The Plant Journal (2004) **40**, 22–34 doi: 10.1111/j.1365-313X.2004.02183.x

Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins

Ilona M. Zimmermann[†], Marc A. Heim, Bernd Weisshaar[‡,*] and Joachim F. Uhrig
Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, D-50829 Koeln, Germany

Received 30 April 2004; revised 18 June 2004; accepted 25 June 2004.
[*]For correspondence (fax +49 0521 106 6423; e-mail bernd.weisshaar@uni-bielefeld.de).
[†]Present address: Botanical Institute III, University of Koeln, Gyrhofstrasse 15, D-50931 Koeln, Germany.
[‡]Present address: Chair of Genome Research, Department of Biology, Bielefeld University, D-33594 Bielefeld, Germany.

Summary

In-depth analysis of protein–protein interaction specificities of the MYB protein family of *Arabidopsis thaliana* revealed a conserved amino acid signature ([DE]Lx$_2$[RK]x$_3$Lx$_6$Lx$_3$R) as the structural basis for interaction between MYB and R/B-like BHLH proteins. The motif has successfully been used to predict new MYB/BHLH interactions for *A. thaliana* proteins, it allows to discriminate between even closely related MYB proteins and it is conserved amongst higher plants. In *A. thaliana*, the motif is shared by fourteen R2R3 MYB proteins and six 1R MYB proteins. It is located on helices 1 and 2 of the R3 repeat and forms a characteristic surface-exposed pattern of hydrophobic and charged residues. Single-site mutation of any amino acid of the signature impairs the interaction. Two particular amino acids have been determined to account for most of the interaction stability. Functional specificity of MYB/BHLH complexes was investigated *in vivo* by a transient *DFR* promoter activation assay. Residues stabilizing the MYB/BHLH interaction were shown to be critical for promoter activation. By virtue of proved and predicted interaction specificities, this study provides a comprehensive survey of the MYB proteins that interact with R/B-like BHLH proteins potentially involved in the TTG1-dependent regulatory interaction network. The results are discussed with respect to multi-functionality, specificity and redundancy of MYB and BHLH protein function.

Keywords: regulatory network, protein–protein interactions, amino acid motif, TTG1, yeast two-hybrid assay.

Introduction

Gene-specific regulation of transcription is of fundamental importance for virtually every aspect of cellular functions. Specificity is provided by the action of transcription factors, modular proteins typically composed of a DNA binding domain and effector domains responsible for activator or repressor activity. In eukaryotes, gene expression frequently is mediated by multi-protein complexes. The formation of these complexes involves the combinatorial action of transcription factors that bind conserved promoter elements in precise spatial orientation and on the basis of both specific protein–DNA and protein–protein interactions. This type of transcriptional regulation, termed combinatorial control, is thought to facilitate the complex regulatory networks found in higher eukaryotes (Wolberger, 1999).

According to the type of their DNA binding domain, transcription factors can be assigned to different families (Pabo and Sauer, 1992). The availability of complete genome sequences has facilitated the comprehensive descriptions of entire transcription factor families with the aim to identify evolutionary relationships, to find common structural features and ultimately to derive functional predictions. In *Arabidopsis thaliana* more than 5% of the genes have been predicted to code for transcription factors (Riechmann *et al.*, 2000). One of the largest families is constituted by proteins characterized by the MYB domain, which consists of up to three imperfect repeats referred to as R1, R2 and R3, each forming a helix helix-turn-helix structure of about 53 amino acids (Frampton *et al.*, 1991). Furthermore, MYB repeats typically contain regularly spaced tryptophan residues, which build a central tryptophan cluster in the three-dimensional helix-turn-helix fold (Kanei-Ishii *et al.*, 1990).

MYB proteins are common to all eukaryotes. However, in higher plants this protein family is extraordinarily amplified (Rabinowicz *et al.*, 1999; Romero *et al.*, 1998; Stracke *et al.*,

 © 2004 Blackwell Publishing Ltd

2001). Particularly, the presence of 126 R2R3-type *MYB* genes in the *A. thaliana* genome indicates plant-specific evolutionary events and led to the idea that R2R3 MYB proteins might predominantly be involved in plant-specific regulatory processes (Jin and Martin, 1999; Martin and Paz-Ares, 1997; Stracke *et al.*, 2001). Systematic large-scale projects for functional analyses of MYB proteins have been initiated and first results are accumulating (Meissner *et al.*, 1999). However, the precise functions of most of the plant MYB transcription factors are still unknown.

Plant MYB proteins are known to be involved in a variety of cellular processes including the regulation of biosynthetic pathways like phenylpropanoid or tryptophan biosynthesis, control of cell fate determination and regulation of the cell cycle. Furthermore, functions as diverse as the more structural role of telomere binding MYB proteins, the involvement in circadian clock-regulated gene expression and a regulatory role in the phosphate starvation response have been described (Meissner *et al.*, 1999; Petroni *et al.*, 2002; Rubio *et al.*, 2001; Stracke *et al.*, 2001).

Both genetic and direct physical interactions suggest an intimate functional relationship between MYB proteins and BHLH proteins. The cooperative action of MYB and BHLH proteins has been most extensively studied with respect to the phenylpropanoid biosynthetic pathways (Winkel-Shirley, 2001). In *Zea mays*, the transcriptional activation of anthocyanin biosynthesis genes by the R2R3 MYB proteins ZmC1 and ZmPl requires the involvement of BHLH proteins from the *R/B* gene family. A direct interaction between the MYB domain of ZmC1 and the N-terminal domain of the BHLH protein ZmB has been described (Goff *et al.*, 1992). The specificity of the interaction was revealed by comparing ZmC1 with the closely related protein ZmP, a MYB protein which in contrast to ZmC1 controls a set of flavonoid biosynthesis genes independent of BHLH transcription factors (Grotewold *et al.*, 1994). The interaction domain itself was analysed in greater detail in maize by constructing chimeric MYB domains of ZmP and ZmC1 showing that four differing amino acids in the otherwise identical MYB R3 repeat are the basis for specificity (Grotewold *et al.*, 2000). The regulation of the phenylpropanoid biosynthetic pathways by MYB proteins in combination with BHLH proteins seems to be conserved throughout the plant kingdom, as exemplified by the MYB proteins PhAN2 and the BHLH proteins PhJAF13 and PhAN1 from *Petunia hybrida*, or strawberry FaMYB1 that is able to interact with the maize BHLH protein ZmR (Aharoni *et al.*, 2001; Quattrocchio *et al.*, 1999). Molecular analysis of the *transparent testa* mutants of *A. thaliana* revealed that several steps of the flavonoid biosynthetic pathway in *A. thaliana* are also controlled by the combinatorial action of MYB and BHLH proteins. Both TT8/AtBHLH042 and TT2/AtMYB123 are involved in the control of the expression of the *DFR* and *BAN* genes in the developing *A. thaliana* seeds (Nesi *et al.*, 2000, 2001). Furthermore, overexpression of PAP1/AtMYB75 and PAP2/AtMYB90, respectively, in *A. thaliana* as well as heterologous overexpression of the BHLH protein EGL3/AtBHLH002 (also named MYC-146) or of GL3/AtBHLH001 in petals of a white-flowered mutant of *Matthiola incana* resulted in activation of anthocyanin biosynthesis (Borevitz *et al.*, 2000; Ramsay *et al.*, 2003). R/B-like BHLH proteins of *A. thaliana* include GL3 as well as EGL3 and TT8 and cluster together in subgroup III of the *BHLH* gene family that contains 162 *BHLH* genes (Bailey *et al.*, 2003; Heim *et al.*, 2003). MYB proteins most similar to ZmC1 cluster in subgroups 5, 6, 7 and 15, respectively, of the R2R3 MYB family (Stracke *et al.*, 2001).

Recent results suggest that in *A. thaliana* certain MYB proteins and R/B-like BHLH proteins work together with the WD40 protein transparent testa glabra1 (TTG1) in a regulatory network which underlies not only the control of phenylpropanoid biosynthesis, but also the regulation of epidermal cell differentiation and cell patterning in root hair and trichome development (Johnson *et al.*, 2002; Schiefelbein, 2003; Zhang *et al.*, 2003). A combinatorial model for TTG1-dependent regulation proposes an activator/repressor system that is based on the competition for binding sites. Direct physical interactions have been shown for GL1/AtMYB0 and the one-repeat MYB protein AtMYBL2, respectively, which both interact with GL3 (Payne *et al.*, 2000; Sawa, 2002). Furthermore, in heterologous binding studies it has been shown that GL1, CPC and WER have the ability to interact with ZmR (Lee and Schiefelbein, 1999; Wada *et al.*, 1997). Recently, EGL3 has been identified as an additional BHLH component of the TTG1-dependent regulatory network and yeast two-hybrid studies showed that it has the capacity to interact with the MYB proteins GL1, PAP1 and PAP2 as well as CPC and TRY (Zhang *et al.*, 2003). An emerging view is that within the TTG1 regulatory network in *A. thaliana*, BHLH proteins affect overlapping subsets of the network, whereas the MYB proteins are the key components providing the specificity for the downstream effects (Zhang *et al.*, 2003). TTG1-related WD40 proteins are conserved throughout the plant kingdom both in terms of sequence and regarding their general role in the anthocyanin pathway. However, with respect to the function of TTG1-related proteins in other traits and the regulatory involvement of MYB and BHLH proteins, recent results from the monocot *Z. mays* indicate a substantial evolutionary divergence of regulatory mechanisms (Carey *et al.*, 2004; Mol *et al.*, 1998).

An important step to understand the complex TTG1-dependent regulatory processes in *A. thaliana* at a genomic level is the comprehensive identification of the components of the MYB and BHLH protein interaction network. In many cases, class boundaries between proteins with different interaction specificities are not easily predictable on the basis of sequence comparisons. Therefore, detailed knowledge of the epitopes involved and the precise identification

of functional amino acid motifs are required. Using the potential to interact with TTG1 as a criterion to define the R/B-like BHLH proteins from *A. thaliana*, we have systematically analysed the interaction specificities of AtMYB proteins with R/B-like BHLH proteins. We have identified and characterized a structural motif underlying the MYB/BHLH interactions, allowing sequence-based prediction of MYB proteins that interact with R/B-like BHLH factors.

Results

A protein interaction matrix of R/B-like BHLH-factors and C1-like MYB transcription factors from A. thaliana

In order to systematically investigate interaction specificities of C1-like *A. thaliana* MYB proteins, we analysed the MYB-proteins from subgroups 5, 6, 15 and 7 and included AtMYB5 and AtMYB11 that cluster in the same area of the phylogenetic tree. These MYB factors were assayed in a yeast two-hybrid matrix against three R/B-like AtBHLH proteins of subgroup IIIf, which were shown to interact with TTG1, and AtBHLH013 from subgroup IIId (Figure 1). BHLH proteins were fused to the GAL4 DNA binding domain (BD) and assayed for their ability to bind the MYB proteins fused to the GAL4 activation domain (AD). Yeast strains of different mating types were transformed with the respective yeast two-hybrid BD- or AD- construct. The combinations were generated by mating the appropriate yeast cells in a systematic fashion. Interaction was studied using two different reporters: yeast growth on triple-dropout media and α-galactosidase activity indicating activation of the MEL1 reporter gene. Yeast growth in the presence of at least 3 mM 3-AT and concomitantly α-galactosidase activity of more than two standard deviations above the background was considered a positive result. EGL3, AtBHLH012 and TT8 consistently activated both reporters when cotransformed with AtMYB5, with the MYB proteins of subgroup 5 (PAP1, PAP2, AtMYB113 and AtMYB114) and with TT2 (subgroup 6), respectively. In contrast, MYB proteins from subgroup 15 (GL1, WER, AtMYB23) interacted with EGL3 and AtBHLH012 but not with TT8 (Figure 1a,b).

None of the tested MYB proteins interacted with AtBHLH013, a member of the BHLH subgroup IIId, nor did any of the AtBHLH factors interact with AtMYB11, AtMYB12 and AtMYB111, members of subgroup 7 of the MYB transcription factors (Figure 1). Similarly, yeast two-hybrid tests were negative when combinations of the abovementioned MYB proteins and BHLH proteins from subgroup IIIe (AtBHLH004, AtBHLH005 and AtBHLH006) were tested.

Because of their inherent transcription activating function, it was not possible to use MYB constructs as binding domain fusions in the yeast two-hybrid system. Because the GL3 binding domain fusion was also auto-activating in yeast, the

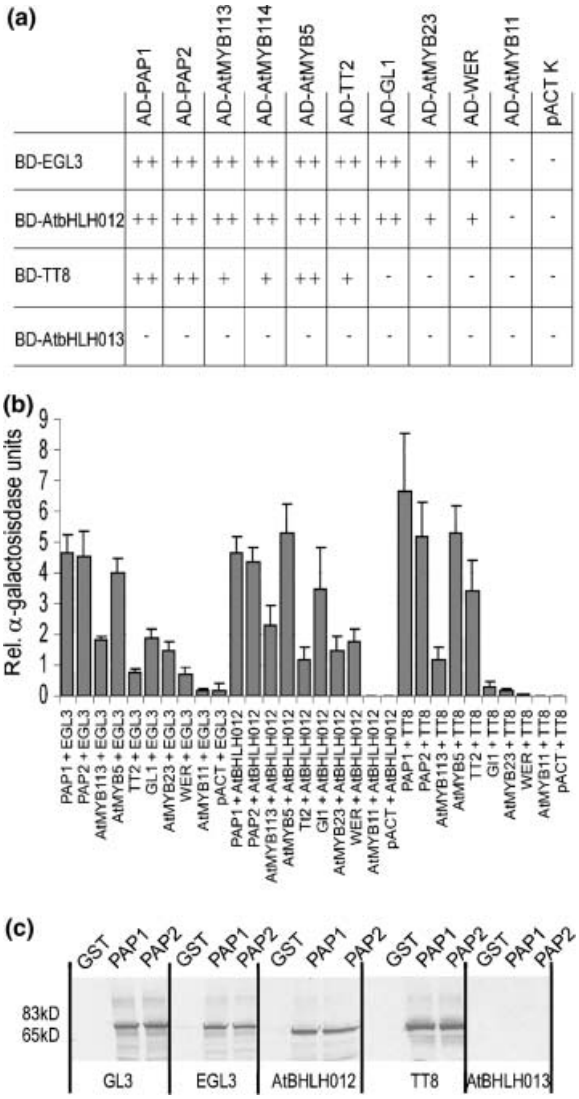

Figure 1. (a) Yeast two-hybrid analysis of MYB-BHLH interactions.
BHLH proteins were fused to the GAL4 binding domain (BD) and assayed for their ability to bind TTG1 and the MYB proteins fused to the GAL4 activation domain (AD). The interaction strength indicated in the matrix was estimated by comparing the growth of yeast cells on triple dropout media supplemented with 3, 20 and 40 mM 3-aminotriazole (3-AT), respectively. pACT K was used in order to check whether the BHLH proteins contain a functional activation domain in yeast. The BHLH proteins of subgroup IIId and IIIe did not interact with TTG1 (exemplarily the test with AtBHLH013 is shown). AtMYB12 and AtMYB111, MYB proteins of subgroup 7 were tested as well but show results similar to AtMYB11 (no interaction with R/B-like BHLH proteins). ++, growth at 20 mM 3-AT; +, growth at 3 mM 3-AT; -, no growth on triple dropout media containing 3 mM 3-AT.
(b) Semiquantitative assay of interaction strength. α-galactosidase assays of yeast strains expressing combinations of BHLH and MYB proteins fused to the GAL4 binding and activation domain, respectively. Results shown represent the mean values of 12 independent α-galactosidase-assays.
(c) *In vitro* pull down assay. Binding of PAP1/AtMYB75 and PAP2/AtMYB90 to GL3/AtBHLH001, EGL3/AtBHLH002, AtBHLH012, TT8/AtBHLH042 and AtBHLH013. MYB proteins were expressed as GST fusion proteins. Fusion proteins were re-immobilized on GSH agarose beads and incubated with [^{35}S]methionine-labelled BHLH proteins synthesized in a coupled transcription-translation system. As a negative control beads were loaded with GST. After extensive washing, bound proteins were analysed by SDS-PAGE. [^{35}S]-labelled proteins were detected by autoradiography.

respective combinations were excluded from the matrix. However, exemplarily the activation domain of PAP1 was mapped to the 58 C-terminal residues of the protein. Interaction of a PAP1 version deleted for this region was shown to interact with all four BHLH proteins from subgroup IIIf including GL3 (data not shown).

Selected interactions were confirmed by *in vitro* pull-down assays. GST-tagged PAP1 and PAP2, respectively, was purified from *Escherichia coli* and bound to GSH agarose beads. As a negative control beads were loaded with GST. The loaded beads were incubated with *in vitro* translated and radioactively labelled GL3, EGL3, TT8, AtBHLH012 and AtBHLH013, respectively. Both, PAP1 and PAP2, specifically interacted *in vitro* with all four R/B-like BHLH proteins from subgroup IIIf. No binding of the related AtBHLH013 from subgroup IIIe was detected, confirming the results of the yeast two-hybrid assays (Figure 1c).

Combinations of R/B-like BHLH proteins with PAP1, PAP2 and TT2 activate the AtDFR *promoter*

To examine whether interactions found in yeast and *in vitro* can be correlated with promoter activation *in vivo*, combinations of MYB proteins from subgroups 5, 6 and 15 with R/B-like BHLH proteins from subgroup IIIf were assayed in a transient protoplast co-transfection system. Constructs were tested in *At7* protoplasts for the activation capacity of their encoded proteins on a co-transfected 520nt fragment of the *AtDFR* promoter fused to the *uidA* ORF encoding β-glucuronidase (GUS, Figure 2). None of the AtMYB or BHLH constructs exhibited a significant activation of the *AtDFR* promoter on its own. In combination with BHLH proteins, PAP1 and PAP2 strongly activated reporter gene expression. Strongest GUS activity was observed in combinations with GL3, EGL3 and TT8, respectively, co-expression with AtBHLH012 resulted in significantly lower reporter activation (approximately 30% compared with the PAP1 + EGL3 combination). TT2 was able to activate expression from the DFR promoter to a much lower degree. Co-expression with GL3, EGL3 and TT8, respectively, resulted in a GUS activity of only 10% compared with the combination PAP1 + EGL3. No reporter activation was observed in co-expression experiments of TT2 with AtBHLH012. MYB proteins of subgroup 15 were found to be unable to activate expression from the *AtDFR* promoter.

R/B-like BHLH proteins interact with 1R MYB proteins

To identify further proteins that interact with R/B-like AtBHLH factors, EGL3, TT8 and AtBHLH012 were used as bait proteins to screen three different activation domain-fused *A. thaliana* cDNA libraries in the yeast two-hybrid system. Five to 10 million zygotes were analysed in each screening resulting in more than 40 different candidates. These candidates included TTG1 and nine different MYB proteins. In addition to five R2R3 MYB proteins that were already present in the interaction matrix mentioned above, four 1R MYB proteins (AtMYBL2, CPC, At1g01380 and At2g30420) were identified in screenings using AtBHLH012 as bait protein. Furthermore, AtMYBL2 and the CPC-homologue At1g01380 were isolated in screenings using EGL3 as bait. Additionally, CPC and AtMYBL2 were detected with TT8 as bait.

The interaction is based on the R3 repeat of the MYB domain

To identify the regions required for the interaction with the R/B-like BHLH proteins, N- and C-terminal deletion constructs of PAP1 were analysed. Protein fragments were fused to the GAL4-AD and tested for interaction with the set of AtBHLH transcription factors. The minimal domain necessary for interaction was confined to amino acids 53 to 114 comprising accurately the complete R3-repeat of the MYB domain of PAP1. These findings were supported by the fact that, from the library screenings, several N-terminally truncated MYB proteins have been isolated. The shortest 1R MYB proteins identified were CPC consisting of amino acids 37–94 and the CPC-homologue At1g01380 comprising amino acids 33–83, respectively (Figure 3).

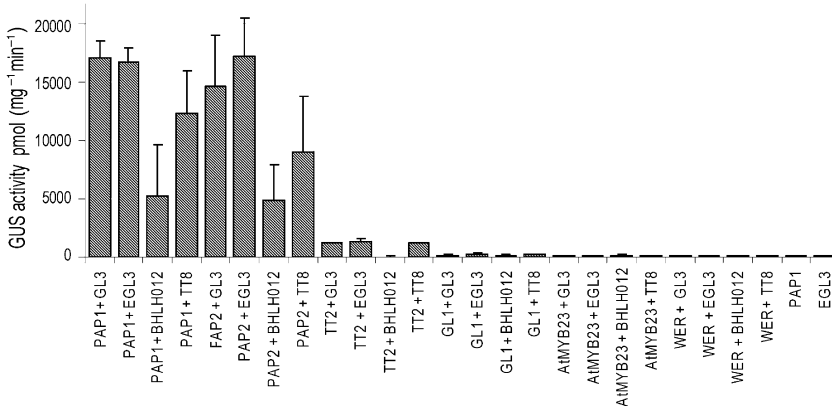

Figure 2. Transient promoter activation assay in *At7* protoplasts using an *AtDFR* promoter and combinations of MYB and BHLH proteins as indicated. Results shown represent the mean values of 10 independent assays. Neither of the constructs activated the *AtDFR* promoter when transfected alone.

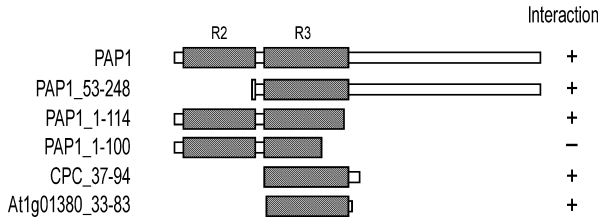

Figure 3. Deletions of MYB proteins fused to the GAL4 activation domain (AD) were assayed for interaction in yeast with EGL3/AtBHLH002. Yeast growth is indicated by (+), no growth is indicated by (−). Grey boxes represent the R2 and R3 MYB repeats.

A random mutagenesis approach was utilized to select PAP1 mutants with temperature-sensitive defects in the ability to interact with R/B-like AtBHLH proteins. The reason for analysing temperature-sensitive mutants was that the maintenance of the interaction activity at permissive temperatures prevents the occurrence of premature stop codons and ensures that each mutant has the potential to encode a viable protein. A total of 60 temperature-sensitive mutants were identified on the basis of reduced growth rate at non-permissive conditions. Sequence analysis revealed an accumulation of amino acid exchanges in the first and second helices of the R3 repeat, affecting 15 different amino acids (highlighted in Figure 4), emphasizing the importance of the N-terminal part of the R3 repeat for interaction with R/B-like AtBHLH proteins.

MYB proteins interacting with R/B-like BHLH factors share the amino acid motif [DE]Lx$_2$[RK]x$_3$Lx$_6$Lx$_3$R

The sample of BHLH-interacting MYB proteins includes members of different clusters that partly have been allocated to phylogenetically distant groups (e.g. the R2R3 MYB protein GL1 and the 1R MYBs AtMYBL2 and CPC). This suggests that interaction is based on conserved amino acid patterns distinguishing this set of proteins from the rest of the MYB family. Using an alignment of the R3 domains (Figure 4), comparison of the relative information content at individual positions of the group of interacting MYB proteins with that of an alignment of the whole MYB family revealed significant differences at six positions within the MYB R3 repeat (Figure 5). High positive values indicate conserved amino acids at positions that are variable in the over-all alignment. These are the positions [DE]12, L13, [RK]16, L20, L29 and R33.

The 3D structure of the MYB domains of the interacting MYB proteins was modelled according to the known structural data of c-MYB. Figure 6 shows the MYB domain of PAP1 as an example. Consistently, the amino acids of the conserved motif are surface-exposed, forming a characteristic pattern of hydrophobic and charged residues. Two hydrophobic residues and a positively charged residue

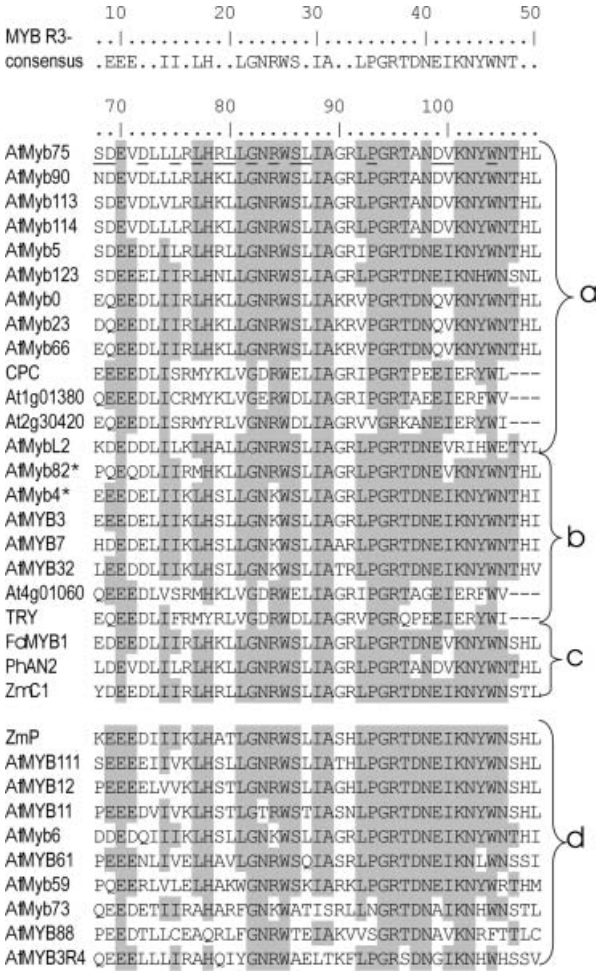

Figure 4. Sequence comparison of the R3 domains of MYB proteins tested for interaction with R/B-like BHLH proteins.
Amino acids that are more than 50% conserved in R3 consensus sequence (Stracke *et al.* (2001) are highlighted in grey. Numbers on top refer to the consensus sequence of the R3 repeat of plant MYB proteins described in Stracke *et al.* (2001), gaps in the consensus sequence (positions 25 and 27), resulting from the alignment of the R2 and R3 repeats were omitted in this figure. Underlined letters in the sequence of PAP1/AtMYB75 show the positions of amino acid exchanges by undirected mutagenesis, which disturb the interactions. The positions D12, L13, R16, R19, L20, L29 and R33 correspond to D72, L73, R76, R79, L80, L87 and R91 in the PAP1/AtMYB75 sequence.
(a) MYB proteins shown in this work to interact with BHLH transcription factors.
(b) MYB proteins predicted to interact with BHLH transcription factors based on the described motif; *the predicted interaction was verified in yeast two-hybrid experiments.
(c) MYB proteins from other plant species shown to interact with R/B like BHLH transcription factors. Fa, *Fragaria ananasa*; Ph, *Petunia hybrida*; Zm, *Zea mays*.
(d) MYB proteins shown to not interact with the tested BHLH transcription factors.

(positions 17, 21 and 24 in the MYB consensus, corresponding to L77, L81 and R84 in the PAP1 sequence) that are to a high degree conserved in the whole MYB family are surrounded by this pattern. Although these amino acids

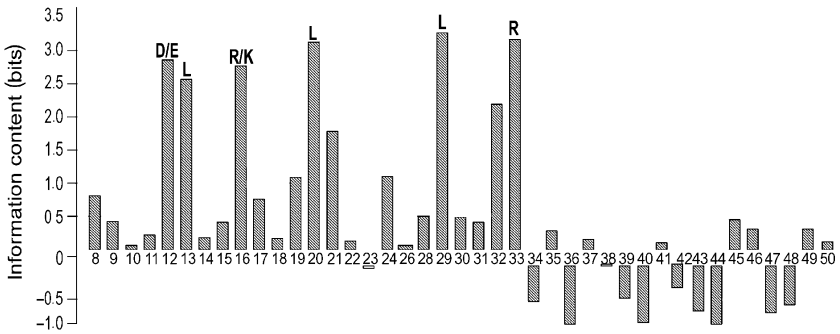

Figure 5. Difference in information content at individual positions of the R3 alignment of the MYB proteins interacting with R/B-like BHLH proteins versus relative entropy in an alignment of the whole MYB family. Numbers refer to the consensus sequence of the R3 repeat of plant MYB proteins as described in Stracke *et al.* (2001) and in Figure 4.

are not characteristic for the BHLH-interacting MYB proteins, and are therefore not included in the motif, they might have a function in stabilizing the interaction. Significantly, four independent temperature-sensitive PAP1 mutants have been isolated that affected two of these positions (L17, R24, highlighted in Figure 4; to allow easy comparison among different MYB proteins, we refer to amino acid positions in PAP1 with numbers from the consensus as outlined in Figure 4).

Two amino acids account for most of the interaction strength

Using site-directed mutagenesis, amino acids of the predicted interaction motif of the PAP1 sequence have been exchanged for amino acids present in related MYB proteins unable to interact with R/B-like AtBHLH proteins. In addition to mutations D12N, L20V and L29I obtained by the random-mutagenesis approach, the single-site mutants L13I, R16E, R16K, R19S, L20A, R33D and R33A were analysed. With the exception of R19S, all of these exchanges led to temperature-sensitive interaction with EGL3, revealed by reduced growth of yeast cells at non-permissive temperatures. However, exchanges at the positions 20 and 33 had the most severe effect, completely abolishing yeast growth. Quantification of the interaction strength by liquid α-galactosidase assays confirmed this result, showing that exchanges at positions 13 and 16 had no or only a minor impact, whereas mutation of either L20 or R33 strongly decreased the interaction strength (Figure 7a).

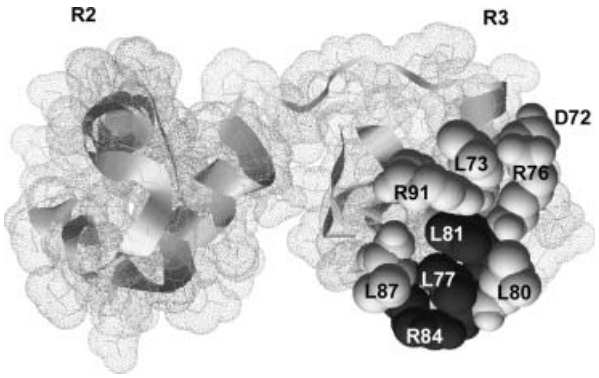

Figure 6. Model of the PAP1/AtMYB75 R2R3-MYB domain on the basis of the NMR structure of the MYB domain of c-MYB (Ogata *et al.*, 1994).
Amino acids which are elements of the motif [DE]Lx_2[RK]$x_3Lx_6Lx_3R$ are highlighted (light grey) and named according to their position in the PAP1/AtMYB75 sequence. The positions 12, 13, 16, 17, 20, 21, 24, 29 and 33 correspond to the amino acids D72, L73, R76, L77, L80, L81, R84, L87 and R91 of PAP1, respectively. Amino acids conserved throughout *Arabidopsis thaliana* MYB proteins are coloured in dark grey. Amino acids involved in interaction are solvent-exposed, forming a characteristic pattern positioned on the surface opposite to the DNA binding site.

Activation of the AtDFR *promoter by PAP1 is impaired by mutations in the interaction motif*

To examine the effects of mutations in the interaction motif under *in vivo* conditions, single-site mutants of PAP1 in combination with R/B-like BHLH proteins were assayed for their capacity to activate the *AtDFR* promoter. The wild-type protein exhibited strongest transcriptional activation together with GL3 and EGL3, respectively (Figures 2 and 7b). Analyses of the mutated PAP1 variants with EGL3 revealed a crucial role of the consensus sequence positions 16, 20 and 33. The mutations L20A, R33D and R33A, shown to severely affect interaction strength, reduced the activation capacity by about 80% in comparison with the wild-type protein. Interestingly, at position 16, the non-conservative change R16E, shown to have only a minor effect on the interaction, correlates with a strong decrease of GUS activity (50%) while the conservative change at the same from arginine to lysine only results in a little lower GUS activity (Figure 7b).

The amino acid motif [DE]Lx_2[RK]$x_3Lx_6Lx_3R$ can be used to predict interaction of MYB proteins with R/B-like BHLH factors

The amino acid motif [DE]Lx_2[RK]$x_3Lx_6Lx_3R$ was used to search the *A. thaliana* genomic sequence to identify additional potential BHLH-interacting MYB proteins. In addition to the already described MYB proteins interacting with R/B-like BHLH factors, seven additional MYB protein sequences were found to fit exactly with the motif: two 1R MYB proteins (TRY and At4g01060) and five R2R3 MYB proteins (AtMYB3,

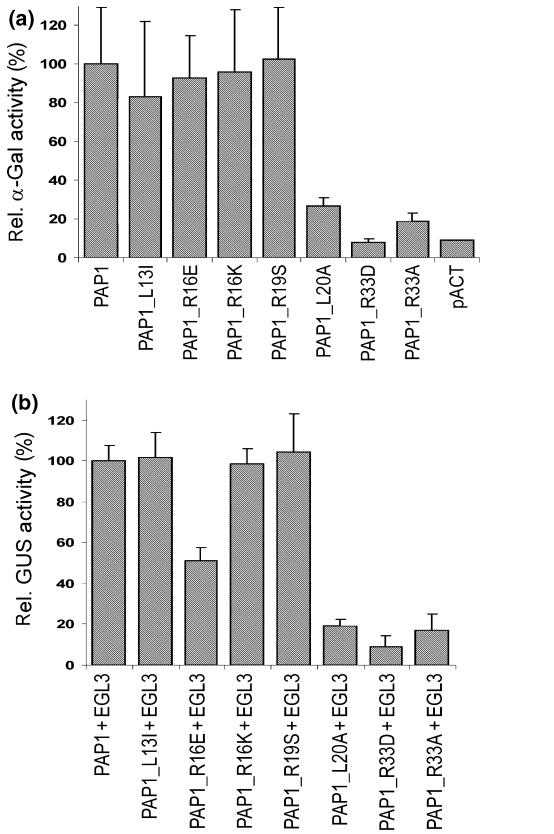

Figure 7. Single site mutations affect interaction strength and potential to activate the *AtDFR* promoter. Quantification of interaction strength of PAP1/AtMYB75 single site mutants with EGL3/AtBHLH002. The labelling of the mutants refers to the MYB consensus sequence. L73, R76, R79, L80, and R91 in PAP1/AtMYB75 correspond to L13, R16, R19, L20 and R33, respectively.
(a) α-Galactosidase assays of mutagenized PAP1/AtMYB75 fused to the GAL4 activation domain (AD) and EGL3/AtBHLH002 as bait protein. Bars indicate mean values of the relative α-galactosidase activity [%] referred to the wildtype PAP1/AtMYB75 interaction with EGL3/AtBHLH002. The empty vector pACT2 was used as a negative control.
(b) Transient expression experiments in At7 protoplasts using an *AtDFR* promoter and various mutated PAP1/AtMYB75 proteins together with EGL3/AtBHLH002 as effectors. Bars indicate mean values of 10 independent experiments. Relative GUS activity [%] refers to the wildtype PAP1/AtMYB75 cotransfected with EGL3/AtBHLH002. The effectors transfected together with the *AtDFR* promoter plasmid are indicated below the diagram.

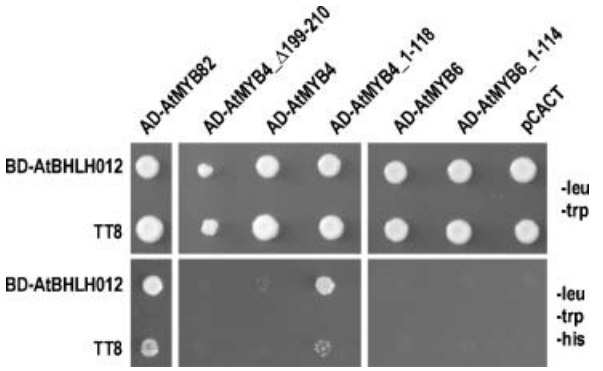

Figure 8. Experimental confirmation of interactions predicted from the *Arabidopsis thaliana* genomic sequence.
The BHLH proteins AtBHLH012 and TT8/AtBHLH042 were fused to the GAL4 binding domain (BD) and assayed for their ability to bind the MYB proteins AtMYB4, AtMYB6 and AtMYB82, as well as C-terminal deletions of AtMYB4 and AtMYB6 fused to the GAL4 activation domain (AD). Interaction was shown by yeast growth on triple dropout media supplemented with 5 mM 3-AT.

AtMYB4, AtMYB7, AtMYB32 and AtMYB82, see Figure 4b). For TRY, interaction with GL3 has recently been demonstrated (Zhang *et al.*, 2003), the CPC-homologue At4g01060 has not been characterized so far. AtMYB82 has not been assigned to any subgroup of the phylogenetic tree of the MYB family, due to the lack of any detectable conserved amino acid motif in the C-terminal region (Stracke *et al.*, 2001). The four other R2R3 MYB proteins predicted to interact with R/B-like BHLH factors have been assigned to subgroup 4, characterized by a C-terminal amino acid motif predicted to be involved in transcriptional repression (Aharoni *et al.*, 2001; Jin *et al.*, 2000; Stracke *et al.*, 2001). Interestingly, two additional MYB proteins belonging to the same cluster as AtMYB4 (AtMYB6 and AtMYB8) do not contain the exact interaction motif and therefore are predicted to be unable to interact with R/B-like BHLH proteins. Exemplarily, we tested the interaction properties of AtMYB82, AtMYB4 and AtMYB6 with the yeast two-hybrid system and could show that the predictions are valid. As indicated by yeast growth on triple-dropout media supplemented with 5 mM 3-AT, AtMYB82 interacts with both TT8 and AtBHLH012 (Figure 8). Interaction of a full-length construct of AtMYB4 protein with TT8 and BHLH012 was detectable but very weak and suppressed already in the presence of low concentrations of 3-AT (Figure 8). To test whether reporter activation in yeast is influenced by the presence of a proposed repressor motif in the C-terminal part of AtMYB4 (Jin *et al.*, 2000), a construct comprising only the R2R3 domain and a second construct lacking specifically the repressor motif were tested. Deletion of the entire C-terminal region (AtMYB4_1-118) resulted in strong interaction with BHLH012 and weaker but significant interaction with TT8. Deletion of only the repressor motif did not result in detectable reporter activation in yeast (Figure 8). Despite the fact that AtMYB4 and AtMYB6 share more than 90% identical amino acids in the MYB domain, full-length AtMYB6 and a deletion construct of AtMYB6 similar to AtMYB4_1-118 did not interact with either EGL3 or AtBHLH012 in yeast. These results demonstrate that the proposed interaction motif allows the prediction of interactions specificities of MYB factors from sequence alone.

Discussion

In *A. thaliana* MYB transcription factors and R/B-like BHLH proteins are connected in a hierarchical regulatory network

controlling several steps of the phenylpropanoid biosynthetic pathway as well as epidermal patterning in trichome and root hair development. On top of the hierarchy, the highly pleiotropic locus, TTG1, controls the whole network by physical interaction with different R/B-like BHLH proteins, each regulating diverse but partially overlapping subsets of the pathways. The factors providing the specificity for individual biosynthetic and developmental processes are activating or repressing MYB proteins, respectively. In root hair and trichome development, direct protein–protein interactions and competition for binding sites are thought to provide a mechanistic basis of this activator/repressor system (Marks and Esch, 2003; Schellmann *et al.*, 2002; Schiefelbein, 2003; Zhang *et al.*, 2003).

As a first step to more comprehensively understand the TTG1-dependent regulatory network, we aimed at a complete picture of the MYB proteins involved. Proteins as diverse as for example the 1R MYB protein CPC and the R2R3 MYB protein GL1/AtMYB0 have been shown to similarly interact with the BHLH protein EGL3/AtBHLH002, indicating a structural and functional property maintained across large phylogenetic distances (Zhang *et al.*, 2003). Furthermore, heterologous interactions of the *A. thaliana* MYB proteins WER/AtMYB66 and CPC and the strawberry FaMYB1, respectively, with the maize BHLH protein R (Aharoni *et al.*, 2001; Lee and Schiefelbein, 1999; Wada *et al.*, 1997) indicate that the interaction domain is functionally conserved across plant species. The potential to interact with R/B-like BHLH proteins is therefore proposed to be based on conserved structural elements distinguishing the MYB proteins that do interact with R/B-like BHLH proteins from those that do not. To precisely identify the amino acids characteristic for the MYB proteins that interact with R/B-like BHLH proteins, we combined bioinformatic analyses with a systematic experimental interaction grid approach.

With the help of yeast two-hybrid arrays, library screenings and *in vitro* pull-down assays, we identified 13 different MYB proteins (nine R2R3 MYB proteins and four 1R MYB proteins) able to specifically interact with BHLH proteins from subgroup IIIf. The subgroups IIId, IIIe and IIIf are characterized by a plant-specific N-terminal stretch of amino acids. Subgroup IIIf contains the functionally characterized R/B-like proteins GL3/AtBHLH001, EGL3/AtBHLH002, TT8/AtBHLH042 and the as yet uncharacterized AtBHLH012 (Heim *et al.*, 2003). We have shown that EGL3/AtBHLH002, TT8/AtBHLH042 and AtBHLH012, members of subgroup IIIf, are able to interact with TTG1, whereas none of the closely related BHLH proteins from subgroups IIId and IIIe interacted with TTG1 or with one of the selected MYB proteins. Therefore, we propose that subgroup IIIf constitutes the complete set of R/B-like BHLH proteins in *A. thaliana* and we consider an interaction of MYB proteins with members of this subgroup as an indication for a potential functional involvement in TTG1-dependent regulatory processes.

In addition to interactions described previously, 33 new interactions between *A. thaliana* MYB and BHLH proteins were detected. With the exception of the MYB proteins of subgroup 15 (GL1, WER and AtMYB23, Stracke *et al.*, 2001) which did not interact with TT8, all R/B-BHLH interacting MYB proteins identified in this study were similarly able to interact with the BHLH proteins EGL3, TT8 and AtBHLH012.

Despite similar interaction properties, functional specificity of MYB proteins from subgroups 6 and 15, respectively, in combination with the four R/B-like BHLH proteins was revealed by their differential ability to activate transcription from the *DFR* promoter in an *in vivo* reporter assay. PAP1/AtMYB75 and PAP2/AtMYB90 strongly activated expression from the *AtDFR* promoter when co-transfected with any of the four R/B-like BHLH proteins. Reporter gene expression was even stronger than the combination TT2/AtMYB123 with TT8, two proteins shown to be required for the normal expression of the *DFR* gene during seed formation (Nesi *et al.*, 2000, 2001). Interestingly, the strongest activation was observed with GL3 and EGL3, respectively, supporting the view that both proteins might have redundant functions not only in regulating trichome development, but also in their ability to directly regulate expression of genes of the anthocyanin biosynthesis. These data are in agreement with the recent finding that double mutants of *gl3* and *egl3* have a discernible anthocyanin-reduced phenotype (Zhang *et al.*, 2003). None of the MYB proteins from subgroup 15 were able to induce transcription from the *AtDFR* promoter in any combination with the four R/B-like BHLH proteins. These results indicate a specific synergistic interplay between MYB proteins and BHLH proteins that does not rely on either protein interaction or promoter binding specificities alone; each component seems to be necessary but not sufficient for activation of transcription.

Given the large number of possible mutual interactions between MYB proteins and R/B-like BHLH proteins revealed in this study, the question of functional specificity or redundancy within the TTG1 regulatory network is getting increasingly complex. The phenotypes of knock-out and over-expressing mutants indicate that both – different spatio-temporal expression patterns and different functional properties of the respective proteins – have to be considered. In a few cases, redundant protein function has been proved by complementation of knock-out mutants. GL1 expressed from the WER promoter, for example, can substitute for WER function and complements the mutant phenotype (Lee and Schiefelbein, 2001). Another example is the finding that constitutive expression of single R/B-like BHLH proteins like ZmR, GL3 or EGL3 can completely or at least in parts complement the pleiotropic phenotypes caused by the TTG1 knock-out, that is supposed to affect the function of all four R/B-like BHLH proteins (Galway *et al.*, 1994; Lloyd *et al.*, 1992; Payne *et al.*, 2000; Zhang *et al.*, 2003). However, the interaction characteristics and the

promoter activation data presented in this study indicate that there is specificity on the level of protein properties as well.

A detailed analysis of deletion mutants confined the R/B-like BHLH interacting domain to the R3 repeat within the PAP1 MYB domain (PAP1 amino acids 53–114) and to the amino acids 33–83 within the single MYB repeat of At1g01380, respectively. These results were supported by the fact that random mutations of PAP1 resulting in temperature-labile interactions clustered in the same region of the protein (Figure 2a). Our data further narrow down the recent findings that interaction with EGL3 is mediated by the MYB domain (amino acids 1–113 of PAP1 and 2, amino acids 1–124 of GL1), and the previous mapping of the interaction domain of CPC necessary for heterologous interaction with the maize BHLH protein ZmR (Wada *et al.*, 2002; Zhang *et al.*, 2003). An analogous region has been described to mediate the specificity of the interaction between the maize MYB protein ZmC1 and the BHLH protein ZmR (Grotewold *et al.*, 2000).

In order to deduce functionally important amino acid positions, the set of sequences of the 13 different MYB proteins interacting with R/B-like BHLH proteins was used to determine a deviance of amino acid frequencies compared with the mean frequencies in an alignment of the whole *A. thaliana* MYB family. Six positions highly variable in the whole MYB family revealed a high degree of conservation within the group of MYB proteins shown to interact with R/B-like BHLH factors. Site-directed mutagenesis of single residues provided evidence that each amino acid included in the motif $[DE]Lx_2[RK]x_3Lx_6Lx_3R$ (positions 12–33 in the MYB R3 consensus) contributes to the interaction strength. However, leucine (L) at position 20 and arginine (R) at position 33 account for most of the interaction stability. *In vivo* reporter assays confirmed the yeast two-hybrid data and supported the essential role of L20, R33 and a positively charged residue at position 16.

Modelling of the sequences of the MYB proteins that interact with R/B-like BHLH proteins on the basis of the NMR structure of the MYB domain of c-MYB (Ogata *et al.*, 1994) indicated that the interaction motif is solvent-exposed, forming a characteristic pattern positioned on the surface opposite to the DNA binding site. These amino acids surround two amino acid positions that are predominantly hydrophobic in all *A. thaliana* MYB proteins, potentially contributing non-specifically to interaction strength. In fact, an exchange of the highly conserved hydrophobic residue L17 to histidine, gave rise to temperature-sensitive interactions in yeast (Figure 2).

Recently, four amino acid residues of the maize MYB protein ZmC1 have been identified, specifying the interaction with the BHLH protein ZmR (Grotewold *et al.*, 2000). Substitution of these residues was shown to be sufficient to transfer the ability to interact with ZmR to the closely related, but functionally distinct ZmP, with every single residue having an essential role in stabilizing the interaction. Substitution of two additional residues made ZmP activity partially dependent on ZmR in maize cells (Grotewold *et al.*, 2000). According to the MYB consensus sequence the four residues responsible for interaction specificity correspond to positions 13, 16, 19 and 20 in the R3 repeat. Our data concerning leucine residues at positions 13 and 20 are in agreement with these results. However, in *A. thaliana*, the impact of these two residues on interaction strength differs considerably. Whereas an exchange at position 13 in PAP1 (L73I) resulted in a temperature-sensitive interaction without significantly affecting activity in the transient promoter activation assay, interaction and promoter activation capacity was almost completely abolished when position 20 was mutated (PAP1 L80A).

In maize, substitutions at positions 16 (arginine to lysine) and 19 (arginine to alanine) impair interaction with ZmP (Grotewold *et al.*, 2000). Our results show that in *A. thaliana* exchanges at the corresponding positions have only a minor impact on interaction strength and on promoter activation in the *in vivo* assay. In fact, a lysine residue at position 16 can be present in both, interacting (AtMYBL2, AtMYB4) and non-interacting MYB proteins. Furthermore, a MYB protein potentially involved in anthocyanin biosynthesis in *Perilla frutescens* (MYB-P1), containing a lysine residue at the corresponding position in the R3 domain, has been shown to interact with an R/B-like BHLH protein (Gong *et al.*, 1999). However, a positively charged amino acid in this position might be important for the function of the MYB/BHLH complex because the exchange of arginine to glutamate significantly reduces the activity in the *in vivo* promoter activation assay. Similarly, our data indicate that in *A. thaliana* the exchange of arginine to alanine in position 19 does not significantly influence the interaction. Position 19 is rather variable in the group of MYB proteins that interact with R/B-like BHLH proteins, including an alanine residue in this position in AtMYBL2. Furthermore, exchange of this particular arginine to serine in PAP1 (PAP1_R79S, Figure 7) does not influence interaction strength and *in vivo* activity, respectively. Our study revealed the involvement of three additional amino acid residues in constituting the specific interaction surface: an acidic residue (aspartate or glutamate) in position 12, leucine 29 and arginine 33. In maize, the latter amino acid position was described to be involved in the transcriptional activation function, rather than interaction of ZmC1 with ZmR (Grotewold *et al.*, 2000). Our findings indicate that in *A. thaliana* this residue is of particular relevance for the interaction surface, as exchanges to alanine or aspartate almost completely abolish both, interaction in yeast and activity in the *in vivo* assay. This position is highly variable in the alignment of the whole *A. thaliana* MYB family. The fact that it is completely conserved within the MYB proteins that interact with

R/B-like BHLH proteins not only from *A. thaliana* but also from other species, including MYB proteins shown to act in repression of transcription rather than activation (AtMYB4, *Fa*MYB1, see below), supports the view that a major function of this residue might be to stabilize interaction with BHLH proteins.

The proposed interaction motif is consistently present in MYB proteins from other plant species described to directly interact with R/B-like BHLH proteins (Figure 4c) and it is sufficiently complex to provide the possibility to identify additional MYB proteins that should be able to interact with R/B-like BHLH factors. In the *A. thaliana* genomic sequence, seven additional AtMYB protein sequences fit exactly with the motif, including the 1R MYB proteins TRY (shown previously to interact with EGL3, Zhang *et al.*, 2003) and the CPC homologue At4g01060, the R2R3 MYB proteins AtMYB3, AtMYB4, AtMYB7, AtMYB32 from subclade 4 and the distantly related AtMYB82, which lacks any conserved sequence pattern in the C-terminus and therefore was not assigned to any subgroup of the *Arabidopsis* MYB family (Stracke *et al.*, 2001). Exemplarily we confirmed interactions of AtMYB82 and AtMYB4, respectively, with two of the R/B-like BHLH proteins, proving that the predictions were valid. AtMYB4 so far is the only R2R3 MYB protein from *A. thaliana* shown to have repressor function. It has been described to be a negative regulator of the cinnamate 4 hydroxylase gene, thus negatively modulating sinapate ester formation in the absence of UVB light (Jin *et al.*, 2000). As predicted, the R2R3 domain of AtMYB4 alone is sufficient to mediate interaction with R/B BHLH proteins. In fact, removal of the C-terminal part of AtMYB4 led to strongly increased reporter activation in yeast. This effect might indicate an inhibitory influence of the C-terminus on the interaction rather than a transcriptional repression activity of AtMYB4 in yeast, because deletion of the proposed repressor motif LNL[ED]L (Jin *et al.*, 2000) alone did not lead to a similar effect. The precise formulation of the interaction motif allows differentiation even of phylogenetically very closely related proteins. AtMYB4 and AtMYB6 share more than 90% identical amino acids in the MYB domain. However, only four of the six required amino acids of the interaction motif are present in the R3 repeat of AtMYB6 and neither the full-length protein nor the R2R3 domain alone interacted with R/B-like BHLH proteins in yeast. The predicted interactions between MYB proteins and R/B-like BHLH proteins, which so far have been confirmed with the yeast two-hybrid system only, provide the basis for future research to investigate a functional relevance of these hetero-dimerizations *in vivo*.

Our findings support the emerging view that, in addition to their function in DNA binding, a conserved function of MYB domains is their involvement in protein–protein interactions with accessory factors. Both, coactivators (e.g. C/EBPβ and p100) and inhibitors (e.g. c.MAF, D-cyclins, Cyp40) have been shown to bind to MYB domains of mammalian c-MYB proteins (Ganter *et al.*, 1998; Kanei-Ishii *et al.*, 1997; Ness, 1999). A partially exposed hydrophobic patch was predicted to be an interaction site for MYB binding proteins (Ogata *et al.*, 1995). Recently, the crystal structure of a complex composed of c-MYB, C/EBPβ and a promoter DNA fragment was solved. Six amino acids within the hydrophobic patch of the MYB R2 repeat were found to make direct contact to C/EBPβ (Tahirov *et al.*, 2002). The general structural similarity of MYB domains and the positions of the conserved tryptophan residues allow the alignment of even distantly related MYB repeats. Interestingly, aligning the R3 repeat of plant MYB proteins to the c-MYB R2 repeat reveals that the MYB/BHLH interaction surface characterized in this study exactly matches the position of the C/EBPβ binding site. The similarities extend even down to the level of the positions of the amino acids of the interaction motif and the central invariable hydrophobic residues, relative to the conserved second tryptophan residue of the MYB consensus sequence. These striking similarities support the idea of a general functional conservation of the MYB domain as a protein–protein interaction module used for the recruitment of a diverse set of accessory proteins.

The precise characterization of amino acid motives is a valuable tool to understand functional specificities within protein families. With the rapid accumulation of genomic data, the prediction of protein properties from sequence is becoming increasingly important for the transfer of knowledge from model species like *A. thaliana* to plants of agricultural interest.

Experimental procedures

Yeast strain, transformation and two-hybrid assays

All yeast two-hybrid analysis was performed in the yeast strain AH109 (James *et al.*, 1996) and Y187 (Harper *et al.*, 1993). The yeast transformation was performed with the LiOAc/single-stranded DNA/PEG method (Gietz *et al.*, 1995). Co-transformed and mated cells were plated onto synthetic dropout medium lacking leucine, tryptophan, and histidine and supplemented with 5–40 mM 3-aminotriazole (3-AT) (Sigma-Aldrich, Munich, Germany) to investigate interaction of the hybrid proteins. Recombinant hybrid proteins were tested for self-activation and non-specific protein-binding properties. α-galactosidase activity was assayed with *p*-nitrophenyl α-D-galactopyranoside (PNPG; Sigma-Aldrich) according to a protocol provided with the Matchmaker system (CLONTECH, Palo Alto, CA, USA).

Yeast two-hybrid screens were performed as described before (Soellick and Uhrig, 2001). The cDNAs encoding EGL3/AtBHLH002 and AtBHLH012 were recombined into the bait vector pCD2attR (J.F. Uhrig, unpublished). The Gal4BD-fused BHLH proteins were used as bait proteins to screen three different *Arabidopsis* cDNA-libraries (two libraries of whole plant, one of an *A. thaliana* cell culture). The screening procedure was performed by mating the bait proteins with pre-transformed frozen prey-library yeast cells. For this approach two yeast strains with different mating type, the yeast

strains AH109 and Y187 were used. The selection was carried out on synthetic dropout medium lacking leucine, tryptophan, and histidine and supplemented with 5 mM 3-AT. After 6 days candidates were analysed by PCR and the insert was sequenced. The candidates were tested in a second transformation by *in vivo* recombination in yeast and cotransformation with the original bait protein.

Constructs and recombinant DNA manipulations

DNA encoding the MYB and BHLH proteins were amplified by PCR using primers containing attB1 and attB2 sites for Gateway recombination as described by Invitrogen and 20 nucleotides of the according cDNA and inserted into pDONR201 (Invitrogen, Carlsbad, CA, USA). These ENTRY-vectors were used for recombination in the yeast shuttle vectors pCD2attR and pCACT2attR (J.F. Uhrig, unpublished), Gateway vectors based on the vectors pCD2 and pCACT (Durfee *et al.*, 1999). Deletion constructs were generated by PCR using site-specific primers and recombination via Gateway technology (Invitrogen). To generate the vector AD-AtMYB4_Δ 199–210 the ENTRY-vector was cut with BsaBI, which cuts twice at +513 bp and +629 bp in the AtMYB4 cDNA coding sequence and religated, as described in Jin *et al.* (2000). The deleted AtMYB4 cDNA was introduced into pCACT2attR via Gateway technology (Invitrogen). Point mutations were inserted by PCR of the PAP1-ENTRY vector using primers containing the appropriate nucleotides and ligation of the PCR products. Accuracy of the constructs was confirmed by sequencing. Random mutagenesis was performed by error-prone PCR in the presence of 0.1 mM $MnCl_2$, a concentration found to result in an average of 4.5 substitutions per allele. Mutants were selected on triple dropout media at 22°C, indicating interaction at the permissive temperature of 22°C. PAP1/AtMYB75 was amplified with *Taq*-Polymerase (Qiagen, Hilden, Germany) in buffer containing 0.1 mM $MnCl_2$ [primers AO015 (5′-GAAAGCAACCTGACCTACAGGAAAGAG-3′) and AO018 (5′-GAGAGACAATTGGTATATAACTATCTATTCGATG-3′) on the vector pCACT-PAP1]. The PCR product was co-transformed with linearized pACT2 (BD Biosciences Clontech, Franklin Lakes, NJ, USA) into yeast (Y187). The obtained colonies were mated with AH109 transformed with EGL3 and TT8, respectively. The growth of the zygotes was examined on media lacking leucine, tryptophan and histidine with 3 mM 3-AT at 22 and 30°C and temperature-sensitive mutants were analysed further.

In vitro *pull down assays*

cDNAs from PAP1 and PAP2 were recombined in frame into pGEX-2T-attR [Gateway compatible vector based on pGEX-2T; Amersham Biosciences (J.F. Uhrig, unpublished)] and transformed into BL21 Codon Plus ^{TM}-DE3 RIL bacteria (Stratagene, La Jolla, CA, USA) and used for purification of glutathione *S*-transferase (GST)-fusion proteins (Ausubel *et al.*, 1994). cDNAs of GL3, EGL3, AtBHLH012 and TT8 were recombined in frame into pDEST14 (Invitrogen) and AtBHLH013 into pET32attR [Gateway compatible vector based on pET32 (Qiagen), J.F. Uhrig, unpublished] which served as templates to synthesize [^{35}S]methionine-labelled proteins in a coupled transcription-translation system (Promega, Madison, WI, USA). Equal amounts of ^{35}S-labelled protein were incubated with glutathione-agarose coupled to GST-PAP1, GST-PAP2 and the GST protein, respectively in 150 μl of binding buffer (20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 0,01% IGEPAL (Sigma-Aldrich)) for 1 h at 4°C. After removal of the supernatant, the beads were washed extensively with binding buffer. The matrix-bound proteins were eluted with 2x SDS-loading buffer, and separated on a 12% SDS-polyacrylamide gel to detect the ^{35}S-labelled proteins by autoradiography.

Sequence analysis and bioinformatics

DNA sequences were determined by the MPIZ DNA core facility on Applied Biosystems (Weiterstadt, Germany) Abi Prism 377 and 3700 sequencers using BigDye-terminator chemistry. Premixed reagents were from Applied Biosystems.

Sequence alignments were generated using the ClustalW program with default settings (Thompson *et al.*, 1994). Information content (relative entropy) of alignments was determined according to Schneider and Stephens (1990) using the sequence part applied in the RNA structure logos [Gorodkin *et al.* (1997), http://www.cbs.dtu.dk/~gorodkin/appl/plogo.html]. Three-dimensional models of MYB domains were built using the 3D-Jigsaw server [Bates *et al.* (2001), http://www.bmm.icnet.uk/3djigsaw/]. The NMR structure of the MYB domain of c-MYB (Ogata *et al.*, 1994) was used as a guideline during modelling.

Co-transfection experiments

The *A. thaliana* At7 cell culture, the protoplast isolation, the co-transfection, and the determination of standardized GUS activity were carried out as described by Hartmann *et al.* (1998) except that 30 μg of plasmid DNA was used for the PEG-mediated DNA transfer into protoplasts: 10 μg of the p *AtDFR*:: *uidA* reporter construct; 0.5 μg of each effector construct; if added, 5 μg of the standardization plasmid pBT8UBI-LUCm3, expressing the luciferase transformation control; and an inactive luciferase expression vector (pBTΔLUC) to complete to the amount of 30 μg of DNA. The p *AtDFR*:: *uidA* reporter construct contains a 520-bp fragment of the DFR promoter (−520 to ATG) and was constructed by Frank Mehrtens. A 1.1-kb promoter fragment was amplified with primers FM31 (5′-GGTTGAAGAAGAAGAAGGAAAGCTTTGAAG-3′) and FM32 (5′-CTGACTAACCATGGTTGTGGTTATATG-3′) containing an *Nco*I restriction site. The purified PCR product was digested with the restriction enzymes *Hin*dIII and *Nco*I. This resulted in two fragments (520 and 560 bp), of which the 520 bp fragment was cloned into pBT10-35S::GUS digested by *Hin*dIII and *Nco*I, replacing the 35S cassette.

The effectors used in this study were constructed using the pBTdest vector (GenBank accession number: AJ551314). Full-length cDNAs were amplified using primer sets containing the attB1 and attB2 recombination sequences (EGL3/AtBHLH002: MJ196: 5′-attB1-CCATGGCAACCGGAGAAAACAGAACGGTG-3′, MJ193: 5′-attB2-TTAACATATCCATGCAACCCTTTGAAGTGCC-3′; PAP1/AtMYB75: GTW-MYB75-1: 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGGAGGGTT CGTCCAAAGGGCTGC-3′, GTW-MYB75-2: 5′-GGGGACCACTTTGTACAAGA AAGCTGGGTTCTAATCAAATTTCACAGTCTCTCCATCG-3′). The amplification products were recombined into the pDONR201 entry vector (Invitrogen), sequenced and then transferred into pBTdest via an LR recombination.

Acknowledgements

The authors gratefully acknowledge Ralf Stracke and Marc Jakoby for providing cDNA clones and for helpful discussions. The PhD works of I.M.Z. and M.H. were funded by a grant of the BMBF-GABI project (German plant genomics program 'Genomanalyse im biologischen System Pflanze') to B.W. and by the Deutsche Forschungsgemeinschaft ('Graduiertenkolleg für Molekulare Analysen von Entwicklungsprozessen' der Universität zu Köln; IIIGK-GRK 306/1), respectively. We also thank Francesco Salamini and Klaus Hahlbrock for constant support, laboratory space and encouragement.

References

Aharoni, A., De Vos, C.H.R., Wein, M., Sun, Z., Greco, R., Kroon, A., Mol, J.N.M. and O'Connell, A.P. (2001) The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *Plant J.* **28**, 319–332.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1994) *Current Protocols in Molecular Biology*. New York: Wiley.

Bailey, P.C., Martin, C., Toledo-Ortiz, G., Quail, P.H., Huq, E., Heim, M.A., Jakoby, M., Werber, M. and Weisshaar, B. (2003) Update on the basic helix-loop-helix transcription factor gene family in *Arabidopsis thaliana*. *Plant Cell*, **15**, 2497–2502.

Bates, P.A., Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins: Structure, Function and Genetics* (Suppl 5), 39–46.

Borevitz, J.O., Xia, Y.J., Blount, J., Dixon, R.A. and Lamb, C. (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell*, **12**, 2383–2393.

Carey, C.C., Strahle, J.T., Selinger, D.A. and Chandler, V.L. (2004) Mutations in the pale aleurone color1 regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in *Arabidopsis thaliana*. *Plant Cell*, **16**, 450–464.

Durfee, T., Draper, O., Zupan, J., Conklin, D.S. and Zambryski, P.C. (1999) New tools for protein linkage mapping and general two-hybrid screening. *Yeast*, **15**, 1761–1768.

Frampton, J., Gibson, T.J., Ness, S.A., Doderlein, G. and Graf, T. (1991) Proposed structure for the DNA-binding domain of the Myb oncoprotein based on model building and mutational analysis. *Protein Eng* **4**, 891–901.

Galway, M.E., Masucci, J.D., Lloyd, A.M., Walbot, V., Davis, R.W. and Schiefelbein, J.W. (1994) The TTG gene is required to specify epidermal cell fate and cell patterning in the *Arabidopsis* root. *Dev. Biol.* **166**, 740–754.

Ganter, B., Fu, S.-L. and Lipsick, J.S. (1998) D-type cyclins repress transcriptional activation by the v-Myb but not the c-Myb DNA-binding domain. *EMBO J.* **17**, 255–268.

Gietz, R.D., Schiestl, R.H., Willems, A.R. and Woods, R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, **11**, 355–360.

Goff, S.A., Cone, K.C. and Chandler, V.L. (1992) Functional analysis of the transcriptional activator encoded by the maize B gene, evidence for a direct functional interaction between two classes of regulatory proteins. *Genes. Dev.* **6**, 864–875.

Gong, Z., Yamazaki, M. and Saito, K. (1999) A light inducible *Myb*-like gene that is specifically expressed in red *Perilla frutescens* and presumably acts as a determining factor of the anthocyanin forma. *Mol. Gen. Genet.* **262**, 65–72.

Gorodkin, J., Heyer, L.J., Brunak, S. and Stormo, G.D. (1997) Displaying the information contents of structural RNA alignments, the structure logos. *Comput. Appl. Biosci.* **13**, 583–586.

Grotewold, E., Drummond, B.J., Bowen, B. and Peterson, T. (1994) The *myb*-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell*, **76**, 543–553.

Grotewold, E., Sainz, M.B., Tagliani, L., Hernandez, J.M., Bowen, B. and Chandler, V.L. (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proc. Natl Acad. Sci. USA*, **97**, 13579–13584.

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K. and Elledge, S.J. (1993) The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell*, **75**, 805–816.

Hartmann, U., Valentine, W.J., Christie, J.M., Hays, J., Jenkins, G.I. and Weisshaar, B. (1998) Identification of UV/blue light-response elements in the *Arabidopsis thaliana* chalcone synthase promoter using a homologous protoplast transient expression system. *Plant. Mol. Biol.* **36**, 741–754.

Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B. and Bailey, P.C. (2003) The basic helix-loop-helix transcription factor family in plants, a genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* **20**, 735–747.

James, P., Halladay, J. and Craig, E.A. (1996) Genomic libraries and a host strain designed for highly efficient two- hybrid selection in yeast. *Genetics*, **144**, 1425–1436.

Jin, H. and Martin, C. (1999) Multifunctionality and diversity within the plant *MYB*-gene family. *Plant. Mol. Biol.* **41**, 577–585.

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., Tonelli, C., Weisshaar, B. and Martin, C. (2000) Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *EMBO J.* **19**, 6150–6161.

Johnson, C.S., Kolevski, B. and Smyth, D.R. (2002) TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell*, **14**, 1359–1375.

Kanei-Ishii, C., Sarai, A., Sawazaki, T., Nakagoshi, H., He, D.N., Ogata, K., Nishimura, Y. and Ishii, S. (1990) The tryptophan cluster, a hypothetical structure of the DNA-binding domain of the *myb* protooncogene product. *J. Biol. Chem.* **265**, 19990–19995.

Kanei-Ishii, C., Tanikawa, J., Nakai, A., Morimoto, R.I. and Ishii, S. (1997) Activation of heat shock transcription factor 3 and nbsp;by c-Myb in the absence of cellular stress. *Science*, **277**, 246–248.

Lee, M.M. and Schiefelbein, J. (1999) WEREWOLF, a MYB-related protein in *Arabidopsis*, is a position-dependent regulator of epidermal cell patterning. *Cell*, **24**, 473–483.

Lee, M.M. and Schiefelbein, J. (2001) Developmentally distinct MYB genes encode functionally equivalent proteins in *Arabidopsis*. *Development*, **128**, 1539–1546.

Lloyd, A.M., Walbot, V. and Davis, R.W. (1992) *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators R and C1. *Science*, **258**, 1773–1775.

Marks, M.D. and Esch, J.J. (2003) Initiating inhibition. *EMBO Rep.* **4**, 24–25.

Martin, C. and Paz-Ares, J. (1997) MYB transcription factors in plants. *Trends Genet.* **13**, 67–73.

Meissner, R.C., Jin, H., Cominelli, E. *et al.* (1999) Function search in a large transcription factor gene family in *Arabidopsis*, assessing the potential of reverse genetics to identify insertional mutations in R2R3 MYB genes. *Plant Cell*, **11**, 1827–1840.

Mol, J., Grotewold, E. and Koes, R. (1998) How genes paint flowers and seeds. *Trends Plant Sci.* **3**, 212–217.

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M. and Lepiniec, L. (2000) The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. *Plant Cell*, **12**, 1863–1878.

Nesi, N., Jond, C., Debeaujon, I., Caboche, M. and Lepiniec, L. (2001) The *Arabidopsis TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell*, **13**, 2099–2114.

Ness, S.A. (1999) Myb binding proteins, regulators and cohorts in transformation. *Oncogene*, **18**, 3039–3046.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, **79**, 639–648.

Ogata, K., Morikawa, S., Nakamura, H. *et al.* (1995) Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Nat. Struct. Biol.* **2**, 309–320.

Pabo, C.O. and Sauer, R.T. (1992) Transcription factors, structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.

Payne, C.T., Zhang, F. and Lloyd, A.M. (2000) *GL3* encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1. *Genetics*, **156**, 1349–1362.

Petroni, K., Tonelli, C. and Paz-Ares, J. (2002) The MYB transcription factor family, from maize to *Arabidopsis* [Review]. *Maydica*, **47**, 213–232.

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J. and Koes, R. (1999) Molecular analysis of the *anthocyanin2* gene of Petunia and its role in the evolution of flower color. *Plant Cell*, **11**, 1433–1444.

Rabinowicz, P., Braun, E., Wolfe, A., Bowen, B. and Grotewold, E. (1999) Maize R2R3 *Myb* genes, sequence analysis reveals amplification in the higher plants. *Genetics*, **153**, 427–444.

Ramsay, N.A., Walker, A.R., Mooney, M. and Gray, J.C. (2003) Two basic-helix-loop-helix genes (MYC-146 and GL3) from *Arabidopsis* can activate anthocyanin biosynthesis in a white-flowered *Matthiola incana* mutant. *Plant. Mol. Biol.* **52**, 679–688.

Riechmann, J.L., Heard, J., Martin, G. *et al.* (2000) *Arabidopsis* transcription factors, genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105–2110.

Romero, I., Fuertes, A., Benito, M.J., Malpica, J.M., Leyva, A. and Paz-Ares, J. (1998) More than 80R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*. *Plant J.* **14**, 273–284.

Rubio, V., Linhares, F., Solano, R., Martin, A.C., Iglesias, J., Leyva, A. and Paz-Ares, J. (2001) A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes Dev.* **15**, 2122–2133.

Sawa, S. (2002) Overexpression of the AtmybL2 gene represses trichome development in *Arabidopsis*. *DNA Res.* **9**, 31–34.

Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., Thumfahrt, J., Jurgens, G. and Hulskamp, M. (2002) TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *EMBO J.* **21**, 5036–5046.

Schiefelbein, J. (2003) Cell-fate specification in the epidermis, a common patterning mechanism in the root and shoot. *Curr. Opin. Plant Biol.* **6**, 74–78.

Schneider, T.D. and Stephens, R.M. (1990) Sequence logos, a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100.

Soellick, T.R. and Uhrig, J.F. (2001) Development of an optimized interaction-mating protocol for large-scale yeast two-hybrid analyses. *Genome Biol.* **2**, 0052.1–0052.7.

Stracke, R., Werber, M. and Weisshaar, B. (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **4**, 447–456.

Tahirov, T.H., Sato, K., Ichikawa-Iwata, E. *et al.* (2002) Mechanism of c-Myb-C/EBP beta cooperation from separated sites on a promoter. *Cell*, **108**, 57–70.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL, W, Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Wada, T., Tachibana, T., Shimura, Y. and Okada, K. (1997) Epidermal cell differentiation in *Arabidopsis* determined by a *Myb* homolog, CPC. *Science*, **277**, 1113–1116.

Wada, T., Kurata, T., Tominaga, R., Koshino-Kimura, Y., Tachibana, T., Goto, K., Marks, M.D., Shimura, Y. and Okada, K. (2002) Role of a positive regulator of root hair development, CAPRICE, in *Arabidopsis* root epidermal cell differentiation. *Development*, **129**, 5409–5419.

Winkel-Shirley, B. (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant. Physiol.* **126**, 485–493.

Wolberger, C. (1999) Multiprotein-DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 29–56.

Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T. and Lloyd, A. (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development*, **130**, 4859–4869.