The structure and function of the DNA binding domain of class B MpARF2 share more traits with class A AtARF5 than to that of class B AtARF1

Structure

Crespo, Isidro; Malfois, Marc; Rienstra, Juriaan; Tarrés-Solé, Aleix; van den Berg, Willy et al

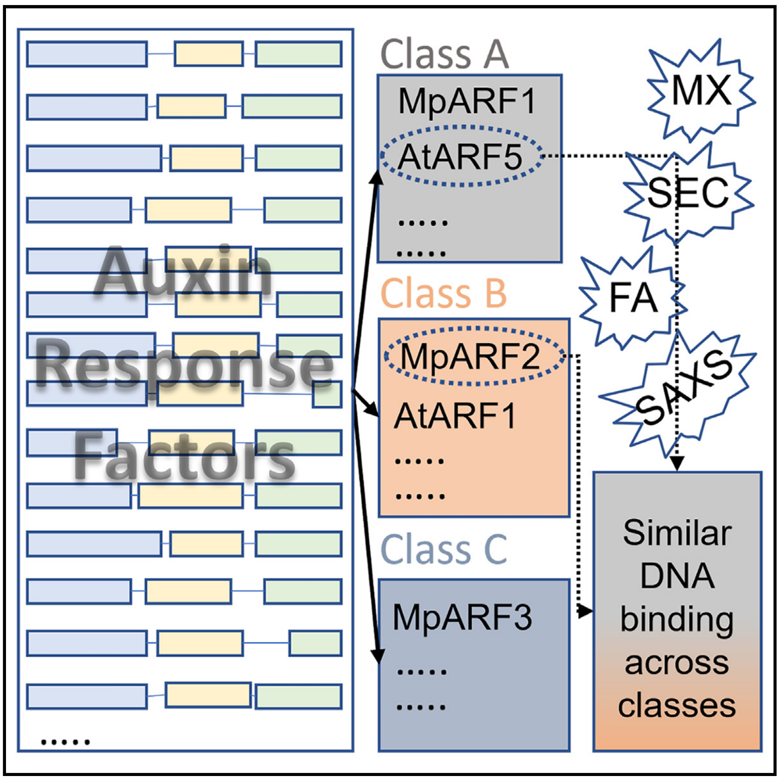
https://doi.org/10.1016/j.str.2025.02.006

Structure

Article

The structure and function of the DNA binding domain of class B MpARF2 share more traits with class A AtARF5 than to that of class B AtARF1

Graphical abstract



Highlights

- Two crystal structures of *M. polymorpha* ARF2 DBD bound to a dsDNA are presented
- Extensive biophysical analyses of the DNA binding to various ARF DBDs are performed
- The data show decoupling of full-length ARF classification from DBD DNA-binding behavior

Authors

Isidro Crespo, Marc Malfois, Juriaan Rienstra, Aleix Tarrés-Solé, Willy van den Berg, Dolf Weijers, Dirk Roeland Boer

Correspondence

rboer@cells.es

In brief

Auxin response factors (ARFs) are proteins that help cells respond to signals. They are grouped into three main types. Crespo et al. studied ARFs from different plants and found that belonging to the same group does not mean they bind to DNA in the same way.

Crespo et al., 2025, Structure *33*, 1–14
May 1, 2025
https://doi.org/10.1016/j.str.2025.02.006

CellPress

Structure

CellPress

Article

The structure and function of the DNA binding domain of class B MpARF2 share more traits with class A AtARF5 than to that of class B AtARF1

Isidro Crespo,[1] Marc Malfois,[1] Juriaan Rienstra,[2] Aleix Tarrés-Solé,[1] Willy van den Berg,[2] Dolf Weijers,[2] and Dirk Roeland Boer[1,3,*]

[1]ALBA Synchrotron Light Source, Carrer de la Llum 2-26, Cerdanyola del Vallès, 08290 Barcelona, Spain

[2]Laboratory of Biochemistry, Wageningen University, Stippeneng 4, 6708 WE Wageningen, the Netherlands

[3]Lead contact

*Correspondence: rboer@cells.es

https://doi.org/10.1016/j.str.2025.02.006

SUMMARY

Plant development is primarily controlled by the auxin phytohormones, which activate the auxin response factors (ARFs). Although the nuclear auxin pathway (NAP) is well studied, little is known on how ARFs specifically select target genes. Here, we investigated the DNA binding mechanism of ARF DNA binding domains (DBDs) from the activator class A and repressor class B in two evolutionary distant plant species, *Marchantia polymorpha* and *Arabidopsis thaliana* using fluorescence anisotropy, size exclusion chromatography, macromolecular crystallography (MX), and small-angle X-ray scattering (SAXS). We find that the previously proposed molecular caliper model, which partially explains the variability in binding of the ARFs to DNA, has been preserved throughout evolution. Our results show that the DBD of class B MpARF2 behaves more like class A AtARF5 than class B AtARF1. These findings suggest that DNA recognition of ARFs has diverged independently of the transcriptional output, which has significant implications for understanding diverse responses to auxin.

INTRODUCTION

Auxins are a prominent class of phytohormones that were discovered during the beginning of the 20th century, which regulate all critical aspects of plant development from early embryogenesis to fruit ripening, organogenesis, cell division, cell expansion/differentiation and root initiation.[1–3] The principal natural auxin in land plants is indole-3-acetic acid (IAA).[4] Cells that sense auxin can respond quickly through non-transcriptional cellular responses,[4] but many of the developmental responses triggered by auxins are mediated through the canonical nuclear auxin pathway (NAP).[5] The NAP consists of three dedicated components, the auxin co-receptors transport inhibitor response 1/auxin-signaling F-box (TIR1/AFB), the auxin/indole-3-acetic acid (Aux/IAA) transcriptional repressors, and the auxin response factor (ARF) transcription factors.[6] The genome of the flowering plant *Arabidopsis thaliana* encodes 6, 29, and 23 genes coding for TIR1/AFB, Aux/IAA, and ARF family members, respectively.[7,8] Differential expression patterns of each of these genes in various tissues increase the complexity of responses to auxins.[7,9,10]

ARF transcription factors regulate gene expression in response to auxins activated by the auxin-dependent degradation of Aux/IAA inhibitor proteins.[11] ARFs are phylogenetically grouped into class A, B, and C,[12–14] and these classes have distinct functions.[5,15–17] Class A ARFs are generally considered transcriptional activators while class B generally repress transcription. *Marchantia polymorpha* encodes a single ARF ortholog for each class, whereas the 23 *Arabidopsis thaliana* orthologs are distributed over the three classes. Based on genetic analysis of the *M. polymorpha* ARF orthologs, it was recently proposed that class C ARFs may not be involved in auxin signaling[10] and their function remains unclear. Despite the divergence of ARFs between different species, their overall sequence has been very much conserved. For example, the percentage of identity between all 23 *Arabidopsis thaliana* and the 3 *Marchantia polymorpha* ARFs ranges from 28.1% to 94.6%. ARFs have a conserved N-terminal DNA binding domain (DBD),[16] important for specific recognition of promoter sequences.[16,18,19] The DBD is followed by a middle region (MR), which is intrinsically disordered and is enriched in specific amino acids.[5,16,17] Most ARFs contain a Phox/Bem1p (PB1) protein-protein interaction domain (domain III/IV) located C-terminal to the MR.[20,21] This domain mediates interaction between ARF transcription factors and Aux/IAA proteins,[5,20,22] and oligomerization of ARFs.[22–24]

The first crystal structures of ARF-DBDs revealed the presence of three subdomains.[18] The N-terminal ARF-DBD consists of a subdomain that allows for DBD dimerization (DD) and is critical for ARF functions *in vivo*.[18] A plant-specific B3 subdomain, which is responsible for DNA binding, is present as an insertion into the DD. The C-terminal residues of the ARF-DBD form the ancillary domain, whose function is unclear.[12,13,18,25] The B3 domain is also found in the ABI3 (abscisic acid insensitive 3, B3-ABI3) and RAV (related to ABI/VP1, B3-RAV) plant transcription factors, but with different DNA binding specificities; whereas B3-ABI3 preferentially binds to 5′-CATGCA-3′ and B3-RAV to 5′-CACCTG-3′, B3-ARFs bind the auxin response element (AuxRE) motif, 5′-TGTCNN-3′,[26–28] in which the identity of the last 2 bases determine affinity of binding.[19] Interestingly, the amino acids involved in B3 DNA binding are mostly conserved among the 23 ARF orthologs found in *A. thaliana*. In fact, this conservation exists even in homologues that diverged 400 million years ago, as is clear when comparing ARFs from *A. thaliana* and those from *M. polymorpha*.[10] As a result, the B3 domain's affinity for target DNA sequences is expected to be invariant. Given the high conservation of B3 DNA-binding residues, the specificity of different ARF orthologs within a plant species cannot be explained by surface changes in the DNA binding interface. It was proposed that structural constraints on the position and spacing of the B3 domains are imposed by intramolecular interactions between non-conserved residues of the loops connecting the DD and B3 subdomains on the one hand, and the lengths and environment-enforced conformational restrictions of these loops on the other. In this way, affinity can be modulated to bind uniquely spaced palindromic motifs by ARF-DBD dimers.[18,20,29] This model is termed the "molecular caliper" hypothesis,[18] and other plant transcription factors, such as the LATERAL ORGAN BOUNDARIES (LOB) domain family of zinc-finger transcription factors, have been proposed to follow a similar mechanism of gene selection.[30] Although *in vivo* results with AtARF5 support this hypothetical model,[31,32] there are unresolved questions. For example, inverted repeat arrangements are not as common as single AuxREs within the promoters of auxin-response genes, and the molecular caliper model does not explain ARF binding to all auxin-responsive promoters.[8] In addition, it is not clear how DBDs from different ARF classes determine the type of response to auxin (i.e., whether the ARF's functional role as activator or repressor is reflected in the function of the DBD alone). Domain swapping of *M. polymorpha* class A and B ARF-DBDs shows that these are largely interchangeable, suggesting that in *M. polymorpha*, the class A and B ARF-DBDs are functionally equivalent, in contrast to the MR and PB1 domains.[10] Duplication of ARF genes in higher land plants has generated many ARF orthologues for both classes, and it is possible that posterior DBD diversification contributed to functional differences between these ARF orthologues.

In this work we address the DNA binding and structural properties of representative members of all three classes of ARF-DBDs from two species, *A. thaliana* and *M. polymorpha*, and their implication on the AuxRE spacing selection. The tested DNA sequences include those found within promoter regions of genes known to be regulated by ARFs, as well as synthetic oligonucleotides that were found to have high affinity for ARFs. We define the oligomeric distributions of ARFs in the absence and presence of DNA and determine the contribution of DBD structural elements to molecular caliper-based DNA recognition using size exclusion chromatography coupled to multi-angle light scattering (SEC and SEC-MALS), small-angle X-ray scattering (SAXS), and macromolecular crystallography (MX). We observed that the DNA-binding preference and behavior of the class B MpARF2-DBD are more consistent with the class A AtARF5 than with the class B AtARF1. Our results show that the length of helix 1 varies with the ARF classes and suggests that this variability is an evolutive strategy to increase gene selection specificity. The results found here can be extrapolated to other plant organisms, as we observed similar tendency for organisms not directly related. Overall, our results unveil an additional layer of ARF gene selection regulation, where target selection and transcriptional output have evolved separately and are not strictly coupled, thereby contributing to the knowledge on the ARF function.

RESULTS

ARF-DBD:DNA binding affinity depends on both sequence and spacing

We investigated the DNA binding preferences of the DBDs of class A (AtARF5), B (AtARF1 and MpARF2), and C (MpARF3) ARFs by analytical SEC and fluorescence anisotropy (FA). For the analyses, we used a set of DNA oligos containing different recognition sequences separated by different spacings. The recognition sequences selected were derived from promoters of *Arabidopsis* genes that are regulated by AtARF5 (*TMO3*, *TMO5*, and *LFY*), and a direct repeat of the canonical 5′-TGTCTC-3′ (AuxRE) motif separated by 5 base pairs (DR5). Furthermore, we analyzed the binding affinity of oligonucleotides containing inverted repeats of the high affinity 5′-TGTCGG-3′ motif with variable spacings (from 5 to 9 base pairs, named HA5 to HA9, respectively). Finally, we also included oligonucleotides containing inverted repeats of the AuxRE motif with spacings of 7 and 8 base pairs (IR7 and IR8) (Figure 1A). The SEC profiles of class A or B ARF-DBDs mixed with double-stranded oligonucleotides IR7, IR8, HA7, and HA8 showed shifts in elution volumes (Data S1; Table S1) compatible with stable ARF-DBD homodimers in complex with dsDNA (Figure 1B). When the spacing differed from the optimal 7–8 base pairs, interactions with some oligonucleotides were still observed for MpARF2-DBD, as shown by peak shifts in SEC chromatograms. However, the elution volume of these species did not correspond to a homodimer:dsDNA complex. In these cases, the lower peak mobility could be explained either by dynamic binding/unbinding of an ARF dimer or by binding of a monomer to the DNA. It should be noted that we only considered a positive interaction if a shift in the peak mobility was of at least the one corresponding to the migration of a monomer bound to a dsDNA. Partial shifts not reaching this threshold were marked as non-binding, but it should be noted that (weak) interactions still occur *in vitro*.

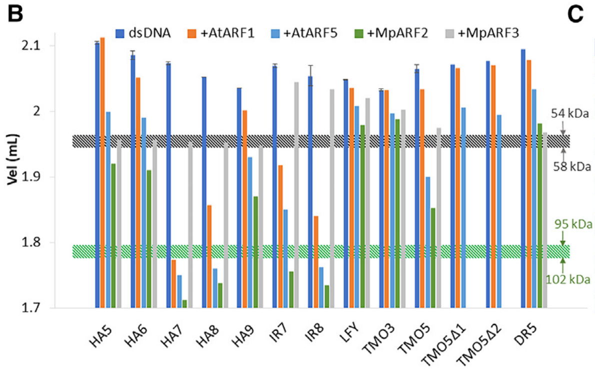
Whereas AtARF1-DBD did not show a significant shift in elution volume in the presence of HA9 (9 base pairs spacing), both AtARF5-DBD and MpARF2-DBD showed partial shifts. No significant shifts were observed for either AtARF1-DBD or AtARF5-DBD with DNAs containing short spacings (5–6 base pairs), suggesting that they require a spacing of at least 7 nucleotides to bind. Partial shifts were observed for these oligonucleotides with MpARF2-DBD. MpARF3-DBD showed partial shifts when binding to any of the spacings in the HA5 to HA9 series. Under the test conditions, none of the ARF-DBDs showed

Please cite this article in press as: Crespo et al., The structure and function of the DNA binding domain of class B MpARF2 share more traits with class A AtARF5 than to that of class B AtARF1, Structure (2025), https://doi.org/10.1016/j.str.2025.02.006

A

```
          AuxRE1-2-4-6-8-AuxRE2
    HA5:  TGTCGGCATTG----CCGACA
    HA6:  TGTCGGCGATCG---CCGACA
    HA7:  TGTCGGCGATTCG--CCGACA
    HA8:  TGTCGGCGATATCG-CCGACA
    HA9:  TGTCGGCGATTATCGCCGACA
    IR7:  TGTCTCCCTTTGG--GAGACA
    IR8:  TGTCTCCCAAAAGG-GAGACA
    LFY:  TGTCAATTTCCCAGCAAGACA
   TMO3:  GGTCAAAAGTAAGACTGGACC
   TMO5:  GGTCTCTGGTCGG--TCGACA
 TMO5Δ1:  GGTCTCTGGTCGG--TTTTTT
 TMO5Δ2:  TTTTTTTGGTCGG--TTTTTT
    DR5:  TGTCTCCCTTT----TGTCTC
```

B



C

ID	AtARF1	AtARF5	MpARF2	MpARF3
HA5	X	X	✔	✔
HA6	X	X	✔	✔
HA7	✔	✔	✔	✔
HA8	✔	✔	✔	✔
HA9	X	✔	✔	✔
IR7	✔	✔	✔	X
IR8	✔	✔	✔	X
LFY	X	X	X	X
TMO3	X	X	X	X
TMO5	X	✔	✔	X
TMO5Δ1	X	X	-	-
TMO5Δ2	X	X	-	-
DR5	X	X	X	X

Figure 1. Size exclusion peak shift analysis of AtARF1-, AtARF5-, and MpARF2-DBDs in complex with DNAs containing various auxin response elements (AuxRE)

(A) DNA sequences designed for *in vitro* testing. The AuxRE-derived sequences are colored in red, while spacers are shown in black.

(B) Elution volume (V_{el}) corresponding to each dsDNA alone (dark blue) or in complex with AtARF1 (green), AtARF5 (light blue), MpARF2 (green) or MpARF3 (gray) DBDs at a constant 2:1 protein:dsDNA molar stoichiometry. Horizontal gray- and green-shaded zones correspond to the theoretical elution volume calculated for a dsDNA in complex with an ARF-DBD monomer or dimer, respectively. For dsDNA alone, the standard deviation in V_{el} of the multiple runs is given.

(C) Qualitative analysis of the SEC results. Positive interactions were assigned when elution volume was lower than that corresponding to a monomer:dsDNA. Error bars represent the standard deviation of mean. See also Data S1 and S3 and Table S1.

complete shifts when binding to DR5, the sequence containing the high affinity 5′-TGTCGG-3′ motif in direct repeat configuration (Figures 1B and 1C). Surprisingly, oligonucleotides derived from *LFY* or *TMO3* promoter sequences, which are known *in vivo* target sequences of AtARF5-DBD[31,32] with 9 base pairs spacing between the putative inverted ARF binding sites, did not show elution shifts when mixed with any of the tested ARF-DBDs (AtARF1, AtARF5, MpARF2, and MpARF3) under the assayed experimental conditions. This is in contrast to surface plasmon resonance data recorded previously,[18] and suggests that ARF proteins behave differently in SEC compared to when immobilized in surface plasmon resonance, probably due to the dilution of the complex experimented during the SEC run, which indicates that the interaction is weak. The class B MpARF2 and class A AtARF5, but not the other ARFs, showed clear elution volume shifts in the presence of oligonucleotides derived from *TMO5* (Figures 1B and 1C). This hybrid oligonucleotide contains both the canonical and non-canonical sequences with only a nucleotide substitution at each 5′ end separated by 7 base pairs. The observed binding of MpARF2 and AtARF5 to *TMO5* indicates an apparent tolerance to these differences with respect to the canonical sequences.

To further test the molecular caliper hypothesis, we performed steady state titration FA assays on AtARF1- and AtARF5-DBDs to quantitatively determine the affinity of ARF-DBDs toward AuxRE sequence and spacing. For these measurements, the FAs of a fluorescein (6-FAM)-labeled dsDNA molecules were measured at different protein concentrations. The FA is a measure of a change in the rotational time of the labeled molecules, which in turn depends on its size and shape. Both AtARF1-DBD and AtARF5-DBD had higher affinity to TGTCGG-based (HA) than to TGTCTC-based (IR) sequences (Table 1; Data S2). In addition, the optimal spacing was found to be 7–8 base pairs, in accordance with the SEC results. The FA data provided information on the cooperativity of binding through calculation of the Hill coefficient. A negative cooperativity indicates a first phase of interaction with one AuxRE, followed by a second phase in which structural rearrangement is required for the interaction of a second ARF monomer with the second AuxRE. A positive cooperativity indicates that binding of the first monomer aids in recruitment of the second monomer to form the complex. A neutral cooperativity indicates either that the dimer had formed before binding and that the complex forms through a single step reaction without the need for structural rearrangements, or that the protein binds as a monomer and that no further interactions between protein molecules occur in binding to the AuxREs. It should be noted that these measurements can only detect cooperativity of dimerization when a second monomer binds to a preformed monomer-DNA complex, but cannot detect the effect of protein dimerization before binding to DNA, due to the fact that the FA of the DNA is measured. Thus, single interaction events of either a protein monomer or preformed dimer binding to the DNA will show neutral cooperativity.

AtARF1-DBD and AtARF5-DBD behaved very differently with the tested oligonucleotides. For oligonucleotides containing the high affinity sequence with spacing of 7 or 8 base pairs and the TMO5 sequence, negative cooperativity was found for AtARF1-DBD, suggesting that binding of a second monomer to a monomer-DNA complex occurs and requires a structural rearrangement. In contrast, AtARF1 displayed no cooperativity for spacings larger than 8 or smaller than 7 base pairs. This implies that these sequences bind AtARF1 in a single step, either as a monomer or a preformed dimer. We also observed neutral cooperativity of AtARF1-DBD for the low affinity IR7/8 sequences, which could be explained by the lower affinity of the B3 domain to the auxREs: formation of the protein-DNA complexes occurs at higher concentrations, which favors protein dimers that bind the DNAs in a single binding event. For AtARF5-DBD, a neutral cooperativity was found for all the assayed DNA sequences, indicating that the protein either binds to the DNA as a monomer, or that the dimer formed before binding to the DNA.

FA showed that AtARF1 and AtARF5 have affinity toward spacings of 5, 6, and 9 nucleotides, in contrast with SEC results.

Table 1. ARF-DBD:DNA interaction parameters obtained from fluorescent anisotropy

	AtARF1-DBD			AtARF5-DBD		
DNA	K_D (μM)	Hill coeff[a]	Cooperativity[b]	K_D (μM)	Hill coeff[a]	Cooperativity[b]
HA5	4.18 ± 1.36	0.84	Neutral	5.74 ± 1.63	MM	Neutral
HA6	3.24 ± 0.78	0.84		5.99 ± 1.52	MM	
HA7	2.87 ± 0.63	0.37	Negative	2.78 ± 0.64	–	
HA8	0.81 ± 0.20	0.47		2.33 ± 0.44	MM	
HA9	3.18 ± 1.00	MM	Neutral	6.31 ± 0.59	MM	
IR7	28.06 ± 7.49	MM		3.73 ± 0.96	MM	
IR8	14.88 ± 3.20	–		1.08 ± 0.26	–	
DR5	9.88 ± 2.67	MM		21.68 ± 2.98	MM	
TMO5	0.71 ± 0.10	0.58	Negative	4.97 ± 0.69	MM	
LFY	0.12 ± 0.10	–	NI*	10.3 ± 2.67	MM	

[a]MM in case the binding follows a Michaelis-Menten fit, a dash indicates that the Hill model did not converge. K_D errors represent standard errors.
[b]The fit that resulted in the lowest variation coefficient for the K_D cooperativity was used to decide whether the binding followed a Michaelis Menten or Hill mechanism. See also Data S2.

As mentioned previously, we only considered interactions in SEC if a significant shift of the peak corresponding to a monomer:dsDNA or dimer:dsDNA were observed. It is likely that the lower affinity for these sequences prevents detection of binding using SEC based on the elution shift.

DNA induces oligomerization of class A and B ARFs in solution

To further analyze the interaction behavior of ARFs with DNAs with inverted sequence separated by a variety of spacings, we prepared ARF-DBD:dsDNA mixtures at variable stoichiometry and analyzed them by SAXS. In addition, we addressed the possible effect of protein concentration on dimerization by testing a range of concentrations from 0.5 to 7 mg/mL of different apo ARF-DBDs (Tables 2, S2, and S3).

For the apo class A and class B ARF-DBDs, the particle size increased with protein concentration as evidenced by an increase of R_g, V_{porod}, and D_{max} (Table 2), which indicates concentration-dependent interactions between ARF-DBDs, consistent with oligomerization. Analysis of apo ARF-DBD SAXS profiles with OLIGOMER[33] suggested that apo class A/B ARFs were able to homodimerize in solution under the conditions tested, but neither purely monomeric nor dimeric models explain the experimental data, with clear inconsistencies in the 0.01–0.15 nm region (Figure S1, green and black lines). The calculated fit using the theoretical curves of both dimer and monomer simultaneously improved the fit dramatically (Figure S1, red line) for concentrations below 3.5 mg/mL, showing that the tested A/B ARFs distribute between dimer and monomer in solution at low concentrations (Figure 2). It should be noted that the dimerization fraction of apo AtARF5-DBD and apo MpARF2-DBD was higher at 0.7 mg/mL compared to apo AtARF1. Thus, the homodimerization dissociation constant (K_D) appears to be higher for apo AtARF1-DBD compared with apo AtARF5-DBD/MpARF2-DBD (Figure 2A), which confirms a previous report.[34] It should be noted that χ^2 values of the fits for the apo MpARF2 and AtARF5 increased significantly at concentrations higher than 1.75 mg/mL (see Table S2), but not for apo AtARF1 and for apo MpARF3-DBD, which is the only ARF-DBD unable to dimerize. This result suggests that other oligomerization states occur at high concentrations.

An *in silico* model of the apo MpARF3-DBD monomer was generated with SwissModel.[35] A theoretical curve was calculated for this model and used to generate a Crysol fit with the experimental SAXS curve (Figures S2B and S2C). The fits for all apo MpARF3-DBD concentrations gave low χ^2 values (see Table S2). This result suggests that the MpARF3-DBD has a similar structure to other ARF-DBD structures. Furthermore, it confirmed that apo MpARF3-DBD is unable to dimerize, which is in line with the prediction that the dimerization interface of MpARF3-DBD is disrupted by an 80 amino acid insertion in α6-helix.[36] The apo MpARF3-DBD shape based on computation from a chain-like *ab initio* model from SAXS data using GASBOR[37] was similar to the *in silico* model (Figure S2A; Table S2). The 80 amino acid insertion in the α6-helix is missing in the apo MpARF3-DBD GASBOR model, which is likely due to the fact that this loop is intrinsically unstructured and therefore does not show contrast in SAXS. For the other apo ARFs, a GASBOR model could not be calculated as the solution contained a mixture of monomers and dimers.

Analysis of SAXS particle size parameters of the protein:dsDNA complexes (Table 2) showed an increase in particle volume and size upon addition of DNA for combinations that were previously determined (by SEC and FA) to have interaction affinity. For protein/DNA combinations that are known to interact weakly, for example for AtARF1-DBD:HA6, or when the protein was unable to form dimers (MpARF3-DBD), a transient increase of the particle volumes (V_{porod}) was observed when increasing the protein:dsDNA stoichiometry from 1:1 to 2:1. Further increases to the protein:dsDNA stoichiometry decreased the V_{porod}, while the maximum particle sizes (R_g and D_{max}) remained nearly constant. A constant R_g and D_{max} combined with a concomitant reduction of the V_{porod} suggest an elongation of the overall shape of the complex, consistent with dsDNA in complex with a protein monomer, which has an elongated rod-like shape. Thus, we interpret that in these cases, the protein binds as a monomer to dsDNA and that the excess protein will remain free on the solution.

Please cite this article in press as: Crespo et al., The structure and function of the DNA binding domain of class B MpARF2 share more traits with class A AtARF5 than to that of class B AtARF1, Structure (2025), https://doi.org/10.1016/j.str.2025.02.006

Table 2. SAXS parameters for Apo- and DNA-bound ARF DBD complexes

			AtARF1			AtARF5		
DNA	[ARF] (mg/mL)	ARF:dsDNA	R_g (nm)	V_{porod} (nm^3)	D_{max} (nm)	R_g (nm)	V_{porod} (nm^3)	D_{max} (nm)
HA6	3.5	4:1	3.87 ± 0.35	153.55	11.23	4.14 ± 1.54	200.16	12.32
	3.5	2:1	3.76 ± 0.99	178.15	11.18	4.15 ± 0.34	209.02	12.56
	3.5	1:1	3.79 ± 0.09	161.59	11.17	5.83 ± 0.91	242.47	11.65
HA7	3.5	4:1	3.61 ± 0.3	130.86	11.03	4.02 ± 0.39	201.11	11.95
	3.5	2:1	3.56 ± 0.21	135.56	10.49	3.85 ± 0.19	188.83	11.64
	3.5	1:1	3.09 ± 0.33	131.41	10.75	3.87 ± 0.41	178.03	12
HA8	3.5	4:1	3.6 ± 0.28	126.25	10.87	3.99 ± 0.13	198.88	12.07
	3.5	2:1	4.42 ± 0.55	142.99	9.87	3.99 ± 0.14	198.88	12.07
	3.5	1:1	3.45 ± 0.14	132.52	10.34	3.85 ± 0.17	166.65	11.55
HA9	3.5	4:1	3.78 ± 0.23	134.59	11.55	5.17 ± 0.31	277.75	14.75
	3.5	2:1	3.72 ± 0.17	144.5	11.25	4.24 ± 0.49	228.69	12.83
	3.5	1:1	3.57 ± 0.77	129.03	10.34	3.83 ± 0.25	173.36	12.07
Apo	7	–	3.48 ± 0.09	111.84	10.89	3.89 ± 0.39	153.94	11.96
	3.5	–	3.17 ± 0.19	105.73	8.8	3.35 ± 0.07	106.23	12.32
	1.75	–	2.92 ± 0.02	83.58	7.6	3.34 ± 0.34	84.64	8.52
	0.7	–	2.63 ± 0.08	85.24	7.36	3.06 ± 0.12	80.48	10.54
			MpARF2			MpARF3		
DNA	[ARF] (mg/mL)	ARF:dsDNA	R_g (nm)	V_{porod} (nm^3)	D_{max} (nm)	R_g (nm)	V_{porod} (nm^3)	D_{max} (nm)
HA6	3.5	4:1	3.89 ± 0.6	163.12	11.09	3.24 ± 0.1	63.65	8.32
	3.5	2:1	3.88 ± 0.18	159.74	11.24	3.06 ± 0.03	67.45	8.42
	3.5	1:1	3.81 ± 0.32	140.32	11.16	3.02 ± 0.2	65.82	8.62
HA7	3.5	4:1	3.75 ± 0.17	151.92	10.87	2.99 ± 0.03	62.95	8.05
	3.5	2:1	3.49 ± 0.32	139.9	9.84	3.07 ± 0.15	68.04	8.52
	3.5	1:1	3.52 ± 0.32	118.44	9.59	3.1 ± 0.15	62.93	8.53
HA8	3.5	4:1	3.85 ± 0.42	155.16	10.67	2.92 ± 0.03	61.95	8.03
	3.5	2:1	3.64 ± 0.35	146.35	9.93	3.14 ± 0.09	65.49	8.82
	3.5	1:1	3.57 ± 0.41	122.77	9.82	3.09 ± 0.2	60.11	8.71
HA9	3.5	4:1	3.94 ± 0.49	185.54	11.43	3.03 ± 0.04	62	8.41
	3.5	2:1	3.76 ± 0.42	163.16	11.01	3.06 ± 0.17	64.8	8.87
	3.5	1:1	3.63 ± 0.41	123.39	10.44	3.23 ± 0.32	60.3	8.95
Apo	7	–	6.52 ± 0.1	181.18	11.01	2.77 ± 0.1±	59.87	7.34
	3.5	–	3.72 ± 0.78	107.92	10.22	2.6 ± 0.03	59.54	6.16
	1.75	–	3.31 ± 0.08	76.71	7.87	2.49 ± 0.09	53.5	6.48
	0.7	–	2.92 ± 0.1	64.74	10.19	2.06 ± 0.05	69.07	6.15

The parameters listed are the radius of gyration (R_g), the Porod volume (V_{porod}), and the maximum particle distance (D_{max}). The standard deviation of the Rg value are included.

In addition to the SAXS parameter analysis, we assessed the contribution of DNA to dimer formation and monomer/dimer distributions of ARF-DBD:dsDNA complexes with the program OLIGOMER.[33] For this analysis, we used two approaches. In the first, we included five molecular species: monomers or dimers of the apo proteins, free DNA, and complexes containing a single monomer or dimer of ARF-DBD bound to DNA. In the second approach, we used only monomer to DNA and dimer to DNA species. Both approaches resulted in similar χ^2 values. In contrast to the apo proteins, fits to the model resulted in high χ^2 values in the case of AtARF5-DBD and MpARF2-DBD (see Table S2). This may be a consequence of the presence of several oligomeric species in the solution, which was not considered in the calculated fits using the crystallographic structural models. Such possible species would include aggregates due to protein dimers binding across different DNA molecules. Despite these difficulties, a χ^2 value of 2.59 was obtained for a mixture of AtARF1-DBD and HA7 and a χ^2 value of 3.26 was obtained for a mixture of AtARF1-DBD and HA8 at stoichiometries of 1:1, showing a displacement of the equilibrium toward dimer formation compared to the apo protein (Table S2). In the case of MpARF3-DBD, χ^2 values in the range of 4.97–14.44 were obtained for a monomer bound to dsDNA for the different DNA concentrations (Figure 2B; Table S2). The SAXS data therefore confirms the dimerization behavior of the studied ARFs in the absence and presence of oligonucleotides, consistent with

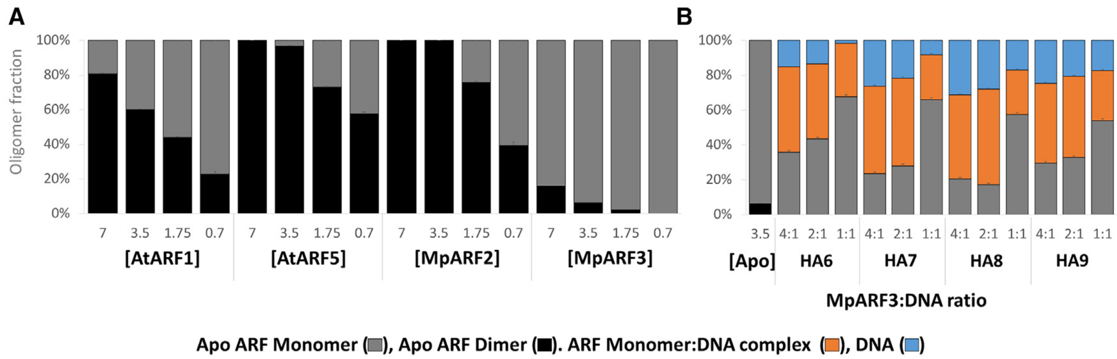


Figure 2. Distribution of oligomeric species in MpARF3-DBD solutions calculated from SAXS data

(A) The distribution of dimers and monomers at 0.7, 1.75, 3.5, and 7 mg/mL as calculated by OLIGOMER[33] for apo MpARF3-DBD. Monomers are represented in gray, and dimers in black.

(B) Mixture of MpARF3-DBD (3.5 mg/mL) with different oligonucleotides (HA recognition sequence with spacings of 6–9 base pairs between AuxRE elements) at different protein:dsDNA ratios generated monomer:dsDNA complexes (orange), but no dimer:dsDNA species were detected. The fraction of dsDNA not bound to protein is represented in blue. For comparison, the first column shows the distribution of monomers and dimers of apo MpARF3-DBD at 3.5 mg/mL as shown in (A), which was the concentration used in the MpARF3:DNA assays. See also Table S2.

observations from SEC and FA (as discussed earlier). Notably, DNA binding appears to be a significant factor driving dimerization in ARF classes A and B. The SAXS data for AtARF1-DBD:HA7, MpARF3-DBD:HA6, and Apo MpARF3-DBD were deposited in the SASBDB with IDs SASBDB: SASDVV5, SASBDB: SASDVU5, and SASBDB: SASDVT5, respectively (Table S3).

In order to confirm the validity of the SEC, SAXS, and FA experiments, we performed SEC-MALS experiments on AtARF1 in complex with DNAs HA5 and HA7, respectively, at different concentrations and stoichiometries (Data S3). For both complexes, we tested stoichiometries 2:1 (AtARF1-DBD:dsDNA) at 40 and 80 μM [AtARF1-DBD] respectively, and a stoichiometry of 10:1 at 200 μM [AtARF1-DBD]. For HA5 at 2:1 stoichiometry at both tested concentrations, we did not detect any dimer bound to DNA; 90% of the sample consisted of apo AtARF1 or monomer AtARF1 bound to dsDNA, resulting in a broad or double peak in the 40–50 kDa elution volume range. The remaining peaks accounting for 12% of the sample corresponded to non-aggregate high Mw multimers of protein:DNA complexes. For AtARF1:HA5 at stoichiometry 10:1, the results are basically the same as those for the 2:1 stoichiometries: 95% of the sample corresponds to monomeric protein bound to DNA and apo protein. Thus, the protein clearly prefers to bind to DNA as a monomer, rather than dimerize in the apo form, as suggested by SAXS with spacings not suitable for ARF-DBD dimerization.

For HA7, a completely different picture emerges. In this case, we find a mixture of AtARF1 monomers and dimers bound to dsDNA at stoichiometries of 2:1. About 50% of the protein is found as the dimer:dsDNA complex. About 35% of the protein is found as apo monomer, or bound to dsDNA, as evidenced by a broad peak eluting between 15.5 and 16.5 mL. Interestingly, we find that about 10% of the DNA is found in free form. This distribution is not dependent on the concentration of the protein and DNA and is perfectly consistent with a cooperative binding model: the dimer:dsDNA is more stable than the monomer: dsDNA complex, leading to a mixture of free protein monomer, free DNA, a small amount of monomer:dsDNA complex, and mostly dimer:dsDNA. It should be noted that in the case of HA5, no free DNA is found in solution. This is due to the lack of dimer:HA5 complex, which leads to a higher effective protein to DNA stoichiometry, and therefore to the increased formation of monomer:HA5 complexes, depleting the DNA. Finally, the cooperativity model for HA7 is confirmed by the complexes found at a stoichiometry of 10:1, where we find that 9% of the sample consists of a protein dimer bound to dsDNA, no free DNA, and 85% of free protein monomer. The remaining 6% is found as higher order structures and aggregates.

Ancestral relationships of ARF sequences and structures

Overall, the results from the SAXS data correlate well with the oligomerization states observed in SEC-MALS and the binding kinetics determined by FA. Interestingly, we consistently found that the class B MpARF2 shared more traits (e.g., DNA binding pattern and homodimerization capability) with the class A AtARF5 than with the class B AtARF1. This suggests that functional clustering does not predict DBD properties in all the cases. To further explore the similarities and dissimilarities between ARFs from the different classes, we determined the structures of *Marchantia polymorpha* ARF2-DBD in complex with HA7 (MpARF2-HA7, 2.32 Å, PDB: 8OJ1) and in complex with IR7 (MpARF2-IR7, 2.56 Å, PDB: 8OJ2) both belonging to space group $I2_12_12_1$ (Tables S4 and S5). We compared these structures with structures of ARF-DBDs from *Arabidopsis thaliana* (AtARF1-HA7, PDB: 6YCQ; AtARF1-IR7, PDB: 4LDX; ApoAtARF1, PDB: 4LDV; ApoAtARF5, PDB: 4LDU) and *Marchantia polymorpha* (PDB: 6SDG) previously described.[10,18,19] The structures showed MpARF2-DBD bound to DNA as a dimer, with dimerization mediated by the α6 helix, similar to that observed in previous ARF-DBD structures (Figure 3A). The DBDs in both *A. thaliana* and *M. polymorpha* structures are composed of a DD (Figure 3A, yellow), in which a B3 DNA interaction domain is embedded (Figure 3A, green), followed by an ancillary domain at the C terminus (Figure 3A, orange).[11,13,17,38] The "taco fold" previously observed in the DD of AtARF1 and AtARF5 structures is

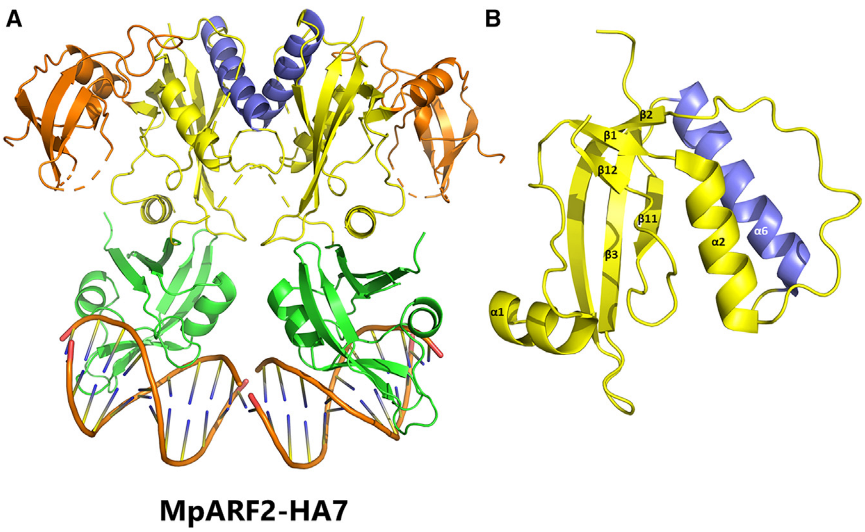


Figure 3. Structural characteristics of MpARF2-DBD in complex with dsDNA HA7

(A) Overall view of an MpARF2-DBD dimer in complex with palindromic high affinity dsDNA HA7. The α6 helix located in the dimerization domain (DD; colored in yellow) is colored in blue, the B3 domain in green, and the ancillary domain in orange.

(B) Cartoon representation with secondary element notation on the "taco" fold in the DD.

conserved in MpARF2-DBD structures (Figure 3B). Finally, the structure of the ancillary domain is similar to what was described for the *Arabidopsis thaliana* and *Marchantia polymorpha* ARFs,[18] showing that the global DBD architecture is conserved since *Marchantiophyta* and *Brassicaceae* diverged over 400 million years ago. The structural similarity of *M. polymorpha* and *A. thaliana* ARF-DBDs occurs in a context where the overall DBD sequence identity remains high, which is 45% on average among the class A and B ARFs.

We observed several similarities between the structures of the class B MpARF2 and class A AtARF5-DBDs that were not found in the class B AtARF1-DBD structures, most importantly the first α-helix (helix 1), which is the contact point between the B3 and the DD (Figure 4). The α1 helices of MpARF2 and AtARF5 are 10 residues long (IDAELWYACA and NSELWHACAG, respectively), whereas the equivalent helix in the AtARF1-HA7 structure is 19 residues in length (PGGVLSDALCRELWHACAG) (Figures 4B–4G). Note that the N-terminal residues preceding the α1 helices of MpARF2 and AtARF5 were present in the protein constructs, but a lack of electron density indicates that they are not ordered. The longer α1 helix of AtARF1-DBD establishes contacts with β-strands 2 and 3 of the DD, and with β-strands 4, 9, and 10 of the B3 subdomain, which provide more anchor points between the DD and the B3 subdomain of AtARF1. In the case of MpARF2 and AtARF5, the shorter helix 1 establishes fewer contacts between these domains. In fact, it is not clear if the formation of the α1 helical conformation in this region is a cause or consequence of the longer β-strands 4, 9, and 10 of the B3 subdomain in AtARF1. In the case of the MpARF2 and AtARF5 B3 domains, the space equivalent to that occupied by the N-terminal residues of helix 1 in AtARF1 is occupied by a loop connecting β-strands 9 and 10 (Figures 4B–4G). The reduced number of contacts between the B3 and DD domains in MpARF2 and AtARF5 due to the shortening of helix 1 allows for more flexibility in the DD:B3 interface, and as a result the B3 domains would turn outwards, yielding structures that are not as planar as in AtARF1. Future binding analysis to the HA5-9 motifs by AtARF1 and AtARF5 with swapped helixes would be needed to show that indeed this has an impact on the specificity for IR spacing.

In order to determine whether the observed differences in helix 1 occur in ARFs from other species, we analyzed the secondary structural elements present on the DBDs of all ARFs from *A. thaliana*, *M. polymorpha*, *A. trichopoda*, *O. sativa*, *P. patens*, *S. lycopersicum*, and *Z. mays*, retrieving the AlphaFold2[39] models for all the ARF sequences available at Phytozome database[40] and computing the secondary structural elements with the Dictionary of Secondary Structure in Proteins (DSSP) implementation from the Phenix's macromolecular toolbox. The obtained structural elements were plotted according to the alignment of the amino acid sequences of the ARFs and clustered according to ARF classes, confirming that the length of helix 1 is invariable for class A and C ARFs, but variable in class B ARFs (Figures 4A and S3). These structural features may explain the difference in spacing preference of AtARF1-DBD and AtARF5-DBD in the SEC and FA analysis presented previously, and in the original formulation of the molecular caliper model.[18]

This analysis also suggests that other differences exist between the secondary structure elements of the classes, particularly at helices 2 and 6. In addition, we find that in some ARFs, a putative helix 2b exists (Figure 5, region 300–400). However, this helix 2b occurs in ARFs from all three classes and consequently would not represent a DNA specificity selector. Helix 6 shows a class C specific insertion (Figure 5, region 500–550), which agrees with previous reports.[36] Helix 2 is positioned in the vicinity of the dimerization helix 6 (Figures 3B and 5, position 350) and its length may alter the flexibility of the dimerization interface and thus the spacing tolerance of the whole ARF dimer. Contrary to the case of helix 1, the lengths of helix 2 in *A. thaliana* and *M. polymorpha* class A and in species *A. trichopoda*, *O. sativa*, *P. patens*, *S. lycopersicum*, and *Z. mays* (Figure S3), this helix is around 10–12 amino acids long. Class B ARFs show more variability in the length of helix, which can decrease to 7 amino acids for some ARFs, although 75% of class B ARFs are in the range from 10 to 12 amino acids (Figure S3). However, for class C, helix 2 consists of 7–8 residues, considerably shorter than class A and B.

Structural superposition of the B3 domain provide insights into the structural restraints of the molecular caliper

In order to analyze the B3 position in the ARF DBDs, we superposed the various structures determined so far. The superposition of the two DNA-bound MpARF2-DBD structures

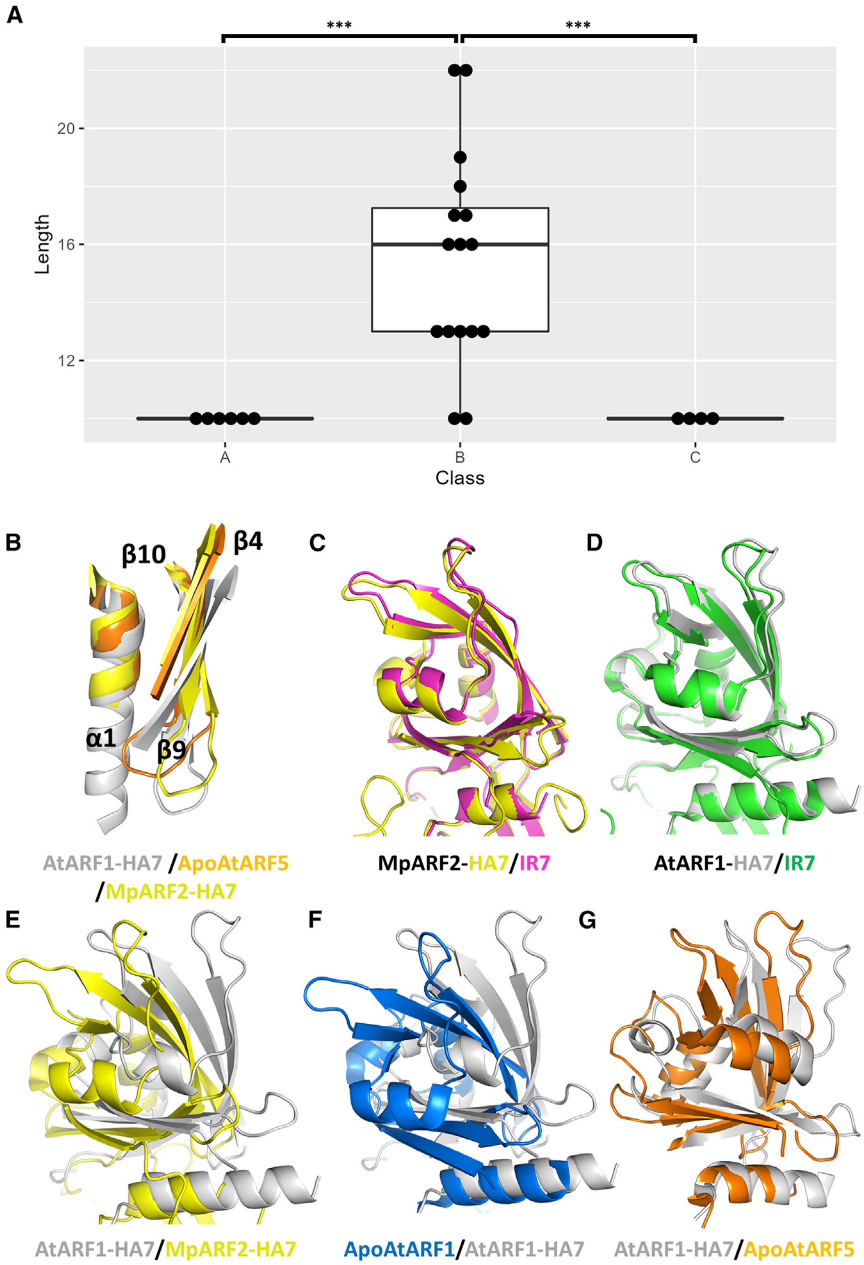


Figure 4. Comparison between helix 1 and B3 domains in IR7- and HA7-bound structures of the DBDs from MpARF2, AtARF1, and AtARF5

(A) Measured length of the α1 helix in *Marchantia polymorpha* and *Arabidopsis thaliana* ARFs, grouped by classes. *** means p value <0.001 obtained from Student's t two-tailed test for two-sample unequal variance.

(B–G) Visualization of the differences in length of the structural elements α1, β4, β9, and β10 after structural superposition of AtARF1, AtARF5, and MpARF2 DDs. The colors of the labels in (B–G) represent the colors of the structure shown. The corresponding RMSD values can be found in Table 3. Structures were superimposed on their dimerization domains (DDs) and their B3 domains are shown as cartoon representations. See also Figure S3.

resulted in a root-mean-square deviation (RMSD) of 1.027 Å for the backbone atoms (Figure 4C). Thus, binding of the HA7 or IR7 oligonucleotides (with similar spacing between binding sequences) led to similar overall structures for MpARF2 despite the difference in the inverted AuxRE elements (Figure 1A). The superposition of the structures of AtARF1-DBD in complex with IR7 or HA7 are also very similar[19] (Figure 4D). We concluded that different AuxRE sequences do not induce stress in the overall conformation of the protein. Locally, comparison of the DNA-binding interface of MpARF2-DBD in complex with HA7 or IR7 revealed that the conformation of His178 alternates between two states depending on the DNA sequence bound (Figure S4), as was previously reported for AtARF1-DBD in complex with HA7 or IR7.[19] This suggests that the selection mechanism for high affinity DNA-binding was established before *Marchantiophyta* and *Brassicaceae* divergence.

The structural comparisons described previously were based on superpositions of the full DBD structures. During our analysis, and especially during model building, we found that the most structurally invariant region among ARFs was the dimerization interface, while the position of the B3 domains was more variable. The superpositions of the full DBDs masked those movements, distributing the differences in B3 positions throughout the structure. To overcome this limitation, we performed a structural superposition of the ARF DDs and calculated RMSDs of the B3 domains alone (Figures 4E–4G; Table 3). The resulting RMSDs revealed that the position of the B3 domains of the MpARF2-HA7 and MpARF2-IR7 determined in this work are very similar to the previously reported structure of MpARF2 in complex with HA7 in a different space group[10] (Table 3, B3 RMSDs, **6SDG**). This shows that crystal packing does not affect the orientation of the B3 domains, and that the observed B3 positions are biologically relevant. Also, the B3 positions of the MpARF2-DBD in complex with HA7 or IR7 were closest to that of ApoAtARF5, but varied when compared to those of AtARF1-DBD (Table 3, B3 RMSDs). This observation is in line with the results of the *in vitro* assays, which show that the class B MpARF2-DBD behaves more like the class A AtARF5-DBD than the class B AtARF1-DBD.

To further quantify the differences in B3 positions among the structures, we measured the displacement of the Cα atoms of equivalent B3 residues from the different structures after the superposition of DDs used previously. The selected residues were

Figure 5. Secondary structure assignment for *M. polymorpha* and *A. thaliana* ARF alignment

The secondary structure layout of each ARF was superposed on a sequence alignment between *M. polymorpha* and *A. thaliana* ARFs. Each ARF is clustered according the functional classification. Green curly lines represent helices, orange arrows sheets, gray lines loops, and discontinuous lines represent gaps on the sequence alignment.

located in the β6 and β8 strands (P160 and R186 in AtARF1) and the last residue of the β7 strand (R181 in AtARF1) of the B3 domain (Figure 6). These residues were chosen because they are conserved between the ARFs analyzed and because they adequately reflect the magnitude of B3 movements. The average RMSD of the Cα atoms from the selected residues of MpARF2-HA7 and AtARF1-HA7 was 8.63 Å, but in superposed structures of MpARF2-HA7 and ApoAtARF1, the average RMSD of the same atoms was reduced to 4.77 Å. The lowest average RMSD (1.07 Å) was found between residues R223, P202, and R228 of MpARF2-HA7 and residues R215, P194, and R220 in ApoAtARF5. This is in line with the backbone RMSD calculation based on all DBD residues (Table 3, B3 RMSDs) calculated previously, which showed that the highest similarity in relative B3 position is found for the MpARF2-HA7 and ApoAtARF5 structures.

The monomers in the dimeric structures are related by a 2-fold symmetry through interactions between the DDs. Since these latter interactions lead to a relatively rigid structure, any variation in B3 position is caused by a change in distance between the B3 domains of the homodimer, or by a change in tilt angle of the B3 domains with respect to the DD domains. Alterations to the distance between B3 domains allow the ARFs to adapt to the spacing between AuxREs in palindromic DNAs, and is thus expected to strongly impact their spacing selectivity. We measured the distance between the B3 domains of homodimers, tracking the distance between residues homologous to Glu217 (in AtARF1) between the different pairs of structures, as these residues are located in the outermost part of the B3 domain of each monomer. The distance between B3 domains in ApoAtARF1 was reduced from 74.9 Å to 56.1 Å or 57.5 Å upon IR7 or HA7 binding, respectively. The B3 distance in ApoAtARF5 was 68.9 Å, decreasing to 54.9 Å in MpARF2-IR7 and 56.9 Å in MpARF2-HA7 (Table 3). Thus, the B3 distances are similar for the complexes binding IR7 and HA7 oligonucleotides, which was expected from their similar AuxRE spacing. It should be noted

Table 3. Quantitative analysis of the similarity of the B3 positions in distinct ARF-DBD structures, after superposition, and of the relative orientation and distances of the B3 domains within each of the dimers

B3 root-mean-square deviation (Å)	
MpARF2-HA7 – ApoAtARF1	3.35
MpARF2-IR7 – ApoAtARF1	3.23
MpARF2-HA7 – AtARF1-HA7	6.64
MpARF2-IR7 – AtARF1-HA7	5.55
ApoAtARF5 – MpARF2-HA7	1.43
ApoAtARF5 – AtARF1-HA7	5.20
ApoAtARF5 – ApoAtARF1	6.98
ApoAtARF1 – AtARF1-HA7	7.54
MpARF2-HA7 – 6SDG	0.579
MpARF2-IR7 – 6SDG	0.563
B3-DD off-axis tilt angle (°)	
ApoAtARF1	11
AtARF1-IR7	3
AtARF1-HA7	2
ApoAtARF5	23
MpARF2-HA7	15
MpARF2-IR7	15
Inter B3 distance (Å)	
ApoAtARF1	74.9
AtARF1-IR7	56.1
AtARF1-HA7	57.5
ApoAtARF5	68.9
MpARF2-HA7	56.9
MpARF2-IR7	54.9

The specified distance and RMSDs of B3s correspond to measures from the superposed dimerization domains (DDs). The twist angles represent the change in B3 position relative to DD with respect to the axis along the DNA duplex.

that ApoAtARF1 had the largest distance between B3 domains, and binding of oligonucleotides with 7 or 8 base pairs spacing required a large change in B3 distance, consistent with cooperativity observed in FA measurements (see previous text).

Another important parameter that facilitates ARF binding to IR sequences is the B3 off-axis tilt angle with respect to the axis along the DNA duplex. As the number of base pairs that separate the two inverted auxRE increases, the orientation between these two elements becomes more skewed leading to a rotation in the relative position of the major grooves of the AuxREs that are recognized by the B3 domains. The B3 positions need to adapt in order to accommodate this change in orientation. To quantify the off-axis tilt angle of the B3 domains relative to the DDs, we structurally superposed the DD of all the structures as described previously. Then, we traced two suitable vectors and determined the angle between them. To track the orientation of the DD in the different structures, a vector connecting the Cαs of the residues homologous to AtARF1 Ile73 was used, located in the DD of both monomers of the ARF structure. For orientation of the B3 domains, a second vector connecting the Cαs of the residues homologous to AtARF1 Glu217 was used. The angle between these two vectors in the plane perpendicular to the dimerization axis represents the off-axis tilt angle of the B3 domain pair with respect to the DD pair (Figure 6), and is also a measure of the planarity of the structure (Table 3). The B3 off-axis tilt angle in the ApoAtARF5 structure was 23°, whereas in ApoAtARF1 this angle was 11°, showing that ApoAtARF1 is more planar than ApoAtARF5, and that the B3 domains of the ApoAtARF5 can adapt to DNA binding sites that are oriented at larger relative angles (i.e., sites that are separated by more nucleotides). The B3 off-axis tilt angles for the AtARF1-DBD in complex with IR7 or HA7 were 3° and 2°, respectively, representing a reduction in tilt angle by 7°–8° for AtARF1 upon DNA binding. The tilt angle of the DNA-bound AtARF1-DBD structures was 20°–21° smaller than that of ApoAtARF5. In contrast, the B3 off-axis tilt angles of the DNA-bound structures of MpARF2-DBD (MpARF2-HA7 and MpARF2-IR7) were 15°, which is much closer to that of ApoAtARF5 (Table 1). This implies that DNA binds to MpARF2 in a slightly different orientation with respect to the plane through the DD domains, consistent with the observation that this structure is not as planar as AtARF1. Thus, the off-axis tilt angles between the B3 domains of the class A ApoAtARF5 and class B DNA-bound structures of MpARF2 are more similar than that of the DNA-bound structures of class B AtARF1. Unfortunately, we were not able to determine DNA-bound AtARF5 or apo MpARF2 structures, which would allow quantification of the magnitude of displacement upon DNA binding, although we can conclude that the DNA-bound structure of MpARF2-DBD resembles the apo AtARF5 structure more closely than of DNA-bound AtARF1.

DISCUSSION

Previous work showed that the DNA binding residues of the ARF B3 domains are highly conserved, and that sequence variation cannot therefore explain the observed specificity of different ARFs. The molecular caliper hypothesis explains how the virtually identical ARF DNA-binding interfaces can differentially regulate a large variety of genes. In this model, the positional flexibility of the B3 subdomains with respect to the DD varies, resulting in differential tolerance to variations in the spacing of the palindromic AuxREs.[18] We aimed to study the ARF-DNA and ARF-ARF interactions from different functional classes in more depth, linked to the evolutionary conservation of their sequences and function. As expected, we found that all DBDs could bind to the oligonucleotides containing inverted repeats of the high affinity HA sequence 5′-TGTCGG-3′ and the canonical IR sequence 5′-TGTCTC-3′ at spacings varying from 5 to 9 base pairs. Interestingly, the results indicate that class A and class B ARFs can bind spacings of 7 and 8 nucleotides as dimers readily, but bind to other spacings as monomers. SEC-MALS analysis on HA5 or HA7-bound AtARF1 confirms these results. Class C ARFs bind as monomers to all DNAs tested.

During the course of the experiments, we noticed consistent differences in behavior between MpARF2 and AtARF1, which belong to the same functional class B, as well as similarities between MpARF2 and AtARF5 (class A), which belong to different functional classes. Thus, AtARF1-DBD (class B) was found predominantly in the monomeric form in solution and homodimerization was induced by DNA binding as determined by SAXS,

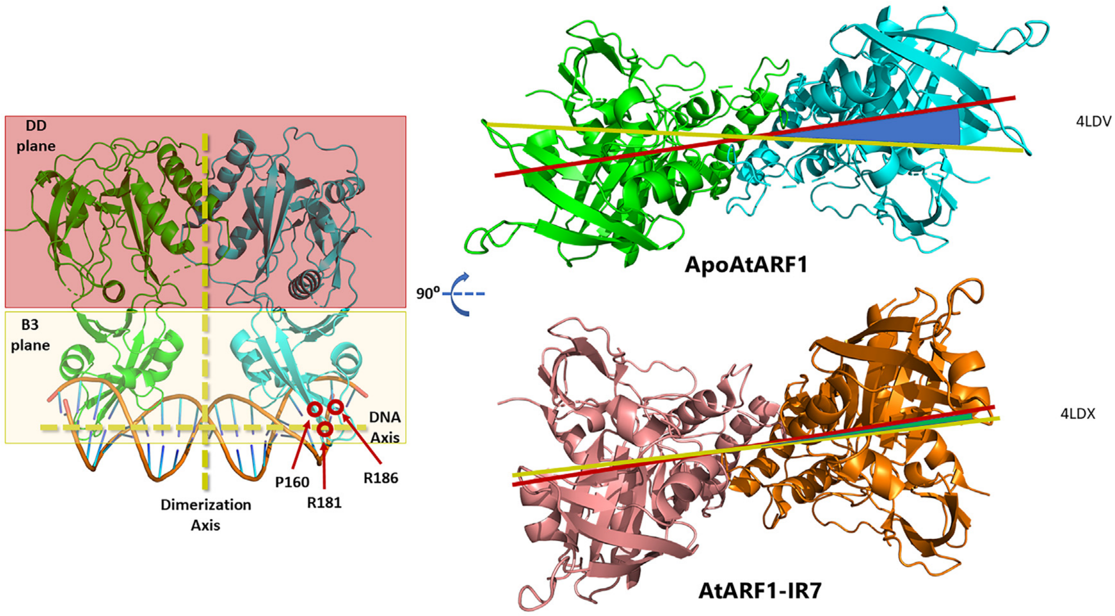


Figure 6. DD and B3 move relative to each other to accommodate for DNA

The left panel illustrates the definition of the two planes that form the B3-DD off-axis tilt angle: the "DD plane" (in red) and the "B3 plane" in yellow. In the right panel, cartoon representations of the structure of ApoAtARF1 (4LDV, *top*) and the structure of AtARF1-IR7 (4LDX, *lower*) are shown. In the case of AtARF1, the apo structure shows a larger angle (represented in blue), while the DNA-bound AtARF1 shows a smaller angle (represented in green), indicating a more planar structure. The DNA in the DNA-bound AtARF1 was omitted for clarity.

whereas the apo forms of the class A AtARF5-DBD and class B MpARF2-DBD were predominantly found as dimers. In addition, SEC analysis showed that MpARF2- and AtARF5-DBDs had similar binding patterns for DNA sequences in terms of spacing and sequence of the AuxREs. Even the protein stability of AtARF5-DBD and MpARF2-DBD were intrinsically lower in the *in vitro* conditions compared to AtARF1-DBD.

The *in vitro* behavioral differences of class A AtARF5 and B AtARF1/MpARF2 DBDs can be explained by structural specificities we found on the X-ray structures determined for the class A and B ARFs. We observed that the position of the B3 domains on the class B MpARF2-DBD in complex with IR7 or HA7 oligonucleotides was more similar to the class A ApoAtARF5 structure than to either the apo or DNA-bound forms of the class B AtARF1-DBD. The distinct positions of the B3 domain in the different structures is likely related to the length of helix 1, which functions as an anchor of the B3 subdomain position relative to the DD. AlphaFold2 generated models of all ARFs from a range of plant species confirm that helix 1 can be longer in some class B ARFs, but is constant in in class A and C ARFs. As the length of helix 1 increases, the number of contacts between the DD and B3 domain also increases, resulting in a more constrained placement of the B3 domain. Therefore, the longer helix 1 in some class B ARFs may induce a separation of the B3 subdomains in an open conformation, which is more amenable to the 8 base pairs spacing. In contrast, the shorter helix 1 in class A ARFs would not restrict DD/B3 movement as in class B, allowing wider range of specificity. Based on this analysis, we propose helix 1 length as an AuxRE spacing selector on ARFs. This analysis did not reveal other systematic differences between the classes with the exception of the dimerization region of class C when compared with class A and B. Helix 2 is shorter in class C ARFs and helix 6 is interrupted by a long unstructured sequence. Both are predicted to disrupt class C ARF dimerization.

Thus, our *in vitro* results using ARF DBDs do not match the ARF functional classification described previously, as class B MpARF2 and class A AtARF5 show more similar DNA-binding and oligomerization behavior than class B AtARF1. Our results therefore indicate that ARF classification is not directly related with DNA binding but reflects a different aspect of their function and possibly alterations in their interactions with other proteins, suggesting that DNA recognition did not coevolve with other functions of the ARF protein. Given that ARFs from the three classes are associated with activation or repression of target genes, it appears that the selection of the specificity of DBD has evolved independently between classes, i.e., of the transcriptional output. The question then is: what are the sequence motifs that determine the classification of the ARFs, and how the function of the ARFs is related with these motifs on a structural level?

The answer to this question should be corroborated in the context of the full-length ARF proteins and has been partially addressed in a previous work,[34] studying detailed DNA-binding kinetics of full-length ARFs using smFRET. This study shows that the results described herein also apply to full-length ARFs. However, the hypothesis that emerges from our work is that class-independent combinations of target selection and distinct repression/activation outcomes would further increase the complexity of the NAP pathway. In the current model of ARF gene selection, dimerization and gene transcription are dependent on the relative protein concentration and AuxRE availability,[10] where fine-tuning of ARF-regulated gene expression is achieved by control of ARF availability in several ways: control of ARF expression[41,42]; posttranslational modifications including phosphorylation[43]; interaction with other partners[44,45]; relative class A/B availability[10,13]; DNA exposure by chromatin compaction[46]; or receptor competition for DNA binding sites.[47,48] Our results unveil an additional regulatory layer based on DBD structural flexibility, where target selection and transcriptional output have evolved separately and are not strictly coupled.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by Dr. Dirk Roeland Boer (rboer@cells.es, or find the current contact using ORCID: 0000-0001-5949-6627).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Atomic coordinates of MpARF2-DBD in complex with HA7 and IR7 DNAs have been deposited at Protein DataBank under codes PDB: 8OJ1 and PDB: 8OJ2, respectively, and are publicly available as of the date of publication. Accession numbers are also listed in Tables S4 and S5 and in the key resources table.
- SAXS data of AtARF1-DBD in complex with HA7, Apo MpARF3-DBD, and MpARF3-DBD in complex with HA6 have been deposited at SASBDB under codes SASBDB: SASDVV5, SASBDB: SASDVT5, and SASBDB: SASDVU5, respectively, and are publicly available as of the date of publication. Accession numbers are also listed in Table S3 and the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Jan-Willem Borst (Wageningen University, The Netherlands) for suggestions and corrections of the manuscript. This work was supported by Ministry of Economy and Competitiveness of the Spanish Government grants PID2020-117028 GB-I00 (AEI/FEDER, EU), BIO2016-77883-C2-2-P (AEI/FEDER, EU), and FIS2015-72574-EXP (AEI/FEDER, EU) to D.R.B.; a PhD fellowship from the graduate school Experimental Plant Sciences to J.R.; and an NWO-VICI grant (number 865.14.001) to D.W. Funding for open access charge: Ministry of Economy and Competitiveness of the Spanish Government PID2020-117028 GB-I00.

AUTHOR CONTRIBUTIONS

I.C., conceptualization, investigation, methodology, formal analysis, writing – original draft, writing – review & editing. M.M., methodology, formal analysis. J.R., investigation, methodology, formal analysis. A.T.-S., investigation, writing – review & editing. W.v.d.B., resources, methodology. D.W., conceptualization, resources, formal analysis, writing – review & editing, project administration. D.R.B., conceptualization, resources, formal analysis, writing – review & editing, project administration.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Microbial strains
- METHOD DETAILS
  - Protein expression and purification
  - Analytical size exclusion chromatography
  - SEC-MALS
  - Fluorescence anisotropy|fluorescent anisotropy assays
  - Crystallization and structure determination
  - Root-mean-square deviation and distance calculation
  - Determination and analysis of secondary structure elements
  - Small-angle X-ray scattering analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2025.02.006.

Received: January 28, 2024
Revised: August 8, 2024
Accepted: February 17, 2025
Published: March 13, 2025

REFERENCES

1. Do, B.H., Phuong, V.T.B., Tran, G.B., and Nguyen, N.H. (2019). Emerging functions of chromatin modifications in auxin biosynthesis in response to environmental alterations. Plant Growth Regul. *87*, 165–174. https://doi.org/10.1007/s10725-018-0453-x.
2. Paque, S., and Weijers, D. (2016). Q&: Auxin: the plant molecule that influences almost anything. BMC Biol. *14*, 67. https://doi.org/10.1186/s12915-016-0291-0.
3. Perrot-Rechenmann, C. (2010). Cellular Responses to Auxin: Division versus Expansion. Cold Spring Harb. Perspect. Biol. *2*, a001446. https://doi.org/10.1101/cshperspect.a001446.
4. Woodward, A.W., and Bartel, B. (2005). Auxin: Regulation, action, and interaction. Ann. Bot. *95*, 707–735. https://doi.org/10.1093/aob/mci083.
5. Weijers, D., and Wagner, D. (2016). Transcriptional Responses to the Auxin Hormone. Annu. Rev. Plant Biol. *67*, 539–574. https://doi.org/10.1146/annurev-arplant-043015-112122.
6. Lavy, M., and Estelle, M. (2016). Mechanisms of auxin signaling. Development *143*, 3226–3229. https://doi.org/10.1242/dev.131870.
7. Kato, H., Nishihama, R., Weijers, D., and Kohchi, T. (2018). Evolution of nuclear auxin signaling: lessons from genetic studies with basal land plants. J. Exp. Bot. *69*, 291–301. https://doi.org/10.1093/jxb/erx267.
8. Chandler, J.W. (2016). Auxin response factors. Plant Cell Environ. *39*, 1014–1028. https://doi.org/10.1111/pce.12662.
9. Vernoux, T., Brunoud, G., Farcot, E., Morin, V., Van Den Daele, H., Legrand, J., Oliva, M., Das, P., Larrieu, A., Wells, D., et al. (2011). The auxin signalling network translates dynamic input into robust patterning at the shoot apex. Mol. Syst. Biol. *7*, 508. https://doi.org/10.1038/msb.2011.39.
10. Kato, H., Mutte, S.K., Suzuki, H., Crespo, I., Das, S., Radoeva, T., Fontana, M., Yoshitake, Y., Hainiwa, E., van den Berg, W., et al. (2020). Design principles of a minimal auxin response system. Nat. Plants *6*, 473–482. https://doi.org/10.1038/s41477-020-0662-y.
11. Guilfoyle, T.J. (2015). The PB1 Domain in Auxin Response Factor and Aux/IAA Proteins: A Versatile Protein Interaction Module in the Auxin Response. Plant Cell *27*, 33–43. https://doi.org/10.1105/tpc.114.132753.
12. Finet, C., Berne-Dedieu, A., Scutt, C.P., and Marlétaz, F. (2013). Evolution of the ARF gene family in land plants: Old domains, new tricks. Mol. Biol. Evol. *30*, 45–56. https://doi.org/10.1093/molbev/mss220.
13. Mutte, S.K., Kato, H., Rothfels, C., Melkonian, M., Wong, G.K.-S., and Weijers, D. (2018). Origin and evolution of the nuclear auxin response system. Elife *7*, e33399. https://doi.org/10.7554/eLife.33399.

14. Flores-Sandoval, E., Eklund, D.M., Hong, S.F., Alvarez, J.P., Fisher, T.J., Lampugnani, E.R., Golz, J.F., Vázquez-Lobo, A., Dierschke, T., Lin, S.S., and Bowman, J.L. (2018). Class C ARFs evolved before the origin of land plants and antagonize differentiation and developmental transitions in Marchantia polymorpha. New Phytol. *218*, 1612–1630. https://doi.org/10.1111/nph.15090.

15. Tiwari, S.B., Hagen, G., and Guilfoyle, T. (2003). The Roles of Auxin Response Factor Domains in Auxin-Responsive Transcription. Plant Cell *15*, 533–543. https://doi.org/10.1105/tpc.008417.

16. Guilfoyle, T.J., and Hagen, G. (2007). Auxin response factors. Curr. Opin. Plant Biol. *10*, 453–460. https://doi.org/10.1016/j.pbi.2007.08.014.

17. Roosjen, M., Paque, S., and Weijers, D. (2018). Auxin Response Factors: output control in auxin biology. J. Exp. Bot. *69*, 179–188. https://doi.org/10.1093/jxb/erx237.

18. Boer, D.R., Freire-Rios, A., Van Den Berg, W.A.M., Saaki, T., Manfield, I.W., Kepinski, S., López-Vidrieo, I., Franco-Zorrilla, J.M., De Vries, S.C., Solano, R., et al. (2014). Structural Basis for DNA Binding Specificity by the Auxin-Dependent ARF Transcription Factors. Cell *156*, 577–589. https://doi.org/10.1016/j.cell.2013.12.027.

19. Freire-Rios, A., Tanaka, K., Crespo, I., van der Wijk, E., Sizentsova, Y., Levitsky, V., Lindhoud, S., Fontana, M., Hohlbein, J., Boer, D.R., et al. (2020). Architecture of DNA elements mediating ARF transcription factor binding and auxin-responsive gene expression in Arabidopsis. Proc. Natl. Acad. Sci. USA *117*, 24557–24566. https://doi.org/10.1073/pnas.2009554117.

20. Korasick, D.A., Westfall, C.S., Lee, S.G., Nanao, M.H., Dumas, R., Hagen, G., Guilfoyle, T.J., Jez, J.M., and Strader, L.C. (2014). Molecular basis for AUXIN RESPONSE FACTOR protein interaction and the control of auxin response repression. Proc. Natl. Acad. Sci. USA *111*, 5427–5432. https://doi.org/10.1073/pnas.1400074111.

21. Kubeš, M., and Napier, R. (2019). Non-canonical auxin signalling: fast and curious. J. Exp. Bot. *70*, 2609–2614. https://doi.org/10.1093/jxb/erz111.

22. Nanao, M.H., Vinos-Poyo, T., Brunoud, G., Thévenon, E., Mazzoleni, M., Mast, D., Lainé, S., Wang, S., Hagen, G., Li, H., et al. (2014). Structural basis for oligomerization of auxin transcriptional regulators. Nat. Commun. *5*, 3617. https://doi.org/10.1038/ncomms4617.

23. Kim, J., Harter, K., and Theologis, A. (1997). Protein-protein interactions among the Aux/IAA proteins. Proc. Natl. Acad. Sci. USA *94*, 11786–11791. https://doi.org/10.1073/pnas.94.22.11786.

24. Dinesh, D.C., Kovermann, M., Gopalswamy, M., Hellmuth, A., Calderón Villalobos, L.I.A., Lilie, H., Balbach, J., and Abel, S. (2015). Solution structure of the PsIAA4 oligomerization domain reveals interaction modes for transcription factors in early auxin response. Proc. Natl. Acad. Sci. USA *112*, 6230–6235. https://doi.org/10.1073/pnas.1424077112.

25. Stigliani, A., Martin-Arevalillo, R., Lucas, J., Bessy, A., Vinos-Poyo, T., Mironova, V., Vernoux, T., Dumas, R., and Parcy, F. (2019). Capturing auxin response factors syntax using DNA binding models. Mol. Plant *12*, 822–832. https://doi.org/10.1016/j.molp.2018.09.010.

26. Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: Structure, function, and evolution. Trends Plant Sci. *18*, 267–276. https://doi.org/10.1016/j.tplants.2012.09.001.

27. Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Tomo, Y., et al. (2004). Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. Plant Cell *16*, 3448–3459. https://doi.org/10.1105/tpc.104.026112.

28. Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). 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. *27*, 470–478. https://doi.org/10.1093/nar/27.2.470.

29. Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J.L., Nishihama, R., and Kohchi, T. (2015). Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. PLoS Genet. *11*, e1005084. https://doi.org/10.1371/journal.pgen.1005084.

30. Chen, W.F., Wei, X.B., Rety, S., Huang, L.Y., Liu, N.N., Dou, S.X., and Xi, X.G. (2019). Structural analysis reveals a "molecular calipers" mechanism for a LATERAL ORGAN BOUNDARIES DOMAIN transcription factor protein from wheat. J. Biol. Chem. *294*, 142–156. https://doi.org/10.1074/jbc.RA118.003956.

31. Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E.H., Schmid, M., Jürgens, G., and Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature *464*, 913–916. https://doi.org/10.1038/nature08836.

32. Yamaguchi, N., Wu, M.F., Winter, C.M., Berns, M.C., Nole-Wilson, S., Yamaguchi, A., Coupland, G., Krizek, B.A., and Wagner, D. (2013). A Molecular Framework for Auxin-Mediated Initiation of Flower Primordia. Dev. Cell *24*, 271–282. https://doi.org/10.1016/j.devcel.2012.12.017.

33. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: A Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. *36*, 1277–1282. https://doi.org/10.1107/S0021889803012779.

34. Fontana, M., Roosjen, M., Crespo García, I., van den Berg, W., Malfois, M., Boer, R., Weijers, D., and Hohlbein, J. (2022). Cooperative action of separate interaction domains promotes high-affinity DNA binding of Arabidopsis thaliana ARF transcription factors. Preprint at bioRxiv. https://doi.org/10.1101/2022.11.16.516730.

35. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., De Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303. https://doi.org/10.1093/nar/gky427.

36. Martin-Arevalillo, R., Thévenon, E., Jégu, F., Vinos-Poyo, T., Vernoux, T., Parcy, F., and Dumas, R. (2019). Evolution of the Auxin Response Factors from charophyte ancestors. PLoS Genet. *15*, e1008400. https://doi.org/10.1371/journal.pgen.1008400.

37. Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from x-ray solution scattering. Biophys. J. *80*, 2946–2953. https://doi.org/10.1016/S0006-3495(01)76260-1.

38. Li, S.-B., Xie, Z.-Z., Hu, C.-G., and Zhang, J.-Z. (2016). A Review of Auxin Response Factors (ARFs) in Plants. Front. Plant Sci. *7*, 47. https://doi.org/10.3389/fpls.2016.00047.

39. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

40. Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. *40*, D1178–D1186. https://doi.org/10.1093/nar/gkr944.

41. Marin, E., Jouannet, V., Herz, A., Lokerse, A.S., Weijers, D., Vaucheret, H., Nussaume, L., Crespi, M.D., and Maizel, A. (2010). miR390, Arabidopsis TAS3 tasiRNAs, and Their AUXIN RESPONSE FACTOR Targets Define an Autoregulatory Network Quantitatively Regulating Lateral Root Growth. Plant Cell *22*, 1104–1117. https://doi.org/10.1105/tpc.109.072553.

42. Yang, J.H., Han, S.J., Yoon, E.K., and Lee, W.S. (2006). Evidence of an auxin signal pathway, microRNA167-ARF8-GH3, and its response to exogenous auxin in cultured rice cells. Nucleic Acids Res. *34*, 1892–1899. https://doi.org/10.1093/nar/gkl118.

43. Han, S., and Hwang, I. (2018). Integration of multiple signaling pathways shapes the auxin response. J. Exp. Bot. *69*, 189–200. https://doi.org/10.1093/jxb/erx232.

44. Oh, E., Zhu, J.Y., Bai, M.Y., Arenhart, R.A., Sun, Y., and Wang, Z.Y. (2014). Cell elongation is regulated through a central circuit of interacting transcription factors in the Arabidopsis hypocotyl. Elife *3*, e03031. https://doi.org/10.7554/eLife.03031.

45. Kelley, D.R., Arreola, A., Gallagher, T.L., and Gasser, C.S. (2012). ETTIN (ARF3) physically interacts with KANADI proteins to form a functional complex essential for integument development and polarity determination in Arabidopsis. Development *139*, 1105–1109. https://doi.org/10.1242/dev.067918.

46. Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Curr. Opin. Genet. Dev. *21*, 175–186. https://doi.org/10.1016/j.gde.2011.01.022.

47. Novac, N., Baus, D., Dostert, A., and Heinzel, T. (2006). Competition between glucocorticoid receptor and NFκB for control of the human FasL promoter. FASEB J. *20*, 1074–1081. https://doi.org/10.1096/fj.05-5457com.

48. Zabet, N.R., and Adryan, B. (2013). The effects of transcription factor competition on gene regulation. Front. Genet. *4*, 197. https://doi.org/10.3389/fgene.2013.00197.

49. Rossi, A.M., and Taylor, C.W. (2011). Analysis of protein-ligand interactions by fluorescence polarization. Nat. Protoc. *6*, 365–387. https://doi.org/10.1038/nprot.2011.305.

50. Juanhuix, J., Gil-Ortiz, F., Cuní, G., Colldelram, C., Nicolás, J., Lidón, J., Boter, E., Ruget, C., Ferrer, S., and Benach, J. (2014). Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the Alba Synchrotron. J. Synchrotron Radiat. *21*, 679–689. https://doi.org/10.1107/S160057751400825X.

51. Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132. https://doi.org/10.1107/S0907444909047337.

52. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221. https://doi.org/10.1107/S0907444909052925.

53. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501. https://doi.org/10.1107/S0907444910007493.

54. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315. https://doi.org/10.1002/pro.3330.

55. González Fernández, J., González, N., Colldelram, C., Ribó, A., Fontserè, A., Jover-Manas, G., Villanueva, J., Llonch, M., Peña, G., Gevorgyan, A., et al. (2018). NCD-SWEET Beamline Upgrade. Mech. Eng. Des. Synchrotron Radiat. Equip. Instrum. 374–376. https://doi.org/10.18429/JACoW-MEDSI2018-THPH17.

56. Gonzalez, N., González, J.B., Colldelram, C., Llonch, M., Ladrera, J., Fontserè, A., Martínez, J.C., Kamma-Lorger, C., Sics, I., Ferrer, S., et al. (2018). Beam Conditioning Optics At The ALBA NCD-SWEET Beamline. Mech. Eng. Des. Synchrotron Radiat. Equip. Instrum. 365–367. https://doi.org/10.18429/JACoW-MEDSI2018-THPH14.

57. Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., and Svergun, D.I. (2017). ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. *50*, 1212–1225. https://doi.org/10.1107/S1600576717007786.

58. Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. *28*, 768–773. https://doi.org/10.1107/S0021889895007047.

59. Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. *34*, 33–41. https://doi.org/10.1107/S0021889800014126.

Structure
Article

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E.coli Rosetta 2	Sigma-Aldrich	Cat#71402-M
Deposited data		
MpARF2DBD-HA7 structure	This paper	PDB: 8OJ1
MpARF2DBD-IR7 structure	This paper	PDB: 8OJ2
AtARF1DBD-HA7	This paper	SASBDB: SASDVV5
Apo MpARF3DBD	This paper	SASBDB: SASDVT5
MpARF3DBD-HA6	This paper	SASBDB: SASDVU5
Oligonucleotides		
dsDNA HA5: 5'-TGTCGG CATTG CCGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA HA6: 5'-TGTCGG CGATCG CCGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA HA7: 5'-TGTCGG CGATTCG CCGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA HA8: 5'-TGTCGG CGATATCG CCGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA HA9: 5'-TGTCGG CGATTATCG CCGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA IR7: 5'-TGTCTC CCTTTCG GAGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA IR8: 5'-TGTCTC CCAAAAGG GAGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA LFY: 5'-TGTCAA TTTCCCAGC AAGACA-'3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA TMO3: 5'-GGTCAA AAGTAAGAC TGGACC-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA TMO5: 5'-GGTCTC TGGTCGG TCGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA TMO5Δ1: 5'-GGTCTC TGGTCGG TTTTTT-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA TMO5Δ2: 5'-TTTTTT TGGTCGG TCGACA-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
dsDNA DR5: 5'- TGTCTC CCTTT TGTCTC-3' and complementary strand (synthetized by Biomers GmbH)	This paper	N/A
Other		
Chitin resin	New England Biolabs	Cat#S6651S

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Microbial strains

AtARF1, AtARF5, MpARF2 and MpARF3 were expressed in *E. coli* Rosetta 2 cells.

METHOD DETAILS

Protein expression and purification

AtARF-DBDs and MpARF-DBDs were expressed and purified using standard protocols as previously reported.[18] In brief, fusions of ARF-DBDs with chitin-binding domain and Intein were expressed in E. coli strain Rosetta 2 (Novagen). Protein expression was induced by addition of 0.3 mM IPTG to the growth medium (TB) for 20 hr at 20°C, and proteins were purified from cell-free extracts in 50 mM HEPES pH 8, 1 mM EDTA, 500 mM NaCl, 2 mM $MgCl_2$, 0.1% (v/v) Tween20 by affinity chromatography on a chitin column (New England Biolabs). Column was washed with 10 column volumes of column buffer (20mM Tris pH 8, 50mM NaCl, 0.1% (v/v)

Tween20), and protein was eluted from the column by 2 hours incubation with column buffer supplemented with 50 mM DTT. Protein from affinity purification was then further purified to homogeneity by size exclusion chromatography on a ProSEC 3-70, 16-600 mm column (Generon Ltd), in 20 mM Tris pH 8, 500 mM NaCl. Both chromatographic steps were run on an ÄKTA Pure 25M (Cytiva) and recording 280 and 260 nm absorbance. Fractions containing the ARF peak were concentrated and filtered through 0.2 μm. Purity was assessed to be >95% by SDS–PAGE, followed by Coomassie Blue staining. Final protein concentration was determined by absorbance at 280 nm using a Nanodrop spectrophotometer and protein was used for assays immediately where possible or stored in aliquots at -80°C.

Analytical size exclusion chromatography

500 picomole of each tested dsDNA was diluted in SEC buffer (15 mM HEPES pH 7.5, 150 mM NaCl) to a final volume of 25 μl (20 μM final DNA concentration) and were injected on a Superdex 200 increase 5/150 (GE Healthcare) equilibrated in SEC buffer. The dsDNA references, in absence of the protein, were repeated at least 2 times and the standard deviation in V_{el} was calculated. Mixtures of protein:DNA at a molar ratio of 2:1 (protein:DNA), at a final protein concentration of 40 μM, were incubated on ice for 1 h, maintaining as final buffer the SEC buffer. For ApoARF control, the same amount of protein used for the assays was injected (25 μl at 40 μM, 1 nmol). The chromatography was run at 0.3ml/min on an ÄKTA Pure 25M and 280 and 260 nm absorbances were recorded.

SEC-MALS

AtARF1 apo was mixed with HA5 and HA7, respectively, at a 2:1 AtARF1:DNA stoichiometry at 40 and 80 μM concentrations of AtARF1. In addition, mixtures of AtARF1 with HA5 and HA7 were prepared at 10:1 stoichiometry of AtARF1:dsDNA, respectively, at a 200 μM concentration of AtARF1. All the samples were prepared in a final buffer consisting on 15 mM HEPES pH 7.5, 150 mM NaCl. The apo protein at 40 μM, both DNAs at 20 μM and the 6 mixtures described previously were subjected to SEC-MALS experiments. To this end, the samples were filtered over a Proteus Clarification Mini Spin Column 0.22 μm and 35 or 48 μl were introduced in an autosampler (SIL-20AC, Shimadzu) connected to a high-performance liquid chromatography system (Prominence, Shimadzu), equipped with a Superdex 200. The samples were eluted at 0.5 ml/min during 50 minutes at a temperature of 25°C. The elution from SEC was monitored by a UV detector (SPD-20A, Shimadzu), differential refractometer (Optilab T-rEX, Wyatt Corp.) and static multi-angle laser LS detector (DAWN HELEOS-II, Wyatt Corp.). The SEC-UV-MALS-RI system was equilibrated at the flow rate of 0.5 mL/min for at least 2 h before carrying out the measurements.

ASTRA 7 software (Wyatt Corp.) was used for data collection and analysis. From the light scattering of selected elution ranges, the molecular weight, polydispersity and total mass were calculated. Finally, the mass fraction of each of the selected elution ranges was calculated with respect to the total analyzed mass.

Fluorescence anisotropy/fluorescent anisotropy assays

Binding of the DNA to the DBDs of AtARF1 and AtARF5 was assayed in a CLARIOstar plate reader (BMG Labtech) on OptiPlate-384 Black well plates (PerkinElmer) in 10 μL final assay volume. DNAs used for FA measurements were synthesized labelled with fluorescein at the 5' end of the forward chain and purified by Biomers GmbH and annealed with the corresponding unlabeled reverse chain. The buffer used for anisotropy assays was 20 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT and 0.01% Triton X-100. The final assay ARF concentration was varied from 0.1 to 200 μM and the labeled DNA concentration was 50 nM. An excitation wavelength of 485 nm and an emission wavelength of 528 nm were used. The data was measured at 25°C and corrected for background by subtracting the free-labeled DNA signal. Plates were read immediately after preparation and later at 4 and 24 h post preparation, storing plates in the dark at 4°C. Plates were warmed at room temperature for 30 minutes before reanalysis. While no differences were detected in the first 4 hours post-preparation, measures at 24h were not considered due to significant fluorescence intensity loss and data accuracy. All the data treatment was done as previously described.[49] Both a Michaelis Menten as well as a Hill model were fitted to the data. The selection of the best model was based on the lowest coefficient of variation for K_D, which represents the fit to the experimental data. In case the Hill model resulted in a coefficient of variation of zero, it was assumed that the model was overparametrized and the Michaelis Menten model was used.

Please cite this article in press as: Crespo et al., The structure and function of the DNA binding domain of class B MpARF2 share more traits with class A AtARF5 than to that of class B AtARF1, Structure (2025), https://doi.org/10.1016/j.str.2025.02.006

Crystallization and structure determination

Purified MpARF2-DBD was concentrated and mixed with the corresponding annealed dsDNA (**HA7:** 5′-d(TTGTCGGCGATTCGCC GACAA)-3′, **IR7:** 5'-d(TTGTCTCCCTTTGGGAGACAA)-3') at a ratio of 2:1 (protein:DNA), to a final protein concentration of 4.5-5 mg/ml in storage buffer (**HA7:** 20 mM Tris (pH 8.0), 500 mM NaCl, 20 mM DTT, **IR7:** 15 mM Sodium HEPES (pH 7.5), 150 mM NaCl, 20 mM DTT). Crystals giving the highest resolution were obtained by sitting-drop vapor-diffusion at 18°C, equilibrating drops of 1 μl protein + 1 μl crystallization buffer (**HA7:** 100 mM Tris pH 7, 27% PEG2KMME, 18.5% (v/v) Glycerol; **IR7:** 0.2 M Ammonium sulfate, 0.1 M MES pH 6.5, 24% PEG 3350) against 100 μl of the crystallization buffer. Cube-shaped crystals took one week to grow to the maximum size. Cryo-cooling in liquid nitrogen did not require using a cryo-protecting solution with the crystals and was performed by direct plunge-freezing in $N_{2(L)}$. Data collection was performed to the indicated resolutions at XALOC beamline (BL13, ALBA Synchrotron Light Source, Barcelona).[50] The crystals belonged to space group $I2_12_12_1$, with one protein molecule and one DNA sequence in the asymmetric unit. Data were processed with XDSgui.[51]

The structure of MpARF2:HA7 at C_2 spacegroup (PDB 6SDG) was used for phasing the data by molecular replacement. The resulting electron density map was processed with phenix.autobuild (v 1.19.2-4158[52]) for reconstructing the whole protein molecule. Both chains of the DNA molecule were manually built with Coot (v0.9.6,[53]) using the 2Fo−Fc and Fo−Fc electron-density maps. All the following refinements were done with Phenix.refine (v 1.19.2-4158[52]) and manual adjustments and reconstructions were made with Coot using the 2Fo−Fc and Fo−Fc electron-density maps. MolProbity[54] was used to validate and improve the final model. Refinement statistics are presented in Tables S4 and S5. Figures were prepared using Pymol (The PyMOL Molecular Graphics System, Version 2.4.0a0, Schrödinger, LLC).

Root-mean-square deviation and distance calculation

RMSD and distances were calculated using PyMOL (The PyMOL Molecular Graphics System, Version 2.4.0a0, Schrödinger, LLC). All structure RMSD were calculated for all the Cα of the corresponding chains using the PyMOL built-in align function. In the case of B3 RMSD calculation, the Dimerization Domains of the analyzed ARFs were aligned as described for all structure RMSD. Then, the RMSD for the Cα atoms of the B3 domains were computed using PyMOL built-in function rms_cur, providing the matchmaker=-1 argument for proper calculations.

Determination and analysis of secondary structure elements

ARF sequences available for the organisms *A. thaliana, M. polymorpha, A. trichopoda*, *O. sativa*, *P. patens*, *S. lycopersicum* and *Z. mays* were retrieved from phytozome[40] and a Blast search against Uniprot.org was used to obtain the accession numbers and AlphaFold2[39] models. For ARF with solved crystallographic structures, the ARF structures were retrieved from the PDB. Determination of secondary structure elements were done in batch with in-house developed Python scripts executing phenix.secondary_structure_restraints on each of the AlphaFold2 models. The search method was selected based on the output of the algorithms available (cablam, from_ca, ksdssp and mmtbx_dssp) compared with the information from the annotated PDBs of the available crystallographic structures. The cablam method yielded the worst results in all the cases, as cablam only predicted helices, not sheets. The calculation using Cα atoms was more accurate, although the best results were obtained with the default KSDSSP and the DSSP implementation from the Phenix's macromolecular toolbox (mmtbx_dssp). We finally chose the mmtbx_dssp implementation as the correlation of the results with the PDB reference (95.96 %) was slightly higher than for the default KSDSSP implementation (94.50 %).

For the generation of the structural alignment, the secondary structural elements generated for all the models were assigned to a sequence alignment of all the analyzed ARFs calculated by MUSCLE. For the calculation of structural elements length, every protein model was structurally aligned to 4ldx structure with an in-house Python script running pymol align command. The alignment object generated by pymol was used to map the position of the aligned residues in the model with respect to the reference, and to compare the mapped positions with the tabulated structural elements generated with mmtbx_dssp for the model and the reference. The reported element length corresponds to the length of the structural feature on the model that at least partially overlaps with the same structural feature on the reference. Statistical significance for the helix 1 length of each class was obtained from students-t two-tailed test for two-sample unequal variance. A p-value < 0.001 was labeled with ***.

Small-angle X-ray scattering analysis

Different concentrations of AtARF1, AtARF5, MpARF2 and MpARF3 ranging from 0.7 mg/ml to 7 mg/ml were tested to record protein dimerization depending on concentration. For protein:DNA assays, a fixed protein concentration of 3.5 mg/ml was used as it gave the best signal, while varying DNA concentration for 4:1, 2:1 and 1:1 protein:DNA stoichiometries. All the samples were prepared in a final buffer consisting on 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT. SAXS data was collected at NCD-SWEET beamline (BL11, ALBA Synchrotron Light Source, Barcelona).[55,56] The buffer and the buffer + DNA in all concentrations were collected for subtraction of protein samples. Measurements were carried out at 293 K in a quartz capillary of 1.5 mm outer diameter and 0.01 mm wall thickness. The data (20 frames with an exposure time of 0.5 sec/frame) was recorded using a Pilatus 1M detector (Dectris, Switzerland) at a sample-detector distance of 2.56 m and a wavelength of λ=1.0 Å.

Buffer subtraction and extrapolation to infinite dilution were performed by using the program package primus/qt from the ATSAS 3.0.4 (r13469) software suite.[57] The forward scattering *I(0)* and the radius of gyration (R_g) were evaluated by the Guinier approximation, and the maximum distance D_{max} of the particle was also computed from the entire scattering patterns with AutoGNOM. The excluded volume V_p of the particle was computed from the Porod invariant. The scattering from the crystallographic models was computed with CRYSOL.[58] The volume fractions of the species that may be present were determined with OLIGOMER,[33] using as template the available PDB structures to generate apo protein monomers and dimers, as well as monomers and dimers bound to the oligonucleotides. The unbound nucleotides were also included in the analysis. The models of MpARF3 were generated from the crystallographic structure of AtARF1 in complex with ER7 (PDB id: 4LDX), since no crystallographic structures are available. In the case of AtARF5, the DNA was positioned by structural alignment with MpARF2. Calculations using all five models did not improve the fit compared to a mixture of monomers and dimers bound to DNA. Therefore, the reported fits are those obtained from a mixture of monomers and dimers bound to DNA.

The low resolution shape of apo MpARF3 was reconstructed *ab initio* with GASBOR[37] run in batch mode, using the curve obtained from 3.5 mg/ml MpARF3. The number of dummy residues was set to 480 and no symmetry or particle shape was imposed for the analysis. A MpARF3 structural model was generated using SWISSModel and the apo model and DNA-bound monomer (based on the MpARF2-HA7 X-ray structure) were superposed onto the low resolution shapes with supcomb.[59]

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of the experiments can be found in the method details section and in figure legends.

X-ray crystallography data collection and refinement statistics are summarized in Tables S4 and S5. SAXS data collection statistics are summarized in Table S3.

The SEC analysis of the DNA samples in absence of protein were repeated at least 2 times and the standard deviation in V_{el} shown in Figure 1B was calculated. For SEC analysis, a positive interaction (Figure 1C) was considered when the elution volume of the sample was lower than 1.964 mL, which corresponds to the calibrated elution volume of an ARF monomer + dsHA5.

The statistical significance of the differences in the helix 1 length between ARF classes (Figures 4A and S4) was obtained from students-t two-tailed test for two-sample unequal variance. A p-value < 0.001 was labeled with ***.

For SEC-MALS, data collection and analysis were performed using ASTRA 7 software (Wyatt Corp.). ASTRA calculates uncertainties for all reported parameters by analyzing baseline data at the beginning and end of the chromatogram. This analysis determines the statistical fluctuations in each detector's output, including all photodiodes and AUX signals. Each detector is assigned a weighting factor based on the noise observed in the first and last 10% of data points, up to a maximum of 100 points. The less noisy end is used to calculate this weighting factor. Then, the analysis involves fitting an nth-order polynomial to the data, and the error bar calculation incorporates the previously calculated weighting factor, the normalized RΘ value, a concentration uncertainty factor, and the Chi-squared value from the fit. These components are combined using standard error propagation rules to calculate a standard deviation. All measurements processed are weighted by the standard deviations of the measured quantities, which are then used to derive the expected standard deviations of all calculated parameters.

Data from FA was analyzed and plotted with OriginPro 2019 (OriginLab Corporation). Data was fitted to the Michaelis-Menten and Hill models, using as iteration algorithm the Orthogonal distance regression. An instrumental error weight ($w_i=1/\sigma_i^2$) considering the standard deviation of each anisotropy data point was applied to the fitting calculations. This was included in the fitting to factor more heavily values with reduced standard error. Standard error of K_D was scaled with square root of reduced chi square. The selection of the Michaelis-Menten or Hill model was done according to the model that yielded the lowest coefficient of variation for K_D ($CV = SE_{(KD)}/ K_D$).