START domain mediates *Arabidopsis* GLABRA2 transcription factor dimerization and turnover independently of homeodomain DNA binding

Thiya Mukherjee,[a,b,c,1] Bibek Subedi,[a,b,1], Aashima Khosla,[a,b,d] Adara L. Warner,[a] Ruben Lerma-Reyes,[a,e] Kyle A. Thompson,[a] Sumedha Gunewardena,[f] and Kathrin Schrick[a,b,2]

[a] Division of Biology, Kansas State University, Manhattan, KS 66506, USA
[b] Molecular, Cellular and Developmental Biology, Kansas State University, Manhattan, KS 66506, USA
[c] Donald Danforth Plant Science Center, Olivette, MO 63132, USA
[d] Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA
[e] Interdepartmental Genetics, Kansas State University, Manhattan, KS 66506, USA
[f] Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, KS 66160, USA

[1] These authors contributed equally
[2] Address correspondence to kschrick@ksu.edu

The author responsible for distribution of materials integral to the findings presented in this article is: Kathrin Schrick (kschrick@ksu.edu).

Short title: **START acts independently of homeodomain**

ABSTRACT

Class IV homeodomain leucine-zipper transcription factors (HD-Zip IV TFs) are key regulators of epidermal differentiation that are characterized by a DNA-binding homeodomain (HD) in conjunction with lipid sensor domain termed START (Steroidogenic Acute Regulatory (StAR)-related lipid Transfer). Previous work demonstrated that the START domain of GLABRA2 (GL2), a HD-Zip IV member from *Arabidopsis*, is required for transcription factor activity. Here, we address the functions and possible interactions of START and the HD in DNA binding, dimerization, and protein turnover. Deletion analysis of the HD and missense mutations of a conserved lysine (K146) result in phenotypic defects in leaf trichomes, root hairs and seed mucilage, similar to those observed for START mutants, despite nuclear localization of the mutant proteins. Gel shift and ChIP-seq experiments demonstrate that while HD mutations impair binding to target DNA, the START domain is dispensable for DNA binding. Vice versa, yeast two-hybrid assays reveal impaired GL2 dimerization for START domain mutants, but not HD mutants. Using *in vivo* cycloheximide chase experiments, we provide evidence for the role of START, but not HD, in maintaining protein stability. This work advances our fundamental understanding of HD-Zip TFs as multidomain regulators of epidermal development in plants.

INTRODUCTION

Class IV homeodomain leucine-zipper transcription factors (HD-Zip IV TFs) are critical regulators of epidermal and sub-epidermal differentiation in plants (Nakamura et al., 2006). In *Arabidopsis*, there are 16 START domain-containing transcription factors belonging to the HD-Zip IV family. The founding member, GLABRA2 (GL2), functions in a complex regulatory network to dictate epidermal cell fate. Genetic studies initially showed that GL2 promotes trichome differentiation in shoots (Rerie et al., 1994) and mucilage biosynthesis in seeds (Western et al., 2001; Western et al., 2004), while it suppresses root hair formation in non-hair root cells (Di Cristina et al., 1996; Masucci et al., 1996). More recent mutant analyses revealed that GL2 regulates the accumulation of seed oil (Shen et al., 2006; Shi et al., 2012) and anthocyanin biosynthesis (Wang et al., 2015). In addition to GL2, the functionally redundant family members *Arabidopsis thaliana* MERISTEM LAYER1 (ATML1) and PROTODERMAL FACTOR2 (PDF2) have been extensively studied for their developmental role in specification of the epidermis in the embryo (Abe et al., 2003; Ogawa et al., 2015; Iida and Takada, 2021). HD-Zip IV proteins are comprised of four defined domains including a DNA-binding homeodomain (HD), a leucine-zipper domain termed Zipper Loop Zipper (ZLZ), a Steroidogenic Acute Regulatory (StAR) protein-related lipid Transfer (START) domain, and a START-associated domain (SAD or STAD) (Schena and Davis, 1994; Di Cristina et al., 1996; Ponting and Aravind, 1999; Riechmann et al., 2000; Schrick et al., 2004).

The HD is a DNA-binding domain of ~60 amino acids that is highly conserved among eukaryotes (Carrasco et al., 1984; McGinnis et al., 1984; Gehring et al., 1990; Bürglin, 1994; Derelle et al., 2007). It was initially discovered in *Drosophila* genes, whose mutation or overexpression cause homeotic transformations affecting the body pattern (Garber et al., 1983; McGinnis et al., 1984; Scott and Weiner, 1984). The 3D structure of the HD shows a flexible amino-terminus and three α-helices, the second and third of which form a helix-turn-helix motif (Gehring et al., 1990; Kissinger et al., 1990; Liu et al., 1990; Otting et al., 1990; Gehring, 1994). DNA binding specificity occurs through major groove contacts via the third helix as well as minor groove contacts via the amino-terminal region (Schofield, 1987; Mann et al., 2009; Burglin and Affolter, 2016). The first reported homeobox gene from plants was maize *Knotted-1* (*Kn-*

1), whose gain-of-function phenotype is marked by uncontrolled cell divisions in lateral veins, forming outpocketings or knots (Vollbrecht et al., 1991). Since the discovery of *Kn-1*, numerous plant homeobox genes have been identified and implicated in a wide variety of developmental processes. In *Arabidopsis,* there are 110 predicted HD proteins that are subdivided into 14 classes including HD-Zip TFs (Classes I-IV) (Mukherjee et al., 2009).

The HD-Zip III and IV TFs contain a START domain that occurs in the middle of the protein, carboxy-terminal to the HD-Zip. Initially discovered in mammalian StAR (Steroidogenic Acute Regulatory) cholesterol-binding transporters (Kallen et al., 1998), START (StAR-related lipid Transfer) domains are modules of ~210 amino acids (Ponting and Aravind, 1999) conserved across plants and animals as well as several protists, bacteria and archaea (Iyer et al., 2001). The domain structure comprises of an α/β helix-grip fold that forms a binding pocket (Tsujishita and Hurley, 2000; Thorsell et al., 2011). Human START domain-containing proteins bind ligands such as sterols, phospholipids, ceramides, bile acids and fatty acids (Kallen et al., 1998; Tsujishita and Hurley, 2000; Roderick et al., 2002; Kudo et al., 2008; Létourneau et al., 2012; Tillman et al., 2020). In mammals, START domains function in lipid/sterol transport or metabolism and several act in intracellular signaling pathways (Clark and Stocco, 1995; Lin et al., 1995; Watari et al., 1997; Stocco, 2001; Soccio and Breslow, 2003).

Studies in plants indicate similar roles for START domains in lipid transfer and/or lipid sensing. In the liverwort *Marchantia polymorpha*, a START domain-containing protein (STAR2) was recently implicated in the incorporation of ER-derived C20 fatty acids into chloroplast glycolipids (Hirashima et al., 2021). Our previous work showed that the START domain is essential for transcription factor activity of HD-Zip IV TF GL2 (Schrick et al., 2014). Several mechanisms have been proposed to explain how the ligand-bound START domain can stimulate transcription factor activity, including conformational changes, stabilization of protein levels, and/or promotion of protein–protein interactions (Schrick et al., 2014). However, lack of detailed information on how START modulates transcription factor activity poses as a major challenge.

The goal of this study was to understand how the START domain modulates HD-Zip IV transcription factor activity either independently or in synergy with the HD. We

generated HD mutations in wild-type and START domain-deleted backgrounds and studied their effect on transcription factor function. HD mutants lacking the predicted DNA-binding α-helices or containing missense mutations in a conserved lysine residue imparted defective phenotypes in trichomes, roots and seeds, similar to those of START mutants. While HD mutations abolished DNA binding, mutations affecting the START domain retained DNA binding affinity *in vitro* and *in vivo*. However, yeast two-hybrid experiments revealed impaired dimerization in START mutants but not in HD mutants. Finally, cycloheximide chase experiments showed that protein levels drop rapidly in several START domain mutants but not in HD mutants, consistent with a unique role of START in maintaining protein stability.

RESULTS

HD and START domain are highly conserved across the HD-Zip IV TFs

The HD is evolutionarily conserved across all major eukaryotic lineages (Carrasco et al., 1984; McGinnis et al., 1984; Derelle et al., 2007; Mukherjee et al., 2009). We found that the crystal structure of HOMEOBOX PROTEIN ENGRAILED-2 (Gg-En-2) from chicken exhibits the closest homology with the HD from HD-Zip IV members. Aligning HDs from all 16 HD-Zip IV TFs including GL2, PDF2, ANL2, ATML1, HDG1-12 with reference to Gg-En-2 revealed highly conserved α-helices as well as amino acid residues (**Figure 1A**). One of the residues displaying 100% conservation in HD-Zip IV members is the lysine at position 45 of the HD (**Figures 1A and 1B**). This conserved lysine selectively recognizes the 3' CC dinucleotide adjacent to the TAAT core sequence (Amendt et al., 1998; Cazorla et al., 2000; Lebel et al., 2001) in the vertebrate Bicoid-like HD transcription factors PITX2 and PITX3 (Semina et al., 1996; Gage and Camper, 1997; Semina et al., 1997; Smidt et al., 1997).

To determine the conserved features of the START domain among HD-Zip IV members, we created a structure-based sequence alignment of HD-Zip IV START domains with STARD5, a mammalian START protein with known crystal structure (Soccio et al., 2002; Rodriguez-Agudo et al., 2005; Thorsell et al., 2011). Despite low sequence identity, the alignment revealed consensus secondary structures (α helices

and β strands) as well as conserved amino acids implicated in ligand binding (**Figure 1C; Supplemental Figure 1**).

To study the role of the HD in HD-Zip IV TFs, we generated several constructs with HD mutations in both wild-type GL2 (and/or PDF2) and START-deleted (ΔSTART) backgrounds (**Figure 1D**). The gl2$^{\Delta ST}$ mutant lacks the 235 amino acids of the START domain between ZLZ and STAD domains (Schrick et al., 2014) (**Figure 1D**). We constructed HD deletion mutants that remove the consensus α-helices in GL2 (HDΔ11, HDΔ20, HDΔ46) and missense mutations (K-E, K-T) targeting a highly conserved lysine residue (K146 in GL2; K107 in PDF2) (**Figures 1A, 1B and 1D**). The positively-charged lysine was replaced with a negatively-charged glutamic acid or with the neutral polar residue, tyrosine. The K-E alteration corresponds to the dominant-negative K88E mutation in PITX2 HD that abolishes DNA binding (Saadi et al., 2001; Saadi et al., 2003). Transgenic lines were constructed in the *gl2-5* null mutant background with wild-type *GL2* or mutant *gl2* expressed under the native *GL2* promoter with an enhanced yellow fluorescent protein (EYFP) translationally fused at the amino-terminus. For each genotype, we screened 28-55 independent T1 transformants to monitor nuclear expression of EYFP and phenotypic complementation of *gl2-5* (**Supplemental Table 1**), and chose representative transgenic lines for further characterization.

Mutations affecting the HD or START domain of GL2 exhibit defects in trichomes, root hairs and seed coat mucilage

Deletion of the START domain in *GL2* was previously shown to cause visible defects in differentiation of the epidermis in the leaf (trichomes), root (non-hair cells), and seed (seed coat mucilage) similar to *gl2-5* null mutant phenotypes (Schrick et al., 2014). In this study, we examined the phenotypic effects of mutations in the HD in presence or absence of START domain deletion (**Figure 2**). Unlike wild-type plants, rosette leaves from HD mutants lacking consensus α-helices (gl2$^{\Delta HD11}$, gl2$^{\Delta HD20}$, gl2$^{\Delta HD46}$) or containing a K-E substitution in the conserved lysine residue (gl2^{K146E}) displayed glabrous phenotypes and were indistinguishable from ΔSTART or *gl2-5* null mutants (**Figure 2A and 2A'**). Another missense mutation in the same lysine residue (K146T) resulted in a

partial rescue phenotype (**Figure 2A'**), suggesting that the HD is not fully functional in this mutant. Importantly, three lines harboring both HD and START domain mutations ($gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146E;\Delta ST}$, $gl2^{K146T;\Delta ST}$) were also defective in trichome differentiation similar to ΔSTART mutants (**Figure 2A'**). Quantification of trichomes on the first leaves revealed a reduction in total number as well as the abnormal presence of unbranched trichome in mutants (**Figure 2D**). These data demonstrate that both the HD and START domain are required for the proper differentiation of trichomes.

Apart from trichome defects, mutation of HD resulted in excessive root hair formation characteristic of *gl2-5* null mutants, irrespective of accompanying START domain mutation (**Figures 2B, 2B' and 2E**). To assess the effect of mutations in the differentiation of mucilage secretory cells, seeds were imbibed in water to allow the release of seed coat mucilage followed by staining with ruthenium red, a dye that binds to pectinaceous polysaccharides. Comparisons of thickness of halos surrounding the stained seeds indicated the reduced seed mucilage production in HD mutants $gl2^{\Delta HD11}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD46}$, and $gl2^{K146E}$, while the K146T mutation resulted in a partial mucilage defect (**Figures 2C and 2C'**) consistent with partial phenotypes observed in trichomes and roots. HD and START domain double mutants ($gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146E;\Delta ST}$, $gl2^{K146T;\Delta ST}$) were also defective in seed mucilage similar to ΔSTART single mutants (**Figures 2C and 2C'**). These results demonstrate that both the HD and START domain are critical for epidermal differentiation in multiple cell types.

HD and START domain mutants retain nuclear localization of GL2

GL2 is expressed in developing trichomes of the leaf epidermis (Rerie et al., 1994; Szymanski et al., 1998), hairless cell files in the epidermis of the root (Di Cristina et al., 1996; Masucci et al., 1996) and hypocotyl (Hung et al., 1998), and seed coat cells (Western et al., 2004). As a transcription factor, GL2 is presumably translocated to the nucleus where it binds DNA to regulate gene expression. Using EYFP fusions to GL2, it was previously shown that the START domain is not necessary for nuclear localization (Schrick et al., 2014). Here we monitored the expression of *EYFP:GL2* transgenes harboring various combinations of HD and START domain mutations in leaf trichomes and developing seeds. Using live imaging of trichomes, we observed nuclear

localization of the EYFP-tagged proteins in all the HD mutants similar to that for wild-type GL2 and gl2$^{\Delta ST}$ (**Figures 3A, 3A', 3B and 3B'**). Nuclear localization remained intact in trichomes of mutants for both the HD and START domains (gl2$^{\Delta HD20;\Delta ST}$, gl2$^{K146E;\Delta ST}$, gl2$^{K146T;\Delta ST}$) (**Figures 3A' and 3B'**). Confocal laser scanning microscopy of developing seeds indicated nuclear localization of HD and START mutant proteins in the seed coat cells (**Figures 3C, 3C', 3D and 3D'**). The results demonstrate that both HD and START domain are dispensable for translocation of GL2 to the nucleus.

HD, but not START domain, is required for DNA binding *in vitro*

HD-Zip IV TFs regulate the expression of various downstream target genes by binding to *cis*-elements in the promoter regions of their respective targets (Nakamura et al., 2006; Khosla et al., 2014). Gel Shift or Electrophoretic Mobility Shift Assay (EMSA) is an efficient tool to detect consensus target sequence binding by plant HD transcription factors (Bolduc and Hake, 2009; Chen et al., 2020). Here, we examined whether DNA binding of HD-Zip IV TFs is regulated by the HD alone or by HD in combination with the START domain. The DNA sequence we used was the L1-Box element (TAAATGTA) that is found upstream of several GL2 gene targets and was previously implicated in GL2-DNA binding (Ohashi et al., 2003; Tominaga-Wada et al., 2009; Khosla et al., 2014). We produced recombinant Halo-tagged transcription factors using wheat germ *in vitro* transcription-translation. Despite repeated attempts, we were unable to obtain full-length recombinant GL2 protein that was functional *in vitro*.

Given the high degree of sequence similarity between HD-Zip IV proteins (**Figure 1**), we selected PDF2 and mutant versions analogous to our gl2 HD and START mutants (pdf2^{K107E}, pdf2^{K107T}, pdf2$^{\Delta ST}$, pdf2$^{K107E;\Delta ST}$, pdf2$^{K107T;\Delta ST}$) for DNA binding studies. ATML1, which functions redundantly with PDF2, is reported to bind the octamer TAAATG(C/T)A (Abe et al., 2001). Thus, we performed EMSAs with Cy3 fluorescently labeled oligonucleotide containing TAAATGTA (**Supplemental Table 3**). Both wild-type PDF2 and pdf2$^{\Delta ST}$ caused a shift in mobility of the Cy3 probe as an indication of DNA binding (**Figure 4A**). In contrast, the two missense mutations K107E and K107T resulted in defective mobility shift of the probe independently of the START domain. K107E completely abolished binding of PDF2 (pdf2^{K107E}) and pdf2$^{\Delta ST}$ (pdf2

$^{K107E;\Delta ST}$) to the target sequence as indicated by a lack of mobility shift. Intriguingly, K107T resulted in partial DNA binding for both wild-type PDF2 (pdf2^{K107T}) and pdf2$^{\Delta ST}$ (pdf2 $^{K107T;\Delta ST}$) (**Figure 4A**). Western blotting confirmed uniform expression of the recombinant proteins used in our EMSA experiments (**Figure 4B**). Surprisingly, by combining EMSA with native polyacrylamide gel electrophoresis, we found that while the wild-type PDF2 protein binds the DNA as a dimer, the pdf2$^{\Delta ST}$ mutant binds the DNA target sequence as a monomer (**Figure 4C; Supplemental Figure 2**). The abnormal monomeric form was also observed in comparing the DNA-binding behavior of the mutant atml1$^{\Delta ST}$ protein to wild-type ATML1 dimer (**Supplemental Figure 2**).

START domain is dispensible for DNA binding *in vivo*

We performed ChIP-seq assays with seedlings expressing EYFP-tagged GL2 wild-type and mutant proteins to test whether the START domain is dispensible for DNA binding *in vivo*. The experiments included three START domain mutants (gl2$^{\Delta ST}$, gl2$^{E375G;R392M}$ gl2^{Amo}) as well as a mutant deleted for the six terminal amino acids of the HD and the entire ZLZ domain (gl2$^{\Delta HD6;\Delta ZLZ}$) (**Figure 1**), all of which displayed leaf trichome defects similar to the *gl2-5* null mutant (**Supplemental Figure 3**). The mutations in gl2$^{E375G;R392M}$ disrupt a salt bridge that is thought to be critical for ligand binding (Roostaee et al., 2009). In gl2^{Amo}, the R386L and E387L missense mutations, which remove charged residues, are followed by an 11-residue duplication in a predicted ligand-contact region (**Figure 1; Supplemental Figure 1**).

Data analysis revealed DNA binding proficiency for wild-type GL2 and all three START domain mutants whereas little or no binding was detected for gl2$^{\Delta HD6;\Delta ZLZ}$ and the *gl2-5* null mutant control (**Figure 4D; Supplemental Table 2**). Genomic distributions of binding sites were similar for wild-type GL2 and the START domain mutants. Most of the ChIP-seq hits were mapped to promoter or intergenic regions, consistent with the role of GL2 as a transcriptional regulator (**Figure 4E; Supplemental Table 2**). The relative positions of binding sites to transcription start sites (TSS) were also similar for the wild-type GL2 and START domain mutants (**Figure 4F; Supplemental Figure 4**). Gene ontology (GO) classification revealed overlapping functional classes of target genes bound by wild-type GL2 and the START domain mutants (**Figure 4G**). In the

category “biological processes”, there was a similar enrichment for genes involved in cellular processes (GO:0009987; p = 4.8e-31) and metabolism (GO:0008152; p = 2.4e-42). In the category of “molecular function”, genes encoding transferase activity (GO:0016740; p = 8e-09), kinase activity (GO:0016301, p = 6.1e-07), and metal ion binding (GO:0046872; *p* = 4.5e-05) appeared as the most enriched group of genes for wild-type GL2 and two of the START mutants ($gl2^{Amo}$, $gl2^{E375G;R392M}$). Although the gene targets bound by the $gl2^{\Delta ST}$ protein were similarly enriched for metal ion binding, they were not enriched for transferase activity and kinase activity (**Figure 4G**). Overall, the data indicate that START domain mutants retain DNA binding activity in a manner that is similar to wild type.

START domain is required for GL2 dimerization independently of the HD

Previous studies demonstrated the important role of the leucine zipper for homo- and hetero-dimerization of HD-Zip transcription factors (Landschulz et al., 1988; Sessa et al., 1993). Our EMSA experiments indicated that the START domain is critical for homodimerization of PDF2 and ATML1 when bound to target DNA (**Figure 4C; Supplemental Figure 2**). We tested GL2 in yeast two-hybrid (Y2H) assays to further address whether the START domain and/or HD play a role in homodimerization. For these experiments, baits were expressed as GAL4 DBD:GL2 (or mutant gl2) fusion proteins and preys were expressed as GAL4 AD:GL2 (or gl2) fusion proteins in yeast strains of opposite mating type. Bait and prey strains were mated to obtain diploids that were assayed for protein-protein interaction and thus homodimerization (**Figure 5A**). Diploids expressing wild-type GL2 and those expressing HD mutants ($gl2^{\Delta HD}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{\Delta HD46}$, $gl2^{K146T}$, $gl2^{K146E}$) grew on selective plates indicating homodimerization (**Figure 5A**). In contrast, diploids expressing ΔSTART mutants ($gl2^{\Delta ST}$, $gl2^{K146T;\Delta ST}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146E;\Delta ST}$) failed to grow under selection, similar to yeast expressing empty bait and prey vectors (**Figure 5A**). These results indicate that GL2 homodimerization is unaffected by HD mutations but is impaired by deletion of the START domain. Western blotting confirmed the expression of wild-type and mutant fusion proteins in yeast (**Figure 5B**).

To rule out the possibility that the observed dimerization is due to autoactivation of bait proteins, yeast cells expressing bait constructs alone were assayed for reporter gene activity. The HD-Zip IV TF members ATML1 and PDF2 served as positive controls that exhibited autoactivation as bait proteins. In contrast, the wild-type GL2 bait (as well as the HD and START mutant baits) did not display autoactivation in yeast (**Supplemental Figure 5A**).

We performed additional Y2H experiments using wild-type GL2 as bait in combination with preys harboring gl2 mutant proteins (**Supplemental Figure 5B**). Converse experiments were also conducted using mutant gl2 proteins as baits with wild-type GL2 as prey (**Supplemental Figure 5C**). In both cases, homodimerization was observed for HD mutants ($gl2^{\Delta HD}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{\Delta HD46}$, $gl2^{K146T}$, $gl2^{K146E}$) but not for ΔSTART mutants ($gl2^{\Delta ST}$, $gl2^{K146T;\Delta ST}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146E;\Delta ST}$) (**Supplemental Figures 5B and 5C**). These results indicate that the START domain regulates GL2 dimerization independently of the HD.

START domain, but not HD, regulates GL2 protein stability

Little is known about how the level of HD-Zip proteins is regulated post-translationally. Recently, it was reported that mutation of a putative lipid contact site in the START domain of ATML1 disrupts expression of the fluorescently tagged protein in lateral roots (Nagata et al., 2021). In our study, we observed that gl2 proteins harboring mutations in either the HD or START domain appeared to be expressed weakly in comparison to the wild-type protein (**Figure 3**).

To determine whether the HD and/or START domain play a role in regulating GL2 protein levels, we monitored protein half-life by treating *Arabidopsis* seedlings with cycloheximide, an inhibitor of cellular protein synthesis. Proteins were extracted from seedlings at various intervals over a 10 or 24 h time course followed by immunodetection and quantification. We analyzed wild-type GL2, $gl2^{\Delta HD20}$ and $gl2^{\Delta ST}$ mutants which we had tested for DNA binding and dimerization (**Figures 4 and 5**). Three additional START domain mutants were included in our experiments: (1) $gl2^{L480P}$, a missense allele affecting a lipid contact site in the C-terminal α-helix of the START domain (Roostaee et al., 2009; Alpy and Tomasetto, 2014; Wojciechowska, 2021), (2)

gl2^{W279R}, a missense allele predicted to add a charged residue within a hydrophobic region of the binding pocket, and (3) gl2^{Amo}, which we also assayed for DNA binding using Chip-seq (**Figures 1C and 1D; Supplemental Figure 1; Figure 4D-G; Supplemental Figure 3**). As revealed by Western blot and protein quantification, wild-type GL2 and gl2$^{\Delta HD20}$ proteins remained stable over time, whereas protein levels of the START domain mutants (gl2$^{\Delta ST}$, gl2^{L480P}, gl2^{W279R}, gl2^{Amo}) dropped rapidly (**Figures 6A and 6B; Supplemental Figure 6**). The gl2$^{\Delta ST}$ and gl2^{W279R} mutant proteins were particularly unstable, having barely detectable protein levels even at the 2-h time point (**Supplemental Figure 6**). In case of gl2$^{\Delta ST}$, a degradation product of ~65 kDa also exhibited a diminished protein half-life (**Supplemental Figure 6**). We additionally examined the stability of proteins that harbored mutations in both the HD and START, but the low levels of expression made it difficult to interpret their half-lives in cycloheximide assays (**Supplemental Figure 7**). Taken together, our results provide strong evidence for the role of START in maintaining protein stability of GL2.

To better understand the degradation pathway used to control protein levels, additional experiments were performed with one of the START mutants (L480P) in presence or absence of proteasome inhibitor MG132. Upon addition of MG132, the gl2^{L480P} protein remained stable for ~5-10 h during CHX treatment whereas the control treatment resulted in disappearance of the protein as early as 2 h (**Figures 6C and 6D**). The data suggest the possible involvement of Ubiquitin/26S proteasome system (UPS) in the regulation of GL2 protein levels.

DISCUSSION

Domain organization in HD-Zip IV TFs and nonoverlapping roles of HD and START

The combination of a DNA-binding HD with a lipid-sensing START domain in a single protein is plant-specific (Schrick et al., 2004) and likely represents a novel mode of action in a multidomain transcription factor. At the outset of this study, we envisioned that the START domain regulates transcription factor activity of HD-Zip IV TFs through its DNA binding domain. Our overall goal was to investigate the possible association or overlap in the roles of the HD and START domains. In addition to assaying phenotypic

effects of genetic mutations *in planta*, we considered several aspects of transcription factor function, including subcellular localization, DNA binding, dimerization, and protein stability. We found that mutations affecting HD activity result in loss-of-function phenotypes, as illustrated by trichome differentiation defects, reduced seed mucilage and ectopic root hair formation similar to *gl2-5* null mutants and *gl2*$^{\Delta ST}$ mutants (**Figure 2**). Thus, both HD and the START domain are critical for epidermal differentiation. The data presented herein show that the molecular mechanisms underlying the loss-of-function phenotypes are different in HD versus START domain mutants. While mutations in the HD abolish DNA binding of the transcription factor, the loss of START domain function results in dimerization defects and unstable protein.

Conserved role of HD in DNA binding

A missense HD mutation changing a conserved lysine to a glutamic acid (K107E) completely disrupted DNA binding activity of PDF2 **(Figure 4A**). The analogous K-E substitution in the vertebrate HD transcription factors PITX2 and PITX3 also abolishes DNA binding (Saadi et al., 2001; Saadi et al., 2003; Sakazume et al., 2007). In contrast, we found that a K-T substitution at the same position (K107T) exerts a partial defect (**Figure 4A**). The effect of either of these two missense mutations on DNA binding was unaltered by the deletion of the START domain **(Figure 4A**), consistent with our model that the HD functions in DNA binding independently of the START domain.

Our results indicate that START domain deletion mutants are proficient in DNA binding under both *in vitro* and *in vivo* conditions. Several independent studies report that HD-Zip IV TFs bind DNA elements as dimers (Sessa et al., 1993; Palena et al., 1999; Tron et al., 2001). Intriguingly, the gel shift experiments (**Figure 4C, Supplemental Figure 2**) and Y2H assays (**Figure 5; Supplemental Figure 5**) suggest that the START domain is important for dimerization, but not for DNA binding. We were surprised by the finding that START-deleted proteins bind DNA as monomers (**Figure 4C; Supplemental Figure 2**). Thus, the START domain behaves differently from the ZLZ domain, which was shown to be required for dimerization as well as DNA binding (Di Cristina et al., 1996).

START domain is critical for protein stability

Our cycloheximide assays revealed that wild-type GL2 protein is relatively stable, while mutations in the START domain dramatically reduce protein half-life (**Figures 6A and 6B; Supplemental Figure 6**). It is noteworthy that gl2^{L480P}, gl2^{W279R} and gl2^{Amo} harbor missense mutations in highly conserved amino acid residues of the START domain predicted to be important for ligand binding (**Figures 1A and 1D; Supplemental Figure 1**). Thus, our findings support the hypothesis that ligand binding mediated via the START domain stabilizes GL2 protein levels. Stabilization of protein levels might aid in activation of the protein, promotion of protein–protein interactions such as dimerization, or correct protein folding. In eukaryotes, protein degradation acts as an important regulatory process that mediates optimum responses of cellular pathways to environmental cues. Our experiments with proteasome inhibitor MG132 point to possible degradation of GL2 via the 26S proteasome (**Figures 6C and 6D**). However, proteasome inhibition is known to cause a global inhibition of translation resulting in proteome-wide effects (Larance et al., 2013). Thus, MG132 treatment might induce other pathways affecting protein stability, such as autophagy (Carlsson and Simonsen, 2015; Kang et al., 2018). Further investigation will be required to characterize the precise mechanism of GL2 protein degradation including possible post-translational modifications.

Model for role of HD and START domain in promoting HD-Zip IV TF activity

We propose a model in which the function of the HD is to recruit the HD-Zip IV TF to the DNA target site (**Figure 7**). In this way, the multidomain transcription factor containing the START domain is correctly positioned on the DNA to activate or repress target genes. If HD function is missing, the HD-Zip IV protein is stably expressed and is capable of dimerization via its START domain. However, if the START domain is missing or non-functional, the protein fails to dimerize and becomes unstable. Since the START domain is required for both dimerization and protein stability, it is possible that these two functions are tightly interlinked. Thus far, we have not identified a mutant or condition in which these two processes are uncoupled. Therefore, START-domain-mediated dimerization likely contributes stabilization of the transcription factor. The

sequence of events leading to transcriptional activity might involve several steps starting with ligand binding of the HD-Zip IV TF through its START domain. Ligand binding could stabilize dimerization which in turn allows the transcription factor to bind the DNA via its HD. It is noteworthy that in the absence of the START domain, the transcription factor binds to its DNA target site through the HD, presumably via its monomeric form (**Figure 4C; Supplemental Figure 2**). In our model, unliganded monomers lacking the START domain are unstable, resulting in premature degradation (**Figure 7**).

Other possible domain interactions in HD-Zip IV TFs

It is tempting to speculate that the START domain modulates the activity of another functional part of the protein in response to ligand binding. Crystallographic and biochemical studies with the mammalian ceramide transfer protein (CERT1/STARD11) revealed that its START domain interacts with its pleckstrin homology (PH) domain to regulate cellular localization and ceramide transfer (Prashek et al., 2017). In another study, mammalian phosphatidylcholine transfer protein (PCTP/STARD2), which binds phosphatidylcholine through its START domain, was shown to interact with and enhance the transcriptional activity of paired-box HD protein PAX3 (Kanno et al., 2007). Although this interaction occurs between two separate proteins, a similar type of interaction could occur between two domains within the same protein.

While our study was focused on the HD and the START domain, several other domains or motifs might also exert regulatory roles either independently or in conjunction with these functional domains. Possible candidates include the ZLZ and the START-associated domain (STAD), with the ZLZ being the most well characterized. The ZLZ is as a plant-specific leucine zipper with two sets of three heptad repeats intervened by a loop of ~10-20 amino acids that includes two highly conserved cysteines (Schrick et al., 2004). One study showed that these cysteines are involved in redox regulation of dimerization as well as the regulation of DNA binding (Tron et al., 2002). Our results demonstrate that both the HD and START domain are dispensable for nuclear localization of the GL2 transcription factor in developing seeds and trichomes (**Figure 3**). An uncharacterized motif that maps elsewhere is likely responsible for nuclear localization of HD-Zip IV TFs. In previous studies, several plant

HD-Zip I proteins were found to contain a nuclear localization signal (NLS) either at their amino- or carboxy-terminus (Arce et al., 2011; Cao et al., 2016; Shao et al., 2018).

Taken together, our study reports distinct roles of the HD and START domain, both of which contribute to the overall activity of HD-Zip IV TFs. These transcription factors are critical for orchestrating plant epidermal development, and they additionally play vital roles in responses to biotic and abiotic stress. Our results not only advance our current understanding of these developmentally important transcription factors but also lay the groundwork for engineering of plant-derived HD and START domains as modular units in synthetic biology applications. Future studies will explore how HD-Zip IV protein levels are tightly regulated during epidermal differentiation. Because these transcription factors also contain a specialized leucine zipper domain (ZLZ) and STAD, a plant-specific domain of unknown function, a promising avenue of future investigation will be to examine whether these domains work in association with HD and/or START to modulate various transcription factor activities.

METHODS

Plant material, growth conditions, and *Arabidopsis* transformation

Wild-type *Arabidopsis thaliana* ecotype Columbia (*Col*) and the *gl2* null allele *gl2-5* (Khosla et al., 2014) were used for this study. Seeds were stratified at 4°C for 3 days and grown on soil containing PRO-MIX PGX, vermiculite and perlite (Hummert International) in a ratio of 4:2:1 at 23°C under continuous light. The *EYFP:GL2* constructs were transformed into *gl2-5* plants using *Agrobacterium tumefaciens* GV3101 (pMP90) mediated floral dip (Clough and Bent, 1998). T1 progeny were screened on 20 µg/ml hygromycin B. Independent T1 transformants (28-55) were characterized for each genotype and 3:1 segregation for EYFP expression were confirmed in T2 progeny. Representative T3 homozygous lines with stable EYFP expression were confirmed by PCR genotyping and were used for further analysis.

HD and START domain structural alignments

The closest structural homolog of the HD from GL2 was identified as the solution structure of *Gallus gallus* HOMEOBOX PROTEIN ENGRAILED-2 (Gg-En-2) (PDB ID =

3ZOB) using the SWISS-MODEL homology modeling server (Arnold et al., 2006; Benkert et al., 2011; Biasini et al., 2014). A structural homolog suitable for HD-Zip IV derived START domain modeling was identified as the crystal structure of human STARD5 (PDB ID = 2R55) using I-TASSER (Yang et al., 2015). Sequences from all 16 HD-Zip IV TFs from *Arabidopsis thaliana* (Nakamura et al., 2006) were aligned to corresponding structural homologs in Clustal O (Madeira et al., 2019). Consensus α-helices and β-sheets were visualized using ESPript 3.0 (Robert and Gouet, 2014).

Plant constructs

The *GL2* cDNA was cloned into pENTR/D-TOPO using the pENTR/D-TOPO Cloning Kit (Invitrogen). Deletion and missense mutations $gl2^{\Delta ST}$, $gl2^{\Delta HD46}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{K146T}$, $gl2^{K146E}$ and $gl2^{W279R}$ were generated using the Q5 Site–Directed Mutagenesis Kit (E0554S, New England Biolabs) following manufacturer's protocol with primers listed in **Supplemental Table 3**. The $gl2^{\Delta ST}$ construct was used to generate $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146E;\Delta ST}$ and $gl2^{K146T;\Delta ST}$ double mutants. After sequence confirmation, the wild-type and mutant pENTR/D-TOPO plasmids were transferred into the SR54 (*ProGL2:EYFP:GL2*) binary vector (Schrick et al., 2014). Inserts were amplified with gene specific primers using Q5 High Fidelity DNA polymerase (M0491, New England Biolabs), followed by purification of amplicons using the Nucleospin Gel and PCR Clean-up (Machery-Nagel) and *Dpn1* digestion. Finally, in a total volume of 10 µl reaction, the vector and insert DNA fragments were assembled in a ratio of 1:5 using NEBuilder HiFi DNA Assembly (E5520S, New England Biolabs). The method used for generation of $gl2^{L480P}$, $gl2^{E375G;R392M}$, and $gl2^{Amo}$ is described in (Wojciechowska, 2021). The Phusion® Site-Directed Mutagenesis Kit (Thermofisher Scientific) was used to generate the $gl2^{\Delta HD}$ and $gl2^{\Delta HD6;\Delta ZLZ}$ deletion mutants using oligonucleotides listed in **Supplemental Table 3**. After verification by sequencing, the GL2 deletion constructs were restriction digested with *Kpn*I and *Sal*I and ligated into the binary vector SR54 (*proGL2:EYFP:GL2*) at the corresponding sites.

Phenotypic assays and microscopy

Arabidopsis lines were assayed for leaf trichomes, root hairs and seed coat mucilage. Trichomes were quantified on first leaves from 20 individual plants. Leaf samples were placed on 0.8% agar-water plates and trichomes were counted manually upon viewing with a stereomicroscope. For root hair density analysis, seeds were vapor sterilized and sown on 0.5X MS media (Murashige and Skoog, 1962), followed by stratification at 4°C for 5 days. Seedlings were germinated in vertical orientation at 23°C under continuous light. Primary roots were imaged with a stereomicroscope 2-3 days after germination. For seed mucilage analysis, seeds were stained with ruthenium red (R2751, Sigma-Aldrich) following an established protocol (McFarlane, 2014) with slight modification. Briefly, 20-30 seeds per line were placed in a 24-well plate followed by addition of 800 µl 50 mM EDTA, and 2 h incubation with gentle shaking at room temperature. Seeds were washed to remove the EDTA and 800 µl 0.01% ruthenium red was added followed by 2 h incubation with gentle shaking. Seeds were washed with water (pH 7.5) and mounted on glass slides for imaging with a Leica DMIL LED compound microscope. Rosettes, roots, and trichomes were imaged with a Leica M165 FC stereomicroscope. Live imaging was performed on trichomes and developing seeds to detect EYFP expression using the stereomicroscope and Zeiss LSM 880 confocal microscope, respectively. Argon laser excitation of EYFP was at 488 nm, and the 515-550 band pass filter captured emission. Images were processed using Adobe Photoshop C4.

In vitro **transcription and translation**

The full-length *PDF2* and *pdf2*$^{\Delta ST}$ cDNAs were transferred from pENTR/D-TOPO into destination vector pIX-HALO (CD3-1742, ABRC) using Gateway LR Clonase II Enzyme mix (ThermoFisher Scientific). The K107E and K107T mutations were generated using the Q5 Site-Directed Mutagenesis Kit (E0552S, New England Biolabs) with primers listed in **Supplemental Table 3** in pENTR/D-TOPO followed by transfer to pIX-HALO. The Halo fusion proteins were produced using 1.2 µg plasmid DNA of the respective pIX-HALO construct in a volume of 15 µl using the TNT SP6 High-Yield Wheat Germ Protein Expression System (L3260, Promega). Protein expression was confirmed by Western blot with Anti-Halo Tag monoclonal antibody (1:2000) (G9211, Promega).

Electrophoretic mobility shift assay (EMSA)

Cy3-labeled and unlabelled dsDNA probes were generated with oligonucleotides listed in **Supplemental Table 3**. Annealing was performed with 25 µM oligonucleotides in 100 mM Tris-Cl (pH 7.5), 1M NaCl, and 10 mM EDTA at 95°C for 2 min, followed by 57°C for 5 min, 57-37°C transition over 90 min and 37°C for 2 min. EMSA reactions (20 µl) were prepared as previously described (Evens et al., 2017) with the following modifications: 6 µl of *in vitro* translated product was pre-incubated with binding buffer at 28°C for 10 min. Binding reactions were initiated by adding 200 nM of Cy3-labeled dsDNA probe, followed by a 20 min incubation. After electrophoresis at 4°C in a 0.6% agarose gel (1X TBE, pH 8.3) at 150 V for 1 h, or a 7.5% polyacrylamide gel (456-1025, Bio-Rad) at 100 V for 4 h, protein-DNA complexes were analyzed with a Typhoon™ Trio Imager (GE Healthcare) using the 580 nm emission filter for BP30Cy3, 600 PMT voltage, 532 nm green laser and high sensitivity.

ChIP-seq assay

Seeds from EYFP-tagged *GL2*, $gl2^{\Delta ST}$, $gl2^{\Delta HD6;\Delta ZLZ}$, $gl2^{Amo}$ and $gl2^{R392M;E375G}$ and *gl2-5* lines were surface sterilized and grown on 0.8% agar containing 1X MS (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 crosslinked with 1% formaldehyde for 10 min in a vacuum chamber. Glycine was added to a final concentration of 0.125 M, and the reaction was terminated by incubation for 5 min under vacuum. Chromatin was fragmented using a Covaris 220 focused 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 of ChIP-seq data

ChIP-sequencing was performed by Illumina HiSeq 2500 (Illumina, San Diego, CA) at a 50 base single read resolution. Read quality was assessed using FastQC (Andrews, 2010). 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) (v 2.2.3; parameters: -q --phred33 --end-to-end --sensitive -p 16 -t). The mapping rates varied from a high of 97% to a low of 3%. Due to the low mapping rates of some replicates, the replicate aligned ChIP samples were merged (using “view” utility of the SAMtools suite (Li et al., 2009)) for greater coverage. The overall number of mapped reads for the merged ChIP samples varied from 20-70 million reads per sample (40 million reads per sample on average) except for the negative control sample, *gl2-5*, which did not contain express any EYFP-tagged protein, and had only 5 million mapped reads. See **Supplemental Data Set 1**. The merged ChIP samples were analyzed along with their corresponding input samples for significant peaks using MACS (Zhang et al., 2008) (v 1.4.2; parameters: -f BAM -g 0.135e9). Peaks calls passing a false discovery rate (FDR) cutoff of 0.1 were considered for downstream analysis.

Target gene ontology analysis

AgriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) was employed to extract functional annotations for genes whose promoter or intergenic regions were bound by the transcription factors (Du et al., 2010). Singular enrichment analysis (SEA) was applied to determine gene ontology (GO) term enrichment in each sample relative to the *Arabidopsis thaliana* TAIR9 reference. Analysis was performed using Fisher’s exact test ($p < 0.01$) with Bonferroni correction.

Yeast two-hybrid (Y2H)

Full-length *GL2, PDF2, ATML1* cDNAs and the various deletion and missense mutations ($gl2^{\Delta ST}$, $gl2^{\Delta HD}$, $gl2^{\Delta HD46}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{K146T}$, $gl2^{K146E}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146E;\Delta ST}$, $gl2^{K146T;\Delta ST}$) were transferred from pENTR/D-TOPO to pDEST32 bait and pDEST22 prey vectors (Proquest Two-Hybrid System, Invitrogen) using Gateway LR

Clonase II Enzyme mix (ThermoFisher Scientific). Yeast strains Y2H Gold (*MATa trp1-901 leu2-3,112 ura3-52 hi3-200 gal4Δ gal80Δ LY2::GAL1*$_{UAS}$*-Gal1*$_{TATA}$*-Hi3 GAL2*$_{UAS}$*-GAL2*$_{TATA}$*-Ade2 URA3::MEL1*$_{UAS}$*-Mel1*$_{TATA}$ *AUR1-C MEL1*) and Y187 (*MATα ura3-52 hi3-200 ade2-101 trp1-901 leu2-3,112 gal4Δ gal80Δmet-URA3::GAL1*$_{UAS}$*-Gal1*$_{TATA}$*-LacZ MEL1*) were transformed with pDEST32 and pDEST22 constructs, using standard LiAc/PEG transformation, and plated on medium lacking leucine and tryptophan to select for the bait and prey plasmids, respectively. Mating experiments were performed following the manufacturer's instruction in Matchmaker Gold Yeast Two-Hybrid, Clontech. Y2HGold/ pDEST32 and Y187/ pDEST22 were combined on the YPDA media plates to produce diploids from the mating reaction. Diploids selected in media lacking leucine and tryptophan were replica plated to media that selects for activity of the reporter genes *HI3*, *AUR1-C*, and *MEL1.* To test for autoactivation of bait plasmids Y2HGold/ pDEST32 were plated in media lacking leucine and replica plated to media that selects for activity of the reporter genes *HI3* and *AUR1-C* respectively. For analyzing bait protein expression, single colonies of yeast expressing GAL4DBD:GL2 in pDEST32 were grown at 30°C with shaking for 48 h or to saturation in media lacking leucine. Protein was extracted using the sodium hydroxide/trichloroacetic acid (NaOH/TCA) method and expression was monitored by Western Blot using a 1:500 dilution of GAL4 (DBD) RK5C1:sc 510-HRP mouse monoclonal antibody (sc-510HRP, Santa Cruz Biotechnology).

***In vivo* protein stability assay**

Wild-type and mutant seedlings were germinated on 0.5X MS media (Murashige and Skoog, 1962) for 4 days. On day 5, 30 seedlings of each genotype were transferred to 24 well-plates containing 1 ml 0.5X MS supplemented with 50 μM MG132 (474787, Sigma-Aldrich). Plates were sealed with surgical tape and placed on a shaker at 55 rpm for 16 h at 23°C under continuous light. After 16 h, the seedlings were washed 3X with 1 ml fresh media. Cycloheximide (C1988, Sigma-Aldrich) (400 μM in DMSO) or DMSO control was added at 0 h, and harvesting occurred at 0, 2, 4, 5, 8, 10 or 24 h, respectively. Samples of 30 seedlings were flash frozen in liquid nitrogen and stored at -

80°C prior to protein extraction. Tissue was homogenized in liquid nitrogen, and hot SDS buffer (8 M urea, 2% SDS, 0.1 M DTT, 20% glycerol, 0.1 M Tris-Cl pH 6.8, 0.004% bromophenol blue) was added prior to SDS-PAGE and Western blotting. Anti-GFP (1:2000;11814460001, Roche) served as the primary Ab, followed by secondary Goat Anti-Mouse IgG [HRP] (1:3000; A00160, Genscript). Proteins were detected with SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific) using an Azure 300 chemiluminescence imager (Azure Biosystems), and blots were stained with Bio-Safe Coomassie Blue G-250 (Bio-Rad) for loading controls. Band intensities were quantified with ImageJ software.

Statistical Analysis

Trichome differentiation on the first leaves: Total number of trichomes and branching patterns were counted on the first leaves of 20 individual plants per genotype. Values are represented as mean ± SD. Ordinary one-way ANOVA using Tukey's multiple comparisons test was applied to establish significant differences between genotypes ($p < 0.05$).

Root hairs of young seedlings: Numbers of root hairs on primary roots of 4-5 day-old seedlings were counted for 6-10 individuals per genotype. Values are represented as mean ± SD. Ordinary one-way ANOVA using Tukey's multiple comparisons test was applied to establish significant differences ($p < 0.05$).

Protein stability assays: The graphed data represent two or three independent cycloheximide experiments with whole seedlings. Numerical values for protein levels were obtained from band intensity quantification and were normalized to the mock 10 (M10) and mock 24 (M24) samples depending on experiment, which were designated a value of 1.0. Protein half-life was determined by the protein level at a Y-axis value of 0.5. Values are represented as mean ± SD. *$p < 0.05$ indicates statistical significance using a two-tailed *t*-test with unequal variance.

Accession Numbers

The ChIP-seq data is deposited in the NCBI Gene Expression Omnibus (Edgar et al., 2002) under the GEO Series accession number GSE186860

(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSExxx). The *Arabidopsis thaliana* genes described in this study are as follows: GL2 (At1g79840); PDF2 (At4g04890); ATML1 (At4g21750); ANL2 (At4g00730); FWA/HDG6 (At4g25530); HDG1 (At3g61150); HDG2 (At1g05230); HDG3 (At2g32370); HDG4 (At4g17710); HDG5 (At5g46880); HDG7 (At5g52170); HDG8 (At3g03260); HDG9 (At5g17320); HDG10 (At1g34650); HDG11 (At1g73360); HDG12 (At1g17920)

SUPPLEMENTAL INFORMATION

Supplemental Figure 1. The START domain is highly conserved among HD-Zip IV family members. Related to **Figure 1C**.

Supplemental Figure 2. START domain is required for *in vitro* DNA binding as a dimer. Related to **Figure 4C**.

Supplemental Figure 3. Leaf trichome phenotypes of lines used for ChIP-seq. Related to **Figure 4D-G**.

Supplemental Figure 4. ChIP-seq peak distance from transcription start site. Related to **Figure 4F**.

Supplemental Figure 5. START domain deletion impairs GL2 homodimerization. Related to **Figure 5**.

Supplemental Figure 6. Cycloheximide chase experiments reveal that the START domain is critical for protein stability of GL2. Related to **Figure 6.**

Supplemental Figure 7. Cycloheximide chase experiments with double mutants affecting both HD and START domain of GL2. Related to **Figure 6** and **Supplemental Figure 5**.

Supplemental Table 1. EYFP expression is independent of phenotypic rescue.

Supplemental Table 2. Distribution and number of significant peaks from ChIP-seq data. Related to **Figure 4D and 4E**.

Supplemental Table 3. Oligonucleotides used in this study.

Supplemental Data Set 1. Mapping statistics of sequencing reads from ChIP-seq data.

ACKNOWLEDGMENTS

This research was funded by the National Science Foundation (MCB1616818), National Institute of General Medical Sciences of the National Institute of Health under Award no. P20GM103418 (K-INBRE) and P2-RR-17686 (Confocal Microscopy Core, Kansas State University), USDA National Institute of Food and Agriculture Hatch/Multi-State project 1013013, and Johnson Cancer Research Center at Kansas State University. This is contribution no. 21-279-J from the Kansas Agricultural Experiment Station. We thank Joel Sannemann for help with confocal microscopy and Timothy Durrett for valuable scientific input.

AUTHOR CONTRIBUTIONS

T.M., B.S. and K.S. designed the research, analyzed data, and wrote the paper. T.M. B.S., and A.K. performed experiments and analyzed data.T.M., R.L-R., A.W., A.K. and K.A.T. generated transgenic lines. S.G. performed ChIP-seq informatic analyses. T.M. B.S. and K.S. oversaw the study.

REFERENCES

Abe, M., Takahashi, T., and 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.

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

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

Amendt, B.A., Sutherland, L.B., Semina, E.V., and Russo, A.F. (1998). The molecular basis of Rieger syndrome. Analysis of Pitx2 homeodomain protein activities. J Biol Chem **273,** 20066-20072.

Andrews, S. (2010). FastQC: a qualtity control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arce, A.L., Raineri, J., Capella, M., Cabello, J.V., and Chan, R.L. (2011). Uncharacterized conserved motifs outside the HD-Zip domain in HD-Zip subfamily I transcription factors; a potential source of functional diversity. BMC Plant Biol **11,** 42.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22,** 195-201.

Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics **27,** 343-350.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res **42,** W252-258.

Bolduc, N., and Hake, S. (2009). The maize transcription factor KNOTTED1 directly regulates the gibberellin catabolism gene *ga2ox1*. Plant Cell **21,** 1647-1658.

Burglin, T.R., and Affolter, M. (2016). Homeodomain proteins: an update. Chromosoma **125,** 497-521.

Bürglin, T.R. (1994). A comprehensive classification of homeobox genes. (Oxford, UK: Oxford University Press).

Cao, L., Yu, Y., DuanMu, H., Chen, C., Duan, X., Zhu, P., Chen, R., Li, Q., Zhu, Y., and Ding, X. (2016). A novel Glycine soja homeodomain-leucine zipper (HD-Zip) I gene, *Gshdz4*, positively regulates bicarbonate tolerance and responds to osmotic stress in *Arabidopsis*. BMC Plant Biol **16,** 184.

Carlsson, S.R., and Simonsen, A. (2015). Membrane dynamics in autophagosome biogenesis. J Cell Sci **128,** 193-205.

Carrasco, A.E., McGinnis, W., Gehring, W.J., and De Robertis, E.M. (1984). Cloning of an *X. laevis* gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes. Cell **37,** 409-414.

Cazorla, P., Smidt, M.P., O'Malley, K.L., and Burbach, J.P. (2000). A response element for the homeodomain transcription factor Ptx3 in the tyrosine hydroxylase gene promoter. J Neurochem **74,** 1829-1837.

Chen, R., Xu, N., Yu, B., Wu, Q., Li, X., Wang, G., and Huang, J. (2020). The WUSCHEL-related homeobox transcription factor OsWOX4 controls the primary root elongation by activating OsAUX1 in rice. Plant Sci **298,** 110575.

Clark, B.J., and Stocco, D.M. (1995). Expression of the steroidogenic acute regulatory (StAR) protein: A novel LH-induced mitochondrial protein required for the acute regulation of steroidogenesis in mouse leydig tumor cells. Endocrine Res **21,** 243-257.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J **16,** 735-743.

Derelle, R., Lopez, P., Guyader, H.L., and Manuel, M. (2007). Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. Evolution & Development **9,** 212-219.

Di Cristina, M., Sessa, G., Dolan, L., Linstead, P., Baima, S., Ruberti, I., and 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, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res **38,** W64-70.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res **30,** 207-210.

Evens, N.P., Buchner, P., Williams, L.E., and Hawkesford, M.J. (2017). The role of ZIP transporters and group F bZIP transcription factors in the Zn-deficiency response of wheat (*Triticum aestivum*). Plant J **92,** 291-304.

Gage, P.J., and Camper, S.A. (1997). Pituitary homeobox 2, a novel member of the bicoid-related family of homeobox genes, is a potential regulator of anterior structure formation. Hum Mol Genet **6,** 457-464.

Garber, R.L., Kuroiwa, A., and Gehring, W.J. (1983). Genomic and cDNA clones of the homeotic locus Antennapedia in Drosophila. EMBO J **2,** 2027-2036.

Gehring, W.J. (1994). Guidebook to the Homeobox Genes. (Oxford, UK: Oxford University Press).

Gehring, W.J., Müller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y.Q., Otting, G., and Wüthrich, K. (1990). The structure of the homeodomain and its functional implications. Trends Gen **6,** 323-329.

Hirashima, T., Jimbo, H., Kobayashi, K., and Wada, H. (2021). A START domain-containing protein is involved in the incorporation of ER-derived fatty acids into chloroplast glycolipids in *Marchantia polymorpha*. Biochem Biophys Res Commun **534,** 436-441.

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

Iida, H., and Takada, S. (2021). A quarter century history of *ATML1* gene research. Plants (Basel) **10**.

Iyer, L.M., Koonin, E.V., and Aravind, L. (2001). Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins **43,** 134-144.

Kallen, C.B., Billheimer, J.T., Summers, S.A., Stayrook, S.E., Lewis, M., and Strauss, J.F. (1998). Steroidogenic acute regulatory protein (StAR) is a sterol transfer protein. J Biol Chem **273,** 26285-26288.

Kang, S., Shin, K.D., Kim, J.H., and Chung, T. (2018). Autophagy-related (ATG) 11, ATG9 and the phosphatidylinositol 3-kinase control ATG2-mediated formation of autophagosomes in *Arabidopsis*. Plant Cell Rep **37,** 653-664.

Kanno, K., Wu, M.K., Agate, D.S., Fanelli, B.J., Wagle, N., Scapa, E.F., Ukomadu, C., and Cohen, D.E. (2007). Interacting proteins dictate function of the minimal START domain phosphatidylcholine transfer protein/StarD2. J Biol Chem **282,** 30728-30736.

Khosla, A., Paper, J.M., Boehler, A.P., Bradley, A.M., Neumann, T.R., and 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.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B., and Pabo, C.O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution:

A framework for understanding homeodomain-DNA interactions. Cell **63,** 579-590.

Kudo, N., Kumagai, K., Tomishige, N., Yamaji, T., Wakatsuki, S., Nishijima, M., Hanada, K., and 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.

Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240,** 1759-1764.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods **9,** 357-359.

Larance, M., Ahmad, Y., Kirkwood, K.J., Ly, T., and Lamond, A.I. (2013). Global subcellular characterization of protein degradation using quantitative proteomics. Mol Cell Proteomics **12,** 638-650.

Lebel, M., Gauthier, Y., Moreau, A., and Drouin, J. (2001). Pitx3 activates mouse tyrosine hydroxylase promoter via a high-affinity binding site. J Neurochem **77,** 558-567.

Létourneau, D., Lorin, A., Lefebvre, A., Frappier, V., Gaudreault, F., Najmanovich, R., Lavigne, P., and LeHoux, J.-G. (2012). StAR-related lipid transfer domain protein 5 binds primary bile acids. J Lipid Res **53,** 2677-2689.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25,** 2078-2079.

Lin, D., Sugawara, T., Strauss, J., Clark, B., Stocco, D., Saenger, P., Rogol, A., and Miller, W. (1995). Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. Science **267,** 1828-1831.

Liu, B., Kissinger, C.R., Pabo, C.O., Martin-Blanco, E., and Kornberg, T.B. (1990). Crystallization and preliminary X-ray diffraction studies of the engrailed homeodomain and of an engrailed homeodomain/DNA complex. Biochem Biophys Res Comm **171,** 257-259.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res **47,** W636-W641.

Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Chapter 3 Hox Specificity: Unique Roles for Cofactors and Collaborators. In Current Topics in Developmental Biology (Academic Press), pp. 63-101.

Masucci, J.D., Rerie, W.G., Foreman, D.R., Zhang, M., Galway, M.E., Marks, M.D., and Schiefelbein, J.W. (1996). The homeobox gene GLABRA2 is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis* thaliana. Development **122,** 1253-1260.

McFarlane, H.E., Gendre,D. and Western,T.L. (2014). Seed coat ruthenium red staining assay. Bio-Protocol **4**.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J. (1984). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell **37,** 403-408.

Mukherjee, K., Brocchieri, L., and Burglin, T.R. (2009). A comprehensive classification and evolutionary analysis of plant homeobox genes. Mol Biol Evol **26,** 2775-2794.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiologia Plantarum **15,** 473-&.

Nagata, K., Ishikawa, T., Kawai-Yamada, M., Takahashi, T., and Abe, M. (2021). Ceramides mediate positional signals in *Arabidopsis thaliana* protoderm differentiation. Development **148**.

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., and Takahashi, T. (2006). Characterization of the class IV homeodomain-leucine zipper gene family in *Arabidopsis*. Plant Physiol **141,** 1363-1375.

Ogawa, E., Yamada, Y., Sezaki, N., Kosaka, S., Kondo, H., Kamata, N., Abe, M., Komeda, Y., and 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., and Aoyama, T. (2003). Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. Science **300,** 1427-1430.

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J., and Wüthrich, K. (1990). Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution. Embo J **9,** 3085-3092.

Palena, C.M., Gonzalez, D.H., and Chan, R.L. (1999). A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein Hahb-4 with DNA. Biochem J **341 (Pt 1),** 81-87.

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

Prashek, J., Bouyain, S., Fu, M., Li, Y., Berkes, D., and Yao, X. (2017). Interaction between the PH and START domains of ceramide transfer protein competes with phosphatidylinositol 4-phosphate binding by the PH domain. J Biol Chem **292,** 14217-14228.

Rerie, W.G., Feldmann, K.A., and Marks, M.D. (1994). The *GLABRA2* gene encodes a homeo domain protein required for normal trichome development in *Arabidopsis*. Genes Dev **8,** 1388-1399.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C.Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., and L. Yu, G. (2000). *Arabidopsis* Transcription Factors: Genome-wide comparative analysis among eukaryotes. Science **290,** 2105.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res **42,** W320-324.

Roderick, S.L., Chan, W.W., Agate, D.S., Olsen, L.R., Vetting, M.W., Rajashankar, K.R., and Cohen, D.E. (2002). Structure of human phosphatidylcholine transfer protein in complex with its ligand. Nat Struct Biol **9,** 507-511.

Rodriguez-Agudo, D., Ren, S., Hylemon, P.B., Redford, K., Natarajan, R., Del Castillo, A., Gil, G., and Pandak, W.M. (2005). Human StarD5, a cytosolic StAR-related lipid binding protein. J Lipid Res **46,** 1615-1623.

Roostaee, A., Barbar, E., Lavigne, P., and 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 alpha-helix in the gating of the binding site. Biosci Rep **29,** 89-101.

Saadi, I., Kuburas, A., Engle, J.J., and Russo, A.F. (2003). Dominant negative dimerization of a mutant homeodomain protein in Axenfeld-Rieger syndrome. Mol Cell Biol **23,** 1968-1982.

Saadi, I., Semina, E.V., Amendt, B.A., Harris, D.J., Murphy, K.P., Murray, J.C., and Russo, A.F. (2001). Identification of a dominant negative homeodomain mutation in Rieger syndrome. J Biol Chem **276,** 23034-23041.

Sakazume, S., Sorokina, E., Iwamoto, Y., and Semina, E.V. (2007). Functional analysis of human mutations in homeodomain transcription factor PITX3. BMC Mol Biol **8,** 84.

Schena, M., and Davis, R.W. (1994). Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families. Proc Natl Acad Sci U S A **91,** 8393-8397.

Schofield, P.N. (1987). Patterns, puzzles and paradigms: the riddle of the homeobox. Trends Neurosci **10,** 3-6.

Schrick, K., Nguyen, D., Karlowski, W.M., and Mayer, K.F.X. (2004). START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. Genome Biol **5,** R41.

Schrick, K., Bruno, M., Khosla, A., Cox, P.N., Marlatt, S.A., Roque, R.A., Nguyen, H.C., He, C., Snyder, M.P., Singh, D., and Yadav, G. (2014). Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity. BMC Biol **12,** 70.

Scott, M.P., and Weiner, A.J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proc Natl Acad Sci U S A **81,** 4115-4119.

Semina, E.V., Reiter, R.S., and Murray, J.C. (1997). Isolation of a new homeobox gene belonging to the Pitx/Rieg family: expression during lens development and mapping to the Aphakia Region on mouse Chromosome 19. Human Mol Gen **6,** 2109-2116.

Semina, E.V., Reiter, R., Leysens, N.J., Alward, W.L., Small, K.W., Datson, N.A., Siegel-Bartelt, J., Bierke-Nelson, D., Bitoun, P., Zabel, B.U., Carey, J.C., and Murray, J.C. (1996). Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. Nat Genet **14,** 392-399.

Sessa, G., Morelli, G., and Ruberti, I. (1993). The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. EMBO J **12,** 3507-3517.

Shao, J., Haider, I., Xiong, L., Zhu, X., Hussain, R.M.F., Övernäs, E., Meijer, A.H., Zhang, G., Wang, M., Bouwmeester, H.J., and Ouwerkerk, P.B.F. (2018). Functional analysis of the HD-Zip transcription factor genes Oshox12 and Oshox14 in rice. PLoS One **13,** e0199248.

Shen, B., Sinkevicius, K.W., Selinger, D.A., and Tarczynski, M.C. (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., and Haughn, G. (2012). *Arabidopsis glabra2* mutant seeds deficient in mucilage biosynthesis produce more oil. Plant J **69,** 37-46.

Smidt, M.P., van Schaick, H.S., Lanctôt, C., Tremblay, J.J., Cox, J.J., van der Kleij, A.A., Wolterink, G., Drouin, J., and Burbach, J.P. (1997). A homeodomain gene *Ptx3* has highly restricted brain expression in mesencephalic dopaminergic neurons. Proc Natl Acad Sci U S A **94,** 13305-13310.

Soccio, R.E., and Breslow, J.L. (2003). StAR-related lipid transfer (START) proteins: Mediators of intracellular lipid metabolism. J Biol Chem **278,** 22183-22186.

Soccio, R.E., Adams, R.M., Romanowski, M.J., Sehayek, E., Burley, S.K., and Breslow, J.L. (2002). The cholesterol-regulated *StarD4* gene encodes a StAR-related lipid transfer protein with two closely related homologues, StarD5 and StarD6. Proc Natl Acad Sci U S A **99,** 6943-6948.

Stocco, D.M. (2001). StAR protein and the regulation of steroid hormone biosynthesis. Annu Rev Physiol **63,** 193-213.

Szymanski, D.B., Jilk, R.A., Pollock, S.M., and Marks, M.D. (1998). Control of *GL2* expression in *Arabidopsis* leaves and trichomes. Development **125,** 1161-1171.

Thorsell, A.-G., Lee, W.H., Persson, C., Siponen, M.I., Nilsson, M., Busam, R.D., Kotenyova, T., Schüler, H., and Lehtiö, L. (2011). Comparative structural analysis of lipid binding START domains. PLoS One **6,** e19521-e19521.

Tillman, M.C., Imai, N., Li, Y., Khadka, M., Okafor, C.D., Juneja, P., Adhiyaman, A., Hagen, S.J., Cohen, D.E., and Ortlund, E.A. (2020). Allosteric regulation of thioesterase superfamily member 1 by lipid sensor domain binding fatty acids and lysophosphatidylcholine. Proc Natl Acad Sci U S A **117,** 22080-22089.

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

Tron, A.E., Bertoncini, C.W., Chan, R.L., and Gonzalez, D.H. (2002). Redox regulation of plant homeodomain transcription factors. J Biol Chem **277,** 34800-34807.

Tron, A.E., Bertoncini, C.W., Palena, C.M., Chan, R.L., and Gonzalez, D.H. (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.

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

Wang, X., Wang, X., Hu, Q., Dai, X., Tian, H., Zheng, K., Wang, X., Mao, T., Chen, J.G., and Wang, S. (2015). Characterization of an activation-tagged mutant

uncovers a role of GLABRA2 in anthocyanin biosynthesis in *Arabidopsis*. Plant J **83,** 300-311.

Watari, H., Arakane, F., Moog-Lutz, C., Kallen, C.B., Tomasetto, C., Gerton, G.L., Rio, M.C., Baker, M.E., and Strauss, J.F., 3rd. (1997). MLN64 contains a domain with homology to the steroidogenic acute regulatory protein (StAR) that stimulates steroidogenesis. Proc Natl Acad Sci U S A **94,** 8462-8467.

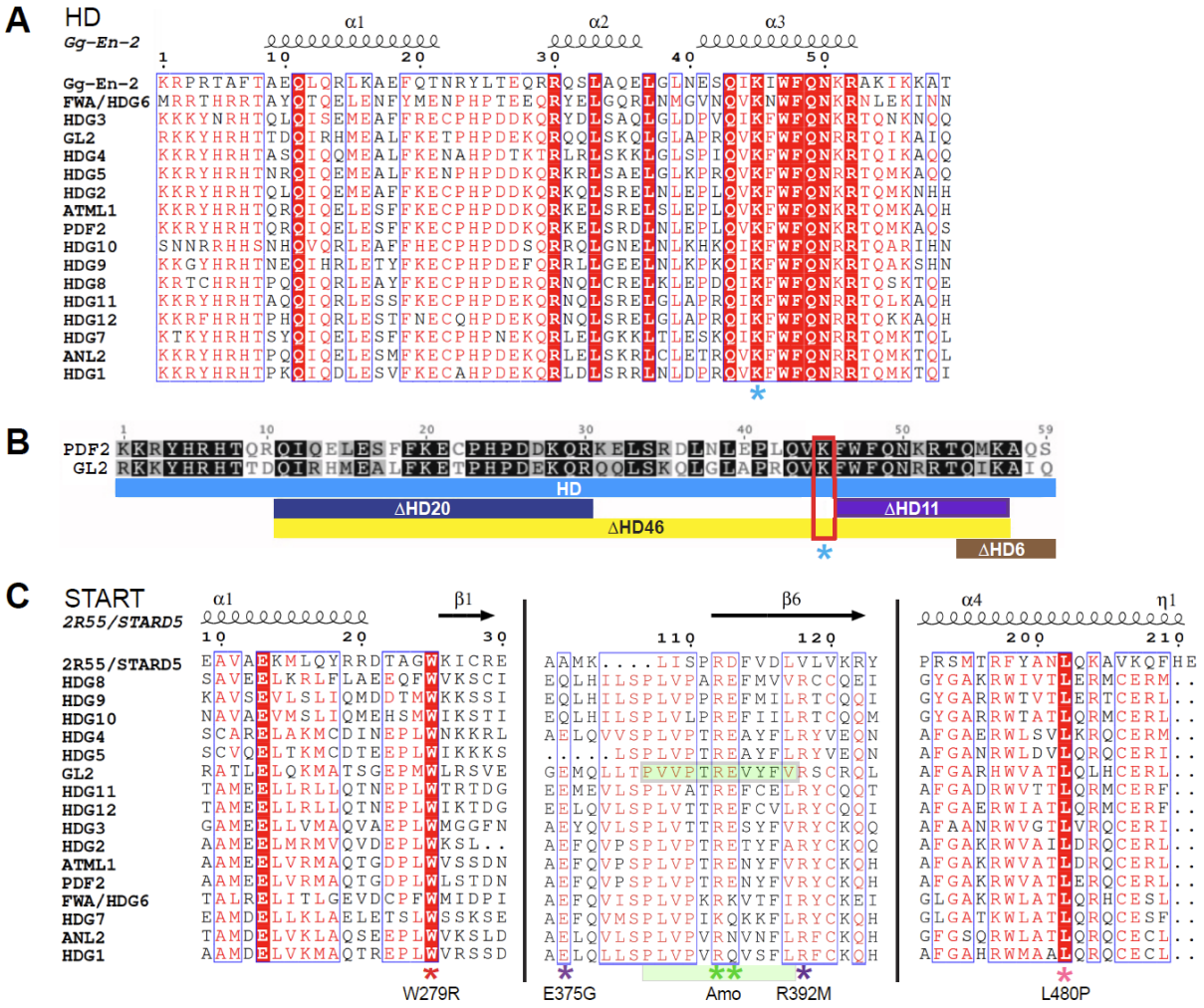
Western, T.L., Young, D.S., Dean, G.H., Tan, W.L., Samuels, A.L., and Haughn, G.W. (2004). MUCILAGE-MODIFIED4 encodes a putative pectin biosynthetic enzyme developmentally regulated by APETALA2, TRANSPARENT TESTA GLABRA1, and GLABRA2 in the *Arabidopsis* seed coat. Plant Physiol **134,** 296-306.

Western, T.L., Burn, J., Tan, W.L., Skinner, D.J., Martin-McCaffrey, L., Moffatt, B.A., and Haughn, G.W. (2001). Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in *Arabidopsis*. Plant Physiol **127,** 998-1011.

Wojciechowska, I., Mukherjee, T., Knox-Brown, P., Hu, X., Khosla, A., Mathews, G.L., Thompson, K.A., Peery, S.T., Szlachetko, J., Thalhammer, A., Hincha, D.K., Skirycz, A., Schrick, K. (2021). *Arabidopsis* PROTODERMAL FACTOR2 binds lysophosphatidylcholines and transcriptionally regulates phospholipid metabolism. bioRxiv doi: 10.1101/2021.10.20.465175

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nat Methods **12,** 7-8.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol **9,** R137.


A
HD
Gg-En-2
α1
α2
α3
Gg-En-2
FWA/HDG6
HDG3
GL2
HDG4
HDG5
HDG2
ATML1
PDF2
HDG10
HDG9
HDG8
HDG11
HDG12
HDG7
ANL2
HDG1
B
PDF2
GL2
HD
ΔHD20
ΔHD46
ΔHD11
ΔHD6
C
START
2R55/STARD5
α1
β1
β6
α4
η1
2R55/STARD5
HDG8
HDG9
HDG10
HDG4
HDG5
GL2
HDG11
HDG12
HDG3
HDG2
ATML1
PDF2
FWA/HDG6
HDG7
ANL2
HDG1
W279R
E375G
Amo
R392M
L480P
D
GL2
HD
ZLZ
START
STAD
747
gl2ΔHD
686
gl2ΔHD6;ΔZLZ
677
gl2ΔHD20
727
gl2ΔST
512
gl2ΔHD20;ΔST
492
K146E
gl2K146E
747
K146T
gl2K146T
747
W279R
gl2W279R
747
R386L;E387L;11
gl2Amo
758
L480P
gl2L480P
747
E375G
R392M
gl2E375G;R392M
747

1039

Figure 1. Schematic of GL2 wild type and mutant proteins.

(A) Structural alignment of homeodomain (HD) across all 16 members of the *Arabidopsis* HD-Zip IV transcription factor family with the solution structure for *Gallus gallus* HOMEOBOX PROTEIN ENGRAILED-2 (Gg-En-2) (PDB ID = 3ZOB) as reference. Consensus α-helices are displayed using ESPript 3.0 (Robert and Gouet, 2014). Similar amino acid residues are enclosed in blue boxes with identical residues highlighted in red, and numbering (1-59) for the HD only. Lysine targeted for missense mutations (K-E or K-T) is indicated (blue asterisk)

(B) Amino acid alignment of GL2 and PDF2 HD domains generated by Geneious 7.0.6 (Biomatters). Identical and similar amino acid residues are indicated by black and grey shading. HD deletions are shown as blue, purple, yellow, or brown tracts. Lysine targeted for missense mutations (K-E or K-T) is indicated by red box (blue asterisk).

(C) Structural alignment of segments of the START domain across the *Arabidopsis* HD-Zip IV family with the crystal structure for *Homo sapiens* STARD5 (PDB ID = 2R55) as reference. Consensus α-helices and β-sheets are displayed using ESPript 3.0 (Robert and Gouet, 2014). Similar amino acid residues are enclosed in blue boxes with identical residues highlighted in red. Positions of gl2 mutants analyzed in Chip-seq (E375G;R392M, Amo) (**Figure 4D-G**) and cycloheximide assays (W279R, Amo, L480P) (**Figure 6; Supplemental Figure 5**) are indicated below alignment (asterisks). See **Supplemental Figure 1** for full-length START domain and corresponding numbering of amino acids (1-211).

(D) Wild-type GL2 and mutant proteins $gl2^{\Delta HD}$ and $gl2^{\Delta START}$ with homeodomain (HD), Zipper-Loop-Zipper (ZLZ), Steroidogenic Acute regulatory protein-Related lipid Transfer (START) and START-Associated Domain (STAD) respectively. Lengths of proteins are indicated in amino acids (right). Vertical lines indicate positions of amino acid changes in the mutants.

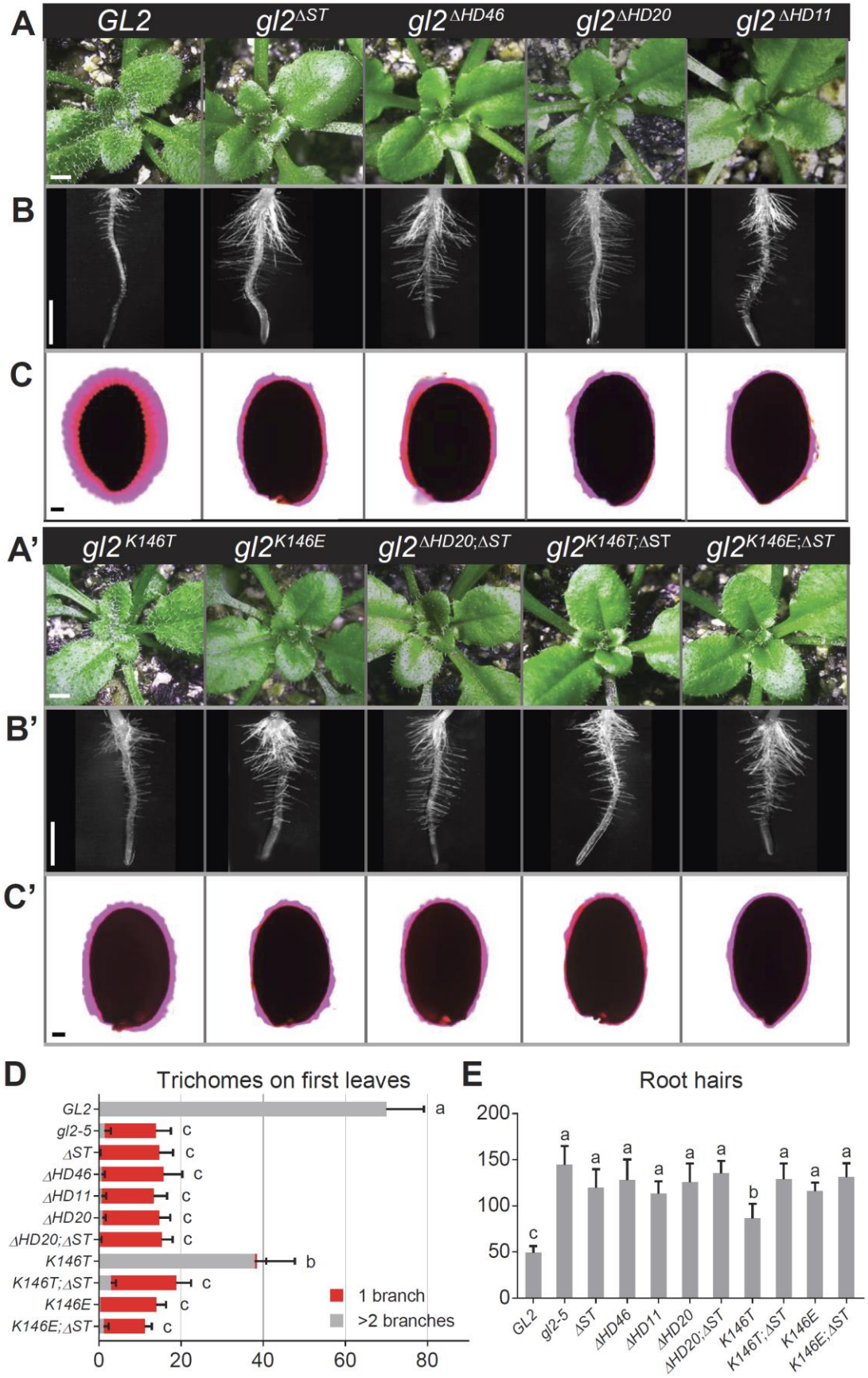

A
GL2
gl2ΔST
gl2ΔHD46
gl2ΔHD20
gl2ΔHD11
B
C
A'
gl2K146T
gl2K146E
gl2ΔHD20;ΔST
gl2K146T;ΔST
gl2K146E;ΔST
B'
C'
D
Trichomes on first leaves
GL2
gl2-5
ΔST
ΔHD46
ΔHD11
ΔHD20
ΔHD20;ΔST
K146T
K146T;ΔST
K146E
K146E;ΔST
1 branch
>2 branches
0
20
40
60
80
E
Root hairs
200
150
100
50
0
GL2
gl2-5
ΔST
ΔHD46
ΔHD11
ΔHD20
ΔHD20;ΔST
K146T
K146T;ΔST
K146E
K146E;ΔST

Figure 2. Leaf trichome, root hair and seed mucilage phenotypes of plants expressing wild-type GL2 and mutant gl2 proteins.

(A) and **(A')** Rosettes were observed for the presence of leaf trichomes. Wild-type *EYFP:GL2* plants display trichomes covering leaf surfaces. In contrast, mutants ($gl2^{\Delta ST}$, $gl2^{\Delta HD46}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{K146E}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146T;\Delta ST}$, $gl2^{K146E;\Delta ST}$) exhibit trichome differentiation defects. $gl2^{K146T}$ displays an intermediate phenotype. Bar = 1 mm.

(B) and **(B')** Root hair phenotypes of 4-5 day-old seedlings. In contrast to wild-type *GL2* and the partial phenotype of $gl2^{K146T}$, all other mutants ($gl2^{\Delta ST}$, $gl2^{\Delta HD46}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{K146E}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146T;\Delta ST}$, $gl2^{K146E;\Delta ST}$) exhibit excessive root hair formation. Bar = 1 mm.

(C) and **(C')** Ruthenium red staining of seeds to assay mucilage production. *GL2* seeds display a normal mucilage layer, while $gl2^{\Delta ST}$, $gl2^{\Delta HD46}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{K146E}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146T;\Delta ST}$ and $gl2^{K146E;\Delta ST}$ exhibit defective mucilage production. $gl2^{K146T}$ displays a partial phenotype. Bar = 100 µm.

(D) Total number and branching pattern of trichomes on the first leaves. Gray bars indicate trichomes with 2-5 branches in the wild-type EYFP-tagged GL2. Mutant plants of the genotypes *gl2-5*, $gl2^{\Delta ST}$, $gl2^{\Delta HD46}$, $gl2^{\Delta HD11}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146T;\Delta ST}$, $gl2^{K146E}$ and $gl2^{K146E;\Delta ST}$ exhibit unbranched trichomes, indicated by red bars. $gl2^{K146T}$ plants display a partial trichome differentiation defects. Values are mean ± SD for n = 20 plants. Letters denote significant differences from ordinary one-way ANOVA ($p < 0.05$) using Tukey's multiple comparisons test.

(E) Numbers of root hairs on primary roots of 4-5 day-old seedlings. Genotypes correspond to those in **(D)**. All the *gl2* mutants display increases in the number of root hairs. The $gl2^{K146T}$ seedlings exhibit an intermediate phenotype. Values are mean ± SD for n = 6-10 seedlings. Letters denote significant differences from ordinary one-way ANOVA ($p < 0.05$) using Tukey's multiple comparisons test.

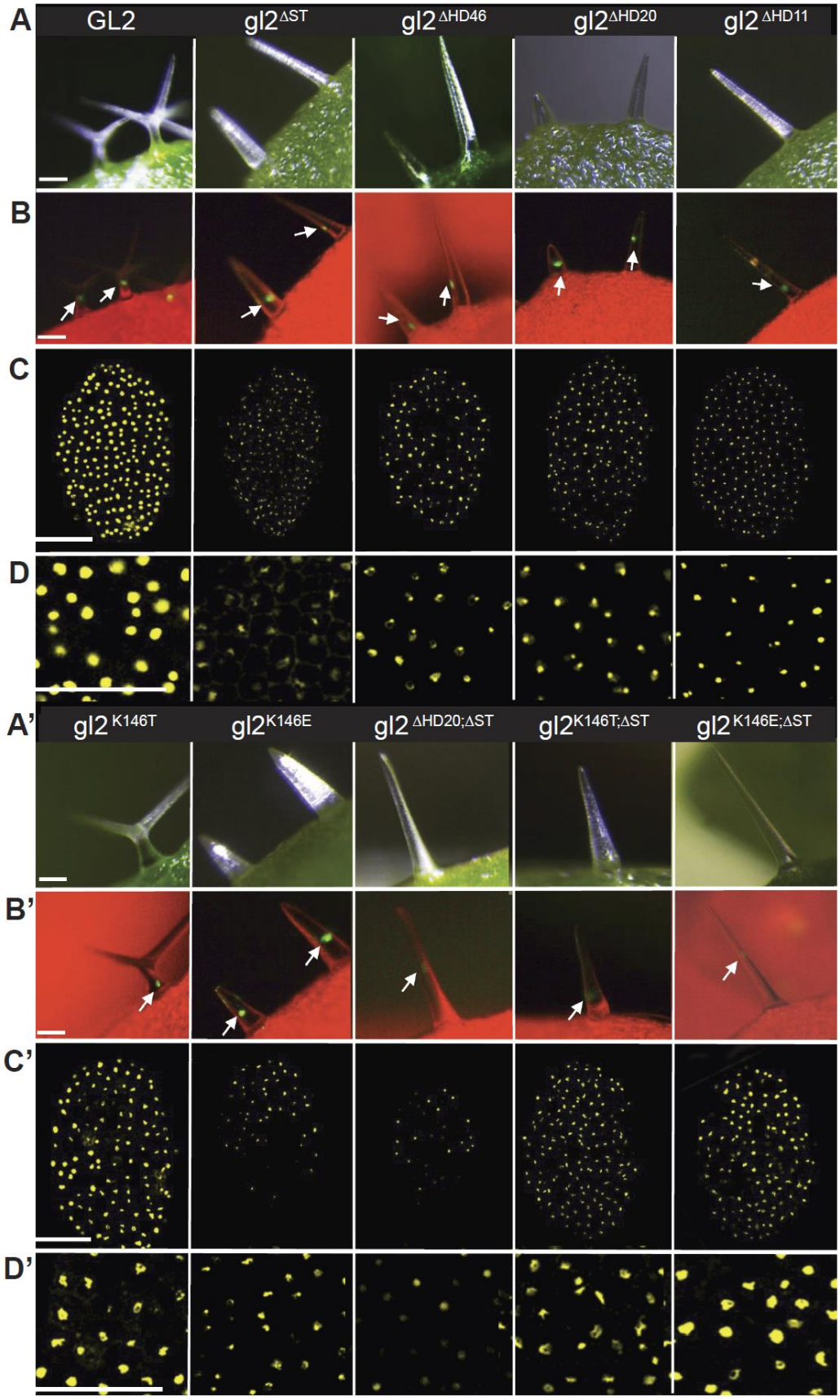

A
GL2
gl2ΔST
gl2ΔHD46
gl2ΔHD20
gl2ΔHD11
B
C
D
A'
gl2K146T
gl2K146E
gl2ΔHD20;ΔST
gl2K146T;ΔST
gl2K146E;ΔST
B'
C'
D'

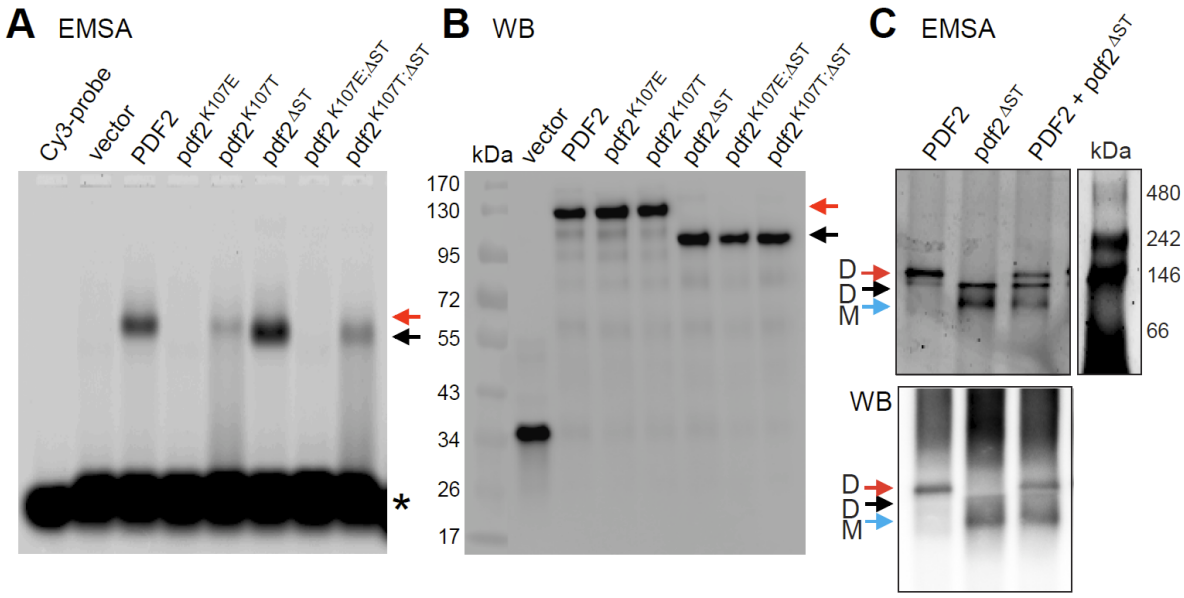
Figure 3. Nuclear localization of wild-type EYFP-tagged GL2 and mutant gl2 transcription factors.

(A) and **(A')** Imaging of leaf trichomes under white light (green, chlorophyll). Bar = 100 µm.

(B) and **(B')** Matching fluorescence images (red, chlorophyll) indicate nuclear localization of EYFP-tagged GL2, $gl2^{\Delta ST}$, $gl2^{\Delta HD46}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{K146T}$, $gl2^{K146E}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146T;\Delta ST}$ and $gl2^{K146E;\Delta ST}$ proteins in trichome cells. Arrows mark the brightest nuclear signal in each image. Bar = 100 µm.

(C) and **(C')** Confocal laser scanning microscopy of developing seeds reveals nuclear expression of EYFP-tagged wild-type and mutant proteins in seed coat cells. Bar = 100 µm.

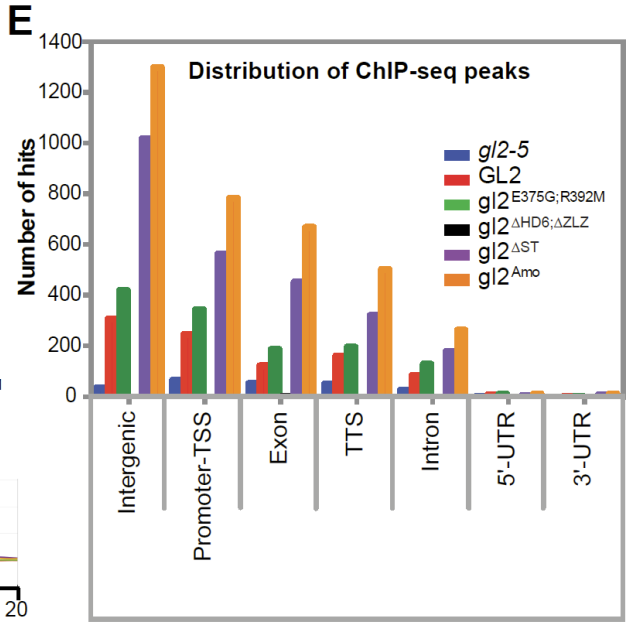
(D) and **(D')** Magnified images from **(C)** and **(C')**. Bar = 50 µm.

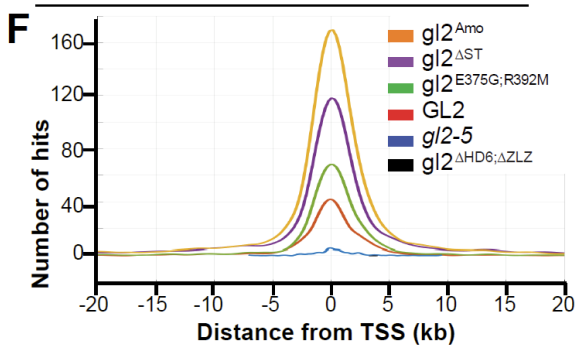


D

Number of significant peaks from ChIP-seq

Sample		Peaks (FDR < 0.1)
Wild type	*EYFP:GL2*	925
Mutants	*EYFP:gl2*^{Amo}	3533
	EYFP:gl2$^{\Delta ST}$	2541
	EYFP:gl2$^{E375G;R392M}$	1284
	EYFP:gl2$^{\Delta HD6;\Delta ZLZ}$	2
	gl2-5	227





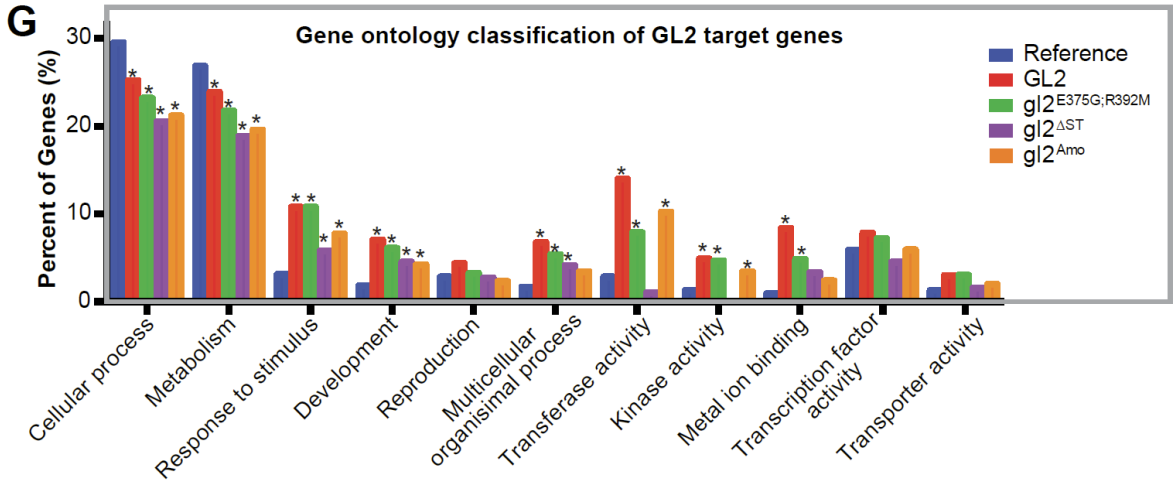


Figure 4. HD is required for DNA binding activity of PDF2 and GL2 while START domain is dispensable for DNA binding.

(A) HD is required for *in vitro* DNA binding of PDF2 and pdf2$^{\Delta ST}$ proteins. Electrophoretic mobility shift assays (EMSA) reveal band shifts for Halo:PDF2 (red arrow) and Halo:pdf2$^{\Delta ST}$(black arrow), but not for proteins harboring HD mutation K107E. The K107T mutation results in partial DNA-protein complex formation indicated by a mixed population of bound and smeared molecules. Migration of the Cy3-labeled probe containing the L1-box TAAATGTA is indicated (asterisk).

(B) Western blot with anti-Halo Ab detects Halo-tagged proteins assayed for DNA binding in (**A**).

(C) The ΔSTART protein binds DNA as a monomer. EMSA coupled with polyacrylamide gel electrophoresis reveals that Halo:PDF2 binds the Cy3-labeled L1-box probe as a dimer (~230 kDa) (red arrow) while Halo:pdf2$^{\Delta ST}$ binds DNA as a monomer (~90 kDa) (blue arrow) and weakly as a dimer (~180 kDa) (black arrow). Western blot (WB) (bottom) of the polyacrylamide gel with anti-Halo Ab detected Halo-tagged PDF2, and pdf2$^{\Delta ST}$ at the expected migration positions matching that of the EMSA (top). See **Supplemental Figure 2**.

(D-G) Chip-seq data reveals that wild-type GL2 and mutants with compromised START domain function (gl2$^{\Delta ST}$, gl2^{Amo}, gl2$^{E375G;R392M}$) are proficient in DNA binding *in vivo*. In contrast, DNA binding was abolished in the gl2$^{\Delta HD6;\Delta ZLZ}$ mutant. See **Supplemental Figure 3**.

(D) Number of significant peaks (FDR < 0.1) from ChIP-seq data indicate that START domain mutants, but not the HD-ZLZ mutant, bind DNA target sites *in vivo*.

(E) Genomic distribution of ChIP-seq binding events for wild-type GL2 and mutant versions are indicated (FDR < 0.1). TSS, transcription start site; TTS, transcription termination site; UTR, untranslated region.

(F) ChIP-Seq hit distribution from gene transcription start sites (TSS). The distribution of significant ChIP-seq peaks (FDR < 0.1) upstream and downstream from the TSS is shown for wild-type GL2 and mutants. See **Supplemental Figure 4**.

(G) Functional classification of target genes from ChIP-seq data indicate over-represented GO terms from target genes whose promoters were bound by wild-type GL2 and mutants $gl2^{\Delta ST}$, $gl2^{Amo}$, and $gl2^{E375G;R392M}$. Genes were classified based on molecular function and biological process, and a given gene may be represented in more than one category. Enrichment analysis was performed using AgriGO (Du et al., 2010). Asterisks indicate significant differences from the *Arabidopsis* genome reference (blue) (Fisher's exact test, $p < 0.01$).

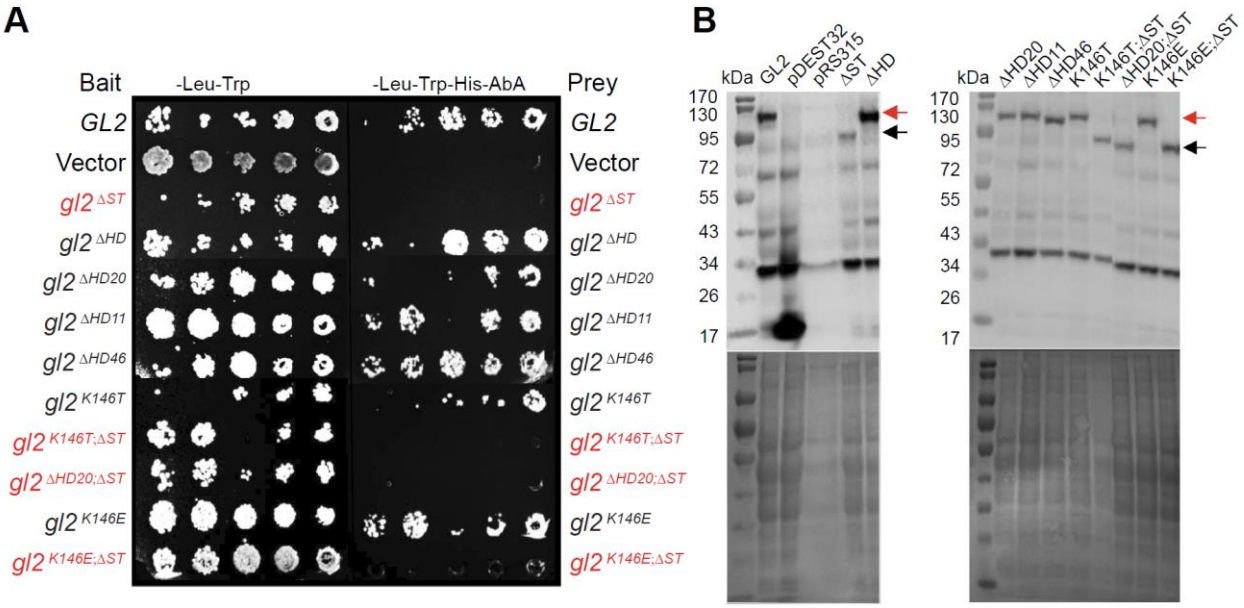


Figure 5. Homodimerization of GL2 is not affected by HD mutations while START domain deletion impairs homodimerization.

(A) Y2H assays show that bait and prey fusion proteins interact to homodimerize only when START domain is present. Bait and prey constructs were transformed into yeast strains of opposite mating type. Haploid yeast were mated and diploids were selected on -Leu-Trp followed by selection on -Leu-Trp-His AbA to assay homodimerization. HD mutants were proficient in homodimerization. START domain deletion resulted in loss of homodimerization in absence or presence of HD mutations (red). Wild-type GL2 bait and prey served as positive controls while empty bait and prey vectors represent negative controls. See **Supplemental Figure 5**.

(B) Western blotting with anti-GAL4(DBD) confirms expression of bait proteins in **(A)**. Migration of wild-type and HD mutant proteins (black arrows) and START deleted proteins (red arrows) are indicated. pDEST32 is the vector control and pRS315 represents a second negative control. Coomassie blue staining after Western blotting indicates protein loading (bottom).

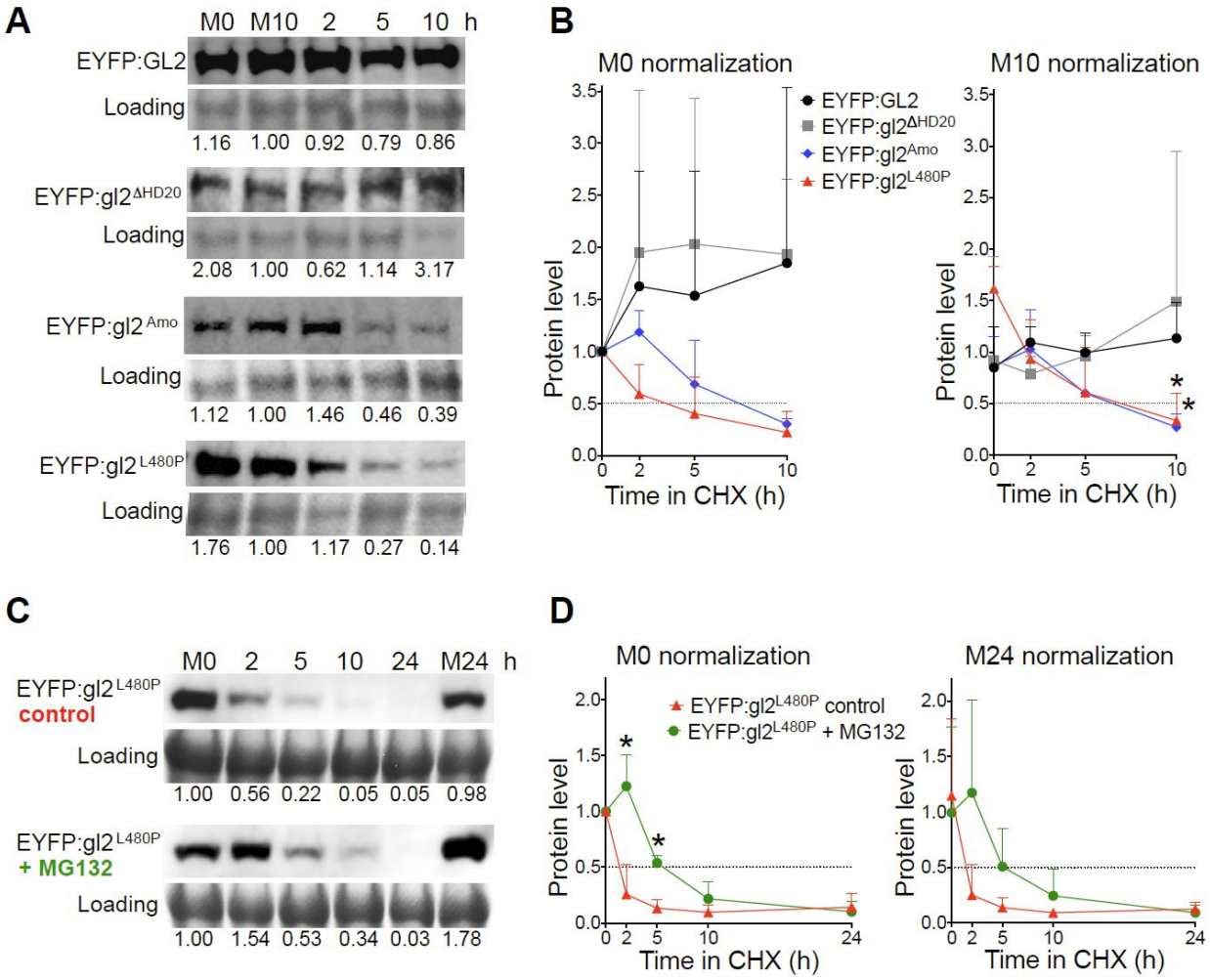


Figure 6. Cycloheximide experiments reveal that the START domain, but not HD, is critical for protein stability of GL2.

(A-B) HD mutant gl2$^{\Delta HD20}$ behaves similarly to wild-type GL2, whereas START domain mutants gl2^{Amo} and gl2^{L480P} exhibit decreases in protein half-life.

(C-D) Treatment with MG132 results in increased protein stability of START domain mutant gl2^{L480P}.

(A and C) Levels of EYFP-tagged proteins were monitored by Western blotting. Immunodetection with anti-GFP Ab was followed by Coomassie blue staining for loading controls. Mock treatments with DMSO at 0 (M0) and 10 (M10) or 24 (M24) h are indicated. Blots for each genotype are representative of three independent experiments.

(A) Five-day-old seedlings expressing wild-type GL2 or mutant gl2 proteins were treated with cycloheximide (400 μM) for a 10 h time course. Values were normalized to loading controls and M10. See **Supplemental Figure 6 and Supplemental Figure 7**.

(B) Five-day-old seedlings expressing EYFP-tagged $gl2^{L480P}$ were treated simultaneously with cycloheximide (400 μM) and DMSO (control) or proteasome inhibitor MG132 (50 μM) for a 24 h time course. Values were normalized to loading controls and M0.

(B and D) Protein quantification is shown for three Western blot experiments including that depicted in **(A)** and **(C)**, respectively. Values were normalized to M0 (left) or M10 (or M24) (right) controls. Protein half-life (0.5) is indicated by intersection with dotted lines. Error bars indicate SD for means from three experiments. Two-tailed *t*-test with unequal variance: *$p \leq 0.05$.

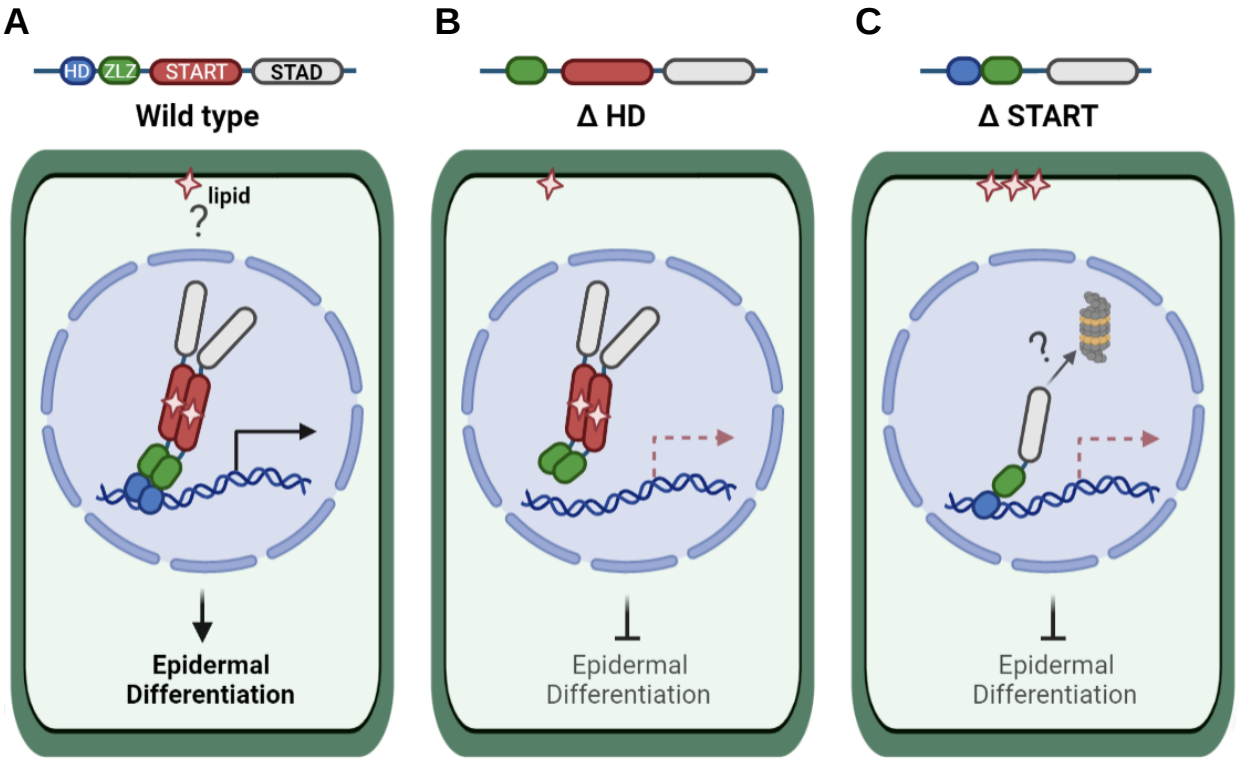


Figure 7. Model of HD and START domain roles in HD-Zip IV transcription factors.

(A) Wild-type HD-Zip IV TFs, as exemplified by GL2, bind a lipid ligand that is possibly retrieved from a membrane compartment. The protein dimerizes and binds to DNA elements in the promoters of target genes, resulting in activation of transcription or repression of gene expression. For simplicity, only activation of gene expression is shown (horizontal arrow). Downstream target genes orchestrate cell differentiation of the plant epidermis.

(B) Deletion or mutation of the HD abolishes DNA binding of the transcription factor, but dimerization and nuclear localization remain intact. Although the protein is found in the nucleus, epidermal defects in the corresponding mutants indicate that it is nonfunctional.

(C) Deletion or mutation of the START domain results in nonfunctional protein that is compromised in lipid binding. Nonetheless, the mutant protein is transported to the nucleus and binds DNA elements. Dimerization of the protein is affected, resulting in monomeric protein that is unstable, possibly due to degradation via a proteasome-dependent pathway.