine:acetic anhydride (v/v)) at a volumetric ratio of 1:20 acetylating reagent:sample. The acetylation process of the samples was performed in order to increase the sensitivity of the assay. Immediately after adding the acetylating reagent, the reaction solutions were mixed vigorously by vortexing at high speed for 2 seconds and their inherent cAMP levels measured using the acetylation protocol in the enzyme immunoassay (EIA) cyclic AMP kit (catalog CA201 Sigma-Aldrich) and in accordance with the manufacturer's protocol (Appendix A). The used kit utilises a conjugated polyclonal antibody to cAMP to bind in a competitive manner with the sample cAMP under investigation. The enzyme immunoassaying samples and standards were all run in triplicate (n = 3) and the means of the triplicates were then subjected to a one-way analysis of variance (ANOVA) test.

Table 5.3: Molecular characterisation of the recombinant GmAC1 protein.

	Tube 1	**Tube 2**	**Tube 3**	**Tube 4**	**Tube 5**	**Tube 6**	**Tube 7**	**Tube 8**	**Tube 9**
Tris	50mM	50mM	50Mm	50mM	50mM	50mM	50mM	50mM	50mM
$MnCl_2$	5mM	5mM	-	5mM	5mM	5mM	5mM	5mM	5mM
Protein	-	25µg	25µg	25µg	25µg	25µg	25µg	25µg	25µg
ATP	1mM	1mM	1mM	1mM	1mM	1mM	-	1mM	1mM
$MgCl_2$	-	-	5 mM	-	-	-	-	-	-
$CaCl_2$	-	-	-	250µM	-	-	-	-	-
GTP	-	-	-	-	-	1mM	1mM	-	-
$NaHCO_3$	-	-	-	-	50mM	-	-	-	-
NaF	-	-	-	-	-	-	-	-	10mM
$AlCl_3$									+30µM

5.3 Results

5.3.1 Determination of the solubility/insolubility status of the recombinant GmAC1 protein

The recombinant GmAC1 protein was expressed in BL21 (DE3) pLysS *E. coli* expression cells in form of insoluble inclusion bodies, which result in an aggregated and non-functional protein product. This was determined by the solubility/insolubility test performed in section 5.2.2, whereby when the expressing cells were raptured in a native buffer followed by resolution (by SDS-PAGE) of the soluble and insoluble fractions of their cellular contents, the bulk of the expressed recombinant GmAC1 protein could solely be found in the insoluble cellular fraction as is shown in Figure 5.1 below.

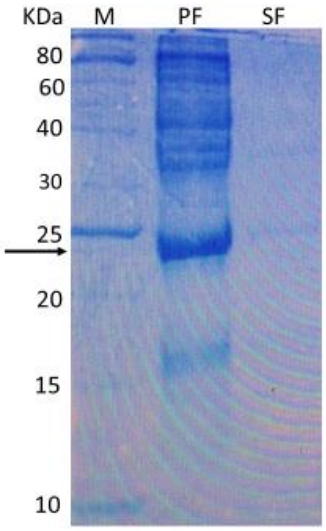


Figure 5.1: Determination of the insolubility/solubility status of GmAC1 protein. The presentation is a 12% SDS-PAGE gel showing the solubility/insolubility status of the recombinant GmAC1 protein. M is the molecular weight marker, PF is the soluble fraction (pellet) of the cell content while SF is the soluble fraction (supernatant) of the cell content. The arrow is marking the recombinant GmAC1 protein, which was wholly expressed as insoluble component of the cell.

5.3.2 Affinity purification of the recombinant GmAC1 protein

The expressed insoluble His-tagged recombinant GmAC1 protein was purified on a HisPur Ni-NTA resin matrix under non-native denaturing conditions, using an 8 M lysis buffer. The protein was purified through its affinity binding onto the HisPur Ni-NTA resin matrix from the cleared lysate and washing off all other contaminants as is shown in Figure 5.2 below.

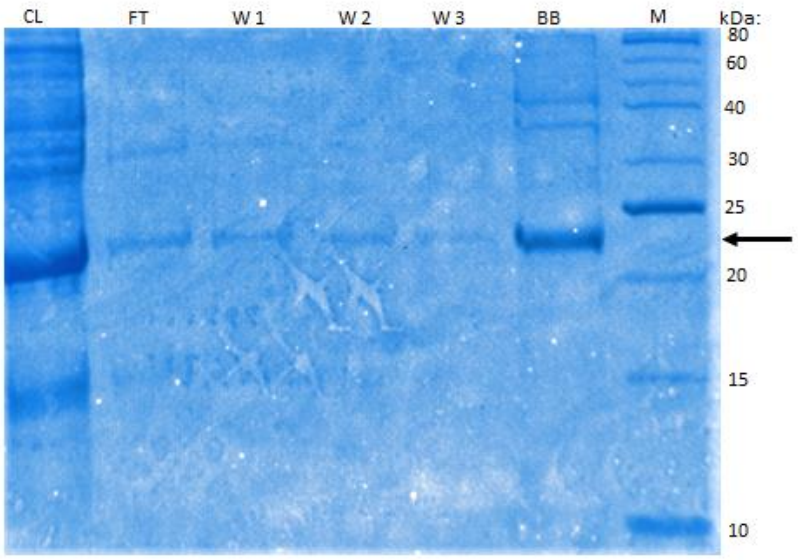


Figure 5.2: Purification of the recombinant GmAC1 protein under non-native denaturing conditions. The presentation is a 12% SDS-PAGE of the purification regimes of the expressed recombinant GmAC1 protein. CL is the cleared lysate before passing through the HisPur Ni-NTA resins, FT is the flow-through of the cleared lysate after it had been passed through HisPur Ni-NTA resins. W1, W2 and W3 are three successive washes of the purified GmAC1 protein bound on the HisPur Ni-NTA resins, and BB is the purified recombinant GmAC1 protein still bound onto the HisPur Ni-NTA resins. M is the molecular weight marker and the arrow is marking the purified recombinant GmAC1 protein still bound onto the HisPur Ni-NTA resins.

5.3.3 Renaturation and elution of the recombinant GmAC1 protein

The denatured and purified recombinant GmAC1 protein was refolded back into its native form whilst it was still bound onto the HisPur Ni-NTA resin matrix to restore its biological properties and then eluted off the HisPur Ni-NTA resin matrix using the BioLogic DuoFlow Chromatography System (Bio-Rad Laboritories Inc., California, USA). The eluted protein was then resolved and analysed by SDS-PAGE as shown in Figure 5.3 below.

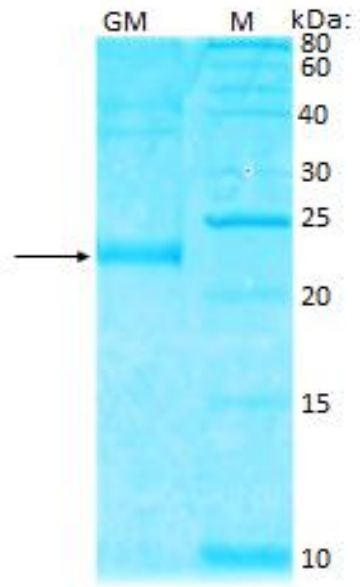


Figure 5.3: Refolding and elution of the purified of the recombinant GmAC1 protein. The presentation is a 12% SDS-PAGE of the eluted version of the recombinant GmAC1 protein after chemical refolding HisPur Ni-NTA resin matrix. M is the molecular weight marker and GM the recombinant GmAC1 protein. The arrow marks the eluted, desalted and concentrated protein fraction of the recombinant GmAC1 protein.

5.3.4 Characterisation of the enzymatic activity of the recombinant GmAC1 protein

After elution, the AC activity of the refolded purified recombinant GmAC1 protein was then assessed *in vitro* and in the presence of other ionic and molecular components that are naturally known to influence and/or affect AC activity. The levels of cAMP generated by these various assessment conditions are shown in Figure 5.4 below.

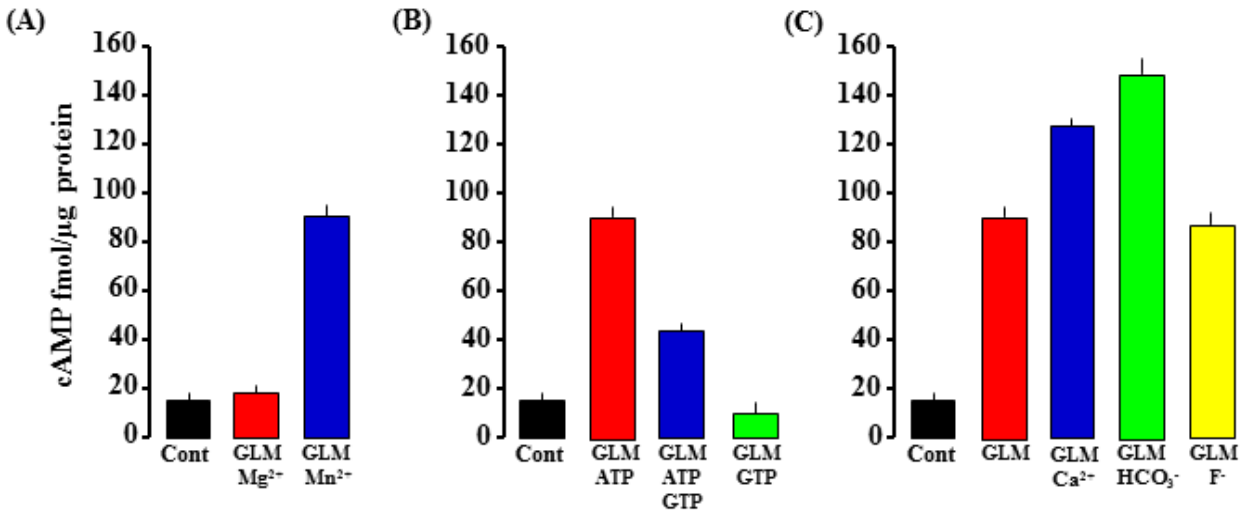


Figure 5.4: Characterisation of the AC activity of the recombinant GmAC1 protein. A reaction mixture containing 10 µg of the refolded purified recombinant GmAC1 protein, 50 mM Tris-Cl, pH: 8.0, 2 mM IBMX, 5 mM Mn^{2+}, 1 mM ATP and/or in the presence of other additives was prepared and incubated at room temperature for 20 minutes. The cAMP levels generated from these reaction mixes were then measured with a cAMP-specific enzyme immunoassaying system (Sigma Aldrich Inc., Missouri, USA) based on its acetylation protocol. (**A**) cAMP levels generated by the purified recombinant GmAC1 protein in the presence of Mg^{2+} or Mn^{2+} ions; (**B**) cAMP levels generated with the purified recombinant GmAC1 protein in the presence of either ATP and/or GTP when Mn^{2+} ion is the co-factor of activity; and (**C**) cAMP levels generated with the recombinant GmAC1 protein in the presence of either calcium, bicarbonate or fluoride ions when Mn^{2+} ion is the co-factor of activity. All assays were undertaken in triplicate sets (n = 3) and error bars represent standard errors of the triplicate means (n = 3).

5.4 Discussion

In this chapter, the C-terminal His-tagged recombinant GmAC1 protein was over expressed following induction of the transformed cells with 4 mM IPTG. Unfortunately, this over-expression then resulted in the production of this protein in form of inclusion bodies (Figure 5.1) that are both insoluble and non-functional. The production of the recombinant GmAC1 protein as inclusion bodies in *E. coli* BL21 (DE3) pLysS expression cells is probably due to the inability of the bacterial system to support all the post-translational modifications required by GmAC1 protein to fold

(Baneyx and Mujacic, 2004). This is because the recombinant GmAC1 protein was synthesised in a microenvironment of *E. coli* (bacterial), which is quite different from that of *G. max* (plant) (Yamaguchi and Miyazaki, 2014). As a result of this insoluble nature of the recombinant GmAC1 protein, its purification process was then undertaken on a HisPur Ni-NTA resin matrix under non-native denaturing conditions (Figure 5.2). This purification process was made possible basically, because of the presence of a C-terminal 6xHis tag on the pTrcHis2 cloning vector that then enabled for successful affinity purification of the expressed GmAC1 protein on the HisPur Ni-NTA affinity matrix. The purification of the recombinant GmAC1 protein was further enabled mainly due to the ability of the histidine tag to viably and stably function under denaturing and non-native experimental conditions (Holzinger *et al*., 1996; Mihic and Harris, 1996).

However, this purification process resulted in the denaturation of the recombinant protein and thus rendering the protein biologically inactive. Therefore, it was very important to first recover a biologically active protein product through refolding (Yamaguchi and Miyazaki, 2014) before any subsequent experimental analysis could be initiated. This refolding process together with its subsequent elution process of the purified denatured recombinant GmAC1 protein then yielded a purified biologically active protein that was visualised on SDS-PAGE as is shown in Figure 5.3 and later used in the *in vitro* functional characterisation assay shown in Figure 5.4.

Results of the *in vitro* characterisation of the enzymatic activity of the recombinant GmAC1 protein showed that the protein was able to generate significant levels of cAMP in the presence of Mn^{2+}, HCO_3^-, and Ca^{2+} molecular ions. Previous research by Al-Younis *et al.* (2015; 2018), Bianchet *et al.* (2019), Chatukuta *et al.* (2018), and Ruzvidzo *et al.* (2013), have all demonstrated the role of these molecular and ionic components in positively regulating the function of novel ACs in higher plants. In our case, the recombinant GmAC1 protein generated around 90 fmols of cAMP per µg protein in the presence of Mn^{2+} as compared to around 20 fmols cAMP per µg protein in the presence of Mg^{2+} (Figure 5.4A). This is an indication that the GmAC1 protein has

a clear preference for the Mn^{2+} and not the Mg^{2+} metal ion as a cofactor of its catalytic activity and consistent with the AC activities of AtKUP7/5 (Al-Younis *et al*., 2015; 2018) and AtLRRAC1 (Bianchet *et al*., 2019). A report made by Londos and Preston (1977) showed that the divalent ions, Mn^{2+} and Mg^{2+}, were capable of activating the AC system through interaction with their metal ion binding sites, however, activation by Mg^{2+} was reported to require concentrations that are at least 50 to 100 folds greater than those for Mn^{2+}. This probably explains the results obtained in this study, where higher cAMP levels were observed in the presence of Mn^{2+} than Mg^{2+}, particularly considering that the two metal ions were used at equimolar levels. More so, the fact that the recombinant GmAC1 protein could generate relatively more cAMP content in the presence of Mn^{2+} than Mg^{2+} perhaps suggests that the protein could be a soluble (sAC) since sAC are active and functional only in the presence Mn^{2+} but not Mg^{2+} (Tresguerres *et al*., 2011).

When comparing ATP and GTP as probable substrates for the recombinant GmAC1, the protein generated around 90 fmols of cAMP with the former compared to approximately 10 fmols of cAMP with the latter (Figure 5.4B). These results, together with the ones in Figure 5.4A, strongly indicate that the novel recombinant GmAC1 protein is strictly dependent upon the Mn^{2+} metal ion as a co-factor and ATP as a sole substrate, and thus implying that the protein is indeed a sAC (Braun and Dods, 1975).

Furthermore, the GmAC1 protein was activated by the HCO_3^- and Ca^{2+} ions, to produce significant cAMP levels of ±140 and ±125 fmols respectively compared to the ~90 fmols of cAMP it generally generates in the absence of both molecular ions. This further confirms that the expressed recombinant GmAC1 protein is indeed a sAC since a prominent feature of all sACs is their direct stimulation by both the HCO_3^- and Ca^{2+} ionic components (Chen *et al*., 2000; Litvin *et al*., 2003). sACs have been reported to require two divalent metal ions in their catalytic active sites for the enzymes to be able to coordinate the binding and ultimately the cyclisation of ATP into cAMP

(Steegborn, 2005), thus explaining the GmAC1's stimulative and modulative sensitivity to both Ca^{2+} and Mn^{2+}.

In our initial preliminary bioinformatic analysis through the GenPRoBis tool (Chapter 2), the GmAC1 protein (as part of the XP_003529590 protein complex) was predicted to possess an analog ligand, 1HAK, that is involved in calcium binding (Table 2.6). This therefore implies that the Ca^{2+} ions could perhaps have been able to bind to some site or sites on the recombinant GmAC1 protein as previously reported by Sadana and Dessauer (2009). More so, it has also been established that divalent ions (Ca^{2+} and Mn^{2+}) result in an increase of cAMP levels for various ACs (Limbird *et al*., 1979; Neer, 1979; Litvin *et al*., 2003) thus these same metal ions could presumably have had a similar effect onto the catalytic site of the studied recombinant GmAC1 protein in our case. According to Cooper *et al.* (1995), this is an indication that ACs are associated with sites of Ca^{2+} sensitisation in all cells and thus the observed coupling of Ca^{2+} and cAMP production in our case, firmly provides evidence for the existence of a Ca^{2+}-cAMP signalling system in plants that is possibly regulated by the GmAC1 protein.

Additionally, our preliminary bioinformatic analysis of the novel GmAC1 gene in Chapter 2 (Table 2.2) also predicted its primary location as mitochondrion. sACs are primarily associated with intracellular organelles such centrioles, microtubules, nuclei (Zippin *et al*., 2003; Schmid *et al*., 2007) and are mostly localised in the mitochondrial matrix (Acin-Perez *et al*., 2009; Tresguerres *et al*., 2014; Valsecchi *et al*., 2014), thus confirming the recombinant GmAC1 protein as a sAC. Ideally, it is believed that activation of sAC in the mitochorion is also primarily through binding by bicarbonate, which directly binds to and activates the sAC in a non pH-dependent manner (Chen *et al*., 2000). The mitochondrion is a site well-known for producing ATP and within this organelle, metabolic respiration results in increase of bicarbonate ions (from the generated CO_2), thus providing ideal conditions for sAC activation and modulation. Incidentally, the information provided here is also very consistent with some other previous reports that have

implicated the sensitivity of sACs to variations in the levels of the physiological ATP and their direct activation by Ca^{2+} and the metabolite bicarbonate (Chen *et al*., 2000; Litvin *et al*., 2003).

Recent literature has also implicated the bicarbonate anion in being capable of eliciting cAMP signals in the matrix by allowing sACs to independently generate cAMP (Acin-Perez *et al*., 2011). The activity of sACs has been reported to increase synergistically in the presence of the bicarbonate and Ca^{2+} ions (Litvin *et al*., 2003; Geng *et al*., 2005) and to be dependant on the physiological changes in ATP concentrations such that when ATP levels are reduced, the activities of sACs are also decreased to due to substrate limitation (Zippin *et al*., 2013).

In our present study, the activation of the novel recombinant GmAC1 protein *in vitro* by both the Ca^{2+} and HCO_3^- ions implies that the protein is indeed an intracellular metabolite and signal intergrator. Though no literature is available on the role of the mitochondrially-generated cAMP in plants or soybean specifically, previous studies have focused on roles of such signalling systems in animals (Valsecchi *et al*., 2014; Pozdniakova and Ladilov, 2018). In mammals, the physiological functions regulated by sACs, included glycolysis and mitochondrial respiration (Acin-Perez *et al*., 2009; 2011), and ACs were also reported to be involved in the regulation of pathogens such as *Candida albacans* and *Pseudomonas aeruginosa* (Hall *et al*., 2010; Topal *et al*., 2012).

Therefore, the possible function of the soluble recombinant GmAC1 protein can be associated with disease resistance and metabolic regulation within the soybean plant. An attempt to understand the regulation of sACs by HCO_3^- and Ca^{2+} in mitochondria was performed by Di Benedetto *et al.* in 2013, where their data also confirmed the importance of ions in activating sACs and cAMP production in mitochondria. Experimental data from this group was also able to confirm that the target of mitochondrial cAMP is a protein kinase A (PKA) or a PKA-like kinase - a sentiment also confirmed by Lefkimmiatis *et al*. in 2013. Therefore, it can be postulated that the presumable

soybean sAC-cAMP signalling system may also involve a PKA or PKA-like kinase pathway. This pathway is important in regulating downstream signalling and gene transcription. The sAC-cAMP-PKA pathway has been implicated in playing a central role in the transduction of environmental signals (Guo *et al*., 2016) and in the un-restrained growth and division of cells (Manning *et al*., 2002). In fungi, the same signalling pathway was reported to be involved in growth, development, stress response and secondary metabolism (Hu *et al*., 2014).

In plants, protein kinases have been implicated in regulation of cell differentiation, growth and development of the plant body (Ho, 2015), therefore, in order to comprehend the physiological roles of the GmAC1 sAC-cAMP-PKA signalling system, the general roles of PKA in plants can be assessed and elucidated. Ca^{2+} as a modulator of sACs has been reported to stimulate protein kinase activity in membranes of the pea shoot (Hetherington and Trewavas, 1984) - a claim that was confirmed by Hepler and Wayne in 1985 that the stimulation of protein kinases is dependent on Ca^{2+} concentration. Therefore, the observed modulation of the GmAC1 sAC by Ca^{2+} in this study shows the importance of this divalent ion in promoting signal transduction in soybean specifically and other plants in general.

Protein kinases have been implicated in abiotic stress regulation such as cold, salt and drought and biotic stresses involving pathogens, pests, herbivores and fungi (Ho, 2015). In response to abiotic and biotic stress exposures, protein phosphorylation by the protein kinases was observed in plant cells of Arabidopsis and the response was believed to be involved in the regulation of physiology, morphology, and gene expression, thus affecting cell proliferation, growth and development (Shinozaki and Yamaguchi-Shinozaki, 2000; Fowler and Thomashow, 2002). The importanc of calcium in modulating sACs to produce cAMP can not be underestimated as previous studies have implicated the increase of Ca^{2+} in the plant cytosol as a crucial event in plant innate immune responses (Sunkar *et al*., 2000; Bridges *et al*., 2005). Therefore, the production of cAMP through

the activation of the recombinant GmAC1 protein by Ca^{2+} can be implicated in plant defense related signalling (Clough *et al*., 2000) within the legume.

5.5 Conclusion

The *in vitro* activity assaying of the recombinant GmAC1 protein confirmed this protein as a sAC, whose physiological roles may involve the sAC-cAMP-PKA signalling system.

CHAPTER 6

EXTENSIVE BIOINFORMATIC ANALYSIS OF THE NOVEL XP_003529590 SOYBEAN GENE

Abstract

Unknown gene function predictions can be established using bioinformatic tools. These web-based and computer-linked tools have made it possible to understand the physiological roles of new genes in plants. In this chapter, the possible physiological roles of a novel soybean gene XP_003529590 were predicted. This gene is annotated as a probable adenylate cyclase (AC - an enzyme that is capable of generating the signal molecule, cAMP) due to the fact that it harbours a putative AC centre, which in various chapters of this study, has been cloned, expressed as a truncated version of the XP_003529590 gene, GmAC1 and ultimately confirmed to be catalytically functional. The major tools used herein include DOMAINATION and Genevestigator. DOMAINATION was used to infer structural domains in the XP_003529590 protein sequence from similarity searches using PSI-BLAST. Genevestigator was used to predict gene function through gene co-expression during soybean development, various perturbations and gene expression profiling through micro arrays experiments from the [Soybean] Affymetrix Soybean Genome Array platform. Gene Ontology of the biological process predictions, molecular function predictions and cellular component predictions were all performed through FFPred 2.0. Results from DOMAINATION showed that the XP_003529590 gene is indeed an AC whereas results from Genevestigator showed that the XP_003529590 gene is relatively expressed during soybean development as was indicated by a Pearson correlation coefficient of 1 for all the top 25 co-expressed genes. Microarray gene expression profiling showed that the XP_003529590 gene

is initially down-regulated due to exposure to the soybean root pathogen *Phytophthora sojae* and later, up-regulated after extended exposures while the opposite is true for it when the plant is exposed to the other root pathogen *Phakopsora pachyrhizi*. On contrary, the XP_003529590 gene is, however, always up-regulated during the entire bud removal, a mimicry of herbivory. Gene ontologies additionally showed localisation of the XP_003529590 gene as being chiefly the mitochondrion. The major biological process predictions included regulation of gene expression, translation and metabolic processes, while molecular process predictions were catalytic activity, nucleic acid binding, ATP binding, RNA binding, kinase binding and kinase activity. From all these aspects, it could therefore, be firmly deduced that the XP_003529590 protein is indeed an RNA binding AC that is largely involved in regulation of development in the soybean plant through expression of genes involved in pathogen defense, drought, cold and salt tolerance, detoxification, and growth and development systems via processes that largely are mediated by cAMP.

6.1 Introduction

One of the key objectives in any biological investigation is to identify all molecules within a living cell and understand their interaction. Unfortunately, the functions of many of the genes are still not understood and such a scenario has been exacerbated with the recent discovery of many novel genes, particularly in soybean, which has been previously reported to have undergone almost over 2.55-fold genome duplications (Turner *et al*., 2012). In Chapter 2, the preliminary bioinformatic analysis of the XP_003529590 gene has shown that the protein product it expresses is an alpha-helical PPR protein, largely involved in RNA and Ca^{2+} binding processes.

Glycine max, being one of the major oil crops world-wide, faces various challenges posed by environmental stressors and in order to cope, this plant evolved sophisticated adaptive response

mechanisms (Yamaguchi-Shinozaki and Shinozaki, 2006). Some of the plant responses to different stresses are defined by regulatory factors that integrate signalling from various pathways (Fujita *et al*., 2006). An attempt to understand such regulatory factors has been made possible through the use of bioinformatic tools. Bioinformatics approaches offer essential tools for the identification of genes and pathways that may be associated with important adaptive responsive mechanisms in crop plants. including soybean. Unravelling the complex resistant mechanisms usually provides fundamental insights into the biological processes involved in environmental stimuli and therefore, helping in alleviating crop losses.

Bioinformatics tools use the Affymetrix microarray platform, which provides a standardised system with a high degree of reproducibility (Hennig *et al*., 2003; Redman *et al*., 2004) for analyses. The Affymetrix system has made it possible to identify biologically significant expression patterns of individual genes (Zimmerman *et al*., 2004) and therefore, helping in illuminating gene functions. The inferring of gene functions from different types of biological data has shown to bring more accurate predictions and therefore, has been widely used in different researches (Pavlidis *et al*., 2001; Mostafavi and Morris, 2010; Gillis and Pavlidis, 2011). Biological functions of unknown genes can be obtained through similarity searches that include multiple sequence alignments (MSAs) and gene co-expression among other methods.

Multiple sequence alignments (MSAs) have been reported to be an essential tool for protein classification, analysis and functional prediction (Hannenhalli and Russell, 2000). These MSAs arrange protein sequences into a rectangular array with the aim that residues in a given column are homologous, superposable or play a common functional role (Edgar and Batzoglou, 2006). The exploitation of similarities in protein sequences to infer gene function has been most widely used (Phuong and Nhung, 2013). Through MSAs, it is possible to view sequence conservation patterns that are indicative of enzyme active sites and secondary structure types as has been performed in some previous analyses by Casari *et al*., in 1995 and Lichtarge *et al*., in 1996. Casari *et al*. (1995)

developed a method for identifying functional residues on proteins based on MSA, through analysis of conserved positions throughout the whole protein family. Lichtarge *et al*. (1996) were also able to determine important positions on protein sequences that were of functional importance. Co-expression network analysis (CNA) is one method that can be used to infer gene function and gene-disease associations from genome-wide gene expression. This CNA method can be used to associate genes of unknown function with biological processes to prioritise candidate disease genes and to understand transcriptional regulatory programmes (van Dam, 2018). Gene CNA has been reported to be effective in identifying correlations, i.e. which genes are active simultaneously, which indicates that they are active in the same biological processes. Over the years, gene expression data has become a powerful resource in describing the molecular state associated with many cellular phenotypes and responses to environmental perturbations. The use of expression profiling has been well demonstrated in predicting class of known genes and in assigning putative functional roles to previously uncharacterised genes based on profile similarity (Carter *et al*., 2004).

In model organisms such as *Saccharomyces cerevisae* and *Arabidopsis thaliana*, there exists a wealth of publicly available gene expression data from experiments designed to perturb the transcriptomes of the organisms by subjecting them to various environmental conditions (Deshmukh *et al*., 2013). However, in *Glycine max*, the transcriptome data is really quite limited when compared to *A. thaliana* and *Oryza sativa* (Wang *et al*., 2014) such that information on proteomics, metabolics and phenomics is reportedly lagging (Deshmukh *et al*., 2013). In order, to elucidate and identify gene functions, the data from these experiments have been subjected to many clustering algorithms in an effort to identify genes that are involved in similar functional roles (Eisen *et al*., 1998; Tamayo *et al*., 1999). The kind of significance analysis has provided insight into specific cellular states through the identification of individual genes that possess corresponding altered expressions (Ideker *et al*., 2000; Rocke and Durbin, 2001). Under varying

environmental conditions, cell transcriptional responses involve a coordinated co-expression of genes encoding proteins that work in concert to achieve an adaptive response (Carter *et al*., 2004).

Therefore, by examining conditional patterns of expression, an insight into the underlying cellular processes that are activated is provided. A study by Carter *et al*. (2004) on the gene co-expression network topology of *S. cerevisiae* and *Caenorhabditi elegans* has demonstrated that molecular characterisation of biological conditions is possible through examination of differential expression profiles in that particular condition and that gene expression network connectivity is a relevant dimension that specifies a gene's involvement in the environmental response program under investigation by microarray experiment. It therefore, follows that the measurement of expression levels of genes through microarray analysis offers great opportunities for probing transcript patterns in plant organs, tissues or cells.

Soybean, an economically important drought sensitive crop, requires extensive study in order to understand its genome responses to the various environmental stresses. Bioinformatic tools have been utilised to understand the function of some of the genes that are involved in combating the effects of stresses. Some of the genes in soybean that have been studied through bioinformatic analyses are the heat shock protein 20 (*Hsp*20) gene family and few others (Lopes-Caitar *et al*., 2013). In this regard, authors employed similarity sequence in phylogenetic analyses to understand relationships within the sub-families of the gene family and Basic Local Alignment Search Tool (BLAST) searches to identify *Hsp*20 domains. A study of the expression profiles of the *Hsp*20 was performed using the gene expression search tools of the soybean database that are available in Genevestigator (https://www.genevestigator.com/gv/plant.jsp) (Hruz *et al*., 2008) and Soybase (http://soybase.org/soyseq/) (Severin *et al*., 2010). In that regard, Lopes-Caitar and co-workers (2013) were able to conclude that the bioinformatic tools were very efficient in precisely identifying the *G. max Hsp*20 family members that were useful in elucidating their roles in the event of heat stress and the tissues in which such genes exert their effects.

Bioinformatic tools have also been employed as well in studying the roles of *G. max WRKY*-type transcription factors in plant defense response and developmental processes by Zhou *et al*. (2008). That kind of a research proved that an understanding of such genes from *G. max* is important in the production of transgenic plants. The study involved identification of 64 *WRKY*-type transcription factors from soybean and characterisation of their expression pattern in response to different abiotic stress factors. The authors employed BLAST to identify the 64 *G. max WRKY*-type transcription factors, from which they were able to identify three members; *GmWRKY13, GmWRKY21* and *GmWRKY54* that were able to show differential effects on abiotic stress tolerance. A better understanding of such plant genes can further be used in crop improvement. The results of the bioinformatic data mining of those *G. max WRKY*-type transcription factors by Zhou *et al*. (2008) provided a baseline for field experiments involving the growth of *GmWRKY21*-transgenic Arabidopsis plants and the subsequent abiotic stress treatments. It can therefore, be concluded that microarray data provides an indication of gene expression patterns hence illuminating targets for further studies. It is because of this that microarray-based studies of plant-pathogen interaction are now prevalent.

Gene expression studies in soybean have also been performed through expressed sequence tags (EST) sequencing, spotted microarrays and the Affymetrix GeneChip technology (Severin *et al*., 2010). These include the laser capture microdissection for study of soybean seed development (Le *et al*., 2007) and iron stress response in soybean (O'Rourke *et al*., 2009). This is in contrast to the RNA-Seq, which was employed by Severin *et al*. (2010) for gene expression studies in soybean - which is limited to analysis of gene expression in tissues and seeds, and thus ideally suitable to plant breeders. The gene expression analyses in seeds focused on the genes with functional roles that aid or complement in economically important seed filling processes.

There is evidence that gene expression microarrays have been used to examine plant and pathogen gene expressions over a time course of post-infection. Moy *et al*. (2004) utilised an amplified-

cDNA microarray that had 3 927 soybean and 969 *Phytophthora sojae* genes for expression studies. Their results indicated that the soybean genes encoding the enzymes of phytoalexin biosynthesis and defense and pathogen-related proteins were strongly up-regulated during infection and their expression peaked at 24 hours after infection. *P. sojae* is an oomycete, narrow range species pathogen that is restricted primarily to soybean (Erwin and Ribiero, 1996). Therefore, *P. sojae* is a major pathogen of *G. max,* which causes root and stem rot (Shan *et al*., 2003; Tyler, 2007; Dou *et al*., 2008) and has been reported to be the major cause of soybean losses (Zhang *et al*., 2011). *Phakopsora pachyrhizi* is another major pathogen of soybean and has been reported to have caused significant economic losses in Africa, Asia and South America (Panthee *et al*., 2007). Inference can therefore, be made that these two pathogens are a serious threat to soybean production, thus requiring proper management.

Therefore, an evaluation of the endogenous up-regulation of genes in response to disease- causing agents is most likely to provide clues to pertinent defense genes. This is possible through microarray analysis, a useful technology for assaying transcriptional responses for abiotic and biotic plant stresses (Moy *et al*., 2004; Mentewab *et al*., 2005). In soybean, microarray analysis has been used to study and understand the root rot disease caused by *P. sojae* (Moy *et al*., 2004), therefore, the technology is useful in identifying fungal resistance genes in plants. Libault *et al*. (2010) used the same technology in the analysis of the soybean root hair cell transcriptome in response to *B. japonicum* inoculation to identify genes that are regulated during root hair infection by the bacteria.

The use of the soybean Affymetrix GeneChips and microarray data was thereof considered in our own study herein since it provides information on transcriptomic analysis for abiotic and biotic stress tolerance in soybean and is more readily available. It is important to note that there still is no published data on microarray gene expression studies on the XP_003529590 gene in soybean.

6.2 Materials and methods

6.2.1 Functional prediction of the XP_003529590 gene through multiple sequence alignments

DOMAINATION (http://mathbio.nir.nimr.ac.uk) (George and Heringa, 2002) or simply domain prediction was used to infer structural domains in the amino acid sequence of the XP_003529590 encoded protein from similarity searches using PSI-BLAST. In other words, a similarity search of the XP_003529590 protein sequences against BLASTP 2.2. 17 (Altschul *et al*., 1997) was made using DOMAINATION and the accuracy of such a search tested using some composition-based statistics (Schaffer *et al*., 2001). To supplement data from this DOMAINATION, PLAZA 4.0 (https://bioinformatics.psb.ugent.be/plaza) (van Bel *et al*., 2018) was used to further functionally characterise the XP_003529590 gene coding sequences. Since a search input of the Gyma.07G251000 in PLAZA 4.0 classified this gene as belonging to the gene family HO04D005110, therefore, a functional clustering analysis of 15 species of economic importance was performed through MSA. The resultant genes from the heatmap were then further characterised and ascertained through Phytozome v12.1.6 (https://phytozome.jgi.doe.gov/pz/portal.html).

6.2.2 Functional prediction of the XP_003529590 gene through co-expression analysis

To assess which genes in soybean are functionally co-expressed with the XP_003529590 gene a similarity search tool, Co-expression from the Compendium Wide Analysis; Genevestigator v3 (http://www.genevestigator.ethz.ch) (Hruz *et al*., 2008) was used. The tool was useful in identifying genes that have the most similar profile to XP_003529590 across the GM_AFFY_SOYBEAN database. Co-expression was tested across two parameters; development

and perturbations. For each parameter, 25 top most co-expressed genes were identified based on the Pearson correlation coefficient as a measure of similarity between the genes.

6.2.3 Functional prediction of the XP_003529590 gene through stimuli expression responses

An exploration of the XP_003529590 gene expression in response to a wide variety of perturbations was investigated using the compendium-wide analysis- condition search tool: perturbations in Genevestigator v3 (http://www.genevestigator.ethz.ch) (Hruz *et al*., 2008). This was done to identify experimental conditions in which the XP_003529590 gene expression can be up-regulated or down-regulated. Expression profiling through arrays experiments from the [Soybean] Affymetrix Soybean Genome Array platform were used to assess the gene expression patterns of the novel gene in soybean, and these included:

(i) Gene expression in *P. sojae* mycelia, germinating zoospores and during its infection of the soybean hypocotyls (Dong *et al*., 2009);

(ii) Expression survey of the Rpp1 soybean line P1200492 resistant to *P. pachyrhizi* (Gregory *et al*., 2007);

(iii) The effects of bud removal on soybean leaf gene expression (Turner *et al*., 2012);

(iv) Soybean root hair cell responses to *B. japonicum* inoculation (Libault *et al*., 2010) and

(v) Analysis of iron deficiency in soybean leaf tissues (O'Rourke *et al*., 2009).

The default values for the fold change and p value were used since there were only 244 perturbations for the *G. max*. The expression levels of the genes were expressed as log2 ratio and statistically computed using the *t*-test for each of the 7 genes against the XP_003529590. The 7 *G. max* genes that were used to compute the expression levels of the gene of interest were obtained

from GM_AFFY_SOYBEAN-0 dataset (based on the selection criteria). These were Glyma.07G184300: Mitochondrial HSO70 2; Glyma.19G242700: Ribosomal protein S24e family protein; Glyma.13G237800: Ribosomal protein S6e (RP-S6e, RPSL); Glyma.05G029300: Uncharacterised protein; Glyma.03G150600: GTP binding elongation Factor Tu family protein; Glyma.08G184300: Ribosomal protein L10 family protein; and Glyma.10G154500: rRNA processing protein-related.

6.2.4 Functional prediction of the XP_003529590 gene through Gene Ontology

The gene ontology predictions for the XP_003529590 gene were performed using FFPred 2.0 which was accessed at the interface of the PSIPRED Protein Analysis Workbench (http://bioinf.cs.ucl.ac/psipred) (Minneci *et al*., 2013). This FFPred runs a Support Vector Machine (SVM) in which GO that are considered highly reliable are denoted H and those which are harder to predict as low reliability denoted L (Minneci *et al*., 2013).

6.3 Results

6.3.1 Functional prediction of the XP_003529590 gene through multiple sequence alignments

A similarity sequence alignment search of the XP_003529590 protein sequence using the BLASTP 2.2.17 PSI-BLAST in DOMAINATION, revealed 7 domains shown in Figure 6.1 below, whose similarity sequences are shown in Appendix B. The domains were identified from *N. tabacum* (AAB87670.1); *O. sativa* ssp *japonica* (NP_00060578.1); and *A. thaliana* (NP_193299.3, BAC42709.1, NP_188783.2, NP_001030740.1 and CAB10342.1). In addition, the composition-based statistics for the observed domains are shown in Table 6.1.

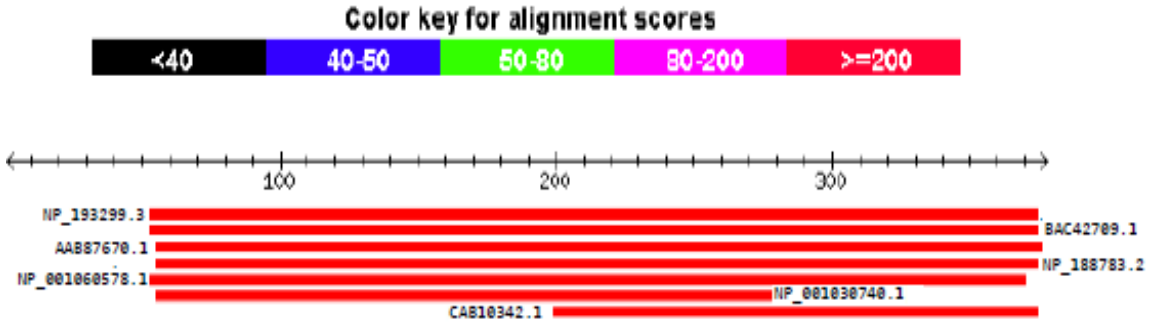


Figure 6.1: Protein domains from BLASTP with sequence similarities to the XP_003529590 gene. DOMAINATION produced 7 domains with significant alignment scores as is indicated in the colour key.

Table 6.1: Gene IDs for the 7 domains with significant sequence scores to the XP_003529590 gene. UP means uncharacterised protein, AC means adenylyl cyclase, and HP means hypothetical protein.

Sequence Locus	Gene ID	Score (bits)	Length Aa	Composition based stats	
				Identities	Positives
NP_193299.3	At4g15640 (AC)	403	390	60	76
BAC42709.1	*A. thaliana* (UP)	402	390	60	76
AAB87670.1	*Axi 141 N. tabacum* (AC)	402	406	60	75
NP_188783.2	At3g21465 (AC)	396	388	59	75
NP_00060578.1	Os07g668200 (AC)	348	413	54	71
NP_001030740.1	*A. thaliana* (UP)	306	312	62	77
CAB10342.1	*A. thaliana* (HP)	222	189	61	78

At4g15640, Os07g668200 and AAB87670.1 have been categorised as AC genes (Altschul *et al*., 1997; Schaffer *et al*., 2001).

The gene IDs shown in Table 6.1 above demonstrate that most of those shared domains are AC catalytic centres. However, for the sequence loci BAC42709.1 (UP), NP_001030740.1 (UP) and CAB10342.1 (HP), since there is a significant sequence similarity in their domain sequences, it can be predicted that they are also catalytic in nature, possessing the annotated AC. The seven

domains sequence loci from Figure 6.1 showed significant similarities to the soybean XP_003529590 gene. In addition, a functional clustering analysis of the gene family HO4D005110 (InterPro-Tetratricopeptide-like helical domain) of some chosen 15 economically important species that was performed in PLAZA 4.0, generated a similarity heatmap. The heatmap was constructed based on the protein sequences of the HO4D005110 tetratricopeptide like helical domain as is shown Figure 6.2 below. The results indicated that within the *G. max* plant, there is a Glyma.17G023300, which shows a close similarity to the tetratricopeptide-like helical domain of the AC gene that was studied herein. The *A. thaliana* gene At4g15640 from Table 6.1 above has since been identified in the gene family search of HO4D005110 together with At3g21465.

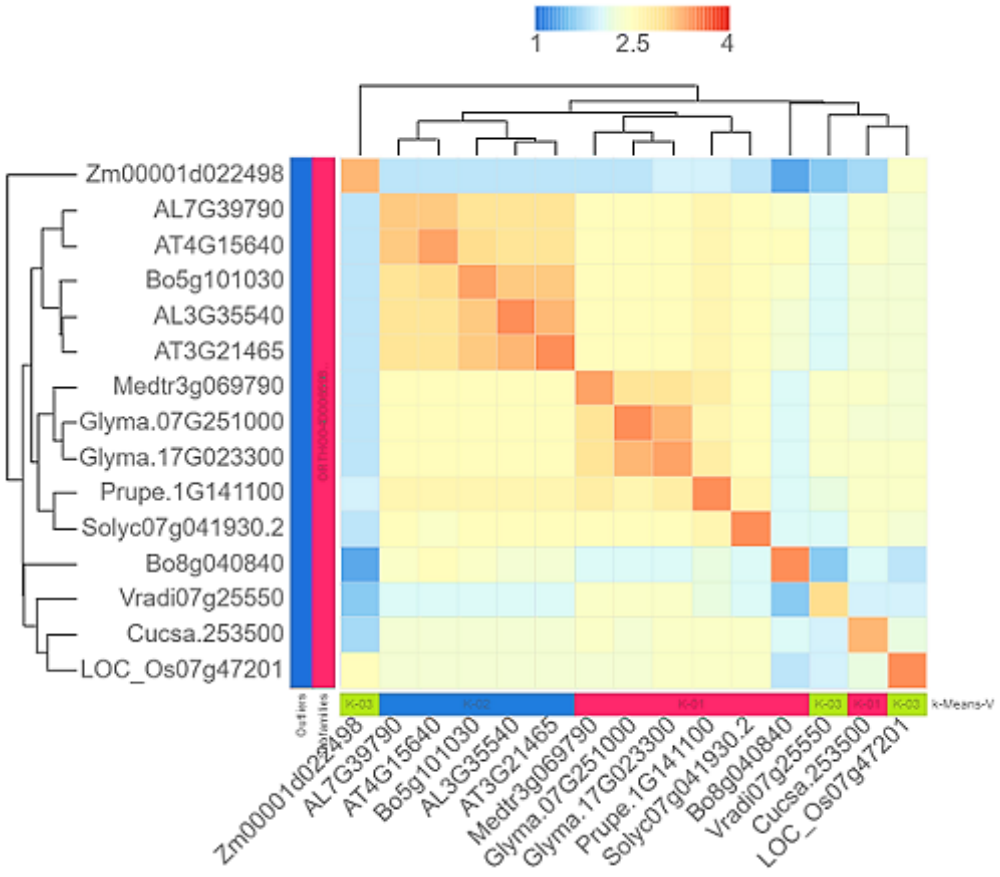


Figure 6.2: Similarity heat map based on the gene family HO04D005110 InterPro tetratricopeptide-like helical protein.

A further inference into the gene families of the genes in Figure 6.2 was made using Phytozome 12. The results of such a further analysis of the genes with close similarity to the XP_003529590 gene are shown in Table 6.2 below.

Table 6.4: A Phytozome 12 inference of the HO4D005110 gene family with close similarity to the XP_003529590 (https://phytozome.jgi.doe.gov/pz/portal.html).

Gene	Naming Platform	Description
Glyma.17G023300	*Glycine max Wm82.a.2v1*	Adenylyl cyclase
AL7g39790	*Arabidopsis lyrata v2.1*	Adenylyl cyclase
AL3g35540	*Arabidopsis lyrata v2.1*	Adenylyl cyclase
Medtr3g069790	*Medigo truncatula Mt4 0v1*	Adenylyl cyclase
Prupe.1G141100	*Prunus persica v2.1*	Adenylyl cyclase
Solyc07G041930.2	*Solanum lycopersicum iTAG2. 4*	Adenylyl cyclase
Cucsa.253500	*Cucumis sativus v1. 0*	Adenylyl cyclase
LOC_Os07g47201	*Oryza sativa v7_JGI*	Adenylyl cyclase, putative, expressed
At3g21465	*Arabidopsis thaliana TAIR 10*	Adenylyl cyclase
At4g15640	*Arabidopsis thaliana TAIR 10*	Adenylyl cyclase
Zm0000b1022498	*Zea mays PH207v1. 1*	RING and CHY zinc finger domain (s)
Bo5g101030	*Brassica oleracea capitate v1. 0.*	Uncharacterised protein
Bo8g040840	*Brassica oleracea capitate v1.0*	Uncharacterised protein
Vradi07g25550	*Vigna angularis*	Uncharacterised protein

An analysis of the tetratricopeptide-like helical domain family gene was able to identify another possible novel AC enzyme from the *G. max*; gene ID: Glyma.17G023300. Most of the genes from the other plants are ACs except for the gene from *Zea mays*, which is a RING finger and CHY finger domain-containing protein.

6.3.2 Functional prediction of the XP_003529590 through co-expression analysis

A similarity search tool Co-expression from the Compendium Wide Analysis; Genevestigator v3 showed 25 top-most genes that are co-expressed with the XP_003529590 during development as is illustrated in Table 6.3 below. All the top 25 co-expressed genes showed a Pearson correlation coefficient of 1, an indication of a strong relation of such development genes to the

XP_003529590. However, under the perturbations tool, not a single gene from the top 25 co-expressed genes had a Pearson correlation coefficient of 1. The top 25 co-expressed genes during perturbations are shown in Table 6.4 below. Interestingly most of the genes that are co-expressed with XP_003529590 gene under perturbations are ribosomal proteins unlike those co-expressed during development.

Table 6.5: The 25 top most genes positively co-expressed with the XP_003529590 gene in soybean during development.

Gene	Description
Glyma.07G251000	
Glyma.04G229000	60s acidic ribosomal protein family
Glyma.05G175900	Protein of unknown function (DUF3550/UPF0682)
Glyma.03G259800	Rho GTPase activating protein with PKA-box /P21-RHO-binding
Glyma.11G231800	Outer membrane OMP85 family protein
Glyma.08G025700	Heat shock protein 70 (Hsp 70) family protein
Glyma.02G266300	Indole-3-butyric acid response 5
Glyma.04G234500	ABL interactor-like protein 2
Glyma.11G925000	Protein of unknown function
Glyma.18G042700	Nonsense-mediated mRNA decay NMD3 family
Glyma.06G074300	CHY-type/CTCHY-type/ RING-type Zinc finger protein
Glyma.14G074900	S-adenosyl-L-methionine-dependent methyltransfarese
Glyma.19G072800	Ubiquitin-conjugating enzyme 22
Glyma.05G148200	Nuclear transporter factor 2B
Glyma.17G105900	Ribosomal protein S11-beta
Glyma.18G025100	Leucine-rich repeat protein kinase family
Glyma.07G017500	Cytochrome c oxidase assembly protein COX19-like
Glyma.08G175200	Glutathione S-transferase TAU 19
Glyma.02G004400	Tetracopeptide repeat (TPR)-like superfamily
Glyma.15G015700	PDI-like 1-6
Glyma.14G065900	Glucose-1- phosphate adenylyltransferase family protein
Glyma.02G073700	Plasma membrane intrinsic protein 2
Glyma.19G184300	ATP binding cassette subfamily B1
Glyma.11G175200	Cyclophilin 5
Glyma.19G207100	B-box type zinc finger protein with CCT domain
Glyma.06G358000	Protein of unknown function

All the top 25 genes co-expressed with the XP_003529590 during the legume development were highly correlated, showing Pearson correlation coefficient of 1.

Table 6.6: The 25 top most genes positively co-expressed with the XP_003529590 gene in response to various perturbations.

Gene	Description
Glyma.07G251000	
Glyma.07G184300*	Mitochondrial HSO70 2
Glyma.19G242700**	Ribosomal protein S24e family protein
Glyma.13G237800**	Ribosomal protein S24e family protein
Glyma.05G029300***	Protein of unknown function
Glyma.03G150600****	GTP binding elongation factor Tu family protein
Glyma.08G184300****	Ribosomal protein L10 family protein
Glyma.10G154500****	rRNA processing protein-related
Glyma.13G276300****	Ribosomal protein L3 plastid
SGlyma.02G263200*****	Ribosomal protein S4 (RPS4A) family protein
Glyma.07G043800*****	40s ribosomal protein SA B
Glyma.07G128300*****	Protein of unknown function
Glyma.02G002300*****	Ribosomal protein L7Ae/L30e/s12e/Gadd45 family
Glyma.20G168100*****	Ribosomal protein 1
Glyma.06G045000*****	Ankyrin repeat family protein
Glyma.17G105700*****	Ribosomal protein S5/Elongation factor G/III/V family
Glyma.05G081300******	Ribosomal protein L25/Gln-tRNA synthetase, anti-codon-binding
Glyma.19G125400******	Nascent polypeptide-associated complex (NAC), alpha subunit family
Glyma.03G055500******	Metallopeptidase M24 family protein
Glyma.03G186500******	Transducin family protein/WD-40 repeat family protein
Glyma.17G023300******	Protein of unknown function
Glyma.02G049300******	Ribosomal protein L2 family
Glyma.08G087200******	Ribosomal L27e protein family
Glyma.11G163000******	Mitochondrial 28s ribosomal protein S2
Glyma.18G278400******	Pentatricopeptide repeat (PPR) superfamily protein
Glyma.02G047800******	Ribosomal protein S4 (RPS4A) family protein

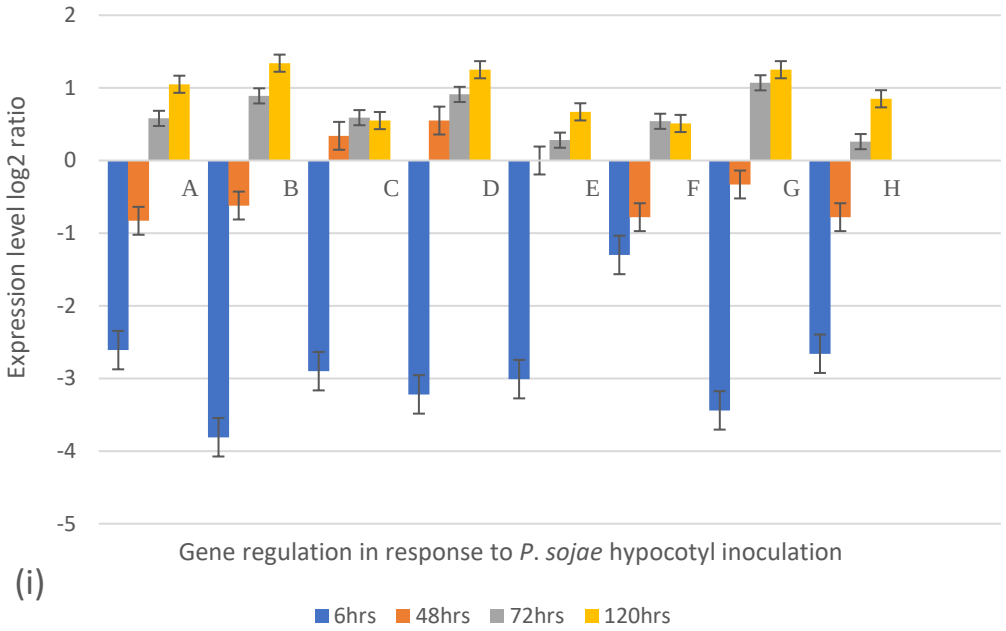
Pearson correlation coefficient *0.88, **0.87, ***0.86, ****0.85, *****0.84, ******0.83. **Glyma.17G023300** is another putative AC gene from *G. max* identified from the HO04D005110 gene family similarity heat map in PLAZA 4.0 Figure 6.2.

6.3.3 Functional prediction of the XP_003529590 gene through stimuli expression responses

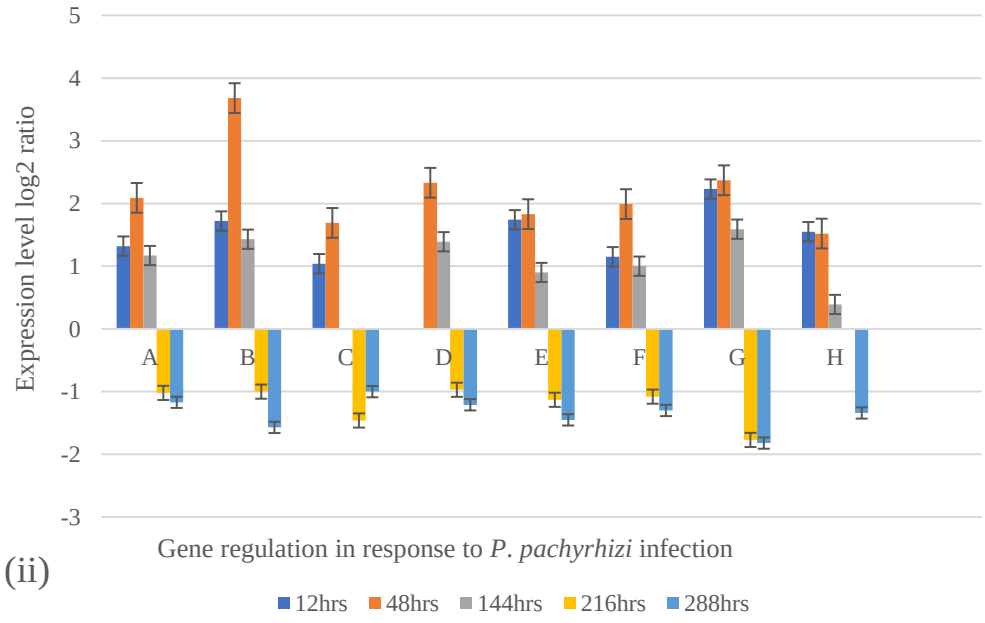
The transcriptional responses of the soybean plant to different perturbations are presented under biotic stress and mineralisation as the two main stimuli categories. The results below show gene expression profiles of the XP_003529590 gene, whereby the Affymetrix Soybean Genome Analysis platform was used to identify expression profiles of the XP_003529590 gene and 7 other co-expressed genes under different experimental perturbations.

6.3.3.1 Expression profile of the XP_003529590 in response to biotic stress

The XP_003529590 gene was found to be down-regulated during the initial exposure of the soybean roots to the pathogen *P. sojae* (within 48 hours), however, a continued exposure of the soybean roots (>72 hours) to the pathogen, resulted in the up-regulation of the XP_003529590 gene (Figure 6.3 (i)). In addition, inoculation of the soybean leaf with the *P. pachyrhizi* resulted in the up-regulation of the gene within 144 hours, however, longer exposure periods to the pathogen then resulted in the down-regulation of the gene, including all the other co-expressed genes (Figure 6.3 (ii)). Apparently, when herbivory biotic stress was mimicked through bud removal and allowing bud growth between time intervals, the XP_003529590 gene was shown to be up-regulated during the bud removal stage (Figure 6.3 (iii)).


2
1
0
-1
-2
-3
-4
-5
Expression level log2 ratio
A
B
C
D
E
F
G
H
Gene regulation in response to P. sojae hypocotyl inoculation
6hrs
48hrs
72hrs
120hrs

(i)


5
4
3
2
1
0
-1
-2
-3
Expression level log2 ratio
A
B
C
D
E
F
G
H
Gene regulation in response to P. pachyrhizi infection
12hrs
48hrs
144hrs
216hrs
288hrs

(ii)

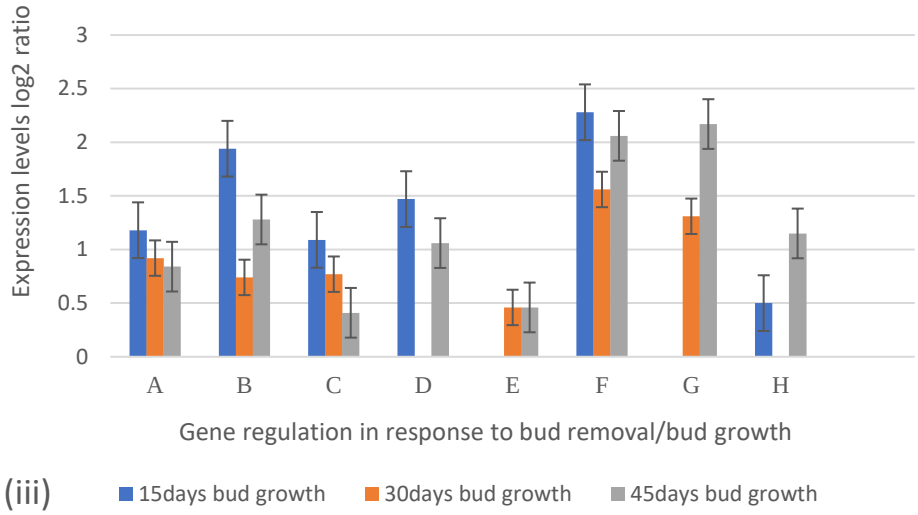


Figure 6.3: Expression profile of the XP_003529590 gene in response to biotic stress. The XP_003529590 gene and its co-expressed genes are represented as follows; A: Glyma.07G251000, B: Glyma.07G184300, C: Glyma.19G242700, D: Glyma.13G237800, E: Glyma.05G029300, F: Glyma.03G150600, G: Glyma.08G184300, H: Glyma.10G154500. Error bars show the standard error. (i) XP_003529590 gene was initially down-regulated and after 48hrs, up-regulated after infection of the soybean hypocotyls with the fungal pathogen *P. sojae*. (ii) The XP_003529590 gene was shown to be initially up-regulated with infection of the soybean plant by the fungal pathogen *P. pachyrhizi* but later down-regulated after 144 hours. (iii) A mimic of herbivory through bud removal and bud growth, showing that the XP_003529590 gene is up-regulated.

6.3.3.2 Gene expression profile of the XP_003529590 gene following mineralisation

Soybean roots that were inoculated with the bacteria *B. japonicum* that is involved in the legume's nodulation, showed an up-regulation of the XP_003529590 gene. However, in root hair cells, the XP_003529590 gene was down-regulated after 48 hours of exposure (Figure 6.4 (i)). In the same Figure, an analysis of the iron deficiency on the regulation of XP_003529590 shows that limiting iron nutrient supply to the legume resulted in the up-regulation of the gene (Figure 6.4 (ii)).

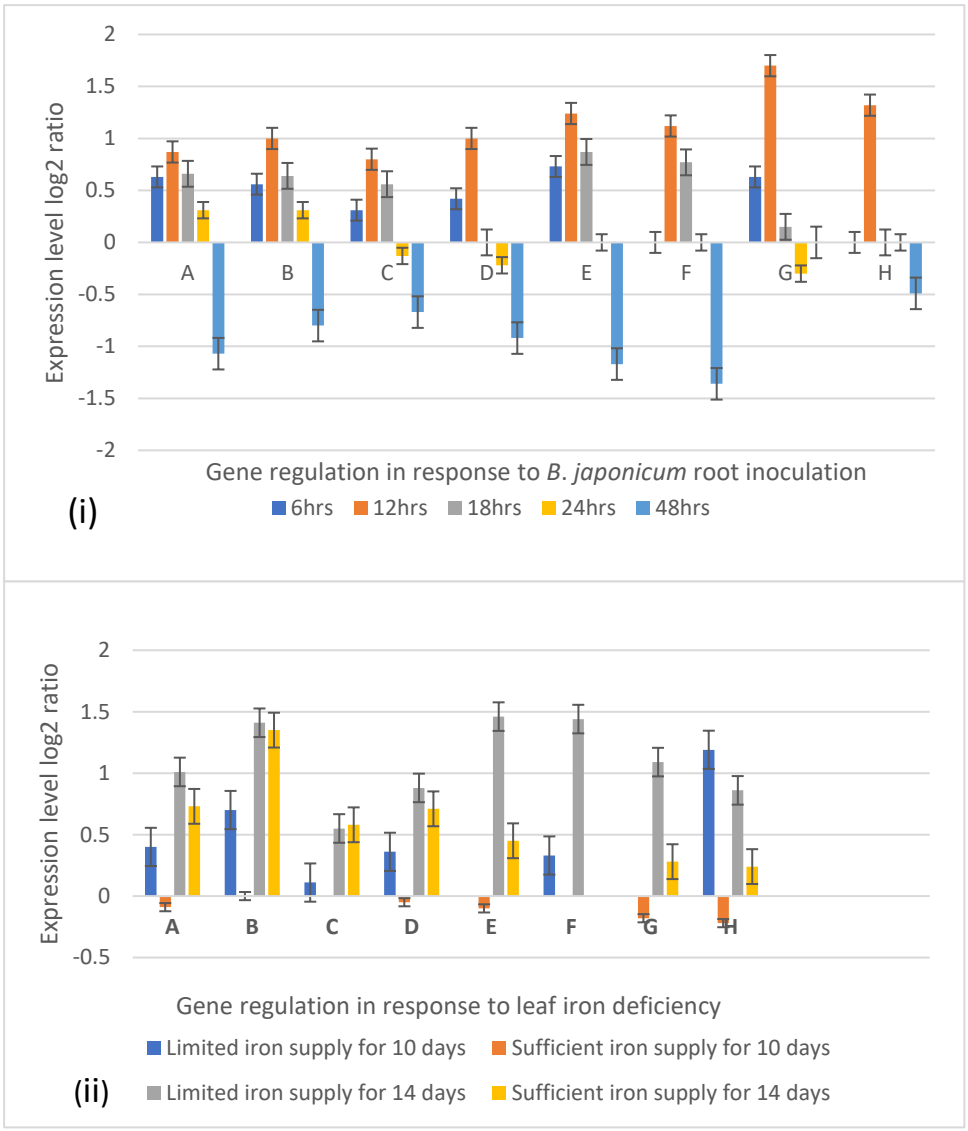


Figure 6.4: Expression profile of the XP_003529590 in response to nutrient supply. The XP_003529590 gene and its co-expressed genes are represented as follows A: Glyma.07G251000, B: Glyma.07G184300, C: Glyma.19G242700, D: Glyma.13G237800, E: Glyma.05G029300, F: Glyma.03G150600, G: Glyma.08G184300, H: Glyma.10G154500. Error bars show the standard error. (i) Soybean bean roots inoculated with the *B. japonicum* resulted in the up-regulation of the gene, however longer period (48 hours) of the soybean inoculation resulted in the down-regulation of the XP_003529590 gene. (ii) Limiting iron nutrient to the soybean plant for 10 and 14 days resulting in the up-regulation of the AC gene, and a sufficient supply for 14 days showed an up-regulation of the XP_003529590 gene.

6.3.4 Gene Ontology FFPred analysis

Results of the FFPred 2.0 component of Phyre2 showed the molecular functions of the XP_003529590 gene to be largely involved in binding; mainly RNA binding (GO:0003723 and GO:0044822), ATP binding (GO:0005524), kinase binding (GO:0019900) and receptor binding (GO:0005102). The cellular component prediction confirmed some other earlier predictions that the XP_003529590 gene is largely localised in the mitochondrion (GO:0005759, GO:0005740, GO:0005739, GO:0031966, GO:0005743). The biological functions predictions then implicated the XP_003529590 gene in regulation of gene expression (GO:0010468), gene expression (GO:0010467), anatomical structure development (GO:0048856) and regulation of signal transduction (GO:0009966). A summary of the GO terms is shown in Table 6.5 below.

Table 6.7: FFPred Gene Ontology Analysis of the XP_003529590 gene.

GO term	Name	SVM reliability
Biological Process Predictions		
GO:0019222	Regulation of metabolic process	High
GO:0030001	Metal ion transport	High
GO:0010468	Regulation of gene expression	High
GO:0009059	Macromolecule biosynthesis process	High
GO:0006412	Translation	High
GO:0045333	Cellular respiration	High
GO:0006396	RNA processing	High
GO:0050896	Response to stimulus	High
GO:0044237	Cellular response to stimulus	High
GO:0023052	Signalling	High
GO:0048856	Anatomical structure development	Low
GO:0007165	Signal transduction	Low
GO:0009966	Regulation of signal transduction	Low
GO:0032502	Developmental process	Low
Molecular Function Predictions		
GO:0003824	Catalytic activity	High
GO:0003676	Nucleic acid binding	High
GO:0000166	Nucleotide binding	High
GO:0005524	ATP binding	High
GO:0005506	Iron ion binding	High
GO:0019900	Kinase binding	High
GO:0016301	Kinase activity	High
GO:0003723	RNA binding	High
GO:0016787	Hydrolase activity	Low
GO:0043169	Cation binding	Low
GO:0005102	Receptor binding	Low
GO:0016772	Transferase activity	Low
Cellular Component Predictions		
GO:0005759	Mitochondrial matrix	High
GO:0005740	Mitochondrial envelope	High
GO:0005739	Mitochondrion	High
GO:0031966	Mitochondrial membrane	High
GO:0005743	Mitochondrial inner membrane	High
GO:0005737	Cytoplasm	Low
GO:0005634	Nucleus	Low

6.4 Discussion

A search through the DOMAINATION web platform, revealed 7 possible domain sequences that are closely related to XP_003529590 (Figure 6.1). The NP_193299.3 domain (gene ID:

At4g15640) possess the highest similarity score (Table 6.1), indicating close functional similarities to XP_003529590. Since sequence similarities are useful in inferring unknown gene function (Phuong and Nhung, 2013), the molecular functions of the XP_003529590 could be predicted from the closest ortholog At4g15640, which is annotated as protein binding adenylyl cyclase (AC). The heat map similarity search also bioinformatically indicated that the XP_003529590 gene is indeed a tetratricopeptide-like (TTP-l) helical AC enzyme, having similar domains to two TPR *Arabidopsis thaliana* AC genes; At4g15640 and At3g21465, (Figure 6.2) which are expressed in the mitochondrion and chloroplast respectively and exclusively in the guard cell. Arabidopsis mitochondrial genes such as At4g15640 have been implicated in DNA/RNA synthesis and processing and protein synthesis to produce defense/stress proteins; proteins involved in detoxification, signalling, transport and structural organisation (Heazlewood *et al*., 2004) while At3g21465 is reportedly involved in response to abiotic stimulus (GO:0009628), response to stress (GO:0006950) and response to biotic stimulus (GO:0045087) (Obulareddy *et al*., 2013).

The domain similarity search using DOMAINATION also revealed an AC from *Nicotiana tabacum*; AAB87670.1 (Figure 6.1), whose gene ID is *Axi 141 N. tabacum* as presented in Table 6.1. The AAB87670.1 was previously identified by Ito *et al*. (2014) as an ortholog to the *N. tabacum* ACR77530, that was expressed and confirmed as an AC with an intracellular signalling role during the tabtoxine-β-lactum cell death and development of necrotic wildfire disease (Ito *et al*., 2014). From this aspect on its own, it can therefore, be presumed that the XP_003529590 can also be putatively involved in intracellular signalling transduction systems during pathogen attack, which normally results in various diseases of the soybean plant, and also in RNA synthesis and processing to produce proteins involved in plant defense. Another *G. max* gene Glyma.17G023300 annotated as an AC was also identified in the similarity heat map (Figure 6.2), implying that this new putative gene closely resembles XP_003529590 in being a putative AC and in possessing a TPR-like domain. An inference of the HO4D005110 gene family used in the

construction of the similarity heat map further confirmed Arabidposis At4g15640 and At3g21465 genes as ACs (Table 6.2). This undoubtably shows that the probable functions of the XP_003529590 in soybean closely resembles those of the two Arabidopsis ACs, particularly the mitochondrial At4g15640 gene which has been reported to be largely involved in RNA synthesis, processing and protein synthesis (Heazlewood *et al*., 2004).

It was also noted that co-expressed genes that are highly correlated with XP_003529590 are expressed during development, of which all the top 25 genes are highly correlated with a Pearson correlation coefficient of 1 (Table 6.3). Therefore, most of the annotated functions for the gene are related to soybean plant growth and development. A *Zea mays* Zm0000b1022498 also possesses the TRP-like domains, thus conferring significant sequence similarity to the XP_003529590 (Figure 6.2; Table 6.2). This gene from *Z. mays* is a RING finger and CHY zinc finger domain containing protein. In soybean, there are two zinc finger development genes with a Pearson correlation co-efficient of 1 to the XP_003529590, thus perhaps suggesting a network of such genes in that plant. The genes are Glyma.06G074300, a CHY-type/CTCHY-type/ RING-type zinc finger protein and Glyma.19G207100, a B-Box type zinc finger protein with CCT domain (Table 6.3). The presence of zinc-finger proteins in plants presents a superfamily that is involved in many aspects of plant growth and development (Guo *et al*., 2009) and there are many zinc finger proteins that have been confirmed to be involved in abiotic and biotic stresses (Sakamoto *et al*., 2004; Oh *et al*., 2006; Ciftci-Yilmaz *et al*., 2007).

A research by Kim *et al*. (2001) has implicated a zinc protein (SCOF 1) in functioning as a positive regulator of cold-regulated gene expression mediated via the ABA response element in soybean. In *A. thaliana*, a CH type zinc finger has been reported to play a central role in reactive oxygen species (ROS) and abiotic stress signalling (Davletova *et al*., 2005) and in a transgenic *N. tabacum*, overexpression of a zinc finger protein gene *OsISP1* from *Oryza sativa* has been reported to confer tolerance to cold, dehydration and salt stress (Mukhopadhyay *et al*., 2004). This therefore implies

that the expression of the XP_003529590 during the development of the soybean plant can be related in promoting tolerance of the young seedlings to cold, dehydration and salt stresses. It has also been previously noted that expression the XP_003529590 gene is high in the early juvenile phase of the legume (Figure 2.2; Chapter 2) of which these young plants are generally very vulnerable to such stresses.

Glyma.03G025400 is another zinc finger CCCH-type/C3HC4-type RING finger family protein 1 with a Pearson correlation of 0.97 that is co-expressed with XP_003529590 as an anatomical gene (Genevestigator v3) (results not shown). The function of the RING finger and CHY zinc finger domain containing protein 1 has been described as mediating E3-dependant ubiquitination and proteasomal degradation of target proteins. Lorick *et al*. (1999) demonstrated that the RING finger containing proteins were involved in modulating protein levels via ubiquitination. Protein ubiquitination is a pathway that involves protein modification via proteasomal degradation, thereby playing a vital role in the control of multicellular organism development (Riechmann *et al*., 2000). Hence, the reason why Glyma.19G072800, a ubiquitin-conjugating enzyme 22 protein, is also highly co-expressed the with Glyma.07G251000 during development (Table 6.3). Ubiquitin modifications are reported to regulate physiological processes in plants such as cell-cycle progression, abscisic acid signalling, development and abiotic and biotic stress responses (Miura and Hasegawa, 2010).

A family of the heat shock proteins 70 (Hsp 70), Glyma.08G025700 is also co-expressed with the Glyma.07G251000 gene during plant development with a Pearson correlation coefficient of 1 (Table 6.3). Heat shock proteins (HSPs) are usually associated with plant responses to cold stress, heavy metals and ROS (Sun *et al*., 2002) amongst others. Recently the same group of proteins have also been found to be correlated with plant response to infection by pathogens such as nematodes (Kandoth *et al*., 2011) and fungi (Panthee *et al*., 2007). Specifically, the major functions of HSP70 proteins include preventing aggregation and assisting refolding of non-native

proteins under normal and stress conditions (Hartl, 1996; Frydman, 2001). They are also regulatory proteins that act as negative repressors of the heat-shock factor mediated transcription (Morimoto, 1998; Miernyk 1997). In Arabidopsis, *HSP70* genes are expressed in response to stress conditions such as heat, cold and drought (Lin *et al*., 2001). Overexpression of the *HPS70* genes also correlated positively with attainment of thermotolerance (Lee and Schöffl, 1996), enhanced tolerance to water, salt and high-temperature stress in plants (Alvin *et al*., 2001; Ono *et al*., 2001; Sung and Guy, 2003). It therefore can be stated that XP_003529590 is indeed a signalling molecule involved in impacting abiotic and biotic stress responses in soybean.

Plant auxins play very crucial roles during plant growth and development. In soybean, the Glyma.02G266300, an indole-3-butyric acid (IBA) response 5 protein is highly co-expressed with XP_003529590 with a Pearson's correlation coefficient of 1 during development (Table 6.3), thus implying the role of the XP_003529590 gene in the initiation of primary root development. IBA is a naturally occurring auxin together with indole-3-acetic acid (IAA), however, IBA has been reported to be more effective in root initiation than IAA since it is more stable under various light and temperature conditions (Nordstrom *et al*., 1991). In Arabidopsis, IBA is believed to provide auxin during the early seedling development, promoting lateral root proliferation (Zolman *et al*., 2001; Rampey *et al*., 2004). The XP_003529590 gene has been reported in Chapter 2 (Figure 2.2) to be highly expressed in the root apical meristem, root tip and root hair. Therefore, its co-expression with Glyma.02G266300 implies its important signalling role in root development.

The XP_003529590 gene has been described previously in Chapter 2 as a compound binding protein, and it follows that during the development of the soybean plant, some binding proteins are co-expressed with the XP_003529590 protein (Table 6.3). A Glyma.19G184300 ATP-binding cassette sub-family B1 and a Glyma.03G259800 Rho GTPase activating protein with PKA-box /P21-RHO-binding capacity (Pearson correlation coefficient = 1) are expressed during soybean plant development (Table 6.3). The ATP-binding cassette sub-family is a family of proteins that

are involved in mediating translocation of structurally unrelated molecules across biological membranes (Higgins, 1992). The ATP-binding cassette family are also reported to transport molecules involved in pathogen resistance in *N. plumbaginifolia* (Stukkens *et al*., 2005) and in the regulation of hypocotyl growth in Arabidopsis (Sidler *et al*., 1998). The Rho GTPase activating protein with PKA-box P21- RHO binding proteins are a family of small guanine nucleotide binding proteins that regulate extracellular stimulus-dependent signalling pathways that affect gene expression, cell proliferation, differentiation, actin reorganisation, cell cycle progression (Berken and Wittinghofer, 2008). Nevertheless, the function of the XP_003529590 can also be further postulated to be extracellular stimulus dependent signalling. Thus, this indicates that the expressed truncated C-terminal end of XP_003529590 or GmAC1 in Chapters 3 and 5 may be involved in a cAMP-PKA signalling pathway.

Glyma.11G175200 (cyclophilin 5 protein) with a Pearson correlation coefficient of 1 to the XP_003529590 is also co-expressed during soybean plant development (Table 6.3). The expression of cyclophilin has been shown to be induced by both the biotic and abiotic stress factors such as salt stress, heat and cold shock (Marivet *et al*., 1994), drought (Sharma and Singh, 2003), wounding and fungal infection (Godoy *et al*., 2000; Kong *et al*., 2001). These cyclophilins are ubiquitous proteins (Galat, 1999) that are located in the subcellular compartments and involved RNA processing (Krzywicka *et al*., 2001), protein maturation (Ferreira *et al*., 1996) and receptor signalling (Yurchenko *et al*., 2002). Mitochondrial cyclophilins have been implicated in accelerating the folding of newly imported proteins within the matrix as part of a complex that includes HPS60 and HSP70 (Rassow *et al*., 1995). This therefore, explains a co-expression of the HSP70 family protein and the ubiquitin conjugation protein during soybean development.

Plant development and growth can never be complete without expression of transferase proteins since it is well known that cellular processes produce toxins such as ROS, endogenous phytochemicals and exogenous toxins. Therefore, for cellular processes to function optimally,

there is need for detoxification and redox buffering. As a result, in the legume development, there are three highly correlated transferase genes that are co-expressed with the Glyma.07G251000 (Pearson correlation coefficient = 1) (Table 6.3). These include Glyma.14G074900 (S-adenosyl-L-methionine-dependent methyltransfareses superfamily), Glyma.08G175200 (Glutathione S-transferase TAU 19) and Glyma.14G065900 (Glucose-1-phosphate adenylyltransferase family protein). Glutathione S-transferases (GST) have been reported to provide crucial detoxification and redox buffering of cells (Noctor and Foyer, 1998), hence providing a protective role during development against oxidative damage. Jha *et al*. (2011) showed that the GST-TAU gene from a halophyte *Salicornia brachiate* in transgenic tobacco was up-regulated under cold, salt and drought stress and it led to enhanced seed germination and growth under salt stress. The S-adenosyl-L-methionine-dependent methyltransfareses are a superfamily of key enzymes in plant metabolic pathways. They are reported to play many roles in the biosynthesis of plant products that are related to disease resistance, growth and development of many plants (Joshi and Chiang 1998). Therefore, the XP_003529590 gene might be basically responsible for signalling the production of proteins that are involved in abiotic and biotic stress tolerance during the early growth stages of the legume as seen by the highly correlated co-expressed genes during development (Table 6.3).

Some ribosomal protein gene families are co-expressed during the legume development. This includes a 60s acidic ribosomal protein family (Glyma.04G229000) (Pearson correlation coefficient = 1) (Table 6.3). This 60s acidic ribosomal protein has been studied in seedling roots of *Z. mays* (Bailey-Serres *et al*., 1997), where they were believed to be involved in the selective translation of mRNA during flooding. It was proposed that when seedling roots of *Z. mays* are exposed to flooding stress, the flooding cause dynamic changes to the acidic phosphoprotein (P-protein) of the ribosome, thus affecting the ribosome associated kinases to phosphorylate P-proteins. In a more recent study by Yin *et al*. (2014) on analysis of changes in protein of soybean

root tip under flooding, it was concluded that flooding stress to soybean induces calcium-related signal transductions, which play important roles in early response to flooding. This is consistent with the earlier prediction that the XP_003529590 gene, is a sAC Ca^{2+}-binding mitochondrial cAMP generating PKA signalling molecule. In addition, this is also consistent with our findings in Chapter 5, where the truncated recombinant version of the XP_003529590 protein was shown *in vitro* to generate enhanced levels of cAMP (Figure 5.3), thus supporting its Ca^{2+} binding capability.

After an analysis of the various literature on the possible physiological and/or biological roles of the highly correlated co-expressed developmental genes (Pearson correlation coefficient = 1) (Table 6.3), it therefore become almost certain to presume that the XP_003529590 gene is largely involved in the growth and development of the legume seedling through impacting tolerance to a number of stress factors. However, the most prominent stresses being cold, salt, drought and flooding.

In response to different perturbations, most ribosomal proteins were observed to be co-expressed with the XP_003529590 (Table 6.4). As previously described, AC are enzymes responsible for the formation of the signalling molecule cAMP, and cAMP is a regulator of various genes in response to different stimuli (Kuo *et al*., 2003). Therefore, genes responsible for the synthesis of effector proteins are transcribed since soybean is a stress sensitive plant (Oh and Komustu, 2015). Soybean plants inevitably interact with climatic factors and therefore, are subjected to diverse abiotic and biotic stresses. They therefore, induce defense mechanism to combat the effects to these environmental stresses. This is mainly through expression of genes that encode for proteins involved in defense or tolerance mechanisms. According to our findings in this chapter, those are the genes that are co-expressed with XP_003529590 in response to various perturbations/stimuli (Table 6.4). Apparently, most of these genes have been shown, as already been discussed above, to be expressed during the juvenile phase of the soybean legume.

It is also notable that most genes co-expressed with the XP_003529590 in response to perturbations were generally ribosomal proteins with Pearson correlation coefficient values ranging between ≤ 0.88 and ≥ 0.83 (Table 6.4). Ribosomal proteins are primarily involved in the translation of mRNA resulting in protein synthesis. Their genes occur as multiple expressed members, which can be incorporated in the cytosolic ribosome under specific conditions such as certain developmental stages, tissues and stress conditions (Schmid *et al*., 2005). In Z. *mays*, expression of a number of ribosomal proteins was found to be up-regulated by ultraviolet-B (UV-B) stress in plants that were exposed to different light regimes (Casati and Walbot, 2003). Work by Falcone-Ferreyra *et al*. (2010) also indicated that ribosomal L10 proteins participate in development and translation during the UV-B stress exposure. The translated proteins include cytoplasmic ribosomal proteins, initiation and elongation factors, and poly(A)-binding proteins (Casati and Walbot, 2003). Liu *et al*. (2014) also provided evidence to the role of ribosomal proteins to stress tolerance. They showed that the 60S ribosomal proteins are associated with tolerance to stress. The ribosomal protein AgRPL44 from a halophilic fungus, *Aspergillus glaucus*, in transgenic tobacco, caused an up-regulation of a 60S sub-unit protein during salt, drought and heavy metals stresses, and thereby moonlighting the role of ribosomal proteins in abiotic stress tolerance. Therefore, the XP_003529590 gene could largely be involved in signalling transduction pathways involving ribosomal proteins that are crucial in promoting stress tolerance and development.

The top 7 co-expressed genes (Table 6.4) were discussed here as they form the basis for gene differential expression using microarray data from the Gevestigator v3 [Soybean] Affymetrix Soybean Genome Array platform. These include Glyma.07G184300, which is a mitochondrial HSO70 2 protein labelled as a flooding responsive (Yin *et al*., 2014) and mitochondrial heat shock protein; Glyma.19G242700 and Glyma.19G242700, which are both ribosomal S24e family proteins (Table 6.4). In a study by Oh and Komatsu, (2015), drought stress to soybean seedling

was shown to increase ribosomal proteins belonging to the S family. These proteins primarily function in protein synthesis. Glyma.03G150600, a GTP-binding elongation factor Tu family protein is also co-expressed with the XP_003529590 gene and the protein family is involved in protein synthesis, however, Yin *et al*. (2014) discovered that the protein family is basically flooding responsive. There are also two ribosomal proteins of the L family (Glyma.08G184300), a ribosomal protein L10 family protein and ribosomal protein L3 plastid. The L family ribosomal proteins are also among the flooding responsive proteins produced in the seedling of the soybean plant (Yin *et al*., 2014). Lastly but not least, are Glyma.10G154500, a rRNA processing protein-related and an uncharacterised protein (Glyma.05G029300). Therefore, the XP_003529590 gene is likely to be involved in imparting flooding tolerance to soybean in its juvenile stages due to co-expression of the L and S family ribosomal proteins together with the HSP protein family.

Results from the differential expression using microarray data showed variable regulation of the XP_003529590 to pathogen attack (Figure 6.3) *P*. *sojae* and *P*. *panchyrhizi*, the two major pathogens of the soybean plant and major causes of crop losses world over. In this regard XP_003529590 was shown to be downregulated upon infection by the *P*. *sojae* pathogen, however, after 48 hours of infection, the gene was up-regulated. The trend was the same for all the other co-expressed genes (Figure 6.3 (i)). *P*. *sojae* pathogen produces free swimming zoospores that are normally attracted to isoflavonoids secreted by the soybean roots (Morris *et al*., 1998), therefore, the noted down-regulation of genes is probably because once the zoospores attach and encyst onto the soybean hypocotyl, the *P*. *sojae* is able to colonise the legume host cells in an initial biotrophic phase of growth, which lasts approximately 12 hours (Ward, 1990). During this phase, *P*. *sojae* RNA increases, accounting for larger quantities (70%) of the large subunit rRNA by 48 hours of infection (Moy *et al*., 2004) and within the same time period, most soybean transcripts would be down-regulated. However, an effective resistance response by the soybean plant occurs within the early hours of attachment is ensured, thereby arresting further pathogen growth (Moy *et al*., 2004),

hence, the observed up-regulation of the soybean genes, including XP_003529590 after 48 hours of infection. XP_003529590 being an enzyme that catalyses the formation of cAMP, which is a signalling molecule, hence assists in alleviating further damage by the pathogen. Most of the co-expressed genes are ribosomal proteins, which are important in the synthesis of proteins and in this case, mostly those related to resistance response. It was observed by Moi *et al*. (2004) that most of the host (soybean) upregulated genes were of defense and signalling proteins and according to Tyler, (2007), soybean genes encoding enzymes of phytolexin biosynthesis and defense and pathogenesis related proteins were strongly upregulated and peaked 24 hours after infection.

Microarray analysis of soybean plant infection by the pathogen *P*. *pachyrhizi* showed an initial up-regulation of the XP_003529590 gene and all the other 7 co-expressed genes with a peak up-regulation at 48 hours after infection (Figure 6.3 (ii)). However, after 144 hours of exposure, the genes were down-regulated. The initial up-regulation of the genes is an attempt of the soybean plant to effect resistance to the rust disease caused by the pathogen. Since most of these co-expressed genes are ribosomal proteins implicated in protein synthesis, it can be hypothesised that the synthesised proteins were involved stress tolerance, plant development, synthesis of secondary metabolites, metabolic pathways, membrane transport and signal transduction (a key function of cAMP) (Panthee *et al*., 2007). However, the down-regulation of the genes after longer periods of exposure is an indication that soybean plants are not able to mount a successful defense against the *P*. *pachyrhizi* pathogen (Panthee *et al*., 2007). The mitochondrial HSO70 (heat shock protein)'s upregulation among the ribosomal proteins plays important roles in stress tolerance and has been reported to be a cell defense-related protein (Patel *et al*., 2004; Luo *et al*., 2005). These results indicate that soybean has a low innate response to *P*. *pachyrhizi*, therefore, there is need for further research to develop transgenic soybean using genes that are tested and shown to confer resistance to *P*. *pachyrhizi*.

In a mimicry microarray analysis of herbivory involving bud removal, there was an up-regulation of the XP_003529590 gene and all other co-expressed soybean genes tested (Figure 6.3 (iii)). The XP_003529590 gene responsible for the synthesis of cAMP signalling molecule was up-regulated, implying its role in conferring resistance to herbivory attack. The other up-regulated protein synthesis ribosomal genes could trigger the synthesis and accumulation of secondary metabolites while the mitochondrial HSO70 enhances oxidative burst. The mitochondria, among other sources such as the chloroplast and peroxisomes, are organs of oxidative burst and have been implicated in the production of ROS in damaged tissues (Gatehouse, 2002, Kessler and Baldwin, 2002). In soybean, a herbivory attack by the insect *Helicoverpa tea* was reported to have induced a shift in the oxidative status of the legume, thus increasing the levels of O_2^- and OH radical formation (Bi and Felton, 1995).

B. japonicum is gram-negative soil bacterium that is capable of establishing a nitrogen-fixing symbiotic relationship with its soybean plant host. The microarray analysis of the soybean root to *B. japonicum* showed an initial up-regulation of the XP_003529590 gene and the co-expressed genes (Figure 6.4 (i)). However, at 48-hour period, the soybean genes were down-regulated. Since XP_003529590 is an AC protein, its initial up-regulation promotes synthesis of cAMP - results that are consistent with findings by Terakado *et al*. (1997) in which they found around 7 pmol g^{-1} f.wt of cAMP in root nodules of soybean inoculated with *B. japonicum*. cAMP is an important signalling molecule, hence probably signalling the production of defense response transcripts, transcription factor genes, hormone related genes (especially during the early stage of soybean inoculation) and stress response-related genes (Colebatch *et al*., 2004; Kouchi *et al*., 2004; Brechenmacher *et al*., 2008). Among the compounds reported to be produced during the *B. japonicum* root treatment are flavonoid compounds that are secreted by the soybean roots and production of these compounds is because of the up-regulation of genes involved in protein synthesis. These flavonoids are secondary metabolites that are also synthesised by most of the

ribosomal proteins. Flavonoid compounds production results in the induction of the nodulation (*nod*) genes (Subramanian *et al*., 2006).

Previous research has provided insight into the role of isoflavonoids in soybean. Isoflavonoids are a group of flavonoid compounds, that were shown to be the chief inducers of *B. japonicum nod* gene expression (Banfalvi *et al*., 1988; Loh *et al*., 1994) and evidence through biochemical analysis of roots, has shown that isoflavone levels increased in response to *B. japonicum* inoculation (Cho and Harper, 1991). In addition, a study by Libault *et al*. (2010) predicted that the up-regulated genes may be involved in plant defense, modification of the cell wall composition, signal transduction and basic metabolic processes. After 48 hours of *B. japonicum* inoculation, the XP_003529590 gene, together with the co-expressed genes (Figure 6.4 (i)) were down-regulated. This observation of down-regulation of genes after their initial up-regulation is consistent with report made by Libault *et al*. (2010) in root hair cell transcriptome analysis using Illumina transcriptome sequencing after *B. japonicum* inoculation. The possible reason for down-regulation could be that the *B. japonicum* has the ability to suppress initial plant defenses that could be detrimental to the infection process.

Adequate or insufficiency iron to the soybean plant resulted in the up-regulation of the XP_003529590 gene and all other co-expressed genes (Figure 6.4 (ii)). In soybean, iron is a critical macronutrient that serves as a co-factor for a wide range of cellular processes (O'Rourke *et al*., 2009). Limited iron supply to soybean leads to iron deficiency termed chlorosis, which is characterised by interveinal chlorosis of the developing trifoliates (Inskeep and Bloom, 1987). Iron is also an important element in the nodulation process (Abdelmajid and Chedly, 2003). Therefore, it is vital that stress related genes are co-expressed and co-translated such that the legume is not adversely impaired by the limited iron supply.

The GO predictions (Table 6.5) have provided key evidence to support the prescribed functions of the recombinantly expressed and functionally characterised C-terminal part of the XP_003529590 gene or GmAC1 in the other chapters of this whole study. The molecular predictions have demonstrated that the XP_003529590 gene is indeed involved in AC catalytic activities, kinase activities and binding of the following molecular components, RNA, ATP, Fe^{2+}, kinase and cations. The biological prediction also supported the assertion that the XP_003529590 gene is involved in regulation of gene expression, RNA processing, signalling, response to stimulus and anatomical structure development.

6.5 Conclusion

The bioinformatic analyses to elucidate the physiological roles of the novel AC protein coding gene, XP_003529590 has revealed that the gene is indeed an AC that is largely located in the mitochondrion of the soybean plant and expressed during the legume's juvenile stages of plant growth and development. Primarily, the gene has been shown to be involved in abiotic stress responses particularly drought, cold, salt and flooding and biotic stress factors such as herbivory and attack from the pathogen *P. sojae*. More so, the XP_003529590 gene is also primarily involved in the essential processes of nodule formation and establishment.

CHAPTER 7

GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

The major goal of this research was to identify and characterise the first ever adenylyl cyclase (AC) molecule in *Glycine max*, accession number: XP_003529590 and Gene ID: Glyma.07G251000, whose truncated version harbouring the AC catalytic motif (GmAC1) was herein cloned, expressed and functionally characterised. *G. max* commonly known as soybean, is a crop of great economic importance and sadly enough, it is also a drought sensitive crop plant. Though a number of transcriptome analyses of the legume responses to drought have been reported (Chen *et al*., 2013; Le *et al*., 2012), information on the role of cAMP and its involvement in conferring drought/stress tolerance in this plant is almost totally lacking and/or nearly non-existent, particularly on the mitochondrial soluble AC- (sAC)-generated cyclic AMP. GmAC1 is one of the expressed C-terminal sAC protein coding genes in *G. max* that was studied in this research. Though several ACs have been confirmed in *Arabidopsis thaliana*, there is still a lot of work that needs to be done in expressing and confirming ACs in crops of economic importance that aim at improving food security and alleviating hunger and poverty. As such, this study has provided the first ever solid evidence of an expressed and functionally confirmed sAC in soybean and in form of the XP_003529590 protein. This XP_003529590 protein is the one that was once reported previously as a putative AC candidate (Ito *et al*., 2014). Notably and besides this XP_003529590, two more soybean putative AC candidates have this far been noticed. The first one, with a Gene ID: Glyma.17G023300, was independently picked up in this study and during our bioinformatic analyses (Chapter 6) while the other one, XP_003547191 (Gene ID:

Glyma.15G090100), was previously reported by Świeżawska *et al*. (2014). Specifically, the Glyma.17G023300 gene featured during our perturbations co-expression analysis of the XP_003529590 using Genevestigator v3 (Table 6.4; Chapter 6) and particularly, during our functional clustering analysis of the gene family HO4D005110 InterPro-tetratricopeptide-like helical domain using the similarity heat map (Figure 6.2; Chapter 6), implying its closest relationship with the novel XP_003529590 gene. Otherwise, these two other additional putative ACs in soybean still need to be closely scrutinised, possibly through cloning and recombinant characterisation.

Presumably, the emergence of these novel AC genes in soybean could be as a result of the reported 2.55 genome duplication process (Turner *et al*., 2012) that ultimately, has led to novel genes in this leguminous plant (Lynch and Conery, 2000). Additionally, the XP_003529590 gene also has an isoform in the form of Glyma07G38080 and together, these two genes are reported to occur on the same chromosome (chromosome 7) with the following transcripts KRH50918 and KRH50917 respectively (Figure 2.1; Chapter 2). However, no data exists as to confirm whether Glyma07G38080 is also a putative AC or not, nevertheless, it can be assumed that the genome duplication of Glyma07G38080 possessing 10 exons, has probably led to a novel gene with new functions - the XP_003529590 with 11 exons. Genome duplication is reported to have contributed to the evolution of novel functions such as induction of disease resistance and adaptation to stress (Panchy *et al*., 2016). The XP_003529590 functions have been predicted to be highly involved in the induction of disease resistance to the common pathogens *Phytophthora sojae* and *Phakopsora pachyrhizi*, defense response to herbivory and adaptation to flooding, drought, cold and salt stress.

Thus in an attempt to try and check if the XP_003529590 gene is indeed an AC molecule, we first cloned its truncated C-terminus end harbouring the annotated catalytic AC centre (GmAC1) and tested for its endogenous activity (Chapter 3) and we noted its ability to produce significant amounts of cAMP in competent *E. coli* BL21 (DE3) pLysS expression cells. After noticing the

endogenous AC activity of this truncated version of XP_003529590, we then proceeded to test for its AC activity in an *in vivo* system through complementation testing and confirmed such activity (Chapter 4). The confirmed AC activity was then characterised *in vitro* using a purified version of the recombinant protein, where it was firmly confirmed that the GmAC1 is indeed a sAC, whose activity is primarily enhanced by presence of the HCO_3^-, Mn^{2+} and Ca^{2+} molecular ions (Chapter 5). This then therefore, confirmed the existence of a functional AC molecule in soybean that is able to catalyse the formation of cAMP from ATP - a first ever outcome of such kind in this leguminous plant.

The fact that the GmAC1 is activated *in vitro* by Ca^{2+} and HCO_3^- suggests that the XP_003529590 protein is an intracellular metabolite and signal intergrator within the legume's mitochondria, where it is probably involved in mitochondrial metabolism (Acin-Perez *et al*., 2009; 2011) and regulation of pathogens (Topal *et al*., 2012). cAMP is now an acceptable key signalling molecule in plants (Gehring and Turek, 2017) and its presence in soybean has been confirmed. Interestingly, we believe, the mitochondrial generated cAMP effects its signalling roles in soybean through a protein kinase A or a PKA-like kinase. This hypothesis can be thought to be true as the GmAC1 protein was shown to be activated by Ca^{2+} to produce cAMP, which is known for modulating sACs in order to stimulate protein kinase activity (Hetherington and Trewavas, 1984; Hepler and Wayne, 1985). The GO molecular function predictions for the XP_003529590 gene have shown that the gene is involved in kinase binding (GO:0019900) and kinase activity (GO:0016301). Therefore, this implies that the soybean mitochondrial cAMP generated by the XP_003529590 protein may follow a PKA or PKA-like kinase pathway. Protein kinases in plants play major roles in the regulation of abiotic stress factors such as cold, salt and drought and biotic stress that involve pathogens and herbivores (Ho, 2015).

The identified *G. max* sAC has been noted bioinformatically to be able to generate cAMP through its binding ability, and through *in vitro* assaying, Ca^{2+}, Mn^{2+} and HCO_3^- were reportedly shown to

be able to bind to this sAC thus generate mitochondrial cAMP and facilitating its signalling through kinase binding and/or kinase activity. The secondary structure of the XP_003529590 gene has indicated that the gene is largely an alpha helical structure that comprises of disordered regions involved in protein binding (amino acids 1-10; 375-377) and disordered regions (amino acids 11-68; 372-374) as in Figure 2.4, Chapter 2. As previously discussed, the presence of long disordered regions on the structure of the XP_003529590 protein are therefore, related to regulation via transcription and translation and the alpha-helical regions being dominated by enzymatic catalysis (Tompa, 2012), thus involved in catalysing the cyclisation of ATP to cAMP, while the disordered regions are involved in binding, thus conferring functional advantage of the protein to stress responses, signalling and regulation (Oldfield *et al*., 2008; Sun *et al*., 2013). Through the further bioinformatic functional analysis of the novel XP_003529590 gene, it has then been established that the gene is primarily involved in stress responses during the legume development, thus involved in regulation of the transcription and translation of proteins involved in such critical processes. It was also noted that a lot of ribosomal proteins are co-expressed with the XP_003529590 gene during perturbations (Table 6.4; Chapter 6).

The XP_003529590 gene was shown to be expressed during early soybean plant development and is largely involved in the subsequent transcript expression of stress defense related genes, particularly drought induced conditions through salt stress, flooding and heavy metals. Usually, plants in their juvenile stages are prone to abiotic stress factors thus hindering further growth and development and ultimately, leading to major crop losses. Bioinformatic studies have indeed provided valuable insights into the fundamental roles of this novel sAC gene in soybean plants. Microarrays expression profiling of the GmAC1 from [Soybean] Affymetrix Soybean Genome Array platform in Genevestigator v3 provided additional support to the role of cAMP during soybean pathogen defense. It has been noted that *P*. *pachyrhizi* and *P*. *sojae* are the major soybean pathogens that threaten production and the overall yield of this legume. In connection with this,

the XP_003529590 gene expression was shown to be up-regulated during infection by *P. sojae*, implying a defense related signalling mechanism. However, exposure of the soybean plant to *P. pachyrhizi*, the XP_003529590 was shown to be downregulated with increased exposure to the pathogen (Figure 6.3 (ii); Chapter 6). This implies that the gene is not very effective in conferring resistance to the pathogen. cAMP production in soybean through the presumable activity of the XP_003529590 protein has also been shown to regulate plant defense against herbivory attack as evidenced by its up-regulated expression during bud removal (Figure 6.3 (iii); Chapter 6). It can therefore, be concluded that the novel XP_003529590 gene in *G. max* is involved in conferring biotic and abiotic stress tolerance in the soybean plant and in increasing nodule formation for increased grain yield.

Apart from playing crucial role in improving food security *G. max* enhances soil fertility through its symbiotic nitrogen fixing ability. Gene expression study using array data has provided insights into the expression of the XP_003529590 during root inoculation with *Bradyrhizobium japonicum* (Figure 6.4 (i) Chapter 6). The gene expression was shown to be up-regulated, implying the importance of cAMP signalling in nodule formation in soybean. Studies have shown that inoculation of soybean seeds with *B. japonicum* prior to sowing significantly enhances the legume's nitrogenase activity and nutrition (Nimnoi *et al*., 2014). The legume has a high nitrogen demand (Kaschuk *et al*., 2016) particularly at the onset of pod filling (Hungria and Mendes, 2015). This high nitrogen demand also explains the upregulation of the XP_003529590 to sufficient iron supply and iron deficiency to soybean (Figure 6.4 (ii); Chapter 6). Iron is an essential component of leghaemoglobin which is important during the soybean nodulation process as it transports oxygen within the nitrogen-fixing cells in the soybean nodules (Abdelmajid and Chedly, 2003). Therefore, the production of cAMP plays vital roles in the biological events that regulate nodule formation and function in *G. max*, hence the noted up-regulated expression of the novel XP_003529590 gene during *B. japonicum* root inoculation and iron supply.

A more interesting discovery is that the XP_003529590 is actually a pentratricopeptide protein, confirming the localisation of the gene as mitochondrial. PPRs are reportedly involved in RNA binding, which ultimately leads to RNA editing, and RNA editing itself is essential for facilitating the expression of functional proteins (Okuda *et al*., 2009), thus facilitating the functional roles of the novel XP_003529590 protein. More, a PPR in the form of an Arabidopsis PPR (AtPPR-AC) has previously been confirmed as a functional AC molecule in plants (Ruzvidzo *et al*., 2013) and thus closely concurring with our overall finding herein for the XP_003529590 protein.

Biotechnology is a branch of science that has provided answers to the improvement of yield in agronomically important crop plants through horizontal transfer of genes that can help in increasing production and yield. As such, the XP_003529590 gene from soybean can perhaps be transformed into other crops, particularly at their seedling stages to help them develop into mature plants since it is an early development gene. Such practices have been done in the production of transgenic tobacco as discussed earlier on in section 3.1 of Chapter 3. An overexpression strategy of this gene in soybean can also be practiced to enhance the production of its own inherent AC gene (s), thus increasing cAMP levels and subsequently conferring the desired tolerance/resistance output. This can lead to much more improved soybean grain production. The fact that soybeans are high protein seeds, it therefore means that this essential characteristic can be strategically exploited in using the legume as an expression system (Hudson *et al*., 2014).

REFERENCES

Abdelmajid K, Chedly A. 2003. Importance of iron use-efficiency of nodules in common bean (*Phaseolus vulgaris* L.) for iron deficiency chlorosis resistance. *Journal of Plant Nutrition and Soil Science* **166:** 525-528.

Acin-Perez R, Russwurm M, Gunnewig K, Gertz M, Zoidl G, Ramos L, Buck J, Levin LR, Rassow J, Manfredi G, Steegborn C. 2011. A phosphodiesterase 2A isoform localised to mitochondria regulates respiration. *Journal of Biological Chemistry* **286:** 30423–30432.

Acin-Perez R, Salazar M, Kamenetsky E, Buck J, Levin LR, Manfredi G. 2009. Cyclic AMP produced inside mitochondria regulates oxidative phosphorylation. *Cell Metabolism* **9:** 265–276.

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

Alvin FC, Carolina SM, Cascardo JC, Nunes CC, Martinez CA, Otoni WC, Fontes EP. 2001. Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress. *Plant Physiology* **126:** 1042–1054.

Al-Younis I, Wong A, Lemtiri 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. *Frontiers in Plant Science* **9:** 1645.

Al-Younis I, Wong A, Gehring C. 2015. *Arabidopsis thaliana* K^+-uptake permease 7 (AtKUP7) contains a functional cytosolic adenylate cyclase catalytic centre. *FEBS Letters* **589:** 3848-3852.

Attwood TK. 2000. The quest to deduce protein function from sequence: the role of pattern databases. *International Journal of Biochemistry Cell Biology* **32:** 139–155.

Bailey-Serres J. Vangala S, Szick K, Lee CH. 1997. Acidic phosphorylation complex of the 60S ribosomal subunit of maize seedling root. *Plant Physiology* **114:** 1293-1305.

Baneyx F. 1999. Recombinant protein expression in *Escherichia coli. Current Opinions in Biotechnology* **10:** 411-421.

Baneyx F. Mujacic M. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology* **22:** 1399-1408.

Banfalvi Z, Nieuwkoop A, Schell M, Besl L, Stacey G. 1988. Regulation of nod gene expression in *Bradyrhizobium japonicum*. *Molecular General Genetics* **214:** 420–424.

Bellalou J, Sakamoto H, Ladant D, Geofioy C, Ullmann A. 1990. Synthesis and secretion of *Bordetella pertussis* adenylate cyclase as a 200-kilodalton protein. *Infection and Immunity* **58:** 1195-1200.

Benz R, Maier E, Ladant D, Ullmann A, Sebo P. 1994. Adenylate cyclase toxin (*CyaA*) of *Bordetella pertussis*. *The Journal of Biological Chemistry* **269:** 27231-27239.

Berken A, Wittinghofer A. 2008. Structure and function of RHO-type molecular switches in plants. *Plant Physiology and Biochemistry* **46:** 380-393.

Beuve A, Boesten B, Crasnier M, Danchin A, O'Gara F. 1990. *Rhizobium meliloti* adenylate cyclase is related to eukaryotic adenylate and guanylate cyclases. *Journal of Bacteriology* **172:** 2614-2621.

Bi JL, Felton GW. 1995. Foliar oxidative stress and insect herbivory: primary compounds, secondary metabolites, reactive oxygen species as components of induced resistance. *Journal of Chemical Ecology* **21:** 1511–1530.

Bianchet C, Wong A, Quaglia M, Alqurashi M, Gehring C, Ntoukakis V, Pasqualini S. 2019. An *Arabidopsis thaliana* leucine-rich repeat harbours an adenylyl cyclase catalytic centre and affects responses to pathogens. *Journal of Plant Physiology* **232:** 12-22.

Bigay J, Deterre P, Pfister C, Chabre M. 1987. Fluoride complexes of aluminium or beryllium act on G-proteins as reversibly bound analogues of the gamma phosphate GTP. *EMBO Journal* **6:** 2907-2913.

Bindschedler LV, Minibayeva F, Gardner SL, Gerrish C, Davies DR, Bolwell GP. 2001. Early signalling events in the apoplastic oxidative burst in suspension cultured French bean cells involve cAMP and Ca^{2+}. *New Phytologist* **151:** 185–194.

Blume B, Nürnberger T, Nass N, Scheel D. 2000. Receptor-mediated increase in cytoplasmic free calcium required for activation of pathogen defense in parsley. *The Plant Cell* **12:** 1425–1440.

Bolwell GP. 1992. A role for phosphorylation in the down-regulation of phenylalanine ammonia-lyase in suspension-cultured cells of French bean. *Phytochemistry* **31:** 4081–4086.

Botsford JL. 1981. Cyclic nucleotides in prokaryotes. *Microbiological Reviews* **45:** 620-642.

Braun T. 1974. Evidence for multiple, cell specific distinctive adenylate cyclase system in rat testis. *Current Topics in Molecular Endocrinology* **1:** 243-264.

Braun T. 1975. The effect of divalent cations on bovine spermatozoal adenylate cyclase activity. *Journal of Cyclic Nucleotide Research* **1:** 271–281.

Braun T, Dods RF. 1975. Development of a Mn^{2+} sensitive, "soluble" adenylate cyclase in rat testis. *Proceedings of the National Academy Science USA* **72:** 1097–1101.

Brechenmacher L, Kim MY, Benitez M, Li M, Joshi T, Calla B, Lee MP, Libault M, Vodkin LO, Xu D, Lee SH, Clough SJ, Stacey G. 2008. Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. *Molecular Plant-Microbe Interactions* **21:** 631–645.

Brevedan RE, Egli DB. 2003. Short periods of water stress during seed filling, leaf senescence and yield of soybean. *Crop Science*. **43:** 2083–2088.

Briand L, Marcion G, Kriznik A, Heydel JM, Artur Y, Garrido C, Seigneuric R, Neiers F. 2016. A self-inducible heterologous protein expression system in *Escherichia coli*. *Scientific reports* **6:** 33037.

Brickman E, Soll L, Beckwith J. 1973. Genetic characterisation of mutations which affect catabolite-sensitive operons in *Escherichia coli*, including deletions of gene for adenyl cyclase. *Journal of Bacteriology* **116:** 582-587.

Bridges D, Fraser ME, Moorhead GB. 2005. Cyclic nucleotide binding proteins in the *Arabidopsis thaliana* and *Oryza sativa* genomes. *BMC Bioinformatics* **6:** 6.

Brownstein MJ, Carpten JD, Smith JR. 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20:** 1004-1010.

Cai W, Okuda K, Peng L, Shikanai T. 2011. Proton gradient regulation 3 recognises multiple targets with limited similarity and mediates translation and RNA stabilisation in plastids. *Plant Journal* **67:** 318–327.

Campos MA, Silva MS, Magalhaes CP, Ribeiro SG, Sarto RPD, Vieira EA, Grosii de Sa MF. 2008. Expression in *Escherichia coli*, purification, refolding and antifungal activity of an osmotin from *Solanum nigrum*. *Microbial Cell Fractions* **7:** 7.

Carricarte VC, Bianchini GM, Muschietti JP, Tellez-Inon MT, Perticari A, Torres N, Flawia M. 1988. Adenylate cyclase activity in a higher plant, alfalfa (*Medicago sativa*). *Biochemical Journal* **249:** 807–811.

Carter SL, Brechbuhler CM, Griffin M, Bond AT. 2004. Gene co-expression network topology provides a framework for molecular characterisation of cellular state. *Bioinformatics* **20:** 2242-2250.

Casari G, Sander C, Valencia A. 1995. A method to predict functional residues in proteins. *Nature Structural Biology* **2:** 171-178.

Casati P, Walbot V. 2003. Gene expression profiling in response to ultraviolet radiation in *Zea mays* genotypes with varying flavonoid content. *Plant Physiology* **132:** 1739–1754.

Catanese CA, Emerich DW, Zahler WL. 1989. Adenylate cyclase and cyclic AMP phosphodiesterase in *Bradyrhizobium japonicum* bacteroides. *Journal of Bacteriology*. **171:** 4531-4531.

Chai WY, Joshi T, Valliyodan B, Prince S, Michel L, Xu D, Ngayen HT. 2015. Soybean transcription factor ORFeome associated with drought resistance: a valuable resource to accelerate research on abiotic stress resistance. *BMC Genomics* **16:** 596.

Chatukuta P, Dikobe TB, Kawadza DT, Sehlabane KS, Takundwa MM, 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.

Chen C, Huang H, Wu CH. 2017. Protein bioinformatics databases and resources. *Methods in Molecular Biology* **1558:** 3-39.

Chen Y, Cann MJ, Litvin TN, Iourgenko V, Sinclair ML, Levin LR, Buck J. 2000. Soluble adenylyl cyclase as an evolutionarily conserved bicarbonate sensor. *Science* **289:** 625-628.

Chen LM, Zhou XA, Li WB, Chang W, Zhou R, Wang C, Sha AH, Zhang CJ, Qiu DZ, Yang ZL, Chen SL. 2013. Genome-wide transcriptional analysis of two soybean genotypes under dehydration and rehydration conditions. *BMC Genomics* **14:** 687.

Cho MJ, Harper JE. 1991. Root isoflavonoid response to grafting between wild-type and nodulation-mutant soybean plants. *Plant Physiology* **96:** 1277–1282.

Choi YE, Xu JR. 2010. The cAMP signalling pathway in *Fusarium verticillioides* is important for conidiation, plant infection, and stress responses but not fumonisin production. *Molecular Plant-Microbe Interactions* **23:** 522-533.

Ciftci-Yilmaz S, Morsy MR, Song L, Coutu A, Krizek BA, Lewis MW, Warren D, Cushman J, Connolly EL, Mittler R. 2007. The EAR-motif of the Cys2/His2-type zinc finger protein Zat7 plays a key role in the defense response of Arabidopsis to salinity stress. *Journal of Biological Chemistry* **282:** 9260–9268.

Clark EDB. 1998. Refolding of recombinant proteins. *Current Opinion in Biotechnology* **9:** 157-163.

Clare A, Karwath H, Ougham R, King D. 2006. Functional bioinformatics for *Arabidopsis thaliana Bioinformatics* **22:** 9 1130–1136.

Clare A, King, RD. 2003. Predicting gene function in *Saccharomyces cerevisiae*. *Bioinformatics* **19:** 1142–1149.

Clough SJ, Fengler KA, YuI C, Lippok B, Smith RKJ, Bent AF. 2000. The Arabidopsis *dnd1* 'defense, no death' gene encodes a mutated cyclic nucleotide-gated ion channel. *Proceedings of National Academy of Sciences, USA* **97:** 9323–9328.

Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J, Udvardi MK. 2004. Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant Journal* **39:** 487–512.

Confer DL, Eaton JW. 1982. Phagocyte impotence caused by invasive bacterial adenylate cyclase. *Science* **217:** 948-950.

Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF. 2014. Distinct copy number, coding sequence, and locus methylation patterns underlie Rhg1 mediated soybean resistance to soybean cyst nematode. *Plant Physiology* 1**65:** 630–647.

Cooke CJ, Smith CJ, Walton TJ, Newton RP. 1994. Evidence that cyclic AMP is involved in the hypersensitive response of *Medicago sativa* to a fungal elicitor. *Phytochemistry* **35:** 889–895.

Cooper DMF. 2005. Compartmentalisation of adenylyl cyclase and cAMP signalling. *Biochemical Society Transactions* **33:** 1399–1322.

Cooper DM, Mon N, Karpen JW. 1995. Adenylyl cyclases and the interaction between calcium and cAMP signalling. *Nature* **374:** 421-424.

Cotta MA, Whitehead TR, Wheeler MB. 1998. Identification of a novel adenylate cyclase in the ruminal anaerobe, *Prevotella ruminicola* D31d. *FEMS Microbiology Letters* **164:** 257-260.

Crowl R, Seamans C, Lomedico P, McAndrew S. 1985. Versatile expression vectors for high-level synthesis of cloned gene products in *Escherichia coli*. Gene **38:** 31-38.

Davanloo P, Rosenberg AH, Dunn JJ. Studier FW. 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Science USA* **81:** 2035–2039.

Davletova S, Schlauch K, Coutu J, Mittler R. 2005. The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signalling in Arabidopsis. *Plant Physiology* **139:** 847–856.

deCrombrugghe B, Chen B, Gottesman M, Pastan I, Varmus HE, Emmer M, Perlman RL. 1971. Regulation of lac mRNA synthesis in a soluble cell free system. *Nature Cell Biology* **230:** 37-40.

Deshmukh RK, Vivancos J, Guérin V, Sonah H, Labbé C, Belzile F, Belanger RR. 2013. Identification and functional characterisation of silicon transporters in soybean using comparative genomics of major intrinsic proteins in Arabidopsis and rice. *Plant Molecular Biology* **83:** 303–315.

Deshmukh R, Sonah H, Patil G, Chen W, Prince S, Mutava R, Vuong T, Valliyodan B, Nguyen HT. 2014. Integrating omic approaches for abiotic stress tolerance in soybean. *Frontiers in Plant Science* **5:** 244.

Di Benedetto G, Scalzotto E, Mongillo M, Pozzan T. 2013. Mitochondrial Ca^{2+} uptake induces cyclic AMP generation in the matrix and modulates organelle ATP levels. *Cell metabolism* **17:** 965–975.

Dong S, Qutob D, Tedman-Jones J, Kuflu K, Wang Y, Tyler BM, Gijzen M. 2009. The *Phytophthora sojae* a virulence locus Avr3c encodes a multi-copy RXLR effector with sequence polymorphisms among pathogen strains. *PLoS One* **4:** e5556.

Dou D, Kale SD, Wang X, Jiang RHY, Bruce NA, Arredondo FD, Tyler BM. 2008. RXLR-mediated entry of *Phytophthora sojae* effector *Avr*1b into soybean cells does not require pathogen-encoded machinery. *The Plant Cell* **20:** 1930-1940.

Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **6:** 197–208.

Edgar RC, Batzoglou S. 2006. Multiple sequence alignment. *Current Opinion in Structural Biology* **16:** 1-6.

Egon A, Brosius J, Ptashne M. 1983. Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene* **25:** 167-178.

Egorov AM, Osipov AP, Dzantiev BB, Gavrilova EM. 1991. Theory and Practice of Enzyme Immunoassay, Moscow: Vysshaya Shkola. Pp. 1-65.

Ehsan H, Reichheld JP, Roef L, Witters E, Lardon F, Van Bockstaele D, Van Montagu M, Inze D, van-Onckelen H. 1998. Effect of indomethacin on cell cycle dependent cyclic AMP fluxes in tobacco BY-2 cells. *FEBS Letters* **422:** 165–169.

Ehsan H, Roef L, Witters E, Reichheld JP, Van Bockstaele D, Inze D, Van Onckelen H. 1999. Indomethacin-induced G1/S phase arrest of the plant cell cycle. *FEBS Letters* **458:** 349–353.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Science USA* **95:** 14863–14868.

Emanuelson O, Nielsen H, Brunak S, von Heijne G. 2000. Prediction of subcellular localisation of proteins based on their N-terminal amino-acid sequence. *Journal of Molecular Biology* **300:** 1005-1016.

Emmer M, deCrombrugghe B, Pastan I, Perlman RL. 1970. Cyclic AMP receptor protein of *E. coli*: its role in the synthesis of inducible enzymes. *Proceedings of National Academy of Science USA* **66:** 480-487.

Eron L, Block R. 1971. An adenosine 3':5'-cyclic monophosphate-binding protein that acts on the transcription process. *Proceedings of National Academy of Science USA* **68:** 1828-1832.

Erwin DC, Ribiero OK. 1996. Phytophthora Diseases Worldwide. St. Paul, MN: APS Press. Pp. 1-28.

Fakruddin M, Mohammad Mazumdar R, Bin Mannan KS, Chowdhury A, Hossain MN. 2013. Critical factors affecting the success of cloning, expression, and mass production of enzymes by recombinant *E. coli*. *ISRN Biotechnology* **2013:** 1–7.

Falcone-Ferreyra ML, Pezza A, Biarc J, Burlinngame AL, Casati P. 2010. Plant L10 ribosomal proteins have different roles during development and translation under ultraviolet-B stress. *Plant Physiology* **153:** 1878-1894.

Ferreira PA, Nakayama TA, Pak WL, Travis GH. 1996. Cyclophilin-related protein RanBP2 acts as chaperone for red/green opsin. *Nature* **383:** 637–640.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research* **39:** W29–W37.

Fowler S, Thomashow MF. 2002. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *The Plant Cell* **14:** 1675-1690.

Franken KLMC, Hiemstra HS, van Meijgaarden KE, Subronto Y, den Hartigh J, Ottenhoff THM, Drijfhout JW. 2000. Purification of His-tagged proteins by immobilised chelate affinity chromatography: The benefits from the use of organic solvent. *Protein Expression and Purification*. **18:** 95-99.

Frezza E, Martin J, Lavery R. 2018. A molecular dynamics study of adenylyl cyclase: The impact of ATP and G-protein binding. *PLoS ONE* **13:** e0196207.

Frydman J. 2001. Folding of newly translated proteins *in vivo*: the role of molecular chaperones. *Annual Review of Biochemistry* **70:** 603–647.

Fujii S, Suzuki T, Giege P, Higashiyama T, Koizuka N, Shikanai T. 2016. The restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291 RNA in Arabidopsis. *The Plant Journal*. **86:** 504-513.

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signalling networks. *Current Opinions in Plant Biology* **9:** 436–442.

Galat A. 1999. Variations of sequences and amino acid compositions of proteins that sustain their biological functions: an analysis of the cyclophilin family of proteins. *Archives of Biochemistry and Biophysics* **371:** 149–162.

Garty NB, Salomon Y. 1987. Stimulation of partially purified adenylate cyclase from bull sperm by bicarbonate. *FEBS Letters* **218:** 148–152.

Gatehouse JA. 2002. Plant resistance towards insect herbivores: a dynamic interaction. *New Phytologis*t **156:** 145–169.

Gattiker A, Gasteiger E, Bairoch A. 2002. ScanProsite: A reference implementation of a PROSITE scanning tool. *Applied Bioinformatics* **1:** 107–108.

Gehring C. 2010. Adenylyl cyclases and cAMP in plant signalling - past and present. *Cell Communication and Signalling* **8:** 15.

George R, Heringa J. 2002. Protein domain identification and improved sequence similarity searching using PSI-BLAST. *Protein: Structure, Function and Genetics* **48:** 672-681.

Gehring C, Turek IS. 2017. Cyclic nucleotide monophosphates and their cyclases in plant signalling. *Frontiers in Plant Science* **8:** 1704.

Geng W, Wang Z, Zhang J, Reed BY, Pak CY, Moe OW. 2005. Cloning and characterisation of the human soluble adenylyl cyclase. *American Journal of Physiology Cell Physiology* **288:** C1305–C1316.

Gill N, Findley S, Walling JG, Hans C, Ma J, Doyle J, Stacey G, Jackson SA. 2009. Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiology* **151:** 1167–1174.

Gillis J, Pavlidis P. 2011. The role of indirect connections in gene networks in predicting function. *Bioinformatics* **27:** 1860-1866.

Gilman AGA. 1970. Protein Binding Assay for Adenosine 3′-5′- Cyclic Monophosphate,

Proceedings of the National Academy of Science USA **67:** 305–312.

Godoy AV, Lazzaro AS, Casalongue CA, San Segundo B. 2000. Expression of a *Solanum tuberosum* cyclophilin gene is regulated by fungal infection and abiotic stress conditions. *Plant Science* **152:** 123–134.

Gordon VM., Young Jr WW, Lechler SM, Gray MC, Leppla SH, Hewlett EL. 1989. Adenylate cyclase toxins from *Bacillus anthracis* and *Bordetella pertussis*. Different processes for interaction and entry into target cells. *Journal of Biological Chemistry* **264:** 14792-14799.

Grant M, Brown I, Adams S, Knight M, Ainslie A, Mansfield J. 2000a. The RPM1 plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. *Plant Journal* **23:** 441–450.

Grant D, Cregan P, Shoemaker RC. 2000b. Genome organisation in dicots: genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. *Proceedings of the National Academy of Science USA* **97:** 4168–4173.

Gregory AW, Roayaei JA, Quiñones OA, Schneider KT. 2007. A microarray analysis for differential gene expression in the soybean genome using Bioconductor and R. *Briefings in Bioinformatics* **8:** 415-31.

Guerinot ML, Chelm BK. 1984. Isolation and expression of the *Bradyrhizobium japonicum* adenylate cyclase gene (cya) in *Escherichia coli. Journal of Bacteriology* **159:** 1068- 1071.

Guo L, Breakspear A, Zhao G, Gao L, Kistler HC, Xu JR, Ma LJ. 2016. Conservation and divergence of the cyclic adenosine monophosphate-protein kinase A (cAMP-PKA) pathway in two plant-pathogenic fungi: *Fusarium graminearum* and *F. verticillioides*. *Molecular Plant Pathology* **17:** 196-209.

Guo Y, Yu Y, Wang D, Wu C, Yang G, Huang J, Zheng C. 2009. GhZFP1, a novel CCCH-type zinc finger protein from cotton, enhances salt stress tolerance and fungal disease resistance in transgenic tobacco by interacting with GZIRD21A and GZIPR5. *New Phytologist* **183:** 62-75.

Gupta SK, Shukla P. 2016. Advanced technologies for improved expression of recombinant proteins in bacteria; perspectives and application. *Critical Reviews in Biotechnology* **36:** 1089-1098.

Hall RA, De Sordi L, Maccallum DM, Topal H, Eaton R, Bloor JW, Robinson GK, Levin LR, Buck J, Wang Y, Gow NA, Steegborn C, Muhlschlegel FA. 2010. CO_2 acts as a signalling molecule in populations of the fungal pathogen *Candida albicans*. *PLoS Pathogens* **6:** e1001193.

Hammes UZ, Schachtman DP, Berg RH, Nielsen E, Koch W, Mclntyre LM, Taylor CG. 2005. Nematode-induced changes of transporter gene expression in Arabidopsis roots. *Molecular Plant-Microbe Interactions* **18:** 1247-1257.

Hannenhalli SS, Russell RB. 2000. Analysis and prediction of functional sub-types from protein sequence alignments. *Journal of Molecular Biology* **303:** 61-76.

Hanoune J, Defer N. 2001. Regulation and role of adenylyl cyclase isoforms. *Annual Review of Pharmacology and Toxicology* **41:** 145–174.

Hartl FU. 1996. Molecular chaperones in cellular protein folding. *Nature* **381:** 571–580.

Heazlewood JL, Howell KA, Whelan J, Millar AH. 2003. Towards the analysis of the rice mitochondrial proteome. *Plant Physiology* **132:** 230–242.

Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH, 2004. Experimental analysis of the Arabidopsis mitochondrial proteome highlights signalling, and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial protein. *The Plant Cell* **16:** 241-256.

Heyman JA, Comthawaite J, Foncerrada L, Gilmore JR, Gontang E, Hartman KJ, Hernandez CL, Hood R, Hull HM, Lee W, Marcil R, Marsh EJ, Mudd KM, Patino MJ, Purcell TJ, Rowland JJ, Sindici ML, Hoeffler JP. 1999. Genome-scale cloning and expression of individual open reading frames using Topoisomerase 1-mediated ligation. *Genome Research* **9:** 383-392.

Hennig L, Menges M, Murray JA, Gruissem W. 2003. Arabidopsis transcript profiling on Affymetrix GeneChip arrays. *Plant Molecular Biology* **53:** 457–465.

Hepler PK, Wayne RO. 1985. Calcium and plant development. *Annual Review of Plant Physiology* **36:** 379 439.

Hetherington A, Trewavas A. 1984. The regulation of membrane-bound protein kinases by phospholipid and calcium. *Annual Proceedings of the Phytochemical Society of Europe* **24:** 181-197.

Higgins CF. 1992. ABC transporters: From microorganisms to man. *Annual Review of Cell and Developmental Biology* **8:** 67–113.

Ho HL. 2015. Functional roles of plant kinases in signal transduction pathways during abiotic and biotic stress. *Journal of Biodiversity, Bioprospecting and development* **2:** 147.

Ho JGS, Middelberg APJ, Ramage P, Kocher HP. 2003. The likelihood of aggregation during protein renaturation can be assessed using the second virial coefficient. *Protein Science* **12:** 708-716.

Holzinger A, Phillips KS, Weaver TE. 1996. Single-step purification/solubilisation of recombinant protein: Application to surfactant protein B. *BioTechniques* **20:** 804-808.

Howard MJ, Karasik A, Klemm BP, Mei C, Shanmuganathan A, Fierke CA, Koutmos M. 2016. Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. *RNA* **22:** 782-792.

Howlett AC, Sternweis PC, Macrik BA, Van Arsdale PM, Gilman AG. 1979. Reconstitution of catecholamine sensitive adenylate cyclase: Association of a regulatory component of the

enzyme with membranes containing the catalytic protein and P-adrenergic receptors. *Journal of Biological Chemistry* **254:** 2287-2295.

Hruz T, Laule O, Szabo G, Wessendrop F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P. 2008. Genevestigator v3: a reference expression database for the meta-analysis of Transcriptomes. *Advances in Bioinformatics* **2008:** 1–5.

Hu S, Zhou X, Gu X, Cao S, Wang C, Xu JR. 2014. The cAMP-PKA pathway regulates growth, sexual and asexual differentiation, and pathogenesis in *Fusarium graminearum*. *Molecular Plant-Microbe Interactions* **27:** 557-566.

Hudson LC, Garg R, Bost KL, Piller KJ. 2014. Soybean seeds: A practical host for the production of functional subunit vaccines. *BioMed Research International* **2014:** 1-13.

Hungria M, Mendes IC, 2015. Nitrogen fixation with soybean: the perfect symbiosis In: de Bruijn F. (Ed.), *Biological Nitrogen Fixation*., v.2, Chapter 99. John Wiley & Sons, New Jersey, Pp. 1005–1019.

Hvidsten TR, Komoroski J, Sandvik AK, Leagreid A. 2001. Predicting gene function from gene expressions and ontologies. *Pacific Symposium on Biocomputing* **2001:** 299–310.

Iakoucheva LM, Brown CJ, Law-son JD, Obradovic Z, Dunker AK. 2002. Intrinsic disorder in cell-signalling and cancer-associated proteins. *Journal Molecular Biology*. **323:** 73– 584.

Ideker T, Thorsson V, Siegel A, Hood L. 2000. Testing for differentially-expressed genes by maximum likelihood analysis of microarray data. *Journal of Computational Biology* **7:** 805–817.

Ito M, Takahashi H, Sawasaki T, Ohnishi K, Hikichi Y, Kiba A. 2014. Novel type adenylyl cyclase participates in tabotoxine-β-lactum-induced cell death and occurrence of wildfire disease in *Nicotiana benthamiana. Plant Signalling and Behaviour* **9:** e27420-1 –e27420-6.

Inskeep WP, Bloom PR. 1987. Soil chemical factors associated with soybean chlorosis in calciaquolls of western Minnesota. *Agronomy Journal* **79:** 779-786.

Isner JC, Maathuis FJ. 2011. Measurement of cellular cGMP in plant cells and tissues using the endogenous fluorescent reporter FlincG. *Plant Journal.* **65:** 329–334.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3:** 318- 328.

Jager R, Russwurm C, Schwede F, Genieser HG, Koesling D, Russwurm M. 2012. Activation of PDE10 and PDE11 phosphodiesterases, *Journal of Biological Chemistry* **287:** 1210–1219.

Jana S, Deb JK. 2005. Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Applications in Microbiology and Biotechnology* **67:** 289-298.

Jha B, Sharma A, Mishra A. 2011. Expression of *SbGSTU* (*tau* class glutathione S-transferase) gene isolated from *Salicornia brachiate* in tobacco for salt tolerance. *Molecular Biology Reports* **38:** 4823-4832.

Jiang J, Fan LW, Wu WH. 2005. Evidences for involvement of endogenous cAMP in Arabidopsis defense responses to Verticillium toxins. *Cell Research* **15:** 585–592.

Jin X, Wu W. 1999. Involvement of cyclic AMP in ABA- and Ca^{2+} -mediated signal transduction of stomatal regulation in *Vicia faba*. *Plant Cell Physiology* **40:** 1127-1333.

Jo B, Choi SS. 2015. Introns: The functional benefits of introns in genomes. *Genomics and Informatics* **13:** 112-118.

Jones DT, Cozzetto D. 2015. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31:** 857-863.

Joshi CP. Chiang VL. 1998. Conserved sequence motifs in plant *S*-adenosyl-L-methionine-dependant methyltransferases. *Plant Molecular Biology* **37:** 663-674.

Kamenetsky M, Middelhaufe S, Bank EM, Levin LR, Buck J, Steegborn C. 2006. Molecular details of cAMP generation in mammalian cells: a tale of two systems. *Journal of Molecular Biology* **362:** 623-639.

Kandoth PK, Ithal N, Recknor J, Maier T, Nettleton D, Baum TJ, Mitchum MG. 2011. The soybean Rhg1 locus for resistance to the soybean cyst nematode *Heterodera glycines* regulates the expression of a large number of stress- and defense-related genes in degenerating feeding cells. *Plant Physiology* **155:** 1960–1975.

Kaneko N, Ago H, Matsuda R. Inagaki E, Miyano M. 1997. Crystal structure of annexin V with ligand K-201 as a calcium activity inhibitor. *Journal of Molecular Biology* **217:** 16-20.

Kasahara M, Suetsugu N, Urano Y, Yamamoto C, Ohmori M, Takada Y, Okuda S, Nishiyama T, Sakayama H, Kohchi T, Takahashi F. 2016. An adenylyl cyclase with a phosphodiesterase domain in basal plants with a motile sperm system. *Nature Scientific Reports* **6:** 39232.

Kaschuk G. Nogueira MA, de Luca MJ, Hungria M. 2016. Responses of determinate and indeterminate soybean cultivars to basal and top-dressing N-fertilisation compared to sole inoculation with Bradyrhizobium. *Field Crops Research* **195:** 21-27.

Kelly LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 portal for protein modelling, prediction and analysis. *Nature Protocols* **10:** 845-858.

Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C, Kumar N, Liu Z, Maurel T, Moore B, McDowall MD, Maheswari U, Naamati G, Newman V, Ong CK, Bolser DM, De Silva N, Howe KL, Langridge N, Maslen G, Staines DM, Yates A. 2018. Ensembl Genomes 2018**:** an integrated omics infrastructure for non-vertebrate species *Nucleic Acids Research* **46**: D802–D808.

Kessler A, Baldwin IT. 2002. Plant responses to insect herbivory: the emerging analysis *Annual Review of Plant Biology* **53:** 299–328.

Kim JC, Lee SH, Cheong YH, Yoo C, Lee SI, Chun HJ, Yun D, Hong JC, Lee SY, Lim CO, Cho MJ. 2001. A novel cold-inducible zinc finger protein from soybean, SCOF-1, enhances cold tolerance in transgenic plants. *The Plant Journal* **25:** 247-259.

King R, Karwath A, Clare A, Dehaspe L. 2001. The utility of different representations of protein sequence for predicting functional class. *Bioinformatics* **17:** 445–454.

King RD, Wise PH, Clare A. 2004. Confirmation of data mining-based predictions of protein function. *Bioinformatics* **20:** 1110–1118.

Kneusel RE, Crowe J, Wulbeck M, Ribbe J. 1998. Procedures for the analysis and purification of His-tagged proteins. *Methods in molecular medicine* **13:** 293–308.

Ko KP, Park SK, Yang JJ, Ma SH, Gwack J, Shin A, Kim Y, Kang D, Chang SH, Shin HR, Yoo KY. 2013. Intake of soy products and other foods and gastric cancer risk: A prospective study. *Journal of Epidemiology* **23:** 337-343.

Kobayashi K, Kawabata M, Hisano K, Kazama T, Matsuoka K, Sugita M, Nakamura T. 2012. Identification and characterisation of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Research* **40:** 2712–2723.

Komatsu S, Hirano H. 1993. Protein-kinase activity and protein phosphorylation in rice (*Oryza sativa*) leaf. *Plant Science* **94:** 127–137.

Konc J, Skrlj B, Erzen N, Kunej T, Janezic D. 2017. GenProBis: web server for mapping of sequence variants to protein binding sites. *Nucleic Acids Research* **45:** W253-W259.

Kong HY, Lee SC, Hwang BK. 2001. Expression of pepper cyclophilin gene is differentially regulated during the pathogen infection and abiotic stress conditions. *Physiology and Molecular Plant Pathology* **59:** 189–199.

Kouchi H, Shimomura K, Hata S, Hirota A, Wu GJ, Kumagai H, Tajima S, Suganuma N, Suzuki A, Aoki T, Hayasha M, Yokoyama T, Ohyama T, Asamizu E, Kuwata C, Shibata D, Tabata S. 2004. Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. *DNA Research* **11:** 263–274.

Kovarova-Kovar K, Egli T. 1998. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed substrate kinetics. *Microbiology and Molecular Biology Reviews* **62:** 646–666.

Krzywicka A, Beisson J, Keller AM, Cohen J, Jerka-Dziadosz M, Klotz C. 2001. KIN241: a gene involved in cell morphogenesis in *Paramecium tetraurelia* reveals a novel protein family of cyclophilin-RNA interacting proteins (CRIPs) conserved from fission yeast to man. *Molecular Microbiol*ogy **42:** 257–267.

Kumudini SV Pallikonda PK, Steele C. 2007. Photoperiod and e-genes influence the duration of the reproductive phase in soybean. *Crop Science* **47:** 1510–1517.

Kuo J, Chang Y, Tseng C. 2003. Growth rate regulation of the lac operon expression in *Escherichia coli* is cyclic AMP dependent *FEBS letters* **553:** 397-402.

Laluk K, AbuQamar S, Mengiste T. 2011. The Arabidopsis mitochondrial-localised pentatricopeptide repeat protein PGN functions in defense against nectrotic fungi and abiotic stress tolerance. *Plant Physiology* **156:** 2053-2068.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E. 2012. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research* **40:** D1202–D1210.

Lauren KT. Fandl JP, Artz SW. 1990. Analysis of sequence Elements important for expression and regulation of the adenylate cyclase gene (*cya*) of *Salmonella typhimurium*. Genetics **125:** 709-717.

Le DT, Nishiyama R, Watanabe Y, Tanaka M, Seki M, Ham LH, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS. 2012. Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. *PLoS One* **7:** e49522.

Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB. 2007. Using genomics to study legume seed development. *Plant Physiology* **144:** 562-574.

Lee JH. Schöffl F. 1996. An Hsp70 antisense gene affects the expression of HSP70/HSC70, the regulation of HSF and the acquisition of thermotolerance in transgenic *Arabidopsis thaliana*. *Molecular and General Genetics* **252:** 11–19.

Lefkimmiatis K, Leronni D, Hofer AM. 2013. The inner and outer compartments of mitochondria are sites of distinct cAMP/PKA signalling dynamics. *Journal of Cell Biology* **202:** 453-462.

Lemtiri-Chlieh F, Thomas L, Marondedze C, Irving H, Gehring C. 2011. Cyclic nucleotides and nucleotide cyclases in plant stress responses. In *Abiotic Stress Response in Plants - Physiological, Biochemical and Genetic Perspectives.* Edited by Shanker A, Venkateswarlu B. InTech - Open Access Publisher; **2011:** 137–182.

Li MW, Qi X, Ni M, Lam HM. 2013. Silicon era of carbon-based life: application of genomics and bioinformatics in crop stress research. *International Journal of Molecular Sciences* **14:** 11444–83.

Li SC, Squires CL, Squires C. 1984. Antitermination of *E. coli* rRNA transcription is caused by a control region segment containing lambda nut-like sequences. *Cell* **38:** 851-860.

Li W, Luan S, Schreiber SL, Assmann SM. 1994. Cyclic AMP stimulates KC channel activity in mesophyll cells of *Vicia faba* L. *Plant Physiology* **106:** 957–961.

Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD, Stacey G. 2010. Complete transcriptome of the soybean root

hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiology* **152:** 541-552.

Lichtarge O, Bourne HR, Cohen FE, 1996. Evolutionarily conserved G-alpha-beta-gamma binding surfaces support a model of the G protein-receptor complex. *Proceedings of the National Academy of Science USA* **93:** 7507-7511.

Limbird LE, Hickey AR, Lefkowitz RJ. 1979. Unique uncoupling of the frog erythrocyte adenylate cyclase system by manganese. *The Journal of Biological Chemistry* **254:** 2677-2683.

Litvin TN, Kamenetsky M, Zarifyan A, Buck J, Levin LR. 2003. Kinetic properties of "soluble adenyl cyclase: Synergism between calcium and bicarbonate *The Journal of Biological Chemistry* **278:** 15922-15926.

Liu F, Andersen MN, Jensen CR. 2003. Loss of pod set caused by drought stress is associated with water status and ABA content of reproductive structures in soybean. *Functional Plant Biology* **30:** 271–280.

Liu S, Kandoth PK, Warren SD Yeckel G, Heinz R, Alden J, Yang C, Jamai A, El-Mellouki T, Juvale PS, Hill J, Baum TJ, Cianzio S, Whitham SA, Korkin D, Mitchum MG, Meksem K. 2012. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens, *Nature* **492:** 256–260.

Liu J, Li Y, Wang W, Gai J, Li Y. 2016. Genome-wide analysis of MATE transporters and expression patterns of a subgroup of *MATE* genes in response to aluminiun toxicity in soybean. *BMC Genomics* **17:** 223.

Liu XD, Xie L, Wei Y, Zhou X, Jia B, Liu J, Zhang S. 2014. Abiotic stress resistance, A novel moonlighting function of ribosomal protein RPL44 in halophilic *Aspergillus glaucus*. *Applied and Environmental Microbiology.* **80:** 4294-4300.

Loh JT, Ho SC, Wang JL. Schindler M. 1994. Carbohydrate binding activities of *Bradyrhizobium japonicum*: IV. Effect of lactose and flavones on the expression of the lectin, BJ38. *Glycoconjugate Journal* **11:** 363–370.

Lomovatskaya LA, Romanenko AS, Filinova NV, Dudareva LV. 2011. Determination of cAMP in plant cells by a modified enzyme immunoassay method. *Plant Cell Reports* **30:** 125-132.

Lomovatskaya LA, Romanenko AS, Krivolapova NV, Kopytchuk VN. 2005. Participation of potato cells cAMP in the transfer of systemic signal in ring rot pathogenesis. *Academy Open Inter*national **15:** 1-7.

Lomovatskaya LA, Romanenko AS, Fylinova NV, Kopytchuk VN. 2007. Functioning of "soluble" and membrane-associated forms of adenylyl cyclase in plant cell organelles under biotic stress. *Biological Membrane* **24:** 370–378.

Londos C, Preston MS. 1977. Activation of the hepatic adenylate cyclase system by divalent cations. *The Journal of Biological Chemistry* **252:** 5957-5961.

Lopes-Caitar VS, de Carvalho MC, Darben LM, Kuwahara MK, Nepomuceno AL, Dias WP, Abdelnoor RV, Marcelino-Guimaraes FC. 2013. Genome-wide analysis of the *Hsp*20 gene family in soybean: a comprehensive sequence, genomic organisation and expression profile analysis under abiotic and biotic stresses. *BMC Genomics* **14:** 577.

Lorick KL, Jensen JP, Fang S, Ong AM, Hatakeyama S, Weissman AM. 1999. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Processing of the National Academy of Science USA* **96:** 11364-11369.

Lurin C, Andre´s C, Aubourg S, Bellaoui M, Bitton F, Bruye`re C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin- Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small. 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16:** 2089–2103.

Luo M, Liang XQ, Dang P, Holbrook CC, Bausher MG, Lee RD, Guo BZ. 2005. Microarray-based screening of differentially expressed genes in peanut in response to *Aspergillus parasiticus* infection and drought stress. *Plant Science* **169:** 695–703.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290:** 1151-1155.

Makrodimitris S, van Ham RCH, Reinders MJT. 2018. Improving protein function prediction using protein sequence and GO-term similarities. *Bioinformatics* **208:** 1-9.

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002. The protein kinase complement of the human genome. *Science* **298:** 1912–1934.

Martinez M. 2013. From plant genomes to protein families: Computational tools. *Computational and Structural Biotechnology Journal* **10:** e20130700.

Marivet J, Margispinheiro M, Frendo P, Burkard G. 1994. Bean cyclophilin gene-expression during plant development and stress conditions. *Plant Molecular Biology* **26:** 1181–1189.

Mathieu-Demaziere C, Poinsot V, Masson-Boivin C, Garnerone A, Batut J. 2013. Biochemical and functional characterisation of SpdA, a 2′,3′-cyclic nucleotide phosphodiesterase from *Sinorhizobium meliloti. BMC Microbiology* **13:** 268.

Mentewab A, Cardoza V, Stewart CNJ. 2005. Genomic analysis of the response of *Arabidopsis thaliana* to trinitrotoluene as revealed by cDNA microarrays. *Plant Science* **168:** 1409–1424.

Miernyk JA. 1997. The 70 kDa stress-related proteins as molecular chaperones. *Trends in Plant Science* **2:** 180–187.

Mihic SJ, Harris RA. 1996. Single step Purification/Solubilisation of recombinant proteins: Application to surfactant protein B. *Biotechniques* **20:** 804-808.

Minneci F, Piovesan D, Cozzetto D, Jones DT. 2013. FFPred 2.0: Improved homology-independent prediction of gene ontology terms for eukaryotic protein sequences. *PLoS ONE* **8:** e63754.

Miura K. Hasegawa PM. 2010. Sumoylation and other ubiquitin-like post-translational modifications in plants. *Trends in Cell Biology* **20:** 223-232.

Morimoto RI. 1998. Regulation of the heat shook transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes and Development* **12:** 3788–3796.

Morris PF, Bone E, Tyler BM. 1998. Chemotropic and contact responses of *Phytophthora sojae* hyphae to soybean isoflavonoids and artificial substrates. *Plant Physiology* **117:** 1171-1178.

Mostafavi S, Morris Q. 2010. Fast integration of heterogeneous data sources for predicting gene function with limited annotation. *Bioinformatics* **26:** 1759-1765.

Moutinho A, Hussey PJ, Trewavas AJ, Malho R. 2001. cAMP acts as a second messenger in pollen tube growth and reorientation. *Proceedings of the National Academy of Sciences USA* **98:** 10481–10486.

Moy P, Qutob D, Chapman BP, Atkinson I, Gijzen M. 2004. Chemoattraction of Zoospores of the soybean pathogen, *P. sojae*, by isoflavones. *Molecular Plant–Microbe Interaction* **17:** 1051–1062.

Mukhopadhyay A, Vij S, Tyagi AK. 2004. Over-expression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco. *Proceedings of the National Academy of Sciences, USA* **101:** 6309–6314.

Müller-Hill B, Crapo L, Gilbert W. 1968. Mutants that make more *lac* repressor. *Proceedings of the National Academy of Science USA* **59:** 1259-1262.

Neer EJ. 1979. Interaction of soluble brain adenylate cyclase with manganese. *The Journal of Biological Chemistry* **254:** 2080-2096.

Nawrath C, Heck S, Parinthawong N, Metraux JP. 2002. EDS5, an essential component of salicylic acid-dependent signalling for disease resistance in Arabidopsis, is a member of the MATE transporter family. *Plant cell* **14:** 275-286.

Neves-Zaph SR, Song RS. 2015. Development of computational models of cAMP signalling. *Methods in Molecular Biology* **1294:** 203- 217.

Nimnoi P, Pongslip N, Lumyong S. 2014. Co-inoculation of soybean with Actinomycetes and *Bradyrhibium japonicum* enhances plant growth, nitrogenase activity and plant nutrition. *Journal of Plant Nutrition* **37:** 432-446.

Noctor G. Foyer CH. 1998. Ascorbate and glutathione: Keeping active oxygen under control. *Annual Review of Plant Physiology and Plant Molecular Biology* **49:** 249–279.

Nordstrom A, Jacobs FA, Eliasson L. 1991. Effect of exogenous indole-3-acetic acid and indole-3-butyric acid on internal levels of the respective auxins and their conjugation with aspartic acid during adventitious root formation in pea cuttings. *Plant Physiology* **96:** 856–861.

Northup JK, Sternweis PC, Gilman AG. 1983. The subunits of the stimulatory regulatory component of adenylate cyclase. Resolution, activity and properties of the 35,000 dalton (beta) subunit. *Journal of Biological Chemistry* **258:** 11361-11368.

Obulareddy N, Panchal S, Melotto M. 2013. Guard cell and RNA isolation suitable for high throughout transcriptional analysis of cell-type responses to biotic stresses. *Molecular Plant-Microbe Interaction* **26:** 844-849.

Oh M, Komatsu S. 2015. Characterisation of proteins in soybean roots under flooding and drought stress. *Journal of Proteomics* **114:** 161-181.

Oh SK, Yi SY, Yu SH, Moon JS, Park JM, Choi D. 2006. CaWRKY2, a chili pepper transcription factor, is rapidly induced by incompatible plant pathogens. *Molecular Cell* **22:** 58–64.

Okuda K, Chateigner-Boutin AL, Nakamura T, Delannoy E, Sugita M, Myouga F, Motohashi R, Shinozaki K, Small I, Shikanai T. 2009. Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell* **21:** 147–156.

Okuda K, Shikanai T. 2012. A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. *Nucleic Acids Research* **11:** 5052-5064.

Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. 2008. Flexible nets: Disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9:** S1.

Olins PO, Devine CS, Rangwala SH, Kavka KS. 1988. T7 Phage gene 10 Leader RNA: A ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene* **73:** 227-235.

O'Rourke JA, Nelson RT, Grant D, Schmutz J, Grimwood J, Cannon S, Vance CP, Graham MA, Shoemaker RC. 2009. Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genomics* **10:** 376.

Omori KJ, Kotera J. 2007. Overview of PDEs and their regulation, *Circulation Research* **100:** 309–32.

Ono K, Hibino T, Kohinata T, Suzuki S, Tanaka Y, Nakamura T, Takabe T, Takabe T. 2001. Overexpression of DnaK from a halotolerant Cyanobacterium *Aphanothece halophytica* enhances the high temperature tolerance of tobacco during germination and early growth. *Plant Science* **160:** 455–461.

Osawa T, Inanaga H, Sato C, Numata T, 2015. Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. *Molecular Cell* **58:** 418-430.

Pacini B, Petrigliano A, Diffley P, Paffetti A, Brown EG, Martelli P, Trabalzini L, Bovalini L, Lusini P, Newton RP. 1993. Adenylyl cyclase activity in roots of *Pisum sativum*. *Phytochemistry* **34:** 899–903.

Panchy N, Lehti-Shiu M, Shiu S. 2016. Evolution of gene duplication in pants. *Plant Physiology* **171:** 2294-2316.

Panthee DR, Yuan JS, Wright DL, Marois JJ, Mailhot D, Stewart CN Jr. 2007. Gene expression analysis in soybean in response to the causal agent of Asian soybean rust (*Phakopsora pachyrhizi* Sydow) in an early growth stage. *Functional and Integrative Genomics* **7:** 291–301.

Patel N, Cardoza V, Christensen E, Rekapalli B, Ayalew M, Stewart CNJ. 2004. Differential gene expression of *Chlamydomonas reinhardtii* in response to 2,4,6-trinitrotoluene (TNT) using microarray analysis. *Plant Science* **167:** 1109–1122.

Pavlidis P, Weston J, Cai J, Grundy WN. 2001. Gene functional classification from heterogeneous data. *In Proceedings of the Fifth Annual International Conference on Computational Biology*, April 22-25, 2001, Montreal S. Istrail, M. Waterman, Clark A. 249-255.

Pazos F, Pietrosemoli N, Garcia-Martin JA, Solano R. 2013. Protein intrinsic disorder in plants. *Frontiers in Plant Science* **4:** 1-5.

Perlman RL, De Crombrugghe B, Pastan I. 1969. Cyclic AMP regulates catabolite and transient repression in *E. coli*. *Nature* **223:** 810–812.

Perlman R, Pastan I. 1973. Pleiotropic deficiency of carbohydrate utilisation in an adenyl cyclase deficient mutant of *Escherichia coli*. *Biochemical and Biophysical Research Communications* **37:** 151-157.

Pfalz J, Bayraktar O, Prikryl J, Barkan A. 2009. Site-specific binding of a PPR protein defines and stabilizes 59 and 39 mRNA termini in chloroplasts. *EMBO Journal* **28:** 2042–2052.

Phuong TM, Nhung NP. 2013. Predicting gene function using similarity learning. *BMC Genomics* **14:** (Suppl 4): S4.

Pozdniakova S, Ladilov Y. 2018. Functional significance of the Adcy10-dependent intracellular cAMP compartments. *Journal of Cardiovascular Development and Disease* **5:** 29.

Prikryl J, Rojas M, Schuster G, Barkan A. 2011. Mechanism of RNA stabilisation and translational activation by a pentatricopeptide repeat protein. *Proceedings of the. National Academy of Science* USA **108:** 415–420.

Rail TW, Sutherland EW, 1958. Formation of a cyclic adenine ribonucleotide by tissue particles *Journal of Biological Chemistry* **232:** 1065-1067.

Rajasekhar VK, Lamb C, Dixon RA. 1999. Early events in the signal pathway for the oxidative burst in soybean cells exposed to avirulent *Pseudomonas syringae* pv. *Glycinea*. *Plant Physiology* **120:** 1137–1146.

Rampey RA, Leclere S, Kowalczyk M, Ljung K, Sandberg G, Bartel B. 2004. A family of auxin conjugate hydrolases that contributes to free indole-3-acetic acid levels during Arabidopsis germination. *Plant Physiology* **135:** 978-988.

Rassow J, Mohrs K, Koid S, Barthelmess IB, Pfanner N, Tropschug M. 1995. Cyclophilin-20 is involved in mitochondrial protein folding in cooperation with molecular chaperones Hsp70 and Hsp60. *Molecular and Cell Biology* **15:** 2654–2662.

Rato C, Monteiro D, Hepler PK, Malho R. 2004. Calmodulin activity and cAMP signalling modulate growth and apical secretion in pollen tubes. *Plant Journal* **38:** 887-897.

Redman JC, Haas BJ, Tanimoto G, Town CD. 2004. Development and evaluation of an Arabidopsis whole genome Affymetrix probe array. *Plant Journal* **38:** 545–561.

Richards H, Das S, Smith CJ, Pereira L, Geisbrecht A, Devitt NJ, Games DE, Geyschem J, Brenton AG, Newton RP. 2002. Cyclic nucleotide content of tobacco BY-2 cells. *Phytochemistry* **61:** 531-537.

Riechmann JL. Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D,

Sherman BK, Yu G. 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290:** 2105–2110.

Robison GA, Butcher RW, Sutherland EW. 1968. Cyclic AMP. *Annual Review of Biochemistry* **37:** 149-174.

Rocke DM, Durbin B. 2001. A model for measurement error for gene expression microarrays. *Journal of Computational Biology* **8:** 557–569.

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

Sadana R, Dessauer CW. 2009. Physiological roles for G protein-regulated adenylyl cyclase isoforms: Insights from knockout and overexpression studies. *Neurosignals* **17:** 5-22.

Saier MH, Reizer J. 1994. The bacterial phosphotransferase system: new frontiers 30 years later. *Molecular Microbiology* **13:** 755–764.

Sakamoto H, Maruyama K, Sakuma Y, Meshi T, Iwabuchi M, Shinozaki K, Yamaguchi-Shinozaki K. 2004. Arabidopsis Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. *Plant Physiology* **136:** 2734–2746.

Schaffer AA, Aravind L, Madden TL, Shavirin S, Spoung JL, Wolf YI, Koonin EV, Altschul SF. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Research* **29:** 2994-3005.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics* **37:** 501–506.

Schmid A, Sutto Z, Nlend MC, Horvath G, Schmid N, Buck J, Levin LR, Conner GE, Fregien N, Salathe M. 2007. Soluble adenylyl cyclase is localised to cilia and contributes to ciliary beat frequency regulation via production of cAMP. *Journal of General Physiology* **130:** 99-109.

Schmitz-Linneweber C, Small I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in Plant Science* **13:** 663–670.

Schmitz-Linneweber C, Williams-Carrier RE, Williams-Voelker PM, Kroeger TS, Vichas A, Barkan A. 2006. A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell* **18:** 2650–2663.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J,

Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **463:** 178–183.

Schoner BE, Belagaje RM, Schoner RG. 1986. Translation of a synthetic two-cistron mRNA in *Escherichia coli*. *Proceedings of the National Academy of Science USA* **83:** 8506-8510.

Severin AJ, Woody JL, Bolon Y, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA, Cannon SB, May GD, Vance CP, Shoemaker RC. 2010. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biology* **10:** 160–175.

Shah S, Peterkofsky A. 1991. Characterisation and generation of *Escherichia coli* adenylate cyclase deletion mutants. *Journal of. Bacteriology* **173:** 3238-3242.

Shan W, Cao M, Leung D, Tyler BM. 2003. The *Avr1b* Locus of *Phytophora sojae* encodes an elicitor and a regulator required for avirulence on soybean plants carrying resistance gene *Rps*1b. *The American Phytopathological Society* **17:** 394-403.

Sharma AD, Singh P. 2003. Effect of water stress on expression of a 20 kD cyclophilin-like protein in drought susceptible and tolerant cultivars of Sorghum. *Journal of Plant Biochemistry and Biotechnology* **12:** 77–80.

Shinozaki K, Yamaguchi-Shinozaki K. 2000. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signalling pathways. *Current Opinion Plant Biology* **3:** 217-223.

Shuman S. 1991. Site-specific interaction of Vaccinia virus topoisomerase I with duplex DNA. Minimal DNA substrate for strand cleavage *in vitro*. *Journal of Biological Chemistry* **266:** 11372–11379.

Shuman S. 1992. Two classes of DNA end-joining reactions catalysed by Vaccinia topoisomerase I. *Journal of Biological Chemistry* **267:** 16755–16758.

Shuman S. 1994. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *Journal of Biological Chemistry* **269:** 32678-32684.

Shuman S, Moss B. 1987. Identification of a vaccinia virus gene encoding a type I DNA topoisomerase. *Proceedings of the National Academy of Science USA.* **84:** 7478–7482.

Shrivastava S, Shukla P, Deepalakshmi PD, Mukhopadhyay K. 2013. Characterisation, cloning and functional expression of novel xylanase from *Thermomyces lanuginosus* SS-8 isolated from self-heating plant material. *World Journal of Microbiology and Biotechnology* **12:** 2407–2415.

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* **144:** 329-338.

Sidler M, Hassa P, Hasan S, Ringli C, Dudler R. 1998. Involvement of an ABC transporter in a developmental pathway regulating hypocotyl cell elongation in the light. *Plant Cell* **10:** 1623–1636.

Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesyley SC, Godzik A. 2007. XtalPred: a web server for prediction of protein crystallisability. *Structural Bioinformatics* **23:** 3403-3405.

Small ID, Peeters N. 2000. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends in Biochemical Sciences* **25:** 46-47.

Song F, Tang DL, Wang XL, Wang YZ. 2011. Biodegradable soy-protein isolate-based materials: A review. *Biomacromolecules* **12:** 3369–3380.

Spadiut O, Capone S, Krainer F, Glieder A, Herwig C. 2014. Microbials for the production of monoclonal antibodies and antibody fragments. *Trends in Biotechnology* **32:** 54-60.

Steegborn C, Litvin TN, Levin LR, Buck J, Wu H. 2005. Bicarbonate activation of adenylyl cyclase via promotion of catalytic active site closure and metal recruitment. *Nature Structural and Molecular Biology* **12:** 32-7.

Sternweis PC, Gilman AG. 1982. Aluminium: A requirement for activation of the regulatory component of adenylate cyclase by fluoride. *Proceedings of the National Academy of Sciences USA* **79:** 4888-4891.

Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **189:** 113–130.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* **185:** 60–89.

Stukkens Y, Bultreys A, Grec S, Trombik T, Vanham D, Boutry M. 2005. NpPDR1, a pleiotropic drug resistance-type ATP-binding cassette transporter from *Nicotiana plumbaginifolia*, plays a major role in plant pathogen defense. *Plant Physiology* **139:** 341-352.

Subramanian S, Stacey G, Yu O. 2006. Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *The Plant Journal* **48:** 261-273.

Sun W, Montagu MV, Verbruggen N. 2002. Review: Small heat shock proteins and stress tolerance in plants. *Biochimica et Biophysica Acta* **1577:** 1–9.

Sun X, Rikkerink EHA, Jones WT, Uversky VN. 2013. Multifarious roles of intrinsic disorder in proteins illustrates its broad impact on plant biology. *The Plant Cell* **25:** 38-55.

Sung DY, Guy CL. 2003. Physiological and molecular assessment of altered expression of Hsc70-1 in Arabidopsis. Evidence for pleiotropic consequences. *Plant Physiology* **132:** 979–987.

Sunkar R, Kaplan B, Bouche N, Arazi T, Dolev D, Talke IN, Maathuis FJ, Sanders D, Bouchez D, Fromm H. 2000. Expression of a truncated tobacco NtCBP4 channel in transgenic

plants and disruption of the homologous Arabidopsis CNGC1 gene confer Pb2+ tolerance. *Plant Journal* **24:** 533–542.

Swartz JR. 2001. Advances in *Escherichia coli* production of therapeutic proteins. *Current Opinions Biotechnology* **12:** 195-201.

Świeżawska B, Jaworski K, Pawelek A, Grzegorzeska W, Szewczuk P, Szmidt-Jaworska A. 2014. Molecular cloning and characterisation of a novel adenylyl cyclase gene, *HpAC1*, involved in stress signalling in *Hippeastrum hybridum. Plant Physiology and Biochemistry* **80:** 41-52.

Syed U, Yona G. 2003. Using a mixture of probabilistic decision trees for direct prediction of protein function. In: *Proceedings of RECOMB ACM,* Berlin, Germany. Pp. 1-6.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. 2017. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* **45:** D362-D368.

Tan J, Tan Z, Wu F, Sheng P, Heng Y, Xinhua W, Ren Y, Wang J, Gou X, Zhang X, Cheng Z, Jiang L, Liu X, Wang H, Wan J. 2014. A novel chloroplast-localised pentatricopeptide repeat protein involved in splicing affects chloroplast development and abiotic stress response in rice. *Molecular Plant* **7:** 1329-1349.

Tan DZ, Tuong HM, Thuy VTT, Son LV, Mau CH. 2015. Cloning and expression of GmDREB2 gene from a Vietnamese drought resistant soybean variety. *Brazilian Archives of Biology and Technology* **58:** 651-657.

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. 1999. Interpreting patterns of gene expression with self-organising maps: Methods and application to hematopoitic differentiation. *Proceeding of the National Academy Science USA*, **96:** 2907–2912.

Tellez-Sosa J, Soberon N, Vega-Segura A, Torre-Marquez ME, Cevallo MA. 2002. The *Rhizobium etli cya*C product: Characterisation of a novel adenylate cyclase class. *Journal of Bacteriology* **184**: 3560-3568.

Terakado J, Okamura M, Fujihara S, Ohmori M, Yoneyama T. 1997. Cyclic AMP in Rhizobia and symbiotic nodules. *Annuals of Botany* **80:** 499–503.

Thomas L, Marondedze C, Ederli L, Pasqualini S, Gehring C. 2013. Proteomic signatures implicate cAMP in light and temperature responses in *Arabidopsis thaliana*. *Journal of Proteomics* **83:** 47–59.

Tijseen P. 1985. Practice and theory of enzyme immunoassays. Elsevier. Amsterdam/New York. Pp. 1-5.

Tompa P. 2005. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett*ers **579:** 3346–3354.

Tompa P. 2012. Intrinsically disordered proteins: A 10-year re-cap. *Trends in Biochemical Sciences* **37:** 509- 516.

Topal H, Fulcher NB, Bitterman J, Salazar E, Buck J, Levin LR, Cann MJ, Wolfgang MC, Steegborn C. 2012. Crystal structure and regulation mechanisms of the *Cya*B adenylyl cyclase from the human pathogen *Pseudomonas aeruginosa*. *Journal of Molecular Biology* **416:** 271–286.

Tresguerres M, Barott MKL, Helman Y, Barron ME, Hess KC, Roa, JN. 2014. Established and potential physiological roles of bicarbonate-sensing soluble adenylyl cyclase (sAC) in aquatic animals. *Journal of Experimental Biology* **217:** 663–672.

Tresguerres M, Levin LR, Buck J. 2011. Intracellular cAMP signalling by soluble adenylyl cyclase. *Kidney International* **79:** 1277-1288.

Tripathi NK, Shrivastva A, Biswal KC, Lakshmana Rao PV. 2009. Optimisation of culture medium for production of recombinant dengue protein in *Escherichia coli*. *Industrial Biotechnology* **5:** 179-183.

Tsumoto K, Ejima D, Kumagai I, Arakawa T. 2003. Practical considerations in refolding proteins from inclusion bodies. *Protein Expression and Purification* **28:** 1-8.

Turnham RE, Scott JD. 2016. Protein kinase A catalytic subunit isoform *PRKACA*: History, function and physiology. *Gene* **577:** 101-108.

Turner GW, Cuthbertson DJ, Voo SS, Settles ML, Grimes HD, Lange BM. 2012. Experimental sink removal induces stress responses, including shifts in amino acid and phenylpropanoid metabolism, in soybean leaves. *Planta* **235:** 939- 954.

Tyler BM. 2007. *Phytophtora sojae*: root pathogen of soybean and model oomycete. *Molecular Plant Pathology* **8:** 1-8.

Udo H. 2015. An alternative method to facilitate cDNA cloning for expression studies in mammalian cells by introducing positive blue white selection in *Vaccinia* topoisomerase 1-mediated recombination. *PLoS ONE* **10:** e0139349.

Uematsu K, Nakajima M, Yamaguchi I, Yoneyama K, Fukui Y. 2007. Role of cAMP in gibberellin promotion of seed germination in *Orobanche minor* Smith. *Journal Plant Growth Regulation* **26:** 245–254.

Ullmann A, Danchin A. 1983. Advances in Cyclic Nucleotide Research. Raven, New York **15:** 1–53.

Upchurch RG, Elkan GH. 1978. The role of ammonia, L glutamate and cyclic adenosine 3′5′ monophosphate in the regulation of ammonia assimilation in *Rhizobium jacanicum*. *Biochimica et Biophysica Acta* **538:** 244-248.

Upchurch RG, Ramirez ME. 2011. Soybean plastidal omega-3 fatty acid desaturase genes: Structure and expression. *Crop Science* **51:** 1673–1682.

Uversky VN, Oldfield CJ, Dunker AK. 2005. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signalling. *Journal of Molecular Recognition.* **18:** 343–384.

Valsecchi F, Konrad C, Manfredi G. 2014. Role of soluble adenylyl cyclase in mitochondria. *Biochim Biophys Acta* **1842:** 2555-2560.

van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, de Peer YV, Coppens F, Vandepoele K. 2018. PLAZA 4.0: an integrative resource for functional evolutionary and comparative plant genomics. *Nucleic Acids Research* **46:** D1190-1196.

van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. 2018. Gene co-expression analysis for functional classification and gene-disease predictions. *Briefings in Bioinformatics* **19:** 575-592.

van Damme T, Blancquaert D, Couturon P, Van Der Straeten D, Sandra P, Lynen F. 2014. Wounding stress causes rapid increase in concentration of the naturally occurring 2′,3′-isomers of cyclic guanosine- and cyclic adenosine monophosphate (cGMP and cAMP) in plant tissues. *Phytochemistry* **103:** 59–66.

Visconti PE, Muschiett JP, Flawin MM, Tezon JG. 1990. Bicarbonate dependence of cAMP accumulation induced phorbol esters in hamster spermatozoa. *Biochimica et Biophysica Acta* **1054:** 231-236.

Volotovski ID, Sokolovsky SG, Molchan OV, Knight MR. 1998. Second messengers mediate increases in cytosolic calcium in tobacco protoplasts. *Plant Physiolology* **117:** 1023–1030.

Walling JG, Shoemaker R, Young N, Mudge J, Jackson S. 2006. Chromosome-level homeology in paleopolyploid soybean (*Glycine max*) revealed through integration of genetic and chromosome maps. *Genetics* **172:** 1893–1900.

Wang L, Cao C, Ma Q, Zeng Q, Wang H, Cheng Z, Zhu G, Qi J, Ma H, Nian H, Wang Y. 2014. RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biology* **14:** 169.

Ward EWB. 1990. The interaction of soya beans with *Phytophthora megasperma* f. sp. *glycinea*: pathogenicity. Pp. 311-327 In: Biological control of soil bourne plant pathogens. Hornby B ed. CAB International, Wallingford, U.K.

Wheeler JI, Freihat L, Irving HR. 2013. A cyclic nucleotide sensitive promoter reporter suitable for bacteria and plant cells. *BMC Biotechnology* **13:** 97.

Witters E, Roef L, Newton RP, Van Dongen W, Esmans EL, Van Onckelen HA. 1996. Quantitation of cyclic nucleotides in biological samples by negative electrospray tandem mass spectrometry coupled to ion suppression liquid chromatography. *Rapid Communications Mass Spectrometry* **10:** 225–231.

Wojtaszek P, Bolwell GP. 1997. Mechanisms for the generation of reactive oxygen species in plant defense: a broad perspective. *Physiological and Molecular Plant Pathology* **51:** 347–366.

Wolf J, Cook GH, Goldhammer AR, Berkowitz SA. 1980. Calmodulin activates prokaryotic adenylate cyclase. *Proceedings of the National Academy of Science USA* **77:** 3841-3844.

Wong A, Gehring C. 2013. The *Arabidopsis thaliana* proteome harbours undiscovered multi-domain molecules with functional guanylyl cyclase catalytic centres. *Cell Communication and Signalling* **11:** 48.

Wuttke MS, Buck J, Levin LR. 2001. Bicarbonate-regulated soluble adenylyl cyclase. *Journal of the Pancreas* **2:** 154–158.

Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lu S, Wu H, Tabata S, Harada K. 2012. Positional cloning and characterisation reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. *Proceeding of the National Academy of Science USA* **109:** E2155–E2164.

Xia Z, Zhai H, Lu S, Wu H, Zhang Y. 2013. Recent achievement in gene cloning and functional genomics in soybean. *The Scientific World Journal* **2013:** 28367.

Yamaguchi H, Miyazaki M. 2014. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **4:** 235-251.

Yamaguchi-Shinozaki K, Shinozaki K. 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annual Review of Plant Biology* **57:** 781-803.

Yin X, Sakata K, Nanjo Y, Komatsu S. 2014. Analysis of initial changes in the proteins of soybean root tip under flooding stress using gel-free and gel-based proteomic techniques. *Journal of Proteomics* **106:** 1-16.

Yunghans WN, Morre, DJ. 1977. Adenylate cyclase activity not found in soybean hypocotyl and onion meristem. *Plant Physiology* **60:** 144-149.

Yurchenko V, Zybarth G, O'Connor M, Dai WW, Franchin G, Hao T, Guo H, Hung HC, Toole B, Gallay P, Sherry B, Bukrinsky M. 2002. Active site residues of cyclophilin A are crucial for its signalling activity via CD147. *Journal of Biological Chemistry* **277:** 22959–22965.

Zeng W, Brutus A, Kremer JM, Withres JC, Gao X, Jones AD, He SY. 2011. A genetic screen reveals Arabidopsis stomatal and/or apoplastic defences against *Pseudomonas syringae* pv. Tomato DC3000. *PLoS Pathogens* **7:** E1002291-E1002291.

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. 2004. GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiology* **136:** 2621-2632.

Zippin JH, Chen Y, Nahirney P, Kamenetsky R, Fischman DA, Levin LR, Buck J. 2003. Compartmentalisation of bicarbonate-sensitive adenylyl cyclase in distinct signalling microdomains *FASEB Journal* **17:** 82–84.

Zippin JH, Chen Y, Straub SG, Hess KC, Diaz A, Lee D, Tso P, Holz GG, Sharp GW, Levin LR, Buck J. 2013. CO_2/HCO_3 and calcium-regulated soluble adenylyl cyclase as a physiological ATP sensor. *Journal of Biological Chemistry* **288:** 33283–33291.

Zhang Y, Zhao J, Xiang Y, Bian X, Zuo Q, Shen Q, Gai J, Xing. 2011. Proteomics study of changes in soybean lines resistant and sensitive to *Phytophthora sojae*. *Proteome Science* **9:** 52

Zhou Q, Tian A, Zou H, Xie Z, Lei G, Huang J, Wang C, Wang H, Zhang J, Chen S. 2008. Soybean WRKY-type transcription factor genes, *GmWRKY13*, *GmWRKY21* and *GmWRKY54*, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. *Plant Biotechnology Journal* **6:** 486-503.

Zolman BK, Silva ID, Bartel B. 2001. Arabidopsis *pxa1* mutant is defective in an ATP-binding cassette transporter-like protein required for peroxisomal fatty acid β- oxidation. *Plant Physiology* **127:** 1266-1278.

Zubay G, Schwartz D, Beckwith J. 1970. Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proceedings of National Academy of Science USA* **66:** 104-110.

APPENDICES

Appendix A: cAMP enzyme immunoassay (EIA) procedure

All reagents were allowed to warm to room temperature for 30 minutes before they were opened. All the standards and samples for the EIA procedure were run in triplicate.

1. The number of wells to be used were determined and 100 µl of Assay Buffer 2 was pipetted into the NBS tube and Bo (0 pmol/ml Standard) wells.
2. 100 µl of Standards 1 through 5 were pipetted into the appropriate wells.
3. 50 µl of yellow cAMP EIA Antibody was pipetted into each well except in the NBS wells and incubated at room temperature for 2 hours on a plate shaker at ~ 500 rpm while they were covered with a plate sealer.
4. After the 2 hours had lapsed, the contents of the plates were emptied, and each well was washed by adding 200 µl of 1x Wash Buffer. The washing procedure was repeated to make a total of 3 washes.
5. After the final wash, the wells were emptied and firmly held on a lint free paper towel to remove any residual wash buffer. This was done to avoid variation in the assay which would result during substrate addition.
6. 200 µl of the *p*-Nitrophenyl Phosphate Substrate Solution was added to every well and the reaction wells were incubated at room temperature for 60 minutes without shaking.
7. After the 1hour incubation at room temperature was over, 50 µl of the Stop solution was added to every well. This process was necessary as it stops the reaction and the readings were done soon after.
8. The plate reader was blanked against the Blank wells and the optical density was read at 405nm.

Appendix B: PSI-BLAST DOMAINATION multiple sequence alignments

At4g15640: *Arabidopsis thaliana*

```
 Score =  403 bits (1035), Expect = e-110,   Method: Composition-based stats.
 Identities = 193/321 (60%), Positives = 246/321 (76%)

Query: 53  LSKAFYSSGVGTVEATPSEDVKELYDKMLDSVKVKRSMPPNAWLWSMIANCKHQPDIRLL 112
           + ++ ++S    +   P+  VK+L+DKML+SV VKRSMPPNAWLW +I NC++Q DI LL
Sbjct: 66  IPRSSFTSEAEKLAGNPTVTVKDLHDKMLNSVNVKRSMPPNAWLWLLIENCQNQDDIHLL 125

Query: 113 FDILQNLRRFRLSNLRIHDNFNCNLCREVAKACVHAGALDFGMKALWKHNVYGLTPNIAS 172
           FD+LQNLRRFRLSNLRIHDNFNCNLC++VAK CV  GA+D G KALWKHNV+GLTP++AS
Sbjct: 126 FDVLQNLRRFRLSNLRIHDNFNCNLCQQVAKTCVRVGAIDSGKKALWKHNVHGLTPSVAS 185

Query: 173 AHHLLTNAKNHNDTXXXXXXXXXXXXNDLPLQPGTADIVFSICYNTDDWELINKYAKRFV 232
           AHHL++ A  H ++             NDLPLQPGTAD+VF IC++TD W+L+ KY+K+F
Sbjct: 186 AHHLMSYALEHKNSELMEEVMQLLKTNDLPLQPGTADLVFRICHDTDKWDLLAKYSKKFS 245

Query: 233 MAGVKLRQTSLETWMEFAAKRGDIHSLWKIEKLRSNSMKQHTLITGFSCVKGLLLERKPS 292
            AGVKLR+T+ + WMEFAAKRGD  SLWK++K RS +  QHTL T FSC KG LLE KP
Sbjct: 246 KAGVKLRKTTFDVWMEFAAKRGDTESLWKVDKQRSETYSQHTLSTAFSCAKGFLLESKPE 305

Query: 293 DAVAVIQVLNQTLSDTKKSGIKGELQKLVSEWPLEVIKHQKEEDRKALAASLKSDILVMV 352
           +A AVIQ++ Q   D KKS I  E +KLV+EWP++VIKHQ +ED+KALAASLKS I  MV
Sbjct: 306 EAAAVIQIICQAYPDEKKSAISTEFEKLVNEWPVDVIKHQTDEDKKALAASLKSVIPSMV 365
```

AAB87670.1: *Nicotiana tabacum*

```
 Score =  402 bits (1032), Expect = e-110,   Method: Composition-based stats.
 Identities = 196/322 (60%), Positives = 243/322 (75%), Gaps = 1/322 (0%)

Query: 55  KAFYSSGVGTVEATPSE-DVKELYDKMLDSVKVKRSMPPNAWLWSMIANCKHQPDIRLLF 113
           KA +S+  GTVE++ +   VKELYDKML SV  +RS PPNAWLWS+I +C ++ D+ LL
Sbjct: 83  KASFSTEAGTVESSAATVSVKELYDKMLKSVVEQRSAPPNAWLWSLIQSCANREDVNLLH 142

Query: 114 DILQNLRRFRLSNLRIHDNFNCNLCREVAKACVHAGALDFGMKALWKHNVYGLTPNIASA 173
           DILQ LR FRLSNLRIH+NFNC LC+++ KACV  GA+D G K LWKHNVYGLTPNI SA
Sbjct: 143 DILQRLRIFRLSNLRIHENFNCALCQDITKACVRVGAIDLGKKVLWKHNVYGLTPNIGSA 202

Query: 174 HHLLTNAKNHNDTXXXXXXXXXXXXNDLPLQPGTADIVFSICYNTDDWELINKYAKRFVM 233
           HHLL  AK HND             NDL LQPGTA+IVFSIC+ TD+W+L+ KY KR V
Sbjct: 203 HHLLLFAKQHNDVKLLVEIMKLVKKNDLNLQPGTAEIVFSICFQTDNWDLMCKYGKRVVK 262

Query: 234 AGVKLRQTSLETWMEFAAKRGDIHSLWKIEKLRSNSMKQHTLITGFSCVKGLLLERKPSD 293
           AGVKLR+TSL+TWMEFA+K GD+ +LWKIEK+RS SMK+HTL +G SC K  L++ KP D
Sbjct: 263 AGVKLRKTSLDTWMEFASKIGDVDALWKIEKIRSESMKEHTLASGLSCAKAFLIDHKPGD 322

Query: 294 AVAVIQVLNQTLSDTKKSGIKGELQKLVSEWPLEVIKHQKEEDRKALAASLKSDILVMVS 353
           A A+IQ LNQT+ D+++     ELQKLV++WPLEVIK QK+E RK LAA+L+ DI  M+S
Sbjct: 323 AAAIIQSLNQTIPDSRRQNFMIELQKLVADWPLEVIKRQKDEKRKELAATLQHDIPAMLS 382

Query: 354 ELLSMGLEANVSLEDLDRKEDI 375
            L + GL  +++LEDL RKE +
Sbjct: 383 ALPNRGLNLDINLEDLTRKEGV 404
```

At3g21465: *A. thaliana*

```
 Score =  396 bits (1017), Expect = e-108,   Method: Composition-based stats.
 Identities = 189/319 (59%), Positives = 242/319 (75%), Gaps = 1/319 (0%)

Query: 55  KAFYSSGVGTVEATPSEDVKELYDKMLDSVKVKRSMPPNAWLWSMIANCKHQPDIRLLFD 114
           +A +SS    VE  P+E VKEL+ K+LDSV VKRSMPPNAWLWS+I NC+++ DI  LFD
Sbjct: 67  RANFSSEAKHVE-NPTEAVKELHSKILDSVNVKRSMPPNAWLWSLIDNCRNEDDISFLFD 125

Query: 115 ILQNLRRFRLSNLRIHDNFNCNLCREVAKACVHAGALDFGMKALWKHNVYGLTPNIASAH 174
           +LQNLRRFRLSNLRIHDNFNCNLC++VAK CV  GA++ G +ALWKHNV+GLTP++ASAH
Sbjct: 126 VLQNLRRFRLSNLRIHDNFNCNLCQQVAKTCVRVGAINHGKRALWKHNVHGLTPSVASAH 185

Query: 175 HLLTNAKNHNDTXXXXXXXXXXXXNDLPLQPGTADIVFSICYNTDDWELINKYAKRFVMA 234
           HLL+ A  H D              N+LPLQPGTAD+VF IC++TD+W+L+ KY+K+F  A
Sbjct: 186 HLLSYALKHKDAKLMDEVMKLLKMNNLPLQPGTADLVFRICHDTDNWDLLVKYSKKFCKA 245

Query: 235 GVKLRQTSLETWMEFAAKRGDIHSLWKIEKLRSNSMKQHTLITGFSCVKGLLLERKPSDA 294
           GVKLR+T+ + WMEFAAKRGD  SLW ++KLRS +  QHTL   FSC KG LLE KP +A
Sbjct: 246 GVKLRKTTFDVWMEFAAKRGDTESLWNVDKLRSETYTQHTLSGAFSCAKGFLLEHKPEEA 305

Query: 295 VAVIQVLNQTLSDTKKSGIKGELQKLVSEWPLEVIKHQKEEDRKALAASLKSDILVMVSE 354
            AVIQ++ Q   D KKS ++ E +KLV+EW +++IKHQ E+D+K +AASLKSDI  MV+
Sbjct: 306 AAVIQIICQAYPDEKKSALEAEFKKLVNEWSVDIIKHQNEQDKKDVAASLKSDIPAMVNA 365

Query: 355 LLSMGLEANVSLEDLDRKE 373
           L++ GL   V L +L++ E
Sbjct: 366 LVNSGLRVRVDLNELNKNE 384
```

Os07g0668200: *Oryza sativa japonica*

```
 Score =  348 bits (893), Expect = 4e-94,   Method: Composition-based stats.
 Identities = 174/317 (54%), Positives = 226/317 (71%), Gaps = 2/317 (0%)

Query: 53  LSKAFYSSGVGTVEATPSEDVKELYDKMLDSVKVKRSMPPNAWLWSMIANCKHQPDIRLL 112
           ++   YS+    ++  P+E V ELY KML SV+ + +MPPNAWLWSMI++C ++ DI+LL
Sbjct: 92  MAATHYSTEASDID-QPTEAVVELYQKMLKSVEAE-TMPPNAWLWSMISSCSNKEDIKLL 149

Query: 113 FDILQNLRRFRLSNLRIHDNFNCNLCREVAKACVHAGALDFGMKALWKHNVYGLTPNIAS 172
             ILQ LR FRLSNLRI  NFN +LC +VA+AC   GALD+G+K LWKHNVYG+TP I S
Sbjct: 150 IQILQKLRVFRLSNLRIDANFNDHLCMKVAEACARVGALDYGLKVLWKHNVYGITPTIGS 209

Query: 173 AHHLLTNAKNHNDTXXXXXXXXXXXXNDLPLQPGTADIVFSICYNTDDWELINKYAKRFV 232
           AH+LL +AK  NDT            N +PLQPGTADIVFSICYN D W+L++KYA+RFV
Sbjct: 210 AHYLLKHAKEKNDTKLMGSIMQVLQRNSMPLQPGTADIVFSICYNADRWDLLSKYARRFV 269

Query: 233 MAGVKLRQTSLETWMEFAAKRGDIHSLWKIEKLRSNSMKQHTLITGFSCVKGLLLERKPS 292
            + VKL   S + WM+FAAK GD  S+W I  LR  S+K++ L TGF+CVKG LLERKP
Sbjct: 270 KSEVKLHGASFDIWMDFAAKVGDSQSIWNINSLRGKSVKRYNLATGFACVKGFLLERKPE 329

Query: 293 DAVAVIQVLNQTLSDTKKSGIKGELQKLVSEWPLEVIKHQKEEDRKALAASLKSDILVMV 352

            A A+I++L++   D KK  +  ELQKLV+EWP EVIK QK++DRKAL  +L +DI  M+
Sbjct: 330 SAAAMIKLLHKHSPDEKKQLVTDELQKLVAEWPAEVIKRQKKDDRKALEEALITDIPQMI 389

Query: 353 SELLSMGLEANVSLEDL 369
           S +  + L+ +V+LE L
Sbjct: 390 SSMSKLRLDISVNLEKL 406
```

NP_001030740.1: *A. thaliana*

```
 Score =  306 bits (785), Expect = 1e-81,   Method: Composition-based stats.
 Identities = 143/229 (62%), Positives = 177/229 (77%), Gaps = 1/229 (0%)

Query: 55  KAFYSSGVGTVEATPSEDVKELYDKMLDSVKVKRSMPPNAWLWSMIANCKHQPDIRLLFD 114
           +A +SS    VE  P+E VKEL+ K+LDSV VKRSMPPNAWLWS+I NC+++ DI  LFD
Sbjct: 67  RANFSSEAKHVE-NPTEAVKELHSKILDSVNVKRSMPPNAWLWSLIDNCRNEDDISFLFD 125

Query: 115 ILQNLRRFRLSNLRIHDNFNCNLCREVAKACVHAGALDFGMKALWKHNVYGLTPNIASAH 174
           +LQNLRRFRLSNLRIHDNFNCNLC++VAK CV  GA++ G +ALWKHNV+GLTP++ASAH
Sbjct: 126 VLQNLRRFRLSNLRIHDNFNCNLCQQVAKTCVRVGAINHGKRALWKHNVHGLTPSVASAH 185

Query: 175 HLLTNAKNHNDTXXXXXXXXXXXXNDLPLQPGTADIVFSICYNTDDWELINKYAKRFVMA 234
           HLL+ A  H D              N+LPLQPGTAD+VF IC++TD+W+L+ KY+K+F  A
Sbjct: 186 HLLSYALKHKDAKLMDEVMKLLKMNNLPLQPGTADLVFRICHDTDNWDLLVKYSKKFCKA 245

Query: 235 GVKLRQTSLETWMEFAAKRGDIHSLWKIEKLRSNSMKQHTLITGFSCVK 283
           GVKLR+T+ + WMEFAAKRGD  SLW ++KLRS +  QHTL   FSC K
Sbjct: 246 GVKLRKTTFDVWMEFAAKRGDTESLWNVDKLRSETYTQHTLSGAFSCAK 294
```

CAB78606.1: *A. thaliana*

```
 Score =  222 bits (566), Expect = 3e-56,   Method: Composition-based stats.
 Identities = 108/175 (61%), Positives = 137/175 (78%)

Query: 199 NDLPLQPGTADIVFSICYNTDDWELINKYAKRFVMAGVKLRQTSLETWMEFAAKRGDIHS 258
           NDLPLQPGTAD+VF IC++TD W+L+ KY+K+F  AGVKLR+T+ + WMEFAAKRGD  S
Sbjct: 11  NDLPLQPGTADLVFRICHDTDKWDLLAKYSKKFSKAGVKLRKTTFDVWMEFAAKRGDTES 70

Query: 259 LWKIEKLRSNSMKQHTLITGFSCVKGLLLERKPSDAVAVIQVLNQTLSDTKKSGIKGELQ 318
           LWK++K RS +  QHTL T FSC KG LLE KP +A AVIQ++ Q   D KKS I  E +
Sbjct: 71  LWKVDKQRSETYSQHTLSTAFSCAKGFLLESKPEEAAAVIQIICQAYPDEKKSAISTEFE 130

Query: 319 KLVSEWPLEVIKHQKEEDRKALAASLKSDILVMVSELLSMGLEANVSLEDLDRKE 373
           KLV+EWP++VIKHQ +ED+KALAASLKS I  MV+ LLS GL+ +V L++L++ E
Sbjct: 131 KLVNEWPVDVIKHQTDEDKKALAASLKSVIPSMVNTLLSSGLKVSVDLDELNKNE 185
```

BAC42707.1: *A. thaliana*

```
 Score =  402 bits (1033), Expect = e-110,   Method: Composition-based stats.
 Identities = 193/321 (60%), Positives = 245/321 (76%)

Query: 53  LSKAFYSSGVGTVEATPSEDVKELYDKMLDSVKVKRSMPPNAWLWSMIANCKHQPDIRLL 112
           + ++ ++S    +   P+  VK+L+DKML+SV VKRSMPPNAWLW +I NC++Q DI LL
Sbjct: 66  IPRSSFTSEAEKLAGNPTVTVKDLHDKMLNSVNVKRSMPPNAWLWLLIENCQNQDDIHLL 125

Query: 113 FDILQNLRRFRLSNLRIHDNFNCNLCREVAKACVHAGALDFGMKALWKHNVYGLTPNIAS 172
           FD+LQNLRRFRLSNLRIHDNFNCNLC++VAK CV  GA+D G KALWKHNV+GLTP++AS
Sbjct: 126 FDVLQNLRRFRLSNLRIHDNFNCNLCQQVAKTCVRVGAIDSGKKALWKHNVHGLTPSVAS 185

Query: 173 AHHLLTNAKNHNDTXXXXXXXXXXXXNDLPLQPGTADIVFSICYNTDDWELINKYAKRFV 232
           AHHL++ A  H ++             NDLPLQPGTAD+VF IC++TD W+L+ KY+K+F
Sbjct: 186 AHHLMSYALEHKNSELMEEVMQLLKTNDLPLQPGTADLVFRICHDTDKWDLLAKYSKKFS 245

Query: 233 MAGVKLRQTSLETWMEFAAKRGDIHSLWKIEKLRSNSMKQHTLITGFSCVKGLLLERKPS 292
            AGVKLR+T+ + WMEFAAKRGD  SLWK++K RS +  QHTL T FSC KG LLE KP
Sbjct: 246 KAGVKLRKTTFDVWMEFAAKRGDTESLWKVDKQRSETYSQHTLSTAFSCAKGFLLESKPE 305

Query: 293 DAVAVIQVLNQTLSDTKKSGIKGELQKLVSEWPLEVIKHQKEEDRKALAASLKSDILVMV 352
            A AVIQ++ Q   D KKS I  E +KLV+EWP++VIKHQ +ED+KALAASLKS I  MV
Sbjct: 306 GAAAVIQIICQAYPDEKKSAISTEFEKLVNEWPVDVIKHQTDEDKKALAASLKSVIPSMV 365

Query: 353 SELLSMGLEANVSLEDLDRKE 373
           + LLS GL+ +V L++L++ E
Sbjct: 366 NTLLSSGLKVSVDLDELNKNE 386
```