REGULATORY ROLE OF THE START LIPID/STEROL BINDING DOMAIN IN HOMEODOMAIN TRANSCRIPTION FACTORS FROM PLANTS

by

AASHIMA KHOSLA

B.Sc., University of Delhi, India, 2005
M.Sc., Maharaja Sayaji Rao University of Baroda, India, 2007

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Graduate Biochemistry Group
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2015

Abstract

Class IV homeodomain leucine-zipper transcription factors (HD-Zip TFs) are master regulators of cell-type differentiation in the plant epidermis. These transcription factors contain a putative START (STeroidogenic Acute Regulatory (StAR)-related lipid Transfer) lipid/sterol-binding domain that is hypothesized to link metabolism to gene expression in plant development. This study is focused on two class IV family members that serve as models in many of the experiments: GLABRA2 (GL2) is a key regulator of differentiation in hair cells called trichomes as well as other epidermal cell types in various plant tissues. The second member addressed in this study is PROTODERMAL FACTOR2 (PDF2), which plays a crucial role in epidermal cell specification in shoots. A leading hypothesis is that the START domain, by binding a ligand, controls transcription factor function, analogously to nuclear receptors from mammals. Domain swap experiments indicated that the START domain from both plants and mammals is a conserved ligand-binding motif that is required for transcription factor activity. To further address its function in ligand binding, mutational analysis of the START domain of GL2 was performed. Several of the mutations remove charged residues in the predicted ligand-binding pocket and resulted in loss-of-function phenotypes, suggesting that ligand binding is critical for HD-Zip TF activity. Chromatin immunoprecipitation–based sequencing (ChIP-seq) revealed that the START domain is dispensable for transcription factor binding to DNA. Using a high throughput thermal shift assay to screen a library of pure natural compounds, specific secondary metabolites were identified as putative START domain ligands for PDF2. Experiments in both yeast and *N. benthamiana* demonstrated that the START domain is required for homodimerization of GL2 through its Zip domain. It was also found that the START domains physically interact with RHAMNOSE SYNTHASE I (RHM1). Further, this work provided evidence for a previously elusive redundancy between GL2 and another class IV HD-Zip TF, and unveils a positive feedback loop in the maintenance of the GL2 activity during trichome differentiation. Taken together, these findings support the premise that START domains are central players in metabolic regulatory networks that can modulate transcription factor activity by binding ligands and mediating protein-protein interactions.

REGULATORY ROLE OF THE START LIPID/STEROL BINDING DOMAIN IN HOMEODOMAIN TRANSCRIPTION FACTORS FROM PLANTS

by

AASHIMA KHOSLA

B.Sc., University of Delhi, India, 2005
M.Sc., Maharaja Sayaji Rao University of Baroda, India, 2007

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Graduate Biochemistry Group
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2015

Approved by:

Major Professor
Kathrin Schrick

Copyright

Abstract

Class IV homeodomain leucine-zipper transcription factors (HD-Zip TFs) are master regulators of cell-type differentiation in the plant epidermis. These transcription factors contain a putative START (STeroidogenic Acute Regulatory (StAR)-related lipid Transfer) lipid/sterol-binding domain that is hypothesized to link metabolism to gene expression in plant development. This study is focused on two class IV family members that serve as models in many of the experiments: GLABRA2 (GL2) is a key regulator of differentiation in hair cells called trichomes as well as other epidermal cell types in various plant tissues. The second member addressed in this study is PROTODERMAL FACTOR2 (PDF2), which plays a crucial role in epidermal cell specification in shoots. A leading hypothesis is that the START domain, by binding a ligand, controls transcription factor function, analogously to nuclear receptors from mammals. Domain swap experiments indicated that the START domain from both plants and mammals is a conserved ligand-binding motif that is required for transcription factor activity. To further address its function in ligand binding, mutational analysis of the START domain of GL2 was performed. Several of the mutations remove charged residues in the predicted ligand-binding pocket and resulted in loss-of-function phenotypes, suggesting that ligand binding is critical for HD-Zip TF activity. Chromatin immunoprecipitation–based sequencing (ChIP-seq) revealed that the START domain is dispensable for transcription factor binding to DNA. Using a high throughput thermal shift assay to screen a library of pure natural compounds, specific secondary metabolites were identified as putative START domain ligands for PDF2. Experiments in both yeast and *N. benthamiana* demonstrated that the START domain is required for homodimerization of GL2 through its Zip domain. It was also found that the START domains physically interact with RHAMNOSE SYNTHASE I (RHM1). Further, this work provided evidence for a previously elusive redundancy between GL2 and another class IV HD-Zip TF, and unveils a positive feedback loop in the maintenance of the GL2 activity during trichome differentiation. Taken together, these findings support the premise that START domains are central players in metabolic regulatory networks that can modulate transcription factor activity by binding ligands and mediating protein-protein interactions.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I'd like to give special thanks, beginning with my advisor Dr. Kathrin Schrick, for allowing me to explore random topics. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. She's always been motivating, encouraging, and enlightening. Her patience, flexibility, genuine caring and concern and faith in me during the dissertation process enabled me to attend to life while also earning my Ph.D. Whenever I needed her, she has always been there for me, both as a friend and a mentor. For all this, I cannot thank her enough. I am forever grateful. Thank You Kathrin!

Thank you to my committee members, S. Muthukrishnan, Jeroen Roelofs, and Richard Todd, for helping me to develop my project and offering good advice when I needed it. Thank you to Michael Herman for serving as my outside chairperson. Another thank you to Brian Geisbrecht for helping with the protein purification. Thank you to Lawrence Davis and Timothy Durrett for providing insightful feedback and assistance when I needed it.

Thank you to the many undergraduates that came through the Schrick Lab during my time here: Preston, James, Erika, Amanda, Allison, Will, Emily, Meghan, and Olivia. In particular, thank you to Erika Peters who has shown me how rewarding being a mentor can be and how valuable my experiences are to an up- and-coming scientist. She really demonstrated a hard-work ethic. Thank you to Daniel Stucky, Janet Paper, and Paige Cox, who I have many fond memories with both in and out of the lab. I feel lucky to have worked with people with such great temperament and work ethics.

My gratitude is also extended to Crystal Sapp and Belinda Banter, who have known the answer to every question I've ever asked regarding Graduate school, filing taxes, you name it. I couldn't have completed all the required paperwork and delivered it to the correct place without them.

Next I'd like to thank Meera. I cannot begin to express my gratitude and feelings for this gregarious woman. We've laughed, cried and among other things, cursed together. Meera has played the part of a friend, confidant, conscience, rear end kicker, humorist, and phone comrade. She's been as tough on me as she's been supportive and caring. In her I have a life-long friend and colleague. For all these reasons and many, many more, I am eternally grateful.

Another staunch supporter and fan is a special friend Kuldeep Shetye. His love, support, and belief in me are a treasure. He understands me in a way no one else ever could. Thank you for your encouragement, support and most of all your humor. You always keep things light and me smiling.

Of course, no acknowledgments would be complete without giving thanks to my parents. Both have instilled many admirable qualities in me and given me a good foundation with which to meet life. They've taught me about hard work and self-respect, about persistence and about how to be independent. Thank you to my mom who always listens to me vent without overwhelming me with solutions to the problem. Thank you to my dad and sister for encouraging me and expressing confidence in my abilities when I could only do the opposite. I thank all three of you for your patience and love you more than you will ever know.

Chapter 1 - Comprehensive analysis of the START domain containing homeodomain-leucine zipper (HD-Zip) transcription factors in plants

INTRODUCTION

Metabolites are defined as small organic molecules produced by a living organism as a result of cellular and physiological metabolism. These molecules constitute significant fraction of a cells dry weight, ranging from 17 to 27% in bacterial and mammalian cells respectively (Charbonnier et al. 2008). They consist of a wide variety of small molecules with a vast chemical diversity, including amino acids, nucleotides, sugars, and fatty acids that are central to all metabolic pathways existing in the cell. Overall metabolites constitute the metabolome of a cell, tissue or organism at a specific time and changes in metabolic profiles have enormous potential to understand cellular function (Vinayavekhin et al. 2010). For example, lipids, in particular signaling lipids such as phosphoinositides (PIs), but also eicosanoids, sphingolipids and fatty acids are known to control critical cellular functions and the alteration of lipid mediated pathways is known to contribute to the development of pathologies, such as chronic inflammation, cancer, neurodegenerative, and metabolic diseases (Pendaries et al. 2003; Wymann and Schneiter 2008; Skwarek and Boulianne 2009). It is becoming increasingly clear that the interactions between lipid metabolites and proteins are highly dynamic in the cells of living organisms. To date, the full extent and biological significance of such interactions remains underexplored, especially in plants. This dissertation focuses on transcription factors that are hypothesized to link metabolism to cell differentiation in plants.

Transcription factors (TFs) are key components underlying mechanisms that control gene expression in all living organisms. They define phenotypic diversity and evolutionary adaptation of organisms (Wang et al. 1999; Bustamante et al. 2005). TFs typically consist of at least two major domains. The first is a DNA-binding domain (BD) that recognizes target DNA sequences and the second is a transcriptional activation domain (AD) that initiates transcription by interacting with general transcription factors (Ma and Ptashne 1987). Several TF families consist

of proteins containing a DNA-binding domain known as the homeodomain (HD). The first HD-containing TF was identified in *Drosophila melanogaster* by Garber et al. (1983). The homeotic gene encoding this TF was found to be responsible for the development of antennae at the position of the second leg pair of *Drosophila*. Proteins with highly conserved HDs have been found in other animals, fungi and plants, indicating that these TFs play an important role in developmental pathways in species across different kingdoms (Gehring et al. 1994).

Class III and class IV Homeodomain leucine-zipper transcription factors (HD-Zip TFs) from plants contain a putative START lipid/sterol binding domains (Ponting and Aravind 1999; Schrick et al. 2004). Although Steriodogenic Acute Regulatory protein (StAR)-related lipid transfer (START) domains are amplified in HD-Zip transcription factors (Figure 1.2), the mechanism underlying function of the plant domain in gene transcription is poorly understood, nor has a biological ligand been characterized for any plant-derived START domain to date. Since HD-Zip proteins are transcription factors, a leading hypothesis is that the START domain, by binding a ligand (lipid/small metabolite), controls gene expression analogously to nuclear receptors from the mammals. One advantage of the proposed mechanism is that the metabolic state of a cell would be linked to cell type differentiation. In a broader perspective, such a mechanism may hold true for many organisms that contain START domain proteins with unknown functions. Additionally, two key pieces of evidence suggest a connection to lipid metabolism. Previously, Ohashi and coworkers (Ohashi et al. 2003) have shown that the HD-Zip transcription factor GL2 functions as a negative regulator of phospholipid signaling. The connection between HD-Zip TF activity and lipid metabolism is further indicated by the finding that transcriptional targets of maize OCL1 include genes involved in lipid metabolism and cuticle-related lipid synthesis and transport (Javelle et al. 2010).

START lipid/sterol binding domains

The Bet v 1-like superfamily (also known as the START superfamily) is a large and ancient family of proteins present in eukaryotes and archaea (Radauer et al. 2008). It contains 11 subfamilies that are structurally related to Bet v I. The subfamily containing highest number of structures is the Bet v 1 family followed by the START subfamily, named after the **Steroidogenic Acute Regulatory (StAR) protein-related lipid transfer (START) domain**. Members of this larger superfamily include the newly identified abscisic acid (ABA) receptors

from plants (Ma et al. 2009; Park et al. 2009). The PYR and PYL proteins belong to the Polyketide Cyc2 subfamily that shares no sequence similarity *per se*, but shares structural features with members of the START subfamily.

The START domain is an evolutionary conserved protein module that binds lipids including sterols (Ponting and Aravind 1999; Schrick et al. 2004). START domains are ubiquitous to plants and animals and also occur in some members of the bacteria and protista, but are absent from archaebacteria. They were first characterized in the mammalian StAR-related protein that binds and transfers cholesterol to the inner mitochondrial membrane for initiation of steroidogenesis (Arakane et al. 1998). In multicellular organisms, START domains are often part of multidomain proteins. They are fused to other signaling domains including pleckstrin homology domain (PH), Rho-GAP domain, SAM/sterile alpha motif domain and the homeodomain (Soccio 2003). Several reports have highlighted their role as a lipid exchange and/or a lipid-sensing domain, and unlike general lipid transporters (Lee et al. 1998; Choinowski et al. 2000) or lipid synthesis enzymes, which also bind lipid or sterols; many members are implicated in activities beyond that of trafficking and metabolism of specific lipids to presumed roles in cell signaling (Iyer et al. 2001; Soccio 2003).

START domains in humans

The mammalian START domain protein family is divided into six subfamilies based on sequence similarities (reviewed in (Alpy and Tomasetto 2005; Alpy et al. 2009; Clark 2012) (see Table 1.1). In total, there are 15 members of the mammalian START protein family. Members of each subfamily share either similar ligand binding specificities or functional domains other than the START domain, such as the cholesterol and oxysterol binding proteins of the STARD1/D3 and STARD4/D5/D6 subfamilies, the phospholipid and sphingolipid binding proteins of the STARD2 (PCTP)/D7/D10/D11 subfamily, the multi-domain proteins containing either putative Rho-GTPase signaling function of the STARD8/12/13 subfamily, thioesterase activity of the STARD14/15 subfamily, or kinesin motor function for STARD9.

Lessons learned from crystal structures of mammalian START domains

The crystal structures of 8 mammalian START domains have been solved thus far. These include ligand-bound co-crystals for PCTP/STARD2 (Roderick et al. 2002) and CERT/STARD11 (Kudo et al. 2008; Kudo et al. 2010), free forms for MLN64/STARD3 (Tsujishita and Hurley 2000), STARD4 (Romanowski 2002), STARD1, STARD5, DLC-2/STARD13, and ACOT11/STARD14 (Thorsell et al. 2011), and the nuclear magnetic resonance (NMR) solution structure of STARD5 in complex with the bile acids (Letourneau et al. 2012; Letourneau et al. 2013). Most recently, the solution structure of apo-STARD6 was also solved (Protein Data Bank (2MOU), unpublished). All of the structures share the same α/β-helix-grip-fold configuration, consisting of a nine-stranded twisted antiparallel β sheet gripped by amino- and carboxy-terminal α helices (α1 and α4). The two Ω loops are inserted between strands β5 and β6 (Ω1) and strands β7 and β8 (Ω2), creating a unique configuration and hydrophobic pocket that is distinct from other classes of lipid binding proteins. Notably, the curved β sheet, three α-helices (α2, α3, α4) and a loop (Ω1) form the walls of this hydrophobic tunnel. The START domains from PCTP/STARD2 and CERT/STARD11 were co-crystallized with ligands, thereby exemplifying that the dimensions and geometry of this tunnel are appropriate to accommodate a single ligand molecule. Additionally, the tunnel possesses two narrow openings, which are too small to allow the entrance or exit of the ligand without major structural rearrangement of the domain. Such major conformational change could involve the carboxy-terminal α4 helix and possibly the Ω1 loop (Roderick et al. 2002; Kudo et al. 2008; Kudo et al. 2010). Kudo et al. (2010) reported the crystal structures of the CERT/STARD11 START domain in complex with (1R,3R)-N-(3-hydroxy-1-hydroxymethyl-3-phenylprpyl) alkanamides (HPAs) of varying acyl chain lengths. HPA12, a synthesized analogue of ceramide, is a selective inhibitor of CERT/STARD11-mediated ceramide trafficking. According to the co-crystal structure data, the Ω1 loop, previously suggested to function as a gate of the cavity, adopts a different conformation when bound to HPA than when bound to ceramide. In the Ω1 loop region, the position of amino acid residue W473 highlights a significant difference between the HPA-bound and the apo-form or ceramide-bound structures. Furthermore, surface plasma resonance experiments convincingly showed that W473 is important for interaction with membranes. With these findings, the co-crystal structures have provided useful insights into not only the molecular mechanism of inhibition by HPAs but also the possible mechanism of

ceramide transfer by CERT/STARD11. It is conceivable that such information may be valuable for prediction of ligand contact points in homologous structural models.

Ligand binding specificity of mammalian START domain proteins

A wealth of knowledge has become available on the ligands for the START domain proteins from mammals. These include sterols such as cholesterol (Tsujishita and Hurley 2000) and 25-hydroxycholesterol (Rodriguez-Agudo et al. 2008), phosphatidylcholine (PC) (Roderick et al. 2002; Horibata and Sugimoto 2009), phosphatidylethanolamine (Olayioye 2005), ceramides (Hanada et al. 2003), bile acids (Letourneau et al. 2012), and steroid hormones (Letourneau et al. 2015).

Ligand specificity appears to be largely conferred by the shape of the cavity and formation of polar interactions with side chains of the residues lining the walls of the hydrophobic cavity (Lavigne et al. 2010). The ligand occupies the centrally located amphiphillic cavity and crystal structure and ligand-binding kinetics data support the idea that START proteins exhibit 1:1 binding stoichiometry of protein and ligand. For example, the co-crystal of PCTP/STARD2 bound to PC identified 28 residues that contact the ligand, of which 15 are aliphatic, 11 aromatic and 2 hydrophobic. The PC head group interacts with hydrophilic side chains groups of Y72, N157 and R78, a feature that explains the striking specificity of PCTP/STARD2 for PC. Further, R78 interacts directly with the phosphoryl group and forms a salt bridge with R82 (Roderick et al. 2002). In several reports on STARD1 and STARD3, the presence of a conserved salt-bridge at the bottom of the cavity between an acidic residue (N169/STARD1 and D332/STARD3) and a basic residue (R188/STARD1 and R351/STARD3) was proposed to be a key determinant for cholesterol binding (Tsujishita and Hurley 2000; Mathieu et al. 2002; Murcia et al. 2006). More precisely, the C3-hydroxyl group of cholesterol could form a specific interaction with the guanidino group of the conserved arginine. The importance of this salt bridge for specific cholesterol binding is further strengthened by the characterization of a StAR/STARD1 salt bridge double mutant E169M; R188M that exhibits complete loss of its cholesterol binding capacity and lacks steroidogenic activity (Roostaee et al. 2009). The carboxy-terminal α4 also appears to be critical for ligand binding, since helix mutations L271N and L275P result in a significant decline in binding affinity and steroidogenic activity (Roostaee et al., 2009). The importance of the shape of ligand binding cavity is revealed

by yet another report, in which CERT/STARD11 was shown to specifically recognize natural D-erythro ceramides having short acyl chain lengths (C_{14}- C_{20}), but not sphingosine, sphingomyelin, cholesterol, or PC. It was observed that the dimensions and geometry of the cavity are appropriate to accommodate one ceramide molecule with short amide-acyl chain length (no longer than C_{20}) (Kumagai et al. 2005; Kudo et al. 2008). Furthermore, it was found that the molecular mechanisms underlying lipid specific recognition and length limits are imparted by specific amino acid residues; R442 (α3), E446 (α3), Q467 (β5), N504 (β7), and Y553 (β9) located at far end of the long amphiphillic cavity (Kudo et al. 2008), and the ceramide head group forms a hydrogen bond network with these amino acid residues. Consistently, the inner wall of the cavity is lined by many non-polar amino acids involved in hydrophobic interactions with the two amphiphilic chains of ceramide.

It should be noted that the residues forming the cavity are by no means strictly hydrophobic; many side chains with H-bond donors and acceptors line the cavity. For example, in the co-crystal of STARD11 bound to ceramides; the polar head group of the ceramide within STARD11 is located at the bottom of the cavity where the cholesterol OH is proposed to be buried in STARD1 and STARD3 (Tsujishita and Hurley 2000; Mathieu et al. 2002; Murcia et al. 2006). Similarly, comparable interactions involving structurally conserved ion pairs and the polar head group of PC are observed in PCTP/STARD2 (Roderick et al. 2002). Previously, Kudo et al. (2010) demonstrated that although ceramides and PC have somewhat similar polar head groups, the STARD11 cavity is too small to accommodate a PC molecule (Kudo et al. 2010). This reinforces the hypothesis that the ligand specificity displayed by START domains relies on the shape of the cavity and on the formation of polar interactions with the side chains inside the cavity.

Human pathologies involving START proteins

In humans, START proteins are linked to various pathologies including cancer, genetic disorders, and autoimmune disease. Expression of several START proteins appears to be perturbed in certain cancers. MLN64/STARD3 was found to be invariably co-amplified and co-expressed with the epidermal growth factor receptor (*ERBB2*) gene in breast cancers (Alpy et al. 2003). Cai et al. (2010) provided evidence that MLN64 expression levels correlate with poor prognosis and overall survival, and it contributes to the progression of breast cancer through cell

proliferation and impairment of cell-matrix adhesion via the integrin-mediated focal adhesion kinase (FAK) signaling pathway. GTT1/STARD7, a novel gestational trophoblastic gene, is frequently up-regulated in many cancer-derived cell lines such as JEG-3, HT29 and HepG2, indicative of its role in the phospholipid mediated tumor signaling (Durand et al. 2004). Similarly, STARD10 is co-expressed with ErbB2 and cooperates with the ErbB pathway in malignant transformation in breast cancers. It is also over-expressed in significant proportion of human colon carcinoma cells, implicating its role in tumorigenesis (Olayioye et al. 2004).

In contrast, genes encoding another class of START domain proteins were reported to be down regulated in tumor cells. DLC-1/STARD12, DLC-2/STARD13 and DLC-3/STARD8, the members of Rho-GAP START family appear to act as tumor suppressor genes, since their expression is frequently lower or absent in various human cancers. Conversely, elevated CERT/STARD11 expression was found to be associated with multi-drug resistance to various cancer cytotoxics such as paclitaxel, doxorubicin and cisplatin (Swanton et al. 2007). Exon array expression profiling revealed a significant representation of genes involved in the autophagy pathway responsible for mediating paclitaxel sensitization following CERT/STARD11 depletion in breast and colorectal cancer cell lines. Thus, pharmacological targeting of CERT/STARD11, as a prognostic marker in breast cancer, may enhance drug response (Lee et al. 2012).

START proteins are also associated with pathologies such as genetic disorders and autoimmune disease. Ectopic expression of StAR leads to increased pregnenolone production in the absence of a stimulus (Lin et al. 1995; Sugawara et al. 1995), while mutations in the StAR gene result in an autosomal recessive disease named lipoid congential adrenal hyperplasia (CAH), characterized by accumulation of cholesterol in the adrenals and gonads coupled with decreased steroid synthesis (Lin et al. 1995; Stocco 2002). Finally, there is evidence that CERT/STARD11 is involved in the autoimmune disease termed Goodpasture syndrome (Hanada et al. 2003).

Homeodomain-leucine zipper (HD-Zip) transcription factors in *Arabidopsis*

In plants, the majority of START domain proteins are members of a plant-specific homeodomain (HD) transcription factor family (HD-Zip family) that is conserved in the flowering plants from *Arabidopsis* to rice (Schrick et al. 2004). HD-Zip genes are found in the moss *Physcomitrella patens* (Sakakibara et al. 2001; Floyd et al. 2006; Prigge and Clark 2006;

Zalewski et al. 2013), suggesting that this plant specific configuration evolved in the earliest ancestor of the land plant lineage. Homologs of HD-Zip genes have subsequently also been identified in charophyte algae, indicating that an ancestral HD-Zip gene evolved in an algal ancestor prior to the origin of land plants (Zalewski et al. 2013).

Class III and IV HD-Zip proteins are comprised of a homeodomain sequence-specific DNA binding domain (HD), a dimerization leucine zipper motif adjacent to the C terminus of the homeodomain (ZLZ), a START domain, and a conserved START-associated domain (SAD) (Mukherjee and Burglin 2006). Little is known about functions of this domain. However, using both EMSA and the yeast 2-hybrid (Y2H) assays, Zhang et al. (2010) demonstrated that simultaneous presence of both the START and SAD domains of cotton class IV HD-Zip TF GbML1 is required for binding to the C-terminal domain of GbMYB25, a key regulator of cotton fiber initiation.

Members of the class III HD-Zip subfamily can be distinguished from class IV HD-Zip subfamily members by an additional domain downstream from the SAD called the PAS-related MEKHLA domain (Mukherjee and Burglin 2006). Different leucine zippers also characterize the two divergent families of HD-Zip proteins in *Arabidopsis* (Figure 1.1). The larger family, known as class IV HD-Zip, contains a leucine zipper with an internal loop, termed zipper-loop-zipper (ZLZ), following the homeodomain (Schrick et al. 2004). Despite disruption of the leucine zipper of the ZLZ motif by an inserted loop sequence, dimerization via this motif can still occur (Di Cristina et al. 1996). Dimerization is a prerequisite for DNA binding by class IV HD-Zip proteins and unlike other HD proteins, which can bind DNA as monomers, class IV HD-Zip TFs can bind DNA efficiently only as either homo- or hetero-dimers (Palena et al. 2001). The smaller group, known as class III HD-Zip, contains a classic leucine zipper domain similar to bZIP and BRLZ found in all eukaryotes (Landschulz et al. 1988; Oshea et al. 1989).

Within class IV HD-Zip proteins, the homeodomain consists of ~60 conserved residues in a DNA binding helix-loop-helix-turn-helix structure (reviewed in Laughon, 1991), while the class III HD-Zip homeodomain contains four amino acid changes in the primary sequence between helix 2 and helix 3 (Mukherjee and Burglin 2006). Homeodomains are highly conserved among eukaryotic transcription factors responsible for regulating cell differentiation and pattern formation (Gehring et al. 1990). As both families share SAD domains, one possibility is that the MEKHLA domain became attached to the carboxy terminus of the class IV HD-Zip TFs,

establishing the class III HD-Zip family (Mukherjee and Burglin 2006; Magnani and Barton 2011). Perhaps the class IV HD-Zip was ancestral gene that subsequently duplicated and modified to produce a class III HD-Zip gene.

Extensive genetic and molecular analyses of the HD-Zip family members have revealed their critical roles in plant development. The class III HD-Zip family includes five *Arabidopsis* proteins implicated in vasculature, meristem initiation, and organ polarity (Prigge et al. 2005; Turchi et al. 2015). By contrast, the larger class IV HD-Zip family includes 16 *Arabidopsis* members primarily involved in cell fate determination and epidermal development in flowering plants (Nakamura et al. 2006).

The class III and IV HD-Zip proteins are among the largest TFs in plants and have a complex domain structure. Unfortunately, attempts to understand function of these TFs through structural studies have so far produced no data.

Towards an understanding of the function of the class IV HD-Zip proteins

The plant epidermis is a multifunctional tissue that serves as an interface of sessile land plants with their environment. Beyond its role as a protective barrier against biotic and abiotic stress, the epidermal cells allow exchange of gas, water, and nutrients with the outside world. Also, some epidermal cells give rise to specialized cell types – hair cells (trichomes) or stomatal guard cells on the aerial parts, root hairs in the root, and aleurone layer in the seed - which are critical for defense, respiration, water and nutrient uptake, and starch degradation, respectively (Glover 2000). In addition to its protective function, the epidermis plays roles in the regulation of organ growth and shoot stem cell maintenance (Savaldi-Goldstein et al. 2007; Knauer et al. 2013; Nobusawa et al. 2013).

The differentiation and maintenance of epidermis involves members of the class IV HD-Zip family, also known as HD-GL2 after its first identified gene, GLABRA2 (GL2) (Rerie et al. 1994; Di Cristina et al. 1996). The *Arabidopsis* genome contains 16 class IV HD-Zip proteins. These include GLABRA2 (GL2), redundant proteins ARABIDOPSIS THALIANA MERISTEM LAYER1 (ATML1) and PROTODERMAL FACTOR2 (PDF2), ANTHOCYANNLESS2 (ANL2), and FWA. Additional characterized family members include HOMEODOMAIN GLABROUS1 (HDG1) through HDG12 (Nakamura et al. 2006). The pivotal role that class IV HD-Zip transcription factors play in various aspects of plant growth and development is exerted

by a fine-tuned regulation of their expression and function operating at multiple levels (discussed below). This dissertation focuses on three members of the class IV HD-Zip family: GLABRA2, PDF2, and HDG11.

GLABRA 2 (GL2)

GLABRA2, also known as Athb-10 (Di Cristina et al. 1996), has been shown to promote trichome formation in the shoot epidermis (Hulskamp et al. 1994; Rerie et al. 1994) and non-root hair cell formation in root (Di Cristina et al. 1996; Masucci et al. 1996; van Hengel et al. 2004). It also positively regulates seed mucilage synthesis, at least in part through positive control of the *MUM4/RHM2,* a rhamnose synthase that is required to produce rhamnose (Western et al. 2004). Consistently, the *gl2* mutant is defective in trichome and seed coat mucilage formation, and has ectopic root hairs in non-root hair cell files. A role of GL2 in seed oil accumulation was also demonstrated by characterization of mutant *Arabidopsis* lines lacking GL2, which revealed a significant increase in seed oil content (Shen et al. 2006). In addition, a recent study in *Arabidopsis* has identified an activation-tagged gain-of-function mutant allele of GL2 with reduced anthocyanins (Wang et al. 2015). Extensive molecular analyses revealed that this phenotype is caused by an elevated expression of GL2, suggesting an additional role for GL2 in regulating anthocyanin biosynthesis.

A large-scale gene expression analyses have revealed a large number genes regulated downstream of GL2 during epidermal cell-fate determination (Lieckfeldt et al. 2008; Marks et al. 2009; Won et al. 2009; Bruex et al. 2012). However, thus far, only a handful of downstream genes have been identified as direct targets of GL2. PLDζ1, an *Arabidopsis* phospholipase D (PLD) gene was the first identified GL2 target (Ohashi et al. 2003). *In vitro* protein-DNA-binding analyses have identified an L1-box sequence TAAATGTT in the promoter region of PLDζ1 as a GL2 binding sequence. Two polysaccharide-metabolizing genes, CELLULOSE SYNTHASE5 (CESA5) and XYLOGLUCAN ENDOTRANSGLUCOSYLASE17 (XTH17) have also been identified as direct targets of GL2 (Tominaga-Wada et al. 2009). Yeast one-hybrid (Y1H) analysis showed that GL2 binds to the L1-box sequence TAAATGTA present in the CESA5 and XTH17 promoters, while chromatin immunoprecipitation (ChIP) assay demonstrated *in vivo* binding to the L1-box sequence in the MYB23 promoter (Khosla et al. 2014). More recently, evidence has been provided that some MYB-bHLH-WD40 (MBW)

component genes, including *MYB113*, *PAP2*, and *TT8* are direct targets of GL2 (Wang et al. 2015). MBW complex regulates the expression of downstream biosynthetic genes in the flavonoid pathway, leading to the production of anthocyanins. Overall, GL2 is emerging as a pivotal transcription factor for epidermal cell-type differentiation in diverse tissues.

ATML1 and PDF2 are functionally redundant

The epidermis of shoot organs in plants develops from the outermost layer (L1) of the shoot apical meristem and plays important roles to not only to protect plants from dehydration and pathogens, but also to ensure their proper organogenesis and growth control (Takada and Lida 2014). In the past decade, molecular genetic studies have uncovered transcription factors that are master regulators of shoot epidermal cell fate. In particular, ATML1 and PDF2, expressed specifically in outermost layer (L1) of shoot apical meristems and the protoderm of developing embryos (Lu et al. 1996), have been shown to play redundant but critical roles in formation of the shoot epidermis. ATML1 was the first transcription factor whose expression was shown to mark the outermost cell layer of the shoot (Sessions et al. 1999). While each loss-of-function mutant of ATML1 and PDF2 displayed no discernible developmental abnormalities, the strong mutant alleles of *atml1;pdf2* showed embryo-lethal phenotypes (Ogawa et al. 2015). Similar to GL2, both ATML1 and PDF2 were shown to bind *in vitro* to an 8 bp cis-element called the L1 box (Abe et al. 2001; Abe et al. 2003). Because L1 box is often found in the promoters of epidermis-specific genes such as *ATML1* and *PDF2*, these TFs have been proposed to positively regulate the expression of epidermis-specific genes, including their own expression by positive auto regulation (Abe et al. 2001; Abe et al. 2003). Accordingly, both PDF1 (Abe et al. 2003), a gene coding for Pro-rich protein and LIP1 (Rombola-Caldentey et al. 2014) encoding GDSL lipase, are directly regulated by ATML1/PDF2 through the L1 box cis-element.

ANL2 and FWA

In situ hybridization and/or gene-specific promoter-GUS fusions demonstrated that the expression of nine of the 16 class IV HD-Zip genes is strongly enriched in the epidermis, whereas expression of one family member is limited to the sub-epidermal layer (Nakamura et al. 2006; Kubo et al. 2008). ANL2 controls anthocyanin pigmentation of the leaf sub-epidermal

layer and cellular organization of the primary root (Kubo et al. 1999). Plant anthocyanins are pigments that are responsible for a range of colors such as red, blue and purple. The accumulation of anthocyanins occurs during specific plant developmental stages and only in certain plant tissues. The *anl2* mutant has an abnormal radial root patterning and produces extra cells known as intervening cells located between the cortical and epidermal layers (Kubo and Hayashi 2011). It was proposed that *ANL2* is involved in the maintenance of sub-epidermal layer identity during plant development and that it controls the process of anthocyanin accumulation in sub-epidermal cells. Another member FWA is expressed exclusively in the endosperm (Kinoshita et al. 2004; Kinoshita et al. 2007) and *fwa* mutant phenotypes display delayed flowering due to ectopic overexpression of the gene in other tissues.

Functional redundancy among the class IV HD-Zip family members

As with the class III HD-Zip, functional overlaps and redundancy seem to be prominent within the class IV HD-Zip family. Single T-DNA insertion mutants of *HDG1-HDG12* are aphenotypic with the exception of *hdg11-1*, which has more branched trichomes (Nakamura et al. 2006) and *hdg2* that has trichomes with abnormal cell walls (Marks et al. 2009) and displays aberrant stomatal development (Peterson et al. 2013). A search for paralogs revealed seven paralogous gene pairs between the class IV HD-Zip family members (Tavares et al. 2000). Double mutants of paralogous gene pairs displayed distinct phenotypes. While *hdg11 hdg12* double mutants displayed enhanced trichome branching phenotype (Nakamura et al. 2006), double mutants of *pdf2-1* with *hdg1-1*, *hdg2-3*, *hdg5-1* and *hdg12-2* produced abnormal flowers with sepaloid petals and carpelloid stamens in association with the reduced expression of the petal and stamen identity gene *APETALA 3* (*AP3*) (Kamata et al. 2013), suggesting that PDF2 and these class IV HD-Zip genes are cooperatively involved in flower development. Similarly, a study by Nakamura et al. (2006) revealed cotyledon defects for *pdf2-1 hdg3-1* and *atml-1 hdg3-1*, suggesting the importance of the interplay in the epidermis between PDF2, ATML1, and HDG3 during cotyledon development. Several reports have highlighted allele specific effects on the mutant phenotypes. In one study, the double mutant combinations between another *pdf2* allele, *pdf2-2*, and *hdg* alleles showed only a reduced number of stamens without homeotic conversions of floral organs (Kamata et al. 2013). The mutant alleles, *pdf2-1* and *pdf2-2*, have a T-DNA inserted in the midst of the START domain coding region, and in the

exon for the more C-terminal part of the START domain, respectively, indicating the importance of the START domain in proper function of the class IV HD-Zip proteins. While in the second study, using two mutant alleles of ATML1 and four mutant alleles of PDF2, authors showed that alteast one copy of these genes are essential for embryo development in *Arabidopsis*. The double mutant combinations of a strong *atml1-3* allele with each *pdf2* allele were found to cause embryonic arrest (Ogawa et al. 2015), whereas, double mutant combinations of a weak *atml1-1* allele with each *pdf2* allele germinated and showed phenotypes defective in shoot epidermal cell differentiation (Abe et al. 2003; San-Bento et al. 2014).

Functional redundancy among nonparalogous members has also been demonstrated. For instance, both *HDG2* and *ATML1* appear to function redundantly in promoting stomatal differentiation, possibly by binding to common *in vivo* target sequences (Peterson et al. 2013). Moreover, a recent study showed that GL2 and HDG11 act redundantly in driving trichome morphogenesis (Khosla et al. 2014). In conclusion, these findings shed light on the cooperative functions of the class IV HD-Zip family members in epidermis; however, due to the limitations of the mutant alleles available for analysis, the reasons for the phenotypes observed in the mutant combinations are still not fully understood.

Multiple mechanisms control the activity of the class IV HD-Zip transcription factors

Several lines of evidence suggest microRNAs play an important role in regulating expression of genes that control pattern formation and cell differentiation in plants (Reinhart et al. 2002). One of the well-characterized cases of miRNA regulation in plants is recognition of a conserved region at the beginning of the START domain in the coding region of class III HD-Zip by miR165/166 (Bowman 2004; Zhou et al. 2007; Du et al. 2015; Jia et al. 2015). Regulation of the class IV HD-Zip genes by miRNAs is supported by the presence of a 17-nucleotide motif located at in the 3'UTR region of the class IV HD-Zip genes. This sequence is conserved throughout land plants and their ability to complementary base pair indicates that the formation of stem-loop is required for function (Nakamura et al. 2006). Two possible functions attributed to this conserved sequence are the generation of miRNAs and posttranscriptional regulation of gene expression via the binding of regulatory proteins mediating translational repression, mRNA stability, and mRNA localization. However, to date, there is no evidence for the production of regulatory small RNAs from any of the *Arabidopsis* genes harboring the conserved 3'-UTR

sequences (Gustafson et al. 2005). Recently, genetic analysis and GFP sensor experiments demonstrated post-transcriptional regulation of maize class IV HD-Zip transcription factor (OCL1) by RDR2 dependent small RNA (small 1), suggesting another level of regulation. It is noteworthy that both the small1 binding site itself and the 19 nt region of the 3′ UTR contribute to *OCL1* translational repression (Klein-Cosson et al. 2015). Since these motifs are predicted to form a stem-loop structure that is conserved, it is tempting to speculate that similar translational repression mechanism exists for the class IV HD-Zip members from other plants.

The recent discovery of physical interactions between ATML1 and PDF2 with DELLA proteins revealed a new layer of post-translational regulation of the class IV HD-Zip activity. Both the proteins were shown to interact directly with DELLA proteins, which blocks gibberellin signaling required for seed germination (Rombola-Caldentey et al. 2014). Thus, upon seed imbibition, increased GA levels mediate the activation of downstream L1 box target genes by releasing the two HD-Zip TFs from their inhibitory interaction with DELLA proteins. Consistent with this, physical interactions between GhHOX3, an HD-Zip TF of cotton, and GhSLR1, a functional orthologue of the *Arabidopsis* DELLA AtGAI, were reported and the domain deletion assay demonstrated that the combination of leucine zipper (ZLZ) and the START domains was required for the interaction (Shan et al. 2014). In this context, it will be of interest to determine whether DELLAs interact with the remaining epidermal class IV HD-Zip proteins to regulate additional plant developmental processes such as trichome differentiation and flower development.

The presence of a putative lipid/sterol binding domain (START) and a dimerization motif (ZLZ) implies regulation of the class IV HD-Zip activities by dimerization and binding of lipid/sterol ligands. In this context, interaction between ATML1 and PDF2, likely by heterodimerization (Rombola-Caldentey et al. 2014), together with our *in planta* data showing heterodimerztion of GL2 with ATML1/PDF2, coupled with the requirement of START domain for homodimerization, indicates a new layer of regulation of the class IV HD-Zip activity. In related transcription factors from sunflower, DNA-binding was inhibited by oxidation of Cys residues in the ZLZ motif, suggesting that redox signals are also involved in the regulation of the class IV HD-Zip activity (Tron et al. 2002).

Class IV HD-Zip genes in Commercially Important Plant Species

Data obtained with *Arabidopsis* on the regulation of trichome development by class IV HD-Zip genes, stimulated attempts to understand roles of these genes in agriculturally important plants with the aim to apply the gained knowledge for plant improvement. Four distinct *GL2* orthologues from *B. napus*, *B. rapa* (*BnaA.GL2.a* and *BnaA.GL2.b*) and *B. oleracea* (*BnaC.GL2.a* and *BnaC.GL2.b*) were identified using an overlapping-PCR strategy (Chai et al. 2010). Suppression of *BnaC.GL2.b* drastically increased oil accumulation, suggesting that BnaC.GL2.b protein can functionally substitute for GL2 to negatively regulate oil accumulation in *Arabidopsis* seeds. Given that *Brassica napus* is the third largest oilseed crop in the world; identification of genes involved in regulation of seed oil accumulation has important economic significance for oilseed rape breeding.

Fu and coworkers (Fu et al. 2009) reported map-based cloning of the gene Yr36 (WKS1), which confers partial resistance to *Pst*, a causative agent of wheat stripe rust disease. *WHEAT KINASE-START* (*WKS1*) gene includes a kinase domain followed by a putative START lipid-binding domain and the most similar protein in *Arabidopsis* to the putative WKS1 START domain is EDR2, a protein that negatively regulates plant defense to the powdery mildew pathogen Golovinomyces cichoracearum (Tang et al. 2005; Vorwerk et al. 2007). Extending these findings, Gou et al. (2015) proposed a working model for WKS1 resistance mechanism highlighting the role of the START domain in conferring partial resistance to *Pst* in wheat. Collectively, these data improve the understanding of the molecular mechanisms associated with WKS1 partial resistance to a devastating wheat pathogen and may have implications in the utilization of this gene in agriculture.

A function of the class IV HD-Zip TFs has also been described in the context of regulation of cotton fiber elongation. Cotton fibers, the most important natural material for textile industry, are unusually long, single-celled epidermal seed trichomes. Comparison with the *Arabidopsis* class IV HD-Zip proteins demonstrated that GaHOX1 from cotton is the closest homolog of GL2, whereas GaHOX2 is more similar to ATML1 and PDF2 than to GL2 (Guan et al. 2008). A recent report on GaHOX3 showed that silencing of the class IV HD-Zip gene dramatically reduces fiber length, whereas its overexpression leads to longer fibers (Shan et al. 2014). Moreover, the study revealed a new molecular mechanism underlying the role of growth hormone gibberellin (GA) in promoting cotton fiber development. These findings could play a

crucial role in the future selective breeding of cotton plants, thereby impacting the billion-dollar textile industry.

To conclude, these genes represent just a part of the class IV HD-Zip family, and we are still far from a comprehensive understanding of the class IV HD-Zip family.

With the broad goal of defining and characterizing the molecular link between metabolism and gene transcription in plants, this dissertation addresses the following objectives: (1) Uncover the regulatory role of the START domain in modulating transcription factor activity, (2) Determine function of the START domain in the class IV HD-Zip transcription factor GLABRA2 (GL2), (3) Identify and characterize protein-metabolite interactions for the START domains, and (4) Elucidate genetic interactions between GL2, HDG11, and MYB23 in the maintenance of cell fate during trichome morphogenesis.

The second chapter of the dissertation describes the regulatory role of the START domain in stimulating transcription factor activity. The data shows that the START domains are versatile ligand-binding motifs that are used by both plants and mammals to modulate transcription factor activity.

The third chapter of the dissertation illustrates functional characterization of the START domain in GLABRA2. Mutational analysis in combination with experiments in yeast and plants reveal a key role of the START domain in regulating HD-Zip transcription factor activity through both ligand binding and via protein-protein interactions. This chapter also describes results from the ChIP-seq experiment. The study was undertaken to identify novel transcriptional targets of GL2 in an unbiased manner.

The fourth chapter uses parallel *in vitro* and *in vivo* approaches to identify the START domain ligands from HD-Zip transcription factors. Using a high throughput differential scanning fluorimetry assay and a small molecule pull down approach coupled to mass spectrometry, evidence is provided for binding to fatty acids and plant secondary metabolites.

The fifth chapter discusses the interaction and functional redundancy among the class IV HD-Zip family members in maintenance of the trichome cell fate. Evidence is provided for a previously elusive redundancy between GL2 and HDG11, and unveils a positive feedback loop in the maintenance of GL2 activity during trichome cell-type differentiation via MYB23.

This dissertation attempts to understand at the mechanistic level, how metabolism is integrated to regulate gene transcription in plant development. These results not only improve the

understanding of the START domains in general, but also provide novel insights into the regulation of HD-Zip TFs via START domain. The class IV HD-Zip family of TFs in unique to the plant kingdom and have been implicated as "master regulators" of epidermal development in flowering plants (Javelle et al. 2011). It is believed that the class IV HD-Zip homologs were instrumental in the evolution of epidermal adaptations required for life on land (Graham et al. 2000; Ligrone et al. 2012). Consistently, loss-of-function phenotype studies have revealed critical roles in differentiation and maintenance of epidermis, regulation of lipid biosynthesis and transport, seed oil content, and drought tolerance. One family member was recently shown to drive fiber elongation in cotton. In addition, insights into the role of these proteins in cuticle-related lipid synthesis and transport, and hence their possible involvement in plant protection from pathogens and abiotic stresses has just started to emerge. Thus, gaining a better understanding of the regulation of *Arabidopsis* class IV HD-Zip genes could have a significant impact in biotechnology and improvement of agriculturally important traits.

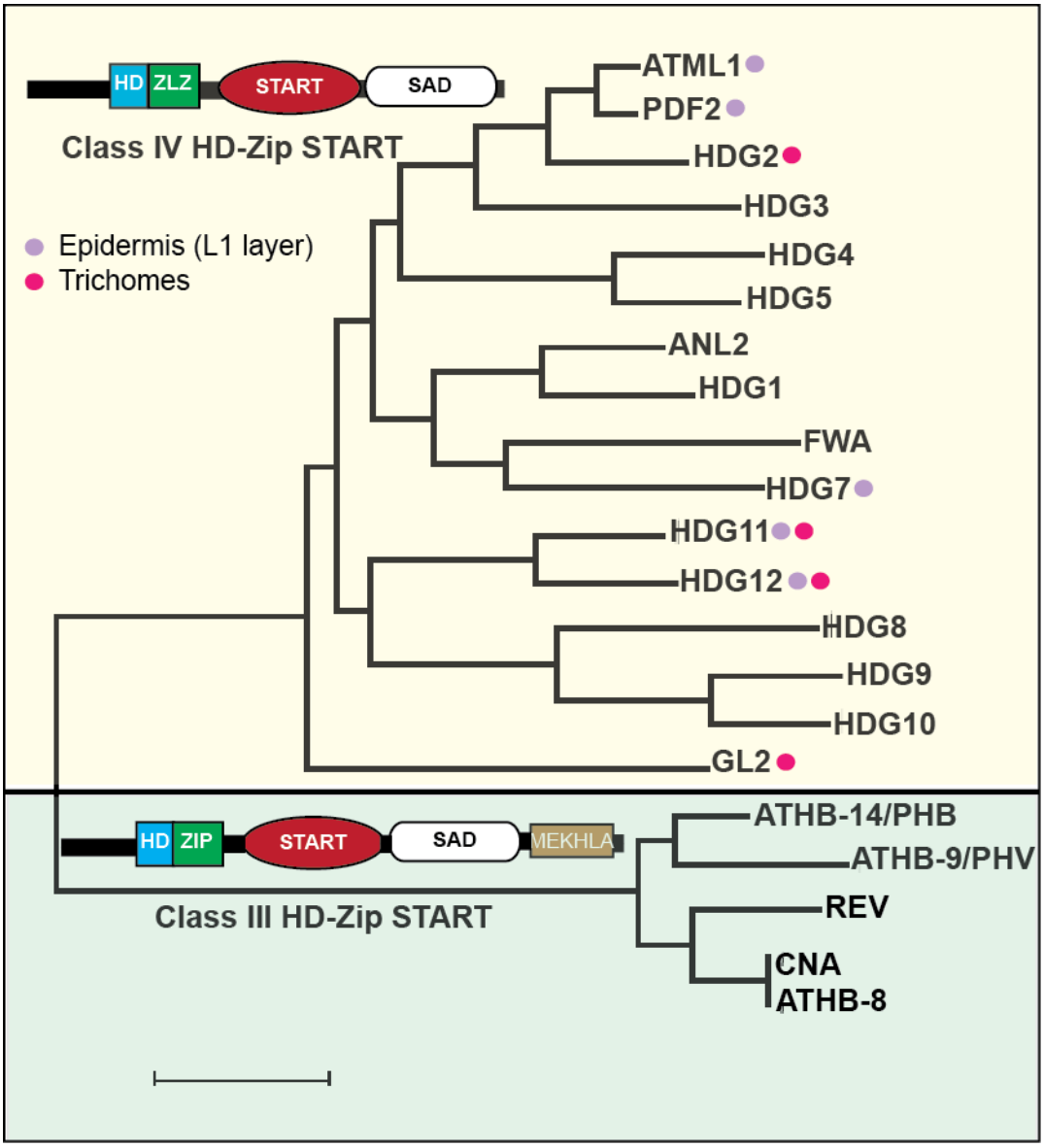


Figure 1.1 Phylogenetic tree of *Arabidopsis* HD-Zip START proteins from alignment of START.

START domain sequences were aligned by the ClustalW2 program (Larkin et al. 2007). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Molecular phylogenetic analysis by maximum likelihood method was conducted in MEGA6 (Tamura et al. 2013). Structures of the two types of HD-Zip START sequences (Class IV and Class III HD-Zip transcription factors) having different leucine zippers. HD, homeodomain; ZLZ, leucine zipper loop zipper; ZIP, basic region leucine zipper; SAD, START adjacent domain. Class III HD-Zip family contains a MEKHLA motif.

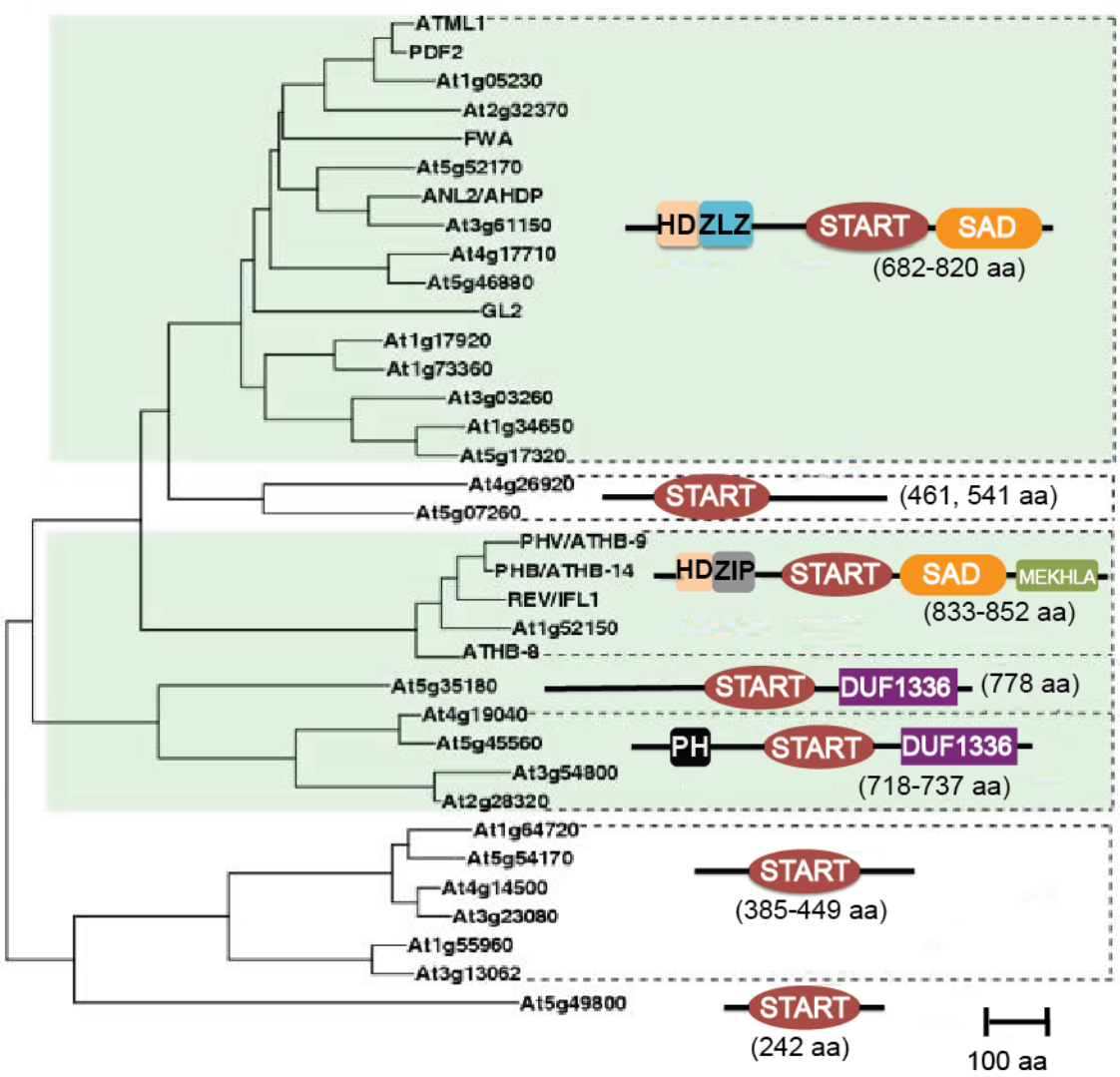


Figure 1.2 Phylogenetic analysis of the START domain proteins in *Arabidopsis.*

START domains from 35 Arabidopsis START-containing proteins are classified into seven subfamilies. The structure and domain organization for each protein or protein subfamily is shown on the right, with START domains in red. HD, homeodomain (orange); ZLZ, leucine zipper-loop-zipper (blue); ZIP, basic region leucine zipper (grey); PH, pleckstrin homology (black), DUF1336, domain of unknown function (purple), MEKHLA (green). Sizes of the corresponding proteins in amino acids (aa) are indicated to the right or below each representation. Figure modified from Schrick et al. (2004).

Table 1-1 Mammalian START domain family.

The START protein subfamily is indicated and each member is designated by protein name followed by other common names in parenthesis. A schematic of the domain structure highlights that the START domain is the C-terminal domain

START Subfamily	START protein	Domain Structure	Tissue Distribution	Cellular location	Lipid binding	Disease association
STARD1/D3	STARD1 (StAR) STARD3 (MLN64)	Mt-START MENTAL-START	Adrenal, ovary, testis, brain, heart, liver Placenta, breast, macrophages	Mitochondria Transmembrane, late endosomes	Cholesterol Cholesterol	Lipoid CAH, HCC NPC
STARD2	STARD2/PCTP STARD7 STARD10 STARD11 (CERT)	START —START -START— -PH—START (FFAT)	Liver, lung Liver Liver, kidney, testis, colon Liver	ER/Golgi Cytosolic, mitochondria Cytosolic ER/Golgi	PC PC PC>PE Ceramide	Insulin resistance Ovarian, lung, colon, liver cancer Fatty liver
STARD8/12/13	STARD8/DLC-3 STARD12/DLC-1 STARD12/DLC-2	—RhoGAP-START —SAM-RhoGAP-START —SAM-RhoGAP-START	Cancer Cancer Endothelial cells	Focal adhesions Focal adhesions, PM Focal adhesions	? ? Charged lipid (?)	 liver cancer angiogenesis
STARD9	STARD9	KM-FHA———cc-START	ubiquitious	nuclear	?	mitosis
STARD14/15	STARD14/ACOT11 STARD15/ACOT12	Hotdog-Hotdog-START -Hotdog-Hotdog-START	Brown adipose tissue Liver	Cytosolic Cytosolic	Fatty acid ?	thermogenesis

PC, Phosphatidylcholine; PE, Phosphatidylethanolamine; NPC, Niemann Pick type C; CAH, congenital lipoid adrenal hyperplasia;PH, Plekstrin Homology; FFAT, peptide EFFDAxE; DLC, Deleted in liver cancer; SAM, sterile α motif ; ACOT, acyl-coenzyme A thioesterase; FHA, forkhead-associated phosphopeptide binding domain

REFERENCES

Abe M, Katsumata H, Komeda Y, Takahashi T. 2003. Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. *Development* **130**: 635-643.

Abe M, Takahashi T, Komeda Y. 2001. Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. *Plant J* **26**: 487-494.

Alpy F, Boulay A, Moog-Lutz C, Andarawewa KL, Degot S, Stoll I, Rio M-C, Tomasetto C. 2003. Metastatic lymph node 64 (MLN64), a gene overexpressed in breast cancers, is regulated by Sp/KLF transcription factors. *Oncogene* **22**: 3770-3780.

Alpy F, Legueux F, Bianchetti L, Tomasetto C. 2009. [START domain-containing proteins: a review of their role in lipid transport and exchange]. *Med Sci (Paris)* **25**: 181-191.

Alpy F, Tomasetto C. 2005. Give lipids a START: the StAR-related lipid transfer (START) domain in mammals. *J Cell Sci* **118**: 2791-2801.

Arakane F, Kallen CB, Watari H, Foster JA, Sepuri NB, Pain D, Stayrook SE, Lewis M, Gerton GL, Strauss JF, 3rd. 1998. The mechanism of action of steroidogenic acute regulatory protein (StAR). StAR acts on the outside of mitochondria to stimulate steroidogenesis. *J Biol Chem* **273**: 16339-16345.

Bowman JL. 2004. Class III HD-Zip gene regulation, the golden fleece of ARGONAUTE activity? *Bioessays* **26**: 938-942.

Bruex A, Kainkaryam RM, Wieckowski Y, Kang YH, Bernhardt C, Xia Y, Zheng X, Wang JY, Lee MM, Benfey P et al. 2012. A gene regulatory network for root epidermis cell differentiation in *Arabidopsis. PLoS Genetics* **8**: e1002446.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD et al. 2005. Natural selection on protein-coding genes in the human genome. *Nature* **437**: 1153-1157.

Cai W, Ye L, Sun J, Mansel RE, Jiang WG. 2010. Expression of MLN64 influences cellular matrix adhesion of breast cancer cells, the role for focal adhesion kinase. *Int J Mol Med* **25**: 573-580.

Chai G, Bai Z, Wei F, King GJ, Wang C, Shi L, Dong C, Chen H, Liu S. 2010. *Brassica GLABRA2* genes: analysis of function related to seed oil content and development of functional markers. *Theor Appl Genet* **120**: 1597-1610.

Charbonnier S, Gallego O, Gavin AC. 2008. The social network of a cell: recent advances in interactome mapping. *Biotechnology Annual Review* **14**: 1-28.

Choinowski T, Hauser H, Piontek K. 2000. Structure of sterol carrier protein 2 at 1.8 A resolution reveals a hydrophobic tunnel suitable for lipid binding. *Biochemistry* **39**: 1897-1902.

Clark BJ. 2012. The mammalian START domain protein family in lipid transport in health and disease. *The Journal of Endocrinology* **212**: 257-275.

Di Cristina M, Sessa G, Dolan L, Linstead P, Baima S, Ruberti I, Morelli G. 1996. The *Arabidopsis* Athb-10 (GLABRA2) is an HD-Zip protein required for regulation of root hair development. *Plant J* **10**: 393-402.

Du Q, Avci U, Li S, Gallego-Giraldo L, Pattathil S, Qi L, Hahn MG, Wang H. 2015. Activation of miR165b represses AtHB15 expression and induces pith secondary wall development in *Arabidopsis*. *Plant J* **11:** 134-140

Durand S, Angeletti S, Genti-Raimondi S. 2004. GTT1/StarD7, a novel phosphatidylcholine transfer protein-like highly expressed in gestational trophoblastic tumour: cloning and characterization. *Placenta* **25**: 37-44.

Floyd SK, Zalewski CS, Bowman JL. 2006. Evolution of class III homeodomain-leucine zipper genes in streptophytes. *Genetics* **173**: 373-388.

Fu D, Uauy C, Distelfeld A, Blechl A, Epstein L, Chen X, Sela H, Fahima T, Dubcovsky J. 2009. A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science* **323**: 1357-1360.

Garber RL, Kuroiwa A, Gehring WJ. 1983. Genomic and Cdna Clones of the Homeotic Locus Antennapedia in Drosophila. *Embo J* **2**: 2027-2036.

Gehring WJ, Affolter M, Burglin T. 1994. Homeodomain Proteins. *Annu Rev Biochem* **63**: 487-526.

Gehring WJ, Muller M, Affolter M, Percival-Smith A, Billeter M, Qian YQ, Otting G, Wuthrich K. 1990. The structure of the homeodomain and its functional implications. *Trends Genet* **6**: 323-329.

Glover BJ. 2000. Differentiation in plant epidermal cells. *Journal of Experimental Botany* **51**: 497-505.

Gou JY, Li K, Wu K, Wang X, Lin H, Cantu D, Uauy C, Dobon-Alonso A, Midorikawa T, Inoue K et al. 2015. Wheat Stripe Rust Resistance Protein WKS1 Reduces the Ability of the Thylakoid-Associated Ascorbate Peroxidase to Detoxify Reactive Oxygen Species. *Plant Cell* **27**: 1755-1770.

Graham LE, Cook ME, Busse JS. 2000. The origin of plants: Body plan changes contributing to a major evolutionary radiation. *Proc Natl Acad Sci USA* **97**: 4535-4540.

Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ, Chen XY. 2008. The HD-Zip IV gene *GaHOX1* from cotton is a functional homologue of the *Arabidopsis GLABRA2*. *Physiol Plant* **134**: 174-182.

Gustafson AM, Allen E, Givan S, Smith D, Carrington JC, Kasschau KD. 2005. ASRP: the *Arabidopsis* Small RNA Project Database. *Nucleic Acids Res* **33**: D637-640.

Hanada K, Kumagai K, Yasuda S, Miura Y, Kawano M, Fukasawa M, Nishijima M. 2003. Molecular machinery for non-vesicular trafficking of ceramide. *Nature* **426**: 803-809.

Horibata Y, Sugimoto H. 2009. StarD7 Mediates the Intracellular Trafficking of Phosphatidylcholine to Mitochondria. *J Biol Chem* **285**: 7358-7365.

Hulskamp M, Misra S, Jurgens G. 1994. Genetic dissection of trichome cell development in *Arabidopsis*. *Cell* **76**: 555-566.

Iyer LM, Koonin EV, Aravind L. 2001. Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins* **43**: 134-144.

Javelle M, Klein-Cosson C, Vernoud V, Boltz V, Maher C, Timmermans M, Depege-Fargeix N, Rogowsky PM. 2011. Genome-Wide Characterization of the HD-ZIP IV Transcription Factor Family in Maize: Preferential Expression in the Epidermis. *Plant Physiology* **157**: 790-803.

Javelle M, Vernoud V, Depege-Fargeix N, Arnould C, Oursel D, Domergue F, Sarda X, Rogowsky PM. 2010. Overexpression of the epidermis-specific homeodomain-leucine zipper IV transcription factor Outer Cell Layer1 in maize identifies target genes involved in lipid metabolism and cuticle biosynthesis. *Plant Physiology* **154**: 273-286.

Jia X, Ding N, Fan W, Yan J, Gu Y, Tang X, Li R, Tang G. 2015. Functional plasticity of miR165/166 in plant development revealed by small tandem target mimic. *Plant Sci* **233**: 11-21.

Kamata N, Okada H, Komeda Y, Takahashi T. 2013. Mutations in epidermis-specific HD-ZIP IV genes affect floral organ identity in *Arabidopsis thaliana*. *Plant J* **75**: 430-440.

Khosla A, Paper JM, Boehler AP, Bradley AM, Neumann TR, Schrick K. 2014. HD-Zip Proteins GL2 and HDG11 Have Redundant Functions in *Arabidopsis* Trichomes, and GL2 Activates a Positive Feedback Loop via MYB23. *Plant Cell* **26**: 2184-2200.

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T. 2004. One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303**: 521-523.

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJ, Koornneef M, Kakutani T. 2007. Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J* **49**: 38-45.

Klein-Cosson C, Chambrier P, Rogowsky PM, Vernoud V. 2015. Regulation of a maize HD-ZIP IV transcription factor by a non-conventional RDR2-dependent small RNA. *Plant J* **81**: 747-758.

Knauer S, Holt AL, Rubio-Somoza I, Tucker EJ, Hinze A, Pisch M, Javelle M, Timmermans MC, Tucker MR, Laux T. 2013. A Protodermal miR394 Signal Defines a Region of Stem Cell Competence in the *Arabidopsis* Shoot Meristem. *Developmental Cell* **24**: 125-132.

Kubo H, Hayashi K. 2011. Characterization of root cells of *anl2* mutant in *Arabidopsis thaliana*. *Plant Sci* **180**: 679-685.

Kubo H, Kishi M, Goto K. 2008. Expression analysis of *ANTHOCYANINLESS2* gene in *Arabidopsis*. *Plant Sci* **175**: 853-857.

Kubo H, Peeters AJ, Aarts MG, Pereira A, Koornneef M. 1999. *ANTHOCYANINLESS2*, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis*. *Plant Cell* **11**: 1217-1226.

Kudo N, Kumagai K, Matsubara R, Kobayashi S, Hanada K, Wakatsuki S, Kato R. 2010. Crystal structures of the CERT START domain with inhibitors provide insights into the mechanism of ceramide transfer. *J Mol Biol* **396**: 245-251.

Kudo N, Kumagai K, Tomishige N, Yamaji T, Wakatsuki S, Nishijima M, Hanada K, Kato R. 2008. Structural basis for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide. *Proc Natl Acad Sci U S A* **105**: 488-493.

Kumagai K, Yasuda S, Okemoto K, Nishijima M, Kobayashi S, Hanada K. 2005. CERT mediates intermembrane transfer of various molecular species of ceramides. *J Biol Chem* **280**: 6488-6495.

Landschulz WH, Johnson PF, McKnight SL. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759-1764.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al. 2007. Clustal W and clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Lavigne P, Najmanivich R, Lehoux JG. 2010. Mammalian StAR-related lipid transfer (START) domains with specificity for cholesterol: structural conservation and mechanism of reversible binding. *Subcell Biochem* **51**: 425-437.

Lee AJX, Roylance R, Sander J, Gorman P, Endesfelder D, Kschischo M, Jones NP, East P, Nicke B, Spassieva S et al. 2012. CERT depletion predicts chemotherapy benefit and mediates cytotoxic and polyploid-specific cancer cell death through autophagy induction. *J Pathol* **226**: 482-494.

Lee JY, Min K, Cha H, Shin DH, Hwang KY, Suh SW. 1998. Rice non-specific lipid transfer protein: the 1.6 A crystal structure in the unliganded state reveals a small hydrophobic cavity. *J Mol Biol* **276**: 437-448.

Letourneau D, Lefebvre A, Lavigne P, LeHoux JG. 2015. The binding site specificity of STARD4 subfamily: Breaking the cholesterol paradigm. *Molecular and Cellular Endocrinology* **408**: 53-61.

Letourneau D, Lorin A, Lefebvre A, Cabana J, Lavigne P, LeHoux JG. 2013. Thermodynamic and solution state NMR characterization of the binding of secondary and conjugated bile acids to STARD5. *Biochimica et Biophysica Acta* **1831**: 1589-1599.

Letourneau D, Lorin A, Lefebvre A, Frappier V, Gaudreault F, Najmanovich R, Lavigne P, LeHoux JG. 2012. StAR-related lipid transfer domain protein 5 binds primary bile acids. *Journal of Lipid Research* **53**: 2677-2689.

Lieckfeldt E, Simon-Rosin U, Kose F, Zoeller D, Schliep M, Fisahn J. 2008. Gene expression profiling of single epidermal, basal and trichome cells of *Arabidopsis thaliana*. *J Plant Physiol* **165**: 1530-1544.

Ligrone R, Duckett JG, Renzaglia KS. 2012. Major transitions in the evolution of early land plants: a bryological perspective. *Ann Bot-London* **109**: 851-871.

Lin D, Sugawara T, Strauss JF, 3rd, Clark BJ, Stocco DM, Saenger P, Rogol A, Miller WL. 1995. Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. *Science* **267**: 1828-1831.

Lu P, Porat R, Nadeau JA, O'Neill SD. 1996. Identification of a meristem L1 layer-specific gene in Arabidopsis that is expressed during embryonic pattern formation and defines a new class of homeobox genes. *Plant Cell* **8**: 2155-2168.

Ma J, Ptashne M. 1987. A New Class of Yeast Transcriptional Activators. *Cell* **51**: 113-119.

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E. 2009. Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324**: 1064-1068.

Magnani E, Barton MK. 2011. A per-ARNT-sim-like sensor domain uniquely regulates the activity of the homeodomain leucine zipper transcription factor REVOLUTA in *Arabidopsis*. *Plant Cell* **23**: 567-582.

Marks MD, Wenger JP, Gilding E, Jilk R, Dixon RA. 2009. Transcriptome analysis of *Arabidopsis* wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. *Mol Plant* **2**: 803-822.

Masucci JD, Rerie WG, Foreman DR, Zhang M, Galway ME, Marks MD, Schiefelbein JW. 1996. The homeobox gene *GLABRA2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana*. *Development* **122**: 1253-1260.

Mathieu AP, Fleury A, Ducharme L, Lavigne P, LeHoux JG. 2002. Insights into steroidogenic acute regulatory protein (StAR)-dependent cholesterol transfer in mitochondria: evidence from molecular modeling and structure-based thermodynamics supporting the existence of partially unfolded states of StAR. *J Mol Endocrinol* **29**: 327-345.

Mukherjee K, Burglin TR. 2006. MEKHLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins. *Plant Physiology* **140**: 1142-1150.

Murcia M, Faraldo-Gomez JD, Maxfield FR, Roux B. 2006. Modeling the structure of the StART domains of MLN64 and StAR proteins in complex with cholesterol. *Journal of Lipid Research* **47**: 2614-2630.

Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T. 2006. Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in *Arabidopsis*. *Plant Physiology* **141**: 1363-1375.

Nobusawa T, Okushima Y, Nagata N, Kojima M, Sakakibara H, Umeda M. 2013. Synthesis of Very-Long-Chain Fatty Acids in the Epidermis Controls Plant Organ Growth by Restricting Cell Proliferation. *Plos Biol* **11**.

Ogawa E, Yamada Y, Sezaki N, Kosaka S, Kondo H, Kamata N, Abe M, Komeda Y, Takahashi T. 2015. ATML1 and PDF2 Play a Redundant and Essential Role in *Arabidopsis* Embryo Development. *Plant Cell Physiol* **56**: 1183-1192.

Ohashi Y, Oka A, Rodrigues-Pousada R, Possenti M, Ruberti I, Morelli G, Aoyama T. 2003. Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. *Science* **300**: 1427-1430.

Olayioye MA. 2005. StarD10, a START Domain Protein Overexpressed in Breast Cancer, Functions as a Phospholipid Transfer Protein. *J Biol Chem* **280**: 27436-27442.

Olayioye MA, Hoffmann P, Pomorski T, Armes J, Simpson RJ, Kemp BE, Lindeman GJ, Visvader JE. 2004. The phosphoprotein StarD10 is overexpressed in breast cancer and cooperates with ErbB receptors in cellular transformation. *Cancer Res* **64**: 3538-3544.

Oshea EK, Rutkowski R, Kim PS. 1989. Evidence That the Leucine Zipper Is a Coiled Coil. *Science* **243**: 538-542.

Palena CM, Tron AE, Bertoncini CW, Gonzalez DH, Chan RL. 2001. Positively charged residues at the N-terminal arm of the homeodomain are required for efficient DNA binding by homeodomain-leucine zipper proteins. *J Mol Biol* **308**: 39-47.

Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF et al. 2009. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324**: 1068-1071.

Pendaries C, Tronchere H, Plantavid M, Payrastre B. 2003. Phosphoinositide signaling disorders in human diseases. *FEBS Letters* **546**: 25-31.

Peterson KM, Shyu C, Burr CA, Horst RJ, Kanaoka MM, Omae M, Sato Y, Torii KU. 2013. *Arabidopsis* homeodomain-leucine zipper IV proteins promote stomatal development and ectopically induce stomata beyond the epidermis. *Development* **140**: 1924-1935.

Ponting CP, Aravind L. 1999. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem Sci* **24**: 130-132.

Prigge MJ, Clark SE. 2006. Evolution of the class III HD-Zip gene family in land plants. *Evol Dev* **8**: 350-361.

Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE. 2005. Class III Homeodomain-Leucine Zipper Gene Family Members Have Overlapping, Antagonistic, and Distinct Roles in *Arabidopsis* Development. *Plant Cell* **17**: 61-76.

Radauer C, Lackner P, Breiteneder H. 2008. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol Biol* **8**: 286.

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. 2002. MicroRNAs in plants. *Genes Dev* **16**: 1616-1626.

Rerie WG, Feldmann KA, Marks MD. 1994. The *GLABRA2* gene encodes a homeo domain protein required for normal trichome development in *Arabidopsis*. *Genes Dev* **8**: 1388-1399.

Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, Rajashankar KR, Cohen DE. 2002. Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat Struct Biol* **9**: 507-511.

Rodriguez-Agudo D, Ren S, Wong E, Marques D, Redford K, Gil G, Hylemon P, Pandak WM. 2008. Intracellular cholesterol transporter StarD4 binds free cholesterol and increases cholesteryl ester formation. *Journal of Lipid Research* **49**: 1409-1419.

Romanowski MJ. 2002. Crystal structure of the Mus musculus cholesterol-regulated START protein 4 (StarD4) containing a StAR-related lipid transfer domain. *Proc Natl Acad Sci USA* **99**: 6949-6954.

Rombola-Caldentey B, Rueda-Romero P, Iglesias-Fernandez R, Carbonero P, Onate-Sanchez L. 2014. *Arabidopsis* DELLA and two HD-ZIP transcription factors regulate GA signaling in the epidermis through the L1 box cis-element. *Plant Cell* **26**: 2905-2919.

Roostaee A, Barbar É, Lavigne P, Lehoux J-G. 2009. The mechanism of specific binding of free cholesterol by the steroidogenic acute regulatory protein: evidence for a role of the C-terminal α-helix in the gating of the binding site. *Bioscience Reports* **29**: 89.

Sakakibara K, Nishiyama T, Kato M, Hasebe M. 2001. Isolation of homeodomain-leucine zipper genes from the moss *Physcomitrella patens* and the evolution of homeodomain-leucine zipper genes in land plants. *Mol Biol Evol* **18**: 491-502.

San-Bento R, Farcot E, Galletti R, Creff A, Ingram G. 2014. Epidermal identity is maintained by cell-cell communication via a universally active feedback loop in *Arabidopsis thaliana. Plant J* **77**: 46-58.

Savaldi-Goldstein S, Peto C, Chory J. 2007. The epidermis both drives and restricts plant shoot growth. *Nature* **446**: 199-202.

Schrick K, Nguyen D, Karlowski WM, Mayer KF. 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol* **5**: R41.

Sessions A, Weigel D, Yanofsky MF. 1999. The *Arabidopsis thaliana* MERISTEM LAYER 1 promoter specifies epidermal expression in meristems and young primordia. *Plant J* **20**: 259-263.

Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ et al. 2014. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. *Nature Communications* **5**: 5519.

Shen B, Sinkevicius KW, Selinger DA, Tarczynski MC. 2006. The homeobox gene *GLABRA2* affects seed oil content in *Arabidopsis*. *Plant Mol Biol* **60**: 377-387.

Skwarek LC, Boulianne GL. 2009. Great expectations for PIP: phosphoinositides as regulators of signaling during development and disease. *Developmental Cell* **16**: 12-20.

Soccio RE. 2003. StAR-related Lipid Transfer (START) Proteins: Mediators of Intracellular Lipid Metabolism. *J Biol Chem* **278**: 22183-22186.

Stocco DM. 2002. Clinical disorders associated with abnormal cholesterol transport: mutations in the steroidogenic acute regulatory protein. *Mol Cell Endocrinol* **191**: 19-25.

Sugawara T, Holt JA, Driscoll D, Strauss JF, 3rd, Lin D, Miller WL, Patterson D, Clancy KP, Hart IM, Clark BJ et al. 1995. Human steroidogenic acute regulatory protein: functional activity in COS-1 cells, tissue-specific expression, and mapping of the structural gene to 8p11.2 and a pseudogene to chromosome 13. *Proc Natl Acad Sci U S A* **92**: 4778-4782.

Swanton C, Marani M, Pardo O, Warne PH, Kelly G, Sahai E, Elustondo F, Chang J, Temple J, Ahmed AA et al. 2007. Regulators of mitotic arrest and ceramide metabolism are determinants of sensitivity to paclitaxel and other chemotherapeutic drugs. *Cancer Cell* **11**: 498-512.

Takada S, Lida H. 2014. Specification of epidermal cell fate in plant shoots. *Front Plant Sci* **5**.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* **30**: 2725-2729.

Tang D, Ade J, Frye CA, Innes RW. 2005. Regulation of plant defense responses in *Arabidopsis* by EDR2, a PH and START domain-containing protein. *Plant J* **44**: 245-257.

Tavares R, Aubourg S, Lecharny A, Kreis M. 2000. Organization and structural evolution of four multigene families in *Arabidopsis thaliana*: AtLCAD, AtLGT, AtMYST and AtHD-GL2. *Plant Mol Biol* **42**: 703-717.

Thorsell AG, Lee WH, Persson C, Siponen MI, Nilsson M, Busam RD, Kotenyova T, Schuler H, Lehtio L. 2011. Comparative structural analysis of lipid binding START domains. *PLoS One* **6**: e19521.

Tominaga-Wada R, Iwata M, Sugiyama J, Kotake T, Ishida T, Yokoyama R, Nishitani K, Okada K, Wada T. 2009. The GLABRA2 homeodomain protein directly regulates *CESA5* and *XTH17* gene expression in *Arabidopsis* roots. *Plant J* **60**: 564-574.

Tron AE, Bertoncini CW, Chan RL, Gonzalez DH. 2002. Redox regulation of plant homeodomain transcription factors. *J Biol Chem* **277**: 34800-34807.

Tsujishita Y, Hurley JH. 2000. Structure and lipid transport mechanism of a StAR-related domain. *Nat Struct Biol* **7**: 408-414.

Turchi L, Baima S, Morelli G, Ruberti I. 2015. Interplay of HD-Zip II and III transcription factors in auxin-regulated plant development. *Journal of Experimental Botany* **66**: 5043-5053.

van Hengel AJ, Barber C, Roberts K. 2004. The expression patterns of arabinogalactan-protein AtAGP30 and GLABRA2 reveal a role for abscisic acid in the early stages of root epidermal patterning. *Plant J* **39**: 70-83.

Vinayavekhin N, Homan EA, Saghatelian A. 2010. Exploring disease through metabolomics. *ACS Chemical Biology* **5**: 91-103.

Vorwerk S, Schiff C, Santamaria M, Koh S, Nishimura M, Vogel J, Somerville C, Somerville S. 2007. EDR2 negatively regulates salicylic acid-based defenses and cell death during powdery mildew infections of *Arabidopsis thaliana*. *BMC Plant Biol* **7**: 35.

Wang RL, Stec A, Hey J, Lukens L, Doebley J. 1999. The limits of selection during maize domestication. *Nature* **398**: 236-239.

Wang XY, Wang XL, Hu QN, Dai XM, Tian HN, Zheng KJ, Wang XP, Mao TL, Chen JG, Wang SC. 2015. Characterization of an activation-tagged mutant uncovers a role of *GLABRA2* in anthocyanin biosynthesis in *Arabidopsis*. *Plant J* **83**: 300-311.

Western TL, Young DS, Dean GH, Tan WL, Samuels AL, Haughn GW. 2004. MUCILAGE-MODIFIED4 encodes a putative pectin biosynthetic enzyme developmentally regulated

by APETALA2, TRANSPARENT TESTA GLABRA1, and GLABRA2 in the *Arabidopsis* seed coat. *Plant Physiology* **134**: 296-306.

Won SK, Lee YJ, Lee HY, Heo YK, Cho M, Cho HT. 2009. cis-Element- and Transcriptome-Based Screening of Root Hair-Specific Genes and Their Functional Characterization in *Arabidopsis*. *Plant Physiology* **150**: 1459-1473.

Wymann MP, Schneiter R. 2008. Lipid signalling in disease. *Nature Reviews Molecular Cell Biology* **9**: 162-176.

Zalewski CS, Floyd SK, Furumizu C, Sakakibara K, Stevenson DW, Bowman JL. 2013. Evolution of the class IV HD-zip gene family in streptophytes. *Mol Biol Evol* **30**: 2347-2365.

Zhang F, Zuo K, Zhang J, Liu X, Zhang L, Sun X, Tang K. 2010. An L1 box binding protein, GbML1, interacts with GbMYB25 to control cotton fibre development. *Journal of Experimental Botany* **61**: 3599-3613.

Zhou GK, Kubo M, Zhong R, Demura T, Ye ZH. 2007. Overexpression of miR165 affects apical meristem formation, organ polarity establishment and vascular development in *Arabidopsis*. *Plant Cell Physiol* **48**: 391-404.

Chapter 2 - Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity

ABSTRACT

Background

Steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domains were first identified from mammalian proteins that bind lipid/sterol ligands via a hydrophobic pocket. In plants, predicted START domains are predominantly found in homeodomain leucine zipper (HD-Zip) transcription factors that are master regulators of cell-type differentiation in development. Here we utilized studies of *Arabidopsis* in parallel with heterologous expression of START domains in yeast to investigate the hypothesis that START domains are versatile ligand-binding motifs that can modulate transcription factor activity.

Results

Our results show that deletion of the START domain from *Arabidopsis* GLABRA2 (GL2), a representative HD-Zip transcription factor involved in differentiation of the epidermis, results in a complete loss-of-function phenotype, although the protein is correctly localized to the nucleus. Despite low sequence similarly, the mammalian START domain from StAR can functionally replace the HD-Zip-derived START domain. Embedding the START domain within a synthetic transcription factor in yeast, we found that several mammalian START domains from StAR, MLN64 and PCTP stimulated transcription factor activity, as did START domains from two *Arabidopsis* HD-Zip transcription factors. Mutation of ligand-binding residues within StAR START reduced this activity, consistent with the yeast assay monitoring ligand-binding. The D182L missense mutation in StAR START was shown to affect GL2 transcription factor activity in maintenance of the leaf trichome cell fate. Analysis of *in vivo* protein-metabolite interactions by mass spectrometry provided direct evidence for analogous lipid-binding activity in mammalian and plant START domains in the yeast system. Structural modeling predicted

similar sized ligand-binding cavities of a subset of plant START domains in comparison to mammalian counterparts.

Conclusions

The START domain is required for transcription factor activity in HD-Zip proteins from plants, although it is not strictly necessary for the proteins nuclear localization. START domains from both mammals and plants are modular in that they can bind lipid ligands to regulate transcription factor function in a yeast system. The data provide evidence for an evolutionarily conserved mechanism by which lipid metabolites can orchestrate transcription. We propose a model in which the START domain is used by both plants and mammals to regulate transcription factor activity.

Contribution

The results of the domain swap experiment and the effect of D182L missense mutation in StAR START on GL2 transcription factor activity are my contribution to the publication: Kathrin Schrick, Michael Bruno, **Aashima Khosla**, Paige N Cox, Sara A Marlatt, Remigio A Rogue, Henry C Nguyen, Cuiwen He, Michael P Snyder. Daljit Singh, and Gitanjali Yadav. (2014). Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity. BMC Biology 12:70.

INTRODUCTION

Steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) is an evolutionarily conserved module of approximately 200 amino acids implicated in lipid/sterol binding (Ponting and Aravind 1999). The prototype for the START domain is found in mammalian StAR proteins that bind cholesterol for the initiation of steroidogenesis (Lavigne et al. 2010). The START domain subfamily belongs to an expansive clan of α/β helix-grip-fold structures that is given the name 'SRPBCC (START/RHO_alpha_C/PITP/Bet_v1/CoxG/CalC) superfamily' in the NCBI Conserved Domains Database, named for six of its major subfamilies. Another subfamily termed PYR/PYL/RCAR-like includes the newly identified abscisic acid receptors from plants (Iyer et al. 2001; Radauer et al. 2008). Members of the SRPBCC superfamily also referred to in the literature as the Bet_v1-like (Ma et al. 2009) or START (Park et al. 2009) superfamily, exhibit the common property of a deep hydrophobic ligand-binding pocket. The focus of the present study is on proteins carrying domains of the START subfamily, and we restrict the term 'START' to the subfamily of proteins that share significant sequence similarity to mammalian StAR, as described in Schrick et al. (2004). START domains vary in size and configuration and occur primarily in animals and plants as well as some species from other taxa. While the StAR protein contains the START domain as the major motif, START domains are modular in that they are found in combination with other functional domains. Ligand-binding by the START domain in a multidomain protein may regulate the activities of other domains such as Rho-GAP and thioesterase domains that occur in human START domain-containing proteins.

Experimental evidence for START domains as ligand-binding motifs derives largely from mammalian proteins (Alpy and Tomasetto 2005; Alpy et al. 2009). The human genome encodes 15 START domain-containing proteins, several of which are implicated in disease (Clark 2012). Human StAR/STARD1 is associated with an inherited disorder known as congenital lipoid adrenal hyperplasia (CAH) (Bhangoo et al. 2006). Identified ligands for mammalian proteins include sterols such as cholesterol and hydroxycholesterol, and other lipids such as phospholipids and sphingolipids. More recently bile acids were shown to bind human STARD5 (Letourneau et al. 2012), expanding the repertoire of possible interactors. Mammalian START domain crystal structures are available, revealing central antiparallel β-sheets and a carboxy-terminal α-helix forming a hydrophobic cavity to accommodate a single ligand

(Tsujishita and Hurley 2000; Roderick et al. 2002; Romanowski et al. 2002; Kudo et al. 2008; Thorsell et al. 2011). Ligand specificity is thought to be conferred by the configuration of amino acids that form the inner lining of the cavity. Amino acid changes that perturb START domain function were uncovered in defective StAR genes from CAH patients, and specific amino acid residues are predicted to affect cholesterol binding (Tsujishita and Hurley 2000). The co-crystal of phosphatidylcholine transfer protein (PCTP)/STARD2 bound to phosphatidylcholine (PtCho) identified multiple residues that contact the ligand in the hydrophobic pocket (Roderick et al. 2002). The structures of CERT/STARD11 in its apo-form and in complex with ceramides of variable acyl chain lengths indicate contact points between amino acids and molecular features of sphingolipids (Kudo et al. 2008).

START domain-containing proteins are abundant in the plant kingdom, where the majority are members of a plant-specific homeodomain leucine zipper (HD-Zip) transcription factor family (Schrick et al. 2004). *Arabidopsis* contains 21 HD-Zip START domain-containing transcription factors of the class III and IV subfamilies. Genetic analysis indicates key roles in cell differentiation and patterning in development, and several family members exhibit striking mutant phenotypes. The class III HD-Zip family contains five proteins implicated in vasculature, meristem initiation and/or organ polarity (Prigge et al. 2005). The larger class IV HD-Zip family comprises 16 members involved in cell fate determination (Nakamura et al. 2006), and includes *Arabidopsis MERISTEM LAYER 1 (ATML1)* and its close relative *PROTODERMAL FACTOR 2 (PDF2)*, which are required for epidermal cell fate of the shoot (Lu et al. 1996; Abe et al. 2003), and *GLABRA2 (GL2*), which is required for the specification of epidermal cells in the shoot (Rerie et al. 1994), root (Di Cristina et al. 1996) and seed (Western et al. 2001).

Although START domains are amplified in HD-Zip transcription factors, the function of the plant domains in lipid/sterol binding has not been verified, nor has the crystal structure been solved for any plant-derived START-subfamily domain to date. Since HD-Zip proteins are transcription factors, one hypothesis is that the START domain, by binding lipid/sterol ligands, controls transcription analogously to steroid hormone receptors in metazoans (Ponting and Aravind 1999; Schrick et al. 2004). One advantage of the proposed mechanism is that the metabolic state of a cell as reflected in lipid/sterol synthesis could dictate cell-type specific mRNA expression. Such a mechanism would be of broad interest since many organisms contain START domain proteins whose functions have not yet been uncovered.

In this study, the overall aim was to investigate the role of the START domain in transcription. We demonstrate that the START domain is required for transcription factor activity of a representative HD-Zip protein, GL2, in *Arabidopsis*. A domain swap experiment indicated that although START domains in mammals and plants exhibit marginal sequence similarity, their function appears to be conserved. We utilized a yeast system to assess the roles of START domains in the context of transcription, and to compare the behavior of known ligand-binding START domains from mammals to 'orphan' START domains from plants. With this approach, ligand-binding activity for several mammalian domains with known ligands was reproduced in yeast. START domains from both plants and mammals were shown to stimulate transcription factor activity when placed in a yeast synthetic transcription factor. Immunoisolation of the synthetic transcription factors followed by mass spectrometry indicated that START domains from two HD-Zip transcription factors participate in protein-metabolite interactions as do the mammalian counterparts. We provide experimental evidence for a model in which ligand-binding to the START domain regulates transcription factor activity, linking metabolism to gene expression.

RESULTS

Deletion of the START domain from HD-Zip transcription factor GL2 results in loss of activity

To probe the function of the START domain within HD-Zip transcription factors we chose *Arabidopsis GL2* for analysis. The *GL2* gene product is dispensable for viability, but *gl2* null mutants exhibit distinct phenotypes in differentiation of the epidermis, including defects in leaf trichome development (Rerie et al. 1994) (Figure 2.1 A; Appendix Figure A-1), excessive root hair formation (Di Cristina et al. 1996) (Figure 2.1 B) and lack of seed mucilage production (Western et al. 2001) (Figure 2.1 C). The molecular nature of *gl2* null activity allele, *gl2-5* is described in Chapter 5 (methods). We deleted the START domain from a GL2 construct in which the cDNA sequence was translationally fused to the enhanced yellow fluorescent protein (EYFP) tag at its amino-terminus (Figure 2.1 D), and transformed homozygous *gl2* plants to examine complementation of the mutant phenotypes. The *EYFP:GL2* transgene was expressed under the native *GL2* promoter (*ProGL2*). Whereas the wild-type *ProGL2:EYFP:GL2* construct rescued all three mutant phenotypes in leaves, roots and seeds, the *ProGL2:EYFP:gl2*Δ*START* construct resulted in null phenotypes indistinguishable from the *gl2* loss-of-function mutant (Figure 2.1 A,B,C,E). Despite the inability of the *EYFP:gl2*Δ*START* transgene to confer phenotypic complementation, we observed nuclear localization in ovules and trichomes, similar to that for the wild-type *EYFP:GL2* transgene (Figure 2.2 A-E).

The START domain from mouse StAR functionally replaces the endogenous START domain from GL2

The 209 amino acid START domain from the mouse StAR domain shares 33% similarity and 13% identity with the 235 amino acid START domain from *Arabidopsis* GL2 (Figure 2.1 F). We used a domain swap experiment to test the ability of this mammalian StAR START domain to replace the *Arabidopsis* START domain functionally in the EYFP:GL2 protein cassette (Figure 2.1 D). Among the *EYFP:GL2-StAR-START* T2 transformants of *gl2*, we characterized several independent lines that exhibited partial rescue of the trichome defects and selected one of these lines for detailed analysis. In addition to partial complementation of the trichome defects

(Figure 2.1 A, E and Appendix Figure A-1), these lines also exhibited partial rescue of the root hair patterning (Figure 2.1 B) and seed mucilage defects (Figure 2.1 C) of *gl2* null mutants. The results indicate that despite relatively low sequence similarity (33%), a mammalian START domain is able to replace the plant START domain from GL2 functionally to a similar level as that of a related class IV HD-Zip family member, ATML1 (Lu et al. 1996) (Figure 2.1 A-E and Appendix Figure A-1). START domain swaps with mouse StAR or *Arabidopsis* ATML1 both resulted in nuclear localization of the EYFP:GL2 protein (Figure 2.2 F,G). The function of the START domain in GL2 was dependent on the encoded domain sequence, since replacement with two additional *Arabidopsis*-derived START domains, namely from ENHANCED DISEASE RESISTANCE 2 (EDR2) (Tang et al. 2005) and from class III HD-Zip member REVOLUTA (REV) (Talbert et al. 1995) , were not found to rescue the *gl2* mutant phenotype (Figure 2.1 A,B,C,E), nor did they display nuclear localization (Figure 2.2 H,I).

A yeast assay for START domain function in transcription

We developed an assay to monitor the activity of the START domain within a synthetic transcription factor using heterologous expression in the yeast *Saccharomyces cerevisiae*. The START domain-coding region from various proteins (Figure 2.3 A) was fused between an amino-terminal GAL4 DNA binding domain (DBD) and a carboxy-terminal VP16 activation domain (AD) (Figure 2.3 B). The GAL4-VP16 (GV) synthetic transcription factor is driven by the *MITOCHONDRIAL RIBOSOMAL PROTEIN 7* (*MRP7*) constitutive promoter, resulting in very low background levels of reporter gene expression. Activities of GAL4-DBD:START:VP16-AD (GSV) synthetic transcription factors were monitored by their ability to activate the *LacZ* reporter (Figure 2.3 C). The yeast strain harbored an *erg6* mutation to facilitate permeability and uptake of small molecules. In a GAL4-DBD:hER:VP16-AD (GEV) construct in which the human estrogen receptor steroid binding domain (hER) is placed between GAL4-DBD and VP16-AD domains, the level of reporter activity, quantified in β-galactosidase (β-Gal) units, was reduced to very low levels. However, these levels were increased ~25-fold by the addition of estradiol (Figure 2.3 D), consistent with a previous report demonstrating that hER is responsive to its biological ligand estradiol when expressed in a GEV transcription factor in yeast (Gao and Pinkham 2000). A GV control construct in which the GAL4-DBD was fused

directly to VP16-AD was tested and shown to display a ~6-fold lower level of activity (Figure 2.3 D).

START domains from mammalian proteins stimulate transcription factor activity in yeast

We studied the function of the START domain in the synthetic transcription factor with a mammalian StAR protein that is known to bind cholesterol. When the START domain from mouse StAR was placed between the GAL4-DBD and VP16-AD domains, we observed significantly elevated levels of β-galactosidase activity in the GSV plasmid in comparison to control plasmids, such as GV without the START domain (~3- and ~6-fold increase; Figure 2.3 D, E), and GEV-expressing yeast in the absence of estradiol (~10- and ~20-fold increase; Figure 2.3 D, E).

To examine the generality of this ability to stimulate transcription factor function, we tested the START domains from two other mammalian proteins, human metastatic lymph node 64 (MLN64/STARD3) and human PCTP. MLN64 is closely related to StAR and has been shown to bind cholesterol (Alpy and Tomasetto 2006), while the more distantly related PCTP binds PtCho (Roderick et al. 2002). Both MLN64 and PCTP behaved similarly to StAR in the ability to stimulate transcription factor activity in the yeast assay (Figure 2.3 E).

START domains from several *Arabidopsis* proteins also boost transcription factor activity

Based on amino acid similarity to the mammalian counterparts, the *Arabidopsis* genome encodes 21 HD-Zip transcription factors that contain START domains out of a total of 35 START domain-containing proteins (Schrick et al. 2004). Then 25 START domains from *Arabidopsis* were tested in the yeast assay and compared with the reporter gene activities of three mammalian START domains: StAR, MLN64 and PCTP (Figure 2.3 E). Although the majority of the *Arabidopsis* START domains assayed resulted in little or no reporter activity, a group of five START domains conferred enhanced transcription factor activity similar to that of the three mammalian proteins tested.

The START domains from class IV HD-Zip transcription factors ATML1 and PDF2 exhibited similar levels of activity (4.4- and 3.1-fold over GV control, and 25- and 17-fold over vector control, respectively; Figure 2.3 E). These proteins display a high degree of sequence

conservation (90% and 95% amino acid identity and similarity, respectively) and have been shown to be functionally redundant *in vivo* (Abe et al. 2003). Similarly, At5g07260, a HD-Zip-related protein of unknown function that lacks the HD domain, exhibited activity. In addition, two PCTP-like proteins, At4g14500 and At3g13062, displayed weak activities that were not significantly different from the GV control but were increased over other plant-derived START domains (Figure 2.3 E).

START domains do not behave as classical activation domains

One possible explanation for the increased transcription factor activity conferred by the START domain is that it acts as a classical transcriptional AD. We examined whether the START domain functions as an AD in combination with GAL4-DBD by deletion of VP16-AD from several GSV constructs. The resulting GAL4-DBD:START (GS) constructs containing mammalian START domains from StAR, MLN64, PCTP or *Arabidopsis* GL2, exhibited a lack of reporter activity in the yeast assay, comparable to that of the empty vector pRS314, about 10-fold lower than that of the GV control (Figure 2.3 C,F). Consistent with these data, when the START domain from *Arabidopsis* PDF2 or GL2 was expressed as a GAL4-DBD fusions in a yeast two-hybrid assay, there was no activation of reporter genes although expression of the myc-tagged GS proteins were detected by Western blot (AK and KS, in preparation). We conclude that the START domain is not a transcriptional activator in the classical sense, since it was found to boost the activity of the transcription factor only when VP16-AD was also present.

Over-expression of positive regulator of sterol biosynthesis increases reporter activity of GSV constructs containing active START domains

One possible interpretation to explain the effect of the START domain on the synthetic transcription factor is that endogenous levels of lipid/sterol metabolites in yeast act as ligands to regulate levels of activity. Over-expression of the yeast transcriptional regulator *Sterol Uptake 1* (*SUT1*) gene elevates sterol biosynthesis activity (Ness et al. 2001). High levels of *SUT1* likely result not only in an increase in sterols, but may also alter the expression profiles of other lipid metabolites needed for membrane synthesis. We found that co-expression of this multi-copy plasmid carrying the *SUT1* gene under control of the strong *PMA1* promoter results in increased

activity of the *LacZ* reporter (~2- to 12-fold) in GSV-expressing yeast (Figure 2.3 G). The reporter activity is only increased for those constructs that contain a functional START domain that exhibits activity in yeast, since constructs containing mutations in the START domain do not yield significant levels of transcriptional activity (see below). These data suggest that *SUT1* over-expression acts directly or indirectly on START domains within the GSV proteins.

Over-expression of the *SUT1* gene has also been shown to enhance sterol uptake (Ness et al. 1998). However, exogenous addition of cholesterol to the medium did not increase the activity of the *LacZ* reporter in *SUT1* over-expressing cells with GSV constructs containing the START domain from StAR (Appendix Figure A-2). With *SUT1* over-expressed, the reporter activity for GSV-StAR did not change with a range of cholesterol concentrations from 0 μM to 50 μM, perhaps due to saturation of ligand-binding.

Site-directed mutagenesis of StAR indicates requirement for ligand binding

Functional analysis of specific amino acids within the START domain of the mammalian StAR protein was carried out to test further the idea that the START domain is a ligand-binding module within the GSV transcription factor. Using site-directed mutagenesis, missense mutants were constructed on the basis of conservation and alignment with known binding sites in PCTP (Figure 2.4 A). Among the eight mutants tested, two were double mutants. The M143R;N147D double mutant has two amino acids altered that are conserved among the mammalian cholesterol-binding START domains. Altering the M/N pair to R/D is predicted to change the specificity of the StAR START domain from cholesterol-binding to zwitterion lipid-binding, making it more similar to that of PCTP (Tsujishita and Hurley 2000). Similarly, the L270M mutant is predicted to alter ligand specificity from cholesterol-binding to phospholipid-binding. The other missense mutants remove or create additional charged residues within the hydrophobic ligand-binding cavity. The L241R mutation disrupts a hydrophobic cluster within a PCTP binding region in the 11th β-sheet (β11) of the predicted structure. Likewise, the F266L mutation affects the charge of a residue that is both highly conserved and is a predicted ligand contact site. R181L, a mutation that corresponds to a substitution in human patients exhibiting lipoid CAH, affects the charge of a highly-conserved residue within β6 of the START domain that contains other residues implicated in direct ligand contact. In the R181L;D182L double mutant affecting the same region, two adjacent charged residues are changed to leucines. C223R, like R181L and

D182L, affects the charge of the cavity, although this residue is also not predicted to be in direct contact with the ligand (Roderick et al. 2002). However, an adjacent residue, Met at position 224, is mutated (M224T) in a subset of lipoid CAH patients (Tsujishita and Hurley 2000).

The activities of the mutants were compared to wild-type StAR (Figure 2.4 B). Although most of the mutations in StAR reduce or completely abolish function, the R181L mutant, which is predicted to affect lipid transfer but not binding (Murcia et al. 2006), displayed wild-type levels of activity. The two mutants, L270M and M143R;N147D, that were predicted to affect specificity, resulted in wild-type and partial activities, respectively.

To examine whether a mutant version of the StAR START domain affects transcription factor activity of the HD-Zip GL2 protein *in planta*, a *ProGL2:EYFP:GL2-StAR-START* transgene harboring the D182L missense mutation was constructed and transgenic lines were analyzed in comparison to the wild-type counterpart. Among several independent lines for which expression of the transgene was verified, fewer branched trichomes were observed compared to a *GL2* transgene containing the wild-type StAR-START domain (Figure 2.4 C-F; Appendix Figure A-3). Nuclear localization of the EYFP:GL2-StAR-D182L-START transcription factor indicated that the mutant protein is expressed and properly localized (Figure 2.4 G-J). Quantification of trichomes on first leaves revealed significantly reduced function for the GL2-StAR protein containing the D182L mutation compared to the wild-type version (Figure 2.4 K). These results indicate that a functional ligand-binding START domain is required for GL2 transcription factor activity.

Protein levels of GSV transcription factors and the relation with activity levels

One possibility for explaining the ability of the START domain to stimulate transcription factor activity is that it may stabilize protein levels when it is ligand-bound. We utilized the yeast system to further probe the mechanism by which the START domain functions to enhance transcription factor activity, by examining protein levels for a representative subset of the GSV synthetic transcription factors. Because we were not successful in detecting the GSV proteins with commercial antibodies against GAL4 or VP16, likely due to very low expression levels from the *MRP7* promoter, the GSV transcription factors were fused to the yeast-enhanced green fluorescent protein (yEGFP3) in plasmid pUG35 (Cormack et al. 1997) (Appendix Figure A-4). Green fluorescent protein (GFP) expression levels were quantified by flow cytometry (Appendix

Figure A-4, see methods). Each of the GSV:yEGFP3 constructs exhibited a GFP level above the negative control (Appendix Figure A-4), and the mean values of fluorescent signals were similar to each other (mean range: 36 to 49) in contrast to that of the control (mean: 12) (Appendix Figure A-4). These data indicate that each of the GSV:yEGFP fusion proteins is properly expressed in yeast.

We assayed the GSV-yEGFP3 constructs for reporter gene activity in the yeast assay (Appendix Figure A-4). Levels of expression, as monitored by % GFP, do not strictly correlate with transcription factor activity levels, since START domains that lacked activity (StAR^{C223R} and StAR^{F266L}) displayed similar % GFP levels (0.12 and 0.10) as START domains that displayed activity (ATML1 and PDF2) (0.11). However, the three GSV:yEYFP3 constructs that displayed the lowest percentage (0.03) of GFP cells (*Arabidopsis* GL2, mouse StAR^{D182L} and StAR^{L241R}), also displayed low levels of transcription factor activity. These results suggest that the ligand-binding activity of a given START domain may result in increased protein stability.

Identification of metabolites bound to START domains in the yeast system

To investigate the molecular nature of the ligands underlying activation of the GSV transcription factors, we used a modified version of a metabolite-protein isolation protocol that was developed for yeast proteins (Li et al. 2010). For these experiments, we chose a subset of GSV transcription factors carrying START domains from mouse StAR, human PCTP, *Arabidopsis* HD-Zip transcription factors PDF2 and GL2, as well as an *Arabidopsis* PCTP-like protein At3g13062, and tagged these GSV proteins with an immunoglobulin G (IgG)-binding domain (Gelperin et al. 2005) (Figure 2.5 A). As negative controls, we included the D182L mutant version of the START domain from mouse StAR that exhibits reduced activity in both yeast and plants (Figure 2.4) and the GV fusion that lacks the START domain.

Small metabolites bound to the tagged proteins were identified by affinity protein purification followed by mass spectrometry. We focused on the metabolites significantly enriched in the GSV transcription factors relative to the GV control. A shortlist of candidate masses were queried against the Human Metabolome Database (HMDB) and METLIN to identify and assign a metabolite name to the enriched features at a high or low confidence level. Purified proteins corresponding to each sample were examined by SDS-PAGE, which indicated expression of GSV-IgG (and GV-IgG control) proteins of expected sizes (Figure 2.5 B). In

parallel, the GSV-IgG constructs were tested for activity in the yeast assay (Figure 2.5 C), and shown to behave in a similar manner as the original GSV constructs (Figures 2.3 E and 2.4 B). While GSV transcription factors containing START domains from mammalian StAR and PCTP, as well as *Arabidopsis* PDF2 and At3g13062 displayed activity levels that were increased over the GV control, the START domains from the $StAR^{D182L}$ mutant and *Arabidopsis* GL2 displayed low levels of activity.

Hierarchical clustering of the mass spectrometry data revealed several distinct clusters of protein-metabolite interactions (Figure 2.5 D). The horizontal dendrogram places the GSV-StAR transcription factor in a separate cluster from the missense mutant $StAR^{D182L}$. The enriched metabolites of $StAR^{D182L}$, which clustered with the *Arabidopsis* At3g13062 START domain, are distinguished by an abundance of several distinct species of diacylglycerols. The remaining proteins are grouped by the horizontal clustering dendrogram such that both *Arabidopsis* PDF2 and GL2 START domains have more similar bound-metabolite preferences, with human PCTP being more closely related to this group than the mouse StAR domain. In addition to several metabolites that were shared with StAR (e.g. lanosterol, squalene, PtCho, diacylglycerol and LysoPtCho), PCTP was shown to bind reproducibly to several types of lipids including sphinganine, sphingosine, behenic acid and linoleamide. The vertical dendrogram reveals a major metabolite cluster comprising StAR, PCTP, PDF2 and GL2 that contains an overrepresentation of LysoPtCho. Metabolites common to PCTP, StAR and the mutant $StAR^{D182L}$ form a distinct cluster, which includes some phospholipids, such as PtCho and LysoPtCho, as well as several sterols including four cholesterol esters (CEs). This cluster also includes two major precursors of the cholesterol biosynthesis pathway, lanosterol and squalene. The mutant $StAR^{D182L}$ is significantly less enriched for CE, PtCho, LysoPtCho, squalene and lanosterol within this cluster.

Although the overall chemical nature of the bound metabolites are diverse in class, glycerolipids and phospholipids together comprise the majority of interacting metabolites identified. We generated a comprehensive protein-metabolite network (Appendix Figure A-5) for all six START domains tested. To compare and contrast metabolite-binding preferences better, we constructed sub-networks (Appendix Figure A-5 and Figure 2.6 A) from a more stringent filtering, and included only high confidence and validated metabolites. The data support a

hypothesis that the START domains from plant and mammalian origin can interact with a common set of metabolic ligands (Figure 2.6).

Structural modeling of *Arabidopsis* START domain ligand-binding cavities

From their X-ray crystal structures, the ligand-binding cavity volumes of mammalian START domains from MLN64 (848 ± 107 $Å^3$) (Tsujishita and Hurley 2000) and PCTP (~ 882 $Å^3$) are highly similar although they bind different hydrophobic ligands. Cholesterol, the ligand of MLN64, occupies a 741 $Å^3$ volume, while PtCho, the ligand of PCTP, occupies 797 $Å^3$. Based on our domain swap experiments in *Arabidopsis* (Figures 2.1 and 2.2) and GSV transcription factor activities in the yeast assay (Figure 2.3 E), and binding activities in the protein-metabolite immunoisolations (Figure 2.5), *Arabidopsis* START domains from class IV HD-Zip transcription factors are predicted to exhibit similar ligand-binding cavities as their mammalian counterparts.

Using available three-dimensional coordinates for crystal structures of mammalian START domains and threading-based fold prediction, the closest structural homolog for each of the *Arabidopsis* START domains tested in the yeast assay was determined to be the mouse STARD4 domain (Figure 2.7 A). Three-dimensional homology models were constructed and the corresponding hydrophobic tunnels were assessed for size, volume and identities of the cavity lining residues. The modeling results indicated that the hydrophobic tunnels of the *Arabidopsis* START domains fall into three categories of volume sizes. Plant START cavities were found to be either small in size (Figure 2.7 B; volumes less than 400 A^3), medium (Figure 2.7 C; volumes between 400 and 700 $Å^3$) or large (Figure 2.7 D; volumes >700 $Å^3$). The five START domains showing activity in the yeast assay fall into the large volume category. For example, START domains from HD-Zip transcription factors ATML1 and PDF2, which exhibit activity in the yeast assay, have predicted volumes of 770 and 720 $Å^3$, respectively. In contrast, START domains from other HD-Zip transcription factors, such as GL2 and CORONA (CNA), which had little or no activity in the yeast assay, fall into the small volume category with tunnels of 284 $Å^3$ and 320 $Å^3$, respectively. Among the plant START domains that exhibited a negative result, three others are predicted to have similarly small cavities.

Detailed examination of the predicted cavity-lining amino acid residues in each of these cases revealed that certain side-chain residues are oriented into the space within the original tunnel, thereby blocking the passage and restricting the cavity size (Figure 2.7 E). Some frequent

blocking positions correspond to N147, A178 and G200 (StAR numbering), all of which are replaced by residues with much larger side chains in *Arabidopsis* START domains, such as Arg, Trp/Tyr/Phe/Gln/Asn and Ile/Leu, respectively. Most of the domains that lack activity in the yeast assay correspond to volumes that fall into the medium category, while two domains from HD-Zip transcription factors (HDG2 and FWA/HDG6) have cavities of the large volume category. In the START domain from HDG2, the opening of the tunnel is slightly shifted, such that a potential ligand would be predicted to encounter steric hindrance in entering the hydrophobic tunnel. Thus, steric factors may explain why some of the plant-derived START domains fail to exhibit activity in the yeast assay. Alternatively, the specific ligands for these START domain proteins may not be present in yeast.

DISCUSSION

A role for the START domain in the regulation of the transcription

Here we chose GL2 to investigate the significance of the START domain in HD-Zip transcription factors. Our results indicate that the START domain is required for transcription factor activity (Figures 2.1 and 2.2). Strikingly, the GL2 START domain could be functionally replaced by the mouse StAR START domain, a mammalian counterpart. The GL2 START domain could similarly be replaced by the START domain from ATML1, a transcription factor from the same class IV HD-Zip subfamily that shares a relatively high degree of amino acid similarity, but other *Arabidopsis* START domains of lower similarity (EDR2 and REV; Figure 2.1 F) could not functionally replace the GL2 START domain, indicating that a specific sequence or structural feature is critical.

The ability of the mouse StAR START domain to confer partial complementation may reflect the presence of the appropriate ligand (i.e. cholesterol) in GL2-expressing cells, whereas the ligands of EDR2 or REV START, which are unknown to date, are potentially not present in these cells. The conclusion that the mouse StAR START domain participates in ligand-binding in plant cells is further supported by our finding that the D182L mutation results in reduced GL2 activity (Figure 2.4). Alternatively, there is a function unrelated to ligand-binding that the START domains of StAR and ATML1 can fulfill. Although our structural modeling data suggest different cavity volumes for StAR and GL2 (as well as ATML1 and GL2), the protein-metabolite immunoisolation data from yeast indicate an overlap in binding partners between GL2 and StAR, as well as between GL2 and PDF2, the closest relative of ATML1. It is noteworthy that the EDR2 and REV START domain swaps are expressed but are not localized in nuclei (Figure 2.2 H, I). One possible explanation is that these START domains yield improperly folded transcription factors, which interfere with transport to the nucleus. However, since the GL2 protein lacking the START domain is found in the nucleus, we conclude that the START domain is not strictly required for correct subcellular localization.

START domains stimulate transcription factor activity via ligand-binding

It was hypothesized that HD-Zip transcription factors from plants are regulated via lipid/sterol ligands that bind to their START domains (Ponting and Aravind 1999; Schrick et al.

2004). The yeast assay described here tests the function and ligand-binding properties of START domains within a synthetic transcription factor, and indicates that START domains of mammalian or plant origin can regulate activity in a similar fashion (Figure 2.3). The START domains do not behave as classical ADs in the absence of VP16 (Figure 2.3 F), and based on our site-directed mutagenesis, protein-metabolite immunoisolation experiments, and structural modeling, we conclude that the likely mode of action is via ligand-binding. Since several START domain-containing proteins from humans have been found in the nucleus, our findings support a possible role for the START domain in regulating gene expression in both plants and animals.

Relevant to our results, mammalian PCTP was previously shown to interact physically with HD transcription factor Pax3 and the two proteins are co-localized in the nuclei of mouse embryos (Kanno et al. 2007) (Figure 2.6 B). The previous study showed that PCTP co-activated the transcriptional activity of Pax3 in tissue culture cells, and it was postulated that the physical interaction of PCTP with Pax3 positively regulates its transcriptional activity. Perhaps the interaction between PCTP and Pax3 is analogous to the function of the START domain in plant HD-Zip transcription factors (Figure 2.6 B). While the previous study examined the function of the START domain in an interacting protein (Kanno et al. 2007), here we demonstrate the regulatory role of the START domain as an integral part of the transcription factor.

We probed three mammalian START domains and 25 *Arabidopsis* START domains for the ability to stimulate transcription factor activity in yeast (Figure 2.3 E). While all three mammalian START domains exhibited activity in comparison to the control, only 5/25 of the *Arabidopsis* START domains displayed similar levels of activity. Using protein-metabolite immunoisolation experiments (Figures 2.5 and 2.6, and Appendix Figure A-5), we tested the idea that lipids/sterols present in yeast bind the corresponding START domains, resulting in the observed activity. We reasoned that START domains that lack activity might require plant-specific metabolites not found in yeast. The *Arabidopsis* GL2-START domain, which displays little or no activity in yeast, was found to bind to a large number of metabolites. The range of metabolites bound to the START domains from *Arabidopsis* PDF2 (showing activity) and GL2 (lacking activity) were overlapping but not identical, leaving the possibility that those specific to GL2 might behave as negative regulators. In general, protein-metabolite interactions may serve to: (a) activate or negatively regulate proteins, (b) stabilize protein levels, (c) facilitate protein-

protein interactions, and/or (d) affect protein sub-cellular localization. In addition to activating or negatively regulating transcription factor function, our experiments with the yEGFP3 fusions provide evidence that ligand-binding stabilizes protein levels of the GSV transcription factors (Appendix Figure A-4). In addition, ligand-binding may also contribute to correct protein folding. Site-directed mutagenesis of the START domain from mouse StAR was utilized in conjunction with the yeast assay to characterize START domain function within the synthetic transcription factor. Our results indicate that transcription factor activity correlates with predicted ligand-binding capacity of the START domain (Figure 2.4). From the protein-metabolite immunoisolation experiments, comparison of the wild-type START domain from StAR versus the mutant StAR^{D182L} reveals that many more metabolites (including CEs) bind to the wild-type StAR, correlating high levels of activity with ligand-binding in the yeast assay. Consistent with the interpretation that our yeast assay measures ligand-binding, the StAR^{R181L} mutation, which displays wild-type levels of activity, was reported not to affect ligand-binding affinity *in vitro* (Baker et al. 2007). In transfected mammalian cells, this missense mutation has an adverse effect on cholesterol transfer and the corresponding conserved Arg may be required for a conformational change that allows release of bound ligand. The StAR mutants that are predicted to change the specificity of ligand-binding had little (StAR^{L270D}) or no affect (StAR$^{M143R; N147D}$) on the activity of the START domain in the yeast assay. This finding was consistent with the results from our protein-metabolite immunoisolation experiments, in which a range of lipid metabolites including PtCho and sterol-related molecules were found associated with the tagged GSV transcription factor containing the START domain from StAR (Figures 2.5 and 2.6 A). We propose that interaction of START domains with accessory proteins, via tissue-specific expression and proper subcellular compartmentalization, are likely to contribute to specificity in binding of selected metabolite ligands during the lifespan of the protein within the cells of the host organism.

Immunoisolation experiments reveal novel protein-metabolite interactions for START domains

The GSV protein-metabolite immunoisolation experiments performed in yeast uncovered potential lipid ligands for mammalian StAR and PCTP that had previously been shown to bind

these START domains, namely cholesterol (CE) and PtCho, respectively (Figure 2.5 and Appendix Figure A-5). In addition, novel metabolite-protein associations were found, including known modulators of lipids and sterols in mammals, such as sphinganine, behenic acid and LysoPtCho. Precursors and intermediates of the sterol biosynthesis pathway, including lanosterol and squalene, were also significantly enriched for the START domain from StAR.

There is mounting evidence for a broad range of potential metabolite binding partners for a given START domain. In a previous study, it was shown that StAR can stimulate the transfer of both cholesterol and β-sitosterol, a plant sterol, from liposomes to mitochondria (Kallen et al. 1998). A structural study showed that in addition to ceramides, the START family member CERT binds diacylglycerol (Kudo et al. 2008), another metabolite that was uncovered here. It is noteworthy that key lipid metabolites in low-density lipoprotein (LDL) transport, namely CE and LysoPtCho, which are converted from cholesterol and PtCho by lecithin-cholesterol acyl transferase (LCAT), an enzymatic reaction that is critical for LDL transport in the cell, were found to be associated with several START domains that we tested. The identified START domain interactor sphinganine is a known cholesterol transport inhibitor that blocks LDL-induced formation of CE, the product of LCAT activity. Behenic acid, which was found enriched only in PCTP and not the other START domains, has been reported to increase cholesterol levels in humans (Cater and Denke 2001). Taken together, our findings support the premise that START domains are central players in metabolic regulatory networks in the cell.

CONCLUSIONS

The START domain has been characterized primarily as a lipid/sterol-binding domain in mammals, but its function in plant HD-Zip transcription factors was previously not known. Here we show that the START domain is required for transcription factor activity of GL2. We report evidence that START domains from the plant kingdom can act as ligand-binding modules, similar to their mammalian counterparts, and that they can modulate transcription factor activity. The protein-metabolite immunoisolation experiments performed in yeast generated a list of small molecules that are candidates for binding the *Arabidopsis* START domains from HD-Zip transcription factors within the physiological context of plant growth and development. In addition, the unexpected protein-metabolite interactions that were detected for the two mammalian START domains tested, StAR and PCTP, may represent additional small molecule

binding partners that regulate these START domains in human cells. In summary, the results presented here provide new insights into the molecular mechanisms underlying START domain function in eukaryotic cells, and reveal novel protein-metabolite interactions that may play critical roles in modulating transcription factor activity in both animals and plants.

MATERIALS AND METHODS

Plants, growth conditions, plant constructs and phenotypic assays

The wild-type *Arabidopsis thaliana* ecotype used was Columbia (Col). The *gl2* null allele was *gl2-5* (Ohashi et al. 2002). Seeds were stratified at 4°C for 3 to 5 days and grown on soil mix containing Supersoil (McLellan Co, Marysville, OH), vermiculite and perlite (McConkey Co, Sumner, WA) in 4:3:2 ratio at 23°C under continuous light. The *GL2* cDNA was sub-cloned into pBluescript via *Kpn*I and *Sal*I restriction sites to serve as a template for the deletion construct. Phusion® Site-Directed Mutagenesis Kit (Thermo Fisher Scientific, Waltham, MA) was used to generate the START domain deletion (amino acids 253-487), following the manufacturer's protocol. After sequence confirmation, the *GL2* START deletion cassette was restriction digested with *Kpn*I and *Sal*I, and ligated into the binary vector SR54 (*ProGL2:EYFP:GL2*), which is derived from pCAMBIA1300. Construction of *ProGL2:EYFP:GL2* is described in Chapter 2 (see methods). For domain swap constructs, the START domain sequences were amplified using PCR from corresponding cDNA clones using Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA). The REV sequence contained mutations in the microRNA binding site as in (Emery et al. 2003). Gene-specific primers with 15-bp extensions homologous to vector ends were used to amplify the insert followed by In-Fusion HD cloning reactions (Clontech, Mountain View, CA) using a PCR-amplified pBluescript plasmid harboring the GL2 cDNA with flanking homologous ends. For the mutant version of GL2-StAR-START, site-directed mutagenesis was performed as described below for the yeast constructs. The reconstructed GL2 cDNAs were cloned into the *Sal*I/*Kpn*I sites of binary vector SR54 carrying *ProGL2:EYFP:gl2-START*Δ. The constructs were transformed into *gl2-5* plants by floral dip transformation using *Agrobacterium* strain GV3101 (pMP90) (Clough and Bent 1998). Transgenic plants were selected with 50 mg/ml hygromycin B. Segregation patterns for EYFP expression were observed in T2 progeny of >30 independent lines. T3 homozygous lines with stable EYFP expression were confirmed by PCR and assayed for leaf trichomes, root hairs and seed coat mucilage. For mucilage analysis, seeds were stained in a 0.25% aqueous solution of ruthenium red (Sigma, St. Louis, MO) for 2 h and replaced with deionized water prior to observation. Seeds, rosettes and roots were imaged with a Leica M125 fluorescence stereo microscope, and images were captured using a Leica DFC295 digital camera

and Leica Application Suite software version 4.1. Ovules were excised from siliques, stained with 1 μg/ml DAPI (Sigma, St. Louis, MO) and imaged with Zeiss LSM 510 meta and chameleon modules with standard settings for visualization of yellow fluorescent protein and DAPI fluorophores. Oligonucleotide sequences used to make DNA constructs are listed in Appendix Table A-1.

Yeast strains and growth media

The genotype of the *Saccharomyces cerevisiae* wild-type strain was *MATα leu2-3,112:: UASGAL10-lacZ-LEU2 his3*Δ *gal4*Δ *ura3-52*. The genotype of *erg6* mutant strain was: *MATα erg6-*Δ*1 leu2-3,112::UASGAL10-lacZ-LEU2 his3*Δ *gal4*Δ *ura3-52*. These strains were constructed from a cross between YGY13 (Gao and Pinkham 2000) and CJY004 (Jackson and Kepes 1994). For protein expression, Y258 (*MATa pep4-3, his4-580, ura3-52, leu2-3,112*) cells were used. Y258 with an integrated *URA3* gene from the pRS426 vector was a negative control for ligand-binding studies (Li et al. 2010). Growth media were made according to standard protocols. DNA constructs were introduced into yeast using a standard lithium acetate transformation method (Gietz and Woods 2002).

Yeast plasmids and DNA constructs

For each GSV construct, the START domain sequence was amplified using PCR from the corresponding cDNA clone or an *Arabidopsis* cDNA library (Grebe et al. 2000) and ligated in-frame into the *Kpn*I/*Sac*I restriction sites of pGEV-HIS3 (Louvion et al. 1993; Gao and Pinkham 2000). cDNA clones of the following genes were obtained from the RIKEN *Arabidopsis* full-length clones collection (Seki et al. 2002): HDG2 (At1g05230), ANL2 (At4g00730), HDG1 (At3g61150), HDG12 (At1g17920), HDG11 (At1g73360), HDG8 (At3g03260), At5g07260, PHB (At2g34710), CNA (At1g52150), ATHB-8 (At4g32880), At5g35180, At5g45560, At2g28320, At1g64720, At5g54170, At4g14500 and At3g13062. When cDNA inserts had internal *Kpn*I and/or *Sac*I sites, site-directed mutagenesis and primer extension PCR modified from (Scott et al. 2002) were used to change the nucleotide sequences without altering the amino acid. To construct GS translational fusions from GSV plasmids, VP16-AD was removed by restriction digestion with *Sac*I, followed by ligation. The empty vector control

was pRS314. The GV control construct lacking the START domain was made using site-directed mutagenesis and primer extension PCR of the GSV-StAR plasmid to insert *Nru*I sites at the junctions of its START domain followed by restriction digestion with *Nru*I, and re-circularization. To construct the control vector for the *SUT1* over-expression experiments, *pSUT1* (pNF1) (Ness et al. 2001) was restriction digested with *Not*I to remove the *SUT1* gene, followed by ligation. To construct translational C-terminal fusions to the yEGFP3 (Cormack et al. 1997), the coding regions for the GSV transcription factors were cloned into the pUG35 plasmid using *Bam*HI/*Eco*RI restriction sites. Site-directed mutagenesis followed by primer extension PCR was used to insert an ATG codon in the polylinker of pUG35 5' to the *Bam*HI site. For protein expression in conjunction with the identification of *in vivo* ligands, the coding regions for GSV transcription factors containing START domains from StAR, $StAR^{D182L}$, PCTP, GL2, PDF2 and At3g13062, as well as the GV control were cloned into pENTRTM-D-TOPO® (Life Technologies, Carlsbad, CA) and then moved into pBG1805 (Gelperin et al. 2005) by Gateway® LR clonase II (Life Technologies, Carlsbad, CA). Oligonucleotides used for plasmid construction, site-directed mutagenesis and sequence confirmation are listed in Appendix Table A-1.

Quantitative β-galactosidase assay

Yeast cultures were grown to exponential phase in 24-well plates (2 ml minimal media per well) at 30°C and 300 rpm using a thermoshaker. OD_{600} measurements were taken prior to centrifugation. Cell pellets were washed and suspended in 100 µl Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4.7H_2O$) followed by two freeze-thaw cycles, and addition of 700 µl Z-buffer containing 50 µM β-mercaptoethanol. Then 160 µl ortho-Nitrophenyl-β-galactoside (4 mg/l) was added to each sample sequentially at 15-s intervals, followed by incubation in a 1,200 rpm thermoshaker at 30°C. Samples were monitored for the development of a yellow color, after which the reaction was stopped by addition of 400 µl of 1 M $NaCO_3$. A_{415} measurements of the supernatants were taken and β-galactosidase activity was calculated in Miller units (Scott et al. 2002).

Site-directed mutagenesis of the START domain of StAR

START domains from ATML1, PDF2, GL2, MLN64/STARD3, StAR/STARD1 and PCTP/STARD2 were aligned using ClustalW (Chenna et al. 2003) and Jalview (Clamp et al. 2004). Site-directed mutagenesis of mStAR-GSV was performed using a one-step PCR-based method (Scott et al. 2002). Reaction mixtures contained 25 ng template DNA, 135 ng of each oligonucleotide, 1 µl of 10 mM deoxynucleotide triphosphates, 1 µl *PfuUltra*™ High-Fidelity DNA Polymerase (Agilent Technologies, Santa Clara, CA), 5 µl buffer supplied with the polymerase, and H_2O in 50 µl. PCR conditions were: 95°C denaturation for 1 min; 16 cycles of 95°C for 50 s, 60°C for 50 s, and 68°C for 1 min/kb; extension at 68°C for 7 min. PCR products were digested with 30 U of *Dpn*I for 3 h at 37°C, followed by ethanol precipitation and transformation of XL10-Gold® Ultracompetent Cells (Agilent Technologies, Santa Clara, CA).

Flow cytometry

GFP levels in live yeast cells were quantified by flow cytometry as in (Niedenthal et al. 1996). Yeast cells transformed with GSV:yEGFP3 constructs were grown to exponential phase (OD_{600} of ~0.500) in selection media containing low-flow fluorescence yeast nitrogen base without riboflavin and folic acid (Sheff and Thorn 2004). GFP positive and negative controls were pUG35 and pNF-1, respectively. For each sample, 2×10^6 cells were washed in 0.5 ml PBS, resuspended in 0.1 ml PBS for sonication, and another 0.9 ml was added prior to sample processing. Flow cytometry was performed using a BD Biosciences FACSAria Flow Cytometer Cell Sorter. Illumination was with a 200 mW 488 nm argon laser. Emission was detected through a 530/30 nm filter (FL1-H filter). 500,000 particles (yeast cells) were gated per sample.

Identification of *in vivo* ligands: affinity purification and metabolite extraction

Y258 *ura3-52* yeast cultures expressing pBG1805 constructs were grown in raffinose until they reached an OD_{600} of 0.5 to 1.0. The Y258 strain with an integrated *URA3* selectable marker that was not transformed with the tagged construct served as a minus protein control for non-specific binding. Galactose was added to induce gene expression for 6 to 8 h at 30°C, and cell pellets were collected from 300 ml cultures. Protein expression was monitored after purification from lysates with IgG Sepharose 6 (GE Healthcare, Little Chalfont, United

Kingdom) using SDS-PAGE. For the identification of protein-bound ligands, proteins were isolated from lysates using rabbit IgG-crosslinked M-270 epoxy Dynabeads (Life Technologies, Carlsbad, CA) as described (Li et al. 2010) with some modifications. After 2 hr incubation at 4°C, the beads were washed in 500 mM NH_4Ac followed by 50 mM NH_4Ac for 5 min each. Using an improved metabolite extraction method (MB and MPS, unpublished data), Dynabeads were briefly (1 min) incubated in 15 µl 95% methanol/5% H_2O with acetic acid added to give a final pH of 4.5, then neutralized by 45 µl of 95% methanol/5% H_2O with ammonium hydroxide and incubated for 8 min at room temperature. The extract was aspirated, stored on ice, and 30 µl fresh 100% methanol was added to the beads with incubation for 10 min at room temperature. Extracts were combined prior to mass spectrometry. To examine protein levels, the beads were boiled in 30 µl 2× SDS sample buffer for 10 min and 15 µl of the supernatant was analyzed by SDS-PAGE.

Mass spectrometry and data analysis

A Thermo LTQ Orbitrap Velos mass spectrometer equipped with a heated electrospray or atmospheric-pressure chemical ionization source was used for ionization and detection of the samples following separation using ultra performance liquid chromatography. A phenyl column was tested with cholesterol first and a gradient method was developed to screen in both positive and negative ionization modes using heated electrospray ionization with optimized parameters in a methanol/water 0.1% formic acid mobile phase system. Extracts were additionally screened using the same chromatography method with an atmospheric-pressure chemical ionization source and optimized parameters. For data analysis, raw files were converted, peaks centroided and parameters optimized for peak identification in MZmine. The data, collected from three technical replicates in each of four biological replicates (12 data points per peak), were median normalized. A two-tailed *t*-test was performed relative to the Y258 *URA3* control to filter out non-essential features and noise. The normalized data representing each sample and GV control were aligned by mass and retention time using the SIMA algorithm (Voss et al. 2011), and peak intensities were adjusted by normalizing for different protein levels . Any peak not present in at least 80% of all replicates and not enriched to greater than fourfold relative to the GV control was excluded from further analysis. Both HMDB and METLIN databases were queried with the remaining M/Z values to identify metabolites within a 2-ppm mass tolerance. Unknown features

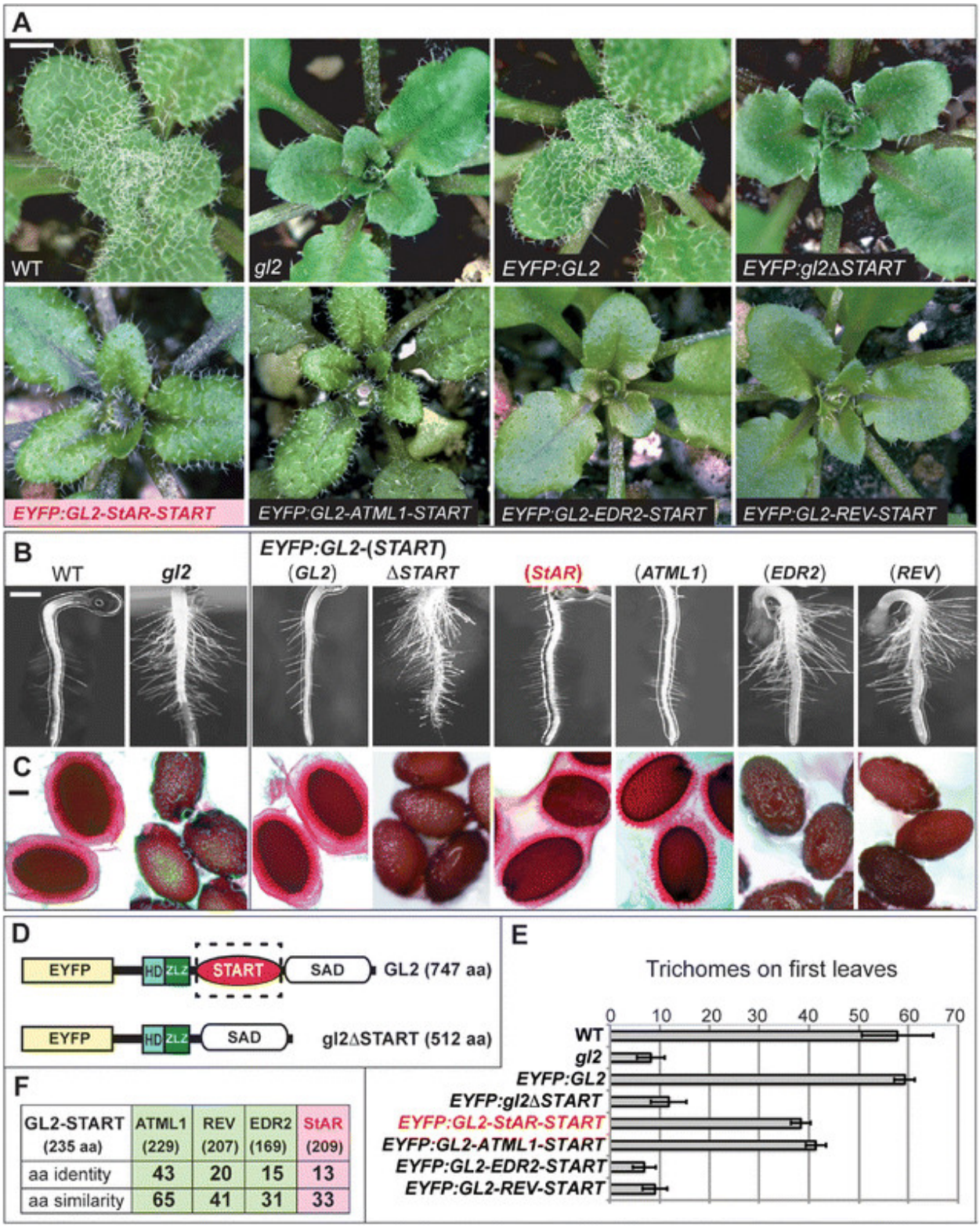
were removed. Features that yielded database hits were compared to the GV control to calculate fold-change enrichment and to facilitate further filtering. Analysis of known chemical standards permitted validation of metabolites identified initially by accurate mass, and of the 27 tested, all 27 matched. Of the 109 features that were significantly enriched, 65 had a high confidence metabolite assignment based upon querying the determined mass against the HMDB and METLIN databases. Hierarchical clustering was performed with Cluster 3.0 (de Hoon et al. 2004) and visualized with Treeview (Page 1996) . Cytoscape (Cline et al. 2007) was employed to generate protein-metabolite networks using the edge-weighted force-directed layout based on interaction weight. The high confidence and standard matched protein-metabolite interactions were assigned the identifier EBI-9687416 and deposited to the IntAct database [PMID:22121220] within the International Molecular Exchange (IMEx) Consortium. The IntAct dataset represents a subset of the total interaction data with corresponding unique PubChem identifiers, as liquid chromatography mass spectrometry was unable to distinguish certain molecules, e.g. PC(20:4(5Z,8Z,11Z,14Z)/21:0) versus PC(21:0/20:4(5Z,8Z,11Z,14Z)).

Three-dimensional structural data analysis

The closest structural homolog for each *Arabidopsis* START domain was identified by fold recognition or threading based predictions using PSIPRED (Jones 1999; Bryson et al. 2005) with 'Fold Recognition' as the prediction method (Jones 1999; McGuffin and Jones 2003). Three-dimensional coordinates for available crystal structures of START domains, either unbound (mouse STARD4 (Protein Database (PDB) 1JSS) and mouse MLN64 (PDB 1EM2)) or ligand-bound (human PCTP (PDB 1LN1)) were obtained from Brookhaven Protein Structure Database. To construct atomic-resolution models of *Arabidopsis* START domains, the closest structural homologs identified by threading were used as templates for model building in alignments using MODELLER 9v2 (Sali and Blundell 1993). Structure validation for homology models was carried out with PROCHECK (Laskowski et al. 1993) and WHAT IF (Vriend 1990). For the models that passed structure validation tests, tunnel architectures were examined using VOIDOO (Kleywegt and Jones 1994) with a probe radius of 1.4 Å. The remaining cavities were rendered by MAPMAN (Bryson et al. 2005) and visualized in PyMOL Molecular Graphics System V.1.3 (Schrödinger, LLC, New York, NY) on an Intel Dual Core iMAC installed with OS X version 10.6.5 (2.4 GHz, 4 GB RAM). VOIDOO identified a large number of cavities, of

which ambiguous and false positives were filtered out by manual visualization of cavities in the context of predicted cavity lining residues.

FIGURES AND TABLES



GL2-START (235 aa)	ATML1 (229)	REV (207)	EDR2 (169)	StAR (209)
aa identity	43	20	15	13
aa similarity	65	41	31	33

Figure 2.1 Function of the START domain in HD-Zip transcription factor GL2 from *Arabidopsis.*

(A) Rosettes exhibiting leaf trichomes in wild-type (WT), and reduction of trichomes in the *gl2* null mutant. Scale bar: 2 mm. **(B)** Roots were germinated on 0.8% agar medium and imaged after 3 to 5 days. *gl2* mutant exhibits excessive root hairs in comparison to WT. Scale bar: 200 μm. **(C)** Mucilage of WT seeds was stained with ruthenium red after imbibition. *gl2* mutants lack mucilage layer. Scale bar: 20 μm. (**A, B, C**) *EYFP:GL2* displays complete rescue of the mutant phenotype while *EYFP:GL2-StAR-START* (pink) and *EYFP:GL2-ATML1-START* exhibit partial rescue. The phenotypes of *EYFP:gl2ΔSTART*, *EYFP:GL2-EDR2-START* and *EYFP:GL2-REV-START* are indistinguishable from *gl2*. **(D)** Schematic diagrams of the EYFP:GL2 and EYFP:gl2ΔSTART proteins depicting enhanced yellow fluorescence protein (EYFP), homeodomain (HD), Zip-Loop-Zipper (ZLZ) and START-adjacent domain (SAD). For specific constructs, the START domain was deleted or replaced with a heterologous START domain. **(E)** Trichomes were counted on the first leaves of 1-week-old plants grown on soil. Error bars indicate standard deviations for n>=20. **(F)** Amino acid similarities and identities of *Arabidopsis* ATML1, REV and EDR2 START domains (green) and mouse StAR START domain (pink) in comparison to the *Arabidopsis* GL2 START domain. Amino acid sizes of corresponding START domains are indicated in parentheses. aa, amino acid; ATML1, *Arabidopsis thaliana* Meristem Layer 1; EDR2, Enhanced Disease Resistance 2; EYFP, enhanced yellow fluorescent protein; GL2, Glabra2; HD, homeodomain; REV, Revoluta; SAD, START adjacent domain; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; WT, wild type.

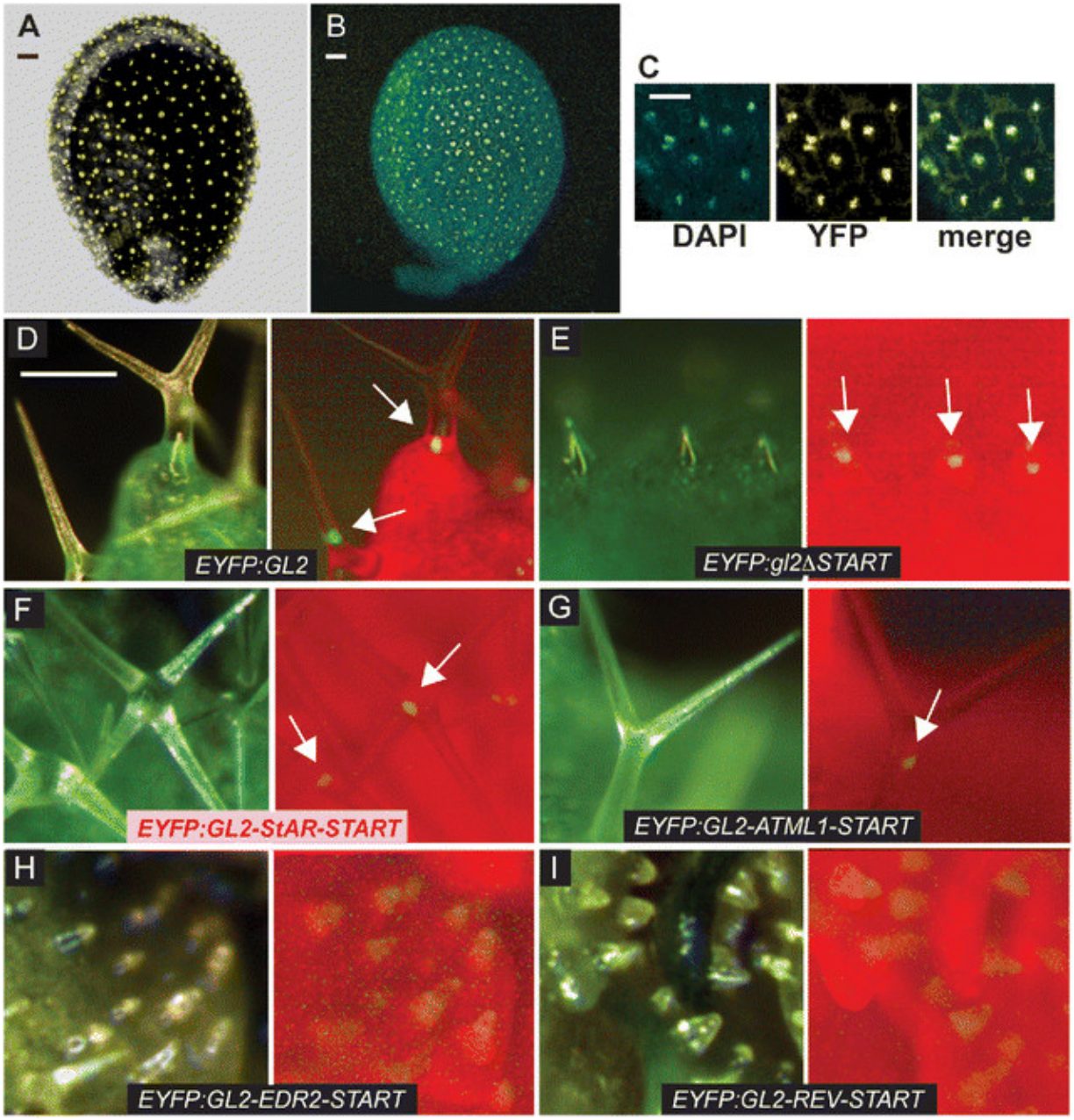


Figure 2.2 Nuclear localization of HD-Zip transcription factor GL2.

Confocal laser scanning images show *Arabidopsis* ovules expressing **(A)** wild-type EYFP:GL2 and **(B)** EYFP:gl2ΔSTART proteins, indicating nuclear localization in immature mucilage secretory cells. **(C)** Nuclear stain DAPI (blue), EYFP (yellow) and merge (green) display nuclear localization of EYFP:gl2ΔSTART from **(B).** (**A, B, C**) Scale bar: 20 μm. Live imaging of leaf trichomes indicates nuclear localization of **(D)** EYFP:GL2 and **(E)** EYFP:gl2ΔSTART, as well as EYFP:GL2-START domain swap proteins containing the **(F)** mouse StAR (pink) or **(G)** ATML1 START domains. Domain swaps with **(H)** EDR2 or **(I)** REV

START domains result in diffuse expression in trichomes. Light (left; green, chlorophyll) and matching fluorescence (right; red, chlorophyll) images are shown for each construct. Scale bar: 100 μm. ATML1, *Arabidopsis thaliana* Meristem Layer 1; DAPI, 4',6-diamidino-2-phenylindole; EDR2, Enhanced Disease Resistance 2; EYFP, enhanced yellow fluorescent protein; GL2, Glabra2; REV, Revoluta; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; YFP, yellow fluorescent protein.

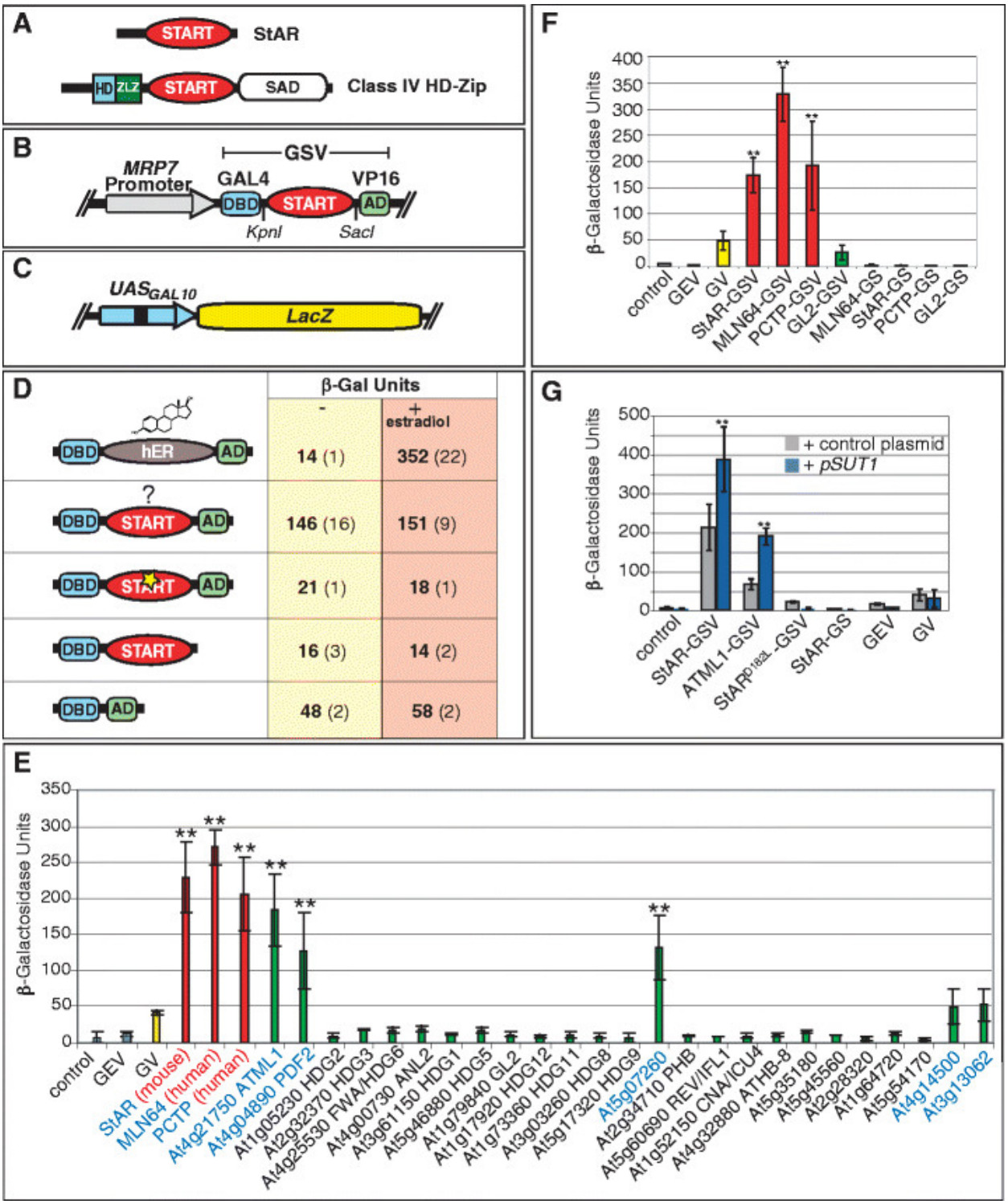


Figure 2.3 The START domain stimulates transcription factor activity in yeast.

(A) Mammalian StAR and a plant-derived HD-Zip transcription factor. **(B)** GSV synthetic transcription factor. **(C)** GAL4-DBD binding to UAS from GAL10 drives a *LacZ* reporter. **(D)** The START domain stimulates activity of the GV transcription factor. A

representative experiment is shown for the START domain derived from mouse StAR. When hER is placed between GAL4-DBD and VP16-AD (GEV), treatment with estradiol (10 μg/ml) leads to increased *LacZ* activity. START domain activity is unchanged by estradiol treatment. START domain missense mutation ($StAR^{D182L}$) (yellow star) results in loss of activity. Removal of VP16-AD similarly results in loss of activity, and GV that lacks the START domain results in a low level of β-galactosidase activity. **(E)** All three mammalian START domains (red) and at least three *Arabidopsis* START domains (green) exhibit significant activity levels in comparison to one or more controls: GV (yellow), GEV (grey) and the vector control (pRS314) (grey). **(F)** The START domain is not a transcriptional AD. Removal of VP16-AD resulted in GS constructs that lack activity. **(G)** Over-expression of *SUT1* results in elevated activity. START domains from mouse StAR or *Arabidopsis* ATML1 exhibit an increase in GSV activity when co-expressed with the *pSUT1* (blue). Asterisks indicate a significant difference over control plasmid (two-tailed *t*-test, $P<0.0005$). Other constructs that do not contain a functional START domain, i.e. $StAR^{D182L}$, GEV and GV, do not display increased activity. In all graphs, error bars indicate standard deviations for two independent transformants in four trials. AD, activation domain; ATML1, *Arabidopsis thaliana* Meristem Layer 1; DBD, DNA binding domain; GEV, Gal4 DNA binding domain:estrogen receptor: VP16 activation domain; GSV, Gal4 DNA binding domain: START domain: VP16 activation domain; GV, GAL4 DNA binding domain: VP16 activation domain; HD, homeodomain; hER, human estrogen receptor; *LacZ*, gene for β-galactosidase; MRP7, Mitochondrial Ribosomal Protein 7; SAD, START adjacent domain; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; SUT1, Sterol Uptake 1; Zip, leucine zipper; ZLZ, Zip-Loop-Zipper (a plant-specific leucine zipper); β -Gal, β-galactosidase.

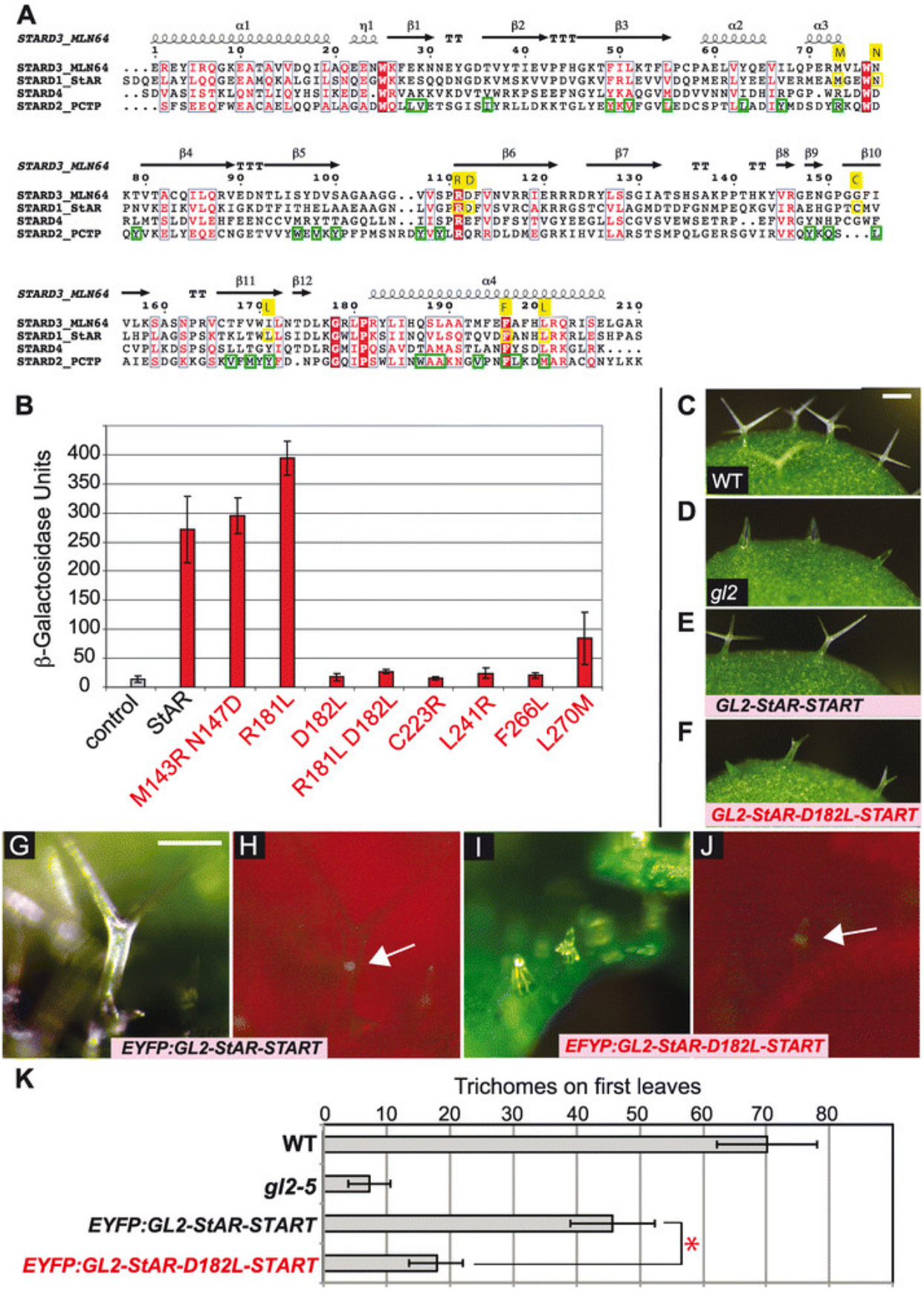

A
STARD3_MLN64
STARD3_MLN64 ...EREYIRQGKEATAVVDQILAQEENMKFEKNNEYGDTVYTIEVPFHGKTFILKTFLPCPAELVYQEVILQPERMVLMN
STARD1_StAR SDQELAYLQQGEEAMQKALGILSNQEGWKESQQDNGDKVMSKVVPDVGKVFRLEVVVDQPMERLYEELVERMEAMGEWN
STARD4 ..SDVASISTKLQNTLIQYHSIKEDE.WRVAKKVKDVTVWRKPSEEFNGYLYKAQGVMDDVVNNVIDHIRPGP.WRLDMD
STARD2_PCTP ....SFSEEQFWEACAELQQPALAGADWQLLVETSGISTYRLLDKKTGLYEYKVFGVIEDCSPTLIADIYMDSDYRKQMD
STARD3_MLN64 KTVTACQILQRVEDNTLISYDVSAGAAGG..VVSPRDFVNVRRIERRRDRYLSSGIATSHSAKPPTHKYVRGENGPGGFI
STARD1_StAR PNVKEIKVLQKIGKDTFITHELAAEAAGN..LVGPRDFVSVRCAKRRGSTCVLAGMDTDFGNMPEQKGVIRAEHGPTCMV
STARD4 RLMTSLDVLEHFEENCCVMRYTTAGQLLN..IISPREFVDFSYTVGYEEGLLSCGVSVEWSETRP..EFVRGYNHPCGWF
STARD2_PCTP QYVKELYEQECNGETVVYWEMKWPFPMSNRDYVYLRQRRDLDMEGRKIHVILARSTSMPQLGERSGVIRVKQYKQSSIT
STARD3_MLN64 VLKSASNPRVCTFVWILNTDLKGRLPRYLIHQSLAATMFEFAFHLRQRISELGAR
STARD1_StAR LHPLAGSPSKTKLTWLLSIDLKGWLPKSIINQVLSQTQVDFANHLRKRLESHPAS
STARD4 CVPLKDSPSQSLLTGYIQTDLRGMIPQSAVDTAMASTLANFYSDLRKGLRK....
STARD2_PCTP AIESDGKKGSKVFMYYFD.NPGGQIPSWLINWAAKNGVPNFLKDMARACQNYLKK
B
β-Galactosidase Units
0
50
100
150
200
250
300
350
400
control
StAR
M143R N147D
R181L
D182L
R181L D182L
C223R
L241R
F266L
L270M
C
WT
D
gl2
E
GL2-StAR-START
F
GL2-StAR-D182L-START
G
EYFP:GL2-StAR-START
H
I
EFYP:GL2-StAR-D182L-START
J
K
Trichomes on first leaves
0
10
20
30
40
50
60
70
80
WT
gl2-5
EYFP:GL2-StAR-START
EYFP:GL2-StAR-D182L-START
*

Figure 2.4 Mutational analysis of amino acids required for StAR START domain activity in yeast assay.

(A) Structural alignment of START domains from MLN64, StAR, STARD4 and PCTP with MLN64 as a reference. Consensus α-helices and β-sheets are displayed with ESPript/ENDscript. Amino acids targeted for site-directed mutagenesis of StAR START are indicated in yellow. Ligand contact points (green), as derived from the PCTP-PtCho co-crystal (Roderick et al. 2002), are shown for PCTP. **(B)** Site-directed mutagenesis of the START domain from mammalian StAR. Activities are indicated for GSV constructs that contain missense mutations in the START domain. The M143R;N147D and R181L mutants display wild-type and slightly elevated activities, respectively. L270M renders partial activity, while the other five mutants exhibit low or no activity. Un-induced GEV served as the control. Error bars indicate standard deviations for two independent transformants in four trials. **(C-F)** Trichomes are shown from second leaves of seedlings. Scale bar: 100 μm. **(C)** Wild-type (WT), (D) *gl2* mutant and *gl2* lines transformed with **(E)** *ProGL2:EYFP:GL2-StAR-START* and **(F)** *ProGL2:EYFP:GL2-StAR-D182L-START*, in which the START domain of StAR contains the D182L missense mutation. Live imaging of leaf trichomes indicates nuclear localization of **(H)** EYFP:GL2-StAR-START and (**J)** EYFP:GL2-StAR-D182L-START (white arrows). **(G, I)** Light (green, chlorophyll) and **(H, J)** matching fluorescence (red, chlorophyll). Scale bar: 100 μm. **(K**) Quantification of trichomes on first leaves. Error bars indicate standard deviations for n>=20. The asterisk marks a significant difference between EYFP:GL2-StAR-START and EYFP:GL2-StAR-D182L-START (two-tailed *t*-test, P<0.00001). EYFP, enhanced yellow fluorescent protein; GEV, Gal4 DNA binding domain:estrogen receptor:VP16 activation domain; GL2, Glabra2; MLN64, Metastatic Lymph Node 64; PCTP, phosphatidylcholine transfer protein; PtCho, phosphatidylcholine; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; WT, wild type.

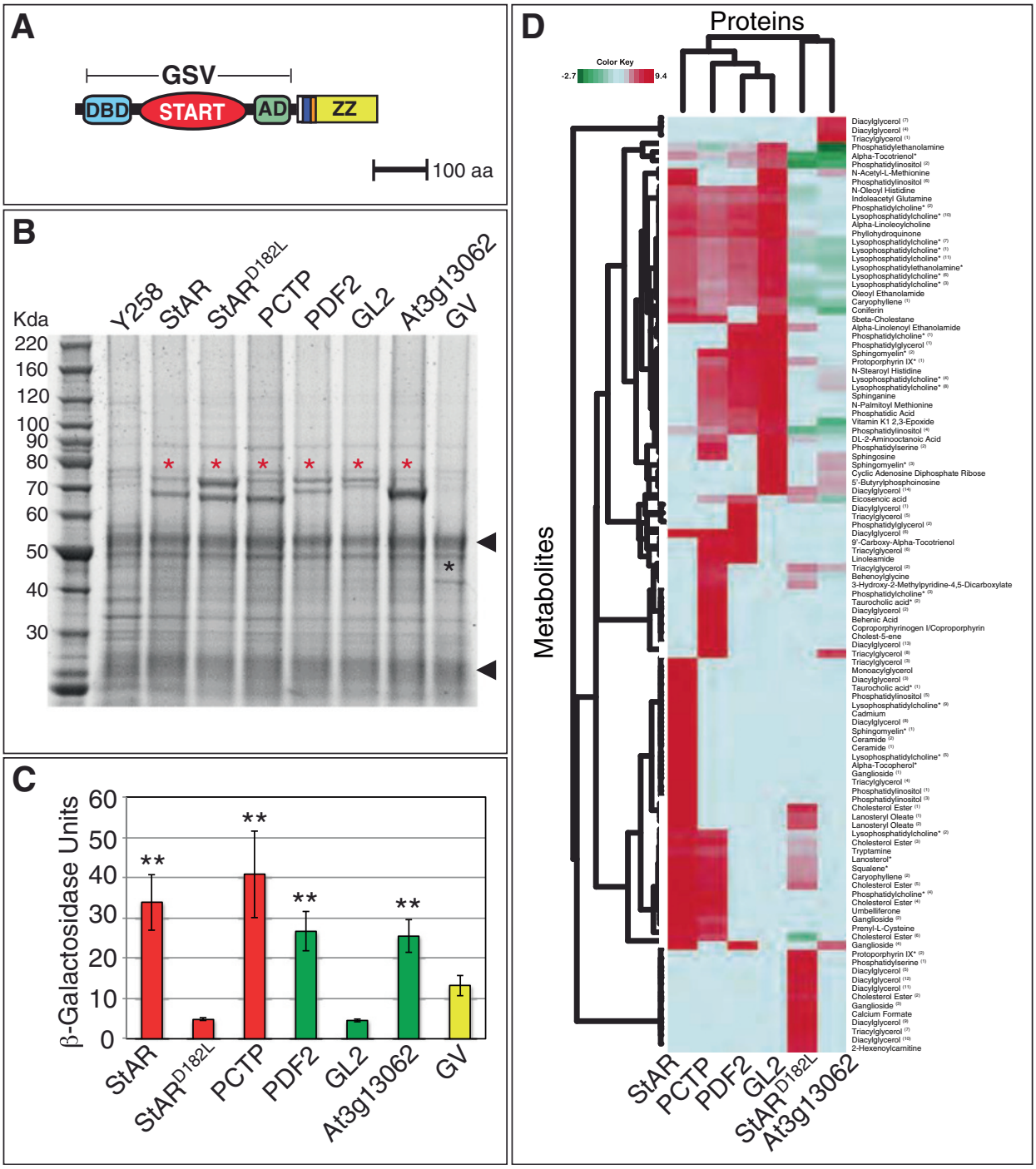


Figure 2.5 Expression and activity of tagged GSV transcription factors and dendrogram from protein-metabolite immunoisolation.

(A) GSV translational fusion to a triple tag (6x-HIS (white), HA-epitope (dark blue), Protease 3C cleavage site (orange), and IgG binding domain ZZ (yellow)) in pBG1805 [33]. **(B)** Immunoisolation of tagged proteins with IgG Sepharose beads followed by SDS-PAGE. One of four biological replicates is shown. Y258 URA3+ (lane 1) served as the negative control, while

GSV proteins are indicated by red asterisks. The GV fusion lacking the START domain is marked by a black asterisk. Arrowheads indicate IgG heavy and light chains. **(C)** Tagged GSV constructs were expressed in *erg6* cells containing the *LacZ* reporter and transcriptional activity was measured in β-galactosidase units. Double asterisks indicate a significant increase in activity over the GV control (two-tailed *t*-test, $P<0.03$). Error bars indicate standard deviations for two independent transformants in four trials. **(D)** Dendrogram from protein-metabolite immunoisolation. Two-way hierarchical clustering was performed on the metabolite enrichment data relative to the GV control with values expressed as the logarithm base 2 (ratio) of normalized metabolite intensities. Horizontal clusters represent START domain clusters based on protein-bound metabolite enrichment. Vertical clusters show groups of protein-bound metabolites clustered by START domain. Metabolite names marked by asterisks (Alpha-tocotrienol, phosphatidylcholine, lysophosphatidylcholine, lysophosphatidylethanolamine, sphinomyelin, taurocholic acid, alpha-tocopherol, and protoporphyrin IX) were validated by mass spectrometry, matching exact mass and retention time to a known standard analyzed with the same experimental conditions. Due to the variety of e.g. CE or PtCho types, metabolites of the same chemical subclass were grouped and given a superscript number to distinguish subclass members that bear distinct masses and retention times due to differing chain lengths.
aa, amino acids; AD, activation domain; DBD, DNA binding domain; GL2, Glabra2; GSV, Gal4 DNA binding domain:START domain:VP16 activation domain; GV, GAL4 DNA binding domain:VP16 activation domain; IgG, immunoglobulin G; PCTP, phosphatidylcholine transfer protein; PDF2, Protodermal Factor 2; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer.

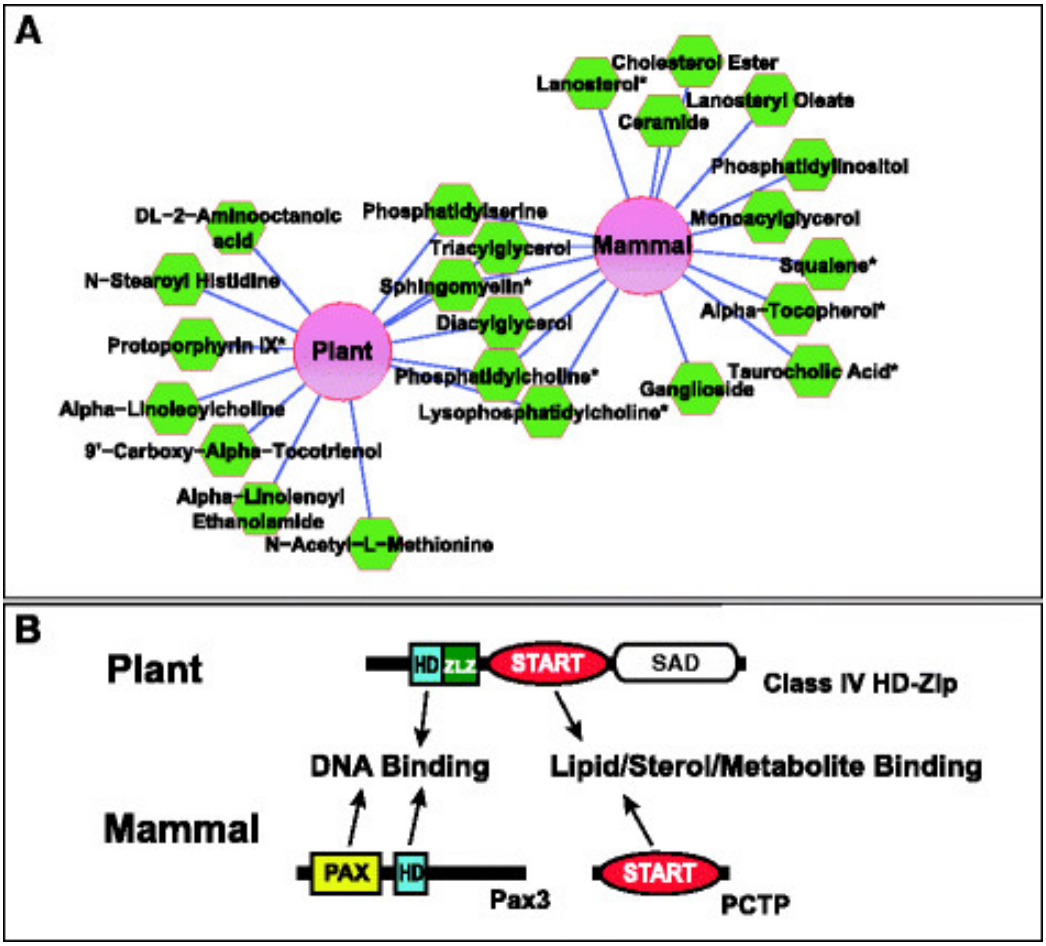


Figure 2.6 Protein metabolite interaction network of mammalian and *Arabidopsis* START domains.

(A) Cytoscape was used to visualize the normalized protein-metabolite enrichment data in an edge-weighted interaction network where metabolite interactions held in common by the *Arabidopsis* START domains were grouped to produce a single node labeled "Plant", and common interactions for mouse and human proteins (StAR, StAR^{D182L*} and PCTP) were grouped to produce a "Mammal" node. Distances between protein and metabolite nodes reflect the interaction strengths based on the magnitude of the fold-change; the shorter the edge the more enriched the metabolite. The network was filtered for interactions with a greater than tenfold change in enrichment relative to the GV control and only high confidence metabolite assignments were included. If a node had multiple interactions with the same chemical sub-class of metabolite (e.g. PtCho) these interactions were combined and weighted to give one interaction. Asterisks designate metabolites that were validated by mass spectrometry, matching exact mass and retention time to a known standard. **(B)** Schematic illustrating how START

domains may modulate transcription factor activity in plants and mammals. In plants, the START domain is found in HD-Zip transcription factors that contain an HD DNA binding domain. In mammals, the START domain and the DNA binding domains are in two separate proteins. A physical interaction between the START protein PCTP and Pax3 transcription factor comprising both paired box (PAX) and HD DNA binding domains was been reported (Kanno et al. 2007). GV, GAL4 DNA binding domain: VP16 activation domain; HD, homeodomain; PCTP, phosphatidylcholine transfer protein; PtCho, phosphatidylcholine; SAD, START adjacent domain; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; Zip, leucine zipper; ZLZ, Zip-Loop-Zipper (a plant-specific leucine zipper).

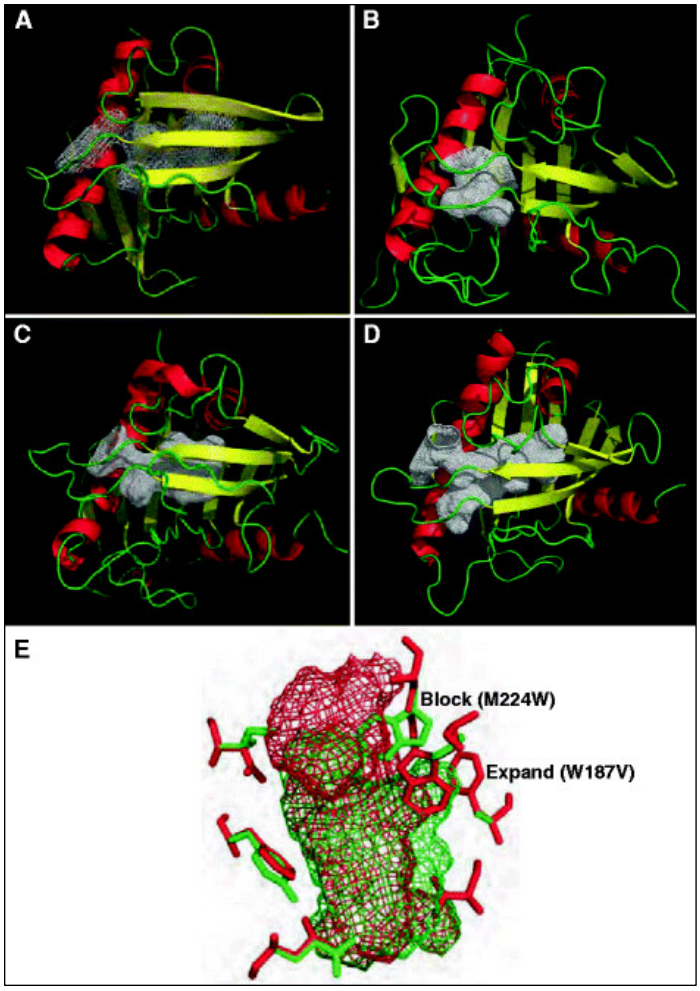


Figure 2.7 Comparative structural modeling of plant START domains.

Ribbon diagrams of three-dimensional models comprised β-sheets (yellow), α-helices (red) and loops (green), showing three categories of cavity volume. The hydrophobic tunnels are represented as a white mesh. **(A)** Crystal structure of the mammalian START domain from STARD4. **(B, C, D)** Examples of small, medium and large cavity volumes, respectively, from modeled *Arabidopsis* START domains. **(E)** Superimposition of cavities from a mammalian domain (red mesh) and a model of an *Arabidopsis* domain (green mesh) to illustrate the amino acid residues that are predicted to change the dimensions of the cavity by blockage or expansion in several plant START domains. A Trp residue (green, marked "block") in the plant structure corresponding to M224 (StAR numbering) results in shortening of the cavity length. In contrast, the mammalian cavity is restricted by W187 (red, marked "expand"), which corresponds to a smaller Val residue in the plant START domain, resulting in an increase in width at the center of the cavity.

REFERENCES

Abe M, Katsumata H, Komeda Y, Takahashi T. 2003. Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. *Development* **130**: 635-643.

Alpy F, Legueux F, Bianchetti L, Tomasetto C. 2009. [START domain-containing proteins: a review of their role in lipid transport and exchange]. *Med Sci (Paris)* **25**: 181-191.

Alpy F, Tomasetto C. 2005. Give lipids a START: the StAR-related lipid transfer (START) domain in mammals. *J Cell Sci* **118**: 2791-2801.

Baker BY, Epand RF, Epand RM, Miller WL. 2007. Cholesterol Binding Does Not Predict Activity of the Steroidogenic Acute Regulatory Protein, StAR. *Journal of Biological Chemistry* **282**: 10223-10232.

Bhangoo A, Anhalt H, Ten S, King SR. 2006. Phenotypic variations in lipoid congenital adrenal hyperplasia. *Pediatr Endocrinol Rev* **3**: 258-271.

Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT. 2005. Protein structure prediction servers at University College London. *Nucleic Acids Res* **33**: W36-38.

Cater NB, Denke MA. 2001. Behenic acid is a cholesterol-raising saturated fatty acid in humans. *Am J Clin Nutr* **73**: 41-44.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**: 3497-3500.

Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. *Bioinformatics* **20**: 426-427.

Clark BJ. 2012. The mammalian START domain protein family in lipid transport in health and disease. *J Endocrinol* **212**: 257-275.

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B et al. 2007. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2**: 2366-2382.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**: 735-743.

Cormack BP, Bertram G, Egerton M, Gow NA, Falkow S, Brown AJ. 1997. Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in *Candida albicans*. *Microbiology* **143 (Pt 2)**: 303-311.

de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. *Bioinformatics* **20**: 1453-1454.

Di Cristina M, Sessa G, Dolan L, Linstead P, Baima S, Ruberti I, Morelli G. 1996. The *Arabidopsis* Athb-10 (GLABRA2) is an HD-Zip protein required for regulation of root hair development. *Plant J* **10**: 393-402.

Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF, Bowman JL. 2003. Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. *Curr Biol* **13**: 1768-1774.

Gao CY, Pinkham JL. 2000. Tightly regulated, beta-estradiol dose-dependent expression system for yeast. *Biotechniques* **29**: 1226-1231.

Gelperin DM, White MA, Wilkinson ML, Kon Y, Kung LA, Wise KJ, Lopez-Hoyo N, Jiang L, Piccirillo S, Yu H et al. 2005. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev* **19**: 2816-2826.

Gietz RD, Woods RA. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**: 87-96.

Grebe M, Gadea J, Steinmann T, Kientz M, Rahfeld JU, Salchert K, Koncz C, Jurgens G. 2000. A conserved domain of the *Arabidopsis* GNOM protein mediates subunit interaction and cyclophilin 5 binding. *Plant Cell* **12**: 343-356.

Iyer LM, Koonin EV, Aravind L. 2001. Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins* **43**: 134-144.

Jackson CL, Kepes F. 1994. BFR1, a multicopy suppressor of brefeldin A-induced lethality, is implicated in secretion and nuclear segregation in *Saccharomyces cerevisiae*. *Genetics* **137**: 423-437.

Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**: 195-202.

Kallen CB, Billheimer JT, Summers SA, Stayrook SE, Lewis M, Strauss JF. 1998. Steroidogenic acute regulatory protein (StAR) is a sterol transfer protein. *Journal of Biological Chemistry* **273**: 26285-26288.

Kanno K, Wu MK, Agate DS, Fanelli BJ, Wagle N, Scapa EF, Ukomadu C, Cohen DE. 2007. Interacting proteins dictate function of the minimal START domain phosphatidylcholine transfer protein/StarD2. *J Biol Chem* **282**: 30728-30736.

Kleywegt GJ, Jones TA. 1994. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr* **50**: 178-185.

Kudo N, Kumagai K, Tomishige N, Yamaji T, Wakatsuki S, Nishijima M, Hanada K, Kato R. 2008. Structural basis for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide. *Proc Natl Acad Sci U S A* **105**: 488-493.

Laskowski RA, Macarthur MW, Moss DS, Thornton JM. 1993. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr* **26**: 283-291.

Lavigne P, Najmanivich R, Lehoux JG. 2010. Mammalian StAR-related lipid transfer (START) domains with specificity for cholesterol: structural conservation and mechanism of reversible binding. *Subcell Biochem* **51**: 425-437.

Letourneau D, Lorin A, Lefebvre A, Frappier V, Gaudreault F, Najmanovich R, Lavigne P, LeHoux JG. 2012. StAR-related lipid transfer domain protein 5 binds primary bile acids. *Journal of Lipid Research* **53**: 2677-2689.

Li X, Gianoulis TA, Yip KY, Gerstein M, Snyder M. 2010. Extensive *in vivo* metabolite-protein interactions revealed by large-scale systematic analyses. *Cell* **143**: 639-650.

Louvion JF, Havaux-Copf B, Picard D. 1993. Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene* **131**: 129-134.

Lu P, Porat R, Nadeau JA, O'Neill SD. 1996. Identification of a meristem L1 layer-specific gene in *Arabidopsis* that is expressed during embryonic pattern formation and defines a new class of homeobox genes. *Plant Cell* **8**: 2155-2168.

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E. 2009. Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324**: 1064-1068.

McGuffin LJ, Jones DT. 2003. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics* **19**: 874-881.

Murcia M, Faraldo-Gomez JD, Maxfield FR, Roux B. 2006. Modeling the structure of the StART domains of MLN64 and StAR proteins in complex with cholesterol. *Journal of Lipid Research* **47**: 2614-2630.

Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T. 2006. Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in *Arabidopsis*. *Plant Physiology* **141**: 1363-1375.

Ness F, Achstetter T, Duport C, Karst F, Spagnoli R, Degryse E. 1998. Sterol uptake in *Saccharomyces cerevisiae* heme auxotrophic mutants is affected by ergosterol and oleate but not by palmitoleate or by sterol esterification. *J Bacteriol* **180**: 1913-1919.

Ness F, Bourot S, Regnacq M, Spagnoli R, Berges T, Karst F. 2001. SUT1 is a putative Zn[II]2Cys6-transcription factor whose upregulation enhances both sterol uptake and synthesis in aerobically growing *Saccharomyces cerevisiae* cells. *Eur J Biochem* **268**: 1585-1595.

Niedenthal RK, Riles L, Johnston M, Hegemann JH. 1996. Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast* **12**: 773-786.

Ohashi Y, Oka A, Ruberti I, Morelli G, Aoyama T. 2002. Entopically additive expression of GLABRA2 alters the frequency and spacing of trichome initiation. *Plant J* **29**: 359-369.

Page RD. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**: 357-358.

Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF et al. 2009. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324**: 1068-1071.

Ponting CP, Aravind L. 1999. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem Sci* **24**: 130-132.

Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE. 2005. Class III Homeodomain-Leucine Zipper Gene Family Members Have Overlapping, Antagonistic, and Distinct Roles in *Arabidopsis* Development. *Plant Cell* **17**: 61-76.

Radauer C, Lackner P, Breiteneder H. 2008. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol Biol* **8**: 286.

Rerie WG, Feldmann KA, Marks MD. 1994. The *GLABRA2* gene encodes a homeodomain protein required for normal trichome development in *Arabidopsis*. *Genes Dev* **8**: 1388-1399.

Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, Rajashankar KR, Cohen DE. 2002. Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat Struct Biol* **9**: 507-511.

Romanowski MJ, Soccio RE, Breslow JL, Burley SK. 2002. Crystal structure of the Mus musculus cholesterol-regulated START protein 4 (StarD4) containing a StAR-related lipid transfer domain. *Proc Natl Acad Sci USA* **99**: 6949-6954.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**: 779-815.

Schrick K, Nguyen D, Karlowski WM, Mayer KF. 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol* **5**: R41.

Scott SP, Teh A, Peng C, Lavin MF. 2002. One-step site-directed mutagenesis of ATM cDNA in large (20kb) plasmid constructs. *Hum Mutat* **20**: 323.

Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y et al. 2002. Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science* **296**: 141-145.

Sheff MA, Thorn KS. 2004. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* **21**: 661-670.

Talbert PB, Adler HT, Parks DW, Comai L. 1995. The REVOLUTA gene is necessary for apical meristem development and for limiting cell divisions in the leaves and stems of *Arabidopsis thaliana*. *Development* **121**: 2723-2735.

Tang D, Christiansen KM, Innes RW. 2005. Regulation of plant disease resistance, stress responses, cell death, and ethylene signaling in *Arabidopsis* by the EDR1 protein kinase. *Plant Physiology* **138**: 1018-1026.

Thorsell AG, Lee WH, Persson C, Siponen MI, Nilsson M, Busam RD, Kotenyova T, Schuler H, Lehtio L. 2011. Comparative structural analysis of lipid binding START domains. *PLoS One* **6**: e19521.

Tsujishita Y, Hurley JH. 2000. Structure and lipid transport mechanism of a StAR-related domain. *Nat Struct Biol* **7**: 408-414.

Voss B, Hanselmann M, Renard BY, Lindner MS, Kothe U, Kirchner M, Hamprecht FA. 2011. SIMA: simultaneous multiple alignment of LC/MS peak lists. *Bioinformatics* **27**: 987-993.

Vriend G. 1990. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**: 52-56, 29.

Western TL, Burn J, Tan WL, Skinner DJ, Martin-McCaffrey L, Moffatt BA, Haughn GW. 2001. Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in *Arabidopsis*. *Plant Physiology* **127**: 998-1011.

Chapter 3 - Functional characterization of the START domain in Class IV HD-Zip Transcription factor GL2

ABSTRACT

The steroidogenic acute regulatory protein (StAR) related lipid transfer (START) domain is an evolutionary conserved module that is found in plant-specific class III and IV homeodomain leucine-zipper (HD-Zip) transcription factors. Recent studies have shown that START domains from both plants and mammals are versatile ligand binding motifs that can stimulate transcription factor activity. One hypothesis is that the START domain by binding a ligand, controls transcription factor function, analogously to nuclear receptors from mammals. To further address its function in ligand binding, mutational analysis of the START domain of GL2 was performed. Using the *proGL2:EYFP:GL2* transgene as a template, missense mutants were constructed in the predicted ligand-binding pocket based on sequence conservation with known ligand binding sites in human START proteins. Several of the mutations remove charged residues in the predicted ligand-binding pocket. Included among the mutants were E375G, V391L, R392M, L480P, and E375G;R392M. While E375G and V391L mutations resulted in little or no effect on GL2 function, the other missense mutations resulted in loss-of-function despite nuclear localization, with exception of R392M, which displays cytoplasmic localization, possibly due to protein misfolding. The study also uncovered $gl2^{Amo}$, a novel semi-dominant missense allele of GL2 having a small duplication in a predicted ligand-contact region of the START domain. Transcriptional profiling of $gl2^{Amo}$ seedlings revealed that gene expression is significantly altered in comparison to wild-type and the *gl2* null mutant, suggesting aberrant DNA-binding properties for transcription factor complexes containing $gl2^{Amo}$. To further dissect the role of the START domain in mediating protein-protein interactions, experiments in both yeast and *N. benthamiana* were carried out. The results suggest that Zip-mediated GL2 homodimerization requires a functional START domain, since neither the gl2ΔSTART nor $gl2^{Amo}$ proteins supported dimerization. Furthermore, yeast two-hybrid screen identified Rhamnose synthase I (RHM1) as an *in vivo* interacting partner for both GL2 and PDF2 START domains. ChIP-seq data reveal that $gl2^{Amo}$ and gl2ΔSTART display enhanced DNA binding

activity, and indicate that the START domain is dispensable for transcription factor binding to DNA. One possibility to explain this data is that deletion or mutation of the START domain prevents DNA binding in homodimers but not in heterodimers. Enhanced DNA binding may further be explained by a model in which the START domain is required for physical interaction with transcriptional repressors. Possible candidates include DELLA proteins that are critical for the gibberellin (GA) response pathway. Collectively, these results indicate that the START plays a crucial role in controlling HD-Zip transcription factor activity through both ligand binding and protein-protein interactions with one or more key co-regulators.

Contribution

The results of protein subcellular localization using confocal microscopy, phenotypic assays for root hair density and seed coat mucilage, Y2H assay using START ligand binding mutants, *in vivo* split GFP assays in *N. benthamiana* for verification of protein-protein interactions, ChIP, and Gene Ontology analysis for microarray and ChIP-seq data are my contribution to the publication: **Aashima Khosla**, Bala K.P. Venkata, Sumedha Gunewardena, Erika M. Peters, Preston M. Stephens, Paige N. Cox, Daniel F. Stucky, and Kathrin Schrick. Functional characterization of the START domain in class IV HD-Zip transcription factor GLABRA2 (manuscript in preparation).

This chapter also includes my unpublished results on the yeast two-hybrid screen using GL2/PDF2-START to identify interacting proteins, bimolecular fluorescence complementation (BiFC) split GFP proteins *N. benthamiana* assays to validate interactions between the class IV HD-Zip class TFs and RHAMNOSE SYNTHASE 1 (RHM1), and identification of novel direct targets of GL2 using ChIP-seq.

INTRODUCTION

The plant epidermis is a multifunctional tissue that serves as an interface of sessile land plants with their environment. Beyond its role as a protective barrier against biotic and abiotic stress, the epidermal cells allow the exchange of gas, water, and nutrients with the outside world. Also, some epidermal cells give rise to specialized cell types – hair cells (trichomes) or stomatal guard cells on the aerial parts, root hairs in the root, and aleurone layer in the seed - which are critical for defense, respiration, water and nutrient uptake, and starch degradation, respectively (Glover 2000). In addition to its protective function, the epidermis plays roles in the regulation of organ growth and shoot stem cell maintenance (Savaldi-Goldstein et al. 2007; Knauer et al. 2013; Nobusawa et al. 2013). In the past decade, extensive genetic and molecular studies have uncovered myriad transcription factors involved in epidermal differentiation. In particular, members of the class IV homeodomain-leucine zipper (HD-Zip) family have been identified as master regulators of epidermal cell fate specification (Ito et al. 2003; Javelle et al. 2011; Takada and Iida 2014). Class IV HD-Zip constitutes a large family of genes, composed of 16 members in *Arabidopsis* that are involved in a number of vitally important plant development processes, including the transcriptional control of epidermal and sub-epidermal cell fate within the layer specific contexts, anthocyanin accumulation, cuticle deposition, and drought tolerance (Nakamura et al. 2006; Wu et al. 2011; Yu et al. 2013). Furthermore, evolutionary studies have identified class IV HD-Zip genes in lycophytes, mosses, and charophycean algae *Coleochaete* and *Spirogyra*, indicating their origin in algal ancestor of land plants (Timme and Delwiche 2010; Zalewski et al. 2013).

Members of the class IV HD-Zip family are characterized by the presence of a plant-specific leucine zipper with an internal loop (ZLZ), immediately downstream of DNA binding homeodomain (HD). Although HD and zipper domain can be found in transcription factors (TFs) from other eukaryotic kingdoms, their combination in a single protein is unique to plants (Ruberti et al. 1991; Schena and Davis 1992). Downstream from HD-Zip, these proteins exhibit an additional domain termed START followed by an adjacent conserved region of unknown function called SAD (START-associated domain) (Ponting and Aravind 1999; Schrick et al. 2004; Mukherjee and Burglin 2006). The proteins belonging to this family have been named HD-Zip GL2 or simply the GL2 family after its founding member, the *Arabidopsis* GLABRA2 proteins (Nakamura et al. 2006).

GLABRA2, whose expression is confined to specialized epidermal cells in the embryo (Lin and Schiefelbein 2001; Costa and Dolan 2003), regulates cell fate by suppressing root-hair formation in the root (Di Cristina et al. 1996), while promoting trichome formation in the shoot (Rerie et al. 1994), and synthesis of mucilage in imbibed seeds (Western et al. 2001). A role of GL2 in seed oil accumulation was also demonstrated by characterization of mutant *Arabidopsis* lines lacking GL2, which revealed a significant increase in seed oil content (Shen et al. 2006). More recently, evidence has been provided that, in *Arabidopsis*, GL2 negatively regulates anthocyanin biosynthesis by directly repressing the expression of some MYB-bHLH-WD40 (MBW) component genes (Wang et al. 2015). Importantly, *GL2* function is dispensable for growth, yet *gl2* mutants exhibit a distinct phenotype that is easily visualized (Schrick et al. 2014). A second member of the class IV HD-Zip family that we address in this work is PDF2, which acts redundantly with ATML1 in the proper differentiation of the epidermis in *Arabidopsis.* While *atml1* and *pdf2* single mutants show the wild-type phenotype, the *atml1-1 pdf2-1* double mutant shows severe defects in the differentiation of shoot epidermal cells and is embryo lethal (Abe et al. 2003).

The Bet v 1-like superfamily (also known as the START superfamily) is a large and ancient family of proteins present in eukaryotes and archaea (Radauer et al. 2008). It contains 11 subfamilies that are structurally related to Bet v I, and the family containing highest number of structures is the Bet v 1 family followed by the START family, named after the Steroidogenic Acute Regulatory (StAR) protein-related lipid transfer (START) domain. It is noteworthy that the members of this larger superfamily include the newly identified PYL/PYR group of abscisic acid (ABA) receptors from plants (Ma et al. 2009; Park et al. 2009), thus indicating a role of START proteins in ABA signaling. The START domain is an evolutionary conserved protein module of ~210 amino acids that binds lipids including sterols and bile acids in animals (Ponting and Aravind 1999; Letourneau et al. 2012). Although the START domains were first identified in mammals, they are more prevalent in plant transcription factors, where the majority is found in class IV HD-Zip TFs (21 out of 35).

Whereas many START-containing proteins found in the animal kingdom have been well characterized and there is ample evidence for their role in trafficking and metabolism of lipids to presumed roles in cell signaling (Alpy et al. 2005; Lavigne et al. 2010), the role of START domains in plants remained elusive until recently. It has now been demonstrated that START

domains can stimulate transcription factor activity in *Arabidopsis* HD-Zip proteins (Schrick et al. 2014). The deletion of the START domain from GL2 results in loss-of-function although the protein displays nuclear localization (Schrick et al. 2014). Interestingly, a mouse START domain can functionally replace the *Arabidopsis* START domain of GL2 *in vivo*, suggesting that a specific ligand-binding sequence and not simply the ~200 amino acid size of the START domain is important for its function (refer Chapter 2). Furthermore, immunoisolation experiments in yeast identified several lipid metabolites that bound plant START domains, suggesting that START domains from mammals and plants exhibit similar functions and that putative ligands of the plant-derived START domains may also be similar (Schrick et al. 2014). In agreement with their ability to bind lipid ligands, it has been recently shown that WHEAT KINASE START 1 has increased affinity for phosphatidic acid (PA) and several forms of phosphatidylinositol phosphates (PIPs), and *in vitro* lipid binding assays indicated binding through START domain (Gou et al. 2015). However, *in planta* ligand partners for plant START domains have not been reported.

Unlike animal homeodomain proteins, HD-Zip proteins bind DNA as dimers *in vitro*, where they bind a palindromic sequence (Sessa et al. 1997; Sessa et al. 1998). It is interesting to note that the removal of the zipper domain absolutely abolishes the binding ability of the TFs, indicating that the relative orientation of the monomers, driven by dimerization, is crucial for efficient DNA recognition (Tron et al. 2004). While the zipper domain appears to be a critical determinant of dimerization specificity, little is known about the regulation of dimerization of the class IV HD-Zip proteins (Di Cristina et al. 1996). However, the predicted structure of these proteins indicates that there are additional mechanisms of post-translational regulation. For example, the recent discovery of the interaction between ATML1 and PDF2 with DELLA proteins revealed a new layer of post-translational regulation of the class IV HD-Zip activity. Both the proteins were shown to interact directly with DELLA proteins, which blocks gibberellin signaling required for seed germination (Rombola-Caldentey et al. 2014). Thus, upon seed imbibition, increased GA levels mediate the activation of downstream L1 box target genes by releasing the two HD-Zip TFs from their inhibitory interaction with DELLA proteins. Consistent with this, physical interactions between GhHOX3, an HD-Zip TF of cotton, and GhSLR1, a functional orthologue of the *Arabidopsis* DELLA AtGAI, were reported and the domain deletion assay demonstrated that both leucine zipper (ZLZ) and START domains were indispensable for

the interaction (Shan et al. 2014). Recently, genetic analysis and GFP sensor experiments demonstrated post-transcriptional regulation of maize class IV HD-Zip transcription factor (OCL1) by RDR2 dependent small RNA (small 1), thus suggesting another level of regulation (Klein-Cosson et al. 2015). Since these motifs are predicted to form a stem-loop structure that is conserved between maize, *Arabidopsis,* and rice, it is tempting to speculate that similar translational repression mechanism exists for the class IV HD-Zip members across species.

Regulatory specificity of TFs is not given only by their interaction with conserved DNA-sequence motifs in the promoters of the genes that they control, but also by the interactions with other factors, which may or may not bind the target promoters. Several reports indicate that class IV HD-Zip TFs form heterodimers with each other, and physically interact with other types of TFs. For example, there is evidence that both the START and SAD domains of the cotton HD-Zip TF GbML1 are necessary for binding to the C-terminal domain of GbMYB25, a transcription factor known to be a key regulator of cotton fiber initiation (Zhang et al. 2010). In addition, Wu et al. (2011) showed that *Arabidopsis* CFL1 negatively regulates cuticle development by affecting the function of HDG1, a class IV homeodomain-leucine zipper transcription factor. Furthermore, ATML1 and PDF2 have been shown to heterodimerize and also physically interact with DELLA proteins as part of a gibberellin (GA) signaling pathway in the seed epidermis (Rombola-Caldentey et al. 2014). Additionally, the SWITCH complex protein 3C1 (ZmSWI3C1) from *Zea Mays* was identified and confirmed as an interacting partner of maize HD-Zip IV outer cell layer 1 (ZmOCL1) (Depege-Fargeix et al. 2011). Only one study so far has shed light on the role of START as a protein-binding domain in humans. Using yeast-two hybrid screen, Kanno et al. (2007) uncovered Pax3 homeodomain transcription factor and thioesterase as potential interactors of PCTP.

In this work, complementary approaches were undertaken to characterize the role of the START domain in ligand binding, dimerization, protein-protein interactions, and DNA binding of the class IV HD-Zip transcription factors. Mutational studies indicate that the ligand binding is critical to HD-Zip transcription factor activity. A dominant-negative allele of *GL2*, $gl2^{Amorphous}$ ($gl2^{Amo}$) that exhibits a novel phenotype was uncovered. Both the experiments in yeast and bimolecular fluorescence complementation (BiFC) assays in *N. benthamiana* demonstrate that a functional START domain is necessary for ZLZ mediated GL2 homodimerization but not for heterodimer assembly. Furthermore, using a yeast two-hybrid (Y2H) library screen, candidates

for proteins that interact with the START domain of PDF2 and GL2 were identified, suggesting its role in mediating protein-protein interactions involving HD-Zip TFs. Genome wide ChIP-seq assays revealed novel direct targets of GL2 and indicated that START domain is dispensable for DNA binding of HD-Zip transcription factor. Taken together, these data support the premise that the START domain plays a crucial role in controlling HD-Zip transcription factor activity through both ligand binding and protein-protein interactions with one or more key co-regulators.

RESULTS

Construction of missense mutations in the predicted ligand binding pocket of the GL2-START domain

A recent study from our lab showed that START domains are versatile ligand binding motifs that are used by both plants and mammals to stimulate transcription factor activity (Schrick et al. 2014). To further address its function in ligand binding, we generated missense mutants in the predicted ligand binding pocket of the START domain. For mutational analysis, we chose *Arabidopsis GL2*, since *gl2* null mutants are viable and display epidermal phenotypes that are easily visualized, including defects in leaf trichome development, excessive root hair formation, and lack of seed coat mucilage production (Schrick et al. 2014). The molecular nature of *gl2* null activity allele, *gl2-5* is described in Chapter 5 (methods). Using site-directed mutagenesis (Scott et al. 2002), and taking the *proGL2:EYFP:GL2* transgene as a template, missense mutants in the predicted ligand-binding pocket were constructed based on conservation and similarity with the known ligand binding residues in the mammalian START domain proteins: STARD1/StAR, STARD2, and STARD4 (Figure 3.1).

Arg188 residue is located in the cholesterol binding pocket of STARD1. It has been shown that R188C, a mutation that corresponds to a substitution in human patients exhibiting lipoid congential adrenal hyperplasia (LCAH) affects the charge of a highly-conserved residue within β6 of the START domain. This region also contains other residues implicated in direct ligand contact (Metherell et al. 2009; Sahakitrungruang et al. 2010). As shown in Figure 3.1, the R392M substitution in GL2 affects the charge of the residue that is both highly conserved and is a predicted ligand contact site. Furthermore, the presence of the conserved salt bridge at the bottom of the cavity between an acidic residue Glu169 and a basic residue Arg188 was proposed as a key determinant for cholesterol binding in STARD1 (Murcia et al. 2006). More precisely, the C3-hydroxyl group of cholesterol could form a specific interaction with the guanidino group of the conserved arginine (Tsujishita and Hurley 2000; Mathieu et al. 2002; Murcia et al. 2006). Consequently, the salt bridge double mutant E169M; R188M (corresponding to E365G;R392M in GL2) exhibited complete loss of its cholesterol binding capacity and lacked steroidogenic activity (Roostaee et al. 2009).

In addititon, there is evidence that carboxy-terminal α4-helix also appears to be critical for ligand binding, since helix mutations L271N (L480P in GL2) and L275P result in significant decline in cholesterol binding affinity and steroidogenic activity (Roostaee et al., 2009). Likewise, clinical and functional studies suggest that LCAH is associated with V187M, a mutation located in the cholesterol binding pocket (Baker et al. 2006; Metherell et al. 2009; Sahakitrungruang et al. 2010). Thus, an analogous mutation (V391L) in the *Arabidopsis* START domain is predicted to restrict the cavity size by replacing valine with a residue having a larger side chain.

The START domain is required for GL2 activity

We compared the ability of the full-length GL2 to complement *gl2* phenotypes with that of GL2-START domain mutants mentioned above (Figure 3.2). Whereas the wild-type *ProGL2:EYFP:GL2* construct rescued mutant phenotypes in leaves, roots and seeds (Figures 3.2 B, B' and Appendix Figure B.1 A') as previously shown (Schrick et al. 2014), lines containing the *ProGL2:EYFP:gl2*^{R392M} construct failed to complement and were indistinguishable from the *gl2* null mutant (Figures 3.2 H, H'). Therefore, the R392M substitution conferred loss-of-function. Similarly the E375G;R392M bridge mutant revealed a phenotype reminiscent of that previously observed for *gl2*Δ*START* (Figures 3.2 G, G' and Appendix Figure B.1 E'). In contrast, the L480P mutant demonstrated partial rescue of the trichome defect (Figure 3.2 F) and partial complementation of seed mucilage and root phenotypes was observed (Figure 3.2 F' and Appendix Figure B.1 D'). In contrast to the loss-of-function mutations, the mutant phenotypes were restored to nearly wild type levels in *gl2* plants transformed with the *EYFP:GL2* transgene harboring E375G and V391L mutations in the START domain (Figures 3.2 E, E', D, D' and Appendix Figures B.1 B', C'). Quantitation of trichomes on first leaves indicated that *gl2* plants expressing *ProGL2:EYFP:gl2*^{E375G} and *ProGL2:EYFP:gl2*^{V391L} have phenotypes that are indistinguishable from the wild type GL2 (Figure 3.2 J and Table 3.1). For mutant lines expressing *ProGL2:EYFP:gl2*^{R392M} and *ProGL2:EYFP:gl2*$^{E375G;R392M}$ transgenes, ~9% of total trichomes were branched, indicating the inability of these START domain mutations to confer phenotypic complementation. Furthermore, in *gl2* mutants transformed with *ProGL2:EYFP:gl2*^{L480P} transgene, an ~3 -fold increase in average trichome numbers and more

branching (~ 51%) on first leaves was observed, but overall trichome formation was partial in comparison with the wild type.

It was previously reported that the mutant GL2 protein lacking the START domain is nonfunctional although the protein correctly localizes to the nucleus (Schrick et al. 2014). To explore this further, the missense mutants were examined for their subcellular localization in the *gl2* mutant background. Most of the mutants displayed nuclear localization in ovules, roots and trichomes (Figures 3.2 K-O, K'-O', Q-U, Q'-U' and Appendix Figures B.1 A-E), indicating that the mutant protein is expressed and properly localized. An exception is the R392M mutant that exhibited cytoplasmic localization of the EYFP:GL2 protein (Figures 3.2 P, P', V, and V'). Combined, these obervations imply that a functional ligand binding START domain is essential for HD-Zip TF activity in plants.

The *gl2*^{Amo} dominant-negative allele exhibits a novel phenotype

Thus far, all of the *gl2* START domain mutants described have either no effect or result in loss-of-function phenotypes in *Arabidopsis*. An exception is a novel mutation in the START domain, which we have named *gl2*^{Amorphous} (*gl2*^{Amo}) due to the abnormal morphology of trichomes in the mutant. The molecular nature of the lesion in *gl2*^{Amo} is that it has the *gl2*$^{R386L;E387L}$ missense mutations followed by an 11-amino acid duplication in a predicted ligand-contact region (Figure 3.3 A). The *gl2*^{Amo} phenotype is more severe than the *gl2* null mutant, since fewer trichomes are found at the margins of the leaves (Figures 3.3 B-C and F-K). Closer examination of trichomes using scanning electron microscopy revealed that the trichomes are grossly malformed or collapsed with cell walls that appeared very thin and fragile (Figures 3.3 L-N). Supporting these observations, toluidine blue (TB) staining revealed cuticular defects in trichomes of the *gl2*^{Amo} trangenic plants compared with wild-type and *gl2* mutant plants (Figures 3.3 O - T). We also observed seed mucilage defects (Figures 3.3 U-W) that differed from the null mutant. The novel phenotype is dependent on two copies of the mutant transgene, and is semi-dominant in presence of the wild-type *GL2* allele (Figure 3.4). The subcellular localization of EYFP:gl2^{Amo} is nuclear (Figures 3.3 D, E, and Appendix Figure B.2).

Next, we conducted a microarray-based comparison of transgenic lines expressing EYFP-tagged wild-type GL2, gl2^{Amo}, and wild-type Columbia ecotype (Col). Three biological

replicates for each sample were collected. RNA was isolated from whole seedlings, processed and labeled using standard protocols and hybridized to custom arrays (Wellmer et al. 2006). Genome-wide expression analysis revealed mRNA expression patterns that are significantly different from both wild-type and the *gl2* null mutant, suggesting that a transcription factor complex containing $gl2^{Amo}$ has altered DNA-binding and/or activation capacity. Using a four-fold difference as a cut-off and passage of a Benjamini and Hochberg multiple testing correction with a P<= 0.01, we found 572 genes expressed only in the wild type, with 211 upregulated and 361 downregulated. Similarly, 956 genes were differentially expressed, with 380 upregulated and 576 downregulated in $gl2^{Amo}$. For the upregulated genes, the gene ontology (GO) classes that were significantly overrepresented in wild type were metabolism, response to stimulus, and transcription factor activity (Appendix Figure B.3 A). In the $gl2^{Amo}$ mutant, the functional classes were more or less the same, with an exception of a small number of genes associated with lipid transport and localization. Compared with the upregulated genes, many more genes were downregulated (Appendix Figure B.3 B). Some of the significantly enriched GO categories found only in $gl2^{Amo}$ corresponded to microtubule based movement (GO:0007018; p-value 6.6e−15), lipid transport and localization (GO:0008289; p-value 2.5e−17), and cytoskeleton protein binding (GO:0008092; p-value 2.1e−29), while genes involved in electron carrier (GO:0009055; p-value 6.2e−08), antioxidant (GO:0016209; p-value 2.6e−27), and oxidoreductase activity (GO:0016491; p-value 1.5e−17) were only enriched in the wild type. There is ample evidence of the link of cuticular lipid composition on trichome development (Gray et al. 2000; Wellesen et al. 2001; Kurdyukov et al. 2006; Wu et al. 2011). Additionally, the roles of plant cytoskeleton (microtubules and microfilaments) and cell wall systems in driving leaf trichome morphogenesis have been elucidated in several reports (Mathur et al. 1999; Basu et al. 2005; Lu et al. 2005; Djakovic et al. 2006). In this study, downregulation of several genes involved in these processes, in particular, Wall-associated kinase 2 (WAK2, AT1G21270, 6.1 fold), SCAR1 (AT2G34150, 6.7 fold), MYO5 (AT2G31900, 8 fold), kinesin-like proteins KICP-02 (AT3G50240, 4.8 fold), and ATK4 (AT5G27000, 4.6 fold) may contribute to the collapsed trichome phenotype and might explain the lack of cuticle defects observed upon toluidine staining.

Characterization of HD (homeodomain), Zip (Leucine zipper), and SAD (START associated) domains

Our previous work revealed that the deletion of the START domain from GL2 results in loss of activity (Schrick et al. 2014). To further understand the function of the START domain in HD-Zip transcription factors, it was sought to determine the roles of the other domains with respect to nuclear localization and ability to rescue *gl2* mutant phenotypes in the epidermis. Starting with the *GL2pro:EYFP:GL2* transgene as a template, site-directed mutagenesis was used to construct deletions of the class IV HD-Zip functional domains: *gl2*Δ*HD*, *gl2*Δ*Zip*, and *gl2*Δ*SAD* (Figure 3.5 A). To test functionality, transgenes for each construct were introduced into the *gl2* null mutant background. Similar to *gl2*Δ*START*, none of these domain deletion constructs were found to restore the *gl2* null mutant phenotypes (Figures 3.5 E-H, E'-H' and Appendix Figures B.1 F'-H'). The gl2ΔZip protein displays nuclear localization similar to that of gl2ΔSTART, except that it is much stronger (Figures 3.5 L, L', P, P' and Appendix Figure B.1 G). In contrast, the EYFP fusions to gl2ΔHD resulted in non-nuclear expression of the protein, indicating that HD is required for nuclear localization. The protein exhibited cytoplasmic expression with numerous bright punctate spots (Figures 3.5 K, K', O, O' and Appendix Figure B.1 H). Interestingly, *in silico* prediction (using nNLS Mapper (Kosugi et al. 2009) suggested the putative NLS, GTNKRK**RKKYHRH** (96-108), in the N-terminus of the protein. The predicted NLS has a high proportion of positively charged amino acids. The underlined residues are part of the HD sequence, consistent with requirement of the HD for nuclear targeting of GL2 (Appendix Figure B.4 A). Punctate localization was further investigated by comparing it with red fluorescent protein (Shaner et al. 2004) tagged markers (Nelson et al. 2007) for mitochondria (yeast cytochrome oxidase subunit IV), Golgi complexes (soybean [*Glycine max*] α -1,2-mannosidase I), or peroxisomes (peroxisome targeting sequence 1). EYFP-gl2ΔHD showed no overlap with the Golgi and peroxisomal markers (Appendix Figures B.4 B, C). However, a significant overlap with the mitchondrial marker was observed, suggesting that a subset of gl2ΔHD protein localizes to mitochondria (Appendix Figure B.4 D). Interestingly, no mitochondrial targeting sequence is predicted for GL2 (Small et al. 2004). No protein expression was detected for EYFP fusions to gl2ΔSAD (data not shown). This may reflect the fact that the transgenic lines used in the study did not express transgenes at high enough levels or that the

mutant protein is unstable. Taken together, results of the domain deletion study further our understanding of other functional domains within the class IV HD-Zip transcription factor GL2.

START domain plays a role in leucine zipper (ZLZ) mediated GL2 homodimerization

It was reported that GL2 can homodimerize through its ZLZ domain (Di Cristina et al. 1996). To explore this further, Y2H analysis was performed. To begin with, a representative set of HD-Zip transcription factors from *Arabidopsis* was surveyed for hetero- and homodimer formation using the ProQuest^{TM} Y2H System (Invitrogen). Six transcription factor cDNAs from the class IV (GL2, HDG1, HDG4, HDG11, HDG12, PDF2) and two from the class III (ATHB8, PHB) HD-Zip family were cloned into Bait and Prey vectors and assayed for physical interaction in yeast. Autoactivation of Bait plasmids (indicating presence of an activation domain) was confirmed for these transcription factors, with the exception of GL2 and HDG4 (Appendix Figure B.5 A). In ONPG β-galactosidase assay, GL2 interacted with itself (Appendix Figure B.5 B). GL2 homodimerization was also confirmed by growth on selection plates (Figure 3.6 B- see methods). To determine whether the START domain might play a role in regulating dimerization, GL2 lacking the START domain (gl2ΔSTART) and GL2-START domain mutants ($gl2^{L480P}$, $gl2^{E375G}$, and $gl2^{R392M}$) were tested. When the START domain was deleted, no interaction above background could be detected (Figure 3.6 B). Similarily, $gl2^{L480P}$ and $gl2^{R92M}$ mutants, previously shown to be required for GL2 activity, did not homodimerize. By contrast, $gl2^{E375G}$ mutation that resulted in little or no effect on GL2 function, strongly homodimerized in Y2H experiments (Figure 3.6 B). These observations support the idea that GL2 homodimerization requires a functional START domain.

To validate GL2 homodimerization *in planta*, Bimolecular Fluorescence Complementation (BiFC) assays were carried out in transiently transformed *Nicotiana benthamiana* leaves. In BiFC experiments, the N-terminal portion of GFP is fused to one of the putative interaction partners and the C-terminal portion of GFP is fused to the other partner. Interaction between the two partners brings the two proteins close enough together that the GFP protein is reconstituted and can fluoresce (Hu et al. 2002). GL2 was able to bring the two halves of GFP together, indicating dimerization. As shown in Figure 3.6 E, a strong fluorecent signal was observed in the nuclei of *N. benthamina* cells. It is thought that homodimerization requires

the Zip domain. To test this idea *in vivo*, gl2ΔZLZ was assayed for its ability to dimerize. Consistent with the role for ZLZ in dimerization, it was found that the ability to homodimerize was compromised in the GL2 proteins lacking the ZLZ domain (Figure 3.6 F). Also, as for the Y2H data, deletion of the START domain resulted in signficantly decreased dimerization *in vivo* (Figure 3.6 G, L). Moreover, analogous to the result in yeast, START domain mutations $gl2^{E375G;R392M}$ and $gl2^{L480P}$ reduced the ability of the mutant protein to interact with the wild-type GL2 (Figures 3.6 I, J, L). Noticeable reduction in the abiltiy of $gl2^{Amo}$ to interact with GL2 was also observed (Figure 3.6 H). Altogether, these findings further lend credence to the idea that START plus ZLZ region is required and sufficient for GL2 homodimerization. No fluorescent signal was observed when different combinations of transcription factors and /or empty vectors were coexpressed (Appendix Figure B.6), thus confirming the specific nature of the interactions. We further examined the ability of gl2ΔSTART and $gl2^{Amo}$ to self-associate. $gl2^{Amo}$ proteins only weakly associate with each other (Figure 3.6 K). Likewise, gl2ΔSTART displayed significantly reduced homodimerization (Figure 3.6 L). Based on these results, it was concluded that deletion of the START domain as well as mutations in the START domain reduces the ability of GL2 to form homodimers.

Further, the Y2H assay indicated that those constructs that include C-terminal half of GL2 (aa 253-747 or START+SAD and aa 494-747 or SAD alone) were able to associate strongly with itself (Figure 3.6 A). Proteins containing amino acids 1-231 (HD+ZLZ) homodimerized, suggesting that these residues are minimally required to mediate this interaction. However, the strength of binding is reduced, implying that flanking residues may contribute to GL2-GL2 interaction particularly those from the C-terminal amino acids 494-747 (Figure 3.6 C). By contrast, the proteins including the START domain alone (aa 253-493) or ZLZ by itself (aa 157-231) were unable to form homodimers, further reinforcing the hypothesis that both START and ZLZ are required for GL2 dimerization (Figure 3.6 C). A simple explanation for this result is that GL2 adopts a conformation such that the START domain occupies space that promotes GL2 monomers to come in contact with one another. Another possibility is that ligand binding to the START domain promotes dimer formation. The inability of putative START ligand binding mutants to interact in our Y2H and BiFC experiments supports this hypothesis. It should be noted that the GL2 protein containing the SAD alone showed substantial autoactivation when used as a bait (Appendix Figure B.7). Consistent with this observation, the result demonstrating

self association of SAD domain (aa 494-747) must be also interpreted with caution (Figure 3.6 C). To our surprise, protein encoding amino acids 1-493, which includes both ZLZ and START domain, failed to form homodimers in yeast, suggesting that the SAD domain is also required for interaction. Western blot analysis confirmed that all bait proteins were expressed and not degraded in the Y2H assays showing no positive interactions (Appendix Figure B.8). Altogether, Y2H data combined with BiFC analysis demonstrate that the START plus ZLZ region is required and sufficient for GL2 homodimerization.

START domain is not required for GL2 heterodimerization

Recent studies have shown that class IV HD-Zip transcription factors can also form heterodimers with other class IV family members. For example, ATML1 and PDF2, that play redundant roles in shoot epidermal differentiation, heterodimerize through their zipper motifs *in planta* (Rombola-Caldentey et al. 2014; San-Bento et al. 2014). Based on these results, it was hypothesized that GL2 also forms heterodimers. To test this, GL2 was examined for its ability to interact with other class IV HD-Zip family members, including HDG1, HDG4, HDG11, HDG12, and PDF2. Two class III HD-Zip family members (PHB and ATHB8) were also included for comparison. Notably, GL2 exhibited heterodimerization in BiFC assays *in planta* but not in Y2H experiments (Appendix Figure B.5). A strong fluorescent signal was observed in nuclei of *N. benthamiana* cells with co-expression of N-YFP:GL2 with C-YFP:ATML1 or C-YFP:PDF2, indicating the ability of GL2 to form heterodimers *in vivo* (Figure 3.7 A). No interaction was detected with Y2H assay, possibly due to an indirect association of ATML1/PDF2 with GL2. Another intriguing explanation for this result is that the START domain from GL2 binds a ligand that promotes heterodimerization *in planta*. Such a ligand may either be absent or present in relatively low concentrations in yeast cells. Additionally, quantification of fluorescence intensity revealed that the BiFC signal of GL2/GL2 homodimer and heterodimers was significantly higher than that observed for the negative controls (Figure 3.7 B). To unravel the role of START domain in regulating heterodimerization, BiFC assays were performed with constructs in which START domain was etiher deleted (gl2ΔSTART) or mutated ($gl2^{Amo}$). Similar to what was observed with the full length GL2, proteins encoded by both constructs interacted with ATML1, indicating that the START domain is not required for heterodimerization (Figures 3.7 C, D).

Negative control combinations gave no signal, thus confirming the specific nature of the interactions (Figure 3.7 A and Appendix Figure B.6). This finding, together with the homodimerization data, indicated a new role for the START domain in homodimer formation but not in the heterodimer assembly of GL2 .

Yeast two hybrid library screen identifies novel interactors of START domains

To further address how START domain functions, a normalized *Arabidopsis* library for was screend for interacting proteins. The Y2H library screen is a high-throughput method for identifying interacting proteins. As employed in this chapter, the Y2H was used to identify proteins that interacted with the PDF2-START and GL2-START domains. I performed PDF2-START screening, while GL2-START domain screen was conducted by Kathrin Schrick. Table 3.3 summarizes the results of extensive sampling of *Arabidopsis* Y2H screen library with two START domain bait proteins.

The expression of PDF2-START bait fusion in yeast was assesed via Western blotting using anti-Myc antibody. The bait plasmid expressed the expected fusion protein (Figure 3.8 A). Some proteins have the inherent ability to activate transcription in the Y2H system. Thus an important step in conducting a Y2H assay is to ensure that the bait fusion proteins do not auto-activate the reporter genes in the absence of an activation domain containing protein fusion partner. Cells containing the bait constructs were independently plated on -Leu-Trp-His and -Leu-Trp/Aureobasidin A/X-α-Gal and did not activate the reporters.

The yeast two-hybrid screen is prone to a variety of artifacts. Thus, one must evaluate each initial positive for these well-known artifacts. In the following section, I have outlined the results for the PDF2-START Y2H cDNA library screen and how I systematically eliminated false positives. Three reporter genes, *AUR1-C*, *MEL-1, and HIS3* were used to detect protein interactions. A*UR1-C* is expressed in Y2HGold Yeast Strain in response to protein-protein interactions that bring the GAL4 transcriptional activation (AD) and DNA binding domains (BD) into close proximity. In *Saccharomyces cerevisiae*, its expression confers strong resistance (*AbA*^{r}) to the otherwise highly toxic drug Aureobasidin A. This drug reporter is preferable to nutritional reporters alone, due to lower background activity. For example, the use of this reporter alone results in considerably less background than a histidine reporter alone. *MEL-1* encodes α -galactosidase, an enzyme occurring naturally in many yeast strains. As a result of

two-hybrid interactions, α-galactosidase (*MEL1*) is expressed and secreted by the yeast cells. Yeast colonies that express *MEL-1* turn blue in the presence of the chromagenic substrate X-α-Gal. Recovery of histidine prototrophy was used at the end of the two-hybrid screen to confirm interactions.

Screen using PDF2-START as bait resulted in 31 diploid colonies on lower stringency DDO/X/A (-Leu-Trp Aba X-α-Gal) plates. For initial identification of false positives, the blue colonies were patched onto higher stringency TDO/X/A (-Leu-Trp-His Aba X- α-Gal) plates. The higher stringency screening reduced the potential binding partners from 31 to 7. All TDO/X/A positive interactions were further analyzed to identify duplicates and to verfiy that the interactions are genuine. The prey plasmids were rescued from yeast and sequenced to determine their identity. Y2H_1, Y2H_2, Y2H_3, and Y2H_24 encoded the same protein corresponding to At2g36530, a predicted 2-phospho-D-glycerate hydroylase. The second positive was plasmid Y2H_17 that encoded a partial cDNA encoding At5g40830 (414 aa), a predicted but uncharacterized SAM (S-adenosyl-L-methionine)-dependent methyltransferase related to bacterial enzymes with substrate specificities to small molecules. The protein has a two domain structure, the N-terminal putative methyltransferase domain (residues 268-364) and a C-terminal DUF248 domain of unknown function (residues 338-414). The cDNA corresponding to At5g40830 detected in the Y2H screen encoded ~187 amino acid residues, including the first 19 residues of the putative methyltransferase domain. Thus, the sequencing results identified two unique prey plasmids. Putative interacting clone At2g36530, a predicted 2-phospho-D-glycerate hydroylase, turned out be a false positive as it alone activated the GAL4-responsive reporters in the absence of the bait (Figure 3.8 B).

Spontaneous suppressor mutations in yeast resulting in the activation of reporters are a common cause of false positive interactions. In these false positives, a mutation in the yeast has resulted in the transcription of one or more reporter genes independent of a protein interaction induced reconstitution of the GAL4 transcriptional activator. To rule out suppressor mutations as the cause of the positive growth phenotype, yeast cells were cotransformed with two unique prey plasmids and a compatible haploid strain harboring the original bait construct (pGBKT7/PDF2-START) and tested for activation of reporters genes. Recapitulation of the interactions for both prey plasmids ruled out suppressor mutations as a source of reporter gene activation.

Another source of false positive, artifactual restoration of prototrophic growth is the situation in which the prey plasmid contains a cDNA clone that encodes a protein that has both DNA binding and transcriptional activation activity independent of the bait construct. To identify such artifactual "positives" it was necessary to cotranform haploid yeast containing each positive prey plasmid with haploid yeast containing empty-BD (BD: pGBKT7) to rule out intrinsic autoactivation of reporter genes by the prey. Thus, diploid yeast capable of growing on selection media independent of the bait protein were categorized as auto- activators and were not investigated further. Putative interacting clone At2g36530, a predicted 2-phospho-D-glycerate hydroylase, tuned out be a false positive as it alone activated the Gal4-responsive reporters in the absence of the bait (Figure 3.8 B).

Collectively, the PDF2-START library screen uncovered one unique interactor corresponding to At5g40830 (henceforth referred as SAM). Interaction between GL2-START and SAM was also confirmed in yeast, albeit weaker than the PDF2-START interaction (Figure 3.8 C).The *Arabidopsis* Electronic Fluorescent Pictograph Browser (eFP Browser- (Winter, 2007) shows At5g40830 expression in shoot epidermis, overlapping the expression pattern of PDF2. This would allow interaction between these two proteins. To confirm these interactions *in planta*, BiFC experiments in *N. benthamiana* leaves were performed. No GFP fluorescence was observed when full length SAM and PDF2-START proteins are present together (data not shown). Thus, the *in vivo* experiment did not offer direct evidence for an association between these two proteins, suggesting that the interaction observed in yeast may not be relevant *in planta* (possibly the posttranslational modifications of SAM protein might hamper the interaction). Two other possible mechanisms that could explain this result are – First, SAM is unstable or expressed at low levels which precludes BiFC complex formation. Another possibility is that presence of fusion tag at the N-terminus of SAM imposes topological constraints on BiFC complex formation. It is estimated at BiFC can occur when VN (N-terminal fragment of YFP) and VC (C-terminal fragment of YFP) are fused to positions that are separated by a distance no greater than ~10nm (Hu et al. 2002). As SAM is tagged with VN at its N-terminal end, if the site of interaction is located away from its N-terminus, the BiFC signal is unlikely to be detected for the protein. One such study by Sung et al. (2013) provided direct evidence for the effects of topological constraints on BiFC complex formation between SUMO and its target proteins. Alternatively, enhanced DNA binding of PDF2 *in planta* could titrate it away from SAM.

In a parallel experiment, the Y2H library screen was repeated using GL2-START protein under reduced stringency conditions. 18 unique interactors were uncovered among 25 strong interactors. After eliminating candidates that interact with the bait plasmid alone, the list was narrowed down to 6 confirmed positive interactors (Table 3.4). One of these candidates, At1g78570 (RHM1), a rhamnose synthase, codes for predicted UDP-glucose 4,6-dehydratase involved in the synthesis of rhamnose (Wang et al. 2009) and flavonoids (Yonekura-Sakakibara et al. 2008). Flavonoids are a class of secondary metabolites that have been shown to play preponderant roles in modulating plant growth and development (Mierziak et al. 2014; Mouradov and Spangenberg 2014).

***At1g78570 (RHM1)*:** Trifunctional UDP-glucose 4,6-dehydratase/UDP-4-keto-6-deoxy-D-glucose 3,5-epimerase/UDP-4-keto-L-rhamnose-reductase enzyme is involved in UDP-β-L-rhamnose biosynthesis. The dehydratase activity is contained in the N-terminal region while the epimerase and reductase activities are in the C-terminal region (Figure 3.9 A). AtRHM1 displays high similarity to the other members of RHM family in *Arabidopsis* (RHM2, RHM3 and UER1) and in other plant species such as rice and grape (Watt et al. 2004; Diet et al. 2006; Oka et al. 2007). In addition, the two conserved domains corresponding to the N-terminal dehydratase and the C-terminal epimerase/reductase in AtRHM1 were also identified in the RHM1 homologs of rice and grape, indicating that both protein architectures are conserved in plants (Wang et al. 2009). L-Rhamnose (Rha) is an important constituent of pectic polysaccharides, a major component of the cell walls of *Arabidopsis* (Ridley et al. 2001). Importantly, UDP-rhamnose is also used for flavonol rhamnosylation and RHM1 has been reported to play a major role in supplying UDP-rhamnose for modification of flavonols (Ringli et al. 2008). Consistent with this, *rhm1* (or *rol1*) mutants exhibit modifications in pectin structure (Diet et al. 2006). Also, *rhm1* mutants display aberrant cell growth and development phenotypes caused by modified flavonol glycosylation profile (Ringli et al. 2008). Furthermore, the 668 aa long RHM1 was shown to be ubiquitously expressed, with strongest expression in roots and expanding primary leaves of the *Arabidopsis* seedling (Wang et al. 2009).

Interaction between RHM1 and Class IV HD-Zip TFs require START domain

Initially, the Y2H system was used to identify the HD-Zip domains involved in interaction. A series of GL2 deletion variants and PDF2-START protein were used as baits and analysis by selection on growth plates was done (Figure 3.9 B). As prey, *RHM1* coding sequence was fused to to the GAL4 activation domain (AD) and empty plasmid was used as the negative control. Domain deletions revealed that RHM1 interacted strongly with the proteins containing the START domain (GL2-START, GL2-START+SAD (S+S) and PDF2-START). The strongest interaction was seen with the PDF2-START. Indeed, as shown in Figure 3.9 B , the GL2-START protein interacts more strongly than the full length GL2. This suggests that the full length protein may be present in a conformation that hinders interaction.

To validate the protein interactions *in planta*, BiFC studies using transient transfections in *N. benthamiana* were performed. BiFC experiments indicate strong interactions between full-length GL2 and RHM1 *in planta* (Figure 3.10 A), but deletion of the START domain in as well as $gl2^{Amo}$, which represents a mutant version of the START domain, result in loss of interaction (Figure 3.10 B). These interactions accord with the results from the homodimerization tests (Figure 3.6 B), suggesting that the formation of GL2 homodimers may favor the interactions between GL2 with RHM1. On the otherhand, the class IV HD-Zip family member HDG11 and class III HD- Zip family member PHB failed to show interaction, but PDF2 exhibited weak interaction (Figure 3.10 B). Strikingly, a strong fluorescent signal was observed in *N.benthamiana* cells co-expressing the START domain of PDF2 and RHM1 (Figures 3.10 A). GFP fluorescence could be seen in cytoplasm when interacting proteins are present together (Figure 3.10), but in some samples showing interaction between GL2 and RHM1, the unexpected appearance of punctate dots was observed (Figure 3.10 A). The background fluorescence signal from the negative controls used is negligible (Figures 3.10 A). The available *Arabidopsis* plasmodesmal proteome identified RHM1 as a novel potential plasmodesmal protein (Fernandez calvino et al, 2011). Based on this finding, it is hypothesized that the punctate dots (Figure 3.10 A) may represent staining at plasmodesmata. Constructs showing weak or no interaction were analyzed by western-blot to determine correct expression of the fusion proteins. While PHB was expressed weakly, expression of HDG11 could not be detected. Together, these data show that RHM1 interacts specifically with GL2 and PDF2, members of the class IV HD-Zip transcription factor family, and this interaction is mediated by the START domain.

RHM1 is localized in the cytoplasm and base of trichomes

Available microarray data indicates co-expression of *RHM1* with *GL2* in trichomes (Marks et al. 2009). In order for these proteins to interact with GL2, they must be present in the same subcellular compartment as well. GFP fusion protein of RHM1 driven by the 35S promoter was created and transient transfection into tobacco leaves was performed to examine subcellular distribution. RHM1 was found in the cytoplasm. This is consistent with previous work that showed cytoplasmic localization of RHM1 in the cotyledon epidermal cells of *Arabidopsis* (Diet et al, 2006). Cytoplasmic localization of GL2 has been demonstrated based on immunolocalization (Di Cristina et al., 1996) and work in our laboratory with GFP fusions support this observation. Although RHM1 is localized specifically to the cytoplasm in all tissues analyzed to date (Diet et al. 2006), no data have been reported on their localization in the trichomes. Interestingly, similar subcellular expression for *35S::RHM1-GFP* and *35S::PDF2 (START)-GFP* fusions in trichomes was observed. RHM1 localizes to the base of the trichome while PDF2 (START) localizes to the base of the trichome as well as the nucleus of the trichome (KS, unpublished). Based on these observations, it is proposed that the interaction between RHM1 and the class IV HD-Zip proteins may be transient and most likely occurs outside the nucleus.

ChIP with *EYFP:GL2* transgenic line

The current understanding of a cell-specific regulation of the class IV HD-Zip gene expression is very limited. There is no global DNA-binding study reported for any class IV HD-Zip family member to date. Using GL2 as a model class IV HD-Zip family member, Chromatin Immunoprecipitation (ChIP) was performed followed by deep sequencing by ChIP-SEQ. In addition to the wild-type *GL2* that is driven by its own 2.1-kb promoter, *proGL2:EYFP:GL2,* several mutant versions including *gl2* null mutant (negative control), *gl2ΔZip* and START domain mutants *gl2ΔSTART,* $gl2^{E375G;R392M}$ and $gl2^{Amo}$ were used to determine the global DNA-binding properties related to START domain function. As mentioned above, these mutants exhibit nuclear expression of the protein, but result in either loss-of-function (*gl2ΔZip, gl2ΔSTART,* $gl2^{E375G;R392M}$) or gain-of-function ($gl2^{Amo}$), making them particularly suitable to study DNA binding profiles. With regards to $gl2^{Amo}$, it is hypothesized that the mutant protein

has an altered DNA-binding profile, given its dominant-negative phenotype. Consequently, a genome-wide approach to addressing its DNA-binding function could be especially revealing. Three biological replicates and one input of each sample were used to compare pre-ChIP DNA with ChIP DNA. After ChIP, the recovered DNA was PCR-amplified and size-fractionated, following the preparation protocol for an Illumina sequencing library. The goals for this experiment were to (i) Delineate global *in vivo* binding sites of GL2, (ii) Identify novel transcriptional targets in an unbiased manner, (iii) link these new candidates to biological processes, and (iv) test whether the START domain is required for DNA binding of GL2. Chromatin immunoprecipitation (ChIP) followed by ultrahigh-throughput Solexa (Illumina) sequencing (ChIP-SEQ) has been shown to be a powerful tool to obtain genome-wide DNA-binding patterns of transcription factors (Johnson et al. 2007; Robertson et al. 2007). The large numbers of short individual sequence reads produced by novel instruments facilitate the digital quantification of DNA sequences that are present in a sample.

Because of the anatomical differences between animal and plant cells, such as rigid cell walls, high levels of cellulose and lignin, and large vacuoles in plant cells, several modifications are needed to establish efficient ChIP protocols for plant systems (Gendrel et al. 2002; Saleh et al. 2008; Kaufmann et al. 2010). In this study, the published ChIP method for the genome-wide identification of transcription factor binding sites in plants (Kaufmann et al. 2010) was optimized and validated prior to Illumina sequencing.

Genome-wide mapping of regions bound by GL2 and mutant versions

DNA libraries were generated for each ChIP sample, followed by ultra-high-throughput Solexa (Illumina) sequencing. From the ChIP-Seq data, 18 to 66 million reads (~50 bp per read) were obtained, which were uniquely mapped to the *Arabidopsis* genome by adopting Model-based Analysis of ChIP-Seq (MACS) software (Zhang et al. 2008). Although the input samples mapped very well, the percentage of uniquely assigned loci were quite low for the ChIP samples. However, merging replicates provided a reasonable number of uniquely mapped reads to proceed with the analysis (Table 3.5). At FDR<0.1, 1284 significantly enriched regions distributed over all five chromosomes was observed for the wild type-GL2 . While the binding profile of $gl2^{E375G;R392M}$ closely resembled GL2, both $gl2^{Amo}$ and gl2ΔSTART transcription factors displayed enhanced binding in comparison with the wild type GL2 (Figure 3.11 A and Table

3.6). Importantly, gl2ΔZip behaved similar to the *gl2* null mutant, as binding of GL2 to DNA appears abolished in ZLZ, indicating that the leucine zipper domain mediated dimerization of GL2 is necessary for DNA binding *in vivo*. Dimerization is a prerequisite for DNA binding by the class IV HD-Zip IV. Unlike other HD proteins, which can bind DNA as monomers, the class IV HD-Zip TFs can bind DNA efficiently only as either homo- or hetero-dimers (Palena et al. 2001).

cis-Regulatory elements controlling gene expression are preferentially found in the promoters of target genes. However, there are also numerous examples of important regulatory sequences downstream of the transcription start site. When we examined the position of binding sites relative to the nearest gene in the ChIP-Seq dataset, most *in vivo* binding sites were found in the promoter or intergenic regions, which constitute bidirectional promoters (Figure 3.11 A). This distribution pattern is consistent with the determined GL2 binding sites in the promoters of *CESA5, XTH17* (Tominaga-Wada et al. 2009), *PLDξ1*(Ohashi et al. 2003), *MYB23* (Khosla et al. 2014) and with the molecular function of GL2 as a transcription factor.

Analysis of GL2 binding motifs in promoters, genic, and intergenic regions

In *Arabidopsis*, the class IV HD-Zip factors bind to the conserved *cis*-element, L1-box sequence (5'-TAAATG(C/T) A-3'), in the promoter of the target genes (Abe et al. 2001; Abe et al. 2003; Nakamura et al. 2006). It has been reported that L1-box region of *PHOSHOLIPASE DΞ1* (*PLDξ1*), *CELLULOSE SYNTHASE 5* (*CESA5*), *XYLOGLUCAN ENDOTRANSGLYCOSYLASE* (*XTH17*), and *MYB23* promoter is a direct target of GL2 (Ohashi et al. 2003; Tominaga-Wada et al. 2009; Khosla et al. 2014). In addition to the asymmetric L1 box described above, a slightly longer palindromic L1-like box 5'-GCATTAAATGC-3' has been defined as the consensus-binding site of the class IV HD-Zip IV HDG7, HDG9, and ATML1 (Nakamura et al. 2006). Consistent with these reports, a moderate portion of the identified GL2 direct target genes contain one or more typical L1 box motifs (Table 3.7). To investigate whether there is any other potential GL2 binding motif, we used a motif searching program, HOMER, to discover statistically overrepresented motifs in GL2 binding regions that were assigned to genes. However, HOMER failed to identify any novel motif (data not shown). The lack of identifiable binding motifs at some of the sites may be explained by the existence of indirect interactions with DNA mediated by other proteins. Since the remaining GL2 direct target

genes do not contain any L1 box, GL2 may bind to these regions through other motifs or it may associate with DNA binding protein. Alternatively, the low alignment statistics did not allow us to detect all L1 box target genes, thereby resulting in false negatives. Strikingly, for several genes, we observed altered binding profile for $gl2^{Amo}$. In one such example, in comparison with the wild type GL2, $gl2^{Amo}$ binds further upstream of the transcription start site (TSS) in the intergenic region upstream of α−*DOX2* gene. Example of binding region profile is shown in Figure. 3.11 B. Together with the fact that deletion of the START domain leads to enhanced DNA binding, our results imply that START domain is dispensable for DNA binding of GLABRA2.

Functional classification of putative GL2 target genes

Gene Ontology (GO) classification system revealed overlapping functional classes of genes bound by GL2. In terms of molecular function, genes encoding transferase activity (GO:0016740; *p*-value 8e-09), kinase activity (GO:0016301, *p*-value 6.1e-07), and metal ion binding (GO:0046872; *p*-value 4.5e-05) are statisitcally enriched compared with the whole genome. When dissecting gene functions according to biological processes, there is a clear enrichment for genes involved in cellular process (GO:0009987; *p*-value 4.8e-31) and metabolism (GO:0008152; *p*-value 2.4e-42). Figure 3.11 C illustrates the eleven most enriched specific GO terms. Notably, the frequency of genes associated with development, multicellular organisimal process, and response to stimulus showed statistically higher occurence than that found within the whole *Arabidopsis* genome, suggesting that these represent direct targets of GL2.

Identification of GL2 targets

Out of over 800 significantly enriched genes, 45 were identified to contain at least one perfect L-1 box in their promoters (Table 3.7), indicating that they could serve as direct targets for GL2. Among them, the majority was classified into biological processes such as transcription factor activity, metabolism, development, and response to hormone stimulus. Putative targets included two lipid binding proteins, α-dioxygenase (α-DOX2) and calcium-dependent lipid-binding protein (AtCLB) containing a C2 domain, transcription factors MYB23, KNAT1, and

genes regulated by different hormones (eg., *ARF19* and *BSL3*) To determine functional relevance of GL2 binding sites to the transcriptional regulation, we queried the available microarray data for differentially expressed genes in the root-hair vs non-root hair cell types (Bruex et al. 2012). We also included the results from genome-wide expression analysis from our lab, which examined gene expression changes using wild-type GL2, $gl2^{Amo}$, and *gl2* in seedlings of the mutant plants.

Evidence for differential expression of the target genes in *gl2* mutants was found for 14% of the direct targets (Table 3.7). Interestingly, several targets are co-expressed or have down-regulated expression relative to GL2 in either root or trichomes (see Table 3.7). Together, our results suggest that GL2 may regulate the expression of these genes by directly binding to the L1 box motifs in their gene promoters. Because it is anticipated that the direct target genes of GL2 would tend to exhibit a similar expression pattern as GL2, this observed degree of enrichment of the GL2 co-expressed genes substantiates this notion.

Transcriptional regulation of genes identified as putative direct targets

Next, we asked whether the expression of these candidate target genes is transcriptionally regulated by GL2. Quantitative real-time PCR (qRT-PCR) showed that the transcript levels of *CHALCONE ISOMERASE LIKE* (*CHIL)*, *PATATIN-RELATED PHOSPHOLIPASE A* (*PLP9)*, *GLUTATHIONE REDUCTASE (GRXC2)*, and *HYDROLASE* (At3g23600) exhibit statistically significant up-regulation in *gl2* roots compared with the wild-type (Figure 3.12 B). Conversely, the expression levels of *MYB23* were sharply decreased in the shoots of the *gl2* mutant, consistent with a previous study (Figure 3.12) (Khosla et al. 2014). Strikingly, qRT-PCR revealed a severe reduction (~35 fold) in *DOX2* transcript levels in roots from *gl2* mutant, indicating negative regulation of *DOX2* expression by GL2 (Figure 3.12 B).

Collectively these findings indicate that GL2 is involved in modulating target genes that are involved in metabolism, thereby reinforcing the hypothesis that the class IV HD-Zip TFs may play key roles in linking metabolism to gene transcription in plants. A detailed description of the target genes will be discussed later in this chapter.

DISCUSSION AND FUTURE PERSPECTIVES

In the last few decades, molecular genetic studies have revealed the class IV HD-Zip transcription factors as key regulators of epidermal cell specification in flowering plants (Javelle et al. 2011; Nadakuduti et al. 2012). Further, they are postulated to be instrumental in the evolution of epidermal adaptations for successful colonization of land plants (Graham et al. 2000; Ligrone et al. 2012). In contrast to the class III HD-Zip class members; understanding of the regulation of the class IV members is limited. The proteins bind L1 box motif in targets genes, suggesting a regulatory role for the L1 layer –specific gene expression (Nakamura et al. 2006). Besides, there is evidence for auto regulation mediated by direct binding to the L1 box in its promoter (Abe et al. 2003; Takada and Juergens 2007). Furthermore, regulation by redox signals has been suggested for related TFs in sunflower (Tron et al. 2001). Most recently, regulation by binding to DELLA inhibitory proteins was uncovered (Rombola-Caldentey et al. 2014; Shan et al. 2014).

It has been postulated that START domains serve as regulatory motifs; however, the functions of this domain have not been extensively studied. A recent report provides compelling evidence for the role of the START domains in stimulating transcription factor via ligand binding (Schrick et al. 2014). Here, these studies are extended to gain further insight into the regulatory function of the START domain in the class IV HD-Zip family members.

A role for the START domain in ligand binding

The ability of the START domain to bind ligands was assayed by testing several ligand binding mutants for their ability to complement *gl2* mutant phenotypes. Both R392M and E375G; R392M bridge mutations failed to complement mutant phenotypes in the root, leaves, and seeds (Figure 3.2). In agreement with this, the salt bridge between an acidic residue (Glu169/ STARD1 and Asp332/STARD3) and a basic residue (Arg188/STARD1 and Arg351/ STARD3) was proposed to be a key determinant for cholesterol binding in human START domain proteins (Tsujishita and Hurley 2000). L480 is a highly conserved residue in the C-terminal α-helix that has been shown to interact with lipid membranes and regulate the accessibility of the lipid ligand to the ligand-binding cavity in human START proteins (Alpy and Tomasetto 2014). In this study, the substitution of leucine with proline, a potent α-helix breaker confers partial complementation

of mutant phenotypes (Figure 3.2). One explanation for this result is that α-helix plays a subsidiary role in structural rearrangements that are necessary to take up the ligand. The finding that molecular dynamic simulations favor the sole involvement of Ω1 loop in driving such structural rearrangements supports this proposition (Murcia et al. 2006; Kudo et al. 2008). Of note, all mutant constructs with the exception of R392M displayed nuclear localization, implying that the protein is correctly folded.

According to the literature, antimorphic alleles, also termed "dominant-negative" alleles, interfere with the wild-type function and are defined as dominant or semi-dominant alleles (Wilkie 1994). Here, we report a novel semi-dominant mutant (*gl2*^{Amorphous}) that confers a severe trichome-less phenotype and intermediate phenotypes (partial trichome formation) depending upon the dosage of the transgene. This dominant negative allele is due to the *gl2*$^{R386L;E387L}$ missense mutation followed by an 11-amino acid duplication in a predicted ligand-contact region (Figure 3.3).The mutant protein exhibited thin-walled trichomes that seem to lack epicuticular wax, a phenotype not observed in the *gl2* null allele. A plausible interpretation of this data is that *gl2*^{Amo} acts as a dominant negative allele of GL2 that interferes with some unknown factor(s) to yield defective wax deposition phenotype –possibly by direct binding and then disabling their function. Alternatively, the mutant protein binds to GL2 target genes but may fail to interact with cofactors required for transcriptional regulation. Further studies, such as a Y2H screen with ^{Amo} could reveal additional components, and thus provide useful insights into the molecular mechanism underlying the *gl2*^{Amo} phenotype.

Strikingly, the *gl2*^{Amo} phenotype resembles the trichome-less phenotype recently published for *gl2 hdg11* (Khosla et al. 2014) and *gl2 myb23* (Kirik et al. 2005) double mutant plants. Based on this finding, it is possible that antimorphic *gl2*^{Amo} interferes with activity of HDG11 and possibly other functionally redundant HD-Zip family members by competing for the same partner in the protein complex or competing for the same DNA binding sites. Given that both *MYB23* and *HDG11* transcripts are downregulated in *gl2*^{Amo} further supports this idea (data not shown). Similar to *gl2*^{Amo}, there are many examples in the literature of semi-dominant or dominant missense alleles that act to interfere with the function of redundant factors to exhibit a stronger phenotype than the corresponding null. For instance, the *Arabidopsis topless-1* (*tpl-1*) mutation transforms the shoot pole into a second root pole, and *tpl1-1* is a dominant-negative

mutation that interferes with the function of multiple TPL-related proteins in embryo development (Long et al. 2006).

The microarray analysis reveals a gene expression profile different from both the wild-type and *gl2* null, suggesting that a transcription factor complex containing $gl2^{Amo}$ has altered function. For example, genes encoding proteins involved in lipid transport and localization, microtubule-based movement, and cytoskeleton protein binding were preferentially expressed in $gl2^{Amo}$ (Appendix Figure B.3). Finally, the enhanced DNA binding observed in our ChIP-seq data lends further credence to the notion that $gl2^{Amo}$ mutant protein behaves differently from both the wild-type and null mutant.

Similar to gl2ΔSTART, nuclear localization is observed upon deletion of the ZLZ domain, suggesting that both START and ZLZ domains are dispensable for nuclear localization of the HD-Zip protein (Figure 3.5). In contrast, deletion of the HD domain resulted in non-nuclear expression of the protein. The higher cytoplasmic fluorescence could have been caused by passive diffusion from the nucleus to the cytoplasm or by inefficient nuclear import. Relevant to this, numerous studies have reported that homeobox-containing transcription factors contain a nuclear localization signal (NLS) motif at either the N-terminus or C-terminus of their HD (Ghaffari et al. 1997; Hessabi et al. 2000). Given that the NLS motif (^{102}RKKYHRH^{108}) is located in the N-terminus of the HD is consistent with the requirement of HD for nuclear localization. Some homeobox proteins require an additional motif within the HD for complete nuclear localization (Christophe-Hobertus et al. 1999). Thus, to exclude the possibility that C-terminal part also contains determinants that enhance nuclear localization of GL2, it will be interesting to determine the nuclear/cytoplasmic distribution of the deletion mutants lacking the putative NLS motif. In addition, future work may include studies to determine whether direct interactions between the NLS motif and nuclear pore complex components drive nuclear localization.

The START domain plays a role in regulating GL2 dimerization

The HD-Zip class of proteins is unique to plants, and is characterized by the presence of a leucine zipper domain (ZLZ) adjacent to HD, which is important for dimerization (Ariel et al. 2007). Unlike other known HD proteins that are able to bind DNA sequence as monomers, HD-Zip TFs bind to DNA and consequently activate/repress downstream gene(s) only as dimers

(Sessa et al. 1993; Sessa et al. 1997). The presence of leucine zipper means that the class IV HD-Zip family members may potentially form homodimers or and heterodimers both between themselves and with other family members, adding a further layer of complexity to their regulatory role. The results form dimerization assays in yeast and plant cells (BiFC assays) demonstrate that ZLZ mediated dimerization requires the START domain. A point mutation in the START domain of WHEAT KINASE START1 protein (WKS1) significantly reduced the ability WKS1.1 to homodimerize, further underscoring the importance of the START domain in regulating dimerization (Gou et al. 2015).

Dimerization among the class IV HD-Zip members has previously been reported. For example, homodimerization of PDF2 and heterodimerization of PDF2 with ATML1 has been shown *in planta* (Rombola-Caldentey et al. 2014; San-Bento et al. 2014). Accordingly, we also observed the interaction between GL2 and two other class IV HD-Zip family members, PDF2 and ATML1 *in planta*, albeit weaker than GL2-GL2 association (Figure 3.7). Based on previous findings and data from this study, we have strong reasons to believe that START domain is a functional ligand-binding module. One intriguing possibility is that dimerization is also ligand dependent. There is evidence in both plants and animals that ligand binding could play a role in dimerization and binding of DNA. For instance, the 9-cis-retinoic acid can increase binding of RXR homodimers to DNA, which in turn leads to unavailability for heterodimer formation with other proteins (Zhang et al. 1992). The observation that GL2 favors homodimerization over heterodimerization accords with this finding. It is worth noting that mutation of putative ligand binding residues in the START domain eliminated homodimerization, further supporting the role of ligands in regulating dimerization directly or indirectly. The result that $gl2^{Amo}$ and gl2ΔSTART proteins form nearly normal levels of heterodimers with ATML1/PDF2 raises an interesting question of why START is only required for homodimerization? One possibility is that in absence of the START domain, GL2 monomer has a ZLZ domain conformation that heterodimerizes in preference to forming homodimers. Alternatively, the homodimers formed are ill defined or unstable, releasing gl2ΔSTART to interact with other proteins, such as ATML1 and PDF2. At this time, we cannot exclude the possibility that factors besides ligands may explain the differential role of the START domain in mediating homo vs. heterodimerization.

Based on these findings, a model is proposed for the ligand-induced homodimerization of GL2. Upon ligand binding to the START domain (Figure 3.13 A), the conformation of the GL2

is changed in such a way that the "hidden" ZLZ domain is exposed, leading to formation of active GL2 monomer. The active monomer is translocated to the nucleus. Once in the nucleus, forms stable homodimers and/or heterodimers capable of DNA binding and activating target gene expression. However, in the absence of the ligand binding START domain (Figure 3.13 B), forms ill-defined unstable homodimers. Instability of homodimers provides a thermodynamic driving force for preferential stable heterodimer formation in the presence of either binding partner (ATML1/PDF2).

GL2 regulates target genes involved in lipid and plant secondary metabolism

Comprehensive genome wide transcriptome data has identified a large number of genes that appear to be regulated by GL2 (Lieckfeldt et al. 2008; Marks et al. 2009; Won et al. 2009; Bruex et al. 2012), but so far only four direct targets of GL2 have been identified. These include *PLDξ1* (Ohashi et al. 2003), *CESA5* , *XTH17* (Tominaga-Wada et al. 2009), and *MYB23* (Khosla et al. 2014) . The ChIP-seq analysis revealed 45 genes that contain at least one L1-box (TAAATGTA or TAAATGTT) in their promoter, indicating they could serve as direct targets of GL2 (Table 3.7). Putative targets included two genes involved in lipid metabolism, α-*DOX2* (Bannenberg et al. 2009) and *CALCIUM-DEPENDENT LIPID-BINDING PROTEIN (AtCLB)* (de Silva et al. 2011) . The genes encoding transcription factors among GL2 targets included *MYB23, WUSCHEL RELATED HOMEOBOX 9 (WOX9), KNAT1, UNE16, SEPALLATA4, and SPATULA (SPT)* (Riechmann et al. 2000). The putative GL2 target genes also included plant secondary metabolism enzymes (such as *PAL2, SSL2, FAH1, 5-FCL*) (Wanner et al. 1995; Fabbri et al. 2000; Pribat et al. 2011; Maruta et al. 2014).

Plant α-dioxygenases are FA-hydroperoxidases, which catalyze the oxygenation of predominantly linoleic acid (18:2) and linolenic acid (18:3) into reactive hydroperoxides. Oxygenated derivatives resulting from secondary transformation of these hydroperoxides include the phytohormone jasmonic acid (JA) as well as reactive oxylipins (Vellosillo et al. 2007). The importance of these oxygenated FA-derived lipid mediators in plant defense signaling in response to different biotic and abiotic stresses is well documented (Seki et al. 2002; Izaguirre et al. 2003; Koeduka et al. 2005; Tirajoh et al. 2005). The role of oxylipins in cell wall modifications required for lateral root development in *Arabidopsis* has also been reported (Vellosillo et al. 2007). Further support for the role of DOX genes in plant development is

derived from results showing that in tomato, mutations of *α-DOX2*, also known as *feebly* or *divaricata* (*div*) cause defects in plant development as well as accumulation of anthocyanins. Also, transcription changes in such mutants showed up-regulation of genes playing roles in lipid and phenylpropanoid metabolism (Bannenberg et al. 2009). The α-DOX proteins identified in different plant species show high amino acid homologies. For example, the amino acid sequence of *N. tabacum* α-DOX shares high amino acid similarities with the proteins in *N. attenuata* (95% identity), *C. annuum* (85% identity), *Solanum lycopersicum* (84% identity), *A. thaliana* (75% identity), and *O. sativa* (63% identity) (Hamberg et al. 2002). A large-scale comparative transcriptional profiling of root epidermis reported the preferential expression of *DOX2* in differentiating hair cells (Bruex et al. 2012), suggesting the idea that GL2 acts mostly as a transcriptional repressor of *DOX2*. Fitting with this notion, the qRT-PCR data revealed a significant reduction in *DOX2* transcript levels in roots from *gl2* mutants (Figure 3.12).

A second GL2 target that codes for phenylalanine ammonia lyase (PAL) is linked to plant secondary metabolism. Phenylpropanoids represent a large group of polyphenolic compounds that comprise an important class of secondary metabolites such as flavonoids, anthocyanin and lignin in plants (Zhang and Liu 2015). Emergence of the phenylpropanoid pathway is among a number of important adaptations that allowed land plants to survive under harsh terrestrial environmental conditions, including desiccation, UV radiation, and microbial attack (Kenrick and Crane 1997; Ferrer et al. 2008). PAL catalyzes the deamination of Phe to give cinnamic acid, which is the first step in the phenylpropanoid pathway and an important regulation point between primary and secondary metabolism (Huang et al. 2010). PAL is encoded by a small gene family in plants with four members in *Arabidopsis* (*Arabidopsis thaliana*; *PAL1–PAL4* (Raes et al. 2003). MYB, LIM-domain containing proteins, and KNOX family of transcription factors have been shown to regulate *PAL* expression (Tsiantis 2001; Zhou et al. 2009). However, thus far, only LIM domain containing proteins are reported to regulate directly PAL expression by binding to the AC-rich motif in its promoter (Kawaoka and Ebinuma 2001). To date, there is no evidence for the regulation of flavonoid biosynthesis enzymes by the class IV HD-Zip transcription factors. Recently, Wang et al. (2015) demonstrated that GL2 could negatively regulate anthocyanin biosynthesis by directly repressing the expression of several R2R3 MYB transcription factors, thus expanding the repertoire of already known direct targets of GL2. Additionally, the available microarray data shows that the genes encoding PLP9 are

preferentially expressed in the root hair cells and trichomes (Marks et al. 2009), consistent with direct regulation by GL2. Moreover, a recent transcriptome data comparing root epidermis of hairy vs. hairless mutants indicated preferential expression of *PAL2* in differentiating hair cells (Bruex et al. 2012). In agreement with these observations, our qRT-PCR data revealed significantly up-regulated transcript levels in *gl2* mutant plants in comparison with the wild type Col, indicating negative regulation by GL2. The ChIP-seq data together with the enrichment of L1 box motif in the *PAL2* promoter suggests that *PAL2* is a direct target of GL2. Further studies, such as luciferase reporter assays (using wt and mutated L1 box) will be necessary to provide evidence for direct transcriptional activation by GL2 through specific L1 box motifs.

Previous work has reported a possible link between the class IV HD-Zip TFs and lipid metabolism. While the *PLDξ1* gene promoter contains an L1 box that is activated by GL2 in *Arabidopsis* (Ohashi et al. 2003), *ZmWBC11a* is transcriptionally activated by OCL1 in the maize kernel, and this transactivation is abolished by mutations in the L1 box (Javelle et al. 2011). In addition to *ZmWBC11a*, OCL1 directly or indirectly regulates several other target genes involved in lipid metabolism or transport. Besides, recent evolutionary and phylogenetic analyses have convincingly argued that among all the known functions of these genes, a role in lipid binding and transport is a promising candidate for an ancestral function, and it is highly likely that the class IV HD-Zip gene in algal ancestor played a role in the origin of cuticle (Zalewski et al. 2013). Our data provides further evidence for the hypothesis that the class IV HD-Zip TFs play important regulatory roles in the differentiation or maintenance of the epidermis in general and lipid and plant secondary metabolism in particular. In this framework, the presence of the putative lipid/sterol-binding START domain leads to an attractive hypothesis that activation of metabolic pathways may depend on the sensing of lipids or small metabolites by the START domain.

GL2 is a transcriptional repressor

Opposite roles for GL2 in the regulation of epidermal development in flowering plants has been previously described. For instance, GL2 positively regulates trichome formation, but functions as a negative regulator of root hair formation (Rerie et al. 1994; Di Cristina et al. 1996; Hung et al. 1998). While both *CESA5* and *XTH17* are direct target genes of GL2, GL2 negatively

regulates the expression of *CESA5* while acting as a positive regulator of *XTH17* (Tominaga-Wada et al. 2009). In addition to GL2, an inhibitory role for two other class IV HD-Zip family members has been demonstrated. Genetic and molecular studies confirmed that in contrast to the published literature, both ATML1 and PDF2 negatively regulate both their own transcription and the transcription of *ACR4* (San-Bento et al. 2014). Recently, using *Arabidopsis* protoplast transfection assays, it was shown that the recruitment of the DNA binding domain of GL2 (HD-GL2) to the promoter of *LexA-GAL4:GUS* reporter results in a significant decrease in the expression of the reporter gene (Wang et al. 2015). Collectively, these findings suggest an inhibitory role for members of the class IV HD-Zip family of transcription factors. The qRT-PCR data showed that 6 out of 7 target genes are up-regulated in roots from *gl2* mutant plants, indicating that GL2 acts a transcriptional repressor to repress the target genes in roots (Figure 3.15). Based on the microarray data available from the *Arabidopsis* eFP browser (Marks et al. 2009), the gene encoding *MYB23* is expressed in the trichomes, similar to the expression observed for GL2. It has been postulated that the repression function of GL2 may be due to the presence of LxLxL motif adjacent to the homeodomain (Wang et al. 2015). The transcriptional repression domain containing a conserved LxLxL motif is required for the repression function of Aux/IAA and ERF transcription factors (Ohta et al. 2001; Tiwari et al. 2004), and the motif is also present in the AtOFP1 protein (Wang et al. 2007). Future work is required to determine whether LxLxL motif mediates the repressive function of GL2.

START domain is dispensable for DNA binding of the class IV HD-Zip proteins

Genome-wide analysis using ChIP-seq revealed enhanced DNA binding for gl2ΔSTART and gl2^{Amo}, suggesting that START domain is not required for DNA binding by the HD-Zip proteins (Figure 3.11). These data raise the question as to why the mutant proteins display augmented binding in comparison with the wild-type protein. This might be explained by their ability to form heterodimers. Possible mechanisms of enhanced binding of heterodimers have been described elsewhere (Weltmeier et al. 2006), including increased affinity for binding target sites and stabilization of dimer structure (as effective DNA binding forms) via equilibrium displacement in dimer-monomer concentrations. Alternatively, improved DNA binding can be explained by the absence of interaction with inhibitory proteins, such as DELLA. DELLA proteins lack typical DNA-binding domain, and negatively regulate gibberellin (GA) signaling

by suppressing activities of transcription factors through protein-protein interaction (Daviere and Achard 2013). This hypothesis draws on from two recent findings. In one study, *Arabidopsis* DELLA proteins were shown to sequester the class IV HD-Zip transcription factors from activating downstream target genes, thereby blocking GA signaling in the epidermis (Rombola-Caldentey et al. 2014). In the second study performed in cotton, Shan et al. (2014) suggested the interaction between GhHOX3, a class IV HD-Zip TF and GhSLR1, a functional orthologue of *Arabidopsis* DELLA requires a functional START domain. Experiments to test this hypothesis in are currently in progress.

Multiple lines of evidence show enhanced DNA binding by heterodimers in comparison with homodimers. For example, in gel shift assays, WRKY18 and WRKY40 heterodimers bind much stronger to W-box consensus sequence than the respective homodimers (Xu et al. 2006). This finding has been extended by another report that showed activation of WRKY60 promoter by WRKY18/WRKY40 heterodimers in reporter gene assay, whereas the homodimers had no effect (Chen et al. 2010). In a second example, heterodimerization of AtbZIP53 and AtbZIP10 or AtbZIP25 enhances DNA binding and synergistically increases target gene activation (Alonso et al. 2009). In addition to enhanced binding, formation of heterodimers is of great functional significance as it may integrate information from different signaling pathways thereby increasing the repertoire of target genes.

RHM1 is a novel protein interacting partner of the START domain

A previous study showed that START+SAD domain of GbML1, a class IV HD-Zip family member from cotton could participate in protein interactions (Zhang et al. 2010). Similarly, Y2H screen performed with PCTP, a human START protein, uncovered novel interactions with the enzyme thioesterase and the Pax3 HD transcription factor (Kanno et al. 2007). Combined, these studies indicate the role of the START domain in mediating protein-protein interactions. The yeast-two hybrid screen using START domain from GL2 and PDF2 revealed candidates for new interacting partners for the START domain. The work described focused on RHM1, UDP-L-Rhamnose synthase involved in the biosynthesis of rhamnose, a major monosaccharide component of pectin (Diet et al. 2006) and flavonoids (Ringli et al. 2008; Yonekura-Sakakibara et al. 2008). In BiFC assays, strong fluorescent signals indicated that the interaction between GL2 and RHM1 is direct and occurs *in planta.* Furthermore*,* the deletion of

the START domain as well as gl2^{Amo} resulted in the loss of interaction, suggesting that the interaction is mediated by the START domain. Given that RHM1:GFP is localized in the cytoplasm (Wang et al. 2009), we found that the interaction between RHM1 and HD-Zip TFs is mostly cytoplasmic. One exception is the appearance of bright punctate dots in some samples showing the interaction between GL2 and RHM1. Consistent with its putative interaction with GL2, epidermal defects including a collapsed trichome phenotype reminiscent of *gl2^{Amo}* are associated with *rol1-2* mutants corresponding to this gene (Ringli et al. 2008). Moreover, available microarray data reveals co-expression with *GL2* in trichomes (Marks et al. 2009). Genetic approaches to establish a link between GL2, RHM1, and epidermal cell-type differentiation are currently underway. Collectively, these findings provide additional evidence for the role of the START domain as a protein-binding motif in plants.

MATERIALS AND METHODS

Plants Material and Growth Conditions

The wild-type *Arabidopsis thaliana* ecotype was Columbia (Col). The *gl2* null allele was *gl2- 5* (Ohashi et al. 2002) . Seeds were stratified at 4˚C for 3-5 days and grown on soil mix containing plug and germination growing medium (PRO-MIX® PGX), vermiculite and perlite (Therm-o-Rock West, Inc.) in 4:1:1 ratio, and 5-6 pellets of fertilizer osmocote 14-14-14 (Everris US), at 23˚C under continuous light. Seeds of *N. benthamiana* were germinated on wet filter paper for 7d, and then the seedlings were planted on above described soil mix. Plants were grown in a growth chamber at 22°C under 14-h-light/10-h-dark cycles. For each set of experiments, we used 3-4 week-old plants that were typically at the five-leaf stage.

Molecular cloning and generation of transgenic plants

To construct the SR54 binary vector containing *proGL2:EYFP:GL2*, the *EYFP* coding region was excised from *pEYFP-C1* (Clontech) with Eco47III/SalI and cloned into the SmaI/SalI site of the SR2 (*proGL2;GL2*) construct. *SR2* was made by cloning the *GL2* coding region from a *GL2pGEMT* clone using *Sal*I/*Kpn*I into C15, a modified *p35S-CAMBIA1300* binary vector that contains the 2115 bp promoter region from the HindIII site upstream of *GL2*. The *GL2* cDNA was sub-cloned into pBluescript via *Kpn*I and *Sal*I restriction sites to serve as a template for making GL2 deletion variants and START domain mutants used in this study. Site-directed mutagenesis was performed using a one-step PCR-based method modified from Scott et al. (2002). The reaction mixture contained 25 ng template DNA, 135 ng of each oligonucleotide, 1 µl of 10 mM dNTPs solution, 1 µl *PfuUltra*™ II Fusion HS DNA polymerase (Agilent Technologies), 5 µl of the buffer supplied with the polymerase, and H_20 in a total volume of 50 µl. PCR was performed under the following conditions: 95°C denaturation for 1 minute; 16 cycles of 95°C denaturation for 50 seconds, 60°C annealing for 50 seconds, and 68°C extension for 1 min/kb; extension at 68°C for 7 minutes. PCR products were digested with 30U of *Dpn*I for 3 hours at 37°C for selective digestion of methylated or hemimethylated parental strands, followed by ethanol precipitation and transformation into 50 µl of XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA). After sequence confirmation, the *GL2* deletion and mutation versions were restriction digested with *Kpn*I and *Sal*I, and ligated into the binary vector SR54.

Following sequence verification, the constructs were transformed into *gl2-5* plants by floral dip transformation using *Agrobacterium* strain GV3101 (pMP90) (Clough and Bent 1998). Transgenic plants were selected with 50 mg/ml hygromycin B. T3 homozygous lines with stable EYFP expression were confirmed by PCR and used for phenotypic assays as described below. See Table 3.8 for a list of plasmid constructs generated in this study and Table 3.9 for a list of primer DNA sequences for molecular cloning.

Phenotypic analysis and microscopy

For mucilage analysis, seeds were stained in a 0.25% aqueous solution of ruthenium red (Sigma) for 1 hour and replaced with deionized water prior to observation. Seeds and rosettes were imaged with a Leica M125 fluorescence stereomicroscope, and images were captured using a Leica DFC295 digital camera and Leica Application Suite software version 4.1. Ovules were excised from siliques and mounted in water under a coverslip for confocal microscopy. For propidium iodide staining of root meristem cell walls, seeds were sown on 0.8% water agar plates, stratified for 3 to 4 d at 4°C to obtain uniform germination, and then transferred to 23°C with the plates in vertical orientation. At day 4 after germination, seedlings were removed and placed in a well of microtiter plate with 1 ml of staining solution for 3 min at room temperature. An aqueous 10 μg/ml solution of propidium iodide (Sigma) was used to stain the cell walls of the *Arabidopsis* root meristem. Seedlings were rinsed once with distilled water and then mounted in water under a coverslip for direct microscopic observation. Confocal images were collected on Zeiss LSM-5 Pascal confocal microscope. GFP/YFP was excited with an argon laser (488 nm), and emitted light was collected between 510 and 540 nm. mCherry and propidium iodide fluorescence was imaged using an excitation wavelength of 543 nm with the emission wavelength at 587-625 nm. Images were processed using LSM 5 Image Browser (Zeiss). The mean pixel intensity of three different circular regions of interest (ROIs) was quantified by the Image J (National Institutes of Health) software. In total, ten different images were analyzed for each construct. Scanning electron microscopy was performed as described previously (Khosla et al. 2014)

Stable and transient expression in *N.benthamiana*

To generate constructs for split GFP, gateway compatible BiFC vectors 125-NXGW and 127-CXGW that carried either fragment of split GFP in N-terminus position were used for LR recombination reaction (kindly provided by Adrienne Roeder). See Table 3.8 for construct information. Agroinoculation of *N. benthamiana* was performed as described by Sparkes et al. (2006). For transient expression of fluorescent protein, leaves of 3–4-week-old *N. benthamiana* plants grown at 22°C under 12 h light/ 12 h dark were used for agroinfiltration. Plasmids were electroporated into *A. tumefaciens* GV3101 and transformants were selected on agar plates containing the required antibiotics. Transformation of *N. benthamiana* leaves was performed with *Agrobacterium* taken directly from agar plates. Cultures were re-suspended in 10 mM $MgCl_2$ (Fisher), 10 mM MES pH 5.6 (Sigma) and 150 µM acetosyringone (ACROS Oganics) to an optical density of 0.2, and incubated for 2–4 h at 28°C in the dark. Leaves were infiltrated with a needle-less syringe into the abaxial side of leaves of 3–4-week-old *N. benthamiana* plants and examined 2–3 d later. Expression levels and duration were substantially improved by co-expression of a virus-encoded suppressor of gene silencing, the p19 protein of tomato bushy stunt virus (TBSV), which prevents post-transcriptional gene silencing (PTGS) in the infiltrated tissues (Voinnet et al. 2003). For co-localization studies, mCherry-tagged peroxisomal (CD3-984), Golgi (CD3-967), and mitochondrial (CD3-991) markers were obtained from TAIR (Nelson et al. 2007) and were introduced into *A. tumefaciens* GV3101.

Yeast media and growth conditions

Media used for yeast growth was prepared as described in the Yeast Protocols Handbook (Clontech). Yeast Peptone Dextrose (YPD) and Yeast Peptone Dextrose Adenine (YPDA) were used for growth of yeast under non-selecting conditions. Synthetic dropout (SD) minimal medium was prepared with appropriate dropout solutions to obtain selection media. SD media lacking tryptophan SD -Trp was used for selection of yeast containing plasmid pGBKT7 (BD). SD lacking leucine, SD -Leu, was used for selection of yeast containing plasmid pGADT7 (AD). SD -Leu-Trp or DDO (double dropout) was used for selection of mated yeast containing both bait and prey plasmids. SD -Leu-Trp-His and SD -Leu-Trp/X-α-Gal/Aureobasidin A was used to select for yeast in interaction screening assays. Liquid cultures were grown at 30°C and 250 rpm.

Growth on solid media was done at 30°C. Library screens were allowed to grow 5 days after plating and pair-wise screens were allowed to grow 3 days after plating.

Quantitative β-galactosidase assay

The relevant pENTR223.1 clones were recombined with pDEST22 and pDEST32 destination vectors that contained the GAL4 activation domain (AD) and DNA binding domain (BD), respectively (ProQuest Two-Hybrid system; Invitrogen). These plasmids were transformed into yeast strain MaV203 (Invitrogen): *MATα; leu2-3,112; trp1-901; his3Δ200; ade2-101; cyh2R; can1R; gla4Δ; gal80Δ;GAL1::lacZ; HIS3*$_{UASGAL1}$*::HIS3LEU2; SPAL10*$_{UASGAL1}$*::URA3* following the protocol published in the Invitrogen Life Technologies, and four independent colonies from each transformation were selected to perform quantitative β-galactosidase liquid assay as described (Schrick et al. 2014). See Tables 3.8 and 3.9 for plasmid construction and primer DNA sequences.

Yeast two-hybrid assays

Table 3.8 contains a list of plasmids used for the yeast two-hybrid assay. The bait constructs were prepared as GAL4-binding domain fusions in plasmid pGBKT7 (Clontech, Mountain View, CA) and transformed into haploid yeast strain Y2H Gold (*MATa*, *trp1-901*, *leu2-3*, *112*, *ura3-52*, *his3-200*, *gal4Δ*, *gal80Δ*, *LYS2:: GAL1*$_{UAS}$*–Gal1*$_{TATA}$*–His3*, *GAL2*$_{UAS}$*–Gal2*$_{TATA}$*–Ade2*, *URA3:: MEL1*$_{UAS}$*–Mel1*$_{TATA}$, *AUR1-C MEL1*) using the LiAc method (Gietz and Woods 2002). This strain contains the auxotrophic markers *Trp1* (required for tryptophan biosynthesis) and *Leu2* (required for leucine biosynthesis), as well as the reporter genes *MEL1* (produces the enzyme α-galactosidase which can be detected by addition chromogenic substrate X-α-gal), *LacZ* (produces the enzyme β-galactosidase which can be detected by addition chromogenic substrate X-α-gal), *AbA*^{r} (confers strong resistance to toxic drug Aureobosidin A), *HIS3* (required for histidine biosynthesis), and *ADE2* (required for adenine biosynthesis). The resulting yeast strains, including an empty vector control, were selected on -Trp. The prey constructs were prepared as GAL4-activation domain fusions in plasmid pGADT7 (Clontech, Mountain View, CA) and transformed into yeast haploid strain Y187 (*MATα*, *ura3-52*, *his3-200*, *ade2-101*, *trp1-901*, *leu2-3*, *112*, *gal4Δ*, *gal80Δ*, *met–*, *URA3:: GAL1*UAS–*Gal1*TATA–

LacZ,*MEL1*). This strain contains the auxotrophic markers *Trp1* and *Leu2*, as well as the reporter genes *MEL1*, *LacZ*, *AbA*^{r}, *HIS3*, and *ADE2*. The resulting yeast strains, including an empty vector control, were selected for on -Leu. All constructs were sequenced to confirm that the binding domain or activation domain was fused N-terminally and inframe with the coding region of the appropriate locus. As a negative control, empty vectors were co-transformed with the desired construct. Bait constructs were tested for the ability to auto-activate reporter genes by plating transformed yeast harboring the desired constructs on -Leu-Trp, -Leu-Trp-His, and -Leu-Trp-His Aba X-α-Gal. Fusion protein expression was confirmed by immunoblotting with Hela-c-Myc (9E10) polyclonal antibody (Gift from Stella Lee, 1:10 dilution). All pair-wise combinations were investigated for protein-protein interaction by growing mated cultures in -Leu-Trp to a density of OD_{600}=1.0. Four serial dilutions at 10x each were prepared from the normalized cultures and spotted onto -Leu-Trp-His, -Leu-Trp-His Aba, and -Leu-Trp-His/Aba/X-α-Gal for detection of interaction and -Leu-Trp for detection of viability using a 48-pin multiplex plating tool ('Frogger', Dankar, Inc.). Plates were incubated at 30°C for 3-5 days.

Yeast two-hybrid library screening

A custom Mate & Plate Library - Universal *Arabidopsis* (Normalized) was obtained from Clontech using mRNA extracted from 11 *Arabidopsis* Colombia tissues chosen to represent a broad range of expressed genes. The library was transformed into haploid yeast strain Y187 containing >5 x 10^7 cfu/ml. The cDNA was normalized prior to library construction to reduce the copy number of abundant cDNAs derived from highly represented mRNAs, thereby increasing the representation of low copy number transcripts. The library was received in 1 ml aliquots and stored at –80°C.

Library screens to identify proteins interacting with PDF2-START and GL2-START were performed according to the Matchmaker Gold Yeast Two-Hybrid System handbook (Clontech). In brief, the haploid yeast strain containing the desired bait construct was grown in 50 ml of -Trp media overnight, centrifuged to pellet and resuspended in the residual liquid. A 1 ml aliquot of the *Arabidopsis* mate and prey library was added to 4 ml of bait culture and grown in approximately 50 ml 2x YPDA + Kanamycin (50 μg/ml) overnight at 30°C at 50 rpm for 20-22 hours. This mating mixture was centrifuged at 3000 rpm for 15 minutes and the pellet was resuspended in 10 ml 0.5x YPDA + Kan. From mating culture, four 10x serial dilutions were

prepared and spread on -Leu, -Trp, and -Leu-Trp plates to calculate mating efficiency. Remainder of the culture, 200 µl per 150 mm was plated on DDO/X/A (50–55 plates). Plates were incubated at 30°C for 3-5 days.

Yeast colonies appearing on selection plates were streaked on DDO/X/A and colonies isolated to ensure a reproducible growth phenotype. Plasmid rescue was performed by crude yeast plasmid preps carried out by the method of Robzyk and Kassir with modifications (Robzyk and Kassir 1992). Briefly, 1.5 ml yeast culture was grown overnight under selective conditions and cells were harvested in a microfuge (13,000 rpm, 5 min). After re-suspending the pellet in 150 µl of STET (1 M Sorbitol, 0.1 mM EDTA, pH 7.5), 10 µl of 2.5 mg/ml zymolyase (Seikagaku Biobusiness corporation, kind gift from Sandy Beeser) was added and cells were incubated at 37˚C with shaking for 30 minutes. Following centrifugation at 13,000 rpm for 1 min, supernatant was discarded and cells were resuspended in a mixture of 150 µl of 50 mM Tris pH 7.5, 20 mM EDTA and 15 µl of 10% SDS. Cells were mixed well by vortexing and mixture was incubated at 65˚C for 15 minutes. After 15 min, 60 µl of 5M Potassium acetate was added and mixture was incubated on ice for 15 minutes. Cells were then spun down at 13,000 rpm in a cold microcentrifuge for 10 minutes. The supernatant was carefully transferred to a fresh microfuge tube containing 200 µl of isopropanol, and DNA was recovered by centrifugation. Supernatant was discarded and pellet was washed with 70% ethanol, dried overnight, and resuspended in 50 µl of sterile water. To characterize the cDNA insert responsible for the positive yeast growth, yeast plasmid preps were transformed into *E. coli* strain DH5α to obtain enough plasmid for DNA sequencing and yeast transformation.

Western blotting

For analysis of BiFC constructs expressing N-terminal half of GFP, total proteins were extracted in hot SDS-sample buffer (100 mM Tris-HCl, pH 6.8, 2% SDS, 20% glycerol, 100 mM DTT, and 0.004% bromophenol blue, 0.48 g Urea per ml) and analyzed by immunoblotting using a Penta-His HRP conjugate (Qiagen, Valencia, CA) at 1:500 dilution and the ECL Western blotting system (GE Healthcare), according to the manufacturer's instructions.

Gene Ontology (GO) analysis for microarray data

To explore the functions of specifically expressed genes, we carried out GO enrichment analysis using the online AgriGO with Fisher's exact test (Du et al. 2010). GO terms having a Fisher's *P*-value<0.01 were considered to be enriched. As a background, *Arabidopsis* genome locus from TAIR9 was used. *P*-values were corrected using the Bonferroni method.

RNA extraction, cDNA synthesis, and quantitative real-time PCR

Arabidopsis seeds were surface-sterilized and grown on 0.8% (plant tissue culture grade) agar containing 1X Murashige and Skoog medium (Murashige and Skoog 1962), 1% Sucrose, and 0.5% MES buffer at pH 5.8. Seeds were stratified for 3 d at 4°C and then transferred to 23°C continuous light for 12 days. Total RNA was isolated using RNeasy Plant Mini Kit (Qiagen) from the shoots of Col wild type, *gl2-5*, and $gl2^{Amo}$ mutant seedlings. The RNA was treated on-column with RNase-Free DNase (Qiagen), and 0.5 g of RNA was used as a template for cDNA synthesis. cDNA synthesis was performed using GoScript Reverse Transcriptase (Promega) according to the manufacturer's protocol. Real- time PCR was performed using IQ SYBR Green Supermix with the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Primers were designed to amplify 90- to 150-bp DNA fragments Each reaction contained 10 µl of 2x SYBR Green Supermix, 1 µl of forward and reverse 10 mM gene-specific primers, and 5 µl of cDNA (diluted 5-fold) in 20 µl. Standard curves were generated from 10-fold dilutions of amplicons for each primer pair. *ACT7* served as the reference gene. The details of oligonucleotides for PCR are described in Table 3.9

ChIP-seq Assays

Transgenic seeds containing EYFP tagged *GL2*, *gl2*Δ*START*, *gl2*Δ*Zip*, $gl2^{Amo}$, *gl2-5*, and $gl2^{R392M;E375G}$ were surface sterilized and grown on 0.8% (plant tissue culture grade) agar containing 1X Murashige and Skoog medium (Murashige and Skoog 1962) and 1% Sucrose. Seeds were stratified for 3 to 5 d at 4°C and then transferred to 23°C and continuous light for 10 days. Seedlings were harvested and cross-linked with 1% formaldehyde for 10 min in a vacuum. Glycine was added to a final concentration of 0.125 M, and the reaction was terminated by incubation for 5 min in a vacuum. Chromatin was fragmented using covaris S220 ultrasonicator

(Duty Cycle: 10%; Intensity: 5; Cycles per Burst: 200) to produce <500 bp fragments. ChIP experiment was performed as previously described (Khosla et al. 2014). Immunoprecipitated DNA was sequenced using Illumina sequencing technology. ChIP DNA was processed, prepared with 6-fold multiplexed Tru-Seq ChiP-seq libraries, and sequenced using short- read Illumina technology at the Kansas University Medical Center (KUMC) Genome Sequencing Facility. Libraries were sequenced (50 cycles) using the TruSeq Single Read Clustering Kit v3 and TruSeq SBS-HS v3 sequencing chemistry with index read.

Bioinformatic Analysis

ChIP-Sequencing was performed in an illumina HiSeq 2500 sequencing machine (Illumina, San Diego, CA) at a 50 base single read resolution. Read quality was assessed using FastQC. Aligned read count statistics were obtained using the "flagstat" utility of the SAMtools suite (Li et al. 2009). Sequences were mapped to the *Arabidopsis thaliana* genome (TAIR10.24) using Bowtie2 (Langmead and Salzberg 2012); parameters: -q --phred33 --end-to-end --sensitive -p 16 -t). Replicate aligned ChIP samples were merged (using "view" utility of the SAMtools suite) for greater coverage and analyzed along with their corresponding input samples for significant peaks using MACS (Zhang et al. 2008); parameters: -f BAM -g 0.135e9). Enriched motifs around a 500bp region of significant peaks were detected using HOMER software suite (Heinz et al. 2010); program: findMotifsGenome.pl; parameters: genome = tair10, size = 500). The HOMER software suite was also used to perform peak annotations (program: annotatePeaks.pl; parameters: genome = tair10).

Target Gene Ontology Analysis

AgriGO (Du et al. 2010) was employed to extract function annotations for genes whose promoter or intergenic regions are bound by the transcription factors. In the present analysis the tool SEA (Singular Enrichment Analysis) was used. SEA determines gene ontology (GO) term enrichment in one group of genes by comparing it with a reference group of genes (Du et al. 2010). As a background, the *Arabidopsis* genome locus from TAIR9 was used. The analysis was

done using Fisher's exact test, and a P-value cut-off of 0.01. *P*-values were corrected using the Bonferroni method.

FIGURES AND TABLES

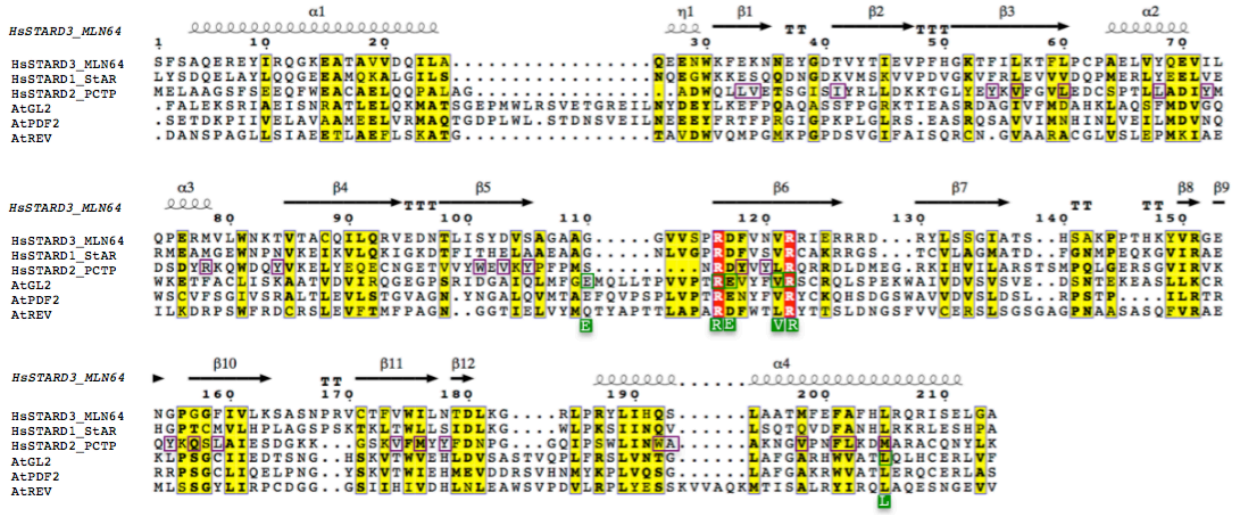


Figure 3.1 GL2 missense mutations in the START domain that are predicted to alter ligand binding.

Sequence alignment of START domains from humans (Hs) and *Arabidopsis* (At). Consensus secondary structure features are displayed with ESPript/ENDscript. GL2 residues targeted for mutagenesis (green) include: E375, R386, E387, V391, R392, and L480. Similar residues are boxed in yellow, while conserved residues are boxed in red. Ligand contact points (purple), as derived from the PCTP-PC co-crystal (Roderick et al. 2002), are shown for PCTP. STARD1StAR, STARD3/MLN64, and STARD2/PCTP represent mammalian START domain proteins. GL2 and PDF2 are members of Class IV HD-Zip family. Class III HD-Zip family member, REVOLUTA (REV) was also included for comparison.

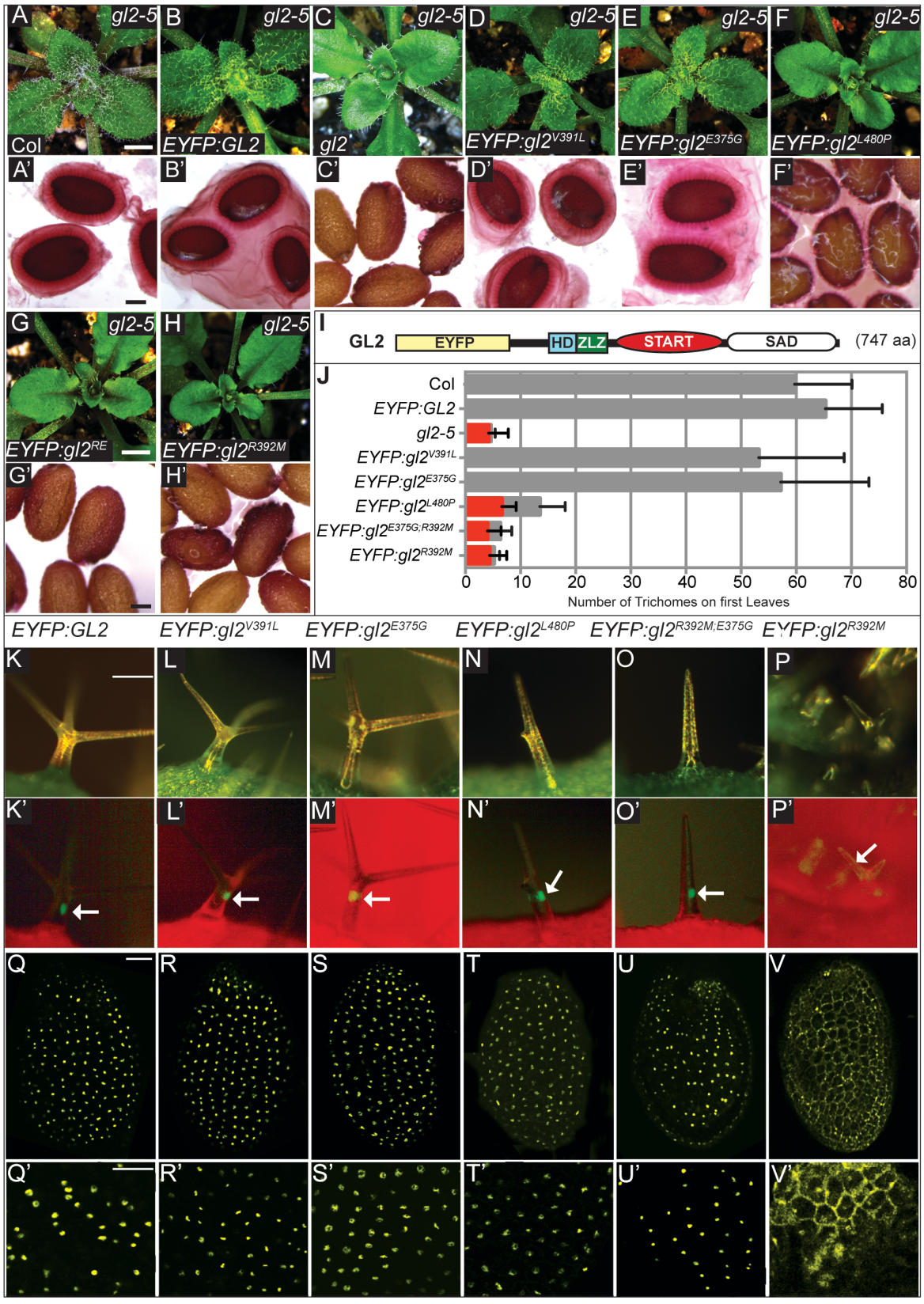

A gl2-5
Col
B gl2-5
EYFP:GL2
C gl2-5
gl2-5
D gl2-5
EYFP:gl2V391L
E gl2-5
EYFP:gl2E375G
F gl2-5
EYFP:gl2L480P
A'
B'
C'
D'
E'
F'
G gl2-5
EYFP:gl2RE
H gl2-5
EYFP:gl2R392M
G'
H'
I
GL2
EYFP
HD ZLZ
START
SAD
(747 aa)
J
Col
EYFP:GL2
gl2-5
EYFP:gl2V391L
EYFP:gl2E375G
EYFP:gl2L480P
EYFP:gl2E375G;R392M
EYFP:gl2R392M
0 10 20 30 40 50 60 70 80
Number of Trichomes on first Leaves
EYFP:GL2
EYFP:gl2V391L
EYFP:gl2E375G
EYFP:gl2L480P
EYFP:gl2R392M;E375G
EYFP:gl2R392M
K L M N O P
K' L' M' N' O' P'
Q R S T U V
Q' R' S' T' U' V'

Figure 3.2 Phenotypes and subcellular localization of GL2-START domain mutants in trichomes, seeds, and ovules.

(A-H) Rosettes exhibiting leaf trichomes. Scale bar = 2 mm. **(A'-H')** Seeds were stained with ruthenium red to detect mucilage release upon imbibition. Scale bar = 200 μm. **(A)** Col wild-type plants display trichomes covering leaf surfaces, **(A')** Normal outer mucilage was observed for Col wild-type seeds, (**B, B'**) *EYFP:GL2,* (**D, D'**) *EYFP:gl2*^{V391L}, and (**E, E'**) *EYFP:gl2*^{E375G} display complete rescue of the mutant phenotype while (**F, F'**) *EYFP:gl2*^{L480P} exhibits partial rescue. The phenotypes of (**G, G'**) *EYFP:gl2*$^{E375G;R392M}$ and (**H, H'**) *EYFP:gl2*^{R392M} are indistinguishable from *gl2* null. **(C,C')** *gl2-5* plants appear glabrous and seeds lack mucilage. **(I)** Schematic representation of the EYFP:GL2 protein depicting epicting enhanced yellow fluorescence protein (EYFP), homeodomain (HD), Zip-Loop-Zipper (ZLZ), START, and START-associated domain (SAD). **(J)** Number of trichomes on first leaves. Grey bars show the number of branched trichomes and red bars indicate the number of unbranched single-spiked trichomes. Error bars indicate standard deviations for n≥ 20 plants.
(K-P) and **(K'-P')** Live imaging of leaf trichomes indicates nuclear localization of **(K, K')** *EYFP:GL2* as well as GL2-START mutants **(L, L')** *EYFP:gl2*^{V391L}, **(M, M')** *EYFP:gl2*^{E375G} **(N, N')** *EYFP:gl2*^{L480P}, and **(O, O')** *EYFP:gl2*$^{E375G;R392M}$. **(P, P')** START mutants with R392M mutation exhibit diffuse expression in trichomes. Matching white light (top; green, chlorophyll) and fluorescence (bottom; red, chlorophyll) images are shown for each construct. Scale bar = 100 μm. **(Q-V)** Confocal laser scanning microscopy of *gl2* ovules indicates nuclear localization of **(Q)** *EYFP:GL2*, **(R)** *EYFP: gl2*^{V391L} , and **(S)** *EYFP:gl2*^{E375G} **(T, U)** *EYFP:gl2*^{L480P} and *EYFP:gl2*$^{E375G;R392M}$ expression is nuclear but disperse compared with wild-type **(V)** EYFP fusion to gl2^{R392M} displays non-nuclear expression of protein. Scale bar = 50 μm
(Q'-V') Magnified view of localization patterns. Scale bar = 100 μm

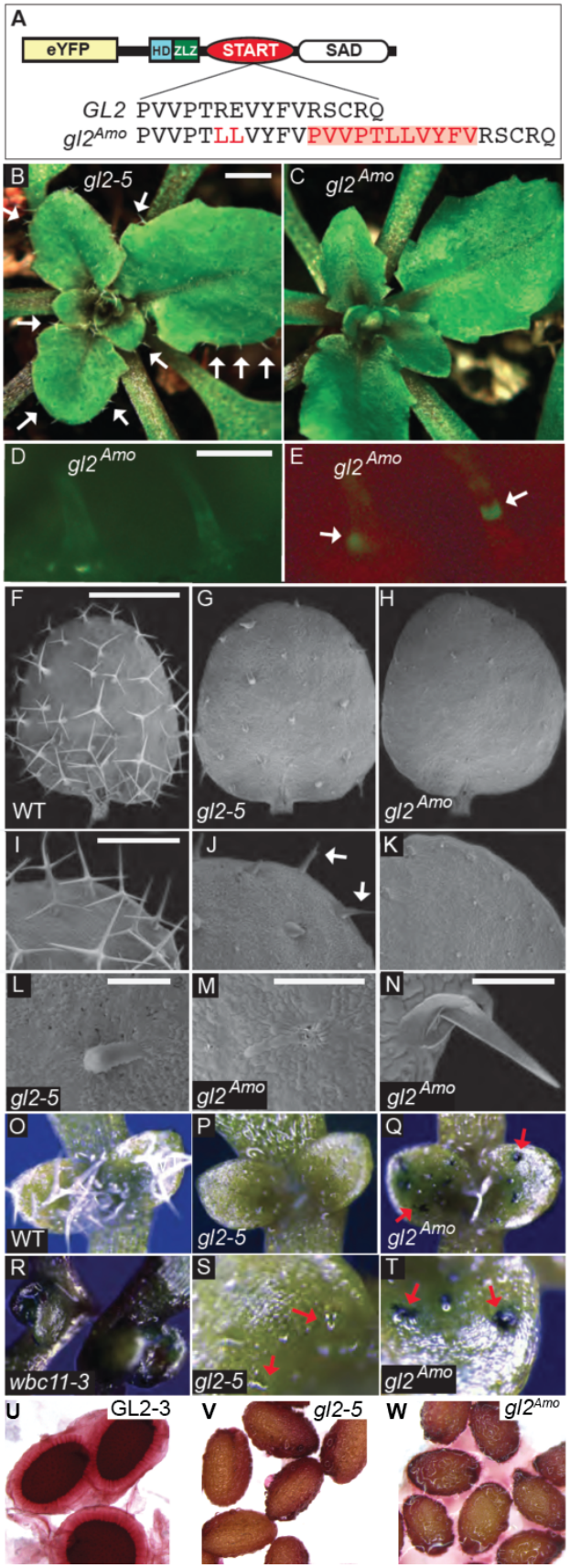

A
eYFP
HD
ZLZ
START
SAD
GL2 PVVPTREVYFVRSCRQ
gl2Amo PVVPTLLVYFVPVVPTLLVYFVRSCRQ
B gl2-5
C gl2Amo
D gl2Amo
E gl2Amo
F WT
G gl2-5
H gl2Amo
I
J
K
L gl2-5
M gl2Amo
N gl2Amo
O WT
P gl2-5
Q gl2Amo
R wbc11-3
S gl2-5
T gl2Amo
U GL2-3
V gl2-5
W gl2Amo

Figure 3.3 Novel antimorphic allele of GL2.

(A) Molecular changes in *gl2*^{Amo}. **(B-C)** Rosettes from **(B)** *gl2-5* null allele and **(C)** *gl2*^{Amo} antimorphic allele. While *gl2-5* leaves exhibit trichomes at the leaf margins (arrows), *gl2*^{Amo} leaves appear to lack trichomes at the macroscopic level. **(D, E)** Matching white light and fluorescence images indicate nuclear localization of EYFP:gl2^{Amo}. SEM images of first leaves of **(F,I)** Col wild-type, **(G,J,L)** *gl2-5* null allele, and **(H,K,M,N)** *gl2*^{Amo} dominant negative allele. *gl2-5* plants exhibit trichomes at leaf margins (arrows) while *gl2*^{Amo} plants do not. **(N)** *gl2*^{Amo} trichomes are thinner and more fragile than in *gl2-5*. Scale bars = 2 mm **(B,C)**; 100 μm. **(D, E)**; 1 mm **(F-H)**; 500microm **(I-K)**; 200 μm **(L-M)**; and. **(O-T)** Toluidine blue staining of **(O)** Col wild-type, **(P,S)** *gl2-5*, **(Q,T)** *gl2*^{Amo}, and **(R)** *wbc11-3,* an allele of *ABCG11* . Wild-type leaves and trichomes fail to take up the dye, while *wbc11-3* mutants that display a defective cuticle turn blue in presence of the dye. The trichomes from *gl2*^{Amo} but not from *gl2-5* take up the dye, suggesting a defect in cuticular lipids on the trichomes. Arrows mark selected trichomes for comparison. **(U-W)** Seeds were stained with ruthenium red to detect mucilage production upon imbibition. Scale bar = 200 μm. **(W)** *gl2* mutants lack mucilage layer. By contrast, **(V)** *gl2*^{Amo} is partially defective in seed coat mucilage production.

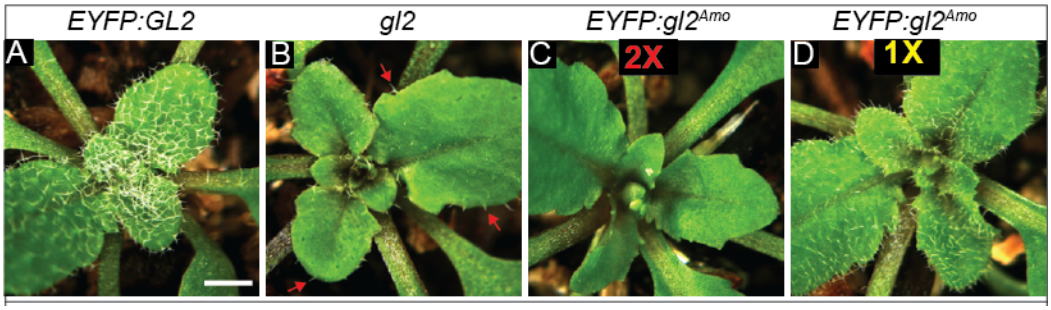


Figure 3.4 gl2^{Amo} phenotype is sensitive to copy number of the transgene.
Rosettes expressing trichome phenotypes. **(A)** *gl2* mutants expressing wild-type *EYFP:GL2* transgene display numerous trichomes. **(B)** The *gl2* null mutant exhibits trichomes mostly at leaf margins (arrows), while **(C)** two copies of *gl2^{Amo}* transgene results in a trichome-less-phenotype. Dosage of the transgene is critical, as **(D)** one copy results in partial trichome formation.

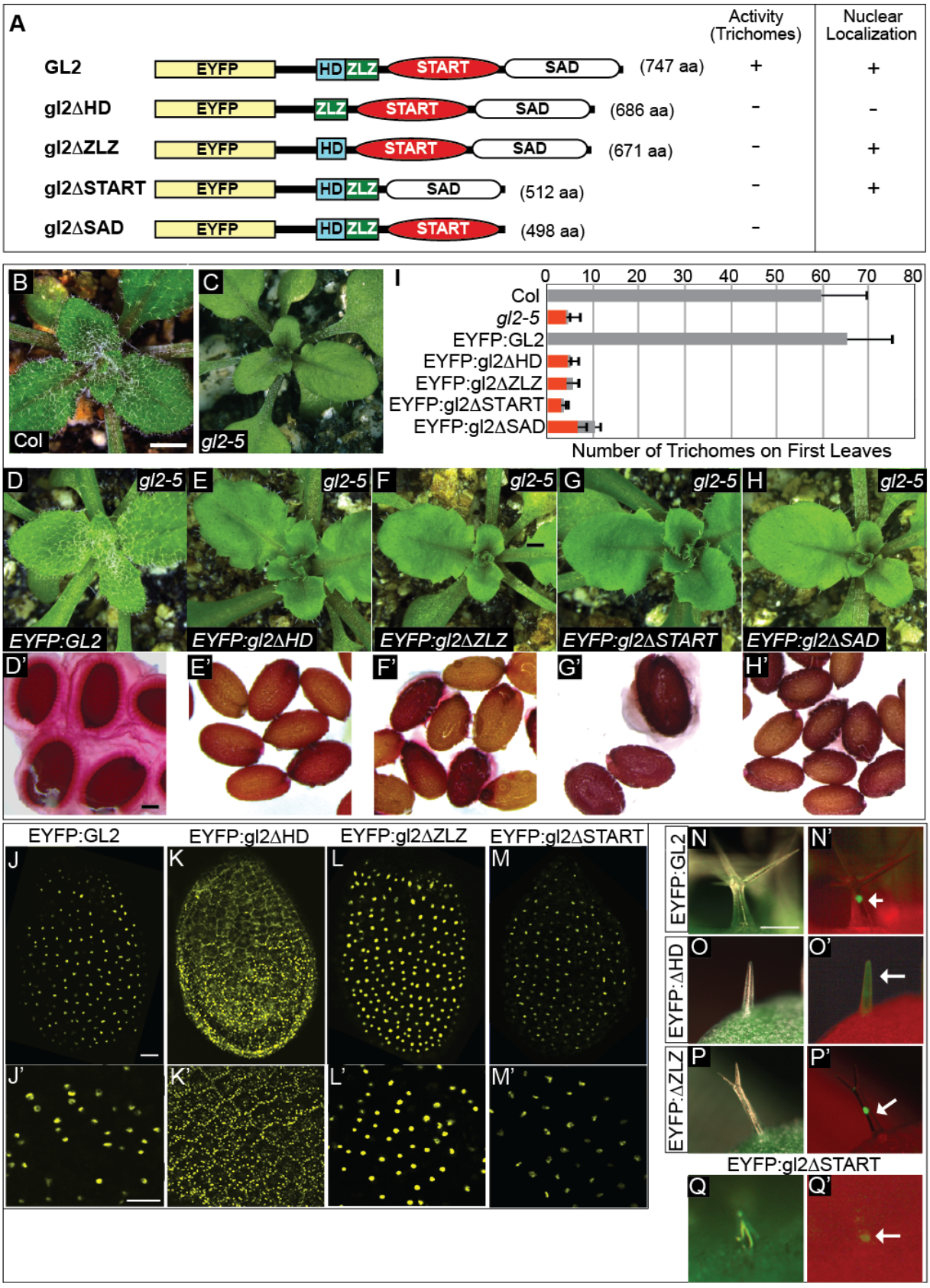

A
Activity (Trichomes)
Nuclear Localization
GL2
EYFP
HD
ZLZ
START
SAD
(747 aa)
+
+
gl2ΔHD
(686 aa)
-
-
gl2ΔZLZ
(671 aa)
-
+
gl2ΔSTART
(512 aa)
-
+
gl2ΔSAD
(498 aa)
-
B
Col
C
gl2-5
I
0 10 20 30 40 50 60 70 80
Col
gl2-5
EYFP:GL2
EYFP:gl2ΔHD
EYFP:gl2ΔZLZ
EYFP:gl2ΔSTART
EYFP:gl2ΔSAD
Number of Trichomes on First Leaves
D gl2-5
EYFP:GL2
E gl2-5
EYFP:gl2ΔHD
F gl2-5
EYFP:gl2ΔZLZ
G gl2-5
EYFP:gl2ΔSTART
H gl2-5
EYFP:gl2ΔSAD
D'
E'
F'
G'
H'
EYFP:GL2
EYFP:gl2ΔHD
EYFP:gl2ΔZLZ
EYFP:gl2ΔSTART
J
K
L
M
J'
K'
L'
M'
EYFP:GL2
N
N'
EYFP:ΔHD
O
O'
EYFP:ΔZLZ
P
P'
EYFP:gl2ΔSTART
Q
Q'

Figure 3.5 Sub-cellular localization and complementation of *gl2* mutant phenotypes by class IV HD-Zip functional domains: gl2ΔSTART, gl2ΔHD, and gl2ΔZip.

(A) A schematic representation of GL2 domain deletion fragments depicting EYFP, homeodomain (HD), Zipper-Loop-Zipper (ZLZ) and START-associated domain (SAD)**. (B-H)** Rosettes exhibiting leaf trichomes. Scale bar = 2 mm. **(D'-H')** Seeds were stained with ruthenium red to detect mucilage production upon imbibition. Scale bar = 200 μm. (**D, D'**) *EYFP:GL2* displays complete rescue of the mutant phenotype in trichomes and seeds. (**E-H and E'-H'**) The phenotypes of *EYFP:gl2*Δ*HD*, *EYFP:gl2*Δ*ZLZ*, *EYFP:gl2*Δ*START*, and *EYFP:gl2*Δ*SAD* are similar to *gl2* null mutant. **(I)** Quantification of trichomes on first leaves. Grey bars indicate branched trichomes, while the red bars show the number of unbranched single-spiked trichomes. Error bars indicate standard deviations for n≥ 20 plants. **(J-M)** *gl2-5* ovules expressing indicated transgenes. Scale bar = 50 μm**. (J'-M')** Magnified view to more clearly show localization patterns. Scale bar = 100 μm**.** Confocal laser scanning microscopy images indicate nuclear localization of **(J, J'**) EYFP:GL2. **(K, K')** EYFP fusion to gl2ΔHD displays non-nuclear expression of the protein with punctate dots. **(L, L')** gl2ΔZip protein displays nuclear localization similar to that of wild-type GL2, except that it is much stronger. **(M, M')** EYFP:gl2ΔSTART displays weak nuclear localization. (**N-Q, N'-Q'**) Matching white light and fluorescence images. Arrows indicate protein localization. Scale bar = 100 μm.

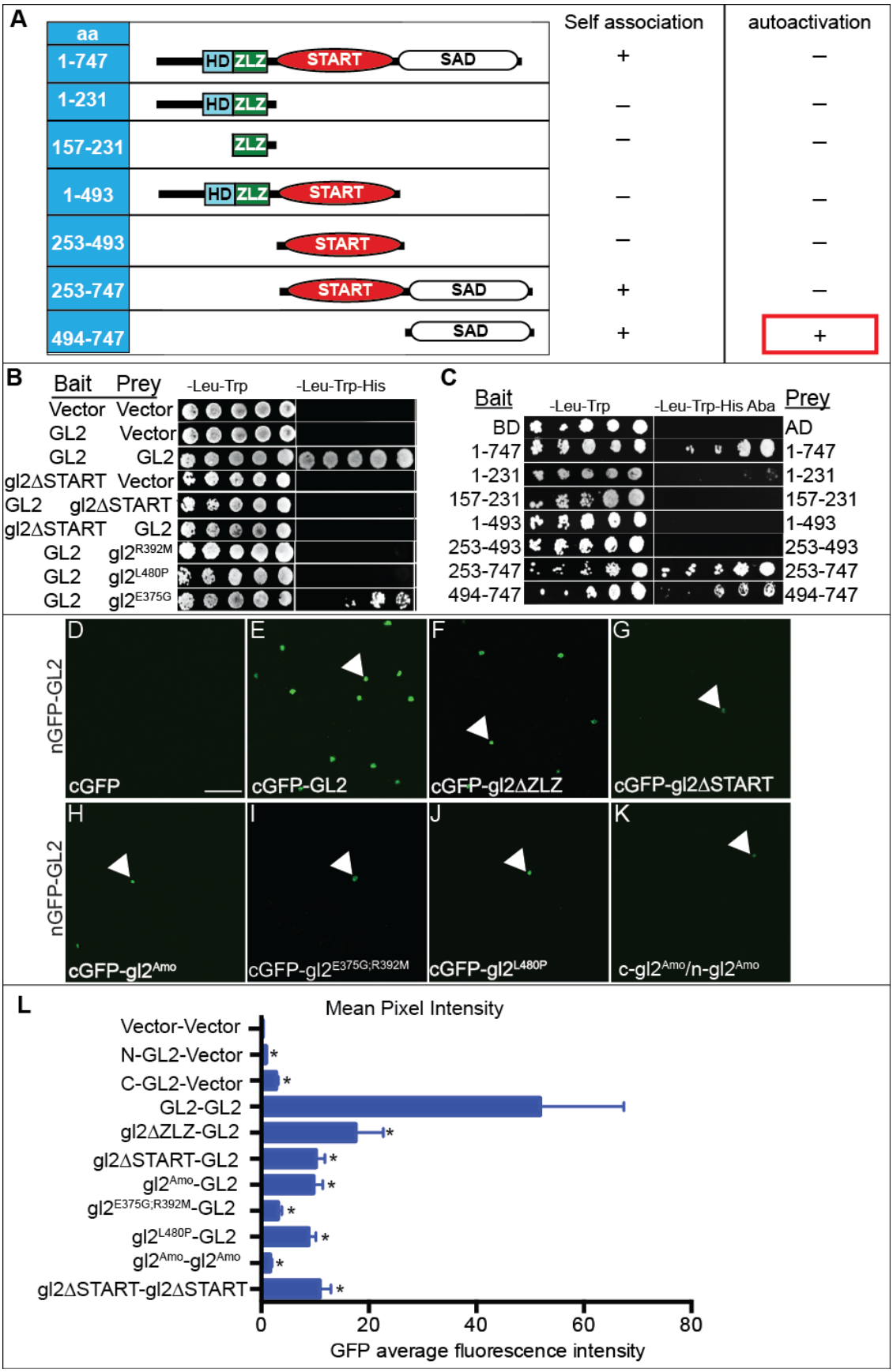

A
aa
Self association
autoactivation
1-747
HD ZLZ START SAD
+
–
1-231
HD ZLZ
–
–
157-231
ZLZ
–
–
1-493
HD ZLZ START
–
–
253-493
START
–
–
253-747
START SAD
+
–
494-747
SAD
+
+
B
Bait Prey
-Leu-Trp
-Leu-Trp-His
Vector Vector
GL2 Vector
GL2 GL2
gl2ΔSTART Vector
GL2 gl2ΔSTART
gl2ΔSTART GL2
GL2 gl2R392M
GL2 gl2L480P
GL2 gl2E375G
C
Bait
-Leu-Trp
-Leu-Trp-His Aba
Prey
BD AD
1-747 1-747
1-231 1-231
157-231 157-231
1-493 1-493
253-493 253-493
253-747 253-747
494-747 494-747
D
E
F
G
nGFP-GL2
cGFP
cGFP-GL2
cGFP-gl2ΔZLZ
cGFP-gl2ΔSTART
H
I
J
K
nGFP-GL2
cGFP-gl2Amo
cGFP-gl2E375G;R392M
cGFP-gl2L480P
c-gl2Amo/n-gl2Amo
L
Mean Pixel Intensity
Vector-Vector
N-GL2-Vector
C-GL2-Vector
GL2-GL2
gl2ΔZLZ-GL2
gl2ΔSTART-GL2
gl2Amo-GL2
gl2E375G;R392M-GL2
gl2L480P-GL2
gl2Amo-gl2Amo
gl2ΔSTART-gl2ΔSTART
0
20
40
60
80
GFP average fluorescence intensity

Figure 3.6 START and ZLZ domain are both required for GL2 homodimerization.

(A) A schematic representation of GL2 constructs used. The DNA binding homeodomain (HD), leucine zipper dimerization (ZLZ), START, and START associated (SAD) domains are indicated. **(B)** GL2 dimerization can support growth on –Leu-Trp-His and –Leu-Trp-His/ Aureobasidin A (Aba) /X-α-Gal media. Deletion and missense mutations in the START domain disrupt homodimer formation. GL2, gl2ΔSTART and the indicated GL2-START missense mutants were cloned into bait (pGBKT7) and prey (pGADT7) vectors. The constructs were co-transformed into yeast strain Y2HGold. Each transformed yeast strain was assayed on the indicated growth media. Empty prey and bait vectors in above shown combinations served as negative controls. **(C)** Homodimerization of GL2 in Y2H assays using complete and partial protein fragments. BD, pGBKT7; AD, pGADT7. **(D-K)** BiFC assays in tobacco leaf abaxial (lower) epidermal cells transiently transformed with the indicated constructs and the silencing suppressor p19 (see methods). Interaction of full length GL2 tagged with split GFP at the N-terminus (nGFP-GL2) with **(D)** vector, **(E)** full-length GL2, **(F)** GL2 with leucine zipper domain (ZLZ) deletion, **(G)** GL2 with START domain deletion, and **(H-J)** START domain missense mutants tagged with the split GFP at C-terminus (cGFP). **(K)** Detection of $gl2^{Amo}$ self-association. Nuclear expression is indicated by white arrows. Scale bar = 50 μm. **(L)** Quantitative analysis. Mean pixel intensities from 10 images are shown. Error bars indicate standard deviations for two independent transformants in two trials. Asterisks indicate statistically significant differences (*, $P < 0.0001$ by *t* test).

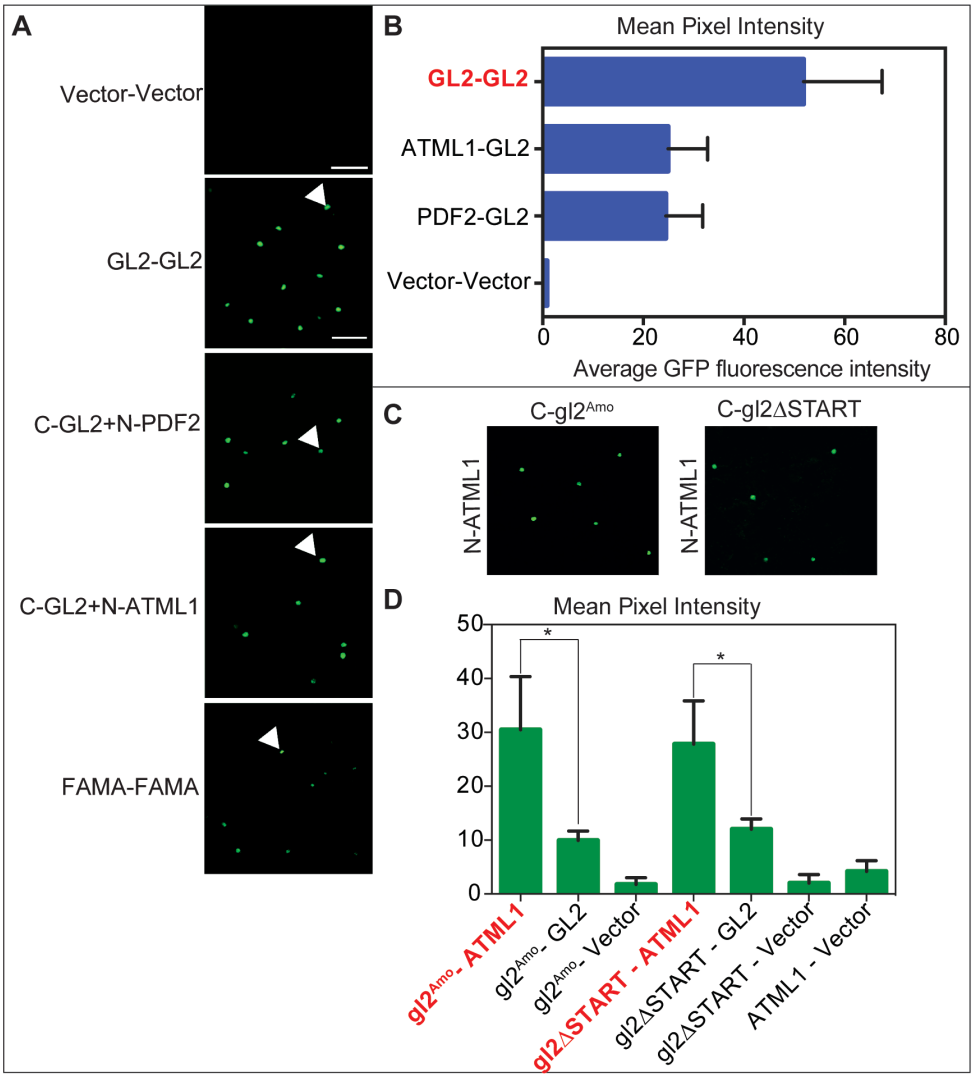


Figure 3.7 GL2 heterodimerizes with two class IV HD-Zip gene family members. **(A and C)** *GL2, ATML1, PDF2, gl2^{Amo},* and *gl2ΔSTART* CDS were fused to the N-GFP (N) or C-GFP (C) CDS and co-expressed in *N. benthamiana* epidermal cells as indicated. Negative control = Empty vector. The documented interaction between basic helix-loop-helix (bHLH) transcription factor FAMA proteins in a homodimer (Kanaoka et al. 2008) was used as a positive control. White arrows indicate nuclei. Scale bar indicates 50 µm. **(B and D)** Quantitative analysis. Mean pixel intensities from 10 images are shown. Error bars indicate standard deviations for two independent transformants in two trials. Asterisks indicate statistically significant differences (*, $P < 0.0001$ by *t* test).

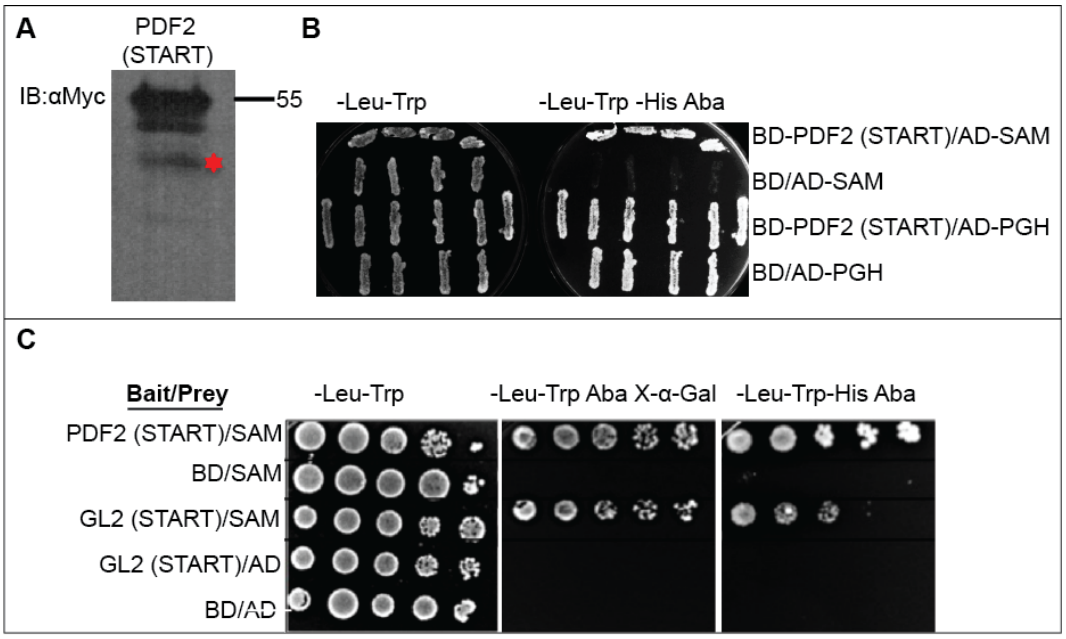


Figure 3.8 Pairwise yeast two-hybrid assay for the confirmation of interaction between PDF2 (START) and selected proteins retrieved in a Y2H screen.

(A) Western blot of fusions of PDF2 (START) full-length to the GAL4 DNA Binding Domain in the pGBK-T7 Bait vector. Asterisk indicates non-specific band**. (B)** Yeast transformants harboring pGBKT7 (BD) and pGADT7 (AD) were grown on -Leu-Trp indicating the presence of each plasmid as well as -Leu-Trp-His Aba indicating interaction between the expressed proteins). Both of the tested proteins (At5g40830; SAM, At2g36530; 2-phospho-D-glycerate hydrolase (PGH) appear to interact with PDF2-START. PGH also interacts with empty bait plasmid (BD). **(C)** Yeast transformants harboring various combinations of bait and prey plasmids as indicated on the left column were grown on -Leu-Trp, -Leu-Trp Aba X-α-Gal, and -Leu-Trp-His Aba. Growth is observed for yeast harboring SAM and either PDF2-START or GL2-START compared with the negative control yeast harboring the empty bait (BD) and prey (AD) plasmid combinations.

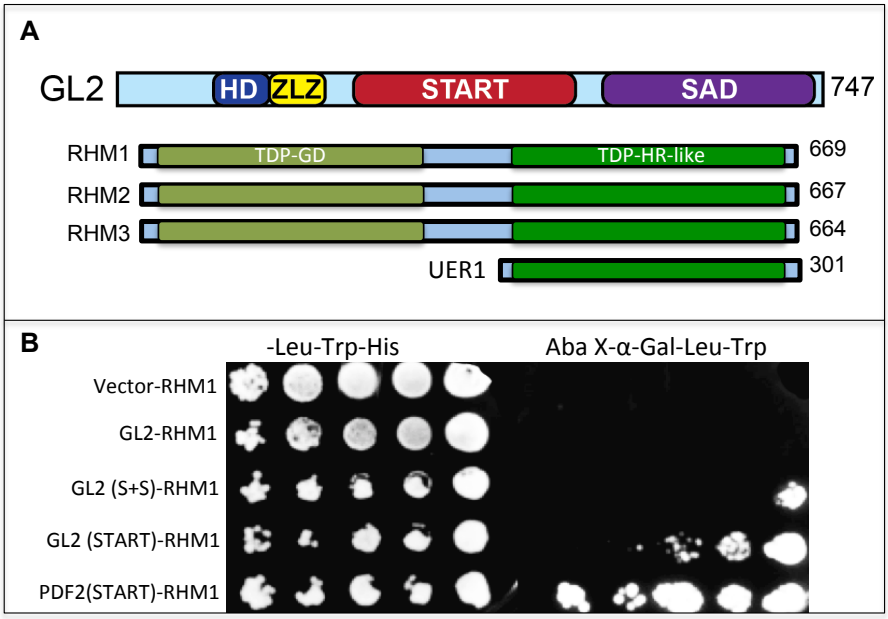


Figure 3.9 Interactions between RHM1 and class IV HD-Zip transcription factors in yeast.

(A) Domain configurations of GL2, RHM1 and other rhamnose synthase family members. The DNA binding homeodomain (HD), leucine zipper dimerization (ZLZ), START, START-Adjacent (SAD), dTDP-glucose 4,6-dehydratase (TDP-GD) catalyzes the dehydration of UDP-glucose to form UDP-4-dehydro-6-deoxy-D-glucose, and dTDP-4-dehydrorhamnose reductase (TDP-HR-like) involved in epimerization of the C3' and C5' positions of UDP-4-dehydro-6-deoxy-D-glucose to form UDP-4-keto-beta-L-rhamnose and the reduction of UDP-4-keto-beta-L-rhamnose to yield UDP-beta-L-rhamnose. **(B)** Y2H assay for the interaction between RHM1 and indicated bait proteins. Yeast harboring GL2 (START+SAD)/RHM1, GL2 (START)/RHM1, and PDF2 (START)/RHM1 grow on selective plates as indicated.

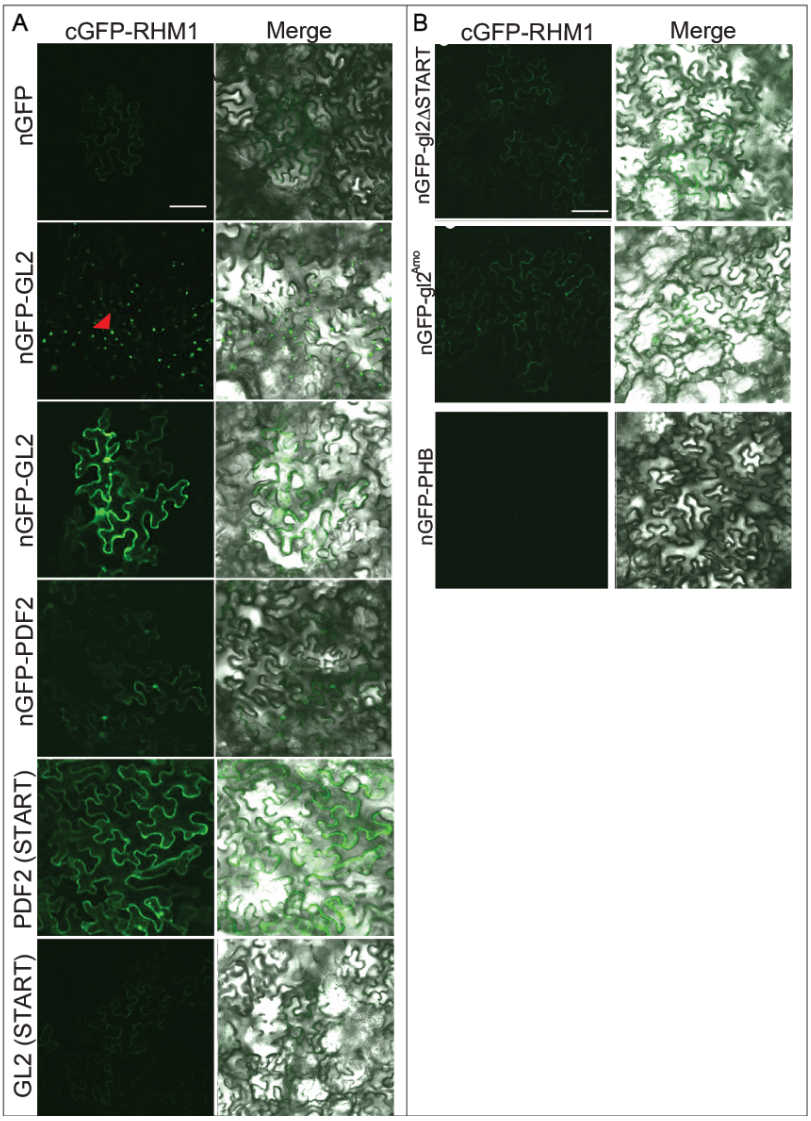


Figure 3.10 RHM1 and class IV HD-Zip transcription factors associate *in planta.*

(A) BiFC of GFP in transiently transformed *N.benthamiana* leaf epidermal cells, Left column, GFP fluorescent signal detection by confocal microscopy; right column, merge of fluorescent signal and differential interference contrast (DIC) microscopy. RHM1 fused to C-terminal half of GFP (cGFP) is co-expressed with interaction partners fused to the N-terminal part of GFP (nGFP). GL2/RHM1 interaction marks cytoplasm and also shows punctate pattern of GFP fluorescence (red arrowhead- see text for details). The negative control consisting of empty vector (cGFP) with RHM1 fusion protein (nGFP-RHM1) shows negligible fluorescence. **(B)** Representative confocal images of the BiFC analysis. Pairwise combinations of constructs, each fused with a complementary, half-GFP molecule (left, nGFP fusion; top, cGFP fusion to RHM1). Both DIC (left panel) and merged (right panel) images are shown. Scale bars = 50 μm.

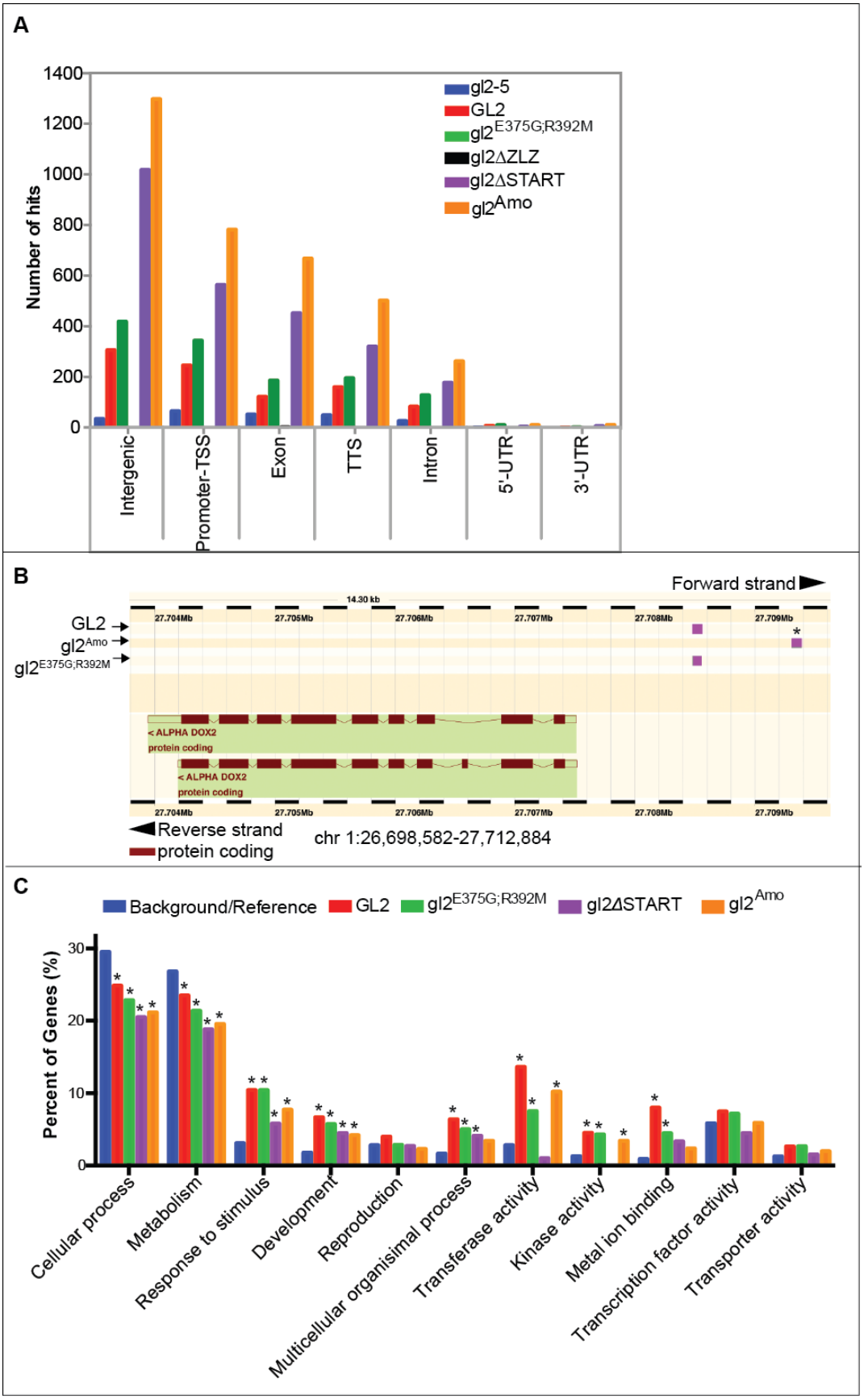

A
Number of hits
1400
1200
1000
800
600
400
200
0
gl2-5
GL2
gl2E375G;R392M
gl2ΔZLZ
gl2ΔSTART
gl2Amo
Intergenic
Promoter-TSS
Exon
TTS
Intron
5'-UTR
3'-UTR
B
Forward strand
14.30 kb
27.704Mb
27.705Mb
27.706Mb
27.707Mb
27.708Mb
27.709Mb
GL2
gl2Amo
gl2E375G;R392M
< ALPHA DOX2
protein coding
< ALPHA DOX2
protein coding
Reverse strand
protein coding
chr 1:26,698,582-27,712,884
C
Background/Reference
GL2
gl2E375G;R392M
gl2ΔSTART
gl2Amo
Percent of Genes (%)
30
20
10
0
Cellular process
Metabolism
Response to stimulus
Development
Reproduction
Multicellular organisimal process
Transferase activity
Kinase activity
Metal ion binding
Transcription factor activity
Transporter activity

Figure 3.11 Deletion of START domain does not interfere with GL2 DNA binding *in vivo.*

(A) Genomic distribution of the binding events for wild-type GL2 and its mutant versions as indicated (FDR ≤0.1). The null mutant *gl2-5* serves as a negative control. TTS, transcription termination site; TSS, transcription start site; UTR, untranslated region. **(B)** Binding profile of $gl2^{Amo}$ to *α-DOX2* genomic region was obtained using *Arabdiposi*s genome browser available from Ensembl Plants database. **(C)** Functional classification of the transcription factor target genes. The bar chart depicts over-represented GO terms obtained from the list of target genes whose promoters are bound by wild-type GL2 and GL2-START mutants; gl2ΔSTART, $gl2^{Amo}$, and $gl2^{E375G;R392M}$. These genes were identified by ChIP-seq and classified based on their molecular function and biological process. Note a given gene may be represented in more than one category. Enrichment analysis was performed using AgriGO (Du et al. 2010). Percentage of each category is compared with the *Arabidopsis* genome locus (TAIR9) background (blue). Asterisks indicate significantly different categories (Fisher's exact test, $P < 0.0001$).

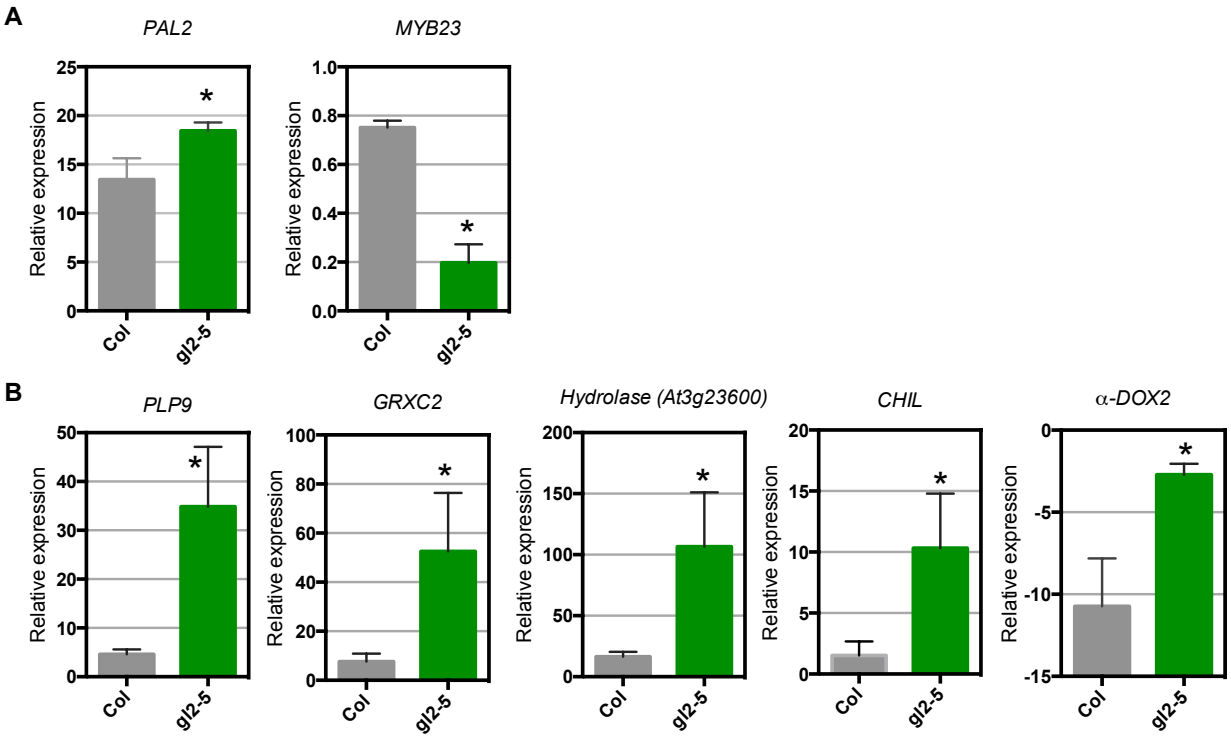


Figure 3.12 qRT-PCR showing expression of candidate GL2 target genes.

(A) qRT-PCR was performed on shoots from wild-type Col and *gl2-5* plants grown under continuous white light for 12 days. *ACT7* was used as the reference gene. Error bars indicate SD of triplicate experiments. Significant differences (*, $P < 0.05$ by unpaired *t*-test) from the wild type (Col) are indicated by asterisks. **(B)** Real time qRT-PCR analysis on roots from 5 day old grown plants. *ACT7* was used as a quantitative control. Transcript levels of *DOX2* were severely attenuated in roots of the *gl2-5* mutants compared with the wild-type Col plants and the scale of Y-axis is presented in Log_2Scale. Error bars represent SD of four biological replicates. Significant differences (*, $P < 0.0001$ by *t*-test) from the wild type (Col) are indicated by asterisks.

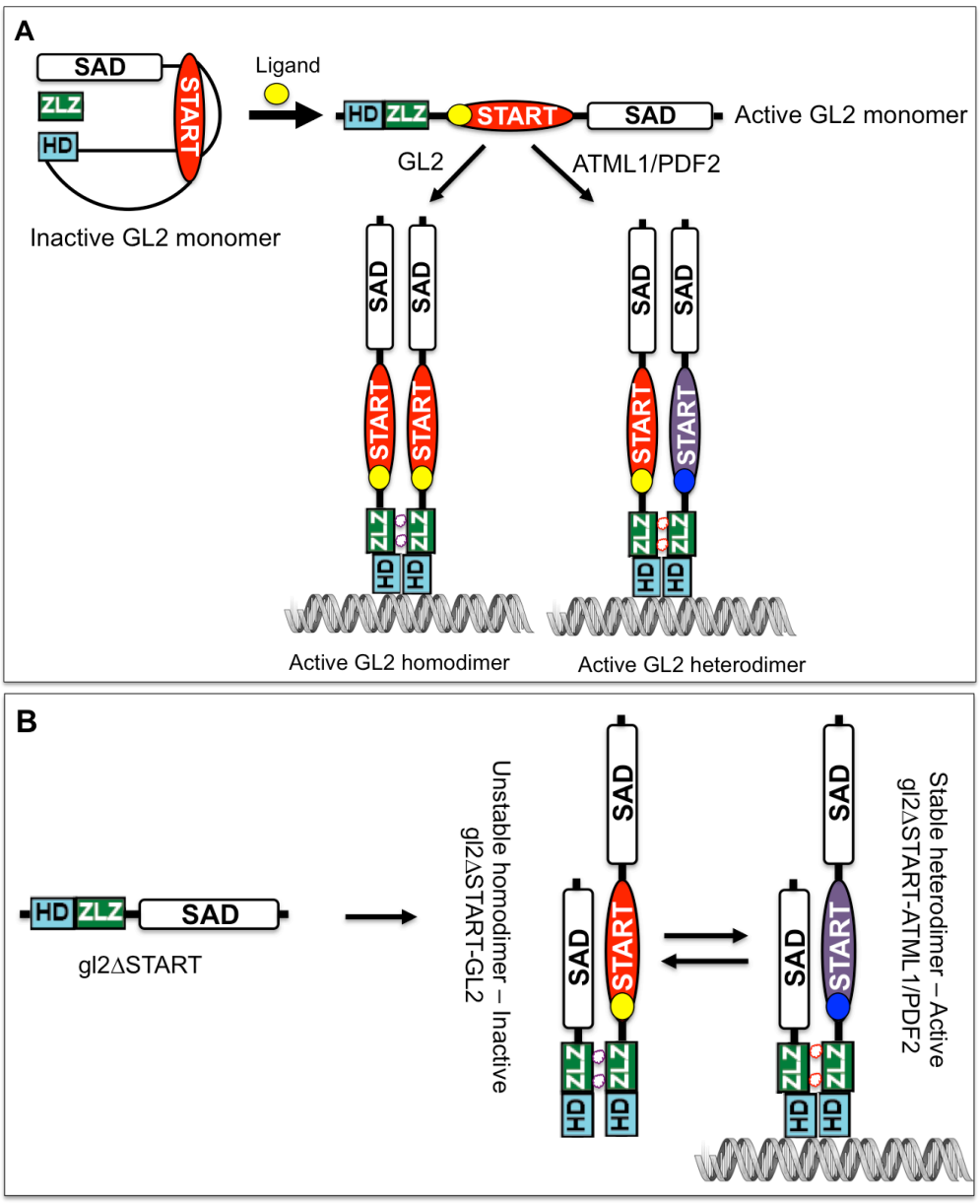


Figure 3.13 Model for regulation of GL2 dimerization by the START domain.

(A) Upon ligand binding to the START domain, the conformation of the GL2 is changed in such a way that the "hidden" ZLZ domain is exposed, leading to formation of active GL2 monomer. **(B)** In the absence of the ligand binding START domain, GL2 forms ill-defined unstable homodimers. Instability of homodimers provides a thermodynamic driving force for preferential stable heterodimer formation in the presence of either binding partner (ATML1/PDF2). The star shapes indicate different dimerization interfaces used for homodimerization (purple) and heterodimerization (red).

Table 3-1 Trichome quantification on the first two leaves. See figures for graphical representations of the data. Standard deviations for n≥20 plants are shown in parentheses.

Genotype	Ave. # trichomes 1 branch	Ave. # trichomes 2-4 branch	Ave. # total trichomes
Col wild type	0	59.7 (10.4)	59.7
GL2	0	65.3 (10.3)	65.3
gl2-5	4.2 (3.6)	0.4 (0.8)	4.6
gl2ΔZLZ	4.3 (3.3)	1.4 (1.7)	5.7
gl2ΔSTART	3.1 (2.2)	0.6 (0.8)	3.7
gl2ΔHD	4.6 (3.0)	0.6 (0.7)	5.2
gl2ΔSAD	6.7 (2.6)	3.8 (1.6)	10.5
$gl2^{L480P}$	6.6 (2.6)	6.9 (4.6)	13.5
$gl2^{R392M}$	4.4 (3.1)	0.8 (1)	5.2
$gl2^{E375G;R392M}$	4 (2.5)	2.3 (2.1)	6.3
$gl2^{E375G}$	0	57.2 (16)	57.2
$gl2^{V391L}$	0	53.2 (15.5)	53.2
$gl2^{Amo}$	0.5 (1.1)	0	0.5
$gl2^{R386L;E387L}$	0	31.9 (12)	31.9

Table 3-2 Genes down-regulated in *gl2*^{Amo} organized by selected GO term classifications.

Lipid binding and transport		
AT5G27630	ACBP5	Acyl-CoA Binding protein 5 involved in fatty acid transport
AT2G45600		Phox (PX) domain-containing protein involved in phosphatidylinositol binding
AT2G45880 AT5G46890 AT5G46900 AT4G22520		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G07560	GRP20	Glycine-rich protein 20 involved in lipid-binding oleosins
AT2G30880	SWAP70	Pleckstrin homology (PH) domain-containing protein involved in phosphoinositide binding
AT4G20260	PCAP1	Plasma membrane-associated cation-binding protein 1 involved in intracellular signaling through interaction with PtdInsPs and calmodulin (CaM)
AT5G58784		Dehydrodolichyl diphosphate synthase 5
AT4G39180	SEC14	Phosphatidylinositol/phosphatidylcholine transfer protein. Catalyzes the transfer of phosphatidylinositol and phosphatidylcholine between membranes in vitro.
AT1G79460	GA2	Ent-kaur-16-ene synthase Catalyzes the conversion of ent-copalyl diphosphate to the gibberellin precursor ent-kaur-16-ene
AT1G78960	LUP2	Amyrin synthase Multifunctional enzyme that converts oxidosqualene to nine different triterpenes
AT4G15370	BARS1	Baruol synthase Converts oxidosqualene to baruol (90%) and 22 minor products
AT5G51810	GA20OX2	Gibberellin 20 oxidase 2
AT5G16230	S-ACP-DES3	Acyl-[acyl-carrier-protein] desaturase 3 Converts stearoyl-ACP to oleoyl-AC. Also able to convert palmitoyl-ACP to palmitoleoyl-ACP at C9 position
AT1G78970	LUP1	Lupeol synthase 1 Multifunctional enzyme that converts oxidosqualene to lupeol and 3,20-dihydroxylupane.
Microtubule based movement		
AT3G44050		Kinesin-like protein
AT2G31900	MYO5	Myosin-12
AT5G60930		Kinesin-like protein
AT3G17360	POK1	PHRAGMOPLAST ORIENTING KINESIN 1, Involved in the

		spatial control of cytokinesis by a proper phragmoplast guidance
AT5G06670 AT3G20150 AT2G21380		Kinesin-like protein
AT1G21730	MKRP1	Kinesin-like protein
AT2G22610		Di-glucose binding protein with Kinesin motor domain
AT5G23910		ATP binding microtubule motor family protein
AT3G50240	KICP-02	Kinesin-like protein
AT5G27000	ATK4 or KINESIN4	Kinesin-like protein that binds microtubules in an ATP-dependent manner.
Hydrolases and transferases		
AT3G20860	NEK5	NIMA-RELATED KINASE 5, Encodes a member of the NIMA-related serine/threonine kinases (Neks)
AT1G10760	GWD1, SEX1	Putative serine/threonine protein kinase STARCH EXCESS 1 Mediates the incorporation of phosphate into starch-like alpha-glucan, mostly at the C-6 position of glucose units.
AT1G56120		Leucine-rich repeat transmembrane protein kinase
AT5G10530	LECRK-IX.1	L-TYPE LECTIN RECEPTOR KINASE IX.1
AT1G64300		Protein serine/threonine/tyrosine kinase activity
AT5G57565		Protein serine/threonine/tyrosine kinase activity
AT2G34180	CIPK13	CBL-INTERACTING PROTEIN KINASE 13
AT2G27760	IPT2	ISOPENTENYLTRANSFERASE 2 Involved in the cis-type cytokinin biosynthesis
AT1G24420		HXXXD-type acyl-transferase family protein
AT4G26600		S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT2G25850	PAPS2	POLY(A) POLYMERASE 2, Nuclear poly(A) polymerase 2
AT5G28080	WNK9	Probable serine/threonine-protein kinase Regulate flowering time by modulating the photoperiod pathway
AT3G17420	GPK1	GLYOXYSOMAL PROTEIN KINASE 1
AT2G19910		Serine/threonine protein kinase-like protein RNA-dependent RNA polymerase family protein
AT3G24715		Protein serine/threonine kinase activity

AT2G19910		RNA-dependent RNA polymerase family protein
AT3G24715 AT1G71530		Protein serine/threonine kinase activity
AT3G48260	WNK3	WITH NO LYSINE (K) KINASE 3 Probable serine/threonine-protein kinase Regulate flowering time by modulating the photoperiod pathway
AT5G12480	CPK7	CALMODULIN-DOMAIN PROTEIN KINASE 7 Role in signal transduction pathways that involve calcium as a second messenger.
AT5G58784		dehydrodolichyl diphosphate synthase activity
AT1G51940	LYK3	LYSM-CONTAINING RECEPTOR-LIKE KINASE 3
AT3G51740	IMK2	INFLORESCENCE MERISTEM RECEPTOR-LIKE KINASE 2
AT5G59650		Putative leucine-rich repeat protein kinase
AT5G10720	HK5	HISTIDINE KINASE 5 Negative regulator of the ETR1-dependent abscisic acid (ABA) and ethylene signaling pathway that inhibits the root elongation
AT5G57630	CIPK21	CBL-INTERACTING PROTEIN KINASE 21, CIPK serine-threonine protein kinases interact with CBL proteins.
AT3G21340		Receptor-like protein kinase
AT1G78650	POLD3	DNA polymerase delta 3
AT1G61460		Putative S-locus protein kinase
AT1G21270	WAK2	WALL-ASSOCIATED KINASE 2 Functions as a signaling receptor of extracellular matrix component.
AT4G05210	LPXD1	Probable UDP-3-O-acylglucosamine N-acyltransferase
AT3G59750	LECRK-V.8	L-TYPE LECTIN RECEPTOR KINASE V.8
AT3G29320	PHS1	ALPHA-GLUCAN PHOSPHORYLASE 1 Encodes a plastidic alpha-glucan phosphorylase
AT5G59260	LECRK-II.1	L-TYPE LECTIN RECEPTOR KINASE II.1
AT3G56960	PIP5K4	PHOSPHATIDYL INOSITOL MONOPHOSPHATE 5 KINASE 4. Encodes a protein with phosphatidylinositol-4-phosphate 5-kinase activity that plays a role in pollen tip growth.
AT1G76040	CPK29	CALCIUM-DEPENDENT PROTEIN KINASE 29
Cytoskeleton protein binding		
AT5G48460		Actin binding Calponin homology (CH) domain-containing protein
AT4G17220	MAP70-5	MICROTUBULE-ASSOCIATED PROTEINS 70-5

AT2G01910	MAP65-6	Microtubule-associated protein that mediates the formation of a mesh-like stable and dense network formed by individual microtubules (MT).
AT5G44610	PCAP2	Plasma membrane-associated cation-binding protein 2 Involved in intracellular signaling through interaction with PtdInsPs and calmodulin (CaM);
AT2G34150	SCAR1	Involved in regulation of actin and microtubule organization. Part of a WAVE complex that activates the Arp2/3 complex. Regulates trichome branch positioning and expansion

Table 3-3 Overview of Y2H library screen.

Bait Protein	Independent clones screened	Initial Positives (# of screened)	Confirmed Interactions	Validated by BiFc
PDF2 (START)	4.6×10^6	31	1	0
GL2 (START)	5.8×10^5	25	6	1

Table 3-4 GL2-START unique interactors from Y2H screen.

Locus	Name
At4g25230	E3 Ubiquitin Ligase (RIN2)
At2g22990	Sinapolyglucose:malate sinapolytransferase (SNG1)
At1g78570	Arabidopsis thaliana Rhamnose Biosynthesis 1 (AtRHM1)
At4g12060	Double Clp-N motif protein
At4g22240	Probable Plastid-lipid-associated protein 2 (PAP2)
At2g28180	Cation/H(+) antiporter 8 (CHX8)

Table 3-5 Summary of sequencing and mapping.

Sample		Total number of reads	Number of mapped reads	Alignment rate (%)	
EYFP:GL2	Input	66,737,881	63,220,725	94.73	
gl2-5	Input	58,011,305	52,306,384	90.17	
EYFP:gl2ΔZLZ	Input	68,008,697	66,051,137	97.12	
EYFP:gl2ΔSTART	Input	21,978,672	20,594,497	93.7	
EYFP:gl2^{Amo}	Input	37,612,990	35,698,137	94.91	
EYFP:gl2$^{E375G;R392M}$	Input	71,183,579	69,698805	97.91	
EYFP:GL2	ChIP	99,681,965	19,652,687	19.71	
gl2-5	ChIP	43,993,538	4,872,999	11.1	
EYFP:gl2ΔZLZ	ChIP	84,023,990	54,369,870	64.7	*
EYFP:gl2ΔSTART	ChIP	68,584,915	24,075,540	35.1	
EYFP:gl2^{Amo}	ChIP	120,221,272	70,376,472	58.5	
EYFP:gl2$^{E375G;R392M}$	ChIP	115,810,973	29,386,624	25.3	

* Number of reads (total and mapped) after merging 2-3 biological replicates
Reads were mapped to the *Arabidopsis* genome by adopting Model-based Analysis of ChIP-Seq (MACS) software.

Table 3-6 Summary of GL2 Binding Sites in the *Arabidopsis* Genome.

Sample		Intergenic	Promoter	Exon	TTS	Intron	5'-UTR	3'-UTR	Total
Wild-type	*EYFP:GL2*	306	246	122	160	83	7	1	925
Mutants	*EYFP:gl2ΔSTART*	1018	564	452	320	177	4	6	2541
	EYFP:gl2^{Amo}	1298	782	668	502	262	10	11	3533
	EYFP:gl2$^{E375G;R392M}$	419	344	186	195	128	10	2	1284
	EYFP:gl2ΔZLZ	0	0	2	0	0	0	0	2
	gl2-5	34	65	52	49	26	1	0	227

UTR, untranslated region

Table 3-7 Putative GL2 target genes in which promoters contain atleast one perfect L1 box motif.

Target	Distance to TSS	Fold enrichment	Co-expression	ChIP'd[a]				GO class
				Wild-type	Mutants			
				GL2	gl2ΔSTART	$gl2^{Amo}$	$gl2^{RE}$	
Leucine rich repeat protein kinase-like protein	-556	17.49		101.36				Kinase activity
α-DOX2	-1000	40.43	Root hair cells	186.455			104.75	Lipid Metabolism
Hydroxyproline rich like glycoprotein	3240	28.35		71.82				Development
Putative membrane lipoprotein	-1043	15.1		90.8				Lipid Metabolism
Serine/threonine-protein phosphatase BSL3	-807	29.36		92.7				Response to hormone stimulus
Unfertilized embryo sac 7	-1210	28.35	Root and Trichomes	69.83				Development
F-box associated ubiquitination effector family protein	-265	35.66	Root	71.41				
AtCLB: Calcium-dependent lipid-binding (CaLB domain) family	-212	42.53	Root and Trichomes	91.1				Lipid metabolism

protein								
Strictosidine synthase-like 2	-281	63.79	Root	137.48				Plant secondary metabolism
PP2C, Protein phosphatase 2C	-2459	28.35	Root	72.99				Response to Abscisic acid stimulus
β-glucosidase 31	-2159	38.88			104.26			Hydrolase activity
FAH1, Ferulic acid 5-hydroxylase 1	3385	28.72			98.19			Phenylpropanoid metabolism
SR1, Calmodulin binding transcription factor	-1613	10.13			95.46	68.39	79.03	Response to stimulus
AtCSLB5, Cellulose synthase–like protein B5	-1434	22.37	Root		52.7			Cell wall and root hair elongation
AtGATL1, Putative galactouronosyltrasnferase-like 1	-2344	75.59	Root and Trichomes		145.51			Carbohydrate Metabolism
AtPRR1, Pinoresinol reductase 1	-9	50.29	Root		112.61			Lignin Biosynthesis
MYB23	-2182	31.59	Root and Trichomes		72.4			Transcription factor activity
UGT72E1, UDP-glucosyltransferase 72E1	-17	31.59	Root and Trichomes		63.73		168.47	Metabolism
DRP5A, Dynamin related protein 5A	-805	35.1			62.35			Cytokinesis
WOX9, WUSCHEL related homeobox 9	-3090	41.99			98.83			Transcription factor
KNAT1, homeobox protein	4721	41.99			81.55			Transcription factor

knotted-1-like-1								
UNE16, Unfertilized embryo sac	-960	50.39			98.83			Transcription factor activity
ALDH10A9,aldehyde dehydrogenase 10A9	-276	30.25	Root and Trichomes			85.61		
SVL3, glycerophosphodiester phoshodiesterase-like protein	-43	25.93				100.8		Glycerophospholipid Metabolism
5-FCL, 5-formotetrahydrofolate cycloligase	-868	12.96	Root			53.79		Thiamin Metabolism
GDSL esterase/lipase EXL3	-1017	5.46				113.3	214.68	Lipase activity
PAL2, phenylalanine ammonia-lyase 2	-708	22.38	Root			64.75		Flavonoid metabolism
CYP81F4, Cytochrome P450	-1790	10.13	Root and Trichomes			74.56		Development
SEPALLATA 4	-1232	30.25				74.6		Transcription factor activity
SPT, SPATULA	-266	25.93	Root and Trichomes			128.8		Transcription factor activity
TINY2, dehydration-responsive element-binding protein 3	2375	15.47	Trichomes			58.31		Response to stimulus
NIMIN1, NIM1-INTERACTING 1	112	15.98	Root			58.9		Metabolism
EXPA4, Expansin A4	-3921	21.61	Trichomes			60.29		Cell wall
EOL2	-106	31.8	Trichomes				97.41	Ethylene biosynthesis

PGA4, polygalacturonase 4	-609	40.99	Root				104.75	Carbohydrate Metabolism
RAB8, RAB GTPase homolog 8	-10	151.99	Root and Trichomes				40.99	GTP binding

[a] $-10*\text{Log}_{10}$ P-values for genes with FDR < 0.05

Genes differentially expressed in publicly available microarray data are highlighted in red

Table 3-8 List of plasmids made in this study and their description.

	Construct	Description	Insert	Vector	Process
Plants					
	EYFP:GL2	GL2 cDNA in SR54	GL2 cDNA in pBluescript	SR54 (proGL2::EYFP:GL2)	PCR, ligation
	EYFP:gl2ΔSTART	GL2-ΔSTART in SR54	GL2-STARTΔ in pBluescript	SR54	Site directed mutagenesis, ligation
	EYFP:gl2ΔZLZ	GL2-ZLZΔ in SR54	GL2-ZLZΔ in pBluescript	SR54	Site directed mutagenesis, ligation
	EYFP:gl2ΔHD	GL2-HDΔ in SR54	GL2-HDΔ in pBluescript	SR54	Site directed mutagenesis, ligation
	EYFP:gl2^{R392M}	GL2-R392M in SR54	GL2-R392M in pBluescript	SR54	Site directed mutagenesis, ligation
	EYFP:gl2$^{E375G;R392M}$	GL2-E375G; R392M in SR54	GL2-E375G; R392M in pBluescript	SR54	Site directed mutagenesis, ligation
	EYFP:gl2^{L480P}	GL2-L480P in SR54	GL2-L480P in pBluescript	SR54	Site directed mutagenesis, ligation
	EYFP:gl2^{E375G}	GL2-E375G in SR54	GL2-E375G in pBluescript	SR54	Site directed mutagenesis, ligation
	EYFP:gl2^{V391L}	GL2-V391L in SR54	GL2-V391L in pBluescript	SR54	Site directed mutagenesis, ligation
	EYFP:gl2^{Amo}	GL2-Amo in SR54	GL2-Amo in pBluescript	SR54	Site directed mutagenesis, ligation
	35S::gl2ΔHD-GFP	GL2-HDΔ in pK7WGF2	GL2-HDΔ in pENTR/D-TOPO	pK7WGF2	TOPO cloning, Gateway LR reaction
	35S::RHM1-GFP	RHM1 cDNA in pK7WGF2	RHM1 cDNA in pENTR/D-TOPO	pK7WGF2	TOPO cloning, Gateway LR reaction
	35S::PDF2 (START)-GFP	PDF2-START in pK7WGF2	PDF2 (START) in pENTR/D-TOPO	pK7WGF2	TOPO cloning, Gateway LR reaction

Yeast Two-hybrid				
GL2 (1-747): Bait	GL2 cDNA in pGBKT7	GL2 cDNA (At1g79840)	pGBKT7	PCR, ligation
GL2 (1-747): Prey	GL2 cDNA in pGADT7	GL2 cDNA (At1g79840)	pGADT7	PCR, ligation
gl2ΔSTART: Bait	GL2-STARTΔ in pGBKT7	GL2-STARTΔ in pBluescript	pGBKT7	PCR, ligation
gl2ΔSTART: Prey	GL2-STARTΔ in pGADT7	GL2-STARTΔ in pBluescript	pGADT7	PCR, ligation
gl2R392M: Prey	GL2-R392M in pGBKT7	GL2-R392M in pBluescript	pGBKT7	PCR, ligation
gl2L480P: Prey	GL2-L480P in pGADT7	GL2-L480P in pBluescript	pGADT7	PCR, ligation
gl2E375G: Prey	GL2-E375G in pGADT7	GL2-E375G in pBluescript	pGADT7	PCR, ligation
GL2 (1-231): Bait	GL2 (1-231 aa) in pGBKT7	GL2 cDNA	pGBKT7	PCR, ligation
GL2 (1-231): Prey	GL2 (1-231 aa) in pGADT7	GL2 cDNA	pGADT7	PCR, ligation
GL2 (157-231): Bait	GL2 (157-231 aa) in pGBKT7	GL2 cDNA	pGBKT7	PCR, ligation
GL2 (157-231): Prey	GL2 (157-231 aa) in pGADT7	GL2 cDNA	pGADT7	PCR, ligation
GL2 (1-493): Bait	GL2 (1-493 aa) in pGBKT7	GL2 cDNA	pGBKT7	PCR, ligation
GL2 (1-493): Prey	GL2 (1-493 aa) in pGADT7	GL2 cDNA	pGADT7	PCR, ligation
GL2 (253-493) or GL2-START: Bait	GL2 (253-493 aa) in pGBKT7	GL2 cDNA	pGBKT7	PCR, ligation
GL2 (253-493) or GL2-START: Prey	GL2 (253-493 aa) in pGADT7	GL2 cDNA	pGADT7	PCR, ligation
GL2 (253-747): Bait	GL2 (253-747 aa) in pGBKT7	GL2 cDNA	pGBKT7	PCR, ligation
GL2 (253-747): Prey	GL2 (253-747 aa) in pGADT7	GL2 cDNA	pGADT7	PCR, ligation
GL2 (494-747): Bait	GL2 (494-747 aa) in pGBKT7	GL2 cDNA	pGBKT7	PCR, ligation
GL2 (494-747): Prey	GL2 (494-747 aa) in pGADT7	GL2 cDNA	pGADT7	PCR, ligation
PDF2-START: Bait	PDF2-START in pGBKT7	PDF2-START in pNIC28-Bsa4	pGBKT7	PCR, ligation

	PDF2-START: Prey	PDF2-START in pGADT7	PDF2-START in pNIC28-Bsa4	pGADT7	PCR, ligation
	RHM1	RHM1 cDNA in pGBKT7	RHM1 cDNA (At1g78570)	pGBKT7	PCR, ligation
	Truncated SAM ($SAM_{100-187}$)	$SAM_{100-187}$ in pGADT7		pGADT7	PCR, ligation
Quantitative β-Galactosidase Assay					
	GL2 (1-747): Bait	GL2 cDNA in pDEST32	GL2 cDNA (At1g79840)	pDEST32	Gateway LR reaction
	GL2 (1-747): Prey	GL2 cDNA in pDEST22	GL2 cDNA	pDEST22	Gateway LR reaction
	PDF2: Prey	PDF2 cDNA in pDEST22	PDF2 cDNA in pENTR2231-Sfi (#U16775)	pDEST22	Gateway LR reaction
	HDG1: Prey	HDG1 cDNA in pDEST22	HDG1 cDNA	pDEST22	Gateway LR reaction
	HDG11: Prey	HDG11 cDNA in pDEST22	HDG11 cDNA	pDEST22	Gateway LR reaction
	HDG12: Prey	HDG12 cDNA in pDEST22	HDG12 cDNA	pDEST22	Gateway LR reaction
	PHB: Prey	PHB cDNA in pDEST22	PHB cDNA	pDEST22	Gateway LR reaction
	ATHB8: Prey	ATHB8 cDNA in pDEST22	ATHB8 cDNA	pDEST22	Gateway LR reaction
	HDG4: Prey	HDG4 cDNA in pDEST22	HDG4 cDNA in pUNI51 (# 103238)	pDEST22	Ligation, Gateway LR reaction
Bimolecular Fluorescence complementation Assay (BiFC)					
	nGFP-GL2	35S::GL2-nGFP in 125-NXGW	GL2 cDNA in pENTR/D-TOPO	125-NXGW	Gateway LR reaction
	cGFP-GL2	35S::GL2-cGFP in 127-CXGW	GL2 cDNA in pENTR/D-TOPO	127-CXGW	Gateway LR reaction
	nGFP-gl2ΔSTART	35S::ΔSTART-nGFP in 125-NXGW	GL2-STARTΔ in pENTR/D-TOPO	125-NXGW	Gateway LR reaction
	cGFP-gl2ΔSTART	35S::ΔSTART-cGFP in 127-CXGW	GL2-STARTΔ in pENTR/D-TOPO	127-CXGW	Gateway LR
	nGFP-$gl2^{Amo}$	35S::GL2-Amo-nGFP in 125-NXGW	GL2-Amo in pENTR/D-TOPO	125-NXGW	Gateway LR reaction
	cGFP-$gl2^{Amo}$	35S::GL2-Amo-nGFP in 127-CXGW	GL2-Amo in pENTR/D-TOPO	127-CXGW	Gateway LR

				reaction
cGFP-gl2ΔZLZ	35S::GL2-ΔZLZ-cGFP in 127-CXGW	GL2-ZLZΔ in pENTR/D-TOPO	127-CXGW	Gateway LR reaction
cGFP-gl2^{L480P}	35S::GL2-L480P-cGFP in 127-CXGW	GL2-L480P in pENTR/D-TOPO	127-CXGW	Gateway LR reaction
cGFP-gl2$^{E375G;R392M}$	35S::GL2-RE-cGFP in 127-CXGW	GL2-E375G;R392M in pENTR/D-TOPO	127-CXGW	Gateway LR reaction
FAMA *				
FAMA *				
PDF2 *				
PDF2 *				
ATML1 *				
ATML1 *				
nGFP-RHM1	35S::RHM1-nGFP in 125-NXGW	RHM1 cDNA in pENTR/D-TOPO	125-NXGW	Gateway LR reaction
cGFP-RHM1	35S::RHM1-cGFP in 127-CXGW	RHM1 cDNA in pENTR/D-TOPO	127-CXGW	Gateway LR reaction
nGFP-PDF2	35S::PDF2-nGFP in 125-NXGW	PDF2 cDNA in pENTR223.1-Sfi (# U16775)	125-NXGW	Triparental Mating
cGFP-PDF2	35S::PDF2-cGFP in 127-CXGW	PDF2 cDNA in pENTR223.1-Sfi (# U16775)	127-CXGW	Triparental Mating
nGFP-PDF2 (START)	35S::PDF2 (START)-nGFP in 125-NXGW	PDF2 (START) in pENTR/D-TOPO	125-NXGW	Gateway LR reaction
cGFP-PDF2 (START)	35S::PDF2 (START)-cGFP in 127-CXGW	PDF2 (START) in pENTR/D-TOPO	127-CXGW	Gateway LR reaction
nGFP-GL2 (START)	35S::GL2 (START)-nGFP in 125-NXGW	GL2-START in pENTR/D-TOPO	125-NXGW	Gateway LR reaction
cGFP-PHB	35S::PHB-cGFP in 127-CXGW	PHB cDNA in pENTR223.1-Sfi (#U25105)	127-CXGW	Triparental Mating
cGFP-HDG11	35S::HDG11-cGFP in 127-CXGW	HDG11 cDNA in pENTR/D-TOPO	127-CXGW	Gateway LR reaction

* Gifts from Adrienne Reoder at Weill Institute of Cell and Molecular Biology, Cornell University.

Table 3-9 Oligonucleotide primers used in this study. Nucleotide bases shown in bold denote restriction sites used for cloning or changed bases from site-directed mutagenesis.

Name	5'-3' sequence
I. Primers for site-directed mutagenesis of GLABRA2 (GL2)	
GL2_E375G_F	CAG CTG ATG TTC GGA **GGG** ATG CAG CTG CTC ACT
GL2_E375G_R	AGT GAG CAG CTG CAT **CCC** TCC GAA CAT CAG CTG
GL2_R386L_F	T CCG GTC GTC CCC ACA **CTA** GAA GTG TAC TTC G
GL2_R386L_R	C GAA GTA CAC TTC **TAG** TGT GGG GAC GAC CGG A
GL2_E387L_F	T CCG GTC GTC CCC ACA AGA **CTA** GTG TAC TTC G
GL2_E387L_R	C GAA GTA CAC TTC TCT **TAG** GGG GAC GAC CGG A
GL2_R386L;E387L_F	CCG GTC GTC CCC ACA **CTA CTA** GTG TAC TTC GTG AGA AGC
GL2_R386L;E387L_R	GCT TCT CAC GAA GTA CAC **TAG TAG** TGT GGG GAC GAC CGG
GL2_V391L_F	AGA GAA GTG TAC TTC **CTG** AGA AGC TGC CGG CAG C
GL2_V391L_R	G CTG CCG GCA GCT TCT **CAG** GAA GTA CAC TTC TCT
GL2_R392M_F	GAA GTG TAC TTC GTG **ATG** AGC TGC CGG CAG CTG
GL2_R392M_R	CAG CTG CCG GCA GCT **CAT** CAC GAA GTA CAC TTC
GL2_L480P_F	G GTC GCC ACC **CCT** CAG CTC CAT TGC G
GL2_L480P_R	C GCA ATG GAG CTG **AGG** GGT GGC GAC C
II. Primers for deletion constructs of GL2 (Primers were phosphorylated [Phos] at their 5' ends.)	
GL2_HD_Δ_F	[Phos] GAACGGCACGAGAACTCCCTGCTC
GL2_HD_ Δ _R	[Phos] CTTATTAGTGCCCTTGTTTCCAGCTGCG
GL2_ZLZ_ Δ _F	[Phos] TCATGCTCCGACGATCAAGAACACCG
GL2_ZLZ_ Δ _R	[Phos] TTGAATAGCCTTGATCTGTGTGCGGCG
GL2_START_ Δ _F	[Phos] GTC TTC TTC ATG GCT ACC AAC GTC CCC ACC
GL2_START_ Δ _R	[Phos] GAG GGC AAA GAC GCC CGT GTA GAA ATC G
GL2_SAD_ Δ _F	[Phos] CAAATCGAAGATTGCTGATGACAAGTCACAGC
GL2_SAD_ Δ _R	[Phos] GGTAGCCATGAAGAAGACAAGGCGTTCG
III. Sequencing primers for proGL2::EYFP:GL2 cassette	
proGL2::EYFP:GL2_seq_B1	CAC AGG AAA CAG CTA TGA CC
proGL2::EYFP:GL2_seq_B2	CGA TCG GTC AAT GCC TCT CGC
proGL2::EYFP:GL2_seq_B3	GAG TCC CAG TTT CCT ATA ATC C
proGL2::EYFP:GL2_seq_B4	CCT TCC CCC CCT GTC TAC AG
proGL2::EYFP:GL2_seq_B5	GTA TCA TTG GAA TTG TAG AGG C
proGL2::EYFP:GL2_seq_B6	CAA GGA CGA CGG CAA CTA C
proGL2::EYFP:GL2_seq_B7	GCC CTC TCT CTA TCT CTC GC
proGL2::EYFP:GL2_seq_B8	GGC TAA TTC CTC CTG CCC C
proGL2::EYFP:GL2_seq_B9	CGA TGT TAT CCG GCA AGG CG
proGL2::EYFP:GL2_seq_B10	GCT TCT ACC GCG CCA TTG C
proGL2::EYFP:GL2_seq_B11	CGT CTA CAC AAG ACG ACA G
proGL2::EYFP:GL2_seq_B12	GGG TTT GCC TCC AAC CCC
IV. Primers for cloning into yeast-two hybrid vectors	
Name	**5'-3' sequence**
GL2_747aa_EcoRI_F	GAC AGG **GAA TTC** *ATG* TCA ATG GCC GTC GAC ATG TCT TCC
GL2_747aa_BamHI_R	TGCTGT**GGATCC**TCATCA GCA ATC TTC GAT TTG TAG AC
GL2_493aa_BamHI_R	TTGGT**GGATCC**CTA GGT AGC CAT GAA GAA GAC AAG GCG
GL2_253aa_EcoRI_F	GTCTTT**GAATTC** GAG AAG TCC CGT ATT GCC GAG ATT TC
GL2_494aa_EcoRI_F	TTCATG**GAATTC** AAC GTC CCC ACC AAA GAC TCT CTC GG
GL2_157aa_EcoRI_F	ACCGCCGCACA**GAATTC** AAG GCT ATT CAA GAA CGG C
GL2_231aa_BamHI_R	ATCGTC**GGATCC**TCA AGC CTG CAG GGG ATA GGG
PDF2_START_EcoRI_F	GAGGTCA**GAATTC**ATTCCTTCTGAGACTGATAAGCC
PDF2_START_BamHI_R	GAAAGATCACCAGGAA**GGATCC**TGGCCATGG

GL2_START_EcoRI_F	GTCTTT**GAATTC**GAGAAGTCCCGTAT
GL2_START_BamHI_R	ATCGTC**GGATCC**TCA AGC CTG CAG GGG ATA GGG
RHM1_669aa_ EcoRI_F	
RHM1_669aa_ BamHI_R	
SAM (At5g40830)_ SmaI_F	AGTAT**CCCGGG**TGTATCGCTGAAGATTGGCG
SAM (At5g40830)_XhoI_R	AA**CTCGAG**ATATCAATCAAATCCGGGCAGG
V. Primers for sequence verification of yeast-two hybrid constructs	
BD_seq_F	AACCGAAGTGCGCCAAGTGTCTG
AD_seq_F	TATAACGCGTTTGGAATCACT
BD & AD_seq_R	AGCCGACAACCTTCATTGGAGAC
VI. Primers for cloning in pENTR™/D-TOPO (Start codons, italics; Stop codons, red)	
GL2_TOPO_F	CACC ***ATG*** TCA ATG GCC GTC GAC ATG
GL2_TOPO_R	TCA TTA GCA ATC TTC GAT TTG TAG ACT TC
RHM1_TOPO_F	CACC ***ATG*** GCT TCG TAC ACT CCC AA
RHM1_TOPO_R	C TCT TCA GGT TTT CTT GTT TGG CCCG
GL2_START_TOPO_F	CACC ***ATG*** GAG AAG TCC CGT ATT GC
GL2_START_TOPO_R	GGG GAC CTA GGT AGC CAT GAA G
PDF2_START_TOPO_F	CACC GAG ACT GAT AAG CCT ATA ATC GTG
PDF2_START_TOPO_R	TCA GGC CAT GGA GCT AGC AAG CC
SAM (At5g40830)_TOPO_F	CACC ***ATG*** GGC TCT GTA TCG CTG AAG
SAM (At5g40830)_TOPO_R	TCA AAT CCG GGC AGG CTT TTG CAG
VII. Primers for sequence verification of Gateway constructs	
GFP-NXGW_seq	GACAAGCAGAAGAACGGCAT
pro35S_pK7WG2_seq	GGTGGCACCTACAAATGCCATC
VIII. Primers used for qRT-PCR	
ACT7_PCR(94)_F	TCGCACATGTACTCGTTTCGCTTTC (amplifies 94 bp)
ACT7_PCR(94)_R	TCGAGAAGCAGCGAGAGAGAAAGATAGA
GL2_PCR_F	ATGAAGCTCGTCGGCATGAGTGGG (amplifies 106 bp)
GL2_PCR_R	TGGATTGCCACTGAGTTGCCTCTG
HDG11_RT(125)_F	TGGGGCTGATCGTTGGGTTACCA (amplifies 125 bp)
HDG11_RT(125)_R	TGCTTCTCTTCCCTTCCGGTGA
MYB23_RT(126)_F	CATCAGACTCCACAAGCTCCTCGG (amplifies 126 bp)
MYB23_RT(126)_R	TCTCCGAGACCAAGTTTCTTGCTGAG
α-DOX2_RT(123)_F	GGGTTGATCACTTAGAAGACAC (amplifies 123 bp)
α-DOX2_RT(123)_R	GTTCCCACCGATGATCACC
LTP4_RT(95)_F	ATCCAAGTCTAGCCTCTGGC (amplifies 95 bp)
LTP4_RT(95)_R	ACTTCACTTGATGGTGGCGC
FLS2_RT(114)_F	CGATCATTGGTTAGATGTCCAC (amplifies 114 bp)
FLS2_RT(114)_R	CCACTATCGCTCTATGCAACAC

REFERENCES

Abe M, Katsumata H, Komeda Y, Takahashi T. 2003. Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. *Development* **130**: 635-643.

Abe M, Takahashi T, Komeda Y. 2001. Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. *Plant J* **26**: 487-494.

Alonso R, Onate-Sanchez L, Weltmeier F, Ehlert A, Diaz I, Dietrich K, Vicente-Carbajosa J, Droge-Laser W. 2009. A Pivotal Role of the Basic Leucine Zipper Transcription Factor bZIP53 in the Regulation of *Arabidopsis* Seed Maturation Gene Expression Based on Heterodimerization and Protein Complex Formation. *Plant Cell* **21**: 1747-1761.

Alpy F, Latchumanan VK, Kedinger V, Janoshazi A, Thiele C, Wendling C, Rio MC, Tomasetto C. 2005. Functional characterization of the MENTAL domain. *J Biol Chem* **280**: 17945-17952.

Alpy F, Tomasetto C. 2014. START ships lipids across interorganelle space. *Biochimie* **96C**: 85-95.

Ariel FD, Manavella PA, Dezar CA, Chan RL. 2007. The true story of the HD-Zip family. *Trends Plant Sci* **12**: 419-426.

Baker BY, Lin L, Kim CJ, Raza J, Smith CP, Miller WL, Achermann JC. 2006. Nonclassic congenital lipoid adrenal hyperplasia: a new disorder of the steroidogenic acute regulatory protein with very late presentation and normal male genitalia. *J Clin Endocrinol Metab* **91**: 4781-4785.

Bannenberg G, Martinez M, Rodriguez MJ, Lopez MA, Ponce de Leon I, Hamberg M, Castresana C. 2009. Functional analysis of *alpha-DOX2*, an active alpha-dioxygenase critical for normal development in tomato plants. *Plant Physiology* **151**: 1421-1432.

Basu D, Le J, El-Essal SED, Huang S, Zhang CH, Mallery EL, Koliantz G, Staiger CJ, Szymanski DB. 2005. DISTORTED3/SCAR2 is a putative *Arabidopsis* WAVE complex subunit that activates the Arp2/3 complex and is required for epidermal morphogenesis. *Plant Cell* **17**: 502-524.

Bruex A, Kainkaryam RM, Wieckowski Y, Kang YH, Bernhardt C, Xia Y, Zheng X, Wang JY, Lee MM, Benfey P et al. 2012. A gene regulatory network for root epidermis cell differentiation in *Arabidopsis*. *PLoS Genetics* **8**: e1002446.

Chen H, Lai ZB, Shi JW, Xiao Y, Chen ZX, Xu XP. 2010. Roles of *Arabidopsis* WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biology* **10**.

Christophe-Hobertus C, Duquesne V, Pichon B, Roger PP, Christophe D. 1999. Critical residues of the homeodomain involved in contacting DNA bases also specify the nuclear accumulation of thyroid transcription factor-1. *Eur J Biochem* **265**: 491-497.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**: 735-743.

Costa S, Dolan L. 2003. Epidermal patterning genes are active during embryogenesis in *Arabidopsis*. *Development* **130**: 2893-2901.

Daviere JM, Achard P. 2013. Gibberellin signaling in plants. *Development* **140**: 1147-1151.

De Silva K, Laska B, Brown C, Sederoff HW, Khodakovskaya M. 2011. *Arabidopsis thaliana* calcium-dependent lipid-binding protein (AtCLB): a novel repressor of abiotic stress response. *Journal of Experimental Botany* **62**: 2679-2689.

Depege-Fargeix N, Javelle M, Chambrier P, Frangne N, Gerentes D, Perez P, Rogowsky PM, Vernoud V. 2011. Functional characterization of the HD-ZIP IV transcription factor OCL1 from maize. *Journal of Experimental Botany* **62**: 293-305.

Di Cristina M, Sessa G, Dolan L, Linstead P, Baima S, Ruberti I, Morelli G. 1996. The *Arabidopsis* Athb-10 (GLABRA2) is an HD-Zip protein required for regulation of root hair development. *Plant J* **10**: 393-402.

Diet A, Link B, Seifert GJ, Schellenberg B, Wagner U, Pauly M, Reiter WD, Ringli C. 2006. The *Arabidopsis* root hair cell wall formation mutant *lrx1* is suppressed by mutations in the *RHM1* gene encoding a UDP-L-rhamnose synthase. *Plant Cell* **18**: 1630-1641.

Djakovic S, Dyachok J, Burke M, Frank MJ, Smith LG. 2006. BRICK1/HSPC300 functions with SCAR and the ARP2/3 complex to regulate epidermal cell shape in *Arabidopsis*. *Development* **133**: 1091-1100.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* **38**: W64-70.

Fabbri M, Delp G, Schmidt O, Theopold U. 2000. Animal and plant members of a gene family with similarity to alkaloid-synthesizing enzymes. *Biochemical and Biophysical Research Communications* **271**: 191-196.

Ferrer JL, Austin MB, Stewart C, Jr., Noel JP. 2008. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry : PPB / Societe francaise de physiologie vegetale* **46**: 356-370.

Gendrel AV, Lippman Z, Yordan C, Colot V, Martienssen RA. 2002. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene *DDM1*. *Science* **297**: 1871-1873.

Ghaffari M, Zeng X, Whitsett JA, Yan C. 1997. Nuclear localization domain of thyroid transcription factor-1 in respiratory epithelial cells. *Biochemical Journal* **328**: 757-761.

Gietz RD, Woods RA. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**: 87-96.

Glover BJ. 2000. Differentiation in plant epidermal cells. *Journal of Experimental Botany* **51**: 497-505.

Gou JY, Li K, Wu K, Wang X, Lin H, Cantu D, Uauy C, Dobon-Alonso A, Midorikawa T, Inoue K et al. 2015. Wheat Stripe Rust Resistance Protein WKS1 Reduces the Ability of the Thylakoid-Associated Ascorbate Peroxidase to Detoxify Reactive Oxygen Species. *Plant Cell* **27**: 1755-1770.

Graham LE, Cook ME, Busse JS. 2000. The origin of plants: Body plan changes contributing to a major evolutionary radiation. *Proc Natl Acad Sci USA* **97**: 4535-4540.

Gray JE, Holroyd GH, van der Lee FM, Bahrami AR, Sijmons PC, Woodward FI, Schuch W, Hetherington AM. 2000. The HIC signalling pathway links CO_2 perception to stomatal development. *Nature* **408**: 713-716.

Hamberg M, Ponce de Leon I, Sanz A, Castresana C. 2002. Fatty acid alpha-dioxygenases. *Prostaglandins & other Lipid Mediators* **68-69**: 363-374.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**: 576-589.

Hessabi B, Schmidt I, Walther R. 2000. The homeodomain of Nkx2.2 carries two cooperatively acting nuclear localization signals. *Biochemical and Biophysical Research Communications* **270**: 695-700.

Hu CD, Chinenov Y, Kerppola TK. 2002. Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* **9**: 789-798.

Huang J, Gu M, Lai Z, Fan B, Shi K, Zhou YH, Yu JQ, Chen Z. 2010. Functional analysis of the *Arabidopsis* PAL gene family in plant growth, development, and response to environmental stress. *Plant Physiology* **153**: 1526-1538.

Hung CY, Lin Y, Zhang M, Pollock S, Marks MD, Schiefelbein J. 1998. A common position-dependent mechanism controls cell-type patterning and GLABRA2 regulation in the root and hypocotyl epidermis of *Arabidopsis*. *Plant Physiology* **117**: 73-84.

Ito M, Sentoku N, Nishimura A, Hong SK, Sato Y, Matsuoka M. 2003. Roles of rice GL2-type homeobox genes in epidermis differentiation. *Breeding Sci* **53**: 245-253.

Izaguirre MM, Scopel AL, Baldwin IT, Ballare CL. 2003. Convergent responses to stress. Solar ultraviolet-B radiation and Manduca sexta herbivory elicit overlapping transcriptional responses in field-grown plants of *Nicotiana longiflora*. *Plant Physiology* **132**: 1755-1767.

Javelle M, Klein-Cosson C, Vernoud V, Boltz V, Maher C, Timmermans M, Depege-Fargeix N, Rogowsky PM. 2011. Genome-Wide Characterization of the HD-ZIP IV Transcription Factor Family in Maize: Preferential Expression in the Epidermis. *Plant Physiology* **157**: 790-803.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* **316**: 1497-1502.

Kanaoka MM, Pillitteri LJ, Fujii H, Yoshida Y, Bogenschutz NL, Takabayashi J, Zhu JK, Torii KU. 2008. SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to arabidopsis stomatal differentiation. *Plant Cell* **20**: 1775-1785.

Kanno K, Wu MK, Agate DS, Fanelli BJ, Wagle N, Scapa EF, Ukomadu C, Cohen DE. 2007. Interacting proteins dictate function of the minimal START domain phosphatidylcholine transfer protein/StarD2. *Journal of Biological Chemistry* **282**: 30728-30736.

Kaufmann K, Muino JM, Osteras M, Farinelli L, Krajewski P, Angenent GC. 2010. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nature Protocols* **5**: 457-472.

Kawaoka A, Ebinuma H. 2001. Transcriptional control of lignin biosynthesis by tobacco LIM protein. *Phytochemistry* **57**: 1149-1157.

Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. *Nature* **389**: 33-39.

Khosla A, Paper JM, Boehler AP, Bradley AM, Neumann TR, Schrick K. 2014. HD-Zip Proteins GL2 and HDG11 Have Redundant Functions in *Arabidopsis* Trichomes, and GL2 Activates a Positive Feedback Loop via MYB23. *Plant Cell* **26**: 2184-2200.

Kirik V, Lee MM, Wester K, Herrmann U, Zheng Z, Oppenheimer D, Schiefelbein J, Hulskamp M. 2005. Functional diversification of *MYB23* and *GL1* genes in trichome morphogenesis and initiation. *Development* **132**: 1477-1485.

Knauer S, Holt AL, Rubio-Somoza I, Tucker EJ, Hinze A, Pisch M, Javelle M, Timmermans MC, Tucker MR, Laux T. 2013. A Protodermal miR394 Signal Defines a Region of Stem Cell Competence in the *Arabidopsis* Shoot Meristem. *Developmental Cell* **24**: 125-132.

Koeduka T, Matsui K, Hasegawa M, Akakabe Y, Kajiwara T. 2005. Rice fatty acid alpha-dioxygenase is induced by pathogen attack and heavy metal stress: activation through jasmonate signaling. *J Plant Physiol* **162**: 912-920.

Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* **106**: 10171-10176.

Kudo N, Kumagai K, Tomishige N, Yamaji T, Wakatsuki S, Nishijima M, Hanada K, Kato R. 2008. Structural basis for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide. *Proc Natl Acad Sci U S A* **105**: 488-493.

Kurdyukov S, Faust A, Nawrath C, Bar S, Voisin D, Efremova N, Franke R, Schreiber L, Saedler H, Metraux JP et al. 2006. The epidermis-specific extracellular BODYGUARD controls cuticle development and morphogenesis in *Arabidopsis*. *Plant Cell* **18**: 321-339.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Lavigne P, Najmanivich R, Lehoux JG. 2010. Mammalian StAR-related lipid transfer (START) domains with specificity for cholesterol: structural conservation and mechanism of reversible binding. *Subcell Biochem* **51**: 425-437.

Letourneau D, Lorin A, Lefebvre A, Frappier V, Gaudreault F, Najmanovich R, Lavigne P, LeHoux JG. 2012. StAR-related lipid transfer domain protein 5 binds primary bile acids. *Journal of Lipid Research* **53**: 2677-2689.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Lieckfeldt E, Simon-Rosin U, Kose F, Zoeller D, Schliep M, Fisahn J. 2008. Gene expression profiling of single epidermal, basal and trichome cells of *Arabidopsis thaliana*. *J Plant Physiol* **165**: 1530-1544.

Ligrone R, Duckett JG, Renzaglia KS. 2012. Major transitions in the evolution of early land plants: a bryological perspective. *Ann Bot-London* **109**: 851-871.

Lin Y, Schiefelbein J. 2001. Embryonic control of epidermal cell patterning in the root and hypocotyl of *Arabidopsis*. *Development* **128**: 3697-3705.

Long JA, Ohno C, Smith ZR, Meyerowitz EM. 2006. TOPLESS regulates apical embryonic fate in *Arabidopsis*. *Science* **312**: 1520-1523.

Lu L, Lee YRJ, Pan RQ, Maloof JN, Liu B. 2005. An internal motor kinesin is associated with the golgi apparatus and plays a role in trichome morphogenesis in *Arabidopsis*. *Mol Biol Cell* **16**: 811-823.

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E. 2009. Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324**: 1064-1068.

Marks MD, Wenger JP, Gilding E, Jilk R, Dixon RA. 2009. Transcriptome analysis of *Arabidopsis* wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. *Mol Plant* **2**: 803-822.

Maruta T, Noshi M, Nakamura M, Matsuda S, Tamoi M, Ishikawa T, Shigeoka S. 2014. Ferulic acid 5-hydroxylase 1 is essential for expression of anthocyanin biosynthesis-associated genes and anthocyanin accumulation under photooxidative stress in *Arabidopsis*. *Plant Sci* **219**: 61-68.

Mathieu AP, Lavigne P, LeHoux JG. 2002. Molecular modeling and structure-based thermodynamic analysis of the StAR protein. *Endocr Res* **28**: 419-423.

Mathur J, Spielhofer P, Kost B, Chua NH. 1999. The actin cytoskeleton is required to elaborate and maintain spatial patterning during trichome cell morphogenesis in *Arabidopsis thaliana*. *Development* **126**: 5559-5568.

Metherell LA, Naville D, Halaby G, Begeot M, Huebner A, Nurnberg G, Nurnberg P, Green J, Tomlinson JW, Krone NP et al. 2009. Nonclassic lipoid congenital adrenal hyperplasia masquerading as familial glucocorticoid deficiency. *J Clin Endocrinol Metab* **94**: 3865-3871.

Mierziak J, Kostyn K, Kulma A. 2014. Flavonoids as important molecules of plant interactions with the environment. *Molecules* **19**: 16240-16265.

Mouradov A, Spangenberg G. 2014. Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front Plant Sci* **5**.

Mukherjee K, Burglin TR. 2006. MEKHLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins. *Plant Physiology* **140**: 1142-1150.

Murashige T, Skoog F. 1962. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**: 473-&.

Murcia M, Faraldo-Gomez JD, Maxfield FR, Roux B. 2006. Modeling the structure of the StART domains of MLN64 and StAR proteins in complex with cholesterol. *Journal of Lipid Research* **47**: 2614-2630.

Nadakuduti SS, Pollard M, Kosma DK, Allen C, Ohlrogge JB, Barry CS. 2012. Pleiotropic Phenotypes of the sticky peel Mutant Provide New Insight into the Role of CUTIN DEFICIENT2 in Epidermal Cell Function in Tomato. *Plant Physiology* **159**: 945-960.

Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T. 2006. Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in *Arabidopsis*. *Plant Physiology* **141**: 1363-1375.

Nelson BK, Cai X, Nebenfuhr A. 2007. A multicolored set of *in vivo* organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J* **51**: 1126-1136.

Nobusawa T, Okushima Y, Nagata N, Kojima M, Sakakibara H, Umeda M. 2013. Synthesis of Very-Long-Chain Fatty Acids in the Epidermis Controls Plant Organ Growth by Restricting Cell Proliferation. *Plos Biol* **11**.

Ohashi Y, Oka A, Rodrigues-Pousada R, Possenti M, Ruberti I, Morelli G, Aoyama T. 2003. Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. *Science* **300**: 1427-1430.

Ohashi Y, Oka A, Ruberti I, Morelli G, Aoyama T. 2002. Entopically additive expression of *GLABRA2* alters the frequency and spacing of trichome initiation. *Plant J* **29**: 359-369.

Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M. 2001. Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell* **13**: 1959-1968.

Oka T, Nemoto T, Jigami Y. 2007. Functional analysis of *Arabidopsis thaliana* RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *Journal of Biological Chemistry* **282**: 5389-5403.

Palena CM, Tron AE, Bertoncini CW, Gonzalez DH, Chan RL. 2001. Positively charged residues at the N-terminal arm of the homeodomain are required for efficient DNA binding by homeodomain-leucine zipper proteins. *J Mol Biol* **308**: 39-47.

Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF et al. 2009. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324**: 1068-1071.

Ponting CP, Aravind L. 1999. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem Sci* **24**: 130-132.

Pribat A, Blaby IK, Lara-Nunez A, Jeanguenin L, Fouquet R, Frelin O, Gregory JF, Philmus B, Begley TP, de Crecy-Lagard V et al. 2011. A 5-formyltetrahydrofolate cycloligase paralog from all domains of life: comparative genomic and experimental evidence for a cryptic role in thiamin metabolism. *Funct Integr Genomic* **11**: 467-478.

Radauer C, Lackner P, Breiteneder H. 2008. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol Biol* **8**: 286.

Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W. 2003. Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiology* **133**: 1051-1071.

Rerie WG, Feldmann KA, Marks MD. 1994. The *GLABRA2* gene encodes a homeo domain protein required for normal trichome development in *Arabidopsis*. *Genes Dev* **8**: 1388-1399.

Ridley BL, O'Neill MA, Mohnen D. 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* **57**: 929-967.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR et al. 2000. *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **290**: 2105-2110.

Ringli C, Bigler L, Kuhn BM, Leiber RM, Diet A, Santelia D, Frey B, Pollmann S, Klein M. 2008. The modified flavonol glycosylation profile in the *Arabidopsis rol1* mutants results in alterations in plant growth and cell shape formation. *Plant Cell* **20**: 1470-1481.

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao YJ, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A et al. 2007. Genome-wide profiles of *STAT1* DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**: 651-657.

Robzyk K, Kassir Y. 1992. A simple and highly efficient procedure for rescuing autonomous plasmids from yeast. *Nucleic Acids Res* **20**: 3790.

Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, Rajashankar KR, Cohen DE. 2002. Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat Struct Biol* **9**: 507-511.

Rombola-Caldentey B, Rueda-Romero P, Iglesias-Fernandez R, Carbonero P, Onate-Sanchez L. 2014. *Arabidopsis* DELLA and two HD-ZIP transcription factors regulate GA signaling in the epidermis through the L1 box cis-element. *Plant Cell* **26**: 2905-2919.

Roostaee A, Barbar É, Lavigne P, Lehoux J-G. 2009. The mechanism of specific binding of free cholesterol by the steroidogenic acute regulatory protein: evidence for a role of the C-terminal α-helix in the gating of the binding site. *Bioscience Reports* **29**: 89.

Ruberti I, Sessa G, Lucchetti S, Morelli G. 1991. A Novel Class of Plant-Proteins Containing a Homeodomain with a Closely Linked Leucine Zipper Motif. *Embo J* **10**: 1787-1791.

Sahakitrungruang T, Soccio RE, Lang-Muritano M, Walker JM, Achermann JC, Miller WL. 2010. Clinical, genetic, and functional characterization of four patients carrying partial loss-of-function mutations in the steroidogenic acute regulatory protein (StAR). *J Clin Endocrinol Metab* **95**: 3352-3359.

Saleh A, Alvarez-Venegas R, Avramova Z. 2008. An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nature Protocols* **3**: 1018-1025.

San-Bento R, Farcot E, Galletti R, Creff A, Ingram G. 2014. Epidermal identity is maintained by cell-cell communication via a universally active feedback loop in *Arabidopsis thaliana*. *Plant J* **77**: 46-58.

Savaldi-Goldstein S, Peto C, Chory J. 2007. The epidermis both drives and restricts plant shoot growth. *Nature* **446**: 199-202.

Schena M, Davis RW. 1992. Hd-Zip Proteins - Members of an *Arabidopsis* Homeodomain Protein Superfamily. *Proc Natl Acad Sci USA* **89**: 3894-3898.

Schrick K, Bruno M, Khosla A, Cox PN, Marlatt SA, Roque RA, Nguyen HC, He C, Snyder MP, Singh D et al. 2014. Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity. *BMC Biology* **12**: 70.

Schrick K, Nguyen D, Karlowski WM, Mayer KFX. 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol* **5**.

Scott SP, Teh A, Peng C, Lavin MF. 2002. One-step site-directed mutagenesis of ATM cDNA in large (20 kb) plasmid constructs. *Hum Mutat* **20**: 323.

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T et al. 2002. Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J* **31**: 279-292.

Sessa G, Morelli G, Ruberti I. 1993. The Athb-1 and Athb-2 Hd-Zip Domains Homodimerize Forming Complexes of Different DNA-Binding Specificities. *Embo J* **12**: 3507-3517.

-. 1997. DNA-binding specificity of the homeodomain leucine zipper domain. *J Mol Biol* **274**: 303-309.

Sessa G, Steindler C, Morelli G, Ruberti I. 1998. The *Arabidopsis* Athb-8, -9 and -14 genes are members of a small gene family coding for highly related HD-ZIP proteins. *Plant Mol Biol* **38**: 609-622.

Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ et al. 2014. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. *Nature Communications* **5**: 5519.

Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nat Biotechnol* **22**: 1567-1572.

Shen B, Sinkevicius KW, Selinger DA, Tarczynski MC. 2006. The homeobox gene *GLABRA2* affects seed oil content in *Arabidopsis*. *Plant Mol Biol* **60**: 377-387.

Sparkes IA, Runions J, Kearns A, Hawes C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc* **1**: 2019-2025.

Sung MK, Lim G, Yi DG, Chang YJ, Yang EB, Lee K, Huh WK. 2013. Genome-wide bimolecular fluorescence complementation analysis of SUMO interactome in yeast. *Genome Res* **23**: 736-746.

Takada S, Iida H. 2014. Specification of epidermal cell fate in plant shoots. *Front Plant Sci* **5**: 49.

Takada S, Juergens G. 2007. Transcriptional regulation of epidermal cell fate in the *Arabidopsis* embryo. *Plant Cell Physiol* **48**: S18-S18.

Timme RE, Delwiche CF. 2010. Uncovering the evolutionary origin of plant molecular processes: comparison of *Coleochaete* (*Coleochaetales*) and *Spirogyra* (*Zygnematales*) transcriptomes. *BMC Plant Biology* **10**: 96.

Tirajoh A, Aung TS, McKay AB, Plant AL. 2005. Stress-responsive alpha-dioxygenase expression in tomato roots. *Journal of Experimental Botany* **56**: 713-723.

Tiwari SB, Hagen G, Guilfoyle TJ. 2004. Aux/IAA proteins contain a potent transcriptional repression domain. *Plant Cell* **16**: 533-543.

Tominaga-Wada R, Iwata M, Sugiyama J, Kotake T, Ishida T, Yokoyama R, Nishitani K, Okada K, Wada T. 2009. The GLABRA2 homeodomain protein directly regulates *CESA5* and *XTH17* gene expression in *Arabidopsis* roots. *Plant J* **60**: 564-574.

Tron AE, Bertoncini CW, Palena CM, Chan RL, Gonzalez DH. 2001. Combinatorial interactions of two amino acids with a single base pair define target site specificity in plant dimeric homeodomain proteins. *Nucleic Acids Res* **29**: 4866-4872.

Tron AE, Welchen E, Gonzalez DH. 2004. Engineering the loop region of a homeodomain-leucine zipper protein promotes efficient binding to a monomeric DNA binding site. *Biochemistry-Us* **43**: 15845-15851.

Tsiantis M. 2001. Control of shoot cell fate: Beyond homeoboxes. *Plant Cell* **13**: 733-738.

Tsujishita Y, Hurley JH. 2000. Structure and lipid transport mechanism of a StAR-related domain. *Nat Struct Biol* **7**: 408-414.

Vellosillo T, Martinez M, Lopez MA, Vicente J, Cascon T, Dolan L, Hamberg M, Castresana C. 2007. Oxylipins produced by the 9-lipoxygenase pathway in *Arabidopsis* regulate lateral root development and defense responses through a specific signaling cascade. *Plant Cell* **19**: 831-846.

Voinnet O, Rivas S, Mestre P, Baulcombe D. 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* **33**: 949-956.

Wang J, Ji Q, Jiang L, Shen S, Fan Y, Zhang C. 2009. Overexpression of a cytosol-localized rhamnose biosynthesis protein encoded by *Arabidopsis RHM1* gene increases rhamnose content in cell wall. *Plant Physiol Biochem* **47**: 86-93.

Wang S, Chang Y, Guo J, Chen JG. 2007. *Arabidopsis* Ovate Family Protein 1 is a transcriptional repressor that suppresses cell elongation. *Plant J* **50**: 858-872.

Wang X, Wang X, Hu Q, Dai X, Tian H, Zheng K, Wang X, Mao T, Chen JG, Wang S. 2015. Characterization of an activation-tagged mutant uncovers a role of GLABRA2 in anthocyanin biosynthesis in *Arabidopsis*. *Plant J* **83**: 300-311.

Wanner LA, Li GQ, Ware D, Somssich IE, Davis KR. 1995. The Phenylalanine Ammonia-Lyase Gene Family in *Arabidopsis thaliana*. *Plant Mol Biol* **27**: 327-338.

Watt G, Leoff C, Harper AD, Bar-Peled M. 2004. A bifunctional 3,5-epimerase/4-keto reductase for nucleotide-rhamnose synthesis in *Arabidopsis*. *Plant Physiology* **134**: 1337-1346.

Wellesen K, Durst F, Pinot F, Benveniste I, Nettesheim K, Wisman E, Steiner-Lange S, Saedler H, Yephremov A. 2001. Functional analysis of the *LACERATA* gene of *Arabidopsis* provides evidence for different roles of fatty acid omega -hydroxylation in development. *Proc Natl Acad Sci U S A* **98**: 9694-9699.

Wellmer F, Alves-Ferreira M, Dubois A, Riechmann JL, Meyerowitz EM. 2006. Genome-wide analysis of gene expression during early *Arabidopsis* flower development. *PLoS Genetics* **2**: e117.

Weltmeier F, Ehlert A, Mayer CS, Dietrich K, Wang X, Schutze K, Alonso R, Harter K, Vicente-Carbajosa J, Droge-Laser W. 2006. Combinatorial control of *Arabidopsis* proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. *Embo J* **25**: 3133-3143.

Western TL, Burn J, Tan WL, Skinner DJ, Martin-McCaffrey L, Moffatt BA, Haughn GW. 2001. Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in *Arabidopsis*. *Plant Physiology* **127**: 998-1011.

Wilkie AOM. 1994. The Molecular-Basis of Genetic Dominance. *J Med Genet* **31**: 89-98.

Won SK, Lee YJ, Lee HY, Heo YK, Cho M, Cho HT. 2009. cis-Element- and Transcriptome-Based Screening of Root Hair-Specific Genes and Their Functional Characterization in *Arabidopsis*. *Plant Physiology* **150**: 1459-1473.

Wu R, Li S, He S, Wassmann F, Yu C, Qin G, Schreiber L, Qu LJ, Gu H. 2011. CFL1, a WW domain protein, regulates cuticle development by modulating the function of HDG1, a class IV homeodomain transcription factor, in rice and *Arabidopsis*. *Plant Cell* **23**: 3392-3411.

Xu XP, Chen CH, Fan BF, Chen ZX. 2006. Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell* **18**: 1310-1326.

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K. 2008. Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. *Plant Cell* **20**: 2160-2176.

Yu L, Chen X, Wang Z, Wang S, Wang Y, Zhu Q, Li S, Xiang C. 2013. *Arabidopsis* enhanced drought tolerance1/HOMEODOMAIN GLABROUS11 confers drought tolerance in transgenic rice without yield penalty. *Plant Physiology* **162**: 1378-1391.

Zalewski CS, Floyd SK, Furumizu C, Sakakibara K, Stevenson DW, Bowman JL. 2013. Evolution of the class IV HD-zip gene family in streptophytes. *Mol Biol Evol* **30**: 2347-2365.

Zhang F, Zuo K, Zhang J, Liu X, Zhang L, Sun X, Tang K. 2010. An L1 box binding protein, GbML1, interacts with GbMYB25 to control cotton fibre development. *Journal of Experimental Botany* **61**: 3599-3613.

Zhang X, Liu CJ. 2015. Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Mol Plant* **8**: 17-27.

Zhang XK, Lehmann J, Hoffmann B, Dawson MI, Cameron J, Graupner G, Hermann T, Tran P, Pfahl M. 1992. Homodimer Formation of Retinoid X-Receptor Induced by 9-Cis Retinoic Acid. *Nature* **358**: 587-591.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Zhou JL, Lee CH, Zhong RQ, Ye ZH. 2009. MYB58 and MYB63 Are Transcriptional Activators of the Lignin Biosynthetic Pathway during Secondary Cell Wall Formation in *Arabidopsis*. *Plant Cell* **21**: 248-266.

Chapter 4 - Characterization of protein-metabolite interactions for the START domain

ABSTRACT

Steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domains were first characterized in the mammalian StAR-related protein that binds and transfers cholesterol to initiate steroidogenesis. Identified ligands for mammalian proteins include sterols such as cholesterol and 25-hydroxycholesterol, phosphatidylcholine, phosphatidylethanolamine, ceramides, bile acids, and steroid hormones. Although START domains are abundant in the plant kingdom, specific ligands for plant-derived START domains from HD-Zip transcription factors are not reported to date. In this chapter, parallel *in vitro* and *in vivo* approaches were used to identify the START domain ligands from the class IV HD-Zip transcription factors GL2 and PDF2. Using a fluorescence-based thermal shift assay, a library of ~1000 pure natural compounds from plants and microbes was screened for PDF2 binding compounds. The assay identified eight hits that thermally stabilized both the full-length PDF2 and PDF2 protein containing the START+SAD domains. Among these, some stabilized the protein in a concentration-dependent manner. Two compounds appear to enhance the soluble expression of PDF2-START protein. When utilized as a culture additive in an *E.coli* expression system, the flavonoids naringin and isoliquiritigenin increased the proportion of recombinant protein expressed in soluble form. Further, IP-based pull down interfaced with QTOF-MS using zipper domain deleted GL2 (gl2ΔZip) protein identified masses consistent with fatty acids, including colnelenic acid as putative class IV HD-Zip TF binders. Collectively, this work has led to the identification of several putative ligands for HD-Zip derived START domains.

INTRODUCTION

Experimental evidence for START domains as a lipid/sterol ligand-binding motif derives exclusively from mammals and the silkworm *Bombyx mori*. The ligands identified for nine human START domains include sterols (4/9), phospholipids (3/9), a ceramide (1/9), bile acids (1/9), and a carotenoid (1/9). STARD1 and MLN64/STARD3 have both been shown by *in vitro* sterol binding assays to bind cholesterol (Tsujishita and Hurley 2000). In contrast to selective cholesterol binding by STARD1 and STARD3, STARD4 also binds 7α-hydroxycholesterol and 7-hydroperoxycholesterol (Rodriguez-Agudo et al. 2008), whereas STARD5 is also reported to bind 25-hydroxycholesterol (Rodriguez-Agudo et al. 2005). Contradicting these observations, using CD, ITC, and NMR-based approach; Letourneau et al. (2013) demonstrated that STARD5 specifically binds bile acids rather than cholesterol. A recent study reported preferential binding of STARD6 to the steroid pregnenolone (Letourneau et al. 2015). Ligand binding specificities for phospholipids have been found for all three members of the PCTP group. Mammalian PCTP/STARD2 and STARD7 both bind the phospholipid PC (Roderick et al. 2002; Horibata and Sugimoto 2009), while the closely related protein, STARD10, was shown to bind both phosphatidylcholine (PC) and phosphatidylethanolamine (PE) (Olayioye 2005). Moreover, CERT/STARD11, a splice variant of the Goodpasture antigen-binding protein specifically binds ceramide (Hanada et al. 2003). Intriguingly, the carotenoid binding protein (CBP1), a silkworm ortholog of STARD3, was demonstrated to bind carotenoid lutein, implying that START domain of STARD3 can bind different ligands. However, unlike STARD1 and STARD3, which possess a mitochondria- and an endosome-targeting signal, respectively, CBP appears not to encode additional motifs (Tabunoki et al. 2004).

Although START domains are amplified in plants (Schrick et al. 2004), the specific ligands for the plant-derived START domains from HD-Zip TFs remain largely unknown. The functional tests described in Chapter 3 suggest that ligand binding is critical to HD-Zip transcription factor activity; however, without knowing the identity of the endogenous ligand, it is not possible to rule out other explanations. This chapter describes parallel *in vitro* and *in vivo* approaches aimed at the identification of the START domain ligands from HD-Zip transcription factors.

Differential scanning fluorimetry

Protein thermal stability testing, also known as thermofluor (TF) or differential scanning fluorimetry (DSF) (Ericsson et al. 2006; Niesen et al. 2007), is based on the measurement of the increase in the thermal stability of the protein induced by ligand binding. The approach consists of applying gradually increasing the temperature on purified proteins, in the presence and absence of a ligand. The protein melting temperature is defined as (Tm), which is known as the midpoint of the protein unfolding position. The Tm value (the inflection point of the transition curve, Figure 4.1), can be calculated using the following equation, derived from the Boltzmann equation:

$$y = LL + \frac{(UL-LL)}{1+exp\left(\frac{Tm-x}{a}\right)}$$

where "LL" and "UL" are the values of minimum and maximum intensities, respectively, "a" denotes the slope of the curve within Tm , "y" and "x" denote the fluorescence intensity and the temperature, respectively (Cummings et al. 2006a; Vedadi et al. 2006; Niesen et al. 2007). At lower limits of fluorescence (LL) the protein exists in its native state and at the upper limit (UL) it becomes fully denatured; the difference in fluorescence between the LL and UL should be high in order to clearly resolve protein unfolding and is strongly dependent on the quantum yield of the employed dye. The inflection point of the transition curve denotes the largest change in unfolding and is arbitrarily set as the melting temperature (T_m) of the protein. Therefore, the simplest way to calculate T_m values is to determine the minimum of the negative first derivative of the melt curve (-d(RFU)/dT, Figure 4.2).

In the DSF approach (Figure 4.3) an "environmentally sensitive" dye, which specifically interacts with non- native protein, is used as an indicator to observe the amount of unfolded protein in solution as a function of temperature (Cummings et al. 2006a; Mezzasalma et al. 2007). This is achieved by measuring the fluorescence from the dye that results from the changes in the dye's emission properties upon interaction with unfolded protein (Vedadi et al. 2006). DSF can be done with various fluorescent dyes, which differ in their optical properties. The dyes fluorescence is quenched in an aqueous solution, whereas, in the non-polar environment, such as the hydrophobic sites on unfolded proteins, the fluorescence intensity becomes significantly higher. With a high signal to noise ratio, SYPRO Orange, the molecular structure of which is not disclosed is currently the dye with most favorable properties for this

approach. It has a relatively high wavelength for excitation (near 500 nm) which reduces the chances of any small molecule causing its fluorescence intensity to quench and, therefore, interfering with its optical properties (Boivin et al. 2013). The emission wavelength of SYPRO Orange is near 600 nm. Some proteins may show no unfolding transition when analyzed using SYPRO Orange, in which case an alternative dye should be tested. Interestingly, the molecular structure of SYPRO Orange is not disclosed.

Most ligands that bind specifically to the native protein will increase the T_m, and a temperature shift (ΔT_m) between the protein's melting temperatures in the two conditions is observed and measured. The extent of this shift is believed to be proportional to the ligand's binding affinity for the protein. Ligands that bind non-specifically are associated with aggregation, which destabilizes the protein tertiary structure and do not test positively in DSF (McMahon et al. 2013). The stability of a protein is related to its Gibbs free energy of unfolding (ΔG_u), which is temperature-dependent and in an idealistic model, the fluorescent dye would exclusively bind to the unfolded parts of the protein (Bruylants et al. 2005). The stability of most proteins decreases with increasing temperature and so the ΔG_u will reach zero at the equilibrium where the concentration of folded and unfolded protein are equal. This temperature is considered as the melting temperature, T_m, and if the protein unfolds in a reversible two- state manner, the equilibrium thermodynamics model will apply. Therefore, a compound that binds to the protein will contribute the free energy of binding (ΔG_b) and usually increase ΔG_u, which may cause an increase in the melting temperature, ΔT_m. It is well known that the stabilizing effect of ligands is proportional to their *in situ* concentration and affinity for the protein. Therefore, plotting ΔT_m against concentration (titration curve, Figure 4.4) provides information about the binding affinity of the compound.

Several articles have demonstrated the use of DSF in screening of small molecules to discover ligands (Parks et al. 2005; Cummings et al. 2006b; Abad et al. 2008; Krishna et al. 2013) for purified protein targets. Most notable are the identification of known natural ligands for estrogen receptor (ER-α and ER-β) from a library of molecules, demonstrating the utility of this assay in characterizing natural ligands (DeSantis et al. 2012; Filgueira et al. 2014). DSF was used to study the signaling events in binding of Abscisic acid (ABA) to receptors PYL1 and PYL2, ABA-induced complex formation of PYL–ABA–PP2C and also to screen and identify SnRK2 inhibitors (Soon et al. 2012). PYL/PYR are members of START cyclase

superfamily that are known to bind ABA endogenously (Park et al. 2009). While DSF provided results that support published data such as PYL1 and PYL2 bind pyrabactin at a lower affinity than the natural ligand ABA, it also provided new knowledge regarding the different thermal stability statuses of SnRK2.2, SnRK2.3, and SnRK2.6. Moreover, the study demonstrated the usefulness of this technique to detect and characterize very weak ($K_D \geq 50$ μM) interactions between receptors and physiological and synthetic agonists.

Chapter 4 demonstrates the utility of DSF assay in screening a natural library for compounds that bind and stabilize PDF2, a representative class IV HD-Zip protein.

Immunoprecipitation based small molecule pull down assay

A number of different types of molecular interactions enable life. These include the interactions between proteins, proteins and nucleic acids, and proteins and metabolites. Elucidating these interactions and understanding how they control biology is an important scientific goal. In recent years, a growing number of approaches have been developed to identify protein-protein interactions (Gingras et al. 2007; Frei et al. 2012) and protein-nucleic acid interactions (Ingolia et al. 2012). While many methods exist for the characterization of biopolymer interactions, far fewer approaches exists to elucidate interactions between proteins and small molecules. Understanding the interactions between small metabolites and proteins can be approached from different perspectives and is important for the advancement of basic science and drug development.

Affinity-based experiments using immobilized proteins are commonly used for detection of protein-metabolite interactions (PMIs). Coupled to mass spectrometry, these approaches allow for the screening of a larger pool of metabolites, including novel metabolites, and can result in the unbiased identification of protein metabolite binding. Importantly, the use of global metabolite profiling enabled the development of a novel, unbiased strategy for the identification of endogenous PMIs (Tagore et al. 2008; Tagore et al. 2009). The method was developed using three different lipid-binding proteins with known ligands: cytosolic retinoic acid- binding protein 2, fatty acid-binding protein 2, and STARD3. The strategy successfully identified specific PMIs for all three of these proteins and, in doing so, created a reliable method to pull down protein metabolite interactions.

The one disadvantage of these methods is the requirement of sufficient quantities of recombinant protein that is immobilized on a solid support. In IP- based small molecule approach, tagged proteins are purified from plant samples using affinity purification. The bottleneck of having sufficient active protein is overcome by using *Nicotiana benthamiana* as an expression system. In comparison with the commonly used heterologous *E. coli* systems, for some proteins higher expression levels are obtained and more importantly, the eukaryotic post-translational modifications are preserved in tobacco (Cunningham and Porter 1998). Furthermore, *N. benthamiana* expression system has proven reliable for the production of large-scale protein arrays (Giddings 2001; Hood 2002; Fischer et al. 2004; Li et al. 2004; Varsani et al. 2006; Liu et al. 2007). Most importantly, no isotope or fluorescent labeling is needed for the ligands, and this allows natural native ligand mixture screening. There are no protein size limitations nor is immobilization on a surface needed. The QTOF-MS-based detection method used in this study is sensitive, allowing detection of unknown bound ligands with only a limited amount of purified protein. Here, this methodology was applied to detect novel protein-interacting ligands for GLABRA2.

RESULTS

Purification of 6xHis-tagged PDF2 protein

Since previous attempts to purify GST-tagged PDF2-START proved problematic (Appendix C), it was sought to examine the expression of recombinant PDF2 fused with a rather smaller but very common 6x-His tag. pNIC28-Bsa4 vector, which includes a TEV protease cleavage site following the N-terminal 6x-His tag, has been used successfully to purify and crystallize mammalian START domains (Thorsell et al. 2011). For this reason, it was chosen for PDF2 expression and purification. 6xHis-tagged full length PDF2 (85 kDa), PDF2-START+SAD (59 kDa), PDF2-START (29 kDa), full-length GL2 (85 kDa), and GL2-START+SAD (59 kDa) constructs were generated and subjected to expression testing in *E.coli*. The expression scale chosen was 500 ml terrific broth (TB), and expression was induced with 0.1 mM IPTG for 20 hours at 18˚C. All three constructs were found to be expressed at high levels (Figure 4.5 A). Lysates were obtained by sonication and subjected to Talon metal affinity resin purification. The eluates were analyzed by SDS-PAGE. Compared

with PDF2-START, both PDF2 and PDF2-START+SAD proteins demonstrated a high level of soluble expression (Figure 4.5 B, C, D). The PDF2 and PDF2 (START+SAD) proteins were eluted with 50 mM imidazole, in contrast to 500 mM imidazole required to elute PDF2-START domain protein. The protein appeared to be relatively pure as judged by SDS-PAGE with moderate yields (~1.5-2 mg purified protein from 0.5 L culture) as determined by the Bradford Assay. The identity of the protein was confirmed by mass spectrometry (Figure 4.6). The analysis revealed no significant matches to *E.coli* proteins, while a PDF2 only search gave a high score (4015) with the N-terminal-most peptide detected at position 42 in the full construct sequence (STSDN…), suggesting partial degradation. Additionally, the full-length GL2 (85 kDa) and GL2-START+SAD (59 kDa) protein constructs were also examined. Unlike PDF2, both the full length and START+SAD versions of GL2 appeared to be insoluble when purification was attempted from *E.coli* BL21(DE3) pRARE cells (data not shown).

Optimization of differential scanning fluorimetry (DSF) assay conditions

Subsequent to purifying correctly folded PDF2 and PDF2-START+SAD proteins, a thermal shift assay was used to identify potential ligands for the START domains. As introduced above, DSF provides a robust, unbiased approach to rapidly identify putative ligands.

Although the thermal stability protocol presented in the literature (Niesen et al. 2007) is suitable for a large number of proteins, some fine-tuning is necessary to optimize the signal/noise ratio and increase accuracy in the measurement of the T_m for a specific protein. Some parameters can be optimized such as the concentration of protein and SYPRO Orange dye. Kean and coworkers tested a range of SYPRO Orange concentrations and found that the amplitude of signal increases linearly with dye concentration. However, they also observed that SYPRO Orange has a destabilizing effect at high concentrations (>20x) (Heinz et al. ; Kean et al. 2008).

Here, four different SYPRO Orange dye concentrations were tested (2x, 5x, 10x, 20x) to determine optimal conditions for the thermal denaturation assay, such as which protein/dye ratio gives higher signal to noise ratio, sharper unfolding transition, and lower background signal. The structural formula and exact concentration is not disclosed by the manufacturer. The dye is supplied as a 5000× concentrated solution dissolved in 100% (v/v) DMSO, so it's

concentration should be kept as low as possible in order to prevent DMSO damage to the protein. 10x dye was found to give a significant gain in fluorescence without damaging the protein (Figure 4.7 A). The dye was diluted in a buffer prior to the addition of the protein to prevent exposure of the protein to localized high concentrations of DMSO, in agreement with the protocol in Niesen et al. (2007). After an appropriate dye concentration had been selected, protein concentration was varied (1 μM, 5 μM, 10 μM) to identify a suitable compromise between maximizing fluorescence gain and minimizing the amount of protein used. 1 μM protein concentration was found to be sufficient to confer an increase in fluorescence intensity to allow accurate determination of T_m (Figure 4.7 B). Sorrell et al. (2010) showed that at high DMSO concentrations the stability of the protein was affected, leading to a negative shift in the T_m and a high initial fluorescence intensity reading, indicating that the protein is destabilized at high DMSO concentrations, relative to the reference. Thus, it is recommended to keep the DMSO concentration less than 2%. Here, 1.5% was used in the final assay. Since protein unfolding is a cooperative process, the unfolding of a small protein region will induce the immediate unfolding of the remaining protein core. Thus, an optimal protein stabilization buffer should result in a sharp and fast thermal denaturation transition between folded and unfolded states, detected through high transition slopes, in parallel with a higher T_m. A sharp melting transition, indicative of a high quality, stable protein preparation was observed for PDF2 (Figure 4.8).

Validation of the DSF assay

For validation of the DSF protocol, an experiment from the literature that shows binding of Pyrabactin to PYL2 receptor was successfully reproduced (Figure 4.8). Some additional compounds, including lipids, hormones, and flavonoids were also checked for their ability to bind and stabilize the PDF2 protein. Barring the two flavonoids; Kaempferol 3,7 dirhamnoside (f1) and Quercetin 3,7 dirhamnoside (f2), none of the compounds showed an effect on protein stability (T_m<2˚C) (Figure 4.9). In comparison with f1, f2 displayed a higher ΔT_m value. To study the effects of f2 concentration on the T_m of PDF2, increasing amounts of f2 were added to the protein. Theoretical DSF dose-response curve showed a sigmoidal dependence of ΔT_m on ligand concentration with a binding constant (K_D) of 12.5 μM (Figure 4.10)

Figure 4.11 illustrates the responses from the three controls used in the DSF experiments. Positive control protein (0.01 mg/ml lysozyme) gives a clear signal (blue curve) in the range of 3000 ΔRFU and a T_m value of 71 °C, consistent with a previous report (Knubovets et al. 1999). The two negative controls (protein-only and dye-only) (red and green respectively) returned flat curves as expected. These negative controls serve as a valuable indicator for contaminated dye or PCR plates, naturally fluorescent proteins.

High-throughput screens of 1000 compounds against PDF2

After optimization of the conditions of the assay, 1000 compounds from the AnalytiCon MEGx library were tested in a 96-well plate format for their ability to bind PDF2 at a single concentration of 300 μM. The average T_m value in the four reference wells or vehicle control (PDF2 protein + 1.5% DMSO) was approximately 41.7 °C.

Based on their effects on the protein thermostability, the compounds tested were classified into four groups: protein stabilizers, destabilizers, denaturants and those, which show no effect on protein stability (Table 4.1). The first group of compounds, which is of the most interest, gave a positive thermal shift, indicating an increase in protein stability. This suggests specific binding to PDF2, increasing the protein structural order and reducing its conformational flexibility. With an initial hit cut-off value of ΔT_m >2˚C, a total of 34 out of the 1000 compounds tested showed a positive thermal shift ranging from +2.2 °C to as high as 18 °C. Detection of 34 putative binders (Table 4.1) from the initial screen corresponded to 3.4% primary hit rate.

For the second group of compounds, there was no apparent significant shift in T_m, indicating no effect on the protein's thermal stability, presumably due to lack of binding. An example of the thermal shift of this second group is shown in Figure 4.12 (A). Compounds in this category were therefore not selected as likely hits in the DSF screening stage.

The third group of compounds gave a negative (ΔT_m), indicating a decrease in protein thermal stability. In total, 88 out of the 1000 compounds tested displayed a negative thermal shift. This could be a sign of misfolding or might suggest that the protein structural changes were induced towards a more disordered conformation, destabilizing the tertiary structure. Non-specific binding i.e. away from protein binding sites is usually associated with such effect (Cummings et al. 2006b; Ericsson et al. 2006). Non-specific binders have no direct effect on

receptor activity, typically binding receptor surface regions, such as hydrophobic patches on the protein surface (Lepre et al. 2004). Many non-specific binders have been found to be promiscuous and act as aggregates. These aggregates were proposed to either adsorb or absorb proteins (McGovern et al. 2002). Thus, these compounds do not form classical inhibitor−enzyme complexes, which promote proteins stability, and results in an increase of T_m. They rather induce the protein towards a more disordered conformation, through their promiscuity and adsorption to hydrophobic patches on protein surfaces, thereby destabilizing the protein. An example of the thermal shift of this third group is shown in Figure 4.12 (B). Cryptopimaric acid (diterpenoid) and other compounds with a similar destabilizing effect on PDF2 were excluded from the secondary screen as such an effect is highly likely to be associated with the non-specific binding.

Some of the prominent features identified in the list of destabilizers include the presence of additional elements such as nitrogen and sulfur in empirical formula, unusually high molecular weight (>900 Da), members of diverse terpenoid family, long chain fatty acids, and antibiotics derived from microorganisms. None of the positive hits displayed such characteristics. The rest of the compounds did not appear to follow any particular pattern.

The fourth group of compounds gave no measurable transition (no peak observed). This is an indication of the potential aggregation/complete denaturation of the protein or its possible destabilization/ partial unfolding (Ericsson et al. 2006). An example of the thermal shift of this fourth group is shown in Figure 4.12 (C) Glycerin Monoisomyristate and other compounds with similar destabilizing effect on PDF2 were also excluded from the screen as such an effect is highly likely to be associated with non-specific binding and does not aid or support future structural studies.

Next, to eliminate false positives and improve confidence in the measured values, the primary hits (34/1000) were retested in triplicate at 300 μM. Visualization of the melting temperature shifts was plotted as observed in Figure 4.13. Of these 34 compounds, 8 were discounted upon retesting as they displayed $\Delta T_m \leq 2.0^{\circ}C$ or the compound exhibited fluorescence values higher than buffer background when screened with dye at standard ligand concentrations. ΔT_m values for the remaining 25 compounds (Figure 4.13, Table 4.2) were calculated after subtracting ligand background fluorescence from the experimental fluorescence values. The extent of the thermal shift (ΔT_m = T_m reference - T_m compound) as

calculated in GraphPad is believed to be proportional to the ligand's binding affinity for the protein (Cummings et al., 2006). Strikingly, the majority of the compounds (15/25) belong to the flavonoid family, followed by stilbenes (2/25), curcuminoid (1/25), and benzyl benzoate (2/25). Benzylated chalcone and pinoquercetin (Figure 4.18, Table 4.2) caused the largest thermal shifts of the library, with remarkable ΔT_m values of + 18.2 ˚C and + 10.6 ˚C, respectively.

To map the ligand-binding domain of PDF2, the positive hits were tested in triplicate against the PDF2-START+SAD protein. Unfortunately, due to relatively low soluble protein yields the assay could not be performed using the PDF2-START domain alone. Out of 25 compounds, 8 were found to induce a positive shift in the T_m of both PDF2 and PDF2-START+SAD proteins, indicating formation of more stable complex that unfolds cooperatively when subjected to thermal melting (Figure 4.14). Eight compounds exhibiting ΔT_m values ranging from 2.35-16.1˚C at 300 μM are listed in Table 4.3. Their chemical structures are depicted in Figure 4.15, and representative normalized melting curves are shown in Figure 4.16 A. All eight compounds belong to the flavonoid family. While addition of isoliquiritigenin (2) and benzylated chalcone (3) to the protein yielded smooth single peak profiles, a biphasic melting profile was observed for licochalcone B (1), pinoquercetin (4), quercetin 4'-O-β-D-glucopyranoside (5), quercetin 3- β-D-mannopyranoside (6), and (2Z)-6-Hydroxy-2-(4-hydroxy-3-methoxybenzylidene)-1-benzofuran-3(2H)-one (7). The two peaks reflect the presence of more stable protein-ligand complex and excess of unbound ligand. As a negative control for ligand binding studies, several positive hits were screened against the PYL2 protein. While most of them induced no apparent significant shift in T_m, pinoquercetin had a destabilizing effect on protein stability (Figure 4.16 B)

In the model plant *Arabidopsis thaliana*, at least 54 flavonoid molecules (35 flavonols, 11 anthocyanins, and eight proanthocyanidins) are reported (Mouradov and Spangenberg 2014). Scaffold structures of flavonoids in *Arabidopsis* are relatively simple that includes kaempferol, quercetin and isorhamnetin for flavonols (Figure 4.17). The chemical diversity of flavonoids has been shown to increase enormously by tailoring reactions, which modify these scaffolds, including glycosylation, methylation, and acylation. Consistently, relatively simple kaempferol and quercetin di-glycosides (di-rhamnosides or glucoside/rhamnoside in -3-O- and -7-O- positions) and tri-glycosides (di-rhamnosides and monoglucoside in -3-O-, -2''-O- and -

7-O- positions) are commonly distributed in all tissues of wild-type *Arabidopsis*. Since the top ligand hits include flavonoid compounds, structural analogs that were readily available were also tested.

All five compounds tested positive against PDF2 (START+SAD) protein while only quercetin and quercetin 3-β-D-glucoside induced a positive shift in the T_m of full-length PDF2 (Figure 4.18). No background fluorescence was observed for all five compounds. Ligand binding data for butein, a structural analog of isoliquiritigenin could not be obtained, as compound precipitated in analysis buffer at 300 μM test concentration.

Concentration-response curves

To further test the interactions, titration of the compound with increasing concentrations from 1 to 300 μM was performed. It was of interest to determine whether the stabilization of protein observed in the initial DSF screen was dependent on the compound concentration. Notably, several types of behavior were observed (Figure 4.19). For some compounds hits such as pinoquercetin, the stabilizing effect was saturable, that is, T_m increased with compound concentrations and reached a maximal value, indicating specific binding (K_D= 12 μM). T_m increases were not detectable at 1:1 stoichiometric protein:pinoquercetin ratios, but rather required high ligand concentration. Naringin showed a marked jump to a large ΔT_m at high concentrations, but was barely or not at all stabilizing at lower concentrations, while quercetin, isoliquiritigenin, and benzylated chalcone showed a linear response with increasing ligand concentrations. One possibility is that this may represent weak binding that requires very high ligand concentrations to reach saturation. However, limits on ligand solubility and requirement to keep DMSO concentration to less than 2% precluded additional experiments involving ligand concentrations higher than 300 μM. Alternatively, this could correspond to non-specific binding. Among all compounds tested, the binding of benzylated chalcone was strongest (Figure 4.19 B). However, contradicting our previous results, it is intriguing that both licochalcone B and naringenin failed to bind and stabilize the protein, resulting in a flat curve. Quercetin 3-β-D-glucoside, on the other hand, displayed weak non-specific binding (ΔT_m at highest ligand concentration was less than 2˚C).

Ligand supplementation as a method to validate putative ligands

There is a substantial literature that shows that addition of ligand during protein expression can result in increased levels of protein expression and significantly increase the proportion of recombinant protein expressed in the soluble form (reviewed in Hassell et al. (2007); (Hozjan et al. 2008). This method was utilized to validate the putative ligands identified by the DSF assay. As mentioned above, the heterologous production of His_6-tagged PDF2-START protein is possible, but yields only small amounts of soluble protein (Figure 4.20 A). The effect of ligand supplementation on soluble protein expression of PDF2-START was tested with 2 compounds. Naringin (#71162, Sigma-Aldrich) and isoliquiritigenin (AK Scientific, Inc., Union City, CA) were selected as culture additives for the expression trials, for the following reasons: First, these compounds are commercially available at a low cost and secondly, we identified them as protein stabilizers in the thermal shift assay. The expression scale chosen was 100 ml TB medium, to detect even low-expressing proteins (see Materials and Methods). The experiment was performed twice; a representative SDS-PAGE result is displayed in Figure 4.20 B. The yield of soluble PDF2-START improved markedly by the inclusion of 100 μM Naringin during protein expression and purification. On the contrary, including naringin dihydrochalcone (negative in DSF assay) had a negligible effect on soluble expression, since almost the whole amount of the recombinant protein was insoluble. These results clearly demonstrate that naringin can enhance soluble expression, most likely by inducing a more stable and correctly folded state of the target protein that is less prone to aggregation.

In vivo ligand pull-down assays from plants to discover natural START domain ligands from HD-Zip transcription factors

To determine the *in vivo* binding partners, IP-based small molecule pull down approach was utilized. As observed in Figure 4.21 A, this technique combines high-affinity protein purification using an HA-tag from *N. benthamiana Agrobacterium* infiltrated leaves, with small molecule pull-down and QTOF-MS detection. A protocol optimized from the unpublished method developed in Sean Cutler's lab was used. Briefly, the tagged protein is introduced into tobacco leaves, resulting in the formation of a protein-metabolite complex

between the protein and its natural binding partner. Following immunoprecipitation (IP) to isolate a single protein-metabolite complex from the proteome, the bound metabolites are eluted and analyzed by QTOF to identify bound ligands. Comparison of experimental samples to pull downs from no protein negative control is used to reveal ligands specifically associated with the protein of interest.

Small molecule pull down using HA tagged gl2ΔZip protein identifies fatty acids as putative ligands

The gl2ΔZip protein N-terminally fused to an HA affinity purification tag was transiently expressed in *N.benthamiana* leaves. Control infiltration of empty HA-tagged vector was performed for the negative control. 10 g of leaf material were harvested, and HA-affinity tag based purification was performed using anti-HA antibody bound to Dynabeads Protein G (Invitrogen). Protein expression was confirmed by Western blot analysis. Bound metabolites were extracted in Chloroform:Methanol:Water, and neutralized with ammonium acetate prior to the analysis by Quadrupole-time of flight tandem mass spectrometer instrument (QTOF-MS). The experiment was performed twice.

Samples were scanned for 10 min (~600 scans) in both positive and negative mode. There was not much difference between control and experimental samples in positive mode. However, in the negative mode, the experimental sample seemed to show an increase in a 291.1946, 293.2129, 277.2184, and 279.2137 peaks. These may correspond to 1) 291.1946; 18:4-O fatty acid such as 12-oxo phytodienoic acid (OPDA) (theoretical m/z ~291.1966), 2) 293.2129; 18:3-O fatty acid (theor m/z ~293.2137), 3) 277.2184; 18:3 linolenic fatty acid, and 4) 279.2137; 18:2 linoleic fatty acid. To characterize each peak fully, four precursor ion peaks were chosen for product ion scanning in tandem MS/MS mode, and the mass spectrums are shown in Figure 4.22. For product ion scanning, it is ideal to fragment the product ion down as much as possible to smaller fragments but still be able to see the product ion. Increasing or decreasing the collision energy alters the fragmentation level. Due to the limited sample amount, collision energy CE-30 was chosen for products of the 277 peak; as higher collision energies removed the signals completely. This collision energy was then used throughout. While 199 peaks in the product scans appeared to match the formula for 12:0 fatty

acid, and the 171 peak matches 10:0, the other fragment ions could not be identified, thereby precluding the confirmation of ions detected in the full scan mode. Since the negative ion fragmentation spectra data is not publicly available for many fatty acids, product scans were performed for fatty acid standards of m/z 291, 293, 277, and 279 (corresponding to OPDA, 13-KOT, colnelenic acid, etherlonic acid, 13(S)-HPOT, 13(S)-HOT, 9(S)-HOT, and linolenic acid respectively) using identical fragmentation conditions as before, 100 cycles and a collision energy of -30, and compare with product ion mass spectra of our samples. Stearic acid, which produces an $(M - H)^-$ parent ion of *m/z* 283.3 and palmitic acid, which produces an $(M - H)^-$ parent ion of *m/z* 255.2 in negative ion mode served as negative controls. The product scans of fatty acid standards did not exactly match the product ion mass spectra of the samples, although the fragmentation pattern of colnelenic acid (m/z 291.26) seemed to show an increase in 171.1039 peak, that resembled 171.1035 product ion peak of m/z 291.2. Thus, the ion at m/z 291.2 appeared to match to colnelenic acid, suggesting that the long chain fatty acid was bound to the HA-tagged gl2ΔZip protein in the pull down experiment.

DISCUSSION AND FUTURE PERSPECTIVES

Structural analysis of the START domain from HD-Zip TFs

A recent survey of the protein data bank has revealed that the crystal structures of eight mammalian START domains have been solved so far. These START domains are those of STARD2 (Roderick et al. 2002), D3 (MLN64, Tsujishita and Hurley (2000)), D4 (Romanowski et al. 2002), D11 (Kudo et al. 2008), and D1, D5, D13, and D14 (Thorsell et al. 2011). STARD3, D4 and D5 have specificity for cholesterol. STARD2/PCTP binds phosphatidylcholine (PC), and STARD11 (CERT) binds ceramides. The lipid specificity of STAR13 (DCL-2) is not known as of yet.

Structural analyses of START domains from mammals have provided detailed insights into how these proteins sequester specific lipids (summarized in Table 1.1). Apart from deletion and insertions in loops, the structures of all the START domains are very similar. The ~210 residue globular START module is a curved β-sheet gripped by two α-helices, known as α/β helix-grip-fold (Alpy and Tomasetto 2005). One striking and common feature of these structures is the presence of an internal hydrophobic cavity large enough to fit one lipid molecule.

Notably, in animals START domains are not found in transcription factors, whereas in plants START is primarily associated with HD-Zip TFs (Schrick et al. 2004). To achieve headway in understanding its function, the structure of the START domain needs to be solved, and natural ligands identified. Indeed, using insights gleaned from high-resolution structures, numerous studies have characterized ligand-binding pockets and shed light on the underlying molecular mechanisms. For example, structural work by Kudo et al. (2010) led to a mechanism by which CERT can distinguish ceramide from other lipids yet still recognize multiple species of ceramides. Similarly, seminal studies on structure of PYR/PYL family of START proteins provided breakthrough insights into previously elusive mechanism of Abscisic acid signaling (Santiago et al. 2012).

In absence of structural data, homology modeling was used as a means to obtain a reasonable 3D model as described in Chapter 2. This work predicted similar sized ligand-

binding cavities of a subset of plant START domains in comparison to mammalian counterparts.

The foremost step in crystal structure determination entails obtaining a highly purified protein in sufficient quantities. In this study, two different *E. coli* fusion expression strategies (the N-terminus GST- and the His_6-tagged expression) were applied under low temperature conditions intending to increase their efficiency for producing soluble recombinant START domain. Insolubility is a typical problem for recombinant proteins expressed in *E. coli*. The reason for this is believed to be their non- native, aggregation-prone conformation and their subsequent expression as inactive forms in inclusion bodies. Important parameters include the interactions between the hydrophobic patches of newly synthesized unfolded polypeptides, which are influenced by the rate of protein synthesis along with the cellular microenvironment of expression. Therefore, factors to consider are usually related with the expression vector used, the induction parameters, cultivation conditions, and fusion partner proteins (Sahdev et al. 2008).

Selection of the appropriate tag for the fusion of the heterologous protein usually depends on the desired method of chromatography and the experimental needs of the protein's utilization. On the other hand, protein solubility depends on the type of the fusion tag to be used. Tags known for their contribution to solubility are the MBP, the NusA and the GST (glutathione S- transferase) (Costa et al. 2014). However, expression of GST-tagged START domains typically resulted in poor solubility (Appendix C). The START domain from PDF2 showed the most promising levels of protein in the soluble fraction. As shown in Appendix Figure C.3, the efforts to upscale the purification led to sufficient quantities of the purified GST:PDF2-START and analysis using ANS binding assay suggested that the protein is correctly folded.

The removal of the fusion partner from the final protein is often necessary because the tag can potentially interfere with the formation of crystals (Waugh 2005; Young et al. 2012). In spite of being widely employed, the removal of fusion partners presents difficulties. It has been shown that recognition of linear amino acid sequence (except for SUMO protease) can result in unspecific cleavage (Butt et al. 2005). Following tag removal, it is not uncommon to obtain low protein yields and poor recovery of active target proteins due to protein precipitation and aggregation (Waugh 2011). Moreover, tag removal often involves

high costs of proteases and tedious optimization of cleavage conditions (Smyth et al. 2003). Consistent with these observations, removal of the GST tag substantially reduced yields of PDF2-START, likely due to protein precipitation and aggregation (Figure 4.7). It was previously shown that target proteins fused to GST can be successfully crystallized (Zhan et al. 2001). Thus, one possibility is to test whether the uncleaved GST:PDF2-START protein can be purified to homogeneity to yield crystals.

Expression of the fused His_6-tagged PDF2:START resulted in poor solubility even at low temperature culture conditions, rendering it unsuitable for purification (Figure 4.5). This is not a surprising result since unlike GST; His_6 is not a solubility enhancing tag. It is commonly employed as an affinity tag and offers several advantages of its own (Li 2010). First, the hexahistidine tag has a small size and charge that rarely interferes with protein function and structure. Second, His tags can be used under both native and denaturing conditions. In addition, oligohistidine-tagged proteins can be eluted under mild conditions by imidazole competition or low pH. Several studies have reported successful crystallization of uncleaved proteins (Carson et al. 2007; Kimple et al. 2013).

Production of soluble protein is not straightforward for START domain proteins since aggregation was observed for most of the HD-Zip-derived START domain sequences using *E. coli* expression. Even under optimal condition, it is unusual to generate active soluble protein on the first attempt. Accordingly, it is important to have several alternative approaches. For example, heterologous over-expression of eukaryotic proteins with chaperones and trigger factor has been shown to improve the solubility of the overexpressed proteins in *E.coli* (Voulgaridou et al. 2013). Multiple lines of evidence argue for an increase in protein solubility by switching to an alternative tag (Hammarstrom et al. 2002). However, in selecting which tag to use, one is confronted with a daunting number of choices. To compound this problem, in addition to the commonly used fusion partners, new solubility enhancing tags are constantly being developed (Costa et al. 2014). Another alternative is to refold the protein *in vitro*. However, even the most robust protocols only refold a small fraction of the input protein, and it is hard to purify the refolded fraction (Vincentelli et al. 2004). Finally, if the bacterial expression is unsuccessful to this point, using other expression hosts could be a viable alternative. Conventional eukaryotic expression systems include baculovirus system in insect cells, the yeasts *Pichia pastoris* and *Saccharomyces cerevisiae*, human cells, or cell-free

systems using prokaryotic or eukaryotic extracts (Graslund et al. 2008). Cell-free systems, which have been used extensively to generate thousands of purified proteins for structural studies, can be utilized to produce proteins that are toxic to *E. coli.* Although the list is certainly not exhaustive, for future work, one or more of these approaches could provide a breakthrough in solving the START domain structure.

The fluorescence-based thermal shift assay (DSF) is tractable for use with PDF2

Chapter 2 shows that the START domains are ligand-binding motifs that can modulate transcription factor activity in both plants as well as mammals while genetic and mutant studies described in Chapter 3 further underscore the importance of ligand binding for HD-Zip TF activity. Together, these results set the stage for the identification of potential small-molecule ligands for the START domain. Although START domains from mammals are reported to bind lipids (Schrick et al. 2014), an unbiased approach could be the best strategy to detect novel START domain ligands. To this end, the availability of soluble and correctly folded His-tagged PDF2 and PDF2-START+SAD proteins enabled a DSF high throughput screen (HTS) for potential START domain ligands (Figure 4.5 C, D).

DSF is a label-free high throughput screening methodology that foregoes the need to develop assay materials or reporters specific for individual protein targets. It is a rapid and cost-effective approach that requires relatively small amounts of partially purified protein (Vedadi et al. 2006). Another advantage of DSF is that it is applicable to different stages of the screening process, from primary screening to hit profiling. It can be a useful selection tool of protein constructs for use in screening and X-ray crystallography (Cummings et al. 2006b). In other cases, DSF aids in the expression and purification of proteins since it allows the detection of a specific ligand that improves the protein stability and hence decreases its potency to unfold and reduces the chances of its aggregation (Vedadi et al. 2006). Using naringin during protein purification resulted in much higher levels of soluble PDF2-START expression, presumably by stabilizing the folded protein (Figure 4.20)

It should be noted that protein unfolding can also be monitored using fairly simple intrinsic fluorescence such as tryptophan or tyrosine residues, albeit in this study it was felt to be advantageous to utilize an external probe. The 230 amino acid PDF2-START domain is intrinsically fluorescent by virtue of 5 Trp residues. However, due to the presence of two many

Trp residues, the small changes of one residue might get buried under the strong and broad signal from the whole protein.

Not all proteins are amenable to DSF measurements. In particular, non-globular and large multi-domain proteins tend not to exhibit distinct phase transitions during thermal denaturation.

As seen in Figure 4.8, melting curve of PDF2 shows good thermal stability, but with a weak denaturation transition (41.7°C). Thermal shift assays, unfortunately, do not always generate single sigmoidal curves, and it is not unusual to observe a complex multiphasic curve in the presence of a multi-domain protein. Occurrence of a single sigmoidal curve for the multidomain PDF2 protein can be explained by the fact that many multidomain proteins have domains that are sufficiently coupled energetically that they melt in a cooperative manner resulting in a single melting transition state (Pantoliano et al. 2001). Further optimization experiments indicated that 10x SYPRO Orange dye concentration and 5 μM protein were suitable conditions to perform high throughput screening (Figure 4.7). Together, the experiment established the assay's tractability, as well as informed the conditions to perform DSF experiments using potential small molecule ligands as discussed below.

PDF2 binds flavonoids

The AnalytiCon MEGx library consists of 1000 pure natural compounds from plants and microbes (plants- 792, microbes- 208). Out of 1000 compounds screened at a concentration of 300 μM, only 25 stabilized while 88 induced a negative shift in T_m of the protein. Whereas 868 had no effect, 19 did not show any discernible peak to allow T_m calculation and were categorized as denaturants. It should be noted that compounds that elicited ΔT_m between 1-2˚C shifts were disregarded. The majority of the hits were flavonoids, a broad class of plant secondary metabolites. As illustrated in Figure 4.13, benzylated chalcone and pinoquercetin displayed the highest thermal shift, with ΔT_m values of 10.8˚C and 13.2˚C, respectively. In general, for 1:1 protein–ligand binding and a simple two-state approximation of the unfolding transition, ΔT_m correlates reasonably with both affinity and ligand concentration (Brandts and Lin 1990). Accordingly, benzylated chalcone and pinoquercetin most likely represent high-affinity protein binding ligands for PDF2.

Flavonoids are widely distributed throughout the plant kingdom and are abundant in many flowers, fruits and leaves. They are characterized by the presence of two benzene rings (rings A and B) that are linked by a 3-carbon bridge to form chalcones or by a pyrane or pyrone ring (ring C), Figure 4.17. On the basis of the position of and the modifications to the A, B and C rings, the >4000 flavonoids discovered to date can be classified into several classes, including flavonols, flavones, isoflavones and anthocyanin pigments (Figure 4.17). While benzylated chalcone is a member of a class of chalcones that exhibits antibacterial, antimalarial and antineoplastic activity, pinoquercetin is a C-methylated flavonol (Figure 4.15). Flavonols have been the subject of intense research interest because of their diverse roles in protection from oxidative damage and UV radiation, inhibition of auxin transport, control of pollen function, and signaling to symbiotic organisms (Wasson et al. 2006; Buer et al. 2010). In addition, dietary flavonols have also been implicated as being protective, regulatory, or cytotoxic molecules in mammals (Williamson and Actis-Goretta 2010).

To further map the binding domain, all 25 compounds were retested using the PDF2-START+SAD protein. Among these, only eight were found to induce a positive shift in the T_m of both PDF2 and PDF2-START+SAD proteins. It is not immediately clear why fewer compounds showed temperature shifts for PDF2-START+SAD as compared to PDF2. Perhaps it is harder to identify stabilizing interactions in proteins with intrinsically higher stability. In line with this idea, under similar purification conditions, more soluble expression was observed for PDF2-START+SAD in comparison with PDF2.

Using a DSF assay, interactions between PDF2-START+SAD protein and various commercially available natural flavonoids such as naringin. naringenin, quercetin, and quercetin 3-β-D-glucoside were also observed. This is an important finding as it may facilitate identification of high affinity ligands that can be obtained at a significantly lower cost and may further enable secondary ligand validation assays requiring large quantities of compounds.

Dose response curves

In ligand titration experiments, lower ΔT_m values were observed for all compounds tested (Figure 4.19). This inconsistency might be explained by batch-to-batch variation in purified protein. Also, both licochalcone B and naringenin elicited $\Delta Tm \leq 1^{\circ}C$, indicating a failure to reproduce the previous results. While the change on T_m may increase sharply as the protein is

saturated by the ligand (particularly for tight binding) the increase of protein T_m continues after the protein is full saturated (Matulis et al. 2005; Cimmperman et al. 2008). The nonlinear T_m dependence on ligand concentrations observed for some strong carbonic acid inhibitors, such as EZA and TFMSA supports this notion (Cimmperman et al. 2008). It may also explain the linear variation exhibited by benzylated chalcone. However, the reason for expected linearity is not fully understood. In fact, the saturation effect seen in some cases (e.g., pinoquercetin) is probably due to ligand solubility issues, where the concentration of the ligand in solution becomes a limiting effect, and in other cases due to ligand binding to the unfolded state of the protein (Kervinen et al. 2006).

Paradoxically, the danger of ranking compounds by relying on T_m alone is highlighted in the literature (Matulis et al. 2005; Niesen et al. 2007; Cimmperman et al. 2008) and determination of binding constants using DSF data is problematic. The binding constant cannot be directly determined from a thermal shift assay since these affinities depend both on enthalpy and entropy, which is not accurately assessed with DSF (Matulis et al. 2005). Further, Cimmperman et al. (2008) postulated that if the ligand binds significantly to the unfolded state of the protein as well as to the folded state, the binding constant may be underestimated, since stabilizing the unfolded state shifts T_m towards a lower temperature. Nevertheless, several studies have successfully used DSF to calculate ligand association constants that are in good agreement with isothermal titration calorimetry (ITC) measurements (Lo et al. 2004) as well as radioactive competition and fluorescence polarization assays (Pantoliano et al. 2001). Vedadi and coworkers (Vedadi et al. 2006) correlated inhibition data for a protein kinase and showed that T_m shifts larger than 4 °C translate into values for $IC_{50} < 1$ μM. For several compounds, this correlation suggests that DSF assays adequately describe binding kinetics despite the irreversible nature of the thermal protein melting (Brandts and Lin 1990; Schwarz et al. 1991).

In conclusion, all eight compounds, except lichochalcone and naringenin represent promising candidates for further validation. It cannot be emphasized enough that applying a combination of orthogonal methods provides the most robust and efficient way to identify a *bona fide* ligand. For instance, while DSF offers a simple platform for rapid identification of interacting ligands, the assay is mainly qualitative and not the best to assess binding specificity. On the other hand, ITC is a gold standard for analyzing weak protein-ligand interacts. However, the technique requires protein and ligand at very high concentrations, which often

puts practical limits on the solubility of protein and ligand at very high K_D. An integrated screening cascade starting from HTS, e.g., DSF or Fluorescence Polarization (FP) followed by validation using NMR and/or SPR secondary screens, and ultimately using more material intensive and information-rich techniques, e.g., ITC to characterize interactions is sensible strategy for ligand screening. In the present study, a HTS assay provided a list of lead compounds to follow up on. ITC experiments are currently underway to verify ligand hits and more accurately estimate binding constants. It is hoped that compounds validated by both assays will become sustainable leads that can be developed further. Finally, co-crystallization experiments will be necessary to determine the binding position of each of the compounds to the protein.

Flavonoids are synthesized via the phenylpropanoid and flavonoid pathways (Figure 4.17). Thus far, only two flavonoid-binding proteins have been identified in plants. These include *Arabidopsis* glutathione S-transferase (GST) (Mueller et al. 2000) and Fra proteins from strawberry (Casanal et al. 2013). Interestingly, Fra proteins belong to START superfamily and similar to mammalian START domains these adopt a helix-grip-fold with an internal cavity capable of binding hydrophobic ligands. Structural analysis of Fra bound to catechin demonstrates conformational flexibility in loop regions surrounding the ligand-binding cavity, suggesting that ligand-induced conformational changes are a conserved feature in the START protein superfamily. Thus, it is conceivable that HD-Zip derived START domains may undergo similar conformational changes upon flavonoid binding. What could be the functional relevance of flavonoids as putative ligands for PDF2? Although the binding still needs to validated, HD-Zip TFs may act as transporters binding to the flavonoid intermediates and making them available to processing enzymes. The evidence has now been provided that at least two of the enzymes of flavonoid metabolism (CHALCONE SYNTHASE [CHS] and CHALCONE ISOMERASE [CHI]) are located not only in the cytoplasm but also in the nuclei (Saslowsky et al. 2005). Another intriguing possibility is that they could play a role as signaling components, monitoring the metabolic flux into the flavonol branch and influencing transcriptional activity of regulatory genes coding for flavonoid enzymes. This hypothesis draws on the flavonol dependent feedback repression of PAL, the gateway enzyme for the synthesis of phenolic and flavonoid compounds. Notably, this inhibition is dependent on the accumulation of aglycosylated flavonoids and can be restored by exogenous application of

naringenin (Yin et al. 2012). The latter supposition is further supported by the ChIP-seq and qRT-PCR data that suggests negative regulation of *PHENYLALANINE AMMONIA-LYASE* (*PAL*) and *CHALCONE ISOMERASE-LIKE* (*CHIL*) gene by GL2 (Figure 3.15). The role of CHALCONE ISOMERASE-LIKE (CHIL) remained elusive until recently. Genetic and metabolic studies have now provided evidence for its role in flavonoid biosynthesis downstream of CHS (Jiang et al. 2015). In summary, it is proposed that the flavonol metabolome could be directly influenced through binding of flavonoids to HD-Zip IV TFs. Whether GL2 also binds to flavonoids still needs to be demonstrated. The protein-metabolite immunoisolation experiments performed in yeast indicate an overlap in binding partners between GL2 and PDF2.

Ligand supplementation

Multiple lines of evidence argue that using ligands during protein purification can lead to improvements in protein stability and solubility. For instance, co-expression with a high-affinity ligand was a key to obtaining purified steroid nuclear receptor ligand-binding domain (LBD) for structural studies (Sack et al. 2001). Likewise, addition of ligand during the cell lysis step and throughout the entire purification process was the key to obtaining pure monomeric kinase 1 (Hassell et al. 2007), while purification without a ligand resulted in a mixture of proteins, lipids and DNA. This raises the question as to what causes increased solubility upon ligand supplementation. A simple explanation is that inclusion of ligand promotes and stabilizes the folded protein, leading to a stable and soluble protein-ligand complex. Wu et al. (2007) found that upon ligand (Carbenoxolone) supplementation and concomitant protein purification, the glucocorticoid-activating enzyme type 1 11β-hydroxysteroid dehydrogenase (11β-HSD) had a ligand bound to its active site. Although less likely, however, another possibility is an increase in RNA stability or reduction of toxic effects on the expression of the host.

As noted previously, His_6-tagged PDF2-START displays extremely poor solubility. One hypothesis is that using ligands as culture additives might improve stability and subsequently increase the solubility of PDF2-START. Indeed, increase in levels of soluble protein was observed in the presence of 100 μM naringin (Figure 4.20). Similarly, a higher level of soluble protein expression was observed by including isoliquiritigenin as a culture

additive (data not shown). These results provide further evidence for the postulate that one can use ligands to stabilize nascently synthesized proteins to promote solubility and prevent aggregation. By contrast, co-expression with naringin dihydrochalcone was not sufficient to provide soluble protein. Structural studies with PDF2-START and naringin will be important in defining the ligand-binding site, which may differ from mammalian START domain binding site.

It should be mentioned that ligand supplementation experiments are very tricky and the design and outcome of this approach is contingent upon several factors. These include the type and potency of the ligand, toxicity in *E.coli*, cellular uptake, the solubility of the ligand, DMSO concentration, interference with the host metabolism, and conversion of compound itself by host. For example, expression studies with GR LBD demonstrated that combination of ligand supplementation and F602S mutation had a dramatic effect on expression levels and was a key to obtaining GR crystals and subsequent structure determination (Bledsoe et al. 2002). Relevant to this, the addition of quercetin induced toxicity in *E.coli* (data not shown). While only narignin, isoliquiritigenin, quercetin, and naringin dihydrochalcone are tested as culture supplements in this study, the effect of other compounds – identified in the DSF screen - remains to be addressed.

Fatty acids represent *in vivo* binding partners for HD-Zip TFs

To achieve full understanding of its function as a putative ligand-binding domain in plants, natural ligands for the START domain need to be identified. An *in vivo* assay for systematic identification of protein- metabolite interactions was developed for yeast proteins (Li et al. 2010), and using this method, candidate ligands for START domains were identified using GSV synthetic transcription factors expressed in yeast (Schrick et al. 2014). This method was modified for immunoisolation of protein- metabolite complexes from plant tissues. The identification of a natural ligand for GL2 helps characterize the START domain, identify a new role for the ligand, and provide insights into physiological regulation of the START. As observed in Figure 4.2`, pull down samples were analyzed by Q-TOF using an untargeted metabolomics approach. In contrast to a targeted approach where only known metabolites of interest are selected for monitoring and quantitation, an untargeted platform allows quantitation of all ionizing metabolites simultaneously, measuring both known and structurally novel

metabolites based on their *m/z* values. Moreover, TOF analyzer has high mass accuracy, which allows molecular formulas to be determined for small molecules.

At least two pieces of evidence suggest that START domain can bind lipid ligands in plants. First, Gou et al. (2015) reported that START domain from WKS1, wheat START domain protein, has binding affinity for phosphatidic acid (PA) and phosphatidylinositol phosphates (PIPs). A recent study identified PA and several phospholipids as potential ligand binding partners for GL2 and PDF2 (Schrick et al. 2014).

Using gl2ΔZLZ in the metabolomics-based ligand-enrichment experiment, elevated levels of four fatty acids corresponding to m/z = 291.2, 293.2, 277.2, and 279.2 were found in comparison to the control. It was suggested that these peaks represent oxylipins and unsaturated fatty acids including linoleic acid (C18:2) and linolenic acid (C18:3). Further detailed analysis comparing the fragmentation data with the MS/MS product ion scans of various standards, revealed colnelenic acid as a potential ligand for GL2-START domain. Although the current data are suggestive, they are insufficient to conclude whether colnelenic acid represents a true ligand. More detailed biochemical and biophysical studies (for example ITC) need to be conducted in to determine binding stoichiometry of the GL2-colnelenic acid interaction. Additionally, structural studies using NMR or X-ray diffraction will be pivotal in defining the ligand-binding site.

One criticism of the assay is the use of transcription factor in which Zip dimerization domain has been deleted. Zip domain deleted version of GL2 was chosen to enhance the throughput. Although protein is defective in DNA binding (refer Chapter 3), it encodes a wild type START domain and is expected to bind similar ligands as the full-length GL2. Also, there is not data in which binding is mapped to the START domain. Further experiments using GL2-START protein will be necessary to elucidate ligand binding to the START domain.

In Hsu and Turk (1999), the authors described distinct fragmentation spectra of isomeric unsaturated fatty acids. Using low energy collisionally activated dissociation (CAD) on a triple stage quadrupole instrument, they found that isomeric polyunsaturated fatty acids: linolenic acid (9,12,15)-18:3 and gamma linolenic acid (6,9,12)-18:3 yield distinguishable tandem spectra. While m/z 99, m/z 195 and m/z 235 reflect the most abundant ions in tandem spectra of (9,12,15)-18:3, the ion spectra of (6,9,12)-18:3 includes members at m/z 193 and m/z 153. These findings suggest that in addition to colnelenic acid, GL2 may bind to a yet

unknown isomer of linoleic or linolenic acid. Since the fragmentation pattern of experimental m/z peaks corresponding to 277.2 and 279.2 does not match that of pure linoleic acid standard (m/z 277.2) used in this study, it is likely that they represent isomeric forms having dissimilar product ion spectra. Thus, whether linoleic and linolenic acid represents a true binding partner for GL2, possibly via its START domain, is still an open question that awaits further investigation.

Taken together, the results provide evidence for the GL2 transcription factor binding to several fatty acids including linoleic, linolenic, and the colnelenic acid, which is an oxylipin that functions in cell wall modifications required for lateral root development and for defense against pathogens (Vellosillo et al. 2007). Results from the homology modeling of *Arabidopsis* START domain ligand binding cavities using mouse STARD4 domain as template further corroborate these findings (Schrick et al. 2014). The modeling studies indicated that the hydrophobic tunnel of START domain from GL2 has a small cavity volume of 270 $Å^3$ (Yadav, unpublished), which may indicate binding to fatty acids.

MATERIALS AND METHODS

Nicotiana benthamiana Growth Conditions

Seeds of *N.benthamiana* were germinated on wet filter paper for 7d, and then the seedlings were planted on soil mix containing plug and germination growing medium (PRO-MIX® PGX), vermiculite and perlite (Therm-o-Rock West, Inc.) in 4:1:1 ratio, and 5-6 pellets of fertilizer osmocote classic 14-14-14 (Everris US). Plants were grown in a growth chamber at 22°C under 14-h-light/10-h-dark cycles. For each set of experiments, used 3-4 week-old plants that were typically at the five-leaf stage were used.

Constructs and cloning

N-terminal GST fusion to START domain sequences from PROTODERMAL FACTOR 2/PDF2$^{E245\text{-}L482}$, HOMEODOMAIN GLABROUS 11/HDG11$^{D228\text{-}F458}$, HOMEODOMAIN GLABROUS12/HDG12$^{E207\text{-}F438}$, ANTHOCYANINLESS 2/ANL2$^{D317\text{-}L544}$, PHABULOSA/PHB$^{N166\text{-}H379}$, GLABRA2/GL2$^{E253\text{-}L487}$, and MERISTEM LAYER 1 /ATML1$^{E254\text{-}L482}$ were generated by cloning PCR products into acceptor pFN2A (GST) flexi vector (Promega, #C8461). RIKEN *Arabidopsis* full-length (RAFL) cDNA clones (Seki et al, 2002): HDG11 (At1g73360), HDG12 (At1g17920), ANL2 (At4g00730), PHB (At2g34710), ATML1 (At4g21750), PDF2 (At4g04890), and GL2 (At1g79840) were used as template to append SgfI and PmeI sites onto the START domain using expand high fidelity PCR system (Roche). Following purification, PCR products were digested with restriction enzymes SgfI and PmeI (NEB). The digested PCR products were subsequently ligated onto an acceptor pFN2A (GST) flexi vector that was digested with SgfI and PmeI. The plasmids were introduced into *E. coli* strain DH5α, the positive clones were selected on solid medium containing 100 mg/ml ampicillin and the constructs were verified by sequencing.

Full length PDF2 (amino acids 1-743), PDF2 (amino acids 232-743) encoding the START+SAD domains, PDF2 (amino acids 245-478) encoding the START domain, full length GL2 (amino acids 1-747), and GL2 (amino acids 234-747) encoding the START+SAD domains were subcloned into expression vector pNIC28-Bsa4 by ligation-independent cloning (Dortay, et al, 2011). The resulting expression constructs contained a hexahistidine tag and a TEV-protease cleavage site (MHHHHHHSSGVDLGTENLYFQS) at the N-terminus.

Linearization of LIC expression vectors for LIC cloning

LIC expression vectors (5 µg) were digested with 25U BsaI in a 50 µl reaction volume and purified from contaminating stuffer fragment and undigested vector by gel-extraction using the NucleoSpin Extract II kit (Macherey & Nagel, Du¨ren, Germany). To generate 5' LIC overhangs (15 and 16 nt, respectively) at both ends the purified vector backbone was treated for 30 min (22˚C) with T4 DNA polymerase in the presence of dCTP, using the following reaction setup: 200 ng of purified vector backbone, 2 µl 10x buffer 2 (NEB), 2 µl dCTP (25 mM), 1 mM dithiothreitol (DTT, 100 mM), 2 µl 10x (10 mg/mL) bovine serum albumin (BSA; NEB), 10U T4 DNA polymerase (NEB) in a volume of 20 µl (filled up with ddH_2O). The reaction mix was heat inactivated for 20 min at 75˚C.

Primer design and preparation of PCR products for LIC cloning

For directional and in-frame cloning of PCR-amplified open reading frames into pNIC-Bsa4 vector, sense primers starting with the following sequence were used: 5'**TACTTCCAATCC** (ATG)-3' (the ATG nucleotides in brackets indicate the gene's start codon). Antisense primers for cloning must start with the sequence 5'-**TATCCACCTTTACTGTTA**XXX-3'. Complementary LIC overhangs within the sense and antisense primers are underlined (for full primer sequences see Table 2). The cDNAs encoding for the *Arabidopsis thaliana* protein PDF2 and GL2 were amplified by high-fidelity polymerase (Phusion DNA Polymerase; Thermo Scientific Inc., Lafayette, CO) using cDNAs of PDF2 (At4g04890) and GL2 (At1g79840) as template. PCR products were treated at 22˚C for 30 min with T4 DNA polymerase in the presence of dCTP, using the following reaction setup: 0.2 pmol purified PCR product, 2 µl 10x buffer 2 (NEB), 2 µl dCTP (25 mM), 1 mM DTT (100 mM), 2ul 10x BSA (10 mg/ml; NEB), 1U T4 DNA polymerase (NEB) in a volume of 20 µl (filled up with ddH_2O). The reaction mix was heat inactivated for 20 min at 75˚C.

LIC cloning of target genes

0.02 pmol (1 µl) and 0.04 pmol (2 µl) of pre-treated LIC vectors and PCR products (see above) were mixed and incubated for 5 min at 22˚C. The reaction mix was supplemented with

1 ml EDTA (25 mM), followed by incubation for 1 hr at 22˚C and transformation of the whole reaction mix into *E. coli* strain DH5α for plasmid amplification. Plasmid constructs were verified by sequencing. Using a PCR-based site-directed mutagenesis method (Scott et al., 2002) and 6xHis-PDF2 as template, 6xHis-pdf2^{L467P} START domain mutant protein was constructed and expressed in *E. coli* following the protocol as described below.

The *GL2* cDNA was sub-cloned into pBluescript via *Kpn*I and *Sal*I restriction sites to serve as a template for the gl2ΔZip deletion construct. Phusion Site-Directed Mutagenesis Kit (Thermo Scientific) was used to generate the Zip domain deletion, following the manufacturer's protocol. For TOPO cloning, *gl2*Δ*Zip* cDNA was recombined into entry vector pENTR/D-TOPO vector (Invitrogen, Carlsbad, CA). Entry clones containing *gl2*Δ*Zip* sequence was transferred from the entry vector to the destination clone vector pEarleyGate 201 with HA tag on N-terminal using the LR reaction (Earley et al, 2006). Oligonucleotides used in this study are listed in Table 4.5.

Test expression in *E.coli*

Each N-terminus GST tagged expression construct was transformed into *E. coli* strain BL21(DE3)pRARE (gift from Philip Gao, COBRE Protein Production Core Laboratory, KU Lawrence). A single colony was used to inoculate 1 ml LB containing both kanamycin (50 mg/ml) and chloramphenicol (34 mg/ml). The culture was incubated at 37˚C with shaking overnight. For test expression, 3 ml of LB media containing kanamycin (50 mg/ml) was inoculated with overnight culture. All cell cultures were grown at 37˚C to an optical density at 600 nm (OD_{600}) of 0.6. Cultures were then induced with 0.1 mM IPTG and incubated at 18˚C. After incubation for 20 h, cells were harvested by centrifugation (Sorvall RC-5B refrigerated superspeed cientrifuge, DuPont Instruments), 6000 rpm, 10 min at 4˚C). Pelleted cells were resuspended in Tris lysis buffer (50 mM Tris-HCl pH 7.5; 500 mM NaCl, 20% (w/v) glycerol, 2 mM DTT, 5mM EDTA) at a ratio of 10 ml per gram of wet pellet. To this mixture, 1X EDTA-free halt protease inhibitor cocktail (Thermo Scientific), 2mM PMSF (Sigma) and 1 mg/ml hen egg-white lysozyme (Sigma) was added, and the cells then lysed by sonication (10 Hz pulses for 15 sec at 30% amplification on an Ultrasonic homogenizer (BioLogics Inc. Australia). Cell debris was removed by centrifugation (Eppendorff 5810R), 12,000rpm, 1 h,

4˚C). The total (crude lysate), soluble (supernatant), and insoluble (pellet) fractions were analyzed by SDS-PAGE.

Large scale expression and purification of GST:PDF2-START

For large-scale protein production, 3 L cultures were used. GST:PDF2-START expression construct was grown in culture as described above using 0.1 mM IPTG and moving cultures from a 37˚C incubator to a 18˚C incubator after OD_{600} 0.5–0.6 was reached. The soluble fraction was incubated with 2.0 ml of Glutathione sepharose 4B (GE Healthcare Life Sciences) pre- equilibrated in Tris wash buffer (50 mM Tris-HCl pH 7.5) o/n at 4˚C to bind the GST-tagged PDF2-START. After overnight incubation, the beads were washed sequentially with 10 bed volumes of Tris wash buffer (50 mM Tris- HCl pH 7.5; 500 mM NaCl, 20% (w/v) glycerol, 1mM EDTA) containing 1%, 0.1%, and 0.01% Triton X-100 and the protein was eluted using Tris elution buffer (50 mM Tris- HCl pH 8.0; 150 mM NaCl, 20% (w/v) glycerol, 20mM Glutathione). The protein was liberated from the hexahistidine tag by incubation with His_6-tagged TEV protease (gift from Karl Nichols, Center for Eukaryotic Structural Genomics) over night at 4˚C and subsequently incubated with glutathione resin to remove uncleaved PDF2-GST fusion protein and GST only. After removal of TEV protease on TALON metal affinity resin (Clontech), the protein was loaded onto 10/300 Superdex-75 column (GE Healthcare). Gel filtration was performed in equilibration buffer (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% glycerol). Fractions corresponding to a major and minor peak were pooled based on gel filtration profiles and purity determined by SDS-PAGE and Coomassie staining. FPLC gel filtration was performed in Jeroen Roelofs lab.

GST:PDF2-START binding assays with 8-anilino-1-naphthalenesulfonic acid (ANS)

The assay was performed as previously described (Downes et al., 2003) with modifications. The GST tagged PDF2-START protein and ANS were prepared or diluted in protein buffer (25 mM Tris, 500 mM NaCl, 10% glycerol, pH 8.0) to desired concentrations, before they were mixed together in a 96-well fluorescence plate to the final volume of 150 µl in each well. The final concentration of protein was held constant in each well at 68 µM (or no protein for control), whereas that of ANS was varied across eight wells at 0, 25, 50, 100, 200,

400, 600, 800 μM. The protein-ANS mixture was allowed to incubate at room temperature for 1 h. After that, the binding was monitored using a Victor 2V microplate reader (Perkin Elmer) at excitation and emission wavelengths of 400 nm and 500 nm, respectively. The experiment was performed in triplicate at each concentration of ANS.

In vitro ligand binding assay by mass spectrometry

An *in vitro* method described in Li et al (2010) was used for the binding assay using equimolar (100 μM) amounts of protein and metabolite. After co-incubation of protein and candidate ligand for 30 min at room temperature, size exclusion chromatography (Zeba spin desalting column, Pierce) was used to remove unbound metabolite and other impurities from the mixture. To extract the protein-bound metabolites, 100 μl of pure methanol was added to the samples and incubated at RT for 15 min. The methanol extract was then immediately transferred to a glass vial and analyzed by mass spectrometry in collaboration with Ruth Welti (KSU). The dissolved metabolites were fragmented and detected by direct infusion ESI-MS/MS (API 4000; Applied Biosystems) using a targeted approach, and quantified as previously described (Devaiah et al., 2007; Lee et al., 2011).

Protein overexpression and purification

Each expression construct was transformed into *E. coli* strain BL21(DE3)pRARE. Cells were grown in Terrific Broth supplemented with 8 g/l of glycerol at 37°C. At an OD_{600nm} of between 1 and 2 the temperature was lowered to 18°C, recombinant protein production was induced by addition of 0.1 mM isopropyl-β-d-thiogalactopyranoside (IPTG), and cell growth was continued for 20 h. Cells were harvested by centrifugation and resuspended in 1.5 ml of buffer 1 per gram of wet cells (50 mM Sodium phosphate pH 7.0, 500 mM NaCl, 10% glycerol, 20 mM imidazole, 0.5 mM TCEP). Before lysis, 4 μl (25 kU) of Pierce Universal Nuclease for Cell Lysis (Thermo Scientific), 1X EDTA-free halt protease inhibitor cocktail (Thermo Scientific), and 1mg/ml hen egg-white lysozyme (Sigma) were added per 50 ml cell suspension, and cells were lysed by sonication (10 Hz pulses for 15 sec at 30% amplification on an Ultrasonic homogenizer (BioLogics Inc. Australia). Cell debris was removed by centrifugation and the soluble fractions were filtered through a syringe filter (0.45 μm pore

size). Cleared cell lysates were added to TALON Metal Affinity Resin (Clontech) pre-equilibrated with buffer 1 containing 5mM imidazole and gently agitated at 4˚C ice for 2 hr on a platform shaker to allow the His_6-tagged protein to bind the resin. Resin was then washed sequentially with 20 bed volumes of buffer 1 and buffer 1 containing 5mM. Bound protein was eluted twice with 10 bed volumes of buffer 1 containing 50 mM imidazole (for PDF2, PDF-START+SAD, GL2, GL2-START+SAD) or 500 mM imidazole for PDF2-START protein. The Bradford protein assay was used to determine protein concentration.

Identification of the protein by Mass spectrometry

The identity of the protein was confirmed by mass spectrometry following the in-gel trypsin digestion protocol. Each band was excised from the SDS-PAGE gel stained with Coomassie blue. Peptides were extracted and dried down for LC-tandem MS. Each data file was searched against *E. coli* protein database and a custom database with only the PDF2 coding sequence. The samples were processed at UMKC Proteomics and Mass Spectrometry facility.

DSF optimization

DSF was carried out using a real-time PCR setup (CFX384 - Biorad Laboratories) fitted with filter sets for measurement of FRET signal. Initially, the general protocol outlined by Niesen et al, 2007 was followed. Conditions of the assay were then optimized for use with PDF2 before high throughput screening (HTS) was carried out to identify ligands from MEGx library (AnlytiCon). Parameters of the assay that were optimized included SYPRO orange dye (Invitrogen) and protein concentration. Each parameter was tested in triplicate on a white, non-skirted 96-well PCR plate (Thermo Scientific), sealed with transparent foil (Bio-rad). Three different protein concentrations (1,5,10 μM) and four different dye concentrations (2x, 5x, 10x, 20x) were tested using the protocol described below. The controls are (1) Two negative controls (protein only and dye only), and (2) a positive control (0.01 mg/ml lysozyme dissolved in 10 mM Tris chloride pH 8, 1 mM EDTA).

AnalytiCon MEGx library

A 1000 compound library at 0.2 µmole each, in 96-well plates was purchased from AnalytiCon Discovery GmbH (Germany). MEGx is a library of pure natural product screening compounds isolated from plants and microorganisms (plants- 792, microbes- 208). The powder compounds were dissolved in 10 µl of DMSO to give stock solution concentrations of 20 mM.

DSF experimental setup for high throughput screening

Purified PDF2 was appropriately diluted in analysis buffer containing 50 mM Tris-Cl, pH 8.0, 150 mM NaCl, 10% Glycerol. Protein-ligand solutions were dispensed into the wells of white, unskirted 96-well PCR plates (Thermo Scientific, AB-0700/W). Each well contained 1 µM protein, 10x SYPRO orange (Invitrogen), and 300 µM library compound (ligand) in a final volume of 20 µl. DMSO concentration was maintained at 1.5% for all assays. In each plate there were 4 reference wells that did not contain any ligand. Since the compounds were in 100% DMSO, the corresponding volume of DMSO (3ul of 10% DMSO stock) was added to the reference wells for accurate comparison (vehicle control). The PCR plates were sealed with optical seal (Thermo Scientific, AB-0558), centrifuged, and incubated at 4˚C for 30 min prior to analysis. Thermal scanning (25 to 95˚C at 0.5˚C/ 15 sec) was performed using a real-time PCR setup (CFX384 - Biorad Laboratories) and fluorescence intensity was measured after every 10 seconds. The excitation/emission filter settings were according to the "FRET" channel that is compatible to the SYPRO Orange fluorescence signal. The positive hits (ΔT_m shift >2˚C) were retested in triplicates using both PDF2 and PDF2-START+SAD proteins using the protocol described above. Each compound was also tested for background fluorescence.

Analysis of DSF Data

Data analysis was performed according to the protocol described by Niesen et al, 2007. For each well solution, the temperature midpoint for the protein unfolding transition, T_m, was determined. The fluorescence intensity is plotted as a function of temperature. This generates a sigmoidal curve that can be described by a two-state transition. The inflection point of the transition curve (T_m) is calculated using the Boltzmann sigmoidal non-linear regression

function in statistical software Graphpad Prism (GraphPad Software, San Diego, CA, USA). Compounds, which do not have a measurable peak, are excluded. For compounds tested in triplicates, relative fluorescence units were averaged and background fluorescence was subtracted. For ease of comparison of thermal shift between wells, graphs are normalized whereby the scale of the amplitude of each curve is between 0 and 1.

DSF titration

Using either 2 mM or 50 mM stock of compound in DMSO, dilutions of each hit compound were made ranging from 1 to 300 μM in analysis buffer (1, 5, 10, 20, 50, 100, 200, 300). Samples (20 μl) were arrayed in white, unskirted 96-well PCR plates (Thermo Scientific) containing 15 μl of 1 μM PDF2-START+SAD in analysis buffer supplemented with 10x SYPRO Orange dye (Invitrogen) and 5 μl of each compound concentration. DSF was carried out via the optimized procedure. ΔT_m was plotted against the concentration of the ligand using GraphPad, and a sigmoidal dose–response curve (variable slope) was fitted to the data to calculate dissociation constant (K_D). Each compound concentration was repeated five times to obtain the standard deviation.

Ligand supplementation to enhance soluble protein production

His_6-tagged PDF2-START proteins were expressed and purified as described above in the presence or absence of the putative ligand, as culture supplement (50 ml culture volume). 100 μM naringin (Sigma, #N1376) and naringin dihydrochalcone (Sigma) were included during protein expression, cell lysis, and downstream purification steps.

Agrobacterium- mediated transient expression in *N. benthamiana* leaves

Agrobacterium tumefaciens GV3101 containing the respective constructs and the p19 silencing suppressor were grown overnight at 30˚C in Luria-Bertani medium (LB) supplemented with the appropriate antibiotics. Cultures were pelleted and suspended in 10 mM MES pH 5.5 (Sigma), 10 mM $MgCl_2$ (Fisher) and 150 μM acetosyringone (ACROS Organics) at OD_{600} = 0.2. The constructs and p19 cultures were mixed in a 1:1 ratio and used for

infiltration of 4-5 leaf *N. benthamiana* plants. Samples were harvested 2 to 3 days after inoculation.

Protein extraction and analysis

For protein extraction, all steps were carried out on ice or at 4˚C. 10 g of leaf tissue were ground in liquid N_2 and resuspended in 3 ml of ice-cold extraction buffer (10% glycerol, 50 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl, 2% w/v PVPP, 10 mM DTT, 1X protease inhibitor cocktail (Sigma), 0.5% NP40) per gram of tissue weight. Debris was removed from the lysate by centrifugation at 10,000 xg, 25 min. The centrifuged supernatant was filtered through four layers of Miracloth (Calbiochem, San Diego, CA, USA) and desalted using zeba spin desalting columns (Thermo Scientific, Rockford, IL). The centrifuged supernatant was incubated with 10 µg of rat monoclonal anti-HA high affinity antibody (Clone 3F10, Roche) coupled to Dynabeads G beads (Invitrogen). The mixture was incubated end-over-end in batch format o/n at 4˚C. After binding, the beads in batch volume were washed three times with IP buffer (10% glycerol, 50 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl) supplemented with 0.5% NP-40 and three times with basal IP buffer (each wash volume = 10x volume of beads). Small metabolites bound to the protein were eluted with 100 µl of chloroform/methanol/water (300:665:17.5) and transferred to a 2ml polypropylene insert vial. The experiment was performed two times. The pEarleyGate201 empty vector served as a negative control.

Western blotting

For analysis of fusion protein expression, total proteins were extracted in hot SDS-sample buffer (100 mM Tris-HCl, pH 6.8, 2% SDS, 20% glycerol, 100mM DTT, and 0.004% bromo- phenol blue, 0.48g Urea per ml) and analyzed by immunoblotting using a standard protocol with mouse anti-hemagglutinin (HA) antibody (Thermo Scientific) and the ECL Western Blotting system (GE, Healthcare), according to the manufacturers instructions.

Q-TOF analysis

Spectra were acquired on an MDS SCIEX/Applied Biosystems QStar Elite hybrid QTOF MS instrument (Applied Biosystems). Samples were introduced into the ESI source by

continuous infusion at a rate of 30 μL/min using the integrated Harvard syringe pump, and the instrument was calibrated daily using a lipid standard mixture. TOF scans of extracts, with no fragmentation, were carried out in positive and negative ion modes, and spectra were examined to identify unique peaks appearing in the experimental vs control samples. In positive mode, the ion spray voltage was 5000 V, the source temperature was 150°C, the curtain gas was 25 (arbitrary units), the ion source gases were 20 and 30 (arbitrary units), the declustering potential was 80 V, the declustering potential 2 was 15 V, the focusing potential was 300 V, and the collision gas, nitrogen, was 3 (arbitrary units). In negative mode, the ion spray voltage was -4500 V and the other listed parameters were the same as in positive mode, with negative voltages. TOF data were collected with 300-600 cumulative scans during 5-10 min acquisitions, over a range of *m/z* 100-2000.

For MS/MS product ion scans of experimental samples, acquisition parameters were the same as listed above for negative mode, except the collision gas was set at 4, and the collision energy was -30 V. Product ion data were collected with 100 cumulative scans during 1.7 min acquisitions, over a range of *m/z* 100-2000.

MS/MS product ion scans in negative ion mode were also performed on purchased fatty acid standards for comparison with experimental sample scans. Standards used were 12-oxo-10,15(Z)-phytodienoic acid (OPDA, $C_{18}H_{28}O_3$, *m/z* 292.2, Larodan), 13-oxo-9(Z),11(E),15(Z)-octadecatrienoic acid (13-KOT, $C_{18}H_{28}O_3$, *m/z* 292.2, Larodan), colnelenic acid ($C_{18}H_{28}O_3$, *m/z* 292.2, Larodan), and etherolenic acid ($C_{18}H_{28}O_3$, *m/z* 292.2, Larodan), all of which produce an $[M - H]^-$ parent ion of *m/z* 291.2 in negative ion mode; 13(S)-hydroperoxy-9(Z),11(E),15(Z)-octadecatrienoic acid (13(S)-HPOT, $C_{18}H_{30}O_4$, *m/z* 310.2, Larodan), which produces an $[M - H - H2O]^-$ fragment ion of *m/z* 291.2 in negative ion mode; 13(S)-hydroxy-9(Z),11(E),15(Z)-octadecatrienoic acid (13(S)-HOT, $C_{18}H_{30}O_3$, *m/z* 294.2, Larodan) and 9(S)-hydroxy-10(E),12(Z),15(Z)-octadecatrienoic acid (9(S)-HOT, $C_{18}H_{30}O_3$, *m/z* 294.2, Cayman Chemical Co.), which produce $[M - H]^-$ parent ions of *m/z* 293.2 in negative ion mode; linolenic acid ($C_{18}H_{30}O_2$, *m/z* 278.2, Nu-Chek Prep), which produces an $[M - H]^-$ parent ion of *m/z* 277.2 in negative ion mode; stearic acid ($C_{18}H_{36}O_2$, *m/z* 284.3, Nu-Chek Prep), which produces an $[M - H]^-$ parent ion of *m/z* 283.3 in negative ion mode; and palmitic acid ($C_{16}H_{32}O_2$, *m/z* 256.2, Nu-Chek Prep), which produces an $[M - H]^-$ parent ion of *m/z* 255.2 in negative ion mode. For these MS/MS fatty acid standard scans, the collision gas was set at 4, the collision energy

ranged from -30 to -45 V, and 100-300 cumulative scans were collected over 1.7-5 min, with a range of *m/z* 50-550. All data were collected and processed using Analyst QS 2.0 software. Accurate masses of the ions were determined to ten thousandths of a mass unit.

FIGURES AND TABLES

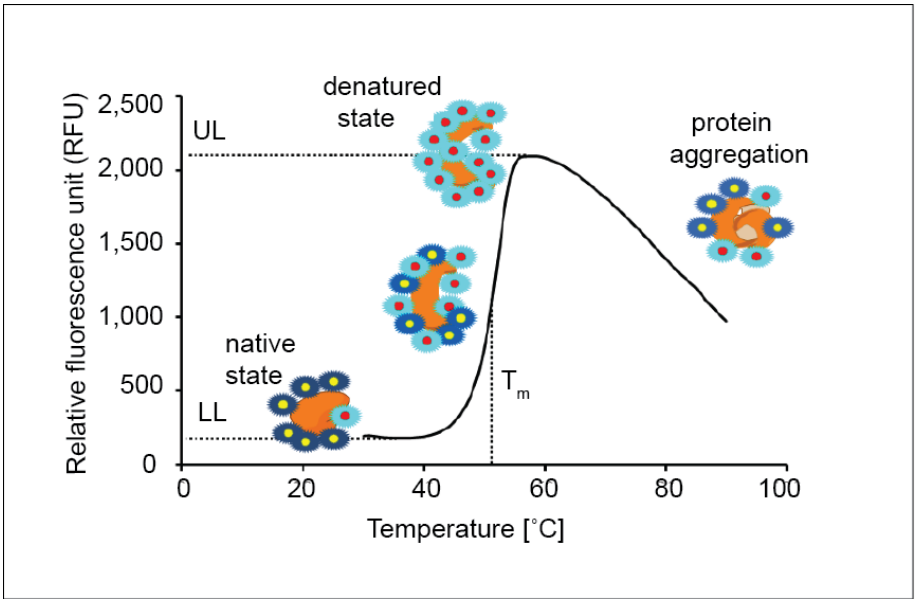


Figure 4.1 Typical DSF melt curve.

Fluorescence intensity plotted as a function of temperature for the unfolding of protein (citrate synthase). The plot has a sigmoidal shape, which is illustrated by a two-state transition depicted in this figure. The lower level (LL) and upper lever (UL) of the fluorescence intensity defined by equation (1) are also demonstrated here. Figure modified from Niesen et al. (2007).

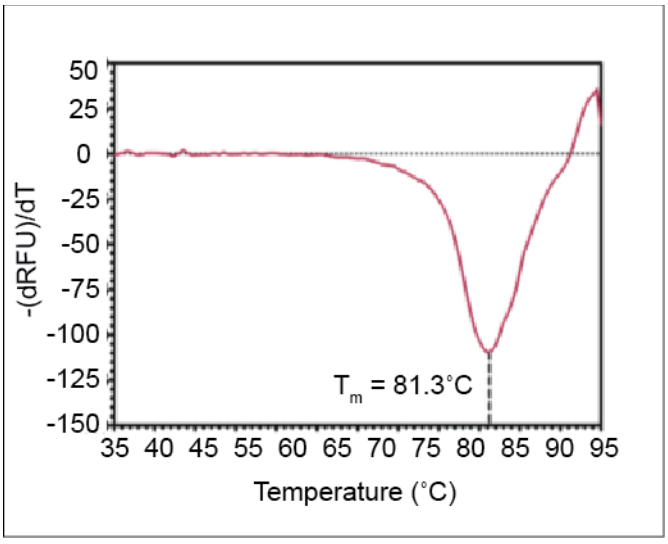


Figure 4.2 Melting curve using the first derivative – (dRFU)/dT of the raw data.
The T_m corresponds to the apex (Boivin et al. 2013).

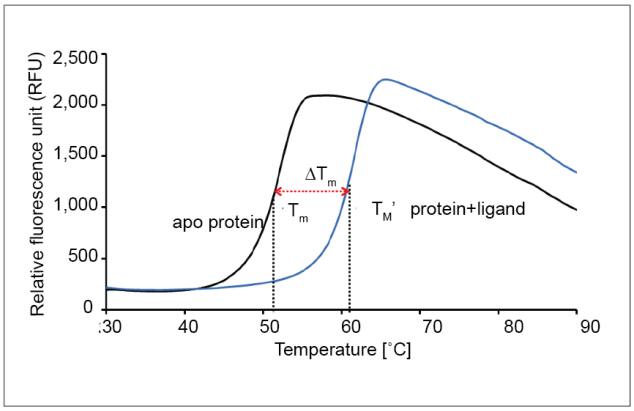


Figure 4.3 Increase in T_m due to ligand binding.

The protein solution is heated in the presence of a hydrophobic dye (SYPRO Orange). Upon denaturation, the dye binds to the internal hydrophobic protein core increasing significantly the fluorescence (left Y axis). Maximal fluorescence intensity is obtained when the protein unfolds completely, then SYPRO Orange signal decreases corresponding to dye-protein dissociation. Usually the fluorescence signal is plotted as a function of temperature to get a sigmoidal curve that shows the fraction of the unfolded protein. The inflection point corresponds to the melting temperature (T_m), at which 50% of the protein is unfolded. A ligand bound to a protein, has the propensity to increase its thermal stability (and hence it's T_m') through newly formed ligand-protein interactions. This difference in melting temperature (ΔT_m) of the protein and of the ligand-protein complex has been shown previously to correlate to measures of the ligand's concentration and binding affinity. Figure modified from Seabrook and Newman (2013)

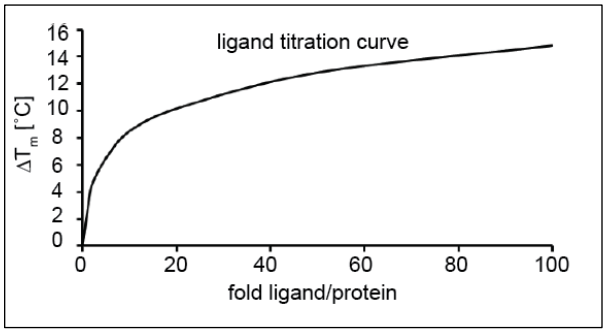


Figure 4.4 Example of DSF titration curve.

The curve represents an exponential fit to the ΔT_m values plotted against the compound concentration.

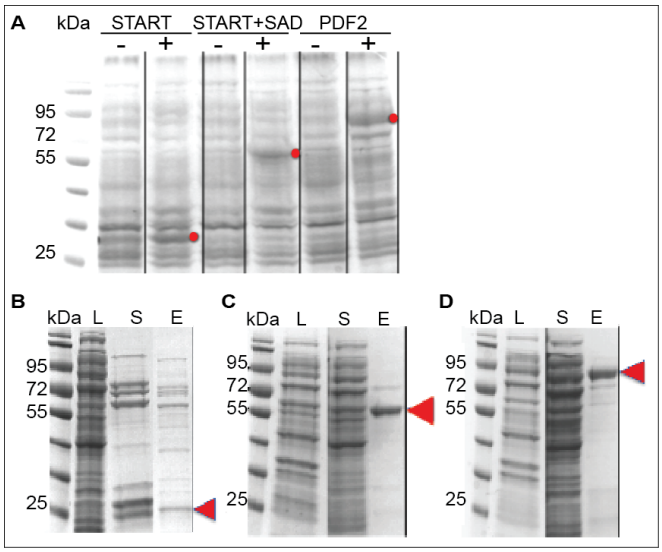


Figure 4.5 Protein purification of His_6-tagged PDF2-START, PDF2-START+SAD, and PDF2.

(A) SDS-PAGE pattern showing induction of proteins at 18˚C (red). Samples were subjected to SDS-PAGE and stained with Coomassie blue. Left lane for each construct is un-induced fraction and right lane is the fraction collected 20 hr post-induction using 0.1 mM IPTG. **(B, C, D)** SDS-PAGE analysis at various stages of purification of recombinant His_6-tagged proteins using Talon-affinity chromatography (Coomassie blue staining). L: Total protein lysate, S: Soluble fraction, E: Elution using 500 mM imidazole for PDF2-START (~29 kDa) **(B)** and 50 mM imidazole for PDF2-START+SAD (~59 kDa) and PDF2 (~85 kDa) **(C, D)**, respectively (red arrows).

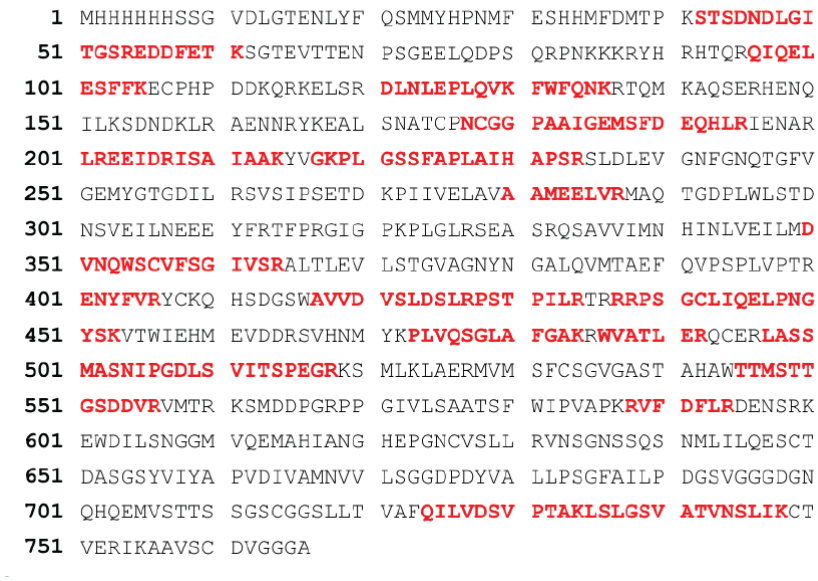


Figure 4.6 PDF2 MASCOT search results.

Each data file was searched against *E. coli* protein database and a custom database with the PDF2 construct sequence. The protein sequence coverage is 32%. Matched peptides are shown in bold red.

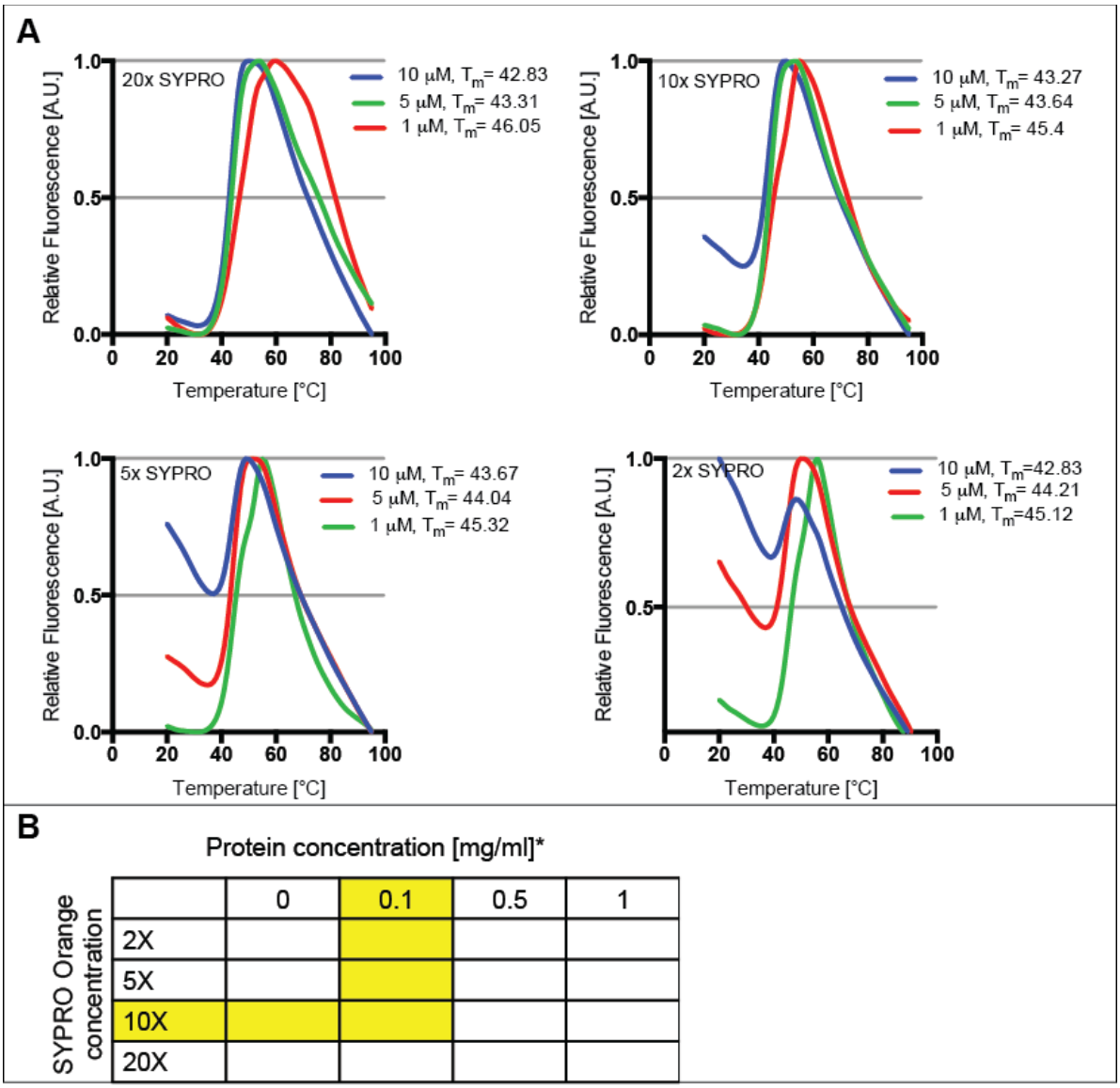


Figure 4.7. DSF optimization for PDF2.

(A) Normalized SYPRO Orange fluorescence plotted as a function of temperature for the unfolding of PDF2. Three different concentrations of protein (1, 5 and 10 µM) and four different dye concentrations (2x, 5x, 10x, 20x) were tested. **(B)** Optimization grid for thermal stability assay using SYPRO Orange. Thermofluor experimental conditions need to be optimized for some proteins The conditions used in this study are colored in yellow.

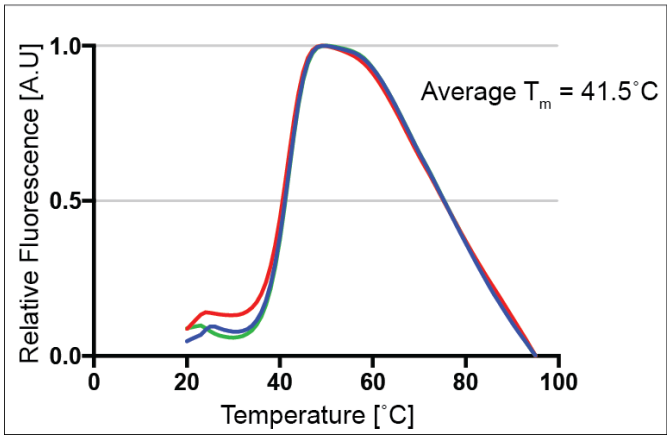


Figure 4.8 Relative SYPRO orange fluorescence plotted as a function of temperature for the unfolding of PDF2.

The average T_m value was 41.5 ± 0.24°C (n=3).

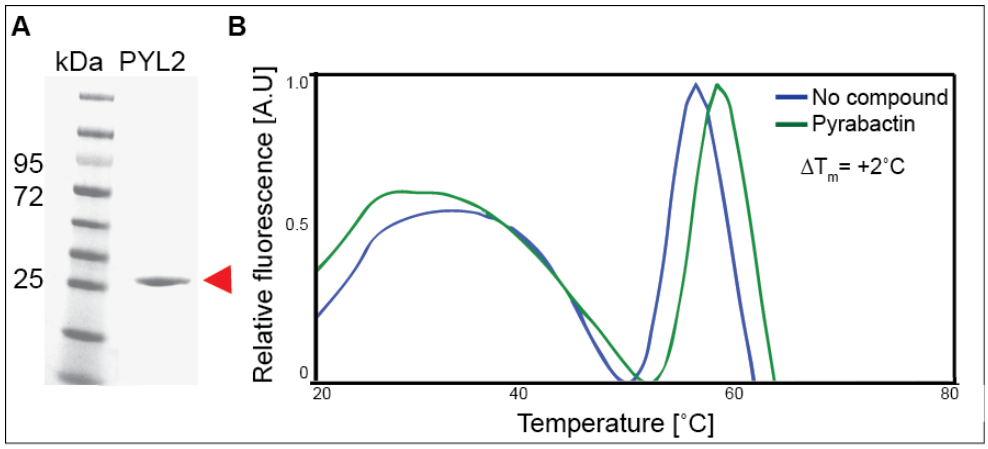


Figure 4.9 DSF assay showing increased thermal stability of PYL2 receptor by pyrabactin.

(A) Purification of His_6-PYL2. SDS-PAGE gel showing the elution fraction of PYL2 during the purification stage. PYL2 (~25 kDa) (red arrowhead). **(B)** Melt curve of 5 µM PYL2 protein in the presence (green) and absence (blue) of 500 µM pyrabactin. The ΔT_m value (as calculated in GraphPad Prism) is indicated.

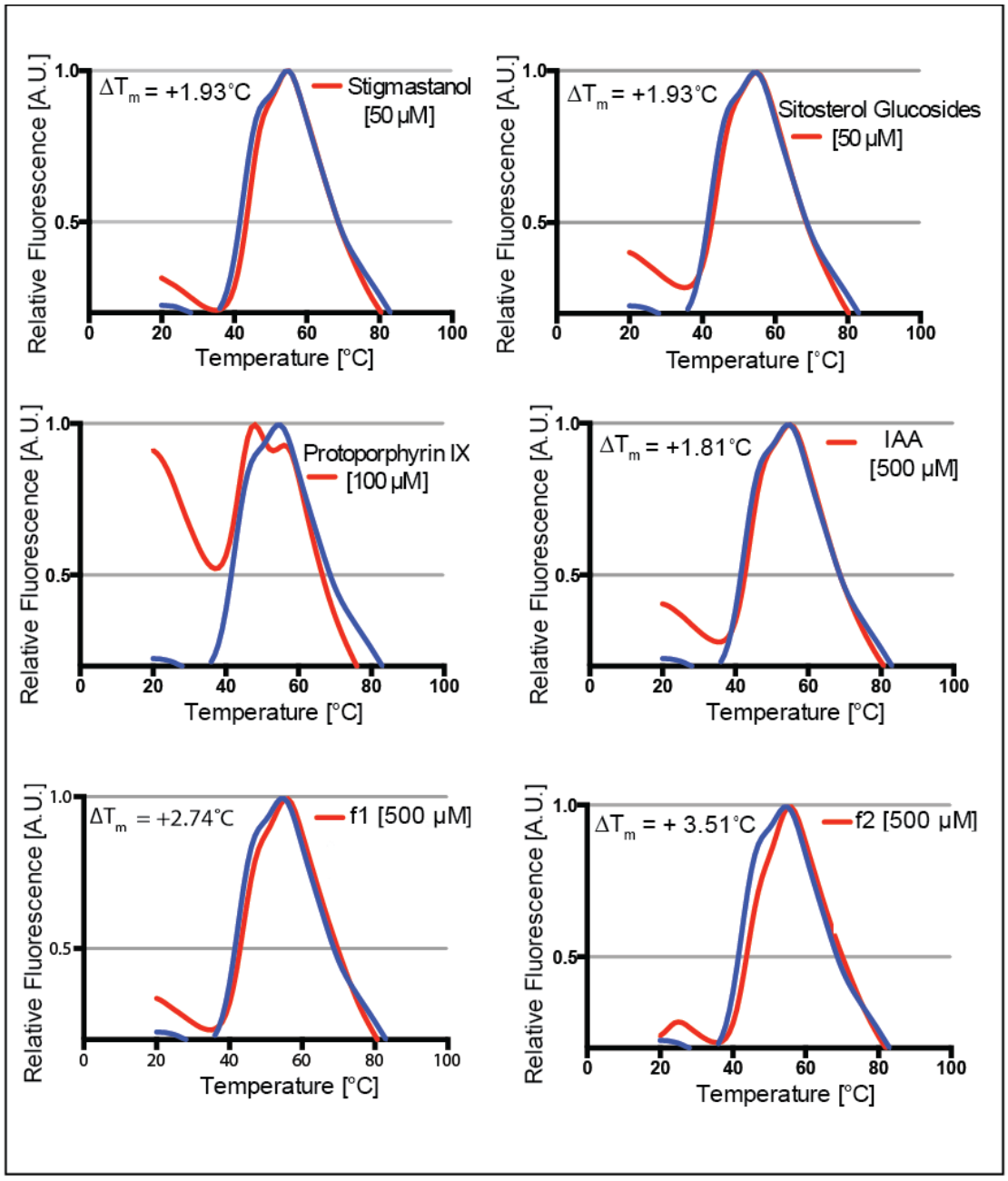


Figure 4.10 Melt curves of PDF2 in the presence of six candidates.

Normalized SYPRO Orange fluorescence plotted as a function of temperature for the unfolding of PDF2. Each graph belonging to the candidate tested (red), is plotted against the DMSO vehicle control (blue). The ΔT_m values (as calculated in GraphPad Prism) are indicated. f1, Kaempferol 3-O-α-L-rhamnoside-7-O-α-L-rhamnoside; f2, Quercetin 3-O-α-L-rhamnoside-7-O-α-L-rhamnoside.

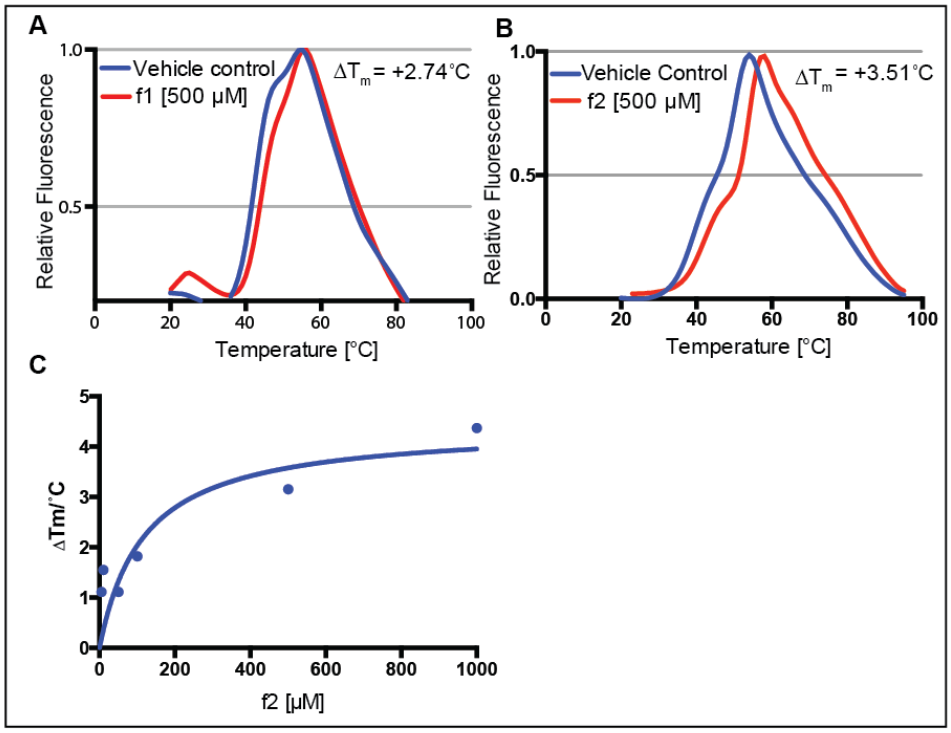


Figure 4.11 Quercetin 3-O-α-L-rhamnoside-7-O-α-L-rhamnoside (f2) induced stabilization of PDF2.

Melt profile of **(A)** 5 µM PDF2 and **(B)** 1 µM PDF2 in the presence of f2. Shift in the melt curve upon addition of 500 µM of f2 suggests a stabilization effect due to binding.

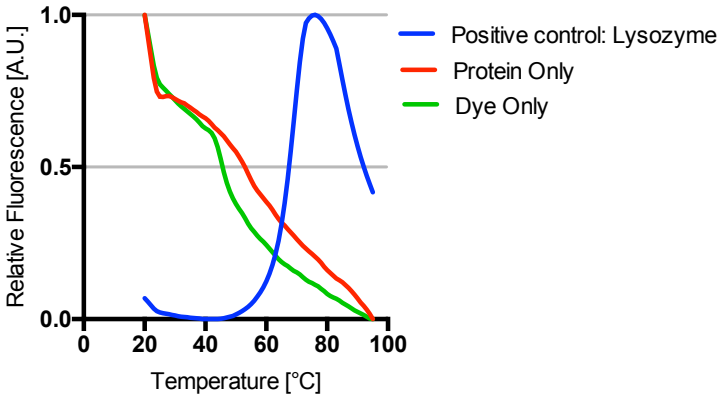


Figure 4.12 Melt curves for the controls.

0.01 mg/ml lysozyme (blue), protein only (red curve), 10x SYPRO Orange dye only (green curve)

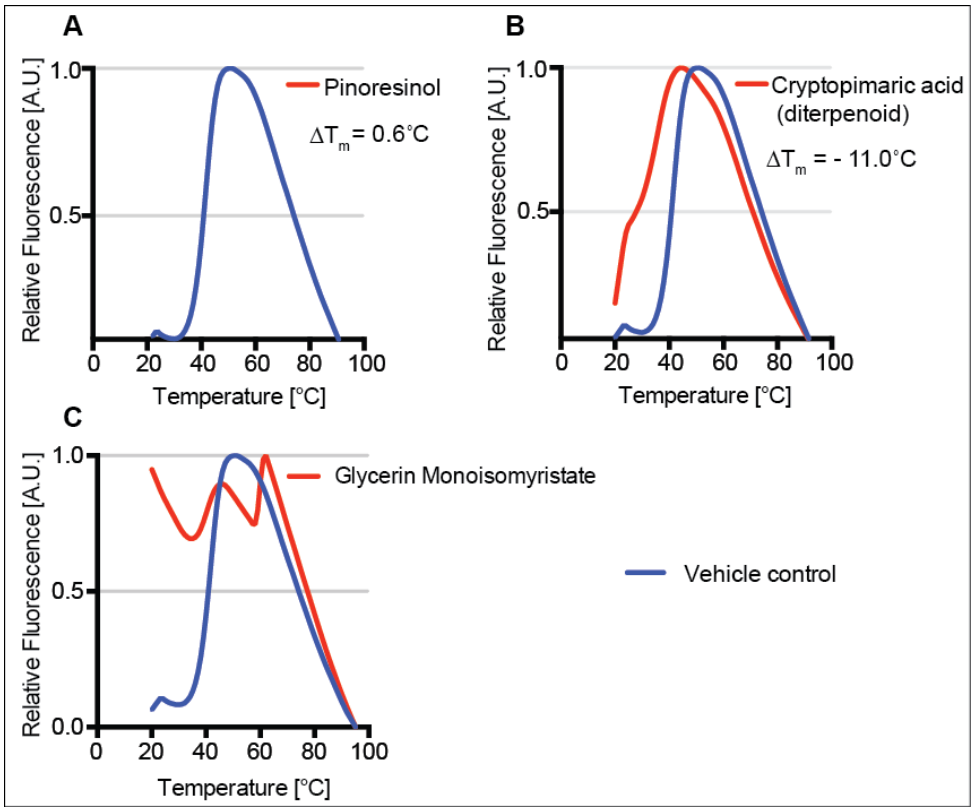


Figure 4.13 Relative SYPRO orange fluorescence plotted as a function of temperature for the unfolding of PDF2.

(A) Pinoresinol resulted in no significant ΔT_m value (0.6˚C), suggesting that this compound had no effect on the protein's thermal stability. (B) Cryptopimaric acid (diterpenoid) resulted in a negative thermal shift (left shift). The ΔT_m value of this compound as calculated is – 11°C which indicates a significant destabilization effect on PDF2. (C) Glycerin Monoisomyristate. The unfolding profile in the presence of this compound shows high initial fluorescence value decreasing with temperature.

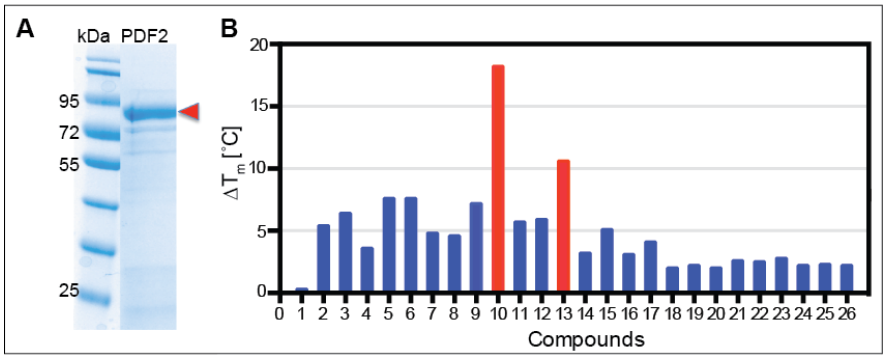


Figure 4.14 DSF assay to screen for natural compounds that increase the stability of PDF2.

(A) SDS-PAGE showing the elution fraction of His_6-tagged PDF2 after purification. A single band at an estimated molecular weight of 85 kDa is observed (red arrowhead). **(B)** DSF identified 25 putative ligands from the AnalytiCon MEGx library of pure natural compounds from plants and microbes. A graph of ΔT_m values of hits determined at compound concentration of 300 µM and protein concentration of 1 µM. **[1]** DMSO vehicle control is indicated. **[10]** benzylated chalcone and **[13]** pinoquercetin exhibit the largest thermal shifts out of 1000 compounds screened (red). The ΔT_m values are listed in Table 4.3

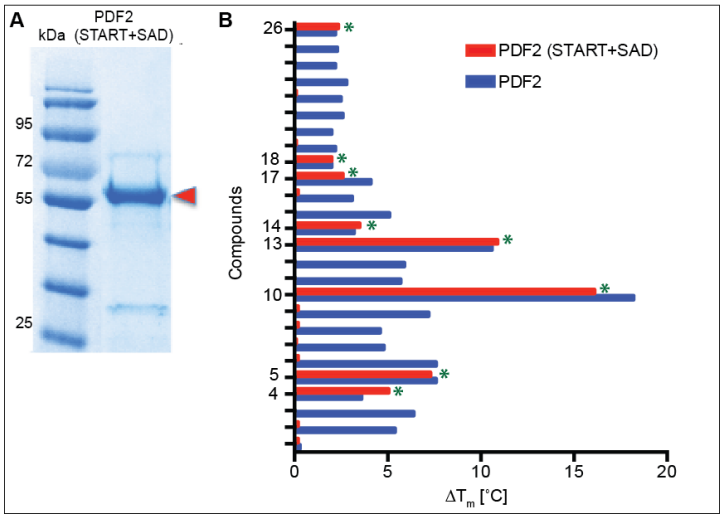


Figure 4.15 DSF assay using PDF2-START+SAD protein.

(A) SDS-PAGE gel showing the elution fraction of His_6-tagged PDF2-START+SAD after purification. A band at an estimated molecular weight of 59 kDa is observed (red arrowhead). **(B)** Comparison of binding profiles of PDF2 and PDF2-START+SAD proteins in the presence of 25 putative ligands. Eight compounds tested positive for both of the proteins (green asterisks, Refer Table 4.4).

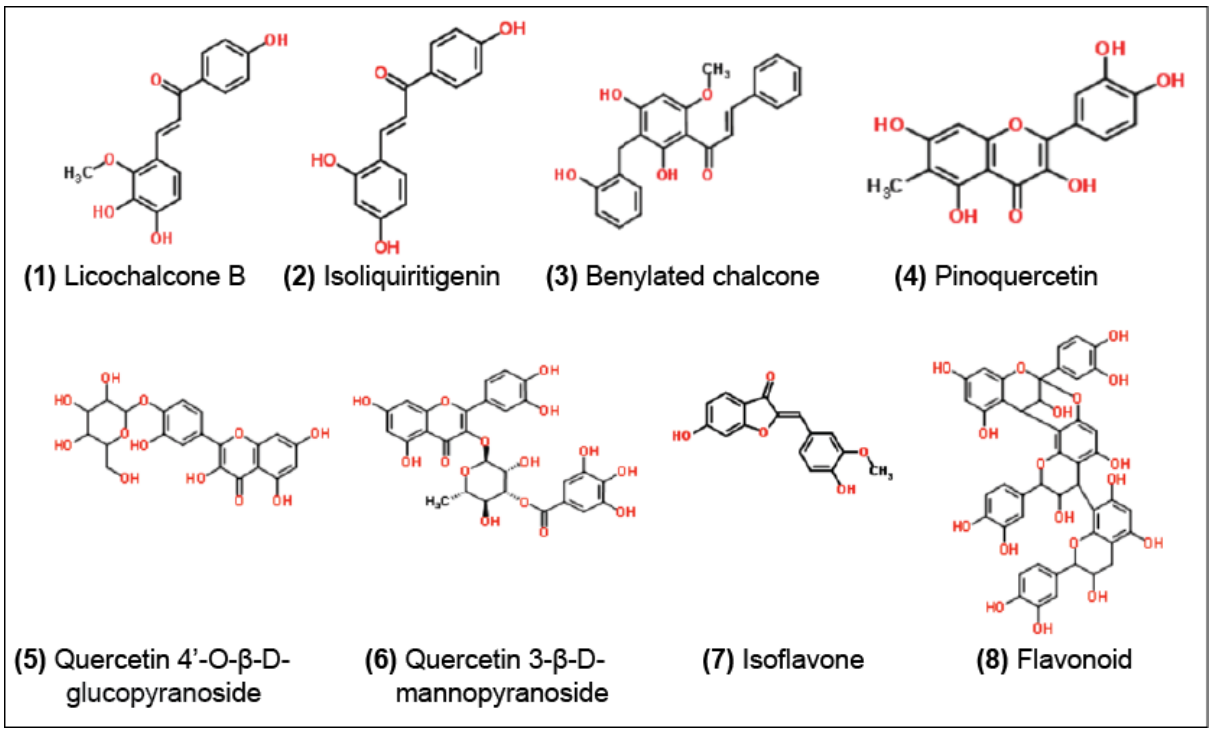


Figure 4.16 Structures of hits identified by screening the AnalytiCon MEGx library using DSF.

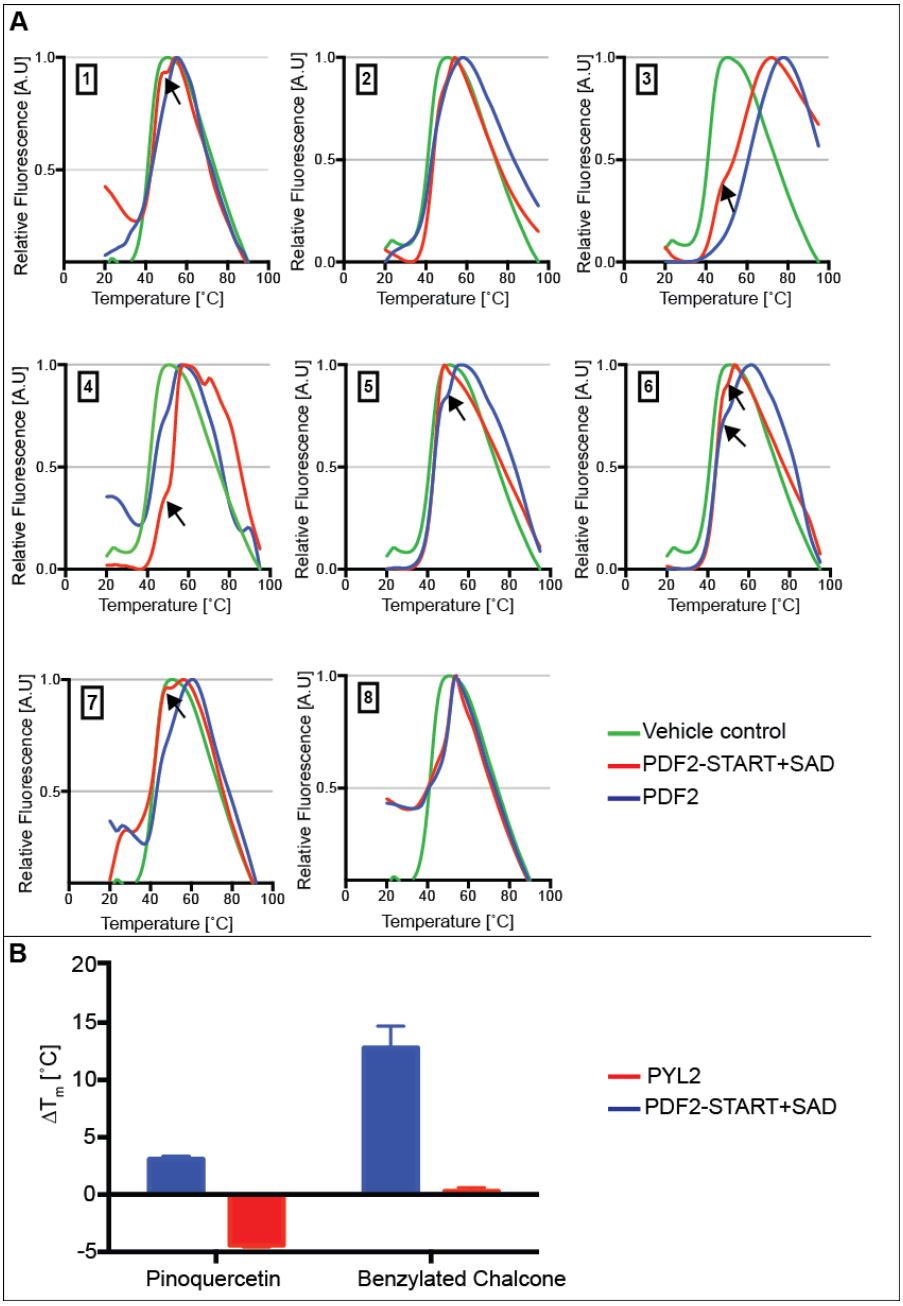


Figure 4.17 Normalized melt curves for PDF2 and PDF2-START+SAD proteins (1 µM) in the presence of the top hits (300 µM) and vehicle control.

(A) The positions of uncomplexed, excess ligands are represented by arrows. The exact ΔT_m values (as calculated in Graphpad) are listed in Table 4.3. **[1]** licochalcone B; **[2]** quercetin 3-β-D-mannopyranoside; **[3]** benzylated chalcone 2; **[4]** pinoquercetin; **[5]** flavonoid; **[6]** (2Z)-6-Hydroxy-2-(4-hydroxy-3-methoxybenzylidene)-1-benzofuran-3(2H)-one; **[7]** quercetin-4'-O- β-D-glucopyranoside; and **[8]** isoliquiritigenin. **(B)** Binding of pinoquercetin and benzylated chalcone to negative control protein (PYL2) is shown.

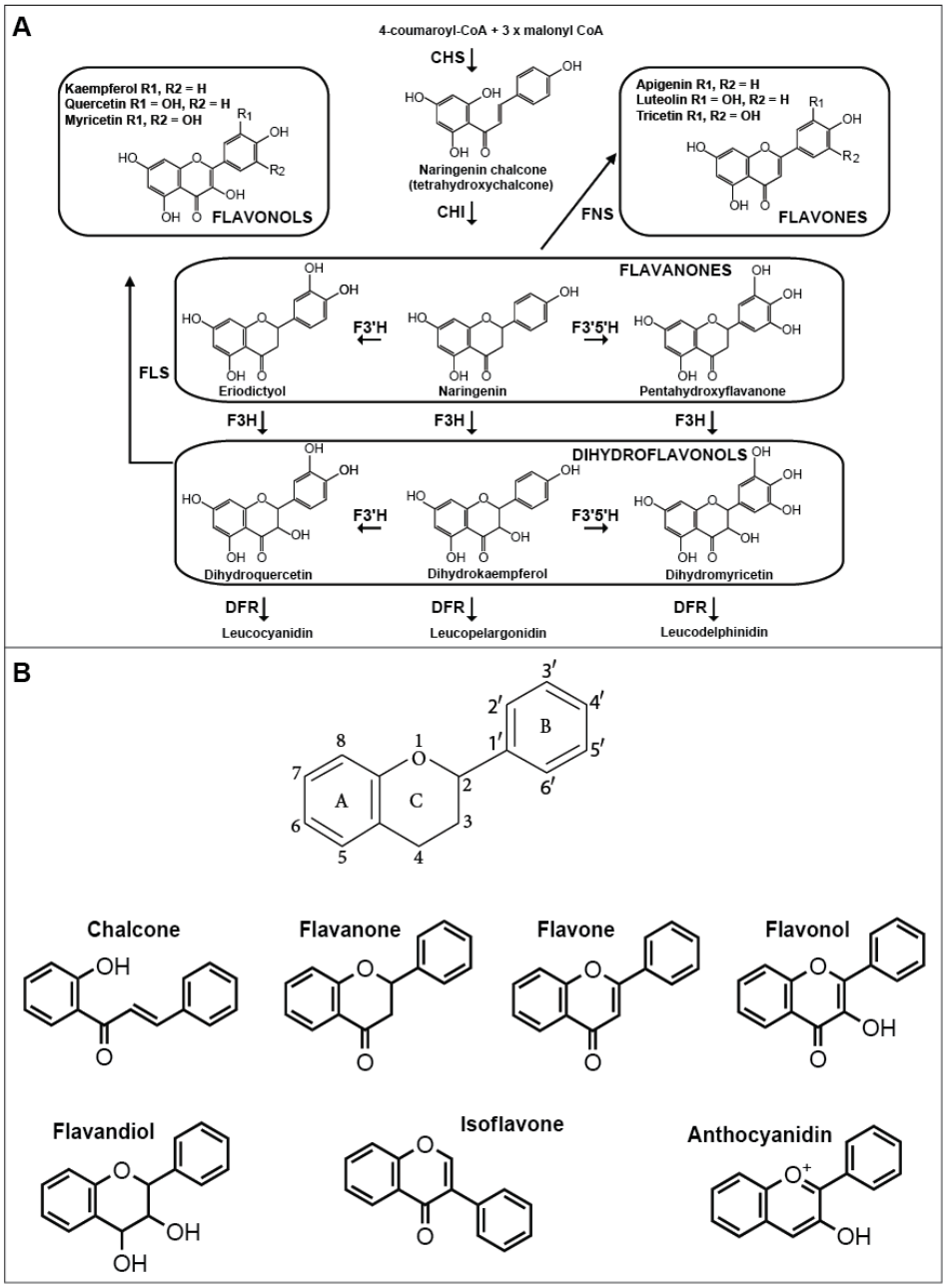


Figure 4.18 The flavonoid biosynthetic pathway in *Arabidopsis*

(A) General biosynthetic pathway of phenylpropanoid and flavonoid skeletons. The enzymes catalyzing each step are indicated in bold. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase (modified from Saito et al. (2013). **(B) Structures of the main classes of flavonoids.** Flavonoids occur as aglycones, glycosides, and

methylated derivatives. The numbering system of the flavonoid skeleton is indicated on the basic flavonoid structure, which is aglycone.

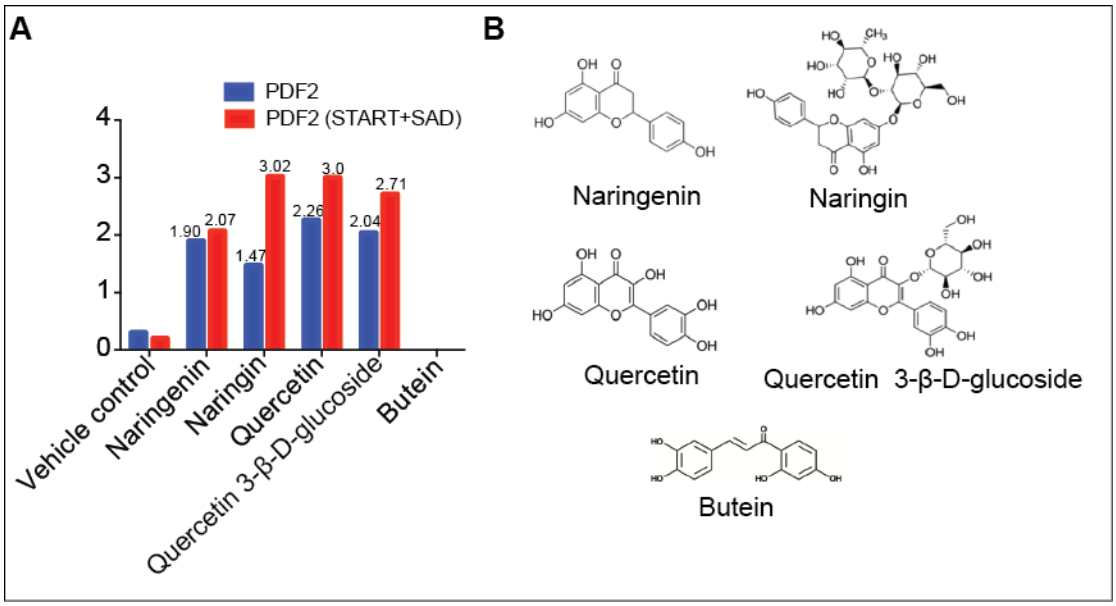


Figure 4.19 Binding of different flavonoid compounds by PDF2 and PDF2-START+SAD proteins.

(A) A graph of ΔT_m values determined at compound concentration of 300 µM for butein, quercetin-3-β-D-glucoside, quercetin, naringin, and naringenin. T_m values were measured by DSF, and the ΔT_m values were calculated. Both quercetin and naringin induced highest thermal shift for the PDF2-START+SAD protein. **(B)** Structures of the hits are depicted.

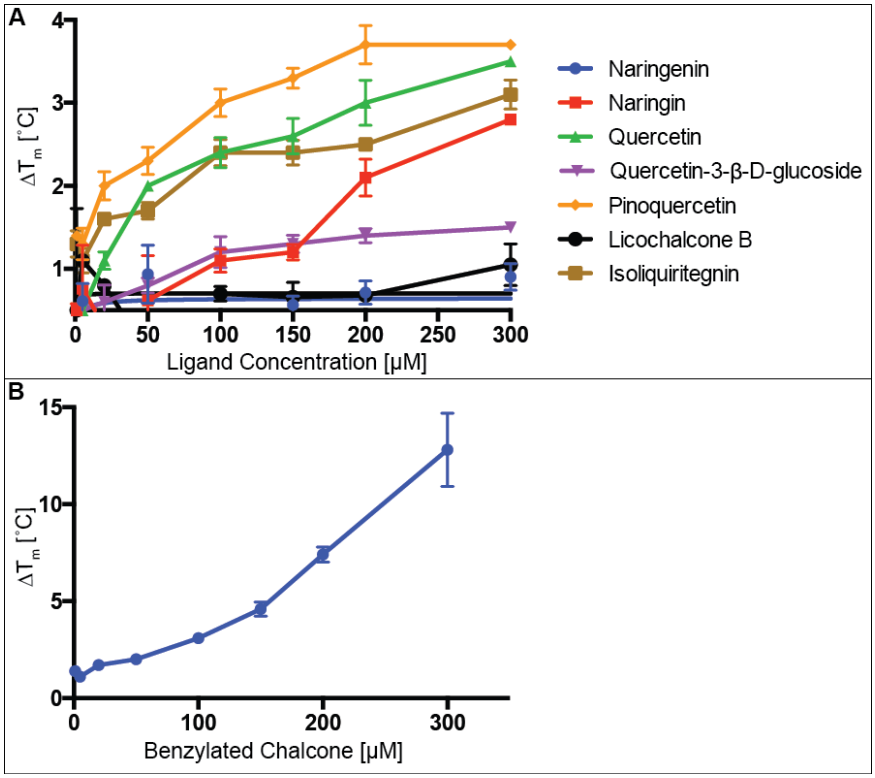


Figure 4.20 Binding of the candidate ligands by PDF2-START+SAD.

(A) Concentration-dependent stabilization of PDF2-START+SAD by the ligands: naringin, naringenin, quercetin, quercetin-3-β-D-glucoside, pinoquercetin, licochalcone B, isoliquiritegnin, and **(B)** Benzylated chalcone ΔT_m was referenced to the T_m value of a control in which the protein was incubated with the corresponding concentration of DMSO. Each point represents average of five measurements. Averages are shown for n=5 samples and error bars indicates standard deviations. Some error bars are smaller than the size of the data point.

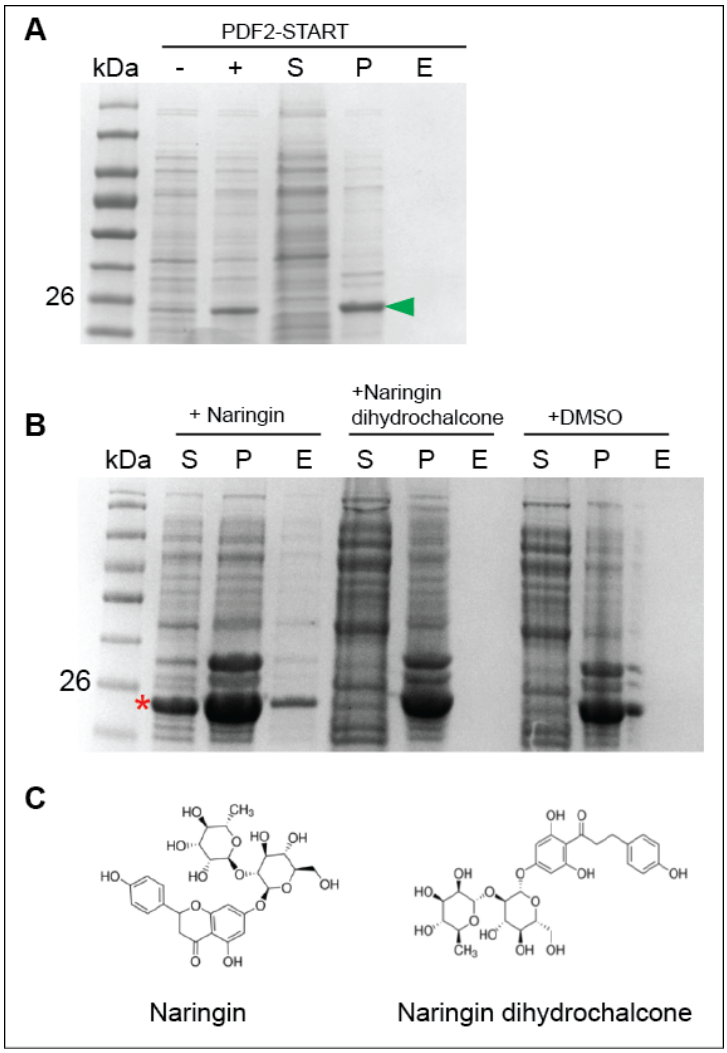


Figure 4.21 Expression of PDF2-START protein analyzed by SDS-PAGE.

(A) His_6-tagged PDF2-START displays poor solubility. **(B)** Ligand supplementation (100 µM during protein purification) leads to increased soluble expression of the PDF2-START protein (asterisk). S, Soluble; P, Pellet; E, Eluted fraction using 500 µM imidazole. **(C)** Structures of the compounds are depicted.

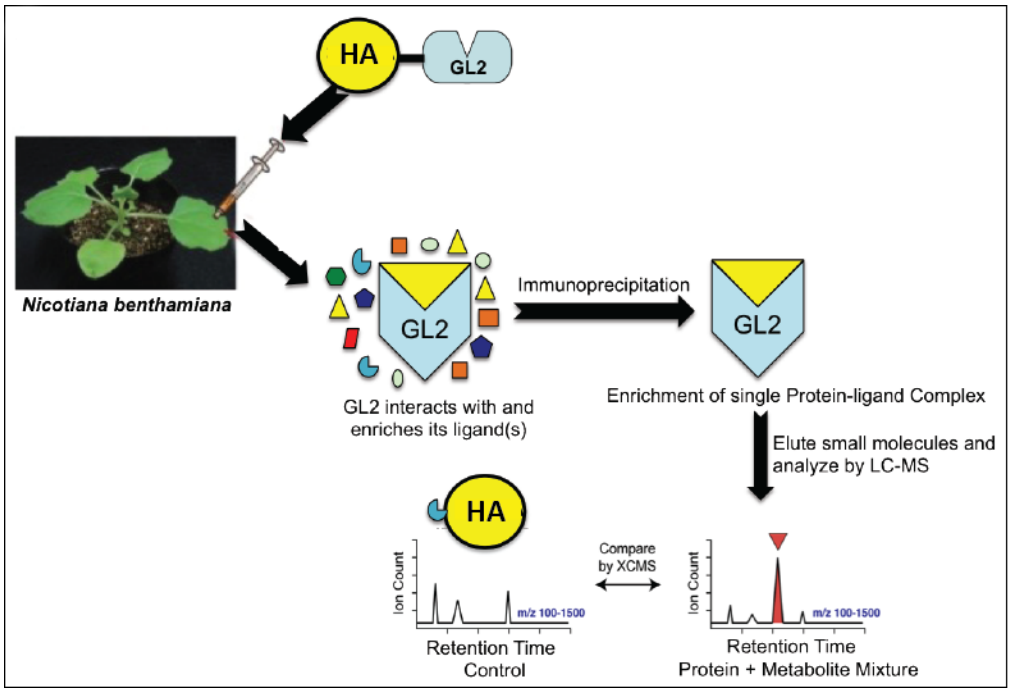


Figure 4.22 Pull down assay for identification of potential START domain ligands. Workflow for the identification of putative protein-metabolite interactors.

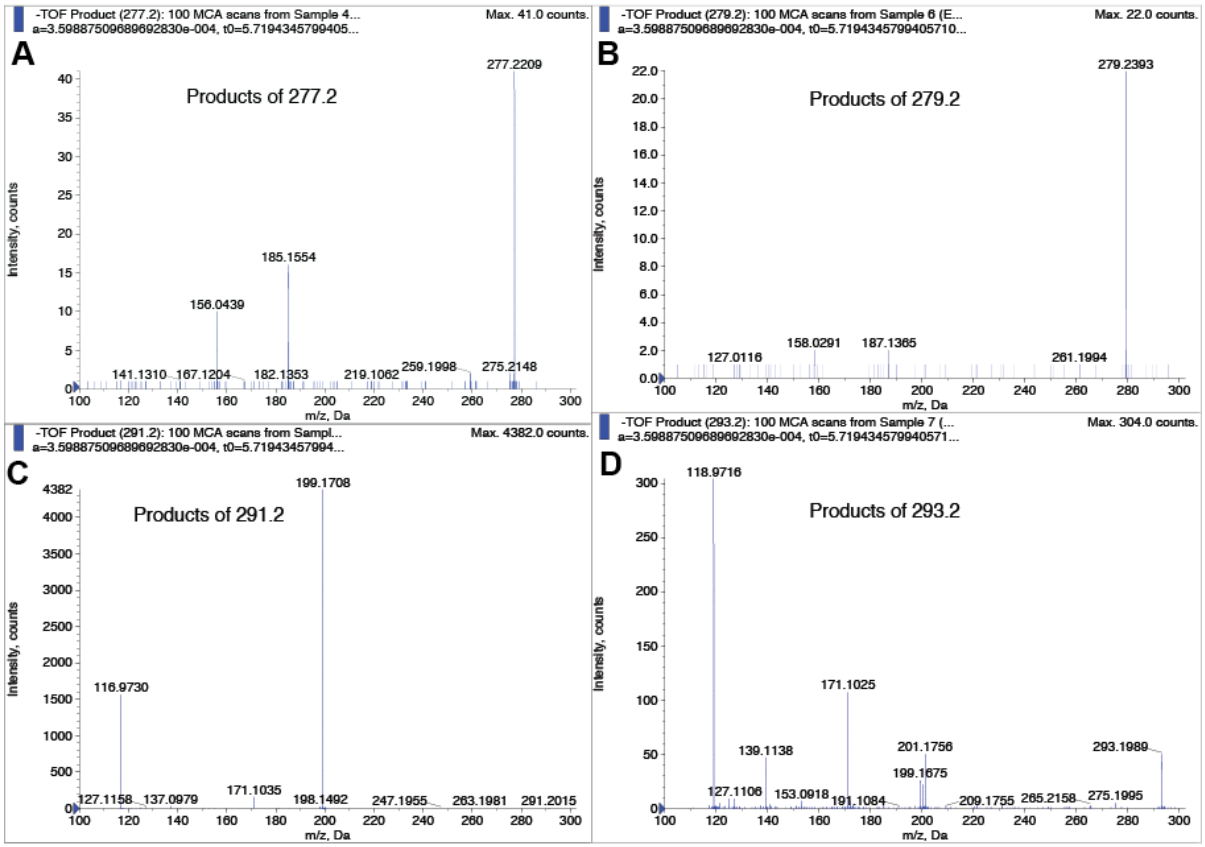


Figure 4.23 ESI-Q-TOF MS/MS showing the negative ion fragmentation spectrum.

Peaks at m/z **(A)** 291.2, **(B)** 293.2, **(C)** 277.2, and **(D)** 279.2. Collision energy -30eV.

Table 4-1 DSF screen results for full-length PDF2.

Results of initial DSF screening of the AnalytiCon MEGx library against PDF2

Type	Effect	# of compounds	ΔT_m (˚C)
Protein stabilizers	Positive thermal shift indicating an increase in protein stability (>or= 2˚C)	25	+2.28- 18˚C
No effect on protein stability	No apparent significant shift in ΔT_m (<1˚C)	868	
Destabilizers	Decrease in protein thermal stability (Negative ΔT_m)	88	
Denaturants	No measurable transition (potential aggregation/complete denaturation of the protein) or destabilization/partial unfolding	19	

Table 4-2 DSF data of 25 hits from the AnalytiCon MEGx library.

#	[a]NP#	[b]CAS #	Class of compounds	ΔT_m (˚C)
2	016976	81905-14-8	Flavonoid Glycoside	5.4
3	012245	89946-11-2	Stilbenes	6.4
4	012489	58749-23-8	Chalcone	3.6
5	004204	NA	Flavonoid	7.6
6	016356	10035-27-5	Lignans	7.6
7	001289	27208-80-6	Stilbene Glycoside	4.8
8	017432	197379-54-7	Flavonoid Glycoside	4.6
9	000759	3952-18-9	Biflavonol	7.2
10	000319	102056-84-8	Chalcones	18.2
11	005115	18942-26-2	Flavonoid	5.7
12	006331	NA	Flavonoid	5.9
13	012356	491-49-6	Flavonol	10.6
14	003298	NA	Flavonoid Glycoside	3.2
15	012592	NA	Flavonoid	5.1
16	010620	480-16-0	Flavonoid	3.1
17	010061	32396-80-8	Flavonoid	4.1
18	016700	20229-56-5	Flavonoid Glycoside	2.0
19	017433	NA	Benzyl Benzoate	2.2
20	017505	160824-52-2	Saponins	2.0
21	000318	120-51-4	Benzyl Benzoate	2.6
22	010021	NA	Curcuminoid	2.5
23	014109	89702-01-2	Flavonoid Glycoside	2.8
24	012957	6920-38-3	Flavonoid Glycoside	2.2
25	017130	NA	Flavonoid Glycoside	2.3
26	004577	13745-20-5	Chalcone	2.2

[a] Arbitrary natural product (NP) library number.

[b] CAS numbers are unique identifiers assigned by the "Chemical Abstracts Service" to describe every chemical described in open access scientific literature.

Table 4-3 DSF data for the eight top hits from AnalytiCon MEGx library.

#	Compound	Compound ID	XLogP3	TPSA	MW	ΔT_m at 1 µM PDF2 (˚C)	ΔT_m at 1 µM PDF2-START+SAD
4	Licochalcone B	NP-012489	2.57	87 A^2	286.279	3.6	5.05
5	Quercetin 3-β-D-mannopyranoside	NP-003298	3.71	253 A^2	600.481	3.2	3.5
10	Benzylated chalcone 2	NP-000319	4.87	87 A^2	376.402	18.2	16.1
13	Pinoquercetin	NP-012356	2.54	127 A^2	316.202	10.6	10.9
14	5,13-Bis(3,4-dihydroxyphenyl)-7-[2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-3,4-dihydro-2H-chromen-8-yl]-4,12,14 trioxapentacyclo[11.7.1.0^{2,11}.0^{3,8}.0^{15,20}]henicosa-2,8,10,15,17,19-hexaene-6,9,17,1 9,21-pentol	NP-004204	1.38	320 A^2	864.757	7.6	7.3
17	(2Z)-6-Hydroxy-2-(4-hydroxy-3-methoxybenzylidene)-1-benzofuran-3(2H)-one	NP-010061	2.2	76 A^2	284.263	4.1	2.6
18	Quercetin 4'-O-β-D-glucopyranoside	NP-016700	-0.38	207 A^2	464.376	2	2
26	Isoliquiritigenin	NP-004577	3.14	78 A^2	256.273	2.2	2.35

XLogP3- The logarithm of the partition coefficient between n-octanol and water (immiscible phases). Measure of lipophilicity of a compound

TPSA (Topological Polar Surface Area.) - Estimate of the area (in Å squared), which is polar. Measure of polarity and hydrophobicity

Compound ID- Arbitrary natural compound (NP) library number

Table 4-4 Oligonucleotide primers used in this study.

I. Primers used for cloning genes into bacterial expression vectors for protein expression and purification. Homologous sequences for in-fusion cloning into pNIC28-Bsa4 are indicated in bold.		
Name	**5'-3' sequence**	**Expression vector**
PDF2 (START)_SgfI_F	GTTTCGGCGATCGCTGAGACTGATAA	pFN2A (GST) flexi
PDF2 (START)_PmeI_R	CACCAGGAGTTTAAACGGCCATGGAG	pFN2A (GST) flexi
HDG11 (START)_SgfI_F	CCTAACTTGGCGATCGCAGACATGGA	pFN2A (GST) flexi
HDG11 (START)_PmeI_R	GAGATCACGAGATGAAGTTTAAACTA	pFN2A (GST) flexi
HDG12 (START)_SgfI_F	CCAAACTTGGCGATCGCAATGGA	pFN2A (GST) flexi
HDG12 (START)_Pme_R	CAAGGGATGATGTTGCAGGTGTTTAA	pFN2A (GST) flexi
PHB (START)_SgfI_F	GCATCAGCAACGTGCGATCGCCAAC	pFN2A (GST) flexi
PHB (START)_PmeI_R	CTGAACTTCTCCAGTTTAAACTTGTGC	pFN2A (GST) flexi
GL2 (START)_SgfI_F	GGGCGTCGCGATCGCCGAGAAGTCC	pFN2A (GST) flexi
GL2 (START)_PmeI_R	CTTTGGTGGTTTAAACGGTAGCCATG	pFN2A (GST) flexi
ATML1 (START)_SgfI_F	GGTTTCGGCGATCGCTGAGGCTGATA	pFN2A (GST) flexi
ATML1 (START)_Pme_R	GATCACAAGCCGGAGTTTAAACGGCC	pNIC28-Bsa4
PDF2 (START+SAD)_F	**TACTTCCAATCCATG**GGAACAGGGGACATTTTGAGGTCAG	pNIC28-Bsa4
PDF2 (START+SAD)_R	**TATCCACCTTTAATGTTA**CTACGCTCCTCCTCCAACATCAC	pNIC28-Bsa4
PDF2 (1-743)_F	**TACTTCCAATCCATG**TACCATCCAAACATGTTTGAGAGC	pNIC28-Bsa4
PDF2 (1-743)_R	**TATCCACCTTTACTGTTA**CTACGCTCCTCCTCCAACATCAC	pNIC28-Bsa4
PDF2 (START)_F	**TACTTCCAATCCATG**GAGACTGATAAGCCTATAATCGTG	pNIC28-Bsa4
PDF2 (START)_R	**TATCCACCTTTACTGTTA**CATGGAGCTAGCAAGCCG	pNIC28-Bsa4
GL2 (1-747)_F	**TACTTCCAATCCATG**TCAATGGCCGTCGACATGTC	pNIC28-Bsa4
GL2 (1-747)_R	**TATCCACCTTTACTGTTA**CATCAGCAATCTTCGATTTGTAGAC	pNIC28-Bsa4
GL2 (START+SAD)_F	**TACTTCCAATCCATG**GACGATCAAGAACACCGTCTCG	pNIC28-Bsa4
GL2 (START+SAD)_R	**TATCCACCTTTACTGTTA**CATCAGCAATCTTCTTCGATTTGTAGAC	pNIC28-Bsa4
II. Primers for site directed mutagenesis of PDF2-START. Mutated residues are shown in red		
PDF2_L467P_F	CGTTGGGTGGCTACACCAGAACGACAATGCGAG	
PDF2_L467P_R	CTCGCATTGTCGTTCTGGTGTAGCCACCCAACG	
III. Primers for sequence verification		
Name	**5'-3' sequence**	**Purpose**
pLIC_F	TGTGAGCGGATAACAATTCC	Sequencing His_6-tagged constructs
pLIC-R	AGCAGCCAACTCAGCTTCC	Sequencing His_6-tagged constructs
pFlexi_R	AGCCGGATCAGCTTGCATGC	Sequencing GST tagged constructs
pro35S_seq	CACGCTCGAGTATAAGAGCTC	Sequencing gl2ΔZLZ-201

REFERENCES

Abad MC, Askari H, O'Neill J, Klinger AL, Milligan C, Lewandowski F, Springer B, Spurlino J, Rentzeperis D. 2008. Structural determination of estrogen-related receptor gamma in the presence of phenol derivative compounds. *The Journal of Steroid Biochemistry and Molecular Biology* **108**: 44-54.

Alpy F, Tomasetto C. 2005. Give lipids a START: the StAR-related lipid transfer (START) domain in mammals. *J Cell Sci* **118**: 2791-2801.

Bledsoe RK, Montana VG, Stanley TB, Delves CJ, Apolito CJ, McKee DD, Consler TG, Parks DJ, Stewart EL, Willson TM et al. 2002. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **110**: 93-105.

Boivin S, Kozak S, Meijers R. 2013. Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification* **91**: 192-206.

Brandts JF, Lin LN. 1990. Study of Strong to Ultratight Protein Interactions Using Differential Scanning Calorimetry. *Biochemistry-Us* **29**: 6927-6940.

Bruylants G, Wouters J, Michaux C. 2005. Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design. *Current Medicinal Chemistry* **12**: 2011-2020.

Buer CS, Imin N, Djordjevic MA. 2010. Flavonoids: New Roles for Old Molecules. *J Integr Plant Biol* **52**: 98-111.

Butt TR, Edavettal SC, Hall JP, Mattern MR. 2005. SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification* **43**: 1-9.

Carson M, Johnson DH, McDonald H, Brouillette C, DeLucas LJ. 2007. His-tag impact on structure. *Acta Crystallogr D* **63**: 295-301.

Casanal A, Zander U, Munoz C, Dupeux F, Luque I, Botella MA, Schwab W, Valpuesta V, Marquez JA. 2013. The strawberry pathogenesis-related 10 (PR-10) Fra a proteins control flavonoid biosynthesis by binding to metabolic intermediates. *The Journal of Biological Chemistry* **288**: 35322-35332.

Cimmperman P, Baranauskiene L, Jachimoviciute S, Jachno J, Torresan J, Michailoviene V, Matuliene J, Sereikaite J, Bumelis V, Matulis D. 2008. A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophysical Journal* **95**: 3222-3231.

Costa S, Almeida A, Castro A, Domingues L. 2014. Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front Microbiol* **5**.

Cummings MD, Farnum MA, Nelen MI. 2006a. Universal screening methods and applications of ThermoFluor. *J Biomol Screen* **11**: 854-863.

Cunningham C, Porter AJR. 1998. *Recombinant proteins from plants : production and isolation of clinically useful compounds*. Humana Press, Totowa, N.J.

DeSantis K, Reed A, Rahhal R, Reinking J. 2012. Use of differential scanning fluorimetry as a high-throughput assay to identify nuclear receptor ligands. *Nuclear Receptor Signaling* **10**: e002.

Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P. 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**: 289-298.

Filgueira CS, Benod C, Lou X, Gunamalai PS, Villagomez RA, Strom A, Gustafsson JA, Berkenstam AL, Webb P. 2014. A screening cascade to identify ERbeta ligands. *Nuclear Receptor Signaling* **12**: e003.

Fischer R, Stoger E, Schillberg S, Christou P, Twyman RM. 2004. Plant-based production of biopharmaceuticals. *Curr Opin Plant Biol* **7**: 152-158.

Frei AP, Jeon OY, Kilcher S, Moest H, Henning LM, Jost C, Pluckthun A, Mercer J, Aebersold R, Carreira EM et al. 2012. Direct identification of ligand-receptor interactions on living cells and tissues. *Nat Biotechnol* **30**: 997-1001.

Giddings G. 2001. Transgenic plants as protein factories. *Curr Opin Biotech* **12**: 450-454.

Gingras AC, Gstaiger M, Raught B, Aebersold R. 2007. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Bio* **8**: 645-654.

Gou JY, Li K, Wu K, Wang X, Lin H, Cantu D, Uauy C, Dobon-Alonso A, Midorikawa T, Inoue K et al. 2015. Wheat Stripe Rust Resistance Protein WKS1 Reduces the Ability of the Thylakoid-Associated Ascorbate Peroxidase to Detoxify Reactive Oxygen Species. *Plant Cell* **27**: 1755-1770.

Graslund S, Nordlund P, Weigelt J, Bray J, Hallberg BM, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R et al. 2008. Protein production and purification. *Nat Methods* **5**: 135-146.

Hammarstrom M, Hellgren N, Van den Berg S, Berglund H, Hard T. 2002. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci* **11**: 313-321.

Hanada K, Kumagai K, Yasuda S, Miura Y, Kawano M, Fukasawa M, Nishijima M. 2003. Molecular machinery for non-vesicular trafficking of ceramide. *Nature* **426**: 803-809.

Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL, 3rd, Deng S-JJ, Gampe RT, Grisard TE, Madauss KP, Nolte RT et al. 2007. Crystallization of protein-ligand complexes. *Acta Crystallographica Section D, Biological Crystallography* **63**: 72-79.

Heinz DW, Ryan M, Smith MP, Weaver LH, Keana JF, Griffith OH. 1996. Crystal structure of phosphatidylinositol-specific phospholipase C from Bacillus cereus in complex with glucosaminyl (alpha 1-->6)-D-myo-inositol, an essential fragment of GPI anchors. *Biochemistry* **35**: 9496-9504.

Hood EE. 2002. From green plants to industrial enzymes. *Enzyme Microb Tech* **30**: 279-283.

Horibata Y, Sugimoto H. 2009. StarD7 Mediates the Intracellular Trafficking of Phosphatidylcholine to Mitochondria. *Journal of Biological Chemistry* **285**: 7358-7365.

Hozjan V, Guo K, Wu X, Oppermann U. 2008. Ligand supplementation as a method to increase soluble heterologous protein production. *Expert Rev Proteomics* **5**: 137-143.

Hsu FF, Turk J. 1999. Distinction among isomeric unsaturated fatty acids as lithiated adducts by electrospray ionization mass spectrometry using low energy collisionally activated dissociation on a triple stage quadrupole instrument. *J Am Soc Mass Spectr* **10**: 600-612.

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols* **7**: 1534-1550.

Jiang W, Yin Q, Wu R, Zheng G, Liu J, Dixon RA, Pang Y. 2015. Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. *Journal of Experimental Botany*.

Kean J, Cleverley RM, O'Ryan L, Ford RC, Prince SM, Derrick JP. 2008. Characterization of a CorA Mg^{2+} transport channel from *Methanococcus jannaschii* using a Thermofluor-based stability assay. *Molecular Membrane Biology* **25**: 653-663.

Kervinen J, Ma HC, Bayoumy S, Schubert C, Milligan C, Lewandowski F, Moriarty K, DesJarlais RL, Ramachandren K, Wang HY et al. 2006. Effect of construct design on MAPKAP kinase-2 activity, thermodynamic stability and ligand-binding affinity. *Arch Biochem Biophys* **449**: 47-56.

Kimple ME, Brill AL, Pasker RL. 2013. Overview of affinity tags for protein purification. *Current protocols in protein science / editorial board, John E Coligan [et al]* **73**: Unit 9 9.

Knubovets T, Osterhout JJ, Connolly PJ, Klibanov AM. 1999. Structure, thermostability, and conformational flexibility of hen egg-white lysozyme dissolved in glycerol. *Proc Natl Acad Sci U S A* **96**: 1262-1267.

Krishna SN, Luan CH, Mishra RK, Xu L, Scheidt KA, Anderson WF, Bergan RC. 2013. A fluorescence-based thermal shift assay identifies inhibitors of mitogen activated protein kinase kinase 4. *Plos One* **8**: e81504.

Kudo N, Kumagai K, Matsubara R, Kobayashi S, Hanada K, Wakatsuki S, Kato R. 2010. Crystal structures of the CERT START domain with inhibitors provide insights into the mechanism of ceramide transfer. *J Mol Biol* **396**: 245-251.

Kudo N, Kumagai K, Tomishige N, Yamaji T, Wakatsuki S, Nishijima M, Hanada K, Kato R. 2008. Structural basis for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide. *Proc Natl Acad Sci U S A* **105**: 488-493.

Lepre CA, Moore JM, Peng JW. 2004. Theory and applications of NMR-based screening in pharmaceutical research. *Chemical Reviews* **104**: 3641-3676.

Letourneau D, Lefebvre A, Lavigne P, LeHoux JG. 2015. The binding site specificity of STARD4 subfamily: Breaking the cholesterol paradigm. *Molecular and Cellular Endocrinology* **408**: 53-61.

Letourneau D, Lorin A, Lefebvre A, Cabana J, Lavigne P, LeHoux JG. 2013. Thermodynamic and solution state NMR characterization of the binding of secondary and conjugated bile acids to STARD5. *Biochimica et Biophysica Acta* **1831**: 1589-1599.

Li XY, Gianoulis TA, Yip KY, Gerstein M, Snyder M. 2010. Extensive *in vivo* Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. *Cell* **143**: 639-650.

Li Y, Geng Y, Song H, Zheng G, Huan L, Qiu B. 2004. Expression of a human lactoferrin N-lobe in Nicotiana benthmiana with potato virus X-based agroinfection. *Biotechnol Lett* **26**: 953-957.

Li YF. 2010. Commonly used tag combinations for tandem affinity purification. *Biotechnol Appl Biochem* **55**: 73-83.

Liu J, Ma P, Sun Y, Yang M, Yang L, Li Y, Wu Y, Zhu X, Wang X. 2007. Expression of human acidic fibroblast growth factor in *Nicotiana benthamiana* with a potato-virus-X-based binary vector. *Biotechnol Appl Biochem* **48**: 143-147.

Lo MC, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, Ellestad G. 2004. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **332**: 153-159.

Matulis D, Kranz JK, Salemme FR, Todd MJ. 2005. Thermodynamic stability of carbonic anhydrase: Measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry-Us* **44**: 5258-5266.

McGovern SL, Caselli E, Grigorieff N, Shoichet BK. 2002. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J Med Chem* **45**: 1712-1722.

McMahon RM, Scanlon MJ, Martin JL. 2013. Interrogating Fragments Using a Protein Thermal Shift Assay. *Aust J Chem* **66**: 1502-1506.

Mezzasalma TM, Kranz JK, Chan W, Struble GT, Schalk-Hihi C, Deckman IC, Springer BA, Todd MJ. 2007. Enhancing recombinant protein quality and yield by protein stability profiling. *J Biomol Screen* **12**: 418-428.

Mouradov A, Spangenberg G. 2014. Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front Plant Sci* **5**.

Mueller LA, Goodman CD, Silady RA, Walbot V. 2000. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiology* **123**: 1561-1570.

Niesen FH, Berglund H, Vedadi M. 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**: 2212-2221.

Olayioye MA. 2005. StarD10, a START Domain Protein Overexpressed in Breast Cancer, Functions as a Phospholipid Transfer Protein. *Journal of Biological Chemistry* **280**: 27436-27442.

Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P et al. 2001. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6**: 429-440.

Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF et al. 2009. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324**: 1068-1071.

Parks DJ, LaFrance LV, Calvo RR, Milkiewicz KL, Gupta V, Lattanze J, Ramachandren K, Carver TE, Petrella EC, Cummings MD et al. 2005. 1,4-benzodiazepine-2,5-diones as small molecule antagonists of the HDM2-p53 interaction: discovery and SAR. *Bioorg Med Chem Lett* **15**: 765-770.

Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, Rajashankar KR, Cohen DE. 2002. Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat Struct Biol* **9**: 507-511.

Rodriguez-Agudo D, Ren S, Hylemon PB, Redford K, Natarajan R, Del Castillo A, Gil G, Pandak WM. 2005. Human StarD5, a cytosolic StAR-related lipid binding protein. The *Journal of Lipid Research* **46**: 1615-1623.

Rodriguez-Agudo D, Ren S, Wong E, Marques D, Redford K, Gil G, Hylemon P, Pandak WM. 2008. Intracellular cholesterol transporter StarD4 binds free cholesterol and increases cholesteryl ester formation. *The Journal of Lipid Research* **49**: 1409-1419.

Romanowski MJ, Soccio RE, Breslow JL, Burley SK. 2002. Crystal structure of the Mus musculus cholesterol-regulated START protein 4 (StarD4) containing a StAR-related lipid transfer domain. *Proc Natl Acad Sci USA* **99**: 6949-6954.

Sack JS, Kish KF, Wang CH, Attar RM, Kiefer SE, An YM, Wu GY, Scheffler JE, Salvati ME, Krystek SR et al. 2001. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc Natl Acad Sci USA* **98**: 4904-4909.

Sahdev S, Khattar SK, Saini KS. 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* **307**: 249-264.

Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR. 2013. The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiology and Biochemistry : PPB / Societe francaise de physiologie vegetale* **72**: 21-34.

Santambrogio C, Grandori R. 2008. Monitoring the Tanford transition in beta-lactoglobulin by 8-anilino-1-naphthalene sulfonate and mass spectrometry. *Rapid Commun Mass Spectrom* **22**: 4049-4054.

Santiago J, Dupeux F, Betz K, Antoni R, Gonzalez-Guzman M, Rodriguez L, Marquez JA, Rodriguez PL. 2012. Structural insights into PYR/PYL/RCAR ABA receptors and PP2Cs. *Plant Sci* **182**: 3-11.

Saslowsky DE, Warek U, Winkel BSJ. 2005. Nuclear localization of flavonoid enzymes in *Arabidopsis*. *Journal of Biological Chemistry* **280**: 23735-23740.

Schrick K, Bruno M, Khosla A, Cox PN, Marlatt SA, Roque RA, Nguyen HC, He C, Snyder MP, Singh D et al. 2014. Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity. *BMC Biology* **12**: 70.

Schrick K, Nguyen D, Karlowski WM, Mayer KFX. 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol* **5**.

Schwarz FP, Puri K, Surolia A. 1991. Thermodynamics of the binding of galactopyranoside derivatives to the basic lectin from winged bean (Psophocarpus tetrogonolobus). *The Journal of Biological Chemistry* **266**: 24344-24350.

Seabrook SA, Newman J. 2013. High-throughput thermal scanning for protein stability: making a good technique more robust. *ACS Combinatorial Science* **15**: 387-392.

Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. 2003. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci* **12**: 1313-1322.

Soon FF, Suino-Powell KM, Li J, Yong EL, Xu HE, Melcher K. 2012. Abscisic Acid Signaling: Thermal Stability Shift Assays as Tool to Analyze Hormone Perception and Signal Transduction. *Plos One* **7**.

Sorrell FJ, Greenwood GK, Birchall K, Chen B. 2010. Development of a differential scanning fluorimetry based high throughput screening assay for the discovery of affinity binders against an anthrax protein. *Journal of Pharmaceutical and Biomedical Analysis* **52**: 802-808.

Tabunoki H, Higurashi S, Ninagi O, Fujii H, Banno Y, Nozaki M, Kitajima M, Miura N, Atsumi S, Tsuchida K et al. 2004. A carotenoid-binding protein (CBP) plays a crucial role in cocoon pigmentation of silkworm (Bombyx mori) larvae. *FEBS Lett* **567**: 175-178.

Tagore DM, Nolte WM, Neveu JM, Rangel R, Guzman-Rojas L, Pasqualini R, Arap W, Lane WS, Saghatelian A. 2009. Peptidase substrates via global peptide profiling. *Nat Chem Biol* **5**: 23-25.

Tagore R, Thomas HR, Homan EA, Munawar A, Saghatelian A. 2008. A global metabolite profiling approach to identify protein-metabolite interactions. *J Am Chem Soc* **130**: 14111-14113.

Thorsell AG, Lee WH, Persson C, Siponen MI, Nilsson M, Busam RD, Kotenyova T, Schuler H, Lehtio L. 2011. Comparative structural analysis of lipid binding START domains. *PLoS One* **6**: e19521.

Tsujishita Y, Hurley JH. 2000. Structure and lipid transport mechanism of a StAR-related domain. *Nat Struct Biol* **7**: 408-414.

Varsani A, Williamson AL, Stewart D, Rybicki EP. 2006. Transient expression of Human papillomavirus type 16 L1 protein in Nicotiana benthamiana using an infectious tobamovirus vector. *Virus Res* **120**: 91-96.

Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ, Wasney GA, Yeung R, Arrowsmith C, Ball LJ, Berglund H et al. 2006. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci USA* **103**: 15835-15840.

Vellosillo T, Martinez M, Lopez MA, Vicente J, Cascon T, Dolan L, Hamberg M, Castresana C. 2007. Oxylipins produced by the 9-lipoxygenase pathway in *Arabidopsis* regulate lateral root development and defense responses through a specific signaling cascade. *Plant Cell* **19**: 831-846.

Vinayavekhin N, Saghatelian A. 2011. Discovery of a protein-metabolite interaction between unsaturated fatty acids and the nuclear receptor Nur77 using a metabolomics approach. *J Am Chem Soc* **133**: 17168-17171.

Vincentelli R, Canaan S, Campanacci V, Valencia C, Maurin D, Frassinetti F, Scappucini-Calvo L, Bourne Y, Cambillau C, Bignon C. 2004. High-throughput automated refolding screening of inclusion bodies. *Protein Sci* **13**: 2782-2792.

Voulgaridou GP, Mantso T, Chlichlia K, Panayiotidis MI, Pappa A. 2013. Efficient E. coli Expression Strategies for Production of Soluble Human Crystallin ALDH3A1. *Plos One* **8**.

Wasson AP, Pellerone FI, Mathesius U. 2006. Silencing the flavonoid pathway in Medicago truncatula inhibits root nodule formation and prevents auxin transport regulation by rhizobia. *Plant Cell* **18**: 1617-1629.

Waugh DS. 2005. Making the most of affinity tags. *Trends Biotechnol* **23**: 316-320. 2011. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification* **80**: 283-293.

Williamson, Actis-Goretta L. 2010. Functional foods for health promotion: state-of-the-science on dietary flavonoids. (vol 67, pg 736, 2009). *Nutr Rev* **68**: 190-190.

Wu XQ, Lukacik P, Kavanagh KL, Oppermann U. 2007. SDR-type human hydroxysteroid dehydrogenases involved in steroid hormone activation. *Molecular and Cellular Endocrinology* **265**: 71-76.

Yin R, Messner B, Faus-Kessler T, Hoffmann T, Schwab W, Hajirezaei MR, von Saint Paul V, Heller W, Schaffner AR. 2012. Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-O-glycosylation. *Journal of Experimental Botany* **63**: 2465-2478.

Young CL, Britton ZT, Robinson AS. 2012. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnology Journal* **7**: 620-634.

Zhan Y, Song X, Zhou GW. 2001. Structural analysis of regulatory protein domains using GST-fusion proteins. *Gene* **281**: 1-9.

Chapter 5 - HD-Zip proteins GL2 and HDG11 have redundant functions in Arabidopsis trichomes, and GL2 activates a positive feedback loop via MYB23

ABSTRACT

The class IV homeodomain leucine zipper transcription factor GLABRA2 (GL2) acts in a complex regulatory circuit that regulates the differentiation of trichomes in Arabidopsis thaliana. We describe a genetic interaction with *HOMEODOMAIN GLABROUS11 (HDG11)*, previously identified as a negative regulator of trichome branching. *gl2 hdg11* double mutants display enhanced trichome cell-type differentiation defects. Transgenic expression of *HDG11* using the *GL2* promoter partially suppresses *gl2* trichome phenotypes. Vice versa, expression of *GL2* under the control of its native promoter partially complements hdg11 ectopic branching. Since *gl2 hdg11* and *gl2 myb23* double mutants and the triple mutant display similar trichome differentiation defects, we investigated a connection to the R2R3-MYB transcription factor MYB23. We show that *MYB23* transcript levels are significantly reduced in shoots from *gl2* mutants and that GL2 can drive the expression of a *MYB23*-promoter fusion to green fluorescent protein. Yeast one-hybrid, chromatin immunoprecipitation, and in planta reporter gene experiments indicate that an L1-box in the *MYB23* promoter acts as a GL2 binding site. Taken together, our findings reveal a functional redundancy between GL2 and HDG11, two homeodomain leucine zipper transcription factors previously thought to mediate opposing functions in trichome morphogenesis. A model is proposed in which *GL2* transcript levels are maintained through a positive feedback loop involving GL2 activation of *MYB23*.

Contribution

DNA constructs and transgenic lines for *GL2* and *HDG11* overexpression experiments and results of the yeast one-hybrid, ChIP, and transient expression in *N. benthamiana* included in this chapter are my contribution to the publication: **Khosla A**, Paper J.M, Boehler AP, Bradley AM, Neumann TR, Schrick K (2014). HD-Zip Proteins GL2 and HDG11 Have Redundant

Functions in Arabidopsis Trichomes, and GL2 Activates a Positive Feedback Loop via MYB23. *Plant Cell* **26**: 2184-2200.

INTRODUCTION

In plants, the epidermis of leaves is derived from the outer L1 layer of the shoot apical meristem. As leaf primordia develop, the epidermal cell layer differentiates to form an organized pattern of distinct cell types, including pavement cells, stomata, and trichomes. In *Arabidopsis thaliana*, mature leaf trichomes are characteristically large (300 to 500 μm) branched hair cells whose nuclei have undergone multiple rounds of endoreduplication and are present on the leaf surface in a nonrandom regular distribution (Hulskamp et al. 1994; Balkunde et al. 2010). Among other plant species, leaf trichomes are highly variable and may be branched or unbranched, single celled or multicellular, and may function as glandular organs. Proposed functions of trichomes include protection from herbivores and shielding from abiotic environmental factors such as frost, evaporation, and UV radiation (Balkunde et al. 2010; Yang and Ye 2013).

Transcription factors of the class IV homeodomain leucine-zipper (HD-Zip) family play central roles in the differentiation of the epidermis during both embryonic and postembryonic development. This protein family is characterized by a homeodomain (HD) N terminal to a plant-specific Leu zipper having an internal loop (zipper-loop-zipper [ZLZ]) (also called truncated leucine zipper (Yang et al. 2002), followed by a steroidogenic acute regulatory protein–related lipid transfer (START) domain (Schrick et al. 2004). The *Arabidopsis* genome encodes a total of 16 class IV HD-Zip family members, several of which are known to be involved in cell fate determination within layer-specific contexts (Nakamura et al. 2006).

The class IV HD-Zip transcription factor GLABRA2 (GL2) plays a pivotal role in regulating the differentiation of the epidermis in numerous tissues, including the root (Di Cristina et al. 1996) and seed coat(Western et al. 2001). Its role in trichomes is the major focus of this article. *gl2* mutants display aborted trichomes with thin cell walls that lack papillae, and some trichomes aberrantly expand in the plane of the pavement cells(Koornneef et al. 1982; Rerie et al. 1994). GL2 is highly expressed in mature trichomes(Marks et al. 2009), and late activity of GL2 seems to be required for proper trichome development. While GL2 is required for the differentiation and maintenance of trichome cell fate, it is dispensable for trichome initiation. In a current model, GL2 acts as a downstream transcription factor in a regulatory circuit together with the mitosis inhibitor SIM that is required for endoreduplication (Grebe 2012). The initiation of trichome cell fate and transcriptional activation of GL2 and SIM involves an activator

complex composed of a WD40 protein encoded by TRANSPARENT TESTA GLABRA1 (TTG1), an R2R3-MYB transcription factor encoded by GL1 and the partially redundant gene MYB23, a basic helix-loop-helix (bHLH) protein encoded by two genes, GL3 and ENHANCER OF GL3 (EGL3), and the possible addition of another component encoded by MYC1 (Grebe, 2012). This complex is counteracted by inhibitory R3 MYB transcription factors encoded by a group of at least six genes, including CAPRICE and TRIPTYCHON, which diffuse between cells, competing with the R2R3-MYB proteins for binding to the bHLH protein.

Based on mutant phenotypes and expression profiling, three other class IV HD-Zip transcription factors in addition to GL2 are implicated in trichome cell-type differentiation. HOMODOMAIN GLABROUS2 (HDG2) is strongly expressed in mature trichomes, similar to the level observed for GL2(Marks et al. 2009). While *hdg2* mutants display no obvious defects in overall trichome cell morphology, their trichome surfaces appear more transparent and their cell wall composition is abnormal(Marks et al. 2009). Based on mutant analysis, HDG11 and HDG12 play negative regulatory roles in trichome branching. *hdg11* mutants result in excess branching of trichomes, a phenotype that is enhanced in combination with *hdg12*(Nakamura et al. 2006). While promoter–β-glucuronidase (GUS) reporter lines for both exhibit strong expression in trichomes(Nakamura et al. 2006), transcript levels of *HDG11* and *HDG12* are low and not detectable, respectively, in mature trichomes(Marks et al. 2009), consistent with expression being mostly restricted to the branching phase of trichome development.

In this article, we uncover a genetic interaction and functional redundancy between the class IV HD-Zip transcription factors GL2 and HDG11 in trichomes. The phenotype of *gl2 hdg11* double mutants led us to investigate a connection to the R2R3-MYB transcription factor MYB23. Members of the expansive R2R3-MYB gene family play key roles in various plant-specific processes, including the regulation of cell fate in development (Stracke et al. 2001; Dubos et al. 2010). MYB23 had previously been found to act redundantly with GL1 in trichome cell-type specification (Kirik et al. 2005). In the root, MYB23 had been shown to function in cell fate determination and to bind its own promoter in a MYB23-mediated positive feedback loop (Kang et al. 2009). Here, we provide evidence for an additional positive feedback mechanism involving GL2 transcriptional activation of MYB23 in the maintenance of cell fate during trichome morphogenesis.

RESULTS

Sequence Analysis of GL2 and HDG11 Reveals Highly Related HDs

GL2 and *HDG11* are linked loci on *Arabidopsis* chromosome 1. Phylogenetic analysis of the 16 members of the class IV HD-Zip family does not reveal a high degree of relatedness between GL2 and HDG11 in comparison with other family members (Nakamura et al., 2006). To further probe sequence similarity between GL2 and HDG11, we divided their protein sequences into four distinct conserved domains: (1) HD, (2) ZLZ, (3) START, and (4) START adjacent domain (SAD) (Figures 5.1 A and B). The N-terminal HD implicated in DNA binding is followed by the ZLZ plant-specific Leu zipper dimerization domain conserved among members of the class IV family (Schrick et al., 2004). The START domain implicated in ligand binding (Ponting and Aravind, 1999; Schrick et al., 2004) lies in the middle, followed by the SAD of unknown function at the C terminus. We determined the percentage amino acid identities and similarities between the four domains of the two proteins in pairwise alignments from protein–protein BLAST (Figure 5.1 B) and found the HD domain to be most conserved, exhibiting 74 and 89% identity and similarity, respectively (Figure 5.1 C). Pairwise alignments were used to compare the HD domain of GL2 with each of the 15 other members of the class IV HD-Zip family (Appendix Table D-1). Ranking by amino acid identity to GL2, HDG11 placed first, followed by HDG5 and PROTODERMAL FACTOR2 (PDF2). Among the other class IV members whose mRNA transcripts are expressed in leaf trichomes (HDG2, HDG11, and HDG12), HDG11 exhibits the highest degree of amino acid identity and similarity to GL2.

Double Mutants for *gl2* and *hdg11* Display Enhanced Trichome Differentiation Defects

Crosses between *gl2-5* and *hdg11-1* plants of the Columbia (Col) ecotype were performed to construct homozygous double mutants. In comparison with *gl2-5* single mutants, the double mutants exhibited enhanced trichome differentiation defects (Figure 5.2). Trichome phenotypes were visualized on rosette leaves by stereomicroscopy (Figures 5.2 A to D). Leaf trichome morphologies were observed under higher magnification using scanning electron microscopy (Figures 5.2 E to T). While trichome initiation appears not to be altered in *gl2-5* single and *gl2-5 hdg11-1* double mutants, outgrowth and differentiation of the trichomes are affected. Trichomes are generally aborted on leaves from *gl2-5* mutants, whereas trichomes at

the leaf margins are more differentiated and typically branched with two (or rarely three) branches. Quantification of trichome outgrowth revealed a significant difference in the number of branched trichomes at the leaf margins in *gl2-5* mutants versus *gl2-5 hdg11-1* double mutants (Figure 5.2 U; Appendix Table D-2). We observed ectopic branching in *hdg11-1* mutants, as reported previously (Nakamura et al. 2006). On first leaves, trichomes with five branches or more were very rare in wild-type Col (~0.4%), while *hdg11-1* mutants exhibited ~13% of the total trichomes with five or more branches. In contrast, in *gl2-5 hdg11-1* double mutants, there was a complete absence of branched trichomes, while *gl2-5* mutants displayed branched trichomes among ~11% of the total trichomes on first leaves. Trichomes also form on the epidermis of sepals from *Arabidopsis* flowers. Analogous deviations in trichome morphogenesis were detected on sepals of *hdg11-1* and *gl2-5* mutants (Appendix Figure D-1). Wild-type sepals exhibited unbranched trichomes, while *hdg11-1* sepals showed both unbranched and branched trichomes. *gl2-5* mutants exhibited fewer and smaller trichomes, while *gl2-5 hdg11-1* mutants typically lacked trichomes on sepals.

Using the Landsberg erecta (Ler) ecotype as the genetic background, we crossed *gl2-1* and *hdg11-3* mutants to produce the corresponding double mutant plants (Appendix Figure D-2). The quantitative difference between trichome branching on first leaves for *gl2-1* single mutants versus *gl2-1 hdg11-3* double mutants was more striking in the Ler genetic background as compared with the Col ecotype (Appendix Figure D-2 M and Appendix Table D-2). We also constructed *gl2-5 hdg11-3* double mutants and found their trichome cell-type differentiation defects enhanced in comparison with *gl2-5* siblings (Appendix Figure D-3).

gl2 hdg11 Trichome Defects Resemble Those of *gl2 myb23* Mutants

Previously it was shown that the trichome differentiation defect of *gl2* mutants is enhanced in a double mutant for the Col alleles *gl2-4AA* and *myb23-2* (Kirik et al. 2005). It was reported that *myb23* mutants exhibit reduced trichome branching but no obvious defect in trichome patterning (Kirik et al. 2005) and display smaller trichomes (Li et al. 2009). We confirmed the previously reported trichome defects using the *myb23-3* mutant allele (Figure 5.3). The *gl2-5 myb23-3* double mutant was constructed, and enhanced trichome defects were confirmed for this new allelic combination in the Col ecotype. Similar to *gl2 hdg11* mutants, *gl2-5 myb23-3* mutants exhibited trichomes that were distinctly smaller and less expanded than *gl2-5*

trichomes, and this was especially apparent at the leaf margins (Figures 5.3 A to D). Using scanning electron microscopy to view trichome morphologies in detail, we observed trichome differentiation defects in *gl2-5 myb23-3* mutants that were analogous to those seen in *gl2 hdg11* mutants (Figures 5.2 and 5.3 E to L). Quantification of the trichome defects on first leaves indicated that while ~11% of the total trichomes in *gl2-5* mutants were branched, none of the trichomes in the *gl2-5 myb23-3* mutants were branched (Figure 5.3 M; Appendix Table D-2).

Construction of the *gl2 hdg11 myb23* Triple Mutant Reveals Genetic Interactions

Given the similarity between the leaf trichome phenotypes of *gl2 hdg11* and *gl2 myb23* double mutants, we investigated genetic interactions between *hdg11* and *myb23* in the *gl2* background. If the *hdg11* and *myb23* mutations affect the same biological pathway in the absence of *GL2*, then the triple mutant *gl2 hdg11 myb23* would be expected to display similar trichome differentiation defects to either double mutant. Indeed, we found that the *gl2-5 hdg11-1 myb23-3* triple mutant displays a trichome phenotype resembling the double mutants (Figures 5.3 N to P). Quantification of trichomes on first leaves indicated that the triple mutant has a phenotype that is indistinguishable from *gl2-5 myb23-3* (Figure 5.3 Q), and examination of first leaves by scanning electron microscopy shows striking similarity between the double and triple mutants (Appendix Figure D-4).

During construction of the triple mutant, we observed trichome defects in the F1 progeny from reciprocal crosses between *gl2 hdg11* and *gl2 myb23*, whose genotype was homozygous for *gl2-5* and heterozygous for both *hdg11-1* and *myb23-3* (Appendix Figure D-5). Unlike *gl2* mutants that exhibit trichomes on leaf margins, the *gl2/gl2 hdg11/+ myb23/+* plants exhibited enhanced trichome defects similar to those of *gl2 hdg11* and *gl2 myb23* double mutants (Appendix Figures D-5 A to F). Unexpectedly, we also uncovered fertility defects that were associated with inviable pollen from plants heterozygous for *hdg11-1* or *myb23-3* or both. However, these phenotypes were found to be independent of *GL2* (Appendix Figures D-5 G to C). Double mutants for *hdg11* and *myb23* were found to exhibit an intermediate trichome phenotype, lacking the ectopic branching of *hdg11* mutants, while showing significantly fewer two-branched trichomes than *myb23* mutants (Appendix Figure D-6 and Appendix Table D-2). This suggests that in the presence of *GL2, HDG11* and *MYB23* act in parallel to orchestrate trichome cell differentiation, and *MYB23* is not required for this inhibitory function of *HDG11*.

Interactions between GL2, HDG11, and MYB23 Are Tissue Specific

After establishing the enhanced trichome cell differentiation defects in *gl2 hdg11* double mutants, we investigated phenotypes in roots and seeds, other tissues that are affected in *gl2* single mutants. In the root epidermis, *gl2* mutants exhibit an increase in the proportion of cells that are specified for root hair fate (Di Cristina et al. 1996). We observed excess root hair formation on primary roots of *gl2 hdg11* double mutants from both the Col and Ler backgrounds but did not detect an enhanced defect as compared with *gl2* single mutants (Figures 5.4 A to D and 5.4 N; Appendix Figure D-7 and Appendix Table D-3). *gl2* mutants also exhibit a defect in the production of seed coat mucilage upon imbibition (Western et al. 2000). Using ruthenium red staining for pectin in the seed coat mucilage, we observed that *gl2 hdg11* double mutants lack the mucilage layer in a manner that is indistinguishable from *gl2* single mutants (Figures 5.4 A′ to D′; Appendix Figure D-7). In addition to defects in the epidermis, *gl2* mutants accumulate higher levels of oil in the seed (Shen et al. 2006), and this phenotype is associated with the lack of seed coat mucilage (Shi et al. 2012). We also found that *gl2 hdg11* double mutants display a similar increase in seed oil levels to *gl2* single mutants (Appendix Figure D-8). Similarly, examination of *gl2 myb23* mutants for root hair and seed mucilage phenotypes revealed no obvious enhanced or novel phenotypes relative to the *gl2* single mutants, consistent with the idea that the genetic interaction with *MYB23* is also specific to trichomes (Figure 5.4; Appendix Table D-3).

Expression of *HDG11* under the Control of the *GL2* Promoter Rescues Ectopic Branching of *hdg11* Mutants and Partially Suppresses Trichome Differentiation Defects of *gl2* Mutants

We used the 2.1-kb promoter upstream of *GL2* to express *HDG11* in trichomes from *hdg11, gl2*, and *gl2 hdg11* mutants (Figure 5.5). An N-terminal enhanced yellow fluorescence protein (EYFP) tag was used to monitor the expression of *HDG11* in transgenic lines. As a control, we utilized an *hdg11* mutant construct that is missing the first 305 residues of the predicted protein and therefore lacking the HD and ZLZ domains as well as part of the START domain (Figure 5.5 A). The trichome phenotypes of representative *hdg11, gl2*, and *gl2 hdg11* lines, as visualized by scanning electron microscopy, were unchanged, indicating that this construct is null for *HDG11* activity (Figures 5.5 B to G). By contrast, scanning electron

microscopy analysis showed that the wild-type *ProGL2:EYFP:HDG11* construct fully complements the trichome-branching defect of *hdg11-1* mutants (Figures 5.5 H and K). Its expression was also sufficient to induce increased branching of trichomes at leaf margins in the *gl2-5 hdg11-1* double mutant, consistent with complementation of the *hdg11* mutant phenotype (Figures 5.5 L and M; Appendix Figures D-9 B, D, and C-F). However, the *ProGL2:EYFP:HDG11* construct only partially rescued the trichome differentiation defect of *gl2-5* mutants on first leaves (Figures 5.5 I and J) and leaves from rosettes (Appendix Figures D-9 A, C, and E). Unexpectedly, novel transient dwarf phenotypes were observed at the rosette stage for both *gl2-5* and *gl2-5 hdg11-1* mutants transformed with *ProGL2:EYFP:HDG11* (Appendix Figures D-9 G to L).

In all three genetic backgrounds, expression of the *ProGL2:EYFP:HDG11* fusion protein was monitored in trichomes and found to be localized in nuclei of trichome cells (Figures 5.5 N to S). Quantification of trichomes on first leaves revealed a significant difference in the branching and/or number of trichomes for mutant lines expressing *ProGL2:EYFP:HDG11* (Figure 5.5 T; Appendix Table D-2). For *hdg11* mutant plants expressing the wild-type *HDG11* transgene, ~0.6% of trichomes had five or more branches in comparison with the control, in which ~17% of the trichomes exhibited five or more branches. Thus, expression of *HDG11* under the control of the *GL2* promoter is sufficient for rescue of the *hdg11* branching phenotype. In *gl2* mutant plants expressing the wild-type *EYFP:HDG11* transgene, an ~3.3-fold increase in average trichome numbers and more branching (~48 versus ~27% in the control) on first leaves were observed, but overall trichome formation was partial in comparison with the wild type. For the *gl2-5 hdg11-1* double mutant, branching and trichome cell-type differentiation were restored to approximately the levels in *gl2-5* mutants, as indicated by an ~35-fold increase in visible trichomes and more branching (~44 versus 0% in the control).

We also expressed *HDG11* in *gl2-5* mutants using the epidermis-specific ~3.4-kb *ATML1* promoter from class IV HD-Zip transcription factor ATML1 (Appendix Figure D-10). *ProATML1:HDG11* transgenic lines exhibited increased trichome cell-type differentiation in comparison with untransformed *gl2-5* control plants. However, the trichome cell-type differentiation was partial, resulting in an increase in mainly one- to two-branched and rarely three-branched trichomes (Appendix Figures D-10 D to G). Additional phenotypes, such as leaf

curling and novel pointy leaf shapes, were also observed in conjunction with trichome formation, suggesting that the ectopic expression of *HDG11* either induces or interferes with other cell-type differentiation pathways in the epidermis.

Expression of *HDG11* under the Control of the *GL2* Promoter Partially Rescues the Root Hair Phenotype of *gl2* Mutants but Negatively Affects Seed Mucilage Production

In roots, we detected a partial rescue of the excess root hair phenotype of *gl2* mutants in transgenic lines expressing *ProGL2:EYFP:HDG11* in comparison with the *ProGL2:EYFP:hdg11Δ* negative control (Figures 5.4 H, L, and O; Appendix Table D-3). A complete rescue of the *gl2* root hair phenotype was conferred by the *ProGL2:EYFP:GL2* transgene (Figures 5.4 J and O). This construct, in which EYFP-tagged GL2 is expressed under the control of its native promoter (Figure 5.6 A), was shown previously to completely rescue the trichome defects of *gl2-5* mutants (K. Schrick, B.P. Srinivas, and M. Hülskamp, unpublished data). The expression of mutant or wild-type *ProGL2:EYFP:HDG11* did not alter the root hair phenotype of *hdg11* mutants, which appears indistinguishable from that of the wild type (Figures 5.4 G, K, and O). However, in seeds, the expression of *ProGL2:EYFP:HDG11* in *hdg11* had a negative effect on mucilage formation in comparison with the control, resulting in a "*gl2*" phenotype (Figures 5.4 G′ and K′). Although *ProGL2:EYFP:GL2* was able to complement the seed mucilage phenotype of *gl2* (Figure 5.4 J′), there was no visible rescue of the seed coat mucilage phenotype in *gl2* mutants by the *ProGL2:EYFP:HDG11* transgene (Figures 5.4 H′, I′, L′, and M′). Taken together, these data indicate that the ability of *HDG11* to partially rescue the gl2 phenotype and/or interfere with *GL2* activity is tissue specific.

GL2 Overexpression Partially Rescues Ectopic Trichome Branching of *hdg11* Mutants

Since *HDG11* overexpression in trichomes can partially rescue the trichome and root hair phenotypes of *gl2*, we asked whether *GL2* overexpression is able to complement the ectopic branching phenotype of *hdg11* mutants. We expressed *ProGL2:EYFP:GL2 in hdg11* mutants and observed partial complementation of the ectopic branching phenotype (Figure 5.6). Scanning electron microscopy analysis showed that in comparison with the control lacking the transgene (Figures 5.6 B and D), *hdg11* mutants expressing *ProGL2:EYFP:GL2* exhibited fewer trichomes

having five or more branches. Expression of the transgene was associated with nuclear localization of the EYFP:GL2 protein in trichomes (Figures 5.6 E and F). Quantification of trichomes on first leaves revealed a significant difference in *hdg11* mutants expressing the transgene (~20% ectopically branched trichomes) versus the controls lacking the transgene (~3% ectopically branched trichomes) (Figure 5.6 H; Appendix Table D-2). Expression of the *GL2* transgene also led to an increase in the average total number of trichomes on first leaves of *hdg11* mutants (~75 versus ~45 in the control), suggesting that *GL2* expression can drive trichome cell-type specification in the absence of *HDG11*.

GL2 Activates *MYB23* Transcription in Trichomes by Binding an L1-Box in the *MYB23* Promoter

To further dissect the genetic interactions between *GL2, HDG11*, and *MYB23*, we monitored their respective mRNA transcripts in seedlings from the Col wild type, *gl2-5, hyd11-1,* and *myb23-3* as well as the *gl2 hdg11* and *gl2 myb23* double mutants. mRNA was isolated from 14-d-old seedling shoots to enrich for transcript levels in trichomes. Quantitative real-time PCR confirmed that the *gl2-5* and *hdg11-1* alleles are RNA knockdown alleles, while *myb23-3* mutants do not exhibit a reduced level of mRNA expression in shoots (Figures 5.7 A to C). It was shown previously that for root tips from *myb23-3*, transcript levels in *myb23-3* are reduced to ~58% of wild-type levels (Kang et al. 2009). The *myb23-3* root hair phenotype was reported to be indistinguishable from that of the transcript-null allele *myb23-1*, indicating that although a transcript is expressed in myb23-3, it is not functional (Kang et al. 2009). HDG11 mRNA levels were decreased only in *hdg11-1* and *gl2-5 hdg11-1* (Figure 5.7 A). By contrast, our analysis revealed that *MYB23* mRNA levels were significantly reduced in *gl2-5* single mutants (~5-fold) as well as in *gl2-5 hdg11-1* and *gl2-5 myb23-3* double mutants (Figure 5.7 B).

The real-time PCR data are consistent with the possibility that *MYB23* is a direct transcriptional target of GL2. Previously, it was shown that a *ProMYB23:MYB23* construct containing the *MYB23* promoter (3.1 kb N terminal to the start codon and 1.0 kb C terminal to the stop codon) was sufficient to rescue the trichome defect in *myb23-1* mutants (Kirik et al. 2005). The 3.1-kb promoter of *MYB23* contains two putative L1-boxes (Figure 5.7 D) that are heptamer/octamer elements shown to be required for DNA binding of GL2 (Ohashi et al. 2003;

Tominaga-Wada et al. 2009). The first, L1-box I, located ~2.4 kb upstream of the start codon of *MYB23*, is a perfect match, in reverse orientation, to the L1-box (TAAATGTA) described for the GL2-regulated *CELLULOSE SYNTHASE5 (CESA5)* and *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE17 (XTH17)* promoters (Tominaga-Wada et al. 2009). The second, L1-box II (TAAATGTG), located 76 bp downstream from L1-box I, has a one-base mismatch (G) adjacent to the conserved heptamer sequence.

We used a yeast one-hybrid assay to test whether GL2 is able to bind to the L1-box motifs found in the *MYB23* promoter (Figures 5.7 E and F). The results indicate that GL2 binds to the wild-type L1-box I sequence but not a mutated version thereof. In contrast, GL2 was not able to bind to the imperfect sequence represented by L1-box II. We also examined whether HDG11 is able to bind to the motifs and failed to detect binding activity to either L1-box. The experiments indicate that GL2 binds one of the L1-boxes in the *MYB23* promoter, whereas HDG11 is unable to bind that same sequence in the yeast system.

The L1-box motifs in the *MYB23* promoter were compared with motifs identified for other known direct targets of GL2 (*CESA5, PHOSPHOLIPASE D ζ1,* and *XTH17*) (Figure 5.8 A) (Ohashi et al. 2003; Tominaga-Wada et al. 2009). To further test the idea that GL2 binds to an L1-box in the *MYB23* promoter, we performed chromatin immunoprecipitation (ChIP) on seedlings from a *ProGL2:EYFP:GL2* transgenic line (Figure 5.8 B). Using an anti–green fluorescent protein (GFP) antibody to purify protein-DNA complexes containing *EYFP:GL2*, we isolated ChIP DNA and performed quantitative real-time PCR. The results indicate that DNA containing the L1-box region of the *MYB23* promoter is enriched to a similar level as a positive control, DNA corresponding to an L1-box–containing region from the *CESA5* promoter. The percentage input values are significantly increased relative to the *ACT7* negative control, indicating that GL2 binds to both the *MYB23* and *CESA5* promoters in seedlings.

GL2 Drives a *MYB23*-Promoter Fusion, and L1-Box I Is Required for Maximal Activation

To test whether L1-box I is required for transcriptional activation of *MYB23* by GL2, we transiently expressed a reporter composed of the 3.1-kb promoter upstream of *MYB23* fused to GFP in *Nicotiana benthamiana*. Low levels of GFP fluorescence were observed in controls that expressed the reporter alone (Figures 5.8 C and I). A marked increase in GFP fluorescence was observed in epidermal cells that coexpressed GL2 (Figures 5.8 D and I). However, no activation

above the basal level was detected for cells that coexpressed HDG11 (Figures 5.8 E, H, and I). An ~2-fold decrease in activation by GL2 was observed for a mutant reporter that lacked L1-box I (Figures 5.8 G and I), indicating that the L1-box is required for maximal activation.

A naturally occurring mutation is found within L1-box I of the *Arabidopsis* accession *Guckingen-0 (Gu-0)* (Appendix Figure D-11). Hierarchical clustering based on pairwise sequence differences among 96 accessions indicated that Col is the most closely related accession to *Gu-0* (Nordborg et al. 2005). Strikingly, we observed significantly fewer four-branched trichomes on first leaves from *Gu-0* in comparison with Col (Appendix Figures D-11 B to C-F), consistent with compromised *MYB23* function. Real-time PCR experiments indicated that *Gu-0* shoots express lower levels of *MYB23* transcripts (Appendix Figures D-11 G to C-I), consistent with a function for L1-box I in the full activation of *MYB23* by GL2.

DISCUSSION

Class IV HD-Zip Transcription Factors Drive Trichome Cell-Type Differentiation in Plants

In this work, we uncovered a functional redundancy between *GL2* and *HDG11* in trichome cell-type differentiation. Both genes are members of the class IV HD-Zip transcription factor family from *Arabidopsis*, and although the full-length sequences are not closely related, we found that their 59-amino acid HD domains display a high degree of amino acid identity (Figure 5.1; Appendix Table D-1). Class IV HD-Zip transcription factors are highly conserved among the land plants (Floyd et al. 2006; Prigge and Clark 2006). The role of these transcription factors in the differentiation of epidermis in relation to trichome formation seems to be maintained in both dicot and monocot angiosperms. Variation in the form and function of leaf trichomes may be attributed to evolutionary adaptation and likely occurred by functional diversification of ancestral members of the class IV HD-Zip family. HOMEOBOX1, a family member from Gossypium arboreum, appears to be a functional homolog of GL2 in trichome development (Guan et al. 2008). In tomato (Solanum lycopersicum), a dominant allele of a PDF2-related gene, Woolly, confers a hairy trichome phenotype (Yang et al. 2011). In an opposing fashion, mutations in maize (Zea mays) OUTER CELL LAYER4 and corresponding RNA interference plants indicate that its function is necessary for the inhibition of trichome formation on leaf blades (Vernoud et al. 2009).

Overexpression of the *Arabidopsis* class IV HD-Zip member ATML1 is sufficient to induce epidermal identity, including the formation of trichomes and stomata in internal cell layers (Takada et al. 2013), providing evidence for its role as a master regulator of shoot epidermis cell fate. Intriguingly, overexpression of the related family member HDG2 is also able to trigger the development of stomata in nonepidermal tissues, and ATML1 and HDG2 bind the same DNA elements to activate transcription in yeast (Peterson et al. 2013). Here, we show that overexpression of HDG11 in the epidermis by either the *GL2* or *ATML1* promoters induces trichome cell fate in *gl2* mutants in a similar manner to GL2 (Figure 5; Appendix Figure D-10). Thus, our study provides additional evidence for combinatorial, overlapping, albeit distinct, functions for members of the class IV HD-Zip transcription factor family.

GL2-Independent Trichome Cell-Type Differentiation Mediated by HDG11 and the R2R3-MYB Transcription Factor MYB23

GL2 was previously identified to play a pivotal role in trichome cell-type differentiation(Rerie et al. 1994). However, *gl2* mutants are not completely devoid of trichomes, and they express trichomes that are partially differentiated and branched at leaf margins, suggesting *GL2*-independent mechanisms for trichome development. Here, we show that one *GL2*-independent pathway requires *HDG11*, which encodes a class IV HD-Zip transcription factor first identified as a negative regulator of trichome branching (Nakamura et al. 2006). Since GL2 appears to act upstream of HDG11 in trichome cell-type differentiation, one expectation would be that *gl2* is epistatic to *hdg11* in the double mutant. In this view, it is surprising that *gl2 hdg11* double mutants exhibit enhanced trichome differentiation defects (Figure 5.2; Appendix Figures D-2 and D-3).

The ability of the *ProGL2:EYFP:HDG11* construct to rescue trichome defects in the *gl2 hdg11* double mutant and partially rescue the *gl2* single mutants demonstrates that *HDG11* can function as a positive regulator of trichome differentiation when *GL2* is absent. *HDG11's* role as a negative regulator appears to be directly or indirectly dependent on wild-type *GL2* activity. The opposite functions of *HDG11* may be explained in a model in which cis-regulatory elements specific for *GL2* may be bound by HDG11 when *GL2* is absent (Figure 5.8 J). However, the affinity of HDG11 for *GL2* binding sites is likely to be significantly lower, since HDG11 did not bind the *MYB23* L1-box I sequence in the yeast one-hybrid assay nor did it activate the *MYB23* promoter–GFP fusion in *N. benthamiana*. It is intriguing that overexpression of *GL2* under the control of its native promoter can rescue the *hdg11* branching phenotype (Figure 5.6). An analogous explanation is that in *hdg11* mutants, GL2 has affinity for *HDG11* binding sites. Functional redundancy at the level of DNA binding is consistent with the high degree of similarity in the HD domains from GL2 and HDG11 (Figure 5.1).

MYB23 Is a Direct Target of GL2 as Part of a Positive Feedback Loop to Maintain Trichome Cell Fate

In *gl2* shoots, transcript levels of *MYB23* are significantly decreased (Figure 5.7 B), indicating that GL2 is required for full transcription of *MYB23*. Indeed, GL2 drives the expression of a *MYB23* promoter fusion to GFP in the *N. benthamiana* leaf epidermis (Figure 5.8 D). The additional data presented here, including ChIP and yeast one-hybrid experiments, strongly suggest that GL2 binds to an L1-box in the *MYB23* promoter (Figures 5.7 and 5.8). A natural mutation in this L1-box that is associated with reduced trichome branching and lower levels of *MYB23* transcripts in shoots (Appendix Figure D-11) further corroborates this conclusion. The GFP reporter experiments indicate that although this L1-box is required for full transcriptional activation of *MYB23*, other factors are involved. One possibility is that GL2 binds other cis-regulatory elements in the 3.1-kb *MYB23* promoter. Alternatively, GL2 may activate transcription via physical interaction with another DNA binding protein. Taken together, our findings are consistent with a previous study in which it was shown that overexpression of MYB23 in trichomes using the *GL2* promoter partially rescues *gl2* trichome defects (Kirik et al. 2005). Contrary to the conventional model, in which GL2 is thought to act downstream of the transcriptional coregulators GL1 and MYB23, these data demonstrate that MYB23 also acts downstream of GL2.

A model for the maintenance of trichome cell fate through the action of GL2 and MYB23 is proposed in Figure 8J. The activator complex is composed of the R2R3-MYB transcription factor GL1, the bHLH proteins GL3/EGL3, and the WD40 protein TTG. Transcriptional activation of GL2 occurs through concurrent binding of GL1 and GL3 (or the redundant protein EGL3) to the GL2 promoter (Wang and Chen 2008). In turn, GL2 activates MYB23 transcription by binding to the MYB23 promoter. Consistent with GL2 regulation of the MYB23 transcript, the activity of the MYB23 promoter–GUS fusion persists in mature trichomes and is similar to the expression pattern seen for GL2 (Kirik et al. 2001). Promoter and protein-coding region swap experiments corroborate that MYB23 plays the dominant role in trichome branching, although GL1 and MYB23 appear functionally redundant with respect to trichome initiation (Kirik et al. 2005). In the root epidermis, it was shown that *MYB23* binds to its own promoter (Kang et al. 2009). In the model presented here (Figure 5.8 J), MYB23 participates in the activator complex

and binds to the GL2 promoter as well as to the promoter of its own encoding gene as part of a positive feedback loop to maintain trichome cell fate. In *gl2* mutants in which GL2 is absent, DNA binding sites specific for GL2 may be bound by HDG11. Alternatively, HDG11 may promote MYB23 activity directly or indirectly by another mechanism. In this model, the phenotypes for *gl2 hdg11* and *gl2 myb23* double mutants are similar because in both types of mutants, MYB23 is not fully transcriptionally activated.

Tissue-Specific Relationships between *GL2, HDG11*, and *MYB23*

Our results point to a tissue-specific regulation of *MYB23* transcript by GL2 in trichomes. The enhanced phenotypes of the *gl2 hdg11* and *gl2 myb23* double mutants relative to the *gl2* single mutants were observed in trichomes but were not apparent in seeds or roots (Figure 5.4; Appendix Figure D-4). *gl2* mutants lack mucilage formation altogether, so it is difficult to envision an enhanced phenotype in the seed. Whereas *MYB23* together with *MYB5* are implicated in mucilage synthesis and seed coat development (Li et al., 2009), *HDG11* expression in seeds is very low. We analyzed the expression pattern of *HDG11* versus *GL2* using the publicly available transcriptome database Bio-Analytic Resource Arabidopsis eFP browser. Differences between *GL2* and *HDG11* expression are striking in developing seeds. While *GL2* expression is strong, especially in mucilage secretory cells that constitute the seed coat, *HDG11* expression is relatively low or absent in developing seeds (Appendix Figure D-12), consistent with our observation that *hdg11* mutants are not defective in mucilage production. Intriguingly, expression of the *ProGL2:EYFP:HDG11* transgene in *hdg11* mutants resulted in a dominant negative inhibition of mucilage formation (Figure 5.4), analogous to its inhibitory role in limiting trichome branching in the shoot epidermis.

In roots, we observed that expression of the *ProGL2:EYFP:HDG11* transgene partially rescued the excess root hair phenotype of *gl2* (Figure 5.4). The differential behavior of *HDG11* in seeds versus roots is likely due to tissue-specific interactions with coexpressed genes. The analysis o*f myb23* mutants has shown that *MYB23* is necessary for precise establishment of the root epidermal pattern (Kang et al. 2009). It is noteworthy that transcript levels of *MYB23* appear not to be regulated by *GL2* in the root. It was previously shown that in roots from *gl2* mutants, *MYB23* mRNA levels are unchanged relative to those of the wild type (Kang et al. 2009). A

genome-wide transcriptome profiling of the *Arabidopsis* root epidermis also failed to show that *MYB23* transcript levels are changed in *gl2* root tissues (Bruex et al. 2012).

Several observations suggest crosstalk between class IV HD-Zip transcription factors in regulatory and developmental pathways. In transgenic lines that expressed *HDG11* under the control of the *GL2* promoter, we observed a novel transient dwarf phenotype (Appendix Figure D-8). Moreover, when *HDG11* was expressed under the control of the epidermis-specific *ATML1* promoter in *gl2* mutants, several of the transgenic lines displayed leaf curling and/or pointy leaves. In relation to these observations, the WW domain protein CURLY FLAG LEAF1, corresponding to a curly leaf mutant phenotype in rice (Oryza sativa), has been shown to physically interact with a class IV HD-Zip family member, HDG1 (Wu et al., 2011). Other striking observations included fertility defects related to pollen viability that were seen in plants heterozygous for either *hdg11* or *myb23* or both (Appendix Figure D-5), indicating that genetic interactions between these transcription factors are highly sensitive to dosage effects. Relevant to these findings, it was shown recently that a special mutant allele of *PDF2 (pdf2-1)* in combination with loss-of-function mutations in other family members (HDG1, HDG2, HDG5, or HDG12) result in developmental defects in floral organs (Kamata et al. 2013). Future studies will unravel the molecular basis for these and other interactions involving transcription factors of the class IV HD-Zip family in relation to regulatory circuits underlying the differentiation of specified cell types in plants.

MATERIALS AND METHODS

Plant Material and Growth Conditions

The wild-type *Arabidopsis thaliana* ecotypes were Col and Ler. The *hdg11-1* (Col) (SAIL_865_G09) (Nakamura et al. 2006) and *hdg11-3* (Ler) alleles were described previously (Roeder et al. 2012). The *myb23-3* (Col) (SALK_018613) allele contains a T-DNA insertion in the 5′ promoter (Kang et al. 2009). *gl2-1* (Ler) is a fast-neutron–induced allele (Koornneef et al. 1982; Di Cristina et al. 1996). gl2-5 (Col) is a En-1 insertion allele of *GL2* (Ohashi et al. 2003). The genomic region was sequenced using En205 and En8130 primers (Baumann et al. 1998) and GL2-specific primers (Wang et al. 2007), revealing an insertion following the Ser-encoded triplet AGC (bases 121 to 123), after the first 37 amino acids of the predicted protein. A TGA stop codon is located 29 amino acids downstream, within the En-1 transposon, indicating that the mutant allele results in a truncated protein of 63 amino acids, of which the first 37 amino acids are derived from the N terminus of GL2. En-1 lies at the beginning (33 bp) of the second exon of the gene, upstream of the HD (amino acids 106 to 157). Appendix Table D-4 lists oligonucleotides used for genotyping. Plants were stratified at 4°C for 3 to 5 d and grown on soil containing Metro-Mix 380, vermiculite, and perlite (Hummert International) in a 7:3:2 ratio at 23°C under continuous light. *Agrobacterium tumefaciens* (GV3101 pMP90)–mediated transformation of plants was performed using the floral dip method as described previously (Clough and Bent 1998).

DNA constructs

Construction of *ProGL2:EYFP:GL2* (SR54, a derivative of the binary vector pCAMBIA1300) will be described elsewhere (A. Khosla, B.K.P. Venkata, P.N. Cox, D.F. Stucky, P.M. Stephens, S.A. Marlatt, B.P. Srinivas, M. Hülskamp, and K. Schrick, unpublished data). An *HDG11* cDNA clone was obtained from the RIKEN *Arabidopsis* full-length clone collection. Internal SalI and KpnI restriction sites interfering with the cloning scheme were removed prior to amplification by site-directed mutagenesis using a one-step PCR-based method as described previously (Scott et al. 2002). To construct *ProGL2:EYFP:HDG11*, the cDNA sequences for wild-type *HDG11* and *hdg11Δ305* were PCR amplified with Phusion High-

Fidelity DNA Polymerase (Thermo Scientific) using primers engineered to contain SalI and KpnI restriction sites at the 5′ and 3′ ends, respectively. *hdg11Δ305* (referred to as hdg11Δ in Figures 5.4 and 5.5) is a deletion mutant missing the N-terminal 305 residues of the predicted protein, resulting in an in-frame 417 amino acid protein that lacks the first 78 residues of the START domain. The amplified PCR fragment was restriction digested with SalI/KpnI and subcloned into the corresponding restriction sites in binary vector SR54 to make *ProGL2:EYFP:HDG11* and *ProGL2:EYFP:hdg11Δ305*. The constructs were transformed into the gl2-5 and *hdg11-1* single mutant and *gl2-5 hdg11-1* double mutant plants. For *ProGL2:EYFP:HDG11*, Mendelian segregation of EYFP fluorescence was seen for the T2 progeny from 6 of 19 *hdg11-1*, 29 of 43 *gl2-5*, and 28 of 34 *gl2-5 hdg11-1* plants. Strong EYFP expression was observed in trichomes from 3 *hdg11-1,* 16 *gl2-5*, and 9 *gl2-5 hdg11-1* lines carrying *ProGL2:EYFP:HDG11*. For the *ProGL2:EYFP:hdg11Δ* construct, none of the transgenic lines from a total of 75 (20 in *hdg11*, 15 in *gl2*, and 40 in *gl2 hdg11*) independent transformants exhibited EYFP expression. The *ProATML1:HDG11* construct was made using pAR176 (Roeder et al. 2010). The *HDG11* cDNA was cloned into pENTR-D-TOPO (Life Technologies) and moved into pAR176 by Gateway LR Clonase II (Life Technologies). For *ProATML1:HDG11*, ~150 T1 plants survived BASTA selection, 24 independent trichome-bearing transformants were characterized, and 8 of 24 T1 progeny segregated in a Mendelian fashion. Oligonucleotides used for plasmid construction, site-directed mutagenesis, and sequence verification are listed in Appendix Table D-4.

RNA Extraction, cDNA Synthesis, and Quantitative Real-Time PCR

Arabidopsis seeds were surface-sterilized and sown onto 0.8% (plant tissue culture grade) agar containing 1× Murashige and Skoog medium (Murashige and Skoog 1962), 1% Suc, and 0.5% MES buffer at pH 5.8. Seeds were stratified for 3 to 5 d at 4°C and then transferred to 23°C and continuous light for 14 d. After removal of roots and hypocotyl, shoots were flash frozen in liquid nitrogen and stored at −80°C. Approximately 50 mg of each sample was used for RNA extraction using the RNeasy Plant Mini Kit and on-column RNase-Free DNase Set (Qiagen). Furthermore, 0.5 to 2 μg of RNA was used as a template for cDNA synthesis. cDNA synthesis was performed using SuperScript II Reverse Transcriptase (Life Technologies) or GoScript Reverse Transcriptase (Promega) according to the manufacturer's protocol. Real-time PCR was

performed using IQ SYBR Green Supermix with the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Each reaction contained 10 µl of 2× SYBR Green Supermix, 1 µl of forward and reverse 10 mM gene-specific primers, and 5 µl of cDNA (diluted 5-fold) in 20 µl. Standard curves were generated from 10-fold dilutions of amplicons for each primer pair. *ACT7* served as the reference gene. The data represent three biological samples of three to four seedling shoots per sample with three technical replicates of each biological replicate. Primers used for PCR are listed in the Appendix Table D-4.

Phenotypic Analysis and Microscopy

Arabidopsis lines were confirmed by PCR and assayed for leaf trichomes, seed coat mucilage, and root hairs. For root hair density analysis, seedlings were grown vertically on 0.8% agar/water medium under continuous light. Primary roots were viewed 3 to 4 d after germination. For quantification of root epidermal cells, seedlings were stained in a 0.05% aqueous solution of toluidine blue, destained in water, mounted in 10% glycerol, and examined for the numbers of trichoblasts and atrichoblasts in the corresponding respective cell files. For seed mucilage analysis, seeds were stained with a 0.02% aqueous solution of ruthenium red (Sigma) for 2 h followed by replacement with deionized water prior to observation. Seeds, rosettes, and roots were imaged with a Leica M125 fluorescence stereomicroscope, and images were captured using a Leica DFC295 digital camera and Leica Application Suite 4.1. For the quantification of trichomes on first leaves, the leaves were placed on 1% agar plates and counting of trichomes on first leaves was performed manually upon viewing with the stereomicroscope. Scanning electron microscopy imaging was performed on fresh leaves and sepals using cold stage (−30°C) with an S-3500N instrument from Hitachi Science Systems equipped with an S-6542 Absorbed Electron Detector, a C1005 SEM Cold/Hot Stage (Oxford Instruments Microanalysis), and backscattered Robinson detector (ETP-USA/Electron Detectors). Images were processed as TIF files using Adobe Photoshop CS4.

Yeast One-Hybrid Analysis

Three tandem copies of L1-box sequences and mutated versions thereof were cloned into the pAbAi vector (Clontech) by annealing of oligonucleotides having KpnI and SalI restriction sites at their ends. Reporter bait constructs were integrated into the yeast genome at the ura3-52

locus using URA3+ selection in the yeast strain *SUB62* (*MATa lys2-801 leu2-3, 2-112 ura3-52 his3-Δ200 trp1-1*). Reporter strains were transformed by standard LiAc transformation with pGAD-T7 (Clontech) containing HD and ZLZ domains from GL2 (amino acids 1 to 231) or HDG11 (amino acids 1 to 169) that been inserted using the EcoRI and BamHI restriction sites. A construct containing the C-terminal START and SAD domains of GL2 (amino acids 253 to 747), also inserted by these restriction sites, was used as a control. Oligonucleotide sequences for this work are listed in Appendix Table D-4.

ChIP–Quantitative PCR

Transgenic seeds expressing *ProGL2:EYFP:GL2* were germinated on 0.8% (plant tissue culture grade) agar containing 1× Murashige and Skoog medium (Murashige and Skoog 1962), and seedlings were harvested at 10 d. Cross-linking of seedlings followed by extraction of chromatin were performed as described (Zhu 2012) with minor modifications. Chromatin was sonicated with the S220 Focused-Ultrasonicator (Covaris) to <500-bp DNA fragments. A 10-fold dilution of total chromatin was prepared to serve as the input control. Immunoprecipitations were performed overnight at 4°C with 2 μL of anti-GFP antibody (ab290; Abcam) bound to Dynabeads Protein A (Life Technologies). After washing sequentially with low-salt wash buffer (150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, and 20 mM Tris-HCl, pH 8), high-salt wash buffer (500 mM NaCl and otherwise identical to the first buffer), LiCl buffer (0.25 mM LiCl, 1% Nonidet P-40, 1% sodium deoxycholate, 1 mM EDTA, and 10 mM Tris-HCl, pH 8), and TE buffer (10 mM Tris-HCl, pH 8, and 1 mM EDTA), the immunocomplexes were eluted. Cross-linking was reversed by overnight incubation at 65°C, followed by proteinase K digestion, phenol-chloroform extraction, and ethanol precipitation. ChIP DNA was quantified using iTaq Universal SYBR Green Supermix with the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). PCR was performed in triplicate with the gene-specific primers listed in Appendix Table C-4. *ACT7* was used for normalization.

Transient Expression in *Nicotiana benthamiana*

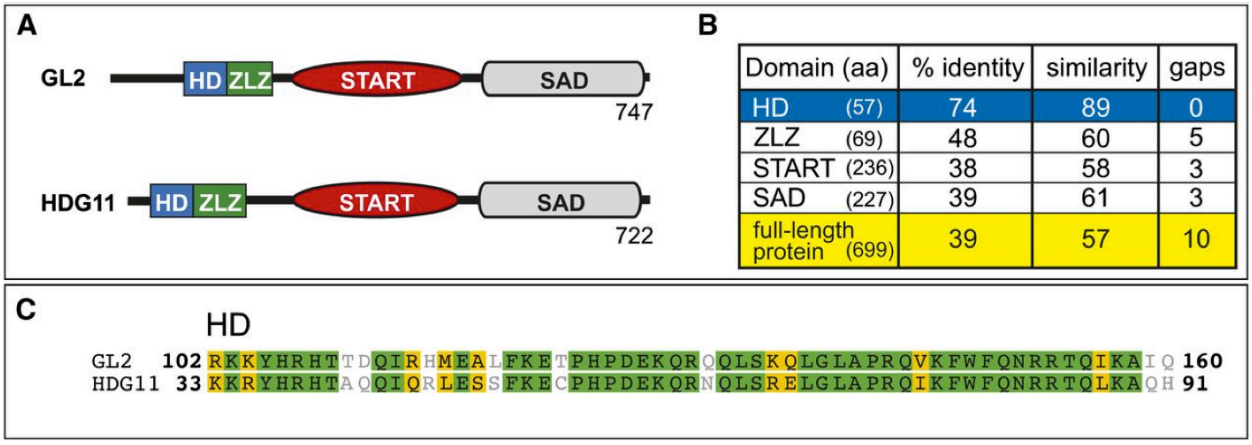
The *ProMYB23:mGFP5-ER* reporter contains the 3.1-kb *MYB23* promoter driving the expression of an endoplasmic reticulum–targeted modified GFP (mGFP-ER) in pCB302 (Kang

et al., 2009). To construct *ProMYB23_ΔL1:mGFP5-ER*, L1-box I was deleted using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) using the oligonucleotides listed in Appendix Table C-4. To construct the *Pro35S:GL2* and *Pro35S:HDG11* effector genes, cDNA sequences for *GL2* and *HDG11* were PCR amplified using Phusion High-Fidelity DNA Polymerase (Thermo Scientific), cloned into pENTR-D-TOPO (Life Technologies), and moved to the pK7WG2 expression vector by Gateway LR Clonase II (Life Technologies). *N. benthamiana* plants (2 to 4 weeks old) were used to express the various construct combinations by *Agrobacterium* (GV3101 pMP90)–mediated transient transformation of lower epidermal leaf cells as described previously (Sparkes et al. 2006). Fluorescence images were captured using a Zeiss LSM 5 PASCAL system at 3 d postinfiltration. GFP was excited at 488 nm and detected at 493 to 530 nm. Images of 512 × 512 pixels over an area of 307.1 × 307.1 μm were analyzed from three biological replicates for each effector and reporter combination. Quantification of the image data was performed by calculating the percentage of the measurement area for which the relative GFP fluorescence exceeded a threshold of 20% of the maximum overall signal intensity.

Accession Numbers

Sequence data from this article can be found in the *Arabidopsis* Genome Initiative or GenBank/EMBL databases under the following accession numbers: At1g79840 (GL2), At1g73360 (HDG11), and At5g40330 (MYB23).

FIGURES AND TABLES



Domain (aa)	% identity	similarity	gaps
HD (57)	74	89	0
ZLZ (69)	48	60	5
START (236)	38	58	3
SAD (227)	39	61	3
full-length protein (699)	39	57	10

Figure 5.1 Sequence relationships between Class IV HD-Zip transcription factors GL2 and HDG11

(A) Domain arrangements of the GL2 and HDG11 proteins composed of 747 and 722 amino acids, respectively.

(B) Percentage amino acid (aa) identities and similarities between GL2 and HDG11 for the four conserved domains and the respective full-length proteins. The HD domains exhibit the highest degree of amino acid identity and similarity.

(C) Amino acid alignment of the HD domains from the GL2 and HDG11 proteins, generated with Geneious 5.4.6 software (Biomatters). Green and yellow shading indicate identical and similar residues, respectively.

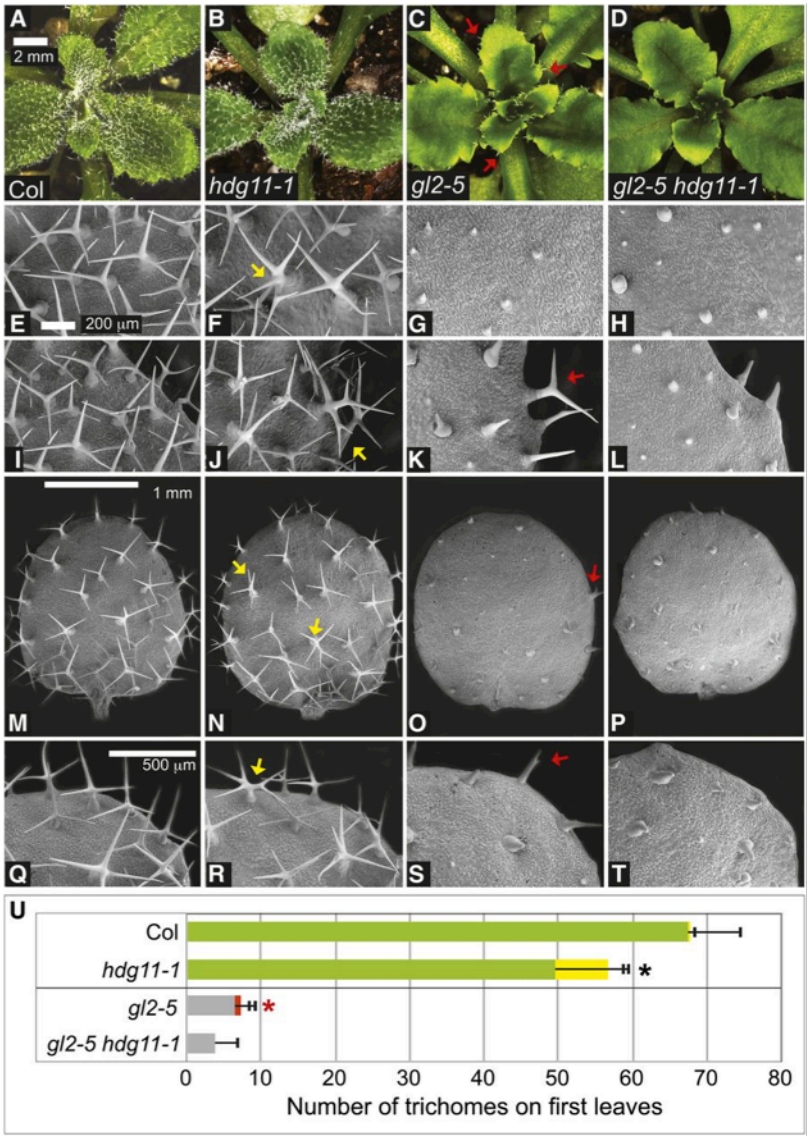


Figure 5.2 *gl2-5 hdg11-1* double mutants display an enhanced trichome differentiation defect

(A) to **(D)** Rosettes from 3-week-old *Arabidopsis* plants.

(A) and **(B)** Col wild-type **(A)** and *hdg11-1* **(B)** plants display trichomes covering leaf surfaces.

(C) and **(D)** *gl2-5* plants appear mostly glabrous except for trichomes at the leaf margins (arrows) **(C)**, while in comparison, *gl2-5 hdg11-1* double mutants display an enhanced glabrous phenotype **(D).**

(E) to **(L)** Scanning electron micrographs of leaves (**[E]** to **[H]**) and leaf margins (**[I]** to **[L]**) reveal detailed trichome morphologies.

(E) and **(I)** Col wild-type trichomes exhibit three to four branches.

(F) and **(J)** *hdg11-1* leaves show five or more branches for some trichomes (arrows).

(G) and **(K)** *gl2-5* trichomes are small and undifferentiated except at leaf margins, where branching (arrow) is observed.

(H) and **(L)** *gl2-5 hdg11-1* double mutants exhibit smaller undifferentiated trichomes at leaf margins.

(M) to **(T)** First leaves (**[M]** to **[P]**) and detailed edges of first leaves **([Q]** to **[T]**).

(M) and **(Q)** Col wild-type trichomes have three to four branches.

(N) and **(R)** *hdg11-1* mutants exhibit ectopic trichome branching (arrows).

(O) and **(S)** *gl2-5* mutants display undifferentiated trichomes except at leaf edges, where some trichomes display branching (arrows).

(P) and **(T)** *gl2-5 hdg11-1* double mutants have undifferentiated trichomes throughout the leaf.

(U) Quantification of trichomes and trichome branching on first leaves. For the Col wild type and *hdg11-1*, green bars show the number of trichomes with three to four branches and yellow bars indicate the number of trichomes with five or more branches. For *gl2* and *gl2 hdg11*, gray bars show the number of unbranched single-spiked trichomes while the red bar indicates branched trichomes. Positive error bars indicate SD for $n \geq 20$ plants. Asterisks indicate significant differences in trichome branching (two-tailed *t* test, $P < 0.0001$).

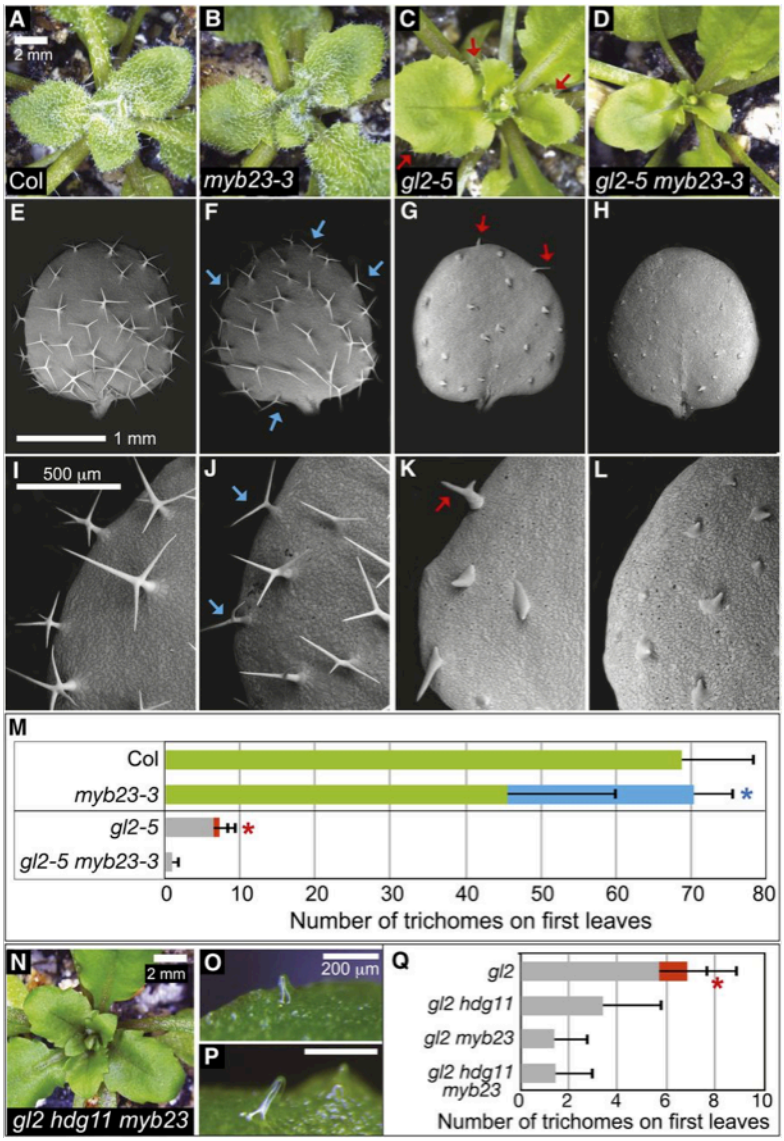


Figure 5.3 Enhanced trichome differentiation defects in the *gl2-5 myb23-3* double mutant and the *gl2 hdg11 myb23* triple mutant

(A) to **(D)** Rosettes from 3-week-old plants.

(A) and **(B)** Col wild-type **(A)** and *myb23-3* **(B)** plants display trichomes covering leaf surfaces.

(C) and **(D)** *gl2-5* plants display trichomes at leaf margins (arrows) **(C),** while *gl2-5 myb23-3* double mutants exhibit glabrous leaves **(D).**

(E) to **(L)** Scanning electron micrographs of first leaves reveal trichome morphology details.

(E) and **(I)** Col wild-type trichomes display three to four branches.

(F) and **(J)** *myb23-3* trichomes exhibit two to three branches. Arrows mark abnormal trichomes with two branches.

(G) and **(K)** *gl2-5* trichomes display differentiation defects, although some trichomes at leaf edges show branching (arrows).
(H) and **(L)** *gl2-5 myb23-3* double mutants exhibit small undifferentiated trichomes that lack branching.
(M) Quantification of trichomes on first leaves. For Col and *myb23-3*, green bars indicate the number of trichomes with three to four branches and the blue bar indicates the number of abnormal trichomes with fewer than three branches. For *gl2-5* and *gl2-5 myb23-3*, gray bars show the number of unbranched trichomes while the red bar indicates branched trichomes. Positive error bars indicate SD for $n \geq 20$ plants, and asterisks indicate significant deviations in trichome branching (two-tailed *t* test, $P < 0.0001$).
(N) Rosette from a *gl2-5 hdg11-1 myb23-3* triple mutant.
(O) and **(P)** Details of leaf trichomes from the triple mutant.
(Q) Quantification of trichomes on first leaves as in **(M)** from *gl2-5, gl2-5 hdg11-1, gl2-5 myb23-3*, and *gl2-5 hdg11-1 myb23-3*.

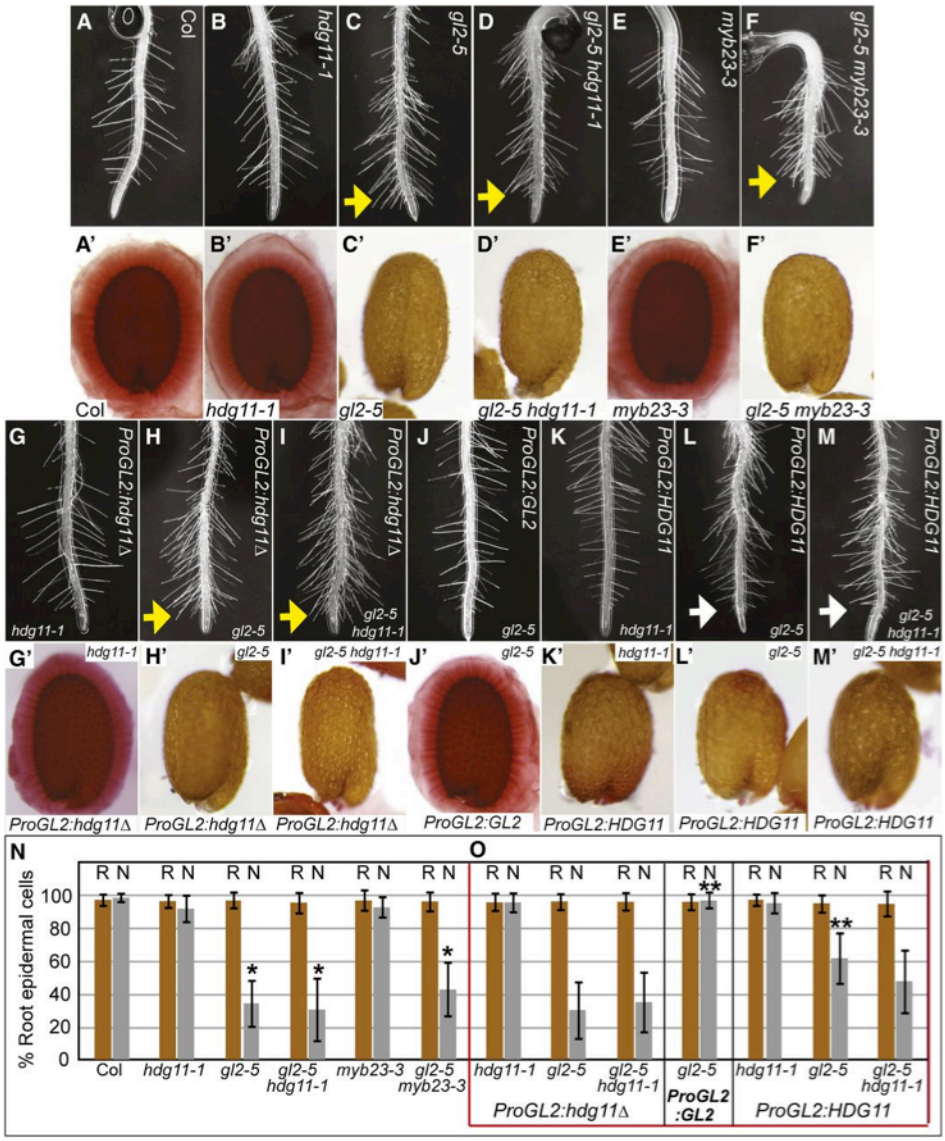


Figure 5.4 Root hair density and seed coat mucilage phenotypes reveal tissue-specific interactions

(A) to **(M)** Primary roots from 3- to 4-d-old seedlings.

(A) to **(F)** Col wild-type **(A)**, *hdg11-1* **(B)**, and *myb23-3* **(E)** root hair density appears normal in comparison with excess root hair formation in *gl2-5* **(C)**, *gl2-5 hdg11-1* **(D)**, and *gl2-5 myb23-3* **(F).**

(G) to **(I)** Transgenic lines expressing *ProGL2:EYFP:hdg11Δ* do not appear to alter root hair density in *hdg11-1* **(G)**, *gl2-5* **(H)**, or *gl2-5 hdg11-1* **(I).**

(J) Expression of *ProGL2:EYFP:GL2* rescues the root hair phenotype of *gl2-5*.

(K) to **(M)** Expression of *ProGL2:EYFP:HDG11* has no effect on root hair density in *hdg11-1* seedlings **(K).** Partial complementation is observed for *gl2-5* **(L)** and possibly *gl2-5 hdg11-1* **(M)**

roots (white arrows). Yellow arrows point to abnormal root hair formation near the tips of the roots.

(A′) to **(M′)** Seeds were stained with ruthenium red to detect mucilage production upon imbibition.

(A′) to **(F′)** Col wild-type **(A′)**, *hdg11-1* **(B′)**, and *myb23-3* **(E′)** seeds exhibit a normal outer mucilage, while *gl2-5* **(C′)**, *gl2-5 hdg11-1* **(D′)**, and *gl2-5 myb23-3* **(F′)** lack mucilage.

(G′) to **(I′)** *ProGL2:EYFP:hdg11Δ* expression does not affect seed mucilage production in *hdg11-1* **(G′)**, *gl2-5* **(H′)**, or *gl2-5 hdg11-1* **(I′).**

(J′) Expression of *ProGL2:EYFP:GL2* rescues the mucilage phenotype of *gl2-5*.

(K′) *ProGL2:EYFP:HDG11* expression negatively affects mucilage production in *hdg11-1* seeds.

(L′) and **(M′)** *gl2-5* **(L′)** and *gl2-5 hdg11-1* **(M′)** seeds expressing *ProGL2:EYFP:HDG11* lack the outer mucilage.

(N) and **(O)** Quantification of root epidermal cells in genotypes lacking transgenes **(N)** and expressing *EYFP:HDG11* or *EYFP:GL2* from transgenes (boxed in red) **(O).** Percentages of trichoblast cells in root hair cell files (R, brown bars) and atrichoblast cells in non-root-hair cell files (N, gray bars) are indicated ($n \geq 20$). Error bars indicate SD. Single asterisks in **(N)** mark significant differences from the Col wild type, while double asterisks in **(O)** mark significant differences between mutants with *ProGL2:EYFP:hdg11Δ* control and *ProGL2:EYFP:GL2* or *ProGL2:EYFP:HDG11* transgenes expressing wild-type *GL2* or *HDG11*, respectively (two-tiered *t* test, $P < 0.0001$).

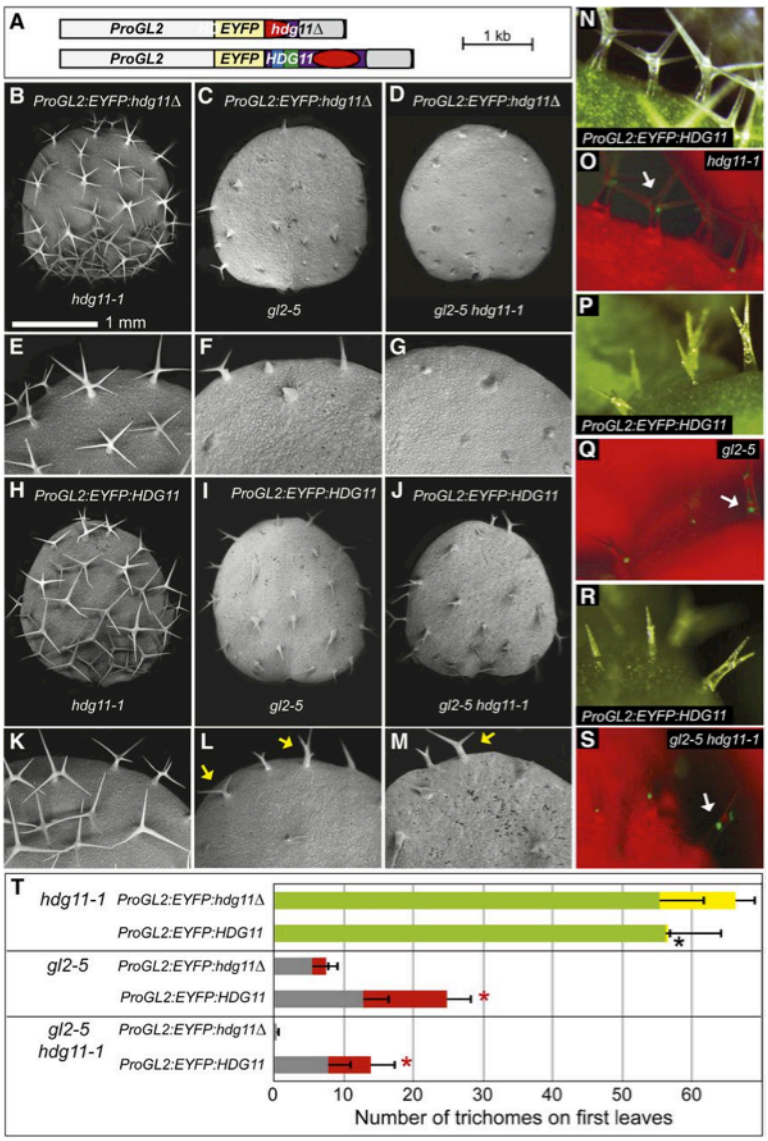


Figure 5.5 *ProGL2:EYFP:HDG11* Rescues *hdg11* Ectopic Branching and Partially Rescues the *gl2* Trichome Differentiation Defects.

(A) Transgene constructs *ProGL2:EYFP:hdg11Δ* and *ProGL2:EYFP:HDG11.*

(B) to **(M)** Scanning electron micrographs of first leaves **([B]** to **[D]** and **[H]** to **[J])** and magnification of leaf margins **([E]** to **[G]** and **[K]** to **[M])** are shown for mutant construct *ProGL2:EYFP:hdg11Δ* **([B]** to **[G])** and wild-type construct *ProGL2:EYFP:HDG11* **([H]** to **[M])** in *hdg11-1* **([B], [E], [H],** and **[K]),** *gl2-5* **([C], [F], [I],** and **[L]),** and *gl2-5 hdg11-1* **([D], [G], [J],** and **[M]**) mutant backgrounds. *ProGL2:EYFP:HDG11* complements *hdg11-1* ectopic branching **([H]** and **[K])** while conferring partial rescue of the *gl2-5* trichome phenotype **([I], [J], [L],** and **[M]**). In **(L),** arrows mark an internal branched trichome and a three-branched trichome,

indicative of partial rescue of the *gl2* phenotype. In **(M)**, the arrow shows branched trichomes on the leaf margin.

(N) to **(S)** Nuclear localization of *EYFP:HDG11* in trichomes. Matching white light and fluorescence images are from *hdg11-1* **([N]** and **[O]**), *gl2-5* **([P]** and **[Q]**), and *gl2-5 hdg11* **([R]** and **[S]**) leaves expressing *EYFP:HDG11*. Arrows indicate the brightest nuclear signal in each image.

(T) Quantification of trichomes and trichome branching on first leaves. For *hdg11-1*, green bars show the number of trichomes with three to four branches and yellow bars indicate the number of trichomes with five or more branches. For *gl2* and *gl2 hdg11*, gray bars show the number of unbranched trichomes while red bars indicate branched trichomes. Positive error bars indicate SD for $n \geq 20$ plants. Asterisks indicate significant differences in branching phenotypes for mutant (hdg11Δ) versus wild-type *HDG11* transgene (two-tiered *t* test, $P < 0.00001$).

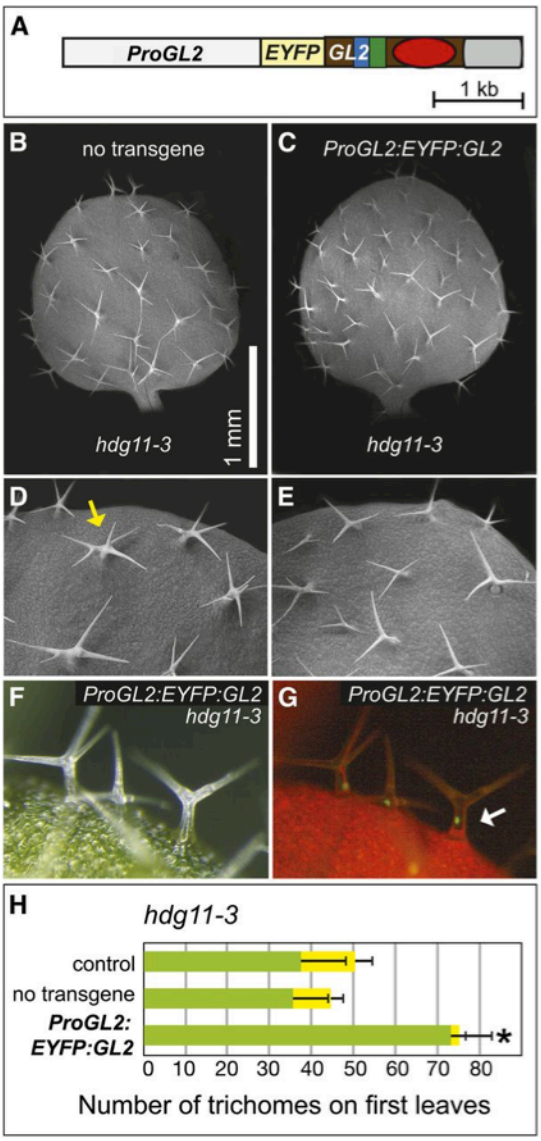


Figure 5.6 *ProGL2:EYFP:GL2* Rescues Ectopic Branching of *hdg11*

(A) Transgene construct *ProGL2:EYFP:GL2*, in which the cDNA of *GL2* is fused to an EYFP tag.

(B) and **(C)** Scanning electron micrographs of first leaves.

(D) and **(E)** Magnification of leaf margins are shown for an *hdg11-3* mutant lacking the transgene **(D)** and an *hdg11-3* mutant expressing *ProGL2:EYFP:GL2* **(E).** In **(D)**, the yellow arrow marks a trichome with ectopic branching.

(F) and **(G)** Matching white light and fluorescence images for the *hdg11-3* mutant expressing *ProGL2:EYFP:GFP*. The white arrow indicates a bright nuclear signal.

(H) Quantification of trichomes and trichome branching on first leaves of *hdg11-3* plants. The control is the *hdg11-3* parent from a cross to a line containing the transgene. Progeny of F2 offspring lacking the transgene were designated as "no transgene," while pure lines expressing the transgene were designated as *ProGL2:EYFP:GL2*. Green bars show the number of trichomes with three to four branches and yellow bars indicate the number of trichomes with five or more branches. Positive error bars indicate SD for $n \geq 20$ plants. The asterisk indicates a significant reduction in branched trichomes for *hdg11-3* plants expressing *ProGL2:EYFP:GFP* (two-tiered *t* test, $P < 0.0001$).

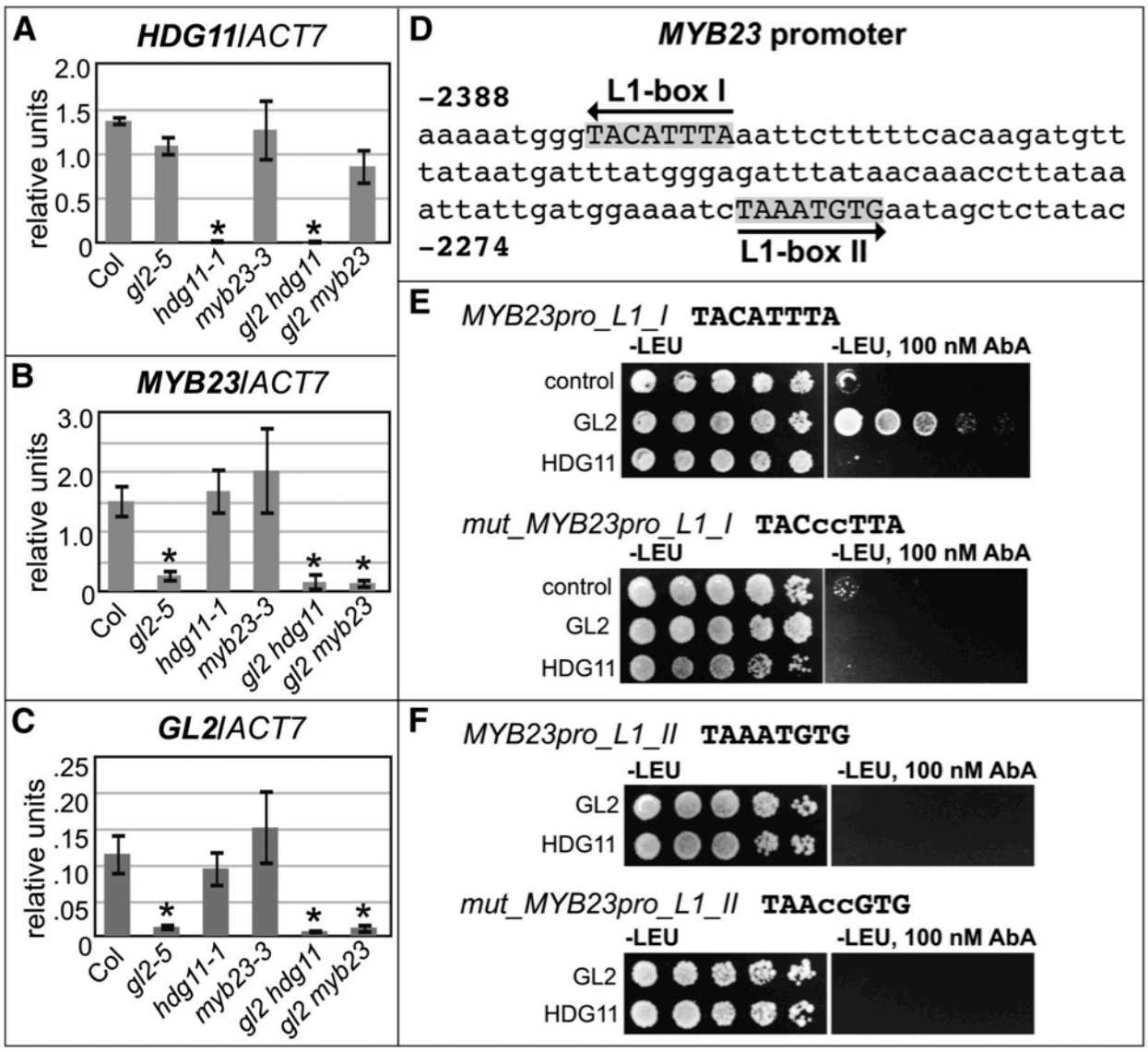


Figure 5.7 *GL2* Is Required for *MYB23* Transcription, and GL2 Binds *MYB23* L1-Box I in Yeast.

(A) to **(C)** Quantitative real-time PCR was performed with cDNA from 14-d-old seedling shoots with *ACT7* as the reference gene. Data represent averages of three biological replicates, and normalized units are plotted on the y axis. Error bars indicate SD. Significant differences (two-tiered *t* test, $P < 0.05$) from the wild type (Col) are indicated by asterisks.

(A) *HDG11* transcript is downregulated in *hdg11-1* mutants.

(B) *MYB23* transcript is downregulated in *gl2-5* mutants.

(C) *GL2* transcript is downregulated in *gl2-5* mutants.

(D) The 3.1-kb *MYB23* promoter region upstream of the ATG start codon contains two predicted L1-box sequences.

(E) and **(F)** Yeast one-hybrid analysis was performed with *MYB23* promoter–derived L1-boxes I and II. Nucleotide base substitutions are indicated in lowercase for the corresponding mutant L1-boxes.

(E) GL2 but not HDG11 binds to the wild-type L1-box I but not its mutant version.

(F) Neither GL2 nor HDG11 binds to the wild-type or mutant L1-box II.

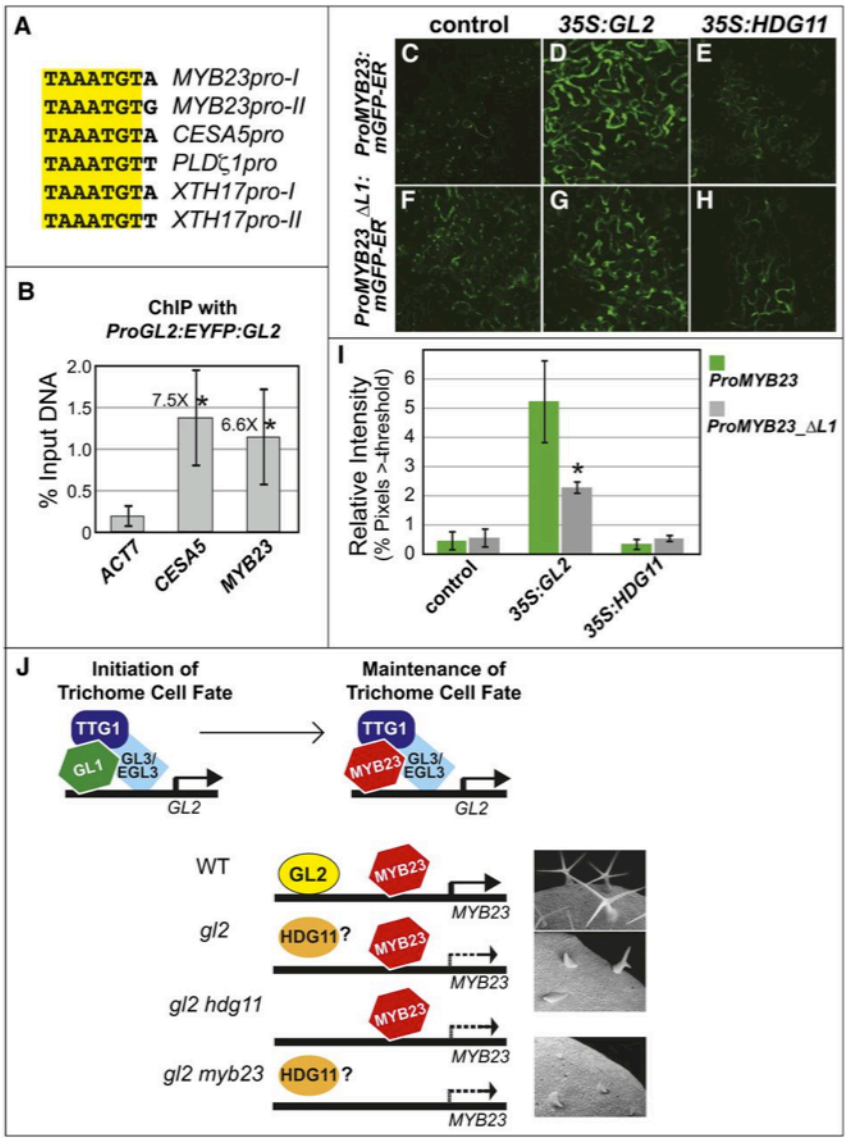


Figure 5.8 *MYB23* Is Transcriptionally Activated by GL2, and Full Activation Requires Binding of GL2 to an L1-Box in the *MYB23* Promoter.

(A) Alignment of L1-box sequences for GL2-regulated genes.

(B) ChIP with seedlings expressing *ProGL2:EYFP:GL2*, followed by quantitative real-time PCR of ChIP DNA. Primers that amplify L1-box regions in the *CESA5* and *MYB23* promoters, and *ACT7* control primers, were used. The average percentage input DNA and corresponding SD were calculated for four biological replicates. Fold differences are indicated, and asterisks mark significant differences from the *ACT7* control (two-tiered *t* test, $P < 0.05$).

(C) to **(H)** Representative images of *N. benthamiana* leaf epidermal cells expressing either a wild-type *MYB23* promoter fusion to GFP, *ProMYB23:mGFP-ER* **([C]** to **[E])**, or a mutant reporter, *ProMYB23_ΔL1:mGFP-ER*, in which the L1-box I is deleted **([F] to [H])**. Controls

([C] and **[F])** had the indicated reporter constructs but lacked effector genes. Reporter constructs in combination with *Pro35S:GL2* **([D]** and **[G]**) or *Pro35S:HDG11* **([E]** and **[H]**) effector genes are indicated.

(I) Quantification of the results shown in **(C)** to **(H)**, using images from three biological replicates. Error bars indicate SD. The asterisk marks a significant decrease in activation of the mutant reporter in comparison with the wild-type reporter by *GL2* (two-tiered t test, $P < 0.005$).

(J) Model for the relationship between the GL2, HDG11, and MYB23 transcription factors in trichome development. An activator complex composed of TTG1, GL1, and GL3/EGL3 positively regulates GL2 transcription for initiation of the trichome cell fate. *GL2* transcription is maintained as trichomes differentiate via the action of *MYB23*, which is functionally redundant with GL1. Both MYB23 and GL2 are required for full transcriptional activation of MYB23. In the absence of GL2, HDG11 can partially activate *MYB23* at leaf margins, either through direct binding to cis-regulatory elements normally occupied by GL2 or via an alternative mechanism. Trichome phenotypes are shown for the Col wild type (top), *gl2-5* (middle), and *gl2-5 myb23-3* (bottom).

REFERENCES

Balkunde R, Pesch M, Hulskamp M. 2010. Trichome patterning in *Arabidopsis thaliana* from genetic to molecular models. *Curr Top Dev Biol* **91**: 299-321.

Baumann E, Lewald J, Saedler H, Schulz B, Wisman E. 1998. Successful PCR-based reverse genetic screens using an En-1-mutagenised *Arabidopsis thaliana* population generated via single-seed descent. *Theor Appl Genet* **97**: 729-734.

Bruex A, Kainkaryam RM, Wieckowski Y, Kang YH, Bernhardt C, Xia Y, Zheng X, Wang JY, Lee MM, Benfey P et al. 2012. A gene regulatory network for root epidermis cell differentiation in *Arabidopsis*. *PLoS Genet* **8**: e1002446.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**: 735-743.

Di Cristina M, Sessa G, Dolan L, Linstead P, Baima S, Ruberti I, Morelli G. 1996. The *Arabidopsis* Athb-10 (GLABRA2) is an HD-Zip protein required for regulation of root hair development. *Plant J* **10**: 393-402.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* **15**: 573-581.

Floyd SK, Zalewski CS, Bowman JL. 2006. Evolution of class III homeodomain-leucine zipper genes in streptophytes. *Genetics* **173**: 373-388.

Grebe M. 2012. The patterning of epidermal hairs in *Arabidopsis*-updated. *Curr Opin Plant Biol* **15**: 31-37.

Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ, Chen XY. 2008. The HD-Zip IV gene GaHOX1 from cotton is a functional homologue of the *Arabidopsis* GLABRA2. *Physiol Plant* **134**: 174-182.

Hulskamp M, Misra S, Jurgens G. 1994. Genetic dissection of trichome cell development in *Arabidopsis*. *Cell* **76**: 555-566.

Kamata N, Okada H, Komeda Y, Takahashi T. 2013. Mutations in epidermis-specific HD-ZIP IV genes affect floral organ identity in *Arabidopsis thaliana*. *Plant J*.

Kang YH, Kirik V, Hulskamp M, Nam KH, Hagely K, Lee MM, Schiefelbein J. 2009. The MYB23 gene provides a positive feedback loop for cell fate specification in the *Arabidopsis* root epidermis. *Plant Cell* **21**: 1080-1094.

Kirik V, Lee MM, Wester K, Herrmann U, Zheng Z, Oppenheimer D, Schiefelbein J, Hulskamp M. 2005. Functional diversification of *MYB23* and *GL1* genes in trichome morphogenesis and initiation. *Development* **132**: 1477-1485.

Kirik V, Schnittger A, Radchuk V, Adler K, Hulskamp M, Baumlein H. 2001. Ectopic expression of the *Arabidopsis AtMYB23* gene induces differentiation of trichome cells. *Dev Biol* **235**: 366-377.

Koornneef M, Dellaert LW, van der Veen JH. 1982. EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat Res* **93**: 109-123.

Li SF, Milliken ON, Pham H, Seyit R, Napoli R, Preston J, Koltunow AM, Parish RW. 2009. The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell* **21**: 72-89.

Marks MD, Wenger JP, Gilding E, Jilk R, Dixon RA. 2009. Transcriptome analysis of *Arabidopsis* wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. *Mol Plant* **2**: 803-822.

Murashige T, Skoog F. 1962. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**: 473-&.

Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T. 2006. Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in *Arabidopsis*. *Plant Physiology* **141**: 1363-1375.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana. PLoS Biol* **3**: e196.

Ohashi Y, Oka A, Rodrigues-Pousada R, Possenti M, Ruberti I, Morelli G, Aoyama T. 2003. Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. *Science* **300**: 1427-1430.

Peterson KM, Shyu C, Burr CA, Horst RJ, Kanaoka MM, Omae M, Sato Y, Torii KU. 2013. *Arabidopsis* homeodomain-leucine zipper IV proteins promote stomatal development and ectopically induce stomata beyond the epidermis. *Development* **140**: 1924-1935.

Prigge MJ, Clark SE. 2006. Evolution of the class III HD-Zip gene family in land plants. *Evol Dev* **8**: 350-361.

Rerie WG, Feldmann KA, Marks MD. 1994. The *GLABRA2* gene encodes a homeo domain protein required for normal trichome development in *Arabidopsis. Genes Dev* **8**: 1388-1399.

Roeder AH, Chickarmane V, Cunha A, Obara B, Manjunath BS, Meyerowitz EM. 2010. Variability in the control of cell division underlies sepal epidermal patterning in *Arabidopsis thaliana. PLoS Biol* **8**: e1000367.

Roeder AH, Cunha A, Ohno CK, Meyerowitz EM. 2012. Cell cycle regulates cell type in the *Arabidopsis* sepal. *Development* **139**: 4416-4427.

Schrick K, Nguyen D, Karlowski WM, Mayer KF. 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol* **5**: R41.

Scott SP, Teh A, Peng C, Lavin MF. 2002. One-step site-directed mutagenesis of ATM cDNA in large (20 kb) plasmid constructs. *Hum Mutat* **20**: 323.

Shen B, Sinkevicius KW, Selinger DA, Tarczynski MC. 2006. The homeobox gene *GLABRA2* affects seed oil content in *Arabidopsis*. *Plant Mol Biol* **60**: 377-387.

Shi L, Katavic V, Yu Y, Kunst L, Haughn G. 2012. *Arabidopsis glabra2* mutant seeds deficient in mucilage biosynthesis produce more oil. *Plant J* **69**: 37-46.

Sparkes IA, Runions J, Kearns A, Hawes C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc* **1**: 2019-2025.

Stracke R, Werber M, Weisshaar B. 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* **4**: 447-456.

Takada S, Takada N, Yoshida A. 2013. ATML1 promotes epidermal cell differentiation in *Arabidopsis* shoots. *Development* **140**: 1919-1923.

Tominaga-Wada R, Iwata M, Sugiyama J, Kotake T, Ishida T, Yokoyama R, Nishitani K, Okada K, Wada T. 2009. The GLABRA2 homeodomain protein directly regulates *CESA5* and *XTH17* gene expression in Arabidopsis roots. *Plant J* **60**: 564-574.

Vernoud V, Laigle G, Rozier F, Meeley RB, Perez P, Rogowsky PM. 2009. The HD-ZIP IV transcription factor OCL4 is necessary for trichome patterning and anther development in maize. *Plant J* **59**: 883-894.

Wang S, Chen JG. 2008. *Arabidopsis* Transient Expression Analysis Reveals that Activation of *GLABRA2* May Require Concurrent Binding of GLABRA1 and GLABRA3 to the Promoter of *GLABRA2*. *Plant Cell Physiol* **49**: 1792-1804.

Wang S, Kwak SH, Zeng Q, Ellis BE, Chen XY, Schiefelbein J, Chen JG. 2007. TRICHOMELESS1 regulates trichome patterning by suppressing GLABRA1 in *Arabidopsis*. *Development* **134**: 3873-3882.

Western TL, Burn J, Tan WL, Skinner DJ, Martin-McCaffrey L, Moffatt BA, Haughn GW. 2001. Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in *Arabidopsis*. *Plant Physiology* **127**: 998-1011.

Western TL, Skinner DJ, Haughn GW. 2000. Differentiation of mucilage secretory cells of the *Arabidopsis* seed coat. *Plant Physiology* **122**: 345-356.

Yang C, Li H, Zhang J, Wang T, Ye Z. 2011. Fine-mapping of the woolly gene controlling multicellular trichome formation and embryonic development in tomato. *TAG Theoretical and Applied Genetics Theoretische und angewandte Genetik* **123**: 625-633.

Yang CX, Ye ZB. 2013. Trichomes as models for studying plant cell differentiation. *Cell Mol Life Sci* **70**: 1937-1948.

Yang JY, Chung MC, Tu CY, Leu WM. 2002. OSTF1: a HD-GL2 family homeobox gene is developmentally regulated during early embryogenesis in rice. *Plant Cell Physiol* **43**: 628-638.

Zhu J-Y, Sun, Y., Wang, Z.-Y. 2012. Genome-wide identification of transcription factor-binding sites in plants using chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq). in *Plant Signaling Networks: Methods and Protocols* (ed. ZY Zhi-Yong Wang), pp. 173-188. Springer Science+Business Media, LLC

Appendix A - Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity

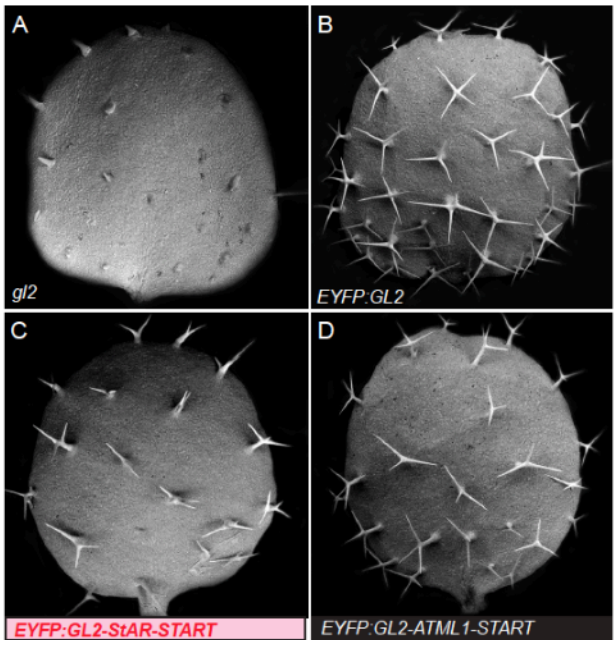


Figure A.1 Trichomes on first leaves of *gl2* mutants transformed with GL2 constructs.

(A-D) Scanning electron micrographs (SEM) of first leaves. **(A)** *gl2* mutants exhibit a defect in differentiation of trichome cells as indicated by short unbranched trichomes that barely emerge from the epidermis. *gl2* mutants transformed with **(B)** ProGL2:EYP:GL2 exhibit branched trichomes, indicating a rescue of the mutant phenotype, while *gl2* mutants transformed with **(C)** ProGL2:EYFP:GL2-StAR-START or **(D)** EYFP:GL2-ATML1-START display a partial rescue of the trichome differentiation defect.

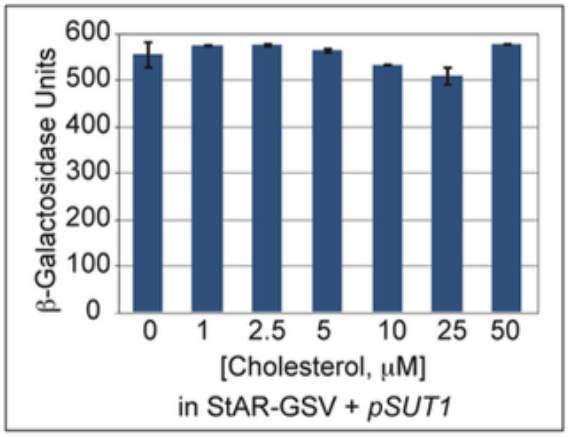


Figure A.2 The addition of cholesterol in the range from 0-50 μM had no effect on the activity levels of yeast cells expressing the GSV construct containing the mouse StAR START domain together with the pSUT1 plasmid. Error bars indicate standard deviations for two independent transformants in two trials.

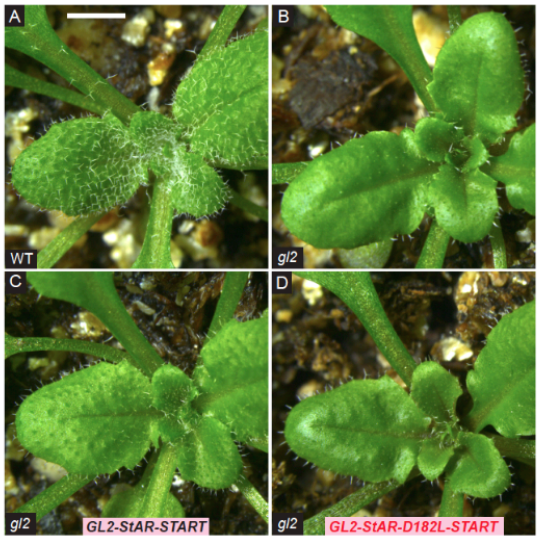


Figure A.3 Rosette phenotypes of StAR-START versus the D182L missense mutant expressed in the GL2 transcription factor.

(A-D) Rosettes exhibiting leaf trichomes. **(A)** Wild-type (WT) level of trichomes in comparison to **(B)** *gl2* null mutant which displays a reduction in leaf trichomes. **(C-D)** Representative *gl2* lines expressing **(C)** *ProGL2:EYFP:GL2-StAR-START* or **(D)** *ProGL2:EYFP:GL2-StAR-D182L-START*. While mouse StAR-START can partially replace the GL2-START domain, the missense mutation D182L results in a reduction in trichome cell differentiation. Scale bar = 2 mm.

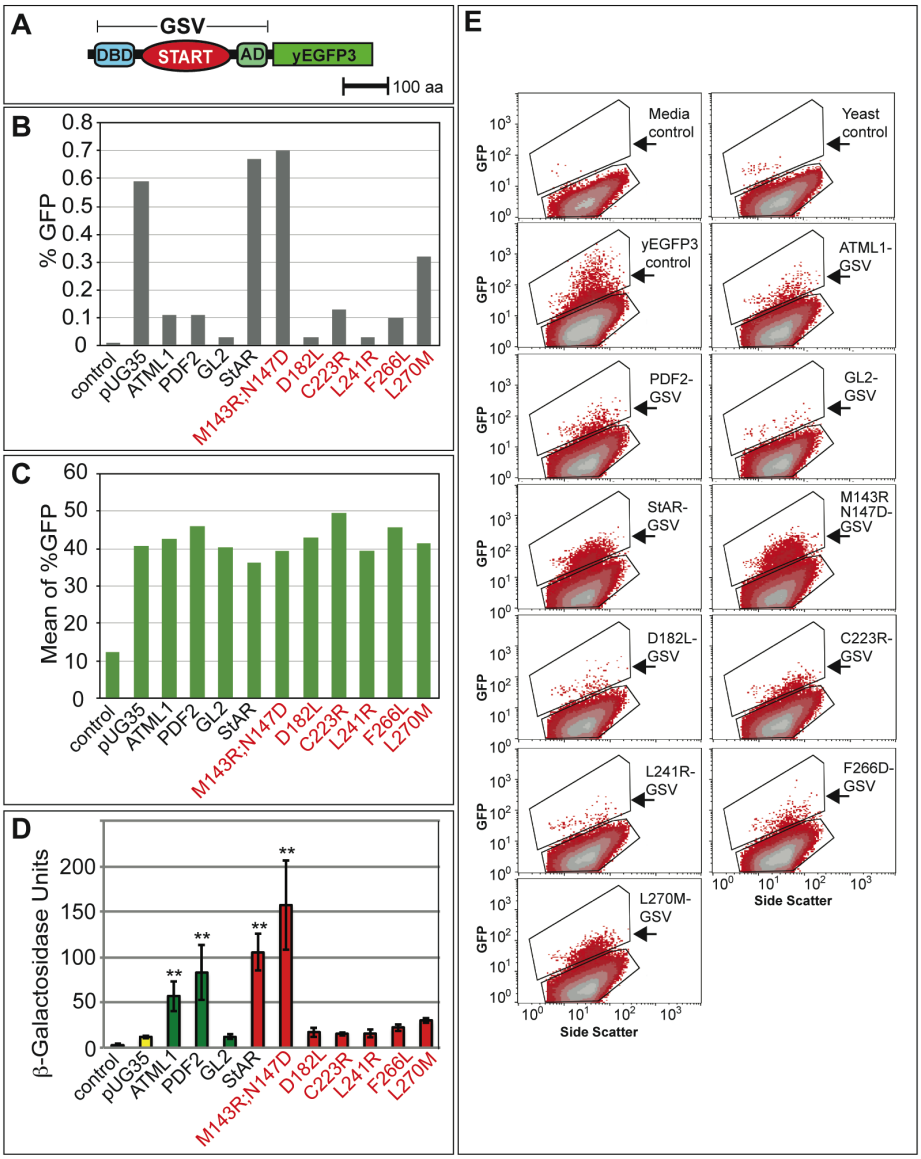


Figure A.4 *In vivo* expression of GSV constructs as yEGFP3 fusions in yeast.

(A) Schematic of GSV translational fusion to yEGFP3. **(B)** Flow cytometry data for the % GFP cells. The negative control which does not contain GFP corresponds to 0.01% GFP positive cells and the positive control which contains yEGFP3 alone (pUG35) corresponds to 0.59% GFP. The GFP-expressing cells exhibit % GFP values ranging from 0.03-0.70%. **(C)** Mean values for % GFP from side scatter plots. The negative control shows a mean value of 12 while the positive control (pUG35) exhibits a mean value of 41. The GFP-expressing cells show mean values in the range from 36-47.**(D)** Activity levels of the corresponding GSV-yEYFP3 constructs containing START domains from *Arabidopsis* ATML1, PDF2, and GL2 (green), and

mammalian StAR and corresponding mutants (red) are indicated. Error bars show standard deviations for two independent transformants in three trials, and double asterisks indicate a significant increase in activity over the pUG35 control (Two-tiered *t*-test, $P < 0.05$). **(E)** Flow cytometry side scatter plots of GFP positive yeast cells expressing yEGFP3. The top polygon from each plot indicates the population of cells that were gated as GFP positive (arrows). Side scatter is indicated on the X-axis and GFP signal is indicated on the Y-axis. "Media control" lacks yeast cells, while the "Yeast control" contains yeast cells that carry the same selectable marker (*URA3*) as the remaining samples albeit no GFP expression. The yEGFP3 control exhibits strong expression of yEGFP3 from the pUG35 plasmid. The sample order of the GSV-yEGFP3 constructs from top to bottom, right to left, corresponds to that in **A- D**. Each of the GSV samples indicates the presence of GFP positive cells in comparison to the negative controls.

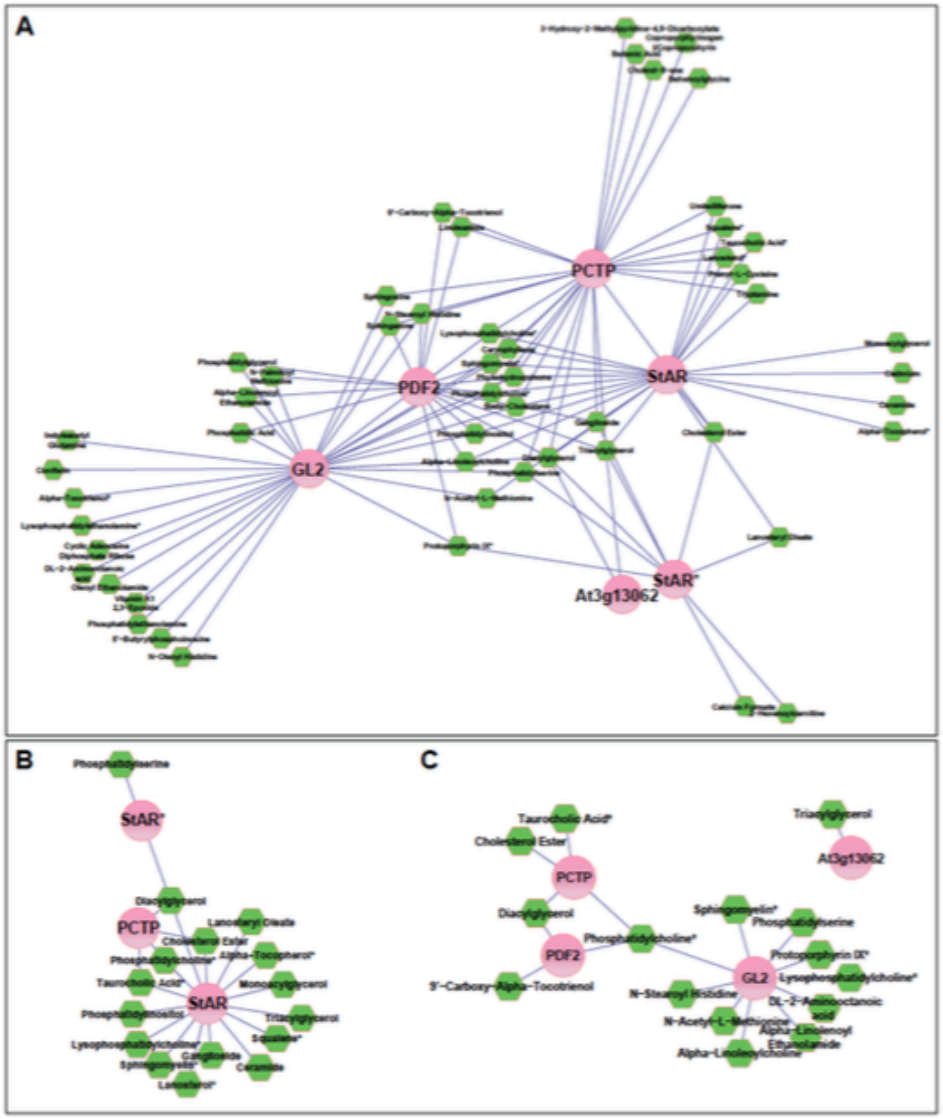


Figure A.5 Protein-metabolite interaction network for mammalian and *Arabidopsis* START domains.

(A) Normalized protein-metabolite enrichment data expressed as the fold-change of domain-bound metabolite relative to the GV control greater than 4 were processed using Cytoscape to produce an edge-weighted interaction network in which larger elliptical nodes represent the different START domains tested and hexagonal nodes represent the interacting metabolites. Distances between protein and metabolite nodes reflect the interaction strengths based upon the magnitude of fold-change – the shorter the edge the more enriched the metabolite. **(B)** A sub-network was generated to compare and contrast the nature of protein-bound metabolites between the mammalian START domains, PCTP (human), StAR (mouse) and

StARD182L* (mouse). **(C)** A sub-network comparing the *Arabidopsis* and human PCTP START domains. The sub- networks **(B, C)** were filtered for interactions with a greater than 10-fold change in enrichment relative to the GV control and only high confidence metabolite assignments were included. For all networks **(A-C)**, in cases where a node had multiple interactions with the same chemical sub-class of metabolite, e.g. PtCho, these interactions were combined and weighted to give one interaction. Metabolite names designated by asterisks were further validated by mass spectrometry, matching exact mass and retention time to a known standard analyzed under the same experimental conditions.

Table A-1 Oligonucleotides used in this study.

Nucleotide bases shown in bold denote restriction sites used for cloning or changed bases from site-directed mutagenesis unless otherwise indicated.

I. Primers for GL2 START domain deletion construct and GL2 START domain swaps. Homologous sequences for domain swap in-fusion cloning are indicated in bold.	
Name	**5'-3' sequence**
GL2_START_Δ_F	[Phos] GTC TTC TTC ATG GCT ACC AAC GTC CCC ACC
GL2_START_Δ_R	[Phos] GAG GGC AAA GAC GCC CGT GTA GAA ATC G
GL2_START_flank_right_F	**GTC TTC TTC ATG GCT** ACC AAC GTC
GL2_START flank_left_R	**GAG GGC AAA GAC GCC** CGT GTA
GL2_ATML1_START_F	**GGC GTC TTT GCC CTC** GAG GCT GAT AAG CCT ATG ATT G
GL2_ATML1_START_R	**AGC CAT GAA GAA GAC** GAG CCG CTC ACA TTG GCG GTC
GL2_EDR2_START_F	**GGC GTC TTT GCC CTC** AAC CAA GCA TTT TCC AGG AA
GL2_EDR2_START_R	**AGC CAT GAA GAA GAC** CCA CCC TTT TAG ATC AAT TTG
GL2_REV_START_F	**GGC GTC TTT GCC CTC** GAG GAG ACT TTG GCA GAG TTC
GL2_REV_START_R	**AGC CAT GAA GAA GAC** CCG CAA CGC GGA AAT GGT CA
GL2_mStAR_START_F	**GGC GTC TTT GCC CTC** GAC CAG GAG CTG TCC TAC ATC C
GL2_mStAR_START_R	**AGC CAT GAA GAA GAC** GCT GGC TTC CAG GCG CTT GC

II. Gene specific primers for PCR amplification and cloning of START domain coding regions in GSV plasmids.	
Name	**5'-3' sequence**
At1g64720_for_KpnI_218	CTCACCACGTTAACCCC**GGTACC**TCTTCCAAAGAG
At1g64720_rev_SacI_945	GTGAGCCATTATGGC**GAGCTC**GGATAAACCTGCTC
At2g28320_for_KpnI_418	TTGAGTAGCTCA**GGTACC**GACCATCACTCAAACTC
At2g28320_rev_SacI_1151	CTTGCACTTCTTG**GAGCTC**CCCCTGACGACAG
At3g13062_for_KpnI_201	CTCGGTTTCTCAATCT**GGTACC**TCCCAATCAGG
At3g13062_rev_SacI_934	AGCTTACAGC**GAGCTC**TGTGGGCCCTTGGGGTCG
At4g14500_for_KpnI_365	TGGCCTCAAGA**GGTACC**GATAACGGG
At4g14500_rev_SacI_1084	CCATTTGGGC**GAGCTC**AGATAGAGATGAGTCTG
At5g07260_for_KpnI_229	CTATATCCC**GGTACC**GCTACGTCTTTGACTG
At5g07260_rev_SacI_952	CTGGTCGAATAT**GAGCTC**ATTGTGACCAATTGAAGG
At5g35180_for_KpnI_634	CAAGGTCCA**GGTACC**CTTTTGAGGCAATCATC
At5g35180_rev_SacI_1382	TGGAACTTGGA**GAGCTC**AACCGTGGCGGAAG
At5g45560_for_KpnI_487	AGGACAACTATT**GGTACC**GGCCCTCCAGAATC
At5g45560_rev_SacI_1229	GATGCCATATT**GAGCTC**AACAGGGATCCTGATCGG
At5g54170_for_KpnI_344	TTTCAAGA**GGTACC**AAAACAAAGGAGAGATTGCC
At5g54170_rev_SacI_1065	CATGAAAGCAGACCG**GAGCTC**CTTGGTTCCTCC
ANL2 outer F_824	CCTCCTTAGAACTCGCTGTCGGCACC
ANL2 outer R_1780	GCTTGCTCCAATTGTGGACCGACG
ANL2_for_KpnI_915	GCAGCAGCAGCAGTCG**GGTACC**ATTAATGGG
ANL2_rev_SacI_1685	GGCGTTATTGATGT**GAGCTC**GTGAGATGTAACGG
ATHB8_for_ KpnI_429	GACCCCT**GGTACC**CAGCCTCGTGATGC
ATHB8_rev_ SacI_	GCCGCTGGTCT**GAGCTC**CCAACCTG
ATML1_for_ KpnI_	ACATTTTGAGGTCG**GGTACC**ATACCTTCTGAGGC
ATML1_rev_ SacI_	CAGGACTCGTTATCACGGA**GAGCTC**ACAAGC
CNA_for_ KpnI_432	GGCATCT**GGTACC**CCTCAGAGAGATGC
CNA_rev_ SacI_1162	GACGCCGTCC**GAGCTC**ATTAACACTAC
FWA outer F_450	GGCTGAGAATGCTAACTTGGAGCGGG
FWA outer R_1440	GCCACTTGTCCACCGAAGGACTCG
FWA_for_KpnI_592	GATTTTAGTGGT**GGTACC**AGGACGTCTGAGAAGG
FWA_rev_SacI_1361	GCAGACAATCC**GAGCTC**AATTTCAGTCAAGTTG
GL2_for_ KpnI_A_	GTCTC**GGTACC**CTCGATTTCTACACGGGCGTC

GL2_rev_ SacI_A_	CTTTGGT**GAGCTC**GTTGGTAGCCATGAAGAAGAC
GL2_for_ KpnI_B_	TCGGCTCTCTC**GGTACC**TACACGGGCGTC
GL2_rev_ SacI_B_	TGTAACTCC**GAGCTC**GTCTTTGGTGGGGACG
GL2_for_KpnI _728	TCTACACG**GGTACC**TTTGCCCTCGAGAAGTCCCG
GL2_rev_SacI_1500	TCCGAGAGA**GAGCTC**GGTGGGGACGTTGGTAG
HDG1_for_KpnI_910	CAACCG**GGTACC**GTTAGTGATTTTGATC
HDG1_rev_SacI_1674	GCAGTTTATAGGGGATGG**GAGCTC**GGAAGTGG
HDG2_for_deltaSacI_759	CGTGGCTGCAATGGAAGAA**A**CTCATGAGGATGGT
HDG2_rev_deltaSacI_791	ACCATCCTCATGAG**T**TCTTCCATTGCAGCCACG
HDG2_for_deltaSacI_1038	AGGAAACTATAATGGAGC**C**CTTCAAGTGATGAGTGC
HDG2_rev_deltaSacI_1073	GCACTCATCACTTGAAG**G**GCTCCATTATAGTTTCCT
HDG2_for_KpnI_712	ATCACTGCA**GGTACC**GAATCTGACAAACC
HDG2_rev_SacI_1415	GTAGCCATGAC**GAGCTC**TAACCGCTCGC
HDG3 outer F (625-650)	CATCCCCGTGTGTCTCCTCCTAATCC
HDG3 outer R (1511-1537)	TGGTCATTCCAGCAAAGAAGGTTCTCG
HDG3_for_KpnI	CCACTCGAGGGAAACC**GGTACC**CCTGCAGATGC
HDG3_rev_SacI	TCTTTCCATGGTTAGTTAGCGC**GAGCTC**GACAG
HDG4 outer F (539-562)	CTTGTGGCCACAATCTCCGCCTCG
HDG4 outer R (1447-1475)	TGTGACAGCTTCATCAAGTTCTTCCTCGC
HDG4_for_ KpnI	AAGAACAACAACGAT**GGTACC**TTGATTGCGG
HDG4_rev_SacI	AGGTAT**GAGCTC**AAGGTCAGTGATGTTTGTAGC
HDG5 outer F (808-836)	GACATGAGTGTATACGCTGGGAACTTTCC
HDG5 outer R (1766-1791)	GGTCCAAGACTGTCCATATGCAGTGC
HDG5_for_KpnI	CAACAAC**GGTACC**TTACTTGCGGATGAAGAAAAGG
HDG5_rev_ SacI	GCAGATGAAATTAC**GAGCTC**ATCAGTTATGTTTCTAGC
HDG8_for_deltaSacI_649	AGTGCGGTTGAAGAGCT**G**AAGCGGCTGTTTTTGGC
HDG8_rev_deltaSacI_683	GCCAAAAACAGCCGCTT**C**AGCTCTTCAACCGCACT
HDG8_for_KpnI_597	ACCACGACCA**GGTACC**GAAACGGATATGAGCC
HDG8_rev_SacI_1322	ATGGAGGA**GAGCTC**CATCCTCTCACAC
HDG9_outer_F_571	TTCTAACCGTCTCCCCGAGCCTTCAAGC
HDG9_outer_R_1547	GACTGTGGCGAGAAGTCGAGTTTGTTAACC
HDG9_deltaSacI_F_1329	CTTTGGCTACGGAGC**C**CGACGTTGGACCG
HDG9_deltaSacI_R_1357	CGGTCCAACGTCG**G**GCTCCGTAGCCAAAG
HDG9_for_KpnI_669	GGAAATGCAGAAT**GGTACC**CCACTATCTCAACTGG
HDG9_rev_SacI_1437	AACTCCGGGATT**GAGCTC**GTTGGGCAAGGC
HDG11_for_deltaKpnI_1000	CAGGAATGGGAGGTAC**G**CATGAGGGTGC
HDG11_rev_deltaKpnI_1028	GCACCCTCATG**C**GTACCTCCCATTCCTG
HDG11_for_KpnI_663	GCCTAACTTGGCT**GGTACC**GACATGGATAAGCC
HDG11_rev_SacI_1400	GAAGACGCTGGTACGGATAG**GAGCTC**AAATCTTTCACAC
HDG12_for_ KpnI_592	CCATCTCAGCCA**GGTACC**GTTTTATCAGAGATGG
HDG12_rev SacI_1361	ACTCCTCC**GAGCTC**AAGGGATGATG
MLN64_deltaSacI_F_867	GCCCTGTCCTGCGGAGCT**T**GTGTACCAGG
MLN64_deltaSacI_R_867	CCTGGTACAC**A**AGCTCCGCAGGACAGGGC
MLN64_for_ KpnI_	TCCTTTGCA**GGTACC**GACAATGAATCAGATGAAGAAG
MLN64_rev_ SacI_	TATCA**GAGCTC**CGCCCGGGCCCCC
PCTP_for_ KpnI_	GACTGC**GGTACC**ATGGAGCTGGCCGCCG
PCTP_rev_ SacI_	TCAACCCATGGATGCAATGTTCC**GAGCTC**TCTTTCATAGG
PDF2_for_KpnI	TTGAGGTCA**GGTACC**ATTCCTTCTGAGACTG
PDF2_rev_SacI	TATCACGGA**GAGCTC**ACCAGGAATGTTGC
PHB_for_ KpnI_463	AACCCAAATCCTCAG**GGTACC**CAACGTGATGC
PHB_rev_ SacI_1198	CAGGTTG**GAGCTC**TCCACCATACTG
REV_for_KpnI_423	GGTCACAACTCCTCAG**GGTACC**CTTAGAGATG
REV_rev_SacI_1162	CAGCAGGCTG**GAGCTC**TAATCCATACACTACT

mStAR_for_ KpnI_	GTCAGTCCTT**GGTACC**CAACTGGAAGCAACACTC
mStAR_rev_ SacI_	TTAACACTG**GAGCTC**AGAGGCAGGGCTGGC

III. Primers for sequencing plasmid inserts, construction of GV plasmid, or cloning of the yEGFP3 expression vector (pUG35) and protein expression vector BG1805

Name	5'-3' sequence
GSV_seq_for	TCCCAAAACCAAAAGGTCTCCGCTG
GSV_seq_rev	CCCCAACATGTCCAGATCGAAATCG
Gal4DBD_for_1	ATGAAGCTACTGTCTTCTATCGAAC
Gal4DBD_rev_276	CAATGCTTTTATATCCTGTAAAGAATCC
Gal4_NruI_for_282	TACCCCTGCAGCTGCG**TCGCGA**CTAGAGGATCC
Gal4_NruI_rev_314	GGATCCTCTAG**TCGCGA**CGCAGCTGCAGGGGTA
VP16_NruI_for_1182	TGCGGGCTCTACTTCATCG**TCGCGA**CACTTAGACGGCG
VP16_NruI_rev_1219	CGCCGTCTAAGTG**TCGCGA**CGATGAAGTAGAGCCCGCA
pUG35_seq_3117R_MET25p	TTCCTTCGTGTAATACAGGGTCG
pUG35_seq_2964F_yEGFP	ACCAAAATTGGGACAACACCAGTG
pUG35_MET25p_for_207	GCACCTTGTCCAATTGAACACGC
pUG35_yEGFP_rev_730	ACCTTCTGGCATGGCAGACTTG
pUG35_for_ATG	CATCCATACTCTAGA**ATG**AGTGGATCCCCCGGGC
pUG35_rev_ATG	GCCCGGGGGATCCACT**CAT**TCTAGAGTATGGATG
pGSV_for_BamHI	AAGCAA**GGATCC**TGAAAGATGAAGCTACTGTC
pGSV_rev_EcoRI	TCGCGC**GAATTC**CCCACCGTACTCG
pGS_rev_EcoRI	ACTATAGGGC**GAATTC**GAGCTCCACC
pG_rev_EcoRI	GTCTAAGTG**GAATTC**GGTACCTAACAATGC
GSV_for_pENTR TOPO	CACCATGAAGCTACTGTCTTCTATCGAAC
GSV_rev_pENTR TOPO	TGCCCCACCGTACTCGTCAATTCCAAG

IV. Primers for site-directed mutagenesis of mouse StAR START domain

Name	5'-3' sequence
StAR M143R;N147D_for (atg->agg;aac->gac)	GC ATG GAG GCC A**G**G GGA GAG TGG **G**AC CCA AAT GTC
StAR M143R;N147D_rev	GAC ATT TGG GT**C** CCA CTC TCC C**C**T GGC CTC CAT GC
StAR R181L;D182L_for (cga->cta;gac->ctc)	CTG GTG GGG CCT C**T**A **CT**C TTC GTG AGC GTG CGC
StAR R181L;D182L_rev	GCG CAC GCT CAC GAA G**AG** T**A**G AGG CCC CAC CAG
StAR R181L_for (cga->cta)	G GGG CCT C**T**A GAC TTC GTG AGC GTG CG
StAR R181L_rev	CG CAC GCT CAC GAA GTC T**A**G AGG CCC C
StAR D182L_for (gac->ctc)	CTG GTG GGG CCT CGA **CT**C TTC GTG AGC GTG CGC
StAR D182L_rev	GCG CAC GCT CAC GAA G**AG** TCG AGG CCC CAC CAG
StAR C224R_for (tgc->cgc)	GAA CAC GGC CCC ACC **C**GC ATG GTG CTT CAT CC
StAR C224R_rev	GG ATG AAG CAC CAT GC**G** GGT GGG GCC GTG TTC
StAR L241R_for (ctg->cgg)	CC AAG ACT AAA CTC ACT TGG C**G**G CTC AGT ATT GAC C
StAR L241R_rev	G GTC AAT ACT GAG C**C**G CCA AGT GAG TTT AGT CTT GG
StAR F266D_for (ttc->gac)	CC TA TCG CAG ACC CAG ATA GAG **GA**C GCC AAC CAC C
StAR F266D_rev	G GTG GTT GGC G**TC** CTC TAT CTG GGT CTG CGA TA GG
StAR L270M_for (ctg->atg)	GAG TTC GCC AAC CAC **A**TG CGC AAG CGC CTG G
StAR L270M_rev	C CAG GCG CTT GCG CA**T** GTG GTT GGC GAA CTC

Appendix B - Functional characterization of START domain in Class IV HD-Zip transcription factor GLABRA2

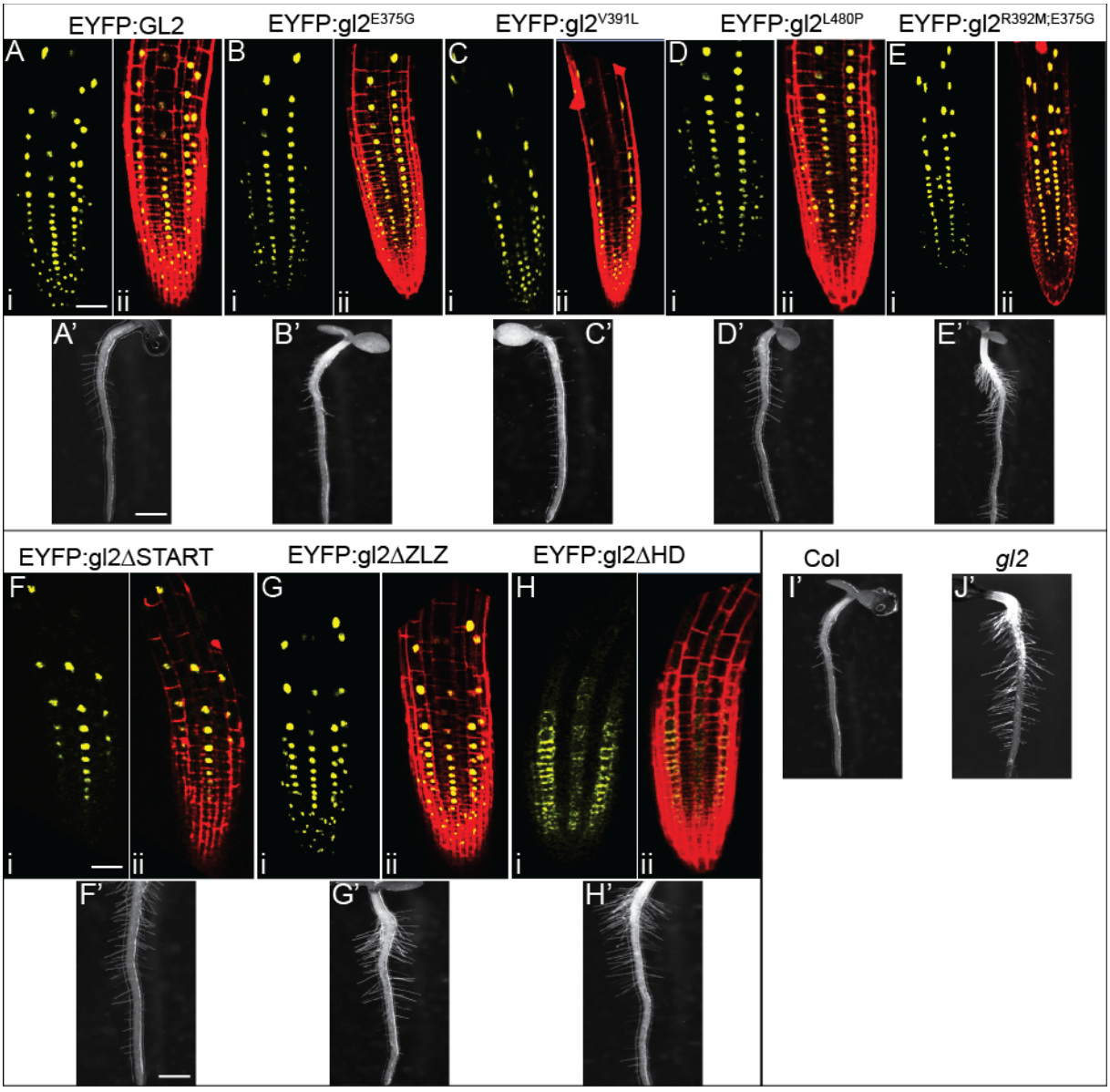


Figure B.1 Sub-cellular localization and root hair phenotypes of GL2 domain deletion and START missense variants.

(A-E and F-H) Confocal images of 4-day-old light grown seedlings expressing indicated proteins. **(i)** Subcellular localization of the EYFP fusion protein (yellow) **(ii)** Propidium iodide staining (red) was performed to visualize cell boundaries. Scale bar = 50 µm. **(I')** Col wild-type

exhibits normal root hair formation in contrast to **(J')** *gl2-5* which displays excess root hair formation. Expression of **(A')** *EYFP:GL2*, **(B')** *EYFP:gl2*^{E375G}, and **(C')** *EYFP:gl2*^{V391L} completely rescues the root hair phenotype of *gl2-5*. Expression of **(E')** *EYFP:gl2*$^{E375G;R392M}$, **(F')** *EYFP:gl2*Δ*START*, **(G')** *EYFP:gl2*Δ*ZLZ*, and **(H')** *EYFP:gl2*Δ*HD* has no effect on root hair density in *gl2-5* seedlings **(D')** while partial complementation is observed for *gl2-5* seedlings expressing *EYFP:gl2*^{L480P}

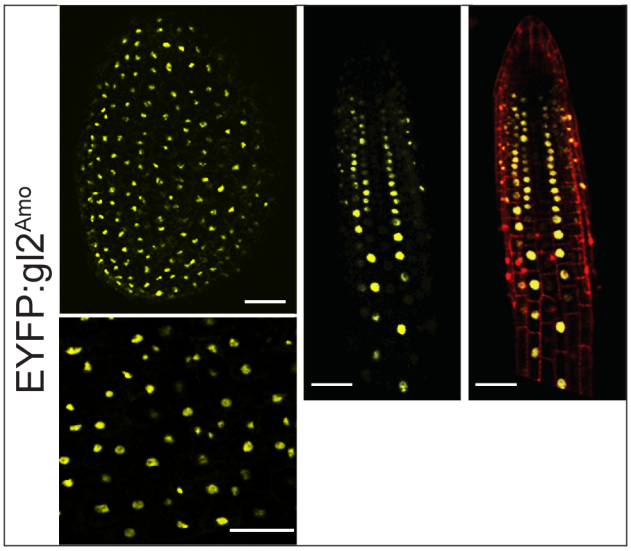


Figure B.2 Subcellular localization of *EYFP:gl2*^{Amo} in ovules and roots of *Arabidopsis gl2* null mutants.

Confocal laser scanning images show nuclear localization in **(A, B)** *Arabidopsis* ovules and **(C, D)** seedlings expressing *EYFP:gl2*^{Amo} **(C)** Subcellular localization of the EYFP fusion protein (yellow), **(D)** Propidium iodide staining (red) was performed to visualize cell boundaries. Scale bars = **(A)** 50 μm**, (B)** 100 μm**, (C-D)** 50 μm

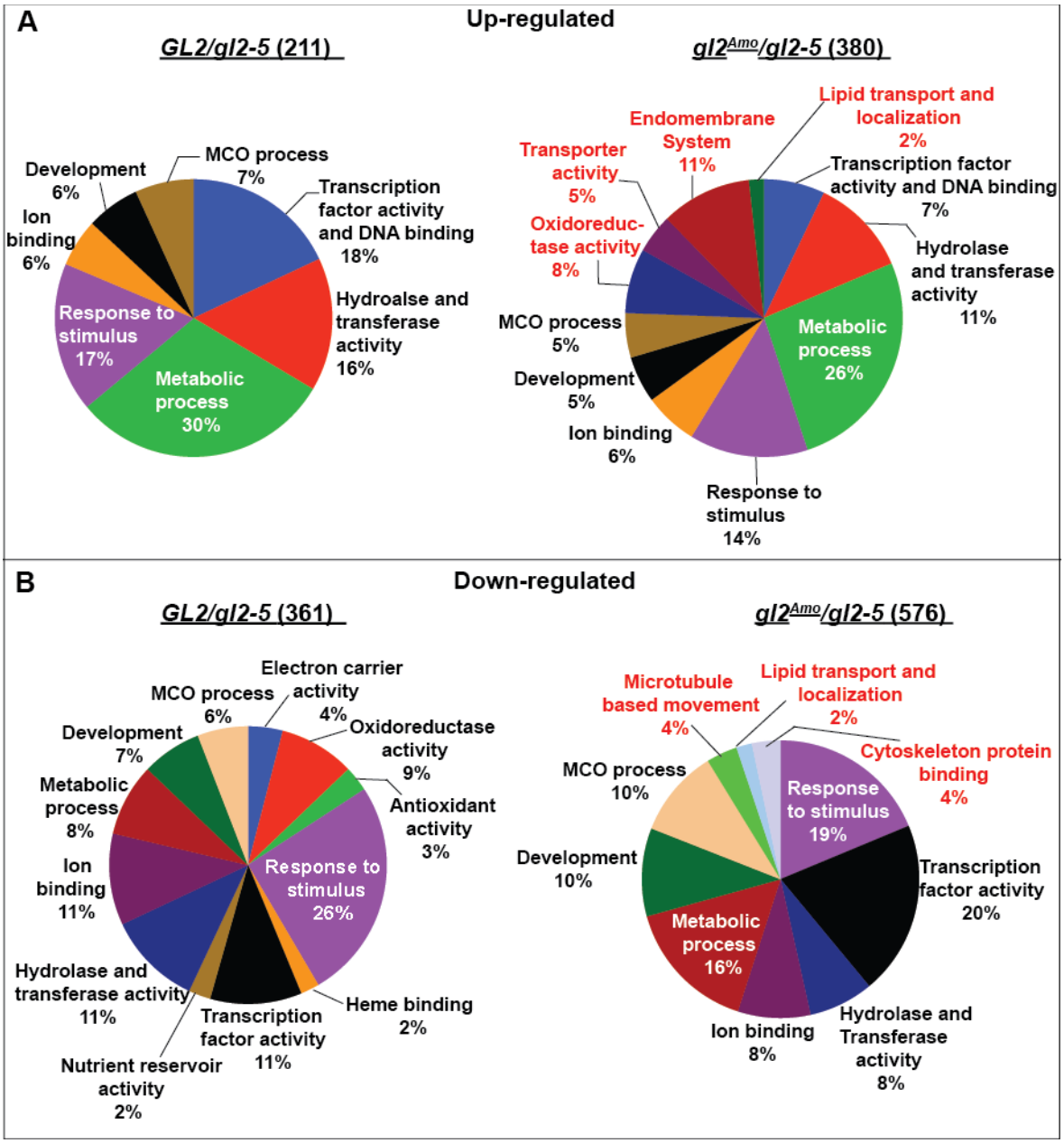


Figure B.3 Gene Ontology (GO) analysis of *gl2^{Amo}*.

Functional characterization of transcripts **(A)** upregulated or **(B)** downregulated in microarray experiment of *gl2^{Amo}* vs *gl2-5* null mutant seedlings. GO terms unique to gl2^{Amo} are highlighted. The number in parentheses represents the number of differentially expressed transcripts. GO term enrichment analysis was performed using the AgriGO Gene Ontology enrichment analysis tool (http://bioinfo.cau.edu.cn/agriGO/). MCO process, Multicellular Organismal Process

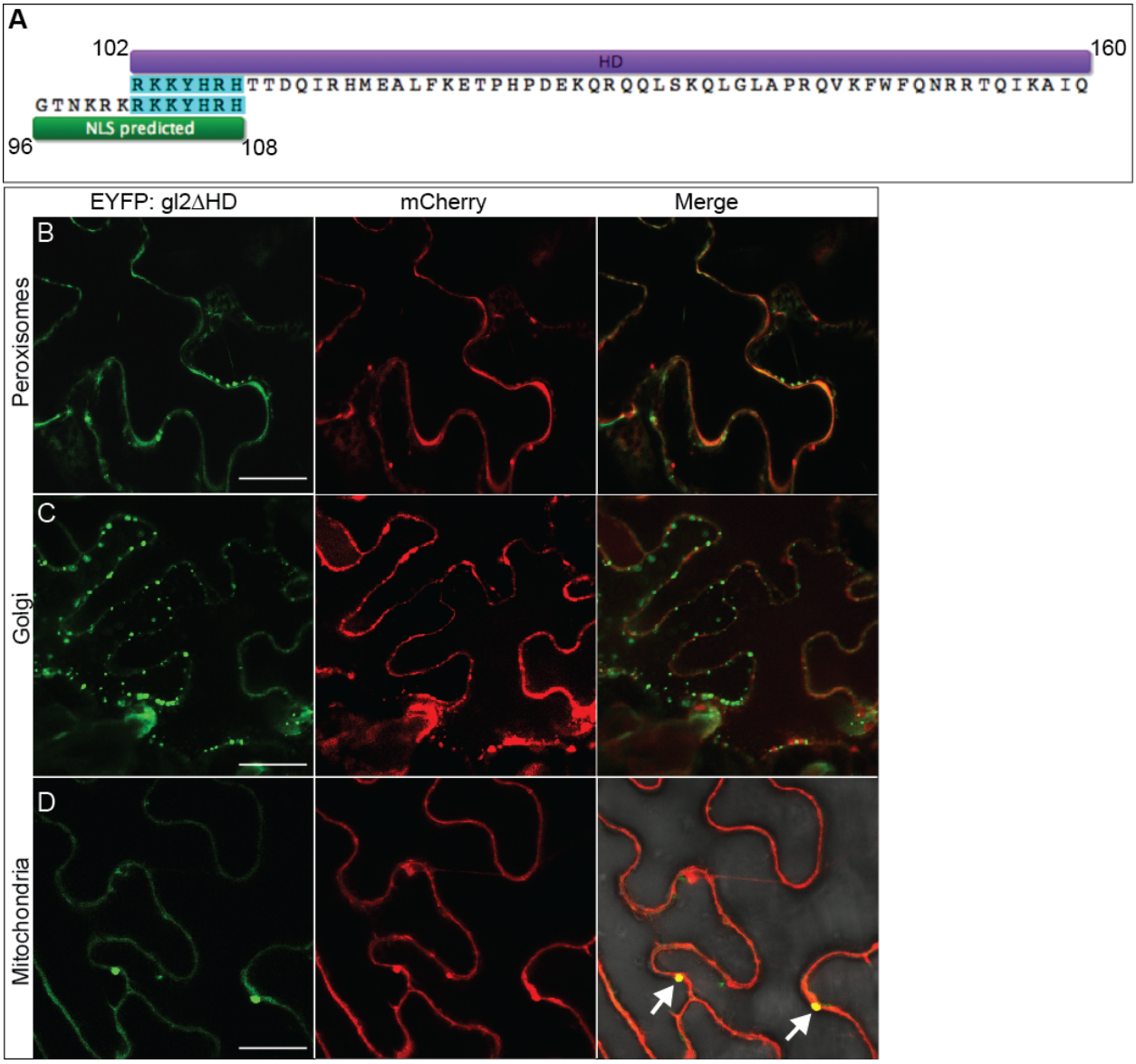


Figure B.4 Homeodomain (HD) is necessary and sufficient for transport into the nucleus.

(A) Nuclear localization signal (NLS) of GL2 (green) overlaps with the beginning of the homeodomain (purple). **(B)** Co-expression of EYFP:gl2ΔHD with different cellular markers in N. benthamiana. Confocal microscopy of N. benthamiana leaves transiently overexpressing EYFP:gl2ΔHD together with the red florescent protein (mCherry)-tagged markers for peroxisomes (**(A)**; peroxisome targeting sequence), the Golgi complex (**(B)**; soybean α-1,2-mannosidase I), or mitochondria (**(C)**; yeast cytochrome oxidase subunit IV). EYFP is shown in green (left panels), mCherry is in red (central panels). Merged images of green and red

fluorescent signals (right panels). Arrows (white) identify sites for which gl2ΔHD and mitochondrial marker overlaps. Bars = 100 μm

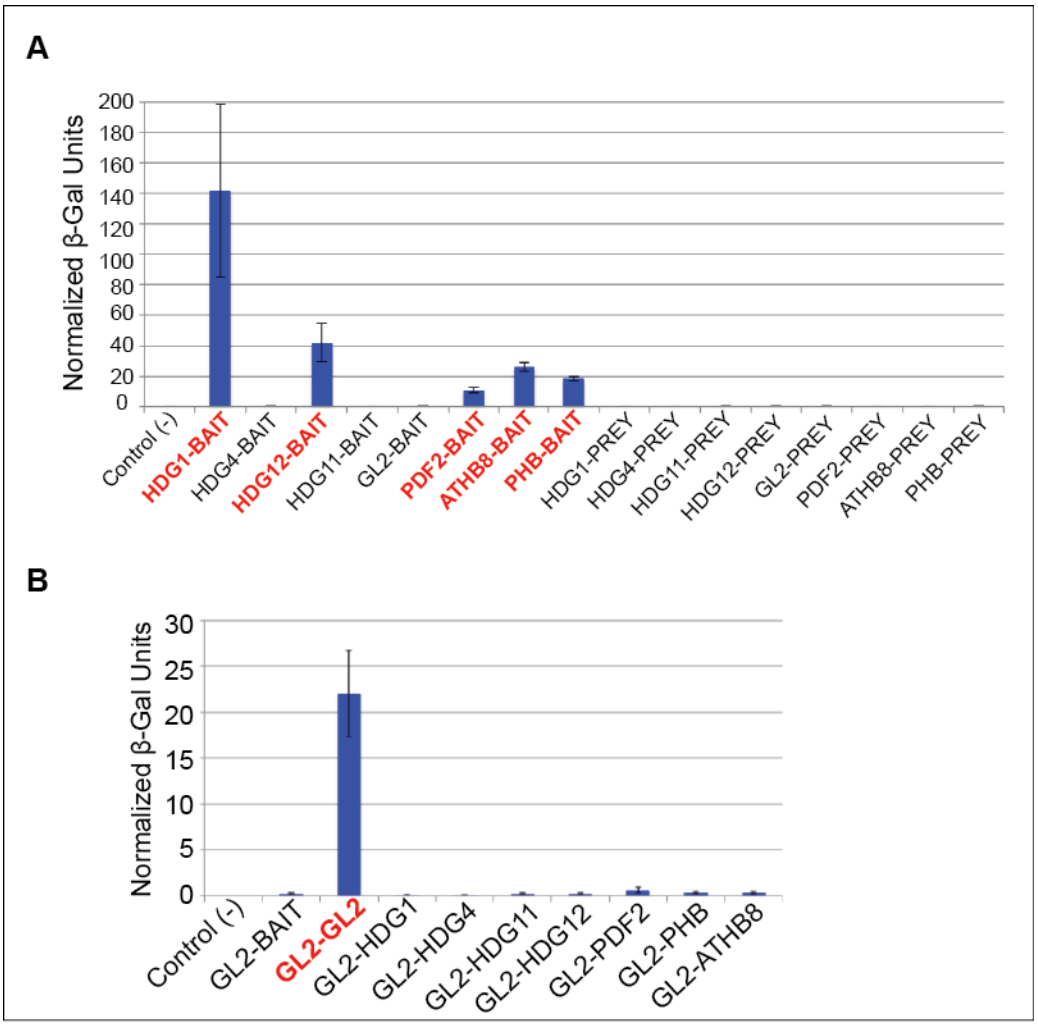


Figure B.5 GL2 does not autoactivate and exhibits only homodimerization in the Y2H assays.

(A) Quantitative yeast ONPG assays for GL2 with various HD-Zip transcription factors. Vector combinations used to co-transform yeast cells are given as bait/prey. Average β-galactosidase values were normalized to a positive control performed at the same time (pEXP^{TM}32/Krev1 and pEXP^{TM}22/RalGDS-wt, set to 100 β units). Combination of empty bait vector with GL2 fusion protein served as a negative control. Error bars indicate standard deviations for two independent transformants. HDG4, HDG12, and PDF2 do not exhibit autoactivation. It should be noted that HDG12 cDNA was found truncated, lacking a portion of

the START domain as well as C-terminal region. **(B)** Quantitative yeast ONPG assays for GL2 with class III and IV HD-Zip transcription factors. Error bars indicate standard deviations for three independent transformants. GL2 only interacts with itself (red).

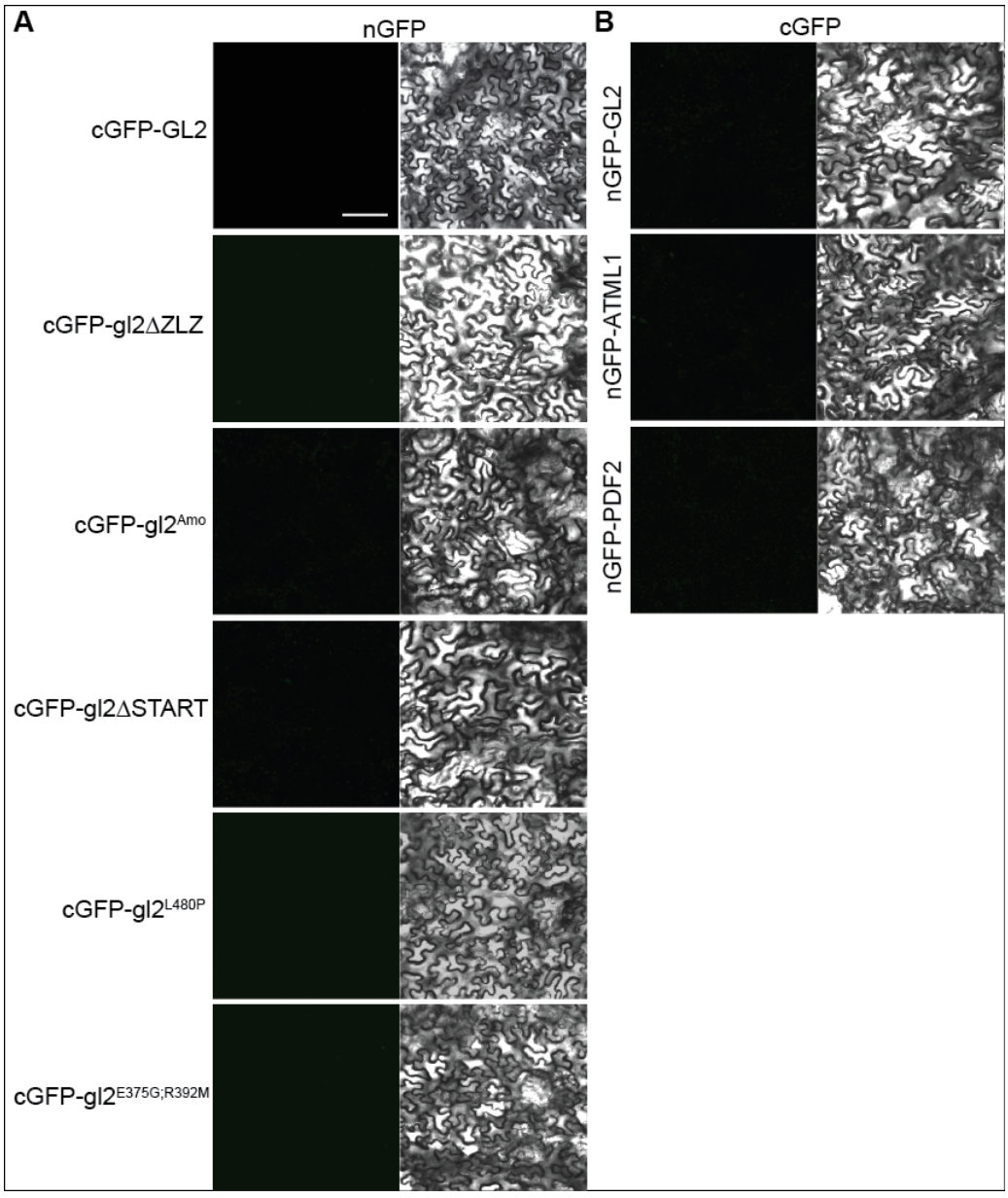


Figure B.6 BiFC negative assays.

Confocal microscopy of *N. benthamiana* leaves. No fluorescence was observed in control experiments coexpressing the indicated proteins with empty vectors **(A)** nGFP (N-terminal half of GFP) and **(B)** cGFP (C-terminal half of GFP)

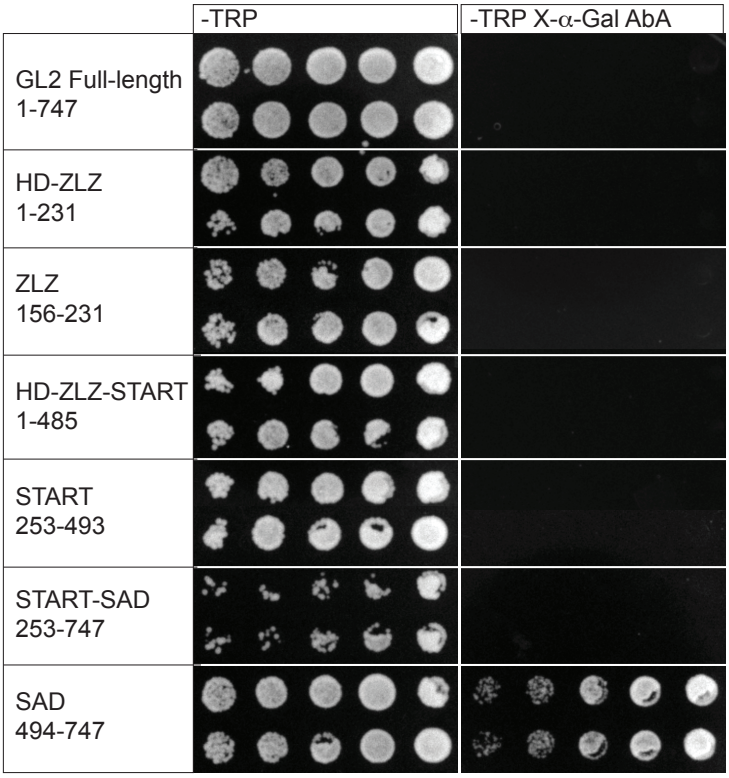


Figure B.7 The SAD domain of GL2 behaves as an activation domain in yeast.

Full-length GL2 and subdomains thereof were expressed in yeast as translational fusions to the GAL4 DNA-binding domain on a Bait plasmid that carries the *TRP1* gene. For each construct, two independent transformants are shown in four-fold serial dilutions on permissive (-TRP) (left panel) and selective (-TRP X-a-Gal AbA) (right panel) plates to assay for the reporter genes *AUR1-C* and *MEL1*. The only Bait plasmids that displayed autoactivation in yeast were those expressing the C-terminal 254 amino acid region of GL2 containing the SAD domain.

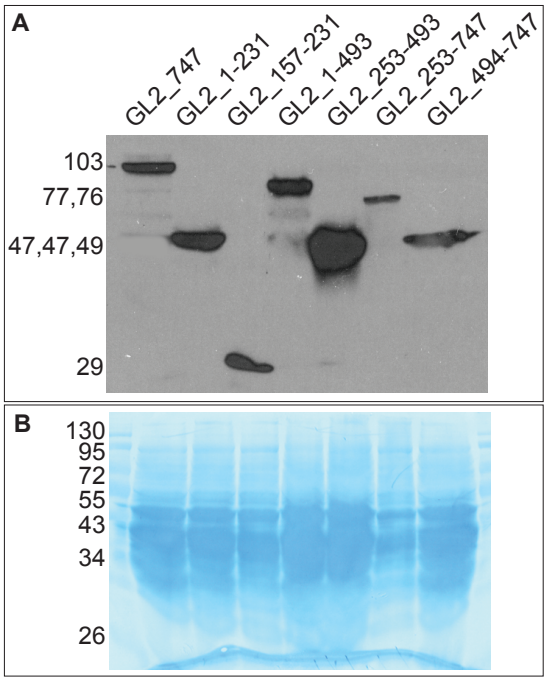


Figure B.8 Protein expression of Bait plasmids in yeast two-hybrid system.

(A) Western blot of fusions of GL2 full-length and subdomains there of to the GAL4 DNA Binding Domain in the pGBK-T7 Bait vector. The primary antibody was c-Myc (9E10) (1:40) and the secondary antibody was Goat Anti-Mouse IgG [HRP] (A00160, GenScript) (1:3000). Detection was with ECL Plus Western Blotting Data System. **(B)** A Coomasie stained image of the blot is shown as a loading control. The Pageruler protein ladder (Fermentas) was loaded to determine the approximate sizes of the proteins. The size of the GAL4 DBD-Myc fusion protein is 184 aa (21.1 kDa). 1-747, 83.2+21.1=103.3 kD, 1-231, 26.2+21.1=47.3 kD, 1-493, 55.5+21.1=76.6 kD, 253-493, 27.0+21.1=47.1, 494-747, 27.7+21.1=48.8 kD, 253-747, 54.6+21.1=75.7 kD, 157-231, 8.3+21.1=29.4 kD.

Appendix C - Purification of GST-tagged START domain

Optimization of the START domain expression and purification

To enable ligand-binding studies, soluble and purified protein is required. Expression constructs encoding START domains from seven of the 21 members of the HD-Zip transcription factor family from *Arabidopsis* were designed based on 2˚ structure predictions. START domain coding regions (~750 bp) were cloned as N-terminal GST fusions in Flexi Vector pFN2A (Promega) which also contains a TEV protease site for cleavage from the GST tag. In comparison to BL21 (DE3)pLysS and BL21(DE3)Arctic Express cells (Stratagene), *E. coli* strain BL21(DE3) pRARE proved most efficient for protein induction and solubility. This was not surprising as the analysis of the PDF2 START domain sequence revealed the presence of 15 rare arginine codons, including arginine double and triple repeats. Protein solubility was tested by centrifugation followed by SDS-PAGE of the supernatant versus pellet fractions. GST-tagged PDF2 gave the most promising levels of soluble protein after induction with 0.1 mM IPTG for 20 hours at 18˚C (Figure C.1 A, B). Thus, the GST-tagged PDF2-START protein was selected for purification. Western blotting using anti-GST antibody and mass spectrometry confirmed the identity of the protein (Figure C.1 C).

ANS binding assay to determine whether GST-tagged PDF2-START is properly folded

The fluorescent dye 8-anilino- 1-napthalenesulfonate (ANS) binds non- specifically to exposed hydrophobic regions in folding intermediates of proteins. It's fluorescence increases when bound proteins transition to folded states (Santambrogio and Grandori 2008). To test whether the GST- tagged PDF2-START domain is folded correctly and contains a hydrophobic binding site, we performed an ANS binding assay according to a published method (Vinayavekhin and Saghatelian 2011). ANS fluorescence yielded saturation binding, indicating that the GST-tagged PDF2-START protein is properly folded and has a hydrophobic binding cavity that can be saturated (Figure C.2). The fluorescence increased non-linearly, indicating a binding equilibrium. The data was fit to a one binding- site model with K_D values of 750μM, which is within the range previously reported for other folded proteins (Tagore et al. 2008)

GST: PDF2-START domain affinity purification

To enable structural studies, protein preparation was upscaled to 3L and purified under the conditions identified for optimal expression as described above. TEV protease cleavage to remove the GST tag was optimal after 24 hours at 23˚C in a Tris-HCl buffer with 10% Glycerol (Figure C.3 A). MALDI-TOF mass spectrometry (MS) indicated that cleavage produced free GST (~26 kDa) and the PDF2-START domain (~26 kDa) as expected. A higher molecular weight band of ~55 kDa was identified as *E. coli* GroEL. Following affinity purification with glutathione to remove the GST protein and Talon resin to remove the His-tagged TEV protease, we utilized fast protein liquid chromatography (FPLC) to separate the PDF2-START domain from co-purifying high molecular weight proteins including GroEL. After affinity purification, the sample was loaded on a Superdex^{TM}75 10/300 gel filtration column using the ÄKTA^{TM} FPLC system. FPLC profile showed two peaks (Figure C.3 B). However, mass spectrometry identified major peak as a GST protein, suggestive of protein aggregation or degradation post tag removal.

In vitro ligand binding studies with GST-tagged PDF2-START

The availability of soluble and correctly folded GST-tagged PDF2-START domain protein enabled *in vitro* ligand binding studies. Previously, using the immunoisolation experiments with GSV synthetic transcription factors expressed in yeast, numerous candidate protein-metabolite interactions for the plant START domains were identified (Schrick et al. 2014) . An in vitro binding assay using purified GST:PDF2-START protein was performed with the following candidates: Vitamin K1, α-tocotrienol (Vitamin E), eicosanoic acids, and a polar lipid mixture extracted from *Arabidopsis* seedlings, comprised mainly of phosphatidic acid (PA) and PC, followed by PE, PI, PG, and PS. GST alone and heat-denatured GST: PDF2-START was used as negative controls. The results indicated binding to polar lipids. Preferential binding of PDF2- START to PC, in comparison to PA was observed, and binding to other phospholipids was also detected (Figure C.4). However, GST alone negative control protein also displayed binding to several polar lipid species, making the interpretation of this experiment problematic.

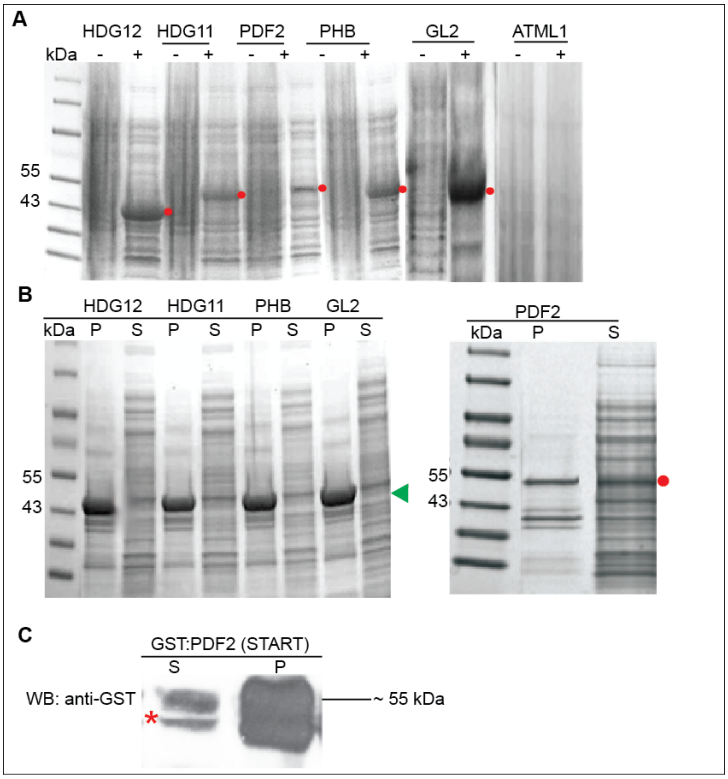


Figure C.1 Test expression of seven GST-tagged START domain constructs in *E.coli*.

(A) Coomassie-blue stained SDS-PAGE of un-induced (-) and induced (+) fractions (red) collected 20 hr post-induction using 0.1 mM IPTG. **(B)** Coomassie-blue stained SDS-PAGE of insoluble (P) and soluble (S) fractions of GST:HDG12-START; GST:HDG11-START; GST:PHB-START; GST:GL2:START, and GST:PDF2-START domain constructs. Green arrow indicates protein in pellet fraction. Red indicates PDF2-START recombinant protein in soluble fraction. **(C)** Western blot confirms identity of 55 kDa GST:PDF2 (START) protein. As shown, the protein is present in both pellet (asterisk) and soluble fraction.

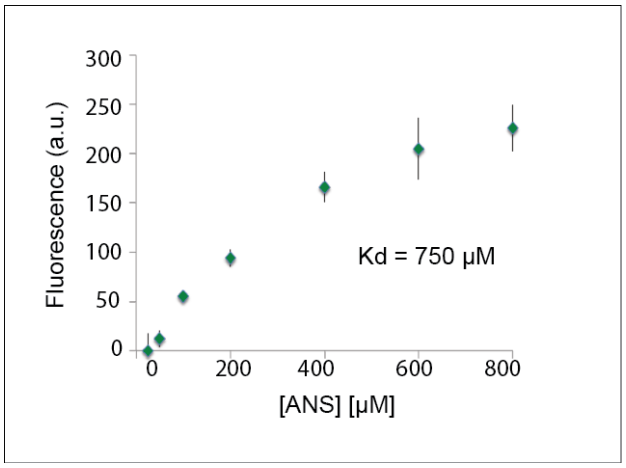


Figure C.2 ANS (8-anilino-1-napthalenesulfonate) binding assay indicates that the GST-tagged PDF2 START is properly folded.

START domains bind ANS with saturation kinetics, indicating that the proteins are properly folded. Binding to ANS was measured as an increase in fluorescence at emission/excitation wavelengths of 405/460 nm. Error bars show standard deviations for three technical replicates.

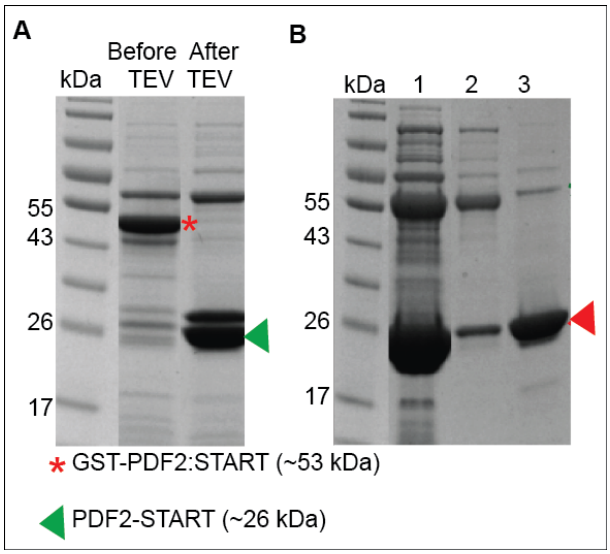


Figure C.3 Expression and large-scale purification of GST-tagged PDF2 START domain.

(A) Purification of PDF2-START before and after TEV protease cleavage. **(B) 1;** Protein after removal of TEV protease and GST; FPLC fractions: **2;** minor and **3;** major peak having PDF2-START. GST tag (~ 26 kDa) (red arrowhead).

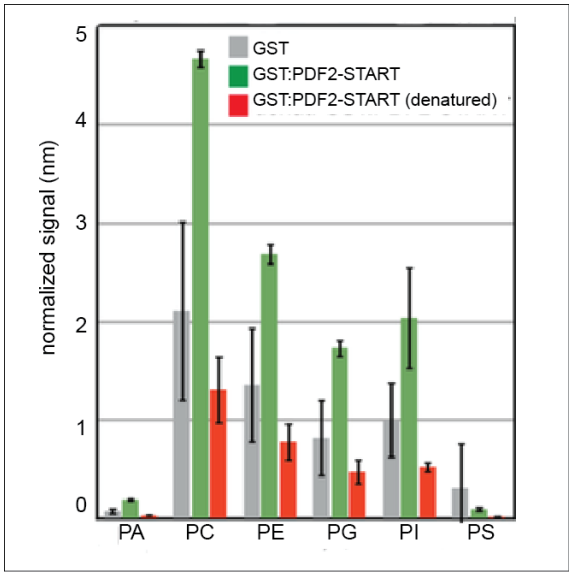


Figure C.4 *In vitro* ligand binding assay showing binding of polar lipids to GST:PDF2-START.

Binding to phospholipid PC appears strongest, followed by PE, PG, and PI. Binding is detected for negative controls (GST and GST:PDF2-START (denatured)) although to a lower extent

Appendix D - HD-Zip Proteins GL2 and HDG11 Have Redundant Functions in Arabidopsis Trichomes, and GL2 Activates a Positive Feedback Loop via MYB23

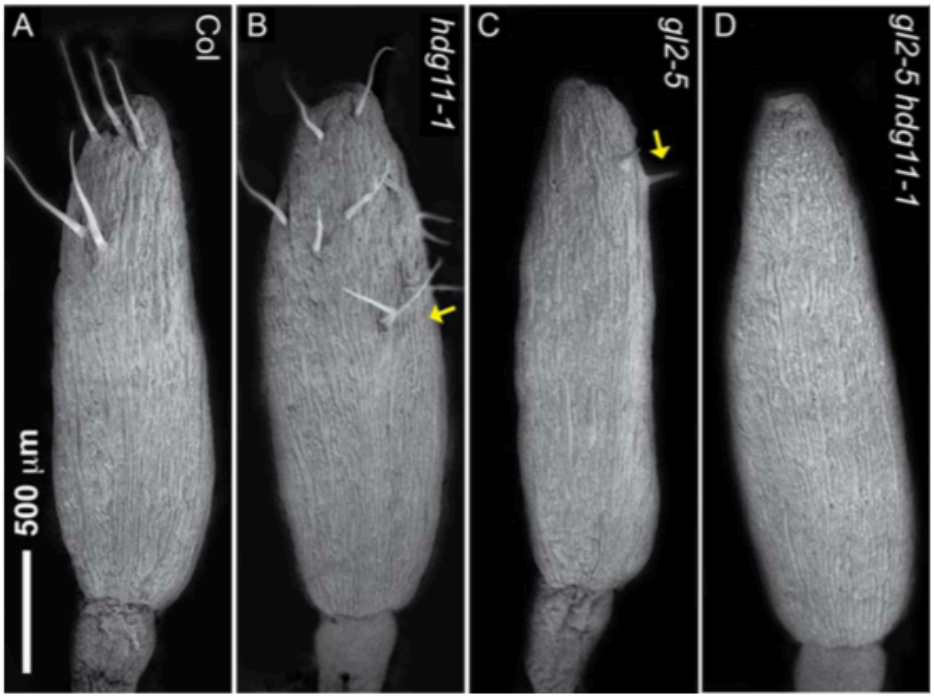


Figure D.1 Trichomes on sepals from *Arabidopsis* flowers.

(A-D) Scanning electron micrographs (SEM) of sepals. In comparison to **(A)** wild-type Col in which unbranched trichomes form on sepals, **(B)** *hdg11-1* sepals display both unbranched and branched trichomes (arrow). **(C)** *gl2-5* mutant sepals display fewer trichomes that are aborted in differentiation (arrow), while **(D)** *gl2-5 hdg11-1* double mutants lack visible trichomes.

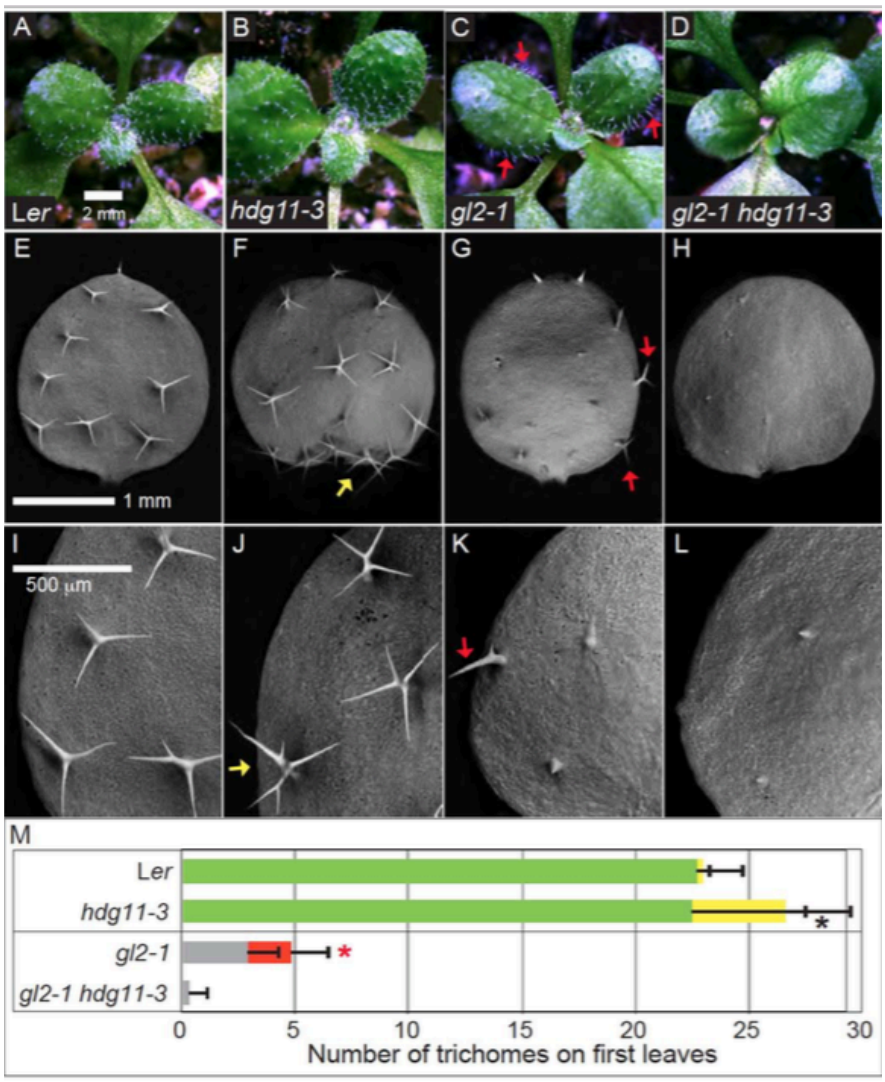


Figure D.2 *gl2-1 hdg11-3* double mutants from the L*er* ecotype exhibit enhanced trichome differentiation defects.

(A-D) Rosettes from 3-week-old plants. **(A)** L*er* wild-type and **(B)** *hdg11-3* plants display evenly-spaced trichomes on young leaves. **(C)** *gl2-1* plants exhibit fewer trichomes. Arrows mark trichomes on leaf margins. **(D)** *gl2-1 hdg11-3* plants have glabrous leaf surfaces. **(E-L)** Scanning electron micrographs of first leaves reveal trichome morphology details. **(E,I)** L*er* wild-type trichomes exhibit a maximum of 3-4 branches, while **(F,J)** *hdg11-3* leaves exhibit some trichomes with ectopic branching (arrows). **(G,K)** *gl2-1* leaves exhibit trichome branching (arrows) at leaf edges. In comparison, **(H, L)** *gl2-1 hdg11-3* trichomes appear smaller and undifferentiated. **(M)** Quantification of trichomes and trichome branching on first leaves as described in Figure 5.2 U. Positive error bars indicate standard deviations for n > 20 plants. Asterisks indicate significant differences in trichome branching (Two- tailed *t*-test, P < 0.00001).

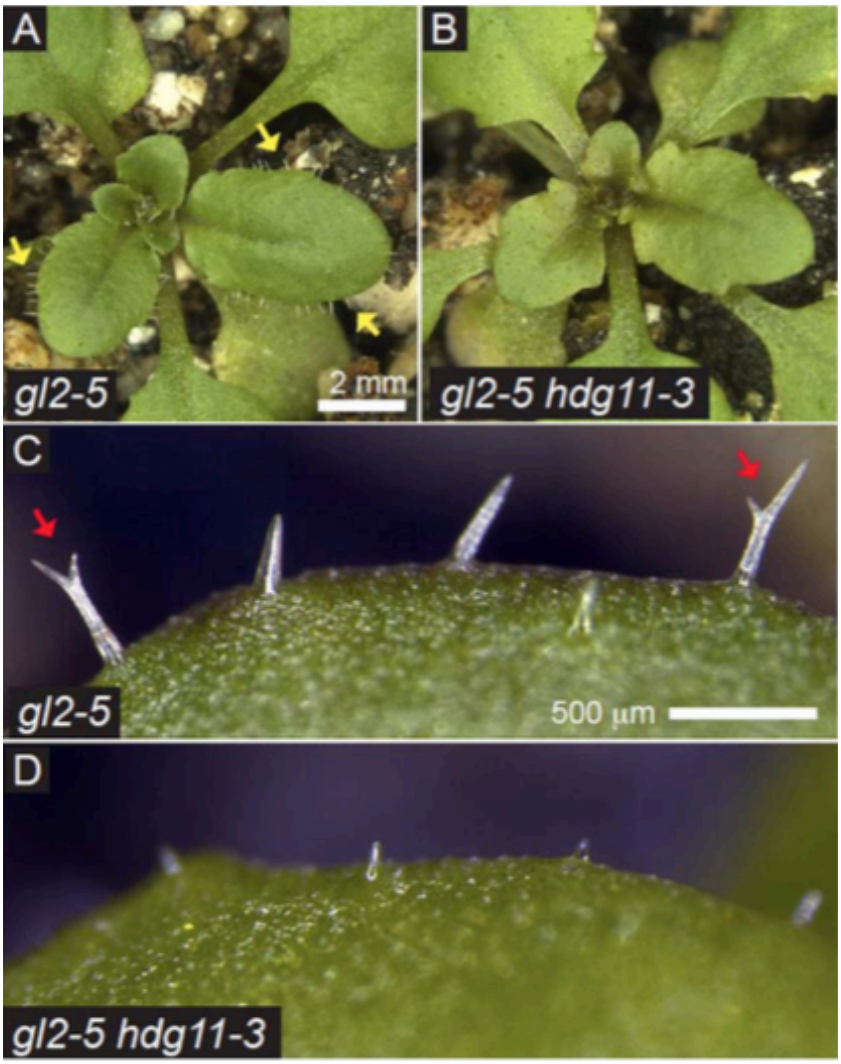


Figure D.3 Trichome phenotype of *gl2-5 hdg11-3* mutants.

F3 progeny from a cross between *gl2-5* plants harboring the *ProGL2:EYFP:GL2* transgene and *hdg11-3* plants were analyzed for phenotypes. *gl2-5* single mutant segregants from an F2 line lacking EYFP expression were compared to the *gl2-5 hdg11-3* double mutants. **(A, B)** Rosette phenotypes of **(A)** *gl2-5* and **(B)** *gl2-5 hdg11-3* reveal trichomes on leaf margins of *gl2-5* plants (arrows) but few or no trichomes on leaf margins of the double mutant. **(C, D)** Magnification of leaf margins shows that **(C)** *gl2-5* plants display some trichomes with branching (red arrows), while **(D)** *gl2-5 hdg11-3* leaves display smaller trichomes that lack branching.

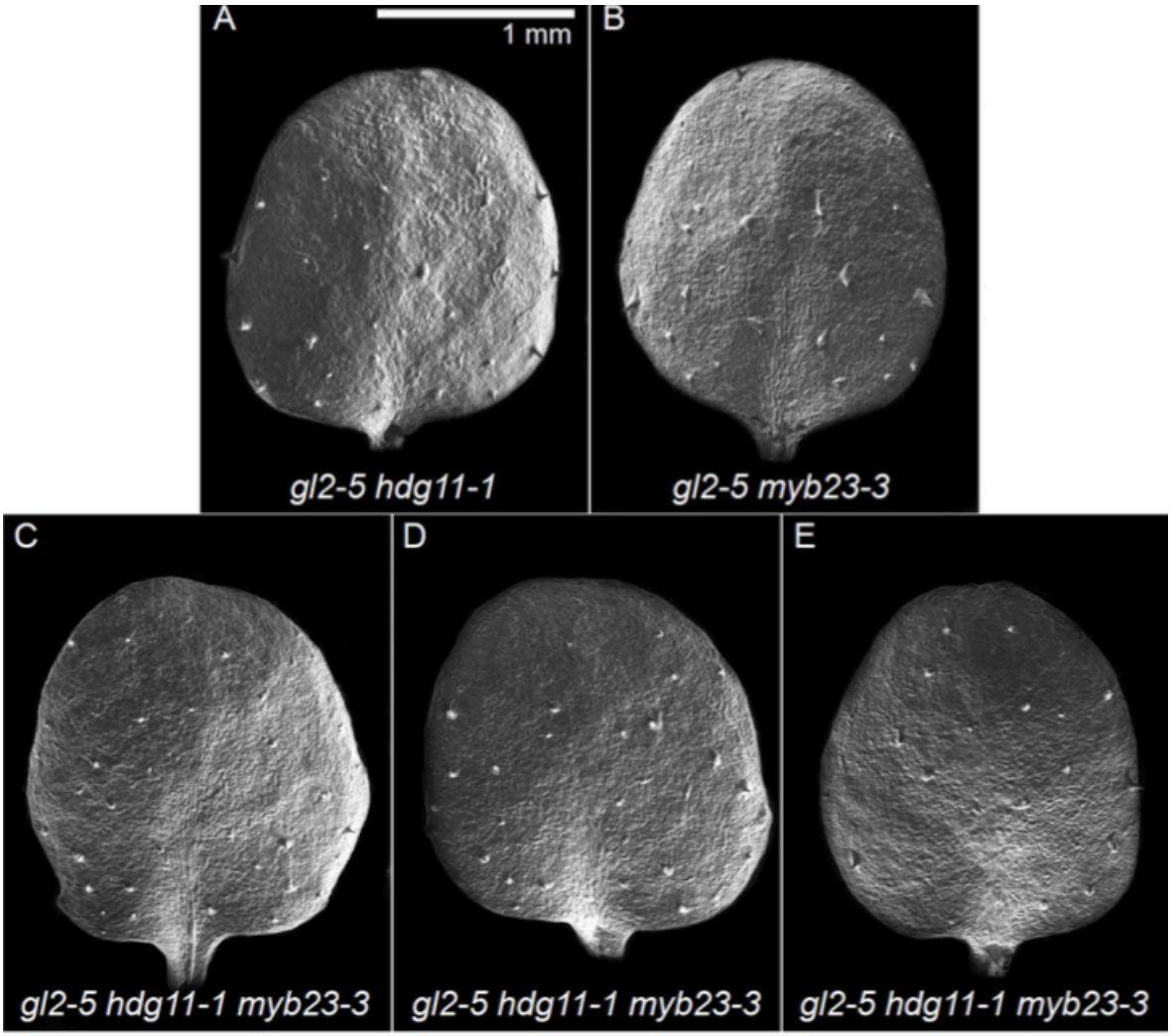


Figure D.4 Trichomes on first leaves of the *gl2-5 hdg11-1 myb23-3* triple mutant. **(A-E)** Scanning electron micrographs of first leaves. The double mutants **(A)** *gl2-5 hdg11-1* and **(B)** *gl2-5 myb23-3* display trichome differentiation defects that are similar to those observed for **(C-E)** *gl2-5 hdg11-1 myb23-3* triple mutants.

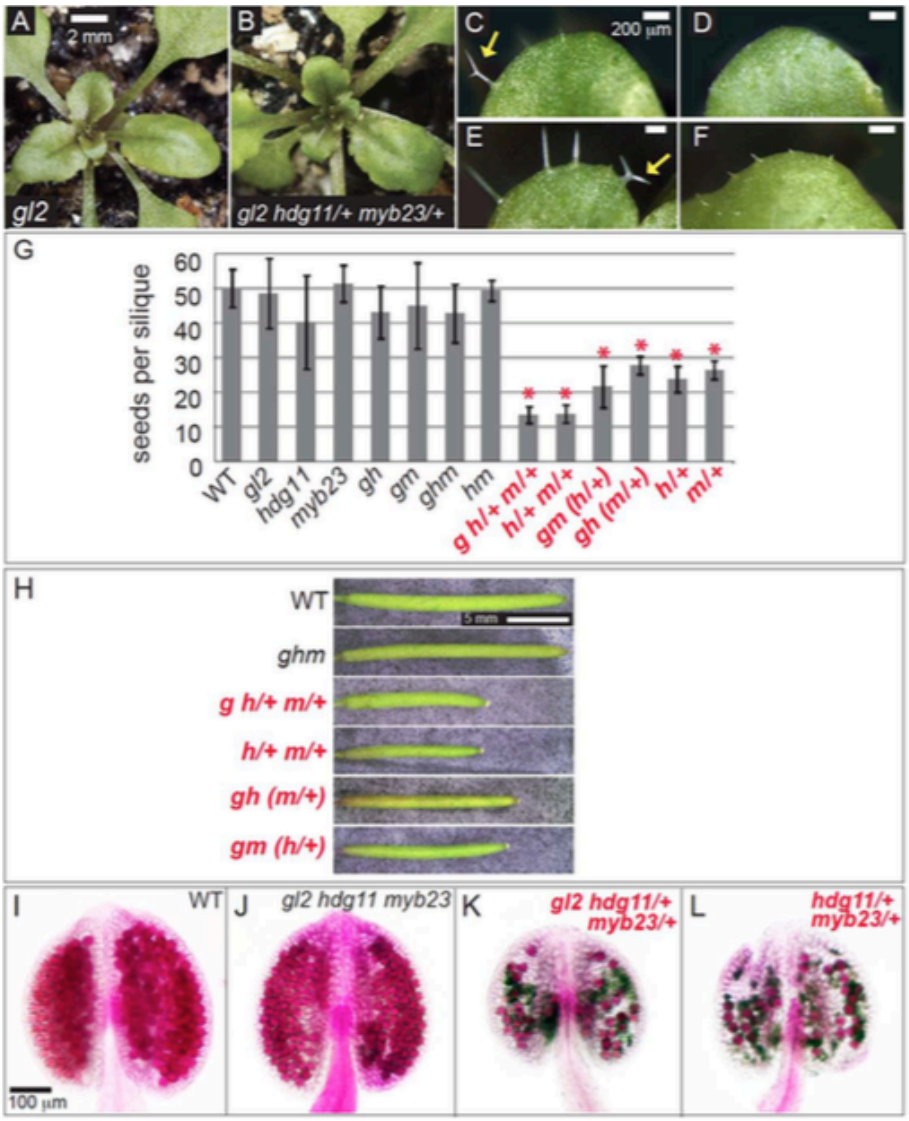


Figure D.5 Developmental defects of plants that are heterozygous for *hdg11-1* and/or *myb23-3*.

(A-F) Trichome phenotypes of *gl2-5* plants heterozygous for *hdg11-1* and *myb23-3* are similar to those of the *gl2-5 hdg11-5* and *gl2-5 myb23-3* double mutants. **(A,C,E)** *gl2-5* plants exhibit both branched and unbranched trichomes on leaf margins. **(C,E)** Arrows indicate branched trichomes. **(B,D,F)** *gl2-5* plants heterozygous for *hdg11-1* and *myb23-3* exhibit leaves with fewer trichomes that are unbranched. **(G)** Seeds per silique were counted for 15 siliques from Col wild-type (WT), *gl2-5* (*g*), *hdg11-1* (*h*), *myb23-3* (*m*), and double mutant and triple mutant combinations, as well as plants heterozygous for *hdg11-1* (*h*/+) and/or *myb23-3* (*m*/+). Standard deviations are indicated by error bars. Asterisks mark significant differences from WT (Two-tailed *t* test, $P < 0.0001$). **(H)** Mature siliques from Col WT, the triple mutant and various

combinations of *hdg11-1/+ myb23-3/+* with and without the *gl2-5* mutation are shown. **(I-L)** Anthers were dissected from buds, fixed in Carnoy's solution, and stained as in (Peterson, 2010). Pollen was imaged with a 10x 0.3NA objective on an Olympus BX51WI microscope using a Canon T3i digital camera and DSLR Remote Pro software. **(I)** Col WT and **(J)** *gl2 hdg11 myb23* anthers exhibit viable pollen as indicated by magenta-red staining. In contrast, anthers from **(K)** *gl2-5 myb23-3/+ hdg11- 1/+* and **(L)** *myb23-3/+ hdg11-1/+* plants show fewer viable pollen grains in addition to aborted pollen grains indicated by blue-green staining.

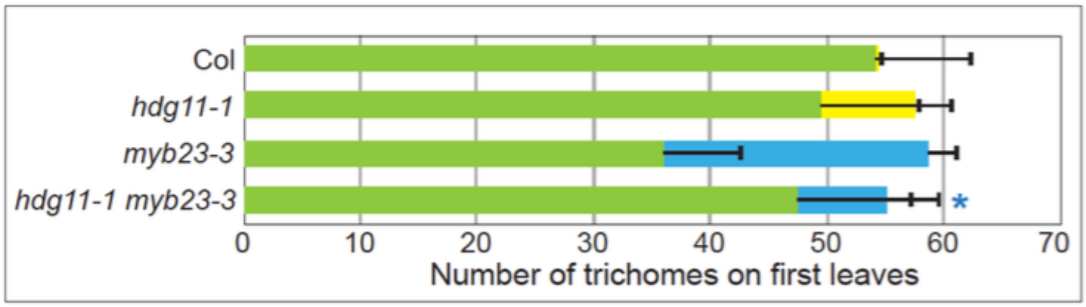


Figure D.6 Intermediate trichome phenotype of *hdg11-1 myb23-3* double mutants. Quantification of trichomes on first leaves. Green bars indicate the number of trichomes with 3-4 branches. Yellow bars show the number of trichomes with greater than four branches. Blue bars indicate the number of abnormal trichomes with fewer than three branches. Positive error bars indicate standard deviations for n > 20 plants, and the blue asterisk points to a significant decrease in the number of abnormal trichomes with fewer than three branches in the *hdg11-1 myb23-3* double mutant in comparison to the *myb23-3* single mutant (Two-tailed *t* test, $P < 0.00001$).

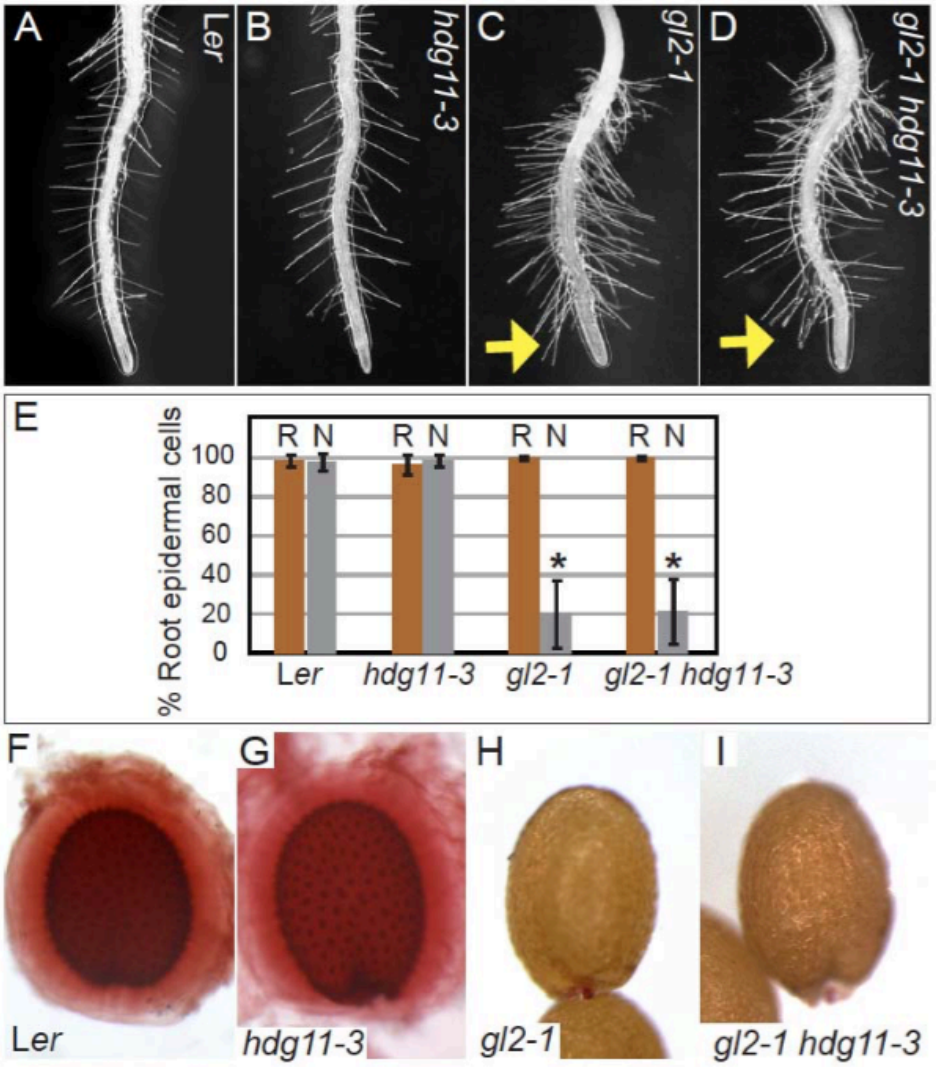


Figure D.7 Root hair density and seed mucilage phenotypes in the L*er* ecotype.

(A) L*er* wild-type and **(B)** *hdg11-3* exhibit normal root hair formation in contrast to **(C)** *gl2-1* single and **(D)** *gl2-1 hdg11-3* double mutants, which display excess root hair formation. Yellow arrows indicate abnormal root hair formation near the tips of the roots in *gl2* mutants. **(E)** Quantification of root epidermal cells. Percentages of trichoblast cells in root hair cell files (R = brown bars), and atrichoblast cells in non root hair cell files (N = grey bars) are indicated (n > 20). Asterisks mark significant differences from L*er* wild-type (Two-tiered *t* test, P > 0.0001). (F-H) Seeds were stained with ruthenium red to detect mucilage production upon imbibition. **(F)** L*er* wild-type and **(G)** *hdg11-3* seeds display mucilage in contrast to **(H)** *gl2-1* single and **(I)** *gl2-1 hdg11-3* double mutants which lack mucilage.

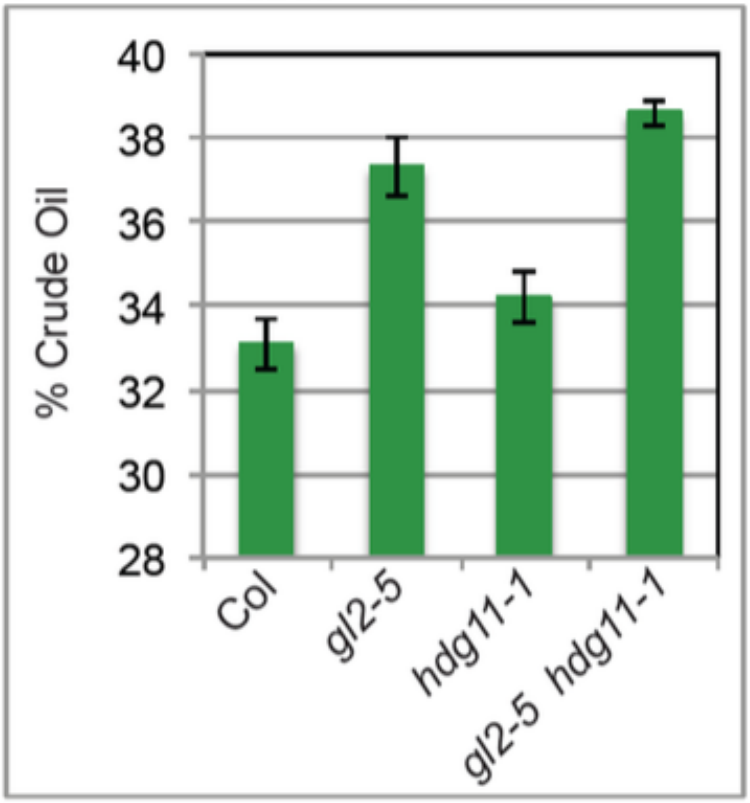


Figure D.8 Crude oil levels in *gl2* single and *gl2 hdg11* double mutant seeds are similar.

In comparison to wild-type Col, *gl2* seeds exhibit higher percentages of crude oil. However, oil levels of *hdg11-1* seeds are similar to those of the wild type, and *gl2-5 hdg11-1* double mutant seeds are comparable to *gl2-5* seeds. A representative experiment from two biological replicates is shown with standard deviations for two 500-mg batches of seeds. The AOAC (Association of Analytical Communities) method 920.39 was used to perform quantitative oil analysis.

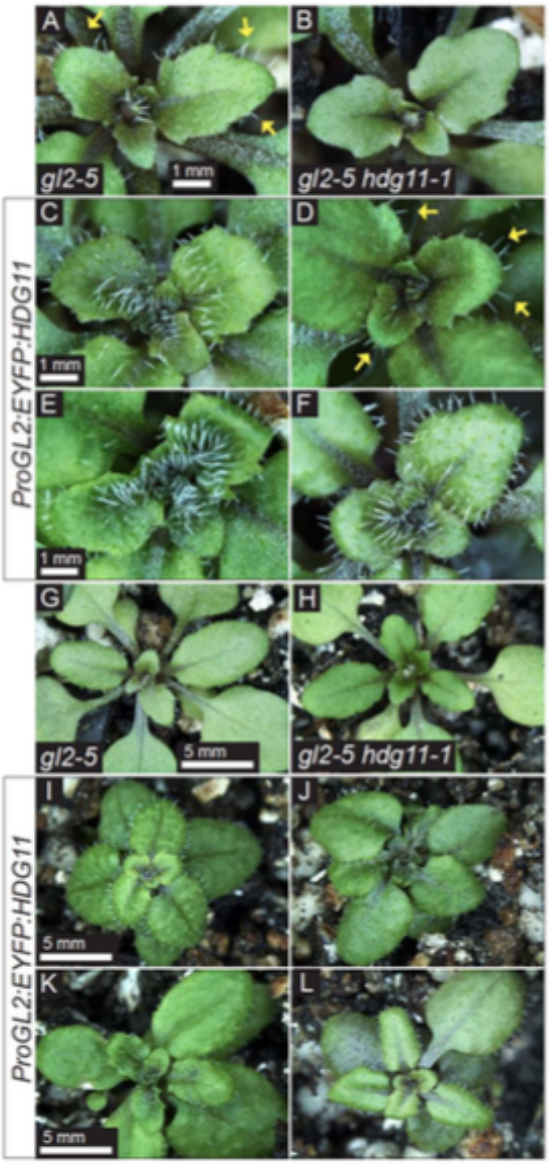


Figure D.9 Leaf and rosette phenotypes of *gl2* and *gl2 hdg11* mutant lines expressing *ProGL2:EYFP:HDG11*.

(A) *gl2-5* rosette leaves exhibit trichomes at leaf margins (yellow arrows). **(B)** *gl2-5 hdg11-1* leaves appear largely devoid of trichomes. **(C-E)** *gl2-5* mutants and **(D,F)** *gl2-5 hdg11-1* mutants expressing *ProGL2:EYFP:HDG11* display increased trichome formation on leaves. The trichomes are typically unbranched. **(D)** *gl2-5 hdg11-1* mutants expressing *ProGL2:EYFP:HDG11* resemble *gl2-5* single mutants (yellow arrows), or **(F)** exhibit increased trichome differentiation phenotypes similar to those conferred by *ProGL2:EYFP:HDG11* in the *gl2-5* background **(C,E)**. **(G)** *gl2-5* and **(H)** *gl2-5 hdg11-1* rosettes are shown. **(I, K)** *gl2-5* mutants and **(J,L)** *gl2-5 hdg11-1* mutants transformed with *ProGL2:EYFP:HDG11* exhibit

transient dwarf rosette phenotypes that appear either **(I,J)** symmetric or **(K,L)** notably asymmetric.

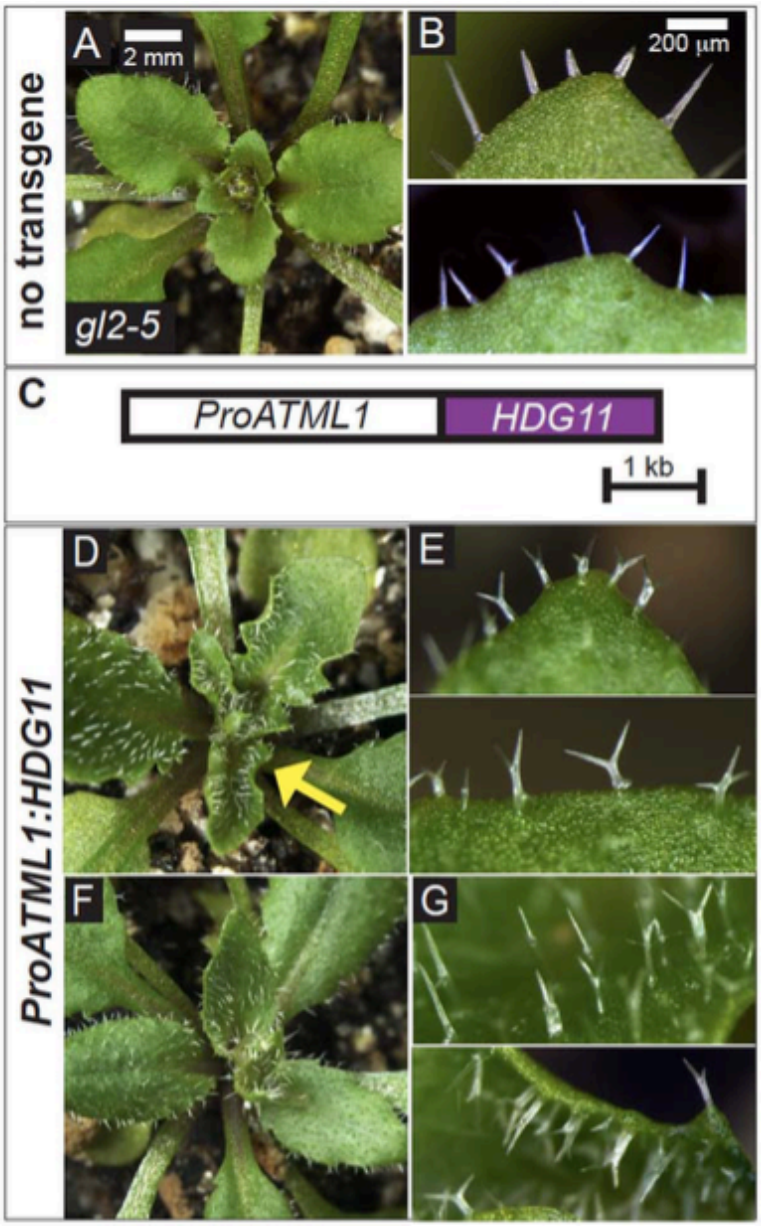


Figure D.10 Expression of *HDG11* under the *ATML1* epidermis-specific promoter induces trichome differentiation in *gl2* mutants.

(A,B) *gl2-5* rosettes exhibit trichomes at the leaf margins but otherwise appear glabrous. **(C)** *ProATML1:HDG11* construct. **(D,F)** Two independent transformants that express *HDG11* under the *ATML* promoter are shown at the rosette stage. **(D)** Leaf curling (yellow arrow) and/or **(F)** pointy leaves were observed for several of the lines. **(E,G)** Magnification of leaf trichomes

from *ProATML1:HDG11* transformants in comparison to **(B)** untransformed *gl2-5* plants indicates a partial rescue of the trichome phenotype.

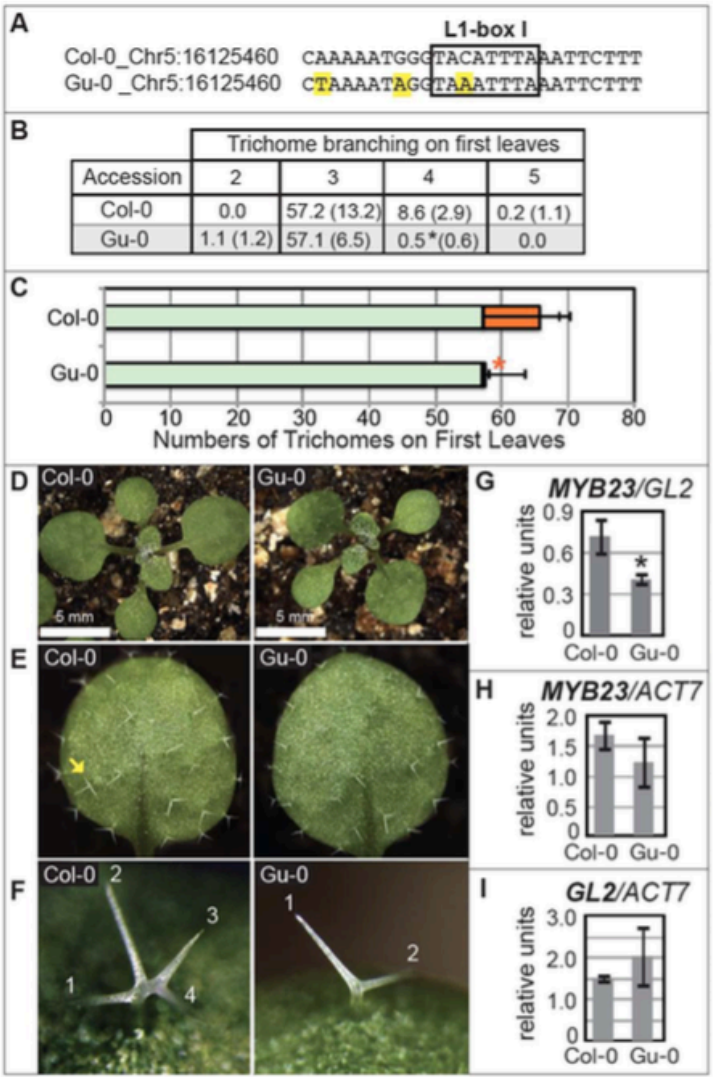


Figure D.11 A natural polymorphism in the *MYB23* L1-box I from the Gu-0 accession is correlated with reduced trichome branching.

(A) Alignment of genomic sequence from the *MYB23* promoter in Col-0 and Gu-0 accessions. Single nucleotide polymorphisms (SNPs) are indicated for Gu-0 (yellow) and the L1-box 1 is boxed. **(B)** Quantification of trichomes and trichome branching on first leaves. Percentages of trichomes with 2, 3, 4, or 5-branches are shown (n > 20) with standard deviations in parentheses. An asterisk marks a significant decrease in the number of 4- branched trichomes in Gu-0 in comparison to Col-0 (Two-tiered *t* test, $P < 0.00001$). **(C)** Graphical representation of the data in (B) illustrates numbers of 3-branched trichomes (green bars) and 4-branched trichomes (orange bars). **(D)** The rosettes of 10-day-old plants and **(E)** first leaves from Col-0

and Gu-0 accessions show similar leaf shapes. Yellow arrow marks a 4-branched trichome in Col-0. **(F)** Four-branched and 2-branched trichomes are shown for Col-0 and Gu-0, respectively. **(G-H)** Quantitative real-time PCR was performed with cDNA from 14-day-old seedling shoots. Data represent 3 biological replicates and normalized units are plotted on the Y-axis. Relative expression levels of **(G)** *MYB23* in comparison to *GL2*, and **(H)** *MYB23* and **(I)** *GL2* in comparison to the *ACT7* reference gene for the Col-0 and Gu-0 accessions. In **(G),** an asterisk marks a significant reduction of *MYB23* levels in Gu-0 in comparison to Col-0 (Two-tiered *t* test, $P < 0.05$).

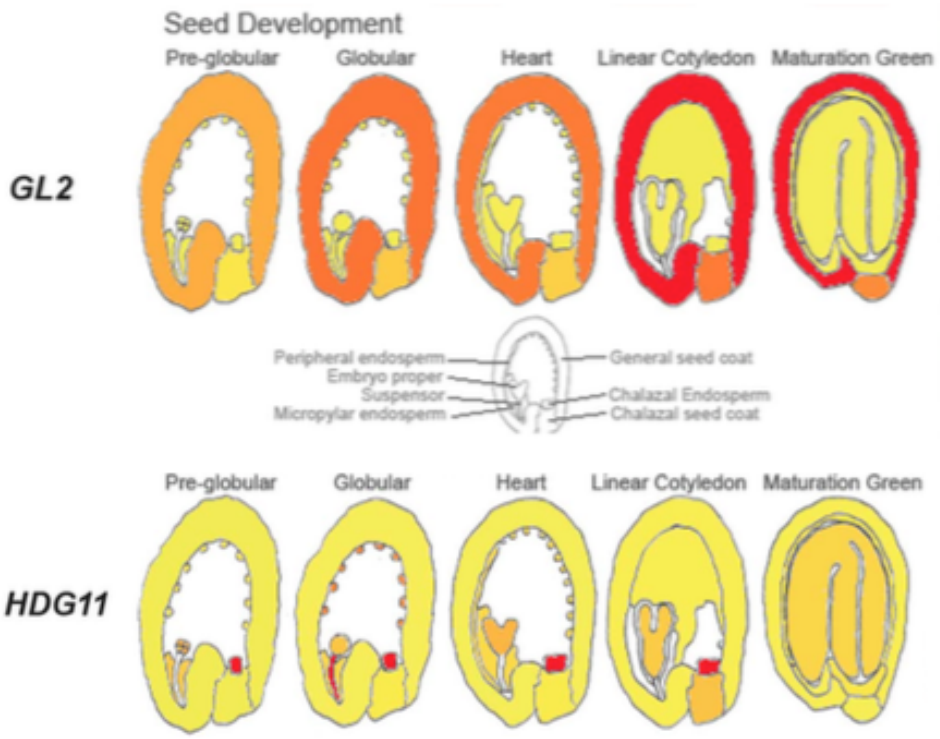


Figure D.12 Comparison of *GL2* and *HDG11* expression in the developing seed.

These data are derived from the Bio-Analytic Resource (BAR) *Arabidopsis* eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). Developing seeds from *Arabidopsis* ecotype Wassilewskija (Ws-0) were harvested from reproductive plants grown under continuous light. Subdomains of the immature seeds were separated by laser capture microdissection, followed by RNA extraction and amplification prior to hybridization on the GeneChip *Arabidopsis* ATH1 Genome Array (Affymetrix, Inc.). The data were generated by the Bob Goldberg and John J. Hanada labs (Le et al., 2010). Images are adapted from Meryl Hashimoto.

Table D.1 Pairwise alignments between the homeodomains from GL2/At1g79840 and each of the 15 other class IV HD-Zip family members.

Protein- protein BLAST was performed with compositional score matrix (BLOSUM62) adjustment. The number of amino acids (aa) in each pairwise alignment is indicated. No gaps were found in the alignments. The HD-Zip derived sequences are sorted by % aa identity and % similarity to GL2/At1g79840. HDG11/At1g73360, HDG5/At5g46880 and PDF2/At4g04890 exhibit the largest degree of aa identity and similarity. HD-Zip transcription factors expressed in trichomes are indicated by an asterisk.

HD Domain	aa	% Identity	% Similarity
*HDG11/At1g73360	57	89	74
HDG5/At5g46880	59	88	75
PDF2/At4g04890	58	87	78
ATML1/At4g21750	57	85	65
*HDG2/At1g05230	56	85	70
*HDG12/At1g17920	57	84	68
ANL2/At4g00730	57	84	68
HDG4/At4g17710	59	83	71
HDG1/At3g61150	57	82	67
HDG7/At5g52170	57	80	61
HDG3/At2g32370	56	78	68
HDG9/At5g17320	57	75	65
HDG8/At3g03260	57	75	60
HDG10/At1g34650	50	74	54
HDG6/FWA/At4g25540	52	63	42

Table D.2 Trichome quantification on the first two leaves of *Arabidopsis* plants.

See Figure 5.5 U, Appendix Figure D.2 M, Figures 5.3 M, 5.3 Q, 5.5 T, 5.6 H, and Appendix Figure D.6 for graphical representations of the data. Standard deviations for n>20 plants are shown in parentheses. Significant differences for double mutants versus *gl2* single mutants are indicated by a single asterisk, significant differences between control and *HDG11* or *GL2* transgenes are indicated by double asterisks, and a significant difference between *myb23-3* and the *hdg11-1 myb23-3* double mutant is indicated by triple asterisks (Two-tiered *t* test, P < 0.00001).

Genotype	Ave. # trichomes 1 branch (*gl2*) 1-2 branch (*myb23*)	Ave. # trichomes 2-4 branch 3 branch (*myb23*)	Ave. # trichomes >4 branches	Ave. # total trichomes	Ave. % ectopic branched trichomes >4 branches	Ave. % branched trichomes >1 branch for *gl2* genotypes 1-2 branch (*myb23*)
Col wild-type	0	67.5 (7.5)	0.3 (0.9)	67.8 (7.4)	0.4 (1.3)	NA
hdg11-1	0	49.6 (9.6)	7.1 (3.0)	56.7 (11.0)	12.6 (5.3)	NA
gl2-5	6.5 (3.1)	0.8 (1.4)	0	7.3 (3.6)	0	11.0 (19.2)
gl2-5 hdg11-1	3.9 (3.4)	0*	0	3.9 (3.4)	0	0*
Col wild-type	0	68.8 (9.8)	NA	68.8 (9.8)	NA	NA
myb23-3	24.8 (5.6)	45.6 (14.7)	NA	70.3 (15.9)	NA	35.3 (8.0)
gl2-5 myb23-3	1.0 (1.2)	0*	0	1.0 (1.2)	0	0*
gl2-5	5.7 (3.3)	1.2 (0.9)	0	6.9 (3.3)	0	19.0 (13.0)
gl2-5 hdg11-1	3.4 (2.5)	0*	0	3.4 (2.5)	0	0*
gl2-5 myb23-3	1.4 (1.5)	0*	0	1.4 (1.5)	0	0*
gl2-1 hdg11-3 myb23-3	1.5 (1.6)	0*	0	1.5 (1.6)	0	0*
Ler wild-type	0	22.7 (2.2)	0.27 (0.46)	23.0 (2.2)	1.2 (2.0)	NA
hdg11-3	0	22.5 (4.0)	4.0 (3.0)	26.5 (5.1)	15.1* (11.3)	NA
gl2-1	2.9 (1.5)	1.9 (1.8)	0	4.8 (2.8)	0	39.6 (37.5)
gl2-1 hdg11-3	0.3* (0.9)	0*	0	0.3 (0.9)	0	0*
hdg11-1 + ProGL2:EYFP:hdg11Δ	0	55.3 (6.6)	10.9 (3.1)	66.2 (6.4)	16.5 (4.7)	NA
hdg11-1 + ProGL2:EYFP:HDG11	0	56.2 (8.1)	0.3** (0.5)	56.5 (8.1)	0.6** (0.9)	NA
gl2-5 + ProGL2:EYFP:hdg11Δ	5.5 (2.6)	2.0 (1.8)	0	7.5 (3.5)	0	26.7 (24)
gl2-5 + ProGL2:EYFP:HDG11	12.8** (3.9)	12.0** (3.8)	0	24.8** (3.9)	0	48.3 (15.3)
gl2-5 hdg11-1 + ProGL2:EYFP:hdg11Δ	0.4 (0.6)	0	0	0.4 (0.6)	0	0
gl2-5 hdg11-1 + ProGL2:EYFP:HDG11	7.8** (3.5)	6.1** (3.8)	0	13.9**(5.1)	0	43.9** (27.3)
hdg11-3 (control)	0	37.5 (10.9)	12.5 (4.3)	50.0 (12.6)	25.0 (8.6)	NA
hdg11-3 (no transgene sibling)	0	35.6 (8.7)	9.0 (3.3)	44.6 (10.3)	20.2 (7.4)	NA
hdg11-3 + proGL2::EYFP:GL2	0	73.2** (10.1)	2.0** (2.0)	75.2** (10.2)	2.7** (2.7)	NA

Genotype	Ave. # trichomes 1 branch (*gl2*) 1-2 branch (*myb23*)	Ave. # trichomes 2-4 branch 3 branch (*myb23*)	Ave. # trichomes >4 branch	Ave. # total trichomes	Ave. % ectopic branched trichomes >4 branch	Ave. % branched trichomes >1 branch for *gl2* genotypes 1-2 branch (*myb23*)
Col wild-type	0	54.1 (8.5)	0.3 (0.5)	54.4 (8.3)	0.5 (0.9)	NA
hdg11-1	0	49.5 (8.6)	8.1 (3.4)	57.5 (10.3)	14.1 (5.9)	NA
myb23-3	22.7 (2.8)	36.0 (6.9)	0	58.6 (7.5)	0	38.7 (4.8)
hdg11-1 myb23-3	7.7*** (2.2)	47.4 (12.3)	0	55.1 (12.4)	0	14.0 ***(4.0)

Table D.3 Quantification of trichoblast and atrichoblast hair cell files on roots.

See Figure 5.4 and Appendix Figure D.7. Ten cells were counted in pairs of root hair (R) and non root hair (N) cell files. Standard deviations for n>20 are shown in parentheses. Significant differences between wild-type and mutant are marked by a single asterisk. Significant differences between control and wild-type *HDG11* or *GL2* transgenes are indicated by a double asterisk (Two-tiered *t* test, P < 0.0001).

Genotype	% Trichoblast in R position	% Atrichoblast in N position
Col wild-type	97.5 (4.4)	98.8 (3.4)
hdg11-1	96.7 (4.8)	92.1 (8.8)
gl2-5	97.1 (5.5)	34.6* (15.0)
gl2-5 hdg11-1	95.8 (7.2)	30.8* (19.9)
myb23-3	97.1 (6.9)	92.9 (6.9)
gl2-5 myb23-3	96.5 (6.7)	43.0* (17.0)
Ler wild-type	98.0 (4.1)	97.5 (5.5)
hdg11-3	96.0 (5.9)	98.0 (4.1)
gl2-1	99.5 (2.2)	20.0* (18.0)
gl2-1 hdg11-3	99.5 (2.2)	21.5* (17.2)
hdg11-1 + ProGL2:EYFP:hdg11Δ	96.0 (6.0)	96.0 (6.8)
hdg11-1 + ProGL2:EYFP:HDG11	97.5 (4.4)	95.5 (6.9)
gl2-5 + ProGL2:EYFP:hdg11Δ	96.5 (5.9)	30.5 (18.2)
gl2-5 + ProGL2::EYFP:GL2	96.3 (5.8)	97.1** (5.5)
gl2-5 + ProGL2:EYFP:HDG11	95.4 (6.4)	62.1** (16.4)
gl2-5 hdg11-1 + ProGL2:EYFP:hdg11Δ	96.3 (6.0)	35.3 (19.0)
gl2-5 hdg11-1 + ProGL2:EYFP:HDG11	95.0 (8.3)	48.0 (20.1)

Table D.4 Oligonucleotides used in this study.

Nucleotide bases shown in bold denote restriction sites used for cloning or changed bases from site-directed mutagenesis.

I. Primers for removing *Kpn*I and *Sal*I restriction sites from *HDG11* cDNA sequence using site-directed mutagenesis. Underlined bases show introduced mutation.	
Name	**5'-3' sequence**
HDG11_SalI_F(1602-32)	C ATG AAT GCT ATG GCA CTT **GTG GAC** ATG TTC ATG G
HDG11_SalI_R(1602-32)	C CAT GAA CAT GTC CAC AAG TGC CAT AGC ATT CAT G
HDG11_KpnI_F(1704-34)	GGA ATG GGA **GGC ACC** CAT GAG GGT GCA TTG C
HDG11_KpnI_R(1704-34)	G CAA TGC ACC CTC ATG GGT GCC TCC CAT TCC
HDG11_KpnI_F(2075-2107)	T GCT TCT CTA TC**G GTG CC**A GCG TCT TCA TCT CG
HDG11_KpnI_R(2075-2107)	CG AGA TGA AGA CGC TGG CAC CGA TAG AGA AGC A
II. Primers for amplification and cloning of *HDG11* cDNA in binary vector SR54	
HDG11_KpnI_F	GAAAGAG**GGTACC**AGAAGAAAGAGGGAAGAGAGC
HDG11_ SalI_R	GAAACATTAA**GTCGAC**AAAATGAGTTTCGTCGTCGGCG
III. Primers for sequence verification of *HDG11* cDNA in *ProGL2:EYFP:HDG11*	
HDG11_(804-23)_seq	CGT TAC CAT CGT CAC ACC GC
HDG11_(1223-45)_seq	TC CAA TGC ACA TCT CAC CGT TGG
HDG11_(1935-56)_seq	TCC AAG GTT ACT TGG GTT GAA C
HDG11_(2471-92)_seq	T GGA TCC AAT GCA ACA CAT AGC
B6_(EYFP)_seq	CAA GGA CGA CGG CAA CTA C
IV. Primers for cloning in pENTR™/D-TOPO (Start codons, italics; Stop codons, red)	
GL2_TOPO_F	CACC ***ATG*** TCA ATG GCC GTC GAC ATG
GL2_TOPO_R	**TCA TTA** GCA ATC TTC GAT TTG TAG ACT TC
HDG11_TOPO_F	CACC ***ATG*** AGT TTC GTC GTC GGC GTC GGC G
HDG11_TOPO_R	GTG **TCA** AGC TGT AGT TGA AGC TGT AGG
V. Primers for real-time PCR to monitor transcript levels of GL2, HDG11, and MYB23	
ACT7_PCR(94)_F	TCGCACATGTACTCGTTTCGCTTTC (amplifies 94 bp)
ACT7_PCR(94)_R	TCGAGAAGCAGCGAGAGAGAAAGATAGA
GL2_PCR_F	ATGAAGCTCGTCGGCATGAGTGGG (ampifies 106 bp) (Ishida et al., 2007)
GL2_PCR_R	TGGATTGCCACTGAGTTGCCTCTG (Ishida et al., 2007)
HDG11_RT(125)_F	TGGGGCTGATCGTTGGGTTACCA (amplifies 125 bp)
HDG11_RT(125)_R	TGCTTCTCTTCCCTTCCGGTGA
MYB23_RT(126)_F	CATCAGACTCCACAAGCTCCTCGG (amplifies 126 bp)
MYB23_RT(126)_R	TCTCCGAGACCAAGTTTCTTGCTGAG
VI. Primers for genotyping mutants	
GL2_F_112	ATGTCAATGGCCGTCGACATGTC (Wang et al., 2007)
GL2_R_1462	TCTCGCAGCTTCTCTAGTTCCG (Wang et al., 2007)
En8130 (3'outward) (*gl2-5*)	GAGCGTCGGTCCCCACACTTCTATAC (Baumann et al., 1998)
oAR283_hdg11-1_F	ATT CTA TCA CCG GAA GGG AAG (Roeder et al., 2012)
oAR284_hdg11-1_R	TGA AGA GAA AGA GAC ACC CAG (Roeder et al., 2012)
SLB1 (SAIL_LB1) (*hdg11-1*)	GCCTTTTCAGAAATGGATAAATAGCCT (Roeder et al., 2012)
oAR300_hdg11-3_F	GTG AAG ATC CTT ACT TTG ATG AT (Roeder et al., 2012)
oAR301_hdg11-3_R	TCA AGC TAT GCA AAA AGA TCA AA (Roeder et al., 2012)
myb23-3_SALK_018613_LP	CAC CGA ACA ACA AAA CAC ATG
myb23-3_SALK_018613_RP	TTT ACG TGG ACG TTT TTG CTC

LbB1 (for *myb23-3*)	GCG TGG ACC GCT TGC TGC AAC T
VII. Primers for sequencing *ProATML1* constructs in *pAR176* (Roeder et al., 2010)	
Name	**5'-3' sequence**
oAR204	CTT CCC TTT CTC CTA AGT TCC T
oAR424	GGA GAA AAA TAG AGA GAG ATA G
VIII. Primers for screening and sequencing inserts in *pAbAi* vector	
pAbAi_Seq_F	GTTCCTTATATGTAGCTTTCGACAT (ampifies 298 bp without insert and 323 with insert)
pAbAi_R	CAGAGCACATGCCTCGAGG
IX. Primers for L1-boxes from *MYB23* promoter (L1-boxes, bold; first L1-box out of three, underlined; mutations, lowercase)	
proMYB23_LI-box-I_F	C**TACATTTATACATTTATACATTTA**G
proMYB23_LI-box-I_R	TCGAC**TAAATGTATAAATGTATAAATGTA**GGTAC
proMYB23_mutLI-box-I_F	C**TACggTTATACggTTATACggTTA**G
proMYB23_mutLI-box-I_R	TCGAC**TAAccGTATAAccGTATAAccGTA**GGTAC
proMYB23_LI-box-II_F	C**TAAATGTGTAAATGTGTAAATGT**GG
proMYB23_LI-box-II_R	TCGAC**CACATTTACACATTTACACATTTA**GGTAC
proMYB23_mutLI-box-II_F	C**TAAccGTGTAAccGTGTAAccGT**G
proMYB23_mutLI-box-II_F	TCGAC**CACggTTACACggTTACACggTTA**GGTAC
X. Primers for construction of pGAD-T7 constructs for HD-Zip expression in yeast (Start codons, italics; Stop codons, red)	
GL2_747aa_EcoRI_F	GACAGG**GAATTC** *ATG* TCA ATG GCC GTC GAC ATG TCT TCC
GL2_231aa_BamHI_R	ATCGTC**GGATCCTCA** AGC CTG CAG GGG ATA GGG
HDG11_722aa_EcoR1_F	TAATTA**GAATTC** *ATG* AGT TTC GTC GTC GGC GTC GGC GG
HDG11_169aa_BamH1_R	TGTAG**GGATCCCTA** TTG CGA TAT CGG TCT TCC CAT GTA C
XI. Primers for construction of pGAD-T7 control construct for expression of GL2 (START + SAD) (Stop codons, red)	
GL2_253aa_EcoRI_F	GTCTTT**GAATTC** GAG AAG TCC CGT ATT GCC GAG ATT TC
GL2_747aa_BamHI_R	TGCTGT**GGATCCTCATCA** GCA ATC TTC GAT TTG TAG AC
XII. Primers for real-time PCR of ChIP DNA	
ACT7_ChIP_F	CGTTTCGCTTTCCTTAGTGTTAGCT
ACT7_ChIP_R	AGCGAACGGATCTAGAGACTCACCTTG
proCESA5_ChIP_F	GTATAGTCATCACTCAGGAGAC
proCESA5_ChIP_R	GACTCACTGAGTCTTTTTATCGG
proMYB23_ChIP_F	GGATCCACACAAGACTAAAATAC
proMYB23_ChIP_R	GCTATTCACATTTAGATTTTCCATC
XIII. Primers for deletion of L1-Box I in *ProMYB23:mGFP5-ER*	
ΔL1-box_F	CTTAAATTCTTTTTCACAAGATGTTTATAATG
ΔL1-box_R	GGTACCCATTTTTGTTGTATTTTAGTC
IX. Primers for sequencing *ProMYB23:mGFP5-ER*	
MYB23pro_2703_F	TCTAGAGTTTGGTATCACGG
MYB23pro_1836_F	TATTAGTATAGTTTGTTGTTCAAACGG
MYB23pro_seq_2082_R	TTCTTACATTTCATCTCATTGC
MYB23pro_3450_F	ACGGGTGTTAACACACACC

REFERENCES

Santambrogio C, Grandori R. 2008. Monitoring the Tanford transition in beta-lactoglobulin by 8-anilino-1-naphthalene sulfonate and mass spectrometry. *Rapid Commun Mass Spectrom* **22**: 4049-4054.

Schrick K, Bruno M, Khosla A, Cox PN, Marlatt SA, Roque RA, Nguyen HC, He C, Snyder MP, Singh D et al. 2014. Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity. *BMC biology* **12**: 70.

Tagore R, Thomas HR, Homan EA, Munawar A, Saghatelian A. 2008. A global metabolite profiling approach to identify protein-metabolite interactions. *J Am Chem Soc* **130**: 14111-14113.

Vinayavekhin N, Saghatelian A. 2011. Discovery of a protein-metabolite interaction between unsaturated fatty acids and the nuclear receptor Nur77 using a metabolomics approach. *J Am Chem Soc* **133**: 17168-17171.