The evolutionary origin of 'floral quartets':

Clues from molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon*

Dissertation

to obtain the academic degree of
doctor rerum naturalium
(Dr. rer. nat.)

at
the Faculty of Biology and Pharmacy
of the Friedrich-Schiller-University Jena

submitted

by M.Sc. in Cell Biology Yongqiang Wang
born on 26.10.1976 in Shandong China

1. Gutachter: Prof. Dr. G. Theißen

2. Gutachter: Prof. Dr. J. Wöstemeyer

3. Gutachter: PD Dr. J. Uhrig

Date of the public defence: 03.05.2010

Summary

The identity of floral organs in angiosperms is specified by multimeric transcription factor complexes composed of floral homeotic MADS-domain proteins that bind to specific *cis*-regulatory DNA-elements ('CArG-boxes') of their target genes, thus constituting floral quartets. Gymnosperms possess orthologues of floral homeotic genes enconding MIKC-type MADS-domain proteins, but when and how the interactions constituting floral quartets were established during evolution has remained unknown. To better understand the 'abominable mystery' of flower origin, in this project a comprehensive study was carried out to detect the dimerization and DNA-binding of several classes of MADS-domain proteins from a gymnosperm, *Gnetum gnemon* of the Gnetales. Determination of protein-protein interactions by pull-down assays revealed complex patterns of heterodimerization among orthologues of class B, class C and class E floral homeotic proteins and B_{sister} proteins, while homodimerization was not observed. In contrast, electrophoretic mobility shift assays (EMSAs) revealed that all proteins tested except one bind to CArG-boxes also as homodimers, suggesting that homodimerization is relatively weak, but facilitated by DNA-binding. Proteins able of DNA-based homodimerization include orthologues of class B and C proteins; B and C proteins also form heterodimers *in vitro* and in yeast, which is in sharp contrast to their orthologues from angiosperms, which require class E floral proteins to 'glue' them together in multimeric complexes. Remarkably, the heterodimers of B and C proteins from *G. gnemon* are not capable of binding to CArG-boxes, suggesting that DNA-binding *in vivo* is based on homodimers, while heterodimerization of B and C proteins may constitute multimeric, DNA-bound complexes by mediating the interaction between two DNA-bound homodimers. EMSAs and DNase I footprint assays indicated that both B with C proteins and C proteins alone but not B proteins alone can induce DNA-looping to form tetrameric protein-DNA complexes similar to floral quartets. These data suggest that at least some of the gymnosperm orthologues of floral homeotic proteins may have the capability of forming higher-order complexes and that gymnosperm B and C proteins control male organ identity and C proteins controls female organ identity, respectively, by forming quartet-like complexes composed of two homodimers, each bound to a CArG-box.

Contents

1 Introduction

The evolution of plants is accompanied by profound morphological changes. The underlying mechanisms of these changes have challenged evolutionary biologists for decades (Theissen 2009). Genetic analyses showed that the development of complex morphological traits is often directed by only a few developmental control genes. Thus, small molecular changes in the network constituted by these genes might lead to great morphological changes, and in its 'extreme' to the origin of evolutionary novelties (Theißen 2006). This assumption can be considered as the rationale of evolutionary developmental biology ('evo-devo' for short), which studies the phylogeny and function of those developmental control genes responsible for morphological changes (Arthur 2002).

The sudden appearance of the angiosperm flowers 90-100 million years ago (MYA) with no gradual fossil record linking flowering plants to an ancestor has been considered as an 'abominable mystery' by Charles Darwin (Crepet 1998; Crepet 2000; Friedman 2009; Frohlich 1999; Frohlich 2003; Frohlich and Parker 2000; Ma and dePamphilis 2000; Pennisi 2009). The nowadays known angiosperm fossil record only traces back to about 130 MYA (Friis *et al.* 2003; Sun *et al.* 2002), which is far after the time when the lineages separated that led to the extant gymnosperms and extant angiosperms about 300 MYA (Bateman *et al.* 2006). Lack of key fossils makes endeavors to solve the 'abominable mystery' and to answer how flowers originated very difficult.

From an evolutionary point of view, the flower is the result of several key innovations, among which are sepals and petals (which, collectively, constitute the perianth) and carpels (Endress 2001). Over the last few decades developmental biology has unraveled the genetic and molecular interactions among the key players directing flower development in model plants such as *Antirrhinum majus* (snapdragon, Antirrhineae family), *Arabidopsis thaliana* (thale cress, Brassicaceae family), *Petunia* × *hybrida* (petunia, Solanaceae family), and *Oryza sativa* (rice, Poaceae family) (Angenent *et al.* 1995; Coen and Meyerowitz 1991; Ditta *et al.* 2004; Goto *et al.* 2001; Pelaz *et al.* 2000; Theissen and Saedler 2001).

1.1Some brief notes on angiosperms and gymnosperms

Seed plants split about 300 MYA in the lineages that led to the extant gymnosperms and extant angiosperms (Bateman *et al*. 2006). Angiosperms (also called flowering plants) are

the dominant plant group in terms of species number in terrestrial habitats, containing more than 260,000 species within 453 families (APG II 2003), with two thirds of the species being eudicots (Magallon *et al.* 1999). On the contrary, extant gymnosperms comprise only conifers (Coniferophytes, ca. 550 species), Gnetales (ca. 70 species), cycads (Cycadophytes, ca. 150 species), and Ginkgo (Ginkgophytes, only one species) (Bowe *et al.* 2000; Chaw *et al.* 2000; Chaw *et al.* 1997; Donoghue and Doyle 2000; Doyle 2008; Frohlich and Parker 2000). Despite the possibility of being sister groups (Fig 1.1), extant

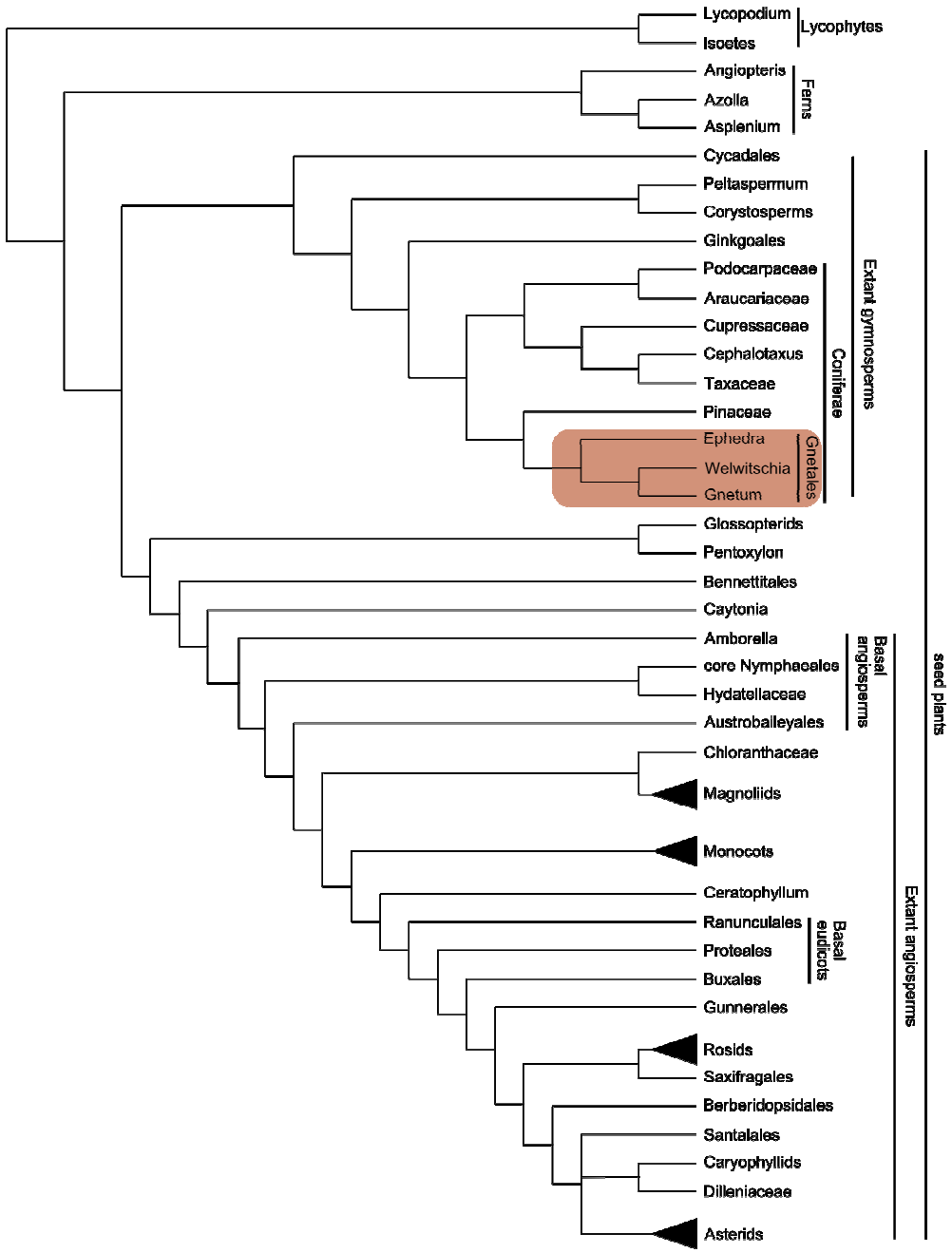

Fig 1.1 Summary topology of current views and recent advances in deep-level angiosperms and gymnosperm relations. (seed plants from Specht and Bartlett, 2009, and non-seed plants from Chaw *et al.*, 2000).

angiosperms are by far more successful than extant gymnosperms in terms of species diversification. These remarkable differences in success are due to morphological differences between angiosperms and gymnosperms. There are a number of morphological differences in vegetative organs, but most evident are differences in the morphology of reproductive organs. A typical eudicot angiosperm flower is composed of four types of organs arranged in different whorls: sterile sepals and petals and fertile stamens and carpels. Sepals are located in the outermost whorl, enclosing and protecting the flower bud before it opens. In the second whorl, petals are located. They are often showy which is important for attracting pollinators. Stamens and carpels are located in the third and fourth whorls, respectively. Stamens (consisting of anther and filament) are the male reproductive structures producing pollen. Carpels (consisting of stigma, style and ovary) are the female reproductive structures of a flower. In contrast to angiosperm flowers, which are primarily bisexual (for review see Theissen and Melzer 2007) (with unisexual species probably evolved several times independently [Renner and Ricklefs 1995]), gymnosperm male and female reproductive structures develop separately on the same plant ('monoecious') or on different plants ('dioecious'). Also, a perianth (i.e. sepals and petals) surrounding male and female reproductive organs in angiosperms is lacking in gymnosperms. Importantly, the ovules of gymnosperms are not enclosed in carpels but are exposed 'naked' on seed cones. The great morphological differences of reproductive organs make homology assessment between gymnosperms and angiosperms difficult and unconvincing (Specht and Bartlett 2009).

Gnetales, consisting of the genera *Gnetum*, *Welwitschia* and *Ephedra* (Arber and Parkin 1908), attracted special attention of scientists because of their controversial phylogenetic position within seed plants (for reviews, see Frohlich 1999; Rothwell *et al*., 2009). Gnetales have some angiosperm-like features, such as perianth-like bracts and double fertilization. Some species (for instance *Gnetum gnemon*) even have bisexual cones in which, however, the female reproductive organs are sterile. Therefore, Gnetales were often regarded as the sister group of angiosperms, with which they were supposed to form the 'anthophyte' clade (for reviews, see Frohlich, 1999; Rothwell *et al*., 2009). However, almost all analyses based on molecular data support that Gnetales are more closely related to conifers than to angiosperms (for example Mathews 2009; Qiu *et al.* 1999; Winter *et al.* 1999).

1.2 Flower development and the ABC model

Flower development in *A. thaliana* can be divided into four stages: floral induction, floral primordia formation, floral organ primordia formation, and floral organ identity specification and differentiation (Meyerowitz *et al.* 1991). Loss of function mutations in genes controlling floral organ identity result in homeotic transformations, i.e. the respective organ is replaced by another type of organ that should not normally appear in this place (Meyerowitz *et al*. 1991). Single and double mutant analysis in *A. thaliana* and *A. majus*, two distantly related species (Bowman *et al.* 1991; Carpenter and Coen 1990; Schwarz-Sommer *et al.* 1990), allowed the proposition of the ABC model which claims that it is the combinatorial interaction of floral homeotic genes that determines floral organ identity (Coen and Meyerowitz 1991). According to the ABC model floral homeotic genes can be grouped into the three functional classes A, B and C. Briefly, class A genes alone determine the identity of sepals, class A and B genes together specify petal identity, class B together with C genes control stamen identity, and class C genes alone function in carpel specification (Coen and Meyerowitz 1991). Class A and C genes repress each other which confines their expression to outer (sepals and petals) and inner (stamens and carpels) whorls, respectively. In *A. thaliana*, class A genes are represented by *APETALA1* (*AP1*) (Mandel *et al.* 1992) and *APETALA2* (*AP2*) (Jofuku *et al.* 1994) and class B genes include *APETALA3* (*AP3*) (Jack *et al.* 1992) and *PISTILLATA* (*PI*) (Goto and Meyerowitz 1994), whilethe only class C gene is *AGAMOUS* (*AG*) (Yanofsky *et al.* 1990). All of these genes, with the exception of *AP2*, encode transcriptional regulators of the MADS-domain protein family.

In ideal class A mutants, sepals are transformed into carpels and petals into stamens while in ideal class B mutants, sepals develop in place of petals and carpels in place of stamens. In ideal class C mutants, stamens are replaced by petals and carpels by sepals, and in addition, there is continued production of mutant organs inside the fourth floral whorl, giving rise to the typical phenotype of a filled flower (Fig 1.2). The ABC model successfully predicts the phenotypes of double and triple mutants of floral homeotic genes and has greatly facilitated studies of flower development (for reviews see Theissen and Saedler 1999; Weigel and Meyerowitz 1994). Meanwhile, evidence has been presented indicating that the basic genetic mechanisms of floral organ determination as proposed by the ABC model apply also to monocots (Ambrose *et al.* 2000; Nagasawa *et al.* 2003; Whipple *et al.* 2007) and probably even to all other angiosperms (Kim *et al.* 2005; Whipple

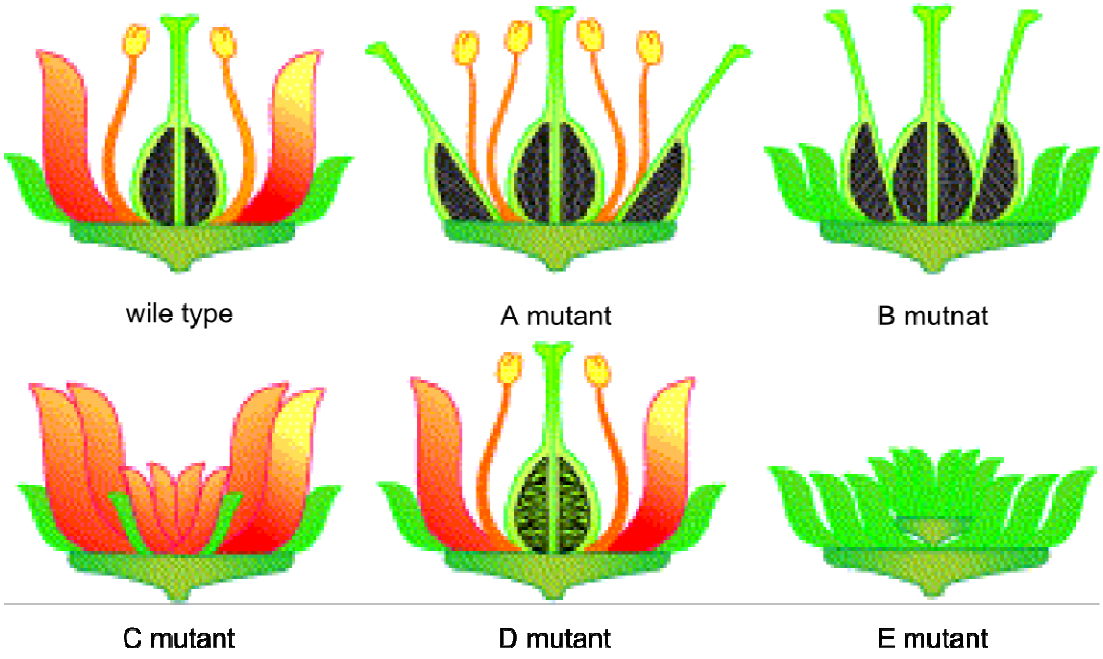

Fig 1.2 Illustration of A, B, C, D and E mutants of *A. thaliana*. In A mutants, carpels are produced in whorl 1 and stamens in whorl 2. In the B mutant, the first two whorls consist of sepals, and the third and fourth whorls of carpels. In the C mutant, petals are formed in both whorl 2 and whorl 3 and the flower becomes indeterminate, resulting in an iteration of the floral program and the production of a new floral bud from the center of the flower. In the D mutant, the ovules in the fifth whorl are converted into carpeloid organs. Finally, in the triple *sep1 sep2 sep3* E mutant, only sepals are produced and the flowers become indeterminate and form a new floral bud from the central meristematic region. (From Ferrario *et al.* 2004)

et al. 2007).

1.3 From the ABC model to the floral quartet model

After identifying the MADS-box gene *FBP11* from *P. hybrida* as an important denominator of ovule identity, the ABC model was expanded into the ABCD model (Colombo *et al.* 1995), with the D function conferring ovule identity (Fig 1.2). The D function in *A. thaliana* is represented by *AG*, *SHATTERPROOF1 (SHP1)*, *SHP2*, and *SEEDSTICK* (*STK*) (Pinyopich *et al.* 2003).

It was by the power of reverse genetics that yet another class of floral homeotic genes was identified. In *A. thaliana*, these are the redundantly functioning genes *SEPALLATA1 (SEP1), SEP2, SEP3* and *SEP4* (formerly known as *AGL2*, *AGL4*, *AGL9* and *AGL3*, respectively). Single mutants of these four *SEP* genes produce only subtle phenotypes, while in *sep1 sep2 sep3* triple mutants (Fig 1.2) organs of the three inner whorls are transformed into sepal like organs, and the flowers loose determinacy (Pelaz *et al*. 2000). In *sep1 sep2 sep3 sep4* quadruple mutants, all the floral organs are transformed into

leaf-like organs (Ditta *et al*. 2004). Consequently, in extension of the ABCD model, a class E floral homeotic function was attributed to the *SEP* genes (Theissen and Saedler 2001).

Transgenic and mutant analyses showed that the class E floral homeotic function of *SEP1*-like genes is, similarly to the A, B and C functions, also conserved across monocots and eudicots (Angenent *et al.* 1994; Kang and An 1997; Pnueli *et al.* 1994).

Very recently, it was shown that *AGL6*-like genes, a subfamily of MADS-box genes that is closely related to the *SEP1*-like genes function partially redundant with them in floral organ specification in species as diverse as petunia, rice and maize (Li *et al.* 2009; Ohmori *et al.* 2009; Rijpkema *et al.* 2009; Thompson *et al.* 2009). Thus, although no loss of function mutant of an *AGL6*-like gene from *A. thaliana* has been described and therefore the function of these genes remains enigmatic in this species, it appears plausible to also designate *AGL6*-like genes as class E floral homeotic genes (Melzer *et al.* 2010).

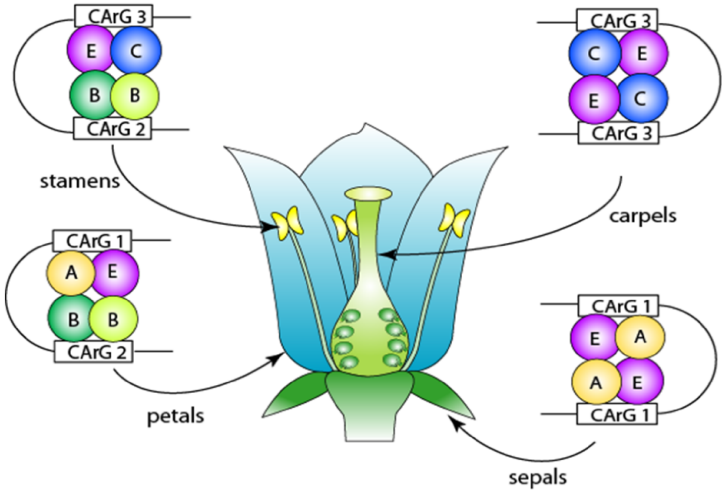

Fig 1.3 The floral quartet model. Different floral organs are specified by different protein complexes binding to genes containing different *cis* regulatory elements (termed CArG boxes). A, B, C, and E represent floral homeotic proteins, and CArG1-3 represent different CArG-boxes. (Adapted from Theissen and Saedler, 2001)

The continuous over-expression of *AP1*, *AP3*, and *PI* or *AP3*, *PI*, and *SEP3* leads to the development of petals from primorida that normally produce vegetative leaves. Likewise, staminoid organs arise from normally leave-developing primorida by the ectopic expression of *AP3*, *PI*, *AG*, and *SEP3*. This indicates that the ABCE genes are not only necessary, but also sufficient for specifying the identities of at least some floral organs (Honma and Goto 2001).

Though they have been extremely valuable for understanding the genetics of floral organ specification, neither the ABC model nor the ABCDE model explains by which mechanism the floral homeotic proteins interact with each other and with their target genes to specify

floral organ identities. The 'floral quartet' model (Fig 1.3), however, proposes that floral homeotic proteins form higher order complexes with each other to specify organ identities. According to this model, the identity of every floral organ is determined by a specific DNA-bound tetrameric complex containing two dimers of floral homeotic MADS-domain proteins. Briefly, sepals are specified by a complex of two class A and two class E proteins, petals by a complex consisting of two class B, one class A and one class E protein, stamens by a tetramer of two class B, one class C and one class E protein, and carpels by a complex of two class C and two class E proteins. Accumulating experimental evidences suggest that floral homeotic proteins can indeed form higher order complexes and that these complexes are indeed tetramers that constitute stable nucleoprotein complexes controlling floral organ identity (Egea-Cortines *et al.* 1999; Ferrario *et al.* 2003; Honma and Goto 2001; Immink *et al.* 2009; Melzer and Theissen 2009; Melzer *et al.* 2009).

1.4 B_{sister}, another clade of MADS-box genes involved in flower development – beyond ABCDE

The floral homeotic A, B, C, D, and E genes fall into distinct phylogenetically conserved clades (Becker and Theissen 2003). The class A genes (except *AP2*-like genes) belong to *SQUA*-like genes, and class B genes are *DEF*/*GLO*-like genes. Both class C and class D genes are *AG*-like genes. Class E genes belong to *SEP1*-like (formerly *AGL2*-like) or *AGL6*-like genes. Beyond these clades, a closely related clade of MADS-box genes is also involved in flower development. These are the B_{sister}-like genes.

B_{sister}-like genes (or *GGM13*-like genes) represent a subfamily that has close phylogenetic affinity to *DEF/GLO*-like genes. Based on their high phylogenetic conservation and their female specific expression pattern, it was hypothesized that they are – together with class C genes – involved in female organ specification (Becker *et al.* 2002). However, mutant analysis of B_{sister}-like genes in *A. thaliana* (Nesi *et al.* 2002) and *P. hybrida* (de Folter *et al.* 2006) showed that these genes are involved in the specification of endothelium identity. A role in female organ specification was not yet revealed. The possibility remains, however, that the situation in *A. thaliana* and *P. hybrida* represents a derived state of female organ specification or ovule specification and seed development, or that genes acting redundantly with them obscure the developmental role of the B_{sister}-like genes in these species.

1.5 Floral organ identity with protein-protein interaction and protein-DNA interaction

All the floral organ identity proteins except AP2 belong to the MADS-domain family of transcription factors. The term MADS is derived from the names of the four founding members of this family: MCM1 in yeast, AGAMOUS in *A. thaliana*, DEFICIENS in *A. majus* and SRF in humans (Schwarz-Sommer *et al*. 1990). Floral homeotic MADS-domain proteins belong to a special subfamily, called MIKC-type MADS-domain proteins (Münster *et al.* 1997). Proteins of this subfamily are characterized by their domain structure. They contain a MADS-domain, an intervening (I) region, a keratin-like (K) domain, and a C-terminal (C) domain (Fig 1.4).

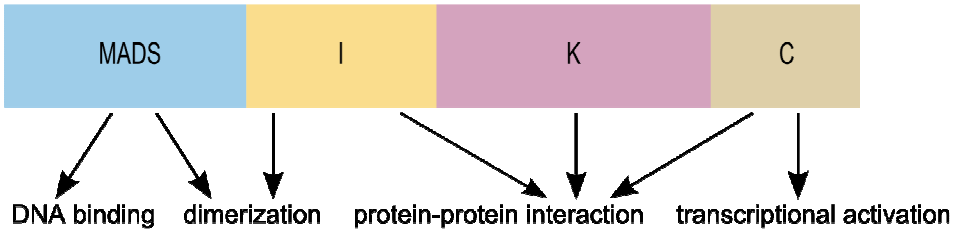

Fig 1.4 Domain structure and their roles in MIKC-type MADS-domain proteins.

The MADS-domain is highly conserved between different members. It contains about 57 amino acids, mainly responsible for DNA binding, but also playing a role in dimerization. The MADS-domain binds to DNA sites termed CArG-boxes with the consensus sequence 5'-CC(A/T)$_6$GG-3' (Pollock and Treisman 1990; Wynne and Treisman 1992). The I-domain has less similarity between different proteins and is variable in length. This domain is important for conferring dimerization (Davies *et al.* 1996; Riechmann *et al.* 1996b). The K-domain is moderately conserved, and is predicted to form coiled-coiled structures that mediate protein-protein interactions (Alvarez-Buylla *et al.* 2000; Davies and Schwarz-Sommer 1994; Yang *et al.* 2003a, Yang and Jack 2004). The C-domain is the most variable region. In some proteins, the C-domain is involved in transcriptional activation (Cho *et al.* 1999; Honma and Goto 2001; Immink *et al*. 2009; Litt and Irish 2003) or protein-protein interactions (Egea-Cortines *et al*. 1999; Tzeng *et al.* 2004).

Protein-protein and protein-DNA interactions of MIKC-type MADS-domain proteins have been the subject of numerous research projects and have been shown to be of vital importance for the function of MIKC-type MADS-domain proteins (Davies *et al*. 1996; de

Folter *et al.* 2005; Immink *et al*. 2009; Immink *et al.* 2003; Leseberg *et al.* 2008; Masiero *et al.* 2004; Riechmann *et al.* 1996a, Riechmann *et al*. 1996b, Schwarz-Sommer *et al.* 1992). The identification of hundreds of dimeric and multimeric protein complexes may indicate that flower development is regulated by a complicated regulatory network (de Folter *et al*. 2005; Kaufmann *et al.* 2005). As mentioned above, the importance of multimer formation among MADS-domain proteins is proposed by the floral quartet model. Meanwhile, the existence of even more multimeric complexes involved in different steps of flower development such as ovule and endothelium formation or the regulation of flowering time has been proposed (Favaro *et al.* 2003; Immink *et al*. 2009; Kaufmann *et al*. 2005).

1.6 MADS-box genes in gymnosperms

As mentioned above, the MADS-box genes involved in floral control fall into distinct clades conserved in all angiosperms. Interestingly, representatives of most of these clades are also found in gymnosperms (Becker *et al*. 2002; Melzer *et al*. 2010; Sundström *et al.* 1999; Tandre *et al.* 1995).

Putative orthologues of class B genes have been reported from different species of gymnosperms, including *Picea abies* (Norway spruce) (Sundström *et al*. 1999), *Pinus radiate* (Monterey pine) (Mouradov *et al.* 1999), *G. gnemon* (Winter *et al*. 1999), and *Cryptomeria japonica* (a conifer) (Fukui *et al.* 2001). In phylogenetic reconstructions, all of these genes appear to be basal to both *DEF*- and *GLO*-like genes from angiosperms (Fukui *et al*. 2001; Winter *et al.* 2002a). Putative orthologues of class C genes have also been reported from *P. abies* (Tandre *et al*. 1995), *P. mariana* (black spruce) (Rutledge *et al.* 1998), *G. gnemon* (Winter *et al*. 1999), *Ginkgo biloba* (Ginkgo)(Jager *et al.* 2003), and *C. japonica* (Futamura *et al.* 2008). Several *AGL6*-like genes have been cloned from *P. abies* (Tandre *et al*. 1995), *P. radiata* (Mouradov *et al.* 1998), *Gnetum parvifolium* (Shindo *et al.* 1999), *G. gnemon* (Winter *et al*. 1999), and *C. japonica* (Futamura *et al*. 2008). The first B_{sister}-like gene, *GGM13*, was reported from *G. gnemon* (Becker *et al*. 2002), and another B_{sister}-like gene has also been cloned from *C. japonica* (Futamura *et al*. 2008).

Despite all proposed similarities in the function of floral homeotic genes and their orthologues between angiosperms and gymnosperms, there are also some profound differences. The most remarkable one concerns *SQUA*-like and *SEP1*-like genes.

Phylogenetic analyses show that *SQUA*-, *SEP1*- and *AGL6*-like genes form a superclade (Becker and Theissen 2003; De Bodt *et al.* 2003; Nam *et al.* 2003). However, in contrast to *AGL6*-like genes, neither *SQUA*- nor *SEP1*-like genes have been isolated from any gymnosperm so far despite numerous attempts. Thus, *SQUA*- and *SEP1*-like genes may have been lost in the lineage that led to extant gymnosperms (Becker and Theissen 2003; Melzer *et al*. 2010; Theissen and Melzer 2007; Zahn *et al.* 2005a), although it cannot be absolutely ruled out that these genes are still to be found in gymnosperms.

For simplicity, gymnosperm *AG*-like genes will henceforth be termed 'C genes', and *DEF/GLO*-like genes will be called 'B genes'. This refers to the phylogenetic relationship of these genes and not necessarily to their function (in contrast, for example, to the term 'class C gene', which designates a certain type of floral organ identity genes).

Along with the phylogenic conservation, also the expression patterns of the MADS-box genes of the different subfamilies are largely conserved between angiosperms and gymnosperms. The gymnosperm B genes were found to be exclusively expressed in pollen cones (Becker *et al.* 2003; Fukui *et al*. 2001; Mouradov *et al*. 1999; Sundström *et al*. 1999), which strongly suggests that gymnosperm B genes have a role in determining male organ identity, similar to class B genes of angiosperms. C genes from gymnosperms are expressed predominantly in male and female reproductive organs (Jager *et al*. 2003; Rutledge *et al*. 1998; Sundström *et al*. 1999; Tandre *et al*. 1995; Winter *et al*. 1999) although expression was also described in vegetative organs (Jager *et al*. 2003). This led to the hypothesis that the genetic mechanisms determining reproductive organ identity are similar in all seed plants, i.e. angiosperms and gymnosperms (Becker *et al*. 2003; Theissen 2002; Theissen *et al.* 2000; Winter *et al*. 2002a). Following this idea, the common ancestor of angiosperms and gymnosperms might have recruited the expression of C genes in reproductive organs to distinguish vegetative from reproductive structures. Furthermore, the male-specific expression of B genes may have served to distinguish male and female reproductive structures (Becker *et al*. 2003; Theissen *et al*. 2000; Winter *et al*. 2002a, Winter *et al.* 2002b). This hypothesis is supported by heterologous expression experiments indicating that B and C genes from gymnosperms can at least partially substitute their putative angiosperm orthologues (Rutledge *et al*. 1998; Sundström and Engström 2002; Tandre *et al.* 1998; Winter *et al*. 2002a; Zhang *et al.* 2004). Further, the female specifically expressed B_{sister}-like genes may have served to specify female reproductive structures since

the duplication of B and B_{sister}-like genes from an ancestral gene (Becker *et al*. 2002). As mentioned previously, *AGL6*-like genes were shown to confer at least partially the class E floral homeotic function in angiosperms. Similar to their angiosperm counterparts, *AGL6*-like genes from gymnosperms are predominantly expressed in reproductive organs (Mouradov *et al*. 1998; Shindo *et al*. 1999). Also, heterologous expression of the *AGL6*-like gene *DAL1* from *P. abies* in *A. thaliana* results in premature flowering, a phenotype similar to that when ectopically expressing the class E gene *SEP3* from *A. thaliana* (Carlsbecker *et al.* 2004). This indicates that *AGL6*-like genes from gymnosperms might also be involved in conferring reproductive organ identity.

1.6.1 Study of MADS-box genes in *Gnetum gnemon*

G. gnemon (Fig 1.5a, b), a species of Gnetales, is one of the best-analyzed gymnosperms in terms of MADS-box genes. cDNAs of nineteen MADS-box genes (Fig 1.6) have been cloned from *G. gnemon* (Becker *et al.* 2000; Winter *et al*. 1999), six of which have clear phylogenetic relationship to genes from angiosperms. Among these genes, *GGM2* and *GGM15* are closely related to class B genes, *GGM3* is closely related to class C genes (Fig 1.6). Additionally, two *AGL6*-like genes, *GGM9* and *GGM11*, related to class E genes, have been isolated. As mentioned above, also a B_{sister}-like gene, *GGM13*, has been cloned from *G. gnemon* (Becker *et al*. 2002). However, no close relatives of class A genes have been isolated from *G. gnemon* so far.

The expression patterns of these genes from *G. gnemon* have been studied by RNA *in situ* hybridization and Northern blotting (Fig 1.5c) (Becker *et al*. 2002; Becker *et al*. 2003; Becker *et al*. 2000; Winter *et al*. 1999). Both *GGM2* and *GGM15* are only expressed in male cones, and not in the sterile ovules of male cones (Becker *et al*. 2003; Winter *et al*. 1999). This similarity suggests that *G. gnemon* B genes, like those of angiosperms, have a role in determining male organ identity. However, the spatial distribution of *GGM15* was more restricted to the antherophore, which indicated that *GGM2* might play a more general role than *GGM15* in specifying male reproductive organ identities (Becker *et al*. 2003).

GGM3 is expressed in both male and female cones. This is similar to its counterparts in angiosperms, which are expressed in the reproductive organs (stamens and carpels). At early stages of development expression of *GGM3* is detected in all reproductive organs, and at later stages, it is localized in the outer envelope of both male and female cones

(Winter *et al*. 1999).

Both *GGM9* and *GGM11* are expressed in male and female cones, but *GGM9* is expressed almost equally strong in male and female cones, while *GGM11* is expressed more strongly in female cones than in male cones (Becker *et al*. 2000). *GGM9* is strongly expressed in very young primordia of male cones, where cells have not differentiated yet, and its expression becomes weaker at later developmental stages of male cones, but its expression is not detected in the sterile ovules of male cones (Becker *et al*. 2003). Initially *GGM11* is expressed in the upper part of the envelope and in a few cell layers surrounding the sporogenic tissue, and at later stages of antherophore development, its expression is localized in the single cell layer surrounding the sporogenic tissue (Becker *et al*. 2003). Unlike *GGM9*, *GGM11* is expressed in the sterile ovules of male cones (Becker *et al*. 2003). The different expression patterns indicate that these two *AGL6*-like genes play different functions in reproductive organ development.

The expression of *GGM13* is restricted to female cones at very early stages of development (Becker *et al*. 2002; Becker *et al*. 2000). This female specific expression pattern led to the

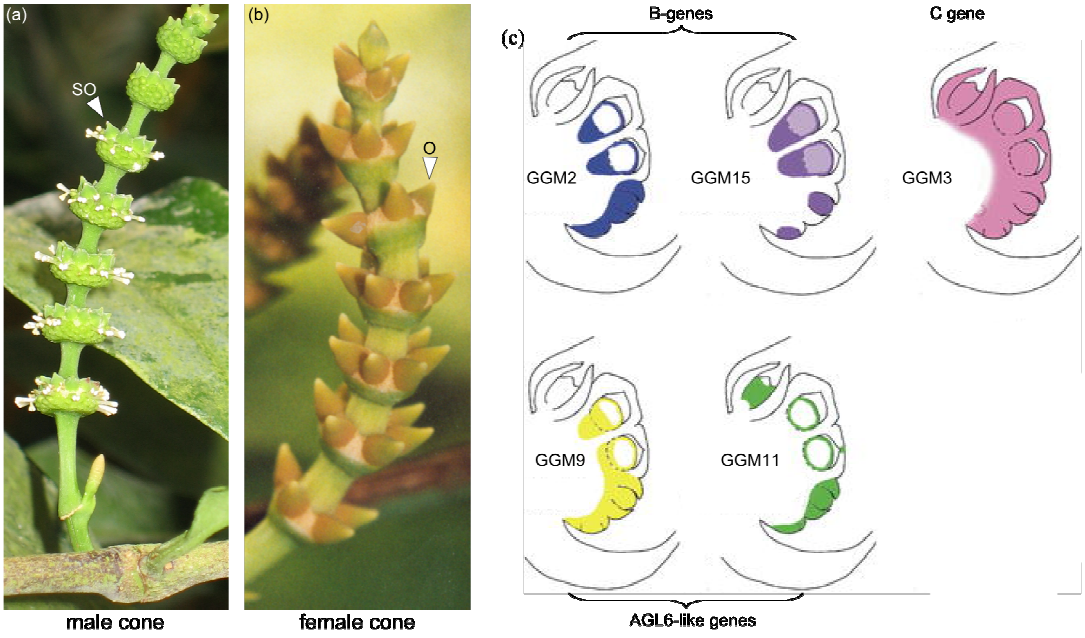

Fig 1.5 Structures of *G. gnemon* and expression of MADS-box genes. (a): male reproductive structures. (b): female reproductive structures. (c): Schematic representation of expression patterns of *Gnetum* MADS-box genes during male cone development (Becker *et al.*, 2003). O: fertile ovule. SO: sterile ovule.

argument that B_{sister}-like genes may have served to specify female reproductive structures since the duplication of B and B_{sister}-like genes from an ancestral gene (Becker *et al*. 2002).

Besides expression patterns, protein-protein interactions of these *G. gnemon* MADS-domain proteins have previously been tested in yeast two-hybrid assays (Melzer 2005). These data indicated that direct protein-protein interactions between members from

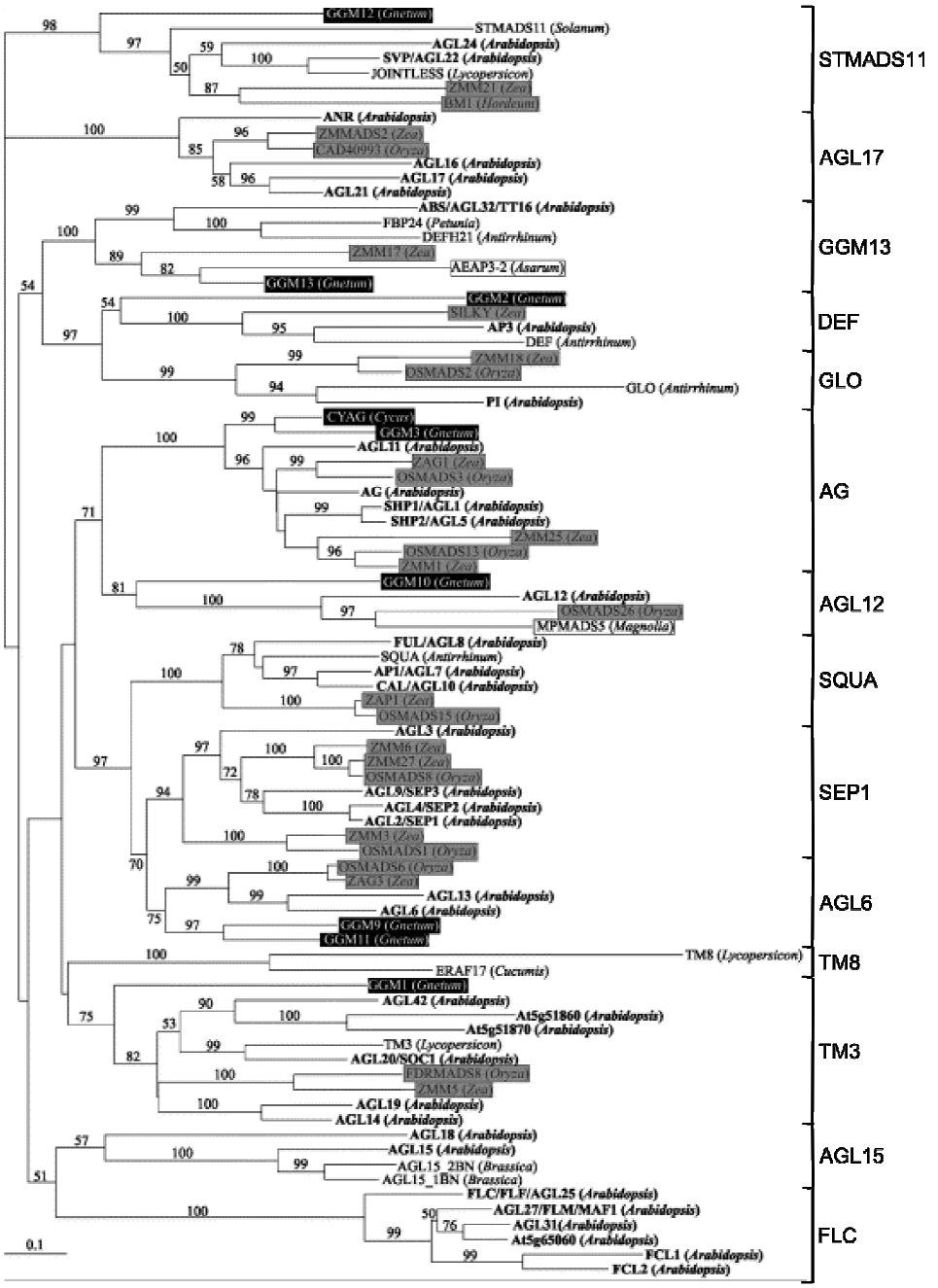

Fig 1.6 Phylogenetic tree showing the relationships between a subset of the MADS-domain proteins known. Genus names of species from which the respective genes were isolated are given in parentheses beside the protein names. (From Becker *et al.* 2000)

different MADS-domain protein clades occur in *G. gnemon*. In contrast, in angiosperms interactions between proteins from different subfamilies appear to be mainly mediated by SEP1-like proteins (Immink *et al*. 2009).

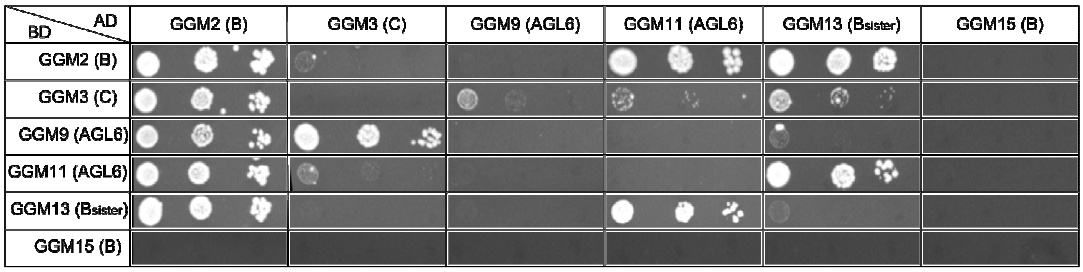

Fig 1.7 Interaction of *G. gnemon* MADS-domain proteins detected by yeast two-hybrid assays) (From Melzer, 2005)

1.7 Aim of the project

The floral organ identities in angiosperms are specified by floral homeotic proteins, whose interactions with each other and with their target genes are pivotally important as indicated by the floral quartet model (Theissen and Saedler 2001). Gymnosperms contain orthologues of floral homeotic genes, but when and how the interactions constituting floral quartets were established during evolution has remained unknown. Therefore, the aim of this project was to obtain a detailed understanding on the molecular evolutionary dynamics of the gene regulatory network controlling seed plant reproductive organ identities and with this to get clues on the molecular mechanisms that enabled the origin of the angiosperm flower. For this purpose, the protein-protein and protein-DNA interactions among gymnosperm MADS-domain proteins were explored. Six putative orthologues of floral homeotic proteins from *G. gnemon* were chosen for the study: the B class related proteins GGM2 and GGM15, the C class related protein GGM3, the E class related proteins GGM9 and GGM11 and the B_{sister}-protein GGM13.

Specifically, the goal was to test all possible protein combinations for dimerization using pull-down assays and EMSAs and to compare these results to the interactions of the respective orthologues proteins from angiosperm. The results from pull down assays and EMSAs should also reveal whether differences in protein interactions can be observed with and without DNA 'scaffolding' the interaction.

Furthermore, angiosperm floral homeotic proteins are known to form multimeric

complexes with each other. Thus, multimerization among *G. gnemon* MADS-domain proteins should be explored by EMSA and DNase I footprint assays using appropriately designed DNA-probes. The aim was to infer whether or not multimerization of MADS-domain proteins is an ancestral feature of seed plants. Based on the data obtained, hypotheses should be developed aiming to explain to which degree the ability of forming multimeric complexes might have been of importance for the origin of the flower.

2 Materials and methods

2.1 Materials

2.1.1 Genes

G. gnemon MADS-box genes *GGM2* (gi 5019428) and *GGM15* (gi 10880310) (both B genes), *GGM3* (gi 5019430 (C gene), *GGM9* (gi 5019455) and *GGM11* (gi 6468291) (both *AGL6*-like E genes), and *GGM13* (gi 5019463) (B_{sister}-like gene) were selected for this study. These genes are hereafter called 'GGM genes' and the corresponding proteins 'GGM proteins'.

2.1.2 Reagents, primers and vectors

The coding sequences of GGM genes were cloned into different host vectors (Table 2.1) for different assays. A list of constructs used can be found in section 5.4. All primers, buffers and other reagents used are listed in sections 5.5 and 5.2. Manufacturers of reagents used are given when the respective methods are described in the text.

Table 2.1 List of vectors used in this project

Host vector	Usage Description	References
pSPUTK & pTNT	Cloning and expression vectors. Used for *in vitro* transcription and translation of GGM genes. Expression is under control of the SP6 (pSPUTK) or SP6 and T7 (pTNT) promoters.	Stratagene (pSPUTK) Promega (pTNT)
pIVEX1.3WG & pIVEX1.4WG	Expression vectors. Used for *in vitro* transcription and translation of GGM genes. Expression is under control of the T7 promoter. Proteins were expressed as fusions with a hexa-histidine tag at the N-terminal (pIVEX1.4WG) or C-terminal (pIVEX1.3WG) end of the protein.	Roche
pBatTL-N-YFP & pBatTL-C-YFP	Transient expression vectors. Used for *in planta* expression of proteins. Proteins were translated as fusion proteins with the yellow fluorescence proteins (YFP) the N- (pBatTL-N-YFP) or C-terminal (pBatTL-C-YFP) part.	Hackbusch *et al*., 2005

2.2 General methods

2.2.1 DNA methods

2.2.1.1 Plasmid extraction

Plasmid DNA was extracted with a plasmid mini kit (OMEGA BIO-TEK) or a plasmid midi kit (nuclebond AX, Machery-Nagel), following the manufacturers' instructions.

2.2.1.2 Restriction endonuclease digestion

One microgram plasmid DNA was digested with 10 U restriction endonuclease (NEB or

Fermentas) in a total volume of 15 µl at the conditions suggested by the manufacturers for two or three hours.

2.2.1.3 PCR

DNA fragments were amplified from plasmids by the polymerase chain reaction (PCR) in a thermocycler (Mastercycler personal, Eppendorf). The PCR reaction mixture included 5 nM forward and reverse primers, 200 µM dNTP, 50 ng template DNA, 2.5 U thermostable DNA polymerase (*Taq*- or *pfu*-polymerase, supplied by Fermentas or Segenetic) and 2 µl 10× DNA polymerase buffer (supplied with the polymerase) in a total volume of 20 µl. When DNA fragments were subjected to restriction endonuclease digestion after PCR, the total volume of the PCR was scaled up to 50 µl. The reactions conditions were as follows:

94°C for 3 minutes	Initial denaturation	1 cycle
94°C for 30 seconds	Denaturation	25-30 cycles
52-56°C (depending on the primer) for 30 seconds	Annealing	
72°C for 30 seconds to 1 minute	Extension	
72°C for 10 minutes	Final extension	1 cycle

2.2.1.4 DNA purification and DNA precipitation

Reaction mixtures of restriction endonuclease digestion were separated on 1% or 2% agarose gels, and gel slices containing target DNA fragments were excised from the gel. DNA fragments in gel slices or amplified by PCR were recovered using the Wizard SV gel and PCR clean-up system (Promega) following the manufacturer's instructions.

DNA was precipitated by adding 1/10 volume of 3 M sodium acetate (NaAC) and 2 volumes of 100% ethanol. The mixture was incubated at -70°C for 20 minutes or at -20°C overnight. Afterwards, the mixture was centrifuged at 10,000× g for 10 minutes. The precipitated DNA was washed two times with 70% ethanol and then dissolved in distilled water or 1× TE buffer.

2.2.1.5 DNA ligation

A ligation reaction was performed at 4°C overnight in a 10 µl reaction volume containing approximately 30 ng insert DNA, 100 ng host vector DNA, 1 µl 10× ligation buffer, and 2.5 U T4 DNA ligase (Fermentas). The ligation product was stored at –20°C until transformation into electrocompetent *E. coli* cells.

2.2.1.6 Preparation of competent *E. coli* cells (XL1-blue strain)

200 ml LB medium was inoculated with *E. coli* strain XL1-blue (Stratagene) from a glycerol stock and incubated overnight at 37°C on a rotary shaker. This overnight culture was used to inoculate 500 ml LB medium to an OD_{600} of 0.2. This culture was grown at 18°C on a rotary shaker (200-250 rpm) until OD_{600}= 0.4 (about 5-7 hours). Cells were centrifuged at 4°C, 5,000 rpm for 20 minutes two times and re-suspended in 350 ml and 250 ml pre-cooled distilled water, respectively, by vortexing. Cells were then centrifuged at 4°C, 7,000 rpm for 20 minutes. After re-suspended in 20 ml pre-cooled distilled water, cells were again centrifuged at 4°C, 7,000 rpm for 20 minutes and finally re-suspended in 800 µl 7% DMSO. Aliquots of 50 µl were made, quickly frozen in liquid nitrogen and stored at -70°C.

2.2.1.7 Transformation of *E. coli* by electroporation and screening for recombinant plasmids

A vial of competent *E. coli* cells was thawed on ice, mixed with 5 µl ligation product, and then pipetted quickly onto the bottom of an ice-cooled electroporation cuvette (1-mm gap width). Electroporation was carried out at 2000 V (Electroporator Model 2510, Eppendorf). After electroporation, 1 ml pre-warmed (37°C) LB medium was added into the cuvette and mixed with the cells. The mixture was transferred into a 1.5 ml microcentrifuge tube. Cells were grown at 37°C for 40 minutes while shaking, and then 150 µl of the cell suspension was spread on LB plates with appropriate antibiotics and incubated at 37°C for 12-15 hours. Grown colonies were checked for the presence of the plasmid with the appropriate insert via PCR or restriction endonuclease digestion. Positive candidate clones were confirmed by sequencing (see section 2.2.1.8).

2.2.1.8 DNA sequencing

Sequencing was carried out on an ABI 3730xl DNA Analyzer using Big Dye Terminator chemistry at the Max Planck Institute for Chemical Ecology.

2.2.2 Protein *in vitro* translation

GGM proteins were *in vitro* translated with the TNT T7/SP6 quick coupled transcription/ translation system (Promega) according to the manufacturer's instruction. For pull-down assays, bait proteins were translated from pIVEX1.3/1.4WG constructs, and prey proteins from pSPUTK constructs. For co-translating proteins, pIVEX1.3/1.4WG constructs were

used together with pTNT constructs. For EMSA and DNase I footprint assays proteins were translated from pSPUTK constructs.

Briefly, the translation reaction was carried out at 30°C for 60-90 minutes in a total volume of 25 µl containing 500-800 ng plasmid DNA, 0.5 µl 1 mM methionine or 0.5 µl [L-^{35}S]-methionine (370 MBq/ml, ~0.01 mM, Hartmann analytic), and 20 µl TNT SP6 or T7 Quick Master Mix (depending on the type of RNA polymerase promoters on the template vectors). If two proteins were tested simultaneously for DNA-binding in EMSA or DNase I footprint assays (see section 2.4.2 and 2.4.3.2), co-translation was performed. In case of co-translation, unless otherwise stated, equal amounts of both plasmid DNA templates were used to translate the two proteins. For pull-down assays, separate translation as well as co-translation of two proteins was tested (see section 2.3.2).

After translation, 6.5 µl 100% glycerol was added to the mixture. The mixture was either used immediately or stored at -70°C. Correct sizes of the proteins produced were checked for every plasmid constructs at least once by including ^{35}S-methionine in the *in vitro* translation reaction and analyzing 2 µl of translated products by SDS-PAGE. A prestained protein marker (broad range [7-175 kDa] from NEB) was used as size reference.

2.2.3 Electrophoresis

2.2.3.1 Agarose gel electrophoresis

To separate DNA fragments 1% or 2% (w/v) agarose gels (depending on the expected size of the DNA fragments) were prepared in 1× TAE containing about 1 µg/ml ethidium bromide. DNA samples were mixed with 1/5 volume of 6× DNA loading dye. Along with the samples, 4-6 µl DNA ladder (GeneRuler Ladder Mix, Fermentas) was loaded to estimate the sizes of DNA fragments. Gels were run in 1× TAE buffer at 80-120 V. Gel images were captured by Gel Doc2000 (Bio-Rad). For vector construction and probe preparation, gel slices containing target DNA fragments were excised under UV light and purified (see section 2.2.1.4).

2.2.3.2 Polyacrylamide gel electrophoresis (PAGE)

2.2.3.2.1 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was used for analyzing proteins, either translated products or elutions from pull-down assays. Gels were 8 cm × 10 cm × 0.75 mm in size. A 12% separating gel

(including 5 ml Rotiphorese NF-Acrylamide/Bis-solution 30% [35.5:1] [Carl Roth], 3.15 ml Tris-HCl [1.5 M, pH 8.8], 125 µl 10% SDS, 125 µl 10% APS, 10 µl TEMED and 4.1 ml distilled water) was prepared and poured to 5.5 cm height. The gel surface was covered with distilled water, and gels were polymerized at room temperature for more than 30 minutes. After polymerization, a 4.5% stacking gel solution (containing 455 µl Rotiphorese NF-Acrylamide/Bis-solution 30% [35.5:1], 670 µl Tris-HCl [0.5 M, pH 6.8], 27 µl 10% SDS, 27 µl 10% APS, 10 µl TEMED and 1.81 ml distilled water) was poured on top of the separation gel and a comb was inserted immediately. Gels were polymerized at room temperature for more than 20 minutes. Gel wells were cleaned just before samples were loaded. Protein samples were mixed with 1/5 volume of 6× SDS loading dye, boiled at 85°C for 3 minutes, chilled on ice for 5 minutes, and then loaded onto the gel. Gels were run in 1× SDS running buffer at 15 mA until bromophenol blue reached the separation gel, then the current was increased to 25-30 mA until bromophenol blue reach the end of the gel. Gels were fixed in gel fixing solution for 15 minutes with gentle shaking and then dried in a gel drier for 1 hour. Signals were visualized by autoradiography or a phosphorimaging system (FLA- 7000, Fujifilm). The exposure time was at least overnight.

2.2.3.2.2 Native PAGE

Native PAGE was used to analyze protein-DNA interactions in EMSA, to separate protein-DNA complexes from free DNA in DNase I footprint assays, or to purify labeled DNA probes for footprint assays. Gel size was 20 cm × 20 cm × 1 mm. Prior to gel pouring, glass plates were washed thoroughly with distilled water, dried glass plates were then assembled in gel casters. A 5% acrylamide gel solution (containing 3.5 ml 5× TBE, 4.375 ml Rotiphorese NF-Acrylamide/Bis-solution 40% [29:1], 350 µl 10% APS, 35 µl TEMED and 27 ml distilled water) was prepared and used to pour the gel. A comb was inserted immediately to create wells, and gels were polymerized at room temperature for at least one hour. For EMSA or to separate protein-DNA complexes from free DNA, before loading samples, gels were pre-run at 100 V for 30-60 minutes. Gel wells were cleaned and then samples were loaded quickly. Gels were run in 0.5× TBE at 120-150 V until bromophenol blue reached the end of the gel. Gels were dried for 1 hour, and signals were visualized by autoradiography or a phosphorimaging system (FLA- 7000, Fujifilm). Exposure time was usually at least overnight. When separating protein-DNA complexes from free DNA or preparing probes for DNase I footprint assays the exposure time was limited to 5-30 minutes due to the high amount of radioactivity on the gel.

2.2.3.2.3 Sequencing Gels

Sequencing gels were used in DNase I footprint assays to separate DNA fragments after DNase I cleavage. Gels were 20 cm × 40 cm × 0.4 mm in size and contained 10% acrylamide and 8 M urea. The glass plates were cleaned with 10% SDS, rinsed thoroughly with distilled water, and the inner surface of the smaller plate was treated with Repel-Silane ES (Amersham Biosciences). Plates were assembled into gel casters, and laid down horizontally on the surface of the bench. The gel solution (containing 10 ml 5× TBE, 12.5 ml Rotiphorese NF-Acrylamide/Bis-solution 40% [29:1], 25 g urea, and 7.5 ml distilled water) was mixed and stirred until urea was completely dissolved. The solution was then mixed with 170 µl 10% APS and 70 µl TEMED. The gel was poured while carefully tapping the plate to avoid bubbles. A comb was inserted immediately and gels were polymerized at room temperature for at least two hours before use. Gels were pre-run at 55-60 W to heat the gel. Gel wells were cleaned and samples were loaded quickly. Gels were run at 40-55 W in 1× TBE until bromophenol blue reached the end of the gel or just ran out of the gel. Gels were dried in a gel dryer for one hour. Signals were visualized by a phosphorimaging system (FLA- 7000, Fujifilm). The exposure time was at least overnight.

2.3 Pull-down assay

2.3.1 Expression vector construction

To produce proteins that carry a hexa-histidine tag (hereafter His_6-tag) at their N-terminal end, cDNA sequences were cut from corresponding pSPUTK constructs with *Nco*I and *Sma*I and then subcloned into *Nco*I/*Sma*I sites of the pIVEX1.4WG vector (Fig 2.1) (Roche).

To produce proteins that carry the His_6-tag at their C-terminal end, reverse primers (listed in section 5.5) were designed to replace stop codons on the respective cDNAs with amino acid codons and to introduce a *Sma*I recognition site. cDNA fragments were amplified from pSPUTK constructs with these reverse primers and plasmid-specific forward primers. Amplified DNA fragments were digested with *Nco*I and *Sma*I and then subcloned into *Nco*I/*Sma*I sites of pIVEX1.3WG vector (Fig 2.1) (Roche).

In both pIVEX1.3WG and pIVEX1.4WG constructs, transcription of the inserted cDNAs is under control of the T7 promoter. In contrast, transcription is under control of the SP6

promoter in pSPUTK constructs. In order to co-translate bait and prey proteins for pull-down assays, the coding sequences of *GGM2*, *GGM11*, *GGM13*, and *GGM15* were subcloned from pTFT1 or pGBKT7 constructs (Melzer, 2005) into *Eco*RI/*Sal*I sites of the pTNT vector. Both SP6 and T7 promoters are located upstream of the cDNA insertion in pTNT vectors, thus enabling co-translation with pIVEX1.3WG and -1.4WG constructs.

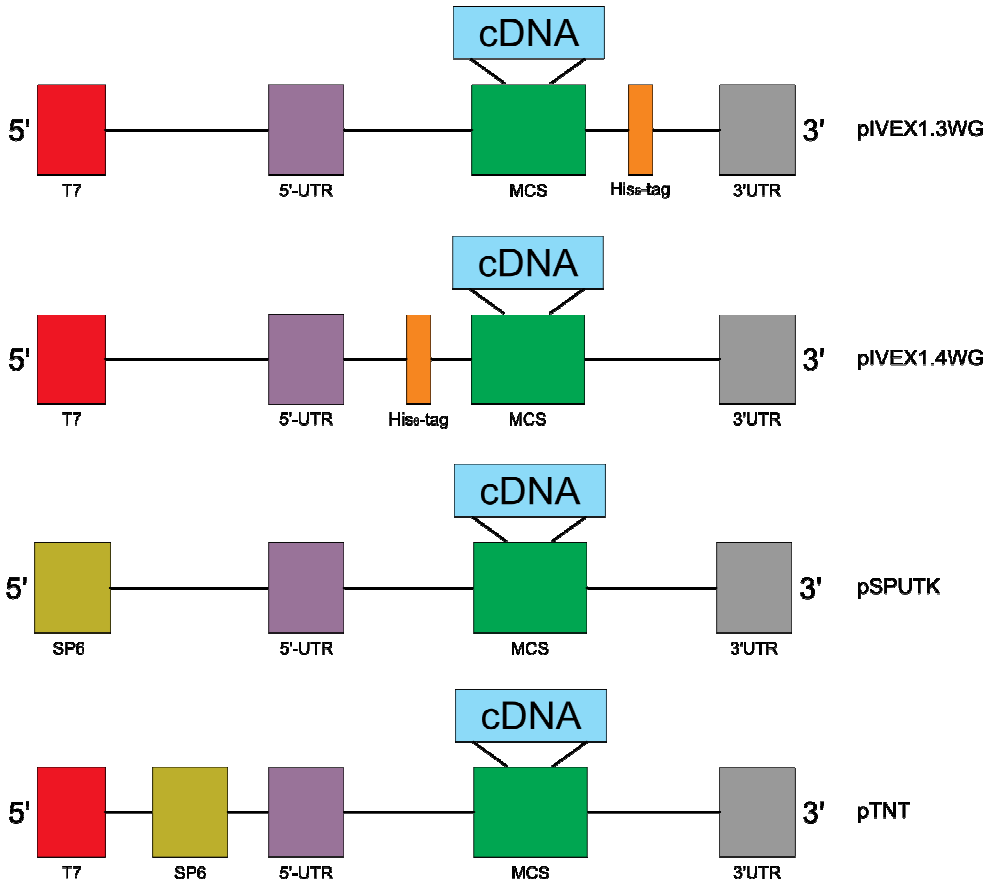

Fig2.1 Scheme of the Linear Template Generation. Backbones of host vecotors (named right-most side) are illustrated. Space and lines are used to align elements of different vectors. His_6-tag: sequence encoding hexa-histidine. cDNA: gene sequences to be inserted. MCS: multiple cloning sites. SP6: SP6 promoter. T7: T7 promoter. UTR: untranslated regulatory regions.

2.3.2 Protein-protein interaction

To set up pull-down assay for GGM proteins, vectors, operation procedures, and buffers were compared. All proteins used for pull-down assays were *in vitro* translated (see section 2.2.2).

As the His_6-tag is short, it may be buried inside the proteins and would then be inaccessible

for affinity purification by Ni-NTA magnetic agarose beads (QIAGEN) (hereafter beads). Thus, binding efficiencies of N- and C-terminal His_6-tagged GGM proteins to the beads were compared. As a starting point, buffers for binding, washing, and elution under native conditions were used as recommended by the bead manufacturer (QIAGEN) (buffer set I, for buffer compositions, see section 5.2.1). 25 µl GGM2 protein carrying a His_6-tag at the N-terminal end (produced from the pIVEX1.4WG construct) or at the C-terminal end (produced from the pIVEX1.3WG construct) were incubated with 10 µl beads in 80 µl binding buffer I for 2 hours. After washing two times with 100 µl washing and interaction buffer I and two times with 100 µl interaction buffer I, bound proteins were eluted with 20 µl elution buffer I and analyzed by SDS-PAGE (Fig 2.2). This way, the full pull-down procedure was mimicked except that no prey protein was added. The results showed that both N- (Fig 2.2b, lane 1) and C-terminal (Fig 2.2a, lane 1) His_6-tagged GGM2 can bind to and be captured by the beads. However, using the construct where the His_6-tag is fused to the N-terminal end of the protein, more protein was captured by and eluted from the beads, indicating a relatively higher efficiency of N-terminal His_6-tag fusions for immobilizing GGM2 protein under the given conditions.

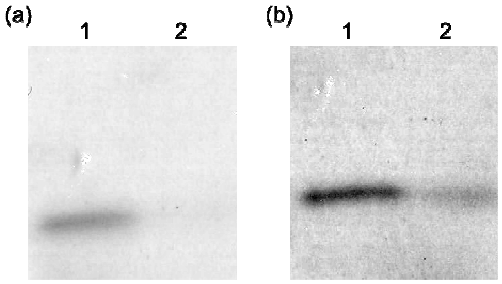

Fig 2.2 Binding efficiency tests of protein fused with N- or C-terminal His_6-tagged protein to beads. GGM2 proteins were fused with a His_6-tag at either the C (a) or N-terminus (b). Lane 1: translated GGM2 protein. Lane 2: eluted GGM2

Next, experiments were carried out to test whether the order in which bait proteins, prey proteins and beads where mixed affects binding of prey proteins (Fig 2.3). Three different variants of mixing the three components were tested. In the first variant (Fig 2.3a) bait and prey proteins were co-translated, incubated with the beads and eluted. In the second variant (Fig 2.3b), bait protein was incubated with the beads, after washing prey protein was added, co-incubated with prey and beads and eluted. In the third variant (Fig 2.3c), bait and prey proteins were translated separately, then co-incubated for interaction, and finally beads were added to bind bait proteins. In all variants, buffer set I was used, and the bound proteins were eluted and analyzed by SDS-PAGE (Fig 2.4).

It was shown in yeast two-hybrid assays that GGM2 and GGM11 interact (Melzer 2005),

so GGM2 and GGM11 were assumed to interact in pull-down assays as well. Therefore, the tests described above were carried out using GGM2 (both the N- and C-terminal His_6-tagged variants) as a bait and GGM11 as a prey. The C-terminal His_6-tag fusion was included to compare again, among the three interaction procedures, if the His_6-tag fusion position influences binding of bait proteins to beads. The results showed that all of the three ways of mixing the different components worked with both N- and C-terminal His_6-tagged GGM2 as a bait. The N-terminal His_6-fusion gave relatively stronger signals in all of the three assays, indicating that the efficiency with which the prey protein is captured is higher when the His_6-tag is fused to the N-terminal end of the protein, which agreed with the results obtained before. Of the three ways of mixing bait protein, prey protein and beads, allowing the bait protein to bind to the beads followed by addition of the prey protein (Fig 2.3b), yielded the relative strongest signal for the prey protein after elution (Fig 2.4a, lane 2). It was therefore decided to use N-terminal His_6-tagged proteins as baits, and to test for interactions by adding the prey protein to the already immobilized bait.

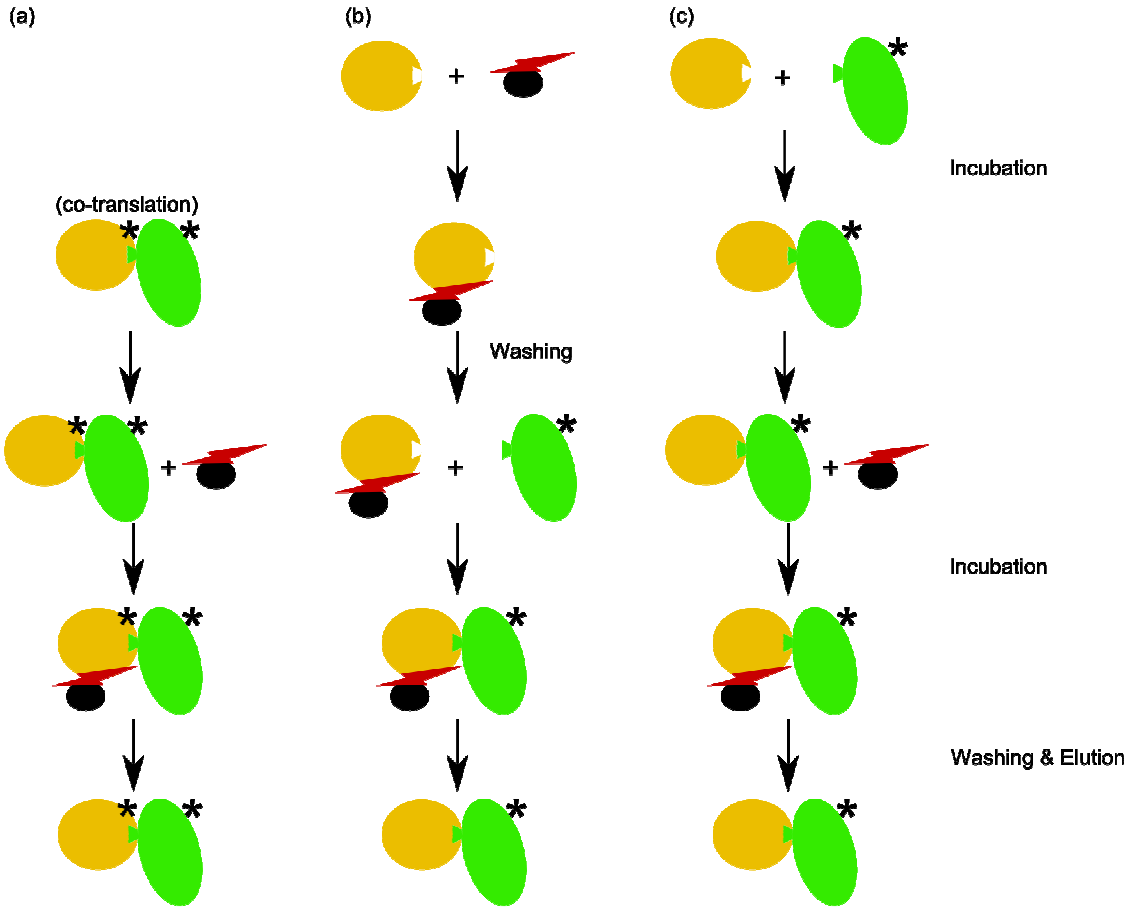

Fig 2.3 Illustration of three different ways tested to mix bait protein, prey protein and magnetic agarose beads coated with Ni-NTA.

Compared to lane 1, signal for bait protein in lane 2 was stronger (Fig 2.4a). This phenomenon might be due to that interaction of bait and prey proteins before the adding of beads (Fig 2.3a) hampered the binding of His_6-tag to beads, while the bait proteins alone had better binding efficiency to beads and thus the eluted bait protein in lane 2 (Fig 2.4a) included a fraction which bound to beads but not to prey protein. It could also be possible that just more bait proteins were efficiently translated and subsequently included in the reaction which contributed to the stronger signals of prey protein in lane 2 (Fig 2.4a), while this scenario was less likely since previously proteins were continuously successfully translated in other assays (EMSA for instance).

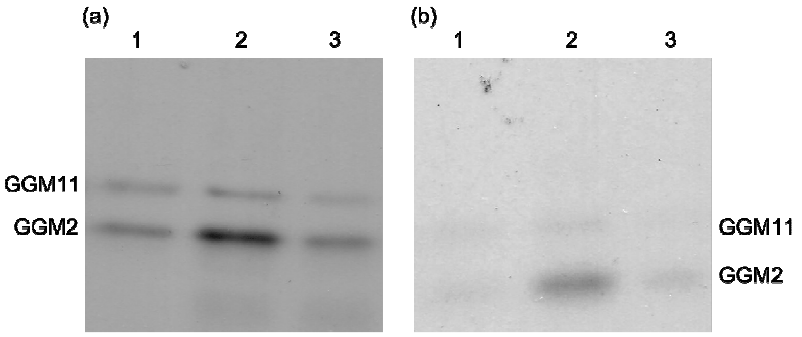

Fig 2.4 Tests of different operation processes for pull-down assays. GGM11 was used as prey to be captured by GGM2 with the His_6-tag at the N-terminal (a) or C-terminal (b) end. Both prey and bait proteins were radioactively labeled and eluted proteins were analyzed in 12% SDS gels. Lane 1: bait and prey proteins were co-translated (see Fig 2.3a). Lane 2: bait proteins were bound to beads before interacting with prey proteins (see Fig 2.3b). Lane 3: bait and prey proteins were translated separately, and they were allowed to interact before the beads were added (see Fig 2.3c).

When buffer set I was used to test other protein-protein interactions, some prey proteins were found to bind to the beads un-specifically (i.e. in the absence of bait proteins). To overcome this problem, different buffers were tested to find conditions that do not lead to unspecific binding of prey proteins to beads. Several protein pairs were used to test this. According to the bead manufacturer's instructions (QIAGEN), mainly combinations of two factors were varied: concentration of sodium chloride and imidazole. The tests showed that using buffer set I, neither GGM2 nor GGM11 exhibited unspecific binding to the beads, confirming that the conclusions drawn before (concerning position of the His_6-tag and the possibilities to mix the different components) were valid. After several tests, buffer set II (for composition see section 5.2.1), where unspecific binding was eliminated, (as shown in the results part [section 3.1]) was chosen for all further studies.

Briefly, the final procedure for pull-down assays was as below:

25 µl of His_6-tagged protein ('bait') was incubated with 5 µl slurry of beads in 100 µl binding buffer II at 15°C for one hour with gentle shaking. *In vitro* translation lysate programmed with DNA templates that did not contain a cDNA insert was used as control to determine non-specific binding of prey proteins to the beads (see section 3.1 or Fig 3.1). Beads were washed two times with Washing & Interaction (W&I) buffer II and resuspended in 100 µl W&I buffer II. Subsequently, 25 µl prey protein was added to the suspension and incubated at 15°C for one hour with gentle shaking. After washing two times with 100 µl with W&I buffer II, bait and prey proteins were eluted with 20 µl elution buffer II and separated on a 12% SDS-polyacrylamide gel (see section 2.2.3.2.1).

Under these assay conditions, however, only a small fraction of input prey proteins were eluted (see section 3.1, or Fig 3.1 comparing signals of input and elution). At least two factors may contribute to this phenomenon. First, imidazole was included in all buffers to prevent non-specific binding while according to the manufacturer of the magnetic beads (QIAGEN) at the same time this chemical also hampers complete binding of bait proteins to beads. Second, a considerable amount of protein was not eluted from the beads. This was revealed by boiling the beads after protein elution and analyzing the mixture by SDS-PAGE (data not shown). Since no quantitative analyses were performed to monitor the protein loss at different steps, it remains unclear which steps during the pull-down procedure were the most critical ones concerning 'loss' of proteins.

2.4 Protein-DNA interaction

2.4.1 Probe labeling and purification

The short DNA-binding probe S contains one CArG-box and some surrounding base pairs (total length 47 bp) and was derived from the regulatory intron of *AG*. Probe Sm has the same base composition as probe S but nucleotides have been reordered randomly. Thus, probe Sm does not contain sequences with obvious similarities to a CArG-box. Probe S and Sm were obtained by annealing two chemically synthesized oligonucleotides. Longer probes containing one (probe L1) or two (probe L2 and L2i) copies of the same CArG-box were excised from a respective pBluescript II SK (+) construct containing the probe sequence using *Xba*I and *Xho*I recognition sites (Melzer *et al*. 2009). Lengths of probe L1 and L2 are 151 bp. Sequences of all probes are listed in section 5.6. The two CArG-boxes encoded on probe L2 were spaced by 6 helical turns (assuming 10.5 base pairs per helical turn). Probe L2i has the same sequence to that of L2 except that a 5-bp insertion was

included between the two CArG-boxes that were thus spaced by 6.5 helical turns.

To label probes used in EMSA (probe S, L1, and L2) with Klenow fill-in reactions, 0.1-0.4 µg of the appropriate DNA fragment, 2 µl 10× Klenow buffer, 2.5 µl dNTP (dTTP/dGTP/dCTP, each 25 mM), 0.5-1 µl [α-^{32}P]-dATP (370 MBq/ml, 0.01-0.03 mM, Hartmann analytic) and 5 U Klenow Fragment exonuclease minus (Fermentas) were incubated in a total volume of 20 µl at 30°C for 15 minutes. The reaction was stopped by heating to 75°C and keeping at this temperature for 10 minutes, and the labeled probes were purified with the QIAquick Nucleotide Removal Kit (QIAGEN) according to the manufacturer's instructions, and stored at -20°C until used.

In DNase I Footprint assay probe L2 and L2i were used. DNA fragments encoding the probes were amplified from the respective pBluescript II SK (+) vectors containing the probe sequence with high fidelity PCR enzyme mix (Fermentas) and subsequently purified (see section 2.2.1.4) (primers used are listed in section 5.5). 3 µg of the amplified DNA was digested with *Xho*I, separated on a 2% agarose gel, and purified from the gel (see section 2.2.1.4). Purified DNA fragments were dephosphorylated at 37°C for 10 minutes in a total volume of 122 µl containing 1× FastAP buffer and 10 U FastAP (Thermosensitive Alkaline Phosphatase, Fermentas). The reaction was stopped by heating to 70°C and keeping at this temperature for 5 minutes, the DNA fragments were precipitated and dissolved in distilled water. Then, DNA fragments were labeled radioactively using T4 Polynucleotide Kinase (T4 PNK, Fermentas) at 37°C for 60 minutes in 40 µl volume containing 1× T4 PNK buffer A, 5 µl [γ-^{32}P]-ATP (370 MBq/ml, ~0.01mM, Hartmann analytic), and 20 U T4 PNK. The reaction was stopped by adding 2 µl 500 mM EDTA and heated to 75°C and keeping at this temperature for 10 minutes. Labeled probes were purified with illustra MicroSpin G-25 Columns (GE Healthcare), digested with *Xba*I, and separated on a 5% native polyacrylamide gel (see section 2.2.3.2.2). Gel slices were excised and DNA fragments were eluted overnight in 500 µl gel elution buffer while shaking. Probes were precipitated by adding 2 volumes of 100% ethanol, then dissolved in distilled water to a specific activity of <50000 cpm/µl, and stored at 4°C until used.

2.4.2 EMSA

2 µl protein (from section 2.2.2) was mixed with the binding cocktail (containing 1-2 ng labeled DNA-probe, 4 µl 3× EMSA buffer [see section 5.2.1], 1.1 µl 86% glycerol, and 1.2

µl 25% CHAPS), and distilled water was added to a total volume of 12 µl. When specificity of DNA-binding was tested, a 100-fold excess (100-200 ng) of non-labeled probe S (as a specific competitor) or mutated probe Sm (as a non-specific competitor) were supplied. Reaction mixtures were incubated on ice for 90 minutes before loading onto a native polyacrylamide gel (see section 2.2.3.2.2).

2.4.3 DNase I footprint assay

2.4.3.1 Preparing G+A and C+T DNA ladder

2.4.3.1.1 G+A ladder

A mixture containing 3-6 ng labeled probe, 1 µg salmon sperm DNA, 1 µl 4% formic acid and distilled water in a total volume of 11 µl was incubated at 37°C for 25 minutes. 150 µl of 10% piperidine (GE Healthcare) were added to the mixture and incubated at 90°C for 30 minutes. The reaction mixture was chilled on ice for 5 minutes, and extracted with chloroform two times. DNA fragments were ethanol-precipitated. The precipitate was washed two times with 70% ethanol. Formamide loading buffer was added to reach a specific activity of 10,000 cpm/µl. The DNA was stored at 4°C until used.

2.4.3.1.2 C+T ladder

A mixture containing 3-6 ng labeled DNA, 10 µg salmon sperm DNA, 30 µl hydrazine monohydrate and distilled water in a total volume of 56 µl was incubated at 30°C for 20 minutes. The reaction was stopped by adding 200 µl stop solution (0.3 M NaAC, 10 mM EDTA, 0.05 mg/ml tRNA), and DNA fragments were ethanol-precipitated. The precipitated DNA was washed two times with 70% ethanol, dissolved in distilled water, and precipitated again with ethanol. Then, the pellet was dissolved in 100 µl distilled water plus 11 µl 98% piperdin, and incubated at 90°C for 30 minutes. The mixture was extracted two times with chloroform, and DNA fragments were precipitated again with ethanol. The pellet was washed two times with 70% ethanol, and finally dissolved in formamide loading buffer to reach a specific activity of 10,000 cpm/µl, and stored at 4°C.

2.4.3.2 DNase I cleavage and signal acquisition

8 µl of *in vitro* translated protein (from section 2.2.2) were mixed with 10,000 cpm of labeled DNA (about 1-1.5 ng) in 1× EMSA buffer (total volume 12 µl), and incubated on ice for 3 hours to enable equilibrium of protein-DNA binding. DNase I (50 U/µl, Fermentas) was diluted to 4 U/µl in DNase I dilution buffer, and 2 µl were added to the

binding mixture, and mixed quickly by pipetting up and down. The DNase I cleavage reaction was carried out on ice for 30 seconds, and stopped by adding 2 µl 0.5 M EDTA. The reaction products were separated by native-PAGE (see section 2.2.3.2.2). Gel slices containing free DNA or protein-DNA complexes were excised and DNA fragments were eluted overnight in 500 µl gel elution buffer while shaking, precipitated by adding 2 volumes of 100% ethanol, and dissolved in formamide loading buffer. DNA samples were loaded on a sequencing gel and subjected to electrophoresis (see section 2.2.3.2.3).

2.4.3.3 Phosphorimager analysis

Changes of sensitivity of DNA to DNase I cleavage after protein binding were calculated using free DNA as reference (Melzer *et al*. 2009). The quantitative analysis was performed using the software ImageQuant 5.0 (Molecular Dynamics) with default settings and parameters. The width of the bands was manually adjusted. Five invariable internal bands were selected as internal references to correct for unequal DNA loading. The change of sensitivity to DNase I digestion after protein binding was calculated for every band by the expression

$$\frac{\text{band intensity of protein-bound DNA} \times \sum\text{band intensities of internal references of free DNA}}{\text{band intensity of free DNA} \times \sum\text{band intensities of internal references of protein-bound DNA}}$$

2.5 BiFC assays

2.5.1 Construction of vectors

The Gateway cloning system (Invitrogen) was used for vector construction. cDNAs were amplified from their pSPUTK constructs with SP6 primer and reverse primers containing an *EcoR*V restriction site (listed in section 5.5) and subcloned into *Nco*I-*EcoR*V sites of the pENTR4 entry vector (entry clones). cDNA inserts in entry clones were transferred to the destination vectors pBatTL-N-YFP and pBatTL-C-YFP (Hackbusch *et al.* 2005) via LR recombination reactions. The reaction mixture contained 70 ng entry vector DNA, 150 ng destination vector DNA, and 0.5 µl Gateway LR Clonase II Enzyme Mix (Invitrogen) in 1× TE buffer (total volume 5 µl). The reaction was incubated at 25°C for 7 hours with gentle shaking and stopped by adding 0.5 µl proteinase K solution (2 µg/µl) and an incubation at 37°C for 10 minutes. Recombinant DNA was transformed into XL1-Blue electrocompetent cells. Cells were plated on medium containing spectinomycine to select for clones containing the appropriate insert in the destination vector. Plasmids were checked for the presence of the appropriate insert by colony PCR and confirmed by sequencing. Destination vectors containing the appropriate cDNA insert were transformed

into *Agrobacterium tumefaciens* strain ABI (Koncz and Schell 1986) by electroporation.

2.5.2 Tobacco transformation mediated by agrobacteria and microscope observation

Tobacco plants (*Nicotiana benthamiana*) were grown in greenhouse under 16 h light (20 °C)/8 h dark (15°C) cycles. When the plants had developed 4 to 5 leaves, they were used for transformation.

Agrobacterium tumefaciens strain ABI containing pBatTL-YFP-N or pBat-TL-YFP-C constructs with the appropriate cDNA inserts were grown separately overnight in 20 ml YEB medium containing chloramphenicol (10 µg/ml), kanamycin (50 µg/ml) and spectinomycin (125 µg/ml). An *A. tumefaciens* strain carrying the p19 viral silencing suppressor gene of *Tomato bushy stunt virus* (Voinnet *et al.* 2003), used to enhance transient expression, was also grown overnight in 20 ml YEB containing tetracycline (8 µg/ml). Cells were centrifuged and re-suspended in 3-5ml re-suspension buffer (10 mM MES, 10 mM $MgCl_2$, 150 µM acetosyringone). The OD value of a 1:100-diluted suspension was measured and the suspension was diluted to an OD_{600}=1. Equal volumes of strains containing pBat-TL-YFP-N and pBat-TL-YFP-C constructs and as well as the strain carrying the p19 suppressor were mixed and incubated in the dark for 2-3 hours before used to transform tobacco leaves. Plants were watered half a day before transformation. Before transformation, leaves not to be infiltrated were removed. Leaves were infiltrated by agrobacteria using a 1-ml syringe through a punched area. The procedure was repeated until the whole leaf acquired a dark green color due to the media infiltrated. Transformed plants were kept dark overnight and then grown in the greenhouse for 2-4 days. The YFP signals were examined using a fluorescence microscope (Leica DM5500B) with YFP fluorescence filter (excited light wavelength 500/520 nm), and signal specificity was checked under fluorescence filters A (wavelength 340-380 nm), I3 (wavelength 450-490 nm) and N2.1 (wavelength 515-560 nm).

3. Results

3.1 Protein-protein interactions as revealed by *in vitro* pull-down assays

To characterize the interactions of GGM proteins, pull-down assays were established and applied. The principle of pull-down assays is that a prey protein that interacts with an immobilized bait protein is captured together with the bait. The bait is usually immobilized by introducing a tag that shows a strong affinity to a matrix that has been coated appropriately. In this study the His_6-tag fusion system was chosen.

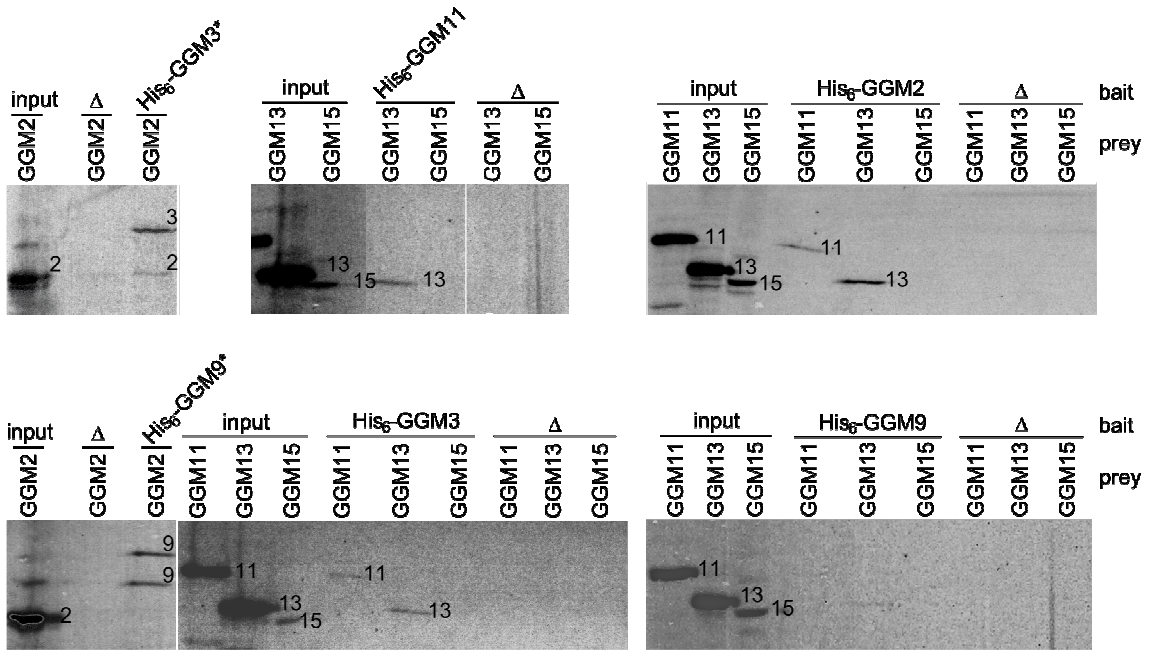

Fig 3.1 Protein-protein interactions tested by *in vitro* pull-down assays. 20 µl of His_6-tagged GGM proteins ('His_6-GGMn') were used as a bait, bound to Ni-NTA beads and incubated with 20 µl of untagged, radioactively labeled prey proteins as noted above the gel. In the left-most lane of each gel, 2 µl of *in vitro* translated prey protein solution were loaded ('input') for size comparison. 'Δ' denotes a negative control in which the 'bait' in vitro translation extract was programed with an expression vector that did not contain a cDNA insert. Bands are numbered according to the names of the GGM proteins. Some bait proteins are marked with an asterisk to indicate that they have also been radioactively labeled and are thus visible on the gel. *In vitro* translation of GGM9 occasionally yielded two bands, one of them probably representing a truncated version of GGM9.

As shown in Fig 3.1, GGM2 interacted with GGM3, GGM11, and GGM13. Besides interacting with GGM2, GGM3 interacted with GGM11 and GGM13. In addition, an interaction between GGM11 and GGM13 was observed. A possible interaction between GGM3 and GGM9 could not be analyzed in the pull-down assay because of unspecific binding of both untagged GGM3 and GGM9 to the beads. Neither GGM9 nor GGM15 interacted with one of the partner proteins tested. None of the proteins tested did interact with itself to form a homodimer (data not shown).

When compared to previous results obtained by yeast two hybrid analyses (Fig 1.7), the two data sets were in good agreement (Fig 3.2). In the 20 cases (where a possible interaction was analyzed with both methods ignoring the undetermined GGM3/GGM9 interaction), 17 cases (85%) are supported by both methods (either interacting or not). One heterodimer (GGM2/GGM9) was detected with the yeast two-hybrid assay only, another one (GGM3/GGM11) only with pull-down assays. Remarkably, homodimerization was not observed with both methods except for GGM2, which was only detected with the yeast two-hybrid assay, however. It is also important to note that none of the methods revealed any interaction partner for the B protein GGM15.

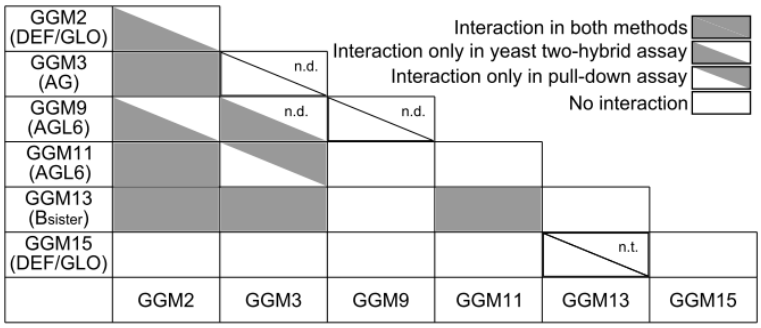

Fig 3.2 Comparison of the interaction data obtained by *in vitro* pull-down and yeast two-hybrid assays. Grey triangles indicate an interaction between two proteins. Interactions that have been found in one direction only in the yeast two-hybrid assay (GGM2/GGM3, GGM2/GGM9, GGM3/GGM9 and GGM3/GGM13) are not differentiated from the ones that have been found reciprocally. The interaction between GGM3 and GGM9 could not be determined in pull-down assays because of unspecific binding of the two proteins to the Ni-NTA agarose beads (n.d., not determined. n.t, not tested).

3.2 Protein-DNA interactions obtained by EMSA

MADS-domain proteins are known to bind as dimers to their target DNA sequence termed CArG-boxes ('CC-A rich-GG' with the consensus sequence $CC(A/T)_6GG$). Floral homeotic proteins are transcription factors and they may regulate their target genes by forming DNA binding complexes. To determine the interaction of the *G. gnemon* MADS-domain proteins with DNA, EMSAs were applied. For homodimerization and heterodimerization, probe S on which one CArG-box is encoded was used.

3.2.1 Homodimer formation

To test homodimerization, full-length proteins, C-terminal deleted proteins and mixtures

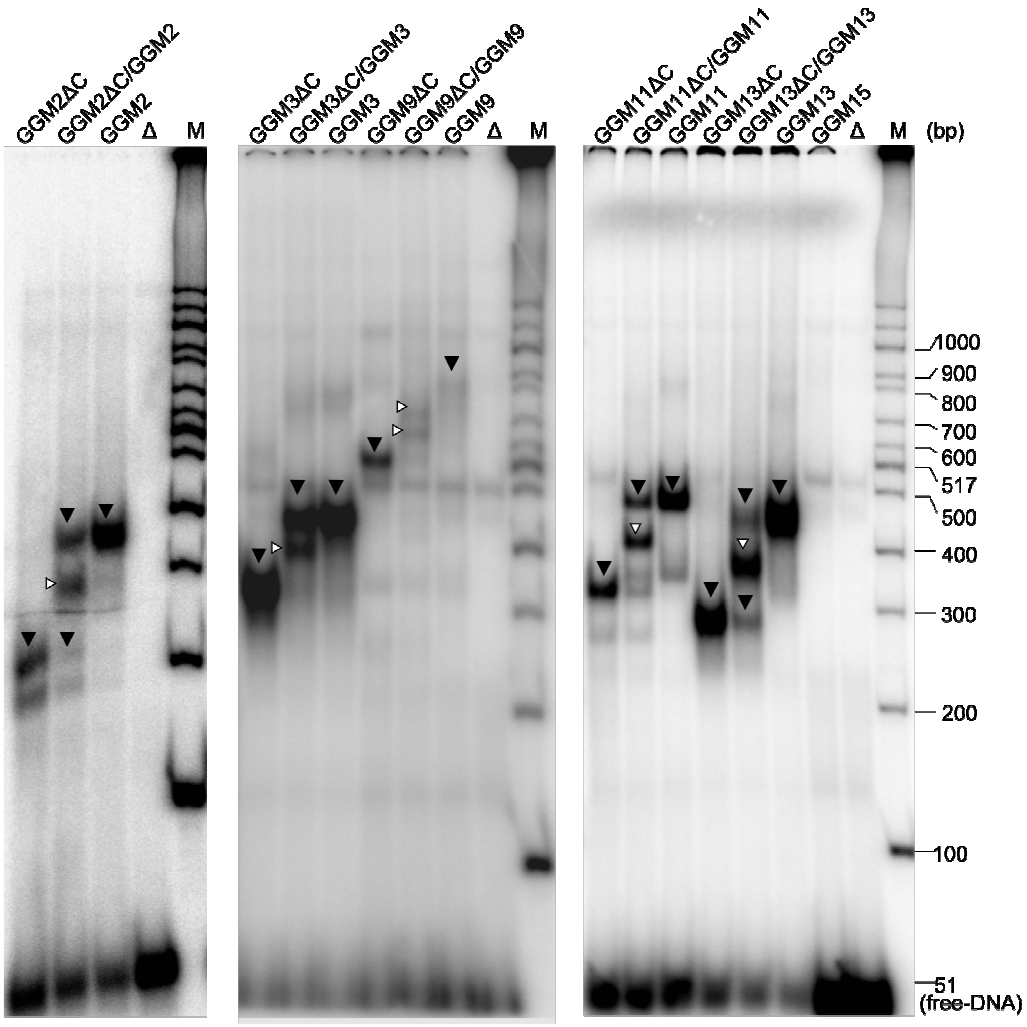

Fig 3.3 Homodimerization test of protein-DNA interactions. Proteins applied are noted above the gel. When C-terminal truncated versions of the proteins were used, 'ΔC' is appended at the end of the name. Major Protein-DNA complexes are indicated by triangles . Heterodimeric complexes are indicated by white triangles. Homodimers and the GGM9-DNA complex are indicated by black triangles. 'Δ' indicates negative controls in which in vitro translation lysate programmed with a vector that did not contain a cDNA insert was added; 'free DNA' indicates the un-bound DNA probes. 'M' denotes marker lanes in which a radioactively labeled DNA ladder (NEB 100 bp DNA ladder) was applied. The property of the relatively weak band beside the GGM2ΔC homodimer is unknown. The composition of complexes formed by GGM9ΔC /GGM9, indicated by while triangles, is unclear. So is the complex formed by GGM9 alone.

obtained by co-translation were used in EMSA. In case where heterodimerization occurs between the full-length and truncated proteins, three different retarded bands will be visible on the gels, representing DNA probes that have bound (from low to high electrophoretic mobility) homodimers of full-length proteins, heterodimers of one full length and one truncated protein, and homodimers of truncated proteins, respectively. For dimerization these are the only combinations. As shown in Fig 3.3, all of the proteins tested except GGM15 were capable of binding to DNA on their own. While in this study GGM15 did not

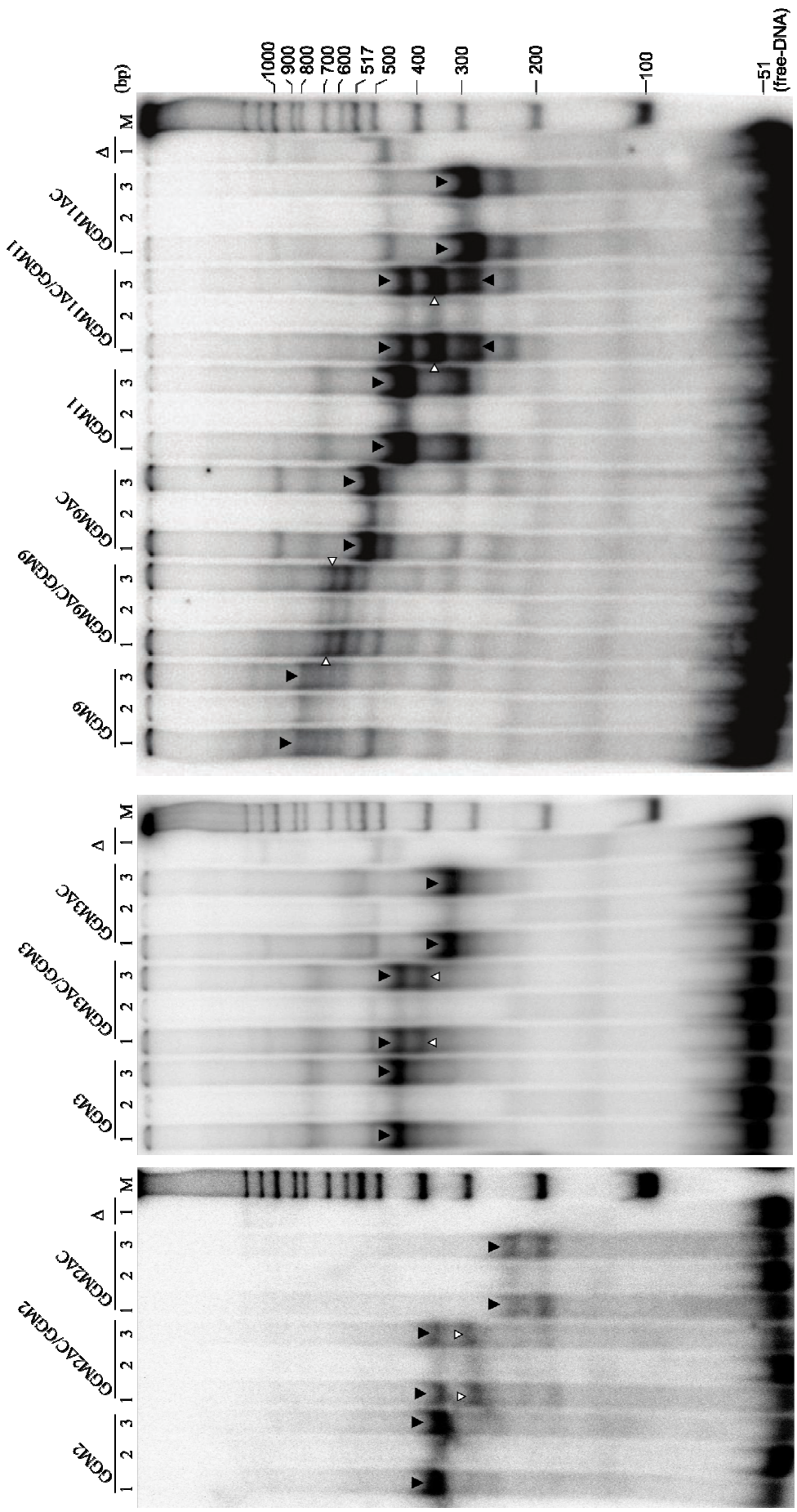

Fig 3.4 Specificity test of homodimerization. Band assignment is as in Fig 3.3. '1' indicates reactions in which only labeled probe S was added to the protein, '2' indicates reactions that contained labeled probe S and a 100 fold excess of unlabeled probe S, and '3' indicates reactions that contained labeled probe S and a 100 fold excess of unlabeled probe Sm. Only one complexes formed by GGM9 is indicated whose property togethers with the rest un-indicated is unclear. The additonal fast mobility complex formed by GGM11 might be composed of truncated proteins.

bind to the probe at all, GGM2, GGM3, GGM11 and GGM13 were found to be able to bind as homodimers (Fig 3.3). In contrast, GGM9 constituted a protein-DNA complex of unusually low electrophoretic mobility when applied without a partner protein (Fig 3.3). This effect cannot be easily attributed to the molecular weight or charge of GGM9, which is very similar to that of the other proteins analyzed. It is unclear whether the low electrophoretic mobility is due to unanticipated changes in DNA conformation upon GGM9 binding, due to an unusual structure of the GGM9 protein, or whether GGM9 binds as a multimer to DNA.

To test for sequence specificity in DNA binding of GGM2, GGM3, GGM9, GGM11 and GGM13, competition tests were carried out (Fig 3.4). In these tests, together with probe S used in above tests, a mutated probe Sm was used which has the same nucleotide composition as the probe S but a randomized order of nucleotides with the sequence not similar to the CArG-box. For such a competition test, *in vitro* translated proteins were incubated with the labeled probe S and with a 100-fold excess of unlabeled probe Sm and, in a parallel reaction, with a 100-fold excess of the unlabeled probe S. When a 100 fold excess of unlabeled probe S was added (lanes labeled with '2' in Fig 3.4), the protein-DNA complexes formed by GGM2, GGM3, GGM9, GGM11 and GGM13 were hardly detectable. In contrast, addition of a 100-fold excess of probe Sm did not result in such a dramatic decrease in signal strength. Thus, it can be concluded that GGM2, GGM3, GGM11, and GGM13 can form homodimers that specifically bind to CArG-box DNA. GGM9 also binds specifically to CArG-boxes. However, the stoichiometry of binding of this complex remains to be determined.

3.2.2 Heterodimer formation

To assess the formation of DNA binding heterodimers, proteins were co-translated and afterwards incubated with the DNA probe S. Heterodimerization was inferred when a band of intermediate mobility compared to the two DNA-bound homodimers was observed. In some cases, the complexes with heterodimers formed by the two co-translated full-length proteins had almost the same electrophoretic mobility, thus hampering detection of potential heterodimers. In these cases, a C-terminal deleted construct of one of the proteins was used. Representative examples of EMSA results are shown in Fig 3.5. The B protein GGM2 was capable of forming DNA-binding heterodimers with several partners (GGM9,

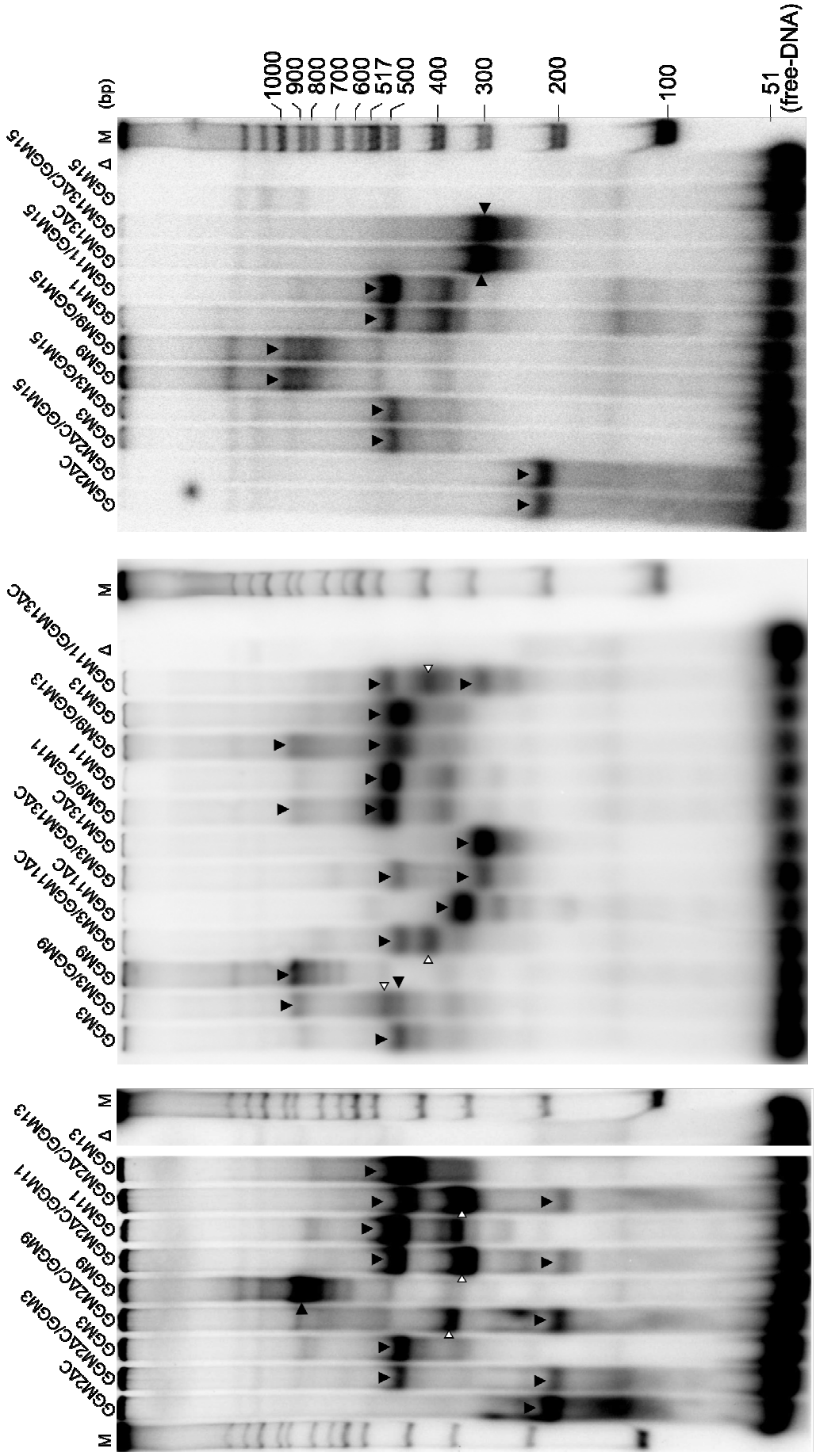

Fig 3.5 Heterodimerization test of protein-DNA interactions. Band assignment is as in Fig 3.3.

GGM11 and GGM13), whereas the C protein GGM3 interacted only with GGM11 to constitute a DNA-binding heterodimer. In addition, an interaction between GGM11 and the B_{sister}-like protein GGM13 was detected. Noteworthy, GGM2 and GGM3, which interacted in yeast two-hybrid as well as pull-down assays, did not form a DNA-binding heterodimer in these tests. The band of intermediate electrophoretic mobility observed when GGM2 and GGM9 are co-incubated indicates that these two proteins can interact with each other and bind to DNA. Considering that GGM9 and GGM11 have a similar molecular weight, and that the DNA-bound GGM2-GGM11 heterodimer has an electrophoretic mobility very similar to that of the GGM2-GGM9-DNA complex, it can be concluded that GGM9 also forms a DNA-bound heterodimer with GGM2. Testing for heteromerization this way, a band of intermediate mobility can thus only indicate that GGM9 interacts with the partner protein but conclusions about the stoichiometry of the protein-DNA complex are difficult to draw.

As for the DNA-bound homodimers, the specificity of the heterodimer binding to DNA was tested and confirmed (Fig 3.6) by adding unlabeled probes S or Sm

Comparison of the EMSA results with that obtained by the yeast two-hybrid analyses and the pull-down assays yielded some notable differences (Fig 3.7). Whereas only GGM2 was able to form homodimers in yeast, and no homodimerization for any protein was observed in pull-down assays, all proteins except GGM15 were able to constitute DNA-bound homomeric complexes in EMSAs. In contrast, heterodimers of GGM2 and GGM3 as well as of GGM3 and GGM13 were reliably observed in yeast two-hybrid and pull-down analyses, but the respective proteins did not form a DNA binding complex in EMSAs.

While GGM2 interacts with all other proteins tested except GGM15, no interaction partner could be determined for GGM15 in neither in pull-down assay nor in EMSA (Fig 3.7). In one test, it was shown that the amount of translated GGM15 was less than that of GGM2 (Fig 3.8). Nevertheless, this could not explain why there was not a signal at all in for GGM15-DNA interaction (Fig 3.3 and 3.5) considering the sensitivity of radioactivity detection. The question remains to be further clarified whether it is that the production of GGM15 in heterologous systems is impaired or that GGM15 functions by interacting with

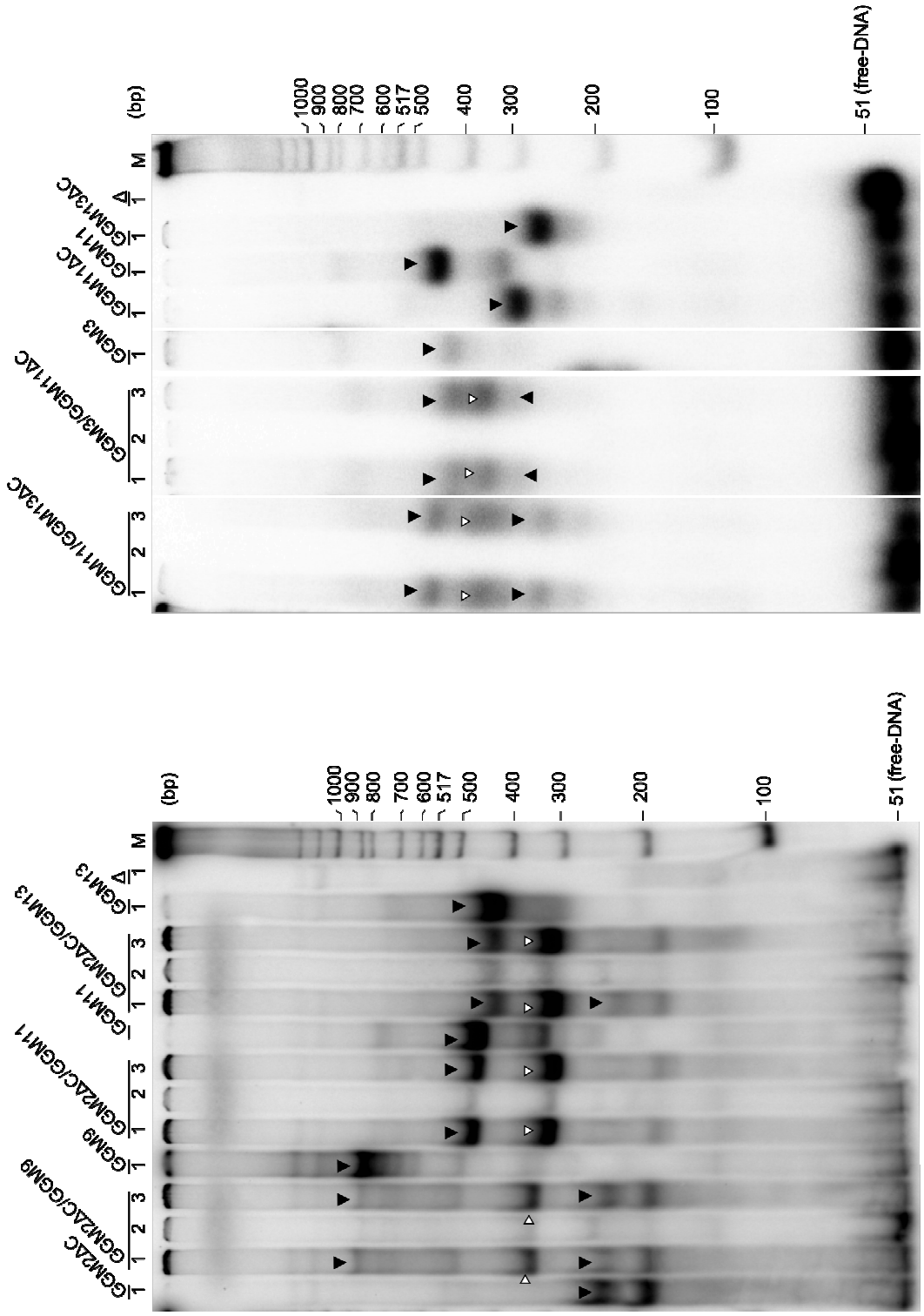

Fig 3.6 Specificity test of heterodimerization. Band assignment and legends see Fig 3.3 and Fig 3.4.

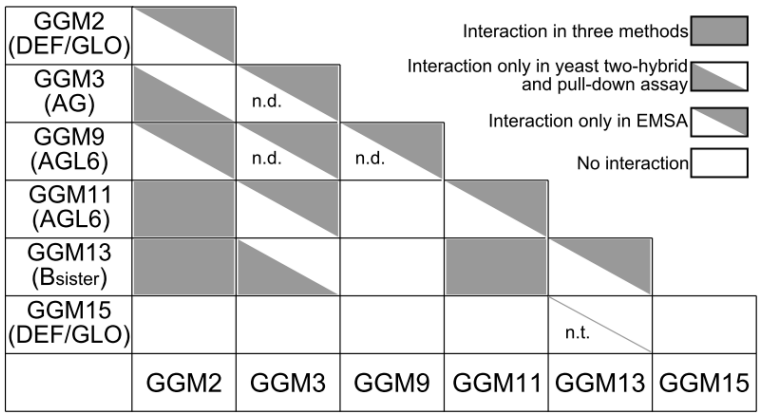

Fig 3.7 Comparison of the data obtained from the yeast two-hybrid and the *in vitro* pull-down assays with those from the EMSA experiments. Grey triangles indicate an interaction between the proteins. For yeast two-hybrid and pull-down assays, an interaction was counted only when observed with both methods.

unknown proteins instead of selected the proteins selected.

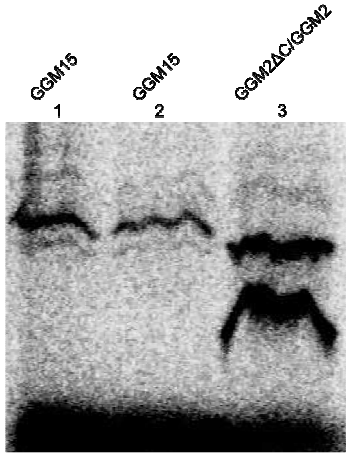

Fig 3.8 Comparison of translation efficiency. Lane 1: translation with 500 ng GGM15 plasmids. Lane 2: translation with 700 ng plasmids. Lane 3: co-translation with 250 ng C-terminal domain deleted GGM2 (GGM2ΔC) and 170 ng GGM2 plasmids. Proteins were analyzed by 12% SDS-PAGE.

3.2.3 Interaction of GGM2 and/or GGM3 with longer CArG-box containing probes

The results that both the B protein GGM2 and the C protein GGM3 can form DNA-binding homodimers (Fig 3.3) while they constitute heterodimers that are not capable of binding to DNA (Fig 3.5) suggest that the interaction between GGM2 and GGM3 may take place in a multimeric complex (see Fig 4.1). To test this possibility, probes L1 and L2, that carry one (probe L1) or two (probe L2) copies of a CArG-box were applied in EMSAs. Both probes were 151 bp in length. The sequence of the CArG-box was the same as that encoded on probe S.

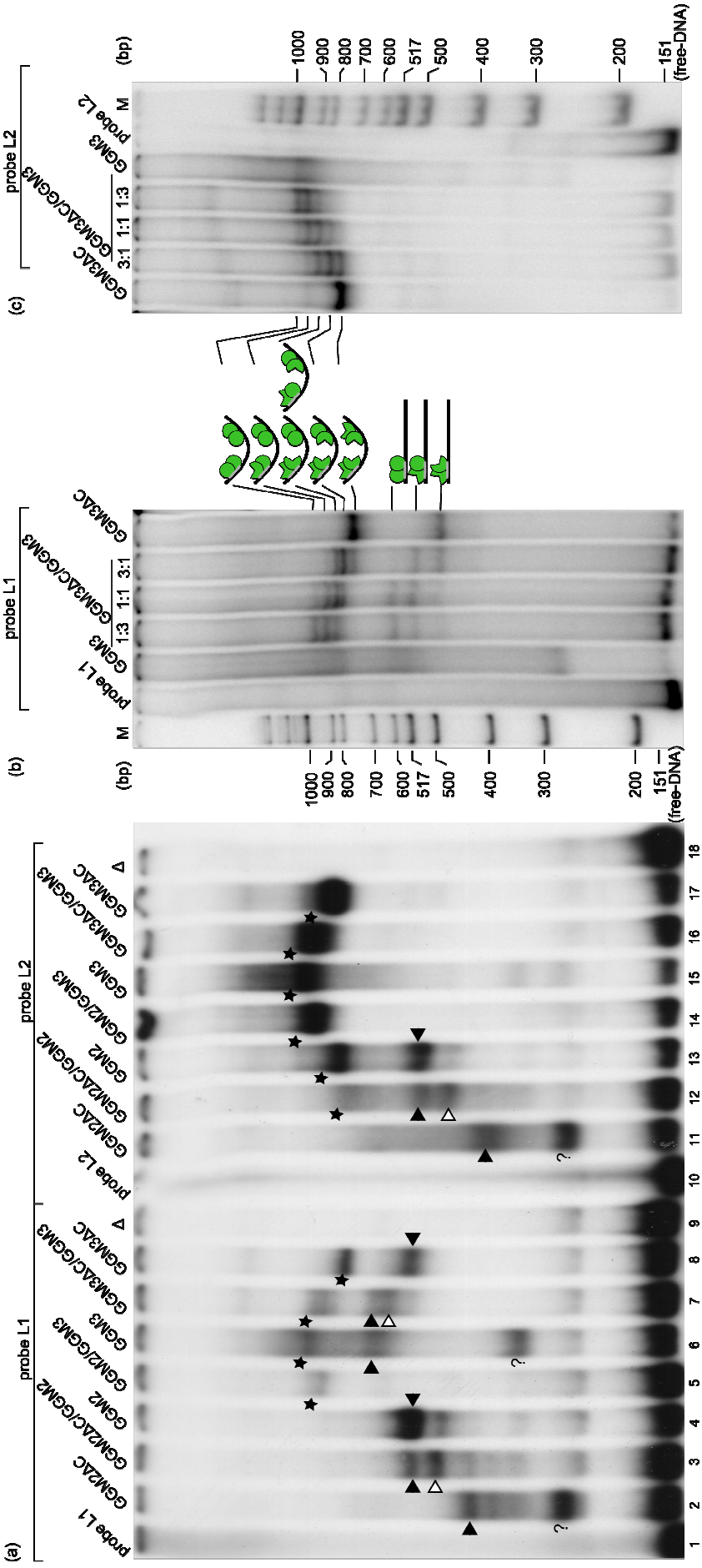

Fig 3.9 Interaction of GGM2 and GGM3 with probes L1 and L2 containing one or two CArG-boxes, respectively. Proteins (and volume ratio) and probes applied are noted above the gel. 'ΔC' indicates that the C-terminal region of the protein was deleted. Major complexes are indicated by black triangles (homodimers), white triangles (heterodimers), and stars (with more than one dimer), respectively. The feature of the fast mobility complex (indicated by question marker) appeared in lane 2 and 11 of (a), where GGM2ΔC was loaded, is unknow since that it was not appeared in lane 3 or 12 where GGM2/GGM2ΔC was loaded. The complex marked with '?' in lane 6 of (a) where GGM3 was loaded did not appeared in lanes of (b), so its characteristics remains unclear.

Co-incubation of GGM2 and GGM2ΔC with probe L1 yielded a complex which electrophoretic mobility was intermediate to that of GGM2-L1 and GGM2ΔC-L1 complexes (Fig 3.9a, lane 2-4). This indicated that GGM2 bound as a homodimer to probe L1. Surprisingly, however, when GGM2 was co-incubated with GGM3 or when GGM3 alone was incubated with probe L1 (that carries only one CArG-box), an additional complex of much lower electrophoretic mobility was formed (Fig3.9a, lane 5 and 6). As the molecular weigth of GGM2 and GGM3 is similar (24.52 kDa and 28.55 kDa, respectively), it appeared likely that these complexes of low mobility consisted of more than two proteins bound to probe L1. Complexes of very similar mobilities were observed when probe L2 was used (Fig 3.9a). However, intensity of the signals corresponding to the complexes of low mobility was stronger in these cases, indicating that DNA-binding efficiency or complex stability was increased (Fig 3.9a, compare lane 5-8 with lane 14-17).

To clarify the stoichiometry of the complex of low mobility formed by GGM3 and probe L1, further tests were made by combining different concentrations of GGM3 and its C-terminal domain deleted version (GGM3ΔC) (Fig 3.9b). From Fig 3.9b, eight bands are observed, and possible compositions were illustrated at the right side. Three of the signals probably represent single dimers bound to DNA, whereas the remaining five signals possibly represent two dimers or one tetramer bound to probe L1 (Fig 3.9b). Similar results were obtained with probe L2 (Fig 3.9c) (single dimers bound to DNA could not be observed here, probably because less DNA was loaded [comparing the free DNA of Fig 3.9b and c]). Therefore, it GGM3 did formed higher-order complexes with probe L1 and L2, but whether the complexes are composed of two dimers or are tetramers remained unclear.

3.3 Protein-DNA binding characteristics revealed by DNase I footprint assays

To further explore the capability of GGM2/GGM3 and GGM3 to form higher-order complexes, DNase I footprint assays were applied with probe L2 that gave stronger signals than L1 did.

DNase I preferentially cuts in widened minor grooves and is thus a sensitive detector of structural changes in DNA. It has been previously used to demonstrate DNA looping upon protein binding in quite a number of cases (Hochschild and Ptashne 1986; Krämer *et al.* 1987), including the formation of floral quartet-like complexes (Melzer and Theissen 2009;

Melzer *et al*. 2009).

First GGM2 and GGM3 were assayed separately to determine their ability to form looped complexes without a partner protein (Fig 3.10a). Incubation of GGM3 with probe L2 that are separated by 6 helical turns (63 bp) yielded a pattern typical for looped DNA, with sites of enhanced and diminished sensitivity spaced in approximately 5 bp intervals (Fig 3.10a, compare lanes 12 and 13; Fig 3.10b). In line with this, such a pattern was not observed for a DNA probe bound to a SEP3 protein that lacks the C-domain and part of K-domain (SEP3ΔK3C) and for which it was shown previously that it does not induce DNA looping (Melzer *et al*. 2009). Similarly, for GGM2-bound DNA, sites of enhanced and diminished sensitivity between the two CArG-boxes were barely detectable (Fig 3.10a, compare lanes 6 and 7; Fig 3.10b). This indicates that GGM2 alone does induce loop formation only very weakly, at best. In contrast, when GGM2 and GGM3 were co-incubated (Fig 3.10a, compare lanes 8 and 7, Fig 3.10b) the pattern of sites with enhanced and diminished sensitivity was more pronounced compared to GGM3 or GGM2 alone (Fig 3.10b). As protein-DNA complexes were separated from free DNA prior to analysis of the digestion pattern, this indicates that a DNA-bound complex containing GGM2 and GGM3 induces DNA loops that are more stable than those induced by GGM3 alone (Fig 3.10b).

If the pattern of sites with enhanced and diminished sensitivity is really caused by protein-induced DNA looping, it is expected to depend on the stereo-specific orientation of the CArG-boxes. If the CArG-boxes are separated by 6.5 instead of 6 helical turns, the two dimers bound are postulated to locate on different sides of the DNA helix. In this case, protein-induced looping requires energetically costly twisting of the DNA and thus is expected to happen less likely. Indeed, for GGM2 co-incubated with GGM3, especially the pattern of sites with diminished sensitivity was weaker when a probe (probe L2i) was used where the two CArG-boxes were spaced by 6.5 helical turns compared with probe L2 where two CArG-boxes were spaced by 6 helical turns (Fig 3.10a, compare lanes 8 and 7 with lanes 9 and 10; Fig 3.10b). This supports the conclusion that GGM2 and GGM3 together induce DNA looping upon binding. In contrast, when GGM2 or GGM3 were separately incubated with a probe where the two CArG-boxes were spaced by 6.5 helical turns, no obvious change in sensitivity was detected compared to experiments where probe L2 was used (Fig 3.10a, compares lanes 5 and 4 to lanes 6 and 7 as well as lanes 11 and 10 to lanes 12 and 13; Fig 3.10b). This provides further evidence that GGM2 alone does only

weakly loop DNA. For GGM3, this lack of stereo-specificity in binding was unexpected. However, as sites of enhanced and diminished sensitivity towards DNase I digestion were equally well pronounced for the two different DNA probes used, the results might indicate that GGM3 loops DNA irrespective of the stereo-specific orientation of the CArG-boxes.

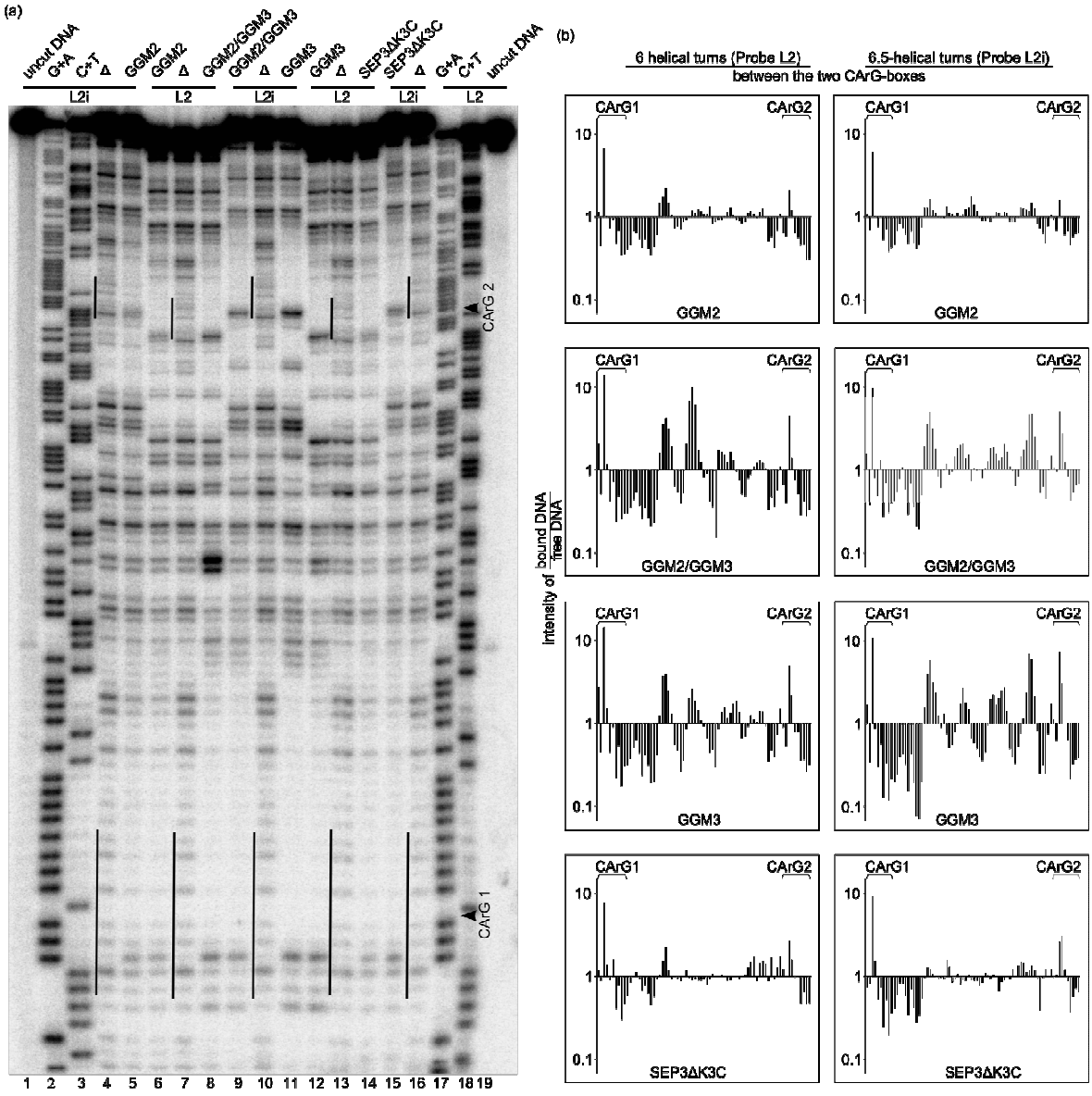

Fig 3.10 DNase I footprint assay.
(a) DNA probes ('L2' and 'L2i' for probe L2 and L2i, respectively) carrying two CArG-boxes were incubated with proteins as noted above the lanes. Lane numbers are marked at the bottom of the figure. 'Δ' denotes lanes in which reticulocyte extract with vector that did not contain a cDNA insert was added to the DNA probe. 'uncut DNA' denotes lanes in which DNA not treated with DNase I was added. G+A and C+T sequencing reactions of the two probes are shown for comparison. The CArG box regions are marked by bars and indicated by CArG1 and CArG2.
(b) Quantitative analysis of DNase I footprint signals shown in (a). The plots show the change of sensitivity to DNase I digestion after protein binding in single base pair-steps, using free DNA as a reference. Values were corrected for differences in DNA loading

3.4 Test for interaction of B and C proteins *in planta* via BiFC assays

The Bimolecular Fluorescence Complementation (BiFC) assay (Hu *et al.* 2002) is a relatively simple, non-invasive fluorescence-based method to detect protein-protein interactions in living cells. This method has been successfully used in plants by co-expressing two proteins in epidermal cells of tobacco leaves or *A. thaliana* cell culture protoplasts (Bracha-Drori *et al.* 2004; Tzfira *et al.* 2004; Verelst *et al.* 2007; Walter *et al.* 2004).

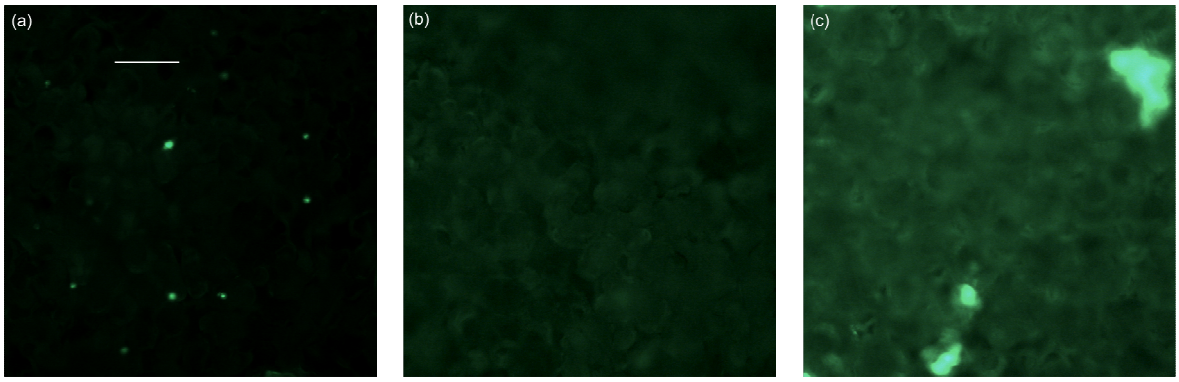

Fig 3.11 Test of protein-protein interaction *in planta* by BiFC. Two strains with inserted genes were transformed into tobacco leaves by syringe injection. Signals were detected under a fluorescent microscope. (a) transformation of pBatTL-N-YFP-SEP3ΔK3C and pBatTL-C-YFP-SEP3ΔK3C. (b) transformation of pBatTL-N-YFP-GGM2 and pBatTL-C-YFP-GGM3. (c) transformation of pBatTL-N-YFP-GGM3 and pBatTL-C-YFP-GGM2. Bar=100 µm.

To test if GGM2 and GGM3 can interact in plant cells, the BiFC assay was performed here. Full-length coding sequences of both of these two genes were subcloned into pBatTL vectors and to be translated into chimeric proteins carrying either the N-terminal or the C-terminal part of YFP. Subsequently, the constructs with the N-terminal and C-terminal part of YFP were transformed together into tobacco leaves via agrobacteria. A construction derived from *SEP3* lacking C-domain and part of the K-domain (SEP3ΔK3C which is known to interact with itself in BiFC (Melzer *et al*. 2009) was used as a positive control. When *SEP3ΔK3C* was transformed, there was a clear fluorescent signal (Fig 3.11a), but when GGM2 and GGM3 were co-transformed no fluorescent signal was detected (Fig 3.11b-c). The reason that the interaction of GGM2 and GGM3 was not observed *in planta* remains to be clarified.

4 Discussion

In recent years, much has been learned about the genetic and molecular mechanisms specifying floral organ identities in angiosperms. The molecular mechanisms of how the transcription factors encoded by floral homeotic genes exert their function are beginning to be elucidated. The floral quartet model proposes that floral homeotic proteins act as DNA-bound tetrameric complexes to specify organ identity. Many studies on protein-protein interaction and protein-DNA interaction of MADS-domain floral homeotic proteins have been carried out revealing not only dimerization but also tetramerization of MIKC-type MADS-domain proteins (Davies *et al*. 1996; de Folter *et al*. 2005; Immink *et al*. 2009; Immink *et al*. 2003; Leseberg *et al*. 2008; Masiero *et al*. 2004; Riechmann *et al*. 1996a; Riechmann *et al*. 1996b; Schwarz-Sommer *et al*. 1992; Tsai *et al.* 2008). Considering the importance of floral homeotic proteins in floral organ specification, it is thus reasonable to study the interaction network of these regulatory proteins in gymnosperms towards clarifying the origin of the flower.

Phylogenetic conservation, expression pattern and mutant analyses suggest that the basic principles of floral organ specification as formulated by the ABC model apply to all angiosperms (Kim *et al*. 2005; Kramer and Irish 2000; Ma and dePamphilis 2000; Soltis *et al.* 2009; Theissen *et al*. 2000; Whipple *et al*. 2007). Furthermore, the basic assumptions of the floral quartet model on the complex formation of floral homeotic proteins have been predicted to be similar in all angiosperms (Kaufmann *et al*. 2005; Theissen and Melzer 2007).

The center of this research focuses on the origin and evolution of the highly conserved regulatory network controlling floral organ specification and -- along with and intrinsically tied to that -- on the origin of the angiosperm flowers. To obtain detailed answers, it is inevitable to study gymnosperms, the closest extant relatives of the angiosperms. Thus, the interaction properties of gymnosperm (*Gnetum gnemon*) MADS-domain transcription factors that are closely related to floral homeotic proteins were explored.

4.1 Method selection for protein-protein interaction and protein-DNA interaction

Dimerization is among the well-characterized features of MADS-domain proteins. Yeast hybrid assays have been widely used in interaction studies of MADS-domain protein, of

e.g. *A. thaliana* (de Folter *et al*. 2005; Immink *et al*. 2009), *P. hybrida* (Immink *et al*. 2003), *S. lycopersicum* (tomato) (Leseberg *et al*. 2008), and also *G. gnemon* (Melzer 2005), and these assays have provided lots of information about protein-protein interactions. Being a heterologous system, however, it is necessary to compare the results obtained with this this assay with those of other methods. An *in vitro* pull-down assay was thus established in this project to study protein-protein interactions of *G. gnemon* MADS-domain proteins.

Glutathione S-transferase (GST) and hexa-histidine tags are two among commonly used affinity tags in pull-down assays. GST is a 26 kDa protein and can be purified by affinity chromatography using immobilized glutathione. His_6-tagged proteins incorporate a hexa-histidine stretch that can be captured by its high-affinity binding to nickel (Ni) metal ions. Usually, matrices that have strongly metal-chelating nitrilotriacetic acid (NTA) groups covalently bound to their surfaces are used for immobilization of His_6-tagged proteins. Although both GST and His_6-tag fusions are widely used, the latter one has several advantages. First, the His_6-tag can be introduced at the N- or C-terminus of the bait proteins, which provides an option when one way of fusion does not work. Second, the interaction of the His_6-tag with Ni-NTA depends on its primary structure which means the tag can bind to immobilized matrices only if it is not buried inside proteins. Last, the tag has little chance to influence the function of fused proteins because of small size (Jacky *et al.* 1993). So in this study the His_6-tag fusion system was employed.

Specific interaction between transcription factors and their *cis*-regulatory elements is the fundament of gene regulation. EMSA (Garner and Revzin 1981) is a well established method. The principle of this method is that the binding of a DNA fragment to a protein usually retards the fragment's mobility in native polyacrylamide gel or agarose gel. The retarded DNA and free DNA can be easily visualized in gel electrophoresis by labeling DNA (reviewed in Lane *et al.* 1992). Several studies have shown the successful application of this method in case of MADS-domain proteins (Egea-Cortines *et al*. 1999; Riechmann *et al*. 1996b; Winter *et al*. 2002a).

MADS-domain proteins bind as dimers to CArG-boxes (Pollock and Treisman 1990; Wynne and Treisman 1992). Sequence selection experiments performed for AG, SHP1, SEP3 and SEP4 show that this also applies to MADS-domain proteins involved in flower development (Huang *et al.* 1993; Huang *et al.* 1996; Huang *et al.* 1995). CArG-boxes that

have the canonical $CC(A/T)_6GG$ motif are among the groups which has high affinity to MADS-domain proteins *in vitro* (Huang *et al*. 1993; Huang *et al*. 1996; Huang *et al*. 1995; Riechmann *et al*. 1996b). So a probe containing a CArG-box derived from the regulatory intron of *AG* was used.

DNA loop formation by two protein dimers is predicted by the floral quartet model. The DNase I footprint assay has been successfully used to illustrate loop formation of the lambda and lac repressors (Hochschild and Ptashne 1986; Krämer *et al*. 1987) and of the floral homeotic protein SEP3 (Melzer *et al*. 2009). The principle underlying this application is that since DNase I preferentially cuts in widened minor grooves, this makes it a sensitive detector of structural changes of DNA. The loop formation is revealed by a periodic change of sensitivity to DNase I cleavage. However, if two dimers are bound on different sides of the DNA helix, additional energy is needed to twist the DNA for loop formation and thus loop formation is often unstable or happens less likely in these cases. Rotating on binding site half a helical turn with respect to the other and thus hamper loop formation can be easily achieved by introducing a short insertion (5-bp in length, namely half DNA helical turn) between the two protein-binding sites of DNA.

4.2 Higher-order complex and the specification of reproductive organ identity in *G. gnemon* - Evidence for tetrameric MADS-domain protein complexes in gymnosperms

To determine the pattern of protein-protein interactions, the results from pull-down assays were compared to that from yeast two-hybrid analyses performed previously (Melzer 2005). As shown in Fig 3.2, largely overlapping results were obtained, suggesting that either of the two methods is suitable for determining interactions between MADS-domain proteins. The disagreement of some protein interactions between yeast two-hybrid and pull-down assays might be partial explained as yeast two-hybrid assays have the capability of detecting some unstable interactions while pull-down assays usually only stable interaction (Jensen 2004). In addition, for all of the proteins that interact (Fig 3.2) their corresponding mRNA spatial expressions also overlapped with the exception of GGM2 and GGM13 (Becker *et al*. 2002; Becker *et al*. 2003; Becker *et al*. 2000; Winter *et al*. 1999) which provides the possibility that two proteins could interact *in planta*. Although interaction of GGM2 and GGM13 was detected by all three methods (Fig 3.7), the mRNA of the B-gene *GGM2* is restricted to male reproductive organs while that of the closely related B_{sister} gene *GGM13* appears to be limited to female reproductive organs. Since B

and B_{sister} genes are sisters to each other (Becker *et al*. 2002), their interaction might be a remnant of a homodimerizing ancestor.

The interaction between GGM2 and GGM3 was sufficiently strong to be detected in yeast two-hybrid and pull-down assays (Fig 3.2) while it is the homodimers but not the heterodimers of these proteins that could bind to probe S (Fig 3.5). However, when using probe L1 and L2, which are tree times as long as probe S, a band of low mobility indicative of a multimeric complex of GGM2 and GGM3 was detected even if the probe contains only one CArG-box (in case of probe L1) (Fig 3.9). Thus, a CArG-box and a second (unknown) site of probe L1 with very weak affinity might suffice for a GGM2-GGM2/GGM3-GGM3-DNA complex to be formed (Fig 4.1), similar to the case of AP3, PI, and SEP3 interaction (Melzer and Theissen 2009). Taken together, these data indicate that the interaction between GGM2 and GGM3 is sufficiently strong to be detected under various conditions. Furthermore, DNase I footprint results suggest that GGM3 alone or in combination with GGM2 is capable of binding to two adjacent CArG-boxes and looping the intervening DNA. For unknown reason, however, interaction of GGM2 and GGM3 was not observed when they were transiently expressed in tobacco leaves (Fig 3.11). Taken together, it thus appears likely that at least some gymnosperm MADS-domain proteins (for instance GGM2 and GGM3, or even GGM3 alone) possess the ability to form tetrameric, DNA-bound complexes that are similar to the ones postulated to confer floral organ identity in angiosperms (Fig 4.1).

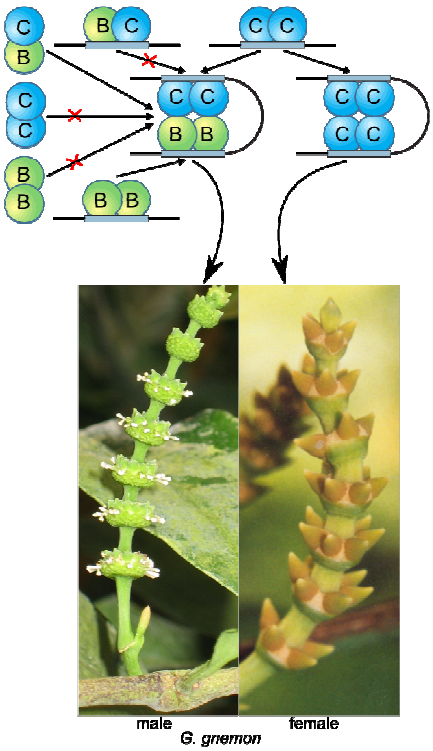

Fig 4.1 Rationale of conclusions concerning dimeric and higher order interactions. *G. gnemon* B- and C-proteins interacted in yeast two-hybrid and pull down assays, but did not form DNA-binding heterodimers with each other. In contrast, DNA-binding B- and C-homodimers were detected. Thus, a heterotetrameric complex of B- and C-proteins is proposed. Such a complex might function in specification of male organs. A homotetramer composed of C-proteins might specify female organs.

The two AGL6-like proteins GGM9 and GGM11 have different interaction partners: GGM11 can interact with both the B protein GGM2 and the B_{sister} protein GGM13 in all

three methods used (Fig 1.7, 3.1 and 3.5; summarized in Fig 3.7) and with the C protein GGM3 in pull-down assays (Fig 3.1) and EMSA (Fig 3.5), while GGM9 only interacts with GGM2 observed in yeast two-hybrid assays (Fig 3.2) and EMSAs (Fig 3.5). The expression of *GGM9* and *GGM11* was found both in male and female reproductive organs (Winter *et al*. 1999). However, *GGM9* is not expressed in the sterile ovules of male reproductive organs. Compared to *GGM9*, *GGM11* showed a restricted expression pattern in male cones, but it is strongly expressed in the sterile ovules of the male cones. The differences in the expression patterns of *GGM9* and *GGM11* indicates that GGM9 is not involved in the general morphological constitution of female organ but rather, maybe together with GGM13, in specifying fertile female reproductive organs.

G. gnemon has three types of reproductive organs, namely sterile ovule, male cones, and fertile female cones. Similar to the floral quartet model in angiosperms, the protein interaction patterns raise two alternative possibilities concerning how reproductive organs are specified in *G. gnemon* (Fig 4.2): if the two AGL6-like proteins (GGM9 and GGM11) do not function to 'glue' other proteins forming higher order complexes as their angiosperm counterparts do, it might be that the complex containing two C protein homodimers (GGM3/GGM3) alone specifies sterile ovules or fertile female cones, B homodimer (GGM2/GGM2) with C homodimer (GGM3/GGM3) male cones, and C homodimer (GGM3) with B_{sister} homodimer (GGM13) fertile female cones (Fig 4.1, 4.2a). Supporting this scenario, the DNA-binding homodimers composing the higher-order complexes were observed in EMSA, and the non-DNA binding dimers GGM2/GGM3 and GGM13/GGM13 which may be involved in protein-protein interaction of two dimers were also detected in yeast two-hybrid and pull-down assays (Fig 3.2). The footprint data support that the C protein GGM3 alone already has the capability to form a tetramer, which indicates the possibility for the GGM3 tetramer to specify female cones (Fig 4.1). In this model, homodimerization plays important roles in forming higher-order complexes specifying reproductive organ identity.

Recently *AGL6*-like genes were shown to play important roles in angiosperm floral organ identity and floral meristem identity (Li *et al*. 2009; Ohmori *et al*. 2009; Rijpkema *et al*. 2009; Thompson *et al*. 2009). Although their function remains to be investigated more deeply, we can see the possibility that the *G. gnmenon* AGL6-like proteins play a role about how reproductive organs are specified in *G. gnemon*: the complex containing C protein and AGL6 protein (GGM3/GGM3 and GGM11/GGM11 dimers, or two GGM3/GGM11

dimers) specifies sterile female ovules; B, C and two AGL6 proteins (GGM2/GGM9 and GGM3/GGM11 dimers) male cones; and B_{sister}, C and two AGL6 proteins (GGM13/GGM9 and GGM3/GGM11) female cones (Fig 4.2b). In supporting of this scenario, the DNA-binding heterodimers constituting different complexes were also detected in EMSA (Fig 3.5), and both the non-DNA binding dimers, namely B/C (GGM2/GGM3) and B_{sister}/C (GGM13/GGM3), are also observed in both yeast two-hybrid and pull-down assays (Fig 3.2). Therefore, the complex which specifies the sterile female cones probably constitutes two heterodimers instead of homodimers. It is then very likely that heterodimerization plays a very important role in specifying reproductive organ identity. At this moment there is no data indicating which hypothesis is more close to the reality.

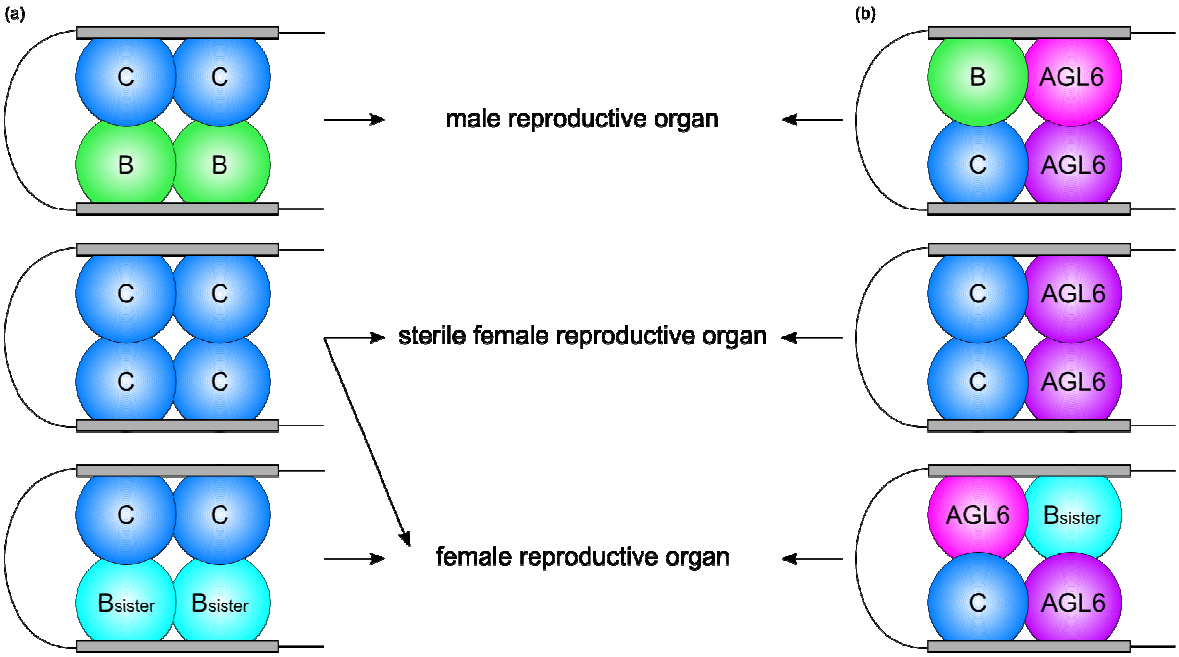

Fig 4.2 Alternative hypotheses of specification of *G. gnemon* reproductive organ identities. *G. gnemon* has male and female cones. Very often in male cones, there are sterile ovules. Two hypotheses suggest different higher-order complexes that specify these three kinds of reproductive organs.

There are supporting evidences of the above two hypotheses about reproductive organs specification in *G. gnemon* from mRNA spatial expression patterns: in sterile ovules *GGM3*, *GGM11* were expressed, in male cones *GGM2*, *GGM3*, *GGM9* and *GGM11* were expressed, and in fertile female cones *GGM3*, *GGM13*, *GGM9* and *GGM11* were expressed.

4.4 Conservation and diversity amongst MADS-domain protein interactions

Members of the DEF/GLO, AG, AGL6, and B_{sister} clades of MADS-domain proteins play

important roles during flower development in eudicots (Becker and Theissen 2003; Coen and Meyerowitz 1991; de Folter *et al*. 2006; Ma and dePamphilis 2000; Zahn *et al.* 2006; Zahn *et al.* 2005b).

In *A. thaliana*, few direct interactions are observed between proteins belonging to the clades studied here (DEF/GLO, AG, AGL6, and B_{sister}), except the only two interactions (AG with AGL6 and AGL13 of *A. thaliana*) detected in yeast two-hybrid assays (de Folter *et al*. 2005; Fan *et al.* 1997). Rather, the interaction network of MADS-domain proteins in *A. thaliana* is organized hierarchically, with proteins belonging to the SQUA- and SEP-like proteins acting as hubs that have many interaction partners (Fig 4.3). In *G. gnemon* and possibly in all gymnosperms, proteins belonging to the SEP-like and AP1-like subfamilies are probably absent. The data presented here indicate that their role as central hubs in the interaction network has not been overtaken by other proteins in gymnosperms but that the network is organized less hierarchically, with extensive direct interactions between the proteins belonging to different clades (Fig 4.3).

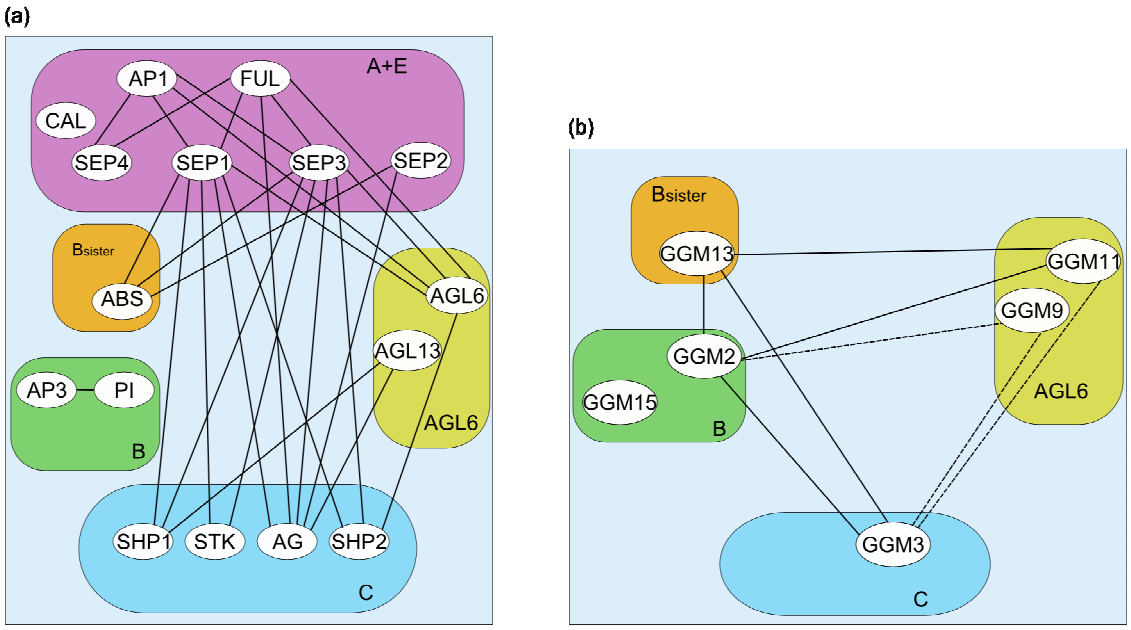

Fig 4.3 Comparison of the heterodimeric MADS-domain protein interaction network from *G. gnemon* and *A. thaliana*. Proteins are indicated by ovals and protein interactions by lines. Boxes enframing the proteins assign their affiliation to different subfamilies. For *G. gnemon*, an interaction is indicated by solid lines when detected in both yeast two-hybrid and pull-down assays or by dashed lines detected only in one method. Data from *A. thaliana* are taken from de Folter *et al*. (2005) and Yang *et al*. (2003).

One line of reasoning to explain the existence of hubs in the MADS interactome of *A. thaliana* is that they function as mediators of multimeric complex formation. Indeed, SEP3,

which is one of the central hubs in *A. thaliana* (Immink *et al*. 2009; Theissen and Saedler 2001), has been shown to be one of the main mediators of multimeric complex formation (Immink *et al*. 2009). These multimeric complexes are assumed to play pivotal roles during flower development. For example, complexes composed of either SEP3 and AG or SEP3, AG, AP3, and PI are postulated to function in stamen and carpel specification respectively in *A. thaliana* (Honma and Goto 2001). The data presented here indicate that in *G. gnemon* the AG-like protein GGM3 alone or in combination with the DEF/GLO-like protein GGM2 is already capable of forming such multimeric complexes (Fig 3.10, 4.1). Thus, the lack of hubs at the level of dimeric interactions is mirrored by the ability of GGM3 and GGM2 to form multimeric complexes in the absence of an additional mediator. It is thus hypothesized that female organ identity in *G. gnemon* is specified by a homotetramer of GGM3 whereas a GGM2-GGM2/GGM3-GGM3-DNA complex functions in male organ specification (Fig 4.2a).

The question arises as to why and how these differences in complex formation were established. It might well be that nothing but neutral changes led to differences in the network topology during evolution. There are examples from other systems, mating in yeast (Tsong *et al.* 2006) for instance, indicating that transcriptional regulation networks specifying certain developmental programs might change substantially during evolution without affecting the developmental program itself. In case of plant MADS-domain proteins, evidence from domain swapping analyses and reverse two-hybrid screens (Yang *et al*. 2003a, Yang *et al*. 2003b) indicate that the regions determining interaction strength and specificity are usually rather small, in some cases single amino acid substitutions have been shown to significantly hamper an interaction that can however be restored by compensatory mutations in the partner protein. Thus, destroying or establishing a heterodimer interaction surface or switching from multimerization to dimerization and back might require only few mutational changes that are easily accomplished.

However, if changes in network topology reflect neutral changes, the question arises as to why the interaction network of MADS-domain proteins appears to be more or less conserved at least in higher eudicots (de Folter *et al*. 2005). This and the intimate link between the floral homeotic genes and the evolution of the flower tempt us to speculate that the hierarchical organized network of MADS-domain protein interactions in eudicots and possibly in all angiosperms provided some selective advantage over the mesh-like

network in gymnosperms.

For example, the need of an additional protein in angiosperms to form multimeric complexes and hence to specify reproductive organ identity might introduce an additional level of robustness. This would be similar to the obligate heterodimerization of class B proteins observed in higher eudicots. In this case, the B function can only be performed if both partners, an AP3-like and a PI-like protein, are present, heterodimerize and establish an autoregulatory feedback loop (Schwarz-Sommer *et al*. 1992). Computational analyses indicate that obligate heterodimerization can, compared to homodimerization, be used to increase the robustness of the system while keeping the sensitivity to input signals constant (Lenser *et al.* 2009). Along these lines, by making the formation of reproductive organs depending from an additional partner, the system might be more robust against random fluctuations of either of the selector proteins. Also, the necessity to express several proteins at once to determine organ identity decrease the risk of heterochrony and heterotopy, (i.e. organ formation at the wrong time or place) and thus contributes to the canalization of flower development (Winter *et al*. 2002b).

Notably, it was recently shown that several angiosperm *AGL6*-like genes can also confer class E floral homeotic functions (Li *et al*. 2009; Ohmori *et al*. 2009; Rijpkema *et al*. 2009; Thompson *et al*. 2009). Although in the same superclade, in contrast to *SEP1*-like and *SQUA*-like genes, *AGL6*-like genes are present in gymnosperms. It is well possible that gymnosperm *AGL6*-like genes are capable of mediating multimeric complex formation, although this remains to be determined experimentally. For instance, one of the AGL6-like protein, GGM9, shows unusual lower mobility in EMSA compared to other proteins (Fig 3.3). The lower mobility complex might comprise more than two GGM9 proteins, indicating GGM9 alone already requires capability forming higher order complex(es), although it cannot be ruled out that the low electrophoretic mobility is due to unanticipated changes in DNA conformation upon GGM9 binding. However, from the network topology in *G. gnemon* it is not evident that they act in a similar way as hubs as SEP1- and SQUA-like proteins do in *A. thaliana* (Fig 4.3). Therefore, even if *G. gnemon* AGL6-like proteins are capable of forming multimeric complexes with AG-like and DEF/GLO-like proteins (Fig 4.2b), there is, in contrast to what is postulated for eudicots, no strict dependence on these additional mediators of complex formation. However, selective advantages like developmental robustness and canalization of development are effective

only if such a strict dependence does exist.

In summary, crucial features of the molecular network controlling floral organ identity in angiosperms are evident in gymnosperms as well. However, assuming that the gymnosperm network represents an ancestral state, it might have been the selective loss of some interactions on the one hand and the establishment of hubs on the other hand that constituted the present day network in flowering plants and that paved the way to the evolution of the flower.

4.5 Homodimerization *vs*. heterodimerization of MADS-domain proteins

With the exception of some DEF- and GLO-like proteins of eudicots (Winter *et al*. 2002b), almost all floral homeotic proteins tested so far form homodimers that bind to CArG-box DNA (Ciannamea *et al.* 2006; Huang *et al*. 1996; Huang *et al*. 1995; Riechmann and Meyerowitz 1997; Riechmann *et al*. 1996b; Shiraishi *et al.* 1993; Tang and Perry 2003; West *et al.* 1998). This also holds true for the MADS-domain proteins from *G. gnemon* tested here. This supports the notion that homodimerization upon DNA binding is a common and ancient feature of MADS-domain proteins (Kaufmann *et al*. 2005). In contrast, except for GGM2, homodimerization was not demonstrated with the yeast two-hybrid system or *in vitro* pull-down assays. Difficulties in detecting homodimerization of MADS-domain proteins using the two-hybrid system have frequently been reported. For instance, in a comprehensive yeast two-hybrid test, de Folter *et al*. (2005) found more than 200 heterodimers but only five homodimers. Thus, in contrast to heterodimers, homodimers of MIKC-type MADS-domain proteins might more strictly depend on CArG-box binding to be formed and detected. However, DNA binding might not be the only factor facilitating homodimer formation, as *in planta* evidence using fluorescence labeled petunia MADS-domain proteins suggests that homodimerization occurs in the cytoplasm prior to nuclear translocation and DNA binding (Immink *et al.* 2002). Based on these findings, it was suggested that homodimerization might, in contrast to heterodimerization, more often require plant specific cofactors (Immink *et al*. 2002).

In summary, these data can be taken as evidence suggesting that homodimeric interactions are often weaker than heterodimeric interactions and thus need additional factors (i.e. either plant specific cofactors or CArG-box DNA) to be stabilized.

There is not a satisfactory explanation for this phenomenon. It should be noted, however, that there is evidence from other proteins that the requirements on interaction surfaces are different for homodimerization and heterodimerization (Jones and Thornton 1996; Zhanhua *et al.* 2005). It is conceivable that during evolution selection acted on the formation of heterodimers to increase the number of possible complexes that can be formed. This way, the possibilities to control gene expression by differential complex formation were maximized. On the other hand, this might have run at the cost of homodimer stability.

4.6 Variations in B protein interactions

Angiosperm B proteins are shown to form DNA-binding dimers with B proteins only (Winter *et al*. 2002b). The paradigms are DEFICIENS (DEF) (Sommer *et al.* 1990) and GLOBOSA (GLO) (Tröbner *et al.* 1992) from *A. majus* that have been shown to interact only with each other but not with other MADS-domain proteins (Davies *et al*. 1996). Both B proteins from *G. gnemon* diverge from this pattern, although in different directions. Whereas GGM2 interacted with all other proteins tested, neither homodimerization nor heterodimerization for GGM15 was detected. As stated above, the interaction properties of MADS-domain proteins from *G. gnemon* appear to be less constrained than that of MADS-domain proteins from angiosperms. The interaction behavior of GGM2 does well integrate into this general pattern.

In one test, it was shown that the amount of translated GGM15 was less than that of GGM2 (Fig 3.8). Nevertheless, this could not explain why there was no signal at all for GGM15-DNA interaction (Fig 3.3 and 3.5) considering the sensitivity of radioactivity detection. It might be that, due to some unknown reasons, only few (if any) functional *in vitro* translated GGM15 protein was obtained. This could potentially explain the results of the pull-down assay and EMSA, but not the results of the yeast two-hybrid assay (Fig 1.7) unless it is argued that GGM15 needs some plant-specific post-translational modification which is not available in yeast or in the *in vitro* translation system which is a coupled rabbit reticulocyte lysate system.

On one hand, it cannot be completely ruled out that the production of GGM15 in heterologous systems is impaired, and that this impeded the detection of interaction partners. On the other hand, the difference of GGM2 and GGM15 in interacting with other

GGM proteins might be due to their different function *in planta.* It has been shown that the expression of GGM15 is spatially more restricted to the antherophore than that of GGM2, and therefore it was proposed that GGM2 has a more general role in specifying male reproductive organs (Becker *et al*. 2003).

One of the putative orthologues of *GGM2* in *Picea abies* (Norway spruce) is *DAL11*, whereas the putative orthologue of *GGM15* is *DAL12* (Winter *et al*. 2002a). Similarly to the results here, DAL11 interacted with several other proteins in yeast two-hybrid assays, whereas DAL12 did not (Sundström and Engström 2002). It could well be, therefore, that lack of interaction with the 'usual suspects' might be a general feature of the proteins belonging to this specific subclade. When ectopically expressed in *A. thaliana*, both, *DAL11* and *DAL12* conferred similar phenotypes (Sundström and Engström 2002), corroborating the view that DAL12 and, by inference, also GGM15 are functional proteins. How they function remains elusive, however. Possibly plant-specific cofactors are required for GGM15 and DAL12 to interact with other proteins**.**

In summary, crucial features of the molecular network controlling floral organ identity in angiosperms are evident in gymnosperms as well. However, assuming that the gymnosperm network represents an ancestral state, it might have been the selective loss of some interactions on the one hand and the establishment of hubs on the other hand that constituted the present day network in flowering plants and that paved the way for the evolution of the angiosperm flower.

5 Appendices

5.1 Abbreviation

APS	Ammonium Persulfate
BiFC	Bimolecular fluorescence complementation
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
cpm	counts per minute
DMSO	Dimethyl Sulfoxide
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assays
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
LB	Luria Bertani media
NaAC	Sodium Acetate
PAGE	Polyacrylamide Gel Electrophoresis
rpm	revolutions per minute
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TAE	Tris-Acetate-EDTA buffer
TBE	Tris-Borate-EDTA buffer
TEMED	N,N,N',N'-Tetramethylethylenediamine

5.2 Buffer and solutions

5.2.1 Home-made buffer and stock solutions

6× DNA loading dye

	Amount	Final concentration
1M Tris-HCl (pH 8.0)	5 ml	10 mM
5% Bromophenol blue	3 ml	0.03%
5% Xylene cyanol FF	3 ml	0.03%
100% Glycerol	300 ml	60%
0.5 M EDTA	60 ml	60 mM
Distilled water	up to 500 ml	
Total volume	500 ml	

Formamide loading buffer

	Amount	Final concentration
Deionized formamide	950 µl	95%
0.5 M EDTA (pH 8.0)	20 µl	10 mM
5% Xylene cyanol FF	10 µl	0.05%
Bromophenol blue	10 µl	0.05%
Distilled water	10 µl	
Total volume	1 ml	

Gel elution buffer (to elute DNA from native polyacrylamide gel)

	Amount	Final concentration
10 M Ammonium acetate	750 µl	500 mM
0.5 M EDTA (pH 8.0)	30 µl	1 mM
Distilled water	14.22 ml	
Total volume	15 ml	

Gel fixing solution

	Amount	Final concentration
Methanol	250 ml	50%
Acetic acid	50 ml	10%
Water	200 ml	40%
Total volume	500 ml	

LB medium

	Amount	Final concentration
Tryptone	10 g	1%
Yeast extract	5 g	0.5%
NaCl	5 g	0.5%
Agar (for solid LB)	10 g	1%
Distilled water	up to 1 L	
Total volume	1 L	

6× SDS loading dye

	Amount	Final concentration
1 M Tris-HCl (pH 8.0)	15 ml	300 mM
100% Glycerol	30 ml	60%
SDS	300 mg	6%
5% Bromophenol blue	600 µl	0.06%
DTT	2.3 g	300 mM
Distilled water	up to 50 ml	
Total volume	50 ml	

10× SDS running buffer

	Amount	Final concentration
Tris base	30 g	250 mM
Glycine	144 g	1.9 M
10% SDS	100 ml	1%
Distilled water	up to 1 L	
Total volume	1 L	

10× STE

	Amount	Final concentration
1 M Tris-HCl (pH 8.0)	100 µl	10 mM
5 M NaCl	100 µl	0.1 M
0.5 M EDTA (pH 8.0)	20 µl	1 mM
Distilled water	780 µl	
Total volume	1 ml	

50× TAE

	Amount	Final concentration
Tris base	216 g	2 M
Acetic acid	57 ml	5.7%
0.5 M EDTA (pH 8.0)	100 ml	50 mM
Distilled water	up to 1 L	
Total volume	1 L	

5× TBE

	Amount	Final concentration
Tris base	54 g	0.5 M
Boric acid	27.5 g	0.44 M
0.5 MEDTA (pH 8.0)	20 ml	10 mM
Distilled water	Up to 1 L	
Total volume	1 L	

10× TE (Tris-HCl-EDTA), pH 8.0

	Amount	Final concentration
1 M Tris-HCl (pH 8.0)	1 ml	100 mM
0.5 M EDTA (pH 8.0)	2 ml	100 mM
Distilled water	7 ml	
Total volume	10 ml	

YEB medium (pH 7.2)

	Amount	Final concentration
Beef extract	5 g	0.5%
Yeast extract	1 g	0.1%
Trypton	5 g	0.5%
Sucrose	5 g	0.5%
2 M $MgSO_4$	1 ml	2 mM
Agar (for solid YEB)	10 g	1%
Distilled water	up to 1 L	
Total volume	1 L	

Buffer for DNase I footprint assay

DNase I dilution buffer

	Amount	Final concentration
0.2 M HEPES, pH7.3	45 µl	9 mM
10 mg/ml BSA	10 µl	0.1 mg/ml
1 M $MgCl_2$	30 µl	30 µM
1 M $CaCl_2$	5 µl	5 µM
Distilled water	910 µl	
Total volume	1 ml	

Buffer for EMSA

3× EMSA buffer

	Amount	Final concentration
BSS buffer	120 µl	-
0.1 M Spermidine	4.8 µl	3.9 mM
Total volume	124.8 µl	

BSS buffer

	Amount	Finial concentration (counted in 3× EMSA buffer)
0.5 M EDTA (pH 8.0)	28.25 µl	4 mM
10 mg/ml BSA	300 µl	0.9 mg/ml
0.2 M HEPES (pH 7.3)	450 µl	27 mM
1 M DTT	9 µl	2.7 mM
10 µg/ µl Salmon sperm DNA	57.8 µl	0.173 µg/ µl
Distilled water	2370 µl	
Total volume	3215 µl	

Buffer for pull-down assay

Binding buffer I

	Amount	Final concentration
1 M NaH_2PO_4	2.5 ml	50 mM
5 M NaCl	3 ml	300 mM
1 M Imidazole	1 ml	20 mM
Distilled water	up to 50 ml	
Total volume	50 ml	

Washing & Interaction (W&I) buffer I

	Amount	Final concentration
1 M NaH_2PO_4	2.5 ml	50 mM
5 M NaCl	3 ml	300 mM
1 M Imidazole	1 ml	20 mM
Tween20	25 µl	0.05%
Distilled water	up to 50 ml	
Total volume	50 ml	

Elution buffer I

	Amount	Final concentration
1 M NaH_2PO_4	2.5 ml	50 mM
5 M NaCl	3 ml	300 mM
1 M Imidazole	12.5 ml	250 mM
Tween20	25 µl	0.05%
Distilled water	up to 50 ml	
Total volume	50 ml	

Binding buffer II

	Amount	Final concentration
1 M NaH_2PO_4	2.5 ml	50 mM
5 M NaCl	10 ml	1 M
1 M Imidazole	2 ml	40 mM
Distilled water	up to 50 ml	
Total volume	50 ml	

W&I buffer II

	Amount	Final concentration
1 M NaH_2PO_4	2.5 ml	50 mM
5 M NaCl	10 ml	1 M
1 M Imidazole	2 ml	40 mM
Tween20	50 µl	0.01%
Distilled water	up to 50 ml	
Total volume	50 ml	

Elution buffer II

	Amount	Final concentration
1 M NaH_2PO_4	2.5 ml	50 mM
5 M NaCl	10 ml	1 M
1 M Imidazole	12.5 ml	250 mM
Tween20	50 µl	0.01%
Distilled water	up to 50 ml	
Total volume	50 ml	

5.2.2 Buffers supplized with enzymes

10× ligation buffer

400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP

10× Klenow buffer

500 mM pH 8.0 Tris-HCl, 50 mM $MgCl_2$, 10 mM DTT

10× FastAP buffer

100 mM Tris-HCl, 50 mM $MgCl_2$, 1 M KCl, 0.2% Triton X-100, 10 mM 2-mercaptoethanol and 1 mg/ml BSA

10× T4 PNK buffer A

500 mM Tris-HCl, 100 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine and 1 mM EDTA.)

10× *Pfu* buffer

200 mM Tris-HCl (pH 8.8), 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1% (v/v) Triton X-100, 1 mg/ml BSA and 20 mM $MgSO_4$

5.3 Antibiotic

	Stock concentration (mg/ml)	Working concentration (μg/ml)
Ampicillin	50	50
Chloramphenicol	25	10
Kanamycin	50	50
Spectinomycin	125	125
Tetracycline	2	8

1

5.4 List of gene constructions

Construct name	Inserted Gene (bp)	Host vector	Cloning sites	Donation vector (primers)	Usage	Source
pSPUTK -GGM2	*GGM2* (633)	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for pull-down assay (prey), EMSA, and footprinting	(Winter *et al.* 2002b)
pSPUTK -GGM3	*GGM3* (744)	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for pull-down assay (prey), EMSA, and footprinting	Previous work of the lab
pSPUTK -GGM9	*GGM9* (762)	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for pull-down assay (prey) and EMSA	Previous work of the lab
pSPUTK -GGM11	*GGM11* (765)	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for pull-down assay (prey) and EMSA	Previous work of the lab
pSPUTK -GGM13	*GGM13* (714)	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for pull-down assay (prey) and EMSA	Previous work of the lab
pSPUTK -GGM15	*GGM15* (678)	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for pull-down assay (prey) and EMSA	Previous work of the lab
pSPUTK -GGM2ΔC	*GGM2* (489), C-terminal deleted	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for EMSA	Previous work of the lab
pSPUTK -GGM3ΔC	*GGM3* (624), C-terminal deleted	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for EMSA	Previous work of the lab
pSPUTK -GGM9ΔC	*GGM9* (612), C-terminal deleted	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for EMSA	Previous work of the lab
pSPUTK -GGM11ΔC	*GGM11* (600), C-terminal deleted	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for EMSA	Previous work of the lab
pSPUTK -GGM13ΔC	*GGM13* (561), C-terminal deleted	pSPUTK	*Nco*I-*BamH*I	-	*In vitro* translation for EMSA	Previous work of the lab
pIVEX1.3WG -GGM2	*GGM2* (633)	pIVEX1.3WG	*Nco*I-*Sma*I	pSPUTK-GGM2 (SP6/GGM2SmaIR)	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.3WG -GGM3	*GGM3* (744)	pIVEX1.3WG	*Nco*I-*Sma*I	pSPUTK-GGM3 (SP6/GGM3SmaIR)	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.3WG -GGM9	*GGM9* (762)	pIVEX1.3WG	*Nco*I-*Sma*I	pSPUTK-GGM9 (SP6/GGM9SmaIR)	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.3WG -GGM13	*GGM13* (714)	pIVEX1.3WG	*Nco*I-*Sma*I	pSPUTK-GGM13 (SP6/GGM13SmaIR)	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.3WG -GGM15	*GGM15* (678)	pIVEX1.3WG	*Nco*I-*Sma*I	pSPUTK-GGM15 (SP6/GGM15SmaIR)	*In vitro* translation for pull-down assay (bait)	This work

pIVEX1.4WG -GGM2	*GGM2* (633)	pIVEX1.4WG	*Nco*I-*Sma*I	pSPUTK-GGM2	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.4WG -GGM3	*GGM3* (744)	pIVEX1.4WG	*Nco*I-*Sma*I	pSPUTK-GGM3	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.4WG -GGM9	*GGM9* (762)	pIVEX1.4WG	*Nco*I-*Sma*I	pSPUTK-GGM9	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.4WG -GGM11	*GGM11* (765)	pIVEX1.4WG	*Nco*I-*Sma*I	pSPUTK-GGM11	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.4WG -GGM13	*GGM13* (714)	pIVEX1.4WG	*Nco*I-*Sma*I	pSPUTK-GGM13	*In vitro* translation for pull-down assay (bait)	This work
pIVEX1.4WG -GGM15	*GGM15* (678)	pIVEX1.4WG	*Nco*I-*Sma*I	pSPUTK-GGM15	*In vitro* translation for pull-down assay (bait)	This work
pTNT-GGM2	*GGM2* (633)	pTNT	*Eco*RI-*Sal*I	pTFT1-GGM2	*In vitro* translation for pull-down assay (co-translation)	This work
pTNT-GGM11	*GGM11* (765)	pTNT	*Eco*RI-*Sal*I	pTFT1-GGM11	*In vitro* translation for pull-down assay (co-translation)	This work
pTNT-GGM13	*GGM13* (714)	pTNT	*Eco*RI-*Sal*I	pGBKT7-GGM13	*In vitro* translation for pull-down assay (co-translation)	This work
pTNT-GGM15	*GGM15* (678)	pTNT	*Eco*RI-*Sal*I	pGBKT7-GGM15	*In vitro* translation for pull-down assay (co-translation)	This work
pENTR4 -GGM2	*GGM2* (633)	pENTR4	*Nco*I-*Eco*RV	pSPUTK-GGM2 (SP6/GGM2EcoRVR)	Entry clone for BiFC	This work
pENTR4 -GGM3	GGM3 (744)	pENTR4	*Nco*I-*Eco*RV	pSPUTK-GGM3 (SP6/GGM3EcoRVR)	Entry clone for BiFC	This work
pENTR4 -GGM2ΔC	GGM2 (489), C-terminal deleted	pENTR4	*Nco*I-*Eco*RV	pSPUTK-GGM2 (SP6/ GGM2_489EcoRVRev)	Entry clone for BiFC	This work
pENTR4 -GGM3ΔC	GGM3 (624), C-terminal deleted	pENTR4	*Nco*I-*Eco*RV	pSPUTK-GGM2 (SP6/ GGM3_624EcoRVRev)	Entry clone for BiFC	This work
pBatTL-N-YFP -GGM2	GGM2 (633)	pBatTL-N-YFP	-	pENTR4-GGM2	Expression clone for BiFC	This work
pBatTL-C-YFP -GGM2	GGM2 (633)	pBatTL-C-YFP	-	pENTR4-GGM2	Expression clone for BiFC	This work
pBatTL-N-YFP	GGM3 (744)	pBatTL-N-YFP	-	pENTR4-GGM3-	Expression clone for BiFC	This work

-GGM3						
pBatTL-C-YFP -GGM3	GGM3 (744)	pBatTL-C-YFP	-	pENTR4-GGM3-	Expression clone for BiFC	This work
pBatTL-N-YFP -SEP3K3C	SEP3, K3- and C-terminal deleted	pBatTL-N-YFP		-	Expression clone for BiFC	(Melzer *et al.* 2009)
pBatTL-C-YFP -SEP3K3C	SEP3, K3- and C-terminal deleted	pBatTL-C-YFP		-	Expression clone for BiFC	(Melzer *et al.* 2009)

5.5 List of primers for vector construction

Name	Sequence (5'→3')	Tm (°C)	Restriction sites	Notes
T7	TAATACGACTCACTATAGGG	56	-	Universe primer of T7 promoter
T3	AATTAACCCTCACTAAAGGG	56	-	Universe primer of T3 promoter
SP6	TATTTAGGTGACACTATAG	56	-	Universe primer of SP6 promoter
T7W	TAATACGACTCACTATAGGC	54	-	forward primers for pIVEX1.3/1.4WG based vectors
pIVEXRev	GAAAAACACTATGCGTTATCG	55	-	Reverse primer for amplifying insertions in pIVEX1.3/1.4WG vectors
pENTRfw	CTACAAACTCTTCCTGTTAGTTAG	56	-	Forward primer for amplifying insertions in pENTR4 based vector.
seql-b	CATCAGAGATTTTGAGACAC	54	-	Reverse primer for pENTR4 based vectors.
GGM2EcoRVR	TAATGATATCTACAAGCAATTCTGGCATACT	55-57	*EcoR*V	Reverse primer for amplifying full-length GGM2 with additional *EcoR*V restriction site.
GGM3EcoRVR	ATAATGATATCTGTAGGTCTGAAATTTGTGGG	57-58	*EcoR*V	Reverse primer for amplifying full-length GGM3 with additional *EcoR*V restriction site.
GGM2_489EcoRVRev	GTCTAGATATCTGCTCATCTGCTCG	58	*EcoR*V	Reverse primer for amplifying C-terminal deleted GGM2 with additional *EcoR*V restriction site.
GGM3_624EcoRVRev	GTCTAGATATCTTATCAGGTTTGCATGC	58	*EcoR*V	Reverse primer for amplifying C-terminal deleted GGM3 with additional *EcoR*V restriction site.
GGM2_489Rev	GCT CAT CTG CTC GTG CAA GG	56	-	Reverse primer for amplifying C-terminal deleted GGM2
GGM3_624Rev	TATCAGGTTTGCATGCATGAAGTTCC	56	-	Reverse primer for amplifying C-terminal deleted GGM3
GGM2SmaIR	ATAATCCCGGGACAAGCAATTCTGGCATACT	53	*Sma*I	Reverse primer replace the stop coden of GGM2 with coding coden and append *Sma*I restriction site.
GGM3SmaIR	ATAATCCCGGGATAGGTCTGAAATTTGTGGG	53	*Sma*I	Reverse primer replace the stop coden of GGM3 with coding coden and append *Sma*I restriction site.
GGM9SmaIR	ATAATCCCGGGACAAACCCACCAAATGTATTG	54	*Sma*I	Reverse primer replace the stop coden of GGM9 with coding coden and append *Sma*I restriction site.
GGM13SmaIR	ATAATCCCGGGACAGAGCTGAAGAACAGGT	54	*Sma*I	Reverse primer replace the stop coden of GGM13 with coding coden and append *Sma*I restriction site.
GGM15SmaIR	ATATACCCGGGATACAAAAACCTTAGCTGAAGA	53	*Sma*I	Reverse primer replace the stop coden of GGM15 with coding coden and append *Sma*I restriction site.

5.6 List of probes for EMSA and footprint assays

Probe Name	**Oligo sequence** (5'→3')	Description
S	AATTCGAAAT TTAATTATAT TCCAAATAAG GAAAGTATGG AACGTTG	Used in EMSA, contains a CArG-box (underlined) from the regulatory intron of *AGAMOUS.*
Sm	AATTCATAAA ACGGCAAGG AGAATTATATT TTTATGATGA ACATATG	Used in EMSA experiments, contains the same nucleotides as the oligonucleotide carrying the CArG-box, but in randomized order.
L1	TCGAGGTCGG AAATTTAATT ATATTCCAAA TAAGGAAAGT ATGGAACGTT CGACGGTATC GATAAGCTTG ATATATGTTC ATCATAAAAA TATAATTCTC CTTGCCGTTT TATATCGAAT TCCTGCAGCC CGGGGGATCC TCGTGGTCTA G	Used in EMSA, contains a CArG-box (underlined) from the regulatory intron of *AGAMOUS.*
L2	TCGAGGTCGG AAATTTAATT ATATTCCAAA TAAGGAAAGT ATGGAACGTT CGACGGTATC GATAAGCTTG ATGAAATTTA ATTATATTCC AAATAAGGAA AGTATGGAAC GTTATCGAAT TCCTGCAGCC CGGGGGATCC AGTAGTTCTA G	Used in EMSA and DNase I footprint assays, contains two CArG-box (underlined) from the regulatory intron of *AGAMOUS.*
L2i	TCGAGGTCGG AAATTTAATT ATATTCCAAA TAAGGAAAGT ATGGAACGTT CGACGGTATC GAT*ccatg*AA GCTTGATGAA A TTTAATTAT ATTCCAAATA AGGAAAGTAT GGAACGTTAT CGAATTCCTG CAGCCCGGGG GATCC AGTAG TTCTAG	Used in DNase I footprint assays, has the same nucleotide acids to probe B except with a 5-bp insertion (italic fonts in lower case) between two CArG-boxes (underlined)

6 References

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas de Pouplana, L., Martinez-Castilla, L. and Yanofsky, M.F. (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. USA.*, **97**, 5328-5333.

Ambrose, B.A., Lerner, D.R., Ciceri, P., Padilla, C.M., Yanofsky, M.F. and Schmidt, R.J. (2000) Molecular and genetic analyses of the *Silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Mol. Cell*, **5**, 569-579.

Angenent, G.C., Franken, J., Busscher, M., van Dijken, A., van Went, J.L., Dons, H. and van Tunen, A.J. (1995) A novel class of MADS box genes is involved in ovule development in petunia. *Plant Cell*, **7**, 1569-1582.

Angenent, G.C., Franken, J., Busscher, M., Weiss, D. and van Tunen, A.J. (1994) Co-suppression of the petunia homeotic gene *FBP2* affects the identity of the generative meristem. *Plant J.*, **5**, 33-44.

APG II (2003) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.*, **141**, 399-436.

Arber, E.A. and Parkin (1908) Studies on the evolution of the angiosperms: The relationship of the angiosperms to the Gnetales. *Ann. Bot.*, **os-22**, 489-515.

Arthur, W. (2002) The emerging conceptual framework of evolutionary developmental biology. *Nature*, **415**, 757-764.

Bateman, R.M., Hilton, J. and Rudall, P.J. (2006) Morphological and molecular phylogenetic context of the angiosperms: Contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *J. Exp. Bot.*, **57**, 3471-3503.

Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M.A., Saedler, H. and Theissen, G. (2002) A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol. Genet. Genomics*, **266**, 942-950.

Becker, A., Saedler, H. and Theissen, G. (2003) Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnemon*. *Dev. Genes Evol.*, **213**, 567-572.

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

Becker, A., Winter, K.-U., Meyer, B., Saedler, H. and Theissen, G. (2000) MADS-box gene diversity in seed plants 300 million years ago. *Mol. Biol. Evol.*, **17**, 1425-1434.

Bowe, L.M., Coat, G. and dePamphilis, C.W. (2000) Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Natl. Acad. Sci. USA.*, **97**, 4092-4097.

Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. (1991) Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development*, **112**, 1-20.

Bracha-Drori, K., Shichrur, K., Katz, A., Oliva, M., Angelovici, R., Yalovsky, S. and Ohad, N. (2004) Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. *Plant J.*, **40**, 419-427.

Carlsbecker, A., Tandre, K., Johanson, U., Englund, M. and Engstrom, P. (2004) The MADS-box gene *DAL1* is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *Plant J.*, **40**, 546-557.

Carpenter, R. and Coen, E.S. (1990) Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus*. *Genes Dev.*, **4**, 1483-1493.

Chaw, S.M., Parkinson, C.L., Cheng, Y., Vincent, T.M. and Palmer, J.D. (2000) Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc. Natl. Acad. Sci. USA.*, **97**, 4086-4091.

Chaw, S.M., Zharkikh, A., Sung, H.M., Lau, T.C. and Li, W.H. (1997) Molecular phylogeny of extant gymnosperms and seed plant evolution: Analysis of nuclear 18s rRNA sequences. *Mol. Biol. Evol.*, **14**, 56-68.

Cho, S., Jang, S., Chae, S., Chung, K., Moon, Y.-H., An, G. and Jang, S. (1999) Analysis of the C-terminal region of *Arabidopsis thaliana* APETALA1 as a transcription activation domain. *Plant Mol. Biol.*, **40**, 419-429.

Ciannamea, S., Kaufmann, K., Frau, M., Tonaco, I.A.N., Petersen, K., Nielsen, K.K., Angenent, G.C. and Immink, R.G.H. (2006) Protein interactions of MADS box transcription factors involved in flowering in *Lolium perenne*. *J. Exp. Bot.*, **57**, 3419-3431.

Coen, E.S. and Meyerowitz, E.M. (1991) The war of the whorls: Genetic interactions controlling flower development. *Nature*, **353**, 31-37.

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H., Angenent, G.C. and van Tunen, A.J. (1995) The petunia MADS box gene *FBP11* determines ovule identity. *Plant Cell*, **7**, 1859-1868.

Crepet, W.L. (1998) Botany: The abominable mystery. *Science*, **282**, 1653-1654.

Crepet, W.L. (2000) Progress in understanding angiosperm history, success, and relationships: Darwin's abominably "Perplexing phenomenon". *Proc. Natl. Acad. Sci. USA.*, **97**, 12939-12941.

Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H. and Sommer, H. (1996) Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J.*, **15**, 4330-4343.

Davies, B. and Schwarz-Sommer, Z. (1994) Control of floral organ identity by homeotic MADS-box transcription factors. *Results Probl. Cell Differ.*, **20**, 235-258.

De Bodt, S., Raes, J., Van de Peer, Y. and Theißen, G. (2003) And then there were many: MADS goes genomic. *Trends Plant Sci.*, **8**, 475-483.

de Folter, S., Immink, R.G.H., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B. and Angenent, G.C. (2005) Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. *Plant Cell*, **17**, 1424-1433.

de Folter, S., Shchennikova, A.V., Franken, J., Busscher, M., Baskar, R., Grossniklaus, U., Angenent, G.C. and Immink, R.G.H. (2006) A B_{sister} MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. *Plant J.*, **47**, 934-946.

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

Donoghue, M.J. and Doyle, J.A. (2000) Seed plant phylogeny: Demise of the anthophyte hypothesis? *Curr. Biol.*, **10**, R106-109.

Doyle, J. (2008) Integrating molecular phylogenetic and paleobotanical evidence on origin of the flower. *Int. J. Plant Sci.*, **169**, 816-843.

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

Endress, P. (2001) Origins of flower morphology. *J. Exp. Zool.*, **291**, 105-115.

Fan, H.-Y., Hu, Y., Tudor, M. and Ma, H. (1997) Specific interactions between the k domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant J.*, **12**, 999-1010.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M. and Colombo, L. (2003) MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell*, **15**, 2603-2611.

Ferrario, S., Immink, R.G. and Angenent, G.C. (2004) Conservation and diversity in flower land. *Curr. Opin. Plant Biol.*, **7**, 84-91.

Ferrario, S., Immink, R.G.H., Shchennikova, A., Busscher-Lange, J. and Angenent, G.C. (2003) The MADS box gene *FBP2* is required for *SEPALLATA* function in petunia. *Plant Cell*, **15**, 914-925.

Friedman, W.E. (2009) The meaning of Darwin's 'abominable mystery'. *Am. J. Bot.*, **96**, 5-21.

Friis, E.M., Doyle, J.A., Endress, P.K. and Leng, Q. (2003) Archaefructus - angiosperm precursor or specialized early angiosperm? *Trends Plant Sci.*, **8**, 369-373.

Frohlich, M.W. (1999) MADS about Gnetales. *Proc. Natl. Acad. Sci. USA.*, **96**, 8811-8813.

Frohlich, M.W. (2003) An evolutionary scenario for the origin of flowers. *Nat. Rev. Genet.*, **4**, 559-566.

Frohlich, M.W. and Parker, D.S. (2000) The mostly male theory of flower evolutionary origins: From genes to fossils. *Syst. Bot.*, **25**, 155-170.

Fukui, M., Futamura, N., Mukai, Y., Wang, Y., Nagao, A. and Shinohara, K. (2001) Ancestral MADS box genes in Sugi, *Cryptomeria japonica* d. Don (taxodiaceae), homologous to the B function genes in angiosperms. *Plant Cell Physiol.*, **42**, 566-575.

Futamura, N., Totoki, Y., Toyoda, A., Igasaki, T., Nanjo, T., Seki, M., Sakaki, Y., Mari, A., Shinozaki, K. and Shinohara, K. (2008) Characterization of expressed sequence tags from a full-length enriched cdna library of *Cryptomeria japonica* male strobili. *BMC Genomics*, **9**, 383.

Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Application to components of the *Escherichia coli* lactose operon regulatory system. *Nucl. Acids Res.*, **9**, 3047-3060.

Goto, K., Kyozuka, J. and Bowman, J.L. (2001) Turning floral organs into leaves, leaves into floral organs. *Curr. Opin. Genet. Dev.*, **11**, 449-456.

Goto, K. and Meyerowitz, E. (1994) Function and regulation of the Arabidopsis floral homeotic gene *PISTILLATA*. *Genes Dev.*, **8**, 1548-1560.

Hackbusch, J., Richter, K., Muller, J., Salamini, F. and Uhrig, J.F. (2005) A central role of *Arabidopsis thaliana* ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. *Proc. Natl. Acad. Sci. USA.*, **102**, 4908-4912.

Hochschild, A. and Ptashne, M. (1986) Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell*, **44**, 681-687.

Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, **409**, 525-529.

Hu, C.-D., Chinenov, Y. and Kerppola, T.K. (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell*, **9**, 789-798.

Huang, H., Mizukami, Y., Hu, Y. and Ma, H. (1993) Isolation and characterization of the binding sequences for the product of the *Arabidopsis* floral homeotic gene *AGAMOUS*. *Nucl. Acids Res.*, **21**, 4769-4776.

Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y. and Ma, H. (1996) DNA binding properties of two *Arabidopsis* MADS domain proteins: Binding consensus and dimer formation. *Plant Cell*, **8**, 81-94.

Huang, H., Tudor, M., Weiss, C.A., Hu, Y. and Ma, H. (1995) The *Arabidopsis* MADS-box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein. *Plant Mol. Biol.*, **28**, 549-567.

Immink, R., Tonaco, I., de Folter, S., Shchennikova, A., van Dijk, A., Busscher-Lange, J., Borst, J.-W. and Angenent, G. (2009) SEPALLATA3: The "Glue" for MADS box transcription factor complex formation. *Genome Biol.*, **10**, R24.

Immink, R.G.H., Ferrario, S., Busscher-Lange, J., Kooiker, M., Busscher, M. and Angenent, G.C. (2003) Analysis of the petunia MADS-box transcription factor family. *Mol. Genet. Genomics*, **268**, 598-606.

Immink, R.G.H., Gadella, T.W.J., Jr., Ferrario, S., Busscher, M. and Angenent, G.C. (2002) Analysis of MADS box protein-protein interactions in living plant cells. *Proc. Natl. Acad. Sci. USA.*, **99**, 2416-2421.

Jack, T., Brockman, L.L. and Meyerowitz, E.M. (1992) The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell*, **68**, 683-697.

Jacky, S., Heike, H. and Hendrik, G.S. (1993) Affinity purification of histidine-tagged proteins. *Mol. Biol. Rep.*, **18**, 223-230.

Jager, M., Hassanin, A., Manuel, M., Guyader, H.L. and Deutsch, J. (2003) MADS-box genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family. *Mol. Biol. Evol.*, **20**, 842-854.

Jensen, L. (2004) Quality analysis and integration of large-scale molecular data sets. *Drug Discovery Today: TARGETS*, **3**, 51-56.

Jofuku, K.D., Boer, B., Montagu, M.V. and Okamuro, J.K. (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell*, **6**, 1211-1225.

Jones, S. and Thornton, J.M. (1996) Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA.*, **93**, 13-20.

Kang, H. and An, G. (1997) Isolation and characterization of a rice MADS box gene belonging to the *AGL2* gene family. *Mol. Cells*, **7**, 45-51.

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

Kim, S., Koh, J., Yoo, M.-J., Kong, H., Hu, Y., Ma, H., Soltis, P.S. and Soltis, D.E. (2005) Expression of floral MADS-box genes in basal angiosperms: Implications for the evolution of floral regulators. *Plant J.*, **43**, 724-744.

Koncz, C. and Schell, J. (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of agrobacterium binary vector. *Mol. Gen. Genet.*, **204**, 383-396.

Kramer, E.M. and Irish, V.F. (2000) Evolution of the petal and stamen developmental programs: Evidence from comparative studies of the lower eudicots and basal angiosperms. *Int. J. Plant Sci.*, **161**, S29-S40.

Krämer, H., Niemöller, M., Amouyal, M., Revet, B., von Wilcken-Bergm, B. and Müller-Hill, B. (1987) *lac* repressor forms loops with linear DNA carrying two suitably spaced lac operators. *EMBO J.*, **6**.

Lane, D., Prentki, P. and Chandler, M. (1992) Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiol. Mol. Biol. Rev.*, **56**, 509-528.

Lenser, T., Theissen, G. and Dittrich, P. (2009) Developmental robustness by obligate interaction of class B floral homeotic genes and proteins. *PLoS Comput. Biol.*, **5**, e1000264.

Leseberg, C.H., Eissler, C.L., Wang, X., Johns, M.A., Duvall, M.R. and Mao, L. (2008) Interaction study of MADS-domain proteins in tomato. *J. Exp. Bot.*, **59**, 2253-2265.

Li, H., Liang, W., Jia, R., Yin, C., Zong, J., Kong, H. and Zhang, D. (2009) The *AGL6*-like gene *OsMADS6* regulates floral organ and meristem identities in rice. *Cell Res.*, **doi:10.1038/cr.2009.143**.

Litt, A. and Irish, V.F. (2003) Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: Implications for the evolution of floral development. *Genetics*, **165**, 821-833.

Ma, H. and dePamphilis, C. (2000) The ABCs of floral evolution. *Cell*, **101**, 5-8.

Magallon, S., Crane, P. and Herendeen, P. (1999) Phylogenetic pattern, diversity, and diversification of

eudicots. *Ann. Mo. Bot. Gard.*, **86**, 297-372.

Mandel, M.A., Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F. (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature*, **360**, 273-277.

Masiero, S., Li, M.-A., Will, I., Hartmann, U., Saedler, H., Huijser, P., Schwarz-Sommer, Z. and Sommer, H. (2004) *INCOMPOSITA*: a MADS-box gene controlling prophyll development and floral meristem identity in *Antirrhinum*. *Development*, **131**, 5981-5990.

Mathews, S. (2009) Phylogenetic relationships among seed plants: Persistent questions and the limits of molecular data. *Am. J. Bot.*, **96**, 228-236.

Melzer, R. (2005) Untersuchungen zur Interaktion von MADS-Domänen Proteinen aus der Gymnosperme *Gnetum gnemon* L. In *Department of Genetics*. Jena, Germany: Friedrich Schiller University Jena.

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

Melzer, R., Verelst, W. and Theissen, G. (2009) The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes *in vitro*. *Nucl. Acids Res.*, **37**, 144-157.

Melzer, R., Wang, Y.-Q. and Theißen, G. (2010) The naked and the dead: The ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin. Cell Dev. Biol.*, **21**, 118-128.

Meyerowitz, E.M., Bowman, J.L., Brockman, L.L., Drews, G.N., Jack, T., Sieburth, L.E. and Weigel, D. (1991) A genetic and molecular model for flower development in *Arabidopsis thaliana*. *Development*, **113**, 157-167.

Mouradov, A., Glassick, T.V., Hamdorf, B.A., Murphy, L.C., Marla, S.S., Yang, Y. and Teasdale, R.D. (1998) Family of MADS-box genes expressed early in male and female reproductive structures of monterey pine. *Plant Physiol.*, **117**, 55-62.

Mouradov, A., Hamdorf, B., Teasdale, R.D., Kim, J.T., Winter, K.-U. and Theißen, G. (1999) A *DEF/GLO*-like MADS-box gene from a gymnosperm: *Pinus radiata* contains an ortholog of angiosperm B class floral homeotic genes. *Dev. Genet.*, **25**, 245-252.

Münster, T., Pahnke, J., Di Rosa, A., Kim, J.T., Martin, W., Saedler, H. and Theissen, G. (1997) Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc. Natl. Acad. Sci. USA.*, **94**, 2415-2420.

Nagasawa, N., Miyoshi, M., Sano, Y., Satoh, H., Hirano, H., Sakai, H. and Nagato, Y. (2003) *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development*, **130**, 705-718.

Nam, J., dePamphilis, C.W., Ma, H. and Nei, M. (2003) Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.*, **20**, 1435-1447.

Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M. and Lepiniec, L. (2002) The *TRANSPARENT TESTA16* locus encodes the Arabidopsis B_{sister} MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell*, **14**, 2463-2479.

Ohmori, S., Kimizu, M., Sugita, M., Miyao, A., Hirochika, H., Uchida, E., Nagato, Y. and Yoshida, H. (2009) *MOSAIC FLORAL ORGANS1*, an *AGL6*-like MADS box gene, regulates floral organ identity and meristem fate in rice. *Plant Cell*, **21**, 3008-3025.

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

Pennisi, E. (2009) Origins: On the origin of flowering plants. *Science*, **324**, 28-31.

Pinyopich, A., Ditta, G.S., Savidge, B., Liljegren, S.J., Baumann, E., Wisman, E. and Yanofsky, M.F. (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*, **424**, 85-88.

Pnueli, L., Hareven, D., Broday, L., Hurwitz, C. and Lifschitz, E. (1994) The *TM5* MADS box gene mediates organ differentiation in the three inner whorls of tomato flowers. *Plant Cell*, **6**, 175-186.

Pollock, R. and Treisman, R. (1990) A sensitive method for the determination of protein-DNA binding specificities. *Nucl. Acids Res.*, **18**, 6197-6204.

Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, F., Soltis, D.E., Soltis, P.S., Zanis, M., Zimmer, E.A., Chen, Z., Savolainen, V. and Chase, M.W. (1999) The earliest angiosperms: Evidence from mitochondrial, plastid and nuclear genomes. *Nature*, **402**, 404-407.

Renner, S. and Ricklefs, R. (1995) Dioecy and its correlates in the flowering plants. *Am. J. Bot.*, **82**, 596.

Riechmann, J. and Meyerowitz, E. (1997) Determination of floral organ identity by Arabidopsis MADS domain homeotic proteins AP1, AP3, PI, and AG is independent of their DNA- binding specificity. *Mol. Biol. Cell*, **8**, 1243-1259.

Riechmann, J.L., Krizek, B.A. and Meyerowitz, E.M. (1996a) Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl. Acad. Sci. USA.*, **93**, 4793-4798.

Riechmann, J.L., Wang, M. and Meyerowitz, E.M. (1996b) DNA-binding properties of *Arabidopsis*

MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res.*, **24**, 3134-3141.

Rijpkema, A.S., Zethof, J., Gerats, T. and Vandenbussche, M. (2009) The petunia *AGL6* gene has a *SEPALLATA*-like function in floral patterning. *Plant J.*, **60**, 1-9.

Rothwell, G.W., Crepet, W.L. and Stockey, R.A. (2009) Is the anthophyte hypothesis alive and well? New evidence from the reproductive structures of Bennettitales. *Am. J. Bot.*, **96**, 296-322.

Rutledge, R., Regan, S., Nicolas, O., Fobert, P., Cote, C., Bosnich, W., Kauffeldt, C., Sunohara, G., Seguin, A. and Stewart, D. (1998) Characterization of an *AGAMOUS* homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis*. *Plant J.*, **15**, 625-634.

Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P.J., Hansen, R., Tetens, F., Lonnig, W.E., Saedler, H. and Sommer, H. (1992) Characterization of the *Antirrhinum* floral homeotic MADS-box gene *DEFICIENS*: Evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.*, **11**, 251-263.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. and Sommer, H. (1990) Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science*, **250**, 931-936.

Shindo, S., Ito, M., Ueda, K., Kato, M. and Hasebe, M. (1999) Characterization of MADS genes in the gymnosperm *Gnetum parvifolium* and its implication on the evolution of reproductive organs in seed plants. *Evol. Dev.*, **1**, 180-190.

Shiraishi, H., Okada, K. and Shimura, Y. (1993) Nucleotide sequences recognized by the AGAMOUS MADS domain of *Arabidopsis thaliana in vitro*. *Plant J.*, **4**, 385-398.

Soltis, P.S., Brockington, S.F., Yoo, M.-J., Piedrahita, A., Latvis, M., Moore, M.J., Chanderbali, A.S. and Soltis, D.E. (2009) Floral variation and floral genetics in basal angiosperms. *Am. J. Bot.*, **96**, 110-128.

Sommer, H., Beltran, J., Huijser, P., Pape, H., Lonnig, W., Saedler, H. and Schwarz-Sommer, Z. (1990) *DEFICIENS*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: The protein shows homology to transcription factors. *EMBO J.*, **9**, 605-613.

Specht, C. and Bartlett, M. (2009) Flower evolution: The origin and subsequent diversification of the angiosperm flower. *Annu. Rev. Eco.l Evol. Syst.*, **40**, 217-243.

Sun, G., Ji, Q., Dilcher, D.L., Zheng, S., Nixon, K.C. and Wang, X. (2002) Archaefructaceae, a new basal angiosperm family. *Science*, **296**, 899-904.

Sundström, J., Carlsbecker, A., Svensson, M.E., Svenson, M., Johanson, U., Theißen, G. and Engström, P. (1999) MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Dev. Genet.*, **25**, 253-266.

Sundström, J. and Engström, P. (2002) Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *Plant J.*, **31**, 161-169.

Tandre, K., Albert, V.A., Sundas, A. and Engstrom, P. (1995) Conifer homologues to genes that control floral development in angiosperms. *Plant Mol. Biol.*, **27**, 69-78.

Tandre, K., Svenson, M., Svensson, M.E. and Engstrom, P. (1998) Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and angiosperms. *Plant J.*, **15**, 615-623.

Tang, W. and Perry, S.E. (2003) Binding site selection for the plant MADS domain protein AGL15: An *in vitro* and *in vivo* study. *J. Biol. Chem.*, **278**, 28154-28159.

Theißen, G. (2006) The proper place of hopeful monsters in evolutionary biology. *Theory Biosci*, **124**, 349-369.

Theissen, G. (2002) Orthology: Secret life of genes. *Nature*, **415**, 741.

Theissen, G. (2009) Saltational evolution: Hopeful monsters are here to stay. *Theory Biosci.*, **128**, 43-51.

Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Münster, T., Winter, K.-U. and Saedler, H. (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.*, **42**, 115-149.

Theissen, G. and Melzer, R. (2007) Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann. Bot.*, **100**, 603-619.

Theissen, G. and Saedler, H. (1999) The golden decade of molecular floral development (1990-1999): A cheerful obituary. *Dev. Genet.*, **25**, 181-193.

Theissen, G. and Saedler, H. (2001) Plant biology: Floral quartets. *Nature*, **409**, 469-471.

Thompson, B., Bartling, L., Whipple, C., Hall, D., Sakai, H., Schmidt, R. and Hake, S. (2009) *BEARDED-EAR* encodes a MADS box transcription factor critical for maize floral development. *Plant Cell*, **21**, 2578-2590.

Tröbner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lönnig, W., Saedler, H., Sommer, H. and Schwarz-Sommer, Z. (1992) *GLOBOSA*: A homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. *EMBO J.*, **11**, 4693-4704.

Tsai, W.-C., Pan, Z.-J., Hsiao, Y.-Y., Jeng, M.-F., Wu, T.-F., Chen, W.-H. and Chen, H.-H. (2008) Interactions of B-class complex proteins involved in tepal development in *Phalaenopsis* orchid. *Plant Cell Physiol.*, **49**, 814-824.

Tsong, A.E., Tuch, B.B., Li, H. and Johnson, A.D. (2006) Evolution of alternative transcriptional circuits with identical logic. *Nature*, **443**, 415-420.

Tzeng, T.Y., Liu, H.C. and Yang, C.H. (2004) The C-terminal sequence of *LMADS1* is essential for the formation of homodimers for B function proteins. *J. Biol. Chem.*, **279**, 10747-10755.

Tzfira, T., Vaidya, M. and Citovsky, V. (2004) Involvement of targeted proteolysis in plant genetic transformation by agrobacterium. *Nature*, **431**, 87-92.

Verelst, W., Saedler, H. and Munster, T. (2007) MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. *Plant Physiol.*, **143**, 447-460.

Voinnet, O., Rivas, S., Mestre, P. and Baulcombe, D. (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of *Tomato bushy stunt virus*. *Plant J.*, **33**, 949-956.

Walter, M., Chaban, C., Schutze, K., Batistic, O., Weckermann, K., Nake, C., Blazevic, D., Grefen, C., Schumacher, K., Oecking, C., Harter, K. and Kudla, J. (2004) Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant Journal*, **40**, 428-438.

Weigel, D. and Meyerowitz, E.M. (1994) The ABCs of floral homeotic genes. *Cell*, **78**, 203-209.

West, A.G., Causier, B.E., Davies, B. and Sharrocks, A.D. (1998) DNA binding and dimerisation determinants of *Antirrhinum majus* mads-box transcription factors. *Nucleic Acids Res.*, **26**, 5277-5287.

Whipple, C.J., Zanis, M.J., Kellogg, E.A. and Schmidt, R.J. (2007) Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals. *Proc. Natl. Acad. Sci. USA.*, **104**, 1081-1086.

Winter, K.-U., Becker, A., Munster, T., Kim, J.T., Saedler, H. and Theissen, G. (1999) MADS-box genes reveal that Gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. USA.*, **96**, 7342-7347.

Winter, K.-U., Saedler, H. and Theissen, G. (2002a) On the origin of class B floral homeotic genes: Functional substitution and dominant inhibition in *Arabidopsis* by expression of an orthologue from the gymnosperm *Gnetum*. *Plant J.*, **31**, 457-475.

Winter, K.-U., Weiser, C., Kaufmann, K., Bohne, A., Kirchner, C., Kanno, A., Saedler, H. and Theißen, G. (2002b) Evolution of class B floral homeotic proteins: Obligate heterodimerization originated from homodimerization. *Mol. Biol. Evol.*, **19**, 587-596.

Wynne, J. and Treisman, R. (1992) SRF and MCM1 have related but distinct DNA binding specificities. *Nucl. Acids Res.*, **20**, 3297-3303.

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

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

Yang, Y., Xiang, H. and Jack, T. (2003b) *PISTILLATA-5*, an *Arabidopsis* B class mutant with strong defects in petal but not in stamen development. *Plant J.*, **33**, 177-188.

Yanofsky, M., Ma, H., Bowman, J., Drews, G., Feldmann, K. and Meyerowitz, E. (1990) The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. *Nature*, **346**, 35-39.

Zahn, L.M., Feng, B. and Ma, H. (2006) Beyond the ABC-model: Regulation of floral homeotic genes. In *Adv. Bot. Res.* (Douglas E. Soltis, J.H.L.M., Pamela S. Soltis, and J. A. Callow ed: Academic Press, pp. 163-207).

Zahn, L.M., Kong, H., Leebens-Mack, J.H., Kim, S., Soltis, P.S., Landherr, L.L., Soltis, D.E., dePamphilis, C.W. and Ma, H. (2005a) The evolution of the *SEPALLATA* subfamily of MADS-box genes: A preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics*, **169**, 2209-2223.

Zahn, L.M., Leebens-Mack, J., dePamphilis, C.W., Ma, H. and Theißen, G. (2005b) To B or not to B a flower: The role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. *J. Hered.*, **96**, 225-240.

Zhang, P., Tan, H.T.W., Pwee, K.-H. and Kumar, P.P. (2004) Conservation of class c function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant J.*, **37**, 566-577.

Zhanhua, C., Gan, J.G., Lei, L., Sakharkar, M.K. and Kangueane, P. (2005) Protein subunit interfaces: Heterodimers versus homodimers. *Bioinformation*, **1**, 28-39.

Acknowledgements

Prof. Günter Theißen, thank you so much being my supervisor. A Chinese saying can express my feeling: “a talk with you saves me a ten-year reading”. Thanks a lot for support me finishing my PhD work in your lab.

Thank you, Prof. Johannes Wöstemeyer, as my co-supervisor. Several discussions with you greatly broaden my view on sciences. Your encouragement gives me confidence on pursuing my favor research topic.

Many thanks are given to Dr. Pia Nutt, who gives me invaluable help when I arrived here. Your patient and enthusiastic help on life, lab rules, usage on chemicals and equipments enable me to carry out bench-work quickly. Great thanks to Dr. Rainer Melzer. Thank you, Rainer, for discussing with me about the project and lots of interesting topics, for sharing experience of experiments and writing, for reviewing and comments on my dissertation, and for helps on life. Many thanks to Dr. Christiane M. Ritz, Dr. Maren Fräger and Dr. Markus Ritz for help before and after the born of my son. Without your help, I might not know how to deal with new life. Lots of thanks to Dr. Mariana Mondragon-Palomino, who shares her experiences on sciences, corrects my English, and many other things. Thanks are given to Ms Lydia Gramzow, Andrea Härter, Teresa Lenser, Susanne Schilling, Janine Ziermann, Lisa Weilandt, Dr. Dajana Lobbes, Dr. Hannelore Simon, Ms Ulrike Wrazidlo, Heidi Kressler, Mr. Christian Gafert, Mr. Christoph Thieme, Ms Conny Bartholmes, Ms Luisa Hiese, Ms Nina Kottenhagen, Dr. Susanne Nolden, who help me on solving experiment problems, seeing doctor, looking for house, reviewing and correcting my dissertations, teaching me German, and many many daily helps.

Thank you for great help, Ms. Sabine Schein, Dr. Dorit Schmidt, and Dr. Petra Franz. Your in-time helps save me lots time and safe me from troubles.

Without any of you, things can be very difficult, and I cannot believe I would survive in many cases where German is the only “official” language.

Thanks are given to Ms Lidan Ye, Dr. Andreas Licht (Jena University), Dr. Robert Winkler (HKI), Dr. Fredd Vergara-Montalvo and Mr. Dahai Yang (Max Planck Institute for Chemical Ecology) who help me with experiments and share plant materials.

I would also give my great thanks to my wife Guilan Yang and my son YangYang, who accompany me here, give me lots of happy time while suffering a lot.

Last but not least, thank you all the people I met in this country who ever help me.

All the best wishes to you.

YongQiang
14-02-2010

Scientific publications & Conference contributions

Publications:

1. **Wang, Y.-Q**., Melzer, R, and Theißen, G. Interaction patterns of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon* provide a clue to the origin of 'floral quartets'. *Plant Journal* (accepted)

2. Melzer, R., **Wang, Y.-Q**. and Theißen, G. (2010) The naked and the dead: The ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin. Cell Dev. Biol.*, **21**, 118-128.

3. ***Wang, Y.-Q**., Tian, H.-Y., Du, X.-Q., Lu, S.-H., Lu, W.-L., Chong, K. and Meng, Z. (2007) Isolation and characterization of a putative class E gene from Taihangia rupestris. J. Integr. Plant Biol., 49, 343-350.*

4. Sun, X.-L., **Wang, Y.-Q**., Wang, B., Gu, S.-J., Wang, Z.-F. and Cao, Q.-W. (2004) Inheritance of the chlorophyll content in immature fruit color of cucumber. *Acta Anaesthesiol. Sin.*, **31**, 327-331.

5. Sun, X.-L., **Wang, Y.-Q**., Wang, Z.-F., Wang, K.-A., Ai, X.-Z., Gao, J.-F., Cao, Q.-W., Zhang, Z.-X. and He, Q.-W. (2002) Studies on adjustment of temperature suitable to cucumber in solar greenhouse in winter. *Shandong Agricultural Sciences*, 10-14.

Verbal presentations:

6. **Wang Y.-Q**. (2009) Evolutionary origin of 'Floral Quartets' as revealed by molecular interactions. JSMC/ILRS Symposium

7. **Wang Y.-Q.** (2008) Interaction of floral homeotic proteins: an evolutionary view. ILRS Symposium

8. **Wang Y.-Q. (2007)** Reconstructing the interaction network of reproductive homeotic proteins in a gymnosperm. ILRS Symposium

CURRICULUM VITAE

Personal data

Name:	Yongqiang Wang
Date of birth:	26.10.1976
Place of birth:	Shandong, China
Marriage status:	Married, one child

Education

2006- present	PhD student; Department of Genetics, Friedrich Schiller University Jena/International Leibniz Research School for Microbial and Biomolecular Interactions (ILRS Jena) (Supervisor Prof. Dr. Günter Theißen and Prof. Dr. Johannes Wöstemeyer) Title of the dissertation: "The evolutionary origin of 'floral quartets': clues from molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon*"
2001-2005	Master in Cell Biology; Graduate University of the Chinese Academy of Sciences (CAS)/ Institute of Botany, CAS
1995-1999	Bachelor in Agronomy; Department of Horticulture, Shandong Agricultural University

Professional and Scientific Career

2006-present	Scientific co-worker; Department of Genetics, Friedrich Schiller University Jena
2005-2006	Assistant Engineer; Center for Signal Transduction and Metabolomics, Institute of Botany, CAS
1999-2001	Research associate; Institute of Vegetables, Shandong Academy of Agricultural Sciences

SELBSTSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, YongQiang Wang, geboren am 26.10.1976, das mir die für die Biologisch-Pharmazeutische Fakultät geltende Promotionsordnung bekannt ist. Ich habe die vorliegende Dissertation selbstständig angefertigt und auβer den angegebenen keine Hilfsmittel, persönliche Mitteilungen oder Quellen eingesetzt. Die Hilfe eines Promotionsberaters wurde nicht in Anspruch genommen. Es wurden von mir keine geldwerten Leistungen erbracht, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Diese Dissertation wurde nur dem Fakultätsrat der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena und keiner anderen Hochschule zur wissenschaftlichen Prüfung oder zur Dissertation eingereicht. Diese Arbeit ist weder identisch noch teilidentisch mit einer Arbeit, welche der Friedrich-Schiller-Universität Jena oder einer anderen Hochschule zur Dissertation vorgelegt worden ist.

Jena,_____ YongQiang Wang