Heterologous expression of *Triticum aestivum* WRKY transcription factors in *Pichia pastoris*: Interaction with promoters of defence genes

Abdulrahman Mohammed Alhudhaibi

A thesis submitted for the Degree of

Doctor of Philosophy

School of Natural and Environmental Sciences

Newcastle University

United Kingdom

March 2022

Thesis Abstract:

WRKY transcription factors are the largest family of plant transcription factors. They are responsible for the regulation of development, growth, senescence, signal transduction, and stress resistance (biotic and abiotic). Three family groups have been characterised based on the number and location of the WRKY conserved domain. Three wheat WRKY transcription factors candidates, and a mutant variant, were selected for this project (*Ta*WRKY53b, *Ta*WRKY19, *Ta*WRKY3, and *Ta*WRKY3m). Previous work from the Molecular Biology and Biotechnology group has shown the differential expression of these *Ta*WRKY genes following periods of biotic and abiotic stress. The study of these proteins would be greatly accelerated if they could be isolated in large quantities. Isolation from source material is not feasible nor are these transcription factors amenable to heterologous expression in prokaryotic expression systems. Here we show that wheat WRKY proteins can be expressed and purified from the methyltrophic eukaryotic yeast *Pichia pastoris*, to provide functional protein with the capacity to bind DNA.

Candidate *Ta*WRKYs were first studied *in silico*. TaWRKY53b and TaWRKY19 were identified as Group I carrying two WRKY DNA binding domains [DBD] (trans and Cis domains) whereas TaWRKY3 a group IIc contained only one WRKY DBD. All were determined to be localized in the nucleus. Expression of the TaWRKY genes was regulated by bZIP, ethylene-responsive and other transcription factors (Chapter 2). Protein to protein interactions using STRING analyses were less informative due to gaps in current protein databases.

In Chapter 3, wheat WRKY transcription factor (*Ta*WRKY53b, *Ta*WRKY19, *Ta*WRKY3, and *Ta*WRKY3m) protein coding sequences were cloned into the pGAPZaA expression vector. The resulting vector was transformed into *Pichia pastoris* for constitutive protein expression. Each WRKY protein carries a 6x His tag for protein identification and purification. Small-scale expression showed the production of all WRKY proteins. *Ta*WRKY19 and *Ta*WRKY53b were expressed in 5 litre bench-top fermenter in basal salt media (BSM) and proteins were purified using a pipeline of hydrophobic interaction chromatography (HIC) and immobilized metal affinity chromatography (IMAC) column. *Ta*WRKY3 and *Ta*WRKY3m were not compatible with HIC purification and phosphates in the BSM precipitate during IMAC. Replacing BSM with M3 media allow one-step IMAC purification of both *Ta*WRKY3 and *Ta*WRKY3m directly from low pH fermentate (Chapter 4).

Purified recombinant WRKYs were used to demonstrate binding to cis-regulatory elements from plant promoters. Electrophoretic mobility shift assays (EMSA) were performed against synthetic W-box and mutated W-box repeats and against promoter fragments of *Ta*PR1-23 and *Pc*PR1 each containing W-box elements. *Ta*WRKY3 dimerized during EMSA thus reducing the conclusively of the interaction. The mutation in the WRKY DNA binding domain of *Ta*WRKY3m resulted in no W-box binding. *Ta*WRKY19 and *Ta*WRKY53b preferentially bound the promoter fragments (*Ta*PR1-23 and *Pc*PR1) suggesting that addition sequences outside of the core W-box element are required for optimal binding. Recombinant WRKYs did not bind the mutated W-box repeats (Chapter 5)

Overall, this project had demonstrated the use of *P. pastoris* to express soluble, correctly folded recombinant *Ta*WRKY transcription factor proteins. It also showed that these recombinant transcription factors interact specifically with target gene promoter elements. Furthermore, the project highlights the regulatory network controlling the expression of WRKYs in wheat. This approach can now be used to generate functional and highly purified recombinant WRKY transcription factors that can be used to identify the genes they regulate and thus elucidate the specific stress responses and provide targets for wheat breeding strategies.

Acknowledgements

Above all, I am thankful to God the almighty who grants me this unique opportunity to learn and process successfully. This thesis, in its current form, is due to the assistance and guidance of many people.

Foremost, I would like to express my sincere gratitude to my supervisory team **Professor Angharad M. R. Gatehouse** for her helpful suggestions and valuable guidance through my PhD. I would be forever grateful to **Dr Martin G. Edwards**, who was very supportive and patient throughout this very challenging PhD project.

My sincere gratitude to **Dr Vasilios Andriotis** for his endless discussions and debates, and support, along with the many jokes we had together.

I would also deeply thank **Aunt Gillian Davison** for the endless support in both work and personal manners. She had been very supportive throughout since I stepped foot in the Devonshire building. I would like to thank my fellows in the lab at Newcastle University: **Rana Alshegaihi**, **Vector Danladi**, **Feras Afifi**, **Chloe Start**, **Oliva Kent**, and **Robert Calvert,** for the stimulating discussions and meetings.

I would also extend my thanks to my dearest friends, **Gracie fozzati**, her husband **Daniel fozzati**, **Abdulaziz Aldosari**, **Abdullah Alkhudari**, **Fares Alasker**, **Khaled Albisher**, **Abdullah Alharbi and Faisal Alharbi**, **Casie Storrs** and **Sean Storrs**, **Mohammed Sindi**, **Abdullah Faghhi**, **Manar Almasri and Daniah Alqahtani**, for their support and encouragement throughout tough times during my studies and stay at Newcastle. This also can be extended throughout my entire journey here in the United Kingdom.

This thesis and the entire journey wouldn't be possible without the full support of my beloved parents **Maha Abdullah Alarifi**, **Mohammed Abdulrahman Alhudhaibi**, My dear sisters, **Hanadi**, **Jory**, and **Jana**, and my brothers **Abdullah**, **Abdulaziz**, and **Azzam**. I also extend my thanks to my lovely aunt **Aljohara Alarifi** as well. I dedicate this achievement to all of them.

I cannot thank you all above named for the full support and love you have given me. Most tremendous thanks to all which no word can truly express the gratitude and appreciation.

Finally, I would like to express my sincere to my sponsor Imam Mohammed Ibn Saud Islamic University, for providing me with the opportunity to study at Newcastle University, UK, to pursue my PhD studies.

Abdulrahman M. Alhudhaibi

Declaration

I declare that this work has not been accepted for any degree or qualification and is not currently being submitted in candidature for any degree other than the degree of Doctor of Philosophy of Newcastle University.

Abdulrahman M. Alhudhaibi

Contents

List of Figures:

List of tables:

List of Abbreviations:

µl	Microliter
µM	Micromolar
*At*WRKY	*Arabidopsis thaliana* WRKY protein
Biotech	Biotechnology
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BSM	Basal Salts Media
Bt	*Bacillus thuringiensis*
bZIP	Basic leucine zipper
$CaCl_2$	Calcium chloride
CAT	Catalase
cDNA	copy DNA
ChIP	Chromatin immunoprecipitation
DBD	DNA Binding Domain
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
ECL	Electrochemiluminescence
EMSA	Electromobility shift assay
ET	Ethylene
ETI	Effector-Triggered Immunity
FAO	Food Agricultural Organisation
FPLC	Fast Protein Liquid Chromatography
gDNA	Genomic DNA
GE	Genetic engineering
GMO	Genetically modified organisms
H_2O_2	Hydrogen Peroxide]
His	Histidine
His-tag	Hexahistidine-tag
HPLC	High-performance liquid chromatography
HR	Hypersensitive Response

HRP	Horseradish peroxidase
JA	Jasmonic Acid
KCL	Potassium chloride
kDa	Kilo dalton
LB	Lysogeny broth
M	Molar; moles per litre
M3	Modified media 3
MAP kinase	Mitogen-activate protein kinase
MDA	Malondialdehyde
MeSA	Methyl Salicylic Acid (Methyl conjugate of SA)
$MgCl_2$	Magnesium chloride
Mha	Million hectares
ml	Millilitre
mol	Mole
MT	Million Tons
MW	Molecular wright
MWCO	Molecular weight cut off
NaCl	Sodium chloride
NLS	Nuclear localization signal
P. pastoris	*Pichia pastoris*
PAMP	Pathogens-Associated Molecular Pattern
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pH	power of hydrogen. It is a measure of the acidity or basicity of an aqueous solution
PMSF	Phenylmethanesulphonyl fluoride
PR gene	Pathogenesis-related protein gene
PRR	Protein Recognition Receptors
PTI	PAMP triggered Immunity
R proteins	Resistance Proteins
RNAi	RNA interference
ROS	Reactive Oxygen Species
SA	Salicylic Acid

SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis
SOD	Superoxide dismutase
TaWRKY	WRKY gene or protein in wheat
TBS	Tris-buffered Saline
TCA	Trichloroacetic acid
TEMED	N, N, N, N-Tetramethylethylendiamine
TF	Transcription Factors
TILLING	Targeting Induced Local Lesions in Genomes
Tris	Tris (hydroxymethyl) aminomethane.
Tween20	Polyethylene (20) sorbitan monolaurate
Uniport	Universal protein resource
UV	Ultraviolet
V	Volts
VIGS	Virus-Induced Gene Silencing
WHO	World Health Organisation

1. Chapter 1. General Introduction

1.1 Crop production:

The global demand for food is increasing, and one of its key drivers is the increasing world population (Ray *et al.*, 2013). According to the United Nations DESA report in 2015, the world population is expected to reach 9.7 billion by 2050. According to the UN SDG2 (2021) report, two billion people already suffer from food insecurity. Since 2014, this number has been rising annually. The essential goal is to reach Zero hunger (SDG2). Prior to the Covid-19 pandemic, the number of undernourished people globally was 650.3 million. As a consequence of Covid-19, food security has been impacted and the number of undernourished has globally increased and is projected to be between 720 to 811 million people, which is an additional 161 million since 2019 (FAO and Unicef, 2020). Undernourishment has been attributed to food supply chain disruptions, income losses, and price hikes. Approximately one in ten people suffer severely from food insecurity (FAO and Unicef, 2020).

"Food security exists when all people, at all times, have physical and economic access to sufficient, safe, nutritious food to meet their dietary need and food preferences for a healthy and active life" (FAO, 1996). Many factors have an impact on global food production and availability, and this includes climate change, population growth due to land urbanization, increasing demand for energy and water (Ajanovic, 2011). These impacts impose stress on agriculture and global food production, leading to a food crisis. Between 2006 and 2008, the global food commodity prices increased up to 85%. Consequently, basic foodstuff was not affordable to many poorest people in many developing countries. Therefore, this had led to hunger, starvation, and malnutrition. Thus, the concept of food security was brought into strong focus globally, which focusses on global food supply (Pinstrup-Andersen, 2009).

Agriculture is one of the primary resources for human consumption. Ninety percent of food calories are provided directly or indirectly from agricultural land (Ajanovic, 2011). However, there are increasing challenges impacting agriculture resulting in food insecurity. These include accelerated urbanization because of demographic growth (FAO, 2017), degradation and soil productivity reduction as a consequence of extensive exploitation of resources (Tóth *et al.*, 2018), and crop failure or destruction as a result of climate change through extreme events such as drought and floods as well as pests and

diseases (Spence;Hill and Morris, 2020). Such challenges severely affect agricultural sustainability.

There are many challenges affecting food production in the long term, leading to rising agricultural commodity prices. Along with the limitations in the availability of arable lands, crop productivity is affected by urbanisation, climate change, lack of sufficient water, pollution, soil erosion (Godfray *et al.*, 2010; Curtis and Halford, 2014) as well as the increasing number of farmers shifting from using lands for food production to biofuel production in many parts of the world (Ajanovic, 2011). These factors put greater competition on arable lands that strongly impact food agriculture. According to Ajanovic (2011), global biofuel production is continuously increasing due to the increasing demand and political support of many countries to replace fossil fuels with biofuels. Corn, wheat, barley, rapeseed, soybean, and sunflower are considered the preferred feedstock for the first generation of biofuel production. Thus, feedstock's that were mainly used for food and feed production are also used for biofuel feedstock which is considered as a major problem in the future (Ajanovic, 2011; McKenzie and Williams, 2015; Tomei and Helliwell, 2016). Significantly, many reports strongly linked the spike in food commodity prices in 2008 with biofuel production (Ajanovic, 2011). The most likely scenario, would occur in the future is that a higher food production rate is needed from the same land or even loss (Tomlinson, 2013).

1.2 Wheat:

Along with Maize, rice and soybean, wheat is one of the major top four crops grown globally. It is considered to be a primary source of staple food in many regions globally (Ray *et al.*, 2013). Wheat is one of the most abundant sources of energy and protein, with 35% of the world population consuming wheat (Edwards and Jennings, 2018). Among the cereals, wheat contains more calories in the form of both starch and proteins (Bold *et al.*, 2015). Historically, wheat cultivation domestication dated 8000 years ago in Europe, West Asia, and North Africa major civilizations. Agronomic adaptability, Ease of grain storage, and conversion of grain into flour ease are the more likely of their adoption of wheat domestication (Curtis;Rajaram and Gómez Macpherson, 2002). Forty percent of the world's population rely on wheat as a stable source of nutrients supplying 20% of human daily proteins and calories. The protein content of wheat is approximately 13%, which is considered to be higher than other types of cereals. In terms of its nutritional content, it

contains micronutrient and dietary fibre as well as vitamins, minerals, and lipids (Shewry and Hey, 2015).

Ninety-five percent of cultivated wheat is hexaploid, while the remaining 5% is tetraploid durum wheat, which is grown mainly in the Mediterranean region and used for making pasta (Shewry, 2009; Curtis and Halford, 2014). Thus, to keep up with the increasing demand by 2050, it is essential that the world needs to increase its global agricultural food production by 60% - 110% (Godfray *et al.*, 2010; Tscharntke *et al.*, 2012; Ray *et al.*, 2013); this equates to be about 2.4% growth rate of crop production per year (Ray *et al.*, 2013).

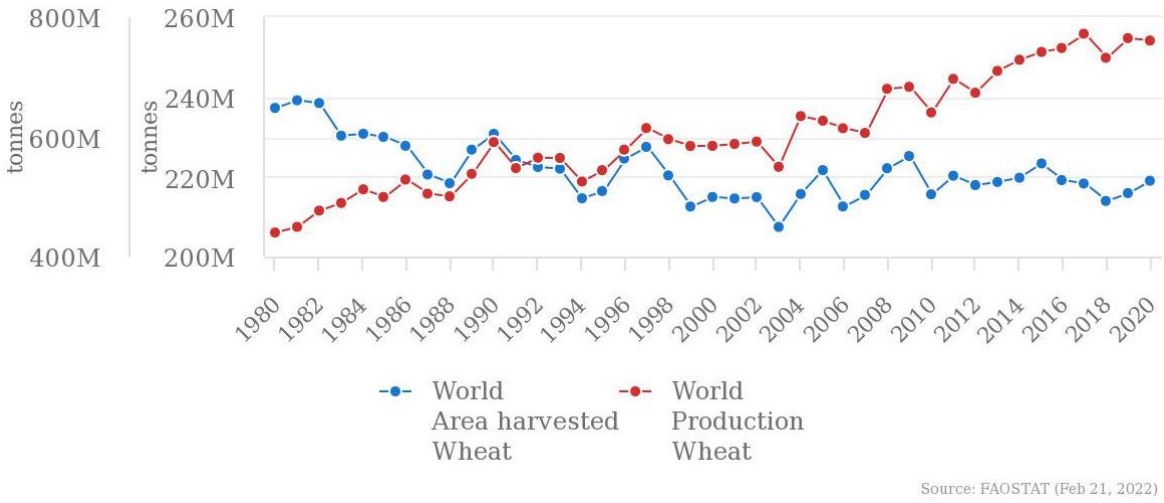


Figure 1-1 Production/Yield quantities of wheat world + total (1980 - 2020). M; Millions (FAOSTAT, 2022)

As one of the most important crops in terms of food security, global wheat production in 1981 was 440 million tons (MT) on about 237 million hectares (Mha). With the increasing demand for wheat production, global wheat production in 2020 increased up to 760 MT with an approximate acreage of about 219 Mha. This shows an increase of 53.3% of wheat production and 7.89% reduction in wheat harvested area worldwide (Figure 1. 1) (FAOSTAT, 2022). With the success of plant breeding e.g genomic selection and marker assisted breeding, wheat high grain yield was achieved in less land cropping area (Voss-Fels *et al.*, 2019)

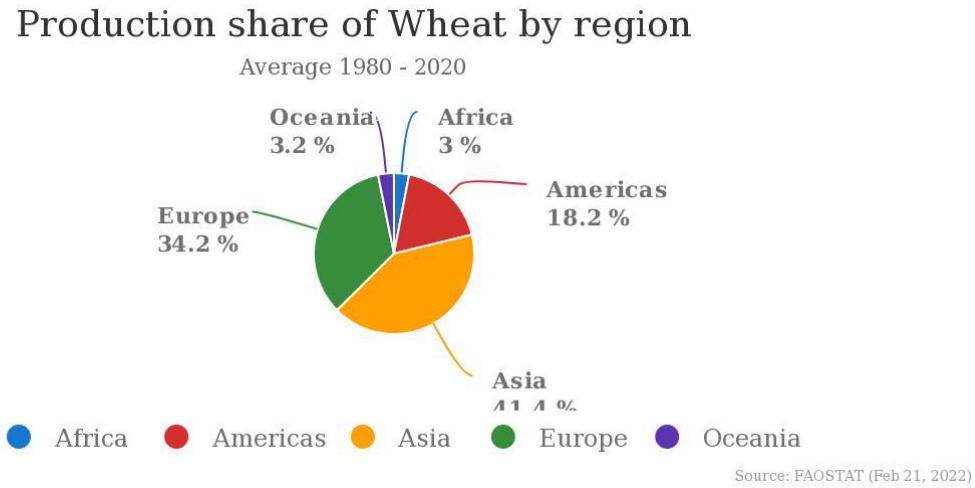


Figure 1-2 Production share of wheat by region (FAOSTAT, 2022)

Asia had the highest wheat production (38.1%), whereas in Europe this was 36.8%, and in the Americas 19%; Oceania and Africa were found to be the lowest wheat producing areas (3.2% and 2.9%, respectively) (FAO, 2022). Whilst there was an apparent increase (59%) in global production, there was, however, a remarkable reduction in the global harvest area (Figure 1. 2) (FAOSTAT, 2022).

1.3 Role of biotechnology to minimise crop loss and increase food yield:

Increasing yield through minimising crop loss is still an important goal in the agricultural industry (Qaim, 2020). Two types of systems are now applied to produce new varieties of crops to reduce the impact of biotic and abiotic stresses, and these are conventional breeding and genetic engineering. Conventional breeding, along with improved farm management practices, have been used since the early 1960s to increase crop productivity for world production (Patnaik and Khurana, 2001). However, the increased impact of biotic and abiotic stresses on food production has led to a continuing agricultural yield loss worldwide. For instance, many insects have evolved resistance to insecticides (Oliveira;Henneberry and Anderson, 2001); in addition, crop productivity has been, and continues to be, constrained by the increasing frequency and severity of extreme weather events (Curtis and Halford, 2014).

Conventional breeding is carried out by the crossing of elite crop lines for desirable traits such as increased yield and disease tolerance (Begna, 2021). Increased productivity has thus been a major focus for plant breeders over the last 100 years (Qaim, 2020). However, the increased use of pesticides, chemical fertilisers, and irrigation water inputs, have also

contributed to an increase in agricultural productivity. Indeed, many high yielding new varieties tend to perform better under highly irrigated conditions and are more responsive to fertilisers, in comparison to traditional landraces, some of elite lines tend to be susceptible to pests and diseases. However, the drawback of conventional breeding, associated with the current agronomic practices, is the considerable environmental effects (Zahoor *et al.*, 2019). Twenty five percent of global greenhouse emissions is associated with the intensity of the overuse of agrochemicals and unsustainable agronomic practices (Qaim, 2020). A major concern is therefor to meet the current demands for increased crop production to feed an ever-growing global population with the current breeding programmes available. The drawback of such programme is the time required to produce a new variety with desired traits for agricultural use. Furthermore, conventional breeding is regarded to have marginal success due to the complexity of stress tolerance traits. Marker assisted breeding had advanced the breeding technology. It is the used of DNA markers for practical selection and breeding with the aim for crop plant genetic improvement. Compared to conventional breeding, it is more effective and efficient in terms of performing plant selection as well as accelerating the progress of breeding. This eventually satisfy crop cultivars changing markets requirements (Jiang, 2015).

As an alternative method for plant breeding, genetic engineering could be used. Since the emergence of genetically engineered (GE) crops in the mid-1990s, many farmers have benefited from this technology through increasing yields accompanied by reduced input of chemical pesticides; and increasingly use of environmentally friendly biopesticides. This technology is considered to be cost-effective (Huang;Pray and Rozelle, 2002). Currently, such crops, also known as biotech crops, are grown over 190 million hectares worldwide (ISAAA Brief; 2019).

Biotechnological approaches can increase the plant or crop gene pool that is available for improvement through the introduction of defined exogenous genes from any organism into the crop's genome (Patnaik and Khurana, 2001). Currently, the level of genetic manipulation in many commercialised GE crops is relatively simple. For example, insertion of genes encoding Crystal (Cry) toxins from the bacterium *Bacillus thuringiensis* (*Bt*) for enhanced resistance to insect pests (Girón-Calva *et al.*, 2020; Lu *et al.*, 2020). With challenges that face crop productivity accompanied by an ever-increasing demand for food production, a combination of desirable traits and the emergence of new technologies and

new traits such as drought tolerance is required to meet the needs of humankind (Godfray *et al.*, 2010; Banerjee and Roychoudhury, 2015).

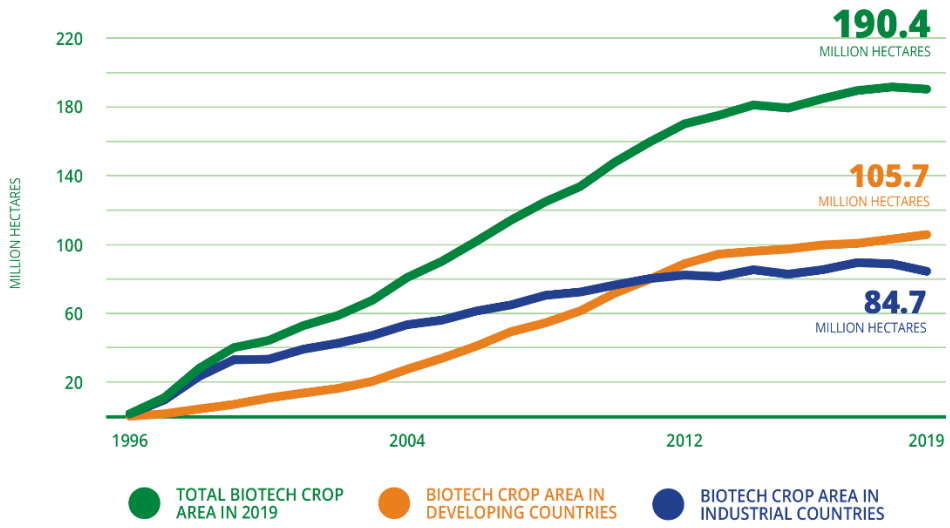


Figure 1-3Biotech crop global area from 1996 to 2019 (ISAAA.Brief.55, 2019).

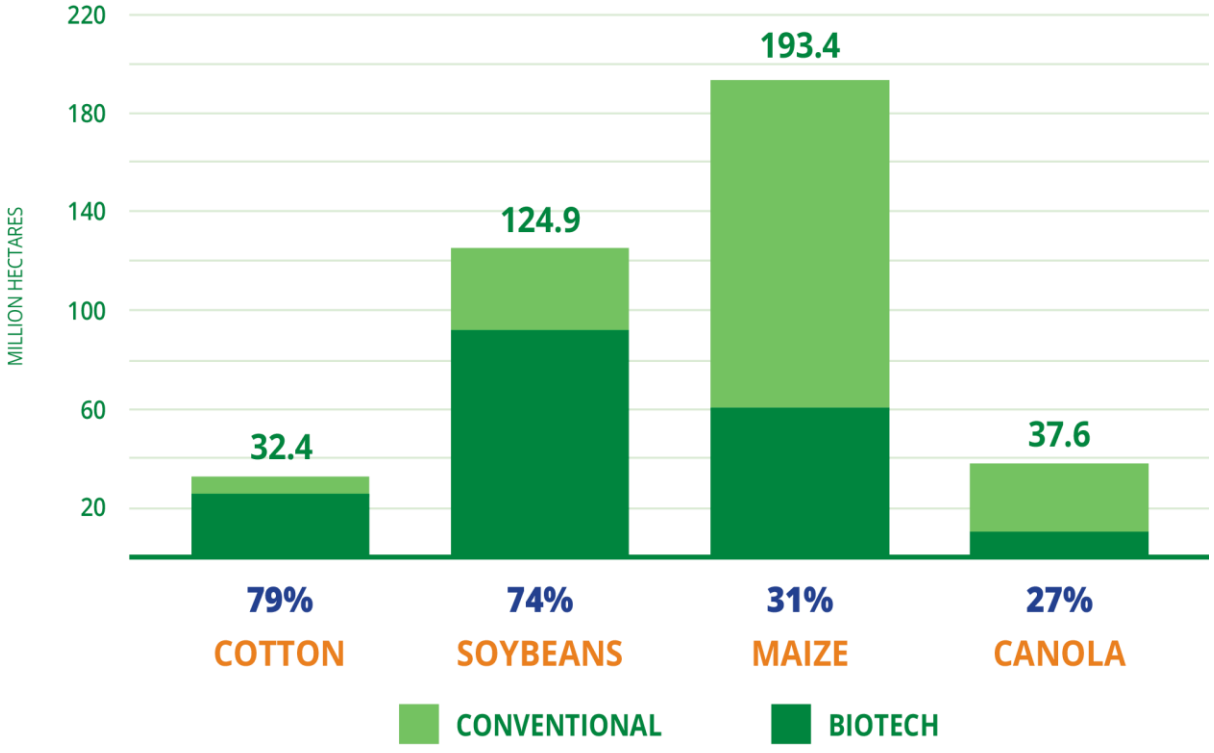


Figure 1-4 Global adoption rates (%) of Biotech crops (ISAAA.Brief.55, 2019)

Biotech crops were first commercialized in the mid-1990s and were then rapidly adopted worldwide. From 1.9 Mha in 1996, biotech crops increased up to 190.4 Mha in 2019, which is about 112-fold increase. Although initially adopted in the industrialised nations, by about 2010 the acreage of such crops was greater in the developing nations than that of the industrialised nations. (105.7 Mha and 84.7 Mha, respectively; Figure 1. 3). Cotton, soybean, maize (corn) and canola are the major biotech crops currently grown, with global adoption rates compared to their conventional counterparts being 79% 74%, 31% and 27%, respectively (Figure 1.4).

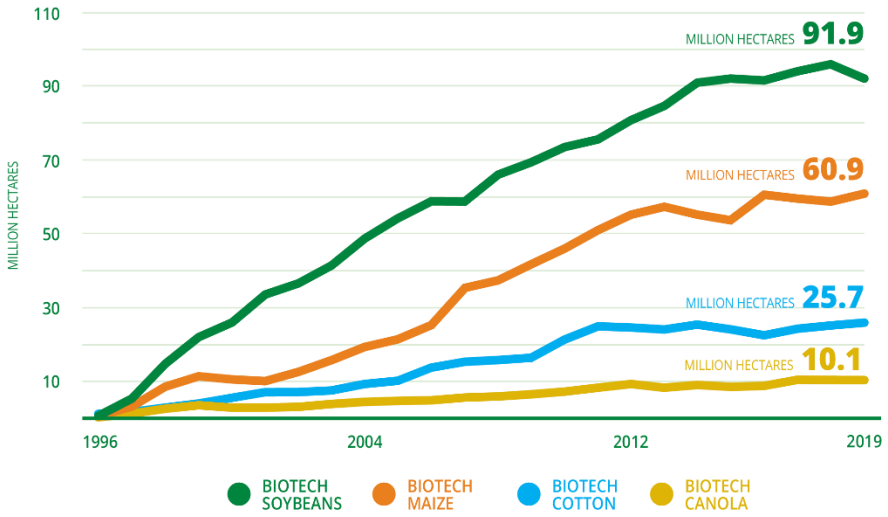


Figure 1-5 Biotech crops global area, by crops, from 1996 to 2019 (ISAAA.Brief.55, 2019)

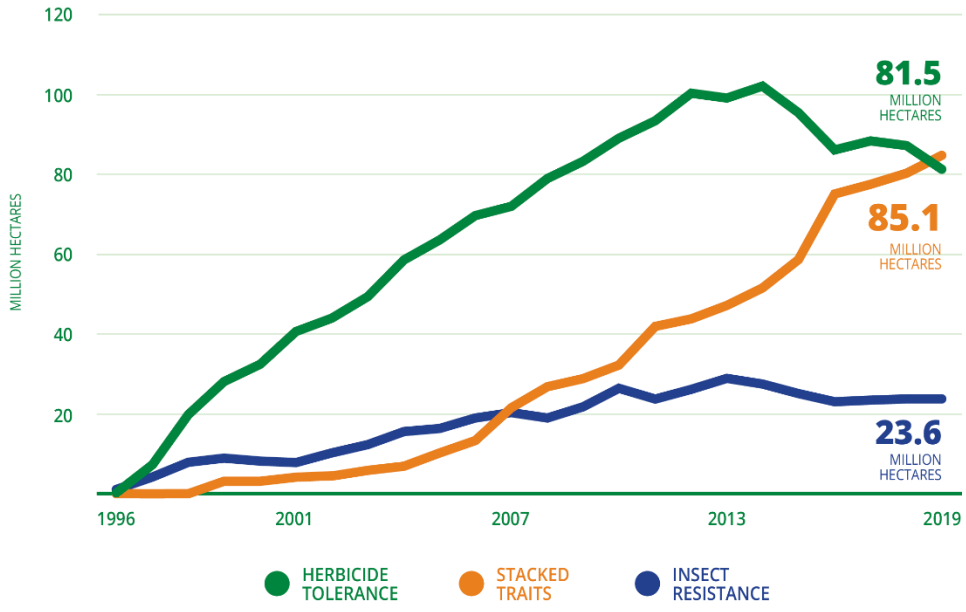


Figure 1-6 Biotech crops global are, by traits, from 1996 to 2019 (ISAAA.Breif.55, 2019)

The global area for biotech soybeans accounted for 91.9 Mha, whereas biotech maize was 60.9 Mha, biotech cotton 25.7 Mha, and canola 10.1 Mha (Figure 1. 5). Since 1996, the two main traits have been herbicide tolerance (HT) and insect resistance (IR), although more recently these two traits have been stacked, and in 2019 represented the greatest acereage (85.1 Mha) (Figure 1.6) (ISAAA.Brief.55, 2019).

Many of biotech companies tend to extensively use transgenic as a tool of transformation of desired foreign gene into crop plants for crop genetic improvement. This is through the use *Argobacterium tumefacaiens*. With this novel traits in crops could be brought to the market in an accelerated manner. The most advanced technique is gene editing such as CRISPR-Cas9 which could be used to delete specific genes at known locations with the aid of engineered nucleases (Georges and Ray, 2017; Anjanappa and Gruissem, 2021)

Many studies over the last 25 years have been conducted to analyse the effects of growing these biotech crops in terms of yield, use of pesticides, and profits to farms. With the use of meta-analysis, the outcome in most situations with their adoption showed benefits to farmers. Considering all biotech crops, yields have increased by 22%, whereas chemical pesticides have been reduced by 37%. The average farm profits from adopting biotech crops increased up to 68%. The benefits of their adoption is more apparent in developing countries than in industrialised countries. Yield increased by 29%, with a 42% reduction in chemical pesticides and a 78% increase in farm profits in developing countries. This contrasts with only an 8% yield increase in the industrialised countries and an 18% reduction in pesticide use with 34% farm profits (Klümper and Qaim, 2014). These difference in the benefits of the adoption of biotech crops between developing and industrialised countries lies with patency. In developing countries, biotech seeds tend not to be patented, leading to seeds prices being lower than industrialised countries (Qaim, 2016).

1.4 Plant defence:

Pathogen-associated molecular patterns (PAMPs) are conserved molecules in the pathogen. After a pathogenic attack, PAMPs are detected by plant protein recognition receptors (PRRs). PRR specific detection of pathogen molecules is what the plant's innate immune system relies on. A signal cascade, passed by Mitogen-activated MAP kinase,

transduce the signal to the nucleus. Then, transcriptional reprogramming occurs in the nucleus, conferring immunity known as PAMP triggered immunity (PTI). To enhance infection and deactivate PTI, pathogens secret protein effectors into host cells. Thus, resistance (R) proteins can be expressed and recognise cytosolic effector molecules resulting in the activation of effector-triggered immunity (ETI) (Dodds and Rathjen, 2010).

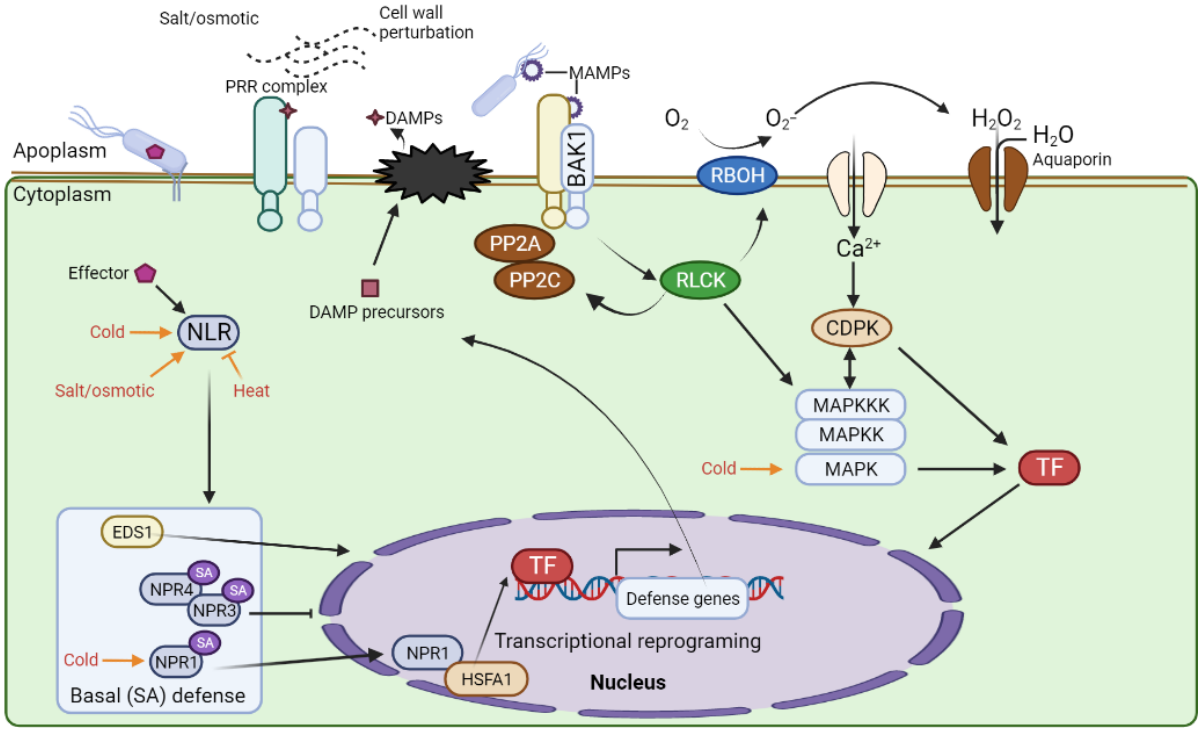


Figure 1.7. A schematic diagram for basic plant immune signalling (Saijo and Loo, 2020).

Phytohormones are another important part of the plant defence system. They are known to regulate resistance responses. "Phytohormones are small molecules which synergistically and/or antagonistically work in a complex network to regulate many aspects of plant growth, development, reproduction, and response to environmental cues" (Denancé *et al.*, 2013). Salicylic acid (SA), Jasmonic acid (JA)/Ethylene (ET) are cellular phytohormones involved in the plant response to biotic and abiotic. They are mainly involved in signalling pathways to facilitate pathogenic resistance. SA was found to regulate, positively, biotrophic pathogen defence. JA/ET, on the other hand, tend to facilitate resistance to both necrotrophic and herbivorous pests. Indeed, there are exceptional circumstances where SA was found to facilitate immune defence against

particular necrotrophic microbes. JA/ET were also found to be essentially triggered by the plant immune system, for some biotrophic pathogen resistance (Glazebrook, 2005).

In response to pathogens, four important signalling pathways are activated. First, the expression of genes involved in the defence response such as pathogenesis-related protein (PR) genes become stimulated by the SA-dependant cascade. These genes encode proteins that localize in the apoplast. The cascade uses both SA and methyl SA (MeSA) which is a SA methyl conjugate (Kombrink and Somssich, 1997). Secondly, Reactive oxygen species (ROS) and Nitric oxide are key players in the second pathway. They tend to accumulate or increase in levels as a consequence of pathogen attack. This promotes a hypersensitive response (HR), synthesis of SA, and the induction of genes involved in the defence-response (McDowell and Dangl, 2000). JA and ET regulate the remaining two pathways. After pathogenic attack, their levels increase, leading to the induction of the expression of defensin and PR genes. These genes encode proteins which are localized in the vascular system (Walling, 2000).

At the cell surface, pattern recognition receptors (PRR) tend to mount the PTI. This is through the detection of molecular structure of microbes as well as damage caused endonously known as Microb – and danger associated molecular pattern (MAMPs, DAMPs). The Leucine –rich-repeat (LRR) receptor like kinases (RLKs) FLS2 and EFR are part of the PRR. They tend to recognise the bacterial flagellin (flg 22 epitope) and EF-Tu (elf18/elf26 epitope), receptively. For fungal chitin-oligomers and bacterial peptidoglycans could be recognised by Lysin-motif (LysM) RLK CERK1 (fig. 1.7) (Couto and Zipfel, 2016).

To prevent microbes and adapted pathogens ligand bound receptors at the apoplastic area contribute for PTI. Ligand bound receptors at the cell was form complexes containing co-receptors/ adapter kinases. Upon activation, a cascade of protein phosphorylation can be triggered though RLK and RLCK. As part of PRR signalling, a hormonal network and extensive transcriptional and translational as well as metabolic reprogramming can also be activated as a consequence apoplastic ROS busts, cytosolic Ca^{2+} as well as Ca^{2+}-dependant protein kinases (CDPKs), and mitogen-activated protein kinase (MAPK) cascades leading to an intensive transcriptional reprogramming (fig. 1.7) (Yu *et al.*, 2017).

Microbial effectors tend to be recognised, directly or indirectly, by intracellular receptors such as nucleotide-binding domain and Leucine rich repeat containing proteins known as NLR. Upon activation of NLR, an amplified form of defence is triggered and it is often involve cell death. The defence signalling pathway can also be activated by NLR that involve EDS1 and different SA-related immune regulators such as NPR4, NPR3, and NPR1 in basal defence (Saijo and Loo, 2020). Bcl-2 homologus antagonist killer protein (BAK1) is a co-receptor which interact with PRR in PTI. It is needed for MAMPs response to pathogens as well as aphid associated microbes which is required for defenses induction (Chaudhary *et al.*, 2014)

Numerous studies elucidating the transcriptional activation of genes that are associated with plant defence responses have been reported, including the role of DNA-binding proteins that bind to the promoter region of target genes involved in the response to biotic attack. Leading to the expression of proteins involved in plant resistance. With regards to crops, many of these studies are aimed at evaluating the potential of these candidate genes for crop improvement. Such candidates can provide either biotic or abiotic tolerance/ resistance to stressors.

1.5 Signal transduction and gene expression:

Genetic transduction has contributed to a better understanding of the plant's physiological responses to its environment regarding abiotic and biotic stresses. Some plant species have a better capability to adapt to unfavourable environmental conditions or biological stresses. They tend to show higher tolerance levels to stress than other species or varieties. This variation, in terms of plant response and tolerance, between species is highly regulated through a complicated network of transcriptional and hormonal crosstalk (Phukan;Jeena and Shukla, 2016).

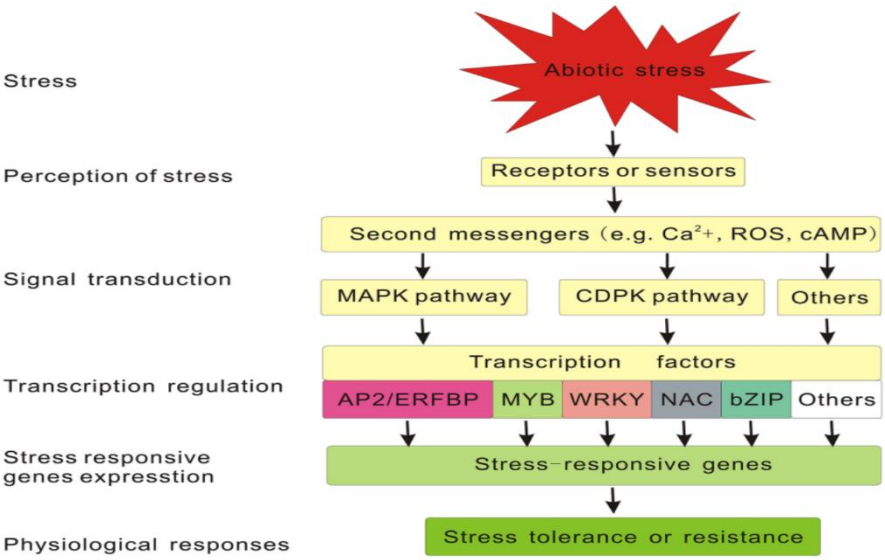


Figure 1-8 Plant abiotic stress signalling pathway (Wang *et al.*, 2016b)

Plant responses to adverse stresses occur at the molecular level. This occurs through the induction of stress-responsive genes, which are activated when stress is recognised and signally transmitted. Simply put, the generic signalling pathway of stress-responsive genes, which in turn are controlled by transcription factors (TFs). At the plant cells' plasma membrane or cell walls, exogenous stress signals are perceived by receptors or sensors. Subsequently, secondary messengers including reactive oxygen species (ROS), cyclic nucleotides (cAMP and cGMP), calcium ions, sugar, and nitric oxide convert these extracellular signals into intracellular ones. Thereafter, the signal transduction pathway is initiated to transmit signals through a signal cascade. During signal transduction, protein kinases and phosphatases mediate a series of molecular events of protein phosphorylation dephosphorylation, respectively. The signal cascade is an important and effective mechanism to transmit signals for the transcription regulation level. At the level of transcription regulation, transcription factors play a crucial role in the stress response. The activation or repression of TFs can be either by protein kinases or phosphatases which occurs at the end of signal transduction as illustrated in Figure 1.8 (Wang *et al.*, 2016b).

1.6 Transcription factors:

A larger portion of the plant genome is devoted to genetic transcription. In Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa*), there are more than 2100 and 2300 genes coding for transcription factors, respectively (Chen *et al.*, 2012; Banerjee and Roychoudhury, 2015). Their role, fundamentally, applies in various aspects in biological

processes in plants which involves growth, development, and environmental factors responses. TFs are employed, in plant cells, to play roles in metabolism, and development, as well as stress response pathways. There are several different families of TFs such as A2P/EREBP, MYB, WRKY, NAC, and bZIP, which are known to play a crucial role in biotic and abiotic stresses (Phukan;Jeena and Shukla, 2016). Generally, TF function exerted by their binding to a defined DNA motif within the regulatory element of their target gene. Their binding can affect the gene expression either positively or negatively. This critically play a role in terms of modulating and fine tuning the plant transcription immune response (Birkenbihl;Liu and Somssich, 2017)

AP2/EREBP TF contain one of the larges member of plant transcription factors. They are identified through their highly conserved AP2/ethylene-responsive element binding factor (ERE) DNA binding domain which interact to GCC box and dehydration-responsive element (DRE)/C-repeat element (CRT) cis- acting regulatory element at the promoter of downstream targeted genes (Riechmann and Meyerowitz, 1998). Their activity in plants is variable which exert on developmental processes such as vegetative and reproductive development as well as cell proliferation. They also act on plants biotic and abiotic stresses and plant hormonal response. The expression of dehydration/cold regulated genes was found to be regulated by DREB by their interaction with DRE/CRT cis-elements (A/GCCCGAC) which is located in the promoter of dehydration/cold genes in response to water deficiency and cold such COR15A, RD29A/COR78, and COR6.6 ((Stockinger;Gilmour and Thomashow, 1997; Liu *et al.*, 1998; Lucas *et al.*, 2011). In response to dehydration, high salinity and heat shock, DREB2 was found to be involved. Overexpression of DREBs had shown an improved tolerance toward dehydration, heat, high salinity. *AtBREB2A* overexpression in transgenic Arabidopsis plants was exhibited an improved tolerance to osmotic stress, and drought. Whereas overexpressed Maize ZmDREB2 had shown an enhanced tolerance to drought in transgenic Arabidopsis. Soybean *GmDREB2* overexpression in Arabidopsis had shown high salinity and drought tolerance.

With regards to ERF subfamily within AP2/EREBP, they are also function in tolerance to plant stress such as high salinity, drought, cold and osmotic stress. This is through their regulation to genes responding to stress by their interaction to cis-regulatory element

GCC box (AGCCGCC). *GmERF3* overexpression in tobacco plants was found to enhance plant resistance against infection as well as exhibiting tolerance to dehydration and high salinity (Zhang *et al.*, 2009).

MYB family form a large family which are widely distributed in plants. They are characterized by MYB, a highly conserved, for DNA binding. One to four imperfect MYP repeats is contained at the N-terminus. However, the activation domain of MYB is located within the C-terminus. MYP TFs have been numerously found to play a function in many plants physiological and biochemical processes such as cell cycle and cell development, metabolism (primary and secondary), signal transduction, synthesis of hormones and biotic and abiotic stresses plant responses (Dubos *et al.*, 2010; Ambawat *et al.*, 2013). *AtMYB15* was found to regulate CBF which exhibited plants tolerance to freezing (Agarwal *et al.*, 2006). An improved tolerance to drought was found by *AtMYB44*, *AtMYB60* and *AtMYB61* by stomatal regulation. In transgenic *Arabidopsis*, soybean *GmMYB76* or *GmMYB177* overexpression enhanced salt and freezing tolerance. (Liao *et al.*, 2008). *AtMYB44* was found to bind directly to the promoter region of *AtWRKY77* leading to the activation of *AtPR1* gene expression as a result salicylic acid activation (Shim *et al.*, 2013).

NAC TF family contain NAC domain, which is a high conserved, in the N-terminal region. In the C-terminal, the transcriptional regulatory region is variable. The DNA binding, nucleus localization and homodimerization or hetrodimerization with other NAC proteins is associated with the NAC domain. The same domain is also function as transcriptional regulation in terms of activation of transcription or repression. NAC proteins interact with NAC recognition sequence (NACRS) with CACG core sequence at the DNA binding motif in their target promoter gene (Olsen *et al.*, 2005). They play role in cell division, flower development, soot apical meristem formation, and responses to biotic and abiotic (Tran *et al.*, 2010). *AtNAC019* from *Arabidopsis* was found to exhibit tolerance to cold (Jensen *et al.*, 2010), while overexpressed rice *OsNAC063* in Arabidopsis improved tolerance to osmotic and salinity stresses (Yokotani *et al.*, 2009). Overexpression of *ZmSNAC1* from *Zea may* in *Arabidopsis* exhibited cold, drought and salinity tolerance (Lu *et al.*, 2012).

bZIP is a basic leucine zipper family protein containing bZIP domain. At the N-terminus, nuclear domain which is highly basic and DNA binding region is contained. At the C-terminus is where the leucine rich motif for dimerization. bZIP, like other transcription factors, play a role in plant developmental processes. It is also play in play stress responses such as high salinity, drought, and cold stresses (Jakoby *et al.*, 2002). Much of studied bZIP TFs had shown that they are induced by ABA and act in regulating the stress-related genes expression in ABA by interacting to ABA-responsive cis acting regulatory elements (ABRE) located at the promoter region (Uno *et al.*, 2000; Zou *et al.*, 2008). *GmbZIP1* overexpression in Arabidopsis exhibited tolerance to drought, cold and high salinity (Gao *et al.*, 2011) while overexpression of TabZIP60 from Triticum aestivum had shown tolerance to salinity, drought and freezing tolerance (Zhang *et al.*, 2015). *OsbZIP71* overexpression in rice plant exhibited a tolerance to drought and salinity (Liu *et al.*, 2014).

TFs are considered as a key regulatory protein (trans-acting proteins) because critical responses are controlled by regulating the transcription of stress-responsive genes. Upon activation of TFs at the end of signal transduction, TFs interact specifically with the promoters of stress response genes (*cis acting*) and regulate their level of transcription, consequently, products of initial downstream of target genes can then participate in different physiological processes which include reactive oxygen species (ROS) scavenging, osmoregulation metabolic synthesis. At the transcriptional level, TFs are regulated by other upstream transcription factors and, in addition, to a series of different levels of modification such as ubiquitination and sumoylation. Therefore, this regulation and modification of TFs form a network of complex regulatory transcription cascades to modulate stress-responsive gene expression which, as a result, determines the activation of both physiological as well as metabolic responses (Liu;Peng and Dai, 2014; Wang *et al.*, 2016b).

Chromatin is a nucleoprotein complex involved in packaging long DNA molecules in eukaryotic genomes. The nucleosome is the first order of chromatin that wraps 146/147 bp of DNA around octamers of histones. This generates a physical barrier to transcription factors to access their target gene. To allow transcription of a target gene, chromatin remodelling occurs disrupting the interaction between histones and DNA. With that

alteration of nucleosome location on the DNA, transcription factors gain access to the promoter region upstream of their target gene. Upon binding of site specific transcription factors onto their cis-regulatory elements upstream of the promoter region of their target gene, general transcription factors and mediator proteins are recruited (Figure 1. 8). This in turn recruits the assembly of RNA polymerase to form a transcription initiation complex. The enhancer sequence is five hundred base pairs in size and located thousands base pairs upstream the target gene transcription start site. It contains multiple activator binding sites for transcriptional regulation. Activator target sequences are also other types of transcription factors. Their binding to the enhancer region initiate DNA bending leading to the interaction with the promoter region bound proteins, thus further ensuring transcription of the target gene (Figure 1. 9) (Spitz and Furlong, 2012; Mobley, 2019; Wan;Marsafari and Xu, 2019).

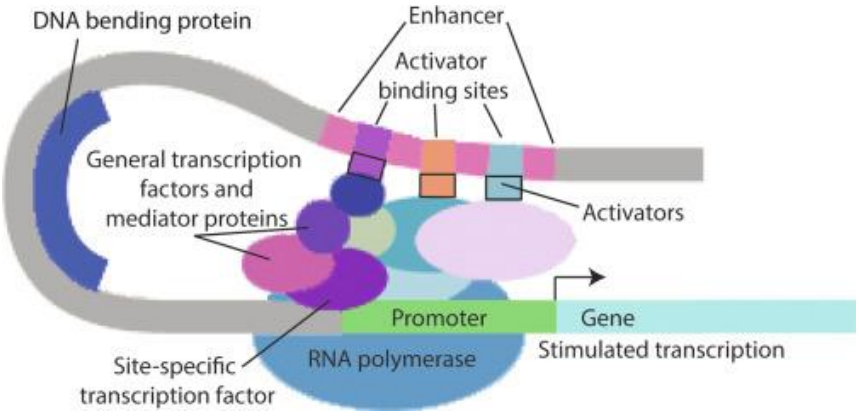


Figure 1-9 Interaction of transcription factors to DNA sequences at the promoter region of target gene transcription stimulation (Mobley, 2019)

1.7 WRKY transcription factors (The WRKY family)

WRKY type transcription factors are regulatory proteins that constitute a large family extremely distributed throughout plants; this family is one of several cellular transcription factors. The name of WRKY is derived from its N-terminus WRKY amino acid sequence (Tryptophan, Arginine, Lysine, and Tyrosine) (Rushton *et al.*, 2012). *Arabidopsis thaliana* contains 74 members of the WRKY family, whereas in rice there are more than 100 members (Ülker and Somssich, 2004). In terms of wheat WRKY, their total number had not yet been entirely explored. To date, only 171 members of wheat WRKY TF (Transcription Factor) have been identified. Homology in wheat WRKYs was associated with their large hexaploid genomic size. Generally, WRKY TFs are involved in abiotic and biotic responses, developmental/ physiological processes including seed coat

and trichrome development, embryogenesis, leaf senescence, hormone signalling, as well as biosynthetic pathways (Chen *et al.*, 2012). Since plants are non-motile, they are subjected to a range of different environmental stresses. Among all TFs, WEKT TFs expression patterns display interconnection, complexity, and flexibility with other cellular components in response to various biotic and abiotic stimuli. Through WRKY TFs, multiple stress responses can be controlled and regulated by genetic manipulation for crop improvement (Franco-Zorrilla *et al.*, 2014; Yu and Zhang, 2021).

The WRKY protein domain is about 60 amino acids and is highly conserved. Some WRKY TFs contain either one or two domains. The N-terminus of the WRKY domain contains a conserved WRKYGQK heptapeptide motif which is considered as DNA binding domain. Whereas at its C-terminus, it contains C2H2 type C2H2 type ($Cx_{4-5}Cx_{22-23}HxH$) or C2HC ($Cx_7Cx_{23}HxC$) zinc finger motif (Eulgem *et al.*, 2000). There are three groups of WRKY TFs which are based on the number of WRKY domains and the structure of their Zinc finger motif. Group I members have Two WRKY domains and C2H2 type zinc finger motif. Group II and III have only one WRKY domain and, C2H2 type for group II and C2HC zinc finger motif for group III (Wang *et al.*, 2016b).

To achieve cellular homeostasis under unfavourable conditions, activated WRKY TFs recognise the W-box (core motif TTGACC/T) located within the promoter of the stress-inducible gene and activate downstream cascade. The presence of the W-box (TTGACT/C) appears to be the minimal consensus requirement for almost all WRKY TFs. The binding affinity of WRKY TFs to cis-acting elements in the W-box is controlled by two motifs present on both N-terminus and C-terminus domains. However, the mode of the binding specificity of WRKY domains is variable among all WRKY TFs. One of which is adjacent DNA sequences outside the TTGACY core motif of the targeted gene which play a role in terms of determining the binding site (Chen et al., 2012). The binding affinity of *Arabidopsis thaliana* (*AtWRKY6* and *AtWRKY11*) is high when G residue is present directly at the 5` elements of the W-box, whereas *AtWRKY26*, *AtWRKY38*, and *AtWRKY43* tend to bind to the same motif as long as it contains T, C, or A residues. Secondly, a cluster of W-boxes can be found at the stress-inducible promoter which plays synergistically during transcription for *PcWRKY1*. Two adjacent W-boxes is required for *HvWRKY38* efficient binding. Thirdly, WRKY proteins can also have an affinity to bind

to non-W box sequences. *HvWRKY46* regulates sucrose in barely can bind to sugar responsive element (SURE) as well as W-boxes. However, *NtWRKY12* which is a tobacco (*Nicotiana tabacum*) WRKY protein binds to only SURE like elements but does not bind to W-boxes. This indicates that the binding affinity for WRKY TFs into the DNA sequence is variable among all WRKY TFs, but each WRKY TF has its specific binding affinity toward the W-box element at their target promoter gene.

A large interconnection network and interaction can define multiple responses regulation. The mediation of various responses toward stress tolerance and plant development had been observed in many WRKY TFs. In terms of biotic and abiotic stresses, *CaWRKY6* in pepper regulates *CaWRKY40* which leads to the regulation of *R. solanacerum* resistance as well as high temperature and high humidity tolerance (Cai *et al.*, 2015). This illustrates a cross talk between WRKYs toward a combined biotic and abiotic stresses tolerance. The SA-signalling was found to be stimulated by *AtWRKY18* and this results to *P. syringae* resistance enhancement. The coexpression of *AtWRKY18* with *AtWRKY40* or *AtWRKY60* can enhance their susceptibility. The plant sensitivity toward high salinity and osmotic stresses could be enhanced by *AtWRKY18* and *AtWRKY60*. However, this enhanced sensitivity can be antagonised by *AtWRKY40* expression. The interaction between *AtWRKY60* and *AtWRKY40* decreases *AtWRKY40* DNA binding ability while *AtWRKY60* and *AtWRKY18* interaction was found to increase *AtWRKY18* binding ability to DNA. A cluster of W-box sequences at the promoter region of *AtWRKY60* was found to be recognised by *AtWRKY18* and *AtWRKY40* which could lead to ABA signalling activation (Xu *et al.*, 2006). A homodimerization and hetrodimerization could be formed between *ThWRKY4* with *ThWRKY2* and *ThWRKY3* which lead to a mediation of various responses to abiotic stresses (Wang *et al.*, 2015a). Both *HvWRKY38* and *HvWRKY1* act as seed germination repressor in barley (Xie *et al.*, 2007). It is apparent that there is a cross-talk between GA and ABA signalling. *OsWRKY51* and *OsWRKY71* were found to be an ABA inducible proteins in which they act synergistically to supress α-amylase expression induced by GA by competing with GAMYB (Xie *et al.*, 2006). With that, there is an apparent cross talk regulation between WRKY transcription factors. It is not limited within the family but it is extended to a wider scope of a cross-talk between WRKYs and other transcription factors to regulate multiple plant developmental and responses toward biotic and abiotic stresses (Phukan;Jeena and Shukla, 2016).

1.8 WRKY transcription factors in response to abiotic and biotic stresses

1.8.1 Abiotic stress:

Drought, salinity, heat, cold, chilling, freezing, light intensity, nutrient, ozone (O_2), and anaerobic stresses are the main forms of abiotic stresses that affect plants and crops (Suzuki *et al.*, 2014). Most of these are a consequence of climate change which hinder the plant's physiological growth and development. For such extreme weather events, it is increased frequency and severity was predicted to be caused by climate change. For instance, Australia and Russia experienced many drought events in the last decade which was due to the lack of rain coupled with high temperature. Curtis and Halford (2014) suggested that it had a significant impact on grain yield production. The negative effect of environmental stress, in terms of extreme temperature and drought, can lead crops grown for commercial production under field conditions to achieve an average yield of 50% of its potential yield (Foyer *et al.*, 2016). Nitrogen stress was also found impact the plant growth (Poll, 2017; Alshegaihi, 2019).

WRKY transcription factors are involved in abiotic, biotic, developmental/ physiological processes including seed coat and trichrome development, embryogenesis, leaf senescence, hormone signalling, as well as regulating biosynthetic pathways (Chen *et al.*, 2012). Since plants are non-motile, they are subjected to a range of different environmental stresses. Among all, WRKY TF expression patterns display interconnection, complexity, and flexibility with other cellular components in response to various biotic and abiotic stimuli. Regarding wheat WRKYs, more such studies have been associated with abiotic stressors rather than biotic stressors. *TaWRKY44* was found to be upregulated in response to salt and drought, abscisic acid (ABA), H_2O_2, and gibberellin. It is localised in the nucleus. Its overexpression in transgenic tobacco was found to confer tolerance to salt and drought. Transgenic lines exhibited a higher survival rate, high relative water content, high soluble sugar, high proline, superoxide dismutase (SOD), catalase (CAT), and peroxidase content. It also showed less ion leakage, lower malondialdehyde content (MDA) content and H2O2 (Wang *et al.*, 2015b). *TaWRKY46* has also been shown to be upregulated in response to drought. It is localised in the nucleus. Overexpression in *Arabidopsis thaliana* has shown an enhancement to osmotic stress (Li *et al.*, 2020b). Yu and Zhang (2021) observed the same findings with transgenic *Arabidopsis*. The survival rate of transgenic plants was increased along with the soluble sugars, proline, superoxide dismutase (SOD), peroxidase, and higher catalase activities, but there was a reduction in malondialdehyde (MDA) and H_2O_2 levels. *TaWRKY10* was upregulated in response to drought, salt, cold and H_2O_2 and is also localised in the

nucleus. Overexpression in transgenic tobacco lines has shown an enhancement in drought and salt stress tolerance as well as increased germination rate, root length survival rate and relative water content under such stressors. Transgenic lines also exhibited high proline content, sugar content and lower reactive oxygen species and MDA (Wang *et al.*, 2013a). *TaWRKY2* and *TaWRKY19* were found to be responsive to drought and salt stress as well as to cold. Both are nuclear proteins. *TaWRKY2* overexpression in transgenic *Arabidopsis* exhibited salt and drought tolerance whereas *TaWRKY19* overexpression conferred tolerance to salt, drought, and freezing stresses. The MDA content was found to be lower in *TaWRKY2* and *TaWRKY19* transgenic lines after salt and freezing stresses. Dehydration responsive binding protein 2A (DREB2A; part of Ethylene-responsive element-binding factor/APETLA2 [ERF/AP2] transcription factor family), salt tolerance zinc finger (STZ), RD29A, RD290B, and cold-regulated (Core6.6) genes were found to be upregulated in response to abiotic stresses in these transgenic *Arabidopsis* lines. *TaWRKY2* was found to bind to the RD29B promoter whereas *TaWRKY19* specifically binds to Cor6.6 and DREB2A. The authors reported that stress tolerance conferred by *TaWRKY19* was due to the function of DREB2A resulting from the activation of RD29A, RD29B, and Cor6.6 (NIU *et al.*, 2012). Further studies on *TaWRKY2* by Gao *et al.* (2018) found that its activity was induced by drought, salt, heat, and ABA. Its overexpression in *A. thaliana* exhibited enhanced tolerance to drought stress. This was due to the higher survival rate and lower water loss rate. Transgenic lines also exhibited higher contents of free proline, soluble sugars, and chlorophyll content. Their findings suggested that it can enhance drought tolerance and increase grain yield in wheat. *TaWRKY1* and *TaWRKY33* were found to be localised in the nucleus and showed responsiveness to abiotic stresses. *TaWRKY1* was upregulated by high temperature and ABA and downregulated by low temperature. *TaWRKY33* gene expression was upregulated in response to high temperature, low temperature, ABA, and jasmonic acids (He *et al.*, 2016) and salt stress (Zhou *et al.*, 2019). Overexpression of both increased germination rates and promote root growth were also observed. *TaWRKY33* transgenic lines exhibited a lower rate of water loss compared to TaWRKY1 transgenic lines. ABA1, ABA2, AB11, AB15 and RD29A expression was upregulated in *TaWRKY1* transgenic, whereas *TaWRKY33* regulates the transcripts of ABA1, ABA2, AB15, DREB2B. *TaWRKY33* overexpressing plants have shown tolerance to heat stress (He *et al.*, 2016). However, further studies conducted on *TaWRKY1* by Ding *et al.* (2016) shows upregulation in response to drought and ABA. Transgenic tobacco with overexpressed *TaWRKY1* conferred tolerance to drought and exhibited greater biomass compared to the

control lines. In addition these transgenic lines also exhibited lower leaf water loss, more osmolyte accumulation and high antioxidant enzyme activities. Furthermore overexpression of this TF caused greater stomata closer upon drought and exogenous ABA treatments. The authors suggested that its overexpression mediated the stomata movement and impacted the leaf water retention capacity. *TaWRKY79* expression was found to be induced by salt and ABA. Overexpression in Arabidopsis showed an enhanced level of tolerance to salinity and ionic stress and reduced ABA sensitivity. Qin *et al.* (2013) suggested that *TaWRKY79* operated in an ABA-dependent pathway. *TaWRKY75-A* was induced by drought and salt stresses and enhanced tolerance to drought and salt was exhibited in transgenic *A. thaliana* overexpressing this particular TF. *TaWRKY93* was found to be induced by salt and ABA and transgenic lines of *A. thaliana* were shown to exhibit tolerance to drought, salt, and low-temperature stresses (Qin;Tian and Liu, 2015; Ye *et al.*, 2021). *TaWRKY70* was induced by cadmium stress in wheat roots and shoot tissues and its role in heavy metal tolerance was verified by overexpression in *A. thaliana,* where the transgenic plants subsequently exhibited tolerance to cadmium. *TaWRKY70* was found to bind to the TaCAT5 promoter which was stated that its transcription was regulated by *TaWRKY70* (Jia *et al.*, 2021).

1.8.2 Wheat WRKY:

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a cytosolic protein and a key enzyme involved in glycolysis. Cytosolic GAPDH (GAPC) plays a role in signal transduction in response to abiotic stress. In Arabidopsis, ROS (Reactive Oxygen Species) interacts with H_2O_2 inducing a signal cascade in which GAPC participates. *TaWRKY40* was found to regulate the expression of *TaGAPC1* through the binding of its promoter region, facilitating tolerance to drought (Zhang *et al.*, 2019). *TaWRKY46*, *TaWRKY92* and *TaWRKY142* were found to confer tolerance to osmotic and salt stresses in transgenic Arabidopsis by the activation of STZ/Zat10 expression (Kuki *et al.*, 2020). RNA-seq conducted by Zhou *et al.* (2019) demonstrated that *TaWRKY13* was upregulated (22-fold) in response to salt stress. The authors also reported that this TF was induced in response to drought, ABA, and cold. Transgenic rice lines with overexpression of *TaWRKY13* have been shown to exhibit enhanced salt tolerance. In Arabidopsis transgenic lines, roots where longer and reported to have a larger surface area. Furthermore, there was an increase in protein content and a decrease in MDA under salt stress. *TaWRKY22* was also reported to be upregulated (20-fold). *TaWRKY41* was found to be downregulated in response to phosphorous deprivation (LI *et al.*, 2014).

1.8.3 Biotic stress:

Under field conditions, plants are also facing the threat of biotic stresses. This includes pathogens infection (bacteria, viruses, fungi, and nematodes) and herbivore pest attacks such as aphids and whitefly (Suzuki *et al.*, 2014; Zandalinas *et al.*, 2021). Globally, it is estimated that 40% of crops are lost due to biotic infection/infestation. Additionally, climate change influences the habitat range of pests and pathogens. Increasing temperatures may lead to the range of many agricultural pest attacks being expanded as well as the capability of pest populations to survive during winter periods would be increased (Schmidhuber and Tubiello, 2007).

Whilst the majority of studies relating to the role of TFs in stress have focussed on abiotic stress, many TFs in wheat have been shown to be associated with biotic stress. For example, *TaWRKY49*, *TaWRKY92*, *TaWRKY112* and *TaWRKY142* have been shown to be upregulated in wheat in response to hybrid necrosis and hybrid chlorosis; all were localised in the nucleus. In transgenic Arabidopsis, *TaWRKY142* enhances *Colletotrichum higginsianum* fungal pathogen resistance through the induction of AtPDF1.2 which is a jasmonic acid pathway marker gene (Kuki *et al.*, 2020). In wheat cultivar ChuanNong 19 (CN19), *TaLHY* is induced by stipe rust; this gene is considered as a disease resistance gene. It is an MYB TF regulated by the induction of salicylic acid. *TaWRKY10* was found to play a role in *TaLHY* expression. The binding *of TaWRKY10* to the *TaLHY* promoter region facilitates the resistance to ChuanNong19 (CN19) to stripe rust pathogen (CYR32). The authors reported that *TaWRKY10* is a stripe rust immune responsive key gene in wheat that binds to the promoter region of *TaLHY* homologs resulting in the regulation of *TaLHY* expression. Interestingly, CN19 susceptible traits to stripe rust were exhibited as a result of *TaWRKY10* silencing (Zhu *et al.*, 2021). With regards to *Puccinia striiformis f. sp. tritici (Pst)*, high-temperature seedling (HTSP) wheat plants exhibited resistance to *Pst* through the upregulation of TaWRKY70 under high temperature. Silencing *TaWRKY70* resulted in these HTSP being susceptible to *Pst.* During high temperatures, infected wheat plants with *Pst* were found to activate salicylic acid and ethylene signalling, resulting in the upregulation of *TaWRKY70* leading to the expression of SA and ET responsive genes such as *TaPR1.1* and *TaPIE1* (Wang *et al.*, 2017). It is important to point out that studies conducted on wheat WRKY responses to biotic stressors are limited. However, this phenomenon is more widely studied in other plants species such as *Arabidopsis thaliana*, rice, and many other species.

1.9 Candidate wheat WRKYs:

TaWRKY53b, *TaWRKY19*, *TaWRKY3* and *TaWRKY3m* were used as candidate wheat WRKY genes. These candidate wheat WRKY genes were selected as a follow up work from previous work conducted by members of gatehouse group. Each of candidate TaWRKY gene showed a differential upregulation under dual stress, abiotic and biotic. Abiotic stress was mainly nitrogen deficiency. Biotic stress was different at each one. *TaWRKY53b* showed an upregulation in wheat plants under nitrogen deficiency and *Zymoseptoria tritic* fungal pathogen (Poll, 2017). *TaWRKY19* had also showed an upregulation in response to nitrogen deficiency and spot blotch infection (Baba, 2019). *TaWRKY3* had also shown an upregulation of its gene expression in response to nitrogen deficiency but with aphid infestation (*Sitobion avenae*) whereas TaWRKY3m from TILLING lines had shown a down regulation under dual stress (Alshegaihi, 2019). On that bases, each of above mentioned WRKY were selected as candidate WRKYs.

1.9.1 *TaWRKY53b*

TaWRKY53b is one of the *TaWRKY53* homolog. To date, two homologs of *TaWRKY53* have been discovered sharing 94% amino acid identity; these are *TaWRKY53a* and *TaWRKY53b* (Chujo *et al.*, 2007; Wu *et al.*, 2008). *TaWRKY53b* belongs to the WRKY group I with the presence of two WRKY DNA binding domains. Its expression is restricted to leaf and crown (Wu *et al.*, 2008).

Phenylalanine ammonia-lyase (PAL) catalyses the deamination of L-phenylalanine to produce trans-cinnamate, which acts as a substrate for the synthesis of several plant secondary metabolites in the phenylpropanoid pathway. Pathogen-derived elicitors induce the expression of PAL genes. These are involved in phytoalexin synthesis which act as antimicrobials, inducing lignification as a means of structural reinforcement, and inducing salicylic acid production (Mauch-Mani and Slusarenko, 1996; Dixon *et al.*, 2002). Van Eck *et al.* (2010) conducted Virus Induced Gene Silencing on resistant wheat lines through the aid of double-stranded RNA for silencing the *TaWRKY53* gene. Their findings indicated that silencing the *TaWRKY53* gene impacted the expression of PAL by suppressing its expression. This resulted in a susceptible phenotypic character in wheat in response to aphid infestation. These authors concluded that this suppression of PAL was due to the silencing of both homologs of *TaWRKY53* and that both genes operate in

the same network of defence response, thus resulting in the susceptibility of transgenic wheat line toward aphid infestation.

In a resistant wheat line, the expression of the *TaWRKY53* gene was upregulated upon *Diuraphis noxia* infestation (Botha;Swanevelder and Lapitan, 2010). Van Eck *et al.* (2014) concluded that *TaWRKY53* regulates apoplastic peroxidase POC1 and Ser/Thr receptor kinase ORK10/LRK10 as downstream targets. This was predicted by the presence of W-box elements in their promoter gene sequence which was verified using electrophoretic mobility shift assay (EMSA). During cereal responses to pathogens both were found to be upregulated. Both also had shown an importance as components of the oxidative burst. This was during the hypersensitive response.

As a way of understanding the function of *TaWRKY53b*, studies conducted on their orthologues could be used. *OsWRKY53* in rice and *AtWRKY33* in *Arabidopsis thaliana* are orthologues of *TaWRKY53*. The Chitin-sensitive MAPK (Mitogen Activated Protein Kinase) signalling pathway was found to upregulate the expression of *AtWRKY33*. Downstream resistance effector genes and MAPK cascades can be connected by their expression. Oxidative stress and SA (Salicylic Acid) were found to induce *AtWRKY33* leading to necrotrophic fungal pathogen resistance. Jasmonic acid and ET- mediated defence signalling was positively regulated by *AtWRKY33*. However, it is considered as a SA mediated response negative regulator. The expression for its pathogenesis induction occurs independently from SA (Zheng *et al.*, 2006; Lippok *et al.*, 2007).

Chitin oligosaccharides have been shown to induce *OsWRKY53*, expression. This stimulates PR (Pathogenesis Related) proteins and peroxidase enzymes expression. *OsWRKY53* was overexpressed in transgenic rice plants and the resulting transgenic lines exhibited enhanced resistance to rice blast fungus *Magnaprthe grisea* and *Xoo*. This indicated its involvement in the basal defence response (Chujo *et al.*, 2007). Both *AtWRKY33* and *OsWRKY53* contain W-box elements at their promoter regions. It is thought that they might be present as a target for their self-regulation or other WRKYs and enhance the activity of their pathogen-specific reactions (Lippok *et al.*, 2007; Chujo *et al.*, 2009).

1.9.2 *TaWRKY19*

TaWRKY19 TF was identified as a group I WRKY containing two WRKY DNA binding domains. In the wheat genome, six *TaWRKY19* homologues have been identified (*TaWRKY19a*, *TaWRKY19b*, *TaWRKY19c*, *TaWRKY19d*, *TaWRKY19e*, and *TaWRKY19f*) (Okay;Derelli and Unver, 2014). NIU *et al.* (2012) reported that overexpression of *TaWRKY19* in *Arabidopsis thaliana* exhibited tolerance to various types of abiotic stresses such as salinity and freezing. In transgenic plants overexpressing this *WRKY*, *DREB2A, Cor6.6, RD29A* and *RD29B* were also upregulated. It was found that *TaWRKY19* facilitates binding to W-box elements in their promoter region. The binding exhibited strong binding to Cor6.6 and DREBA2A promoter regions which was very specific. It was found that DREBA2A constitutive expression induces the expression of stress-responsive genes such as *Cor6.6, RD29.A, RD29B* resulting in Arabidopsis abiotic stress tolerance. Baba (2019) reported its comparative expression between biotic susceptible Iraqi varieties (Rashida) to resistant variety (Latifa). The rate of expression *TaWRKY19* was higher in susceptible variety than resistant ones 72 hours post-infection (hpi). Pathogenesis-related protein 1 gene was shown to be upregulated in response to fungal infection. Under dual stress (nitrogen input severity and spot blotch infection), *TaWRKY19* exhibited down-regulation in the Cadenza variety. TILLING line containing a mutation in *TaWRKY19* at Histidine (H334) to Tyrosine (Y) on the second WRKY DNA binding domain exhibited significantly higher defence against spot blotch compared to Cadenza wheat variety. Additionally, knockdown and/or knockout of *TaWRKY19* was reported to enhance stipe rust wheat susceptibility. This was through the repression of *TaNOX10* transcription, increasing reactive oxygen species (ROS) production for enhanced tolerance (Wang *et al.*, 2022).

1.9.3 *TaWRKY3* and its mutant form.

TaWRKY3 TF was found to be differentially expressed following aphid (*Sitobion avenae*) infestation under nitrogen-limited conditions (Alshegaihi, 2019). According to NIU *et al.* (2012) and (Okay;Derelli and Unver, 2014) *TaWRKY3* transcription factor belongs to group IIc with only one WRKYGQK DNA binding domain. Up to date, few studies have been conducted to elucidate the role of *TaWRKY3* in response to biotic and abiotic stresses apart from the recent study by Alshegaihi (2019). This study had shown its role in reducing aphid's fecundity at low nitrogen input. This aphid species exhibited a

significant reduction in their number when exposed to severe/low nitrogen input in wheat plants. *TaWRKY3* expression was comparatively higher during aphid and low nitrogen input. Furthermore, *TaWRKY3* expression was upregulated in response to drought (He *et al.*, 2016)

TaWRKY3m is a wheat WRKY transcription factor mutated at its DNA binding domain. The mutation occurs within the WRKY itself. The polar Tyrosine (Y160) amino acid was mutated to a negatively charged Aspartic acid (D), forming a WRKD transcription factor. Thus, it was named as *TaWRKY3* mutant (*TaWRKY3m*). This mutation arose by chemically mutating wheat using ethyl methanesulfonate (EMS) ; this line is referred to as TILLING line 1996. This mutation can result in loss-of-function of the gene homolog. This redundancy of genes can be studied to identify alternate responses of mutant wheat plants to stressors, which may show positive phenotypic characteristics and some negative in response to stressors.

Under conditions of low nitrogen, the mutant TILLING line 1996 harbouring a mutation in *TaWRKY3* was significantly more susceptible to aphid infestation compared to wild type plants and Cadenza, suggesting a role for this WRKY in aphid tolerance (Alshegaihi, 2019).

1.10 Genetic homology

From three ancestral grasses (*Triticum Urartu, Triticum turgidum, and Triticum tauschii*), the wheat genome had formed a hexaploid (AABBDD) genome, which contains subgenomes (A, B, and D). This occurs through hybridisation between ancestral grasses. The first genomic hybridisation occurred between *T. Urartu (AA)* and *Ae. Speltoides* (BB) forming *T. turgidum* (AABB). Through time, *T. tauschii* (DD) hybridised with *T. urgidum*, forming *T. spelta* (AABBDD). With domestication and time, *T.spelta* developed into the modern hexaploid wheat crop type known as *T. aestivum* (AABBDD) (Krasileva *et al.*, 2017).

In terms of genetic homology, the hexaploid wheat genome constitutes AABBDD derived from their diploid ancestral grasses with triplicate homologous genes. Thus, the number

of WRKY transcription factors is greater in wheat than found in *Arabidopsis*. Many of the more homologous WRKY transcription factors are found within wheat genomes. *TaWRKY53* transcription factor contains two pairs of homologous WRKYs known as *TaWRKY53-a* and *TaWRKY53-b* sharing 94% amino acid identify (Duan *et al.*, 2007; Wu *et al.*, 2008). Both *TaWRKY53b* and *TaWRKY19* were identified in WRKY group I (Okay;Derelli and Unver, 2014). Regarding *TaWRKY3*, two homologues were also found in the wheat hexaploid genome (*TaWRKY3-a and TaWRKY3-b*) (Okay;Derelli and Unver, 2014).

The size of the bread wheat genome is 16 gigabases (bread wheat) (Walkowiak *et al.*, 2020) which is considered to be massive in comparison to the model plant *Arabidopsis thaliana* (approximately 135 megabases) (TAIR, 2010). Transcription factors tend to have orthologsacross species. Wheat *WRKY53* counterpart in rice is *WRKY53*, and in Arabidopsis is *WRKY33*. The hybridization between wheat-related species allowed bread wheat (*Triticum aestivum*, AABBDD) to have multiple homologs WRKYs within its genome. Thus, the existence of more genomic homologous WRKYs can be presumed.

1.11 Aims and objectives:

The overarching aim is to use eukaryotic host (Pichia pastoris) to produce a full length of recombinant wheat (*Triticum aestivum*) WRKY proteins (*TaWRKY53b*, *TaWRKY19*, *TaWRKY3*, *TaWRKY3m*) and demonstrate function by validating protein DNA binding to promoters of target gene.

Objectives

The specific objectives are to:

1. Identify the transcription regulatory elements of wheat WRKY transcription factors *TaWRKY53b*, *TaWRKY19*, *TaWRKY3*, and *TaWRKY3m* using bioinformatics tools (Chapter 2).
2. Transform *P. pastoris* with WRKY cDNA for their expression using the pGAPZα expression system (Chapter 3).
3. Produce recombinant WRKY proteins by large-scale fermentation of transformed *P. pastoris* followed by purification of His-tagged WRKY proteins using chromatographic methods (Chapter 4).

4. Determine recombinant WRKY - DNA interactions using electrophoretic mobility shift assays (Chapter 5).

2. Chapter 2: *In silico* Candidate *Ta WRKY* Transcription Factors Investigation

2. 1 Introduction

Transcription factors were found to play an essential role in the development of plants and responses to stresses (biotic and abiotic). Their role is to bind to the promoter region of their target gene, therefore regulating expression. WRKY TF also binds to other components of the transcription machinery. WRKY transcription factors are one of many different transcription factors playing a crucial role in plant responses and development. WRKY protein nomenclature was obtained from its highly conserved DNA-binding domain WRKYGQK motif located at the N-terminus of the motif. Variations of this canonical sequence are found in many other plants such as WRKYGEK, WRKYGQK, WRKYGKK, and WKKYGQK (Rushton *et al.*, 2010). A zing-finger motif is located at the C-terminus of the domain, which was found to be crucial for protein structure and DNA-binding functionality. WRKY TFs bind to W-boxes, which are *cis*-regulatory elements (5'-TTGAC-C/T-3') found in the promoter region of their target genes (Ciolkowski *et al.*, 2008a). Upon stress, WRKY proteins can be activated by binding other transcription factors to their *cis*-acting regulatory elements found at the promoter site of the target gene.

In wheat, 171 WRKY transcription factors were identified from whole-genome analysis with the use of Hidden Markov Model (HMM) profile of the WRKY domain (Ning *et al.*, 2017). WRKY transcription factors are known to play a role in biotic and abiotic stresses in many other species such as *Arabidopsis*, rice, and barley (Rushton *et al.*, 2010). Therefore, stress response mechanisms can be understood through the functional characterisation of WRKY transcription factors.

In silico studies were found to be a valuable tool in determining proteins function and characterising interactions (Hassan *et al.*, 2019) . With the help of wheat genome recently updated from international wheat genome sequencing consortium (IWGSC) database, phylogenetic analysis of candidate WRKY proteins would performed, motif determination, as well as mapping DNA-binding residues. Protein structure, orthologues, activity and responses can be determined from protein sequences, and genomic DNA sequences can

provide data on the potential binding and interaction site to facilitate mechanistic studies. Regulatory motif elements within WRKY promoter genes can be obtained using PlantCARE. This scans the upstream promoter region of the gene and locates *cis*-acting elements. Once located the requirements for gene expression can be hypothesised based on the interaction with specific transcription factors.

The rational of this study is that to collectively gather much knowledge of candidate WRKYs and their domains as well as binding specificities. This would allow us to further our understanding of their roles and responses under stress. Understanding candidate WRKYs molecular network as well as predicting WRKY binding sites and collecting also produced information such as DNA sequences, protein solubility and molecular weight that would be needed for successful expression of the protein in later chapters.

This chapter aim is to identify the transcriptional regulatory elements of four selected WRKY transcription factors (*TaWRKY53b*, *TaWRKY19*, *TaWRKY3*, and *TaWRKY3m*). Previous work from the research group has demonstrated that these WRKY genes are differentially regulated in response to biotic (insects and pathogens) and/or abiotic (nitrogen availability and elevated salinity). The following objectives will be achieved to meet this aim and provide a comprehensive understanding of WRKY gene regulation.

- To construct a phylogenetic tree analysis to determine the grouping classification within WRKY family and to also determine their as determining ng their conserved domains.
- To Investigate orthologues of candidate wheat WRKY in other plant species
- To predict subcellular localization of candidate WRKYs within the cell using online bioinformatics tools.
- Elucidating cis-acting regulatory elements at the promoter region of each candidate wheat WRKY using PlantCARE database.
- To investigate protein regulatory maps of each WRKY to predict their molecular interaction using STRING online software.

2. 2 Methodology:

2. 2. 1 Sequences retrieval:

Triticum aestivum WRKY transcription factors DNA and amino acid sequences were retrieved from our library WRKY TF factors database obtained from Lee *et al.* (2015). EnsemblPlants (https://plants.ensembl.org/index.html), UniPort (https://www.uniprot.org/uniprot/A0A0D3QTE8), Panther classification system (http://www.pantherdb.org/) NCBI (National Center for Biotechnology Information) database (https://www.ncbi.nlm.nih.gov/). All sequences were downloaded in FASTA format for local computational analysis.

Chromosomal organisation:

Location of targeted *TaWRKY* proteins within a wheat genome (*Triticum aestvium*) was obtained using EnsemblPlants (https://plants.ensembl.org/index.html). A gene mapping was obtained by a BLAST of protein sequences in Sequence data. The search was against *Triticum aestivum* as well as selecting protein data based using a BLASTX as a search tool.

2.2.2 Protein sequence alignment:

ClustaO protein alignment of all protein sequences (*TaWRKY3*, *TaWRKY3* mutant, *TaWRKY53b*, and *TaWRKY19*). Clustal O was used to generate aliments between input protein sequences. Homologous WRKY protein sequences from different plant WRKY families such as *Arabidopsis thaliana* (*At*), *Glycine max* (*Gm*), *Populus tremula* (*Pt*), *Brachypodium distachyou* (Bd) as a comparative sequence containing WRKY sequence from the same family of candidate wheat WRKYs. This would be used for the follow up anlalysis which are showing conserved domains within protein sequences, relative amino acid abundance, and finally wold be used for phylogeny analysis.

2.2.3 Conserved domains within protein sequences:

Proteins sequences were first converted into FASTA format using open EMBOSS seqret (https://www.ebi.ac.uk/Tools/sfc/emboss_seqret/). Sequences were submitted as unknown format to obtain as a FASTA output format. Converted sequences were saved in a Text file. To obtain the number of conserved domains within protein sequences, amino acid sequences were entered in FASTA format at NCBI search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)

2.2.4 Relative amino acid abundance:

Protein sequence were aligned to give an OUTPUT FORMAT as EMBL-EBI. Aligned protein amino acid sequences were submitted in WebLogo (http://weblogo.berkeley.edu). Multiple logo (symbols per line) were kept at 32 and the maximum up to 50.

2. 2. 5 Phylogenetic analysis:

WRKY protein sequences were submitted to protein BLAST at NCBI to identify homologous proteins (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The resulting hits were exported in FASTA format, and headers were edited so each WRKY protein sequence would have its plant short Latin initial with WRKY followed by its number. To ensure the preservation of sequences, accession numbers of each sequence were saved in a separate text document. Sequences were aligned using the Clustal Omega site (https://www.ebi.ac.uk/Tools/msa/clustalo/), sequences were aligned, and the output format was in Pearson/FASTA format. Aligned sequences were downloaded and saved in a text document which was used in the GBlocks website to remove any divergent informative positions for resulted alignment (http://molevol.cmima.csic.es/castresana/Gblocks_server.html). GBlock aligned sequences were saved and used for sequence format converter (http://phylogeny.lirmm.fr/phylo_cgi/data_converter.cgi) and the output format was selected as Phylip (sequential). The resulting GBlock aligned Phylip were saved into a separate plain text file. SeaView software was used to the phylogenetic tree. To obtain phylogenetic tree on SeaView, PhyML was used with an approximate Likelihood-Ratio Test (aLRT) values.

2. 2. 6 Protein subcellular localisation:

WRKY protein's subcellular localisation was obtained by entering WRKY proteins sequences in Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/#) in a FASTA format. Identification of nucleus signal peptide from protein primary sequence (INSP) had also been used to identify nucleus signal peptide from protein primary sequence (http://www.csbio.sjtu.edu.cn/bioinf/INSP/). cNLS mapper (prediction of importin a-dependant nuclear localisation signals (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) and SeqNLS (nuclear localisation prediction based on frequent pattern mining and linear motif scoring) were also used additionally used for protein subcellular localisation (http://mleg.cse.sc.edu/seqNLS/). DeepLoc – 1.0 eukaryotic

protein subcellular localisation prediction using deep learning (https://services.healthtech.dtu.dk/service.php?DeepLoc-1.0). Protein amino acid sequences were submitted and results were obtained in tables and figures. TaWRKY protein were submitted in FASTA format.

2. 2. 7 Protein promoter regions:

Transcription factors promoter regions were obtained using WRKY proteins DNA sequences in FASTA format. Using ENSEMBLE databases (http://ensembl.gramene.org/Triticum_aestivum/Tools/Blast?db=core), proteins were searched against *Triticum asetivum*. The data was exported as 2000 bp upstream 5' Flanking sequence. The data was saved in Text file. Using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), The 2000 pb upstream region of DNA was submitted and promoter regions of each protein would be received at an email address.

2. 2. 8 Protein regulatory map:

The interaction of TaWRKY proteins (*TaWRKY53b*, *TaWRKY19*, *TaWRKY3*, and *TaWRKY3m*) with other closely related proteins was determined by employing STRING V10.0 (https://string-db.org/) server for predicting TaWRKY protein network interaction. The selection of organism was *Triticum aestivum*. The input amino acid sequence was in FASTA format. The interactive scores were evaluated at a high confidence level (0.700).

2. 3. Results:

2.3.1 Search Candidate wheat WRKY proteins

The protein sequence of each WRKY (*TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and *TaWRKY3m*) was used to Blast in each EnsmblPlants and PlantTFDP database. Each database recorded the WRKYs with different IDs. To assure that both IDs correspond to the same WRKY, protein sequences from each database were aligned using NCBI Blastp. The percentage alignment (Table 2.1) indicates the percentage of shared amino acids between the two sequences. There are minor amino acid variations between proteins retrieved from both databases. However, this variation is little and is not found at the conserved WRKY domain. Additionally, amino acid sequences had been used rather than nucleic acid sequences. This is due to high specificity that comes along with protein sequences than nucleic acid sequences. Each amino acid monomer of a polypeptide chain comprises of three nucleotides. Thus, proteins sequences were employed for the purpose of an increased specificity.

Table 2-1 Description of wheat WRKY transcription factors.

Gene	EnsmblPlants ID	WGSC	Genomic location	Orientation	PlantTFDP ID	GeneBank	% alignment
TaWRKY53b	TraesCS1A02G070400.2	1A:531872744:53189460:-1	1A:53187274-53189460	Reverse	Traes_1AS_F3EAEC435.1	EF368364	99%
TaWRKY19	TraesCS2B02G209200	2B:190968612:190972154:1	2B:190968612-190972154	Forward	Traes_2BS_380EC4D1E.1	EU665430	100%
TaWRKY3	TraesCS2D02G390200	2D:497922168:497926465:-1	2D:497922168-497926465	Reverse	Traes_2DL_4F9F8F1F0.1	EU665432	100%

From **Error! Reference source not found.**, *TaWRKY53b* is located at 1A chromosome and it stretch from 53,187,274 to 53,189,460 with a given 440 amino acids and 1,631 base pairs. The protein sequence retrieved from EnsmblePlant database (TraesCS1A02G070400.2) was aligned with PlantTFDP (Traes_1AS_F3EAEC435.1) and showed 99% alignment. *TaWRKY19* is located at 2B chromosome with a forward orientation. It stretch from 190,966,612 to 190,972,154 forward orientation with a 1,771 bp and 468 aa. The alignment between two sequences had shown 98% alignment (TraesCS2B02G209200 and Traes_2BS_380EC4D1E.1, respectively). *TaWRKY3* is located at 2D chromosome which stretch from 497,922,168 to 497,926,465 reverse orientation with 1.069 bp and 229 aa. The alignment between protein sequences from both

databases had shown 100% alignment (TraesCS2D02G390200 and Traes_2DL_4F9F8F1F0.1, respectively). *Ta WRKY3* mutant was not included in the table because it is *Ta WRKY3* protein backbone with only a mutation within the WRKY conserved domain. The mutation occurs the tyrosine amino acid into aspartic acid and becomes WRKD. Thus, for the purpose of bioinformatic study, *Ta WRKY3* upstream would be used for both *Ta WRKY3* and *Ta WRKY3m*.

2.3.2 Conserved domain analysis:

Candidate WRKY proteins were aligned with amino acid sequence of conserved WRKY domains from different plants species (*Triticum aestivum*, *Arabidopsis thaliana*, *Oryza sativa*, *Barachypodium distachyon*, and *Hordeum vulgare*). TaWRKY53b and TaWRKY19 has shown 2 WRKY conserved domains. Once of which at its amino terminus and the other one is located at the c terminus (firgue 2.2 A and B). TaWRKY3 shows only one WRKY domain at it c-terminus (figure 2.2 A and B).

Figure 2-1 sequence analysis of *Ta WRKY53b*, *Ta WRKY19*, *Ta WRKY3* and *Ta WRKY3m*. Sequence alignment of *Ta WRKY53b*, *Ta WRKY19*, *Ta WRKY3* and *Ta WRKY3m* and their homologous. WRKY domains alignment. Domains of WRKY from different plant WRKY family, including WRKYs from Arabidopsis thaliana (At), *Glycine max* (*Gm*), *Populus tremula* (Pt), *Brachypodium distachyou* (Bd), and *Hordeum vulgare* (Hv). are analyized by ClustalO. Yellow highlight represents 100% similarity of WRKYGQK domain and its zinc finger motif and Green highlight represent amino acids mutation.

Relative amino acid abundance at WRKY domain (Carboxyl-terminal):

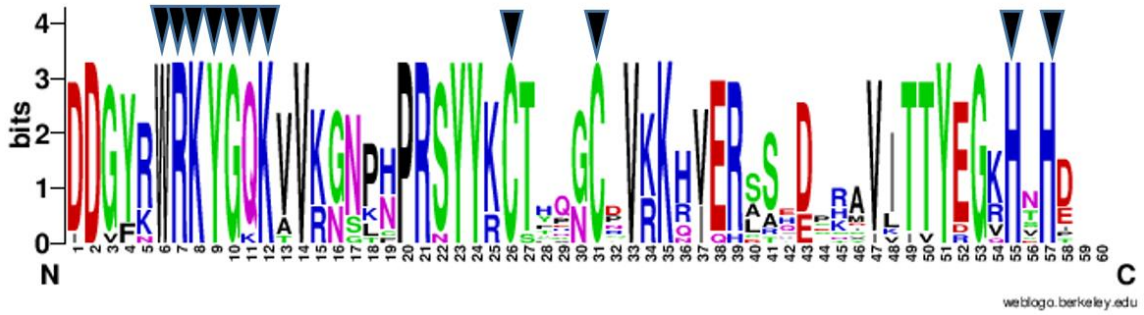


Figure 2-2 WebLogo (https://weblogo.berkeley.edu/logo.cgi) analysis of WRKY domain sequences. Multiple sequences alignment representation of WRKY transcription factors. Black, hydrophobic amino acid; Green, polar amino acids; blue, positively charged amino acid; red, negatively amino acid; purple, neutral; amino acid. Bits (y-axis) represent the conservation of sequence at a position. Triangular labelled used to highlight WRKYGQK domain, two cysteine, and Histidine amino acids. WRKY domains obtained from *TaWRKY53b*, *AtWRKY25*, *OsWRKY30*, *HvWRKY19*, *TaWRKY19*, *BdWRKY3*, *OsWRKY3*, *BdWRKY4*, and *AtWRKY58*.

Approximately 60 amino acid sequence of conserved WRKY domains from different plants species (*Triticum aestivum*, *Arabidopsis thaliana*, *Oryza sativa*, *Barachypodium distachyon*, and *Hordeum vulgare*), list of WRKY protein sequences as phylogenetic tree which contained thirty four protein sequences to have a diversity of proteins from different familites. The higher and a singular a codon position is an indication that it is well conserved. As occurs with WRKY (positions 6-9, Figure 2.3), it appears singular indicating it is well conserved within the forty protein sequences. At these positions, it is a combination of one hydrophobic amino acids (W; Tryptophan), two basic amino acids (R; arginine and K; lysine), and one polar amino acid (Y; Tyrosine). Multiple codons at the same position indicates that this position is less conserved as it occurs in GQK or GKK position (positions 10-12). However, between these selected protein sequences, Glutamine (Q) at position 11 is highly predominant than Lysine (K). So at this specific position, it is more neutral position than basic base. Two cysteine (position 26 and 31) and two histidine (position 55 and 57) in the WRKY amino acid sequences were predicted as dominant among all submitted sequences.

2.3.3 Phylogenetic analysis:

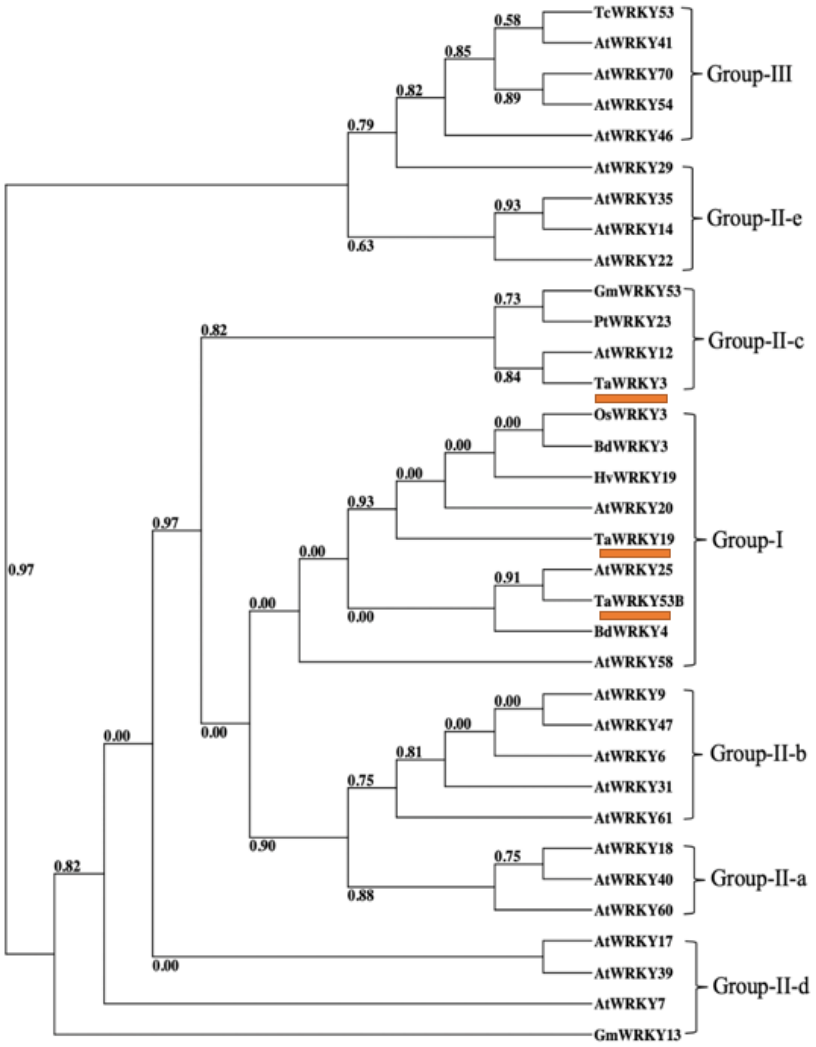


Figure 2-3 Phylogenetic analysis of TaWRKY domains with orthologs in various plants. WRKY amino acids sequences were subjected to PhyML with an approximate likelihood-Ratio Test (aLRT) values. Proteins with one WRKY domain at the C-terminal for Group II and III, and two WRKY domains were at N-terminal and C-terminal for Group I were analysed together. Soybean, *Glycine max* (*Gm*); wheat, *Triticum aestivum* (*Ta*); Rice, *Oryza sativa* (*Os*); Cacao, *Theobroma cacao* (*Tc*); *Populus trichocarpa* (*Pt*); *Arabidopsis thaliana* (*At*); *Barachypodium distachyon* (*Bd*), Barely, and *Hordeum vulgare* (*Hv*).

The predicted molecular weight (MW) and isoelectric point (IPC) has shown a variability among our selected TaWRKYs. Results had shown that the length of *TaWRKY53b* was 440 residues with 47.49 kDa and 7 IPC protein. *TaWRKY19* was 469 residues with 50.79 kDa and 5.73 IPC protein whereas *TaWRKY3* and its mutant form were 229 residues in sequence and 24.97 kDa and 7.65 IPC protein.

WRKY transcription factors were clustered into three major groups (I-III). Group II can also be sub-grouped into five subgroups (IIa- e). A phylogenetic tree was constructed using amino acid sequences from Soybean (*Glycine max* [GM]), Wheat (*Triticum aestivum* [Ta]), Rice (*Oryza sativa* [Os]), cacao (*Theobroma cacao* [Ta]), *Populous trichocarpa* (Pt), *Arabidopsis thaliana* (At), and *Barachypodium distachyon* (Bd) WRKYs. The tree was constructed on the bases on published identified classification of each WRKY sequence to group each of *Ta WRKY53b*, *Ta WRKY19*, and *Ta WRKY3*. *Ta WRKY53b* and *Ta WRKY19* were clustered in Group I in the phylogenetic tree. Both *Ta WRKY53b* and *Ta WRKY19* contain two WRKY domain. One of which at its Amino-terminal domain (178-235 and 199-254, respectively) and the other one is located at its carboxylic-terminal domain (324-381 and 363-420, respectively). This indicates that both genes belong to Group l WRKY transcription factors. The only difference between *Ta WRKY53b* and *Ta WRKY19* was in their zinc finger motif (Cx5 Cx25 HxH and Cx3 Cx22 HxH, respectively). *Ta WRKY3* was found to be clustered into group IIc. Their classification is and/or grouping is based on the number of WRKY domains as well as differences in their hetapeptide and Zinc-finger motif. *Ta WRKY3* and its mutant (indicated as a green colour, Figure 2.2B) contains only one conserved domain at its carboxylic region (151-199) and Cx4 Cx23 HxH at its zinc finger motif (Figure 2. 2 A and B).

2.3.4 Nuclear localisation:

Using Plant-mPLoc and INSP for predicting subcellular localisation of plant proteins. Each protein sequence had been submitted in to FASTA format. Results show that *Ta WRKY53b*, *Ta WTKY19*, and *Ta WRKY3* and its mutant most likely to be located within the nucleus. The nuclear localisation signal by INSP had shown that *Ta WRKY53b* contains 3 merged signals on its protein sequence (36-56, 276-295, and 328-338) (threshold 0.929). In terms of *Ta WRKY3*, two merged sequences (120-146 and 181-190) (threshold 0.929). Finally, *Ta WRKY19* contains only one nuclear localisation signal (195-232) (threshold 0.929) (Table 2. 2)

Table 2-2 Nuclear localisation signal prediction by INSP

Protein	Predicted NLS peptide	Start	END	SCORE
TA WRKY19	DRPADDGYNWRKYGQKAVKGGKYPRSYYKCTLNCPARK	195	232	0.936
Ta WRKY3	WWKGAEKGKMKVRRKMREPRFCFQTRS	120	146	0.94078
	NCRVKKRVER	181	190	0.93127

TaWRKY53b	AERSPRGFNRGGRAGAPKFKS	36	56	0.93269
	DDEADKPETKRRKEHGDNEG	276	295	0.95731
	RWRKYGQKVVK	328	338	0.93259

* The model predicts NLS by the consensus model combined by large-scale frequent pattern mining model and statistical knowledge-based and machine learning SVM-based model with merging (threshold: 0.929).

* Highlighted sequences indicate there is a sharing between two or more prediction tools.

Two other bioinformatic tools had also been used as a confirmatory elements which are cNLS and SeqNLS. Their output results had shown similar results between all. *TaWRKY53b* had shown a distinctive nuclear prediction among all prediction tool with sequences containing PETKRRKEHG amino acids with scores at INSP; 0.957, cNLS; 8, and SeqNLS;0.897. Two sequences were common for *TaWRKY3* across prediction tools, the sequence containing KGKMKVRRKMREPR amino acids had the highest score for *TaWRKY3* with 0.94 at INSP and 0.697 at SeqNLS. However, the NLS peptide region identified by INSP for *TaWRKY19* was not identified by any of the other analysis tools. The highest score was found at INSP with 0.936.

Protein subcellular localisation prediction using TargetP1.1 server which is based on the pre-sequence N-terminal presence predication in different locations within the cell. This includes chloroplast Transit Peptide (cTP), mitochondrial targeting peptide (mTP), secretory pathway signal peptide (SP), or "others" is an indication of other subcellular compartments such as nucleus. When protein sequences had been submitted, plants was selected as an organism for the purpose of predicting chloroplast as a possible location. The cut-off had been selected as a default. The output shows the Reliability Class which is a measure of the size difference between the highest, and the second highest output scores and it is divided into scale of 1 as the highest and 5 as low reliability class (RC). For cTP, mTP, SP, and "other", scores closer to 1 considered to be as strong score. *TaWRKY53b* had shown scores in both cTP and nucleus (0.675 and 0.546, respectively). The RC score for *TaWRKY53b* was 5 which is low RC. *TaWRKY19* scored 0.839 at the nucleus and the RC was 2 which is considered to be high. This aligns with results obtained from INSP (table 2. 2). *TaWRKY3* had shown localisation at the nucleus (0.414) and the RC was 4 which is considered to be low (Table 2. 3)

Table 2-3 TargetP-1.1 subcellular location of protein prediction.

Protein	Length	cTP	mTP	SP	Other	loc	RC
TaWRKY53b	429	0.675	0.018	0.014	0.546	C	5
TaWRKY19	465	0.113	0.119	0.074	0.839	-	2
TaWRKY3	229	0.187	0.086	0.073	0.414	-	4

SignalP-5.0 online server had also been used to test the likelihood of proteins to have a signal peptide and cleavage sites in eukaryotic amino acid sequences. This predicts the signal peptides presence and the location of their cleavage site in proteins. Sec/SPI indicate a standard secretory peptide transport by the sec translocon and cleaved by signal peptidease I (Lep). Table 2. 4 shows the signal peptide prediction of *TaWRKY53b*, *TaWRKY19*, and *TaWRKY3* proteins. The prediction tool result indicate a very week result for all TaWRKY proteins. Strong signal is 1.0. *TaWRKY53b* was 0.0015 sec/SPI, *TaWRKY19* was 0.0032 sec/SPI, and *TaWRKY3* was 0.0015 sec/SPI. These results shown very low number which was far from 1.0. This indicates that there is no signal peptide in each of TaWRKY protein. The signal peptide prediction from taregtP-1.1 and SignalIP had very similar results. Both results that each of TaWRKY proteins had no signal peptides on their protein sequence. On the other hand, all of test proteins (*TaWRKY19*, *TaWRKY53b* and *TaWRKY3*) had shown a high score at other.

Table 2-4 SignalIP - 5.0 for predicting signal peptide and cleavage sites.

Protein	Signal Peptide (sec/SPI)	Other
TaWRKY53b	0.0015	0.9985
TaWRKY19	0.0032	0.9968
TaWRKY3	0.0015	0.9985

Deeploc – 1.0 was used for prediction of eukaryotic protein subcellular localisation using deep learning. This tool can differentiate proteins prediction from 10 different localisations (nucleus, cytoplasm, extracellular, mitochondrion, endoplasmic reticulum, Chloroplast, Golgi apparatus, Lysosome/vacuole and peroxisome). This to further scan each protein using deep learning. Results had shown that each of Wheat WRKY protein (*TaWRKY53b*, *TaWRKY19*, and *TaWRKY3*) were primarily localised in the nucleus (0.9873, 0.9986, and 0.9977, respectively) (Figure 2. 4). Very weak results subcellular localisation prediction were found in the cytoplasm (*TaWRKY53b*; 0.0126, *TaWRKY19*; 0.0013, and *TaWRKY3*; 0.0022) (Table 2. 5). In terms of their solubility prediction using

Deeploc – 1.0, TaWRKY proteins were found to be soluble protein (*TaWRKY53b*; 0.789, *TaWRKY19*; 0.6813, and *TaWRKY3*; 0.815) (Table 2. 6).

To predict a eukaryotic protein subcellular localization, Deeploc-1.0 uses neural network algorithm. It is trained on Uniport proteins with experimental evidence of subcellular localizations. To preform prediction, it rely on protein sequence information. The accuracy rate is 78% for 10 categories and 92% for membrane bound or soluble proteins.

In comparison, TargetP1.1 uses N-terminals sequence information to discriminate protein target in the cell (mitochondria, chloroplast, secretory pathway, and "others" localizations). The success rate is 85% for plant proteins and 90% for non-plant proteins.

Table 2-5 protein prediction using DeepLoc – 1.0 for their subcellular localisation using deep learning.

TaWRKY	Nucleus	Cytoplasm	Mitochondrion	Endoplasmic reticulum	Golgi apparatus	Plastid	Cell membrane	Extracellular	peroxisome	Lysosome / Vacuole
TaWRKY53b	0.9873	0.0126	0	0	0	0	0	0	0	0
TaWRKY19	0.9986	0.0013	0	0	0	0	0	0	0	
TaWRKY3	0.9977	0.0022	0.0001	0	0	0	0	0	0	0

Table 2-6 6 DeepLoc - 1.0 protein solubility predication.

TaWRKY	Soluble	Membrane
TaWRKY53b	0.789	0.211
TaWRKY19	0.6813	0.3187
TaWRKY3	0.815	0.185

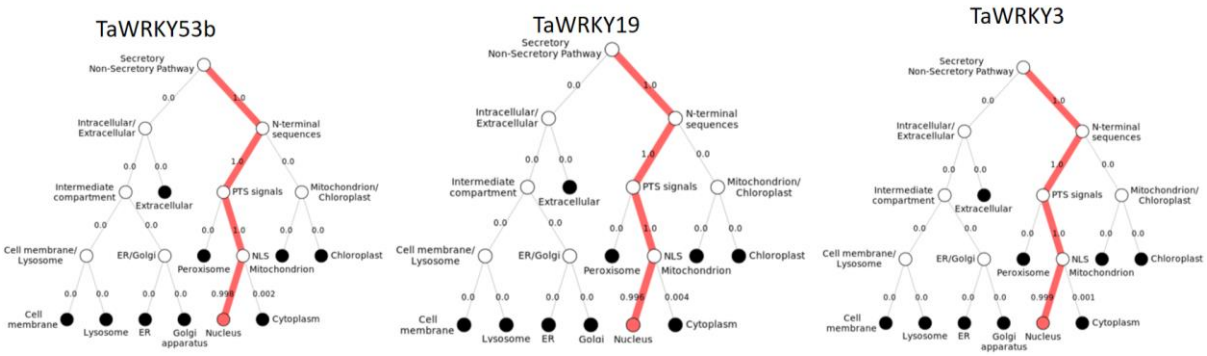


Figure 2-4 DeepLoc presentation of protein secretory/ non-secretory pathway in eukaryotic cell. *TaWRKY53b*, *TaWRKY19*, and *TaWRKY3*. Redline indicate the protein pathway in the cell.

2.3.5 Arabidopsis Ortholog:

Table 2-7 Identity of orthologous sequences in *Arabidopsis*. TaWRKY protein sequences were compared with BlastP to entries in the TAIR database (https://www.arabidopsis.org/Blast/).

TaWRKY	Arabidopsis counterpart	TAIR ID	Ensemble ID%
TaWRKY19	*AtWRKY3*	AT2G03340	83.0%
TaWRKY53b	*AtWRKY26*	AT5G07100	82.9%
TaWRKY3	*AtWRKY12*	AT2G44745	85.7%

Ensemble ID percentage is a percentage of protein homology between the query (as wheat protein) and the subject (as Arabidopsis protein). The higher the percentage the more homologous that protein become. Thus, the highest percentage had been selected for each protein orthologue. As it shown in Table 2. 7, *AtWRKY3* is an orthologue of *TaWRKY19* with 83.0% homology. *TaWRKY53b* is orthologous to *AtWRKY26* (82.9%) and *TaWRKY3* to *AtWRKY12* was 85.7% which is the highest among all (Table 2. 7).

2.3.6 Transcriptional regulation via protein:protein interactions

The transcriptional regulatory map will show the interaction between candidate wheat WRKY transcription factors and other plant transcription factors.

2.3.6.1 *TaWRKY19*:

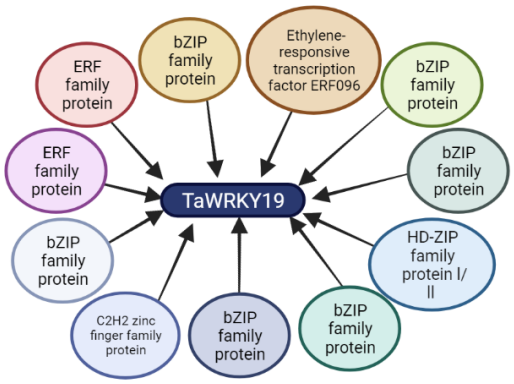


Figure 2-5 Illustrates the upstream regulator of *TaWRKY19*.

Using PlantRegMap/PlantTFDB (v5.0.) *TaWRKY19* Blast has shown a highest result to Traes_2BS_380EC4D1E.1 (Figure 2. 5) (GeneBank: EU665430). The protein properties showed its length is 468 amino acid (aa), molecular weight is 50.688 kDa, and Isoelectric point (IP) is 6.2328. It contains two WRKY conserved WRKY domains amino terminus WRKY domain and carboxyl terminus WRKY domain (199-254 aa and 362-420 aa, respectively). Outcome results showed that its functional description by Uniport as a Transcription factor interacts specifically with the W-box (5'-(T)TGAC(CT)-3'), a frequently occurring elicitor-responsive cis-acting element. The regulation of this protein, retrieved by Uniport, is an induction by salicylic acid and during leaf senescence. Its counterpart in *Arabidopsis thaliana* is *AtWRKY3* (AT2G03340.1).

The transcriptional regulatory map STRING had shown a network involved 13 proteins transcriptionally regulating *TAWRKY19* (Figure 2. 5). Six of the interacting proteins belong to bZIP family proteins which are considered as a set of protein transcription factors playing critically in the same way as WRKY transcription factors such as plant growth, development, and abiotic and biotic stress responses. bZIP proteins interacting with *TaWRKY19* had shown a wide verity of functions and regulations.. The overall shows that all interacting bZIP family proteins to *TaWRKY19* involved in many processes including plant development, stresses responses to either biotic and/or abiotic. There are only one C2H2-zinc finger family protein and two ERE family proteins function as a transcriptional activator that bind to GCC-box pathogenesis-related promoter element. These results indicate that this protein might be involved in the regulation of gene

expression by stress factors and by components of stress signal transduction (Appendix E. Table 2. 8).

2.3.6.2 *TaWRKY53b:*

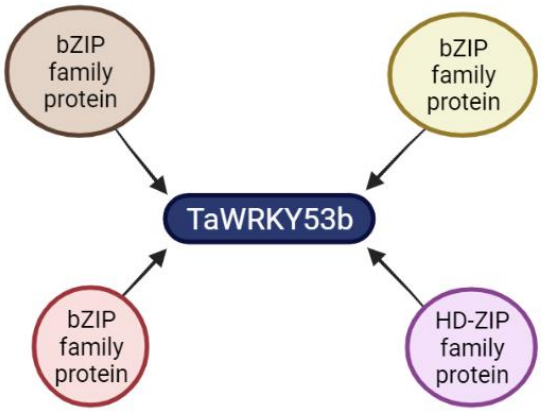


Figure 2-6 Illustrates the upstream regulator of *TaWRKY53b.*

TaWRKY53b STRING had shown a highest result to Traes_1AS_F3EAEC435.1. The protein properties showed its length as 440 aa, MW: 47.54 kDa, and IP: 7.9393. It contains conserved WRKY domains one of which at its amino terminus and the second is located at the WRKY carboxylic terminus (178-237 and 323-381, respectively). Outcome results showed that its functional description by Uniport as a Transcription repressor (By similarity) (Figure 2. 6). Interacts specifically with the W box (5'-(T)TGAC[CT]-3'), a frequently occurring elicitor-responsive cis-acting element. Negative regulator of both gibberellic acid (GA) and abscisic acid (ABA) signalling in aleurone cells, probably by interfering with GAM1, via the specific repression of GA- and ABA-induced promoters (By similarity) ECO:0000250|UniProtKB:Q6IEQ7, ECO:0000250|UniProtKB:Q6QHD1. Protein regulation, retrieved by Uniport, as an induction which is Induced by abscisic acid (ABA) in aleurone cells, embryos, roots and leaves (PubMed:25110688). Slightly down-regulated by gibberellic acid (GA) (By similarity). Accumulates in response to jasmonic acid (MeJA) ECO:0000269|PubMed:16919842, ECO:0000269|PubMed:25110688. Annotation by nucleotide showed that it is a *Triticum aestivum* WRKY transcription Factor (*WRKY53b*) mRNA, completed cds (GeneBank; EF368364). *TaWRKY53b* counterpart *in Arabidopsis thaliana* is *AtWRKY26* (AT5G07100.1).

In terms of *Ta WRKY53b*, four bZIP proteins involved within its transcriptional regulatory map. These proteins mainly function as transcriptional regulators involved in defence, plant signalling, and for some might be induced by drought stress (Appendix E.Table 2. 9)

2.3.6.3 *Ta WRKY3*:

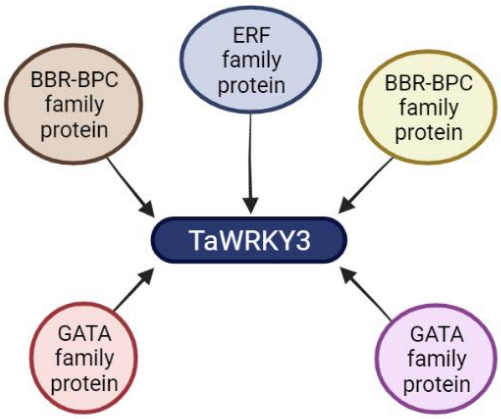


Figure 2-7 illustrates the upstream regulator of *Ta WRKY3*.

STRING had shown a highest result to Traes_2DL_4F9F8F1F0.1. The protein properties showed its length as 229 aa, MW: 24.963 kDa, and IP: 8.5563. It contains a one conserved WRKY domain located at its carboxylic terminus (151-209). Outcome results showed that its functional description by Uniport as a transcription factor which Interacts specifically with the W box (5'-(T)TGAC[CT]-3'), a frequently occurring elicitor-responsive cis-acting element (By similarity). Nucleotide annotation that it is a *Triticum aestivum* WRKY transcription Factor (*WRKY3*) mRNA, completed cds (GeneBank; EU665432). *Ta WRKY3* counterpart in *Arabidopsis thaliana* is *At WRKY12* (AT2G44745.1).

The transcriptional regulatory map of *Ta WRKY3* showed that there are 5 proteins involved (Figure 2. 7). Two BBR-BPC family proteins functioning as a transcriptional regulator that specifically bind to GA-rich elements (GAGA-repeats) present in regulatory sequences of genes involved in developmental processes. Two GATA family protein functioning as a transcriptional activator that specifically bind to 5'-GATA-3' or 5'-GAT-3' motifs within gene promoters and its reaction is an induction by abscisic acid (ABA), and drought and salt stress. It is down-regulated by Jasmonate and wounding. One ERF family protein function as probability of acting as a transcriptional activator through binding to GCC-boc pathogenesis-related promoter element. It might be involved in the

regulation of gene expression by stress and by components of stress signal transduction (Appendix E.Table 2. 10).

2.3.7 Protein-protein interaction:

Protein to protein interaction of all selected TaWRKY proteins was done by STRING V10.0 (https://string-db.org/) online server. The run was done by two methods. One of which was by protein through its sequences. This method looks at a specific protein's interaction with the protein of interest. *TaWRKY19* amino acid sequence was used as an input and *Triticum aestivum*. The output results had shown that this protein ID is Traes_2BS_380EC4D1E.1. Its annotation was uncharacterised protein with identity of 97.9%. It had been found that there are 10 proteins interact to *TaWRKY19*. The minimum score of interaction was 0.561 and the maximum was 0.675. Two proteins were found to be small ubiquitin-related modifier (Traes_3B_BC6EF5032.1 and Traes_3B_E077414F2.1; score 0.661) and two proteins belong to the thioredoxin family (Traes_2AL_0ADBF3D27.1 and Traes_2BS_9C3ACD499.1; score 0.561). The remaining of interacting proteins were found to be uncharacterised proteins. *TaWRKY53b* protein sequence was also run using the same method, the results showed that the protein is Traes_1AS_F3EAEC435.1 and its annotation was uncharacterised protein with identity of 96.7%. 10 proteins were found to highly interact with *TaWRKY53b* at a score 0.938 as a minimum. There were 3 MAPK interacting proteins (Traes_4BL_2CEFDE904.1 named as MAPK, Traes_4BS_9285C0809.1 and Traes_4DL_15045954F.1). Only one WRKY transcription factor interacts with *TaWRKY53b* which is *TaWRKY27* (Traes_3B_990298FF5). The remaining of proteins were defined as uncharacterised proteins.

The final sequence was *TaWRKY3* and its result showed that it is identified as Traes_2DL_4F9F8F1F0.1 and there were 10 highly interacting proteins. Two of which were identified as they belong to cytochrome P450 family proteins (score: 0.507) (Traes_2BS_F8ED79291.3 and Traes_7DL_33BB5BE33.3) (Figure 2. 8).

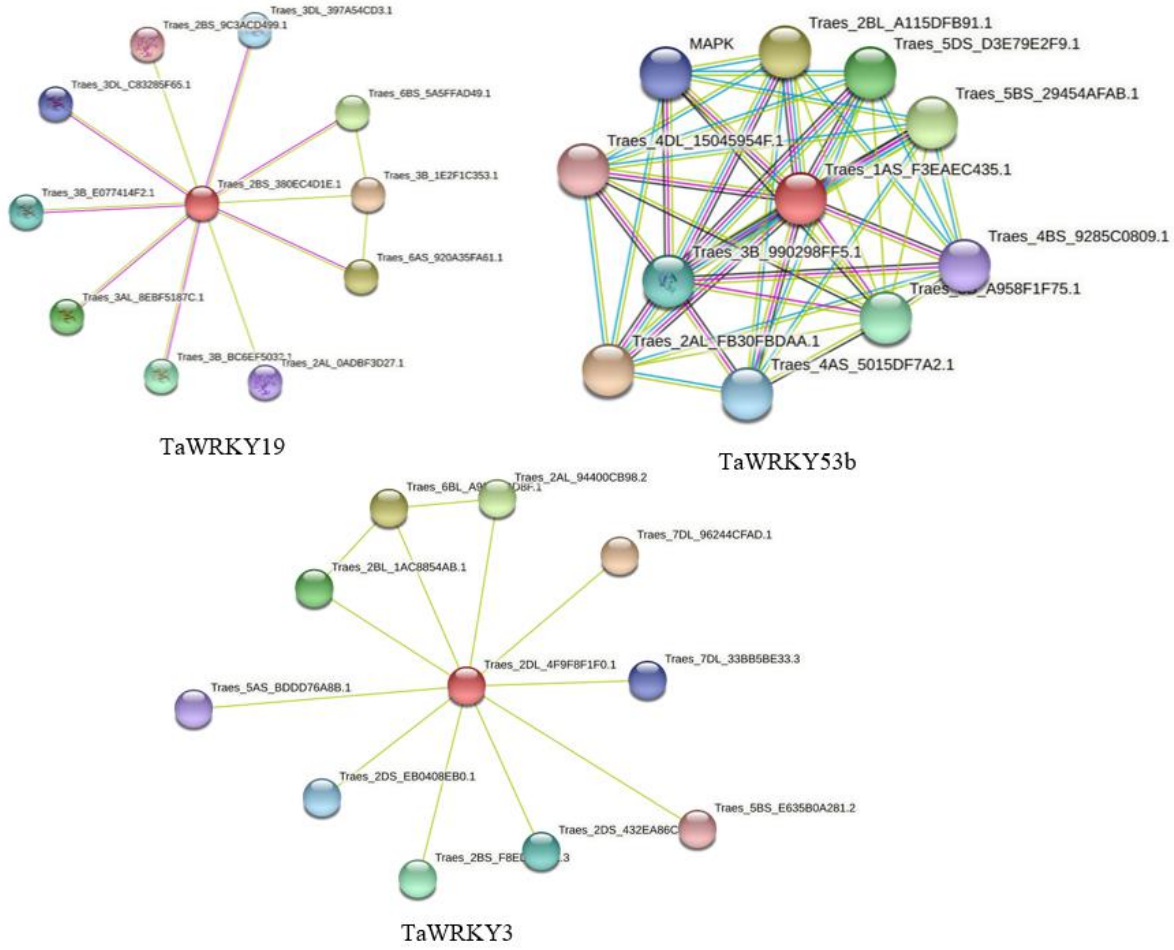


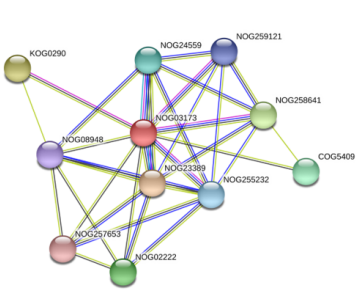
Figure 2-8 String protein – protein interaction network for *TaWRKY53b*, *TaWRKY19* and *TaWRKY3*. Coloured lines between proteins is an indication of various types of interaction evidence. Green line – neighbourhood evidence, Blue line – occurrence evidence, Purple line – experimental evidence, Yellow line – textmining evidence, light blue line – data base evidence, and Black line – coexpression evidence.

The other type of method used to identify protein-protein interaction was based on protein families (COGs) which is based on clusters of orthologous groups. The output of results would show families and functions of interacting proteins. In teams of *TaWRKY19*, there were 10 predicted functional partners. Their scores was no less than 0.953. There were an interacting of 4 valine-glutamine motif proteins (NOG23389; 0.997, NOG258641; 0.988, NOG24559; 0.982, NOG259121; 0.970). There was also an interaction of conserved WD40 repeat-containing protein AN11, phosphate ion transport protein, protein that is positive regulation of DNA-binding transcription factor activity, and WRKY transcription factor. Response proteins were also interacting to *TaWRKY19* such as response to chitin and response to ozone protein. *TaWRKY53b* shows interaction the same families of proteins as *TaWRKY19*. *TaWRKY3* interacted proteins had shown 10 highly functional predicated partners. Their score was accepted at 0.827. The highest interacting proteins was sequence-specific DNA binding protein (score 0.961). There was also cellular response

to boron-containing substance deprivation protein, response to ozone protein, and response to chitin protein, ADP binding protein, negative regulation of leaf senescence, ethylene-responsive transcription factor, MAP2K protein, fruit dehiscence protein, and predicted Zn-finger protein (Figure 2. 9).

Using STRING online software, protein to protein interactions had been tested based on their protein amino acid sequence. The figure shows that there is an interaction of each protein toward other types of protein such as VQ motif and other types of proteins. The test general assesses protein interaction to other proteins in terms of their neighbourhood, gene function, co-occurrence, coexpression, experiments as well as databases. The score had be integrated as confidence per prediction. In terms of *Ta WRKY53b* and *Ta WRKY19*, both had identical protein to protein interaction. Experiments prediction for both shows dominance of their prediction score in terms of VQ motif, conserved WD40 repeated-containing, and positive regulation of DNA-binding transcription factor activity. In terms of coexpression, response of chitin, phosphate ion transport, WRKY transcription factor and response to ozone proteins had the most dominant prediction score. Databases results showed only VQ motifs.

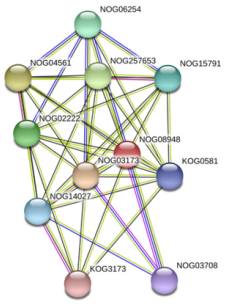
In terms of *Ta WRKY3* and its mutant form had identical protein to protein interaction results. Their predicted interaction lays down at co-expression that include sequence-specific DNA binding, cellular response to boron-containing substance deprivation, response to ozone, response to chitin, ADP binding, negative regulation of leaf senescence, ethylene-responsive transcription factor, mitogen-activated protein kinase kinase (MAP2K), fruit dehiscence and predicted Zn-finger protein. Sequence-specific DNA binding had the highest predicted score among all with 0.961 whereas fruit dehiscence had the lowest predicted score with 0.827 (Figure 2.9).



A

TaWRKY53b and TaWRKY19

Predicted Functional Partners		Score
NOG23389	VQ motif	0.997
KOG0290	Conserved WD40 repeat-containing protein AN11	0.991
NOG258641	VQ motif	0.988
NOG02222	Response to chitin	0.984
COG5409	Phosphate ion transport	0.982
NOG24559	VQ motif	0.982
NOG255232	Positive regulation of DNA –binding transcription factor activity	0.973
NOG259121	VQ motif	0.970
NOG08948	WRKY transcription factor	0.961
NOG257653	Response to ozone	0.953



B

TaWRKY3

Predicted Functional Partners		Score
NOG03173	Sequence-specific DNA binding	0.961
NOG04561	Cellular response to boron-containing substance deprivation	0.892
NOG257653	Response to ozone	0.883
NOG02222	Response to chitin	0.883
NOG06254	ADP binding	0.862
NOG15791	Negative regulation of leaf senescence	0.859
NOG14027	Ethylene-response transcription factor	0.841
KOG0581	Mitogen-activated protein kinase (MAP2K)	0.832
NOG03708	Fruit dehiscence	0.827
KOG3173	Predicted Zn-finger protein	0.826

Figure 2-9 protein to protein interaction using COG's. A; *TaWRKY53b* and *TaWRKY19*, B; *TaWRKY3*. Green line – neighbourhood evidence, Blue line – occurrence evidence, Purple line – experimental evidence, Yellow line – textmining evidence, light blue line – data base evidence, and Black line – coexpression evidence.

2.3.8 Regulation of gene expression via protein:DNA interactions

2.3.8.1 *TaWRKY19* (-2000 bp upstream) promoters:

From the translation start of each WRKY gene, the promoter sequences up to 2 kbp upstream was retrieved from EnsmblPlants and scanned using PlantCARE database to identify cis-acting regulatory elements (CAREs). *TaWRKY19* upstream promoter region had shown 10 types of known cis-acting regulatory elements. Along with core cis-acting elements such as CAAT box and TATA box, there is one ABRE (ABA-responsive element) functioning as a cis-acting element involved in the abscisic acid (TACGGTC [-618]) and 3 ARE cis-acting regulatory element essential for anaerobic induction (AAACCA [-38, -876, and -278]). There is also a presence of one Box 4 functioning as part of conserved DNA involved in light responsiveness (ATTAAT [-525]) and one MRE element which function as MYB binding site involved in light responsiveness (AACCTAA [-174]). There is one MBS element that also function as MYB binding site involved in drought inducibility (CAACTG [-403]). Five elements of CGTCA-motif (CGTCA [-29, -1010, -974, -935, and -1001,]) and five elements of TGACG-motif (TGACG [-29, -1010, -974, -935, and -1001]) both function as a cis-acting regulatory element involved in MeJA-responsiveness. There is a presence of one TC-rich repeats functioning as cis-acting element involved in defiance and stress responsiveness (GTTTTCTTAC [-909]), and one TCA-element functioning as cis-acting element involved in salicylic acid responsiveness (CCATCTTTTT [-351]). There are six W-box motifs with core sequence (TTGACG [-621, -1113, -1041, -1293, -9, and -744])) functioning as WRKY binding site. The W-box is well known for WRKY binding this suggests that the involvement of other WRKY TFs in regulating *TaWRKY19* (Table 2. 11).

Table 2-8 *TaWRKY19* (-2000 bp upstream) promoters.

Motif	Position	Signal sequence	Function
ABRE	618 (+)	TACGGTC	cis-acting element involved in the abscisic acid responsiveness.
ARE	38 (–); 876 (+); 278 (–)	AAACCA	cis-acting regulatory element essential for the anaerobic induction
Box 4	525 (+)	ATTAAT	Part of conserved DNA module involved in light responsiveness.
CGTCA-motif	29 (–); 1010 (–);974 (–); 935(–); 1001 (–)	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness.
MBS	403 (–)	CAACTG	MYB binding site involved in drought-inducibility.
MRE	174 (–)	AACCTAA	MYB binding site involved in light responsiveness.
TC-rich repeats	909 (+)	GTTTTCTTAC	cis-acting element involved in defense and stress responsiveness
TCA-element	351 (+)	CCATCTTTTT	cis-acting element involved in salicylic acid responsiveness
TGACG-motif	29 (+), 1010 (+),974 (+), 935 (+), 1001 (+)	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
W box	621 (–);1113 (–); 1041 (–); 1293 (–); 9 (–); 744 (–)	TTGACG	WRKY binding site.

It contains also 8 TATA box and 32 CAAT-box

TATA-box is a core promoter element around ~30 of transcription start.

CAAT-box is a common cis-acting element in promoter and enhancer regions.

Position numbers represent how far the motif from the start codon i.e 1 is close to start codon, 1000 is far from the start codon.

(+); is 5'-3' direction, (-); 3' – 5'direction.

2.3.8.2 *TaWRKY53b* (-2000 bp upstream) promoters:

In terms of *TaWRKY53b*, its upstream region was also scanned cis-acting elements. Its upstream had shown 11 types of known cis-acting regulatory elements. Two A-box, cis acting regulatory elements (CCGTCC [-214 and -536]). There are two elements of ABRE functioning as cis-acting element involved in the abscisic acid responsiveness (ACGTC, ACGTG, and GCCGCGTGGC [-1141 and -38, -62, and -1264]). Three CGTCA-motif, cis-acting regulatory elements involved in the MeJA-responsiveness (CGTCA [-812, -1199, and -1174]). G-Box (CACGTT [-38]) and G-box (CACGTC and CACGAC [-62, -1137, -1092, and -1140]), both cis-acting regulatory elements involved in light responsiveness. Gc-motif enhancer-like element involved in anoxic specific inducibility (CCCCCG [-253]). LTR cis-acting element involved in low temperature (CCGAAA [-818]).Two Sp1 elements for light responsiveness and one G-box which is a cis-acting regulatory element involved in light responsiveness (GGGCGG [-174]). TGA-element function as auxin-responsive element (AACGAC [-750]). TGACG-motif, cis-acting regulatory element involved in the MeJA-responsiveness (TGAGC [-812, -1199, and -1174]). One w-box element (TTGACC [-416]) (Table 2. 12).

Table 2-9 Regulatory motifs found within *TaWRKY53b* promoter (-2000 bp upstream) promoters.

Factor or site name	Site	Signal sequence	Function
A-box	214 (+); 536 (–)	CCGTCC	cis-acting regulatory element.
ABRE	1141 (+); 38 (–); 62 (–); 1264 (–)	ACGTC ACGTG GCCGCGTGGC	cis-acting element involved in the abscisic acid responsiveness.
CGTCA-motif	812 (+); 1199 (+) ; 1174 (+)	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness.
G-Box	38 (+)	CACGTT	cis-acting regulatory element involved in light responsiveness.
G-box	62 (+); 1137 (+); 1092 (+); 1140 (–)	CACGTC CACGAC	cis-acting regulatory element involved in light responsiveness.
GC-motif	253 (+)	CCCCCG	Enhancer-like element involved in anoxic specific inducibility
LTR	818 (–)	CCGAAA	cis-acting element involved in low-temperature responsiveness.
Sp1	174 (–)	GGGCGG	Light responsive element.
TGA-element	750 (–)	AACGAC	Auxin-responsive element.
TGACG-motif	812 (–); 1199 (–); 1174 (–)	TGACG	cis-acting regulatory element involved in the MeJA-responsivness.
W-box	416 (–)	TTGACC	WRKY binding site.

Position numbers represent how far the motif from the start codon i.e 1 is close to start codon, 1000 is far from the star codon.

(+); is 5'-3' direction, (-); 3' – 5'direction.

2.3.8.3 *TaWRKY3* (-2000 bp upstream) promoters:

Similarly to the other WRKYs, the promoter region of *TaWRKY3* was also analysed for the presence of cis-acting elements. Contrary to *TaWRKY53b* and *TaWRKY19*, analysis showed that the promoter region of *TaWRKY3* only contained 5 types of known cis-acting regulatory elements. There are two elements of ABRE functioning as cis-acting element involved in the abscisic acid responsiveness (AACCCGG and ACGTG [-241 and -445]). CAT-box cis acting regulatory element related to meristem expression (GCCACT [-211]). G-box cis-acting regulatory element involved in light responsiveness (CACGTC [-444]), MBS which is a MYB binding site involved in drought inducibility (CAACTG [-539]). Two Sp1 elements for light responsiveness and one G-box which is a cis-acting regulatory element involved in light responsiveness (GGGCGG [-257 and -367]). (Table 2. 13).

Table 2-10 Regulatory motifs found within *TaWRKY3* promoter (-2000 bp upstream) promoters.

Factor or site name	Site	Signal sequence	Function
ABRE	241 (–); 445 (+)	AACCCGG ACGTG	cis-acting element involved in the abscisic acid responsiveness
CAT-box	211 (+)	GCCACT	cis-acting regulatory element related to meristem expression
G-box	444 (–)	CACGTC	cis-acting regulatory element involved in light responsiveness.
MBS	539 (+)	CAACTG	MYB binding site involved in drought-inducibility.
Sp1	257 (+); 367 (+)	GGGCGG	Light responsive element.

Position numbers represent how far the motif from the start codon i.e 1 is close to start codon, 1000 is far from the star codon.

(+); is 5'-3' direction, (-); 3' – 5'direction.

The mutation of *TaWRKY3m* was located within the gene coding sequence and not the regulatory motifs of the promoter. Therefore, *TaWRKY3m* motif sequence was not included in the analysis.

2. 4 Discussion:

Based on the number of WRKY domains, *TaWRKY53b* (1A chromosome) and *TaWRKY19* (2B chromosome) exhibited two WRKY domains and *TaWRKY3* (2D chromosome) contains only one WRKY domain at the carboxylic part of the protein. Several studies had indicated that the carboxylic WRKY domain is the domain that facilitates the binding of the W-box motif within the promoter region. Yet, the N-terminus WRKY domain function is unknown. It was interesting to know which group of protein they lay on. We had constructed a phylogenetic tree containing several WRKY proteins from different groups and plant species. It covers the major WRKY TF groups and their sub-groups. It appears that both *TaWRKY19* and *TaWRKY53b* belong to group l WRKY proteins and *TaWRKY3* belongs to group-ll-c WRKY proteins. Indeed, there are many WRKY motifs such as WRKYGQK, WRKYGKK, and WRKYGEK. However, all tested proteins belong to the WRKYGQK motif and this was visible when WRKY proteins were subjected to WebLogo to look at the abundance of the WRKY motif. Our *in silico* findings aligns with Hassan *et al.* (2019) with regards to the WRKY genomic location, a number of domains, genomic coordinates as well as a gene identification number. This meets the first objective which is providing a phylogenetic tree analysis as well as determining their classification within the WRKY family and determine their conserved domains.

In addition, there was limited information with regards to wheat WRKY proteins. It was found that determining their counterpart in *Arabidopsis thaliana* might be helpful to get more information with regards to their function and activity. Their counterpart can be determined through submitting wheat WRKY protein sequences in databases such as Ensemble and/or Tair. The latter database is very specific for *Arabidopsis thaliana* species. The former was found as a very useful tool to assuring that both databases results aligned. It had been found that the *TaWRKY19* counterpart was *AtWRKY3* (AT2G03340). At Ensemble database, the coverage ID percentage between both was 83%. For *TaWRKY53b* was *AtWRKY26* (83.9%) [AT5G07100] and *TaWRKY3* was *AtWRKY12* (85.7%) [AT2G44745] which was the highest. With the help of PlantTFDB (v5.0), the transcription factor information could be found.

To understand the role of unknown wheat WRKY TFs, it is crucial to investigate orthologues of candidate wheat WRKY in other plant species and their role and function

could be obtained based on their known data from other plant species WRKY orthologues such as *Arabidopsis thaliana* and rice. *TaWRKY53b* orthologue in rice are *OsWRKY53* and both belong to Group I (Van Eck *et al.*, 2010; Satapathy;Kumar and Mukhopadhyay, 2017). *OsWRKY53* expressed in both roots and leaves and were found to be induced by drought stress as well as chitinous elicitors (Akimoto-Tomiyama *et al.*, 2003; Ramamoorthy *et al.*, 2008). In wheat, the Russian wheat aphid (*Diuraphis noxia* kurdjumov) (Botha;Swanevelder and Lapitan, 2010; Smith *et al.*, 2010) and leaf senescence (Wu *et al.*, 2008) were found to induce *TaWRKY53*.

Two orthologues of *TaWRKY3* were found which are *AtWRKY12* in *Arabidopsis thaliana* and *HvWRKY12* in barley (*Hordeum vulgare*). Barley percentage alignment to the wheat orthologue was higher, 99.1%, than *AtWRKY12*. *HvWRKY12* was found to be highly expressed in root tissues in response to salt stress (Uçarlı and Gürel, 2020). In terms of leaves, it was found to be highly expressed in response to drought stress (Janack *et al.*, 2016). Hollmann;Gregersen and Krupinska (2014) study showed that *HvWRKY12* was found to be upregulated during leaf development. The authors hypothesised that it might be involved in age-related processes. Their microarray study showed that *HvWRKYY12* was the only WRKY transcription factor upregulated on leaf flag during senescence. Tufan *et al.* (2017) found that *HvWRKY12* was responsive following the crown rot disease causative which was *Fusarium culmorum*. With regards to *AtWRKY12*, it had been demonstrated to localise within the nucleus of plant cells. It is typically expressed in flowers and silique as well as young, matured, and senescent leaves. Through its description, it was found to have a positive role in resistance to necrotrophic pathogens and it is induced by salicylic acid and during leaf senescence (Han *et al.*, 2019).

In terms of *TaWRKY19*, *AtWRKY3* was found to be its orthologue using blast alignment. *AtWRKY3* was found to be responsive to pathogen infection and salicylic acid (SA) (Lai *et al.*, 2008). Lai *et al.* (2008) had also shown that an overexpression of *AtWRKY3* results in an increase resistance to necrotrophic pathogens. It had been found to be oppositely modulated in Arabidopsis during flowering time under short-day conditions.

As WRKY TF had been first reported for their role in regulating stress reactions in plants, some of which tend to interact with proteins as part of stress response machinery in the plant nucleus. Like all transcription factors, WRKYs are found within the nucleus where they interact with the W-box promoter sequence of target genes and associate with other regulatory proteins to form the transcription complex with RNA polymerase. WRKY protein localisation is an important factor in terms of determining its activity in the cell. One of many useful methods was determining its localisation in the cell through availability *in silico* tools which can be used to predict cellular localisation based on protein sequence. Our finding using *in silico* tools (INSP, cNLS, SeqNLS, and DeepLoc) predicted that these wheat WRKYs were localised in the nucleus. *TaWRKY53b* had shown that it contains 3 nucleus location signals (NLS) within the amino acid backbone. *TaWRKY3* had 2 NLS and *TaWRKY19* contains only one. To confirm *in silico* prediction of WRKY cellular localisation, an experimental approach can be taken where putative WRKY genes can be cloned into a GFP vector and transformed into a plant cell such as onion epidermal cells. If the putative WRKY was a functional TF, transformants harboring the hybrid WRKY:GFP construct would emit a green fluorescence localised in the nucleus. Using such a method, *TaWRKY19* was found to be localised in the nucleus (NIU *et al.*, 2012) and its orthologues *AtWRKY3* also localised in the nucleus (Lai *et al.*, 2008). There is no available transformation conducted on *TaWRKY53b* and *TaWRKY3* proteins, nor was it possible to generate such plants within the scope of this project. However, *TaWRKY3* counterpart *AtWRKY12* was exclusively localised in the nucleus of *Arabidopsis* (Wang *et al.*, 2010) whereas *TaWRKY53b* counterpart *OsWRKY53* had shown a localisation in the nucleus (Chujo *et al.*, 2007). Based on *in silico* and *in vivo* results, *TaWRKY53b*, *TaWRKY19*, and *TaWRKY3* would be localised within and therefore function as TF in the nucleus and this satisfies our third objective.

Promoter regions and cis-regulatory elements:

The promoter regions found upstream of gene coding regions (ORFs) contain many diverse cis-acting regulatory elements. It is important to elucidate cis-regulatory elements at the promoter region of each candidate wheat WRKY and that was done by PlantCARE database. Such elements are considered to be a binding site for proteins involved in the initiation and regulation of transcription. Core promoter elements are located upstream (~40 bp) such as TATA-box which is an essential binding site for the transcription

initiation factor TFIID subunit (TATA-box-binding protein). Proximal and distal regions of promoters are found upstream of the core promoter region. Promoters contain different regulatory sequences (silencers, enhancers, insulators, and cis-acting elements). At a transcriptional level, all contribute to the regulation of gene expression (Hernandez-Garcia and Finer, 2014). The ability to control temporal and spatial gene expression is considered to be essential for defense response, in particular for plants that do not have the ability to move away from the stressor. Identifying which proteins interact with these promoter regions provides insights into gene regulation. PlantCARE in a bioinformatic tool that can identify *cis*-acting regulatory elements upstream of target genes. Among All WKRY proteins, it was found, when testing for regulatory motifs 2000 bp upstream, that they contain numerous types of *cis*-acting regulatory elements. This includes ABRE, TGACG-motif, CGTCA-motifs, and MBS etc. The presence of such *cis*-acting regulatory elements at the promoter of each WRKY TF gene, indicate their involvement at different biotic and abiotic stresses. To demonstrate their involvements, NUI et al (2012) had illustrated an improved tolerance to salt, drought, and freezing by an overexpression of *TaWRKY2* and *TaWRKY19*. They had shown that activation of key genes for stress was through their involvement in responsive biological systems such as Abscisic acids signaling (ABA) pathway. For example, RD29B gene had shown a fold change increase in its relative gene expression in overexpressed lines under unstressed conditions (3 fold and 7 fold, respectively). At the upstream promoter of RD29B, there had been a similar finding by Kim *et al.* (2011) that it contains ABRE motif. They had demonstrated that ABRE motif is found in the promoter region of ABA-inducible genes. At their model of ABA signaling regulation by ABRE, it had been shown that upon drought stress, ABA hormone tends to be activated leading to the expression of AREB1 gene. With splicing translation, protein modification and conformational change of ARBE1, this protein finally binds to ABRE motif at the promoter region of target genes leading to stress response. With the presence of such motif within the promoter region of *TaWRKY53b*, *TaWRKY19* and *TaWRKY3*. This suggests that they could be a target gene in response to ABA. As *TaWRKY53b* contains 4 ABRE (one at positive strand and 3 at negative strand), *TaWRKY3* contains 2 ABRE (one at positive strand and one negative strand) and *TaWRKY19* contains only one ABRE at its positive strand. This suggests their involvement in Abscisic acid (ABA) responsiveness. Zheng *et al.* (2013) studied the promoter region of *Tamarix hispida* WRKY4 (*ThWRKY4*) cis-elements at its promoter region. Their findings showed that its promoter region contain two ABRE motifs. In response to salt and drought, ABA stress signals are rapidly produced resulting in the

expression of *ThABF8* which bind to ABRE motif present upstream of *ThWRKY4*. This lead to the expression *ThWRKY4* and increasing stress tolerance.

With regards to CGTCA-motif and TGACG motif, both were contained upstream region of TaWRKY53b and TaWRKY19 which they considered as methyl Jasmonate responsiveness (MeJA) elements. PR5 and PR9 in *Arabidopsis thaliana* and *Oryza sativa* contain a conserved CGTCA-motif which was suggested to play a role in the activation of different abiotic defense mechanisms such as salinity, drought, and low temperature (Kaur *et al.*, 2017). It was found that the transcription of PR genes mediated by the binding of bZIP TGA transcription factor to TGACG (Wang *et al.*, 2013b).

There are six W-box elements present at the promoter region of *TaWRKY19* and one W-box element at the promoter region *TaWRKY53b*. As it is known that these W-box elements function as WRKY binding sites for regulating their expression. It can be either a target of other WRKY proteins or a self-regulator. For the latter, three W-box elements were found at the promoter region of *ThWRKY4*. It was found that *ThWRKY4* binds to its W-box element as self-regulator (Zheng *et al.*, 2013). This suggests that the presence of w-box element at the promoter region of *TaWRKY53b* and *TaWRKY19* could be a target of other WRKY transcription factors of a self-target as self-regulation of its expression.

Upon cis-acting elements identification using PlantCARE for candidate wheat WRKYs, the upstream promoter region contained multiple cis-regulatory elements. However, the orientation was either forward or reverse, such as ARE (-30, +876, and -278) at the promoter region of *TaWRKY19* (Table 2. 11). It is also found across promoters of *TaWRKY53b*, *TaWRKY3* as well (Table 2.12 and Table 2. 13). The transcription factor needs to bind to the promoter region of the target gene to induce or repress the transcription of the downstream gene. It is crucial to know whether the promoter element's binding orientation impacts the transcription downstream gene or not. Lis and Walther (2016) reported that the transcription factor binding orientation does not matter to the downstream gene transcription.

Protein to protein interactions:

It had been widely known that transcription factors will either activate and/or deactivate the transcription of targeted genes at their promoter side and/or interact with other cellular proteins as a response to a stressor. We had demonstrated the transcriptional regulatory map of our selected TaWRKYs. However, it is crucial to investigate the interaction of proteins to proteins interaction as well. Through the use of an online bioinformatic tool (STRING), two methods had been applied in order to investigate their interaction. The first one was through looking at their interaction to other proteins. However, much of outcome results showed protein ID's and no details indicate that they were uncharacterised proteins. For *TaWREKY19*, there were two ubiquitin-related modifier and two thioredoxin family interacting proteins. *TaWRKY53b*, three Mitogen Activated Protein Kinase (MAPK) interacting proteins and one *TaWRKY27* whereas *TaWRKY3* two cytochrome P450.

Small ubiquitin-related modifiers (SUMO) are conjugated proteins. Their function is to alter protein properties of modified proteins and increasing their proteomic complexity. This is considered as post-translational modification of target proteins. Their conjugation mediates protein trafficking as well as protein to protein interactions (Hay, 2005). With regard to gene expression, SUMO modification involved in the gene expression regulation. It acts on transcription factors by sumolyation which in result affects its ability of binding to target DNA. It can either affect transcription factor association or dissociation from its target promoter. SUMO was found to target WRKY transcription factor and control their ability of binding to their target DNA. SUMO overexpression was found to target *WRKY4*. As a result, the expression pathogen-induced *PR1* was found to be suppressed. In response to salt stress, the sumolyation of *MYB30* transcription factor was found to be critical to enable it to bind to the promoter of AOX1a and resulting an up-regulation of its expression (Roy and Sadanandom, 2021). Two SUMO proteins were found to interact with *TaWRKY19*. This might indicate that SUMO protein control *TaWRKY19* activity.

Thioredoxin are a small class of redox proteins which play a role in redox signaling. Their regulation extends to photosynthesis, plant growth, flowering and development (Meng *et al.*, 2010). The Redox pathway is fundamental for plant tolerance to intracellular oxidative stress caused by external biotic and abiotic stresses (Freeborough;Gentle and Rey, 2021). Thioredoxin acts on protein reduction through an exchange of cysteine thiol-disulfide

bonds (Holmgren, 1989). Using STRING (protein to protein interaction tool), *TaWRKY19* was found to interact with two Thioredoxin family proteins (figure 2.8). This might suggest their role in plant tolerance to biotic and abiotic stresses.

Among eukaryotes, MAPK are conserved protein kinases. They are involved in multiple environmental stresses signaling as well as developmental programming. Their mode of action is to phosphorylate their substrates. As a result, this post-translational modification (PTM) can contribute to protein regulation. Modified protein activity, subcellular localisation, stability, or translocation might be affected by post-translational modification. Generally, MAPKs localise in the cytosol of the cell but can also be found in the nucleus. These MARKs are associated with the transduction of signals detected at the cell surface to the nucleus as a result of environmental changes. Signal transduction occurs through the phosphorylation and de-phosphorylation of their target proteins resulting defense response activity. Several studies found that many WRKY transcription factors function as a downstream cascade of MAPKs resulting in biotic and/ or abiotic response (Bigeard and Hirt, 2018). With regards to biotic stresses, *Arabidopsis WRKY46* was identified as a MAPK3 substrate causing *WRKY46* to be positively regulating in plants defense (Sheikh *et al.*, 2016). Phytoalexins are a secondary antimicrobial compound induced in reponse to invading pathogen. MAPK3/MAPK6 plays in the induction of camalexin, a major phytoalexin in Arabidopsis thaliana. *AtWRKY33* was found to play a role in camalexin biosynthesis functioning as a downstream of MAPK3/MAPK6. Mutation of *AtWRKY33* phosphorylation sites for MAPK3/MAPK6 compromised the production of pathogen-induced camalexin (Mao *et al.*, 2011). *AtWRKY22* and *AtWRKY29* were identified as downstream of MAP kinase signaling cascade. Flagellin, as a pathogenic invading signal, activates MAPK pathway. Consequently, *AtWRKY22* and *AtWRKY29* gene express levels increase, leading to increased resistance to bacterial and fungal pathogens (Asai *et al.*, 2002). *OsWRKY30* also function as a downstream of MAPK cascade. The phosphorylation of *OsWRKY30* by several MAP kinases (*OsMPK3*, *OsMPK7*, *OsMPK14*) activated its activity of transcription allowing it to function as a transcription factor. Overexpression of *OsWRKY30* was found to improve drought tolerance in rice through its phosphorylation and this was proven by developing *OsWRKY30* mutated transgenic lines (Shen *et al.*, 2012). As it was found by protein to protein interaction test that *TaWRKY3* interact with

mitogen-activated protein kinase kinase (MAP2K) (figure 2.9). This might suggest that *TaWRKY3* function as a downstream MAPK cascade.

Cytochrome P450 is one of the plants largest enzymatic protein which is also exists ion other organisms such as mammals, fungi, bacteria, and insects. Their activity involved in many cellular metabolic pathways. The synthesis of secondary metabolites, as a result of cytochrome P450, function as plants growth, signals involved in development, biotic and biotic stresses protection (Jun;WANG and GUO, 2015). The interaction of *TaWRKY3* with two Cytochrome P450 proteins might suggests their activity and/or involvement in many other plants processes such as growth, development, and protection from plants biotic and abiotic stresses.

The protein to protein interaction network results was found to be a useful tool to identify which protein that interact with candidate TaWRKY protein. However, much of interacted proteins were found to be non-identified protein based on their EnsmblePlants ID. Thus, an alternative way is to investigate their interaction to which family proteins. In other words, what families of proteins our TaWRKYs tend to interact with. This was also done by STRING with a protein COGs mode. *TaWRKY53b* and *TaWRKY19* had shown an identical outcome. It had shown that both can have an interaction with VQ motif family proteins, response to chitin, response to ozone as well as WRKY transcription factors. For *TaWRKY53b* with regards to chitin, there might be a direct interaction as protein to protein interaction but no binding to the promoter W-box element in chitinase gene (Van Eck *et al.*, 2010). In terms of VQ proteins, they tend to be considered to be involved in responsiveness to stress (Zhu *et al.*, 2020). *TaWRKY3* had shown a diverse interaction to protein families in terms of Coexpression. The highest score was contained in sequence-specific DNA binding (0.961). This might suggest TaWRKY3 interacts with other protein involved in specific DNA binding activity.

With protein to protein interaction STRING *in silico* test, many interacting proteins were identified but were not defined properly. This might be caused by the lack of studies associated with TaWRKY proteins. Thus, no studies were detected toward identified interacting proteins. Commonly, only the Protein Ensemble IDs were returned with no information of the protein function or properties being reported. Further studies, *in silico*,

in vitro, and/or *in vivo* could be conducted to define their roles and functions toward plants defense against biotic and abiotic stresses.

In the present work, wheat WRKY proteins (*Ta WRKY53b*, *Ta WRKY19*, and *Ta WRKY3*) were studied *in silico* to determine their WRKY family grouping, subcellular localisation, promoter cis-acting elements, and protein to protein interactions. The outcome results for candidate WRKYs align with the literature in terms of their WRKY grouping and subcellular localisation. With regards to protein to protein interactions, there is a lack of knowledge and literature, at this stage, to fully understand the potential interactions with other cellular components. Further studies on the protein interaction will provide vital evidence to elucidate the complex interactions governing gene regulation.

3. Chapter 3: Cloning of selected TaWRKY transcription factors and small-scale production

3.1 Introduction:

WRKY transcription factors (TF) are largely distributed among plants. The nomenclature of WRKY TF's proteins is derived from its highly conserved 60 amino acid domain containing WRKYGQK amino acids sequence motif as a DNA binding domain (trans-terminus). A slight variation had been shown at this motif in various plant species in term of its amino acid sequence such as WRKYGKK and WRKYGEK. WRKY proteins can be divided into three main groups, depending on domain structures and zinc-finger (ZF)-motif at the C-terminus. Group I members have two WRKY domains and C2H2 (Cx4-5Cx22-23HxH) type ZF-motif whereas group II and III have only single WRKY domain and C2H2 and C2HC (Cx7Cx23HxC) ZF-motif, respectively (Wang *et al.*, 2016a). WRKY show a high binding affinity to cis-acting elements present on DNA named as W box elements, 5' (C/T) TGAC (T/C) 3'. Binding of WRKY representing the minimal consensus required for specific DNA binding.

Up to date, 171 wheat WRKY transcription factors have been identified and WRKY TF's function has been studied and identified in many plants, specifically crops, such as rice, barley, and soybeans, along with Arabidopsis as a model plant (Li *et al.*, 2020a). However, in wheat little is known in terms of their regulation, cellular mechanisms, and responses. Many of such studies were concerned on differential expression of WRKY TF in crops (Phukan;Jeena and Shukla, 2016). This is associated with more evolving understanding of their activity under different stresses either biotic or abiotic. However, there is less knowledge of the specific amino acid to DNA interactions that govern the highly specific DNA-binding activity of *WRKY*s required to initiate gene expression.

Our knowledge on plants responses to stressors had been enriched by the study of genetics. Since the first report published on 1994 on WRKY transcription factors (TFs), substantial progress has been achieved describing the role of plant *WRKY* TFs and their responses to different stresses, mainly on Arabidopsis as a model plant (Rushton *et al.*, 2010). However, the knowledge on the role of wheat WRKY (*Ta WRKY* TFs) under various stresses is small and very limited. Our group had made a great effort on working on various numbers of WRKY genes in wheat.

As explored in detail in chapter 1, many of such *TaWRKY*s had been found to change in its gene expression under stress are *TaWRKY53b*, *TaWRKY19*, as well as *TaWRKY3* and its mutant form. All had been exposed to dual stress; all share nitrogen reduction as the abiotic stress. The only difference lay with biotic stress. *TaWRKY53b* gene expression showed a change with exposure to *Zymoseptoria tritici*, the causative agent of Septoria leaf blotch (SLB) (Poll, 2017), changes *in TaWRKY3* and 3 mutant expressions were quantified in response to aphid, *Sitobion avenae,* infestation (Alshigeihi, 2019), and the expression *TaWRKY19* gene was determined following *Bipolaris sorokiniana* infection (Baba, 2019). *TaWRKY3* mutant is a *TaWRKY3* but a tyrosine to aspartic acid substitution within the WRKY domain, resulting in WRKYGQK to WRKDGQK. The aim is to perform a cloning and expression analysis of candidate *Ta*WRKYs. There are many tools of methods for analysis of such gene which can be either through delivering over expressed *Ta*WRKY genes into model plants such as *Arabidopsis thaliana* and study its role by intensively stressing mutant plants and studying its phenotypic characteristics. The downside of this system is its time consumption and intensive labour work for generating homozygous transformed plant (Kihara *et al.*, 2006). However, other systems can be used such as *in vivo* systems which can be done through expressing *Ta*WRKY proteins through cloning and using a microbial heterologous expression host to produce desired proteins. Such proteins can be extensively used for many studies such as DNA:protein binding and protein:protein binding assays.

Due to problems encountered with the use of prokaryotic system for WRKY fusion protein expression encountered with Ciolkowski *et al.* (2008b), Alshegaihi (2019), and Poll (2017) work, alternative method of expression could be used for expressing WRKY fusion proteins heterologously in their native structure using yeast. *Saccharomyces cervisiae* and *Pichia pastoris* are the two most utilised yeast strains for heterologous protein expression. Generally, yeasts could be considered as suitable organisms for expressing recombinant eukaryotic proteins due to its secretory pathway enabling recombinant proteins to be secreted to the media, favoured disulphide bonds and glycosylation, however, a misfolded protein might by produced by *E. coli*, as a prokaryotic expression organism, leading to either inactive or insoluble proteins (Demain and Vaishnav, 2009). Thus, solubility and stability of expressed proteins are important and these can be, with yeast, enhanced as well as higher chances of potential functional proteins (Lambertz *et al.*, 2014). In contrast to prokaryotic expression systems, yeast systems combine the

feasibility of genetic manipulation as well as the rapid cellular growth found with prokaryotic systems with the presence of a necessary subcellular machinery for post-translational modifications (Martínez-Alarcón;Blanco-Labra and García-Gasca, 2018).

Glycosylation, disulphide bonds and proteolytic processing are some of many advantages that made yeast systems to overcome prokaryotic systems. In addition, with the use of proper signalling sequences, yeast are capable of secreting extracellularly active heterologous proteins into broth which can be easily harvested. Thus, it had been proposed, initially, that *S. cerevisiae* to be a suitable host platform for protein expression due to the long history of using *S. cerevisiae* in industrial fermentation systems as well as its molecular understanding of physiology and its genetics. However, glycosylation of *S. cerevisiae* was unacceptable, in terms of glycoproteins for mammalian proteins (Demain and Vaishnav, 2009). Thus, it was replaced with *P. pastoris* as an alternative expressing host candidate. There are many advantages of *P. pastoris* over *S. cerevisiae* among of which are the following: i) high protein production; ii) hyperglycosylation avoidance; iii) reasonable growth in solution containing strong methanol that inhibits most other microbial organisms ; iv) construction of multi-copies of DNA constructs that can integrate into chromosomal DNA which can yield stable transformants (Gellissen *et al.*, 1992; Demain and Vaishnav, 2009).

Up to date, there are no attempts in the scientific community to utilise the eukaryotic, yeast, system for the heterologous expression of plant WRKY proteins. Thus, one of our objectives is to express WRKY proteins in *P. pastoris* to take advantage of the benefits of using a eukaryotic system compared to a prokaryotic system. For expressing WRKY TFs in *P. pastoris*, it is important to select a vector which can be used as a backbone and driver for expression within *P. pastoris*. There are many vectors that can be utilised in *P. pastoris* providing both constitutive and inducible protein expression. The yeast can utilise either methanol or glycerol as a carbon source depending on the type of vector used. Selecting the type of vector is mainly dependant on desired end of product as an expressed protein. Our intension was to select a *P. pastoris* protein expressing vector that could allow us to clone WRKY genes as well as express it heterologously with an efficient protein production to allow to harvest proteins for analyses and further processes.

pGAPZαA is a member of pGAPZ family (pGAPZ A, B, and C, and pGAPZα A, B and C) that are sold commercially for protein expression for *Pichia*. Each member of this family has its own distinct features. However, all have some common features that they share together. For the sake of simplicity, pGAPZ version lack encoded *S. cervisie* protein α-factor secretion signal while pGAPZα versions contained it within its backbone. α-factor secretion signal is infused within the protein N-terminal peptide of the secreted protein and it allows for efficient secretion of most proteins allowing for harvesting proteins from growth medium. This can be said that pGAPZ versions express intracellular proteins while pGAPZα express extracellular proteins. All of which contain Zeocin™ resistance as an antibiotic gene which allows for selecting transformants in both *E. coli* and *Pichia* cells. Expressed protein can be fused, in both pGAPZ and pGAPZα, to C-terminal peptide containing the *myc* epitope and enables for detection of the fusion protein by the Anti *myc* Antibody. The C-terminal polyhistidine (6x-His) tag encodes six histidine residues that form a metal binding site for affinity purification of recombinant protein as well as used detection. Both contain Glycerldehyde-3-phosphate (GAP) promoter encodes for glycerldehyde-3-phosphate dehydrogenase which had shown to express recombinant proteins constitutively in *P. pastoris* to high levels and this is dependent on carbon source. With similar features between pGAPZ and pGAPZα, the latter series had been selected for their ability to secret heterologous proteins whereas the former express intracellular recombinant proteins (Cregg *et al.*, 2000). Not much of differences between pGAPZα A, B, and C. The only differences lay with additional restriction sites occurs only in versions B and C, between α-factor signal and the *Eco*R I site. For cloning our genes, restrictions sites that are between *Xho I* within α-factor and *Xba I* located prior to *myc* epitope were needed for cloning. Thus, pGAPZαA was selected as an expression candidate due to its ability to satisfy our intentions for expression and secretion of WRKY TF proteins.

Generally, genes desired for cloning (WRKY genes) can be inserted into the vector between the α-factor signal and *myc* epitope with the use multiple of restriction sites that is integrated within that specific region. With this cloning, the expressed protein would be carried out of the cell with the help of the infused secretion signal at its N-terminal (a-factor) into the medium broth. The presence of polyhistidine tag (6-His) at the protein C-terminus would be suitable for use as a metal binding site for affinity purification of the desired protein. Also, the presence of protein, *myc* epitope at its C-terminus would be then employed with Anti-*myc* Antibody for protein detection. Cloning of WRKY genes had to

be infused within vector backbone without affecting the reading frame of α-factor signal at its N-terminus as well as *myc* epitope tag and 6- His tag at its C-terminus for proper expression of these tags. Impacting tags reading frame at C-terminus would lead, as a consequence, a loss of its functionality for detection and purification at later stages. Thus, it was necessary to, as a critical control point, to make sure that cloned WRKY genes fits within reading frame of pGAPZαA.

Our cloning and analysis would be directed for studying cloning and producing *Ta*WRKY protein for analysing its role in terms of its binding to promoter genes in wheat plants. Our strategy had been divided into three main parts; I) cloning and confirmation of expression; II) large scale expression and protein purification; III) DNA-protein binding assays through Electrophoretic mobility shift assays. This part of the project aims to satisfy the first strategic point via the cloning and expression of candidate *Ta*WRKY proteins using the methyltrophic yeast, *Pichia pastoris* as the expression platform.

To achieve this aim the following objectives will be met;

1) Insert TaWRKY coding sequences into pGAPZαA expression vector
2) Transformed recombinant vectors to *P. pastoris*
3) Small scale expression tests

3.2 Materials and methods:

3.2.1 Cloning strategy:

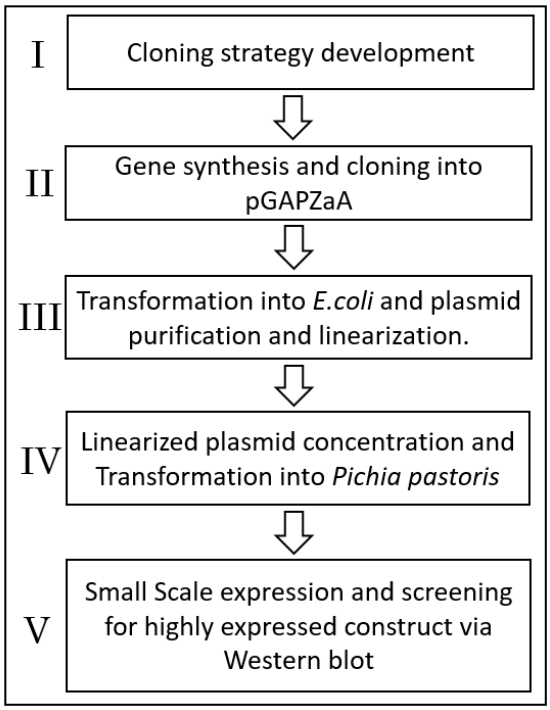


Figure 3-1 Summary of cloning strategy and overall experimentation.

Ta WRKY53b (1320 bp), *Ta WRKY19* (1404 bp), *Ta WRKY3* (687bp) and *Ta WRKY3m* (687 bp) coding sequences (Appendix A). At the multiple cloning site, *Ta*WRKY genes would be cloned to create fusion proteins with *myc* epitope and 6x His tag at each recombinant protein C-terminus. Each WRKY gene would be cloned into pGAPZαA (appendix A) with the Kex2 cleavage site. This through the use of *Xho I* site at 736 – 741 to clone each WRKY gene with Kex2 cleavage site and also *Xba I* restriction site (823). The reading frame of each recombinant protein was assured to be started with Leu. Arg. Glu. Ala. (L R E A) amino acids from *Xho I* restriction site at the Kex2 signal cleavage. It is then followed gene of interest started with Met (M) and completion of the reading frame with poly-Histidine tag (Figure 3. 1, I).

pGAPZaA was selected as it enables candidate TaWRKY proteins to be expressed and released externally to the culture medium with the aid of Kex2 signal. The presence of His-tag enables recombinant proteins to be detected and purified.

Synthesised genes would be cloned into *pGAPZαA* (Figure 3. 1, II) which would be then transformed in *E. coli*. Cloned plasmids in *E. coli* would be then propagated in Zoecin™ containing Low Salt Lysogeny Broth (LSLB) plates to be purified. *Bgl II* restriction

enzymes was effectively used to linearize constructs (Figure 3. 1, III) for transformation into *Pichia pastoris* (Figure 3. 1, IV). Small scale expression in 50 ml baffled flasks to screen for highly expressed transformants using western blot (Figure 3. 1, V).

3.2.2 Summary of transformation into *Pichia pastoris*:

Expression constructs were designed using Benchling, a cloud-based informatics platform for life sciences R&D. *Ta WRKY53b*, *Ta WRKY19*, *Ta WRKY3* and *Ta WRKY3m* genes were synthesised using Genscript and directly cloned, individually, into pGAPZαA (Figure 3. 2, I, II, III, and IV, respectively)(appendix C). Cloned construct were initially transformed into TOP10 *E. coli* and propagated. Constructs were then purified and linearized using *Bgl II* restriction enzyme (figure 3.1 III). Linearized constructs were then transformed into *P. pastoris*. Transformants were selected on Zoecin™ containing plates (Figure 3. 1, IV). WRKY proteins were expressed in 50 ml YPG for 72 hours which were precipitated using Trichloroacetic acid (TCA) precipitation and run in SDS-PAGE and western blot (Figure 3. 1, V). Positive constructs were only selected for bench top expression.

3.2.3 Detailed materials and methods:

For generating new vector constructs and plasmid extractions, DH5α *E. coli* strain had be used in all cloning experiments. To generate *Ta WRKY53b*, *Ta WRKY19*, *Ta WRKY3*, and *Ta WRKY3m* expression transformants, wild type *P. pastoris* (X-33) host strain was used in these studies. The advantage of this strain is that is can make an expression of recombinant proteins from vectors with Zeocin™ Resistance as the only selectable marker. This is to selectively allow the growth of transformed *Pichia* cells with pGAPZaA constructs.

E. coli competent cells were obtained from Agilent Technologies marked as StrataClone SoloPack competent cells (Cat. 200185) and *P. pastoris* wild type strain from Invitrogen corp. (Carlsbad, USA).

3.2.4 Expression plasmids:

pGAPZαA plasmid was used mainly as a parent plasmid for the construction of *Ta WRKY53b* (pGAPZαA/*Ta WRKY53b*), *Ta WRKY19* (pGAPZαA/*Ta WRKY19*), *Ta WRKY3* (pGAPZ*a*A/*Ta WRKY3*) and *Ta WRKY3m* (pGAPZ*a*A/*Ta WRKY3m*) which was obtained from Invitrogen corp. (Cat. no. V205-20. Carlsbad, USA). Each vector uses *GAP* promoter

to constitutively express recombinant proteins in *P. pastoris*. For cloning in this vector, it is important to care for protein expression reading frame using Expasy (https://web.expasy.org/translate/). Otherwise, *myc* epitope tag and 6x His tag expression would be interrupted which would be a problematic to reading frame.

3.2.5 Oligonucleotide Synthesis and cloning:

Ta WRKY53b, *Ta WRKY19*, *Ta WRKY3* and *Ta WRKY3m* genes were designed to be cloned into pGAPZαA between Kex2 cleavage site and *myc* epitope using *Xho I* (736 bp) and *Xba I* (823 bp) in which have to be fitted with the expression reading frame for pGAPZaA for appropriate synthesis (Figure 3. 3; Figure 3.4 A, B, C, and D, respectively). Design of constructs was conducted on Benchling (https://www.benchling.com/). Benchling was then used to predict endogenous restriction enzymes specifically for BspH I, Avr II, and Bgl II with all constructs. The coding sequence of each candidate TaWRKYs were then synthesised by GenScript and cloned into pGAPZaA plasmid backbone between *Xho I* and *Xba I* (figure 3-3). This to ensure that all constructs would contain myc epitope and His-tag. As it had been synthesised and cloned, there was no need for sequencing each construct. This is because each construct was designed to fit in frame prior to the synthesis and cloning. All constructs were received at 4 µg dried form

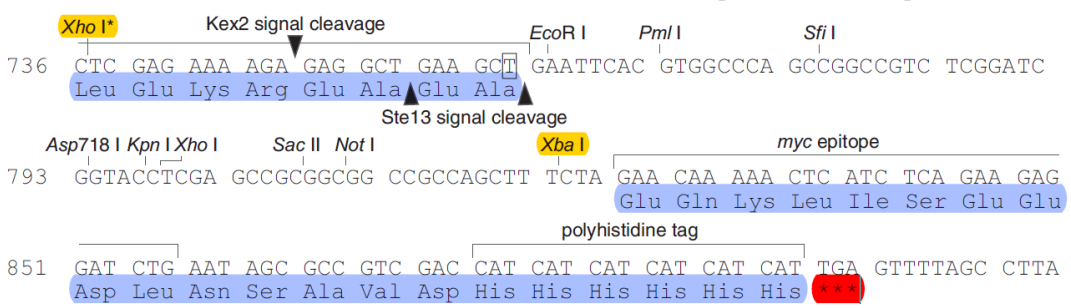


Figure 3-2 pGAPZαA partial DNA sequence. Restriction sites labelled to indicate the site of cleavage for the insertion of gene of interest into the expression vector. α-factor signal sequence contains native code *S. cerevisiae* α-factor secretion signal allowing for protein section efficiently in *P. pastoris*. C-terminal *myc* epitope permitting fusion protein detection by Anti *myc*-epitope. C-terminal polyhistidine tag contains six histidine residues coding that form a metal binding site for affinity purification for recombinant protein and detection by ant-His tag. Arrows indicate Kex2 signal cleavage and Ste13 signal cleavage that cleaved from expressed fusion protein by yeast intracellular protein expression machinery. *Xho I* and *Xba I* restriction sites highlighted in yellow, fusion reading frame highlighted in blue, and stop codon for fusion protein highlighted red.

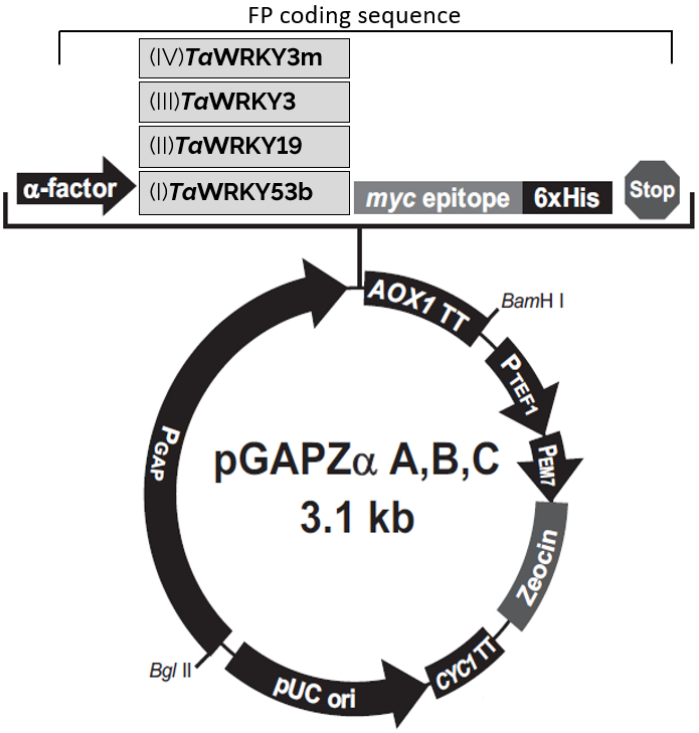


Figure 3-3 Diagrammatic representation of consecutive expression of pGAPZαA (3147 bp). Box indicate α-Factor secretion signal, (I) *TaWRKY53b* gene (1320 bp; 50.16 kDa), (II) *TaWRKY19* (1404 bp; 53.69 kDa), (III) *TaWRKY3* (687 bp; 24.96 kDa), and (IV) *TaWRKY3m* (687 bp; 24.96 kDa), myc epitope and 6x His tag followed by stop codon. AOX1 TT; AOX1 transcription termination region, P TEF1; transcription elongation factor 1 for expressing Zeocin™ resistance gene, CYC1 TT transcription termination, pUC for replication and maintenance of plasmid in *E. coli*, pGAP allow constitutive high level expression in *Pichia pastoris*. FP is an indication of fusion protein.

3.2.6 Low salt Luria broth (LSLB):

For liquid medium, 10 g tryptone (sigma, T7293), 5 g NaCl (sigma, S9888), and 5 g Yeast Extract (sigma 70161) were combined together and 950 ml deionised water. pH was adjusted to 7.5 with 1N NaOH and the volume brought to 1 liter. For plate 15g/l agar was added before autoclaving. 25 µg/ml Zeocin™ was added to the solution after autoclaving when the medium cooled down to at least 55°C. Plates were stored at 4°C in the dark. Zeocin ™ is stable for 1-2 weeks as maximum.

3.2.7 Transformation into *E. coli*:

50 µl competent cells, *Escherichia coli* (Invitrogen) and/or DH5α (New England Biolabs) strains were mixed with 200 ng of plasmid DNA and incubated on ice for 30 minutes. Heat shock reaction was started as cells were incubated in 42 °C water bath for 30 seconds and placed on ice for 2 minutes. 250 µl was added to the mixture and incubated at 37 °C with shaking at 225 rpm for 1 hour and then followed by spreading on selection plate containing low salt LB agar and 25 µg/ml Zeocin™. Cells were incubated at 37 °C overnight to allow only transformed cells to grow on the selection plate.

3.2.8 Plasmid purification:

Plasmid DNA extraction from bacterial cells, QIAGEN® miniprep DNA purification was used. Cells were harvested from overnight bacterial cultures by centrifugation for 3 minutes at the maximum speed (13,000 x g) in a conventional benchtop centrifuge machine at room temperature. The pellet was re-suspended, by vortexing, using 250 µl of Resuspension Solution (P1) buffer. 250 µl of Cell Lysis Solution (P2) was added and mixed the lysate by inverting 4 – 6 times. 350 µl of Neutrislation Solution (N3) was added to stop the reaction and mixed by inverting 4- 6 times. It is then followed by an immediate centrifugation of the bacterial lysate at 13,000 g for 10 minutes.

Into a combined spin column and collection tube, the clear lysate was transferred which is then centrifuged at 13,000 g for one minute. The flow through was discarded and 750 µl of Wash Solution (PE) was added and centrifuged for one minute. Again, the flow-through was discarded and the sample was then further centrifuged for 1 minute at 13,000 g to remove any remaining residuals of Wash solution. The spin column was then transferred into a sterile 1.5 µl microcentrifuge tube and 15 µl of Elution Buffer (EB) was pipetted directly onto the column surface and incubated for 3 minutes at room temperature. For higher concentration of plasmid, EB was heated at 70 °C. Finally, the samples was then centrifuges at 13,000 g and extracted plasmid was quantified using Nanodop 1000.

3.2.9 Linearization of plasmids for transformation:

Prior to transformation, pGAPZαA constructs must be linearized with restriction enzyme Bsp HI (PagI). The enzyme was supplied by Thermo Fisher Scientific Ltd. As structured by Invitrogen for pGAPZαA linearization, about 5 to 10 µg of plasmid DNA had to be digested with a single cut using *Bsp HI* (356), *Avr II* (191 bp), or *Bgl*II (1bp). Each of which would cut the plasmid DNA at a single site in the GAP promoter region to linearise the vector. Thus, it had been recommended to select a restriction enzyme that does not cut within inert gene. 2 µl of restriction enzyme (either one of previously mentioned), 2 µl of 10 X buffer and 5 – 10 µg plasmid DNA were mixed were added; the mixture was made up to 20 µl with nuclease free water. Incubation is 37°C and inactivation time varies each enzyme (Table 3. 1). Prior to transformation, 1 µl of linearized plasmid was used to run on 1% agarose gel to confirm digestion completion and then followed by purification of DNA using QIAquick PCR purification kit (cat nos. 28104) for transformation.

Table 3-1 List of enzymes for construct linearization

Enzyme	Incubation time and temperature	Inactivation time and temperature.
BspHI	37 °C for 16 hours	80 °C for 20 minutes
AvrII	37 °C for 3 hours	80 °C for 20 minutes
BglII	37 °C for 20 minutes	non

3.2.10 Gel electrophoresis:

Gel electrophoresis was used to detect the size of linearized constructs. The concentration percentage was variable [0.8-2% (w/v)] which was dependently selected upon the size of the analysed fragment. Agarose gel is prepared melting in x 1 TAE using a microwave oven. As the gel cools down to 50 °C, ethidium bromide is added to concentration 0.5 µg/ml and mixed thoroughly by swirling the flask. Using platform with a comb in place, the gel is then poured and left to solidify. After removing the comb from the solidified gel, the gel is placed in the electrophoresis tang immersed with 0.5 x TAE. DNA loading Dye was added to the DNA mix and added to the gel electrophoreses well. DNA molecular weight markers were added alongside with PCR and run samples. The size of ladder to be added was dependent on the analysed fragment of nucleic acid. For smaller fragments up to 800 bp, 50 bp is generally being used and 1 kp ladder is being used for larger fragments. Thus, the size of ladder was chosen depending upon the analysed DNA fragment. Gel electrophoresis was carried out at 100 voltage. The time was conducted 40 - 45 minutes. To visualise the DNA fragment size, ultra violet transilluminator was used.

3.2.11 PCR purification kit:

Based on agarose gel results, PCR purification kit was conducted on the linearized plasmids to remove restriction enzymes using QIAquick PCR purification kit (cat nos. 28104) which is to purify 10 µg of each construct. The range of amplified fragments to be purified that can be obtained is from 100 bp to 10 kb in size. The aim of this is to remove PCR reaction impurities and to get as much pure DNA as possibly. Following the manufacturing protocol, 5 volumes of PB buffer was added to 1 volume of PCR reaction and mixed thoroughly by pipetting up and down to ensure mixing. The mixture was then applied on the QIAquick column to bind DNA and centrifuged at maximum speed for 60 seconds and discarded the flow through. To wash the column, 750 µl of PE buffer was added to column and centrifuged for 60 seconds and followed by a second centrifugation

after changing the collection tube into a fresh one. This is to ensure that all residual wash buffer had been removed. The QIAquick column then placed into a new 1.5 ml microcentrifuge tube and 30 µl of EB buffer was added to the centre of the column to elute the DNA. The elusion buffer was left on the column and left to stand for 10 minutes which was then followed by centrifugation at maximum speed for 2 minutes. This was to ensure that all DNA had been eluted from the column. Finally, 1 µl of the eluted DNA had been used on the Nano drop to measure the concentration of DNA in ng/µl.

3.2.12 DNA concentration:

DNA was concentrated via a vacuum speed to reduce the volume of water volume and concentrate the DNA sample. This could be either circular (non-digested plasmid) or linearized plasmid.

3.2.13 Transformation of constructs into yeast competent cells:

3.2.13.1 Preparation of competent cells:

A single colony was inoculated into 10 ml of Yeast peptone dextrose (YPD) (1% yeast; 2% peptone; 2% dextrose [D-glucose) of wild-type *Pichia pastoris* (X-33) and grown overnight at 28 – 30 °C on a shaking incubator at 250 – 300 rpm. Firstly, cells were diluted from the overnight culture to OD_{600} of 0.1 – 0.2 in 10 ml of YPD. Cells were left 4 – 6 hours to grow at 28 – 30 °C on a shaking incubator until the OD_{600} reached 0.6 – 1.0. Cells were pelleted by centrifugation at 500 x g for 5 minutes at room temperature. Cell pellet was then re-suspended in 10 ml Solution I which is then followed by pelleting cells at 500 x g for 5 minutes at room temperature. Pellet was then re-suspended in 1 ml of solution I. At this stage cells are now considered as competent cells which they were aliquot in 50 µl into labelled sterile microcentrifuge tubes. Cells were the freezed down slowly and stored at -80 °C until use. This method of perpetration is obtained from *Pichia* EasyComp kit (Invirtogen).

3.2.13.2 Transformation:

Linearized constructs were integrated into the host chromosome using *Pichia* EasyComp kit (Invirtogen) through a heat shock protocol. For each transformation, 50 µl of fresh competent cells were thawed at room temperature. 3 µg of linearized Pichia expression vector DNA were added to competent cells. As recommended from the manufacturing

protocol, the volume should not exceed 5 µl. Thus, the volume of DNA was reduced and concentrated using speed vacuum. Following that, 1 ml of solution II was added to DNA/cell mixture and mixed by vortexing and incubated for 1 hour at 30 °C which was mixed every 15 minutes sequentially for an increased transformation efficiency. The reaction mixtures was exposed to a heat shock at 42 °C for 10 minutes. As plasmid constructs considered as Zeocin™ resistant plasmids, the reaction mixture was split into 2 microcentrifuge tubes, each contains approximately 525 µl per tube. 1 ml of YPD was then added into each tube and cells were incubated at 30 °C for 1 to 3 hours. This will allow expression Zeocin™ resistance. Tubes were then centrifuged at 3000 x g for 5 minutes at room temperature to harvest only pellet cells. 1 ml of solution III and tubes were combined together into one 2 ml tube. Cells were re-pelleted again by centrifugation at 3000 x g for 5 minutes at room temperature and discarded the supernatant. Cells were finally re-suspended in 100 – 150 µl of solution III, The entire transformation were plated on YPD containing 100 µg Zeocin™ and incubated for 2 to 4 days at 30 °C.

As strongly recommended by Invitrogen for pGAPZαA vectors to purify Zeocin™ resistant transformants at least once. This is because that at later steps it would not require the use of Zeocin™ selection. Also, mixed colony tansformants may impact the expression clone. Thus, Zeocin™ resistant *Pichia* transformants were streaked on YPD (1% Yeast extracts, 2% peptone, 2% dextrose, and 20g/l agar) containing 100 µg/ml Zeocin™ and incubated for 2 to 4 days at 30 °C. Transformed cultures were routinely cultured.

3.2.14 Expression of protein of interest:

Prior to fermentation for expressing the gene of interest, it is important to identify and confirm a recombinant *Pichia* clones that were expressing the correct protein. This was done through small-scale expression (shake flask protein expression). The primary purpose of this step is made to for identification and conformation of expressing *Pichia* transformants. However, expression conditions might not be optimal for the expressed protein but, in general, is usable and useful for detecting protein of interest though a suitable detection method (i.e. SDS-PAGE and/or western blot etc.).

3.2.15 Media and culture conditions:

In order identify that colonies expressing recombinant protein, cells were grown in several media recombination's in this study. Yeast peptone dextrose (YPD) medium (1% yeast extract. 2% peptone, 2% d-glucose), Yeast peptone glycerol (YPG) medium (1% yeast extract. 2% peptone, 2% glycerol). All of which were used to test the expression of transformed *Pichia pastoris* cells. At all experiments, cell cultivation was held baffled flasks (250ml) with working volume of 50 ml at 28 – 30 °C on a shaking incubator at 250 – 300 rpm. Cultures were incubated at different time points ((0h, 24 h, 48 h, and 72 h) which were collected (50ml) at each time point. The main purpose is to analyse protein expression as well to get an idea of its optimal time point to harvest. Thus, samples were collected and centrifuges at 4,300 x g speed in a centrifuge for 10 minutes at 4 °C. For secreted protein, the supernatant was transferred into a separate tube. Both supernatant and cells pellets were flash freeze using liquid nitrogen and stored at -80 °C until ready to assay.

3.2.16 Detection of recombinant proteins in *Pichia* cells:

Following growing in a small scale expression, it is important to analyse the expression of protein of interest. Analysis can be up taken from either cell pellets and/or medium. This is for the presence of desired recombinant protein. For unprocessed protein that had been collected from the cell pellets, it had be accounted the additional proteins approximately 9.3 kDa from the pGAPZαA that would be added into the protein of interest from the a-factor signal sequence. There is also additional 2.5 kDa would be added from C-terminal tag to protein of interest containing *myc* epitope and His-tag.

Several methods that can be used for detecting protein of interest and one of which is SDS-PAGE through a Coomassie-stained blue. The sensitivity of the method can detect as little as 100 ng in a single band. However, a more precise method can be used as well is western blot analysis. The method of detection comes from Anti-His (C-term) antibody, which can detect as little as 1 – 10 pg.

3.2.17 Analytical methods of expressed proteins:

3.2.17.1 TCA/Acetone Protein precipitation:

To analyse the expression of supernatant samples obtained from both small scale and/or large scale fermentation, Trichloroacetic acid/Acetone (TCA/Acetone) protein precipitation is commonly used (Hao *et al.*, 2015). The main objective of this method is to

remove contaminants, interfering substances (lipids, nucleic acids, salts etc.) and to concentrate protein samples. The supernatant was collected and added TCA at 10% of the total supernatant volume and mixed by vortexing which is the left to stand on ice for 15 minutes. The mixture was then centrifuge at maximum speed for 10 minutes at 4 °C and discarded the supernatant. Pellet washed twice with ice cold acetone at 10% of the total supernatant (same volume initially used for TCA) and centrifuge using a benchtop at maximum speed for 5 minutes at 4 °C and discarded the supernatant. Washing with acetone was repeated once again and centrifuged at the same speed for 5 minutes. the pellet was then dried on a heat block at 90°C for 3-5 minutes and resuspended on 1X PBS (10 – 20 µl) and stored at -20 °C until needed.

For protein analysis, precipitated proteins was mixed with 6X SDS sample loading buffer (4x Tris-HCL [pH6.8], 30% glycerol, 10% SDS, dithiothreitol [DTT], and bromophenol blue [β-me]) in 4/1 ratio and boiled at 100 °C for 5 – 10 minutes. The sample mix let and spin after being cooled down.

3.2.17.2 SDS-PAGE:

15% separating gel was prepared with 0.3 M Tris (pH8.8), 30% acrylamide/Bis Assay, 0.1% SDS, 0.1% ammonium persulfate (APS), and 0.01% *N,N,N',N'*-tetramethyllenediamine (TEMED). The gel was mixed thoroughly and poured in gel cast immediately, which was then left for about 30 – 45 minutes to solidify. 4% stacking gel was prepared with 0.125 M Tris (pH 6.8), 4 % acrylamide/Bis Assay, 0.1% SDS, 0.1, APS, 0.001% TEMED. The staking gel was left for 1-2 hours for complete solidification. Gels were normally prepared 3 – 4 days in advance for better gel results which is kept at 4 °C.

4X tris-HCL/SDS pH8.8 (Tris base and 10% SDS) running buffer was used. Samples were poured in wells at a maximum of 20 µl using a gel loading tip. The electrophoresis started at 50-60 V and increased to 100 V when the dye front had reached the running gel. The estimated time of running 60-90 minutes. After running, the gel was extracted from the gel glass cast using a spatula to separate the glass and cut the edge of the gel to remove the stacking gel. The gel was then soaked in a tray containing transfer buffer for western blot (48 mM Tris, 39 mM Glycine, and 20% methanol, pH 9.2). The nitrocellulose membrane and filter papers were also soaked on transfer buffer as well. On the semi-dry

blot, a sandwich of gel, 0.2 µm Nitrocellulose membrane and filter paper as assembled as the with the assurance of no air bubbles contained within.

For gel staining, the gel was immersed in staining solution (40% methanol, 10% acetic acid, and 0.1% Coomassie brilliant blue) for 1.30 hours – overnight at room temperature. The gel was de-stained using a de-staining solution (20% methanol and 10% acetic acid) for 2-3 hours at room temperature until a clear ground was obtained. The gel image was recorded using gel doc

3.2.17.3 Immunostaining of Western Blot membrane:

To detect expressed protein, Western blot was traditionally used by utilizing specific antibodies to bind to expressed proteins. The nitrocellulose membrane was first rinsed with 1X PBS to the remaining of SDS and immersed in a fresh blocking solution containing 5% skimmed milk, 1X PBS, and 0.05% Tween20 and kept shaking overnight at room temperature. Later, the membrane was then washed with 1X PBS and 0,1% Tween20 to remove any residues of the blocking solution. The membrane was then immersed in a primary antibody solution (5% skimmed milk, 10% 10X PBS, and 1/1000 diluted primary antibody (6X –His Tag Monoclonal Antibody (HIS.H8), ThermoFisher) for overnight with shaking at 4 °C. Before immersing the membrane into a secondary antibody solution (5% skimmed milk, 10% 10X PBS, and 1/5000 diluted secondary antibody [Goat Anti-mouse IgG (H L)-HRP, Conjugate BIO-RAD]), it was washed with antisera buffer to remove any residues of primary antibody 3 washings each 5 minutes long. The membrane was kept in a secondary membrane shaking for at least 2 hours at room temperature or overnight at 4 °C. Finally, the membrane was washed with 1X PBS and 0.1% Tween20 for 15 minutes once and twice at 5 minutes with final wash for 15 minutes and rinsing with distilled water.

3.2.17.4 Develop Signal and Detection:

The specificity bound secondary antibody of the target protein can be detected using Enhanced Chemiluminescence (ECL) reagents (SuperSignal™ West Pico PLUS Chemiluminescence substrate, Thermo Scientific). ECL mainly contains two reagents stable peroxide solution and Luminol/Enhancer solution. A working solution was made by mixing an equal parts of stable peroxide solution and Luminol/Enhancer solution. 0.1 ml

working solution per cm^2 of membrane which was incubated for 5 minutes. The signal duration 6 – 24 hours.

3.3 Results:

In this chapter, *Ta WRKY53b*, *Ta WRKY19*, *Ta WRKY3* and *Ta WRKY3m* would be used for protein expression through the cloning into pGAPZαA vector backbone and expressed by *Pichia pastoris* yeast as an expression platform. The cloning of WRKY transcription factors and expression in *P. pastoris* would express desired proteins in its native structure. This would allow us to use WRKY proteins for protein-DNA binding assays using Electrophoretic Mobility Shift Assay (EMSA, Chapter 5).

3.3.1 Cloning of WRKY transcription factors:

Four TaWRKY transcription factors (*Ta WRKY53b* [1320 bp], *Ta WRKY19* [1404 bp], *Ta WRKY3* [727 bp], and *Ta WRKY3m* [727 bp]) were cloned into pGAPZαA which permits constitutive expression of recombinant genes from the *GAP* promoter. *Ta WRKY53b* and *Ta WRKY19* were synthesised and cloned into pGAPZαA between Xhol site in position 736 bp and *Xba 1* site in position 824 bp. Constructed plasmids were received as 4 µg freeze dried samples. Upon arrival, 20 µl of DNase and RNase free water was added to rehydrate it and 1 µl which is roughly 200 ng was used for transformation into *E. coli*. Transformed plasmids were then grown in low salt lysogenyl broth medium with 25 µg/ml Zeocin™ as a selection antibiotic. Transformed competent cells were then purified by reselection and grown on liquid LSLB medium with 25 µg/ml Zeocin overnight for plasmid extraction as shown in figure 4. The quality of plasmid extraction was tested through running extracted plasmid in Agarose gel. Four types of bands appear on each extract (nicked circular, linear, supercoiled, and circular single stranded). Out of all, the desired type of DNA from extracts was a highly condensed and with thick band of supercoiled DNA as appears in pGAPZαA/*Ta WRKY53b* sample 1 (Figure 3. 5). This indicates that high quality plasmids were isolated. Thus, sample 1 was selected for plasmid linearization either *BspH 1* (355 bp), *Avr1* (190 bp), or *BglII* (1 bp) restriction enzymes. When designing pGAPZαA/*Ta WRKY53b*, it was crucial to select restriction site that cuts only in a single site. So *BspH 1*, *Avr 1*, and *Bgl II* were selected for restriction digestion for linearization. *Avr I* and *BspH I* were located in α-factor of pGAPZαA. However, *BspH 1* restriction enzyme was first used at different plasmid concentration (5 µg and 7.5 µg) (figure 3.5). The linearization showed that there was an appearance of three bands. The first band was faint and located at around 6 kb was non-digested plasmid. The second thick band was around 4.4 kb and it was the linearized pGAPZαA/*Ta WRKY53b*. However, unexpected faint band appears down below and its size was around 400 bp (Figure 3. 5).

The digestion of all plasmids using *BglII* showed clean digestion with no additional bands (Figure 3. 6). All of which digested plasmids were used for transformation into *P. pastoris* for small scale protein expression.

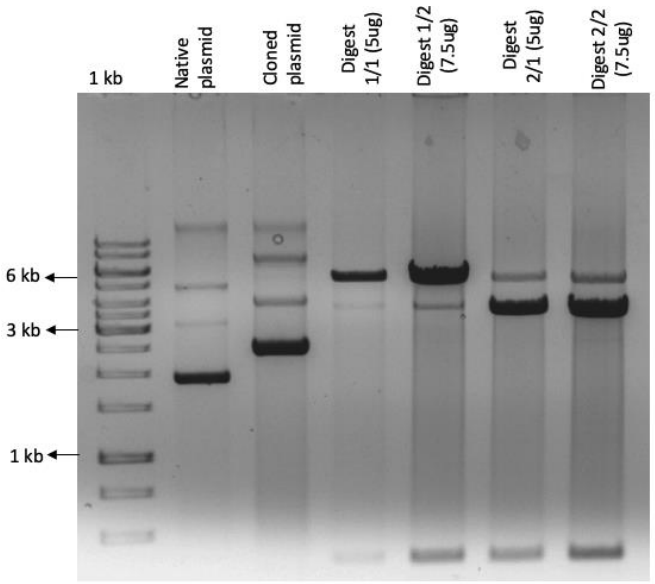


Figure 3-4 1% Agarose gel of linearized plasmids with *BspH I* restriction enzyme. A) pGAPZαA/*TaWRKY53b* digested plasmid. Native plasmid contains no insert with no cut (3.1 kb); cloned plasmid is pGAPZαA with *TaWRKY53b* with no cut (4406 bp); digest 1/1 colony (5 µg) is pGAPZαA with *TaWRKY53b* insert; digest 1/2 colony (7.5 µg) is pGAPZαA with *TaWRKY53b* insert; digest 2/1 colony (5 µg) is pGAPZαA with *TaWRKY53b* insert; digest 2/2 colony (7.5 µg) is pGAPZαA with *TaWRKY53b* insert. . Non-digest is an indication of plasmid with insert that had not been cut. 5 µg and 10 µg is the total concentration of plasmids used for linearization. Black arrows indicate the position the size of ladder that is correspondent to constructs. 1kb DNA ladder (Thermofisher).

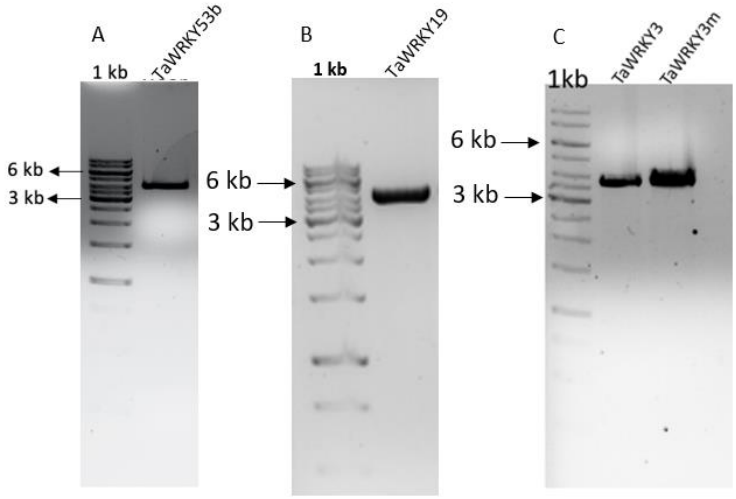


Figure 3-5 1% Agarose gel of linearized plasmids. A) Bgl II linearization enzyme at all constructs. A, pGAPZαA/*TaWRKY53b* (4406 bp, B, pGAPZαA/*TaWRKY19* (4488 bp); C, pGAPZαA/*TaWRKY3* (3774 bp); pGAPZαA/*TaWRKY3m* (3774 bp) plasmids. Black arrows indicate the position the size of ladder that is correspondent to constructs. 1kb DNA ladder (Thermofisher).

Once fragments of recombinant plasmids were linearized and confirmed by agarose gel (Figure 3. 6 A, B, and C). Each linearized construct was then PCR purified using QIAquick PCR purification kit to remove non-desired contaminants such as restriction enzymes, buffers etc. samples were eluted using elution buffer and then concentrated using speed vacuum to remove the excess of water. This to prepare each sample (linearized construct) for transformation into *Pichia* cells. The desired concentration for transformation was 5 to 7 µg in 5 µl.

3.3.2 Detection of expressed proteins:

Pichia competent cells were prepared as recommended by Invitrogen protocol and stored in -80 °C. Linearized constructs were integrated into the host chromosome using *Pichia* EasyComp kit as recommended by Invitrogen manufacturing protocol. The transformation/ homologous recombination of linearized constructs was mainly based on a heat shock protocol. Transformed constructs were then streaked onto YPG containing Zoecin™ as each construct carries Zoecin™ resistance gene as selection gene. 3 – 4 days incubation at 28 °C. Five to six colonies were the picked up and grown in 50 ml YPG in baffled flask and incubated for 72 hours at 28 °C. Later on, cells were collected and centrifuged to remove cells and preserve broth media for protein precipitation using TCA method. This was to be used in SDS-PAGE and Western Blot.

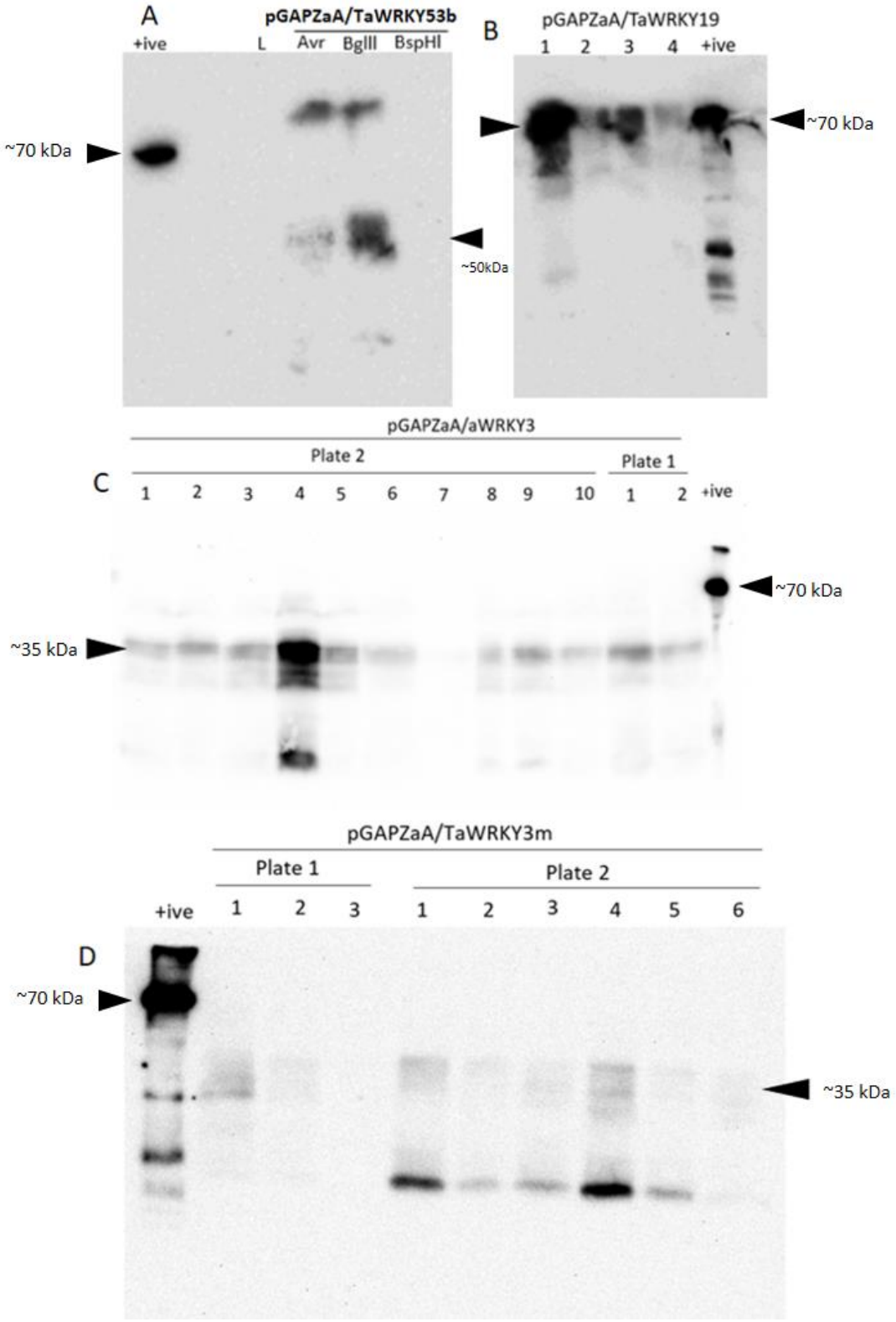


Figure 3-6 TaWRKY constructs transformed into *Pichia pastoris* (X-33) and precipitated and visualized on western blot. A; pGAPZαA/*TaWRKY53b* western blot with different restriction enzyme digestion and protein expression (47.4 kDa), B; pGAPZαA/*TaWRKY19* different colonies protein expression (50 kDa), C; pGAPZαA/*TaWRKY3* different colonies protein expression (24.9 kDa), D; pGAPZαA/*TaWRKY3m* different colonies protein expression (24.9 kDa). Positive control (+ive); SF16:BSA (70 kDa). Immunodetection at western blot was carried out using 6x His-tag antibodies. Arrows indicate the expected recombinant fusion protein.

During expression of protein, there was a difficulty of expressing fusion proteins from cells that had been transformed with constructs linearized using *Bsp HI* restriction endonuclease. Thus, all pGAPZαA/*Ta WRKY53b* constructs were linearized with *AvrI* and *Bgl II* restriction endonucleases as well. As clearly shown in Figure 3. 7. A, an appearance of differential expression of protein among all constructs that had been linearized with *Bsp HI*, *Avr I* and *Bgl II*. Among all constructs there was no expression of fusion protein detected with constructs digested with *Hsp HI*. However, *Avr I* restriction endonuclease showed that there is a low expression. Most expressive constructs were apparent with those that had been linearized with *Bgl II*. Thus, constructs that had been linearized with *Bgl II* restriction endonuclease were selected for the analysis of protein expression levels and to determine the optimal time to harvest. With regards to *pGAPZαA/Ta WRKY19*, *pGAPZαA/Ta WRKY3*, *pGAPZαA/Ta WRKY3m*, different expression was clearly noticed using different transformed colonies. Colony 1 in *pGAPZαA/Ta WRKY19* was shown the most expressive protein. Hence it was selected. In terms of *pGAPZαA/Ta WRKY3* and *pGAPZαA/Ta WRKY3m*, colony 4 at plate 2 at each was found the most abundant expression of proteins. Thus, they had been selected (Figure 3. 7, C and D, respectively). When looking at *Ta WRKY3* and *Ta WRKY3m* expressed proteins in Figure 3. 7, there was an apparent of low band at each precipitated protein. This is an indicative a truncated protein.

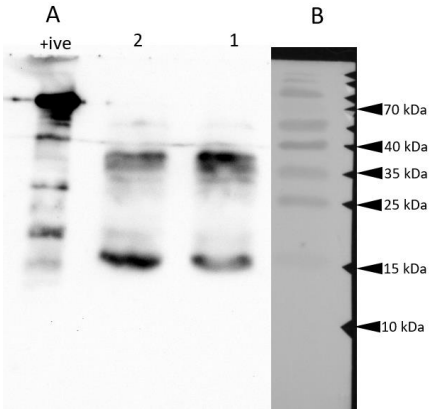


Figure 3-7 *Ta WRKY3* Expressed and TCA precipitated proteins from colony 4 for size determination. A; bands appeared from western blot chemiluminescent assay, B; nitrocellulose membrane picture western blot at the same position visualizing PageRule prestained protein ladder.

Glycosylation of proteins was apparent in expressed proteins, as shown by a shift in predicted molecular weight. The calculated molecular weight of *Ta WRKY19* was 50 kDa. However, it appears similar to or more than 70 kDa in comparison to positive control (70 kDa). Figure 3.8 shows a comparison between western blot membrane under

chemiluminescent assay and the nitrocellulose picture of the same run. This was to determine the size of expressed *TaWRKY3* protein using the pre-stained ladder. *TaWRKY3* calculated molecular weight was 24.9 kDa. When its western blot bands were compared to the membrane image showing the pre-stained ladder, it showed that the molecular weight of *TaWRKY3* had shifted to above 35 kDa. Truncated proteins were 15 kDa.

3.4 Discussion:

In this chapter we had focused on constructing and cloning selected *TaWRKYs* (*TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and *TaWRKY3m*) and transform into *Pichia* pastoris for expressing fusion proteins. *TaWRKY* genes were aimed to be cloned into pGAPZαA which is plasmid vector for *Pichia pastoris*. Two different methods were applied for cloning. One of which was firstly was amplifying fragments of *TaWRKY*s as well as pGAPZαA vector and combined together using Gibson Isothermal assembly (appendix B). For that method, difficulties encountered during amplification DNA fragments. For *TaWRKY53b*, a repetitive amplification of DNA fragment, using specific primers, had been conducted with many modifications at amplification conditions resulted to amplification of non-specific bands. Studying the complexity of *TaWRKY53b* using IDT (Integrated DNA technologies web site) showed that *TaWRKY53b* was highly complex. The complexity occurs due to variations of GC content composition at different parts of DNA (80% at 84 bases and 62% at position 488 to 1287 bp), presence of 45.9% of the overall *TaWRKY53b* sequence had one or more repeated sequences greater than 8 bases, and hairpin exists at 167 and 738 bp positions. This had given us an insight of our inability of amplifying *TaWRKY53b* using designed primed through conventional PCR conditions (appendix B).

With the development of WRKY plasmids constructs, it was necessary to transform into *pichia* genome. Transformation of plasmids into competent cells can be said as a straight forward thought because they can be transformed in their circular form easily into cells. However, *Pichia* cells it is quite different. In order to transform constructs into *pichia*, it was crucial to linearize plasmid constructs prior to transformation which would allow for constructs to be transformed into *pichia* through homologous transformation. Generally, *Bsp HI* (356 bp) and *Avr I* (191 bp) are the two recommended restriction sites at *pGAPZαA* plasmid backbone for linearization (figure 3.7). Each enzyme cut once in the GAP promoter region to linearize the vector. Using bioinformatics analysis, it was necessary when cloning plasmids that each insert gene does not contain any of these two restriction sites. This was for only linearizing such plasmids at a singular restriction cut that only present on *pGAPZαA* for successful transformation and protein expression. In theory, each restriction site serves as a site of cut and their specificity at recognizing sites is very high (Burrell, 1993). Thus, digestion (cutting) *pGAPZαA* backbone with restriction enzyme should not interfere to the expression of proteins. However, otherwise was found.

Initial attempts for linearizing constructs was performed at *Bsp HI*. Gel electrophoresis, in terms of *pGAPZαA/TaWRKY53b*, results showed a band located at 4406 bp which is the right size. However, a faint band that was apparent at roughly 375 bp (Figure 3. 5). This an indicative of non-specific digestion of the plasmid. With Downstream processing analysis using western blot showed that there was no expression of proteins using *Bsp HI* site of cut. This could be justified by the occurrence of 'Star Activity' which is, under non-standard reactions conditions, a cleavage of DNA sequences no identical to its defined recognition site. High glycerol concentration (>5% v/v), high concentration enzyme-DNA ratio, non-optimal buffer, prolonged reaction time, presence of organic solvents (DMSO, ethanol, ethylene glycol, dimethylacetamide, sulphalane), or Mg^{2+} substitutions with other divalent cations (Mn^{2+}, Cu^{2+}, Co^{2+}, Zn^{2+}) are conditions that contribute to star activity. With the use of *BspH I* restriction enzyme to linearize pGAPZαA/*TaWRKY53b*, two factors that had raised star activity issue are high concentration enzyme/µg DNA ratio as well as prolonged reaction time. This issue could be solved by reducing the concentration of DNA within the reaction mixture and reduce reaction time. However, our aim was to linearize 5-10 µg of plasmid DNA to be used for a successful transformation. Thus, alternative restriction enzymes could be used to overcome Star Activity. *Bgl II* was used for plasmid linearization and results showed a clear linearization (Figure 3. 6). Our investigation on the impact of restriction site showed that protein expression results through western blot showed differences on protein expression at each site of cut for all pGAPZαA/*WRKY* constructs. *Bgl II* construct cut showed higher expression compared to *Avr I* whereas *Bsp HI* showed the least or no heterologous expression at all constructs. With extracting intracellular proteins to examine whether transformed clones had been expressing intracellularly. Western blot for tagged proteins show also negative. This indicating that constructs linearized with *Bsp HI* restriction enzyme had a major impact protein expression by deactivating its ability to express proteins in *Pichia*. This might be justified by the exposure of high concentration of restriction enzyme to obtain high concentration of linearized plasmid at extended time which led to cut at sites that might differ with only one base pair from the canonical site (Burrell, 1993). Thus, it was important to select a linearization site of cut for obtaining high expression and the best candidate was *Bgl II* (figure 3.7).

Studies on DNA binding selectivity of *WRKY* transcription factors into *WRKY* domains had been ongoing for the past 20 year but limited. However, many of such studies were

using, for studying WRKY DNA-binding activity, bacterial system for expressing fusion proteins. In *in vitro* tests for heterologous protein production, *E. coli* was the considered as the first and most widely used as an expression system for recombinant proteins. This was due to its many advantages: i) rapid protein expression; ii) high yields; iii) fast mass production and cost effective. However, there are a drawback of this system: i) production of unglycosylated protein; ii) production of proteins with endotoxins; iii) proteins with disulphide bonds that are difficult to express; iv) formation of acetate in media resulting in cell toxicity. In addition, proteins are produced as inclusion bodies within the cell which are often inactive, insoluble and require refolding. To obtain structurally active protein, inclusion bodies must be extracted from the cell and solubilized by in vitro downstream denaturing and renaturing processes. This major disadvantage is the uncertainty of protein refolding to its native structure (Martínez-Alarcón;Blanco-Labra and García-Gasca, 2018). These issues, in terms of protein expression through bacterial system, had occurred in many studies. For example, Ciolkowskie et al (2008) cloned five *At*WRKY TFs into bacterial expression vectors pGEX2T and pQE30 (GST fusions and epitope tagged variants, respectively) in *E. coli.* Two technical difficulties were mainly faced with using this system i) impact of expressed proteins on bacterial growth; ii) protein purification. For bacterial growth, they had found that zinc homeostasis in bacteria was negatively influenced by expressing WRKY protein. They had suspected that due to the presence of Zinc finger motif in WRKYs might cause this problem. Even with the addition of exogenous zinc into the culture could not relived this problem. The second issue was the expression of protein with inclusion bodies which led to use total soluble bacterial protein extracts. This was justified by the impact of denaturing and renaturing steps leads to protein misfolding resulting in loss of its ability of binding to W-box. Even though with affinity purification under mild conditions of epitope tagged WRKY proteins had proved problematic as well. Thus, they finally compromised by using total soluble bacterial protein extracts for DNA-binding experiments rather than purified recombinant protein. This issue also occurs in many studies such as expressing *Aa*WRKY1 protein using pRT-30a vector, (Ma *et al.*, 2009) *TaWRKY2* and *TaWRKY19* using pMAL-c2X vector (NIU *et al.*, 2012), *TaWRKY10* using pET32a vector (Wang *et al.*, 2013a), *TaWRKY53* using pET SUMO vector (Van Eck *et al.*, 2014), *Jc*WRKY using pET28a vector (Agarwal;Dabi and Agarwal, 2014), *AtWRKY50* using pGEX-KG vector (Hussain *et al.*, 2018), *MdWRKY9* using pGEX-6P-1 vector (Zheng *et al.*, 2018), *VviWRKYIIa_3* and *VviWRKYIII_3* using pTrcHisA vector (Romero *et al.*, 2019), and *GhWRKY27* using pET-28a (+) vector (Gu *et al.*, 2019). All of which have in common is that the use of prokaryotic system (*E. coli*). They

also had faced the same issues of fusion protein expression purification mentioned above. However, this study shows that WRKY transcription factors can be expressed using *Pichia pastors.* This is represented by the ability of Pichia cells to express recombinant proteins of candidate TaWRKYs (figure 3.7 A,B, C, and D). With their expression, there was no impact of cloned *TaWRKY*s on Pichia cells with regards to Zinc-homeostasis. *TaWRKY* proteins were secreted into the media which were collected for analysis by western blot for His-tagged *TaWRKY* proteins. No need for cell lysis to release recombinant proteins.

In conclusion, *TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and its mutant form (*TaWRKY3m*) were designed to be cloned in frame of pGAPZaA for recombinantly expressing proteins. Each protein contained His-tag at its C-terminus and that was very useful for protein expression determination and would be very useful for protein purification at later stages. Candidate TaWRKY coding sequences were synthesised and cloned into pGAPZαA plasmid vector for protein expression.

BglII restriction enzyme was used to linearize each construct as it was found to be best in linearization without causing star activity. Recombinant protein expression for each construct was tested post-transformation into *Pichia*. The outcome results illustrated that each construct successfully expressed recombinant protein at a small-scale level, and western blot results approved this.

Next, expressing constructs could be moved into large-scale protein production. This is where cells carrying constructs would be inoculated in a bench-top bioreactor and run for 92 hours. It is then could be collected and passed through chromatography columns for purification (chapter 4).

4. Chapter 4: Bench-top expression of *Ta*WRKY transcription factors using *Pichia pastoris* and purification

4.1 Introduction

From previous experimental chapter 3, wheat WRKY transcription factors were cloned into pGAPZαA plasmids carrying 6x His-tag at its C-terminus and transferred into *P.pastoris*. Following successful cloning, *TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and *TaWRKY3m* were expressed using shake flask. The expression of each was confirmed via western blotting with 6x His-tag monoclonal antibody binding to 6x His-tag present in the expressed wheat WRKY protein.

In this experimental chapter, the four recombinant proteins were expressed in a selection of media capable of maintaining high cell densities during fermentation in benchtops bioreactors and recovered in a four-stage purification and concentration pipeline to recover correctly folded and functional transcription factors. The four recombinant proteins (*TaWRKY53b, TaWRKY19, TaWRKY3,* and *TaWRKY3m*) were expressed in high density medium volume fermentation bioreactors and recovered in a four-stage purification and concentration pipeline to recover correctly folded and functional transcription factors. Our strategy of purifying *Ta*WRKY expressed proteins was through; 1) Hydrophobic interaction Chromatography (HIC), 2) Nickel column (Ni-column), 3) protein tube dialysis, 4) column ultrafiltration and concentration, and 5) Bradford assay for protein concentration.

The overarching aim was to produce full-length recombinant wheat WRKY proteins and validate their binding to promoters of target genes. Chapter 3 demonstrated the ability of *Pichia pastoris* to express candidate wheat WRKY proteins in a shake flask. For a protein to DNA binding assays, obtaining a high concentration of TaWRKY proteins is crucial. Previous work in the literature had demonstrated a variable amount of WRKY proteins for the binding assays. Van Eck *et al.* (2014) and Hussain *et al.* (2018) used 0.5 µg of purified *TaWRKY53* and *AtWRKY50*, respectively. For *TaWRKY2* and *TaWRKY19*, 2- 4 µg was used in the binding assay (NIU *et al.*, 2012). Cheng *et al.* (2019) used 7.7 µg of purified proteins. In relation to this work, high amounts of candidate proteins were not recoverable in the shake flask. As the requirement for electrophoretic mobility shift assay

to use high protein concentration, it was necessary to scale up the production of recombinant TaWRKY protein expression. This was achieved by a high-density fermentation unit that can handle up to 5 L of culture media as a laboratory test (Cregg *et al.*, 2000).

For bench-top fermentation (large scale), Basal Salts Media (BSM) is considered as a most widely used media. It is also the main media's recommended by Invitrogen™ as the manufacturing company and supplier for *P.pastoris*. Therefore, *Ta*WRKY proteins were initially expressed in BSM media and purified through our proposed purification strategy.

With regards to TaWRKYs protein purification, chromatography system could be used. Each of candidate TaWRKY was cloned in pGAPZaA to contain His-tag at its C-terminus. The aim for performing hydrophobic interaction chromatography, initially, was to remove host cell impurities, high molecular weight aggregates as well as to concentrate the sample for being used in Ni-column. With such method, hypothetically, proteins at high salt concentration adsorb to HIC matrix resin at the stationary phase in which later on could be eluted using gradient elution through decreasing salt concentration. *TaWRKY53b* and *TaWRKY19* expression and purification were following this protein strategy which were then purified using Ni-column.

For approaching our aim in and purifying *TaWRKY3* heterologous proteins and its mutant form, it was thought to purify proteins directly from Ni-column. Another issue raised with this method was salts precipitation when mixing Ni-column sample buffer with the fermentation broth. The difference within pH between sample buffer (pH 8.0) and fermentation broth (pH 4.6) was found to cause this issue in terms of salts precipitation. Thus, different media recipe for bench-top fermentation was selected as an alternative medial for expression and purification. Our criteria for such media was its ability to express *Ta*WRKY proteins in levels comparable to BSM and can with stand pH conditions more than 8.0 for protein purification in Ni-column as one step purification.

M3 was firstly proposed by (Zhang;Sinha and Meagher, 2006) along with other recipes for protein expression using *P.pastoris* through fermentation. Their study concerned with

phosphate precipitation in BSM at pH above 5.5. They had investigated multiple modified media containing different glycerophosphates as a source of phosphorous (sodium (Na_2GP), potassium (K_2GP), calcium (CaGP), and magnesium (MgGP) glycerophosphates) to eliminate precipitation. Throughout their results, Zhang *et al.* (2006) excluded CaGP and MgGP due to their low solubility (only around 2% w/v) at room temperature which was considered to be an issue for use in fermentation. Thus, these glycerophosphates were excluded by the authors. On the other hand, Na_2GP and K_2GP were soluble above 65% w/v at room temperature as well as no appearance of precipitation up to pH 10.5 and 7.8, respectively. For M3 capability to remain soluble with no precipitation at high alkaline conditions, Na2GP was then selected. For comparison, BSM media tends to precipitate at pH 8.0 whereas M3 had shown no precipitation at pH above 10.0.

Zhang;Sinha and Meagher (2006) had also observed a comparable growth level between M3 medium and BSM in terms of cell growth and protein production. In their study, they had used a methylotrophic *P.pastoris* GS115 strain to express α-galactosidase recombinant protein. This strain uses methanol utilization plus (Mut^+) to express the recombinant protein. The cell growth in M3 was 498 g wet cell/L and BSM medium was 433g wet cell/L which is 15% higher in M3 medium than BSM medium. In terms of α-galactosidase recombinant protein, 12 U/mL was found in M3 medium and 3.8 U/mL in BSM which is 215% higher in M3 medium than BSM medium throughout 30 hours incubation time. They had summarised their study as the capability of *P.pastoris* strains to assimilate sodium glycerophosphate which can be employed, reliably, as phosphorous source for the growth of cells as well as the production of recombinant proteins. Thus, M3 medium containing sodium glycerophosphate was mainly suggested for use benchtop fermentation for expressing *TaWRKY3* and its mutant form recombinant proteins in *P.pastoris* as well as purification.

Aim and objectives:

Our aim for this experimental chapter was to obtain a purified recombinant expressed wheat WRKY transcription factor protein using *Pichia pastoris* as a eukaryotic expression system.

- Determine optimal expression conditions of all wheat WKRY using basal salts medium.
- Enrich expressed proteins using hydrophobic interaction chromatography.

- Specifically purify wheat WRKY proteins using immobilized metal affinity chromatography (IMAC).
- Compare basal salts media and M3 media on recombinant protein expression.
- Optimise a purification strategy for 1 step purification of proteins by IMAC using M3 media.

4.2 Materials and methods:

Following the small-scale protein expression and confirmation using western blotting, the protein contained in culture medium would undergo several processes to get a purified protein. An overview of the process is detailed (in the chapter introduction) indicating that proteins would be largely being expressed through large scale fermentation followed by supernatant collection using centrifugation. Once it has been collected, proteins would first undergo hydrophobic interaction chromatography and followed by Nickel-column chromatography to collect only His-tagged proteins. As final product, his-tagged proteins would be dialyzed and concentrated using concentration columns. .

4.2.1 Fermentation Basal Salt Medium (BSM)

For large scale fermentation, positive colonies were inoculated as starting culture into 4 baffled flasks each contains 50 ml YPG and grown with shaking (250 rpm) at 30 °C for 72 hours. Each of transformed constructs were inoculated into a 1L benchtop fermentation unit containing a sterile Basal salts media (85% phosphoric acid [H3PO4, 26.7ml], calcium sulphate [CaSo4, 0.93g], potassium sulphate [K2SO4, 18.2 g], Magnesium sulphate 7-hydralate, [MgSO4.7H2O, 14.9g], Potassium hydroxide [KOH, 4.13g], and glycerol [40ml] as described in " *Pichia Fermentation process Guidelines*" of Invitrogen Corporation (Corp., 2002). 1 ml of Antifoam was added afterwards and autoclaved at 121 °C 15 lbs pressure for 20 minutes. Post autoclaving, the fermentation vessel was connected to the fermentation unit and adjusted temperature to 30 °C. The medium pH was adjusted to 5.0 with 30% ammonia (NH_4OH). Trace elements (filter sterilized PTM1 contains, Cupric sulphate-$_5H_2O$ [$CuSO_{4.5}H_2O$, 24.031mM], Sodium iodide [Nal, 533.725 µM], Manganese(II) sulphate-H2O [$MnSo_{4.}H_2O$, 19.867 mM], Sodium molybdate-$_2H_2O$ [$Na_2MoO_{4.2}H_2O$, 971.251 µM], Boric acid [H2BO2, 323.468 µM], Cobalt chloride-$_6H_2O$ [CoCl2.6H2O, 6.932 mM], Zinc chloride [ZnCl, 146.750 mM], Ferrous sulphate-$_7H_2O$ [$FeSO_{4.7}H_2O$, 427.891 mM], Biotin [818.599 µM], and Sulphuric acid [H_2SO_4, 50.979 mM]) were added once the media was saturated with O_2 and maintained at 30% dissolved oxygen. PTM1 salts were also added to the 50% glycerol feeding solution (4.35 ml/L). Once fermentation initiated by inoculating the starting culture into the vessel. The medium was kept stirring and the cultivation was standardized at 30% dissolved oxygen, pH 4.5 to 5.0 and at 27 °C using input settings. The glycerol feed was first set at 5 ml/h on day one (120 ml) and increased to 10 ml/h on day two (240 ml) and 20 ml/h the third day (480 ml). Fermentation reaction was terminated once glycerol medium content was depleted

and this was addressed by the increase of dissolved oxygen above 30% and the reduction of the rotor speed to its minimum (~200 rpm).

Culture media was collected and centrifuged at 7,000 x g for 30 minutes at 4 °C to collect culture broth. 50 ml of each batch was collected and the proteins were precipitated using TCA-precipitation method [details at previous chapter] for testing protein expression using western blot.

4.2.2 M3 Fermentation medium composition (Zhang, Sinha *et al.* 2006)

An alternative fermentation medium was used for expressing protein named as M3 which was developed by Zhang, Sinha *et al.* (2006). M3 medium composed of two parts for preparation. Part A contains CaSO4.4H2O (1g/L), K2SO4 (27.3 g/L). MgSo4.7H2O (16.6 g/L), (NH4) SO4 (11.1 g/L), and glycerol (44.4 ml) in 900 ml. Part B contains only 100 ml Sodium glycerophosphate hydrate (12g/L). 1 ml of Antifoam was added to Part A. Both Parts were autoclaved separately and let to cool down and mixed together using a feed bottle to the fermentation vessel. The vessel was let to be aerated at 600 rpm for a 2 of hours. 4.4 ml PTM1 was then injected into the vessel and followed with an inoculation of 200 ml seed culture. Dissolved oxygen, glycerol, and input was set the same settings as BSM fermentation. 30% of ammonia input was also used to maintain the pH at 4.6. The fermentation process was kept the same as BSM though out the 72 hours of fermentation. Prior harvesting, the pH was raised to 8.0 using 30% ammonia.

Culture media was collected and centrifuged at 7,000 x g for 30 minutes at 4 °C to collect culture broth. 50 ml of each batch was collected and its proteins were precipitated using TCA-precipitation method [details at previous chapter] for testing protein expression using western blot.

4.2.3 Medium and proteins stability tests:

To measure the extent of protein degradation, both culture broth supernatants were used in this experiment. 5 time points was selected (0, 12, 24, 48, and 72h). Control samples was taken from each supernatant corresponding to points and were subjected to the same conditions without the use of PhenylMethylSulfonyl Fluoride (PMSF). The experiment was run under room temperature and was subjected to shaking using a rotor during

experimental time course. Samples were collected and conducted TCA protein precipitation for western blot as described previously.

4.2.4 Hydrophobic interaction chromatography purification:

To purify proteins, culture broth was run through hydrophobic interaction chromatography (HIC) using Phenyl Sepharose column (GE Healthcare). Prior running sample to HIC, 4M sodium chloride was added to the supernatant and filtered using 1.6 µm, 1.2 µm and 0.7 µm. process was done at 4 stages. Firstly, the column was saturated by running a filtered 4M NaCl at 2ml/min until a baseline for both conductivity and absorbance (280 nm) was achieved. The sample load (~900 ml) containing expressed proteins was then loaded into the column at the same speed until almost its volume finished. This would show that the UV-absorbance increase at the optimum. The column then washed with 4M NaCl until the UV absorbance returned to baseline levels. To elute proteins, a gradient between 4M NaCl and Di-H_2O was set up 80 ml elution. Once protein fully eluted, the UV absorbance would start to increase. Indicating that protein elution starts, and fractions of proteins were collected and stored at -20 °C until use.

4.2.5 Nickel column protein purification:

The second purification step of proteins was by Nickel column purification (Ni-column) (IMAC SepFast, Biotoolomics). This was to only bind His-tagged protein and remove untagged proteins. Three types of buffers were prepared in advanced, 4x Sample buffer (lysis buffer) [200 mM NaH2PO4, 300 mM NaCl and 10 mM Imidazole], wash buffer contains [200 mM NaH2PO4, 300 mM NaCl and 20 mM Imidazole], and Elution buffer [200 mM NaH2PO4, 300 mM NaCl and 250 mM Imidazole]. The pH of each buffer was set at 8.0. HIC eluted products (sample) was mixed with sample buffer to make 1x sample buffer. The reaction run was done at 5 steps. Equalization step which is equalizing the column with 1X of sample buffer (without proteins). The UV absorbance would show a plateau (~ 0 mAu). Then 1X sample buffer containing sample proteins would then loading into the column and this would be clearly be shown on the rapid increase of UV absorbance. Once the sample is almost fully loaded into the column, 1X buffer would be loaded to wash the column from loosely contaminants until the UV absorbance drops down. A wash buffer was used as a second wasting step to remove the remaining of strong contaminants leading to the UV absorbance to be completely flat down. Finally, an elution buffer was used to elute His-tagged proteins that was attached to the Nickel. Once loaded

to the column, the UV absorbance would rapidly increasing leading to His-tagged proteins to be released and eluted. All fractions were collected and stored at -20 °C until use.

For *TaWRKY3* and *TaWRKY3m* purification, a glycerophosphate based buffer was used to perform the purification process. 4x buffer of glycerophosphate buffer was made containing glycerophosphate (200 mM), NaCl (1.2 M) and Imidazole (40 mM) pH 8.0. Fermentation supernatant was mixed with the buffer to make a 1 X buffer. Phenylmethylsulfonyl fluoride (PMSF) was added to the sample buffer to make 1 mM concentration and filtered using 1.6 µm, 1.2 µm and 0.7 µm.

Equilibration buffer was made as 1 X of glycerophosphate buffer which was used to equilibrate Nickel column. The sample was then loaded and circulated through the column overnight at room temperature. The column was then washed multiple times using sample buffer and the wash buffer. Proteins were then eluted with Elution buffer containing 50 mM glycerophosphate, 300 mM NaCl, and 250 mM high purity of Imidazole.

4.2.6 Protein dialysis and ultrafilteration

Once protein were fully eluted, they were dialysed to remove the high concentration of Imidazole. A dialysis membrane was used using 8,000 molecular weight cut-off (MWCO) to remove excess of salts and Imidazole. Eluted proteins were loaded into dialysis membrane and sealed properly and submerged in a d-H2O using magnetic stirrer for two days with a regular change daily.

Proteins were finally concentrated using Pierce™ protein concentrators, PES column (10K MWCO). The protein sample was loaded into the concentration sample chamber and placed into 4 °C rotor and centrifuged at 4,000 x g for 15 minutes. Concentrated proteins was then collected and loaded into a 2 ml Eppendorf tube and stored at – 20 °C.

4.2.7 Protein concentration determination

The concentration of all purified proteins were done via Bradford Assay (Sigma, USA). The reagent used in this method was Bradford reagent solution (Sigma Aldrich, B6916).

The spectrophotometer (SpectraMax) was set to read at 595 mm 15 minutes prior reading and this to equilibrate the instrument. A standard of Bovine Serum Albumin (BSA) proteins was prepared in water. Standards was made with an eight incremental concentration from 0 µg to 2.50 µg. The assay was performed in a 96 well plate. 10 µl of each standard was loaded in a triplicate manner and mixed with 150 µl ddH_2O and 40 µl Bradford Reagent. Protein Samples were loaded in 1 µl and mixed with 159 µl ddH_2O and 40 µl Bradford Reagent. All proteins were loaded also in triplicate manner. The plate was then covered with a parafilm and mixed well using a vortex. The parafilm was removed and air bubbled were assured to not being formed. The plate was placed on the SpectraMax and reading the absorbance was started. Absorbance reading was collected and using an Excel sheet the concentration of proteins was determined through the standard curve chart trend line equation displayed.

4.3 Results:

4.3.1 Basal salts media expression:

pGAPZαA/*Ta*WRKY vectors that were transformed into *P.pastoris* were previous grown at small-scale to confirm protein expression (Chapter 3). Western blot analysis confirmed expression of the desired proteins from the *Pichia* clones after which they were grown large scale in bench-top fermenters for large scale protein expression and purification. At this stage, two types of media were used to successfully express *Ta*WRKY proteins.

Centrifuged supernatant from each fermentation run was collected and stored in -20 C until processed. Each recombinant protein was purified firstly with hydrophobic interaction chromatography (in case of *Ta WRKY53b* and *Ta WRKY19*) and Ni-column for purifying specifically binding His-tagged proteins which was used for *Ta*WRKY proteins.

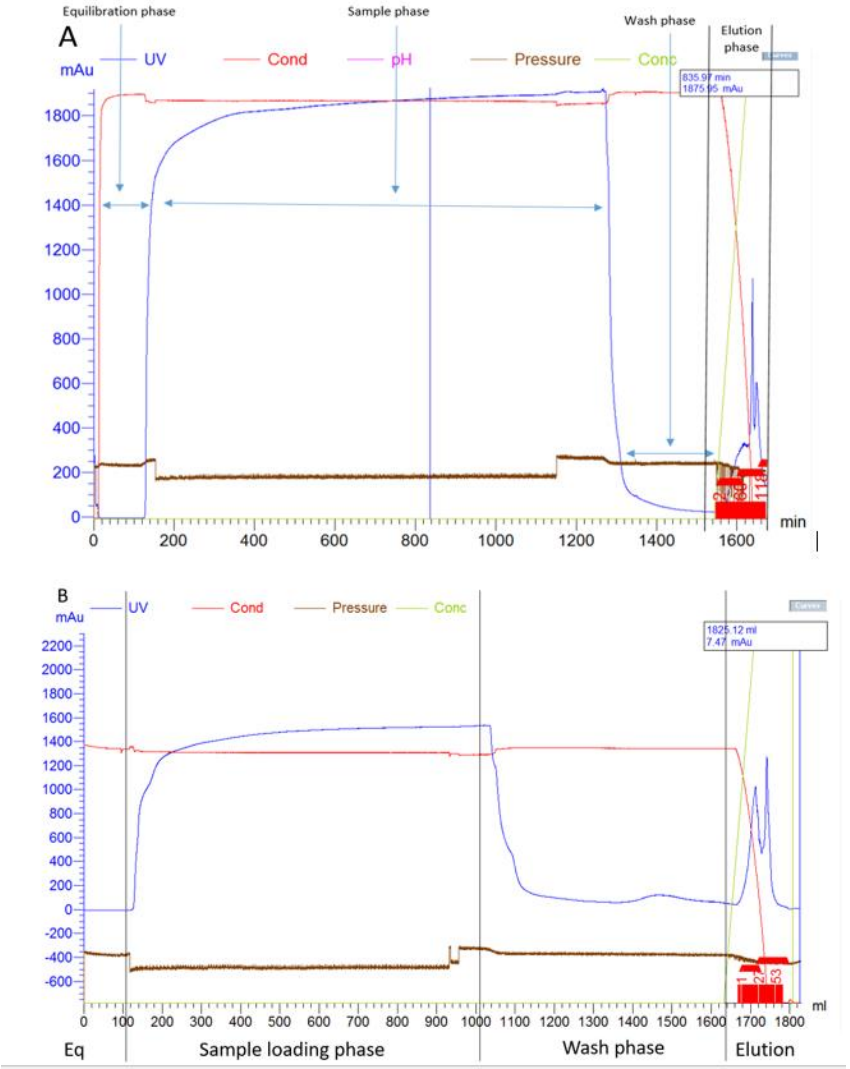


Figure 4-1 *Ta WRKY53* and *Ta WRKY19* purified proteins hydrophobic interaction chromatography, A and B respectively. Eq is equilibration phase. mAu indicate milli-Absorbance unit, Cond; conductivity; pressure, and Conc indicate concentration. Both *Ta WRKY53b* and *Ta WRKY19* were expressed in BSM.

Following fermentation in BSM, initial purification of *TaWRKY53b* and *TaWRKY19* was performed via hydrophobic interaction chromatography. As shown in Figure 4.1 (A and B) each purification run contains 4 distinct phases. Equilibration was the phase of running 4M NaCl buffer into the column to calibrate the column to be ready for sample loading. The sample of fermentation supernatant was adjusted to 4M NaCl. Sample loading phase was the sample containing 4M NaCl running through the column for protein binding. As soon as the sample started loading through the column, the absorbance (mAu) started to rapidly increase until reaching plateau. Once the whole sample was completely loaded into the column, then washing phase was initiated using 4M NaCl to remove weakly bound or associated proteins. During this stage of the purification the absorbance of each sample rapidly decreased until the absorbance baseline is achieved. This indicate that much of contaminating, unbound protein was removed from the column. Once the column was fully washed, then elution phase was initiated by 4M NaCl:ddH_2O gradient (100% to 0%) until the sample fully eluted from the column and this is clearly shown in figure 4.1A and B. As it clearly seen that two adjoining peaks during elution in both *TaWRKY53b* and *TaWRKY19* (Figure 1 A and B). After the elution of these peaks the absorbance decreased and flattened to a plateau, indicate that the majority of bound proteins were eluted through the column. Each of eluted products were collected and were stored at -20 °C until used in IMAC purification.

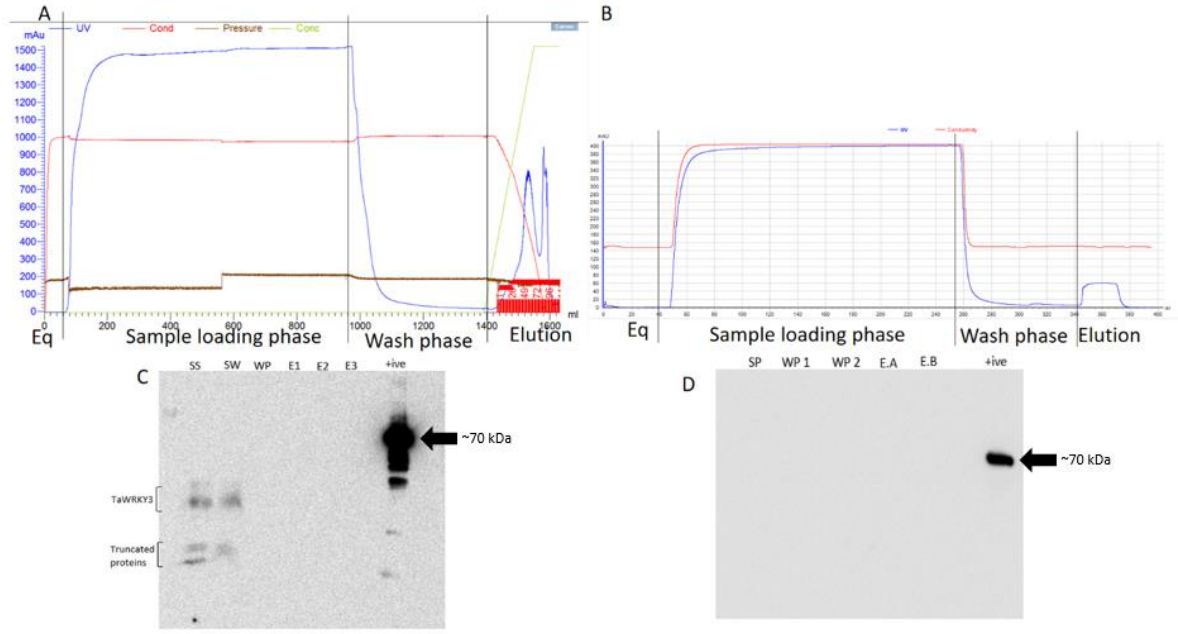


Figure 4-2 *TaWRKY3* protein expressed using BSM media and subjected to purification. A; HIC purification, B; Ni-column purification, C; western blot of HIC each purification fractions, D; western blot membrane of Ni-column fractions. SS; Sample phase, SW; Sample Waste, WP; Wash phase, E1-E3; fraction of elations. SP; Sample phase, W1-W2; Wash phase, EA-EB; Elution fractions.

TaWRKY3 fermentation run was initially expressed using BSM medium. The total culture medium was collected and centrifuged to collect the culture broth for protein purification. As with both *TaWRKY53b* and *TaWRKY19*, HIC was the first purification step and the filter culture broth was adjusted to 4M NaCl before the chromatography step. Figure 4. 2 (A) shows an increase in absorbance to approximately 1400 mAU. Once the culture broth fully passed through the HIC column, it was washed using 4M NaCl to remove the cellular debris and any other contaminants and this can be seen on the absorbance level which was dropping down to nearly to 10 mAU. Bound proteins were collected through 80 ml gradient with distilled water. As long as the conductivity decreased though the addition of water gradient, the more proteins could be released from the HIC column. From each HIC phase a sample of 50 ml was collected for analysis at western blot (Figure 4. 2 C). Protein fractions collected were precipitated and run in western blot to assess the success of HIC. Following TCA, proteins were run 10% denaturing acrylamide gel and transferred into 0.2 µm nitrocellulose membrane and exposed to 6X His-tag monoclonal Antibody as a primary antibody and Goat anti-mouse IgG conjugate as a secondary antibody. Subsequent band florescence was observed with the gel doc.

Each fraction of HIC were run in western blot, the membrane had shown a detection of sample supernatant and sample waste from the HIC column (Figure 4.2 C). Two groups of protein bands (~35 kDa and ~15 kDa) are present in the original sample (SS and the sample waste (SW) respectively. The upper group indicate the intact *TaWRKY3* protein, and with glycosylation, and the lower group of bands indicate truncated *TaWRKY3* proteins. At washing phase of the column, no proteins were identified. Elution fractions (E1, E2, and E3) also showed no detection of *TaWRKY3*. From figure 4.2 C, it appears that *TaWRKY3* expressed protein passes through HIC column and not being hydrophobically being bound to the Phenyl Sepharose resin column.

For IMAC column protein purification, eluted products collected from each HIC were combined and mixed with mono-phosphate sample buffer (10 mM imidazole). The sample was run in an equilibrated IMAC SepFast resin. As it clearly shown in Figure 4. 2 B that the absorbance increases as the sample infuses onto the column. Once the load volume sample passes through the column, then washing of the column was conducted to remove non-bound proteins from the column using sample buffer. The second washing was done using sample buffer with 20 mM imidazole. Bound proteins were eluted using 250 mM imidazole which was collected in to two fractions. Each purification phase was collected and typically 50 ml of eluted protein solution was collected to be precipitated and evaluated via western blot (Figure 4. 2 D). As above, protein were precipitated and loaded into 10% acrylamide gel and transferred into nitrocellulose membrane which then treated with primary and secondary antibodies. Following that, band florescence was observed with the gel doc. Unfortunately the western blot analysis for *TaWRKY3* did not show any bands corresponding to the recombinant protein, blots only showed the presence of the 6xHis tagged control protein. This indicated that eluted parts from HIC purification step did not contain 6x His tagged *TaWRKY3* protein species for being captured using IMAC column.

To overcome this issue of no binding to the HIC column, we proposed extracting the *TaWRKY3* protein directly from clarified BSM culture using IMAC. This required adjusting the pH of the clarified supernatant to pH8.0 in mono-phosphate buffer with 10 mM imidazole. However, once mixed and the pH increased from 4.6 to 8.0 a heavy

precipitate was formed. This precipitate blocked the FPLC disabled us from running the sample through the Ni-column.

4.3.2 M3 media protein expression:

Due to the difficulties identified within the *TaWRKY3* purification through HIC using BSM media. M3 medium adopted from Zhang *et al.* (2006) was used to look at its ability for constitutively express target proteins (*TaWRKY3* and *TaWRKY3m*), with the view of utilising a media that was compatible with direct purification with IMAC. Fermentation conditions and control loop set points were replicated for previous experiments with BSM.

The fermented culture was collected and centrifuged to collect only the supernatant which would be used for salts stability under high pH conditions. Fermentation based on BSM was also included side by side with M3 medium to run both at the same conditions.

Phosphate stability in BSM and M3 media was assessed for both in triplicates (50 ml each). Initial pH for BSM based supernatant was 5.0. 10 N NaOH was used to increase the pH to 10.0 under magnetic stirring. It had been observed that precipitate started to form above pH 5.8 (figure 4. 3 A). At pH 8.0, there was a clear appearance of a heavy precipitate. Reducing the pH to 5.0 lead to re-solubilize the precipitate in the media.

M3 based supernatant initial pH was 4.6 No precipitation was observed at pH 10.0. Precipitate were observed at pH above 11.0 (Figure 4. 3 A). Reducing the pH back to 5.0 had shown a re-solubilisation of precipitates back again.

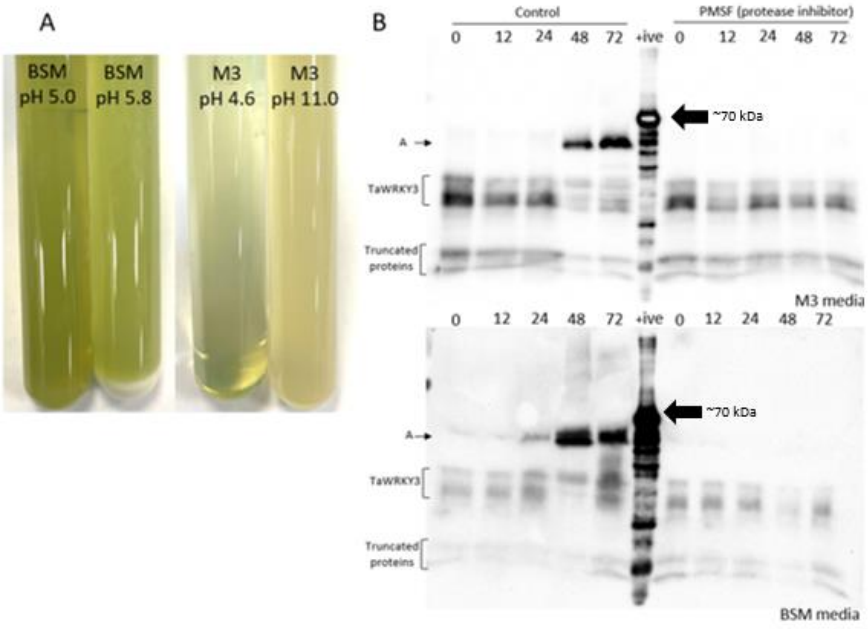
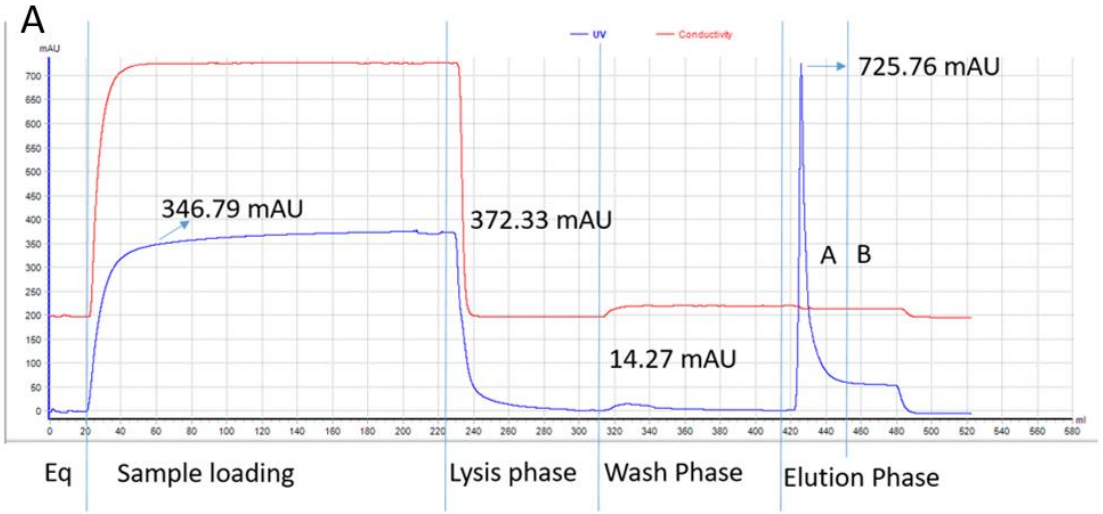


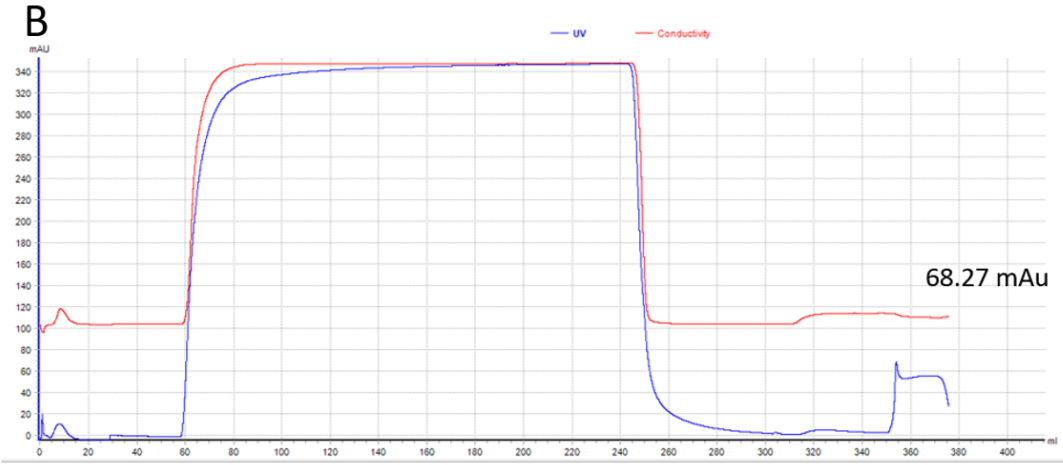
Figure 4-3 His-tagged *TaWRKY3* protein expressed by *P.pastoris* using M3 media and Basal media (BSM). A; precipitation test of BSM and M3 media. B; stability test against time with 1 mM PMSF and without. Samples were collected at different time points from 0 – 72 h and precipitated using TCA precipitation. SF16:BSA protein was used as a positive control (~70 kDa).

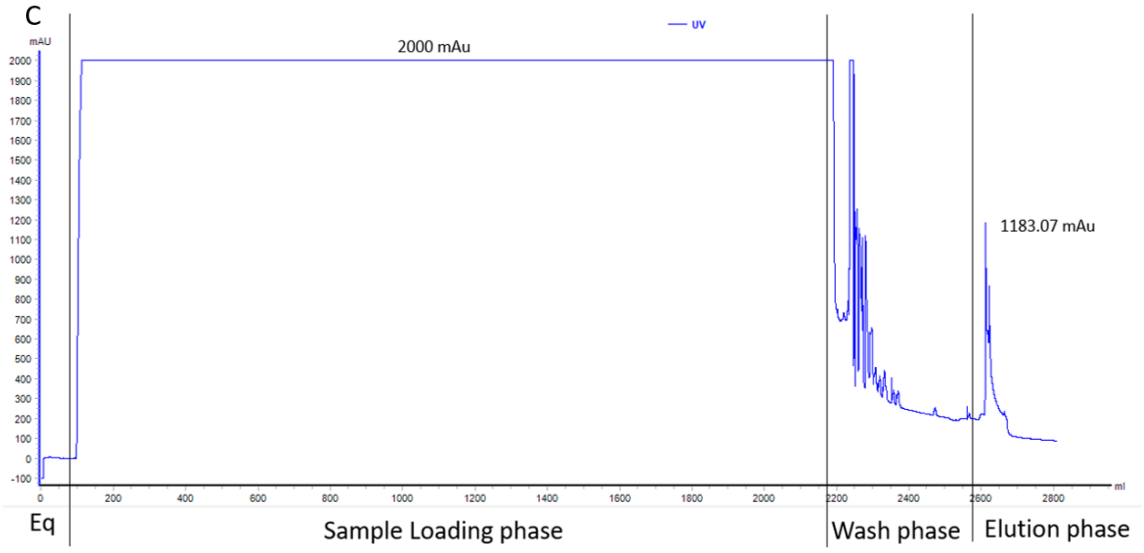
For the protein stability, PMSF, as irreversible inhibitor of trypsin, chymotrypsin, thrombin and papain enzymes was used to assess protein degradation after fermentation. BSM and M3 medium supernatants were used at different time points. PMSF was added after the 12, 24, 48, and 72 hours time points. It was observed that both mediums with no PMSF shad shown a development of cloudiness though all time points. All samples were precipitated and run on western blot.

As it is shown in Figure 4. 3 B, proteins with PMSF were found to be stable and not being exposed to degradation through tested time frame. In comparison, control test was found to be degraded from time point 0 h to 24 h. A high molecular weight protein band appeared from time point 48 h to 72 h. These higher molecular weight bands are too large to be the recombinant proteins. At these time points, no appearance of this higher molecular weight band shift when treated with 1 mM PMSF. The intensity of secreted proteins in M3 was found to be higher than BSM. There was an appearance of truncated proteins right below each band when using M3 and BSM.

4.3.3 His-tag protein purification







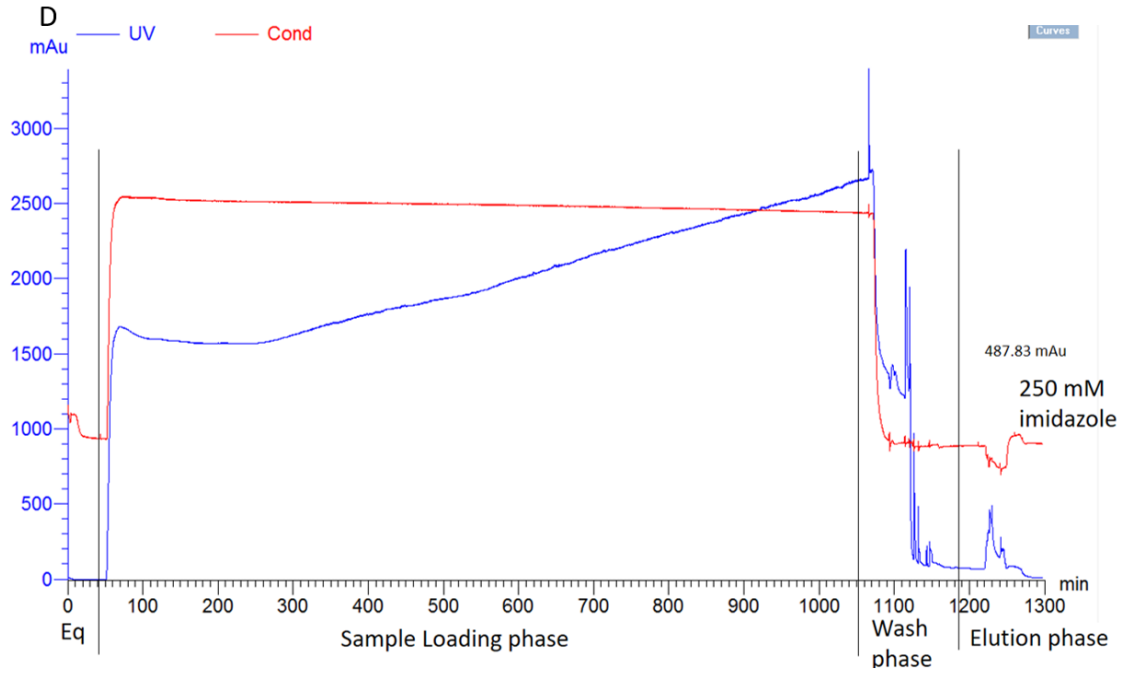


Figure 4-4 Ni-column protein purification of TaWRKY proteins. A; *TaWRKY53b*, B; *TaWRKY19*, C; *TaWRKY3*, and D; *TaWRKY3m*. Each set of protein was eluted with 250 mM imidazole. *TaWRKY53b* and TaWRKY19 were expressed using BSM medium and *TaWRKY3* and *TaWRKY3m* were expressed using M3 medium.

TaWRKY53b and *TaWRKY19* 6x His-tagged proteins were run in a charged IMAC SepFast resin with 0.1 mM Nickel sulfate after HIC. The IMAC column was, firstly, equilibrated with sample buffer. For *TaWRKY53b* and *TaWRKY19*, monosodium phosphate with 10 mM imidazole-based buffer was used as a sample buffer (Figure 4. 4 A and B). Each was run in the column and washed using 10 mM imidazole buffer and a second was also added using 20 mM imidazole buffer. His-tagged *TaWRKY53b* and *TaWRKY19* were eluted using 250 mM imidazole. The absorbance for *TaWRKY53b* was 725.76 mAu whereas *TaWRKY19* was 68.72 mAu (Figure 4.4 A and B). With such lower absorbance, this might indicate lower concentration of *TaWRKY19* present in figure 4.4 B. thus, multiple fermentation runs were conducted and purified using HIC and Ni-column. All eluted products post-Ni-column were dialysed and ultrafiltrated which was then combined together.

In terms of *Ta*WRKY3 and its mutant form, sodium glycerophosphate based buffer was used a sample buffer (Figure 4. 4 C and D). Generally, 1x sample buffer was run in the column to equilibrate it. Sample containing sample buffer was run through the column and this called sample loading phase. Sample solution was allowed to circulate though the column multiple times at low flow rate (~ 0.5 ml/mins). For large volumes such as

TaWRKY3 and *3m*, the sample were left circulating overnight. The loading phase was ended by the initiation of wash phase. The wash phase was washing the column with 10 mM imidazole sample buffer and a second was also added using 20 mM imidazole to remove loosely bound proteins, the Elution phase was initiated with 250 mM imidazole. As clearly seen among all eluted proteins. A high peak present in *TaWRKY3* (1183.07 mAu), and *TaWRKY3m* (487.83 mAu). Each of these proteins had presented a very high peak, and this was an indication of high protein content was eluted off the column. Each eluted protein was collected and stored in -20 °C until use. The volume of collected eluted products was roughly between 10 – 15 ml each.

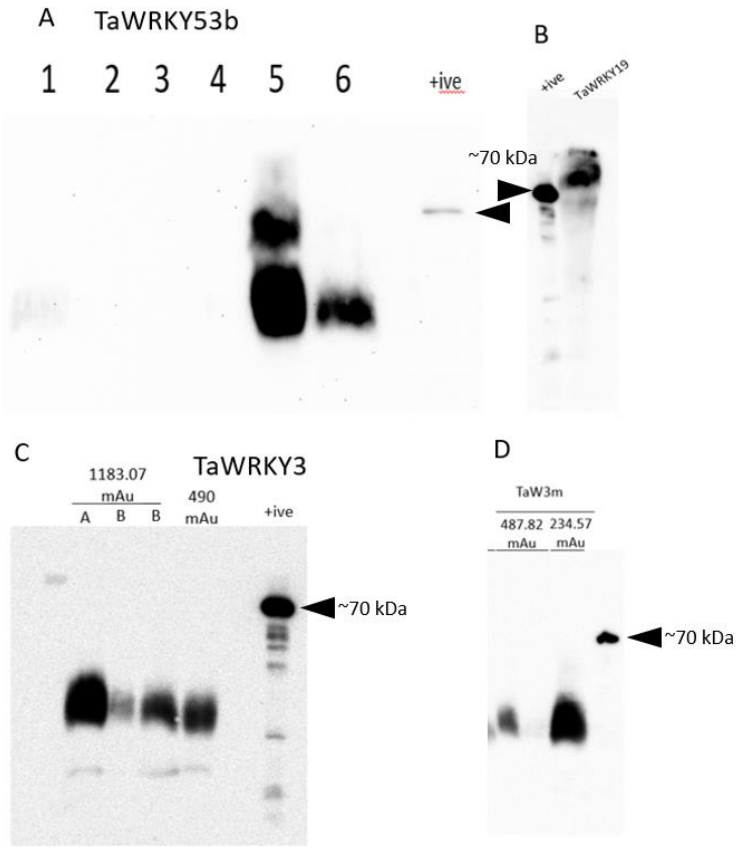


Figure 4-5 Ni-column purified, and column concentrated (10 kDa MWCO) TaWRKY protein run using western blot. A; illustrates Ni-column fractions. 1; sample loading, 2; sample loading waste, 3 sample wash 1, 4; sample wash 2, 5; Elution A fraction, and 6; Elution B fraction. B is *TaWRKY19* purified protein. C and D all purified proteins of *TaWRKY3* and *TaWRKY3m*, respectively. +ive indicate positive control. All purified proteins were eluted with 250 mM imidazole. SF16: BSA protein was used as a positive control (70 kDa).

At this far of protein purification, eluted products were dialysed using 8 kDa molecular weight cut off (MWCO) over night at cold against ddH_2O. 10 kDa MWCO ultrafilteration columns were used to concentrate proteins to make it in a sample volume but with high concentration as well as removal of truncated proteins (Figure 4. 3 B). Ni-column purified, and column concentrated (~10 kDa MWCO) TaWRKY protein run using western blot (figure 4.5). A; illustrates Ni-column fractions 1; sample loading, 2; sample loading waste, 3; sample wash 1, 4: sample wash 2, 5; Elution A fraction, and 6; Elution B fraction. B is *Ta WRKY19* purified protein. C and D all purified proteins of *Ta WRKY3* and *Ta WRKY3m*, respectively. +ive indicate positive control (SF16: BSA, ~70 kDa) which was previously cloned in pGAPZaA and expressed by Pichia in YPG media. SF16:BSA was precipitated using TCA method and used as a positive control. All purified proteins were eluted with 250 mM imidazole. Figure 4.5 shows a western blot of each protein post column concentration.

Each TaWRKY protein were run in western blot. The calculated molecular weight for *Ta WRKY53b* was 47.4 kDa. However, western blot revealed a double band in the membrane at line 5 (figure 4.5 A). The upper band approximately 55 kDa. The lower band was between 35 to 50 kDa which appeared as a very large cluster of protein species. In comparison, Line 6 in Figure 4.5 A shows only a single band which was clearly ~ 40 kDa. In terms of *Ta WRKY19*, the calculated molecular weight was 50 kDa. In figure 4.5 B, there was only a presence of only a single band. The size of *Ta WRKY19* appeared was ~ 80 kDa when compared to SF16: BSA control (70 kDa). This was an indication of hyperglycosylation of *Ta WRKY19*. This phenomena was found with GS11mIL10 produced protein which showed a hyperglycosylation and resulted to an increased of molecular weight (Jacobs *et al.*, 2009). With regards to *Ta WRKY3*, its calculated molecular weight was 24.9 kDa. However, there was a presence of two band species in western blot membrane. The upper band was ~ 35 kDa and the lower band was ~ 15 kDa (Figure 4.5 C). The presence of double bands in *Ta WRKY3* occurred in two separate fermentation and purifications (1183.7 mAu and 490 mAu). The shift of the molecular weight to the calculated molecular weight could be caused by glycosylation of the protein. The lower band can be indicated as truncated *Ta WRKY3* proteins carrying 6x His-tag. *Ta WRKY3m* had shown only a single band at the size of ~ 35 kDa (Figure 4.5 C and D).

The protein concentration of each TaWRKY protein was determined using Bradford Assay. The concentration of *TaWRKY53b* was found 1.45 µg/µL, *TaWRKY19* was 3.2 µg/µL, *TaWRKY3* was 0.938 µg/µL and *TaWRKY3m* was 0.963 µg/µL.

4.4 Discussion:

As part of protein-DNA interactions studies (chapter 5), it is essential to express WRKY transcription factors. For that, bacterial systems had been very commonly used to express such proteins. Many had encountered several problems when using a full length of cDNA to express WRKY transcription factors in *E. coli*. Ciolkowski *et al.* (2008b) attempted to express *AtWRKY26*, *AtWRKY6*, *AtWRKY43*, *AtWRKY11*, and *AtWRKY38* in *E. coli*. Expression was found to be detrimental to bacterial growth. WRKY transcription factors contain zinc finger motif. Thus, expression was found to impact negatively bacterial zinc homeostasis. Exogenous addition of zinc was not helpful. They also reported that even if the bacterial growth was not affected by the expression of WRKY proteins, some were found to be exclusively in inclusion bodies. From which, purification of protein from this source would involve denaturation and renaturation (refolding) steps, leading to misfolding of protein and resulting in loss of activity from binding to W-box element. Under mild conditions, affinity purification of WRKY proteins was also problematic. This occurs with *NtWRKY12* (van Verk *et al.*, 2008) and *AtWRKY50* (Hussain *et al.*, 2018). It could also be related to the fact that bacterial expression mechanisms lack correct structural folding of proteins leading to WRKY proteins' inability to be functional.

Protein fold to a three-dimensional structure, which becomes biologically functional. In eukaryotic cells, proteins undergo posttranslational modifications that are critical to the functionality of proteins. This includes N-terminal formyl methionine residue elimination, disulphide bond formation between cysteine, hydroxylation, covalent modification, acetylation, carboxylation, methylation, deamination, amidation, and phosphorylation (Ramazi and Zahiri, 2021). Mis-folding of proteins leads to affect their activity in the cell, such as inhibiting proteins from binding or interacting with their target cellular components such as DNA or proteins (Blanco and Blanco, 2017)

To overcome WRKY protein expression and purification issues and acquire properly folded WRKY proteins for their functionality. *Pichia pastoris* expression system was proposed in this project to express WRKY proteins. It is a eukaryotic expression system and offers the quality of *E. coli* in terms of simplicity of genetic modification, high cell density in costly effective media, and capability to scale up the protein expression production. Additionally, it has advantages of a eukaryotic organism, such as proper protein folding, post-translational modification, and the ability to secret protein to the medium. *P.pastoris*

protein expression confers not much concern with regards to protein folding, formation of disulphide bridge, post-translational modification, and secretory cleavage that is well supported in yeast (Lee *et al.*, 2021).

Following a previous chapter 3, small scale protein expression was brought up to a large-scale production (bench-top) for wheat WRKY protein expression. Recovery of proteins was dictated by their performance at the purification stages. Different proteins had different behaviour toward purification.

As described by '*Pichia* Fermentation Process Guidelines' of Invitrogen™, BSM was recommended to be used as a sole medium for pGAP constitutive expression promoter for expressing proteins using *P.pastoris*. It uses phosphoric acid as a phosphorus source. During fermentation pH and temperature were kept constant at 27 °C and 4.6, respectively. Growth at temperature above 32 °C can be detrimental to the expression of proteins. The typical bench-top fermentation was divided into two phases: the glycerol batch phase and the glycerol fed-batch phase. Both phases result in rapid biomass accumulation and constitutively expression of recombinant proteins. The main usage of glycerol was to be consumed by cells as a carbon source during fermentation. According to dissolved oxygen concentration, fermentation can be switched from the glycerol batch phase to the glycerol fed batch phase. Once dissolved oxygen rapidly increases, this indicates that medium/ batch glycerol had been completely consumed. Hence, glycerol fed-batch would be initiated to increase the cell biomass and more production of proteins though out 72 hours run. The level of dissolved oxygen was generally kept at 30% at all times which was regulated by agitation provided by impellors speed. To terminate the reaction run, the glycerol fed-batch was stopped, and once glycerol was completely consumed and dissolved, oxygen was rapidly increased to 100%. At this stage, the culture was then collected and centrifuged to collect culture broth for protein purification.

A pure protein purification must be obtained for analytical experiments to have a high signal to noise ratio. However, there is no generalized purification method for proteins. Selecting the type of chromatography is entirely dependent on the target protein's physical and chemical properties (Lee, 2017). Hydrophobic interaction column chromatography, size exclusion column chromatography, ion-exchange column

chromatography, and affinity chromatography. In terms of TaWRKY recombinant expressed proteins, our purification strategy lays on two-step purification chromatography: hydrophobic interaction column chromatography and affinity column chromatography. The former was mainly used to hydrophobically trap expressed proteins from broth medium and salting out. This method was affinity column chromatography using Immobilized metal-ion affinity chromatography which was mainly through the Nickel chelating resin. This was only to purify TaWRKY recombinant expressed protein containing polyHistidine (6x His) tag sequence within the protein using Nickel-charged affinity.

TaWRKY53b and *TaWRKY19* transformed clones were run using BSM medium and the culture broth was collected for purification using HIC and Ni-column. Initially, the aim for performing HIC was to remove host cell impurities high molecular weight aggregates and concentrate the sample. With that, *TaWRKY53b* and *TaWRKY19* proteins were firstly purified by HIC (fig. 4.1). *TaWRKY3* and its mutant form, on the other hand, were not enabled to be purified through HIC (fig. 4.4). Both proteins HIC purification were performed the same as *TaWRKY53* and *TaWRKY19* with 4M NaCl as salting out buffer.

With regards to *TaWRKY3* and *TaWRKY3m*, they were intended to be purified, firstly, by HIC. As the sample was loaded to an equilibrated resin matrix, *TaWRKY3* had shown a tendency to pass through the column toward sample waste, which is clearly shown in Figure 4.4. Purification fractions were analysed to investigate whether excess proteins get eluted or if all *TaWRKY3* protein does not bind to the resin matrix. Western blot analysis for all purification fractions had shown that *TaWRKY3* passes through the column matrix toward sample waste as it clearly appears on western blot membrane in both HIC and N-column. This indicates that protein binding to resin matrix was not facilitated, even with the existence of high salt content to promote high ionic strength for hydrophobic interaction between protein and Sepharose ligands.

As there was high salt concentration added to the protein sample for further understanding of HIC setting. Proteins generally are considered to be nonpolar molecules. As a polar solvent, water is considered to be a poor solvent for proteins macromolecules (non-polar). Proteins molecules favour to self-associate or aggregate under polar

conditions and this to achieve the lowest thermodynamic energy state. Prior to that, around individual protein macromolecules, a hydration shell is formed by highly ordered structures of water molecules that forms a strong shell. High molarity salts can be added to promote hydrophobic interactions between protein-ligand and precipitation, leading to salting out effect. At this effect, the introduction of high molarity of salts causes an interruption of the water hydration shell scrounging individual protein macromolecule. This leads water molecules to favour the interaction with salt ions than protein amino acid side chains (Fleming, 2020). Consequently, hydrophobic amino acids side chains to be exposed, allowing protein-ligand interactions to occur, facilitated by thermodynamic forces. The interaction is considered to be reversible as well which can be reversed by the reduction of the column's salt ionic strength. The hydration shell can be gradually restored and allow proteins to be released from resin ligand interaction because this interaction is considered weak and can be easily broken for elution (Queiroz;Tomaz and Cabral, 2001). In case of *Ta*WRKY3, protein-ligand interaction seemed as never been facilitated and many parameters could impact the hydrophobic interaction between protein- ligand interactions which had to be taken in consideration. Protein properties, ligand type, ligand saturation bead, salt type and concentration, pH, temperature and matrix composition are all parameters influencing HIC. Salt type and concentration could be said the most important parameters among all, since they facilitate the hydrophobicity of proteins. Generally, ammonium sulfate a magnesium sulfate are the most commonly used types of lypotopic salts due to their wider range of proteins to facilitate binding to HIC.

NaCl is also commonly used as lyotropic salts for facilitating hydrophobic interaction (McCue, 2009). However, it is considered to be not very effective type of salt to be used in HIC because not all proteins bind to HIC column even at high concentration (3.5M) (Yang;Koza and Chambers, 2015), which was also considered to be a weak salting-out type of salt (Tsumoto *et al.*, 2007). A change of salt type could be used to facilitate the binding protein-ligand binding for purification such as ammonium sulfate.

At Large scale of proteins expression of *TaWRKY3* using Basal Salts Medium (BSM), a sample of each fermentation run used to be tested at western blot to check for the protein expression using His-tag. Once it had been confirmed, the purification process starts with

Hydrophobic Interaction Chromatography (HIC) as it was performed with *Ta WRKY53b* and *Ta WRKY19* proteins. Then the process followed with Nickel column (Ni-column) to specifically purify His-tagged proteins. However, with *Ta WRKY3* protein, it appeared no presence of His-tagged proteins bands on western Blot membrane when taking purification stages fraction of HIC and Ni-column apart of sample supernatant and sample waste of HIC. Thus, it appeared that *Ta WRKY3* proteins tends to pass though the column resin and does not hydrophobically bind to the Phenyl Sepharose beads column and gets eluted before binding to the column beads for purification. Hence, no His-tagged proteins were available for binding for Ni-column to be eluted and tested. As part of troubleshooting, a fresh Phenyl Sepharose breads was packed to the column and a new fermentation supernatant was run for *Ta*WRKY3 and the same issue appeared. A changing in the purification strategy design was mandatory by passing the whole sample through the Ni-column. However, when mixing the sample buffer (monosodium phosphate-based buffer, pH 8.0) with the sample itself, immediate precipitation of salts appeared, which had a massive impact on the purification process using Ni-column by clogging the column itself and disabling the sample to run through. This precipitation was thought by phosphates (dication or trication phosphate) in BSM, causing phosphate precipitates out of the medium at pH above more than 5.5 (Huang *et al.*, 2013). Thus, an alternative fermentation medium was used to overcome these issues.

M3 medium was selected as a sole recipe for protein expression using *P.pastoris* through fermentation. It was firstly proposed by Zhang;Sinha and Meagher (2006) along with other modified media for protein expression. Generally, BSM or FM22 recipe was modified by replacing the phosphate salts with glycerolphosphate. They aimed to solve phosphate precipitation issue in *P.pastoris* fermentation through the use of glycerolphosphate as a phosphorus source. Sodium (Na_2GP), potassium (K_2GP), calcium (CaGP), and magnesium (MgGP) glycerolphosphates were used in their investigation. Throughout their findings, they had excluded CaGP and MgGP due to their solubility (only around 2% w/v) at room temperature. On the other hand, Na_2GP and K_2GP were soluble above 65% w/v at room temperature. Na_2GP (M3) medium was selected because its ability to withstand pH up to 10.5 compared to K_2GP, which showed precipitation at pH above 7.5.

M3 media was run using *P.pastoris* expressing *Ta*WRKY3 along with BSM media expressing the same construct to test their salt stability side by side. Our precipitation tests had shown that the M3 medium could be intact up to pH 9.8. Precipitation was found to start at a pH above 10.0.

As part of troubleshooting and experimentation for both M3 and BSM, before downstream processing, it was crucial to study the stability of expressed proteins against time and with and without the addition of PMSF as a protease inhibiter. *Ta WRKY3* protein was selected for this study. Thus, *Pichia* expressing *Ta WRKY3* was inoculated in both types of medium. Harvested supernatant was incubated at room temperature and compared between both media and with the addition of PMSF and without. *Ta WRKY3* was found to be stable and showed no degradation throughout the 72h time course with the addition of (1mM) PMSF (fig. 4.5.). However, both media without the addition of PMSF was degrading throughout time. This might be due to the presence of Serine proteases expressed by *Pichia* and released to the medium as the used *Pichia* strain in this study was a wild-type (X-33 strain) which is not protease deficient as SMD1168H strain. On the other hand, the addition of PMSF to collected media prevents proteins' degradation, which is clearly presented in figure (fig.4.5). For instance, it had been suggested by many studies in the literature that post-secretory proteolytic degradation provoked by serine and aspartic proteases was found to be one of the major drawbacks of the *P.pastoris* expression system. This specifically could be said to be a wild type of *Pichia.* However, protease-deficient host strains, such as SMD1163, SMD1165, and SMD1168, had shown an effective reduction in some foreign protein's degradation. The Drawback of protease-deficient host strains was low protein yield compared to wild-type strains (Higgins and Cregg, 1998).

For protein expression, many strategies were identified to control protease activity during bench-top fermentation and maintain protein stability, such as increase operating pH between 5.5 to 8, the addition of protease inhibitors in the yeast culture supernatant (Potvin;Ahmad and Zhang, 2012), as well as reducing temperature to 23 °C (Li *et al.*, 2001). With such methods, the degradation of proteins by extracellular proteases could be decreased. Casein conjugated to a florescent marker as substrate could be used as an activity assay to measure supernatant protease activity. This could give an indication of

the severity of proteases being expressed in the medium. For that, method development is thought to be required to decrease expressed of proteases (Potvin;Ahmad and Zhang, 2012).

There was a clear appearance of higher intensity of proteins expressed in M3 medium compared to Basal medium in western blot membrane image. Yet, this higher intensity was not measured in this experiment. However, it could be measured with either through analysing western blot images with ImageJ software or total protein purification and protein concentration using Bradford assay. Proteins expressed using M3 medium and Basal media medium could be loaded in SDS-PAGE and a standard control protein containing the same tag as *Ta*WRKY3, which is 6xHis-tag. Through that, the standard control protein concentration and band intensity at the membrane image, the protein concentration of *Ta*WRKY3 expressed with either M3 medium and/or basal medium could be measured. Zhang Sinha et al (2006) of α-galactosidase recombinant protein expression comparison between BSM and M3 found M3 medium expressed α-galactosidase recombinant protein 215% higher than BSM. This could confirm our comparison between M3 and BSM using *Ta*WRKY3 TF protein.

A higher molecular weight band was present from time point 48 h to 72 h at both M3 medium and BSM medium (Fig4.5 A). This could be thought of as a dimer aggregation of proteins containing 6xHis-tag. This prospect could be easily rejected by the nature of TCA protein precipitation prior to loading in SDS-PAGE for western blot. Precipitated proteins were subjected to heat temperature at 100 °C for 10 minutes and loaded in a denaturing SDS-PAGE gel for separation. This would eventually lead to proteins, when exposed to extreme conditions, to denature and dissociate from aggregated proteins, if it is that the case.

However, samples containing PMSF had shown no such development of higher molecular weight proteins. A similar phenomenon was also present at *TaWRKY53b* post-Ni-column purification. This could be said to be dimerization, which bonds between two WRKY transcription factors to form one bonded molecule. Cheng *et al.* (2019) identified the structural bases of *OsWRKY42*-DBD dimerization. They found that WRKY transcription factors have a capability to form homo-dimerization through hydrogen bonds between

antiparallel ß-strands of proteins. They had indicated that this anti-parallel binding between two *OsWRKY42*-DBD monomers was hard to separate. Even with the denaturation of proteins and exposure to a denaturing acrylamide gel, proteins dimerization was conserved. We hypothesised that *TaWRKY3* formation homo-dimerization throughout time enabled it to withstand denaturation treatment. This protein stability could be attributed to the higher glycosylation and disulphide bond formation within each monomer, facilitated by P.pastoris expression and the hydrogen bond between antiparallel proteins. However, this could not be a major concern for protein- DNA interaction experimentation. Cheng *et al.* (2019) findings on rice WRKY protein had shown that each protein monomer can interact with DNA at homo-dimerization state. Meaning that two proteins with homo-dimerization would interact with two W-box DNA (Marianayagam;Sunde and Matthews, 2004).

Downstream processing of secreted protein was a major concern prior to using the M3 medium. Unlike BSM, there is only a minimal amount of literature reporting the benefits of Pichia fermentation with M3 media Due to difficulties encountered during *Ta*WRKY3 protein HIC purification as well as broth salts precipitation at alkaline pH using BSM. It was thought best to eliminate the HIC purification step and substitute BSM with M3 medium for using broth media in Ni-column at alkaline conditions to prevent the occurrence of salts precipitation. For the purpose of optimisation at Ni-column, monosodium phosphate-based buffer (pH8.0) was found not to enable to purify of His-tagged proteins. Thus, it was substituted by glycerolphosphate (glycerolphosphate diasodium salt hydrate salt) at pH 8.0. This allowed the purification of *Ta*WRKY3 proteins to be much more feasible.

One of the main purposes of using HIC is to remove cellular impurities as well as concentrate expressed proteins from large bulk volumes into small volumes. It is not selective and does not purify target proteins, but it purifies the total protein content present in the sample supernatant. It is later on eluted with water from the column by an increasing gradient from 0% to 100% for full elution. Eluted products could then be mixed with sample buffer (glycrophosphate based buffer) which can then passed to be purified through Ni-column. The Drawback of eliminating HIC in this study was accumulation of precipitates in Ni-column resin when passing a circulating bulk volume overnight at room

temperature. Firstly, this could pose a higher risk of clogging the purification system. Secondly could have extensive column washing using washing buffer required to be used to ensure the removal of culture growth remnants. Column clogging was mainly caused by the accumulation of precipitate of culture inside the column, and this was remarkably noticed from the circulating overnight column. Thus, it is recommended to use HIC to eliminate macromolecules contaminants. Using HIC purification of proteins using M3 medium was found to be successful in this study. So, it is crucial to use HIC to prevent system clogging by contaminants and to remove undesired culture macromoleculesof culture inside the column, and this was remarkably noticed from the circulating overnight column. Thus, it is recommended to use HIC to eliminate macromolecules contaminants. Using HIC purification of proteins using M3 medium was found to be successful in this study. So it is crucial to use HIC to prevent system clogging by contaminants and to remove undesired culture macromolecules

Post protein collection, it was mandatory to dialyse eluted proteins, remove excessive salts and imidazole, and concentrate using protein ultrafiltration columns to remove excessive water content within the protein sample. Despite difficulties encountered previously, *TaWRKY3* and its mutant form were successfully enabled to be specifically purified from Ni-column using 250 mM imidazole and collected as well as concentrated for being used, along with *TaWRKY53b* and *TaWRKY19* proteins, at Electrophoretic Mobility shift Assay study (EMSA) as a crucial part of protein-DNA interaction study.

In the Biotechnology industry, Tangential Flow Filteration (TFF) had been used to separate cell-protein, virus-Protein, and protein-buffer. This type of separation uses a membrane to separate molecules based on their size and molecular weight (Steen *et al.*, 2019). For industrial scales, biological samples, including protein, could be processed by utilizing the capability of TFF in terms of scalability. Time efficiency, reproducibility, and reusability are also associated with TFF (Yehl and Zydney, 2020). Compared to chromatography, its operation is considered to be cost-effective in large scale protein production. For large scale proteins production, M3 media could be used for recombinant protein expression fermentation as a sole media. Its ability to substantially produce proteins using *P. pastoris* could be coupled with TFF for downstream processing in a large-scale setting.

To surmise our findings, *Ta*WRKY proteins expressed by *P.pastoris* through pGAPZαA were successfully expressed and purified in large scale fermentation (bench-top). M3 medium was found to express desired proteins at comparable levels and/or higher than BSM. Purification methods had shown to be conserved to protein characteristics. *TaWRKY53b* and *TaWRKY19* were found to be simply expressed using BSM and purified through HIC, for cellular impurities removal and protein concentration, and then passed through Ni-column purification for trapping His-tagged proteins. On the other hand, *TaWRKY3* and *TaWRKY3m* protein purification were difficult using HIC due to protein degradation and medium salt suitability at alkaline pH. Thus, M3 medium was used to express *TaWRKY3* and *TaWRKY3m* proteins and be purified directly from Ni-column. It was found that the M3 medium could hold its salts contents at high alkaline conditions (~ pH 10.0) for protein purification using Ni-column.

For such His-tagged native proteins to be expressed and purified as a future aspect, it might be excellent to use the M3 medium as the main source of phosphorous for *P.pastoris* cellular growth and recombinant protein expression. Regarding purification, HIC with different salt types such as ammonium sulphate and/ or magnesium sulphate could be used for the sake of cellular impurities removal and bulk protein concentration to smaller volumes. Glycerophosphate-based buffer for Ni-column for His-tagged proteins purification could also be employed due to its compatibility with the column and better binding of His-tagged proteins compared with mono-sodium phosphate or mono-potassium phosphate-based binding buffer. Protein ultrafiltration columns could be much recommended to be used as they tend to reduce the Ni-column elute to in volume and increase the concentration of protein post-dialysis.

5. Chapter 5: Binding of Wheat WRKY Transcription Factors Proteins on Wheat DNA Defence Promoters

5.1 Introduction:

Typically, cellular responses are mediated by transcription factors that can be triggered by plant exogenous stressors. This mediation occurs through specific recognition of *Cis*-regulator DNA sequences found in the promoter regions of target genes. This work focuses on Zinc-finger motif transcription factors known as WRKY. The family of transcription factors comprise of a large family which they actively on DNA-binding proteins in plants. It is not only restricted in plants but also appear in yeast and animals (Ülker and Somssich, 2004). WRKY proteins are one of many transcription factors that were found to play role in biotic and abiotic stress responses (Franco-Zorrilla *et al.*, 2014). To regulate gene expression most identified WRKY transcription factors proteins with the characteristic WRKYGQK amino acid sequence can bind to W-box (TTGAC[C/T]) elements of located within promoters (Eulgem *et al.*, 2000; Ciolkowski *et al.*, 2008a). As it was widely known that most WRKY protein share approximately 60 amino acid DNA binding domain (DBD) and conserved Cys and His residues composing zinc finger motif. Through their DBD, their classification grouping were found as well. Their specificity of binding to the Cis-regulatory W-box element is facilitated (Yamasaki *et al.*, 2013). The C-terminal WRKY domain of WRKY transcription factors was firstly reported from Arabidopsis AtWRKY4 in complex with DNA using NMR structure (Yamasaki *et al.*, 2005; Duan *et al.*, 2007; Yamasaki *et al.*, 2012). The DBD structure of WRKY protein consists of four to five antiparallel strands of ß-sheet structure. Two Cys and His residues were also found to form a pocket for the Zinc binding. The second ß-sheet consists of the WRKYGQK sequence which can penetrate the major groove of the DNA strand. The specificity recognition as well as binding of target sequences is therefore facilitated by this amino acid sequence located in the second ß-sheet (Yamasaki *et al.*, 2012). The interaction occurs between WRKYGQK contained within the ß-sheet as a positively charged and nucleobases and the negatively charged phosphate backbone within the DNA nucleotides (Duan *et al.*, 2007).

Different WRKY proteins had been shown by gel shift experiments, DNA-ligand binding screens, random binding site selection, co-transfection and yeast one-hybrid studies that there is a stereotypic preference of binding to 5'-TTGAC-C/T-3' known as W-box. There is

a stereotypic preference of binding to 5'-TTGAC-C/T-3' known as W-box. This feature represents the minimal consensus requirement for DNA binding specificity (Rushton *et al.*, 1996). The specificity of WRKY binding toward individual promoters is not yet fully understood. However, it had been suggested that the specificity of such proteins might be conferred by flanking regions surrounding W-box elements. It is assumed that there is a possibility of WRKY transcription factors being involved with a higher order nucleoprotein which might be for promoter selectivity determination as well as transcriptional output (Ciolkowski *et al.*, 2008a).

Ta WRKY3 (group II) with one WRKY domain. This domain is located at its Cis-terminus domain of the protein. *Ta WRKY19* and *Ta WRKY53b* are group I with 2 domains. WRKY proteins has Trans- and Cis- terminals domains and their zinc motif. All of which have WRKYGQK and zinc finger motif (Cx4C23HxH) at their *Cis*-terminus domain (Chpater2).

From our previous chapter, wheat WRKY transcription factors (*Ta WRKY53b*, *Ta WRKY19*, *Ta WRKY3* and *Ta WRKY3m*) were cloned into pGAPZαA and transformed into *Pichia pastoris*. The aim was to express proteins through bench-top fermentation unit and later been purified using IMAC Nickel column to purify 6xHis tagged WRKY proteins. Western blot results had shown that wheat WRKY proteins were expressed and also purified. With that WRKY proteins could now be used for protein-DNA interactions using electrophoretic mobility shift assay (EMSA).

To detect protein-nucleic interactions, electrophoretic mobility shift assay (EMSA) is used as a rapid and sensitive. It is based on the transcription factors ability to bind to specific radiolabelled or non-radiolabelled sequence of DNA. As a result, this interaction retard their migration through native polyacrylamide gel. Crude nuclear proteins or purified proteins could be used a source of binding to DNA. Generally, EMSA is considered to be a qualitative method which can be easily determine of the binding of known transcription factor protein to a known fragment of DNA. Generally in this method, fragments of DNA used tend to be short (20-25 bp) to prevent multiple protein-DNA interactions that might occur in long oligonucleotides. Labelling of probes is essential for detecting the shifted protein DNA interactions from free DNA. Biotin is a non-radioactive substance that

commonly used to end labelling of oligonucleotides to be run in non-denaturing polyacrylamide gel (Smith and Delbary-Gossart, 2001; Holden and Tacon, 2011).

For protein-DNA binding assays, the majority of WRKY proteins in literature were found to be expressed using a prokaryotic expression system. Most of which encountered difficulties expressing the complete protein sequence in *E. coli* as occurred with Ciolkowski *et al.* (2008a). With that, they had to use truncated protein containing only the Cis-terminus of the protein containing only the DNA-binding domain of the protein. The second issue was encountered at the purification and folding of the protein though the impact of denaturing and renaturing of protein. Both issues can impact the protein folding itself which might lead to skew the binding assay as a result of over exposure of the DBD to the w-box element DNA. Thus, as we had successfully expressed wheat WRKY proteins in their native structure using a eukaryotic expression system. This might introduce an alternative expression system for assays such as EMSA as well as chromatin immunoprecipitation assay.

Our aim was to determine the capability of natively expressed wheat WRKY proteins to target promoter genes:

- Binding of wheat WRKY proteins to target W-box element (parsley *PcPR1-1*) obtained from Rushton *et al.* (1996).
- Assess their binding to wheat pathogenesis related protein gene (*TaPR1-23*) and a mutant form of TaPR1-23.
- Determine their binding capability to bind to tandem repeats of W-box elements and a mutant form as well.

5.2 Materials and methods:

The aim of this experiment is to demonstrate that purified TaWRKY proteins binds to w-box elements. This is demonstrate its binding to W-box elements and would allow us to move forward to Chip DNA-protein binding which is fishing through wheat genome to see its binding throughout the wheat genome.

5.2.1 Target genes selection:

The method of selecting target genes were performed using two strategies. Firstly, identifying W-box elements from plant genes through looking at their promoter region. Parsley 2xW-box element was selected which is an identical sequence to parsley *PR1-1* promoter region containing only 1 W-box element (Rushton *et al.*, 1996).

Triticum aestivum Pathogen Related Protein gene was selected from NCBI database. Its nucleotides were searched by typing on the search side 'Pathogen Related protein *Triticum*' and selected only *Triticum aestivum* as an organism. Based on NCBI output results, *TaPR1-23* was selected in FASTA format for our testing. Using EnsemblePlants https://plants.ensembl.org/index.html, the full sequence of *TaPR1-23* DNA sequence was run on blast to obtain upstream sequence. Based on %ID alignment results, the gene was exported in FASTA sequence format, featured strand, and 2000 bp 5' Flanking sequence (upstream). The sequence was saved in TEXT file. For upstream promoter genes, sequence was submitted in PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Selected W-box element w was chosen as a forward sequence. Within such sequence containing W-box element, 30 bp length was selected for probe synthesis and experimentation in EMSA. The sequence was named as *TaPR1* (appendix D).

For developing mutant form of *TaPR1*, the w-box element (TTGACC) region was only selected and plotted in a readily available free random modification website (https://onlinerandomtools.com/shuffle-letters). The latter were only used as negative control. A second strategy was synthesizing only a multiple W-box elements next to each other and it was named as w-box probe. A mutation was also synthesized by replacing the fifth nucleotide (cytosine) by Adenine, and this was named as mW-box.

For all obtained probes, reverse complement counterpart sequences were also acquired. Forward probe sequences were submitted in a freely available bioinformatics website 'Reverse Complement' (https://www.bioinformatics.org/sms/rev_comp.html) in FASTA format. Output reverse complement counterparts were then saved. For which, each probe would then contain two sets of sequences forward and its reverse complement counterpart.

5.2.2 Probe synthesis:

For synthesis, all oligonucleotides were submitted to Sigma Aldrich. Each probe would be synthesized in two forms. One of which would be Biotin labelled and this is called hot probe. For the same probe, a non-labelled synthesis would be synthesized and this to act as a cold competitor in EMSA (table 5.1).

Table 5-1 dsDNA fragments were generated with and without addition of Biotin to generate the labelled probe and the unlabelled probe.

Gene source:	Forward	Reverse
PcPR1-1	TTATTCAGCCATCAAAGTTGACCAATAAT	ATTATTGGTCAACTTTGATGGCTGAATAA
TaPR1-23	TCTGTTTGGTTTGACCAACTAGATTGATTT	AAATCAATCTAGTTGGTCAAACCAAACAGA
Multiple W-box	CGTTGACCTTGACCTTGACTTT	AAAGTCAAGGTCAAGGTCAACG
mWb	CGTTGAACTTGAACTTGAATTT	AAATTCAAGTTCAAGTTCAACG
(*mTaPR1-23*)	TCTGTTTGGTGTCTACAACTAGATTGATTT	AAATCAATCTAGTTGTAGACACCAAACAGA

Nucleic acids highlighted in red represent W-box element. Green highlight represent a nucleic acid mutation.

5.2.3 Probe annealing:

Complimentary strands of nucleic acids had to be firstly annealed before being used in EMSA's reactions. Each strand of nucleic acid probe was at 100 µM concentration. Each of complementary oligonucleotides were mixed together at a 1:1 molar ratio in a micro-centrifuge. Mixtures were diluted in dilution buffer (10 mM Tris, 1 mM EDTA, 50 mM NaCl (pH 8.0). Non-labelled complimentary mixtures were diluted to final concentration 40 pMol. Biotin labelled oligonucleotides were diluted down to 10 fmol/µL. Each pair of complimentary oligonucleotides were transferred into a thermocylcer for binding (Table 5.2). Annealed complementary oligonucleotides were stored at – 20 °C.

Table 5-2 Thermocycler program for annealing oligonucleotides.

	Cycles	Temperature	Time
Step 1	1	95 ºC	5 min
Step 2	70	95 ºC (-1 ºC/cycle)	1 min
Step 3		4 ºC	Hold

5.2.4 Control Epstein Barr Nuclear Antigen system

Control Epstein-Barr nuclear antigen system was included with the kit. The total volume for each sample was 20 µL (Table 5.3). Each binding reaction contained 20 fmol of Biotin-EBNA control DNA. The reaction was electrophoresed and transferred according to the manufacturing manual kit protocol. This was perfumed when using the kit for the first time with the control EBNA System reaction to verify that the kit components and overall procedure were adequately working.

Table 5-3 Binding reaction for control EMSA system

Component	Final Amount	Control Reactions		
		#1	#2	#3
Ultrapure water	----	12 µL	11 µL	9 µL
10X Binding Buffer	1X	2 µL	2 µL	2 µL
50% glycerol	2.5%	1 µL	1 µL	1 µL
100 mM $MgCl_2$	5mM	1 µL	1 µL	1 µL
1 µg/µL Poly (dl●dC)	50 ng/µl	1 µL	1 µL	1 µL
1% NP-40	0.05%	1µL	1µL	1 µL
Unlabelled EBNA DNA	4 pmol	----	----	2 µL
EBNA Extract	1 Unit	----	1 µL	1 µL
Biotin-EBNA Control DNA	20 fmol	2 µL	2 µL	2 µL
Total volume	----	20 µL	20 µL	20 µL

Optimized EBNA control supplemented with the light shift chemiluminescent EMSA kit was firstly used to verify kit components and to make sure that the system works properly.

Control EBNA system results in Figure 5. 1 shows three different reaction that were electrophoresed in 5% native acrylamide gel in 0.5 x TBS and transferred into supercharged nylon membrane and detected using chemiluminesce kit. Presented results indicate no shifting observed without the protein extract for DNA to binding in line #1. In terms of line #2, there was an occurrence of band shift of the biotin-EBNA DNA and this due to the presence of sufficient target protein to initiate and promote binding. The last line (#3) demonstrate that signal shift observed in line #2 was presented by the addition and competition of non-labelled DNA. With such results, this indicate that the overall of control system works as expected (Figure 5.1).

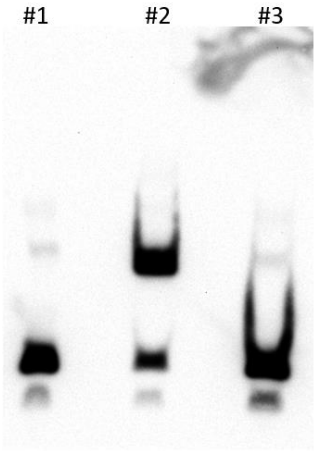


Figure 5-1 EBNA control reaction for EMSA kit verification. . #1 Biotin control DNA, #2 Biotin-EBNA Control DNA with EBNA extract, and #3 Biotin-EBNA Control DNA with EBNA extract and 200 fold molar excess of unlabelled EBNA DNA

Binding reactions for TaWRKY proteins were mixed as Table 5. 4 accordingly. The volume of ultra-pure water was firstly calculated for each sample. This was to complete reaction mixture up to 20 µL. 10X binding buffer, 50% glycerol, 100 mM MgCl2, 1 µg/µL Poly (dl●dC), and 1% NP-40 were firstly mixed together. Unlabelled W-box DNA, Protein Extract, Biotin W-box DNA (20 fmol) were added to their subsequence samples (table 5.4). Samples were incubated for 20 – 30 minutes at room temperature. Binding reactions were stopped by the addition of 5 x loading buffer to each sample. Control 1 (C1) indicate the binding reaction mixture with only 2 µg protein. As control 2, binding reaction with only Biotin labelled DNA (20 fmol). Sample 1 indicate the binding reaction mixture with protein extract as well as biotin labelled DNA. Samples 2, 3, 4 indicate binding reaction mixture with protein extract, Biotin labelled DNA and unlabelled DNA in 4 pmol, 8 pmol and 16 pmol, respectively.

Table 5-4 TaWRKY protein-DNA binding reaction

Component	Samples					
	C1	**C2**	**1**	**2**	**3**	**4**
Ultrapure water						
10X Binding Buffer	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
50% glycerol	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
100 mM $MgCl_2$	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
1 µg/µL Poly (dl●dC)	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
1% NP-40	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
Unlabeled DNA	****	****	****	4 pmol	8 pmol	16 pmol
Protein Extract	2 µg	****	2 µg	2 µg	2 µg	
Biotin labelled DNA (20 fmol)	******	20 fmol	20 fmol	20 fmol	20 fmol	20 fmol
Total volume	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL

For EMSA gel run, 5% native polyacrylamide gel in 0.5 X TBE (0.5x TBE, 5% acrylamide, 0.1% APS and 0.01% TEMED) was pre-electrophoresing for 1 hour at 100 V, wells were flushed with running buffer to remove contaminants present prior loading. To each sample, 5 µl of 5X loading buffer was added and mixed. Once samples were loaded in to each well, the current was switched on at 100 V until electrophoresis samples migrated approximately 2/3 to 3/4 down the length of the gel. The free DNA duplex migrates just behind the bromophenol blue.

Using a pre-soaked supercharged Nylon membrane (0.45 µm pore size) in 0.5 X TBE for 10 minutes, the sandwich gel containing nylon membrane and blotting paper wet in 10 °C 0.5X TBE buffer was placed in Trans-blot turbo transfer system. The transfer was conducted at 1.2 A constant and up to 20 V for 7 minutes at turbo setting. Once the transfer was completed, the membrane was cross-linked at 120 mJ/cm^2 using a commercial UV-light cross linker instrument equipped with 254 nm bulbs for 45 -60 seconds exposure (GS GENE LINKER UV CHAMBER, BIO-RAD).

For detecting Biotin-labelled DNA by Chemiluminescence (Thermo Scientific, 89880), the membrane was blocked using 1 X wash buffer for 15 minutes with gentle shaking. The blocking buffer was decanted and replaced with a conjugate/blocking solution containing Streptavidin-Horseradish Peroxidase conjugate mixed with 1 x buffer (1:300 dilution). The membrane was gently shaking for 15 minutes. It was then followed by 4 times washing using wash buffer, each for 5 minutes. Once done, the membrane was then transferred into a new container containing Substrate Equilibration buffer and incubated to be washed for 5 minutes with gentle shaking. The membrane was then subjected to an equal amount 1:1 ratio of luminol/enhancer solution and Stable Peroxide solution according to the manufacturer's instruction of LightShift Chemiluminescent EMSA Kit. The membrane was then visualized using Gel Doc (UVP ChemStudio, analytikjena) for 5 minutes up to 15 minutes for high resolution.

5.2.5 Native gel for western blot:

10% separating gel was prepared with 0.375 M Tris (pH8.8), 30% acrylamide/Bis Assay, 0.1% ammonium persulfate (APS), and 0.01% *N,N,N',N'*-tetramethyllenediamine (TEMED). The gel was mixed thoroughly and poured in gel cast immediately which is then left for about 30 – 45 minutes to solidify.

4% stacking gel was prepared with 0.375 M Tris (pH 6.8), 4 % acrylamide/Bis Assay, 0.1, APS, 0.001% TEMED. The staking gel was left for 1-2 hours for a complete solidification. Gels where normally prepared in 3 – 4 days in advance for better gel results which is kept at 4 °C.

25 mM Tris-HCl pH8.8 with 192 mM glycine (pH 8.3) running buffer was used. Samples were poured in wells at maximum of 20 µl using a gel loading tip. For sample buffer, 62.5 mM Tris-HCl (pH 6.8), 25% glycerol, and 1% bromophenol blue were mixed together and kept at - 20 °C until use.

The electrophoresis started at 50-60 V and increased to 100 V until the dye front had reached the running gel. The estimate time of running 60-90 minutes. After running, the gel was extracted from the gel glass cast using a spatula to separate the glass and cut the edge of the gel to remove the staking gel.

The gel was then soaked in a tray containing transfer buffer for western blot (48 mM Tris, 39 mM Glycine, and 20% methanol, pH 9.2). The nitrocellulose membrane and filter papers were also soaked on transfer buffer. On the semi-dry blot, a sandwich of gel, membrane and filter paper was assembled with the assurance of no air bubbles contained within.

Western blot was traditionally used for protein detection by utilizing specific antibodies to bind to expressed proteins. The nitrocellulose membrane was first rinsed with 1X PBS to the remaining running buffer and immersed in a fresh blocking solution containing 5% skimmed milk, 1X PBS, and 0.05% Tween20 and kept shaking overnight at room temperature. Later, the membrane was then washed with 1X PBS and 0,1% Tween20 to remove any residues of the blocking solution. The membrane was then immersed in a primary antibody solution (5% skimmed milk, 10% 10X PBS, and 1/1000 diluted primary antibody (6X –His Tag Monoclonal Antibody (HIS.H8), ThermoFisher) for overnight with shaking at 4 °C. Before immersing the membrane into a secondary antibody solution (5% skimmed milk, 10% 10X PBS, and 1/5000 diluted secondary antibody [Goat Anti-mouse IgG (H L)-HRP, Conjugate BIO-RAD]), it was washed with antisera buffer to remove any residues of primary antibody 3 washings each 5 minutes long. The membrane was kept in a secondary membrane shaking for at least 2 hours at room temperature or overnight at 4 °C. Finally, the membrane was washed with 1X PBS and 0.1% Tween20 for 15 minutes once and twice at 5 minutes with final wash for 15 minutes and rinsing with distilled water.

The specificity bound secondary antibody of the target protein can be detected using Enhanced Chemiluminescence (ECL) reagents (SuperSignal™ West Pico PLUS Chemiluminescence substrate, Thermo Scientific). ECL mainly contains two reagents stable peroxide solution and Luminol/Enhancer solution. A working solution was made by mixing equal parts of Stable Peroxide solution and Luminol/Enhancer solution. 0.1 ml

working solution per cm^2 of the membrane which was incubated for 5 minutes. The signal duration is stable 6 – 24 hours.

5.3 Results:

5.3.1 Target probe

In order to select the target gene for synthesised *Ta WRKY* proteins, pathogen-related protein sequences were searched by NCBI. The selected pathogen-related protein gene was *TaPR1-23* isoform (*Triticum aestivum* cultivar BR34 pathogenesis-related protein 1-23 gene, [GeneBank HQ700377.1]). The cDNA was 495 bp. To obtain the upstream region of *TaPR1-23*, the cDNA sequence was submitted in EnsemblePlants. The output result showed that one sequence had got all traits *of TaPR1-23*. The genomic location of the obtained sequence was 5A:59563139:59565633:1, and the overlapping gene was TraesCS5A02G059000.1. Orientation of gene was forward, its score was 495 in sequence, and %ID Alignment was 100.0%. From such output, 2000 bp upstream region was extracted and submitted in PlantCARe for Cis-Acting Regulatory Element. The output results had shown there was only a single W-box element at 171bp position (1829 bp upstream). Its direction was forward (**Error! Reference source not found.**).

```
>PlantCARE_7749
+ TCTTCTTGGA TTTATATCCG ATGTAATCTT CTTTTGTTGT GCATTTGTTG GGATCTAATA AATTGTGGGT
- AGAAGAACCT AAATATAGGC TACATTAGAA GAAAACAACA CGTAAACAAC CCTAGATTAT TTAACACCCA

+ TTATGATCAG ATTATTCATT GAAAGTAATT GAGTCTTTTC TGAACTTTAT TATGCATGAT TATTATAGTT
- AATACTAGTC TAATAAGTAA CTTTCATTAA CTCAGAAAAG ACTTGAAATA ATACGTACTA ATAATATCAA

+ TTATATTTAT CTTCGATCTA TCTGTTTGGT TTGACCAACT AGATTGATTT ATCTTCAGTG GGAGAGGTGC
- AATATAAATA GAAGCTAGAT AGACAAACCA AACTGGTTGA TCTAACTAAA TAGAAGTCAC CCTCTCCACG

+ TTTGTGATGG GTTCAATCTT GTGGTGTCCT CACCTCCTGA TAGAAGGGGT AGCGAGGCAT ACATTGTATT
- AAACACTACC CAAGTTAGAA CACCACAGGA GTGGAGGACT ATCTTCCCCA TCGCTCCGTA TGTAACATAA

+ GTTTCCAATA AGGGTAAAAT GGTGGAGTTT ATTCTTATTG CTTGAGTTTA CTTTGTCTAC ATCATATCAT
- CAAAGGTTAT TCCCATTTTA CCACCTCAAA TAAGAATAAC GAACTCAAAT GAAACAGATG TAGTATAGTA
```

Figure 5-2 *TaPR1-23* nucleotide sequence promoter region. Highlighted sequence indicate W-box element as candidate Cis-acting regulatory element. W-box is located at -1829 upstream *TaPR1-23*. The Cis-acting regulatory element was obtained by PlantCARE

5.3.2 Electrophoretic mobility shift Assay:

Candidate wheat WRKY proteins that had been expressed using *Pichia Pastoris* and purified were concentrated and employed for Electrophoretic Mobility Shift Assay. Biotin Labelled probes (hot probe), and non-labelled probes (cold probe) were selected as described in methodology and synthesised in the form of forward and reverse oligonucleotides. As Recommended by Thermo Scientific Tech Tip #45 for annealing complementary pairs of oligonucleotides, 20 fmol of hot oligos and 40 pmol of cold oligos were diluted with annealing buffer (Tris containing salt, pH 7.5) and annealed using a

Thermocycler (BioRad, T100). Gel mobility shift assay was performed using *TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and *TaWRKY3m* mutants.

Purified proteins were used to interact with DNA sequences containing W-box elements. DNA sequences were selected and synthesised in forward and reverse sequences to be annealed to form a complementary DNA sequence. From each, two sequences were synthesised. One set was Biotin labelled serving as a hot probe. The second set of the same complementary DNA was non-labelled, serving as a cold probe. 200– 800 fold molar excess of cold competitor probe was added to test the speciality of hot probe binding to proteins. The addition of a cold probe is to compete with the hot probe in binding to the protein. As the competitor cold probe concentration increase in the reaction mixture, the less binding of Biotin labelled probe to the protein. This would show that the reduction in the binding signal and protein binding specificity can be indicated by the loss of binding to the Biotin labelled probe.

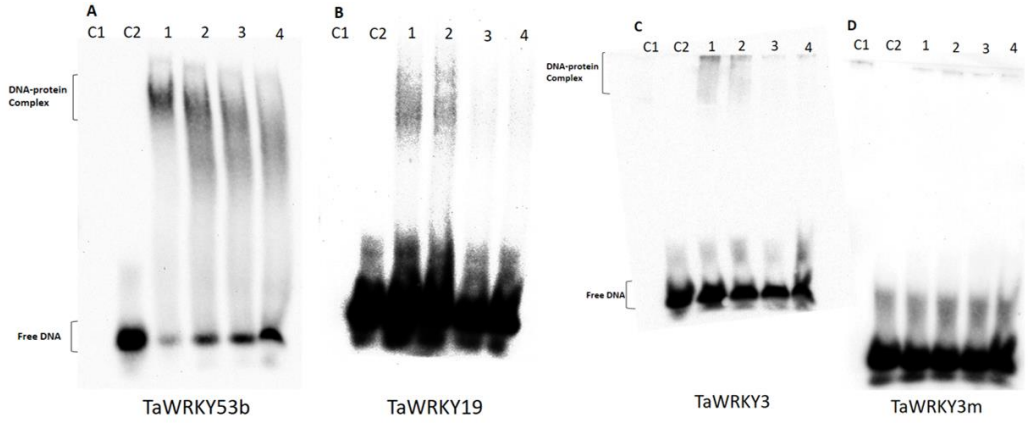


Figure 5-3 Parsley Pathogen related gene promoter region synthesised probe (*PcPR1*) complex with TaWRKY proteins (A, *TaWRKY53b*; B, *TaWRKY19*; C, *TaWRKY3*; *TaWRKY3m*). The binding activity of *PcPR1* was demonstrated by 20 fmol of hot probe and competition assay using 200-800-fold excess unlabelled probe. Protein concentration was roughly 2 µg at each protein. C1, TaWRKY protein with no probe (Control 1); C2, only labelled probe (control 2); 1, TaWRKY protein with 20 fmol Biotin labelled DNA; 2, TaWRKY protein with 20 fmol Biotin labelled DNA and 4 pmol cold DNA; 3, TaWRKY protein with 20 fmol Biotin labelled DNA and 8 pmol cold DNA; 4, TaWRKY protein with 20 fmol Biotin labelled DNA and 16 pmol cold DNA.

Parsley pathogenesis-related oligonucleotide (*PcPR1-1*) promoter containing one W-box was employed to test its binding ability to *TaWRKY53b*, *TaWRKY19*, *TaWRKY3*, and *TaWRKY3m* proteins. In each protein-DNA binding test, ~ 2 ng of purified protein was

used. 20 fmol of biotin labelled probe and the corresponding competitor probe at a range of concentrations were used (4 pmol, 8 pmol, and 16 pmol) (Figure 5. 3). Controls had been added into each protein-DNA interaction experiment. Control 1 was contained only protein with the reaction mixture. Control 2 contained oligonucleotide probe with the reaction mixture. In *TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and its mutant form, all C1 wells had shown clear wells in each tested protein. Indicating that the assays were working correctly and that there was no contamination. In each C2, there was no presence of an unshifted band and only free Biotin labelled DNA. This indicates that the probe labelling does work, with no protein contamination within C2 control (Figure 5. 3).

Among all proteins, *TaWRKY53b* protein had shown the highest level of binding to the parsley PR1 oligonucleotide (Figure 5. 3 A). When using competitor probes at different concentrations, a reduction in binding the labelled probe was clearly shown as the cold probe concentration increases from 4 pmol to 16 pmol (Figure 5. 3 A).

In terms of *TaWRKY19*, there was a faint binding between *TaWRKY19* and *PcPR1*. As 4 pmol of the cold probe was added in the reaction mixture in well 2, the level of binding showed a reduction compared to well 1. The shade of binding was remarkably reduced once 8 pmol of cold probe was added in well 3. 16 pmol had shown the same reduction of shade of binding 8 pmol (figure 5.3.B). The intensity of binding between *TaWRKY19* and hot probe of *PcPR1* in competitor presence shows a strong competition across well 2, 3 and 4 (figure 5.3 B).

TaWRKY3 protein had shown an interaction with *PcPR1*. The interaction was clearly present on the top of the gel. The increase of cold competitor probe reduces the binding signal across well 2 and 3 and 4. Signal reduction indicates that the binding of *TaWRKY3* and *PcPR1* was specific (Figure 5.3 C).The mutated version of *TaWRKY3* (*TaWRKY3m*) had shown no binding activity across all wells containing *PcPR1* (Figure 5.3 D).

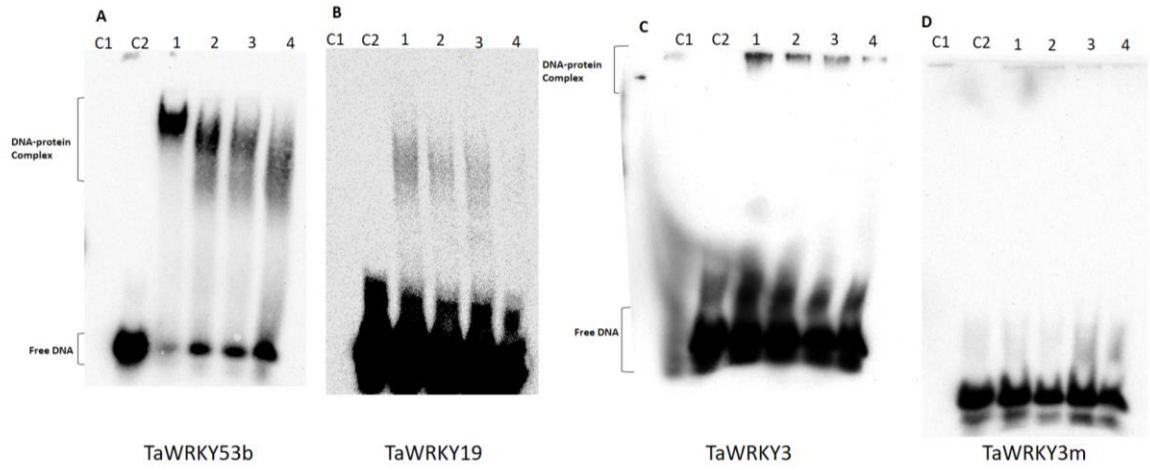


Figure 5-4 Wheat Pathogenesis related gene promoter region synthesised probe (*TaPR1-23*) complex with TaWRKY proteins (A, *TaWRKY53b*; B, *TaWRKY19*; C, *TaWRKY3*; *TaWRKY3m*). The binding activity of PcPR1 was demonstrated by 20 fmol of hot probe and competition assay using 200-800 fold excess unlabelled probe. Protein concentration was roughly 2 µg at each protein. C1, TaWRKY protein with no probe (Control 1); C2, only labelled probe (control 2); 1, TaWRKY protein with 20 fmol Biotin labelled DNA; 2, TaWRKY protein with 20 fmol Biotin labelled DNA and 4 pmol cold DNA; 3, TaWRKY protein with 20 fmol Biotin labelled DNA and 8 pmol cold DNA; 4, TaWRKY protein with 20 fmol Biotin labelled DNA and 16 pmol cold DNA.

In terms of *TaPR1* probe in Figure 5. 4, the binding signal shift of *TaWRKY53b* did exhibit a similar strong binding signal shift compared to *PcPR1* probe. As cold competitor probe was added into the reaction, the binding signal was reduced sequentially as the concentration of cold probe increased (Figure 5. 4 A). The majority of the labelled probe was bound by the recombinant *TaWRKY53b*.

Similar attributes of *TaWRKY19* occurred when interacting with *TaPR1-23* as occurred in *PcPR1*. The interaction could be considered to be very weak as only a small amount of labelled probe is bound. With the addition of competitor, the binding signal shows a reduction in wells 2 and 3. There was no signal interaction at well 4 (16 pmol) indicating that there is a specificity of binding between *TaWRKY19* and *TaPR1-23*. The binding to *TaPR1-23* is more sensitive than *PcPR1* as *TaWRKY19* as it can bind to the probe at a high competitor (Figure 5.4 B).

TaWRKY3 had shown a signal on the top of the membrane at the side of the wells (figure 5.4C). The signal on *TaWRKY3-TaPR1* interaction was visible. This attribute also occurs with protein-DNA and non-labelled DNA. As it shown in from 2-4, the signal decreases as

the competitor DNA increases. The protein-DNA complex did not migrate very far into the native gel (Figure 5.4 C). No *TaWRKY3m* was bound to *TaPR1* probe (Figure 5. 4 D).

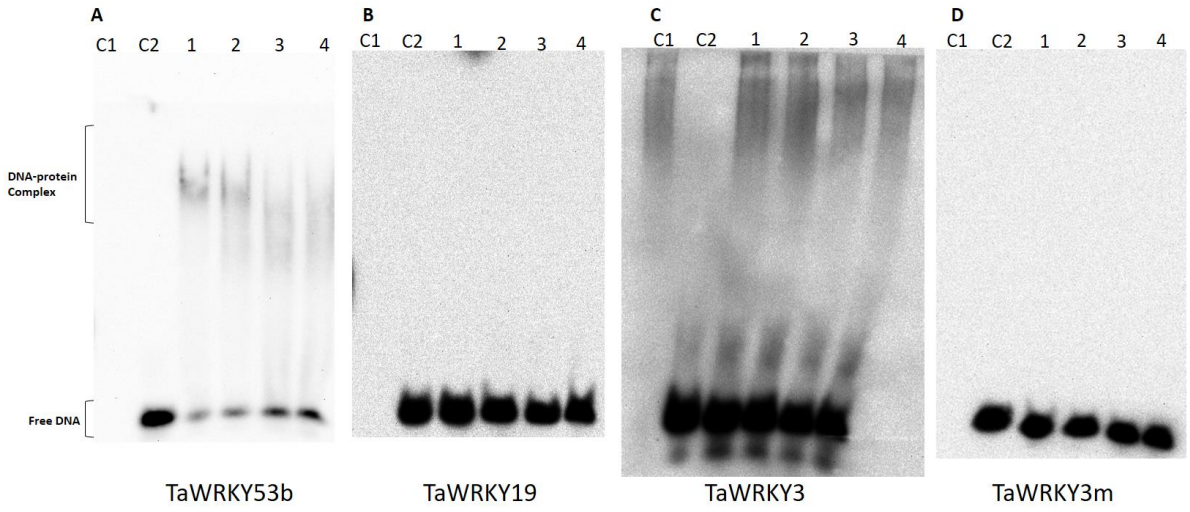


Figure 5-5 Mutant wheat Pathogen related gene promoter region synthesised probe (*TaPR1*) complex with TaWRKY proteins (A, *TaWRKY53b*; B, *TaWRKY19*; C, *TaWRKY3*; *TaWRKY3m*). The binding activity of PcPR1 was demonstrated by 20 fmol of hot probe and competition assay using 200-800 fold excess unlabelled probe. Protein concentration was roughly 2 µg at each protein. C1, TaWRKY protein with no probe (Control 1); C2, only labelled probe (control 2); 1, TaWRKY protein with 20 fmol Biotin labelled DNA; 2, TaWRKY protein with 20 fmol Biotin labelled DNA and 4 pmol cold DNA; 3, TaWRKY protein with 20 fmol Biotin labelled DNA and 8 pmol cold DNA; 4, TaWRKY protein with 20 fmol Biotin labelled DNA and 16 pmol cold DNA.

mTaPR1-23 probe is a mutated form of *TaPR1-23*. The mutation was done within the W-box element nucleotides (TTGACC) into GTCTAC. This was to test the protein specificity of binding into the mutant W-box element.

The combination of *TaWRKY53b* with mutant *TaPR1-23* (*mTaPR1-23*) DNA had shown a very weak binding shift (Figure 5.5 A). *TaWRKY19*, *TaWRKY3* and *TaWRKY3m* reaction mixtures had shown no binding shift of *mTaPR1-23* (Figure 5.5 B, C, and D respectively).

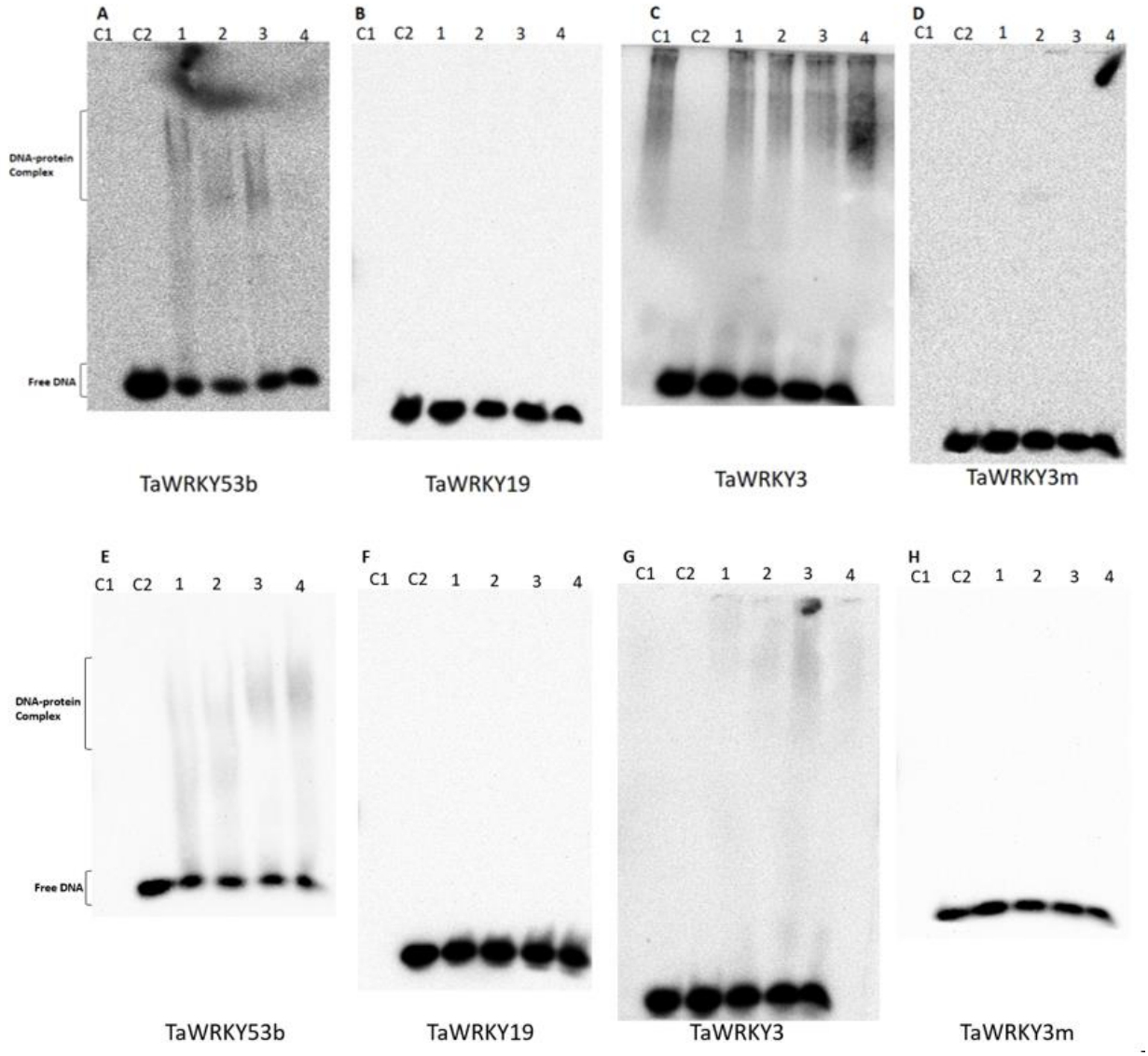


Figure 5-6 multiple W-box element probe complex with TaWRKY proteins (A, *TaWRKY53b*; B, *TaWRKY19*; C, *TaWRKY3*; D, *TaWRKY3m*). Mutant (multiple) W-box element probe complex with TaWRKY proteins (E, TaWRKY53b; F, TaWRKY19; G, TaWRKY3; H, TaWRKY3m). The binding activity of W-box and mW-box was demonstrated by 20 fmol of hot probe and competition assay using 200-800 fold excess unlabelled probe. Protein concentration was roughly 2 µg at each protein. C1, TaWRKY protein with no probe (Control 1); C2, only labelled probe (control 2); 1, TaWRKY protein with 20 fmol Biotin labelled DNA; 2, TaWRKY protein with 20 fmol Biotin labelled DNA and 4 pmol cold DNA; 3, TaWRKY protein with 20 fmol Biotin labelled DNA and 8 pmol cold DNA; 4, TaWRKY protein with 20 fmol Biotin labelled DNA and 16 pmol cold DNA.

In terms of the synthetic multiple W-box (W-box) and synthetic multiple mutated W-box oligonucleotides, *TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and *TaWRKY3m* proteins had shown no binding shift at both probes (Figure 5.6).

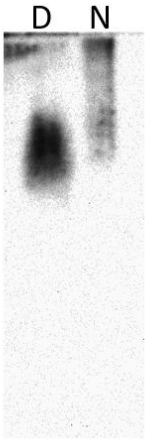


Figure 5-7 *TaWRKY3* protein western blot. D; Denatured protein at 100 °C for 10 minutes, N; Native protein.

To further identify the upper band appeared in *TaWRKY3* protein binding to *PcPR1* and *TaPR1* oligonucleotide probe present in Figure 5. 3 C and Figure 5.4 C, *TaWRKY3* protein was run in 10% native gel. Denatured proteins were used as a tool of comparison to native protein. In order to prevent denatured proteins from renaturing, SDS-loading buffer was added prior to heat treatment. Native sample buffer with was added to *TaWRKY3* protein and both were loaded into native gel for electrophorese. As it appears in figure 5.7, Native *TaWRKY3* tend to be aggregated on the top of the nitrocellulose membrane and smearing along the way similarly to protein-DNA complex present in Figure 5. 3 C and Figure 5. 4 C. On the other hand, denatured *TaWRKY3* tends to shown a migration through the gel and that had reflected, by its position, on the nitrocellulose membrane.

5.4 Discussion:

It had been reported that plant species contain DNA-binding proteins with WRKY domains that interact with Cis-regulatory w-box elements found in the promoter region of target gene. Binding of transcription factors proteins to specific sequences of DNA was found to be crucial for transcriptional regulation. This can be either induce expression and/ or repression of the target gene. Studies on protein-DNA binding mainly conducted to determine their interaction through either *in silico* and *in vitro* as well. This study evaluated the protein-DNA interaction *in vitro* using heterologously expressed *TaWRKY* transcription factors proteins and short regions of DNA homolgous to known promoter regions containing W-box elements and also to synthetic oligonucleotides containing one or more W-box repeats. From previous chapters, *TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and *TaWRKY3m* proteins were cloned into pGAPZαA and transformed into *Pichia pastoris* for fusion protein expression. Proteins where purified in their native form using 6xHis-tag before being used in Electrophoretic mobility shift assay (EMSA).

In this study, *TaWRKY* proteins expressed in a eukaryotic heterologous system were examined to determine their ability to natively bind to promoter regions. It had been widely demonstrated in literature that WRKYGQK region of amino acids in these transcription factors was found to be an important protein region for DNA-binding. All previous studies attempted to produce recombinant WRKY transcription factors in prokaryotic host (*E. coli*) had was problematic. This includes inability to express a full sequence of protein due to its detrimental impact to their host by negatively impacting zinc homeostasis. Even with the success of expression, WRKY proteins were found to be expressed in inclusion bodies. This required harsh purification steps to release WRKY proteins from cells. This involved denaturation and renaturation of WRKY protein causing the misfolded protein to loss their binding activity to W-box (Ciolkowski *et al.*, 2008b). This present study is the first to employ the eukaryotic yeast, *P. pastoris*, as the expression host.

5.4.1 Binding Assays:

To determine wheat WRKY proteins (*TaWRKY53b*, *TaWRKY19*, *TaWRKY3*, and *TaWRKY3m*) binding activity, five short (22-30 bp) DNA sequences containing W-box

elements were designed. One of which was parsley pathogenesis related (PR) gene (*Petroselinum crispum* [*PcPR1*]) containing one W-box (W2) element. It had shown a specific binding to WRKY transcription factors (Rushton *et al.*, 1996; Rushton *et al.*, 2002; Ciolkowski *et al.*, 2008a). Thus, it was selected as a positive control to determine the capability of eukaryotic expressed recombinant wheat WRKY proteins to actively bind to w-box elements. In addition, a wheat pathogenesis related gene (*TaPR1-23*) was also selected for binding assay. -2000 bp upstream starting nucleic acids motif scan of *TaPR1-23* using PlantCARE database for cis-acting elements determination had shown only one W-box motif located at -1823 bp with a core sequence TTGACC. As a result, 30 bp oligonucleotides was designed and synthesised for binding assay. A mutation within *TaPR1-23* (*mTaPR1-23*) was conducted within the W-box motif was also designed and synthesised. In addition, multiple W-box elements in one oligonucleotides as a three tandem repeats and a mutant form were also synthesised (NIU *et al.*, 2012). Each of sequence was synthesised with biotin label for binding determination and also, for the same sequences, synthesised with no label acting as competitor sequences. Each of oligonucleotides were designed as forwards and reverse strand to form a double stranded DNA which was later annealed for EMSA assay.

The gel shift assay using *TaWRKY* proteins were firstly tested on *PcPR1* probe. *TaWRKY53b* had shown a very clear binding among all tested proteins. Its binding was confirmed with the addition of competitor *PcPR1* probe. This is an indicator of true binding of *TaWRKY53b* to *PcPR1* oligonucleotides (figure 5.3 A). *TaWRKY19* and *TaWRKY3* had shown also binding to *PcPR1* but not as strong and specific binding as *TaWRKY53b* (Figure 5.3 B and C, respectively). In terms of *TaWRKY3m* did not show binding (figure 5.3 D). In terms of *TaPR1-23*, the same binding pattern was also remarkably found as *PcPR1* with all tested proteins (figure 5.4 A, B, C, and D). However, *mTaPR1-23* a partial binding of *TaWRKY53b* was only found but no binding of *TaWRKY19*, *TaWRKY3* and *TaWRKY3m* (figure 5.5 A, B, C, and D, receptively). No binding among all recombinant wheat WRKY proteins using tandem repeats of W-box and its mutant form (figure 5.6). Such binding suggests that eukaryotic recombinant expressed native proteins were capable of binding to W-box elements.

As *TaWRKY53b* was found to be bound to *PcPR-1* and *TaPR1-23* strongly. In comparison to our work with regards to *TaWRKY53b*, a study was conducted by Van Eck (2014) looking at an *in vitro* the binding assay of an expressed *TaWRKY53* protein to three rice genes with different functional categorise. Their studies were to assess Ser/Thr-type receptor kinase ORK10 (ORK1) as an induced gene in infected cereals by biotrophic rust fungi. Apoplastic cationic peroxidase (POK1) which is induced, as part of oxidative rust, by the infestation of *X.oryzae pv. Oryzae* and aphids. *Chitinase-2* which is a pathogenesis-related (PR) protein that expressed in response to biotic stressors as aphids and fungus. The biotinylated oligonucleotides were selected from 1 kb upstream region. Each gene contained four or more W-box. Their protein-DNA binding assays had shown that *TaWRKY53* binds to promoter fragments of POC1 and ORK10 whereas *chitinase-2* has shown no binding to *TaWRKY53*. This is an indicative of *TaWRKY53b* tendency to contribute in plant defence against biotic stressors.

A partial or fainted binding of *TaWRKY19* toward *TaPR1-23*. This indicate there might be a partial involvement of *TaWRKY19* to defence. NIU *et al.* (2012) expressed truncated *TaWRKY19* and *TaWRKY2* using *Escherichia coli* to test their binding toward synthetic multiple W-box elements in one oligo as a three tandem TTGACC repeats. A mutant form of such oligo was also synthesised. In their study, *TaWRKY19* and *TaWRKY2* were found to bind to the labelled multiple W-box element. This indicated that *TaWRKY19* bind specifically to multiple W-box element. Other promoter sequences for RD29A (-580 to -541), RD29B (-904 to -870), Cor6.6 (-761 to -710), and DREB2A l (-2113 to -2077), DREB2A ll (-2005 to -1972), and DREB2A lll (-1798 to -1762) were also tested. Their findings had shown a specific strong binding of *TaWRRKY19* toward W-box element of Cor6.6 but very weak binding to RD29A and RD29B promoter sequences. A specific binding of *TaWRKY19* to DREB2A gene and that was DBEB2A l and lesser in DBER2A ll. The least binding or could be said a weaker binding was found in DBER2A ll. In comparison to our findings, much of its specificity bindings were found to be abiotic genes rather than biotic. So when we exposed *TaWRKY19* to *TaPR1-23*, the binding was partial. This could suggest that *TaWRKY19* tend to play a role in abiotic stressors rather than biotic.

TaWRKY3 was found to be bound to *PcPR1* and *TaPR1-23* in our binding assay. However, no binding was observed using *TaWRKY3m*. In a similar bindings assay conducted by

Alshgaihi, et al (2018) (unpublished data) using unpurified *TaWRKY3* expressed using *E. coli* had shown potential binding to all *PcPR1*, *TaPR1*, 3X W-box and 3X mutant W-box. The mutant form did not shown any binding with any of oligonucleotides.

5.4.2 W-box elements:

Following that many papers were using such oligonucleotides to test the ability of expressed WRKY transcription factors binding. Our findings had shown no binding of natively expressed proteins to repeated W-box motifs. With the partial binding of *TaWRKY53b*, it is an indicative of flanking regions contribution of binding specificity and selectivity of promoter regions. With W-box repeats, no flanking regions was involved between each W-box element. This inhibited the binding of the TaWRKY proteins. Indeed, W-box motif is an important element for recognition by TaWRKY proteins. However, our results shown that their inability to be recognised by recombinant proteins. In comparison to our work, Zhou *et al.* (2008) was firstly included multiple W-box elements as tandem repeat of TTGACT in their binding assays. They found that WRKY proteins such as *GmWRKY27*, *GmWRKY54*, *GmWRKY13* were capable of binding repeated W-box element but not with its mutant form. a similar W-box tandem repeats was experimented by (NIU *et al.*, 2012) which had shown binding by *TaWRKY2* and *TaWRKY19* proteins. In our lab, *TaWRKY3* and its mutant form were also tested for their binding to 3x w-box element and its mutant form. There was a very strong binding to both 3x W-box and 3x mutant W-box. For justification, bindings assay conducted by Zhou *et al.* (2008) and (NIU *et al.*, 2012) were done by using truncated expressed WRKY proteins containing only DNA binding domain where as *TaWRKY3* and *TaWRKY3m* expressed by *E. coli* was full protein expression. In both truncated and full protein expression, WRKY proteins were found to be bound to w-box element. Due to the nature of proteins using *E. coli* by non-post-translational modification as well as exposure of proteins to denaturation and renaturations, this could be resulted WRKYGQK to be exposed to DNA and lead to binding. in other words, the exposure of WRKY region due to the lack of posttranslational modification and denaturation and renaturation had enable proteins to bind to 3X W-box. This can lead to reduce the specificity of WRKY proteins to be bind to their target genes. In our case proteins were expressed fully by *Pichia pastoris* and purified in its native form. As a results, this had improved the selectivity of such proteins to be bound to target DNA by binding to *PcPR1* and *TaPR1-23* and not to W-box tandem repeats. This improvement

was contributed by the post-translation modification through its expression by Pichia as well as purification in its native folding.

5.4.4 Dimerization

With the protein-DNA binding assay, a peculiar shifting of *TaWRKY3* when it was bound to *PcPR1* and *TaPR1-23*. It was clearly noticed that the migration of protein-DNA complex was difficult to move across the native gel. In order to understand that issue, a comparison between native proteins against denatured proteins was done in a native gel. Denatured *TaWRKY3* proteins were heated at 100 °C for 10 minutes and SDS-loading dye was added prior to heating. This was to make sure denatured *TaWRKY3* protein do not refold back again at the native electrophoresis. On the other hand, native proteins were loaded into the gel with the addition of native loading dye. Right after running in the gel proteins were transferred into the nitrocellulose membrane. Western blot images of such comparison shown the migration of denatured *TaWRKY3* proteins across the gel. Native proteins (10% acrylamide) were aggregated at the top of the membrane. This could be justified by the presence of hyper-glycosylation formed by the Pichia post-translational processing and/ or multi-dimerization of TaWRKY proteins. Both can increase the molecular weight of proteins and causes it to migrate less in the acrylamide gel. A reduction in acrylamide gel less than 4% can lead to a fracture in the gel and lose the sample.

WRKY proteins were previously studied by Cheng *et al.* (2019) though looking at its structural bases and the capability of WRKY proteins to form dimerization. A truncated form of *OsWRKY45-DBD* (rice WRKY proteins) was cloned and expressed by E.coli and purified by Nickel chelating beads. Their protein crystallization extermination had shown a formation of homodimerization of *OsWRKY45-DBD*. As they had illustrated the molecules itself consists of five ß strands (ß1 – ß5). Two molecules of the *OsWRKY45-DBD* tend to form a dimerization by exchanging ß4- ß5 in antiparallel manner. They as also added that WRKY proteins can also form a tetramers. In link to our results, the shift of *TaWRKY3* proteins occurred can be said that it was caused by the formation of dimerization between proteins. This had resulted to the less migration of proteins in the native gel. In terms of protein-DNA binding, the authors also looked at the impact of dimerization of proteins toward its ability to bind to W-box elements in oligonucleotides. Their findings had illustrated such dimerization of proteins can also promote W-box DNA

recognition as well as binding. The binding of proteins to DNA can be correlated to the number of dimers. Two dimers of proteins can bind to a dual W-box DNA. This indicated that the formation of proteins DNA binding in the form of dimerized proteins could not be affected. However, such formation tend to greatly increase the molecular weight of protein-DNA and this, as result, impact their migration in native gel electrophoresis. As it can be clearly seen, *TaWRKY3-TaPR1* binding image apparently visible at the top of the membrane. It was thought, firstly, as an aggregation of proteins. The addition of non-biotinylated *TaPR1-23* probe had illustrated that protein-DNA binding was a true binding. This was thought the reduction of the signal of biotinylated with the increases of the competitor in the reaction. Thus, we could conclude that this reaction was a true reaction and the issue of migration of protein-DNA complex was greatly caused by the multi-dimerization of proteins as well as the binding of *TaPR1-23*.

In order to prevent a dimerization of proteins, it could be useful to express WRKY proteins at their DNA binding domain. This would include WRKY and its zinc finger element. This method was adopted by NIU *et al.* (2012) for their *TaWRKY2* and *TaWRKY19* expression, Cheng et al. (2019) for *OsWRKY45-DBD* and many other in the literature. Through the use of truncated proteins, it could be possible to express and WRKY proteins monomers with no dimerization occurrence.

5.4.5 Binding affinity:

The binding affinity of WRKY transcription factors to w-box elements was found to occur on both sides. This were WRKYGQK binding region recognise W-box element specifically based on their sequence (T-TGAC-C). However, the specificity of such proteins occurs in biochemical interaction. It was reported that the electrostatic potential molecular surface of the DNA binding domain of WRKY TF was found to be positive change. This was dictated by the presence of at least of eight positively charged amino acids at within the WRKY 60 amino acid sequence. However, the backbone of DNA double helix negatively charge (Pandey;Grover and Sharma, 2018).

With such interactions, it was reported that the presence of WRKYGQK amino acids binding site was crucial for the specificity of recognition of W-box element in target gene promoter. The specificity was merely determined with the presence of interaction between WRKYGQK amino acids and TTGACC nucleotides (Eulgem et al., 2000). Maeo et al. (2001) and Duan et al. (2007) had concluded that certain amino acid residues in WRKY TF play a crucial role for w-box recognition. This was mainly determined by multiple

mutations at WRKYGQK protein restudies in their binding assays. Their findings had shown that mutation on WRKYGQK residues replaced by alanine had abolished the binding of protein toward W-box element. . Duan et al. (2007) investigated further the role of mutations on WRKYGQK as well. They had mutated the WRKY-DBD with different amino acids. Their findings had presented a detrimental impact on the binding when substituting K to A, Y to R, and G to F. They had concluded that K, G, K were found to be crucial for DNA-specific recognition in their EMSA assays.

With protein-DNA crystallization, more detailed information was extracted with regards to WRKY-DBD to W-box element. The interaction between WRKY-DBD and W-box occurred in 3 distinct interactions which are hydrophobic interaction, H-bond, electrostatic interaction. Much of such interactions occurs between K-G-K (Duan *et al.*, 2007) and the W-box element at both forward and revers strands. With regards to tyrosine (Y) and glutamine (Q) amino acids, it was thought mutation at these residues might reduce binding or have no effect to binding (Duan *et al.*, 2007).

For tyrosine (Y), mutation to arginine and alanine had also shown no binding to DNA in both Maeo *et al.* (2001) and (Duan *et al.*, 2007) assays . With regards to *TaWRKY3m*, a single mutation occurred in tyrosine to Aspartic acid. It had shown also no binding at all in our EMSA results to parsley *PR1-1* promoter whereas *TaWRKY3* showed a binding. Generally, the charge of tyrosine is polar charge (neutral) whereas aspartic acid (D) is – negative charge. The presence of a negative charge amino acid in the DNA binding domain could cause a repel between negative protein binding domain and negative charge DNA sequence. Additionally, tyrosine binding interaction was shown by Yamasaki *et al.* (2012), Cheng *et al.* (2019), and Xu *et al.* (2020) had 5 hydrophobic interactions and 4 hydrogen bonds to the antiparallel W-box element, specifically to TGAC nucleotides. Mutation to a negatively charged aspartic acid led to not binding to *PcPR1-1* and *TaPR1-23*. This suggests that preservation electrostatic bonds, as well as hydrophobic interaction and H-bonds with present tyrosine, sustains binding to both *PcPR1-1* and *TaPR1-23* in TaWRKY3. As Duan *et al.* (2007) concluded that (K-G-K) in the WRKYGQK were found to be crucial for DNA specific recognition. Our findings suggest tyrosine in WRKYGQK also found to be essential for the specific recognition of DNA.

As Yamasaki *et al.* (2012), Cheng *et al.* (2019), and (Xu *et al.*, 2020) indicated that Tyrosine mostly dominate the binding into 5'-TT-3' and 3'-CT-5' of the W-box element (TTGACC) but both in its forward strand (5' to 3') and reverse strand (3' to 5'). The binding of the protein would implies on both Waltson strand ($T_6T_7G_8A_9C_{10}C_{11}$) and Crick strand ($A_{11'}A_{10'}C_{9'}A_{8'}T_{7'}G_{6'}G_{5'}$). Ciolkowski *et al.* (2008a) was testing multiple WRKY-DBD to multiple mutated *PcPR1-1* (W2) genes. Their mutation work covered W-box element nucleotides and expanded to flanging regions as well. To link the role of Tyrosine for group II and Group III Arabidopsis WRKY proteins (*AtWRKY6-IIb*, *AtWRKY43-IIC*, and *AtWRKYIId* and *AtWRKY38-III*). Mutations occurs at TTG nucleotides of TTGACC of W-box element such as M18 (TTCACC), M19 (TAGACC), M5 (CTGACC), M6 (GTGACC), and M7 (ATGACC). As T_6T_7 and $C_{9'}T_{8'}$ of the W-box element was found to be crucial tyrosine binding. Mutation at such nucleotides had shown to be detrimental for WRKY-DNA specific recognition at their EMSA assays. This is suggests the importance of conserved of KYG-K of the WRKYGQK and the W-box element (TTGACC) for an absolute specific recognition of the W-box element by WRKY transcription factor.

Despite the importance of the W-box elements, flanking regions could partially participate to the WRKY-DBD to its target promoter gene. Our EMSA assays had shown a very strong binding of *TaWRKY53b* to *TaPR1-23*. Using a complete mutation of *TaPR1-23* at its W-box element but with a preservation of flanking regions sequence. Our test assay had shown a partial binding of *TaWRKY53b*. Yamasaki *et al.* (2012), Cheng *et al.* (2019), and Xu *et al.* (2020) had shown in their crystallisation of WRKY-DBD and W-box binding that both Tryptophan (W) and arginine (R) of WRKYGQK tends to bind to upstream nucleotides of w-box element (TTGACC). In terms of *TaWRKY53b*, indeed *mTaPR1-23* mutation occurred at its W-box element. However, the binding could had been facilitated partially by the Tryptophan and Arginine of the WRKYGQK that bound to 5'-GT-3' upstream of the mutated W-box. This study also illustrated the binding capability of natively expressed Wheat WRKY proteins to bind to promoter regions of their target genes.

6. General discussion and future perspectives:

The overarching aim of the current study was to better understand how WRKY transcription factors in wheat (*Triticum aestivum*) regulate gene expression by investigating the ability of recombinant WRKY proteins (*TaWRKY53b*, *TaWRKY19*, *TaWRKY3*, and *TaWRKY3m*), expressed in *Pichia pastoris*, to bind to promoter of target genes. The following chapter discusses the most important outcome of this work.

Candidate wheat type WRKY transcription factors (TaWRKY53b, TaWRKY19, TaWRKY3, and TaWRKY3m) were studied in *In silico*. This was to identify their transcriptional regulatory elements. The study was followed by the cloning and expression of candidate WRKY using the methylotrophic yeast *Pichia pastoris* as the expression platform. Purified recombinant expressed WRKY TFs proteins were successfully obtained using conventional chromatography methods. Purified TaWRKY proteins were subjected to wheat defence promoter DNA sequences using electrophoretic mobility shift assay. Our findings illustrated that wheat WRKY proteins could be recombinantly expressed using *P. pastoris.* It had been presented that their expression does not have an impact on the host expressing organism and also the ease of purification as well. In addition, the protein-DNA binding had demonstrated the activity of fully folded protein interaction to promoters of target genes. With such findings, the expression of total TaWRKY proteins using a eukaryotic expression system could be an initial step for scientists to express and purify WRKY transcription factors in a eukaryotic system rather than prokaryotic.

Candidate wheat WRKY genes (*TaWRKY53b*, *TaWRKY19*, *TaWRKY3*, and *TaWRKY3m*) were firstly studied *in silico* (chapter2). Based on their WRKY DNA binding domains (DBD), *TaWRKY53b* and *TaWRKY19* were identified as Group I WRKYs carrying trans and Cis domains WRKY DBD. TaWRKY3 belongs to group IIc and contained one WRKY DBD. All of which were found to be localized in the nucleus. Their promoter region determines that their expression was regulated by transcription factors such as bZIP and ethylene responsive transcription factors. TaWRKY protein-protein interaction using STRING analysis was less informative. This is because of gaps in current databases.
The protein coding sequence of *TaWRKY53b*, *TaWRKY19*, *TaWRKY3* and *TaWRKY3m* were cloned into pGAPZαA expression vector and then transformed into *Pichia pastoris*

for protein expression. For identification and purification, each construct was carrying 6x His-tag. All WRKY proteins expression was assessed in small scale production which had shown an expression (chapter 3). Large-scale expression of *TaWRKY53b* and *TaWRKY19* proteins was conducted in 5 liter bench-top fermentation unit using basal salts media (BSM). The purification of secreted proteins was following a pipeline which was hydrophobic interaction and nickel column. With regards to *TaWRKY3* and *TaWRKY3m*, hydrophobic interaction chromatography purification was found to be incompatible. When purifying *TaWRKY3* and *TaWRKY3m* in nickel-column, phosphates in BSM were found to precipitate. Thus, M3 media was used as an alternative media in fermentation and this was to use fermentation broth as one step purification (chapter 4).

Purified recombinant WRKY proteins were used to demonstrate their capability to bind to W-box from plant promoters using electrophoretic mobility shift assay (EMSA). Purified WRKY proteins were interacted against synthetic W-box and mutated w-box repeats. It also interacted promoter fragments from wheat *PR1-23* and parsley *PR1* as well. Each plant promoters contain W-box elements. The interaction of was reduced during EMSA between *TaWRKY3* and promoters due to dimerization. No W-box binding was found in the *TaWRKY3m* which was due to the mutation in the WRKY DNA binding domain. *TaWRKY53b* and *TaWRKY19* were found to bind to *TaPR1-23* and *PcPR1* promoter fragments. This suggested that flanking regions outside the W-box element core are required for optimal binding. No binding was found between recombinant WRKYs and mutated W-box repeats (chapter 5).

With the use *P. pastoris* in this project, WRKY transcription factors soluble correctly folded recombinant proteins were expressed. This showed that their interaction capability to specifically binds to target gene promoter elements. The regulatory network controlling the expression of wheat WRKYs were highlighted in this project. This approach of expression soluble correctly folded WRKY proteins could be used to generate functional and highly purified WRKY transcription factors. This could be used for gene identification; they regulate and elucidate specific stress responses.

All previous researchers have used an *E. coli* system to produce recombinant WRKYs. Whilst this system is highly economical and rapid there are many drawbacks when attempting to produce eukaryotic protein in this prokaryotic system. The section of this thesis with the most impact for research and translational science are those where we

demonstrate the expression of functional WRKY transcription factors in *Pichia pastoris* and show how modifications to the culturing conditions can be exploited to produce a wider range of proteins for industrial and commercial stakeholders.

6.1 Expression of TaWRKYs in Pichia pastoris overcomes challenges of bacterial inclusion body formation

Previous researchers have expressed recombinant WRKY transcription factors proteins using *E. coli*. Ciolkowski *et al.* (2008b) show that TaWRKY proteins in *E. coli* are problematic due to the zinc-finger proteins negatively affecting zinc homeostasis resulting in poor bacterial growth. The *P. pastoris* expression platform did not suffer the same problems and allowed for the recovery of functional WRKY proteins. The benefit of the *Pichia* system is that recombinant proteins are secreted to the culture media. This overcame the problems highlight by Ciolkowski *et al.* (2008) where their recombinant WRKY proteins were often found in inclusion bodies, and they resulted to using soluble bacterial lysates for DNA binding studies rather than highly purified proteins.

Many of WRKYs proteins expressed using *E. coli* were found to be exclusively in inclusion bodies (Ciolkowski *et al.*, 2008b; Romero *et al.*, 2019). Purification of proteins from this source was found to impact W-box binding ability due to extraction through denaturation and subsequent renaturation steps. This could induce protein misfolding. Purification using affinity under mild conditions was also found problematic. Thus, many WRKYs were either expressed as full-length protein and total extracts of *E. coli* were used for subsequent experimentations or expression of truncated WRKY proteins containing only DNA binding domain. Either of these two types of expressions and purifications was found to be subjected to trial and error for WRKY proteins expression. Ciolkowski *et al.* (2008) reported that not all WRKY proteins expression would impact bacterial zinc homeostasis. However, their extraction from inclusion bodies and subsequent purification by affinity chromatography was problematic. Thus*, P. pastoris* was proposed in this project to be used as an alternative WRKY protein expression that could allow proteins to be easily recombinantly expressed in the medium without impacting their host cellular zinc homeostasis as well as enabling proteins to be purified using affinity purification with not impacting protein folding.

The benefit is using *Pichia* as a eukaryotic system instead of using *E. coli*, as a prokaryotic system. One of the main advantages is its capability of expressing proteins as its expression mechanism is close to other eukaryotic cells. Growth speed, post-translational modification, secretory expression, and easy genetic manipulation are other significant advantages of *Pichia* cells. It can be used for laboratory settings and industrial levels as it is cost-effective, and high yields of recombinant proteins can be produced with high similarity of protein, in terms of glycosylation, to other eukaryotic cells (Karbalaei;Rezaee and Farsiani, 2020).

One of the main concerns in *E. coli* was its folding capabilities of WRKY proteins. The benefit of *Pichia* is its ability to fold protein appropriately, which occurs in the endoplasmic reticulum. Proteins in *E. coli* tend to be expressed in inclusion bodies. With that, protein had to undergo a harsh extraction method that impacted protein folding. Pichia vectors are equipped with Kex2 as a signal peptide for secretion. With that, recombinant proteins can be expressed to the cell's external environment. Culture broth can be used to purify proteins using chromatography purification setting for laboratory scale. Regarding industrial scale, tangential flow filtration could be used as well. Thus, soluble *TaWRKY* proteins were successfully expressed and were purified using chromatography columns.

Many eukaryotic native proteins are glycosylated. Protein glycosylation effectively determines or drives the correct folding. This essentially affects the three-dimensional structures of proteins and plays a critical role in determining structure, function, and stability. In Pichia, one of its most common post-translational modifications is glycosylation. There are two main forms of protein glycosylation in *Pichia* cells, N-glycosylation and O-linked glycosylation. In co-translational modification events, N-linked oligosaccharides are translocated to the nascent protein. N-linked glycosylation is initiated at the luminal side of the endoplasmic reticulum (ER) membrane. It specifically recognizes asparagine residues Asn-X-Ser/Thr hydroxyl groups of nascent glycoproteins (Bretthauer, 2007).

The first assembly of the highly conserved oligosaccharide structures occurs at the cytosol (Li *et al.*, 2007). It consists of 14 sugar residues of sugars of three glucose, nine mannose and two *N*-acetylglucosamine ($Glc_3Man_9GlcNAc_2$). This core structure is then entered inside ER to transfer onto proteins. During the transportation of proteins along the

secretory pathway in the ER, three glucose residues are removed by glucosidase I and II, resulting in $Man_9GlcNAc_2$. Further trimming one mannose (α-1,2-mannose residue) by mannosidase I ($Man_8GlcNAc_2$) before transferring the glycoprotein into the Golgi apparatus where the elongation occurs of *N*-glycans take place. Once transferred, further modification of glycoproteins by the Och1P is mannosyltransferase of the cis-Golgi apparatus. This adds α-1,6-mannose residue, which forms the first block for further elongation of mannose by mannosyltransferase leading to hypermannosylation. It ranges between Man 8 to Man 14 (Vervecken *et al.*, 2007).

O-linked glycosylation also occurs in the serine and threonine of the protein starting at the ER. It is a short chain of oligosaccharides. In ER, O-mannosylation plays a vital role in folding and protein quality (Hart;Housley and Slawson, 2007; Janik;Lityńska and Vereecken, 2010; Rambaruth and Dwek, 2011). It had been found that the extent of o-linked-glycosylation in *Pichia pastoris* is much less than *Saccharomyces cerevisiae (Macauley-Patrick et al., 2005).* Mannose residues are transferred to Ser/Thr to form mannosyl- α-O-Ser/Thr. This o-mannosylation is initiated by the catalysis by O-mannosyl transferase to donate mannose residue from dolichylphsophoryl ß-D-mannosylpryanoside precursor, which is GDP-mannose. Further elongation reactions by the catalysis of other mannosyl transferases through the addition of mannose residues from GDP-mannose occur in the Golgi apparatus. This would eventually form unbranched oligosaccharides, which can get only up to five mannose residues (Bretthauer, 2007).

To generate recombinant and native proteins, Pichia pastoris are commonly utilized. In relation to glycosylation, it appears that *Pichia pastoris* tend to glycosylate recombinant protein differently to mammalian or plants as an original organism. The biosynthetic pathway of $Man_8GlcNAc_2$ is highly conserved in mammals and yeast. However, after the formation of $Man_8GlcNAc_2$ the biosynthesis pathway of N-glycosylation at Golgi apparatus diverges in both (mammals and yeast). Mannose residues are removed in mammalians whereas mannose residues are added in yeast and fungi. This addition of mannose by α-1,6-mannose residue leading to the formation of $Man_9GlcNAc_2$. In mammals, $Man_8GlcNAc_2$ is trimmed by Golgi mannosidase (α-1,2-mannosidase) to $Man_5GlcNAc_2$ structures. It is then elongated by GlcNAc transferase I to $GlcMan_5GlcNAc_2$. Mannosidase II (α-1,3/1,6-mannosidase) further process $GlcMan_5GlcNAc_2$ by trimming it into $GlcMan_3GlcNAc_2$. It is then elongated by GlcNAc transferase II to form $Glc_2Man_3GlcNAc_2$ which is further elongated by glycosyl transferase

to form a complex biantennary glycan which is a complex type of oligosaccharides (Vervecken *et al.*, 2007)

With the apparent differences between mammalian and yeast glycosylation. Indeed, *Pichia pastoris* had been popular for its ability to extensively produce foreign proteins at high levels and its simplicity to genetically manipulate with its ability to perform post-translational modifications. For therapeutic proteins to be used in humans, it may produce immunogenic protein in man. This is due to the high mannolysation in post-translational modification (Demain and Vaishnav, 2009). However, genetic engineering attempts aimed to humanize glycosylation biosynthesis pathway in *Pichia pastoris*.

Glycosylation of proteins is essentially crucial for WRKY proteins transcription factors. Previous studies attempted to express the full length of WRKY proteins. However, they were found to be non-functional (non-active). Glycosylation benefited the protein in terms of proper folding, stability, solubility, and proper biological activity (Cereghino *et al.*, 2002). This also includes the functionality of glycosylated WRKY. Thus, WRKY TFs (Transcription Factors) expressed by *Pichia pastoris* were found to be functional. This is based on their ability to bind to wheat *PR1* promoter sequence (Figure 5.4 A).

This demonstrate the first example of WRKY TFs being recombinantly expressed in *Pichia pastoris* as a eukaryotic system which resulted in expression of total soluble active proteins and purification. This improves the work of expression of WRKY TFs in a eukaryotic system than prokaryotic system.This work would provide additional information benefiting stakeholders such as researchers in this area. As a future work, native recombinant proteins could be employed for chromatin immune precipitation for gene identification which lead to elucidating a potential candidate gene for crop improvement. It can also be useful for protein structure which is for fundamental science on protein: protein interaction in the transcription complex.

6.2 Phosphate replacement provides a better media for protein production across a broad range of pH

For *Pichia pastoris* growth in a very high density, BSM was recommended as a sole medium during fermentation. It is a well-defined medium for cell growth and protein production. As a phosphorous source, it contains phosphoric acids. The medium pH is 1.5,

and 28% ammonium hydroxide must be used to adjust the working pH. The drawback of such a medium is that it precipitates magnesium and calcium phosphates at pH above 5.5. Even though the optimal pH for *Pichia* fermentation for recombinant protein production varies between 5.5 to 7.0. In addition, there are also much more adverse effects of precipitates formation during fermentation above 5.5. An occurrence of unbalanced nutrient supply or nutrient starvation. Due to its abrasive nature, cellular disruption leads to secreted products contamination with intracellular materials. The impact also extends to the fermentation unit. The gas sparger the fermentation unit could be clogged and also causing a deterioration in the mechanical seal and bearing. With regards to purification or downstream processing, extra steps could be considered. Thus, it is crucial to use a medium that overcomes the above-mentioned issues associated with pH above 5.5 and can also enable a good variability of pH range above 5.5 for recombinant protein production during fermentation. Zhang;Sinha and Meagher (2006) proposed M3 medium that does not result in precipitation at pH above 5.5 and it can have a tremendous positive impact to secreted proteins downstream processing. This was with the replacement of phosphates using glycerophosphates to serve as sole of phosphorus for Pichia growth and recombinant protein expression during fermentation. This provides a better media for protein production across a broad range of pH, neutral to high pH, as well as comparable level of protein expression level to BSM. A simplified purification method could be provided to purify high purity protein for research with this new approach.

It is crucial to develop a downstream processing strategy for an ideal downstream processing approach to purify protein from Pichia fermentation. It had been suggested that produced proteins should contain affinity tags which can aid protein detection and purification (Kimple;Brill and Pasker, 2013). Their existence on the protein cascade does not adversely affect its biological or biochemical activity (Gräslund *et al.*, 2008). The ideal approach for purifying fermented proteins could undergo simplified chromatography process. The first stage of purification, with the aid of affinity tags such as hexahistidine tag, is the use of Immobilized metal chelating affinity chromatography. It separates proteins according to their affinity to metal ions which can be carried out under undenaturing conditions. Thus, His-tagged secreted proteins could only be selectively trapped and eluted. However, truncated His-tagged proteins could also be eluted as well. Thus, a second stage of purification could be employed for high purification and this with the use of size exclusion chromatography. With this, proteins could be purified according

to their size. Fully secreted proteins could only be collected at the end of the purification. For concentrating eluted products, ultrafiltration centrifugation columns could be used.

Untagged secreted proteins could also be purified and processed as well. Pollet *et al.* (2021) expressed candidate SARS-CoV-2 RBD219-WT as SARS-CoV-2 receptor binding domain (RBD) protein using *Pichia pastoris*. SARS-CoV-2 RBD219-WT as His-tagged protein was purified were purified using three steps of purification. This involved buffer exchange using Pellicon 2 cassette and followed with IMAC (Immobilized Metal Affinity Chromatography) for His-tagged proteins capturing and then size exclusion chromatography for high purity of proteins. With the use of M3 media for development, this could reduce the purification process from 3 step purification into 2 step purification. This could be done by applying the culture broth directly to the IMAC and followed with size exclusion chromatography.

For high value of proteins, it is beneficial to use low cost, high production, and quality of proteins. M3 media could be used at this level as it showed, by Zhang;Sinha and Meagher (2006), a very comparable secretion level of protein and wide range of pH in M3 to BSM. The combination of high through output secretion of proteins with an effective strategy of protein purification could provide a low cost and high production and quality of proteins. With regards to industrial level, secreted proteins processing could be proceeded by the use of Tangential flow filtration as one step of purification, which separates proteins according to their size and molecular weight. Its advantages include scalability, time efficiency, reproducibility, and reusability of the separation membrane. This would make the secretion and downstream processing of secreted proteins more efficient and therefore cheaper.

6.3 Future perspectives:

This work is the first of its kind that expressed recombinant WRKY transcription factors proteins in a eukaryotic system with the use of *Pichia pastoris*. Our results had demonstrated that WRKY transcription factors can be expressed and purified easily using chromatography systems with no impact on proteins solubility and folding. This had resulted in a functional purified WRKY proteins which demonstrated their ability to bind

to the DNA sequence of promoter region of pathogenesis related gene. With that, this work could be place forward in the future to elucidate other target genes through ChIP assay.

What we currently know about wheat WRKY transcription factors is very little and mainly concerned, in literature, with osmotic stresses such as drought and salinity. Little understanding of their role in biotic stresses such as insect and fungal stresses (Satapathy *et al.*, 2014). Indeed, many WRKY transcription factors were thoroughly studied in other plant species such as *Arabidopsis thaliana* and rice. WRKY Orthologues from other species to wheat WRKY can provide a certain understanding of their role which can be used to expand our knowledge in wheat WRKY's in response to biotic and abiotic stresses. With the aid of bioinformatics, we identified candidate wheat WRKYs counterparts in other species which to provide us information on their role and/or function. As these candidate WRKY genes were responsive to biotic stresses. We acquired to understand their binding to DNA sequences at the promoter regions. Despite the general way of expression of WRKY's attempted previously for such interaction experiments using prokaryotic expression system. We found that a eukaryotic protein expression system was found could be a particularly useful tool to express WRKY proteins without drawbacks found in prokaryotic expression. This includes protein secretion as well as maintenance of native folding after purification with no exposure to harsh purification conditions. Additionally, their binding activity to W-box elements could be said more specific and selective compared to prokaryotic expressed proteins. At this point expressed WRKY protein could be used at chromatin immunoprecipitation which is a protein interaction with wheat genomic DNA. This will allow the protein to bind to their target promoter genes within the genome. Following that, bound proteins to target genes could be selectively purified and sequenced. Identified sequences could be then compared to genomic databases to determine their target genes. This could help us to understand their role and function in plant molecular machinery in response to stresses.

To elucidate plant tolerance/ resistance or susceptibility towards various types of biotic and stresses, overexpression of these wheat WRKYs could be conducted in model plants such as Arabidopsis thaliana or tobacco plants. This will allow further understanding of their functional activity in different plant aspects such as physiological, biochemical, and molecular. Many examples in literature had been conducted which identified their role in

plant stress response. *TaWRKY19*, as an example, was overexpressed and transformed into *Arabidopsis thaliana* by NIU *et al.* (2012). They had reported that *TaWRKY19* overexpression, in transgenic plants, confers tolerance to salt, drought, and freezing stresses. Biotic stresses could be b applied to such transgenic plants to look at their response in such stress conditions.

A gene knockdown/ knockout candidate WRKY genes in wheat could provide more data in response to biotic and abiotic stresses. Knockdown genes are essentially a suppression of genes by reducing their expression or silencing which can effectively alter the plant response which could be done by RNA interference (Han, 2018). Knockout gene is the removal of a specific gene, in other words, a target gene (Hall;Limaye and Kulkarni, 2009). Both mainly study the loss of function of an individual gene. This could affect the wheat plant positively or negatively toward stresses. WHIRLY1, as an example, is a plastid-nucleus-located protein that is described as a leaf senescence upstream regulator that binds to the promoter region of senescence-associated genes such as HcS40. Knockdown of WHIRLY1 was found to delay the expression of senescence-associated genes as well as drought stress-responsive genes (Janack *et al.*, 2016). Up-regulation of ethylene biosynthesis with downregulated *TaWRKY51* using RNAi whereas overexpression of *TaWRKY51* in *Arabidopsis* had shown a downregulation of ethylene biosynthesis (Hu *et al.*, 2018). In both experimentations with regards to knockdown/ knockout genes and overexpression, much of the work had been done on *Arabidopsis*. This is due to the simplicity of its genome in comparison to wheat ones (Ding *et al.*, 2014). CRISPR/Cas9 system could also be applied to introduce mutation to WRKY genes. Li;Li and Jiang (2021) generated a double mutation at *AtWRKY3* and *AtWRKY4* using CRISPR/Cas9 to study *Arabidopsis* response to salt and ME-JA stresses. Their findings demonstrated, with mutated WRKY's, a decrease of *Arabidopsis* tolerance toward salt and Me-JA stresses. All of which can be applied as a future work with the aim of being used in the crop protection. This is in either through conventional breeding, marker assisted breeding, and/or genetic selection. This could also be extended in plants genetic manipulation for the sake of crop protection.

With this type of protein recombinant expression, protein can be used to interact with genomic DNA from wheat plants. This is to elucidate their target promoter genes. For

that, a Chromatin immunoprecipitation assay could be used for that specific interaction. It is manly to capture proteins of interest at the state of their binding with DNA. Following that, DNA could be then released and analyzed. DNA sequencing could be used to identify which target gene promoter is present in that reaction. Protein to protein interaction is another potential method that could be used. This is to elucidate the interaction of WRKY proteins with other cellular proteins. Both methods could provide a larger understanding of WRKY protein machinery in the cell. This could be a very useful tool for plant breeders in crop protection through the use of marker assisted breeding.

7. References:

Agarwal, M., Hao, Y., Kapoor, A., Dong, C.-H., Fujii, H., Zheng, X. and Zhu, J.-K. (2006) 'A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance', *Journal of Biological Chemistry*, 281(49), pp. 37636-37645.

Agarwal, P., Dabi, M. and Agarwal, P.K. (2014) 'Molecular cloning and characterization of a group II WRKY transcription factor from Jatropha curcas, an important biofuel crop', *DNA and cell biology*, 33(8), pp. 503-513.

Ajanovic, A. (2011) 'Biofuels versus food production: Does biofuels production increase food prices?', *Energy*, 36(4), pp. 2070-2076.

Akimoto-Tomiyama, C., Sakata, K., Yazaki, J., Nakamura, K., Fujii, F., Shimbo, K., Yamamoto, K., Sasaki, T., Kishimoto, N. and Kikuchi, S. (2003) 'Rice gene expression in response to N-acetylchitooligosaccharide elicitor: comprehensive analysis by DNA microarray with randomly selected ESTs', *Plant molecular biology*, 52(3), pp. 537-551.

Alshegaihi, M.A.R. (2019) *Role of WRKY transcription factors in tolerance of winter wheat 'Triticum aestivum' to aphid 'Sitobion avenae' infestation under different nitrogen regimes*. Newcastle University [Online]. Available at: http://theses.ncl.ac.uk/jspui/handle/10443/4665.

Ambawat, S., Sharma, P., Yadav, N.R. and Yadav, R.C. (2013) 'MYB transcription factor genes as regulators for plant responses: an overview', *Physiology and Molecular Biology of Plants*, 19(3), pp. 307-321.

Anjanappa, R.B. and Gruissem, W. (2021) 'Current progress and challenges in crop genetic transformation', *Journal of Plant Physiology*, 261, p. 153411.

Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.-L., Gomez-Gomez, L., Boller, T., Ausubel, F.M. and Sheen, J. (2002) 'MAP kinase signalling cascade in Arabidopsis innate immunity', *Nature*, 415(6875), pp. 977-983.

Baba, S.H.A. (2019) *Molecular response of wheat to Bipolaris sorokiniana under nitrogen stress*. Newcastle University [Online]. Available at: http://theses.ncl.ac.uk/jspui/handle/10443/4623.

Banerjee, A. and Roychoudhury, A. (2015) 'WRKY proteins: signaling and regulation of expression during abiotic stress responses', *The Scientific World Journal*, 2015.

Begna, T. (2021) 'Conventional Breeding Methods Widely used to Improve Self-Pollinated Crops'.

Bigeard, J. and Hirt, H. (2018) 'Nuclear signaling of plant MAPKs', *Frontiers in plant science*, 9, p. 469.

Birkenbihl, R.P., Liu, S. and Somssich, I.E. (2017) 'Transcriptional events defining plant immune responses', *Current Opinion in Plant Biology*, 38, pp. 1-9.

Blanco, A. and Blanco, G. (2017) *Medical biochemistry*. Academic Press.

Bold, O., Jeevan, R.J., Lim, Y.P. and Vanjildorj, E. (2015) 'Differatation of wheat WRKY transcription factor TAWRKY10 gene expression in abiotic stress resistance', *Mongolian Journal of Agricultural Sciences*, 13(2), pp. 136-140.

Botha, A.-M., Swanevelder, Z.H. and Lapitan, N.L. (2010) 'Transcript profiling of wheat genes expressed during feeding by two different biotypes of Diuraphis noxia', *Environmental entomology*, 39(4), pp. 1206-1231.

Bretthauer, R.K. (2007) 'Characterization of O-Linked Saccharides on Glycoproteins', in Cregg, J.M. (ed.) *Pichia Protocols*. Totowa, NJ: Humana Press, pp. 107-118.

Burrell, M.M. (1993) *Enzymes of molecular biology*. Springer.

Cai, H., Yang, S., Yan, Y., Xiao, Z., Cheng, J., Wu, J., Qiu, A., Lai, Y., Mou, S. and Guan, D. (2015) 'CaWRKY6 transcriptionally activates CaWRKY40, regulates Ralstonia solanacearum resistance, and confers high-temperature and high-humidity tolerance in pepper', *Journal of Experimental Botany*, 66(11), pp. 3163-3174.

Cereghino, G.P.L., Cereghino, J.L., Ilgen, C. and Cregg, J.M. (2002) 'Production of recombinant proteins in fermenter cultures of the yeast Pichia pastoris', *Current opinion in biotechnology*, 13(4), pp. 329-332.

Chaudhary, R., Atamian, H.S., Shen, Z., Briggs, S.P. and Kaloshian, I. (2014) 'GroEL from the endosymbiont Buchnera aphidicola betrays the aphid by triggering plant defense', *Proceedings of the National Academy of Sciences*, 111(24), pp. 8919-8924.

Chen, L., Song, Y., Li, S., Zhang, L., Zou, C. and Yu, D. (2012) 'The role of WRKY transcription factors in plant abiotic stresses', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(2), pp. 120-128.

Cheng, X., Zhao, Y., Jiang, Q., Yang, J., Zhao, W., Taylor, I.A., Peng, Y.-L., Wang, D. and Liu, J. (2019) 'Structural basis of dimerization and dual W-box DNA recognition by rice WRKY domain', *Nucleic Acids Research*, 47(8), pp. 4308-4318.

Chujo, T., Sugioka, N., Masuda, Y., Shibuya, N., Takemura, T., Okada, K., Nojiri, H. and Yamane, H. (2009) 'Promoter analysis of the elicitor-induced WRKY gene OsWRKY53, which is involved in defense responses in rice', *Bioscience, biotechnology, and biochemistry*, 73(8), pp. 1901-1904.

Chujo, T., Takai, R., Akimoto-Tomiyama, C., Ando, S., Minami, E., Nagamura, Y., Kaku, H., Shibuya, N., Yasuda, M. and Nakashita, H. (2007) 'Involvement of the elicitor-induced gene OsWRKY53 in the expression of defense-related genes in rice', *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1769(7-8), pp. 497-505.

Ciolkowski, I., Wanke, D., Birkenbihl, R.P. and Somssich, I.E. (2008a) 'Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function', *Plant molecular biology*, 68(1), pp. 81-92.

Ciolkowski, I., Wanke, D., Birkenbihl, R.P. and Somssich, I.E. (2008b) 'Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function', *Plant molecular biology*, 68(1-2), pp. 81-92.

Corp., i. (2002) *Pichia Fermentation process Guidlines*. Available at: https://tools.thermofisher.com/content/sfs/manuals/pichiaferm_prot.pdf.

Couto, D. and Zipfel, C. (2016) 'Regulation of pattern recognition receptor signalling in plants', *Nature Reviews Immunology*, 16(9), pp. 537-552.

Cregg, J.M., Cereghino, J.L., Shi, J. and Higgins, D.R. (2000) 'Recombinant protein expression in Pichia pastoris', *Molecular biotechnology*, 16(1), pp. 23-52.

Curtis, B., Rajaram, S. and Gómez Macpherson, H. (2002) 'Wheat in the world'.

Curtis, T. and Halford, N.G. (2014) 'Food security: the challenge of increasing wheat yield and the importance of not compromising food safety', *Annals of Applied Biology*, 164(3), pp. 354-372.

Demain, A.L. and Vaishnav, P. (2009) 'Production of recombinant proteins by microbes and higher organisms', *Biotechnol Adv*, 27(3), pp. 297-306.

Denancé, N., Sánchez-Vallet, A., Goffner, D. and Molina, A. (2013) 'Disease resistance or growth: the role of plant hormones in balancing immune responses and fitness costs', *Frontiers in plant science*, 4, p. 155.

Ding, B., Wang, J., Song, N., Li, M., Cheng, Q., Huang, G., Guo, Y., Fu, Y., Xie, C. and Sun, Q. (2014) 'TaWRKY68 responses to biotic stresses are revealed by the orthologous genes from major cereals', *Genetics and Molecular Biology*, 37(1), pp. 73-80.

Ding, W., Fang, W., Shi, S., Zhao, Y., Li, X. and Xiao, K. (2016) 'Wheat WRKY Type Transcription Factor Gene TaWRKY1 is Essential in Mediating Drought Tolerance Associated with an ABA-Dependent Pathway', *Plant Molecular Biology Reporter*, 34(6), pp. 1111-1126.

Dixon, R.A., Achnine, L., Kota, P., Liu, C.J., Reddy, M.S. and Wang, L. (2002) 'The phenylpropanoid pathway and plant defence—a genomics perspective', *Molecular plant pathology*, 3(5), pp. 371-390.

Dodds, P.N. and Rathjen, J.P. (2010) 'Plant immunity: towards an integrated view of plant–pathogen interactions', *Nature Reviews Genetics*, 11(8), pp. 539-548.

Duan, M.-R., Nan, J., Liang, Y.-H., Mao, P., Lu, L., Li, L., Wei, C., Lai, L., Li, Y. and Su, X.-D. (2007) 'DNA binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana WRKY1 protein', *Nucleic acids research*, 35(4), pp. 1145-1154.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. and Lepiniec, L. (2010) 'MYB transcription factors in Arabidopsis', *Trends in plant science*, 15(10), pp. 573-581.

Edwards, S.G. and Jennings, P. (2018) 'Impact of agronomic factors on Fusarium mycotoxins in harvested wheat', *Food Additives & Contaminants: Part A*, 35(12), pp. 2443-2454.

Eulgem, T., Rushton, P.J., Robatzek, S. and Somssich, I.E. (2000) 'The WRKY superfamily of plant transcription factors', *Trends in plant science*, 5(5), pp. 199-206.

FAO (1996) *Rome Declaration on World Food Security and World Food Summit Plan of Action. World Food Summit*

FAO (2022). 17/02/2022. Food and Agricultural Organisation of the United Nations. Available at: https://www.fao.org/faostat/en/#home.

FAO, F. (2017) 'The future of food and agriculture–Trends and challenges', *Annual Report*, 296, pp. 1-180.

FAO, I. and Unicef (2020) 'WFP, WHO (2020) The state of food security and nutrition in the world 2020', *Transforming food systems for affordable healthy diets. FAO, Rome*, 10.

FAOSTAT (2022) 'Food and agriculture data'. 21/02/2022. Available at: https://www.fao.org/faostat/en/#home.

Fleming, R. (2020) 'ADC analysis by hydrophobic interaction chromatography', in *Antibody-Drug Conjugates*. Springer, pp. 147-161.

Foyer, C.H., Rasool, B., Davey, J.W. and Hancock, R.D. (2016) 'Cross-tolerance to biotic and abiotic stresses in plants: a focus on resistance to aphid infestation', *Journal of Experimental Botany*, 67(7), pp. 2025-2037.

Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P. and Solano, R. (2014) 'DNA-binding specificities of plant transcription factors and their potential to define target genes', *Proceedings of the National Academy of Sciences*, 111(6), pp. 2367-2372.

Freeborough, W., Gentle, N. and Rey, M.E. (2021) 'WRKY Transcription Factors in Cassava Contribute to Regulation of Tolerance and Susceptibility to Cassava Mosaic Disease through Stress Responses', *Viruses*, 13(9), p. 1820.

Gao, H., Wang, Y., Xu, P. and Zhang, Z. (2018) 'Overexpression of a WRKY transcription factor TaWRKY2 enhances drought stress tolerance in transgenic wheat', *Frontiers in plant science*, 9, p. 997.

Gao, S.-Q., Chen, M., Xu, Z.-S., Zhao, C.-P., Li, L., Xu, H.-j., Tang, Y.-m., Zhao, X. and Ma, Y.-Z. (2011) 'The soybean GmbZIP1 transcription factor enhances multiple abiotic stress tolerances in transgenic plants', *Plant molecular biology*, 75(6), pp. 537-553.

Gellissen, G., Janowicz, Z.A., Weydemann, U., Melber, K., Strasser, A.W. and Hollenberg, C.P. (1992) 'High-level expression of foreign genes in Hansenula polymorpha', *Biotechnology advances*, 10(2), pp. 179-189.

Georges, F. and Ray, H. (2017) 'Genome editing of crops: a renewed opportunity for food security', *GM Crops & Food*, 8(1), pp. 1-12.

Girón-Calva, P.S., Twyman, R.M., Albajes, R., Gatehouse, A.M. and Christou, P. (2020) 'The impact of environmental stress on Bt crop performance', *Trends in Plant Science*, 25(3), pp. 264-278.

Glazebrook, J. (2005) 'Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens', *Annu. Rev. Phytopathol.*, 43, pp. 205-227.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M. and Toulmin, C. (2010) 'Food Security: The Challenge of Feeding 9 Billion People', *Science*, 327(5967), pp. 812-818.

Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C. and Hui, R. (2008) 'Protein production and purification', *Nature methods*, 5(2), pp. 135-146.

Gu, L., Dou, L., Guo, Y., Wang, H., Li, L., Wang, C., Ma, L., Wei, H. and Yu, S. (2019) 'The WRKY transcription factor GhWRKY27 coordinates the senescence regulatory pathway in upland cotton (Gossypium hirsutum L.)', *BMC plant biology*, 19(1), p. 116.

Hall, B., Limaye, A. and Kulkarni, A.B. (2009) 'Overview: generation of gene knockout mice', *Current protocols in cell biology*, 44(1), pp. 19.12. 1-19.12. 17.

Han, H. (2018) 'RNA interference to knock down gene expression', *Disease Gene Identification*, pp. 293-302.

Han, Y., Fan, T., Zhu, X., Wu, X., Ouyang, J., Jiang, L. and Cao, S. (2019) 'WRKY12 represses GSH1 expression to negatively regulate cadmium tolerance in Arabidopsis', *Plant molecular biology*, 99(1-2), pp. 149-159.

Hao, R., Adoligbe, C., Jiang, B., Zhao, X., Gui, L., Qu, K., Wu, S. and Zan, L. (2015) 'An optimized trichloroacetic acid/acetone precipitation method for two-dimensional gel electrophoresis analysis of Qinchuan cattle longissimus dorsi muscle containing high proportion of marbling', *PloS one*, 10(4), p. e0124723.

Hart, G.W., Housley, M.P. and Slawson, C. (2007) 'Cycling of O-linked ß-N-acetylglucosamine on nucleocytoplasmic proteins', *Nature*, 446(7139), pp. 1017-1022.

Hassan, S., Lethin, J., Blomberg, R., Mousavi, H. and Aronsson, H. (2019) 'In silico based screening of WRKY genes for identifying functional genes regulated by WRKY under salt stress', *Computational biology and chemistry*, 83, p. 107131.

Hay, R.T. (2005) 'SUMO: a history of modification', *Molecular cell*, 18(1), pp. 1-12.

He, G.-H., Xu, J.-Y., Wang, Y.-X., Liu, J.-M., Li, P.-S., Chen, M., Ma, Y.-Z. and Xu, Z.-S. (2016) 'Drought-responsive WRKY transcription factor genes TaWRKY1 and TaWRKY33 from wheat confer drought and/or heat resistance in Arabidopsis', *BMC plant biology*, 16(1), pp. 1-16.

Hernandez-Garcia, C.M. and Finer, J.J. (2014) 'Identification and validation of promoters and cis-acting regulatory elements', *Plant Science*, 217, pp. 109-119.

Higgins, D.R. and Cregg, J.M. (1998) 'Introduction to Pichia pastoris', in *Pichia protocols*. Springer, pp. 1-15.

Holden, N.S. and Tacon, C.E. (2011) 'Principles and problems of the electrophoretic mobility shift assay', *Journal of pharmacological and toxicological methods*, 63(1), pp. 7-14.

Hollmann, J., Gregersen, P.L. and Krupinska, K. (2014) 'Identification of predominant genes involved in regulation and execution of senescence-associated nitrogen remobilization in flag leaves of field grown barley', *Journal of Experimental Botany*, 65(14), pp. 3963-3973.

Holmgren, A. (1989) 'Thioredoxin and glutaredoxin systems', *Journal of Biological Chemistry*, 264(24), pp. 13963-13966.

Hu, Z., Wang, R., Zheng, M., Liu, X., Meng, F., Wu, H., Yao, Y., Xin, M., Peng, H. and Ni, Z. (2018) 'Ta WRKY 51 promotes lateral root formation through negative regulation of ethylene biosynthesis in wheat (Triticum aestivum L.)', *The Plant Journal*, 96(2), pp. 372-388.

Huang, J., Pray, C. and Rozelle, S. (2002) 'Enhancing the crops to feed the poor', *Nature*, 418(6898), pp. 678-684.

Huang, J., Yang, Z., Guan, F.-f., Zhang, S.-s., Cui, D., Guan, G.-h. and Li, Y. (2013) 'A novel mono- and diacylglycerol lipase highly expressed in Pichia pastoris and its application for food emulsifier preparation', *Process Biochemistry*, 48, pp. 1899-1904.

Hussain, R.M., Sheikh, A.H., Haider, I., Quareshy, M. and Linthorst, H.J. (2018) 'Arabidopsis WRKY50 and TGA transcription factors synergistically activate expression of PR1', *Frontiers in plant science*, 9, p. 930.

ISAAA.Brief.55 (2019) *Global Status of Commercialized Biotech/GM Crops in 2019*.International Service for the Aquisition of Agri-biotech Applications (ISAAA). [Online]. Available at: http://www.isaaa.org.

Jacobs, P.P., Geysens, S., Vervecken, W., Contreras, R. and Callewaert, N. (2009) 'Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology', *Nature protocols*, 4(1), pp. 58-70.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T. and Parcy, F. (2002) 'bZIP transcription factors in Arabidopsis', *Trends in plant science*, 7(3), pp. 106-111.

Janack, B., Sosoi, P., Krupinska, K. and Humbeck, K. (2016) 'Knockdown of WHIRLY1 affects drought stress-induced leaf senescence and histone modifications of the senescence-associated gene HvS40', *Plants*, 5(3), p. 37.

Janik, M.E., Lityńska, A. and Vereecken, P. (2010) 'Cell migration—the role of integrin glycosylation', *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1800(6), pp. 545-555.

Jensen, M.K., Kjaersgaard, T., Nielsen, M.M., Galberg, P., Petersen, K., O'shea, C. and Skriver, K. (2010) 'The Arabidopsis thaliana NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling', *Biochemical Journal*, 426(2), pp. 183-196.

Jia, Z., Li, M., Wang, H., Zhu, B., Gu, L., Du, X. and Ren, M. (2021) 'TaWRKY70 positively regulates TaCAT5 enhanced Cd tolerance in transgenic Arabidopsis', *Environmental and Experimental Botany*, 190, p. 104591.

Jiang, G.-L. (2015) 'Molecular marker-assisted breeding: a plant breeder's review', in *Advances in plant breeding strategies: breeding, biotechnology and molecular tools*. Springer, pp. 431-472.

Jun, X., WANG, X.-y. and GUO, W.-z. (2015) 'The cytochrome P450 superfamily: Key players in plant development and defense', *Journal of Integrative Agriculture*, 14(9), pp. 1673-1686.

Karbalaei, M., Rezaee, S.A. and Farsiani, H. (2020) 'Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins', *Journal of cellular physiology*, 235(9), pp. 5867-5881.

Kaur, A., Pati, P.K., Pati, A.M. and Nagpal, A.K. (2017) 'In-silico analysis of cis-acting regulatory elements of pathogenesis-related proteins of Arabidopsis thaliana and Oryza sativa', *PloS one*, 12(9), p. e0184523.

Kihara, T., Zhao, C.-R., Kobayashi, Y., Takita, E., Kawazu, T. and Koyama, H. (2006) 'Simple identification of transgenic Arabidopsis plants carrying a single copy of the integrated gene', *Bioscience, biotechnology, and biochemistry*, 70(7), pp. 1780-1783.

Kim, J.-S., Mizoi, J., Yoshida, T., Fujita, Y., Nakajima, J., Ohori, T., Todaka, D., Nakashima, K., Hirayama, T. and Shinozaki, K. (2011) 'An ABRE promoter sequence is involved in osmotic stress-responsive expression of the DREB2A gene, which encodes a transcription factor regulating drought-inducible genes in Arabidopsis', *Plant and Cell Physiology*, 52(12), pp. 2136-2146.

Kimple, M.E., Brill, A.L. and Pasker, R.L. (2013) 'Overview of affinity tags for protein purification', *Current protocols in protein science*, 73(1), pp. 9.9. 1-9.9. 23.

Klümper, W. and Qaim, M. (2014) 'A meta-analysis of the impacts of genetically modified crops', *PloS one*, 9(11), p. e111629.

Kombrink, E. and Somssich, I. (1997) 'Pathogenesis-related proteins and plant defense', in *Plant relationships*. Springer, pp. 107-128.

Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R.H., Wang, X. and Borrill, P. (2017) 'Uncovering hidden variation in polyploid wheat', *Proceedings of the National Academy of Sciences*, 114(6), pp. E913-E921.

Kuki, Y., Ohno, R., Yoshida, K. and Takumi, S. (2020) 'Heterologous expression of wheat WRKY transcription factor genes transcriptionally activated in hybrid necrosis strains alters abiotic and biotic stress tolerance in transgenic Arabidopsis', *Plant Physiology and Biochemistry*, 150, pp. 71-79.

Lai, Z., Vinod, K., Zheng, Z., Fan, B. and Chen, Z. (2008) 'Roles of Arabidopsis WRKY3 and WRKY4 transcription factors in plant responses to pathogens', *BMC plant biology*, 8(1), pp. 1-13.

Lambertz, C., Garvey, M., Klinger, J., Heesel, D., Klose, H., Fischer, R. and Commandeur, U. (2014) 'Challenges and advances in the heterologous expression of cellulolytic enzymes: a review', *Biotechnology for biofuels*, 7(1), p. 135.

Lee, C.-H. (2017) 'A simple outline of methods for protein isolation and purification', *Endocrinology and metabolism*, 32(1), pp. 18-22.

Lee, J., Liu, Z., Chen, W.-H., Wei, J., Kundu, R., Adhikari, R., Rivera, J.A., Gillespie, P.M., Strych, U. and Zhan, B. (2021) 'Process development and scale-up optimization of the SARS-CoV-2 receptor binding domain–based vaccine candidate, RBD219-N1C1', *Applied microbiology and biotechnology*, 105(10), pp. 4153-4165.

Lee, J., Orosa, B., Millyard, L., Edwards, M., Kanyuka, K., Gatehouse, A., Rudd, J., Hammond-Kosack, K., Pain, N. and Sadanandom, A. (2015) 'Functional analysis of a W heat H omeodomain protein, TaR1, reveals that host chromatin remodelling influences the dynamics of the switch to necrotrophic growth in the phytopathogenic fungus Z ymoseptoria tritici', *New Phytologist*, 206(2), pp. 598-605.

Li, P., Li, X. and Jiang, M. (2021) 'CRISPR/Cas9-mediated mutagenesis of WRKY3 and WRKY4 function decreases salt and Me-JA stress tolerance in Arabidopsis thaliana', *Molecular Biology Reports*, 48(8), pp. 5821-5832.

Li, W., Pang, S., Lu, Z. and Jin, B. (2020a) 'Function and mechanism of WRKY transcription factors in abiotic stress responses of plants', *Plants*, 9(11), p. 1515.

LI, X.-j., GUO, C.-j., LU, W.-j., DUAN, W.-w., Miao, Z., GU, J.-t. and Kai, X. (2014) 'Expression pattern analysis of zinc finger protein genes in wheat (Triticum aestivum L.) under phosphorus deprivation', *Journal of integrative agriculture*, 13(8), pp. 1621-1633.

Li, X., Tang, Y., Zhou, C., Zhang, L. and Lv, J. (2020b) 'A wheat WRKY transcription factor TaWRKY46 enhances tolerance to osmotic stress in transgenic Arabidopsis plants', *International journal of molecular sciences*, 21(4), p. 1321.

Li, Z., Xiong, F., Lin, Q., d'Anjou, M., Daugulis, A.J., Yang, D.S. and Hew, C.L. (2001) 'Low-temperature increases the yield of biologically active herring antifreeze protein in Pichia pastoris', *Protein Expression and Purification*, 21(3), pp. 438-445.

Liao, Y., Zou, H.-F., Wang, H.-W., Zhang, W.-K., Ma, B., Zhang, J.-S. and Chen, S.-Y. (2008) 'Soybean GmMYB76, GmMYB92, and GmMYB177 genes confer stress tolerance in transgenic Arabidopsis plants', *Cell research*, 18(10), pp. 1047-1060.

Lippok, B., Birkenbihl, R.P., Rivory, G., Brümmer, J., Schmelzer, E., Logemann, E. and Somssich, I.E. (2007) 'Expression of AtWRKY33 encoding a pathogen-or PAMP-responsive WRKY transcription factor is regulated by a composite DNA motif containing W box elements', *Molecular Plant-Microbe Interactions*, 20(4), pp. 420-429.

Lis, M. and Walther, D. (2016) 'The orientation of transcription factor binding site motifs in gene promoter regions: does it matter?', *BMC genomics*, 17(1), pp. 1-21.

Liu, C., Mao, B., Ou, S., Wang, W., Liu, L., Wu, Y., Chu, C. and Wang, X. (2014) 'OsbZIP71, a bZIP transcription factor, confers salinity and drought tolerance in rice', *Plant molecular biology*, 84(1), pp. 19-36.

Liu, J.-H., Peng, T. and Dai, W. (2014) 'Critical cis-acting elements and interacting transcription factors: key players associated with abiotic stress responses in plants', *Plant molecular biology reporter*, 32(2), pp. 303-317.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1998) 'Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought-and low-temperature-responsive gene expression, respectively, in Arabidopsis', *The Plant Cell*, 10(8), pp. 1391-1406.

Lu, M., Ying, S., Zhang, D.-F., Shi, Y.-S., Song, Y.-C., Wang, T.-Y. and Li, Y. (2012) 'A maize stress-responsive NAC transcription factor, ZmSNAC1, confers enhanced tolerance to dehydration in transgenic Arabidopsis', *Plant cell reports*, 31(9), pp. 1701-1711.

Lu, Z., Dang, C., Wang, F., Liu, Z., Chen, J., Wang, Y., Yao, H., Fang, Q., Peng, Y. and Gatehouse, A.M. (2020) 'Does long-term Bt rice planting pose risks to spider communities and their capacity to control planthoppers?', *Plant Biotechnology Journal*, 18(9), p. 1851.

Lucas, S., Durmaz, E., Akpınar, B.A. and Budak, H. (2011) 'The drought response displayed by a DRE-binding protein from Triticum dicoccoides', *Plant Physiology and Biochemistry*, 49(3), pp. 346-351.

Ma, D., Pu, G., Lei, C., Ma, L., Wang, H., Guo, Y., Chen, J., Du, Z., Wang, H. and Li, G. (2009) 'Isolation and characterization of AaWRKY1, an Artemisia annua transcription factor that regulates the amorpha-4, 11-diene synthase gene, a key gene of artemisinin biosynthesis', *Plant and cell physiology*, 50(12), pp. 2146-2161.

Macauley-Patrick, S., Fazenda, M.L., McNeil, B. and Harvey, L.M. (2005) 'Heterologous protein production using the Pichia pastoris expression system', *yeast*, 22(4), pp. 249-270.

Maeo, K., Hayashi, S., Kojima-Suzuki, H., Morikami, A. and Nakamura, K. (2001) 'Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins', *Bioscience, biotechnology, and biochemistry*, 65(11), pp. 2428-2436.

Mao, G., Meng, X., Liu, Y., Zheng, Z., Chen, Z. and Zhang, S. (2011) 'Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis', *The Plant Cell*, 23(4), pp. 1639-1653.

Marianayagam, N.J., Sunde, M. and Matthews, J.M. (2004) 'The power of two: protein dimerization in biology', *Trends in biochemical sciences*, 29(11), pp. 618-625.

Martínez-Alarcón, D., Blanco-Labra, A. and García-Gasca, T. (2018) 'Expression of Lectins in Heterologous Systems', *Int J Mol Sci*, 19(2).

Mauch-Mani, B. and Slusarenko, A.J. (1996) 'Production of salicylic acid precursors is a major function of phenylalanine ammonia-lyase in the resistance of Arabidopsis to Peronospora parasitica', *The Plant Cell*, 8(2), pp. 203-212.

McCue, J.T. (2009) 'Theory and use of hydrophobic interaction chromatography in protein purification applications', *Methods in enzymology*, 463, pp. 405-414.

McDowell, J.M. and Dangl, J.L. (2000) 'Signal transduction in the plant immune response', *Trends in biochemical sciences*, 25(2), pp. 79-82.

McKenzie, F.C. and Williams, J. (2015) 'Sustainable food production: constraints, challenges and choices by 2050', *Food Security*, 7(2), pp. 221-233.

Meng, L., Wong, J.H., Feldman, L.J., Lemaux, P.G. and Buchanan, B.B. (2010) 'A membrane-associated thioredoxin required for plant growth moves from cell to cell, suggestive of a role in intercellular communication', *Proceedings of the National Academy of Sciences*, 107(8), pp. 3900-3905.

Mobley, A.S. (2019) 'Chapter 4 - Induced Pluripotent Stem Cells', in Mobley, A.S. (ed.) *Neural Stem Cells and Adult Neurogenesis*. Academic Press, pp. 67-94.

Ning, P., Liu, C., Kang, J. and Lv, J. (2017) 'Genome-wide analysis of WRKY transcription factors in wheat (Triticum aestivum L.) and differential expression under water deficit condition', *PeerJ*, 5, p. e3232.

NIU, C.F., Wei, W., ZHOU, Q.Y., TIAN, A.G., HAO, Y.J., ZHANG, W.K., Ma, B., Lin, Q., ZHANG, Z.B. and ZHANG, J.S. (2012) 'Wheat WRKY genes TaWRKY2 and TaWRKY19 regulate abiotic stress tolerance in transgenic Arabidopsis plants', *Plant, cell & environment*, 35(6), pp. 1156-1170.

Okay, S., Derelli, E. and Unver, T. (2014) 'Transcriptome-wide identification of bread wheat WRKY transcription factors in response to drought stress', *Molecular Genetics and Genomics*, 289(5), pp. 765-781.

Oliveira, M.R.V., Henneberry, T.J. and Anderson, P. (2001) 'History, current status, and collaborative research projects for Bemisia tabaci', *Crop Protection*, 20(9), pp. 709-723.

Olsen, A.N., Ernst, H.A., Leggio, L.L. and Skriver, K. (2005) 'NAC transcription factors: structurally distinct, functionally diverse', *Trends in plant science*, 10(2), pp. 79-87.

Pandey, B., Grover, A. and Sharma, P. (2018) 'Molecular dynamics simulations revealed structural differences among WRKY domain-DNA interaction in barley (Hordeum vulgare)', *BMC genomics*, 19(1), pp. 1-16.

Patnaik, D. and Khurana, P. (2001) 'Wheat biotechnology: A minireview', *Electronic Journal of Biotechnology*, 4, pp. 7-8.

Phukan, U.J., Jeena, G.S. and Shukla, R.K. (2016) 'WRKY transcription factors: molecular regulation and stress responses in plants', *Frontiers in plant science*, 7.

Pinstrup-Andersen, P. (2009) 'Food security: definition and measurement', *Food Security*, 1(1), pp. 5-7.

Poll, A. (2017) *Role of WRKY Transcription Factors in Coordinating Nitrogen Stress Responses in Wheat: Implications for Septoria tolerance*. Newcastle University.

Pollet, J., Chen, W.-H., Versteeg, L., Keegan, B., Zhan, B., Wei, J., Liu, Z., Lee, J., Kundu, R. and Adhikari, R. (2021) 'SARS-CoV-2 RBD219-N1C1: A yeast-expressed SARS-CoV-2 recombinant receptor-binding domain candidate vaccine stimulates virus neutralizing antibodies and T-cell immunity in mice', *Human vaccines & immunotherapeutics*, 17(8), pp. 2356-2366.

Potvin, G., Ahmad, A. and Zhang, Z. (2012) 'Bioprocess engineering aspects of heterologous protein production in Pichia pastoris: a review', *Biochemical Engineering Journal*, 64, pp. 91-105.

Qaim, M. (2016) *Genetically modified crops and agricultural development*. Springer.

Qaim, M. (2020) 'Role of new plant breeding technologies for food security and sustainable agricultural development', *Applied Economic Perspectives and Policy*, 42(2), pp. 129-150.

Qin, Y., Tian, Y., Han, L. and Yang, X. (2013) 'Constitutive expression of a salinity-induced wheat WRKY transcription factor enhances salinity and ionic stress tolerance in transgenic Arabidopsis thaliana', *Biochemical and Biophysical Research Communications*, 441(2), pp. 476-481.

Qin, Y., Tian, Y. and Liu, X. (2015) 'A wheat salinity-induced WRKY transcription factor TaWRKY93 confers multiple abiotic stress tolerance in Arabidopsis thaliana', *Biochemical and Biophysical Research Communications*, 464(2), pp. 428-433.

Queiroz, J., Tomaz, C. and Cabral, J. (2001) 'Hydrophobic interaction chromatography of proteins', *Journal of biotechnology*, 87(2), pp. 143-159.

Ramamoorthy, R., Jiang, S.-Y., Kumar, N., Venkatesh, P.N. and Ramachandran, S. (2008) 'A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments', *Plant and cell physiology*, 49(6), pp. 865-879.

Ramazi, S. and Zahiri, J. (2021) 'Post-translational modifications in proteins: resources, tools and prediction methods', *Database*, 2021.

Rambaruth, N.D. and Dwek, M.V. (2011) 'Cell surface glycan–lectin interactions in tumor metastasis', *Acta histochemica*, 113(6), pp. 591-600.

Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A. (2013) 'Yield Trends Are Insufficient to Double Global Crop Production by 2050', *PLOS ONE*, 8(6), p. e66428.

Riechmann, J.L. and Meyerowitz, E.M. (1998) 'The AP2/EREBP family of plant transcription factors', *Biological chemistry*, 379, pp. 633-646.

Romero, I., Alegria-Carrasco, E., de Pradena, A.G., Vazquez-Hernandez, M., Escribano, M.I., Merodio, C. and Sanchez-Ballesta, M.T. (2019) 'WRKY transcription factors in the response of table grapes (cv. Autumn Royal) to high CO2 levels and low temperature', *Postharvest biology and technology*, 150, pp. 42-51.

Roy, D. and Sadanandom, A. (2021) 'SUMO mediated regulation of transcription factors as a mechanism for transducing environmental cues into cellular signaling in plants', *Cellular and Molecular Life Sciences*, 78(6), pp. 2641-2664.

Rushton, D.L., Tripathi, P., Rabara, R.C., Lin, J., Ringler, P., Boken, A.K., Langum, T.J., Smidt, L., Boomsma, D.D. and Emme, N.J. (2012) 'WRKY transcription factors: key components in abscisic acid signalling', *Plant biotechnology journal*, 10(1), pp. 2-11.

Rushton, P.J., Reinstadler, A., Lipka, V., Lippok, B. and Somssich, I.E. (2002) 'Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen-and wound-induced signaling', *The Plant Cell*, 14(4), pp. 749-762.

Rushton, P.J., Somssich, I.E., Ringler, P. and Shen, Q.J. (2010) 'WRKY transcription factors', *Trends in Plant Science*, 15(5), pp. 247-258.

Rushton, P.J., Torres, J.T., Parniske, M., Wernert, P., Hahlbrock, K. and Somssich, I. (1996) 'Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes', *The EMBO journal*, 15(20), pp. 5690-5700.

Saijo, Y. and Loo, E.P.i. (2020) 'Plant immunity in signal integration between biotic and abiotic stress responses', *New Phytologist*, 225(1), pp. 87-104.

Satapathy, L., Kumar, D. and Mukhopadhyay, K. (2017) 'WRKY transcription factors: Involvement in plant–pathogen interactions', in *Recent Advances in Applied Microbiology*. Springer, pp. 229-246.

Satapathy, L., Singh, D., Ranjan, P., Kumar, D., Kumar, M., Prabhu, K.V. and Mukhopadhyay, K. (2014) 'Transcriptome-wide analysis of WRKY transcription factors in wheat and their leaf rust responsive expression profiling', *Molecular genetics and genomics*, 289(6), pp. 1289-1306.

Schmidhuber, J. and Tubiello, F.N. (2007) 'Global food security under climate change', *Proceedings of the National Academy of Sciences*, 104(50), pp. 19703-19708.

SDG2, U. (2021) *The Sustainable Development Goals report* [Online]. Available at: https://unstats.un.org/sdgs/report/2021/.

Sheikh, A.H., Eschen-Lippold, L., Pecher, P., Hoehenwarter, W., Sinha, A.K., Scheel, D. and Lee, J. (2016) 'Regulation of WRKY46 transcription factor function by mitogen-activated protein kinases in Arabidopsis thaliana', *Frontiers in plant science*, 7, p. 61.

Shen, H., Liu, C., Zhang, Y., Meng, X., Zhou, X., Chu, C. and Wang, X. (2012) 'OsWRKY30 is activated by MAP kinases to confer drought tolerance in rice', *Plant molecular biology*, 80(3), pp. 241-253.

Shewry, P.R. (2009) 'Wheat', *Journal of Experimental Botany*, 60(6), pp. 1537-1553.

Shewry, P.R. and Hey, S.J. (2015) 'The contribution of wheat to human diet and health', *Food and energy security*, 4(3), pp. 178-202.

Shim, J.S., Jung, C., Lee, S., Min, K., Lee, Y.W., Choi, Y., Lee, J.S., Song, J.T., Kim, J.K. and Choi, Y.D. (2013) 'A t MYB 44 regulates WRKY 70 expression and modulates antagonistic interaction between salicylic acid and jasmonic acid signaling', *The Plant Journal*, 73(3), pp. 483-495.

Smith, C.M., Liu, X., Wang, L.J., Liu, X., Chen, M.-S., Starkey, S. and Bai, J. (2010) 'Aphid feeding activates expression of a transcriptome of oxylipin-based defense signals in wheat involved in resistance to herbivory', *Journal of chemical ecology*, 36(3), pp. 260-276.

Smith, M.F. and Delbary-Gossart, S. (2001) 'Electrophoretic mobility shift assay (EMSA)', in *Colorectal Cancer*. Springer, pp. 249-257.

Spence, N., Hill, L. and Morris, J. (2020) 'How the global threat of pests and diseases impacts plants, people, and the planet', *Plants, People, Planet*, 2(1), pp. 5-13.

Spitz, F. and Furlong, E.E. (2012) 'Transcription factors: from enhancer binding to developmental control', *Nature reviews genetics*, 13(9), pp. 613-626.

Steen, J., Kozlov, M., Lutz, H., Parrella, J., Goodrich, E.M. and Raghunath, B. (2019) 'Single pass tangential flow filtration systems and tangential flow filtration systems with recirculation of retentate'. Google Patents.

Stockinger, E.J., Gilmour, S.J. and Thomashow, M.F. (1997) 'Arabidopsis thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit', *Proceedings of the National Academy of Sciences*, 94(3), pp. 1035-1040.

Suzuki, N., Rivero, R.M., Shulaev, V., Blumwald, E. and Mittler, R. (2014) 'Abiotic and biotic stress combinations', *New Phytologist*, 203(1), pp. 32-43.

TAIR, G.A.-. (2010) 'Genome Assembly' 2010, April 5. 2022 Feburary 3. Available at: https://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/agicomplete.jsp#:~:text=The%20Arabidopsis%20thaliana%20genome%20was,size%20of%20approximately%20135%2Dmegabases.

Tomei, J. and Helliwell, R. (2016) 'Food versus fuel? Going beyond biofuels', *Land Use Policy*, 56(Supplement C), pp. 320-326.

Tomlinson, I. (2013) 'Doubling food production to feed the 9 billion: A critical perspective on a key discourse of food security in the UK', *Journal of Rural Studies*, 29(Supplement C), pp. 81-90.

Tóth, G., Hermann, T., da Silva, M.R. and Montanarella, L. (2018) 'Monitoring soil for sustainable development and land degradation neutrality', *Environmental Monitoring and Assessment*, 190(2), pp. 1-4.

Tran, L.-S.P., Nishiyama, R., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2010) 'Potential utilization of NAC transcription factors to enhance abiotic stress tolerance in plants by biotechnological approach', *GM crops*, 1(1), pp. 32-39.

Tscharntke, T., Clough, Y., Wanger, T.C., Jackson, L., Motzke, I., Perfecto, I., Vandermeer, J. and Whitbread, A. (2012) 'Global food security, biodiversity conservation and the future of agricultural intensification', *Biological Conservation*, 151(1), pp. 53-59.

Tsumoto, K., Ejima, D., Senczuk, A.M., Kita, Y. and Arakawa, T. (2007) 'Effects of salts on protein–surface interactions: applications for column chromatography', *Journal of pharmaceutical sciences*, 96(7), pp. 1677-1690.

Tufan, F., Uçarlı, C., Tunalı, B. and Gürel, F. (2017) 'Analysis of early events in barley (Hordeum vulgare L.) roots in response to Fusarium culmorum infection', *European Journal of Plant Pathology*, 148(2), pp. 343-355.

Uçarlı, C. and Gürel, F. (2020) 'Differential physiological and molecular responses of three-leaf stage barley (Hordeum vulgare L.) under salt stress within hours', *Plant Biotechnology Reports*, 14(1), pp. 89-97.

Ülker, B. and Somssich, I.E. (2004) 'WRKY transcription factors: from DNA binding towards biological function', *Current opinion in plant biology*, 7(5), pp. 491-498.

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000) 'Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions', *Proceedings of the National Academy of Sciences*, 97(21), pp. 11632-11637.

Van Eck, L., Davidson, R.M., Wu, S., Zhao, B.Y., Botha, A.-M., Leach, J.E. and Lapitan, N.L. (2014) 'The transcriptional network of WRKY53 in cereals links oxidative responses to biotic and abiotic stress inputs', *Functional & integrative genomics*, 14(2), pp. 351-362.

Van Eck, L., Schultz, T., Leach, J.E., Scofield, S.R., Peairs, F.B., Botha, A.M. and Lapitan, N.L. (2010) 'Virus-induced gene silencing of WRKY53 and an inducible phenylalanine ammonia-lyase in wheat reduces aphid resistance', *Plant biotechnology journal*, 8(9), pp. 1023-1032.

van Verk, M.C., Pappaioannou, D., Neeleman, L., Bol, J.F. and Linthorst, H.J. (2008) 'A novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors', *Plant Physiology*, 146(4), pp. 1983-1995.

Vervecken, W., Callewaert, N., Kaigorodov, V., Geysens, S. and Contreras, R. (2007) 'Modification of the N-Glycosylation Pathway to Produce Homogeneous, Human-Like Glycans Using GlycoSwitch Plasmids', in Cregg, J.M. (ed.) *Pichia Protocols*. Totowa, NJ: Humana Press, pp. 119-138.

Voss-Fels, K.P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., Chen, T.-W., Zetzsche, H., Seddig, S. and Majid Baig, M. (2019) 'Breeding improves wheat productivity under contrasting agrochemical input levels', *Nature Plants*, 5(7), pp. 706-714.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J., Ramirez-Gonzalez, R.H., Kolodziej, M.C., Delorean, E. and Thambugala, D. (2020) 'Multiple wheat genomes reveal global variation in modern breeding', *Nature*, 588(7837), pp. 277-283.

Walling, L.L. (2000) 'The myriad plant responses to herbivores', *Journal of plant growth regulation*, 19(2), pp. 195-216.

Wan, X., Marsafari, M. and Xu, P. (2019) 'Engineering metabolite-responsive transcriptional factors to sense small molecules in eukaryotes: current state and perspectives', *Microbial Cell Factories*, 18(1), pp. 1-13.

Wang, C., Deng, P., Chen, L., Wang, X., Ma, H., Hu, W., Yao, N., Feng, Y., Chai, R. and Yang, G. (2013a) 'A wheat WRKY transcription factor TaWRKY10 confers tolerance to multiple abiotic stresses in transgenic tobacco', *PloS one*, 8(6).

Wang, H., Avci, U., Nakashima, J., Hahn, M.G., Chen, F. and Dixon, R.A. (2010) 'Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants', *Proceedings of the National Academy of Sciences*, 107(51), pp. 22338-22343.

Wang, H., Wang, H., Shao, H. and Tang, X. (2016a) 'Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology', *Frontiers in plant science*, 7, p. 67.

Wang, H., Wang, H., Shao, H. and Tang, X. (2016b) 'Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology', *Frontiers in plant science*, 7.

Wang, J., Tao, F., An, F., Zou, Y., Tian, W., Chen, X., Xu, X. and Hu, X. (2017) 'Wheat transcription factor TaWRKY70 is positively involved in high-temperature seedling plant resistance to Puccinia striiformis f. sp. tritici', *Molecular plant pathology*, 18(5), pp. 649-661.

Wang, L., Zheng, L., Zhang, C., Wang, Y., Lu, M. and Gao, C. (2015a) 'ThWRKY4 from Tamarix hispida can form homodimers and heterodimers and is involved in abiotic stress responses', *International journal of molecular sciences*, 16(11), pp. 27097-27106.

Wang, N., Fan, X., He, M., Hu, Z., Tang, C., Zhang, S., Lin, D., Gan, P., Wang, J. and Huang, X. (2022) 'Transcriptional Repression of TaNOX10 by TaWRKY19 Compromises ROS Generation and Enhances Wheat Susceptibility to Stripe Rust', *The Plant Cell*.

Wang, P., Du, Y., Zhao, X., Miao, Y. and Song, C.-P. (2013b) 'The MPK6-ERF6-ROS-responsive cis-acting Element7/GCC box complex modulates oxidative gene transcription and the oxidative response in Arabidopsis', *Plant physiology*, 161(3), pp. 1392-1408.

Wang, X., Zeng, J., Li, Y., Rong, X., Sun, J., Sun, T., Li, M., Wang, L., Feng, Y. and Chai, R. (2015b) 'Expression of TaWRKY44, a wheat WRKY gene, in transgenic tobacco confers multiple abiotic stress tolerances', *Frontiers in plant science*, 6, p. 615.

Wu, H., Ni, Z., Yao, Y., Guo, G. and Sun, Q. (2008) 'Cloning and expression profiles of 15 genes encoding WRKY transcription factor in wheat (Triticum aestivem L.)', *Progress in Natural Science*, 18(6), pp. 697-705.

Xie, Z., Zhang, Z.-L., Hanzlik, S., Cook, E. and Shen, Q.J. (2007) 'Salicylic acid inhibits gibberellin-induced alpha-amylase expression and seed germination via a pathway involving an abscisic-acid-inducible WRKY gene', *Plant molecular biology*, 64(3), pp. 293-303.

Xie, Z., Zhang, Z.L., Zou, X., Yang, G., Komatsu, S. and Shen, Q.J. (2006) 'Interactions of two abscisic-acid induced WRKY genes in repressing gibberellin signaling in aleurone cells', *The Plant Journal*, 46(2), pp. 231-242.

Xu, X., Chen, C., Fan, B. and Chen, Z. (2006) 'Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors', *The Plant Cell*, 18(5), pp. 1310-1326.

Xu, Y.-p., Xu, H., Wang, B. and Su, X.-D. (2020) 'Crystal structures of N-terminal WRKY transcription factors and DNA complexes', *Protein & cell*, 11(3), pp. 208-213.

Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T. and Tomo, Y. (2005) 'Solution structure of an Arabidopsis WRKY DNA binding domain', *The Plant Cell*, 17(3), pp. 944-956.

Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K. and Yokoyama, S. (2013) 'DNA-binding domains of plant-specific transcription factors: structure, function, and evolution', *Trends in plant science*, 18(5), pp. 267-276.

Yamasaki, K., Kigawa, T., Watanabe, S., Inoue, M., Yamasaki, T., Seki, M., Shinozaki, K. and Yokoyama, S. (2012) 'Structural basis for sequence-specific DNA recognition by an Arabidopsis WRKY transcription factor', *Journal of Biological Chemistry*, 287(10), pp. 7683-7691.

Yang, H., Koza, S.M. and Chambers, E.E. (2015) 'Method Development for Hydrophobic Interaction Chromatography (HIC) Based Protein Separations on Waters Protein-Pak Hi Res HIC Columns', *Application Note https://www. waters. com/webassets/cms/library/docs/720005514en. pdf (accessed 26 November 2019)*.

Ye, H., Qiao, L., Guo, H., Guo, L., Ren, F., Bai, J. and Wang, Y. (2021) 'Genome-wide identification of wheat WRKY gene family reveals that TaWRKY75-A is referred to drought and salt resistances', *Frontiers in Plant Science*, 12, p. 812.

Yehl, C.J. and Zydney, A.L. (2020) 'Single-use, single-pass tangential flow filtration using low-cost hollow fiber modules', *Journal of Membrane Science*, 595, p. 117517.

Yokotani, N., Ichikawa, T., Kondou, Y., Matsui, M., Hirochika, H., Iwabuchi, M. and Oda, K. (2009) 'Tolerance to various environmental stresses conferred by the salt-responsive rice gene ONAC063 in transgenic Arabidopsis', *Planta*, 229(5), pp. 1065-1075.

Yu, X., Feng, B., He, P. and Shan, L. (2017) 'From chaos to harmony: responses and signaling upon microbial pattern recognition', *Annual review of phytopathology*, 55, p. 109.

Yu, Y. and Zhang, L. (2021) 'Overexpression of TaWRKY46 enhances drought tolerance in transgenic wheat', *Cereal Research Communications*, pp. 1-10.

Zahoor, S.A., Ahmad, S., Ahmad, A., Wajid, A., Khaliq, T., Mubeen, M., Hussain, S., Din, M.S.U., Amin, A. and Awais, M. (2019) 'Improving water use efficiency in agronomic crop production', *Agronomic crops*, pp. 13-29.

Zandalinas, S.I., Sengupta, S., Fritschi, F.B., Azad, R.K., Nechushtai, R. and Mittler, R. (2021) 'The impact of multifactorial stress combination on plant growth and survival', *New Phytologist*, 230(3), pp. 1034-1048.

Zhang, G., Chen, M., Li, L., Xu, Z., Chen, X., Guo, J. and Ma, Y. (2009) 'Overexpression of the soybean GmERF3 gene, an AP2/ERF type transcription factor for increased tolerances to salt, drought, and diseases in transgenic tobacco', *Journal of experimental botany*, 60(13), pp. 3781-3796.

Zhang, L., Xu, Z., Ji, H., Zhou, Y. and Yang, S. (2019) 'TaWRKY40 transcription factor positively regulate the expression of TaGAPC1 to enhance drought tolerance', *BMC genomics*, 20(1), pp. 1-12.

Zhang, L., Zhang, L., Xia, C., Zhao, G., Liu, J., Jia, J. and Kong, X. (2015) 'A novel wheat bZIP transcription factor, TabZIP60, confers multiple abiotic stress tolerances in transgenic Arabidopsis', *Physiologia plantarum*, 153(4), pp. 538-554.

Zhang, W., Sinha, J. and Meagher, M.M. (2006) 'Glycerophosphate as a phosphorus source in a defined medium for Pichia pastoris fermentation', *Applied microbiology and biotechnology*, 72(1), pp. 139-144.

Zheng, L., Liu, G., Meng, X., Liu, Y., Ji, X., Li, Y., Nie, X. and Wang, Y. (2013) 'A WRKY gene from Tamarix hispida, ThWRKY4, mediates abiotic stress responses by modulating reactive oxygen species and expression of stress-responsive genes', *Plant molecular biology*, 82(4), pp. 303-320.

Zheng, X., Zhao, Y., Shan, D., Shi, K., Wang, L., Li, Q., Wang, N., Zhou, J., Yao, J. and Xue, Y. (2018) 'Md WRKY 9 overexpression confers intensive dwarfing in the M26 rootstock of apple by directly inhibiting brassinosteroid synthetase Md DWF 4 expression', *new phytologist*, 217(3), pp. 1086-1098.

Zheng, Z., Qamar, S.A., Chen, Z. and Mengiste, T. (2006) 'Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens', *The Plant Journal*, 48(4), pp. 592-605.

Zhou, Q.Y., Tian, A.G., Zou, H.F., Xie, Z.M., Lei, G., Huang, J., Wang, C.M., Wang, H.W., Zhang, J.S. and Chen, S.Y. (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(5), pp. 486-503.

Zhou, S., Zheng, W.-J., Liu, B.-H., Zheng, J.-C., Dong, F.-S., Liu, Z.-F., Wen, Z.-Y., Yang, F., Wang, H.-B. and Xu, Z.-S. (2019) 'Characterizing the role of TaWRKY13 in salt tolerance', *International journal of molecular sciences*, 20(22), p. 5712.

Zhu, C., Li, Z., Tang, Y., Zhang, L., Wen, J., Wang, Z., Su, Y. and Zhang, H. (2021) 'TaWRKY10 plays a key role in the upstream of circadian gene TaLHY in wheat', *Plant Science*, 310, p. 110973.

Zhu, H., Zhou, Y., Zhai, H., He, S., Zhao, N. and Liu, Q. (2020) 'A novel sweetpotato WRKY transcription factor, IbWRKY2, positively regulates drought and salt tolerance in transgenic Arabidopsis', *Biomolecules*, 10(4), p. 506.

Zou, M., Guan, Y., Ren, H., Zhang, F. and Chen, F. (2008) 'A bZIP transcription factor, OsABI5, is involved in rice fertility and stress tolerance', *Plant molecular biology*, 66(6), pp. 675-683.

8. Appendix

8.1 Appendix A

Candidate wheat WRKY transcription factors coding sequence:

TaWRKY53b coding sequence (1320 bp):

ATGAGTTCCTCCACTGGTTCCCTTGACCACGCCGGATTCACCTTTACCCCACCTCCATTCATCACATCATTCACAGAGTTGTT
GAGTGGTTCTGGTGCTGGAGATGCTGAGAGATCCCCAAGAGGTTTTAACAGAGGTGGTAGAGCTGGTGCTCCTAAATTCAAGT
CTGCTCAACCACCTTCTTTGCCAATTTCTTCTCCTTTTTCTTGTTTCTCTGTTCCAGCTGGTTTGTCTCCTGCTGAGTTGTTG
GATTCTCCAGTTTTGTTGAACTACTCTCATATCTTGGCTTCTCCAACTACTGGTGCTATTCCTGCTCAAAGATGTGATTGGCA
AGTTTCTGCTGATTTGAACACTTTCCAACAAGATGAATTGGGTTTGTCTGGTTTTTCTTTCCACGCTGTTAAATCTAACGCTA
CTGTTAATGCTCAAGCTAACTGTTTGCCATTGTTCAAGGAACAACAAGAGCAACAACAAGAAGAGGTTGTTCAAGTTTCTAAC
AAATCTTCTTCTTCTTCTGGTAACAATAAGCAAGTTGAGGATGGTTACAACTGGAGAAAGTACGGTCAAAAGCAAGTTAAGGG
TTCTGAAAACCCAAGATCCTACTACAAGTGTACTTACAACAACTGTTCTATGAAGAAAAAGGTTGAGAGATCCTTGGCTGATG
GTAGAATCACTCAAATCGTTTACAAGGGTGCTCACGATCACCCAAAGCCTTTGTCTACTAGAAGAAATTCTTCTGGTTGTGCT
GCTGTTGTTGCTGAAGATCATACTAATGGTTCTGAGCACTCTGGTCCAACTCCTGAAAACTCTTCTGTTACTTTTGGAGATGA
TGAAGCTGATAAACCAGAGACTAAAAGAAGAAAGGAACATGGAGATAACGAGGGTTCTTCTGGTGGTACTGGTGGTTGTGGTA
AACCAGTTAGAGAACCTAGATTGGTTGTTCAAACTTTGTCTGATATCGATATTTTGGATGATGGTTTCAGATGGAGAAAATAC
GGTCAAAAAGTTGTTAAGGGTAACCCAAATCCAAGATCCTACTATAAATGTACTACTGTTGGTTGTCCTGTTAGAAAGCATGT
TGAGAGAGCTTCTCACGATAACAGAGCTGTTATTACTACTTACGAAGGTAAACATTCTCACGATGTTCCAATTGGTAGAGGTA
GAGCTTTGCCTGCTTCTTCTTCTTCTGATTCTTCTGCTGTTATTTGGCCAGCTGCTGCTGTTCAAGCTCCTTGTACTTTGGAA
ATGTTGGCTGGTCACCCTGGATACGCCGCTAAGGATGAGCCAAGAGATGATATGTTTGTTGAAAGTTTGTTGTGCCCTCTA

Amino acid sequence:

MSSSTGSLDHAGFTFTPPPFITSFTELLSGSGAGDAERSPRGFNRGGRAGAPKFKSAQPPSLPISSPFSCFSVPAGLSPAELL
DSPVLLNYSHILASPTTGAIPAQRCDWQVSADLNTFQQDELGLSGFSFHAVKSNATVNAQANCLPLFKEQQEQQQEEVVQVSN
KSSSSSGNNKQVEDGYNWRKYGQKQVKGSENPRSYYKCTYNNCSMKKKVERSLADGRITQIVYKGAHDHPKPLSTRRNSSGCA
AVVAEDHTNGSEHSGPTPENSSVTFGDDEADKPETKRRKEHGDNEGSSGGTGGCGKPVREPRLVVQTLSDIDILDDGFRWRKY
GQKVVKGNPNPRSYYKCTTVGCPVRKHVERASHDNRAVITTYEGKHSHDVPIGRGRALPASSSSDSSAVIWPAAAVQAPCTLE
MLAGHPGYAAKDEPRDDMFVESLLCPL

TaWRKY19 coding sequence (1404bp):

ATGGCTGCAGGTCAATGGTCTGGTATTGGTGACGGTGGTGGTTTGTGGGCTCCACCAGCATTGGATTCATTATTTCCAGATGA
TCAACCATCTCCAGCTGCATCAGCTTTGGGTTTCTTTGGTGGTTCTTTGGCACAATTACCATCTCCACCACCATTGTGTGGTA
CAGCTTTGTTGGGTTACCCACAAGATAACTTCGATGTTTTCCATGAAAGAGATTTGGCTCAATTAGCTGCACAAGTTGCACAA
AAGAAAGAATTGAGAGAAAAACAAGGTGCTGGTTTACATCATAAAATTGGTCCACAATTGGCATTTTCAAAGTACTCTATTTT
AGATCAAGTTGATAATTCTTCATCTTTTTCTTTGGCTACTTCAGTTTTAACACCACAACATGTTTCATCTTCAGTTGGTGCTG
CATTGATGCAAGGTAGAACTTTACCATCTCATACAGGTTCAGGTTCTGTTAATACTGGTCCAACAGGTGTTTTGCAAGCATTA
CAAGATTCTTCAACTACATTGGATTCTATTAATACTGGTTCAACAGGTGTTTTGGAAGCTTTGCAAGGTTCTTCAATCACATT
AGATAGACCAGCAGATGATGGTTATAATTGGAGAAAATACGGTCAAAAAGCTGTTAAAGGTGGTAAATACCCAAGATCATACT
ACAAGTGTACTTTGAACTGTCCAGCTAGAAAGAATGTTGAACATTCAGCAGATAGAAGAATTATTAAGATCATCTATAGAGGT
CAACATTGTCATGAACCACCATCTAAAAGATTCAAAGATTGTGGTGACTTGTTGAACGAATTGAACGATTTCGATGATGCTAA
AGAACCATCAACAAAGTCTCAATTGGGTTGTCAAGGTTACTACGGTAAACCAATTACTCCAAATGGTATGATGACAGATGTTT
TGTTACCAACTAAAGAAGAAGGTGACGAACAATTGTCTTCATTGTCTGATATCAGAGAAGGTGACGGTGAAATTAGAACTGTT
GATGGTGACGATGGTGACGCTGATGCAAATGAAAGAAATGCTCCAGGTCAAAAGATTATCGTTTCAACTACATCTGATGCAGA
TTTGTTAGATGATGGTTATAGATGGAGAAAGTATGGTCAAAAAGTTGTTAGAGGTAATCCACATCCTAGATCATATTACAAGT
GTACATACCAAGGTTGTGATGTTAAGAAACATATCGAAAGATCATCTGAAGAACCACATGCTGTTATTACTACATACGAAGGT
AAACATACTCATGATGTTCCAGAATCAAGAAATAGATCACAAGCAACAGGTCAACATCATTGTAAGGAACAAACTTACTCAGA
ACAATCTGCTGCATCATTCTGTTCTTCATCTGAAAAGAGAAAGTACGGTACTGCTATCTTGAACGATTTGGCATT

Amino acid sequence:

MAAGQWSGIGDGGGLWAPPALDSLFPDDQPSPAASALGFFGGSLAQLPSPPPLCGTALLGYPQDNFDVFHERDLAQLAAQVAQ
KKELREKQGAGLHHKIGPQLAFSKYSILDQVDNSSSFSLATSVLTPQHVSSSVGAALMQGRTLPSHTGSGSVNTGPTGVLQAL
QDSSTTLDSINTGSTGVLEALQGSSITLDRPADDGYNWRKYGQKAVKGGKYPRSYYKCTLNCPARKNVEHSADRRIIKIIYRG
QHCHEPPSKRFKDCGDLLNELNDFDDAKEPSTKSQLGCQGYYGKPITPNGMMTDVLLPTKEEGDEQLSSLSDIREGDGEIRTV
DGDDGDADANERNAPGQKIIVSTTSDADLLDDGYRWRKYGQKVVRGNPHPRSYYKCTYQGCDVKKHIERSSEEPHAVITTYEG
KHTHDVPESRNRSQATGQHHCKEQTYSEQSAASFCSSSEKRKYGTAILNDLA

TaWRKY3 Coding sequence (687bp):

ATGGAAGGTGGTAGTCAATTAGGTGCCTGTTTGCCAAGTTTATACGCTTTAGACCCATACGCATCCCCTCCTTTGTTAGCCCC
ATTACCAAACCAACATAAGTTGCATCAATTGCCATTAGTTTTGCAAGAACAACCAGGTAATCATGGTGTTATGTTTTCTTCAG
ATCATGGTGGTGGTTTGTATCCATTGTTACCAGGTATTCCATTTTGTCATTCTGCTGCAGCTTGTGAAAAATCAACAGGTTTT
GCACCATTGGGTGGTACAGGTGAAGCAGGTACTTCTGCAGCTAGAGCTGGTAATGAATTTGCATCTGCTACTACAACTACAAC
TGCATCATGTCATGGTCCATCTTCATGGTGGAAAGGTGCTGAAAAGGGTAAAATGAAAGTTAGAAGAAAGATGAGAGAACCAA
GATTCTGTTTCCAAACAAGATCAGAAGTTGATGTTTTAGATGATGGTTACAAGTGGAGAAAGTACGGTCAAAAGGTTGTTAAA
AATTCTTTGCATCCAAGATCATACTACAGATGTACTCATTCTAACTGTAGAGTTAAGAAAAGAGTTGAAAGATTGTCAGAAGA
TTGTAGAATGGTTATTACAACTTACGAAGGTAGACATACACATACTCCTTGTTCTGATGACGACGCAGGTGGTGACCATACAG
GTTCTTGTGCTTTTACTTCTTTTTATCTAGA

Amino acid sequence:

MEGGSQLGACLPSLYALDPYASPPLLAPLPNQHKLHQLPLVLQEQPGNHGVMFSSDHGGGLYPLLPGIPFCHSAAACEKSTGF
APLGGTGEAGTSAARAGNEFASATTTTTASCHGPSSWWKGAEKGKMKVRRKMREPRFCFQTRSEVDVLDDGYKWRKYGQKVVK
NSLHPRSYYRCTHSNCRVKKRVERLSEDCRMVITTYEGRHTHTPCSDDDAGGDHTGSCAFTSFYL

TaWRKY3m coding sequence (687bp):

ATGGAAGGTGGTAGTCAATTAGGTGCCTGTTTGCCAAGTTTATACGCTTTAGACCCATACGCATCCCCTCCTTTGTTAGCCCC
ATTACCAAACCAACATAAGTTGCATCAATTGCCATTAGTTTTGCAAGAACAACCAGGTAATCATGGTGTTATGTTTTCTTCAG
ATCATGGTGGTGGTTTGTATCCATTGTTACCAGGTATTCCATTTTGTCATTCTGCTGCAGCTTGTGAAAAATCAACAGGTTTT
GCACCATTGGGTGGTACAGGTGAAGCAGGTACTTCTGCAGCTAGAGCTGGTAATGAATTTGCATCTGCTACTACAACTACAAC
TGCATCATGTCATGGTCCATCTTCATGGTGGAAAGGTGCTGAAAAGGGTAAAATGAAAGTTAGAAGAAAGATGAGAGAACCAA
GATTCTGTTTCCAAACAAGATCAGAAGTTGATGTTTTAGATGATGGTTACAAGTGGAGAAAGGACGGTCAAAAGGTTGTTAAA
AATTCTTTGCATCCAAGATCATACTACAGATGTACTCATTCTAACTGTAGAGTTAAGAAAAGAGTTGAAAGATTGTCAGAAGA
TTGTAGAATGGTTATTACAACTTACGAAGGTAGACATACACATACTCCTTGTTCTGATGACGACGCAGGTGGTGACCATACAG
GTTCTTGTGCTTTTACTTCTTTTTATCTAGA

Amino acid sequence:

MEGGSQLGACLPSLYALDPYASPPLLAPLPNQHKLHQLPLVLQEQPGNHGVMFSSDHGGGLYPLLPGIPFCHSAAACEKSTGF
APLGGTGEAGTSAARAGNEFASATTTTTASCHGPSSWWKGAEKGKMKVRRKMREPRFCFQTRSEVDVLDDGYKWRKDGQKVVK
NSLHPRSYYRCTHSNCRVKKRVERLSEDCRMVITTYEGRHTHTPCSDDDAGGDHTGSCAFTSFYL

pGAPZaA DNA sequence:

AGATCTTTTTTGTAGAAATGTCTTGGTGTCCTCGTCCAATCAGGTAGCCATCTCTGAAATATCTGGCTCCGTTGCAACTCCGA
ACGACCTGCTGGCAACGTAAAATTCTCCGGGGGTAAAACTTAAATGTGGAGTAATGGAACCAGAAACGTCTCTTCCCTTCTCTC
TCCTTCCACCGCCCGTTACCGTCCCTAGGAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCCCTTGCAGCAATGCTCTT
CCCAGCATTACGTTGCGGGTAAAACGGAGGTCGTGTACCCGACCTAGCAGCCCAGGGATGGAAAAGTCCCGGCCGTCGCTGGC
AATAATAGCGGGCGGACGCATGTCATGAGATTATTGGAAACCACCAGAATCGAATATAAAAGGCGAACACCTTTCCCAATTTT
GGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTATTTGTCCCTATTTCAATCAATTGAACAACTATTTCGAAACGATGAGA
TTTCCTTCAATTTTTACTGCTGTTTTATTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAAC
GGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACA
GCACAAATAACGGGTTATTGTTTATAAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGA
GAGGCTGAAGCTGAATTCACGTGGCCCAGCCGGCCGTCTCGGATCGGTACCTCGAGCCGCGGCGGCCGCCAGCTTTCTAGAAC
AAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATTGAGTTTTAGCCTTAGACATGACT
GTTCCTCAGTTCAAGTTGGGCACTTACGAGAAGACCGGTCTTGCTAGATTCTAATCAAGAGGATGTCAGAATGCCATTTGCCT
GAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCTC
GTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCTGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGAT
GTTTTTCTTGGTATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCCCCACACACCAT
AGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCGTACCACTTCAAAACACCCA
AGCACAGCATACTAAATTTTCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAGAAAAAAGA
GACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTGAAATTTTTTTTTTTAG
TTTTTTTCTCTTTCAGTGACCTCCATTGATATTTAAGTTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCATTTTTCTT
GTTCTATTACAACTTTTTTTACTTCTTGTTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGGGCGGTGTTGACAATTAATC
ATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGC
TCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTC
GCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGT
GTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCA
TGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAG
GAGCAGGACTGACACGTCCGACGGCGGCCCACGGGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTTTGTCGATATCATGT
AATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTC

TAGGTCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAG
ACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAGCTGG
AGACCAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGC
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCC
CCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCC
CCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGC
AGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT
ACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGC
AAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGCATGAGATC

8.2 Appendix B

TaWRKY53b cloning design

Table1: primer for TaWRKY53b

				200mM primer conc.	400mM primer conc.
Primer Nomenclature	Primer sequence	No. bp	C. Annal @	Comp. Ann @	Comp. Ann @
W53A-FP	GCTGAATTCATGTCCTCCTCCACGGGGAGC	30	65C	59C	60C
W53A-RP	TGTTCTAGAGGGCAGAGGAGCGACTCGACG	30			

DNA to amino acid sequence conversion website:

ExPASY Bioinformatics Resource Portal.

https://web.expasy.org/translate/

list for Shorthand Symbols for Amino acids:

http://web.sonoma.edu/users/t/thatcher/biol480/amino.htm

For reverse sequence:

https://www.bioinformatics.org/sms/rev_comp.html

primer Tm Calculation from:

http://tmcalculator.neb.com/#!/main

Ta WRKY53b gene had been cloned into pET28 plasmid. Primers were used to amplify the coding sequence of taWRKY53b along with flanking regions for pGAPZaA.

Ta WRKY53b amplification results:

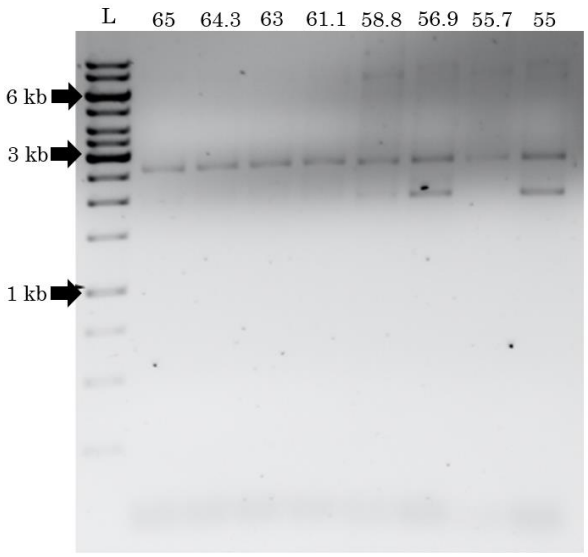


Figure 1. Amplification of *TaWRKY53b* amplification with a temperature gradient from 55 °C to 65 °C. Plasmid backbone pET28. Extension time up to 2 minutes with an addition of 3% DSMO. 10 kb Ladder.

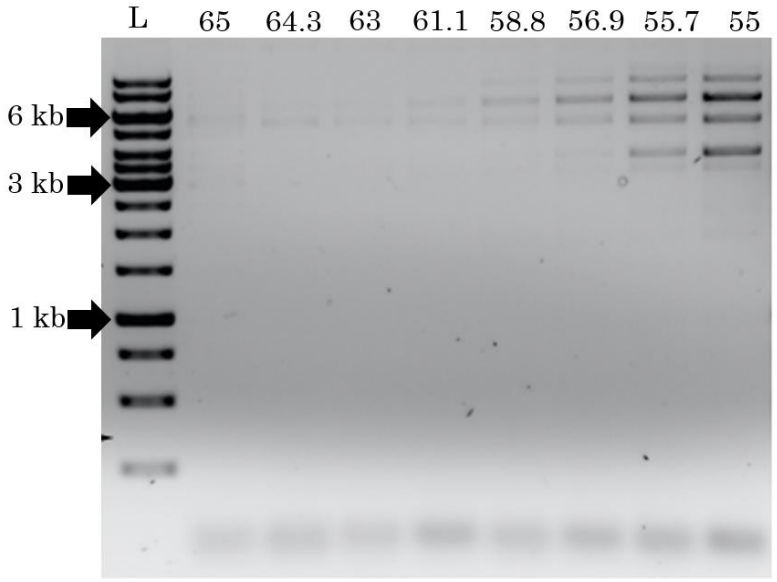


Figure 2. Amplification of *TaWRKY53b* amplification with a temperature gradient from 55 °C to 65 °C. Plasmid backbone pET28. Extension time up to 4 minutes. 10 kb Ladder.

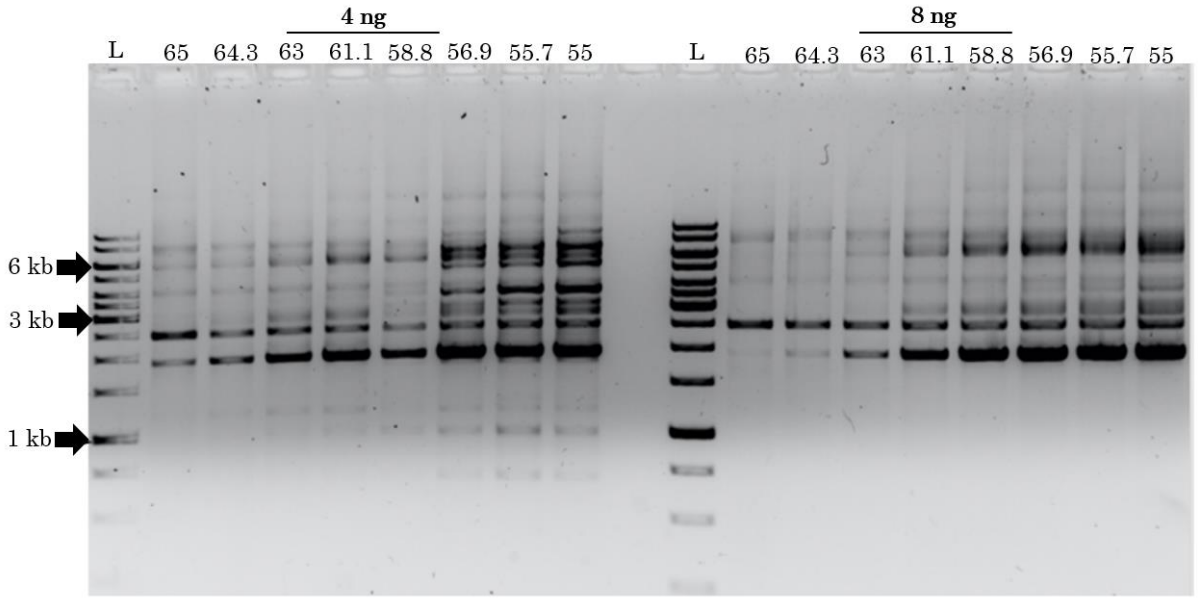


Figure 3. Amplification of *TaWRKY53b* amplification with a temperature gradient from 55 °C to 65 °C with 4 ng and 8 ng plasmid concentration. Plasmid backbone pET28. Extension time up to 2 minutes with an addition of 3% DSMO. 10 kb Ladder.

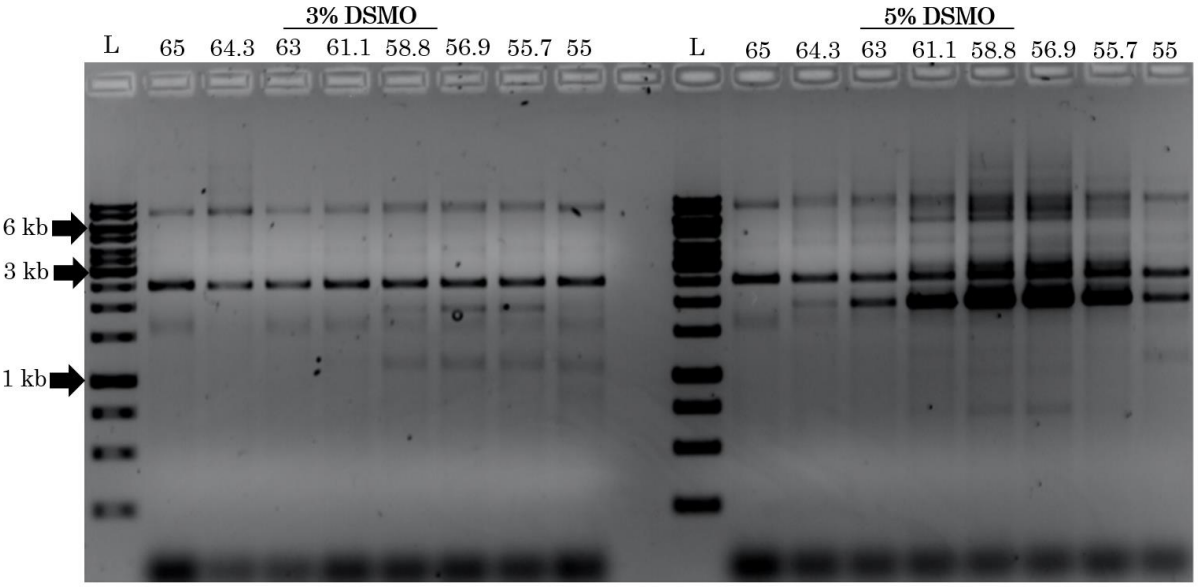


Figure 4. Amplification of *TaWRKY53b* amplification with a temperature gradient from 55 °C to 65 °C with 3% DSMO and 5% DSMO. Plasmid backbone pET28. Extension time up to 2 minutes with an addition of 3% DSMO. 10 kb Ladder.

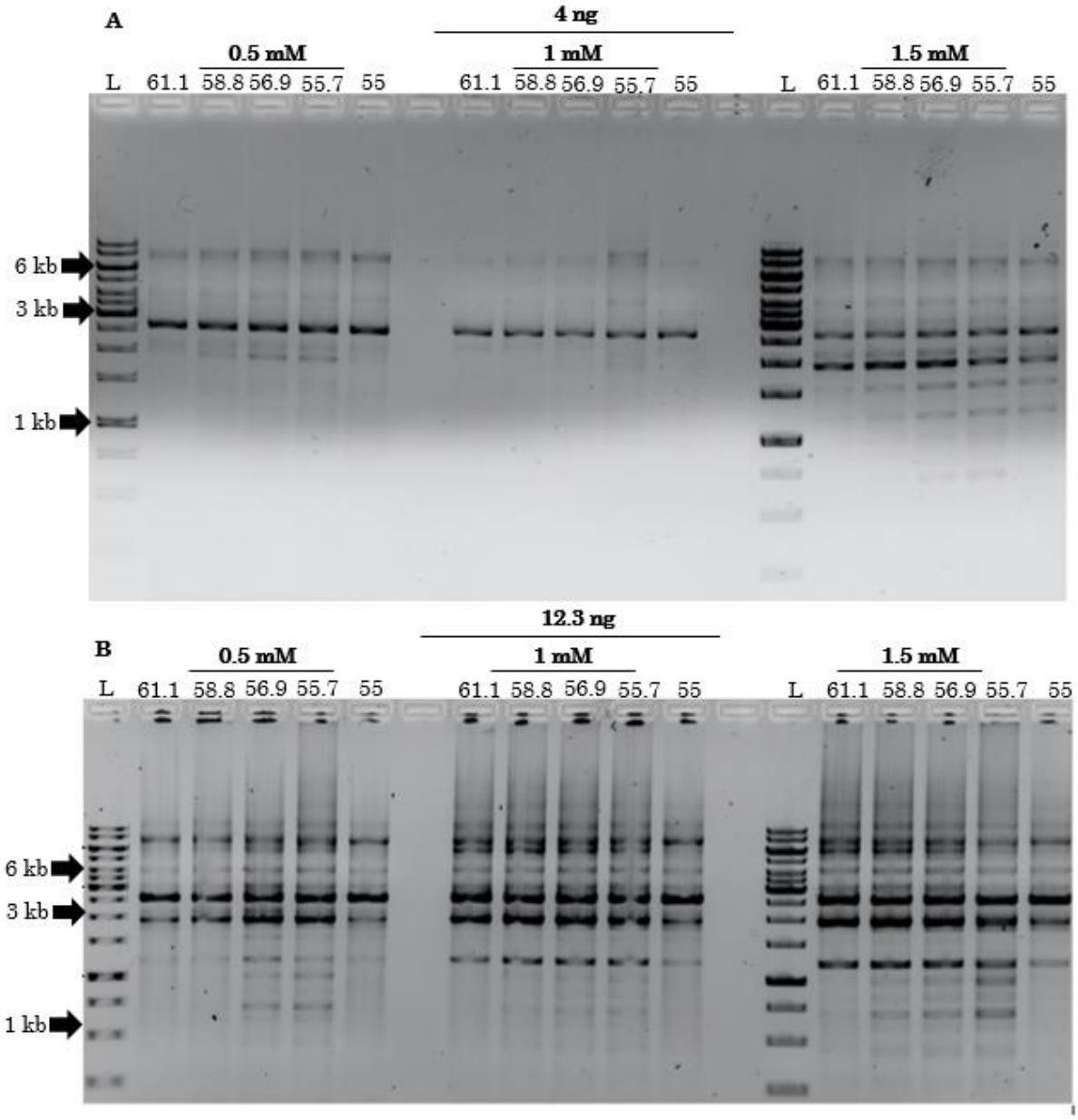


Figure 5. Amplification of *Ta WRKY53b* amplification with a temperature gradient from 55 °C to 61.1 °C with 4 ng and 12.3 ng plasmid concentration and change in MgCl2. Plasmid backbone pET28. Extension time up to 2 minutes with an addition of 3% DSMO. 10 kb Ladder.

8.3 Appendix C : Cloning candidate TaWRKYs into pGAPZaA:

TaWRKY53b protein sequence cloned in pGAPZaA expressing 6x His tag.

>pGAPZaA/TaWRKY53b DNA sequence:

CTCGAGAAAAGAGAGGCTGAAGCTATGAGTTCCTCCACTGGTTCCCTTGACCACGCCGGATTCACCTTTACCCCACCTCCATT
CATCACATCATTCACAGAGTTGTTGAGTGGTTCTGGTGCTGGAGATGCTGAGAGATCCCCAAGAGGTTTTAACAGAGGTGGTA
GAGCTGGTGCTCCTAAATTCAAGTCTGCTCAACCACCTTCTTTGCCAATTTCTTCTCCTTTTTCTTGTTTCTCTGTTCCAGCT
GGTTTGTCTCCTGCTGAGTTGTTGGATTCTCCAGTTTTGTTGAACTACTCTCATATCTTGGCTTCTCCAACTACTGGTGCTAT
TCCTGCTCAAAGATGTGATTGGCAAGTTTCTGCTGATTTGAACACTTTCCAACAAGATGAATTGGGTTTGTCTGGTTTTTCTT
TCCACGCTGTTAAATCTAACGCTACTGTTAATGCTCAAGCTAACTGTTTGCCATTGTTCAAGGAACAACAAGAGCAACAACAA
GAAGAGGTTGTTCAAGTTTCTAACAAATCTTCTTCTTCTTCTGGTAACAATAAGCAAGTTGAGGATGGTTACAACTGGAGAAA
GTACGGTCAAAAGCAAGTTAAGGGTTCTGAAAACCCAAGATCCTACTACAAGTGTACTTACAACAACTGTTCTATGAAGAAAA
AGGTTGAGAGATCCTTGGCTGATGGTAGAATCACTCAAATCGTTTACAAGGGTGCTCACGATCACCCAAAGCCTTTGTCTACT
AGAAGAAATTCTTCTGGTTGTGCTGCTGTTGTTGCTGAAGATCATACTAATGGTTCTGAGCACTCTGGTCCAACTCCTGAAAA
CTCTTCTGTTACTTTTGGAGATGATGAAGCTGATAAACCAGAGACTAAAAGAAGAAAGGAACATGGAGATAACGAGGGTTCTT
CTGGTGGTACTGGTGGTTGTGGTAAACCAGTTAGAGAACCTAGATTGGTTGTTCAAACTTTGTCTGATATCGATATTTTGGAT
GATGGTTTCAGATGGAGAAAATACGGTCAAAAAGTTGTTAAGGGTAACCCAAATCCAAGATCCTACTATAAATGTACTACTGT
TGGTTGTCCTGTTAGAAAGCATGTTGAGAGAGCTTCTCACGATAACAGAGCTGTTATTACTACTTACGAAGGTAAACATTCTC
ACGATGTTCCAATTGGTAGAGGTAGAGCTTTGCCTGCTTCTTCTTCTTCTGATTCTTCTGCTGTTATTTGGCCAGCTGCTGCT
GTTCAAGCTCCTTGTACTTTGGAAATGTTGGCTGGTCACCCTGGATACGCCGCTAAGGATGAGCCAAGAGATGATATGTTTGT
TGAAAGTTTGTTGTGCCCTCTAGAACAAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATC
ATTGA

>pGAPZaA/TaWRKY53b protein sequence.

MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPF
SNSTNNGLLFINTTIASIAAKEEGVSLEKREAEAMSSSTGSLDHAGFTFTPPPFI
TSFTELLSGSGAGDAERSPRGFNRGGRAGAPKFKSAQPPSLPISSPFSCFSVPAG
LSPAELLDSPVLLNYSHILASPTTGAIPAQRCDWQVSADLNTFQQDELGLSGFSF
HAVKSNATVNAQANCLPLFKEQQEQQQEEVVQVSNKSSSSSGNNKQVEDGYNWRK
YGQKQVKGSENPRSYYKCTYNNCSMKKKVERSLADGRITQIVYKGAHDHPKPLST
RRNSSGCAAVVAEDHTNGSEHSGPTPENSSVTFGDDEADKPETKRRKEHGDNEGS
SGGTGGCGKPVREPRLVVQTLSDIDILDDGFRWRKYGQKVVKGNPNPRSYYKCTT
VGCPVRKHVERASHDNRAVITTYEGKHSHDVPIGRGRALPASSSSDSSAVIWPAA
AVQAPCTLEMLAGHPGYAAKDEPRDDMFVESLLCPLEQKLISEEDLNSAVDHHHH
HH

>pGAPZaA/TaWRKY19 (DNA sequence)

CTCGAGAAAAGAGAGGCTGAAGCTatggctgcaggtcaatggtctggtattggtgacggtggtggtttgtgggctccaccagcattgg
attcattatttccagatgatcaaccatctccagctgcatcagctttgggtttctttggtggttctttggcacaattaccatctccaccaccattgtgtg
gtacagctttgttgggttacccacaagataacttcgatgttttccatgaaagagatttggctcaattagctgcacaagttgcacaaaagaaagaat
tgagagaaaaacaaggtgctggtttacatcataaaattggtccacaattggcattttcaaagtactctattttagatcaagttgataattcttcatc
ttttttctttggctacttcagttttaacaccacaacatgtttcatcttcagttggtgctgcattgatgcaaggtagaactttaccatctcatacaggttc
aggttctgttaatactggtccaacaggtgttttgcaagcattacaagattcttcaactacattggattctattaatactggttcaacaggtgttttgg
aagctttgcaaggttcttcaatcacattagatagaccagcagatgatggttataattggagaaaatacggtcaaaaagctgttaaaggtggtaa
atacccaagatcatactacaagtgtactttgaactgtccagctagaaagaatgttgaacattcagcagatagaagaattattaagatcatctata
gaggtcaacattgtcatgaaccaccatctaaaagattcaaagattgtggtgacttgttgaacgaattgaacgatttcgatgatgctaaagaacca
tcaacaaagtctcaattgggttgtcaaggttactacggtaaaccaattactccaaatggtatgatgacagatgttttgttaccaactaaagaaga
aggtgacgaacaattgtcttcattgtctgatatcagagaaggtgacggtgaaattagaactgttgatggtgacgatggtgacgctgatgcaaatg
aaagaaatgctccaggtcaaaagattatcgtttcaactacatctgatgcagatttgttagatgatggttatagatggagaaagtatggtcaaaaa

gttgttagaggtaatccacatcctagatcatattacaagtgtacataccaaggttgtgatgttaagaaacatatcgaaagatcatctgaagaacc
acatgctgttattactacatacgaaggtaaacatactcatgatgttccagaatcaagaaatagatcacaagcaacaggtcaacatcattgtaag
gaacaaacttactcagaacaatctgctgcatcattctgttcttcatctgaaaagagaaagtacggtactgctatcttgaacgatttggcattTCTA
GAACAAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATCATTGA

Protein expression reading frame:

>pGAPZaA/TaWRKY19 (protein sequence)

LEKREAEAMAAGQWSGIGDGGGLWAPPALDSLFPDDQPSPAASALGFFGGSLAQLP
SPPPLCGTALLGYPQDNFDVFHERDLAQLAAQVAQKKELREKQGAGLHHKIGPQL
AFSKYSILDQVDNSSSFSLATSVLTPQHVSSSVGAALMQGRTLPSHTGSGSVNTG
PTGVLQALQDSSTTLDSINTGSTGVLEALQGSSITLDRPADDGYNWRKYGQKAVK
GGKYPRSYYKCTLNCPARKNVEHSADRRIIKIIYRGQHCHEPPSKRFKDCGDLLN
ELNDFDDAKEPSTKSQLGCQGYYGKPITPNGMMTDVLLPTKEEGDEQLSSLSDIR
EGDGEIRTVDGDDGDADANERNAPGQKIIVSTTSDADLLDDGYRWRKYGQKVVRG
NPHPRSYYKCTYQGCDVKKHIERSSEEPHAVITTYEGKHTHDVPESRNRSQATGQ
HHCKEQTYSEQSAASFCSSSEKRKYGTAILNDLAFLEQKLISEEDLNSAVDHHHH
HH

>2_pGAPZaA/TaWRKY3

CTCGAGAAAAGAGAGGCTGAAGCTATGGAAGGTGGTAGTCAATTAGGTGCCTGTTTGCCAAGTTTATACGCT
TTAGACCCATACGCATCCCCTCCTTTGTTAGCCCCATTACCAAACCAACATAAGTTGCATCAATTGCCATTAGTT
TTGCAAGAACAACCAGGTAATCATGGTGTTATGTTTTCTTCAGATCATGGTGGTGGTTTGTATCCATTGTTACC
AGGTATTCCATTTTGTCATTCTGCTGCAGCTTGTGAAAAATCAACAGGTTTTGCACCATTGGGTGGTACAGGTG
AAGCAGGTACTTCTGCAGCTAGAGCTGGTAATGAATTTGCATCTGCTACTACAACTACAACTGCATCATGTCAT
GGTCCATCTTCATGGTGGAAAGGTGCTGAAAAGGGTAAAATGAAAGTTAGAAGAAAGATGAGAGAACCAAG
ATTCTGTTTCCAAACAAGATCAGAAGTTGATGTTTTAGATGATGGTTACAAGTGGAGAAAGTACGGTCAAAAG
GTTGTTAAAAATTCTTTGCATCCAAGATCATACTACAGATGTACTCATTCTAACTGTAGAGTTAAGAAAAGAGT
TGAAAGATTGTCAGAAGATTGTAGAATGGTTATTACAACTTACGAAGGTAGACATACACATACTCCTTGTTCTG
ATGACGACGCAGGTGGTGACCATACAGGTTCTTGTGCTTTTACTTCTTTTTATCTAGAACAAAAACTCATCTCA
GAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATTGA

Expressed protein:

LEKREAEAMEGGSQLGACLPSLYALDPYASPPLLAPLPNQHKLHQLPLVLQEQPGN
HGVMFSSDHGGGLYPLLPGIPFCHSAAACEKSTGFAPLGGTGEAGTSAARAGNEF
ASATTTTTASCHGPSSWWKGAEKGKMKVRRKMREPRFCFQTRSEVDVLDDGYKWR
KYGQKVVKNSLHPRSYYRCTHSNCRVKKRVERLSEDCRMVITTYEGRHTHTPCSD
DDAGGDHTGSCAFTSFYLEQKLISEEDLNSAVDHHHHHH–

>3_pGAPZaA/TaWRKY3m

CTCGAGAAAAGAGAGGCTGAAGCTATGGAAGGTGGTAGTCAATTAGGTGCCTGTTTGCCAAGTTTATACGCT
TTAGACCCATACGCATCCCCTCCTTTGTTAGCCCCATTACCAAACCAACATAAGTTGCATCAATTGCCATTAGTT
TTGCAAGAACAACCAGGTAATCATGGTGTTATGTTTTCTTCAGATCATGGTGGTGGTTTGTATCCATTGTTACC
AGGTATTCCATTTTGTCATTCTGCTGCAGCTTGTGAAAAATCAACAGGTTTTGCACCATTGGGTGGTACAGGTG
AAGCAGGTACTTCTGCAGCTAGAGCTGGTAATGAATTTGCATCTGCTACTACAACTACAACTGCATCATGTCAT
GGTCCATCTTCATGGTGGAAAGGTGCTGAAAAGGGTAAAATGAAAGTTAGAAGAAAGATGAGAGAACCAAG
ATTCTGTTTCCAAACAAGATCAGAAGTTGATGTTTTAGATGATGGTTACAAGTGGAGAAAGGACGGTCAAAA
GGTTGTTAAAAATTCTTTGCATCCAAGATCATACTACAGATGTACTCATTCTAACTGTAGAGTTAAGAAAAGAG
TTGAAAGATTGTCAGAAGATTGTAGAATGGTTATTACAACTTACGAAGGTAGACATACACATACTCCTTGTTCT
GATGACGACGCAGGTGGTGACCATACAGGTTCTTGTGCTTTTACTTCTTTTTATCTAGAACAAAAACTCATCTC
AGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATTGA

Expressed protein:

LEKREAEAMEGGSQLGACLPSLYALDPYASPPLLAPLPNQHKLHQLPLVLQEQPGN
HGVMFSSDHGGGLYPLLPGIPFCHSAAACEKSTGFAPLGGTGEAGTSAARAGNEF
ASATTTTTASCHGPSSWWKGAEKGKMKVRRKMREPRFCFQTRSEVDVLDDGYKWR
KDGQKVVKNSLHPRSYYRCTHSNCRVKKRVERLSEDCRMVITTYEGRHTHTPCSD
DDAGGDHTGSCAFTSFYLEQKLISEEDLNSAVDHHHHHH-

Appendix D:

TaPR1-23 (forward strand) (HQ700377.1)

>TraesCS5A02G059000.1 TraesCS5A02G059000:TraesCS5A02G059000.1

>5A dna:chromosome chromosome:IWGSC:5A:59563139:59565633:1

ATCTTCTTGGATTTATATCCGATGTAATCTTCTTTTGTTGTGCATTTGTTGGGATCTAAT

AAATTGTGGGTTTATGATCAGATTATTCATTGAAAGTAATTGAGTCTTTTCTGAACTTTA

TTATGCATGATTATTATAGTTTTATATTTATCTTCGATCTATCTGTTTGGT**TTGACC**AAC

TAGATTGATTTATCTTCAGTGGGAGAGGTGCTTTGTGATGGGTTCAATCTTGTGGTGTCC

TCACCTCCTGATAGAAGGGGTAGCGAGGCATACATTGTATTGTTTCCAATAAGGGTAAAA

TGGTGGAGTTTATTCTTATTGCTTGAGTTTACTTTGTCTACATCATATCATCGTACTTCA

TGCATTACTCTATTTGTCATGAACTTAATACGTAGAGAGGCAAGCATCGAAGCGCTCTCG

AAGTGGAGTAGTAGTAGATGCAGAATCGTTTCGGTCTACTTGTCACGGACATGATGCATA

TATACATGACCATTGACATGAATAACATAATAACTTTGTGTTTTCTATCAATTATCCAAG

AGTAACTTGTCTACCCACCGTATGCTATTGTTTCGGGAGAGAAGCATCTAGTGAAAACTA

TGGCCCCTGGGTCTATCTTAATCATATTATTAAAACCAAAAATACCTTACTGTAATTTAT

TTACTTTTCTTTTTTCTATCTACCGCTATCAGGTTTGCAAGTAACGATTTCAAGGGGATT

GACAACCCTTTTTCCCACGTTGAGTGTAAGTATTTGCTCTTTTGTGCGCAGGTATTCTTC

ACTAGGATTTGCGTGGTTCTCCTATTGGTTCGATAATCTTGGTTCTCGCTAAGGAAAATA

CTTATCAGCTACTATACTGTTTCACCCTTTCTCTTCGGGGAAAATCCCGACGCATCTCAC

AAGTAGCAACATCGTCGCATCACCAACTCGTAGCAACCCCACATCCCCCTCACTCAACTC

CTCACGGCGTTGTTGACGACCATCACAACACCAACTTGCATCAATACTATAGGCACCTCA

CGCTACTGCTTGCAGCATCTCCGGCAGCCGTCGCCGATGACATCACAGTAGCAACATGCA

ACAACAACAGCACGGTGGCGCCTTGCACATCGTTGTCTTGCATCATGGATACAGCACCCC

CATGTCATATTGTTTGCAACACGGATGCAATGAGCGGCGCTGGATTTGCAAAACCGATCC

CCAACTTGTAGCATCGCCGGTCGCAGTGGGATTATATCCACCAAGCATGGGTGAGGTGCA

TCTCGTAGCACCATGTGCCGCCTTGCTGATGTTGTGACAACAACAAGATCCATTGGCGGC

TGCTGTAGTCGCCCTCGTCCAGTTCACTTGGACGGCGATACCACAACTCAAACGGGAGAG

GTAGCTGAAAGCATTCCACTGGCGAGGGAAAATGTAAGGAAGAGGGAGAGGGCACATAAG

TAGGGCTCACTGAATGTGTGTCGCTCGCGCGGGGCGGTTGTTGATGCGAGAGTCTACCGC

CGGTTGAAACGAATCAATTTCCTTGTTTCATACAGTGCTTATTGGTTAGTTTTACATAAA

AATTTTGTCATAATCTTGTGTGTGCAAACATATGCTTCAAACATACTAAGACAGACATTA

TTTTGTTGTTCTTTTATCTTGCTTTTCGTTAGTATCTTATTAATCAATATTTCATATTAA

ATTTTGCCAAAAATATTAACTTAGGTGATCTTTTAATGGTAATGGACATTTTTTGGATTT

CCATTCCTTCATTAATACTTTAGAGTATTTACATCAAATTTATTTCTAAATATAAGTGCG

ATTCTATGGTAGACACTAATTCCCGGTGGTTGTTCAATATCCTTTCATTACATCCTGCCA

GCCATAGTTACGTATACTCGATAATTTCGGCTTATCCTTCAACAAAATTTCTCTATAAAT

ACACTGCGAAAGTGGACGCTACCACACAAGAGTACGAAACAAAAACATAAGCAATATATT

GCACATTCAAAGGCGTACACATGGAGGCATCTAAACTCGCCCTTTTGGTCATGCTCGCAA

TGGCAGCCGTGATGTCTGACCCCTGCAATGCCCAGAACTCGCCACACGACTACGTCGTCG

CCCACAACGTCGCTCGCGCCGCCGTGGGCTTGGGCCTTGTGACCTGGGACGCGTCCGTGG

CGGCCTATGCGGCGAGCTACGCAAGGCAGCGCTCCGGCGATTGCAAGCTGGTGCACTCAA

AGGCACCACAGTACGGGGAGAACCTCTTCTGGGGCTCCGGCGAGGACTGGACGGCTGCGC

AGGCCGTGAAGATATGGGCCGACGAGAAGGCCAACTACAACTACGCCTCCAACAGCTGCG

CCGCCGGGAAGCAGTGCGGGCACTACACGCAGATAGTGTGGCGCAACTCAACGCATATCG

GCTGCGCGCGTCTGCTCTGCGACCACAACGCCGGCGTGTTCATCACTTGCAATTACAGCC

CTCCGGGCAACTACATTGGGCAGAGGCCATATTGA

Green highlight indicate the W-box region that had been selected for EMSA test.

Red highlight indicate the starting site for TaPR1-23.

Yellow highlight indicate TaPR1-23 gene that had been obtained from NCBI.

This TaPR1-23 had been taken from NCBI search.

This is our methodology we used for obtaining W-box element from TaPR1-23

Pathogen related protein for *Triticum aestivum* gene selected from NCBI database. From the data base, its nucleotides gene was searched by plotting on the search side “Pathogen related protein triticum” and selected only Triticum aestivum as an organism. The output results had shown a list of pathogens related genes (in wheat). A number of these genes had been selected for analyzing their 2000 bp upstream promoter region using PlantCARE which was obtained from Ensemble (TraesCS5A02G059000.1) [5A dna:chromosome chromosome:IWGSC:5A:59563139:59565633:1]. Based on their upstream promoter region, TaPR1-23 gene (HQ700377.1) was selected due to the presence of a singular W-box element was found at it 171 bp.

Appendix E:

Table 8-1 Plant regulation map for *TaWRKY19*.

Protein ID	Uniport Description
Traes_1DL23D036A85	bZIP family protein
Traes_2BS_B1A73C7A8.1	Ethylene-responsive transcription factor ERF096
TRAES3BF001900070CFD_t1	bZIP family protein. Function: Binds to the embryo specification element and the ABA-responsive (ABRE) of the Dc3 gene promoter. Could participate in abscisic acid-regulated gene expression during seed development.
TRAES3BF099600130CFD_g	bZIP family protein, Function: Involved in abscisic acid (ABA) signalling pathway. Binds to the G-box motif 5'-CACGTG-3' of TRAB1 gene promoter Involved in the regulation of pollen maturation. May act as negative regulator of salt stress response. Together with PYL5, PP2C30 and SAPK2, is part of an ABA signaling unit that modulates seed germination and early seedling growth. Regulation: INDUCTION: Induced by abscisic acid (ABA). Induced by salt stress. Down-regulated by cold and drought stresses.
Traes_4BL_78DD63002	HD-ZIP family protein l/ll which is induced in leaves by drought stress.
Traes_5BL_A5532B750	bZIP family protein. Function: probable transcription factors that may be involved in responses to fungal pathogen infection and abiotic stresses. Regulation: Induced during incompatible interaction with the fungal pathogen Puccinia striiformis. Induced by abscisic acid (ABA), ethylene, cold stress, salt stress and wounding
Traes_5BS_FF44610EF1	bZIP family protein Function: May contribute to developmentally specific patterns of gene expression. Binds specifically to cis elements which are transcriptional enhancer found in the promoters of several plant genes. OCSBF-1 is able to bind to a site within each half of the ocs element as well as to animal AP-1 and CREB sites.
Traes_6DS_2B2F9C290	C2H2 zinc finger family protein
Traes_7DL_68B814464	bZIP family protein. Function: Transcription factor that promotes photomorphogenesis in light. Acts downstream of the light receptor network and directly affects transcription of light-induced genes. Specifically involved in the blue light specific pathway, suggesting that it participates in transmission of cryptochromes (CRY1 and CRY2) signals to downstream responses. In darkness, its degradation prevents the activation of light-induced genes (Probable). Acts co-ordinately with SPL7 to regulate the microRNA miR408 and its target genes in response to changes in light and copper conditions. Regulates the abscisic acid (ABA) signaling pathway. Also involved in root gravitropism.
Traes_7DS_433704E8E1	ERF family protein Function: Probably acts as a transcriptional activator. Binds to the GCC-box pathogenesis-related promoter element. May be involved in the regulation of gene expression by stress factors and by components of stress signal transduction pathways.
Traes_7DS_433704E8E	ERF family protein Function: Probably acts as a transcriptional activator. Binds to the GCC-box pathogenesis-related promoter element. May be involved in the regulation of gene expression by stress factors and by components of stress signal transduction pathways.

* Proteins the map had been named by their protein ID due to the lack of common name of each. However, family name of each protein was found in PlantTFDB data base. Their functions and regulations were mostly found which in some is based in similarity with other plant species.

Table 8-2 Plant regulation map for *TaWRKY53b*.

Protein ID	Uniport Description
Traes_2BL_9528AAD7C	bZIP family protein Function: Transcriptional regulator involved in defence response.
Traes_4AL_5E7F93445.2	bZIP family protein Function: Transcriptional regulator involved in defence response.
Traes_4BL_78DD63002.1	HD-ZIP family protein Regulation: induction in leaves by drought stress.
Traes_5DL_1950C1FC2.2	bZIP family protein Function: Transcription activator that binds to as1-like elements (5'-TGACGTAAgggaTGACGCA-3') in promoters of target genes. Regulates transcription in response to plant signalling molecules salicylic acid (SA), methyl jasmonate (MJ) and auxin (2,4D) only in leaves. Prevents lateral branching and may repress defence signalling.

* Proteins the map had been named by their protein ID due to the lack of common name of each. However, family name of each protein was found in PlantTFDB data base. Their functions and regulations were mostly found which in some is based in similarity with other plant species.

Table 8-3 Plant regulation map for *TaWRKY3*.

Protein ID	**Uniport Description**
Traes_2BL_9CD6E043A.2	ERF family protein Function: Probably acts as a transcriptional activator. Binds to the GCC-box pathogenesis-related promoter element. May be involved in the regulation of gene expression by stress factors and by components of stress signal transduction pathways (By similarity).
Traes_4AL_A02408BC8	BBR-BPC family protein Function: Transcriptional regulator that specifically binds to GA-rich elements (GAGA-repeats) present in regulatory sequences of genes involved in developmental processes.
Traes_5BL_F258582BB	GATA family protein
Traes_7AL_EA6F4FFDE	GATA family protein Function: Transcriptional activator that specifically binds 5'-GATA-3' or 5'-GAT-3' motifs within gene promoters. Regulation: INDUCTION: By abscisic acid (ABA), and drought and salt stresses. Down-regulated by Jasmonate and wounding.
Traes_7DS_FE8BC1125	BBR-BPC family protein Function: Transcriptional regulator that specifically binds to GA-rich elements (GAGA-repeats) present in regulatory sequences of genes involved in developmental processes..

* Proteins the map had been named by their protein ID due to the lack of common name of each. However, family name of each protein was found in PlantTFDB data base. Their functions and regulations were mostly found which in some is based in similarity with other plant species