

1 The Intervening Domain Is Required For DNA-binding and Functional Identity of Plant 2 MADS Transcription Factors

3
4 Xuelei Lai^{1,*}, Rosario Vega-Leon^{2,*}, Veronique Hugouvieux^{1,*,#}, Romain Blanc-Mathieu¹,
5 Froukje van der Wal³, Jérémy Lucas¹, Catarina S. Silva^{1,4}, Agnès Jourdain¹, Jose Muino⁵, Max
6 H. Nanao⁶, Richard Immink^{3,7}, Kerstin Kaufmann², François Parcy¹, Cezary Smaczniak^{2,#},
7 Chloe Zubieta^{1,#}
8

9 ¹ Laboratoire Physiologie Cellulaire et Végétale, Univ. Grenoble Alpes, CNRS, CEA, INRAE, IRIG-DBSCI-
10 LPCV, 17 avenue des martyrs, F-38054, Grenoble, France.

11 ² Plant Cell and Molecular Biology, Humboldt-Universität zu Berlin, Institute of Biology, Berlin, Germany.

12 ³ Bioscience, Wageningen Plant Research, Wageningen University and Research, 6708 PB, Wageningen, The
13 Netherlands.

14 ⁴ European Molecular Biology Laboratory, Grenoble, France

15 ⁵ Systems Biology of Gene Regulation, Humboldt-Universität zu Berlin, Institute of Biology, Berlin, Germany.

16 ⁶ European Synchrotron Radiation Facility, Structural Biology Group, Grenoble, France

17 ⁷ Laboratory of Molecular Biology, Wageningen University and Research, 6708 PB, Wageningen, The
18 Netherlands.

19 *first

20 #correspondence
21
22

23 **Abstract**

24 The MADS transcription factors (TF) are an ancient protein family with a high degree of
25 sequence identity that bind almost identical DNA sequences across all eukaryotic kingdoms of
26 life, yet fulfill dramatically different physiological roles. In plants, the family is divided into
27 two main lineages, type I and II, based on sequence conservation of the DNA-binding MADS-
28 box domain (M domain) with yeast and animal M domains. Here, we demonstrate that DNA
29 binding in both lineages absolutely requires a short amino acid sequence C-terminal to the M
30 domain called the Intervening domain (I domain) in type II MADS. Structural elucidation of
31 the MI domains from the floral regulator, SEPALLATA3 (SEP3), shows a highly conserved
32 MADS-box fold with the I domain forming an alpha helix and acting to stabilize the M domain.
33 Based on secondary structure prediction, sequences fulfilling the same function as the SEP3 I
34 domain can be found in both lineages of plant MADS TFs, suggesting the I domain is a
35 conserved and required part of the DNA-binding domain. Using the floral organ identity MADS
36 TFs, SEP3, APETALA1 (AP1) and AGAMOUS (AG), domain swapping demonstrate that the
37 I domain alters DNA-binding specificity based on seq-DAP-seq experiments. Yeast 2-hybrid
38 experiments further revealed the role of the I domain in dimerization specificity. Surprisingly,
39 introducing AG carrying the I domain of AP1 in the Arabidopsis *ap1* mutant, resulted in a high
40 degree of complementation and restoration of first and second whorl organs. Taken together,
41 these data demonstrate that the I domain acts both as an integral part of the DNA-binding
42 domain and strongly contributes to the functional identity of the MADS TF.
43

44 **Introduction**

45 The MADS-box genes, named after founding members *MINICHROMOSOME*
46 *MAINTENANCE1* (*MCMI*, *Saccharomyces cerevisiae*), *AGAMOUS* (*Arabidopsis thaliana*),
47 *DEFICIENS* (*Antirrhinum majus*), and *SERUM RESPONSE FACTOR* (*SRF*, *Homo sapiens*),

48 is an ancient gene family present before the ancestral split of animals, plants and fungi¹⁻⁴. Prior
49 to the divergence of eukaryotes into different kingdoms of life, the MADS gene family
50 underwent a duplication event giving rise to two main lineages, the SRF-like and MEF2-like
51 lineages, which correspond to the type I and II MADS genes in plants, respectively^{6,7}. While
52 these two MADS lineages both encode a highly conserved ~60 amino acid MADS-box DNA-
53 binding domain (DBD) that recognizes a CArG box motif (CC-“Adenine rich”-GG), DNA-
54 binding preferences are slightly altered due to amino acid changes in the MADS-box domain
55 specific to each lineage⁸⁻¹¹. In addition to changes in amino acids important for direct base
56 readout, the region C-terminally adjacent to the MADS-box domain varies based on sequence
57 alignments and available structural data for mammalian and fungal SRF-like and MEF2-like
58 MADS TFs. These carboxyl terminal sequences have been shown to be important for
59 dimerization and DNA binding although they do not directly contact the DNA^{4,8,12}. While the
60 MADS genes are ubiquitous in extant eukaryotes, they have undergone a significant expansion
61 in plants, with angiosperms possessing tens of type I SRF-like and type II MEF2-like MADS
62 genes that fulfill diverse physiological roles.

63 The plant type I MADS genes generally consist of one or two exons and encode
64 transcription factors (TFs) comprised of the MADS-box DBD and a variable C-terminal
65 domain. The type I genes are further subdivided into three subfamilies, $M\alpha$, $M\beta$ and $M\gamma$, with
66 preferential interactions between subfamilies forming heterodimeric complexes^{13,14}. While
67 generally expressed at very low levels in a tissue specific manner, the type I genes have been
68 shown to be important in plant reproduction and speciation, with crucial roles in female
69 gametophyte, embryo, and endosperm development^{15,16}. In contrast to the simple one or two
70 exon composition and protein structure of the type I MADS, the type II genes consist of 5-8
71 exons and encode TFs with a modular four-domain structure, called “MIKC”¹⁷⁻¹⁹. The MIKC
72 domains refer to M, for MADS DBD, I for the dimerization specifying Intervening domain, K
73 for the coiled-coil keratin-like domain involved in dimer and tetramer formation and C for the
74 variable C-terminal domain important for transactivation and higher order complex formation.
75 The type II MADS are further subdivided into the MIKC* and MIKC^C subfamilies based on
76 the structure of the I domain, with dimerization mainly occurring within the same subfamily,
77 but not between subfamilies²⁰⁻²². While all MADS TFs are believed to form dimeric complexes
78 in order to bind DNA, the type II MADS TFs are able to tetramerise via the K domain, greatly
79 expanding the number of possible MADS complexes that can be formed²³⁻²⁵. The diversity of
80 heteromeric complexes the type II MADS form is hypothesized to be directly tied to their
81 functional diversity²⁶. The type II MADS play well established roles in many developmental

82 processes including meristem identity, flowering time, fruit and seed development and floral
83 organ identity²⁷.

84 The MADS TFs involved in floral organ specification are arguably the most well studied
85 MADS family members and provide an ideal model to examine function *in planta*. Elegant *in*
86 *vitro* and *in vivo* experiments dating back decades have tried to address the question of how
87 plant MADS TF achieve their DNA-binding specificity and functional diversity in the context
88 of flower and floral organ development^{28,29}. Four classes of MIKC MADS TFs, class A, B, C
89 and E are necessary for floral organ identity. The A+E class specify sepals, the A+B+E petals,
90 A+C stamen and C+E carpels. In Arabidopsis, these correspond to APETALA1 (AP1, A class),
91 APETALA3 (AP3, B class) and PISTILLATA (PI, B class), AGAMOUS (AG, C class) and
92 the four SEPALLATAs (SEP1 -4, E class). The generation of chimeric MADS TFs with
93 swapped M domains from AP1, AP3, PI and AG demonstrated that the MADS domains of these
94 proteins were interchangeable with respect to their *in vivo* function^{29,30}. Even swapping in
95 portions of human or yeast M domains, which did alter *in vitro* DNA-binding, did not alter the
96 function of the MADS proteins *in vivo*, suggesting that the physiological function of these
97 proteins seems to be independent of the DNA-binding domain³⁰. Thus, tetramerisation in MIKC
98 MADS TFs came under scrutiny as a potential determinant of DNA-binding syntax by selecting
99 for two-site DNA-binding constrained by specific intersite distances^{31,32}. More recent studies
100 of a tetramerisation mutant of SEP3 revealed that while tetramerisation contributed to *in vivo*
101 function, *in vitro* changes in DNA-binding specificity genome-wide were more limited,
102 putatively affecting a relatively small number of key gene targets whose regulation depends on
103 specific spacings between CARG box binding sites^{32,33}. Thus, the fundamental question of how
104 plant MADS TFs are able to recognize different DNA sequences critical for the regulation of
105 different target genes, with apparently interchangeable DBDs, is still not fully resolved.

106 In order to address this, we performed structural, biochemical, genome-wide binding
107 and *in vivo* studies focusing on the role of the Intervening domain. We demonstrate that an I-
108 like domain is present in both type I and type II plant MADS TFs and that this region is required
109 for DNA-binding, affects DNA-binding specificity and alters dimerization specificity *in vitro*
110 and in yeast assays. Seq-DAP-seq experiments comparing SEP3-AG and SEP3-AG^{IAP1}, a
111 chimera with the I domain from AG replaced by the I domain of AP1, revealed unique binding
112 sites for each complex, changes in preferred site spacing and a loss of carpel and fourth whorl-
113 specific targets for the chimeric protein. Using these same floral organ identity MADS TFs to
114 probe *in vivo* function of the I domain, different chimeric constructs were introduced into the
115 *apl* background. Interestingly, replacing the I domain of *AG* with that of *AP1* was sufficient to

116 confer the majority of *API* functions *in planta* including sepal and petal identity in the first and
117 second whorls. Taken together, these data illustrate the importance and multiple roles of the I
118 domain in DNA-binding and *in planta* function.

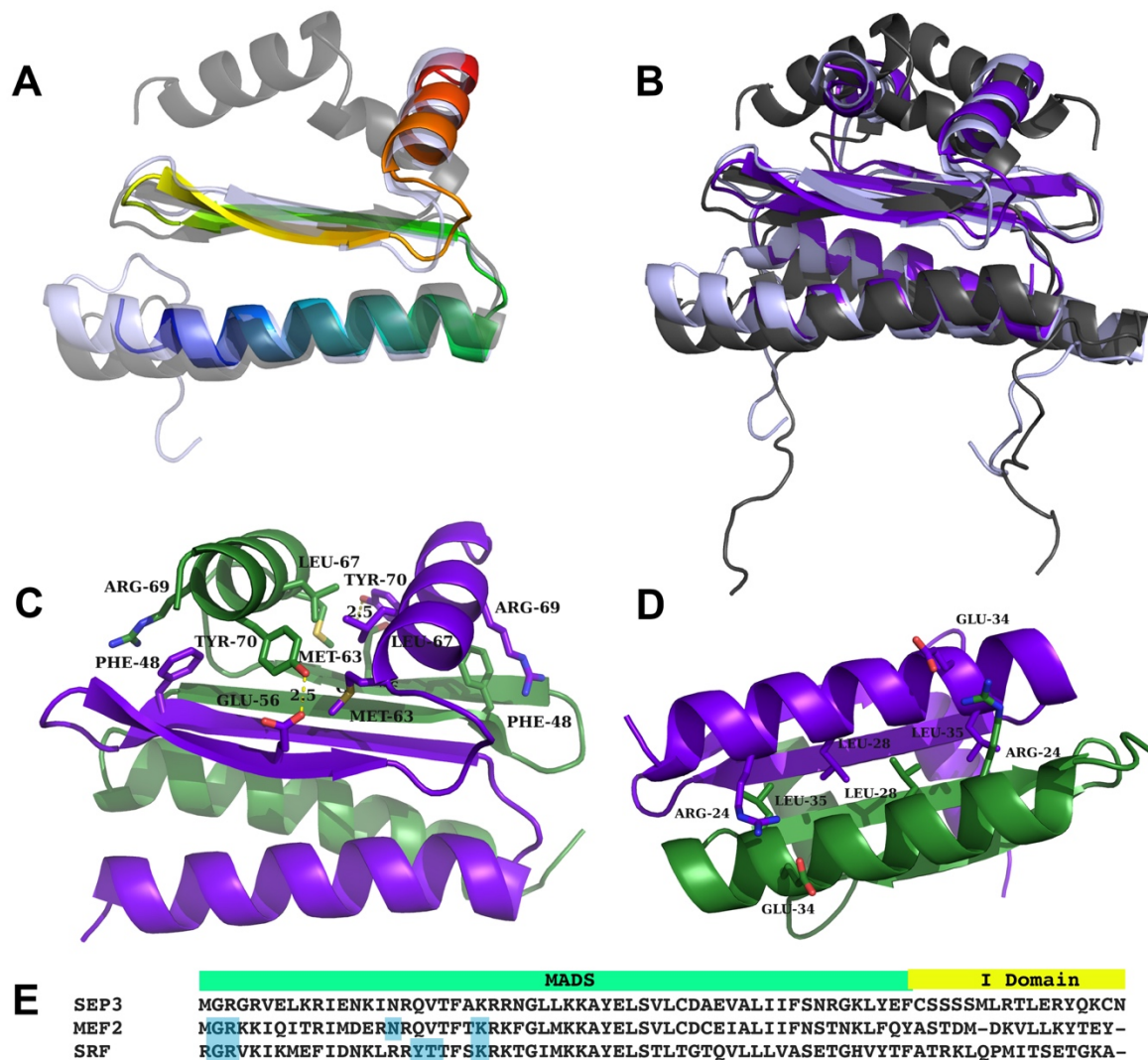
119

120 **Results**

121 **Structure of the SEPALLATA3 MI domain**

122 The MADS-box and I domain of SEP3 (SEP3^{MI}, residues 1-90) was used in structural
123 studies as this construct has been shown to be dimeric in solution and to bind DNA in a sequence
124 specific manner³⁴. SEP3^{MI} crystallized in spacegroup C222₁ with 4 monomers per asymmetric
125 unit and diffraction to 2.1 Å (**Table 1**). The first 18 amino acids, which are predicted to have
126 no defined secondary structure, were disordered as were the 17 C-terminal amino acids, and no
127 electron density was interpretable for these portions of the protein in any of the molecules. Each
128 monomer adopts the same tertiary structure with an N-terminal alpha helix (α 1; aa 18-40) and
129 two antiparallel beta strands (β 1; aa 44-48 and β 2; aa 55-58). These elements make up the core
130 MADS DBD and are conserved in the MEF2 and SRF-like MADS TFs (**Figure 1A and B**). C-
131 terminal to the beta strands is a short loop followed by an alpha helix which is contributed by
132 the I domain (α 2; aa 63-73). The orientation of the C-terminal alpha helix differs between MEF2
133 (PDB codes 1EGW, 1C7U, 1TQE, 3KOV) and SRF-like MADS TFs (PDB code 1HBX), with
134 SEP3^{MI} resembling the MEF2 structure. Overall, SEP3^{MI} adopts the MADS/MEF2 fold, as
135 predicted for type II plant MADS TFs.

136 Each SEP3^{MI} monomer dimerizes via extensive contacts between the pairs of alpha
137 helices and beta strands from each partner, with dimerization required for DNA binding. In the
138 structure, all dimers were formed by crystallographic symmetry, with dimerization burying over
139 25% of the total surface area of the molecule (**Figure 1**)³⁵. The dimer interface includes
140 hydrophobic interactions between residues in the N-terminal DNA-binding alpha helices,
141 Leu28 and Leu35, and pairs of salt bridges between Glu34 and Arg24 (**Figure 1D**). The
142 formation of a 4-stranded antiparallel beta sheet further stabilizes the quaternary structure of
143 SEP3 with interactions bridging the N-terminal DNA binding helices and the C-terminal I
144 domain helices. The I domain helices sandwich the beta sheet and lie perpendicular to the DNA-
145 binding helices of the MADS-box domain. The I domain is anchored to the beta sheet via
146 hydrogen bonding interactions between Glu56 and Tyr70 and pi-cation interactions between
147 Phe48 and Arg69 of the partner monomers. Each I domain helix also interacts with



148
149 **Figure 1. Structure and sequence of the SEP3 MI domain and corresponding SRF and MEF2 regions.** (A) Overlay of
150 SEP3 dimer (rainbow) and SRF (1HBX) dimer (gray). The I region is alpha helical for both structures, however the contacts
151 with the M domain differ. (B) Overlay of SEP3 (dark purple) and MEF2A (3KOV, light purple) and SRF (gray) demonstrating
152 the different conformations of the I region in the type II (MEF2-like) and I (SRF-like) MADS TFs. (C) SEP3 dimer with one
153 monomer in purple and one in green. Amino acids important for interactions between I domains and between M and I domains
154 are labelled and drawn as sticks. Hydrogen bonds are shown as dashed yellow lines. (D) View of the I domain with
155 intermolecular interactions from the N-terminal alpha helices shown. (E) Partial sequence alignment of SEP3 (accession
156 NP564214.2), MEF2 (accession AAB25838.1) and SRF (accession P11831) corresponding to the region in the crystal structure.
157 Residues directly contacting the DNA for MEF2 and SRF are highlighted in blue.

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

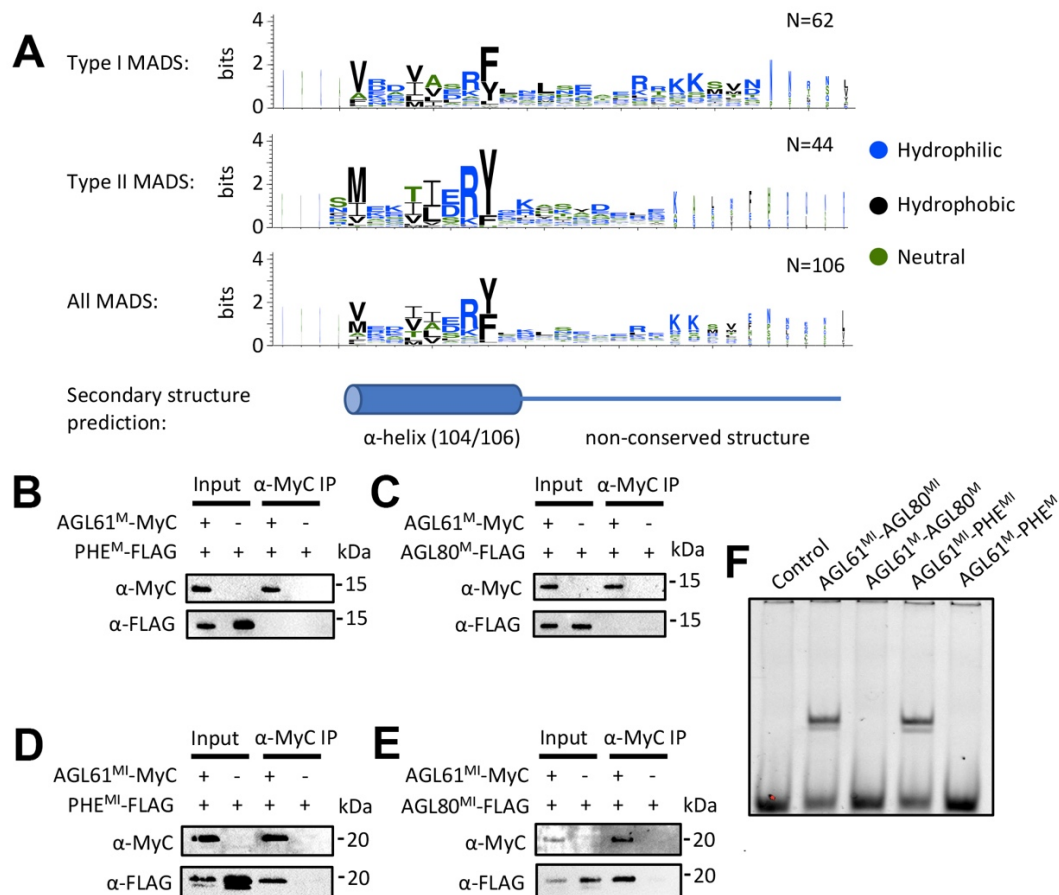
164 SEP3^{MI} was crystallised without DNA, however, multiple DNA-bound structures of the
165 closely related MEF2 protein are available for comparison. Structural alignment of SEP3^{MI} with
166 the corresponding MEF2 structure reveals a rmsd of 0.942 Å², underscoring the conservation

167 in the MADS/MEF2 domains with the MADS/I domains of SEP3^{M1} (**Figure 1A, B**). The
168 residues directly contacting the DNA are contributed from the flexible N-terminus and the N-
169 terminal helix of the MADS-box domain, with no direct contacts from the MEF2/I domain
170 (**Figure 1E**). These DNA-binding residues are highly conserved in SEP3 even though MEF2
171 recognizes a YTA(A/T)₄TAR sequence (Y=pyrimidine, R=purine), whereas SEP3 recognizes
172 a classic CArG box motif (CC(A/T)₆GG). The differences in DNA-binding specificity of the
173 MEF2-like MADS TFs coupled with the high conservation of sequence and structure of the
174 MADS-box domain itself, suggest allosteric contributions from the MEF2/I domain may play
175 a role in DNA binding. Indeed, examining the DNA-binding of the I domain mutants R69L,
176 R69P and Y70E demonstrated that these single point mutations were sufficient to abrogate the
177 DNA binding activity of the full-length SEP3 based on electrophoretic mobility shift assays
178 (EMSAs) (**Supplementary Figure 1**).

179

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

181 Structural and mutagenesis data suggest that the MEF2/I domain is critical for stability
182 and DNA-binding of type II MADS TFs. The type I plant MADS TFs do not possess an I or
183 MEF2-domain and are more closely related to the SRF-type MADS TFs. However, the structure
184 of SRF (PDB 1HBX) reveals a similar alpha helical region C-terminal to the MADS-box
185 domain, albeit with a different orientation of helices from the MEF2/I domain (**Figure 1A**).
186 Protein sequence alignment of the 106 MADS TFs from Arabidopsis reveal that 104 of the 106
187 sequences possess ~15 amino acids C-terminal to the MADS domain that are structurally
188 conserved and predicted to form alpha helices with similar physicochemical properties (**Figure**
189 **2A**). In particular, positions including Arg69 and Tyr70 in SEP3 that are important for I domain
190 anchoring to the M domain are conserved in both type I and type II MADS, while a high degree
191 of sequence variation in the non-alpha helical region is observed (**Figure 2A and**
192 **Supplementary Figure 2**). This suggests that these ~15 amino acids in type I MADS TFs may
193 act as an I-like alpha helical domain and stabilize the structure of the DBD, influence DNA-
194 binding specificity or both.

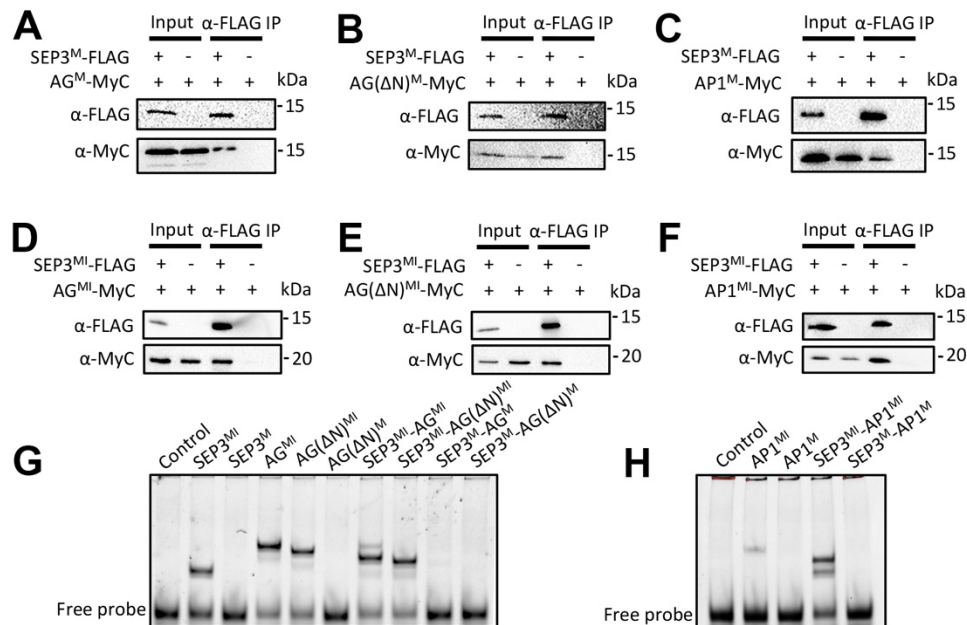


195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206

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) Pull-down 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) Pull-down assay showing that the M domain plus the I like region of AGL61 (AGL61^M) interacts with the MI region of PHE (PHE^M) and AGL80 (AGL80^M). (F) EMSA assay showing that heterodimers AGL61^M-AGL80^M and AGL61^M-PHE^M 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.

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

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



218
 219 **Figure 3. Type II MADS TFs require the I domain for DNA binding but not dimerization.** (A-C) Pull-down assay showing
 220 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
 221 deleted (AG(ΔN)^M) and AP1 (AP1^M), respectively. (D-F) Pull-down assay showing that the MI domain of SEP3 (SEP3^{MI})
 222 interacts with the MI domain of AG^{MI}, AG(ΔN)^{MI} and AP1^{MI}, respectively. (G and H) EMSA assay showing that homodimers
 223 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
 224 DNA sequence containing a canonical CArG-box binding site (as per Figure 2), while their corresponding constructs without
 225 the I domain can not, suggesting that the I domain in type II MADS TFs is required for DNA binding.

226 For the type II MADS TFs, we generated M domain and MI domain constructs for
 227 SEP3, AP1 and AG. In Arabidopsis, SEP3-AP1 containing MADS TF complexes define first
 228 and second whorl flower organs (sepal and petals) whereas SEP3-AG containing complexes
 229 are required for third and fourth whorl (stamen and carpel) flower organ identity. We focused
 230 on these MADS TFs as their heterocomplex formation, DNA-binding specificity and *in vivo*
 231 activity are well understood^{11,14,23,39-43}. As expected, the MI domain of SEP3 interacts with that
 232 of AP1 and AG (**Figure 3D-F**). However, in contrast to the type I MADS, the M domain of
 233 SEP3 is able to pull down the M domains of AP1 and AG (**Figure 3A and C**). As AG possesses
 234 an N-terminal 16 amino acid extension, constructs with and without these amino acids were
 235 tested. Both constructs were able to interact with the M domain of SEP3 (**Figure 3A and B**).
 236 Thus, unlike the type I MADS TFs tested, the M domain alone of the type II MADS TFs tested
 237 was sufficient for protein-protein interaction. However, the M domains alone of SEP3-AP1 and
 238 SEP3-AG were unable to bind a canonical CArG DNA motif in EMSA experiments (**Figure**
 239 **3G and H**), suggesting that the M domain of type II without the I domain is not sufficient to
 240 form a stable dimer for DNA binding. In contrast, homo and heterodimers of the MI domains

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

245

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

247 To further explore the role of the I domain in DNA-binding specificity, we focused on
248 the floral organ identity-specifying MADS TFs, SEP3, AP1, AG and a chimeric version of AG,
249 in which the AG I-domain was replaced with the AP1 I-domain (AG^{IAP1}). Using sequential
250 DNA-affinity purification followed by sequencing (seq-DAP-seq ref), we determined the
251 DNA-binding of the heteromeric complexes of SEP3-AG³² and SEP3-AG^{IAP1} (this study).
252 Unfortunately, the SEP3-AP1 heteromeric complex under the same experimental conditions
253 did not yield data of sufficient quality for analysis. However, SEP3-AG and SEP3-AG^{IAP1}
254 strongly bound DNA and each produced highly similar replicates for data analysis
255 (**Supplementary Fig. 3**). We used a highly stringent peak calling for each replicate and peaks
256 were kept only when present in all replicates. This yielded 6,347 peaks for SEP3-AG and 3,552
257 peaks for SEP3-AG^{IAP1}. These peak lists were merged and the binding intensities for the two
258 complexes were compared for all resulting peaks. This profiling indicates that while most peaks
259 have a similar binding intensity, several hundred exhibited at least a two-fold difference in
260 binding intensity between SEP3-AG and SEP3-AG^{IAP1} (**Figure 4A-B**). The same profiling
261 between replicas for a given dataset demonstrated no such variability in binding intensity
262 (**Supplementary Figure 3**).

263 The top 15% sequences (1,073 sequences for each complex) displaying the most
264 extreme binding difference were used to search for specific DNA patterns that potentially code
265 for SEP3-AG and SEP3-AG^{IAP1} differential response. The TF binding sites were modeled using
266 position weight matrices, transcription factor flexible models and k-mer set memory (KSM)
267 analysis. Due to the highly similar binding modes of the MADS family TFs, only KSM was
268 sensitive enough to differentiate the MADS heterocomplexes (**Figure 4C and D**)⁴⁴. For the
269 KSM, pipeline, KMAC and KSM tools from the GEM package ⁴⁵ were used to search for
270 clusters of short (4-20 bp) overlapping sequences (with allowances and penalties for gaps and
271 mismatches) that are over-represented in the 600 sequences bound best by either SEP3-AG or
272 SEP3-AG^{IAP1}. This analysis yielded 55 clusters for SEP3-AG and 22 clusters for SEP3-AG^{IAP1}
273 and served as kmer-based models for prediction analysis. With AUC of 0.79 for SEP3-AG and
274 0.84 for SEP3-AG^{IAP1}, the kmer-based models were highly predictive. Importantly, each kmer-

275 based model did not perform well (AUC \sim 0.7) in predicting bound regions from the other
276 heterocomplex (**Figure 4D**). This suggests that specific sequence patterns code for differential
277 binding between the two heterocomplexes.

278 Previously, using seq-DAP-seq, we showed that SEP3-AG demonstrates preferential
279 intersite spacing of \sim 47 and \sim 57 bp between two CArG boxes due to its cooperative binding, a
280 contributing factor for DNA-binding specificity of MADS TF complexes³². Using specific
281 bound regions of SEP3-AG^{IAP1} when compared with SEP3-AG, we found that SEP3-AG^{IAP1}
282 gained new preferential intersite spacings of 25 and 34 bp and lost the SEP3-AG preferential
283 intersite spacings (**Figure 4E**). This further suggests that the I domain plays a role in DNA-
284 binding specificity by affecting the conformation of the tetramer and selecting for specific
285 intersite distances.

286 In order to determine whether these observed *in vitro* differences in binding were
287 potentially relevant *in vivo*, we examined published ChIP-seq datasets. Previous studies have
288 shown that the *in vitro* binding of SEP3-AG relative to SEP3-AP1 in SELEX-seq correlates
289 with the *in vivo* binding intensity of AG relative to AP1 in ChIP-seq for the 1500 most enriched
290 SEP3 ChIP-seq peaks (**Figure 4F**)⁴⁴. This result suggested that specific sequence patterns
291 detected *in vitro* were able to at least weakly discriminate sequences either more likely to be
292 bound by SEP3-AG or by SEP3-AP1 in the Arabidopsis genome. Applying the same analysis
293 to our seq-DAP-seq data, we found a similar weak positive correlation between the SEP3-AG
294 binding affinity relative to that of SEP3-AG^{IAP1} and the corresponding ChIP-seq derived *in vivo*
295 binding, suggesting that SEP3-AG^{IAP1} behaves more similarly to SEP3-AP1 than to SEP3-AG
296 (**Figure 4F**). Although modest, this correlation is supported by the observation that the genes
297 corresponding to the 250 peaks most differentially bound by SEP3-AG relative to SEP3-AG^{IAP1}
298 exhibited a 7.6-fold enrichment for genes involved in “carpel development” GO term (FDR =
299 4.2×10^{-3}). However, no clear GO term enrichment was detected for genes corresponding to
300 the 250 regions best bound by SEP3-AG^{IAP1} relative to SEP3-AG. This may be due to the
301 relatively small number of genes corresponding to GO terms for “petal development” (4 genes)
302 and “sepal development” (13 genes), suggesting an incomplete list of genes for these GO terms.
303 Taken together, the binding differences between the SEP3-AG^{IAP1} and SEP3-AG complex
304 suggest that the AG^{IAP1} protein has lost some AG identity and gained AP1-like identity.

305

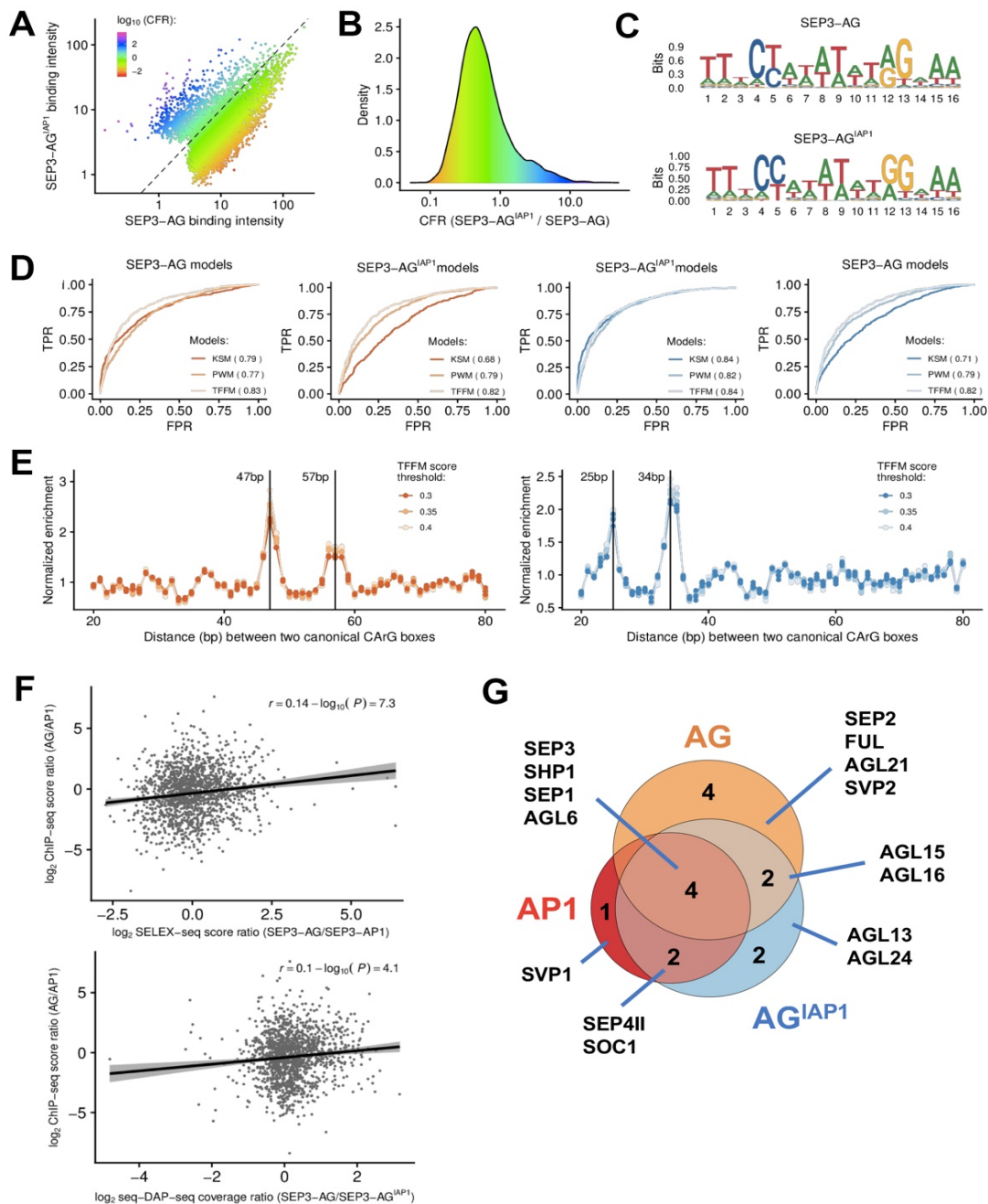


Figure 4. DNA-binding and protein interactions patterns. (A) Comparison of SEP3-AG and SEP3-AG^{IAP1} seq-DAP-seq binding intensity (log₁₀ of reads per kb per million of reads mapped in bound regions) and color-coded by purple-blue (SEP3-AG^{IAP1}-specific) to orange-red (SEP3-AG-specific) according to log₁₀ 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^{IAP1} specific regions. (F) Top, published SELEX-seq for SEP3-AP1 and SEP3-AG⁵ 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 SEP3-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^{IAP1} 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 MIKCC^C MADS TFs in Arabidopsis. Data shows that AG^{IAP1} loses AG interactors and gains AP1 interactors.

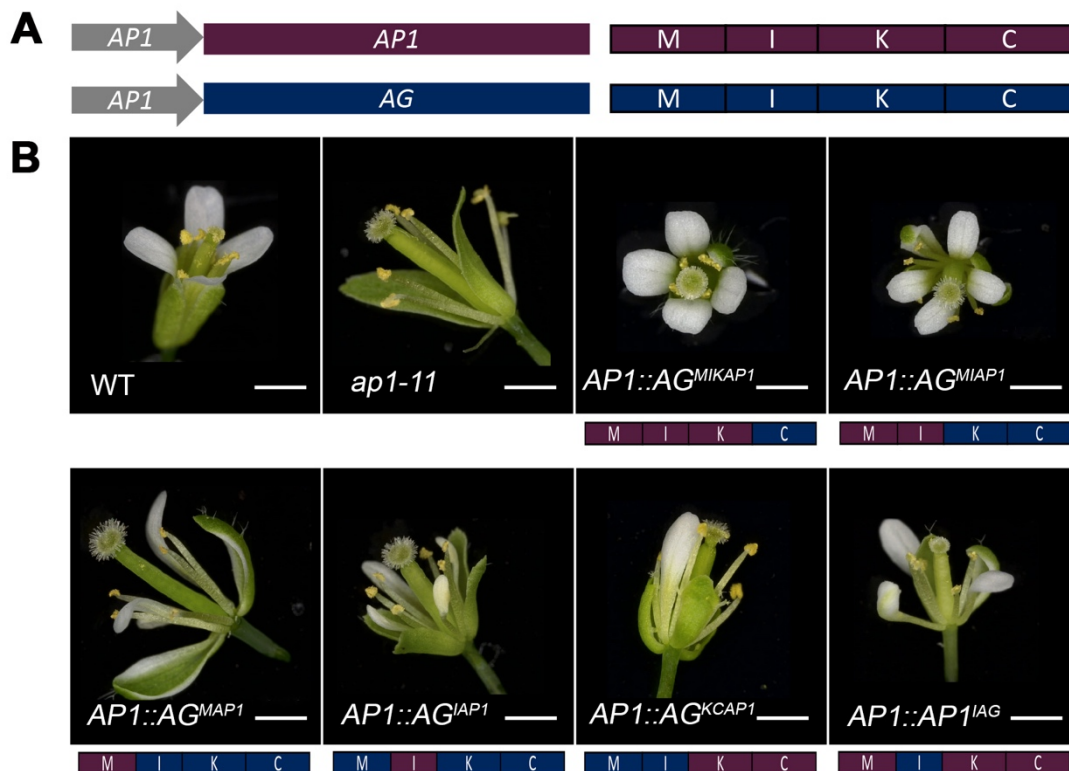
307 **Impact of the I domain on protein interaction specificity**

308 Next, we sought to understand to what extent the I domain could confer protein-protein
309 interaction specificity as it plays an important role in dimerization and dimer stability²⁸. AP1
310 and AG share SEP3 as a binding partner in the molecular model predicting flower organ
311 specification, but do exhibit differential heterodimerisation capabilities with other MIKC^c
312 MADS TFs^{14,23}. In order to probe the function of the I domain in dimerization specificity, a
313 matrix-based yeast two-hybrid screen for AP1, AG and AG^{IAP1} against all Arabidopsis type II
314 MIKC^c MADS TFs was performed. The targets were cloned into both bait and prey vectors and
315 only interactors for both replica and in all three selection media were scored as positive
316 interactions (**Supplementary Figure 4**). This may underestimate the true number of binding
317 partners, however, it reduces the number of false positives. Using this strict cut-off for protein-
318 protein interaction events, AP1 interacted with seven different partners, of which three were
319 specific for AP1 and four were in common with AG, including the SEP clade members SEP1
320 and SEP3. AG interacted with ten partners of which six were exclusive with respect to AP1
321 interacting partners. AG^{IAP1} interacted with ten partners, including two out of the three AP1-
322 specific interactions and lost four AG-specific partners (**Figure 4G**). Interestingly, both SOC1
323 and AGL24, known AP1 binding partners¹⁴ interacted with the AG^{IAP1} chimera but not with
324 AG. These data indicate that the I domain of AP1 contributes to heterodimerization specificity
325 of the chimera.

326

327 **In planta function of the I domain**

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



337
338
339
340
341
342

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 ($n \geq 10$). 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.

343 *AG^{MAP1}* expressing plants were poorly complemented with a reduced number of first
344 and second whorl organs and only partial organ identity recovery with petaloid and sepaloid
345 organs in 43% of the plants. *AG^{KCAP1}* also complemented poorly, with most of the plants
346 showing carpeloid sepals and with only partial recovery of 1 to 3 sepals in 60% of the plants.
347 In contrast, almost complete complementation was present in plants expressing *AG^{MIAP1}*, with
348 92% of the transformants showing 4 sepals or petaloid sepals in the first whorl and 2 to 4 petals
349 in the second whorl. This was similar to plants expressing *AG^{MIKAP1}*, in which first and second
350 whorl organ identity was restored for 69% of the plants with occasional reduced number of
351 organs in the second whorl and lateral flower formation. *AP1* plants with the I domain of AG
352 (*AP1^{IAG}*), showed a similar phenotype as *AG^{MAP1}*. Strikingly, *AG^{IAP1}* expressing plants exhibited
353 4 sepaloid organs and 3-4 petals or petaloid organs in over 40% of the plants, suggesting that
354 the I domain acts as the smallest unit conferring the most AP1 function to the *AG^{IAP1}* chimera
355 (**Supplementary Table 2**).

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

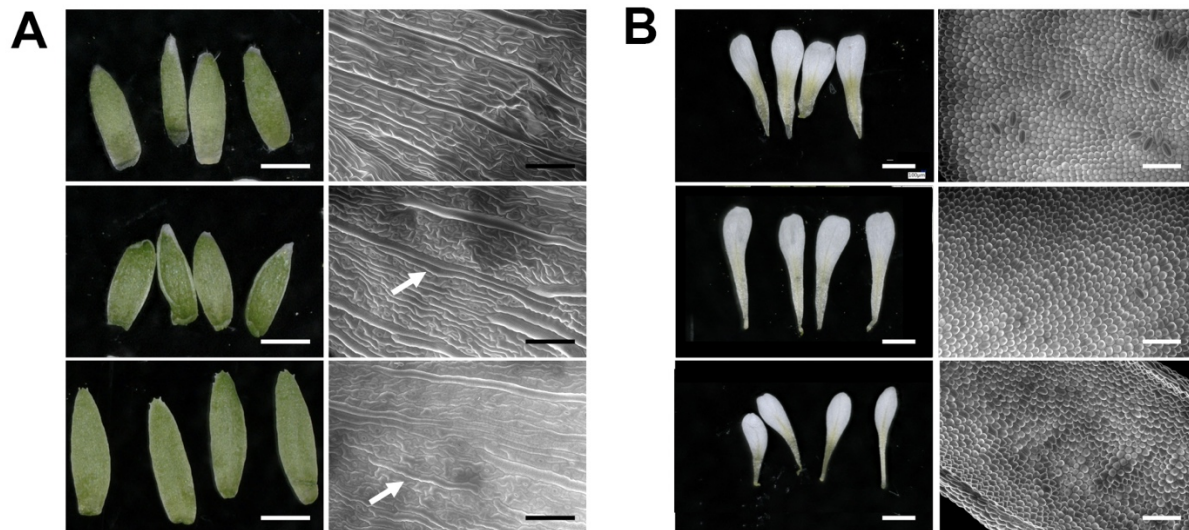
358 promoter in the strong *ap1-7* mutant background (**Figure 6 and Supplementary Figures 6-8**).
359 *AG^{LAP1}* *ap1-7* lines produced flowers very similar to WT (**Figure 6**).



Figure 6. *AP1:AG^{LAP1}* expression largely complements the *ap1-7* flower phenotype. *ap1-7* lines expressing either *AP1* (*AP1::AP1*), *AG* (*AP1::AG*) or *AG^{LAP1}* (*AP1::AG^{LAP1}*) 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^{LAP1}* 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^{LAP1}* 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.

360
361 In these plants, whorl 1 of each flower was made of four sepal-like organs and axillary
362 buds were never observed as opposed to the *ap1-7* mutant (**Figure 6 and 7**). In addition,
363 between 66% to 82% of the flowers in three independent lines homozygous for the transgene
364 showed 4 petal-like organs (**Figure 6 and 7**). Scanning electron microscopy (SEM)
365 observations of epidermal cells from whorls 1 and 2 of *AG^{LAP1}* expressing lines showed
366 elongated cells characteristic of sepals and conical cells characteristic of petals, respectively
367 (**Figure 7 and Supplementary Figure 7**). Detailed inspection of sepal and petal surfaces by
368 SEM from the *AG^{LAP1}* expressing plants revealed the presence of a few clusters of leaf-like and
369 stamen identity cells in whorls 1 and 2, respectively (**Figure 7 and Supplementary Figure 8**).
370 In these 3 lines, 10 to 25% of the flowers showed petals with only petal identity cells while 75
371 to 90% of the flowers show at least one petal with a small cluster of stamen identity cells
372 (**Supplementary Figure 8**). This indicates incomplete complementation of *AP1* function by
373 the chimera. However, *AP1:AG* was unable to even partially complement *AP1* function in the
374 first and second whorl. Inspection of the cell surface in *ap1-7 AP1:AG* plants of the first whorl
375 showed characteristic carpel cells (replum, style and papilla) and an absence of organs in the

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



380
381 **Figure 7. First and second whorls cell identity are complemented in $ap1-7$ plants expressing AG^{IAP1} .** First (A) and second
382 (B) whorl organs were removed from WT, and flowers from API or AG^{IAP1} expressing $ap1-7$ plants. First whorl organs in
383 AG^{IAP1} expressing plants are slightly longer compared to WT and API expressing plants (A, left panel). Second whorl organs
384 in AG^{IAP1} expressing plants are slightly smaller compared to WT and API expressing plants (B, left panel). Epidermal cell
385 identity, observed by SEM, show characteristic WT elongated sepal cells in API and AG^{IAP1} expressing plants (A, right panel)
386 and characteristic WT conical cells in petals from API and AG^{IAP1} expressing plants (B, right panel). A small number of
387 epidermal cells typical of leaves are also seen in AG^{IAP1} expressing plants and to a lesser extent in API expressing plants
388 (arrow). Scale bars indicate 100 μ M for SEM images and 1 mm for organ photographs.

389 Discussion

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

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

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

415 With the introduction of the floral quartet model, the role of the I domain in determining
416 MADS TF function specificity was supplanted by the importance of specific tetramer
417 components triggering DNA-looping as being required for gene regulation²⁴. Tetramerisation
418 allows cooperative two-site binding and looping of DNA, thus selecting for DNA-binding sites
419 that exhibit preferred spacing^{31,34,49,50}. Examination of a tetramerisation mutant of SEP3
420 revealed that fourth whorl carpel development and meristem determinacy require efficient
421 tetramerisation of SEP3³³. Recent studies comparing the genome-wide binding patterns of
422 tetrameric versus dimeric SEP3-AG complexes have further shown that tetramerisation both
423 increases binding affinity and plays a role in determining DNA-binding specificity via
424 preferential binding of specific intersite distances³². Comparisons of genome-wide binding of
425 SEP3-AG and SEP3-AG^{IAP1}, presented here, reveals that the I domain alters preferred intersite
426 spacing between CArG-type binding sites. These data suggest a role of the I domain as a
427 modulator of tetramerisation and highlight an additional mechanism mediated by the I domain
428 for changing DNA binding patterns of MADS TF complexes.

429 The recruitment of ternary factors, non-MADS family protein partners, has been
430 postulated to be a major determinant of DNA-binding specificity⁵¹. Based on yeast 2-hybrid
431 screening, mass spectrometry and pulldown assays, additional TF partners from the NF-Y and
432 homeodomain families as well as the orphan TF, LEAFY, have been identified for a few type
433 II MADS TFs including OsMADS18 from rice and SHATTERPROOF, SEEDSTICK, AG and
434 SEP3 from Arabidopsis^{25,48,52,53}. While ternary complex formation likely plays a role in
435 differential gene regulation, the formation of such complexes has only been shown for a
436 relatively small number of MADS TFs and how much it accounts for the diversity of function

437 in the MADS TF family requires further studies. As shown here, swapping the I domain of the
438 well-studied floral organ identity MADS TF, AP1 (first and second whorl organs) into the third
439 and fourth whorl specific TF, AG, results in strong complementation of the *apl* loss of function
440 phenotype by the chimeric construct. The experiments keep the K and C-terminal domains of
441 AG intact and these domains have been postulated to be important for neo-functionalization⁵⁴
442 and/or the recruitment of ternary factors⁵⁵. We cannot completely exclude that ternary factors
443 recruited via the I domain could be responsible for the *API*-like activity of *AG^{IAP1}* in planta.
444 For example, OsMADS18 uses the MI domains to recruit OsNF-YB1 based on yeast 2-hybrid
445 assays⁴⁸. However, it is more likely that the combination of tuning DNA-binding specificity
446 and heterodimer partner selection by the I domain accounts for the majority of AP1 activity of
447 the *AG^{IAP1}* chimera.

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

456 **Methods**

457

458 **SEP3 MI domain protein expression and purification**

459 SEP3 MI (aa 1-90) was cloned into the pETM41 vector using the NcoI-NotI restriction sites to
460 obtain His6-MBP translational protein fusion. Recombinant SEP3 MI protein was expressed in
461 *E. coli* BL 21 Codon Plus cells. Cells were grown at 37 °C to an OD₆₀₀ of 0.8-1 after which time
462 the temperature was reduced to 15 °C and protein expression induced by addition of 1 mM of
463 IPTG for 12 h. Cells were harvested by centrifugation and the cell pellet resuspended in 25 mM
464 KH₂PO₄ pH 7 buffer containing 10 % glycerol, 500 mM NaCl, 2 mM Tris(2-
465 carboxyethyl)phosphine (TCEP) and cOmplete protease inhibitors (Roche). Cells were lysed
466 by sonication and cell debris pelleted at 25,000 rpm for 40 min. The soluble fraction was applied
467 to a 1 ml Ni-NTA column and the protein eluted with resuspension buffer + 200 mM imidazole.
468 The protein was dialysed against 25 mM Tris, pH 8.0, 300 mM NaCl and 1 mM TCEP. The
469 protein was then applied to a heparin column to remove any bound DNA using a salt gradient
470 from 300 mM to 2 M NaCl. The protein was dialysed overnight in the presence of TEV protease
471 (10:1) to cleave the His-MBP tag at 4 °C. The protein was then passed over a Ni-NTA column
472 to deplete the His-MBP and any uncleaved protein. A second heparin column was run to remove
473 any His-MBP and to obtain pure SEP3 MI. The protein was concentrated to ~6 mg/ml and used
474 for crystallization trials.

475

476 **Protein crystallization, data collection and refinement**

477 SEP3 MI at a concentration of 6 mg/ml was mixed at a 1:1 ratio with potassium sodium tartrate
478 tetrahydrate (0.2 M), bis-tris propane (0.1 M, pH 7.5) and 20 % PEG 3350. The protein
479 crystallized over 20 d at 4 °C forming a diamond shaped single crystal. 15 % glycerol was used
480 as a cryoprotectant and the crystal flash frozen in liquid N₂. Diffraction data were collected at
481 100 K at the European Synchrotron Radiation Facility (ESRF), Grenoble France on ID29 at a
482 wavelength of 1.0 Å. Indexing was performed using MXCube⁵⁶ and the default optimized
483 oscillation range and collection parameters used for data collection. The data set was integrated
484 and scaled using the programs *XDS* and *XSCALE*⁵⁷. Data collection and refinement statistics
485 are given in Table 1. The structure is deposited under PDB code 7NB0.

486

487 **In vitro pull-down assays**

488 The following constructs from type I and type II MADS TFs are cloned into pTnT vector for in
489 vitro pulldown assays. Type I MADS TFs include AGL61^{MI}(63-122)-5MyC, AGL61^{MI}(63-
490 155)-5MyC, PHE^{MI}(1-60)-3FLAG, PHE^{MI}(1-95)-3FLAG, AGL80^{MI}(1-61)-3FLAG and
491 AGL80^{MI}(1-88)-3FLAG, and type II MADS TFs include SEP3^{MI}(1-57)-3FLAG, SEP3^{MI}(1-90)-
492 3FLAG, AG^{MI}(1-73)-5MyC, AG^{MI}(1-107)-5MyC, AG(ΔN)^{MI}(17-73)-5MyC, AG(ΔN)^{MI}(17-
493 107)-5MyC, AP1^{MI}(1-57)-5MyC and AP1^{MI}(1-92)-5MyC. Proteins were produced using in
494 vitro transcription and translation system using wheat germ extract following manufactures'
495 instructions (SP6 High Yield Expression System from Promega). Briefly, purified plasmids
496 were used as input in a 2 hr 25 °C incubation reaction. For single input reactions, 1 μg of
497 plasmid was used in a 25 μl reaction volume. For double input reactions, 1 μg plasmid of each
498 construct was used in a 25 μl reaction volume. 10 % of each input reaction (i.e. 2.5 μl) was
499 used as input for western blots. The rest of the reaction was used for pull-down experiments.
500 Briefly, each reaction was completed to 100 μl using PBS buffer (150 mM sodium phosphate,
501 pH 7.2, and 150 mM NaCl), and added with 10 μl of the appropriate magnetic beads (anti-MyC
502 beads, Thermo Fisher or anti-FLAG beads, Merck Millipore) and incubated for 1 hr at 4 °C.
503 The beads plus protein solution suspension was then placed on magnets, and the supernatant
504 was discarded. The beads were washed with PBS buffer four times, and SDS-PAGE loading
505 dye was added to the beads and boiled for 5 min. at 95 °C. Western blots were used to assess
506 the pulldown results using anti-MyC (Thermo Fisher) and anti-FLAG antibodies (Merck
507 Millipore).

508

509 **EMSA experiments**

510 Proteins for EMSAs experiments were produced as described above. EMSAs were performed
511 as described with 10 nM DNA labeled with Cy5 (Eurofins) using a 103-bp DNA fragment
512 containing 2 CARG boxes belonging to the *SEP3* promoter³³. For each EMSA, a negative
513 control was run, labeled 'DNA alone' in which the in vitro translation assay was done with
514 pTNT vector without any insert and incubated with the DNA probe.

515

516 **Yeast two-hybrid screening**

517 AP1, AG and the chimeric construct, AG^{IAP1}(residues 74-107 of AG replaced by residues 58-
518 92 of AP1) were cloned into pENTR/D-TOPO® (Kan) gateway entry vector. Subsequently,
519 Gateway LR reactions were performed using pDEST32 (pBDGAL4) and pDEST22
520 (pADGAL4) destination vectors. The resulting pDEST32-AG, pDEST32-AP1 or pDEST32-
521 AG^{IAP1} expression construct was transformed into yeast strain PJ69-4a, and pDEST22-AG,
522 pDEST22-AP1 or pDEST22-AG^{IAP1} into PJ69-4A, followed by an autoactivation screen for the
523 bait vector as described⁵⁸. Upon confirmation that the AG^{IAP1} bait did not possess any
524 autoactivation capacity, a matrix-based Y2H screening was performed following the protocol
525 described in⁵⁸. Bait and prey clones of all native Arabidopsis type II MIKC^c MADS TFs were
526 generated previously¹⁴ and were screened against AP1, AG and AG^{IAP1}. Growth of yeast and

527 hence, protein-protein interaction events, were scored after seven days of incubation at 20 °C
528 on selective medium. Three different selective media were used: SD lacking Leucine,
529 Tryptophan, and Histidine (-LWH) and supplemented with 1 mM 3-Amino-1,2,4-triazole (3-
530 AT); SD -LWH + 5 mM 3-AT and SD lacking L,W and Adenine (-LWA). Mating was
531 performed twice in independent experiments and only combinations of MADS TFs giving
532 growth on all three selection media and in both replica were scored as interaction events.
533

534 **Seq-DAP-seq**

535 Seq-DAP-seq for SEP3-AG^{IAP1} complex was performed as described previously³². Briefly, 2 µg
536 of each purified plasmid (pTnT-SEP3-3FLAG and pTnT- AG^{IAP1}-5Myc) were used as input in
537 a 50 µl TnT reaction incubated at 25 °C for 2 hr (Promega). The reaction solution was then
538 combined with 50 µl IP buffer (PBS supplemented with 0.005 % NP40 and proteinase inhibitors
539 (Roche)) and mixed with 20 µl anti-FLAG magnetic beads (Merck Millipore). Following 1 hr.
540 incubation at room temperature, the anti-FLAG magnetic beads were immobilized, and washed
541 three times with 100 µl IP buffer. Protein complexes were eluted with 100 µl IP buffer
542 supplemented with 200 µg/ml 3xFLAG peptide (Merck Millipore). The eluted protein was then
543 immobilized on anti-c-Myc magnetic beads (Thermo Fisher) and washed three times with 100
544 µl IP buffer to isolate homogeneous SEP3-AG^{IAP1} complex. The purified protein complex, while
545 still bound on anti-c-Myc magnetic beads, was incubated with 50 ng DAP-seq input library pre-
546 ligated with Illumina adaptor sequences. The reaction was incubated for 90 mins, and then
547 washed 6 times using 100 µl IP buffer. The bound DNA was heated to 98 °C for 10 min and
548 eluted in 30 µl EB buffer (10 mM Tris-Cl, pH 8.5). The eluted DNA fragments were PCR
549 amplified using Illumina TruSeq primers for 20 cycles, and purified by AMPure XP beads
550 (Beckman). The libraries were quantified by qPCR, pooled and sequenced on Illumina HiSeq
551 (Genewiz) with specification of paired-end sequencing of 150 cycles. Each library obtained 10
552 to 20 million reads. The seq-DAP-seq was performed in triplicate.
553

554 **Seq-DAP-seq data analysis**

555 *Reads processing and peak calling*

556 SEP3-AG and SEP3-AG^{IAP1} reads processing and peak calling was performed as previously
557 described³². Briefly, reads were checked using FastQC
558 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and adapter sequences removed
559 with NGmerge⁵⁹ and mapped with bowtie2⁶⁰ onto the TAIR10 version of the *A. thaliana*
560 genome (www.arabidopsis.org), devoid of the mitochondrial and the chloroplast genomes. The
561 duplicated reads were removed using the samtools rmdup program⁶¹. The resulting alignment
562 files for each sample were input to MACS2⁶² to call peaks using the input DNA as control.
563 Consensus peaks between replicates were defined using MSPC⁶³ (P-value cutoff = 10⁻⁴) for
564 each experiment. Each consensus peak was scanned for possible subpeaks, split into several
565 peaks if needed, and the peak widths were then re-sized to ± 200 bp at both side of the peak
566 maximal height. For all the resulting peaks, coverage was computed as the mean of the
567 normalized read coverage for each replicate. This normalized coverage defines the binding
568 intensity of a heterocomplex at a bound region.
569

570 *Comparison between SEP3-AG and SEP3-AG^{IAP1}*

571 SEP3-AG or SEP3-AG^{IAP1} were merged according to the following procedure: peaks were
572 considered common if at least 80 % of two peaks overlapped with < 50 % of either peak non-
573 overlapping. Peaks that did not overlap by > 50 % of their length were considered new peaks.
574 These values were chosen empirically based on visual inspection of the peaks in the Integrated
575 Genome Browser⁶⁴. The averaged normalized coverage from each experiment, divided by the
576 peak size, was computed for each peak. Figure 4 was computed using R

577 (<https://www.Rproject.org>) and the ggplot library⁶⁵. The coverage fold reduction (CFR) was
578 computed as the ratio between the mean normalized coverages in SEP3–AG^{IAP1} and SEP3–AG
579 seq-DAP-seq. The top 15 % sequences with extremes CFR were considered SEP3–AG^{IAP1} and
580 SEP3–AG specific and used to search for differential DNA patterns that potentially direct
581 transcription factor differential binding using sequence modelling (see next paragraph).
582 Detection of preferred spacings between canonical CA_nG boxes was performed as in ³².

583
584 *Position weight matrices (PWM), transcription factor flexible model (TFFM) and k-mer set*
585 *memory (KSM).*

586 For each experiment, PWM, TFFM and KSM models were reconstructed out of the 600 best
587 peaks (judged according to their averaged coverage). PWM were generated by the meme-suite,
588 using meme-chip⁶⁶ with options -meme-minw 16, -mememaxw 16, -meme-nmotifs 1 –meme-
589 pal. TFFM were generated using the TFFM-framework package and the PWM obtained in the
590 meme-chip output ⁶⁷. KSM is a recently developed TF binding motif representation that consist
591 of a set of aligned k-mers that are overrepresented at TF binding sites. KSM were generated
592 using the KMAC tool ⁴⁵ with options set to search k-mers, with size ranging from 4 to 20bp in
593 a 300bp sliding windows, that are enriched compare to TF unbound sequences. TF binding sites
594 were predicted by searching the PWM, TFFM and KSM models against TF bound sequences
595 and unbound sequences. PWM and TFFM scan was performed using in house scripts. KSM
596 were searched using the KSM tool ⁴⁵. The best TFFM/PWM scores and the sum of KSM scores
597 (because distinct KSM can hit to a same subject sequence) obtained for each bound regions and
598 for unbound regions are retained to assess the prediction power of each model in an AUROC
599 (Area Under the Receiver Operating Characteristic Curve). In this assessment, the unbound set
600 of regions are chosen with similar GC content, size, and origin (promoter, intron, exon and
601 intergenic) than the set of bound regions ⁶⁸. The set of unbound sequences fed to KMAC to
602 detect enriched k-mer is different than the one used to evaluate the KSM model prediction.

603

604 **Plant material and growth conditions**

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

611

612 **Plasmid constructs for plant transformation**

613

614 Constructs used for the primary transformant phenotypic analysis were cloned into a modified
615 version of pMDC32 binary vector⁶⁹, where the 2x35S promoter was replaced with the *API*
616 promoter. *AG^{MAP1}*, *AG^{MIAP1}*, *AG^{MIKAP1}*, *AG^{KCAP1}*, *AG^{CAP1}*, *API^{IAG}* and *AG^{IAP}* coding sequences
617 were amplified from pTNT-*AG^{MAP1}*, pTNT-*AG^{MIAP1}*, pTNT-*AG^{MIKAP1}*, pTNT-*AG^{KCAP1}*, pTNT-
618 *AG^{CAP1}*, pTNT-*API^{IAG}* and pTNT-*AG^{IAP}* vectors and subsequently inserted by XbaI restriction
619 enzyme downstream of the *API* promoter. pTNT-CDS vectors were generated through Gibson
620 assembly (NEB).

621

622 *API::API*, *API::AG* and *API::AG^{IAP1}* vectors were generated using Gibson assembly in the
623 vector backbone pFP100⁷⁰. *API* promoter and *OCT* terminator were PCR amplified from pML-
624 BART-AP1pro:AP1-AR plasmid⁷¹. *API*, *AG* and *AG^{IAP1}* coding sequences were amplified
625 from pTNT-*API*, pTNT-*AG*, pTNT- *AG^{IAP1}* vectors.

626

627 **Plant transformation and floral phenotype analysis**

628

629 All the plant transformations were performed using *apl* mutant plants following the floral dip
630 method⁷². For primary resistant analysis in *apl-11* intermediate allele, T1 seeds were sown on
631 the 0.5X MS medium supplemented with 20 µg/ml hygromycin. Resistant plants were
632 transferred to soil. Flowers arising from the primary shoot were analyzed by light microscopy.
633 For the detailed analysis of the *API::API*, *API::AG* and *API::AG^{API}*, the strong allele *apl-7*
634 was used as transformation background and T1 seeds were selected based on the seed specific
635 GFP selection marker⁷⁰. Flower phenotypic analyses were performed by light microscopy on
636 flower number ~10 to 19 based on their order of emergence on more than 20 T1 plants for each
637 construct, growing in parallel to *apl-7* plants. Three independent lines with one insertion
638 expressing *AG*, *API* and *AG^{API}* were selected for further phenotype analyses.

639

640

641 **Environmental scanning electron microscopy**

642

643 SEM experiments were performed at the Electron Microscopy facility of the ICMG Nanobio-
644 Chemistry Platform (Grenoble, France) as described³³.

645

646

647 **Table 1.**

648

649 **Data collection and refinement statistics**

650

	SEP3 ^{M1}
Data collection	
Space group	C222 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	67.4, 67.4, 122.6
α , β , γ (°)	90, 90, 90
Resolution (Å)	48.-2.10 (2.16-2.10) *
<i>R</i> _{sym} or <i>R</i> _{merge}	8.4 (130)
<i>CC</i> _{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
<i>R</i> _{work} / <i>R</i> _{free} (%)	20.4/23.6 (32.8/34.2)*
No. atoms	1880
Protein	1814
Ligand/ion	0
Water	66
<i>B</i> -factors	
Protein	55.5
Water	53.6
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.24

651

- refers to highest resolution shell

653

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

661

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

665

666

667 **References**

- 668 1 Passmore, S., Maine, G. T., Elble, R., Christ, C. & Tye, B. K. Saccharomyces
669 cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT
670 alpha cells. *J Mol Biol* **204**, 593-606, doi:10.1016/0022-2836(88)90358-0 (1988).
- 671 2 Yanofsky, M. F. *et al.* The protein encoded by the Arabidopsis homeotic gene
672 agamous resembles transcription factors. *Nature* **346**, 35-39, doi:10.1038/346035a0
673 (1990).

- 674 3 Sommer, H. *et al.* Deficiens, a homeotic gene involved in the control of flower
675 morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription
676 factors. *EMBO J* **9**, 605-613 (1990).
- 677 4 Norman, C., Runswick, M., Pollock, R. & Treisman, R. Isolation and properties of
678 cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum
679 response element. *Cell* **55**, 989-1003, doi:10.1016/0092-8674(88)90244-9 (1988).
- 680 5 Smaczniak, C., Angenent, G. C. & Kaufmann, K. SELEX-Seq: A Method to
681 Determine DNA Binding Specificities of Plant Transcription Factors. *Methods Mol*
682 *Biol* **1629**, 67-82, doi:10.1007/978-1-4939-7125-1_6 (2017).
- 683 6 Martinez-Castilla, L. P. & Alvarez-Buylla, E. R. Adaptive evolution in the
684 Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny.
685 *Proc Natl Acad Sci U S A* **100**, 13407-13412, doi:10.1073/pnas.1835864100 (2003).
- 686 7 Alvarez-Buylla, E. R. *et al.* An ancestral MADS-box gene duplication occurred before
687 the divergence of plants and animals. *Proc Natl Acad Sci U S A* **97**, 5328-5333,
688 doi:97/10/5328 [pii] (2000).
- 689 8 Pollock, R. & Treisman, R. Human SRF-related proteins: DNA-binding properties and
690 potential regulatory targets. *Genes Dev* **5**, 2327-2341 (1991).
- 691 9 Huang, H., Mizukami, Y., Hu, Y. & Ma, H. Isolation and characterization of the
692 binding sequences for the product of the Arabidopsis floral homeotic gene
693 AGAMOUS. *Nucleic Acids Res* **21**, 4769-4776 (1993).
- 694 10 Wynne, J. & Treisman, R. SRF and MCM1 have related but distinct DNA binding
695 specificities. *Nucleic Acids Res* **20**, 3297-3303, doi:10.1093/nar/20.13.3297 (1992).
- 696 11 Shiraishi, H., Okada, K. & Shimura, Y. Nucleotide sequences recognized by the
697 AGAMOUS MADS domain of Arabidopsis thaliana in vitro. *Plant J* **4**, 385-398
698 (1993).
- 699 12 Pellegrini, L., Tan, S. & Richmond, T. J. Structure of serum response factor core
700 bound to DNA. *Nature* **376**, 490-498, doi:10.1038/376490a0 (1995).
- 701 13 De Bodt, S. *et al.* Genomewide structural annotation and evolutionary analysis of the
702 type I MADS-box genes in plants. *J Mol Evol* **56**, 573-586, doi:10.1007/s00239-002-
703 2426-x (2003).
- 704 14 de Folter, S. *et al.* Comprehensive interaction map of the Arabidopsis MADS Box
705 transcription factors. *Plant Cell* **17**, 1424-1433, doi:tpc.105.031831 (2005).
- 706 15 Masiero, S., Colombo, L., Grini, P. E., Schnittger, A. & Kater, M. M. The emerging
707 importance of type I MADS box transcription factors for plant reproduction. *Plant*
708 *Cell* **23**, 865-872, doi:10.1105/tpc.110.081737 (2011).
- 709 16 Bemmer, M., Heijmans, K., Airoidi, C., Davies, B. & Angenent, G. C. An atlas of type I
710 MADS box gene expression during female gametophyte and seed development in
711 Arabidopsis. *Plant Physiol* **154**, 287-300, doi:pp.110.160770 (2010).
- 712 17 Saedler, H., Becker, A., Winter, K. U., Kirchner, C. & Theissen, G. MADS-box genes
713 are involved in floral development and evolution. *Acta Biochim Pol* **48**, 351-358
714 (2001).
- 715 18 Becker, A. & Theissen, G. The major clades of MADS-box genes and their role in the
716 development and evolution of flowering plants. *Mol Phylogenet Evol* **29**, 464-489,
717 doi:S1055790303002070 [pii] (2003).
- 718 19 Kaufmann, K., Melzer, R. & Theissen, G. MIKC-type MADS-domain proteins:
719 structural modularity, protein interactions and network evolution in land plants. *Gene*
720 **347**, 183-198, doi:S0378-1119(04)00762-0 (2005).
- 721 20 Kwantes, M., Liebsch, D. & Verelst, W. How MIKC* MADS-box genes originated
722 and evidence for their conserved function throughout the evolution of vascular plant
723 gametophytes. *Mol Biol Evol* **29**, 293-302, doi:10.1093/molbev/msr200 (2012).

- 724 21 Adamczyk, B. J. & Fernandez, D. E. MIKC* MADS domain heterodimers are
725 required for pollen maturation and tube growth in Arabidopsis. *Plant Physiol* **149**,
726 1713-1723, doi:10.1104/pp.109.135806 (2009).
- 727 22 Verelst, W., Saedler, H., Munster, T. MIKC* MADS-Protein complexes Bind Motifs
728 Enriched in the Proximal Region of Pollen-Specific Arabidopsis Promoters. *Plant*
729 *Physiol* **143**, 447-460, doi:<https://doi.org/10.1104/pp.106.089805> (2007).
- 730 23 Immink, R. G. *et al.* SEPALLATA3: the 'glue' for MADS box transcription factor
731 complex formation. *Genome Biol* **10**, R24, doi:gb-2009-10-2-r24 (2009).
- 732 24 Theissen, G. & Saedler, H. Plant biology. Floral quartets. *Nature* **409**, 469-471,
733 doi:10.1038/35054172 (2001).
- 734 25 Smaczniak, C. *et al.* Characterization of MADS-domain transcription factor
735 complexes in Arabidopsis flower development. *Proc Natl Acad Sci U S A* **109**, 1560-
736 1565, doi:10.1073/pnas.1112871109 (2012).
- 737 26 Gramzow, L. & Theissen, G. A hitchhiker's guide to the MADS world of plants.
738 *Genome Biol* **11**, 214, doi:gb-2010-11-6-214 (2010).
- 739 27 Smaczniak, C., Immink, R. G., Angenent, G. C. & Kaufmann, K. Developmental and
740 evolutionary diversity of plant MADS-domain factors: insights from recent studies.
741 *Development* **139**, 3081-3098, doi:10.1242/dev.074674 (2012).
- 742 28 Riechmann, J. L., Krizek, B. A. & Meyerowitz, E. M. Dimerization specificity of
743 Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3,
744 PISTILLATA, and AGAMOUS. *Proc Natl Acad Sci U S A* **93**, 4793-4798,
745 doi:10.1073/pnas.93.10.4793 (1996).
- 746 29 Krizek, B. A. & Meyerowitz, E. M. Mapping the protein regions responsible for the
747 functional specificities of the Arabidopsis MADS domain organ-identity proteins.
748 *Proc Natl Acad Sci U S A* **93**, 4063-4070, doi:10.1073/pnas.93.9.4063 (1996).
- 749 30 Riechmann, J. L. & Meyerowitz, E. M. Determination of floral organ identity by
750 Arabidopsis MADS domain homeotic proteins AP1, AP3, PI, and AG is independent
751 of their DNA-binding specificity. *Mol Biol Cell* **8**, 1243-1259,
752 doi:10.1091/mbc.8.7.1243 (1997).
- 753 31 Jetha, K., Theissen, G. & Melzer, R. Arabidopsis SEPALLATA proteins differ in
754 cooperative DNA-binding during the formation of floral quartet-like complexes.
755 *Nucleic Acids Res* **42**, 10927-10942, doi:10.1093/nar/gku755 (2014).
- 756 32 Lai, X., Stigliani, A., Lucas, J., Hugouvieux, V., Parcy, F., Zubieta, C. Genome-wide
757 binidng of SEPALLATA3 and AGAMOUS complexes dertermined by sequential
758 DNA-affinity purification sequencing. *Nucleic Acids Res* **48**, 9637-9648,
759 doi:10.1093/nar/gkaa729 (2020).
- 760 33 Hugouvieux, V. *et al.* Tetramerization of MADS family transcription factors
761 SEPALLATA3 and AGAMOUS is required for floral meristem determinacy in
762 Arabidopsis. *Nucleic Acids Res* **46**, 4966-4977, doi:10.1093/nar/gky205 (2018).
- 763 34 Puranik, S. *et al.* Structural basis for the oligomerization of the MADS domain
764 transcription factor SEPALLATA3 in Arabidopsis. *Plant Cell* **26**, 3603-3615,
765 doi:10.1105/tpc.114.127910 (2014).
- 766 35 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline
767 state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).
- 768 36 Crooks, G. E., Hon, G., Chandonia, J-M., Brenner, S.E. WebLogo:a sequence logo
769 generator. *Genome Res* **14**, doi:10.1101/gr.849004 (2004).
- 770 37 Bemer, M., Wolters-Arts, M., Grossniklaus, U. & Angenent, G. C. The MADS
771 domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central
772 cell in Arabidopsis ovules. *Plant Cell* **20**, 2088-2101, doi:tpc.108.058958 (2008).

- 773 38 Steffen, J. G., Kang, I.-H., Portereiko, M. F., Lloyd, A. & Drews, G. N. AGL61
774 Interacts with AGL80 and Is Required for Central Cell Development in Arabidopsis.
775 *Plant Physiology* **148**, 259-268 (2008).
- 776 39 Pelaz, S., Gustafson-Brown, C., Kohalmi, S. E., Crosby, W. L. & Yanofsky, M. F.
777 APETALA1 and SEPALLATA3 interact to promote flower development. *Plant J* **26**,
778 385-394 (2001).
- 779 40 Kaufmann, K. *et al.* Target genes of the MADS transcription factor SEPALLATA3:
780 integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS*
781 *Biol* **7**, e1000090, doi:08-PLBI-RA-4384 (2009).
- 782 41 O'Maoileidigh, D. S. *et al.* Control of reproductive floral organ identity specification
783 in Arabidopsis by the C function regulator AGAMOUS. *Plant Cell* **25**, 2482-2503,
784 doi:10.1105/tpc.113.113209 (2013).
- 785 42 Mizukami, Y., Huang, H., Tudor, M., Hu, Y. & Ma, H. Functional domains of the
786 floral regulator AGAMOUS: characterization of the DNA binding domain and
787 analysis of dominant negative mutations. *Plant Cell* **8**, 831-845,
788 doi:10.1105/tpc.8.5.831 (1996).
- 789 43 Riechmann, J. L., Wang, M. & Meyerowitz, E. M. DNA-binding properties of
790 Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3,
791 PISTILLATA and AGAMOUS. *Nucleic Acids Res* **24**, 3134-3141, doi:6t0273 [pii]
792 (1996).
- 793 44 Smaczniak, C., Muino, J. M., Chen, D., Angenent, G. C. & Kaufmann, K. Differences
794 in DNA Binding Specificity of Floral Homeotic Protein Complexes Predict Organ-
795 Specific Target Genes. *Plant Cell* **29**, 1822-1835, doi:10.1105/tpc.17.00145 (2017).
- 796 45 Guo, Y., Tian, K., Zeng, H., Gifford, D.K. A novel k-mer set memory (KSM) motif
797 representation improves regulatory variant prediction. *Genome Res* **28**, 891-900,
798 doi:10.1101/gr.226852.117 (2018).
- 799 46 Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E.M., Smyth, D.R. Control of
800 Flower development in Arabidopsis thaliana by APETALA1 and interacting genes.
801 *Development* **119**, 721-743 (1993).
- 802 47 Irish, V. F. & Sussex, I. M. Function of the apetala-1 gene during Arabidopsis floral
803 development. *Plant Cell* **2**, 741-753, doi:10.1105/tpc.2.8.741 (1990).
- 804 48 Masiero, S. *et al.* Ternary complex formation between MADS-box transcription
805 factors and the histone fold protein NF-YB. *J Biol Chem* **277**, 26429-26435,
806 doi:10.1074/jbc.M202546200 (2002).
- 807 49 Mendes, M. A. *et al.* MADS domain transcription factors mediate short-range DNA
808 looping that is essential for target gene expression in Arabidopsis. *Plant Cell* **25**,
809 2560-2572, doi:10.1105/tpc.112.108688 (2013).
- 810 50 Melzer, R., Verelst, W. & Theissen, G. The class E floral homeotic protein
811 SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro.
812 *Nucleic Acids Res* **37**, 144-157, doi:10.1093/nar/gkn900 (2009).
- 813 51 Davies, B., Sommer, H., Schwarz-Sommer, Z. *Flower Development: Genetic Views*
814 *and Molecular News*. 167-184 (Springer Science and Business Media, 1999).
- 815 52 Brambilla, V. *et al.* Genetic and molecular interactions between BELL1 and MADS
816 box factors support ovule development in Arabidopsis. *Plant Cell* **19**, 2544-2556,
817 doi:10.1105/tpc.107.051797 (2007).
- 818 53 Liu, C., Xi, W., Shen, L., Tan, C. & Yu, H. Regulation of floral patterning by
819 flowering time genes. *Developmental cell* **16**, 711-722,
820 doi:10.1016/j.devcel.2009.03.011 (2009).

- 821 54 Vandebussche, M., Theissen, G., Van de Peer, Y. & Gerats, T. Structural
822 diversification and neo-functionalization during floral MADS-box gene evolution by
823 C-terminal frameshift mutations. *Nucleic Acids Res* **31**, 4401-4409 (2003).
- 824 55 Egea-Cortines, M., Saedler, H. & Sommer, H. Ternary complex formation between
825 the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in
826 the control of floral architecture in *Antirrhinum majus*. *EMBO J* **18**, 5370-5379,
827 doi:10.1093/emboj/18.19.5370 (1999).
- 828 56 Gabadinho, J. *et al.* MxCuBE: a synchrotron beamline control environment
829 customized for macromolecular crystallography experiments. *J Synchrotron Radiat*
830 **17**, 700-707, doi:10.1107/S0909049510020005 (2010).
- 831 57 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta*
832 *Crystallogr D Biol Crystallogr* **66**, 133-144, doi:10.1107/S0907444909047374 (2010).
- 833 58 de Folter, S. a. I., R.G.H. in *Plant Transcription Factors. Methods in Molecular*
834 *Biology (Methods and Protocols)* Vol. 754 (ed L. Yuan., Perry, S.) 145-165 (Humana
835 Press, 2011).
- 836 59 Gaspar, J. M. NGmerge: merging paired-end reads via novel empirically-derived
837 models of sequencing errors. *BMC Bioinformatics* **19**, 536, doi:10.1186/s12859-018-
838 2579-2 (2018).
- 839 60 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
840 *Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 841 61 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
842 Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup. The
843 Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079,
844 doi:10.1093/bioinformatics/btp352 (2009).
- 845 62 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137,
846 doi:10.1186/gb-2008-9-9-r137 (2008).
- 847 63 Jalili, V., Matteucci, M., Masseroli, M. & Morelli, M. J. Using combined evidence
848 from replicates to evaluate ChIP-seq peaks. *Bioinformatics* **31**, 2761-2769,
849 doi:10.1093/bioinformatics/btv293 (2015).
- 850 64 Freese, N. H., Norris, D. C. & Loraine, A. E. Integrated genome browser: visual
851 analytics platform for genomics. *Bioinformatics* **32**, 2089-2095,
852 doi:10.1093/bioinformatics/btw069 (2016).
- 853 65 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag 2016).
- 854 66 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic*
855 *Acids Res* **37**, W202-208, doi:10.1093/nar/gkp335 (2009).
- 856 67 Mathelier, A. & Wasserman, W. W. The next generation of transcription factor
857 binding site prediction. *PLoS Comput Biol* **9**, e1003214,
858 doi:10.1371/journal.pcbi.1003214 (2013).
- 859 68 Stigliani, A. *et al.* Capturing Auxin Response Factors Syntax Using DNA Binding
860 Models. *Mol Plant* **12**, 822-832, doi:10.1016/j.molp.2018.09.010 (2019).
- 861 69 Curtis, M. D. G., U. A Gateway Cloning Vector Set for High-throughput Functional
862 Analysis of Genes in *Planta*. *Plant Physiol* **133**, 462-469 (2003).
- 863 70 Bensmihen, S. *et al.* Analysis of an activated ABI5 allele using a new selection
864 method for transgenic *Arabidopsis* seeds. *FEBS Lett* **561**, 127-131,
865 doi:10.1016/S0014-5793(04)00148-6 (2004).
- 866 71 O'Maoileidigh, D. S., Thomson, B., Raganelli, A., Wuest, S.E., Ryan, P.T.,
867 Kwasniewska, K., Carles, C.C., Graciet, E., Wellmer, F. Gene network analysis of
868 *Arabidopsis thaliana* flower development through dynamic gene perturbations. *The*
869 *Plant Journal* **83**, 344-358 (2015).

870 72 Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-
871 mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**, 735-743 (1998).
872
873