The Intervening Domain Is Required For DNA-binding and Functional Identity of Plant MADS Transcription Factors

Xuelei Lai
CNRS
Rosario Vega-Leon
Humboldt-Universität zu Berlin
Veronique Hugouvieux
Commissariat à l´Energie Atomique et aux Energies Alternatives
Romain Blanc-Mathieu
CEA Grenoble https://orcid.org/0000-0002-9485-6330
Froukje van der Wal
Wageningen University and Research
Jeremy Lucas
CEA
Catarina Silva
EMBL
Agnes Jourdain
CEA
Jose Muino
Humboldt-Universität zu Berlin
Max Nanao
European Synchrotron Radiation Facility
Richard Immink
Wageningen University & Research https://orcid.org/0000-0002-0182-4138
Kerstin Kaufmann
Humboldt University of Berlin https://orcid.org/0000-0001-7960-6256
Francois Parcy
Laboratoire Physiologie Cellulaire et Végétale, Univ. Grenoble Alpes, CNRS, CEA, INRAE,
https://orcid.org/0000-0003-2191-500X
Cezary Smaczniak
Humboldt University of Berlin https://orcid.org/0000-0002-4663-8275
Chloe Zubieta (chloe.zubieta@cea.fr)
CEA https://orcid.org/0000-0003-4558-9333

Article

Keywords: MADS transcription factors (TF), proteins, physiological roles, plant physiology

Posted Date: March 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-288619/v1

Version of Record: A version of this preprint was published at Nature Communications on August 6th, 2021. See the published version at https://doi.org/10.1038/s41467-021-24978-w.

The Intervening Domain Is Required For DNA-binding and Functional Identity of Plant MADS Transcription Factors

Xuelei Lai[1,*], Rosario Vega-Leon[2,*], Veronique Hugouvieux[1,*,#], Romain Blanc-Mathieu[1], Froukje van der Wal[3], Jérémy Lucas[1], Catarina S. Silva[1,4], Agnès Jourdain[1], Jose Muino[5], Max H. Nanao[6], Richard Immink[3,7], Kerstin Kaufmann[2], François Parcy[1], Cezary Smaczniak[2,#], Chloe Zubieta[1]

[1] Laboratoire Physiologie Cellulaire et Végétale, Univ. Grenoble Alpes, CNRS, CEA, INRAE, IRIG-DBSCI-LPCV, 17 avenue des martyrs, F-38054, Grenoble, France.
[2] Plant Cell and Molecular Biology, Humboldt-Universität zu Berlin, Institute of Biology, Berlin, Germany.
[3] Bioscience, Wageningen Plant Research, Wageningen University and Research, 6708 PB, Wageningen, The Netherlands.
[4] European Molecular Biology Laboratory, Grenoble, France
[5] Systems Biology of Gene Regulation, Humboldt-Universität zu Berlin, Institute of Biology, Berlin, Germany.
[6] European Synchrotron Radiation Facility, Structural Biology Group, Grenoble, France
[7] Laboratory of Molecular Biology, Wageningen University and Research, 6708 PB, Wageningen, The Netherlands.

*first
#correspondence

Abstract

The MADS transcription factors (TF) are an ancient protein family with a high degree of sequence identity that bind almost identical DNA sequences across all eukaryotic kingdoms of life, yet fulfill dramatically different physiological roles. In plants, the family is divided into two main lineages, type I and II, based on sequence conservation of the DNA-binding MADS-box domain (M domain) with yeast and animal M domains. Here, we demonstrate that DNA binding in both lineages absolutely requires a short amino acid sequence C-terminal to the M domain called the Intervening domain (I domain) in type II MADS. Structural elucidation of the MI domains from the floral regulator, SEPALLATA3 (SEP3), shows a highly conserved MADS-box fold with the I domain forming an alpha helix and acting to stabilize the M domain. Based on secondary structure prediction, sequences fulfilling the same function as the SEP3 I domain can be found in both lineages of plant MADS TFs, suggesting the I domain is a conserved and required part of the DNA-binding domain. Using the floral organ identity MADS TFs, SEP3, APETALA1 (AP1) and AGAMOUS (AG), domain swapping demonstrate that the I domain alters DNA-binding specificity based on seq-DAP-seq experiments. Yeast 2-hybrid experiments further revealed the role of the I domain in dimerization specificity. Surprisingly, introducing AG carrying the I domain of AP1 in the Arabidopsis *ap1* mutant, resulted in a high degree of complementation and restoration of first and second whorl organs. Taken together, these data demonstrate that the I domain acts both as an integral part of the DNA-binding domain and strongly contributes to the functional identity of the MADS TF.

Introduction

The MADS-box genes, named after founding members *MINICHROMOSOME MAINTENANCE1* (*MCM1*, *Saccharomyces cerevisiae*), *AGAMOUS* (*Arabidopsis thaliana*),

DEFICIENS (*Antirrhinum majus*), and *SERUM RESPONSE FACTOR* (*SRF*, *Homo sapiens*), is an ancient gene family present before the ancestral split of animals, plants and fungi[1-4]. Prior to the divergence of eukaryotes into different kingdoms of life, the MADS gene family underwent a duplication event giving rise to two main lineages, the SRF-like and MEF2-like lineages, which correspond to the type I and II MADS genes in plants, respectively [6,7]. While these two MADS lineages both encode a highly conserved ~60 amino acid MADS-box DNA-binding domain (DBD) that recognizes a CArG box motif (CC-"Adenine rich"-GG), DNA-binding preferences are slightly altered due to amino acid changes in the MADS-box domain specific to each lineage[8-11]. In addition to changes in amino acids important for direct base readout, the region C-terminally adjacent to the MADS-box domain varies based on sequence alignments and available structural data for mammalian and fungal SRF-like and MEF2-like MADS TFs. These carboxyl terminal sequences have been shown to be important for dimerization and DNA binding although they do not directly contact the DNA[4,8,12]. While the MADS genes are ubiquitous in extant eukaryotes, they have undergone a significant expansion in plants, with angiosperms possessing tens of type I SRF-like and type II MEF2-like MADS genes that fulfill diverse physiological roles.

The plant type I MADS genes generally consist of one or two exons and encode transcription factors (TFs) comprised of the MADS-box DBD and a variable C-terminal domain. The type I genes are further subdivided into three subfamilies, Mα, Mβ and Mγ, with preferential interactions between subfamilies forming heterodimeric complexes [13,14]. While generally expressed at very low levels in a tissue specific manner, the type I genes have been shown to be important in plant reproduction and speciation, with crucial roles in female gametophyte, embryo, and endosperm development [15,16]. In contrast to the simple one or two exon composition and protein structure of the type I MADS, the type II genes consist of 5-8 exons and encode TFs with a modular four-domain structure, called "MIKC"[17-19]. The MIKC domains refer to M, for MADS DBD, I for the dimerization specifying Intervening domain, K for the coiled-coil keratin-like domain involved in dimer and tetramer formation and C for the variable C-terminal domain important for transactivation and higher order complex formation. The type II MADS are further subdivided into the MIKC* and $MIKC^C$ subfamilies based on the structure of the I domain, with dimerization mainly occurring within the same subfamily, but not between subfamilies [20-22]. While all MADS TFs are believed to form dimeric complexes in order to bind DNA, the type II MADS TFs are able to tetramerise via the K domain, greatly expanding the number of possible MADS complexes that can be formed [23-25]. The diversity of heteromeric complexes the type II MADS form is hypothesized to be directly

82 tied to their functional diversity [26]. The type II MADS play well established roles in many
83 developmental processes including meristem identity, flowering time, fruit and seed
84 development and floral organ identity [27].
85 The MADS TFs involved in floral organ specification are arguably the most well
86 studied MADS family members and provide an ideal model to examine function *in planta*.
87 Elegant *in vitro* and *in vivo* experiments dating back decades have tried to address the
88 question of how plant MADS TF achieve their DNA-binding specificity and functional
89 diversity in the context of flower and floral organ development[28,29]. Four classes of MIKC
90 MADS TFs, class A, B, C and E are necessary for floral organ identity. The A+E class
91 specify sepals, the A+B+E petals, A+C stamen and C+E carpels. In Arabidopsis, these
92 correspond to APETALA1 (AP1, A class), APETALA3 (AP3, B class) and PISTILLATA
93 (PI, B class), AGAMOUS (AG, C class) and the four SEPALLATAs (SEP1 -4, E class). The
94 generation of chimeric MADS TFs with swapped M domains from AP1, AP3, PI and AG
95 demonstrated that the MADS domains of these proteins were interchangeable with respect to
96 their *in vivo* function[29,30]. Even swapping in portions of human or yeast M domains, which did
97 alter *in vitro* DNA-binding, did not alter the function of the MADS proteins *in vivo*,
98 suggesting that the physiological function of these proteins seems to be independent of the
99 DNA-binding domain[30]. Thus, tetramerisation in MIKC MADS TFs came under scrutiny as a
100 potential determinant of DNA-binding syntax by selecting for two-site DNA-binding
101 constrained by specific intersite distances[31,32]. More recent studies of a tetramerisation mutant
102 of SEP3 revealed that while tetramerisation contributed to *in vivo* function, *in vitro* changes in
103 DNA-binding specificity genome-wide were more limited, putatively affecting a relatively
104 small number of key gene targets whose regulation depends on specific spacings between
105 CArG box binding sites[32,33]. Thus, the fundamental question of how plant MADS TFs are able
106 to recognize different DNA sequences critical for the regulation of different target genes, with
107 apparently interchangeable DBDs, is still not fully resolved.
108 In order to address this, we performed structural, biochemical, genome-wide binding
109 and *in vivo* studies focusing on the role of the Intervening domain. We demonstrate that an I-
110 like domain is present in both type I and type II plant MADS TFs and that this region is
111 required for DNA-binding, affects DNA-binding specificity and alters dimerization specificity
112 *in vitro* and in yeast assays. Seq-DAP-seq experiments comparing SEP3-AG and SEP3-
113 AG^{IAP1}, a chimera with the I domain from AG replaced by the I domain of AP1, revealed
114 unique binding sites for each complex, changes in preferred site spacing and a loss of carpel
115 and fourth whorl-specific targets for the chimeric protein. Using these same floral organ

identity MADS TFs to probe *in vivo* function of the I domain, different chimeric constructs were introduced into the *ap1* background. Interestingly, replacing the I domain of *AG* with that of *AP1* was sufficient to confer the majority of *AP1* functions *in planta* including sepal and petal identity in the first and second whorls. Taken together, these data illustrate the importance and multiple roles of the I domain in DNA-binding and *in planta* function.

Results

Structure of the SEPALLATA3 MI domain

The MADS-box and I domain of SEP3 ($SEP3^{MI}$, residues 1-90) was used in structural studies as this construct has been shown to be dimeric in solution and to bind DNA in a sequence specific manner [34]. $SEP3^{MI}$ crystallized in spacegroup $C222_1$ with 4 monomers per asymmetric unit and diffraction to 2.1 Å (**Table 1**). The first 18 amino acids, which are predicted to have no defined secondary structure, were disordered as were the 17 C-terminal amino acids, and no electron density was interpretable for these portions of the protein in any of the molecules. Each monomer adopts the same tertiary structure with an N-terminal alpha helix (α1; aa 18-40) and two antiparallel beta strands (β1; aa 44-48 and β2; aa 55-58). These elements make up the core MADS DBD and are conserved in the MEF2 and SRF-like MADS TFs (**Figure 1A and B**). C-terminal to the beta strands is a short loop followed by an alpha helix which is contributed by the I domain (α2; aa 63-73). The orientation of the C-terminal alpha helix differs between MEF2 (PDB codes 1EGW, 1C7U, 1TQE, 3KOV) and SRF-like MADS TFs (PDB code 1HBX), with $SEP3^{MI}$ resembling the MEF2 structure. Overall, $SEP3^{MI}$ adopts the MADS/MEF2 fold, as predicted for type II plant MADS TFs.

Each $SEP3^{MI}$ monomer dimerizes via extensive contacts between the pairs of alpha helices and beta strands from each partner, with dimerization required for DNA binding. In the structure, all dimers were formed by crystallographic symmetry, with dimerization burying over 25% of the total surface area of the molecule (**Figure 1**) [35]. The dimer interface includes hydrophobic interactions between residues in the N-terminal DNA-binding alpha helices, Leu28 and Leu35, and pairs of salt bridges between Glu34 and Arg24 (**Figure 1D**). The formation of a 4-stranded antiparallel beta sheet further stabilizes the quaternary structure of SEP3 with interactions bridging the N-terminal DNA binding helices and the C-terminal I domain helices. The I domain helices sandwich the beta sheet and lie perpendicular to the DNA-binding helices of the MADS-box domain. The I domain is anchored to the beta sheet via hydrogen bonding interactions between Glu56 and Tyr70 and pi-cation interactions between Phe48 and Arg69 of the partner monomers. Each I domain helix also interacts with

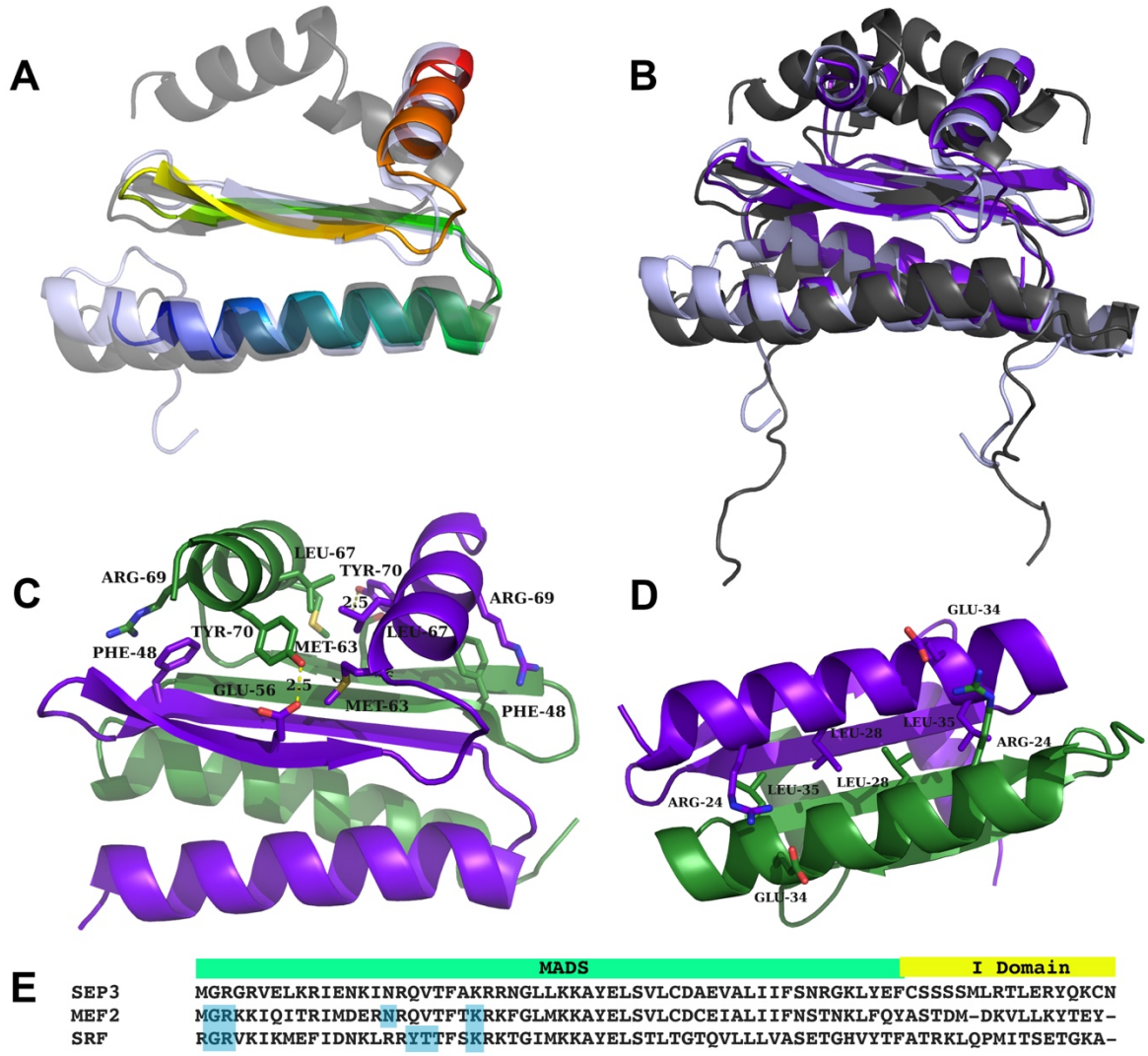

Figure 1. Structure and sequence of the SEP3 MI domain and corresponding SRF and MEF2 regions. (A) Overlay of SEP3 dimer (rainbow) and SRF (1HBX) dimer (gray). The I region is alpha helical for both structures, however the contacts with the M domain differ. (B) Overlay of SEP3 (dark purple) and MEF2A (3KOV, light purple) and SRF (gray) demonstrating the different conformations of the I region in the type II (MEF2-like) and I (SRF-like) MADS TFs. (C) SEP3 dimer with one monomer in purple and one in green. Amino acids important for interactions between I domains and between M and I domains are labelled and drawn as sticks. Hydrogen bonds are shown as dashed yellow lines. (D) View of the I domain with intermolecular interactions from the N-terminal alpha helices shown. (E) Partial sequence alignment of SEP3 (accession NP564214.2), MEF2 (accession AAB25838.1) and SRF (accession P11831) corresponding to the region in the crystal structure. Residues directly contacting the DNA for MEF2 and SRF are highlighted in blue.

its dimer partner helix via hydrophobic interactions mediated by Leu67, Tyr70 and Met63 (**Figure 1C**). These residues are highly conserved in all MIKC MADS TFs from Arabidopsis, suggesting that the interactions are also conserved and likely important for structural stability of plant MADS TFs. Mutations R69L (present in APETALA3), R69P (present in FLOWERING LOCUS M-delta) and Y70E (present in FLM-delta) destabilized SEP3^{MI} based on thermal shift assays, as predicted from the structural data (**Supplementary Figure 1**).

SEP3^{MI} was crystallised without DNA, however, multiple DNA-bound structures of the closely related MEF2 protein are available for comparison. Structural alignment of SEP3^{MI} with the corresponding MEF2 structure reveals a rmsd of 0.942 Å^2, underscoring the

conservation in the MADS/MEF2 domains with the MADS/I domains of $SEP3^{MI}$ (**Figure 1A, B**). The residues directly contacting the DNA are contributed from the flexible N-terminus and the N-terminal helix of the MADS-box domain, with no direct contacts from the MEF2/I domain (**Figure 1E**). These DNA-binding residues are highly conserved in SEP3 even though MEF2 recognizes a $YTA(A/T)_4TAR$ sequence (Y=pyrimidine, R=purine), whereas SEP3 recognizes a classic CArG box motif ($CC(A/T)_6GG$). The differences in DNA-binding specificity of the MEF2-like MADS TFs coupled with the high conservation of sequence and structure of the MADS-box domain itself, suggest allosteric contributions from the MEF2/I domain may play a role in DNA binding. Indeed, examining the DNA-binding of the I domain mutants R69L, R69P and Y70E demonstrated that these single point mutations were sufficient to abrogate the DNA binding activity of the full-length SEP3 based on electrophoretic mobility shift assays (EMSAs) (**Supplementary Figure 1**).

Type I and II MADS possess I-like domain sequences required for DNA-binding

Structural and mutagenesis data suggest that the MEF2/I domain is critical for stability and DNA-binding of type II MADS TFs. The type I plant MADS TFs do not possess an I or MEF2-domain and are more closely related to the SRF-type MADS TFs. However, the structure of SRF (PDB 1HBX) reveals a similar alpha helical region C-terminal to the MADS-box domain, albeit with a different orientation of helices from the MEF2/I domain (**Figure 1A**). Protein sequence alignment of the 106 MADS TFs from Arabidopsis reveal that 104 of the 106 sequences possess ~15 amino acids C-terminal to the MADS domain that are structurally conserved and predicted to form alpha helices with similar physicochemical properties (**Figure 2A**). In particular, positions including Arg69 and Tyr70 in SEP3 that are important for I domain anchoring to the M domain are conserved in both type I and type II MADS, while a high degree of sequence variation in the non-alpha helical region is observed (**Figure 2A and Supplementary Figure 2**). This suggests that these ~15 amino acids in type I MADS TFs may act as an I-like alpha helical domain and stabilize the structure of the DBD, influence DNA-binding specificity or both.

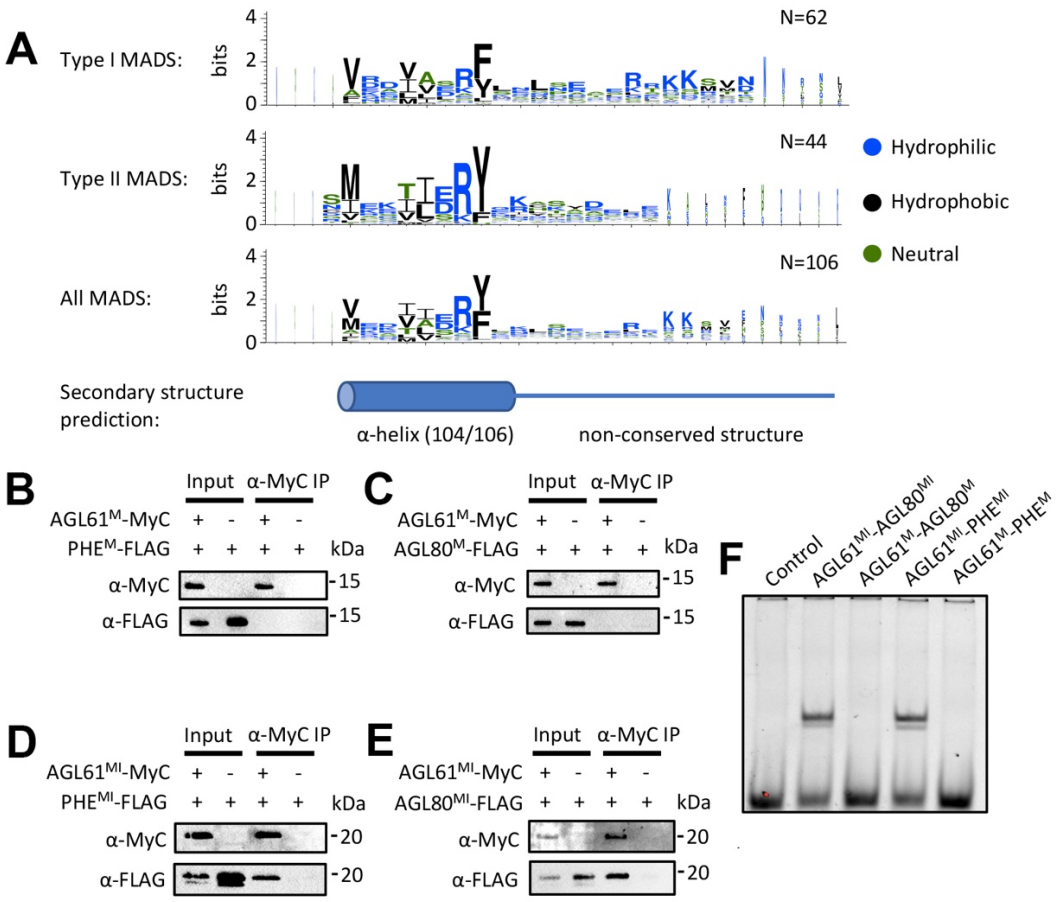

Figure 2. Type I MADS possesses I domain like region which is required for both dimerization and DNA binding. (A) Amino acid enrichment of the I region (~30 amino acids C-terminal to the M domain) of type I, type II and all MADS TFs, logos generated with WebLogo[36]. The overall height of the stack in each position indicates the sequence information content at that position, while the height of the amino acid symbols within the stack indicates the relative frequency at each position. The MADS TF sequences are taken from The Arabidopsis Information Resource (www.arabidopsis.org). **(B-C)** Pulldown assay showing that M domain of AGL61 ($AGL61^{M}$) does not interact with the M domain of PHE (PHE^{M}) or AGL80 ($AGL80^{M}$). **(D-E)** Pulldown assay showing that the M domain plus the I like region of AGL61 ($AGL61^{MI}$) interacts with the MI region of PHE (PHE^{MI}) and AGL80 ($AGL80^{MI}$). **(F)** EMSA assay showing that heterodimers $AGL61^{MI}$-$AGL80^{MI}$ and $AGL61^{MI}$-PHE^{MI} shift a DNA sequence containing a canonical CArG-box binding site from the *SEP3* promoter, while their corresponding constructs without the I region do not exhibit any binding.

To further determine the role of the alpha helical region adjacent to the M domain in both type I and II MADS (called for simplicity the I region hereafter), we performed pull down experiments and EMSAs for representative MADS TFs from Arabidopsis. We chose AGAMOUS-LIKE 61 (AGL61) from subclass Mα and AGAMOUS-LIKE 80 (AGL80) and PHERES1 (PHE1) from subclass Mγ for type I MADS TFs, as interaction patterns of these have been previously studied [37,38]. Pull-down experiments revealed that the M+I region of AGL61 ($AGL61^{MI}$) interacts with the M+I region of AGL80 ($AGL80^{MI}$) or PHE1 ($PHE1^{MI}$), while the M domains alone are not sufficient for interaction (**Figure 2B-E**), suggesting that the I region in type I MADS TFs is required for stable dimerization. Consistent with this,

EMSAs show that while heterodimers of AGL61^{MI}-AGL80^{MI} and AGL61^{MI}-PHE1^{MI} bind to a canonical CArG box motif, the M domains alone exhibit no DNA binding (**Figure 2F**).

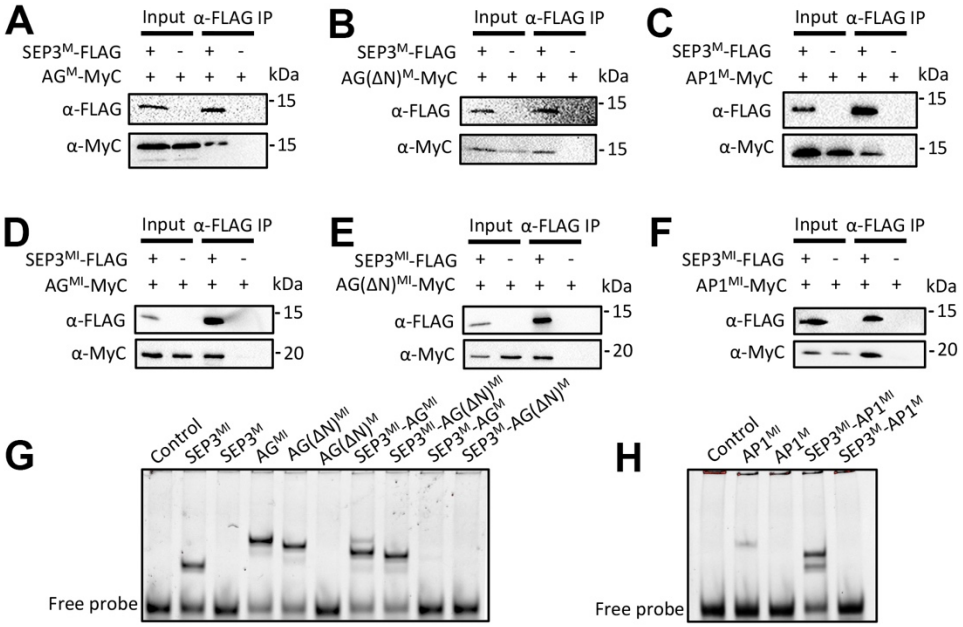

Figure 3. Type II MADS TFs require the I domain for DNA binding but not dimerization. (A-C) Pulldown assay showing that the M domain of SEP3 (SEP3^{M}) interacts with the M domains of AG (AG^{M}), AG with the first 16 N-terminal amino acids deleted (AG(ΔN)^{M}) and AP1 (AP1^{M}), respectively. (D-F) Pulldown assay showing that the MI domain of SEP3 (SEP3^{MI}) interacts with the MI domain of AG^{MI}, AG(ΔN)^{MI} and AP1^{MI}, respectively. (G and H) EMSA assay showing that homodimers from SEP3^{MI}, AG^{MI}, AG(ΔN)^{MI} and AP1^{MI}, and heterodimers SEP3^{MI}-AG^{MI}, SEP3^{MI}-AG(ΔN)^{MI} and SEP3^{MI}-AP1^{MI} shift a DNA sequence containing a canonical CArG-box binding site (as per Figure 2), while their corresponding constructs without the I domain can not, suggesting that the I domain in type II MADS TFs is required for DNA binding.

For the type II MADS TFs, we generated M domain and MI domain constructs for SEP3, AP1 and AG. In Arabidopsis, SEP3-AP1 containing MADS TF complexes define first and second whorl flower organs (sepal and petals) whereas SEP3-AG containing complexes are required for third and fourth whorl (stamen and carpel) flower organ identity. We focused on these MADS TFs as their heterocomplex formation, DNA-binding specificity and *in vivo* activity are well understood [11,14,23,39-43]. As expected, the MI domain of SEP3 interacts with that of AP1 and AG (**Figure 3D-F**). However, in contrast to the type I MADS, the M domain of SEP3 is able to pull down the M domains of AP1 and AG (**Figure 3A and C**). As AG possesses an N-terminal 16 amino acid extension, constructs with and without these amino acids were tested. Both constructs were able to interact with the M domain of SEP3 (**Figure 3A and B**). Thus, unlike the type I MADS TFs tested, the M domain alone of the type II MADS TFs tested was sufficient for protein-protein interaction. However, the M domains alone of SEP3-AP1 and SEP3-AG were unable to bind a canonical CArG DNA motif in EMSA experiments (**Figure 3G and H**), suggesting that the M domain of type II without the I domain is not sufficient to form a stable dimer for DNA binding. In contrast, homo and

heterodimers of the MI domains of SEP3, SEP3-AP1, AP1, AG and SEP3-AG were all able to bind DNA under our experimental conditions (**Figure 3G and H**). Taken together, these data suggest that the I domain, while possessing no direct interactions with DNA, is essential for DNA-binding for both type I and type II MADS TFs, likely by stabilizing the MADS TF dimer.

Impact of the I domain on DNA-binding specificity and intersite spacing

To further explore the role of the I domain in DNA-binding specificity, we focused on the floral organ identity-specifying MADS TFs, SEP3, AP1, AG and a chimeric version of AG, in which the AG I-domain was replaced with the AP1 I-domain (AG^{IAP1}). Using sequential DNA-affinity purification followed by sequencing (seq-DAP-seq ref), we determined the DNA-binding of the heteromeric complexes of SEP3-AG[32] and SEP3-AG^{IAP1} (this study). Unfortunately, the SEP3-AP1 heteromeric complex under the same experimental conditions did not yield data of sufficient quality for analysis. However, SEP3-AG and SEP3-AG^{IAP1} strongly bound DNA and each produced highly similar replicates for data analysis (**Supplementary Fig. 3)**. We used a highly stringent peak calling for each replicate and peaks were kept only when present in all replicates. This yielded 6,347 peaks for SEP3-AG and 3,552 peaks for SEP3-AG^{IAP1}. These peak lists were merged and the binding intensities for the two complexes were compared for all resulting peaks. This profiling indicates that while most peaks have a similar binding intensity, several hundred exhibited at least a two-fold difference in binding intensity between SEP3-AG and SEP3-AG^{IAP1} (**Figure 4A-B**). The same profiling between replicas for a given dataset demonstrated no such variability in binding intensity (**Supplementary Figure 3)**.

The top 15% sequences (1,073 sequences for each complex) displaying the most extreme binding difference were used to search for specific DNA patterns that potentially code for SEP3-AG and SEP3-AG^{IAP1} differential response. The TF binding sites were modeled using position weight matrices, transcription factor flexible models and k-mer set memory (KSM) analysis. Due to the highly similar binding modes of the MADS family TFs, only KSM was sensitive enough to differentiate the MADS heterocomplexes (**Figure 4C and D**)[44]. For the KSM, pipeline, KMAC and KSM tools from the GEM package [45] were used to search for clusters of short (4-20 bp) overlapping sequences (with allowances and penalties for gaps and mismatches) that are over-represented in the top 15% of SEP3-AG and SEP3-AG^{IAP1} specific bound sequences. This analysis yielded 55 clusters for SEP3-AG and 22 clusters for SEP3-AG^{IAP1} and served as kmer-based models for prediction analysis. With

AUC of 0.79 for SEP3-AG and 0.84 for SEP3-AG^{IAP1}, the kmer-based models were highly predictive. Importantly, each kmer-based model did not perform well (AUC ~0.7) in predicting bound regions from the other heterocomplex (**Figure 4D**). This suggests that specific sequence patterns code for differential binding between the two heterocomplexes.

Previously, using seq-DAP-seq, we showed that SEP3-AG demonstrates preferential intersite spacing of ~47 and ~57 bp between two CArG boxes due to its cooperative binding, a contributing factor for DNA-binding specificity of MADS TF complexes[32]. Using specific bound regions of SEP3-AG^{IAP1} when compared with SEP3-AG, we found that SEP3-AG^{IAP1} gained new preferential intersite spacings of 25 and 34 bp and lost the SEP3-AG preferential intersite spacings (**Figure 4E**). This further suggests that the I domain plays a role in DNA-binding specificity by affecting the conformation of the tetramer and selecting for specific intersite distances.

In order to determine whether these observed *in vitro* differences in binding were potentially relevant *in vivo*, we examined published ChIP-seq datasets. Previous studies have shown that the *in vitro* binding of SEP3-AG relative to SEP3-AP1 in SELEX-seq correlates with the *in vivo* binding intensity of AG relative to AP1 in ChIP-seq for the 1500 most enriched SEP3 ChIP-seq peaks (**Figure 4F**) [44]. This result suggested that specific sequence patterns detected *in vitro* were able to at least weakly discriminate sequences either more likely to be bound by SEP3-AG or by SEP3-AP1 in the Arabidopsis genome. Applying the same analysis to our seq-DAP-seq data, we found a similar weak positive correlation between the SEP3-AG binding affinity relative to that of SEP3-AG^{IAP1} and the corresponding ChIP-seq derived *in vivo* binding, suggesting that SEP3-AG^{IAP1} behaves more similarly to SEP3-AP1 than to SEP3-AG (**Figure 4F**). Although modest, this correlation is supported by the observation that the genes corresponding to the 250 peaks most differentially bound by SEP3-AG relative to SEP3-AG^{IAP1} exhibited a 7.6-fold enrichment for genes involved in "carpel development" GO term (FDR = 4.2 x 10^{-3}). However, no clear GO term enrichment was detected for genes corresponding to the 250 regions best bound by SEP3-AG^{IAP1} relative to SEP3-AG. This may be due to the relatively small number of genes corresponding to GO terms for "petal development" (4 genes) and "sepal development" (13 genes), suggesting an incomplete list of genes for these GO terms. Taken together, the binding differences between the SEP3-AG^{IAP1} and SEP3-AG complex suggest that the AG^{IAP1} protein has lost some AG identity and gained AP1-like identity.

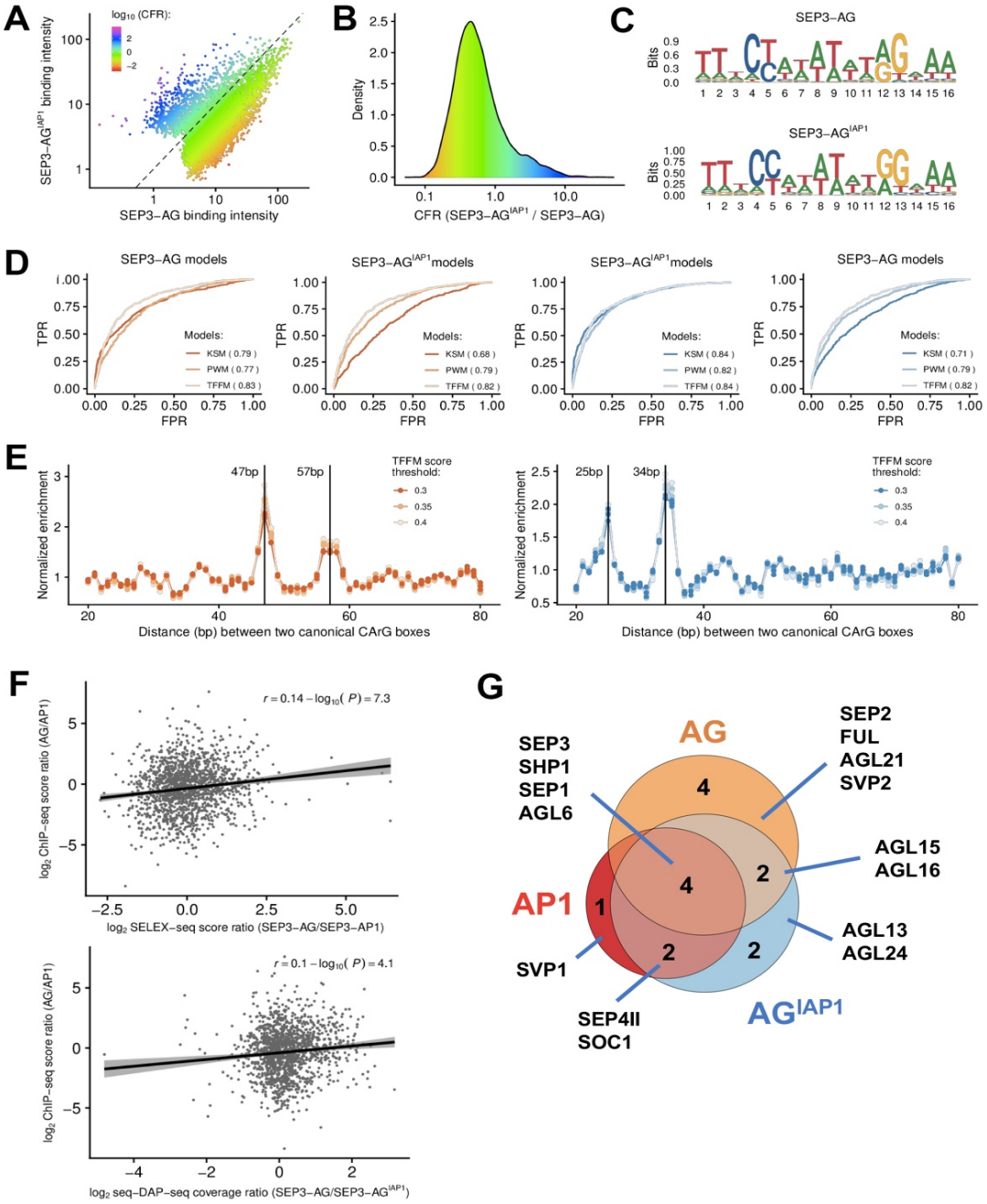

Figure 4. **DNA-binding and protein interactions patterns. (A)** Comparison of SEP3-AG and SEP3-AGI^{AP1} seq-DAP-seq binding intensity (log_{10} of reads per kb per million of reads mapped in bound regions) and color-coded by purple-blue (SEP3-AG^{IAP1}-specific) to orange-red (SEP3-AG-specific) according to log_{10} of SEP3-AG^{IAP1}/SEP3-AG. **(B)** Density plot showing data as per A. **(C)** Logos derived from PWM-based models obtained for SEP3-AG and SEP3-AG^{IAP1}. **(D)** Predictive power of TFBS models. Models are built using 600 sequences best bound by each of the two heterocomplexes and are searched against 1,073 SEP3-AG (orange) and 1,073 SEP3-AG^{IAP1} (blue) specific regions, defined as the top 15% of sequences that are most strongly bound by one complex relative to the other. Matrix-based models (PWM and TFFM) are not able to differentiate SEP3-AG and SEP3-AG^{IAP1} binding whereas k-mer-based analysis is able to better predict binding for the respective datasets. **(E)** SEP3-AG favors intersite spacings of 47 and 57 bp based on SEP3-AG specific regions. SEP3-AG^{IAP1} favors intersite spacings of 25 and 34 bp based on SEP3-AG^{IAP1I} specific regions. **(F)** Top, published SELEX-seq for SEP3-AP1 and SEP3-AG [5] comparing the normalized score ratios (SEP3-AG/SEP3-AP1) for SELEX-seq and score ratios (AG/AP1) ChIP-seq at 1,500 SEP3 best bound loci in ChIP-seq show a positive correlation, suggesting that SEP3-AP1 and SEP-AG bind different sequences *in vivo* and that *in vitro* binding is able to differentiate bound sequences that are more SEP3-AP1-like versus SEP3-AG-like. Bottom, SEP3-AG and SEP3-AG^{IAP1I} seq-DAP-seq coverage as per SELEX-seq scores. A positive correlation is observed suggesting that, *in vitro*, the swap of AP1 I domain in AG is able to recover some of the binding specificity of SEP3-AP1. **(G)** Yeast two-hybrid assays using AG, AP1 and AG^{IAP1} as bait against MIKC^{C} MADS TFs in Arabidopsis. Data shows that AG^{IAP1} loses AG interactors and gains AP1 interactors.

Impact of the I domain on protein interaction specificity

Next, we sought to understand to what extent the I domain could confer protein-protein interaction specificity as it plays an important role in dimerization and dimer stability[28]. AP1 and AG share SEP3 as a binding partner in the molecular model predicting flower organ specification, but do exhibit differential heterodimerisation capabilities with other MIKC^c MADS TFs[14,23]. In order to probe the function of the I domain in dimerization specificity, a matrix-based yeast two-hybrid screen for AP1, AG and AG^{IAP1} against all Arabidopsis type II MIKC^c MADS TFs was performed. The targets were cloned into both bait and prey vectors and only interactors for both replica and in all three selection media were scored as positive interactions (**Supplementary Figure 4**). This may underestimate the true number of binding partners, however, it reduces the number of false positives. Using this strict cut-off for protein-protein interaction events, AP1 interacted with seven different partners, of which three were specific for AP1 and four were in common with AG, including the SEP clade members SEP1 and SEP3. AG interacted with ten partners of which six were exclusive with respect to AP1 interacting partners. AG^{IAP1} interacted with ten partners, including two out of the three AP1-specific interactions and lost four AG-specific partners (**Figure 4G**). Interestingly, both SOC1 and AGL24, known AP1 binding partners[14] interacted with the AG^{IAP1} chimera but not with AG. These data indicate that the I domain of AP1 contributes to heterodimerization specificity of the chimera.

In planta function of the I domain

Ap1 mutants (strong *ap1-7* and intermediate *ap1-11*) exhibit a non-ambiguous flower phenotype including the homeotic conversion of the first whorl sepals into bract-like organs, as characterized by their leaf-like epidermal morphology and the formation of buds in their axils and the absence of second whorl organs (**Supplementary Figure 5**)[46,47]. To examine the role of the I domain in *in vivo* MADS function, complementation of the *ap1* loss-of-function mutants with different AG/AP1 chimeric constructs was examined. Primary transformants expressing a battery of chimeric domain swap constructs, AG^{MIKAP1}, AG^{MIAP1}, AG^{MAP1}, AG^{IAP1}, AG^{KCAP1} and $AP1^{IAG}$, under the control of the *AP1* promoter, revealed a spectrum of complementation for the first and second whorls in the intermediate *ap1-11* mutant background (**Figure 5**).

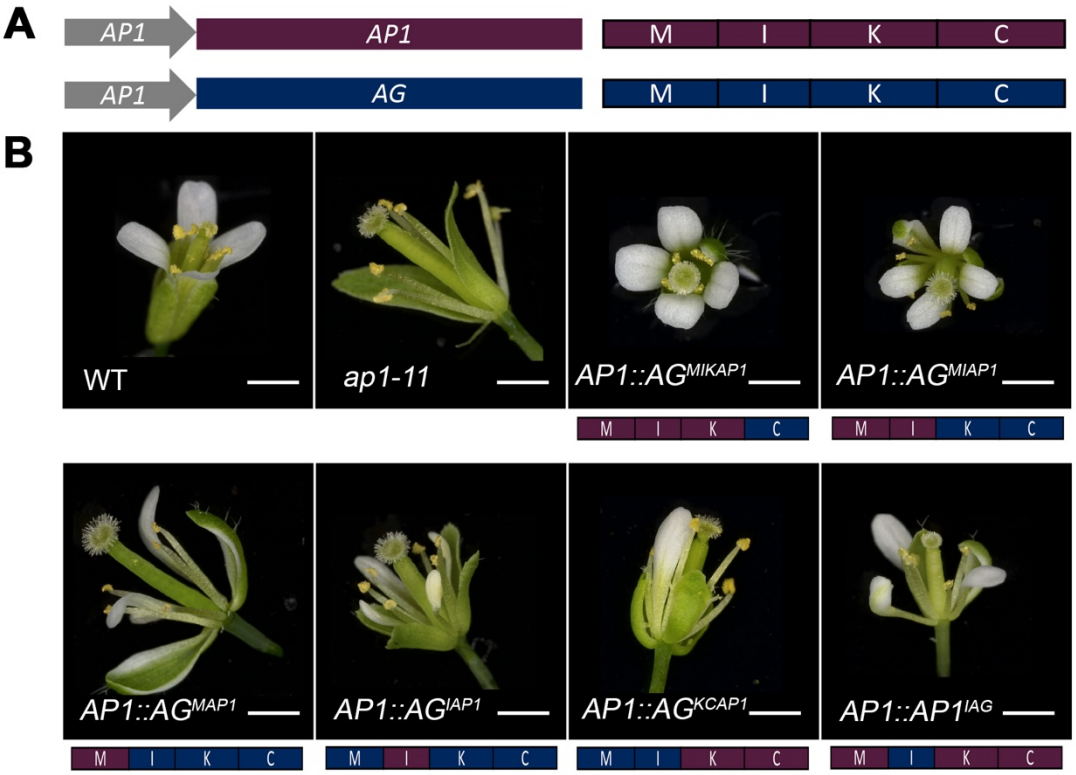

Figure 5. Primary transformants in the *ap1*mutant background exhibit a spectrum of complementation. (A) Schematic of the constructs and proteins produced with AP1 protein in purple and AG in dark blue. Domains are labeled MIKC. **(B)** Flower phenotypes of T1 transformants (n ≥ 10). All transformants were in the *ap1-11* background. Phenotypes for WT and *ap1-11* are shown. The domains corresponding to AP1 and AG are colored as per (A) and shown schematically below each panel.

AG^{MAP1} expressing plants were poorly complemented with a reduced number of first and second whorl organs and only partial organ identity recovery with petaloid and sepaloid organs in 43% of the plants. AG^{KCAP1} also complemented poorly, with most of the plants showing carpeloid sepals and with only partial recovery of 1 to 3 sepals in 60% of the plants. In contrast, almost complete complementation was present in plants expressing AG^{MIAP1}, with 92% of the transformants showing 4 sepals or petaloid sepals in the first whorl and 2 to 4 petals in the second whorl. This was similar to plants expressing AG^{MIKAP1}, in which first and second whorl organ identity was restored for 69% of the plants with occasional reduced number of organs in the second whorl and lateral flower formation. *AP1* plants with the I domain of AG ($AP1^{IAG}$), showed a similar phenotype as AG^{MAP1}. Strikingly, AG^{IAP1} expressing plants exhibited 4 sepaloid organs and 3-4 petals or petaloid organs in over 40% of the plants, suggesting that the I domain acts as the smallest unit conferring the most AP1 function to the AG^{IAP1} chimera (**Supplementary Table 2**).

To confirm the phenotype of the AG^{IAP1} expressing plants, three independent homozygous lines were generated and examined for *AP1*, *AG* and AG^{IAP1} under the *AP1*

promoter in the strong *ap1-7* mutant background (**Figure 6 and Supplementary Figures 6-8**). *AG*^{IAP1} *ap1-7* lines produced flowers very similar to WT (**Figure 6).**

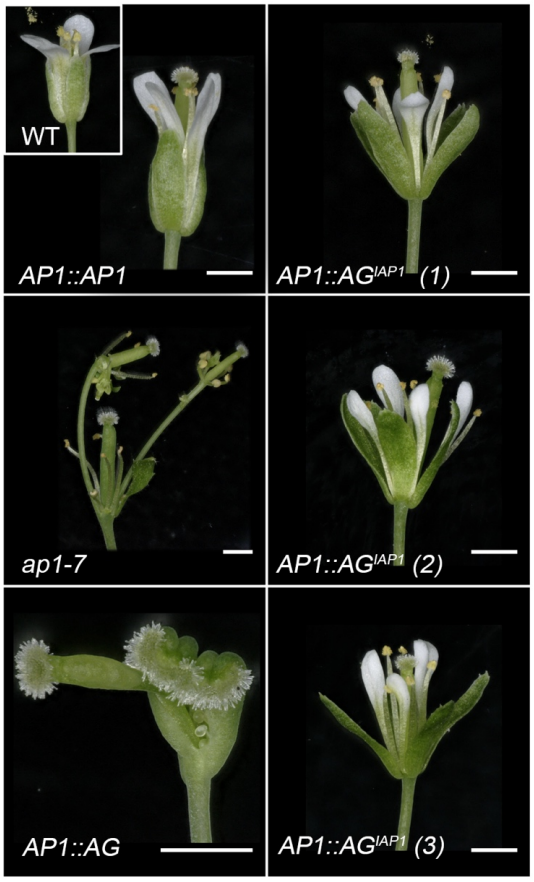

Figure 6. AP1:AG^{IAP1} expression largely complements the *ap1*-7 flower phenotype. *ap1*-7 lines expressing either *AP1* (*AP1::AP1*)), *AG* (*AP1::AG)* or *AG*^{IPA1} (*AP1::AG*^{IAP1}*)* under the control of the *AP1* promoter were grown at 22 °C in long days. A typical WT flower is shown as an insert in the upper left panel. Representative flowers for 3 independent *AP1::AP*^{IAP1} lines are presented and one representative flower from *AP1::AG* and *AP1::AP1* expressing lines. While the first whorl in *ap1-7* is replaced by bract or stipule with axillary buds and petals are missing in the second whorl, *AP1::AG*^{IAP1} expression restores WT first and second whorl organs, while *AG* expression triggers carpel development in the first whorl and absence of petals in the second whorl. Scales bars = 1 mm.

In these plants, whorl 1 of each flower was made of four sepal-like organs and axillary buds were never observed as opposed to the *ap1-7* mutant (**Figure 6 and 7**). In addition, between 66% to 82% of the flowers in three independent lines homozygous for the transgene showed 4 petal-like organs (**Figure 6 and 7**). Scanning electron microscopy (SEM) observations of epidermal cells from whorls 1 and 2 of *AG*^{IAP1} expressing lines showed elongated cells characteristic of sepals and conical cells characteristic of petals, respectively (**Figure 7 and Supplementary Figure 7**). Detailed inspection of sepal and petal surfaces by SEM from the *AG*^{IAP1} expressing plants revealed the presence of a few clusters of leaf-like and stamen identity cells in whorls 1 and 2, respectively (**Figure 7 and Supplementary Figure 8**). In these 3 lines, 10 to 25% of the flowers showed petals with only petal identity cells while 75 to 90% of the flowers show at least one petal with a small cluster of stamen identity cells (**Supplementary Figure 8**). This indicates incomplete complementation of *AP1* function by the chimera. However, *AP1:AG* was unable to even partially complement *AP1* function in the first and second whorl. Inspection of the cell surface in *ap1-7 AP1:AG* plants of the first whorl showed characteristic carpel cells (replum, style and papilla) and an absence

of organs in the second whorl (**Supplementary Figure 6**). AG^{IAP1} therefore possesses the ability to trigger *AP1* specific developmental programs and has lost much of its ability to trigger *AG*-specific organ development, indicating that the I domain of AP1 plays a major role in conferring AP1 functional identity to AG^{IAP1} *in vivo*.

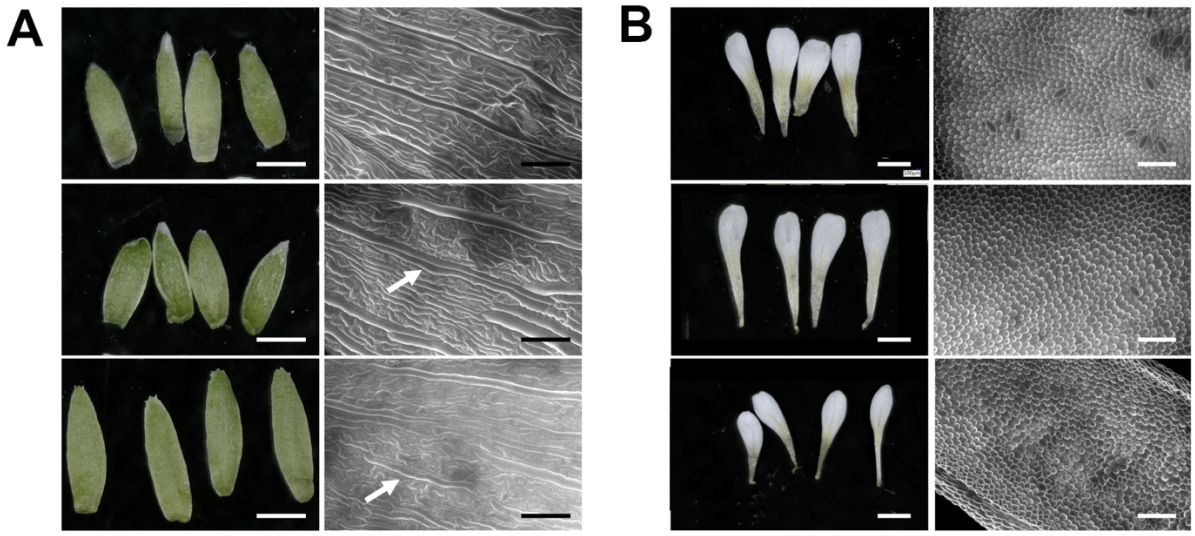

Figure 7. First and second whorls cell identity are complemented in *ap1-7* plants expressing AG^{IAP1}. First (A) and second (B) whorl organs were removed from WT, and flowers from *AP1* or AG^{IAP1} expressing *ap1*-7 plants. First whorl organs in AG^{IAP1} expressing plants are slightly longer compared to WT and *AP1* expressing plants (A, left panel). Second whorl organs in AG^{IAP1} expressing plants are slightly smaller compared to WT and *AP1* expressing plants (B, left panel). Epidermal cell identity, observed by SEM, show characteristic WT elongated sepal cells in *AP1* and AG^{IAP1} expressing plants (A, right panel) and characteristic WT conical cells in petals from *AP1* and AG^{IAP1} expressing plants (B, right panel). A small number of epidermal cells typical of leaves are also seen in AG^{IAP1} expressing plants and to a lesser extent in *AP1* expressing plants (arrow). Scale bars indicate 100 µM for SEM images and 1 mm for organ photographs.

Discussion

Transcription factor function depends on its selective binding of specific sequences in the genome. The MADS TF family has dramatically expanded in the green lineage over the course of evolution and fulfills many roles throughout the lifecycle of the plant. MADS constitute one of the largest plant TF families and are almost evenly divided into two types, I and II, represented structurally by SRF and MEF2. The MADS-box DBD, encoded by a single exon, is either SRF-like or MEF2-like and exhibits little sequence variation even amongst different eukaryotic kingdoms of life. How the MADS TF family is able to regulate diverse developmental processes and different target genes with such a highly conserved DBD has been a fundamental question leading to the speculation that regions distal to the DBD may play a role in DNA-binding specificity either via the formation of higher order complexes, the recruitment of ternary factors or via allosteric alterations in DNA-binding specificity of the MADS-box domain.

Domain swap experiments performed over 25 years ago suggested that the I domain

played a key role in DNA-binding specificity and dimerization in floral organ identity MADS TFs, although the mechanism of this was unclear[29,43,48]. Based on the *in vitro* and *in vivo* data presented here, the I region or I domain should be considered as a fundamental component of the DNA-binding module as it is essential for DNA-binding for both type I and II MADS TFs. Examination of the structure of the MI domains of SEP3 in comparison with MEF2 and SRF coupled with secondary structure predictions, reveals that the I region is always alpha helical and interacts extensively with the beta sheet of the M domain, albeit with different orientations. This suggests that allosteric effects of the I region in type I and II MADS TFs, due to variability in amino acid composition, length and secondary structure orientation, will influence the MADS-domain conformation and tune DNA-binding and specificity. In addition, the I region alters dimer partner specificity, altering the repertoire of heteromeric MADS complexes able to be formed and resulting in additional functional diversity *in vivo*.

With the introduction of the floral quartet model, the role of the I domain in determining MADS TF function specificity was supplanted by the importance of specific tetramer components triggering DNA-looping as being required for gene regulation[24]. Tetramerisation allows cooperative two-site binding and looping of DNA, thus selecting for DNA-binding sites that exhibit preferred spacing[31,34,49,50]. Examination of a tetramerisation mutant of SEP3 revealed that fourth whorl carpel development and meristem determinacy require efficient tetramerisation of SEP3[33]. Recent studies comparing the genome-wide binding patterns of tetrameric versus dimeric SEP3-AG complexes have further shown that tetramerisation both increases binding affinity and plays a role in determining DNA-binding specificity via preferential binding of specific intersite distances[32]. Comparisons of genome-wide binding of SEP3-AG and SEP3-AG^{IAP1}, presented here, reveals that the I domain alters preferred intersite spacing between CArG-type binding sites. These data suggest a role of the I domain as a modulator of tetramerisation and highlight an additional mechanism mediated by the I domain for changing DNA binding patterns of MADS TF complexes.

The recruitment of ternary factors, non-MADS family protein partners, has been postulated to be a major determinant of DNA-binding specificity[51]. Based on yeast 2-hybrid screening, mass spectrometry and pulldown assays, additional TF partners from the NF-Y and homeodomain families as well as the orphan TF, LEAFY, have been identified for a few type II MADS TFs including OsMADS18 from rice and SHATTERPROOF, SEEDSTICK, AG and SEP3 from Arabidopsis[25,48,52,53]. While ternary complex formation likely plays a role in differential gene regulation, the formation of such complexes has only been shown for a relatively small number of MADS TFs and how much it accounts for the diversity of function

in the MADS TF family requires further studies. As shown here, swapping the I domain of the well-studied floral organ identity MADS TF, AP1 (first and second whorl organs) into the third and fourth whorl specific TF, AG, results in strong complementation of the *ap1* loss of function phenotype by the chimeric construct. The experiments keep the K and C-terminal domains of AG intact and these domains have been postulated to be important for neo-functionalization[54] and/or the recruitment of ternary factors[55]. We cannot completely exclude that ternary factors recruited via the I domain could be responsible for the *AP1*-like activity of AG^{IAP1} in planta. For example, OsMADS18 uses the MI domains to recruit OsNF-YB1 based on yeast 2-hybrid assays [48]. However, it is more likely that the combination of tuning DNA-binding specificity and heterodimer partner selection by the I domain accounts for the majority of AP1 activity of the AG^{IAP1} chimera.

Based on the *in vitro* and *in vivo* experiments presented here, the I region is absolutely required for DNA-binding with the minimal DNA binding domain consisting of the M plus the I region, as opposed to the M domain alone. As shown here, the I domains of AP1 and AG also play an important role in DNA-binding and dimerization specificity. The I domain is able to tune the function of plant MADS TFs and, in the case of AP1, acts as a major determinant of functional identity in concert with the M domain. Further experiments will be required to determine how generalizable these results are to the wider MADS TF family (type I and II) and whether the I region broadly acts as a primary factor in MADS TF functional identity *in planta*.

Methods

SEP3 MI domain protein expression and purification

SEP3 MI (aa 1-90) was cloned into the pETM41 vector using the NcoI-NotI restriction sites to obtain His6-MBP translational protein fusion. Recombinant SEP3 MI protein was expressed in *E. coli* BL 21 Codon Plus cells. Cell were grown at 37 °C to an OD_{600} of 0.8-1 after which time the temperature was reduced to 15 °C and protein expression induced by addition of 1 mM of IPTG for 12 h. Cells were harvested by centrifugation and the cell pellet resuspended in 25 mM KH_2PO_4 pH 7 buffer containing 10 % glycerol, 500 mM NaCl, 2 mM Tris(2-carboxyethyl)phosphine (TCEP) and cOmplete protease inhibitors (Roche). Cells were lysed by sonication and cell debris pelleted at 25,000 rpm for 40 min. The soluble fraction was applied to a 1 ml Ni-NTA column and the protein eluted with resuspension buffer + 200 mM imidazole. The protein was dialysed against 25 mM Tris, pH 8.0, 300 mM NaCl and 1 mM TCEP. The protein was then applied to a heparin column to remove any bound DNA using a salt gradient from 300 mM to 2 M NaCl. The protein was dialysed overnight in the presence of TEV protease (10:1) to cleave the His-MBP tag at 4 °C. The protein was then passed over a Ni-NTA column to deplete the His-MBP and any uncleaved protein. A second heparin column was run to remove any His-MBP and to obtain pure SEP3 MI. The protein was concentrated to ~6 mg/ml and used for crystallization trials.

Protein crystallization, data collection and refinement

SEP3 MI at a concentration of 6 mg/ml was mixed at a 1:1 ratio with potassium sodium tartrate tetrahydrate (0.2 M), bis-tris propane (0.1 M, pH 7.5) and 20 % PEG 3350. The protein crystallized over 20 d at 4 °C forming a diamond shaped single crystal. 15 % glycerol was used as a cryoprotectant and the crystal flash frozen in liquid N_2. Diffraction data were collected at 100 K at the European Synchrotron Radiation Facility (ESRF), Grenoble France on ID29 at a wavelength of 1.0 Å. Indexing was performed using MXCube[56] and the default optimized oscillation range and collection parameters used for data collection. The data set was integrated and scaled using the programs *XDS* and *XSCALE* [57]. Data collection and refinement statistics are given in Table 1. The structure is deposited under PDB code 7NB0.

In vitro pull-down assays

The following constructs from type I and type II MADS TFs are cloned into pTnT vector for in vitro pulldown assays. Type I MADS TFs include AGL61^{M}(63-122)-5MyC, AGL61^{MI}(63-155)-5MyC, PHE^{M}(1-60)-3FLAG, PHE^{MI}(1-95)-3FLAG, AGL80^{M}(1-61)-3FLAG and AGL80^{MI}(1-88)-3FLAG, and type II MADS TFs include SEP3^{M}(1-57)-3FLAG, SEP3^{MI}(1-90)-3FLAG, AG^{M}(1-73)-5MyC, AG^{MI}(1-107)-5MyC, AG(ΔN)^{M}(17-73)-5MyC, AG(ΔN)^{MI}(17-107)-5MyC, AP1^{M}(1-57)-5MyC and AP1^{MI}(1-92)-5MyC. Proteins were produced using in vitro transcription and translation system using wheat germ extract following manufactures' instructions (SP6 High Yield Expression System from Promega). Briefly, purified plasmids were used as input in a 2 hr 25 °C incubation reaction. For single input reactions, 1 µg of plasmid was used in a 25 µl reaction volume. For double input reactions, 1 ug plasmid of each construct was used in a 25 µl reaction volume. 10 % of each input reaction (i.e. 2.5 µl) was used as input for western blots. The rest of the reaction was used for pull-down experiments. Briefly, each reaction was completed to 100 µl using PBS buffer (150 mM sodium phosphate, pH 7.2, and 150 mM NaCl), and added with 10 µl of the appropriate magnetic beads (anti-MyC beads, Thermo Fisher or anti-FLAG beads, Merck Millipore) and incubated for 1 hr at 4 °C. The beads plus protein solution suspension was then placed on magnets, and the supernatant was discarded. The beads were washed with PBS buffer four times, and SDS-PAGE loading dye was added to the beads and boiled for 5 min. at 95 °C. Western blots were used to assess the pulldown results using anti-MyC (Thermo Fisher) and anti-FLAG antibodies (Merck Millipore).

EMSA experiments

Proteins for EMSAs experiments were produced as described above. EMSAs were performed as described with 10 nM DNA labeled with Cy5 (Eurofins) using a 103-bp DNA fragment containing 2 CArG boxes belonging to the *SEP3* promoter[33]. For each EMSA, a negative control was run, labeled 'DNA alone' in which the in vitro translation assay was done with pTNT vector without any insert and incubated with the DNA probe.

Yeast two-hybrid screening

AP1, AG and the chimeric construct, AG^{IAP1}(residues 74-107 of AG replaced by residues 58-92 of AP1) were cloned into pENTR/D-TOPO® (Kan) gateway entry vector. Subsequently, Gateway LR reactions were performed using pDEST32 (pBDGAL4) and pDEST22 (pADGAL4) destination vectors. The resulting pDEST32-AG, pDEST32-AP1 or pDEST32-AG^{IAP1} expression construct was transformed into yeast strain PJ69-4α, and pDEST22-AG, pDEST22-AP1 or pDEST22-AG^{IAP1} into PJ69-4A, followed by an autoactivation screen for the bait vector as described[58]. Upon confirmation that the AG^{IAP1} bait did not possess any autoactivation capacity, a matrix-based Y2H screening was performed following the protocol described in [58]. Bait and prey clones of all native Arabidopsis type II MIKC^{c} MADS TFs were

generated previously [14] and were screened against AP1, AG and AG^{IAP1}. Growth of yeast and hence, protein-protein interaction events, were scored after seven days of incubation at 20 ˚C on selective medium. Three different selective media were used: SD lacking Leucine, Tryptophan, and Histidine (-LWH) and supplemented with 1 mM 3-Amino-1,2,4-triazole (3-AT); SD -LWH + 5 mM 3-AT and SD lacking L,W and Adenine (-LWA). Mating was performed twice in independent experiments and only combinations of MADS TFs giving growth on all three selection media and in both replica were scored as interaction events.

Seq-DAP-seq

Seq-DAP-seq for SEP3-AG^{IAP1}complex was performed as described previously[32]. Briefly, 2 µg of each purified plasmid (pTnT-SEP3-3FLAG and pTnT- AG^{IAP1}-5MyC) were used as input in a 50 µl TnT reaction incubated at 25 °C for 2 hr (Promega). The reaction solution was then combined with 50 µl IP buffer (PBS supplemented with 0.005 % NP40 and proteinase inhibitors (Roche)) and mixed with 20 µl anti-FLAG magnetic beads (Merck Millipore). Following 1 hr. incubation at room temperature, the anti-FLAG magnetic beads were immobilized, and washed three times with 100 µl IP buffer. Protein complexes were eluted with 100 µl IP buffer supplemented with 200 µg/ml 3xFLAG peptide (Merck Millipore). The eluted protein was then immobilized on anti-c-Myc magnetic beads (Thermo Fisher) and washed three times with 100 µl IP buffer to isolate homogeneous SEP3-AG^{IAP1}complex. The purified protein complex, while still bound on anti-c-Myc magnetic beads, was incubated with 50 ng DAP-seq input library pre-ligated with Illumina adaptor sequences. The reaction was incubated for 90 mins, and then washed 6 times using 100 µl IP buffer. The bound DNA was heated to 98 °C for 10 min and eluted in 30 µl EB buffer (10 mM Tris-Cl, pH 8.5). The eluted DNA fragments were PCR amplified using Illumina TruSeq primers for 20 cycles, and purified by AMPure XP beads (Beckman). The libraries were quantified by qPCR, pooled and sequenced on Illumina HiSeq (Genewiz) with specification of paired-end sequencing of 150 cycles. Each library obtained 10 to 20 million reads. The seq-DAP-seq was performed in triplicate.

Seq-DAP-seq data analysis

Reads processing and peak calling

SEP3-AG and SEP3-AG^{IAP1} reads processing and peak calling was performed as previously described[32]. Briefly, reads were checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adapter sequences removed with NGmerge[59] and mapped with bowtie2[60] onto the TAIR10 version of the *A. thaliana* genome (www.arabidopsis.org), devoid of the mitochondrial and the chloroplast genomes. The duplicated reads were removed using the samtools rmdup program [61]. The resulting alignment files for each sample were input to MACS2 [62] to call peaks using the input DNA as control. Consensus peaks between replicates were defined using MSPC [63] (P-value cutoff = 10^{-4}) for each experiment. Each consensus peak was scanned for possible subpeaks, split into several peaks if needed, and the peak widths were then re-sized to ± 200 bp at both side of the peak maximal height. For all the resulting peaks, coverage was computed as the mean of the normalized read coverage for each replicate. This normalized coverage defines the binding intensity of a heterocomplex at a bound region.

Comparison between SEP3–AG and SEP3-AG^{IAP1}

SEP3–AG or SEP3–AG^{IAP1} were merged according to the following procedure: peaks were considered common if at least 80 % of two peaks overlapped with < 50 % of either peak non-overlapping. Peaks that did not overlap by > 50 % of their length were considered new peaks. These values were chosen empirically based on visual inspection of the peaks in the

Integrated Genome Browser[64]. The averaged normalized coverage from each experiment, divided by the peak size, was computed for each peak. Figure 4 was computed using R (https://www.Rproject.org) and the ggplot library[65]. The coverage fold reduction (CFR) was computed as the ratio between the mean normalized coverages in SEP3–AG^{IAP1} and SEP3–AG seq-DAP-seq. The top 15 % sequences with extremes CFR were considered SEP3–AG^{IAP1} and SEP3–AG specific and used to search for differential DNA patterns that potentially direct transcription factor differential binding using sequence modelling (see next paragraph). Detection of preferred spacings between canonical CArG boxes was performed as in [32].

Position weight matrices (PWM), transcription factor flexible model (TFFM) and k-mer set memory (KSM).

For each experiment, PWM, TFFM and KSM models were reconstructed out of the 600 best peaks (judged according to their averaged coverage). PWM were generated by the meme-suite, using meme-chip[66] with options -meme-minw 16, -mememaxw 16, -meme-nmotifs 1 –meme-pal. TFFM were generated using the TFFM-framework package and the PWM obtained in the meme-chip output [67]. KSM is a recently developed TF binding motif representation that consist of a set of aligned k-mers that are overrepresented at TF binding sites. KSM were generated using the KMAC tool [45] with options set to search k-mers, with size ranging from 4 to 20bp in a 300bp sliding windows, that are enriched compare to TF unbound sequences. TF binding sites were predicted by searching the PWM, TFFM and KSM models against TF bound sequences and unbound sequences. PWM and TFFM scan was performed using in house scripts. KSM were searched using the KSM tool [45]. The best TFFM/PWM scores and the sum of KSM scores (because distinct KSM can hit to a same subject sequence) obtained for each bound regions and for unbound regions are retained to assess the prediction power of each model in an AUROC (Area Under the Receiver Operating Characteristic Curve). In this assessment, the unbound set of regions are chosen with similar GC content, size, and origin (promoter, intron, exon and intergenic) than the set of bound regions [68]. The set of unbound sequences fed to KMAC to detect enriched k-mer is different than the one used to evaluate the KSM model prediction.

Plant material and growth conditions

All experiments were performed using *A. thaliana* Col-0 accession. *Ap1*-7 allele was obtained from cal/cal al1-7 plants kindly provided by Justin Goodrich and the *ap1-11* (N6231) mutant was ordered from the Nottingham Arabidopsis Stock Centre. Seedlings were grown in controlled growth chambers in long day conditions (16 h light / 8 h dark) at 22 °C for plant transformation and phenotype analysis. WT, *ap1* mutant and complemented lines were always grown in parallel.

Plasmid constructs for plant transformation

Constructs used for the primary transformant phenotypic analysis were cloned into a modified version of pMDC32 binary vector[69], where the *2x35S* promoter was replaced with the *AP1* promoter. AG^{MAP1}, AG^{MIAP1}, AG^{MIKAP1}, AG^{KCAP1}, AG^{CAP1}, $AP1^{IAG}$ and AG^{IAP} coding sequences were amplified from pTNT-AG^{MAP1}, pTNT-AG^{MIAP1}, pTNT-AG^{MIKAP1}, pTNT-AG^{KCAP1}, pTNT-AG^{CAP1}, pTNT-$AP1^{IAG}$ and pTNT-AG^{IAP} vectors and subsequently inserted by XbaI restriction enzyme downstream of the *AP1* promoter. pTNT-CDS vectors were generated through Gibson assembly (NEB).

AP1::*AP1*, *AP1*::*AG* and *AP1*::AG^{IAP1} vectors were generated using Gibson assembly in the vector backbone pFP100[70]. *AP1* promoter and *OCT* terminator were PCR amplified from pML-BART-AP1pro:AP1-AR plasmid[71]. *AP1*, *AG* and AG^{IAP1} coding sequences were amplified from pTNT-*AP1*, pTNT-*AG*, pTNT- AG^{IAP1} vectors.

Plant transformation and floral phenotype analysis

All the plant transformations were performed using *ap1* mutant plants following the floral dip method[72]. For primary resistant analysis in *ap1-11* intermediate allele, T1 seeds were sown on the 0.5X MS medium supplemented with 20 µg/ml hygromycin. Resistant plants were transferred to soil. Flowers arising from the primary shoot were analyzed by light microscopy. For the detailed analysis of the *AP1::AP1*, *AP1::AG* and *AP1::*AG^{IAP1}, the strong allele *ap1-7* was used as transformation background and T1 seeds were selected based on the seed specific GFP selection marker[70]. Flower phenotypic analyses were performed by light microscopy on flower number ~10 to 19 based on their order of emergence on more than 20 T1 plants for each construct, growing in parallel to *ap1*-7 plants. Three independent lines with one insertion expressing *AG, AP1* and AG^{IAP1} were selected for further phenotype analyses.

Environmental scanning electron microscopy

SEM experiments were performed at the Electron Microscopy facility of the ICMG Nanobio-Chemistry Platform (Grenoble, France) as described [33].

Table 1.

Data collection and refinement statistics

	$SEP3^{MI}$
Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	67.4, 67.4, 122.6
α, β, γ (°)	90, 90, 90
Resolution (Å)	48.-2.10 (2.16-2.10) *
R_{sym} or R_{merge}	8.4 (130)
$CC_{1/2}$	99.8 (56.4)
Completeness (%)	99.0 (95.6)
Redundancy	6.0 (4.1)
Refinement	
Resolution (Å)	47.7-2.10
No. reflections	16519
R_{work} / R_{free} (%)	20.4/23.6 (32.8/34.2)*
No. atoms	1880
Protein	1814
Ligand/ion	0
Water	66
B-factors	
Protein	55.5
Water	53.6
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.24

- refers to highest resolution shell

Funding: This project received support from the ANR (ANR-16-CE92-0023) and the GRAL LabEX (ANR-10-LABX-49-01) with the frame of the CBH-EUR-GS (ANR-17-EURE-0003). This work benefited from access to the MX-Grenoble, an Instruct-ERIC centre within the Grenoble Partnership for Structural Biology (PSB). Financial support was provided by Instruct-ERIC (PID 13317), FRISBI (ANR-10-INBS-05-02) and GRAL, a programme from the Chemistry Biology Health (CBH) Graduate School of University Grenoble Alpes (ANR-17-EURE-0003).

Author contributions: KK, FP, VH, CS and CZ conceived the study, CS, VH, RI, FP and CZ designed experiments, XL, RVL, CSS, AJ and VH performed experiments, RBM, JL and JM analyzed data, CZ wrote the manuscript with the help of all authors.

References

1 Passmore, S., Maine, G. T., Elble, R., Christ, C. & Tye, B. K. Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. *J Mol Biol* **204**, 593-606, doi:10.1016/0022-2836(88)90358-0 (1988).

2 Yanofsky, M. F. *et al.* The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* **346**, 35-39, doi:10.1038/346035a0 (1990).

683 3 Sommer, H. *et al.* Deficiens, a homeotic gene involved in the control of flower
684 morphogenesis in Antirrhinum majus: the protein shows homology to transcription
685 factors. *EMBO J* **9**, 605-613 (1990).
686 4 Norman, C., Runswick, M., Pollock, R. & Treisman, R. Isolation and properties of
687 cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum
688 response element. *Cell* **55**, 989-1003, doi:10.1016/0092-8674(88)90244-9 (1988).
689 5 Smaczniak, C., Angenent, G. C. & Kaufmann, K. SELEX-Seq: A Method to
690 Determine DNA Binding Specificities of Plant Transcription Factors. *Methods Mol*
691 *Biol* **1629**, 67-82, doi:10.1007/978-1-4939-7125-1_6 (2017).
692 6 Martinez-Castilla, L. P. & Alvarez-Buylla, E. R. Adaptive evolution in the
693 Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny.
694 *Proc Natl Acad Sci U S A* **100**, 13407-13412, doi:10.1073/pnas.1835864100 (2003).
695 7 Alvarez-Buylla, E. R. *et al.* An ancestral MADS-box gene duplication occurred before
696 the divergence of plants and animals. *Proc Natl Acad Sci U S A* **97**, 5328-5333,
697 doi:97/10/5328 [pii] (2000).
698 8 Pollock, R. & Treisman, R. Human SRF-related proteins: DNA-binding properties and
699 potential regulatory targets. *Genes Dev* **5**, 2327-2341 (1991).
700 9 Huang, H., Mizukami, Y., Hu, Y. & Ma, H. Isolation and characterization of the
701 binding sequences for the product of the Arabidopsis floral homeotic gene
702 AGAMOUS. *Nucleic Acids Res* **21**, 4769-4776 (1993).
703 10 Wynne, J. & Treisman, R. SRF and MCM1 have related but distinct DNA binding
704 specificities. *Nucleic Acids Res* **20**, 3297-3303, doi:10.1093/nar/20.13.3297 (1992).
705 11 Shiraishi, H., Okada, K. & Shimura, Y. Nucleotide sequences recognized by the
706 AGAMOUS MADS domain of Arabidopsis thaliana in vitro. *Plant J* **4**, 385-398
707 (1993).
708 12 Pellegrini, L., Tan, S. & Richmond, T. J. Structure of serum response factor core
709 bound to DNA. *Nature* **376**, 490-498, doi:10.1038/376490a0 (1995).
710 13 De Bodt, S. *et al.* Genomewide structural annotation and evolutionary analysis of the
711 type I MADS-box genes in plants. *J Mol Evol* **56**, 573-586, doi:10.1007/s00239-002-
712 2426-x (2003).
713 14 de Folter, S. *et al.* Comprehensive interaction map of the Arabidopsis MADS Box
714 transcription factors. *Plant Cell* **17**, 1424-1433, doi:tpc.105.031831 (2005).
715 15 Masiero, S., Colombo, L., Grini, P. E., Schnittger, A. & Kater, M. M. The emerging
716 importance of type I MADS box transcription factors for plant reproduction. *Plant*
717 *Cell* **23**, 865-872, doi:10.1105/tpc.110.081737 (2011).
718 16 Bemer, M., Heijmans, K., Airoldi, C., Davies, B. & Angenent, G. C. An atlas of type I
719 MADS box gene expression during female gametophyte and seed development in
720 Arabidopsis. *Plant Physiol* **154**, 287-300, doi:pp.110.160770 (2010).
721 17 Saedler, H., Becker, A., Winter, K. U., Kirchner, C. & Theissen, G. MADS-box genes
722 are involved in floral development and evolution. *Acta Biochim Pol* **48**, 351-358
723 (2001).
724 18 Becker, A. & Theissen, G. The major clades of MADS-box genes and their role in the
725 development and evolution of flowering plants. *Mol Phylogenet Evol* **29**, 464-489,
726 doi:S1055790303002070 [pii] (2003).
727 19 Kaufmann, K., Melzer, R. & Theissen, G. MIKC-type MADS-domain proteins:
728 structural modularity, protein interactions and network evolution in land plants. *Gene*
729 **347**, 183-198, doi:S0378-1119(04)00762-0 (2005).
730 20 Kwantes, M., Liebsch, D. & Verelst, W. How MIKC* MADS-box genes originated
731 and evidence for their conserved function throughout the evolution of vascular plant
732 gametophytes. *Mol Biol Evol* **29**, 293-302, doi:10.1093/molbev/msr200 (2012).

21 Adamczyk, B. J. & Fernandez, D. E. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiol* **149**, 1713-1723, doi:10.1104/pp.109.135806 (2009).

22 Verelst, W., Saedler, H., Munster, T. MIKC* MADS-Protein complexes Bind Motifs Enriched in the Proximal Region of Pollen-Specific Arabidopsis Promoters. *Plant Physiol* **143**, 447-460, doi:https://doi.org/10.1104/pp.106.089805 (2007).

23 Immink, R. G. *et al.* SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biol* **10**, R24, doi:gb-2009-10-2-r24 (2009).

24 Theissen, G. & Saedler, H. Plant biology. Floral quartets. *Nature* **409**, 469-471, doi:10.1038/35054172 (2001).

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

26 Gramzow, L. & Theissen, G. A hitchhiker's guide to the MADS world of plants. *Genome Biol* **11**, 214, doi:gb-2010-11-6-214 (2010).

27 Smaczniak, C., Immink, R. G., Angenent, G. C. & Kaufmann, K. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* **139**, 3081-3098, doi:10.1242/dev.074674 (2012).

28 Riechmann, J. L., Krizek, B. A. & Meyerowitz, E. M. Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc Natl Acad Sci U S A* **93**, 4793-4798, doi:10.1073/pnas.93.10.4793 (1996).

29 Krizek, B. A. & Meyerowitz, E. M. Mapping the protein regions responsible for the functional specificities of the Arabidopsis MADS domain organ-identity proteins. *Proc Natl Acad Sci U S A* **93**, 4063-4070, doi:10.1073/pnas.93.9.4063 (1996).

30 Riechmann, J. L. & Meyerowitz, E. M. Determination of floral organ identity by Arabidopsis MADS domain homeotic proteins AP1, AP3, PI, and AG is independent of their DNA-binding specificity. *Mol Biol Cell* **8**, 1243-1259, doi:10.1091/mbc.8.7.1243 (1997).

31 Jetha, K., Theissen, G. & Melzer, R. Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Res* **42**, 10927-10942, doi:10.1093/nar/gku755 (2014).

32 Lai, X., Stigliani, A., Lucas, J., Hugouvieux, V., Parcy, F., Zubieta, C. Genome-wide binidng of SEPALLATA3 and AGAMOUS complexes dertermined by sequential DNA-affinity purification sequencing. *Nucleic Acids Res* **48**, 9637-9648, doi:10.1093/nar/gkaa729 (2020).

33 Hugouvieux, V. *et al.* Tetramerization of MADS family transcription factors SEPALLATA3 and AGAMOUS is required for floral meristem determinacy in Arabidopsis. *Nucleic Acids Res* **46**, 4966-4977, doi:10.1093/nar/gky205 (2018).

34 Puranik, S. *et al.* Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. *Plant Cell* **26**, 3603-3615, doi:10.1105/tpc.114.127910 (2014).

35 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).

36 Crooks, G. E., Hon, G., Chandonia, J-M., Brenner, S.E. WebLogo:a sequence logo generator. *Genome Res* **14**, doi:10.1101/gr.849004 (2004).

37 Bemer, M., Wolters-Arts, M., Grossniklaus, U. & Angenent, G. C. The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. *Plant Cell* **20**, 2088-2101, doi:tpc.108.058958 (2008).

782 38 Steffen, J. G., Kang, I.-H., Portereiko, M. F., Lloyd, A. & Drews, G. N. AGL61
783 Interacts with AGL80 and Is Required for Central Cell Development in Arabidopsis.
784 *Plant Physiology* **148**, 259-268 (2008).
785 39 Pelaz, S., Gustafson-Brown, C., Kohalmi, S. E., Crosby, W. L. & Yanofsky, M. F.
786 APETALA1 and SEPALLATA3 interact to promote flower development. *Plant J* **26**,
787 385-394 (2001).
788 40 Kaufmann, K. *et al.* Target genes of the MADS transcription factor SEPALLATA3:
789 integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS*
790 *Biol* **7**, e1000090, doi:08-PLBI-RA-4384 (2009).
791 41 O'Maoileidigh, D. S. *et al.* Control of reproductive floral organ identity specification
792 in Arabidopsis by the C function regulator AGAMOUS. *Plant Cell* **25**, 2482-2503,
793 doi:10.1105/tpc.113.113209 (2013).
794 42 Mizukami, Y., Huang, H., Tudor, M., Hu, Y. & Ma, H. Functional domains of the
795 floral regulator AGAMOUS: characterization of the DNA binding domain and
796 analysis of dominant negative mutations. *Plant Cell* **8**, 831-845,
797 doi:10.1105/tpc.8.5.831 (1996).
798 43 Riechmann, J. L., Wang, M. & Meyerowitz, E. M. DNA-binding properties of
799 Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3,
800 PISTILLATA and AGAMOUS. *Nucleic Acids Res* **24**, 3134-3141, doi:6t0273 [pii]
801 (1996).
802 44 Smaczniak, C., Muino, J. M., Chen, D., Angenent, G. C. & Kaufmann, K. Differences
803 in DNA Binding Specificity of Floral Homeotic Protein Complexes Predict Organ-
804 Specific Target Genes. *Plant Cell* **29**, 1822-1835, doi:10.1105/tpc.17.00145 (2017).
805 45 Guo, Y., Tian, K., Zeng, H., Gifford, D.K. A novel k-mer set memory (KSM) motif
806 representation improves regulatory variant prediction. *Genome Res* **28**, 891-900,
807 doi:10.1101/gr.226852.117 (2018).
808 46 Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E.M., Smyth, D.R. Control of
809 Flower development in Arabidopsis thaliana by APETALA1and interacting genes.
810 *Development* **119**, 721-743 (1993).
811 47 Irish, V. F. & Sussex, I. M. Function of the apetala-1 gene during Arabidopsis floral
812 development. *Plant Cell* **2**, 741-753, doi:10.1105/tpc.2.8.741 (1990).
813 48 Masiero, S. *et al.* Ternary complex formation between MADS-box transcription
814 factors and the histone fold protein NF-YB. *J Biol Chem* **277**, 26429-26435,
815 doi:10.1074/jbc.M202546200 (2002).
816 49 Mendes, M. A. *et al.* MADS domain transcription factors mediate short-range DNA
817 looping that is essential for target gene expression in Arabidopsis. *Plant Cell* **25**,
818 2560-2572, doi:10.1105/tpc.112.108688 (2013).
819 50 Melzer, R., Verelst, W. & Theissen, G. The class E floral homeotic protein
820 SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro.
821 *Nucleic Acids Res* **37**, 144-157, doi:10.1093/nar/gkn900 (2009).
822 51 Davies, B., Sommer, H., Schwarz-Sommer, Z. *Flower Development: Genetic Views*
823 *and Molecular News*. 167-184 (Springer Science and Business Media, 1999).
824 52 Brambilla, V. *et al.* Genetic and molecular interactions between BELL1 and MADS
825 box factors support ovule development in Arabidopsis. *Plant Cell* **19**, 2544-2556,
826 doi:10.1105/tpc.107.051797 (2007).
827 53 Liu, C., Xi, W., Shen, L., Tan, C. & Yu, H. Regulation of floral patterning by
828 flowering time genes. *Developmental cell* **16**, 711-722,
829 doi:10.1016/j.devcel.2009.03.011 (2009).

54 Vandenbussche, M., Theissen, G., Van de Peer, Y. & Gerats, T. Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Res* **31**, 4401-4409 (2003).
55 Egea-Cortines, M., Saedler, H. & Sommer, H. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. *EMBO J* **18**, 5370-5379, doi:10.1093/emboj/18.19.5370 (1999).
56 Gabadinho, J. *et al.* MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *J Synchrotron Radiat* **17**, 700-707, doi:10.1107/S0909049510020005 (2010).
57 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133-144, doi:10.1107/S0907444909047374 (2010).
58 de Folter, S. a. I., R.G.H. in *Plant Transcription Factors. Methods in Molecular Biology (Methods and Protocols)* Vol. 754 (ed L. Yuan., Perry, S.) 145-165 (Humana Press, 2011).
59 Gaspar, J. M. NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics* **19**, 536, doi:10.1186/s12859-018-2579-2 (2018).
60 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
61 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
62 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
63 Jalili, V., Matteucci, M., Masseroli, M. & Morelli, M. J. Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics* **31**, 2761-2769, doi:10.1093/bioinformatics/btv293 (2015).
64 Freese, N. H., Norris, D. C. & Loraine, A. E. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* **32**, 2089-2095, doi:10.1093/bioinformatics/btw069 (2016).
65 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag 2016).
66 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-208, doi:10.1093/nar/gkp335 (2009).
67 Mathelier, A. & Wasserman, W. W. The next generation of transcription factor binding site prediction. *PLoS Comput Biol* **9**, e1003214, doi:10.1371/journal.pcbi.1003214 (2013).
68 Stigliani, A. *et al.* Capturing Auxin Response Factors Syntax Using DNA Binding Models. *Mol Plant* **12**, 822-832, doi:10.1016/j.molp.2018.09.010 (2019).
69 Curtis, M. D. G., U. A Gateway Cloning Vector Set for High-throughput Functional Analysis of Genes in Planta. *Plant Physiol* **133**, 462-469 (2003).
70 Bensmihen, S. *et al.* Analysis of an activated ABI5 allele using a new selection method for transgenic Arabidopsis seeds. *FEBS Lett* **561**, 127-131, doi:10.1016/S0014-5793(04)00148-6 (2004).
71 O'Maoileidigh, D. S., Thomson, B., Raganelli, A., Wuest, S.E., Ryan, P.T., Kwasniewska, K., Carles, C.C., Graciet, E., Wellmer, F. Gene network analysis of Arabidopsis thaliana flower development through dynamic gene perturbations. *The Plant Journal* **83**, 344-358 (2015).

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

Figures

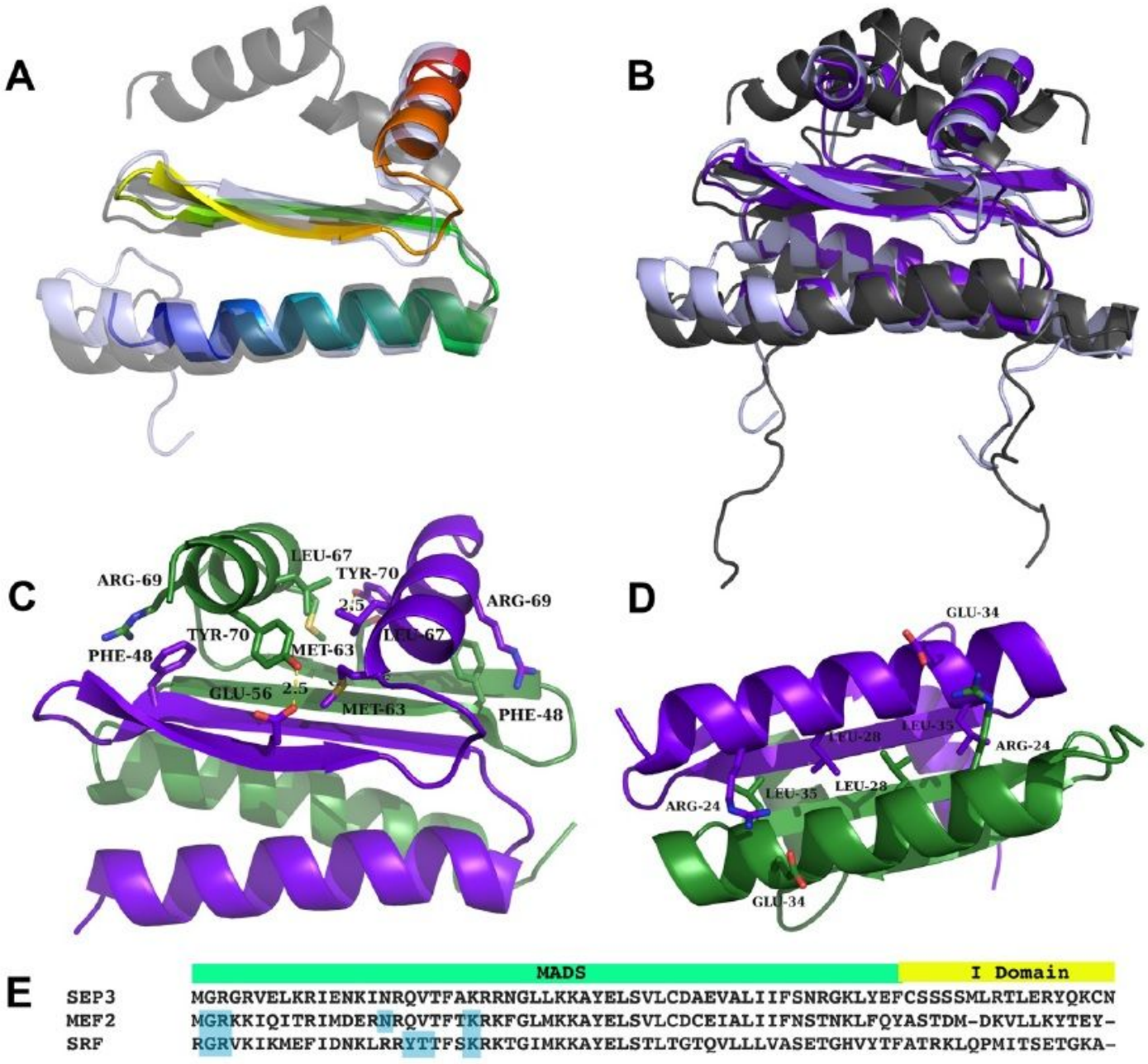

Figure 1

Structure and sequence of the SEP3 MI domain and corresponding SRF and MEF2 regions. (A) Overlay of SEP3 dimer (rainbow) and SRF (1HBX) dimer (gray). The I region is alpha helical for both structures, however the contacts with the M domain differ. (B) Overlay of SEP3 (dark purple) and MEF2A (3KOV, light purple) and SRF (gray) demonstrating the different conformations of the I region in the type II (MEF2-like) and I (SRF-like) MADS TFs. (C) SEP3 dimer with one monomer in purple and one in green. Amino acids important for interactions between I domains and between M and I domains are labelled and drawn as sticks. Hydrogen bonds are shown as dashed yellow lines. (D) View of the I domain with intermolecular

interactions from the N-terminal alpha helices shown. (E) Partial sequence alignment of SEP3 (accession NP564214.2), MEF2 (accession AAB25838.1) and SRF (accession P11831) corresponding to the region in the crystal structure. Residues directly contacting the DNA for MEF2 and SRF are highlighted in blue.

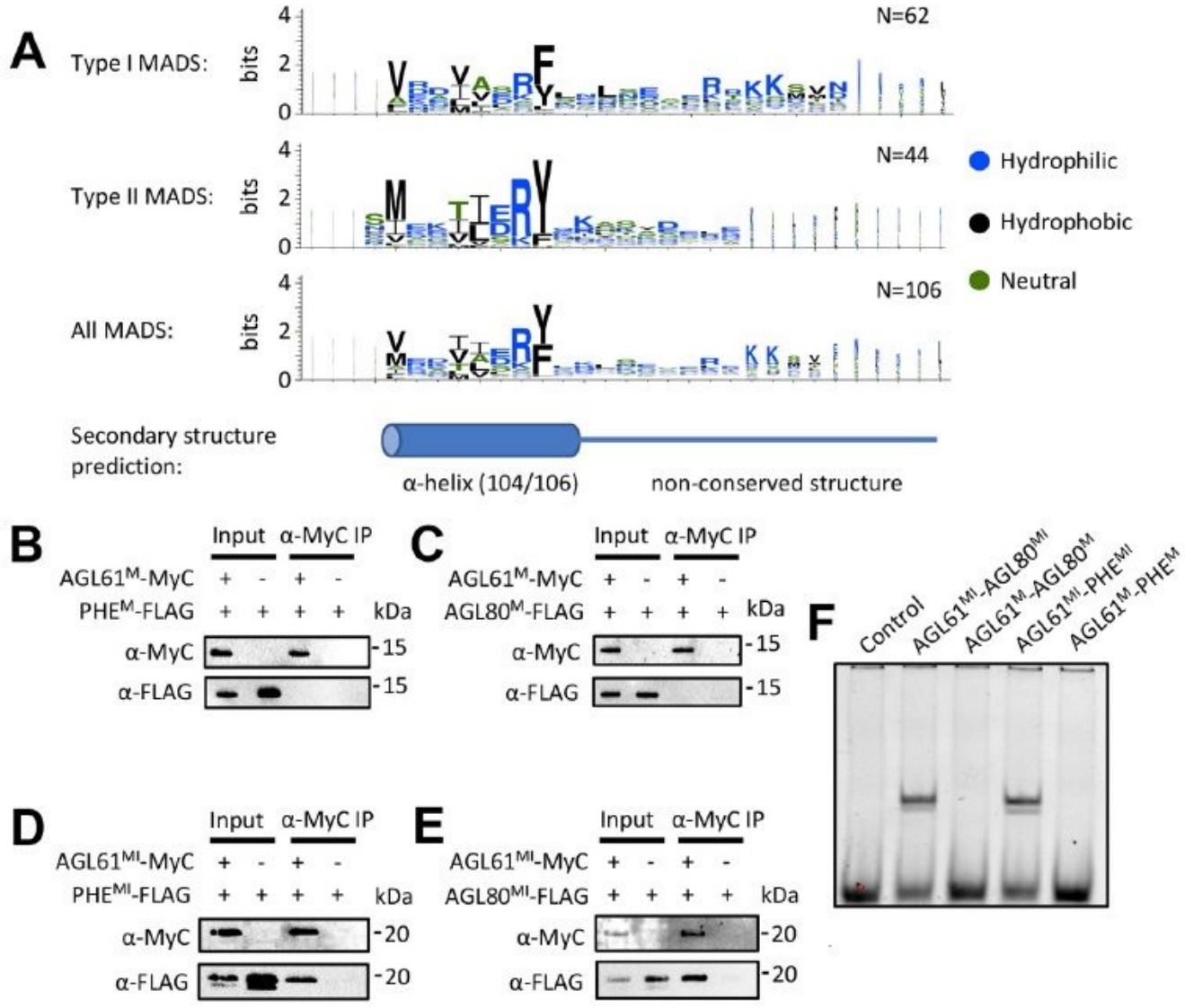

Figure 2

Type I MADS possesses I domain like region which is required for both dimerization and DNA binding. (A) Amino acid enrichment of the I region (~30 amino acids C-terminal to the M domain) of type I, type II and all MADS TFs, logos generated with WebLogo36. The overall height of the stack in each position indicates the sequence information content at that position, while the height of the amino acid symbols within the stack indicates the relative frequency at each position. The MADS TF sequences are taken from The Arabidopsis Information Resource (www.arabidopsis.org). (B-C) Pulldown assay showing that M domain of AGL61 (AGL61M) does not interact with the M domain of PHE (PHEM) or AGL80 (AGL80M). (D-E) Pulldown assay showing that the M domain plus the I like region of AGL61 (AGL61MI) interacts with the MI region of PHE (PHEMI) and AGL80 (AGL80MI). (F) EMSA assay showing that heterodimers

AGL61MI-AGL80MI and AGL61MI-PHEMI shift a DNA sequence containing a canonical CArG-box binding site from the SEP3 promoter, while their corresponding constructs without the I region do not exhibit any binding.

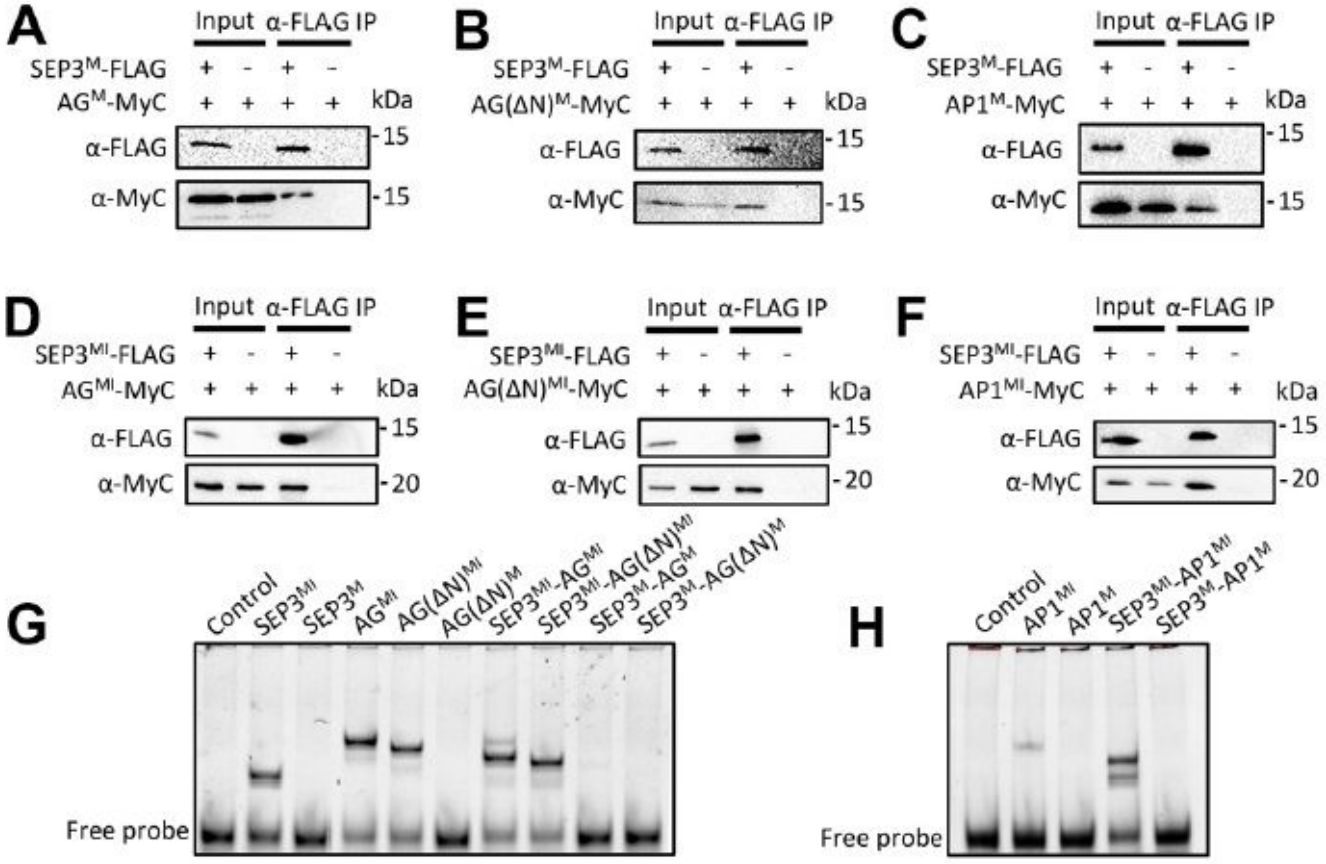

Figure 3

Type II MADS TFs require the I domain for DNA binding but not dimerization. (A-C) Pulldown assay showing that the M domain of SEP3 (SEP3M) interacts with the M domains of AG (AGM), AG with the first 16 N-terminal amino acids deleted (AG(ΔN)M) and AP1 (AP1M), respectively. (D-F) Pulldown assay showing that the MI domain of SEP3 (SEP3MI) interacts with the MI domain of AGMI, AG(ΔN)MI and AP1MI, respectively. (G and H) EMSA assay showing that homodimers from SEP3MI, AGMI, AG(ΔN)MI and AP1MI, and heterodimers SEP3MI-AGMI, SEP3MI-AG(ΔN)MI and SEP3MI AP1MI shift a DNA sequence containing a canonical CArG-box binding site (as per Figure 2), while their corresponding constructs without the I domain can not, suggesting that the I domain in type II MADS TFs is required for DNA binding.

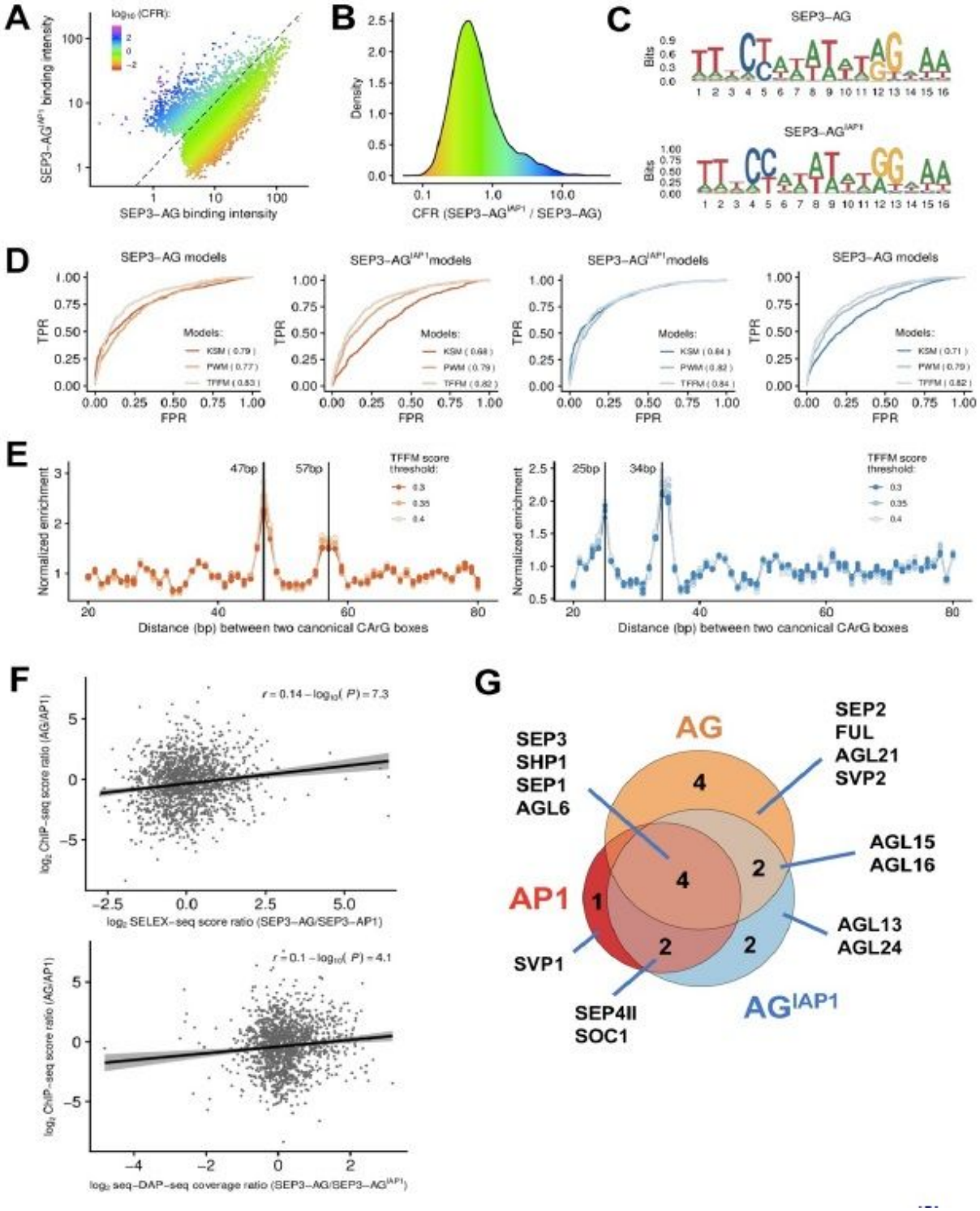

Figure 4

DNA-binding and protein interactions patterns. (A) Comparison of SEP3-AG and SEP3-AGIAP1 seq-DAP-seq binding intensity (log10 of reads per kb per million of reads mapped in bound regions) and color-coded by purple-blue (SEP3-AGIAP1-specific) to orange-red (SEP3-AG-specific) according to log10 of SEP3-AGIAP1/SEP3-AG. (B) Density plot showing data as per A. (C) Logos derived from PWM-based models obtained for SEP3-AG and SEP3-AGIAP1. (D) Predictive power of TFBS models. Models are built

using 600 sequences best bound by each of the two heterocomplexes and are searched against 1,073 SEP3-AG (orange) and 1,073 SEP3-AGIAP1 (blue) specific regions, defined as the top 15% of sequences that are most strongly bound by one complex relative to the other. Matrix-based models (PWM and TFFM) are not able to differentiate SEP3-AG and SEP3-AGIAP1 binding whereas k-mer-based analysis is able to better predict binding for the respective datasets. (E) SEP3-AG favors intersite spacings of 47 and 57 bp based on SEP3-AG specific regions. SEP3-AGIAP1 favors intersite spacings of 25 and 34 bp based on SEP3-AGIAP1I specific regions. (F) Top, published SELEXseq for SEP3-AP1 and SEP3-AG 5 comparing the normalized score ratios (SEP3-AG/SEP3-AP1) for SELEX-seq and score ratios (AG/AP1) ChIP-seq at 1,500 SEP3 best bound loci in ChIP-seq show a positive correlation, suggesting that SEP3- AP1 and SEP-AG bind different sequences in vivo and that in vitro binding is able to differentiate bound sequences that are more SEP3-AP1-like versus SEP3-AG-like. Bottom, SEP3-AG and SEP3-AGIAP1I seq-DAP-seq coverage as per SELEXseq scores. A positive correlation is observed suggesting that, in vitro, the swap of AP1 I domain in AG is able to recover some of the binding specificity of SEP3-AP1. (G) Yeast two-hybrid assays using AG, AP1 and AGIAP1 as bait against MIKCC MADS TFs in Arabidopsis. Data shows that AGIAP1 loses AG interactors and gains AP1 interactors.

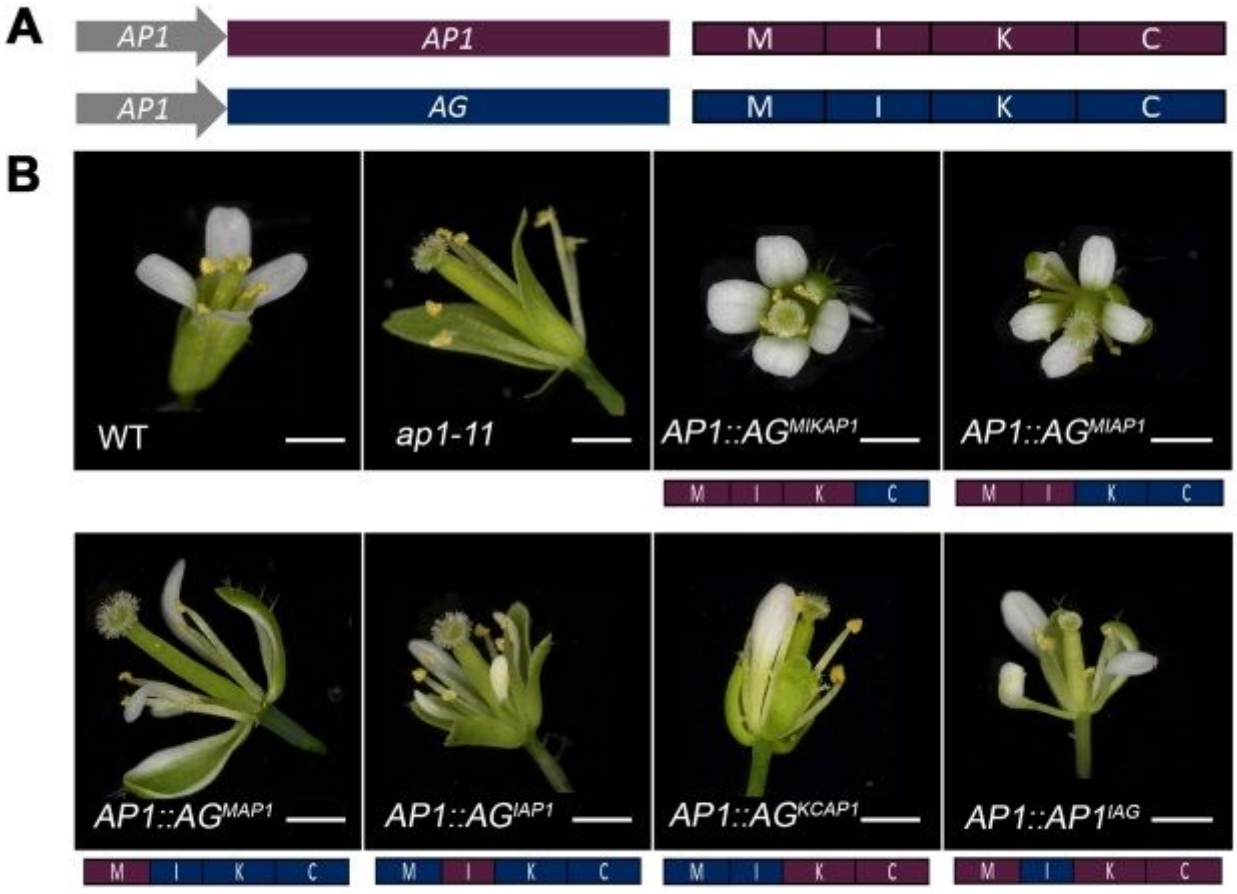

Figure 5

Primary transformants in the ap1mutant background exhibit a spectrum of complementation. (A) Schematic of the constructs and proteins produced with AP1 protein in purple and AG in dark blue. Domains are labeled MIKC. (B) Flower phenotypes of T1 transformants (n ≥ 10). All transformants were in the ap1-11 background. Phenotypes for WT and ap1-11 are shown. The domains corresponding to AP1 and AG are colored as per (A) and shown schematically below each panel.

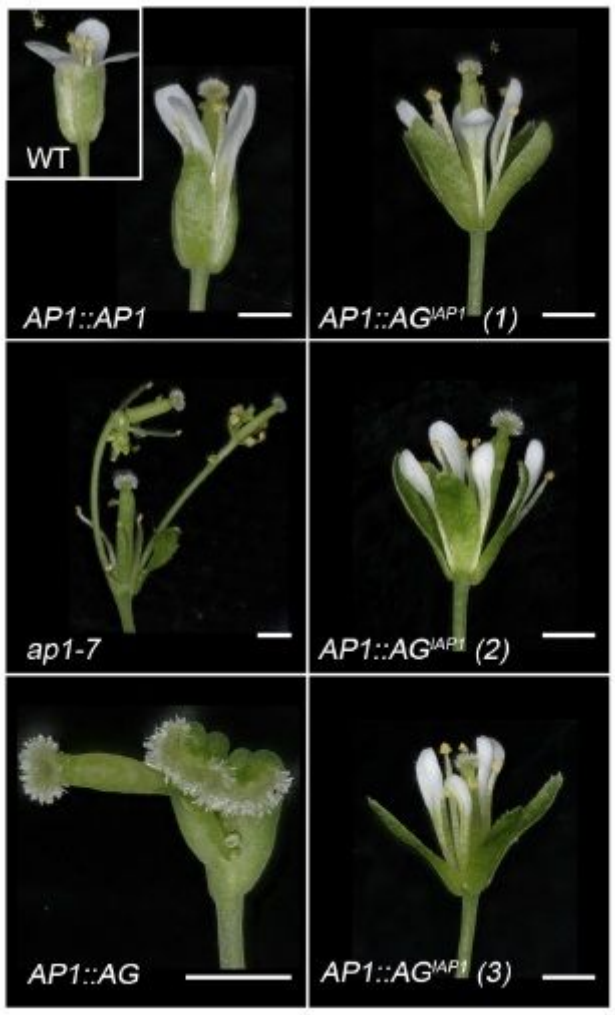

Figure 6

AP1:AGIAP1 expression largely complements the ap1-7 flower phenotype. ap1-7 lines expressing either AP1 (AP1::AP1)), AG (AP1::AG) or AGIPA1 (AP1::AGIAP1) under the control of the AP1 promoter were grown at 22 °C in long days. A typical WT flower is shown as an insert in the upper left panel. Representative flowers for 3 independent AP1::APIAP1 lines are presented and one representative flower from AP1::AG and AP1::AP1 expressing lines. While the first whorl in ap1-7 is replaced by bract or stipule with axillary buds and petals are missing in the second whorl, AP1::AGIAP1 expression restores WT first and second whorl organs, while AG expression triggers carpel development in the first whorl and absence of petals in the second whorl. Scales bars = 1 mm.

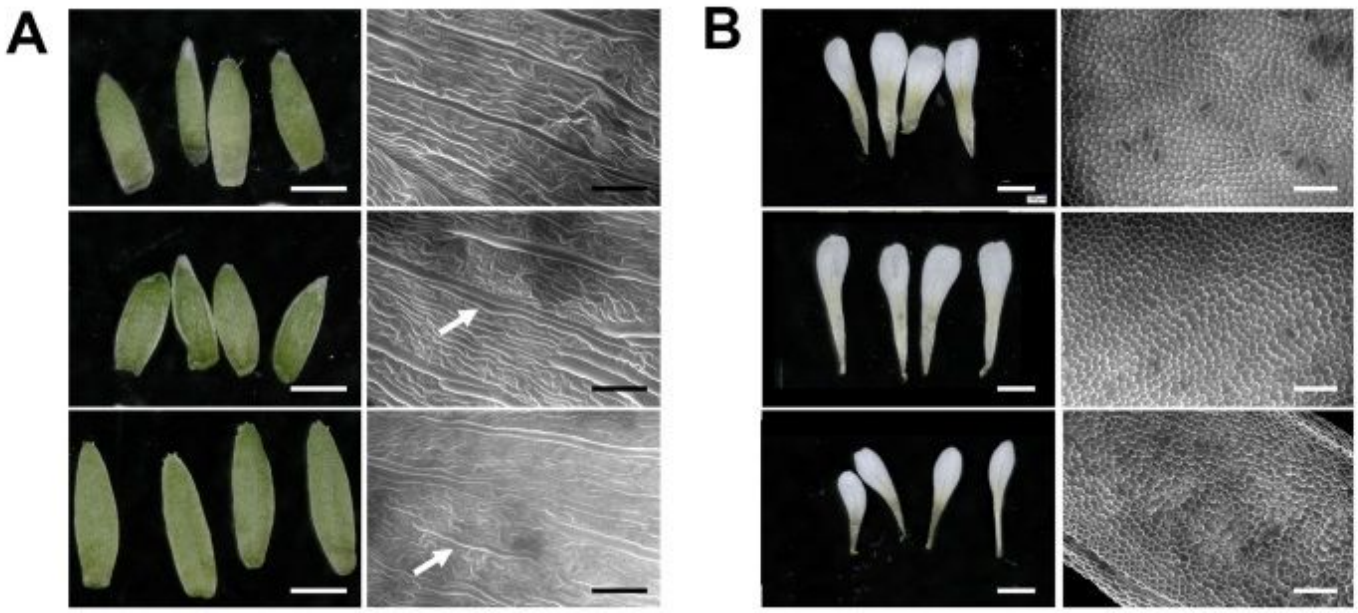

Figure 7

First and second whorls cell identity are complemented in ap1-7 plants expressing AGIAP1. First (A) and second (B) whorl organs were removed from WT, and flowers from AP1 or AGIAP1 expressing ap1-7 plants. First whorl organs in AGIAP1 expressing plants are slightly longer compared to WT and AP1 expressing plants (A, left panel). Second whorl organs in AGIAP1 expressing plants are slightly smaller compared to WT and AP1 expressing plants (B, left panel). Epidermal cell identity, observed by SEM, show characteristic WT elongated sepal cells in AP1 and AGIAP1 expressing plants (A, right panel) and characteristic WT conical cells in petals from AP1 and AGIAP1 expressing plants (B, right panel). A small number of epidermal cells typical of leaves are also seen in AGIAP1 expressing plants and to a lesser extent in AP1 expressing plants (arrow). Scale bars indicate 100 μM for SEM images and 1 mm for organ photographs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarysubmissioncompressed.pdf