Computational and Structural Biotechnology Journal 17 (2019) 946–953

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Structural Basis for Plant MADS Transcription Factor Oligomerization

Xuelei Lai *, Hussein Daher, Antonin Galien, Veronique Hugouvieux, Chloe Zubieta *

Laboratoire de Physiologie Cellulaire et Végétale, CNRS, Univ. Grenoble Alpes, CEA, INRA, IRIG, Grenoble, France

ARTICLE INFO

Article history:
Received 20 March 2019
Received in revised form 6 June 2019
Accepted 11 June 2019
Available online 14 June 2019

Keywords:
MADS transcription factors
Oligomerization
Floral development
Arabidopsis

ABSTRACT

MADS transcription factors (TFs) are DNA binding proteins found in almost all eukaryotes that play essential roles in diverse biological processes. While present in animals and fungi as a small TF family, the family has dramatically expanded in plants over the course of evolution, with the model flowering plant, *Arabidopsis thaliana*, possessing over 100 type I and type II MADS TFs. All MADS TFs contain a core and highly conserved DNA binding domain called the MADS or M domain. Plant MADS TFs have diversified this domain with plant-specific auxiliary domains. Plant type I MADS TFs have a highly diverse and largely unstructured Carboxy-terminal (C domain), whereas type II MADS have added oligomerization domains, called Intervening (I domain) and Keratin-like (K domain), in addition to the C domain. In this mini review, we describe the overall structure of the type II "MIKC" type MADS TFs in plants, with a focus on the K domain, a critical oligomerization module. We summarize the determining factors for oligomerization and provide mechanistic insights on how secondary structural elements are required for oligomerization capability and specificity. Using MADS TFs that are involved in flower organ specification as an example, we provide case studies and homology modeling of MADS TFs complex formation. Finally, we highlight outstanding questions in the field.
This is an open access article under the CC BY license. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Contents

1. Introduction . . . 946
2. Structure of MIKC MADS TFs . . . 948
   2.1. The Highly Conserved MADS Domain . . . 948
   2.2. The Flexible and Variable Intervening Domain . . . 948
   2.3. The K Domain, a Key Oligomerization Module . . . 948
      2.3.1. The K Domain Is a Leucine Zipper . . . 950
      2.3.2. K Domain Tetramerization *Versus* Dimerization . . . 950
   2.4. Carboxy-Terminal C Domain . . . 951
3. Conclusions and Perspectives . . . 951
Acknowledgement . . . 951
References . . . 951

1. Introduction

MADS transcription factors (TFs) are found in almost all eukaryotes, playing important roles from pheromone sensing in yeast, to muscle development and cell proliferation and differentiation in animals [1] and to floral development in plants [2]. The core MADS DNA-binding domain (DBD) likely evolved from a domain of topoisomerase IIA, an enzyme involved in DNA maintenance [3]. MADS TFs are characterized by this canonical and evolutionarily conserved DBD. The MADS family draws its name from the members, MAINTENANCE OF MINICHROMOSOME1 (MCM, *S. cerivisae*), AGAMOUS (AG, *A. thaliana*), DEFICIENS (DEF *A. majus*) and Serum Response Factor (SRF, *H. sapiens*) [2,4–10]. While present in animals and fungi as a small TF family, plants have dramatically expanded the MADS family from a few MADS TFs in basal phyla such as charophytes (green algae) to over a hundred different members in Arabidopsis [3,11]. In higher plants, MADS TFs act as key regulators in diverse developmental processes including seed germination [12–15], vegetative growth [16,17], transition to flowering [18,19], flower development [9,20,21] and senescence [22,23]. How plant MADS TFs achieve

* Corresponding authors.
E-mail addresses: xuelei.lai@cea.fr (X. Lai), chloe.zubieta@cea.fr (C. Zubieta).

https://doi.org/10.1016/j.csbj.2019.06.014
2001-0370/This is an open access article under the CC BY license. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

this functional diversity is a fundamental question we are now starting to answer [24–26].

All MADS TFs bind DNA as obligate dimers, recognizing a "CArG" box motif (CC[A/T]$_6$GG) [27]. Allowing a single mismatch, this motif is represented over 340,000 times in the Arabidopsis genome and it has been hypothesized that MADS TFs could potentially control up to 30% of genes in Arabidopsis [28]. The complex gene regulation directed by the MADS TFs in plants requires highly specific targeting of the TF to different target genes. In order to achieve this, the canonical MADS DNA binding domain (DBD) is elaborated with additional domains. The plant MADS TF family can be divided into two distinct types, type I and type II [29]. Type I MADS TFs possess the core MADS domain (M domain) and an extended and highly variable Carboxy-terminal domain. Apart from the M domain, type I plant MADS TFs do not have

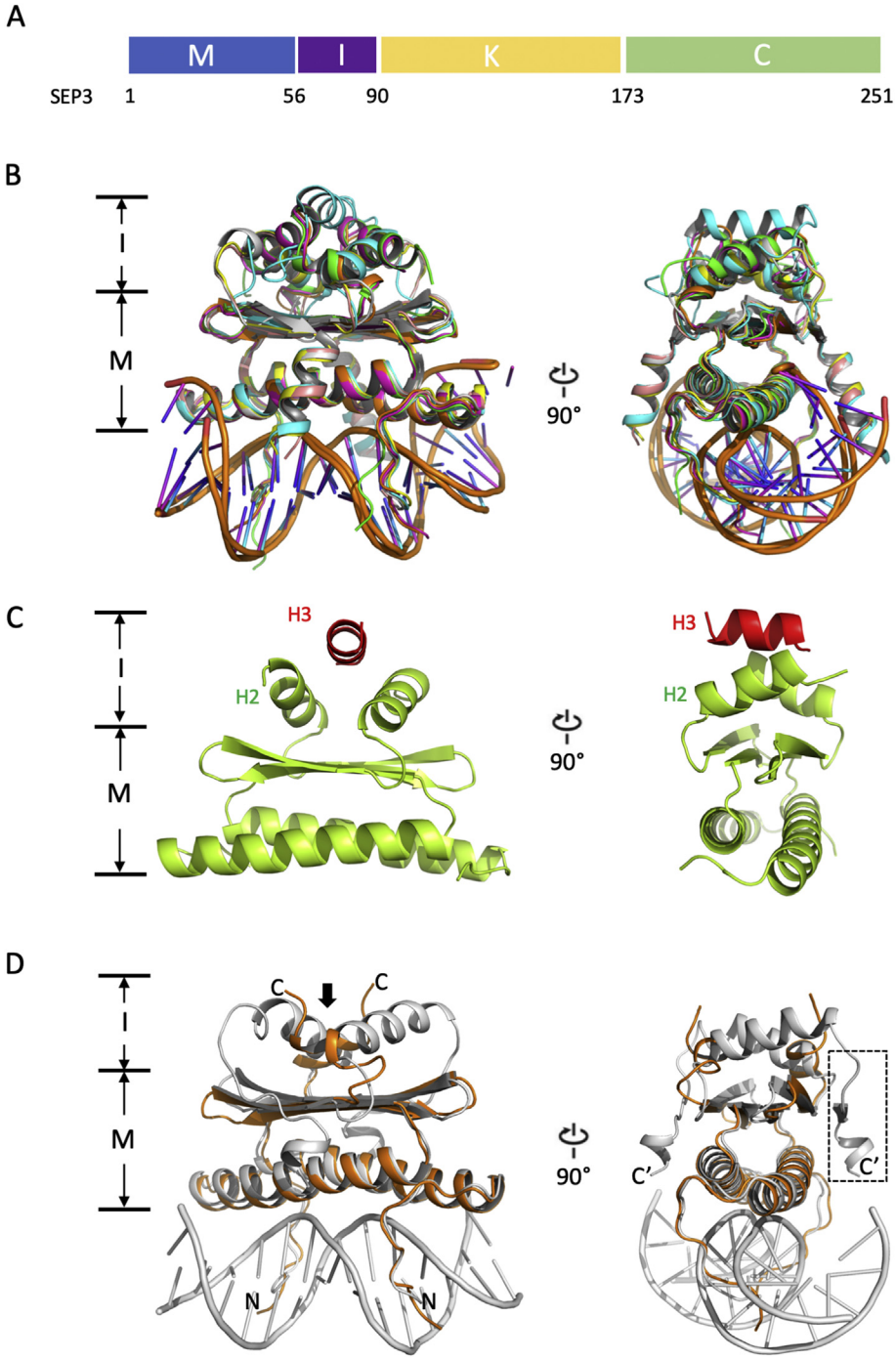

Fig. 1. 3D model superposition of M domains and I domain extensions of MADS family proteins from yeast and human with plant MI domains modeled. (A) Schematic representation of domain compositions of MIKC type MADS TF using SEP3 amino acid numbering as an example. (B) The coordinates were taken from yeast MCM1 (green, PDB 1MNM; residues 18–91) [47], human SRF (magenta, PDB 1SRS, residues 132–223) [46], human myocyte enhancer factor 2A (MEF2A) (yellow, PDB 3KOV, residues 2–95) [45], human MEF2B (cyan, PDB 1N6J, residues 2–58) [54], human MEF2 chimera (pink, PDB 6BYY, residues MEF2A (1–64 and 91–95) and MEF2B (65–90)) [55]. Two SEP3 models (SEP3-M1 in orange (residues 1–72) and SEP3-M2 in gray (residues 2–89)) were generated from SWISS-MODEL [56], using PDB 1SRS and 6BYY as templates, respectively. (C) Crystal structure of human MEF2B (PDB 6C9L) showing the presence of α-helix, H3 (highlighted in red), on top of H2 helices; Note that H3 in this conformation is only present in one dimer (formed by chain E and F) of the three dimers in the asymmetric unit. (D) Superposition of SEP3 M + I domain models showing different conformation possibilities of the I domain (indicated by a bold arrow). M, M domain; I, I domain; N, N-terminus; C, C-terminus of SEP3-M1; C′, C-terminus of SEP3-M2 (part of I domain is indicated in dash box).

additional conserved and predicted folded domains and will not be further discussed in this review. In contrast, the type II MADS TFs are characterized by a conserved four-domain structure, termed "MIKC" that includes the M domain, the I domain (Intervening domain, a short helical domain involved in dimerization specificity), the K domain (Keratin-like domain, a predicted coiled-coil important for tetramerization), and the C domain (Carboxy-terminal domain, which is a sequence variable and largely unstructured domain involved in transactivation and/or recruitment of other factors) (Fig. 1A). The MIKC MADS are further divided into $MIKC^C$ and $MIKC^*$ based on phylogeny and sequence conservation but retain this modular domain structure [30]. The I and the K domains were acquired during evolution and functionally diversify the conserved MADS DBD by facilitating oligomerisation with other MADS partners [31]. Adding these domains resulted in the development of a large protein–protein interaction (PPI) network consisting of homo- and hetero- dimers and tetramers. The tetrameric MADS complexes formed primarily by $MIKC^C$, are plant specific and can bind to DNA cooperatively at two sites [32,33], unlike other eukaryotic MADS TFs that only bind as dimers to a single CArG box [34]. For example, in Arabidopsis, the SEPALLATA (SEP) clade functions as a hub within the MADS PPI network and drives the formation of distinct tetrameric complexes [35,36]. Each complex is sufficient to determine a specific floral organ fate [2,6,20,37]. Extensive studies of the MADS PPI network using yeast "N"-hybrid experiments have revealed complex interaction patterns at both the level of the dimer as well as the tetramer [38,39].

In this review, we describe the overall architecture and structure of Type II MIKC MADS TFs, with a focus on the oligomerization domain based on the recent structural characterization of the SEP3 K domain and extensive mutagenesis studies [40–42]. We further address the structural determinants of K domain oligomerization, key structural motifs and K domain evolution.

2. Structure of MIKC MADS TFs

2.1. The Highly Conserved MADS Domain

The MIKC MADS TFs are modular multi domain proteins with highly conserved secondary structures (Fig. 1A). The DBD, or M domain, consists of ~55–60 residues and is highly conserved from animals to plants (with sequence identity as high as 40% between *H. sapiens* and *A. thaliana*). In addition to binding DNA, most plant M domains contain a predicted bipartite nuclear localization sequence (NLS) with the consensus sequence $(K/R)(K/R)X_{10}(K/R)_{3-5}$ [43,44]. The M domain has been structurally characterized in human and yeast and reveals a conserved fold and highly conserved DNA-binding residues (Fig. 1B) [45–47]. Domain swap experiments in Arabidopsis underscore how highly conserved the M domain is over the course of evolution. These experiments have shown that *APETALA1 (AP1)*, *APETALA3 (AP3)*, *PISTILLATA (PI)* and *AGAMOUS (AG)* chimeras possessing the M domain of *SRF* (yeast) or *MEF2* (human) are able to complement the appropriate mutant phenotype as well as the native gene, and can reproduce the phenotypes exhibited by plants expressing *35S::AP1*, *35S::AP3*, *35S::PI* and *35S::AG*, respectively [48]. These data suggest that either 1) the M domain is dispensable for DNA-binding specificity and ternary factors are required for *in vivo* binding specificity, 2) other MADS TF partners in the plant MADS heteromeric complexes may compensate for the switch in M domain of a single partner and/or 3) the oligomerization domains (*i.e.*, I and K domains) strongly contribute to DNA-binding specificity, perhaps *via* allosteric interactions. It is likely that a combination of these factors plays a part in determining DNA-binding specificity and function. For example, *in vivo* specificity studies using ChIP-seq demonstrated enrichment of non-CArG box motifs for MADS TF binding sites, suggesting a role for ternary factors [49]. Domain swap experiments described below for the I domain have been shown to alter DNA-binding specificity. *In vitro* and *in vivo* studies of the K domain have also demonstrated a role in DNA-binding specificity and *in vivo* function for the plant MIKC MADS [50].

2.2. The Flexible and Variable Intervening Domain

The I domain, a small (~25–30 amino acid) linker domain between M and K domains (Fig. 1A), is involved in dimerization specificity based on *in vitro* domain swaps, band shift assays, computational models and yeast 2-hybrid studies [51,52]. This domain exhibits greater sequence diversity than the M or K domains within the plant MIKC family. Based on secondary structure predictions, the I domain is largely α-helical. While the I domain does not exhibit high sequence homology to mammalian MADS, this domain may be structurally homologous its mammalian counterpart.

In the mammalian MADS box containing proteins, such as MEF2B, adjacent to the M domain is a small domain, consisting of two α-helices, called H2 and H3, that helps to stabilize the M domain by sandwiching the β sheet adjacent to the DNA-binding α-helix (Fig. 1C). The presence of flexible loops between H2 and H3 α-helices allows different conformations for the C-terminal α-helix, H3, as has been shown by a recent crystal structure in which one C-terminal MEF2 H3 is flipped up to stack on top of the H2 helices, a novel conformation not see in previous structural studies (Fig. 1C) [53]. This conformation, in which the H3 helix occupies the same protein-binding region as cofactors such as Cabin1, is observed for only one dimer out of three in the asymmetric unit. In addition, the second H3 helix from the partner monomer is completely missing from the structure due to either disorder or possibly *in situ* proteolysis. While the H2 helices form a protein-protein interface for a single helix to bind, this interface is not sufficient to bind two H3 helices and is unlikely to mimic plant MIKC dimers which would present two helices. The different conformations for MEF2 are likely the result of truncated constructs and crystal packing due to inherent flexibility between H2 and H3 and may have limited physiological relevance.

Based on human and yeast structures of MADS TFs, different homology models for plant MI domains can be generated. Using the 3D structure of human MADS TFs (PDB 1SRS and 6BYY), we generated two M domain (with partial I domain) models for SEP3, called SEP3-M1 and SEP3-M2, respectively (Fig. 1D). In both cases, the H2 helix of the I domain is present but with different conformations (indicated by bold arrow in Fig. 1D); for the SEP3-M2 model, a second helix (H3) with a flexible linker is modeled (in dashed box, Fig. 1D). Depending on the conformation and extension direction of the I domain, two distinct structural modes can be modeled that could potentially affect DNA binding through direct DNA contacts *via* the C-terminal extensions of the I domain [40]. Whether the I domain of plant MIKC TFs adopts one of these MEF2 conformations remains to be determined as no MI domain structures from plants are currently available.

2.3. The K Domain, a Key Oligomerization Module

MIKC type MADS genes have been reported even in the charophytes, extant green algae that are the closest relatives to the land plants [57]. This suggests that the MIKC type MADS genes evolved by the addition of a K domain to an ancestral MADS gene present in the most recent common ancestor of charophytes and chlorophytes which contained a type II MEF2 MADS gene with an M domain but lacked the IKC domains [29,31,58]. In seed plants, it has been described that the K domain sequences of MIKC type MADS TFs are highly conserved even among sequences sampled from high evolutionary distance [42], suggesting that the overall structure of the K domain is retained among MIKC type MADS TFs over the course of evolution. Thus, the K domain is a highly conserved and defining feature of the MIKC plant MADS TFs.

Based on secondary structure predictions, the K domain was divided into three α-helices, K1, K2 and K3. Extensive mutagenesis studies determined that hydrophobic residues in the K domain were of particular importance for mediating protein-protein interactions and critical for

dimerization and tetramerization of MADS TFs involved in floral organ identity. Leucine-zipper motifs in the K1 (position 87–108, PI numbering) and K2 (position 121–135, PI numbering) helices, for example, were shown to be important for AP3/PI heterodimerization, based on extensive mutagenesis studies and yeast 2-hybrid assays [59,60]. Further mutagenesis studies and yeast 3-hybrid assays incorporating the additional MADS partner, SEP3, helped identify residues in the K2 and K3 (position 142–167, PI) helices contributing to tetramerization [59]. These studies provided an important basis for predicting dimer and tetramer interactions. The subsequent 3D crystal structure of the K domain from SEP3 demonstrates that the residues identified through previous mutagenesis studies are either directly contacting the dimer/tetramer interface or are required for maintaining the structural integrity of the K domain helices [40].

The crystal structure of the K domain from SEP3 was the first and only structurally characterized oligomerization module of a plant MADS TFs available to date (Fig. 2B) [40]. The crystal structure (residues 75–178) contains the complete K domain (residues 91–173), a small portion of the I domain (residues 75–90) and several residues of the C domain (residues 174–178). Each monomer folds into two amphipathic

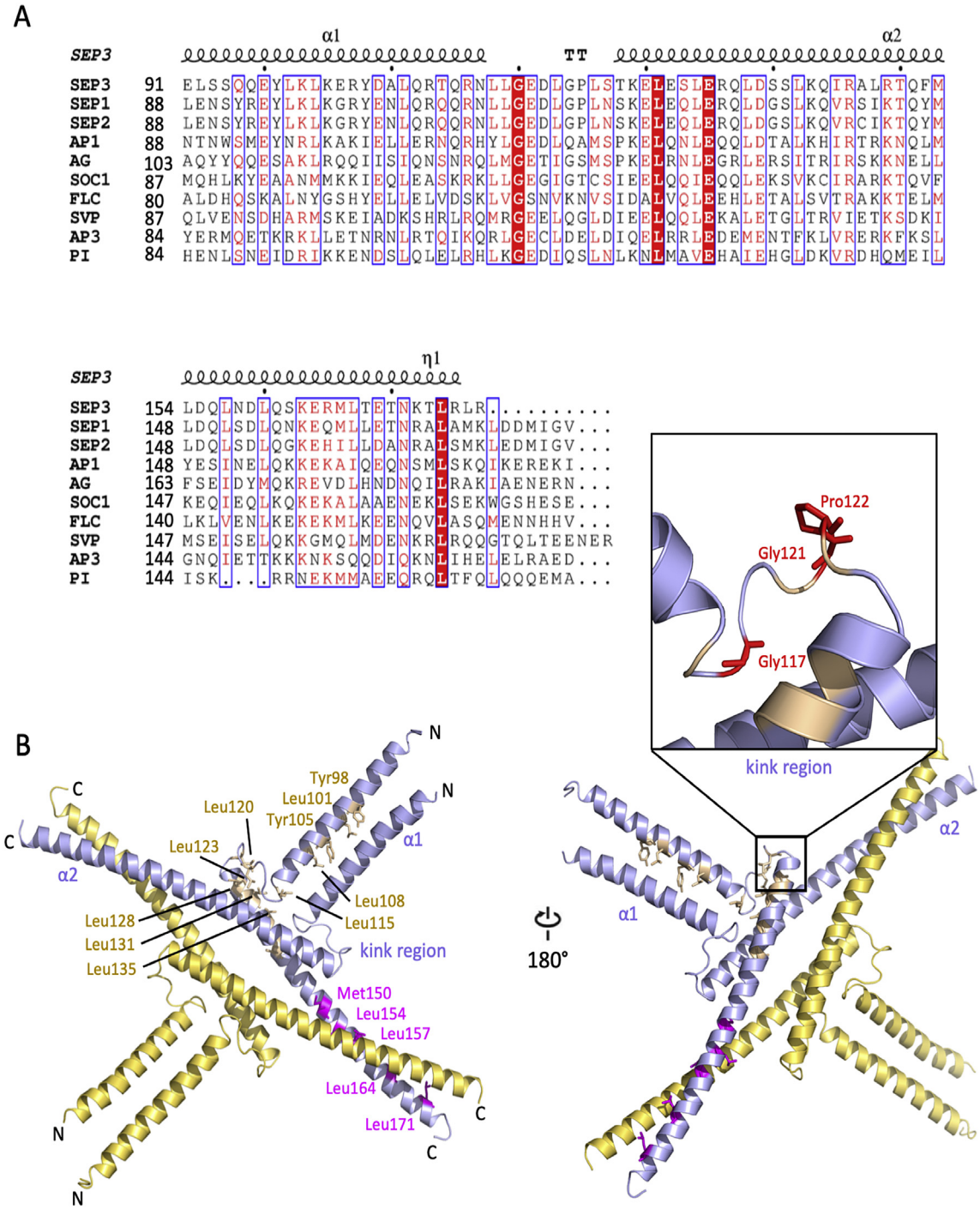

Fig. 2. K domain sequence alignment of some representative MADS TFs from Arabidopsis and SEP3 K domain structure. (A) K domain sequence alignment of SEP3, SEP1, SEP2, AP1, AG, SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), FLOWERING LOCUS C (FLC), SHORT VEGETATIVE PHASE (SVP), AP3 and PI, using ESPript 3 [65]. TT indicates the kink region that connects α1 and α2. (B) Crystal structure of SEP3 K domain (PDB 4OX0). The residues involved in dimer and tetramer formation are indicated and colored by wheat and magenta, respectively. The insert shows key residues in the kink region colored red.

α-helices, α1 (residues 75–114) and α2 (residues 115–178) separated by a glycine and/or proline rich spacer (referred to as the "kink region") of approximately 5–10 amino acids (Fig. 2B). Two monomers interact *via* their N-terminal α-helices to form a dimer and two dimers form a tetramer *via* their C-terminal α-helices. Thus, the biological assembly of SEP3 in the crystal structure is a homotetramer or a dimer of dimers. Sequence alignments of representative type II MADS TFs from Arabidopsis reveal that the amino acids in the kink region are variable but contain glycines and/or proline residues, resulting in a "break" between the α1 and α2 helices (Fig. 2A). This break is characterized by either added flexibility due to the presence of glycine residues, which are able to sample more conformational space, or proline residues which break the α–helical secondary structure and help force the two helices apart in a more rigid conformation. Thus, the α1 and α2 helices of the K domain are able to form two distinct coiled coils, allowing dimerization and tetramerization. The coiled coil motif is a well-established secondary structure element that can be found in many different protein families and is often involved in protein–protein interactions and/or DNA-binding [61–64]. Plants have thus co-opted a versatile domain to expand their repertoire of protein-protein interactions.

2.3.1. The K Domain Is a Leucine Zipper

The K domain of SEP3 has a characteristic leucine zipper heptad repeat pattern, $(abcdefg)_n$, in which the 'a' and 'd' positions are occupied preferentially by leucines but may comprise other hydrophobic residues (*e.g.* isoleucine and methionine). The leucine zipper is typical for coiled coils, which can be well predicted by bioinformatic algorithms, such as MultiCoil [66]. Using MultiCoil, we predicted the coiled coil formation probability of MADS TFs that are well studied in flower organ development in Arabidopsis, including SEP3, AG, AP3, PI and AP1 (Fig. 3A). These TFs are known to form the so-called floral "quartets", in which specific heterotetrameric complexes formed by different MADS TFs specify the identity of the floral organs in each of the four whorls of a flower (Fig. 3B) [6,20]. Despite highly conserved primary sequences in the K domain (Fig. 2A), these MADS TFs show highly divergent probability to form coiled coils (due in large part to differences in key leucine residues in critical positions of the heptad repeats). SEP3, a well characterized MADS TF that acts as a hub for flower quartet formation [36], shows the highest probability to form coiled coils, consistent with its known stable coiled coil structure and promiscuity in forming complexes with other MADS TFs (Fig. 2A). This is followed by AP1 and PI (medium probability), and AG and AP3 (low probability) (Fig. 3A). Interestingly, PI coiled coil probability is high for the α1 helix involved in dimerization but low for the α2 helix important for tetramerization. Indeed, PI is not a hub MADS TF and seems to require additional partners such as SEP3 in order to form tetramers. At least one monomer in each potential quartet must have a strong coiled coil formation propensity in the α2 helix in order to drive tetramerization (Fig. 3B). Thus, the ability to form a stable coiled coil structural motif, at least in part, contributes to MADS TF complex formation. This modeling approach is potentially useful to analyze and explain why certain MADS TF complexes are able to form and are more stable/favored than others.

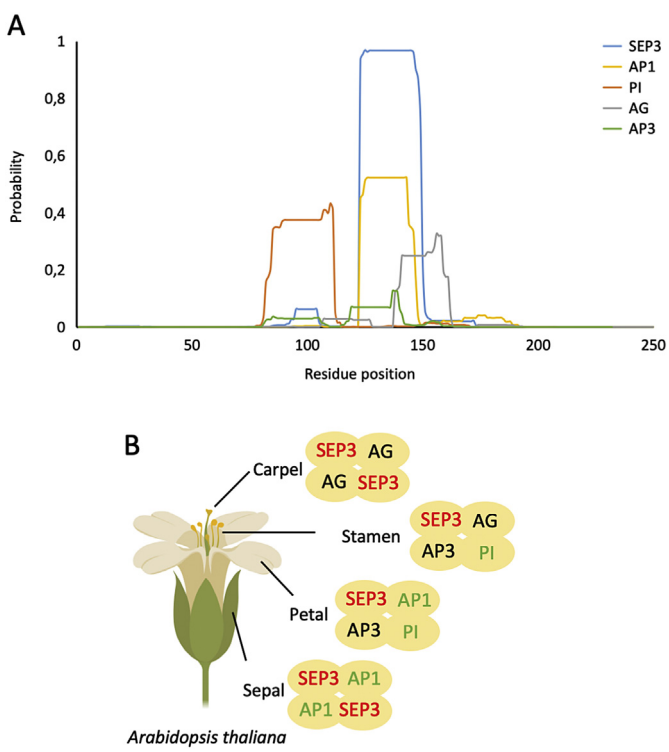

Fig. 3. Coiled coil prediction of MADS TFs involved in flower organ specification in Arabidopsis. (A) The coiled coil probability was predicted by MultiCoil [66] using the full-length protein sequence of SEP3, AP1, PI, AG and AP3, respectively, as input. The prediction was set for coiled coil dimer formation with a prediction window of 21 residues. The peak regions correspond to the K domains of the corresponding sequences. (B) Schematic representation of flower quartet model in organ specification. SEP3 labels colored in red indicates strong probability of coiled coil formation, AP1 and PI in green medium, and AP3 and AG in black weak.

Based on the positions of hydrophobic residues in the heptad repeats of the K domain, recent studies have experimentally assessed the effects of these residues by means of mutagenesis coupled with size exclusion chromatography and electrophoretic mobility shift assays (EMSAs) [25,40,42]. Strikingly, even a single point mutation in SEP3 is sufficient to completely abrogate dimer/tetramer formation. For example, tetramerization was abrogated with alanine or valine mutations at positions L164 [42], L154 or L131, and dimerization abrogated with the leucine to arginine mutation, L115R [41]. In addition, full-length SEP3 with mutations to residues important for tetramerization was dramatically affected in its DNA binding cooperativity-instead of binding to two CArG boxes on a single DNA probe as a tetramer, the mutated proteins only bound to one site as a dimer. In contrast, increasing the stability of the tetramer through addition of leucine residues at the "a" position of the heptad repeat at positions 161 and 168 resulted in increased cooperativity [42]. Furthermore, a chimeric SEP3 construct consisting of part of the AP3 α2 helix mutated to preserve leucine residues at the "a" positions of the heptad repeat was able to homotetramerize and cooperatively bind DNA, unlike the AP3 wild type protein that is unable to homotetramerize [42]. The effect of tetramerization has also been tested *in planta* using a splice variant of SEP3 that lacks residues 161–174 in the α2 helix, which showed abrogated tetramerization activity *in vitro*. The transgenic plant expressing the tetramerization mutant of SEP3 in the *sep1sep2sep3* triple mutant background showed defects in flower meristem termination [50]. Taken together, these studies suggest that subtle changes in the MADS TF oligomerization domain affects their DNA-binding activity both *in vitro* and *in vivo* with accompanying changes in regulation of target genes.

2.3.2. K Domain Tetramerization Versus Dimerization

A fundamental question remains as to what determines whether a MADS TF dimer will form an intramolecular coiled coil or recruit a second MADS TF dimer *via* the K domain to form a tetrameric complex. Based on extensive EMSAs studies and small-angle x-ray scattering experiments [41], it is becoming clear that only certain MADS TFs are able to act as drivers for tetramerization and that many MADS exist primarily as dimers. Based on sequence alignments, coiled coil predictions and homology modeling, certain MADS TFs likely act as tetramerization determinants due to specific residues between helices α1 and α2 which either favor or disfavor recruitment of a second MADS dimer for tetramer formation (Fig. 2B). SEP3 is able to form tetramers *in vitro* and *in vivo* and is one of the most promiscuous complex-forming MADS TFs [42]. Hub proteins, such as SEP3, act as central connecting nodes with many interaction partners in protein-protein interaction networks. Tetramerization capability may be a factor contributing to

the characteristics of a hub protein for SEP3 and other MADS TFs [67]. Ancestral reconstructions of MADS genes reveals that SEP3 has retained this hub characteristic and even lost interactions over the course of evolution [68]. The SEP clade has highly conserved proline and glycine residues between α1 and α2 helices, which may favor an open conformation of the α1 and α2 helices and thus encourage tetramer formation with other MADS partners. SOC1, SVP, FUL and AP1 are additional examples of proteins that may be able to form tetramers in the MADS TF network and all have at least two glycine residues or a glycine and proline residue between the α1 and α2 helices [38,39,69]. The conservation of flexible amino acid motifs between the α1 and α2 helices may correlate with the propensity for inter *versus* intra molecular coiled coils to form (*i.e.* tetramerization), although this remains to be experimentally confirmed.

2.4. Carboxy-Terminal C Domain

The C-terminal domain of plant MIKC TFs has a variable primary sequence with few conserved structural motifs based on secondary structure prediction. However, some short amino acid motifs are present which likely play roles in target gene regulation and/or higher order complex formation. For example, the EAR motif (ethylene-responsive element binding factor-associated amphiphilic repression motif) with a consensus sequence LxLxL or DLNxxP is recognized by the *co*-repressors TOPLESS and SIN3-ASSOCIATED POLYPEPTIDE P 18 (SAP18). This motif is present in over 20 MADS TFs from Arabidopsis [70]. A direct interaction dependent on the EAR motif between the MADS TF, AGAMOUS-LIKE 15, and the corepressor, SAP18, has been demonstrated, suggesting that this interaction is important for the recruitment of a histone deacetylase complex [71]. Direct interactions with SEUSS, a corepressor that forms a complex with LEUNIG, have been shown to require the C domain of SEP3 and AP1 [72]. Interestingly, transactivation activity has also been linked to the C-terminal domain for MADS TFs such as AP1 and SEP3 [73–75]. The presence of acidic, glutamine-rich and/or proline-rich clusters are often present in transcriptional activators and these types of clusters are present in both the AP1 and SEP3 C domains. These studies highlight the versatility of the C domain in mediating both activation and repression of target genes depending on the nature of the MADS TF complex and ternary factors.

For certain MADS TFs, the C domain does not appear to be required for function. For example, *AP3* or *PI* lacking the entire C domain were able to rescue *ap3* and *pi* mutant plants, respectively [76]. This further highlights the lack of conserved function of the C-terminal region. Higher order complex formation or stabilization of tetrameric complexes is another described role of the C domain [77]. Whether this is generally true in the MADS TF family requires further investigation.

3. Conclusions and Perspectives

The plant MADS TFs possess the highly conserved DBD that has exhibited little change over the course of evolution. A ubiquitous structural motif- the coiled coil- has been acquired by the MIKC plant MADS TFS that alters the oligomerization state and interaction patterns of the protein family, helping to explain the rich regulatory and functional diversity of these TFs. The K domain relies on leucine zipper-type interactions to alter complex formation in a myriad of ways *via* subtle changes in the energetics and kinetics of coiled coil formation. Initial formation of K domain α-helices for each monomer may depend on stabilizing hydrogen-bonding and salt bridge interactions, as has been experimentally shown for some coiled-coil forming proteins by NMR, time-resolved CD spectroscopy and mutagenesis [78]. These studies have not been performed to date for a K domain or K domain peptide from plant MADS TFs, but based on available structural data (SEP3 K domain, PDB 4OX0) residues on the hydrophilic face of the amphipathic K domain α-helices form a hydrogen-bonding network (R162/Q158/N155 and D152/Q148). In addition, the amino acids at these positions vary, suggesting a possible mechanism for controlling α-helix folding and stability. Additional amino acid substitutions on the hydrophobic face in each monomer α-helix in turn will stabilize or destabilize the formation of a dimeric coiled coil. Based on sequence alignments the "a" and "d" positions in MIKC MADS TFs vary, with leucine residues at the "d" position most strongly contributing to free energy stabilization.

Alternative splicing (AS) is another factor that increases plant MADS protein diversity by affecting primarily the oligomerization domains. The MIKC MADS proteins arise from multi-exonic genes, which undergo extensive AS resulting in alterations in the I and K domains that affects the PPI network and thus function. A number of predicted AS sites for the MADS family have been annotated, and are likely to be functionally relevant based on their location in the I/K domains and cross-species conservation [79]. Indeed, numerous transcripts corresponding to AS genes have been identified and physiological function of some of these isoforms has been determined. For example, in addition to the canonical transcript, SVP1, the SVP gene produces a shorter isoform, SVP3, that exhibits loss of interaction partners [52]. Overexpression of SVP1 and SVP3 in Arabidopsis result in different phenotypes, highlighting the function of AS in modifying PPI network and MADS activity [79]. Further examples include FLOWERING LOCUS M (FLM), in which different FLM isoforms are produced in a temperature dependent manner resulting in strong or highly attenuated DNA-binding when in complex with SVP [80]. Recent work from our team on SEP3 isoforms has also shown that AS changes the oligomerization capacity of the protein and results in changes in meristem determinacy and carpel development [50]. As more and more splice variants are being identified in RNA-seq datasets, it is becoming clear that the diversity of the MADS PPI network is even more complex than previously thought and changes in MADS function due to AS remains a rich subject to be explored.

Key questions remain as to MADS TF function and specificity. How do the different oligomerization patterns affect DNA binding and are there allosteric interactions which may further tune the direct interactions between the MADS DNA-binding domain and its cognate DNA? Is heterodimer or heterotetramer formation sufficient to functionally diversify the MADS TFs or are ternary factors necessary? How important is MADS tetramer cooperativity for gene regulation? Are additional oligomerization states accessible through extended coiled coil formation? Answers to these questions remain elusive, but the availability of structural, computational, modeling and genome wide binding data provide the necessary foundation for addressing these outstanding questions in the field.

Acknowledgement

This work was supported by the Agence Nationale de la Recherche (project FloPiNet, ANR-16-CE92-0023) (CZ, VH, HD), Grenoble Alliance for Integrated Structural and Cell Biology (GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003)) (XL, VH, CZ) and CFR CEA thesis fellowship (AG).

References

[1] Shore P, Sharrocks AD. The MADS-box family of transcription factors. Eur J Biochem 1995;229:1–13. https://doi.org/10.1111/j.1432-1033.1995.0001l.x.
[2] Melzer R, Wang Y-Q, Theißen G. The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. Semin Cell Dev Biol 2010;21:118–28. https://doi.org/10.1016/j.semcdb.2009.11.015.
[3] Gramzow L, Ritz MS, Theißen G. On the origin of MADS-domain transcription factors. Trends Genet 2010;26:149–53. https://doi.org/10.1016/j.tig.2010.01.004.
[4] Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM. The emerging importance of type I MADS box transcription factors for plant reproduction. Plant Cell 2011;23:865–72. https://doi.org/10.1105/tpc.110.081737.
[5] Litt A, Kramer EM. The ABC model and the diversification of floral organ identity. Semin Cell Dev Biol 2010;21:129–37. https://doi.org/10.1016/j.semcdb.2009.11.019.
[6] Thomson B, Wellmer F. Molecular regulation of flower development. Curr Top Dev Biol 2019;131:185–210. https://doi.org/10.1016/bs.ctdb.2018.11.007 Elsevier.
[7] Yan W, Chen D, Kaufmann K. Molecular mechanisms of floral organ specification by MADS domain proteins. Curr Opin Plant Biol 2016;29:154–62. https://doi.org/10.1016/j.pbi.2015.12.004.

[8] Dinneny JR, Yanofsky MF. Floral development: an ABC gene chips in downstream. Curr Biol 2004;14:R840–1. https://doi.org/10.1016/j.cub.2004.09.037.
[9] Callens C, Tucker MR, Zhang D, Wilson ZA. Dissecting the role of MADS-box genes in monocot floral development and diversity. J Exp Bot 2018;69:2435–59. https://doi.org/10.1093/jxb/ery086.
[10] Sommer H, Beltrán JP, Huijser P, Pape H, Lönnig WE, Saedler H, et al. Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors. EMBO J 1990;9:605–13. https://doi.org/10.1002/j.1460-2075.1990.tb08152.x.
[11] Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, et al. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci 2000;97:5328–33. https://doi.org/10.1073/pnas.97.10.5328.
[12] Victoria BM, Claudia B, Cecilia D, Mauricio H-C. MADS-box genes expressed during tomato seed and fruit development. Plant Mol Biol 2003;52:15.
[13] Ehlers K, Bhide AS, Tekleyohans DG, Wittkop B, Snowdon RJ, Becker A. The MADS Box Genes ABS, SHP1, and SHP2 are essential for the coordination of cell divisions in ovule and seed coat development and for endosperm formation in Arabidopsis thaliana. PLOS ONE 2016;11:e0165075. https://doi.org/10.1371/journal.pone.0165075.
[14] Fan C-M, Wang X, Wang Y-W, Hu R-B, Zhang X-M, Chen J-X, et al. Genome-wide expression analysis of soybean MADS genes showing potential function in the seed development. PLoS ONE 2013;8:e62288. https://doi.org/10.1371/journal.pone.0062288.
[15] Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, et al. The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development: the role of STK and ABS in ovule and seed development. Plant J 2012;70:409–20. https://doi.org/10.1111/j.1365-313X.2011.04878.x.
[16] Carmona MJ, Ortega N, Garcia-Maroto F. Isolation and molecular characterization of a new vegetative MADS-box gene from Solanum tuberosum L. Planta 1998;207:181–8. https://doi.org/10.1007/s004250050471.
[17] Rosin FM. Suppression of a vegetative MADS box gene of potato activates axillary meristem development. Plant Physiol 2003;131:1613–22. https://doi.org/10.1104/pp.102.012500.
[18] Immink RGH. A petunia MADS box gene involved in the transition from vegetative to reproductive development. Development 1999;126:10.
[19] Yu H, Goh CJ. Identification and characterization of three orchid MADS-box genes of the AP1/AGL9 subfamily during floral transition. Plant Physiol 2000;123:1325–36. https://doi.org/10.1104/pp.123.4.1325.
[20] Theißen G, Melzer R, Rümpler F. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. Development 2016;143:3259–71. https://doi.org/10.1242/dev.134080.
[21] Nam J. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. Mol Biol Evol 2003;20:1435–47. https://doi.org/10.1093/molbev/msg152.
[22] Chen M-K, Hsu W-H, Lee P-F, Thiruvengadam M, Chen H-I, Yang C-H. The MADS box gene, *FOREVER YOUNG FLOWER*, acts as a repressor controlling floral organ senescence and abscission in Arabidopsis: FYF regulates flower senescence and abscission. Plant J 2011;68:168–85. https://doi.org/10.1111/j.1365-313X.2011.04677.x.
[23] Xie Q, Hu Z, Zhu Z, Dong T, Zhao Z, Cui B, et al. Overexpression of a novel MADS-box gene SlFYFL delays senescence, fruit ripening and abscission in tomato. Sci Rep 2014;4:1–10. https://doi.org/10.1038/srep04367.
[24] Smaczniak C, Muiño JM, Chen D, Angenent GC, Kaufmann K. Differences in DNA-binding specificity of floral homeotic protein complexes predict organ-specific target genes. Plant Cell 2017. https://doi.org/10.1105/tpc.17.00145 (tpc.00145.2017).
[25] Hugouvieux V, Zubieta C. MADS transcription factors cooperate: complexities of complex formation. J Exp Bot 2018;69:1821–3. https://doi.org/10.1093/jxb/ery099.
[26] Bartlett ME. Changing MADS-box transcription factor protein–protein interactions as a mechanism for generating floral morphological diversity. Integr Comp Biol 2017;57:1312–21. https://doi.org/10.1093/icb/icx067.
[27] Tilly JJ, Allen DW, Jack T. The CArG boxes in the promoter of the Arabidopsis floral organ identity gene APETALA3 mediate diverse regulatory effects. Development 1998;125:1647–57.
[28] Folter SD, Angenent GC. Trans meets cis in MADS science. Trends Plant Sci 2006;11:224–31. https://doi.org/10.1016/j.tplants.2006.03.008.
[29] Gramzow L, Theissen G. A hitchhiker's guide to the MADS world of plants. Genome Biol 2010;11:214. https://doi.org/10.1186/gb-2010-11-6-214.
[30] Becker A, Theißen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol 2003;29:464–89. https://doi.org/10.1016/S1055-7903(03)00207-0.
[31] Thangavel G, Nayar S. A survey of MIKC type MADS-box genes in non-seed plants: algae, bryophytes, lycophytes and ferns. Front Plant Sci 2018;9. https://doi.org/10.3389/fpls.2018.00510.
[32] Melzer R, Verelst W, Theißen G. The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro. Nucleic Acids Res 2009;37:144–57. https://doi.org/10.1093/nar/gkn900.
[33] 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 2014;42:10927–42. https://doi.org/10.1093/nar/gku755.
[34] Han A, He J, Wu Y, Liu JO, Chen L. Mechanism of recruitment of class II histone deacetylases by myocyte enhancer Factor-2. J Mol Biol 2005;345:91–102. https://doi.org/10.1016/j.jmb.2004.10.033.
[35] Malcomber ST, Kellogg EA. SEPALLATA gene diversification: brave new whorls. Trends Plant Sci 2005;10:427–35. https://doi.org/10.1016/j.tplants.2005.07.008.
[36] Immink RG, Tonaco IA, de Folter S, Shchennikova A, van Dijk AD, Busscher-Lange J, et al. SEPALLATA3: the "glue" for MADS box transcription factor complex formation. Genome Biol 2009;10:R24. https://doi.org/10.1186/gb-2009-10-2-r24.
[37] Smaczniak C, Immink RGH, Angenent GC, Kaufmann K. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development 2012;139:3081–98. https://doi.org/10.1242/dev.074674.
[38] de Folter S. Comprehensive interaction map of the Arabidopsis MADS box transcription factors. Plant Cell 2005;17:1424–33. https://doi.org/10.1105/tpc.105.031831.
[39] Espinosa-Soto C, Immink RG, Angenent GC, Alvarez-Buylla ER, de Folter S. Tetramer formation in Arabidopsis MADS domain proteins: analysis of a protein-protein interaction network. BMC Syst Biol 2014;8:9. https://doi.org/10.1186/1752-0509-8-9.
[40] Puranik S, Acajjaoui S, Conn S, Costa L, Conn V, Vial A, et al. Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. Plant Cell 2014;26:3603–15. https://doi.org/10.1105/tpc.114.127910.
[41] Silva CS, Puranik S, Round A, Brennich M, Jourdain A, Parcy F, et al. Evolution of the plant reproduction master regulators LFY and the MADS transcription factors: the role of protein structure in the evolutionary development of the flower. Front Plant Sci 2016;6. https://doi.org/10.3389/fpls.2015.01193.
[42] Rümpler F, Theißen G, Melzer R. A conserved leucine zipper-like motif accounts for strong tetramerization capabilities of SEPALLATA-like MADS-domain transcription factors. J Exp Bot 2018;69:1943–54. https://doi.org/10.1093/jxb/ery063.
[43] Krizek BA, Meyerowitz EM. Mapping the protein regions responsible for the functional specificities of the Arabidopsis MADS domain organ-identity proteins. Proc Natl Acad Sci 1996;93:4063–70. https://doi.org/10.1073/pnas.93.9.4063.
[44] Dingwall C, Laskey RA. Nuclear targeting sequences – a consensus? Trends Biochem Sci 1991;16:478–81 (n.d.).
[45] Wu Y, Dey R, Han A, Jayathilaka N, Philips M, Ye J, et al. Structure of the MADS-box/MEF2 domain of MEF2A bound to DNA and its implication for Myocardin recruitment. J Mol Biol 2010;397:520–33. https://doi.org/10.1016/j.jmb.2010.01.067.
[46] Pellegrini L, Tan S, Richmond TJ. Structure of serum response factor core bound to DNA. Nature 1995;376:490–8. https://doi.org/10.1038/376490a0.
[47] Tan S, Richmond TJ. Crystal structure of the yeast MATa2/MCM1/DNA ternary complex. Nature 1998;391:660–6. https://doi.org/10.1038/35563.
[48] Riechmann JL, Meyerowitz EM. Determination of floral organ identity by Arabidopsis MADS domain homeotic proteins API, AP3, PI, and AG is independent of their DNA-binding specificity. Mol Biol Cell 1997;8:17.
[49] Aerts N, de Bruijn S, van Mourik H, Angenent GC, van Dijk ADJ. Comparative analysis of binding patterns of MADS-domain proteins in Arabidopsis thaliana. BMC Plant Biol 2018;18. https://doi.org/10.1186/s12870-018-1348-8.
[50] Hugouvieux V, Silva CS, Jourdain A, Stigliani A, Charras Q, Conn V, et al. Tetramerization of MADS family transcription factors SEPALLATA3 and AGAMOUS is required for floral meristem determinacy in Arabidopsis. Nucleic Acids Res 2018;46:4966–77. https://doi.org/10.1093/nar/gky205.
[51] Riechmann JL, Krizek BA, Meyerowitz EM. Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proc Natl Acad Sci 1996;93:4793–8. https://doi.org/10.1073/pnas.93.10.4793.
[52] van Dijk ADJ, Morabito G, Fiers M, van Ham RCHJ, Angenent GC, Immink RGH. Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. PLoS Comput Biol 2010;6:e1001017. https://doi.org/10.1371/journal.pcbi.1001017.
[53] Lei X, Shi H, Kou Y, Rajashekar N, Wu F, Sen C, et al. Crystal structure of Apo MEF2B reveals new insights in DNA binding and cofactor interaction. Biochemistry 2018;57:4047–51. https://doi.org/10.1021/acs.biochem.8b00439.
[54] Han A, Pan F, Stroud JC, Youn H-D, Liu JO, Chen L. Sequence-specific recruitment of transcriptional co-repressor Cabin1 by myocyte enhancer factor-2. Nature 2003;422:730–4. https://doi.org/10.1038/nature01555.
[55] Lei X, Kou Y, Fu Y, Rajashekar N, Shi H, Wu F, et al. The Cancer mutation D83V induces an α-Helix to β-Strand conformation switch in MEF2B. J Mol Biol 2018;430:1157–72. https://doi.org/10.1016/j.jmb.2018.02.012.
[56] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46:W296–303. https://doi.org/10.1093/nar/gky427.
[57] Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, et al. Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. Proc Natl Acad Sci 2005;102:2436–41. https://doi.org/10.1073/pnas.0409860102.
[58] Kaufmann K, Melzer R, Theißen G. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene 2005;347:183–98. https://doi.org/10.1016/j.gene.2004.12.014.
[59] Yang Y, Jack T. Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. Plant Mol Biol 2004;55:45–59. https://doi.org/10.1007/s11103-004-0416-7.
[60] Yang Y, Fanning L, Jack T. The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. Plant J 2003;33:47–59. https://doi.org/10.1046/j.0960-7412.2003.01473.x.
[61] Wang Y, Zhang X, Zhang H, Lu Y, Huang H, Dong X, et al. Coiled-coil networking shapes cell molecular machinery. Mol Biol Cell 2012;23:3911–22. https://doi.org/10.1091/mbc.e12-05-0396.
[62] Rose A, Meier I. Scaffolds, levers, rods and springs: diverse cellular functions of long coiled-coil proteins. Cell Mol Life Sci 2004;61. https://doi.org/10.1007/s00018-004-4039-6.
[63] Rhys GG, Wood CW, Lang EJM, Mulholland AJ, Brady RL, Thomson AR, et al. Maintaining and breaking symmetry in homomeric coiled-coil assemblies. Nat Commun 2018;9. https://doi.org/10.1038/s41467-018-06391-y.
[64] Lupas AN, Bassler J. Coiled coils – a model system for the 21st century. Trends Biochem Sci 2017;42:130–40. https://doi.org/10.1016/j.tibs.2016.10.007.

[65] Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 2014;42:W320–4. https://doi.org/10.1093/nar/gku316.
[66] Wolf E, Kim PS, Berger B. MultiCoil: a program for predicting two-and three-stranded coiled coils: MultiCoil. Protein Sci 1997;6:1179–89. https://doi.org/10.1002/pro.5560060606.
[67] Vandereyken K, Van Leene J, De Coninck B, Cammue BPA. Hub protein controversy: taking a closer look at plant stress response hubs. Front Plant Sci 2018;9:694. https://doi.org/10.3389/fpls.2018.00694.
[68] Alhindi T, Zhang Z, Ruelens P, Coenen H, Degroote H, Iraci N, et al. Protein interaction evolution from promiscuity to specificity with reduced flexibility in an increasingly complex network. Sci Rep 2017;7. https://doi.org/10.1038/srep44948.
[69] Immink RGH, Pose D, Ferrario S, Ott F, Kaufmann K, Valentim FL, et al. Characterization of SOC1's central role in flowering by the identification of its upstream and downstream regulators. Plant Physiol 2012;160:433–49. https://doi.org/10.1104/pp.112.202614.
[70] Kagale S, Links MG, Rozwadowski K. Genome-wide analysis of ethylene-responsive element binding factor-associated amphiphilic repression motif-containing transcriptional regulators in Arabidopsis. Plant Physiol 2010;152:1109–34. https://doi.org/10.1104/pp.109.151704.
[71] Hill K, Wang H, Perry SE. A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components. Plant J 2008;53:172–85. https://doi.org/10.1111/j.1365-313X.2007.03336.x.
[72] Sridhar VV. APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development. Development 2006;133:3159–66. https://doi.org/10.1242/dev.02498.
[73] Cho S, Jang S, Chae S, Chung KM, Moon Y-H, Jang SK. Analysis of the C-terminal region of Arabidopsis thaliana APETALA1 as a transcription activation domain. Plant Mol Biol 1999;40:419–29 (n.d.).
[74] Kaufmann K, Muiño JM, Jauregui R, Airoldi CA, Smaczniak C, Krajewski P, et al. Target genes of the MADS transcription factor sepallata3: integration of developmental and hormonal pathways in the arabidopsis flower. PLoS Biol 2009;7:0854–75. https://doi.org/10.1371/journal.pbio.1000090.
[75] Honma T, Goto K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature 2001;409:525–9.
[76] Piwarzyk E, Yang Y, Jack T. Conserved C-terminal motifs of the Arabidopsis proteins APETALA3 and PISTILLATA are dispensable for floral organ identity function. Plant Physiol 2007;145:1495–505. https://doi.org/10.1104/pp.107.105346.
[77] Egea-Cortines M. 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 1999;18:5370–9. https://doi.org/10.1093/emboj/18.19.5370.
[78] Steinmetz MO, Jelesarov I, Matousek WM, Honnappa S, Jahnke W, Missimer JH, et al. Molecular basis of coiled-coil formation. Proc Natl Acad Sci 2007;104:7062–7. https://doi.org/10.1073/pnas.0700321104.
[79] Severing EI, van Dijk ADJ, Morabito G, Busscher-Lange J, Immink RGH, van Ham RCHJ. Predicting the impact of alternative splicing on plant MADS domain protein function. PLoS ONE 2012;7:e30524. https://doi.org/10.1371/journal.pone.0030524.
[80] Posé D, Verhage L, Ott F, Yant L, Mathieu J, Angenent GC, et al. Temperature-dependent regulation of flowering by antagonistic FLM variants. Nature 2013;503:414–7. https://doi.org/10.1038/nature12633.