7-31-2013 12:00 AM

Characterization of a Putative Activation Domain in the Hulk Gene Family

Christopher Doan, *The University of Western Ontario*

Supervisor: Dr. Vojislava Grbic, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

CHARACTERIZATION OF A PUTATIVE ACTIVATION DOMAIN IN THE *HULK* GENE FAMILY

(Thesis format: Monograph)

by

Christopher Doan

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

The *HULK* gene family participates in regulation of both flowering time and development in the plant *Arabidopsis thaliana*. The proteins encoded by these genes share conserved domain structures including a proline-rich region (PRR) in the carboxyl-terminus. Based on sequence analysis and the presence of a proline-rich domain, it has been suggested that the HULKs are putative transcription factors in which HUA2 is known to regulate several late-flowering genes: *FLC*, *FLM* and *MAF2*.

To investigate the putative transcriptional activation domain in the carboxyl-terminus of the HULKs, full-length HULKs and deletion constructs were 3-AT titrated in yeast-one hybrid experiments. The transcriptional activity varied between both the full-length and carboxyl-terminus of the HULKs as well as between the HULKs themselves. However, HULK2 carried the strongest transcriptional activation domain, which was active in both the full-length protein and when expressed as just in the carboxyl-terminus.

A domain swap was then performed with HULK2's PRR and HULK3's significantly weaker PRR to see if the transcriptional activity observed in HULK2 was localised to the PRR. While it was established that the PRR found at the carboxyl-terminus did contribute to the transcriptional activity, it was determined that the domain is not solely responsible for the transcriptional activity. The data suggests that there are multiple transcriptional activation domains working in tandem in the HULKs.

Keywords

Arabidopsis, HULK, HULK2, HULK3, PRR, transcriptional activation domain, yeast-one hybrid

Acknowledgments

This Master of Science project has been a significant event in my life. I have furthered my understanding of biology and also developed as person. This growth was not done alone and I take this opportunity to thank all of those who gave me guidance, support, and showed much patience.

I begin by thanking my supervisor, Dr. Vojislava Grbic. She took me into the lab and provided ideas and support throughout my project. I also would like to thank my advisory committee members, Dr. Rima Menassa and Dr. Priti Krishna, for their suggestions and expertise.

I thank Dr. Andre Lachance for reading through this thesis. Without his help, this thesis would not be here today.

In the Grbic lab, I thank past and present lab members who have helped me in both lab work, and in areas outside the lab, mainly the Grad Club. Special note goes to Dr. Vladimir Zhurov, Dr. Preetam Janakirama and Ingrid Fung whose immense help and friendship saved me countless times.

For all my friends in London and Mississauga, I love you all! I am thankful for the support I felt from friends back home and grateful for the friends that I have made in London for their company.

Last but not least, I thank my family who gave me unconditional love and care. Their sacrifices and love helped me get here and produce this thesis. Cheers!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3-AT	3-Amino-1,2,4-triazole
^{o}C	degree Celsius
µg	microgram
µl	microliter
aa	amino acid
AD	activation domain
bp	base pair
cDNA	complementary deoxyribonucleic acid
CTD	c-terminal domain of RNA polymerase II
CO	*CONSTANS*
DNA	deoxyribonucleic acid
DBD	DNA binding domain
FLC	*FLOWER LOCUS C*
FRI	*FRIGIDA*
FT	*FLOWERING LOCUS T*
FWD	forward
g	grams
GA	gibberellin or gibberellic acid
GTF	general transcription factor
HULK	HUA2-like gene
kb	kilo base pair
L	liter
LB	Lysogeny broth medium

LFY	*LEAFY*
MCS	multiple cloning site
Min	minute
mRNA	messenger ribonucleic acid
ml	milliliters
NLS	nuclear localisation signal
OD_{600}	optical density at 600 nm
OE-PCR	overlapping extension PCR
ORF	open reading frame
PCR	polymerase chain reaction
PPLP	proline-proline-leucine-proline
PWWP	proline-tryptophan-tryptophan-proline
REV	Reverse
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RPR	regulation of nuclear pre-mRNA processing
SAM	shoot apical meristem
SOC1	*SUPPRESSOR OF OVEREXPRESSION OF CO 1*
SDS	sodium dodecyl sulphate
TAD	transcriptional activation domain
WW	tryptophan-tryptophan

Chapter 1 : Introduction

1.1 *Arabidopsis thaliana* as a model organism

Arabidopsis thaliana is a dicotyledonous member of the Brassicaceae family, native to Europe, Asia and northwestern Africa. While it is not agriculturally relevant, *A. thaliana* has been used as a model organism for research in plant sciences, including plant development and genetics, for over 20 years (Meyerowitz, 2001). Although the earliest non-taxonomic report of *A. thaliana* was performed in 1873 by Alexander Braun, it wasn't until the late 1940s that Freidrich Laibach proposed *A. thaliana*'s utility as a model organism (Meyerowitz, 2001). By the 1980s, numerous works highlighted the value and properties of *A. thaliana* as a small plant. It is easily cultivated with limited space requirements, its quick life cycle of 5 – 6 weeks, and the large number of seeds produced that can arise from either self-pollination or crosses, make it a desirable for laboratory studies (Meinke, 1998; Meyerowitz, 2001). *A. thaliana* can also be easily transformed by *Agrobacterium tumefaciens* and has a small genome of 135 megabases that is amenable to genetic engineering. Mutations can be isolated, resulting in large collections of mutant lines available (Meinke, 1998). Furthermore, by the end of 2000, an initiative by the Arabidopsis Genome Initiative (AGI) led to the sequencing of *A. thaliana*'s genome. As of 2012, the AGI has catalogued over 27,000 protein coding genes of *A. thaliana*, organized into 5 chromosomes (Lamesch et al., 2012).

1.2 Flowering time in *A. thaliana*

The transition from vegetative to reproductive development, flowering time, is a major phase in a plant's life cycle and has been widely studied in *A. thaliana*. The mature *A. thaliana* flower consists of four types of floral organs: sepals, petals, stamens, and carpels (Chen and Meyerowitz, 1999). Plants develop from meristems, analogous to stem cells in animals, which proliferate to maintain meristematic cells apically and form derivative populations at the periphery that differentiate into primordia that give rise to organs (Irish, 2010). During vegetative development, the root system is developed from root apical meristems while the above ground shoot is derived from the shoot apical meristem, which produces leaves and axillary buds (Irish, 2010). At the initiation of flowering time, the shoot apical meristem permanently converts to a reproductive inflorescence meristem and in *A. thaliana*, floral meristems develop on the flank giving rise to flowers (Irish, 2010).

1.2.1 Pathways of flowering time

Initiation of flowering is dependent on both exogenous and endogenous cues (Srikanth and Schmid, 2011). Exogenous signals include detection of the photoperiod and temperature, while endogenous signals are dictated by the developmental state of the plant. The combination of exogenous and endogenous cues allows a plant to monitor its environment in order to flower at the most optimal times, namely spring and summer. Four molecular pathways that regulate timing of flowering in *A. thaliana* have been identified: the photoperiodic pathway, the autonomous pathway, the vernalization pathway, and the gibberellic-acid (GA) pathway (Moon et al., 2005; Srikanth and Schmid, 2011). In general, the photoperiodic and vernalization pathways respond to

environmental cues, while the autonomous and GA pathways monitor the developmental state of the plant, independent of the environment (Mouradov, 2002; Moon et al., 2005).

1.2.2 Photoperiod pathway

The photoperiod pathway monitors the duration and quality of the daily light period. Genes identified as part of this pathway regulate both flowering time as well as light transduction pathways or circadian clock function (Mouradov, 2002). *A. thaliana* is a facultative long-day plant with the gene *CONSTANS* (*CO*) playing a key role as the central regulator in integrating the photoperiodic pathway with flowering (Mouradov, 2002; Moon et al., 2005; Srikanth and Schmid, 2011).

1.2.3 Vernalization pathway

Vernalization is the process by which seeds or seedlings are subjected to cold temperatures, usually 1 to 7 ^{o}C, for 1 – 3 months, in order to promote development and flowering (Srikanth and Schmid, 2011). In contrast to summer annuals, which germinate and flower the same summer, winter annual varieties germinate in autumn or winter, grow vegetatively throughout the winter, and flower in the following spring or summer in response to daylight cues (Mouradov, 2002). Analysis of crosses between *A. thaliana* winter and summer annual varieties identified two loci required for vernalization, *FLOWERING LOCUS C* (*FLC*) and *FRIGIDA* (*FRI*) (Koornneef and Vries, 1994; Mouradov, 2002; Srikanth and Schmid, 2011). *FRI* upregulates *FLC* that encodes a MADS box transcription factor known to suppress flowering (Mouradov, 2002; Srikanth and Schmid, 2011). Cold temperatures during vernalization decrease levels of *FLC*

transcripts, which are continued to be epigenetically repressed after a return to warmer temperatures, thus allowing flowering to occur (Srikanth and Schmid, 2011).

1.2.4 GA pathway

The GA pathway was first discovered when gibberellins (gibberellic acids or GA) produced during a fungal infection caused rice seedlings to grow quickly (Srikanth and Schmid, 2011). Active GAs control a variety of processes ranging from development to promoting flowering (Olszewski et al., 2002). During flowering, GAs are crucial for the development of stamens and petals; GA mutants grow as sterile dwarfs with misshapen pistils and sepals (Olszewski et al., 2002; Mouradov, 2002; Srikanth and Schmid, 2011). Further studies also showed that the application of GA increased mRNA levels of both the floral meristem identity gene *LEAFY* (*LFY*) and *FLOWERING LOCUS T* (*FT*), both of which are major contributors to flowering as a long distance signal between leaves and shoot meristem (Mouradov, 2002; Srikanth and Schmid, 2011).

1.2.5 Autonomous pathway

The autonomous pathway was identified in a group of mutants characterized as being late flowering regardless of the photoperiod length (Mouradov, 2002; Srikanth and Schmid, 2011). Genes identified in the autonomous pathway, such as *FCA*, generally encode either chromatin remodelling factors or proteins involved in RNA processing (Doyle et al., 2005; Srikanth and Schmid, 2011). As all of these genes work to repress *FLC*, it has been concluded that the redundant genes of the autonomous pathway work in parallel with environmental cues to promote flowering (Komeda, 2004; Srikanth and Schmid, 2011).

1.2.6 Activation of floral pathway integrators

Although the four flowering pathways can independently promote or repress flowering, they form an interconnected network and all converge on common downstream target genes, the floral pathway integrators (Mouradov, 2002; Moon et al., 2005; Srikanth and Schmid, 2011). *FLC* integrates the autonomous and vernalization pathways; however, it also negatively regulates *FT* and *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*). *SOC1* is a positive flower regulator found at the shoot apical meristem. Both *FT* and *SOC1* interact with *CO* (*CONSTANS*) of the photoperiodic pathway (Srikanth and Schmid, 2011). The GA pathway works in parallel with the *CO* of the photoperiodic pathway and also regulates *SOC1* (Moon et al., 2005; Srikanth and Schmid, 2011). Furthermore, *LFY* is acted upon by the photoperiodic, autonomous, and GA pathways (Mouradov, 2002; Moon et al., 2005; Srikanth and Schmid, 2011).

Pathways are also dependent on each other: GAs are required by the autonomous pathway to induce flowering (Olszewski et al., 2002). While the crosstalk between the pathways is much more complicated, it can be concluded that all four flowering pathways, the floral repressing and the floral promoting pathways, converge on at least one of the key downstream pathway integrators *SOC1*, *FT*, or *LFY* (Figure 1.1) (Moon et al., 2005; Irish, 2010; Srikanth and Schmid, 2011).

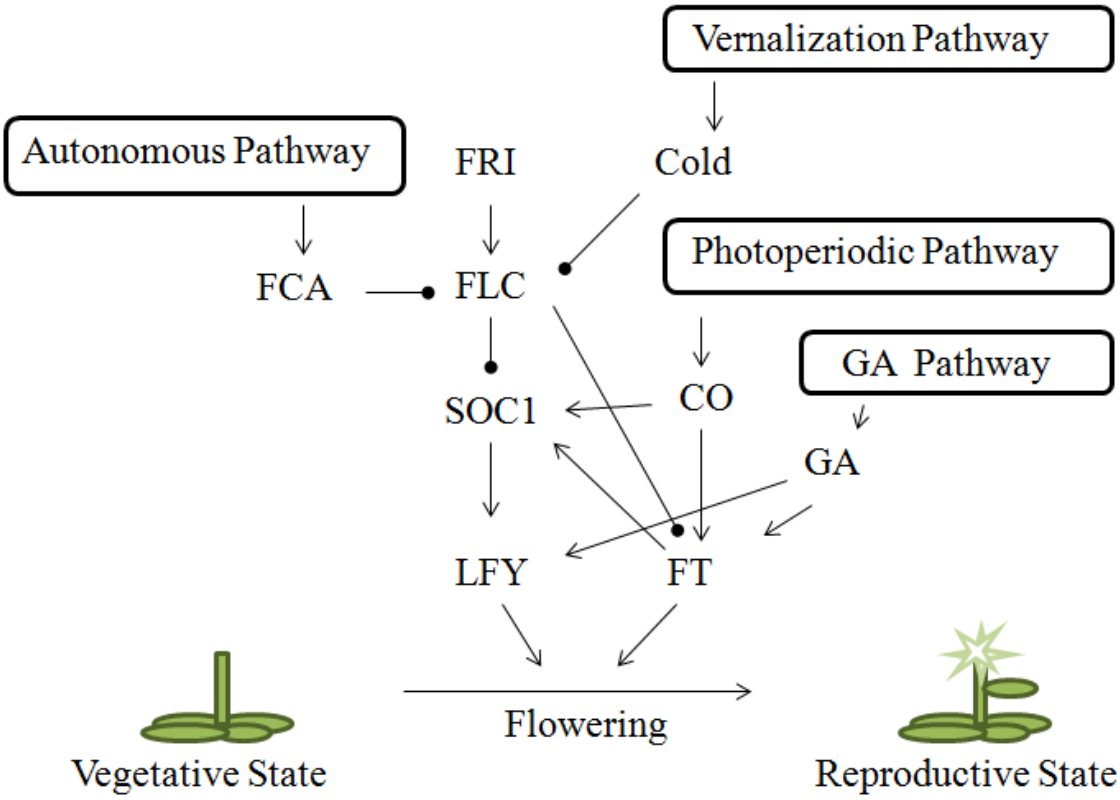

Figure 1.1: A simplified diagram of flowering pathways in *A. thaliana*.

This model does not present all genes involved in flowering time; however, it highlights the interactions of several key genes, the four flowering pathways and three prominent downstream integrators which activate meristem identity genes.

1.3 Floral development

Floral pathway integrators regulate floral meristem identity genes, including LFY, which in turn regulate floral organ identity (Weigel et al., 1992). The floral organ identity genes control the development of the sepals, petals, stamens, and carpels, which develop sequentially in four concentric rings, or whorls (Chen and Meyerowitz, 1999; Irish, 2010; Posé et al., 2012). These homeotic genes control floral development according to a model known as the ABCE model of flower development (Irish, 2010; Posé et al., 2012). A-class genes include *APETALA 1* (*AP1*) and *APETALA 2* (*AP2*) which direct sepal and petal development (Irish, 2010; Posé et al., 2012). B-class genes include *APETALA 3* (*AP3*) and PISTILLATA (*PI*), which establish petal and stamen identity (Irish, 2010; Posé et al., 2012). *AGAMOUS* (*AG*) is the only member of the C-class genes, which specifies the stamen and carpel development (Chen and Meyerowitz, 1999; Irish, 2010; Posé et al., 2012). The E-class genes, SEPALLATA genes 1 – 4 (*SEP1-4*), are co-regulators of the other three classes (Posé et al., 2012). A combination of the A-, B-, and C-class genes determines the development of organs in the whorls. In first whorl, A-class genes function alone, in the second whorl, both A- and B-classes function together, in the third whorl, B- and C-class genes are active, and in the fourth whorl, only C-class genes are present (Irish, 2010; Posé et al., 2012). E-class genes are found throughout all four whorls (Irish, 2010; Posé et al., 2012). Interestingly, A- and C-class genes mutually repress each other from the first and fourth whorls respectively (Alvarez and Smyth, 1999).

1.4 Identification of *HUA2*

Irish (2010) stated that *AG* as the most vital component in determinacy, flowering time, and promotion of the development of the floral meristem. *AG* also directs development of the stamen and carpel, the main reproductive organs of *A. thaliana*. In studies of a severe loss-of-function *ag* mutant, petals formed instead of stamens in the third whorl and another *ag*-like flower formed instead of carpels in the fourth whorl (Alvarez and Smyth, 1999; Chen and Meyerowitz, 1999). A subsequent mutagenesis experiment with a partial *AG*-loss of function strain of *A. thaliana*, *ag-4*, identified two additional genes that act with *AG*: *HUA1* and *HUA2* (Chen and Meyerowitz, 1999). Further screening for cofactors of the C-class genes identified the *HUA ENHANCER* (*HEN*) *1-5* group of genes which also contribute to floral organ identity and floral meristem determinacy (Cheng et al., 2003). These genes were found to regulate the same functions as *AG* and were proposed to be additional C-class genes. Recessive mutations in *HUA1* and *HUA2* enhanced the weak *ag-4* phenotype and led to stamen to petal transformation and indeterminacy of floral meristem (Chen and Meyerowitz, 1999). The *HUA1* and *HUA2* gene products also facilitate *AG* pre-mRNA processing (Chen and Meyerowitz, 1999; Cheng et al., 2003; Irish, 2010). However, it is also believed that the HUA2 and FRI pathways converge to activate FLC expression, repressing flowering (Poduska et al., 2003; Doyle et al., 2005; Wang et al., 2007). HUA2 also potentially interacts with FLOWERING LOCUS M (FLM/MAF1) and MAF2, both of which are related to FLC at the amino acid level and represses flowering in response environmental cues (Doyle et al., 2005).

1.5 Transcription of eukaryotic genes

A great number of proteins and factors involved in the flowering pathways and in floral organ development work through the regulation of other genes, turning the expression of their target genes on or off. In eukaryotes, gene expression follows Crick's central dogma of molecular biology, which can simply be stated as "DNA makes RNA makes protein". The first stage of this multistep process begins at transcription, in which mRNA is synthesized from DNA (Barberis and Petrascheck, 2003; Ma, 2011). Transcription can be divided into three classes based on the enzymes primarily responsible for the RNA synthesis: RNA polymerase I transcribes rRNA, RNA polymerase II generates protein coding mRNA and RNA polymerase III transcribes tRNA and other small RNA (Ptashne and Gann, 1997; Kornberg, 2005). While the three RNA polymerases are principally responsible for the transcription of DNA, they are unable to initiate transcription by themselves or even bind to DNA unaided (Ptashne and Gann, 1997; Barberis and Petrascheck, 2003; Ma, 2011). Instead, transcription initiation, and thus, gene transcription, is regulated by a number of *cis*-acting elements, promoter or enhancer sequences of DNA, as well as *trans*-acting elements, such as general transcription factors or co-activators (Ptashne and Gann, 1997; Barberis and Petrascheck, 2003; Kornberg, 2005).

1.5.1 Mechanisms of transcription

While a variety of *cis*- and *trans*-acting elements are associated with transcription, the basal transcription machinery only requires a core promoter where specific general transcription factors and RNA polymerase assemble (Ptashne and Gann, 1997; Barberis and Petrascheck, 2003). The core promoter is a set of short conserved DNA sequences that are typically found upstream of a gene's transcription start site (Ptashne and Gann, 1997; Ma, 2011). These elements are situated close to the transcription start site with different core promoters containing functionally similar, but not identical sets of elements; common elements are the TATA box, with the core DNA sequence of 5'-TATAAA-3', and the INR element, with the consensus 5'-YYAN(T/A)YY-3' (Ptashne and Gann, 1997; Kornberg, 2005; Xi et al., 2007). These *cis*-acting elements are recognized by *trans*-acting elements, referred to as general transcription factors (GTFs). GTFs are generally required at promoters used by RNA polymerase II and perform a variety of functions required for transcription initiation such as melting DNA and recruiting RNA polymerase II (Barberis and Petrascheck, 2003; Kornberg, 2005).

1.5.2 Basal transcription machinery

At the onset of transcription initiation with promoters for RNA polymerase II, the TATA-binding protein (TBP), a general transcription factor, recognizes and binds to the TATAAAA sequence of the TATA box (Figure 1.2A) (Barberis and Petrascheck, 2003; Ma, 2011). TBP forms with TBP-associated factors (TAFs) form a multi-subunit GTF called TFIID (Barberis and Petrascheck, 2003; Ma, 2011). These stepwise interactions continue as the complex associates with various GTFs, such as TFIIB, and the Mediator complex, which allows the growing complex to bind tightly with DNA at other promoter

elements, such as the INR element (Figure 1.2B) (Barberis and Petrascheck, 2003; Cantin et al., 2003; Kornberg, 2005). Through the TFIIF, another GTF, RNA polymerase II is finally brought to the promoter region and stabilised through TFIIB and the Mediator complex (Figure 1.2B) (Ptashne and Gann, 1997; Barberis and Petrascheck, 2003; Kornberg, 2005). The culminating product of several GTFs bound simultaneously to RNA polymerase and promoter regions in DNA is a complex known as the preinitiation complex (Green, 2005; Ma, 2011). The inactive preinitiation complex, or closed complex, positions the active site of RNA polymerase II over the transcription start site, ready for transcription (Green, 2005; Ma, 2011). As the preinitiation complex melts the DNA at the promoter through the multi-subunit GTF TFIIH, it transitions into a state called the open complex and transcription may begin (Figure 1.2C) (Ptashne, 1988; Kim, 2000; Barberis and Petrascheck, 2003). Once transcription begins, the complex is called the initial transcribing complex. After the initial transcribing complex produces a transcript of more than ten nucleotides, the complex is said to have escaped from the promoter; transcription initiation is complete and the elongation phase begins (Ptashne, 1988).

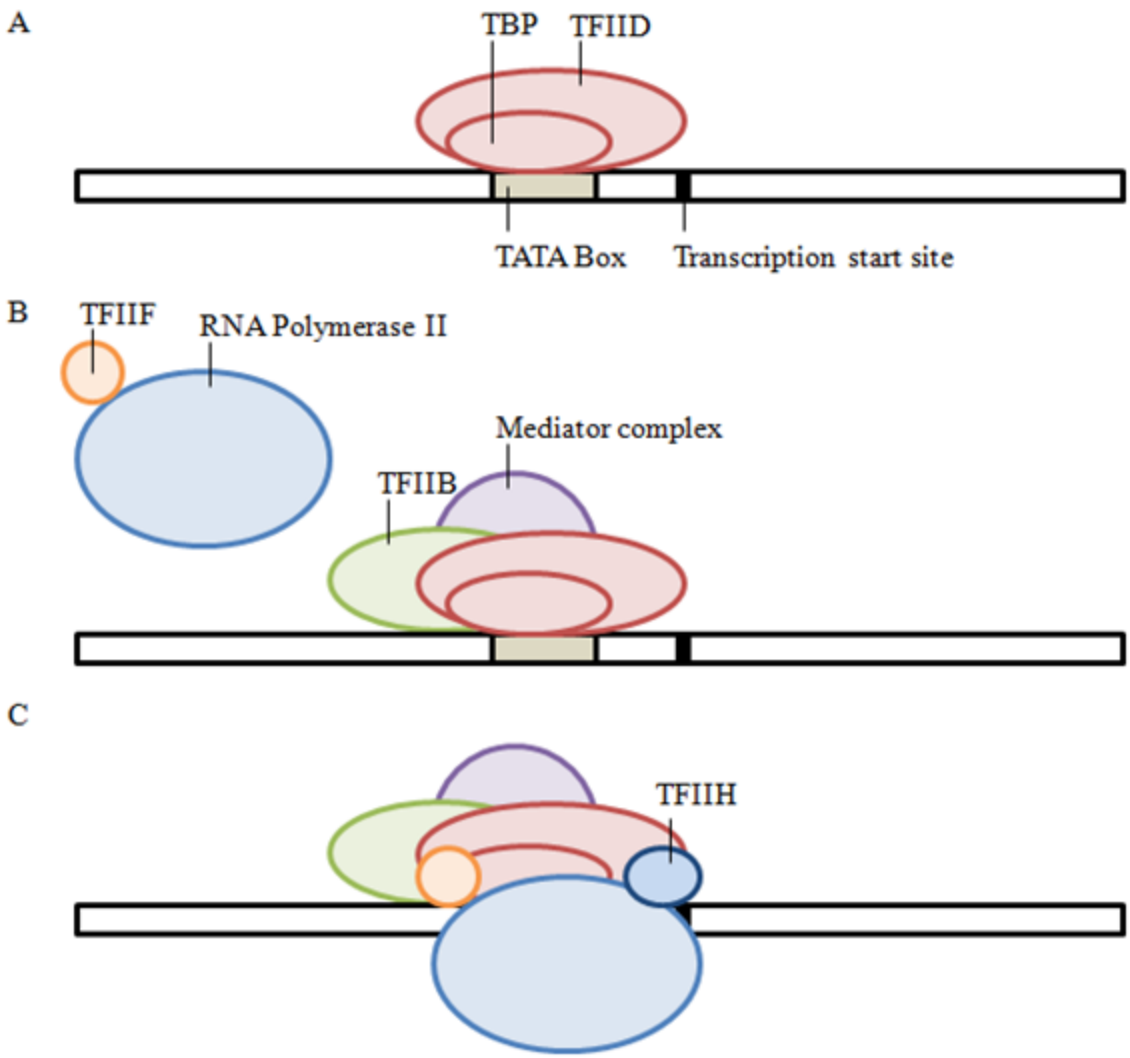

Figure 1.2: Simplified schematic of the formation of preinitiation complex during transcription initiation by RNA Polymerase II.

A) The GTF TFIID binds to the TATA box at the promoter region through the TBP subunit. B) TFIID provides a platform for which other GTFs, such as TFIIB, and the Mediator complex may bind to the DNA. TFIIF is recruited to the promoter region together with RNA Polymerase II. C) TFIIH is one of the last proteins recruited to the preinitiation complex and is responsible for unwinding the promoter DNA to allow transcription to begin (Ptashne and Gann, 1997; Kornberg, 2005; Barberis and Petrascheck, 2003).

1.6 Transcription factors

With just the core promoter and the basal transcription machinery active, transcription occurs at a low level, at what is known as a basal rate or constitutive transcription (Barberis and Petrascheck, 2003). This is due to the infrequent and unstable binding of unaided RNA polymerases to promoters during transcription initiation (Ptashne and Gann, 1997). Since the process of transcription initiation is a relatively difficult step, it stands to reason that it is a rate-limiting step in gene transcription. Therefore, it is also an ideal point for gene regulation; gene transcription can be turned "on", or "off" (Mitchell and Tjian, 1989; Ma, 2011). In the "off" state, gene transcription can be practically non-existent at a basal rate or suppressed (Ma, 2011). In the "on" state, high levels of transcription can be achieved through a number of other *cis*- and *trans*-acting elements that are required for accurate and efficient binding of RNA polymerase to promoters (Biddick and Young, 2005). The c*is*-acting elements can be divided into promoters and upstream activation sequences (UAS), in yeast, or enhancers, in higher eukaryotes (Ma, 2011). Whereas promoters act as binding sites for GTFs, enhancers are clusters of DNA sequences that are bound by regulatory proteins such as transcription factors (Barberis and Petrascheck, 2003; Kadonaga, 2004; Ma, 2011). Transcription factors are one type of *trans*-acting elements that can either activate gene transcription or may increase the rate of transcription of genes that are transcribed at a low basal rate (Biddick and Young, 2005; Ma, 2011).

1.6.1 Properties of transcription factors

Transcription factors were initially studied in yeast, but have been found to be completely functional when expressed in non-native cells such as those of plants, insects, and mammals (Ptashne, 1988; Ptashne and Gann, 1997). Typical transcriptional activators minimally contain a DNA-binding domain (DBD) and at least one transcriptional activation domain (TAD) (Ptashne, 1988; Biddick and Young, 2005; Titz et al., 2006). The DBD recognizes enhancers, specific sites that may be adjacent to genes they regulate or even several thousand base pairs up- or down-stream of the transcription start site (Kornberg, 2005; Biddick and Young, 2005; Titz et al., 2006). It provides gene-targeting specificity and localizes the transcription factor to the promoter of the gene to be regulated. DBDs have been well characterized both functionally and structurally; the different DBD families can be categorized based on the distinct three-dimensional structure they form (Kadonaga, 1990; Titz et al., 2006; Ma, 2011). These structures include zinc fingers, zinc clusters, leucine zippers and the helix-turn-helix motif (Mitchell and Tjian, 1989; Garvie and Wolberger, 2001; Kadonaga, 2004). It is important to note that DBDs do not play a role in controlling the levels of transcription to which a transcription factor upregulates a particular gene (Ptashne, 1988). In fact, transcription factors do not always have a DBD, such as in herpes simplex virus transcription factor VP16 (Sadowski et al., 1988). TADs may also be separated from a DBD and continue to function as a chimeric transcription factor, upregulating a different gene when fused with a non-native DBD (Ptashne and Gann, 1997; Titz et al., 2006; Ma, 2011). These observations suggest that transcriptional activation occurs solely through

TADs and that the mechanism of gene upregulation is similar across eukaryotes while specificity is dictated by the DBD (Ptashne and Gann, 1997; Ma, 2011).

1.6.2 Properties of transcriptional activation domains (TADs)

TADs mediate transcription through one of two processes (Cantin et al., 2003; Green, 2005). Some TADs function by recruitment of enzymes that affect chromatin structure (Cantin et al., 2003; Kornberg, 2005; Biddick and Young, 2005). Eukaryotic DNA is packaged into nucleosomes, a structure in which DNA is wrapped around a histone core that blocks interactions of the basal transcription machinery with DNA (Ma, 2011). However, chromatin remodelling complexes recruited by TADs can either modify histones to loosen the histone-DNA interaction or unpack nucleosomes to open promoters that may otherwise be inaccessible (Cantin et al., 2003; Kornberg, 2005; Ma, 2011). The exposed DNA is then considered to be in an active chromatin state poised for transcription. The other process involves TADs recruiting members of the basal transcription machinery and coactivators to promoters (Ptashne and Gann, 1997; Kornberg, 2005; Ma, 2011). A variety of yeast and mammalian TADs directly interact with TATA-binding protein and at least two GTFs that are parts of the RNA polymerase II preinitiation complex, TFIIB and TFIIH (Blau et al., 1996; Ptashne and Gann, 1997; Piskacek et al., 2007). Coactivators are a variety of *trans*-acting elements that are involved in various steps of transcription, but are not essential to the basal transcription machinery. However, some coactivators, such as the Mediator complex, can aid in the formation of the preinitiation complex by acting as bridge between transcription factors or other regulatory proteins and the basal transcription machinery (Cantin et al., 2003; Kadonaga, 2004; Ma, 2011). The increased concentration of subunits of the basal

transcription machinery and coactivators combined with their proximity to promoters can have a synergistic effect on recruitment and cooperative binding of other components and accelerates the formation of a stable preinitiation complex (Blau et al., 1996; Barberis and Petrascheck, 2003; Ma, 2011). Thus, transcription factors most likely increase transcription levels through recruitment of other proteins and complexes to promoter regions (Ma, 2011; Lin et al., 2010).

While DNA-binding domains (DBDs) are well defined and their mechanism of function is known, much less is known regarding TADs. Other than requiring hydrophobic residues, TADs do not share easily recognizable motifs or structures; they tend to be short protein sequences, with as few as nine amino acids, with low complexity (Mitchell and Tjian, 1989; Titz et al., 2006; Ma, 2011). Between transcription factors, there is little or no sequence conservation, and so it is difficult to predict a TAD based on sequence alone (Garvie and Wolberger, 2001; Piskacek et al., 2007; Lin et al., 2010). As a result, TADs are classified by their amino acid sequence composition. Most TADs can be assigned to three well established classes: acidic activation domains, glutamine-rich activation domains and proline-rich activation domains (Ptashne, 1988; Barberis and Petrascheck, 2003; Titz et al., 2006).

1.6.3 Acidic activation domains

Acidic activation domains were the first to be discovered through the yeast transcription factor *GAL4* (Sadowski et al., 1988). As the name suggests, these domains are predominantly aspartic and glutamic acid residues, but members of the class are known to carry a significant negative charge and form amphiphatic α-helical structures when bound to target proteins (Mitchell and Tjian, 1989; Sullivan et al., 1998; Tell et al.,

1998). The minimal active sequence required of acidic TAD has been defined as short segments of acidic residues, and so, while *GAL4* was found to have two separate acidic domains of approximately 100 amino acids each, with no sequence homology, the minimal sequence was mapped to 17 residues (Hermann et al., 2001). Many TADs in this class stimulate transcription initiation by interacting with the preinitiation complex through binding with various subunits including TBP and TFIIB (Tell et al., 1998; Hermann et al., 2001; Piskacek et al., 2007). The mechanism of activation seems to be conserved as a number of well-known TADs have been observed to function in yeast, mammalian, and plant cells (Remacle et al., 1997). However, studies of a number of acidic TADs have shown that while overall acidity is important, the presence of interspersed hydrophobic amino acids is more important (Sullivan et al., 1998; Hermann et al., 2001).

1.6.4 Glutamine-rich activation domains

The glutamine-rich activation domain has been widely studied in the transcription factor Sp1. Sp1 has been described as containing two short stretches of glutamine residues, but unfortunately, no sequence homology exists between the two domains (Bouwman and Philipsen, 2002). Short sequences of two to four residues, interspersed with hydrophobic amino acids, that are able to stimulate transcription are common in other glutamine-rich TADs (Mitchell and Tjian, 1989; Künzler et al., 1994; Mayr et al., 2005). Glutamine-rich TADs tend to activate transcription from proximal enhancers, receiving cues from remote enhancers (Remacle et al., 1997; Bouwman and Philipsen, 2002; Gehring and Henikoff, 2008). As well, the target of glutamine-rich TADs is strongly suspected to be a member of the preinitiation complex; however, unlike acidic

TAD, the mechanism of activation does not seem to be conserved between different organisms; glutamine-rich TADs of higher eukaryotes are unable to activate transcription in yeast cells (Emili et al., 1994; Remacle et al., 1997; Escher et al., 2000). The incompatibility of yeast glutamine-rich TADs in yeast cells is due to the fact that the regulatory sequences controlling gene expression in yeast differs from that of other eukaryotes and highlights the dependency of glutamine-rich activation domains on cues from other enhancers (Escher et al., 2000).

1.6.5 Proline-rich activation domains

Proline-rich activation domains were first identified in the carboxyl-terminus of the transcription factor CTF/NFI (Mitchell and Tjian, 1989; Prado et al., 2002; Lin et al., 2010). While there is a lack of significant sequence homology between TADs, the addition of tracts of proline or glutamine residues to TADs can result in enhanced transcription rates (Gerber et al., 1994; Remacle et al., 1997; Schwechheimer et al., 1998). Similar to glutamine-rich TADs, proline-rich TADs tend to act from promoter proximal enhancers in conjunction with cues from remote enhancers (Gerber et al., 1994). The properties of polyproline sequences result in their tendency to form α-helix structures; polyproline stretches present a large hydrophobic surface with a good hydrogen-binding site (Kay et al., 2000). The proline-rich regions are exposed, allowing for a great number of protein-protein interactions, potentially some with factors involved in transcription initiation (Li, 1999; Kay et al., 2000). Furthermore, the exposed binding sites tend to have quick on- and off-rates of binding and can be found in interactions requiring the quick recruitment of several proteins, such as during transcription initiation (Kay et al., 2000).

1.7 *FRI*: A plant transcription factor

Transcription factors are found in all types of cells and organisms, and many floral organ identity genes, previously mentioned, encode MADS domain transcription factors (Chen and Meyerowitz, 1999; Irish, 2010). A well characterized transcription factor in the flowering pathways of *A. thaliana* is FRIGIDA (FRI). FRI upregulates transcription rates of *FLC*, a repressor of floral pathway integrator genes *FT* and *SOC1*. While several nonspecific regulators were identified in FRI-mediated *FLC* regulation, six *FLC*-specific regulators were identified. These are FRI, FRIGIDA LIKE1 (FRL1), FRL2, FRIGIDA ESSENTIAL1 (FES1), SUF4 AND FLC EXPRESSOR (FLX) (Choi et al., 2011). The regulators act as components of a FRI-containing protein complex (FRI-C) with FRI itself also acting as a scaffold and FES1 and FRL1 stabilizing the complex (Choi et al., 2011). FRI-C utilizes SUF4 to bind to DNA, bringing the complex to the *FLC* promoter region (Choi et al., 2011). To activate transcription, FRI-C recruits several chromatin modification factors, such as the SWR1 complex and EFS, both of which catalyze chromatin modification leading to active transcription (Choi et al., 2011). FRI-C also attracts a range of general transcription factors such as TAF14, a component of both TFIID and TFIIF, and RNA polymerase III itself (Choi et al., 2011).

1.8 *HULK* gene family

Another putative transcription factor that regulates the expression of *FLC* is HUA2. *HUA2* was identified as a member of a family of four plant specific genes, the *HUA2-LIKE* (*HULK*) gene family: *HUA2, HULK1*, *HULK2* and *HULK3* (Challa, 2009). While the exact function of these genes is not known, studies with loss-of-function mutants revealed that they are redundant in their function, and that they are essential for development in *A. thaliana* (Challa, 2009). HULK protein family members share conserved domain structures and as a result share a high amino acid sequence similarity (Challa, 2009).

1.8.1 Structure of the HULK protein family

The four protein domains found in the HULK proteins are the PWWP domain, nuclear localization signals (NLS), a regulation of nuclear pre-mRNA processing (RPR) domain, and a putative proline-rich region (PRR) (Figure 1.3) (Challa, 2009). PWWP domains are involved in histone binding and also interact with chromatin-associated factors (Slater et al., 2003; Wang et al., 2009). HUA2 has four NLS domains (Challa, 2009). The NLS domain's ability to localize the protein to the nucleus has been demonstrated in our lab for all four HULK members (Janakirama and Grbic, unpublished). The RPR domain is present in proteins that interact with the carboxyl-terminal domain (CTD) of RNA Polymerase II (Doerks et al., 2002). This function is accomplished specifically through the DSI motif (aspartic acid (D), serine (S), and isoleucine (I)) and work in our lab has confirmed that mutations within this motif disrupt the ability of HUA2 to restore a late-flowering phenotype to an early-

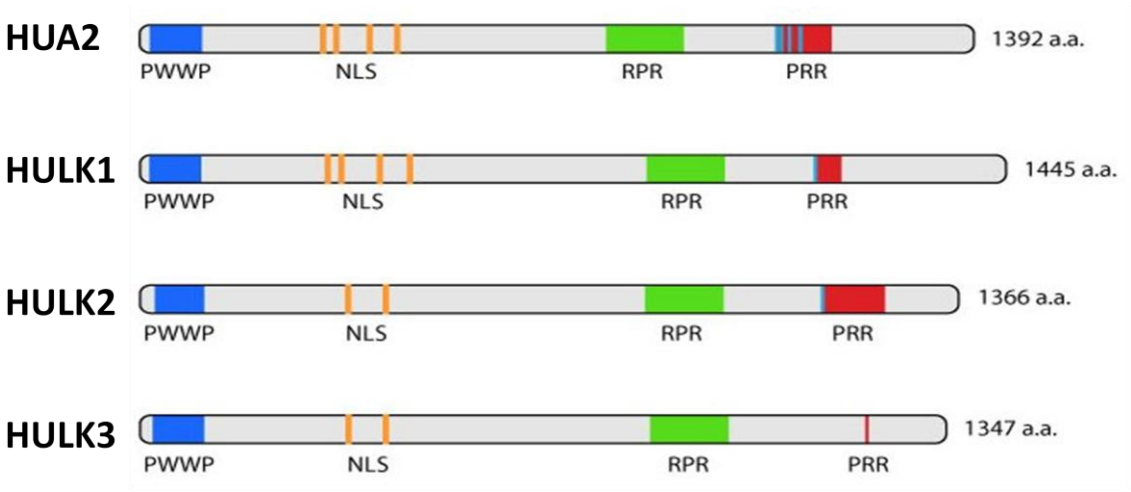

Figure 1.3: A schematic representing the HULK protein family.

Various structures of the HULK proteins are represented. The PWWP domain is shown in dark blue near the amine terminus. Each HULK carries a number of nuclear localization signals (orange) as well as an RPR domain (green). In the carboxyl-terminus, the PRR is highlighted in red with PPLP represented by the light blue boxes (Sajja, 2009).

flowering *hua2* mutant (Sajja, 2009). The PRR is found towards the carboxyl-terminus of the HULKs and is rich in proline residues. With the exception of HULK3, other HULKs also contain at least one proline-proline-leucine-proline (PPLP) repeat within the PRR (Challa, 2009). PPLP repeats are known to be involved in a variety of protein-protein interactions (Bedford and Leder, 1999; Macias et al., 2002).

In addition to the four previous structures, Chen and Meyerowitz (1999) have highlighted two additional putative domains which flank the PRR. They suggest that within the carboxyl-terminus of HUA2, both an acidic residue rich domain (amino acids 964 – 1018) and a proline/serine/asparagine residue rich domain (amino acids 1129 – 1392) flank either side of the PRR (amino acids 1056 – 1128), both of which could be putative transcriptional activation domains (Chen and Meyerowitz, 1999).

1.8.2 The HULK protein family: Putative plant transcription factors

Chen and Meyerowitz (1999) predicted that HUA2 could be a transcription factor based on sequence analysis and the presence of acidic residue rich, proline residue rich, and proline/serine/asparagine residue rich domains. Further, Doyle et al. (2005), Poduska et al. (2003), and Wang et al., (2007) concluded that HUA2 was required for the enhanced expression of several late-flowering genes: *FLC*, *FLM* and *MAF2*. The NLS domain localizes the protein to the nucleus where transcription occurs (Janakirama and Grbic, unpublished). HUA2 has the PPLP repeats that could allow it to interact with members of the basal transcription machinery or cofactors. Tracts of prolines, such as those found in the PRR, increase rates of transcription (Schwechheimer et al., 1998; Gerber et al., 1994; Lin et al., 2010). Taken together, these observations provide

circumstantial evidence that HUA2 and other members of the HULK family may be involved in transcriptional regulation.

1.9 Principles of yeast-two hybrid assays

The common method to test a protein's ability to act as transcriptional regulator is the yeast-one hybrid assays. Yeast-one hybrid assays are modified from the more common yeast-two hybrid assays (Uetz et al., 2000; Chen and DenBoer, 2008). The yeast-two hybrids are mainly used today to identify protein-protein interactions as well as protein-DNA interactions, but the principles of the assay also lends its ability to be used in the study of transcription factors. Early in its inception, it was used to study the yeast transcription factor GAL4 (Ma and Ptashne, 1987; Mermod et al., 1989). Fields and Song (1989) capitalized on the modularity of the DNA-binding domain (DBD) and activation domain (AD) of transcription factors and performed the first yeast-two hybrid using two known interactors, SNF1 and SNF4.

Both SNF1 and SNF4 are yeast proteins that bind physically to each other. The amino-terminus of GAL4 is able to bind to the upstream activation sequence (UAS) of the *GAL1-lacZ* fusion gene while the carboxyl-terminus contains the potent AD (Figure 1.4A) (Fields and Song, 1989). In Field and Song's study, SNF1 was translationally fused to the GAL4 DBD (the bait) and SNF4 was translationally fused to the GAL4 AD (the prey). Both were expressed in a yeast strain missing the *GAL4* gene and with the *GAL1-lacZ* UAS, the target for the GAL4 DBD, integrated at the *URA3* locus (Fields and Song, 1989). In this system, the GAL4-SNF4 chimeric protein binds to the UAS, and since SNF1 physically interacts with SNF4, it brings the GAL4-SNF1 into the proximity

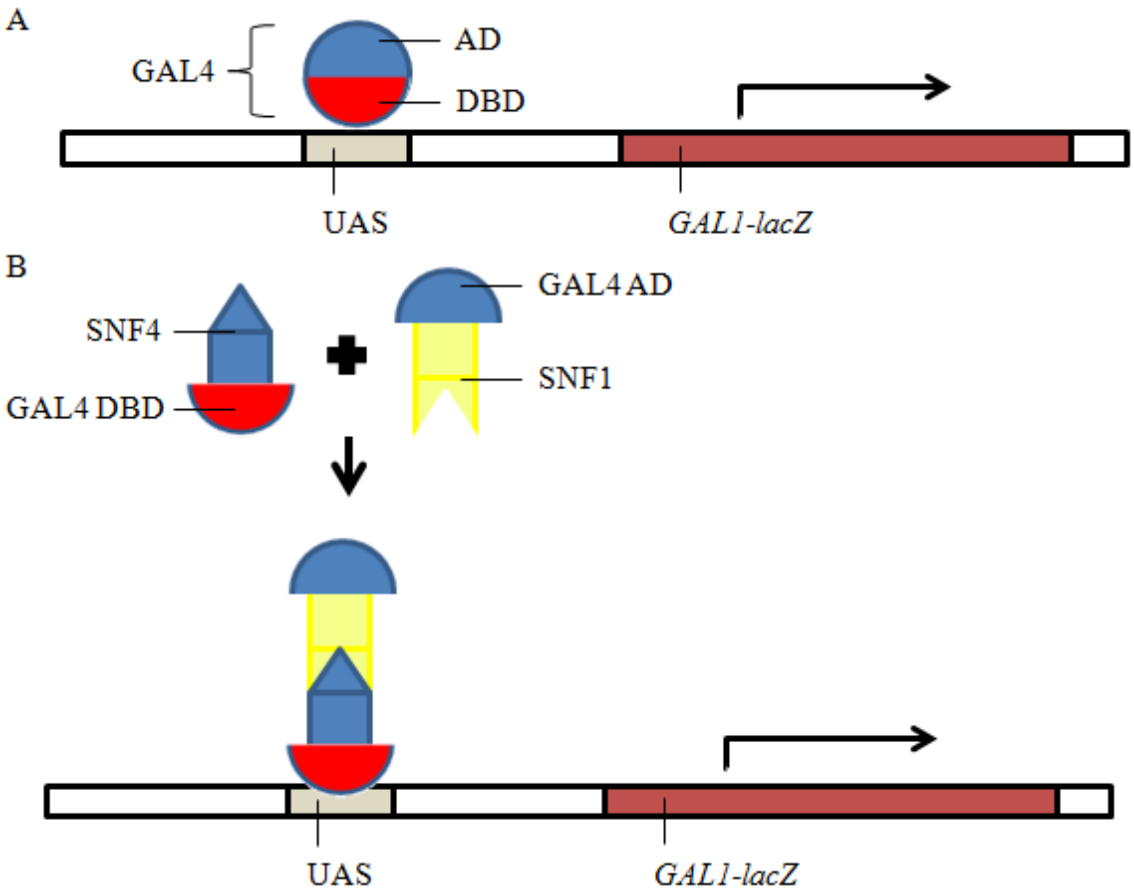

Figure 1.4: Schematic of transcriptional activation by reconstituting GAL4 activity.

A) GAL4 protein binds to the UAS of *GAL1-lacZ* through a DBD. The GAL4 AD induces production of β-galactosidase, which in the presence of X-gal results in the formation of blue yeast colonies. B) SNF4 is translationally fused to GAL4 DBD and SNF1 is translationally fused to GAL4 AD. Both are expressed in the same yeast colonies and the protein-protein interactions between SNF1 and SNF4 brings the GAL4 domains into close proximity of the *GAL1-lacZ* promoter and activate transcription.

of the *GAL1-lacZ* promoter (Figure 1.4B) (Fields and Song, 1989) (Fields and Song, 1989). This results in the induced production of β-galactosidase, which in the presence of X-gal can be visualized as blue yeast colonies. Thus, two proteins not regularly involved in the process of transcription can be tested together as putative interactors if they are able to reconstitute the transcription factor and activate a reporter gene. These experiments also highlighted the modularity of ADs: ADs able to be separated from their native proteins and recombined as functional chimeric proteins.

Researchers then began to investigate ADs that are not native to yeast while using the yeast two-hybrid system. The potent activation domain of the herpes simplex virus protein, VP16, has been well studied within yeast in conjunction with GAL4 DBD (Sadowski et al., 1988; Schwechheimer et al., 1998). A study with the tomato Myb-like activator, THM18, also performed in a yeast two-hybrid system with the DBD of GAL4 indicated the ability of chimeric transcription factors to be interchangeable between higher eukaryotes and yeast (Schwechheimer et al., 1998).

Thus, the yeast two-hybrid assay is used to examine protein-protein, the basis of the assay are rooted in transcriptional activity.

1.10 Yeast one-hybrid assays as a test of activation strength

Modified from the yeast two-hybrid assay, the yeast one-hybrid removes the need of a prey protein. Originally developed to study protein-DNA interactions, it has a single bait protein, the AD-fusion protein with a putative DBD that targets a specific DNA sequence cloned upstream of a reporter gene (Titz et al., 2006). If the putative DBD did associate with the specified DNA sequence, the fusion protein would function as an autoactivator, able to trigger transcription by itself (Uetz et al., 2000; Titz et al., 2006).

Yeast one-hybrid assays were then also applied to measuring the strength of putative transcription factors, known as the activation strength (Chen and Meyerowitz, 1999; Titz et al., 2006). Instead of using a putative DBD fused to a known AD, a known DBD was used to target the UAS of a reporter gene and the fused putative AD was then measured for its ability to upregulate expression of the reporter gene (Van Criekinge and Beyaert, 1999). While this was usually an observable marker, the use of certain reporter genes allowed for quantification of the activation strength.

Qualitatively, activation strength could be measured with the use of 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of histidine biosynthesis, and the reporter gene *HIS3*. In a yeast one-hybrid assay, the strength of the activation strength directly influences the level of expression of the *HIS3* reporter gene. However, since *HIS3* is a rather sensitive reporter, it has a leaky expression, and even the smallest level of expression is detected (Gietz et al., 1997; Van Criekinge and Beyaert, 1999). Thus, 3-AT is typically used to quench the basal level expression of the gene before testing bait-prey interactions.

By titration with 3-AT in yeast one-hybrid assays, a point will be achieved at which the ability of the autoactivator to upregulate the *HIS3* gene will be insufficient for yeast cultures to survive due to the inhibitory nature of 3-AT. The lowest concentration of 3-AT that inhibits growth can be considered the activation strength of the TAD tested. In 2006, Titz et al. using this principle performed yeast-one hybrid assays in the yeast strain YULH. Proteins classified as weak autoactivators allowed yeast growth in the 3 – 25 mM 3-AT range; medium strength activation domains are capable of inducing

transcription in the 50 – 200 mM 3-AT range; and strong auto-activators induce transcription at 3-AT concentrations greater than 200 mM (Titz et al., 2006).

1.11 Hypothesis

I hypothesize that the HULK family proteins have a transcriptional activation domain and that they will be able to activate reporter genes in yeast.

1.12 Objectives

1) To determine whether HULKS contain an activation domain;
2) To determine whether PRR/PPLPs are responsible for transcriptional activation

Chapter 2 : Materials and Methods

Different methods exist for scanning transcriptional activity, but the simplest and most common techniques involve the expression and upregulation of a non-native reporter gene in a yeast system. In my study, yeast one-hybrid assays were used to investigate the putative transcriptional activation domains (TADs) in the HULK protein family as well as to measure their strength.

2.1 Testing transcriptional activity of the HULK protein family

For the purpose of studying the HULK protein family as transcription factors, yeast-one hybrid assays were performed to measure their activation strength. Colonies of the *S. cerevisiae* yeast strain MaV203 (MATα, *leu*2-3,112, *trp*1-901, *his*3Δ200, *ade*2-101, *gal*4Δ, *gal*80Δ, *SPAL*10*::URA*3, *GAL*1::*lacZ*, *HIS*3$_{\text{UAS GAL1}}$*::HIS*3*@LYS*2, *can*1R, *cyh*2R) (Walhout and Vidal, 1999) were grown on a selective medium lacking leucine and histidine. Alone, the yeast itself is unable to synthesize leucine and histidine and would not survive on the medium. Thus, the culture was transformed with the plasmid pDBLeu. For the experiment, pDBLeu carried the GAL4 DBD-putative transcription factor construct as well as the reporter gene *LEU2*. *LEU2* encodes beta-isopropylmalate dehydrogenase, a protein involved in the third step of leucine biosynthesis in yeast (Brisco and Kohlhaw, 1990). The presence of this gene lends MaV203 the ability to biosynthesize leucine. To allow the biosynthesis of histidine, the GAL4 DBD must upregulate the *HIS3* gene, an auxotrophic marker found specifically in yeast two-hybrid strains, such as MaV203 (Invitrogen, 2005). *HIS3* encodes imidazoleglycerol-phosphate dehydratase, a protein that catalyzes the sixth step of histidine biosynthesis (Titz et al.,

2006). Furthermore, the yeast strain MaV303 has been specifically designed to include regulatory regions that contain DNA binding sites (UAS) for the GAL4 DBD (Invitrogen, 2005).

The constructs to be carried by pDBLeu and assayed are the full length HULK family proteins. Once *S. cerevisiae* MaV203 has been transformed with the plasmids, the proteins expressed will be translationally fused to the GAL4 DBD.

2.1.1 Testing for transcriptional activity in the carboxyl-terminus of the HULK protein family

Since the PRR is found at the carboxyl-terminal end, constructs of the carboxyl-terminal ends of the HULK family proteins (CT-HUA2, CT-HULK1-3) translationally fused to the GAL4 DBD were made and subsequently expressed in the yeast strain MaV203 and tested for 3-AT titration.

2.2 Procedures for 3-AT titration for full-length and carboxyl-terminus ends of HULK gene family

2.2.1 Subcloning in pDBLeu

Plasmids carrying the *HULK* gene family were acquired from lab stocks. The constructs obtained were of full length *HUA2* (4179 base pairs), *HULK2* (4107 bp) and *HULK3* (4050 bp) genes and the carboxyl-terminus ends of the *HULK* gene family in the vector pEXP-AD. The carboxyl-terminal ends of the *HULK* gene family are as follows: CT-HUA2 (1182 bp, nucleotides 2997 - 4179), CT-HULK2 (903 bp, nucleotides 3204 –

4107) and CT-HULK3 (930 a.a., nucleotides 3120 – 4050) in the vector pEXP-AD. The plasmids were kept in *E.coli* DH5α glycerol stocks.

A small sample, 5 µl, of each glycerol stock was grown in liquid LB (plus kanamycin 100 µg/ml) overnight at 30 ^{o}C. Plasmids were harvested from 3 ml of each overnight culture through a QIAprep Spin Miniprep Kit (QIAGEN). A small sample of the product of each miniprep was checked on a 1% agarose gel after restriction enzyme digestion: vectors with the carboxyl-terminus ends of the *HULK* gene family were digested with *Hin*dIII while vectors with the full length genes were digested with *Ava*I, *Hin*dIII, or *Xho*I. The reactions were kept at 37 ^{o}C overnight.

For subcloning into the vector pDBLeu, pEXP-AD containing either full length HUA2, HULK2, or HULK3, or CT-HUA2, CT-HULK2, or CT-HULK3 was then digested with *Sal*I for the excision at the N-terminus and *Not*I at the carboxyl-terminus. Simultaneously, pDBLeu was digested with *Sal*I and *Not*I. *Sal*I and *Not*I were used as the HULKs were previously cloned into pEXP-AD. Again, these digestions were kept at 37 ^{o}C overnight. The digestion products were then loaded onto a 0.8% agarose gel and run at a low voltage to separate the plasmids and genes. Using a UV transilluminator to mark a guide, a small sample of the digestion products run on the side, the bands of full length and carboxyl-terminus of the *HULK* gene family were cut out of the gel using a razor blade. The linearised pDBLeu was also cut out from the gel and all the samples were gel eluted using GeneJET Gel Extraction Kit (Fermentas).

For ligation of the genes into pDBLeu, the linearised vector and either full-length or carboxyl-terminus of the *HULK* genes were combined (3:1 molar ratio of vector and insert) and incubated overnight at 16 ^{o}C with T4 DNA ligase (Fermentas).

2.2.2 *E.coli* transformation

Following the ligation of the *HULK*s into pDBLeu, the product was purified with a GeneJET Gel Extraction Kit (Fermentas) and used to transform chemi-competent DH5α *E.coli* cells. Competent DH5α cells (200 µl) were placed in a 1.5 ml tube with 2 µl of plasmid for each construct and chilled on ice for 10 minutes. Cells were then heat shocked for 90 seconds at 42 ^{o}C with gentle shaking. Subsequently, the tubes were placed on ice for 2 minutes and 800 µl of liquid LB was added to each tube. Tubes were then incubated for 60 minutes at 37 ^{o}C and were centrifuged at 7000 RPM for 3 minutes. The supernatant (800 µl) was removed and the rest was used to resuspend the pellet. The resuspended transformation mixture (100 µl) was spread onto an LB agar plate with kanamycin (100 µg/ml). The plates were incubated at 37 ^{o}C overnight. The following day, colonies representing successful transformants, were picked, and transferred to liquid LB (plus kanamycin, 100 µg/ml), and incubated overnight at 37 ^{o}C. Plasmids were then harvested using a QIAprep Spin Miniprep Kit (QIAGEN).

2.2.3 Yeast transformation

The constructs in pDBLeu recovered from the DH5α transformation were then used to transform yeast *S. cerevisiae*, strain MaV203. Colonies of yeast from a YPD plate were used to inoculate 2X YPAD broth and were grown for 3 days in a shaker (200 rpm) at 30 ^{o}C. On the third day, the transformation mix was prepared. A tube of SS-carrier DNA (2 mg/ml) was placed in a boiling water bath for 5 minutes and quickly chilled in ice. While the SS-carrier DNA was cooling, 3 ml of MaV203 cultures was centrifuged twice (13,000 rpm, 1 minute each time) and the supernatant was removed.

To a MaV203 pellet, 240 µl of PEG 3500 50% w/v was added, followed by 36 µl of LiAC (1 M) and 50 µl of the boiled SS-carrier DNA. Finally, 8 µl of plasmid was added along with 26 µl of water. The MaV203 pellet was resuspended in this transformation mix. Each tube was then incubated in a water bath at 42 ^{o}C for 2 hours. After incubation, the tubes were centrifuged for 30 seconds and the supernatant removed. Sterile double distilled water (200 µl) was used to resuspend each pellet. The resuspended cells (50 µl) were then spread on YNB –leucine agar and allowed to grow for 2 days at 30 ^{o}C.

Three transformant colonies of each construct were then inoculated in separate falcon tubes of liquid YNB –leucine overnight. The following morning, to determine concentration and growth, the optical density of the overnight cultures was measured using a spectrophotometer set at a wavelength of 600 nm. Based on the OD_{600} readings, inocula from the overnight cultures were diluted to a uniform OD_{600} reading of 0.5 in separate 1.5 ml microcentrifuge tubes; 3 tubes for each construct.

2.2.4 3-AT titration

Plates were made using YNB -leucine -histidine medium, glucose and various amounts of 3-AT on gridded square Petri dishes. Concentrations of 3-AT on the plates were made to range from 0 mM to 300 mM. To perform 3-AT assays, 1 µl of diluted overnight culture from each microcentrifuge tube was spotted in separate 13 mm × 13 mm squares on each plate. Each construct was spotted 6 times, 1 µl in each square. Plates were then wrapped in plastic paraffin film and incubated at 30 ^{o}C for 4 days, after which images of the plates were obtained with a scanner. The lowest concentration of 3-AT which was considered to inhibit growth was when none of the spotted colonies grew.

2.3 Checking 3-AT titration for domain swap of HULK2's and HULK3's PRR

In order to synthesize the domain-swapped proteins, CTHULK3hulk2PRR (CT-HULK3 with its PRR swapped for CT-HULK2's PRR) and CTHULK2hulk3PRR (CT-HULK2 with its PRR swapped for CT-HULK3's PRR), the method of overlapping extension-PCR (OE-PCR) was used. An OE-PCR is a variant of PCR and can be used to insert mutations or synthesize a gene. For purposes of constructing the domain-swapped genes, OE-PCR was used to ligate PCR synthesized fragments of the PRRs and fragments of the carboxyl-terminus of either *HULK2* or *HULK3*. The primers used to generate those fragments extend beyond fragments, creating an overhang that overlaps with and compliments the specific sequences of the other fragments. This allows one PCR reaction to ligate and amplify the domain-swapped genes. Both chimeric constructs were produced in the same fashion, but the synthesis of CTHULK3hulk2PRR will be described more thoroughly.

The proline rich region (PRR) sequences in both *HULK2* and *HULK3*, which were initially identified by a previous lab member, Sathya Challa, were located in the full length CDS nucleotide sequences obtained from the TAIR (The Arabidopsis Information Resource) database. Three fragments to be synthesized were then identified; the fragment of *HULK3*'s carboxyl-terminus before *HULK3*'s PRR (*HULK3* pre-PRR, nucleotides 1 to 237), the PRR of *HULK2* (*HULK2* PRR, nucleotides 171 to 297) and the fragment of *HULK3* after *HULK3*'s PRR (*HULK3* post-PRR, nucleotides 342 to 924). To synthesize these fragments, PCR reactions were performed using *HULK2* and *HULK3* cDNA.

2.3.1 OE-PCR primer design

For the first fragment, *HULK3* pre-PRR was amplified with two primers, CT-HULK3 FWD and DH101, and 1 µl of *HULK3* cDNA (Table 2.1). The forward primer, CT-HULK3 FWD, was designed by another student (Preetam Janakirama) and was tailored to allow PCR amplification from the beginning of the carboxyl-terminus of *HULK3* while maintaining an open reading frame in the chimeric protein; it also included a *Sal*I cut site to allow ligation into either cloning or yeast expression vectors. The reverse primer for *HULK3* pre-PRR, DH101, was designed to begin replication 21 base pairs upstream of *HULK3*'s PRR. The resulting PCR product was a 238 bp fragment spanning the beginning of *HULK3*'s carboxyl-terminus to the beginning of *HULK3*'s PRR. *HULK3* cDNA (1 µl) was used for amplification of the second fragment, *HULK3* post-PRR. The fragment spanned 620 bp, starting from the end of *HULK3*'s PRR and extending a few amino acids past the end of *HULK3*. The forward primer, DH102, was designed from the first 7 amino acids in sequence following *HULK3*'s PRR. The reverse primer, G7705, was designed by a previous lab member, Sathya Challa. It extends into the noncoding region after *HULK3* and also includes a *Not*I cut site to allow ligation into cloning and yeast expression vectors. The final fragment to be amplified, *HULK2*'s PRR, was 175 bp in length and amplified from 1 µl of *HULK2* cDNA. The forward primer, DH103f, was composed of 2 parts, the reverse primer of *HULK3* pre-PRR as well as the first eight amino acids of *HULK2*'s PRR. The reverse primer, DH104r, was composed of the last 7 amino acids of *HULK2*'s PRR followed by the forward primer of *HULK3* post-PRR. The inclusion of sequences from *HULK3* will allow for overlaps during the ligation step of the PCR fragments in the second round of PCR.

2.3.2 Domain swap fragment PCR

Three separate PCR reactions were performed in a Stratagene RoboCycler Gradient 40 using 1 µl of Phusion high fidelity DNA polymerase (Thermo Scientific) for each reaction. The PCR conditions for the domain swap fragments were as follows:

1) 3-minute denaturation step at 95 ^{o}C; 1 cycle
2) 45-second denaturation step at 95 ^{o}C,

   45-second annealing step at 55 ^{o}C,

   60-second extension step at 72 ^{o}C; 32 cycles
3) 5-minutes extension step at 72 ^{o}C; 1 cycle

The amplified fragments were then run on a 0.8% agarose gel and purified using a GeneJET Gel Extraction Kit (Fermentas).

2.3.3 OE-PCR

Following the purification, to form the CTHULK3hulk2PRR construct, an OE-PCR reaction was performed to ligate all three gel purified PCR products. Primers CT-HULK3 FWD and G7705 were used in conjunction with Phusion high fidelity DNA polymerase. The OE-PCR conditions differed from the conditions for the domain swap as a larger product, 990 bp, was synthesized:

1) 5-minute denaturation step at 95 ^{o}C; 1 cycle
2) 45-second denaturation step at 94 ^{o}C,

   45-second annealing step at 55 ^{o}C,

   80-second extension step at 72 ^{o}C; 32 cycles
3) 5-minutes extension step at 72 ^{o}C; 1 cycle

The results of the OE-PCR PCR reaction were ran on a 0.8% agarose gel and subsequently purified with the Fermentas gel extraction kit.

2.3.4 Cloning of domain swap constructs

For the cloning of CTHULK3hulk2PRR into pGEMT-easy, the purified OE-PCR product was first treated with a standard tailing procedure and then ligated to the vector. 222 ng of purified PCR product was incubated for 30 minutes at 70 ^{o}C with 1 µl of *Taq* DNA polymerase (NEB) to add on a single adenine residue to the 3' ends of the construct. For ligation, 100 ng of linearised vector and 77 ng of CTHULK3hulk2PRR were combined (3:1 ratio of vector and insert) in an overnight incubation at 16 ^{o}C with 1 µl T4 DNA ligase (Fermentas). After ligation, the product was purified and transformed into the chemicompetent DH5α. Colonies were screened on LB agar plates containing 50 µg/ml of ampicillin and 40 µl of X-Gal. Successful transformants, white colonies, were picked and regrown in LB overnight at 37 ^{o}C. A GeneJet Plasmid Miniprep Kit (Fermentas) was then used for plasmid extraction. To check for errors in the cloned products, CTHULK3hulk2PRR pGEMT-easy, purified plasmids were sent to the DNA Sequencing facility to be sequenced using standard primers SP6 and T7. Sequencing results were assembled manually and analyzed using DNAMAN (Lynnon Corporation) and ClustalW. After confirming the accuracy of inserts, both the cloned products and yeast expression vector pDBLeu were digested with *Sal*I and *Not*I, cut sites existing in the forward and reverse primers. A GeneJET Gel Extraction Kit (Fermentas) was used to isolate and purify CTHULK3hulk2PRR and linearised pDBLeu from a 0.8% gel. Another ligation reaction was performed using both products, a 3:1 ratio of linearised pDBLeu and CTHULK3hulk2PRR, and T4 DNA ligase (Fermentas).

Table 2.1: Primer list for cloning and domain swap experiments

Primers used for cloning of constructs used in the yeast one-hybrid experiments. The list also includes the primers used to create the domain swapped chimeric proteins.

Primer Name	Sequence (5' to 3')	Function
CT-HULK3 FWD	GTCGACTGTCTCGTCATCCACGGCTG	Cloning of HULK3's pre-PRR. Includes *Sal*I cut site.
DH101	GAGTCTGCAAATTGTCTGCTT	Cloning of *HULK3*'s pre-PRR (reverse primer)
DH102	GACTACCGCAGAAATCCCAGC	Cloning of *HULK3*'s post-PRR
G7705	GTGGGGAGACAAGAGATGAAGAGCGGCCGC	Cloning of *HULK3*'s post-PRR (reverse primer). Includes *Not*I cut site.
DH103f	GAGTCTGCAAATTGTCTGCTTGTCTTTGGCACTTCACATCAGCAT	Cloning of *HULK2*'s PRR
DH104r	GATTCCTACTTGAATGGGTTTTACCGCAGAAATCCCAGCATG	Cloning of *HULK2*'s PRR (reverse primer)
CT-HULK2 FWD	GTCGACTGTCTCGTCATCCACGGCTG	Cloning of carboxyl-terminus of *HULK2*. Includes a *Sal*I cut site.
DSPD	GGACAGCAGCATCGGCCC	Cloning of *HULK2*'s pre-PRR (reverse primer)
DSPE	GAAAATGGAGGATATCGC	Cloning of *HULK2*'s post-PRR
G7705	AGAAACTTGAGAGAGAGTTATAAAGCGGCCGC	Cloning of HULK2's post-PRR (reverse primer). Includes *Not*I cut site.
DSPA	GGACAGCAGCATCGGCCCGTCCCTGGAACTTCACATCAG	Cloning of *HULK3*'s PRR
DSPB	TCCTACTCAAATGGCTTTGAAAATGGAGGATATCGC	Cloning of *HULK3*'s PRR (reverse primer)

After an overnight incubation at 16 ^{o}C, the products were purified (GeneJET Gel Extraction Kit, Fermentas). CTHULK3hulk2PRR was then used to transform into *S. cerevisiae* strain MaV203 chemically using the method described for the 3-AT titration of full length and CT-ends of the *HULK* gene family, in Section 2.2.3.

2.3.5 3-AT titration

To perform 3-AT checks, 3 transformant colonies were picked and inoculated in separate Falcon tubes of liquid YNB –leucine overnight. The following morning, OD_{600} measurements were taken and inocula from the overnight cultures were diluted down to a uniform OD_{600} reading of 0.5 in separate 1.5 ml eppendorph tubes. From each tube of overnight culture, 1 µl was spotted twice on 2 separate 13 mm × 13 mm squares on 8 different gridded square Petri plates. The plates contained YNB -leucine -histidine medium and 3-AT concentration ranging from 0 mM to 300 mM, with each plate increasing by 50 mM. Plates were then incubated at 30 ^{o}C. Afterwards, each day, for the following 4 days, plates were scanned using a scanner.

2.4 Bioinformatics analysis

The protein sequences of the carboxyl-terminus of HULK2 and HULK3 were further analysed using bioinformatics tools to find unidentified putative domains through sequences comparisons with other sequences in databases of well-known and characterised conserved domains. In particular, the sequences of both proteins' pre-PRR and post-PRR were submitted as queries; CT-HULK2 residues 1 - 58 and residues 103 – 298, CT-HULK3 residues 1 – 79 and residues 114 – 308.

Three web-based programs were used to detect protein domains in the protein sequence. NCBI's BLAST tool was used to find regions of similarities within the non-redundant protein sequences (nr) database. Searches used the default parameters of the blastp (protein-protein BLAST) algorithm. The NCBI Conserved Domain Database (CDD) was also searched using the CD-Search Tool. All databases were searched using the default options. EMBL SMART software was also used to detect novel protein domains as well as their putative function. All algorithm options of identifying additional protein domains were used in each analysis.

Chapter 3 : Results

3.0 Transcriptional activity in the HULK protein family

The amino acid sequences of the HULK protein family carry sequences typical of transcriptional activation domains, namely, large tracts of proline residues which could indicate a proline-rich TAD and the presence of PPLP repeats, which could facilitate recruitment and binding of basal transcription machinery. The interest in a putative TAD was developed after initial yeast two-hybrid assays with various constructs of the HULK protein family indicated the presence of a strong autoactivator. Thus, to investigate further a putative TAD in members of the HULK protein family, yeast one-hybrid assays were performed.

The presence of putative TADs was verified by testing the ability of constructs of the HULK protein family to contribute to autoactivation in a yeast one-hybrid system, while titration of 3-AT in the yeast assays allowed measurement of the strength of the autoactivation, and consequently the strength of the TADs. Since the sequences identified as putative TADs are found in the carboxyl-terminus of the proteins, the yeast one-hybrid assays were performed with full-length proteins as well as truncated proteins, the carboxyl-terminus ends of the HULK protein family.

3.0.1 Subcloning full-length and carboxyl-terminus of the HULK gene family

Full-length *HUA2* (Figure 3.1), *HULK2* (Figure 3.2), and *HULK3* (Figure 3.3) and their truncated iterations, *CT-HUA2* (Figure 3.4), *CT-HULK2* (Figure 3.5), and *CT-HULK3* (Figure 3.6), were subcloned from either pEXP-AD or pGEM-T Easy into

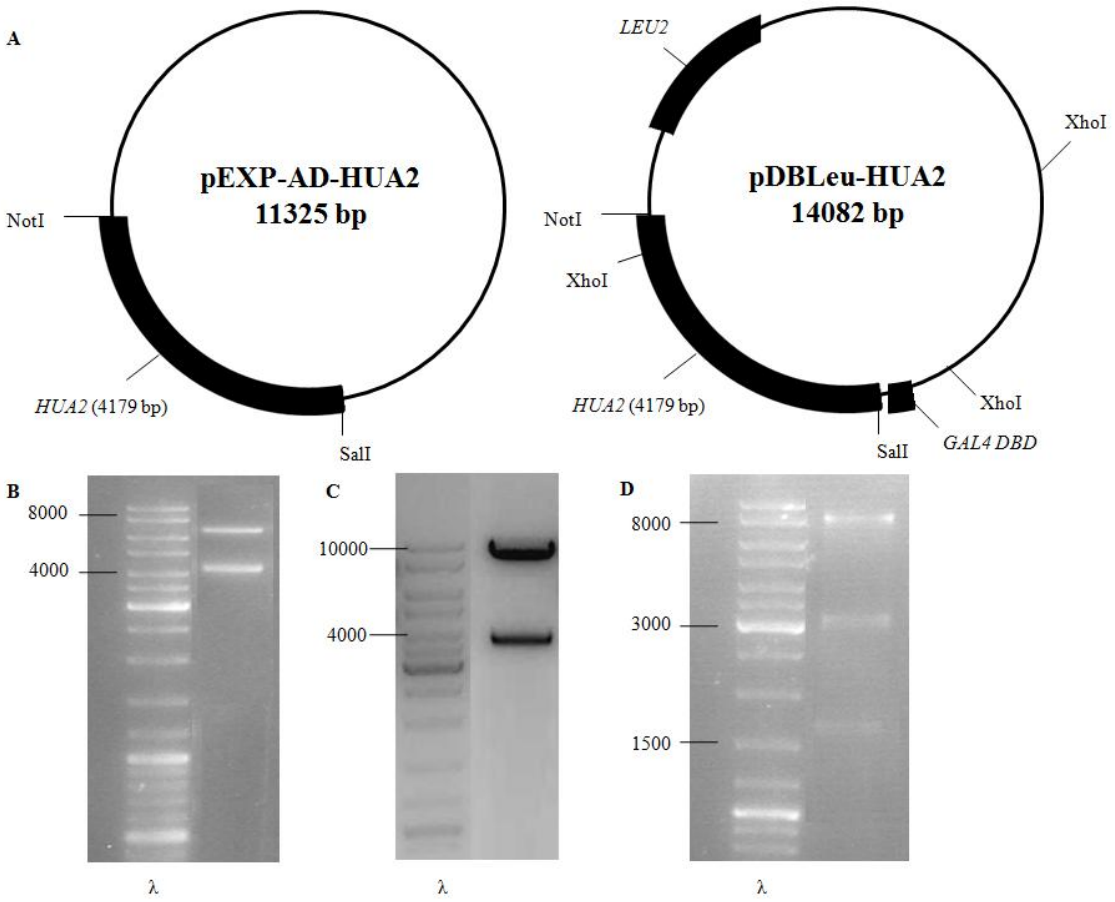

Figure 3.1: Steps in subcloning full-length HUA2 from pEXP-AD into pDBLeu.

A) Schematic of HUA2 in both pEXP-AD and pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of the restriction digestion to excise HUA2 (4179 bp) from pEXP-AD (7146 bp) with restriction enzymes *Sal*I and *Not*I. C) Digestion of HUA2 from pDBLeu (9903 bp) after ligation of HUA2 and pDBLeu. D) Digestion of pDBLeu-HUA2 with *Xho*I (expected bands 9.1 kbp, 3.2 kbp, 1.7 kbp).

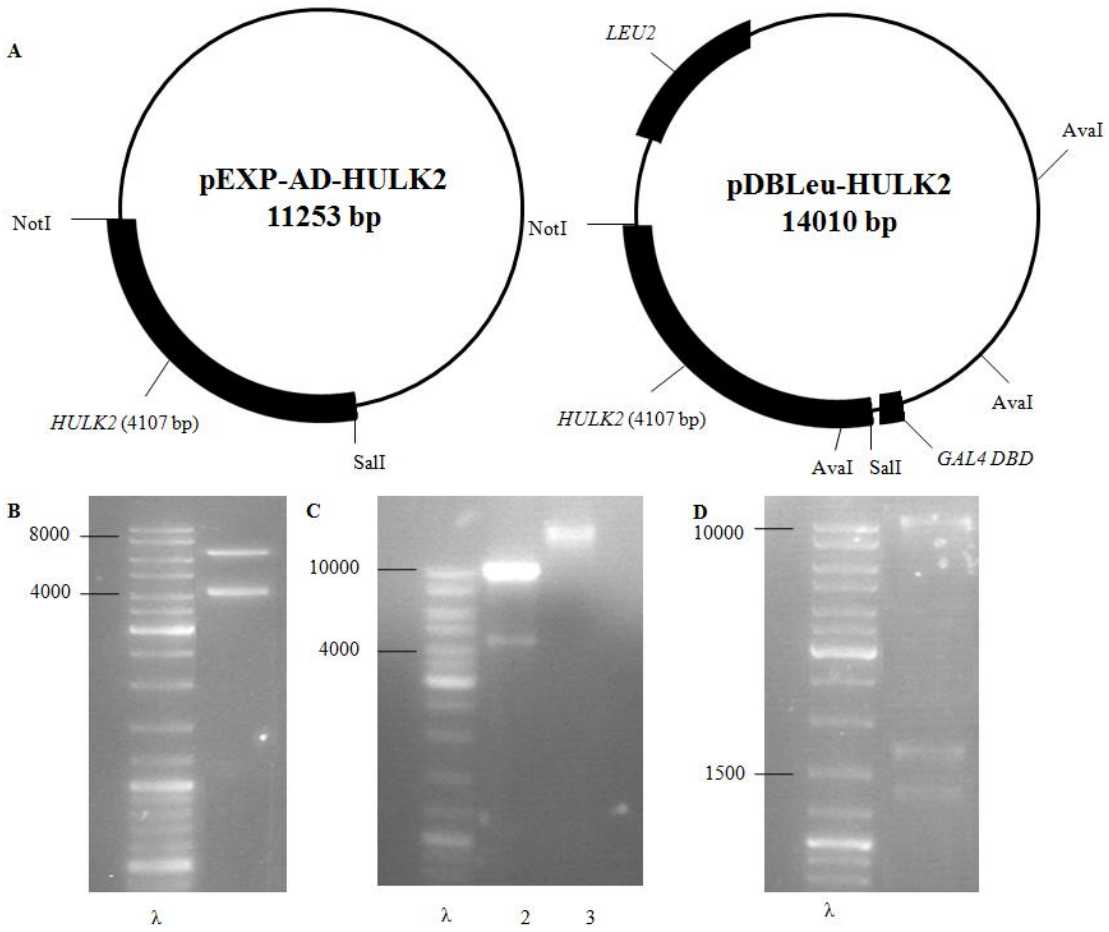

Figure 3.2: Steps in subcloning full-length *HULK2* from pEXP-AD in pDBLeu.

A) Schematic of *HULK2* in both pEXP-AD and pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of A) Products of the restriction digestion to excise *HULK2* (4107 bp) from pEXP-AD (7146 bp) with *Sal*I and *Not*I. C) Ligation of *HULK2* into pDBLeu (9903 bp) (lane 3) as well as a control reaction without T4 DNA ligase (lane 2). D) Digestion of pDBLeu-HULK2 with *Ava*I (expected bands 10.8 kbp, 1.7 kbp, 1.4 kbp).

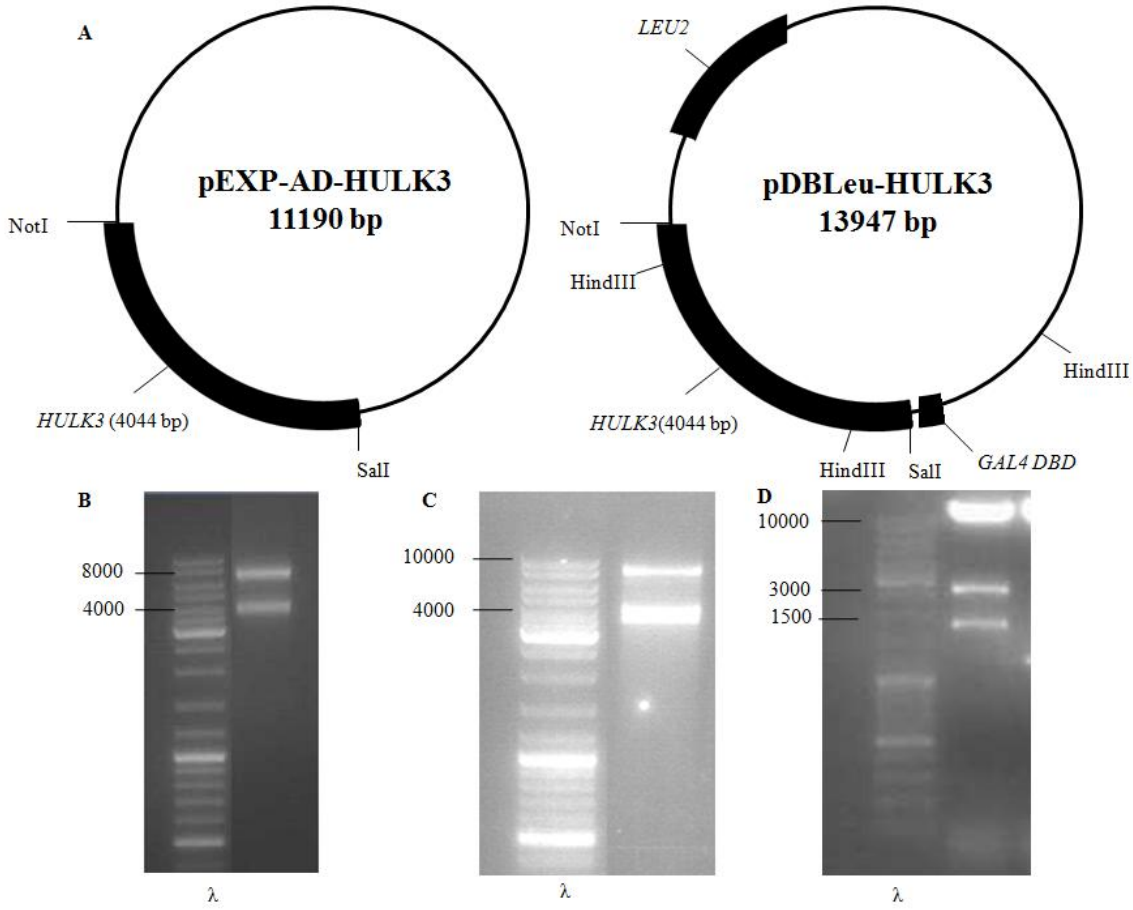

Figure 3.3: Steps in subcloning full-length *HULK3* from pEXP-AD into pDBLeu.

A) Schematic of *HULK3* in both pEXP-AD and pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of the restriction digestion to excise *HULK3* (4044 bp) from pEXP-AD (7146 bp) with *Sal*I and *Not*I. C) Digestion of *HULK3* from pDBLeu (9903 bp) after ligation of *HULK3* and pDBLeu with *Sal*I and *Not*I. D) Digestion of pDBLeu-HULK3 with *Hin*dIII (9.8 kbp, 2.7 kbp, 1.4 kbp).

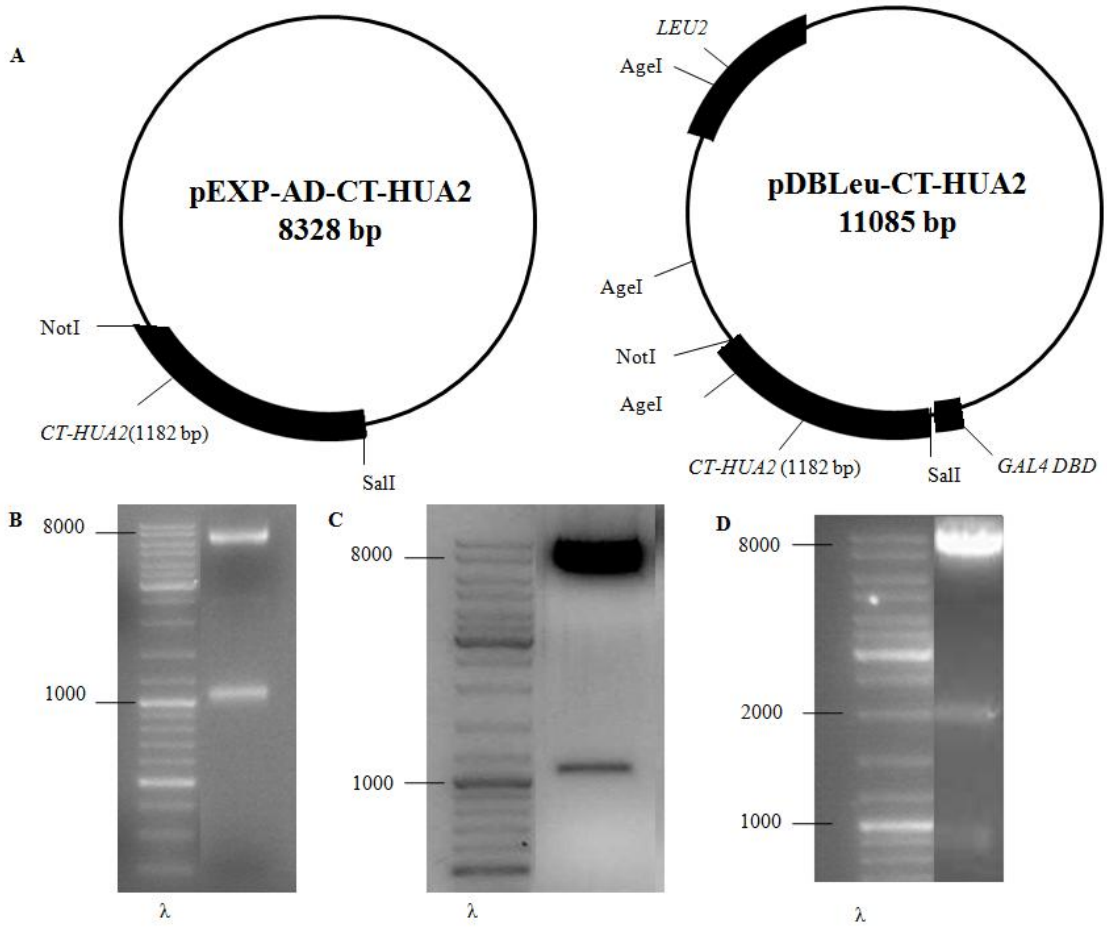

Figure 3.4: Steps in subcloning carboxyl-terminus *HUA2* from pEXP-AD into pDBLeu.

A) Schematic of *CT-HUA2* in both pEXP-AD and pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of the restriction digestion to excise *CT-HUA2* (1182 bp) from pEXP-AD (7146 bp) with *Sal*I and *Not*I. C) Digestion of *CT-HUA2* from pDBLeu (9903 bp) after ligation of *CT-HUA2* and pDBLeu with *Sal*I and *Not*I. D) Digestion of pDBLeu-HULK3 with *Age*I (8.2 kbp, 1.9 kbp, 0.94 kbp).

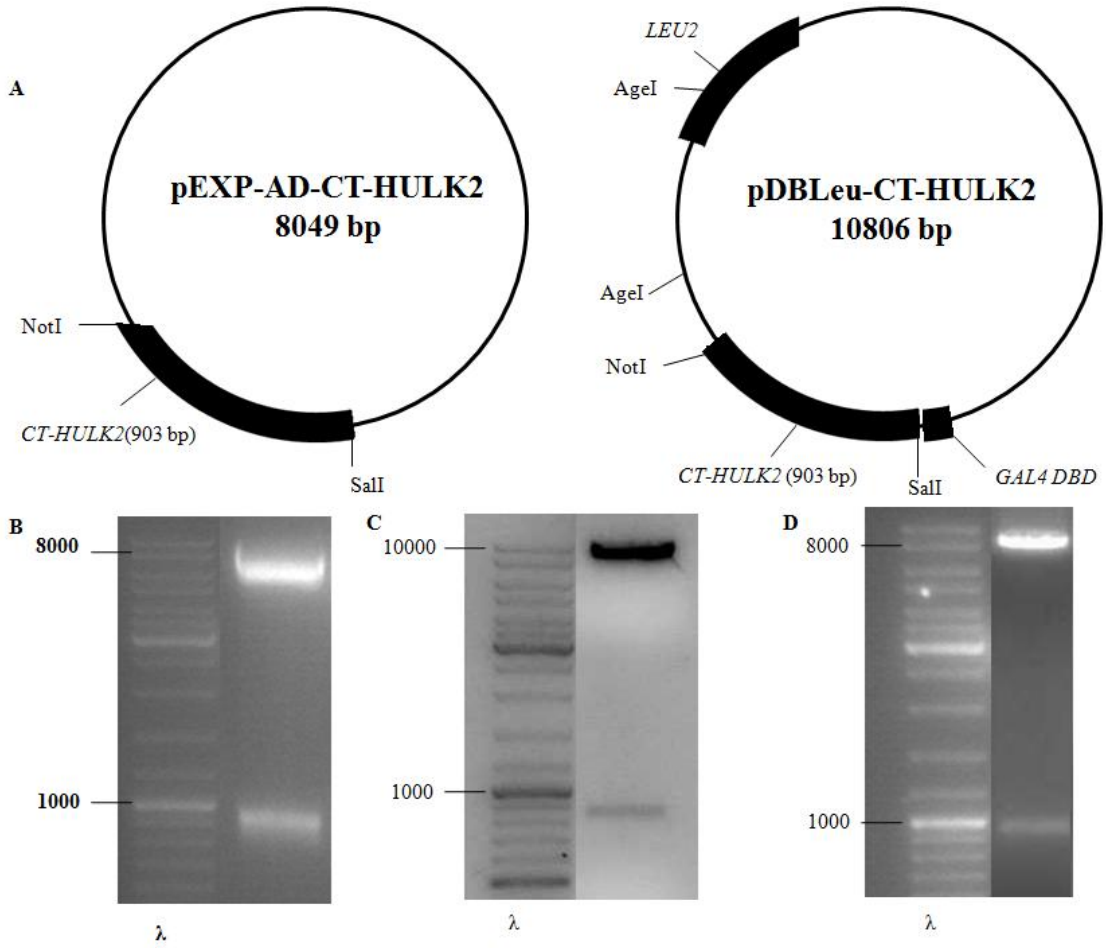

Figure 3.5: Steps in subcloning carboxyl-terminus *HULK2* from pEXP-AD into pDBLeu.

A) Schematic of *CT-HULK2* in both pEXP-AD and pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of the restriction digestion to excise *CT-HULK2* (903 bp) from pEXP-AD (7146 bp) with *Sal*I and *Not*I. C) Digestion of *CT-HULK2* from pDBLeu (9903 bp) after ligation of *CT-HULK2* and pDBLeu with *Sal*I and *Not*I. D) Digestion of pDBLeu-HULK2 with *Age*I (9.8 kbp, 0.9 kbp).

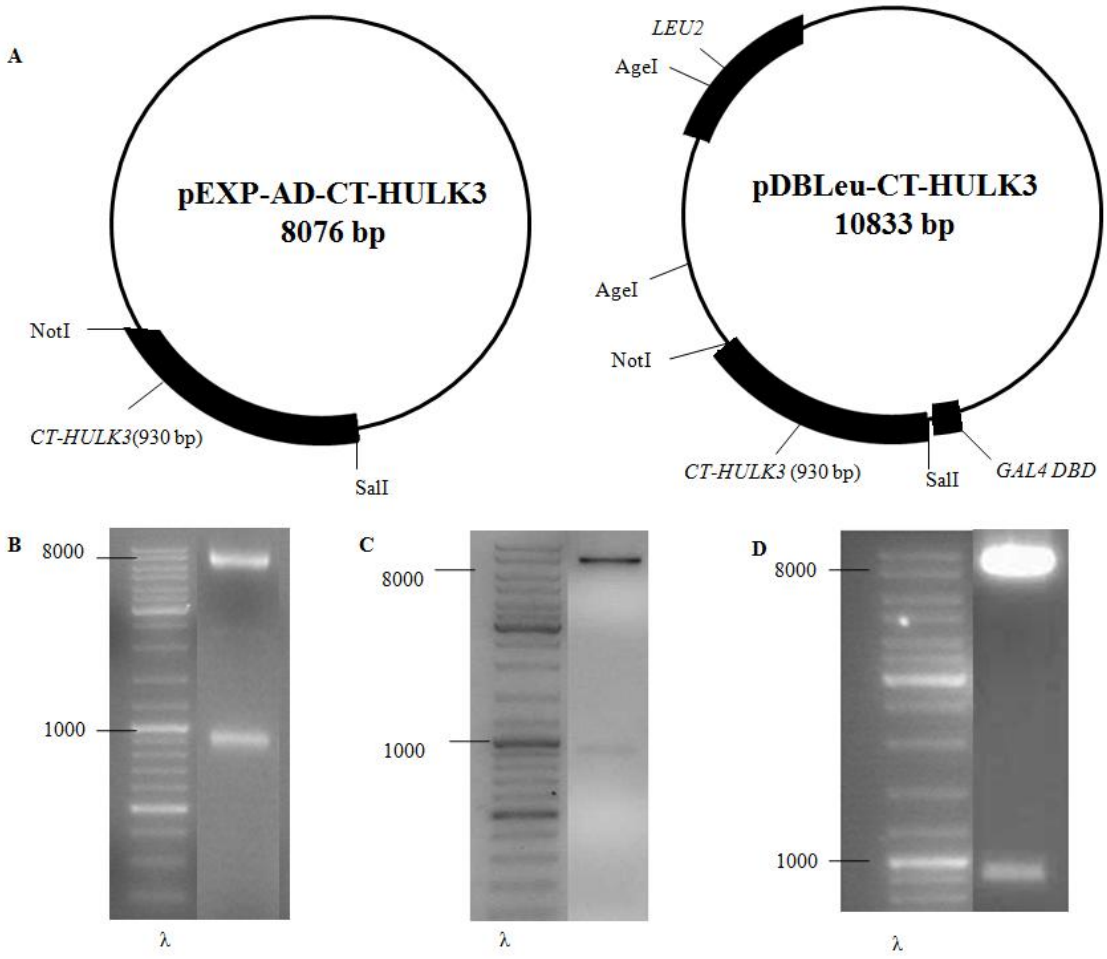

Figure 3.6: Steps in subcloning carboxyl-terminus *HULK3* from pEXP-AD into pDBLeu.

A) Schematic of *CT-HULK3* in both pEXP-AD and pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of the restriction digestion to excise *CT-HULK3* (930 bp) from pEXP-AD (7146 bp) with *Sal*I and *Not*I. C) Digestion of *CT-HULK3* from pDBLeu (9903 bp) after ligation of *CT-HULK2* and pDBLeu with *Sal*I and *Not*I. D) Digestion of pDBLeu-HULK3 with *Age*I (9.8 kbp, 0.9 kbp).

pDBLeu, a plasmid suitable for transformation of both *E.coli* (DH5α) and *S. cerevisiae* (MaV203).

Overnight cultures of glycerol stocks of *E.coli* strain DH5α containing HULKs in either pEXP-AD or pGEM-T Easy were grown in LB and DNA was isolated. The *HULK* genes were excised from the cloning vectors while the destination vector, pDBLeu, was linearised, both reactions using the restriction enzymes *Not*I and *Sal*I. The *HULK* genes were then ligated into pDBLeu and the products of that reaction was used to transform DH5α cells.

DNA was isolated from positive colonies grown overnight. Plasmid DNA was isolated and digested with either *Age*I, *Ava*I or *Hin*dIII to confirm the presence of the genes in pDBLeu as they cut both the insert and vector.

3.0.2 Subcloning *HULK1* into pDBLeu

The subcloning of *HUA2*, *HULK2*, and *HULK3* into pDBLeu was successful. Unfortunately, for *HULK1*, the procedure was met with no success.

Numerous attempts were made to subclone *HULK1* into pDBLeu. Initially, attempts with HULK1 used the same procedure for the other members of the HULK gene family, see Section 2.2. However, when the product of the ligation of *HULK1* into pDBLeu was checked with restriction digestions, the expected products were not observed. Later attempts at subcloning *HULK1* into pDBLeu then employed PCR reactions to create fragments for ligation into pDBLeu (Figure 3.7A). Despite trials with various ratios of insert (*HULK1*) to vector (pDBLeu) during the ligation step, none of the checks with restriction digestions yielded the expected fragments for a successful subcloning (Figure 3.7B).

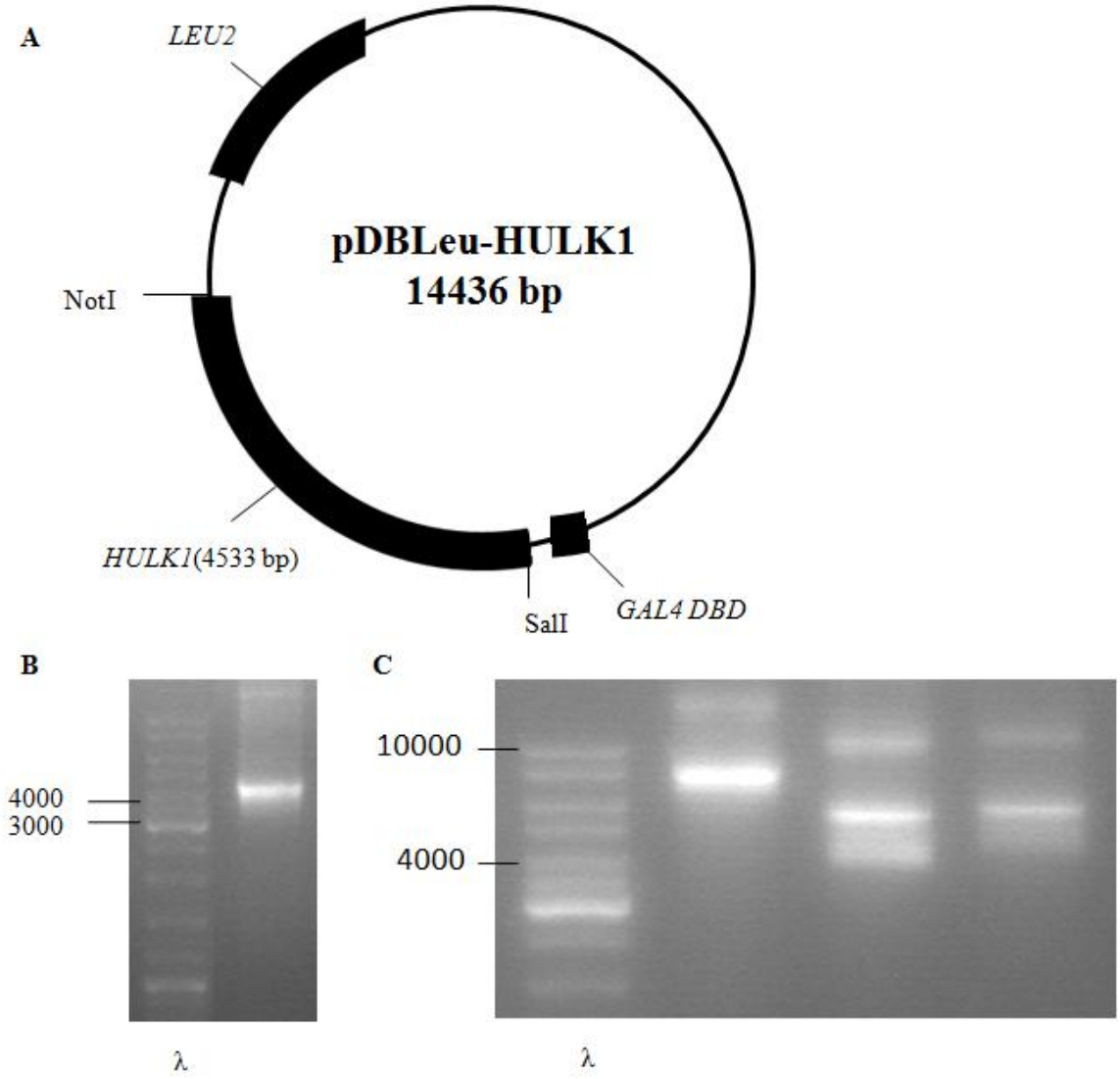

Figure 3.7: Attempt at subcloning HULK1 into pDBLeu.

A) Schematic of *HULK1* in pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of a PCR reaction to generate HULK1 fragments (4533 bp) for ligation. C) Digestion DNA plasmids recovered from a ligation reaction of *HULK1* and pDBLeu (9903 bp) with restriction enzymes *Not*I and *Sal*I (9.9 kbp, 4.5 kbp).

3.1 Transcriptional activity in the full-length HULK protein family

Transcriptional activity in full-length HULKs was checked utilising yeast one-hybrid assays and titration with 3-AT. pDBLeu carrying HUA2, HULK2 and HULK3, along with empty pDBLeu were subsequently used to transform MaV203. Therefore, for the yeast one-hybrid assays, the proteins expressed were HUA2, HULK2, and HULK3, each translationally fused to GAL4 DBD, while in the empty pDBLeu vector, only the GAL4 DBD was expressed.

Colonies of MaV203 expressing these proteins were plated on amino acid drop out medium, YNB -leucine -histidine. The plasmids carried the reporter gene *LEU2*, which allowed identification of successful yeast transformants capable of surviving on media lacking leucine. In line with the hypothesis that the HULKs contain a putative TAD, it was expected that they would upregulate the reporter gene, resulting in the expression of *HIS3* and the survival of transformed MaV203 colonies on the -histidine medium due to their ability to biosynthesize histidine. On the other hand, colonies transformed with an empty vector would be able to biosynthesize both leucine and very low levels of histidine. The low levels of histidine were a result of a known leaky *HIS3* promoter found in this system. However, with the addition of 3-AT, growth of the colonies will be limited.

3-AT was used in the yeast one-hybrid assays, for two reasons. Since growth is expected across all transformants, including the control, 3-AT is used to quench the basal transcription levels of the *HIS3* gene. As well, this allows the strength of TADs to be tested. Through titration with 3-AT, survival of transformant colonies on plates with increasing concentrations of 3-AT was dependent on the strength of the TAD carried on

the plasmids. HULKs with stronger TADs will be able to activate *HIS3* and biosynthesize histidine, required for growth, at levels high enough to overcome the quenching effects of 3-AT.

The control, MaV203 transformed with empty pDBLeu, was expected to grow on YNB -leucine –histidine without 3-AT. However, at the lowest level of 3-AT to be used, 10 mM, it was expected that no growth would be observed as concentrations of less than 10 mM 3-AT are sufficient to quench the low basal expression of the *HIS3* gene (Guthrie, 2002; Walder et al., 2002; Prodoehl, 2007).

The HUA2-GAL4 DBD construct created for this experiment (Figure 3.1) was expected to carry a strong TAD based on its large PRR, relative to other HULKs, of 45 proline residues and the presence of four PPLP repeats. The HULK2-GAL4 DBD construct (Figure 3.2) was expected to have high transcriptional activity, but not as great as that of HUA2. In its PRR, it has 11 proline residues, but only one PPLP. The HULK3-GAL4 DBD construct (Figure 3.3) was expected to have the least amount of growth out of the full-length constructs tested. It does not have any PPLP repeats, and the smallest PRR which contains only four proline residues.

Once plated, all the transformants exhibited various levels of growth. These included a control, MaV203 transformed with an empty vector (Figure 3.8), highlighting both the leaky nature of the MaV203 yeast system and the need for 3-AT. The control MaV203 transformed with an empty vector had growth only on plates lacking 3-AT.

MaV203 colonies expressing HUA2 grew on plates with less than 50 mM of 3-AT. At this concentration, HUA2 could be considered a weak autoactivator. Colonies expressing HULK2 and HULK3 grew on 3-AT plates at concentrations in the range of

strong transcription factors, over 200 mM of 3-AT. HULK2 allowed growth on plates with up to 200 mM of 3-AT and HULK3 allowed growth on plates with up to 250 mM of 3-AT (Figure 3.9). While not all HULKs have strong TADs, HULK2 and HULK3 had strong TADs. Interestingly, HUA2 is the richest of the HULKs in proline residues and the greatest number of PPLP repeats, but had the lowest transcriptional activity. On the other hand, HULK3, which is the least abundant in proline residues and lacks PPLP repeats, had the greatest level of transcriptional activity.

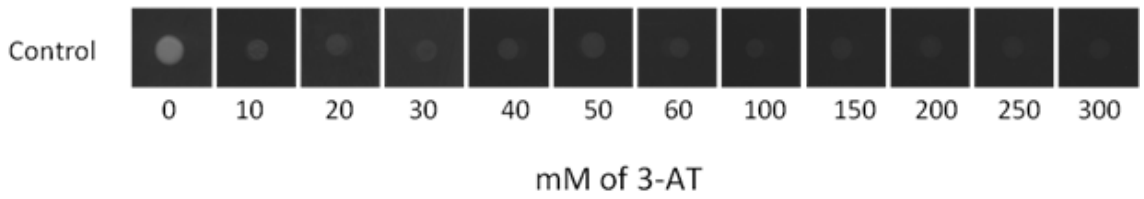

Figure 3.8: 3-AT titrations without a transcriptional activator in pDBLeu.

Yeast transformed with unmodified pDBLeu were grown on media with increasing concentrations of 3-AT. Due to the low basal expression of the *HIS3* reporter gene, untransformed colonies were able to grow on YNB -leucine -histidine medium at 0 mM of 3-AT. However, without the presence of an autoactivator, colonies were unable to grow in the presence of 10 mM 3-AT.

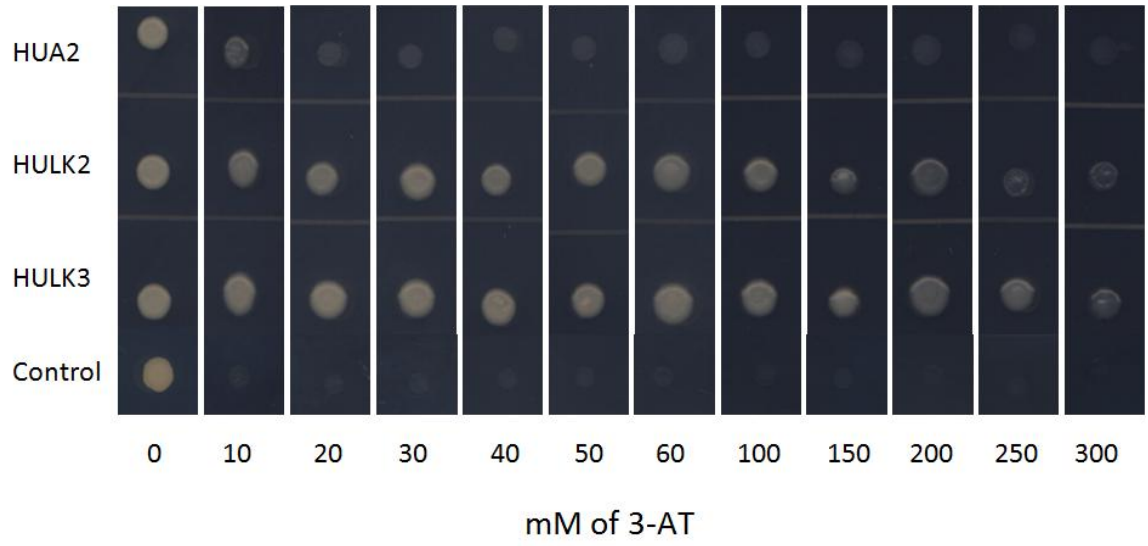

Figure 3.9: 3-AT titrations with full length members of HULK protein family.

Yeast expressing the full-length proteins of the HULK protein family were grown on media with increasing concentrations of 3-AT to identify the concentration at which autoactivation of the *HIS3* gene would be quenched, resulting in no growth. The control was empty pDBLeu vector.

3.2 Transcriptional activity in the carboxyl-terminus end of the HULK protein family

Once it was confirmed that there is transcriptional activation within members of the HULK family, the PRR toward the carboxyl-terminus was the focal point of the next set of experiments. Unlike the rest of the protein, where there are defined structures (the PWWP, NLS, and RPR domains), this region does not contain any known domains aside from proline-rich tracts, a trait of many transcription factors, as well as PPLP repeats, which could aid in transcriptional activity (Williamson, 1994; Kay et al., 2000). Thus, truncated iterations of *HUA2* (*CT-HUA2*), *HULK2* (*CT-HULK2*), and *HULK3* (*CT-HULK3*) were cloned into pDBLeu and subsequently used to transform MaV203, along with empty pDBLeu. Translationally fused to GAL4 DBD, these proteins, CT-HUA2 (Figure 3.4), CT-HULK2 (Figure 3.5), and CT-HULK3 (Figure 3.6), were titrated with 3-AT in yeast one-hybrid assays to establish the transcriptional activity of the putative TADs.

If the PRR is responsible for the transcriptional activity observed during the yeast one-hybrid assays with full-length HULKs, then it was expected that the truncated proteins would have the same levels of transcriptional activity as their full-length counterparts: CT-HUA2 should have the weakest levels of transcriptional activity while CT-HULK2 and CT-HULK3 should allow growth of MaV203 at the greatest concentrations of 3-AT in –leucine –histidine drop out medium.

The transcriptional activity, as measured by growth on 3-AT plates revealed unexpected results: CT-HUA2 allowed MaV203 colonies to grow at high concentrations of 3-AT, up to 150 mM of 3-AT as opposed to 50 mM of 3-AT as observed for the full-

length HUA2 (Figure 3.10). Expression of CT-HULK2 in MaV203 showed an increase in autoactivation strength compared to full-length HULK2, which allowed growth in 3-AT concentrations of up to 250 mM (Figure 3.10). On the contrary, MaV203 colonies expressing CT-HULK3 experienced a massive reduction in autoactivation strength with colonies not growing beyond 10 mM of 3-AT (Figure 3.10).

When isolated through truncations assayed for transcriptional activity, the PRR of the HUA2 and HULK3 did not maintain the levels of transcriptional activity observed in yeast one-hybrids with full-length proteins. The transcriptional activity observed in the carboxyl-terminus of HUA2 indicates that a TAD is present in the PRR but is hindered by other factors in the full-length protein. HULK3's carboxyl-terminus produced a large decrease in transcriptional activity implying that its PRR might not be responsible for the transcriptional activity observed in full-length protein. HULK2's PRR maintained a high level of transcriptional activity when assayed both with the full-length protein and as the carboxyl-terminus region. Thus, its PRR is a good candidate for being a strong TAD.

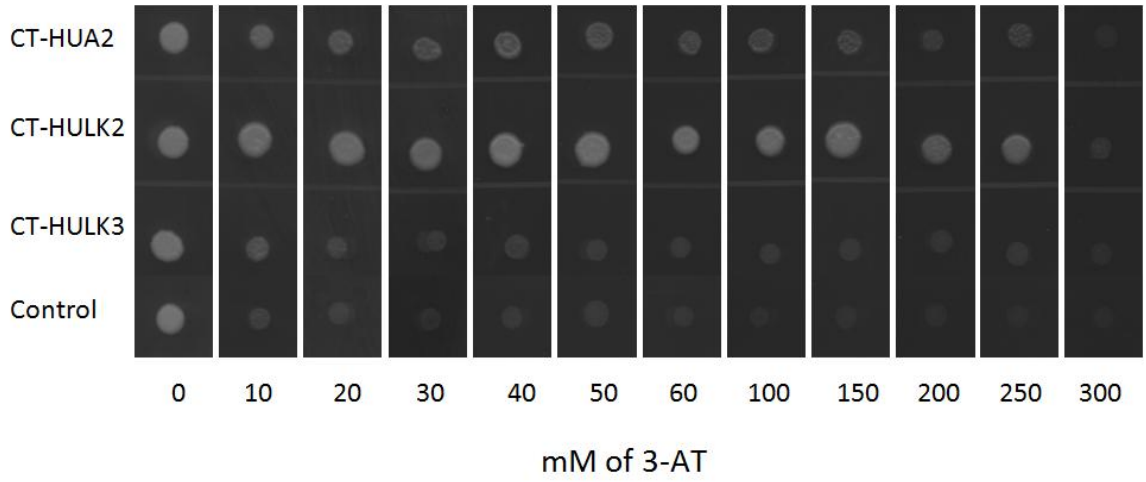

Figure 3.10: 3-AT titrations with carboxyl-terminus of HULK protein family.

Yeast expressing the carboxyl-terminal ends of the HULK protein family were grown on media with increasing concentrations of 3-AT to identify the concentration at which autoactivation of the *HIS3* gene would be quenched, resulting in no growth. The control was empty pDBLeu vector.

3.3 Transcriptional activity in PRR swapped chimeric proteins

Through yeast one-hybrid assays with full-length HULKs, I found that both HULK2 and HULK3 contain potent activation domains. However, when examining the carboxyl-terminus of the HULKs for transcriptional activity in the PRR, I observed that CT-HULK2 confers a high level of autoactivation while CT-HULK3 does not. This difference in transcriptional activity could be attributed to differences between the HULK2 and HULK3 PRR.

Compared to HULK3, HULK2 has a larger PRR with 44 amino acids, of which a quarter (11) are proline residues. Furthermore, these proline residues are grouped together in clusters of one to five proline residues in sequence. Nestled in this putative domain is a single PPLP repeat. In contrast, HULK3 has a PRR that is shorter and has 34 amino acids. It contains only four proline residues that are not in series and does not carry a single PPLP motif.

To investigate further if the PRR in HULK2 was responsible for its strong autoactivation strength, a follow-up yeast one-hybrid assay was performed with two chimeric proteins: CT-HULK2's larger PRR was exchanged with CT-HULK3's smaller PRR and vice versa.

3.4 Preparation of domain-swapped constructs

Two chimeric proteins were produced through OE-PCRs (Figure 3.11): CTHULK3hulk2PRR and CTHULK2hulk3PRR. For each gene involved in the domain exchange, *CT-HULK2* and *CT-HULK3*, three separate PCRs were performed for each leading up to the OE-PCR. The three PCR-generated fragments prepared for the constitution of the final chimeric genes were: the PRR fragment, the pre-PRR fragment,

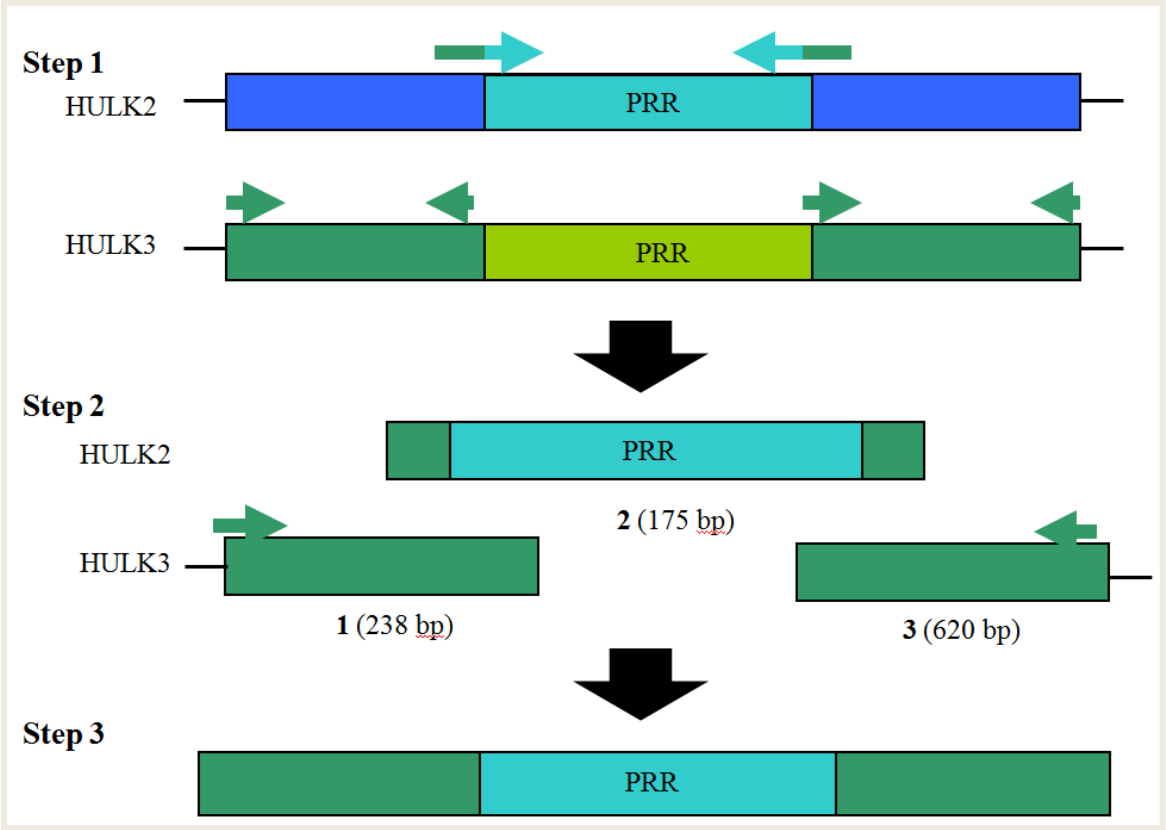

Figure 3.11: Steps of an OE-PCR to produce domain swapped proteins.

Step 1) Three separate PCRs are performed. In this example, HULK2's PRR is amplified and HULK3's pre-PRR and post-PRR regions are amplified . Step 2) The PRR fragment is amplified so that at its ends are sequences which overlap with the sequences of the pre-PRR and post-PRR of other gene. Step 3) All three fragments are combined in one final PCR reaction to produce the PRR-swapped CTHULK3hulk2PRR. Maps are not drawn to scale.

and the post-PRR fragment. The products of these PCR reactions were checked on agarose gels and purified with the GeneJET Gel Extraction Kit (Fermentas) before being recombined in two separate OE-PCRs to form *CTHULK2hulk3PRR* (Figure 3.12) and *CTHULK3hulk2PRR* (Figure 3.13).

For ligation into the cloning vector, pGEMT-Easy, a standard tailing procedure was performed, namely incubation of the PCR products and Taq DNA polymerase (NEB) with dATP at 70 ^{o}C for 30 minutes. The resulting 3' A-tailed fragments were then ligated into the cloning vector pGEMT-Easy and used to transform DH5α. Plasmids were then isolated from transformed colonies and the constructs excised from pGEMT-Easy and separated on a 0.8% agarose gel (Figure 3.12C and 3.13C). The constructs were subsequently recovered, purified and subcloned into pDBLeu, behind *GAL4DB* (Figure 3.12D and 3.13D). These plasmids were finally used to transform DH5α cells and were sequenced (Figure 3.14 and 3.15). The sequences were then compared with the desired sequence prepared in the laboratory to ensure accuracy of the cloning process. The chimeric proteins had the correct sequences, and thus were ready for the assays of their putative transcriptional activity

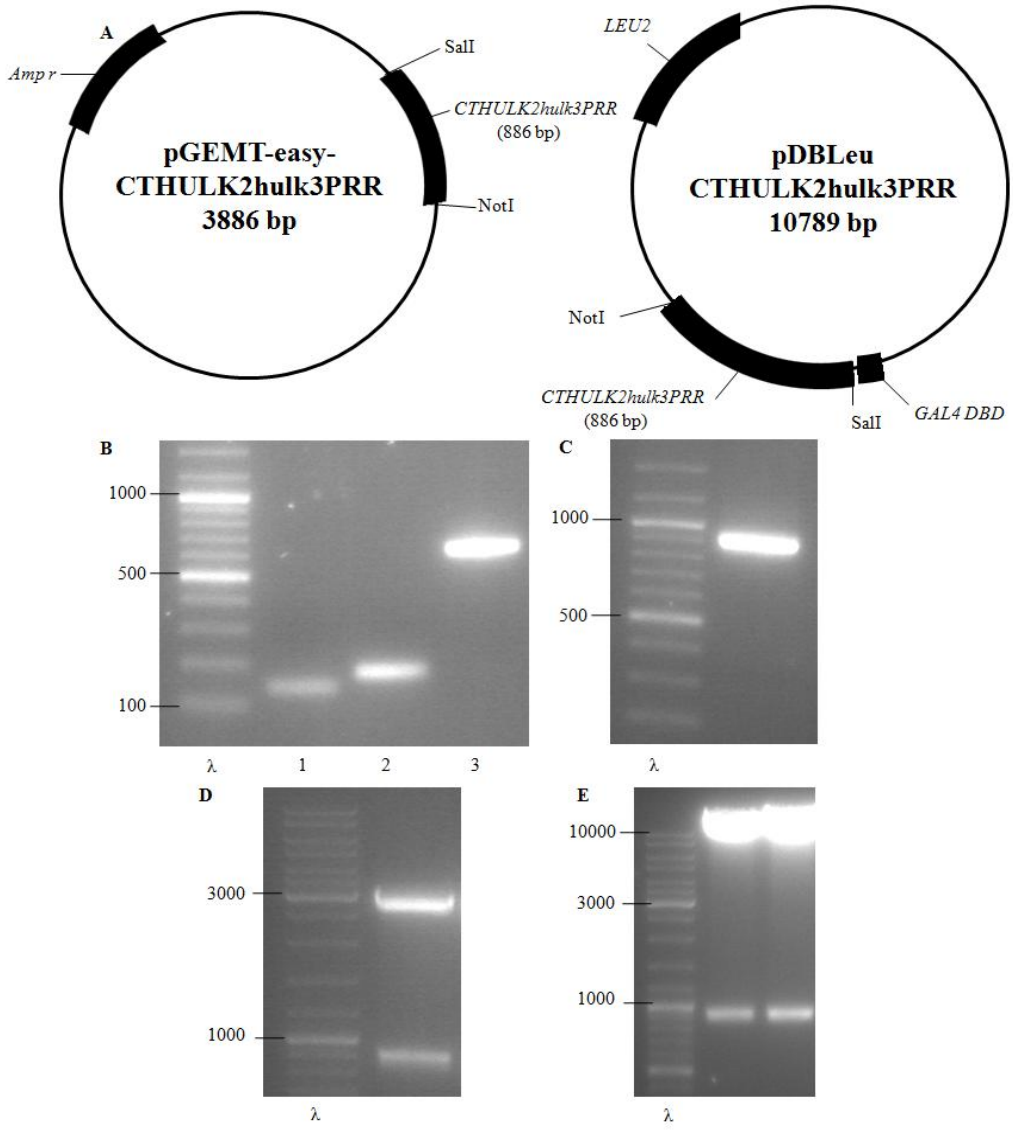

Figure 3.12: Cloning of domain swapped gene *CTHULK2hulk3PRR*.

A) Schematic of *CTHULK2hulk3PRR* cloned into pGEMT-Easy and pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of PCRs to amplify segments of the carboxyl-termini (1: *HULK2* pre-PRR 120 bp, 2: *HULK3* PRR 169 bp, 3: *HULK2* post-PRR 629 bp) C) OE-PCR reaction ligating all 3 fragments of *CTHULK2hulk3PRR* (10.7 kbp) D) Restriction digestion using *Sal*I and *Not*I after ligation of *CTHULK2hulk3PRR* into pGEMT-Easy (3000 bp). E) Restriction digestion with *Sal*I and *Not*I after ligation of *CTHULK2hulk3PRR* into pDBLeu (9903 bp).

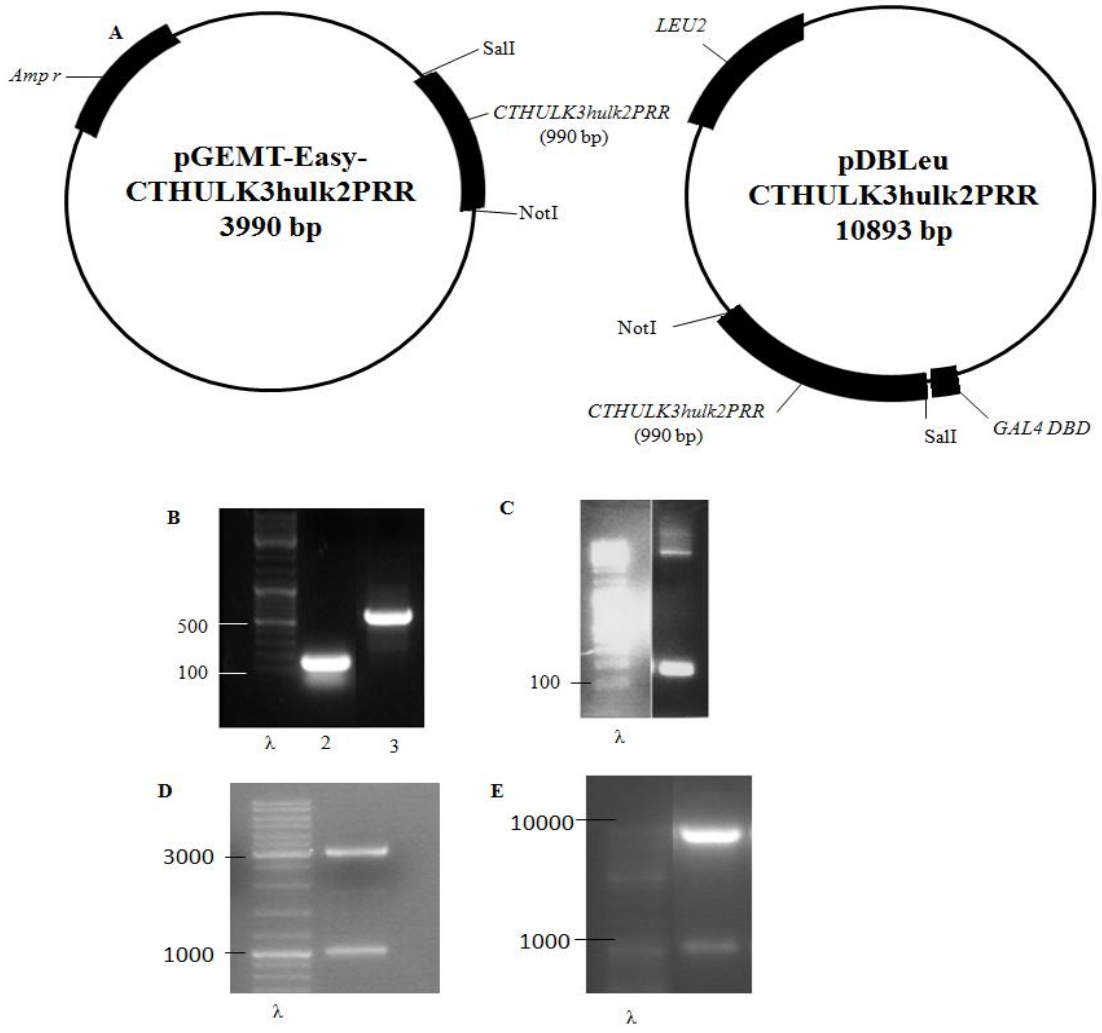

Figure 3.13: Cloning of domain swapped gene *CTHULK3hulk2PRR*.

A) Schematic of *CTHULK3hulk2PRR* cloned into pGEMT-Easy and pDBLeu. Locations of genes of interest and restriction enzyme cut sites are indicated. B) Products of PCRs to amplify segments of *HULK3* carboxyl-terminus (lane 2: *CT-HULK3* pre-PRR 238 bp, lane 3: *CT-HULK3* post-PRR 620 bp). C) Product of a PCR to amplify the PRR of *HULK2*: 175 bp. D) Restriction digestion using *Sal*I and *Not*I after ligation of *CTHULK3hulk2PRR* into pGEMT-Easy (3000 bp). E) Restriction digestion with *Sal*I and *Not*I after ligation of *CTHULK3hulk2PRR* into pDBLeu (9903 bp).

1	S	T	V	S	S	S	T	A	E	R	H	T	L	I	L	E	D	V	D	G
21	E	L	E	M	E	D	V	A	P	P	W	G	T	E	N	C	T	H	T	D
41	Q	A	D	N	T	K	V	S	N	C	Q	L	G	Q	Q	H	R	P	V	P
61	G	T	S	H	Q	N	V	T	S	S	S	P	P	A	R	P	S	Q	N	A
81	Q	L	A	M	S	N	S	Y	S	N	G	F	E	N	G	G	Y	R	N	V
101	H	G	D	Q	Q	A	G	P	L	R	M	N	P	P	L	S	G	S	T	M
121	H	Y	Q	G	P	E	S	S	Y	I	S	G	V	Q	L	T	N	S	I	P
141	Q	A	D	G	S	N	F	Q	H	R	P	Y	P	S	H	P	H	P	H	P
161	P	P	P	P	P	P	P	Q	H	Q	F	S	F	R	E	P	G	H	V	L
181	K	S	H	R	D	A	P	S	Y	S	H	R	S	H	Y	V	P	N	C	D
201	E	R	N	F	H	D	N	H	E	R	M	R	H	A	P	F	E	N	R	D
221	N	W	R	Y	P	P	S	S	S	Y	G	S	R	Y	Q	D	E	H	K	A
241	P	Y	P	S	S	S	Y	N	G	V	R	W	D	N	P	P	R	X	Y	N
261	N	R	P	S	F	H	P	K	P	H	S	E	G	P	A	P	V	G	M	R
281	D	P	G	M	W	H	Q	X	S	D	*									

Figure 3.14: Amino acid sequence of CTHULK2cthulk3PRR.

Amino acid sequence of the CTHULK2hulk3PRR protein. Red sequences represent the C-terminus of HULK2 and the blue sequences represents the PRR of CT-HULK3.

1	S	T	E	G	S	D	S	D	G	G	D	F	E	S	V	T	P	E	H	E
21	S	R	S	L	E	E	H	V	T	P	S	I	T	E	R	H	T	R	I	L
41	E	D	V	D	G	E	L	E	M	E	D	V	A	P	P	W	E	G	G	S
61	S	A	S	A	I	T	D	Q	A	D	N	R	E	S	A	N	C	L	L	V
81	F	G	T	S	H	Q	H	M	S	L	S	S	P	P	L	P	S	S	S	P
101	P	P	P	P	A	P	P	S	Q	Q	G	E	C	A	M	P	D	S	Y	L
121	N	G	F	Y	R	R	N	P	S	M	Q	G	D	Y	H	A	G	P	P	R
141	M	N	P	P	M	H	Y	G	S	P	E	P	S	Y	S	S	R	V	S	L
161	S	K	S	M	P	R	G	E	G	S	N	F	Q	H	R	P	Y	P	S	S
181	H	P	P	P	P	P	P	S	H	H	Y	S	Y	M	E	P	D	H	H	I
201	K	S	R	R	E	G	L	S	Y	P	H	R	S	H	Y	T	L	E	F	D
221	E	R	N	Y	Q	D	S	Y	E	R	M	R	P	E	P	C	E	N	R	D
241	N	W	R	Y	H	P	P	S	S	H	G	P	R	Y	H	D	R	H	K	G
261	P	H	Q	S	S	S	Y	S	G	H	H	R	D	S	G	R	L	Q	N	N
281	R	W	S	D	S	P	R	A	Y	N	N	R	H	S	Y	H	Y	K	Q	H
301	S	E	G	P	V	P	V	G	M	R	D	P	G	T	W	H	Q	R	*	

Figure 3.15: Amino acid sequence of CTHULK3cthulk2PRR.

Amino acid sequence of the CTHULK3hulk2PRR protein. Red sequences represent the C-terminus of HULK2 and the blue sequences represents the PRR of CT-HULK3.

3.5 Transcriptional activity of the domain-swapped proteins

Colonies expressing CTHULK3hulk2PRR, which carries HULK2's potent PRR, were expected to grow at 3-AT concentrations similar to that observed for CT-HULK2, namely 250 mM. On the contrary, colonies expressing CTHULK2hulk3PRR, carrying HULK3's PRR, were only expected grow at concentrations of 3-AT up to 10 mM.

Colonies of MaV203 were transformed with CTHULK3hulk2PRR, CTHULK2hulk3PRR, CT-HULK2, and CT-HULK3. They were then plated on YNB – leucine –hisitidine medium with varying concentrations of 3-AT. Colonies expressing CTHULK3hulk2PRR, CT-HULK3 carrying HULK2's potent PRR, were able to grow in 100 mM of 3-AT, and are considered medium strength activation domains. Similarly, colonies expressing CTHULK2hulk3PRR grew on medium containing 75 mM of 3-AT, thus were in the range on medium strength activation domains (Figure 3.16). Colonies with CT-HULK3 were only able to grow on plates with up to 10 mM of 3-AT while colonies with CT-HULK2 were able to grow on plates with up to 300 mM of 3-AT (Figure 3.16).

These results show that while the PRR contributes to the transcriptional activity of the proteins, since large changes of autoactivation strength were observed, they are not the only source of this activity. If the PRR of HULK2 was solely responsible for the transcriptional activity, it should have been able to allow growth of MaV203 transformed with HULK3hulksPRR on plates with much higher levels of 3-AT than observed by colonies with CT-HULK2.

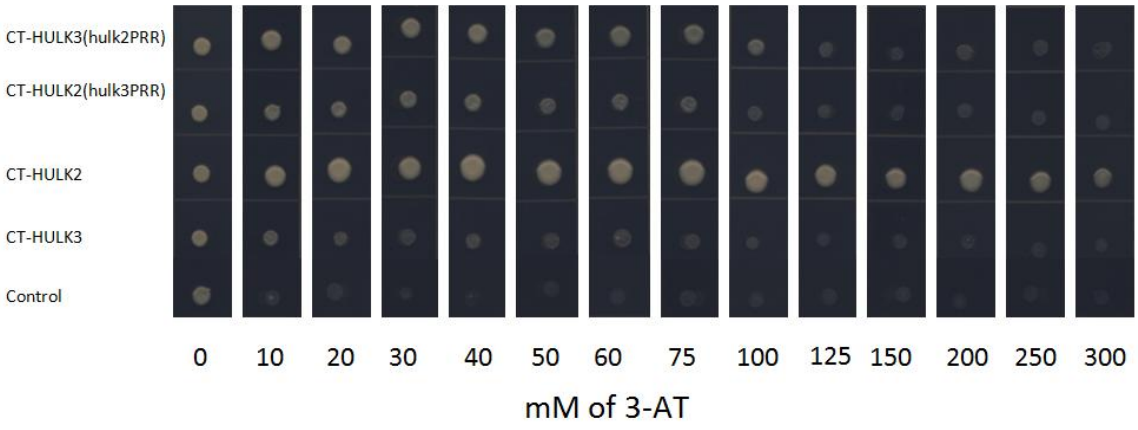

Figure 3.16: 3-AT titration of PRR domain-swapped CT-HULK2, CTHULK2hulk3PRR, and CT-HULK3, CTHULK3hulk2PRR.

Yeast expressing CT-HULK2 and CT-HULK3 as well as CT-HULK2 with its PRR switched with CT-HULK3's PRR and vice versa were grown on media with increasing concentrations of 3-AT to identify the concentration at which autoactivation of the *HIS3* gene would be quenched, resulting in no growth. The control was empty pDBLeu vector.

Similarly, CT-HULK2 without its PRR, CTHULK2hulk3PRR should not have been able to grow in the presence of up to 100 mM of 3-AT if its absent strong PRR was solely responsible for transcriptional activity.

Therefore, the total transcriptional activity observed in HULK2 and CT-HULK2 could be the combined strength of multiple transcriptional activation domains in the carboxyl-terminus, in the same manner that the reduction of autoactivation strength between full-length HULK3 and CT-HULK3 could have been attributed to another unidentified TAD outside of the carboxyl-terminus.

3.6 Bioinformatics analysis of carboxyl-terminus ends of CT-HULK2 and CT-HULK3

After investigating the autoactivation strength of the PRR of CT-HULK2 and CT-HULK3 through the domain swap experiment, the data suggested that the transcriptional activity observed could be attributed to another unidentified TAD in the carboxyl-terminus of HULK2. Sequences around the PRR of both CT-HULK2 and CT-HULK3 were further examined for the presence of putative transcriptional activation domains.

A blastp search (NCBI-BLAST) was used initially to see if the pre-PRR and post-PRR sequences of both CT-HULK2 and CT-HULK3 would match any known TAD. The blastp compares a given protein sequence to the NCBI protein database and returns the most similar sequence but can also locate known conserved domains in the sequence. Both the pre-PRR and post-PRR sequences of both CT-HULK2 and CT-HULK3 were input as search queries, but the sequences did not yield any putative conserved domains. The best matched sequences (highest total score, highest query coverage and lowest E-value) to the pre-PRR and post-PRR sequences of CT-HULK2 are the full length HULK2

sequence. The best match for the pre-PRR and post-PRR sequences of CT-HULK3 is the full length HULK3 sequence.

Analysis of the pre-PRR of CT-HULK2 and pre-PRR of CT-HULK3 with SMART (EMBL) to identify and analyze protein domains did not lead to the identification of domains, repeats, motifs, or other features. On the other hand, for the sequences of both the post-PRR sequences of CT-HULK2 and CT-HULK3, various regions of low complexity were identified (Table 3.1). These sequences were detected by the SEG program, which filters out regions of low complexity for searches with protein sequences in SMART. Subsequently, another blastp search was performed with the sub-sequences to see if they correspond to any conserved domain or protein. Unfortunately, this second round did not yield any hits.

Similar to the SMART search, the Conserved Domain- (CD-)Search Tool (NCBI) compares a protein sequence to a large database of conserved domains and full-length proteins. Inputting the pre- and post-PRR sequences of CT-HULK2 and CT-HULK3 did not return any identified conserved domains. Since the CD-Search Tool uses the same SEG program as SMART to filter out segments of query sequence with low compositional complexity, the same sequence segments were identified (Table 3.1).

While no conserved domains were located around the PRRs of CT-HULK2 and CT-HULK3, 3 sequences of low compositional complexity were identified. These sequences are normally not included in searches as they tend to be homopolymeric runs or short repeats, such as poly-A tails or proline-rich regions. However, as the HULKs are purported to carry proline-rich TADs, these sequences picked outside the PRR could attribute to some of the transcriptional activity observed during the yeast one-hybrids.

Table 3.1: SMART and CD-Search Tool results for identification and analysis of protein domains.

The amino acid sequences of the pre-PRR and post-PRR of both CT-HULK2 and CT-HULK were analyzed using the normal mode of SMART and the CD-Search. The sequences identified were considered to have low compositional complexity and were detected by the SEG program in SMART and the low-complexity filter of CD-Search.

Query sequence	Sequence identified
CT-HULK2 post-PRR	QHRPYPSHPHPHPPPPPPPPQHQ
	RYPPSSSYGSRY
CT-HULK3 post-PRR	PYPSSHPPPPPPSHHYSY

Chapter 4 : Discussion

In yeast one-hybrid assays, the strength of putative TADs, the autoactivation strength, is tested through the detection and quantification of upregulated reporter genes which is proportionate to the transcriptional activity of TADs. The putative TAD in the HULK protein family was tested through 3-AT titrations in a yeast one-hybrid system. Furthermore, the PRRs of both HULK2 and HULK3 were further examined in their capacity to be responsible for the transcriptional activity found in the HULK protein family.

4.1 Sensitivity of *HIS3* reporter gene

Titz et al. (2006) performed a study involving approximately 6000 yeast proteins in a yeast one-hybrid system to identify 451 previously uncharacterized transcription factors. In this study studying the autoactivation strength of the HULK protein family, the levels of autoactivation were compared to standards that were set by the work done by Titz et al. (2006), in which weak autoactivation was transcriptional activity in 3-AT concentrations of 3 – 25 mM, medium strength activation domains induced transcription between 50 – 200 mM 3-AT and strong auto-activators induced transcription at 3-AT concentrations greater than 200 mM.

However, it should be noted that Titz et al.'s (2006) putative activators were tested in the yeast strain YULH, versus the strain MaV203 used in this study. While both the MaV203 and YULH strains share the *HIS3* reporter, there is a difference in the amount of 3-AT required to quench autoactivation between different yeast strains (Gietz et al., 1997; Van Criekinge and Beyaert, 1999).

While no direct comparisons are found between the concentrations of 3-AT required for quenching of YULH compared to MaV203, MaV203 generally requires more 3-AT since it was designed to express a low level of the *HIS3* gene product. Concentrations of up to 100 mM of 3-AT have been commonly used previously and in many cases, concentrations of about 10 mM 3-AT were sufficient to quench the low basal expression of the *HIS3* gene for mAV203 (Guthrie, 2002; Walder et al., 2002; Prodoehl, 2007). On the other hand, studies with YULH have used approximately 3 mM of 3-AT to quench the basal expression of *HIS3* (Uetz et al., 2000; Titz et al., 2006).

While the *HIS3* gene was used as the reporter gene for my study, there are a variety of other reporter genes available. Other assays may use β-galactosidase or green fluorescent protein (GFP) to report expression. In measuring luminescence, the intensity of the colour produced by GFP, the production of the GFP can be associated with activation strength (engineer). More quantitative in its measurement, the expression of β-galactosidase in the presence of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) induces a blue colour change in yeast (Titz et al., 2006). The time for colour change to reach a certain optical density, can be mathematically correlated to activation strength (Titz et al., 2006).

4.2 High transcriptional activity in members of the HULK protein family

Through the yeast one-hybrid assay a large variation was observed in the autoactivation strength of full length HUA2, HULK2, and HULK3 proteins. While HUA2 had a low activation strength of up to 10 mM of 3-AT, neither HULK2 and

HULK3's autoactivation of the *HIS3* gene was quenched until 200 mM and 250 mM of 3-AT respectively. Comparing these observations to the standard, set by Titz et al. (2006), HUA2 is classified as a weak transcription factor and both HULK2 and HULK3 are classified as strong transcriptional activators. While the 3-AT concentrations at which growth were observed in the yeast one-hybrids are inflated when compared to the standard based on 3-AT titrations in YULH, concentrations upwards of 200 mM of 3-AT are considered extremely high since the vast majority of assays use less than 10 mM of 3-AT.

Since the HULKs are homologous proteins, with a shared domain structure, and are part of the same family (Challa, 2009), it was expected that their putative transcriptional activation domain would have the same autoactivation strength. However, a previous phylogenetic analysis of the gene family showed that the *HULK*s are split into two paralogous pairs; *HUA2* and *HULK1* are grouped in one clade, and *HULK2* and *HULK3* in another clade (Challa, 2009). Given that *HULK2* and *HULK3* belong to the same clade and share 60.5% sequence identity and 77% sequence similarity (Challa, 2009), this could explain the similar levels of autoactivation between *HULK2* and *HULK3* as compared to the lower autoactivation of *HUA2* which shares with *HULK2* 27.1% sequence identity and 37.7% sequence identity with HULK3. The differences in amino acid sequences may affect the physical process by which the linear amino acid chains fold into the functional three-dimensional tertiary structure. With large differences in sequence identity, it is expected that there will be differences in conformation and function. As a result, differences in autoactivation strength between members of the

HULK protein family may occur, especially since HUA2 contains larger stretches of proline residues than both HULK2 and HULK3.

Although large tracts of prolines can increase the performance of transcription factors (Gerber et al., 1994; Lin et al., 2010), as observed for HULK2, it is interesting to observe that full-length HUA2 does not have a strong level of autoactivation in the yeast one-hybrid. The weak level of autoactivation in full-length HUA2 could be a result of the physical properties of the amino acids. Due to the bulkiness of the proline ring structure, the amino acid's own flexibility is restricted as well as the preceding residue's conformation, limiting the set of possible conformations in the tertiary structure of the protein (Williamson, 1994; Kay et al., 2000). This trait results in rigid stretches of proline residues that are more favourable for the binding found in protein-protein interactions, but at the cost of being a weaker binding site (Kay et al., 2000). Thus, while HUA2 has larger polyproline stretches and may have binding sites, it may not be able to bind long enough to members of the basal transcription machinery to upregulate transcription as effectively as HULK2 or HULK3.

4.2.1 Subcloning of *HULK1* in pDBLeu

A large amount of time was used to attempt to troubleshoot the subcloning of *HULK1* into pDBLeu. Often, after transformation of DH5α with the ligation product of *HULK1* and pDBLeu, there were few or no colonies. The colonies that were found were either carrying empty pDBLeu or did not actually carry *HULK1* when checked with a digestion with restriction enzymes.

Of the HULKs, *HULK1* is the largest gene at 4533 bp. *HULK1*'s large size is a possible cause for the difficulties during the subcloning. Large molecules of DNA are

more likely to experience mutations. As well, pDBLeu is a fairly large vector which makes transformation more difficult as compared to smaller plasmids.

4.2.2 Transcriptional activity in the HULK protein family carboxyl-terminus

The 3-AT titrations of the carboxyl-terminus ends of the HULK protein family members (Figure 3.3) showed a strong increase in the autoactivation strength of CT-HUA2 while CT-HULK3's ability to enhance transcription of the *HIS3* was dramatically reduced.

What is significant about the carboxyl-terminus of HULK proteins is the abundance of proline residues. Analysis of the amino acid content of CT-HUA2 reveals that prolines are the most abundant residue, making up over 22% of the sequence. In CT-HULK2, prolines are again the most abundant residue at over 15%, and in CT-HULK3, prolines are the second most abundant residue constituting close to 12% of the carboxyl-terminus. Furthermore, there are four PPLP repeats in CT-HUA2, one PPLP repeat in CT-HULK2, and none in CT-HULK3.

In associating autoactivation strength and abundance of proline residues and PPLP repeats, a trend appears in that the protein fragment with the highest amount of proline residues and number of PPLP repeats exhibits the strongest autoactivation strength. Both polyproline stretches and PPLP repeats are known for being involved in protein-protein interactions and a higher abundance of either would result in more potential binding sites for members of the basal transcription machinery.

While the large increase in transcriptional activity in HUA2 could be attributed to both its high concentration of proline residues and its PPLP repeats, it could also be due

to a potential decrease in steric hinderance. A truncation of the HUA2 protein, CT-HUA2, may have opened more active binding sites that were once blocked by the conformation of the full-length protein. Similarly, the dramatic reduction of HULK3's autoactivation strength could be attributed to a weak proline-rich activation domain, but also the truncation may have removed a more potent acidic or glutamine-rich activation domain found outside of the carboxyl-terminus. A reduction in the size of the truncation in a future experiment may result in a larger fragment of HULK3 that has recovered its high autoactivation strength.

While changes were observed in the autoactivation strength of both CT-HUA2 and CT-HULK3, CT-HULK2 had similar autoactivation strength to full-length HULK2, with perhaps a small increase in autoactivation strength. CT-HULK2's PRR does not contain the greatest percentage of proline residues; however, in terms of number of residues, it has the highest number of prolines, 11, out of all members of the HULKs. The literature does not specify a requisite number of proline residues for a domain to be considered a proline-rich activation domain, but HULK2's PRR considerable and consistent strength could indicate that this PRR could be a transcriptional activation domain.

4.3 Transcriptional activity of domain-swapped proteins

If the PRR and the PPLP repeats are responsible for the transcriptional activation observed in the yeast one-hybrids, then switching PRRs with different transcriptional activation strengths should also change the autoactivation strength of new constructs in subsequent yeast one-hybrids. The domain swap experiment replaced the strongest PRR, HULK2's, with the weakest PRR, that of HULK3's. This resulted in an increase in the

autoactivation strength of the chimeric protein carrying the HULK2 PRR, CTHULK3hulk2PRR, and a decrease in the autoactivation strength of the chimeric protein carrying the HULK3 PRR, CTHULK2hulk3PRR (Figure 3.4).

The transcriptional activity was attributed to the amount of proline residues and the number of PPLP repeats in the PRR. CT-HULK2's 43 amino acid PRR is composed of 25% proline residues and one PPLP motif, resulting in high autoactivation strength. Conversely, CT-HULK3's PRR 33 amino acid PRR is approximately 11% proline and does not carry a PPLP motif and is one of the weakest putative transcription factors in the HULK protein family.

If transcriptional activity were solely controlled by the PRR, then by exchanging the PRR of CT-HULK2 and CT-HULK3, one would expect that the autoactivation strength of the new chimeric proteins would reflect that of the protein originally carrying the PRR. However, this was not the case as both CTHULK3cthulk2PRR and CTHULK2cthulk3PRR were found to be medium strength transcription factors (50 mM – 200 mM 3-AT) rather than being strong or weak autoactivators like CT-HULK2 and CT-HULK3.

The results show that while the PRR does contribute to the transcriptional activity of the proteins, since large changes of autoactivation strength were observed, they are not the sole source of this activity. Therefore, the total transcriptional activity observed in HULK2 and CT-HULK2 could be the result of multiple transcriptional activation domains working in concert, similar to how the reduction of autoactivation strength between full-length HULK3 and CT-HULK3 could have been attributed to the absence of other unidentified TADs.

The chimeric protein CTHULK2hulk3PRR carries the HULK3 PRR, but is able to upregulate the *HIS3* reporter gene to a higher level than CT-HULK3. Thus, another segment of the HULK2 carboxyl-terminus may also code for a yet unidentified TAD. Several programs were used to analyze the protein sequence, and did detect any conserved domain. However, two sequences of low compositional complexity were identified (Table 3.1). Low compositional complexity regions are filtered out because they often represent non-random compositional biased sequences (Wootton and Federhen, 1993; Altschul et al., 1994). Yet, these sequences are often acidic-, basic- or proline-rich sequences (Wootton and Federhen, 1993). Even though a subsequent search with the sub-sequences identified in both the SMART search and CD-Search Tool did not match any known conserved domain, the sequences have the potential to be transcriptional activation domains. One sequence contains over 50% proline residues and so could be another putative proline-rich transcriptional activation domain and may contribute to an increased capacity for binding (Williamson, 1994). Interestingly, it has been noted in the literature that proteins with sequences of low compositional complexity found towards the centre of the protein (more than 25 amino acids from either sequence ends) tend to have more protein-protein binding partners than proteins without low compositional complexity (Coletta et al., 2010). Furthermore, within *S. cerevisiae*, proteins with sequences of low compositional complexity have a tendency to have more roles in transcription, transcription regulation, and translation (Coletta et al., 2010).

While the yeast one-hybrid assays with CTHULK3hulk2PRR showed an increase in transcriptional activity (weak transcription factor to a medium strength transcription factor), CTHULK3hulk2PRR did not induce transcriptional activation to the extent that

CT-HULK2 could, despite carrying the stronger PRR. While the conformation of the chimeric protein could be involved and affect the transcriptional activity of the PRR of CT-HULK2, both SMART and CD-Search Tool searches indicate that the carboxyl-terminus of CT-HULK2 has a low compositional complexity region not found in CT-HULK3 (Table 3.1). Furthermore, the low compositional complexity regions of CT-HULK2 have more proline residues (one region with 12 proline residues and one region with 2 proline residues) than CT-HULK3's single low compositional complexity region (containing 7 proline residues). These regions, which are absent in CT-HULK3, could be contributing to the high transcriptional activity observed in CT-HULK2.

4.4 Conclusion and further studies

My project involved studying the transcriptional activating ability of HULK proteins of *A. thaliana* in the yeast, *S. cerevisiae* through the modularity and ability of chimeric transcription factors to be interchangeable between a plant anNeast. Members of the HULK protein family do show transcriptional activation through upregulation of the *HIS3* reporter gene in a yeast one-hybrid system at varying levels. While it was established that the PRR found at the carboxyl-terminus ends of the proteins contributed to the transcriptional activity, it was determined that the domain is not solely responsible for the transcriptional activity. The presence of multiple transcriptional activation domains working in tandem are likely explanation for the high levels of autoactivation.

Currently, no conserved domains can be detected using several web-based programs. However, several sequences not associated with the previously identified PRR were highlighted in the carboxyl-terminus (Table 3.1). While these sequences do not cover the total sequences of the putative acidic residue-rich or proline/serine/asparagine

residue-rich domains as predicted by Chen and Meyerowitz (1999), they may contribute to the transcriptional activity found in the HULKs. In the deletion constructs used for yeast one-hybrids, constructs either did not include the acidic residue-rich domain or began part way through the putative domain. Not only would this make it difficult to predict the presence of the domain, but could also render the protein useless.

Applying yeast one-hybrid assays with 3-AT titration using constructs that include the putative acidic activation domain could be used to assess whether all the HULKs have a strong transcription factor. The assays may also determine if the reduction in autoactivation strength in the carboxyl-terminus of HULKs was due to the exclusion of this putative domain. Thus, while the PRR of the HULK may contribute to their transcriptional activation ability, a number of other sequences in the carboxyl-terminus may also be transcriptional activation domains waiting to be found in HULK2. Also, in HULK3, the activation domain may be in the amino-terminus of the protein.

References

Altschul, S. et al. (1994). Issues in searching molecular sequence databases. *Nat. Genet.* **6**: 119–129.

Alvarez, J. and Smyth, D.R. (1999). CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. *Development* **126**: 2377–86.

Barberis, A. and Petrascheck, M. (2003). Transcription Activation in Eukaryotic Cells. *eLS*.**58**: 138 - 189

Bedford, M.T. and Leder, P. (1999). The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem Sci.* **24**: 264–265.

Biddick, R. and Young, E.T. (2005). Yeast mediator and its role in transcriptional regulation. *C R Biol.* **328**: 773–82.

Blau, J. et al. (1996). Three Functional Classes of Transcriptional Activation Domains. *Mol Cell Biol.* **16**: 2044–2055.

Bouwman, P. and Philipsen, S. (2002). Regulation of the activity of Sp1-related transcription factors. *Mol Cell Endocrinol.* **195**: 27–38.

Brisco, P.R.G. and Kohlhaw, G.B. (1990). Regulation of Yeast LEU2. *J. Biol. Chem.* **265**: 11667–11675.

Cantin, G.T. et al. (2003). Activation domain-mediator interactions promote transcription preinitiation complex assembly on promoter DNA. *Proc. Natl. Acad. Sci. USA* **100**: 12003–8.

Challa, S.S. (2009). Genetic and Phenotypic Analysis of the HUA2-LIKE (HULK) Gene Family in Arabidopsis thaliana. London, Ont.: Department of Biology, University of Western Ontario.

Chen, G. and DenBoer, L. (2008). Design of a single plasmid-based modified yeast one-hybrid system for investigation of in vivo protein-protein and protein-DNA interactions. *BioTechniques* **45**: 295–304.

Chen, X. and Meyerowitz, E.M. (1999). HUA1 and HUA2 are two members of the floral homeotic AGAMOUS pathway. *Mol. Cell.* **3**: 349–360.

Cheng, Y. et al. (2003). Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in Arabidopsis thaliana. *Dev cell* **4**: 53–66.

Choi, K. et al. (2011). The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *Plant Cell* **23**: 289–303.

Coletta, A. et al. (2010). Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst Biol.* **4**: 43.

Van Criekinge, W. and Beyaert, R. (1999). Yeast Two-Hybrid: State of the Art. *Biol Proced Online* **2**: 1–38.

Doerks, T. et al. (2002). Systematic Identification of Novel Protein Domain Families Associated with Nuclear Functions. *Genome Res.* **12**: 47–56.

Doyle, M.R. et al. (2005). HUA2 is required for the expression of floral repressors in Arabidopsis thaliana. *Plant J.* **41**: 376–85.

Emili, A. et al. (1994). Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. *Mol Cell Biol.* **14**: 1582–93.

Escher, D. et al. (2000). Conservation of Glutamine-Rich Transactivation Function between Yeast and Humans Conservation of Glutamine-Rich Transactivation Function between Yeast and Humans. *Mol Cell Biol.* **20**: 2774–2782.

Fields, S. and Song, O. (1989). A novel genetic system to detect protein protein interactions. *Nature* **20**: 245–246.

Garvie, C.W. and Wolberger, C. (2001). Recognition of Specific DNA Sequences Review. *Mol. Cell* **8**: 937–946.

Gehring, M. and Henikoff, S. (2008). DNA Methylation and Demethylation in Arabidopsis. *The Arabidopsis Book* **58**.

Gerber, A.H. et al. (1994). Activation Modulated by Transcriptional Homopolymeric Glutamine and Proline Stretches. *Science* **263**: 808–811.

Gietz, R.D. et al. (1997). Identification of proteins that interact with a protein of interest: applications of the yeast two-hybrid system. *Mol. Cell. Biochem.* **172**: 67–79.

Green, M.R. (2005). Eukaryotic transcription activation: right on target. *Mol. Cell* **18**: 399–402.

Guthrie, C. (2002). Guide to Yeast Genetics and Molecular Cell Biology, Volume 350 (Academic Press).

Hermann, S. et al. (2001). How transcriptional activators bind target proteins. *J. Biol. Chem.* **276**: 40127–32.

Invitrogen (2005). ProQuest ™ Two-Hybrid System Manual.

Irish, V.F. (2010). The flowering of Arabidopsis flower development. *Plant J* **61**: 1014–28.

Kadonaga, J. (2004). Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. *Cell* **116**: 247–257.

Kadonaga, J.T. (1990). Gene transcription :. basal and regulated RNA polymerase II transcription by. *Curr Opin Cell Biol* **2**: 496 – 501.

Kay, B. et al. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB Journal*: 231–241.

Kim, T. (2000). Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH. *Science* **288**: 1418–1421.

Komeda, Y. (2004). Genetic regulation of time to flower in Arabidopsis thaliana. *Annu. Rev. Plant Biol.* **55**: 521–35.

Koornneef, M. and Vries, H. (1994). The phenotype of some late • flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg erecta wild- • type. *Plant J.* **6**: 911–919.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* **30**: 235–9.

Künzler, M. et al. (1994). Functional differences between mammalian transcription activation domains at the yeast GAL1 promoter. *EMBO J.* **13**: 641–5.

Lamesch, P. et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**: D1202–10.

Li, R. (1999). Stimulation of DNA replication in Saccharomyces cerevisiae by a glutamine- and proline-rich transcriptional activation domain. *J. Biol. Chem.* **274**: 30310–4.

Lin, C.-H. et al. (2010). A tryptophan-rich peptide acts as a transcription activation domain. *BMC Mol. Biol.* **11**: 85.

Ma, J. (2011). Transcriptional activators and activation mechanisms. *Protein & Cell* **2**: 879–888.

Ma, J. and Ptashne, M. (1987). Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48**: 847–53.

Macias, M.J. et al. (2002). WW and SH3 domains , two different scaffolds to recognize proline-rich ligands. *FEBS Letters* **513**: 30–37.

Mayr, B.M. et al. (2005). Glutamine rich and basic region/leucine zipper (bZIP) domains stabilize cAMP-response element-binding protein (CREB) binding to chromatin. *J. Biol. Chem.* **280**: 15103–10.

Meinke, D.W. (1998). Arabidopsis thaliana: A Model Plant for Genome Analysis. *Science* **282**: 662–682.

Mermod, N. et al. (1989). The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58**: 741–53.

Meyerowitz, E.M. (2001). Prehistory and history of Arabidopsis research. *J. Plant Physiol.* **125**: 15–9.

Mitchell, P.J. and Tjian, R. (1989). Transcriptional Regulation in Mammalian Cells by DNA Binding Proteins. *Science* **245**: 371–378.

Moon, J. et al. (2005). Analysis of flowering pathway integrators in Arabidopsis. *Plant Cell Physiol.* **46**: 292–9.

Mouradov, A. (2002). Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* **14**: 111–131.

Olszewski, N. et al. (2002). Gibberellin signaling: biosynthesis, catabolism, and response pathways. *Plant biotechnology journal* **14**: 61–81.

Piskacek, S. et al. (2007). Nine-amino-acid transactivation domain: establishment and prediction utilities. *Genomics* **89**: 756–68.

Poduska, B. et al. (2003). The synergistic activation of FLOWERING LOCUS C by FRIGIDA and a new flowering gene AERIAL ROSETTE 1 underlies a novel morphology in Arabidopsis. *Genetics* **163**: 1457–1465.

Posé, D. et al. (2012). The end of innocence: flowering networks explode in complexity. *Curr Opin Plant Biol* . **15**: 45–50.

Prado, F. et al. (2002). Differential role of the proline-rich domain of nuclear factor 1-C splice variants in DNA binding and transactivation. *J. Biol. Chem.* **277**: 16383–90.

Prodoehl, M.J. (2007). Functional Analysis of the Deubiquitylating Enzyme Fat Facets in Mouse in Protein Trafficking (University of Adelaide, School of Molecular and Biomedical Sciences, Discipline of Biochemistry).

Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature* **335**: 683–689.

Ptashne, M. and Gann, A. (1997). Transcriptional activation by recruitment. *Nature* **386**: 569 – 577.

Remacle, J.E. et al. (1997). Three classes of mammalian transcription activation domain stimulate transcription in Schizosaccharomyces pombe. *EMBO J.* **16**: 5722–9.

Sadowski, I. et al. (1988). GAL4- VP16 is an unusually potent transcriptional activator. *Nature* **335**: 563–564.

Sajja, U.B. (2009). Molecuarl Characterization of Arabidopsis Flowering Regulator HUA2, a Putative Pre mRNA Processing Factor. London, Ont.: Department of Biology, University of Western Ontario.

Schwechheimer, C. et al. (1998). The activities of acidic and glutamine-rich transcriptional activation domains in plant cells: design of modular transcription factors for high-level expression. *Plant Mol. Biol.* **36**: 195–204.

Slater, L.M. et al. (2003). Structural variation in PWWP domains. *J. Mol. Biol* **330**: 571–576.

Srikanth, A. and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. *Cell Mol Life Sci.* **68**: 2013–37.

Sullivan, S.M. et al. (1998). Mutational analysis of a transcriptional activation region of the VP16 protein of herpes simplex virus. *Nucleic Acids Res.* **26**: 4487–96.

Tell, G. et al. (1998). Structural and functional properties of the N transcriptional activation domain of thyroid transcription factor-1: similarities with the acidic activation domains. *Biochem J.* **329**: 395–403.

Titz, B. et al. (2006). Transcriptional activators in yeast. *Nucleic Acids Res.* **34**: 955–67.

Uetz, P. et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**: 623–7.

Walder, K. et al. (2002). Tanis: a link between type 2 diabetes and inflammation? *Diabetes* **51**: 1859–66.

Walhout, A. and Vidal, M. (1999). A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. *Genome Res.*: 1128–1134.

Wang, Q. et al. (2007). HUA2 caused natural variation in shoot morphology of A. thaliana. *Curr. Biol.* **17**: 1513–9.

Wang, Y. et al. (2009). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Mol. Cell* **33**: 428–37.

Weigel, D. et al. (1992). LEAFY controls floral meristem identity in Arabidopsis. *Cell* **69**: 843–59.

Williamson, M.P. (1994). The structure and function of proline-rich regions in proteins. *Biochem J.* **297**: 249–260.

Wootton, J. and Federhen, S. (1993). Statistics of local complexity in amino acid sequences and sequence databases. *Comput. Chem.* **17**: 149–163.

Xi, H. et al. (2007). Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res.* **17**: 798–806.

Appendices

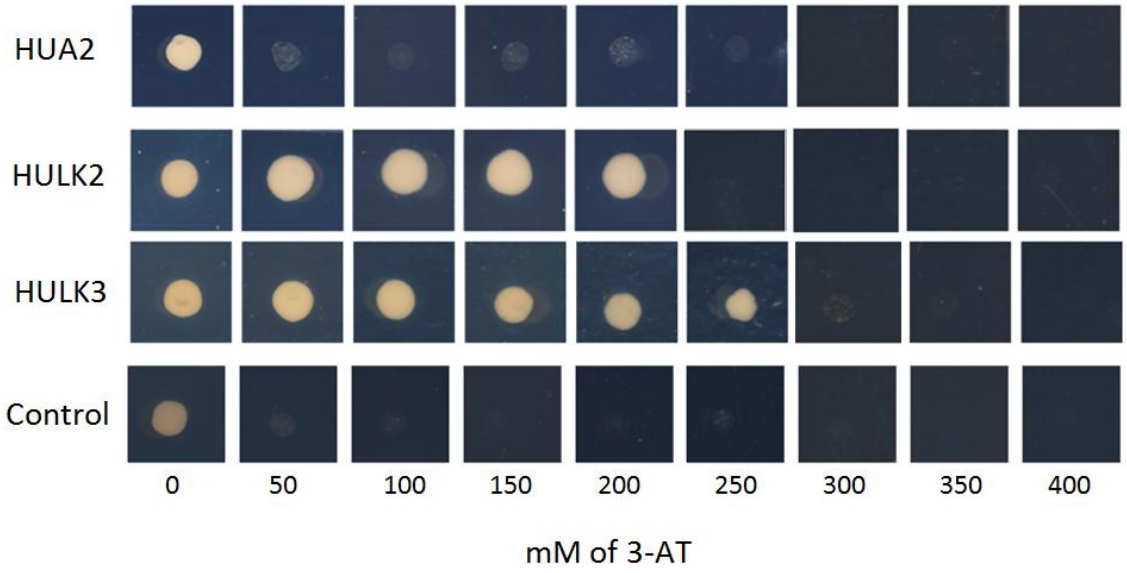

Figure A1.9: 3-AT titrations with full length members of HULK protein family.

Yeast expressing the full-length proteins of the HULK protein family were grown on media with increasing concentrations of 3-AT to identify the concentration at which autoactivation of the *HIS3* gene would be quenched, resulting in no growth. The control was empty pDBLeu vector.

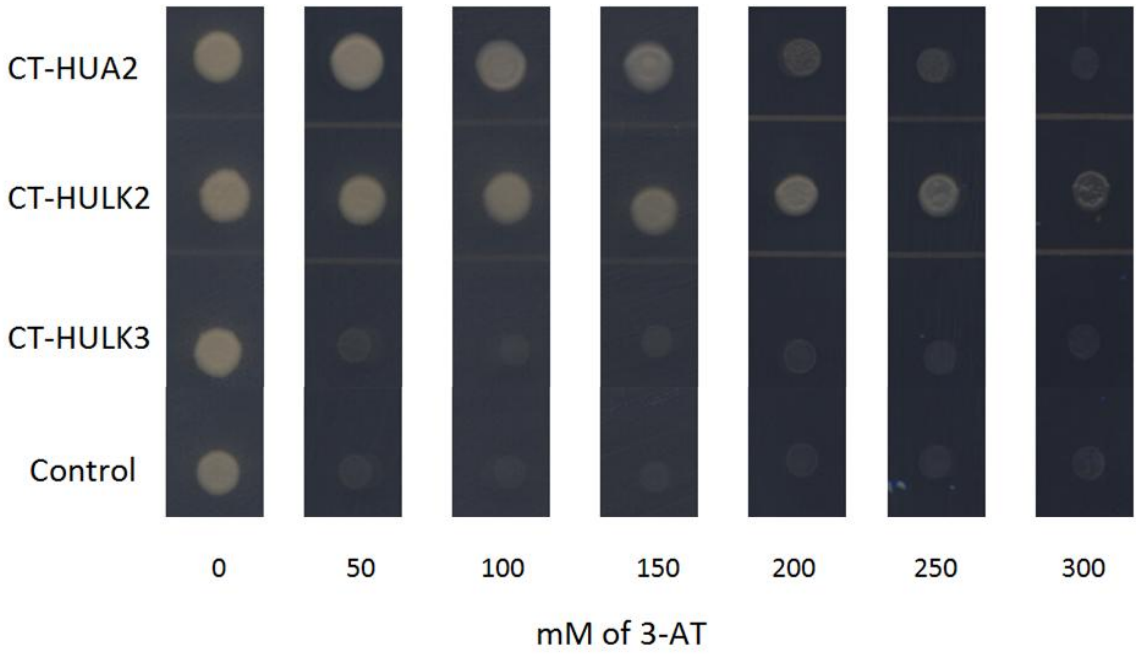

Figure A2: 3-AT titrations with full length members of HULK protein family.

Yeast expressing the C-terminal ends of the HULK protein family were grown on media with increasing concentrations of 3-AT to identify the concentration at which autoactivation of the *HIS3* gene would be quenched, resulting in no growth. The control was empty pDBLeu vector.

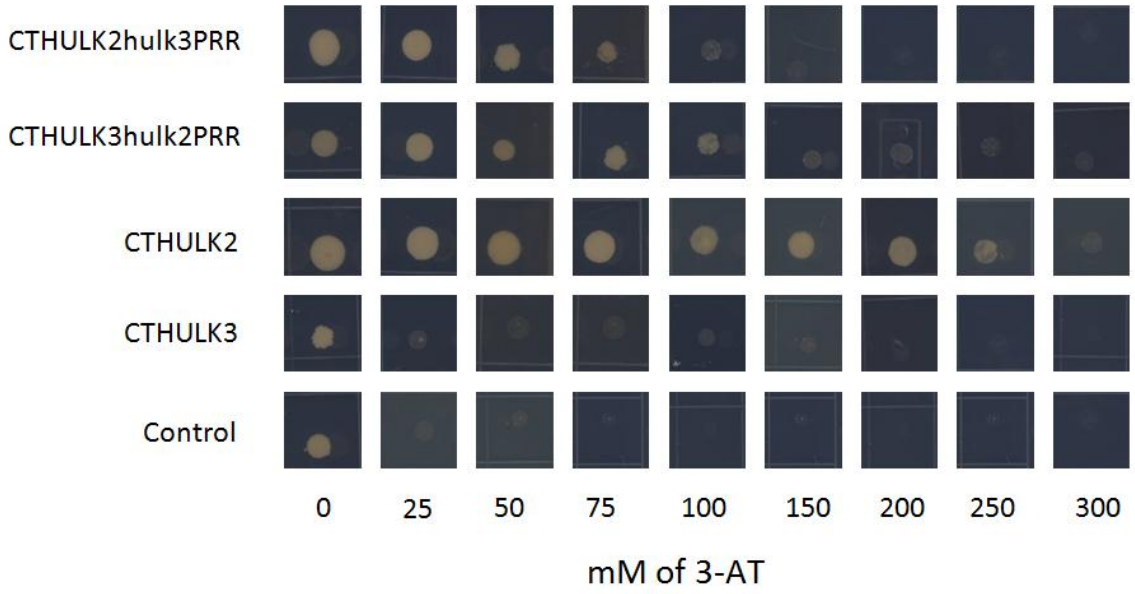

Figure A3: 3-AT titration of PRR domain-swapped CT-HULK2, CTHULK2hulk3PRR, and CT-HULK3, CTHULK3hulk2PRR.

Yeast expressing CT-HULK2 and CT-HULK3 as well as CT-HULK2 with its PRR switched with CT-HULK3's PRR and vice versa were grown on media with increasing concentrations of 3-AT to identify the concentration at which autoactivation of the *HIS3* gene would be quenched, resulting in no growth. The control was empty pDBLeu vector.

Curriculum Vitae

Name:	Christopher Doan
Post-secondary Education and Degrees:	University of Toronto Toronto, Ontario, Canada 2004-2008 Hon. B.Sc. (Biology Specialist) The University of Western Ontario London, Ontario, Canada 2009-2013 M.Sc (Cell and Molecular Biology)
Honours and Awards:	Western Graduate Research Scholarship 2009 - 2011
Related Work Experience	Teaching Assistant The University of Western Ontario 2009 - 2013