Structural Basis for Specific Gene Regulation by Auxin Response Factors

Alejandra Freire Rios

Thesis committee

Promotor
Prof. Dr D. Weijers
Personal chair at the Laboratory of Biochemistry
Wageningen University

Other members
Prof. Dr G.C. Angenent, Wageningen University
Prof. Dr J. van der Oost, Wageningen University
Prof. Dr L. Østergaard, John Innes Centre, United Kingdom
Dr Y. Stahl, Heinrich Heine University, Düsseldorf, Germany

This Research was conducted under the auspices of the Graduate School of Experimental Plant Sciences

Structural Basis for Specific Gene Regulation by Auxin Response Factors

Alejandra Freire Rios

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Thursday 22 December 2016
at 8:30 a.m. in the Aula.

Contents

CHAPTER 1

Introduction

Regulation of gene expression is a very important process for all live forms. All the information needed for an organism to complete its life cycle is contained in its genome, but the genome must be translated into a functional organism with the ability to grow, to specify new cell types/tissues and to respond to its environment. It is of crucial importance that the correct genes are switched on or off according to the developmental stage and needs of organisms. This critical step starts by selecting genes in the genome. As the central dogma of molecular biology tells: "DNA makes RNA and RNA makes proteins"[1]. The process of making RNA from a DNA template is called transcription and is the first line for gene regulation. Transcription is mediated by RNA polymerases. There are four RNA polymerases, of which RNA polymerase II (Pol II) is in charge of producing messenger RNA (mRNA) for protein-coding genes in both prokaryotes and eukaryotes. However, this enzyme cannot by itself recognize specific DNA sequences, and for site-specific initiation of transcription, Pol II must form a complex with other proteins called transcription factors. There are both general and specific transcription factors that aid in specific gene regulation. There are 6 general transcription factors (GTFs) required for correct initiation of transcription by pol II : TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH[2–8] . The complex formed by Pol II and the general transcription factors is called the transcription preinitiation complex (PIC). Eukaryotic promoters contain core sequences close to the translational start codon, such as the TATA box, that are recognized by subunits of the PIC, dictating its proper assembly and orientation[9]. TFIID is the first unit of the complex to bind the DNA and serves as a scaffold for the rest[10]. It either recognizes the core sequences in the promoters or mediates the interaction between the general transcription machinery and specific transcription factors to enhance the assembly of the PIC at their specific targets. Other subunits of the complex can also interact with specific transcription factors (reviewed in [9]). Part of the general transcription machinery is composed of general cofactors that facilitate the interaction between the specific, DNA-binding transcription factors and the general transcription machinery, one example being the Mediator complex [11].

For organisms to respond properly to the demands of their developmental stage or environmental stimuli, genes must be specifically regulated. Changes in the transcription of key genes can lead to dramatic developmental or physiological phenotypes[12], but can also be the driving force for morphological innovation during evolution[13–15]. Thus, precise control of transcription is of fundamental importance for proper growth and development. Similar to the general transcription factors, also specific transcription factors need to recognize DNA sequences in the promoter of their target genes. Hence the presence of a DNA sequence that allows stable biophysical interactions with the DNA binding domain(s) of a transcription factor gives a first way to achieve specific gene regulation. In prokaryotes, these sequences are generally long enough to guarantee specific binding of a TF to its target[16]. This means that one recognition/binding event is sufficient for gene regulation. Eukaryotes instead, despite having larger genomes, generally use shorter sequences as recognition elements[16]. This by necessity means that for any transcription factor there will be many unspecific binding events that may not lead to gene regulation. To overcome this, eukaryotes can use multiple strategies to ensure specific binding of TFs to their target genes. First, many more copies of the TF protein are produced to ensure that they will find the right site[17–19]. Then, recognition

9

sequences are clustered and cooperative and synergistic events might be necessary at these clusters ([20–23]). Different combinations of recognition elements might also be important to distinguish between different clusters of recognition sequences [16]. Increasing the complexity of the recognition element implies that that transcription factors may need to form higher order protein complexes to be able to bind the regulatory elements in the promoter of the target genes. Thus, protein-protein interactions are a second way to facilitate specific gene regulation in eukaryotes. Such interactions can be with transcription factors of the same, or of different families. Protein-protein interactions may also regulate the activity of the transcription factor itself as they can interact with other proteins that inhibit their activity[24]. Furthermore, protein-protein interactions may be involved in a third way to increase specific gene regulation: Transcription factors may interact with chromatin remodelers[25]. Changing the state of chromatin can either hide or expose genes, making them available for transcription only at specific moments or specific cell types. So in order to achieve specificity: the gene to be expressed must be exposed, the correct combination of proteins that will drive or repress transcription must be present and finally the DNA sequences must interact stably with the specific transcription factor.

Eukaryote proteins are generally composed by combinations of different well conserved domains[26]. In the case of transcription factors they typically will be composed of at least one DNA binding domain (DBD), an activation or repression domain and protein-protein interaction domain(s)[27]. Modularity has been important to generate a quick way to adapt and hence to generate evolutionary diversity, as it is advantageous to use existing pieces to adapt to changing environments. Because domains in modular proteins are likely to fold and function independently of the rest of the protein, modularity allows to study each domain's function separately, and understanding the function of the isolated different pieces may help us understand the process as a whole. Eukaryotic transcription factors have been classified into families based on phylogenetic classification of the domains they are composed of. In the case of plants, specifically *Arabidopsis thaliana*, there are approximately 1500 probable TFs classified in approximately 30 families. About half of these families are plant-specific. Among the plant specific transcription factors the following DNA binding domains (DBDs) are observed: AP2/ERF, WRKY, NAC, SBP and B3[28].

An important family of plant-specific transcription factors are the Auxin Response Factors (ARF). As its name indicates, the members of this family regulate genes in response to the plant signaling molecule auxin. Auxin (Indole-3-acetic acid or IAA) is a small molecule but also a very important one. Auxin regulates multiple biological processes during plant life. This structurally simple, tryptophan-like molecule regulates diverse events such as embryogenesis, vascular development, root growth, lateral root formation, tropic responses, leaf and flower initiation, etc. Auxin acts as a signal that triggers a cascade of events that will ultimately lead to changes in gene regulation: the nuclear auxin signaling pathway. When auxin enters the nucleus it finds the other elements that play a role in this pathway, three main families of proteins: the ARFs, the Aux/IAAs and the TRANSPORT INHIBITOR RESISTANT1/ AUXIN F-BOX (TIR1/AFB) auxin receptors, which are a subunit of the SKP1–CULLIN1–F-BOX (SCF)$^{TIR1/AFB}$. ARFs bind sequences in the regulatory regions

of their target genes to activate or repress transcription[29,30], but in the absence of auxin they are bound by Aux/IAA proteins that inhibit their activity[24,31]. Auxin binds to both the SCF(TIR1/AFB) receptor complex and the Aux/IAA protein, increasing the binding affinity between the two, leading to ubiquitination and degradation of the Aux/IAA proteins[32,33]. This liberates ARFs from inhibition and allows these transcription factors to regulate gene expression.

Each of these families of proteins has multiple members in most land plants studies[34–40]. In Arabidopsis there are 23 ARFs, 29 Aux/IAAs and 6 auxin receptors[41]. It has been reported that these proteins interact among each other with different affinities ([42,43]). In addition, most ARFs and Aux/IAAs have unique and different expression patterns[44]. These properties allow for the establishment of different combinations of interacting proteins in different cells. It has been proposed that this may contribute to generating different local responses to the same hormonal signal[45]. While this may be a source of divergence in the response to auxin, the question of how the same small signaling molecule can trigger so many different responses is still not answered. From all the components of the pathway, ARFs will be the ones ultimately selecting the target genes to be locally regulated by auxin. As indicated above, ARF genes are expressed in specific spatial patterns. For example, the overlapping expression of ARFs in the *Arabidopsis* embryo divides it in groups of cells with specific subsets of ARFs: a pre-pattern. Interestingly, each of these groups of cells correlates with the precursors of different tissues[44]. Importantly, ARF proteins are generally not interchangeable. *arf5* mutants cannot form roots[46]; *arf7arf19* mutants cannot make lateral roots[47]; and *arf6arf8* will turn into plants with flowering defects [48]. In each of these cases none of the other co-expressed ARFs are able to take over the function of the defective one. Furthermore, in promoter-swap experiments it was shown that ARF1 cannot replace ARF5 when expressed from the ARF5 promoter[44], and that expressing ARF5 from the ARF13 promoter causes developmental defects[49]. Thus, it can be concluded that indeed, individual members of the ARF family have distinct biochemical properties that contribute to generating different local responses to the same hormonal input signal. We postulate that differences in target gene recognition are an important component of the biochemical divergence between ARF proteins. In this thesis, the ARF family is used as a case study for how specific gene regulation can be achieved by plant transcription factors.
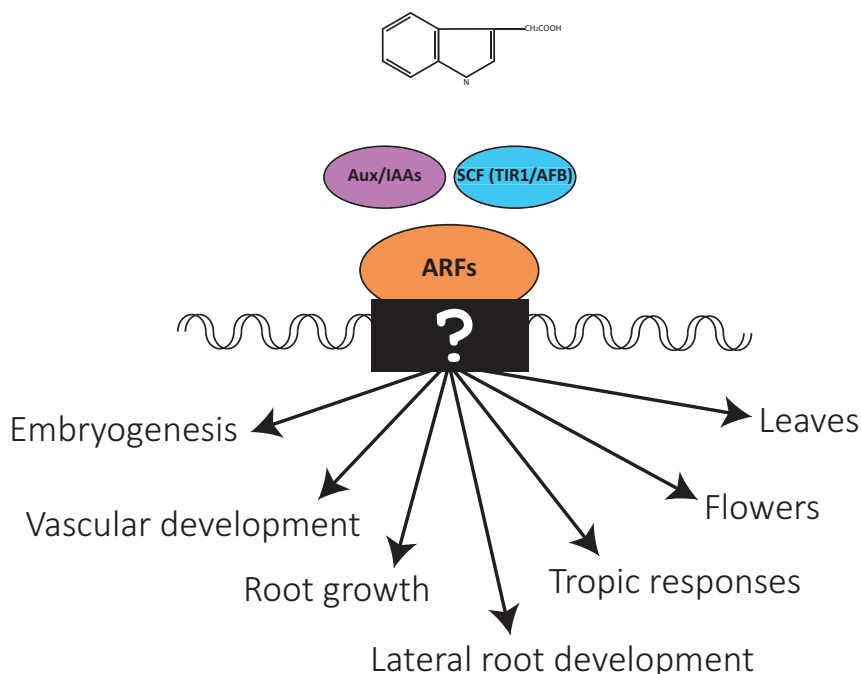
Scope of this thesis

Auxin Response Transcription Factors (ARFs) form a major plant transcription factor family. As their name indicate they are in charge of regulating gene expression in response to auxin stimuli. Auxin is a very important phytohormone. Auxin regulates multiple developmental processes during the life of plants starting with the formation of an embryo, correct specification of different tissues, flowering, etc. Due to its importance in so many crucial processes this hormone and its signaling pathway, ending in gene regulation, has been extensively studied for decades. All the key players in this pathway have been identified, but details on how ARFs can specifically select targets for so many important and diverse responses still remains a black box. The work presented in this thesis aims to shed some light into that black box.

Chapter 2 reviews and discusses in detail what is known on auxin signaling: everything that happens before and after regulation of genes by ARFs in context of an important phase of life, embryogenesis.

Chapter 3 reports the crystal structure of the DNA binding domain (DBD) of two representative members of the ARF family. These structures are the starting point to begin to understand how ARFs recognize and bind DNA. This chapter introduces the "caliper model" for specific DNA binding by ARFs. In this model the separation between the DNA recognition elements provide an extra layer for generating specificity and divergence

between the different ARFs. This chapter also identifies a novel high-affinity DNA sequence discovered through systematic screening for ARF binding sites.

Chapter 4 shows the structural properties of ARFs that explain the high binding affinity to the newly described binding sequence. In this chapter we also test the validity and biological relevance of the structure-based caliper model *in vivo* using diverse methods. We present a detailed bioinformatics analysis correlating complex AuxREs to auxin response, and explore their role *in vivo*. These genome-wide data, as well as experiments involving the known ARF target TMO5 demonstrate the critical importance of the inverted repeat binding site – predicted by the ARF protein structure – for auxin-dependent expression. Finally, this chapter explores the role of differences in spacing between two adjacent ARF binding sites in driving gene expression patterns during development.

While the previous chapters focus exclusively on divergence in the DNA-binding domain as a source of differences between ARFs and in their biological function, in **Chapter 5** we explore the contribution of other domains to biological function. Using ARF5 as a model, we show that all domains are required for biological function, and through swap experiments it is shown that even the closely related ARF6 protein cannot replace ARF5, or its domains. We next use a proteomics approach to identify proteins interacting with domains of ARF5 to test whether protein complex formation is another element giving target specificity to ARFs. The results demonstrate the validity of this domain-specific interactome approach and show contributions of each domain in recruiting co-factors.

In this thesis, protein-protein interactions are a central component in developing models for ARF function. **Chapter 6** describes an update in the methodology used to test protein-protein interactions *in vivo*: Förster Resonance Energy Transfer – Fluorescence Lifetime Imaging (FRET-FLIM).

Finally, **Chapter 7** summarizes of all main findings and conclusions from this thesis, and gives directions for future research.

13

CHAPTER 2

Auxin Regulation of Embryo Development

* Alejandra Freire Rios
* Saiko Yoshida
Dolf Weijers

*These authors contributed equally to this work

Abstract

Important steps in plant development are made shortly after fertilization. In a brief succession of cell divisions, the zygote is transformed into an embryo, a multicellular structure carrying all fundamental tissue types and the meristems. Hence, embryogenesis offers excellent opportunities to dissect the molecular control and cellular mechanisms underlying plant development. In the past decades, forward and reverse genetics studies have revealed that the plant hormone auxin plays a central role in the establishment of pattern and polarity in the Arabidopsis embryo. Here, we review the roles that localized auxin biosynthesis, directional transport and cell type-specific response play in embryo development. We focus on the molecular mechanisms, as well as the feedbacks that connect these disparate levels of regulation. Finally, we discuss the potential for hormonal cross-talk in auxin-dependent control of the key events during the earliest, formative phase of plant life.

Early plant embryogenesis

Multicellular organisms begin life as a single zygote cell. While the animal embryo is a miniature form of the adult body and thus has a relatively complex structure, the plant mature embryo has a rather simple structure: an embryonic root, hypocotyl and one or two embryonic leaves. This miniature encompassed meristems in the shoot and root tips. The meristems will create all the other parts of the mature plant body after germination [50]. In Arabidopsis embryos, these meristem primordia consist only of a few stem cells. These stem cells will divide and spatially coordinate the acquisition of different cell identities during the post-embryonic generation of a functional body. Stem cell niches are an excellent example for the importance of spatial coordination in cell specification where, in order to keep a functional niche, stem and organizer cells need to be in direct vicinity [51]. Additionally, as early embryos consist of few cells, yet different layers, tissues and organs are established, there is a general need for tightly coordinated development. In this chapter we discuss how the plant hormone mediates the coordinated acquisition of cell types during this formative phase of Arabidopsis life.

Unlike in animal embryos, plant cells do not migrate during embryogenesis because they are surrounded by a rigid cell wall. Therefore, oriented cell division and directional expansion plays an important role in morphogenesis. Division patterns during embryogenesis have been studied and described based on the observations of sections [52,53]. After fertilization, the apical-basal polarity is forecast when the zygote divides asymmetrically to create an apical embryonic cell and a basal extra embryonic cell (Figure 1). The embryonic cell further divides three times to generate 8 embryonic cells organizer in two tiers. While the cells in the upper tier will generate the shoot (shoot apical meristem and cotyledons), the cells in the lower tier will make the hypocotyl, root and root apical meristem. Subsequently, all embryonic cells divide periclinally and generate inner and outer cells corresponding to the
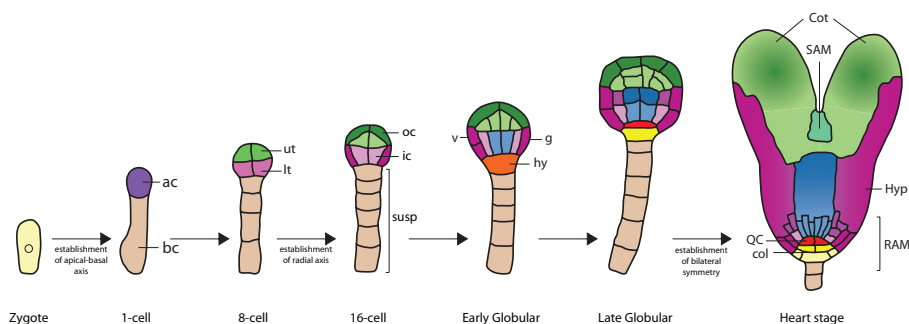
Figure 1: Embryo development in Arabidopsis.
Sequential cell division events as observed in 2D sections of embryos starting in the zygote until the heart stage. Points in the process where symmetries are established are indicated. Cells in different colours will differentiate in different tissues/organs.
ac=apical cell; bc=basal cell; ut=upper tier; lt=lower tier; oc=outer cell; ic=inner cell; susp= suspensor; v=vascular initial; g=ground tissue initial; hy=hypophysis; Cot=cotyledon; SAM=shoot apical meristem; Hyp=hypocotyl; RAM=root apical meristem; QC=quiescent center; col=columella cells

first establishment of a radial axis. Next, lower tier inner cells divide periclinally to generate the initials for ground tissue and vascular cells at early globular stage. On the other hand, the extra-embryonic cell divides to create the suspensor. Its uppermost cell is specified into hypophysis and then divides asymmetrically to create the upper lens-shaped cell and lower cell which will respectively become the initials for quiescent center (QC) and columella cells in the root. From transition to heart stage, primordia of the two cotyledons are formed and the structure of embryo obtains bilateral symmetry. Thus, during this morphogenesis phase the basic body pattern is established and the meristem of shoot/root and embryonic organs are generated. Arguably, this morphogenetic phase is of great importance for the establishment of a new plant from a single fertilized egg cell. Hence, understanding of the molecular and cellular mechanisms underlying these events is a key goal for plant developmental biology. In the past decades, much progress has been made, and interestingly, the plant hormone auxin has surfaced repeatedly as a central regulator. Here we will discuss the various aspects of auxin regulation that contribute to regulating its activity in embryo development.

Auxin biosynthesis pathways in embryogenesis

The major natural auxin, Indole-3-acetic acid (IAA) is biosynthesized from tryptophan via a two-step pathway (Figure 2). Several key enzymes are known to be involved in this pathway. TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS 1 (TAA1) and its closest homologs TRYPTOPHAN AMINOTRANSFERASE RELATED 1,2 (TAR1, 2) are transaminases that convert tryptophan to indole-3-pyruvate (IPA)(Tao et al. 2008). YUCCA is a flavin monooxygenase that catalyses oxidative decarboxylation of IPA to produce IAA [54]. Eleven YUC homologues are known in the Arabidopsis genome [55]. *TAA1* and some of the *YUC* genes are expressed during embryogenesis. *YUC3*, *YUC4* and *YUC9* are expressed in the suspensor from 8-cell stage on. At the globular stage, *YUC1* and *YUC4* are expressed in the cells around the future shoot apical meristem, while *YUC8* is expressed around the
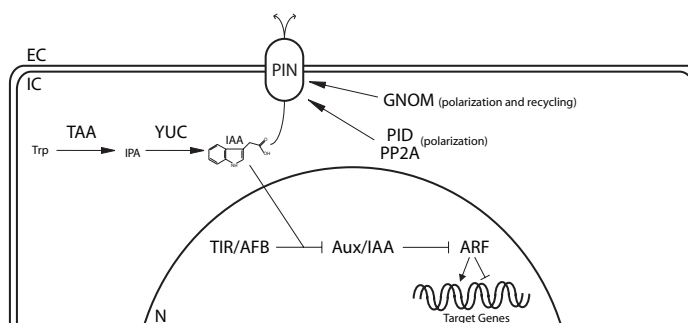
Figure 2: Core auxin pathways.
After being synthesized by the TAA/YUC pathway, intracellular (IC) auxin can either be transported to the extracellular space (EC) or initiate the signaling cascade inside the nucleus (N). Auxin in the nucleus will allow the recognition of Aux/IAAs by the TIR/AFB complex. This will lead to Aux/IAA degradation, and hence de-repression of ARF activity. Transport of auxin is carried out by PIN proteins. The polarized position of PINs within the cells is controlled by proteins like PID, PP2A and GNOM.

hypophysis. From late globular to heart stage, *YUC4* is not only expressed in apical cells around the shoot meristem but is also expressed in the basal hypophysis. *TAA1* is expressed in the apical epidermal cells from 16-cell stage (Robert et al., in press). From transition to heart stage on, it is expressed in the L1 layer of shoot apical meristem [55]. Consistent with the TAA/TAR and YUC proteins acting in a linear biosynthetic pathway, higher-order mutants in each family lead to nearly indistinguishable phenotypes. Embryos of *yuc1 yuc4 yuc10 yuc11* quadruple mutant and the *taa1 tar1 tar2* triple mutant display abnormal cell division in embryonic cells. Seedlings of these mutants lack root/hypocotyl and often have aberrant number of cotyledons, which strongly suggests that auxin activity is required for the normal establishment of these embryonic structures [55,56]. Interestingly, these severe phenotypes cannot be rescued by ectopic production of auxin, which indicates the importance of spatial and temporal regulation of auxin local biosynthesis (Robert et al., in press).

If local, rather than ubiquitous auxin biosynthesis is important for auxin-dependent embryo development, a key question is what activates the expression of the biosynthesis genes. Transcriptional regulators of the SHORT INTERNODES/STYLISH (SHI/STY) family have been identified as activators of *YUC4* and *YUC8* [57–60]. The expression pattern of *STY1* and *YUC* genes overlap [61] and *STY1* is expressed in the future cotyledon primordia of early globular embryos [61]. Overexpression of *STY1* increases auxin biosynthesis while in the *sty1 sty2* mutant the level of free IAA is low. This mutant displays severe defects in the development of leaf and floral organs, especially the style. The style phenotype of *sty1 sty2* can be restored by exogenous application of auxin [60]. A dominant negative transgenic line of *STY1* does not make a shoot apical meristem, indicating that STY1-dependent auxin biosynthesis has a role in the formation of the shoot apical meristem during embryogenesis or its post-embryonic maintenance. STY1 is in turn activated by DORNRÖSCHEN-LIKE (DRNL), an AP2/ERF family transcription factor [62]. DRNL and its homologue DORNRÖSCHEN (DRN) interact with class III HD-ZIP family transcription factors PHAVOLUTA, PHABULOSA, REVOLUTA, CORONA and ATHB8 to regulate patterning of apical embryo [63]. As a quintuple mutant among these HD-ZIP transcription factor genes does not develop an embryonic shoot meristem [64], this suggests the outlines of a genetic network that directs morphogenesis of shoot apical meristem during embryogenesis and involves several interacting transcription factors to control local auxin biosynthesis. Interestingly, *DRN* as well as the HD-ZIP gene *ATHB8* are direct transcriptional target genes of the auxin response factor MONOPTEROS (see below; [65,66]), which suggests that this control network is not linear, but likely involves feedback regulation by auxin.

Auxin redistribution through directional transport

The expression pattern of auxin biosynthesis genes does not always match the locations that are marked by auxin response reporter genes such as DR5-GFP [67]. This suggests that locally produced auxin should be transported to where auxin is needed. Indeed, expression analysis and protein localization, as well as genetic analysis supports an important role for the PIN auxin efflux facilitators in embryo development. The PIN proteins are well known

to be involved in the efficient transport system of auxin [68]. Of the 8 Arabidopsis *PIN* genes, 4 are expressed during embryogenesis [67]. From the 1-cell stage onward, the *PIN7* gene is expressed and its protein becomes polarly localized in the apical side of extra-embryonic cells. PIN1 is also expressed from 1-cell stage and subsequent stages, although it does not show clear subcellular polarity until mid-globular stage. Auxin response activity (DR5-GFP reporter) is also detected in the embryonic cells, but not the extra-embryonic cells at these stages. In *pin7 and pin1 pin2 pin4 pin7* quadruple mutants, ectopic DR5-GFP expression is observed in the suspensor [67]. Considering that few *YUC* genes are expressed in the suspensor, it suggests that PIN7 transports the biosynthesized auxin from suspensor to embryonic cells [69]. At globular stage, the polarity of PIN7 is changed from apical to basal in the suspensor cells, while PIN1 becomes basally polarized in the vascular initials. This likely promotes accumulation of auxin in the hypophysis. Consistently, strong DR5-GFP signal is observed in the hypophysis [67]. At the transition stage, PIN1 polarity localizes towards the flanks of the apical embryonic cells. This promotes accumulation of auxin in the cotyledons primordial [70]. On the other hand, PIN4 is expressed in the hypophysis at the globular stage embryo. After the division of the hypophysis, it is expressed in the upper lens-shaped cell. PIN3 is expressed in the columella precursors in the heart stage embryo. The *pin7* mutant displays abnormal cell division as well as the quadruple mutants of *pin1 pin3 pin4 pin7* and *pin2 pin3 pin4 pin7* indicating that coordinated polar localization of PINs regulates embryo patterning [67,71].

Recent work has suggested the existence of a connection between local auxin biosynthesis and polarization of PIN proteins. In auxin biosynthesis mutants, localization of PIN1 is apolar and its expression level is reduced in the later stage (Robert et al., in press). The mechanisms underlying this link are unclear, but will likely involve control of the cellular mechanisms that target the polar localization of PIN proteins. The polar membrane localization of PIN proteins is the result of a continuous exocytosis/endocytosis cycle, where regulation can act on either step [68]. Several proteins are known as regulators of PIN trafficking and polarity (Figure 2). A serine-threonine kinase, PINOID (PID) and PROTEIN PHOSPHATASE 2A (PP2A) antagonistically regulate polarity of PIN proteins by regulating their phosphorylation status [72–75]. While PIN proteins are targeted to the apical plasma membrane by phosphorylation, PINs are targeted to the basal plasma membrane by dephosphorylation. PID, three PID homologs (PID2, WAG1, WAG2) and PP2A are expressed during embryogenesis [75,76]. As expected, mutations in any of these polarity regulators cause defects similar to loss of PIN proteins or auxin biosynthesis [73,75].

The recycling of PIN proteins from endocytic vesicles (endosomes) to the plasma membrane requires GNOM, an ADP ribosylation factor-guanine exchange factor [77,78]. GNOM regulates the recycling of PIN1 protein by controlling its polar localization on the membrane [77]. GNOM is ubiquitously expressed [79] and the mutant seedling is rootless and makes fused cotyledons [80], which phenocopies the *pin1* mutant and the *pin1,3,4,7* quadruple mutant [67]. The *gnom* mutant embryo defects appear from the zygote onward and the mutant embryo fails to establish an apical-basal axis [80]. At the globular stage, orientation of cell division plate and cell division pattern of the entire embryo become abnormal [81]. Establishment of the bilateral symmetric structure at heart stage is also disturbed and results in a ball–shaped

embryo [80]. In *gnom* embryos, PIN1 is no longer polarized and auxin transport is reduced causing accumulation in the apical region of globular embryos [81]. When GNOM expression is driven by the promoter of a provascular gene, polar localization of PIN1 is restored, auxin accumulation in the apical embryo is reduced and formation of primary root is rescued. When GNOM is expressed in the hypophysis, primary root formation is also restored. This suggests that GNOM acts non-autonomously to regulate root apical meristem formation, which is consistent with its function in intercellular auxin transport. Furthermore, GNOM expression in apical epidermis restores the formation of two cotyledons. Thus GNOM-dependent polar auxin transport is important for the establishment of meristems [81].

Studies on the HANABA TARANU (HAN) GATA transcription factor have revealed a critical role for PIN gene regulation in embryo patterning [82]. *HAN* is first expressed in the zygote, and expression is retained in all embryonic cells until the 16-cell stage. From globular to heart stage, *HAN* expression becomes restricted to the provasculature cells [82,83]. HAN regulates the transcription of the genes regulating the development of the basal embryonic cells. Therefore, in the *han* mutant embryo, establishment of the apical-basal axis fails. The basal cells of *han* embryos from 16-cell stage resemble suspensor cells with large vacuoles and lower cell divisions. In these cells, marker genes of suspensor and hypophysis (*SUC3, WOX5*) are expressed, whereas *SHR,* which is normally expressed in provasculature of basal cells, disappears. The expression patterns of apical marker genes (*WUS, ML1*) are not affected in the *han* mutant. Distribution of auxin is also changed in the *han* mutant. While DR5-GFP is expressed in hypophysis and neighbouring suspensor cells in the wild-type embryo, the expression domain is expanded to basal embryonic cells in *han* mutant. The expression domain of PIN7 is also expanded to the basal embryonic cells, while the expression of PIN1 is restricted to the apical embryonic cells. Thus the lack of root meristem in the *han* mutant is strongly correlated with of the disruption of the establishment of an auxin maxima in hypophysis likely due to PIN gene misexpression [82]. Whether the regulation is direct remains to be determined, but this finding opens new avenues for understanding the regulation of PIN gene activity during embryo development.

In addition to the PIN proteins, other regulators of auxin transport have been identified. Notably, the PGP/ABCB transporters facilitate non-polar auxin transport [84], but their activity has not yet been proven essential for normal embryo development [85]. Recently, a new family of auxin transporters, the PIN-LIKES (PILS) proteins, was shown to mediate intracellular auxin transport between cytosol and the endoplasmic reticulum [86]. PILS are auxin efflux carriers with a similar topology to PINs. The PILS family consist of 7 proteins containing an auxin carrier domain and six (PILS1,2,3,5,6,7) are localized to the endoplasmic reticulum [86]. Among the 7 PILS, PILS2 and PILS5 are the most abundantly expressed in seedlings. Unlike PIN proteins, the family is conserved throughout the plant kingdom and even exists in algae, suggesting that PILS can be evolutionarily older than PINs. PILS are uniformly expressed in various tissues and some of them are auxin-inducible. Over expression and loss of function of PILS2 and/or PILS5 affects hypocotyl and root growth, lateral root organogenesis, and root hair length. It will be interesting to see if this novel mode of auxin partitioning is also important for embryo development.

Transcriptional response to auxin

After auxin biosynthesis and transport, hormone accumulation triggers transcriptional changes to effect altered cell division and identity. The auxin mediated transcriptional responses are mainly controlled by the interaction of two families of plant transcriptional regulators: the Auxin Response Factors (ARFs) and the Aux/IAA proteins. In a general mechanism of action, the Aux/IAA proteins, together with transcriptional co-repressors like TOPLESS (TPL), form a complex with the ARFs in auxin low levels. When auxin levels rise in the cell, the Aux/IAA proteins are targeted to the 26S-proteosome by an SCF E3 ubiquitin ligase complex. Upon degradation of Aux/IAAs, ARFs are released and can act activating or repressing their target genes (Figure 2; reviewed in [87]).

The Arabidopsis genome encodes 29 Aux/IAA proteins that share conserved domains. Domain I is necessary for transcriptional repression and it has been shown to recruit the TPL corepressor in most of the Aux/IAAs [88,89]. Domain II contains the degron motif, a 13 amino acids sequence responsible for the Aux/IAAs' instability by mediating their interaction with the TIR1/AFB receptor [90,91]. There might be other sequences outside Domain II contributing to this interaction; for example, the affinity between the Aux/IAAs and the TIR1/AFB can drop dramatically when a KR conserved motif between Domain I and Domain II is mutated [42]. Finally, domains III/IV in the C-terminal region, considered as the interaction domain, is thought to mediate homo- and heterodimerization between the Aux/IAAs and with the ARFs [92]. There are also Aux/IAAs that lack one of the domains. They are considered non-canonical Aux/IAAs and their function remain unknown. Over expression of a subclade of Aux/IAAs which lack domain II (IAA20, IAA30 and IAA31), results in auxin-related phenotypes suggesting that they may interfere with endogenous ARF-Aux/IAA interactions [93].

The degradation of Aux/IAAs is mediated by the family of E3 ligases called SCF[TIR1/AFB1-5] which targets them for ubiquitination. The F-box protein TIR1 (TRANSPORT INHIBITOR RESPONSE 1) and related proteins AFB 1-5 (AUXIN SIGNALLING F-BOX PROTIEN 1,2,3,4,5) are the substrate receptor of the SCF [90,91]. All the 6 members of the TIR1/AFB family are auxin receptors although they show individual distinct biochemical properties and biological functions [94]. For the SCF to recognize the Aux/IAAs, the F-box protein needs to be directly bound to auxin. The structure of TIR1 has been determined in the presence of auxin and the degron peptide of IAA7, and it shows that its C-terminal 18 Leucine-Rich-Repeats (LRRs) is essential for Aux/IAA and auxin binding [95]. There are no conformational changes of TIR1 upon auxin binding, which sits in a binding pocket underneath the Aux/IAA binding site [95,96]. Recent experiments using TIR1-ASK1 and IAA7 showed that both proteins act as co-receptors and are necessary and sufficient for auxin binding. With 29 Aux/IAAs and 6 TIR/AFBs, many qualitatively different co-receptor pairs may exist. Recent data suggests that this is the case. In a qualitative yeast two-hybrid assay, different receptor pairs showed different auxin dose-response and, furthermore, an inverse correlation between the Aux/IAAs' stability and the strength of the interactions. Also quantitative biochemical assays (saturation and/or homologue competitive IAA binding assays) showed that affinity of different co-receptor pairs for auxin ranges from 10 nM to >1 uM[288]. Given that the

complex TIR1/AFB-auxin-AUX/IAA is the first step toward Aux/IAA degradation, hence de-repression of ARFs; this complex formation may be a part of a control mechanism of differential auxin responses. While mutations in the degron of many Aux/IAA proteins have been reported to cause distinct auxin-related defects [97], only few were shown to affect embryo development. The *iaa12/bdl* and *iaa13* mutations cause defects in root initiation [98,99], while *iaa18* mutants have defects in cotyledon formation [100]. In contrast, *iaa10* and *iaa11* mutations affect suspensor and hypophysis development [49]. Rather than distinct protein capacities, these unique phenotypes likely reflect the highly specific gene expression patterns of Aux/IAA genes. Misexpression of the unrelated Aux/IAA protein iaa3/shy2 in the IAA12/BDL expression domain causes bdl-like root defects [99].

As Aux/IAAs are degraded, ARFs are free to elicit gene expression response to auxin. There are 23 ARFs in Arabidopsis [101], which adds to the possible combinatorial logic of auxin signalling considering that the members of this family also have different biochemical properties and biological functions. All auxin-dependent processes in the embryo seem to be mediated by ARF activity, as mutations in individual or multiple ARFs, or misexpression of the Aux/IAA proteins disrupt all these processes [49,102]. ARFs have three protein domains. At the N-terminus there is the conserved DNA binding domain, which binds to the Auxin Response Elements in the promoter regions of the direct targets. The second domain, named Middle Region, is the non-conserved part and is proposed to determine the activity of the ARFs. ARFs have been classified as gene activators or repressors based on the amino acid composition of their middle region. Experiments in protoplasts showed that some ARFs with a relatively glutamine rich MR (ARF5—8 and 19) can activate synthetic auxin promoters, and that ARFs with less glutamines (ARF1—4 and 9) can repress the same promoters. The rest of the ARFs have been arbitrarily classified based on these results [92]. Since there is experimental data for some ARFs suggesting that they can act as both activators and repressors of different genes [103,104], this classification might not be entirely accurate and the regulation of transcription by ARFs might be more complex. Indeed, triple mutants between ARF1, ARF2 and ARF6 show a phenotype that none of the single or double mutants shows [49]. This suggests that redundancy among ARFs is not limited to close related members with similar domains III/IV.

Interestingly, recognition motifs for corepressor proteins have been found in several ARFs [105], and yeast two-hybrid assays have shown strong interactions between the transcriptional corepressor TPL and ARF2, 9, 17 and 18; and putative weak interactions with ARF1, 3, 4, and 19 [88]. Finally at the C-terminus, we find the Domains III/IV which mediates the ARFs interaction with the Aux/IAAs and with other ARFs [92].

Generating cell type-specific responses

Various developmental processes during embryogenesis are dependent on proper auxin response [49], and each of these is marked by activity of the DR5-GFP auxin response reporter [67]. Therefore, whether or not there is an auxin response does not seem to define

what developmental output is triggered. A key question is what defines the nature of the auxin output. A plausible explanation lies in the biochemically distinct properties of the transcriptional regulators. Importantly, these proteins are differentially expressed in embryos: different sets of Aux/IAAs and ARFs are expressed in different cell types. Tissue types are established in the embryo soon after fertilization and the differential expression of ARFs can already be observed at this level in *Arabidopsis*. An expression map of all ARFs at different embryonic developmental stages was recently described [44]. All the embryonic cells express at least one *ARF* gene and most cell types express a unique combination. These expression patterns are dynamic, changing between developmental stages, suggesting that as the embryo gets more complex, and as the different cell types gain identities, different subsets of ARFs need to be active. In the octant cells two subsets of ARFs can be observed; ARFs expressed in all cells of pro-embryo and suspensor (ARF1, 6 and 18) and ARFs expressed only in the suspensor (ARF2, 9 and 13). During the globular stage, seven ARFs are expressed in the embryo in partially overlapping patterns; ARF1 and 18 are expressed in every cell, ARF6 is expressed un the basal tier and suspensor, ARF5/MP is strongly expressed in the lower tier of the pro-embryo, ARF13 is expressed in the suspensor and surrounding endosperm, and finally ARF2 and 9 are expressed in the suspensor and in the protoderm of the lower tier of the pro-embryo. In heart stage, when cotyledons and meristems are established, more ARFs are expressed. Nine ARFs (1, 2, 4, 5, 6, 7, 10, 11 and 18) are expressed in the vascular cylinder. ARF3 is restricted to the abaxial side of the cotyledons. ARF9 and 10 are expressed in the protoderm. ARF1, 2, 6 and 18 are expressed in the quiescent center (QC) and columella cells. ARF5 and 7 are expressed in the QC and ARF9 is expressed in the columella cells. Co-expression of ARFs in specific tissues may imply that they act in the same biological processes, but this still needs to be determined [44] (Figure 3).
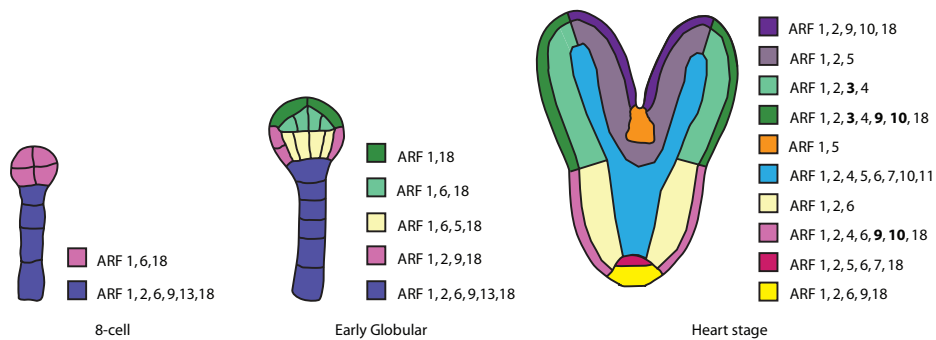
Figure 3: Expression of Auxin Response Transcription factors during embryogenesis.
Schematic depiction of gene expression patterns of *ARF* genes as determined by promoter-GFP reporters. Expression can be observed already early at embryogenesis. The number of *ARFs* expressed increases in time. Unique combinations of *ARFs* correlate with the identities obtained by each group of cells which are depicted in different colours.

In any event, the cell type-specific ARF complements have the potential to generate unique cellular auxin output if the ARF proteins are not biochemically equivalent. In promoter-

swap and misexpression experiments, it was shown that ARF16 cannot fully replace the embryonic function of ARF5/MP [99]. Furthermore, driving ARF9 from the ARF5/MP promoter enhances weak *mp* phenotypes [49], and while an *arf6* mutation enhances the *mp* phenotype, *arf1* mutation suppressed it [49]. Finally, expression of MP in the suspensor from the ARF13 promoter interfered with suspensor development while and extra ARF13 dose did not. These data together suggest that some ARFs are interchangeable (ARF6 and ARF5/MP), while others act differently (ARF16 and ARF5/MP) or even antagonistically (ARF1 or ARF9 and ARF5/MP). This supports a role for an ARF prepattern in establishing cell-specific auxin output.

Downstream effectors of auxin signalling

The key to understanding the cellular mechanisms for auxin-dependent development lies in the identification of the genes that are controlled in each cell type. The only ARF targets identified in the context of the embryo are regulated by ARF5/MP. Three direct target genes of MP expressed in cells relevant for root initiation have been identified through a microarray approach, and were named *TARGET OF MONOPTEROS* (*TMO*)[103]. MP acts in inner, lower tier embryonic cells to promote root intiation, and does this in part non-cel-autnomously in the case of hypophysis specification [106]. All three TMO genes (*TMO3,5,7*) are expressed in the cells adjacent to the hypophysis where MP acts. *TMO3* encodes an AP2 transcription factor that becomes broadly expressed in later stages. *TMO5* and *TMO7* encode basic helix-loop-helix (bHLH) transcription factors. At later stages *TMO5* is expressed specifically in vascular tissues and *TMO7* becomes restricted to the future root stem cells. These three TMOs are all individually able to partially rescue root initiation in a weak *mp* mutant when misexpressed, but the largest effect was seen with *TMO5* and *TMO7*. Since these TMOs are transcription factors we can assume that root initiation is a process of successive transcriptional steps. It is to be noted that the TMO7 protein moves directionally to the hypophysis where it is presumed to act as a cofactor for other bHLH transcription factors, and control hypophysis specification or division. The mobilisation of this may be taken as the signal of intercellular communication used by MP to regulate hypophysis divisions [103].

In contrast to the mobile TMO7, TMO5 acts cell-autonomously in vascular cells. MP is required for oriented divisions in these cells that allow the development of a vascular bundle with more than 30 cell files from 4 precursor cells. TMO5 was recently shown to mediate these local, oriented divisions [107]. TMO5 acts in a complex with its bHLH partner LONESOME HIGHWAY. The activity of the TMO5/LHW complex in promoting oriented division is normally restricted to a small domain by transcriptional control through MP (on TMO5) and other pathways that restrict LHW expression. When ectopically expressed, the TMO5/LHW dimer is able to trigger the same oriented division in a variety of cell types in the root. Hence, the diverse functions of MP in activating root formation bifurcate at the level of its *TMO* target genes.

Another identified direct target of MONOPTEROS is *DRN*. *DRN* encodes an AP2 transcription factor that acts redundantly with its paralogue *DRNL* upstream auxin polar transport and synthesis (see above). The expression of *DRN* can be observed from the two-to four-cell stage in the embryo proper, then it focuses in the emerging cotyledons and then it gets restricted to the SAM at torpedo stage. Loss-of-function *drn* mutant phenotype, affects the apical and the basal embryo domains. In *drn* mutants localization of the PIN1 protein is altered, being randomly distributed instead of being located basally as in the wild type. This can already be observed at 32-cell stage, where it is found to be localized laterally. In this mutant the expression of the auxin distribution and response reporter DR5-GFP at different embryonic stages is also abnormal. In the *drn* mutant abnormal cell division can be observed from the globular stage onwards. This abnormal division affects specially cotyledon organogenesis. Also phenotypes resembling the one of *mp* and *bdl* can be observed [63,65], which suggests that MP acts in part by controlling PIN1 protein localization through the DRN/DRNL genes.

Auxin plays a fundamental role not only in determining the location of the distal stem niche but also in the specification of the QC and entire embryonic root. Downstream MP and other ARFS, other genes are activated. The AP2 putative transcription factors *PLETHORA* (*PLT*) are transcribed in response to auxin. In the embryo ARF5-MP and ARF7-NHP4 are necessary for the transcription of *PLETHORA1 (PLT1)* and *PLETHORA2 (PLT2)*, genes that act in the specification and maintenance of the QC and the stem cell niche (Aida et al. 2004). The expression of both PLT genes can be detected already at octant stage and it is restricted to the basal half of the embryo [108,109]. At globular stage it is expressed in the provascular cells and the QC progenitor and later on it is expressed only in the QC and surrounding stem cells. It has been observed that misexpression of PLT1 and PLT2 leads to the development of ectopic roots, and that PLT activity is critical for embryonic root formation [109]. The regulation by MP and ARF7 may not be direct given the slow activation after auxin treatment0 [108]. Recently, an intruiging aspect of PLT activity was revealed. The expression of the HD-ZIP III genes (see above), master regulators of the embryonic apical fate, is expanded to the root domain in *plt* mutants [110]. PLT misexpression suppresses HD-Zip gene expression, and conversely, HD-Zip misexpression suppressed PLT expression. In these cases, misexpression induced the formation of a second shoot in the root domain (HD-Zip misexpression), or a second root in the shoot position (PLT misexpression). This suggests that part of the network downstream of auxin acts through PLT genes to suppress shoot development in the root pole [110].

Interactions of auxin with other hormonal pathways

Auxin is a major hormone that has been extensively studied. However, recent studies have shown that auxin interacts with other hormones to regulate developmental pattering and growth in various tissues [104,111–114]. Therefore, understanding how auxin controls development will have to include a description of its cross-talk with other hormones. We briefly review recent insights in cross-talks relevant to auxin-dependent embryo development.

While auxin promotes cytokinin signalling in the shoot apical meristem [104], cytokinin acts antagonistically to auxin signalling in the root. It was shown that auxin promotes meristematic activity whereas cytokinin promotes differentiation of stem cells in root. Furthermore, cytokinin regulates redistribution of auxin in the root apical meristem. A primary cytokinin-response transcription factor, ARR1, activates the gene SHY2/IAA3 (SHY2), a repressor of auxin signalling that negatively regulates the PIN genes [112]. First evidence for auxin control of cytokinin signalling in embryo development has come from the analysis of the cytokinin signalling reporter pTCS-GFP [115]. Its activity is first detected in the hypophysis and suspensor at the 16 cell stage embryo [115]. After the division of the hypophysis, TCS expression is only maintained in the apical cell that is specified to become the quiescent center, whereas it is repressed in the basal cell that becomes the distal root cap. Consistently, the negative regulators of cytokinin signalling, type-A *ARR7* and *15* are up-regulated in the basal hypophysis cell. Importantly, expression of these ARR genes is promoted by auxin. Furthermore, expression of DR5-GFP is maintained in the basal hypophysis cell while it is suppressed in apical hypophysis cell. These data suggest that auxin antagonizes cytokinin signalling via activating the type A ARRs. This appears to be biologically meaningful as altering cytokinin signalling in embryos through manipulating ARR gene expression causes defects in embryonic root formation similar to those found in several auxin-related mutants.

In the root vasculature, cytokinin signalling markers TCS-GFP and *ARR5* are expressed in procambial cells adjacent to the xylem whereas auxin signaling markers such as DR5-GFP and *IAA2* promoters are expressed in the xylem. This indicates that cytokinin activity is correlated with procambium division while auxin promotes xylem differentiation [116]. Indeed, cytokinin treatment inhibits protoxylem formation, and in the cytokinin receptor mutant *wol*, all the vascular cells differentiate into protoxylem [117]. Mutation in the ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN 6 (AHP6) gene was able to restore phloem and procambium in *wol* mutant. AHP6 belongs a member of histidine phosphotransfer proteins that transduce cytokinin signal by phosphotransfer. However, AHP6 has a mutation in the conserved histidine residue, which is a target of phosphorylation of AHPs. Therefore, AHP6 cannot participate in phosphotransfer and is considered as pseudo AHP. In the *ahp6* mutant, differentiation of protoxylem is disrupted. However, the phenotype is restored by expressing cytokinin oxidase CKX2 from the *AHP6* promoter, which is expressed in protoxylem and pericycle cells [117]. The expression domain of auxin signalling markers overlaps with that of *AHP6* expression and auxin promotes transcription of AHP6, likely mediated by MP [116]. Thus, *AHP6* is an auxin inducible, negative regulator of cytokinin signalling that promotes protoxylem differentiation [117].

In procambium cells adjacent to the xylem, PIN1 was localized on both basal and lateral side of the plasma membrane. This lateral localization of PIN1 promotes auxin transport from procambium to protoxylem. Accumulated auxin in the protoxylem likely creates a bisymmetric AHP6 expression domain. During embryogenesis, *AHP6* is expressed in the cotyledon tips of the heart stage embryo and the expression domain migrates from cotyledon to vasculature cells. At the same time, expression pattern of the auxin signalling marker, IAA2 changes from symmetric to bisymmetric in vasculature cells. This suggests that AHP6

is required to the establishment of bisymmetric pattern of protoxylem during embryogenesis [116]. Likely, these examples are the first of many more that show intimate connections between auxin and cytokinin, as well as perhaps other hormones, that dynamically control cell fate and division during embryo development.

Concluding remarks

The morphogenetic potential of the plant hormone auxin has been discovered many decades ago (Skoog and Miller), but the mechanisms by which it controls embryo development have only been revealed in the last decade. In this chapter we have discussed the developmental progression of early embryogenesis, and have reviewed which steps are under auxin control. We show that a network involving the regulation of auxin biosynthesis, transport and cell-type-specific response allows this generic hormone to control a variety of processes during embryogenesis. While the outlines of this network have been drafted, important questions remain. These include how local biosynthesis is activated, how PIN protein polarity regulation leads to precise auxin accumulation patterns, and finally, how these accumulation patterns in turn trigger the activation of specific sets of developmental effector genes. With the current pace of progress, we anticipate that the next years will have much in store for our understanding of how this hormone directs multicellular plant development.

Acknowledgements

CHAPTER 3

Structural Basis for DNA Binding Specificity by the Auxin-Dependent ARF Transcription Factors

[1,2] * D. Roeland Boer
* Alejandra Freire Rios
* Willy van den Berg
Terrens Saaki
[3] Iain W. Manfield
[4] Irene López-Vidriero
[4] Jose Manuel Franco Zorrilla
Sacco C. de Vries
[4] Roberto Solano
Dolf Weijers
[1,2] Miguel Coll

These authors contributed equally to this work

1. Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10-12, 08028 Barcelona, Spain
2. Institut de Biologia Molecular de Barcelona (IBMB-CSIC), Baldiri Reixac 10-12, 08028 Barcelona, Spain
3. Astbury Centre for Structural Molecular Biology (IWM) & Centre for Plant Sciences (SK), Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, U.K.
4. Genomics Unit and Department of Plant Molecular Genetics, Centro Nacional de Biotecnología-Consejo Superior de Investigaciones Científicas, Campus Universidad Autónoma, 28049 Madrid, Spain

Abstract

Auxin regulates numerous plant developmental processes by controlling gene expression via a family of functionally distinct DNA-binding AUXIN RESPONSE FACTORS (ARFs). Despite the central importance of ARF-mediated transcriptional control in development, the mechanistic basis for the generation of specificity in auxin response is unknown. Here, we address this key question by solving high-resolution crystal structures of the pivotal *Arabidopsis* developmental regulator ARF5/MONOPTEROS (MP), its divergent paralog ARF1, and a complex of ARF1 and a generic auxin response DNA element (AuxRE). We show that ARFs homodimerize to generate cooperative DNA-binding that is critical for *in vivo* ARF5/MP function. Strikingly, DNA-contacting residues are conserved between ARFs, and we discover that monomers have the same intrinsic specificity. ARF1 and ARF5 homodimers, however, differ in spacing tolerated between binding sites. Our data suggest that ARF dimers bind complex sites as molecular calipers with ARF-specific spacing preference and provide an atomic-scale mechanistic model for specificity in auxin response.

Introduction

The plant hormone auxin controls numerous growth and developmental processes, and is a key determinant of plant architecture [68]. Physiological approaches in the early 20th century have led to the identification of indole-3-acetic acid as the main natural auxin [118]. In the past decades, genetic studies have revealed mechanisms of hormone biosynthesis[119], transport[120] and response[87]. The cellular response to auxin involves ubiquitin-proteasome-dependent degradation of Aux/IAA proteins, transcriptional co-repressors [89] that act by binding AUXIN RESPONSE FACTORs (ARFs) [92]. The latter are DNA-binding transcription factors that control the expression of the large set of auxin-dependent genes that mediate hormone-dependent growth and development [121]. A central, yet unanswered question in auxin biology is how the simple tryptophan-like indole-3-acetic acid can trigger a wide variety of cellular responses. As the last step in auxin signaling prior to gene regulation, the ARF transcription factors are likely components to confer specificity to auxin response through selection of target genes. Consistent with a role in response diversification, the ARF family consists of 23 members in *Arabidopsis thaliana*, and contains >10 members even in the moss *Physcomitrella patens* [34]. *ARF* genes are expressed in dynamic, and different patterns during development [44], and genetic studies have shown that individual ARFs control distinct developmental processes [44,49,99,121]. For example, ARF5 (also named MONOPTEROS [MP], but for consistency referred to as ARF5 here) is critically required for several developmental auxin responses, including embryonic root and flower formation [46,122], while ARF1 and 2 control senescence and floral organ abscission [123]. These differences between ARFs are at least in part due to differences in protein sequence, as misexpression and promoter-swap studies demonstrated that ARF proteins are not equal [49,99]. Instead, double mutant analysis suggested that ARF1 and ARF5 act antagonistically [44]. ARFs are modular transcription factors, consisting of several domains that have remained conserved despite hundreds of millions of years of evolution [34]. At their N-terminus, all ARFs have a DNA-binding domain (DBD), followed by a Middle Region (MR) that determines if ARFs activate or repress target genes [92] and a C-terminal interaction domain (domain III/IV). The latter has been shown to mediate interactions between ARFs and their Aux/IAA inhibitors, as well as between ARFs [124]. Several lines of evidence suggest that ARF domains are functionally autonomous, i.e. they act in isolation. First, both the DBD and the C-terminal interaction domains are found in other protein families. The DBD harbors a B3 DNA binding motif that is also found in many other plant transcription factors [125]. Similarly, the C-terminal interaction domain III/IV is also found in the interacting Aux/IAA proteins [124,92]. Secondly, transient expression assays and domain swaps have demonstrated that each of the three domains can act in isolation [92].

Here we address the atomic basis for sequence-specific DNA binding by ARF transcription factors, and explore mechanisms by which variation in the ARF DBD selects different target genes. Most ARFs tested have been shown to bind a generic auxin response element (AuxRE) ; [127] that was identified based on its occurrence in auxin-dependent promoters [30]. However, since target sites have not been screened exhaustively, it is unknown whether different ARFs prefer distinct binding sites and if so, what the molecular basis for such differences is. Here,

we have determined high-resolution crystal structures of the DNA-binding domains of two divergent ARFs, as well as an ARF-DNA complex. Structure-function analysis and saturating binding site selection lead to a re-defined ARF binding motif, as well as novel DNA-binding mechanism in which dimerization of ARF DNA binding domains generates cooperative binding to adjacent sites where spacing determines ARF binding affinity. Our study provides an atomic-level explanation for DNA-binding specificity in the auxin pathway.

Results

Crystal structures of ARF DNA-binding domains

All ARFs carry a conserved DNA-binding domain (DBD) at their N-terminus (Fig. 1A). This domain is often followed by a middle region (MR) that directs transcriptional changes and a C-terminal domain (III/IV) that mediates protein-protein interactions [92]. It is well-established that domain III/IV is essential for the heterotypic ARF-Aux/IAA interactions that render ARF activity auxin-dependent (Fig. 1A) [92]. The same domain has been proposed to mediate ARF-ARF interactions [124], but whether this is biologically meaningful has not been established. The ARF DBD is sufficient for binding auxin-responsive promoters [92]. Its B3 subdomain is found in other transcription factors (Fig. 1A) [125], and was shown to bind DNA in RAV1 [128], which suggests domain modularity. Interestingly, phylogenetic trees based on sequence alignments of only the DBD strongly resemble those derived from entire ARF proteins (Fig. S1), raising the possibility that variations in this domain contribute to the distinct properties of ARFs.

To gain insight into the mechanism of DNA binding by ARFs, we expressed and purified the DBD of ARF1 and ARF5/MONOPTEROS (MP). These two ARFs are phylogenetically distant (Fig. S1), and their divergence occurred early in land plant evolution hundreds of millions of years ago [34]. Based on mutant phenotypes and misexpression analysis, ARF1 and ARF5 are functionally divergent [44,103], although both proteins are able to bind the same generic core DNA motif [127]. Both proteins were purified to homogeneity and crystallized. The structure of ARF1-DBD was solved using single-wavelength anomalous diffraction (SAD) on a seleno-methionine (SeMet) derivative, and the resulting model was used for molecular replacement on the other structures (Table S1). We obtained two distinct structures of ARF1-DBD (solved to 1.45 and 2.67 Å resolution), and one of ARF5-DBD (2.15 Å). All three models subtly differed in the topology of several loops, but showed the same overall structure (Fig 1B-D; Fig. S2A,B). The crystal structures reveal that the ARF DBDs are composed of three distinct structural domains. One is the B3 domain (residues 120-226 in ARF1-DBD; 154-260 in ARF5-DBD; Fig. S2C), which folds in a 7-stranded open b-barrel structure (Fig. 1B,D), similar to the B3 domains of *Arabidopsis* RAV1 [128] and Atlg16640 [129]. Remarkably, the B3 domain in ARF1-DBD and ARF5-DBD is embedded in a larger fold context. The regions N- and C-terminal to the B3 domain together form a single second domain (Fig. 1B; Fig. S2A-C), that is very similar between ARF1-DBD and ARF5-DBD (Fig. 1B,C), and constitutes a dimerization domain (DD, see below). Thus, the
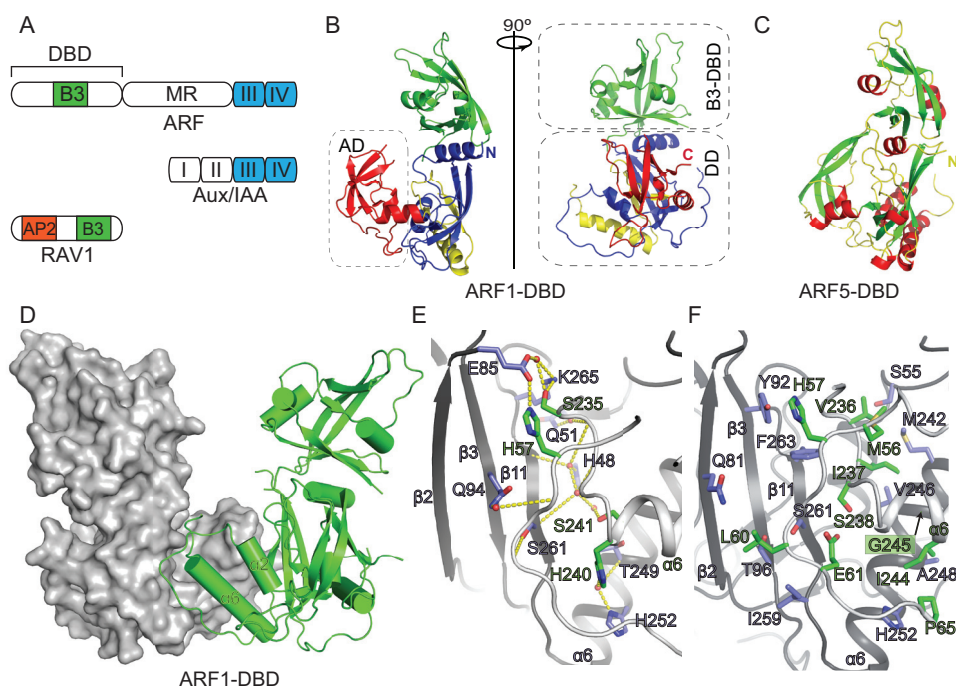
Figure 1: Structure of the ARF DNA binding domain.
(A) Modular domain organization of ARF, Aux/IAA and RAV1 proteins. B3-type DNA-binding domain (DBD) is also found in other B3 transcription factors such as RAV1. The C-terminal domains III/IV also occur in Aux/IAA proteins. Each of these families has other, family-specific domains (Middle Region [MR] in ARFs; domains I and II in Aux/IAAs; AP2 domain in RAV1).
(B) Crystal structure of an ARF1-DBD monomer in two orientations (90° rotation along indicated axis), colored by subdomains (B3 DBD in green; Dimerization Domain [DD] in blue and yellow [blue and yellow parts occur in sequence N- and C-terminal of the B3-DBD, respectively]; Ancillary Domain [AD] in red).
(C) Structure of ARF5-DBD monomer in orientation analogous to left image in (B). Secondary structure elements are colored in red (alpha helix) and green (beta strand).
(D) Crystal structure of ARF1-DBD dimer with one monomer shown in surface rendering, and the other monomer as cartoon with secondary structure elements indicated.
(E,F) Molecular interactions at the ARF1-DBD dimerization interface. Side chains and secondary structure elements are labeled in different colors for each monomer (blue/green). (E) Hydrogen bonds between the two monomers, indicated as yellow dashed lines. Water molecules are indicated as red spheres. (F) Hydrophobic interactions. N- and C-termini of proteins are indicated in (B,C). See also Figures S1, S2, S3 and S5 and Table S1.

B3 domain appears to be an insertion in the DD. Structure similarity searches (DALI [130]; PDBeFold [131]) on the DD show that there are no homologs in the Protein Data Bank, hence it defines a novel fold, characterized by an antiparallel 5-stranded central b-sheet which is highly curved (~100 °), resulting in a taco-like shape (Fig. 1B-D; Fig. S2D).

Finally, the last 80 C-terminal residues form a third separate ancillary domain (AD) that tightly interacts with the DD. The AD folds in a small 5-stranded b-barrel-like structure. Structural similarity searches identified the Tudor domain of the human PHD-finger protein 20 (PDB entry 3QII) but the hydrophobic cage that recognizes methylated lysine residues [132,133] is missing from the ARF-DBD-AD.

ARF dimerization through the DBD

Strikingly, in all crystal structures, ARF-DBDs homodimerized through their DD (Fig. 1D; Fig. S3A-C). The dimer interface contacts include hydrophobic interactions (Fig. 1F) between several highly conserved residues (Fig. S3D,E), which indicates that this is most probably a physiologically meaningful interaction. It is stabilized by a network of hydrogen bonds, some mediated by water molecules (Fig. 1E). Interestingly, alpha helix 6 (α6) of both monomers are juxtaposed centered at a conserved (Fig S3D,E) glycine residue (G245 in ARF1-DBD and G279 in ARF5-DBD). Other residues of this helix (A248, T249, A253; in ARF5: A282, A283 and A287) engage in hydrophobic interactions (Fig. 1E,F), while the P233-S238 loop (ARF5: P267-S272) fits into a groove of the opposite monomer, and involves interactions between S235 (ARF5: S269) on one monomer and K265 and E85 (ARF5: N299 and D118) on the other. (Fig. 1E; Fig. S3D,E).

To address if dimerization is induced by the crystallization conditions, or whether this also occurs in solution, small-angle X-ray scattering (SAXS) was used for ARF1-DBD. While neither monomer or dimer models explained the scattering data, a monomer:dimer equilibrium improved the fit dramatically (Fig. 2A). Hence, homodimerization also occurs in solution, and given that both ARF1 and ARF5 dimerize, this is likely a general property of ARF-DBDs.

To next determine if ARF-DBD homodimerization is required for biological function, we mutated several amino acids in the dimerization interface of ARF5/MP (Table 1), and tested the ability of mutant proteins to replace the wild-type protein *in vivo*. The *arf5/mp* mutant is unable to establish an embryonic root, and as a consequence, forms rootless seedlings [102]. Adventitious roots can however be induced post-embryonically, and mutant plants have distinctive growth defects, including aberrant flowers or even naked, pin-like inflorescences [122]. Importantly, while S269N, G279A and N299S mutations did not impair ARF5 activity during embryonic root formation (Table 1), G279E, G279I, A282N and A287N mutations all compromised ARF5 function *in vivo* (Table 1; Fig. 2B). In some cases, these mutated ARF5/MP proteins even induced dominant-negative defects in wild-type plants (Fig. 2C). To ascertain that the failure of these mutant proteins to complement the *arf5/mp* mutant is due to alterations in dimerization properties, rather than abnormal folding behavior, secondary structures were determined using circular dichroism (CD) spectroscopy. Consistent with the solubility of purified mutant proteins, none showed deviations in the CD spectrum (Fig. S4), suggesting that each folds normally. Hence, this analysis shows that amino acids at the dimerization interface, in particular in the alpha-6 helix (G279, A282, A287), are required
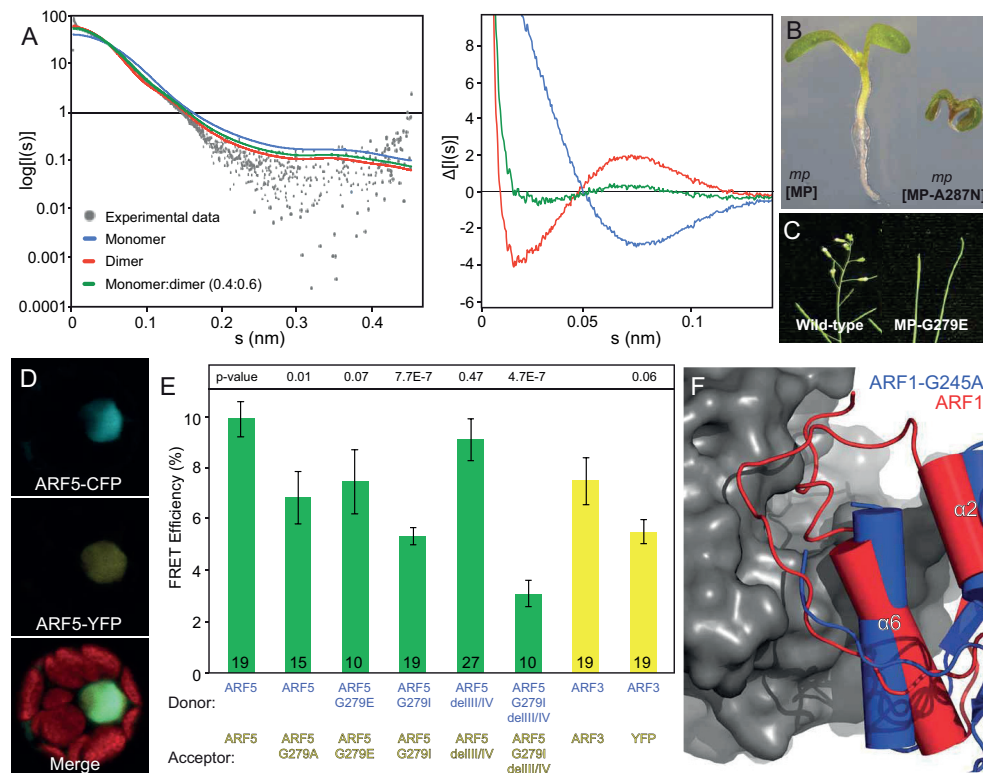
Figure 2. DBD dimerization is critical for in vivo ARF5 function.
(A) Small-Angle X-ray Scattering curve of native ARF1-DBD in solution. Left panel: Experimental data are shown as grey dots, the monomer fit in blue, the dimer fit in red and the fit of a monomer:dimer mixture (0.4:0.6 stoichiometry) in green. The right panel shows the deviations of the fitted curves from experimental data in the 0-0.15 nm range.
(B) Homozygous *mp/arf5* mutants carrying a wild-type ARF5/MP (left) or MP-A287N mutated transgene.
(C) Floral apex in wild-type (left) and MP-G279E transgenic plants.
(D) Expression and nuclear accumulation of ARF5-CFP (top) and ARF5-YFP (middle) proteins in a transfected mesophyll protoplast. Merge with chlorophyll fluorescence is shown in lower panel.
(E) Dimerization of ARF proteins as measured by FRET-FLIM in live protoplasts. Interaction is expressed as average FRET efficiency (+/- SEM), number of protoplasts is indicated in each bar, and the p-value for significance of difference between mutants and wild-type is shown (Student's t-test). The CFP donor is indicated in blue and YFP acceptor in yellow.
(F) Differences in the dimerization interface between ARF1 (red) and ARF1-G245A (blue). In the structural comparison, the position of one monomer (surface rendered) was fixed, and the relative orientation of wild-type and mutant protein is shown. Alpha helices are shown as cylinders. See also Table S1.

for ARF5/MP function *in vivo*.

To address whether these mutations indeed interfere with homodimerization in the context of a full-length ARF protein that also carries the C-terminal interaction domain (III/IV), we

Table 1. Complementation of *mp-B4149* mutant phenotypes with wild-type and engineered mutated ARF/MP transgenes.

Corresponding positions in ARF1 are indicated. Rescue was tested in T1 and T2 generations by scoring for the absence of rootless, transgenic seedlings (resistant to PPT). In addition, vegetative and floral defects were observed.

ARF5 Mutation	ARF1 position	Number of lines	Embryonic rescue (%)			Adult phenotype
			Full rescue	Partial rescue	No rescue	
DNA Binding Domain						
Wild-type		5	100			Normal
H170A	H136	8	75	25		Small, bushy, flower defects
R215A	R181	3		100		Sterile, small plants, small fruits
P218A	P184	6	100			Normal
R220A	R186	8	12.5	62.5	25	Flower defects, sterile
Dimerization Domain						
S269N	S235	5	100			Small, bushy
N299S	K265	3	100			Mild flower defects
A287N	A253	4	25	75		Curled leaves, flower defect
A282N	A248	5	40	60		Flower defect, sterile
G279E	G245	4	25	75		Pins/small plants
G279I	G245	2			100	Small plants, small fruits, sterile
G279A	G245	3	100			Normal

employed a FRET-based interaction assay. Here, interactions between CFP- and YFP-tagged ARF5 are quantified in mesophyll protoplasts (Fig. 2D; [134], and we have previously used this assay to demonstrate ARF-Aux/IAA interactions *in vivo* [49]. In this assay, wild-type ARF5 showed clear homodimerization as measured by a decrease in the average lifetime of the ARF5-CFP donor (Expressed as FRET efficiency; Fig. 2E). As expected from the position of the glycine in the dimerization interface, G279A, G279I and G279E mutations significantly decreased the FRET efficiency, and hence impair dimerization (Fig. 2E). To determine the relative contribution of the DBD and domain III/IV in homodimerization, a truncated ARF5 protein was generated, in which domain III/IV was deleted [106,135]. Homodimerization still occurred, albeit at lower efficiency (Fig 2E). Similarly, the ARF3 protein, which naturally lacks domain III/IV [127] was also able to homodimerize (Fig 2E; compare with ARF3-CFP / free YFP control). Hence, the DBD is sufficient for dimerization *in vivo*, but interactions through domain III/IV may help to stabilize dimers. Indeed, when the G279I mutation was introduced in the truncated ARF5 protein lacking domains III/IV, FRET efficiency dropped to background levels (Fig. 2E). Collectively, this data shows that ARF proteins form dimers through interactions between their DNA-binding domains, and that this DBD dimerization is required for ARF function *in vivo*. Sequence alignments show that the amino acids at the dimerization surface are deeply conserved in the ARF family (Fig S3D,E), suggesting

that this capacity is both widespread and ancient. The interaction surface is composed of many intermolecular interactions (Fig. 1E,F), which suggests that the interaction may be robust. Indeed, several mutations in residues at the surface (S269N, G279A, N299S) are tolerated *in vivo* (Table 1). To test if such mutations indeed affect the interaction surface, we purified and crystallized the ARF1-G245A mutant (analogous to ARF5-G279A) and solved its structure to 2.3 Å resolution (Table S1). This showed an overall dimeric structure similar to wild-type ARF1-DBD (Fig. S3C). As predicted, the structure showed a disturbed dimerization interface (Fig 2F), in which the two-fold symmetry is broken and the 233-238 loop of one monomer could not bind the opposing monomer. In summary, the ARF interaction is robust and adaptive, as it tolerates a mutation that significantly affects the interaction surface. Yet, as more drastic mutations (G279E, G279I) in the same residue that measurably interfere with dimerization (Fig. 2E) impair biological function (Table 1), we conclude that DBD dimerization is essential for *in vivo* function of ARF5.

Mechanism of DNA binding by ARFs

Based on promoter analysis of an auxin-responsive gene in soybean, a canonical auxin-response element has been defined as TGTCTC [30]. ARF1 was first identified in a screen for factors binding this motif [29]. An inverted repeat of the same element, spaced by 7 nucleotides (Fig. 3A; ER7), was shown to be efficiently bound by ARF1 [29]. To determine the structural basis for DNA binding, we co-crystallized ARF1-DBD and a double-stranded ER7 oligonucleotide and solved its structure to 2.9 Å (Table S1; Fig. 3B). The DNA binding interface is located at the tips of the U-shaped dimer. The two B3 domains bind to the inverted AuxRE elements located at both extremes of the oligonucleotide, and the connecting DNA sequence bridges the gap between the B3 domains (Fig. 3B). The DNA adopts a B-DNA conformation and is bent by 40º.

The structures of apo-ARF1-DBD (without DNA) and DNA-bound ARF1-DBD are very similar, except that the B3 domains are rotated relative to the DDs by 25º (Fig 3C,D). As a similar conformational difference is seen between the different apo structures of ARF1 and ARF5 (Fig. 1D; Fig. S3A), it appears that the B3 domain displays an intrinsic flexibility with respect to the DD, and that DNA binding locks the protein into a conformation. Comparing apo- and DNA-bound structures shows that the N-terminal α1 helix functions as a pivot point on which the B3 domain is balanced, and the loops that connect the B3 domain to the DD run down on both sides of the helix (Fig. 3C). Interestingly, these loops are mostly disordered in the structures, which indicates flexibility. Given the dimerization of the DBD, and the binding of each TGTCTC element to one of the monomers, this structure now explains the efficient binding of ARF1 to an inverted repeat sequence, as well as the constraints of the spacing between repeats [29].

Binding of two AuxRE sites by an ARF dimer suggests that DNA binding may be cooperative. To test if this is the case, we used surface plasmon resonance (SPR) with immobilized oligonucleotides. Both ARF1-DBD and ARF5-DBD showed binding to
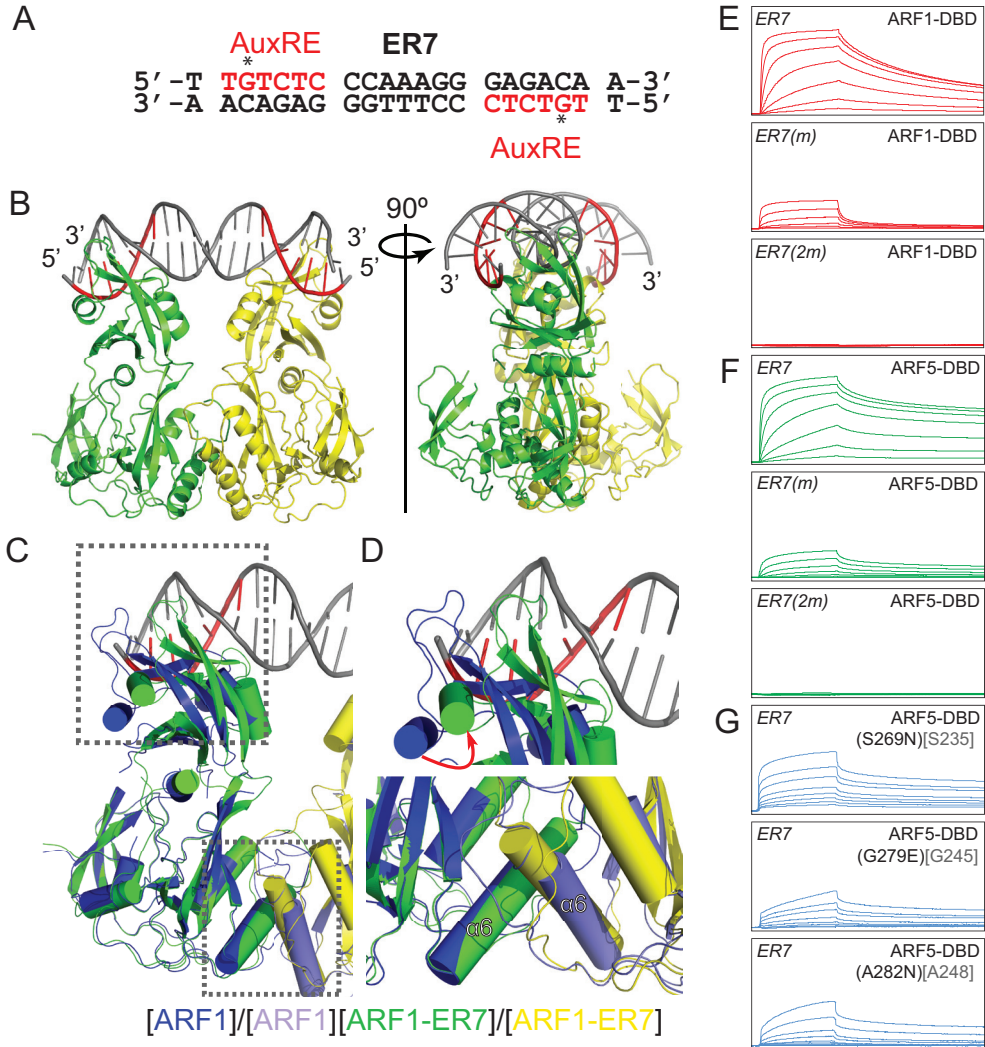
Figure 3: Cooperative DNA binding through ARF-DBD dimerization.

(A) ER7 nucleotide sequence with auxin response elements (AuxRE) indicated. Bases marked with an asterisk were mutated in (E and F).

(B) Crystal structure of an ARF1-DBD/ER7 complex, shown in two orientations (rotation axis indicated). ARF1-DBD monomers are differently colored.

(C) Superposition of apo-ARF1-DBD dimer (dark blue and light blue) and ARF1-DBD/ER7 (green and yellow) structures, showing the rearrangement of the B3 domains upon DNA binding.

(D) Details of the dimer interface (lower panel) and the protein-DNA interface (upper panel) from structure in (C).

(E,F) SPR binding profiles of ARF1-DBD (E) and ARF5-DBD (F) on ER7, as well as ER7 with either one (ER7(m)) or two (ER7(2m)) Guanines mutated to Adenine as indicated in (A). Protein concentrations are 800 nM in the upper lines, and 6.25 nM in the lower line, and increase by steps of 2.

(G) SPR binding profiles of ARF5-DBD(S269N), (G279E) and (A282N) mutant protein to ER7 oligonucleotide. Corresponding amino acids in ARF1-DBD are indicated in gray. Scales on X- and Y-axis is identical in all panels in (E) and (F). See also Figures S3, S4 and S5 and Table S1 and S2.
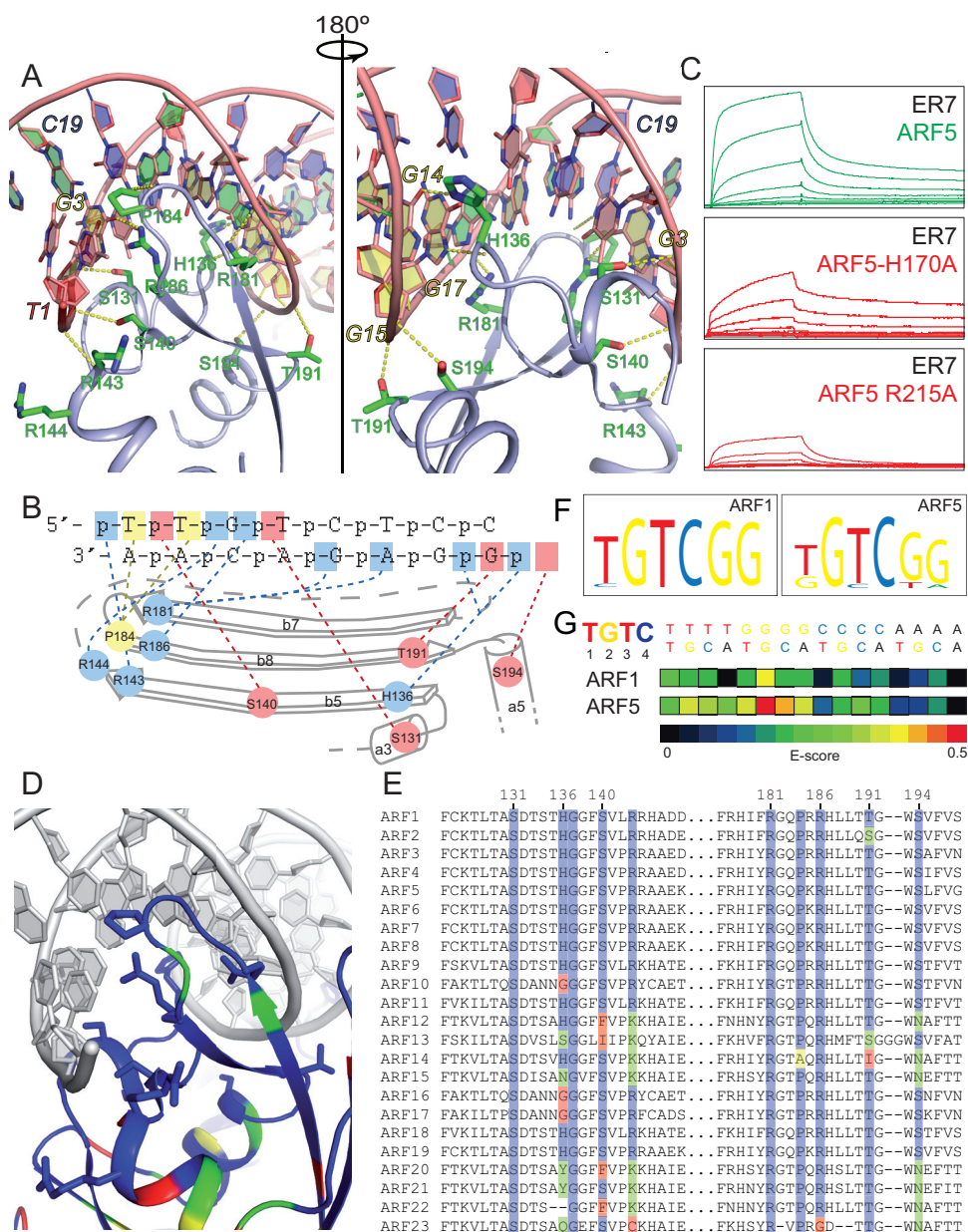
Figure 4. Sequence-specific DNA binding by ARF-DBDs.

(A) Detail of DNA-protein interface of the ARF1-DBD/ER7 complex showing the residues involved in DNA recognition. The two views show the same interaction surface, rotated by 180 degrees. Bases are colored and labeled in italics, and DNA-contacting residues are labeled.

(B) Scheme of intermolecular contacts between ARF1 protein and ER7 DNA bases (A,C,G,T) or backbone phosphates (p). Positively charged amino acids are marked in light blue, polar residues in pink and Proline in yellow. Dashed lines indicate atomic interactions.

(C) SPR binding profiles of wild-type ARF5-DBD and H170A (ARF1: H136) and R215A (ARF1: R181) mutants on ER7 oligonucleotides. Scales are identical in all three panels.

(legend continued on next page)

ER7 (Fig. 3E,F; Table S2). Binding was sequence-specific, as mutating both AuxRE sites completely abrogated binding (Fig. 3E,F; Table S2). Consistent with cooperative binding, mutating only one of the two AuxRE sites reduced the affinity more than 2-fold (Fig. 3E,F; Fig. S3F,G; Table S2). This cooperative binding behavior of wild-type ARFs predicts that mutations that disturb dimerization should affect DNA binding affinity. We tested this prediction by performing SPR measurements using S269N, G279E and A282N mutant proteins. Indeed, all three proteins showed a clear reduction in DNA binding affinity to the ER7 oligonucleotide (Fig. 3G; Table S2). These findings show that dimerization of the ARF DBD generates cooperative DNA-binding behavior. As dimerization is important for *in vivo* function of ARF5, this suggests that cooperative DNA binding is essential for normal ARF function.

Recognition and specificity of DNA binding

We next addressed the structural basis for specific DNA binding. The B3 domain recognizes the DNA largely at the major groove of both TGTCTC elements (Fig. 4A,B; Fig. S5C,D). The B3 b-barrel is positioned laterally to the DNA with the axis of the barrel almost parallel to the axis of the DNA double helix. Two adjacent b-strands (b5 and b8) run over the major groove, parallel to the two sugar-phosphate backbones. The loops connecting these strands (R181-R186 and H136-G137; ARF5: R215-R220 and H170-G171), located on either side of the barrel, further penetrate the major groove, and make interactions that contribute to specific DNA base recognition (Fig. 4A). In addition to these base contacts, DNA binding involves interactions of the DNA backbone with residues S131, S140, T191 and S194 (Fig. 4B; ARF5: S165, S174, T227, S230). In summary, ARF1 binding to the canonical TGTCTC motif involves both base contacts and backbone interactions, and the specific contacts involve only the 5' bases on one strand, which explains why this part of the motif is critical for ARF DNA binding [127].

To determine if the residues that mediate DNA binding in the crystal structure are also required for DNA binding in solution, and for ARF function *in vivo*, several residues were individually mutated to alanines. Even though CD analysis showed that these mutant proteins showed normal overall structure (Fig S4), SPR analysis of ARF5-H170A (ARF1: H136) and ARF5-R215A (ARF1:R181) proteins revealed that in both cases ER7 binding was significantly reduced (Fig. 4C; Table S2), which supports a role in DNA binding.

(D) Conservation of amino acids between ARF1 and ARF5 mapped onto the ARF1-DBD/ER7 protein-DNA interface. Blue: identical; Green: conserved; Yellow: semi-conserved; Red: non-conserved.
(E) Sequence alignment of DNA-contacting loops in the 23 Arabidopsis ARF proteins. Positions that directly contact DNA in ARF1 are colored, and conservation is marked using the same color code as in (D).
(F) Logos of binding motifs identified for ARF1-DBD and ARF5-DBD in protein-binding microarrays (PBM).
(G) Binding preference according to PBM (in E-score according to color scale) of ARF1-DBD and ARF5-DBD to all possible hexamers starting with TGTC. See also Figure S4 and Table S2.

Correspondingly, neither of the H170A and R220A (ARF1: R186) mutants could restore normal development to the *arf5/mp* mutant, while the P218A (ARF1: P184) mutant was still partially functional (Table 1). These findings support the validity of the protein-DNA contacts observed in the complex structure and show the atomic basis for DNA recognition by ARF proteins.

Intriguingly, when comparing the DNA-binding amino acids among and between ARF1 and ARF5, all appear to be almost completely conserved (Fig. 4D,E; Fig. S2C). This finding raises the question whether ARF1 and ARF5 bind qualitatively different sequences. Even though a generic AuxRE has been defined, no systematic exploration of sequence space has been reported. To determine the spectrum of binding motifs of each protein, we carried out a protein-binding microarray (PBM)[136] analysis with recombinant ARF1-DBD and ARF5-DBD. This analysis is saturating for 6-mers [136] and allows statistical and quantitative evaluation of intrinsic binding site preference for the ARF-DBDs. However, longer motifs, such as inverted AuxRE repeats that would be expected for ARF dimers, can not be reliably identified. Strikingly, the preferred binding site of both ARF1 and ARF5 appears to be TGTCGG (Fig. 4F,G), rather than the "canonical" AuxRE TGTCTC [30]. It should be noted however, that the latter is found as an enriched motif, although the former is strongly preferred (Fig. 4G). We did not observe a significant difference in the motifs bound by ARF1-DBD or ARF5-DBD (Fig. 4G), which is consistent with the invariance of DNA-contacting residues (Fig. 4D,E).

We therefore conclude that the B3 domains of ARF1 and ARF5 do not have qualitatively distinct DNA binding specificity, but bind the same spectrum of motifs with quantitatively different efficiency.

We next tested which residues in the ARF5 protein contribute to sequence-specific DNA binding, by performing PBM analysis on mutant proteins. Sequence-specific binding was lost when either P218 or R215 residues were mutated (Fig. 5), while H170A and G171A mutations did not affect binding specificity. This analysis thus helps identify residues within the B3 domain that confer binding specificity (R215, P218), and distinguish these from residues that contribute to DNA affinity (H170; Fig. 4C). We next analyzed DNA-binding specificity of proteins impaired in dimerization. Neither ARF5-S269N nor ARF5-G279E altered the PBM binding profile (Fig. 5). This shows that dimerization contributes to DNA binding affinity, but not to the specificity of DNA motif recognition.

Motif spacing constrains specific ARF binding

We found that ARF proteins are extremely conserved at both their dimerization interface and their DNA-contacting residues. As a consequence, both ARF1 and ARF5 dimerize, and bind qualitatively similar sequences. A key unanswered question therefore remains how different genes can be selected by different ARFs. We noticed that the largest variation between ARF1 and ARF5 DBDs is in the loops that connect the B3 and DD domains (Fig. 6A). Therefore, in addition to quantitative differences in binding of the two ARF
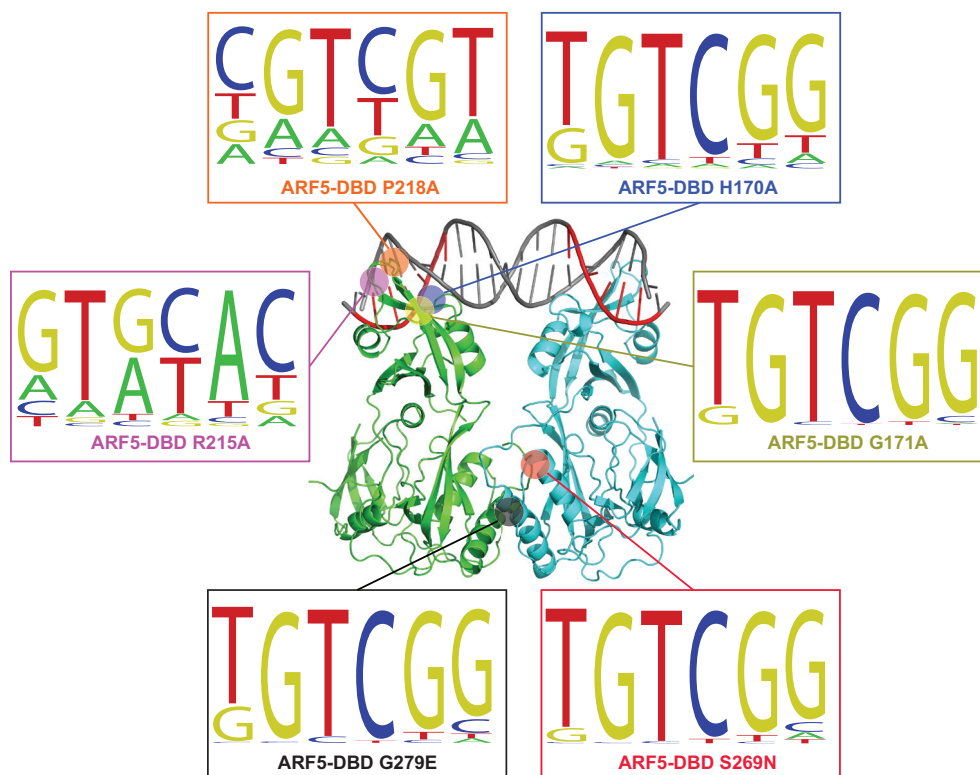
Figure 5. DNA binding specificity through local protein-DNA interaction.
PBM binding profiles of mutant ARF5-DBD proteins. The positions of mutated amino acids are indicated in the ARF1-DBD-ER7 protein-DNA complex. Note that mutations in the dimerization interface do not affect DNA binding specificity, while two DNA-contacting amino acids are indispensable for correct DNA binding specificity. See also Figure S4.

DBDs to distinct sequence motifs, one could envisage differences in binding of ARF dimers to complex motifs with varying spacing between AuxRE sites. To test this hypothesis, we performed SPR experiments using ER7 oligonucleotides in which the spacing between the two inverted TGTCTC sites was changed to 5 (ER5), 6 (ER6), 8 (ER8) or 9 (ER9). Both ARF1-DBD and ARF5-DBD bound to ER8 with similar efficiency as ER7 binding (Fig. 6B,C; Table S2). Affinity of ARF1-DBD to ER5, ER6, and ER9 was strongly reduced (Fig. 6B; Table S2), to a level comparable to that of ER7 with one TGTCTC site mutated (Fig. 3E; Table S2), or to that of a mutant impaired in dimerization (Fig. 3G; Table S2). In contrast, ARF5-DBD retained significant binding to all ER versions, although binding efficiency to ER5, ER6 and ER9 was slightly reduced compared with ER7 and 8 (Fig. 6C; Table S2). Hence, in addition to quantitative differences at the level of binding sites, ARF1 and ARF5 markedly differ in their ability to bind complex motifs depending on the spacing of the two binding sites. Such complex sites, with appropriate spacing, are indeed found in
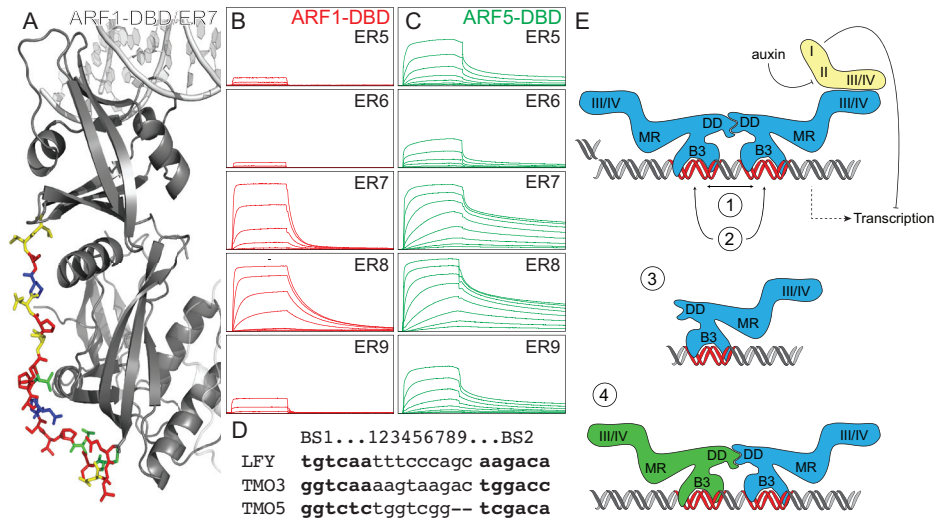
Figure 6: A spacing-based model for ARF DNA-binding specificity.
(A) Conservation of amino acids between ARF1 and ARF5 mapped onto the external loop in the ARF1-DBD/ER7 structure. Blue: identical; Green: conserved; Yellow: semi-conserved; Red: non-conserved.
(B,C) SPR binding profiles of ARF1-DBD (B) and ARF5-DBD (C) proteins to ER5, ER6, ER7, ER8 and ER9 (number indicates spacing between two inverted TGTCTC elements). Values were normalized to the highest value of that same protein on the ER7 oligonucleotide. Scales are identical in all panels.
(D) Complex ARF binding sites in ARF5/MP target genes LFY, TMO3 and TMO5. Binding sites are in bold and intervening bases are numbered.
(E) Model for auxin-dependent transcription. ARF proteins bind DNA as dimers, mediated by interactions in the DNA-binding domain. The main determinant of specificity is the spacing between the two binding sites (1), although quantitative differences in preferences for binding sites may exist (2). Auxin controls ARF activity by promoting degradation of Aux/IAA proteins that bind to the distant domains III/IV. ARFs may also act through low-affinity DNA-binding as monomers (3), and heterodimerization (4) may further extend the range of binding preference. See also Table S2.

the promoters of direct, and physiologically relevant ARF5/MP target genes (Fig. 6D) [103,137].
In the case of *LEAFY*, mutating this site abrogated MP-dependent gene regulation[137], which suggest that it represents a physiologically relevant binding site

Discussion

Our work provides the molecular basis for sequence-specific DNA binding by the ARF transcription factors (Fig. 6E). Strikingly, homodimerization of the DNA-binding domain generates cooperative binding behavior. Mutations that affect this dimerization do not qualitatively affect DNA binding specificity, but reduce affinity, and their inability to rescue the *arf5/mp* mutant demonstrates the biological relevance of cooperativity for ARF function.

This behavior introduces a non-linear element in auxin response, which may contribute to the switch-like mode suggested for auxin activity [138], and could confer robustness against small fluctuations in auxin levels [139]. Furthermore, the ability of monomers to bind DNA, albeit with much lower affinity, suggests that genomes may harbor distinct high-affinity and low-affinity ARF binding sites. As dimer-bound high-affinity sites are more constrained due to specific spacing requirements, our study predicts that low-affinity sites will outnumber high-affinity sites.

Our results also show that domains III/IV are not likely to be important for ARF dimerization upon DNA binding. Previously, a systematic yeast 2-hybrid-based interaction analysis among and between ARF and Aux/IAA protein families concluded that ARF homodimerization is limited [140]. However, this study used only the previously known C-terminal domain III/IV. Our finding that the DBD represents a critical ARF dimerization domain calls for re-evaluation of this and other studies. Given the high degree of conservation of residues at the ARF-ARF interaction interface (Fig. S3E), along with the notion that the phylogenetically diverse ARF1, ARF3 and ARF5 all homodimerize, it is likely that most, if not all ARFs homodimerize through their DBD. Whether or not ARFs can also heterodimerize and if this would be biologically meaningful is another interesting open question.

The identification of the ARF DBD as a dimerization domain in addition to the previously known domain III/IV also has interesting implications for both DNA recognition and auxin regulation. First, given domain modularity, it is possible that DNA-bound ARF dimers interact through their domains III/IV with other ARF dimers to build higher-order complexes, analogous to what has been suggested for MADS-box transcription factors [141]. A testable prediction from such interactions would be that ARF complexes can bind more distantly spaced sites, and induce DNA looping. Secondly, the ability of domains III/IV to mediate both homotypic (Aux/IAA or ARF-ARF) and heterotypic (ARF-Aux/IAA) interactions suggested that Aux/IAAs may obstruct ARF dimerization (reviewed in [45]). The existence of an additional dimerization domain in the ARFs suggests a different mechanism of Aux/IAA function. Although Aux/IAAs could in principle modulate the stability of ARF-ARF dimers formed through their DBDs, an attractive alternative hypothesis is that they act as competitive inhibitors to prevent the formation of domain III/IV-mediated higher-order DNA-bound ARF complexes.

ARFs can have overlapping [44,101,123,142], different [44,49,99], or even opposing [44] functions, and an important question is how these different activities are encoded in their structures. Often, for example in the Homeodomain family [143,144], variation in sequence-specific DNA binding in transcription factor families is generated by substitutions in the DNA-contacting residues. In contrast, intrinsic DNA-binding specificity among ARF proteins is highly similar, even between the phylogenetically diverse family members ARF1 and ARF5, a finding that is consistent with the limited sequence divergence at the DNA-binding surface. One potential caveat is that, even though ARF domains can fold and act in isolation (this study and [92]), it is possible that other domains alter DNA binding specificity by intramolecular interactions with the DBD. Particularly the Middle Region is very divergent between ARFs, and correlated with the ability of ARFs to either activate or repress transcription [92].

Nonetheless, we find that dimerization allows for variation of ARF DNA recognition at the level of spacing between two adjacent inverted binding sites. This mechanism, in which ARFs act as "molecular calipers" to bind uniquely spaced motifs, can at least account for differences between ARF1 and ARF5, and is consistent with *in vivo* binding sites for ARF5 [103,137]. The divergence in the loops connecting the B3 and DD domains extends beyond ARF1 and ARF5, and it is therefore conceivable that other ARFs also have distinct inter-domain flexibility that allows unique binding site spacing. It will be interesting to address what distance can be accommodated by ARF complexes, and if two binding sites can be separated by larger DNA loops or nucleosomes. In this context, it is important to note that the distance of 7 bases between AuxRE sites in the ER7 substrate requires little or no protein conformational change or torsion of the DNA (Fig. 3B). In contrast, increasing or decreasing this distance will also rotationally displace the two binding sites. Hence the different potential in binding between ARF1 and ARF5 depending on site spacing may either be a consequence of different flexibility of the dimer, or a difference in the capacity of the two proteins to induce DNA bending or torsion. Given the different biophysical properties of A:T and G:C pairs, sequence within the spacer may also contribute to binding affinity.

Transcription factor dimerization is a common element in transcriptional control. Often, dimerization is required for binding a single site, such as is the case in basic Helix-Loop-Helix (bHLH) factors (e.g. MyoD [145]), or by bZip factors (e.g. AP-1/CREB [146]). Unlike many other examples however, ARF DNA binding can involve either one or two binding sites, where the latter case involves cooperativity. Conceptually, the mechanism underlying sequence-specific DNA binding in the ARF family is similar to that found in the animal Nuclear Receptor (NR) family. Members of this family of transcription factors, whose nuclear localization and activity is modulated by membrane-permeable hormones such as Retinoic Acid or Estrogen [147], bind DNA either as monomers or dimers [148]. When bound as dimers, the choice of the partners determines the optimal spacing (3, 4 or 5 bases) between two tandem binding sites, a phenomenon that led to the formulation of the 3,4,5-rule [149]. The case with ARFs is distinct as symmetric homodimers bind an inverted repeat rather than a tandem repeat as bound by the NR dimers, and in addition the space between two binding sites is large for ARFs. Nonetheless, both NR [150] and ARFs can bind DNA cooperatively, and contribute to generating specific responses to hormonal signals in the animal and plant kingdom.

A key question in auxin biology is how this structurally simple molecule can elicit such a wide range of growth and developmental responses. Our study suggests a model where diversification of gene expression responses follows from the distinct properties of dimeric complexes formed by the DNA-binding ARF transcription factors. This model, as well as the ARF structures presented here will now open new avenues to define the mechanistic basis for context-dependent gene regulation in the auxin pathway.

Materials and Methods

Protein expression and purification.

Regions corresponding to the DNA binding domain (DBD) of Arabidopsis ARF1 (At1g59750; residues 1-354) and ARF5 (At1g19850; residues 1-390) were amplified from cDNA clones using Phusion Flash polymerase (Finnzymes), and cloned in an expression vector pTWIN1 (New England Biolabs) to generate fusions with Chitin Binding Domain (CBD) and Intein. ARF-DBD-CBD fusion proteins were expressed in *E. coli* strain Rosetta DE3 (Novagen). Protein expression was induced by 0.3 mM IPTG for 20 hours at 20 °C, and proteins were purified from cell-free extracts by affinity chromatography on a chitin column followed by size exclusion chromatography on a Superdex 200PG column, both using an Akta Explorer 100 (GE Healthcare). Full details on expression and purification are described in the Supplemental Experimental Procedures.

X-ray crystallography and structure refinement

All crystals were grown at 20 °C using sitting drop vapor diffusion experiments. Initial screens were performed using 80-200 nl droplets on 96-well plates, using a Cartesian robot. Additive screens on initial hits showed improved crystal size and longevity with GSH/GSSG. Additional trials using the reducing agents GSH and DTT confirmed the dependence of crystal growth and stability on the reduction potential of the environment. Oligonucleotides used for crystallization were obtained from Biomers (Ulm, Germany). Full details on crystallization conditions, data collection (ESRF beamlines ID14-1, ID14-4; ALBA beamline XALOC) and analysis is given in Supplemental Experimental Procedures.

Small-Angle X-ray Scattering measurements.

SAXS data of ARF1-DBD (concentration 3.2 mg/ml) were collected at beamline BM29 (ESRF, Grenoble). Bovine Serum Albumin (BSA) references were used for calculating the molecular mass of ARF1-DBD. Measurements were carried out at 293 K, within a momentum transfer range of 0.01 Å^{-1} < s < 0.45 Å^{-1}. Calculation of the theoretical scattering curves of monomeric and dimeric ARF1 against the scattering data was performed using CRYSOL [151].

Surface Plasmon Resonance and Circular Dichroism Spectroscopy

SPR measurements were performed using eight two-fold dilution steps (800nM, 400 nM, 200 nM, 100 nM, 50 nM, 25 nM, 12.5 nM, 6.25 nM) of purified ARF-DBD proteins on a Biacore 3000 platform using double-stranded biotin-labeled oligonucleotides (5'-biotin; Eurogentec) immobilized on SA chips (GE Healthcare). Data were analyzed with Scrubber2-T200 (BioLogic Software Pty Ltd).

Circular Dichroism was performed on 0.1 mg/ml dilutions of purified ARF DBD proteins in 0.1 M Sodium Borate buffer (pH 7.4) using a 1 mm quartz cell in a J-715 CD spectropolarimeter (Jasco). Traces are averages of 20 spectra and smoothened over 3 nm windows.

Site-directed mutagenesis and cloning

Mutations were introduced into cDNA fragments corresponding to the ARF DBD's through PCR, and fragments were cloned into pTWIN1. The wild-type and mutated cDNA of ARF5-DBD were amplified and used to replace the genomic DBD in an 8.5 kb *MP* genomic fragment [106] using the unique restriction sites *Xho*I and *Bam*HI. Wild-type and mutant versions of the full length ARF5 cDNA, or a fragment truncated after T794 (after [135]) were LIC-cloned [152] into the PMON999 (Monsanto) vector and fused to sCFP3A or sYFP2 and transiently expressed in *A. thaliana* Columbia ecotype mesophyll protoplasts under the 35S promoter for FRET-FLIM assays. The ARF3 plasmids were previously described [49].

Protein-binding microarrays

PBM11 was performed on ARF1-DBD, ARF5-DBD and their mutated versions H170A, G171A, R215A, P218A, S269N and G279E according to [136] with modifications detailed in the Supplemental Experimental Procedures.

Plant growth and rescue experiments

Heterozygous plants of the *mp-B4149* strong allele [99] were transformed with a pGREEN vector carrying the construct pMP::MP and its different mutated versions by floral dipping with *A.tumefaciens*. Seeds carrying the transgene were screened on MS media with 15 mg/ml phosphinotricin (PPT). Segregation of the *monopteros* phenotype in the T2 generation was checked to determine the genotype of the T1 plants. The percentage of rootless seedlings observed in the progeny of heterozygous T1 plants was used to determine if the transgene could rescue the phenotype.

FRET-FLIM

Transfections were performed as described (Russinova et al., 2004) using Arabidopsis (Columbia wild-type) mesophyll protoplasts. FLIM images were acquired with a Leica TCS SP5 X system equipped with a 63X 1.20 NA water-immersion objective lens. sCFP fluorophore was excited using a pulse diode laser (40nm) at a frequency of 40mHz. Donor fluorescence was recorded via an external fiber output connected to the Leica SP5 X scan head and coupled to a Hamamatsu HPM-100-40 Hybrid detector (Becker & Hickl), which has a time resolution of 120 ps. Donor fluorescence emission was obtained using a 470-500nm band pass filter. Images of 128x128 pixels were acquired with acquisition times of 120 seconds. FRET-FLIM analysis in Arabidopsis leaf mesophyll protoplasts was performed as described previously [49].

Acknowledgements

Supplemental Figures and Tables

Figure S1, related to Figure 1: Phylogenetic relationship of Arabidopsis ARFs and their DNA-binding domains.
Phylogenetic trees of the 23 full-length Arabidopsis ARF proteins (left) or only the DNA-binding domain (right) were generated using the ClustalW program. The ARF1 and ARF5 proteins are labeled in red and green, respectively.
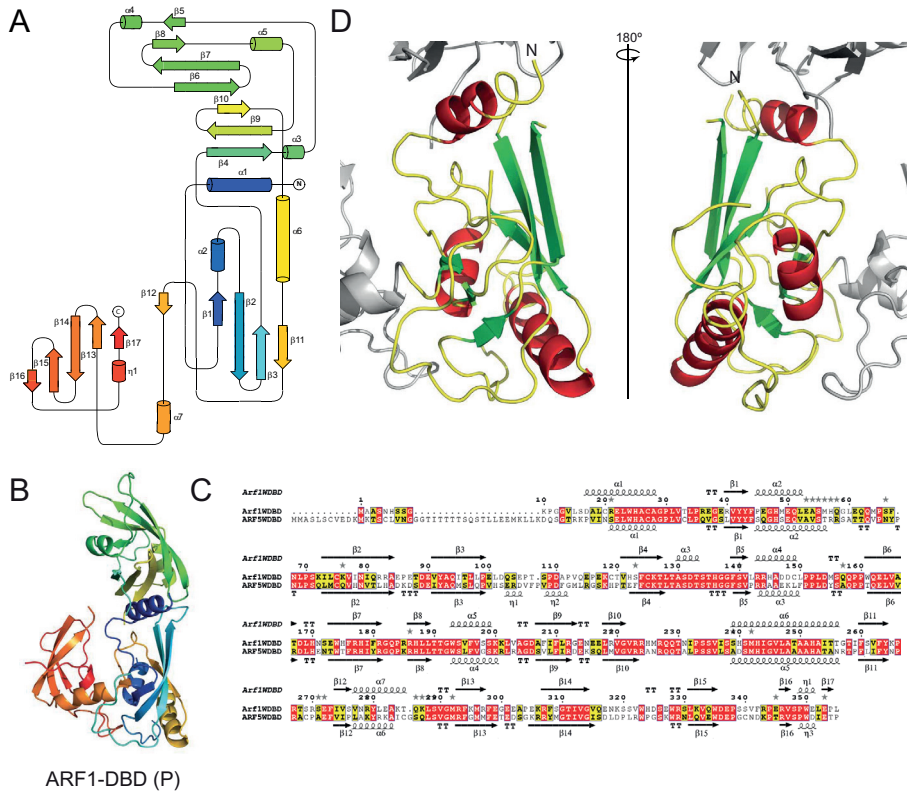
Figure S2, related to Figure 1: Structures of ARF1-DBD and ARF5-DBD, and new "taco" fold.

(A) Secondary structure topology of ARF1-DBD, with color scale running from blue (N-terminus) to red (C-terminus). Alpha helices are depicted as cylinders and beta strands as arrows. Secondary structure elements are numbered successively.

(B) Crystal structure of ARF1-DBD-P, with colors corresponding to (A).

(C) Alignment of ARF1 and ARF5 DBD sequences showing positions of secondary structure elements.

(D) The novel "taco"-like fold of the dimerization domain of ARF1-DBD, shown as a cartoon drawing in two orientations. The B3 domain (top) and ancillary domain (side) are shown in light grey
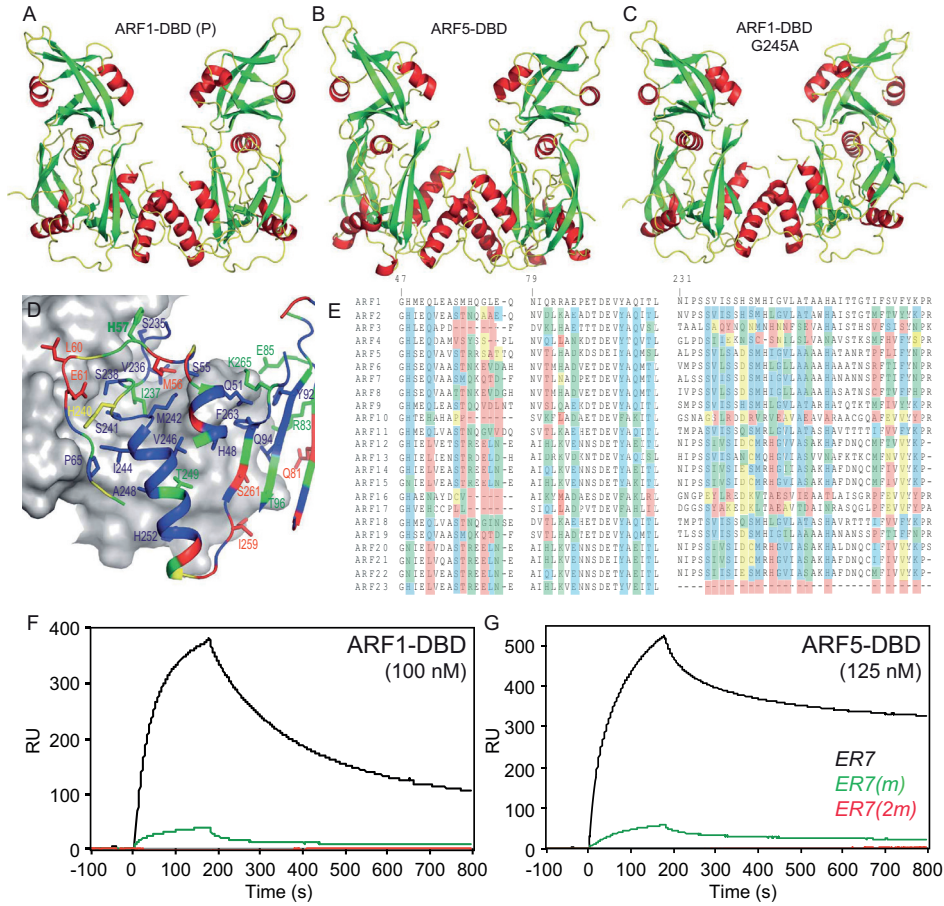
Figure S3, related to Figure 1 and 3: Dimerization and cooperative DNA binding by ARF-DBDs

(A,B,C) Crystal structures of ARF1-DBD-P (A), ARF5-DBD (B) and ARF1-DBD-G245A mutant (C) dimers colored by secondary structure elements (helices: red; sheets: green).

(D) Conservation of dimer contacts. Residues in the dimer interface are colored by homology in the *Arabidopsis* ARF family. Blue: conserved; Green: conservative; Yellow: similar; Red: non-conserved. (B) Alignment of *Arabidopsis* ARF sequences, with conservation of residues colored as in (D).

(F,G) Binding of ARF1-DBD (F) and ARF5-DBD (G) to ER7 oligonucleotide, as well as ER7 with 1 (ER7(m)) or 2 (ER7(2m)) AuxRE sites mutated. In all cases, the same concentration of ARF1-DBD (100 nM) or ARF5-DBD (125 nM) protein was used, and all measurements were performed on the same SPR chip. RU: Response Units.
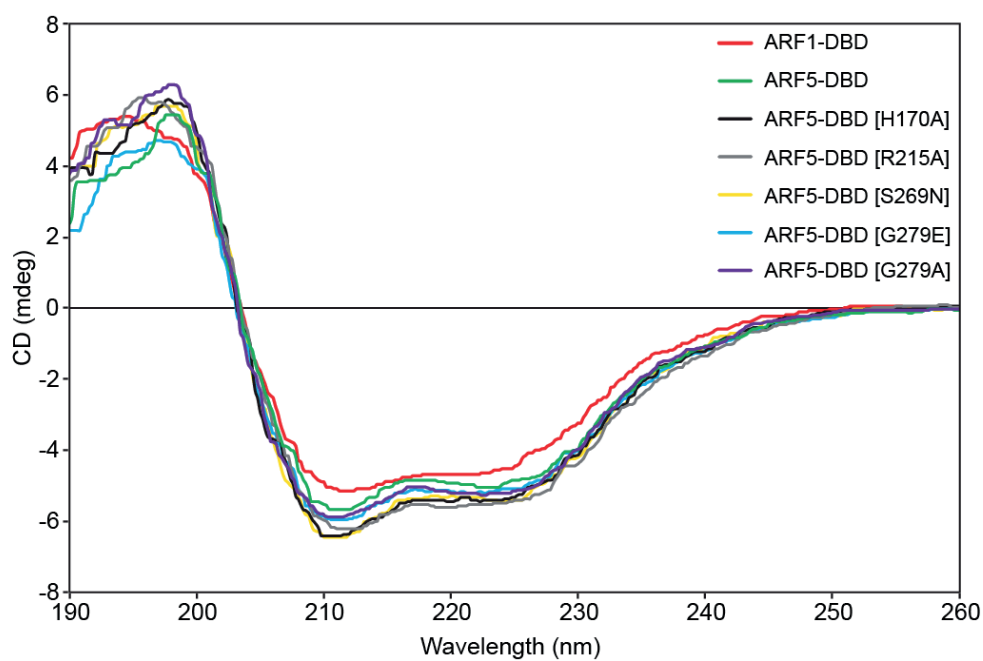
Figure S4, related to Figure 3, 4 and 5: Circular Dichroism spectra of ARF DBD proteins
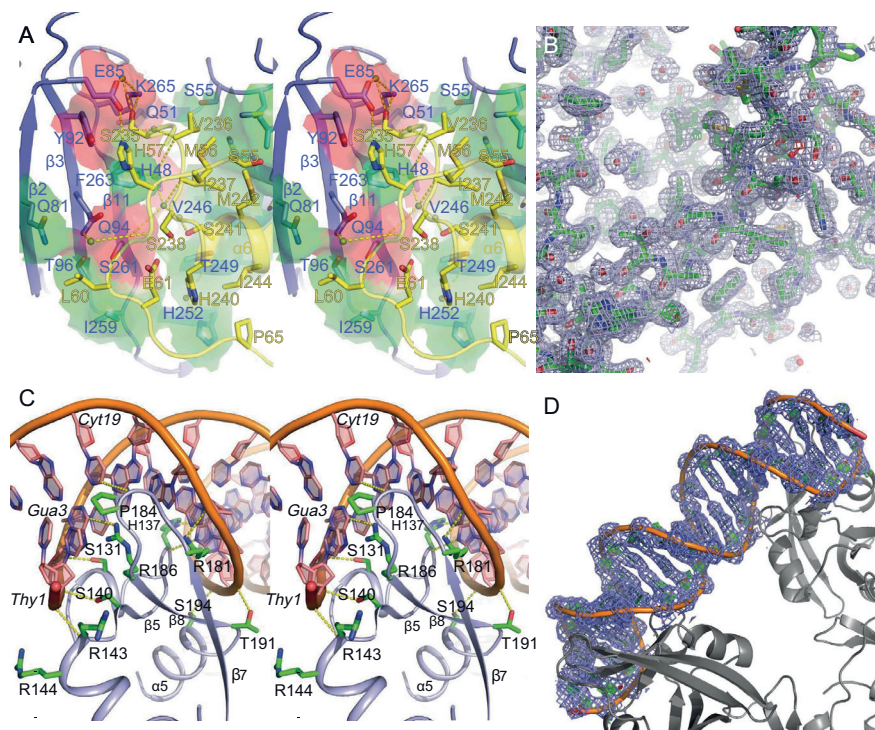
Figure S5, related to Figure 1 and 3: Dimerization and DNA binding by ARF1-DBD
(A,C) Stereo view of the ARF1-DBD dimer interface and (A) and the ARF1-DBD-ER7 protein-DNA interface (C).
(B,D) 2Fo-Fc electron density map (blue) of the dimer interface of the apo ARF1-DBD (B) and the protein-DNA interface in the ARF1-DBD-ER7 complex (D), contoured at 1s. The difference density maps Fo-Fc are also shown, contoured at -3.5s (red) and 3.5s (green).

Table S1, related to Figure 1, 2 and 3: Crystallographic statistics.

	ARF1-DBD-SeMet	ARF1-DBD	ARF1-DBD-P	ARF5-DBD	ARF1-DBD-ER7	ARF1-DBD-G245A
Data collection						
Data source	ID14-4 (ESRF)	ID14-4 (ESRF)	ID14-4 (ESRF)	ID14-1 (ESRF)	ID14-1 (ESRF)	XALOC (ALBA)
Space group	C2221	C2221	P21	C2	P21	P21
Cell dimensions						
a, b, c (Å)	88.21 126.49 84.00	88.36 126.26 83.67	73.26 83.67 79.03	69.61 91.50 77.18	43.52 105.19 127.91	73.10 84.24 78.96
α, β, γ (°)	90 90 90	90 90 90	90 117.61 90	90 114.28 90	90 98.14 90	90 114.76 90
Resolution (Å)	39.05-1.61	39.1-1.45	70.0-2.67	45.0-2.15	35-2.9	71.7-2.30
High resolution shell	1.70-1.61	1.53-1.45	2.81-2.67	2.27-2.15	3.06-2.9	2.42-2.30
Rmerge (%)[a,c]	8.1 (80.0)	9.3 (57.8)	9.8 (65.0)	6.3 (42.5)	15.6 (63.4)	10.2 (62.8)
Rmeas (%)[b,c]	8.9 (87.7)	10.5 (72.5)	11.7 (80.1)	7.8 (52.2)	19.9 (81.3)	12.4 (75.4)
$<I> / \sigma<I>$ [c]	11.5 (2.1)	7.3 (1.4)	6.9 (2.2)	11.0 (2.9)	6.2 (1.7)	5.9 (1.8)
Completeness (%)[c]	99.7 (100)	98.9 (99.9)	96.6 (96.2)	99.8 (100)	98.7 (98.8)	99.9 (100)
Multiplicity c	5.9 (6.0)	3.2 (2.7)	3.8 (3.7)	2.8 (2.8)	2.6 (2.6)	3.1 (3.2)
Refinement						
Resolution (Å)	Not refined	72.4-1.45	35.9-2.67	38.4-2.15	33.8-2.90	63.9-2.30
No. Reflections						
(Rwork / Rfree)	-	77670/4079	22140/1173	22772/1211	23815/1203	36804/1943
Rwork / Rfree (%)[d]	-	20.9/24.3	21.9/27.9	19.5/25.2	21.4/25.9	21.0/25.9
No. atoms	-	2912	5160	2842	6274	5350
Protein	-	591	5104	2651	5386	5239
DNA	-	-	-	-	855	0
Ligands/water	-	321	56	191	33	111
B-factors	-					
Wilson plot		18.4	73.1	30.1	54.5	41.1
All atoms		18.7	69.4	26.4	47.4	61.1
Protein	-	16.8	69.6	25.1	44.3	61.3
DNA	-	-	-	-	67.81	-
Ligands/water	-	35.5	58.8	44.8	27.7	49.3
RMSd from target values	-					
Bond lengths (Å)	-	0.014	0.011	0.013	0.011	0.015
Bond angles (°)	-	1.57	1.48	1.63	1.50	1.64
MolProbity scores						
Overall score (%ile)	-	1.6 (71%)	2.4 (91%)	2.1 (77%)	2.2 (98%)	2.0 (89%)
All-atom clashscore (%ile)	-	6.9 (82%)	7.3 (99%)	6.5 (97%)	5.6 (100%)	6.5 (98%)

Bad rotamers (%)	-	0	28 (4.9%)	18 (6.2%)	41 (6.8%)	18 (3 %)
Ramachandran outlier (%)	-	2 (0.6%)	3 (0.5%)	2(0.6%)	0	2 (0.3%)
Ramachandran favoured (%)	-	299 (96%)	573 (93.2%)	323 (97%)	631 (95.6%)	608 (95.6%)
PDB code		**4LDV**	**4LDW**	**4LDU**	**4LDX**	**4LDY**

[a] $Rsym = \Sigma_i | \hat{I}_h - I_{h,i} | / \Sigma_h \Sigma_i I_{h,i}$, where $\hat{I}_h = (1/n_h) \Sigma i \, I_{h,i}$ and nh is the number of times a reflection is measured.

[b] $Rmeas = [\Sigma_h (n_h/[n_{h-1}])\frac{1}{2} \Sigma_i | \hat{I}_h - I_{h,i} |] / \Sigma_h \Sigma_i I_{h,i}$, where $\hat{I}_h = (1/n_h) \Sigma_i I_{h,i}$ and n_h is the number of times a reflection is measured.

[c] Number in brackets represent highest resolution shell

[d] $Rcryst = \Sigma_{hkl} [|F_{obs}| - k |F_{calc}|] / \Sigma_{hkl} |F_{obs}|$ and $R_{free} = \Sigma_{hkl \, T} [|F_{obs}| - k |Fc_{alc}|] / \Sigma_{hkl \, T} |F_{obs}|$ where T represents the test set excluded during refinement.

Table S2, related to Figure 3, 4 and 6: Affinity parameters of Surface Plasmon Resonance assays.

Figure	Protein	DNA	KD in Fig (nM)	Average KD (nM)	SD	N
3E	ARF1	ER7	147	435	374	7
3E	ARF1	ER7(m)	1090	1620	1126	4
3E	ARF1	ER7(2m)	>5 µM	>5 µM	N/A	2
3F	ARF5	ER7	118	122	51	17
3F	ARF5	ER7(m)	868	708	226	2
3F	ARF5	ER7(2m)	>5 µM	>5 µM	N/A	2
3G	ARF5(S269N)	ER7	160	350	269	2
3G	ARF5(G279E)	ER7	1490	1347	113	3
3G	ARF5(A282N)	ER7	970	2230	1786	4
4C	ARF5	ER7	109	122	51	17
4C	ARF5(H170A)	ER7	1760	1920	226	2
4C	ARF5(R215A)	ER7	681	810	478	3
6B	ARF1	ER5	>5 µM	>5 µM	N/A	2
6B	ARF1	ER6	>5 µM	>5 µM	N/A	2
6B	ARF1	ER7	930	435	374	7
6B	ARF1	ER8	320	320	N/A	1
6B	ARF1	ER9	>5 µM	>5 µM	N/A	2
6C	ARF5	ER5	727	666	87	4
6C	ARF5	ER6	2900	1976	1209	3
6C	ARF5	ER7	114	122	51	17
6C	ARF5	ER8	71	80	13	3
6C	ARF5	ER9	389	369	29	4

Figure: Figure panel representing the experiment
KD in Fig (nM): Apparent equilibrium dissociation constant in experiment selected for figure
Average KD (nM): Average apparent equilibrium dissociation constant based on N experiments
SD: Standard Deviation (nM)
N: number of repeats

Table S3: Primers used in this study

Site-directed Mutagenesis and FRET		
AF-009	CATAGCATGCATATTGCGGTCCTTGCAACAGC	ARF1 G245A
AF-72	GCTGTTGCAAGGACCGCAATATGCATGCTATG	ARF1 G245A
AF-010	GATAGTATGCACATCGCGGTTCTTGCTGCTG	ARF5 G279A
AF-019	CAGCAGCAAGAACCGCGATGTGCATACTATC	ARF5 G279A
AF-020	GTAGAGCTGCAGAGAAGCTATTTCCACCATT	ARF5 G279A
AF-026	GGATCCAAATCGCTGATTCCAACAAT	ARF5 G279A
AF-024	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTT-GAAGAC	ARF5 (FRET)
AF-59	AGTATGGAGTTGGGTTCTGAAACAGAAGTCTTAAGATCGT-TAATG	ARF5 (FRET)
AF-36	CTACCGAGGGCAAGCAAAGAGACATCTCCTAAC	ARF5 P218A
AF-37	GTTAGGAGATGTCTCTTTGCTTGCCCTCGGTAG	ARF5 P218A
AF-38	CGAGGGCAACCAAAGGCACATCTCCTAACTACAG	ARF5 R220A
AF-39	CTGTAGTTAGGAGATGTGCCTTTGGTTGCCCTCG	ARF5 R220A
AF-42	CAAACAGCACTTCCTTCAAATGTTCTCTCAGCG	ARF5 S269N
AF-43	CGCTGAGAGAACATTTGAAGGAAGTGCTGTTTG	ARF5 S269N
AF-44	CACATCGGTGTTCTTAATGCTGCTGCTCACGC	ARF5 A282N
AF-45	GCGTGAGCAGCAGCATTAAGAACACCGATGTG	ARF5 A282N
AF-46	GCTGCTGCTGCTCACAATACCGCCAACCGTACT	ARF5 A287N
AF-47	AGTACGGTTGGCGGTATTGTGAGCAGCAGCAGC	ARF5 A287N
AF-48	CTTTTTTGATATTCTATAGTCCAAGAGCTTGTCC	ARF5 A299S
AF-49	GGACAAGCTCTTGGACTATAGAATATCAAAAAAG	ARF5 A299S
AF-50	CGCCATATCTACGCAGGGCAACCAAAG	ARF5 R215A
AF-51	CTTTGGTTGCCCTGCGTAGATATGGCG	ARF5 R215A
AF-52	GACACAAGCACAGCGGGAGGTTTCTCAG	ARF5 H170A
AF-53	CTGAGAAACCTCCCGCTGTGCTTGTGTC	ARF5 H170A
AF-99	GATAGTATGCACATCGATGTTCTTGCTGCTG	ARF5 G279D
AF-100	CAGCAGCAAGAACATCGATGTGCATACTATC	ARF5 G279D
AF-101	GATAGTATGCACATCGAAGTTCTTGCTGCTG	ARF5 G279E
AF-102	CAGCAGCAAGAACTTCGATGTGCATACTATC	ARF5 G279E
AF-103	GATAGTATGCACATCATTGTTCTTGCTGCTG	ARF5 G279I
AF-104	CAGCAGCAAGAACAATGATGTGCATACTATC	ARF5 G279I
AF-115	AGTATGGAGTTGGGTTCGGTTCGGACGCGGGGTGTCG	ARF5 T794 truncation
Protein Expression		
WB017	CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGAATTCT-TACTCCCAATCAATG	
WB018	GCACATTGATTGGGAGTAAGAATTCTCCTTCTTAAAGT-TAAACAAAATTATTT	pTWIN_LIC
WB019	TTTAAGAAGGAGAATTCATGGCAGCTTCCAATCATTC	Arf1Dbd
WB037	ATTGATTGGGAGTAAGAATTCAGGGGCTCAAGTTCCCAAGG	Arf1Dbd

WB023	TTTAAGAAGGAGAATTCATGATGGCTTCATTGTCTTG	Arf5Dbd
WB024	ATTGATTGGGAGTAAGAATTCGGTGTTTCGATATCCCATG	Arf5Dbd
Surface Plasmon Resonance		
WB193	CCGGTAGGTTGTCTCCCAAAGGGAGACAACCGGTAGG	5'-biotin labeled ER7
WB194	CCTACCGGTTGTCTCCCTTTGGGAGACAACCTACCGG	complement ER7
WB195	CCGGTAGGTTATCTCCCAAAGGGAGACAACCGGTAGG	5'-biotin labeled hlfmtER7
WB196	CCTACCGGTTGTCTCCCTTTGGGAGATAACCTACCGG	complement hlfmtER7
WB199	CCGGTAGGTTGTCTCCCAGGGAGACAACCGGTAGG	5'-biotin labeled ER5
WB200	CCTACCGGTTGTCTCCCTGGGAGACAACCTACCGG	complement ER5
WB201	CCGGTAGGTTGTCTCCCAAAAAGGGAGACAACCGGTAGG	5'-biotin labeled ER9
WB202	CCTACCGGTTGTCTCCCTTTTTGGGAGACAACCTACCGG	complement ER9
IM005	CCGGTAGGTTATCTCCCAAAGTGGGATAACCGGTAGG	5'-biotin labeled dblmtER7
IM006	CCTACCGGTTATCTCCCTTTGGGAGATAACCTACCGG	complement dblmtER7
IM013	CCGGtAGG T TGTCTC CCAA GG GAGACA A CCGGtAGG	5'-biotin labeled ER6
IM014	CCGGtAGG T TGTCTC CCAAAAGG GAGACA A CCGGtAGG	5'-biotin labeled ER8
IM015	CCTaCCGG T TGTCTC CCTT GG GAGACA A CCTaCCGG	complement ER6
IM016	CCTaCCGG T TGTCTC CCTTTTGG GAGACA A CCTaCCGG	complement ER8

Supplemental Experimental Procedures

Protein expression and purification

Regions corresponding to the DNA binding domain (DBD) of Arabidopsis ARF1 (At1g59750; residues 1-354) and ARF5 (At1g19850; residues 1-390) were amplified from cDNA clones using Phusion Flash polymerase (Finnzymes), and cloned in an expression vector pTWIN1 (New England Biolabs), modified for Ligation-Independent Cloning (LIC)[152] to generate fusions with Chitin Binding Domain (CBD) and Intein. An adaptor, generated by annealing primers WB017 (5'-ctagaaataattttgtttaactttaagaaggagaattcttactcccaatcaatg-3') and WB018 (5'- gcacattgattgggagtaagaattctccttcttaaagttaaacaaaattattt -3'), was ligated into *Xba*I/*Sap*I-digested pTWIN1vector. ARF1-DBD and ARF5-DBD were LIC-cloned in the *Eco*RI site introduced through the adaptor. All clones were confirmed by sequencing.

ARF-DBD-CBD fusion proteins were expressed in *E. coli* strain Rosetta DE3 (Novagen). For all expect the ARF1-DBD-SeMet protein, three liters of Difco Terrific Broth (BD), supplemented with 100 mg/L ampicillin was inoculated with 30 ml of an overnight culture and after growth to OD_{600} of 0.5-0.7, protein expression was induced by adding 0.3 mM IPTG and switching temperature from 37 °C to 20 °C and continue growth for 20 hours. For ARF1-DBD-SeMet protein, cells were grown over night at 37 °C in LB medium, and diluted twice 1:100 in M9 medium supplemented with amino acids (L-Lys, L-Phe and L-Thr at 100 mg/L; L-Iso, L-Leu and L-Val at 50 mg/L) and 50 mg/L L-Selenomethionine, and grown to OD_{600} of 0.5. Next, expression was induced as above. Cells (appr. 75 gram) were harvested by centrifugation at 6000xg and resuspended in 50 ml extraction buffer (20 mM Tris, 500 mM NaCl, 1 mM EDTA, 0.1% NP-40 and 2 mM $MgCl_2$ pH 7.8, 10 mg of DNAse and 0.2 mM PMSF). Cells were lysed by passing the suspension twice through a French Pressure cell at 10000 psi and cell free extract was generated by centrifugation for one hour at 50000xg. The supernatant was loaded onto a chitin column (New England Biolabs) with 25 ml bed volume, and washed with 10 column volumes wash buffer (20 mM Tris, 500 mM NaCl, pH 7.8) using an AKTA explorer 100 (GE Healthcare). ARF-DBD proteins were eluted by 1 hour incubation with 40 mM DTT in wash buffer. Proteins were concentrated into a final volume of approximately 1 ml using Amicon ultra-15 10K spin filters, and next passed over a Superdex 200PG size-exclusion chromatography column with dimensions 26 mm by 100 cm. Protein was eluted in washing buffer with 1 mM DTT. ARF-DBD proteins typically eluted at approximately 350 ml, and were concentrated using Amicon ultra-15 10K spin filters, and stored until use at -80 °C. Protein concentration was measured at 280 nm (value of 1.2 corresponded to 1 mg/ml) using a Nanodrop 2000C (Thermo scientific). About 20 mg of ARF1-DBD and 10 mg of ARF5-DBD was obtained from a typical isolation, and purity was typically >95% based on Coomassie staining after SDS-PAGE.

Crystal growth and data collection.

ARF1-DBD-SeMet. Crystals were grown from a 1+1 µl drop of, respectively, 7 mg/mL protein solution and crystallization buffer (2.0-2.3 M NaFormate, 100 mM MES 6.5, 50

mM KI, 10 mM GSH/GSSG). Prism-shaped crystals grew to their maximum size after 2 days and were immediately frozen in crystallization buffer supplemented with 10% glycerol. Data were collected at beamline ID14-4 (ESRF) using radiation of near-Se edge energy.

Native ARF1-DBD (A1D). Crystals were grown from a 0.75+0.75 μl mixture of, respectively, a 4mg/mL protein solution and crystallization buffer (2.9 M NaFormate, 0.1 M MES 6.0, 50 mM KI). Crystals grew to maximum size within one day and were immediately frozen using crystallization buffer supplemented with 10% glycerol and diffraction experiments were performed at beamline ID14-4 (ESRF).

Native ARF1-DBD, monoclinic crystals (A1D-P). Crystals were grown from a 1+1 μl drop of, respectively, 4 mg/mL protein solution and crystallization buffer (100 mM MES 6.5, 20%PEG 5K MME), in the presence of dsDNA of sequence 5'-d(TGTCTC)-3'. Prism-shaped crystals grew to their maximum size in 3 days and were immediately frozen in crystallization buffer enriched with PEG5KMME (to 25%) and supplemented with 10% glycerol. X-ray diffraction experiments were performed at beamline ID14-4 (ESRF, Grenoble). DNA was not observed in the final structure.

Native ARF5-DBD (A5D). Crystals were grown from a 100+100 nl drop of, respectively, 13 mg/mL protein solution and crystallization buffer (20 mM TRIS 7.5, 12.5% PEG MME 2000). Block-shaped crystals appeared after 2-3 days and reached their maximum size after a week. They were subsequently frozen in crystallization buffer supplemented with 20% glycerol. Diffraction experiments were performed at beamline ID14-1 (ESRF, Grenoble).

ARF1-DBD complexed with ER7 (A1D-ER7). The oligonucleotide 5'-d(TTGTCTCCCTTTGGGAGACAA)-3' was annealed with an equimolar amount of the complementary strand. A 1:1 mixture of ARF1DBD and ER7, respectively was prepared with final concentration of ARF1DBD of 4 mg/mL. Crystals were grown from a 100 nL+100 nL drop of, respectively, complex and crystallization buffer (10% PEG 20K, Glycine pH 8.5, 7.5% propanediol). The needle-shaped crystals appeared after 5 days and reached maximum size after 15 days. The crystals were frozen in crystallization buffer supplemented with 15% glycerol. X-ray diffraction experiments were performed at beamline ID14-1 (ESRF, Grenoble).

ARF1-DBD G245A mutation (A1D-G245A). Crystals were grown from a 0.75+0.75 μl drop of, respectively, 13 mg/mL protein solution and crystallization buffer (100 mM bis-tris-propane pH 7.0, 0.7 M succinic acid). Block-shaped crystals appeared after 1 day and grew to their maximum size in 2 days and were immediately frozen using in crystallization buffer supplemented with 20% glycerol. X-ray diffraction experiments were performed at beamline XALOC (ALBA, Cerdanyola del Vallès, Spain).

Data Analysis

All data were processed using iMosflm[153] and SCALA[154]. The initial structure was solved using the hkl2map interface[155] to SHELXC/D/E (Sheldrick, 2002) using SAD data measured at the Se K absorption edge on SeMet ARF1-DBD. Refinement was performed

using REFMAC v. 5.7.0032 [156] of the CCP4 program suite[157]. Visualization and manual adjustments were done using Coot v. 0.7 (Emsley et al., 2010). The refinement statistics and model validation parameters of the crystal structures are given in Table S1. Figures were prepared using Pymol[158] and the ESPript Server[159]. The ClustalW online server was used for sequence alignments[160].

SAXS measurements.

A volume of 25 μl of ARF1-DBD, purified as described above, was diluted with 50 μl of buffer BS (20 mM TRIS 7.5, 0.5 M NaCl, 1 mM EDTA and 10 mM DTT) and immediately concentrated to a final concentration of 3.2 mg/mL over a 10 kDa centricon (Sigma-Aldrich). SAXS data were immediately collected at beamline BM29 (ESRF, Grenoble), using the concentrators flow-through as buffer. BSA references were used for calculating the molecular mass of ARF1DBD. Measurements were carried out at 293 K, within a momentum transfer range of 0.01 Å^{-1} < s < 0.45 Å^{-1}, where s = $4\pi\sin(\theta)/\lambda$ and 2θ is the scattering angle. No measurable radiation damage was detected on comparison of three successive time frames of 10 second exposure. Buffer subtraction and extrapolation to infinite dilution, as well as evaluation of the Guinier region and calculation of the Porod volume were performed by the program PRIMUS[161] in accordance with standard procedures. Calculation of the theoretical scattering curves of monomeric and dimeric ARF1 against the scattering data was performed using CRYSOL[151].

Surface plasmon resonance

SPR measurements were performed on a Biacore 3000 platform. Labeled oligonucleotides (1 μM 5'-biotin; Eurogentec) were made double stranded by incubation with 1.2 μM non labeled complement at 99 °C and slowly cooled before being immobilized on an SA chip (GE Healthcare) to 100 to 120 response units. Proteins were diluted in SPR buffer (20 mM HEPES, 150 mM NaCl, 1 mM EDTA, 0.01% Tween20 and 5 mM β-mercaptoethanol, pH 7.4) in eight two-fold dilution steps (800nM, 400 nM, 200 nM, 100 nM, 50 nM, 25 nM, 12.5 nM, 6.25 nM). Flow during the measurement of the sensorgram was 40 μl per minute. Samples were injected starting with a buffer injection and then the lowest (6.25 nM) concentration. After each injection the chip was regenerated by injection of 40 μl 0.05% (w/v) SDS. Flow channel 1 was always underivatized and used as a reference blank in analysis. Data were analyzed with Scrubber2-T200 (BioLogic Software Pty Ltd) with normalization to zero over 10 to 30 seconds before injection, cropped to remove SDS wash and aligned at injection start. The signal from the closest buffer injection was subtracted. Equilibrium binding constants (K_D) were derived by plotting the average RU values of the last 10 seconds of protein injection prior to washing, relative to protein concentrations (Table S2).

Protein-binding microarrays

PBM11[162] was performed on ARF1-DBD, ARF5-DBD and their mutated versions H170A, G171A, R215A, P218A, S269N and G279E. The synthesis *in situ* of double-stranded DNA and processing of slides were as described [136] but omitting the blocking steps. The

binding mixture volume was adjusted to 170 μl and contained 2% milk, 0.5 μg of denatured herring sperm DNA and 135 μl of the protein extract supernatant. This was obtained by dissolving the pellet of a 25 ml bacterial culture (as described above) in 1.5 ml of AP2 binding buffer 4X [10 mm 2-amino-2-(hydroxymethyl)-1,3-propanediol (TRIS)-HCl pH 8, 60 mm KCl, 4 mm MgCl2, 0.1 mm EDTA, 10% glycerol, 200 mg ml–1 BSA, 0.2% nonyl phenoxypolyethoxylethanol(NP40)] with 1mM PMSF. The pellets were dissolved by vortexing then sonicated and centrifuged for 15 minutes at 14000xg, after which the supernatant was used. The binding mixture was deposited onto the slide, covered with a LifterSlip (22 × 65 mm; Erie Scientific, http://www.eriescientific.com/) and incubated in a humid chamber for 2.5 h at room temperature. Slides were washed three times in 50 ml of phosphate-buffered saline (PBS)-1% Tween 20 (5 min), 3× in PBS-0.01% Triton X-100 (5 min) and dried. DNA–protein complexes were incubated with 20 μg of rabbit polyclonal serum against Chitin Binding Domain (MBL, https://ruo.mbl.co.jp) in PBS–2% milk for 16 h at room temperature. Slides were washed 3× in PBS-0.05% Tween 20, 3× in PBS-0.01% Triton X-100 (5 min each wash) and dried. Labeling of DNA–protein complexes was performed by incubating the microarrays with 0.4 μg of goat anti-rabbit IgG (H+L) DyLight 550 conjugated (ThermoFisher Scientific) in PBS–2% milk for 3 h at room temperature, followed by the same washes as before and the slides were then dried for scanning. Two different images were obtained for each microarray at DNA Microarray scanner (Agilent Technologies) and quantified in the Feature Extraction software (Agilent Technologies). One corresponded to double-stranded DNA at 635 nm. The second image was obtained after labeling of DNA–protein complexes at 532 nm. The combination of the files, normalization and adjustment of the probe intensities and transformation to a list of scores for all the *k*-mers considered was carried out with the PBM Analysis Suite[162]. The best-scored motif, represented as an energy-based matrix, was converted into a graphical logo with the on-line tool enoLOGOS (http://chianti.ucsd.edu/cgi-bin/enologos/enologos.cgi).

CHAPTER 4

In vitro Requierements for High-Affinity ARF Binding and Auxin-Dependent Gene Expression

Alejandra Freire Rios
[1] Victoria Mironova
[2] Roeland Boer
[a] André Kuhn
Willy van den Berg
Keita Tanaka
Dolf Weijers

1. Novosibirsk State University; Institute of Cytology and Genetics, Novosibirsk, Russia
2. Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10-12, 08028 Barcelona, Spain
a. Present address: John Innes Centre, Department of Crop Genetics. Norwich, Norkfolk, United Kingdom

Abstract

The plant hormone auxin controls numerous growth and developmental processes in most plant species investigated. Specificity for the many responses is likely generated by functional diversification and cell type-specific expression of the DNA-binding ARF transcription factors that are activated by auxin. Based on crystal structures and *in vitro* protein analysis, we recently proposed a "caliper" model in which ARFs recognize target genes as dimers, and where space between two inverted monomer binding sites contributed to different binding affinities among the ARF family. A key question however, is if and how these structural properties translate to biological reality. Here we explore a number of predictions derived from ARF structural properties. First, *in vitro* analysis indicated the highest binding affinity for the novel TGTCGG motif by multiple members of the ARF family, and it was later shown that this element indeed conferred increased auxin-dependent expressio*n in vivo*. Here we define the structural basis for ARF binding preference to the TGTCGG motif above the canonical AuxRE TGTCTC. We show that the same conserved Histidine that interacts with the G opposing the C6 position in TGTCTC can now interact with either the G5 or G6 base in the TGTCGG sequence. In addition the backbone is rotated such that an extra hydrogen bond is made with the C opposing G6. Secondly, dimerization of the ARF DBD predicts high-affinity binding to inverted repeat elements, and we here first used a well-known ARF5 target gene (*TMO5*) to show that a complex, inverted repeat is indeed critical for *TMO5* gene expression. We next used a bioinformatics approach to show that presence of bipartite Direct Repeat5 (DR5) and Inverted Repeat8 (IR8) elements in gene promoters are highly correlated with auxin responsiveness. We next experimentally validated that 75% of genes that harbor DR5 or IR8 element are indeed auxin-regulated, and found that the DR5 element only mediates gene activation through exclusive interaction with class A ARFs. Finally we use novel fluorescent spacing reporter transgenes to test if inter-AuxRE spacing contributes to differences in auxin-dependent regulation *in vivo*, and find evidence for a subtle contribution. In summary, this work demonstrates that several ARF properties predicted from *in vitro* analysis reflect *in vivo* activity, and provides a structure-based explanation for the cis elements involved in genome-wide auxin regulation.

Introduction

Many efforts are invested in understanding how (plant) transcription factors can specifically recognize their target elements [136,163–165]. This important question is particularly challenging when transcription factors belong to families with closely related members. Members of a family of eukaryotic transcription factors may share affinity for the same short DNA motif, yet regulate different genes [166–168], as seems to be the case for the ARF family in *Arabidopsis* ([126,166]). The first element that confers specificity to DNA recognition by a transcription factor is the sequence of the DNA itself. Eukaryotic transcription factors recognize small specific DNA sequences present in the regulatory regions of the target genes [16]. Specific binding to these sequences depends on the biophysical interactions permitted by the protein structure of the transcription factor [169]. Because eukaryotic transcription factors recognize small fragments of DNA, the chances of finding this fragments along the whole genome are significant, and hence specific DNA binding may need more elements to fine tune it. One way to ensure specific binding is to hide "unwanted" binding sites while making accessible the target genes that need to be regulated according to the developmental needs of the organism. Thus, chromatin structure/DNA accessibility can be considered as another element in specific DNA binding. The next element to further fine tune specific DNA binding is the ability of a transcription factor to interact and cooperate with other proteins or other transcription factors. In this way the number of cis regulatory elements needed for specific binding increases. In the case that a transcription factor forms heterotypic interactions with a protein from a different family, not only the number of recognition sequences matters, but also the combination of different recognition sequences. Whether the interaction of the transcription factor is homo- or heterotypic, the spatial conformation of the complex will demand that the cis regulatory elements be at accessible positions for all the players involved/needed for the regulation of the target gene(s).

Plant transcription factors have been classified in approximately 30 families based on the structural properties of their DNA binding domains. One of these is the Auxin Response Factor (ARF) family, which has 23 members in *Arabidopsis thaliana*. These proteins contain B3 DNA binding domain and are therefore part of the B3 superfamily of plant transcription factors. ARFs are responsible for regulating many developmental processes during the lifespan of plants. As the name indicates they regulate transcription of genes in response to the plant hormone auxin (indol-3-acetic acid), a small molecule that triggers a well-studied cascade of events that ultimately leads to gene regulation by the ARFs. Incorrect response to auxin by ARFs can mean an early death for a plant, as is the case for the *arf5* mutants, which cannot form a root and will die shortly after germination [46]. Loss of ARF activity can also mean the loss of new generations since mutations in other ARFs will lead to severe flower defects [48,170]. ARF proteins bind cis-regulatory elements composed of 6 base pairs and are called Auxin Response Elements (AuxREs). The AuxRE 6mer has the form of TGTCNN. In the 1990's, TGTCTC was described as the recognition sequence for ARFs [30]. TGTCTC was found in the promoter region of the soybean GH3 auxin responsive gene, and was next used as bait to isolate the first members of the ARF family. This element has been considered ever since as the canonical AuxRE. However, *in vitro* evidence reported in Chapter 3 showed

that *Arabidopsis* ARF1 and ARF5 have the highest binding affinity not for TGTCTC, but for TGTC**GG** [126]. This result was obtained using protein binding microarray (PBM)[136]. The design of the PBM used allowed to expose the DBDs to every possible 6-mer and in every case TGTCGG was the preferred binding sequence. Liao *et al* later reported *in vivo* evidence that supports the higher affinity of ARFs for the TGTCGG sequence. They redesigned the extensively used DR5 auxin activity reporter by replacing the TGTCTC elements by TGTCGG ones. By doing this, the sensitivity to auxin treatment increased more than 10-fold, and as a consequence showed auxin activity where the original DR5 could not detect it. Indeed, bioinformatic analysis showed that, among all TGTCNN sites, TGTCGG is the motif most strongly associated with auxin response [171]. Furthermore, using a DNA Affinity Purification (DAPseq) method, it was shown that this element is the preferred binding site when ARFs are exposed to fragments of *Arabidopsis* genomic DNA[172]. While there are multiple lines of evidence suggesting the high affinity of ARFs for the TGTCGG, 6-mer, a critical and unanswered question is whether this higher binding affinity can be explained by the biophysical properties of ARF DBDs. In other words, why do ARFs prefer TGTCGG over any other TGTCNN element?

Similar to protein-DNA interactions at the individual base or amino acid level, the topology of binding sites at larger space scale follows from structural properties of the DNA-binding proteins. Initial structural characterization of the ARF-DNA interaction revealed that these proteins can dimerize through their DNA-binding domain (DBD). This DBD dimerization allows high-affinity binding to two AuxRE's that are oriented in an inverted repeat, with a spacing dictated by the structural constraints of the ARF protein (Chapter 3)[126]. We indeed showed that mutations in the DBD that compromise dimerization of the protein fail to properly rescue the *arf5* mutant, and impair high-affinity DNA binding *in vitro*. Thus, it appears that dimerization of the DBD is indeed a requirement for ARF function. Interestingly, a natural variant in the *Brassica rapa* that affects seed size was shown to be caused by a mutation that affected ARF18 DBD dimerization [173], which suggests that this property may be generally important for ARF function.

In addition to the DBD dimerization which would facilitate high-affinity binding to inverted repeat AuxRE's, it was shown that the C-terminal PB1 domain found in most ARFs can mediate head-to-tail multimerization [174]. Thus, ARFs have at least 2 different modes of higher-order interactions, each of which would be expected to impact upon the mode of DNA binding. A critical question thus becomes what topology of complex AuxRE's is associated with auxin response, or in other words, what is the dominant mechanism for auxin regulation across the genome. In this chapter, we address the *in vivo* importance of complex AuxRE elements to determine the degree by which *in vitro* ARF structures report biologically meaningful regulatory properties. We thereby validate a number of predictions made from the ARF structures and provide structural insight into auxin-dependent gene regulation.

Results

Structural basis for high affinity ARF-DNA binding

In order to understand DNA binding specificity by ARF members, the first step was to define their affinity for particular DNA sequences. In Chapter 3 we approach this issue using Protein Binding Microarray (PBM) for two representative members of the ARF family[126]. Both ARFs did show affinity for the canonical TGTCTC AuxRE but they show the highest affinity for the TGTCGG sequence. In order to understand the biophysical causes for this significant higher binding affinity, we crystalized the ARF1 DNA binding domain in complex with the TGTCGG-IR7 (inverted repeat) double stranded oligonucleotide. The overall structure is highly reminiscent (RMSd 0.75Å for all protein atoms of the dimer) of the structure of the ARF1 DBD in complex with the TGTCTC-IR7 double stranded oligonucleotide previously published [126]. The improved resolution of the A1D-GG structure presented here, in which the TGTCTC sequences were mutated to TGTCGG, allowed for the tracing of additional loops for which the density was not clear in the original IR7 model, in particular of base domain residues 230-232, which sit at the dimer interface, and 301-305, which interact with the B3 domain. However, the most interesting changes occur near the GG bases and their respective pairs, which interact with H136 and G137 of the protein. On both binding sites, constituted by the two protein monomers and the two inverted repeats, the H136 sidechain turns inwards into the major groove, thus placing the Nε2 of the imidazole ring at hydrogen bond distance to the O6 atoms of the two guanidine bases. An additional hydrogen bond is formed between the carbonyl oxygen atom of H136 with the cytosine base complementary to the second G in the TGTCGG sequence (Figure 1). This hydrogen bond, which is not possible in the TGTCTC-IR7 structure, is formed through a peptide flip, which occurs as the result of the peptide bond flexibility between histidine 136 and glycine 137. The inward movement of H136 is hindered in the TGTCTC-IR7 structure due to the presence of the thymine at position 5 of the TGTCTC AuxRE and by the bulky complementary guanidine of the cytosine 6. The formation of the additional hydrogenbond provides a stronger binding and therefore results in higher affinity between ARF and TGTCGG than TGTCTC.

An inverted AuxRE repeat motif mediates TMO5 expression

The ARF DBD structure shows that ARFs can bind an inverted AuxRE motif in which the individual motifs are spaced by optimally 7 or 8 bases. However, it remains a question if such motifs mediate ARF function *in vivo*. We focused on a well-known ARF target gene to addres this question. The ARF5/MP target gene *TARGET OF MONOPTEROS5* (*TMO5*) is involved in vascular development in *Arabidopsis* and its expression pattern has been well characterized [107]. It has been previously shown that a 2.3 kb fragment upstream of the *TMO5* start codon is sufficient for *TMO5* expression [107]. This region contains 12 TGTCNN sites, of which 7 are in the same orientation as the gene and 5 are in reverse orientation (Figure 2A). From them only one inverted repeat with a binding permissive 7 nucleotides spacing between
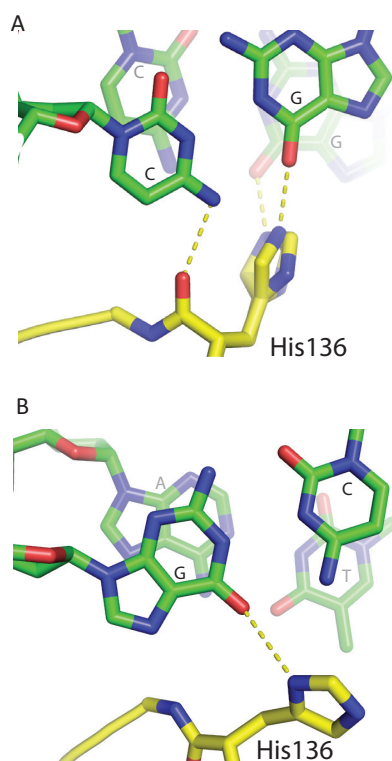
Figure 1. Atomic basis for high-affinity ARF-DNA interactions
(A) ARF1 Histidine 136 (yellow) is able to form 2 Hydrogen bonds with the TGTCGG sequence: 1 with the nitrogen atom of either G5 or G6 (a superposition of both options is shown in the figure) and 1 with the oxygen atom of the opposing C6 (green).
(B) In contrast, the same His136 can make only a single Hydrogen bond with the nitrogen atom of the opposing G6 in the complex between ARF1 DBD and the TGTCTC-containing ER7 oligonucleotide.

the AuxREs is formed 1589 bps upstream of the start codon (Figure 2A and Supplemental Figure 1). Based on data derived from the structure, our hypothesis is that from all the AuxRE sites present in the promoter of *TMO5*, this bipartite element might have the highest binding affinity with ARFs, and therefore contribute significantly in TMO5 regulation. We mutated (replaced by TTTTTT) either a single AuxRE or both, in the context of the 2,3 kb promoter, and fused these to TMO5-3xGFP. Next, we analyzed fluorescence pattern and quantified intensity in transgenic roots. The independent AuxRE sites of this bipartite element could act in 3 possible ways. First, they might act redundantly. In this case only one of them would be sufficient for proper ARF binding and TMO5 regulation and the loss of one would have no effect in TMO5 expression. Secondly, the elements may act additively. In this case both sites would contribute equally and independently to TMO5 regulation and the loss of one of them would reduce the expression of TMO5 by 50%. Finally, the sites may act cooperatively. In this case both sites would be necessary for proper ARF binding and the loss of one of them would strongly reduce TMO5 expression (more than 50%).

The intensity of the GFP signal was measured in root tips for multiple independent lines of each construct (Figure 2B, C and D). The first striking observation is that complete elimination of the bipartite element reduced the amount of fusion protein to an almost
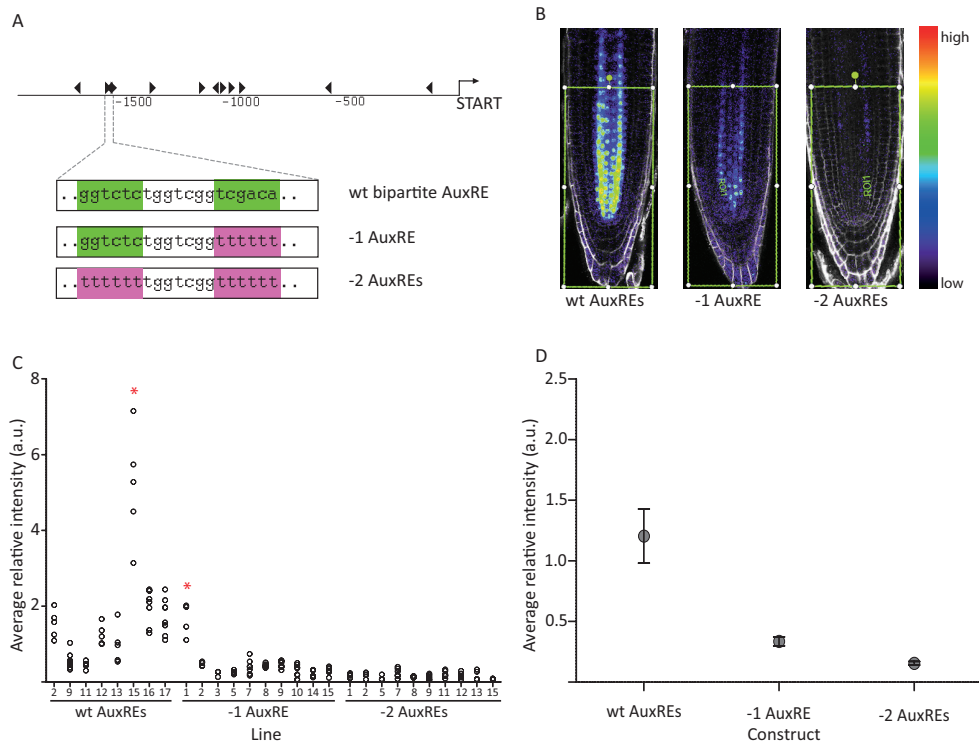
Figure 2. Requirement of in inverted AuxRE element for *TMO5* gene expression

(A) All TGTCNN-like elements found in the 2.3kb promoter of *TMO5*. > indicates direct AuxREs and < indicates reverse AuxREs. The region of the promoter containing the only bipartite element is magnified to show the sites and mutant versions used.

(B) Expression of the TMO5-3xGFP fusion protein driven by the three different versions of the *TMO5* promoter used. A root tip of one representative line per construct is shown. The green box is the region where GFP intensity was measured. Fluorescence intensity is depicted on a false-color scale.

(C) Average intensities of TMO5-3xGFP fusion protein fluorescence in root tips of seedlings carrying the wild-type *TMO5* promoter, or versions in which either 1 or 2 AuxRE sites were mutated. Each point represents the average pixel intensity value per root tip in the area indicated by a green box in (B). All root images were taken under the same microscope settings. The number of roots analyzed per line were: WT #2 (5), #9 (8), #11 (5), #12 (5), #13 (6), #15 (5), #16 (8), #17 (9); -1 AuxRE #2 (8), #3 (3), #5 (6), #7 (7), #8 (6), #9 (7), #10 (9), #14 (6), #15 (5); -2 AuxREs #1 (5), #2 (4), #5 (4), #7 (6), #8 (5), #9 (6), #11 (6), #12 (6), #13 (4), #15 (5). The values obtained for wt line #15 and -1AuxRE line #1 (marked with a red asterisk) were considered outliers and not included in further analysis.

(D) Average fluorescence intensities of multiple independent transgenic lines per construct (7 lines for non-mutated construct, 9 lines for -1AuxRE construct and 10 lines for -2AuxRE construct). Major statistical outliers were not included and error bars show the standard error of the mean (SEM).

undetectable level. The drop in intensity was highly significant compared to the wild type reporter (p-value 0.0002). This means that indeed, from all the AuxREs, the bipartite element seems to be the main binding site for ARFs and therefore responsible for correct levels of

protein expression. When the reverse AuxRE of the repeat was mutated, the expression of the fusion protein also dropped significantly (p-value 0.0007). The amount of protein detected was less than half of the observed in the wild type construct, implying that the two inverted AuxRE elements act cooperatively to control *TMO5* gene expression. Given that these are canonical ARF binding sites, that *TMO5* is a direct ARF target and regulated by auxin, and consistent with the structure of the ARF dimers, we conclude that ARF5 not only binds as dimer *in vivo*, but also dimerizes to cooperatively regulate expression of *TMO5*.

Genome-wide correlation of bipartite AuxREs with auxin responsive expression

The occurrence and relevance of an inverted AuxRE repeat, spaced by 7 nucleotides, in the *TMO5* promoter confirms predictions from the structure. However, this may be a unique case, and it is important to determine if this binding mode is representative for ARF function. If ARFs indeed bind to DNA as a dimer, one would expect an enrichment of AuxRE repeats in the promoters of auxin-regulated genes. Here, we have bioinformatically searched all possible TGTCNN bipartite AuxREs: direct (DR), everted(ER) and inverted (IR) repeats (Figure 3) with spacing between the elements ranging from 0 to 15 bp in the -1500 regions (relative to the first start codon) of all genes in the *A. thaliana* genome. First, we determined the distribution of all variants in the whole genome and found that they are nearly uniformly distributed (Figure 4). We next tested the association of each of these bipartite elements with auxin-responsive gene expression. We collected publicly available auxin-related microarray data sets (35 in total) and pre-processed these to identify the differentially expressed genes in response to auxin (>1.5 times, FDR adjusted p-value<0.05). For further analysis, we took those microarrays which showed more than 50 differentially up- or down-regulated genes (17 in total, Supplemental table 1). In each of the 17 experiments we tested if the genes having a certain bipartite AuxRE variant are enriched among auxin up- or down-regulated genes (p<0,05 by Fisher's exact test). Figure 4 shows the number of microarrays in which certain bipartite AuxRE were found associated with auxin response. As multiple tests were carried out, the significance level was adjusted using a Bonferroni correction to an alpha level of 0.05 for multiple comparisons (45; 3 types of repeat with 15 variants of the spacer length). Under this correction there was no TGTCNN repeat associated to gene down-regulation. However, three variants were significantly associated to gene up-regulation upon auxin treatments. These include DR1 (TGTCNNnTGTCNN), DR5 (TGTCNNnnnnnTGTCNN) and IR8 (TGTCNNnnnnnnnnNNGACA) (Figure 5A).
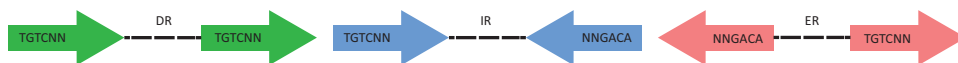

Figure 3. Bipartite AuxRE variants.
Schematic representation of the 3 possible variants of bipartite AuxREs considered in bioinformatics analysis.
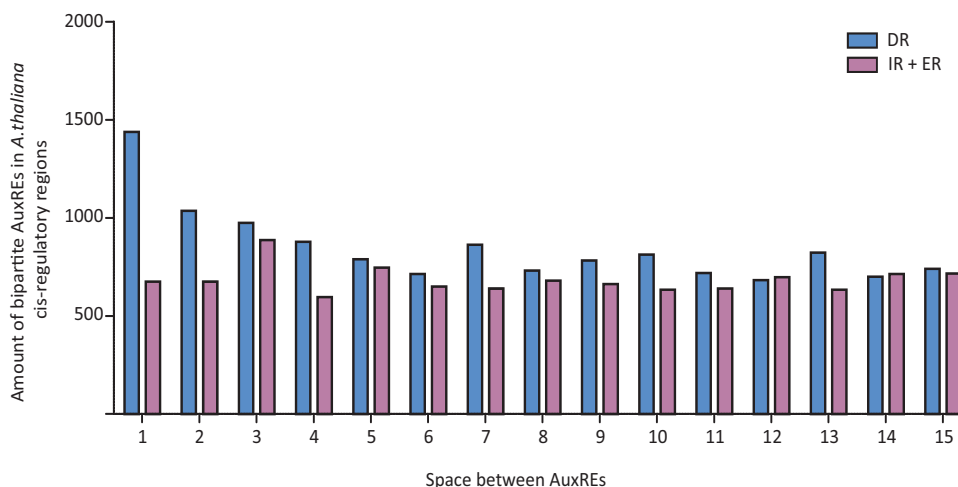
Figure 4. Genome-wide occurrence of AuxRE repeats
Distribution of bipartite AuxRE variants in the 1,500 bp region upstream of all *Arabidopsis thaliana* genes. With the exception DR0, DR1 and ER3 all the other variants are observed uniformly in the -1500 cis regulatory regions of all genes.

Although other bipartite AuxRE variants were not significant under the stringent Bonferroni criterion, their association with auxin response in more than one microarray suggest that they might be also functional. In this context, it should be noted that the transcriptome experiments used for the bioinformatic analysis were highly diverse, and the stringent statistics biases toward very strong and generic auxin regulation. It is likely that for example IR7, whose association with auxin response was detected in two microarrays, is biologically meaningful. Likewise, IR1, DR2, ER13-14 were comparable to IR7 (Figure 5B).

To determine if there is a functional distinction between the response controlled by the two main AuxREs, DR5 and IR8, we used the data of a time-course auxin treatment (GSE42007; DEGs >1,5 times, p-value <0.05) and determined if the two elements associated with early or late genes. The analysis showed that IR8 was significantly associated with primary auxin response; response up to 2 hours after auxin treatment. DR5 instead was associated with gene up-regulation from 1 until 24 hours after auxin treatment (Figure 6). This suggests that the two elements, though both strongly associated with auxin-dependent gene regulation are regulated differently.

The ARF structures presented in Chapter 3 showed ARFs binding as homodimers to an AuxRE inverted repeat with a 7 nucleotide spacer (IR7). This bipartite binding element was described previously and used for ARF binding tests for years [29]. Interestingly, in this bioinformatic analysis, its association with auxin response was less notable than that of IR8. The structures also suggested that the DBD is flexible and rotates to accommodate to the position of the AuxREs within a permissive range and other *in vitro* experiments showed a
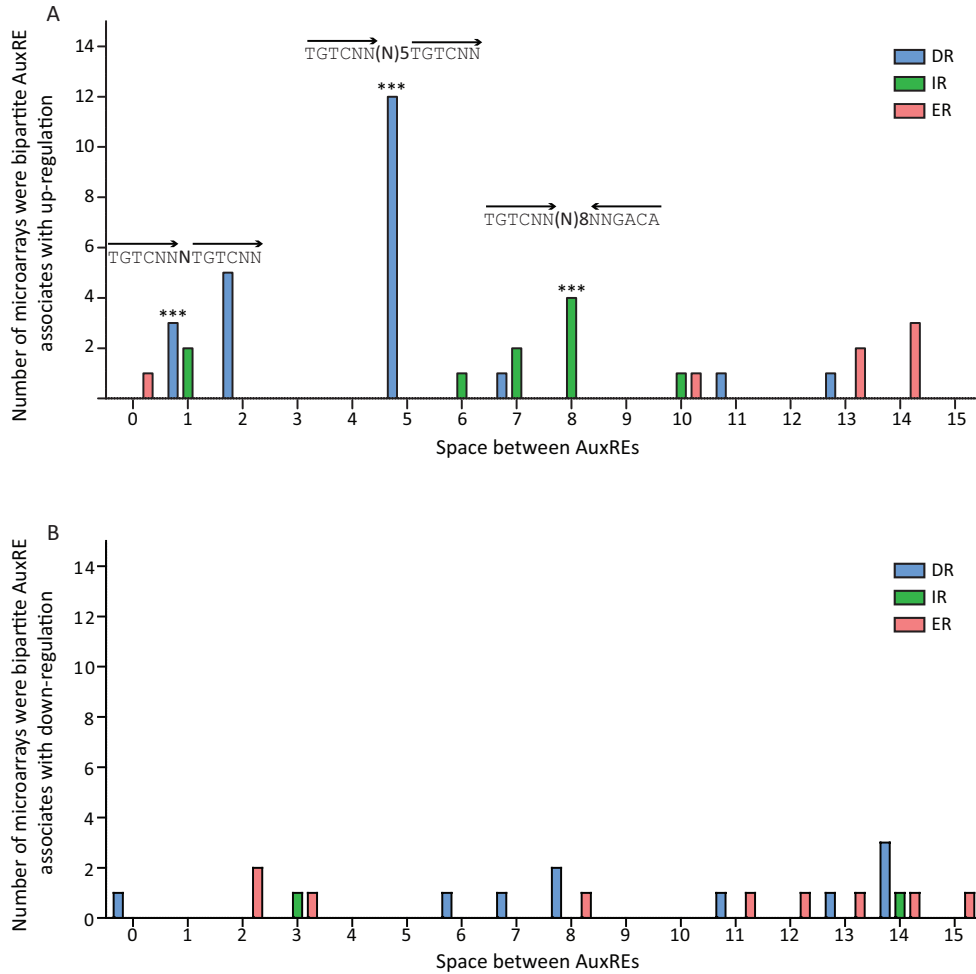
Figure 5. Genome-wide correlation of bipartite AuxRE elements with auxin-regulated gene expression
Association betwwen the presence of direct (DR), inverted (IR) and everted (ER) repeats in the 1,500 bp upstream region of genes with up-regulation (A) or down-regulation (B) in response to auxin. The number of nucleotides between the TGTCNN elements is shown in the x axis. The number of independent transcriptome experiments (out of a total of 17) in which the repeat was enriched in the auxin-regulated gene set (p-value<0.05; Fisher's exact test) is shown on the y axis.
*** The association was found significant under multiple test correction (p-value<0.05/15*3) at least in one microarray.

high affinity of ARFs for the IR8 binding site [126]. Thus, the association of IR8 motifs with auxin response can be explained by the dimerization of ARFs through their DBD.

In contrast, the dimer structure of ARFs DNA binding domains cannot explain binding to a direct repeat. Interestingly, it is precisely this direct repeat with a spacing of 5 nucleotides

Figure 6. Temporal dynamics of DR5- and IR8-containing genes
On the x axis is time of treatment by exogenous auxin (GSE42007 microarray). On the y axis is the fraction of upregulated genes normalized to the average value. The statistical significance by t-test for proportions is shown by asterisks: * p<0.05; ** p<0,01; *** p<0,001

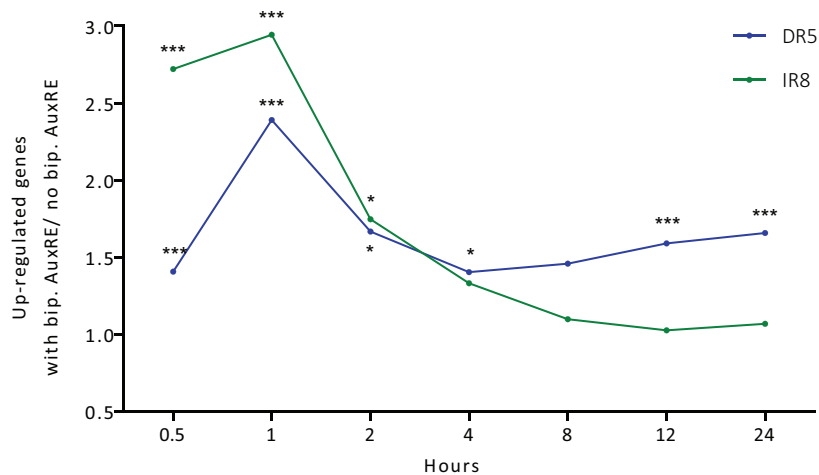(DR5) that was used to generate the widely used auxin activity reporter "DR5" and its improved version "DR5v2" [175]. We asked if binding to the DR5 element may also be based on cooperativity of ARF-DNA interactions. We used Surface Plasmon Resonance to quantify binding of various ARF DBD's to a DR5 oligonucleotide, compared to binding to a single AuxRE (Figure 7). We included two ARFs of each subclass. ARF5 and ARF6 belong to the Class A ("activator") ARFs, while ARF1 and ARF4 belong the Class B ("repressor") ARFs and ARF10 and ARF16 belong to the atypical Class C ARFs. We found that all ARFs showed little binding to a single AuxRE. Strikingly, only the Class A ARFs ARF5 and ARF6 showed increased binding to the DR5 element. Particularly the protein-DNA dissociation was strongly decreased in the ARF5-DR5 and ARF6-DR5 interactions, when compared to all other interactions tested. Thus, while the structural basis remains unclear, the strong association of DR5 elements with auxin-responsive genes is likely based on cooperative binding of the Class A ARF DBD.

Bipartite AuxRE's predict auxin-responsiveness

The genome-wide association of DR5 and IR8 elements with auxin responsive gene expression is based on statistical enrichment. However, it is not clear how predictive the presence of these elements in the regulatory region of a gene is of auxin regulation. To determine predictiveness, we first generated a list of all *A.thaliana* genes that contain at least one of these elements in their promoter regions (-1500 bp). The DR5 motif was found in
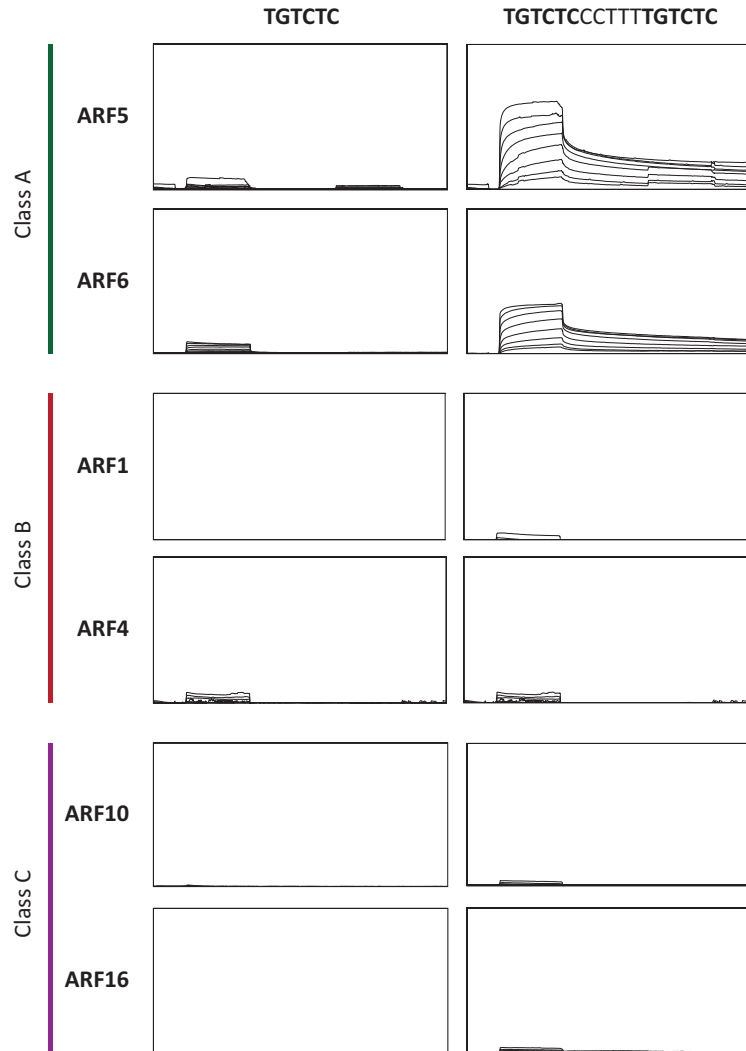
Figure 7. Exclusive cooperative binding of Class A ARFs to a DR5 element
Surface Plasmon Resonance binding profiles of the DNA binding domains of ARF5 and ARF6 (Class A – activators), ARF1 and ARF4 (Class B – repressors) and ARF10 and ARF16 (Class C) proteins to a single AuxRE (left panel) or the bipartite DR5 AuxRE (right panel). Traces represent successive two-fold dilutions of the ARF protein solution (from upper to lower). Values were normalized to the highest value of that same protein on the ER7 oligonucleotide. Scales are identical in all panels.

701 genes and the IR8 motif in 297 genes. We next tested auxin-regulated gene expression in a subset of these genes. Since both lists contained known auxin-regulated genes, at least 3 of those were included as positive controls for each class, the rest were blindly picked. Thus, both selections are semi-random. For the experiment, 5 days old seedlings were treated with

Table 1: Auxin regulation of genes containing a bipartite AuxRE in their promoter region.

qPCR analysis of a random selection of genes containing a bipartite AuxRE in their promoter region was performed on cDNA from roots of 5 day-old seedlings that had been treated with 1 µM 2,4D during 15 minutes, 1 hour, 2 hours or 6 hours. qPCR reactions were performed on 3 technical replicates and the experiment was performed on 2 biological replicates. The results for one of the biological replicates is shown. Expression of all genes was normalized to that of the EEFα4, ACT2 and CDKA1 reference genes, and expression levels were compared to a No-treatment control for each time point to determine the fold-change (Regulation). Statistically significant (p-value<0.05) fold-changes are highlighte Technical replicates were used to calculate the p-value in a Student's t-test. Genes showing up-regulation are in green, down-regulation in red, combined effect in time in orange and not affected in blue. To fall into one of this categories the trend was observed in both replicates.

| Gene ID/Name | # TGTCs | Bipartite element | | | 15 minutes | | 1 hour | | 2 hours | | 6 hours | |
		Variant	Position from start codon	Sequence (5' – 3')	Regulation	p-value	Regulation	p-value	Regulation	p-value	Regulation	p-value
AT4G28640 / IAA11	12	IR8	-1023 to -1004	TGTCTTCCCTCCAGACGACA	3.50	0.0002	4.98	0.0018	12.00	0.0001	6.55	0.0011
AT1G80240 / DGR1	17	IR8	-168 to -149	TGTCTGCGTGGCACCTGACA	1.58	0.0005	2.12	0.0003	6.35	0.0000	5.16	0.0000
AT4G30080 / ARF16	12	IR8	-1131 to -1112	TGTCGGAAGCACCGACGACA	-1.03	0.6182	-1.57	0.0070	1.78	0.0001	2.22	0.0002
AT3G56810	14	IR8	-291 to -272	TGTCTAGCCAAAAAATGACA	1.26	0.0011	-1.17	0.1262	1.68	0.0002	1.56	0.0290
AT1G18400 / BEE1	21	IR8	-1555 to -1536	TGTCATTGTGAAATGGGACA	1.26	0.0143	1.13	0.1260	2.75	0.0000	5.30	0.0000
AT5G67250 / SKIP2	10	IR8	-165 to -146	TGTCGACTACGATTATGACA	1.01	0.8274	-1.31	0.0198	1.18	0.2408	1.322	0.0420
AT1G80280	13	IR8	-1879 to -1860	TGTCATTTAACTCGATGACA	1.18	0.0086	1.11	0.0040	1.06	0.0525	1.01	0.8065
AT5G45260 / RRS1	20	IR8	-1442 to -1423	TGTCGCTTGTTGCAATGACA	1.54	0.0404	1.58	0.0000	2.05	0.0001	-1.07	0.5347
AT2G30990 / DUF688	12	IR8	-975 to -956	TGTCTTAAATTATTTCGACA	-1.12	0.0189	-1.53	0.0050	-1.47	0.0001	-3.25	0.0000
AT3G62980 / TIR1	12	IR8	-336 to -317	TGTCATCATCAGAATCGACA	-1.35	0.0004	-1.65	0.0006	1.21	0.0668	1.34	0.0013
AT1G11320	13	IR8	-897 to -878	TGTCCAATGGGCTAGAGACA	-1.81	0.0034	-3.95	0.0046	-1.03	0.4914	1.23	0.0945
AT4G03110 / BRN1	14	IR8	-806 to -787	TGTCGGCGTTTGAATACGACA	1.00	0.9960	-2.15	0.0000	-1.63	0.0037	-2.36	0.0001
AT1G28200 / FIP1	7	IR8	-618 to -599	TGTCCGAAATGATAAAGACA	-1.11	0.1954	1.44	0.0045	1.05	0.4410	-1.64	0.0060
AT3G59420 / ACR4	15	IR8	-54 to -35	TGTCTCAGAAAGTGAAGACA	-1.68	0.0044	-1.81	0.0004	1.10	0.2687	1.73	0.0021
AT4G20400 / JMJ14	10	IR8	-1826 to -1807	TGTCACTAAACAAAGAGACA	-1.04	0.5389	-1.21	0.0467	1.00	0.8511	1.30	0.0104
AT1G71080	14	IR8	-660 to -641	TGTCAATGAAGAAGAAGACA	-1.28	0.0005	-1.35	0.0149	1.13	0.0144	1.35	0.0071
AT3G02875 / ILR1	14	IR8	-55 to -36	TGTCCTTTTCTTTCATTGACA	-1.08	0.0404	-1.01	0.7495	1.11	0.1695	-1.21	0.0883
AT2G42590 / GRF9	18	IR8	-281 to -203	TGTCATGATATGCGTTAGACA	-1.13	0.0671	1.15	0.0446	-1.02	0.5386	-1.36	0.0033
AT1G65380 / CLV2	12	IR8	-1327 to -1308	TGTCTTTCTTGTTAAAGACA	-1.43	0.0110	1.06	0.5458	1.54	0.0056	1.01	0.8795

Gene	n	Element	Position	Sequence								
AT1G19220 / ARF19	11	DR5	-419 to -403	CCGACAAGAAGAAGACA	3.29	0.0002	5.24	0.0000	5.32	0.0000	5.29	0.0001
AT3G23030 / IAA2	13	DR5	-240 to -224	TGTCCCACTTTTGTCCC	4.75	0.0000	6.99	0.0002	10.95	0.0000	4.60	0.0000
AT2G26710 / BAS1	12	DR5	-1132 to -1116	TGTCACTATTTTGTCAA	1.19	0.0632	1.63	0.0034	1.66	0.0085	2.75	0.0001
AT4G60450 / ARF4	14	DR5	-476 to -460	TGTCTTGGCTTTGTCTC	1.41	0.5619	1.87	0.0009	4.16	0.0000	3.68	0.0008
AT1G14350 / FLP	15	DR5	-72 to -56	GAGACAGAAAGAGGACA	1.18	0.0437	1.52	0.0004	2.47	0.0001	9.85	0.0005
AT5G24590 / TTP	12	DR5	-920 to 904	AAGACATAAGCTCGACA	2.08	0.0011	-1.28	0.0611	1.66		1.15	0.0565
AT3G50060 / MYB77	17	DR5	-847 to -831	TGTCAGTCTAATGTCGT	4.53	0.0001	-1.02	0.7783	2.86	0.0019	1.25	0.0222
AT3G55610 / P5CS2	13	DR5	-325 to -314	AAGACAAAATGCAGACA	1.47	0.0026	1.48	0.0022	1.61	0.0041	1.47	0.0059
AT3G54000	11	DR5	-1264 to -1248	TAGACATGATATTGACA	1.24	0.0092	1.84	0.0012	4.09	0.0008	13.35	0.0000
AT4G39100 / SHL1	11	DR5	-353 to -337	AAGACAAAGTCAGGACA	1.08	0.3698	1.87	0.0003	1.15	0.0723	1.13	0.3714
AT1G07080	15	DR5	-774 to -758	TGTCAGAATGGTGTCCT	1.14	0.0071	1.89	0.0026	1.21	0.0767	-1.09	0.3784
AT3G47810 / MAG1	13	DR5	-315 to -299	TGTCTACTTTGTGTCAA	1.19	0.0300	1.87	0.005	1.17		-1.25	0.0646
AT1G17330	16	DR5	-1446 to -1430	TGTCTTCCTCGTGTCGT	-1.03	0.6424	1.21	0.0002	1.06	0.3420	1.27	0.0594
AT1G15670 / KMD2	10	DR5	-1089 to -1073	TGTCATTTTCGTGTCAG	-1.18	0.0013	-1.13	0.0015	1.25	0.0137	1.30	0.0244
AT5G07100 / WRKY26	20	DR5	-1355 to -1339	TGTCTTCCTCTTGTCCT	1.34	0.0183	1.09	0.2183	1.09	0.2039	-1.44	0.0922
AT1G04140	13	DR5	-1223 to -1207	TTGACACATTTCCGACA	1.12	0.0003	1.36	0.0028	1.07	0.002	-1.12	0.2015
AT4G00760 / APRR8	25	DR5	-37 to -21	TAGACAGCGATGCGACA	1.00	0.9275	1.29	0.0600	1.87	0.2175	-1.10	0.1401
AT2G43710 / SSI2	18	DR5	-825 to -809	TGTCCCCTCTCTGTCTC	1.16	0.1055	1.23	0.0288	1.10	0.271	-1.33	0.0549
AT2G44150 / ASHH3	15	DR5	-457 to -441	ATGACATATAAACGACA	-1.33	0.0360	1.09	0.1212	-1.13	0.0674	-1.19	0.0153
AT3G26760	18	DR5 IR8	-1307 to -1291 / -1105 to -1086	TGTCTAAAGTCTGTCTC / TGTCTCTAGAAATGTAGACA	1.41	0.0018	6.97	0.0000	8.02	0.0000	4.65	0.0000
AT4G29910 / ORC5	18	DR5 IR8	-950 to -934 / -34 to -15	TGTCAAACCTTTGTCCT / TGTCTTGTTTCATTTTGACA	-1.29	0.2832	-1.37	0.0194	1.33	0.0381	1.26	0.00515
AT2G37390 / NAKR2	19	DR5 IR8 IR8	-1613 to -1597 / -1602 to-1583 / -688 to -669	TGTCTTCGCCTTGTCTC / TGTCTCGCAATCTCGTGGACA / TGTCTCGTCTCTCTGTCGACA	-1.30	0.1895	-2.84	0.0001	-1.25	0.0193	-2.60	0.0010
AT1G69310 / WRKY57	12	DR5 IR8	-1297 to -1281 / -968 to -949	TAGACATCACGTAGACA / TGTCATATCCAACACGGACA	-1.52	0.0005	1.02	0.8187	-1.34	0.4410	-2.37	0.0000

auxin and roots were taken for RNA extraction after different incubation times (Table 1 and Supplemental Figure 2 and 3). We used qPCR to determine transcript levels for each gene, normalized to reference genes, and to the expression in non auxin-treated controls. From the 19 tested genes containing an IR8 site, 8 were significantly upregulated in at least one of the timepoints, while 4 were down-regulated. In addition, 1 gene was first up- then down-regulated, another vice versa, and 5 were not affected by auxin (Suppl. Fig 2). From the 19 tested genes containing a DR5 site: 14 were upregulated and 5 were not affected by auxin (Suppl. Fig 3). In summary, expression levels of 74% (14 of 19) genes containing a DR5 or IR8 motif were affected by auxin. This data suggests that the presence of these bipartite elements serves as a reliable indicator that gene expression is regulated by auxin via ARFs. We checked the position of the elements relative to the translational start of each gene (Figure 8) but there seems to be no correlation of the position to its regulatory functionality. The number of TGTC elements in the 2Kb upstream of these genes was quite homogenous (14 ±3) and didn't have correlation with their up- or down-regulation upon auxin treatment. In fact, the average number of TGTC elements in a random sample of 55 genes was not significantly different (13+/-5).
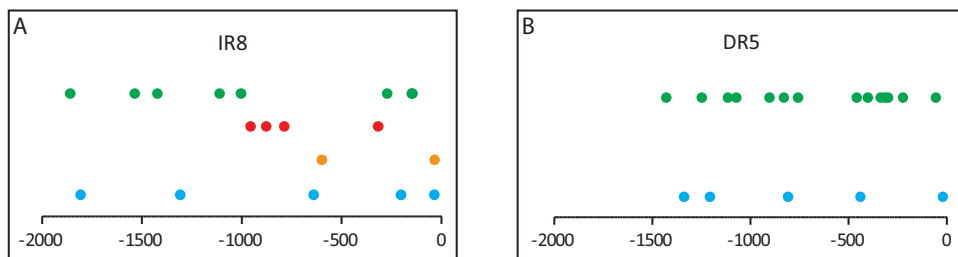
Figure 8. Position of bipartite AuxREs in gene regulatory regions.
Positions of bipartite IR8 (A) or DR5 (B) AuxREs upstream the translational start codon of genes for which expression was tested by qPCR after auxin treatment. Green shows the position for up-regulated genes, red for down-regulated genes, orange for genes which regulation changed in time (up -> down or vice versa) and blue genes for which auxin had no effect in their expression.

The meta-analysis of DR5 and IR8 showed a correlation of the motifs to auxin response in time, with IR8 being associated with early, and DR5 with later responses. This trend could not be confirmed by qPCR. A bigger sample might be necessary to draw conclusions on this matter. Four genes out of the whole selection turned out to have both elements in their promoters, but this was not correlated neither to the kind nor the amplitude of the response. We have to consider that because of the nature of the sampling we might be losing information of auxin effect as some might not be expressed in root tissue or at the developmental stage from which the cDNA was taken. Despite this, the data obtained here show that the presence of the bipartite element in promoter regions can be used as an indicator of auxin-dependent expression, with more confidence than the presence of

single AuxREs that are found in nearly all gene promoters. While both IR8 and DR5 are associated with gene up-regulation after auxin treatment, only IR8 is also correlated with gene down-regulation. This is in strong agreement that both ARF1 (Class B) and ARF5 (Class) bind cooperatively to IR8 elements [126], while only Class A ARFs ARF5 and ARF6 could cooperatively bind to DR5.

Testing the *in vivo* impact of spacing in bipartite AuxRE's

ARF DBDs bind strongly as a head-to-head dimer to an inverted repeat of AuxREs. Both ARF1 and ARF5 DBDs bind strongly to IR7 and IR8, but ARF1 bound poorly to IR5/6/9, while ARF5 bound these sites with high affinity. We thus proposed that ARFs act as molecular calipers and that specificity of binding depends on the space between two inverted repeats (Chapter 3 [126]). Bio-informatics only confirmed strong association of IR8 with auxin response, while no other IR motifs were enriched. However, since other IR repeats will likely be bound by fewer ARF proteins, it is likely that such ARF targets are expressed only in specific tissues or conditions, and would thus be missed in our stringent analysis. Whether each ARF has a preference for a specific spacing *in planta* is not known but, given that ARFs have distinct expression patterns (described in Rademacher et al. 2011), one would predict different spacing to drive different spatial patterns of gene expression. We therefore designed gene expression reporters in which this prediction could be tested. We generated a series of reporters where the red-fluorescent tandem Tomato (tdTom) is driven by synthetic promoters consisting of an inverted AuxRE repeat followed by a minimal 35S promoter (TGTCGG-n-CCGACA : min35S :: tdTom). In these reporters, the space between the two TGTCGG elements was varied. As a control, DR5v2 driving nucleus localized green fluorescent protein (n3GFP) was placed on the same transgene such that the expression of the "spacing reporters" could be compared to a reference auxin response reporter. In addition, the expression of DR5v2-n3GFP was used to select lines with comparable expression levels. Because ARF expression patterns are highly diverse in the heart stage embryo [44], the expression of the reporters was observed during this stage of development (Figure 9). In this chapter, we include the results obtained for reporters driven by DR5, IR0, IR6, IR8, and IR9. It is notable that none of these reporters, even the DR5 version, showed the same expression pattern as the DR5v2 reporter. All reporters displayed a very similar expression pattern with only subtle differences. All were mainly expressed in the shoot apical meristem (SAM) region and the inner part of the future cotyledons. DR5 has the narrowest expression around this area while the IR6 and IR9 patterns were broader. IR0 is also narrow around the SAM region but it is also expressed in the epidermal layer. IR8 expression overlaps greatly with the one of DR5v2 and is also expressed in the inner part of the cotyledons and SAM area. It is possible that these reporters reveal the activity of a single AuxRE rather than a repeat, as IR0 (TGTCGGTGTCGG) is unlikely to allow binding of a dimer given structural constraints. Despite this, differences in the expression of the reporters, though subtle, can be observed. This is especially the case for IR8. From the reporters analyzed, IR8 would be the one giving the most optimal binding site predicted by the ARF DBD structure, and it is precisely this element that best resembles the DR5v2 reporter. None of the reporters directly mirrors
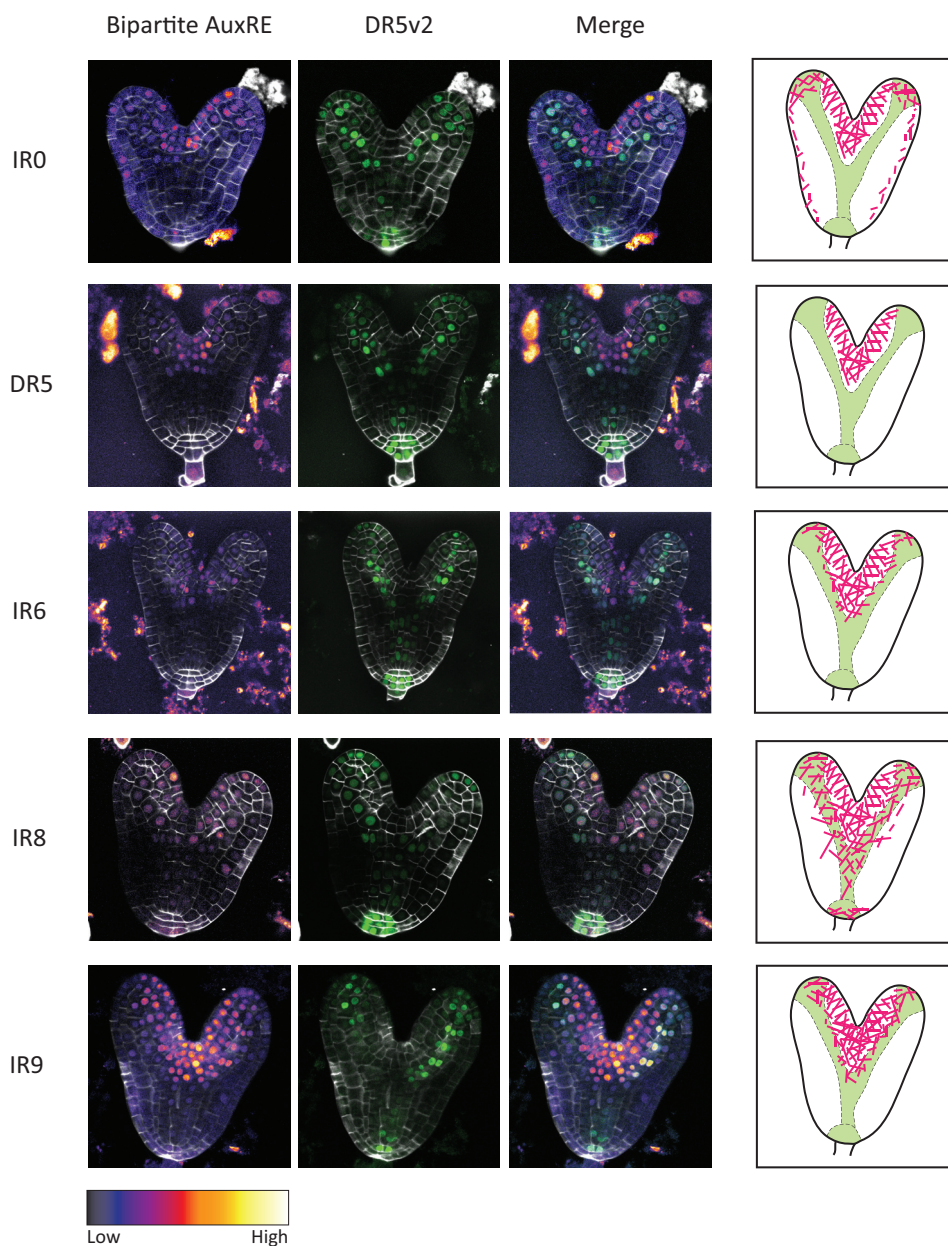
Figure 9. Expression pattern of reporters driven by synthetic promoter containing bipartite AuxREs.
Fluorescence patterns of heart-stage embryos carrying transgenes with tdTomato driven by a minimal 35S promoter carrying IR or DR element with variable spacing as indicated on the left, and 3xGFP driven by the DR5v2 promoter. The first column shows false colored images of tdTomato fluorescence (color scale is shown below), while the second column shows GFP fluorescence which was used to normalize transgene expression levels between lines. The third column shows the merge of both fluorophores. White signal in all images is cell wall staining of Renaissance RS2200. The right column shows an interpretation of the expression patterns of the bipartite AuxRE reporters in pink and the DR5v2 expression pattern in green.

the expression of individual ARFs at the same stage of development[44]. It is possible that the configuration of a single inverted repeat, connected to a minimal 35S promoter, is not sufficient to recapitulate the activity of the same element in a native promoter. Thus, further optimization will be required to develop spacing-based reporters for ARF activity.

Discussion

Auxin controls numerous plant growth and developmental processes through a simple signaling pathway with only three dedicated components: a receptor F-box protein, an auxin-degradable transcriptional inhibitor and a DNA-binding transcription factor. Based on a range of genetic and biochemical data, the DNA-binding ARF transcription factors appear to generate specificity for gene regulation, and diversity in ARF properties likely allows different sets of genes to be regulated by auxin in different cell and organ types [46–48,170,176]. In Chapter 3, we determined the structures of the functionally diverse *Arabidopsis* ARF1 and ARF5 DNA-binding domains. We found that both proteins recognized an essentially identical DNA motif, different from the one previously considered as the "canonical" ARF binding site. In addition, we found that both ARF1 and ARF5 DBD's dimerize, and that this dimerization allows high-affinity binding to inverted repeats of ARF binding sites (AuxRE's). Surface Plasmon Resonance experiments showed that ARF1 and ARF5 differ in the space tolerated between the two inverted AuxRE sites, and we thus postulated the ARFs to acts as molecular calipers, selecting target sites both by specific DNA interactions, and by "measuring" the space between these sites. This structural, and mostly *in vitro*, work made several predictions about ARF-dependent gene regulation. In this Chapter, we have tested these predictions.

First, Protein-Binding Microarray experiments had identified TGTCGG as the preferred high-affinity binding sites for both ARF1 and ARF5, and this result was confirmed by sequencing of *Arabidopsis* genomic fragments that bound *in vitro* to ARF1 and ARF5 proteins [126]. Furthermore, re-design of the auxin response reporter DR5 to encompass the high-affinity TGTCGG site showed that this motif indeed conferred high-affinity auxin-responsive gene expression and activity in an expanded range of cells [175]. Here, we first addressed the structural basis for this high-affinity binding. The crystal structure of an ARF1-DBD/DNA complex with a TGTCGG-(N8)-CCGACA oligonucleotide revealed that the same Histidine in the B3 moiety of the DBD that engages in interactions with the G opposing the terminal C in the medium affinity TGTCTC site can now interact with G's in position 5 and 6 of TGTCGG. The electron density in the crystal showed that the same Nitrogen in the Histidine side chain can either form a Hydrogen bond with G5 or G6. However, in the complex with the TGTCGG element, the backbone of the Histidine rotated such that the backbone Oxygen formed a second Hydrogen bond with the C opposing the G6 position. Thus, the high-affinity binding observed with TGTCGG can be explained by a single additional Hydrogen bond formed by the same DNA-contacting amino acid. This Histidine is conserved among most Class A and Class B ARFs (except in the cluster of recently duplicated ARFs near the centromere of chromosome 1), yet is notoriously absent

in Class C ARFs (ARF10, 16 and 17). A prediction from this finding is that Class C ARFs either do not bind the TGTCGG site with affinity, or use a mode of binding that is different. Further biochemical and structural studies should answer this question.

A second prediction from the structural data in Chapter 3 is that ARFs bind inverted repeat sites in their *in vivo* target genes. We first scrutinized the regulatory regions of *TMO5* as this gene is directly regulated by ARF5 [107]. Among the many TGTC core elements in the *TMO5* promoter, there was only one that made an inverted repeat with a spacing of 7 nucleotides between the AuxREs, spacing that should allow binding of an ARF5 dimer. Elimination of only one of the inverted AuxREs in the IR7 element led to significant drop, of more than half, in the expression level of the TMO5-3GFP fusion protein. Elimination of both inverted elements reduced the expression of the protein to almost undetectable levels. We concluded from this experiments that indeed this kind of bipartite AuxRE – ARF dimer interaction does occur *in vivo*, and that it can be the main source of functional gene regulation by ARFs.

The presence of a single TGTCNN element is often a hint that a gene may be under ARFs regulation. Given that any 4-mer, unless specifically depleted, will be found on average in every 256 basepairs, most plant gene promoters will carry at least one TGTCNN element. Thus, the presence of such a motif hardly predicts auxin response. The structures presented in Chapter 3 indicate that composite AuxREs may be more predictive of the auxin responsiveness of genes. An important question was whether regulation through a composite AuxRE is not only valid for TMO5, but also on a genome-wide scale. In other words, can we use the presence of composite AuxREs to predict auxin responsiveness of genes? To answer these questions, we performed a whole genome analysis of composite AuxREs. We included in the analysis all possible conformations of bipartite AuxREs: direct repeats, inverted repeats and everted repeats with spacing ranging for 0 to 15 between the TGTCNN elements. Our analysis showed a very high correlation of two bipartite AuxREs to gene regulation by auxin: DR5 and IR8. However, neither of these motifs is unexpected. IR8 would be the optimal binding site as predicted by the DBD structures of both ARF5 and ARF1. DR5 is precisely the AuxRE conformation in the widely used DR5 auxin activity reporter. This reporter consists of 9 TGTCGG direct repeats, each spaced by 5 nucleotides. Although the current structural information of the DNA binding domains of ARFs cannot explain how a dimer could bind this element, SPR experiments revealed that DR5 can indeed be bound by DBD dimers of ARF5 and ARF6, both belonging to the Class A "activator" ARFs. Dimer binding to a DR5 site would require a head-to-tail dimerization, like the one mediated by the PB1 domain in the C-terminus of ARFs, but so far there has been no evidence suggesting that dimerization via this domain is involved in functional DNA binding. The mechanistics of such event are still an open question that deserves further study.

In order to test whether the presence of IR8 and DR5 in regulatory regions can point to ARF regulated genes, we took a sample set of 19 genes of each class and tested their expression in seedling roots after auxin treatment. We found that the expression of 74% of the genes of each class was significantly altered by auxin treatment. These results are very encouraging as the presence of DR5 and IR8 might be used with high confidence to predict ARF targets. An interesting observation is that DR5-containing genes were only

up-regulated in our qPCR experiments. SPR experiments showed that only ARF5 and ARF6 could cooperatively bind DR5, likely as dimers. ARF5 and ARF6 are closely related and known activator ARFs. Recent studies have suggested a role for ARF5 in chromatin unlocking as it interacts with chromatin remodelers of the SWIF/SNF family [25]. Unlocking chromatin makes genes available for transcription and is a plausible mechanism for gene activation. The association of DR5 only with auxin-dependent gene upregulation may thus be causally connected to its ability to recruit ARF5 and 6, which next help recruit SWI/SNF complexes for chromatin remodeling. This model makes a number of predictions, i.e. that DR5 activity depends strictly on SWI/SNF factors, and that inclusion of a DR5 motif in a promoter may be sufficient to recruit ARF5 and 6 and SWI/SNF factors. These predictions can now be tested in future experiments.

A final prediction from the structural analyses in Chapter 3 is that the space between the inverted AuxREs contributes to specific DNA recognition: the caliper model. The DNA binding interface of ARF DBDs are highly conserved, and therefore might recognize the same bases but the loop connecting this interface with the dimerization interface is highly variable and may allow each ARF to position the DNA binding interfaces of a dimer at different distances, each ARF having an optimal one. To start addressing this question and its biological significance, we designed a set of artificial reporters consisting of AuxRE repeats with different spacing coupled to a minimal 35S promoter and observed the expression pattern of the reporter proteins in heart stage embryos of *Arabidopsis thaliana*. Five reporters were tested: DR5, IR0, IR6, IR8 and IR9. Although none of the reporters had the same expression pattern as the DRv2 reporter, IR8 overlapped greatly with it. The others not only did not overlap but also only very subtle differences could be observed between them. It is interesting to notice that the expression of all our bipartite reporters do overlap with the regions were the YUC1 and YUC4 auxin synthesis enzymes are expressed in heart stage embryos[55]. Thus, these sites are likely sources of auxin and may have high endogenous auxin concentrations. Furthermore, their expression overlaps with the regions were auxin activity is seen by the R2D2 auxin activity reporter [175]. This indicates that the bipartite AuxRE reporters respond to the presence of auxin. Indeed, post-embryonic patterns of expression were similar to the DR5v2 reporter (not shown), suggesting that the elements indeed report auxin activity. The similarity in expression of the different reporters may indicate that the activity is largely the response conferred by a single AuxRE, except for IR8 where both AuxREs may provide a more specific binding by some ARFs. It is not possible to correlate the expression pattern of the reporters with individual ARFs. The difference of responsiveness to auxin between the bipartite AuxRE reporters was not tested yet, but it could provide with more information about the role of this kind of regulatory elements in an efficient auxin response. Thus, this preliminary analysis into the involvement of inter-site spacing to ARF-dependent gene regulation is not yet conclusive, and further optimization of the reporter design will likely be required to fully test the importance of spacing.

In summary, in this Chapter we present a multi-faceted approach to test a number of predictions made by the structure-guided model of ARF action. We reveal the structural basis for high-affinity binding of ARF monomers to optimal binding sites. We further show

that bipartite AuxREs are important for auxin mediated gene regulation both at the single gene level and at genome-wide scale. The data presented in this and the previous chapter give a biophysical explanation for ARF preferences for specific sequence and topography of AuxREs, as well as their *in vivo* relevance. On the other hand, the contribution of site spacing remains a question.

Material and Methods

Crystallization and structure determination

The synthetic oligonucleotide of sequence TTGTCGGCCTTTGGCCGACAA and its complementary strand were ordered from Biomers (Ulm, Germany). An equimolar mix of both strands at 3 mM was heated to 85° and left to cool slowly overnight. ARF1 DBD, purified as described previously (Boer et al, 2014), was slowly mixed in 2:1 stochiometry with the resulting solution of dsDNA. Crystals of the complex (A1D-GG) were obtained with condition B10 of the Morpheus crystallization screen (Molecular Dimensions) as crystallization buffer, using the sitting and hanging drop vapour diffusion method. Crystals appeared after 1-2 days and reached maximum size after one week. Crystals belonged to space group $P2_1$ and were typically harvested after 3 days and frozen by flash freezing into bulk liquid nitrogen. A full dataset (Supplemental Table 2) was measured at the XALOC beamline [177] at the ALBA synchrotron. The data were processed using IMosflm [153]. The structure was solved through molecular replacement with PHASER [178] from the CCP4 package [157], using the ARF DBD – ER7 complex (PDB structure 4LDX). The structure was manually adjusted and refined iteratively using Coot 0.7.1 [179] and Refmac 5.8.0131 [156] from the CCP4 package, respectively. The final model was refined to an R_{work} and R_{free} of 18.5 and 22.8 respectively (Sup. Table 2). The model was validated using Coot and Molprobity [180] with an overall score of 99%.

Search for bipartite AuxREs in *A. thaliana* genome

TAIR10 *Arabidopsis thaliana* genome and its annotation were taken from the Plant Ensembl database (MySQL server, http://www.ensembl.org/info/data/mysql.html). [-1500; 5'UTR] regulatory regions (1500 bp upstream transcription start site and full-length 5'untranslated region) of the genes that had a unique probe on the ATH1 microarray platform were taken into analysis. Bipartite AuxREs with two joint TGTCNN motifs spacing one of other on the distance ranging between 0 to 15bp were searched in the regulatory regions. Direct (DR), Inverted (IR) and everted (ER) repeats were considered separately, so that 3*16=48 variants of bipartite AuxREs were into the analysis. On this step, we prepared 48 overlapping gene sets having certain bipartite AuxRE variants in their upstream regions.

Microarray data analysis

Seventeen publicly available microarray datasets on auxin treatment of *Arabidopsis thaliana* were selected from the GEO database (listed in the Supplemental Table 1). These datasets were chosen for having at least 50 differentially expressed genes. Differentially expressed

genes were defined as those expressed at least 1.5 times higher or lower after auxin treatment compared to control, with false discovery rate (FDR) adjusted p-value < 0.05 (Welch t-test with Benjamini–Hochberg correction). To reveal the associations between the presence of a particular bipartite AuxRE variant in a gene regulatory region and its auxin responsiveness we used the exact Fisher's test of a 2x2 contingency table that reflected the association between the presence of a certain bipartite AuxRE variant in the upstream region [-1500; 5'UTR] and auxin up(down)-regulation status of the gene. The significance level threshold was adjusted by Bonferroni correction, so that the association found in the exact microarray was considered significant if p-value < 0.05/48=0,001042.

Protein expression and purification.

Regions corresponding to the DNA binding domain (DBD) of *Arabidopsis* ARF1 (At1g59750; residues 1-354), ARF4 (AT5G60450; residues 1-406), ARF5 (At1g19850; residues 1-390), ARF6 (AT1G30330; residues 1-359), ARF10 (AT2G28350; residues 1-387) and ARF16 (AT4G30080; residues 1-390) were amplified from cDNA clones using Phusion Flash polymerase (Finnzymes), and cloned in an expression vector pTWIN1 (New England Biolabs) to generate fusions with Chitin Binding Domain (CBD) and Intein. ARF-DBD-CBD fusion proteins were expressed in *E. coli* strain Rosetta DE3 (Novagen). Protein expression was induced by 0.3 mM IPTG for 20 hours at 20 °C, and proteins were purified from cell-free extracts by affinity chromatography on a chitin column followed by size exclusion chromatography on a Superdex 200PG column, both using an Akta Explorer 100 (GE Healthcare).

Surface Plasmon Resonance

Surface Plasmon Resonance was performed as described in Boer et al.[126]. Purified DNA binding domains of ARF1, ARF4, ARF5, ARF6, ARF10 and ARF16 were bound to the following oligos: No AuxRE: CCGGtAGGTT, One AuxRE: CCGGtAGGT**TGTCTC**CCTT and DR5 AuxRE: CCGGtAGGT**TGTCTC**CCTTT**TGTCTC**CCCTT. All traces were normalized against binding traces to the No AuxRE oligo.

Quantitative RT-PCR analysis

Primers were designed using Beacon Designer 8 software (Premier Biosoft International) RNA isolation was done with TRIzol reagen (Invitrogen) and the RNeasy kit (Qiagen). cDNA was prepared with the iScript cDNA Synthesis kit (BioRad) according to the protocol provided by the manufacturer. qRT-PCR reactions were performed with iQ SYBR Green Supermix (BioRad) and run in a CFX384 Real-Time PCR detection system (BioRad). Reactions were done in triplicate with two biological replicates. Data were analyzed using the Bio-Rad CFX Manager3.1 software and gene expression levels were normalized relative to ACT, CDKA1.1 and EEFa4. All the primers used for qRT-PCR are listed in Supplemental Table 3.

Cloning and Mutagenesis

All cloning was done using the LIC method described in [152]

Mutagenesis of TMO5 AuxREs was done by overlapping PCR. The amplified fragment with mutated sites was replaced in the original vector using the XhoI and EcoRI unique restriction sites that flank the bipartite element.

IR synthetic reporters were designed with spacers ranging from 0 to 20 nucleotides and followed by a minimal promoter. These fragments were LIC cloned into the vector pGIIM/LIC_SwaI-ntdTomato-DR5v2::n3EGFP described in Liao et al., 2015.

Plant material

All *Arabidopsis thaliana* plants used were from the Columbia-0 ecotype.

Seeds were surface sterilized and plated on half-strength Murashige-Skoog (MS) medium and appropriate antibiotic for selection of transgenic plants. Kanamycin (50mg/L) for TMO5 lines or MTX (50 mg/L) for double AuxRE reporter lines. No antibiotics were used when plants where grown for RNA extraction. Seeds were vernalized for 48 hours at 4°C and then grown under long-day (16h light, 8h dark) conditions at 22°C.

For auxin treatments, 5 days-old seedlings were transfer into MS media containing 2,4-Dichlorophenoxyacetic acid (2,4D) (dissolved in DMSO) at a 1μM concentration and roots were harvested at the appropriate times. For each time point, control groups were taken where seedlings have been transferred to MS containing DMSO.

Microscopy and TMO5-GFP quantification

All the root and embryo confocal imaging was done in the Leica SP5 II system (HyD detector) microscope with 20× numerical aperture (NA) = 0.75 and 63× NA = 1.20 water-immersion objective and pinhole equivalent to 1.0× the Airy disk diameter. Roots were stained with a solution of Propidium Iodide (PI) and embryos with SCRI Renaissance Stain 2200 (R2200 Renaissance Chemicals, UK). Excitation and detection of fluorophores were configured as follows: EGFP was excited at 488 nm and detected at 498–530 nm; tdTomato was excited at 561 nm and detected at 571–630 nm.

Fluorophore intensities were measured using the LAS-AF 2.6.3 software (Leica Microsytems CMS GmbH). Acquisition settings were set on the brightest sample and kept for all the samples. The relative average pixel intensities were measured using the same region of interest in all samples.
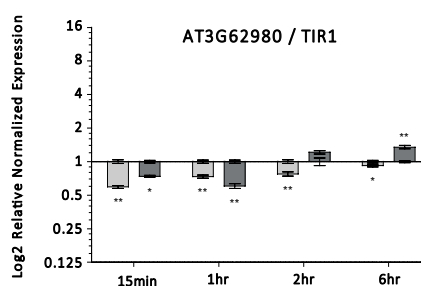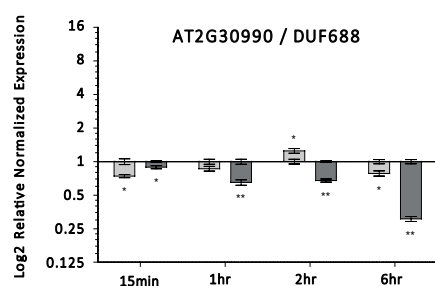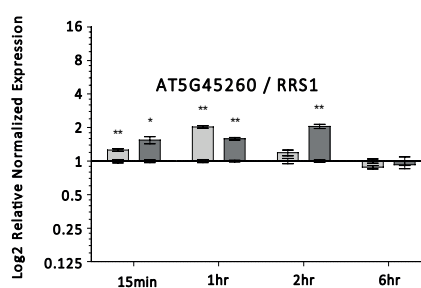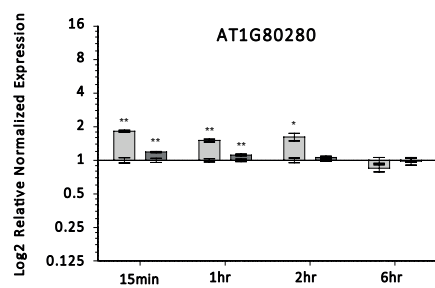
Acknowledgements

Supplementary Figures and Tables

gtcgtgtgggcttcactctgcaccaccaccccaaaaccctagttcataccctatcctataatttatactcaaaattttctcat
ttcttctattatttatcaccattcatatggaagtatatatttaatgaagaattttcggatctatagatcattatatgtaaaat
cagtttttgataaaacatgtgacagtttcacaaaaatagttcatctttcacaaaaatagtactcgaatccggtagcattttag
tatcctactatttagtttttccactccttttcattatatatataatgatataagtctttttacctctttttcataagtctttt
taactctttttttctttaaaaatgattatatgcacatctacttgataaactggcggatggtaaacttgagttaggtttgagaa
ttttcactaattattttaggatttgaatcttctttagtatgagacaattaagctaacacactcgaagatatattcagatcata
atcaaaatttatccgttggatatgtatgaatagattaaataaaagagagaatctcttcggtgcactt<u>gagacaaatc</u>tataac
atagaaaatcttgattagattaaaaactagattaaagaaggaaagaaaaacaagaaagtgggtttggctatacgaaacgagtg
ggattattaatgattaaaagtaaaagtctttt<mark>ggtctctggtcggtcgaca</mark>atccgtg<u>tgtcct</u>gtggtcagtgtttgtttt
ttattcttttggttaattttcgagtattattatgctttccactgaaaaaacaaaaacattaagcagccaaatgaaattctttt
aacgcgttgtttcagaacctgatgaaaaaacaataatcatgacctctttcttcttcccctttaaatgc<u>atgtcaaat</u>tgtgta
ggagtgtttgattatataaattaaatctactttagtcttctctcttctttgtagaattgtagagtcgtggtataaactcgaaaat
ttcttattatacttatagaaattaggcagacgaaatgctcttgatttgcttacaatcttagattaggtttaacccaagataaa
aaaaaactttgtttttgatgaagaaaaacacctttagttacaattt<u>tggacat</u>catcttttattttcacagttagcatctgagac
tttttttttctttttcttaaatagagttct<u>tagaca</u>atatcttttgggtccctgatca<u>tgtcaaat</u>gccaaaaaaataaataagt
atacatgattg<u>tgtcta</u>tttgaccattctaacttctaaaagtttcaaatgaaataacttt<u>tgtcaac</u>tattatttttgcctatg
aattaggtgctgaccagaatcatcttgctaaaatcatgattttctactattgcatattattcttgatcgcatttcacagatct
taatatatacgaaaataactgtatatattatcaactgttacgattaatatataaccatgcatatcgaaatattttgaattagt
atttttttttgctaaataatattttgatatagtggtggcgtggtattcctcatttacaaatcgcatgaactttatcctaaatac
gattaacttaagtagaaaaggcacaaaaatcttctattcattaacatgaaatcaattcttaagaacataaataaataataaata
ttcataggaagccaaatccatcagagaaatccacaagatg<u>atgacag</u>ctccccatttagaagtcgtaaaagaaaagcaaagaa
gatgtgcagcacttgcagtttcatttgttttgttagatataccagaaaaaaaaaaagagtggtcccataaaaacctgtaatat
ataacattttaatcaaagaaaatgatagatttttttattaaacaggaaaaaccactaactttttaattaaaaaatttggaatac
agcaatacatcaaaaatccattaacggtacattatagttgaaaataaattcaggttacataatatttttttgcgtaatcactat
actattctgccattttatatggaactaagttcatacgttatttctcacaattctccttgatgtgatcaaagataatggcggtc
tgattgtataacccggtgaaaaaggtttttttggaaagaattctttaatattaaaatatcaaataaattttgtttaaaaattaa
<u>ggaca</u>cattatatagcactatatatacacatgtgtatgcattatcaaatgggatgtttgtgaccttttgtctcttcttctcta
aaaattctctttctcccaaaaactaaaaaaaccaaaaaaaccaaaaa**ATG**

Supplemetal Figure 1: 2.3kb TMO5 promoter region.
AuxRE-like sites, direct or inverted, are shown with an underline. Inverted repeat with spacer distance
permissive of ARF dimer binding (IR7) is highlighted.

AT4G28640 / IAA11

AT1G80240 / DGR1

AT4G30080 / ARF16

AT3G56810

AT1G18400 / BEE1

AT5G67250 / SKIP2

AT1G80280

AT5G45260 / RRS1

AT2G30990 / DUF688

AT3G62980 / TIR1
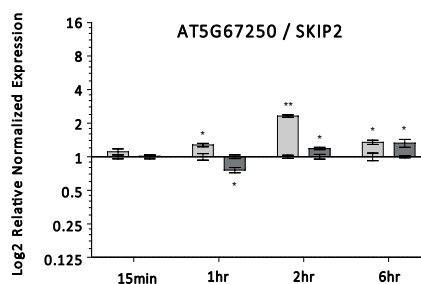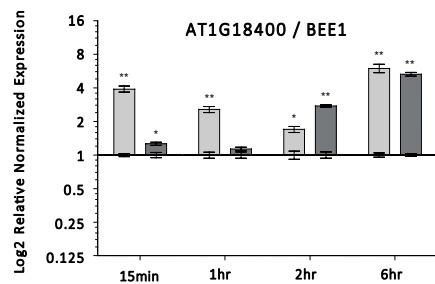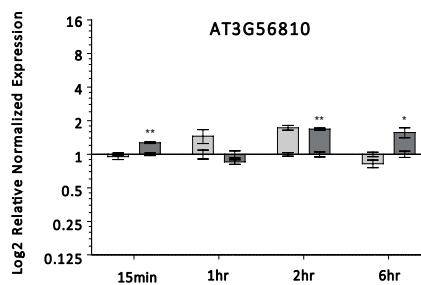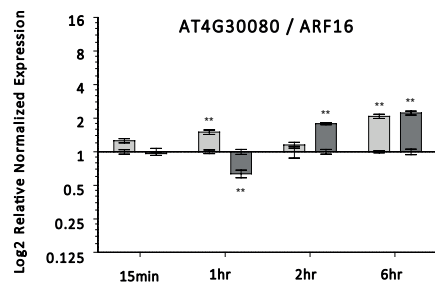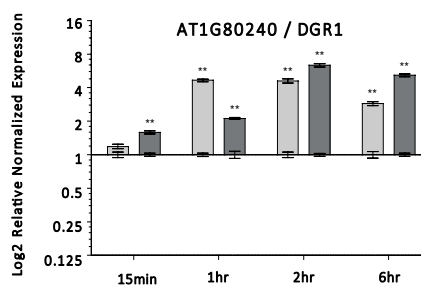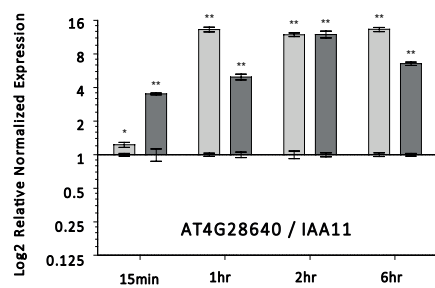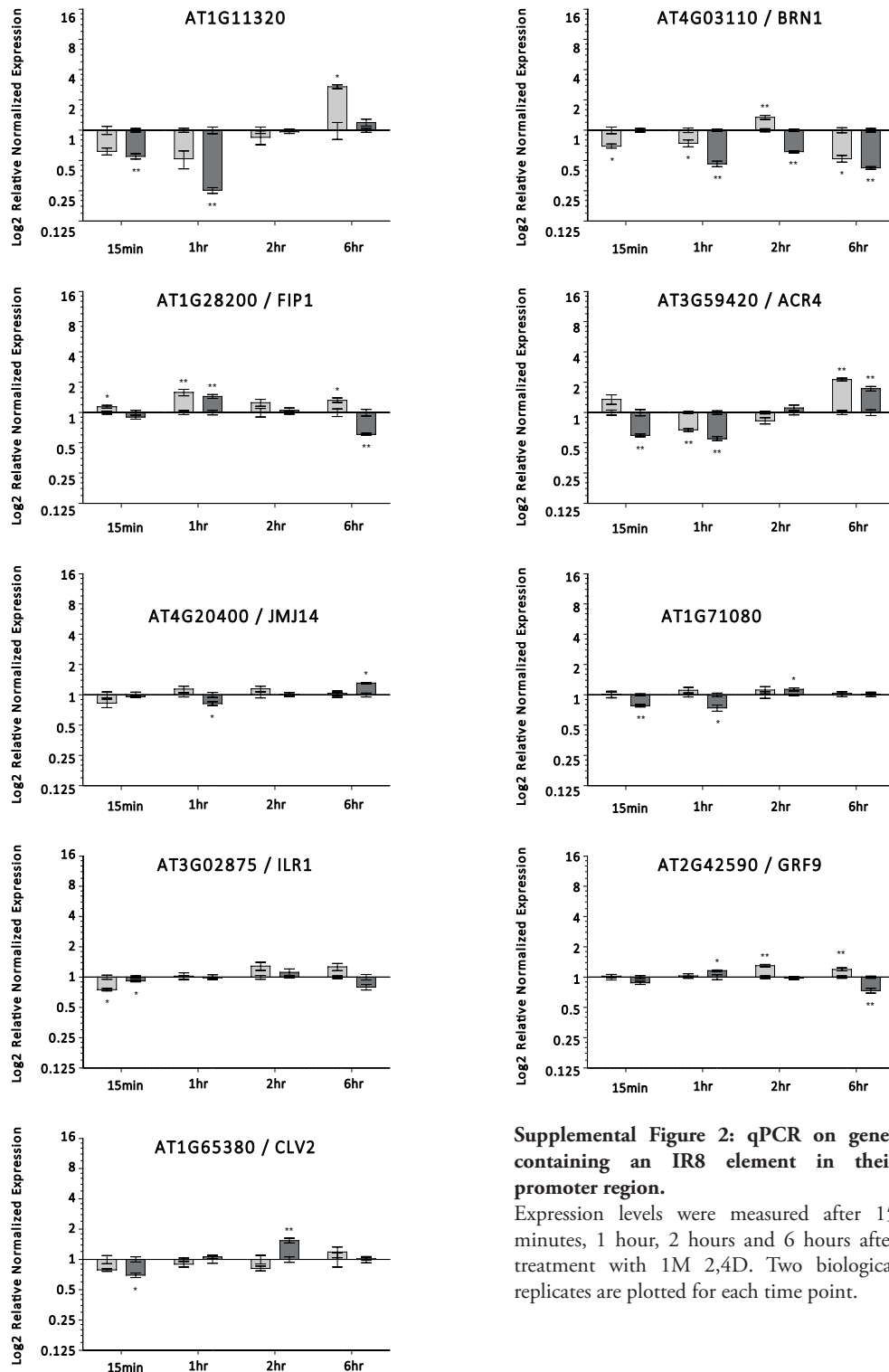
Supplemental Figure 2: qPCR on genes containing an IR8 element in their promoter region.

Expression levels were measured after 15 minutes, 1 hour, 2 hours and 6 hours after treatment with 1M 2,4D. Two biological replicates are plotted for each time point.

AT1G19220 / ARF19

ATG323030 / IAA2

AT2G26710 / BAS1

AT5G60450 / ARF4

AT1G14350 / FLP

AT5G24590 / TIP

AT3G50060 / MYB77

AT3G55610 / P5CS2

AT3G54000

AT4G39100 / SHL1

Supplemental Figure 3: qPCR on genes containing an DR5 element in their promoter region.

Expression levels were measured after 15 minutes, 1 hour, 2 hours and 6 hours after treatment with 1M 2,4D. Two biological replicates are plotted for each time point.
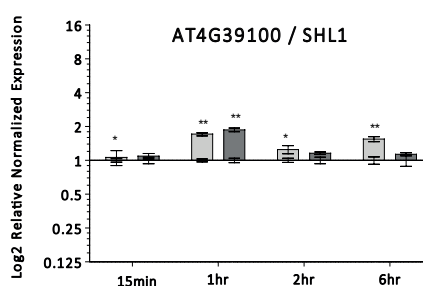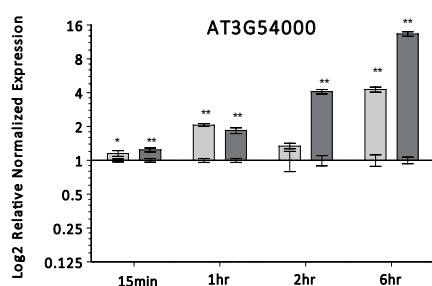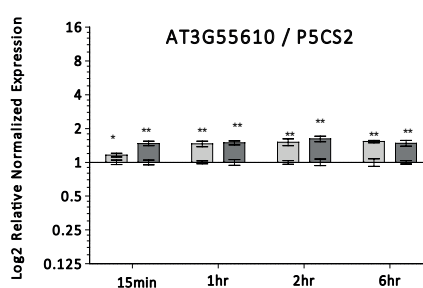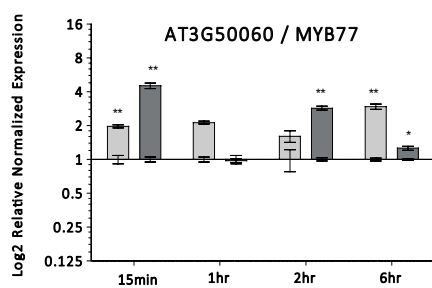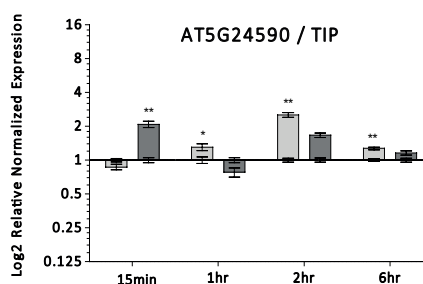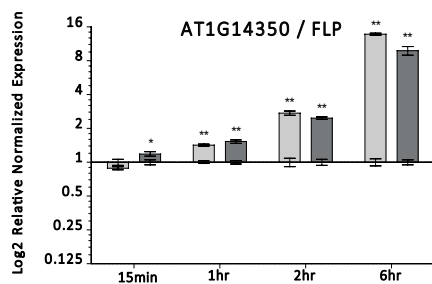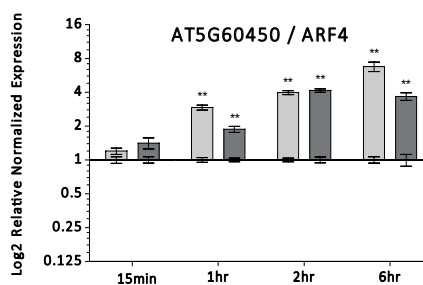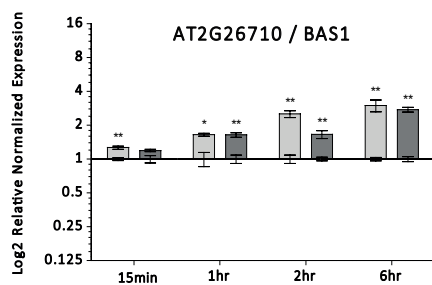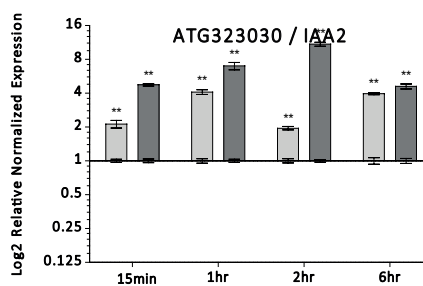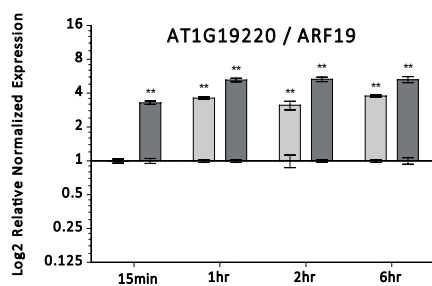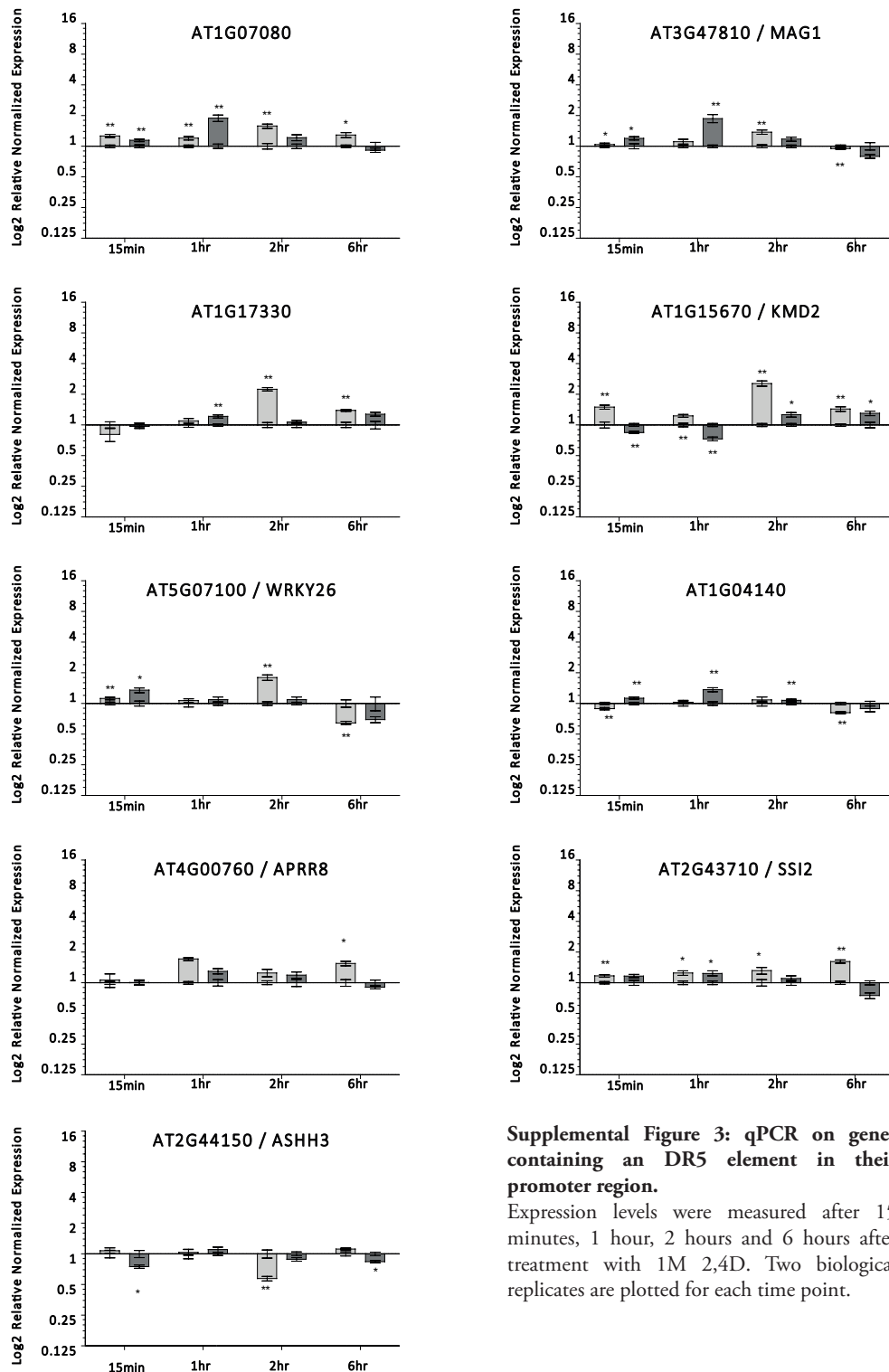
Supplemental Table 1. Publicly available microarray datasets used in this study.

No	Microarray ID	Tissue	Treatment [concentration; time]	Number of DEGs (up/down)	Reference
1	GSE35580	Roots of 7 dag seedlings	5 μM IAA, 2 h	571/577	Bargmann et al., 2013
2		Root, epidermis		129/138	
3		Root, stele		279/254	
4	GSE627	7 dag seedling	5 μM IAA, 2 h	109/39	Okushima et al., 2005
5	GSE18975	7 dag seedling	1 μM IAA, 30 mins	256/64	Delker et al., 2010
6			1 μM IAA, 1 h	327/112	
7			1 μM IAA, 3 h		
8	GSE42896	Root segments of 3 dag seedling	10 μM NAA, 2 h	790/709	De Rybel et al., 2012
9			10 μM NAA, 6 h	3412/4018	
10	GSE59426	Root segments of 3 dag seedling	10 μM IBA, 6 h	786/332	Xuan et al., 2015
11	GDS672	10 dag seedlings	0.1 μM IAA, 1 h	90/17	Redman et al., 2004
12			0.1 μM IAA, 3 h	88/49	
13			1 μM IAA, 1 h	242/116	
14			1 μM IAA, 3 h	289/282	
15	GSE42007	Roots of 6 dag seedlings	1 μM IAA, 4 h	520/260	Lewis et al., 2013
16			1 μM IAA, 8 h	663/516	
17			1 μM IAA, 12 h	539/263	

Note: All the experiments were performed on the same ATH1 microarray platform and have at least 2 biological replicates.

Supplemental Table 2 . Table XRAY: X-ray data collection and structure refinement statistics.

Structure	A1D-GG
Data collection	
Space group	$P2_1$
Cell dimensions	a =43.36 Å,
	b =102.96 Å,
	c =126.98 Å,
	β =126.98°
Resolution (Å)	125.79-1.980
	(1.986-1.980)[a]
R_{merge} [b]	4.6 (64.3)
$<I / \sigma(I)>$	16.9 (2.0)
Completeness (%)	97.3 (99.1)
Multiplicity	3.4 (3.4)
Refinement	
Resolution (Å)	50-1.94
No. reflections	82168
R_{work}[c]/ R_{free}[d]	18.5 / 22.8
No. atoms	
Protein	6891
DNA	855
Ligands/water	515
B-factors	
Wilson	-
Overall	34.9
Protein	33.9
DNA	36.6
Ligands/water	41.75
R.M.S. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.79
MolProbity scores	
Overall score	1.48 (96th %)
All-atom clashscore	2.87 (99th %)
Ramachandran outlier (%)	4 (0.6%)
Ramachandran favoured (%)	559 (90.8%)

[a] Throughout the table, the values in parentheses are for the outermost resolution shell.
[b] $R_{sym} = \Sigma_{h,i} \mid \hat{I}_h - I_{h,i} \mid / \Sigma_h \Sigma_i I_{h,i}$, where $\hat{I}_h = (1/n_h) \Sigma_i I_{h,i}$ and n_h is the number of times a reflection is measured.
[c] $R_{work} = \Sigma_{hkl} \mid \mid F_{obs} \mid - k \mid F_{calc} \mid \mid / \Sigma_{hkl} \mid F_{obs} \mid$
[d] $R_{free} = \Sigma_{hkl \subset T} \mid \mid F_{obs} \mid - k \mid F_{calc} \mid \mid / \Sigma_{hkl \subset T} \mid F_{obs} \mid$ where T represents a test set comprising ~5% of all reflections excluded during refinement.

Chapter 4

Supplemental Table 3: Primers used in this study

Gene	Primer orientation	Sequence 5'- 3'
ACT2	Sense	CTCCATTTGTTTGTTTCATT
	Antisense	TCAATTCGATCACTCAGA
CDKA1	Sense	ATTGCGTATTGCCACTCTCATAGG
	Antisense	TCCTGACAGGGATACCGAATGC
EEFα4	Sense	CTGGAGGTTTTGAGGCTGGTAT
	Antisense	CCAAGGGTGAAAGCAAGAAGA
AT4G28640 IAA11	Sense	TTGGTCGCAAAGGGTATC
	Antisense	TCTTGATCCCAGCAATTACG
AT1G80240 DGR1	Sense	TATATCCGTGACGCATGA
	Antisense	AAGCCCATGAGTATGAGT
AT4G30080 ARF16	Sense	CATCAAATACGCAGGAAA
	Antisense	TTATACTACAACGCTCTCA
AT3G56810	Sense	TTGATGGAGAGCGTGAAC
	Antisense	ATGAAGAGCCAAACCGATT
AT1G18400 BEE1	Sense	GATCACAACATCGTCGAT
	Antisense	CAGTGGCAATAACATTCG
AT5G67250 SKIP2	Sense	TAGATTAGCGATGATTAGGT
	Antisense	AACGAACGAGGATTACAA
AT1G80280	Sense	GTACTTAGTGAGGAGTGA
	Antisense	GCTTATAATGAATACAGACTTAC
AT5G45260 RRS1	Sense	AATATGGAGGTGGATAATGA
	Antisense	AATCCTTGGATCTCACTG
AT2G30990 DUF688	Sense	GTTCAAGGCTTCATCATC
	Antisense	TGGTGGTATTGTTGGTAA
AT3G62980 TIR1	Sense	ATTGATAACCATTGCTAGG
	Antisense	CAATATCCAGTGGCTCTA
AT1G11320	Sense	TTAGGAGGAGAGTTGTCT
	Antisense	AGTAGAAGCCTCAATGTC
AT4G03110 BRN1	Sense	AACCAAGGTAATAACAATG
	Antisense	ATAGCAGGATAACCAGAT
AT1G28200 FIP1	Sense	TGTCCTCTTGGCGTTTACT
	Antisense	CGAATCCAACTTAGAACACGAATA
AT3G59420 ACR4	Sense	GGAGGAAGCAACAAGAAG
	Antisense	TAAGGAACACATTATACATCAGA
AT4G20400 JMJ14	Sense	GACACAAAGGCTCAAACT
	Antisense	CCCATTTATACTCTCCAAAGG
AT1G71080	Sense	ATAAGAACAGGTCAGGAT

	Antisense	CCTCACTACTACCTTCAA
AT3G02875 ILR1	Sense	ACTCAGTGTTCATGTCTT
	Antisense	TGAACCGTGACTGTAAAC
AT2G42590 GRF9	Sense	GCGGTGAATAATAGTGAA
	Antisense	ATAGTAAGACGACAGTAGT
AT1G65380 CLV2	Sense	TCCTTCTTGGTTGTTCAA
	Antisense	TTAAGTTATCGTCTGGTATGA
AT1G19220 ARF19	Sense	GACTGCCAATCTCATCATC
	Antisense	AGAATAGAGTGGATCAGAAGT
AT3G23030 IAA2	Sense	GTACGAGAAAGTCAACGA
	Antisense	AATAGACGCTTGTTGTTAC
AT2G26710 BAS1	Sense	TGGTCAGAATCTTGCTATAC
	Antisense	TAAGTAGGAGCCAAGTGAA
AT4G60450 ARF4	Sense	GAATGCGAATGACGATAA
	Antisense	TAGTGATTGTAGGAGAAGAA
AT1G14350 FLP	Sense	TTTAAGTTTCTTTCCTGGT
	Antisense	GTAGATGGCTTCCTTATG
AT5G24590 TIP	Sense	TCAATGGCTTACAATCTG
	Antisense	ATCTGTTCTGGCTCATAG
AT3G50060 MYB77	Sense	TTGAGTTTGTCACTACCT
	Antisense	CATTAATCTTCATCTGACT
AT3G55610 P5CS2	Sense	GGCTGTTGAGAATGGAAT
	Antisense	GCTGCTAAACATTTCACTATT
AT3G54000	Sense	ACGAACCATACCATAGTG
	Antisense	AGAGATTGATGATTGTTGTAG
AT4G39100 SHL1	Sense	GTGAGTCTGAGTATCAATCG
	Antisense	GCCAACAAACCTAAACCC
AT1G07080	Sense	TCATTCGTAGTGTCAAATTG
	Antisense	CAGGAAGTGTTCATAGCA
AT3G47810 MAG1	Sense	CTTACAGCAGCATAAACC
	Antisense	GCTCATAGACATAGACTACA
AT1G17330	Sense	AAGGAGTTCTATGAAGAGT
	Antisense	GTAGGTACTGCTTACTCA
AT1G15670 KMD2	Sense	TGGGATATGAGCAATTCT
	Antisense	CAAAATCAACCGACCAAA
AT5G07100 WRKY26	Sense	CTTGTGAAGGGTCAGATG
	Antisense	TCTCCATTACTGCTGTTC
AT1G04140	Sense	CTATGCGTTAGAATGCTA
	Antisense	AGAAACAAAGGAAGGTAAA
AT4G00760 APRR8	Sense	GACACATGAAGCATTAGA

	Antisense	ATCGGTGAATATCTTATCG
AT2G43710 SSI2	Sense	GATACACGACAGAGAAGT
	Antisense	TTAAGAGCAGACAGATGAA
AT2G44150 ASHH3	Sense	TCTAACTTGTGACCTGTGTT
	Antisense	GCAAGACATCATCATTTGAGA
AT3G26760	Sense	ACTGAGGAGGAACAAATC
	Antisense	GCACTATTGAGCATTACG
AT4G29910 ORC5	Sense	AACCATTCTGTAGAGTCA
	Antisense	ATCCCTAAATCATCAAGTG
AT2G37390 NAKR2	Sense	TAACGATGCTAAAGGTGA
	Antisense	CAGTAAGACCAAGGCTAT
AT1G69310 WRKY57	Sense	AGAGACAATAATGCTCCT
	Antisense	ACAATATCACCAAGTAAGC
pTMO5	Sense	GGGCCCCCCCTCGAGTCGTGTGGGC
	Antisense	AATATTAAAGAATTCTTTCC
pTMO5 -1AuxRE	Sense	GGTCGGTTTTTTATCCGTGTGTCCTGTGG
	Antisense	CGGATAAAAAACCGACCAGAGACC
pTMO5 -2AuxRE	Sense	GTCTTTTTTTTTTTTGGTCGGTTTTTTATCCGTG-TGTCCTGTGG
	Antisense	CGGATAAAAAACCGACCAAAAAAAAAAAAAGACTTTTAC
IR6	Sense	TAGTTGGAATAGGATTTCTGTCGGAAAGGTCCGA-CAAAAGGGGGCAGGCC
	Antisense	AGTATGGAGTTGGAATTTCCCTGTAATTGTAATTG
IR0	Sense	TAGTTGGAATAGGATTTCTGTCGGCCGACAAAAGGGGG-CAGGCC
	Antisense	AGTATGGAGTTGGAATTTCCCTGTAATTGTAATTG
IR8	Sense	TAGTTGGAATAGGATTTCTGTCGGAAAGGTTTCCGA-CAAAAGGGGGCAGGCC
	Antisense	AGTATGGAGTTGGAATTTCCCTGTAATTGTAATTG
IR9	Sense	TAGTTGGAATAGGATTTCTGTCGGAAAGGTTTCCCGA-CAAAAGGGGGCAGGCC
	Antisense	AGTATGGAGTTGGAATTTCCCTGTAATTGTAATTG
DR5	Sense	TAGTTGGAATAGGATTTCTGTCGGAAAGGTGTCG-GAAAGGGGGCAGGCC
	Antisense	AGTATGGAGTTGGAATTTCCCTGTAATTGTAATTG
spacer sequence 5'-biotin	Sense	CCGGtAGGTT
DR5 (1x) 5'-biotin	Sense	CCGGtAGGTTGTCTCCCTT
DR5 (2x) 5'-biotin	Sense	CCGGtAGGTTGTCTCCCTTTTGTCTCCCCTT

CHAPTER 5

Contribution of ARFs Domains to *In Vivo* Activity and Protein-Protein Interaction

Alejandra Freire Rios
[a] André Kuhn
Mark Roosjen
Dolf Weijers

a. Present address: John Innes Centre, Department of Crop Genetics. Norwich, Norkfolk, United Kingdom

Abstract

Auxin-dependent gene regulation is mediated by the ARF family of DNA-binding transcription factors. The ARF family in *Arabidopsis thaliana* consists of 23 members that share a conserved topology of domains. Most ARFs are composed of three domains: an N-terminal DNA binding Domain (DBD), a C-terminal protein-protein interaction domain (III/IV) and a middle region (MR) that connects the N- and C-terminal domains. The DBD harbors the conserved B3 DNA binding motif and domains III and IV form the conserved PB1 protein-protein interaction domain, while no conserved motif has been described for the MR of any ARF. ARFs have different biological functions that are correlated with different protein properties. While diversification of the DBD contributes to recognition of different target genes, a key question is if and how the various domain contribute to (differences in) biological activity. In this chapter we address the question of how its (sub)domains contribute to the *in vivo* function and specificity of the ARF5 protein. Using DBD swaps between the closely related ARF5 and ARF6 proteins, we first establish that ARF DBDs are an important source of specificity in target gene activation, although not completely sufficient for ARF-specific function. We further show by generating a series of domain deletions that all ARF5 domains are critical for *in vivo* function of the protein. As protein interaction interfaces have been identified in all three ARFs domains, we take advantage of the modular nature of ARFs and use a proteomic approach to dissect the contribution of the individual domains to the formation of protein complexes. By IP-MS/MS we found that the isolated DBD can interact with active transcription machinery via the Mediator complex and the isolated CT can mediate interactions with other ARFs and Aux/IAAs. We also propose that the MR of (activator) ARFs are intrinsically disordered, and that a Tudor domain within the DBD is necessary for ARFs activity. Taken together, we demonstrate contributions of ARF5 domains to biological function and provide a proof of concept for dissecting the mechanistic contributions of each domain.

Introduction

Eukaryotic (plant) transcription factors are usually composed of a combination of independent protein domains or modules: DNA binding domain(s), an activation or repression domain, protein-protein interaction domain(s), etc. These modules are generally conserved and may be shared by transcription factors of the same or different families and can even be interchanged between transcription factors from different species[27]. The generation of new combinations of existing conserved domains has an important evolutionary role since it allows for increased phenotypic complexity without the necessity of the appearance of new protein domains[26]. This is the case for the Auxin Response transcription factor family (ARF). The ARF family in the flowering plant *Arabidopsis thaliana* consists of 23 members that share a particular architecture of domains. In general, ARFs are composed of three domains: an N-terminal DNA binding Domain (DBD), a C-terminal protein-protein interaction domain (III/IV) and a middle region (MR) that connects the N- and C-terminal domains.

In chapter 3, we have described the crystal structure of the DBD of two *Arabidopsis* ARFs, ARF1 and ARF5. The DBD of ARFs is highly conserved, and the residues that interact with the DNA backbone or bases in ARF1 are completely conserved within the family, with a few exceptions[126]. The DNA-interacting subdomain within the DBD is a B3 domain, which is a plant specific DNA binding motif found also in other plant transcription factors[22,181]. The B3 domain is connected to a dimerization domain that mediates ARF-ARF interactions needed for cooperative DNA binding to inverted repeats of auxin responsive elements (AuxREs)[126]. Examination of the DBD showed revealed to contain a Tudor-like domain. The Tudor domain has been described in animal proteins and has a role in the recognition and binding to methylated lysisnes and arginines, which are highly present in histones[182,183]. Whether the Tudor-like domain present in the DBD of ARFs shares this function of recognizing histone methylation is not known.

The C-terminal part of ARFs is formed by domains III and IV which fold towards each other to form a Phox/Bem1 (PB1) domain. PB1 is a protein module that can be found in animals, fungi, amoebas and plants[184]. A PB1 domain is about 80 amino acids long and it was first identified in the budding yeast protein Bem1p and in the mammalian protein p67[phox].[185] A PB1 domain can contain a cluster of acidic amino acids (OPCA motif), a cluster of positively charged (basic) residues, or both. PB1s can interact via these clusters in a head to tail manner where the acidic cluster of one interacts with the basic cluster of another allowing multiple PB1 domains to oligomerize[186]. Many specific protein-protein interactions necessary for signal transduction cascades are mediated by PB1 domains[187]. For example, in mammals PB1 domain containing proteins are involved in host-defense[188,189], cell polarization[190–192] and early cardiovascular development[193]. The structure of ARF PB1 domain was recently elucidated[174]. This domain is highly conserved within the ARF family and also within the Aux/IAA family[194,195]. This domain can form dimeric or oligomeric head-to-tail interactions to mediate protein-protein interactions: ARF-ARF or ARF-Aux/IAA[194]. The later has an important role in the auxin signaling pathway as Aux/IAA proteins repress ARF activity when bound[24,196,197]. Repression of ARFs by Aux/IAA proteins stops when, in presence of auxin, the Aux/IAAs are targeted for degradation[32].

The DBD and the PB1 domains in ARFs are connected by a Middle Region (MR). In contrast to the other domains, the MR shows very little conservation[30]. A correlation of the amino acid composition of the MR with ARFs activation or repression properties has been suggested[198]. In the MR of some ARFs with gene repression activity, the LxLxL motif can be found. The LxLxL motif is necessary for interaction with the TOPLESS protein, which in turn recruits Histone deacetylases (HDACs) to repress transcription [89]. Recently, it was described that the MR of ARF5 mediates interaction with chromatin remodelers of the SWI/SNF family [25]. This kind of interaction has been described for other transcription factors and would contribute to specific gene regulation by making genes available for transcription only in specific cell types or at defined developmental stages[199]. So far no structure of the ARF MR has been described or predicted. The contribution of the MR to ARF functioning is not yet fully understood, but as it is the more variable part, it may play a role in differentiating the biological functions of individual ARFs.

In the previous chapters we have described the biophysical properties by which the DNA binding domain of ARFs recognizes specific DNA sequences. Despite sharing high affinity for the same sequence *in vitro*, subtle differences may have an impact in target specificity *in vivo*. This includes preference for certain unique AuxRE motifs by the minor differences at residues surrounding the DNA-contacting amino acids, but also the intrinsic flexibility of each ARF DBD to accommodate to different distances within bipartite AuxREs. Although this mechanism may contribute to specific gene regulation, a critical question is whether the differences in the DBDs are enough to generate specificity and if the other domains contribute in the generation of specific gene regulation.

Transcription factors can increase their target specificity by interacting with other proteins. Above we describe that each ARF domain is capable of protein-protein interactions[200,201]. DBDs can mediate homodimerization; DBDs may also mediate interactions with histones via the Tudor-like domain; the MR of some ARFs facilitate interaction with TOPLESS for transcriptional repression and the MR of other ARFs may interact with chromatin remodelers to enable transcription; and finally domains III/IV mediate interactions with Aux/IAA proteins to regulate the transcriptional activity of ARFs. It is necessary also for specific transcription factors to be able to interact with the general transcription machinery, usually via the Mediator complex [202]. Furthermore, interactions of ARFs with other transcription factors to regulate specific biological processes have been described, mostly based on protein interactions found in yeast 2-hybrid screens: ARF8 interacts with the bHLH BIGPETALp through its C-terminal domain, and together they regulate petal growth[203]. ARF7 interacts with with MYB77 also through its C-terminal domain and together have a role in lateral root formation [204]. ARF6 interacts with the bHLH PIF4 and the brassinazole resistant transcription factor BZR1 through its MR and C-terminal domains and together they regulate hypocotyl elongation, integrating auxin, BR and environmental signals[176]. Furthermore, analysis of promoter region of auxin inducible genes showed that they are enriched in recognition elements for other transcription factor families: MYB related elements (MREs), bZIP response elements (ZREs) and G-box related elements (GREs)[205].

Protein-protein interaction of transcription factors play a very important role in the

determination of target specificity. Transcription factor complexes likely require specific combinations of cis-regulatory elements to be able to efficiently bind their targets. Complexes formed by transcription factors with chromatin remodelers will also contribute in the generation of specificity as they generate open chromatin environments making genes available for transcription only at specific stages of cell types. And finally, interactions with proteins that regulate the activity of the transcription factor itself would ultimately contribute to specific metabolic outputs.

In this chapter we address the question of how its (sub)domains contribute to the *in vivo* function and specificity of the ARF5 protein. Furthermore, we take advantage of the modular nature of ARFs and use a proteomic approach to dissect the contribution of the individual domains to the formation of protein complexes.

Results

Patterns of divergence in ARF protein domains

ARFs contain two conserved structural domains, B3 and PB1, separated by a poorly conserved MR. Diversity in ARFs originated by gene duplications, presumably during whole genome duplications, and posterior modifications that include some domain truncations (ARF3, ARF13, ARF17 and ARF23 lack the C-terminal PB1 domain)[34]. Given that ARFs in *Arabidopsis* have different functions, an important question is whether functional differences can be attributed to individual domains, or whether multiple domains contribute to divergence. To address this question, we first generated phylogenetic trees with either full-length ARF proteins sequences, as well as with those of each of the domains: DBD, MR and CT (Figure 1).

The N-terminal DBD approximately corresponds to the first 350 amino acids, and is the largest domain in the ARFs. There is a high overall similarity between ARF DBDs, likely due to structural constrains on sequence divergence, and the phylogenetic tree derived from DBD sequences largely resembles that of the full length proteins. Thus, overall ARF divergence is captured by sequence variation in the DBD. The other highly conserved part of ARFs is located in the C-terminus of the proteins (PB1) and is about 100 amino acids in size. Again, despite strong conservation, the variations in the C-terminus can also resolve the divergence of the full-length ARF sequences. It should be noted that some ARFs lack the C-terminus, which leads to some deviations in tree architecture.

Compared to the DBD and CT, the MR is more variable. This is also seen in the MR phylogenetic tree in which branches are much longer than in the other three trees, indicating high divergence between even closely related ARFs. As a consequence, the MR alone does not accurately reflect overall ARF divergence patterns.

Recently structures of the DBD[126] and the CT[174] domains of ARFs have been obtained, but so far the structure of the MR has been elusive. In order to obtain more information about the MR of ARF5, we used a secondary structure prediction server to determine what secondary
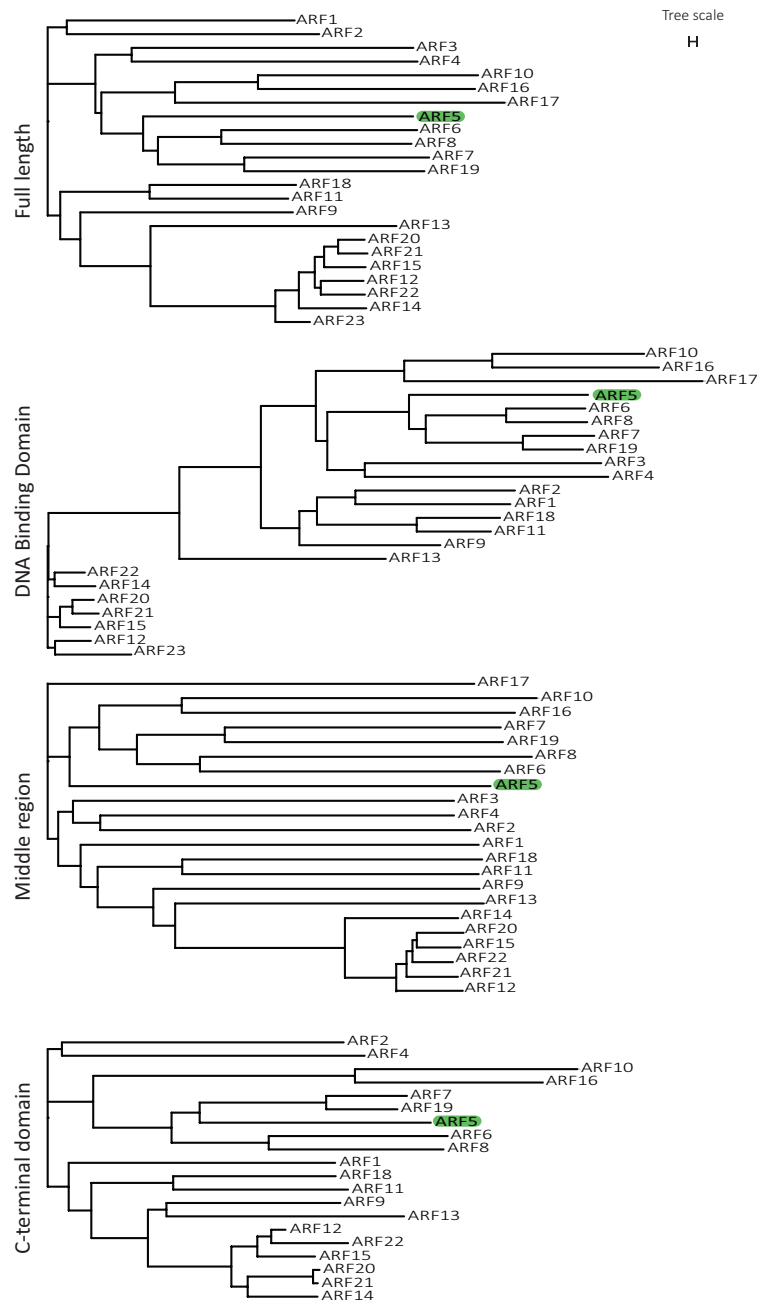
Figure 1. Phylogenetic tree of *Arabidopsis* ARFs' domains.
ARFs full length and independent domains were been clustered according to their amino acid sequence homology (ClustalOmega). It shows that the most conserved domain is the DNA binding domain followed by the C-terminal domain. The middle region is the main source of variability within the family.

109

Figure 2. ARF5 domains secondary structure prediction.
Predicted and known secondary structure elements are indicated above and below the sequence of ARF5, respectively. Helicesa re shown in blue, and strands in yellow. Note that elements were accurately predicted for the DNA binding domain and C-terminal region, but no significant structure was predicted for the Middle Region.

structure elements would be predicted based upon its sequence[206]. To infer the accuracy of the prediction, we separately analyzed the sequences corresponding to the DBD and CT of ARF5, for which crystal structures have been solved. The server accurately predicted most of the secondary structures of both DBD and CT domains (Figure 2). Despite its large size (378 amino acids), only one small alpha-helix was predicted for the MR. Assuming good accuracy of this prediction, it appears that either the MR is intrinsically disordered or that its structure is induced by interactions between the flanking domains, with the DNA or with other co-factors. We then subjected the sequences of full length ARFs to disorder prediction using the DISpro[207] server that utilizes the PONDR-FIT algorithm, which is used in the prediction of intrinsically disordered domains (IDPs). Interestingly we found that disorder of the MR seems to be a prominent characteristic of ARFs belonging to the class A (activators) (Figure 3) compared with other ARF classes. Strikingly, these patterns of

disorder are shared between the Arabidopsis ARFs and their co-orthologues in the liverwort *Marchantia polymorpha*[208] (Figure 3). It is worth noting the accurate prediction of structured regions (DBD and CT) by the server.
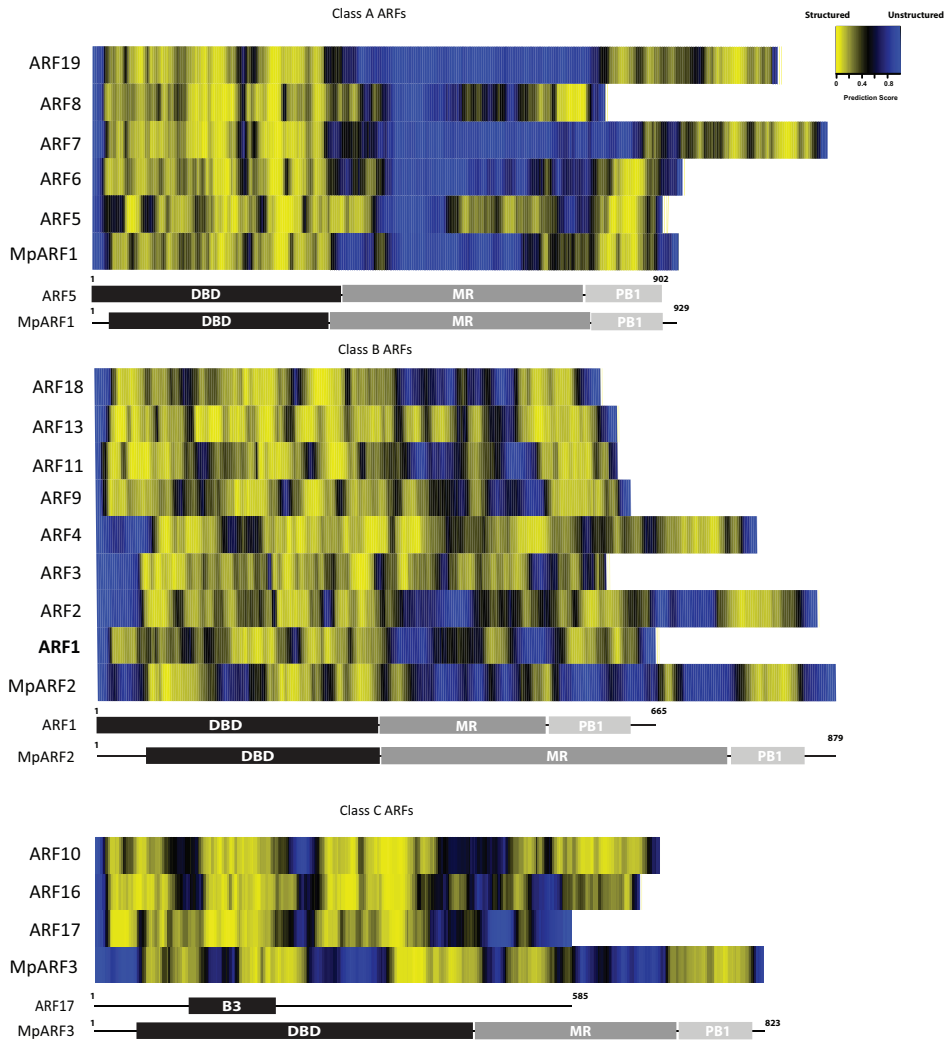
Figure 3. Protein disorder prediction of ARFs.
Analysis of disordered protein regions in *Arabidopsis* ARFs from clades A, B and C. An evolutionary ancestor of each clade (*Marchantia polymorpha*) was included in the predictions. Color key (yellow: structured to blue: unstructured) is shown in the top right. Predictions showed that the Middle regions of class A ARF might be intrinsically disordered.

In summary, two of the ARF domains (DBD and CT) show divergence typical of the entire proteins, while the MR does not. This likely correlates with structural constraints related to the biological activity of each domain.

Functional divergence of closely related ARF5 and ARF6 proteins

ARF5/MONOPTEROS is important for several aspects of *Arabidopsis* development. Its expression early during embryogenesis is required for correct root initiation[46,102]. The *arf5/mp* strong mutant allele B4149 fails in this process and as a result the plants are unable to form a root[99]. ARF5 is expressed early during embryogenesis with a very specific pattern[44]. ARF6 is one of the closest homologues to ARF5, and ARF6 is co-expressed with ARF5 during early embryogenesis. Furthermore, an *arf6* mutation can slightly enhance the weak phenotype of a partial loss of function allele *arf5-S319*[44], suggesting that ARF6 may partially substitute for reduced ARF5 function in a mutant background. Nevertheless, its presence in the *arf5* mutant is not sufficient to avoid the rootless phenotype. This indicates functional differences between the ARF5 and ARF6 genes or proteins. As such differences may be related to the pattern or level of expression of each gene, we designed a dedicated experiment to compare ARF5 and ARF6 proteins, or chimaeras between both. For this, we expressed cDNAs of either ARF5 or ARF6 under the control of ARF5 promoter in the *arf5-B4149* mutant background. The homozygous *arf5* mutant is unable to form a root. If the proteins we introduce are functional then we would expect to see a rescue of this strong phenotype. As it had not previously been excluded that elements in ARF introns are involved in gene regulation, we first confirmed that the expression of the ARF5 cDNA was able to fully complement the *arf5* mutant phenotype (Table 1). In contrast, the pARF5::ARF6 transgene was unable to restore root development in *arf5* mutants (Table 1), demonstrating that ARF5 and ARF6 proteins differ in function. We cannot exclude the possibility that proper expression is affected by the removal of regulatory elements present in the ARFs coding regions.

Table1. Complementation of *mp-B4149* rootless phenotype with cDNA wt ARF5, ARF6 and DBD domain swaps.
Constructs were the promoter of ARF5 was driving the cDNA of ARF5, ARF6, ARF5DBD/ARF6 and ARF6DBD/ARF5; were introduced into the strong *monopteros* mutant background. Rescue was tested by scoring the absence of rootless resistant seedlings in the T2 generation.

Construct	Number of lines	Embryonic rescue (%)		
		Full rescue	Partial rescue	No rescue
pARF5::ARF5	7	100		
pARF5::ARF6	8			100
pARF5::ARF5(DBD)/ARF6	8		37.5	62.5
pARF5::ARF6(DBD)/ARF5	6			100

To further explore what differences between ARF5 and ARF6 proteins underlie the distinct biological activity, we first focused on the DBD (Figure 4). Plotting the conservancy between ARF5 and ARF6 onto the ARF5 DBD structure reveals the high degree of similarity. All the DNA-contacting residues are completely conserved, as well as the dimerization interface. It is very unlikely that these two proteins would recognize different AuxREs as monomers, but divergence in the connecting loops may generate different binding affinities for complex motifs. To test if differences in the DBD contribute to differences in their biological activity, we expressed a chimaera of the ARF6 DBD and the remainder of ARF5 (ARF6DBD-ARF5), and found that this could not rescue the *arf5* mutant. Conversely, the ARF5DBD-ARF6 chimaera showed partial activity, which suggests that the ARF5 DBD is both necessary and partially sufficient for determining biological functionality of ARF5. Nonetheless, the MR and CT domains of ARF6 could not completely replace the ARF5 MR and CT domains, which shows that these domains also contribute to protein-specific functions.
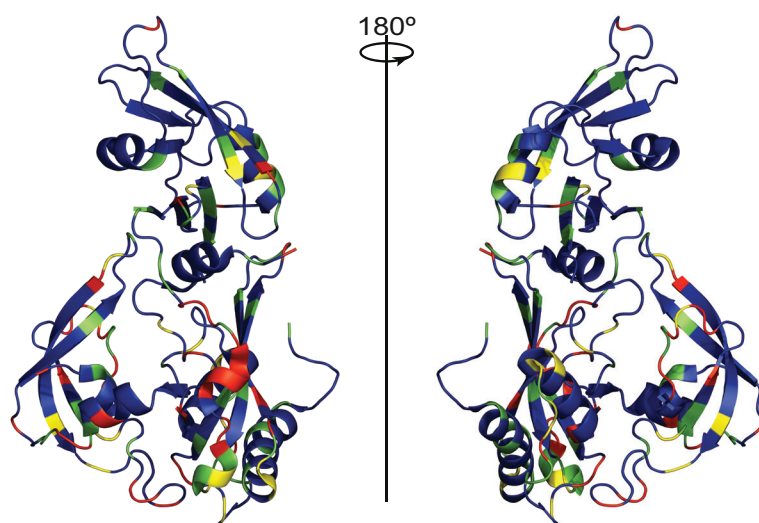
Figure 4. Sequence conservation between DNA-binding domains of ARF5 and ARF6.
Structure of the DNA binding domain of ARF5 where amino acids are colored according to their conservation between ARF5 and ARF6. Green = identical, blue = conserved, yellow = semi-conserved and red = nonconserved.

In summary, these results suggests that the subtle differences in ARF DBDs are indeed an important source of specificity in gene targeting. DBDs are necessary for ARF specific function but not sufficient, and the remainder of the protein contributes to final product of auxin induced transcriptional regulation.

Domain requirements for ARF5 function

While biochemical functions can be attributed to the various (sub) domains in ARFs, requirement for domains in the biological context has not been studied in detail. Here, we used the *arf5* mutant complementation assay to address the importance of individual domains (Table 2). We expressed N-terminal or C-terminal truncations (DBD alone; DBD/MR; MR alone; MR/CT; CT alone), as well as an internal deletion to remove the Tudor-like domain within the DBD in the *mp-B4149* background. Unfortunately, no transgenic plants could be obtained for the MR alone and MR/CT versions, likely due to strong dominant phenotypes. For all other truncations, rescue was tested in multiple independent transgenic lines. Strikingly, none of the deletions was able to complement the *arf5* mutant phenotype. Thus, each subdomain, including the Tudor-like domain is required for ARF5 function *in vivo*. Previously, it was shown that deletion of the CT from ARF5 caused a dominant auxin-independent phenotype, yet such mutants did complement the loss-of-function phenotype[209,210]. It is unclear why the results between these studies differ but the smallest of those truncations was 28 amino acids longer than our C-terminal truncations.

Table2. Complementation of *mp-B4149* rootless phenotype with ARF domains.
Constructs where the promoter of ARF5 was driving the cDNA of ARF5 full length (FL), DNA binding domain (DBD), middle region (MR), C-terminal domain (CT), DBD-MR and MR-CT; were introduced into the strong *monopteros* mutant background. Rescue was tested by scoring the absence of rootless resistant seedlings in the T2 generation.

Construct	Number of lines	Embryonic rescue (%)		
		Full rescue	Partial rescue	No rescue
pARF5::ARF5 FL	7	100		
pARF5::ARF5 DBD	2			100
pARF5::ARF5 MR	5			100
pARF5::ARF5 CT	0			
ARF5::ARF5 DBD/MR	5			100
ARF5::MR/CT	0			
ARF5::ARF5dT*	7			100

*Construct carrying the deletion of the Tudor domain was done in the genomic sequence while other constructs in the coding sequence.

We next expressed individual ARF5 domains as fusion proteins to Yellow Fluorescent Protein (YFP) to visualize the subcellular localization and accumulation level of these domains. To ensure sufficient amounts of protein, we expressed these fusions from the RPS5A promoter that mediates strong expression in meristematic cells[211]. The Full length and CT domain fusions were exclusively localized in the nuclei (Figure 5), consistent with earlier localization of ARF5-GFP expressed from its own promoter[44]. Fusions of the DBD and the DBD without
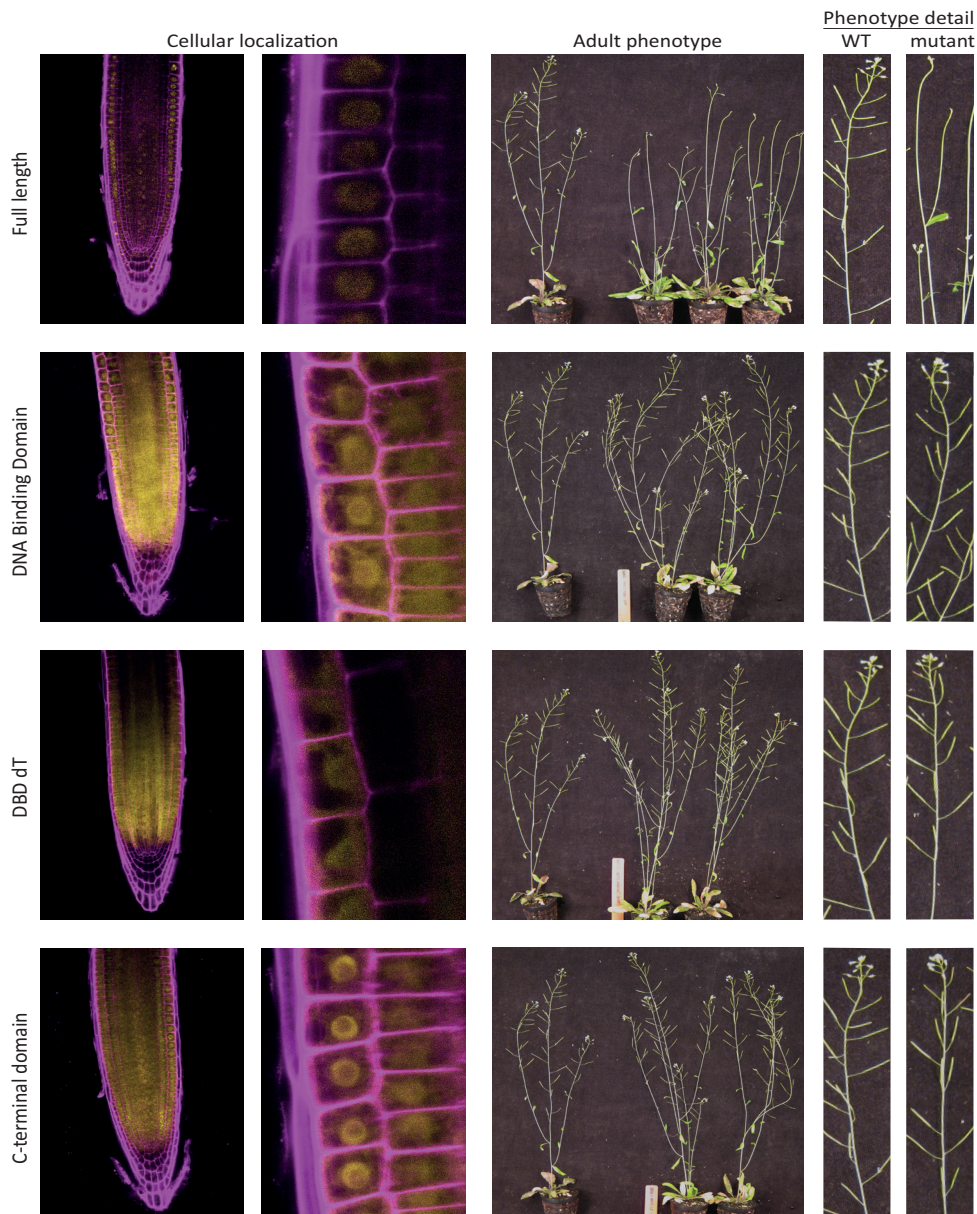
Figure 5: Localization and biological activity of ARF5 domains.
Localization of YFP fluorescence in root tips of transgenic lines expressing either full-length ARF5, of individual subdomains as YFP-fusions from the RPS5A promoter. Left panels show the entire root meristem, whiel the second column show a detail that highlights subcellular localization. Magenta signal are Propidium Iodide-stained cell walls. DBDdT and CT lines showed similar expression levels and confocal images were taken under identical settings. FL and DBD lines showed much lower expression levels and detector gain was increased to facilitate detection. Right panels show the adult phenotype of the same transgenic lines compared to a wild-type plant of the same age (left in each panel. Rightmost panels show details of inflorescences in each line.

the Tudor-like domain (DBDdT) were both localized in nuclei and cytosol (Figure 5). In contrast, the MR, as well as DBD-MR and MR-CT fragments, accumulated in vesicles (data not shown). Thus, nuclear localization of ARF5 can be conferred both by the DBD and the CT domain. The MR, even when connected to the DBD or CT, that by themselves localize to the nucleus, causes punctate vesicle-like localization, which can be interpreted as an artefact of expressing protein fragments.

Adult plants miss-expressing the full length ARF5 from the RPS5A promoter showed auxin-related phenotypes that affected flowering and seed production (Figure 5). This was also the case for lines misexpressing the DBD fragment. Similar to the Full-length ARF5 misexpression lines, the majority of those plants were infertile but few lines with very low expression of the fusion proteins developed normally. In contrast, plants expressing only the CT or the DBDdT showed no phenotype. This suggests that ARF5 misexpression phenotypes depend on the protein's ability to bind DNA. In addition, the Tudor-like domain in the DBD is required for biological activity in this misexpression experiment.

In summary, isolated ARF5 domains are unable to fulfill ARF5 biological activity, and all domains are necessary for its proper activity. Each domain confers certain characteristics to the full length protein. Misexpression of the Full length results in dominant phenotypes replicated by the misexpression of the DBD alone. In the first case phenotype may be a result of incapacity of the endogenous system to control the excess amount of ARF5. In the second case, as the DBD alone maintains its DNA binding capability, it simply cannot be control due to the lack of the ARF-Aux/IAA interaction interface. Furthermore, it seems that the Tudor-like domain present in the DBD is necessary for functional DNA binding. The self-standing CT domain cannot exert gene regulation and the MR seems to be incapable of proper folding without the flanking domains.

ARF5 domain contributions to *in vivo* protein-protein interactions

The functions of ARF (sub)domains involve DNA binding, but likely also the interaction with other transcription factors or the recruitment of other proteins to ARF targets. Some ARF interactors have previously been identified, mostly by *in vitro* or yeast assays[25,89,92,212,213], and the contribution of ARF domains to such interactions *in vivo* is not clear. To address this we used a proteomic approach to identify proteins interacting with each of the individual ARF domains. Based on subcellular localization patterns (Figure 5), we included the YFP fusions with Full-length ARF5, the DBD, the DBD without the Tudor-like domain and the CT domain. As each protein is tagged with the same YFP protein, and because biological activity of these fusion proteins was evident from overexpression phenotypes (Figure 5), we immunoprecipitated the ARF:YFP fusion proteins from seedling roots of pRPS5A::ARF:YFP lines. To allow quantitative and statistical analysis, triplicate immunoprecipitations were carried out, and in parallel triplicate immunoprecipitations were performed on non-transgenic wild-type control seedling roots. Proteins in complexes were identified and quantified using nano Liquid Chromatography followed by tandem Mass Spectrometry and analysis using

the MaxQuant software. This analysis allows to measure the fold-enrichment of proteins in immunoprecipitations compared to the control, as well as its statistical significance.

Peptides unique to ARF5 (Table 3) were enriched after IP for all of the samples. This is a good indication that immunoprecipitation worked correctly and that the ARF5 domains were correctly expressed as fusions to the YFP tags. In the FL sample, we identified peptides belonging to the DBD and the CT but none belonging to the MR. In the CT sample we identified one peptide that belongs to the DBD, suggesting that the CT alone had homodimerized with the endogenous ARF5 protein. In the DBD and DBDdT truncation, only peptides belonging to the respective DBD were found. Thus, homodimerization taking place through the DBD with endogenous ARF5 could not be demonstrated in this experiment. However, due to the strong overexpression phenotype in FL and DBD misexpression plants (Figure 5), only lines with relatively low expression levels could be generated, and these yielded few seeds. Thus, the amount of plant material collected for FL and DBD samples was considerably lower than for the CT and truncated DBD.

We next analyzed the proteins that co-precipitated with the ARF5 fusions to determine if domains were able to form meaningful interactions (Figure 6). Although the ARF5FL-YFP could be recovered from the sample, no significant interactions were found. ARF5FL was only enriched about 25 times compared to the wild-type background control. All the other samples were enriched more than a 1000-fold compared to their respective controls, and several significant interactors were recovered. The DBD recovered subunits of the Mediator complex. Given that Mediator is part of the machinery that connects specific transcription factors to the general transcription machinery[11,202], this finding indicates that the DBD domain alone is able to engage in active transcription complexes. Interactions of ARFs to subunits of the Mediator complex have been reported before[213]. Our data suggests that the DBD may be involved in mediating this interactions. Deletion of the Tudor-like domain from the DBD did not lead to loss of interaction with Mediator subunits, which shows that this sub-domain is dispensable for at least these interactions. The CT domain sample showed the highest number of significant interactions. In addition to endogenous ARF5 (see above), we found ARF2 and 3 different Aux/IAA proteins (IAA4, IAA16 and IAA8) to be highly enriched in this sample. ARF-ARF and ARF-Aux/IAA interactions have been previously reported by crystallography[174,195], and inferred from Yeast2-Hybrid[196] and protoplasts essays[174]. To our knowledge this is the first demonstration of *in vivo* ARF-Aux/IAA interactions through mass spectrometry. Our results, while optimization and scaling-up will be required to exhaustively identify ARF-domain interactors, demonstrate that individual domains can interact with co-factors without the need for the remaining ARF domains. Hence, these results serve as a proof of concept that individual isolated domains can be used in proteomic experiments to gain an idea of interactions proper of their role within their full length context.

In summary, proteomics of isolated ARF domains is feasible and gives us information about the contribution of each domain to the transcriptional process mediated by ARFs. DNA binding domains seem to be in charge of interacting with the general transcription machinery via the Mediator complex. Meanwhile the CT domain mediate interactions, with

Table 3. Unique peptides of ARF5 identified in by IP/MS-MS per domain pulldown.
Peptides identified per sample are marked with an X.

Identified Peptides		Sample			
Domain	Sequence	DBD	CT	FL	DBDdT
DBD	MMASLSCVEDK	X			X
DBD	TSCLVNGGGTITTTTSQSTLLEEMK	X			X
DBD	DSDEIYAQMSLQPVHSER	X			X
DBD	DVFPVPDFGMLR	X			X
DBD	GSKHPTEFFCK	X		X	X
DBD	HPTEFFCK	X			X
DBD	TLTASDTSTHGGFSVPRR	X			
DBD	TLTASDTSTHGGFSVPR	X	X	X	X
DBD	AAEKLFPPLDYSAQPPTQELVVR	X			X
DBD	LFPPLDYSAQPPTQELVVR	X		X	X
DBD	DLHENTWTFR	X			X
DBD	RHLLTTGWSLFVGSK	X			X
DBD	HLLTTGWSLFVGSKR	X			X
DBD	HLLTTGWSLFVGSK	X			X
DBD	LRAGDSVLFIRDEK	X			X
DBD	AGDSVLFIR				X
DBD	AGDSVLFIRDEK	X		X	X
DBD	SQLMVGVR	X		X	X
DBD	QQTALPSSVLSADSMHIGVLAAAAHATANR	X			X
DBD	TPFLIFYNPR	X		X	X
DBD	ACPAEFVIPLAKYR				X
DBD	ACPAEFVIPLAK	X			
DBD	KAICGSQLSVGMR	X			
DBD	AICGSQLSVGMR	X			
DBD	FGMMFETEDSGKR	X			
DBD	FGMMFETEDSGK	X			
DBD	RYMGTIVGISDLDPLR	X			
DBD	YMGTIVGISDLDPLRWPGSK	X			
DBD	YMGTIVGISDLDPLR	X		X	
DBD	NLQVEWDEPGCNDK	X			
DBD	NLQVEWDEPGCNDKPTR	X		X	
CT	TYTKVQKTGSVGR		X		
CT	SIDVTSFKDYEELK		X		
CT	SIDVTSFK		X		
CT	LVYVDYESDVLLVGDDPWEEFVGCVR		X		
CT	ILSPTEVQQMSEEGMK		X		
CT	LLNSAGINDLK		X	X	
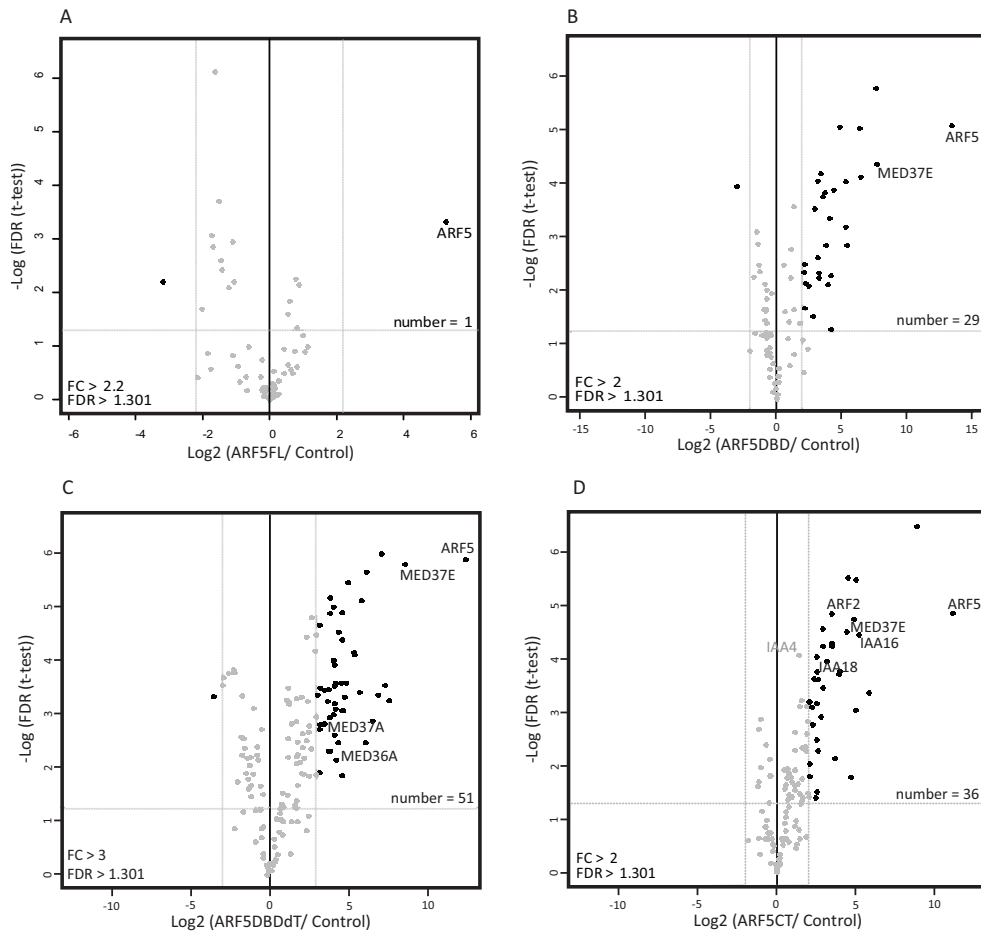
Figure 6. Co-immunoprecipitated proteins identified by MS-MS.
Significantly enriched proteins in immunoprecipitates of the different transgenic lines pRPS5A-ARF5(FL)-YFP (A), pRPS5A-ARF(DBD)-YFP (B), pRPS5A-ARF5(DBD-delta Tudor)-YFP (C) and pRPS5A-ARF5(CT)-YFP (D) are shown in the right upper quadrant of each Vulcano plot. The X axis indicates enrichment of proteins in triplicates of the transgenic line compared to triplicates of Control sample (Col-0 wild-type). The Y axis indicates the p-value for significance of enrichment in a Student's t-test. For all samples, YFP=tagged proteins were successfully recovered. FC= fold change, FDR=false discovery rate, number= significant hits

other ARFs and Aux/IAAs, important for the regulation of ARF activity within the auxin signaling pathway

119

Discussion

The small plant molecule auxin is capable of generating a wide variety of responses in plant cells. The same molecular signal can trigger many different biological processes all through the life of a plant (embryogenesis[214], vascular development[215], root growth[216], lateral root formation[217], tropic responses[218], leaf and flower initiation[70], etc.). It does so by triggering a cascade of events known as the nuclear auxin signaling pathway where three main families of proteins are involved. The ultimate effectors of this cascade are the Auxin Response Factors (ARFs). In *Arabidopsis* this family consists of 23 members that regulate specific genes to drive auxin-dependent development. In the previous chapters we explored the DNA sequence requirements for target genes to be recognized by ARFs. So far results indicate that most ARFs recognize the same DNA sequence (AuxRE), but structural constrains inherent to each ARF may be responsible to give different binding affinities for complex (bipartite) AuxREs, and thus contribute to target specificity. In this work we address the question of what other mechanisms ARF members use to generate/increase specificity. For this purpose we explored properties of ARFs beyond the DNA binding domain. ARFs are modular proteins. This means they are composed by independent modules that are conserved and found in other proteins as well. The general architecture of ARFs is: **DBD**[B3]-**MR-CT**[PB1]. We analyze the divergence of these domains between *Arabidopsis* ARFs. As the domains co-evolved in each ARF, the general clustering of ARF clades was kept for all the domains but the length of the branches showed that while the DBD and CT where highly conserved, the MR was extremely variable within the family. How the variations of each domain contributes to the variation in ARF functionality is not fully known. Furthermore, protein structure of the conserved DBD[126] and CT[174] domains have been recently published while so far it seems that the MR lacks one. Studies in the origin and evolution of the ARF family may serve to correlate sequence divergences to the appearance of new biological functions.

To determine the extent by which the DBD confers specificity, DBDs were swapped between ARF5 and ARF6 in the context of the rootless *arf5* mutant. Based on the enhanced rootless phenotype of a weak *arf5* allele by the *arf6* mutation, it was previously shown that ARF6 may have some redundancy with ARF5[44], but here we tested if ARF6 can replace ARF5 when expressed in its place. To facilitate the interpretation of swap experiments, cDNAs were used for ARF5 and ARF6. This however leads to the risk of eliminating regulatory elements present in intronic regions that may be necessary for gene expression. The impact of intronic regulatory elements in expression has been described for other plant transcription factors and it may also be important for ARFs[219,220]. ARF5 cDNA was indeed sufficiently expressed to rescue the rootless phenotype of the *arf5* mutant. Plants expressing ARF6 cDNA did not rescue the rootless phenotype and we interpret this as the inability of ARF6 to regulate ARF5 targets. Furthermore, a chimera that included the ARF5 DBD and the MR and CT of ARF6, but not the inverse chimera, could partially rescue the rootless phenotype. Thus, the ARF5 DBD is necessary for specific gene regulation but not entirely sufficient. ARF5 DBD needs more than a generic ARF activation domain to regulate its target genes. Likely, ARF5 MR and CT domains engage in protein interactions that contribute to ARF5's unique activity. It should be noted however, that our experiments cannot exclude the possibility that

the coding sequence of ARF5 contains cis-regulatory elements (enhancers) required for gene expression. Replacing parts of the ARF5 cDNA by corresponding parts from the ARF6 cDNA may therefore not only replace protein domains, but also remove cis-regulatory elements causing chimera's to be expressed but inadequately regulated[221,222].

Strategies used by plant transcription factors to increase target specificity involve working cooperatively with other proteins[23,223–225]. Such proteins can also be transcription factors of the same or different families, they can be part of the general transcriptional machinery, or they can be involved in modifying chromatin states. We asked if ARFs engage in such protein-protein interactions, and if they do so, what part(s) of the protein mediates the interaction. To tackle this question, we used a proteomic approach for full-length ARF5 as well as for its isolated domains. ARF5 Full length, DBD, DBDdT and CT were properly expressed and localized to the nuclei; fragments containing the MR seemed to aggregate and were seen as punctuated bodies dispersed in the cytoplasm (data not shown). This may be due to the lack of intrinsic structure of this domain. It is possible that the lack of order in this domain contributes to functionality as seen in other eukaryotic proteins involved in cell signaling and transcription regulation. Disordered domains have been proposed to allow the transcription factor to fold in many ways, exposing different binding interfaces and thus giving them the ability to interact with multiple different partners[226–229]. Selective change of interaction partners may explain how ARF5 (and other activator ARFs) is involved in the regulation of multiple biological processes. Another possibility is that the MR structure is induced by conformational changes of the whole protein upon DNA interaction[230–232], or upon interaction between the flanking DBD and CT domains. In either case, misexpression of the MR alone, or in conformations that do not protect the MR from spurious interactions may lead to aggregation. Indeed, the FL, DBD and CT fragments fold and localize properly while the MR does not. The DBD-MR and the MR-CT fragments also aggregated inside the cells. A 28 amino acids longer DBD-MR fragment[209] has been shown to be functional. It is possible that those amino acids are involved in interacting with amino acids in the DBD to stabilize the conformation to the protein and prevent spurious interactions.

Besides the proper localization of the domains, the misexpression of the Full length and DBD induced auxin-related phenotypes that affected flowering and thus seed production, and thereby limiting the amount of plant material for the analysis. One solution for this problem will be to induce the nuclear localization just before sample collection, for example by adding a rat Glucocorticoid Receptor (GR) domain[233,234] to the fusion proteins. In this way, lines expressing higher levels of protein could be used, eliminating the phenotypes caused by prolonged unregulated ARF accumulation in the nuclei.

Despite the need for further optimization of domain-specific protein interaction analysis, several conclusions could be drawn:

Firstly, ARF domains can indeed be studied separately, as they seem to interact with expected proteins in the absence of the other domains. Unfortunately, interactions with other transcription factors were not observed, although they have previously been reported for ARF5[213] and for other ARFs[176,203,204]. Interactions with other transcription factors were

neither observed for individual domains nor for the full length protein. An explanation could be that the low amount of full length protein that could be recovered limited identification of low-abundant interacting proteins. If ARFs are part of multiple complexes, or when not all ARF protein molecules engage in such complexes at all times, all interactors are expected to be sub-stoichiometric. The same could be true for the DBD sample, but an alternative explanation could be that the DBD alone cannot mediate interactions with transcription factors.

In chapter 3 we showed dimerization of ARFs via the DNA binding domain. Unfortunately we could not demonstrate binding of the DBD to endogenous ARFs as we only identify peptides belonging to the same AF5 DBD fragments used for the pull down. Again it is also possible that the endogenous levels or ARF5 are too low to be detected in these assays. The DBD could be shown to engage in interactions with subunits of the Mediator complex. This kind of interactions have been reported for Full-length ARFs and it seems now that the DBD is the domain mediating this important interaction. Mediator complex forms a bridge between active transcription factors and the general transcriptional machinery, and therefore it is common to see TFs interacting with subunits of the complex. In our case it is a confirmation of the functionality of the isolated ARF DBD. It has been hypothesized that the Mediator complex may not be so "general" in its function and that its association with transcription factors may be specific and in response to environmental stimuli[235,236]. Because specific Mediator subunits are involved in the regulation of specific biological processes in plants (MED25 and MED8 in cell elongation and cell wall composition[237], MED12 and MED13 in *Arabidopsis* pattern formation[237], MED15 in lipid metabolism[239]), dissection of specific ARF-Mediator subunits may provide information over specific gene regulatory events by this complexes.

Another kind of interaction that would mediate ARF transcriptional activity is the ones it holds with Aux/IAAs via the PB1 domain in the CT. We could confirm this interactions taken place with our proteomic approach, and furthermore that the isolated ARF5 CT is able to mediate these interactions without the need of the remainder of the protein. There have been studies about the pairing affinity of ARFs with specific members of the Aux/IAA family[99,212]. Our approach could be used to recognize these differences in an *in planta* context with the advantage of not generating phenotypic distress caused by the misexpression of either Full length ARFs or Aux/IAA proteins. The Aux/IAA proteins recovered are IAA4, IAA6 and IAA18. It should be noted that pull-downs were done in root tips of misexpression lines, and these 3 Aux/IAA proteins might not be the most abundant in the tissue. Based on expression profiles, other Aux/IAAs are at least equally or more abundant[212,240]. Thus, a specific set of Aux/IAA proteins was recovered by the CT domain of ARF5. Interestingly gain of function mutations in IAA4, IAA6 and IAA18 show defects similar to loss of ARF5 function. IAA4 mutants are rootless (Yunde Zhao, personal communication), IAA6 mutants have short hypocotyls and upcurled leaves[241,240], and the iaa18 mutation causes cotyledon[100] and root[242] defects similar to those in *arf5* mutants[243]. Thus, it is possible that the pull-down experiment reveals preferential interactions between Aux/IAA proteins and ARF5.

The last kind of possible interactions that could contribute to specificity of gene regulation

are the ones with chromatin remodelers. Unfortunately we did not identify such proteins in any of the pull-downs. It has been reported that the SWI/ SNF chromatin remodeling proteins do interact with ARF5 and they seem to do so via the MR[25]. We could not test ARF5 fragments containing the MR, and redesign of the limits of the fragment should be done. To further explore the role of ARF5 in chromatin remodeling we focused on the Tudor-like domain described within the DBD in chapter 3. The Tudor domain is involved in protein-protein interactions where they recognize methylated arginines. Methylated arginines are common in histones and thus Tudor domains have been involved with DNA-templated biological processes[182,183]. It is possible though that the Tudor domain facilitates the location of ARFs in the histones that need to be "unfolded" for genes to be transcribed. A deletion in this domain seems to impair the proper regulatory functions of ARF5 *in vivo* but the proteomic analysis revealed that it does not impair the capacity of the DBD to interact with the transcriptional machinery. Rescue experiments using the genomic sequence of ARF5 where the Tudor-like domain was deleted showed that this protein was not functional. A possibility is that indeed the Tudor domain is required for transcriptional regulation mediated by ARF5, in particular by targeting ARF5 to chromatin domains enriched in histones with methylated Arginines, but we cannot exclude the possibility that the deletion impaired the proper expression or folding of the protein. Nevertheless there is evidence of the role of ARFs in chromatin remodeling and it deserves further studies.

To conclude, in this work we have presented evidence that modularity of ARFs can be used to dissect contribution of ARF domains to specific gene regulation, and that divergence in the DBD sequences alone may not be sufficient to account for specific gene selection. The contribution of the MR remains an open question, and being this the main source of divergence between ARFs should be further characterized.

Material and Methods

Sequence Alignments

Protein sequences were retrieved from the UniProt database and alignments and phylogeneitcs trees were generated using the ClustalOmega software.

Secondary structure predictions

Fasta files of the independent domains were uploaded into the Jpred4 server[206].

For protein disorder prediction, Fasta files of the full length *Arabidopsis* and *Marchanthia polymorpha* ARFs were uploaded into the DISpro server[207].

Cloning

All fragments were cloned using the LIC system described in De Rybel et al [152]. Primers used are listed in Supplemental Table 1. For rescue experiments, cDNA fragments were cloned

into the LIC vector pMP::LIC-tNOS, except for the deletion of Tudor domain in the Full length context were the genomic sequence was used. For IP experiments cDNA fragments were cloned into the LIC vector PRPS5A::LIC-sYFp-tNOS.

Plant material

For rescue experiments constructs were introduced in the *mpB4149* background[46]. For IP experiments, constructs were introduced in the Columbia-0 ecotype of *Arabidopsis thaliana*. T3 homozygous seeds were surface-sterilized in 25% bleach/75% ethanol for 10 minutes, washed twice with 70% ethanol and once with 100% ethanol. When dried, seeds were plated on half-strength Murashigues-Skoog (MS) medium, vernalized for 2 days at 4 degrees C and grown for 5 days under log-day conditions at 22 degrees C.

Confocal microscopy

Roots of 5-day old seedlings expressing domain-sYFP were stained with 10ug/ml propidium iodide (PI). sYFP was excited with 514nm. Imaging was performed on a Leica SP5 II system equipped with Hybrid detectors.

IP/MS-MS

IP/MS-MS was performed as described in Wendrich *et al* [244]. Transgenic plants were grown on MS media for 5 days, long day conditions. Root tips were harvested. 0.69 g of roots were used for DBD and FL samples and 1.8 g for DBDdT and CT samples.

Acknowledgements

Supplementary information

Supplemental Table 1: Primers

For Domain Swaps

Gene	Primer orientation	Sequence 5'- 3'
ARF5FL	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	AGTATGGAGTTGGGTTCTTATGAAACAGAAGTC
ARF6FL	Sense	TAGTTGGAATAGGTTCATGAGATTATCTTCAGCTGG
	Antisense	AGTATGGAGTTGGGTTCCTAGTAGTTGAATGAACCCC
ARF5 DBD	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	CATAGGGAATGTTGTTAAAGGTGTTTCGATATCCC
ARF5 MR-CT	Sense	CATTGTGGGAGATTGAGCCTGAAAGTCTCTTCATTTTTCC
	Antisense	AGTATGGAGTTGGGTTCTTATGAAACAGAAGTC
ARF6 DBD	Sense	TAGTTGGAATAGGTTCATGAGATTATCTTCAGCTGG
	Antisense	GGAAAAATGAAGAGACTTTCAGGCTCAATCTCCCACAATG
ARF6 MR-CT	Sense	GGGATATCGAAACACCTTTAACAACATTCCCTATG
	Antisense	AGTATGGAGTTGGGTTCCTAGTAGTTGAATGAACCCC

For Rescue

Gene	Primer orientation	Sequence 5'- 3'
ARF5FL	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	AGTATGGAGTTGGGTTCTTATGAAACAGAAGTC
ARF5 DBD	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	AGTATGGAGTTGGGTTCTTAAGGTGTTTCGATATCCCATGGACTGAC
ARF5 MR	Sense	TAGTTGGAATAGGTTCATGGAAAGTCTCTTCATTTTTCCTTCTCT-GACCTCA
	Antisense	AGTATGGAGTTGGGTTCTTAAACATTGCTTGAAGATGTACCAGTGCC
ARF5 CT	Sense	TAGTTGGAATAGGTTCATGGATTTTGATGATTGTAGTCTGCG-GCAAAAT
	Antisense	AGTATGGAGTTGGGTTCTTATGAAACAGAAGTC
ARF5 DBD-MR	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	AGTATGGAGTTGGGTTCTTAAACATTGCTTGAAGATGTACCAGTGCC
ARF5 MR-CT	Sense	TAGTTGGAATAGGTTCATGGAAAGTCTCTTCATTTTTCCTTCTCT-GACCTCA
	Antisense	AGTATGGAGTTGGGTTCTTATGAAACAGAAGTC

For IP

Gene	Primer orientation	Sequence 5'- 3'
ARF5FL	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	AGTATGGAGTTGGGTTCCCTGAAACAGAAGTCTTAAGATCGTTAAT-GCCTG
ARF5 DBD	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	AGTATGGAGTTGGGTTCCCAGGTGTTTCGATATCCCATGGACTGAC
ARF5 DBDdT	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	AGTATGGAGTTGGGTTCCCGAGCTGAGACCCGCATATCG
ARF5 MR	Sense	TAGTTGGAATAGGTTCATGGAAAGTCTCTTCATTTTTCCTTCTCT-GACCTCA
	Antisense	AGTATGGAGTTGGGTTCCCAACATTGCTTGAAGATGTACCAGTGCC
ARF5 CT	Sense	TAGTTGGAATAGGTTCATGGATTTTGATGATTGTAGTCTGCG-GCAAAAT
	Antisense	AGTATGGAGTTGGGTTCCCTGAAACAGAAGTCTTAAGATCGTTAAT-GCCTG
ARF5 DBD-MR	Sense	TAGTTGGAATAGGTTCATGATGGCTTCATTGTCTTGTGTTGAAGAC
	Antisense	AGTATGGAGTTGGGTTCCCAACATTGCTTGAAGATGTACCAGTGCC
ARF5 MR-CT	Sense	TAGTTGGAATAGGTTCATGGAAAGTCTCTTCATTTTTCCTTCTCT-GACCTCA
	Antisense	AGTATGGAGTTGGGTTCCCTGAAACAGAAGTCTTAAGATCGTTAAT-GCCTG

CHAPTER 6

FRET-FLIM for Visualizing and Quantifying Protein Interactions in Life Plant Cells

Methods in Molecular Biology (in press)

Alejandra Freire Rios
Tatyana Radoeva
Bert De Rybel
Dolf Weijers
Jan Willem Borst

Abstract

Proteins are the workhorses that control most biological processes in living cells. Although proteins can accomplish their functions independently, the vast majority of functions require proteins to interact with other proteins or biomacromolecules. Protein interactions can be investigated through biochemical assays such as co-immunoprecipitation (co-IP) or Western blot analysis, but such assays lack spatial information. Here we describe a well-developed imaging method, Förster resonance energy transfer (FRET) analyzed by Fluorescence lifetime imaging microscopy (FLIM), that can be used to visualize protein interactions with both spatial and temporal resolution in live cells. We demonstrate its use in plant developmental research by visualizing *in vivo* dimerization of AUXIN RESPONSE FACTOR (ARF) proteins, mediating auxin responses.

1. Introduction

In the mid-nineties, commercial confocal microscopes were introduced, which has revolutionized the investigation of biomolecules in living cells [245]. Visualizing biological processes has made great progress due to significant advances in instrument and detector design as well as the introduction of genetically encoded fluorescent proteins *in vivo*. Although confocal microscopy provides very-high-quality multi colored images, spatial resolution is limited by the diffraction of light [246]. Since 2006, several fluorescence imaging techniques have been developed known as super-resolution imaging technology, which provide fine detailed and high contrast images create images down up to 20 nm spatial resolution [247,248]. However, these imaging techniques are still unable to visualize protein interactions in vivo [249,250]. Förster resonance energy transfer (FRET) is a phenomenon of radiation-less energy transfer from a fluorescent donor molecule to an acceptor molecule (either fluorescent or not) through dipole-dipole coupling. This process only takes place when the donor and acceptor molecules are in very close proximity (typically 2-10 nm) [251]. The energy transfer rate is inversely proportional to the sixth power of the distance between donor and acceptor, making FRET very sensitive to small changes in distance and useful as a molecular ruler [252] (See Fig. 1). In order to accomplish FRET, the fluorescent donor and acceptor molecules have to fulfill several prerequisites; 1) a spectral overlap between donor emission and acceptor absorption spectra, 2) close proximity between donor and acceptor (<10 nm) and 3) finite dipole orientation factor [251]. FRET can be quantified using different imaging methods of which fluorescence lifetime imaging microscopy (FLIM) is the most robust and straightforward approach [253]. A fluorescence lifetime can be defined as the average time a fluorophore remains in the excited state before returning to the ground state. The fluorescence lifetime is an intrinsic property of a fluorophore and is independent of concentration, absorption by the sample, sample thickness, photobleaching and laser excitation intensity [254]. However, the fluorescence lifetime is very sensitive to the microenvironment of the fluorophore such as pH, ion or oxygen concentration, molecular binding or the proximity of energy acceptors [255], which makes this method unique for FRET analysis. FRET-FLIM measurements are based on determination of the fluorescence lifetime of the donor molecule. The close proximity of acceptor molecules creates an additional relaxation path of the donor-excited fluorophores resulting in a decreased donor fluorescence lifetime [253] (see Fig. 1). The amount of donor fluorescence lifetime reduction is directly correlated with the FRET efficiency (E) via $E = (1 - t_{DA}/t_D)$ where t_{DA} is the fluorescence lifetime of the donor in the presence of acceptor and t_D is the fluorescence lifetime of the donor alone. In order to obtain a comprehensive overview and critical essentials concerning the fundamentals of FRET and the methods of measurements, we refer to a detailed description [256].

In this chapter, a detailed description of FRET-FLIM technique and analysis will be presented by showing recent data on the dimerization of the transcription factor family AUXIN RESPONSE FACTORs (ARFs) [126] in living plant cells. ARFs act dependent on the signalling molecule auxin, which is a small organic molecule important for cell division, elongation and differentiation [119]. Specificity in auxin responses is generated through the DNA-binding domain (DBD) of ARFs. Crystal structures of the ARF5 and ARF1 DBDs
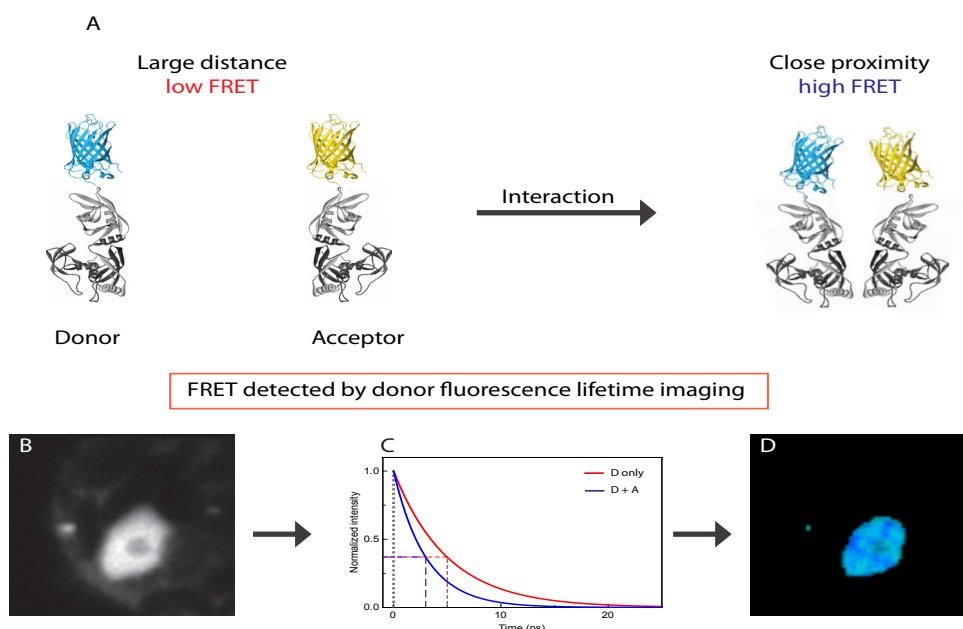
Figure 1: Illustration of FRET principle and FRET analysis.
A) Two DNA binding domains of ARF5 are coupled to CFP (donor) and YFP (acceptor). Large distance between CFP and YFP results in low FRET whereas high FRET signals are obtained if ARF domains are in close proximity and correct orientation. FRET can be quantified by fluorescence lifetime analysis. A fluorescence intensity image.
(B) shows the nuclear expression of the donor ARF5-sCFP3A in a protoplast. Per pixel the time resolved photon distribution is plotted
(C) of which the fluorescence lifetime is calculated and shown as a false color-coded fluorescence lifetime image
(D). Interaction of a donor with an acceptor results in a decrease of the fluorescence lifetime (C).

were solved showing homodimerization of DBDs generating cooperative DNA binding, which is critical for *in vivo* ARF5 function [126]. ARF5 dimerization was also demonstrated in *Arabidopsis* protoplasts using a FRET-FLIM analysis [126]. In this assay ARF5-sCFP3A and ARF5-sYFP2 fusion proteins were expressed and donor fluorescence lifetime analysis revealed specificity in homodimerization of ARF5 proteins. Full-length ARF5 homodimerization revealed a decrease of donor fluorescence lifetime resulting in a FRET efficiency of 10% (see section 3) [126], whereas mutations within the dimerization domain of the DBD of ARF5 resulted in FRET efficiencies around background level. These FRET data excluded the possibility of ARF DBD dimerization being an artifact of the crystallization process but confirmed that these interactions take place *in vivo*. Furthermore, site directed mutations and subsequent phenotypic analysis shed light on the importance of dimerization for the ARF function *in vivo*.

Here we present a stepwise protocol to perform FRET-FLIM analysis and show data on

ARF5 homodimerization as an example. Protein expression in plant protoplasts, FRET-FLIM analysis, quantification of FRET efficiencies, use of proper controls and different methods to display protein interactions will be discussed.

2. Materials

2.1. Cloning of Transient Expression Vectors

1. Plant expression vector: LIC vectors based on pMON999 (Monsanto, USA) [152].

2. The cDNA of ARF5 was cloned into LIC vectors harboring either SCFP3A fluorescent protein (sCFP3A) or Yellow Fluorescent Protein (sYFP2) (see Note 1).

3. Constructs were generated using Ligase Independent Cloning procedure (see Note 2).

2.2. Protoplast Isolation - *Protocol adapted from references [257] and [258]*

1. Plant material: Rosette leaves of *Arabidopsis thaliana* plants (ecotype Columbia) grown for ± 3 weeks under long-day conditions (16 h light/8 h dark) at ± 22°C.

2. Mannitol solution: 0.4 M mannitol, 20 mM KCl, and 20 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 5.7 in Milli-Q water.

3. Calcium chloride solution: 1 M $CaCl_2$ in Milli-Q water.

4. Enzyme solution: 1% (w/v) cellulose "Onozuka" R10 (Serva Electrophoresis, GmbH, Germany) and 0.2% (w/v) Macerozyme R10 from Rhizopus sp. (Serva Electrophoresis, GmbH, Germany) dissolved in mannitol solution; subsequent addition of $CaCl_2$ to a final concentration of 10 mM.

5. Plastic round-bottom tubes.

2.3 Transfection - *Protocol adapted from [257] and [258]*

1. PEG/Ca^{2+} solution: 40% (w/v) Polyethyleneglycol 4000 (PEG) 0.2 M mannitol, and 100 mM $Ca(NO_3)_2$ in Milli-Q water (see Note 3).

2. W5 solution: 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, and 2 mM MES pH 5.7.

3. W5/Glucose solution: W5 solution containing 1 mM glucose.

4. MMg solution: 0.2 M mannitol, 15 mM $MgCl_2$, and 4 mM MES pH 5.7.

5. Plastic round-bottom tubes.

6. Microscope 8-well slides

2.4. Confocal imaging and FLIM

1. Leica SP5X (see Note 4).

2. B&H SPC 730/830 module (Becker & Hickl, Germany).

3. Method

3.1 Protoplast Isolation

1. All steps are carried out at room temperature unless otherwise stated.

2. Rosette leaves of ±3 week old *A. thaliana* plants are fixed on Time tape adhered to the upper epidermis [259].

3. 3M Magic tape is fixed to the lower epidermis to make a "Tape-*Arabidopsis* Sandwich" [259] (see Note 5).

4. A clean Petri dish (9 cm diameter) is covered with 3 stripes of tape covered with leaves sufficient for 6 independent transfections.

3. Add 25 mL of enzyme solution, and swirl to dampen all plant material.

4. The Petri dish is transferred on a platform shaker (shaking ± 60 rpm) and incubated at 27°C for 20 minutes (Note 6).

5. Protoplasts are released from the leaf matrix by carefully pipetting the enzyme solution on the taped leaves.

6. Transfer the protoplasts to a clean plastic round-bottom tube.

7. Protoplasts are collected by centrifugation for 3 min at 50×g using a tabletop centrifuge; wash once with 5 mL of W5 solution.

3.2. Protoplast Transfection

1. All steps are carried out at room temperature.

2. Pipette plasmid DNAs in round bottom tubes; for single transfections 10–20 μg of the respective DNA is required and for double transfections, 10–15 μg of each construct. Single transfection is used to obtain protoplasts expressing the donor construct only, whereas the transfections of two plasmids yield protoplasts for the interaction studies.

3. Collect the isolated and washed protoplasts (see Section 3.1) by centrifugation for

135

3 min at 50×g in a tabletop centrifuge and remove the supernatant carefully with a pipette.

4. Resuspend the protoplasts in 1.2 mL of MMg (see Note 6) and transfer 200 μL aliquots to round bottom tubes that contain plasmid DNA already (see Note 7).

5. Add 220 μL of PEG/Ca^{2+} solution to the protoplasts and mix well but carefully and incubate for 5 min (see Note 8).

6. Stop the transfection by addition of 800 μL of W5 solution to the protoplast solution and collect the protoplasts by centrifugation at 50×g for 90 seconds in a tabletop centrifuge.

7. Remove the supernatant with a pipette and wash with 5 mL of W5 solution.

8. Collect protoplasts by centrifugation at 50×g for 3 min and remove the supernatant with a pipette and resuspend the protoplasts in 1 mL of W5 solution containing 1 mM glucose.

9. Protoplasts can stay in round bottom tubes incubated at 22°C in the dark for 16 hours (see Note 9).

3.3 Image acquisition

1. In this chapter the FRET-FLIM experimental procedure is optimized for a Leica SP5. Single photon excitation pulses of 40 Mhz (see Note 10) are generated by a diode pulsed laser resulting in excitation pulses of ±10 ps. (see Note 11)

2. Image acquisition is performed using a 60x/1.2 water-immersion objective.

3. Set objective for optimal glass thickness (see Note 12).

4. Cyan fluorescence emission was selected from 450–495 nm using spectral detection (see Note 13).

5. Single photons are captured using internal hybrid detector (see Note 14).

6. Fluorescence images of 128 x 128 pixel size are set for both the Leica acquisition software as well as the B&H SPC 830 FLIM module (see Note 15).

7. The ADC (analog-digital converter) of TCSPC module is set at 256 channels (see Note 16).

8. FLIM acquisition is performed using line scanning at 400 Hz (see Note 17) using an average count rate around 10^4 photons per second for an acquisition time of 90 s (see Note 18).

9. Single transfected protoplasts expressing C-terminally sCFP3A-tagged ARF5 results in the donor fluorescence lifetime required as reference. To reveal dimer formation of ARF5, protoplasts expressing ARF5 sCFP3A and sYFP2 tagged proteins are investigated (see Note 18).

10. The protoplasts are transferred into an 8-well chamber for imaging.

3.4 Analysis of FLIM Data with SPCImage 5.2

1. SPCImage is a software package, which is included with the B&H acquisition card. This software determines the instrumental response function from the rise of the decay curve and performs data analysis using an exponential model function (see Note 20).

2. After importing the raw data, a fluorescence intensity image will appear. A blue crosshair is visible that can be used to point anywhere in the image for displaying the fluorescence decay of that pixel.

3. Within the histogram showing the fluorescence decay, the limits in between the fluorescence lifetime analysis should be performed, can be defined. Typical values are around 2 ns (before the rising edge) for the left and about 20 ns for the right border.

4. A threshold of fluorescence intensity can also be set. Pixels will be excluded from the analysis if the numbers of photons in that pixel is below that value.

5. In case the number of photons is too low for fluorescence lifetime analysis, static pixel binning can be applied. This specific binning option is a procedure where the selected pixel is analyzed including the photon statistics of neighboring pixels for calculation of the fluorescence lifetime. The binning factor can be calculated according to the following formula: Binning factor = $(2n+1)^2$, where n is number of pixels. For the FLIM images shown in Figure 2, a binning factor of 1 has been applied (see Note 21).

6. Before fitting the raw data, the number of components for the underlying exponential fit function has to be defined (here one or two). All other parameters should remain unfixed independently of the fit model.

7. After the fit procedure, the fluorescence lifetimes are calculated per pixel and displayed as a false-colored image (see Fig. 1 D and Fig 2 A, C). During the fitting process, the chi-square value between model function and data is minimized [260]. For a good fit, the chi-square value should be around one and the displayed residuals in the box below the fitted curve should scatter around zero.

8. To reveal if protein interactions take place one compares donor fluorescence lifetime in the absence and presence of acceptor. Single transfected protoplasts (donor

137
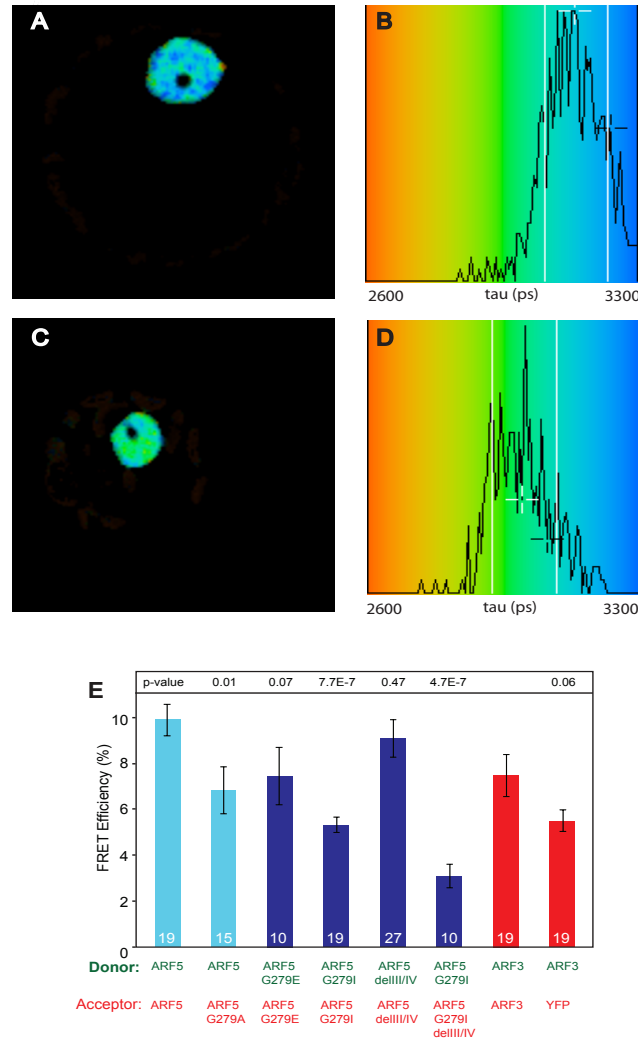
Figure 2; FRET-FLIM analysis of ARF5 homodimerization in plant protoplasts.
Donor fluorescence lifetime images of ARF5-sCFP3A alone
(A) and ARF5-sCFP3A in presence of ARF5-sYFP2 proteins (C). Histograms of fluorescence lifetime values of the nucleus (B and D) show a change of color (from blue to green) when donor fluorescence lifetime is reduced. (E) Quantification of *in vivo* ARF dimerization measured by FRET-FLIM in protoplasts. Protein interaction is expressed as average FRET efficiency (± SEM), number of protoplasts is indicated in each bar, and the p-value for significance of difference between mutants and wild-type is shown (two-sided Student's t test). Mutations within the dimerization domain of the ARF5 DBDs abolished the interactions, resulting in FRET efficiencies around background level. The donor is indicated in green, and acceptor in red. The histogram is reproduced from reference [126].

only) are evaluated using one- component analysis (see Note 22).

9. Double-transfected protoplasts (donor and acceptor) are analyzed based on a two-component model. The fluorescence lifetimes and amplitudes obtained can be assigned to two donor populations – one transferring energy to an adjacent acceptor molecule resulting in a reduced fluorescence lifetime, the second showing no FRET and hence exhibiting fluorescence decay kinetics as donor alone (see Note 23).

10. After FLIM analysis, a fluorescence intensity image and a false colored donor fluorescence lifetime image (see Fig 2 A and C) is shown. The color distribution of the FLIM image is automatically set. To compare donor alone versus donor-acceptor samples it is possible to set the color limits of fluorescence lifetime (see Fig. 2 C and D).

11. The nucleus of the fluorescence intensity image can be selected using the "region of interest" (ROI) tool. The corresponding fluorescence lifetime distribution will directly be adapted.

12. After performing the calculations, the mean fluorescence lifetime, the distribution of the single pixel values as well as a false-color coded lifetime image will be displayed for the selected ROI.

13. In order to determine if a particular protein couple is interacting, three independent transfections of 10-15 samples of each construct are measured and average fluorescence lifetimes are determined for all cells. In this example we show a histogram of the FRET efficiencies of different combinations tested (see Fig 2 E).

4. Notes

1. Over the years many variants of the Cyan family were developed. For fluorescence lifetime analysis mTurquoise2 is the most optimal fluorophore, as it shows the longest fluorescence lifetime (see more information in [261])

2. An extended protocol for ligation independent cloning can be found in Wendrich and coworkers [262].

3. Heat twice for 6 s at 300 Watt in a microwave to dissolve PEG.

4. FLIM can be performed on any confocal microscope (or multi photon microscope).

The only requirement is the photon registration of the detector, which is coupled to a time correlated single photon counting card

5. The Time tape supports the top side of the leaf during manipulation, while tearing off the 3 M Magic tape allows easy removal of the lower epidermal layer and exposes mesophyll cells to cell wall digesting enzymes when the leaf is later incubated in an enzyme solution. The protoplasts released into solution are collected and washed for further use.

6. Incubation can be performed at lower temperatures but this will reduce the enzyme activity and duration should be adapted

7. MMg solution is stressful for protoplasts therefore the transfection procedure should take place quickly.

8. Use tips with enlarged openings for pipetting to reduce shear forces.

9. In general, imaging can be performed about 16 h after transfection but this depends on the protein expression. It is recommended to check the expression levels at several time points to determine the duration for optimal expression.

10. Pulsed diode lasers have the possibility to vary the repetition rate of the laser, which is set depending on the fluorescence lifetime. Rule of thumb is the fluorescence lifetime multiplied by 5 to know what time window is appropriate.

11. Fluorophores absorbing in the range of 470 -670 nm can be excited using a super continuum laser, which have typical excitation pulses of 1 ps.

12. This can be achieved by selecting 488 nm laser, setting AOBS in reflection mode and select regular PMT.

13. In case filters are used, narrow band-pass filters should be used to avoid contamination of acceptor signal in donor channel.

14. External Hybrid (HyD) or MCP-PMT (Hamamatsu Photonics, Japan) detectors can also be used for time correlated single photon counting experiments. HyD detectors show high quantum efficiencies, low dark noise and large dynamic range and have a typical time transient of 100 ps, which offers opportunities like fluorescence lifetime imaging, but can also be used for single molecule type of experiments like Fluorescence correlation spectroscopy. Hamamatsu R3809U MCP photomultiplier suffer in quantum efficiency but exhibit best time resolution (30 ps).

15. FLIM analysis is optimal at sufficient photon distribution per pixel. Imaging at higher pixel resolution requires a prolonged data acquisition time to ensure the detection of a statistically relevant number of photons.

16. Time resolution of the ADC can be set ranging from 16 to 4096 channels. In case of low photon counts, the ADC can be set to 64 channels.

17. Scanning should not be slower than 200 Hz and bidirectional scanning is not appropriate.

18. TCSPC principle only holds when single photons are counted. Higher count rates can cause so called pile up effects, which affects fluorescence lifetime estimation.

19. In order to obtain information about the expression levels of the sCFP3A and sYFP2 tagged proteins a sequential confocal image of donor and acceptor should be taken. Before a FLIM measurement, confocal images of donor and acceptor should be taken to select cells that show similar expression levels.

20. Other analysis software like TIMP developed by [263] or FLIMfit [264] can be used.

21. In these studies, the effect of binning has been checked between binning 0 (no pixel binning) and 1. No significant difference of the average fluorescence lifetime distribution was observed. However, image acquisition was taking 5 times longer in case binning 0 was used for obtaining sufficient number of photons for data analysis.

22. Donor samples are also evaluated using two-component analysis to check for false positives.

23. Using 2 component analysis one can choose to fix the long lifetime to value found for donor only sample.

Acknowledgements

CHAPTER 7

General Discussion

General Discussion

Hormones are signaling molecules that can elicit many different responses in an organism. For example, in animals, perception of the hormone acetylcholine triggers contraction in striated muscles, reduces contractions in heart muscle and promotes secretion of digestive enzymes in pancreatic acinar cells [265]; the hormone progesterone is involved in control of reproduction and sexual behavior but also has effects in the nervous system where it is involved in adult neurogenesis, astroglial and synaptic plasticity, and the myelination process [266]; vitamin D is involved in calcium-phosphorus maintenance and bone homeostasis, cell differentiation, cell-cycle control, glucose metabolism, muscle and adipose tissue function[267]. This variety of responses is seen also for plant hormones, like auxin. As discussed in **Chapter 1** and **Chapter 2**, auxin is a chemically simple, small molecule that can trigger many different responses in different plant organs through all the stages of a plant's life. The different outcome of hormone action between organs lies not in the hormone itself, but in the proteins responding to it in the different cells. In all cell types, response to auxin follows a conserved order of events, referred to as the "Auxin nuclear signaling pathway" that involves three dedicated protein families. The first family is in charge of perceiving the presence of auxin in the nucleus, the TIR/AFB family of 6 members. In the presence of auxin, these proteins bind the members of the second protein family involved, the Aux/IAA family of 29 members, and promote their degradation through the ubiquitin-proteasome system. The Aux/IAA proteins are repressors of the transcriptional activity of the third family, the Auxin Response Factor (ARF) family of 23 members, which they bind to through a shared Phox/Bem1 (PB1) protein-protein interaction domain at their C-termini. The number of protein members of each family corresponds to *Arabidopsis,* which is the model plant used for most auxin-related studies, and also in this thesis. As the response to the hormone auxin will be highly dependent on the cellular context where it is perceived, expression patterns of the proteins involved in the auxin signaling pathway [99,44] and their combinatorial effect in the auxin output [99,42,49] have been studied. Differences in expression levels and patterns can be observed between the auxin receptors[94] and two of them (AFB4 and AFB5) seem to be required for response to auxin analogs[42,268]. Despite these differences, single mutants show very subtle effects on auxin sensitivity suggesting genetic redundancy between the receptors[94,269]. They all have a binding pocket for auxin but stable binding requires the formation of a co-receptor complex with the Aux/IAAs[42]. These co-receptors have shown different sensitivities for auxin[42]. As the immediate effect of the formation of theses complexes is the ubiquitination and degradation of Aux/IAAs, the differences in sensitivity may result in some Aux/IAA not being degraded in cells were the auxin levels are below a certain threshold, furthermore Aux/IAAs have been proven to have different degradation rates (some can be degraded in minutes while others take hours). Finally the presence of "survivor" Aux/IAAs will directly affect the activity of the ARFs. Pairs of ARFs-Aux/IAAs have been suggested to have a role in specific processes but swap experiments of Aux/IAAs revealed that they have comparable activity, if expressed at the same cell type and at the same levels[99]. An interaction network model also suggested that virtually Aux/IAAs can interact with almost any ARF[43]. In conclusion the first steps of the nuclear auxin signaling pathway will contribute to specific outputs by regulating "how fast" ARFs will be free to act in specific cell types (amount of auxin + Aux/IAAs). The ultimate

effectors of the pathway are the ARFs. ARFs are transcription factors and will regulate gene expression in response to the hormone. The variation in gene responses will then depend in the intrinsic variation within the ARF family member's activity.

ARF transcription factors as the diversity hub in auxin action

In *Arabidopsis* there are 23 different ARFs and (with few exceptions) they all share a general domain architecture: an N-terminal DNA binding domain (DBD) that harbors a B3 motif and a dimerization interface, followed by a Middle Region (MR) that connects it to the C-terminal domain that harbors a PB1 protein-protein interaction motif. Despite their similarities, ARFs can have different effects in the transcription of their target genes either activating or repressing them. Based on phylogenetic distribution and transcriptional activity, the ARFs have been divided in three classes – A, B and C[34]. It was shown in protoplasts that the class A ARFs that were tested (5, 6, 7 and 8) activated expression of a reporter gene while the tested class B ARFs (1, 2, 3 and 9) repressed expression [92]. The activation/repression properties of ARFs were correlated with the overall amino acid composition of their MR: activator ARFs are rich in Glutamine and repressor ARFs are rich in Proline and/or Serine[92]. The rest of the family members were then arbitrarily classified as activators (class A) or repressors (class B and C) according to their amino acid composition and without experimental evidence, and exceptions to the 'rule' are seen. According to this classification ARF19 (Q-rich MR) is an activator ARF but it is interesting to note that its MR carries an LxLxL motif that mediates interactions with the transcription repressor TOPLESS[45]. Furthermore *arf7arf19* double mutant shows loss of regulation of auxin up and down regulated genes, suggesting that these ARFs can affect gene regulation in both ways[101]. ARF5, which acted as an activator in protoplasts, and indeed acts in development in part by activating some of its target genes (TMO5[103], TMO7[103], ATHB8[66], LEAFY[137]), can also act by repressing genes in a different cellular context[270,271]. Amino acid composition of the MR is not the only intrinsic difference between ARF proteins. Sequence alignments of their DNA binding domains reveal a subset of ARFs (3, 4, 10, 16 and 17) where the dimerization interface is not conserved ([126], Chapter 3), repressor ARFs from the class C are targeted by micoRNA160[272] and finally some ARFs (3 and 17) are truncated in their C-terminus and thus their activity may be controlled in an alternate way that might not involve Aux/IAAs.

ARFs have been considered unique to land plants[34] but they can be tracked back to one ARF in green algae (Sumanth Mutte, personal communication). The earliest diverging land plants, the liverworts represented by *Marchantia polymorpha*[208], contains three ARFs and from there their numbers increased as land plants became more complex[34]. A clear link between the appearance of 'new' ARFs with 'new' functions has not been clearly established. It has been proposed that cell-type specific response to auxin goes in hand with cell-type specific ARF combinations[44]. Early in plant embryo development ARFs start expressing differently giving rise to an ARF pre-pattern, and the presence of specific ARFs is absolutely necessary for specific processes in specific cells [122,142,273,47]. It will be interesting to see if the expansion of the ARF family in land plants can be causally connected to increased complexity

in auxin-dependent gene regulation and morphological diversity.

Taking into account the fact that ARFs are different in their amino acid composition, domain architecture and regulatory effects on their targets; that they are expressed in different cell types[44] ; and that some of them are involved in specific biological processes [142,47,46,170]; it would be expected that they are able to specifically regulate different genes. This can be seen in the little redundancy in function between members of the family, phenotypes can be observed in many cases mutating only two members[142,47,123], and in the case of ARF5 its sole mutation already has a great effect in plant development[102], despite being co-expressed with other ARFs.

The atomic basis of ARF-DNA interactions

ARFs, as other eukaryote proteins, are modular. Modularity of proteins can be used to study properties of independent protein domains as they retain functionality when isolated. In this thesis we took advantage of ARFs' modular properties in order to study their different domains and how they may contribute to each ARFs specific activities. With some exceptions, where deletions occurred, ARFs have the general architecture of: an N-terminal DNA binding domain (DBD) followed by a Middle Region (MR) that connects it with a C-terminal domain (CT). As the main question we address in this work is how ARFs manage to give such variety of outputs to the same signal by specifically regulating transcription of different genes, we first addressed the structural basis for functional diversity in their DNA binding domain.

ARF DBD's harbor a highly conserved B3 DNA binding domain which is not exclusive to ARFs, but also found in three other transcription factor families of the B3 superfamily (LEAFY COTYLEDON2-ABI3-VAL (LAV); RELATED TO ABI3/VP1 (RAV); and REPRODUCTIVE MERISTEM (REM))[28]. All these families are related to hormone response[125]. Consensus binding sequences have been also identified for the B3 domains of ABI3/VP1 members of the LAV family (5'-CATGCA-3')[29], and for RAV family members (5'-CACCTG-3')[181]. Furthermore, structures of the domain of RAV1 and a member of the REM family were solved and resembled the non-catalytic part of bacterial restriction enzymes EcoRII and BfiI (reviewed in [28]), but none of these structures were obtained in complex with DNA. Understanding how the ARF DBD contacts the DNA could help to understand how different ARFs regulate different genes. There was the possibility that DNA contacting residues were different between ARFs, thus conferring variation in the way they bind DNA. As there was no reported structure of ARF DBDs, in **Chapter 3** we addressed this issue by solving the crystal structure of the DBDs of the functionally divergent ARF1 and ARF5 proteins, in the case of ARF1 also in the presence of DNA. This work allowed identifying the DNA-contacting residues and by sequence alignments of the family members we could confirm that they were mostly conserved.

An obvious question became whether different ARFs had different specificities for the DNA sequence they recognize and bind. We first address this question in **Chapter 3** using Protein

Binding Microarray (PBM) for the DBD of two functionally different ARFs: ARF1, a class B (repressor)and ARF5, class A(activator). Strikingly, both ARFs showed the highest affinity to the exact same DNA sequence TGTCGG. Later it was reported that ARF3 – another class B ARF - also has the highest affinity for the same motif determined by the same technique[166]. Before our study, the canonical Auxin Responsive Element (AuxRE) had been described as TGTCTC[29] and the sequence was the base for the widely used auxin activity reporter DR5. Our re-defined AuxRE was used to construct a new version of this reporter, DR5v2 by *Liao et al* [175]. DR5v2 proved to be more sensitive to auxin *in vivo,* showing expression in cells where DR5 did not. Later, TGTCGG was confirmed as the preferred ARF binding site by other methodologies [171,172]. To understand the biophysics for this base preference in **Chapter 4** we obtained the structure of ARF1 DBD bound to the redefined TGTCGG AuxRE and compared it with the structure of ARF1 DBD bound to the canonical TGTCTC AuxRE presented in **Chapter 3**. The DNA-contacting His136 from the B3 moiety, forms 1 hydrogen bond with the G opposed to the last C of the TGTCTC sequence, but it can form hydrogen bonds with either one of the last Gs plus the C opposing the last G of the TGTCGG sequence. His136 in ARF1 corresponds to His170 in ARF5. In **Chapter 3,** we show that a His170 to Ala mutation in ARF5 caused reduced binding affinity to DNA *in vitro* (shown by Surface Plasmon Resonance), which was reflected in its reduced *in vivo* activity as the His170Ala mutant of ARF5 could only partially complement the phenotype of the *arf5/monopteros* mutant. PBM showed that the mutant protein still preferentially bound the TGTCGG sequence, but it became less stringent in the choice of the last two bases compared to wild-type ARF5 protein. While amino acids contacting the first bases of the AuxRE are critical for specific ARF-DNA interactions, His170 affects the strength of the binding as well as the specificity at the last two DNA positions. The last bases of the AuxRE will affect how His170 accommodates in space and how many bonds it can form, thus fine tuning specificity. A weak binding (influenced by the last AuxRE bases) will result in little effect of auxin on such gene probably by reducing its occupancy time. Direct measurements of on- and off-rates, as well as dwelling and occupancy times of wild-type and mutated ARF proteins on various DNA elements, for example by single-molecule methods, would allow addressing this question.

It is not uncommon for different members of eukaryote transcription factor families to show high affinity for the same DNA sequence. This makes specific gene regulation a much more complex event that goes beyond a simple 1:1 relation where only one transcription factor and one regulatory element in the promoter region are involved in defining the regulatory output. Furthermore, optimal *in vitro* binding sites usually do not always reflect *in vivo* DNA – transcription factor binding[274] and specific preferences by members of a transcription factor family may be seen only at low-affinity binding sites[166] or may be influenced by the DNA flanking the recognition elements[274]. *In vivo* binding data (ChIP) could shed light on specific preferences and thus on the biologically meaningful differences in binding site selection between ARFs. Unfortunately such data is not available for ARFs and it should be a priority in order to obtain a clearer understanding of ARF gene regulation.

Molecular calipers recognize complex DNA motifs

A novel feature of ARFs DBDs came to light with their structural determination and it was their ability to form dimers, and to bind DNA as dimers. We could confirm with further experiments that ARF DBDs cooperatively bind as dimers to DNA that have an inverted repeat of the AuxRE binding site. Furthermore, mutations impairing dimerization reduced DNA-binding affinity and rendered proteins non-functional *in vivo*. Interestingly, a natural variant of the *Brassica napus ARF18* gene that deletes 55 amino acids from the dimerization domain within the DBD prevents dimerization and causes changes in seed size and seed pod length[173]. *Brassica* ARF18 is co-orthologous to *Arabidopsis* ARF11 and ARF18, which belong to the class B (repressor) ARFs. Thus, from this natural variant it can be concluded that the biological relevance of DBD-mediated ARF dimerization extends to class A (ARF5) and class B ARFs, and to multiple species. It will be important asses to what degree DBD dimerization is a fundamental property of ARF proteins.

Dimerization of the DBD is mediated by an alpha-helix that is conserved among most members of the family, but that is notably absent in the class C ARFs. Thus, it would be expected that most, if not all Class A and B ARFs can form homodimers. Given the close structural conservation at the dimerization interface between ARF1 and ARF5, it is plausible that ARFs would also form heterodimers. Evidence for ARF hetero-dimerization has been previously reported by FRET-FLIM in *Arabidopsis* protoplasts [213]. However, in those assays, full-length proteins were expressed and given that the C-terminal PB1 domain can also mediate dimerization, it remains to be determined is such heterodimerisation occurs at the DBD, and what its relevance could be. As discussed above, ARFs can have different ways of acting on genes (activating/repressing) and heterodimerization of different classes of ARFs may affect transcriptional output. Heterodimerization may serve for example as a way to control ARF activity by coupling a repressing ARF to an activating one, and these may simply happen as a result of the stoichiometry of ARFs in specific cell types without further need for regulation beyond protein levels.

An implication of the cooperative binding of ARF dimers is that, for high-affinity binding, two AuxRE sites are needed instead of one. Furthermore, the two AuxRE sites need to have the correct orientation and distance. The structures we obtained predicted that the binding site needs to be an inverted repeat of AuxREs. The structures of ARF1 and ARF5 also showed that there is a loop connecting the dimerization interface to the DNA binding interface that may contribute to differences between these two proteins. This loop is likely flexible and is the most variable part in ARF DBDs. This intrinsic flexibility restricts the distance between the two ARF monomers within the dimer, and thus represent a structural constraint for the distance tolerated between the inverted AuxREs. In this way each ARF homo- or heterodimer may have an optimal AuxRE spacing that will confer a degree of target specificity between the ARFs. Indeed, Surface Plasmon Resonance analyses confirmed that the ARF1 and ARF5 DBD differed in the space tolerated between the inverted AurRE repeats. We thus likened this mode of action to a caliper, and proposed the "caliper model for ARF-DNA specific binding". It is common among eukaryotic and prokaryotic transcription factors to form dimers and bind cooperatively to DNA. Such properties are for example

151

found among the basic Helix-Loop-Helix (bHLH), basic Leucine Zipper (bZIP), Nuclear Hormone Receptor (NR), MADS-box and Homeodomain-Leucine Zipper (HD-ZIP) families[275]. Often, dimerization is a requirement for DNA binding, such as for example in the bHLH family[276]. The use of inter-site spacing as a way to fine tune specificity among family members had not been described for plant transcription factors yet. A "caliper" has been described for the human Steroid/Thyroid hormone Receptor (ER/TR) family. These transcription factors also bind DNA as homo- or heterodimers, and the spacing between the same recognition elements increases target specific binding: 3 bp for vitamin D receptors, 4 bp for the thyroid hormone receptor and 5 bp for the retinoic acid receptor[149,277]. These TFs bind direct repeats of their RE, but another subfamily, the Estrogen Receptors, binds to inverted repeats of the same ER[278]. Distance and orientation of recognition elements can thus determine binding of a transcription factor to their target genes.

Based on the ARF DBD structures presented in **Chapter 3**, we would expect the presence of inverted repeats of AuxRE elements in the promoter regions of ARF targets. We indeed found inverted repeat with appropriate spacing of 7 bases, in the promoter regions of the ARF5/MP targets *TARGET OF MONOPTEROS5 (TMO5), TMO3* and *LEAFY* [107]. We used this bipartite element in the *TMO5* gene promoter to test relevance of motif topology. Elimination of even one of the binding sites within this repeat brings the expression of TMO5 to significantly low levels, showing the power of composite AuxREs for gene regulation by ARFs.

As distance and orientation of recognition elements may be important to define target specificity, we used a bioinformatics approach to mine for composite AuxREs that correlate to auxin responsiveness in genes. With this analysis, presented in **Chapter 4**, we found a strong correlation of auxin response with two bipartite AuxREs: an inverted repeat with a separation of 8 bases (IR8) and a direct repeat with a separation of 5 bases (DR5). It was shown with DAP-seq that ARFs bind both bipartite motifs[172]. Furthermore we tested the presence of either of these elements in promoter regions as indicators of auxin responsiveness in small sets of diverse genes and found that around 75% of them did actually show a change in expression levels after auxin treatment. This means that these bipartite elements are highly predictive of auxin responsiveness, and thus can be used as a tool to identify auxin regulated genes. Researchers often take the presence of single AuxRE as an indication of possible auxin responsiveness but the chances of finding a TGTCNN element in a 1 kb promoter fragment are close to 100%, and this element alone therefore has little predictive power. The structure-based IR8 motif instead, has a low background occurrence, and therefore serves as an excellent predictor of generic ARF-dependence and auxin-responsiveness. We propose that this motif is adopted as the gold standard element for predicting auxin-dependence.

As discussed above, we propose that each ARF dimer may have preference for specific spacing between the TGTCNN elements of a bipartite AuxRE. Establishing which those preferences are for the different ARF dimers could further improve the predictive power of the bipartite elements that could be used to find direct links between a specific ARF and genes co-expressed in the same cell-types. A systematic approach to reveal distance preferences of ARF dimers is yet to be performed. A proper system for this should be chosen as transactivation of bipartite

AuxRE promoters in a system where multiple ARFs are present will likely give noisy results. One possibility would be to exploit the simple *Marchantia* system for such investigations.

Interestingly genes carrying the DR5 element in their promoter were only upregulated (not downregulated) when regulated by auxin. Binding of ARFs to inverted repeats is explained by the structural conformation of symmetric DBD dimers, but how multiple ARFs would physically fit upon a DR5 motif remains unclear. It is a possibility that ARF dimerization via the CT plays a role here, although the SPR experiments showing cooperative binding were done using only the DBDs, and truncated versions of ARF do not lose their DNA binding / regulatory capacity *in vivo* [210,135,194]. We tested binding of ARFs from all three different classes to DR5, finding that only class A (activator) ARFs can bind this sequence in a cooperative manner. Interestingly in the DAP-seq experiment mentioned above, where both ARF5 and ARF2 were tested, the DR5 peak was only observed for ARF5, consistent with our results. Together with the expression analysis we conclude that the DR5 motif can only be bound by certain (activator) ARFs and only to promote transcription of the target.

As a result of the studies done in the relationship between ARFs DBDs and the DNA, we can postulate that: 1) ARFs bind cooperatively as dimers to bipartite AuxREs in the promoters of their target genes and hetero-dimerization of ARFs may contribute in the generation of diversity in binding specificity, 2) distance and orientation of the AuxREs are important factors in defining ARF target specificity, and 3) DBDs of (activator) ARFs may acquire a different spatial configuration induced by the topography of the cis-regulatory regions.

ARF anatomy: domain requirements for function

The DBD has several features that contribute to ARF specific binding to DNA. ARF5-ARF6 DBD swap experiments presented in **Chapter 5** confirmed that the DBD is absolutely necessary and partially sufficient for regulation of ARF5 target genes, although the rest of the protein also contributes to fine tuning of gene specific regulation. Again we took advantage of ARF modularity to learn more about the attributes of the MR and CT. Using each ARF domain in proteomics experiments we found that the DBD alone can engage in active transcription complexes; and we could confirm the previously reported ARF-ARF ARF-Aux/IAA interactions mediated by the C-terminal domain important for the regulation of ARFs' activity. We also took a closer look at the Tudor-like domain within the DBD (described in Chapter 3) and concluded that it is not required for ARF5 to interact with the transcription machinery, but it is essential for biological function and may thus be required for correct regulation of the target genes.

We contemplate the possibility of ARFs being able to act as pioneer transcription factors. Pioneer transcription factors are usually involved in reprogramming the fate of cells during development[279,280], like ARFs. Transcription of pioneer target genes is usually delayed as they are hidden in nucleosomes and first need to become accessible[280]. Although it is not a common feature of transcription factors to be able to access condensed DNA, pioneers can bind with specificity their recognition elements even when the DNA elements are wrapped in

nucleosomes[279] and are able to then open up chromatin for other factors to bind and initiate transcription. They can do this in different ways; some may have, besides their specific DNA binding domain, a second protein domain that allow them to interact directly with histones. These kind of interactions can destabilize the nucleosome making the DNA accessible for trasncription[281]. Furthermore, interactions with histones may depend on their Lysine and Arginine methylation state[282] and domains that can read these methylations help to bind nucleosomes with specificity. ARF DBDs have a Tudor-like domain; in animals, Tudor domains are involved in binding methylated Lysines in histones[282]. It is therefore possible that ARF dimers cluster (via their PB1 domain) on nucleosomes and destabilize these with the aid of their Tudor domain. Direct assays for ARF binding to *in vitro* reconstituted nucleosomes (see for example [283,284]) and analysis of the stability of such nucleosomes in the presence of ARF proteins would allow determining if ARFs can act as pioneer factors.

Another way for making chromatin accessible is to recruit other proteins involved in chromatin remodeling. It has been proposed that ARF5 induces chromatin remodeling by recruiting a SWI/SNF complex[25]. In this context, it is conceivable that the Tudor-like domain would aid positioning an ARF-SWI/SNF complex in nucleosomes. This would not be the first report of a plant transcription factor involved in inducing chromatin changes as a similar scenario has been proposed for MADS-domain transcription factors[285]. Changing the chromatin state can affect greatly the specific action of a transcription factor giving, for example, a 'temporal' specificity as in the case of mammal FoxA transcription factors. FoxA factors have been involved in making accessible for regulation genes necessary for hepatocyte differentiation and hence their mutation early in development is lethal but it has no effect in already differentiate tissue[286].

As genes carrying the DR5 motifs seem to be only up-regulated (for which they should be accessible for transcription) and bioinformatics correlated the presence of the DR5 motif to delayed response (**Chapter 3**), it could be interesting to correlate 'DR5 genes' to chromatin accessibility (DNAse footprint) of different tissues or different developmental stages.

ARF5 interaction with the remodeling complex is mediated by the MR. Also the MR region of repressor ARFs carry a small LxLxL motif involved in interactions with TOPLESS transcriptional repressor. TOPLESS likely acts by recruiting a Histone Deacetylase (HDAC) enzyme[287], and thus presumably alters DNA accessibility in chromatin context. Unfortunately we could not look into other interactions mediated by this domain, as all truncations of ARF5 that contained this domain located in punctate cellular structures, and were absent from the nucleus. From structural predictions it seems that this domain lacks structure. Intrinsically disordered regions in transcription factors have been proposed to be important as their flexibility allows them to expose different interaction surfaces depending in the cellular context. It could be that the MR of ARFs plays an important role in defining context-dependent gene regulation *via* interactions with other proteins, such as transcription factors. There is not enough information about the role of the MR for ARF functions. It is easier to study ARFs DBDs and CTs because they are highly conserved and data obtained from determination of the properties or structure of a few representatives can be easily extrapolated to the whole family. The MR in contrast is highly variable and to understand

its role it would be necessary to study each of the 23 separately. This added to its intrinsic disorganization makes it a difficult study subject. The effort though appears justified as MRs are the major source of variability between ARFs (**Chapter 5**).

In this thesis we addressed the question of how the 23 members of the ARF family can generate the broad variety of responses to auxin in plant development. We employed multiple approaches in order to gain more insight in the process. Structural biology gave us the starting point and the use of protein binding microarrays, bioinformatics, *in vitro* binding assays, *in vivo* mutant complementation assays, *in vivo* FRET-FLIM and proteomics came together to fill some of the blanks. The results provided here can be taken as a starting point for the elucidation of ARF's singular properties influencing specific gene regulation.
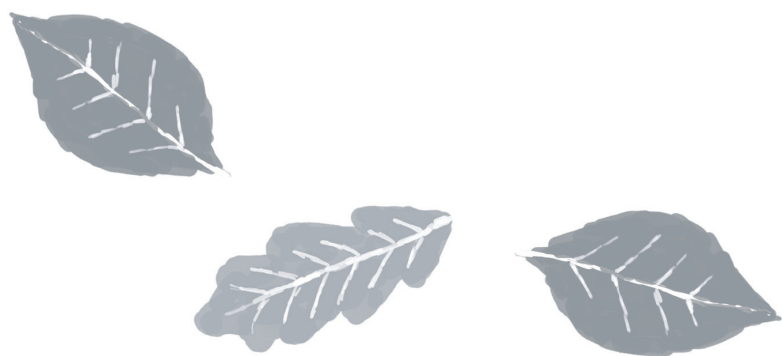
List of References

1.   Crick F. Central Dogma of Molecular Biology. *Nature*. 1970;227:561-563.

2.   Matsui T, Segall J, Weil PA, Roeder RG. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem*. 1980;255(24):11992-11996. http://www.ncbi.nlm.nih.gov/pubmed/7440580. Accessed July 25, 2016.

3.   Samuels M, Fire A, Sharp PA. Separation and characterization of factors mediating accurate transcription by RNA polymerase II. *J Biol Chem*. 1982;257(23):14419-14427. http://www.ncbi.nlm.nih.gov/pubmed/7142220. Accessed July 25, 2016.

4.   Sawadogo M, Roeder RG. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc Natl Acad Sci U S A*. 1985;82(13):4394-4398. http://www.ncbi.nlm.nih.gov/pubmed/3925456. Accessed July 25, 2016.

5.   Reinberg D, Roeder RG. Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains. *J Biol Chem*. 1987;262(7):3331-3337. http://www.ncbi.nlm.nih.gov/pubmed/3818644. Accessed July 25, 2016.

6.   Flores O, Maldonado E, Reinberg D. Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J Biol Chem*. 1989;264(15):8913-8921. http://www.ncbi.nlm.nih.gov/pubmed/2566609. Accessed July 25, 2016.

7.   Flores O, Lu H, Reinberg D. Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J Biol Chem*. 1992;267(4):2786-2793. http://www.ncbi.nlm.nih.gov/pubmed/1733973. Accessed July 25, 2016.

8.   Ge H, Martinez E, Chiang C-M, Roeder RG. Activator-dependent transcription by mammalian RNA polymerase II: In vitro reconstitution with general transcription factors and cofactors. *Methods Enzymol*. 1996;274:57-71. doi:10.1016/S0076-6879(96)74008-9.

9.   Thomas MC, Chiang C-M. The General Transcription Machinery and General Cofactors. *Crit Rev Biochem Mol Biol*. 41(3):105-178. doi:10.1080/10409230600648736.

10.   Buratowski S, Hahn S, Guarente L, et al. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*. 1989;56(4):549-561. doi:10.1016/0092-8674(89)90578-3.

11.   Wu S-Y, Zhou T, Chiang C-M. Human mediator enhances activator-facilitated recruitment of RNA polymerase II and promoter recognition by TATA-binding protein (TBP) independently of TBP-associated factors. *Mol Cell Biol*. 2003;23(17):6229-6242. http://www.ncbi.nlm.nih.gov/pubmed/12917344. Accessed August 12, 2016.

12.   Schneuwly S, Klemenz R, Gehring WJ. Redesigning the body plan of Drosophilaby ectopic expression of the homoeotic gene Antennapedia. *Nature*. 1987;325(6107):816-818. doi:10.1038/325816a0.

13.   Doebley J, Wang R-L, Stec A, Hey J, Lukens L. The limits of selection during maize domestication. *Nature*. 1999;398(6724):236-239. doi:10.1038/18435.

14.   Durbin ML, Lundy KE, Morrell PL, Torres-Martinez CL, Clegg MT. Genes that determine flower color: the role of regulatory changes in the evolution of phenotypic adaptations. *Mol Phylogenet Evol*. 2003;29(3):507-518. doi:10.1016/S1055-7903(03)00196-9.

15.   Yoon H-S, Baum DA. Transgenic study of parallelism in plant morphological evolution. *Proc Natl Acad Sci*. 2004;101(17):6524-6529. doi:10.1073/pnas.0401824101.

16. Wunderlich Z, Mirny LA. Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet*. 2009;25(10):434-440. doi:S0168-9525(09)00165-6 [pii]\r10.1016/j.tig.2009.08.003.

17. Fisher WW, Li JJ, Hammonds AS, et al. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proc Natl Acad Sci*. 2012;109(52):21330-21335. doi:10.1073/pnas.1209589110.

18. Biggin MD, Arnosti DN, Barolo S, et al. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Dev Cell*. 2011;21(4):611-626. doi:10.1016/j.devcel.2011.09.008.

19. Simicevic J, Schmid AW, Gilardoni PA, et al. Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. *Nat Methods*. 2013;10(6):570-576. doi:10.1038/nmeth.2441.

20. Xu X, Chen C, Fan B, Chen Z. Physical and Functional Interactions between Pathogen-Induced Arabidopsis WRKY18, WRKY40, and WRKY60 Transcription Factors. *Plant Cell*. 2006;18(5):1310-1326. doi:10.1105/tpc.105.037523.

21. Olsen ANAN, Ernst HAHAHA, Leggio LL Lo, et al. NAC transcription factors: Structurally distinct, functionally diverse. *Trends Plant Sci*. 2005;10(2):79-87. doi:10.1016/j.tplants.2004.12.010.

22. Kagaya Y, Ohmiya K, Hattori T. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res*. 1999;27(2):470-478. doi:10.1093/nar/27.2.470.

23. Singh KB. Transcriptional Regulation in Plants: The Importance of Combinatorial Control. *PLANT Physiol*. 1998;118(4):1111-1120. doi:10.1104/pp.118.4.1111.

24. Tiwari SB, Wang XJ, Hagen G, Guilfoyle TJ. AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. *Plant Cell*. 2001;13(12):2809-2822. doi:10.1105/tpc.010289.

25. Wu M-F, Yamaguchi N, Xiao J, et al. Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate. *Elife*. 2015;4:e09269. doi:10.7554/eLife.09269.

26. Bhattacharyya RP, Reményi A, Yeh BJ, Lim W a. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem*. 2006;75:655-680. doi:10.1146/annurev.biochem.75.103004.142710.

27. Frankel AD, Kim PS, Brent R, et al. Modular structure of transcription factors: implications for gene regulation. *Cell*. 1991;65(5):717-719. doi:10.1016/0092-8674(91)90378-c.

28. Yamasaki K, Kigawa T, Seki M, et al. DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends Plant Sci*. 2013;18(5):267-276. doi:10.1016/j.tplants.2012.09.001.

29. Ulmasov T. ARF1, a Transcription Factor That Binds to Auxin Response Elements. *Science (80-)*. 1997;276(5320):1865-1868. doi:10.1126/science.276.5320.1865.

30. Ulmasov T, Liu ZB, Hagen G, Guilfoyle TJ. Composite structure of auxin response elements. *Plant Cell*. 1995;7(10):1611-1623. doi:10.1105/tpc.7.10.1611.

31. Ulmasov T, Murfett J, Hagen G, Guilfoyle TJ. Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell*. 1997;9(11):1963-1971. doi:10.1105/tpc.9.11.1963.

32. Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M. Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature*. 2001;414(6861):271-276. doi:10.1038/35104500.

33. Zenser N, Ellsmore a, Leasure C, Callis J. Auxin modulates the degradation rate of Aux/IAA proteins. *Proc Natl Acad Sci U S A*. 2001;98(20):11795-11800. doi:10.1073/pnas.211312798.

34. Finet C, Berne-Dedieu A, Scutt CP, Marletaz F. Evolution of the ARF Gene Family in Land Plants: Old Domains, New Tricks. *Mol Biol Evol*. 2013;30(1):45-56. doi:10.1093/molbev/mss220.

35. Hagen G, Guilfoyle T. Auxin-responsive gene expression: Genes, promoters and regulatory factors. *Plant Mol Biol*. 2002;49(3-4):373-385. doi:10.1023/A:1015207114117.

36. Kalluri UC, DiFazio SP, Brunner AM, et al. Genome-wide analysis of Aux/IAA and ARF gene families in Populus trichocarpa. *BMC Plant Biol*. 2007;7(1):59. doi:10.1186/1471-2229-7-59.

37. Wang D, Pei K, Fu Y, et al. Genome-wide analysis of the auxin response factors (ARF) gene family in rice (Oryza sativa). *Gene*. 2007;394(1):13-24. doi:10.1016/j.gene.2007.01.006.

38. Rensing SA, Lang D, Zimmer AD, et al. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*. 2008;319(5859):64-69. doi:10.1126/science.1150646.

39. Wang Y, Deng D, Shi Y, Miao N, Bian Y, Yin Z. Diversification, phylogeny and evolution of auxin response factor (ARF) family: insights gained from analyzing maize ARF genes. *Mol Biol Rep*. 2012;39(3):2401-2415. doi:10.1007/s11033-011-0991-z.

40. Banks JA, Nishiyama T, Hasebe M, et al. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science*. 2011;332(6032):960-963. doi:10.1126/science.1203810.

41. Weijers D, Wagner D. Transcriptional Responses to the Auxin Hormone. *Annu Rev Plant Biol*. 2016;67(1):annurev-arplant-043015-112122. doi:10.1146/annurev-arplant-043015-112122.

42. Calderón Villalobos LIA, Lee S, De Oliveira C, et al. A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. *Nat Chem Biol*. 2012;8(5):477-485. doi:10.1038/nchembio.926.

43. Vernoux T, Brunoud G, Farcot E, et al. The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol Syst Biol*. 2011;7:508. doi:10.1038/msb.2011.39.

44. Rademacher EH, Möller B, Lokerse AS, Llavata-Peris CI, van den Berg W, Weijers D. A cellular expression map of the Arabidopsis AUXIN RESPONSE FACTOR gene family. *Plant J*. 2011;68(4):597-606. doi:10.1111/j.1365-313X.2011.04710.x.

45. Lokerse AS, Weijers D. Auxin enters the matrix—assembly of response machineries for specific outputs. *Curr Opin Plant Biol*. 2009;12(5):520-526. doi:10.1016/j.pbi.2009.07.007.

46. Berleth T, Jürgens G. Embryonic pattern formation in flowering plants. *Development*. 1993;118:575-587. doi:10.1146/annurev.genet.28.1.351.

47. Okushima Y, Fukaki H, Onoda M, Theologis A, Tasaka M. ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in Arabidopsis. *Plant Cell*. 2007;19(1):118-130. doi:10.1105/tpc.106.047761.

48. Tabata R, Ikezaki M, Fujibe T, et al. Arabidopsis auxin response factor6 and 8 regulate jasmonic acid biosynthesis and floral organ development via repression of class 1 KNOX genes. *Plant Cell Physiol*. 2010;51(1):164-175. doi:10.1093/pcp/pcp176.

161

49. Rademacher EH, Lokerse AS, Schlereth A, et al. Different Auxin Response Machineries Control Distinct Cell Fates in the Early Plant Embryo. *Dev Cell*. 2012;22(1):211-222. doi:10.1016/j. devcel.2011.10.026.

50. Weigel D, Jürgens G. Stem cells that make stems. *Nature*. 2002;415(6873):751-754. doi:10.1038/415751a.

51. Scheres B. Stem-cell niches: nursery rhymes across kingdoms. *Nat Rev Mol Cell Biol*. 2007;8(5):345-354. doi:10.1038/nrm2164.

52. Scheres B, Wolkenfelt H, Willemsen V, et al. Embryonic origin of the Arabidopsis primary root and root meristem initials. *Development*. 1994;120(9):2475-2487.

53. Jurgens G, Mayer U. Arabidopsis. In: Jonathan Bard, ed. *EMBRYOS, Color Atlas of Development*. London: Wolfe Publishing; 1994:7-21.

54. Mashiguchi K, Tanaka K, Sakai T, et al. The main auxin biosynthesis pathway in Arabidopsis. *Proc Natl Acad Sci*. 2011;108(45):18512-18517. doi:10.1073/pnas.1108434108.

55. Cheng Y, Dai X, Zhao Y. Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in Arabidopsis. *Plant Cell*. 2007;19(8):2430-2439. doi:10.1105/tpc.107.053009.

56. Stepanova AN, Robertson-Hoyt J, Yun J, et al. TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell*. 2008;133(1):177-191. doi:10.1016/j. cell.2008.01.047.

57. Sohlberg JJ, Myrenås M, Kuusk S, et al. *STY1* regulates auxin homeostasis and affects apical-basal patterning of the Arabidopsis gynoecium. *Plant J*. 2006;47(1):112-123. doi:10.1111/j.1365-313X.2006.02775.x.

58. Eklund DM, Thelander M, Landberg K, et al. Homologues of the Arabidopsis thaliana SHI/STY/LRP1 genes control auxin biosynthesis and affect growth and development in the moss Physcomitrella patens. *Development*. 2010;137(8):1275-1284. doi:10.1242/dev.039594.

59. Eklund DM, Ståldal V, Valsecchi I, et al. The Arabidopsis thaliana STYLISH1 protein acts as a transcriptional activator regulating auxin biosynthesis. *Plant Cell*. 2010;22(2):349-363. doi:10.1105/tpc.108.064816.

60. Ståldal V, Sohlberg JJ, Eklund DM, Ljung K, Sundberg E. Auxin can act independently of *CRC*, *LUG*, *SEU*, *SPT* and *STY1* in style development but not apical-basal patterning of the *Arabidopsis* gynoecium. *New Phytol*. 2008;180(4):798-808. doi:10.1111/j.1469-8137.2008.02625.x.

61. Kuusk S, Sohlberg JJ, Long JA, Fridborg I, Sundberg E. STY1 and STY2 promote the formation of apical tissues during Arabidopsis gynoecium development. *Development*. 2002;129(20):4707-4717. http://www.ncbi.nlm.nih.gov/pubmed/12361963. Accessed August 16, 2016.

62. Eklund DM, Cierlik I, Ståldal V, et al. Expression of Arabidopsis SHORT INTERNODES/STYLISH family genes in auxin biosynthesis zones of aerial organs is dependent on a GCC box-like regulatory element. *Plant Physiol*. 2011;157(4):2069-2080. doi:10.1104/pp.111.182253.

63. Chandler JW, Cole M, Flier A, Grewe B, Werr W. The AP2 transcription factors DORNROSCHEN and DORNROSCHEN-LIKE redundantly control Arabidopsis embryo patterning via interaction with PHAVOLUTA. *Development*. 2007;134(9):1653-1662. doi:10.1242/dev.001016.

64. Emery JF, Floyd SK, Alvarez J, et al. Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol*. 2003;13(20):1768-1774. http://www.ncbi.nlm.nih.gov/pubmed/14561401. Accessed August 16, 2016.

65. Cole M, Chandler J, Weijers D, Jacobs B, Comelli P, Werr W. DORNROSCHEN is a direct target of the auxin response factor MONOPTEROS in the Arabidopsis embryo. *Development*. 2009;136(10):1643-1651. doi:10.1242/dev.032177.

66. Donner TJ, Sherr I, Scarpella E. Regulation of preprocambial cell state acquisition by auxin signaling in Arabidopsis leaves. *Development*. 2009;136(19):3235-3246. doi:10.1242/dev.037028.

67. Friml J, Vieten A, Sauer M, et al. Efflux-dependent auxin gradients establish the apical–basal axis of Arabidopsis. *Nature*. 2003;426(6963):147-153. doi:10.1038/nature02085.

68. Vanneste S, Friml J. Auxin: a trigger for change in plant development. *Cell*. 2009;136(6):1005-1016. doi:10.1016/j.cell.2009.03.001.

69. Robert HS, Friml J. Auxin and other signals on the move in plants. *Nat Chem Biol*. 2009;5(5):325-332. doi:10.1038/nchembio.170.

70. Benková E, Michniewicz M, Sauer M, et al. Local, Efflux-Dependent Auxin Gradients as a Common Module for Plant Organ Formation. *Cell*. 2003;115(5):591-602. doi:10.1016/S0092-8674(03)00924-3.

71. Blilou I, Xu J, Wildwater M, et al. The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. *Nature*. 2005;433(7021):39-44. doi:10.1038/nature03184.

72. Bennett SRM, Alvarez J, Bossinger G, Smyth DR. Morphogenesis in pinoid mutants of Arabidopsis thaliana. *Plant J*. 1995;8(4):505-520. doi:10.1046/j.1365-313X.1995.8040505.x.

73. Benjamins R, Quint A, Weijers D, Hooykaas P, Offringa R. The PINOID protein kinase regulates organ development in Arabidopsis by enhancing polar auxin transport. *Development*. 2001;128(20):4057-4067. http://www.ncbi.nlm.nih.gov/pubmed/11641228. Accessed August 16, 2016.

74. Friml J, Yang X, Michniewicz M, et al. A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. *Science*. 2004;306(5697):862-865. doi:10.1126/science.1100618.

75. Michniewicz M, Zago MK, Abas L, et al. Antagonistic regulation of PIN phosphorylation by PP2A and PINOID directs auxin flux. *Cell*. 2007;130(6):1044-1056. doi:10.1016/j.cell.2007.07.033.

76. Cheng Y, Qin G, Dai X, Zhao Y. NPY genes and AGC kinases define two key steps in auxin-mediated organogenesis in Arabidopsis. *Proc Natl Acad Sci*. 2008;105(52):21017-21022. doi:10.1073/pnas.0809761106.

77. Geldner N, Anders N, Wolters H, et al. The Arabidopsis GNOM ARF-GEF Mediates Endosomal Recycling, Auxin Transport, and Auxin-Dependent Plant Growth. *Cell*. 2003;112(2):219-230. doi:10.1016/S0092-8674(03)00003-5.

78. Steinmann T, Geldner N, Grebe M, et al. Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF. *Science*. 1999;286(5438):316-318. http://www.ncbi.nlm.nih.gov/pubmed/10514379. Accessed August 16, 2016.

79. Geldner N, Richter S, Vieten A, et al. Partial loss-of-function alleles reveal a role for GNOM in auxin transport-related, post-embryonic development of Arabidopsis. *Development*. 2004;131(2):389-

400. doi:10.1242/dev.00926.

80. Mayer U, Buttner G, Jurgens G. Apical-basal pattern formation in the Arabidopsis embryo: studies on the role of the gnom gene. *Development*. 1993;117(1):149-162.

81. Wolters H, Anders N, Geldner N, Gavidia R, Jürgens G. Coordination of apical and basal embryo development revealed by tissue-specific GNOM functions. *Development*. 2011;138(1):117-126. doi:10.1242/dev.059147.

82. Nawy T, Bayer M, Mravec J, Friml J, Birnbaum KD, Lukowitz W. The GATA factor HANABA TARANU is required to position the proembryo boundary in the early Arabidopsis embryo. *Dev Cell*. 2010;19(1):103-113. doi:10.1016/j.devcel.2010.06.004.

83. Zhao Y, Medrano L, Ohashi K, et al. HANABA TARANU is a GATA transcription factor that regulates shoot apical meristem and flower development in Arabidopsis. *Plant Cell*. 2004;16(10):2586-2600. doi:10.1105/tpc.104.024869.

84. Geisler M, Blakeslee JJ, Bouchard R, et al. Cellular efflux of auxin catalyzed by the Arabidopsis MDR/PGP transporter AtPGP1. *Plant J*. 2005;44(2):179-194. doi:10.1111/j.1365-313X.2005.02519.x.

85. Mravec J, Kubes M, Bielach A, et al. Interaction of PIN and PGP transport mechanisms in auxin distribution-dependent development. *Development*. 2008;135(20):3345-3354. doi:10.1242/dev.021071.

86. Barbez E, Kubeš M, Rolčík J, et al. A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants. *Nature*. 2012;485(7396):119-122. doi:10.1038/nature11001.

87. Chapman EJ, Estelle M. Mechanism of auxin-regulated gene expression in plants. *Annu Rev Genet*. 2009;43:265-285. doi:10.1146/annurev-genet-102108-134148.

88. Causier B, Ashworth M, Guo W, Davies B. The TOPLESS Interactome: A Framework for Gene Repression in Arabidopsis. *PLANT Physiol*. 2012;158(1):423-438. doi:10.1104/pp.111.186999.

89. Szemenyei H, Hannon M, Long JA. TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis. *Science*. 2008;319(5868):1384-1386. doi:10.1126/science.1151461.

90. Dharmasiri N, Dharmasiri S, Estelle M. The F-box protein TIR1 is an auxin receptor. *Nature*. 2005;435(7041):441-445. doi:10.1038/nature03543.

91. Kepinski S, Leyser O. The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature*. 2005;435(7041):446-451. doi:10.1038/nature03542.

92. Tiwari SB, Hagen G, Guilfoyle T. The roles of auxin response factor domains in auxin-responsive transcription. 2003;15(2). doi:10.1105/tpc.008417.

93. Sato A, Yamamoto KT. Overexpression of the non-canonical Aux/IAA genes causes auxin-related aberrant phenotypes in Arabidopsis. *Physiol Plant*. 2008;133(2):397-405. doi:10.1111/j.1399-3054.2008.01055.x.

94. Parry G, Calderon-Villalobos LI, Prigge M, et al. Complex regulation of the TIR1/AFB family of auxin receptors. *Proc Natl Acad Sci*. 2009;106(52):22540-22545. doi:10.1073/pnas.0911967106.

95. Tan X, Calderon-Villalobos LIA, Sharon M, et al. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature*. 2007;446(7136):640-645. doi:10.1038/nature05731.

96.   Calderon-Villalobos LI, Tan X, Zheng N, Estelle M. Auxin Perception--Structural Insights. *Cold Spring Harb Perspect Biol*. 2010;2(7):a005546-a005546. doi:10.1101/cshperspect.a005546.

97.   Reed JW. Roles and activities of Aux/IAA proteins in Arabidopsis. *Trends Plant Sci*. 2001;6(9):420-425. http://www.ncbi.nlm.nih.gov/pubmed/11544131. Accessed August 16, 2016.

98.   Hamann T, Benkova E, Bäurle I, Kientz M, Jürgens G. The Arabidopsis BODENLOS gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning. *Genes Dev*. 2002;16(13):1610-1615. doi:10.1101/gad.229402.

99.   Weijers D, Benkova E, Jäger KE, et al. Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators. *EMBO J*. 2005;24(10):1874-1885. doi:10.1038/sj.emboj.7600659.

100.   Ploense SE, Wu M-F, Nagpal P, Reed JW. A gain-of-function mutation in IAA18 alters Arabidopsis embryonic apical patterning. *Development*. 2009;136(9):1509-1517. doi:10.1242/dev.025932.

101.   Okushima Y, Overvoorde PJ, Arima K, et al. Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19. *Plant Cell*. 2005;17(2):444-463. doi:10.1105/tpc.104.028316.

102.   Hardtke CS, Berleth T. The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J*. 1998;17(5):1405-1411. doi:10.1093/emboj/17.5.1405.

103.   Schlereth A, Möller B, Liu W, et al. MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature*. 2010;464(7290):913-916. doi:10.1038/nature08836.

104.   Zhao Z, Andersen SU, Ljung K, et al. Hormonal control of the shoot stem-cell niche. *Nature*. 2010;465(7301):1089-1092. doi:10.1038/nature09126.

105.   Ikeda M, Ohme-Takagi M. A novel group of transcriptional repressors in Arabidopsis. *Plant Cell Physiol*. 2009;50(5):970-975. doi:10.1093/pcp/pcp048.

106.   Weijers D, Schlereth A, Ehrismann JS, Schwank G, Kientz M, Jürgens G. Auxin triggers transient local signaling for cell specification in Arabidopsis embryogenesis. *Dev Cell*. 2006;10(2):265-270. doi:10.1016/j.devcel.2005.12.001.

107.   De Rybel B, Möller B, Yoshida S, et al. A bHLH Complex Controls Embryonic Vascular Tissue Establishment and Indeterminate Growth in Arabidopsis. 2013;24(4). doi:10.1016/j.devcel.2012.12.013.

108.   Aida M, Beis D, Heidstra R, et al. The PLETHORA Genes Mediate Patterning of the Arabidopsis Root Stem Cell Niche. *Cell*. 2004;119(1):109-120. doi:10.1016/j.cell.2004.09.018.

109.   Galinha C, Hofhuis H, Luijten M, et al. PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature*. 2007;449(7165):1053-1057. doi:10.1038/nature06206.

110.   Smith ZR, Long JA. Control of Arabidopsis apical–basal embryo polarity by antagonistic transcription factors. *Nature*. 2010;464(7287):423-426. doi:10.1038/nature08843.

111.   Marhavý P, Bielach A, Abas L, et al. Cytokinin modulates endocytic trafficking of PIN1 auxin efflux carrier to control plant organogenesis. *Dev Cell*. 2011;21(4):796-804. doi:10.1016/j.devcel.2011.08.014.

112.   Dello Ioio R, Nakamura K, Moubayidin L, et al. A genetic framework for the control of cell divi-

sion and differentiation in the root meristem. *Science*. 2008;322(5906):1380-1384. doi:10.1126/science.1164147.

113. Mouchel CF, Osmont KS, Hardtke CS. BRX mediates feedback between brassinosteroid levels and auxin signalling in root growth. *Nature*. 2006;443(7110):458-461. doi:10.1038/nature05130.

114. Shani E, Yanai O, Ori N. The role of hormones in shoot apical meristem function. *Curr Opin Plant Biol*. 2006;9(5):484-489. doi:10.1016/j.pbi.2006.07.008.

115. Müller B, Sheen J. Cytokinin and auxin interaction in root stem-cell specification during early embryogenesis. *Nature*. 2008;453(7198):1094-1097. doi:10.1038/nature06943.

116. Bishopp A, Help H, El-Showk S, et al. A Mutually Inhibitory Interaction between Auxin and Cytokinin Specifies Vascular Pattern in Roots. *Curr Biol*. 2011;21(11):917-926. doi:10.1016/j.cub.2011.04.017.

117. Mähönen AP, Bishopp A, Higuchi M, et al. Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science*. 2006;311(5757):94-98. doi:10.1126/science.1118875.

118. THIMANN K V., KOEPFLI JB. Identity of the Growth-Promoting and Root-Forming Substances ofPlants. *Nature*. 1935;135(3403):101-102. doi:10.1038/135101a0.

119. Ljung K. Auxin metabolism and homeostasis during plant development. *Development*. 2013;140(5):943-950. doi:10.1242/dev.086363.

120. Grunewald W, Friml J. The march of the PINs: developmental plasticity by dynamic polar targeting in plant cells. *EMBO J*. 2010;29(16):2700-2714. doi:10.1038/emboj.2010.181.

121. Guilfoyle TJ, Hagen G. Auxin response factors. *Curr Opin Plant Biol*. 2007;10(5):453-460. doi:10.1016/j.pbi.2007.08.014.

122. Przemeck GH, Mattsson J, Hardtke C, Sung ZR, Berleth T. Studies on the role of the Arabidopsis gene MONOPTEROS in vascular development and plant cell axialization. *Planta*. 1996;200(2):229-237. doi:10.1007/BF00208313.

123. Ellis CM, Nagpal P, Young JC, Hagen G, Guilfoyle TJ, Reed JW. AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission in Arabidopsis thaliana. *Development*. 2005;132(20):4563-4574. doi:10.1242/dev.02012.

124. Kim J, Harter K, Theologis A. Protein-protein interactions among the Aux/IAA proteins. *Proc Natl Acad Sci U S A*. 1997;94(22):11786-11791. http://www.ncbi.nlm.nih.gov/pubmed/9342315. Accessed August 22, 2016.

125. Swaminathan K, Peterson K, Jack T. The plant B3 superfamily. *Trends Plant Sci*. 2008;13(12):647-655. doi:10.1016/j.tplants.2008.09.006.

126. Boer DR, Freire-Rios A, van den Berg WAM, et al. Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell*. 2014;156(3):577-589. doi:10.1016/j.cell.2013.12.027.

127. Ulmasov T, Hagen G, Guilfoyle TJ. Dimerization and DNA binding of auxin response factors. *Plant J*. 1999;19(3):309-319. http://www.ncbi.nlm.nih.gov/pubmed/10476078. Accessed August 22, 2016.

128. Yamasaki K, Kigawa T, Inoue M, et al. Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. *Plant Cell*. 2004;16(12):3448-3459.

doi:10.1105/tpc.104.026112.

129. Waltner JK, Peterson FC, Lytle BL, Volkman BF. Structure of the B3 domain from Arabidopsis thaliana protein At1g16640. *Protein Sci*. 2005;14(9):2478-2483. doi:10.1110/ps.051606305.

130. Holm L, Rosenström P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010;38(Web Server issue):W545-9. doi:10.1093/nar/gkq366.

131. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 12 Pt 1):2256-2268. doi:10.1107/S0907444904026460.

132. Adams-Cioaba MA, Li Z, Tempel W, et al. Crystal structures of the Tudor domains of human PHF20 reveal novel structural variations on the Royal Family of proteins. *FEBS Lett*. 2012;586(6):859-865. doi:10.1016/j.febslet.2012.02.012.

133. Huang Y, Fang J, Bedford MT, Zhang Y, Xu R-M. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science*. 2006;312(5774):748-751. doi:10.1126/science.1125162.

134. Russinova E, Borst J-W, Kwaaitaal M, et al. Heterodimerization and endocytosis of Arabidopsis brassinosteroid receptors BRI1 and AtSERK3 (BAK1). *Plant Cell*. 2004;16(12):3216-3229. doi:10.1105/tpc.104.025387.

135. Krogan NT, Ckurshumova W, Marcos D, Caragea AE, Berleth T. Deletion of MP/ARF5 domains III and IV reveals a requirement for Aux/IAA regulation in Arabidopsis leaf vascular patterning. *New Phytol*. 2012;194(2):391-401. doi:10.1111/j.1469-8137.2012.04064.x.

136. Godoy M, Franco-Zorrilla JM, Pérez-Pérez J, Oliveros JC, Lorenzo Ó, Solano R. Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. *Plant J*. 2011;66(4):700-711. doi:10.1111/j.1365-313X.2011.04519.x.

137. Yamaguchi N, Wu M-F, Winter CM, et al. A molecular framework for auxin-mediated initiation of flower primordia. *Dev Cell*. 2013;24(3):271-282. doi:10.1016/j.devcel.2012.12.017.

138. Lau S, Smet I De, Kolb M, Meinhardt H, Jürgens G. Auxin triggers a genetic switch. *Nat Cell Biol*. 2011;13(5):611-615. doi:10.1038/ncb2212.

139. Masel J, Siegal ML. Robustness: mechanisms and consequences. *Trends Genet*. 2009;25(9):395-403. doi:10.1016/j.tig.2009.07.005.

140. Brunoud G, Wells DM, Oliva M, et al. A novel sensor to map auxin response and distribution at high spatio-temporal resolution. *Nature*. 2012;482(7383):103-106. doi:10.1038/nature10791.

141. Smaczniak C, Immink RGH, Muiño JM, et al. Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc Natl Acad Sci U S A*. 2012;109(5):1560-1565. doi:10.1073/pnas.1112871109.

142. Nagpal P, Ellis CM, Weber H, et al. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development*. 2005;132(18):4107-4118. doi:10.1242/dev.01955.

143. Berger MF, Badis G, Gehrke AR, et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*. 2008;133(7):1266-1276. doi:10.1016/j.cell.2008.05.024.

144. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. Analysis of

167

homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*. 2008;133(7):1277-1289. doi:10.1016/j.cell.2008.05.023.

145. Ma PC, Rould MA, Weintraub H, Pabo CO. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell*. 1994;77(3):451-459. http://www.ncbi.nlm.nih.gov/pubmed/8181063. Accessed August 22, 2016.

146. Kim J, Struhl K. Determinants of half-site spacing preferences that distinguish AP-1 and ATF/CREB bZIP domains. *Nucleic Acids Res*. 1995;23(13):2531-2537. http://www.ncbi.nlm.nih.gov/pubmed/7630732. Accessed August 22, 2016.

147. Mangelsdorf DJ, Thummel C, Beato M, et al. The nuclear receptor superfamily: the second decade. *Cell*. 1995;83(6):835-839. http://www.ncbi.nlm.nih.gov/pubmed/8521507. Accessed August 22, 2016.

148. Khorasanizadeh S, Rastinejad F. Nuclear-receptor interactions on DNA-response elements. *Trends Biochem Sci*. 2001;26(6):384-390. http://www.ncbi.nlm.nih.gov/pubmed/11406412. Accessed August 22, 2016.

149. Umesono K, Murakami KK, Thompson CC, Evans RM. Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell*. 1991;65(7):1255-1266. http://www.ncbi.nlm.nih.gov/pubmed/1648450. Accessed August 22, 2016.

150. Zechel C, Shen XQ, Chambon P, Gronemeyer H. Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements. *EMBO J*. 1994;13(6):1414-1424. http://www.ncbi.nlm.nih.gov/pubmed/8137825. Accessed August 22, 2016.

151. Svergun D, Barberato C, Koch MHJ, IUCr. CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Crystallogr*. 1995;28(6):768-773. doi:10.1107/S0021889895007047.

152. De Rybel B, van den Berg W, Lokerse AS, et al. A Versatile Set of Ligation-Independent Cloning Vectors for Functional Studies in Plants. *PLANT Physiol*. 2011;156(3):1292-1299. doi:10.1104/pp.111.177337.

153. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 4):271-281. doi:10.1107/S0907444910048675.

154. Evans P, IUCr, T. KE, et al. Scaling and assessment of data quality. *Acta Crystallogr Sect D Biol Crystallogr*. 2006;62(1):72-82. doi:10.1107/S0907444905036693.

155. Thomas Pape, Thomas R. Schneider. HKL2MAP: a graphical user interface for phasing with SHELX programs. *Applyed Crystallogr*. 2004;37:843-844. http://proteincrystallography.org/programs/HKL2MAP/.

156. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997;53(Pt 3):240-255. doi:10.1107/S0907444996012255.

157. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 4):235-242. doi:10.1107/S0907444910045749.

158. DeLAno W.L. The Pymol Molecular Graphics System. http://www.pymol.org. Published 2002.

159.   Gouet P, Courcelle E, Stuart DI, Métoz F. ESPript: analysis of multiple sequence alignments in Post-Script. *Bioinformatics*. 1999;15(4):305-308. http://www.ncbi.nlm.nih.gov/pubmed/10320398. Accessed September 6, 2016.

160.   Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947-2948. doi:10.1093/bioinformatics/btm404.

161.   Konarev P V., Volkov V V., Sokolova A V., et al. PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr*. 2003;36(5):1277-1282. doi:10.1107/S0021889803012779.

162.   Berger MF, Bulyk ML. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc*. 2009;4(3):393-411. doi:10.1038/nprot.2008.195.

163.   Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. 1990;249(4968):505-510. doi:10.1126/science.2200121.

164.   Jolma A, Kivioja T, Toivonen J, et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res*. 2010;20(6):861-873. doi:10.1101/gr.100552.109.

165.   Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*. 2012;13(12):840-852. doi:10.1038/nrg3306.

166.   Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci*. 2014;111(6):2367-2372. doi:10.1073/pnas.1316278111.

167.   Yamamoto H, Yamamoto H, Nishihara T, Imagawa M. DNA Binding Specificity of the CCAAT/Enhancer-binding Protein Transcription Factor Family. *J Biol Chem*. 1996;271(7):3891-3896. doi:10.1074/jbc.271.7.3891.

168.   Merika M, Orkin SH. DNA-Binding Specificity of GATA Family Transcription Factors. *Mol Cell Biol*. 1993;13(7):3999-4010.

169.   Coulocheri SA, Pigis DG, Papavassiliou KA, Papavassiliou AG. Hydrogen bonds in protein–DNA complexes: Where geometry meets plasticity. *Biochimie*. 2007;89(11):1291-1303. doi:10.1016/j.biochi.2007.07.020.

170.   Sessions A, Nemhauser JL, McColl A, Roe JL, Feldmann KA, Zambryski PC. ETTIN patterns the Arabidopsis floral meristem and reproductive organs. *Development*. 1997;124(22):4481-4491. http://www.ncbi.nlm.nih.gov/pubmed/9409666. Accessed August 20, 2016.

171.   Zemlyanskaya E V., Wiebe DS, Omelyanchuk NA, Levitsky VG, Mironova V V. Meta-analysis of transcriptome data identified TGTCNN motif variants associated with the response to plant hormone auxin in Arabidopsis thaliana L. *http://dx.doi.org/101142/S0219720016410092*. 2016.

172.   O'Malley RC, Huang SC, Song L, et al. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*. 2016;165(5):1280-1292. doi:10.1016/j.cell.2016.04.038.

173.   Liu J, Hua W, Hu Z, et al. Natural variation in ARF18 gene simultaneously affects seed weight and silique length in polyploid rapeseed. doi:10.1073/pnas.1502160112.

174.   Korasick DA, Westfall CS, Lee SG, et al. Molecular basis for AUXIN RESPONSE FACTOR

protein interaction and the control of auxin response repression. *Proc Natl Acad Sci U S A*. 2014;111(14):5427-5432. doi:10.1073/pnas.1400074111.

175. Liao C-Y, Smet W, Brunoud G, Yoshida S, Vernoux T, Weijers D. Reporters for sensitive and quantitative measurement of auxin response. *Nat Methods*. 2015;12(3):207-210. doi:10.1038/nmeth.3279.

176. Oh E, Zhu JY, Bai MY, Arenhart RA, Sun Y, Wang ZY. Cell elongation is regulated through a central circuit of interacting transcription factors in the Arabidopsis hypocotyl. *Elife*. 2014;2014(3):1-19. doi:10.7554/eLife.03031.

177. Juanhuix J, Gil-Ortiz F, Cuní G, et al. Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J Synchrotron Radiat*. 2014;21(Pt 4):679-689. doi:10.1107/S160057751400825X.

178. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40(Pt 4):658-674. doi:10.1107/S0021889807021206.

179. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 12 Pt 1):2126-2132. doi:10.1107/S0907444904019158.

180. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 2007;35(Web Server issue):W375-83. doi:10.1093/nar/gkm216.

181. Suzuki M, Kao CY, McCarty DR. The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *Plant Cell*. 1997;9(5):799-807. doi:10.1105/tpc.9.5.799.

182. Lu R, Wang GG. Tudor: A versatile family of histone methylation "readers." *Trends Biochem Sci*. 2013;38(11):546-555. doi:10.1016/j.tibs.2013.08.002.

183. Gayatri S, Bedford MT. Readers of histone methylarginine marks. *Biochim Biophys Acta*. 2014;1839(8):702-710. doi:10.1016/j.bbagrm.2014.02.015.

184. Sumimoto H, Kamakura S, Ito T. Structure and function of the PB1 domain, a protein interaction module conserved in animals, fungi, amoebas, and plants. *Sci STKE*. 2007;2007(December 2007):re6.

185. Ito T, Matsui Y, Ago T, et al. Novel modular domain PB1 recognizes PC motif to mediate functional protein-protein interactions. *EMBO J*. 2001;20(15):3938-3946. doi:10.1093/emboj/20.15.3938.

186. Noda Y, Kohjima M, Izaki T, et al. Molecular Recognition in Dimerization between PB1 Domains. *J Biol Chem*. 2003;278(44):43516-43524. doi:10.1074/jbc.M306330200.

187. Moscat J, Diaz-Meco MT, Albert A, Campuzano S. Cell Signaling and Function Organized by PB1 Domain Interactions. *Mol Cell*. 2006;23(5):631-640.

188. Nauseef WM. The NADPH-dependent oxidase of phagocytes. *Proc Assoc Am Physicians*. 111(5):373-382. http://www.ncbi.nlm.nih.gov/pubmed/10519156. Accessed September 1, 2016.

189. Clark RA. Activation of the neutrophil respiratory burst oxidase. *J Infect Dis*. March 1999:S309-17. doi:10.1086/513849.

190. Ohno S. Intercellular junctions and cellular polarity: the PAR–aPKC complex, a conserved core cassette playing fundamental roles in cell polarity. *Curr Opin Cell Biol*. 2001;13(5):641-648. doi:10.1016/S0955-0674(00)00264-7.

191. Wodarz A. Establishing cell polarity in development. *Nat Cell Biol*. 2002;4(2):E39-E44. doi:10.1038/ncb0202-e39.

192. Etienne-Manneville S, Hall A. Cell polarity: Par6, aPKC and cytoskeletal crosstalk. *Curr Opin Cell Biol*. 2003;15(1):67-72. doi:10.1016/S0955-0674(02)00005-4.

193. Etienne-Manneville S, Hall A, Allen WE, et al. Integrin-Mediated Activation of Cdc42 Controls Cell Polarity in Migrating Astrocytes through PKCζ. *Cell*. 2001;106(4):489-498. doi:10.1016/S0092-8674(01)00471-8.

194. Guilfoyle TJ, Hagen G. Getting a grasp on domain III/IV responsible for Auxin Response Factor-IAA protein interactions. *Plant Sci*. 2012;190:82-88. doi:10.1016/j.plantsci.2012.04.003.

195. Nanao MH, Vinos-Poyo T, Brunoud G, et al. Structural basis for oligomerization of auxin transcriptional regulators. *Nat Commun*. 2014;5:3617. doi:10.1038/ncomms4617.

196. Wang S, Hagen G, Guilfoyle TJ. ARF-Aux/IAA interactions through domain III/IV are not strictly required for auxin-responsive gene expression. *Plant Signal Behav*. 2013;8(6):e24526. doi:10.4161/psb.24526.

197. Guilfoyle TJ. The PB1 domain in auxin response factor and Aux/IAA proteins: a versatile protein interaction module in the auxin response. *Plant Cell*. 2015;27(1):33-43. doi:10.1105/tpc.114.132753.

198. Ulmasov T, Hagen G, Guilfoyle TJ. Activation and repression of transcription by auxin-response factors. *Proc Natl Acad Sci U S A*. 1999;96(10):5844-5849. http://www.ncbi.nlm.nih.gov/pubmed/10318972. Accessed August 31, 2016.

199. Peterson CL, Workman JL. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr Opin Genet Dev*. 2000;10(2):187-192. doi:10.1016/S0959-437X(00)00068-X.

200. Espinosa-Soto C, Immink RG, Angenent GC, et al. Tetramer formation in Arabidopsis MADS domain proteins: analysis of a protein-protein interaction network. *BMC Syst Biol*. 2014;8(1):9. doi:10.1186/1752-0509-8-9.

201. Huang N, Chelliah Y, Shan Y, et al. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science*. 2012;337(6091):189-194. doi:10.1126/science.1222804.

202. Björklund S, Gustafsson CM. The yeast Mediator complex and its regulation. *Trends Biochem Sci*. 2005;30(5):240-244. doi:10.1016/j.tibs.2005.03.008.

203. Varaud E, Brioudes F, Szécsi J, et al. AUXIN RESPONSE FACTOR8 regulates Arabidopsis petal growth by interacting with the bHLH transcription factor BIGPETALp. *Plant Cell*. 2011;23(3):973-983. doi:10.1105/tpc.110.081653.

204. Shin R, Burch AY, Huppert KA, et al. The Arabidopsis transcription factor MYB77 modulates auxin signal transduction. *Plant Cell*. 2007;19(8):2440-2453. doi:10.1105/tpc.107.050963.

205. Berendzen KW, Weiste C, Wanke D, et al. Bioinformatic cis-element analyses performed in Arabidopsis and rice disclose bZIP- and MYB-related binding sites as potential AuxRE-coupling elements in auxin-mediated transcription. *BMC Plant Biol*. 2012;12(1):125. doi:10.1186/1471-2229-12-125.

206. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*. 2015;43(W1):W389-W394. doi:10.1093/nar/gkv332.

207. Cheng J, Sweredoski MJ, Baldi P. Accurate Prediction of Protein Disordered Regions by Mining

Protein Structure Data. http://www.ics.uci.edu/~pfbaldi. Accessed September 2, 2016.

208. Kato H, Ishizaki K, Kouno M, et al. Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. *PLoS Genet*. 2015;11(5). doi:10.1371/journal.pgen.1005084.

209. Garrett JJT, Meents MJ, Blackshaw MT, et al. A novel, semi-dominant allele of MONOPTEROS provides insight into leaf initiation and vein pattern formation. *Planta*. 2012;236(1):297-312. doi:10.1007/s00425-012-1607-0.

210. Ckurshumova W, Krogan NT, Marcos D, Caragea AE, Berleth T. Irrepressible, truncated auxin response factors: natural roles and applications in dissecting auxin gene regulation pathways. *Plant Signal Behav*. 2012;7(8):1027-1030. doi:10.4161/psb.20366.

211. Weijers D, Franke-van Dijk M, Vencken RJ, Quint A, Hooykaas P, Offringa R. An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. *Development*. 2001;128(21):4289-4299. http://www.ncbi.nlm.nih.gov/pubmed/11684664. Accessed August 21, 2016.

212. Piya S, Shrestha SK, Binder B, Stewart CN, Hewezi T. Protein-protein interaction and gene co-expression maps of ARFs and Aux/IAAs in Arabidopsis. *Front Plant Sci*. 2014;5:744. doi:10.3389/fpls.2014.00744.

213. Llavata-Peris CI. Proteomic and mechanistic analysis of Auxin Response Factors in the Arabidopsis embryo. 2013.

214. Smit ME, Weijers D. The role of auxin signaling in early embryo pattern formation. *Curr Opin Plant Biol*. 2015;28:99-105. doi:10.1016/j.pbi.2015.10.001.

215. Scarpella E, Marcos D, Friml J, Berleth T. Control of leaf vascular patterning by polar auxin transport. *Genes Dev*. 2006;20(8):1015-1027. doi:10.1101/gad.1402406.

216. Overvoorde P, Fukaki H, Beeckman T. Auxin control of root development. *Cold Spring Harb Perspect Biol*. 2010;2(6):a001537. doi:10.1101/cshperspect.a001537.

217. Casimiro I, Marchant A, Bhalerao RP, et al. Auxin Transport Promotes Arabidopsis Lateral Root Initiation. *Plant Cell*. 2001;13:843-852.

218. Friml J, Wiśniewska J, Benková E, Mendgen K, Palme K. Lateral relocation of auxin efflux regulator PIN3 mediates tropism in Arabidopsis. *Nature*. 2002;415(6873):806-809. doi:10.1038/415806a.

219. Krizek BA. Intronic sequences are required for AINTEGUMENTA-LIKE6 expression in Arabidopsis flowers. *BMC Res Notes*. 2015;8:556. doi:10.1186/s13104-015-1537-6.

220. Schauer SE, Schlüter PM, Baskar R, et al. Intronic regulatory elements determine the divergent expression patterns of AGAMOUS-LIKE6 subfamily members in Arabidopsis. *Plant J*. 2009;59(6):987-1000. doi:10.1111/j.1365-313X.2009.03928.x.

221. Hirsch N, Birnbaum RY. Dual Function of DNA Sequences: Protein-Coding Sequences Function as Transcriptional Enhancers. *Perspect Biol Med*. 2015;58(2):182-195. doi:10.1353/pbm.2015.0026.

222. Thumma BR, Matheson BA, Zhang D, et al. Identification of a Cis-acting regulatory polymorphism in a Eucalypt COBRA-like gene affecting cellulose content. *Genetics*. 2009;183(3):1153-1164. doi:10.1534/genetics.109.106591.

223. Chi Y, Yang Y, Zhou Y, et al. Protein–Protein Interactions in the Regulation of WRKY Transcription Factors. *Mol Plant*. 2013;6(2):287-300. doi:10.1093/mp/sst026.

224. Mann RS, Chan SK. Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet*. 1996;12(7):258-262. http://www.ncbi.nlm.nih.gov/pubmed/8763497. Accessed August 31, 2016.

225. Gonzales DH. General aspects of plant transcription factors. In: Gonzales DH, ed. *Plant Transcription Factors: Evolutionary, Structural and Functional Aspects*. Elsevier; 2016.

226. Sun X, Xue B, Jones WT, Rikkerink E, Dunker AK, Uversky VN. A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol Biol*. 2011;77(3):205-223. doi:10.1007/s11103-011-9803-z.

227. Sun XL, Jones WT, Rikkerink EHA, et al. GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling. *Biochem J*. 2012;442(1):1-12. doi:10.1042/BJ20111766.

228. Dai J-R, Liu B, Feng D-R, et al. MpAsr encodes an intrinsically unstructured protein and enhances osmotic tolerance in transgenic Arabidopsis. *Plant Cell Rep*. 2011;30(7):1219-1230. doi:10.1007/s00299-011-1030-1.

229. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic Disorder in Transcription Factors. *Biochemistry*. 2006;45(22):6873-6888. doi:10.1021/bi0602718.

230. Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K. Folding transition in the DMA-binding domain of GCN4 on specific binding to DNA. *Nature*. 1990;347(6293):575-578. doi:10.1038/347575a0.

231. Sigler PB. Acid blobs and negative noodles. *Nature*. 1988;333(6170):210-212. doi:10.1038/333210a0.

232. O'Neil KT, Hoess RH, DeGrado WF. Design of DNA-binding peptides based on the leucine zipper motif. *Science*. 1990;249(4970):774-778. doi:10.1126/science.2389143.

233. Beato M, Adler S, Waterman ML, et al. Gene regulation by steroid hormones. *Cell*. 1989;56(3):335-344. doi:10.1016/0092-8674(89)90237-7.

234. Htun H, Barsony J, Renyi I, Gould DL, Hager GL. Visualization of glucocorticoid receptor translocation and intranuclear organization in living cells with a green fluorescent protein chimera. *Proc Natl Acad Sci U S A*. 1996;93(10):4845-4850. http://www.ncbi.nlm.nih.gov/pubmed/8643491. Accessed August 31, 2016.

235. Fan X, Struhl K. Where Does Mediator Bind In Vivo? Jin D-Y, ed. *PLoS One*. 2009;4(4):e5029. doi:10.1371/journal.pone.0005029.

236. Fan X, Chou DM, Struhl K. Activator-specific recruitment of Mediator in vivo. *Nat Struct Mol Biol*. 2006;13(2):117-120. doi:10.1038/nsmb1049.

237. Seguela-Arnaud M, Smith C, Uribe MC, et al. The Mediator complex subunits MED25/PFT1 and MED8 are required for transcriptional responses to changes in cell wall arabinose composition and glucose treatment in Arabidopsis thaliana. *BMC Plant Biol*. 2015;15(1):215. doi:10.1186/s12870-015-0592-4.

238. Gillmor CS, Park MY, Smith MR, Pepitone R, Kerstetter RA, Poethig RS. The MED12-MED13 module of Mediator regulates the timing of embryo patterning in Arabidopsis. *Development*.

2010;137(1):113-122. doi:10.1242/dev.043174.

239.   Kim MJ, Jang I-C, Chua N-H. The Mediator Complex MED15 Subunit Mediates Activation of Downstream Lipid-Related Genes by the WRINKLED1 Transcription Factor. *Plant Physiol*. 2016;171(3):1951-1964. doi:10.1104/pp.16.00664.

240.   Abel S, Nguyen MD, Theologis A. The PS-IAA4/5-like family of early auxin-inducible mRNAs in Arabidopsis thaliana. *J Mol Biol*. 1995;251(4):533-549. doi:10.1006/jmbi.1995.0454.

241.   Kim B, Soh M, Kang BJ, Furuya M, Nam H. Two dominant morphogenetic mutations of Arabidopsis thaliana identified as suppressor mutations of hy2. *Plant J*. 1996;9(4):441-456.

242.   Uehara T, Okushima Y, Mimura T, Tasaka M, Fukaki H. Domain II mutations in CRANE/IAA18 suppress lateral root formation and affect shoot development in Arabidopsis thaliana. *Plant Cell Physiol*. 2008;49(7):1025-1038. doi:10.1093/pcp/pcn079.

243.   Odat O, Gardiner J, Sawchuk MG, Verna C, Donner TJ, Scarpella E. Characterization of an allelic series in the MONOPTEROS gene of arabidopsis. *genesis*. 2014;52(2):127-133. doi:10.1002/dvg.22729.

244.   Wendrich JR, Boeren S, Moller BK, Wiejers D, De Rybel B. In vivo identification of plant protein complexes using IP-MS/MS. *Methods Mol Biol*.

245.   Amos WB, White JG. How the confocal laser scanning microscope entered biological research. *Biol Cell*. 2003;95(6):335-342. http://www.ncbi.nlm.nih.gov/pubmed/14519550. Accessed August 30, 2016.

246.   Cox G, Sheppard CJR. Practical limits of resolution in confocal and non-linear microscopy. *Microsc Res Tech*. 2004;63(1):18-22. doi:10.1002/jemt.10423.

247.   Betzig E, Patterson GH, Sougrat R, et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*. 2006;313(5793):1642-1645. doi:10.1126/science.1127344.

248.   Hess ST, Girirajan TPK, Mason MD. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys J*. 2006;91(11):4258-4272. doi:10.1529/biophysj.106.091116.

249.   Bastiaens PI, Pepperkok R. Observing proteins in their natural habitat: the living cell. *Trends Biochem Sci*. 2000;25(12):631-637. http://www.ncbi.nlm.nih.gov/pubmed/11116191. Accessed August 30, 2016.

250.   Bücherl CA, Bader A, Westphal AH, Laptenok SP, Borst JW. FRET-FLIM applications in plant systems. *Protoplasma*. 2014;251(2):383-394. doi:10.1007/s00709-013-0595-7.

251.   Robert M Clegg. No TitleFluorescence resonance energy transfer. In: XF W, B H, eds. *Fluorescence Imaging Spectroscopy and Microscopy (Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications)*. John Wiley & Sons, Ltd; 1996:179-252.

252.   Stryer L. Fluorescence energy transfer as a spectroscopic ruler. *Annu Rev Biochem*. 1978;47:819-846. doi:10.1146/annurev.bi.47.070178.004131.

253.   Borst JW, Visser AJWG. Fluorescence lifetime imaging microscopy in life sciences. *Meas Sci Technol*. 2010;21(10):102002. doi:10.1088/0957-0233/21/10/102002.

254.   Lakowicz J. *Principles of Fluorescence Spectroscopye*. New York, NY: Springer; 2006.

255.   Valeur B, Berberan-Santos MN. *Molecular Fluorescence Principles and Applications*. Wiley-VCH;

2013.

256.   Clegg RM. Chapter 1 Förster resonance energy transfer—FRET what is it, why do it, and how it's done. In: ; 2009:1-57. doi:10.1016/S0075-7535(08)00001-6.

257.   Sheen J. Signal Transduction in Maize and Arabidopsis Mesophyll Protoplasts. *PLANT Physiol*. 2001;127(4):1466-1475. doi:10.1104/pp.010820.

258.   Bücherl C, Aker J, de Vries S, Borst JW. Probing protein-protein Interactions with FRET-FLIM. *Methods Mol Biol*. 2010;655:389-399. doi:10.1007/978-1-60761-765-5_26.

259.   Wu F-H, Shen S-C, Lee L-Y, Lee S-H, Chan M-T, Lin C-S. Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods*. 2009;5:16. doi:10.1186/1746-4811-5-16.

260.   Becker W, Bergmann A, Hink MA, König K, Benndorf K, Biskup C. Fluorescence lifetime imaging by time-correlated single-photon counting. *Microsc Res Tech*. 2004;63(1):58-66. doi:10.1002/jemt.10421.

261.   Goedhart J, von Stetten D, Noirclerc-Savoye M, et al. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat Commun*. 2012;3:751. doi:10.1038/ncomms1738.

262.   Wendrich JR, Liao C-Y, van den Berg WAM, De Rybel B, Weijers D. Ligation-independent cloning for plant research. *Methods Mol Biol*. 2015;1284:421-431. doi:10.1007/978-1-4939-2444-8_21.

263.   Laptenok SP, Snellenburg JJ, Bücherl CA, Konrad KR, Borst JW. Global analysis of FRET-FLIM data in live plant cells. *Methods Mol Biol*. 2014;1076:481-502. doi:10.1007/978-1-62703-649-8_21.

264.   Warren SC, Margineanu A, Alibhai D, et al. Rapid global fitting of large fluorescence lifetime imaging microscopy datasets. *PLoS One*. 2013;8(8):e70687. doi:10.1371/journal.pone.0070687.

265.   Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Overview of Extracellular Signaling. 2000.

266.   Giatti S, Melcangi RC, Pesaresi M. The other side of progestins: effects in the brain. *J Mol Endocrinol*. 2016;57(2):R109-R126. doi:10.1530/JME-16-0061.

267.   Caprio M, Infante M, Calanchini M, Mammi C, Fabbri A. Vitamin D: not just the bone. Evidence for beneficial pleiotropic extraskeletal effects. *Eat Weight Disord - Stud Anorexia, Bulim Obes*. August 2016:1-15. doi:10.1007/s40519-016-0312-6.

268.   Prigge MJ, Greenham K, Zhang Y, et al. The Arabidopsis Auxin Receptor F-Box Proteins AFB4 and AFB5 Are Required for Response to the Synthetic Auxin Picloram. *G3 Genes|Genomes|Genetics*. 2016;6(5):1383-1390. doi:10.1534/g3.115.025585.

269.   Dharmasiri N, Dharmasiri S, Weijers D, et al. Plant Development Is Regulated by a Family of Auxin Receptor F Box Proteins. *Dev Cell*. 2005;9(1):109-119. doi:10.1016/j.devcel.2005.05.014.

270.   Zhang J-Y, He S-B, Li L, Yang H-Q. Auxin inhibits stomatal development through MONOPTEROS repression of a mobile peptide gene STOMAGEN in mesophyll. *Proc Natl Acad Sci U S A*. 2014;111(29):E3015-23. doi:10.1073/pnas.1400542111.

271.   Zhao Z, Andersen SU, Ljung K, et al. Hormonal control of the shoot stem-cell niche. *Nature*. 2010;465(7301):1089-1092. doi:10.1038/nature09126.

272.   Mallory AC, Bartel DP, Bartel B. MicroRNA-directed regulation of Arabidopsis AUXIN RE-SPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell*. 2005;17(5):1360-1375. doi:10.1105/tpc.105.031716.

273.   Hawkins C, Liu Z. A model for an early role of auxin in Arabidopsis gynoecium morphogenesis. *Front Plant Sci*. 2014;5:327. doi:10.3389/fpls.2014.00327.

274.   Muiño JM, Smaczniak C, Angenent GC, Kaufmann K, van Dijk ADJ. Structural determinants of DNA recognition by plant MADS-domain transcription factors. *Nucleic Acids Res*. 2014;42(4):2138-2146. doi:10.1093/nar/gkt1172.

275.   Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci*. 2008;33(5):220-229. doi:10.1016/j.tibs.2008.02.002.

276.   Davis RL, Cheng P-F, Lassar AB, et al. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell*. 1990;60(5):733-746. doi:10.1016/0092-8674(90)90088-V.

277.   Näär AM, Boutin J-M, Lipkin SM, et al. The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. *Cell*. 1991;65(7):1267-1279. doi:10.1016/0092-8674(91)90021-P.

278.   Gronemeyer H. Transcription Activation by Estrogen and Progesterone Receptors. *Annu Rev Genet*. 1991;25(1):89-123. doi:10.1146/annurev.ge.25.120191.000513.

279.   Soufi A, Garcia MF, Jaroszewicz A, et al. Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell*. 2015;161(3):555-568. doi:10.1016/j.cell.2015.03.017.

280.   Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012;13(9):613-626. doi:10.1038/nrg3207.

281.   Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*. 2011;25(21):2227-2241. doi:10.1101/gad.176826.111.

282.   Bartke T, Vermeulen M, Xhemalce B, et al. Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. *Cell*. 2010;143(3):470-484. doi:10.1016/j.cell.2010.10.012.

283.   Utley RT, Owen-Hughes TA, Juan L-J, Côté J, Adams CC, Workman JL. In vitro analysis of transcription factor binding to nucleosomes and nucleosome disruption/displacement. *Methods Enzymol*. 1996;274:276-291. doi:10.1016/S0076-6879(96)74024-7.

284.   Perlmann T, Wrange O. Specific glucocorticoid receptor binding to DNA reconstituted in a nucleosome. *EMBO J*. 1988;7(10):3073-3079. http://www.ncbi.nlm.nih.gov/pubmed/2846275. Accessed September 9, 2016.

285.   Pajoro A, Madrigal P, Muiño JM, et al. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol*. 2014;15(3):R41. doi:10.1186/gb-2014-15-3-r41.

286.   McPherson CE, Shim E-Y, Friedman DS, et al. An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. *Cell*. 1993;75(2):387-398. doi:10.1016/0092-8674(93)80079-T.

287.   Krogan NT, Hogan K, Long JA. APETALA2 negatively regulates multiple floral organ iden-

tity genes in Arabidopsis by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19. *Development*. 2012;139(22):4180-4190. doi:10.1242/dev.085407.

288.   Calderón Villalobos LIA, Lee S, De Oliveira C, et al. A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. Nat Chem Biol. 2012;8(5):477-485. doi:10.1038/nchembio.926.

Summary

Auxin is a plant hormone that triggers a broad variety of responses during plant development. These responses range from correct cell division patterns during embryogenesis to formation and growth of different organs. Due to its importance for plant growth and development, many aspects of the biology of auxin have been studied. In **Chapter 2**, we use *Arabidopsis* embryogenesis as a stage to describe generalities about its biosynthesis, transport, components of its signaling pathway and transcriptional control of some know target genes.

As most of the players involved in transcriptional regulation in response to auxin have been identified, the question of how the same signal can elicit so many different responses remains open. In this thesis we approach this issue by focusing on the ultimate effectors of the auxin signaling pathway: the ARF family of transcription factors. In **Chapter 3** we present the crystal structure of the DNA binding Domain (DBD) of two divergent members of the family: ARF1 and ARF5. Careful observation of the structures, followed by *in vitro* and *in vivo* experiments led to the following conclusions: 1) ARF DBDs dimerize through a conserved alpha-helix, and bind cooperatively to an inverted repeat of the canonical TGTCTC AuxRE. Dimerization of this domain is important for high-affinity DNA binding and *in vivo* activity. 2) Monomeric ARFs have the same binding preference for the DNA sequence TGTCGG (determined by protein binding microarray). 3) DNA-contacting residues are almost completely conserved within the ARF family members. 4) The distance between the AuxREs may play a role for binding of specific ARF dimers as for example, ARF5 can accommodate and bind to different spacing (6-9 bp) compared to ARF1 which is more rigid (7-8 bp).

In **Chapter 4** we follow up on the observations made. First we again used structural biology to determine the reason of the high binding affinity to the TGTCGG sequence compared to the previously identified canonical TGTCTC element. We found that in complex with TGTCGG, His137 (ARF1) could rotate and make hydrogen bonds with either G5 or G6, as well as a hydrogen bond with the C opposing to G6. This rotation is not possible when in complex with TGTCTC and there the same histidine can make only one hydrogen bond with the G opposing to C6. We conclude then that this histidine plays a role in determining the strength of binding to TGTCNN elements and that this also reflects in its specific transcriptional activity as mutating the corresponding histidine in ARF5 renders a semi-functional protein *in vivo* (Chapter 3).

The next observation we followed up in **Chapter 4** is the biological meaning of ARF DBD cooperative binding to DNA. We identified AuxRE inverted repeats (IR) in the promoter of the *TMO5* gene and mutated them. This brought the expression of the gene to very low levels despite the presence of other multiple single AuxREs. Thus, the single inverted AuxRE repeat in the *TMO5* promoter is essential for ARF5 binding and gene regulation. Importantly, mutating only a single AuxRE element within the inverted repeat led to very pronounced loss of activity, consistent with requirement of both AuxRE sites for high-affinity ARF5 binding. We then concluded that IR AuxREs have a significant effect in gene regulation by ARFs. Next we search the genome for bipartite AuxREs that correlated to

auxin response and found two main elements: inverted repeat with 8 bases of spacing (IR8) and direct repeat with 5 bases of spacing (DR5). As this kind of bipartite AuxREs are rarer to find than single AuxREs, we tested their presence in promoters as predictors of auxin responsiveness by qPCR. We found that about 75% of the selected genes containing either IR8 or DR5 responded to auxin. The expression study also show that genes containing the DR5 sequence were only up-regulated when regulated. Interestingly, Surface Plasmon Resonance study showed that only class A (activator) ARFs can bind the DR5 sequence cooperatively.

As the structural differences of ARFs DBDs are subtle, we then asked if specific gene targeting is determined by this domain alone. In **Chapter 5** we used a DBD swap experiment and conclude that the DBD is necessary for specific gene targeting but not sufficient and the other domains of an ARF also contribute in its specific activity.

In **Chapter 5** we expand our focus from the DBD to the other ARF domains, Middle Region (MR) and C-terminal (CT). As ARFs have protein-protein interaction interfaces in all three domains, we expressed the isolated domains of ARF5 and perform immuno-precipitation followed by tandem mass-spectrometry. Although the procedure needs optimization, some interactions expected for each domain could be identified. The DBD showed to interact with the general transcription machinery and the CT could interact with another ARF and 3 Aux/IAA. These interactions seem to be specific as the Aux/IAA recovered are not the most abundant in the sampled tissue.

Finally, in **Chapter 6** all the obtained results are put in a broader context and new questions derived from our results are proposed.

Samenvatting

Auxine is een hormoon dat veel verschillende reacties in gang zet tijdens de ontwikkeling van een plant. Deze reacties variëren van juiste celdelingspatronen tijdens de embryogenese tot de vorming en groei van diverse organen. Omdat het zo belangrijk is voor plantengroei en –ontwikkeling is de werkwijze van auxine uitvoerig bestudeerd. In **Hoofdstuk 2** gebruiken we de embryogenese van *Arabidopsis* om de algemene principes te beschrijven van auxine biosynthese, transport, de componenten van de signaleringsroute en transcriptionele controle van enkele bekende target genen. Hoewel de meeste eiwitten die betrokken zijn bij auxine-gereguleerde transcriptionele respons bekend zijn, blijft de vraag hoe een enkel molecuul zoveel verschillende processen in gang kan zetten. In dit proefschrift is gepoogd een antwoord op deze vraag te vinden door te focussen op de eiwitten die uiteindelijk de signalen van de auxine pathway omzetten in daden: de familie van ARF transcriptie-factoren. In **Hoofdstuk 3** beschrijven we de kristalstructuur van het DNA-bindende domein (DBD) van twee verschillende ARFs: ARF1 en ARF5. Nauwkeurige observatie van de structuren, gevolgd door *in vitro* and *in vivo* experimenten, heeft tot de volgende conclusies geleid: 1) ARF DBD's vormen een dimeer door middel van een evolutionair geconserveerde alfa-helix en binden als dimeer aan een complexe bindingsplaats in het DNA – een "inverted repeat" van de bekende TGTCTC AuxRE sequentie. Dimerizatie van dit domein is belangrijk voor goede binding aan het DNA en voor *in vivo* activiteit. 2) Verschillende ARF monomeren hebben dezelfde bindingsvoorkeur voor de DNA sequentie TGTCGG. 3) De aminozuren die contact maken met het DNA zijn vrijwel volledig geconserveerd binnen de ARF familie. 4) De afstand tussen de AuxRE's in de inverted repeat zou een rol kunnen spelen in de specificiteit van ARF dimeer binding, omdat bijvoorbeeld ARF5 zich kan aanpassen aan verschillende afstanden (6-9bp), terwijl ARF1 minder flexibel is (7-8 bp).

In **Hoofdstuk 4** bouwen we voort op de observaties uit het voorgaande werk. Allereerst hebben we opnieuw structuurbiologie gebruikt om te bepalen waarom ARFs sterker aan de TGTCGG sequentie binden dan aan het bekende TGTCTC element. We vonden dat in complex met TGTCGG de His137 van ARF1 kan draaien en waterstofbruggen kan vormen met ofwel G5 of G6. Daarnaast kan het een waterstofbrug vormen met de C tegenover G6. Deze draaiing is niet mogelijk in complex met TGTCTC en dezelde Histidine kan slechts één waterstofbrug maken met de G tegenover C6. We concluderen daarom dat deze Histidine een rol speelt in het bepalen van de bindingssterkte met TGTCNN elementen. Dit heeft ook invloed op specifieke transcriptionele activiteit, omdat het muteren van de histidine in ARF5 resulteert in een eiwit met onvolledige *in vivo* activiteit (Hoofdstuk 3).

De volgende observatie waar we in **Hoofdstuk 4** mee verder zijn gegaan is het biologische belang van ARF-DBD dimeer binding aan DNA. In de promoter van het *TMO5* gen – een direct target van ARF5 – komt een inverted repeat (IR) voor. Door het aanbrengen van mutaties in deze Inverted Repeat konden we vaststellen dat het gen nog maar erg zwak tot expressie kwam, ondanks de aanwezigheid van meerdere enkelvoudige AuxRE's. De inverted repeat AuxRE in de *TMO5* promoter is dus essentieel voor ARF5-binding en genregulatie. Zelfs mutatie van slechts één van de twee AuxRE sites in de repeat leidde tot sterk verlies van activiteit, wat overeenkomt met het model dat beide AuxRE sites in de inverted repeat nodig

zijn voor sterke binding van een ARF5 dimeer. Vervolgens doorzochten we het *Arabidopsis* genoom voor soortgelijke dubbele AuxRE's die samenhingen met auxine respons en vonden twee veel-voorkomende elementen: inverted repeats met 8 baseparen ertussen (IR8) en een direct repeat met 5 baseparen ertussen (DR5). Omdat deze dubbele AuxRE's moeilijker te vinden zijn dan enkele AuxRE's, hebben we met qPCR getest of hun aanwezigheid in promoters voorspelt hoe de promoter reageert op auxine. We vonden dat ongeveer 75% van de geselecteerde genen met ofwel IR8 of DR5 reageerden op auxine. De expressiestudie liet ook zien dat, terwijl genen met een IR8 element zowel geactiveerd of onderdruk kunnen worden door auxine, genen met DR5 alleen opgereguleerd werden als ze reageerden. Surface Plasmon Resonance studies lieten zien dat alleen klasse A (activator) ARFs het DR5 motief kunnen binden, het geen de exclusieve activatie door auxine kan verklaren.

Omdat de structurele verschillen tussen ARF DBD's subtiel zijn, is de vraag of genspecificiteit alleen bepaald wordt door dit domein. In hoofdstuk 5 hebben we een DBD uitwisselingsexperiment gebruikt (tussen de nauw verwanta ARF5 en ARF6 eiwitten), waaruit bleek dat het DBD nodig is voor het selecteren van specifieke genen, maar niet genoeg is en dat andere domeinen van ARF ook aan de specificiteit bijdragen.

In **Hoofdstuk 5** breiden we onze focus uit van het DBD naar de andere ARF domeinen: Middle Region (MR) en C-terminal (CT). Omdat ARF's eiwit-eiwit interactieplekken hebben in al deze domeinen, zijn individuele domeinen van ARF5 tot expressie gebracht en gebruikt voor immuun-precipitatie gevolgd door massa-spectrometrie. Ondanks dat deze methode moet worden geoptimaliseerd voor ARF eiwitten, konden verschillende bekende interactoren worden geïdentificeerd voor ieder domein. Het DBD interacteert met de algemene transcriptie complexen en de CT kan interacteren met een andere ARF en 3 Aux/IAA eiwitten. Deze interacties lijken specifiek, omdat de Aux/IAA die gevonden zijn niet de meest veel-voorkomende zijn in het weefsel dat bestudeerd is.

Tot slot worden alle gevonden resultaten in **Hoofdstuk 6** in een bredere context geplaatst en worden op basis van deze resultaten nieuwe vragen voorgesteld.

Resumen

La auxina es una hormona vegetal que desencadena una gran variedad de respuestas durante el desarrollo de una planta. Estas respuestas van desde los patrones correctos de división celular durante la embriogénesis, hasta la formación y crecimiento de diferentes órganos. Debido a su importancia para el crecimiento y desarrollo de plantas, muchos aspectos de la biología de la auxina han sido estudiados. En el **Capítulo 2**, usamos la embriogénesis de *Arabidopsis* como marco para describir generalidades sobre su biosíntesis, transporte, los componentes de su cascada de señalización y el control transcripcional de algunos de sus genes blanco conocidos.

A pesar de que muchos de los componentes involucrados en la regulación transcripcional en respuesta a la auxina han sido identificados, cómo la misma señal puede desencadenar tal variedad de respuestas aún se desconoce. En esta tesis buscamos la respuesta a esta pregunta enfocándonos en los efectores finales de la cascada de señalización de auxina: la familia de factores de transcripción, ARF. En el **Capítulo 3** presentamos la estructura cristalográfica del Dominio de unión al ADN (DBD) de dos miembros divergentes de la familia: ARF1 y ARF5. La observación cuidadosa de las estructuras seguida de experimentos *in vivo* e *in vitro* llevaron a las siguientes conclusiones: 1) El DBD de los ARFs forman dímeros a través de una hélice-alfa conservada y se unen de forma cooperativa a la repetición invertida del AuxRE canónico, TGTCTC. La dimerización de este dominio es importante para la unión al ADN con alta afinidad y actividad *in vivo*. 2) Los monómeros de ARFs tienen la misma preferencia de unión por la secuencia de ADN TGTCGG (determinado por micro arreglos de unión de proteína). 3) Los residuos que hacen contacto con el AND están casi completamente conservados entre los miembros de la familia. 4) La distancia entre los AuxREs podría tener un rol en la unión de dímeros específicos de ARFs ya que por ejemplo ARF5 se puede acomodar y unir a distintos espacios (6-9 pb) comparado con ARF1 que es más rígido (7-8 pb).

En el **Capítulo 4** profundizamos en las observaciones hechas. Primero hacemos uso nuevamente de biología estructural para determinar la razón de la alta afinidad de unión a la secuencia TGTCGG en comparación con la secuencia canónica previamente identificada TGTCTC. Encontramos que en complejo con TGTCGG, la His137 (ARF1) puede rotar y formar puentes de hidrogeno con G5 o G6, así como también un puente de hidrogeno con la C opuesta a G6. Esta rotación no es posible en complejo con TGTCTC, donde la misma histidina sólo puede formar un puente de hidrógeno con la G opuesta a C6. Concluimos que esta histidina juega un rol en determinar la fuerza de unión a los elementos TGTCNN y que esto se refleja en su actividad trasncripcional específica, ya que la mutación de la histidina correspondiente en ARF5 resulta en una proteína parcialmente funcional *in vivo* (**Capitulo 3**).

La siguiente observación en la que profundizamos en el **Capítulo 4** es el significado biológico de la unión cooperativa del DBD de ARF al ADN. Identificamos repeticiones invertidas de AuxREs en la región promotora del gen *TMO5* y las mutamos. Esto resultó en la expresión del gen a niveles muy bajos a pesar de la presencia de muchos otros AuxRE individuales. Por

lo tanto, la única repetición invertida de AuxRE en el promotor de *TMO5* es esencial para la unión de ARF5 y regulación génica. Cabe recalcar que la mutación de sólo uno de los elementos AuxRE de la repetición invertida resulta en una pronunciada caída de actividad, consistente con el requerimiento de ambos elementos AuxRE para la unión de ARF5 con alta afinidad. Con esto concluimos que las repeticiones invertidas de AuxRE tienen un efecto significativo en la regulación génica por ARFs. Seguidamente buscamos en el genoma elementos AuxRE bipartita que tengan correlación con respuesta a auxina y encontramos dos elementos principales: repeticiones invertidas con un espacio de 8 bases (IR8) y repeticiones directas con un espacio de 5 bases (DR5). Ya que estos tipos de AuxRE bipartita son más difíciles de encontrar que AuxRE individuales probamos su presencia en promotores como indicadores de sensibilidad a auxina por qPCR. Encontramos que aproximadamente 75% de los genes seleccionados que contienen IR8 o DR5 responden a la auxina. Los estudios de expresión también mostraron que los genes que contienen la secuencia DR5 sólo respondían a la auxina elevando sus niveles. Cabe mencionar que estudios de resonancia plasmónica de superficie (SPR) mostraron que sólo ARFs de clase A (activadores) pueden unirse a la secuencia DR5 de forma cooperativa.

Debido a que las diferencias estructurales de los DBDs de ARFs son sutiles, nos preguntamos si la especificidad por genes es determinada solamente por este dominio. En el Capítulo 5 usamos un experimento donde intercambiamos DBDs y concluimos que el DBD es necesario pero no suficiente para la selección específica de genes y que los otros dominios de los ARFs también contribuyen a su actividad específica.

En el **Capítulo 5** expandimos nuestro enfoque del DBD hacia los otros dominios de los ARFs, la Región Media (MR) y el dominio C-terminal (CT). Ya que los ARFs tienen interfaces para interacciones proteína - proteína en sus tres dominios, expresamos los dominios aislados de ARF5 y realizamos inmuno - precipitación seguida por espectrometría de masa en tándem. A pesar de que el procedimiento necesita optimización, algunas interacciones esperadas para cada dominio pudieron ser identificadas. EL DBD mostró interacciones con la maquinaria general de transcripción y el CT pudo interactuar con otro ARF y 3 Aux/IAAs. Estas interacciones parecen ser específicas ya que las Aux/IAAs recuperadas no son las más abundantes en el tejido muestreado.

Finalmente, en el Capítulo 6 todos los resultados obtenidos son puestos en un contexto más amplio y nuevas interrogantes resultantes de nuestros resultados son propuestas.

Acknowledgements

I want to thank every person that helped me in the process of getting the PhD. It would have been impossible to complete without the support system of colleagues, friends and family.

First of all, thanks to Dolf Weijers for the supervision and guidance. It was a privilege to learn from you. As a master student I already had a lot of fun doing my thesis in your group. You managed to create a good group of people and a very positive atmosphere. It was an honor to be part of the team that you lead, with your example, on how to be a good professional and person. Thank you for always having your door open.

To all the members of the WUR Plant Development group, I would need a second book to list all the nice memories and stories I have with each one of you. Thank you for the great atmosphere, the scientific and non-scientific conversations. All of you have wonderful minds, big hearts and a fantastic sense of humor. I consider you all as friends. I specially want to thank Tanya, Maritza, Bert and Jos for being there to give me a hand when expecting Jacob didn't let me do all by myself at the lab. And of course Willy, it was fun sharing a project with you and I learned a lot from it.

To the rest of the Biochemistry family thank you too. We had very good times together that I will always remember.

To all the students I had the honor to work with. Specially André, with whom I worked during the last year of the PhD. You were a great support, I was very lucky to have you working in my project and I wish you the best in your scientific career and in life!

For all the people who helped me get the necessary time to write this thesis, especially Marja and Arjan. I wouldn't have been able to finish this without all your help. Niels, thank you very much for the time you put in the design of the cover, I love it.

To my friends in Wageningen, my family away from home: Gracias por todo el apoyo y por darme siempre ánimos. Especialmente Mylu, Claudia, Gabi y Lore "cara de cocou"; fue diver vivir con ustedes. Chino, gracias a tí tambien por todos los buenos ratos.

To my men, Jeroen and Jacob, you are my motivation to be a better me every day. Jeroen, you have been there since the start. You've patiently listen all my pros and cons lists and helped me make every decision.

Y por último, muchas gracias a mi familia. Papitos, es difícil estar lejos pero ustedes me ensenaron a esforzarme y ser valiente, y aquí estoy. No hubiese obtenido estos logros académicos sino fuera por ustedes. Gracias por siempre darme lo mejor y ser el punto de apoyo necesario para seguir este camino. Mane, gracias por siempre cuidar de mi. Siempre estaré ahi para tí como sé que tú estarás para mi. Abuelita Consuelo, gracias por siempre darme motivación para aspirar a más, esto es para tí.

Curriculum vitae

Alejandra Freire Rios was born on the 29 of July of 1981 in Lima, Peru. In 1998 she finished her secondary education at the school "Villa Maria – La Planicie". In 1999 she went to medical school and discovered after a couple of years that her true call was for biological research. In 2002 she started studying Biology with a specialization in Biotechnology at La Molina Agrarian University – Peru, where she obtained her Bachelor degree in 2007. During her last summer as a bachelor, she volunteered at Dr. Carl Nathan's Immunology lab at Weill Cornell, New York; and made her mind to pursue post-graduate studies abroad. After completing her bachelor she did a year - long internship at the Virology Laboratory of the International Potato Center (CIP) - Peru. Then she did another year -long internship at the Bacteriology laboratory of the US Navy Medical Research Institute Detachment (NAMRID) in Lima, Peru. In 2009 she went to The Netherlands to study a master in Plant Biotechnology at Wageningen University. As part of her master studies she did an intership at Keygene in Wageningen. She did the thesis for her master degree at Dr Dolf Weijers' Plant Development group, where she studied ARF-ARF interactions using FRET-FLIM. She obtained her MSc in 2011 and immediately joined Dr Dolf Weijers' group at Wageningen University to pursue a PhD degree. In her project she tried to understand how the ARF family of transcription factors can translate a single signal, auxin, into a great variety of responses during the life of a plant. Multiple approaches were used in order to undestand how these transcription factors recognize and bind DNA which will lead to specific regulation of genes. The results obtained are presented in this thesis. Alejandra wants to pursue a career in academic science and will start a post-doc in the group of Dr Joop Vermeer at Wageningen University.

Publications

Simonini S, Deb J, Moubayidin L, Stephenson P, Valluru M, **Freire-Rios A**, Sorefan K, Weijers D, Friml J, Ostergaard L. (2016). A non-canonical auxin-sensing mechanism is required for organ morphogenesis in Arabidopsis. *Genes Dev.* 30:2286-2296. doi: 10.1101/gad.285361.116.

Freire-Rios A, Radoeva T, De Rybel B, Weijers D, Borst JW. (in press). FRET-FLIM for visualizing and quantifying protein interactions in live plant cells. *Methods Mol Biol.*

Boer DR*, **Freire-Rios A***, van den Berg WA*, Saaki T, Manfield IW, Kepinski S, López-Vidrieo I, Franco-Zorrilla JM, de Vries SC, Solano R, Weijers D, Coll M. (2014). Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell.*156(3):577-89. doi: 10.1016/j.cell.2013.12.027.(* equal contribution)

Freire-Rios A*, Yoshida S*, Weijers D. (2014). Auxin and Its Role in Plant Development. Chapter in book: *Auxin and Its Role in Plant Development.* pp.171-189. doi: 10.1007/978-3-7091-1526-8_9. (* equal contribution)

Rademacher EH, Lokerse AS, Schlereth A, Llavata-Peris CI, Bayer M, Kientz M, **Freire Rios A**, Borst JW, Lukowitz W, Jürgens G, Weijers D. (2012). Different auxin response machineries control distinct cell fates in the early plant embryo. *Dev Cell.* 22(1):211-22. doi: 10.1016/j.devcel.2011.10.026.

187

Education Statement of the Graduate School

Experimental Plant Sciences

The Graduate School
EXPERIMENTAL PLANT SCIENCES

Issued to: **Alejandra Freire Rios**
Date: **22 December 2016**
Group: **Biochemistry**
University: **Wageningen University & Research**

1) Start-up phase	*date*
► **First presentation of your project**	
The structural basis of DNA binding specificity within a transcription factor family.	Apr 19, 2012
► **Writing or rewriting a project proposal**	
Writing a review or book chapter	
Book Chapter: Auxin regulation of embryo development, *IN Auxin and its Role in Plant Development (E. Zazímalová et al. (eds.))* , Spriger_Verlag Wien 2014, p.p. 171. doi 10.1007/978-3-7091-1526-8_9	2013
► **MSc courses**	
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*7.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
Get2Gether PhD student days, Soest, The Netherlands	Jan 29-30, 2015
► **EPS theme symposia**	
EPS theme 1 Symposium 'Developmental Biology of Plants', Wageningen, NL	Jan 19, 2012
EPS theme 1 Symposium 'Developmental Biology of Plants', Leiden, NL	Jan 17, 2013
EPS theme 1 Symposium 'Developmental Biology of Plants', Wageningen, NL	Jan 24, 2014
EPS theme 1 Symposium 'Developmental Biology of Plants', Leiden, NL	Jan 08, 2015
► **Lunteren days and other National Platforms**	
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 02-03, 2012
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 22-23, 2013
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 14-15, 2014
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 13-14, 2015
► **Seminars (series), workshops and symposia**	
Thursday Seminars Biochemisrty: Dr Jürgen Kleine-Vehn (BOKU, Vienna, Austria)	Apr 24, 2012
Thursday Seminars Biochemisrty: Prof. Hong Ma (Fudan University, Shanghai, China)	May 29, 2012
Thursday Seminars Biochemisrty: Prof. Doris Wagner (University of Pensylvania, Philadelphia, USA)	Jul 10, 2012
Thursday Seminars Biochemisrty: Prof. David Robinson (Heidelberg University, Germany)	Sep 27, 2012
EPS Flying Seminar: Dr. Detlef Weigel (Max Planck Institute for Developmental Biology, Germany)	Feb 27, 2013
Thursday Seminars Biochemisrty: Dr M.G. Smits (Hospital Gelderse Valley, Ede, The Netherlands)	Mar 28, 2013
Thursday Seminars Biochemisrty: Dr. Marcus Grebe (Umeá Plant Science Center, Sweden)	May 29, 2013
Thursday Seminars Biochemisrty: Dr. Pierre Hilson (IJPB Versailles, France)	Nov 13, 2013
Thursday Seminars Biochemisrty: Dr. Marcus Heisler (EMBL, Heidelberg, Germany)	Nov 28, 2013
Thursday Seminars Biochemisrty: Dr. Daniël Van Damme (VIB, Ghent, Belgium)	Jan 30, 2014
Thursday Seminars Biochemisrty: Dr. Yvonne Stahl (Heinrich Heine University, Düsseldorf, Germany)	Mar 11, 2014
Thursday Seminars Biochemisrty: Dr. Goerge Bassel (University of Birmingham, United Kingdom)	Apr 24, 2014
Thursday Seminars Biochemisrty:Prof. Cyril Zipfel (Sainsbury Laboratory, United Kingdom)	Jun 04, 2014
Symposium Systems Biology for Food, Feed, and Health	Jun 24, 2014
EPS Symposium 'Omics Advances for Academia and Industry - Towards True Molecular Plant Breeding'	Dec 11, 2014
Thursday Seminars Biochemisrty: Prof. Doris Wagner (University of Pensylvania, Philadelphia, USA)	Dec 11, 2014
► **Seminar plus**	
► **International symposia and congresses**	
Auxin Sailing, Leiden, The Netherlands	Jun 08-09, 2013
Auxcentric, Norwich, UK	May 24-25, 2014
Auxin and Citokynins in Plant Development , Prague, Czech Republic	Jun 29-Jul 04, 2014
► **Presentations**	
Poster at ALW meeting "Experimental Plant Sciences" , Lunteren, NL	Apr 23, 2013
Oral presentation at ALW meeting "Experimental Plant Sciences" , Lunteren 2014	Jun 15, 2014
Poster at ACPD International Symposium , Prague 2014	Jun 30, 2014
► **IAB interview**	
► **Excursions**	
Excursion to Rijk Zwaan	Sep 27, 2013
Excursion to Enza Zaden	Jun 12, 2015
Subtotal Scientific Exposure	*12.5 credits**

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
Transcription Factors and Transcriptional Regulation	Dec 17-19, 2013
EPS PhD Course: Bioinformtics: A User's approach	Aug 25-29, 2014
► **Journal club**	
Participate in weekly journal club (Dolf Weijers group)	2011-2015
► **Individual research training**	
Training in Protein Binding Microarray, Lab of Dr. Roberto Solano - Universidad Autonoma de Madrid,	May 07-18, 2012
Subtotal In-Depth Studies	*8.5 credits**

4) Personal development	*date*
► **Skill training courses**	
Dutch course - NT2-II	2012
WGS course Competence Assessment	Mar 2012
Advanced course guide to Scientific Artwork	May 06-07, 2013
Project and Time Management	Nov 06-Dec 18, 2014
Techniques for Writing and Presenting Scientific Papers	Apr 21-24, 2015
Last Stretch of the PhD	May 22, 2015
► **Organisation of PhD students day, course or conference**	
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*6.1 credits**

TOTAL NUMBER OF CREDIT POINTS*	**34.6**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements
set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

* A credit represents a normative study load of 28 hours of study.