

Post-translational regulation of HD-Zip IV transcription factors in Arabidopsis

by

Bibek Subedi

B.S., Tribhuvan University, 2012

M.S., Tribhuvan University, 2016

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

Class IV homeodomain leucine-zipper (HD-Zip IV) transcription factors (TFs) are master regulators of epidermal cell fate during embryonic and post-embryonic development in plants. GLABRA2 (GL2), the founding member of the HD-Zip IV TFs, plays critical roles in trichome and root hair development as well as in biosynthesis of anthocyanin, seed coat mucilage and seed oil. It has four defined domains including a DNA-binding homeodomain (HD), a leucine-zipper dimerization domain termed Zipper Loop Zipper (ZLZ), a Steroidogenic Acute Regulatory (StAR) protein-related lipid Transfer (START) domain, and a START-associated domain. Previous work demonstrated that the START domain of GL2 is involved in ligand binding and is required for TF dimerization, stability and activity. However, a detailed mechanism controlling the level of HD-Zip IV TFs in plants remains unknown. The overall goal of this project is to identify and characterize post-translational regulatory mechanisms affecting the turnover and activity of HD-Zip IV TFs in Arabidopsis.

To determine protein half-lives, *in vivo* cycloheximide chase experiments were performed using Arabidopsis seedlings, uncovering a role for the ZLZ domain in destabilizing GL2. In addition, the Ubiquitin/26S proteasome system was implicated in the degradation of unstable GL2 protein resulting from START domain mutation. Next, bioinformatics tools predicted multiple, high confidence conserved ubiquitination and SUMOylation sites in the ZLZ domain of GL2 and the modifications were verified *in vivo* by performing immunoprecipitation experiments. Site-directed mutagenesis of the candidate amino acids resulted in a gain-of-function phenotype characterized by an increased frequency of abaxial leaf curling in lysine-to-arginine (K-R) mutants. In contrast to EYFP-tagged protein, plants expressing GL2 with a smaller 6x His 3x FLAG tag did not display leaf curling, suggesting that this phenotype is

associated with increased protein stability. Ongoing experiments focus on RNA-Seq experiments to characterize the molecular mechanisms of abaxial leaf curling.

In addition to ubiquitin and SUMO (small ubiquitin-like modifier) sites, two high confidence SUMO-interacting motifs (SIMs) were bioinformatically predicted in the GL2 START domain from *Arabidopsis*. SIMs are short hydrophobic sequences present in many proteins that can non-covalently bind to SUMO-conjugated proteins. Mutational analysis with confocal microscopy revealed defects in trichome branching and nuclear localization of the protein in SIM mutants. Yeast two-hybrid experiments failed to detect GL2-SUMO interactions as most of the *Arabidopsis thaliana* SUMO isoforms were not expressed in yeast. However, unlike wild-type protein, SIM mutants displayed impaired dimerization. These results indicate the role of SIMs in the subcellular localization and TF activity of GL2. Further experiments will be required to identify proteins that associate with GL2 through SUMO-SIM interactions.

Overall, the findings suggest that regulatory mechanisms involving ubiquitin and/or SUMO contribute to GL2 levels and activity in plants. This work provides new insight into how the protein levels of critical transcriptional regulators of epidermal cell differentiation are fine-tuned in plants. The new knowledge will ultimately guide molecular toolkits for engineering HD-Zip IV TFs to establish more resilient and high yielding cultivars.

Post-translational regulation of HD-Zip IV transcription factors in Arabidopsis

by

Bibek Subedi

B.S., Tribhuvan University, 2012

M.S., Tribhuvan University, 2016

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Kathrin Schrick

Copyright

© Bibek Subedi 2022.

Abstract

Class IV homeodomain leucine-zipper (HD-Zip IV) transcription factors (TFs) are master regulators of epidermal cell fate during embryonic and post-embryonic development in plants. GLABRA2 (GL2), the founding member of the HD-Zip IV TFs, plays critical roles in trichome and root hair development as well as in biosynthesis of anthocyanin, seed coat mucilage and seed oil. It has four defined domains including a DNA-binding homeodomain (HD), a leucine-zipper dimerization domain termed Zipper Loop Zipper (ZLZ), a Steroidogenic Acute Regulatory (StAR) protein-related lipid Transfer (START) domain, and a START-associated domain. Previous work demonstrated that the START domain of GL2 is involved in ligand binding and is required for TF dimerization, stability and activity. However, a detailed mechanism controlling the level of HD-Zip IV TFs in plants remains unknown. The overall goal of this project is to identify and characterize post-translational regulatory mechanisms affecting the turnover and activity of HD-Zip IV TFs in Arabidopsis.

To determine protein half-lives, *in vivo* cycloheximide chase experiments were performed using Arabidopsis seedlings, uncovering a role for the ZLZ domain in destabilizing GL2. In addition, the Ubiquitin/26S proteasome system was implicated in the degradation of unstable GL2 protein resulting from START domain mutation. Next, bioinformatics tools predicted multiple, high confidence conserved ubiquitination and SUMOylation sites in the ZLZ domain of GL2 and the modifications were verified *in vivo* by performing immunoprecipitation experiments. Site-directed mutagenesis of the candidate amino acids resulted in a gain-of-function phenotype characterized by an increased frequency of abaxial leaf curling in lysine-to-arginine (K-R) mutants. In contrast to EYFP-tagged protein, plants expressing GL2 with a smaller 6x His 3x FLAG tag did not display leaf curling, suggesting that this phenotype is

associated with increased protein stability. Ongoing experiments focus on RNA-Seq experiments to characterize the molecular mechanisms of abaxial leaf curling.

In addition to ubiquitin and SUMO (small ubiquitin-like modifier) sites, two high confidence SUMO-interacting motifs (SIMs) were bioinformatically predicted in the GL2 START domain from *Arabidopsis*. SIMs are short hydrophobic sequences present in many proteins that can non-covalently bind to SUMO-conjugated proteins. Mutational analysis with confocal microscopy revealed defects in trichome branching and nuclear localization of the protein in SIM mutants. Yeast two-hybrid experiments failed to detect GL2-SUMO interactions as most of the *Arabidopsis thaliana* SUMO isoforms were not expressed in yeast. However, unlike wild-type protein, SIM mutants displayed impaired dimerization. These results indicate the role of SIMs in the subcellular localization and TF activity of GL2. Further experiments will be required to identify proteins that associate with GL2 through SUMO-SIM interactions.

Overall, the findings suggest that regulatory mechanisms involving ubiquitin and/or SUMO contribute to GL2 levels and activity in plants. This work provides new insight into how the protein levels of critical transcriptional regulators of epidermal cell differentiation are fine-tuned in plants. The new knowledge will ultimately guide molecular toolkits for engineering HD-Zip IV TFs to establish more resilient and high yielding cultivars.

Table of Contents

List of Figures	xiii
List of Tables	xvi
List of Abbreviations	xvii
Acknowledgements	xix
Dedication	xxi
Chapter 1 - Introduction.....	1
Homeodomain and homeodomain transcription factors.....	1
HD-Zip transcription factors.....	2
Roles of HD-Zip IV transcription factors in plants	2
Research questions, objectives and hypotheses	3
Outline of dissertation.....	4
References.....	7
Chapter 2 - START domain mediates Arabidopsis GLABRA2 dimerization and turnover	
independently of homeodomain DNA binding.....	11
Abstract	11
Introduction.....	12
Materials and methods	15
Plant material, growth conditions, and <i>Arabidopsis</i> transformation.....	15
HD and START domain structural alignments.....	16
Plant constructs	16
Phenotypic assays and microscopy of <i>Arabidopsis</i>	17
<i>In vitro</i> transcription and translation	18
Electrophoretic mobility shift assay (EMSA).....	18
ChIP-seq assay	19
Bioinformatic analysis of ChIP-seq data	20
Target gene ontology analysis.....	20
Yeast two-hybrid (Y2H)	21
Y2H quantitative β -galactosidase assay.....	22
Bimolecular fluorescence complementation (BiFC) in <i>Nicotiana benthamiana</i>	23

<i>In vivo</i> protein stability assay.....	24
Statistical Analysis.....	24
Accession Numbers	25
Results.....	26
HD and START domain are highly conserved across the HD-Zip IV TFs	26
Mutations affecting the HD of GL2 exhibit defects in leaf, root and seed epidermis	27
HD and START domain mutants retain nuclear localization of GL2.....	28
HD, but not START domain, is required for DNA binding <i>in vitro</i>	30
START domain is dispensable for DNA binding <i>in vivo</i>	31
START domain is required for GL2 homodimerization, while the HD is dispensable for this activity.....	33
START domain, but not HD, regulates GL2 protein stability	34
Discussion.....	36
Domain organization in HD-Zip IV TFs and nonoverlapping roles of HD and START	36
Conserved role of HD in DNA binding	37
START domain is critical for protein stability	37
Model for role of HD and START domain in promoting HD-Zip IV TF activity	38
Other possible domain interactions in HD-Zip IV TFs	39
Figures	41
.....	43
.....	45
.....	47
.....	50
References.....	55
Chapter 3 - Post-translational regulation of HD-Zip IV transcription factors in Arabidopsis.....	64
Abstract.....	64
Introduction.....	65
Materials and methods	69
Plant materials and growth conditions.....	69
Generation of DNA constructs.....	70
Arabidopsis transgenics	71

Bioinformatical analysis	71
Protein extraction and immunoblot analysis.....	72
Cycloheximide chase experiments.....	73
Pull-down experiments	74
Phenotypic characterization and microscopy of Arabidopsis.....	75
Statistical analysis.....	76
Leaf biometry and trichome count.....	76
Protein stability assays.....	76
Results.....	77
GL2 protein level is negatively regulated by the ZLZ in Arabidopsis	77
The ZLZ downregulates GL2 levels independently of START mutations.....	78
ZLZ is dispensable for nuclear localization but indispensable for activity of GL2.....	79
Role of UPS in the degradation of HD-Zip IV TFs is more evident under conditions affecting START domain lipid binding	80
GL2 is weakly modified by ubiquitination in young seedlings	82
GL2 contains multiple, conserved predicted Ubiquitination and SUMOylation sites.....	84
Genetic analyses of the predicted PTM sites in GL2 to verify their functions in vivo.....	86
Mutation of putative modification sites cause phenotypic changes in Arabidopsis	87
K-R and C-A mutants retain nuclear localization of GL2 with altered protein stability	89
Discussion.....	90
Role of ZLZ in GL2 turnover in addition to dimerization.....	90
Challenges associated with detection of GL2 PTMs in vivo	92
Insights from bioinformatic analysis of PTM sites.....	94
Association between GL2 levels and abaxial leaf curling	95
Proposed model.....	99
Conclusion and implications.....	99
Figures	101
References.....	116
Chapter 4 - EYFP fusions to HD-Zip IV transcription factors enhance their stability and lead to phenotypic changes in <i>Arabidopsis</i>	135
Abstract.....	135

Introduction.....	136
Materials and methods	139
Plant materials and growth conditions	139
Generation of transgenic lines	139
In vivo protein stability assay	140
Phenotypic assays, leaf biometry and microscopy.....	140
Results.....	141
EYFP fusions to HD-Zip IV TFs enhance their protein stability	141
Expression of EYFP:GL2 results in downward leaf curling in <i>Arabidopsis</i>	143
Characterization of the curly leaf trait associated with EYFP:GL2 expression	144
Discussion	145
Figures	149
.....	153
References.....	155
Chapter 5 - The SUMO-interacting motifs of the START domain are required for GLABRA2	
transcription factor activity	164
Abstract.....	164
Introduction.....	165
Materials and methods	169
Plant materials and growth conditions	169
Generation of constructs and transgenic lines.....	169
Machine learning bioinformatic tools	170
Western blotting.....	171
Phenotypic assays and microscopy	172
Yeast two-hybrid (Y2H) experiments.....	172
Accession Numbers	173
Results.....	174
Two conserved SUMO-interacting motifs (SIMs) are predicted in the GL2 START domain	
.....	174
GL2 SIMs can be genetically interrogated without affecting the predicted tertiary structure	
of the START domain.....	175

Mutations affecting the GL2 SIMs cause defects in epidermal differentiation	176
Nuclear localization of GL2 is impaired in SIM mutants	177
GL2 SIM-SUMO interactions are not detected in Y2H assays	178
Homodimerization of GL2 is impaired in SIM mutants	179
Discussion	180
Figures	185
References	195
Chapter 6 - Perspectives.....	206
Conclusions and future directions.....	206
References	210
Appendix A - Chapter 2 Supplemental Data	212
Supplemental Figures	212
Supplemental Tables.....	227
ChIP-seq Data Set 1. Mapping statistics of sequencing reads from ChIP-seq data.....	232
ChIP-seq Data Set 2. Functional descriptions of GL2 targets from ChIP-seq data.....	232
References	233
Appendix B - Chapter 3 Supplemental Data.....	235
Supplemental Figures	235
Supplemental Tables.....	241
Appendix C - Chapter 5 Supplemental Data.....	244
Supplemental Figures.....	244
Supplemental Tables.....	247

List of Figures

Figure 2.1. HD mutants exhibit loss-of-function phenotype in leaves, seeds, and roots.....	42
Figure 2.2. Nuclear localization of wild-type EYFP-tagged GL2 and mutant gl2 transcription factors.....	44
Figure 2.3. START domain missense mutants exhibit loss-of-function phenotypes despite nuclear localization of the corresponding EYFP-tagged proteins.	46
Figure 2.4. HD is required for DNA binding activity of PDF2 and GL2 while START domain is dispensable for DNA binding.	48
Figure 2.5. START domain is required for homodimerization of GL2.....	51
Figure 2.6. Cycloheximide experiments reveal that the START domain, but not HD, is critical for protein stability of GL2.	52
Figure 2.7. Model of HD and START domain roles in HD-Zip IV transcription factors.	54
Figure 3.1. ZLZ negatively affects the stability of GL2.	101
Figure 3.2. GL2 is nuclear localized despite phenotypic defect in ZLZ deletion mutants.	102
Figure 3.3. Degradation of HD-Zip IV TFs is favored by conditions impairing lipid binding.	104
Figure 3.4. Detection of ubiquitination and SUMOylation in epitope tagged wild-type and mutant gl2 proteins.	106
Figure 3.5. GL2 contains multiple predicted lysine conjugation sites that are conserved in related proteins.....	108
Figure 3.6. Experimental evaluation of potential PTM sites in GL2.....	110
Figure 3.7. . Mutation of the predicted PTM sites in GL2 leads to increased frequency of abaxially curled leaves in <i>Arabidopsis</i>	112
Figure 3.8. K-R and C-A mutants retain nuclear localization of GL2 with altered protein stability.....	113
Figure 3.9. Model showing the effects of mutations in START domain, ZLZ domain and PTM sites of GL2.....	115
Figure 4.1. EYFP fusions to HD-Zip IV TFs enhance their protein stability.....	149
Figure 4.2. Expression of EYFP:GL2 leads to development of abaxially curled leaves in <i>Arabidopsis</i>	151

Figure 4.3. Quantitative characterization of curly leaf plants in comparison to normal/flat leaf plants.	153
Figure 5.1. GL2 contains multiple potential SUMO-interacting motifs (SIMs) that are conserved in HD-Zip IV family members.....	186
Figure 5.2. Mutant analysis of predicted SIMs in the GL2 START domain.....	188
Figure 5.3. Phenotypic characterization of <i>gl2^{SIM353}</i> and <i>gl2^{SIM404}</i> mutants.....	189
Figure 5.4. Expression and subcellular localization of EYFP-tagged GL2 wild-type and <i>gl2^{SIM}</i> mutant proteins.....	191
Figure 5.5. Detection of GL2 protein interactions mediated by its predicted SIMs.	194
Figure A.1. Structural alignments of HD and START domain from HD-Zip IV TFs (related to Figures 2.1A and 2.3A).....	213
Figure A.2. Root hair phenotypes and nuclear localization in various START domain missense mutants (related to Figure 2.3).....	214
Figure A.3. Characterization of START domain missense mutants in the N-terminus of the START domain (related to Figure 2.3).....	216
Figure A.4. START domain is required for in vitro DNA binding as a dimer (related to Figure 2.4C).....	217
Figure A.5. GL2 binding motif and ChIP-seq peak distance from TSS (related to Figure 2.4, D-G).	219
Figure A.6. The START domain is required for GL2 homodimerization (related to Figure 2.5C).	220
Figure A.7. Protein expression of Bait plasmids in Y2H experiments (related to Figure 2.5)..	222
Figure A.8. BiFC negative assays (related to Figure 2.5, D-F).	223
Figure A.9. Cycloheximide chase experiments reveal that the START domain is critical for protein stability of GL2 (related to Figure 2.6).....	224
Figure A.10. Cycloheximide chase experiments with double mutants affecting both HD and START domain of GL2 (related to Figure 2.6; Figure A.9).....	226
Figure B.1. Western blot analysis of leaves from ~one-month-old T1 progeny of <i>FLAG:gl2^{ΔZLZ}</i> , <i>FLAG:gl2^{L480}</i> and <i>FLAG:gl2^{ΔZLZ;L480P}</i> lines confirming the protein expression (Related to Figure 3.2).....	235
Figure B.2. PDF2 is destabilized in absence of its START domain (Related to Figure 3.3). ...	236

Figure B.3. Detection of ubiquitination and SUMOylation in epitope tagged wild-type and mutant gl2 proteins (Related to Figure 3.4).	238
Figure B.4. Alignment of ZLZ from Arabidopsis HD-Zip IV TFs illustrating the conservation of residues in the leucine zippers and the loop (Related to Figure 3.6).	239
Figure B.5. Screening of T1 transformants for rescue of null phenotype, EYFP expression and abaxial leaf curling (Related to Figure 3.7).	240
Figure C.1. Detection of GL2 protein interactions mediated by its predicted SIMs (Related to Figure 5.5).	244
Figure C.2. Western blot showing the expression of AtSUMO1, AtSUMO2, and AtSUMO3 fusion proteins in diploid yeast from Y2H assays using α -SUMO3 antibodies (Related to Figure 5.5).	245
Figure C.3. Bait autoactivation assay corresponding to the Y2H experiments (Related to Figure 5.5).	246

List of Tables

Table A.1. EYFP expression is independent of phenotypic rescue (related to Figures 2.1 and 2.2).	227
Table A.2. Distribution and number of significant peaks from ChIP-seq data (related to Figure 2.4F).	228
Table A.3. Oligonucleotides used in this study.	229
Table B.1. List of oligonucleotides used in this study	241
Table B.2. Strategies used to generate PTM mutants with double/triple/multiple amino acids substitutions	243
Table C.1. List of oligonucleotides used in this study	247
Table C.2. List of constructs generated in this study	248
Table C.3. Accession numbers used for building alignments.....	250
Table C.4. Bacterial stocks ordered from ABRC for SUMO genes	251

List of Abbreviations

Aba	Aureobasidin
ABRC	Arabidopsis Biological Resource Center
AD	Activation domain
ANOVA	Analysis of variance
ATML1	<i>Arabidopsis thaliana</i> MERISTEM LAYER1
BiFC	Bimolecular fluorescence complementation
CaMV	Cauliflower mosaic virus
ChIP	Chromatin immunoprecipitation
CHX	Cycloheximide
Col	Columbia
ConcA	Concanamycin A
DBD	DNA-binding domain
DMSO	Dimethyl sulphoxide
DPSS	Diode-pumped solid state
DTT	Dithiothreitol
DUB	Deubiquitinase
EDTA	Ethylene diamine tetraacetic acid
EGFP	Enhanced green fluorescent protein
EMSA	Electrophoretic Mobility Shift Assay
EYFP	Enhanced yellow fluorescent protein
FP	Fluorescent protein
GAL4	Galactose-responsive transcription factor
GFP	Green fluorescent protein
GL2	GLABRA2
GO	Gene Ontology
GPS-SUMO	Group-based Prediction System-SUMO
HA	Haemagglutinin
HD	Homeodomain
HD-Zip	Homeodomain leucine-zipper
HD-Zip IV	Class IV homeodomain leucine-zipper
HiFi	High fidelity
His	Histidine
HRP	Horseradish peroxidase
HygB	Hygromycin B
IAA	Iodoacetamide
IB	Immunoblot/Immunoblotting
IP	Immunoprecipitate/Immunoprecipitation
I-TASSER	Iterative Threading ASSEMBly Refinement
JASSA	Joined Advanced SUMOylation site and SIM Analyzer
kDa	Kilodaltons
Leu	Leucine
LRI	Leaf Rolling Index
LZ	Leucine-zipper
MES	2-(N-morpholino)ethanesulfonic acid

MS	Murashige and Skoog/ Mass Spectrometry
NEM	N-methylmaleimide
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDF2	PROTODERMAL FACTOR2
PEG	Polyethylene glycol
PI	Propidium iodide
PMSF	Phenylmethylsulfonylfluoride
PONDR	Predictor of Natural Disordered Regions
PTM	Post-translational modification
PVDF	Polyvinylidene fluoride
RING	Really Interesting New Gene
RuBiSCo	Ribulose-1,5-bisphosphate carboxylase-oxygenase
SAD	START-adjacent domain
SBM	SUMO-binding domain
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SIM	SUMO-interacting motif
STAD	START-adjacent domain
StAR	Steroidogenic Acute Regulatory
START	StAR protein-related lipid Transfer
STUbL	SUMO-targeted ubiquitin ligase
SUMO	Small ubiquitin-like modifier
TAIR	The Arabidopsis Information Resource
TBE	Tris-borate-EDTA
TBS	Tris-buffered saline
TF	Transcription factor
Trp	Tryptophan
TSS	Transcription start site
Ub	Ubiquitin
UPS	Ubiquitin proteasome system
WB	Western blot/Western blotting
WT	Wild type
Y2H	Yeast two-hybrid
Zip	Leucine-zipper
ZLZ	Zipper-loop-zipper

Acknowledgements

First of all, I would like to acknowledge my Ph. D. advisor Dr. Kathrin Schrick for continuous mentoring and guidance and for providing necessary resources to run the experiments smoothly. I am also grateful to my Ph. D. supervisory committee members Drs. Thomas Platt, Richard Todd and Ruth Welte who efficiently reviewed my progress reports, provided valuable feedbacks during regular committee meetings and guided me properly for my preliminary as well as final examination for the doctoral degree. Similarly, I thank Dr. Shahideh Nouri for agreeing to serve as the outside chair in my final examination. I thank my former advisor Dr. Govind Vedyappan for training me on molecular cloning techniques that were very useful in my current Ph. D. project. I would also like to remember Drs. Stephanie Shames and Brian Geisbrecht as my former graduate committee members. Next, I want to mention Drs. Timothy Durrett and Lawrence Davis for their valuable inputs during my presentations in combined lab meetings and for their roles as my instructors in Biochemistry I (BIOCH 755) course.

I express my sincere gratitude to Johnson Cancer Research Center (JCRC) at Kansas State University and its director for generously supporting my research by providing me summer stipend for three consecutive years from 2020 to 2022. As an instructor of Cellular and Molecular Immunology (BIOL840) course, Dr. Fleming also motivated me to contribute to a review paper as a co-first author which was my first publication during the graduate career. I would like to thank International Student and Scholar Services and K-State Alumni Center for awarding me with various scholarships. I am also grateful to K-State Graduate School and College of Arts and Sciences for providing me travel awards to attend various conferences and workshops.

I acknowledge the teaching faculty at the Division of Biology (DOB), especially Dr. Martha Smith Caldas, Dr. Christopher Herren, Dr. Eva Horne, Dr. Eve McCulloch and Robert Bear for their effective role as coordinators of the courses that I taught for my GTA employment. Special thanks to Dr. Caldas who also nominated me for the Michael Scotts Watkins Teaching Award for excellence in teaching during graduate program in 2021. I also acknowledge the effort of Graduate Affairs Committee at the DOB in recognizing my research accomplishments and awarding me with John C. Frazier Award for Excellence in Graduate Student Research in Plant Science this year based on my advisor's recommendation. Many thanks to the current and former staffs at the DOB, especially Melissa Bruce, Teresa Wille, Becki Bohnenblust, Sarah Hacker, Matt Spooner, Chad Starr and Bob LeHew for their help in various routine works.

Furthermore, I appreciate the effort of Biology Graduate Student Association (BGSA) and fellow graduate students for organizing social events, providing travel and research grants and for conveying the voice of graduate students to the administration. Also thanks to all of my lab members who were available for socializing during birth day celebrations and ice-cream parties. Last but not least, I thank my better half Shuva Laxmi Pandey for continuous encouragement and support to navigate ups and downs during my graduate career.

Dedication

I dedicate this dissertation to my ever-loving, ever-caring mother Saraswoti Subedi.

Chapter 1 - Introduction

Homeodomain and homeodomain transcription factors

Transcription factors (TFs) are proteins that activate or repress target genes by directly binding to specific promoter elements (Latchman, 1993; Franco-Zorrilla et al., 2014; Weidemüller et al., 2021). TFs are important for regulating the expression of genes in response to endogenous and environmental signals (Weidemüller et al., 2021). Homeodomain (HD) TFs are critical regulators of eukaryotic development (Gehring et al., 1994; Chan et al., 1998). Their name is derived from homeotic transformation that is characterized by defective development or organization of body segments or structures in animals (McGinnis et al., 1984a; McGinnis et al., 1984b; Scott and Weiner, 1984; Capellini et al., 2011). These developmental anomalies occur as a result of overactivation or inactivation of the corresponding genes termed homeotic (Hox) genes which have been extensively studied in *Drosophila* (Garber et al., 1983; McGinnis et al., 1984a; McGinnis et al., 1984b). All homeodomain TFs contain a highly conserved DNA binding region of ~60 amino acids known as the homeodomain (HD) (Carrasco et al., 1984; McGinnis et al., 1984a; Gehring et al., 1990; Bürglin, 1994; Derelle et al., 2007). In plants, HD TFs are involved in various developmental processes such as stem cell maintenance, lateral outgrowth and response to environmental factors (Hay and Tsiantis, 2010; Brandt et al., 2014; Tsuda and Hake, 2015). Plant HD proteins are classified into 14 families including homeodomain leucine-zipper (HD-Zip), Wuschel-related homeobox (WOX) and Knotted-like homeobox (KNOX) (Mukherjee et al., 2009; Romani et al., 2018).

HD-Zip transcription factors

HD-Zip TFs constitute a plant-specific TF family that are involved in various biological processes including development, growth and stress responses as well as light and hormone signaling (Yu et al., 2008; Harris et al., 2011; Shan et al., 2014; Merelo et al., 2017; Roodbarkelari and Groot, 2017; Sessa et al., 2018; Gong et al., 2019; Sharif et al., 2021). This family comprises about half of all the HD TFs in plants. HD-Zip TFs are characterized by the presence of a leucine-zipper (LZ) dimerization domain C-terminal to the HD (Ruberti et al., 1991; Schena and Davis, 1992; Schrick et al., 2004) and are further divided into four subfamilies or classes, HD-Zip I-IV (Schena and Davis, 1994; Di Cristina et al., 1996; Ponting and Aravind, 1999; Riechmann et al., 2000; Schrick et al., 2004).

Roles of HD-Zip IV transcription factors in plants

HD-Zip III and HD-Zip IV subfamily proteins contain a lipid binding START (steroidogenic acute regulatory protein (StAR)-related lipid transfer) domain and a START adjacent domain (STAD or SAD) (Ponting and Aravind, 1999; Schrick et al., 2004; Mukherjee et al., 2022). Previously it was reported that HD-START proteins are expanded in plants and presumably function to sense a variety of hydrophobic ligands that could govern transcription in different cell types (Schrack et al., 2004; Schrick et al., 2014). HD-Zip IV TFs play vital roles in the specification of the epidermis, anthocyanin accumulation, and plant responses to environmental stresses (Kubo et al., 1999; Chen et al., 2017; Zhang et al., 2019; Wang et al., 2020a; Wang et al., 2020b). Three well-studied members of this group include GLABRA2 (GL2) and the functionally redundant members PROTODERMAL FACTOR2 (PDF2) and *Arabidopsis thaliana* MERISTEM LAYER1 (ATML1). GL2 dictates the specification of epidermal cells in

leaves, roots, and seeds (Rerie et al., 1994; Di Cristina et al., 1996; Western et al., 2001). Impaired functions of GL2 lead to visible defects in trichome differentiation and seed coat mucilage production in addition to ectopic root hair formation from non-hair epidermal cells of the root (Rerie et al., 1994; Di Cristina et al., 1996; Masucci et al., 1996; Western et al., 2001; Western et al., 2004). Similarly, PDF2 and ATML1 are involved in maintenance of epidermal (L1) identity of the vegetative, floral and inflorescence meristem (Abe et al., 2003). The double knockout mutants of *PDF2* and *ATML1* result in embryonic lethality (Abe et al., 2003; Ogawa et al., 2015).

Research questions, objectives and hypotheses

Despite the importance of HD-Zip IV proteins as multidomain TFs, the molecular mechanism by which they elegantly control gene expression during cell-type differentiation remains elusive. The START domain was previously reported to be essential for TF activity of HD-Zip IV TFs (Schrack et al., 2014). Specifically, ligand binding to the HD-Zip IV START domain was shown to control transcriptional activity in both yeast and plant cells (Schrack et al., 2014; Wojciechowska et al., 2021). However, the precise mechanism of START-mediated regulation of TF activity is unknown. Similarly, there is a gap in our current understanding regarding the functions of the HD and ZLZ domain in addition to their primary roles in DNA binding and TF dimerization, respectively. The overall goal of this dissertation work is to elucidate the functions of key domains and associated regulatory mechanisms governing the activity of HD-Zip IV TFs using GL2, PDF2 and ATML1 as representative family members in the plant model *Arabidopsis*. Some of the important hypotheses tested are:

1. The homeodomain and START domain play distinct roles in DNA binding, dimerization and protein turnover of HD-Zip IV TFs.

2. The ZLZ dimerization domain is important in the regulation of HD-Zip IV protein levels.
3. The HD, START domain and ZLZ domain are all dispensable for nuclear localization but indispensable for HD-Zip IV TF activity.
4. Protein levels of HD-Zip IV TFs are maintained by post-translational regulatory mechanisms involving ubiquitin, SUMO or related molecules driving optimal development of the plant epidermis.

Outline of dissertation

The first chapter of this dissertation provides general background information on HD-Zip TFs. In addition, it presents an overview of the research questions, objectives and hypotheses of the overall work as well as an outline of the dissertation. Shorter introductions are additionally included in Chapters 2, 3, 4 and 5, and these provide more specific information relevant to each chapter.

The goal of the second chapter is to understand whether the START domain modulates HD-Zip IV TF activity independently or cooperatively with the HD. Mutations were introduced in key residues of the HD in wild-type plants and START domain mutants, and the resulting mutants were characterized for trichome, root hair and seed coat phenotypes as well as protein subcellular localization. The contribution of the START domain in DNA binding by the HD was probed using both in vitro (EMSA) and in vivo (ChIP-Seq) assays. Protein-protein interaction studies in yeast and *Nicotiana benthamiana* were used to discern the possible roles of these two domains in protein dimerization. Finally, a series of cycloheximide chase experiments were conducted to determine the role of the HD versus the START domain in protein turnover, and the possible involvement of the 26S proteasome was tested.

The third chapter aims to identify and characterize post-translational regulatory mechanisms governing HD-Zip IV TFs in Arabidopsis. It presents a series of results from protein stability experiments that were carried out to decipher the regulation of protein levels by the ZLZ domain. Whether the START domain and the ZLZ act independently of one another to affect protein turnover was addressed by examining *gl2^{ΔZLZ}* mutants that carry a mutation in the START domain. Another set of experiments performed with wild-type PDF2 and a mutant lacking the START domain link protein stability with phosphate starvation, ligand binding and involvement of the ubiquitin 26S proteasome system. To detect the PTMs of GL2, pull-down assays were conducted under normal and various stress conditions followed by immunoblotting with α -ubiquitin and α -SUMO antibodies. Additionally, candidate PTM sites in GL2 were identified by using machine learning tools. To evaluate the roles of the predicted sites in vivo, several PTM-related mutants were generated using site-directed mutagenesis, followed by characterization for rescue of the trichome phenotype, leaf curl and nuclear localization of the EYFP-tagged proteins.

The fourth chapter addresses the limitations of using fluorescent protein (EYFP) tags in terms of the effect on protein stability, using the HD-Zip IV TFs GL2 and PDF2 as examples. The data show that while EYFP:GL2 is functional, its overexpression can result in additional phenotypic effects, namely abaxially curled leaves that are associated with enhanced protein stability. A detailed phenotypic characterization of the related Arabidopsis lines was conducted, and the implications of curly leaf phenotypes are discussed.

In the fifth chapter, the possible regulation of GL2 through protein-protein interactions with a small ubiquitin modifier (SUMO), was investigated. Machine learning algorithms predicted two SUMO-interacting motifs (SIMs) that map to the surface of the GL2 START

domain. Mutagenesis of the predicted SIMs along with phenotypic analysis of the resulting mutants are discussed. The subcellular localization of the SIM mutant proteins was addressed by performing confocal microscopy of lines expressing EYFP-tagged proteins. Yeast two-hybrid assays were used to assay possible protein-protein interactions of GL2 with SUMO, as well as the ability of the SIM mutants to form homodimers.

Finally, the sixth chapter provides an overall summary and conclusions for the entire dissertation. The novelty of the findings and their implications are discussed, with relevance to the fields of plant biology and eukaryotic gene regulation. Critical remarks and future directions are highlighted.

References

- 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.
- Brandt, R., Cabedo, M., Xie, Y., and Wenkel, S.** (2014). Homeodomain leucine-zipper proteins and their role in synchronizing growth and development with the environment. *Journal of Integrative Plant Biology* **56**, 518-526.
- Bürglin, T.R.** (1994). A comprehensive classification of homeobox genes. Oxford Univ. Press, Oxford, UK.
- Capellini, T.D., Zappavigna, V., and Selleri, L.** (2011). Pbx homeodomain proteins: TALEnted regulators of limb patterning and outgrowth. *Developmental Dynamics* **240**, 1063-1086.
- 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.
- Chan, R.L., Gago, G.M., Palena, C.M., and Gonzalez, D.H.** (1998). Homeoboxes in plant development. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1442**, 1-19.
- Chen, E., Zhang, X., Yang, Z., Wang, X., Yang, Z., Zhang, C., Wu, Z., Kong, D., Liu, Z., Zhao, G., Butt, H.I., Zhang, X., and Li, F.** (2017). Genome-wide analysis of the HD-ZIP IV transcription factor family in *Gossypium arboreum* and GaHMG11 involved in osmotic tolerance in transgenic Arabidopsis. *Mol Genet Genomics* **292**, 593-609.
- 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. *The Plant J* **10**, 393-402.
- Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R.** (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences* **111**, 2367-2372.
- Garber, R.L., Kuroiwa, A., and Gehring, W.J.** (1983). Genomic and cDNA clones of the homeotic locus Antennapedia in *Drosophila*. *The EMBO J* **2**, 2027-2036.
- Gehring, W.J., Affolter, M., and Bürglin, T.** (1994). HOMEODOMAIN PROTEINS. *Annual Review of Biochemistry* **63**, 487-526.

- 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 in Genetics* **6**, 323-329.
- Gong, S., Ding, Y., Hu, S., Ding, L., Chen, Z., and Zhu, C.** (2019). The role of HD-Zip class I transcription factors in plant response to abiotic stresses. *Physiol Plant* **167**, 516-525.
- Harris, J.C., Hrmova, M., Lopato, S., and Langridge, P.** (2011). Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. *New Phytol* **190**, 823-837.
- Hay, A., and Tsiantis, M.** (2010). KNOX genes: versatile regulators of plant development and diversity. *Development* **137**, 3153-3165.
- Kubo, H., Peeters, A.J., Aarts, M.G., Pereira, A., and Koornneef, M.** (1999). ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis*. *The Plant Cell* **11**, 1217-1226.
- Latchman, D.S.** (1993). Transcription factors: an overview. *Int J Exp Pathol* **74**, 417-422.
- 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.
- McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J.** (1984a). A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* **37**, 403-408.
- McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A., and Gehring, W.J.** (1984b). A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* **308**, 428-433.
- Merelo, P., Paredes, E.B., Heisler, M.G., and Wenkel, S.** (2017). The shady side of leaf development: the role of the REVOLUTA/KANADI1 module in leaf patterning and auxin-mediated growth promotion. *Curr Opin Plant Biol* **35**, 111-116.
- 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.
- Mukherjee, T., Subedi, B., Khosla, A., Begler, E.M., Stephens, P.M., Warner, A.L., Lerma-Reyes, R., Thompson, K.A., Gunewardena, S., and Schrick, K.** (2022). The START domain mediates *Arabidopsis* GLABRA2 dimerization and turnover independently of homeodomain DNA binding. *Plant Physiol* **kiac383**.
- 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.

- Ponting, C.P., and Aravind, L.** (1999). START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends in Biochemical Sciences* **24**, 130-132.
- Rerie, W.G., Feldmann, K.A., and Marks, M.D.** (1994). The GLABRA2 gene encodes a homeodomain 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). Transcription factors: Genome-wide Comparative analysis among eukaryotes. *Science* **290**, 2105.
- Romani, F., Reinheimer, R., Florent, S.N., Bowman, J.L., and Moreno, J.E.** (2018). Evolutionary history of HOMEODOMAIN LEUCINE ZIPPER transcription factors during plant transition to land. *New Phytol* **219**, 408-421.
- Roodbarkelari, F., and Groot, E.P.** (2017). Regulatory function of homeodomain-leucine zipper (HD-ZIP) family proteins during embryogenesis. *New Phytol* **213**, 95-104.
- Ruberti, I., Sessa, G., Lucchetti, S., and Morelli, G.** (1991). A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif. *The EMBO J* **10**, 1787-1791.
- Schena, M., and Davis, R.W.** (1992). HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily. *PNAS USA* **89**, 3894-3898.
- Schena, M., and Davis, R.W.** (1994). Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families. *PNAS USA* **91**, 8393-8397.
- 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. *PNAS USA* **81**, 4115-4119.
- Sessa, G., Carabelli, M., Possenti, M., Morelli, G., and Ruberti, I.** (2018). Multiple Links between HD-Zip Proteins and Hormone Networks. *International journal of molecular sciences* **19**, 4047.
- Shan, C.-M., Shanguan, X.-X., Zhao, B., Zhang, X.-F., Chao, L.-m., Yang, C.-Q., Wang, L.-J., Zhu, H.-Y., Zeng, Y.-D., Guo, W.-Z., Zhou, B.-L., Hu, G.-J., Guan, X.-Y.,**

- Chen, Z.J., Wendel, J.F., Zhang, T.-Z., and Chen, X.-Y.** (2014). Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. *Nat Comm* **5**, 5519.
- Sharif, R., Raza, A., Chen, P., Li, Y., El-Ballat, E.M., Rauf, A., Hano, C., and El-Esawi, M.A.** (2021). HD-ZIP gene family: Potential roles in improving plant growth and regulating stress-responsive mechanisms in plants. *Genes (Basel)* **12**.
- Tsuda, K., and Hake, S.** (2015). Diverse functions of KNOX transcription factors in the diploid body plan of plants. *Current Opinion in Plant Biology* **27**, 91-96.
- Wang, X., Bi, S., Wang, L., Li, H., Gao, B.A., Huang, S., Qu, X., Cheng, J., Wang, S., Liu, C., Jiang, Y., Zhang, B., Liu, X., Zhang, S., Fu, Y., Zhang, Z., and Wang, C.** (2020a). GLABRA2 regulates actin bundling protein VILLIN1 in root hair growth in response to osmotic stress. *Plant Physiol* **184**, 176-193.
- Wang, Z., Wang, S., Xiao, Y., Li, Z., Wu, M., Xie, X., Li, H., Mu, W., Li, F., Liu, P., Wang, R., and Yang, J.** (2020b). Functional characterization of a HD-ZIP IV transcription factor NtHDG2 in regulating flavonols biosynthesis in *Nicotiana tabacum*. *Plant Physiol Biochem* **146**, 259-268.
- Weidemüller, P., Kholmatov, M., Petsalaki, E., and Zaugg, J.B.** (2021). Transcription factors: Bridge between cell signaling and gene regulation. *PROTEOMICS* **21**, 2000034.
- 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., Hinch, D.K., Skirycz, A., and Schrick, K.** (2021). PROTODERMAL FACTOR2 binds lysophosphatidylcholines and transcriptionally regulates phospholipid metabolism. *bioRxiv*, 2021.2010.2020.465175.
- Yu, H., Chen, X., Hong, Y.-Y., Wang, Y., Xu, P., Ke, S.-D., Liu, H.-Y., Zhu, J.-K., Oliver, D.J., and Xiang, C.-B.** (2008). Activated expression of an Arabidopsis HD-START protein confers drought tolerance with improved root system and reduced stomatal density. *The Plant Cell* **20**, 1134-1151.
- Zhang, H., Ma, X., Li, W., Niu, D., Wang, Z., Yan, X., Yang, X., Yang, Y., and Cui, H.** (2019). Genome-wide characterization of NtHD-ZIP IV: different roles in abiotic stress response and glandular trichome induction. *BMC Plant Biol* **19**, 444.

Chapter 2 - START domain mediates Arabidopsis GLABRA2 dimerization and turnover independently of homeodomain DNA binding

This chapter has been published as my co-first author research paper in *Plant Physiology*. For this project, I performed a part of cycloheximide chase experiments (Figures 2.6C and 2.6D) and made a major contribution in the preparation of manuscript by writing abstract, introduction, results and discussion sections which were revised by Dr. Schrick.

“Thiya Mukherjee, Bibek Subedi, Aashima Khosla, Erika M Begler, Preston M Stephens, Adara L Warner, Ruben Lerma-Reyes, Kyle A Thompson, Sumedha Gunewardena, Kathrin Schrick, The START domain mediates Arabidopsis GLABRA2 dimerization and turnover independently of homeodomain DNA binding, *Plant Physiology*, 2022;, kiac383.”

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 a lipid-binding domain termed START (Steroidogenic Acute Regulatory (StAR)-related lipid Transfer). Previous work established that the START domain of GLABRA2 (GL2), a HD-Zip IV member from Arabidopsis (*Arabidopsis thaliana*), is required for transcription factor activity. Here, we addressed 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) resulted in phenotypic defects in leaf trichomes, root hairs and seed mucilage, similar to those observed for START domain mutants, despite nuclear localization of the respective proteins. *In vitro* and *in vivo* experiments demonstrated that while HD mutations impair binding to target DNA, the START domain is dispensable for DNA binding. Vice versa, protein interaction assays revealed impaired GL2 dimerization for multiple alleles of START mutants, but not HD mutants. Using *in vivo* cycloheximide chase experiments, we provided evidence for the role of START, but not HD, in maintaining protein stability. This work advances our mechanistic 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., 2006a). In *Arabidopsis* (*Arabidopsis thaliana*), 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 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 composed 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 in the eukaryotes (Carrasco et al., 1984; McGinnis et al., 1984; Gehring et al., 1990; Bürglin, 1994; Derelle et al., 2007). It was 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 earliest reported homeobox gene from plants was maize (*Zea mays*) *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).

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; L  tourneau 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 (Schr  ck 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 (Schr  ck et al., 2014). However, lack of detailed information on how START modulates transcription factor activity poses 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 imparted defective phenotypes in trichomes,

roots and seeds, similar to those of START mutants. While HD mutations abolished DNA binding, multiple mutations affecting the START domain retained DNA binding affinity *in vitro* and *in vivo*. Gel shift experiments led to the surprising discovery that HD-Zip IV proteins lacking the START domain can bind DNA as monomers. Using interaction studies in yeast and *Nicotiana benthamiana* we confirmed dimerization defects from mutations affecting START but not HD. 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.

Materials and methods

Plant material, growth conditions, and *Arabidopsis* transformation

Wild-type *Arabidopsis* (*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 wild-type *proGL2:EYFP:GL2* and mutant *proGL2: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 (21-55) 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 HD from GL2 was found to exhibit structural similarity with the solution structure of *Gallus gallus* HOMEBOX 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., 2006a) 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 $gl2^{\Delta ST}$, $gl2^{\Delta HD46}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{K146T}$, $gl2^{K146E}$, $gl2^{A258P}$, $gl2^{M271I}$, $gl2^{W279R}$, $gl2^{I289N}$, and $gl2^{E294Q}$ were generated using the Q5 Site-Directed Mutagenesis Kit (E0554S, New England Biolabs) following the manufacturer's protocol with primers listed in **Table A.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 (Schrack 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. The vector and insert DNA fragments were assembled in a ratio of 1:5 using NEBuilder HiFi DNA Assembly (E5520S, New England Biolabs). The $gl2^{L480P}$, $gl2^{E375G}$, $gl2^{R392M}$, $gl2^{E375G;R392M}$, and $gl2^{Amo}$ mutants were generated by one-step PCR-based site-directed mutagenesis (Scott et

al., 2002) using PfuUltra II Fusion HS DNA polymerase (Agilent Technologies). The Phusion[®] Site-Directed Mutagenesis Kit (ThermoFisher Scientific) was used to generate the *gl2^{AHD}* and *gl2^{AHD6;AZLZ}* deletion mutants. Oligonucleotides for generation of mutants are listed in **Table A.3**. After verification by sequencing, the GL2 deletion constructs were restriction digested with *KpnI* and *SalI* and ligated into the binary vector SR54 (*proGL2:EYFP:GL2*) at the corresponding sites.

Phenotypic assays and microscopy of *Arabidopsis*

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% w/v 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 2-5 days after germination. For seed mucilage analysis, seeds were stained with ruthenium red (R2751, Sigma-Aldrich) following an established protocol (McFarlane et al., 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% w/v 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 (10% Intensity), and a 519-607 nm

band pass filter captured emission (Gain = 700). To image EYFP-tagged proteins in primary roots, 4-d-old seedlings were stained with propidium iodide (10 µg/ml) to visualize cell boundaries. Seedlings were rinsed with distilled water and mounted in water under a coverslip. Images were collected using Zeiss LSM-5 Pascal confocal microscope. The EYFP excitation was performed using an 488 nm Argon laser (30% Intensity), collecting emission with a 510-540 nm band pass filter (Gain = 800). Propidium iodide fluorescence was detected using a 543 nm laser (10% Intensity) for excitation and a 587-625 nm band pass filter (Gain = 800) for emission. Images were processed using Adobe Photoshop C4.

***In vitro* transcription and translation**

The full-length *PDF2* and *pdf2^{ΔST}* 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 **Table A.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 **Table A.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% w/v agarose gel (1X TBE, pH 8.3) at 150 V for 1 h, or a 7.5% w/v 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*^{ΔST}, *gl2*^{ΔHD6;ΔZLZ}, *gl2*^{Amo} and *gl2*^{R392M;E375G} and *gl2-5* lines were surface sterilized and grown on 0.8% w/v agar containing 1X MS (Murashige and Skoog, 1962) and 1% w/v 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% v/v 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 **ChIP-seq 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)

For **Figure 2.5A**, Bait and Prey constructs were prepared by restriction enzyme cloning with *EcoRI* and *BamHI* inserts into the pGBKT7 and pGADT7 plasmids (Clontech). Bait constructs were transformed into the haploid yeast strain Y2HGold (*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 prey constructs were transformed into Y187 (*MATa ura3-52 hi3-200 ade2-101 trp1-901 leu2-3,112 gal4Δ gal80Δ met-URA3::GAL1_{UAS}-Gal1_{TATA}-LacZ MEL1*), using the LiAc method (Gietz and Woods, 2002), followed by selection on -Trp and -Leu, respectively. As a negative control, empty vectors were co-transformed with the desired constructs. Mating experiments were performed following Matchmaker Gold Yeast Two-Hybrid System (Clontech). Diploids were replica plated to media that selects for reporter gene activity of *HI3*, *AUR1-C*, and *MEL1*. To test for autoactivation of bait plasmids Y2HGold/pDEST32 were plated in -Trp media and replica plated to media that selects for activity of the reporter genes *HI3* and *AUR1-C* respectively. All pair-wise combinations were investigated for protein-protein interaction by growing mated cultures to a density of $OD_{600}=1.0$. Four serial dilutions at 4x each were prepared from the normalized cultures and spotted onto permissive and selective media for detection of interaction using a 48-pin multiplex plating tool ('Frogger', Dankar, Inc.). Plates were incubated at 30°C for 3-5 days, followed by imaging. Bait protein expression was confirmed by immunoblotting with Hela-c-Myc (9E10) polyclonal antibody (Gift from Stella Lee, 1:10 dilution).

For **Figure 2.5B-C** and **Figures A.6C-E**, full-length *GL2*, *PDF2*, *ATML1* cDNAs and the various deletion and missense mutations 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). For **Figure 2.5B**, both bait and prey constructs were transformed into Y2HGold, and interaction was assayed in this haploid strain. For **Figure 2.5B-C** and **Figures A.6C-E**, Y2HGold and the opposite mating type, Y187, were transformed with pDEST32 and pDEST22 constructs, using LiAc/PEG transformation, and plated on –Leu and -Trp media to select for the bait and prey plasmids, respectively. Mating experiments and selection for diploids were performed as described above. To test for autoactivation of bait plasmids Y2HGold/pDEST32 were plated in -Leu media and replica plated to media that selects for activity of the reporter genes *HIS3* and *AUR1-C* respectively. Y2H assays for imaging were performed as described above. For analyzing bait protein expression, single colonies expressing pDEST32 constructs were grown at 30°C 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).

Y2H quantitative β -galactosidase assay

For **Figure A.6A-B**, Gateway cloning was used to transfer genes for HD-Zip III and IV TFs from pENTR223.1 clones to pDEST22 and pDEST32 destination vectors (ProQuest Two-Hybrid System, Invitrogen). The bait and prey plasmids were transformed into yeast strain MaV203 (Invitrogen): *MAT α leu2-3,112 trp1-901 his3 Δ 200; ade2-101 cyh2R can1R gla4 Δ gal80 Δ GAL1::lacZ HIS3_{UASGAL1}::HIS3LEU2 SPAL10_{UASGAL1}::URA3* following the Clontech

Yeast Protocol Handbook. Four independent colonies from each transformation were selected to perform quantitative β -galactosidase liquid assay as previously described (Schrack et al., 2014).

Bimolecular fluorescence complementation (BiFC) in *Nicotiana benthamiana*

To generate constructs for split GFP, Gateway compatible BiFC vectors 125-NXGW and 127-CXGW for N-terminal fusions of split GFP (a gift from Adrienne Roeder) were used for LR recombination reactions of the corresponding wild-type and mutant genes in pENTR/D-TOPO plasmids. Agroinoculation of *N. benthamiana* was performed as described in (Sparkes et al., 2006). Seeds of *N. benthamiana* were germinated on wet filter paper for 7d, and then the seedlings were planted on above-described soil mix. For transient expression of fluorescent protein, leaves of 3–4-week-old 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 MgCl₂, 10 mM 2-(N-morpholino)ethanesulfonic acid pH 5.6 and 150 μ M acetosyringone to an optical density of 0.2, and incubated for 2–4 h at 28°C in the dark. BiFC constructs were co-expressed with the p19 suppressor of gene silencing, derived from tomato bushy stunt virus (TBSV) (Voinnet et al., 2003). Leaves were infiltrated with a needleless syringe into the abaxial side of leaves of 3–4-week-old *N. benthamiana* plants and examined after 2–3 d by confocal laser scanning microscopy using a Zeiss LSM-5 Pascal microscope with an 488 nm Argon laser for excitation (30% Intensity), collecting emission with a 510-540 nm band pass filter (Gain = 800). Mean pixel intensities of three different circular regions of interest (ROIs) were quantified from 10 images for each construct combination using Image J software.

***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 μ M 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 μ M 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 $\pm$ 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 $\pm$ SD. Ordinary one-way ANOVA using Tukey's multiple comparisons test was applied to establish significant differences ($p < 0.05$).

BiFC assays: For each combination of paired constructs, ten images were analyzed. Mean pixel intensity of three different circular regions of interest (ROIs) was quantified by the Image J software. An unpaired t -test was used to indicate statistically significant differences between wild-type and mutant using data from two independent experiments in two trials ($p < 0.0001$).

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 (DMSO controls) 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 $\pm$ 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)

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 HOMEODOM PROTEIN ENGRAILED-2 (Gg-En-2) from chicken exhibits predicted structural similarity with the HD from HD-Zip IV members. Aligning HDs from all 16 HD-Zip IV TFs (GL2, PDF2, ANL2, ATML1, HDG1-12) with reference to Gg-En-2 revealed highly conserved α -helices as well as amino acid residues (**Figure A.1A**). The lysine at position 45 of the HD displays 100% conservation in HD-Zip IV members (**Figure 2.1A; Figure A.1A**). 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 transcription factors PITX2 and PITX3 (Semina et al., 1996; Gage and Camper, 1997; Semina et al., 1997; Smidt et al., 1997).

The conserved features of the START domain among HD-Zip IV members were studied using a structure-based sequence alignment 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 α helices and β strands, as well as conserved amino acids implicated in ligand binding (**Figure A.1B**).

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 (Δ ST)

backgrounds (**Figure 2.1A**). The $gl2^{\Delta ST}$ mutant lacks 235 residues encoding the START domain (Schrack et al., 2014). We constructed HD deletion mutants that remove the consensus α -helices in GL2 ($\Delta HD11$, $\Delta HD20$, $\Delta HD46$) and missense mutants (K-E, K-T) targeting a highly conserved lysine (K146 in GL2; K107 in PDF2) (**Figure 2.1A**). This positively charged lysine was replaced with a negatively-charged glutamic acid or with a neutral polar residue, threonine. 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 the enhanced yellow fluorescent protein (EYFP) translationally fused at the amino-terminus. For each genotype, we screened >20 independent T1 transformants to monitor nuclear expression of EYFP and phenotypic complementation of $gl2-5$ (**Table A.1**), and chose representative homozygous T3 lines for further characterization.

Mutations affecting the HD of GL2 exhibit defects in leaf, root and seed epidermis

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 (Schrack et al., 2014). We examined the phenotypic effects of mutations in the HD in the presence or absence of the START domain deletion (**Figure 2.1**). 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 ($gl2^{K146E}$) displayed trichome defects and were indistinguishable from ΔST or $gl2-5$ null mutants (**Figure 2.1B and 2.1B'**). Another missense mutation in the same lysine (K146T) resulted in partial rescue (**Figure 2.1B'**). 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 ΔST mutants (**Figure 2.2A'**). Quantification of trichomes on

the first leaves revealed a reduction in total number as well as the abnormal presence of unbranched trichome in these mutants (**Figure 2.1D**).

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 2.1C, 2.1C' and 2.1E**). To assess the effect of mutations on differentiation of the mucilage secretory cells, imbibed seeds were stained with ruthenium red, a dye that binds to pectinaceous polysaccharides. Reduced seed coat mucilage was observed in HD mutants *gl2*^{ΔHD11}, *gl2*^{ΔHD20}, *gl2*^{ΔHD46}, and *gl2*^{K146E}, while the K146T mutation resulted in a partial mucilage defect (**Figures 2.1C and 2.1C'**) consistent with partial phenotypes observed in trichomes and roots (**Figures 2.1D and 2.1E**). HD and START domain double mutants (*gl2*^{ΔHD20;ΔST}, *gl2*^{K146E;ΔST}, *gl2*^{K146T;ΔST}) were also defective in seed mucilage similar to ΔST mutants (**Figures 2.1C and 2.1C'**). These results demonstrate that both the HD and START domain are critical for epidermal differentiation in multiple cell types, and that mutation in one domain is not able to suppress the loss-of-function phenotype in the other domain.

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), non-hair cell files in the root epidermis (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 (Schrack 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 2.2A, 2.2A', 2.2B and 2.2B'**). 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 2.2A' and 2.2B'**). Confocal laser scanning microscopy of developing seeds indicated nuclear localization of HD and START mutant proteins in the seed coat cells (**Figures 2.2C, 2.2C', 2.2D and 2.2D'**). The results demonstrate that both HD and START domain are dispensable for translocation of GL2 to the nucleus.

START domain missense mutants exhibit loss-of-function despite nuclear localization of GL2

We next tested whether missense mutations in the START domain result in phenotypic defects similar to those seen in the START deletion mutant, $gl2^{\Delta ST}$ (**Figure 2.3A**). Transgenic lines carrying *EYFP:gl2* constructs were established according to the strategy described above for the HD mutants and $gl2^{\Delta ST}$. Phenotypic analysis showed that three START missense mutants ($gl2^{E375G;R392M}$, $gl2^{Amo}$, $gl2^{L480P}$) and a mutant deleted for the six terminal amino acids of the HD and ZLZ domain ($gl2^{\Delta HD6;\Delta ZLZ}$), display leaf trichome defects and seed mucilage defects (**Figure 2.3B, 2.3C and 2.3C'**), as well as excess root hair formation (**Figure A.2A**) similar to the $gl2$ -5 null mutant. The mutations in $gl2^{E375G;R392M}$ disrupt a salt bridge that is thought to be critical for ligand binding (Roostae 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 2.3A and Figure A.1B**). In $gl2^{L480P}$, the mutation affects a lipid contact site in the C-terminal α -helix of the START domain (Roostae et al., 2009; Alpy and Tomasetto, 2014; Wojciechowska, 2021).

We identified another missense allele, W279R, that maps to the N-terminus of the START domain, which also results in loss-of-function (**Figure A.3**). In *gl2*^{W279R}, a highly conserved tryptophan is replaced with a charged residue within a predicted hydrophobic region of the binding pocket (**Figure A.1B and Figure A.3A**). In contrast, missense mutations in the predicted N-terminal α -helix of START (*gl2*^{A258P}, *gl2*^{M271I}, *gl2*^{I289N}, *gl2*^{E294Q}) do not significantly impair function (**Figures A.3A-E**)

Despite loss-of-function phenotypes, the four missense START alleles (*gl2*^{E375G;R392M}, *gl2*^{Amo}, *gl2*^{L480P}, *gl2*^{W279R}) and the *gl2* ^{Δ HD6; Δ ZLZ} mutant retained nuclear localization of the corresponding EYFP:tagged proteins (**Figure 2.3, Figure A.2B and Figure A.3F**). These results provide additional evidence that the HD and START domain are not required for nuclear location of GL2.

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 respective promoter regions (Nakamura et al., 2006b; Khosla et al., 2014). We performed electrophoretic mobility shift assays (EMSA) to examine whether DNA binding is regulated by the HD alone or by HD in combination with the START domain. For the DNA probe, we used the L1-Box element (TAAATGTA) implicated in GL2-DNA binding that is found upstream of several GL2 targets (Ohashi et al., 2003; Tominaga-Wada et al., 2009; Khosla et al., 2014). We produced recombinant Halo-tagged proteins using wheat germ *in vitro* transcription-translation. Despite repeated attempts, we were unable to obtain full-length recombinant GL2 protein that was functional in EMSA.

Given the high degree of sequence similarity between HD-Zip IV proteins (**Figure A.1**), we selected PDF2 and mutant versions analogous to our *gl2* HD and START

mutants (pdf2^{K107E}, pdf2^{K107T}, pdf2^{ΔST}, pdf2^{K107E;ΔST}, pdf2^{K107T;ΔST}) 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 a Cy3-labeled probe containing three tandem repeats of TAAATGTA. Both wild-type PDF2 and pdf2^{ΔST} caused a shift in probe mobility as an indication of DNA binding (**Figure 2.4A**). In contrast, the two missense mutations K107E and K107T resulted in defective mobility shift independently of the START domain. K107E completely abolished binding to the target sequence as indicated by a lack of mobility shift. While K107T resulted in partial DNA binding for both wild-type PDF2 (pdf2^{K107T}) and pdf2^{ΔST} (pdf2^{K107T;ΔST}) (**Figure 2.4A**). Western blotting confirmed uniform expression of the recombinant proteins used in our EMSA experiments (**Figure 2.4B**).

Combining EMSA with native polyacrylamide gel electrophoresis, we observed that the wild-type PDF2 protein binds the DNA as a dimer. In contrast, the pdf2^{ΔST} mutant binds the DNA target sequence as a monomer (**Figure 2.4C; Figure A.4**). The abnormal monomeric form was also seen in comparing the DNA-binding behavior of the mutant atml1^{ΔST} protein to wild-type ATML1 dimer (**Figure A.4**).

START domain is dispensable for DNA binding *in vivo*

We performed ChIP-seq assays with seedlings expressing EYFP-tagged wild-type GL2 and mutant proteins to test whether the START domain is dispensable for DNA binding *in vivo*. The experiments included three START domain mutants (gl2^{ΔST}, gl2^{E375G;R392M} gl2^{Amo}) as well as the HD-ZLZ mutant (gl2^{ΔHD6;ΔZLZ}), all of which display loss-of-function phenotypes similar to the *gl2-5* null mutant, but whose proteins are nuclear-localized (**Figures 2.1-2.3; Figure A.2; ChIP-seq Data Sets 1 and 2**).

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 2.4D; Table A.2**). The consensus DNA binding motif TAAATGT(A/T) was found in >85% of ChIP-seq peaks for wild-type GL2 and the START mutants (**Figure 2.4D and 2.4E; Figure A.5A**). 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 2.4F; Table A.2**). The relative positions of binding sites to transcription start sites (TSS) were similar for the wild-type GL2 and START domain mutants (**Figure 2.4G; Figure A.5B**).

Gene ontology (GO) classification revealed overlapping functional classes of target genes bound by wild-type GL2 and the START domain mutants (**Figure 2.4H**). 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 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 2.4H**). A list of target genes with ChIP-seq peaks in their respective promoter regions is presented in **ChIP-seq Data Set 2**. The *MYB23/At5g40330* gene, which was previously identified as a direct target of GL2 (Khosla et al. 2014) was found to contain ChIP-seq peaks in wild-type GL2 as well as $gl2^{\Delta ST}$, $gl2^{E375G;R392M}$ and $gl2^{Amo}$ samples (**ChIP-seq Data Set 2**).

Overall, the data indicate that START domain mutants retain DNA-binding activity in a manner that is comparable to wild type.

START domain is required for GL2 homodimerization, while the HD is dispensable for this activity

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 and heterodimerization of PDF2 and ATML1 when bound to target DNA (**Figure 2.4C; Figure A.4**). We tested GL2 in yeast two-hybrid (Y2H) assays to further address the role of START in homodimerization of HD-Zip IV TFs. Unlike several other HD-Zip family members, GL2 does not show bait autoactivation in the Y2H assay (**Figures A.6A and A.6C**). Y2H experiments confirmed that GL2 undergoes homodimerization but we did not detect heterodimerization of GL2 with other HD-Zip IV members nor was heterodimerization detected with two HD-Zip III members that were tested (**Figure A.6B**).

Next, we assayed self-interaction between different segments of GL2. Only the full-length GL2 protein and the segment encoding START+STAD displayed homodimerization. The START domain alone did not exhibit homodimerization, while the STAD domain alone could not be tested because of its bait autoactivation (**Figure 2.5A**). Consistent with the importance of the START domain for self-interaction, deletion of START ($gl2^{\Delta ST}$) and two loss-of-function missense mutants ($gl2^{R392M}$ and $gl2^{L480P}$) abolished homodimerization. The $gl2^{E375G}$ mutant, whose phenotype is similar to wild-type (**Figure 2.2**) supported dimerization (**Figure 2.5B**).

We examined whether the HD plays a role in homodimerization. All of the HD mutants tested ($gl2^{\Delta HD}$, $gl2^{\Delta HD20}$, $gl2^{\Delta HD11}$, $gl2^{\Delta HD46}$, $gl2^{K146T}$, $gl2^{K146E}$) exhibited homodimerization but not bait autoactivation similar to wild-type *GL2* (**Figure 2.5C; Figures A.6C-E**). In contrast, double mutants with deletion of START ($gl2^{K146T;\Delta ST}$, $gl2^{\Delta HD20;\Delta ST}$, $gl2^{K146E;\Delta ST}$) abolished homodimerization (**Figure 2.5C**). These results show that *GL2* dimerization is unaffected by HD mutations but is impaired by deletion of the START domain. Western blotting confirmed the expression of the various wild-type and mutant fusion proteins in yeast (**Figure A.7**).

Bimolecular fluorescence complementation (BiFC) assays in *N. benthamiana* (**Figure 2.5D-F**) verified our *GL2* homodimerization results in yeast. Controls expressing either the N- or C-terminal half of GFP showed little or no fluorescence (**Figure A.8**). In contrast, strong fluorescence signals were observed in nuclei for samples expressing both halves of GFP fused to wild-type *GL2*, indicating homodimerization. Consistent with the reported role of the ZLZ domain in dimerization, the fluorescence intensity was significantly reduced when wild-type *GL2* was tested for interaction with $gl2^{\Delta HD6;\Delta ZLZ}$ (**Figure 2.5E**). Reduced fluorescence intensity was similarly seen when wild-type *GL2* was tested for interaction with four START domain mutant proteins ($gl2^{\Delta ST}$, $gl2^{Amo}$, $gl2^{E375G;R392M}$, $gl2^{L480P}$), as well as for self-dimerization for both of the START mutants we tested ($gl2^{\Delta ST}$, $gl2^{Amo}$) (**Figure 2.5F**). Taken together, the interaction assays in yeast and *N. benthamiana* demonstrate the critical importance of START for homodimerization of *GL2*, whereas the HD domain is not required for this activity.

START domain, but not HD, regulates *GL2* protein stability

Little is known about how the level of HD-Zip TFs is regulated post-translationally. It was reported that mutation of a putative lipid contact site in the START domain of ATML1 completely disrupts expression of the fluorescently tagged protein in lateral roots (Nagata et al.,

2021). In our study, we observed that some EYFP-tagged gl2 proteins harboring mutations in either the HD or START, particularly gl2^{W279R}, appeared weakly expressed in comparison to the wild-type protein (**Figure 2.3; Figure A.3F**).

To determine whether the HD and/or START domain play a role in regulating GL2 levels, we monitored protein half-life by treating *Arabidopsis* seedlings with cycloheximide, an inhibitor of 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^{ΔHD20} and gl2^{ΔST} along with three of the START missense mutants: gl2^{W279R}, gl2^{L480P}, and gl2^{Amo}. As revealed by Western blot and protein quantification, wild-type GL2 and gl2^{ΔHD20} proteins remained stable over 24 h, whereas protein levels of various START domain mutants (gl2^{ΔST}, gl2^{L480P}, gl2^{W279R}, gl2^{Amo}) dropped rapidly (**Figures 2.6A and 2.6B; Figure A.9**). The gl2^{ΔST} and gl2^{W279R} mutant proteins were particularly unstable, having barely detectable protein levels at the 2-h time point (**Figure A.9**). In case of gl2^{ΔST}, a degradation product of ~65 kDa also exhibited a diminished protein half-life (**Figure A.9**). 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 (**Figure A.10**). 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 the 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 2.6C and 2.6D**). The data suggest the possible involvement of

the Ubiquitin/26S proteasome system (UPS) in the turnover of GL2 when the START domain is nonfunctional.

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 (Schrack 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 the null mutant *gl2-5*, Δ ST, and multiple START missense mutants (**Figures 2.1 and 2.3; Figure A.3**). Both HD and the START domain are critical for epidermal differentiation, but the molecular mechanisms underlying their loss-of-function phenotypes are different. 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 mutation in the HD, changing a conserved lysine to a glutamic acid (K107E), completely disrupted DNA binding activity of PDF2 (**Figure 2.4A**). 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 2.4A**). The effect of either of these two missense mutations on DNA binding was unaltered by the deletion of the START domain, consistent with our model that the HD functions in DNA binding independently of the START domain.

Our results indicate that START domain mutants are proficient in DNA binding under both *in vitro* and *in vivo* conditions (**Figure 2.4**). Several studies report that HD-Zip IV TFs bind DNA elements as dimers (Sessa et al., 1993; Palena et al., 1999; Tron et al., 2001). Surprisingly, our EMSA data suggested that START-deleted proteins bind DNA as monomers (**Figure 2.4C**, **Figure A.4**). While the START domain is not required for DNA binding, we confirmed the importance of the START domain for dimerization using self-interaction/homodimerization assays in yeast (**Figure 2.5B-C**; **Figure A.6**), and BiFC experiments *in planta* (**Figure 2.5D-F**; **Figure A.8**). 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 2.6A and 2.6B**; **Figure A.9**). It is noteworthy that gl2^{L480P}, gl2^{W279R} and gl2^{Amo} harbor missense mutations in highly conserved amino acid residues predicted to be important for ligand binding (**Figure**

A.1B). Thus, our findings support the hypothesis that ligand binding through the START domain stabilizes GL2 protein levels. Stabilization of protein levels might aid in activation of the protein, promotion of protein–protein interactions, or correct protein folding. It is noteworthy that deletion of the START domain results in very unstable protein (**Figure A.9**), but the *gl2*^{AST} protein binds to the promoters of target genes *in planta* (**Figure 2.4D-H; Figure A.5**), arguing against a general defect in 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 MG132 point to possible degradation of GL2 via the 26S proteasome (**Figures 2.6C and 2.6D**). 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 turnover 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 2.7**). In this way, the multidomain transcription factor 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 nonfunctional due to mutation, the protein fails to dimerize and becomes unstable. Since START 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 START. 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 monomeric form of the transcription factor binds to its DNA target site via its HD (**Figure 2.4C; Figure A.4**). In our model, unliganded monomers lacking the START domain are unstable, resulting in premature degradation.

Other possible domain interactions in HD-Zip IV TFs

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. It is intriguing that while the START domain is important for homodimerization of GL2, it was unable to form dimers on its own and appears to require cooperative action with STAD (**Figure 2.5A**), a C-terminal domain of unknown function. Another candidate is the ZLZ domain, a plant-specific leucine zipper that contains ~10-20 amino acid loop with two highly conserved cysteines (Schrack et al., 2004). One study suggested that these cysteines are involved in redox control 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 (**Figures 2.2 and 2.3**). 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.

Figures

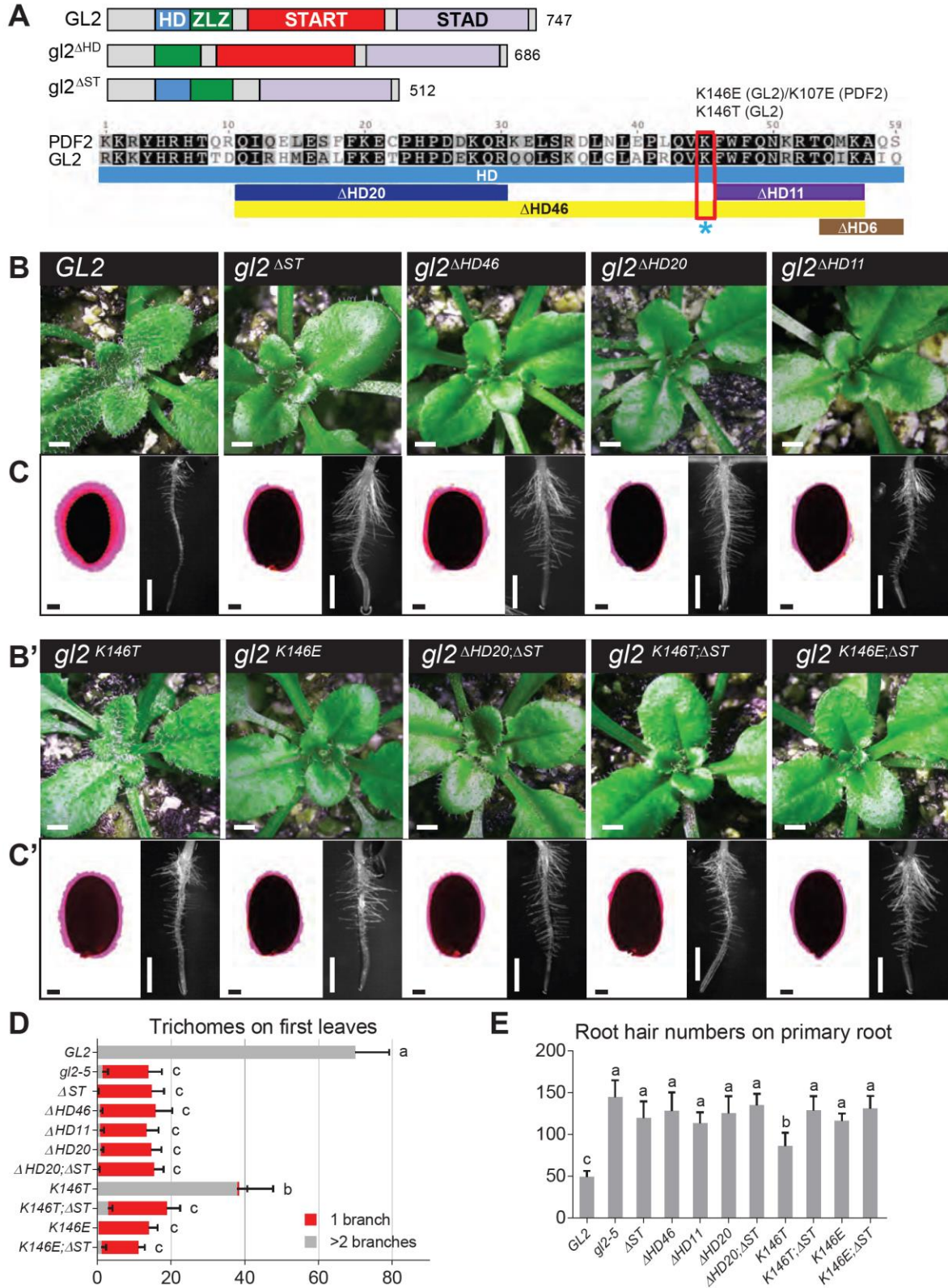


Figure 2.1. HD mutants exhibit loss-of-function phenotype in leaves, seeds, and roots.

(A) Schematic of wild-type GL2 and mutant $gl2^{\Delta HD}$ and $gl2^{\Delta START}$ proteins containing HD, ZLZ, START and STAD. Amino acid lengths of proteins are indicated (right). 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 gray shading, respectively. See **Figure A.1A**. 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 and blue asterisk. **(B)** and **(B')** Wild-type *GL2* rosettes 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. **(C)** and **(C')** Left: Ruthenium red staining of seeds to assay mucilage production. Wild-type *GL2* seeds display a normal mucilage layer, while the various *gl2* mutants exhibit defective mucilage production. $gl2^{K146T}$ displays a partial phenotype. Bar = 100 μ m. Right: 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 exhibit excessive root hair formation. Bar = 1 mm. **(D)** Total number and branching pattern of trichomes on the first leaves. Gray bars indicate trichomes with 2-5 branches in plants expressing wild-type EYFP-tagged GL2. Mutant plants of the genotypes *gl2-5*, and the other *gl2* mutants exhibit unbranched trichomes, indicated by red bars. $gl2^{K146T}$ plants display partial trichome differentiation defects. Values are mean $\pm$ 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. The $gl2^{K146T}$ seedlings exhibit an intermediate phenotype. Values are mean $\pm$ SD for n = 6-10 seedlings. Letters denote significant differences from ordinary one-way ANOVA ($p < 0.05$) using Tukey's multiple comparisons test.

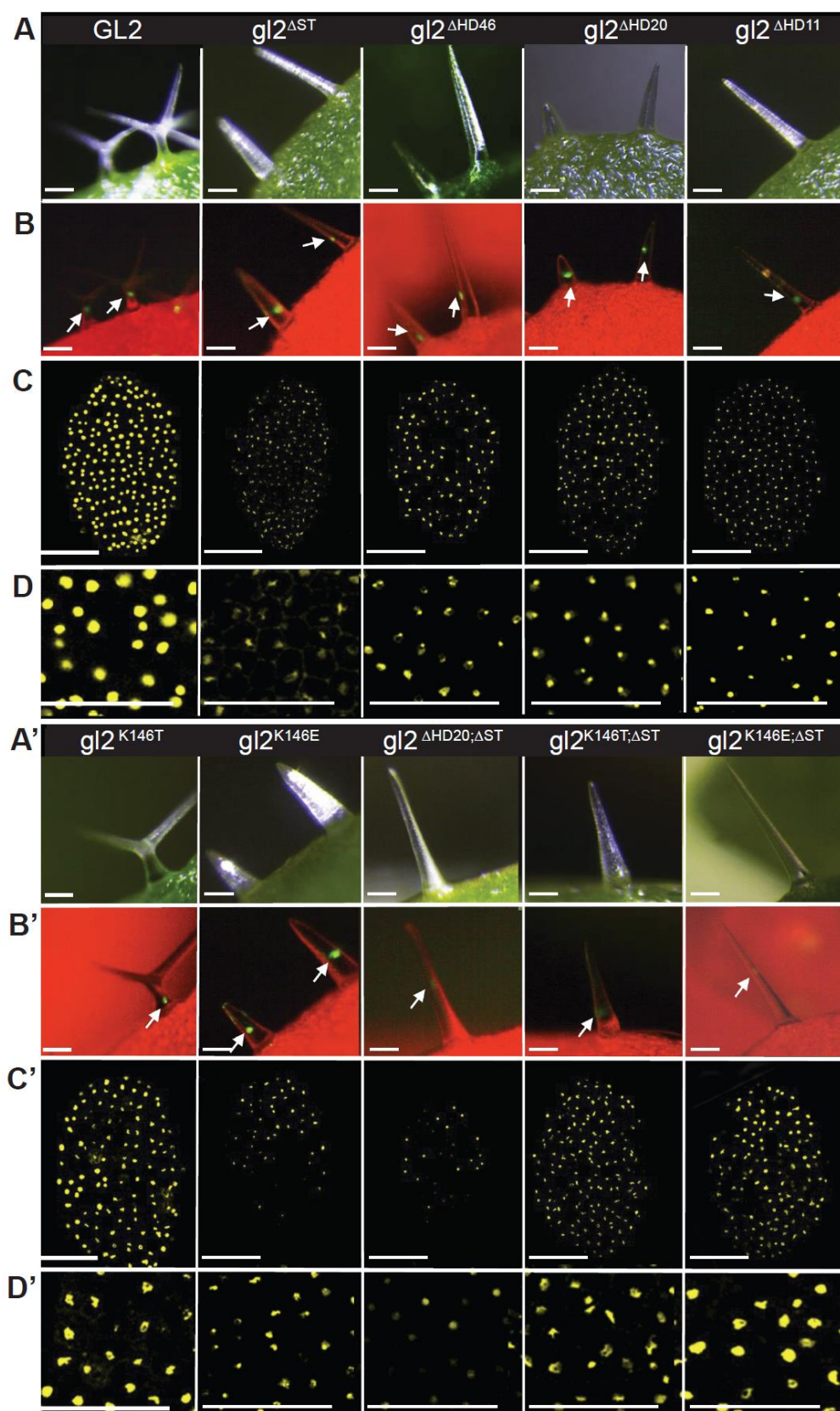


Figure 2.2. Nuclear localization of wild-type EYFP-tagged GL2 and mutant gl2 transcription factors.

(A) and (A') Imaging of leaf trichomes under white light (green, chlorophyll). Bar = 100 μ m. (B) and (B') Matching fluorescence images (red, chlorophyll) indicate nuclear localization of EYFP-tagged GL2, gl2 Δ ST, gl2 Δ HD46, gl2 Δ HD20, gl2 Δ HD11, gl2 Δ K146T, gl2 Δ K146E, gl2 Δ HD20; Δ ST, gl2 Δ K146T; Δ ST and gl2 Δ K146E; Δ ST proteins in trichome cells. Arrows mark the brightest nuclear signal in each image. Bar = 100 μ m. (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 μ m. (D) and (D') Magnified images from (C) and (C'). Bar = 50 μ m.

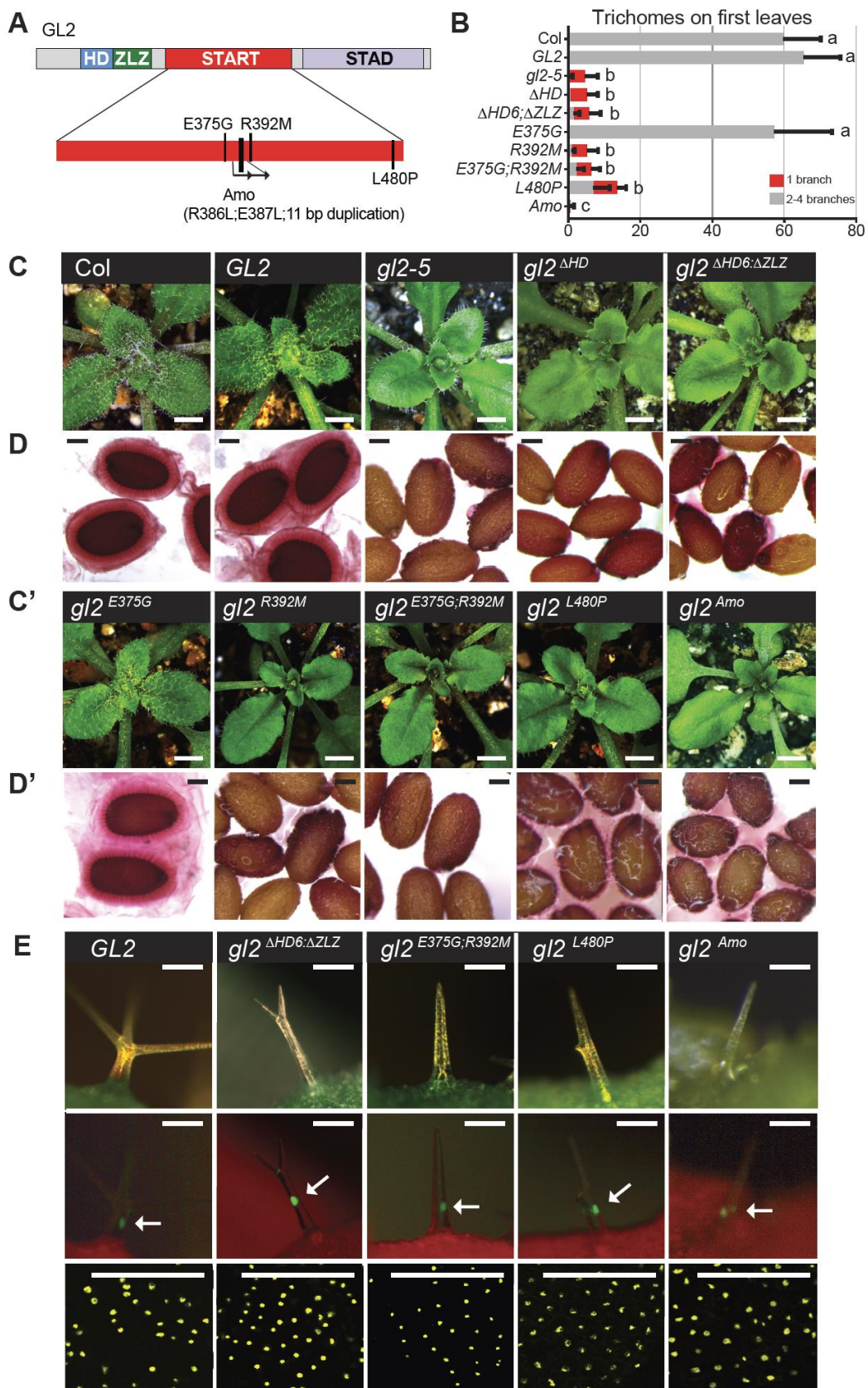


Figure 2.3. START domain missense mutants exhibit loss-of-function phenotypes despite nuclear localization of the corresponding EYFP-tagged proteins.

(A) Schematic showing the functional domains of GL2. Positions of mutations within the START domain are indicated. See **Figure A.1B**. **(B)** Total number and branching pattern of trichomes on the first leaves. Gray bars indicate normal trichomes in plants expressing Col wild type, and EYFP-tagged GL2 and $gl2^{E375G}$. Mutant $gl2-5$ plants and those expressing EYFP-tagged $gl2^{\Delta HD}$, $gl2^{\Delta HD6;\Delta ZLZ}$, $gl2^{R392M}$, $gl2^{E375G;R392M}$, $gl2^{L480P}$, and $gl2^{Amo}$ exhibit unbranched trichomes (red bars). Values are mean $\pm$ SD for $n = 20$ plants. Letters denote significant differences from one-way ANOVA ($p < 0.05$) using Tukey's multiple comparisons test. **(C and C')** Wild-type Col, *EYFP:GL2*, and *EYFP:gl2^{E375G}* expressing rosettes display trichomes covering leaf surfaces. In contrast, $gl2-5$ mutants as well as the EYFP-tagged HD mutants ($gl2^{\Delta HD}$, $gl2^{\Delta HD6;\Delta ZLZ}$) and START domain mutants ($gl2^{R392M}$, $gl2^{E375G;R392M}$, $gl2^{L480P}$, and $gl2^{Amo}$) exhibit trichome differentiation defects. Bar = 1 mm. **(D and D')** Ruthenium red staining of seeds to assay mucilage production. Wild-type Col, *EYFP:GL2*, and *EYFP:gl2^{E375G}* seeds display a normal mucilage layer, while $gl2-5$ mutants as well as the EYFP-tagged HD mutants ($gl2^{\Delta HD}$, $gl2^{\Delta HD6;\Delta ZLZ}$) and START mutants ($gl2^{R392M}$, $gl2^{E375G;R392M}$, $gl2^{L480P}$, and $gl2^{Amo}$) exhibit defective mucilage production. Bar = 100 μ m. **(E)** Nuclear localization of EYFP-tagged GL2 and mutant $gl2$ transcription factors. Top panel: Imaging of leaf trichomes under white light (green, chlorophyll). Bar = 100 μ m. Middle panel: Matching fluorescence images (red, chlorophyll) indicate nuclear localization of EYFP-tagged proteins in trichomes. Arrows mark the brightest nuclear signal in each image. Bar = 100 μ m. Bottom panel: Confocal laser scanning microscopy of developing seeds reveals nuclear expression of EYFP-tagged GL2 and mutant proteins in seed coat cells. Bar = 50 μ m.

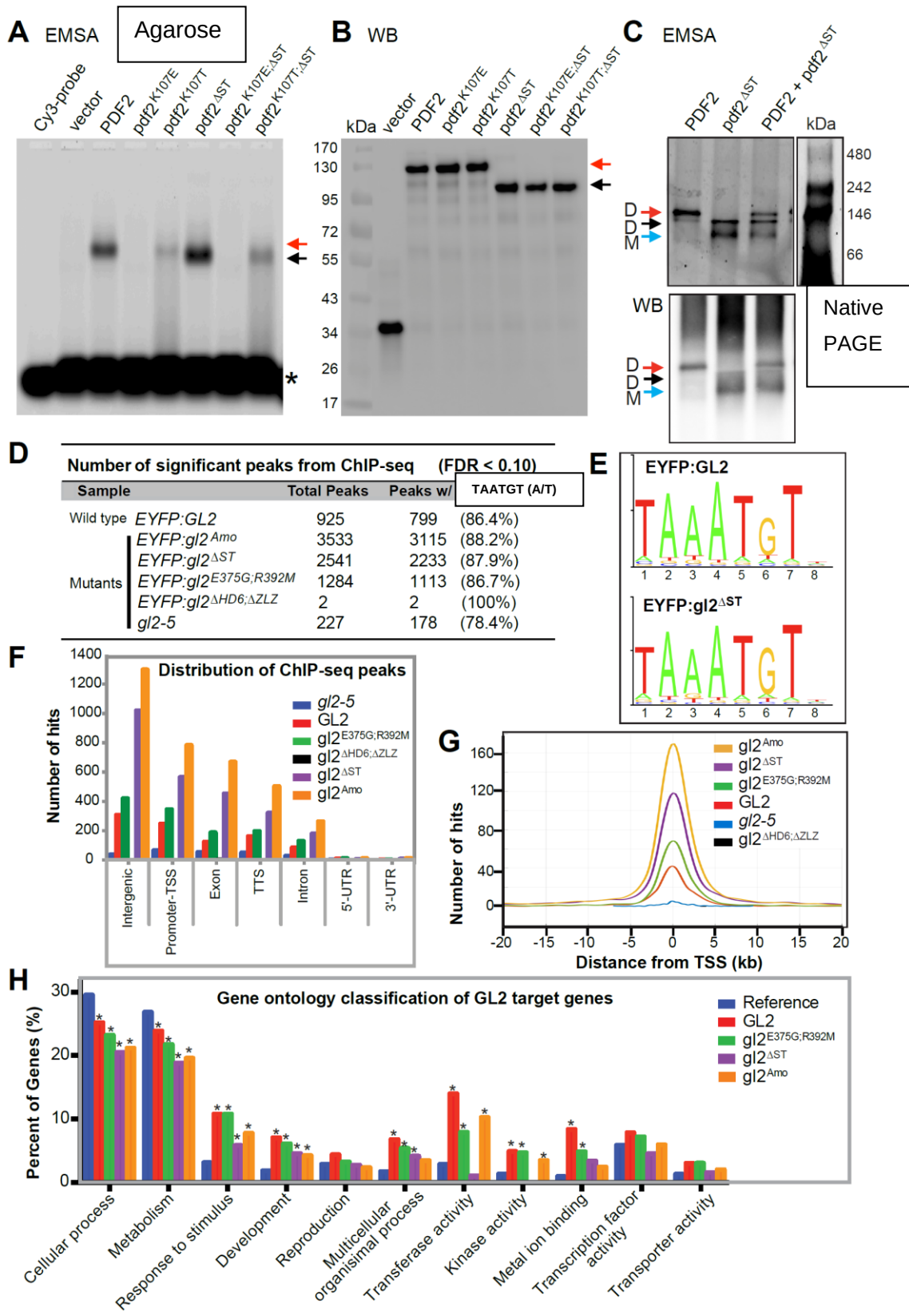


Figure 2.4. HD is required for DNA binding activity of PDF2 and GL2 while START domain is dispensable for DNA binding.

(A) HD is required for *in vitro* DNA binding of PDF2 and pdf2^{ΔST} proteins.

Electrophoretic mobility shift assays (EMSA) reveal band shifts for Halo:PDF2 (red arrow) and Halo:pdf2^{ΔST} (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 probe containing the L1-box TAAATGTA is indicated (asterisk). Representative experiment from three independent experiments. **(B)**

Western blot with anti-Halo Ab detects Halo-tagged proteins assayed for DNA binding in **(A)**, representative of three independent experiments. **(C)** The pdf2^{ΔST} 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^{ΔST} 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^{ΔST} at the expected migration positions matching that of the EMSA (top). Representative blot of two replicates. See **Figure A.4.** **(D-H)** Chip-seq data reveals that wild-type GL2 and mutants with compromised START domain function (gl2^{ΔST}, gl2^{Amo}, gl2^{E375G;R392M}) are proficient in DNA binding *in vivo*. In contrast, DNA binding was abolished in the gl2^{ΔHD6;ΔZLZ} mutant. **(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*. Second column shows number and percentage of ChIP-seq peaks containing one or more binding motifs with at least 80% sequence identity to DNA sequence TAAATGT(A/T). **(E)** Sequence logo representing the information content of the GL2 consensus DNA binding motif identified in 86.4% of the wild-type GL2 and 87.9% of the mutant gl2^{ΔST} ChIP-seq peak regions, respectively. See **Figure**

A.5A. (F) 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. See **Table A.2. (G)** 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 **Figure A.5B. (H)** 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, $p < 0.01$).

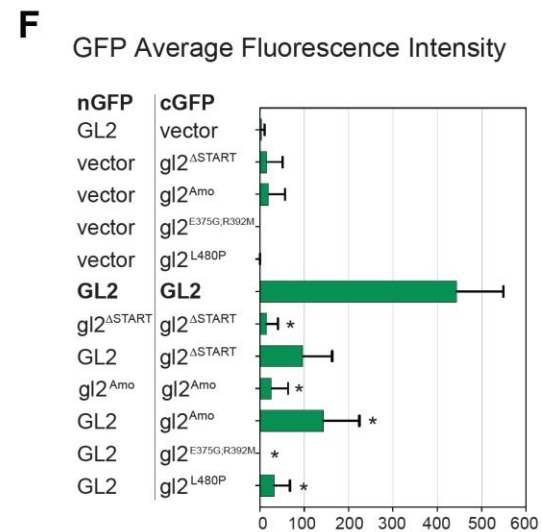
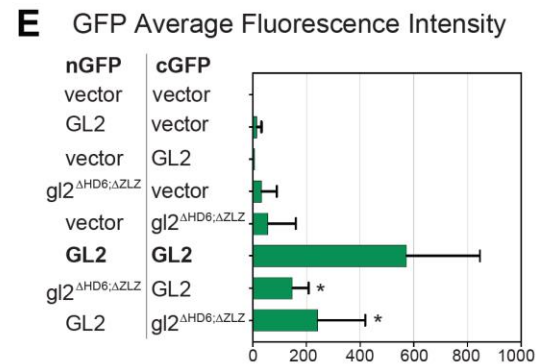
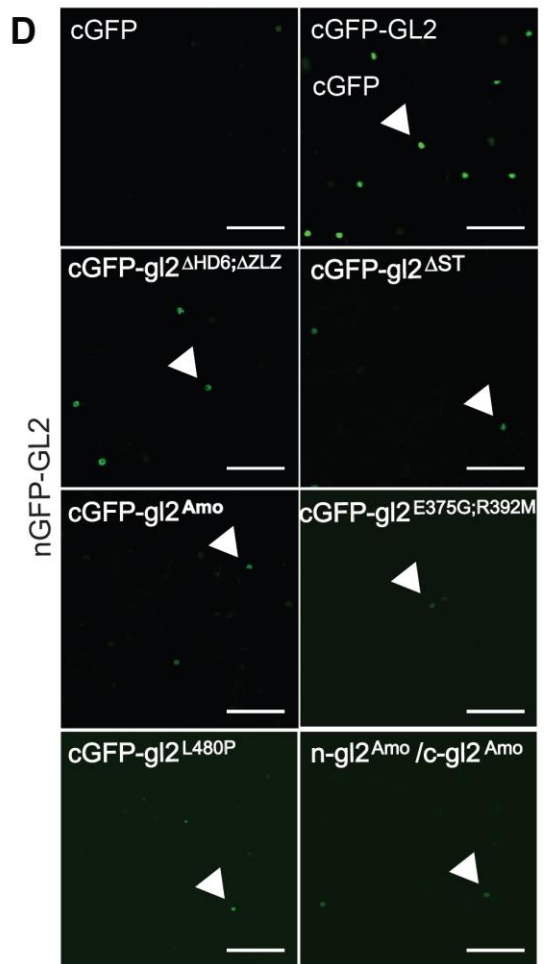
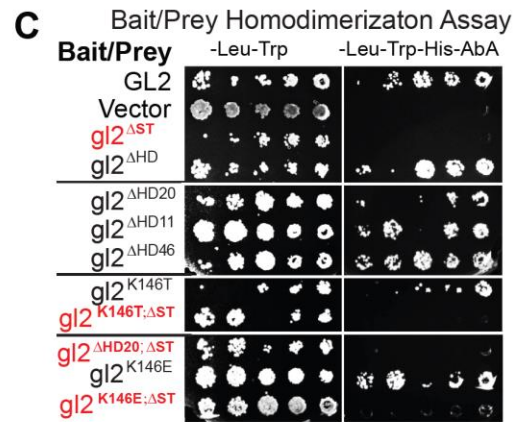
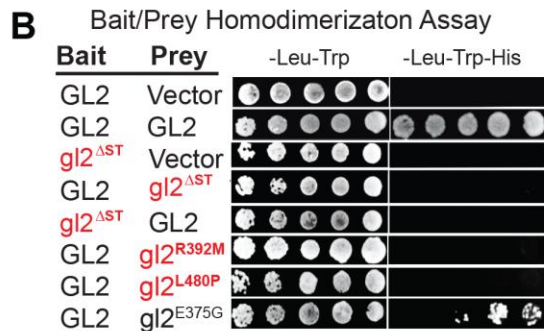
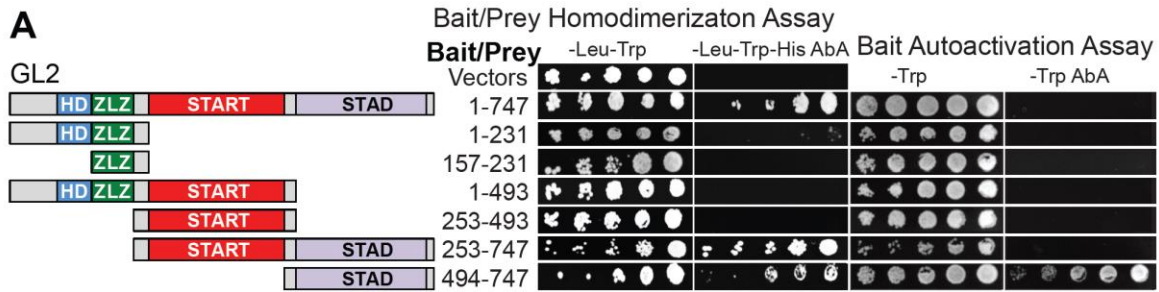


Figure 2.5. START domain is required for homodimerization of GL2.

(A) Schematic representation of full-length GL2 and segments for Y2H assays (left). GL2 homodimerization of full-length protein (1-747) and START-STAD fragment (253-747) was observed (middle). The STAD fragment exhibits bait autoactivation (right). Homodimerization was assayed on -Leu-Trp-His Aureobasidin A (AbA) while bait autoactivation was assayed on -Trp AbA media. See **Figure A.7A**. **(B)** Deletion or missense mutations in the START domain disrupt homodimer formation in Y2H assays. GL2, gl2^{ΔST} and the indicated missense mutants were cloned into bait and prey vectors, co-expressed in yeast strain Y2HGold and assayed on the indicated growth media. Baits in combination with empty prey vectors served as negative controls. **(C)** Y2H assays show that bait and prey fusion proteins interact to homodimerize only when START domain is present. Homodimerization was assayed on selective media as in **(A)**. 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 **Figures A.6C-E and A.7B**. **(D-F)** BiFC assays indicate homodimerization of wild-type GL2 but not HD-ZLZ mutant nor START domain mutants. **(D)** *N. benthamiana* leaf abaxial epidermal cells were transiently transformed with the indicated constructs and silencing suppressor p19. Interaction of full-length GL2 tagged with split GFP at the N-terminus (nGFP-GL2) with vector with the split GFP at C-terminus (cGFP), full-length GL2 and various gl2 mutants. Detection of gl2^{Amo} self-interaction is also shown. Nuclear expression is indicated by white arrows. Scale bar = 50 μm. See **Figure A.8**. **(E-F)** Quantitative analysis of BiFC assays in **(D)**. 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).

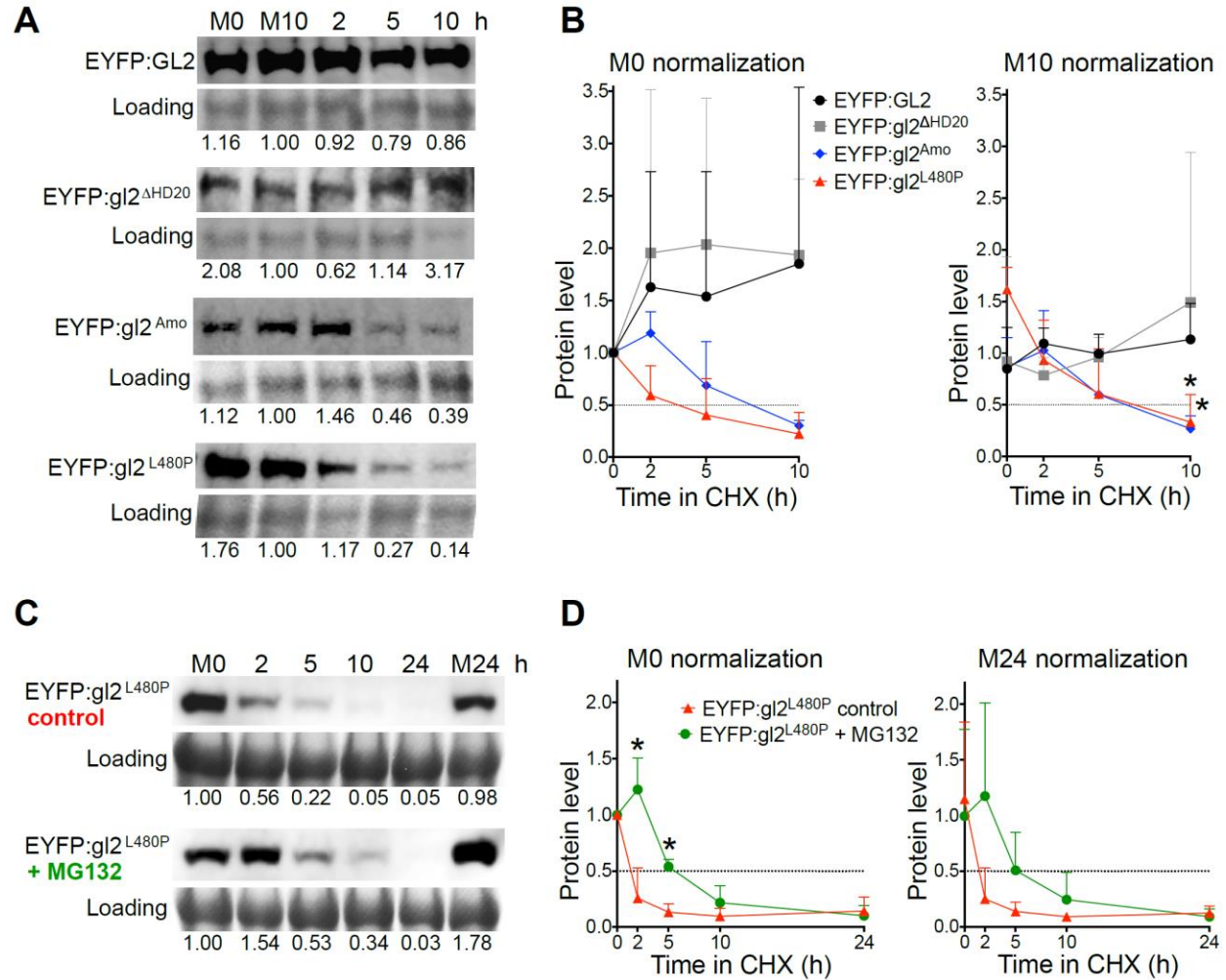


Figure 2.6. Cycloheximide experiments reveal that the START domain, but not HD, is critical for protein stability of GL2.

(A-B) HD mutant gl2^{ΔHD20} 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 μ M) for a 10 h time course. Values were normalized to loading controls and M10. See **Figures A.9 and A.10. (B)** Five-day-old seedlings expressing EYFP-tagged gl2^{L480P} were treated simultaneously with cycloheximide (400 μ M) and DMSO (control) or proteasome inhibitor MG132 (50 μ M) 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: * $p \leq 0.05$.

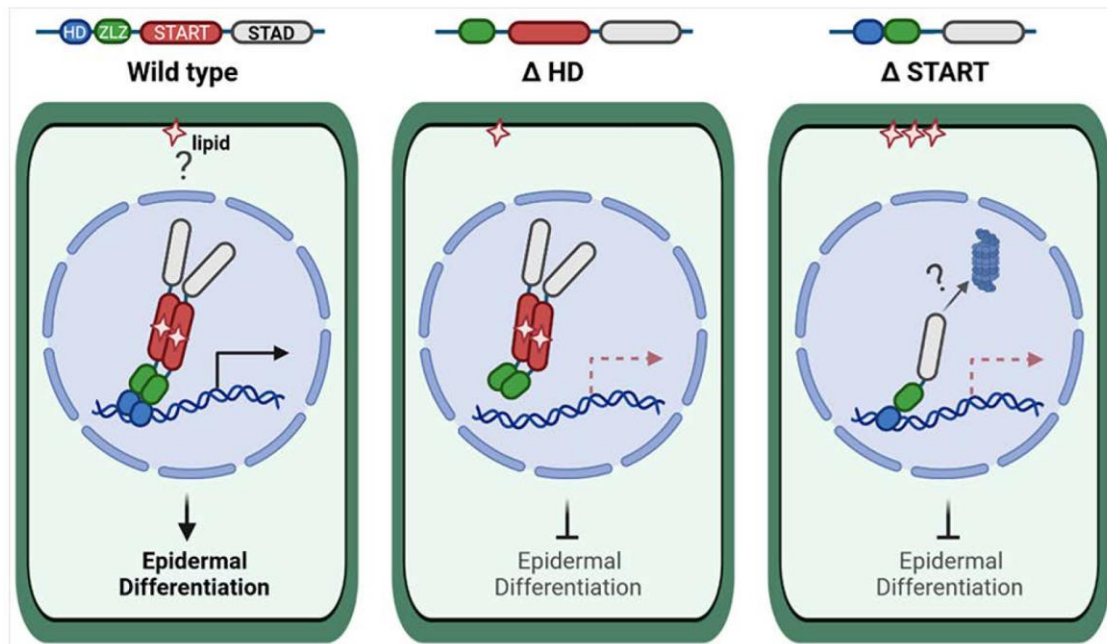


Figure 2.7. Model of HD and START domain roles in HD-Zip IV transcription factors.

Wild-type HD-Zip IV TFs, as exemplified by GL2 (left), 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. Deletion or mutation of the HD (K146E in GL2) (middle) abolish 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. Deletion of or missense mutations in the START domain (right) result 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. This figure was created with BioRender.com.

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. *The 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 quality 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.
- 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 Univ. Press, Oxford, UK.
- 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.
- 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*. *The 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. *The 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*). *The 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*. *The 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 in Genetics* **6**, 323-329.

- Gietz, R.D., and Woods, R.A.** (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**, 87-96.
- 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 Reports* **37**, 653-664.
- 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. *The 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. *PNAS USA* **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. *Molecular & Cellular 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. *Biochemical and Biophysical Research Communications* **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 research* **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 Protoc* **4**, e1096.
- 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.** (2006a). Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in Arabidopsis. *Plant Physiol* **141**, 1363-1375.
- Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., and Takahashi, T.** (2006b). 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. *The 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 in Biochemical Sciences* **24**, 130-132.
- Rerie, W.G., Feldmann, K.A., and Marks, M.D.** (1994). The GLABRA2 gene encodes a homeodomain 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). 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.
- Roostaei, 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. *PNAS USA* **91**, 8393-8397.
- Schofield, P.N.** (1987). Patterns, puzzles and paradigms: the riddle of the homeobox. *Trends in Neurosciences* **10**, 3-6.
- Schrack, 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.
- Schrack, 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*. *PNAS USA* **81**, 4115-4119.
- Scott, S.P., Teh, A., Peng, C., and Lavin, M.F.** (2002). One-step site-directed mutagenesis of ATM cDNA in large (20kb) plasmid constructs. *Hum Mutat* **20**, 323.
- 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 Genetics* **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. *The 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. *The 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. *PNAS USA* **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. *PNAS* **99**, 6943-6948.
- Sparkes, I.A., Runions, J., Kearns, A., and Hawes, C.** (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc* **1**, 2019-2025.
- 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., Schöler, 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. *Nature Structural Biol* **7**, 408-414.
- Voinnet, O., Rivas, S., Mestre, P., and 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 Journal* **33**, 949-956.
- 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. *The 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. *PNAS USA* **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., Hinch, 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.**

Chapter 3 - Post-translational regulation of HD-Zip IV transcription factors in Arabidopsis

Abstract

Class IV homeodomain leucine-zipper (HD-Zip IV) transcription factors (TFs) contain a plant-specific domain configuration comprised of a DNA-binding HD, a Zipper-Loop-Zipper (ZLZ) dimerization domain and a lipid-sensing START domain. Several HD-Zip IV members control formation of the epidermis, the outer layer that is critical for protection of plants from environmental stressors. GLABRA2 (GL2), a representative HD-Zip IV family member, is required for proper differentiation of trichome cells, non-root hair cells, and mucilage secretory cells in seeds. Previously, the START domain was shown to be important for TF dimerization, stability and activity. In eukaryotes, post-translational modifications (PTMs) serve as marks for a protein's subcellular localization, turnover, and activity. However, the precise mechanism controlling the level of HD-Zip IV TFs at the post-translational level is poorly characterized. Using *in vivo* cycloheximide chase experiments, we observed increased stability of GL2 lacking the ZLZ in both wild type and START domain mutants, uncovering a role for the ZLZ in modulating GL2 protein levels. Live imaging of trichomes and roots revealed nuclear localization of gl2^{AZLZ} proteins despite a loss-of-function phenotype for the corresponding mutants. Additionally, we used machine learning tools to predict multiple, high confidence conserved ubiquitination and SUMOylation sites, and evaluated possible GL2 modifications *in vivo* by performing immunoprecipitation experiments. Site-directed mutagenesis of the candidate amino acids resulted in a gain-of-function phenotype characterized by an increased frequency of

abaxial leaf curling that correlated with expression levels. Our findings support the hypothesis that regulatory mechanisms involving ubiquitin and/or SUMO contribute to GL2 protein levels and activity in plants.

Keywords: HD-Zip IV transcription factors, Arabidopsis, Zipper-Loop-Zipper, stability, cycloheximide, ubiquitination

Introduction

Homeodomain transcription factors (TFs) are key regulators of gene expression during development of multicellular organisms. The Arabidopsis genome encodes 16 TFs from the HD-Zip IV subfamily that are preferentially expressed in plant epidermal or sub-epidermal cells and are involved in differentiation of outer cell layers (Rerie et al., 1994; Abe et al., 2001; Abe et al., 2003; Ito et al., 2003; Gao et al., 2015), anthocyanin pigment accumulation (Kubo et al., 1999), lipid transport (Javelle et al., 2010), cuticle deposition (Javelle et al., 2010; Wu et al., 2011) and plant responses to stress factors (Yu et al., 2008; Cao et al., 2009; Wei et al., 2010; Yu et al., 2013; Li et al., 2016; Yu et al., 2016; Zhu et al., 2016; Chen et al., 2017; Banavath et al., 2018; Guo et al., 2019; Zhang et al., 2019; Sharif et al., 2021; Wang et al., 2022). The founding member of this subfamily is GLABRA2 (GL2) which promotes trichome initiation in shoots (Rerie et al., 1994) but negatively regulates hair formation from hairless cell files in roots (Di Cristina et al., 1996; Masucci et al., 1996; Wang et al., 2020). In seeds, it is involved in the differentiation of mucilage secretory cells (Western et al., 2001; Western et al., 2004) as well as regulation of seed oil production (Shen et al., 2006; Shi et al., 2012). Additionally, GL2 negatively regulates anthocyanin accumulation, an important component of the plant stress response (Wang et al., 2015). PROTODERMAL FACTOR2 (PDF2), another member of this

subfamily, functions redundantly with its paralog *Arabidopsis thaliana* MERISTEM LAYER1 (ATML1) in shoot epidermal differentiation during embryogenesis and in hormonal regulation of cell expansion (Abe et al., 2003; Kamata et al., 2013; Rombolá-Caldentey et al., 2014; Ogawa et al., 2015). Among agriculturally important crops, HD-Zip IV TFs are involved in the regulation of fiber differentiation and seed coat development in cotton (Guan et al., 2008; Zhang et al., 2010; Walford et al., 2012; Shan et al., 2014), anther and kernel development in maize (Ingram et al., 1999; Khaled et al., 2005; Vernoud et al., 2009) and bulliform cell development as well as bulliform cell-dependent leaf rolling in rice (Zou et al., 2011; Fang et al., 2021).

HD-Zip IV transcription factors have a complex domain topology in which each domain has a distinct structure and function contributing to overall TF activity. The HD is a 60 amino-acid helix-turn-helix domain comprised of three α -helices and a disordered amino-terminal arm located upstream of the first helix (Gehring et al., 1990; Kissinger et al., 1990; Liu et al., 1990; Otting et al., 1990). The helix III (recognition helix) binds the major groove of target DNA with high affinity while the amino-terminus contacts the DNA minor groove ((Schofield, 1987; Gehring, 1994); for review see (Burglin and Affolter, 2016)). Leucine zippers (LZs) are coiled coil motifs that are formed by two or more alpha helices winding around each other (Landschulz et al., 1988; Vinson et al., 1993). In HD-Zip IV TFs, the LZ is located carboxy-terminal to the HD, and it is intervened by a disordered loop of 10–20 amino acids (hence termed zipper-loop-zipper, ZLZ) (Yang et al., 2002) comprising two sets of three heptad repeats (Ruberti et al., 1991; Di Cristina et al., 1996; Chan et al., 1998; Nakamura et al., 2006; Ciarbelli et al., 2008). The heptad repeats consist of alternating three- and four-residue sequences with the leucine residues in the seventh position (Krylov and Vinson, 2001). The ZLZ is thought to be required for the dimerization of HD-Zip IV TFs (Di Cristina et al., 1996). Upon dimerization, the LZ adds

extra specificity in the HD-mediated DNA binding by correct positioning of the HD (Vinson et al., 1989; Palena et al., 1999; Tron et al., 2004). Accordingly, the $gl2^{\Delta ZLZ}$ mutant fails to bind DNA as indicated by ChIP-seq, and thus lack transcriptional activity (Mukherjee et al., 2022). Two invariant cysteine residues (CXXC) within the loop of the ZLZ were previously implicated in the redox-dependent regulation of oligomerization and DNA binding activity of HD-Zip TFs (Tron et al., 2002).

Carboxy-terminal to the ZLZ, the Steroidogenic Acute Regulatory (StAR) protein-related lipid Transfer (START) domain comprises of ~210-amino acids in the middle of the protein (Ponting and Aravind, 1999; Schrick et al., 2004; Mukherjee et al., 2009). It adopts an α/β helix-grip fold, creating a hydrophobic pocket that is implicated in binding to lipophilic ligands (Kallen et al., 1998; Tsujishita and Hurley, 2000; Soccio and Breslow, 2003; Thorsell et al., 2011; L  tourneau et al., 2012). Several reports indicate the role of plant-derived START domains in lipid transport and sensing (Gou et al., 2015; Hirashima et al., 2021; Nagata et al., 2021). A couple of studies have also highlighted the importance of the START domain in controlling the protein stability of HD-Zip IV transcription factors. In the case of ATML1, mutation of a candidate lipid contact site in the START domain was shown to abolish the expression of the fluorescently tagged protein in lateral roots (Nagata et al., 2021). In our recent study concerning GL2, the expression of EYFP-tagged proteins with START domain missense mutations predicted to disrupt ligand binding was much weaker in comparison to WT (Mukherjee et al., 2022). Quantification of protein-half-lives in young seedlings revealed highly reduced stability in these mutants with a possible involvement of the Ubiquitin 26S Proteasome System (UPS) in protein degradation (Mukherjee et al., 2022).

Post-translational modifications (PTMs) dynamically regulate diverse aspects of TFs including their stability, sub-cellular localization, protein-protein/DNA interactions (Schütze et al., 2008; Filtz et al., 2014). Two PTMs extensively used by cells are ubiquitination and SUMOylation in which small polypeptides called Ubiquitin and SUMO (Small Ubiquitin-like Modifier) are covalently attached to the target protein, respectively (Matunis et al., 1996; Kerscher et al., 2006; Vierstra, 2012). Ubiquitination has a major role in the regulation of protein degradation through the UPS, which is responsible for degradation of up to 90% proteins in eukaryotic cells (Moon et al., 2004; Smalle and Vierstra, 2004; Finley, 2009; Sharma et al., 2016). In *Arabidopsis*, more than 1,600 genes encode components of the UPS, comprising about 6% of the whole genome (Vierstra, 2009, 2012). Similarly, SUMOs regulate subcellular localization, nuclear transport, stability and protein-protein interactions of target proteins and they are also involved in DNA replication and repair (Lois et al., 2003; Novatchkova et al., 2004; Elrouby et al., 2013; Miller et al., 2013; Nelis et al., 2015; Benlloch and Lois, 2018; Accossato et al., 2020; Bernula et al., 2021; Han et al., 2021; Clark et al., 2022). One hypothesis is that PTMs serve as marks to degrade HD-Zip IV TFs in absence of a stabilizing ligand. However, a detailed mechanism controlling HD-Zip IV protein levels during epidermal differentiation in plants has not been identified to date. This presents a gap in our knowledge of how these developmentally important TFs regulate gene expression.

In this study, we monitored the half-life of $gl2^{\Delta ZLZ}$ proteins in presence or absence of a functional START domain, uncovering a novel role for the ZLZ in destabilizing protein levels. GL2 PTMs were detected by immunoisolation from plants followed by immunoblotting with ubiquitin- or SUMO-specific antibodies. In addition, we identified multiple, conserved ubiquitination and SUMOylation sites in GL2 by in silico analysis. Site-directed mutagenesis of

the selected lysines resulted in increased prevalence of abaxial leaf curling despite nuclear localization of protein and rescue of trichome defects in most of the mutants. Intriguingly, the curly leaf phenotype correlated with the expression of transgenes containing the EYFP tag, suggesting a novel role for EYFP-induced GL2 stabilization in leaf polarity establishment.

Materials and methods

Plant materials and growth conditions

All plants used in this study were of the *Arabidopsis thaliana* Columbia (Col-0) ecotype. *gl2-5* represents an En-1 insertion allele of *GL2* (Ohashi et al., 2003; Khosla et al., 2014). *Arabidopsis* plants were grown at 23°C under continuous light on soil comprising Berger BM6, vermiculite and perlite (Hummert International) in a ratio of 4:2:1. For protein stability and pull-down experiments, seeds were vapour-sterilized and normally sown onto 0.5X MS media containing 0.215% Murashige and Skoog (MS) basal salt mixture (Sigma-Aldrich, M5524) (Murashige and Skoog, 1962), 1% Sucrose, 0.05% MES buffer at pH 5.8, 0.01% Myo-inositol and 0.5% agar (final pH adjusted to 5.8). To study the effects of salt stress, 100mM NaCl was incorporated in 0.5X MS media. For phosphate starvation experiments, vapor-sterilized seeds were sown onto Pi sufficient (P⁺) media containing 20.6 mM NH₄NO₃, 2.26 mM CaCl₂ dihydrate, 0.759 mM MgSO₄ heptahydrate, 18.8 mM KNO₃ and 1.25 mM KH₂PO₄ monobasic, MS micronutrient solution (PhytoTech Labs, M529), 1% Sucrose, 0.05% MES at pH 5.7 and 0.8% agar or Pi limiting (P⁻) media with similar composition but lacking KH₂PO₄ monobasic. After stratifying the seeds at 4°C for 4-5 days, the plates were placed vertically along the sidewalls in the growth chamber at 23°C under continuous light for various lengths of time depending upon the nature of experiment.

Generation of DNA constructs

The *proGL2:EYFP:GL2* and *proGL2:6xHis 3xFLAG:GL2* constructs were generated as previously described (Schrack et al., 2014; Subedi and Schrack, 2022). The K-R, C-A and deletion mutation of the ZLZ domain were introduced in the TOPO-cloned *GL2* using Q5 Site-Directed Mutagenesis Kit (New England Biolabs, E0554S) or Phusion High Fidelity DNA polymerase (ThermoFisher Scientific) and the primers listed in **Table B.1**. Multiple K-R mutations in ZLZ domain were mostly generated using mutant templates containing lower order K-R mutations but a few multiple mutants were also generated using WT template and the primers with multiple mutations. The 8^{KR} mutation was generated by combining K129R and 7^{KR} mutations. The *gl2 Δ ZLZ; W279R*, *gl2 Δ ZLZ; L480P* and *gl2 Δ ZLZ; Δ ST* double mutants were generated by introducing Δ ZLZ mutation in the corresponding START domain single mutants obtained from the previous studies (Wojciechowska et al., 2021; Mukherjee et al., 2022). The mutations were confirmed by sequencing TOPO constructs using the primers listed in **Table B.1**. The 42^{KR} mutation was generated through synthetic approach (Twist Bioscience). Briefly, a silicon-based, high-throughput DNA synthesis platform was used to synthesize a clonal gene which is a double-stranded DNA cloned into the D-TOPO/pENTR vector and NGS sequence verified.

To transfer the sequences from the D-TOPO vector into the SR54 binary vector, the inserts were PCR amplified with Q5 High Fidelity DNA polymerase (New England Biolabs, M0491S) using the primers listed in **Table B.1**. The NEBuilder HiFi DNA Assembly kit (New England Biolabs, E5520S) was used to insert the purified amplicons into the empty vectors obtained by cleaving *proGL2:EYFP:GL2* and *proGL2: 6xHis 3xFLAG:GL2* with *Sall* and *KpnI*. The *EYFP:gl2^{N-term KR}*, *EYFP:gl2^{START KR}* and *EYFP:gl2^{STAD KR}* constructs were generated by amplifying corresponding fragments of *GL2* using WT and 42^{KR} templates and combining WT

and KR-containing fragments appropriately. While the WT and all mutant *GL2* sequences were used to generate binary vector constructs containing EYFP tag, only selected sequences were used with 6xHis 3xFLAG tag. All the mutant constructs with FLAG tag were exclusively generated in this study. The binary vector constructs were further verified by sequencing using the primers listed in **Table B.1**.

Arabidopsis transgenics

The binary vector DNA constructs were transformed into *Agrobacterium* strain GV3101 (MP90) by electroporesis using BioRad. The transgenic lines were generated by *Agrobacterium*-mediated transformation of *Arabidopsis* using floral dip technique (Clough and Bent, 1998). The *proGL2:EYFP:GL2*, *proGL2:FLAG:GL2* and corresponding mutants were generated in *gl2-5* background as described in (Schrack et al., 2014; Subedi and Schrack, 2022). The generation of *proGL2:EYFP:PDF2* and corresponding Δ START mutant in *gl2-5* background and *HA:PDF2* in Col-O background is described elsewhere (Wojciechowska et al., 2021). The T1 transformants were selected on agar medium containing 20 μ g/ml Hygromycin B. T2 progeny derived from independent transformants were screened for Mendelian segregation patterns of 3:1. Finally, homozygosity was confirmed in selected T3 lines that were genotyped by PCR using the primers are listed in **Table B.1** and mutations were further confirmed by Sanger sequencing.

Bioinformatical analysis

Ubiquitination sites in HD-Zip IV TFs were probabilistically determined by using the random forest-based predictor UbPred (Radivojac et al., 2010). SUMOylation sites were predicted in HD-Zip IV TFs using web servers Abcepta SUMOplotTM (Abcepta, 2013), JASSA (Joined Advanced SUMOylation Site and SIM Analyser) (Beauclair et al., 2015) and

GPS-SUMO 2.0 (Zhao et al., 2014). Scoring of candidate SUMO attachment sites was done using Abcepta SUMOplot™ (Abcepta, 2013). Protein disorder in GL2 was predicted using PONDR® (Predictor of Natural Disordered Regions) VSL2 Predictor (Obradovic et al., 2003; Peng et al., 2006). GL2 orthologs from representative monocots and dicots obtained with BlastP (Altschul et al., 1990) were aligned using Clustal Omega (Madeira et al., 2019). Sequence alignments were visualized using Jalview (Waterhouse et al., 2009). Structural modeling of GL2 START domain with or without K-R mutations was performed using I-TASSER (Iterative Threading ASSEmbly Refinement) server (Roy et al., 2010; Yang and Zhang, 2015a; Yang and Zhang, 2015b; Zhang et al., 2017). The images were created using RasWin Molecular Graphics program (Bernstein, 2009).

Protein extraction and immunoblot analysis

For protein extraction, frozen plant tissues (leaves or whole seedlings) were homogenized in liquid nitrogen followed by addition of hot SDS buffer (8 M urea, 2% SDS, 0.1 M DTT, 20% glycerol, 0.1 M Tris pH 6.8, 0.004% bromophenol blue) and boiling for 10 mins. The protein samples were resolved on Bio-Rad Mini-PROTEAN TGX gels (10%) by performing SDS-PAGE, transferred into PVDF membranes. The membranes were blocked with 5% (w/v) non-fat milk in TBS with 0.1% Tween-20 at room temperature and probed with the respective antibodies. To visualize the proteins, membranes were incubated with SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific, 34095) for 5 minutes and the blots were developed with Azure 300 chemiluminescence imager (Azure Biosystems).

To strip the membranes for reprobing, they were first placed in a mild stripping buffer containing 200 mM glycine, 0.1% (w/v) SDS and 1% (v/v) Tween 20 at pH 2.2 twice with each incubation done for 5-10 mins. The membranes were washed twice with each of 1x PBS and 1x

TBS-T, allowing 10 mins for each wash. After developing the blot images, the membranes were stained with Bio-Safe Coomassie Blue G-250 (Bio-Rad) to confirm the equal loading of the protein samples in each lane.

Cycloheximide chase experiments

Protein half-lives were measured in Arabidopsis following a protocol developed by (Guo and Yin, 2019). 4-5 days old seedlings (20-30 seedlings per genotype) were transferred from 0.5X MS, P+ or P- solid media into corresponding liquid media in 24-well plates. The seedlings were allowed to grow for another 16 h at 23°C under continuous light on a shaking platform at 55 rpm. Cycloheximide (Sigma-Aldrich, C1988) (final concentration 400 μ M) or 5 μ l of DMSO (control treatment) was added to each well. For the inhibition of proteasome and autophagy (Lee and Goldberg, 1998), cycloheximide was added together with MG132 (carbobenzoxy-L-leucyl-L-leucyl-L-leucinal; EMD Millipore Corp., 474787) (final concentration 50 μ M) and Concanamycin A (Santa Cruz Biotechnology, SC-202111A) (final concentration 1 μ M) respectively. Seedlings were collected at various intervals over a 10, 24 or 48 h time course which were frozen in liquid nitrogen and stored at -80°C in screw-cap microcentrifuge tubes prior to protein extraction. For Western blot analysis, the membranes were incubated with primary antibodies against GFP (1:2000; Roche, 11814460001), FLAG (1:10,000; Sigma, F1804) or HA (1:10,000; Pierce, 26183) followed by goat anti-mouse secondary IgG HRP antibodies (1:3300; GenScript, A00160). Band intensities were quantified with ImageJ after blot imaging.

Pull-down experiments

Arabidopsis seedlings were grown on normal growth media (0.5X MS or P+) or under various stress conditions such as phosphate starvation (P-), high salinity (100 or 150 mM NaCl) or exposure to oxidative stress (10 mM H₂O₂ for 4 h) and heat shock (37 °C for 30 mins) prior to harvesting them. Growth was carried out for 7-10 days under continuous light or for ~2 weeks under 12 h light/12 h dark conditions. For experiments with FLAG-tagged proteins, seedlings were treated with 100 µM MG132 (EMD Millipore Corp., 474787) for 4 h prior to their harvest. Whole seedlings were harvested in the liquid nitrogen and stored at -80 °C until further use. For total protein extraction, about 7-10 g of plant materials were ground in liquid nitrogen and resuspended in 7-10 ml of lysis buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 5% glycerol, 1% (v/v) Triton X-100, 1 mM DTT, 1 mM PMSF or 1x protease inhibitor cocktail (Sigma, P9599), 50 mM N-ethylmaleimide (NEM) (Sigma, E3876) or Iodoacetamide (IAA) (Sigma, 16125) and 100 µM MG132. The lysates were centrifuged at 20,000x g for 15 min at 4°C to obtain supernatants that were clarified by filtration through miracloth when necessary. Next, 25 µl of GFP-Trap® Magnetic Agarose beads (ChromoTek, gta-20) or DYKDDDDK Fab-Trap™ Agarose (ffa-2 or ffa-ts) equilibrated with the lysis buffer were added to the clear supernatants in 15-ml facon tubes that were then rotated end-over-end for 1 hr at 4°C allowing protein binding. The beads were harvested by centrifugation at 2,500x g for 5 min at 4°C and washed with the lysis buffer for 5 times at 4°C. For experiments in **B.3D-F**, highly stringent washing steps under denaturing conditions (8M Urea and 1% SDS) were applied to remove unspecifically bound proteins. The beads were boiled in 70-100 µl of 2x SDS-sample (Laemmli) buffer for 5 min to dissociate immunocomplexes, and clear supernatants were obtained by separating the beads magnetically or by centrifugation.

The lysate (input), supernatant (flow-through/non-bound fraction) and eluate (bound fraction) were analysed by Western blot. To detect ubiquitination, nondiscriminating anti-Ub P4D1 antibody (1:2000; Santa Cruz Biotechnology, 8017) or eBioP4D1 (biotin conjugated) (1:2000, Invitrogen, 13-6078-82) recognizing both monoubiquitin and polyubiquitin chains and FK1 (1:2000; EMD Millipore Corp., 04-262) and FK2 antibodies (1:2000; EMD Millipore Corp., ST1200) binding preferentially to polyUb chains (Fujimuro et al., 1994; Barberon et al., 2011; Martins et al., 2015) were used. To detect SUMOylation, anti-SUMO1 (1:10,000; Abcam, ab5316 or 1:10,000; Agrisera, AS08 308) and anti-SUMO3 (1:10,000; Abcam, ab5317) antibodies were used as primary antibodies. The secondary antibodies used against anti-Ub and anti-SUMO primary antibodies were goat anti-mouse IgG HRP secondary antibodies (1:3300; GenScript, A00160) and goat anti-rabbit IgG HRP secondary antibodies (Enzo, ADI-SAB-300-J) respectively except for eBioP4D1 that was detected by HRP-conjugated avidin (1:10,000; ThermoFisher Scientific, 21123).

Phenotypic characterization and microscopy of Arabidopsis

The images of rosettes, trichomes and EYFP signals were taken with a Leica M165 fluorescence stereomicroscope fitted with a GFP2 filter set and a Leica DFC295 digital camera using Leica Application Suite 4.1 software. The trichomes and their branches were counted on the first leaves of 20 plants for each genotype using the same stereomicroscope. The subcellular localization of EYFP-tagged proteins was detected in 4-d-old seedlings using a Zeiss LSM 880 confocal microscope after staining with propidium iodide (10 µg/ml) to delineate the cell boundaries. While a 561 nm DPSS excitation laser with 575–645 nm emission filter was used to detect propidium iodide, a 488 nm Argon excitation laser with 500–550 nm emission filter was used for capturing EYFP signals. Raw images were processed using Adobe Photoshop.

The distances between the leaf edges halfway from the leaf blades were measured on the 6th leaves (Farmer et al., 2013) of ~25-d-old plants. The extent of leaf curling was quantified by calculating Leaf Rolling Index (LRI) or Degree of epinasty using the formula $LRI = (1 - W_n/W_m) \times 100$ (Shi et al., 2006; Fierro et al., 2015), where W_m is the maximum possible width of the flattened leaf blade and W_n is the natural width of leaves.

Statistical analysis

Leaf biometry and trichome count

Leaf dimensions were measured on the 6th leaves from 25-d-old homozygous plants. The trichome count and quantification of trichome branching defects were done on the first leaves from 10-d-old homozygous plants. For each genotype, 15-20 plants were included. Values reflect mean $\pm$ SD. The significance of differences was tested by applying ordinary one-way analysis of variance (ANOVA) using Tukey's multiple comparisons test (* $p < 0.05$, **** $p < 0.0001$, **** $p < 0.0001$).

Protein stability assays

The cycloheximide experiment data were analyzed using two or three biological replicates except the one obtained by growth in presence of 100 mM NaCl for which only one biological replicate was used. Protein bands from blot images were quantified based on their intensity and the resulting numerical values were normalized to loading control and a mock sample that was assigned a value of 1.0. The time in x-axis corresponding to the protein level of 0.5 in y-axis was considered as the half-life of a particular protein. Values reflect mean $\pm$ SD.

The statistical significance was tested using unpaired *t*-tests (two-tailed distribution) or ordinary two-way ANOVA using Tukey's multiple comparisons test (**p* < 0.05, ***p* < 0.005).

Results

GL2 protein level is negatively regulated by the ZLZ in Arabidopsis

Previously, we showed a rapid drop of GL2 levels in several START domain mutants unlike wild-type (WT) plants or HD mutants in which the protein levels remained fairly constant (Mukherjee et al., 2022). However, the possible involvement of other regions of GL2 including its 65-amino-acid ZLZ in maintaining protein stability has not been investigated. To determine the role of ZLZ in GL2 protein turnover in Arabidopsis, we treated WT and *gl2^{ΔZLZ}* mutant seedlings with cycloheximide (CHX) in a time-course experiment followed by Western blotting (WB) of the lysates with tag-specific antibodies. CHX inhibits eukaryotic translational elongation by interfering with the translocation step in protein synthesis and is broadly used in cell biology to determine the half-life of proteins (Kao et al., 2015; Buchanan et al., 2016; Guo and Yin, 2019).

Quantification of blot images showed that EYPG:GL2 and EYFP:*gl2^{ΔZLZ}* proteins remain stable for >48 hours of CHX treatment (**Figures 3.1A and 3.1B**). However, EYFP:*gl2^{ΔZLZ}* was significantly more stable with a half-life of >48 h, in contrast to EYFP:GL2 that had a half-life of ~48 h (**Figures 3.1A and 3.1B**). Furthermore, we carried out additional protein stability experiments by growing the seedlings in the presence of 100 mM NaCl under 12/12-hour light/dark conditions to determine the role of stress conditions in GL2 stability. Under these conditions, the difference between the protein levels of EYPG:GL2 and EYFP:*gl2^{ΔZLZ}* were more pronounced which was evident after 10 h (**Figures 3.1C and 3.1D**). When time-course

experiments were performed with Arabidopsis lines expressing FLAG-tagged proteins, overall protein levels in both WT seedlings and *gl2*^{ΔZLZ} mutant were lower; however, the latter still showed significantly higher stability with the protein half-life of >24 hours compared to the former that had the protein half-life of ~18 h (**Figures 3.1E and 3.1F**). These results indicate that the ZLZ is required for the degradation of GL2 with the influence of stress conditions in further modulating the protein levels.

The ZLZ downregulates GL2 levels independently of START mutations

Since the START and ZLZ domains appear to regulate GL2 protein levels in opposite ways, we asked whether the deletion of ZLZ can restore the protein stability of START domain mutants. For this, we combined ΔZLZ mutation with three separate START domain mutations ΔST, W279R and L480P. The first represents the START domain deletion and the remaining two are missense mutations in the START domain. Specifically, the L480P mutation affects a predicted lipid binding site in the C-terminal α-helix of the START domain (Alpy and Tomasetto, 2005; Roostaei et al., 2009; Wojciechowska et al., 2021). Similarly, the W279R mutation affects a predicted hydrophobic region of the ligand binding pocket (Mukherjee et al., 2022). All these double mutations were expressed using EYFP tag under the native *GL2* promoter with the *gl2*^{ΔZLZ;L480} mutation additionally expressed using 6x His 3x FLAG tag.

To quantify the protein levels, we selected the double mutant with the L480P mutation. EYFP:gl2^{L480P} was previously shown to be detectable in WBs despite being more unstable than EYFP:GL2 (Mukherjee et al., 2022). Here, the *gl2*^{ΔZLZ;L480P} double mutant exhibited less steep drop in protein levels compared to *gl2*^{L480P} mutant which was more evident in the FLAG tag background (**Figures 3.1G-3.1J**). Specifically, the protein half-lives of EYFP:gl2^{L480P} and EYFP:gl2^{ΔZLZ;L480P} were ~4 h and ~12 h respectively (**Figures 3.1G and 3.1H**) whereas those of

FLAG:gl2^{L480P} and FLAG:gl2^{ΔZLZ;L480P} were ~5 h and ~20 h respectively (**Figures 3.1I and 3.1J**). These results suggest that ZLZ regulates GL2 protein turnover irrespective of START domain mutation.

ZLZ is dispensable for nuclear localization but indispensable for activity of GL2

The precise mechanism regulating the nuclear import of HD-Zip IV TFs is currently unknown. However, previous studies reported the presence of a nuclear localization signal either at amino- or carboxyl-terminus of several HD-Zip I proteins (Arce et al., 2011; Cao et al., 2016; Shao et al., 2018). Using EYFP fusions to GL2, we previously showed that both the HD and START domain are dispensable for the nuclear localization of GL2 in developing seeds and trichomes (Schrack et al., 2014; Mukherjee et al., 2022). Here, we asked whether the ZLZ plays a role in regulating the subcellular localization of GL2 in addition to investigating the phenotypic outcome of deleting this domain from GL2. Despite the inability of both EYFP:gl2^{ΔZLZ} and FLAG:gl2^{ΔZLZ} to rescue null phenotype in leaves (**Figures 3.2A and 3.2D**), epifluorescence microscopy indicated that EYFP:gl2^{ΔZLZ} was nuclear localized in leaf trichomes similarly to the WT protein (**Figure 3.2C**). Moreover, the fluorescence signals in the leaf trichomes of *EYFP:gl2^{ΔZLZ}* plants were much stronger compared to those of WT plants (**Figure 3.2C**), consistent with a role for the ZLZ domain in protein turnover.

All three double mutants with mutations in both ZLZ and START domain (*gl2^{ΔZLZ;ΔST}*, *gl2^{ΔZLZ;W279R}* and *gl2^{ΔZLZ;L480P}*) also displayed trichome defects irrespective of the type of amino-terminal tag (**Figures 3.2A and 3.2D**) and screening of the corresponding EYFP-tagged T1 transformants revealed mostly undetectable fluorescence signals, although a few showed either medium or strong expression (**Figure 3.2B**). Western blot analysis of leaf samples from T1 progeny of *FLAG:gl2^{ΔZLZ;L480P}* mutant revealed signals for most of the plants, comparable to

those of *FLAG:gl2^{ΔZLZ}* and *FLAG:gl2^{L480}* single mutants (**Figure B.1**). The double mutant proteins still appeared to be nuclear localized; however, the fluorescence signals were very weak in the homozygous lines. To confirm the subcellular localization patterns of the *ΔZLZ* mutants, we observed the EYFP tagged proteins in the primary root of *Arabidopsis* seedlings using confocal laser scanning microscopy. The nuclear localization of GL2 was intact in both *EYFP:gl2^{ΔZLZ}* and *EYFP:gl2^{ΔZLZ;L480P}* mutants similar to *EYFP:GL2* or *EYFP:gl2^{L480P}* plants (**Figures 3.2E**) confirming that the ZLZ is dispensable for translocation of GL2 to the nucleus.

Role of UPS in the degradation of HD-Zip IV TFs is more evident under conditions affecting START domain lipid binding

Previously, use of the proteasome inhibitor MG132 in a CHX time-course experiment resulted in significantly increased accumulation of protein for the START domain mutant *gl2^{L480P}* that is predicted to be defective in lipid binding (Mukherjee et al., 2022). To determine the role of the UPS in the degradation of wild-type HD-Zip IV TFs, we conducted similar experiments using *Arabidopsis* lines expressing EYFP:GL2, FLAG:GL2 and HA:PDF2. For the latter two lines, we also used Concanamycin A (ConcA) (Kinashi et al., 1984) in addition to MG132 to determine the involvement of autophagy in the turnover of HD-Zip IV TFs. Autophagy is a proteasome-independent proteolytic system in which cytoplasmic components are engulfed into the vacuole or lysosome for degradation (Klionsky and Emr, 2000; Reggiori and Klionsky, 2002; Avin-Wittenberg et al., 2012). ConcA increases the internal vacuolar pH and thus inhibits vacuolar H⁺-ATPase (Dröse et al., 1993; Nishihara et al., 1995; Huss et al., 2002) resulting in an accumulation of autophagosomes with undegraded proteins in the vacuole (Yoshimoto et al., 2004; Xiong et al., 2007).

Presence of MG132 did not stabilize EYFP- or FLAG-tagged GL2 but rather caused the protein levels to drop slightly during the 24 h time course (**Figures 3.3A-3.3D**). In the case of HA:PDF2, the overall trend of protein levels was unaffected by MG132 treatment although quantification of the data showed significantly higher protein level in treated seedlings at the 2 h time point (**Figures 3.3E and 3.3F**). Similarly, neither FLAG:GL2 nor HA:PDF2 exhibited any elevated protein levels in ConcA treated seedlings (**Figures 3.3C-3.3F**). Rather, FLAG:GL2 levels dropped significantly but HA:PDF2 levels remained about the same in the presence of ConcA (**Figures 3.3C-3.3F**). These results indicate no clear involvement of proteolytic systems in the turnover of WT HD-Zip IV TFs in seedlings under normal growth conditions.

PDF2 was recently reported to bind lysophosphatidylcholines via its START domain consistent with the role of START domain as a lipid sensor (Wojciechowska et al., 2021). In addition, EYFP:GL2, HA:GL2 and HA:PDF2 exhibited reduced stability under phosphate (Pi) limiting conditions suggesting a link between Pi sensing and HD-Zip IV TF-mediated gene expression (Wojciechowska et al., 2021). Here, we performed additional cycloheximide assays using EYFP:PDF2 and EYFP:pdf2^{ΔST} seedlings to determine whether the START domain affects protein stability in the setting of Pi starvation. In line with previous results, quantification of blot images showed reduced half-life of EYFP:PDF2 under Pi limitation, although it remained stable throughout the 10 h time course in both conditions (**Figures 3.3G and 3.3H**). The stability of the EYFP:gl2^{L480P} mutant protein was also reported to be reduced by phosphate starvation and the presence of MG132 was shown to stabilize the protein (Wojciechowska et al., 2021). Here, we explored the response of EYFP:pdf2^{ΔST} mutant protein to phosphate starvation using similar experiments. As expected, we observed a drop in half-life of this protein from >10 h under Pi sufficiency to ~3.25 h under Pi limitation (**Figures 3.3I and 3.3J**). Consistent with the previous

reports (Wojciechowska et al., 2021; Mukherjee et al., 2022), the half-life of this protein was also lower compared to WT protein under Pi limitation conditions (**Figures B.2A and B.2B**). In addition, a degradation product of 90 kDa was also detected corresponding to EYFP:pdf2^{ΔST} (**Figure B.2C**). Treatment of seedlings with proteasome inhibitor MG132 increased the half-life of EYFP:pdf2^{ΔST} to ~8.3 h under Pi limitation (**Figures 3.3K and 3.3L**) supporting the previous observation that START domain mutants are degraded via the 26S proteasome (Wojciechowska et al., 2021; Mukherjee et al., 2022). These results emphasize the role of Pi limitation in destabilizing HD-Zip IV TFs with a contribution of START domain in maintaining protein levels.

GL2 is weakly modified by ubiquitination in young seedlings

Protein stability experiments carried out in this and previous (Wojciechowska et al., 2021; Mukherjee et al., 2022) studies show that HD-Zip IV protein levels are affected by mutations in key domains, by altered growth conditions, or by addition of a proteasome inhibitor. Moreover, our preliminary proteomics data from immunoisolation of EYFP-tagged GL2 indicates that several proteins annotated as ubiquitin or SUMO are candidates for interacting partners of GL2. These results suggest involvement of the 26S Ubiquitin Proteasome or other degradation pathway in the turnover of HD-Zip IV TFs.

To detect ubiquitination and/or SUMOylation of EYFP-tagged proteins, Arabidopsis seedlings were grown in normal 0.5X MS media as well as in phosphate sufficient (P+) vs phosphate limiting (P-) conditions and under 12 h light/12 h dark conditions, with the harvesting of seedlings at the start of day or night. EYFP:GL2 and EYFP:gl2^{ΔZLZ} proteins were immunoprecipitated from whole seedlings using GFP-trap magnetic agarose beads, which were analyzed by immunoblotting and probing with antibodies against ubiquitin (P4D1) or SUMO1/2

in addition to anti-GFP antibodies. Here, we used EYFP-only expressing plants as negative control. The anti-GFP antibody revealed bands of expected sizes for both *EYFP:GL2* and *EYFP:gl2^{AZLL}* lines (**Figures 3.4A, B.3A, B.3D and B.3G**). With anti-Ub (P4D1) antibody, faint smear ladders of sizes comparable to those of ubiquitinated proteins were detected in eluates corresponding to both *EYFP:GL2* and *EYFP:gl2^{AZLL}* lines grown under the day/night cycle (**Figures 3.4B**). However, no detectable smears were observed for normal or phosphate limiting conditions, irrespective of the stringency of washes done during elution step (**Figures B.3B, B.3E and B.3H**). The anti-SUMO antibody detected bands that corresponded to molecular weights of ~150 kDa and ~70 kDa which were stronger in the *EYFP:gl2^{AZLL}* mutant compared to WT plants for the lysate, eluate and non-bound fractions in all conditions (**Figures 3.4C, B.3C, B.3F and B.3I**).

Immunoprecipitation of 6x His 3xFLAG-labelled wild-type GL2 or mutant gl2 proteins was carried out using FLAG-trap (DYKDDDDK Fab-Trap Agarose). Here, we used EYFP non-expressing WT (*Col*) plants as well as *FLAG:gl2^{42KR}* mutant lacking lysines (See mutagenesis section, **Figure 3.6**) as negative controls. In addition to growth in normal 0.5X MS media or in presence of high salinity (100 or 150 mM NaCl), seedlings were exposed to oxidative stress (10 mM H₂O₂ for 4 h) or heat shock (37 °C for 30 mins) prior to harvesting. For all growth conditions, seedlings were treated with MG132 for 4 h prior to harvesting. Subsequent immunoblot analysis of immunoprecipitates with anti-FLAG antibody confirmed the immunoisolation of FLAG-tagged proteins (**Figures 3.4D and B.3L**). Use of P4D1 or FK1 anti-Ub antibody indicated the high molecular mass smears corresponding to both *FLAG:GL2* and *FLAG:gl2^{42KR}*, although weaker non-specific bands were also observed for *Col* seedlings (**Figures 3.4E and B.3M**). Importantly, no substantial differences were observed between the

signals corresponding to different growth conditions (**Figures 3.4E and B.3M**). Additionally, FK2 antibody yielded no detectable signals indicative of ubiquitinated proteins (**Figure B.3J**). With anti-SUMO1 and anti-SUMO3 antibodies, *FLAG:GL2* and *FLAG:gl2^{42KR}* samples exhibited two diffuse bands with molecular weights of ~150 kDa and ~70 kDa similar to those seen with EYFP-tagged proteins (**Figures 3.4F, B.3K and B.3N**). Probing with anti-SUMO1 antibody also resulted in additional distinct band of ~130 kDa for both *FLAG:GL2* and *FLAG:gl2^{42KR}* (**Figures 3.4F and B.3N**). Taken together, the results suggest that the GL2 protein is weakly modified with signal intensities varying with the growth conditions or the tags used to label the proteins.

GL2 contains multiple, conserved predicted Ubiquitination and SUMOylation sites

Ubiquitination occurs by reversible conjugation of a highly conserved, 76-amino acid and 8.5-kDa protein called ubiquitin to the target protein in a series of enzymatic reactions involving E1 (ubiquitin activating), E2 (ubiquitin conjugating) and E3 (ubiquitin ligating) enzymes (Kerscher et al., 2006; Preston and Brodsky, 2017). This involves the formation of an isopeptide bond between the C-terminal glycine of ubiquitin and the ϵ -amino group of lysine residues in the target protein (Preston and Brodsky, 2017; Romero-Barrios and Vert, 2018). Similarly, SUMOs are ~100 amino acid, ~12 kDa proteins that are also covalently attached to lysines in target proteins through an enzymatic cascade similar to that of ubiquitination (Kurepa et al., 2003; Johnson, 2004; Colby et al., 2006; Wilkinson and Henley, 2010). To predict the modification sites, we performed bioinformatics analysis of GL2 using various web-based predictor tools including UbPred (Radivojac et al., 2010), Abcepta SUMOplotTM (Abcepta, 2013), JASSA v4 (Beauclair et al., 2015) and GPS-SUMO 2.0 (Zhao et al., 2014) which revealed multiple potential lysine conjugation sites throughout its amino acid sequence (**Figures 3.5A and 3.5B**).

High confidence modification sites were found mainly within the 65-amino acid ZLZ (**Figures 3.5A and 3.5B**), consistent with the role of this domain in destabilizing GL2. The lysines 169 and 211 were predicted for both ubiquitination and SUMOylation (**Figures 3.5A and 3.5B**). In addition, the predicted sites appeared in clusters (e.g. K169, K174, K180, K188 and K209, K211, K216) (**Figures 3.5A and 3.5B**), consistent with the possibility of “ubiquitination hot spots” (Walton et al., 2016).

The consensus sequence associated with lysines undergoing ubiquitination is not yet known. However, SUMOylation was reported to occur on a consensus motif Ψ KxD/E, where Ψ is a hydrophobic residue such as A, I, L, M, P, F, V or W; K is the Lys where SUMO is linked; X is any amino acid residue and D or E is an acidic residue (Melchior, 2000; Rodriguez et al., 2001; Sampson et al., 2001; Wilkinson and Henley, 2010). Here, two major SUMOylation sites predicted at positions 169 and 211 of GL2 were located on LKAE motif that matches to the consensus motif Ψ KxD/E (**Figure 3.5B**). Previously, ubiquitination sites were found to be associated with higher degree of disorder and flexibility of the substrates (Prakash et al., 2004; Singh et al., 2006; Radivojac et al., 2007; Gsponer et al., 2008; Tompa et al., 2008; Tsvetkov et al., 2008). To determine the disorder status of GL2 with respect to its potential to undergo ubiquitination, we used PONDR[®] VSL2 Predictor (Obradovic et al., 2003; Peng et al., 2006). The results indicated that most of the highly predicted conjugation sites are located in intrinsically disordered regions of GL2 (**Figure 3.5C**).

To determine which predicted amino acids are conserved across plants species, amino acid sequences of related proteins were aligned. We found that most of the high confidence lysines are conserved as predicted ubiquitination sites in the majority of the GL2 orthologs from representative dicot and monocot plants (**Figure 3.5D**). The predicted sites were also conserved

among the 16 HD-Zip IV transcription factors from *Arabidopsis*, although the extent of conservation was lower compared to that among GL2 orthologs (**Figure 3.5E**).

Genetic analyses of the predicted PTM sites in GL2 to verify their functions in vivo

Substituting the lysines predicted to be involved in ubiquitination or SUMOylation with arginine is a commonly used genetic tool to understand their biological functions in a given protein. Such changes preserve the charge status of the polypeptide while removing the potential conjugation site. Accordingly, we first created single or multiple K-R mutants by mutating predicted lysines from the ZLZ (K169, K174, K180, K188, K209, K211, K216) and the HD (K129) (**Figures 3.6A and 3.6B**). One of these mutants contain K-R point mutations for all lysines in the ZLZ which is termed 7^{KR} (**Figure 3.6A**). However, we envisioned pitfalls of only mutating selected residues since consensus information is not available in predicting ubiquitination sites. Moreover, experimental data suggest that about 32-40% of known SUMOylation sites do not conform to the consensus motif ψ KxD/E (Matic et al., 2010; Hendriks et al., 2014; Zhao et al., 2014; Beauclair et al., 2015). Intriguingly, our pull-down experiments hinted towards ubiquitination occurring in the gl2^{ΔZLZ} mutant (**Figure 3.4B**). To account for roles of the remaining lysines in GL2, a gene synthesis approach was used to generate additional mutants that encompassed K-R changes for the whole GL2 sequence (42^{KR}) and its amino-terminus (K10, 14R), START domain (12^{KR}) and STAD domain (11^{KR}) (**Figures 3.6A and 3.6C**). To investigate the impact of the K-R substitutions on protein structure, we modeled the GL2 START domain using I-TASSER (Iterative Threading ASSEmbly Refinement) (Roy et al., 2010; Yang and Zhang, 2015a; Yang and Zhang, 2015b; Zhang et al., 2017) and did not detect changes in the overall 3-D structure of START domain upon introducing mutations (**Figure 3.6D**).

Similar to the HD-Zip II TFs, members of HD-Zip IV family contain a plant-specific leucine zipper dimerization domain truncated by a loop forming zipper-loop-zipper (ZLZ) (Schena and Davis, 1992; Di Cristina et al., 1996; Nakamura et al., 2006; Ciarelli et al., 2008). Here, we hypothesized that the conserved cysteines in the loop region are involved in the formation of disulfide linkage to stabilize the GL2 structure. To analyze the role of these residues, we substituted the cysteines (C) with alanines (A), generating three mutants: C193A (AXXC), C196A (CXXA) and C193, 196A (AXXA) (**Figure B.4**). These mutations were introduced into a binary vector and expressed under the GL2 native promoter with an enhanced yellow fluorescence protein (EYFP) or a 6x His 3x FLAG tag amino-terminally fused to GL2 (*proGL2:EYFP/FLAG:gl2*) (**Figure 3.6E**). Transgenic lines were generated by transforming *gl2-5* null mutant plants with the binary vector constructs using *Agrobacterium*-mediated floral dip (**Figure 3.6E**).

Mutation of putative modification sites cause phenotypic changes in Arabidopsis

The *GL2* gene is required for normal development of leaf trichomes, the root epidermis, and the seed coat, but it is not essential for viability, making it useful for genetic analyses. When we assayed ~2-weeks-old T1 transformants for functional rescue of the *gl2* mutant trichome phenotype, we observed a repertoire of rescue patterns ranging from complete rescue to no rescue corresponding to most of the mutations except 42KR that caused trichome branching defects in all plants (**Figure B.5A and 3.7G**). Similarly, we observed varying degrees of EYFP expression levels in all transformants including those with WT *GL2* reflecting the outcomes of random integration of transgenes into the genome (**Figure B.5B**). Further analysis of rescue and expression data showed a highly positive correlation of EYFP expression with rescue of

trichome phenotype (**Figure 3.7B**). This suggests that most of the K-R and C-A mutations do not cause visible defects in differentiation of the epidermis in the leaf.

Recently, we reported that stabilization of GL2 by amino-terminal fusion with EYFP results in a gain-of-function phenotype characterized abaxial (downward) leaf curling (Subedi and Schrick, 2022). Interestingly, most of the K-R and C-A mutants generated here also displayed a similar phenotype but at higher frequencies, indicating alterations in adaxial-abaxial polarity of leaves (**Figure B.5C**). Examination of leaf trichomes for the expression of transgenes revealed stronger fluorescent signals corresponding to curly leaves compared to flat leaves (**Figure 3.7D**), consistent with the previous result (Subedi and Schrick, 2022). This was also supported by a positive correlation between the leaf curling and EYFP expression (**Figure 3.7C**) which was obtained by statistical analysis of T1 transformant data. One exception was EYFP:gl2^{42KR}, which yielded only weak or undetectable signals (**Figure B.5B and 3.7D**). More careful observation of rosettes of EYFP:gl2^{42KR} plants also indicated that the appearance of leaf curling in this mutant is distinct from rest of the mutants or EYFP:GL2 plants (**Figure 3.7D and 3.7E**). Comparisons of selected EYFP lines with their FLAG-tagged counterparts demonstrated the absence of curly leaf phenotype in FLAG-tagged lines (**Figure 3.7A**) as reported previously (Subedi and Schrick, 2022).

Leaf architecture directly impacts various functions of leaves such as photosynthesis, respiration and transpiration (O'Toole and Cruz, 1980; Turner et al., 1986; Matschi et al., 2020). Prior reports have implicated leaf rolling in leaf erectness and enhanced light transmission (Govaerts et al., 1996; Moon and Hake, 2011; Saglam et al., 2014) in addition to regulation of water loss and canopy temperature by reducing the exposed leaf surface area (O'Toole et al.,

1979; Kadioglu and Terzi, 2007; Kadioglu et al., 2012)). To quantify the extent of leaf curl, leaf width was measured in natural and flattened state followed by calculation of the rolling leaf index (LRI) or degree of epinasty for 25-d old WT and mutant lines. The results indicated higher values of LRI in curly looking plants from each genotype (**Figure 3.7F**), confirming our phenotypic observations. Lower LRI values were obtained for all FLAG lines with the *gl2-5* null mutant displaying the least epinastic leaves (**Figure 3.7F**). Trichome quantification demonstrated no significant change in the total number of trichomes as well as the number of their branches on the first leaves of most of the transgenic lines with or without abaxially curled leaves (**Figure 3.7H**). However, 42KR mutants produced defective unbranched and two-branched trichomes in addition to normal trichomes with ≥ 3 branches (**Figure 3.7H**). Such trichome defects were seen in both EYFP and FLAG lines irrespective of leaf curl (**Figure 3.7H**), confirming distinct phenotypic defects associated with 42KR mutant lines.

K-R and C-A mutants retain nuclear localization of GL2 with altered protein stability

Previous studies using fluorescent protein tagging revealed that GL2 is a nuclear protein, and that it is strongly expressed during trichome cell differentiation (Schrack et al., 2014; Mukherjee et al., 2022). Using live imaging of trichomes, we observed nuclear localization of fluorescently labelled GL2 in K-R and C-A mutants similar to WT plants (**Figure 3.7D**). In addition, confocal laser scanning microscopy also confirmed the nuclear localization of GL2 in the primary root of *Arabidopsis* seedlings in selected mutants including EYFP:gl2^{7KR}, EYFP:gl2^{42KR} and EYFP:gl2^{C193; C196A} (**Figure 3.8A**). These results suggest that mutations of the putative PTM sites in GL2 do not alter its subcellular localization.

In contrast to other K-R mutants, the expression of EYFP:gl2^{42KR} was mostly weak or undetectable in corresponding T1 transformants or homozygous lines (**Figures B.5B and 3.7D**). To confirm the expression of the gl2^{42KR} protein, Western blot analysis of leaf samples from FLAG-tagged T1 transformants was carried out using anti-FLAG antibodies. Most of the plants with 42KR mutation gave signals in blot images (**Figure 3.8B**). However, additional bands of molecular weight comparable to that of GL2 dimer were also observed (**Figure 3.8B**), suggesting the stabilization of dimer, even under denaturing conditions. To further explore the effect of PTM related mutations in the stability of GL2, we carried cycloheximide experiments using two representative mutants *FLAG:gl2^{7KR}* and *FLAG:gl2^{C193A; C196A}*. Unexpectedly, quantification of protein levels over a time course of 24 h showed a drop in stability of gl2^{7KR} protein compared to WT protein with statistically significant differences at 5 and 10 h time points (**Figure 3.8C**). As shown in the graph, the half-life of FLAG:gl2^{7KR} was only ~2 h that was substantially lower compared to ~24 h for FLAG:GL2 (**Figure 3.8C**). The levels of FLAG:gl2^{C193A;C196A} also appeared slightly lower than those of FLAG:GL2 although the difference was not statistically significant (**Figure 3.8C**).

Discussion

Role of ZLZ in GL2 turnover in addition to dimerization

HD-Zip START TFs are highly conserved proteins occurring across plant species from freshwater green algae to agronomic crops (Romani et al., 2018). They are key regulators indispensable for growth and development of plants. Studies have shown vital roles of HD-Zip IV TFs in protection of plants from pathogens and abiotic stresses in addition to their involvement in developmental processes. However, how HD-Zip IV protein levels are tightly regulated during epidermal differentiation in plants has not been completely understood. Protein

degradation is one of the important cellular mechanisms in eukaryotes which governs protein levels in response to specific intracellular or environmental signals. In this study, we thoroughly characterized the relative roles of ZLZ and START domains in the degradation of HD-Zip IV TFs from Arabidopsis under various growth conditions. As cycloheximide treatment affects only translation, we were able to monitor the protein half-lives ruling out the contribution from transcription. While the previous study identified the role of START domain in stabilizing GL2 (Mukherjee et al., 2022), our current study unveiled a novel role of LZ dimerization domain in destabilizing GL2 (**Figure 3.1**). Deletion of the ZLZ domain was previously shown to enhance the accumulation and nuclear localization of ATML1 (Iida et al., 2019), consistent with our current finding that ZLZ domain destabilizes GL2. Compared to WT protein, START domain mutants exhibit dramatically lower protein half-lives suggesting a major mode of regulation of protein turnover by START domain (Mukherjee et al., 2022). In contrast, deletion of ZLZ did not alter the protein levels drastically although the Δ ZLZ protein levels were significantly higher compared to WT protein (**Figures 3.1A-3.1F**). This was further supported by our experiments with ZLZ and START double mutant where the drop in protein levels in $gl2^{L480P}$ mutant was only partially rescued by ZLZ deletion (**Figures 3.1G-3.1H**).

Although the addition of MG132 was previously shown to partially stabilize $gl2^{L480}$ mutant protein, we did not observe this effect in WT GL2 and PDF2 proteins (**Figures 3.3A-3.3F**). Rather, the proteins appear to be unstable in the presence of inhibitors of protein degradation although the differences were very small (**Figures 3.3A-3.3F**). It is likely that unstable proteins such as the one resulting from START domain mutation are noticeably degraded through cellular degradation machineries. Since WT GL2 is inherently more stable, we were possibly not able to discern the subtle difference caused by MG132. Recently, it was

proposed that PDF2 participates in a positive feedback loop that is activated by phospholipid binding to its START domain promoting de novo membrane synthesis required for plant growth (Schrack et al., 2014; Wojciechowska et al., 2021). Importantly, the function of PDF2 in phospholipid homeostasis and elongation growth was shown to be affected by Pi limitation (Wojciechowska et al., 2021). Under Pi starvation, overall phospholipid levels decrease in plants (Li et al., 2006). Here, we observed a Pi deficiency-induced reduction in half-life of PDF2 that was further reduced by START domain mutation (**Figures 3.3G-3.3J and B.2**), supporting the hypothesis that START-dependent mechanisms link Pi sensing with transcriptional control. Apart from this, the presence of salt in growth media also seems to have a major impact in the GL2 protein level in Arabidopsis (**Figures 3.1D and 3.1E**). In future studies, it will be important to carry out more experiments involving salinity, drought, heat, oxidative stress and other abiotic stresses to better understand the effects of stress factors in HD-Zip IV protein turnover.

Challenges associated with detection of GL2 PTMs in vivo

Analysis of the immunisolated samples with anti-Ub antibody revealed high molecular smear ladders typical of ubiquitinated proteins in the eluate of EYFP-tagged proteins under certain conditions (**Figures 3.4B, 3.4E and B.3M**). Interestingly, FLAG:gl2^{42KR} mutant also yielded high molecular weight smear with anti-Ub antibodies similar to FLAG:GL2 protein (**Figures 3.4E and B.3M**), indicating possible non-canonical ubiquitination. Although ubiquitination most commonly occurs in lysine residues, studies have shown that ubiquitin can also get attached to the free N-terminal α -amino group (Scaglione et al., 2013; Tatham et al., 2013; Vittal et al., 2015) or the side chains of internal cysteine, serine and threonine residues in some proteins (Iwai and Tokunaga, 2009; Shimizu et al., 2010; Okumoto et al., 2011). Another possibility is modification of lysines present in the FLAG tag itself, adding another layer of

complexity in the detection of PTMs. Similarly, positive results seen with Col samples may be because of very high concentration of proteins in lysates causing non-specific binding of proteins to beads.

In case of SUMOylation, at least two bands of ~150 and ~70 kDa were observed in all the samples including lysates and non-bound fractions (**Figures 3.4 and B.3**). It is possible that the ~150 kDa band represents the sumoylated form of either GL2 or one of its interacting partners. In many instances, only a small proportion (less than 1 %) of target proteins are in modified state as ubiquitination and SUMOylation are transient (Johnson, 2004; Park et al., 2011). Ub and SUMO molecules are rapidly cleaved from the target proteins by deubiquitinases and SUMO proteases respectively (Li and Hochstrasser, 1999; Barroso-Gomila et al., 2021). We attempted to overcome this problem by using the 26S proteasome inhibitor MG132 (Lee and Goldberg, 1998) and N-ethylmaleimide (NEM) or iodoacetamide (IAA), the inhibitors of deubiquitinases and de-sumoylation (Dahlmann et al., 1985). However, MG132 does not inhibit the deubiquitinase activity of 19S lid of proteasome and NEM or IAA also inhibits other enzymes with active site cysteines such as E1 and E2 required for modification.

SUMOylation is also strongly induced by several stress factors including heat shock, cold, drought, high salt, oxidants and ethanol (Kurepa et al., 2003; Saracco et al., 2007; Tempé et al., 2008; Golebiowski et al., 2009; Miller and Vierstra, 2011b; Seifert et al., 2015; Augustine et al., 2016; Augustine and Vierstra, 2018). Among them, heat shock is most efficient as illustrated by reversible upregulation of SUMO conjugation within minutes of mild heat treatment at 37°C with the concomitant depletion of the free SUMO pool (Kurepa et al., 2003; Saracco et al., 2007; Miller and Vierstra, 2011a; Miller et al., 2013; Rytz et al., 2018). However, our results indicate no effect of stress conditions on the extent of SUMOylation of WT GL2 or its mutants (**Figures**

3.4E, B.3K and B.3N). In case of EYFP-tagged proteins, the light/dark cycles appear to have an effect on protein PTM (**Figures 3.4B**), although whether there is involvement of circadian clock in this process remains to be further investigated. Whether our protein of interest is ubiquitinated or SUMOylated may be further confirmed by performing protein-protein interaction assays and *in vitro* ubiquitination or SUMOylation experiments. An alternative approach for identification of modifications and/or modification sites in GL2 would be mass spectrometry-based Ub remnant profiling where the Ub attachment sites are revealed by di-glycine remnants (Ub footprints) resulting from the trypsin digest of immunoprecipitated proteins (Okada et al., 2009; Martins et al., 2015).

Insights from bioinformatic analysis of PTM sites

The prediction of both ubiquitination and SUMOylation in GL2 (**Figures 3.5A and 3.5B**) implies a possible cross talk between the two systems. SUMO conjugation can inhibit ubiquitination by sequestering the lysine residues in the target proteins (Desterro et al., 1998; Taupitz et al., 2017). It was also found that Ub can be directly attached to the SUMO moieties of the SUMOylated proteins during stress conditions (Uzunova et al., 2007; Miller et al., 2010; Miller and Vierstra, 2011a; Barroso-Gomila et al., 2021). In addition, SUMO conjugation can contribute to proteasome-mediated degradation of the modified protein by recruiting E3 ubiquitin ligases called SUMO-targeted Ubiquitin ligase (STUbLs). Intriguingly, phosphorylation can mediate the antagonism or co-operation between the ubiquitination and SUMOylation, adding an additional layer of complexity (Desterro et al., 1998; Filtz et al., 2014). Although our bioinformatic analyses primarily focused on the ubiquitination and SUMOylation, the lysine residues in a protein also serve as targets for other PTMs such as acetylation,

formylation, methylation and biotinylation (Brooks and Gu, 2003; Freiman and Tjian, 2003) which need to be addressed in future studies.

Furthermore, the presence of potential ubiquitination and SUMOylation sites in the intrinsically disordered regions of GL2 (**Figures 3.5C**) conforms to the previous findings that these modifications require significant flexibility without a fixed tertiary structure in the target protein (Prakash et al., 2004; Tompa et al., 2008; Tsvetkov et al., 2008; Wright and Dyson, 2015). Since the E3 ligases have relatively rigid structures with large cavities and gaps (Gieffers et al., 2001; Zheng et al., 2002; Hao et al., 2005; Hao et al., 2007), they are likely to favor the substrates with unstructured regions. Although PONDR® algorithm predicted HD to be disordered (**Figures 3.5C**), this domain is likely to undergo disorder-to-order transitions upon DNA binding as shown for other protein regions involved in protein-protein or protein-nucleic acid interactions (Dunker et al., 2001; Dunker et al., 2002; Dyson and Wright, 2002; Vucetic et al., 2003; van der Lee et al., 2014).

Association between GL2 levels and abaxial leaf curling

To examine the biological role of predicted modification sites, we generated several transgenic plants expressing mutant gl2 protein with various combinations of K-R substitutions throughout its amino acid sequence. Here, our intention was to replace lysines with the positively charged but nonubiquitinatable arginine residue. To our surprise, the abaxially curled leaves appeared in a subset of almost all PTM mutants generated using the EYFP tag. Strikingly, the frequency of this phenotype was higher in majority of the PTM mutants compared to the EYFP:GL2 plants (**Figures B.5C and 3.7A**). However, we did not see a phenotype specific to particular lysines except the unique form of leaf curling in *EYFP:gl2^{42KR}* lines (**Figures 3.7D and 3.7E**) and the substantial trichome branching defects in all *gl2^{42KR}* lines regardless of the

nature of fusion tags (**Figures B.5A, 3.7G and 3.7H**). Previously, the protein levels of EYFP:GL2 were found to be significantly higher in curly leaves versus flat leaves (Subedi and Schrick, 2022). Thus, our current results might indicate functional redundancies of several lysine residues in the degradation of GL2 protein although it is less likely that all 42 lysines in GL2 have roles in its PTM. Since GL2 also plays critical developmental roles in other tissues such as root hairs and seed, it will be helpful to determine whether a particular PTM related mutation can result in the phenotypic change(s) in these tissues.

Besides elimination of putative ubiquitination sites, other factors may also contribute to the leaf curling observed in this study. One hypothesis is that leaf curling is specific to EYFP component of the transgene (Subedi and Schrick, 2022); however, we never saw such phenotypes in *gl2-5* lines or the plants expressing EYFP-fused non-functional proteins such as *gl2^{ΔZLZ}*, *gl2^{L480P}* or other START domain mutants (**Figures 3.2**). Relatively higher epinasty of leaves in WT Col or FLAG:GL2 lines compared to *gl2-5* lines (**Figure 3.7F**;(Subedi and Schrick, 2022)) also support the idea that a functional GL2 protein is a prerequisite for the appearance of curly leaves. Another possible reason for leaf curl is stabilization of GL2 stemming from protein geometry changes in R variants. Unlike lysine that shows unidirectional interaction, the guanidinium group in arginine permits interactions in three possible directions resulting in greater extent of electrostatic interactions such as salt bridges and H-bonds (Borders Jr. et al., 1994; Donald et al., 2011; Platts and Falconer, 2015). As a result, mutation of lysines to arginine can lead to increased protein stability but unfavorable protein folding (Turunen et al., 2002; Sokalingam et al., 2013). In *gl2^{42KR}* mutant, such effect is likely to accumulate to a level high enough to stabilize dimers even under denaturing conditions (**Figure 3.8B**) or to cause trichome branching defects by impairing protein folding (**Figures B.5A, 3.7G and 3.7H**).

Nevertheless, our results correlate the expression levels of GL2 with appearance of abaxial leaf curling in Arabidopsis (**Figures 3.7C and 3.7D**), suggesting a novel role of GL2 in leaf polarity establishment. Leaf polarity along the adaxial/abaxial (upper/lower) axis is primarily maintained by polarized expression of adaxial-determining class III homeodomain leucine zipper (HD-Zip III) genes *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*) and *REVOLUTA* (*REV*) (Emery et al., 2003; Prigge et al., 2005; Rong et al., 2019) which in turn are targets of several microRNAs (Juarez et al., 2004; Song et al., 2012; Merelo et al., 2016). The HD-Zip III genes are also downregulated by the expression of *KANADI* (Kerstetter et al., 2001; Emery et al., 2003; Zhang et al., 2009), *YABBY* (Siegfried et al., 1999; Bowman, 2000; Kumaran et al., 2002; Yang et al., 2019) and auxin-responsive genes (Pekker et al., 2005; Kelley et al., 2012; Takahashi et al., 2013), which promote abaxial fate. It is possible that alteration in GL2 protein levels may affect one of these pathways or phospholipid mediated osmoregulation or cuticle development resulting in leaf curling phenotype.

Unlike *gl2^{ΔZLZ}* mutants, *gl2^{7KR}* mutant displayed decrease in protein half-life compared to WT (**Figure 3.8C**). This is in line with a slight decrease in the stability of FLAG:GL2 when the seedlings were treated with the proteasome inhibitor (**Figure 3.3C and 3.3D**). Both observations do not match to our initial hypothesis that elimination of either ubiquitination or degradation of GL2 causes elevated protein levels. We reasoned that the cause of this discrepancy could be the use of young seedlings in which the signs of altered leaf polarity are not yet established. In addition, we used the whole seedlings grown on normal media for these experiments, but the protein turnover may differ depending upon whether the tissues are taken from leaves, roots or seed coats and whether there is any change in the nutrient composition of the growth media. In our previous study, the level of EYFP:GL2 was dramatically higher in curly leaves compared to

normal leaves when protein samples were used from ~25 days old plants (Subedi and Schrick, 2022).

Additional findings from mutational analysis

Our microscopy results demonstrated no change in nuclear localization of GL2 in K-R and C-A mutants (**Figures 3.7D and 3.8A**) along with the phenotypic complementation of null phenotype in most of the PTM mutants (**Figures B.5A, 3.7B, 3.7G and 3.7H**). One exception was 42KR mutant which showed partial trichome branching defects (**Figures B.5A, 3.7G and 3.7H**). In our previous studies, HD and START domain mutants also displayed defects in epidermal differentiation despite nuclear localization of proteins (Mukherjee et al., 2022). The underlying causes of the phenotypic defects in these mutants were loss of DNA binding ability of GL2 in HD mutants and impaired dimerization and protein stability of GL2 in START domain mutants (Mukherjee et al., 2022). In this work, we also reported super strong nuclear expression (Comelli and Gonzalez, 2007; Dietz, 2013; Mukherjee, 2021) of ZLZ mutant protein although the corresponding mutants exhibited null phenotypes possibly resulting from defective protein dimerization and DNA binding.

Mutation of two conserved cysteine residues predicted to form disulfide linkage in the ZLZ of GL2 did not result in any phenotypic changes other than leaf curl phenotype in EYFP-tagged lines (**Figures B.5C, 3.7A, 3.7D-F**). Similar cysteines were previously implicated in redox sensing by several proteins (Akamatsu et al., 1997; Tron et al., 2002; Couturier et al., 2013; Skryhan et al., 2015). Importantly, changes in redox conditions directly influence the development of root and root hair development in plants ((De Tullio et al., 2010); for review see

(Considine and Foyer, 2014)). It will be interesting to carry out further experiments using redox agents such as reduced glutathione to determine the possible redox-dependent regulation of GL2 dimerization and DNA binding in the growth of root hairs.

Proposed model

In our model, mutation of START domain disrupts ligand binding by GL2 resulting in loss of dimerization (**Figure 3.9**). The resulting mutant still binds to DNA as monomer but fails to orchestrate proper epidermal differentiation (**Figure 3.9**). Similarly, we proposed that removal of the ZLZ stabilizes the GL2 but the resulting protein is defective in dimerization and DNA binding leading to phenotypic defect in plants (**Figure 3.9**). In contrast, the K-R mutants are able to regulate epidermal differentiation as they are proficient in ligand binding, dimerization as well as DNA binding (**Figure 3.9**). Additionally, their protein levels are elevated resulting in increased abaxial leaf curling that can be considered as a gain-of-function phenotype (**Figure 3.9**).

Conclusion and implications

In this study, we uncovered post-translational regulatory aspects that govern developmentally important functions of HD-Zip IV transcription factors. Our results support the hypothesis that post-translational regulatory mechanisms involving ubiquitin and/or SUMO contribute to GL2 levels/activity in plants. Follow-up studies to verify these results will include identification of various components of HD-Zip IV degradation pathways such as E3 ubiquitin ligases and characterization of underlying protein-protein interactions. The identification of protein interactors will be needed to form a mechanistic model for how HD-Zip transcription factors undergo post-translational modifications leading to changes in subcellular localization,

protein stability, and ultimately their activity. In many cases, phosphorylation regulates the other PTMs such SUMOylation (Colignon et al., 2019). Thus, it will be also interesting to explore possible cross-talks between multiple PTM in the regulation of HD-Zip IV TFs. Additionally, future works should compare the transcriptional levels of GL2 or its targets among WT and various mutant lines displaying different protein half-lives or phenotypes. In summary, our findings lead to an improved understanding of how regulatory proteins such as HD-Zip IV transcription factors are themselves regulated post-translationally. Outcomes of this work will provide mechanistic information on how HD-Zip IV transcription factors may be engineered for enhanced stress resistance and productivity of agriculturally important plants.

Figures

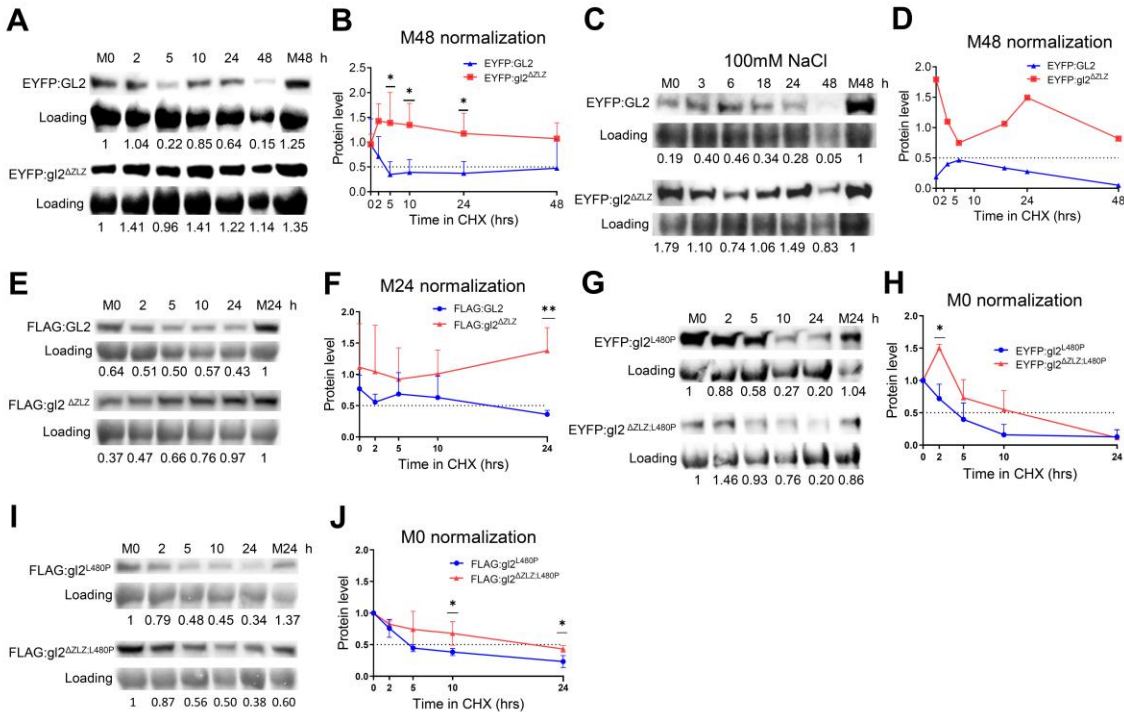


Figure 3.1. ZLZ negatively affects the stability of GL2.

(A-J) Representative blot images and corresponding data quantification from CHX experiments for EYFP:GL2 vs EYFP:gl2^{ΔZLZ} (A, B), EYFP:GL2 vs EYFP:gl2^{ΔZLZ} with 100 mM NaCl (C, D), FLAG:GL2 vs FLAG:gl2^{ΔZLZ} (E, F), EYFP:gl2^{L480P} vs EYFP:gl2^{ΔZLZ;L480P} (G, H) and FLAG:gl2^{L480P} vs FLAG:gl2^{ΔZLZ;L480P} (I, J) using antibodies directed against corresponding tags. Protein half-lives were measured using 4-5-day-old seedlings grown in 0.5x MS media. M0, M24 and M48, Mock (DMSO) treatments for 0, 24 and 48 h; Loading, Coomassie Blue stained total protein bands. Graph shown in (D) corresponds to a single biological replicate shown in (C). For other experiments, graphs represent n=3 biological replicates where the values are expressed as mean ± SD. The significance of the differences between protein levels at each time point was tested using an unpaired *t*-test (two-tailed) with unequal variance (**p* < 0.05, ***p* < 0.005).

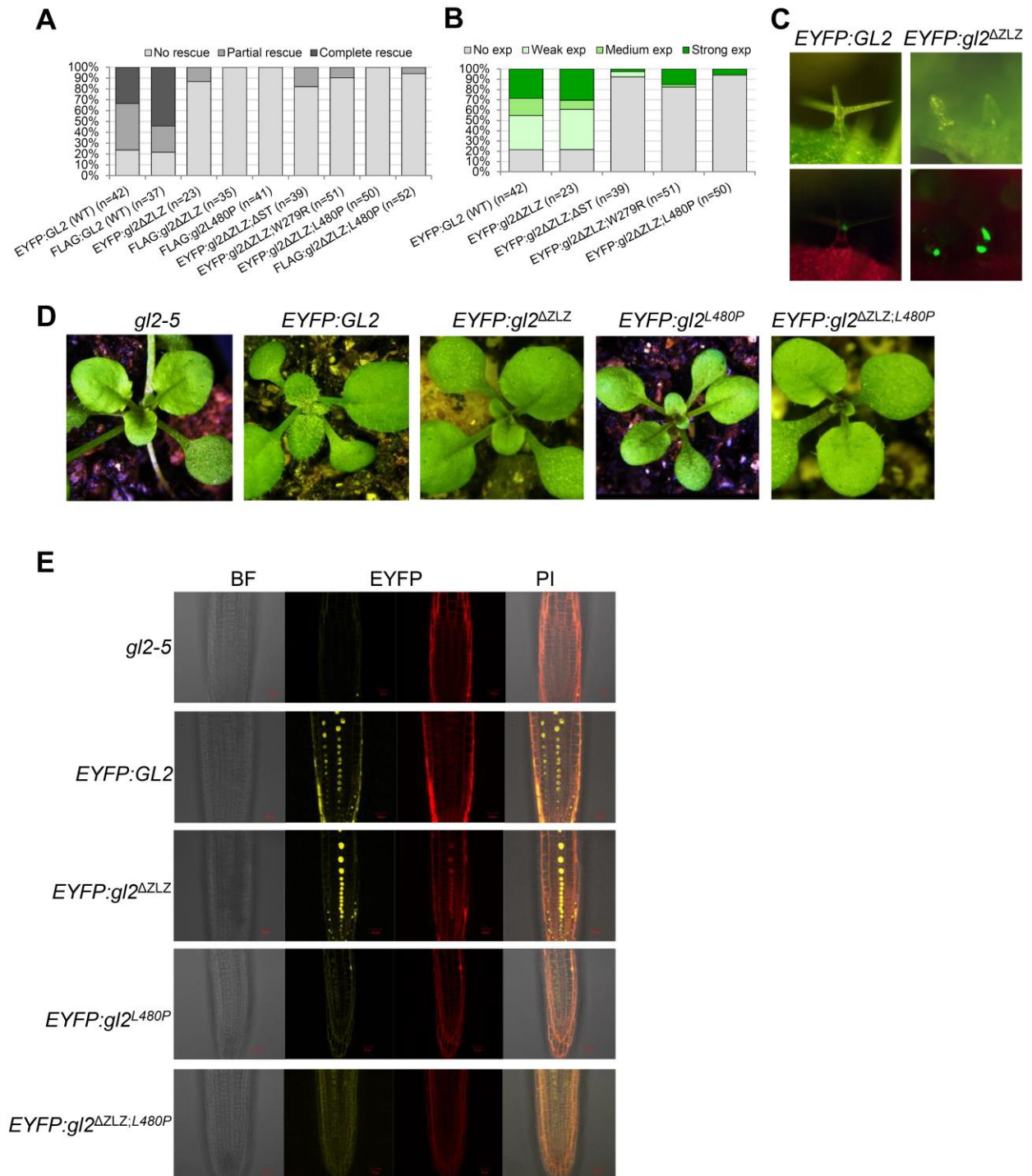


Figure 3.2. GL2 is nuclear localized despite phenotypic defect in ZLZ deletion mutants.

(A-B) Graphs showing complementation of trichome phenotype (A) and EYFP expression (B) in T1 progeny of the indicated genotypes. The transgenes were expressed in the

gl2-5 null mutant background with N-terminally fused EYFP or 6x His 3x FLAG tags under the native *GL2* promoter. **(C)** Light images (upper) and matching fluorescence images (lower) of leaf trichomes showing the expression of EYFP:GL2 and EYFP:gl2^{ΔZLZ} in 15-d-old homozygous plants. Size bars = 100 μm. **(D)** Rosettes from 15-d-old *gl2-5*, *EYFP:GL2*, *EYFP:gl2^{ΔZLZ}*, *EYFP:gl2^{L480}* and *EYFP:gl2^{ΔZLZ;L480P}* plants. Size bar = 2 mm. **(E)** Representative confocal images confirming the nuclear localization of EYFP:gl2^{ΔZLZ}, EYFP:gl2^{L480P} and EYFP:gl2^{ΔZLZ;L480P} in root tips of Arabidopsis seedlings stained with propidium iodide. Right panel: Magnified images. Size bar = 20 μm.

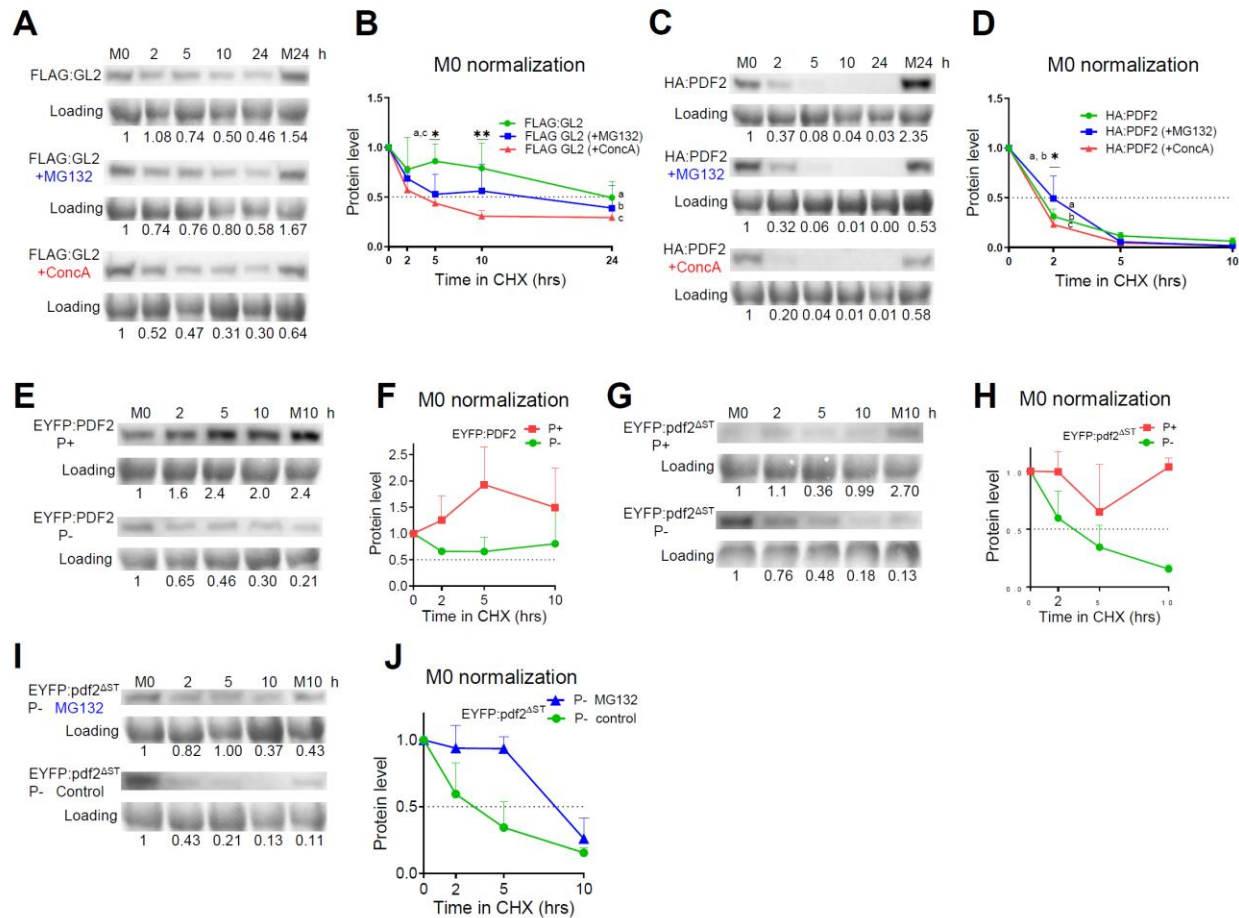


Figure 3.3. Degradation of HD-Zip IV TFs is favored by conditions impairing lipid binding.

(A-F) Representative blot images and corresponding data quantification from CHX experiments with or without MG132 (50 μ M)/ConcA (1 μ M) treatment for EYFP:GL2 (A, B), FLAG:GL2 (C, D) and HA-PDF2 (E, F) using antibodies directed against corresponding tags. Protein half-lives were measured using 4-5-day-old seedlings grown in 0.5x MS media. M0 and M24, Mock (DMSO) treatments for 0 and 24 h; Loading, Coomassie Blue stained total protein bands. Graphs represent $n=3$ biological replicates where the values are expressed as mean $\pm$ SD. The significance of the differences between protein levels at each time point was tested by applying an unpaired t -test (two-tailed) with unequal variance or an ordinary two-way ANOVA

using Tukey's multiple comparisons test (* $p < 0.05$, ** $p < 0.005$). **(G-J)** Representative blot images and corresponding data quantification from CHX experiments for EYFP:PDF2 **(G, H)** and EYFP:pdf2^{ΔST} **(I, J)** using seedlings grown either on Pi sufficiency (P+) or limitation (P-) for 4-5 days. **(K, L)** CHX experiment results showing reversal of the instability of EYFP:pdf2^{ΔST} upon MG132 (50 μM) treatment for seedlings grown on Pi limitation (P-). M0 and M10, Mock (DMSO) treatments for 0 and 10 h; Loading, Coomassie Blue stained total protein bands. Graphs represent n=2 biological replicates.

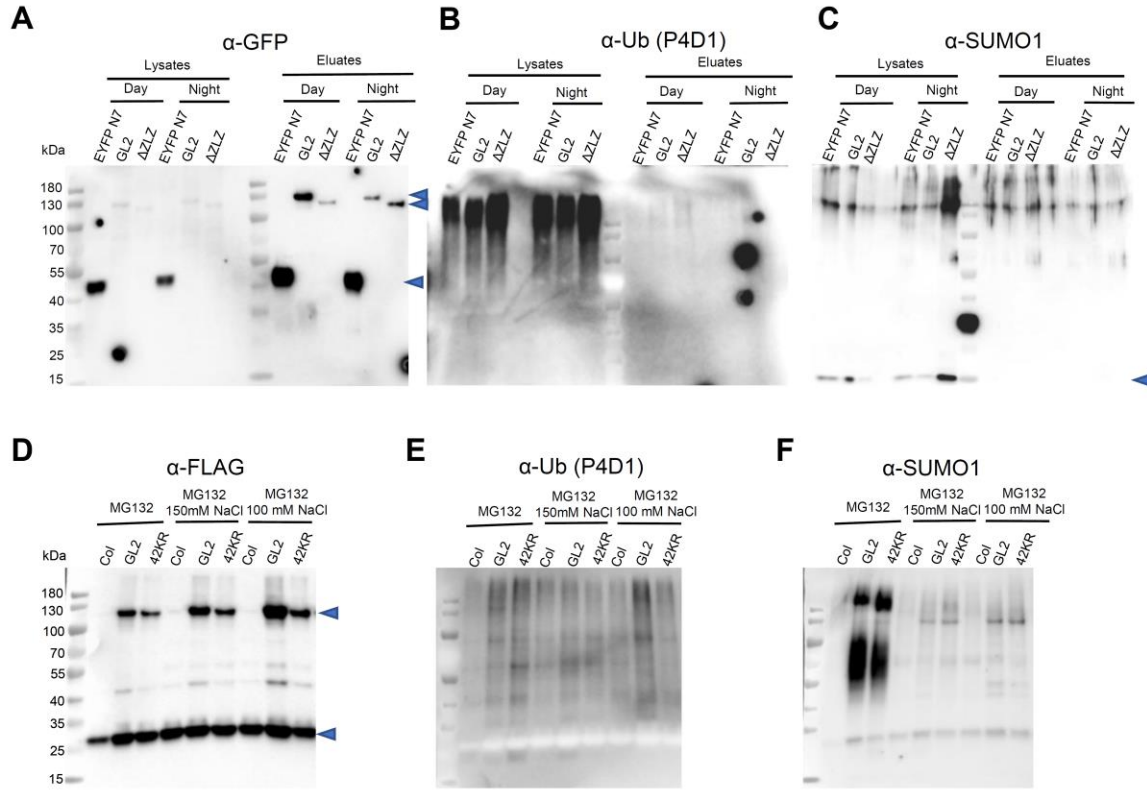


Figure 3.4. Detection of ubiquitination and SUMOylation in epitope tagged wild-type and mutant *gl2* proteins.

(A-C) Immunoblotting of *EYFP N7* (negative control), *EYFP:GL2* and *EYFP:gl2^{ΔZLZ}* samples immunoprecipitated with GFP-trap magnetic agarose beads using anti-GFP (A), P4D1 anti-Ub (B) and anti-SUMO1 (C) antibodies. (D-F) Immunoblotting of *Col* (WT), *FLAG:GL2* and *FLAG:gl2^{42KR}* samples immunoprecipitated with FLAG-trap using anti-FLAG (D), P4D1 anti-Ub (E) and anti-SUMO1 (F) antibodies. Conditions used to grow the seedlings are specified. Day and night indicate the samples corresponding to seedlings harvested at the middle of day and night respectively. For *EYFP N7* eluates, 1:20 dilution was used. Bands representing immunoprecipitated proteins, Fab-fragment of the DYKDDDDK Fab-Trap and free Ub or SUMO molecules at the bottom of corresponding blots are indicated. Molecular weights (kDa): GL2, 83.2 (747 aa); *gl2^{ΔZLZ}*, 76 (65 aa); EYFP, 30 (238 aa); 6xHis 3x FLAG, 5.8 (46 aa); Ubiquitin, 8.6 (76 aa); SUMO1 11 (100 aa); Fab, 25.

A

GL2

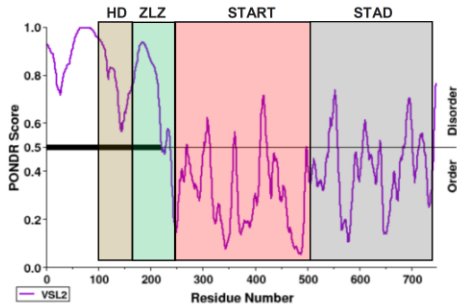
MSMAVDMSSKQPTKDFSSPALSLSLAGIFRNASSGSTNPEEDFLGRRVVD
 DEDRTVMSSSENSGPTRSRSEEDLEGEDHDEEEEDGAAGNKGTKNRKR
 KKYHRHTDQIRHMEALFKETPHPDEKQRQQLSKQLGLAPRQVKFWFQNR
 TQIKAIQERHENSILKAELEKLEENKAMRESFSKANSSCPNCGGPDLLH
 LENSILKAELEKLEENKAMRESFSKANSSCPNCGGPDLLH
 RIAEISNRATLELQKMATSGEPMWLRSVETGREILNYDEYLKEFPQAQASS
 FPGRKTIEASRDAGIVFMDAHKLAQSFMDVGQWKETFACLISKAATVDVIR
 QGEGPSRIDGAIQLMFGEMQLTPVVPTRVYFVRSCRQLSPEKWAIVDVS
 VSVEDSNTEKEASLLKCRKLPSCGCIIEDTSNGHSKVTWVEHLDSASTVQP
 LFRSLVNTGLAFGARHWVATLQLHCERLVFFMATNVPKDSLGVTTLAGRK
 SVLKMAQRMQTQSFYRAIAASSYHQWTKITTKTGQDMRVSSRKNLHDPGEPT
 GVIVCASSSLWLPVSPALLDFDFRDEARRHEWDALSNGAHVQSIANLSKGQ
 DRGNSVAIQTVKSREKSWVLQDSSTNSYESVVYAPVDINTQLVLAGHD
 PSNIQILPSPGFSIIPDGVESRPLVITSTQDDRSQGGSLTLALQTLINPS
 PAAKLNMEVESVNTLVSVTLHNIRKSLQIEDC

Ubiquitination **SUMOylation**
 High confidence Medium confidence Low confidence Motifs with high probability Motifs with low probability

B

Residue	Ubiquitination confidence score	SUMOylation	
		Consensus motif (ψ Kx/D/E)	Confidence score
K129	0.88		
K188	0.88		
K169	0.87	HENSLKAELEKLR	0.91
K180	0.84		
K209	0.83		
K174	0.82		
K211	0.82	LENSLKAELEKLR	0.91
K14	0.81		
K10	0.78		
K216	0.78		
K121	0.77		
K418	0.72		
K718	0.72		
K498	0.71		
K311	0.67		
K552	0.64		
K297		NYDEYLKEFPQAQA	0.56
K427		SLLKCRKLPSCGCI	0.34

C



D

	HD			ZLZ								
	127D	167L	177E	187S	210L							
<i>Z.mays</i> /1-655	SFF	ECPPHDD	QQRK	SQRAENE	LRAENMR	YKEALS	---SASC	ENARL	REEIDR	ISA		
<i>O.sativa</i> /1-779	ALF	ESPHPDER	QRQ	SLL	SELE	LQDEHR	AMRELAK	---PSRC	EKA	L	AEIER	LRG
<i>B.napus</i> /1-750	ALF	ETPHPDE	QRQ	SLL	AELE	LREEN	AMRESFS	ANSSSC	ENT	L	AELD	LRA
A.thaliana /1-747	ALF	ETPHPDE	QRQ	SLL	AELE	LREEN	AMRESFS	A-NSSC	ENS	L	AELD	LRA
<i>C.sativa</i> /1-744	ALF	ETPHPDE	QRQ	SLL	AELE	LREEN	AMRESFS	AANSSC	ENS	L	AELD	LRA
<i>C.sativus</i> /1-781	ALF	ESPHPDE	QRQ	TLL	AEME	LREEN	AMREISK	---KIGC	KNA	L	AEVE	LRA
<i>H.annuus</i> /1-749	ALF	ESPHPDE	QRQ	SLL	SELD	LGEEN	LLRETIK	---GTC	ENA	L	RAEIE	LRT
<i>P.dactylifera</i> /1-744	ALF	ESPHPDE	QRQ	SLL	SEID	LRDENR	AMRQMIN	---ACC	ENARL	AEIE	LRA	
<i>N.tabacum</i> /1-767	ALF	ESPHPDE	QRQ	SLL	AEIE	LREEN	GLRETLK	---NPSC	ENARL	MQVE	LRA	
<i>S.tuberosum</i> /1-772	ALF	ESPHPDE	QRQ	SLL	AEIE	LREEN	GLRGTSK	---NPSC	ENARL	RAEVE	LRA	
<i>A.hypogaea</i> /1-823	ALF	ESPHPDE	QRQ	SLL	TEIE	LKEKN	GLRETIN	---ACC	ENA	L	AEIE	LRG
<i>G.hirsutum</i> /1-753	ALF	ESPHPDE	QRQ	SLL	QELD	LRDEN	AMRETIN	---ACC	ENA	L	AEVE	LRT
<i>M.domestica</i> /1-761	ALF	ESPHPDE	QRQ	SLL	GEME	LRDES	AMREQIN	---ACC	ENARL	SEVE	LRA	
<i>G.max</i> /1-751	ALF	ESPHPDE	QRQ	SLL	SEIE	LKEKN	TLRETIN	---ACC	ENA	L	AEVE	LRA
<i>J.regia</i> /1-754	ALF	ESPHPDE	QRQ	SLL	SEMD	LRDEN	AMREQIN	---ACC	ENARL	AEVE	LRA	
<i>H.brasiliensis</i> /1-759	ALF	ESPHPDE	QRQ	SLL	TEME	LREEN	AMRETIN	---ACC	ENA	L	AEVE	LRA
	K121	K129	K169	K174	K180	K188	K209	K211	K216			

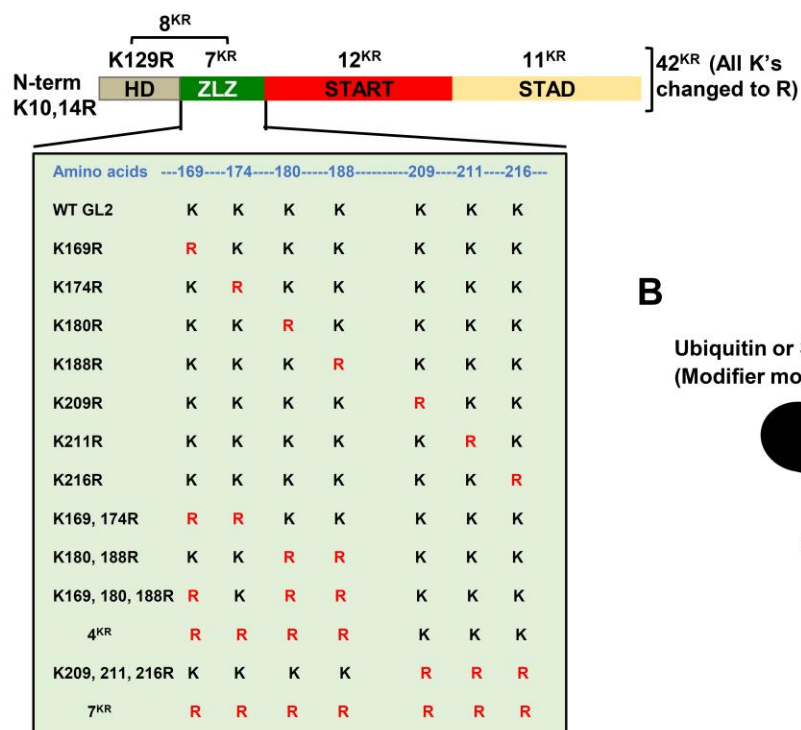
E

	HD			ZLZ								
	126P	166S	176R	186F	206E	216K						
HDG8/1-699	AYF	ECPPHDERQRN	VLLRG	ENETLQSDNE	AMLDAK	---SVL	ENARL	DHRDR	ISN			
HDG9/1-718	TYF	ECPPHDEFQRR	AALRA	ENI	IRRENES	MEDALN	---NVV	QNAYL	DEYERVS			
HDG10/1-708	AFH	ECPPHDDSQRR	IALRV	NM	IRC	NEAMEKALE	---TVL	KNVIL	TEYERLSS			
HDG4/1-709	ALF	ENAHPT	TRL	AKL	AE	NETLKT	ESQNIQSNFQ	---CLF	ENARL	RQELDR		
HDG5/1-826	ALF	ENPHPD	QRK	VMLRA	ENDNLK	SENCH	LQAELE	---CLS	ENCR	LR		
GL2/1-747	ALF	ETPHPDE	QRQ	SLL	AELE	LREEN	AMRESFS	ANSS	ENS	L	AELD	LRA
FWA/1-686	NFY	MENPHPT	EEQRY	VTLRE	HDRL	LATQD	QLRSAML	---RSL	ENAN	LERE	IDQFNS	
HDG7/1-682	SFF	ECPPHNE	QRL	VIL	QENE	LRL	ENSLFKESMR	---GSL	ENA	L	EELDR	ICA
ANL2/1-802	SMF	ECPPHDE	QRL	ALLRQ	END	LRA	ENMSIREAMR	---NPI	ENARL	DELDRVCN		
HDG1/1-808	SVF	ECAPHDE	QRL	ALLRQ	END	LRA	ENMSVREAMR	---NPM	ENSRL	DELDRVCA		
HDG11/1-722	SSF	ECPPHDE	QRN	SAL	AEND	IRCEN	IAIREALK	---HAI	ENAH	LR	EELERMST	
HDG12/1-687	STF	NECHPDE	QRN	CAL	LEND	IRCEN	IAIREAIK	---HAI	ENAH	LR	DELERVSS	
HDG3/1-725	AFRE	CHPPD	QRY	SEL	RNLNNH	LRSENOR	LREAIH	---QAL	LNAR	L	TEEIKQLSV	
HDG2/1-721	AF	ECPPHDD	QRK	SHLRA	ENE	LRNDN	LYREALA	---NAS	ENAR	L	REEIDRISA	
PDF2/1-743	SFF	ECPPHDD	QRK	QIL	SDND	LRA	ENNNRYKEALS	---NAT	ENAR	L	REEIDRISA	
ATML1/1-762	SFF	ECPPHDD	QRK	QIL	SEND	LRA	ENNNRYKDALS	---NAT	ENAR	L	REEIDRISA	
	K121	K129	K169	K174	K180	K188	K209	K211	K216			

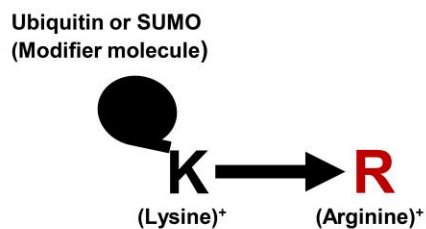
Figure 3.5. GL2 contains multiple predicted lysine conjugation sites that are conserved in related proteins.

(A) 747-amino acid sequence of GL2 from *Arabidopsis thaliana* showing the ubiquitination and SUMOylation sites predicted using UbPred (Radivojac et al., 2010) and Abcepta SUMOplotTM (Abcepta, 2013) respectively. Confidence for ubiquitination sites: red, high; blue, medium; green, low. Probability of SUMOylation motifs: yellow, high; grey, low. **(B)** The prediction scores of modification sites shown in **(A)** and matching of amino acids between the predicted SUMOylation sites and SUMO consensus sequence (ψ KxD/E, where ψ is a hydrophobic residue, K is the lysine conjugated to SUMO, x is any amino acid and D or E is an acidic residue). **(C)** Graph showing intrinsic disorder of GL2 as calculated by the VSL2 algorithm of PONDR® prediction (Obradovic et al., 2003; Peng et al., 2006). The PONDR® output combined with PTM prediction reveal the presence of potential ubiquitination sites in intrinsically disordered regions of GL2. **(D)** Amino acid alignment of GL2 from *Arabidopsis thaliana* with orthologous proteins from *Zea mays*, *Oryza sativa*, *Brassica napus*, *Camelina sativa*, *Cucumis sativus*, *Helianthus annuus*, *Phoenix dactylifera*, *Nicotiana tabacum*, *Solanum tuberosum*, *Arachis hypogaea*, *Gossypium hirsutum*, *Malus domestica*, *Glycine max*, *Juglans regia* and *Hevea brasiliensis*. The conserved lysines at various positions are highlighted in blue. **(E)** Alignment of all 16 members of HD-Zip IV subfamily from *Arabidopsis thaliana* showing the conservation of predicted PTM sites. The conserved lysines at various positions are highlighted in blue.

A



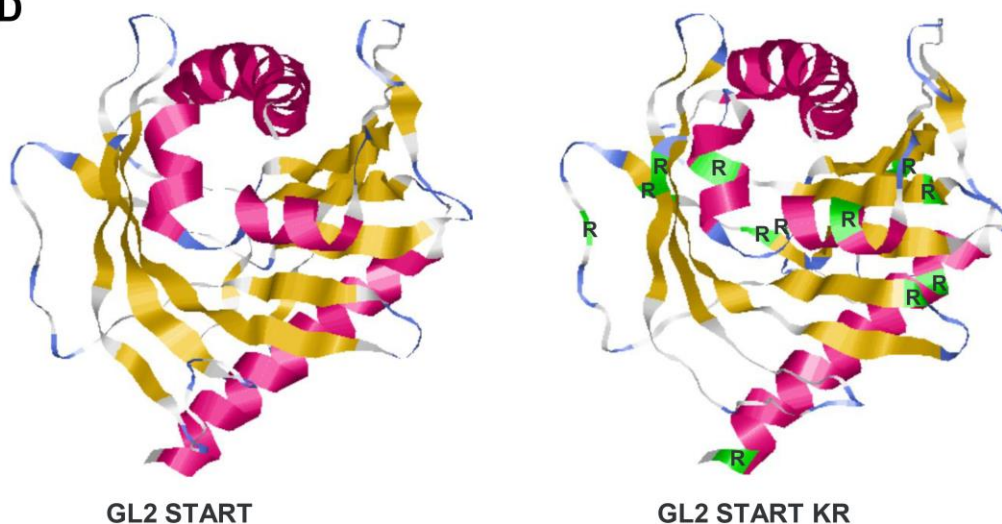
B



C

Sequences	Length (bp)	Lys/Arg codons	%GC	% Lysine (K)	% Arginine (R)
GL2	2247	AAA, AAG	52.60%	5.60%	6.40%
<i>gl2</i> ^{42KR}	2247	AGA, AGG	54.40%	0.00%	12.00%

D



E



Figure 3.6. Experimental evaluation of potential PTM sites in GL2.

(A) and **(B)** Schematic representation of strategy used for genetic analysis of modification sites. Single or multiple lysines from major domains of GL2 were mutated to arginines (shown in red) abolishing the possibility of ubiquitination but retaining the positive charge of amino acids. **(C)** Sequence information of WT *GL2* vs *gl2*^{42KR}. Modified sequence was synthetically generated using the indicated codon changes. **(D)** Structural homology models of WT START domain and its mutant counterpart containing 12 K-R substitutions (shown in green) generated in I-TASSER. The integrity of overall tertiary structure including the lipid binding cleft is unaffected in K-R mutant. **(E)** Strategy used to generate transgenic lines. Mutations were first transferred into a binary vector using EYFP or 6x His 3x FLAG tags and were subsequently introduced into *gl2-5* null mutant plants using *Agrobacterium*-mediated transformation.

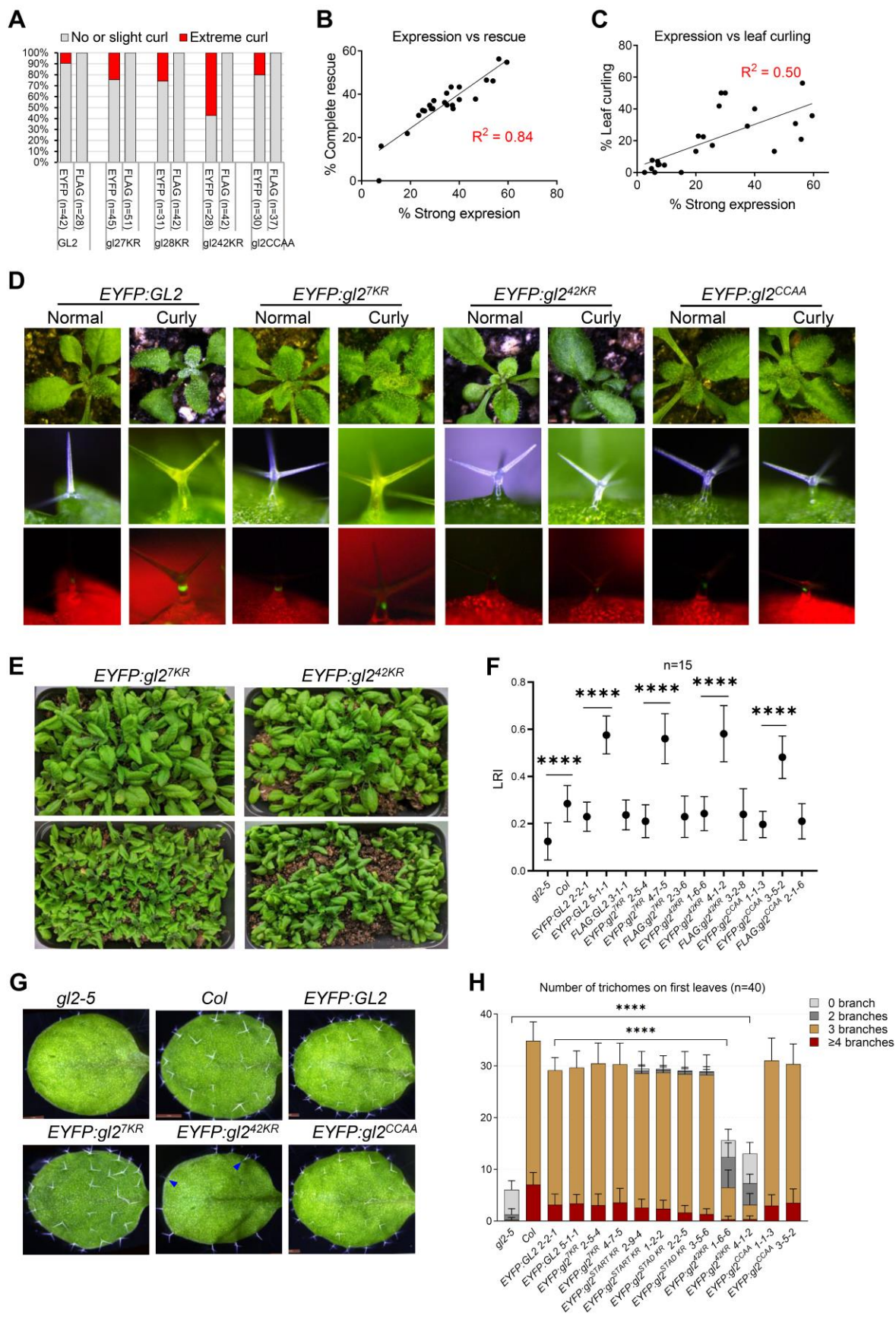


Figure 3.7. . Mutation of the predicted PTM sites in GL2 leads to increased frequency of abaxially curled leaves in *Arabidopsis*.

(A) Graph showing the occurrence of abaxially curled leaves in T1 progeny of EYFP vs FLAG lines. (B-C) Graphs depicting positive correlation of EYFP expression with rescue of trichome phenotype (B) and abaxial leaf curling (C) in available lines. (D) Rosettes of 15-d-old *EYFP:GL2*, *EYFP:gl2^{C193A;C196A}*, *EYFP:gl2^{7KR}* and *EYFP:gl2^{42KR}* plants corresponding to two types of leaves (Size bar = 2 mm) and EYFP expression in their trichomes (Size bar = 500 μ m). (E) 25-d-old *EYFP:gl2^{7KR}* and *EYFP:gl2^{42KR}* plants with two types of leaves (Size bar = 2 cm). (F) Leaf Rolling Index (LRI) of the 6th leaves from 25-d-old plants of various genotypes bearing normal and curly leaves. (G) First leaves of 10-d-old homozygous plants expressing *EYFP:GL2*, *EYFP:gl2^{C193A;C196A}*, *EYFP:gl2^{7KR}* and *EYFP:gl2^{42KR}* in comparison to *Col* wild type (WT) and *gl2-5* null mutant plants. Arrows indicate trichome defects in *EYFP:gl2^{42KR}*. Size bar = 2 mm. (H) Number of trichomes and their branches on the first leaves of 10-d-old plants. Means are shown for n= 15 (leaf biometry) or 20 (trichome quantification) plants, and error bars signify standard deviations (SDs). Significant differences between genotypes were determined using ordinary one-way ANOVA with Tukey's multiple comparisons test (*p < 0.05, ****p < 0.0001, ****p < 0.0001).

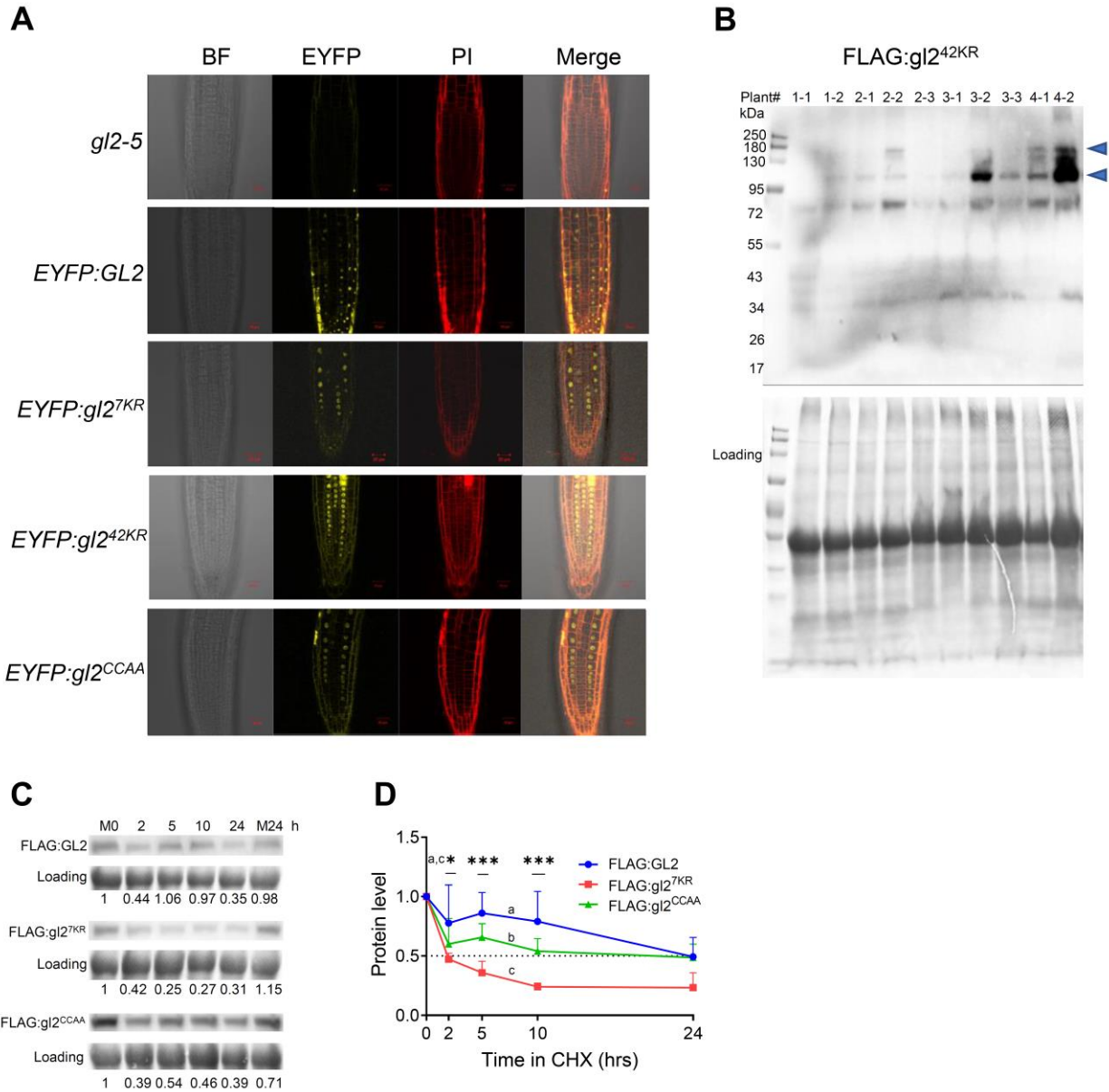


Figure 3.8. K-R and C-A mutants retain nuclear localization of GL2 with altered protein stability.

(A) Representative confocal images confirming the nuclear localization of EYFP:gl2^{7KR}, EYFP:gl2^{42KR} and EYFP:gl2^{C193A; C196A} GL2 in root tips of Arabidopsis seedlings stained with propidium iodide. Right panel: Magnified images. Size bar = 20 μ m. (B) Western blot analysis of leaves from T1 progeny of *FLAG:gl2^{42KR}* lines confirming the protein expression. The indicated lower and upper bands represent GL2 monomer and dimer respectively. Grayscale

image of Coomassie blue staining after Western blotting indicates protein loading (bottom). **(C)** Representative blot images and corresponding quantification data from CHX experiments for FLAG:GL2, FLAG:gl2^{7KR} and FLAG:gl2^{C193; C196A} over a 24 h time course using 4-5-day-old seedlings. M0 and M24, Mock (DMSO) treatments for 0 and 24 h; Loading, Coomassie Blue stained total protein bands. Graphs represent n=3 biological replicates where the values are expressed as mean $\pm$ SD. The significance of the differences between WT and mutant protein levels at each time point was tested using an unpaired *t*-test (two-tailed) with unequal variance (**p* < 0.05, ***p* < 0.005).

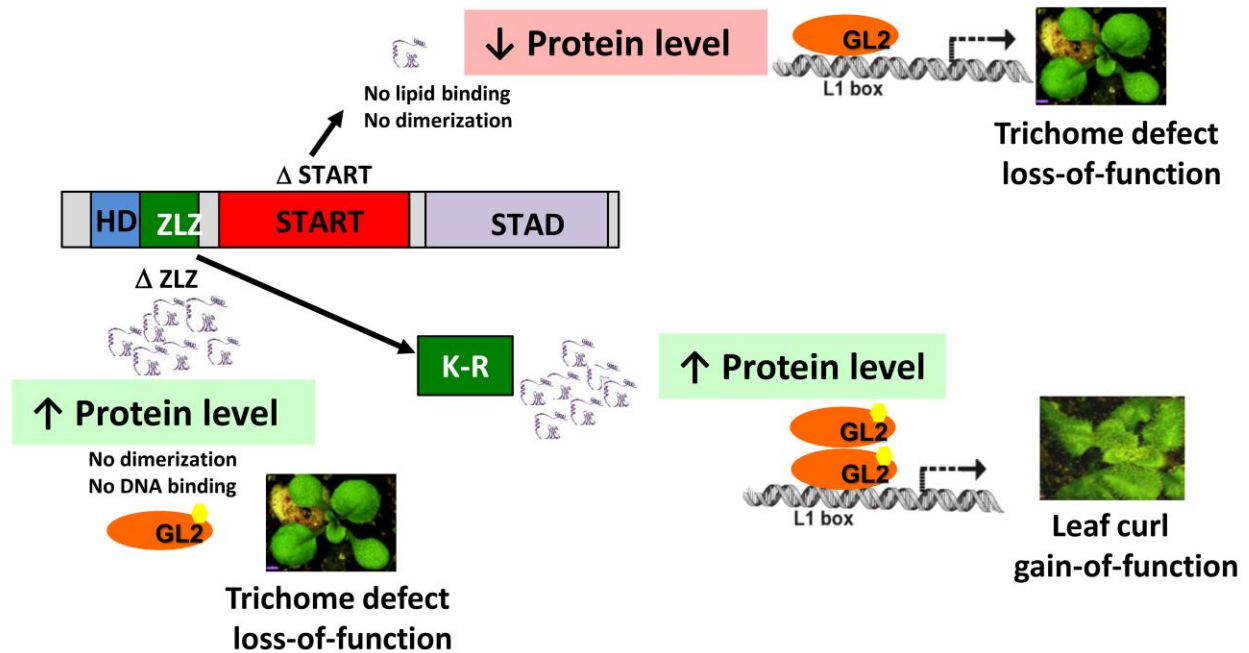


Figure 3.9. Model showing the effects of mutations in START domain, ZLZ domain and PTM sites of GL2.

Deletion of the START renders GL2 unable to bind lipids, negatively affecting its dimerization and stability. Binding of unstable and monomeric START mutants to the promoters of target genes does not suffice to regulate gene expression. In case of Δ ZLZ mutant, the protein fails to dimerize and can not bind to target DNA sequence despite its increased accumulation, resulting in phenotypic defects similar to those of START domain mutants. Contrary to this, K-R mutants are still functional in terms of their ability to dimerize and bind DNA in response to ligand binding. In addition, their overactivation resulting from elevated protein levels give rise to gain-of-function phenotype characterized by abaxially curled leaves. Yellow hexagon represents the lipid ligand bound to the START domain of GL2.

References

- Abcepta.** (2013). SUMOplot™ Analysis Program for prediction and scoring of SUMOylation sites in protein.
- 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. *The 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.
- Accossato, S., Kessler, F., and Shanmugabalaji, V.** (2020). SUMOylation contributes to proteostasis of the chloroplast protein import receptor TOC159 during early development. *eLife* **9**, e60968.
- Akamatsu, Y., Ohno, T., Hirota, K., Kagoshima, H., Yodoi, J., and Shigesada, K.** (1997). Redox regulation of the DNA binding activity in transcription factor PEBP2: The roles of two conserved cysteine residues. *J Biol Chem* **272**, 14497-14500.
- Alpy, F., and Tomasetto, C.** (2005). Give lipids a START: the StAR-related lipid transfer (START) domain in mammals. *J Cell Sci* **118**, 2791-2801.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.** (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.
- 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.
- Augustine, R.C., and Vierstra, R.D.** (2018). SUMOylation: re-wiring the plant nucleus during stress and development. *Curr Opin Plant Biol* **45**, 143-154.
- Augustine, R.C., York, S.L., Rytz, T.C., and Vierstra, R.D.** (2016). Defining the SUMO System in Maize: SUMOylation Is Up-Regulated during Endosperm Development and Rapidly Induced by Stress *Plant Physiol* **171**, 2191-2210.
- Avin-Wittenberg, T., Honig, A., and Galili, G.** (2012). Variations on a theme: plant autophagy in comparison to yeast and mammals. *Protoplasma* **249**, 285-299.
- Banavath, J.N., Chakradhar, T., Pandit, V., Konduru, S., Guduru, K.K., Akila, C.S., Podha, S., and Puli, C.O.R.** (2018). Stress inducible overexpression of AtHDG11 leads to improved drought and salt stress tolerance in peanut (*Arachis hypogaea* L.). *Front Chem* **6**, 34.

- Barberon, M., Zelazny, E., Robert, S., Conejero, G., Curie, C., Friml, J., and Vert, G.** (2011). Monoubiquitin-dependent endocytosis of the iron-regulated transporter 1 (IRT1) transporter controls iron uptake in plants. *PNAS USA* **108**, E450-458.
- Barroso-Gomila, O., Trulsson, F., Muratore, V., Canosa, I., Merino-Cacho, L., Cortazar, A.R., Pérez, C., Azkargorta, M., Iloro, I., Carracedo, A., Aransay, A.M., Elortza, F., Mayor, U., Vertegaal, A.C.O., Barrio, R., and Sutherland, J.D.** (2021). Identification of proximal SUMO-dependent interactors using SUMO-ID. *Nat Comm* **12**, 6671.
- Beauchair, G., Bridier-Nahmias, A., Zagury, J.-F., Saïb, A., and Zamborlini, A.** (2015). JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. *Bioinformatics* **31**, 3483-3491.
- Benlloch, R., and Lois, L.M.** (2018). Sumoylation in plants: mechanistic insights and its role in drought stress. *J Exp Bot* **69**, 4539-4554.
- Bernstein, H.J.** (2009). RasMol Molecular Graphics Visualisation Tool.
- Bernula, P., Pettkó-Szandtner, A., Hajdu, A., Kozma-Bognár, L., Josse, E.-M., Ádám, É., Nagy, F., and Viczián, A.** (2021). SUMOylation of PHYTOCHROME INTERACTING FACTOR 3 promotes photomorphogenesis in *Arabidopsis thaliana*. *New Phytologist* **229**, 2050-2061.
- Borders Jr., C.L., Broadwater, J.A., Bekeny, P.A., Salmon, J.E., Lee, A.S., Eldridge, A.M., and Pett, V.B.** (1994). A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Science* **3**, 541-548.
- Bowman, J.L.** (2000). The YABBY gene family and abaxial cell fate. *Curr Opin Plant Biol* **3**, 17-22.
- Brooks, C.L., and Gu, W.** (2003). Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Current Opinion in Cell Biology* **15**, 164-171.
- Buchanan, B.W., Lloyd, M.E., Engle, S.M., and Rubenstein, E.M.** (2016). Cycloheximide Chase Analysis of Protein Degradation in *Saccharomyces cerevisiae*. *J of visualized experiments : JoVE*, 53975.
- Burglin, T.R., and Affolter, M.** (2016). Homeodomain proteins: an update. *Chromosoma* **125**, 497-521.
- 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.
- Cao, Y.J., Wei, Q., Liao, Y., Song, H.L., Li, X., Xiang, C.B., and Kuai, B.K.** (2009). Ectopic overexpression of AtHDG11 in tall fescue resulted in enhanced tolerance to drought and salt stress. *Plant Cell Rep* **28**, 579-588.

- Chan, R.L., Gago, G.M., Palena, C.M., and Gonzalez, D.H.** (1998). Homeoboxes in plant development. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1442**, 1-19.
- Chen, E., Zhang, X., Yang, Z., Wang, X., Yang, Z., Zhang, C., Wu, Z., Kong, D., Liu, Z., Zhao, G., Butt, H.I., Zhang, X., and Li, F.** (2017). Genome-wide analysis of the HD-ZIP IV transcription factor family in *Gossypium arboreum* and GaHMG11 involved in osmotic tolerance in transgenic Arabidopsis. *Mol Genet Genomics* **292**, 593-609.
- Ciarbelli, A.R., Cioffi, A., Salvucci, S., Ruzza, V., Possenti, M., Carabelli, M., Fruscalzo, A., Sessa, G., Morelli, G., and Ruberti, I.** (2008). The Arabidopsis Homeodomain-leucine Zipper II gene family: diversity and redundancy. *Plant Mol Biol* **68**, 465-478.
- Clark, L., Sue-Ob, K., Mukkavar, V., Jones, Andrew R., and Sadanandom, A.** (2022). Understanding SUMO-mediated adaptive responses in plants to improve crop productivity. *Essays in Biochemistry* **66**, 155-168.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium* - mediated transformation of *Arabidopsis thaliana*. *The Plant J* **16**, 735-743.
- Colby, T., Matthäi, A., Boeckelmann, A., and Stuible, H.P.** (2006). SUMO-conjugating and SUMO-deconjugating enzymes from Arabidopsis. *Plant Physiol* **142**, 318-332.
- Colignon, B., Delaive, E., Dieu, M., Demazy, C., Muhovski, Y., Antoine, A., Raes, M., and Mauro, S.** (2019). Dual coordination of the SUMOylation and phosphorylation pathways during the response to heat stress in *Solanum tuberosum*. *Environmental and Experimental Botany* **162**, 192-200.
- Comelli, R.N., and Gonzalez, D.H.** (2007). Conserved homeodomain cysteines confer redox sensitivity and influence the DNA binding properties of plant class III HD-Zip proteins. *Archives of Biochem and Biophys* **467**, 41-47.
- Considine, M.J., and Foyer, C.H.** (2014). Redox regulation of plant development. *Antioxid Redox Signal* **21**, 1305-1326.
- Couturier, J., Chibani, K., Jacquot, J.-P., and Rouhier, N.** (2013). Cysteine-based redox regulation and signaling in plants. *Frontiers in Plant Science* **4**.
- Dahlmann, B., Kuehn, L., Rutschmann, M., and Reinauer, H.** (1985). Purification and characterization of a multicatalytic high-molecular-mass proteinase from rat skeletal muscle. *Biochem J* **228**, 161-170.
- De Tullio, M.C., Jiang, K., and Feldman, L.J.** (2010). Redox regulation of root apical meristem organization: connecting root development to its environment. *Plant Physiol Biochem* **48**, 328-336.
- Desterro, J.M., Rodriguez, M.S., and Hay, R.T.** (1998). SUMO-1 modification of I κ B α inhibits NF- κ B activation. *Mol Cell* **2**, 233-239.

- 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. *The Plant J* **10**, 393-402.
- Dietz, K.-J.** (2013). Redox regulation of transcription factors in plant stress acclimation and development. *Antioxidants & redox signaling* **21**.
- Donald, J.E., Kulp, D.W., and DeGrado, W.F.** (2011). Salt bridges: Geometrically specific, designable interactions. *Proteins: Structure, Function, and Bioinformatics* **79**, 898-915.
- Dröse, S., Bindseil, K.U., Bowman, E.J., Siebers, A., Zeeck, A., and Altendorf, K.** (1993). Inhibitory effect of modified bafilomycins and concanamycins on P- and V-type adenosinetriphosphatases. *Biochemistry* **32**, 3902-3906.
- Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradović, Z.** (2002). Intrinsic Disorder and Protein Function. *Biochemistry* **41**, 6573-6582.
- Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., and Obradovic, Z.** (2001). Intrinsically disordered protein. *J Mol Graph Model* **19**, 26-59.
- Dyson, H.J., and Wright, P.E.** (2002). Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**, 54-60.
- Elrouby, N., Bonequi, M.V., Porri, A., and Coupland, G.** (2013). Identification of Arabidopsis SUMO-interacting proteins that regulate chromatin activity and developmental transitions. *PNAS USA* **110**, 19956-19961.
- Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L.** (2003). Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol* **13**, 1768-1774.
- Fang, J., Guo, T., Xie, Z., Chun, Y., Zhao, J., Peng, L., Zafar, S.A., Yuan, S., Xiao, L., and Li, X.** (2021). The URL1-ROC5-TPL2 transcriptional repressor complex represses the ACL1 gene to modulate leaf rolling in rice. *Plant physiol* **185**, 1722-1744.
- Farmer, E., Mousavi, S., and Lenglet, A.** (2013). Leaf numbering for experiments on long distance signalling in Arabidopsis. *Protocol Exchange*.
- Fierro, A.C., Leroux, O., De Coninck, B., Cammue, B.P.A., Marchal, K., Prinsen, E., Van Der Straeten, D., and Vandenbussche, F.** (2015). Ultraviolet-B radiation stimulates downward leaf curling in *Arabidopsis thaliana*. *Plant Physiol and Biochem* **93**, 9-17.
- Filtz, T.M., Vogel, W.K., and Leid, M.** (2014). Regulation of transcription factor activity by interconnected post-translational modifications. *Trends in Pharmacological Sciences* **35**, 76-85.

- Finley, D.** (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem* **78**, 477-513.
- Freiman, R.N., and Tjian, R.** (2003). Regulating the regulators: lysine modifications make their mark. *Cell* **112**, 11-17.
- Fujimuro, M., Sawada, H., and Yokosawa, H.** (1994). Production and characterization of monoclonal antibodies specific to multi-ubiquitin chains of polyubiquitinated proteins. *FEBS Lett* **349**, 173-180.
- Gao, Y., Gao, S., Xiong, C., Yu, G., Chang, J., Ye, Z., and Yang, C.** (2015). Comprehensive analysis and expression profile of the homeodomain leucine zipper IV transcription factor family in tomato. *Plant Physiol and Biochem* **96**, 141-153.
- 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 in Genetics* **6**, 323-329.
- Gieffers, C., Dube, P., Harris, J.R., Stark, H., and Peters, J.M.** (2001). Three-dimensional structure of the anaphase-promoting complex. *Mol Cell* **7**, 907-913.
- Golebiowski, F., Matic, I., Tatham, M.H., Cole, C., Yin, Y., Nakamura, A., Cox, J., Barton, G.J., Mann, M., and Hay, R.T.** (2009). System-wide changes to SUMO modifications in response to heat shock. *Sci Signal* **2**, ra24.
- Gou, J.Y., Li, K., Wu, K., Wang, X., Lin, H., Cantu, D., Uauy, C., Dobon-Alonso, A., Midorikawa, T., Inoue, K., Sánchez, J., Fu, D., Blechl, A., Wallington, E., Fahima, T., Meeta, M., Epstein, L., and Dubcovsky, J.** (2015). Wheat Stripe Rust Resistance Protein WKS1 Reduces the Ability of the Thylakoid-Associated Ascorbate Peroxidase to Detoxify Reactive Oxygen Species. *The Plant Cell* **27**, 1755-1770.
- Govaerts, Y.M., Jacquemoud, S., Verstraete, M.M., and Ustin, S.L.** (1996). Three-dimensional radiation transfer modeling in a dicotyledon leaf. *Appl Opt* **35**, 6585-6598.
- Gsponer, J., Futschik, M.E., Teichmann, S.A., and Babu, M.M.** (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* **322**, 1365-1368.
- Guan, X.Y., Li, Q.J., Shan, C.M., Wang, S., Mao, Y.B., Wang, L.J., and Chen, X.Y.** (2008). The HD-Zip IV gene *GaHOX1* from cotton is a functional homologue of the Arabidopsis *GLABRA2*. *Physiol Plant* **134**, 174-182.
- Guo, H., and Yin, Y.** (2019). Measuring Protein Half-life in *Arabidopsis thaliana*. *Bio Protoc* **9**, e3318.

- Guo, X.-Y., Wang, Y., Zhao, P.-X., Xu, P., Yu, G.-H., Zhang, L.-Y., Xiong, Y., and Xiang, C.-B.** (2019). AtEDT1/HDG11 regulates stomatal density and water-use efficiency via ERECTA and E2Fa. *New Phytologist* **223**, 1478-1488.
- Han, D., Lai, J., and Yang, C.** (2021). SUMOylation: A critical transcription modulator in plant cells. *Plant Science* **310**, 110987.
- Hao, B., Oehlmann, S., Sowa, M.E., Harper, J.W., and Pavletich, N.P.** (2007). Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol Cell* **26**, 131-143.
- Hao, B., Zheng, N., Schulman, B.A., Wu, G., Miller, J.J., Pagano, M., and Pavletich, N.P.** (2005). Structural Basis of the Cks1-Dependent Recognition of p27Kip1 by the SCFSkp2 Ubiquitin Ligase. *Molecular Cell* **20**, 9-19.
- Hendriks, I.A., D'Souza, R.C., Yang, B., Verlaan-de Vries, M., Mann, M., and Vertegaal, A.C.** (2014). Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* **21**, 927-936.
- 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.
- Huss, M., Ingenhorst, G., König, S., Gassel, M., Dröse, S., Zeeck, A., Altendorf, K., and Wieczorek, H.** (2002). Concanamycin A, the specific inhibitor of V-ATPases, binds to the V(o) subunit c. *J Biol Chem* **277**, 40544-40548.
- Iida, H., Yoshida, A., and Takada, S.** (2019). ATML1 activity is restricted to the outermost cells of the embryo through post-transcriptional repressions. *Development* **146**.
- Ingram, G.C., Magnard, J.-L., Vergne, P., Dumas, C., and Rogowsky, P.M.** (1999). ZmOCL1, an HDGL2 family homeobox gene, is expressed in the outer cell layer throughout maize development. *Plant Mol Biol* **40**, 343-354.
- Ito, M., Sentoku, N., Nishimura, A., Hong, S.-K., Sato, Y., and Matsuoka, M.** (2003). Roles of rice GL2-type homeobox genes in epidermis differentiation. *Breeding Science* **53**, 245-253.
- Iwai, K., and Tokunaga, F.** (2009). Linear polyubiquitination: a new regulator of NF-kappaB activation. *EMBO Rep* **10**, 706-713.
- Javelle, M., Vernoud, V., Depège-Fargeix, N., Arnould, C., Oursel, D., Domergue, F., Sarda, X., and Rogowsky, P.M.** (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 Physiol* **154**, 273-286.

- Johnson, E.S.** (2004). Protein Modification by SUMO. *Annual Rev Biochem* **73**, 355-382.
- Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C.** (2004). microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. *Nature* **428**, 84-88.
- Kadioglu, A., and Terzi, R.** (2007). A dehydration avoidance mechanism: Leaf rolling. *The Botanical Review* **73**, 290-302.
- Kadioglu, A., Terzi, R., Saruhan, N., and Saglam, A.** (2012). Current advances in the investigation of leaf rolling caused by biotic and abiotic stress factors. *Plant Sci* **182**, 42-48.
- 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.
- Kamata, N., Okada, H., Komeda, Y., and Takahashi, T.** (2013). Mutations in epidermis-specific HD-ZIP IV genes affect floral organ identity in *Arabidopsis thaliana*. *The Plant J* **75**, 430-440.
- Kao, S.-H., Wang, W.-L., Chen, C.-Y., Chang, Y.-L., Wu, Y.-Y., Wang, Y.-T., Wang, S.-P., Nesvizhskii, A.I., Chen, Y.-J., Hong, T.-M., and Yang, P.-C.** (2015). Analysis of protein stability by the cycloheximide chase assay. *Bio Protoc* **5**, e1374.
- Kelley, D.R., Arreola, A., Gallagher, T.L., and Gasser, C.S.** (2012). ETTIN (ARF3) physically interacts with KANADI proteins to form a functional complex essential for integument development and polarity determination in *Arabidopsis*. *Development* **139**, 1105-1109.
- Kerscher, O., Felberbaum, R., and Hochstrasser, M.** (2006). Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol* **22**, 159-180.
- Kerstetter, R.A., Bollman, K., Taylor, R.A., Bomblies, K., and Poethig, R.S.** (2001). KANADI regulates organ polarity in *Arabidopsis*. *Nature* **411**, 706-709.
- Khaled, A.S., Vernoud, V., Ingram, G.C., Perez, P., Sarda, X., and Rogowsky, P.M.** (2005). Engrailed-ZmOCL1 fusions cause a transient reduction of kernel size in maize. *Plant Mol Biol* **58**, 123-139.
- 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. *The Plant Cell* **26**, 2184-2200.
- Kinashi, H., Someno, K., and Sakaguchi, K.** (1984). Isolation and characterization of concanamycins A, B and C. *J Antibiot (Tokyo)* **37**, 1333-1343.

- 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.
- Klionsky, D.J., and Emr, S.D.** (2000). Autophagy as a regulated pathway of cellular degradation. *Science* **290**, 1717-1721.
- Krylov, D., and Vinson, C.R.** (2001). Leucine Zipper. In eLS, (Ed.) (
- Kubo, H., Peeters, A.J., Aarts, M.G., Pereira, A., and Koornneef, M.** (1999). ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in Arabidopsis. *The Plant Cell* **11**, 1217-1226.
- Kumaran, M.K., Bowman, J.L., and Sundaresan, V.** (2002). YABBY polarity genes mediate the repression of KNOX homeobox genes in Arabidopsis. *The Plant Cell* **14**, 2761-2770.
- Kurepa, J., Walker, J.M., Smalle, J., Gosink, M.M., Davis, S.J., Durham, T.L., Sung, D.Y., and Vierstra, R.D.** (2003). The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress. *J Biol Chem* **278**, 6862-6872.
- 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.
- Lee, D.H., and Goldberg, A.L.** (1998). Proteasome inhibitors: valuable new tools for cell biologists. *Trends in Cell Biol* **8**, 397-403.
- 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[S]. *J Lipid Res* **53**, 2677-2689.
- Li, L., Zheng, M., Deng, G., Liang, J., Zhang, H., Pan, Z., Long, H., and Yu, M.** (2016). Overexpression of AtHDG11 enhanced drought tolerance in wheat (*Triticum aestivum* L.). *Molecular Breeding* **36**, 23.
- Li, M., Welti, R., and Wang, X.** (2006). Quantitative profiling of Arabidopsis polar glycerolipids in response to phosphorus starvation. Roles of phospholipases D zeta1 and D zeta2 in phosphatidylcholine hydrolysis and digalactosyldiacylglycerol accumulation in phosphorus-starved plants. *Plant Physiol* **142**, 750-761.
- Li, S.J., and Hochstrasser, M.** (1999). A new protease required for cell-cycle progression in yeast. *Nature* **398**, 246-251.
- 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. *Biochemical and Biophysical Res Comm* **171**, 257-259.

- Lois, L.M., Lima, C.D., and Chua, N.-H.** (2003). Small Ubiquitin-Like Modifier Modulates Absciscic Acid Signaling in Arabidopsis. *The Plant Cell* **15**, 1347-1359.
- 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.
- Martins, S., Dohmann, E.M., Cayrel, A., Johnson, A., Fischer, W., Pojer, F., Satiat-Jeunemaitre, B., Jaillais, Y., Chory, J., Geldner, N., and Vert, G.** (2015). Internalization and vacuolar targeting of the brassinosteroid hormone receptor BRI1 are regulated by ubiquitination. *Nat Comm* **6**, 6151.
- 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.
- Matic, I., Schimmel, J., Hendriks, I.A., van Santen, M.A., van de Rijke, F., van Dam, H., Gnad, F., Mann, M., and Vertegaal, A.C.** (2010). Site-specific identification of SUMO-2 targets in cells reveals an inverted SUMOylation motif and a hydrophobic cluster SUMOylation motif. *Mol Cell* **39**, 641-652.
- Matschi, S., Vasquez, M.F., Bourgault, R., Steinbach, P., Chamness, J., Kaczmar, N., Gore, M.A., Molina, I., and Smith, L.G.** (2020). Structure-function analysis of the maize bulliform cell cuticle and its potential role in dehydration and leaf rolling. *Plant Direct* **4**, e00282.
- Matunis, M.J., Coutavas, E., and Blobel, G.** (1996). A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J Cell Biol* **135**, 1457-1470.
- Melchior, F.** (2000). SUMO--nonclassical ubiquitin. *Annu Rev Cell Dev Biol* **16**, 591-626.
- Merelo, P., Ram, H., Caggiano, M.P., Ohno, C., Ott, F., Straub, D., Graeff, M., Cho, S.K., Yang, S.W., Wenkel, S., and Heisler, M.G.** (2016). Regulation of *MIR165/166* by class II and class III homeodomain leucine zipper proteins establishes leaf polarity. *PNAS* **113**, 11973-11978.
- Miller, M.J., and Vierstra, R.D.** (2011a). Mass spectrometric identification of SUMO substrates provides insights into heat stress-induced SUMOylation in plants. *Plant Signal Behav* **6**, 130-133.
- Miller, M.J., and Vierstra, R.D.** (2011b). Mass spectrometric identification of SUMO substrates provides insights into heat stress-induced SUMOylation in plants. *Plant Signal Behav* **6**, 130-133.

- Miller, M.J., Barrett-Wilt, G.A., Hua, Z., and Vierstra, R.D.** (2010). Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in Arabidopsis. *PNAS USA* **107**, 16512-16517.
- Miller, M.J., Scalf, M., Rytz, T.C., Hubler, S.L., Smith, L.M., and Vierstra, R.D.** (2013). Quantitative proteomics reveals factors regulating RNA biology as dynamic targets of stress-induced SUMOylation in Arabidopsis. *Mol Cell Proteomics* **12**, 449-463.
- Moon, J., and Hake, S.** (2011). How a leaf gets its shape. *Curr Opin Plant Biol* **14**, 24-30.
- Moon, J., Parry, G., and Estelle, M.** (2004). The ubiquitin-proteasome pathway and plant development. *The Plant Cell* **16**, 3181-3195.
- 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.
- Mukherjee, S.** (2021). Cysteine modifications (oxPTM) and protein sulphenylation-mediated sulfenome expression in plants: evolutionary conserved signaling networks? *Plant Signal Behav* **16**, 1831792.
- Mukherjee, T., Subedi, B., Khosla, A., Begler, E.M., Stephens, P.M., Warner, A.L., Lerma-Reyes, R., Thompson, K.A., Gunewardena, S., and Schrick, K.** (2022). The START domain mediates Arabidopsis GLABRA2 dimerization and turnover independently of homeodomain DNA binding. *Plant Physiol* **kiac383**.
- Murashige, T., and Skoog, F.** (1962). A revised medium for rapid growth and bio assays with Tobacco tissue cultures. *Physiologia Plantarum* **15**, 473-497.
- 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.
- Nelis, S., Conti, L., Zhang, C., and Sadanandom, A.** (2015). A functional Small Ubiquitin-like Modifier (SUMO) interacting motif (SIM) in the gibberellin hormone receptor GID1 is conserved in cereal crops and disrupting this motif does not abolish hormone dependency of the DELLA-GID1 interaction. *Plant Signal Behav* **10**, e987528.
- Nishihara, T., Akifusa, S., Koseki, T., Kato, S., Muro, M., and Hanada, N.** (1995). Specific inhibitors of vacuolar type H(+)-ATPases induce apoptotic cell death. *Biochem Biophys Res Comm* **212**, 255-262.
- Novatchkova, M., Budhiraja, R., Coupland, G., Eisenhaber, F., and Bachmair, A.** (2004). SUMO conjugation in plants. *Planta* **220**, 1-8.

- O'Toole, J.C., and Cruz, R.T.** (1980). Response of leaf water potential, stomatal resistance, and leaf rolling to water stress. *Plant Physiol* **65**, 428-432.
- O'Toole, J.C., Cruz, R.T., and Singh, T.N.** (1979). Leaf rolling and transpiration. *Plant Science Letters* **16**, 111-114.
- Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., and Dunker, A.K.** (2003). Predicting intrinsic disorder from amino acid sequence. *Proteins: Structure, Function, and Bioinformatics* **53**, 566-572.
- 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.
- Okada, S., Nagabuchi, M., Takamura, Y., Nakagawa, T., Shinmyozu, K., Nakayama, J., and Tanaka, K.** (2009). Reconstitution of *Arabidopsis thaliana* SUMO pathways in *E. coli*: functional evaluation of SUMO machinery proteins and mapping of SUMOylation sites by mass spectrometry. *Plant Cell Physiol* **50**, 1049-1061.
- Okumoto, K., Misono, S., Miyata, N., Matsumoto, Y., Mukai, S., and Fujiki, Y.** (2011). Cysteine ubiquitination of PTS1 receptor Pex5p regulates Pex5p recycling. *Traffic* **12**, 1067-1083.
- 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. *The 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.
- Park, H.J., Kim, W.-Y., Park, H.C., Lee, S.Y., Bohnert, H.J., and Yun, D.-J.** (2011). SUMO and SUMOylation in plants. *Molecules and Cells* **32**, 305-316.
- Pekker, I., Alvarez, J.P., and Eshed, Y.** (2005). Auxin response factors mediate *Arabidopsis* organ asymmetry via modulation of KANADI activity. *The Plant Cell* **17**, 2899-2910.
- Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., and Obradovic, Z.** (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**, 208.
- Platts, L., and Falconer, R.J.** (2015). Controlling protein stability: Mechanisms revealed using formulations of arginine, glycine and guanidinium HCl with three globular proteins. *Int J Pharm* **486**, 131-135.

- Ponting, C.P., and Aravind, L.** (1999). START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends in Biochemical Sciences* **24**, 130-132.
- Prakash, S., Tian, L., Ratliff, K.S., Lehotzky, R.E., and Matouschek, A.** (2004). An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat Struct Mol Biol* **11**, 830-837.
- Preston, G.M., and Brodsky, J.L.** (2017). The evolving role of ubiquitin modification in endoplasmic reticulum-associated degradation. *Biochem J* **474**, 445-469.
- Prigge, M.J., Otsuga, D., Alonso, J.M., Ecker, J.R., Drews, G.N., and Clark, S.E.** (2005). Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. *The Plant Cell* **17**, 61-76.
- Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N., and Dunker, A.K.** (2007). Intrinsic disorder and functional proteomics. *Biophys J* **92**, 1439-1456.
- Radivojac, P., Vacic, V., Haynes, C., Cocklin, R.R., Mohan, A., Heyen, J.W., Goebel, M.G., and Iakoucheva, L.M.** (2010). Identification, analysis, and prediction of protein ubiquitination sites. *Proteins: Structure, Function, and Bioinformatics* **78**, 365-380.
- Reggiori, F., and Klionsky, D.J.** (2002). Autophagy in the eukaryotic cell. *Eukaryot Cell* **1**, 11-21.
- Rerie, W.G., Feldmann, K.A., and Marks, M.D.** (1994). The GLABRA2 gene encodes a homeodomain protein required for normal trichome development in Arabidopsis. *Genes Dev* **8**, 1388-1399.
- Rodriguez, M.S., Dargemont, C., and Hay, R.T.** (2001). SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J Biol Chem* **276**, 12654-12659.
- Romani, F., Reinheimer, R., Florent, S.N., Bowman, J.L., and Moreno, J.E.** (2018). Evolutionary history of HOMEODOMAIN LEUCINE ZIPPER transcription factors during plant transition to land. *New Phytol* **219**, 408-421.
- Rombolá-Caldentey, B., Rueda-Romero, P., Iglesias-Fernández, R., Carbonero, P., and Oñate-Sánchez, L.** (2014). Arabidopsis DELLA and Two HD-ZIP Transcription Factors Regulate GA Signaling in the Epidermis through the L1 Box-Element. *The Plant Cell* **26**, 2905.
- Romero-Barrios, N., and Vert, G.** (2018). Proteasome-independent functions of lysine-63 polyubiquitination in plants. *New Phytol* **217**, 995-1011.
- Rong, F., Chen, F., Huang, L., Zhang, J., Zhang, C., Hou, D., Cheng, Z., Weng, Y., Chen, P., and Li, Y.** (2019). A mutation in class III homeodomain-leucine zipper (HD-ZIP III) transcription factor results in curly leaf (cul) in cucumber (*Cucumis sativus* L.). *Theor Appl Genet* **132**, 113-123.

- Roostaei, 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.
- Roy, A., Kucukural, A., and Zhang, Y.** (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**, 725-738.
- Ruberti, I., Sessa, G., Lucchetti, S., and Morelli, G.** (1991). A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif. *The EMBO J* **10**, 1787-1791.
- Rytz, T.C., Miller, M.J., McLoughlin, F., Augustine, R.C., Marshall, R.S., Juan, Y.T., Charng, Y.Y., Scalf, M., Smith, L.M., and Vierstra, R.D.** (2018). SUMOylation profiling reveals a diverse array of nuclear targets modified by the SUMO ligase SIZ1 during heat stress. *The Plant Cell* **30**, 1077-1099.
- Saglam, A., Kadioğlu, A., Demiralay, M., and Terzi, R.** (2014). Leaf rolling reduces photosynthetic loss in maize under severe drought. *Acta Botanica Croatica* **73**.
- Sampson, D.A., Wang, M., and Matunis, M.J.** (2001). The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *J Biol Chem* **276**, 21664-21669.
- Saracco, S.A., Miller, M.J., Kurepa, J., and Vierstra, R.D.** (2007). Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. *Plant physiol* **145**, 119-134.
- Scaglione, K.M., Basrur, V., Ashraf, N.S., Konen, J.R., Elenitoba-Johnson, K.S., Todi, S.V., and Paulson, H.L.** (2013). The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates. *J Biol Chem* **288**, 18784-18788.
- Schena, M., and Davis, R.W.** (1992). HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily. *PNAS USA* **89**, 3894-3898.
- Schofield, P.N.** (1987). Patterns, puzzles and paradigms: the riddle of the homeobox. *Trends in Neurosciences* **10**, 3-6.
- Schrack, 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.
- Schrack, 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.
- Schütze, K., Harter, K., and Chaban, C.** (2008). Post-translational regulation of plant bZIP factors. *Trends in Plant Science* **13**, 247-255.

- Seifert, A., Schofield, P., Barton, G.J., and Hay, R.T.** (2015). Proteotoxic stress reprograms the chromatin landscape of SUMO modification. *Science Signaling* **8**.
- Shan, C.-M., Shangguan, X.-X., Zhao, B., Zhang, X.-F., Chao, L.-M., Yang, C.-Q., Wang, L.-J., Zhu, H.-Y., Zeng, Y.-D., Guo, W.-Z., Zhou, B.-L., Hu, G.-J., Guan, X.-Y., Chen, Z.J., Wendel, J.F., Zhang, T.-Z., and Chen, X.-Y.** (2014). Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. In *Nature communications*, pp. 5519.
- 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.
- Sharif, R., Raza, A., Chen, P., Li, Y., El-Ballat, E.M., Rauf, A., Hano, C., and El-Esawi, M.A.** (2021). HD-ZIP gene family: Potential roles in improving plant growth and regulating stress-responsive mechanisms in plants. *Genes (Basel)* **12**.
- Sharma, B., Joshi, D., Yadav, P.K., Gupta, A.K., and Bhatt, T.K.** (2016). Role of ubiquitin-mediated degradation system in plant biology. *Front Plant Sci* **7**, 806.
- 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. *The Plant J* **69**, 37-46.
- Shi, Z., Wang, J., Wan, X., Shen, G.-z., Wang, X.-q., and Zhang, J.-l.** (2006). Over-expression of rice OsAGO7 gene induces upward curling of the leaf blade that enhanced erect-leaf habit. *Planta* **226**, 99-108.
- Shimizu, Y., Okuda-Shimizu, Y., and Hendershot, L.M.** (2010). Ubiquitylation of an ERAD substrate occurs on multiple types of amino acids. *Mol Cell* **40**, 917-926.
- Siegfried, K.R., Eshed, Y., Baum, S.F., Otsuga, D., Drews, G.N., and Bowman, J.L.** (1999). Members of the YABBY gene family specify abaxial cell fate in *Arabidopsis*. *Development* **126**, 4117-4128.
- Singh, G.P., Ganapathi, M., Sandhu, K.S., and Dash, D.** (2006). Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins* **62**, 309-315.
- Skryhan, K., Cuesta-Seijo, J.A., Nielsen, M.M., Marri, L., Mellor, S.B., Glaring, M.A., Jensen, P.E., Palcic, M.M., and Blennow, A.** (2015). The role of cysteine residues in redox regulation and protein stability of *Arabidopsis thaliana* starch synthase 1. *PLoS One* **10**, e0136997.
- Smalle, J., and Vierstra, R.D.** (2004). The ubiquitin 26S proteasome proteolytic pathway. *Annu Rev Plant Biol* **55**, 555-590.

- 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.
- Sokalingam, S., Madan, B., Raghunathan, G., and Lee, S.-G.** (2013). In silico study on the effect of surface lysines and arginines on the electrostatic interactions and protein stability. *Biotechnology and Bioprocess Engineering* **18**, 18-26.
- Song, J.B., Huang, S.Q., Dalmay, T., and Yang, Z.M.** (2012). Regulation of leaf morphology by MicroRNA394 and its target LEAF CURLING RESPONSIVENESS. *Plant and Cell Physiol* **53**, 1283-1294.
- Subedi, B., and Schrick, K.** (2022). EYFP fusions to HD-Zip IV transcription factors enhance their stability and lead to phenotypic changes in Arabidopsis. *Plant Signal Behav* **17**, 2119013.
- Takahashi, H., Iwakawa, H., Ishibashi, N., Kojima, S., Matsumura, Y., Prananingrum, P., Iwasaki, M., Takahashi, A., Ikezaki, M., Luo, L., Kobayashi, T., Machida, Y., and Machida, C.** (2013). Meta-analyses of microarrays of Arabidopsis asymmetric leaves1 (as1), as2 and their modifying mutants reveal a critical role for the ETT pathway in stabilization of adaxial-abaxial patterning and cell division during leaf development. *Plant Cell Physiol* **54**, 418-431.
- Tatham, M.H., Plechanovová, A., Jaffray, E.G., Salmen, H., and Hay, R.T.** (2013). Ube2W conjugates ubiquitin to α -amino groups of protein N-termini. *Biochem J* **453**, 137-145.
- Taupitz, K.F., Dörner, W., and Mootz, H.D.** (2017). Covalent capturing of transient SUMO–SIM interactions using unnatural amino acid mutagenesis and photocrosslinking. *Chemistry – A European Journal* **23**, 5978-5982.
- Tempé, D., Piechaczyk, M., and Bossis, G.** (2008). SUMO under stress. In *Biochemical Society Transactions*, pp. 874-878.
- Thorsell, A.-G., Lee, W.H., Persson, C., Siponen, M.I., Nilsson, M., Busam, R.D., Kotenyova, T., Schöler, H., and Lehtiö, L.** (2011). Comparative structural analysis of lipid binding START domains. *PLoS One* **6**, e19521-e19521.
- Tompa, P., Prilusky, J., Silman, I., and Sussman, J.L.** (2008). Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins* **71**, 903-909.
- Tron, A.E., Welchen, E., and Gonzalez, D.H.** (2004). Engineering the loop region of a homeodomain-leucine zipper protein promotes efficient binding to a monomeric DNA binding site. *Biochemistry* **43**, 15845-15851.
- 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.
- Tsujishita, Y., and Hurley, J.H.** (2000). Structure and lipid transport mechanism of a StAR-related domain. *Nature Structural Biol* **7**, 408-414.

- Tsvetkov, P., Asher, G., Paz, A., Reuven, N., Sussman, J.L., Silman, I., and Shaul, Y.** (2008). Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. *Proteins* **70**, 1357-1366.
- Turner, N.C., O'Toole, J.C., Cruz, R.T., Namuco, O.S., and Ahmad, S.** (1986). Responses of seven diverse rice cultivars to water deficits I. Stress development, canopy temperature, leaf rolling and growth. *Field Crops Res* **13**, 257-271.
- Turunen, O., Vuorio, M., Fenel, F., and Leisola, M.** (2002). Engineering of multiple arginines into the Ser/Thr surface of *Trichoderma reesei* endo-1,4-beta-xylanase II increases the thermotolerance and shifts the pH optimum towards alkaline pH. *Protein Eng* **15**, 141-145.
- Uzunova, K., Götttsche, K., Miteva, M., Weisshaar, S.R., Glanemann, C., Schnellhardt, M., Niessen, M., Scheel, H., Hofmann, K., Johnson, E.S., Praefcke, G.J.K., and Dohmen, R.J.** (2007). Ubiquitin-dependent proteolytic control of SUMO conjugates. *J Biol Chem* **282**, 34167-34175.
- van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E., and Babu, M.M.** (2014). Classification of intrinsically disordered regions and proteins. *Chemical Reviews* **114**, 6589-6631.
- Vernoud, V., Laigle, G., Rozier, F., Meeley, R.B., Perez, P., and Rogowsky, P.M.** (2009). The HD-ZIP IV transcription factor OCL4 is necessary for trichome patterning and anther development in maize. *Plant J* **59**, 883-894.
- Vierstra, R.D.** (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat Rev Mol Cell Biol* **10**, 385-397.
- Vierstra, R.D.** (2012). The expanding universe of ubiquitin and ubiquitin-like modifiers. *Plant Physiol* **160**, 2-14.
- Vinson, C.R., Sigler, P.B., and McKnight, S.L.** (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246**, 911-916.
- Vinson, C.R., Hai, T., and Boyd, S.M.** (1993). Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes & development* **7** 6, 1047-1058.
- Vittal, V., Stewart, M.D., Brzovic, P.S., and Klevit, R.E.** (2015). Regulating the Regulators: Recent Revelations in the Control of E3 Ubiquitin Ligases. *J Biol Chem* **290**, 21244-21251.
- Vucetic, S., Brown, C.J., Dunker, A.K., and Obradovic, Z.** (2003). Flavors of protein disorder. *Proteins* **52**, 573-584.

- Walford, S.-A., Wu, Y., Llewellyn, D.J., and Dennis, E.S.** (2012). Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. *The Plant J* **71**, 464-478.
- Walton, A., Stes, E., Cybulski, N., Van Bel, M., Inigo, S., Durand, A.N., Timmerman, E., Heyman, J., Pauwels, L., De Veylder, L., Goossens, A., De Smet, I., Coppens, F., Goormachtig, S., and Gevaert, K.** (2016). It's time for some "site"-seeing: Novel tools to monitor the ubiquitin landscape in *Arabidopsis thaliana*. *The Plant Cell* **28**, 6-16.
- Wang, F., Yan, H., Fang, P., Ji, X., and Peng, Y.** (2022). Heterologous overexpression of ZmHDZIV13 enhanced drought and salt tolerance in *Arabidopsis* and Tobacco. *Agronomy* **12**, 2378.
- 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*. *The Plant J* **83**, 300-311.
- Wang, X., Bi, S., Wang, L., Li, H., Gao, B.A., Huang, S., Qu, X., Cheng, J., Wang, S., Liu, C., Jiang, Y., Zhang, B., Liu, X., Zhang, S., Fu, Y., Zhang, Z., and Wang, C.** (2020). GLABRA2 regulates actin bundling protein VILLIN1 in root hair growth in response to osmotic stress. *Plant Physiol* **184**, 176-193.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J.** (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.
- Wei, Q., Song, H.L., Xiang, C.B., and Kuai, B.K.** (2010). Preliminary study on improving drought tolerance of ryegrass (*Lolium perenne*) by overexpressing a HD-START transcription factor AtHDG11 from *Arabidopsis*. *Plant Physiol Comm* **46**, 1140-1146.
- 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.
- Wilkinson, K.A., and Henley, J.M.** (2010). Mechanisms, regulation and consequences of protein SUMOylation. *Biochem J* **428**, 133-145.
- Wojciechowska, I., Mukherjee, T., Knox-Brown, P., Hu, X., Khosla, A., Mathews, G.L., Thompson, K.A., Peery, S.T., Szlachetko, J., Thalhammer, A., Hinch, D.K., Skirycz, A., and Schrick, K.** (2021). PROTODERMAL FACTOR2 binds lysophosphatidylcholines and transcriptionally regulates phospholipid metabolism. *bioRxiv*, 2021.2010.2020.465175.

- Wright, P.E., and Dyson, H.J.** (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Mol Cell Biol* **16**, 18-29.
- Wu, R., Li, S., He, S., Wassmann, F., Yu, C., Qin, G., Schreiber, L., Qu, L.J., and 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*. *The Plant Cell* **23**, 3392-3411.
- Xiong, Y., Contento, A.L., Nguyen, P.Q., and Bassham, D.C.** (2007). Degradation of oxidized proteins by autophagy during oxidative stress in *Arabidopsis*. *Plant Physiol* **143**, 291-299.
- Yang, H., Shi, G., Li, X., Hu, D., Cui, Y., Hou, J., Yu, D., and Huang, F.** (2019). Overexpression of a soybean YABBY gene, GmFILA, causes leaf curling in *Arabidopsis thaliana*. *BMC Plant Biol* **19**, 234-234.
- Yang, J.-Y., Chung, M.-C., Tu, C.-Y., and Leu, W.-M.** (2002). OSTF1: A HD-GL2 family homeobox gene is developmentally regulated during early embryogenesis in rice. *Plant and Cell Physiol* **43**, 628-638.
- Yang, J., and Zhang, Y.** (2015a). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* **43**, W174-W181.
- Yang, J., and Zhang, Y.** (2015b). Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics* **52**, 5.8.1-5.8.15.
- Yoshimoto, K., Hanaoka, H., Sato, S., Kato, T., Tabata, S., Noda, T., and Ohsumi, Y.** (2004). Processing of ATG8s, ubiquitin-like proteins, and their deconjugation by ATG4s are essential for plant autophagy. *The Plant Cell* **16**, 2967-2983.
- Yu, H., Chen, X., Hong, Y.-Y., Wang, Y., Xu, P., Ke, S.-D., Liu, H.-Y., Zhu, J.-K., Oliver, D.J., and Xiang, C.-B.** (2008). Activated expression of an *Arabidopsis* HD-START protein confers drought tolerance with improved root system and reduced stomatal density. *The Plant Cell* **20**, 1134-1151.
- Yu, L., Chen, X., Wang, Z., Wang, S., Wang, Y., Zhu, Q., Li, S., and Xiang, C.** (2013). *Arabidopsis* enhanced drought tolerance1/HOMEODOMAIN GLABROUS11 confers drought tolerance in transgenic rice without yield penalty. *Plant Physiol* **162**, 1378-1391.
- Yu, L.H., Wu, S.J., Peng, Y.S., Liu, R.N., Chen, X., Zhao, P., Xu, P., Zhu, J.B., Jiao, G.L., Pei, Y., and Xiang, C.B.** (2016). *Arabidopsis* EDT1/HDG11 improves drought and salt tolerance in cotton and poplar and increases cotton yield in the field. *Plant Biotechnol J* **14**, 72-84.
- Zhang, C., Freddolino, P.L., and Zhang, Y.** (2017). COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic Acids Res* **45**, W291-W299.

- Zhang, F., Zuo, K., Zhang, J., Liu, X., Zhang, L., Sun, X., and Tang, K.** (2010). An L1 box binding protein, GbML1, interacts with GbMYB25 to control cotton fibre development. *J Exp Bot* **61**, 3599-3613.
- Zhang, G.H., Xu, Q., Zhu, X.D., Qian, Q., and Xue, H.W.** (2009). SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. *The Plant Cell* **21**, 719-735.
- Zhang, H., Ma, X., Li, W., Niu, D., Wang, Z., Yan, X., Yang, X., Yang, Y., and Cui, H.** (2019). Genome-wide characterization of NtHD-ZIP IV: different roles in abiotic stress response and glandular trichome induction. *BMC Plant Biol* **19**, 444.
- Zhao, Q., Xie, Y., Zheng, Y., Jiang, S., Liu, W., Mu, W., Liu, Z., Zhao, Y., Xue, Y., and Ren, J.** (2014). GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res* **42**, W325-330.
- Zheng, N., Schulman, B.A., Song, L., Miller, J.J., Jeffrey, P.D., Wang, P., Chu, C., Koepp, D.M., Elledge, S.J., Pagano, M., Conaway, R.C., Conaway, J.W., Harper, J.W., and Pavletich, N.P.** (2002). Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**, 703-709.
- Zhu, Z., Sun, B., Xu, X., Chen, H., Zou, L., Chen, G., Cao, B., Chen, C., and Lei, J.** (2016). Overexpression of AtEDT1/HDG11 in Chinese kale (*Brassica oleracea* var. alboglabra) enhances drought and osmotic stress tolerance. *Front Plant Sci* **7**, 1285.
- Zou, L.-p., Sun, X.-h., Zhang, Z.-g., Liu, P., Wu, J.-x., Tian, C.-j., Qiu, J.-l., and Lu, T.-g.** (2011). Leaf rolling controlled by the homeodomain leucine zipper class IV gene Roc5 in rice. *Plant Physiol* **156**, 1589-1602.

Chapter 4 - EYFP fusions to HD-Zip IV transcription factors enhance their stability and lead to phenotypic changes in *Arabidopsis*

This chapter has been published in *Plant Signaling and Behavior* Journal.

“Subedi, B. and K. Schrick, *EYFP fusions to HD-Zip IV transcription factors enhance their stability and lead to phenotypic changes in Arabidopsis*. *Plant Signaling & Behavior*, 2022. **17(1)**: p. 2119013.”

Abstract

Green fluorescent protein (GFP) and its derivatives are extensively used for labeling cells, monitoring gene expression and/or tracking the localization or interactions of proteins. Previous reports of detrimental effects of fluorescent protein (FP) expression include cytotoxicity and interference with fusion protein function or localization. Only a few studies have documented the fluorescent tag-specific effects in plants. Here, we show that placing an enhanced yellow FP (EYFP) tag on the amino-terminus of GLABRA2 (GL2) and PROTODERMAL FACTOR2 (PDF2), two developmentally important HD-Zip IV transcription factors from *Arabidopsis*, enhances their protein stability. Additionally, expression of EYFP:GL2 not only rescued the *gl2* null mutant but also resulted in the abnormal development of abaxially curled leaves associated with EYFP-tag induced GL2 overexpression. Our study raises concerns on the use of FPs regarding their effects on the native properties of target proteins as well as biological consequences of fusion protein expression on morphology.

Keywords: EYFP, HD-Zip IV transcription factors, *Arabidopsis*, protein stability, abaxial curly leaf, GLABRA2

Introduction

First observed in the jellyfish *Aequorea victoria* in 1962 (Shimomura et al., 1962; Prasher et al., 1992), Green Fluorescent Protein (GFP) and its derivatives have revolutionized the molecular biosciences. Using fluorescent proteins (FPs) as tags, it is possible to visualize proteins and to mark organelles, intact cells or whole organisms. FPs are used extensively for analyzing gene expression, reporting promoter activity, monitoring protein subcellular localization, studying protein-protein interactions, protein purification, designing biosensors or cell markers and lineage tracing (Chalfie et al., 1994; Leffell et al., 1997; Chudakov et al., 2010; Blatt and Grefen, 2014). Their widespread popularity is based on fast and spontaneous formation of chromophores by posttranslational folding in the presence of oxygen without the need of any additional co-factors, substrates or enzymes (Heim et al., 1994). FPs are generally compatible in fusion proteins, resistant to proteolytic enzymes and can be easily visualized with standard filters (Kain et al., 1995; Sheen et al., 1995).

Despite their above-mentioned useful properties, limitations of FPs have been increasingly recognized. GFP and its homologs are ~28 kDa proteins that form an eleven stranded beta-barrel structure with a length of ~4.2 nm and a diameter of ~2.4 nm (Prasher et al., 1992; Ormö et al., 1996; Yang et al., 1996). Addition of these relatively bulky tags can potentially interfere with protein folding and function as well as trafficking of proteins. However, limited evidence is available for such effects of FPs on fusion proteins. Some

undesirable properties of FPs include their propensity to oligomerize, pH sensitivity, photobleaching, and overlap of their spectral signals with autofluorescence of cellular components (Patterson et al., 1997; Lisenbee et al., 2003; Surribas et al., 2006; Costantini et al., 2012; Cranfill et al., 2016; Botman et al., 2019). Their expression can also have deleterious effects on cells, as exemplified by formation of cytotoxic aggregates (Lisenbee et al., 2003; Lenassi Zupan et al., 2004; Chudakov et al., 2010), induction of apoptosis (Hanazono et al., 1997; Liu et al., 1999), and organ pathology in mice (Huang et al., 2000; Krestel et al., 2004). One report showed that GFP expression in endothelial cells causes transcriptional upregulation of a heat shock protein, resulting in induction of cyclooxygenase-2, prostaglandin production and increase in regional blood flow (Zhang et al., 2003). In *Arabidopsis*, the presence of GFP at the amino- or carboxyl-terminus of a virus movement protein sterically hindered its ordered aggregation into tubules required for virus spread (Thomas and Maule, 2000). Similarly, overexpression of GFP fused to the actin-binding domain of mouse talin (GFP:mTalin) which is typically used to study the actin cytoskeleton, attenuated blue-light-induced chloroplast movements in *Arabidopsis* leaves (Luesse et al., 2006). In a separate study, overexpression of the same fusion protein caused a reduction in actin depolymerization, resulting in severe defects in root hair expansion (Ketelaar et al., 2004).

Here, we report the effects of Enhanced Yellow Fluorescent Protein (EYFP) tags on class IV homeodomain leucine-zipper (HD-Zip IV) transcription factors in *Arabidopsis*. These regulatory proteins are specific to epidermal or sub-epidermal cells of plants and play an important role in differentiation of these cell layers, anthocyanin accumulation, lipid transport, cuticle biosynthesis and responses to stress ((Nakamura et al., 2006; Yu et al., 2008; Javelle et al., 2011; Wu et al., 2011); reviewed in (Chew et al., 2013)). GLABRA2 (GL2), the founding

member of this family from *Arabidopsis*, is expressed in developing trichomes and neighboring epidermal cells in leaves during early stages of trichome development (Rerie et al., 1994; Szymanski et al., 1998). Later, GL2 is mainly expressed in trichomes and is required for the subsequent stages of their morphogenesis including cell expansion, branching and maturation of the trichome cell wall (Szymanski et al., 1998). GL2 is also preferentially expressed in the differentiating hairless cells in the root and hypocotyl epidermis (Masucci et al., 1996; Hung et al., 1998; Lin and Schiefelbein, 2001) and is involved in the inhibition of root hair initiation in non-hair cell files (Di Cristina et al., 1996; Masucci et al., 1996; Hung et al., 1998). It also regulates mucilage biosynthesis in the seed coat (Western et al., 2001; Western et al., 2004). Another HD-Zip IV member PROTODERMAL FACTOR2 (PDF2) functions redundantly with *Arabidopsis thaliana* MERISTEM LAYER1 (ATML1) in shoot epidermal differentiation during embryonic development (Abe et al., 2003; Ogawa et al., 2015). Both genes positively regulate the expression of a subset of genes specific to the outermost layer (L1) of the shoot apical meristem (Abe et al., 2001; Abe et al., 2003; Nakamura et al., 2006; Rombolá-Caldentey et al., 2014).

In numerous previous studies, FP tags were translationally fused to HD-Zip IV TFs to study their tissue specific gene expression patterns, determine protein subcellular localization, and analyze mutant variants (Takada and Jürgens, 2007; Yu et al., 2008; Schrick et al., 2014; Lin et al., 2015; Iida et al., 2019; Nagata et al., 2021). In particular, our previous work demonstrated that an EYFP-tagged GL2 protein rescues the null mutant phenotype of *gl2* plants, including rescue of defects in three different cell types: trichomes, non-hair root epidermal cells, and seed coat cells (Schrack et al., 2014; Mukherjee et al., 2022). However, tag-induced side-effects have not been described to date. In this study, we report the effect of amino-terminal EYFP fusions on

the stability of two HD-Zip IV transcription factors, GL2 and PDF2, along with an abnormal curly leaf phenotype associated with elevated expression of EYFP-tagged GL2.

Materials and methods

Plant materials and growth conditions

The wild-type (WT) *Arabidopsis thaliana* ecotype Columbia (Col), *gl2-5* null mutant (Khosla et al., 2014) and transgenic plants expressing EYFP or FLAG-tagged GL2 were grown on soil containing Berger BM6, vermiculite and perlite (Hummert International) in a ratio of 4:2:1 at 23°C under continuous light.

Generation of transgenic lines

The *proGL2:EYFP:GL2* construct was generated using SR54 binary vector as previously described (Schrack et al., 2014). For FLAG tagging, the *proGL2:EYFP:GL2* construct was restriction digested with *Bam*HI and *Sal*I to remove the EYFP tag and the 6xHis 3xFLAG tag (Huang et al., 2016) was added to amino-terminus of *GL2* using NEBuilder HiFi DNA Assembly (New England Biolabs, E5520S). Construction of *HA:GL2*, *EYFP:PDF2* and *HA:PDF2* is described elsewhere (Wojciechowska et al., 2021). The binary vector constructs were used to transform *gl2-5* plants by *Agrobacterium tumefaciens* (GV3101) mediated floral dip (Clough and Bent, 1998). Transgenic lines were screened in presence of 20 µg/ml Hygromycin B. For each genotype, 37-42 independent T1 transformants were characterized for the rescue of trichome phenotype, EYFP expression and leaf curling. T2 progeny were analyzed for 3:1 segregation of the transgene. T3 homozygous lines confirmed by PCR genotyping and sequencing were used for phenotypic analysis and further experiments.

In vivo protein stability assay

Seeds from EYFP-, FLAG- and HA-tagged GL2 or PDF2 lines were vapor sterilized, sown on 0.5X MS agar (Murashige and Skoog, 1962), stratified at 4°C for 4-5 days, and germinated in vertical orientation at 23°C under continuous light. After 4 days, samples of ~30 seedlings each were transferred to 24-well plates containing 1 ml 0.5X liquid MS media, followed by the incubation on a shaker at 60 rpm for ~16 h under continuous light at 23°C. Next, cycloheximide (Sigma-Aldrich, C1988) (400 μ M in DMSO) or DMSO control was added to the wells. Seedling samples were harvested in screw-cap microcentrifuge tubes at the designated times by immediate freezing in liquid nitrogen followed by storage at -80°C until further use. Proteins were extracted from the seedlings by adding Laemmli sample buffer at 95°C after grinding in liquid nitrogen. For Western blotting, samples were separated on Bio-Rad Mini-PROTEAN TGX gels (10%), transferred into PVDF membranes and incubated with primary antibodies against GFP (1:2000; Roche, 11814460001), FLAG (1:10,000; Sigma, F1804) or HA (1:10,000; Pierce, 26183) followed by goat anti-mouse secondary IgG HRP antibodies (1:3300; GenScript, A00160). Proteins were detected with SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific) using an Azure 300 chemiluminescence imager (Azure Biosystems). The membranes were stained with Bio-Safe Coomassie Blue G-250 (Bio-Rad) for loading controls, and band intensities were quantified with ImageJ software.

Phenotypic assays, leaf biometry and microscopy

Rosettes and trichomes of ~15-d-old plants and EYFP expression were imaged with a Leica M165 fluorescence stereomicroscope fitted with a GFP2 filter set and a Leica DFC295

digital camera. To measure seedling lengths, sterile seeds were sown on 0.8% agar media, stratified at 4°C for 5 days and the plates were placed vertically at 23°C under continuous light for 4 days. The leaf blade lengths and widths were measured on the 6th leaves (Farmer et al., 2013) of 25-d-old plants of each genotype. Digital images of seedlings, whole plants and leaves were captured using a Google Pixel 3XL (12MP f/1.8) camera. To determine the subcellular localization of EYFP-tagged proteins, the primary roots of 4-d-old seedlings stained with propidium iodide (10 µg/ml) were imaged with a Zeiss LSM 880 confocal microscope. For acquiring EYFP signals, a 488 nm Argon laser and 500–550 nm emission filter was used. For propidium iodide, a 561 nm DPSS laser in combination with 575–645 nm emission filter was used.

Leaf Rolling Index (LRIs)/ Degree of epinasty was calculated following the protocol described in (Shi et al., 2006) and (Fierro et al., 2015).

$LRI = (W_m - W_n)/W_m \times 100$, where W_m is the greatest width of the leaf blade in its expanded state and W_n is the natural distance of the leaf blade margins at the same site.

Interpretation of LRI values:

0= Flat leaf blade (no curling)

+ve= Epinastic leaf blade (downward curling)

-ve= Hyponastic leaf blade (upcurling)

Results

EYFP fusions to HD-Zip IV TFs enhance their protein stability

EYFP is a GFP variant comprising of S65G, V68L, S72A and T203Y substitutions relative to wild-type GFP (Ormö et al., 1996; Cubitt et al., 1998). Its excitation and emission

peaks at 514 nm and 527 nm, respectively, are associated with stacking of Tyr66 with Tyr203 (Ormö et al., 1996; Rizzo et al., 2004). Here, we used transgenic *Arabidopsis* lines expressing amino-terminal fusions of EYFP to the HD-Zip IV transcription factors GL2 and PDF2 in comparison to amino-terminal fusions using 3xFLAG or HA tags (**Figure 4.1A**). The FLAG tag, also known as the DYKDDDDK-tag, is a short, hydrophilic epitope tag of ~1.013 kDa (Hopp et al., 1988) that is commonly implemented in a tandem configuration as the 3xFLAG peptide (Ueda et al., 2011). Similarly, the HA tag is a ~1.102 kDa peptide corresponding to residues 98-106 of human influenza virus hemagglutinin (HA) (Wilson et al., 1984). For the EYFP and 3xFLAG tags, the fusion proteins were expressed under the *GL2* native promoter, whereas the cauliflower mosaic virus (CaMV) 35S promoter was used to drive the HA-tag fusions.

To determine the effect of EYFP fusions on protein stability, we performed Western blot (WB) analysis of protein samples taken at various time points after treatment of *Arabidopsis* seedlings with cycloheximide (CHX) (**Figure 4.1B**). CHX is an inhibitor of eukaryotic translational elongation that is routinely used to study protein turnover in the absence of translation. Quantification of blot images showed that while the protein levels of GL2 fused to the smaller tags, 3xFLAG and HA, dropped over time with half-lives of ~6.5 and ~22 h respectively, EYFP:GL2 remained stable throughout the 24 h time course (**Figures 4.1C and 4.1D**). A more dramatic difference was observed between the half-lives of EYFP- and HA-tagged PDF2. The EYFP:PDF2 protein remained significantly more stable compared to HA:PDF2, which exhibited hardly detectable protein levels after 2 h (**Figures 4.1D and 4.1E**). These data indicate that the amino-terminal fusion to EYFP confers increased stability to HD-Zip IV TFs in plants.

Expression of EYFP:GL2 results in downward leaf curling in *Arabidopsis*

Previous studies revealed the role of GL2 in the differentiation of leaf epidermal cells, especially in the formation of trichomes (Rerie et al., 1994; Szymanski et al., 1998), but GL2 has not been implicated in the leaf polarity establishment to date. In our study, screening T1 transformants for the rescue of null phenotypes indicated complete rescue of trichome defects in 30.95% (n=42) and 54.05% (n=37) of *EYFP:GL2* and *FLAG:GL2* transformants, respectively (**Figure 4.2A**). The *EYFP:GL2* lines exhibited a range of EYFP expression including strong (17.07%), medium (31.71%), weak (21.95%) and no (29.27%) expression (**Figure 4.2B**). Strikingly, 9.76% of *EYFP:GL2* plants displayed abnormal curling of leaves, a phenotype that was absent in *FLAG:GL2* lines (**Figure 4.2C**). Confocal laser scanning microscopy confirmed the nuclear localization of the EYFP:GL2 protein in the primary root of *Arabidopsis* seedlings (**Figure 4.2D**) indicating that fusing EYFP at the amino-terminus of GL2 does not alter its subcellular localization, consistent with previous reports (Schrack et al., 2014; Mukherjee et al., 2022). Taken together, our observations led us to hypothesize that the elevated protein stability of the EYFP:GL2 protein causes abnormal leaf curling during development.

The curly leaves in *EYFP:GL2* plants display a narrow and pointed morphology associated with downward rolling and altered adaxial-abaxial leaf polarity (**Figures 4.2E, 4.2F and 4.3A**). The leaf curling phenotype largely contributed to the distinct appearance of *EYFP:GL2* rosettes in comparison to *FLAG:GL2*, *gl2-5* null mutant or Col wild-type lines (**Figure 4.2E**). The onset of this phenotype occurred at around day 12-13 after germination, with rolling occurring in either the longitudinal or transverse direction (**Figure 4.2E**). However, longitudinal curling became predominant later in development and the intensity of leaf curling increased at the stages leading up to flowering (**Figure 4.2F**). We next asked whether the

EYFP:GL2 protein expression level is altered in curly leaf plants in comparison to normal leaf plants. Under an epifluorescence stereomicroscope, fluorescent signals in leaf trichome nuclei of curly leaf plants appeared much stronger compared to those of normal leaf plants (**Figure 4.2G**). To confirm this, WB analysis was performed with leaves from 25-d-old curly leaf and normal leaf plants expressing EYFP:GL2 (**Figure 4.2H**). Quantification of protein bands revealed significantly higher expression of EYFP:GL2 in curly leaves compared to normal/flat leaves (**Figure 4.2I**).

Characterization of the curly leaf trait associated with EYFP:GL2 expression

Leaf size and shape play important roles in overall plant architecture and physiology (Govaerts et al., 1996; Zhang et al., 2009; Moon and Hake, 2011). Appropriate leaf rolling is of agronomic value as it is associated with erect leaf canopies, higher photosynthetic efficiency, optimum canopy temperature, and protection against UV light (Turner et al., 1986; Saglam et al., 2014; Fierro et al., 2015). Leaf rolling is also critical for adaptation to drought as it affects osmotic adjustment and reduces transpirational water loss (O'Toole et al., 1979; O'Toole and Cruz, 1980; Hsiao et al., 1984). To quantify the effect of EYFP:GL2 expression on the three-dimensional architecture of leaves, we measured the distances between the leaf edges halfway from the leaf blades in curly leaf versus normal/flat plant lines. Measurement of the leaf dimensions revealed significantly reduced leaf blade width in curly leaf lines compared to the normal/flat leaf lines although both displayed similar leaf lengths (**Figure 4.3B**). In accordance, *EYFP:GL2* plants exhibited the highest values for the rolling leaf index (LRI) or degree of epinasty, followed by Col wild type, *FLAG:GL2* and *gl2-5* null mutant plants (**Figure 4.3C**). These quantitative data confirm our original observations that the leaf blade edges curl

downwards to a greater extent in *EYFP:GL2* curly leaf lines (**Figures 4.2D, 4.2E and 4.3A**). The curly leaf plants also exhibited short stature with somewhat stunted growth (**Figures 4.3D and 4.3E**) and slightly early flowering (**Figure 4.3F**) compared to plants with normal leaf morphology. In contrast, there were no significant differences in the lengths of the hypocotyl and root in young seedlings (**Figures 4.3G and 4.3H**).

Discussion

In this chapter, we quantified the stability of HD-Zip IV transcription factors fused to various tags and reported the phenotypic changes in *Arabidopsis* lines expressing highly stable EYFP:GL2 fusion proteins. There are several possible reasons behind altered stability of fluorescently tagged HD-Zip IV proteins. Addition of the EYFP tag could potentially change protein conformation, interfere with binding partners or impact post-translational modifications (PTMs) such as phosphorylation or ubiquitination altering the extent of protein activation or degradation. In a related study using mammalian cells, expression of EGFP inhibited ubiquitination by RING-type E3 ubiquitin ligases TRAF6 and API2, resulting in impaired NF- κ B and JNK signaling and enhanced stability of the tumor suppressor p53 (Baens et al., 2006). Intriguingly, it was found that EGFP contains a consensus motif (PNEKRD) implicated in TRAF6 binding (Baens et al., 2006). In another study, the expression of EGFP impaired actin-myosin interactions and thus contractile properties of cardiac muscle cells by competitive binding to the actin-binding site of myosin (Agbulut et al., 2006; Agbulut et al., 2007). Proteins also have targeting signals and sites for PTMs such as palmitoylation, myristylation and farnesylation (Martoglio and Dobberstein, 1998; Hughes et al., 2001; Tian et al., 2004) which are reported to be obscured by fusion of fluorescent tags (Tian et al., 2004; DeBlasio et al., 2010).

Although the EYFP:GL2 protein is functional in terms of its ability to rescue *gl2* null mutants, amino-terminal tagging may nonetheless cause subtle changes in GL2 trafficking or turnover. Fusion to the EYFP tag could additionally disrupt another aspect of GL2 function when it is overexpressed, such as slightly altering its DNA-binding specificity, thus causing misexpression of target genes responsible for leaf curling.

In a previous study, it was reported that ectopic expression of *GL2* using the CaMV 35S promoter disrupts the endogenous function of *GL2* and fails to rescue the *gl2-1* mutant (Ohashi et al., 2002). In the same study, additive expression of *GL2* under its native promoter in the wild-type genetic background resulted in an increase in trichome initiation along with altered spacing of trichomes (Ohashi et al., 2002). However, a curly leaf phenotype was not reported in either scenario that was associated with overexpression of untagged GL2. Therefore, the occurrence of leaf curl in our study seems to be attributed to the EYFP component of the fusion protein. Furthermore, the data presented here show that the more pronounced abaxial leaf curl or epinasty of leaf blade edges in a subset of *EYFP:GL2* lines was associated with the enhanced protein level and stability of GL2 (**Figures 4.1C, 4.1D, 4.2G, 4.2H and 4.2I**). Not all plants expressing *EYFP:GL2* displayed this phenotype (**Figures 4.2A, 4.2E, 4.3A, 4.3B, 4.3C and 4.3D**) because each transformation event represents a random insertion of the transgene into the genome and positions of the insertion and/or tandem insertions can impact expression levels (Figure 2B). Leaf curling is usually associated with polarity changes in the leaf abaxial–adaxial axis (McConnell and Barton, 1998; Byrne, 2005). A complex regulatory network consisting of various transcription factors including HD-Zip III (McConnell et al., 2001; Emery et al., 2003; Prigge et al., 2005; Merelo et al., 2016), YABBY (Siegfried et al., 1999; Bowman, 2000; Yang et al., 2019) and KANADI (KAN) (Eshed et al., 2001; Kerstetter et al., 2001; Zhang et al., 2009;

Merelo et al., 2017; Zumajo-Cardona et al., 2019), small non-coding/microRNAs (Williams et al., 2005; Song et al., 2012; Merelo et al., 2016) and components of hormonal signaling including the *Auxin/Indole-3-Acetic Acid (Aux/IAA)* gene family and the redundantly acting *AUXIN RESPONSE FACTOR (ARF)* genes *ETTIN (ETT)/ARF3 ARF4* (Pekker et al., 2005; Hunter et al., 2006; Li et al., 2007; Fierro et al., 2015) are known to control establishment of leaf polarity. Auxin biosynthesis, polar transport, and signal transduction are important for plant growth and development (Teale et al., 2006; Mockaitis and Estelle, 2008) and defects in these processes can lead to leaf curling and dwarf/semi-dwarf phenotypes (Reed, 2001; Overvoorde et al., 2005; Zhao et al., 2019). Since we observed dwarf stature and reduced plant growth in association with leaf curling in plants expressing EYFP:GL2 (**Figure 4.3D and 4.3E**), it is possible that GL2 overexpression exerts its effect through auxin signaling or another hormone pathway.

A few HD-Zip IV genes were implicated in leaf curling previously. In one study, knockout of *Rice Outermost Cell-Specific 5 (ROC5)*, an ortholog of *GL2*, resulted in abaxial leaf rolling because of abnormal bulliform cell formation in the adaxial leaf blade (Zou et al., 2011). Vice versa, overexpression of *ROC5* suppressed the development of bulliform cells giving rise to adaxially rolled leaves in rice (Zou et al., 2011). However, we observed the opposite effect with *GL2* overexpression in *Arabidopsis* (**Figures 4.2C, 4.2E, 4.2F, 4.3A, 4.3B, 4.3C and 4.3D**). A more recent study showed that HD-Zip IV member *ROC8* heterodimerizes with *ROC5* in transcriptional repression of *Abaxially Curled Leaf 1 (ACL1)*, a positive regulator of bulliform cell development (Fang et al., 2021). Overexpression of *GL2* seems to induce leaf curling by a different mechanism since bulliform cells are specific to monocots. One approach to identify the leaf polarity determinants affected by EYFP:GL2 expression might be transcriptomic analysis of

differentially expressed genes in curly versus normal/flat leaves. Similarly, mRNA *in situ* hybridization could be used to identify tissue-specific or polarized distribution of specific transcripts that regulate leaf polarity (Juarez et al., 2004; Fahlgren et al., 2006; Hunter et al., 2006). Since changes in epidermal cell shape, size or arrangement may be associated with leaf curling (Yang et al., 2019), anatomical examination of leaves with curly leaf morphologies will be useful in future studies.

Although the use of fluorescent tags is critical for specific experiments, it is often helpful to use alternative tags of different sizes and/or properties to rule out any tag-specific observations. If feasible, using specific antibodies against the proteins of interest will also aid in resolving tag-related issues, e.g., to determine whether the fusion proteins are overexpressed relative to their endogenous counterparts. The position of the tag may also influence protein activity. While our analysis was restricted to amino-terminal fusions, it would be interesting to see whether attaching EYFP to the carboxyl-terminus or inserting it internally between domains affects GL2 function and/or leaf curling. Nonetheless, chances of affecting protein structural features or disrupting interactions with a ligand or other proteins are high if the site of insertion is not selected properly. In summary, being aware of the limitations of fluorescent tags and available alternatives is important when generating fusion proteins in research studies.

Figures

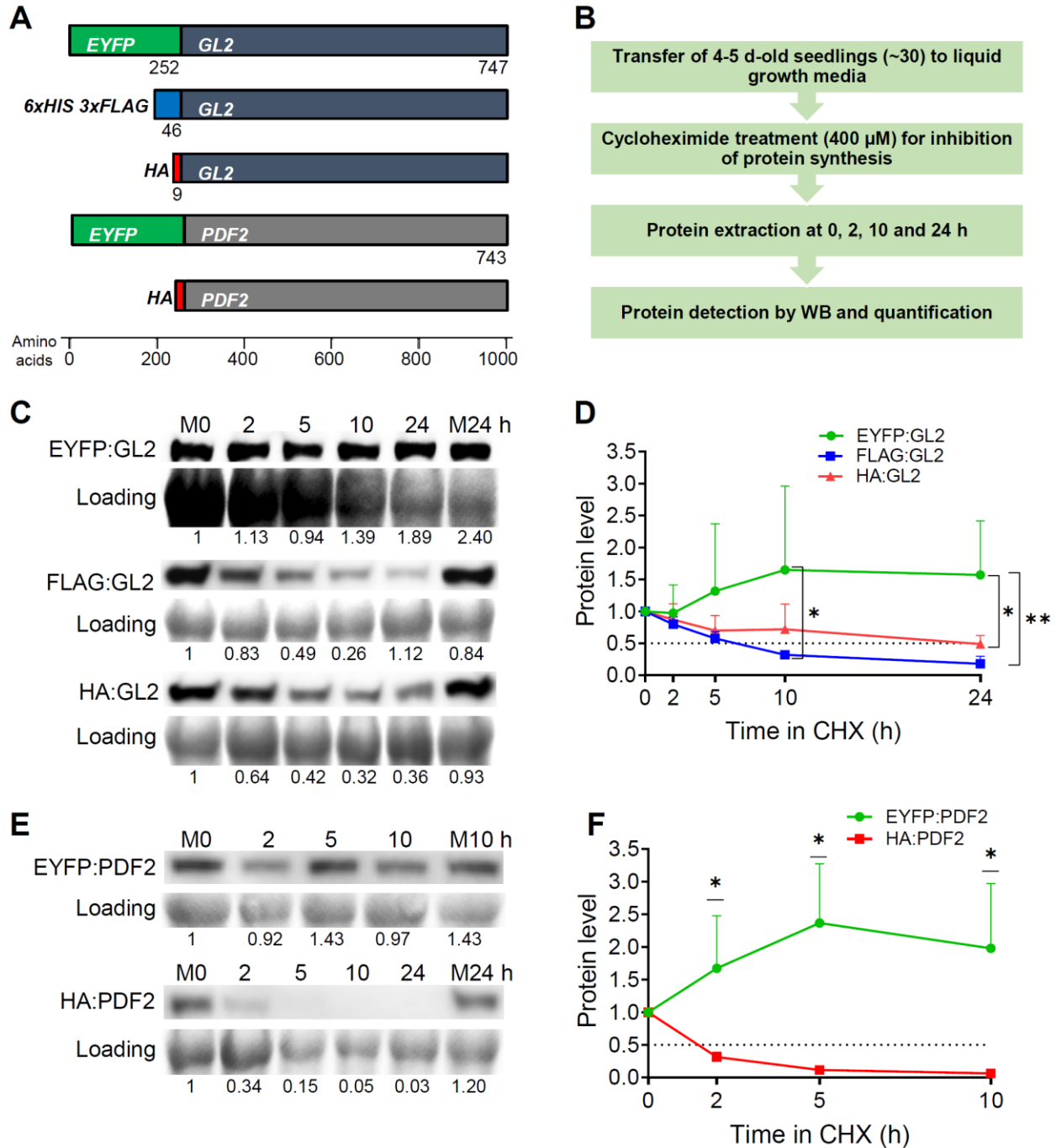


Figure 4.1. EYFP fusions to HD-Zip IV TFs enhance their protein stability.

(A) Schematic of constructs (drawn to scale) used in this study including EYFP-, FLAG- and HA-tagged GL2 and EYFP- and HA-tagged PDF2. EYFP- and FLAG-tagged proteins were

expressed under the GL2 native promoter while HA-tagged proteins were expressed under the CaMV 35S promoter. **(B)** Flow diagram of the *in vivo* CHX chase experiment used to monitor protein turnover. **(C-F)** Blot images and corresponding data quantification from CHX experiments for GL2 **(C, D)** and PDF2 **(E, F)** proteins. M0, M10 and M24, Mock treatments for 0, 10 and 24 h; Loading, RuBisCo bands from Coomassie Blue stained total protein. Means are shown for n=3 biological replicates, and error bars signify standard deviations (SDs). Unpaired *t*-test or ordinary two-way ANOVA using Tukey's multiple comparisons test was applied to test the significance of the difference between protein levels (**p* < 0.05, ***p* < 0.005).

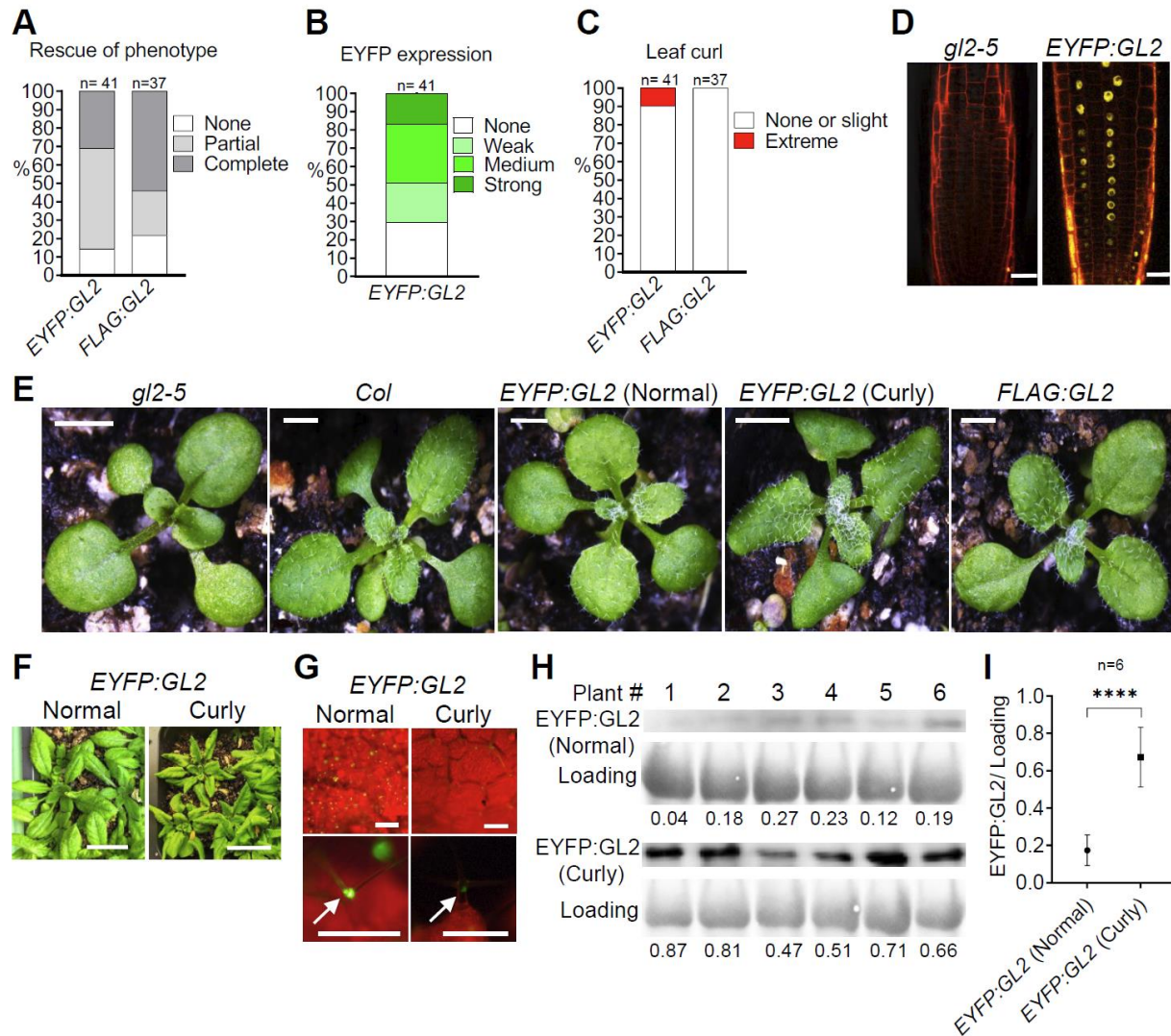


Figure 4.2. Expression of EYFP:GL2 leads to development of abaxially curled leaves in *Arabidopsis*.

Graphs showing (A) rescue of trichome phenotype, (B) EYFP expression and (C) occurrence of curly leaves in T1 progeny of plants stably transformed with indicated construct. (D) Confocal laser scanning microscopy indicating the nuclear localization of EYFP:GL2 in the primary root of a 4-d-old seedling stained with propidium iodide. Size bar = 20 μ m. (E) Rosettes of 15-d-old wild type (*Col*), *gl2-5*, EYFP:GL2 and FLAG:GL2 plants. Size bar = 2 mm. (F) 25-d-old EYFP:GL2 plants with two types of leaves. Size bar = 2 cm. (G) Expression of EYFP:GL2

in rosettes and leaf trichomes of 15-d-old plants. Size bar = 500 μm . **(H)** Blot images and **(I)** corresponding quantification data of EYFP:GL2 expression in the leaves of 25-d-old plants. Numerical values indicating EYFP:GL2 protein levels are represented as mean $\pm$ SD (n=6 biological replicates). Unpaired *t*-test (two-tailed) was applied to test the significance of the difference between protein levels (**** $p < 0.0001$).

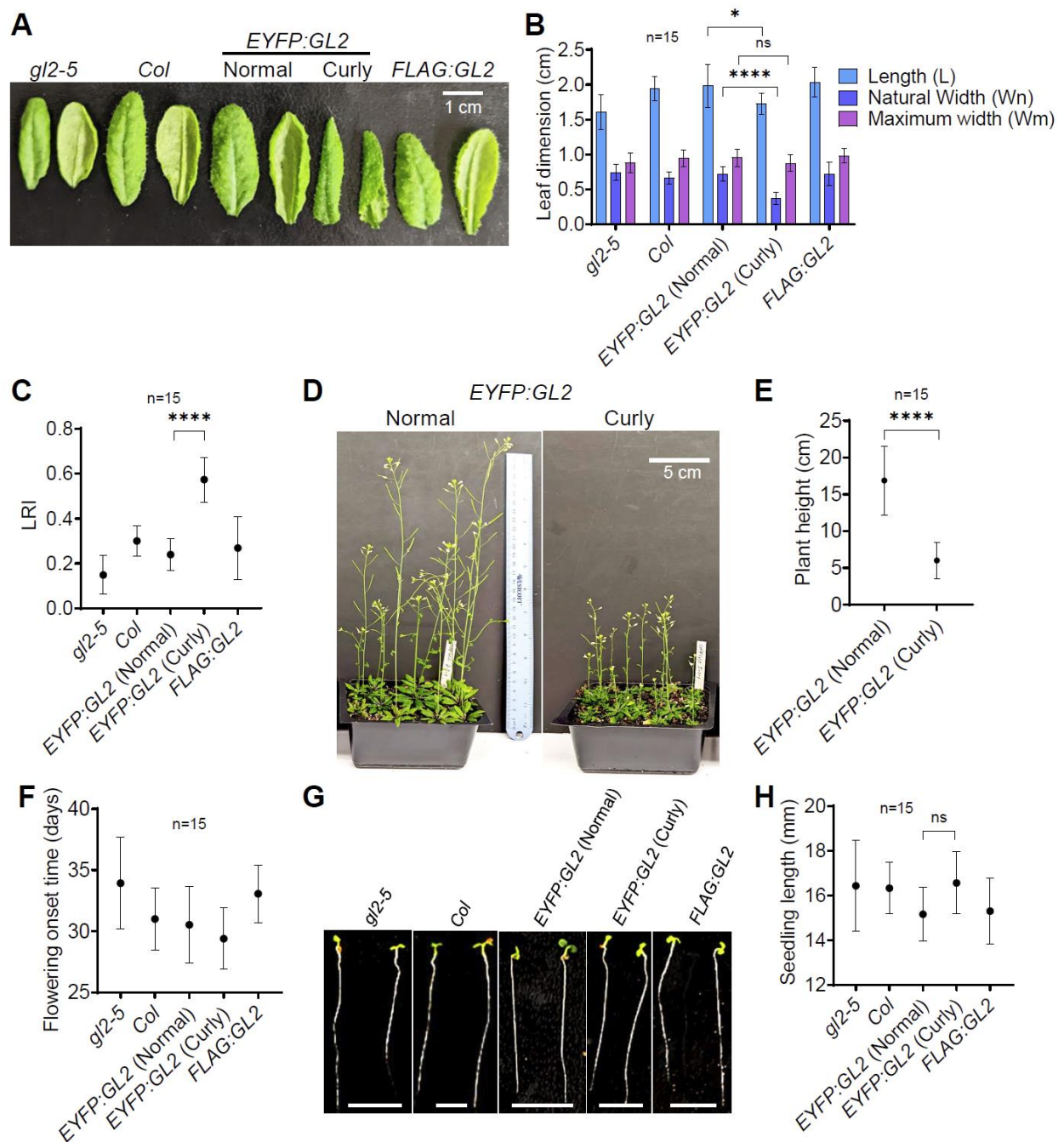


Figure 4.3. Quantitative characterization of curly leaf plants in comparison to normal/flat leaf plants.

(A) The phenotypes of adaxial (left) and abaxial (right) faces, and (B) leaf blade length and width statistics of the 6th leaves of 25-d-old plants. (C) Degree of leaf 6 epinasty in 25-d-old

plants of various genotypes. **(D)** Images of 45-d-old *EYFP:GL2* plants with two types of leaves and **(E)** corresponding graph indicating their heights. **(F)** Time of onset of flowering in *EYFP:GL2* plants bearing normal and curly leaves. **(G)** Images of 4-d-old seedlings of various genotypes and **(H)** graph showing their lengths. Size bar = 5 mm. For all measurements, means are shown for n=15 plants, and error bars indicate SDs. Depending upon the number of variables, ordinary one- or two-way ANOVA using Tukey's multiple comparisons test was used to establish significant differences between genotypes (*p < 0.05, ****p < 0.0001).

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. *The 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.
- Agbulut, O., Huet, A., Niederländer, N., Puceat, M., Menasché, P., and Coirault, C.** (2007). Green fluorescent protein impairs actin-myosin interactions by binding to the actin-binding site of myosin. *J Biol Chem* **282**, 10465-10471.
- Agbulut, O., Coirault, C., Niederländer, N., Huet, A., Vicart, P., Hagège, A., Puceat, M., and Menasché, P.** (2006). GFP expression in muscle cells impairs actin-myosin interactions: implications for cell therapy. *Nature Methods* **3**, 331-331.
- Baens, M., Noels, H., Broeckx, V., Hagens, S., Fevery, S., Billiau, A.D., Vankelecom, H., and Marynen, P.** (2006). The dark side of EGFP: defective polyubiquitination. *PLoS One* **1**, e54.
- Blatt, M.R., and Grefen, C.** (2014). Applications of fluorescent marker proteins in plant cell biology. *Methods Mol Biol* **1062**, 487-507.
- Botman, D., de Groot, D.H., Schmidt, P., Goedhart, J., and Teusink, B.** (2019). In vivo characterisation of fluorescent proteins in budding yeast. *Sci Rep* **9**, 2234.
- Bowman, J.L.** (2000). The YABBY gene family and abaxial cell fate. *Curr Opin Plant Biol* **3**, 17-22.
- Byrne, M.E.** (2005). Networks in leaf development. *Curr Opin in Plant Biol* **8**, 59-66.
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C.** (1994). Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805.
- Chew, W., Hrmova, M., and Lopato, S.** (2013). Role of Homeodomain leucine zipper (HD-Zip) IV transcription factors in plant development and plant protection from deleterious environmental factors. *Int J Mol Sci* **14**, 8122-8147.
- Chudakov, D.M., Matz, M.V., Lukyanov, S., and Lukyanov, K.A.** (2010). Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev* **90**, 1103-1163.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant J* **16**, 735-743.

- Costantini, L.M., Fossati, M., Francolini, M., and Snapp, E.L.** (2012). Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions. *Traffic* **13**, 643-649.
- Cranfill, P.J., Sell, B.R., Baird, M.A., Allen, J.R., Lavagnino, Z., de Gruiter, H.M., Kremers, G.-J., Davidson, M.W., Ustione, A., and Piston, D.W.** (2016). Quantitative assessment of fluorescent proteins. *Nature methods* **13**, 557-562.
- Cubitt, A.B., Woollenweber, L.A., and Heim, R.** (1998). Chapter 2: Understanding structure—function relationships in the *Aequorea victoria* green fluorescent protein. In *Methods in Cell Biology*, K.F. Sullivan and S.A. Kay, eds (Academic Press), pp. 19-30.
- DeBlasio, S.L., Sylvester, A.W., and Jackson, D.** (2010). Illuminating plant biology: using fluorescent proteins for high-throughput analysis of protein localization and function in plants. *Brief Funct Genomics* **9**, 129-138.
- 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. *The Plant J* **10**, 393-402.
- Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L.** (2003). Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol* **13**, 1768-1774.
- Eshed, Y., Baum, S.F., Perea, J.V., and Bowman, J.L.** (2001). Establishment of polarity in lateral organs of plants. *Curr Biol* **11**, 1251-1260.
- Fahlgren, N., Montgomery, T., Howell, M., Allen, E., Dvorak, S., Alexander, A., and Carrington, J.** (2006). Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in Arabidopsis. *Curr Biol* **16**, 939-944.
- Fang, J., Guo, T., Xie, Z., Chun, Y., Zhao, J., Peng, L., Zafar, S.A., Yuan, S., Xiao, L., and Li, X.** (2021). The URL1-ROC5-TPL2 transcriptional repressor complex represses the ACL1 gene to modulate leaf rolling in rice. *Plant physiol* **185**, 1722-1744.
- Farmer, E., Mousavi, S., and Lenglet, A.** (2013). Leaf numbering for experiments on long distance signalling in Arabidopsis. *Protocol Exchange*.
- Fierro, A.C., Leroux, O., De Coninck, B., Cammue, B.P.A., Marchal, K., Prinsen, E., Van Der Straeten, D., and Vandenbussche, F.** (2015). Ultraviolet-B radiation stimulates downward leaf curling in *Arabidopsis thaliana*. *Plant Physiol and Biochem* **93**, 9-17.
- Govaerts, Y.M., Jacquemoud, S., Verstraete, M.M., and Ustin, S.L.** (1996). Three-dimensional radiation transfer modeling in a dicotyledon leaf. *Appl Opt* **35**, 6585-6598.

- Hanazono, Y., Yu, J.M., Dunbar, C.E., and Emmons, R.V.** (1997). Green fluorescent protein retroviral vectors: low titer and high recombination frequency suggest a selective disadvantage. *Hum Gene Ther* **8**, 1313-1319.
- Heim, R., Prasher, D.C., and Tsien, R.Y.** (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *PNAS USA* **91**, 12501-12504.
- Hopp, T.P., Prickett, K.S., Price, V.L., Libby, R.T., March, C.J., Pat Cerretti, D., Urdal, D.L., and Conlon, P.J.** (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* **6**, 1204-1210.
- Hsiao, T.C., O'Toole, J.C., Yambao, E.B., and Turner, N.C.** (1984). Influence of osmotic adjustment on leaf rolling and tissue death in rice (*Oryza sativa* L.). *Plant Physiol* **75**, 338-341.
- Huang, H., Alvarez, S., Bindbeutel, R., Shen, Z., Naldrett, M.J., Evans, B.S., Briggs, S.P., Hicks, L.M., Kay, S.A., and Nusinow, D.A.** (2016). Identification of evening complex associated proteins in Arabidopsis by affinity purification and mass spectrometry. *Mol Cell Proteomics* **15**, 201-217.
- Huang, W.-Y., Aramburu, J., Douglas, P.S., and Izumo, S.** (2000). Transgenic expression of green fluorescence protein can cause dilated cardiomyopathy. *Nature Medicine* **6**, 482-483.
- Hughes, T.E., Zhang, H., Logothetis, D.E., and Berlot, C.H.** (2001). Visualization of a functional Galpha q-green fluorescent protein fusion in living cells. Association with the plasma membrane is disrupted by mutational activation and by elimination of palmitoylation sites, but not by activation mediated by receptors or ALF4. *J Biol Chem* **276**, 4227-4235.
- 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.
- Hunter, C., Willmann, M.R., Wu, G., Yoshikawa, M., de la Luz Gutiérrez-Nava, M.a., and Poethig, S.R.** (2006). Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in Arabidopsis. *Development* **133**, 2973-2981.
- Iida, H., Yoshida, A., and Takada, S.** (2019). ATML1 activity is restricted to the outermost cells of the embryo through post-transcriptional repressions. *Development* **146**.
- Javelle, M., Klein-Cosson, C., Vernoud, V., Boltz, V., Maher, C., Timmermans, M., Depège-Fargeix, N., and Rogowsky, P.M.** (2011). Genome-wide characterization of the HD-ZIP IV transcription factor family in maize: preferential expression in the epidermis. *Plant Physiol* **157**, 790-803.
- Juarez, M.T., Twigg, R.W., and Timmermans, M.C.P.** (2004). Specification of adaxial cell fate during maize leaf development. *Development* **131**, 4533-4544.

- Kain, S.R., Adams, M., Kondepudi, A., Yang, T.T., Ward, W.W., and Kitts, P.** (1995). Green fluorescent protein as a reporter of gene expression and protein localization. *Biotechniques* **19**, 650-655.
- Kerstetter, R.A., Bollman, K., Taylor, R.A., Bomblies, K., and Poethig, R.S.** (2001). KANADI regulates organ polarity in Arabidopsis. *Nature* **411**, 706-709.
- Ketelaar, T., Anthony, R.G., and Hussey, P.J.** (2004). Green fluorescent protein-mTalin causes defects in actin organization and cell expansion in Arabidopsis and inhibits actin depolymerizing factor's actin depolymerizing activity in vitro. *Plant Physiol* **136**, 3990-3998.
- 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. *The Plant Cell* **26**, 2184-2200.
- Krestel, H.E., Mihaljevic, A.L., Hoffman, D.A., and Schneider, A.** (2004). Neuronal co-expression of EGFP and beta-galactosidase in mice causes neuropathology and premature death. *Neurobiol Dis* **17**, 310-318.
- Leffell, S.M., Mabon, S.A., and Stewart, C.N., Jr.** (1997). Applications of green fluorescent protein in plants. *Biotechniques* **23**, 912-918.
- Lenassi Zupan, A., Trobec, S., Gaberc-Porekar, V., and Menart, V.** (2004). High expression of green fluorescent protein in *Pichia pastoris* leads to formation of fluorescent particles. *Journal of Biotechnology* **109**, 115-122.
- Li, L.-C., Kang, D.-M., Chen, Z.-L., and Qu, L.-J.** (2007). Hormonal regulation of leaf morphogenesis in Arabidopsis. *J Integ Plant Biol* **49**, 75-80.
- Lin, Q., Ohashi, Y., Kato, M., Tsuge, T., Gu, H., Qu, L.-J., and Aoyama, T.** (2015). GLABRA2 directly suppresses basic helix-loop-helix transcription factor genes with diverse functions in root hair development. *The Plant Cell* **27**, 2894.
- Lin, Y., and Schiefelbein, J.** (2001). Embryonic control of epidermal cell patterning in the root and hypocotyl of Arabidopsis. *Development* **128**, 3697-3705.
- Lisenbee, C.S., Karnik, S.K., and Trelease, R.N.** (2003). Overexpression and mislocalization of a tail-anchored GFP redefines the identity of peroxisomal ER. *Traffic* **4**, 491-501.
- Liu, H.S., Jan, M.S., Chou, C.K., Chen, P.H., and Ke, N.J.** (1999). Is green fluorescent protein toxic to the living cells? *Biochem Biophys Res Commun* **260**, 712-717.
- Luesse, D.R., DeBlasio, S.L., and Hangarter, R.P.** (2006). Plastid movement impaired 2, a new gene involved in normal blue-light-induced chloroplast movements in Arabidopsis. *Plant Physiol* **141**, 1328-1337.

- Martoglio, B., and Dobberstein, B.** (1998). Signal sequences: more than just greasy peptides. *Trends Cell Biol* **8**, 410-415.
- 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.
- McConnell, J.R., and Barton, M.K.** (1998). Leaf polarity and meristem formation in *Arabidopsis*. *Development* **125**, 2935-2942.
- McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K.** (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* **411**, 709-713.
- Merelo, P., Paredes, E.B., Heisler, M.G., and Wenkel, S.** (2017). The shady side of leaf development: the role of the REVOLUTA/KANADI1 module in leaf patterning and auxin-mediated growth promotion. *Curr Opin Plant Biol* **35**, 111-116.
- Merelo, P., Ram, H., Caggiano, M.P., Ohno, C., Ott, F., Straub, D., Graeff, M., Cho, S.K., Yang, S.W., Wenkel, S., and Heisler, M.G.** (2016). Regulation of *MIR165/166* by class II and class III homeodomain leucine zipper proteins establishes leaf polarity. *PNAS* **113**, 11973-11978.
- Mockaitis, K., and Estelle, M.** (2008). Auxin receptors and plant development: a new signaling paradigm. *Annu Rev Cell Dev Biol* **24**, 55-80.
- Moon, J., and Hake, S.** (2011). How a leaf gets its shape. *Curr Opin Plant Biol* **14**, 24-30.
- Mukherjee, T., Subedi, B., Khosla, A., Begler, E.M., Stephens, P.M., Warner, A.L., Lerma-Reyes, R., Thompson, K.A., Gunewardena, S., and Schrick, K.** (2022). The START domain mediates *Arabidopsis* *GLABRA2* dimerization and turnover independently of homeodomain DNA binding. *Plant Physiol* **kiac383**.
- Murashige, T., and Skoog, F.** (1962). A revised medium for rapid growth and bio assays with Tobacco tissue cultures. *Physiologia Plantarum* **15**, 473-497.
- 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.
- O'Toole, J.C., and Cruz, R.T.** (1980). Response of leaf water potential, stomatal resistance, and leaf rolling to water stress. *Plant Physiol* **65**, 428-432.

- O'Toole, J.C., Cruz, R.T., and Singh, T.N.** (1979). Leaf rolling and transpiration. *Plant Science Letters* **16**, 111-114.
- 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., Ruberti, I., Morelli, G., and Aoyama, T.** (2002). Entopically additive expression of GLABRA2 alters the frequency and spacing of trichome initiation. *The Plant J* **29**, 359-369.
- Ormö, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., and Remington, S.J.** (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392-1395.
- Overvoorde, P.J., Okushima, Y., Alonso, J.M., Chan, A., Chang, C., Ecker, J.R., Hughes, B., Liu, A., Onodera, C., Quach, H., Smith, A., Yu, G., and Theologis, A.** (2005). Functional genomic analysis of the AUXIN/INDOLE-3-ACETIC ACID gene family members in *Arabidopsis thaliana*. *The Plant Cell* **17**, 3282-3300.
- Patterson, G.H., Knobel, S.M., Sharif, W.D., Kain, S.R., and Piston, D.W.** (1997). Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys J* **73**, 2782-2790.
- Pekker, I., Alvarez, J.P., and Eshed, Y.** (2005). Auxin response factors mediate *Arabidopsis* organ asymmetry via modulation of KANADI activity. *The Plant Cell* **17**, 2899-2910.
- Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J.** (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **111**, 229-233.
- Prigge, M.J., Otsuga, D., Alonso, J.M., Ecker, J.R., Drews, G.N., and Clark, S.E.** (2005). Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *The Plant Cell* **17**, 61-76.
- Reed, J.W.** (2001). Roles and activities of Aux/IAA proteins in *Arabidopsis*. *Trends in Plant Science* **6**, 420-425.
- Rerie, W.G., Feldmann, K.A., and Marks, M.D.** (1994). The GLABRA2 gene encodes a homeodomain protein required for normal trichome development in *Arabidopsis*. *Genes Dev* **8**, 1388-1399.
- Rizzo, M.A., Springer, G.H., Granada, B., and Piston, D.W.** (2004). An improved cyan fluorescent protein variant useful for FRET. *Nat Biotechnol* **22**, 445-449.
- Rombolá-Caldentey, B., Rueda-Romero, P., Iglesias-Fernández, R., Carbonero, P., and Oñate-Sánchez, L.** (2014). *Arabidopsis* DELLA and Two HD-ZIP Transcription Factors

- Regulate GA Signaling in the Epidermis through the L1 Box-Element. *The Plant Cell* **26**, 2905.
- Saglam, A., Kadioğlu, A., Demiralay, M., and Terzi, R.** (2014). Leaf rolling reduces photosynthetic loss in maize under severe drought. *Acta Botanica Croatica* **73**.
- Schrack, 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.
- Sheen, J., Hwang, S., Niwa, Y., Kobayashi, H., and Galbraith, D.W.** (1995). Green-fluorescent protein as a new vital marker in plant cells. *The Plant J* **8**, 777-784.
- Shi, Z., Wang, J., Wan, X., Shen, G.-z., Wang, X.-q., and Zhang, J.-l.** (2006). Over-expression of rice OsAGO7 gene induces upward curling of the leaf blade that enhanced erect-leaf habit. *Planta* **226**, 99-108.
- Shimomura, O., Johnson, F.H., and Saiga, Y.** (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol* **59**, 223-239.
- Siegfried, K.R., Eshed, Y., Baum, S.F., Otsuga, D., Drews, G.N., and Bowman, J.L.** (1999). Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. *Development* **126**, 4117-4128.
- Song, J.B., Huang, S.Q., Dalmay, T., and Yang, Z.M.** (2012). Regulation of leaf morphology by MicroRNA394 and its target LEAF CURLING RESPONSIVENESS. *Plant and Cell Physiol* **53**, 1283-1294.
- Surribas, A., Resina, D., Ferrer, P., and Valero, F.** (2006). Limitations using GFP as a protein expression reporter in *Pichia pastoris*. *Microbial Cell Factories* **5**, P56.
- 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.
- Takada, S., and Jürgens, G.** (2007). Transcriptional regulation of epidermal cell fate in the Arabidopsis embryo. *Development* **134**, 1141-1150.
- Teale, W.D., Paponov, I.A., and Palme, K.** (2006). Auxin in action: signalling, transport and the control of plant growth and development. *Nature Reviews Mol Cell Biol* **7**, 847-859.
- Thomas, C.L., and Maule, A.J.** (2000). Limitations on the use of fused green fluorescent protein to investigate structure-function relationships for the cauliflower mosaic virus movement protein. *J Gen Virol* **81**, 1851-1855.
- Tian, G.W., Mohanty, A., Chary, S.N., Li, S., Paap, B., Drakakaki, G., Kopec, C.D., Li, J., Ehrhardt, D., Jackson, D., Rhee, S.Y., Raikhel, N.V., and Citovsky, V.** (2004). High-

- throughput fluorescent tagging of full-length Arabidopsis gene products in planta. *Plant Physiol* **135**, 25-38.
- Turner, N.C., O'Toole, J.C., Cruz, R.T., Namuco, O.S., and Ahmad, S.** (1986). Responses of seven diverse rice cultivars to water deficits I. Stress development, canopy temperature, leaf rolling and growth. *Field Crops Res* **13**, 257-271.
- Ueda, M., Manabe, Y., and Mukai, M.** (2011). The high performance of 3XFLAG for target purification of a bioactive metabolite: A tag combined with a highly effective linker structure. *Bioorganic & Medicinal Chemistry Letters* **21**, 1359-1362.
- 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.
- Williams, L., Grigg, S.P., Xie, M., Christensen, S., and Fletcher, J.C.** (2005). Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes. *Development* **132**, 3657-3668.
- Wilson, I.A., Niman, H.L., Houghten, R.A., Cherenson, A.R., Connolly, M.L., and Lerner, R.A.** (1984). The structure of an antigenic determinant in a protein. *Cell* **37**, 767-778.
- Wojciechowska, I., Mukherjee, T., Knox-Brown, P., Hu, X., Khosla, A., Mathews, G.L., Thompson, K.A., Peery, S.T., Szlachetko, J., Thalhammer, A., Hinch, D.K., Skirycz, A., and Schrick, K.** (2021). PROTODERMAL FACTOR2 binds lysophosphatidylcholines and transcriptionally regulates phospholipid metabolism. *bioRxiv*, 2021.2010.2020.465175.
- Wu, R., Li, S., He, S., Wassmann, F., Yu, C., Qin, G., Schreiber, L., Qu, L.J., and 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. *The Plant Cell* **23**, 3392-3411.
- Yang, F., Moss, L.G., and Phillips, G.N., Jr.** (1996). The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**, 1246-1251.
- Yang, H., Shi, G., Li, X., Hu, D., Cui, Y., Hou, J., Yu, D., and Huang, F.** (2019). Overexpression of a soybean YABBY gene, GmFILA, causes leaf curling in *Arabidopsis thaliana*. *BMC Plant Biol* **19**, 234-234.
- Yu, H., Chen, X., Hong, Y.-Y., Wang, Y., Xu, P., Ke, S.-D., Liu, H.-Y., Zhu, J.-K., Oliver, D.J., and Xiang, C.-B.** (2008). Activated expression of an Arabidopsis HD-START

protein confers drought tolerance with improved root system and reduced stomatal density *The Plant Cell* **20**, 1134-1151.

Zhang, F., Hackett, N.R., Lam, G., Cheng, J., Pergolizzi, R., Luo, L., Shmelkov, S.V., Edelberg, J., Crystal, R.G., and Rafii, S. (2003). Green fluorescent protein selectively induces HSP70-mediated up-regulation of COX-2 expression in endothelial cells. *Blood* **102**, 2115-2121.

Zhang, G.H., Xu, Q., Zhu, X.D., Qian, Q., and Xue, H.W. (2009). SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. *The Plant Cell* **21**, 719-735.

Zhao, B., Wang, B., Li, Z., Guo, T., Zhao, J., Guan, Z., and Liu, K. (2019). Identification and characterization of a new dwarf locus DS-4 encoding an Aux/IAA7 protein in *Brassica napus*. *Theoretical and Applied Genetics* **132**, 1435-1449.

Zou, L.-p., Sun, X.-h., Zhang, Z.-g., Liu, P., Wu, J.-x., Tian, C.-j., Qiu, J.-l., and Lu, T.-g. (2011). Leaf rolling controlled by the homeodomain leucine zipper class IV gene Roc5 in rice *Plant Physiol* **156**, 1589-1602.

Zumajo-Cardona, C., Vasco, A., and Ambrose, B. (2019). The evolution of the KANADI gene family and leaf development in lycophytes and ferns. *Plants* **8**, 313.

Chapter 5 - The SUMO-interacting motifs of the START domain are required for GLABRA2 transcription factor activity

Abstract

The homeodomain transcription factor GLABRA2 (GL2) is a key regulator in the specification of epidermal cell identity in Arabidopsis. In addition to DNA-binding homeodomain, it consists of a zipper-loop-zipper dimerization domain and a lipid sensing START domain that is also involved in its dimerization and stability. However, only a few motifs or functional elements lying within or outside of these structural domains have been identified to date. SUMO-interacting motifs (SIMs) are short hydrophobic sequences that mediate protein-protein interactions by non-covalently binding to SUMO (small ubiquitin-like modifier) conjugated proteins. We used machine learning tools to predict two high confidence, conserved SIMs that map to the outer surface of the GL2 START domain. Phenotypic analysis of SIM mutants revealed defects in trichome branching. The corresponding EYFP-tagged mutant proteins displayed defects in nuclear localization. Yeast two-hybrid assays revealed protein dimerization defects in SIM mutants, similar to those previously reported for other START domain mutants. Overall, our findings indicate a role for putative SIMs in the subcellular localization and homodimerization of GL2. Further experiments will be required to identify proteins that associate with GL2 through SUMO/SIM interactions.

Keywords: GLABRA2, START domain, SUMO-interacting motifs, Arabidopsis, nuclear localization, Yeast two-hybrid, dimerization

Introduction

GLABRA2 (GL2) is a member of the class IV homeodomain leucine-zipper (HD-Zip IV) transcription factor family that is deeply conserved in the plant lineage from freshwater green algae to the angiosperms (Sakakibara et al., 2001; Zalewski et al., 2013). As a group, the HD-Zip IV transcription factors play critical roles in the differentiation of cell layers during plant development ((Nakamura et al., 2006); reviewed in (Chew et al., 2013)). GL2 is required for the specification of epidermal cell identity in the shoot, root and seed (Rerie et al., 1994; Di Cristina et al., 1996; Western et al., 2001). It is expressed in trichomes from the earliest stage of specification and plays a key role in promoting trichome initiation and morphogenesis (Rerie et al., 1994; Szymanski et al., 1998). In roots, GL2 is preferentially expressed in hairless cell files and functions as a negative regulator of root hair development (Di Cristina et al., 1996; Masucci et al., 1996; Hung et al., 1998). In seeds, it regulates seed coat mucilage biosynthesis (Western et al., 2001; Western et al., 2004) and seed oil production (Shen et al., 2006a; Shi et al., 2012). GL2 is also involved in the regulation of hypocotyl stomatal development (Hung et al., 1998) and anthocyanin accumulation (Wang et al., 2020). The expression of GL2 is transcriptionally regulated by the MYB-bHLH-WD40 (MBW) trans-activating complexes comprised of the R2R3-type MYB transcription factor GL1, MYB23 or WER, the basic helix-loop-helix (bHLH) transcription factor GL3 or EGL3, and the WD40-repeat protein TRANSPARENT TESTA GLABRA1 (TTG1) (Lee and Schiefelbein, 1999; Payne et al., 2000; Bernhardt et al., 2003; Zhang et al., 2003; Morohashi et al., 2007; Wang and Chen, 2008; Zhao et al., 2008).

Similar to other HD-Zip IV members, GL2 contains four domains that include an amino-terminal homeodomain (HD) involved in DNA binding, a plant-specific leucine zipper with an internal loop forming a zipper-loop-zipper (ZLZ), a Steroidogenic Acute Regulatory (StAR)

protein-related lipid Transfer (START) domain involved in lipid sensing, and an uncharacterized carboxy-terminal domain known as the START-associated domain (STAD) (Landschulz et al., 1988; Sessa et al., 1993; Schena and Davis, 1994; Di Cristina et al., 1996; Ponting and Aravind, 1999; Riechmann et al., 2000; Schrick et al., 2004; Mukherjee et al., 2009). Two highly conserved cysteines (CXXC) found within the variable loop of the ZLZ were previously implicated in redox regulation of dimerization and DNA binding of HD-Zip TFs (Ciarbelli et al., 2008). Mutant analysis showed that both the HD and the START domain are required for the transcription factor activity of GL2 (Mukherjee et al., 2022). While the HD is critical for DNA binding, the START domain is required for dimerization. The START domain is also important for protein stability as revealed by a rapid drop of GL2 protein level in START mutants (Mukherjee et al., 2022). Apart from these defined domains and their major structural features, little is known about other motifs involved in GL2 transcription factor function.

Small Ubiquitin-like Modifiers (SUMOs) are ~12 kDa proteins that are post-translationally and covalently attached to a lysine (K) residue in target proteins by a process termed SUMOylation (Melchior, 2000; Hay, 2001; Müller et al., 2001). In contrast, SUMO-interacting motifs (SIMs) mediate non-covalent interactions of proteins with SUMOs (Minty et al., 2000; Song et al., 2004; Hannich et al., 2005; Song et al., 2005; Hecker et al., 2006; Kerscher, 2007; Santiago et al., 2009). SIMs are characterized by the presence of a hydrophobic core sequence of three-four hydrophobic residues (I, V or L) flanked by a variable length stretch of acidic (E, D) and/or phosphorylatable (S, T) residues in flanking sequences (Minty et al., 2000; Song et al., 2004; Hannich et al., 2005; Hecker et al., 2006; Kerscher, 2007; Zhao et al., 2014). The best characterized SIM is V/I/L-x-V/I/L-V/I/L (or V/I/L-V/I/L-x-V/I/L), where x is any amino acid (Song et al., 2004; Hannich et al., 2005; Song et al., 2005; Hecker et al., 2006;

Kerscher, 2007). The hydrophobic core of SIMs adopts an extended β -strand conformation that interacts with the α 1-helix and β 2-strand of SUMO (Song et al., 2005; Hecker et al., 2006; Kerscher, 2007; Chang et al., 2011; Namanja et al., 2012). The adjacent negatively charged residues interact electrostatically with the basic residues of SUMO to dictate the affinity and/or specificity of SIM binding in addition to influencing the orientation of SUMO-SIM interactions (Reverter and Lima, 2005; Song et al., 2005; Hecker et al., 2006; Meulmeester and Melchior, 2008; Chang et al., 2011; Namanja et al., 2012). SIMs can bind to SUMO in parallel or antiparallel orientations (Reverter and Lima, 2005; Song et al., 2005; Gareau and Lima, 2010; Chang et al., 2011; Namanja et al., 2012).

SIMs are present in various proteins including SUMO ligases (E3), transcription factors, and transcriptional corepressors (Song et al., 2004; Hecker et al., 2006; Lin et al., 2006; Du et al., 2010; Liu et al., 2011; Sæther et al., 2011; Kolesar et al., 2012) and are essential for diverse cellular processes such as transcriptional regulation (Lin et al., 2006; Ouyang et al., 2009; Sahoo et al., 2017; Ryu et al., 2020), DNA repair (Guzzo et al., 2012; Poulsen et al., 2013), cancer cell proliferation (Sæther et al., 2011), formation of phase-separated membraneless organelles (Shen et al., 2006b; Banani et al., 2016; Keiten-Schmitz et al., 2021) and host-pathogen interactions (Chang et al., 2010; Arriagada et al., 2011; Cuchet-Lourenço et al., 2011; Gareau et al., 2012; Lukic et al., 2013; Wang et al., 2018; Nair et al., 2020). SIMs are also involved in the formation of E3 ubiquitin ligase complexes (Werner et al., 2012). A family of RING-type Ub ligases termed SUMO-targeted Ub ligases (STUbL) use SIMs to mediate targeting of poly-SUMOylated substrates for ubiquitin-dependent proteasomal degradation (Prudden et al., 2007; Uzunova et al., 2007; Perry et al., 2008; Geoffroy and Hay, 2009; Abed et al., 2011; Elrouby et al., 2013; Erker et al., 2013; Poulsen et al., 2013; Kumar et al., 2017; Sriramachandran et al., 2019). Several

documented SIM-SUMO interactions involve in positive or negative regulation of SUMOylation in SIM-containing proteins (Takahashi et al., 2005; Tatham et al., 2005; Lin et al., 2006; Kerscher, 2007; Knipscheer et al., 2008; Kim et al., 2009; Kim et al., 2010; Kolesar et al., 2012; Zheng et al., 2019). In some cases, SIMs are engaged in intra- or intermolecular SUMO–SIM binding, preventing the interaction of a protein with other proteins (Baba et al., 2005; Steinacher and Schär, 2005).

Although most examples of known SUMO-SIM interactions are from yeast or mammalian cells, several studies indicate important roles for SIMs in plants. A conserved SIM-like motif in the amino-terminus of SCE1, a SUMO conjugating enzyme from *Arabidopsis*, was found to be involved in SUMO1 binding (Mazur et al., 2017). Similarly, the SUMO-modified form of COP1, an E3 ubiquitin ligase that regulates plant growth, interacts with its substrates through their SIMs, leading to ubiquitin-mediated degradation (Holm et al., 2001; Kim et al., 2016; Uljon et al., 2016; Mazur et al., 2019). In plants, SUMO-SIM interactions are also implicated in blocking ubiquitination of target proteins. One example relates to DELLA-mediated growth control (Conti et al., 2014; Nelis et al., 2015). SUMOylated DELLA repressor sequesters the gibberellin receptor GID1 by binding to its SIM, stabilizing the larger pool of unmodified DELLA during stress conditions (Conti et al., 2014; Nelis et al., 2015). In another example, binding of SUMO-conjugated JAZ repressors to the jasmonic acid (JA) receptor COI1 inhibits COI1-mediated degradation of non-SUMOylated JAZ, negatively regulating JA signaling during necrotrophic infections (Srivastava et al., 2018).

In this study, we used machine learning tools to predict several SIMs within the GL2 transcription factor, and selected the top two candidates to test our hypothesis. Structural homology modeling indicates that the START domain of GL2, like that of other family

members, consists of an α/β helix-grip-fold which forms the ligand-binding pocket (Schrick et al., 2014). Both of the predicted SIMs map to the outer surface of the GL2 START domain, and these motifs are conserved in GL2 orthologs. Site-directed mutagenesis of the hydrophobic residues to alanine in predicted SIMs resulted in trichome branching defects and impaired nuclear localization of the EYFP-tagged protein. Using yeast two-hybrid experiments, we failed to detect any GL2-SUMO interactions but unlike wild-type GL2 transcription factors, SIM mutants were defective in dimerization. These results characterize putative SIMs as important motifs regulating the subcellular localization and activity of GL2 in Arabidopsis.

Materials and methods

Plant materials and growth conditions

The plant materials included the wild-type *Arabidopsis thaliana* ecotype Columbia (*Col*), the *gl2-5* null mutant (Khosla et al., 2014) and transgenic lines expressing EYFP- or FLAG-tagged GL2 (Schrick et al., 2014; Subedi and Schrick, 2022) and *gl2*^{SIM353} and *gl2*^{SIM404} mutants in the *gl2-5* background. Seeds were sown on soil containing Berger BM6, vermiculite and perlite (Hummert International) in a ratio of 4:2:1 followed by stratification at 4°C for 5 days and transfer to a growth chamber maintained at 23°C under continuous light.

Generation of constructs and transgenic lines

Construction of *proGL2:EYFP:GL2* and *proGL2:6xHis 3xFLAG:GL2* are described elsewhere (Schrick et al., 2014; Subedi and Schrick, 2022). The SIM353 and SIM404 mutations were introduced in the GL2 cDNA cloned into pENTR/D-TOPO using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs, E0554S) with oligonucleotide primers listed in **Table**

C.1. Mutant *gl2* sequences were transferred into the SR54 binary vectors containing *EYFP* and *6xHis 3xFLAG* tags. The inserts were amplified with gene-specific primers (**Table C.1**) using the Q5 High Fidelity DNA polymerase (New England Biolabs, M0491S), followed by purification of the amplicons using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, 740609.250) and *Dpn1* digestion. The vectors were prepared by restriction digestion of the *proGL2:EYFP:GL2* and *proGL2:6xHis 3xFLAG:GL2* constructs with *SalI* and *KpnI*. DNA fragments of the corresponding genes were assembled with their vectors using NEBuilder HiFi DNA Assembly (New England Biolabs, E5520S) following manufacturer's protocol. The mutations were verified by sequencing before and after cloning of the mutant *gl2* coding sequences into the binary vector. The constructs were transformed into *Agrobacterium tumefaciens* GV3101 (pMP90) followed by transformation of *gl2-5* plants via floral dip (Clough and Bent, 1998). Transgenic lines were selected on agar media with 20 µg/ml Hygromycin B. For each genotype, >20 independent T1 transformants were screened for phenotypic complementation of *gl2-5* and nuclear expression of EYFP. Representative T3 homozygous lines from T2 progeny showing 3:1 segregation of the transgene were confirmed by PCR genotyping and sequencing and were used for further characterization.

Machine learning bioinformatic tools

Joint Analyzer of SUMOylation sites and SIMs (JASSA v4) (Beauclair et al., 2015) was used for identification and scoring of candidate SIMs in GL2. Scoring is based on a Position Frequency Matrix obtained by sequence alignment with experimental SIMs. The GL2 orthologs from various plant species were obtained using BlastP (Altschul et al., 1990). Multiple sequence alignments were performed with Clustal Omega (Madeira et al., 2019) and visualized in Jalview (Waterhouse et al., 2009). Structural models of wild-type and mutant *gl2* START domains were

rendered using I-TASSER (Iterative Threading ASSEmblY Refinement) (Yang and Zhang, 2015a; Yang and Zhang, 2015b; Zhang et al., 2017). The three-dimensional structures were analyzed using RasWin Molecular Graphics (Bernstein, 2009).

Western blotting

Proteins were extracted by homogenization of 1-2 leaves from ~one-month-old T1 plants in liquid nitrogen and addition of 100 µl of hot (95°C) Laemmli buffer (8 M urea, 2% SDS, 0.1 M DTT, 20% glycerol, 0.1 M Tris-HCl pH 6.8, 0.004% bromophenol blue). Proteins were separated in 10% Mini-PROTEAN TGX Precast Protein Gels (Bio-Rad, 4561036) followed by transfer to PVDF membranes. Incubation with anti-FLAG primary antibodies (1:10,000; Sigma, F1804) was followed by goat anti-mouse IgG HRP secondary antibodies (1:3300; GenScript, A00160). Chemiluminescence was detected using the Azure 300 imager (Azure Biosystems) after adding SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific, 34096). Stripping of the membranes was performed by incubating membranes twice with a mild stripping buffer (200 mM glycine, 0.1% (w/v) SDS, 1% (v/v) Tween 20) at pH 2.2 for 5-10 mins. Membranes were washed twice (10 mins) with 1x PBS and 1x TBS-T, blocked and reprobed. Finally, membranes were stained with Bio-Safe Coomassie Blue G-250 (Bio-Rad) to indicate protein loading.

To monitor the expression of bait proteins in yeast, isolated colonies expressing pDEST32 constructs were grown in 2 ml selective media in screw-capped tubes at 30 °C for 3-5. Proteins were extracted by a standard sodium hydroxide/trichloroacetic acid (NaOH/TCA) method and were detected by Western blot as described above using a 1:500 dilution of GAL4 (DBD) (RK5C1) HRP sc-510 mouse monoclonal antibodies (Santa Cruz Biotechnology, 786-003). To monitor the expression of AtSUMO1-3, cells with SUMO bait and prey constructs were

grown in selective media followed by protein extraction as above. Proteins were detected with 1:10,000 dilutions of primary antibodies against AtSUMO1 (Abcam, ab5316 or Agrisera, AS08308) or AtSUMO3 (Abcam, ab5317) followed by a 1:3300 dilution of goat anti-rabbit IgG HRP secondary antibodies (Enzo, ADI-SAB-300-J).

Phenotypic assays and microscopy

Images of rosettes and trichomes, as well as EYFP signals from ~15-d-old plants were captured by a Leica M165 epifluorescence stereomicroscope fitted with a GFP2 filter set and a Leica DFC295 digital camera using Leica Application Suite 4.1. Numbers of trichomes and branching patterns were quantified on first leaves of 20 plants for each genotype. Ordinary one-way ANOVA using Tukey's multiple comparisons test was applied to test significance of differences. The subcellular localization patterns of EYFP-tagged proteins in the primary roots of 4-d-old seedlings were imaged with a Zeiss LSM 880 confocal microscope after staining with propidium iodide (10 µg/ml). To acquire EYFP signals, a 488 nm Argon laser and 500–550 nm emission filter was used. For propidium iodide stain, a 561 nm DPSS laser in combination with 575–645 nm emission filter was used. All images were processed in Adobe Photoshop.

Yeast two-hybrid (Y2H) experiments

AtSUMO1-3 cDNA sequences from pUNI51 plasmids (U17495, U21609, U83833; Arabidopsis Biological Resource Center) were cloned into pENTR/D-TOPO using a pENTR Directional TOPO Cloning Kit (Invitrogen, K2400-20) and confirmed by sequencing. The pENTR/D-TOPO constructs containing cDNAs for AtSUMO1-3, full-length GL2, GL2 START domain, gl2^{SIM353} and gl2^{SIM404} were transferred to pDEST32 and pDEST22 destination vectors (ProQuest Two-Hybrid System, Invitrogen) using Gateway LR Clonase II Enzyme mix (Thermo

Fisher Scientific, 11791-100), and confirmed by sequencing. Oligonucleotides for TOPO cloning and sequence verification are listed in **Table C.1**. A list of constructs is provided in **Table C.2**. Bait and prey constructs were transformed into the haploid yeast strains Y2HGold (*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*) and Y187 (*MATa ura3-52 his3-200 ade2-101 trp1-901 leu2-3,112 gal4Δ gal80Δ met-URA3::GAL1_{UAS}-Gal1_{TATA}-LacZ MEL1*), respectively, using standard LiAc/PEG transformation (Daniel Gietz and Woods, 2002). Empty vectors served as negative controls. The pDEST32 and pDEST2 transformants were selected in media lacking Trp and Leu, respectively. To test for autoactivation of bait constructs, pDEST32 transformants were replica plated to selective media (-Leu-His AbA) to monitor reporter genes *HIS3* and *AUR1-C*. The yeast strains of opposite mating types were mated on nonselective (YPDA) media, and diploids were selected on -Leu-Trp, followed by replica plating on -Leu-Trp-His and -Leu-Trp-His X-gal/AbA to monitor the reporter genes *HIS3*, *AUR1-C*, and *MEL1*. Protein-protein interactions were assayed in both directions. Liquid cultures were grown for 3-5 days to OD600 0.8-1.0, and diluted with water to normal cell densities, followed by four-fold serial dilutions and spotting onto permissive and selective media using an 8 x 6-pin multiplex plating tool (Frogger, Dankar Inc.). Plates were incubated at 30°C for 3-5 days prior to imaging with a Gel Doc XR+ Gel Documentation System (Bio-Rad).

Accession Numbers

See **Table C.3**.

Results

Two conserved SUMO-interacting motifs (SIMs) are predicted in the GL2 START domain

GL2 is a 747 amino acid-protein that likely interacts with other proteins for its nuclear localization, lipid binding or other aspects of its transcription factor activity and regulation. To predict SIMs in GL2, we used the machine learning web servers JASSA (Joined Advanced SUMOylation Site and SIM Analyser) (Beauclair et al., 2015) and GPS-SUMO (Group-based Prediction System-SUMO) (Zhao et al., 2014). The prediction results showed the highest confidence score for the SIM at amino acid position 404-407 (IVDV) followed by the one at position 353-356 (VDVI) both of which mapped to the GL2 START domain (**Figure 5.1A**). Two additional SIMs with considerable predictive scores were predicted in the GL2 STAD domain at amino acid positions 667-670 (IQIL) and 685-688 (PLVI) (**Figure 5.1A**). The hydrophobic patches of the predicted SIMs contain the characteristic three hydrophobic residues (I, V or L), along with a variable residue (**Figure 5.1A, 5.1B and 5.1C**). In case of high confidence SIMs in the START domain, the variable residue was glutamate (E) at the 2nd or 3rd position (**Figure 5.1A**). Both SIMs mapped to predicted β sheets ($\beta 4$ and $\beta 7$, respectively) of the GL2 START domain (**Figure 5.2B**). In 3-D view, the hydrophobic residues in the IVDV patch appear to project from the outer surface of the START domain with the side chains of the negatively charged D facing inward. The sequences surrounding SIM404, the SIM with the highest confidence, additionally showed enrichment for phosphorylatable (S) and acidic (D/E) residues (**Figure 5.1A**).

Next, we asked whether the putative SIMs in GL2 START domain are conserved among the GL2 orthologs from representative plant species and among the 16 HD-Zip IV members from

Arabidopsis. Multiple sequence alignment indicated that SIM404 is conserved in all 15 selected GL2 orthologs (15/15 or ~100%) (**Figure 5.1B**). Similarly, SIM353 was conserved in most (14/15 or ~93%) of the selected GL2 orthologs with one of them showing variation in the SIM type (**Figure 5.1B**). Within the HD-Zip IV family from Arabidopsis, SIM404 was conserved in majority (11/16 or ~69%) of GL2 paralogs (**Figure 5.1C**). Interestingly, majority (11/16 or ~69%) of HD-Zip IV members contained putative SIMs at positions corresponding to GL2 SIM353 but 9 of them were of different SIM types (**Figure 5.1C**). These results indicate that predicted SIMs are highly conserved in HD-Zip IV TFs in a position-specific manner.

GL2 SIMs can be genetically interrogated without affecting the predicted tertiary structure of the START domain

SUMO/SIM complexes display dissociation constants in the micromolar or submicromolar range, and thus mutagenesis of the SIM core hydrophobic amino acids is predicted to significantly destabilize binding (Song et al., 2005; Chang et al., 2011). In accordance with this, mutation of these hydrophobic residues was used as a strategy to characterize SUMO-SIM interactions in several studies (Hecker et al., 2006; Santiago et al., 2009; Kim et al., 2010; Arriagada et al., 2011; Taupitz et al., 2017; Zheng et al., 2019). To determine the importance of predicted SIMs for GL2 activity, we substituted the three conserved hydrophobic residues with alanines in SIM353 (VDVI→ADAA) and SIM404 (IVDV→AADA) to produce two mutant variants in which the SUMO binding motif is removed (**Figure 5.2A**). The overall tertiary structure and lipid binding pocket of the START domain appears to be unaffected by the alanine substitutions as revealed by protein structural modeling using I-TASSER (Yang and Zhang, 2015b) (**Figure 5.2B**). Transgenic Arabidopsis lines were generated by transforming *gl2-5* null mutants with constructs expressing wild-type GL2 protein or the

gl2^{SIM353} or *gl2*^{SIM404} mutant proteins under the *GL2* native promoter. In one set of transgenic lines, the wild-type and mutant proteins were fused to an amino-terminal enhanced yellow fluorescent protein (EYFP), while in another set, the proteins were fused to an amino-terminal 6x His 3x FLAG tag (**Figure 5.2C**).

To confirm expression of the mutant proteins in T1 transformants, we performed Western blot analysis of leaf samples from the FLAG-tagged lines. Similar to wild-type plants (*FLAG:GL2*), most of the plants carrying SIM mutant constructs yielded signals in blot images (**Figure 2D**). The results indicate that the *gl2*^{SIM353} and *gl2*^{SIM404} mutations do not interfere with protein expression.

Mutations affecting the GL2 SIMs cause defects in epidermal differentiation

Deletion of the GL2 HD or START domain or missense mutations in their key residues were previously shown to result in leaf trichome developmental defect similar to that seen in *gl2-5* null mutant (Schrack et al., 2014; Mukherjee et al., 2022). Therefore, we used the trichome phenotype, which is readily visible and quantifiable, to assess GL2 transcription factor activity in our study. Screening of >20 T1 transformants revealed that both the *gl2*^{SIM353} and *gl2*^{SIM404} constructs failed to rescue the trichome defects of *gl2-5* mutants (**Figure 5.3A and 5.3B**). In contrast, numerous T1 transformants corresponding to the wild-type *GL2* construct displayed normal trichome development. Overall, 33.3% (n=42) of *EYFP:GL2* and 60.7% (n=28) of *FLAG:GL2* transformants showed phenotypic rescue of the *gl2-5* null phenotype (**Figure 5.3B**). For the SIM353 mutation, 19.4% (n=31) of *EYFP:gl2*^{SIM353} and 37.0% (n=27) of *FLAG:gl2*^{SIM353} displayed partial rescue of trichome defects (**Figure 5.3B**). The SIM404 mutants exhibited a more severe phenotype, with only 4.2% (n=24) of *EYFP:gl2*^{SIM404} showing partial rescue while none (0%) (n=26) of the *FLAG:gl2*^{SIM404} transformants exhibited rescue (**Figure**

5.3B). Strikingly, the glabrous phenotypes resulting from the loss of trichomes in either of the SIM mutants were indistinguishable from those of *gl2-5* null mutants (**Figure 5.3A**).

We quantified trichomes on the first leaves of plants that were homozygous for the wild-type or mutant transgenes. The results confirmed the reduction in total trichome number as well as the trichome branching defects in both of the SIM mutants (**Figure 5.3C**). The leaves of wild-type plants harbored trichomes with mostly three branches, some with four branches and a few with five or six branches (**Figure 5.3C and 5.3D**). However, SIM mutants mostly produced either unbranched or double branched trichomes with only a few trichomes showing more branches (**Figure 5.3C and 5.3D**). Compared to *gl2^{SIM353}* mutants, *gl2^{SIM404}* mutants displayed more severe phenotypic defect, especially with regard to the number of trichomes with two or more branches (**Figure 5.3C**). These results indicate the critical roles of these putative GL2 SIMs in epidermal differentiation of leaves.

Nuclear localization of GL2 is impaired in SIM mutants

Transcription factors are translated in the cytoplasm of cells but they are localized to the nucleus to bind DNA elements upstream of target genes. In addition, their levels also dictate the extent of regulation of gene expression (Subedi and Schrick, 2022). When we examined ~2-week-old T1 transformants for the expression of EYFP using an epifluorescence microscope, a large proportion (67.74%) of SIM353 mutants (n=31) showed no detectable expression and the remainder showed either weak (19.35%), medium (6.45%) or strong (6.45%) expression (**Figure 5.4A**). Similarly, for SIM404 transformants, 66.67% (n=24) displayed no detectable EYFP expression, whereas 20.83% and 12.50% showed weak and medium expression, respectively. None of the SIM404 mutants showed strong expression (**Figure 5.4A**). These results contrast with the statistics for EYFP:GL2 wild-type transformants (n=42) in which only 21.43% of the

plants had no detectable EYFP expression (**Figure 5.4A**). The proportions of EYFP:GL2 T1 transformants with weak, medium and strong expression were 33.33%, 16.67% and 28.57% respectively (**Figure 5.4A**).

Previous studies uncovered a role for SIMs in the subcellular localization of SIM-containing proteins (Lin et al., 2006; Cuchet-Lourenço et al., 2011; Zheng et al., 2019). When we used epifluorescence microscopy to assay for expression of EYFP-tagged proteins in the leaves of T1 transformants or T3 homozygous lines, it appeared to be non-nuclear in both SIM353 and SIM404 mutants (**Figure 5.4B**). To further investigate subcellular localization of the EYFP-tagged mutant proteins, we stained the roots of 4-d-old seedlings with propidium iodide and imaged them using confocal laser scanning microscopy. Here, *gl2-5* null mutant seedlings were used as negative controls for EYFP expression (**Figure 5.4C**). The results showed that the wild-type GL2 protein was localized exclusively in the nucleus in non-root hair cell files of the root, as previously reported (Subedi and Schrick, 2022). In contrast, both the SIM353 and SIM404 mutant proteins were uniformly distributed throughout the cells from alternating non-root hair cell files (**Figure 5.4C**). These findings indicate the involvement of these putative SIMs in the proper subcellular localization of GL2 to the nucleus.

GL2 SIM-SUMO interactions are not detected in Y2H assays

The *Arabidopsis* genome encodes eight SUMO isoforms (AtSUMO1-8) although only AtSUMO1, 2, 3, and 5 appear to be expressed under standard growth conditions (Kurepa et al., 2003; Novatchkova et al., 2004; Castaño-Miquel et al., 2011; Hammoudi et al., 2016; Benlloch and Lois, 2018). AtSUMO1 and AtSUMO2 are highly expressed isoforms sharing 83% amino acid sequence identity. In comparison, AtSUMO3 and AtSUMO5 share only 42% and 30% sequence identity with AtSUMO1, respectively ((Kurepa et al., 2003; Castaño-Miquel et

al., 2011; Hammoudi et al., 2016); **Figure C.1**). To test whether GL2, via its SIM motifs, can physically interact with *Arabidopsis* SUMO proteins, we performed yeast two-hybrid (Y2H) assays. Full-length GL2 and the GL2 START domain were fused to the yeast GAL4 DNA-binding domain (DBD) in bait plasmids whereas AtSUMO1, AtSUMO2, and AtSUMO3 were fused to the GAL4 activation domain (AD) in prey plasmids. The assays were repeated in the opposite direction in which AtSUMO1-3 served as baits and GL2 and GL2 START served as preys (**Figure 5.5A, Figure C.2**). Assaying yeast cells with both bait and prey plasmids for the reporter gene activity indicated no interaction between GL2 or the GL2 START domain with either of the SUMO isoforms (**Figure 5.5A, Figure C.2**). Given these negative results, we asked whether the *Arabidopsis* SUMO isoforms were adequately expressed in the yeast cells. Western blot analysis using antibodies against GAL4-DBD, AtSUMO1/2 and AtSUMO3 revealed very strong expression of AtSUMO2 but no detectable expression for AtSUMO1 or AtSUMO3 (**Figure 5.5B and 5.5C, Figure C.3**). In contrast, the proteins for full-length GL2 and the GL2 START domain were readily detected by Western blot (**Figure 5.5B**). These results suggest that GL2 does not interact with AtSUMO2 in the yeast system. The negative results for AtSUMO1 and AtSUMO3 can be explained by their lack of expression in the Y2H assays.

Homodimerization of GL2 is impaired in SIM mutants

In HD-Zip IV transcription factors, both the ZLZ domain and the START domain are implicated in dimerization (Nagata and Abe, 2021; Mukherjee et al., 2022). Deletion of and certain missense mutations in the GL2 START were previously shown to result in homodimerization defects (Mukherjee et al., 2022). To determine whether the gl2^{SIM} mutants are proficient in dimerization, we tested for interaction with wild-type GL2 in both direct directions in Y2H assays. In contrast to wild-type GL2, both gl2^{SIM353} and gl2^{SIM404} mutants

failed to interact with GL2 (**Figure 5.5A** and **Figure C.2**). These results suggest that mutations in the conserved hydrophobic core of either SIM abolish the homodimerization of GL2. As indicated by Western blot analysis of the bait proteins with anti-GAL4 DBD antibodies, the SIM mutant proteins were expressed at levels comparable to that of wild-type GL2 (**Figure 5.5B**). Therefore, the loss of dimerization in the SIM mutants is not due to the inability of gl2^{SIM353} and gl2^{SIM404} to be expressed in yeast cells. BBait autoactivation assays confirmed that none of the proteins that we tested in our Y2H experiments, including wild-type GL2, activated reporter gene expression when fused to the GAL4 DBD in the absence of the prey plasmid (**Figure C.4**).

Discussion

In this study, we used machine learning tools to probe the GL2 amino acid sequence for putative SIMs. We focused on the top two high confidence hits, experimentally assessing the role of these predicted motifs in various aspects of GL2 function. Because of the complex features of SIMs, their verification is a challenging task. Moreover, we cannot completely rule out the possibility of other functional sequences deviating from the consensus motifs as not all SIMs match the consensus sequence (Yau et al., 2021). The presence of hydrophobic amino acids in the two predicted GL2 SIMs (SIM353 and SIM404) and their 3D positions on the outer surface of the START domain is consistent with the prediction that these motifs bind SUMO protein.

Although the primary determinant of SUMO-SIM interactions is the core hydrophobic consensus sequence, the flanking amino acid composition also influences SUMO binding (Beauclair et al., 2015). The over-representation of potential phosphorylation sites in adjacent regions of the GL2 SIMs suggests the regulation of SUMO-interacting activity by phosphorylation. Phosphorylation was previously implicated in the regulation of SUMOylation

efficiency and affinity of SIM binding to SUMO (Hecker et al., 2006; Chang et al., 2011; Cappadocia et al., 2015). Thus, genetic or structural analyses of these residues will be useful for further characterization of SIM-mediated regulation of GL2. Intriguingly, most of the proteins undergoing SUMOylation also contain SIMs (Song et al., 2004; Kerscher, 2007). In addition to SIMs, multiple SUMOylation sites are also predicted in GL2 (unpublished data). One possible scenario is that covalent SUMO modification and noncovalent SUMO binding mutually regulate each other to control GL2 transcription factor activity.

In our mutational analysis, both $gl2^{SIM353}$ (VDVI→ADAA) and $gl2^{SIM404}$ (IVDV→AADA) mutants displayed either complete or partial loss of trichomes in addition to branching defects. Phenotypic outcomes of SIM mutations were similar in transgenic lines generated with EYFP and FLAG tags, strengthening our data on the effects of these mutations in epidermal differentiation. Furthermore, the EYFP-tagged mutant proteins were also examined for nuclear versus cytoplasmic localization in comparison to wild-type EYFP:GL2, which is strongly nuclear localized (**Figure 5.4C**). Live imaging using confocal laser scanning microscopy indicated the defective subcellular partitioning of the $gl2^{SIM353}$ and $gl2^{SIM404}$ mutants to both the cytoplasm and the nucleus (**Figure 5.4C**). In addition, it is possible that the mutant proteins are abnormally localized to the plasma membrane. Protein fractionation studies could help to clarify the proportion of nuclear, cytoplasmic and plasma membrane localization relative to wild type.

Similar to other nuclear proteins, transcription factors often contain nuclear localization signals (NLSs) that mediate their import into the nucleus (Kalderon et al., 1984; Robbins et al., 1991; Moroianu et al., 1995; Lowe et al., 2015; Bernhofer et al., 2018). Previous studies implicate SUMO-SIM interactions in the nuclear import of proteins via nuclear pore complexes

or importins (Mahajan et al., 1997; Matunis et al., 1998; Lin et al., 2006; Cuchet-Lourenço et al., 2011; Wang et al., 2018; Zheng et al., 2019). Thus, the defective nuclear localization of $gl2^{SIM353}$ and $gl2^{SIM404}$ mutants may reflect their inability to bind to adaptor proteins such as importin α . Importantly, several studies indicate that SUMO-interacting proteins function predominantly in the nucleus, playing important roles in transcriptional regulation, protein sub-nuclear localization and nuclear body formation (Lin et al., 2006; Ouyang et al., 2009; Sahoo et al., 2017; Ryu et al., 2020). Based on our structural modeling of GL2 START domain, mutation of the hydrophobic core residues to alanines does not disrupt the structural integrity of the GL2 START domain (**Figure 5.2B**). Nonetheless, it is possible that the mutations result in misfolding of the protein, leading to inactive protein that is mislocalized. However, misfolded proteins are usually targeted for degradation. The relatively high levels of expression of the $gl2^{SIM353}$ and $gl2^{SIM404}$ mutant proteins, together with the partial phenotypes observed for $gl2^{SIM353}$ argue against the possibility that the loss of activity is due to protein misfolding. Additionally, the intensity of phenotypic defect also correlated with the prediction scores (**Figures 5.1A, 5.3A, 5.3B and 5.3C**), supporting the hypothesis that these mutations only affect SIM-mediated functions of GL2.

Based on our phenotypic and nuclear localization data, we hypothesize that SIMs play critical roles in mediating protein interactions required for GL2 transcription factor activity. To identify SUMO proteins that potentially interact with GL2 SIMs, we conducted Y2H assays. However, our experiments failed to show any interactions between GL2 and the *Arabidopsis* SUMO paralogs, AtSUMO1, AtSUMO2 and AtSUMO3 (**Figures 5.5A and Figure C.2**). While AtSUMO2 was expressed at high levels in yeast, AtSUMO1 expression was not detectable (**Figures 5.5B**), despite the high DNA sequence identity (89%) between the two isoforms (Kurepa et al., 2003). It is possible that yeast SUMO interfered with the expression of AtSUMO1

and AtSUMO3 in yeast. Previously it was found that in *Arabidopsis*, the *sumo1;sumo2* double mutant is lethal. Each single mutant is indistinguishable from wild type, suggesting highly redundant functions of SUMO1 and SUMO2 (Saracco et al., 2007). Thus, GL2 would have likely interacted with one isoform had it interacted with the other. Another consideration is that SUMO-SIM interactions are of low affinity with dissociation constants of 1–100 μ M (Keusekotten et al., 2013; Taupitz et al., 2017). Therefore, interactions may be difficult to detect by traditional protein-protein interaction assays. It is also possible that GL2 SIM-SUMO interactions only occur when the SUMOs are conjugated to their respective substrates. Even under physiologically relevant conditions in plants, the SUMO-SIM interactions may be difficult to study because SUMOylation is often infrequent, transient, and reversible due to the action of SUMO proteases (Johnson, 2004; Tempé et al., 2008). Relevant to our study, such limitations might be overcome by employing more sensitive tools such as SUMO-ID based on Split-TurboID (Barroso-Gomila et al., 2021) or nuclear magnetic resonance (NMR) spectroscopy (Song et al., 2004; Song et al., 2005). Alternatively, incorporation of genetically encoded crosslinker close to the SIM-interaction groove in candidate SUMOs followed by isolation and purification of the covalently trapped complex could allow the identification of GL2 as a binding partner (Taupitz et al., 2017).

Mutations in $gl2^{SIM353}$ and $gl2^{SIM404}$ resulted in the failure of the proteins to homodimerize in Y2H assays (**Figures 5.5A** and **Figure C.2**). Like wild-type GL2, the SIM mutant proteins were expressed at levels sufficient to allow detection of protein-protein interactions (**Figures 5.5B**). Thus, the negative results cannot be explained by the lack of protein expression. So far, all tested START domain mutations that impair GL2 function result in dimerization defects (Mukherjee et al., 2022). Thus, dimerization defects of the $gl2^{SIM}$ mutants

are not surprising. It is possible that introducing three point mutations in each SIM might have altered the dimerization property of GL2. Further experiments will be required to confirm whether the dimerization defects in $gl2^{SIM}$ are due to direct effect of amino acid substitutions or via SUMO.

In conclusion, this work identified SIMs as amino acid motifs that play important roles in the function of GL2 in differentiation of trichome cells of the epidermis. We showed that two highly conserved putative SIMs in GL2 ($gl2^{SIM353}$ and $gl2^{SIM404}$) are critical for its nuclear localization, and that their mutation abolishes homodimerization of the GL2 transcription factor. These results expand our current understanding of the regulation of GL2 and other developmentally important HD-Zip IV transcription factors that additionally play vital roles in response to environmental stress in agricultural crops ((Cao et al., 2009; Zhang et al., 2019; Wang et al., 2022); reviewed in (Sharif et al., 2021)). It will be important to pursue structural studies to further assess the role of these predicted regions as bona fide SUMO binding motifs in mediating SUMO interactions, or other possible protein-protein interactions involving GL2.

Figures

A

Amino acid position	Domain	Sequence	SIM type	Acidic/Ser stretch	Predictive score (PS)
404-407	START	QLSPEKWAIVDVSVSVEDSN	β	[N] [SIM] [Y]	4.191
353-356	START	CLISKAATVDVIRQGEGPSR	α	[N] [SIM] [N]	0.665
667-670	STAD	LAGHDPSNIQILPSGFSIIP	2	[N] [SIM] [N]	0.535
685-688	STAD	IPDGVESRPLVITSTQDDR	4	[N] [SIM] [N]	0.222
702-705	STAD	DRNSQGGSLTLALQTLINP	1	[N] [SIM] [N]	0.053
730-733	STAD	ESVTNLVSVTLHNIKRS	1	[Y] [SIM] [N]	0.044
630-633	STAD	TVKSREKSIWVLQDSSTNSY	2	[Y] [SIM] [Y]	0.002
562-565	STAD	LHDPGEPTGVIVCASSLWL	4	[N] [SIM] [Y]	0.001
643-646	STAD	DSSTNSYESVVVYAPVDINT	4	[Y] [SIM] [N]	0.001
656-659	STAD	APVDINTTQLVLAGHDPSNI	4	[N] [SIM] [N]	0.001

B

		GL2 orthologs					
		START					
		359G	368A	378L	388V	398S	408S
<i>Z.mays</i> /1-655	AAT	LEVLS	STGVAGN	- - YNGAL	- - - - -	- - - - -	- - - - -
<i>O.sativa</i> /1-779	AAT	LEMIS	SNR - EDDG	- RDGVL	QLMYAEL	QTLTP	MPVPTRELYFARYCKKLAAERWA
<i>B.napus</i> /1-750	AVT	VDV	RQG - EGPSR	IDGAI	QLMFGEM	QQLTP	VVPTRVYFVRSCRQLSPEKWW
A.thaliana /1-747	AAT	VDV	RQG - EGPSR	IDGAI	QLMFGEM	QQLTP	VVPTRVYFVRSCRQLSPEKWA
<i>C.sativa</i> /1-744	AAT	VDV	RQG - EGPSR	IDGAI	QLMFGEM	QQLTP	VVPTRVYFVRSCRQLTPEKWA
<i>C.sativus</i> /1-781	AAT	VDV	CNG - EAAKWN	NGAVQL	MFAEVQML	TPLVP	TREMYFIRHCKQLDAEQNA
<i>H.annuus</i> /1-749	AAT	LDV	CNG - EGPNR	- KGAVQL	MFAELQML	TPLVAT	REYFVRYSKQLTAEKWA
<i>P.dactylifera</i> /1-744	AST	IDV	YNE - NGE - SGDGT	IQLMFAEL	QMLTP	IVPTRE	FYFVRYSKQLSPEKWA
<i>N.tabacum</i> /1-767	AAT	VDV	CNG - EGANC	WDGAVQL	MFAEVQML	TPVVG	TREYFVRYSKQLSVGQNA
<i>S.tuberosum</i> /1-772	AAT	VDV	CNGTEGANS	WDGAVQL	MFAEVQML	TPVVG	TREYFVRYSKQMSAAQAG
<i>A.hypogaea</i> /1-823	AAT	VDV	CNG - EGNSR	- NGAVQL	MFAEVQML	TPMVPT	REYFVRYSKQLSSEQNA
<i>G.hirsutum</i> /1-753	AAT	VDV	CHG - EAPNK	- NGAVQL	MFAELQML	TPLVP	TREYFVRYSKQLSAEQNA
<i>M.domestica</i> /1-761	AAT	VDV	NNG - EGDDR	- NGAVQL	MFAELQML	TPLVP	TREYFVRYSKQLSPEQNA
<i>G.max</i> /1-751	AAT	VDV	CNG - EGPR	- NGAVQL	MFAELQML	TPMVPT	REYFVRYSKQLSAEQNA
<i>J.regia</i> /1-754	AAT	VDV	SIG - EGPNG	- NGAVQL	MFAELQML	TPMVPT	REYFVRYSKQLSAEQNA
<i>H.brasiliensis</i> /1-759	AAT	VDV	SNG - EGPNK	- NGAVQL	MFAEVQML	TPMVPT	REYFVRYSKQLSAEKWA

C

		GL2 paralogs					
		START					
		354D	364R	374G	384P	394C	403A
<i>HDG8</i> /1-699	ANT	IHLV	GSGLP	IRGNCNV	LQVMVEQLH	ILSPLVP	AREFMVVRCCQE
<i>HDG9</i> /1-718	AKT	IHLV	DSVDHRG	- - KTFSRV	IYEQLH	ILSPLVP	PPREFMILRTCQQI
<i>HDG10</i> /1-708	AKT	IHLV	DSMDHPR	- - QTFSRV	IYEQLH	ILSPLVP	LPREFIILRTCQQM
<i>HDG4</i> /1-709	AKTAQ	I	SSGASG	- - PSGLTLL	MFAELQV	SPLVPTRE	AYFLRYVEQNAEEGKMMVDFPID
<i>HDG5</i> /1-826	AKT	VQI	SSGVSG	- - ASGSLLL	MFAELQV	SPLVPTRE	AYFLRYVEQNAETGNNA
GL2 /1-747	AAT	VDV	RQGEGPSR	IDGAI	QLMFGEM	QQLTP	VVPTRVYFVRSCRQL - SPEKWA
<i>FWA</i> /1-686	AST	HKVL	STSGG	- TKSGL	LQQIQAEFQV	ISPLVP	KRVTFIRYCKEI
<i>HDG7</i> /1-682	AST	LEV	SNGSDG	- SRNGS	ILLMQAEFQV	MSPLVP	IKQKKFLRYCKQH
<i>ANL2</i> /1-802	ATTTDV	I	SSGMAG	- TINGAL	QLMNAELQV	SPLVPVRNVN	FLRFCKQH
<i>HDG1</i> /1-808	TSTTE	I	SSGMGG	- - RNGALH	LMHAELQV	SPLVPVRQVS	FLRFCKQH
<i>HDG11</i> /1-722	SKT	LAV	SSGMGG	- THEGAL	HLLYEEMV	LSPLVATRE	FCVLRYSQQT
<i>HDG12</i> /1-687	SKT	LAV	SSGLRG	- NHGDAL	HLMIIEELQV	SPLVTTRE	FCVLRYSQQT
<i>HDG3</i> /1-725	ARTHEQ	I	MADAAG	- NFNGNL	QIMSAYQV	LSPLVTTRE	SFVRYSKQ
<i>HDG2</i> /1-721	AMT	LAV	STGVAG	- NYNGAL	QVMSAEFQV	PSPLVPTRE	TYFARYCKQ
<i>PDF2</i> /1-743	ALT	LEV	STGVAG	- NYNGAL	QVMTAEFQV	PSPLVPTRE	NYFVRYSKQ
<i>ATML1</i> /1-762	ALT	LEV	STGVAG	- NYNGAL	QVMTAEFQV	PSPLVPTRE	NYFVRYSKQ

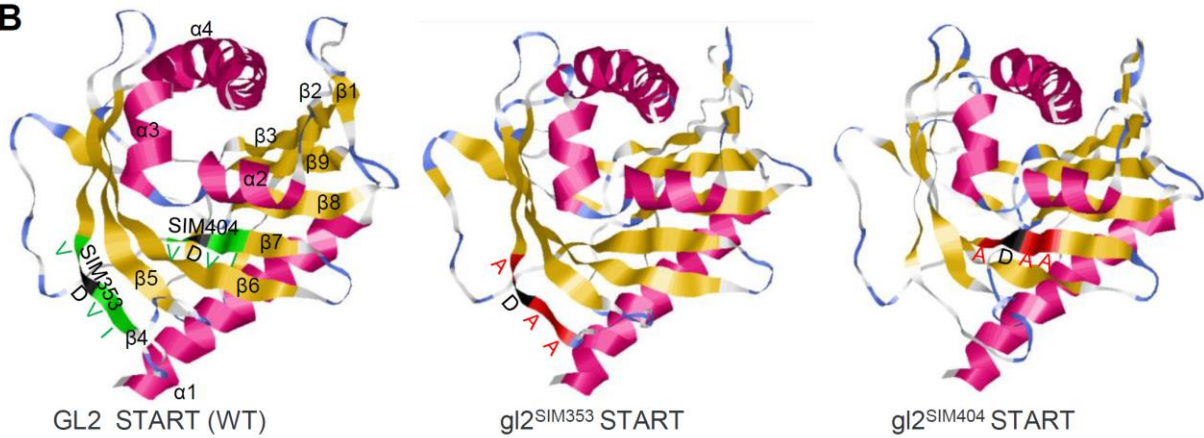
Figure 5.1. GL2 contains multiple potential SUMO-interacting motifs (SIMs) that are conserved in HD-Zip IV family members.

(A) Identification and scoring of candidate SIMs (green) in GL2 using JASSA. The presence or absence of acidic and/or serine residues flanking the putative SIMs are indicated by ‘Y’ or ‘N’, respectively. **(B)** Amino acid alignment of GL2 from *Arabidopsis thaliana* with homologous proteins from *Zea mays*, *Oryza sativa*, *Brassica napus*, *Camelina sativa*, *Cucumis sativus*, *Helianthus annuus*, *Phoenix dactylifera*, *Nicotiana tabacum*, *Solanum tuberosum*, *Arachis hypogaea*, *Gossypium hirsutum*, *Malus domestica*, *Glycine max*, *Juglans regia* and *Hevea brasiliensis*. **(C)** Amino acid alignment of all 16 members of the HD-Zip IV family from *Arabidopsis thaliana*. The conserved SIMs corresponding to positions 353-556 and 404-407 of GL2 are highlighted. Blue, [V/I]-[x]-[V/I]-[V/I]; purple, [V/I]-[V/I]-[x]-[V/I/L]; grey, other SIM types.

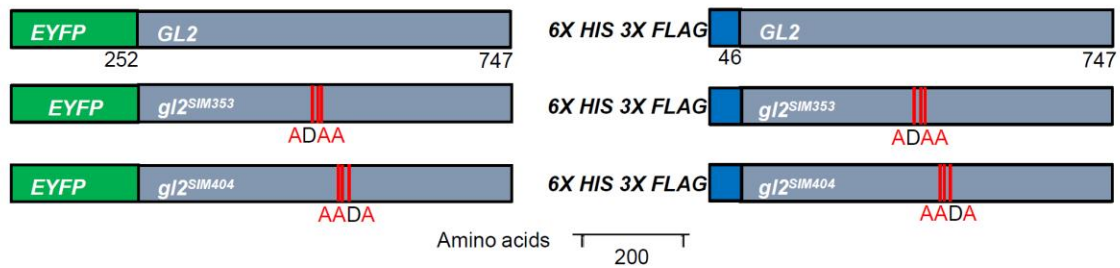
A

SIM position (aa)	GL2	gl2 ^{SIM353}	gl2 ^{SIM404}
353-356	VDVI	ADAA	VDVI
404-407	IVDV	IVDV	AADA

B



C



D

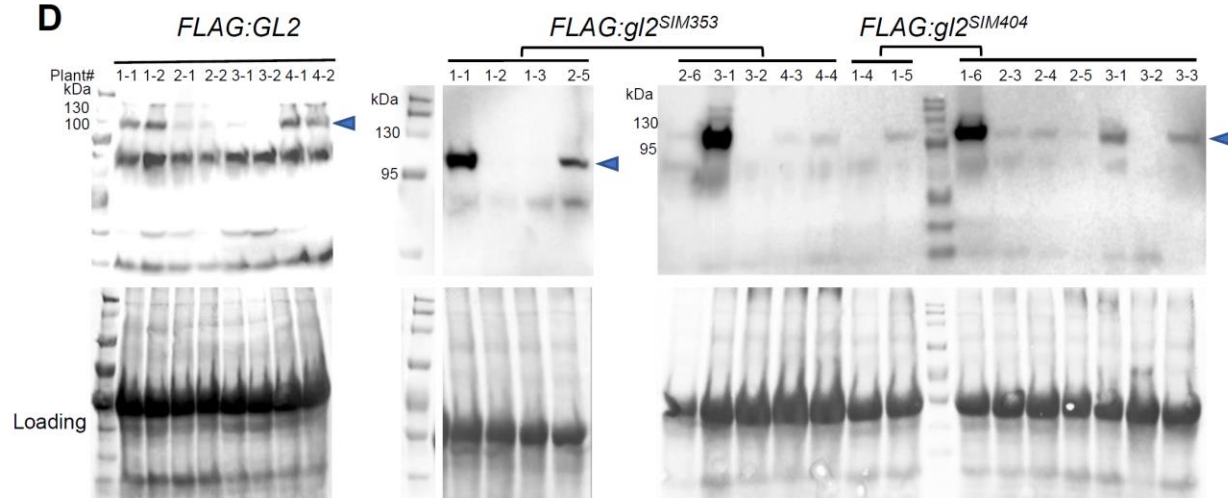


Figure 5.2. Mutant analysis of predicted SIMs in the GL2 START domain.

(A) Core sequence of wild-type and mutant SIMs. Alanine substitutions introduced in each SIM are shown in red. **(B)** Predicted structures of wild-type and mutant GL2 START domains. Secondary structural elements are depicted in ribbon diagrams. Amino acid residues and alanine substitutions within the SIMs are color coded (green, hydrophobic residues; black, aspartate; red; alanine). **(C)** Schematic of constructs (drawn to scale) showing EYFP- and FLAG-tagged $gl2^{SIM353}$ and $gl2^{SIM404}$ proteins expressed under the GL2 native promoter. **(D)** Western blot analysis of leaves from T1 progeny indicate expression of the FLAG-tagged GL2, $gl2^{SIM353}$ and $gl2^{SIM404}$ proteins.

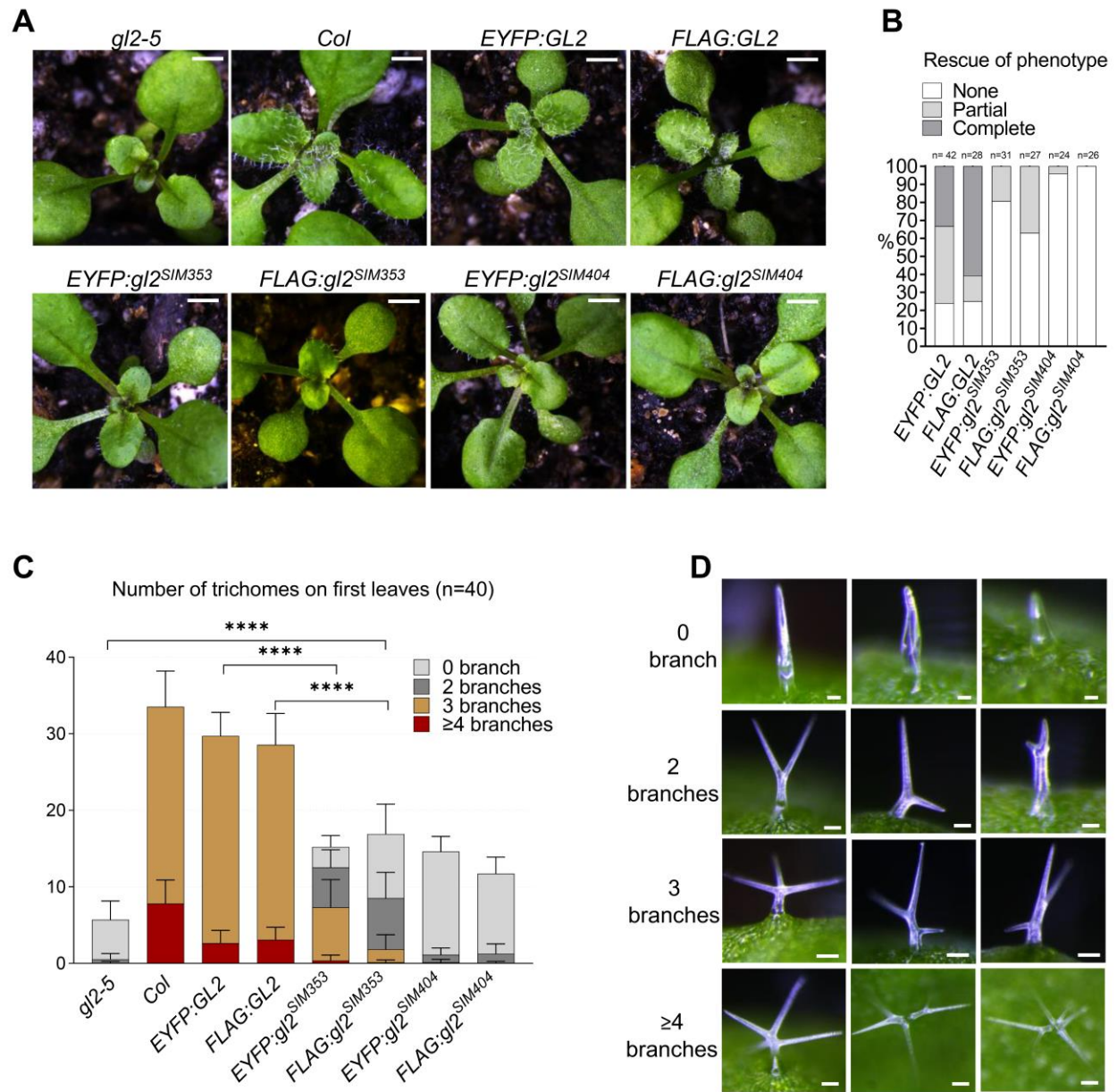


Figure 5.3. Phenotypic characterization of $gl2^{SIM353}$ and $gl2^{SIM404}$ mutants.

(A) Rosettes of 15-d-old homozygous plants of indicated genotypes. The SIM mutants display defects in leaf trichome differentiation. Size bar = 2 mm. **(B)** Graph showing extent of phenotypic complementation in T1 progeny of plants stably transformed with indicated constructs. **(C)** Quantification of trichome defects in SIM mutants: Total numbers of trichomes and numbers of branches on first leaves. Values are mean $\pm$ SD for $n = 20$ plants. Significant

differences between genotypes were determined using ordinary one-way ANOVA with Tukey's multiple comparisons test (**** $p < 0.0001$). **(D)** Examples of trichome branching patterns quantified in **(C)**. Size bar = 100 μm .

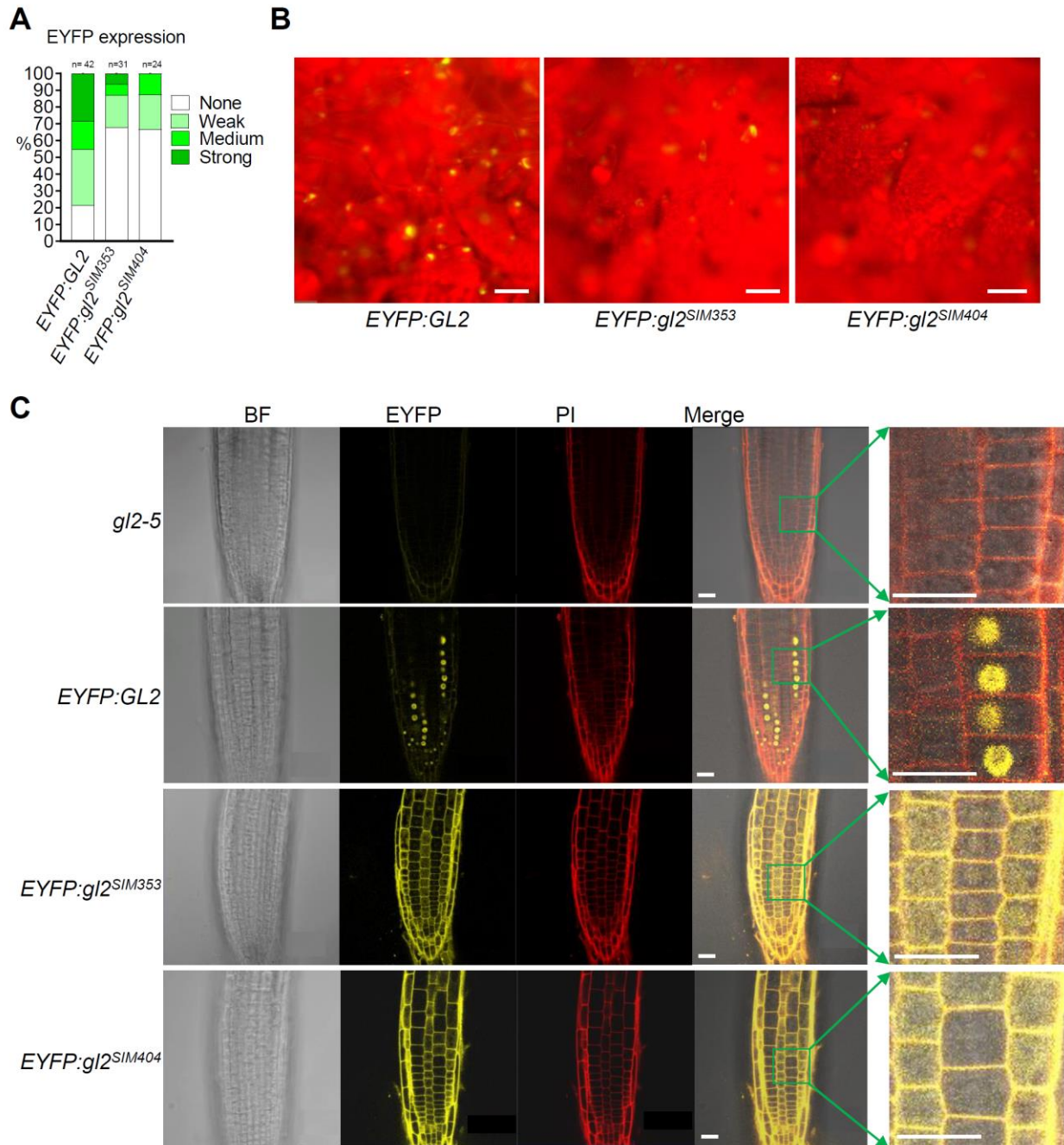


Figure 5.4. Expression and subcellular localization of EYFP-tagged GL2 wild-type and $gl2^{SIM}$ mutant proteins.

(A) Percentage of T1 lines stably transformed with the indicated constructs having strong, medium, weak or no EYFP expression. (B) Epifluorescence microscopy of EYFP expression in leaf trichomes from ~15-d-old homozygous plants (EYFP, yellow; chlorophyll,

red). Size bar = 100 μm . **(C)** Confocal laser scanning microscopy shows subcellular localization of EYFP-tagged proteins (green) in the primary root of 4-d-old homozygous seedlings stained with propidium iodide (red). Magnified views of merged images (right) indicate the presence of EYFP-tagged proteins in alternating cell files. In contrast to nuclear localization for wild-type GL2, the mutant $\text{gl2}^{\text{SIM353}}$ and $\text{gl2}^{\text{SIM404}}$ proteins appear diffuse throughout the cells, displaying both nuclear and non-nuclear expression. Size bar = 20 μm .

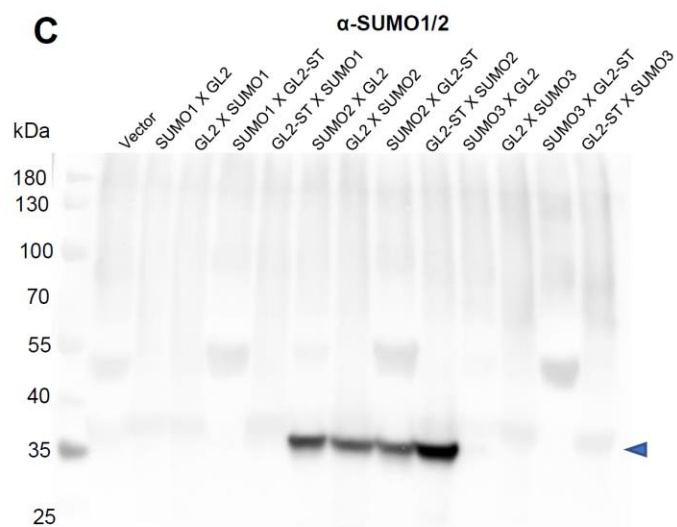
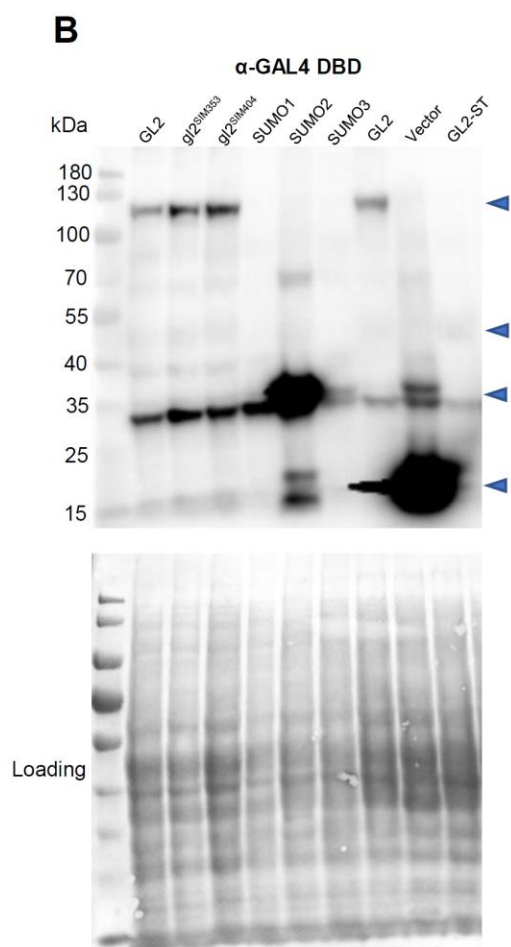
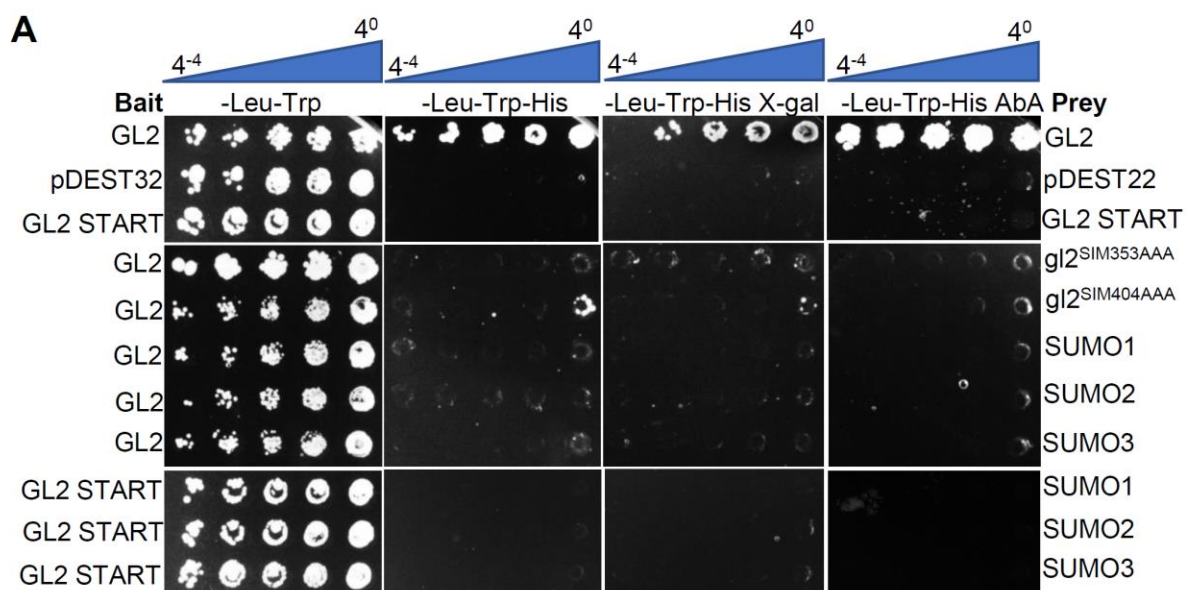


Figure 5.5. Detection of GL2 protein interactions mediated by its predicted SIMs.

(A) Y2H screening of diploid yeast cells expressing bait constructs with full-length GL2 or the GL2 START domain and prey constructs with gl2^{SIM} mutants or SUMO isoforms on permissive versus selective (-Leu-Trp-His and -Leu-Trp-His+X-gal/AbA) media. Unlike wild-type GL2, gl2^{SIM} mutants failed to exhibit homodimerization. No interactions were detected between full-length GL2 or the GL2 START domain and any of the AtSUMO isoforms (1-3). Empty prey and bait vectors represent negative controls. Leu, Leucine; Trp, Tryptophan; His, Histidine; AbA, Aureobasidin A. **(B)** Detection of bait protein expression in yeast by Western blot using antibodies against GAL4 DBD. Baits were expressed as GAL4 DBD fusion proteins in Y2HGold. **(C)** Western blot analysis of the AtSUMO1, AtSUMO2, and AtSUMO3 fusion proteins in diploid yeast using α -SUMO1/2 antibodies. AtSUMO2 displays strong expression while the other two isoforms are not detected. See **Figures C.1-C.4**.

References

- Abed, M., Barry, K.C., Kenyagin, D., Koltun, B., Phippen, T.M., Delrow, J.J., Parkhurst, S.M., and Orian, A.** (2011). Degringolade, a SUMO-targeted ubiquitin ligase, inhibits Hairy/Groucho-mediated repression. *The EMBO J* **30**, 1289-1301.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.** (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.
- Arriagada, G., Muntean, L.N., and Goff, S.P.** (2011). SUMO-interacting motifs of human TRIM5 α are important for antiviral activity. *PLoS Pathog* **7**, e1002019.
- Baba, D., Maita, N., Jee, J.-G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M.** (2005). Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **435**, 979-982.
- Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., and Rosen, M.K.** (2016). Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663.
- Barroso-Gomila, O., Trulsson, F., Muratore, V., Canosa, I., Merino-Cacho, L., Cortazar, A.R., Pérez, C., Azkargorta, M., Iloro, I., Carracedo, A., Aransay, A.M., Elortza, F., Mayor, U., Vertegaal, A.C.O., Barrio, R., and Sutherland, J.D.** (2021). Identification of proximal SUMO-dependent interactors using SUMO-ID. *Nat Comm* **12**, 6671.
- Beauchair, G., Bridier-Nahmias, A., Zagury, J.-F., Saïb, A., and Zamborlini, A.** (2015). JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. *Bioinformatics* **31**, 3483-3491.
- Benlloch, R., and Lois, L.M.** (2018). Sumoylation in plants: mechanistic insights and its role in drought stress. *J Exp Bot* **69**, 4539-4554.
- Bernhardt, C., Lee, M.M., Gonzalez, A., Zhang, F., Lloyd, A., and Schiefelbein, J.** (2003). The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. *Development* **130**, 6431-6439.
- Bernhofer, M., Goldberg, T., Wolf, S., Ahmed, M., Zaugg, J., Boden, M., and Rost, B.** (2018). NLSdb-major update for database of nuclear localization signals and nuclear export signals. *Nucleic Acids Res* **46**, D503-D508.
- Bernstein, H.J.** (2009). RasMol Molecular Graphics Visualisation Tool.
- Cao, Y.J., Wei, Q., Liao, Y., Song, H.L., Li, X., Xiang, C.B., and Kuai, B.K.** (2009). Ectopic overexpression of AtHDG11 in tall fescue resulted in enhanced tolerance to drought and salt stress. *Plant Cell Rep* **28**, 579-588.
- Cappadocia, L., Mascle, X.H., Bourdeau, V., Tremblay-Belzile, S., Chaker-Margot, M., Lussier-Price, M., Wada, J., Sakaguchi, K., Aubry, M., Ferbeyre, G., and Omichinski, J.G.** (2015). Structural and functional characterization of the

- phosphorylation-dependent interaction between PML and SUMO1. *Structure* **23**, 126-138.
- Castaño-Miquel, L., Seguí, J., and Lois, L.M.** (2011). Distinctive properties of Arabidopsis SUMO paralogues support the in vivo predominant role of AtSUMO1/2 isoforms. *The Biochem J* **436**, 581-590.
- Chang, C.-C., Naik, Mandar T., Huang, Y.-S., Jeng, J.-C., Liao, P.-H., Kuo, H.-Y., Ho, C.-C., Hsieh, Y.-L., Lin, C.-H., Huang, N.-J., Naik, Nandita M., Kung, Camy C.H., Lin, S.-Y., Chen, R.-H., Chang, K.-S., Huang, T.-H., and Shih, H.-M.** (2011). Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation. *Mol Cell* **42**, 62-74.
- Chang, P.-C., Izumiya, Y., Wu, C.-Y., Fitzgerald, L.D., Campbell, M., Ellison, T.J., Lam, K.S., Luciw, P.A., and Kung, H.-J.** (2010). Kaposi's Sarcoma-associated Herpesvirus (KSHV) encodes a SUMO E3 ligase that is SIM-dependent and SUMO-2/3-specific. *J Biol Chem* **285**, 5266-5273.
- Chew, W., Hrmova, M., and Lopato, S.** (2013). Role of Homeodomain leucine zipper (HD-Zip) IV transcription factors in plant development and plant protection from deleterious environmental factors. *Int J Mol Sci* **14**, 8122-8147.
- Ciarbelli, A.R., Ciolfi, A., Salvucci, S., Ruzza, V., Possenti, M., Carabelli, M., Fruscalzo, A., Sessa, G., Morelli, G., and Ruberti, I.** (2008). The Arabidopsis Homeodomain-leucine Zipper II gene family: diversity and redundancy. *Plant Mol Biol* **68**, 465-478.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant J* **16**, 735-743.
- Conti, L., Nelis, S., Zhang, C., Woodcock, A., Swarup, R., Galbiati, M., Tonelli, C., Napier, R., Hedden, P., Bennett, M., and Sadanandom, A.** (2014). Small Ubiquitin-like Modifier protein SUMO enables plants to control growth independently of the phytohormone gibberellin. *Dev Cell* **28**, 102-110.
- Cuchet-Lourenço, D., Boutell, C., Lukashchuk, V., Grant, K., Sykes, A., Murray, J., Orr, A., and Everett, R.D.** (2011). SUMO pathway dependent recruitment of cellular repressors to herpes simplex virus type 1 genomes. *PLoS Pathog* **7**, e1002123.
- Daniel Gietz, R., and Woods, R.A.** (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. In *Methods in Enzymology*, C. Guthrie and G.R. Fink, eds (Academic Press), pp. 87-96.
- 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. *The Plant J* **10**, 393-402.

- Du, J.X., McConnell, B.B., and Yang, V.W.** (2010). A small ubiquitin-related modifier-interacting motif functions as the transcriptional activation domain of Krüppel-like factor 4. *J Biol Chem* **285**, 28298-28308.
- Elrouby, N., Bonequi, M.V., Porri, A., and Coupland, G.** (2013). Identification of Arabidopsis SUMO-interacting proteins that regulate chromatin activity and developmental transitions. *PNAS USA* **110**, 19956-19961.
- Erker, Y., Neyret-Kahn, H., Seeler, J.S., Dejean, A., Atfi, A., and Levy, L.** (2013). Arkadia, a novel SUMO-targeted ubiquitin ligase involved in PML degradation. *Mol Cell Biol* **33**, 2163-2177.
- Gareau, J.R., and Lima, C.D.** (2010). The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **11**, 861-871.
- Gareau, J.R., Reverter, D., and Lima, C.D.** (2012). Determinants of Small Ubiquitin-like Modifier 1 (SUMO1) protein Specificity, E3 ligase, and SUMO-RanGAP1 binding activities of nucleoporin RanBP2. *J Biol Chem* **287**, 4740-4751.
- Geoffroy, M.C., and Hay, R.T.** (2009). An additional role for SUMO in ubiquitin-mediated proteolysis. *Nat Rev Mol Cell Biol* **10**, 564-568.
- Guzzo, C.M., Berndsen, C.E., Zhu, J., Gupta, V., Datta, A., Greenberg, R.A., Wolberger, C., and Matunis, M.J.** (2012). RNF4-dependent hybrid SUMO-ubiquitin chains are signals for RAP80 and thereby mediate the recruitment of BRCA1 to sites of DNA damage. *Science Signaling* **5**, ra88-ra88.
- Hammoudi, V., Vlachakis, G., Schranz, M.E., and van den Burg, H.A.** (2016). Whole-genome duplications followed by tandem duplications drive diversification of the protein modifier SUMO in Angiosperms. *New Phytol* **211**, 172-185.
- Hannich, J.T., Lewis, A., Kroetz, M.B., Li, S.J., Heide, H., Emili, A., and Hochstrasser, M.** (2005). Defining the SUMO-modified proteome by multiple approaches in *Saccharomyces cerevisiae*. *J Biol Chem* **280**, 4102-4110.
- Hay, R.T.** (2001). Protein modification by SUMO. *Trends in Biochemical Sciences* **26**, 332-333.
- Hecker, C.M., Rabiller, M., Haglund, K., Bayer, P., and Dikic, I.** (2006). Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* **281**, 16117-16127.
- Holm, M., Hardtke, C.S., Gaudet, R., and Deng, X.-W.** (2001). Identification of a structural motif that confers specific interaction with the WD40 repeat domain of Arabidopsis COP1. *The EMBO J* **20**, 118-127.
- 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.

- Johnson, E.S.** (2004). Protein Modification by SUMO. *Annual Rev Biochem* **73**, 355-382.
- Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E.** (1984). A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499-509.
- Keiten-Schmitz, J., Röder, L., Hornstein, E., Müller-McNicoll, M., and Müller, S.** (2021). SUMO: Glue or solvent for phase-separated ribonucleoprotein complexes and molecular condensates? *Front Mol Biosci* **8**, 673038.
- Kerscher, O.** (2007). SUMO junction-what's your function? New insights through SUMO-interacting motifs. *EMBO Rep* **8**, 550-555.
- Keusekotten, K., Bade, Veronika N., Meyer-Teschendorf, K., Sriramachandran, Annie M., Fischer-Schrader, K., Krause, A., Horst, C., Schwarz, G., Hofmann, K., Dohmen, R.J., and Praefcke, Gerrit J.K.** (2013). Multivalent interactions of the SUMO-interaction motifs in RING finger protein 4 determine the specificity for chains of the SUMO. *Biochem J* **457**, 207-214.
- 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. *The Plant Cell* **26**, 2184-2200.
- Kim, E.T., Kim, K.K., Matunis, M.J., and Ahn, J.-H.** (2009). Enhanced SUMOylation of proteins containing a SUMO-interacting motif by SUMO-Ubc9 fusion. *Biochem Biophys Res Comm* **388**, 41-45.
- Kim, E.T., Kim, Y.-E., Huh, Y.H., and Ahn, J.-H.** (2010). Role of noncovalent SUMO binding by the human cytomegalovirus IE2 transactivator in lytic growth. *J Vir* **84**, 8111-8123.
- Kim, J.Y., Jang, I.C., and Seo, H.S.** (2016). COP1 controls abiotic stress responses by modulating AtSIZ1 function through its E3 ubiquitin ligase activity. *Front Plant Sci* **7**, 1182.
- Knipscheer, P., Flotho, A., Klug, H., Olsen, J.V., van Dijk, W.J., Fish, A., Johnson, E.S., Mann, M., Sixma, T.K., and Pichler, A.** (2008). Ubc9 sumoylation regulates SUMO target discrimination. *Mol Cell* **31**, 371-382.
- Kolesar, P., Sarangi, P., Altmannova, V., Zhao, X., and Krejci, L.** (2012). Dual roles of the SUMO-interacting motif in the regulation of Srs2 sumoylation. *Nucleic Acids Res* **40**, 7831-7843.
- Kumar, R., González-Prieto, R., Xiao, Z., Verlaan-de Vries, M., and Vertegaal, A.C.O.** (2017). The STUbL RNF4 regulates protein group SUMOylation by targeting the SUMO conjugation machinery. *Nat Comm* **8**, 1809-1809.

- Kurepa, J., Walker, J.M., Smalle, J., Gosink, M.M., Davis, S.J., Durham, T.L., Sung, D.Y., and Vierstra, R.D.** (2003). The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress. *J Biol Chem* **278**, 6862-6872.
- 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.
- Lee, M.M., and Schiefelbein, J.** (1999). WEREWOLF, a MYB-related protein in Arabidopsis, Is a position-dependent regulator of epidermal cell patterning. *Cell* **99**, 473-483.
- Lin, D.Y., Huang, Y.S., Jeng, J.C., Kuo, H.Y., Chang, C.C., Chao, T.T., Ho, C.C., Chen, Y.C., Lin, T.P., Fang, H.I., Hung, C.C., Suen, C.S., Hwang, M.J., Chang, K.S., Maul, G.G., and Shih, H.M.** (2006). Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol Cell* **24**, 341-354.
- Liu, Y.-C., Lin, M.-C., Chen, H.-C., Tam, M.F., and Lin, L.-Y.** (2011). The role of small ubiquitin-like modifier-interacting motif in the assembly and regulation of metal-responsive transcription factor 1. *The J Biol Chem* **286**, 42818-42829.
- Lowe, A.R., Tang, J.H., Yassif, J., Graf, M., Huang, W.Y.C., Groves, J.T., Weis, K., and Liphardt, J.T.** (2015). Importin- β modulates the permeability of the nuclear pore complex in a Ran-dependent manner. *eLife* **4**, e04052.
- Lukic, Z., Goff, S.P., Campbell, E.M., and Arriagada, G.** (2013). Role of SUMO-1 and SUMO interacting motifs in rhesus TRIM5 α -mediated restriction. *Retrovirology* **10**, 10.
- 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.
- Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F.** (1997). A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**, 97-107.
- 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.
- Matunis, M.J., Wu, J., and Blobel, G.** (1998). SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. *J Cell Biol* **140**, 499-509.

- Mazur, M.J., Kwaaitaal, M., Mateos, M.A., Maio, F., Kini, R.K., Prins, M., and van den Burg, H.A.** (2019). The SUMO conjugation complex self-assembles into nuclear bodies independent of SIZ1 and COP1. *Plant physiol* **179**, 168-183.
- Mazur, M.J., Spears, B.J., Djajasaputra, A., van der Gragt, M., Vlachakis, G., Beerens, B., Gassmann, W., and van den Burg, H.A.** (2017). Arabidopsis TCP transcription factors interact with the SUMO conjugating machinery in nuclear foci. *Front Plant Sci* **8**, 2043.
- Melchior, F.** (2000). SUMO—Nonclassical Ubiquitin. *Annu Rev Cell Dev Biol* **16**, 591-626.
- Meulmeester, E., and Melchior, F.** (2008). SUMO. *Nature* **452**, 709-711.
- Minty, A., Dumont, X., Kaghad, M., and Caput, D.** (2000). Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif. *J Biol Chem* **275**, 36316-36323.
- Morohashi, K., Zhao, M., Yang, M., Read, B., Lloyd, A., Lamb, R., and Grotewold, E.** (2007). Participation of the Arabidopsis bHLH factor GL3 in trichome initiation regulatory events. *Plant Physiol* **145**, 736-746.
- Moroianu, J., Blobel, G., and Radu, A.** (1995). Previously identified protein of uncertain function is karyopherin alpha and together with karyopherin beta docks import substrate at nuclear pore complexes. *PNAS* **92**, 2008-2011.
- 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.
- Mukherjee, T., Subedi, B., Khosla, A., Begler, E.M., Stephens, P.M., Warner, A.L., Lerma-Reyes, R., Thompson, K.A., Gunewardena, S., and Schrick, K.** (2022). The START domain mediates Arabidopsis GLABRA2 dimerization and turnover independently of homeodomain DNA binding. *Plant Physiol* **kiac383**.
- Müller, S., Hoege, C., Pyrowolakis, G., and Jentsch, S.** (2001). Sumo, ubiquitin's mysterious cousin. *Nat Rev Mol Cell Biol* **2**, 202-210.
- Nagata, K., and Abe, M.** (2021). The lipid-binding START domain regulates the dimerization of ATML1 via modulating the ZIP motif activity in *Arabidopsis thaliana*. *Dev Growth Differ* **63**, 448-454.
- Nair, A., Chatterjee, K.S., Jha, V., Das, R., and Shivaprasad, P.V.** (2020). Stability of Begomoviral pathogenicity determinant β C1 is modulated by mutually antagonistic SUMOylation and SIM interactions. *BMC Biol* **18**, 110.
- 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.

- Namanja, A.T., Li, Y.J., Su, Y., Wong, S., Lu, J., Colson, L.T., Wu, C., Li, S.S., and Chen, Y.** (2012). Insights into high affinity small ubiquitin-like modifier (SUMO) recognition by SUMO-interacting motifs (SIMs) revealed by a combination of NMR and peptide array analysis. *J Biol Chem* **287**, 3231-3240.
- Nelis, S., Conti, L., Zhang, C., and Sadanandom, A.** (2015). A functional Small Ubiquitin-like Modifier (SUMO) interacting motif (SIM) in the gibberellin hormone receptor GID1 is conserved in cereal crops and disrupting this motif does not abolish hormone dependency of the DELLA-GID1 interaction. *Plant Signal Behav* **10**, e987528.
- Novatchkova, M., Budhiraja, R., Coupland, G., Eisenhaber, F., and Bachmair, A.** (2004). SUMO conjugation in plants. *Planta* **220**, 1-8.
- Ouyang, J., Shi, Y., Valin, A., Xuan, Y., and Gill, G.** (2009). Direct Binding of CoREST1 to SUMO-2/3 Contributes to Gene-Specific Repression by the LSD1/CoREST1/HDAC Complex. *Mol Cell* **34**, 145-154.
- Payne, C.T., Zhang, F., and Lloyd, A.M.** (2000). GL3 encodes a bHLH protein that regulates trichome development in arabidopsis through interaction with GL1 and TTG1. *Genetics* **156**, 1349-1362.
- Perry, J.J., Tainer, J.A., and Boddy, M.N.** (2008). A SIM-ultaneous role for SUMO and ubiquitin. *Trends Biochem Sci* **33**, 201-208.
- Ponting, C.P., and Aravind, L.** (1999). START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends in Biochemical Sciences* **24**, 130-132.
- Poulsen, S.L., Hansen, R.K., Wagner, S.A., van Cuijk, L., van Belle, G.J., Streicher, W., Wikström, M., Choudhary, C., Houtsmuller, A.B., Marteijn, J.A., Bekker-Jensen, S., and Mailand, N.** (2013). RNF111/Arkadia is a SUMO-targeted ubiquitin ligase that facilitates the DNA damage response. *J Cell Biol* **201**, 797-807.
- Prudden, J., Pebernard, S., Raffa, G., Slavin, D.A., Perry, J.J.P., Tainer, J.A., McGowan, C.H., and Boddy, M.N.** (2007). SUMO-targeted ubiquitin ligases in genome stability. *The EMBO J* **26**, 4089-4101.
- Rerie, W.G., Feldmann, K.A., and Marks, M.D.** (1994). The GLABRA2 gene encodes a homeodomain protein required for normal trichome development in Arabidopsis. *Genes Dev* **8**, 1388-1399.
- Reverter, D., and Lima, C.D.** (2005). Insights into E3 ligase activity revealed by a SUMO–RanGAP1–Ubc9–Nup358 complex. *Nature* **435**, 687-692.
- 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). Transcription factors: Genome-wide Comparative analysis among eukaryotes. *Science* **290**, 2105.

- Robbins, J., Dilworth, S.M., Laskey, R.A., and Dingwall, C.** (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615-623.
- Ryu, H.-Y., Zhao, D., Li, J., Su, D., and Hochstrasser, M.** (2020). Histone sumoylation promotes Set3 histone-deacetylase complex-mediated transcriptional regulation. *Nucleic Acids Res* **48**, 12151-12168.
- Sæther, T., Pattabiraman, D.R., Alm-Kristiansen, A.H., Vogt-Kielland, L.T., Gonda, T.J., and Gabrielsen, O.S.** (2011). A functional SUMO-interacting motif in the transactivation domain of c-Myb regulates its myeloid transforming ability. *Oncogene* **30**, 212-222.
- Sahoo, M.R., Gaikwad, S., Khuperkar, D., Ashok, M., Helen, M., Yadav, S.K., Singh, A., Magre, I., Deshmukh, P., Dhanvijay, S., Sahoo, P.K., Ramtirtha, Y., Madhusudhan, M.S., Gayathri, P., Seshadri, V., and Joseph, J.** (2017). Nup358 binds to AGO proteins through its SUMO-interacting motifs and promotes the association of target mRNA with miRISC. *EMBO Reports* **18**, 241-263.
- Sakakibara, K., Nishiyama, T., Kato, M., and 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.
- Santiago, A., Godsey, A.C., Hossain, J., Zhao, L.Y., and Liao, D.** (2009). Identification of two independent SUMO-interacting motifs in Daxx: Evolutionary conservation from *Drosophila* to humans and their biochemical functions. *Cell Cycle* **8**, 76-87.
- Saracco, S.A., Miller, M.J., Kurepa, J., and Vierstra, R.D.** (2007). Genetic analysis of SUMOylation in *Arabidopsis*: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. *Plant Physiol* **145**, 119-134.
- Schena, M., and Davis, R.W.** (1994). Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families. *PNAS USA* **91**, 8393-8397.
- Schrack, 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.
- Schrack, 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.
- Sessa, G., Morelli, G., and Ruberti, I.** (1993). The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. *The EMBO J* **12**, 3507-3517.

- Sharif, R., Raza, A., Chen, P., Li, Y., El-Ballat, E.M., Rauf, A., Hano, C., and El-Esawi, M.A.** (2021). HD-ZIP gene family: Potential roles in improving plant growth and regulating stress-responsive mechanisms in plants. *Genes (Basel)* **12**.
- Shen, B., Sinkevicius, K.W., Selinger, D.A., and Tarczynski, M.C.** (2006a). The homeobox gene *GLABRA2* affects seed oil content in *Arabidopsis*. *Plant Mol Biol* **60**, 377-387.
- Shen, T.H., Lin, H.K., Scaglioni, P.P., Yung, T.M., and Pandolfi, P.P.** (2006b). The mechanisms of PML-nuclear body formation. *Mol Cell* **24**, 331-339.
- Shi, L., Katavic, V., Yu, Y., Kunst, L., and Haughn, G.** (2012). *Arabidopsis glabra2* mutant seeds deficient in mucilage biosynthesis produce more oil. *The Plant J* **69**, 37-46.
- Song, J., Zhang, Z., Hu, W., and Chen, Y.** (2005). Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *The J Biol Chem* **280**, 40122-40129.
- Song, J., Durrin, L.K., Wilkinson, T.A., Krontiris, T.G., and Chen, Y.** (2004). Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *PNAS USA* **101**, 14373-14378.
- Sriramachandran, A.M., Meyer-Teschendorf, K., Pabst, S., Ulrich, H.D., Gehring, N.H., Hofmann, K., Praefcke, G.J.K., and Dohmen, R.J.** (2019). Arkadia/RNF111 is a SUMO-targeted ubiquitin ligase with preference for substrates marked with SUMO1-capped SUMO2/3 chain. *Nat Comm* **10**, 3678.
- Srivastava, A.K., Orosa, B., Singh, P., Cummins, I., Walsh, C., Zhang, C., Grant, M., Roberts, M.R., Anand, G.S., Fitches, E., and Sadanandom, A.** (2018). SUMO suppresses the activity of the jasmonic acid receptor CORONATINE INSENSITIVE1. *The Plant Cell* **30**, 2099-2115.
- Steinacher, R., and Schär, P.** (2005). Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr Biol* **15**, 616-623.
- Subedi, B., and Schrick, K.** (2022). EYFP fusions to HD-Zip IV transcription factors enhance their stability and lead to phenotypic changes in *Arabidopsis*. *Plant Signal Behav* **17**, 2119013.
- 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.
- Takahashi, H., Hatakeyama, S., Saitoh, H., and Nakayama, K.I.** (2005). Noncovalent SUMO-1 binding activity of thymine DNA glycosylase (TDG) is required for its SUMO-1 modification and colocalization with the promyelocytic leukemia protein. *J Biol Chem* **280**, 5611-5621.

- Tatham, M.H., Kim, S., Jaffray, E., Song, J., Chen, Y., and Hay, R.T.** (2005). Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection. *Nat Struct Mol Biol* **12**, 67-74.
- Taupitz, K.F., Dörner, W., and Mootz, H.D.** (2017). Covalent capturing of transient SUMO–SIM interactions using unnatural amino acid mutagenesis and photocrosslinking. *Chemistry – A European Journal* **23**, 5978-5982.
- Tempé, D., Piechaczyk, M., and Bossis, G.** (2008). SUMO under stress. In *Biochemical Society Transactions*, pp. 874-878.
- Uljon, S., Xu, X., Durzynska, I., Stein, S., Adelmant, G., Marto, J.A., Pear, W.S., and Blacklow, S.C.** (2016). Structural basis for substrate selectivity of the E3 ligase COP1. *Structure* **24**, 687-696.
- Uzunova, K., Götsche, K., Miteva, M., Weisshaar, S.R., Glanemann, C., Schnellhardt, M., Niessen, M., Scheel, H., Hofmann, K., Johnson, E.S., Praefcke, G.J.K., and Dohmen, R.J.** (2007). Ubiquitin-dependent proteolytic control of SUMO conjugates. *J Biol Chem* **282**, 34167-34175.
- Wang, F., Yan, H., Fang, P., Ji, X., and Peng, Y.** (2022). Heterologous overexpression of ZmHDZIV13 enhanced drought and salt tolerance in Arabidopsis and Tobacco. *Agronomy* **12**, 2378.
- Wang, P.Y., Hsu, P.I., Wu, D.C., Chen, T.C., Jarman, A.P., Powell, L.M., and Chen, A.** (2018). SUMOs mediate the nuclear transfer of p38 and p-p38 during *Helicobacter pylori* infection. *Int J Mol Sci* **19**.
- Wang, S., and Chen, J.-G.** (2008). Arabidopsis transient expression analysis reveals that activation of GLABRA2 may require concurrent binding of GLABRA1 and GLABRA3 to the promoter of GLABRA2. *Plant and Cell Physiol* **49**, 1792-1804.
- Wang, Z., Wang, S., Xiao, Y., Li, Z., Wu, M., Xie, X., Li, H., Mu, W., Li, F., Liu, P., Wang, R., and Yang, J.** (2020). Functional characterization of a HD-ZIP IV transcription factor NtHDG2 in regulating flavonols biosynthesis in *Nicotiana tabacum*. *Plant Physiol Biochem* **146**, 259-268.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J.** (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.
- Werner, A., Flotho, A., and Melchior, F.** (2012). The RanBP2/RanGAP1*SUMO1/Ubc9 complex is a multisubunit SUMO E3 ligase. *Mol Cell* **46**, 287-298.
- 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.
- Yang, J., and Zhang, Y.** (2015a). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* **43**, W174-W181.
- Yang, J., and Zhang, Y.** (2015b). Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics* **52**, 5.8.1-5.8.15.
- Yau, T.-Y., Sander, W., Eidson, C., and Courey, A.J.** (2021). SUMO Interacting Motifs: Structure and Function. *Cells* **10**, 2825.
- Zalewski, C.S., Floyd, S.K., Furumizu, C., Sakakibara, K., Stevenson, D.W., and Bowman, J.L.** (2013). Evolution of the class IV HD-Zip gene family in Streptophytes. *Mol Biol Evol* **30**, 2347-2365.
- Zhang, C., Freddolino, P.L., and Zhang, Y.** (2017). COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic Acids Res* **45**, W291-W299.
- Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T., and Lloyd, A.** (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. *Development* **130**, 4859-4869.
- Zhang, H., Ma, X., Li, W., Niu, D., Wang, Z., Yan, X., Yang, X., Yang, Y., and Cui, H.** (2019). Genome-wide characterization of NtHD-ZIP IV: different roles in abiotic stress response and glandular trichome induction. *BMC Plant Biol* **19**, 444.
- Zhao, M., Morohashi, K., Hatlestad, G., Grotewold, E., and Lloyd, A.** (2008). The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development* **135**, 1991-1999.
- Zhao, Q., Xie, Y., Zheng, Y., Jiang, S., Liu, W., Mu, W., Liu, Z., Zhao, Y., Xue, Y., and Ren, J.** (2014). GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res* **42**, W325-330.
- Zheng, Y., Jayappa, K.D., Ao, Z., Qiu, X., Su, R.C., and Yao, X.** (2019). Noncovalent SUMO-interaction motifs in HIV integrase play important roles in SUMOylation, cofactor binding, and virus replication. *Virol J* **16**, 42.

Chapter 6 - Perspectives

Conclusions and future directions

HD-START proteins are deeply conserved in the plant lineage with their origin traced back to charophycean green algae (Sakakibara et al., 2001; Floyd et al., 2006; Prigge and Clark, 2006; Zalewski et al., 2013). Thus, this domain configuration represents a highly successful evolutionary innovation in plants. This dissertation mainly focuses on the role of HD-Zip IV TFs as the developmental regulators of trichomes, root hairs and seed coat mucilage; however, recent studies have highlighted emerging roles of these TFs in plant stress responses (Yu et al., 2008; Chen et al., 2017; Banavath et al., 2018; Guo et al., 2019; Zhang et al., 2019; Wang et al., 2022). One of the important aspects of this work is identifying the contributions of major domains including HD, START and ZLZ to the overall activity of HD-Zip IV TFs. Here, the START domain was characterized as a versatile module that is important not only for ligand binding but also for protein dimerization, stability and possible protein-protein interactions of HD-Zip IV TFs (Mukherjee et al., 2022). Previous studies reported the dimerization of HD-Zip TFs as a prerequisite for DNA binding (Tron et al., 2001; Tron et al., 2004); however, we found that HD-Zip IV TFs with mutations in their START domain can bind DNA as monomers (Mukherjee et al., 2022). Although the presence of a START domain within a transcription factor is unique to plants, START domains occur in proteins from diverse organisms, indicating a broader impact of this study.

Another important finding from this work is that the ZLZ dimerization domain negatively affects protein stability and is enriched in potential post-translational modification sites. This is in contrast to the HD whose role appears to be confined to DNA binding (Mukherjee et al.,

2022). The enhanced stability of the $gl2^{\Delta ZLZ}$ mutant protein that is defective in DNA binding points towards the possibility that the lack of DNA binding makes the proteins less susceptible to degradation. However, mutations in the HD did not alter the protein stability of GL2 (Mukherjee et al., 2022), suggesting more complex association between the DNA binding and protein stability. Although the data presented here suggest that each of these domains functions independently of the other, it is still possible that interactions between one or more of the other domains results in coordination of TF activities. One possible scenario is that the START domain regulates ZLZ-mediated homodimerization via lipid ligand binding, as was recently suggested for ATML1 (Nagata and Abe, 2021). This raises the possibility of control of ZLZ function by START domain in GLABRA2 and other HD-Zip IV TFs.

Interestingly, START domain mutations affecting amino acid residues involved in lipid binding caused impaired dimerization and stability of HD-Zip IV TFs leading to phenotypic defects in Arabidopsis (Mukherjee et al., 2022). Ligand binding to the START domain has been postulated to induce conformational changes favorable for protein dimerization and stabilization, ultimately leading to optimal differentiation of epidermal layers. However, we currently do not have strong evidence for the exact sequence of these events. Therefore, it will be important to investigate how lipid binding, protein dimerization and stability are coupled together to appropriately regulate the expression of target genes downstream of these TFs. One of the challenges associated with these studies is the lack of experimental structural information (i.e. crystal structure) of GL2 or any other HD-Zip IV members despite having a long history of their molecular research. In addition, we do not know whether the ligand binding by the START domain can affect specific protein-protein interactions of these TFs. In this dissertation, the START domain was predicted to be involved in the protein interactions of GL2 via SIMs.

Intriguingly, the predicted SIMs were found to be critical for GL2 nuclear localization and dimerization, both of which are required for TF activity of GL2. However, more sophisticated and sensitive biophysical methods such as NMR will be required to characterize the possible weak and transient interaction(s) of GL2 with other proteins. It will also be important to determine whether these interactions are relevant to dimerization and TF activity of GL2.

The microscopy results showed that all three major domains including ZLZ, HD and START are dispensable for the proper subcellular localization of GL2 (Mukherjee et al., 2022). Thus, a yet uncharacterized motif located towards either amino- or carboxyl- terminal end of the HD-Zip IV TFs seems to function as a nuclear localization signal. Similarly, the Y2H assays revealed an important role for the carboxyl-terminal STAD domain in START domain-mediated homodimerization of GL2 (Mukherjee et al., 2022). Recently, there has been some interest in considering this domain as a second START domain based on AlphaFold structural predictions (Jumper et al., 2021). In future studies, it will be important to understand the function of the two START domains, which also can be referred to as a ‘didomain’, and to determine whether STAD binds lipid ligands in a similar fashion as START.

Furthermore, we mostly relied on bioinformatic predictions, mutational analysis and immunoprecipitation experiments for the identification and characterization of GL2 PTMs. The reasoning behind the K-R mutations was to remove the sites predicted for modifications, but such missense mutations may also affect protein side-chain geometry. Although the genetic analysis presented here could not identify the specific lysine(s) undergoing modification, faint smears of molecular weights comparable to those of ubiquitinated GL2 were detected in WBs of immunoprecipitated proteins under daytime conditions. Strikingly, these observations were further supported by the results from protein stability experiments using the proteasome inhibitor

MG132 (Mukherjee et al., 2022). Mass spectrometry-based proteomic approaches are likely to be more sensitive and promising to correctly identify modification sites in HD-Zip IV TFs. Similarly, future studies of day-night differences with respect to modifications could be fruitful in understanding whether the GL2 protein itself or a component mediating its modification is under the influence of the circadian clock.

Last but not the least, the effects of the EYFP tag on protein stability and biological function which was documented here (Subedi and Schrick, 2022) provides mounting evidence for the limitations of fluorescent tags in molecular biology experiments. The curly leaf phenotype was associated with EYFP-induced GL2 stabilization (Subedi and Schrick, 2022) which occurred at an elevated frequency in PTM mutants. This finding uncovers a novel role for GL2 in leaf polarity establishment, presenting an interesting topic for further investigation.

Overall, the scope of this work provides mechanistic insights into HD-Zip IV transcription factor function and regulation, with relevance for agricultural productivity as well as for biotechnology and synthetic biology. Future experiments are poised to focus on details of how each regulatory mechanism contributes to the overall function of HD-Zip IV TFs in plant growth and development, as well as in response to environmental factors.

References

- Banavath, J.N., Chakradhar, T., Pandit, V., Konduru, S., Guduru, K.K., Akila, C.S., Podha, S., and Puli, C.O.R.** (2018). Stress inducible overexpression of AtHDG11 leads to improved drought and salt stress tolerance in peanut (*Arachis hypogaea* L.). *Front Chem* **6**, 34.
- Chen, E., Zhang, X., Yang, Z., Wang, X., Yang, Z., Zhang, C., Wu, Z., Kong, D., Liu, Z., Zhao, G., Butt, H.I., Zhang, X., and Li, F.** (2017). Genome-wide analysis of the HD-ZIP IV transcription factor family in *Gossypium arboreum* and GaHDG11 involved in osmotic tolerance in transgenic Arabidopsis. *Mol Genet Genomics* **292**, 593-609.
- Floyd, S.K., Zalewski, C.S., and Bowman, J.L.** (2006). Evolution of Class III Homeodomain–Leucine Zipper Genes in Streptophytes. *Genetics* **173**, 373-388.
- Guo, X.-Y., Wang, Y., Zhao, P.-X., Xu, P., Yu, G.-H., Zhang, L.-Y., Xiong, Y., and Xiang, C.-B.** (2019). AtEDT1/HDG11 regulates stomatal density and water-use efficiency via ERECTA and E2Fa. *New Phytologist* **223**, 1478-1488.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstern, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., and Hassabis, D.** (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589.
- Mukherjee, T., Subedi, B., Khosla, A., Begler, E.M., Stephens, P.M., Warner, A.L., Lerma-Reyes, R., Thompson, K.A., Gunewardena, S., and Schrick, K.** (2022). The START domain mediates Arabidopsis GLABRA2 dimerization and turnover independently of homeodomain DNA binding. *Plant Physiol* **kiac383**.
- Nagata, K., and Abe, M.** (2021). The lipid-binding START domain regulates the dimerization of ATML1 via modulating the ZIP motif activity in *Arabidopsis thaliana*. *Dev Growth Differ* **63**, 448-454.
- Prigge, M.J., and Clark, S.E.** (2006). Evolution of the class III HD-Zip gene family in land plants. *Evolution & Development* **8**, 350-361.
- Sakakibara, K., Nishiyama, T., Kato, M., and 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.
- Subedi, B., and Schrick, K.** (2022). EYFP fusions to HD-Zip IV transcription factors enhance their stability and lead to phenotypic changes in Arabidopsis. *Plant Signal Behav* **17**, 2119013.

- Tron, A.E., Welchen, E., and Gonzalez, D.H.** (2004). Engineering the loop region of a homeodomain-leucine zipper protein promotes efficient binding to a monomeric DNA binding site. *Biochemistry* **43**, 15845-15851.
- 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.
- Wang, F., Yan, H., Fang, P., Ji, X., and Peng, Y.** (2022). Heterologous overexpression of ZmHDZIV13 enhanced drought and salt tolerance in Arabidopsis and Tobacco. *Agronomy* **12**, 2378.
- Yu, H., Chen, X., Hong, Y.-Y., Wang, Y., Xu, P., Ke, S.-D., Liu, H.-Y., Zhu, J.-K., Oliver, D.J., and Xiang, C.-B.** (2008). Activated expression of an Arabidopsis HD-START protein confers drought tolerance with improved root system and reduced stomatal density. *The Plant Cell* **20**, 1134-1151.
- Zalewski, C.S., Floyd, S.K., Furumizu, C., Sakakibara, K., Stevenson, D.W., and Bowman, J.L.** (2013). Evolution of the class IV HD-Zip gene family in Streptophytes. *Mol Biol Evol* **30**, 2347-2365.
- Zhang, H., Ma, X., Li, W., Niu, D., Wang, Z., Yan, X., Yang, X., Yang, Y., and Cui, H.** (2019). Genome-wide characterization of NtHD-ZIP IV: different roles in abiotic stress response and glandular trichome induction. *BMC Plant Biol* **19**, 444.

Appendix A - Chapter 2 Supplemental Data

Supplemental Figures

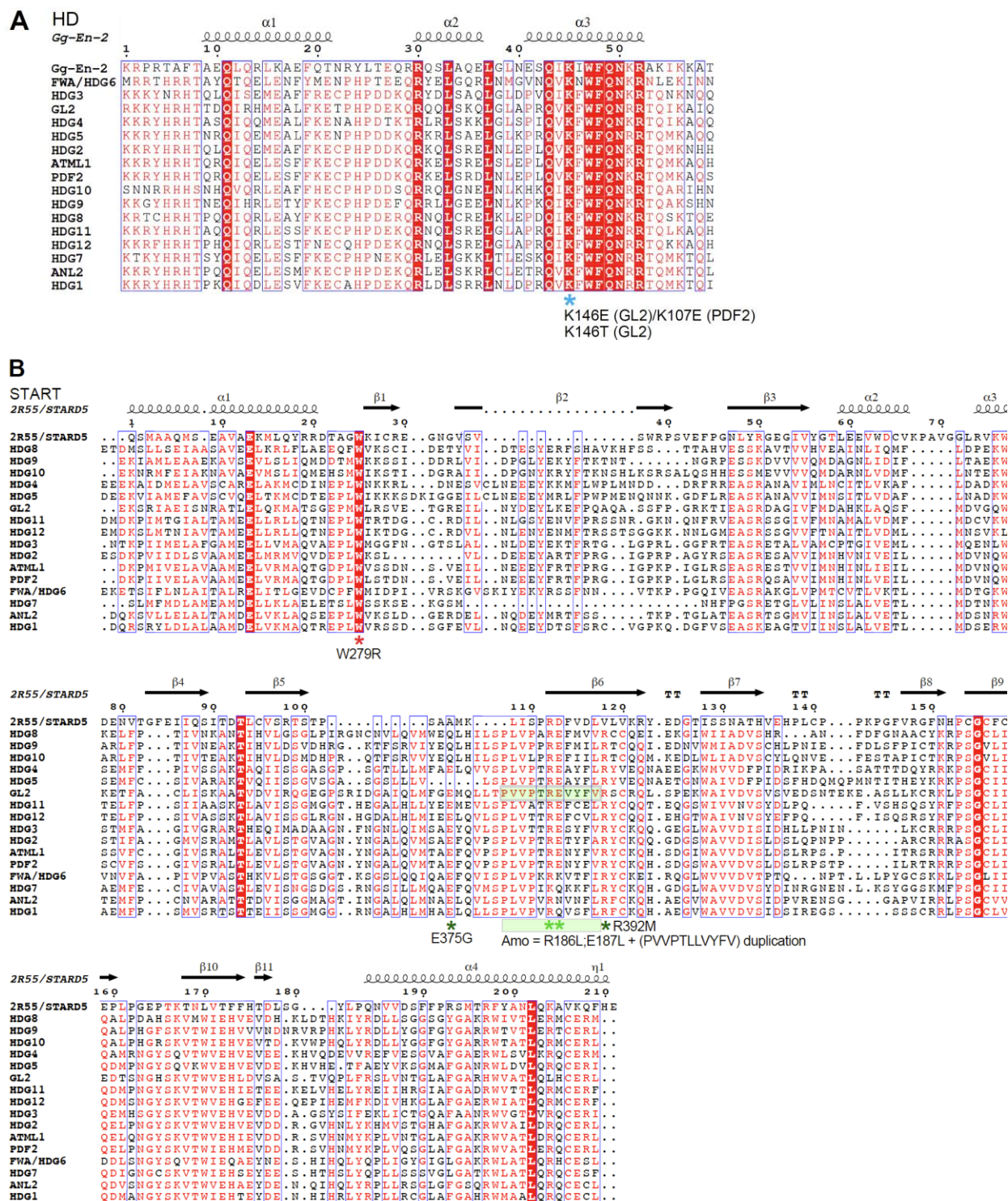


Figure A.1. Structural alignments of HD and START domain from HD-Zip IV TFs (related to Figures 2.1A and 2.3A).

(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* HOMEODOMAIN 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)** Structural alignment indicates positions of START domain mutations examined in this study. Related to **Figure 2.3A**. Structural alignment of the START domain across the HD-Zip IV family in *Arabidopsis* 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. Numbering (1-211) corresponds to the START domain only. Positions of START mutations of GL2 analyzed in ChIP-seq (*E375G;R392M* and *Amo*) (**Figure 2.4, D-G**) and the cycloheximide chase assays (*W279R*, *Amo*, and *L480P*) (**Figure 2.6; Figures A.9 and A.10**) are indicated below alignment (asterisks): *W279R* is a missense allele that adds a charged residue within a hydrophobic region of the binding pocket. *L480P* affects a lipid contact site in the C-terminal α -helix (Roostae et al., 2009; Alpy and Tomasetto, 2014; Wojciechowska, 2021). In the *E375G;R392M* double mutant, a predicted salt bridge critical for ligand binding is disrupted (Roostae et al., 2009). In *Amo*, the *R386L* and *E387L* missense mutations, which replace charged residues with leucines, are followed by an 11-residue duplication in a predicted ligand-contact region.

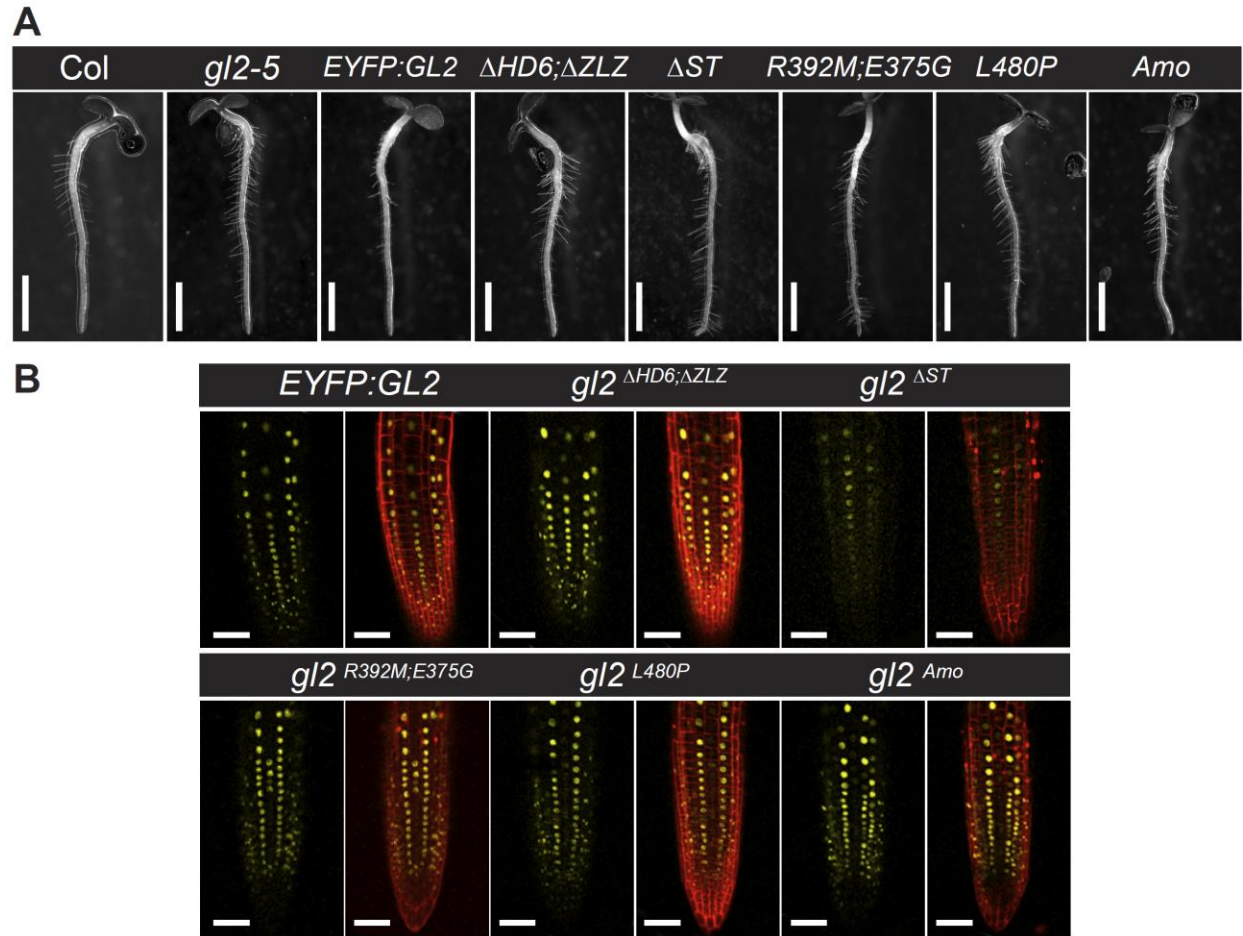


Figure A.2. Root hair phenotypes and nuclear localization in various START domain missense mutants (related to Figure 2.3).

(A) Root hair phenotypes of 4–5-day-old seedlings. Bar = 1 mm. In contrast to wild-type Col, *gl2-5* null mutants display excess root hair formation. Expression of *EYFP:GL2* rescues the root hair phenotype of *gl2-5*. In contrast, aberrant root hair formation is observed for lines expressing EYFP-tagged *gl2* ^{$\Delta HD6;\Delta ZLZ$} , *gl2* ^{ΔST} , *gl2*^{*E375G;R392M*}, *gl2*^{*L480P*} and *gl2*^{*Amo*}. Bar = 1 mm.

(B) Confocal laser scanning images of primary roots from 4-day-old seedlings expressing the EYFP-tagged proteins for lines depicted in **(A)**. Nuclear localization of the EYFP:GL2 wild-type and *gl2* mutant proteins (yellow). Propidium iodide staining (red) was performed to visualize cell boundaries. Bar = 50 μ m.

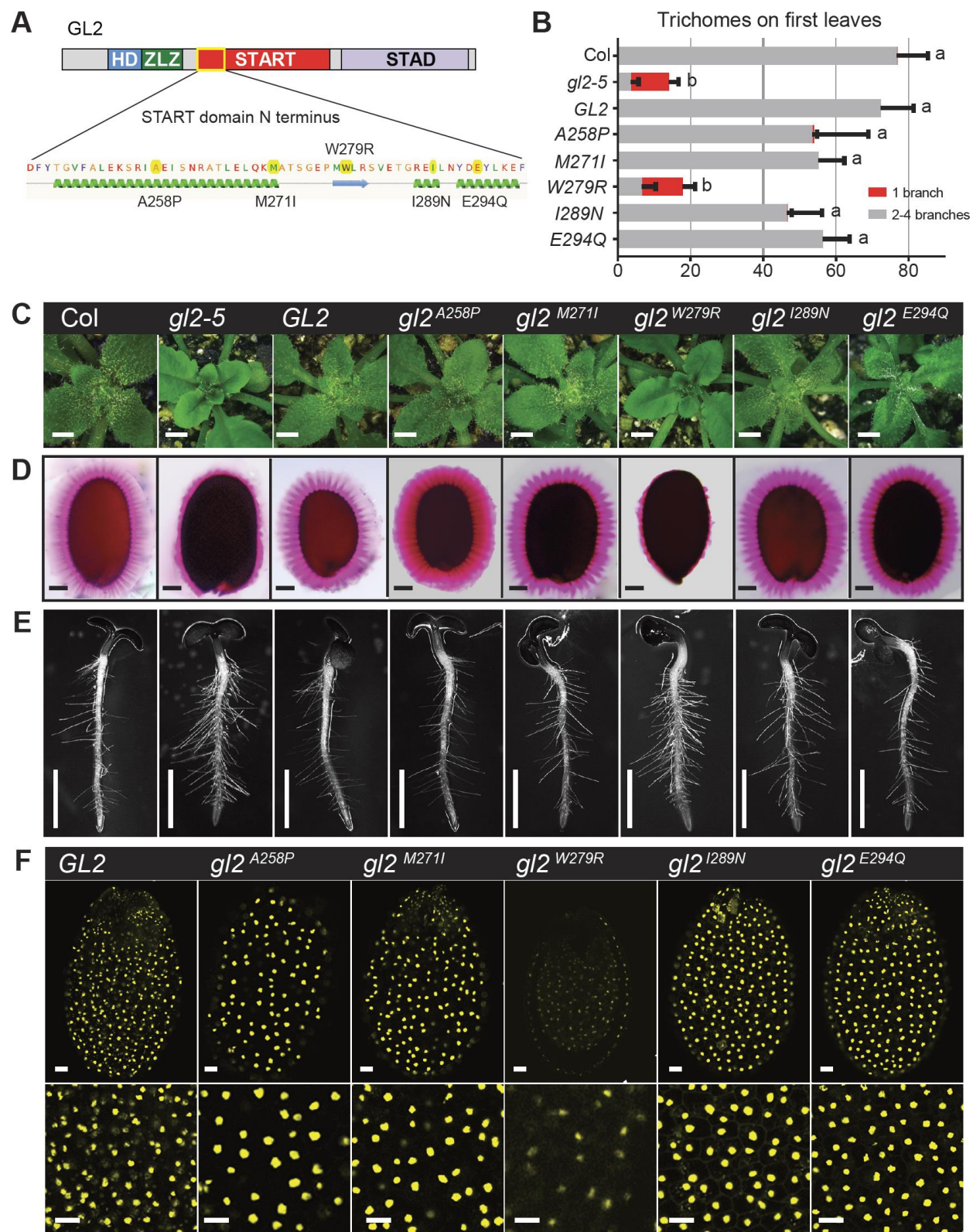


Figure A.3. Characterization of START domain missense mutants in the N-terminus of the START domain (related to Figure 2.3).

(A) Schematic showing the functional domains of the GL2 protein. Positions of mutations within the N-terminus of the START domain are indicated in relation to predicted helical (green) and strand (blue) secondary structures. See also **Figure A.1B**. **(B)** Total number and branching pattern of trichomes on the first leaves. Gray bars indicate normal trichomes in plants expressing Col wild type, EYFP-tagged GL2 and the majority of the *gl2* missense mutants shown in **(A)**. In contrast, mutant *gl2-5* plants and those expressing EYFP-tagged *gl2*^{W279R} exhibit mostly unbranched trichomes (red bars). Values are mean ± SD for n = 20 plants. Letters denote significant differences from one-way ANOVA (p < 0.05) using Tukey's multiple comparisons test. **(C)** Rosettes from wild-type Col, *EYFP:GL2*, and the majority of the *gl2* missense mutants shown in **(A)** display trichomes covering leaf surfaces. The *gl2-5* mutants and the EYFP-tagged START domain mutant *gl2*^{W279R} exhibit trichome differentiation defects. Bar = 1 mm. **(D)** Ruthenium red staining of seeds to assay mucilage production. Wild-type Col, *EYFP:GL2*, and the majority of *EYFP:gl2* mutant seeds display a normal mucilage layer, while *gl2-5* mutants and the EYFP-tagged START domain mutant *gl2*^{W279R} exhibit defective mucilage production. Bar = 100 µm. **(E)** Root hair phenotypes of 4-5 d-old seedlings. In contrast to wild-type *GL2* and the majority of the *gl2* mutants, *gl2-5* and *gl2*^{W279R} mutants exhibit excessive root hair formation. Bar = 1 mm. **(F)** Nuclear localization of EYFP-tagged GL2 and mutant *gl2* transcription factors. Top panel: Confocal laser scanning microscopy of developing seeds reveals nuclear expression of EYFP-tagged GL2 and mutant proteins in seed coat cells. Bar = 100 µm. Bottom panel: Magnified images. The EYFP:*gl2*^{W279R} mutant protein exhibits weak expression in comparison to wild-type EYFP:GL2 and the other missense mutants. Bar = 50 µm.

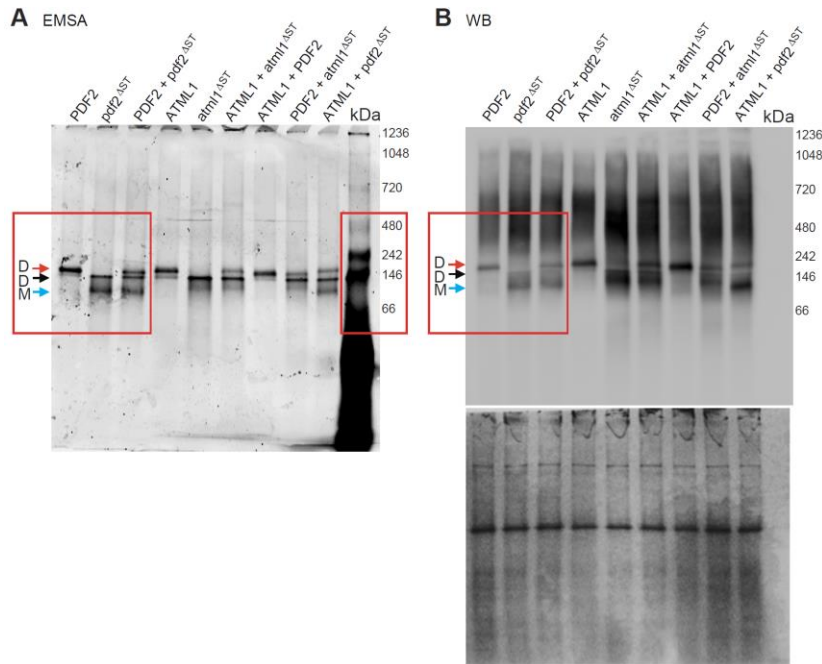


Figure A.4. START domain is required for in vitro DNA binding as a dimer (related to Figure 2.4C).

(A) Electrophoretic mobility shift assay (EMSA) coupled with polyacrylamide gel electrophoresis reveals that Halo:PDF2 and Halo:ATML bind the Cy3-labeled L1-box probe as a dimer (~230 kDa) (red arrow) while Halo:pdf2^{ΔST} and Halo:atml1^{ΔST} bind DNA as a monomer (~90 kDa) (blue arrow) and weakly as a dimer (~180 kDa) (black arrow). A 1:1 mixture of Halo-tagged PDF2 and pdf2^{ΔST}, PDF2 and atml1^{ΔST}, as well as ATML1 and pdf2^{ΔST}, indicate additive banding patterns, suggesting that dimerization is affected when at least one of the two partners harbors a START domain deletion. Parts identified by red boxes are displayed in **Figure 2.4C**.

(B) Western blot (WB) (top) of the polyacrylamide gel in **(A)** with anti-Halo Ab detected Halo-tagged PDF2, ATML, pdf2^{ΔST} and atml1^{ΔST} at the expected migration positions as indicated from the EMSA **(A)**. After Western blotting, the blot was stained with Coomassie blue (bottom, in grayscale) to monitor equal loading of each EMSA reaction. The major band across all samples likely originates from wheat-germ extract used for in vitro transcription/translation of the Halo-tagged proteins.

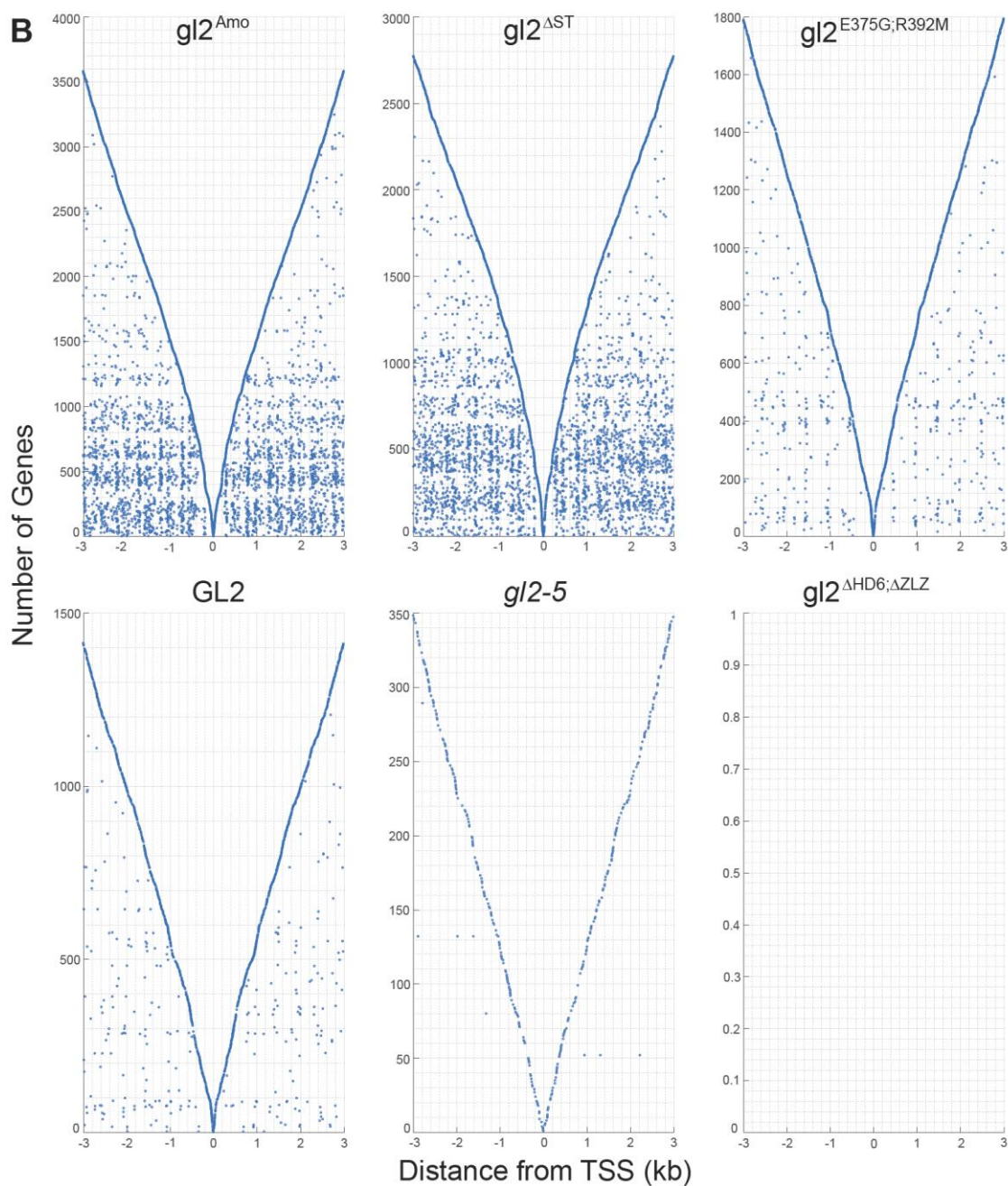
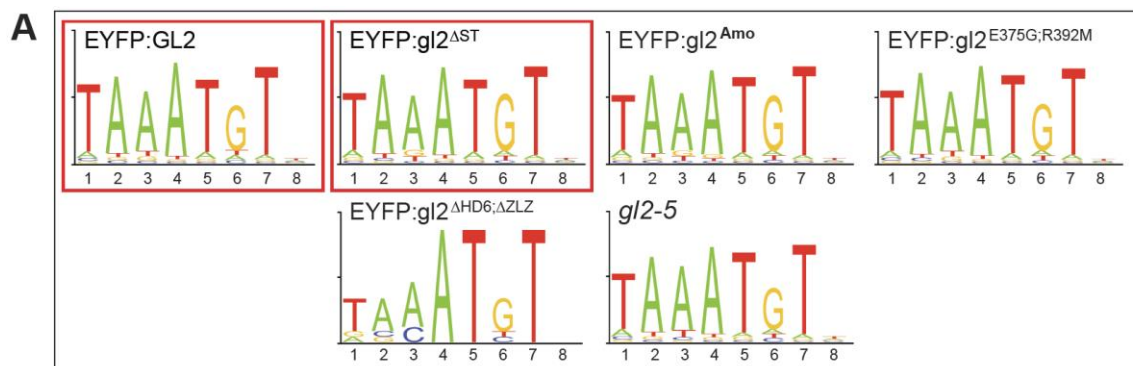


Figure A.5. GL2 binding motif and ChIP-seq peak distance from TSS (related to Figure 2.4, D-G).

(A) Sequence logo representing the information content of the GL2 consensus DNA binding motif TAAATGT(A/T) identified wild-type and mutant ChIP-seq peak regions as indicated in **Figure 2.4D**. Motifs for wild-type GL2 and mutant $gl2^{\Delta ST}$ (red boxes) are highlighted in **Figure 2.4E**. **(B)** The significant ChIP-seq peaks ($FDR < 0.1$) from EYFP-tagged wild-type GL2 and mutant lines were plotted from gene transcription start sites (TSS). The genotypes are ordered according to the number of ChIP-seq peaks from left to right and top to bottom, with *gl2-5* representing the null mutant control that lacks EYFP. The Y axis represents individual genes ordered on the closest first hit to the furthest first hit in any of the samples within a 3 kilobase (kb) region from the TSS. Each dot represents a hit plotted for each gene along the X axis representing the distance from the TSS.

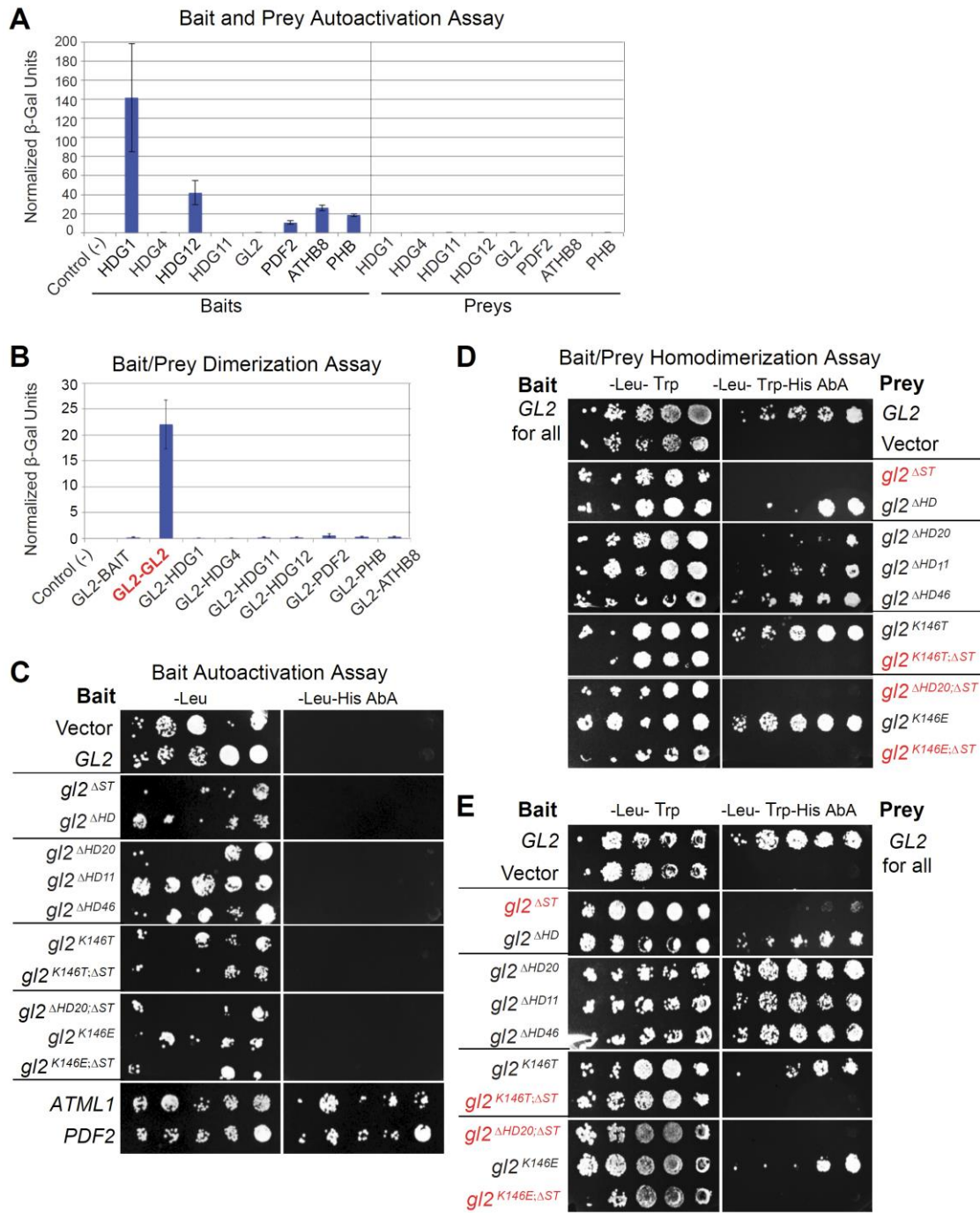


Figure A.6. The START domain is required for GL2 homodimerization (related to Figure 2.5C).

(A and B) Quantitative yeast ONPG assays for GL2 and other members of the HD-Zip TF family. Vector combinations used to co-transform yeast cells are given as prey/bait. Average β -galactosidase values were normalized to a positive control performed at the same time

(pEXPTM32/Krev1 and pEXPTM22/RalGDS-wt, set to 100 β -units). The empty bait vector served as a negative control. Error bars indicate standard deviations for four independent transformants.

(A) GL2 does not display bait or prey autoactivation in Y2H assays. GL2, HDG4 and HDG12 do not exhibit bait autoactivation whereas all prey constructs tested do not exhibit autoactivation.

Note: The HDG12 clone tested in this experiment was found to be truncated, lacking a portion of the START domain as well as C-terminal region. **(B)** GL2 exhibits only homodimerization in Y2H assays. Heterodimerization was tested with other HD-Zip IV family members (HDG1, HDG4, HDG11, HDG12, PDF2) and HD-Zip III members (PHB, ATHB8). **(C)** GL2 expression does not lead to bait autoactivation in contrast to expression of ATML1 or PDF2. GL2 wild-type and gl2 mutant bait constructs, as well as ATML1 or PDF2 bait constructs were transformed into haploid yeast cells. Autoactivation was monitored by growth on selective media (-Leu-His AbA). Lack of autoactivation was observed for wild-type GL2 and each of the mutants, while ATML1 and PDF2 exhibited autoactivation. **(D)** Wild-type GL2 bait fusion protein homodimerizes with preys harboring HD mutations but not START domain deletion. The GL2 bait construct and indicated preys were transformed into yeast strains of opposite mating type. **(E)** Wild-type GL2 prey fusion protein homodimerizes with baits harboring HD mutations but not START domain deletion. The indicated bait constructs and wild-type GL2 prey were transformed into yeast strains of opposite mating type. **(D and E)** Haploid yeast were mated and diploids were grown on -Leu-Trp followed by selection on -Leu-Trp-His AbA to assay 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, and empty bait and prey represent negative controls.

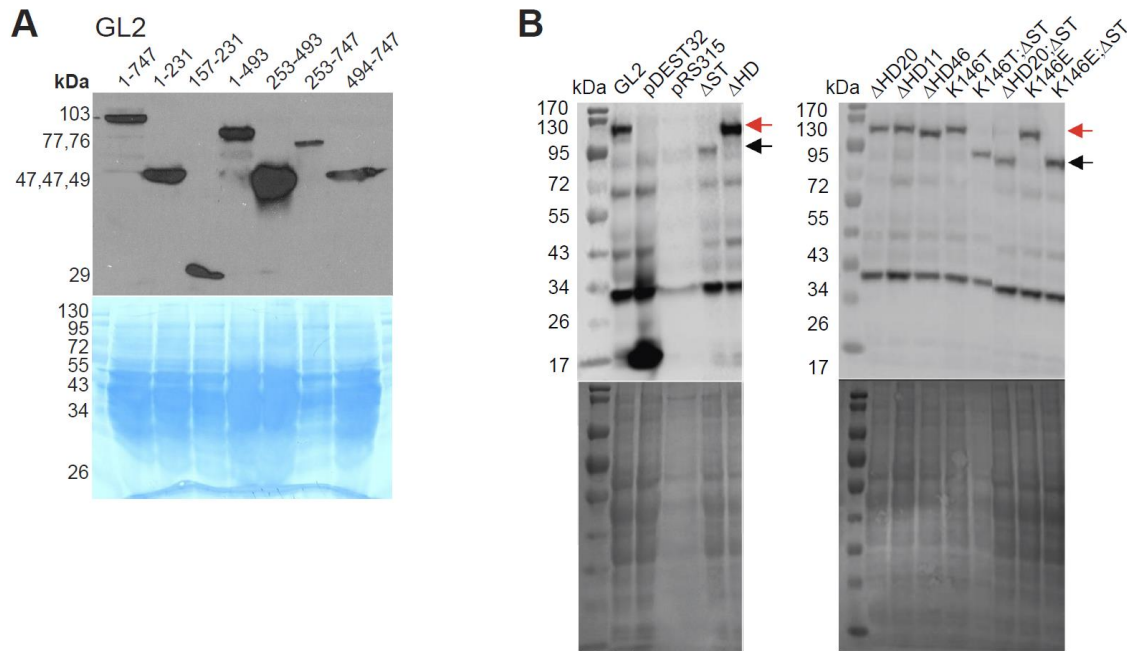


Figure A.7. Protein expression of Bait plasmids in Y2H experiments (related to Figure 2.5).

(A) Western blotting confirms expression of bait proteins in **Figure 2.5A**. GL2 full-length protein and subdomains thereof are fused to the GAL4 DNA Binding Domain (DBD) and the human c-Myc epitope 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). Bottom: Coomassie stained blot indicates loading of protein samples. Expected sizes of the GAL4 DBD-Myc fusion proteins are: 1-747, 103.3 kDa; 1-231, 47.3 kDa; 1-493, 76.6 kDa; 253-493, 47.1 kDa; 494-747, 48.8 kDa; 253-747, 75.7 kDa, 157-231, 29.4 kDa. **(B)** Western blotting with anti-GAL4(DBD) confirms expression of GL2 wild-type and mutant bait proteins in **Figure 2.5C**. Migration of wild-type and HD mutant proteins (red arrows) and START deleted proteins (black arrows) are indicated. pDEST32 is the vector control and pRS315 represents a second negative control. Grayscale image of Coomassie blue staining after Western blotting indicates protein loading (bottom).

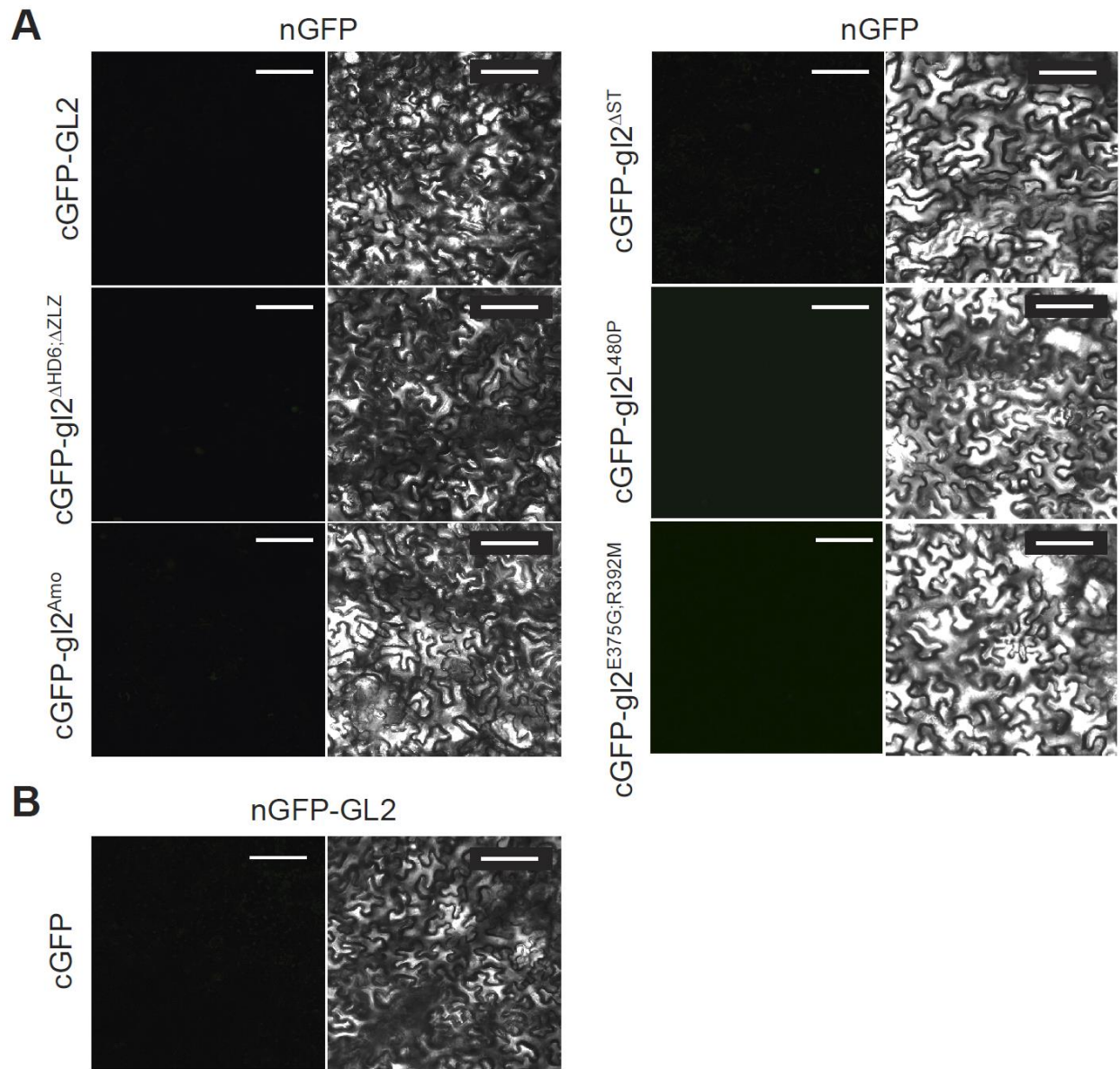


Figure A.8. BiFC negative assays (related to Figure 2.5, D-F).

(**A and B**) Confocal laser scanning microscopy of *N. benthamiana* leaves from BiFC experiments. Little or 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). Bar = 50 μ m.

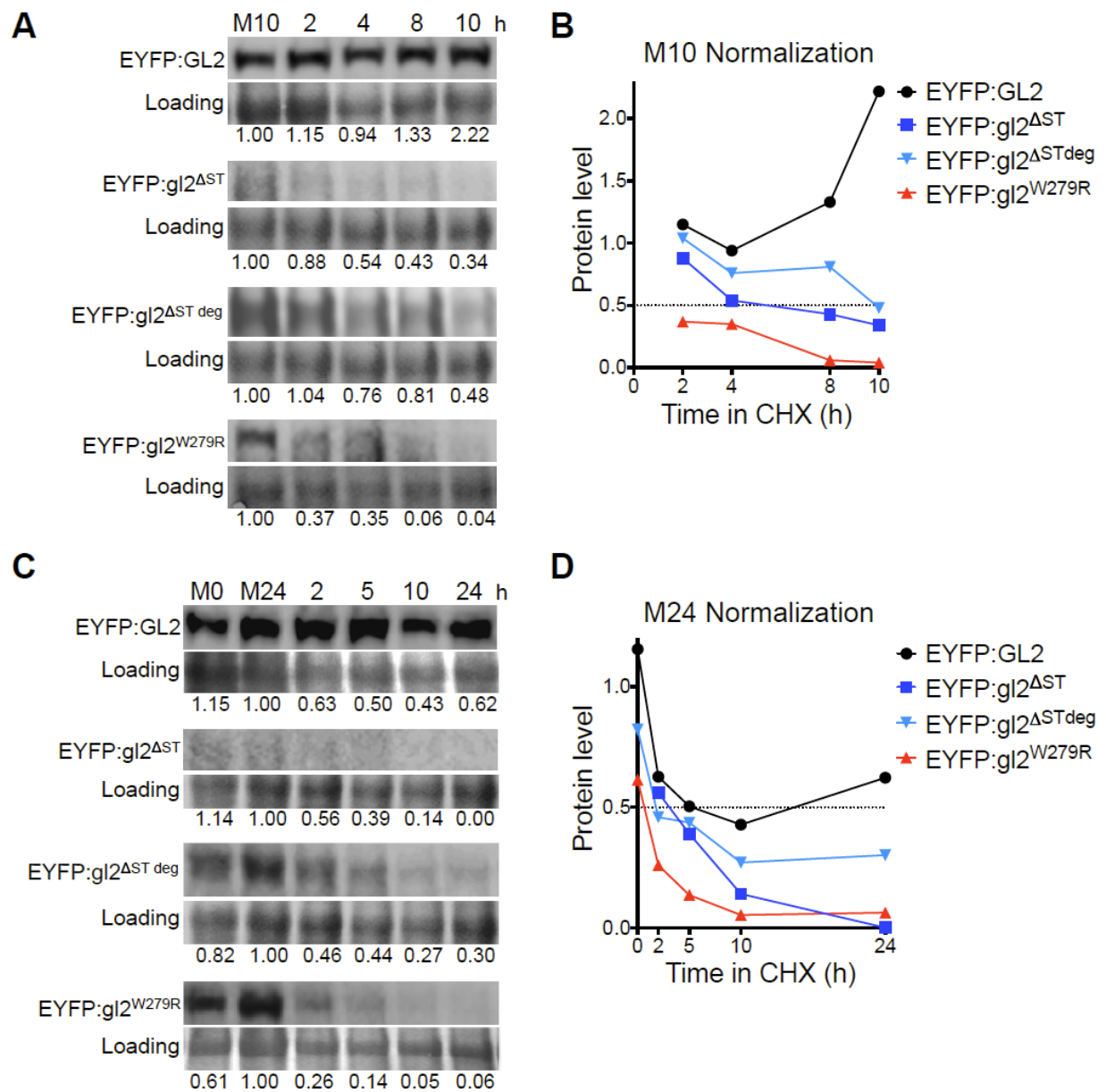


Figure A.9. Cycloheximide chase experiments reveal that the START domain is critical for protein stability of GL2 (related to Figure 2.6).

In comparison to wild-type GL2, START domain mutant protein gl2^{ΔST} (and its degradation product gl2^{ΔST deg}) as well as gl2^{W279R} exhibit a decrease in protein half-life. Two independent experiments are shown in (A-B) and (C-D), respectively. Five-day-old seedlings expressing EYFP-tagged wild-type GL2 or START domain mutant gl2 proteins were treated

with cycloheximide (400 μ M) for a 10 h **(A-B)** or 24 h **(C-D)** time course. **(A and C)** Levels of 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), 10 (M10) or 24 (M24) h are indicated. The $gl2^{\Delta ST \text{ deg}}$ protein was observed as a 65 kDa degradation product of $gl2^{\Delta ST}$. The loading controls for $gl2^{\Delta ST}$ and $gl2^{\Delta ST \text{ deg}}$ are the same since these proteins were detected on the same blot. **(B and D)** Graphs illustrate protein quantification for the experiments depicted in **(A)** and **(C)**, respectively. Values were normalized to loading controls and M10 **(A-B)** or M24 **(C-D)** controls. Protein half-life (0.5) is indicated by intersection with the dotted lines.

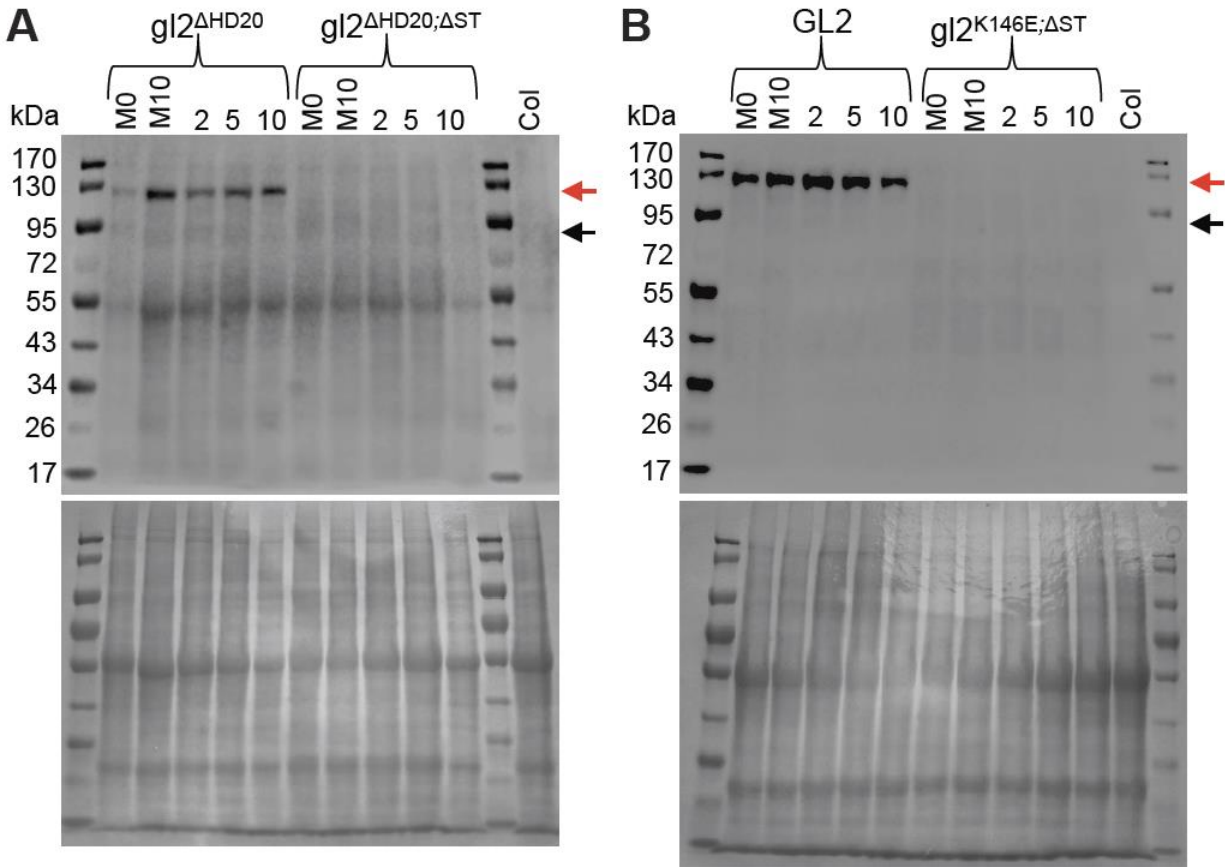


Figure A.10. Cycloheximide chase experiments with double mutants affecting both HD and START domain of GL2 (related to Figure 2.6; Figure A.9).

Five-day-old seedlings expressing EYFP-tagged mutant *gl2* or wild-type GL2 proteins were treated with cycloheximide (400 μ M) for a 10 h time course. Mock treatments with DMSO at 0 (M0) or 10 (M10) h are indicated. Protein levels were monitored by Western blotting. Immunodetection with anti-GFP Ab (top) was followed by Coomassie blue staining for loading controls (bottom, in grayscale). **(A)** The *gl2*^{ΔHD20} protein (expected size ~109 kDa; red arrow) is stable whereas the *gl2*^{ΔHD20;ΔST} protein (expected size ~83 kDa; black arrow) is barely visible or not detected over the 10 h time course. **(B)** Wild-type GL2 protein (expected size ~111 kDa; red arrow) is stable while the *gl2*^{K146E;ΔST} protein (expected size ~83 kDa; black arrow) is barely visible or not detected over the 10 h time course.

Supplemental Tables

Table A.1. EYFP expression is independent of phenotypic rescue (related to Figures 2.1 and 2.2).

Screening of T1 *gl2-5* transgenic plants for wild-type *EYFP:GL2* and mutant *EYFP:gl2* constructs indicated no simple correlation between the trichome rescue phenotype and the intensity of EYFP expression in the nucleus. All wild-type *EYFP:GL2* lines exhibited phenotypic rescue although 5 out of 28 lines exhibited low levels of EYFP expression. Both HD mutants and START deletion mutants failed to rescue but showed varying levels of EYFP expression. The *gl2*^{K146T} mutants displayed a partial phenotype. n = 28-50 T1 plants for each of the constructs.

Genotype	n	Level of phenotypic rescue			EYFP nuclear expression			
		Complete	Partial	None	High	Medium	Low	Undetected
<i>GL2</i>	28	27	1	-	22	1	5	
<i>gl2</i> ^{ΔST}	50			50			30	20
<i>gl2</i> ^{ΔHD46}	37			37	37			
<i>gl2</i> ^{ΔHD20}	30			30	11	14	5	
<i>gl2</i> ^{ΔHD11}	34			34	11	22	1	
<i>gl2</i> ^{K146T}	35	19	7	9		32	3	
<i>gl2</i> ^{K146E}	55			55		41	13	1
<i>gl2</i> ^{K146T;ΔST}	47			47			32	15
<i>gl2</i> ^{K146E;ΔST}	46			46			24	22
<i>gl2</i> ^{ΔHD20;ΔST}	55			55			28	27

Table A.2. Distribution and number of significant peaks from ChIP-seq data (related to Figure 2.4F).

Genomic distribution of ChIP-seq peaks from binding events for wild-type GL2 and mutant versions are indicated (FDR < 0.1). Most of the significant peaks occur in the intergenic and promoter regions. TSS, transcription start site; TTS, transcription termination site; UTR, untranslated region.

Summary of GL2 binding sites in the <i>Arabidopsis thaliana</i> genome									
Sample		Intergenic	Promoter -TSS	Exon	TTS	Intron	5'- UTR	3'- UTR	Total
Wild type	<i>EYFP:GL2</i>	306	246	122	160	83	7	1	925
Mutants	<i>EYFP:gl2^{Amo}</i>	1298	782	668	502	262	10	11	3533
	<i>EYFP:gl2^{ΔST}</i>	1018	564	442	320	177	4	6	2541
	<i>EYFP:gl2^{E375G;R392M}</i>	419	344	186	195	128	10	2	1284
	<i>EYFP:gl2^{ΔHD6;ΔZLZ}</i>	0	0	2	0	0	0	0	2
	<i>gl2-5</i>	34	65	52	49	26	1	0	227

Table A.3. Oligonucleotides used in this study.

I. Primers for constructing mutations in <i>GL2</i>. Nucleotide bases in red denote changes leading to amino acid substitutions.	
Name	5'-3' sequence
gl2_K146T_F	CT CGC CAG GTC ACG TTC TGG TTC
gl2_K146E_F	CT CGC CAG GTC GAG TTC TGG TTC
gl2_K146T or E_R	G GGC CAG CCC TAG TTG CTT GC
gl2_ΔHD11 or 46_F	GCT ATT CAA GAA CGG CAC GAG AAC
gl2_ΔHD20_F	CAG CAG CTG AGC AAG CAA C
gl2_ΔHD20 or 46_R	ATC GGT GGT GTG ACG ATG ATA C
gl2_ΔHD11_R	CTT GAC CTG GCG AGG GG
GL2_START_Δ_F	[Phos] GTC TTC TTC ATG GCT ACC AAC GTC CCC ACC (Schrack et al., 2014)
GL2_START_Δ_F	[Phos] GAG GGC AAA GAC GCC CGT GTA GAA ATC G (Schrack et al., 2014)
GL2_HD_Δ_F	[Phos] GAA CGG CAC GAG AAC TCC CTG CTC
GL2_HD_Δ_R	[Phos] CTT ATT AGT GCC CTT GTT TCC AGC TGC G
GL2_ZLZ_Δ_F	[Phos] TCA TGC TCC GAC GAT CAA GAA CAC CG
GL2_HD-ZLZ_Δ_R	[Phos] TGT GCG GCG GTT TTG GAA CCA GAA CTT G
GL2_W279R_F	A GGC GAA CCT ATG AGG CTC CGC AGC G
GL2_W279R_R	GA GGT GGC CAT CTT CTG GAG TTC AAG GGT G
GL2_E375G_F	CAG CTG ATG TTC GGA GCG ATG CAG CTG CTC ACT
GL2_E375G_R	AGT GAG CAG CTG CAT CCC TCC GAA CAT CAG CTG
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 (Wojciechowska et al, 2021)
GL2_L480P_R	C GCA ATG GAG CTG AGG GGT GGC GAC C (Wojciechowska et al, 2021)
GL2_A258P_F	GAAGTCCCGTATT CCC GAGATTTCTAACCG
GL2_A258P_R	TCGAGGGCAAAGACGCCCGTGTAGAAATC
GL2_I289N_F	GACTGGCCGTGAG AAT CTCAACTACGATG
GL2_I289N_R	TCAACGCTGCGGAGCCACATAGGTTTCG
GL2_E294Q_F	TCTCAACTACGAT CAG TACCTCAAGGAG
GL2_E294Q_R	ATCTCACGGCCAGTCTCAACGCTG
GL2_M271I_F	TGAACTCCAGAAG ATT GCCACCTCAGGC
GL2_M271I_R	AGGGTGGCTCGGTTAGAAATCTCGGC
II. 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_B9_5	GTC AAC ACC GGT TTG GCC
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
proGL2::EYFP:GL2_seq_B13	GG AGT GCG AGG AGA AGA GGG
proGL2::EYFP:GL2_seq_B14	GTC GTG ATC CTC ACC CTC C
III. Primers for amplification of GL2 cDNA from pENTR/D-TOPO vector to clone into binary vector SR54 and for sequence verification.	
EYFP:GL2_F	TCA AGC TTC GAA TTC TGC AGT CGA CAT GTC AAT GGC CGT CGA CAT G
p1300_GL2_R	CCA CCG CGG TGG AGC TCG TCA TTA GCA ATC TTC GAT TTG
IV. Primers for genotyping mutants	
To confirm <i>gl2-5</i> mutation	
GL2_F_112	ATGTCAATGGCCGTCGACATGTC (Wang et al., 2007; Khosla et al., 2014)
En8130	GAGCGTCGGTCCCCACACTTCTATAC (Baumann, 1998; Khosla et al., 2014)
To confirm transgene (<i>GL2</i> cDNA in <i>proGL2:EYFP:GL2</i>)	
SR54_GL2_747F	GTC GAC ATG TCA ATG GCC GTC G
SR54_GL2_747R	AGC TCG TCA TTA GCA ATC TTC GAT TTG
V. Primers for constructing K107T, K107E and START deletion in PDF2 and ATML1. Nucleotide bases in red denote changed bases underlying respective amino acid substitutions.	
pdf2_K107T_F	CT CTT CAA GTT AAG TTT TGG TTC C
pdf2_K107E_F	CT CTT CAA GTT GAG TTT TGG TTC C
pdf2_K107E_R	G CTC TAA ATT GAG ATC ACG GC
pdf2ΔSTART_F	GCT AGC TCC ATG GCC AGC
pdf2ΔSTART_R	AGA AGG AAT CGA AAC TGA CCT C
atml1ΔSTART_F	GCC AGT TCC ATG GCC AGC
atml1ΔSTART_R	AGA AGG TAT CGA AAC CGA CCT C
VI. Primers for confirming Gateway cloning of pIX-HALO constructs	
pIX-Halo_Seq	GCCTAACTGCAAGGCTGTGG
P1300_PDF2_R	ACCGCGGTGGAGCTCG CTA CGC TCC TCC TCC AAC
P1300_ATML1_R	CACCGCGGTGGAGCTCG TTA GGC TCC GTC GCA GGC C
VII. Oligonucleotides used in EMSA. Three repeats of the L1 motif are depicted in red.	
L1_EMSA_F_Cy3	/5Cy3/GGG TACATTTA TACATTTA TACATTTA AAT

L1_EMSA_R	ATT TAAATGTA TAAATGTA TAAATGTA CCC
VIII. Primers for restriction cloning into pGBK and pGAD vectors (Clontech Matchmaker)	
GL2_747aa_EcoRI_F	GAC AGG <u>GAA TTC</u> ATG TCA ATG GCC GTC GAC ATG TCT TCC
GL2_747aa_BamHI_R	TGCTGT <u>GGATCC</u> TCATCA GCA ATC TTC GAT TTG TAG AC
GL2_493aa_BamHI_R	TTGGT <u>GGATCC</u> CTA GGT AGC CAT GAA GAA GAC AAG GCG
GL2_253aa_EcoRI_F	GTCTTT <u>GAATTC</u> GAG AAG TCC CGT ATT GCC GAG ATT TC
GL2_494aa_EcoRI_F	TTCATG <u>GAATTC</u> AAC GTC CCC ACC AAA GAC TCT CTC GG
GL2_157aa_EcoRI_F	ACCGCCGCACAG <u>GAATTC</u> AAG GCT ATT CAA GAA CGG C
GL2_231aa_BamHI_R	ATCGTC <u>GGATCC</u> TCA AGC CTG CAG GGG ATA GGG
IX. Primers for sequence verification of Y2H constructs	
BD_seq_F	AACCGAAGTGCGCCAAGTGTCTG
AD_seq_F	TATAACGCGTTTGGAATCACT
BD & AD_seq_R	AGCCGACAACCTTCATTGGAGAC
X. Primers for cloning in pENTRTM/D-TOPO (Start and 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
XI. Primers for sequence verification of BiFC Gateway constructs	
GFP-NXGW_seq	GACAAGCAGAAGAACGGCAT
pro35S_pK7WG2_seq	GGTGGCACCTACAAATGCCATC

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

The number of reads, alignment statistics, and total mapped reads to the *Arabidopsis thaliana* genome are indicated for individual input and immunoprecipitated (IP) samples. The total mapped reads of merged replicate IP samples are shown at the bottom. The ChIP-seq assays were performed with seedlings expressing EYFP-tagged GL2 wild-type and mutant proteins, including three START domain mutants (gl2_ΔST (listed as gl2_dSTART in the GEO data GSE186860), gl2_Amo, and gl2_E375G;R392M) as well as a mutant deleted for the six terminal amino acids of the HD and the entire leucine zipper (gl2_ΔHD6;ΔZLZ (listed as gl2_dZip in the GEO data GSE186860)). **See xls file – Sheet 1.** See **ChIP-seq Data Set 2** for additional information.

ChIP-seq Data Set 2. Functional descriptions of GL2 targets from ChIP-seq data

Arabidopsis thaliana genes with significant ($\text{FDR} \leq 0.10$) GL2 binding in the promoter region along with corresponding functional annotation obtained from the Arabidopsis Information Resource (TAIR) database (Berardini et al., 2015). ChIP-seq assays were performed with seedlings expressing EYFP-tagged wild-type GL2 and mutant gl2 proteins, including three START domain mutants (gl2_ΔST, gl2_Amo, and gl2_E375G;R392M) as well as a mutant deleted for the six terminal amino acids of the HD and the entire leucine zipper (gl2_ΔHD6;ΔZLZ). The *gl2-5* genotype lacks the EYFP tag and contains the *gl2-5* null mutation and served as a negative control. **See xls file – Sheet 2.** See **ChIP-seq Data Set 1** for additional information.

References

- Alpy, F., and Tomasetto, C.** (2014). START ships lipids across interorganelle space. *Biochimie* **96**, 85-95.
- Baumann, E., Lewald, J., Saedler, H., Schulz, B., and 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.
- Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E.** (2015). The *Arabidopsis* information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* **53**, 474-485.
- 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. *The Plant Cell* **26**, 2184-2200.
- Robert, X., and Gouet, P.** (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324.
- Roostae, 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.
- Schrack, 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.
- Wang, S., Kwak, S.H., Zeng, Q., Ellis, B.E., Chen, X.Y., Schiefelbein, J., and Chen, J.G.** (2007). TRICHOMELESS1 regulates trichome patterning by suppressing GLABRA1 in *Arabidopsis*. *Development* **134**, 3873-3882.
- Wojciechowska, I., Mukherjee, T., Knox-Brown, P., Hu, X., Khosla, A., Mathews, G.L., Thompson, K.A., Peery, S.T., Szlachetko, J., Thalhammer, A., Hinch, 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

Appendix B - Chapter 3 Supplemental Data

Supplemental Figures

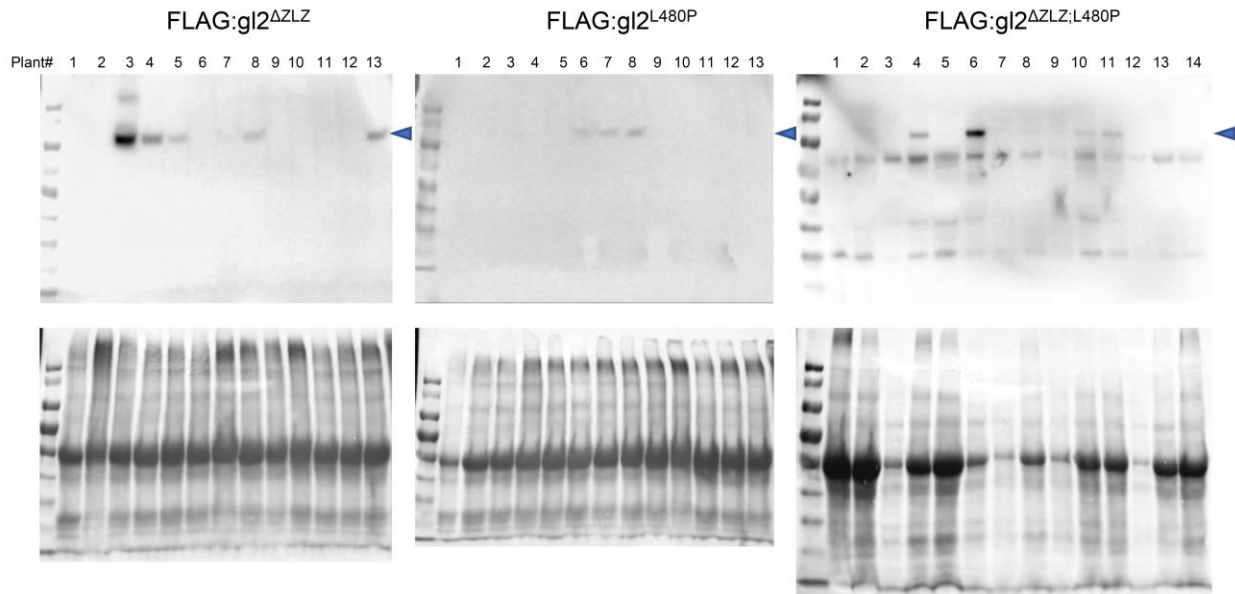


Figure B.1. Western blot analysis of leaves from ~one-month-old T1 progeny of *FLAG:gl2^{ΔZLZ}*, *FLAG:gl2^{L480}* and *FLAG:gl2^{ΔZLZ;L480P}* lines confirming the protein expression (Related to Figure 3.2).

Grayscale image of Coomassie blue staining after Western blotting indicates protein loading (bottom). Molecular weights (kDa): FLAG:GL2, 88.0; FLAG:gl2^{ΔZLZ} and FLAG:gl2^{ΔZLZ;L480P}, 82.0.

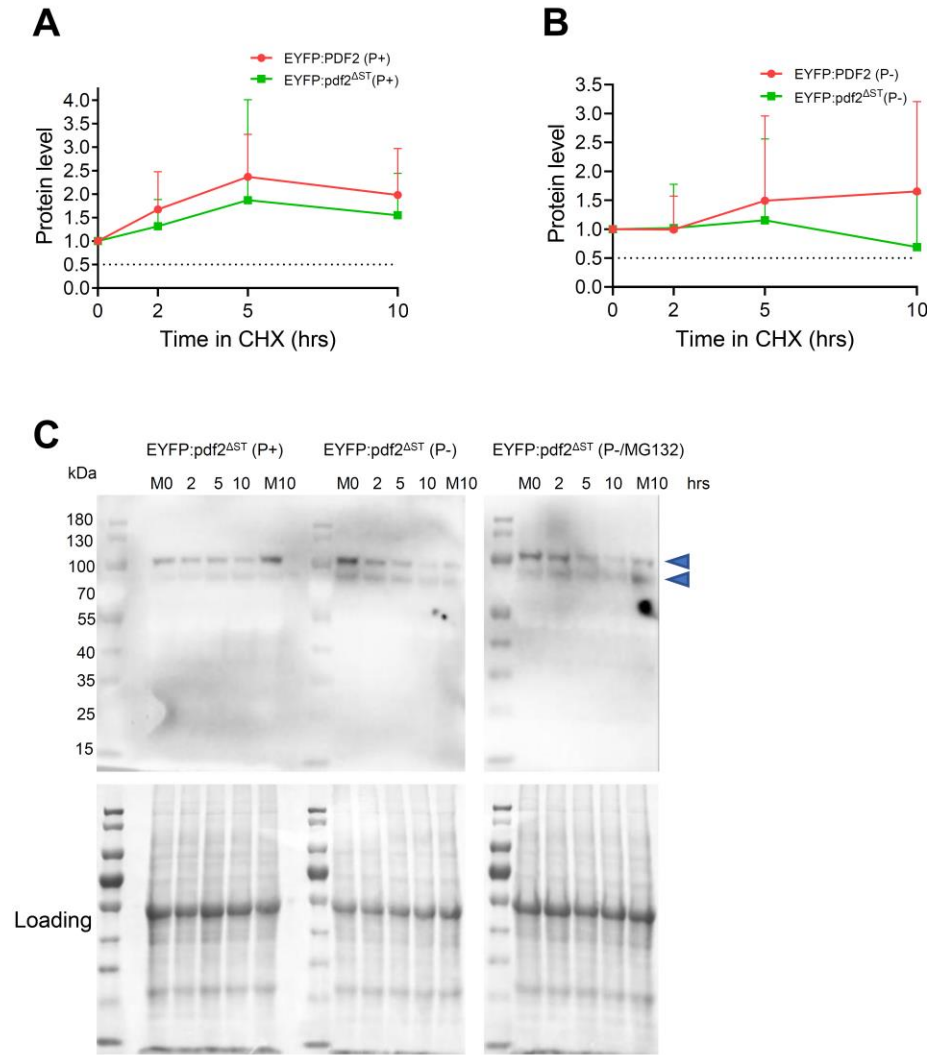


Figure B.2. PDF2 is destabilized in absence of its START domain (Related to Figure 3.3).

(A-B) Graphs corresponding to **Figures 3.3G-3.3J** comparing the half-lives of EYFP:PDF2 and EYFP:pdf2^{ΔST} under Pi sufficiency (P+) **(A)** and limitation (P-) **(B)**. Compared to EYFP:PDF2, EYFP:pdf2^{ΔST} exhibited a reduction in protein levels that was enhanced under Pi limitation. **(C)** Full-size blot images corresponding to **Figures 3.3I** and **3.3K** showing the degradation product (lower band) of EYFP:pdf2^{ΔST} (upper band) under all growth conditions. Grayscale image of Coomassie blue staining after Western blotting indicates protein loading (bottom).

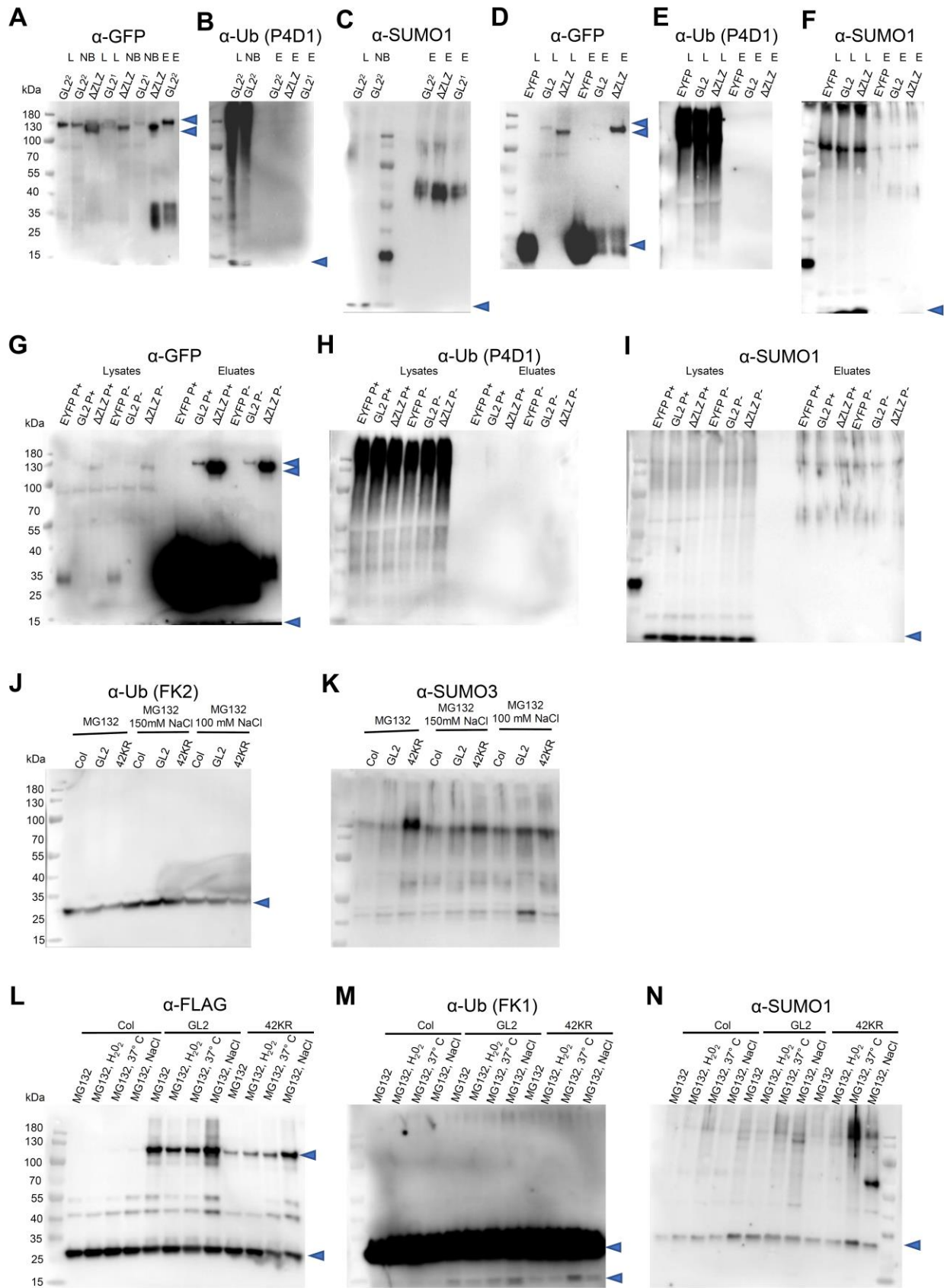


Figure B.3. Detection of ubiquitination and SUMOylation in epitope tagged wild-type and mutant gl2 proteins (Related to Figure 3.4).

(A-I) Immunoblotting of *EYFP* (negative control), *EYFP:GL2* and *EYFP:gl2^{ΔZZ}* samples immunoprecipitated with GFP-trap magnetic agarose beads using anti-GFP **(A, D, G)**, P4D1 anti-Ub **(B, E, H)** and anti-SUMO1 **(C, F, I)** antibodies. **D-F** correspond to samples obtained by harsh washing during elution step of immunoprecipitation. **(J-K)** Immunoblotting of samples used in **Figures 3.4D-F** with FK2 anti-Ub **(J)** and anti-SUMO3 **(K)** antibodies. **(L-N)** Immunoblotting of *Col* (WT), *FLAG:GL2* and *FLAG:gl2^{42KR}* samples immunoprecipitated with FLAG-trap using anti-FLAG **(L)**, FK1 anti-Ub **(M)** and anti-SUMO1 **(N)** antibodies. Conditions used to grow the seedlings are specified. Bands representing immunoprecipitated proteins, Fab-fragment of FLAG-trap and free Ub or SUMO molecules at the bottom of corresponding blots are indicated. L, lysate (input); NB, supernatant (flow-through/non-bound fraction); E, eluate (bound fraction).

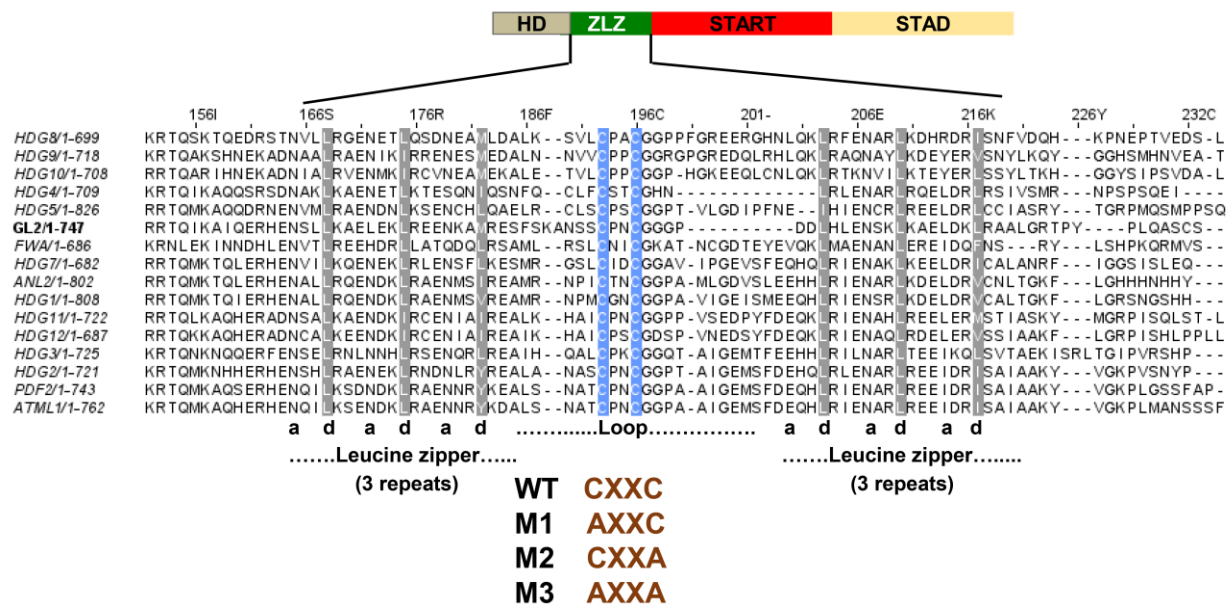


Figure B.4. Alignment of ZLZ from Arabidopsis HD-Zip IV TFs illustrating the conservation of residues in the leucine zippers and the loop (Related to Figure 3.6).

M1, M2 and M3 indicate the mutants affecting the invariant cysteine(s) present in loop region.

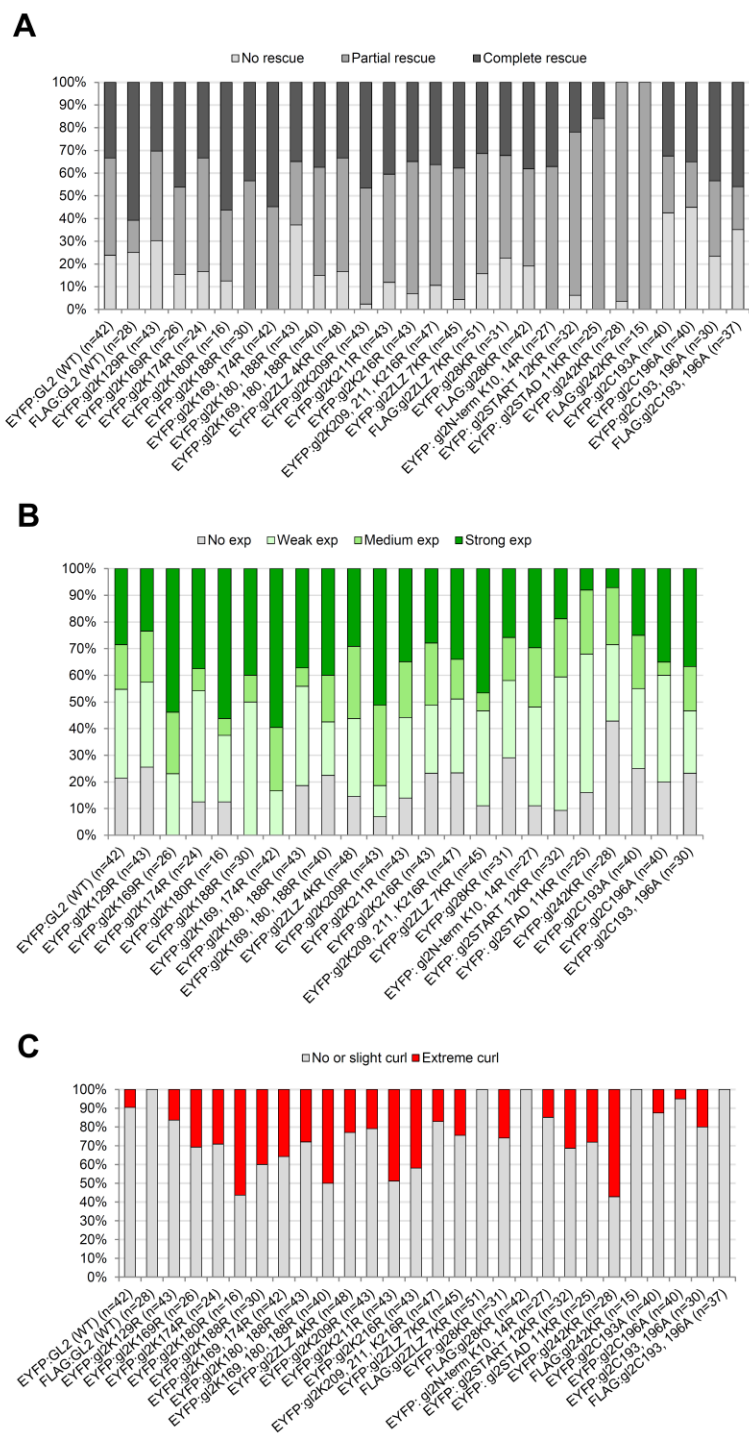


Figure B.5. Screening of T1 transformants for rescue of null phenotype, EYFP expression and abaxial leaf curling (Related to Figure 3.7).

Graphs showing (A) rescue of trichome phenotype, (B) EYFP expression and (C) occurrence of abaxially curled leaves in T1 progeny of WT plants and GL2 PTM mutants.

Supplemental Tables

Table B.1. List of oligonucleotides used in this study

I. Primers used to generate PTM mutations (K-R and C-A) in ZLZ/HD of GL2. Nucleotide bases in red denote changes leading to amino acid substitutions.	
gl2 K129R F	CAT CCG GAC GAG A GG CAA AGA CAG CAG
gl2 K129R R	CTG CTG TCT TTG C CT CTC GTC CGG ATG
gl2_K169R F	A GG GCG GAA CTA GAG AAG CTG
gl2_K169R R	GAG CAG GGA GTT CTC GTG C
gl2 K174R F	CG GAA CTA GAG A GG CTG CGA GAG GAA AAC
gl2 K174R R	GTT TTC CTC TCG CAG C CT CTC TAG TTC CG
gl2 K180R F	GA GAG GAA AAC A GA GCC ATG AGG GAG TC
gl2 K180R R	GA CTC CCT CAT GGC T CT GTT TTC CTC TC
gl2 K188R F	GAG TCT TTT TCC A GG GCT AAT TCC TCC TG
gl2 K188R R	CA GGA GGA ATT AGC C CT GGA AAA AGA CTC
gl2_K209R F	A GA CTG AAA GCC GAG CTC G
gl2_K209R R	GGAGTTTTCGAGGTGGAGATC
gl2_K211R F	A GA GCC GAG CTC GAT AAG C
gl2_K211R R	CAGTTTGGAGTTTTCGAGGTG
gl2_K216R F	A GG CTT CGT GCA GCT CTT G
gl2_K216R R	ATCGAGCTCGGCTTTCAGTTTG
gl2_K209R_K211R F	A GA CTG A GA GCC GAG CTC
gl2_K211R_K216R F	A GA GCC GAG CTC GAT A GG
gl2_K209R_K211R_K216R R	CAGT CT GGAGTTTTCGAGGTG
gl2_C193A_F	GCA CCC AAC TGC GGA GGA GG
gl2_C193A_R	GGAGGAATTAGCCTTGAAAAAGACTCC
gl2_C196A_F	GCT GGA GGA GGC CCC GAT GAT C
gl2_C196A_R	GTTGGGGCAGGAGGAATTAGCCTTGG
gl2_C193A_C196A_R	GTT GGG G GC GGA GGA ATT AGC C
II. Primers used to derive domain wise K-R mutations from <i>gl2</i>^{42KR}	
GL2_N-term KR-F	GGATTTCTGCGGAGAGAG
GL2_N-term KR-R	CTCTTCTGCCCAGGAAATCC
GL2_START KR_5'-F	GCTCTCTCGATTCTACACGG
GL2_START KR_5'-R	CCGTGTAGAAATCGAGAGAGC
GL2_START KR_3'-F	CCTTGTCTTCTTCATGGCTACC
GL2_START KR_3'-R	GGTAGCCATGAAGAAGACAAGG
EYFP-GL2-F	TCAAGCTTCGAATTCTGCAGTCGACATG TCA ATG GCC GTC GAC ATG
p1300-GL2-R	CCACCGCGGTGGAGCTCG TCA TTA GCA ATC TTC GAT TTG
III. Primers used to generate ZLZ, START double mutants (using START domain	

mutants as templates)	
GL2_ZLZ_Δ_F	[Phos] TCATGCTCCGACGATCAAGAACACCG
GL2_ZLZ_Δ_new_R	[Phos] TTGAATAGCCTTGATCTGTGTGCGGCG
IV. Primers used to generate FLAG-tagged GL2 (pB7_HFC3_0 used to amplify the 6x His 3x FLAG tag)	
SR54_FLAG_F	ATTAA GCAAA GAAGC TAG AG ATG CAT CAC CAT CAC CAT CAC
SR54_FLAG_R_65 Ultra	T GGG TTG TTT GGA AGA CAT GTC CAC GGC CAT TGA CAT GTC GAC CTT GTC ATC GTC ATC CTT GTA G
V. Primers used to amplify GL2 cDNA from pENTR/D-TOPO vector to clone into binary vector SR54	
EYFP-GL2-F	TCAAGCTTCGAATTCTGCAGTCGACATG TCA ATG GCC GTC GAC ATG
FLAG-GL2-F	C AAG GAT GAC GAT GAC AAG GTC GACATG TCA ATG GCC GTC GAC ATG
p1300-GL2-R	CCACCGCGGTGGAGCTCG TCA TTA GCA ATC TTC GAT TTG
p1300-GL2-KpnI_R	CCACCGCGGTGGAGCTCGGTACC TCA TTA GCA ATC TTC GAT TTG
VI. Primers used for sequence verification of <i>gl2</i> mutant constructs	
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_B9_5	GTC AAC ACC GGT TTG GCC
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
proGL2::EYFP:GL2_seq_B13	GG AGT GCG AGG AGA AGA GGG
proGL2::EYFP:GL2_seq_B14	GTC GTG ATC CTC ACC CTC C
SR54_EYFP_F_seq	GTC CTG CTG GAG TTC GTG
VII. Primers used for genotyping of <i>gl2</i> mutants	
SR54_GL2_747F	GTC GAC ATG TCA ATG GCC GTC G
SR54_GL2_747R	AGC TCG TCA TTA GCA ATC TTC GAT TTG
SR54_FLAG_F	ATT AAG CAA AGA AGC TAG AGA TGC ATC ACC ATC ACC ATC AC

Table B.2. Strategies used to generate PTM mutants with double/triple/multiple amino acids substitutions

GL2 mutations (K to R)	Template	Primers
Mutations generated by site-directed mutagenesis		
K169, 174R	K169R	gl2 K174R F, gl2 K174R R
K180, 188R	K188R	gl2 K180R F, gl2 K180R R
K169, 180, 188R	K180, 188R	gl2_K169R F, gl2_K169R R
K169, 174, 180, 188R	K169, 180, 188R	gl2 K174R F, gl2 K174R R
K209, 211R	K211R	gl2_K209R_K211R F, gl2_K209R R
K211, 216K	K216R	gl2_K211R_K216R F, gl2_K211R R
K209, 216R	K216R	gl2_K209R F, gl2_K209R R
K209, 211, 216R	K211, 216R	gl2_K209R_K211R F, gl2_K209R R
K169, 174, 180, 188, 209, 211, 216R (7KR)	K169, 174, 180, 188R	gl2_K211R_K216R F, gl2_K209R_K211R_K216R R
K129, 169, 174, 180, 188, 209, 211, 216R (8KR)	169, 174, 180, 188, 209, 211, 216 (7KR)	gl2 K129R F, gl2 K129R R
C193, 196A	WT <i>GL2</i>	gl2_C196A_F, gl2_C193A_C196A_R
Mutations derived from <i>gl2</i>^{42KR}		
N-term KR (two fragments assembled)	<i>gl2</i> ^{42KR}	EYFP-GL2-F, GL2_N-term KR-R
	WT <i>GL2</i>	GL2_N-term KR-F, p1300-GL2-R
START KR (three fragments assembled)	WT <i>GL2</i>	EYFP-GL2-F, GL2_START KR_5'-R
	<i>gl2</i> ^{42KR}	GL2_START KR_5'-F, GL2_START KR_3'-R
	WT <i>GL2</i>	GL2_START KR_3'-F, p1300-GL2-R
STAD KR (two fragments assembled)	WT <i>GL2</i>	EYFP-GL2-F, GL2_START KR_3'-R
	<i>gl2</i> ^{42KR}	GL2_START KR_3'-F, p1300-GL2-R

Appendix C - Chapter 5 Supplemental Data

Supplemental Figures

S1

Arabidopsis SUMO paralogs



Figure C.1. Amino acid alignment of the AtSUMO1, 2, 3 and 5 protein sequences (Related to Figure 5.5).

S2

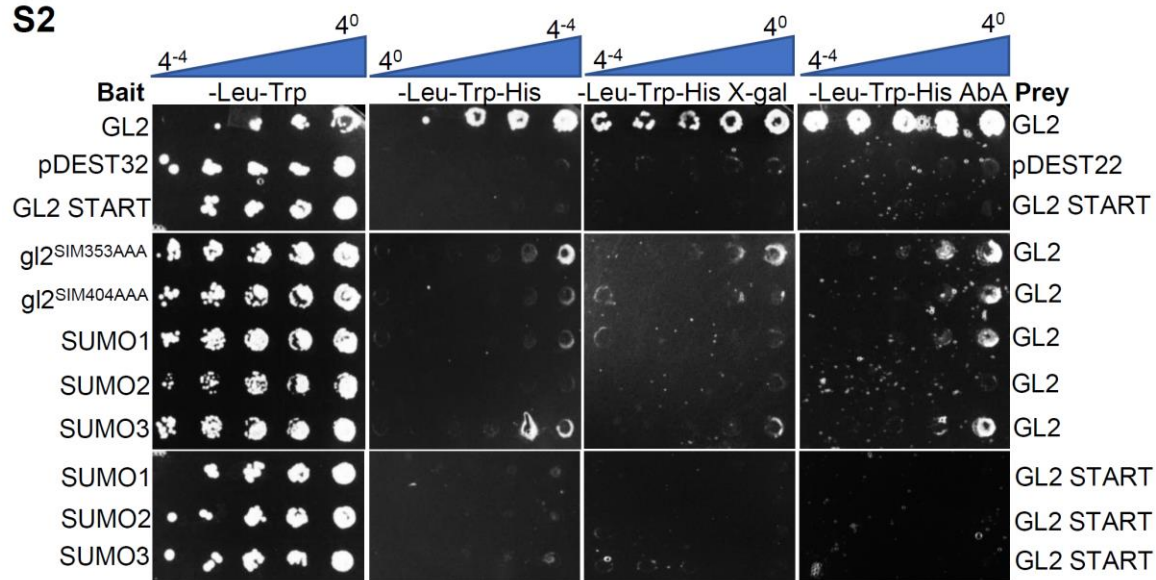


Figure C.1. Detection of GL2 protein interactions mediated by its predicted SIMs (Related to Figure 5.5).

Y2H screening of diploid yeast cells expressing bait constructs with $gl2^{SIM}$ mutants or SUMO isoforms and prey constructs with full-length GL2 or the GL2 START domain on permissive versus selective media. Unlike wild-type GL2, the $gl2^{SIM}$ mutants failed to show homodimerization. No interactions were detected between full-length GL2 or the GL2 START domain and any of the AtSUMO isoforms (1-3). Empty prey and bait vectors represent negative controls. Leu, Leucine; Trp, Tryptophan; His, Histidine; AbA, Aureobasidin A.

S3

α -SUMO3 recognizes SUMO2

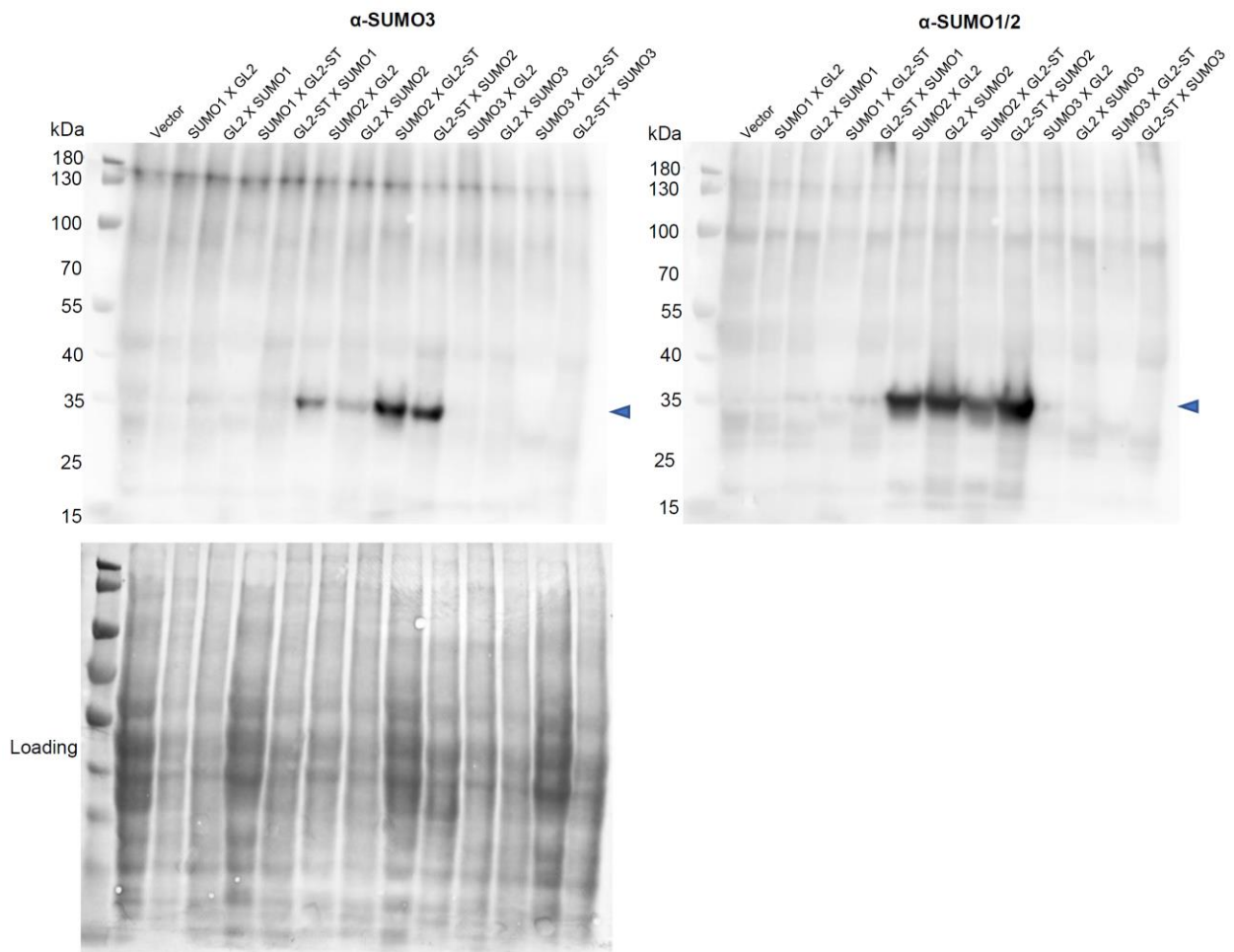


Figure C.2. Western blot showing the expression of AtSUMO1, AtSUMO2, and AtSUMO3 fusion proteins in diploid yeast from Y2H assays using α -SUMO3 antibodies (Related to Figure 5.5).

No proteins were visible for SUMO3 but SUMO2 was detected due to its cross-reactivity with α -SUMO3 antibody. The stripped blot was reprobed with the α -SUMO1/2 antibody, which also detected SUMO2. Bait and prey proteins were expressed as GAL4 DBD and GAL4 AD fusions in Y2HGold and Y187, respectively.

S4 Bait Autoactivation Assay

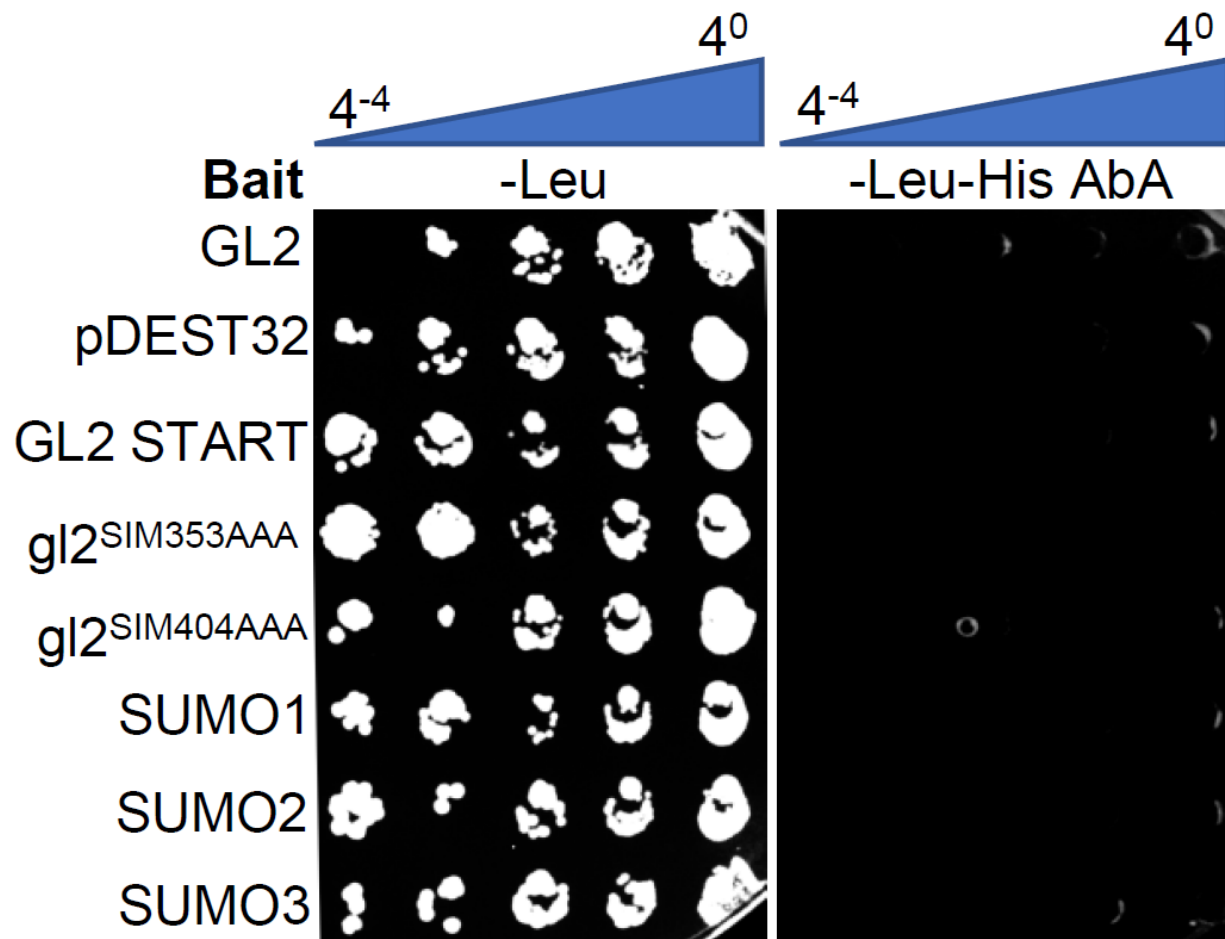


Figure C.3. Bait autoactivation assay corresponding to the Y2H experiments (Related to Figure 5.5).

Haploid yeast cells (Y2HGold) expressing bait constructs were grown on permissive (-Leu) versus selective (-Leu-His AbA) media to monitor reporter gene activity. None of the proteins expressed in the bait constructs, including full-length GL2, the GL2 START domain, gl2^{SIM353}, gl2^{SIM404} and the AtSUMO1-3 isoforms resulted in bait autoactivation.

Supplemental Tables

Table C.1. List of oligonucleotides used in this study

I. Primers used for generating GL2 SIM mutants. Nucleotides in red indicate amino acid substitutions.	
Name	5'-3' sequence
GL2_SIM 353_F	ATG CTG CCC GGC AAG GCG AAG G
GL2_SIM 353_R	CGG CCG TTG CAG CCT TTG AGA TCA AGC
GL2_SIM 404_F	CCT CGG TCT CCG TGG AGG ACA GC
GL2_SIM 404_R	CGT CCG CTG CTG CCC ATT TCT CAG GGC
II. Primers used for amplifying <i>gl2</i>^{SIM353} and <i>gl2</i>^{SIM404} from pENTR/D-TOPO vector to clone into binary vector SR54.	
Name	5'-3' sequence
EYFP_GL2_F	TCA AGC TTC GAA TTC TGC AGT CGA CAT GTC AAT GGC CGT CGA CAT G
p1300_GL2_R	CCA CCG CGG TGG AGC TCG TCA TTA GCA ATC TTC GAT TTG
FLAG-GL2-F	CAA GGA TGA CGA TGA CAA GGT CGA CAT GTC AAT GGC CGT CGA CAT G
III. Primers used for sequence verification of D-TOPO/SR54 SIM mutant constructs and genotyping of homozygous plants	
Name	5'-3' sequence
M13F	GTA AAA CGA CGG CCA GT
M13R	CAG GAA ACA GCT ATG AC
proGL2::EYFP:GL2_seq_B8	GGC TAA TTC CTC CTG CCC C
proGL2::EYFP:GL2_seq_B9_5	GTC AAC ACC GGT TTG GCC
proGL2::EYFP:GL2_seq_B13	GG AGT GCG AGG AGA AGA GGG
SR54_EYFP_F_seq	GTC CTG CTG GAG TTC GTG
SR54_GL2_747F	GTC GAC ATG TCA ATG GCC GTC G
SR54_GL2_747R	AGC TCG TCA TTA GCA ATC TTC GAT TTG
SR54_FLAG_F	ATT AAG CAA AGA AGC TAG AGA TGC ATC ACC ATC ACC ATC AC
IV. Primers for TOPO cloning of SUMO genes used in Y2H assays	
SUMO1_cDNA_F	CAC CAT GTC TGC AAA CCA GG
SUMO1_cDNA_R	TCA GGC CGT AGC ACC ACC
SUMO2_cDNA_F	CAC CAT GTC TGC TAC TCC GG
SUMO2_cDNA_R	CTA AAA GCA GAA GAG CTT CAG GCC
SUMO3_cDNA_F	CAC CAT GTC TAA CCC TCA AGA TG
SUMO3_cDNA_R	TTA AAG CCC ATT ATG ATC GAA AAG CA
V. Primers used for sequence verification of constructs used for Y2H assays	
T3	ATT AAC CCT CAC TAA AGG G
M13F	GTA AAA CGA CGG CCA GT
Bait_and_Prey_Rev	AGC CGA CAA CCT TGA TTG GAG AC
proGL2::EYFP:GL2_seq_B8	GGC TAA TTC CTC CTG CCC C

Table C.2. List of constructs generated in this study

Gene	Vector	Template	Primers used	Procedure
<i>gl2^{SIM353}</i>	pENTR/D-TOPO	<i>GL2</i>	GL2_SIM 353_F and GL2_SIM 353_R	Q5 site-directed mutagenesis
<i>gl2^{SIM404}</i>	pENTR/D-TOPO	<i>GL2</i>	GL2_SIM 404_F and GL2_SIM 404_R	Q5 site-directed mutagenesis
<i>EYFP:gl2^{SIM353}</i>	SR54	<i>gl2^{SIM353}</i>	EYFP_GL2_F and p1300_GL2_R	Q5 HiFi PCR and NEBuilder Assembly
<i>FLAG:gl2^{SIM353}</i>	SR54	<i>gl2^{SIM353}</i>	FLAG-GL2-F and p1300_GL2_R	Q5 HiFi PCR and NEBuilder Assembly
<i>EYFP:gl2^{SIM404}</i>	SR54	<i>gl2^{SIM404}</i>	EYFP_GL2_F and p1300_GL2_R	Q5 HiFi PCR and NEBuilder Assembly
<i>FLAG:gl2^{SIM404}</i>	SR54	<i>gl2^{SIM404}</i>	FLAG-GL2-F and p1300_GL2_R	Q5 HiFi PCR and NEBuilder Assembly
<i>(CACC)SUMO1</i>	pENTR/D-TOPO	<i>SUMO1</i>	SUMO1_cDNA_F and SUMO1_cDNA_R	TOPO cloning
<i>(CACC)SUMO2</i>	pENTR/D-TOPO	<i>SUMO2</i>	SUMO2_cDNA_F and SUMO2_cDNA_R	TOPO cloning
<i>(CACC)SUMO3</i>	pENTR/D-TOPO	<i>SUMO3</i>	SUMO3_cDNA_F and SUMO3_cDNA_R	TOPO cloning
<i>GL2</i>	pDEST32/pDES T22	<i>TOPO:GL2</i>	N/A	Gateway cloning
<i>GL2 START</i>	pDEST32/pDES T22	<i>TOPO:GL2 START</i>	N/A	Gateway cloning

$gl2^{SIM353}$	pDEST32/pDES T22	$TOPO:gl2^{SI}$ <i>M353</i>	N/A	Gateway cloning
$gl2^{SIM404}$	pDEST32/pDES T22	$TOPO:gl2^{SI}$ <i>M404</i>	N/A	Gateway cloning
<i>SUMO1</i>	pDEST32/pDES T22	$TOPO:SUM$ <i>O1</i>	N/A	Gateway cloning
<i>SUMO2</i>	pDEST32/pDES T22	$TOPO:SUM$ <i>O2</i>	N/A	Gateway cloning
<i>SUMO3</i>	pDEST32/pDES T22	$TOPO:SUM$ <i>O3</i>	N/A	Gateway cloning

Table C.3. Accession numbers used for building alignments

GL2 paralogs from HD-Zip IV subfamily		
GL2 (AT1G79840)	PDF2 (AT4G04890)	ATML1 (AT4G21750)
ANL2 (AT4G00730)	FWA (AT4G25530)	HDG1 (AT3G61150)
HDG2 (AT1G05230)	HDG3(AT2G32370)	HDG4 (AT4G17710)
HDG5 (AT5G46880)	HDG7 (AT5G52170)	HDG8 (AT3G03260)
HDG9 (AT5G17320)	HDG10 (AT1G34650)	HDG11 (AT1G73360)
HDG12 (AT1G17920)		
GL2 orthologs across species		
ONM13423.1 (<i>Zea mays</i>)		BAB61212.1 (<i>Oryza sativa</i>)
XP_013680822.1 (<i>Brassica napus</i>)		XP_010430051.1 (<i>Camelina sativa</i>)
XP_031737751.1 (<i>Cucumis sativus</i>)		XP_021984354.1 (<i>Helianthus annuus</i>)
XP_008789333.1 (<i>Phoenix dactylifera</i>)		XP_016501674.1 (<i>Nicotiana tabacum</i>)
XP_006343080.1 (<i>Solanum tuberosum</i>)		XP_025611006.1 (<i>Arachis hypogaea</i>)
NP_001314068.1 (<i>Gossypium hirsutum</i>)		NP_001280927.1 (<i>Malus domestica</i>)
XP_003546423.2 (<i>Glycine max</i>)		XP_035548052.1 (<i>Juglans regia</i>)
XP_021652755.1 (<i>Hevea brasiliensis</i>)		
Arabidopsis SUMO isoforms		
SUMO1 (AT4G26840)	SUMO2 (AT5G55160)	SUMO3 (AT5G55170)
SUMO5 (AT2G32765)		

Table C.4. Bacterial stocks ordered from ABRC for SUMO genes

Stock #	Vector	Genes
U17495	pUNI51 (Kan ^R)	<i>SUMO1 (At4g26840)</i>
U21609	pUNI51 (Kan ^R)	<i>SUMO2 (At5g55160)</i>
U83833	pUNI51 (Kan ^R)	<i>SUMO3 (At5g55170)</i>