Sequence features of MADS-domain proteins that act as hubs in the protein-protein interaction network controlling flower development

Florian Rümpler[a], Günter Theißen[a,1,*] and Rainer Melzer[a,b,1,*]

[a] *Department of Genetics, Friedrich Schiller University Jena, Philosophenweg 12, D-07743 Jena, Germany*

[b] *School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland*

[1] *Contributed equally to this project*

[*] *For correspondence (e-mail: guenter.theissen@uni-jena.de; rainer.melzer@ucd.ie)*

Key Words

Amborella trichopoda, *Arabidopsis thaliana*, Coiled-coil; Floral quartet; Keratin-like domain; Leucine-zipper; MIKC-type protein; SEPALLATA3; Transcription factor

ABSTRACT

Protein-protein interaction networks (PPIs) are usually scale-free networks that contain a small number of highly connected nodes (hubs) and many poorly connected nodes. However, the molecular mechanisms that underlie the promiscuous interactions of hub proteins remain largely unknown. Here, we show that the floral homeotic MADS-domain transcription factor SEPALLATA3 from *Arabidopsis thaliana* can act as a hub in the PPI controlling flower development because it contains leucine residues at inter- and intramolecular interaction interfaces. Comprehensive sequence analyses of diverse MADS-domain proteins indicate exceedingly high conservation of the identified leucine residues within SEPALLATA-subfamily proteins, whereas non-hub MADS-domain proteins exhibit preferences for other amino acids at homologous sites. Our results indicate that the conservation of leucine residues at positions critical for protein-protein interactions contributed significantly to the present-day structure of the PPI and may have facilitated the evolution of the flower.

Complexity of biological systems is often achieved by the combinatorial activity of a small number of factors [1]. One important example are protein-protein interaction networks (PPIs) that are based on transcription factors (TFs) that act in a combinatorial manner to accomplish the required degree of e.g. morphological complexity. PPIs are often scale-free networks. They contain a small number of hub proteins with many interaction partners and a large number of poorly connected nodes. Though combinatorial control is of eminent importance for almost all developmental processes, the molecular determinants that are underlying the specific combinatorial interactions remain poorly understood. This is especially true for protein-protein interactions among TFs belonging to the same family. The respective TFs are often very similar in terms of sequence and biochemical properties yet fulfill highly distinct and specific functions which are at least partially determined by distinct protein-protein interactions. The PPI controlling flower development in angiosperms is a good case in point. Floral organ specification is regulated by so called floral quartets which are organ specific tetrameric complexes of MIKC-type MADS-domain TFs bound to two adjacent DNA-binding sites while looping the DNA to regulate target genes [2, 3, 4, 5, 6, 7]. In the model plant species *Arabidopsis thaliana* the floral homeotic protein SEPALLATA3 (SEP3) together with its paralogs SEP1, SEP2 and SEP4 from the closely related LOFSEP-subfamily bears a central role by forming tetrameric complexes with numerous other MIKC-type MADS-domain TFs [5, 8, 9, 10]. The four SEP proteins act in a largely redundant manner but in agreement with their central position in the PPI controlling flower development *sep* multiple mutants show severe developmental defects [3, 4]. *sep1 sep2 sep3* triple mutant plants develop sepals from primordia that would normally develop into petals, stamens and carpels and *sep1 sep2 sep3 sep4* quadruple mutants develop vegetative leaves instead of floral organs [3, 4].

Among the four SEP genes, SEP3 has been studied best [5, 6, 8, 9, 11, 12]. Beyond the formation of complexes that determine floral organ identity SEP3 is also involved in controlling flowering time, floral transition and ovule development [8, 11, 13, 14]. It does therefore constitute one of the major hub proteins within the PPI controlling reproductive development [8, 9, 11, 13]. However, it is unclear which biochemical and biophysical properties enable SEP3 to interact with numerous partners whereas other MIKC-type MADS-domain TFs show a much narrower interaction spectrum. For example, the floral homeotic proteins APETALA3 (AP3) and PISTILLATA (PI) from *A. thaliana* that are involved in the developmental specification of petals and stamens do only form obligate heterodimers and require SEP proteins for tetramer formation [5, 8, 15].

The protein-protein interactions that allow for tetramer formation are mainly mediated by the about 80 amino acids long keratin-like domain (K-domain), which is shared by all MIKC-type MADS-domain TFs [5, 16, 17]. The amino acid sequence within the K-domain of most MADS-domain proteins shows three characteristic heptad repeat patterns (K1-; K2-; K3-subdomain repeat) of the form $[abcdefg]_n$, where most 'a' and 'd' positions are occupied by highly hydrophobic residues [16, 17, 18]. This sequence feature is typical for coiled-coils, a common and intensively studied type of protein-protein interaction domains [19, 20, 21, 22] (Fig. 1). Within a coiled-coil, an α-helix is formed and the amino acids on heptad repeat 'a' and 'd' positions form a stripe of hydrophobic residues that runs along the α-helix and facilitates hydrophobic interaction with a partner coiled-coil [20, 21].

Recently the crystal structure of the complete K-domain of SEP3 was reported [23]. Based on the crystal structure, the K-domain forms two amphipathic α-helices separated by a kink region which prevents intramolecular association of both helices. Helix one comprises the first heptad repeat (K1-subdomain) and is involved in dimerisation of two SEP3 monomers. Helix two spans heptad repeat two (K2-subdomain) that further stabilizes the interaction of two SEP3 monomers and heptad repeat three (K3-subdomain) which constitutes an interface for the interaction of two SEP3 dimers i.e. tetramerisation (Fig. 1). In this study, we determine sequence features that enable SEP3 to form tetrameric complexes and identify the amino acid patterns that distinguish members of the SEP3- and LOFSEP-protein subfamilies (termed SEP subfamily henceforth for simplicity) from other MIKC-type MADS-domain TFs with more restricted interaction capabilities. Our data suggest that leucine residues at intramolecular contact points and at the interaction interface of the K3-subdomain are indispensable for tetrameric complex formation. Those leucines are highly conserved in the SEP subfamily but much less frequent in e.g. AP3- and PI-subfamily MIKC-type MADS-domain TFs. They may thus be a critical denominator that determines the ability of SEP-subfamily proteins to act as a hub protein in the scale free PPI controlling flower development.

RESULTS

Leucine residues in the K-domain strongly influence cooperative DNA-binding of SEP3

To investigate the relevance of the different K-subdomains for cooperative DNA-binding and tetramer formation, single and double amino acid substitutions to proline were performed. Proline was chosen because it is known to possess helix-breaking properties [24, 25]. For each of the three K-subdomains two substitution mutants were created (Fig. 2a, Supplementary Fig. 1). Based on coiled-coil predictions one substitution mutant was supposed to destroy the K-subdomain coiled-coil (L115P for K1-, L131P-L135P for K2- and L164P for K3-subdomain, respectively) whereas the other one was expected to not alter the formation of the respective coiled-coil (S94P (K1); L145P (K2); G178P (K3)). Beyond the three K-subdomains, we also introduced proline substitutions at positions occupied by two conserved hydrophobic amino acids in the interhelical region between the K1- and the K2-subdomain (L120P and L123P, Fig. 2a) because positions homologous to L120 and L123 have been shown to be important for the interaction of MADS-domain proteins [17].

We used electrophoretic mobility shift assays (EMSAs) to study the DNA-binding and tetramerisation behavior of the mutant SEP3 proteins. Based on previous studies it is known that SEP3 binds as homodimer to a DNA-element termed CArG-box (for CCArichGG; consensus sequence 5'-CC(A/T)$_6$GG-3') and that four SEP3 proteins bind to a DNA probe containing two CArG-boxes [6]. To first investigate whether DNA-binding affinities of individual dimers were affected by the different amino acid substitutions, we performed saturation binding EMSA experiments using increasing amounts of a DNA probe containing only one CArG-box together with constant amounts of protein as previously described [12]. The estimated affinities for binding of the altered SEP3 proteins to a single DNA-binding site varied slightly but did not considerably differ from the values obtained for SEP3 wild type protein (Supplementary Fig. 2, Supplementary Table 1), indicating that the different amino acid substitutions did not or only marginally affect DNA-binding of individual dimers.

If increasing amounts of SEP3 were incubated together with constant amounts of a DNA probe containing two CArG-boxes, three bands of different electrophoretic mobility were observed (Fig. 2b left side). As determined previously [6] the band of high electrophoretic mobility constitutes unbound DNA (indicated with '0' in Fig. 2b), the band of intermediate electrophoretic mobility constitutes a DNA probe bound by two SEP3 proteins ('2') and the band of low electrophoretic mobility constitutes a DNA probe bound by four SEP3

proteins (‘4’). By analyzing the signal intensities of the three different fractions the ability of SEP3 to form DNA-bound tetrameric complexes can be quantified and expressed via the cooperativity constant k_{coop} (equation (4) in Methods). k_{coop} equals 1 for non-cooperative binding and increases with increasing tetramer formation capabilities of the examined protein. SEP3 wild type protein always showed a highly cooperative DNA-binding although the degree of cooperativity varied between different experiments and was slightly higher than previously estimated [6, 12], probably owing to difficulties to precisely determine high k_{coop} values (Fig. 2b and d, Supplementary Table 1).

In contrast to the wild type protein, all of the leucine-to-proline substitution mutants of SEP3 (L115P; L120P-L123P; L131P-L135P; L145P; L164P) showed a considerably reduced ability to bind cooperatively to DNA *in vitro*, independent of whether the formation of coiled-coils was predicted to be affected or not (Fig. 2c and d, Supplementary Table 1). Only the two proline substitutions S94P and G178P, located at the N- and C-terminal borders of the K-domain, respectively, did not strongly reduce cooperative binding of SEP3.

To test the effect of amino acid substitutions that are supposed to have a less severe effect on helix formation than proline, we substituted a subset of the previously selected leucines (L115; L145; L164) by alanine. Surprisingly, of these 3 substitutions only L145A showed a cooperative DNA-binding ability comparable to that of SEP3 wild type protein, whereas substitutions L115A and L164A caused an almost complete loss of cooperative DNA-binding, comparable to the proline substitutions at the respective positions (Fig. 2d, Supplementary Table 1). We further substituted position L164 by three additional amino acids (L164E; L164W; L164I) comprising glutamate and tryptophan which occur at position 164 in several members of the SEP subfamily and isoleucine which has very similar physicochemical properties to leucine. However, none of the resulting mutants was able to approach SEP3 wild type cooperative binding strength (Fig. 2d, Supplementary Table 1). Our results indicate that the examined leucine residues are of critical importance for tetramer formation and cooperative binding of SEP3.

Within the $[abcdefg]_n$ heptad repeat of the K3-subdomain of SEP3 two neighboring ‘a’ positions (E161; N168) are not occupied by hydrophobic amino acids. Substituting these positions by leucine (E161L-N168L) resulted in a higher probability for the formation of the K3-subdomain coiled-coil *in silico* (Supplementary Fig. 1). The respective mutant protein showed a cooperativity at least as high as the wild type protein in EMSAs. In contrast to the wild type protein, repeated measurements yielded k_{coop} values that consistently were above

200 (Fig. 2d, Supplementary Table 1). In fact, in none of the performed EMSAs a signal of a DNA probe bound by only one protein dimer was detected, an observation that was different from the other proteins for which high cooperativity in DNA-binding was detected (e.g. SEP3-WT and SEP3-L145A) indicating that cooperative binding was increased by the E161L-N168L substitutions (Supplementary Fig. 3). Surprisingly, when we performed saturation binding EMSA experiments using increasing amounts of a DNA probe containing only one CArG-box the mutant protein SEP3-E161L-N168L exhibited no binding of individual dimers. Instead a signal of low electrophoretic mobility occasionally occurred for high amounts of applied DNA probe that might constitutes a protein DNA complex consisting of more than two proteins (Supplementary Fig. 4).

Mutations in a distantly related ortholog of SEP3 have very similar effects on cooperative DNA-binding as in SEP3

The SEP3 ortholog AMtrAGL9 from the early diverging angiosperm *Amborella trichopoda* [15] forms homotetrameric protein-DNA complexes with a cooperative binding affinity comparable to SEP3 (Fig. 2e, Supplementary Fig. 5). AMtrAGL9 amino acid position I141 is homologous to SEP3 L145 and is thus located in the K2-subdomain heptad repeat of AMtrAGL9 (Fig. 2a). Substitution to alanine at that position interfered to some extend with cooperative binding capabilities whereas substitution to proline at position I141 results in an almost complete loss of cooperative binding (Fig. 2e, Supplementary Table 1). If amino acid position L160 of AMtrAGL9, which is homologous to position L164 in the center of the K3-subdomain of SEP3, is exchanged by proline or alanine, the ability of AMtrAGL9 to cooperatively bind to DNA is almost completely lost in either case, a behavior that is similar to that observed for SEP3 (compare Fig. 2d and e).

Interacting sites are more often occupied by leucine in SEP-subfamily proteins than in proteins of other MIKC-type subfamilies

The importance of leucine residues for the tetramerisation ability of SEP3 and AMtrAGL9 raised the question as to which extent these positions are conserved within the SEP subfamily and which amino acid preferences members of other MIKC-type protein subfamilies show at homologous sites. We therefore created a multiple sequence alignment based on 1,325 sequences of MIKC-type MADS-domain proteins belonging to 14 subfamilies and comprising sequences from a diverse array of seed plants. Despite the high evolutionary distance of the sampled taxa, the sequences aligned almost without gaps throughout the

complete K-domain (i.e. without potential insertions or deletions). The only exception were PI-subfamily protein sequences, among which a deletion of four amino acids within the C-terminal half of the K-domain was very common. This deletion within the PI-linage most likely occurred after early diverging angiosperms branched off, as most of the sampled PI-subfamily sequences from early diverging angiosperms still possess those four amino acids.

We first compared the conservation of sites that are homologous to the 15 residues that (based on the crystal structure of SEP3) mediate the hydrophobic intra- and intermolecular interactions in the SEP3 homotetramer [23] to the overall conservation of the K-domain. We found that within the SEP3 subfamily, sites that are homologous to interacting sites in the SEP3 homotetramer are significantly less variable than the remaining residues of the K-domain (Fig. 3a). This conservation pattern also holds true for sequences of all other 13 subfamilies of MIKC-type MADS-domain proteins (Fig. 3b, Supplementary Fig. 6a) as well as for sequences from gymnosperms to core eudicots (Supplementary Fig. 6b). Beyond this similar pattern of conserved positions also the amino acid properties in terms of hydrophobicity at homologous sites appear highly similar among all examined subfamilies (Supplementary Fig. 7), suggesting that the overall structure of the K-domain as determined for SEP3 is conserved among MIKC-type proteins of most if not all subfamilies and throughout seed plants.

Next we analyzed the amino acid distribution at sites homologous to the 12 leucine residues (L101, L108, L115, L120, L123, L128, L131, L135, L154, L157, L164 and L171) that contribute to inter- and intramolecular interactions in a SEP3 homotetramer (Fig. 4a) [23]. All these residues were found to be highly conserved within the 78 examined SEP3-subfamily sequences; 8 out of 12 positions were completely invariable (Fig. 4b). In contrast to this, members of other subfamilies (e.g. AP3- and PI-subfamily proteins) often show preferences for other amino acids on equivalent sites (Fig. 4b, Supplementary Fig. 8). Especially positions equivalent to L154, L157 and L164 of SEP3 that are located within the center of the tetramerisation interface are often not occupied by leucines in AP3- and PI-subfamily proteins. The high conservation of leucines also becomes apparent within LOFSEP-subfamily proteins (comprising SEP1, SEP2 and SEP4 from *A. thaliana*) which form the sister group of SEP3-subfamily proteins and that are assumed to function in a mostly redundant manner with SEP3 during flower development (Fig. 4c) [3, 4, 10]. The closest relatives of the SEP subfamily are AGL6-subfamily proteins followed by AP1-subfamily proteins [26]. However, despite the close relationship AGL6- as well as AP1-subfamily proteins display a considerably lower

leucine frequency especially on sites within the tetramerisation interface (Fig. 4c, Supplementary Fig. 8). Instead, these positions are more frequently occupied by other hydrophobic amino acids such as isoleucine and methionine. It has previously been shown that within a coiled-coil, leucine packs very well at heptad repeat 'd' positions and enables the formation of a tight dimeric coiled-coil as it becomes apparent in a leucine-zipper [22, 27]. In contrast other hydrophobic amino acids such as isoleucine or valine lead to steric hindrance at heptad repeat 'd' positions [22] (Fig. 4d).

Insertion of leucine residues into the K3-subdomain of AP3 facilitates homotetramerisation of the chimeric protein $SEP3_{AP3chim}$

Based on our data we hypothesized that the overall structure of the K-domain is conserved throughout most if not all subfamilies of MIKC-type MADS-domain proteins and that preferences for different amino acids on interacting sites account for subfamily specific interaction patterns. We aimed to test our hypothesis with help of the chimeric protein $SEP3_{AP3chim}$, in which we substituted the K3-subdomain (i.e. tetramersiation interface) of SEP3 (residues 150-181) by the homologous sites of AP3 (Fig. 5a and b). AP3 is known to form obligate heterodimers with PI and is thus not able to form DNA-binding homodimers or homotetramers [28, 29]. As expected, the chimeric protein $SEP3_{AP3chim}$ showed a complete loss of homotetramerisation capabilities compared to SEP3 wild type protein in EMSA experiments (Fig. 5c and d right side, Supplementary Table 1). Although the K3-subdomains of SEP3 and AP3 share only four identical residues at homologous sites the sequence similarity in terms of hydrophobicity on most heptad repeat 'a' and 'd' positions is comparatively high (Fig. 5a). However, two heptad repeat 'd' positions occupied by leucine in SEP3 (L157 and L164) are occupied by threonine and glutamine in AP3, respectively (Fig. 5a). Both leucines are highly conserved throughout SEP3-subfamily proteins whereas homologous sites in AP3-subfamily proteins are almost exclusively occupied by residues other than leucine (Fig. 4b). We thus substituted positions T157 and Q164 of the chimeric protein by leucine and tested the ability of the resulting mutants to form homotetramers. Both single amino acid substitutions could not improve tetramerisation ability of the chimeric protein (Supplementary Table 1). However, the insertion of both leucine residues into the K3-subdomain of $SEP3_{AP3chim}$ sufficed to fully 'restore' the ability to form DNA-binding homotetramers (Fig. 5e right side). Visualizing the amino acid sequence of the tetramerisation interface of SEP3 and AP3 in a helical wheel diagram illustrates how residues M150, L157, L164 and L171 form a strong hydrophobic stripe within the tetramerisation interface of SEP3, whereas the hydrophobic

stripe is interrupted by threonine and glutamine in AP3 (Fig. 5c and d left side). Substituting both residues by leucine closes the gap within the hydrophobic stripe and most likely thereby facilitates homotetramerisation (Fig. 5e left side).

DISCUSSION

Tetramer formation among MIKC-type MADS-domain transcription factors is of central importance for flower development [5, 7, 8, 9, 30]. However, knowledge about the molecular determinants facilitating tetramer formation remains scarce. Our data indicate that substitution of leucines in the K-domain of SEP3 did almost invariably lead to a strong reduction in tetramer formation abilities (Fig. 2). This was expected for leucine to proline substitutions within the helical regions of the K-domain as proline has helix-breaking properties. However, also the rather conservative substitution from leucine to alanine in the tetramerization interface (L164A) affected cooperative binding and tetramerization strongly. Similar results have been obtained for substituting other leucine residues in the tetramerization interface by alanine [23].

The question arises as to why specifically leucine residues are favoured over other hydrophobic amino acid residues. The tetramerization interface forms coiled-coils and it is well established that complex 'knobs-into-holes' side chain interactions within the hydrophobic core determine the strength of the interaction between coiled-coils [19]. Numerous studies on energetic contributions of different hydrophobic amino acids inside the hydrophobic core revealed that β-branched amino acids (e.g. isoleucine or valine) as well as amino acids with small side chains (e.g. alanine) in heptad repeat 'd' positions have a strong destabilizing effect on formation of parallel dimeric coiled-coils [27, 31]. The local stereochemical environment at heptad repeat 'd' positions instead strongly favours γ-branched amino acids for intermolecular interactions, making leucines uniquely suited at these sites [22, 27, 29, 31]. This is in line with the observation that L145, which is located at a heptad repeat 'd' position but according to structural data not involved in intermolecular interactions [23] can be mutated to alanine without a decrease in tetramer formation capabilities. In contrast, L164 (also at a heptad repeat 'd' position but involved in intermolecular interactions) mutation to alanine leads to a strong decrease in tetramerization. In addition, L145 is by far not as conserved as leucines involved in interactions (Supplementary Fig. 8).

A decrease in tetramer formation was also observed for substitution of leucines in the kink region between the two helices, where an effect on helix formation was not predicted (Supplementary Fig. 1). However, although the leucine residues in the kink are not directly involved in tetramer formation, they interact intramolecularly with each other to stabilize the kink and thus bring the tetramer interface in a favourable position for protein-protein interactions [23]. It is likely that substitutions to proline or alanine in the kink region altered or

destabilized the orientation of the tetramerisation interface and thus impeded tetramer formation indirectly. Similar to the leucines at interacting sites within the helical regions of the K-domain stereochemical restrictions may also in this case favour leucines over other hydrophobic amino acids. This may explain why the L115A mutation in the kink region, which presumably only affects intramolecular interactions, caused a decrease in tetramer formation capabilities.

Taken together, these findings indicate that inter- and intramolecular hydrophobic interactions specifically among leucines are of critical importance for SEP3 homotetramerization. This principle does very likely apply to the entire subfamily of SEP proteins, as leucines at interaction positions are evolutionarily highly conserved throughout this subfamily (Fig. 4). The evolutionary conserved and important role of leucines is further highlighted by the observation that in the SEP3 ortholog AMtrAGL9 from *A. trichopoda* leucines at positions homologous to those in SEP3 were also of critical importance for tetramer formation (Fig. 2).

The K-domain is the second highest conserved domain of MIKC-type proteins (the most highly conserved domain is the MADS-domain) [32]. Previous structural predictions indicated that the K-domain is forming coiled-coils in most if not all MIKC-type proteins [18, 23, 33]. Our analyses indeed strongly support this view. The chemical properties of amino acids that are of importance for intra- and intermolecular interactions in SEP3 are conserved in MIKC-type proteins from all of the 14 subfamilies analysed here. This indicates that most K-domains fold in a structure similar to that determined for SEP3 and that residues that are homologous to interacting sites in the SEP3 homotetramer may also constitute intra- and intermolecular contact points in most other protein family members.

However, although the chemical properties of amino acids important for interactions were conserved in subfamilies other than SEP, their identity was not always. Whereas the vast majority of leucine residues important for intra- and intermolecular interactions is highly conserved within the SEP subfamily, leucine residues are observed at a clearly lower frequency in other subfamilies (Fig. 4, Supplementary Fig. 4). This indicates that, although the overall structure of the K-domain is conserved in all MIKC-type proteins, their tetramerization capabilities may vary depending on the presence of leucines on critical interaction sites. For example, AP3 and PI, who do not possess leucines on all inter- and intramolecular contact points, are unable to form tetramers not involving SEP3 [5, 9]. Indeed, the K3-subdomain of AP3, which is not capable of mediating homotetramer formation, gained this ability when placed in the SEP3 protein context and two leucines were introduced

(Fig. 4). Thus, we speculate that leucines at intra- and intermolecular contact points may not only be necessary but also sufficient for tetramer formation of MIKC-type proteins.

Intriguingly, the high conservation of leucines in the K-domain of SEP-subfamily proteins and their importance for homotetramer formation correlates very well with the crucial function of those proteins as hubs within the PPI controlling flower development. In addition, proteins like AP3 and PI that have less central positions within the interaction network also lack leucines at several positions critical for tetramerization. It thus appears plausible that leucines in SEP-subfamily proteins are not only important for homotetramer formation but also play a pivotal role in the formation of heterotetrameric complexes. For example, though a lack of leucines in the kink region of many MIKC-type proteins may destabilizes the orientation of the teramerization interface and prevents homotetramer formation, the high structural stability of the K-domain of SEP-subfamily proteins that is brought about by intramolecular leucine interactions may serve as a scaffold that helps to align the interaction interface of partner proteins and hence facilitate heterotetramer formation.

The pattern of leucines at the tetramerization interface may be explained in a similar manner. Though data on the interaction of leucines at heptad repeat ‘d’ positions with other amino acids at ‘d’ positions in a heteromeric coiled-coil are scarce, data from leucine zippers indicate that beyond leucine-leucine interactions also interactions of leucines with a number of other amino acids are more favourable than most other interactions not involving any leucine [34].

Taken together, we propose that the leucine residues in SEP-subfamily proteins serve to facilitate heterotetrameric interactions while at the same time the absence of leucines in the interaction partners prevents homotetramer formation or formation of heterotetramers not involving SEP-subfamily proteins. This way, SEP-subfamily proteins could act as hubs in the scale free PPI controlling flower development: tetramerization of many proteins depends on them and probably cannot occur in the absence of SEP-subfamily proteins.

We previously proposed that the dependence of other MIKC-type proteins on SEP-subfamily proteins for tetramer formation facilitated the concerted development of the different floral organs and the evolution of the flower as a single reproductive entity [15]. The evolutionary conservation of leucines in the SEP subfamily as opposed to most other subfamilies may be thus one important molecular mechanism that fostered the evolution of the flower.

Importantly, however, coiled-coil interactions are very complex, with amino acids occupying heptad repeat ‘a’, ‘d’, ‘e’ and ‘g’ position playing key roles in determining the affinity and specificity of an interaction [20, 21, 35] and we are far from completely understanding the implications of sequence variations on the different positions for MIKC-type protein interactions. For example, polar and charged residues are observed at heptad repeat ‘d’ positions in a number of MIKC-type protein subfamilies and those would be expected to not only hinder homotetramerization but also heterotetramerization with SEP-subfamily proteins. Furthermore, subfamily specific patterns of charged residues at heptad repeat ‘e’ and ‘g’ positions can be observed that may account for differences in interaction specificity. Although our findings bring us one step closer towards solving the code for floral quartet-like complex formation, additional structural and biophysical analyses are required to more completely understand the molecular mechanisms and evolutionary patterns of MIKC-type protein interactions. This will eventually also lead to a better understanding as to why this transcription factor family expanded in seed plants and plays a role in virtually every reproductive developmental process.

METHODS

Cloning procedures and site-directed mutagenesis.

The plasmids for *in vitro* transcription/translation of *SEP3*, *AP3*, *PI* and *AMtrAGL9* (pTNT-SEP3, pSPUTK-AP3, pSPUTK-PI and pSPUTK-AMtrAGL9) have been generated previously [6, 15]. The cDNA sequences for the single- and double amino acid substitution mutants of SEP3 were created by site-directed mutagenesis PCR according to the Q5 Site-Directed Mutagenesis Manual (New England Biolabs). The cDNA sequence for the chimeric protein $SEP3_{AP3chim}$ was created by megaprimer-mediated mutagenesis PCR for domain substitutions according to [36].

Design of DNA probes and radioactive labeling.

Design and preparation of DNA probes have been described previously [6]. The CArG-box sequence 5'-CCAAATAAGG-3' that was used for all DNA probes was derived from the regulatory intron of *AGAMOUS*. For studies on homotetramer formation a 151 nt long DNA probe was used that contained two CArG-boxes in a distance of 63 bp, i.e. 6 helical turns (sequence: 5'- TCGAG GTCGG AAATT TAATT ATATT <u>CCAAA TAAGG</u> AAAGT ATGGA ACGTT CGACG GTATC GATAA GCTTG ATGAA ATTTA ATTAT ATT<u>CC AAATA AGG</u>AA AGTAT GGAAC GTTAT CGAAT TCCTG CAGCC CGGGG GATCC ACTAG TTCTA G -3', CArG-box sequences are underlined). Saturation binding assays to quantify dimer binding affinities were performed with a 51 nt long DNA probe harboring a single CArG-box in the center (sequence: 3'- AATTC GAAAT TTAAT TATAT T<u>CCAA ATAAG G</u>AAAG TATGG AACGT TGAAT T- 5', CArG-box sequence is underlined). The DNA probes were radioactively labeled via Klenow fill-in reaction of 5'-overhangs with [α-^{32}P] dATP.

In vitro transcription/translation and electrophoretic mobility shift assay.

Proteins were produced *in vitro* using the TNT SP6 Quick Coupled Transcription/Translation System (Promega) according to the manufacturer's instructions and used directly without freezing and thawing. The composition of the protein-DNA binding reaction buffer was essentially as described [37], with final concentration of 1.6 mM EDTA, 10.3 mM HEPES, 1 mM DTT, 1.3 mM spermidine, 33.3 ng/µl Poly dI/dC, 2.5 % CHAPS, 4.3 % glycerol, and a minimum of 1.3 µg/µl BSA. The amounts of protein, DNA probe and BSA were varied according to the performed assay. For cooperative DNA-binding studies to infer tetramer

formation capabilities a constant amount of 0.1 ng of a DNA probe containing two CArG-boxes in a distance of six helical turns was co-incubated with variable amounts of *in vitro* translated protein ranging from 0.05 µl to 3 µl. Variable amounts of applied *in vitro* translated protein were compensated by adding appropriate volumes of BSA (10 µg/µl). For saturation binding assays to quantify dimer binding affinities a constant amount of 2 to 5 µl *in vitro* translated protein was co-incubated with variable amounts of a DNA probe containing one CArG-box in the center, ranging from 0.05 to 32 ng as previously described in [12]. Binding reactions had a total volume of 12 µl, were incubated overnight at 4°C and subsequently loaded on a polyacrylamide (5 % acrylamide, 0.1725 % bisacrylamid) 0.5x TBE gel that has been pre-run for 30 min. The gel was run with 0.5x TBE buffer for 2.5 h at 7.5 V/cm and afterwards dried and exposed onto a phosphorimaging screen to quantify signal intensities.

Quantification of cooperative DNA-binding.

For each lane of the EMSA gel relative signal intensities of all fractions were measured using Multi Gauge 3.1 (Fujifilm). The equations that were used to quantify the ability for cooperative DNA-binding of two dimers to a DNA probe carrying two CArG-boxes have been described previously [6, 38]. Briefly, if the relative concentration of unbound DNA probe $[Y_0]$ (signal of high electrophoretic mobility), a DNA probe bound by two proteins $[Y_2]$ (signal of intermediate electrophoretic mobility), and a DNA probe bound by four proteins $[Y_4]$ (signal of low electrophoretic mobility) are described as a function of applied protein [P2],

$$[Y_0] = \frac{1}{1+\left(\frac{2}{k_{d1}}\right)\times[P2]+\left(\frac{1}{k_{d1}\times k_{d2}}\right)\times[P2]^2} \quad \textbf{(1)}$$

$$[Y_2] = \frac{\left(\frac{2}{k_{d1}}\right)\times[P2]}{1+\left(\frac{2}{k_{d1}}\right)\times[P2]+\left(\frac{1}{k_{d1}\times k_{d2}}\right)\times[P2]^2} \quad \textbf{(2)}$$

$$[Y_4] = \frac{\left(\frac{1}{k_{d1}\times k_{d2}}\right)\times[P2]^2}{1+\left(\frac{2}{k_{d1}}\right)\times[P2]+\left(\frac{1}{k_{d1}\times k_{d2}}\right)\times[P2]^2} \quad \textbf{(3)}$$

then k_{d1} is the dissociation constant for binding of a protein dimer to a DNA probe with two unoccupied binding sites and k_{d2} is the dissociation constant for binding of a second protein dimer to a DNA probe where one of the two binding sites is already occupied. By nonlinear regression of the measured signal intensities of the three fractions to equation (1) to (3), k_{d1}

and k_{d2} were estimated using GraphPad Prism 5 (GraphPad Software). As we used *in vitro* transcription/translation for protein production, the exact protein concentrations were unknown. Therefore the amount of applied *in vitro* transcription/translation mixture was used as proxy for [P2], as previously described [6]. As a result of the unknown protein concentrations the estimated values for k_{d1} and k_{d2} depend on the *in vitro* transcription/translation efficiency and can only be considered as relative values. However, estimating a cooperativity constant k_{coop} (defined as the ratio of k_{d1} and k_{d2}) is still possible:

$$k_{coop} = \frac{k_{d1}}{k_{d2}} \quad \textbf{(4)}$$

As described earlier, k_{coop} values of ≈ 200 were the upper limit that could be determined with our experimental setup [12].

Saturation binding assay. To estimate the dissociation constant for binding of a protein dimer to a single DNA-binding site k_d, saturation binding assays with a DNA probe carrying a single CArG-box were performed. The equation that was used to infer k_d has been described previously [12]. k_d can be defined as

$$k_d = \frac{([\mathrm{P_t}]-[\mathrm{PD}])\times[\mathrm{D}]}{[\mathrm{PD}]} \quad \textbf{(5)}$$

with [PD], [P_t], and [D] being the concentration of the protein-DNA complex, total protein, and unbound DNA probe, respectively. By expressing [PD] as a function of [D] for increasing concentrations of applied DNA probe, [P_t] and k_d were determined via nonlinear regression using GraphPad Prism 5.

Multiple sequence alignments and *in silico* sequence analysis.

For analyses on amino acid preferences of different MIKC-type MADS-domain protein subfamilies throughout the K-domain a comprehensive sequence collection was compiled. Via BLAST search [39] representatives of all 14 subfamilies [40, 41] of MIKC-type proteins present in *A. thaliana* (AP1-, AP3-, PI-, AG-, ABS-, SEP3-, LOFSEP-, AGL6-, AGL12-, AGL15-, AGL17-, FLC-, TM3-, and SVP-subfamily) were collected using the amino acid sequences of *A. thaliana* AP1, AP3, PI, AG, ABS, SEP3, SEP1, AGL6, AGL12, AGL15, AGL17, FLC, SOC1, and SVP, respectively, as query. To cover a broad set of species, six individual searches were performed for each subfamily. Each of those searches was restricted to a different group of seed plants: core eudicots, early diverging eudicots, monocots, magnoliids,

early diverging angiosperms, and gymnosperms. For sequences from core eudicots the search queries were restricted to asterids (BLAST tax-ID: 71274), Dilleniaceae (24942), Caryophyllidea (108240), Santalales (41947), Berberidopsidales (403664), Saxifragales (41946), rosids (71275), and Gunnerales (232382); for sequences from early diverging eudicots the search queries were restricted to Proteales (232378), Buxales (280577), and Ranunculales (41768); for sequences from monocots and magnoliids, respectively, the queries were restricted to the corresponding predefined organism groups implemented in BLAST (tax-ID: 4447 and 232347, respectively); for sequences from early diverging angiosperms the queries were restricted to Austrobaileyales (82956), Hydatellaceae (178426), Nymphaeales (261007), and Amborella (13332); and for sequences from gymnosperms the queries were restricted to Gnetales (3378), Pinaceae (3318), Taxaceae (25623), Cephalotaxus (50178), Cupressaceae (3367), Araucariaceae (25664), Podocarpaceae (3362), Ginkgoales (3308), and Cycadales (3297). For each of the 84 resulting BLAST searches the amino acid sequences of all hits were downloaded (if more than 100 sequences were found, only top 100 hits according to the total score calculated by BLAST were downloaded). The results of all BLAST searches were combined into a single data set, all completely redundant sequences as well as all sequences that did not constitute MIKC-type proteins were removed and the remaining sequences were aligned with Mafft applying E-INS-i mode using Jalview [42, 43]. The subfamily assignment of each sequence was performed according to its clustering within a phylogenetic tree calculated with MrBayes (based on MADS-, I- and K-domain sequences, applying mixed AA model with 20 million generations, 50% burn-in, and a sample frequency of 1000) [44]. All sequences with uncertain subfamily assignment were removed. To optimize the alignment quality of the K-domain 133 further sequences were removed that produced gaps and that did not appear to be representative for the respective subfamily. The final sequence collection comprised 1325 MIKC-type protein sequences.

Relative sequence similarities at homologous sites were calculated with R (https://www.R-project.org/). Each pair of amino acids at equivalent sites were assigned a similarity score based on BLOSUM40 values normalized to 1. Subsequently, all pairwise similarity scores were averaged to calculate the mean relative sequence similarity for all amino acid positions within the K-domain. BLOSUM40 was chosen because the average sequence identity within the K-domain of all examined sequences was about 40 %. Box-plots and line graphs of sequence similarity scores were created with SPSS (IBM). Statistical significance of sequence similarity differences were tested via Mann-Whitney U-tests implemented in SPSS.

Subfamily specific amino acid frequencies and mean hydrophobicity values for positions within the K-domain were calculated with R. SEP3 K-domain crystal structure pictures were created with Swiss-PdbViewer (SIB). Helical wheel diagrams were created with R. Coiled-coil predictions to preselect potential positions for single and double amino acid substitutions were performed with COILS [45].

ACKNOWLEDGEMENTS

We are grateful to Fredo-Torpedo (Fred Ferber), Chris-Master (Christian Gafert) and Tanja Schulze for valuable help with some experiments. This research was supported by the DFG (Deutsche Forschungsgemeinschaft) grant to G.T. and R.M. (TH417/5-3). R.M. received a post-doctoral fellowship from the Carl Zeiss Foundation.

AUTHOR CONTRIBUTIONS

G.T., R.M. and F.R. designed the study. F.R. performed the experiments. G.T., R.M. and F.R. wrote the manuscript.

COMPETING FINANCIAL INTEREST

The authors declare no competing financial interests.

REFERENCES

1. Remenyi A, Scholer HR, Wilmanns M. Combinatorial control of gene expression. *Nat Struct Mol Biol* **11**, 812-815 (2004).

2. Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31-37 (1991).

3. Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. *Curr Biol* **14**, 1935-1940 (2004).

4. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405**, 200-203 (2000).

5. Melzer R, Theißen G. Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. *Nucleic Acids Res* **37**, 2723-2736 (2009).

6. 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* **37**, 144-157 (2009).

7. Theißen G, Saedler H. Plant biology. Floral quartets. *Nature* **409**, 469-471 (2001).

8. Immink RGH*, et al.* SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biol* **10**, R24 (2009).

9. Smaczniak C*, et al.* Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc Natl Acad Sci USA* **109**, 1560-1565 (2012).

10. Zahn LM*, et al.* The evolution of the SEPALLATA subfamily of MADS-Box genes: A preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169**, 2209-2223 (2005).

11. Favaro R*, et al.* MADS-box protein complexes control carpel and ovule development in Arabidopsis. *Plant Cell* **15**, 2603-2611 (2003).

12. Jetha K, Theißen G, Melzer R. Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Res* **42**, 10927-10942 (2014).

13. Liu C, Xi W, Shen L, Tan C, Yu H. Regulation of floral patterning by flowering time genes. *Dev Cell* **16**, 711-722 (2009).

545 14. Lopez-Vernaza M, Yang SX, Muller R, Thorpe F, de Leau E, Goodrich J. Antagonistic roles of
546 SEPALLATA3, FT and FLC genes as targets of the polycomb group gene CURLY LEAF. *Plos One*
547 **7**, e30715 (2012).

548
549 15. Melzer R*, et al.* DEF- and GLO-like proteins may have lost most of their interaction partners
550 during angiosperm evolution. *Ann Bot* **114**, 1431-1443 (2014).

551
552 16. Yang YZ, Fanning L, Jack T. The K domain mediates heterodimerization of the Arabidopsis
553 floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J* **33**, 47-59 (2003).

554
555 17. Yang YZ, Jack T. Defining subdomains of the K domain important for protein-protein
556 interactions of plant MADS proteins. *Plant Mol Biol* **55**, 45-59 (2004).

557
558 18. Riechmann JL, Meyerowitz EM. MADS domain proteins in plant development. *Biol Chem* **378**,
559 1079-1101 (1997).

560
561 19. Parry DA, Fraser RD, Squire JM. Fifty years of coiled-coils and alpha-helical bundles: a close
562 relationship between sequence and structure. *J Struct Biol* **163**, 258-269 (2008).

563
564 20. Mason JM, Arndt KM. Coiled coil domains: Stability, specificity, and biological implications.
565 *Chembiochem* **5**, 170-176 (2004).

566
567 21. Mason JM, Hagemann UB, Arndt KM. Role of hydrophobic and electrostatic interactions in
568 coiled coil stability and specificity. *Biochemistry* **48**, 10380-10388 (2009).

569
570 22. Betz SF, Bryson JW, Degrado WF. Native-like and structurally characterized designed alpha-
571 helical bundles. *Curr Opin Struc Biol* **5**, 457-463 (1995).

572
573 23. Puranik S*, et al.* Structural basis for the oligomerization of the MADS domain transcription
574 factor SEPALLATA3 in Arabidopsis. *Plant Cell* **26**, 3603-3615 (2014).

575
576 24. Nilsson I, Saaf A, Whitley P, Gafvelin G, Waller C, von Heijne G. Proline-induced disruption of
577 a transmembrane alpha-helix in its natural environment. *J Mol Biol* **284**, 1165-1175 (1998).

578
579 25. Richardson JS. The anatomy and taxonomy of protein structure. *Adv Protein Chem* **34**, 167-
580 339 (1981).

581
582 26. Kim S, Soltis PS, Soltis DE. AGL6-like MADS-box genes are sister to AGL2-like MADS-box
583 genes. *J Plant Biol* **56**, 315-325 (2013).

584
585 27. Zhu BY, Zhou NE, Kay CM, Hodges RS. Packing and hydrophobicity effects on protein folding
586 and stability - effects of beta-branched amino-acids, valine and isoleucine, on the formation

587 and stability of 2-stranded alpha-helical coiled coils leucine zippers. *Protein Sci* **2**, 383-394
588 (1993).

589
590 28. de Folter S*, et al.* Comprehensive interaction map of the Arabidopsis MADS Box transcription
591 factors. *Plant Cell* **17**, 1424-1433 (2005).

592
593 29. Moitra J, Szilak L, Krylov D, Vinson C. Leucine is the most stabilizing aliphatic amino acid in
594 the d position of a dimeric leucine zipper coiled coil. *Biochemistry* **36**, 12567-12573 (1997).

595
596 30. Theißen G, Melzer R, Rümpler F. MADS-domain transcription factors and the floral quartet
597 model of flower development: linking plant development and evolution. *Development* **143**,
598 3259-3271 (2016).

599
600 31. Takei T*, et al.* The effects of the side chains of hydrophobic aliphatic amino acid residues in
601 an amphipathic polypeptide on the formation of a helix and its association. *J Biochem* **139**,
602 271-278 (2006).

603
604 32. Kaufmann K, Melzer R, Theißen G. MIKC-type MADS-domain proteins: structural modularity,
605 protein interactions and network evolution in land plants. *Gene* **347**, 183-198 (2005).

606
607 33. Ma H, Yanofsky MF, Meyerowitz EM. AGL1-AGL6, an Arabidopsis gene family with similarity
608 to floral homeotic and transcription factor genes. *Genes Dev* **5**, 484-495 (1991).

609
610 34. Fong JH, Keating AE, Singh M. Predicting specificity in bZIP coiled-coil protein interactions.
611 *Genome Biol* **5**, R11 (2004).

612
613 35. Potapov V, Kaplan JB, Keating AE. Data-driven prediction and design of bZIP coiled-coil
614 interactions. *Plos Comput Biol* **11**, (2015).

615
616 36. Perez K, Yeam I, Jahn MM, Kang BC. Megaprimer-mediated domain swapping for
617 construction of chimeric viruses. *J Virol Methods* **135**, 254-262 (2006).

618
619 37. Egea-Cortines M, Saedler H, Sommer H. Ternary complex formation between the MADS-box
620 proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture
621 in Antirrhinum majus. *Embo J* **18**, 5370-5379 (1999).

622
623 38. Senear DF, Brenowitz M. Determination of binding constants for cooperative site-specific
624 protein-DNA interactions using the gel mobility-shift assay. *J Biol Chem* **266**, 13661-13671
625 (1991).

626
627 39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol*
628 *Biol* **215**, 403-410 (1990).

629

630 40. Becker A, Theißen G. The major clades of MADS-box genes and their role in the development
631 and evolution of flowering plants. *Mol Phylogenet Evol* **29**, 464-489 (2003).

632
633 41. Gramzow L, Theißen G. Phylogenomics of MADS-Box genes in plants - two opposing life styles
634 in one gene family. *Biology (Basel)* **2**, 1150-1164 (2013).

635
636 42. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple
637 sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

638
639 43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
640 improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).

641
642 44. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees.
643 *Bioinformatics* **17**, 754-755 (2001).

644
645 45. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science* **252**,
646 1162-1164 (1991).

FIGURES

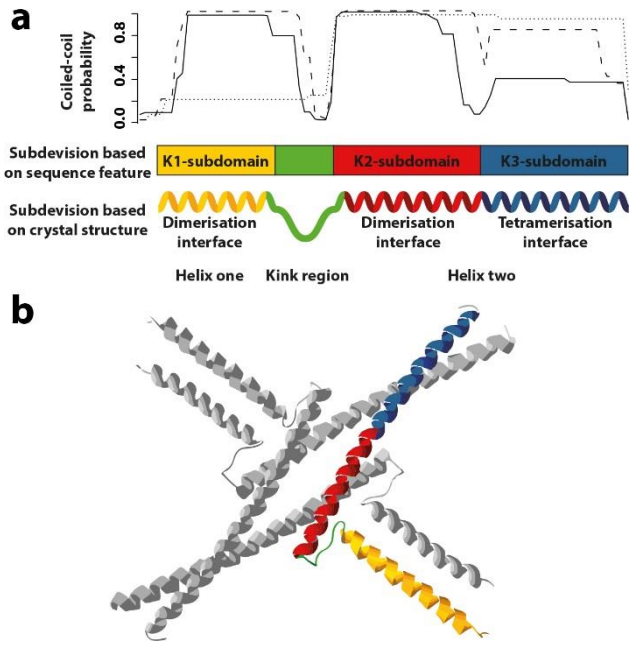

Figure 1. Domain architecture of the K-domain of SEP3 based on sequence and structural features. (**a**) Based on coiled-coil predictions (top) the K-domain was assumed to fold into three separate coiled-coils and was thus subdivided into three subdomains K1-, K2- and K3-subdomain (middle). The crystal structure of the K-domain of SEP3 revealed that it folds into two α-helices separated by a kink region (bottom). The first helix spans the K1-subdomain (color coded in yellow) and is involved in the dimerisation of two SEP3 monomers (i.e. dimerisation interface). The second helix spans the K2- and K3-subdomains and constitutes an N-terminal interaction interface that further stabilizes dimerisation of two SEP3 monomers (red) and a second C-terminal interaction interface that mediates the interaction of two SEP3 dimers (i.e. tetramerisation interface, blue). Coiled-coil predictions were performed with COILS [45]. The solid, dashed and dotted lines correspond to a sliding window size of 14, 21 and 28 amino acids used for the prediction, respectively. (**b**) Crystal structure of a SEP3 K-domain homotetramer [23]. The dimerisation interface of helix one, the kink region, the dimerisation interface of helix two and the tetramerisation interface of one K-domain are color coded in yellow, green, red and blue, respectively.

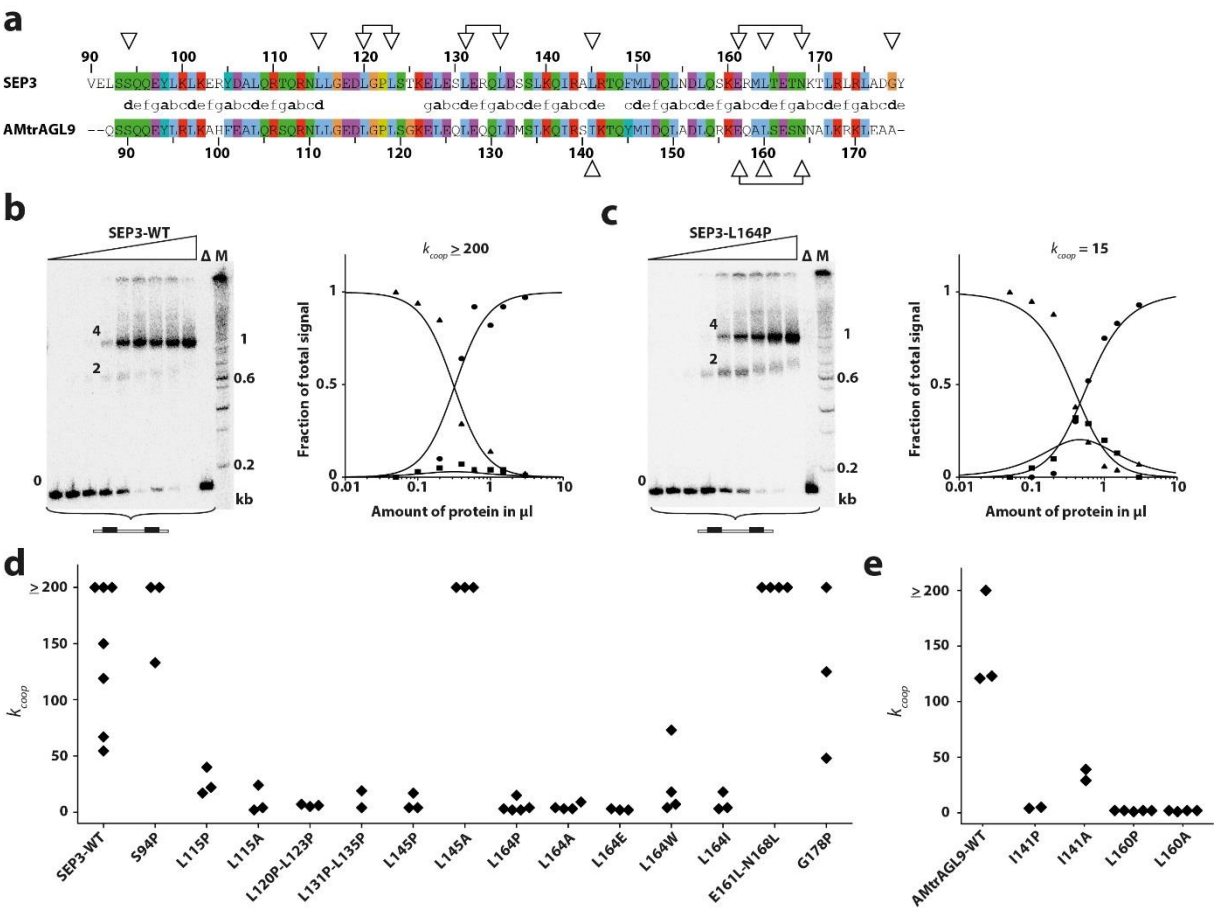

Figure 2. Ability of SEP3 and AMtrAGL9 wild type proteins and different amino acid substitution mutants to cooperatively bind to DNA. (**a**) Pairwise sequence alignment of the K-domains of SEP3 and AMtrAGL9. The heptad repeat pattern is depicted in the center. Positions at which amino acids were substituted are indicated by open triangles. (**b** and **c**) Binding of SEP3 wild type (**b**) and SEP3-L164P (**c**) to a DNA probe containing two CArG-boxes. Increasing amounts of *in vitro* translated protein were incubated with constant amounts of DNA probe. As negative control the empty pTNT vector without any cDNA insert was used as template DNA for the *in vitro* translation (lane Δ). For size comparison a radioactively labeled DNA ladder (100 bp Ladder, NEB) was applied (lane M). The labeling of the three different fractions ‘0’, ‘2’ and ‘4’ corresponds to the number of proteins bound to one DNA molecule. Quantified signal intensities of the different fractions and graphs, fitted according to equation (1) to (3) described in Methods, are shown next to the gel pictures (▲ free DNA; ■ DNA probe bound by two proteins; ● DNA probe bound by four proteins). The k_{coop} value inferred from this particular measurement is depicted above the diagram. (**d** and **e**) k_{coop} values for the wild type protein and all examined single and double amino acid substitution mutants of SEP3 (**d**) and AMtrAGL9 (**e**). k_{coop} values above 200 could not be determined reliably (see Methods).

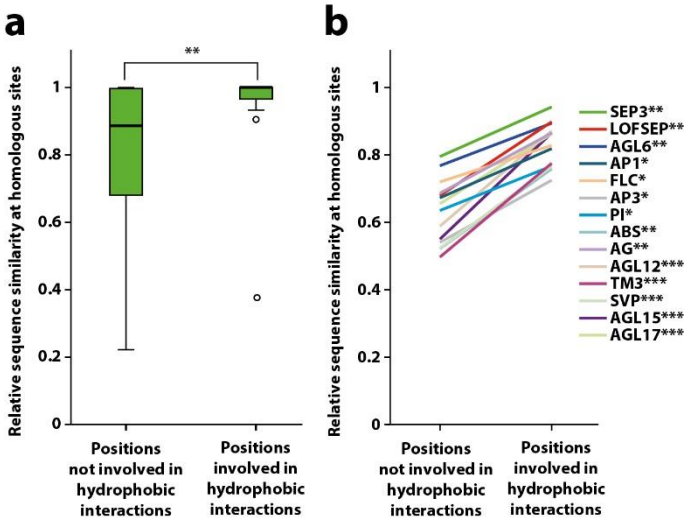

Figure 3. Sequence similarity analysis of SEP3-subfamily proteins and members of other MIKC-type MADS-domain protein subfamilies. (**a**) Box-plot showing relative sequence similarity at homologous sites of SEP3-subfamily proteins for positions that are involved in hydrophobic interactions within the SEP3 homotetramer and positions that are not involved in hydrophobic interactions. (**b**) Line graph showing the same analysis as in (**a**) but for all MIKC-type protein subfamilies. For all subfamilies amino acid positions that are homologous to sites involved in hydrophobic interactions are significantly less variable than positions that are homologous to non-interacting sites (Mann-Whitney-U-test; * $p = 0.01$-0.05; ** $p = 0.001$-0.01; *** $p < 0.001$).

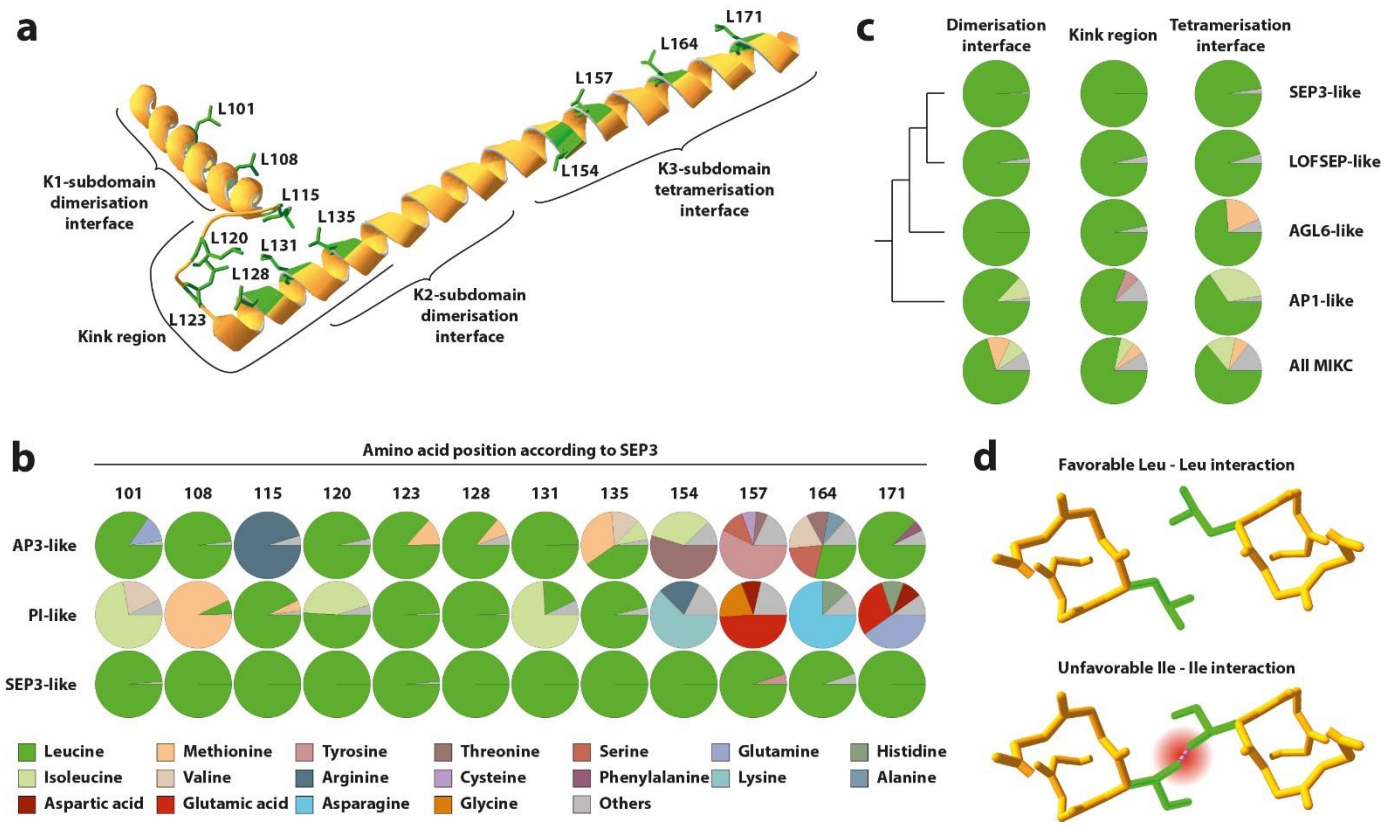

Figure 4. Amino acid preferences of SEP3-subfamily proteins and members of other MIKC-type MADS-domain protein subfamilies. (**a**) Picture of the crystal structure of a single K-domain of SEP3. Leucine side chains that are involved in inter- and intramolecular interactions are shown in green. (**b**) Amino acid frequencies at sites homologous to leucine residues that are involved in inter- and intramolecular interactions in the SEP3 homotetramer shown for SEP3-, AP3- and PI-subfamily proteins. Amino acids that occurred in less than 5 % of the examined subset of sequences were condensed as 'others'. The vast majority of the positions shown vertically are homologous to each other. The only exception are positions 154, 157, 164 and 171 of PI-like proteins. In this case, a gap was detected in the alignment but amino acids directly following the gap were included here. (**c**) Amino acid preferences at sites homologous to leucine residues that contribute to dimerisation interface (L101, L108), kink region (L115, L120, L123, L128, L131, L135) and tetramerisation interface (L154, L157, L164, L171) in the SEP3 homotetramer, shown for SEP3-, LOFSEP-, AGL6- and AP1-subfamily proteins and all MIKC-type proteins, respectively, following the color coding of panel B. (**d**) Part of the crystal structure of two interacting tetramerisation interfaces within a SEP3 homotetramer. The picture illustrates the favorable Leu-Leu interaction at heptad repeat 'd' positions as it becomes apparent several times within a SEP3 homotetramer (upper part). In contrast to the γ-branched leucine a β-branched amino acid such as isoleucine would potentially lead to steric hindrance at heptad repeat 'd' positions (lower part).

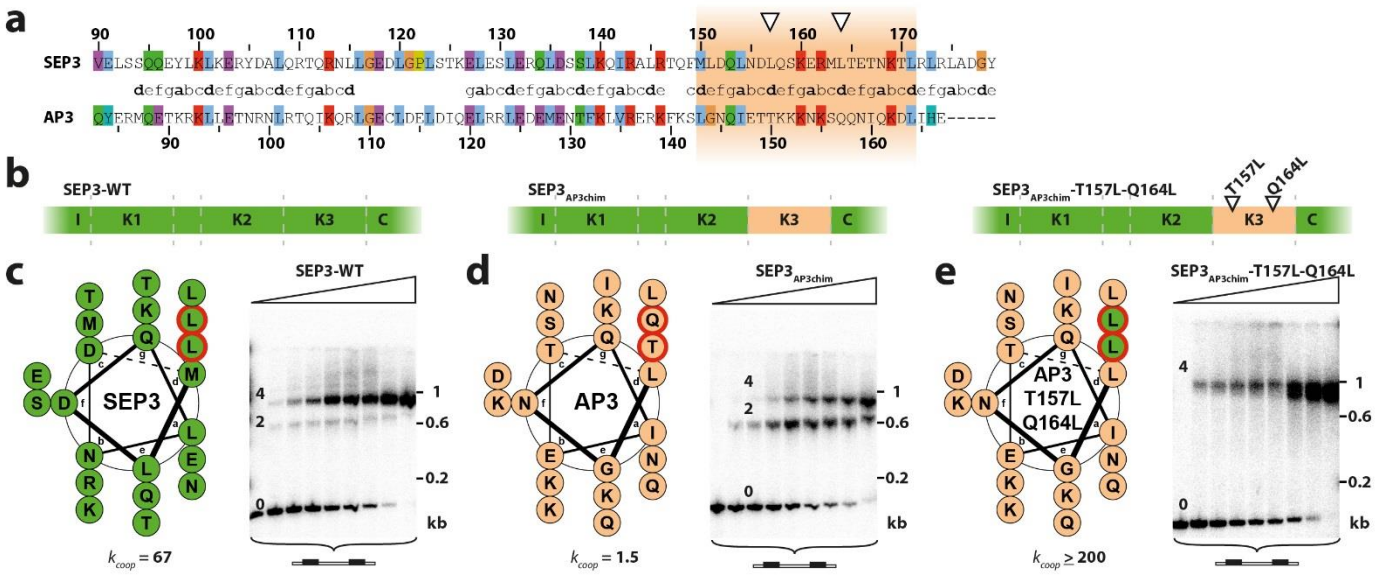

Figure 5. Design and cooperative DNA-binding capabilities of the chimeric protein SEP3$_{AP3chim}$. (**a**) Pairwise sequence alignment of the K-domains of SEP3 and AP3. The heptad repeat pattern is depicted in the center. Orange background marks the region that was substituted to create the chimeric protein SEP3$_{AP3chim}$. Open triangles mark the positions of the subsequently introduced amino acid substitutions. (**b**) Experimental setup to test for the ability of leucines to restore tetramerisation ability of SEP3$_{AP3chim}$. First the complete tetramerisation interface (i.e. the K3-subdomain) of SEP3 was substituted by the equivalent positions of AP3. Subsequently the two residues T157 and Q164 were substituted back to leucine to reestablish the hydrophobic stripe. (**c**-**e**, left) Helical wheel diagram of the tetramerisation interface of SEP3 wild type (**c**), SEP3$_{AP3chim}$ (**d**) and SEP3$_{AP3chim}$-T157L-Q164L (**e**), respectively, to illustrate the presumed position of amino acids 157 and 164 (framed in red) within the hydrophobic stripe of the K3-subdomain coiled-coil. (**c**-**e**, right) Binding of SEP3 wild type (**c**), SEP3$_{AP3chim}$ (**d**) and SEP3$_{AP3chim}$-T157L-Q164L (**e**) to a DNA probe containing two CArG-boxes. Increasing amounts of *in vitro* translated protein were incubated together with constant amounts of DNA probe. k_{coop} values inferred from this particular measurement are depicted below the helical wheel diagrams.