MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution

Günter Theißen1,*, Rainer Melzer2 and Florian Rümpler1

ABSTRACT

The floral quartet model of floral organ specification poses that different tetramers of MIKC-type MADS-domain transcription factors control gene expression and hence the identity of floral organs during development. Here, we provide a brief history of the floral quartet model and review several lines of recent evidence that support the model. We also describe how the model has been used in contemporary developmental and evolutionary biology to shed light on enigmatic topics such as the origin of land and flowering plants. Finally, we suggest a novel hypothesis describing how floral quartet-like complexes may interact with chromatin during target gene activation and repression.

KEY WORDS: ABC model, Floral homeotic gene, MIKC-type MADS-domain transcription factor, Nucleosome, Pioneer transcription factor, Plant evolution

Introduction

Flowers are frequently composed of four different classes of organs arranged in whorls, with sepals in the first floral whorl, petals in the second whorl, stamens (male reproductive organs) in the third whorl and carpels (female organs) in the fourth whorl. Understanding how these distinct floral organs are specified has been a long-standing challenge in plant developmental genetics (Meyerowitz et al., 1989; Schwarz-Sommer et al., 1990; Coen and Meyerowitz, 1991; Irish, 2010). According to the floral quartet model (FQM), which was proposed in 2001, the identity of the different floral organs is specified during development by quaternary (tetrameric) protein complexes composed of MIKC-type MADS-domain proteins (see Glossary, Box 1; Theißen, 2001). These quartets are assumed to function as transcription factors by binding to the DNA of their target genes, which they either activate or repress to control the development of the respective floral organs (Theißen, 2001).

MIKC-type MADS-box genes (see Glossary, Box 1; Münster et al., 1997; Kaufmann et al., 2005b) encode proteins that exhibit a characteristic domain organization that includes (from N- to C-terminus): a MADS (M) domain, an intervening (I) domain, a keratin-like (K) domain, and a C-terminal (C) domain (Theißen et al., 1996; Kaufmann et al., 2005b). The MADS domain is by far the most highly conserved region of all kinds of MADS-domain proteins, including MIKC-type proteins. It represents a DNA-binding domain but is also important for the dimerization and nuclear localization of MADS-domain transcription factors (Gramzow and Theißen, 2010). The I domain, by contrast, is only relatively weakly conserved and contributes to the selective formation of DNA-binding dimers (Kaufmann et al., 2005b). The K domain is characterized by a conserved, regular spacing of amphipathic helices involved in protein dimerization and multimeric complex formation (Yang et al., 2003; Puranik et al., 2014). Finally, the C domain is quite variable and, in some MADS-domain proteins, is involved in transcriptional activation or multimeric complex formation (for a review on structural and phylogenetic aspects of MIKC-type proteins, see Kaufmann et al., 2005b; Theißen and Gramzow, 2016).

MADS-domain proteins bind as dimers to DNA sequences termed ‘CArG-boxes’ (see Glossary, Box 1; reviewed by Theißen and Gramzow, 2016). According to the FQM, two protein dimers of each tetramer recognize two different CArG-boxes and bring them into close vicinity by looping the DNA between the CArG-boxes (Theißen, 2001; Theißen and Saedler, 2001). In recent years, the remarkable capacity of MIKC-type proteins to constitute multimeric transcription factor complexes, together with the importance of these complexes in plant development and evolution, has been increasingly recognized. However, the heuristic value of the FQM in plant developmental and evolutionary biology has not yet been fully explored. To stimulate further research, we revisit the FQM and review the current status of the field. We first provide a short history of the FQM, summarize its recent experimental support, and outline its use in current research. We also propose a simplified and generic version of the FQM that helps to harmonize genetic and molecular models of floral organ identity specification. Finally, we discuss major open questions regarding floral quartet-like protein complexes (FQCs; see Glossary, Box 1), concerning their molecular mode of action during the activation or repression of target genes.

A brief history of the floral quartet model

The scientific journey that eventually led to the development of the FQM started with the analysis of mutants in which all or some organs of the flower had been replaced with organs of another identity, a phenomenon known as ‘homeosis’. Mutational changes in floral organ identity have been known from many species and have fascinated humans for over centuries (Meyerowitz et al., 1989). It turned out that many of the respective mutants, termed floral homeotic mutants, including those of Arabidopsis thaliana (thale cress) and Antirrhinum majus (snapdragon), fall into three classes, termed A, B and C (Bowman et al., 1991; Coen and Meyerowitz, 1991). In ideal class A mutants, sepals are replaced by petals and carpels are substituted by sepals. In class B mutants, sepals instead of petals and carpels instead of stamens develop. In class C mutants, stamens are replaced by petals and carpels are substituted by sepals. The typical determinate growth of flowers is also often abolished in class C mutants, so that a potentially unlimited series of additional mutant flowers develops inside the primary mutant flower.
Based on these frequently found classes of homeotic mutants, simple genetic models were proposed and successfully tested by analysing double and triple mutants (for a historical perspective, see Theißen, 2001; Theißen and Melzer, 2006; Causer et al., 2010; Irish, 2010; Bowman et al., 2012). Arguably the most well-known of these models is the ‘ABC model’ as outlined by Bowman et al. (1991) and Coen and Meyerowitz (1991). It maintains that organ identity in each whorl is specified by a unique combination of three homeotic functions, termed A, B and C, which are accomplished by floral organ identity genes. Expression of the A function alone specifies sepal formation. The combination AB specifies the development of petals, while the combination BC specifies the formation of stamens. Expression of C alone determines the identity of the flower. The ABC model was extended to an MADS-box domain protein. A MADS-domain protein that exhibits a characteristic domain structure including a DNA-binding MADS (M) domain, an Intervening (I) domain, a keratin-like (K) domain and a C-terminal (C) domain.

**Pioneer transcription factor (PTF).** A transcription factor that can bind to nucleosome-associated DNA sites, possibly by eviction nucleosomes.

ABC genes were expressed outside the floral context they could not, in most cases, induce floral organ development from leaf primordia (Krizek and Meyerowitz, 1996a,b; Mizukami and Ma, 1992; Pelaz et al., 2001). It turned out that additional homeotic functions had escaped forward genetic approaches. Indeed, based on studies in petunia (Petunia hybrida), the ABC model was extended to an ‘ABCD’ model by addition of a D function specifying ovule identity (Angenent and Colombo, 1996). In A. thaliana, three genes closely related to AG, namely SEEDSTICK (STK; formerly known as AGL11), SHATTERPROOF1 (SHP1; formerly known as AGL1) and SHATTERPROOF2 (SHP2; formerly known as AGL5) (Favaro et al., 2003; Pinyopich et al., 2003) were identified as D-function genes. Several lines of data (Honma and Goto, 2001; Pelaz et al., 2001) strongly suggested that the ABCDE genes are not only necessary, but also sufficient to superimpose floral organ identity upon vegetative developmental programs of angiosperms (see Glossary, Box 1), even though it has remained unclear up to now as to whether C+D genes suffice to generate carpels (Battaglia et al., 2006). Like the ABC model, the ABCDE model relied mainly on genetic data. This raised questions with regards to the molecular mechanism by which the different floral homeotic genes interact. For example, how B and C class proteins interact to specify stamen identity remained elusive (Riechmann et al., 1996), and all attempts to fully explain the interactions of the floral homeotic genes and functions just by the dimerization of floral homeotic proteins were not successful. The inability to answer these questions was seen as another major shortcoming of the ABC model and its derivatives (Theißen, 2001). Overcoming this limitation required a switch from...
The floral quartet model and the underlying ABCDE model of organ identity determination in Arabidopsis thaliana. The top part of the figure depicts a version of the floral quartet model, which maintains that the five floral organ identities (sepals, petals, stamens, carpels and ovules) are specified by the formation of floral organ-specific tetrameric complexes of MIKC-type MADS-domain transcription factors that bind to two adjacent cis-regulatory DNA binding sites (CArG-boxes, green) and loop the DNA (blue) in between. A complex of two APETALA1 (AP1) class A proteins and two SEPALLATA (SEP) class E proteins determines sepal identity, a complex of one AP1 and one SEP protein together with one of each of the class B proteins APETALA3 (AP3) and PISTILLATA (PI) determines petal identity, a complex of one SEP, one AP3, one PI protein and the class C protein AGAMOUS (AG) determines stamen identity, a complex of two SEP proteins together with two AG proteins determines carpel identity, and a complex of one SEP and one AG protein together with one of each of the class D proteins SHATTERPROOF (SHP) and/or SEEDSTICK (STK) controls ovule identity. The bottom part of the figure illustrates the genetic ABCDE model. According to this model, organ identity during flower development in A. thaliana is controlled by five sets of floral homeotic genes providing overlapping floral homeotic functions: A, B, C, D and E. Class A genes are expressed in the organ primordia of the 1st and the 2nd whorl of the flower, class B genes in the 2nd and 3rd whorl, class C genes in whorls 3 and 4, class D genes in parts of the 4th whorl (ovule primordia), and class E genes are expressed throughout all four whorls. Class A and E genes specify first whorl sepals, class A, B and E genes specify second whorl petals, class B, C, and E genes specify third whorl stamens, class C and E genes specify fourth whorl carpels, and class C, D and E genes control the development of the ovules within the fourth whorl carpels.

Recent experimental evidence supporting the floral quartet model

The FQM was rapidly accepted in the literature (see, e.g. Jack, 2001; Eckardt, 2003; Ferrario et al., 2004; Jack, 2004; Krizek and Fletcher, 2005; Baum and Hileman, 2006), suggesting that it was plausible and not in conflict with major evidence at the time of its inception. In addition, a number of protein interaction studies in yeast using proteins from different flowering plant species, such as...
tomato, petunia, chrysanthemum, gerbera and rice, demonstrated that floral homeotic proteins could form multimers (e.g. Ferrario et al., 2003; Favaro et al., 2003; Shchennikova et al., 2004; Yang and Jack, 2004; Kaufmann et al., 2005a; Leseberg et al., 2008; Ruokolainen et al., 2010; Seok et al., 2010). Additional experimental evidence supporting the FQM, however, remained scarce for a while. In recent years, this has changed considerably. Diverse experimental approaches comprising analyses in vitro, in vivo, in planta and in silico have contributed to the view that floral quartets really exist and play an important role in controlling plant development.

An early experiment that provided evidence for the formation of multimeric complexes of MIKC-type MADS-domain proteins in plant cells employed different fusions between petunia MIKC-type proteins and yellow fluorescent protein (YFP) or cyan fluorescent protein (CFP). When two fusion proteins that dimerize only weakly were coexpressed in petunia protoplasts with a third, unlabelled MIKC-type protein, strong fluorescence resonance energy transfer (FRET) was observed, suggesting that a higher-order complex, as predicted by the FQM, had been formed (Nougali-Tonaco et al., 2006). Next, in a series of experiments employing electrophoretic mobility shift assays (EMSA) and DNase I footprint assays, it was demonstrated that FQCs can be reconstituted from a limited number of components in vitro. Initial experiments revealed that not even a combination of different MIKC-type proteins is required to obtain FQCs; the class E floral homeotic protein SEP3 from A. thaliana shows an intrinsic capacity to cooperatively bind as a tetramer to two CArG-boxes (Melzer et al., 2009). The spacing and phasing of CArG-boxes influence the efficiency of FQC binding: binding occurs better if the CArG-boxes are separated by an integer number of helical turns (Melzer and Theißen, 2009; Melzer et al., 2009). In this context, the two CArG-boxes are in the same orientation, so that bending and looping, but not twisting, of the DNA is required when a MIKC-type protein tetramer binds. The ability of a SEP3 homotetramer to loop the DNA sequence separating the two binding sites supports some of the major tenets of the FQM. In follow-up experiments, it was shown that the other three SEP proteins (SEP1, SEP2 and SEP4) of A. thaliana also constitute FQCs involving protein homotetramers under suitable conditions in vitro (Jetha et al., 2014). All four SEP proteins bind to CArG-boxes in a similar way, and yet they also show subtly distinct DNA-binding properties. For example, the cooperativity of DNA binding differs among the different SEP proteins, with SEP3 often showing the least cooperativity (Jetha et al., 2014). It was also shown that all SEP proteins prefer surprisingly short distances of 4-6 helical turns (~42-63 nucleotides) between the CArG-boxes (Jetha et al., 2014). Remarkably, the optimal distance was shown to differ in in vitro experiments, with SEP2 preferring relatively large distances and SEP4 preferring small distances; SEP1 binds well to CArG-box pairs separated by a relatively broad range of distances (Jetha et al., 2014). It is conceivable that FQCs involving SEP proteins alone have a function in the development of flowering plants, but conclusive evidence for that is missing so far (Melzer et al., 2009; Melzer and Theißen, 2009) and other studies instead suggest that SEP proteins act as a kind of ‘glue’ in interactions of MIKC-type proteins (Immink et al., 2009). It was further shown that complexes composed of SEP3, AP3 and PI form preferentially over SEP3 homotetramers, suggesting a mechanism that would allow different target genes to be activated at different developmental stages (Melzer and Theißen, 2009). In addition, the ectopic expression of SEP3, together with the class B proteins AP3 and PI, is sufficient to induce the development of petals from primordia that would normally develop into vegetative leaves (Honma and Goto, 2001), highlighting that FQCs involving SEP3, AP3 and PI represent a minimal set of master control elements governing floral organ (petal) identity (Melzer and Theißen, 2009).

Data supporting FQC formation in planta has also been published. Identifying protein complexes isolated from transgenic plants by affinity purification followed by mass spectrometry and label-free quantification, Smaczniak et al. (2012b) collected data strongly suggesting that the five major floral homeotic MIKC-type proteins that were tested as baits – AP1 (A function), AP3 and PI (B function), AG (C function) and SEP3 (E function) – interact in floral tissues as proposed by the FQM, even though the data do not provide unequivocal evidence that exactly tetramers form in planta. Moreover, some tetramers of MIKC-type proteins appear to be able to bind to single CArG-boxes (see, e.g. Melzer et al., 2009; Smaczniak et al., 2012b). As such, an important aspect of the FQM – the looping of regulatory DNA of target genes bound by tetramers of MIKC-type proteins (Theißen, 2001; Theißen and Saedler, 2001) – remained untested in planta. Not much later, however, Mendes et al. (2013) reported a series of experiments in favour of FQC formation involving DNA looping. Employing the single-molecule in vitro method of tethered particle motion (TPM), the authors studied binding of the floral homeotic proteins STK (class D) and SEP3 (class E) to a fragment of the promoter region of VERDANDI (VDD), which is a direct target gene of STK that contains three CArG-boxes up to 444 bp apart. The data strongly suggested that loop formation indeed occurs and that FQC formation clearly favours one pair of CArG-boxes (CArG-box 1+CArG-box 3) over alternative combinatorial possibilities for protein binding (Mendes et al., 2013). Using promoter-reporter gene fusions, the authors also studied the functional importance of different CArG-boxes in transgenic A. thaliana plants, demonstrating that single CArG-boxes are not sufficient to drive VDD gene expression in planta, and that both CArG-boxes 1 and 3 are required to establish the typical VDD gene expression pattern. Together with chromatin immunoprecipitation (ChIP) studies demonstrating that STK and SEP3 preferentially bind to CArG-boxes 1 and 3 in the VDD promoter region, these findings suggest that FQCs involving STK, SEP3 and CArG-boxes 1 and 3 assemble in the VDD promoter region and are involved in controlling gene expression in planta. These findings provide remarkable in vivo evidence for the FQM, even though alternative scenarios have not been completely ruled out so far.

Additional support for the FQM has been provided by structural biology studies. Some EMSA experiments had demonstrated that the C-terminal half of the K domain, which was assumed to form an amphipathic α-helix involved in the formation of a coiled-coil, is of crucial importance for MIKC-type protein tetramerization (Melzer et al., 2009; Melzer and Theißen, 2009). Recent X-ray crystallography studies of the K domain of A. thaliana SEP3 revealed that the K domain forms two amphipathic α-helices separated by a rigid kink, which prevents intramolecular association (Puranik et al., 2014). The K domain thus provides two separate interaction interfaces to facilitate dimerization and tetramerization with other K domains (Puranik et al., 2014). Atomic force microscopy (AFM) further demonstrated the looping of target DNA by SEP3 and even allowed FQCs to be ‘seen’ for the first time (Puranik et al., 2014).

Last, but not least, recent in silico analyses have provided support for the FQM. Network-based analyses of the known physical interactions between MADS-domain proteins from A. thaliana (as
revealed by yeast two-hybrid and three-hybrid assays) indicated that the formation of functional tetramers is a widespread property of *A. thaliana* MIKC-type proteins, but not of non-MIKC-type MADS-domain proteins i.e. those that lack a K domain (Espinoza-Soto et al., 2014). Given that all floral organ identity proteins (ABCD/E proteins) of the MADS-domain family are MIKC-type proteins, and that MIKC-type proteins have a tendency to tetramerize (even though not all of them may actually do so), it appears even more likely that the combinatorial interactions of the different homeotic genes predicted by the ABCDE model are indeed realized by the tetramerization of MIKC-type floral homeotic proteins.

The findings reviewed above, however, do not imply that all MIKC-type proteins exert their function only as constituents of tetrameric complexes. Several lines of evidence, such as ChIP-seq data of protein binding *in vivo* (Kaufmann et al., 2009; Kaufmann et al., 2010a,b), suggest that dimers of MIKC-type proteins are also of functional importance, and it also appears likely that at least some dimers and tetramers exist in dynamic equilibria.

**The FQM as guiding model in current research**

The heuristic value of the FQM is revealed by its use as a guiding model in current research. For example, the destruction of floral quartets has been proposed to cause the development of the often bizarre symptoms observed in plants infected by the bacterial pathogen phytoplasma (Maejima et al., 2014). One characteristic phenotype (‘phyllody’) of phytoplasma-infected plants from diverse species, including *A. thaliana*, resembles the phenotype of class E floral homeotic mutants, with floral organs unable to develop proper floral organ identity. This was recently shown to be due to proteasome-mediated degradation of the class A and E floral homeotic proteins AP1, CAULIFLOWER (CAL) and SEP3, which is initiated by interaction of the floral homeotic proteins with phytoplasma-secreted effector proteins termed SAP54 or PHYL1 (Maejima et al., 2014; MacLean et al., 2014). An *in silico* study suggests that the PHYL1 structure resembles that of the K domain, thus facilitating dimerization between some floral homeotic proteins and PHYL1 (Rümpler et al., 2015). The authors hypothesized that the similarity between PHYL1 and the K domain represents a case of convergent evolution (‘molecular mimicry’) that evolved to enable phytoplasmas to manipulate their host plants according to their needs. Maejima et al. (2014) noted that the strength of phenotype in the different floral whorls of plants expressing PHYL1 (severe in the first whorl, medium in whorl 2, weak in whorl 3 and again medium in the fourth whorl) correlates perfectly with the number of class A and E proteins (4, 2, 1, 2) in the floral quartets of whorls 1, 2, 3, and 4, respectively; they thus argue that the floral quartet model provides the basis for an explanation of the whorl-specific differences in the strength of phenotype in affected plants.

Diversification of the floral quartet that specifies petal identity has been used to explain the differences in the petaloid organs of orchids. Orchids typically have four paralogous *AP3*-like genes, in contrast to the one *AP3*-like class B gene found in *A. thaliana* and *A. majus*. According to the ‘orchid code hypothesis’, sub- and neo-functionalization involving differential expression of these genes led to a combinatorial system that specifies the identity of the different petaloid perianth organs, i.e. outer tepals (also called ‘sepals’) in the first floral whorl, and inner lateral tepals (‘petals’) and the labellum (‘lip’) in the second whorl (Mondragón-Palomino and Theißen, 2008, 2011). Recently, Hsu et al. (2015) reported several lines of evidence suggesting that competition between two floral quartets decides whether outer and inner lateral tepals (‘sepaloids’) or lips develop. Both floral quartets contain one protein encoded by the single *PI*-like (class B) gene, but different paralogs of *AP3*-like (class B) and *AGL6*-like genes (that may function in orchids as class E genes).

Floral quartets have also been used to explain how the interesting phenomenon of paralog interference can affect the evolutionary dynamics of genes after duplication (Kaltenegger and Ober, 2015). When proteins function in obligate homomeric complexes of identical subunits, duplication of the gene encoding these proteins generates paralogous genes whose gene products may cross-interact when co-expressed, thus resulting in paralog interference. Since independent mutations in the different gene copies may interfere with protein interaction and function, and hence may bring about a dominant negative effect, both copies are expected to remain under purifying selection during a prolonged time window. This increases the chance that they accumulate mutations that lead to novel properties of the different paralogous proteins. In line with this, positive selection may occur, creating asymmetric protein dimers or multimers that may contribute to evolutionary novelties or innovations. While Kaltenegger and Ober (2015) focused their discussion on the obligate heterodimerization of class B proteins within some floral quartets of angiosperms, it is tempting to speculate that paralog interference played an important role during the expansion and diversification of all kinds of MIKC-type genes and FQCs throughout the evolution of land plants.

**The (A)B(C)s of floral quartets**

When the simple and elegant ABC model developed into the more elaborate ABCDE model, the FQM was proposed to explain the interactions between floral homeotic genes and proteins, but it also intended to resimplify matters (Theißen, 2001). Recent improvements to the ABCDE model enable the FQM to further harmonize the genetic and the molecular models.

In contrast to the genetically, phylogenetically and developmentally quite well-defined B and C floral homeotic functions, the concept of A function has been considered controversial for almost as long as the ABC model itself (Theißen et al., 2000; Litt, 2007; Causier et al., 2010). One reason is that in almost all plants that have been investigated so far, with *A. thaliana*
A. majus

In fact, an early alternative to the ABC model that was focused on a more fundamental function in specifying floral meristem identity. Antagonizing the C function is difficult to separate genetically from function in specifying perianth (sepal and petal) organ identity and ill-defined (Litt, 2007; Causier et al., 2010). For example, an A function in combination with a B function, respectively, of the ABC model (Causier et al., 1990), with A and B being equivalent to the class B and C function, respectively, of the ABC model (Causier et al., 2010). In this alternative model, sepal development represents the ‘default state’ of floral organ development and hence does not require a specific floral homeotic function (Schwarz-Sommer et al., 1990).

To resolve controversies surrounding A function, Causier et al. (2010) suggested an (A)BC model with (A) function controlling both floral meristem identity (the ‘floral ground state’) and floral organ identity in the first two floral whorls. According to the (A)BC model, (A) function also comprises the E function of the ABCDE model i.e. (A)=A+E. According to Causier et al., (A) function is provided by a group of genes, but if one focusses on the MADS-box genes involved – in the case of A. thaliana the class A gene AP1 and the class E genes (senso lato), i.e. the SEP genes and the AGL6-like genes (Mandel et al., 1992; Pelaz et al., 2000; Ditta et al., 2004; Rijpkema et al., 2009; Hsu et al., 2014) – one finds some support for the new (A) function in gene phylogeny. All of these genes are
relatively closely related members of a gene superclade (Gramzow and Theißen, 2010, 2013, 2015; Ruelens et al., 2013), and it is thus conceivable that the A and E functions known from flowering plants trace back to an ancestral function in specifying reproductive meristem identity (Box 2). Similarly, there is evidence that the C and D functions of angiosperms trace back to a combined C/D function provided by 4G-like genes in extant gymnosperms and stem group seed plants (Box 2; Gramzow et al., 2014). Hence in analogy to (A) function, one may define (C) function, with (C)=C+D, yielding an (A)B(C) model for the angiosperm flower. The (C) function may specify reproductive organ identity, and its expression may distinguish reproductive from non-reproductive organs. Based on these considerations, one can transform a generalized ABCDE model into a more simple (A)B(C) model (Fig. 2). Note that the model shown is a generic model, and that the genes contributing to these functions may have been differentially sub- and neo-functionalized in different groups of angiosperms. This hampers interspecific comparisons and might be one reason for some controversies about A/E and C/D functions in the literature (see, e.g., Litt, 2007).

‘Translating’ the (A)B(C) model into a model based on FQCs, one gets a generic floral quartet model (Fig. 2, perspective 2) with four (A) proteins specifying floral meristem identity and sepal, two (A)+two B proteins specifying petals, one (A)+two B+one (C) proteins specifying male reproductive organs (stamens), and two (A)+two (C) proteins specifying female reproductive organs (carpels including ovules). Thus, after somewhat of a detour, the ABC model regains its simplicity as an (A)B(C) model, and the FQM has also been generalized and simplified (e.g. Fig. 2, compare 1 and 2). Given that (A), B and C genes probably already existed in the most recent common ancestor (MRCA) of extant seed plants, the generic FQM has obvious consequences for understanding the origin of the angiosperm flower.

**Box 3. FQCs: beyond floral organ identity**

Given that the formation of functional tetramers is a widespread property of *A. thaliana* MIKC-type proteins (Puranik et al., 2014; Espinosa-Soto et al., 2014), we hypothesize that FQCs play important roles beyond floral organ identity specification. Indeed, several studies have suggested that MIKC-type proteins other than the canonical class A-E floral homeotic proteins of the FQM can form FQCs. For example, members of the B superclass are involved in specifying the endothelium, and in case of *A. thaliana*, there is evidence that SHP proteins and/or STK, SEP3 and the B superclass protein ARABIDOPSIS BSISTER (ABS, also known as TT16 and AGL32) are components of a FQC that specifies this identity (Becker et al., 2002; Nesi et al., 2002; Kaufmann et al., 2005a; Mizzoti et al., 2012; Theißen and Gramzow, 2016). Several MIKC-type genes have also been implicated in fruit development, and it is reasonable to assume that FQCs also play a role in this context. For example, FQCs are involved in the development of the fleshy fruits of tomato (*Solanum lycopersicum*); such proposed FQCs were shown to contain MIKC-type proteins, such as the SIMADS1-like protein JOINTLESS, that are part of clades not belonging to those containing floral homeotic proteins (Liu et al., 2014; Fujisawa et al., 2014). Smaczniak et al. (2012b) also identified complexes of several other MIKC-type proteins, in line with the hypothesis that they too are involved in protein tetramerization and FQC formation. For example, complexes of AP1 and the TM3-like protein SUPPRESSOR OF OVEREXPRESSON OF CONSTANS 1 (SOC1), and of SOC1 and FUL were detected; both complexes might be part of FQCs involved in the transition to flower (Smaczniak et al., 2012b). These findings support the view that FQCs with protein compositions other than those described by the FQM play a role in processes other than floral organ identity specification.

Floral organ identity does not develop without the proper activity of floral homeotic genes. It appears reasonable, therefore, that understanding the evolution of floral quartets is key to understanding the origin of the angiosperm flower – a scientific problem closely related to the origin of the angiosperms, which has been popularized as Darwin’s ‘abominable mystery’ (Theißen and Saedler, 2001; Friedman, 2009). So how did floral homeotic genes of the MIKC type originate, and how did they start to constitute floral quartets? Early studies had already documented a strong correlation between the evolution of MIKC-type genes and the origin of evolutionary novelties, including floral organs, in land plants (Theißen and Saedler, 1995; Purugganan et al., 1995; Theißen et al., 1996, 2000; Becker and Theißen, 2003). The phylogeny of MIKC-type genes is characterized by the formation of ancient paralogs, many of which originated by whole genome duplications, preferential gene retention after duplication, and sequence divergence resulting in sub- and neo-functionalization (Gramzow and Theißen, 2013, 2015; Theißen and Gramzow, 2016). Radiations of genes occurred independently in different groups of land plants. Even though diverse MIKC-type MADS-box genes are involved in the control of many developmental processes in angiosperms, and probably also in all other land plants (Smaczniak et al., 2012a; Gramzow and Theißen, 2010), the floral homeotic genes are all members of gene clades that are seed plant- or flowering plant-specific.

Recent phylogeny reconstructions involving the first whole-genome sequence data from conifers (gymnosperms) suggest that MIKC-type genes of seed plants are all members of 11 seed plant-specific superclades that were present in the MRCA of extant seed plants about 300 million years ago (MYA), but that did not yet exist in the MRCA of monilophytes (ferns and their allies such as horsetails) and seed plants (gymnosperms and angiosperms) about 400 MYA (Nystedt et al., 2013; Gramzow et al., 2014). Among species plants (angiosperms and gymnosperms) together with ferns represent the clade of vascular plants; vascular plants and mosses represent the clade of land plants; and all land plants together with charophytes (a division of freshwater green algae) build up the clade of streptophytes. The K domain most likely emerged (indicated by *) prior to the split between extant charophytes and land plants and is thus a synapomorphy of streptophytes. There is experimental evidence that at least some MIKC-type proteins of angiosperms and gymnosperms can form FQCs, whereas MADS-domain proteins from charophytes lack a K domain and thus may bind to DNA only as dimers, as indicated by the icons next to the names of the different plant groups. It is not yet known if FQCs exist in charophytes, mosses or ferns (indicated by ?). Whether FQC formation is directly linked to the emergence of the K domain and is a synapomorphy of streptophytes, or if it occurred later during land plant evolution, is therefore also unknown.
these superclades are those containing, besides other genes, genes providing floral homeotic A function (FLC/SQUA-like, or FLC/AP1-like genes), class B genes (DEF/GLO/OsMADS12-like, or AP3/PI/OhMADS2-like genes), class C genes (AG-like genes) and class E genes (SEP/AGL6-like genes) (Box 2). Based on gene or even whole genome duplications in the stem group of angiosperms, the 11 superclades evolved into 17 clades that had already been established in the MRCA of extant angiosperms, including distinct DEF (AP3)- and GLO (PI)-like genes (class B), the AG-like and STK-like genes (classes C and D), the AGL2-like (SEP-like) and AGL6-like genes (class E), and the SQUA (API)-like genes (class A) (Theißen et al., 1996, 2000; Becker and Theißen, 2003; Ruelens et al., 2013; Gramzow et al., 2014).

It has been shown that some putative DEF/GLO-like (class B) and AG-like (class C/D) MIKC-type proteins from gymnosperms can alone constitute FQCs that may specify male and female reproductive cone development (Wang et al., 2010). Moreover, early phylogeny reconstructions suggested that combined DEF/GLO-like (class B) and AG-like (class C) genes, but no SQUA-like (class A) and SEP-like (class E) genes, existed in the MRCA of extant seed plants. Even though AGL6-like genes had been found in diverse extant gymnosperms (Winter et al., 1999), the function of these genes was, at that time, unknown even in angiosperms. These findings led to the view that the origin of SEP-like genes and the incorporation of SEP-like proteins into FQCs have been important steps during the origin of floral quadrants, and hence floral organ identity and flower development (Fig. 2, perspective 1; Zahn et al., 2005; Baum and Hileman, 2006; Silva et al., 2016). However, recent experimental data from different species suggest that not only SEP-like but also AGL6-like genes can exert the E function (Thompson et al., 2009; Rijpkema et al., 2009; Hsu et al., 2014) and phylogeny reconstructions suggest that the genomes of extant conifers and the MRCA of extant seed plants contain(ed) orthologs of floral homeotic class A and E genes (Gramzow et al., 2014). It is conceivable, therefore, that FQCs quite similar to those of extant floral quadrants also exist in extant gymnosperms and were already established in the MRCA of extant seed plants (Fig. 2, perspective 2). Specifically, and in contrast to previous views (Fig. 2, perspective 1; Zahn et al., 2005; Theißen and Melzer, 2007) that proposed that the incorporation of SEP-like proteins into FQCs played an important role during the origin of the flower, we consider it more likely now that the FQCs specifying male and female reproductive cone identity in ancestral and extant gymnosperms very much resemble(d) those of angiosperms, in that they contain(ed) AG-like proteins [(C) function] (female cones) or AG-like and DEF/GLO-like proteins [(B) function] (male cones) as well as SEP/AGL6-like and/or FLC/AP1-like proteins [(A) function] (Fig. 2, perspective 2).

The differences between the two hypotheses on the origin of floral quadrants are obviously of heuristic relevance. Assuming that changes in the composition of FQCs played an essential role during the evolution of the flower may inspire investigations into the evolution of MIKC-type proteins interactions in seed plants (e.g. Wang et al., 2010; Melzer et al., 2014). However, if one hypothesizes that the FQCs specifying male and female reproductive organ identity in gymnosperms and angiosperms did not change substantially during the origin of the flower, one may conclude that changes in the interactions between the FQCs specifying reproductive organ identity and their target genes have been of special importance during the origin of the flower. If so, comparison of the target genes of the MIKC-type proteins in FQCs in extant gymnosperms and angiosperms would be most revealing. While target genes for several A. thaliana floral homeotic proteins have already been determined, for example by employing techniques such as ChIP-seq (Kaufmann et al., 2009, 2010b; Ö Maoilléidigh et al., 2013; Wuest et al., 2012), respective data for gymnosperms are still missing.

**The origin of FQCs: a MIKC blessing**

The floral quartet model describes the interaction of the floral homeotic proteins at the molecular level. However, floral homeotic proteins represent only a minor fraction of the MIKC-type protein family, which comprises 45 members in A. thaliana alone (Becker and Theißen, 2003; Parenicová et al., 2003). Therefore, the question arises as to whether multimerization is restricted to the floral homeotic proteins of eudicots (an extreme hypothesis) or is a common feature of all MIKC-type proteins in all kinds of land plants (another extreme hypothesis) (Kaufmann et al., 2005b). Based on rapidly growing empirical evidence, we hypothesize that FQCs play an important role far beyond floral organ identity specification in A. thaliana (see Box 3). This raises the question as to when and where during evolution FQC formation began. Intriguingly, in contrast to many other multimeric complexes of transcription factors, the key protein constituents of floral quadrants are all encoded by paralogous MIKC-type genes. This corroborates the view that duplications of ancestral MIKC-type genes are intimately interlinked with the evolution of developmental complexity in plants. Thus, the question arises as to how the origin of MIKC-type proteins and of their ability to constitute FQCs are linked.

A variety of *in vitro* experimental data has demonstrated that the K domain is essential for mediating the interactions that are necessary for FQC formation (Yang et al., 2003; Yang and Jack, 2004; Melzer and Theißen, 2009; Melzer et al., 2009). As explained above, the K domain provides the structural basis on which FQC formation takes place (Purakí et al., 2014). It thus appears that the emergence of the K domain – with two distinct interaction interfaces that facilitate both

**Box 4. Why quartets?**

Why do many MIKC-type transcription factors bind to the DNA of their target genes as tetramers (quartets) rather than as independent dimers, as is the case for many other MADS-domain proteins? One important difference between tetramers and two dimers binding to DNA is the increased cooperativity in DNA binding. This cooperativity creates a sharp transcriptional response, i.e. even small increases in protein concentration can lead to drastic changes in regulatory output (Georges et al., 2010). Floral homeotic proteins as well as many other MIKC-type proteins act as genetic switches that control discrete developmental stages, and cooperative DNA binding might be one important mechanism that translates the quantitative nature of biomolecular interactions into discrete phenotypic outputs (Theißen and Melzer, 2007; Kaufmann et al., 2010a). Tetramer formation could also potentially incorporate different signals and thereby increase the robustness of the system. If one protein component of the tetramer is missing, the entire complex will not form or will be greatly destabilized and the developmental switch will not occur (Whitty, 2008). The formation of tetramers might also, in principle, contribute to an increase in target gene specificity. It was previously shown that different tetramers have different DNA-binding affinities, and that different tetramers may prefer different CArG-box distances for maximum binding (Melzer and Theißen, 2009; Jetha et al., 2014). This offers the possibility to differentially regulate target genes even in the absence of differential DNA-binding of MIKC-type protein dimers (Georges et al., 2010). We are still in the early days of exploring the developmental and evolutionary relevance of cooperative DNA binding and FQC formation. Plants expressing mutant proteins defective specifically in cooperative DNA binding will hopefully yield additional insights into how and why quartet formation is essential for MIKC-type proteins to act as developmental switches.
Box 5. The ‘nucleosome mimicry’ model of FQC action

We hypothesize that FQCs represent sequence-specific transcription factors with (half-) nucleosome-like properties that help to establish permissive or repressive chromatin modifications at CArG-box-containing promoters. A permissive, gene-activating case is illustrated below. In the first step, a nucleosome in inactive chromatin near to a transcription start site (TSS) is substituted by a FQC, resulting in a poised state of the chromatin. The FQC can then recruit histone-modifying factors such as acetylases and methylases, leading eventually to recruitment of the basal transcriptional machinery. The FQC and its co-factors may also be involved in substitution of a canonical nucleosome immediately upstream of the TSS (−1 position) by a labile, non-canonical one with modified histones (such as H2A.Z and H3.3). For simplicity, only histone acetylation is shown as symbol of gene activation here.

Our model is based on similarities between FQCs on the one hand, and (half-) nucleosomes and the transcription factor NF-Y, which mimics H2A/H2B-DNA nucleosome assembly (Nardini et al., 2013), on the other hand. Both FQCs and half-nucleosomes are composed of tetramers of similar proteins. Moreover, DNA might be wrapped around FQCs in a similar way as in nucleosomes, including similar loop sizes [about 42-94 base pairs in the case of FQCs and 86 (147:1.7) base pairs in the case of nucleosomes]. Like NF-Y, MADS-domain proteins insert a stretch of their sequence into the minor groove, and they bind to remarkably similar DNA sequences (note that a CCAAT box, to which NF-Y binds, is one half of a perfect CArG-box). Also, DNA containing short AT-rich sequences spaced by an integral number of DNA turns is easiest to bend around the nucleosome, and the same criterion is fulfilled by two CArG-boxes separated by an integer number of helical turns, an arrangement known to facilitate FQC formation (Jetha et al., 2014). In fact, the central region of the CArG-box largely resembles an ‘A-tract’ (sequence motif A_{n+m}T_n with n+m>3) and periodically spaced A-tracts outside the CArG-box have also been detected (Muino et al., 2014). Thus, the DNA binding of FQCs and nucleosomes is facilitated by similar structural motifs.
Conclusions

Much has been learned about FQCs and their role in plant development in recent years. However, two major questions that were not addressed by the original QM remain largely unanswered. First, how do FQCs acquire target gene specificity? Second, by what molecular mechanisms do they activate or repress the expression of their target genes? These topics are highly inter-related, with chromatin structure and nucleosome activities providing an obvious link.

As is the case for many transcription factors, how MIKC-type proteins achieve target gene specificity still represents a major conundrum. The problem actually has at least two layers of complexity. First, DNA-sequence elements similar to CARG-boxes occur thousands of times in the A. thaliana genome, so that almost every gene possesses a potential binding site for MIKC-type transcription factors (de Folter and Angenent, 2006). This strongly indicates that the CARG-box motif alone is not sufficient to explain the target gene specificity of MIKC-type proteins. Second, all of the at least 45 different MIKC-type proteins encoded in the A. thaliana genome share the highly conserved DNA-binding MADS domain (Parenicová et al., 2003) and studies indicate that, for many of these proteins, DNA-binding specificity might be quite similar, although subtle differences can be detected (Huang et al., 1996; Riechmann et al., 1996). Yet, mutant phenotypes of different floral homeotic genes (and other MIKC-type proteins) differ drastically, suggesting a considerable level of target gene specificity among different floral homeotic proteins. Indeed, recent investigations suggest a complex picture in which the CARG-box sequence, structural features of the CARG-box (e.g. a narrow minor groove, the number, distance and orientation of CARG-boxes), sequences beyond the CARG-box and transcriptional cofactors all play a role in FQC target gene recognition (Melzer et al., 2006; O’Maoléidigh et al., 2013; Jetha et al., 2014; Muino et al., 2014; Yan et al., 2016). Chromatin structure may also play a role in target site specificity. In line with this, chromatin-remodelling and -modifying factors were identified as interactors of MIKC-type proteins (Smacziński et al., 2012b). For example, A. thaliana AP1 was suggested to recruit the H3K27 demethylase RELATIVE OF EARLY FLOWERING 6 (REF6) to the promoter of SEP3. This may explain the observed removal of the H3K27me3 inhibitory histone mark and, consequently, activation of SEP3, possibly by antagonizing Polycomb Group (PcG)-mediated transcriptional repression (Smacziński et al., 2012b). It was also shown that AP1 and SEP3 bind to enhancer sites very early during flower development and that chromatin accessibility changes only subsequently, suggesting that SEP3 acts as a pioneer transcription factor (PTF, see Glossary, Box 1) that modifies chromatin accessibility (Pajoro et al., 2014). PTFs are by definition able to bind to inaccessible, nucleosome-associated DNA sites, thus creating an open chromatin environment that is permissive for the binding of non-pioneer factors that can only bind to accessible sites (termed ‘setters’ if they almost always bind to sites matching their DNA-binding motif, and ‘migrants’ if they are more selective, e.g. because their binding requires co-factors) (Slattery et al., 2014; Todeschini et al., 2014). This raises the question as to what enables AP1 and SEP3 to function as PTFs. Jetha et al. (2014) calculated that the ability of cooperative DNA binding of SEP proteins during FQC formation could facilitate their invasion of nucleosomal DNA and thus their activity as PTFs. It is also known that nucleosomes are most efficiently ejected by DNA-binding proteins whose binding sites are spaced by up to 74 bp from each other (Polach and Widom, 1996; Moyle-Heyman et al., 2011); this distance is close to the CARG-box distances for which the highest cooperativity was observed by Jetha et al. (2014).

The analysis of nucleosome-mediated control of gene expression has also provided clues into how FQCs might function. Nucleosomes are composed of an octamer of H2A, H2B, H3 and H4 histones, all of which are present in two copies, wrapped around with DNA almost exactly 147 base pairs long. However, nucleosomes are all but static systems, and chromatin is frequently reorganized at multiple levels (Henikoff, 2008). For example, nucleosomes near transcription start sites may continuously cycle between a repressed canonical form and an unstable, noncanonical form that contains histone variants such as H2A.Z and H3.3 substituting the standard histones H2A and H3, respectively (Soboleva et al., 2014). There is also experimental evidence for the existence of subnucleosomal particles such as half-nucleosomes that contain just one copy of H2A, H2B, H3 and H4. Again, especially at the 5’ end of genes, such dynamic nucleosomes may increase accessibility to transcription start sites and transcription factor binding sites (Rhee et al., 2014). Such dynamic half-nucleosomes (or even full nucleosomes) bear similarities to FQCs and, based on these similarities, we suggest a ‘nucleosome mimicry’ model of FQC action. Specifically, we hypothesize that FQCs represent sequence-specific transcription factors with (half-) nucleosome-like properties that help to establish permissive or repressive chromatin modifications at CARG-box-containing promoters (see Box 5 for details). This molecular mimicry might enable FQCs to evict nucleosomes from positions at which they are already quite labile, e.g. promoter regions with A-tracts (Henikoff, 2008) and hence to act as PTFs.

We hope that the ‘nucleosome mimicry’ model that we propose here will be rigorously tested in the near future. We have the same hope for the QFM itself and more general functions of FQCs. We feel that FQCs provide a useful framework for studying many more processes in plant development and evolution than just the specification of floral organ identity.

Acknowledgements

We are grateful to three anonymous reviewers for their helpful comments on a previous version of the manuscript. G.T. thanks Mia and Lydias Gramzow for their patience. We apologize to all authors whose publications could not be cited owing to space constraints.

Competing interests

The authors declare no competing or financial interests.

Funding

G.T. and R.M. received funding from the Deutsche Forschungsgemeinschaft (DFG) [TH417/5-3].

References


