COMBINATORIAL TRANSCRIPTIONAL REGULATION OF THE MAIZE
FLAVONOID PATHWAY: UNDERSTANDING THE OLD PLAYERS AND
DISCOVERING NEW ONES.

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University.

By

Julia Marcela Hernandez, M.S.

*****

The Ohio State University

2006

Dissertation Committee:

Approved by

Professor Erich Grotewold, Adviser

Professor David M. Bisaro

Professor Vicki L. Chandler

Professor Mark R. Parthun

Professor Michael C. Ostrowski

Ohio State Biochemistry Graduate Program
ABSTRACT

The maize flavonoid biosynthetic pathway has two distinct branches regulated by two very similar MYB domain transcription factors. C1 regulates the accumulation of anthocyanins while P1 regulates the synthesis of phlobaphenes. These two proteins exhibit more than 70% identity in the amino acid sequence of their MYB domain, bind the same DNA sequences in vitro, albeit with different affinity, and share a set of common target genes. One difference between these proteins is that C1 is absolutely dependent on the bHLH cofactor R for activity while P1 is not. We have shown that it is through the interaction with R that C1 achieves specificity. We have also found that the function of R can be dissected into an R-enhanced activity and an R dependent activity. The R-enhanced activity is achieved through the ARE, a cis-DNA element present in all the C1 target genes. I propose that R recruits a DNA binding factor that recognizes and binds the ARE. The R-dependent activity is an effect that R has on the C1 protein that allows it to bind and activate its target promoters. Due to its multidomain structure, R is likely to recruit additional factors to the regulatory complex. We have identified an ENT domain protein, RIF-1C, which specifically interacts with the bHLH region of R, and uncovered evidence that suggests that the bHLH and RIF-1C are involved in chromatin remodeling.
Dedicated to my children, Nicolás and Andrés
ACKNOWLEDGMENTS

I would like to thank my adviser Erich Grotewold, who gave me the opportunity to rediscover science and helped me tremendously in my scientific growth and development.

I also would like to thank Vicki Chandler and Ben Bowen for a very fruitful collaboration and for the many materials they shared with us.

I am very grateful to the members of my committee, Dave Bisaro, Vicki Chandler, Mike Ostrowski, and Mark Parthun, for their input and their guidance.

Thanks to Daan van Aalten for helping us with the modeling of MYB domains, Tien-Hsien Chang for the DED1p antibody, Carmelita Bautista and Margaret Falkowski at the Antibody Facility at Cold Spring Harbor Laboratory, for the monoclonal C1 antibody, Sue Wessler for providing us the cDNA of the D12 mutant, and Daniel Janies for help with the phylogeny of the plant ENT proteins. I would also like to thank Brian Kemmenoe at the Campus Microscopy facility, and Herbert Auer at the Columbus Children’s Research Institute for excellent technical assistance.

I am also very grateful to the faculty, students, postdocs, and staff of the Plant Biotechnology Center for the many scientific discussions, reagents, advice, and support. I
consider my interaction with this group one of the most productive experiences during my graduate school career.

A very special thank you also to my labmates Ed Braun, Lila Pooma, Todd Matulnik, Min-Gab Kim, George Heine, Antje Feller, and Kengo Morohashi with whom I teamed up for many aspects of this work. I would also like to acknowledge the help of many undergraduate students including Kevin Johnson, Jamie Smialek, Maddalena Pizzirusso, George Wang, Kris Nemeth, and Brett Hirsch, who worked under my supervision and whose help was critical for the timely completion of this work.

I am very grateful for all the computer support that Sarat C. Subramaniam provided me as well as for all the help with the Rightmire Hall and greenhouse facilities provided by Joe Takayama, Dave Long and Scott Hines. A special thanks to Melinda Parker and Diane Furtney for their excellent administrative support.

Finally I would like to thank the OSBP staff, director, former director, faculty and students for providing me with an excellent interactive atmosphere that complemented my scientific and professional development.

This research was supported by grants from the National Science Foundation.
VITA

June 13, 1969  ………….. Born - Bogota, Colombia

1991  ……………… B.S. Molecular Genetics, The Ohio State University

1995  ………………… M.S. Biochemistry, The Ohio State University

1995 - 2002  …………… Research Associate, Department of Molecular Genetics, and Plant Biotechnology Center, The Ohio State University.

2002 - Present  …………… Research Associate, Plant Biotechnology Center, The Ohio State University.

PUBLICATIONS


FIELDS OF STUDY

Major Field: Biochemistry Program
### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td><strong>Chapters:</strong></td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Combinatorial transcriptional control</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Examples of combinatorial transcriptional control in eukaryotes.</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Additional determinants of specificity of transcription factors.</td>
<td>4</td>
</tr>
<tr>
<td>1.4 The flavonoid biosynthetic pathway: A model system to understand</td>
<td>5</td>
</tr>
<tr>
<td>combinatorial control of gene transcription in plants.</td>
<td></td>
</tr>
<tr>
<td>1.5 Regulatory genes of the flavonoid biosynthetic pathway.</td>
<td>6</td>
</tr>
<tr>
<td>1.5.1 The MYB domain factors.</td>
<td>7</td>
</tr>
<tr>
<td>1.5.2 The bHLH domain factors.</td>
<td>8</td>
</tr>
<tr>
<td>1.5.2.1 The MYB interacting region.</td>
<td>9</td>
</tr>
<tr>
<td>1.5.2.2 The bHLH domain.</td>
<td>10</td>
</tr>
<tr>
<td>1.5.3 A new class of WD repeat proteins regulates anthocyanin biosynthesis.</td>
<td>11</td>
</tr>
<tr>
<td>1.5.4 The MYB-bHLH-WD-repeat regulatory network.</td>
<td>13</td>
</tr>
<tr>
<td>1.6 Dissertation aims</td>
<td>16</td>
</tr>
<tr>
<td>2. Identification of the residues in the Myb domain of maize C1 that</td>
<td>21</td>
</tr>
<tr>
<td>specify the interaction with the bHLH cofactor R</td>
<td></td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>21</td>
</tr>
<tr>
<td>2.2 Materials and methods</td>
<td>24</td>
</tr>
<tr>
<td>2.2.1 Plasmids used in transient expression experiments.</td>
<td>24</td>
</tr>
<tr>
<td>2.2.2 Plasmids used in yeast two-hybrid experiments.</td>
<td>25</td>
</tr>
<tr>
<td>2.2.3 Microprojectile bombardment and gene expression experiments.</td>
<td>26</td>
</tr>
<tr>
<td>2.2.4 Yeast two-hybrid experiments.</td>
<td>26</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>27</td>
</tr>
</tbody>
</table>
2.3.1 Regulatory activity of chimeric MYB domains ............................................ 27
2.3.2 Four residues are sufficient to transfer the interaction with R from C1 to P. 30
2.3.3 P1 mutants interact with R in plant cells. ..................................................... 32
2.3.4 Interaction with the cofactor R is necessary for transcriptional activation of $BzI$ ................................................................................................................. 34
2.4 Discussion ......................................................................................................... 34
2.4.1 Specificity of the interaction between MYB domains and co-activators..... 34
2.4.2 Four amino acid changes are sufficient to transfer the interaction with R from C1 to P1......................................................................................................... 35
2.4.3 R contributes to the regulatory specificity of MYB transcription factors.... 36

3. Different Mechanisms participate in the R-dependent activity of the R2R3 MYB transcription factor C1 ................................................................................................................. 43

3.1 Introduction ....................................................................................................... 43
3.2 Materials and methods ..................................................................................... 46
3.2.1 Plasmids used in yeast experiments .............................................................. 46
3.2.2 Plasmids used in transient expression experiments ..................................... 47
3.2.3 Microprojectile bombardment and gene expression experiments ............... 48
3.2.4 Microprojectile bombardment and gene expression experiments ............... 49
3.2.5 Expression and purification of proteins expressed in bacteria .................... 50
3.2.6 Generation of C1 monoclonal antibodies and Western analyses ................ 51
3.2.7 Electrophoretic mobility shift assays (EMSA). ............................................ 52
3.3 Results ............................................................................................................... 54
3.3.1 The ARE is required for R-enhanced activity............................................... 54
3.3.2 Contrary to P1, C1 does not activate transcription in yeast from the $haPBS$. 55
3.3.3 C1$^{SH}$, a mutant of C1 that activates transcription in yeast ......................... 56
3.3.4 C1$^{SH}$ binds DNA with high affinity. ......................................................... 57
3.3.5 High-affinity DNA-binding is not sufficient for R-independent transcription by C1 .............................................................. 58
3.3.6 The fusion of C1 with the Gal4 DNA-binding domain makes Gal4-mediated transcription dependent on R ........... 59
3.4 Discussion ......................................................................................................... 60

4. Characterization of RIF-1C, a novel R-interacting factor that harbors a BRCA2-binding like domain .............................................................................................. 77

4.1 Introduction ....................................................................................................... 77
4.2 Materials and methods ..................................................................................... 79
4.2.1 Plasmids used in yeast two-hybrid experiments and GST pull-downs ....... 79
4.2.2 Plasmids used in transient expression experiments .................................. 80
4.2.3 Plasmids for GFP transient assays ............................................................. 81
4.2.4 Plasmids for RNAi experiment ................................................................. 81
4.2.5 Plant material and seed stocks ................................................................. 82
4.2.6 Yeast two-hybrid experiments ............................................................... 82
4.2.7 GST pull-down experiments ................................................................. 83
4.2.8 Microprojectile Bombardment and Gene Expression Experiments ....... 85
4.2.9 GFP transient assays ............................................................................ 85
4.2.10 RNAi analysis ...................................................................................... 86
4.2.11 Silique analysis ................................................................................... 86
4.3 Results ...................................................................................................... 87
   4.3.1 Mutations in the bHLH of R result in a decrease of activation of its target genes ............................................................... 87
   4.3.2 RIF-1C interacts specifically with the bHLH region of R and homodimerizes through its ENT domain ......................... 89
   4.3.3 RIF-1C expression profile in maize .................................................... 91
   4.3.4 RIF-1C localizes to the nucleus ........................................................ 91
   4.3.5 The plant ENT domain proteins ......................................................... 92
   4.3.6 Arabidopsis has an RIF-1C homolog .................................................. 93
   4.3.7 Phenotype of AtAck1 mutants ......................................................... 94
4.4 Discussion ................................................................................................. 98
   5 Final discussion ....................................................................................... 124
       5.1 The MYB-bHLH interaction and the implications for identifying other regulatory complexes .............................................. 125
       5.2 The R-enhanced and the R-dependent activities may be conserved in all plant species ...................................................... 127
       5.3 R-interacting factors ........................................................................ 129
       5.4 Are RIF-1C and R involved in gene silencing? ................................. 130
       5.5 Final remarks .................................................................................... 131
List of references ........................................................................................... 133
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Identified regulators of the anthocyanin biosynthetic pathway identified in plants</td>
<td>20</td>
</tr>
<tr>
<td>2.1 Dependence on R of C1, P1 and a P1 mutant for the activation of the $A1$ and $Bz1$ genes [130]</td>
<td>42</td>
</tr>
<tr>
<td>3.1 Comparison of the apparent equilibrium dissociation constants ($K_d$) for the $A1$ high affinity P1 binding site by the MYB domains of P ($P^{MYB}$), C1 ($C1^{MYB}$) and C1$^{SH}$ ($C1^{SHMYB}$) [119]</td>
<td>76</td>
</tr>
<tr>
<td>4.1 RIF-1C expression profile according to available EST data</td>
<td>122</td>
</tr>
<tr>
<td>4.2 Results of crosses between homozygous wild type background plants (WT) or wild type Columbia plants (Col), and homozygous mutant $ack1-1$, $ack1-2$, $ack1-3$ plants</td>
<td>123</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 General chemical structure of flavonoids.</td>
<td>18</td>
</tr>
<tr>
<td>2.1 Activation of the <em>A1</em> and <em>Bz1</em> promoters.</td>
<td>38</td>
</tr>
<tr>
<td>2.2 MYB-domain sequences that contribute to the specificity of the interaction with the bHLH cofactor R.</td>
<td>39</td>
</tr>
<tr>
<td>2.3 Transfer of the interaction with R from C1 to P1 <em>in vivo</em>.</td>
<td>40</td>
</tr>
<tr>
<td>2.4 Solvent-exposed surfaces of the C1 and P MYB domains.</td>
<td>41</td>
</tr>
<tr>
<td>3.1 Sequence of the MYB domains of C1 and P1 and structure of the plasmids used for transient expression experiments.</td>
<td>67</td>
</tr>
<tr>
<td>3.2 Activation of transcription by P1, C1 and P1*.</td>
<td>68</td>
</tr>
<tr>
<td>3.3 P1, but not C1, activates transcription in yeast.</td>
<td>70</td>
</tr>
<tr>
<td>3.4 C1&lt;sup&gt;SH&lt;/sup&gt; activates transcription in yeast and interacts with R.</td>
<td>71</td>
</tr>
<tr>
<td>3.5 C1&lt;sup&gt;SH&lt;/sup&gt; binds DNA with high affinity.</td>
<td>72</td>
</tr>
<tr>
<td>3.6 The activity of C1&lt;sup&gt;SH&lt;/sup&gt; remains R-dependent in maize cells.</td>
<td>73</td>
</tr>
<tr>
<td>3.7 C1 confers R-dependency upon the Gal4 DNA-binding domain.</td>
<td>74</td>
</tr>
</tbody>
</table>
3.8 Mechanisms of cooperation between C1 or P1* and R. ............................................. 75

4.1 Effect of mutations in the bHLH domain of R in transient expression assays. ....... 103

4.2 Chromatin immunoprecipitation analysis of the promoters of the maize anthocyanin biosynthetic genes................................................................. 105

4.3 Yeast two-hybrid analysis of RIF-1C. ................................................................. 106

4.4 Interaction of RIF-1C and the C-terminus of R in vitro. ............................... 108

4.5 Domain structure and dissection of RIF-1C. ................................................. 109

4.6 Nuclear localization of RIF-1C ................................................................. 110

4.7 ENT domain proteins in plants. ......................................................... 111

4.8 Expression of ACK1 and closely related genes in Arabidopsis organs and developmental stages. ................................................................. 113

4.9 Yeast two-hybrid analysis of the ENT domain of ACK1 ......................... 114

4.10 Mutant alleles of ACK1 ........................................................................ 115

4.11 Silique and seed analysis of ACK1 mutants ........................................... 117

4.12 Silique and seed analysis of ACK1 RNAi transgenic lines ...................... 118

4.13 Analysis of ACK1 expression in RNAi lines by real time PCR ............ 120

4.14 Protein alignment of the PHD domain of ATX1 and ACK1 ..................... 121

xiii
CHAPTER 1

INTRODUCTION

1.1 Combinatorial transcriptional control.

Families of transcription factors are generally defined by the presence of conserved DNA-binding motifs[1]. Often, transcription factors belonging to the same family recognize a common core consensus DNA sequence. Thus, how transcription factors with similar DNA-binding domains control the expression of specific sets of genes has puzzled scientists for a long time. The notion that is emerging from a number of studies in eukaryotes is that specificity is provided by the association of transcription factors with other proteins through protein-protein interactions. A given transcription factor may be part of a protein complex that regulates a specific set of genes and form a different complex when regulating another set of genes. Different complexes may also determine if a transcription factor is functioning as an activator or repressor of transcription. This strategy of involving partner proteins is also known as combinatorial transcriptional regulation. Multiple proteins binding the same promoter and interacting with one another imply additional requirements such as the presence of multiple binding sites in the promoter, and co-expression of all the protein partners in the cells where the target gene is transcribed.
There are different ways in which combinatorial transcriptional regulation influences the transcriptional activity or the ability of transcription factors to activate specific sets of target genes (reviewed in [2], [3]). A given family of transcription factors may exhibit more than one of these modes of regulation.

1.2 Examples of combinatorial transcriptional control in eukaryotes.

The ETS gene family, which includes factors regulating cell growth, differentiation and apoptosis in metazoans, exhibit a different strategy in the GABPα/β complex than in the SAP-1/SRF or the Ets-1/Pax-5 complexes. The non-DNA-binding protein GABPβ forms a heterodimer with the ETS protein GABPα (reviewed in [4]), which associates with itself forming a heterotetramer, to bind DNA [5]. GABPβ provides an activation domain, a nuclear localization signal, increases the affinity of GABPα for DNA, and allows binding of the complex to two tandem binding sites [6, 4]. Other ETS proteins form complexes with other DNA-binding factors, as exemplified by the SAP-1/SRF and the Ets-1/Pax-5 complexes. In both cases, the DNA-binding domains for the partner proteins are positioned in a way such that it allows them to bind to opposite faces of the DNA helix. In the case of SAP-1, an additional interaction is provided by a motif located at the C-terminal. This interaction has been shown to dominate the partnership with the MADS-box protein SRF, which in turn binds as a dimer [7, 8]. The Ets-1/Pax-5 partnership illustrates how an interaction can allow a transcription factor to bind sites with a minimal core, which otherwise it would not bind. The Ets-1 and Pax-5 DNA-
binding sites are overlapping which results in a suboptimal ETS binding site, yet favors Pax-5 binding. Ets-1 is able to bind DNA due to a conformational change induced by the presence of Pax-5 ([9], reviewed in [10]).

The VP16-HCF-1-Oct-1 provides another excellent example of combinatorial regulation of transcription. VP16 is a viral protein from the herpes simplex virus which interacts with host transcription factors to mediate activation of the necessary viral genes to enter the lytic cycle. Association of VP16 with HCF-1 is necessary to form a complex with the POU domain protein Oct-1 at the target promoters (reviewed in [11]). Recently, it has been shown that in Oct-1 deficient cells, the viral infection is arrested, unveiling a second role of Oct-1 in viral replication [12]. HCF-1 is involved in transcriptional regulation by altering chromatin structure [11]. The association of HCF-1 with VP16 mimics some of the interactions that occur when HCF-1 is at work during chromatin remodeling. The absolute need for HCF-1 was recently demonstrated in vivo by Narayanan et al. [13]. Three requirements need to be fulfilled in order for the VP16-HCF-1-Oct-1 complex to activate target genes: a) presence of the correct cis element flanking the core sequence, which determines what isoform of the VP16 protein can bind and activate [14, 15], b) ability of VP16 to discriminate related POU domain activators, as in the case of Oct-1 and Oct-2 [16], and c) association of VP16 with HCF-1 and not HCF-2, a related protein that can associate with VP16 and stabilize the Oct-1 complex, but does not promote transcription [17].
1.3 Additional determinants of specificity of transcription factors.

Post-translational modifications of transcription factors add another layer of regulation, allowing regulatory proteins to interact and control gene expression only in particular cellular environments. This ‘context-specific modification code’ has been proposed to explain the biological specificity of R1R2R3 MYB proteins in animals [18]. In vertebrates, three MYB proteins are expressed: A-MYB, B-MYB and c-MYB. All three of them can activate a mim-1 gene reporter construct in transient assays but only c-MYB activates the endogenous gene [18]. While A-MYB is essential for male germ cell and mammary gland development in mice [19] and B-MYB is involved in cell proliferation during mouse embryogenesis [20] and its cell cycle regulation requires its degradation by ubiquitination [21], c-MYB is involved in the regulation of hematopoiesis in vertebrates [22]. In addition to its MYB domain, c-MYB also has activation and negative regulatory domains, which are interdependent for transcriptional control [23]. Furthermore, c-MYB has been shown to interact with C/EBPβ and cooperatively activate the mim-1 gene, which has binding sites for both transcription factors separated by about 80 nucleotides, and which form a loop when the two proteins interact [24]. Chimeras of c-MYB and A-MYB revealed that the target gene specificity is given by small subdomains within the activation domains of these proteins [25]. It is also known that the activation domain of c-MYB is the target of many post-translational modifications [18]. Thus, the specificity is very likely to be the combined result of interactions with other factors through the subdomains, and the post-translational modifications. This allows for
a wide range of possibilities as the specificity of this transcription factor will change depending on what conditions are met in any given cellular environment [18, 25].

1.4 The flavonoid biosynthetic pathway: A model system to understand combinatorial control of gene transcription in plants.

The flavonoid biosynthetic pathway is one of the best characterized secondary metabolic pathways in plants (recently reviewed in Grotewold 2006 [26]). The origin of the flavonoid pigments pathway precedes the split between monocots and dicots as all flowering plants accumulate these compounds, albeit with some chemical variations depending on the species. Both the structural and the regulatory genes of the pathway have been evolutionarily conserved in all plants that have been examined to date. Flavonoid pigments can function as defense against pathogens, protection against UV light, fertility determinants, or as attractants for pollinating insects [27, 28]. There is also a substantial amount of evidence that points to these compounds playing a role in the plant stress response (reviewed in [29]), in development, and as signal molecules (reviewed in Taylor and Grotewold, 2005 [30]).

The most common flavonoid-derived pigments are the anthocyanins (Fig. 1.1), which range from red to blue in the flowers, fruits, and/or vegetative tissues of many higher plants. Due to their antioxidant, antiproliferative, and anti-inflammatory effects, anthocyanins and flavonols are gaining importance as potential therapeutic agents in a wide variety of human diseases including chronic inflammatory disease, coronary artery disease and cancer (reviewed in [31, 32, 33]). In maize, in addition to anthocyanins which
accumulate in the aleurone of the kernel as well as in the vegetative and root tissues, 3-deoxy flavonoid compounds accumulate in the pericarp of the kernel, the cob glumes, and other few floral tissues as brick red pigments denominated phlobaphenes. Thus the flavonoid pathway in maize is comprised of these two main branches (Fig 1.2).

Flavonoid pigments have also been an excellent genetic tool, since they are easy to score for, and failure to synthesize them is not lethal. As a result, a wealth of genetic data has been accumulated for genes in this pathway, not only in maize, but also in petunia, snapdragon and Arabidopsis. Many structural and regulatory mutants are available in these and other species. This allows the opportunity to use genetics extensively and to conduct in vivo experiments very easily, an advantage that is not easily found in animal systems. In addition, the flavonoid pathway is an excellent system to address transcriptional regulation as it provides an array of co-regulated target genes one could study.

1.5 Regulatory genes of the flavonoid biosynthetic pathway.

Two separate R2R3 MYB-domain proteins regulate each branch of the flavonoid pathway in maize: C1/PL1 regulates the anthocyanin branch [34, 35] while P1 regulates the accumulation of phlobaphenes [36]. In order to activate the genes that lead to the accumulation of anthocyanins, C1/PL1 needs to interact with a bHLH cofactor belonging to the R/B family. In contrast, P1 activates a subset of the same genes (c2, chi1, and al) independently of R/B [37, 36, 38].
1.5.1 The MYB domain factors.

MYB transcription factors may have one, two (R2R3), three (R1R2R3) or more MYB repeats arranged head to tail followed by a C-terminal domain that may contain an activation domain as in C1, which includes an acidic activation domain [39]. P1 also has a C-terminal region where a putative acidic activation domain is located [40]. Although plants have MYB domain proteins of the R1R2R3 class [41, 42], the great majority of MYB proteins belong to the R2R3 class. The two-repeat MYB proteins are the largest group in the MYB family of transcriptions factors. This dramatic expansion in the plant kingdom suggests that they are involved in plant specific functions [43, 44].

The maize \textit{C1/PL1} genes differ in their expression pattern as \textit{C1} is only expressed in the kernel (aleurone and embryo), and at low levels in the husk [45], while \textit{PL1} is expressed in the plant body. The petunia gene \textit{AN2} is the orthologue of \textit{C1/PL1} and it is expressed in the flower except in the anthers where a paralogous gene, \textit{AN4}, is expressed [46, 47]. In snapdragon, three different MYB genes regulate anthocyanins with differential expression throughout the flower: \textit{ROSEA1}, \textit{ROSEA2}, and \textit{VENOSA} [48]. In \textit{Arabidopsis}, \textit{PRODUCTION OF ANTHOCYANIN PIGMENT 1} (PAP1) AND PAP2 have been shown to activate early as well as late genes in the pathway when expressed constitutively, resulting in the accumulation of anthocyanins throughout the plant [49]. TT2, another \textit{Arabidopsis} MYB protein, is expressed only in seeds and regulates the \textit{DFR}, \textit{LDOX}, \textit{BAN}, and \textit{TT12} genes involved in the production of tannins (proanthocyanidins: a branch of the anthocyanin pathway), but it cannot activate the early genes of the pathway [50]. As shown in table 1.1, many other MYB domain regulators of
anthocyanin biosynthesis in plants have been identified, including one in black spruce, which indicates that the regulation of this pathway is also conserved in gymnosperms [51]. Even though all of the above MYB genes are positive regulators of the pathway, other MYB genes have been shown to repress transcription of this pathway as in the case of FaMYB1 which has been shown to repress AS and FLS in red ripe strawberry fruits [52].

All of the above orthologous MYB domain proteins are highly identical in the MYB domain but their C-terminal regions share very little if any similarity. This is typical of MYB genes of which only a few have functionality assigned to their C-terminals. As mentioned above, all of the C1 orthologs are R2R3 MYB domain proteins, but interestingly a single MYB repeat (R3) gene, MYB-P1 isolated from Perilla frutescens has been shown to activate the DFR gene in yeast [53].

1.5.2 The bHLH domain factors.

Proteins containing bHLH motifs in plants are numerous as exemplified in Arabidopsis, which has 131-147 proteins that belong to this family (reviewed in [54, 55, 56, 57]). Co-factors belonging to this family of transcription factors are required for the activation of the genes in the anthocyanin branch of the pathway. In maize the bHLH genes regulating the anthocyanin pathway belong to the R/B family. Similar to their MYB partners, R and B are differentially expressed in the plant. A third bHLH protein involved in the anthocyanin pathway has been identified in maize, but unlike the R/B proteins, INTENSIFIER 1 (In1) negatively regulates the pathway [58]. In petunia, two bHLH proteins have been identified as regulators of the anthocyanin pathway: JAF13 which is
the orthologue of R and interacts with AN2, and AN1 another activator that seems to be under the control of AN2 and JAF13, and that is also able to interact with AN2 [59]. In spite of being an activator, AN1 shares more identity with the maize IN1 bHLH repressor. AN1 activates DFR, and it plays an important role in the control of vacuolar pH as well seed coat cell morphology [59]. Another bHLH protein that is closely related to IN1 but that acts as an activator is TT8 in Arabidopsis. TT8 interacts with TT2 to activate the “late” genes in the branch of the pathway that leads to proanthocyanidin synthesis. TT8 is expressed in seed coats as well as in young seedlings, so it may interact with PAP1 during early developmental stages. MYC1 is another closely related bHLH protein in Arabidopsis but little is known about its function [60]. Two other proteins closely related to R in Arabidopsis are GL3 and EGL3 which interact with the MYB GL1 to regulate trichome formation in addition to anthocyanin biosynthesis [61]. DELILA is the bHLH responsible for flower pigmentation in snapdragon [62], and it is most closely related to JAF13. Related bHLH proteins that interact with a MYB partner have been identified in other species as well (Table 1.1).

1.5.2.1 The MYB interacting region.

The N-terminal region in the maize B protein that has been shown to be necessary and sufficient for the interaction with C1 [37] is also present in the bHLH proteins mentioned above. Interestingly, this region is only present in a small group of bHLH proteins (twelve bHLH proteins in Arabidopsis), and it has not been found in any other type of proteins in non-plant genomes (Braun and Grotewold, unpublished). In fact, analysis of the AN1 mutant allele G621 has shown that a truncated protein that lacks the
bHLH region but that retains the N-terminal region is still able to significantly activate the *DFR* gene [59]. Deletion of the first 115 amino acids in *P. frutescens* MYC-GP results in a truncated protein that is no longer able to increase the levels of anthocyanins in transgenic tobacco plants [63]. We are tentatively calling this N-terminal region the “MYB interacting region” or MIR. GL3 and EGL3 in *Arabidopsis* also contain a MIR region that is sufficient for their interactions with the MYB domain protein GL1 [64, 61].

### 1.5.2.2 The bHLH domain.

The function of the bHLH domain in these proteins remains elusive. It has been shown in transient expression assays that the bHLH region in R/B can be deleted and still obtain moderate activation of the target genes [65, 37]. The AN1 G6121 mutant that lacks the bHLH domain as well as AN1 mutants with insertions in the bHLH region are dramatically impaired in their ability to control vacuolar pH and seed coat morphogenesis while they exhibit modest activation of *DFR* [59]. It is unlikely that this region of the protein is completely dispensable or else it would not have been well conserved. Another possible function for the bHLH is the formation of hetero or homodimers. It has been shown that R can form dimers through the C-terminal region, which includes the bHLH domain; however, R homodimerizes through the C-terminal region downstream from the bHLH (Feller et al. Manuscript submitted) indicating that dimerization does not require the bHLH either. The same C-terminal region of GL3 and EGL3 allows them to homo and heterodimerize [64, 61].
1.5.3 A new class of WD repeat proteins regulates anthocyanin biosynthesis.

WD repeat proteins are a large family of proteins that contain tandem repeats of 40-43 amino acids, with a GH dipeptide 11-24 residues from the N-terminus, and punctuated with WD (Trp-Asp) [66]. These proteins are classified into different groups according to the number of repeats present in the protein [67]. The structure of the WD repeat proteins has been shown to assume a β-propeller fold and it is thought that this fold allows these proteins to participate in protein-protein interactions [66].

A recently discovered group of WD repeat proteins characterized by the presence of five-repeats include proteins that participate in the regulation of the anthocyanin biosynthetic pathway in plants. This group is comprised of proteins present not only in plants but also in other eukaryotes [68, 69]. The first of these proteins to be reported as playing a regulatory role in the production of anthocyanins was the petunia AN11 protein [68]. Mutations in an11 result in no visible phenotype other than unpigmented flowers, even though normally it is expressed in all tissues. AN11 is not regulated by AN2 or AN1, as mutations in AN2 and AN1 did not affect the levels of transcription of AN11. Overexpression of an2, of maize C1 and R, and C1 alone, complement an an11 mutant in transient expression assays with the latter being a partial complementation [68, 46].

Ironically, the first regulatory protein with a WD repeat that was shown to control anthocyanin biosynthesis was the Arabidopsis TTG1, but because ttg1 mutants could be complemented by ectopic expression of maize R, it was thought that TTG1 encoded a bHLH [70]. TTG1 is actually a WD repeat protein that is expressed in all major organs of Arabidopsis [71], and ttg1 mutants exhibit pleiotropic effects: lack of anthocyanins and
tannins, no trichome formation, excess root hair formation, and no seed coat mucilage. There are at least two more closely related WD-repeat genes in *Arabidopsis* but it is still not clear if they have overlapping functions with TTG1 [68, 72, 73]. In maize the *pale aleurone color1* (*PAC1*) mutant exhibits a reduction in the mRNA of the structural genes *a1, bz1* and *c2* but not of the regulators *C1* and *B*. In transient assays *PAC1* mutants could not be complemented by *C1* and *B* [74]. Later *PAC1* was cloned and it was revealed that it encodes a WD repeat protein closely related to AN11 and TTG1 that was capable of complementing the *ttg1* mutation in *Arabidopsis*. *PAC1* transcripts are present in immature tassel but apparently the phenotype is only visible in the aleurone and the roots [73]. In maize there is at least one more gene related to *PAC1* although it does not seem to play any role in anthocyanin biosynthesis [73, 72]. It is clear that these proteins are acting in concert with the MYB and bHLH regulators rather than controlling their expression. This is supported by the fact that in *PAC1* mutants the levels of expression of the regulators is not affected [74].

The WD repeat proteins interact with the bHLH proteins through the region located between the MIR and the bHLH [61] (Oh and Grotewold unpublished observation). This acidic domain is one of the most divergent regions among R-like proteins. It has also been shown that TTG1 can interact with TT2 [75] and PAP1 (Norambuena and Grotewold unpublished observation) and based on the fact that the MYB domain of C1, the maize ortholog of PAP1, does not interact with PAC1 (Wang and Grotewold unpublished data), we postulate that the interaction with TTG1 occurs through the C-terminal region of TT2 and PAP1. These WD repeat proteins are expressed
in all tissues and homologs are also present in animals and other eukaryotes, all of which suggests that they are high in the regulatory hierarchy and are probably involved in many cellular processes [68]. The fact that *ANII* mutants do not exhibit pleiotropic phenotypes and that *PACI* mutants show only subtle pleiotropy could be due to the fact that these two species have been genetically manipulated for centuries by man and during this process these proteins may have lost the ability to control some traits, which may have led to an exclusive regulation of these functions by a duplicate gene.

One possible role for these proteins is to mediate interactions of the bHLH and/or MYB regulators with the degradation machinery of the cell. COP1 is an example of a WD repeat protein that targets transcription factors involved in photomorphogenesis to proteosome-mediated degradation [76]. Although COP1 has other functional domains, it is through the WD repeat domain that this protein interacts with the target transcription factors HY5 and HYH [77].

1.5.4 The MYB-bHLH-WD-repeat regulatory system.

The three best characterized anthocyanin biosynthesis regulatory networks to date are those present in maize, petunia, and *Arabidopsis*. It is clear that in all three systems the regulatory triad of MYB/bHLH/WD-repeat regulators have common ancestors and are in fact interchangeable: maize C1 and R can function in Petunia and tobacco [46, 70], MYC-RP/GP and FaMYB1 can activate and repress respectively in tobacco [63, 52], PAC1 can complement *ttg1* [73], among others. Interestingly this is not limited to the regulators; in fact, the structural genes of maize can complement *Arabidopsis* structural gene mutants in spite of their moderate sequence similarity [78]. In addition interactions
between proteins from heterologous systems have been shown in yeast two-hybrid assays: FaMYB1 interacts with petunia AN1 and JAF13 [52], TTG interacts with R (Oh and Grotewold, unpublished). This network is now being reported for other species as well as in the case of *Perilla frutescens* MYB P1/MYC-RP/PFWD [63, 53, 79]. It is clear that the anthocyanin biosynthetic pathway arose prior to the split of dicots and monocots and even before the split between angiosperms and gymnosperms, and that the regulators may have initially participated in other cellular processes as evidenced by TTG1 in *Arabidopsis*. Although a WD repeat component of the regulatory network has not been found in snapdragon where the MYB/bHLH regulation has been described (ROSEA1, 2/DELILA) [48], and only MYB or bHLH domain proteins have been identified in systems such as Gerbera hybrida [80], or strawberry (*Fragaria x ananasa*) [52], we suspect that the MYB/bHLH/WD regulatory triad will be found to be present in these species as well.

Another trend that is easily observable is that in snapdragon, petunia, as well as in the tannin branch of *Arabidopsis* the regulation of the early genes of the pathway (prior to DFR; see Fig. 1.1) is independent from the MYB/bHLH/WD repeat regulators described here. In maize and *Arabidopsis* (as evidenced by PAP1) the regulators activate transcription of the entire anthocyanin pathway and they even activate the phenylalanine ammonia lyase (PAL) gene [49, 81]. Interestingly, C1 and R are not able to activate the early genes when overexpressed in petunia suggesting that the promoters of these genes do not contain the sequences for these MYB/bHLH partners [46].
In *Arabidopsis* there are about 125 MYB genes [44], about a dozen MIR-bHLH encoding genes (Braun and Grotewold unpublished), and at least three closely related WD repeat proteins including TTG1 (and probably a handful more of more distantly related WD repeat proteins). These numbers suggest that the specificity of the WD repeat proteins is very broad followed by the bHLHs and the MYBs in that order. This is supported by the fact that the WD repeat homologs of TTG1 are found in animals where no anthocyanins or compounds of a similar nature are synthesized. In fact it seems that the most closely related human homolog of these proteins is involved in interacting with a 3a protein kinase that inactivates glycogen synthase by phosphorylating it [69].

Other regulatory proteins have been implicated in pigment biosynthesis which are not from the MYB, bHLH or WD repeat families. For example, ANTHOCYANINLESS2 (ANL2) is an *Arabidopsis* homeo-domain transcription factor of the HD-Glabra 2 type that appears to control accumulation of anthocyanins in the subepidermal tissues of seedlings and mature plants, as well as cellular organization of the primary root [82]. Also in *Arabidopsis*, mutations in *TT1* and *TTG2* result in normal accumulation of anthocyanins in the plant but fail to make condensed tannins in the seed coat [83, 84]. *TT1* is a WIP plant zinc finger, and *TTG2* is a WRKY zinc finger-like transcription factor containing two tandem WRKY repeats. *TTG2* seems to be downstream from *TT2* and *TTG1*. In spite of their impact in the synthesis of pigments, these transcription factors appear to be involved more directly in developmental processes.
1.6 Dissertation aims.

Given the many advantages of the flavonoid biosynthetic pathway, we have used this system to investigate the mechanisms of transcriptional regulation that control the production of flavonoids in maize. This dissertation is organized into three results chapters which address three specific questions.

In chapter two, we addressed the question of biological specificity. We investigated how the two MYB domain proteins that control the expression of the structural genes in the two branches of the maize pathway, C1 and P1, activate only their target genes in spite of the fact that they exhibit high levels of amino acid identity in their DNA binding domains (MYB domains), bind the same \textit{cis} DNA elements \textit{in vitro}, and even share a few common target genes. We determined that the interaction with the bHLH R enables C1 to activate the genes that lead to the accumulation of anthocyanins. Since P1 cannot interact with R, it is unable to activate many of the genes in that branch of the pathway. We made mutations in P1 to allow it to interact with R and found that a mutant, P1*, was able to activate the anthocyanin biosynthetic genes. Thus, we demonstrated that the regulatory specificity of these MYB transcription factors is given by their interactions with other proteins, making this another example of combinatorial transcriptional regulation.

In chapter three, we unveiled two mechanisms of action of the bHLH protein R. Using our P1* mutant we dissected out two major functions for R. One role of R is to contribute to the activation of its target genes through an additional \textit{cis} DNA-element which is present in all the anthocyanin biosynthetic genes. It is not yet clear if R binds
this element directly or through another recruited protein. The other role for R involves
the abolishment of an inhibitory effect on the C1 protein which prevents its function. We
attempted to understand this effect by generating an R-independent C1 mutant. Our
mutant was able to bind DNA independently of R but this ability was not enough to allow
C1 to activate any of its target genes. These results have led to a model for how the
transcription of the anthocyanin biosynthetic genes is regulated.

In chapter four we addressed the question of how R achieves its multiple roles by
interacting with other factors. We focused on a protein that specifically interacts with the
bHLH of R and is putatively involved in chromatin remodeling, which we have recently
discovered is a factor in the regulation of some of the genes in the pathway (Morohashi
and Grotewold, unpublished). This protein, which we have named RIF-1C, has an ENT
and an Agenet domain. In order to understand the mechanism of action of RIF-1C, we
analyzed mutations in a homologous gene in Arabidopsis (ACK1) and found that it
causes a seed morphology phenotype.
Figure 1.1: General chemical structure of flavonoids.

Flavonoids have a three ring structure (A, B, and C). Rings A and B are benzene rings while ring C is a pyrane. Modifications to this structure by substitutions at the R positions determines the classification for a given flavonoid.
Figure 1.2: Maize flavonoid biosynthetic pathway.

Structural genes are shown with the traditional maize nomenclature and the enzyme name in parenthesis: chalcone synthase (CHS), chalcone isomerase (CHI), dehydroflavonol reductase (DFR), flavonone 3 hydroxylase (F3H), Leucoanthocyanidin dioxygenase (LDOX), UDP-glucose flavonoid 3-o-glucosyl transferase (UGFT), and glutathione S-transferase (GST). The brackets indicate the branches regulated by either P1 or C1 + R.
<table>
<thead>
<tr>
<th>Species</th>
<th>Regulators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species</strong></td>
<td>MYB</td>
</tr>
<tr>
<td>Maize</td>
<td>C1, PL1 [34, 35]</td>
</tr>
<tr>
<td>Petunia</td>
<td>AN2, AN4 [46]</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>PAP1, PAP2,</td>
</tr>
<tr>
<td>SnapDragon</td>
<td>ROSEA1, ROSEA2,</td>
</tr>
<tr>
<td>Perilla frutescens</td>
<td>MYB-P1 [53]</td>
</tr>
<tr>
<td>Gerbera hybrida</td>
<td></td>
</tr>
<tr>
<td>Rice</td>
<td>OsC1 [88]</td>
</tr>
<tr>
<td>Ipomoea purpurea</td>
<td>IpMYB1 [90]</td>
</tr>
<tr>
<td>Pepper</td>
<td>A [91]</td>
</tr>
<tr>
<td>Kyoho grape</td>
<td>VIMYBA [92]</td>
</tr>
<tr>
<td>Black Spruce</td>
<td>MBF1 [51]</td>
</tr>
<tr>
<td>Strawberry</td>
<td>FaMYB1 [52]</td>
</tr>
<tr>
<td>Tomato</td>
<td>AN1 [93]</td>
</tr>
<tr>
<td>Sweet Potato</td>
<td></td>
</tr>
<tr>
<td>Cotton</td>
<td>IbMADS10 (MADS) [94]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Table 1.1 Identified regulators of the anthocyanin biosynthetic pathway identified in plants.</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 2
IDENTIFICATION OF THE RESIDUES IN THE MYB DOMAIN OF MAIZE C1
THAT SPECIFY THE INTERACTION WITH THE bHLH COFACTOR R

2.1 Introduction.

Combinatorial interactions between transcription factors are of central importance to regulation of gene expression in eukaryotes. These interactions can either modulate transcription factor activity or contribute to the biological specificity of factors with very similar DNA-interaction motifs. Elucidation of the mechanisms by which proteins with very similar DNA-binding domains achieve regulatory specificity remains a fundamental question in biology today.

Proteins containing the MYB-homologous DNA-binding domain are widespread in eukaryotes (reviewed in [96, 97]). The vertebrate c-MYB gene plays an essential regulatory role in the proliferation and differentiation of hematopoietic cells. Besides c-MYB, at least two other MYB-related genes (A-MYB and B-MYB) are present in vertebrates [98]. The products of these genes have MYB domains, each consisting of three head-to-tail MYB motifs (R1, R2 and R3). Oncogenic versions of c-MYB, such as v-MYB, contain only R2 and R3, as do hundreds of plant MYB-domain proteins [99]. MYB domains formed by the R2 and R3 MYB motifs bind DNA. Each MYB motif
contains three $\alpha$-helices, and the third helix of each MYB motif makes sequence-specific DNA contacts. The second and third helices of each MYB motif form a helix-turn-helix structure when bound to DNA, similar to motifs found in the $\lambda$ repressor and in homeo domains [100]. In addition to their well-established roles in DNA binding, MYB domains are also emerging as important protein-protein interaction motifs. These MYB-domain-mediated protein-protein interactions play key roles in the biological specificity of the corresponding factors [37,101-107]. However, the mechanisms by which protein-protein interactions contribute to the regulatory specificity of MYB-domain proteins are poorly understood.

The control of flavonoid biosynthesis by the maize R2R3 MYB-domain proteins C1 and P1 is an excellent system to investigate how regulatory specificity by MYB-domain proteins is achieved through combinatorial interactions with additional regulatory proteins. Anthocyanin accumulation in maize is controlled by two classes of regulatory proteins acting in concert: those with a MYB domain (C1 or P1, two closely-related homologs [35]), and those with a basic helix-loop-helix (bHLH) domain (R or B, members of the R/B family [108]). Extensive studies have shown that the MYB-homologous C1 or P1 genes (do not confuse P1 with P1) require a member of the bHLH-containing R or B gene family to activate transcription of the anthocyanin biosynthetic genes (reviewed in [109]). Indeed, the R/B- and C1-encoded proteins physically interact, and this interaction is mediated by the MYB domain of C1 and the N-terminal region of B [37]. The maize P1 gene controls the accumulation of 3-deoxy flavonoids and red phlobaphene pigments by activating a subset of the anthocyanin biosynthetic genes.
controlled by C1 and R/B. P1 and C1 activate the expression of some common genes in the flavonoid biosynthetic pathway such as A1, by interacting with different affinities to identical cis-acting regulatory elements in the A1 gene promoter [36,38]. In contrast, C1, but not P1, binds and activates transcription of the Bz1 gene, specific for anthocyanin biosynthesis [36,38]. Furthermore, the ability of P1 to activate gene expression is independent of the R/B co-activators, despite the fact that the MYB domains of P and C1 are over 70% identical [36]. Thus, the interaction of C1 or PL with R or B is very specific, providing these two MYB proteins with unique regulatory activities.

Here, we have used chimeric MYB domains of P and C1 to determine which residues specify the interaction of C1 with the bHLH co-activator R. We show that residues in the first helix of the R3 MYB repeat of C1 are sufficient for the specificity of this interaction. Replacement of four solvent-exposed residues in the MYB domain of P1 with the corresponding residues in the MYB domain of C1 is sufficient to transfer the interaction with R to P1 in yeast two-hybrid experiments. The replacement of six residues in the R3 MYB repeat of P1 for the corresponding residues of C1 allows R to enhance the regulatory activity of P1 in maize cells. In addition, when P1 is altered in this way, it can now activate the Bz1 promoter, normally regulated only by C1+R. Together, these findings identify for the first time amino acids residues that allow MYB-domains to act as efficient protein-protein interaction motifs, which help confer unique regulatory specificity to closely related transcription factors containing MYB domains.
2.2 Materials and methods.

2.2.1 Plasmids used in transient expression experiments.

As previously described, all P1 and C1 plant expression vectors include the CaMV 35S promoter, the tobacco TMV Ω' leader and maize first Adhl-S intron in the 5'UTR, and potato proteinase II (pinII) termination signal. Previously described plasmids [38,110] include pBz1Luc, containing 2.3 kb of the Bz1 promoter and intron upstream of luciferase; pA1Luc, containing 1.4 kb of the A1 promoter and the Adhl-S intron; PPP (35S::P-cDNA1) corresponding to plasmid pPHP1962; CCC (35S::C1) (pPHP665); PPC (35S::P^MYB-C1^act) (pPHP4884) and R (35S::R) (pPHP471). 35S::BAR (pPHP611) was used for normalizing the concentration of 35S sequences delivered in each bombardment and was also previously described [111]. Ubi::GUS (PHP3953) was used to normalize the efficiency of each bombardment [112]. Standard site-directed mutagenesis procedures were used to introduce restriction sites or modify the amino acid sequence in the following constructs: CCP (35S::C1^MYB-P1^act) (pPHP4885) was obtained by introducing a translationally silent SnaBI restriction site at the 3’ end of the MYB region at Y119 in P1 and blunt-end ligating the non-MYB region of P1 to the MYB region of C1 at a PvuII site at G124. CPP (35S::C1^R2-PMYB^R3-P1^act) (pPHP6186), was obtained by introducing a silent AatII restriction site between the R1 and R2 MYB motifs at P63 in C1 (Fig. 2A) and ligating to the AatII site in P at D64. PCP (35S::P^R2-C1^R3-P^act) (pPHP6194 or pPHP7663) was obtained by fusing P1^R2 to C1^R3 using the AatII sites described in CPP and fusing C1^R3 to the non-MYB region of P1 using the SnaBl/PvuII sites described in CCP. CP^C:65-84-P (35S::C1^R2-C1^R3h1-P1^R3h2-P1^act) (pPHP7726) was
obtained by introducing a silent BstEII restriction sites at the turn between the first helix of the R3 MYB motif and the second helix of the R3 MYB motif at L85 in C1 and P1. The upstream C1 regions were fused to the downstream P1 regions at the BstEII sites to create the R3 MYB motif chimera.

2.2.2 Plasmids used in yeast two-hybrid experiments.

The MYB domains of wild type C1 and P1 as well as mutant P1 MYB domains were generated by PCR and then cloned into the pAD-GAL4 vector (Stratagene, CA) as EcoRI/SalI fragments. The long P1 cDNA [113] was used to obtain the wild type and mutant P1 MYB domain (P1\textsuperscript{MYB}). Wild type P1\textsuperscript{MYB} was generated using primers p5pAD, which includes the first 7 amino acids of P1 with EcoRI and NcoI sites in the 5’ end, and p3pET, which includes amino acids 114-116 of P1 [113] with SnaBI, SalI, and BamHI sites in the 5’ end. Mutant P1\textsuperscript{MYB} were generated by ligating 2 independent PCR fragments at an AgeI site which was engineered as a silent mutation at amino acids 87-88 from P1, changing these codons from AAC AGG to AAC CGG [113], and then amplifying the ligation product with p5pAD and p3pET. All of the resulting PCR products were subsequently cloned into pTAdvantage (Clontech, CA) and sequenced. Inserts were digested with EcoRI and SalI and cloned into pAD-GAL4. C1 MYB domains (C1\textsuperscript{MYB}) were generated from plasmid pPHP687 [36]. Wild type C1 MYB domain was generated using primers C1N1 which corresponds to amino acids 2-8 with an XhoI and a EcoRI at the 5’ end, and C13pET which corresponds to amino acids 113-118 with SnaBI, SalI, and BamHI sites at the 5’ end. R\textsuperscript{1-252} was generated by PCR using plasmid DNA pPHP687 as template and primers LcN1, which corresponds to amino
acids 1-7 with a XhoI site at the 5’ end and an EcoRI and a GCG codon inserted between amino acids 1 and 2, and LeC1, which corresponds to amino acids 142-151 with a BamHI site at the 5’ end and a SalI site resulting from a mutation that changes amino acid 148 from V to D. After sequencing, the insert was cloned as an EcoRI/SalI fragment into pBD-GAL4 (Stratagene, CA).

2.2.3 Microprojectile bombardment and gene expression experiments.

Bombardment conditions of suspension cells and transient expression assays for luciferase and GUS were performed essentially as previously described [36]. For each microprojectile preparation, the mass of DNA was adjusted to 10 µg with 35::BAR [111] to equalize the amount of 35S promoter in each bombardment. One µg of each of the regulators and 3 µg of reporter plasmid (pA1Luc or pBz1Luc) were used in each bombardment. To normalize luciferase activity to GUS activity, 3 µg of UBI::GUS was included in every bombardment. Each treatment was done in triplicate and entire experiments were repeated at least twice. The assays for luciferase and GUS, and the normalization of the data were done as described [36]. Data are expressed as the ratio of arbitrary light units (luciferase) to arbitrary units of fluorescence (GUS).

2.2.4 Yeast two-hybrid experiments.

The plasmid containing R1-252 in the pBD-GAL4 (TRP+) vector, and each of the constructs containing the mutant MYB domains of P1 or C1 cloned into pAD-GAL4 (LEU+) vector were cotransformed into yeast strain PJ69.4a with the genotype Mata trp1-901 leu2-3, 112 ura3-52 his3-200 gal 4Δ gal 80Δ LYS2::GAL1-HIS3 GAL2-ADE2
met::GAL7-lacz [114] which is and plated on SC-LEU-TRP medium. Colonies were then screened for growth on SC-LEU-TRP, SC-LEU-TRP-HIS, and SC-LEU-TRP-HIS-ADE.

2.3 Results.

2.3.1 Regulatory activity of chimeric MYB domains.

To identify regions in the MYB domain of C1 required for the specificity of the interaction with R, chimeras between P1 and C1 were generated. Transcriptional activation by P1/C1 chimeras was assayed by transient expression in maize callus cells in the presence and absence of R. Two luciferase reporter constructs containing promoters from different flavonoid biosynthetic promoters were used: A1 (activated by C1+R and P1) and Bz1 (activated by C1+R, but not by P1). In Figure 2.1, chimeric proteins are named according to the origin of the R2, R3 and C-terminal regions. For example, PCP corresponds to a protein containing the R2 MYB motif of P1, the R3 MYB motif from C1 and the C-terminal region of P1. As previously shown [36], P1 (PPP, Fig. 2.1) activates transcription of the A1 promoter independently of R, but fails to activate Bz1, either with or without R. C1 (CCC, Fig. 1) does not activate A1 or Bz1 expression alone, but the presence of R makes C1 a potent activator of these two promoters [36]. A chimeric protein containing the MYB domain of P1 and the C-terminal region of C1 (PPC, Fig. 2.1) activates the A1 promoter very poorly, albeit in an R-independent fashion, and fails to activate Bz1, in the presence or absence of R. The chimera consisting of the MYB domain of C1 fused to the C-terminal non-MYB region of P1 (CCP, Fig. 2.1) is
inactive on either \textit{A1} or \textit{Bz1} in the absence of R, but with R activates both of these promoters. These results demonstrate that, as previously shown with B [37], the MYB domain of C1 is the region that mediates the interaction with R.

To further identify the region within the MYB domain of C1 that makes transcription by C1 R-dependent, chimeras between the MYB domains of C1 and P1 were analyzed. A chimeric protein in which the R2 MYB repeat of P1 was replaced by the corresponding region of C1 (CPP, Fig. 2.1) activates the \textit{A1} promoter in the absence of R. Unlike P1, CPP also weakly activates the \textit{Bz1} promoter (compare PPP and CPP, Fig. 2.1), although the activity of CPP was not significantly increased by the presence of R. In contrast, a chimeric factor containing the R2 MYB repeat of P1 fused to the R3 MYB repeat of C1 (PCP) was unable to effectively activate \textit{A1} or \textit{Bz1} in the absence of R. However, the presence of R increased transcriptional activation over 40 fold on both the \textit{A1} and the \textit{Bz1} promoters (Fig. 2.1).

The above results suggest that either the R3 repeat, the linker between the R2-R3 repeats (residues 63-66, Fig. 2.2A) or both were mediating the interaction with R. The linker region plays an important role in the DNA-binding activity of MYB domains [115, 116]. Thus, it was possible that the different linker regions in P1 and C1 influenced their independence or dependence on R for transcriptional activation, respectively. To test this hypothesis, the PCP construct used in the experiment shown in Figure 2.1, which contains the amino acid sequence linker characteristic of C1 (PNIR, see Materials and methods), was compared to a construct identical to PCP but with the amino acid sequence of the linker region corresponding to P1 (ADVK). The PCP protein with the ADVK
linker region activated \( Al \) 10.3 +/- 1 fold without R, and 50.3 +/- 12.4 fold with R, and \( Bz1 \) 1.5 +/- 1 fold without R, and 73.8 +/- 12.2 fold with R. These results were very similar to those shown for PCP in Figure 2.1, suggesting that the linker region between R2 and R3 does not contribute to R-dependent or R-independent transcription. This indicates that only the R3 MYB repeat of C1 (residues 67-115, Fig. 2.2A) is responsible for the inability of C1 to activate transcription on its own, and that the same R3 MYB repeat mediates the specificity of the interaction with R, which is required as a cofactor for activation.

To determine which residues in the R3 MYB repeat of C1 are responsible for the functional dependence of C1 on R, residues 65 to 84 in the CPP chimera were replaced by the corresponding region of C1 (CPC:65-84P, Fig. 2.1). In contrast with CPP, activation by CPC:65-84P was increased by R, on both the \( Al \) as well as the \( Bz1 \) promoters. Thus, the specificity of the interaction between C1 and R is provided by a region of the R3 MYB repeat of C1 between residues 67 and 84 (Fig. 2.2A).

Interestingly, PPP, CPP and CPC:65-84P (in contrast to CCC, CCP or PCP) can each activate the \( Al \) promoter without R. These results indicate that a region of the MYB domain of P1 between amino acids 85 and the end of the MYB domain allows transcription activation in an R-independent fashion. Together, these data show that the R3 MYB repeat of C1 (residues 67-115) contains a sequence that allows C1 to interact with R (within residues 67-84), as well as a region that makes C1 transcription R-dependent (within residues 84-115), suggesting that these two activities map to separate regions in R3.
A comparison of the MYB domain of C1 with the MYB domains of two other proteins that interact with R, and that show R-dependent activity reveals a high level of conservation in the R3 MYB repeat, further limiting the potential functionally relevant residues (Fig. 2.2A). The *Arabidopsis* GL1 protein interacts with R in co-precipitation [117] and yeast two-hybrid experiments [113], and a model has been proposed in which the interactions between GL1 (and similar proteins) with bHLH co-factors is essential for trichome and root hair formation [114]. The Petunia AN2 protein is the C1 ortholog that also requires a bHLH protein for flower pigmentation [109, 115] and physically interacts with R in yeast two-hybrid experiments (Kroon, Koes, Grotewold and Mol, unpublished results).

### 2.3.2 Four residues are sufficient to transfer the interaction with R from C1 to P.

To determine whether R-dependent activation of transcription by the P1 and C1 chimeras tested in figure 2.1 reflect the ability of these proteins to physically interact, we conducted yeast two-hybrid experiments with the P/C1 chimeric MYB domains fused to the Gal4 activation domain (Gal4\(^\text{AD}\)). If these proteins were able to interact with a protein containing the first 252 amino acids of R fused to the Gal4 DNA-binding domain (R\(^{1-252}\)-Gal4\(^\text{DBD}\)), this would result in the activation of two selectable markers (HIS3 and ADE3) driven by Gal4 binding sites, thereby conferring growth in synthetic media lacking adenine and histidine (see Materials and methods). As previously shown with B [37], the N-terminal region of R (Gal4\(^\text{DBD}\)-R\(^{1-252}\)) physically interacts with the MYB domain of C1 (Gal4\(^\text{AD}\)-C1\(^\text{MYB}\), Fig. 2.2B1). Under similar conditions, however, the MYB domain of P1 does not interact with R (Gal4\(^\text{AD}\)-P1\(^\text{MYB}\), Fig. 2.2B2), thus providing direct
evidence of the specificity of the interaction of related MYB domains with R. As deduced from the transient expression experiments, the PC, but not the CP, chimeric MYB domain mediates interaction with R (compare Gal4^{AD-PC^{MYB}} and Gal4^{AD-CP^{MYB}} in Fig. 2.2B3 and 2.2B4). These results show that the region of C1 that specifically interacts with R is the same as that required for R-dependent activity.

Based on the transient expression experiments (Fig. 2.1), the 67-84 region of the R3 MYB repeat of C1 is essential for the specificity of the interaction with R. To further determine which of the six residue differences between P1 and C1 in this region (Fig. 2.2A) are important for the interaction with R, two residues in C1 (Y72 and D73) were simultaneously changed to the corresponding residues in P1 (K and E). The resulting C1 mutant MYB domain was capable of interacting with R in yeast two-hybrid experiments (data not shown), indicating that those two residues are unnecessary for interaction with R.

To determine which residues in C1 confer specificity in the interaction with R, residues I77, K80, A83 and T84 in P1 (Fig. 2.2A) were replaced with the corresponding L, R, R and L residues present in C1. The P1 MYB domain with these four changes, when fused to the Gal4^{AD} (Gal4^{AD-P1^{MYB}\text{I77L,K80R,A83R,T84L}}), was capable of interacting with R like C1 (Fig. 2.2B). However, when each one of these residues in P1 was mutated independently or in combinations of two or three, no interaction with R was observed (see for example Gal4^{AD-P1^{MYB}\text{I77L,K80R,T84L}} in Fig. 2.2B). These results demonstrate that four amino acids in the MYB R3 motif of C1 are sufficient to confer upon the P1 MYB domain the ability to specifically interact with the cofactor R. Consistent with this, when
R83 in C1 was changed to the P1 residue A, or L84 in C1 to T, interaction between C1 and R was lost (data not shown), suggesting that R83 and L84 are necessary for C1 to interact with R.

2.3.3 P1 mutants interact with R in plant cells.

To investigate the regulatory activity in plant cells of P1 mutants able to interact with R in yeast, the I77L, K80R, A83R and T84L changes were introduced into the full-length P1 protein. Driven from the constitutive CaMV 35S promoter, P1\textsuperscript{177L,K80R,A83R,T84L} was assayed in transient expression experiments for activation of the A1 and Bz1 promoters. Surprisingly, no activation of A1 or Bz1 was observed, regardless of the presence or absence of R (Fig. 2.3). Since some of these factors show a dose-dependent response (Grotewold and Bowen, unpublished), different concentrations of 35S::P1\textsuperscript{177L,K80R,A83R,T84L} were tested, with identical negative results (not shown). A negative result like this could be indicative of a loss of DNA-binding activity caused by the four residue changes. To test for this, we expressed the P1\textsuperscript{177L,K80R,A83R,T84L} protein in a yeast strain containing a reporter gene controlled by the high-affinity P1-binding sites [118]. In this assay, P1\textsuperscript{177L,K80R,A83R,T84L} activates transcription as efficiently as P1 does (Matulnik and Grotewold, unpublished), suggesting that, at least in yeast cells, the mutant P1\textsuperscript{177L,K80R,A83R,T84L} binds DNA.

Although we can not rule out from our findings the possibility that the P1\textsuperscript{177L,K80R,A83R,T84L} protein is unstable in maize cells, another possibility is that other residues from C1 are required for transcription activation. There are five differences between P1 and C1 in the 84-115 region (Fig. 2.2A). The G94 and R95 residues in C1
were particularly interesting because they are conserved in other R-dependent MYB domains (Fig. 2.2A). In addition, the simultaneous change of G94S and R95H allowed a C1\textsuperscript{MYBG94S,R95H}-Gal4\textsuperscript{AD} chimeric protein to activate transcription in yeast from a promoter containing the previously-described high-affinity P1-binding sites, in contrast to C1\textsuperscript{MYB}-Gal4\textsuperscript{AD} which is inactive in this assay [119]. Thus, we investigated the effect of changing S94 to G and H95 to R in the context of P1\textsuperscript{I77L,K80R,A83R,T84L} on the activation of A1 and BzI in transient expression experiments in plant cells. 35S::P1\textsuperscript{*} (35S::P1\textsuperscript{*}) activated transcription of A1 independently of R, similar to P1 (Fig. 2.3). However, when R is co-bombarded, a significant enhancement of this activity was observed (Fig. 2.3), providing strong evidence that this mutant of P1 interacts with R in maize cells. The requirement for the six residue changes in plant cells relative to four in yeast could reflect that additional residues are necessary for interaction with R in plant cells, or that the two additional residues are required for R to mediate transcriptional activation. The first hypothesis is unlikely for several reasons. First, a C1 derivative with the G94S and R95H residue changes interacts with R in yeast two-hybrid experiments (not shown). Second, the C1\textsuperscript{MYBG94S,R95H}-Gal4\textsuperscript{AD} fusion has R-dependent transcriptional activity in maize cells like C1\textsuperscript{MYB}-Gal4\textsuperscript{AD} [119]. Finally, a P1\textsuperscript{MYBS94G,H95R}-Gal4\textsuperscript{AD} construct does not interact with R in yeast two-hybrid experiments (not shown).

Together, these findings demonstrate that four residues in R3 are responsible for the specificity of the interaction of C1 with R, and that in the context of P1 two additional residues need to be altered to enable R to enhance P1 transcriptional activation in plant cells.
2.3.4 Interaction with the cofactor R is necessary for transcriptional activation of \textit{Bz1}.

In contrast to \textit{A1}, which is activated by both P1 and C1+R, \textit{Bz1} is activated by C1+R, but not by P1 [36]. The P1* factor, which can activate \textit{A1} without R but interacts with R \textit{in vivo}, provides an unique opportunity to investigate the contribution of R to the activation of \textit{Bz1}. In the absence of R, P1* does not activate \textit{Bz1}, similar to P1 and to C1 (Fig. 2.3). However, in the presence of R, a dramatic activation of \textit{Bz1} is observed. This demonstrates that R is still required for \textit{Bz1} activation, even with a MYB-domain protein that can activate \textit{A1} without R.

2.4 Discussion.

In this study, we have used the independent regulation of two branches of maize flavonoid biosynthesis by the related MYB-domain transcription factors C1 and P1 to elucidate the participation of the MYB domain in co-activator-dependent transcription. We identified the residues in the MYB domains of C1 that specify the interaction with the bHLH co-activator R. By replacing residues in P1 with the corresponding amino acids present in C1, we transferred the interaction with R to P1, resulting in an activator with novel regulatory functions. Finally, we demonstrate a central role of R in the regulatory specificity of the MYB-domain proteins C1 and P1.

2.4.1 Specificity of the interaction between MYB domains and co-activators.

Despite the higher than 70% identity between the MYB domains of C1 and P1 [113], our findings demonstrate that only the MYB domain of C1 interacts with the N-
terminal region of the bHLH cofactor R. Transient expression and yeast two-hybrid experiments revealed that the R3 MYB repeat of C1 is responsible for the specificity of the interaction with R. While R3 is necessary for the interaction of C1 with R, it is probably not sufficient. Truncation analyses indicate that in yeast two-hybrid experiments, the last 13 amino acids of R2 are also required (not shown). This includes most of the DNA-recognition helix of R2, opening the possibility that the need for these R2 sequences involves correct folding of R3, exposing the right surface in R3 for R interaction. A *Perilla frutescens* MYB-domain with only an R3 interacts with Myc-rp, an R-like bHLH factor [53]. In addition, it has been recently proposed that the competitive effect of the WER (an R2R3 MYB-domain protein) and CPC (which only has an R3 MYB repeat [120]) proteins in *Arabidopsis* root epidermal cell patterning is mediated by the ability of these two proteins to interact with an as yet unidentified bHLH transcription factor [121]. These findings suggest that the R3 MYB repeat of R2R3 MYB domain proteins may provide a general surface for protein-protein interactions.

### 2.4.2 Four amino acid changes are sufficient to transfer the interaction with R from C1 to P1.

Our findings demonstrate that the L77, R80, R83 and L84 residues in C1 specify the interaction with R. Replacement of the corresponding residues in P1 for the residues present in C1 is sufficient to transfer the interaction with R from C1 to P1 in yeast two-hybrid experiments. A model of the MYB domain of C1 (Fig. 2.4A), based on the NMR structure of the R2R3 MYB domain of c-MYB [100,122], indicates that these four residues are solvent-exposed, providing a surface for the interaction with R. The smaller
residues found in the MYB domain of P1 (Fig. 2.4B) make the corresponding surface of P1 significantly different. All four residues appear to be necessary for the specificity of the interaction, as single, double or triple changes did not transfer the interaction with R to P1 (not shown). Consistent with our findings, the L77, R80, R83 and L84 residues are also conserved in the AN2 and GL1 proteins that physically interact with R and that require R for regulatory activity.

This surface may play an important role in other protein-protein interactions. The MYB domains of c-MYB and A-MYB, but not of B-MYB, were shown to interact with nucleolin in animal cells, and the solvent-exposed R161 residue present in c-MYB and A-MYB, but not B-MYB, is crucial for this interaction [105]. Strikingly, the R161 in c-MYB coincides in position with L84 in C1 (Fig. 2.2A), one of the key residues in the interaction of C1 with R. Thus, a variety of co-activators may recognize similar regions with distinct surface-exposed residues in MYB domains to modulate MYB protein activity.

### 2.4.3 R contributes to the regulatory specificity of MYB transcription factors.

R is absolutely essential for C1 to activate transcription of all the genes in the anthocyanin pathway, including *A1* and *Bz1*, while P1 activates transcription of a subset of the C1-regulated genes (including *A1* but not *Bz1*) independently of R [36]. Is the function of R to make C1 active, or does R contribute to the regulatory specificity of C1? The P1* protein provided us with unique tools to addresses these issues. Similar to P1, P1* activates transcription of *A1* independently of R (Table 2.1). However, R can interact with P1* enhancing its activity on the *A1* promoter. When tested on the *Bz1*
promoter, P1* does not activate transcription, like P1. However, very robust activation of BzI is observed by P1* in the presence of R. Thus, P1* has a regulatory specificity that is different from either P1 or C1 (Table 2.1). Because P1* does not activate BzI in the absence of R, we can conclude that the six residue changes in P1 do not allow P1* to interact with the BzI promoter in a productive manner. Rather, the ability of P1* to activate BzI is completely dependent on its interaction with R. These results suggest that R does not simply activate C1, but rather that it plays a key role in the regulatory specificity of C1.

Together, our results demonstrate that, although C1 and P1 have very similar DNA-binding specificity [38], their ability to control the accumulation of different pigments by activating distinct sets of target genes is given by the specific interaction of the MYB-domain of C1 with R. These findings are of particular significance given the very large number of R2R3 MYB transcription factors expressed in the higher plants which have very similar DNA-binding domains [43,123]. The regulatory specificity of these MYB factors might be largely provided by combinatorial interactions with other cellular factors, rather than by different DNA-binding preferences.
Figure 2.1: Activation of the A1 and Bz1 promoters.

Results of transient expression following co-bombardment of cultured maize BMS cells with different chimeras of P1 and C1 together with A1Luc (red) or Bz1Luc (blue) reporter constructs, in the absence (-) or presence (+) of a vector that expresses R from the constitutive CaMV 35S promoter. Sequences derived from P1 are shown in yellow with the helices characteristic of each of the two MYB repeats (R2 and R3) in orange. Sequences derived from C1 are shown in gray with the helices in green. The first two letters indicate the constitution of the MYB domain, the first indicating the constitution of R2, the second of R3; and the last indicates the origin of the C-terminal region containing the transcriptional activation motif. A UBI::GUS construct was included in every bombardment as a normalization control. Each treatment was done in triplicate, and the data were normalized for GUS activity as described [36]. The fold activation was calculated as the ratio between each particular treatment and the treatment with pA1Luc or pBz1Luc constructs without activator. The average values are shown and the error bars indicate the standard deviation of the samples [130].
Figure 2.2: MYB-domain sequences that contribute to the specificity of the interaction with the bHLH cofactor R.

A. Sequence comparison between the MYB domains of P1 [113], and other proteins shown to interact with R, including C1 [34], AN2 [47], and GL1 [124]. The position of the three α-helices that form each MYB repeat are marked, with helix 3 of each motif involved in DNA interaction. Residue numbers are based on the sequences of P1 and C1. Dark shading indicates identical residues, light shading indicates conservative changes. Residues focused on in this study are marked with asterisks. B. Yeast two-hybrid interactions of the MYB domains of C1, P1 or mutant versions of the MYB domain of P fused to the Gal4 activation domain (GAL4\text{AD}) with the N-terminal 252 amino acids of R fused to the Gal4 DNA-binding domain (GAL4\text{DBD}). The simultaneous change of the I77, K80, A83, T84 residues in P1 for the corresponding residues of C1 (Gal4\text{AD-}P1\_MYB177L, K80R, A83R, T84L) allow P1 to interact with R [130].
Figure 2.3: Transfer of the interaction with R from C1 to P1 \textit{in vivo}.

Results of transient expression following co-bombardment of cultured maize cells with P1, C1 and mutants of P1 together with A1Luc (clear bars) or Bz1Luc (dark bars) reporter constructs, in the absence (-) or presence (+) of 35S::R. All other experimental details are described in the legend of Fig. 1.1 [130].
The four amino acids in C1 (L77, R80, R83 and L84) sufficient to transfer the interaction with R from C1 to P1 are shown in red. The position of G94 and R95 could not be precisely determined, although the polar nature of R95 makes it a candidate for a surface-exposed residue [130].
<table>
<thead>
<tr>
<th></th>
<th>R interaction</th>
<th>A1 activation</th>
<th>Bz1 activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Yes</td>
<td>R-dependent</td>
<td>R-dependent</td>
</tr>
<tr>
<td>P1</td>
<td>No</td>
<td>R-independent</td>
<td>None</td>
</tr>
<tr>
<td>P1*</td>
<td>Yes</td>
<td>R-enhanced</td>
<td>R-dependent</td>
</tr>
</tbody>
</table>

Table 2.1: Dependence on R of C1, P1 and a P1 mutant for the activation of the A1 and Bz1 genes [130].
CHAPTER 3
DIFFERENT MECHANISMS PARTICIPATE IN THE R-DEPENDENT ACTIVITY OF THE R2R3 MYB TRANSCRIPTION FACTOR C1

3.1 Introduction.

Flowering plants express a large number of proteins containing the conserved R2R3 MYB DNA-binding domain. About 125 R2R3 Myb genes are present in the Arabidopsis genome [44], and many more are predicted to be expressed in maize and related monocots [43, 125]. Similar to other transcription factor families, the R2R3 MYB factors show exquisite regulatory specificity in vivo, while recognizing very similar DNA sequences in vitro [36, 38, 123, 126-128]. Thus, mechanisms other than discrimination between similar DNA-binding sites are at play in the control of specific sets of target genes by each R2R3 MYB transcription factor in vivo.

The regulation of flavonoid biosynthetic gene expression by the cooperation of R2R3 MYB and basic helix-loop-helix (bHLH) transcription factors provides one of the best-described examples of combinatorial gene regulation in plants [109, 129]. Anthocyanin accumulation in maize is controlled by two classes of regulatory proteins that act in concert: C1 or PL1, two closely related R2R3 MYB domain proteins [35], and R or B, which are members of the R/B family of bHLH-domain proteins [108].
Extensive genetic and molecular studies have shown that the \textit{C1} or \textit{Pl1} genes require a member of the bHLH-containing \textit{R} or \textit{B} gene family to activate transcription of the anthocyanin biosynthetic genes \cite{109}. The \textit{C1}- and \textit{R/B}-encoded proteins physically interact, and this interaction is mediated by the MYB domain of C1 and the N-terminal region of \textit{B} \cite{37} or \textit{R} \cite{130}.

The maize \textit{P1} gene, encoding another R2R3 MYB transcription factor, controls the accumulation of 3-deoxy flavonoids and red phlobaphene pigments by activating a subset of the anthocyanin biosynthetic genes controlled by C1 and R, without a need for a known bHLH partner. \textit{P1} and C1 activate the expression of some common genes in the flavonoid pathway such as \textit{A1}, and they interact with different affinities to the same \textit{cis}-acting regulatory elements in the \textit{A1} gene promoter \cite{36, 38}. The \textit{A1} promoter has a modular structure in which the proximal high-affinity \textit{P1} binding sites (\textit{haPBS}) and the distal low affinity \textit{P1}-binding sites (\textit{laPBS}) are separated by the ARE (\textit{Anthocyanin Regulatory Element}) \cite{131}. In transient expression experiments, these three elements contribute to the regulation of \textit{A1} by \textit{P1} or by C1+R \cite{36, 38, 132}. In addition, transposon insertions and mutations in the ARE differentially affect the \textit{in vivo} regulation of \textit{A1} by \textit{P1} or C1+R \cite{133}. Unlike \textit{P1}, C1 activates the transcription of the \textit{A2}, \textit{Bz1} and \textit{Bz2} genes, which are specific for the anthocyanin branch of the pathway \cite{36, 38, 131}. The \textit{A2}, \textit{Bz1} and \textit{Bz2} gene promoters are also modular \cite{131, 134}, containing ARE and C1-binding sites.

Even though the MYB domains of \textit{P1} and C1 are over 70\% identical \cite{113}, and they recognize very similar DNA sequences \cite{38}, only C1 has an absolute requirement
for the bHLH factor R to activate transcription of the anthocyanin biosynthetic genes (R-dependent transcription). In contrast, P1 controls gene expression independently of R (R-independent transcription). The substitution of six residues in the MYB domain of P1 with the corresponding residues from C1 generates the P1* (formerly P1*) protein (Fig. 3.1A), which, unlike P1, is able to physically interact with R [130]. Similar to P1, P1* activates A1 but not Bz1 in the absence of R, indicating that the DNA-binding properties of P1* have not been altered. Interestingly, however, in the presence of R, P1* mediates a robust activation of the Bz1 promoter [130].

The cooperation between bHLH and R2R3 MYB transcription factors is not limited to the regulation of flavonoid biosynthesis by C1- and R-like proteins. The GL1 R2R3 MYB protein interacts with the GL3 and EGL3 bHLH factors to regulate the accumulation of trichomes in Arabidopsis [61,64]. Similarly, the bHLH rd22BP1 and R2R3 MYB AtMYB2 Arabidopsis proteins cooperate for drought- and abscisic acid-regulated gene expression [135]. While the Arabidopsis genome contains more than 120 genes encoding bHLH proteins [54, 55, 56, 57], the factors that cooperate with R2R3 MYB proteins belong to a small subgroup of bHLH proteins that share a common motif in their N-termini [56], (Braun and Grotewold, unpublished). This motif corresponds to the region in R that interacts with C1 [37, 130]. These findings suggest a general mechanism of cooperation between R2R3 MYB proteins and bHLH factors in transcriptional regulation. How this cooperation contributes to the regulatory specificity of R2R3 MYB proteins is the subject of this study.
Herein, we have investigated the cooperation between C1 and R for the regulation of flavonoid biosynthetic genes. Using P1*, we uncovered two components for this synergy. One component, manifested by the R-enhanced activity of C1 and P1* on the Al promoter requires, in addition to the high-affinity P1-binding sites, cis-regulatory sequences within the ARE. The second component is manifested by the R-dependent activity of C1 on promoters containing only the hypBS. We generated a mutant of C1 (C1SH) that binds DNA with a higher affinity than C1 and comparable to P1. Using C1SH, we demonstrate that the R-dependent activity of C1 is not solely due to the intrinsic low DNA-binding affinity of C1, as the C1SH transcriptional activity continues to be R-dependent. The differential activity in yeast and maize cells of chimeras of C1 with the yeast Gal4 DNA-binding domain suggests that part of the function of R is to relieve C1 from an inhibitor. Together, our findings uncover two distinct and separable mechanisms by which R cooperates with C1 for transcriptional activity. In addition, our results provide a model to explain how R2R3 MYB factors with related DNA-binding preferences achieve regulatory specificity through combinatorial interactions with accessory factors.

3.2 Materials and methods.

3.2.1 Plasmids used in yeast experiments.

The P1 and C1 cDNAs [34, 113] were cloned under the control of the glyceraldehyde 3-phosphate dehydrogenase (GPD1) promoter in the YEplac112 vector [136]. A control vector was also constructed with the GPD1 promoter with no
downstream gene. The fusions of the MYB domains of P1 and C1 to the Gal4 activation domain (Gal4\textsuperscript{AD}) were constructed by synthesizing the MYB domains and the Gal4 activation domain by PCR and then ligating them at an engineered Not I site. The fusion was then cloned under the GPD1 promoter in YEplac112. The yeast expression constructs corresponding to the N-terminal region of R fused to the Gal4 DNA-binding domain (Gal4\textsuperscript{DBD}-R\textsuperscript{1-252}) was the previously described fusion [130], but with the yeast selection marker changed from TRP1 to LEU2 by replacing the Gal4\textsuperscript{AD} with the Gal4\textsuperscript{DBD}-R\textsuperscript{1-252} fusion in pAD-GAL4 (Stratagene).

The yeast strain used for the two-hybrid experiments was PJ69.4A, Mat a trp1–901 leu2–3, ura3–52 his3–200 gal 4D gal 80D LYS2::GAL1-HIS3 GAL2-ADE2 met::GAL7-lacz [114]. For the yeast one-hybrid experiments, a strain (YEG102) was constructed by transforming strain AY926 [137] with a PvuII linearized pTH plasmid [138] in which the double stranded oligonucleotide primer APB1X2: 5'-GATCGT GTA CCT ACC AAC CTT AAA CGT GTA CCT ACC AAC CTT AAA C-3' containing two copies of the haPBS was cloned upstream of a minimal GAL1 yeast promoter driving the expression of the HIS3 gene. Transformations were done using the method described in http://tto.trends.com.

3.2.2 Plasmids used in transient expression experiments.

As previously described, all P1 and C1 plant expression vectors include the CaMV 35S promoter, the TMV Ω’ leader, the first intron of maize Adh1-S in the 5’UTR, and the potato proteinase II (pinII) termination signal. Plasmid described previously include p35SP1, p35SC1, p35SR, p35SP1* (corresponding to
p35SP^{I77L,K80R,A83R,T84L,S94G,H95R} [130]), pBz1Luc, containing 2.3 kb of the Bz1 promoter and intron upstream of luciferase; and pA1Luc, containing 1.4 kb of the A1 promoter and the Adh1-S intron ([38, 130, 36]). The p35SC1SH construct was generated by PCR mutagenesis of p35SC1 using the QuickChange® XL kit (Stratagene). p35SBAR [111] was used for normalizing the concentration of the 35S promoter delivered in each bombardment. The p35SGal4DBD-C1C-term construct was previously described [139]. The pBz2Luc construct was also previously described [140] and was kindly provided by Dr. Bodeau. The pA2Luc construct was obtained from Dr. Lesnick [131]. The construct containing three copies of the haPBS present in the A1 gene upstream of the luciferase reporter [p(haPBS)3Luc] corresponds to the 3xAPB1-35S construct previously described [36]. The pGal4BSLuc reporter construct consists of four CGGAGTACTGTCCCTCCGAG motifs in tandem upstream of a minimal CaMV 35S promoter. The pA1315LLuc construct is identical to pA1Luc, except for the six base pairs insertion present in the a1315L allele, left as a consequence of the excision of Spm from the a1-m2-7991A::Spm-s allele [133]. pUbiGUS [130] was used to normalize the efficiency of each bombardment.

3.2.3 Microprojectile bombardment and gene expression experiments.

Bombardment conditions of maize Black Mexican Sweet (BMS) suspension cells and transient expression assays for luciferase and GUS were performed essentially as previously described [130]. For each microprojectile preparation, the mass of DNA was adjusted to 10 µg with p35SBAR [111] to equalize the amount of 35S promoter in each bombardment. One µg of each of the regulators and 3 µg of reporter plasmid were used
in each bombardment. To normalize luciferase activity to GUS activity, 3 µg of pUbiGUS was included in every bombardment. Each treatment was done in triplicate and entire experiments were repeated at least twice. The assays for luciferase and GUS, and the normalization of the data were done as described [36]. Data are expressed as the ratio of arbitrary luciferase light units to arbitrary GUS light units. Fold activation is calculated as the ratio between the Luc/GUS units of the reporter construct with transcriptional activator divided by the Luc/GUS ratio without the activator. Luciferase for a typical bombardment with just pA1Luc (no activator) gives between 1,000 and 5,000 units, and GUS (from pUbiGUS) between 150,000 and 700,000 units.

3.2.4 Microprojectile bombardment and gene expression experiments.

Bombardment conditions of maize Black Mexican Sweet (BMS) suspension cells and transient expression assays for luciferase and GUS were performed essentially as previously described [130]. For each microprojectile preparation, the mass of DNA was adjusted to 10 µg with p35SBAR [111] to equalize the amount of 35S promoter in each bombardment. One µg of each of the regulators and 3 µg of reporter plasmid were used in each bombardment. To normalize luciferase activity to GUS activity, 3 µg of pUbiGUS was included in every bombardment. Each treatment was done in triplicate and entire experiments were repeated at least twice. The assays for luciferase and GUS, and the normalization of the data were done as described [36]. Data are expressed as the ratio of arbitrary luciferase light units to arbitrary GUS light units. Fold activation is calculated as the ratio between the Luc/GUS units of the reporter construct with transcriptional activator divided by the Luc/GUS ratio without the activator. Luciferase
for a typical bombardment with just pA1Luc (no activator) gives between 1,000 and 5,000 units, and GUS (from pUbiGUS) between 150,000 and 700,000 units.

3.2.5 Expression and purification of proteins expressed in bacteria.

The plasmids for the expression of the MYB domains of P1, C1 and C1\textsuperscript{SH} in \textit{E. coli} were obtained by cloning into the pTYB2 vector (New England Biolabs). The MYB domains (residues 1-119) were synthesized by PCR adding an \textit{NdeI} restriction site at the \textit{N}-terminus and an \textit{XhoI} restriction site at the \textit{C}-terminus which adds a Cys residue at the \textit{N}-terminus of the protein splicing element Intein from \textit{S. cerevisiae}. For expression, \textit{E. coli} BL21 (DE3) PlyS cells bearing the corresponding plasmids were grown, induced and purified essentially as described [141], with the following modifications. After induction of a 1 L culture with 1 mM IPTG, the cells were harvested by centrifugation and stored at −80°C until further use. The cells were resuspended in 40 ml of resuspension buffer (20 mM Tris pH 8, 300 mM NaCl, 1 mM EDTA, 0.1% Triton X-100 and 1 mM PMSF) and passed twice through a French press. The cell lysate was centrifuged at 14,000 x \textit{g} for 20 minutes; the supernatant was filtered through two layers of Miracloth (Calbiochem). The chitin beads (New England Biolabs) were equilibrated in CB1 (20 mM Tris pH 8, 300 mM NaCl and 1 mM EDTA) and resuspended to give a 50 % slurry. Ten ml chitin bead slurry was added to the cell lysate supernatant and incubated for two and a half hours with rocking at 4°C. The beads were gently pelleted by centrifugation and the pellet was then re-suspended with 5 ml of CB2 (20 mM Tris pH 8, 500 mM NaCl, 1 mM EDTA and 0.1% TritonX-100), loaded onto a column and washed twice with the same buffer. The column was washed three additional times with CB3 (20 mM Tris pH 8, 1 M NaCl 1 mM
EDTA and 0.1% Triton X-100). The column was incubated in CB3 plus 50 mM DTT overnight in order to allow the self-cleavage of the intein and therefore allow elution of the target protein. The elution was then dialyzed against A-0 buffer (10 mM Tris pH 7.5; 50 mM NaCl; 1 mM DTT; 1 mM EDTA; 5 % glycerol) at 4°C and stored at –80°C until further use. Each wash and elution fraction was collected and analyzed by SDS-PAGE, followed by staining with Coomassie Brilliant Blue R-250.

3.2.6 Generation of C1 monoclonal antibodies and Western analyses.

Monoclonal antibodies were generated using affinity purified N_{10}His-C1 protein by the Antibody Facility at Cold Spring Harbor Laboratory. For Western analyses, PJ69.4a yeast cultures were grown in 1.5 ml of SC-Leu-Trp medium at 30°C to an OD_{600} 0.7. Cells were harvested by centrifugation at 14,000 rpm and washed once with 1.5 ml of distilled water. The pellet was then resuspended in 100 μl of SDS loading buffer (0.06 M Tris-HCl, pH 6.8, 10 % glycerol, 2 % w/v SDS, 5 % β-mercaptoethanol, 0.0025 % w/v bromophenol blue), and heated to 95°C for 5 min. The suspension was then centrifuged at 14,000 rpm, and 15 μl of extract was loaded and analyzed by 12 % SDS-PAGE. The gels were transferred to PVDF membrane by electrophoresis using the BioRad Mini Trans-Blot transfer cell in 39 mM glycine, 48 mM Tris-HCl and 20% methanol. The transfer was done at 4°C and 100V for 1 hour. Membranes were stained with Ponceau Red to verify transfer and then blocked in 1 % non-fat dried milk dissolved in 1X TBS (10 mM Tris-HCl pH 8.0, 150 mM NaCl) with 0.05 % Tween-20 overnight at 4°C. Following blocking, the membranes were incubated with the primary antibodies diluted
in 1 % non-fat dried milk dissolved in 1X TBS with 0.05 % Tween-20 for 2 hours at room temperature. The C1 monoclonal antibody C1 40/1-249 was used at 1:150 dilution. For normalization, the blot was probed with an antibody that recognizes the yeast DED1p protein at a dilution of 1:4000. After washing the blots three times for ten minutes in 1X TBS, blots were incubated in anti-rabbit IgG, horseradish peroxidase secondary antibody (Amersham) for DED1p and anti-mouse IgG, horseradish peroxidase (Amersham) for C1 40/1-249 diluted 1:2000 in 1 % non-fat dried milk dissolved in 1X TBS with 0.05 % Tween-20 for 2 hours at room temperature. The blots were then rinsed again with 1X TBS as before and then they were visualized by using the Amersham ECL kit.

3.2.7 Electrophoretic mobility shift assays (EMSA).

End labeling of synthetic oligonucleotide probes (APB01, APB05) was carried out using T4 Polynucleotide Kinase (Invitrogen) in the presence of a two-molar excess of γ32P-ATP (>8,000Ci/mmol, ICN). The labeled oligos were then annealed to equal amounts of complementary oligos (APB10, APB50) by heating to 95°C and slowly cooling to room temperature. The oligonucleotide pairs APB01-APB10 and APB05-APB50 were annealed to generate APB1 and APB5 respectively. The oligonucleotides were then precipitated on glass filters for quantification in a scintillation counter. The probes used correspond to:

APB1: APB10 5’-GATCCGGGTCAGTGTACCTACCAACCTAATGACGAGG-3’,
APB01 5’-GATCGTGTTGTTAGGTTTTAGGTACACTGACGGC-3’

APB5: APB50 5’-GATCCGGGTCAGTGTACCGATCGTCTAATGACGAGG-3’,
APB05 5’-GATCGTGTTGTTAAGACGATCGGTACGCTCCTTAAACACG-3’
DNA-binding assays were performed on ice for 30 minutes in 25 μl total volume in A-0 buffer with 0.8 μg poly d(I)/d(C) and 1 mM DTT (unless otherwise indicated). After incubation on ice, approximately 10,000 cpm of end-labeled oligonucleotide probe was added and incubated on ice for an additional 30 minutes. Protein-DNA complexes were resolved on a 8 % polyacrylamide gel (80:1, acrylamide:bis-acrylamide) with 0.25X Tris borate-EDTA (TBE) running buffer at 415 V for 55 minutes at 4°C. After electrophoresis, gels were dried onto Whatman paper and subjected to autoradiography at –70°C overnight using X-ray film or a Kodak phosphorimager for 2 hours and quantified using BioRad imaging system.

The apparent dissociation constants were estimated by carrying out the DNA-binding assays as described above with a fixed amount of protein (35 ng) and in the presence of varying amounts of cold double stranded oligonucleotide (ranging from 0 to 5.240 nmoles). After incubation on ice, approximately 10,000 cpm of end-labeled oligonucleotide probe was added and incubated on ice for an additional 30 minutes. Quantification of the free and bound APB1 oligo was estimated from the radioactivity present in each of the corresponding bands, and calibrated against known standards. To compare our studies with others [38, 141] data was assumed to fit to a linear relationship, although comparable RMS values are obtained using a hyperbolic fit (not shown). It should be noted that the APB1 probe is formed by two overlapping binding sites [36], and while only one protein can bind to APB1 at a time, thus ruling out a cooperative interaction (not shown), our studies cannot rule out the possibility that each binding site is recognized with different affinities by the different proteins.

53
3.3 Results.

3.3.1 The ARE is required for R-enhanced activity.

Previously, we showed that transferring the interaction with R from C1 to P1 resulted in the P1* protein with novel regulatory activities [130]. In the absence of R, P1* activated the A1 promoter (pA1Luc, Fig. 3.1B) just like P1 (Fig. 3.2A) (R-independent transcription, Table 2.1), but not the Bz1 promoter (Fig. 3.2B). In the presence of R P1* activated Bz1, similar to C1 (R-dependent transcription, Table 2.1) (Fig. 3.2B), and induced anthocyanin accumulation (not shown). P1* displayed a 2-3 fold enhanced activity on the pA1Luc in the presence of R, (Fig. 3.2A, compare P1* and P1*+R). To ascertain whether this R-enhanced activity required cis-elements other than the MYB binding sites, we tested whether P1* displays R-enhanced activity on a promoter containing three copies of the haPBS upstream of a minimal CaMV 35S promoter (p(haPBS)3Luc, Fig. 3.1B). This synthetic promoter is activated by P1 and C1 + R, but not by C1 alone (Fig. 3.2C). On this promoter, however, the activity of P1* is not enhanced by R (Fig. 3.2C, compare P1* and P1*+R). These results suggest that other cis-regulatory elements in the A1 promoter participate in the R-enhanced transcription of A1 by P1*.

The ARE element was identified as a conserved motif present in several flavonoid biosynthetic gene promoters [131]. A six base pair insertion in the A1 ARE generated by the excision of Spm from the a1-m2 allele dramatically reduced aleurone anthocyanin pigmentation controlled by C1 and R, without a significant effect on the pericarp
phlobaphene pigmentation specified by P1 [133]. To establish whether the ARE is responsible for the R-enhanced activity of P1*, we tested the promoter containing the six base pair insertion (pA1315L-Luc, Fig. 3.1B) for activation by P1, P1*, P1* + R and C1 + R (Fig. 3.2D). Supporting the significantly reduced anthocyanin pigmentation of the a1315L allele in vivo [133], C1+R activate pA1315L-Luc significantly weaker than pA1Luc (Fig. 3.2D), while P1 activates pA1Luc and pA1315L-Luc at similar levels (Fig. 3.2D, P1). P1* activates pA1Luc and pA1315L-Luc at similar levels, but in contrast to the activation of pA1Luc, the activation of pA1315L-Luc by P1* is not enhanced by R (Fig. 3.2D). These results indicate that one function of R is to enhance the transcriptional activity of C1 and P1*. This R-enhanced activity requires the ARE cis-regulatory element.

3.3.2 Contrary to P1, C1 does not activate transcription in yeast from the \(^{ha}\)PBS.

In addition to the ARE-mediated R-enhanced activity, R is essential for the activity of C1. To investigate what makes C1 activity R-dependent, we developed a yeast system in which two copies of the \(^{ha}\)PBS were introduced upstream of a minimal yeast promoter driving the HIS3 selectable gene (Fig. 3.3). The p\(^{ha}\)PBS\(_{2\times}\)HIS3 construct was integrated as a single copy in the yeast genome (linked to the URA3 marker), to generate the YEG102 yeast strain (see Materials and methods). The P1 and C1 proteins were expressed in YEG102 from the GPD promoter in a plasmid permitting selection in synthetic media lacking tryptophan (-Trp). When grown in –Ura –Leu –Trp – His media, a robust growth was observed for the yeast cells expressing P1 (Fig. 3.3A, P1). In contrast, cells expressing C1 verified by Western analysis using the C1 monoclonal antibody C1 40/1-249 (Fig. 3.3B) did not grow in the –Ura –Leu –Trp – His media (Fig. 55
3.3A, C1), even when in the presence of a version of the R paralog B, in which the bHLH domain was deleted (B^ΔbHLH) to permit the interaction with C1 [37]. These results suggest that C1 is unable to activate transcription on its own, even in a heterologous, non-plant system. In contrast, if provided with a promoter containing high-affinity binding sites, P1 activates transcription independently of other plant specific factors.

3.3.3 C1^SH, a mutant of C1 that activates transcription in yeast.

Previously, we showed that the second and third helices of R3 in C1 (Fig. 3.1A) participated in making the activity of C1 R-dependent [130]. In addition, we established that the G94 and R95 residues in C1 (Fig. 3.1A) needed to be transferred to P1 to create the P1* protein [130]. Thus, we targeted G94 and R95 as candidate residues that participate in making the C1 activity R-dependent. We generated the C1^SH protein containing the G94 and R95 amino acids replaced by S and H respectively, as found in P1 (Fig. 3.1A). The MYB domains of P1, C1 and C1^SH were fused to the Gal4 activation domain (Gal4^AD) and the resulting P1^MYB-Gal4^AD, C1^MYB-Gal4^AD, and C1^SHMYB-Gal4^AD proteins were expressed from the GPD1 promoter in the yeast strain YEG102. Cells expressing the fusion constructs were assayed for growth in media without histidine (-His), which would be indicative of transcription of the p(haPBS)2xHIS3 reporter. In these conditions, cells expressing P1^MYB-Gal4^AD and C1^SHMYB-Gal4^AD grew in the selective media (Fig. 3.4A1, and 3.4A5), while cells expressing C1^MYB-Gal4^AD did not (Fig. 3.4A3). The inability of C1^MYB-Gal4^AD to activate transcription of the p(haPBS)2xHIS3 reporter is not complemented by the co-expression of either R or B^ΔbHLH (not shown). The expression levels of C1^MYB-Gal4^AD and C1^SHMYB-Gal4^AD were comparable, as shown
by Western analyses using the C1 monoclonal antibody C1 40/1-249 (Fig. 3.4C). Thus, the G94S and R95H substitutions are sufficient to allow C1 SHMYB-Gal4 AD to activate transcription in yeast, independently of R. No activation by P1, C1 nor the respective fusions to the Gal4 AD was observed in a yeast strain in which the haPBS sites in p(haPBS)2xHis3 were mutated to a DNA sequence to which neither P1 nor C1 bind in vitro (not shown).

To investigate whether the G94S and R95H amino acid changes affected the interaction of C1 with R, we fused the N-terminal region of R (R1-252), which is sufficient for the interaction with the MYB domain of C1 [130], to the Gal4 DNA-binding domain (Gal4DBD) to create R1-252-Gal4DBD. We then co-expressed P1MYB-Gal4 AD, C1MYB-Gal4 AD or C1 SHMYB-Gal4 AD with R1-252-Gal4DBD in the PJ69.4a yeast strain and assayed the ability of cells co-expressing these proteins to grow in selective media (Figure 3.4B). Because PJ69.4a has the HIS3 and ADE2 selectable genes regulated by Gal4 binding sites [114], successful interaction would result in growth in –Trp –Leu –His –Ade media. C1MYB-Gal4 AD and C1 SHMYB-Gal4 AD (Fig. 3.4B4 and 3.4B46) but not P1MYB-Gal4 AD (Fig. 3.4B2), provide robust interactions with R1-252-Gal4DBD. Thus, the G94S and R95H replacements in C1 do not interfere with the ability of C1 to interact with R, and allow the corresponding MYB domain to activate transcription in yeast, when fused to the Gal4 activation domain.

3.3.4 C1 SH binds DNA with high affinity.

The ability of C1 SHMYB-Gal4 AD to activate transcription from the haPBS in yeast suggested that the G94S and R95H substitutions increased the normally low DNA-
binding affinity of C1 [38]. To determine whether C1^{SH} binds DNA better than C1, we expressed the corresponding MYB domains in *E. coli* utilizing the IMPACT™ (Intein Mediated Purification with an Affinity Chitin-binding Tag) system to obtain 80-90% pure P1^{MYB}, C1^{MYB} and C1^{SHMYB} proteins (Fig. 3.5A) without any need for denaturation, and only modified by the addition of a Cys residue at their C-termini (see EXPERIMENTAL PROCEDURES). Electrophoretic mobility shift assays (EMSA) showed that the C1^{MYB} protein (Fig. 3.5B, lane 4) binds the APB1 probe, representing the \(^{ha}\)PBS present in the *a1* promoter, much weaker than P1^{MYB} (Fig. 3.5B, lane 2). Interestingly, C1^{SHMYB} bound DNA with a much higher affinity than C1^{MYB} (Fig. 3.5B, lane 3). Similar to P1 and C1, C1^{SH} did not bind to a mutant sequence in which the \(^{ha}\)PBS have been mutated (APB5, not shown).

To quantitatively compare the DNA-binding affinity of C1^{SHMYB} with those of P1 and C1, we determined the apparent equilibrium dissociation constant (K\(_d\)) of the C1^{SHMYB}-APB1 complex. The results of Scatchard analyses (Fig. 3.5C) carried out in the conditions previously described for P1 and C1 [38, 36, 141] show that C1^{SHMYB} binds APB1 with an affinity comparable to that of P1 (Table 3.1) or P1^{MYB} (Fig. 3.5C), significantly stronger than C1 (Table 3.1) or C1^{MYB} (Fig. 3.5C).

### 3.3.5 High-affinity DNA-binding is not sufficient for R-independent transcription by C1.

Having obtained a C1 mutant, C1^{SH}, capable of binding DNA with high affinity, we were posed to answer the important question of whether the need of C1 for R resides in the normal inability of C1 to bind DNA with high affinity. For this purpose, we
assayed the activity of p35SC1\textsuperscript{SH} on the pA1Luc and pBz1Luc reporter constructs (Fig. 1B) in maize BMS cells by transient expression experiments. C1\textsuperscript{SH} does not efficiently activate transcription of either promoter on its own (Fig. 3.6, C1\textsuperscript{SH}) yet in the presence of R, C1\textsuperscript{SH} efficiently activates pA1Luc and pBz1Luc (Fig. 3.6, C1\textsuperscript{SH}+R). The inability of C1\textsuperscript{SH} to provide a strong activation of pA1Luc or pBz1Luc in the absence of R is not compensated by increasing the concentration of the protein in the transient expression experiments (not shown). These results show that a C1 mutant capable of binding DNA with high affinity remains dependent on R for function in maize cells. Thus, the function of R cannot be solely increasing the \textit{in vivo} affinity of C1 for DNA.

3.3.6 The fusion of C1 with the Gal4 DNA-binding domain makes Gal4-mediated transcription dependent on R.

Previous studies have shown that the C-terminal activation domain of C1, when fused to the DNA-binding domain of the yeast Gal4 transcription factor (Gal4\textsuperscript{DBD}–C1\textsuperscript{C-term}), was able to activate transcription independently of R [38, 39, 139], as shown in Figure 3.7A. To investigate whether the full-length C1 protein remains R-dependent in the presence of another DNA-binding domain, we introduced the Gal4\textsuperscript{DBD} between the MYB domain of C1 and its C-terminal region, to create C1\textsuperscript{MYB}–Gal4\textsuperscript{DBD}–C1\textsuperscript{C-term}. The activity of this protein was assayed on the pA1Luc reporter construct as well as on a reporter containing four tandem copies of the Gal4-binding sites upstream of a minimal \textit{CaMV 35S} promoter driving luciferase (pGal4\textsuperscript{BS}Luc), in the presence or absence of R, in maize BMS cells by transient bombardment. The ability of C1\textsuperscript{MYB}–Gal4\textsuperscript{DBD}–C1\textsuperscript{C-term} to activate both pA1Luc and pGal4\textsuperscript{BS}Luc was dependent on R (Fig. 3.7A). These results...
suggest that the presence of the MYB domain of C1, or an interaction of the MYB domain with the C1 C-terminal region, block the ability of the $Gal4^{DBD}$ to direct transcription from the Gal4-binding sites in maize cells, in the absence of $R$.

To determine why $C1^{MYB} \cdot Gal4^{DBD} \cdot C1^{C-term}$ is unable to activate transcription using either reporter construct in the absence of $R$, we expressed $C1^{MYB} \cdot Gal4^{DBD} \cdot C1^{C-term}$ in the yeast PJ69.4a and YEG102 yeast strains. As shown in Fig. 3.7B (right panel) $C1^{MYB} \cdot Gal4^{DBD} \cdot C1^{C-term}$ activates transcription from the Gal4-binding sites in PJ69.4a, manifested by a robust growth in –His –Ade media. In contrast, $C1^{MYB} \cdot Gal4^{DBD} \cdot C1^{C-term}$ did not activate transcription from the $^{ha}PBS$ in the YEG102 strain (Fig. 3.7B, left panel), similar to the results with $C1^{MYB} \cdot Gal4^{AD}$ (Figure 3.4A). Together, these results demonstrate that the $C1^{MYB} \cdot Gal4^{DBD} \cdot C1^{C-term}$ protein has the ability to bind to Gal4-binding sites and activate transcription in yeast, but not in maize cells. Thus, it is likely that a plant specific factor is involved in making the $C1^{MYB} \cdot Gal4^{DBD} \cdot C1^{C-term}$ activity dependent on $R$.

3.4 Discussion.

In this study we investigated the mechanisms by which the bHLH factor $R$ cooperates with the R2R3 MYB protein C1 to activate transcription of flavonoid biosynthetic genes. By comparing the regulatory activities of C1 and P1* on different promoters, we identified two mechanisms for R-dependent transcription. On synthetic promoters that contain only C1/P1 binding sites, $R$ functions as an essential co-activator of C1. This function of $R$ is unlikely to result solely in an increased C1 DNA-binding
affinity, since a C1 mutant (C1\textsuperscript{SH}) that binds DNA with comparable affinity as P1 remains R-dependent. A second function of R requires \textit{cis}-regulatory elements in addition to the C1/P1 DNA-binding sites. We hypothesize that R functions in this mode by binding or recruiting additional factors that bind to the ARE element conserved in the promoters of several anthocyanin genes. Together, these findings suggest a model in which combinatorial interactions with co-activators enable MYB factors with very similar DNA-binding preferences to discriminate between target genes \textit{in vivo}.

While P1 and P1* can activate transcription of a promoter containing the \textsuperscript{ha}PBS independently of R (Fig. 3.2C), the activity of P1* is enhanced by R on the pA1Luc reporter, but not on a promoter containing just the \textsuperscript{ha}PBS (Fig. 3.2A, C). The ARE, present in several maize flavonoid biosynthetic genes [131], provides a good candidate \textit{cis}-regulatory element that participates in R-enhanced transcription. We show here that an insertion of six base pairs in the ARE which dramatically affects A1 activation \textit{in vivo} by C1+R but not by P1 [133] also abolishes R-enhanced transcription by P1* in transient expression experiments (Fig. 3.2D). While this inhibitory effect of the insertion could be caused by altering the spacing between two \textit{cis}-regulatory elements, three lines of evidence suggest that it is the ARE element itself and not the spacing between ARE flanking elements that is important. First, a four base pair insertion in the ARE does not result in a decrease of A1 activation \textit{in vivo} [133] or in transient expression experiments (not shown). Second, linker-scan analyses in which the sequence but not the distance of the ARE element was altered also showed an inhibitory effect on C1+R activity [131,132,134]. And third, a conserved ARE element is necessary for the expression of a
flavonoid biosynthetic gene in *Gerbera hybrida* [142]. The participation of the ARE in R-enhanced transcription can be explained by proposing that R binds to the ARE, or that the ARE is recognized by a DNA-binding R-interacting factor (RIF in Fig. 3.8). However, no factor(s) has yet been identified that may participate in the recruitment of R to DNA. The *PAC1* gene, discovered as a mutation that reduces R+C1-specified anthocyanin pigmentation in maize aleurones [73, 74], may encode a key player in this complex, but is unlikely to recruit R to DNA because it contains no recognizable DNA-binding domains [73].

This R-mediated recruitment to the ARE would also participate in allowing P1 and C1 to activate different sets of target genes, thus contributing to their regulatory specificity, by providing R-dependent activation of promoters lacking DNA-binding sites for the R2R3 MYB factors. Indeed, the putative C1 binding sites in the *Bz1* promoter were replaced with the *ha*PBS present in the *A1* promoter, and neither P1* or C1 were able to activate independently of R from the mutant promoter (Hernandez, Heine, and Grotewold, unpublished). Moreover, a complete disruption of the site did not impair R-dependent transcription by P1* or C1, indicating that in this promoter, the correct location of the MYB binding site is unknown (Hernandez, Heine, and Grotewold, unpublished). Our results suggest that, in the absence of P1-binding sites, R can recruit P1* to the *Bz1* promoter, explaining why P1* and C1, but not P1, activate this gene (Fig. 3.8, see activation of *Bz1* by P1*). This ability of R to recruit interacting R2R3 MYB factors to DNA was previously observed using a mutant *A1* promoter lacking the *ha*PBS and *ha*PBS. This mutant promoter can still be activated by C1+R, albeit at a reduced level.
Similarly, a C1 protein defective in DNA binding also retains about 20% of the transcriptional activity of wild type C1 [38]. In these cases, the role of the R2R3 MYB (P1* or C1) would be to provide the activation domains lacking in R.

While the R-enhanced activity of P1* depends on the ARE, the R-dependent activity of C1 is observed on promoters containing just the $^\text{a} \text{PBS}$ (Fig. 3.2C), suggesting a direct effect of R on C1. There are several possible mechanisms by which the activity of C1 could be dependent on R. R could increase the affinity of C1 for DNA, it could mediate C1 localization to the nucleus, relieve the effect of an inhibitory domain of C1 or an external C1 inhibitor, or it could stabilize the C1 protein. Our results showing that the activity of C1$^{\text{SH}}$, a C1 mutant that binds DNA with an affinity comparable to that of P1 (Table 2), remains R-dependent strongly suggests that R is not required simply to increase the intrinsically low DNA-binding affinity of C1. The G94 and R95 residues are not required for the interaction with R (Fig. 3.7B), and given their positions in the MYB domain [100], are unlikely to involve direct contacts with DNA. Thus, G94 and R95 could be involved in maintaining a specific MYB domain conformation that modulates DNA-binding.

The C1$^{\text{MYB-Gal4}}$-DBD–C1$^{\text{C-term}}$ chimeric protein is dependent on R for the activation of promoters containing C1- or Gal4-binding sites in maize cells (Fig. 3.7A). However, C1$^{\text{MYB-Gal4}}$-DBD–C1$^{\text{C-term}}$ activates transcription in yeast from the Gal4-binding sites in the absence of other plant factors. This result suggests that the inability of C1$^{\text{MYB-Gal4}}$-DBD–C1$^{\text{C-term}}$ to activate transcription in maize cells is not due to the presence of an auto-inhibitory domain, as found in ETS-1 and other transcription factors [10,143].
Furthermore, the ability of C1\textsuperscript{SH}-Gal4\textsuperscript{AD} and C1\textsuperscript{MYB}-Gal4\textsuperscript{DBD}-C1\textsuperscript{C-term} to activate transcription in yeast also suggests that these proteins contain nuclear localization signals (NLS) that function in yeast and therefore are likely to function in maize cells. Thus, it is unlikely that the R-dependent function of C1 resides solely in the R NLSs [144] driving the R/C1 complex to the nucleus. Together, these results suggest that there is a plant factor, absent in yeast cells, which either (i) inhibits C1 activity by masking the DNA-binding or activation domains, (ii) retains C1 in the cytoplasm, or (iii) destabilizes the C1 protein. The role of R would be to either overcome the effect of this plant factor, or displace it from its binding to C1. The \textit{in1} gene, encoding a protein with a bHLH domain, acts as an inhibitor of the pathway [58]. It is not currently known whether IN1 is responsible for making the C1 activity R-dependent. However, IN1-like proteins have not been identified as inhibitors of anthocyanin accumulation in other plants in which the action of a C1-like R2R3 MYB factors is dependent on R-like bHLH proteins for function. For example in petunia, a single MYB repeat protein (MYBx) functions as an inhibitor of anthocyanin pigmentation [145], in a fashion that resembles the activity of the CPC and TRY single MYB repeat proteins in trichome and root hair differentiation in \textit{Arabidopsis} [146]. The FaMYB1 R2R3 MYB protein inhibits transcription of flavonoid genes in strawberry [52]. Thus, it is possible that different plants utilize different mechanisms for the modulation of the activity of the C1 or R regulators. Alternatively, a common factor, as yet unidentified, could be responsible for making C1 and C1-like proteins dependent of R-like proteins for function.
Plants express hundreds of R2R3 MYB-domain proteins with very related DNA-binding activities [43,44]. Our results with P1 and C1 provide a framework to start to understand how these proteins achieve regulatory specificity. Clearly, R plays a key role in enabling related factors to recognize and activate distinct subsets of promoters, evidenced by the ability of P1* to activate anthocyanin biosynthesis, a function only gained through its interaction with R. We propose that, in addition to slight differences in DNA-binding preferences by the R2R3 MYB factors, gene target specificity is provided by 1) the specific recruitment of R-like factors by the MYB domains, and 2) the modular organization of the promoters formed by binding sites for the MYB and other factors recruited by the R-like co-activators. The *Arabidopsis* genome encodes 12-16 bHLH factors that share conserved N-termini with R (Braun, E.L. and Grotewold, E. unpublished), which contain the conserved MYB interaction region. It is likely that specific interactions of MYB proteins with particular R-like bHLH factors provides the first level of specificity, with the correct organization of *cis*-elements in the corresponding target genes providing a second level of specificity.

In summary, this study provides novel insights into the mechanisms by which two branches of maize flavonoid biosynthesis are independently regulated by transcription factors with very similar R2R3 MYB DNA-binding domains. Our results indicate that the interaction between C1 and R is essential for the regulatory specificity of C1, and that multiple *cis*-regulatory elements are required for robust and specific gene activation.
These findings emphasize the important role of combinatorial control of gene expression in allowing transcription factors with very similar DNA-binding domains to regulate specific cellular processes.
Figure 3.1: Sequence of the MYB domains of C1 and P1 and structure of the plasmids used for transient expression experiments.

A, Sequence alignment of the MYB domains of C1 (C1\textsuperscript{MYB}), P1 (P1\textsuperscript{MYB}) and P1* (P1*\textsuperscript{MYB}). P1* corresponds to a mutant of P1 in which six residues different in C1 (marked by asterisks) were replaced in P1, conferring P1* the ability to interact with R [130]. Shaded areas indicate amino acid identity, and the position of the $\alpha$-helices that form each of the two MYB repeats (R2 and R3) that characterize R2R3 MYB domains.

B, Structure of the reporter constructs utilized for the transient expression in maize Black Mexican Sweet (BMS) cells. The high- and low-affinity P1-binding sites are indicated as $^{ha}$PBS and $^{lo}$PBS respectively, ARE indicates the Anthocyanin Regulatory Element, CBS indicates C1-binding sites in the Bz1 promoter [147, 119].
Figure 3.2: Activation of transcription by P1, C1 and P1*.

Results of transient expression following co-bombardment of cultured maize BMS cells with P1, C1 and P1*. All the regulators are expressed from the CaMV 35S promoter (p35S). A, Activation of pA1Luc. B, Activation of pBz1Luc. C, Activation of p(hoPBS)3Luc. D, Activation of pA1Luc or pA1315Luc. A pUbiGUS construct was included in every bombardment as a normalization control. Triplicates were done for each treatment, and the data were normalized for GUS activity as described. The fold activation was calculated as the ratio between each particular treatment and the treatment with the reporter constructs without activator. The average values are shown and the error bars indicate the standard deviation of the samples [119].
Figure 3.2
Figure 3.3: P1, but not C1, activates transcription in yeast.

A. Activation of transcription of the (\(^{3\text{haPBS}}\)\(2\times\)HIS3 construct (\(URA^+\)) by P1 and C1 expressed from the constitutive GPD promoter (\(TRP^+\)). Growth in media without histidine (-His) is indicative of transcriptional activation of the HIS3 gene. B. Western blot analysis of yeast strains expressing empty vectors (indicated as Empty Vector) and C1 + B\(^\Delta bHLH\). The blot was probed with the C1 monoclonal antibody C1 40/1-249. As a loading control, an identical gel was stained with Coomassie Brilliant Blue (indicated as Coomassie) [119].
Figure 3.4: C1$^{SH}$ activates transcription in yeast and interacts with R.

A, Activation of transcription in yeast of the (haPBS)$_2$x HIS3 construct (URA') by fusions of the MYB domains of P1, C1 or C1$^{SH}$ (C1 with G94S and R95H substitutions) to the Gal4-activation domain (P1$^{MYB}$-Gal4$^{AD}$, C1$^{MYB}$-Gal4$^{AD}$ and C1$^{SHMYB}$-Gal4$^{AD}$ respectively). Growth in –Ura –Leu –Trp –His plate is indicative of activation. B, Yeast two hybrid experiment to verify interaction between P1$^{MYB}$-Gal4$^{AD}$, C1$^{MYB}$-Gal4$^{AD}$ or C1$^{SHMYB}$-Gal4$^{AD}$ and the N-terminal region of R fused to the Gal4 DNA-binding domain (R$^{1-252}$-Gal4$^{DBD}$) in the yeast strain PJ69.4a [114] containing the HIS3 and ADE2 genes under the control of Gal4-binding sites. C, Western blot of yeast protein extracts from the YEG102 strain expressing C1$^{SHMYB}$-Gal4$^{AD}$ (5, corresponding to section 5 in Fig 3.4A) or C1$^{MYB}$-Gal4$^{AD}$ (3, corresponding to section 3 in Fig 3.4A). The blot was probed with the C1 monoclonal antibody C1 40/1-249, and with an antibody against the yeast cytoplasmic helicase DED1p protein (yDED1p) to normalize [119].
Figure 3.5: C1<sup>SH</sup> binds DNA with high affinity.

A, SDS-PAGE analysis of the purified MYB domains of P1<sup>MYB</sup>, C1<sup>MYB</sup>, and C1<sup>SHMYB</sup> used in the EMSA experiments. B, Comparison of the DNA-binding activities of recombinant, purified C1<sup>MYB</sup>, C1<sup>SHMYB</sup> and P1<sup>MYB</sup> domains by EMSA, using a fixed amount of protein, <sup>32</sup>P-labeled APB1 probe containing the <sup>ha</sup>PBS of the A1 promoter, and titrating it with varying amounts of non-labeled APB1 DNA. C, Scatchard analysis of P1<sup>MYB</sup>, C1<sup>MYB</sup>, and C1<sup>SHMYB</sup> binding to the A1 high-affinity P1-binding sites (<sup>ha</sup>PBS). Scatchard plots are shown with the x-axis representing the bound/free APB1 ratio estimated from the amount of radioactivity in these two fractions, and the y-axis representing the amount of bound APB1, estimated from the amount of radioactivity in the bound fraction. For each experiment, the RMS and apparent equilibrium dissociation constants (K<sub>d</sub>), deduced from the slopes assuming a linear relationship (see Materials and methods), are indicated [119].
Figure 3.6: The activity of C1\textsuperscript{SH} remains R-dependent in maize cells.

A, Activation of transcription of pA1Luc in BMS maize cells by p35SC1 and p35SC1\textsuperscript{SH} in the absence and presence of p35SR. B, Activation of transcription of pBz1Luc in BMS maize cells by p35SC1 and p35SC1\textsuperscript{SH} in the absence and presence of p35SR. The average values are shown and the error bars indicate the standard deviation of the samples. The fold activation was calculated as in Fig. 3.2 [119].
Figure 3.7: C1 confers R-dependency upon the Gal4 DNA-binding domain.

A, Activation of transcription of pGal4BSLuc and pA1Luc by p35SC1 and p35SC1MYB-Gal4DBD-C1C-term in the absence and presence of p35SR. The activation of transcription of pGal4BSLuc by p35SGal4DBD-C1C-term in the absence of R is indicated as well. The average values are shown and the error bars indicate the standard deviation of the samples. The fold activation was calculated as in Fig. 3.2. B, Activation of transcription in the yeast strains YEG102 and PJ69.4a by the P1 and C1MYB-Gal4DBD-C1C-term proteins expressed from the constitutive GPD promoter. Selection was done as described in Fig 3.4 [119].
Model for the regulation of the $A1$ and $Bz1$ genes by P1, C1+R or P1*+R. The cooperation between C1 and R has at least two components: The first component is dependent on C1 making contacts with the C1/P1 binding sites (PBS) and is likely involved in relieving C1 from the effect of an inhibitory plant cellular factor (shown as a red circle). The second component of R is to make direct or indirect (through an as yet unidentified R-Interacting factor, RIF) contacts with the DNA through the ARE cis-regulatory elements. A similar mechanism of regulation by C1 and R on the $A1$ and $Bz1$ promoters is suggested, the latter containing only C1-binding sites (CBS) to which P1 or P1* do not bind. P1* activates the $Bz1$ promoter through its interaction with R and recruitment to DNA by the ARE. The thickness of the arrows is proportional to the levels of activation by the transcription factors [119].
<table>
<thead>
<tr>
<th>Protein</th>
<th>K\textsubscript{d} (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P\textsuperscript{MYB}</td>
<td>28\textsuperscript{a}</td>
</tr>
<tr>
<td>C\textsuperscript{MYB}</td>
<td>N.D.</td>
</tr>
<tr>
<td>C\textsuperscript{SHMYB}</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

Table 3.1: Comparison of the apparent equilibrium dissociation constants (K\textsubscript{d}) for the \textit{A1} high affinity P1 binding site by the MYB domains of P (P\textsuperscript{MYB}), C1 (C\textsuperscript{MYB}) and C\textsuperscript{SH} (C\textsuperscript{SHMYB}) \cite{119}.

\textsuperscript{a} Values reported for N\textsubscript{10}His-P1\textsuperscript{MYB} in Williams and Grotewold (1997).
\textsuperscript{b} Values reported for N\textsubscript{10}His-P1\textsuperscript{MYB} and N\textsubscript{6}His-C1\textsuperscript{MYB} in Sainz et al. (1997a)
\textsuperscript{c} Values obtained for P1\textsuperscript{MYB}, C1\textsuperscript{MYB}, and C\textsuperscript{SHMYB} in this work. N.D. indicates not determined.
CHAPTER 4

CHARACTERIZATION OF RIF-1C, A NOVEL R-INTERACTING FACTOR THAT HARBORS A BRCA2-BINDING LIKE DOMAIN.

4.1 Introduction.

The regulation of the maize anthocyanin biosynthetic pathway in maize is one of the best characterized cases of combinatorial transcriptional control in plants. The MYB domain protein C1 and its bHLH partners the B or R, regulate the transcription of the enzymes that are involved in the synthesis of the purple-blue anthocyanin pigments [35, 85]. We have examined the mechanism by which the C1-R complex activates transcription of their target genes and found that R provides additional DNA contacts, although it is not clear yet whether they are direct or indirect, and it also relieves C1 from a plant inhibitor that prevents it from activating transcription [119]. We have proposed that R achieves these functions by recruiting additional factors to the complex, which is supported by the presence of multiple domains that could provide protein-protein interaction surfaces.

The first of these domains to be identified was the bHLH, which suggested that R had the potential to bind DNA; however, in spite of a number of attempts in multiple laboratories, R has yet to be shown to bind DNA directly. Another role that has been proposed for the R bHLH is to mediate homodimerization, as other bHLH transcription
factors are known to dimerize through the HLH domain. R is known to homodimerize but it does so through an ACT-like domain located at the very C-terminus of the protein and not through the bHLH (Feller et al. 2006. Submitted manuscript). On the other hand, the bHLH region is necessary to heterodimerize with RIF-2C and RIF-3C, two other bHLH proteins (Feller and Grotewold, unpublished). In the work presented in this chapter, I will describe data that supports that the bHLH plays a critical role in R function and that its effects are only fully apparent when activating genes in a chromatin context.

It is well known that chromatin structure makes DNA accessibility a challenge that the cellular machinery has to overcome in order to execute processes such as transcription, DNA repair, and DNA replication. For this purpose the cell has multi-protein remodeling complexes that are recruited to chromatin to allow other factors to bind DNA. These remodeling complexes mainly act by modifying the histone proteins thus affecting the DNA-histone interactions which in turn results in changes in the position of the nucleosomes [148, 149]. The activity of chromatin remodeling complexes is ATP-dependent, and they are classified into distinct groups depending on what other domains are present in the ATPase subunit. The domains that have been found in these ATPases are the bromodomain, the chromodomain, the latter of which belong to the Royal super-family, and the Sant domain [149].

One mechanism by which transcription factors turn on genes is by recruiting, directly or indirectly, specific ATP-dependent remodeling complexes to their target promoters, followed by modifying enzymes such as histone acetyl transferases (HAT), resulting in an altered chromatin structure that allows the RNA polymerase II complex to
initiate transcription. Some repressors use a similar strategy, by recruiting modifying enzymes such as histone deacetylases (HDAC), histone methyltransferases, and DNA methyltransferases, which silence gene expression by generating inactive forms of chromatin [150].

We have identified an R interacting protein that specifically interacts with the R bHLH, which we have named R interacting factor 1C (RIF-1C). RIF-1C has an ENT domain and an Agenet domain. The presence of the latter domain suggests a role in chromatin remodeling as this domain belongs to the Royal super-family, which is comprised of domains that seem to have evolved from a common ancestor that binds methylated substrates. Many members of this family, as the ATPases discussed above, have been shown to be involved in chromatin remodeling complexes [151].

Here I present the characterization of RIF-1C in maize and its closest homolog in Arabidopsis. Our results suggest that it is possible that RIF-1C is involved in chromatin remodeling associated with transcriptional regulation, and perhaps with DNA repair mechanisms as well. The latter may or may not involve the participation of R-like proteins.

4.2 Materials and methods.

4.2.1 Plasmids used in yeast two-hybrid experiments and GST pull-downs.

The following fragments were generated by PCR and then cloned into pBD-GAL4 (Stratagene, CA) as EcoRI/SalI fragments: the C-terminus of R which is comprised of amino acids 411 to 610, the bHLH regions of R and R^{LeD12} comprising
amino acids 411 to 464 and 411 to 467, respectively, the region of R corresponding to amino acids 462 to 610, the ENT domain of RIF-1C and the ENT domain of ACK1 which corresponds to amino acids 44 to 127, and amino acids 45 to 128, respectively, fragments of RIF-1C corresponding to amino acids 1-127, 230-452, 1-236, 44-299, and 44-452. The RIF-1C and ACK1 constructs were also cloned as an *EcoRI/SalI* fragment into pAD-GAL4 (Stratagene, CA). The RIF-1C full cDNA fused to the Gal4^AD^ was obtained from a screen using a yeast two hybrid cDNA library from immature tassel cloned into pAD-GAL4-2.1 (Stratagene, CA) obtained from the laboratory of Dr. Robert Schmidt. The full cDNA of At5g06480 fused to the GAL4^AD^ was obtained from a screen using a yeast-two hybrid library cloned into pAD-GAL4-2.1 (Stratagene, CA) from inflorescence meristem, floral meristem and floral buds up to stage 8 or 9, donated to the Arabidopsis Biological Resource Center (ABRC) by Drs. Hong Ma and Benjamin Horwitz.

The GST pull-down bait was made by cloning the R C-terminal region expanding amino acids 411 to 610 into the vector pGEX-KG [152] as an *XhoI, HindIII* fragment. The pGEX-KG vector was used as a negative control. The Gal4^AD^-RIF-1C cDNA was excised from the library plasmid with *HindIII* and *SalI* and then cloned into pBluescript (Stratagene, CA) to be used for *in vitro* transcription and translation reactions.

### 4.2.2 Plasmids used in transient expression experiments.

As previously described [130], all plant expression vectors include the *CaMV 35S* promoter, the tobacco TMV Ω’ leader and maize first Adh1-S intron in the 5’UTR, and potato proteinase II (pinII) termination signal, unless otherwise specified. Previously
described plasmids [36, 38, 130] include p35SC1, p35SR, pA1Luc, containing 1.4 kb of the A1 promoter and the Adh1-S intron; p35SBAR (pPHP611), and Ubi::GUS (PHP3953). The p35SR\textsuperscript{AbHLH} (pPHP5660) construct is identical to p35SR except that is missing amino acids 406 to 474. We obtained the p35SR\textsuperscript{LeD12} [65] construct from the laboratory of Dr. Sue Wessler.

4.2.3 Plasmids for GFP transient assays.

The cDNAs of R, R\textsuperscript{AbHLH}, and RIF-1C were generated by PCR and cloned into pENTR/D-TOPO (Invitrogen, CA). Then the cDNAs were mobilized to the pGWB5, a binary vector derived from pABH-Hm1 [153] in which a HindIII-Sacl fragment containing the beta amylase promoter driving GUS has been replaced with a GATEWAY cassette fused to sGFP [154], to obtain C-terminal fusions to GFP, via LR clonase reactions (Invitrogen, CA). The pGWB5 was a kind gift from Dr. Tsuyoshi Nakagawa (Research Institute of Molecular Genetics, Shimane University). A construct of the silencing suppressor p19 protein of tomato bushy stunt virus in a binary vector driven by the 35S CaMV promoter (pBIN61) [155], was used to boost expression.

4.2.4 Plasmids for RNAi experiment.

The RNAi AtAck1 500 bp fragment was synthesized with primers AtRIF-1C-KOF, TTATCTAGAGGCGCGCCAAATTCATCTCATTGAGCAAG, and AtRIF-1C-KOR, TTAGGATCCATTTAAATTTCTTGGTCTTTGGACCTGGT. The PCR product was then cloned as XbaI/Asc-I and BamHI/SwaI fragments into pFGC5941.
4.2.5 Plant material and seed stocks.

We obtained seeds through the ABRC that were deposited by the Arabidopsis TILLING project after we requested them to target gene At5g13020, corresponding to the ack1-1 (CS93489; D74N mutation), ack1-2 (CS96315; Q150* mutation), and ack1-3 (CS93973; D233N mutation) mutant alleles for the ACK1 gene. Columbia ecotype plants were used for the RNAi experiments. These plants were transformed with either the RNAi construct for AtAck1, or the empty vector pFGC5941. Plant transformation and growth were done as previously described [78].

4.2.6 Yeast two-hybrid experiments.

The screen using the Gal4\textsuperscript{DBD-}R\textsuperscript{411-610} as a bait was done using a maize cDNA library from immature tassel fused to the Gal4\textsuperscript{AD}. Gal4\textsuperscript{DBD-}R\textsuperscript{411-610} was transformed into yeast strain PJ69.4a [114], to obtain a bait strain that was then transformed with 10 \( \mu \)g of library and plated on SC-LEU-TRP-HIS medium. Colonies were then screened for growth on SC-LEU-TRP, SC-LEU-TRP-HIS, and SC-LEU-TRP-HIS-ADE. Positives were then grown in 5 ml of rich YEPD liquid medium overnight. The culture was pelleted and then resuspended in 200 \( \mu \)l of 2\% Triton X-100, 1\% SDS, 100 mM Tris pH 8.0, 100 mM NaCl, and 1 mM EDTA. The resuspended pellet was then mixed with 200 \( \mu \)l of phenol/chloroform/isoamyl alcohol (25:24:1) and 100 \( \mu \)l of glass beads by vortexing 2 minutes. The aqueous phase was separated after centrifugation at 14,000 rpm for 5 minutes and then the DNA was precipitated by adding 2 volumes of 100\% ethanol and 1/10 volume of sodium acetate pH 5.2 followed by a 10 minute incubation in ice. The DNA was pelleted by centrifugation at 14,000 rpm for 20 minutes. After drying the pellet
for 20 minutes at room temperature, the pellet was resuspended in 50 μl of water. This was followed by transformations into *E. coli* by electroporation with 5-10 μl of the yeast DNA preparation. Six colonies from each transformation were analyzed by restriction digest. Prey constructs were then re-transformed into PJ69.4a [114] with the bait construct and with empty vector and plated on SC-LEU-TRP medium. Colonies were further screened for growth on SC-LEU-TRP, SC-LEU-TRP-HIS, and SC-LEU-TRP-HIS-ADE. Constructs that resulted in growth on the SC-LEU-TRP-HIS, and SC-LEU-TRP-HIS-ADE with the bait but not with the empty vector were subsequently sequenced. A total of screened 2.66 x 10^6 clones were screened, and identified five positive clones. Another small screen (7 x 10^5 clones of which two were positive) was done using Gal4<sup>DBD</sup>-AtACK1<sup>ENT</sup> and a cDNA library from inflorescence meristem, floral meristem and floral buds up to stage 8 or 9 fused to the Gal4<sup>AD</sup>. For the directed interactions we co-transformed fusions to Gal4<sup>DBD</sup> and Gal4<sup>AD</sup> into yeast strain PJ69.4a [114], and plated on SC-LEU-TRP medium. Colonies were then screened for growth on SC-LEU-TRP, SC-LEU-TRP-HIS, and SC-LEU-TRP-HIS-ADE.

4.2.7 GST pull-down experiments.

The GST-R<sup>411-610</sup> and pGEX-KG were each transformed into *E. coli* BL21 (DE3) PlyS cells for expression. Cultures were grown, induced and purified essentially as described [141], with the following modifications. After induction of a 500 ml culture with 1 mM IPTG, the cells were harvested by centrifugation and stored at –80°C until further use. The cells were resuspended in 10 ml of PBS buffer (137mM NaCl, 2.7mM KCl, 10mM Na<sub>2</sub>HPO<sub>4</sub>, 2mM KH<sub>2</sub>PO<sub>4</sub>, 1mM DTT, 1% PMSF) and passed twice through
a French press. The cell lysate was centrifuged at 14,000 x g for 20 minutes; the supernatant was filtered through two layers of Miracloth (Calbiochem).

We used the pull-down protocol previously described by Frangioni and Neel (1993) [156]. GST sepharose beads (Novagen) were pre-washed with NETN150 buffer (0.5% NP40, 0.1 mM EDTA, 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1mM DTT, and 1mM PMSF) 4 times and equilibrated in NETN150 buffer. The beads were coated by incubating the bacterial cell extract (100 - 200 μl) containing GST (used as negative control) or GST-R\textsuperscript{411-610} with 20 μl, glutathione sepharose beads at 4°C and nutating for 1 hour. The beads were washed 4 times by nutating with 1 ml of NETN150 at 4°C (20 min/wash), and then resuspended in 20μl of NETN150.

The Gal4\textsuperscript{AD}-RIF-1C was in vitro transcribed and translated in the presence of Redivue \textsuperscript{35}S-Methionine in vitro translation grade (Amersham, Sweden) using the TNT T7/T3 coupled wheat germ extract system (Promega, WI).

The pull-down reactions were done with 20 μl samples of coated beads to which 100 μl of NETN150 buffer was added together with 5 μl of labeled Gal4\textsuperscript{AD}-RIF-1C and incubated at 4°C nutating for 1 hour. Reactions included beads loaded with GST-R\textsuperscript{411-610} and GST, as a negative control. The beads were washed 4 times (nutation at 4°C for 5 min) with 1 ml of NETN150 buffer. The bound proteins are eluted by boiling in 10 μl of 2X SDS sample buffer and visualized using Coomassie blue staining, followed by autoradiography.
4.2.8 Microprojectile bombardment and gene expression experiments.

Bombardment conditions of suspension cells and transient expression assays for luciferase and GUS were performed essentially as previously described [36, 38, 130]. For each experiment, the mass of DNA was adjusted to 10 µg with p35S::BAR to equalize the amount of 35S promoter in each bombardment. One µg of each of the regulators and 3 µg of reporter plasmid (pA1Luc or pBz1Luc) were used in each bombardment. To normalize for transformation efficiency, 3 µg of UBI::GUS was included in every bombardment, and luciferase activity was normalized to GUS activity. Each treatment was done in triplicate and entire experiments were repeated at least twice. The assays for luciferase and GUS, and the normalization of the data were done as described [36, 38, 130]. Data are expressed as the ratio of arbitrary light units (luciferase) to arbitrary units of luminescence (GUS). For the red cell counts, the cells were allowed to stay at room temperature in the dark for 48 hours. The cells accumulating anthocyanins were counted under a dissecting microscope, and then the cells were lysed and GUS assays were performed as above. The ratio of the number of red cells and units of luminescence (GUS) was multiplied by 10,000 and then plotted.

4.2.9 GFP transient assays.

Binary constructs of C-terminal GFP fusions of R, R^AbHLH, and RIF-1C, free GFP, and P19 were transformed into Agrobacterium strain GV3101 by electroporation and plated on LB gentamycin-kanamycin medium.

Agrobacterium cultures of each construct were grown in selective medium overnight at 30°C. The overnight cultures were diluted in 10mM MES, 20 µM
acetosyringone LB gentamycin-kanamycin medium and grown overnight at 30°C. Cultures were then further diluted to obtain an OD$_{600}$ 0.9-1.1 in 10 mM MgCl$_2$, 10 mM MES, 100 μM acetosyringone, and were incubated at room temperature for 2-3h. Cultures of P19 and each GFP construct was mixed 1:1 and used to infiltrate leaves of 3-4 week old *N. benthamiana* plants grown at 26°C. The leaves were then examined by confocal fluorescent microscopy 2-4 days later.

4.2.10 **RNAi analysis.**

RNA from transgenic plants was isolated using the RNAeasy kit (Qiagen, Germany) following the manufacturer’s protocol. Effectiveness of the RNAi construct was confirmed by real time PCR at the Columbus Children’s Research Institute. Briefly, the reverse transcription reaction was done with 500 ng of total RNA using SuperScript First-Strand Synthesis-System for RT-PCR (Invitrogen, CA) and the Quantitative PCR was performed using SYBR Green PCR Core Reagents (Applied Biosystems, CA). Reactions were done in triplicate and were analyzed using the ABI PRISM 7000 Sequence Detection System (Applied Biosystems, CA).

4.2.11 **Silique analysis.**

Mature siliques were harvested, their length measured, and then opened to release the seeds. The seeds were counted under the dissecting microscope and scored for abnormal phenotypes. The number of seeds per cm of silique, and the percentage of abnormal seeds per silique were calculated.
4.3 Results.

4.3.1 Mutations in the bHLH of R result in a decrease of activation of its target genes.

In order to understand the function of the bHLH we conducted transient expression assays in BMS cells to test two R mutants we had available with mutations in the bHLH region. The first was a construct of the R cDNA with a complete deletion of the bHLH (R\(^{\Delta bHLH}\)). The second was a construct with a three amino acid insertion (QPY) in the second helix of the bHLH domain, which mimics the mutation in the \(D12\) maize allele (R\(^{LcD12}\)). \(D12\) is a stable mutant allele that resulted from a footprint left over after the excision of a \(Ds\) transposable element from the bHLH region of the \(r-m1\) allele [65]. The aleurone tissue of \(D12\) kernels exhibits pale pigmentation that is the result of a dramatic decrease in the activation of the anthocyanin biosynthetic genes.

Each mutant R cDNA was co-bombarded with C1, all driven by a 35S constitutive promoter. We performed two different sets of transient assays. The first was assaying the activation of a reporter construct that had the \(A1\) promoter driving luciferase. Previous reports [37, 132, 65] have shown only a modest decrease in the activation of the reporter constructs, but we actually see an increase in activation (Fig. 4.1B). The second assay was to count the number of cells that accumulate anthocyanins 48 hours after the constructs were introduced by particle bombardment (Fig. 4.1C). The production of the pigment is a result of the activation of all the endogenous structural genes by the transcription factors introduced, and this activation occurs in the context of native chromatin. The data was normalized to the GUS activity resulting from the co-
bombardment of a GUS construct driven by the Ubiquitin promoter. As it is shown in Figure 4.1D, both the deletion and the insertion cause a dramatic reduction in the number of cells that are able to accumulate anthocyanins. This is consistent with the phenotype observed in $D12$ mutants, and indicates that the $Al$ promoter and potentially other endogenous genes that are in a chromatin context behave differently than the artificial reporter constructs in this instance.

These results also suggest that the role of the bHLH may be related to activation in the chromatin context. One way in which the bHLH could regulate chromatin structure is by recruiting a chromatin remodeling complex to the promoters bound by C1 and R. Our lab has gathered chromatin immunoprecipitation data with antibodies against methylated or acetylated Lys 9 in histone 3 (K9H3) that indicates that the promoters of these genes exhibit different histone modifications depending on whether they are actively transcribed or not. In cells and tissues where the pathway is on, the promoters of the structural genes seem to be enriched in acetylated K9H3, a hallmark of transcriptionally active chromatin. At the same time, the levels of methylated K9H3, which is usually associated with transcriptionally silent chromatin, do not show a significant change as it was expected (Morohashi and Grotewold, unpublished; Fig. 4.2). These results show that the genes of this pathway go through chromatin remodeling events as part of their regulatory mechanism.
4.3.2 RIF-1C interacts specifically with the bHLH region of R and homodimerizes through its ENT domain.

As a result of a yeast two-hybrid screen we conducted using the C-terminal region of R starting at the bHLH (amino acids 411-610), R411-610, fused to the Gal4 DNA-binding domain (Gal4DBD), and an immature tassel cDNA maize library obtained from the laboratory of Dr. Robert Schmidt (Feller and Grotewold, unpublished results), we recovered one cDNA three times, which represented the same clone and that corresponded to a protein we initially identified as R-interacting factor 1C (RIF-1C) (Fig 4.3A1). We analyzed the sequence and found that it harbors an ENT domain [157] towards the N terminus of the protein, and an Agenet domain [151] towards the C terminal region (Fig 4.5).

To further narrow down the region of interaction between R and RIF-1C, we tested by yeast two hybrid the interaction between the cDNA we obtained from the screen and constructs containing only the bHLH (amino acids 458 to 464, R458-464) or the remaining C-terminal region of R (amino acids 462 to 610, R462-610). We found that RIF-1C was able to interact with the bHLH region of R, but not with the C terminal region (Fig 4.3A).

We also generated a bHLH construct with the three amino acid insertion in the second helix of the bHLH domain present in the D12 maize allele [65]. Yeast two hybrid assays with the R1cD12458-464 showed that this mutation interferes with the RIF-1C interaction (Fig. 4.3B).
RIF-1C was able to interact with itself in yeast two hybrid experiments, suggesting that this protein is able to homodimerize (Fig 4.3C). We tested the ENT domain to ascertain if this region of the molecule was the one responsible for this homodimerization. We found that the ENT domain was sufficient to obtain interaction with the full length RIF-1C protein. Furthermore, we were able to detect this interaction using yeast two-hybrid constructs that contained only the ENT domain. Thus the ENT domain is necessary and sufficient to mediate the homodimerization of RIF-1C. This is consistent with the fact that the human protein EMSY, from which the ENT domain takes its name, has also been shown to homodimerize through this domain in crystals and in solution [158, 159].

I tried to find the region of RIF-1C that was necessary and sufficient for the interaction with the bHLH region of R (R\textsuperscript{458-464}), but all the deletion constructs tested so far have failed to interact with R\textsuperscript{458-464} (Fig. 4.5). Neither the ENT domain nor the Agenet domains were sufficient to provide an interaction with R\textsuperscript{458-464} (1 and 5 in Fig. 4.5). A construct that includes the region between these domains with either the ENT domain, or both the ENT and the Agenet domains also failed to give interaction with R\textsuperscript{458-464} (2 and 3 in Fig. 4.5). We were also unable to detect an interaction with a construct in which the only region missing was the N-terminal region upstream of the ENT domain (4 in Fig 4.5). All constructs that included the ENT domain homodimerized when we tested them with the ENT domain construct indicating that the proteins are being expressed in yeast (not shown).
To further verify the yeast two-hybrid interaction between RIF-1C and R (Fig 4.3A1), I tested the in vitro interaction of GST-R$_{411-610}$ with the product of an in vitro transcription-translation of the Gal4$^{AD}$-RIF-1C construct recovered from the screen. We found that the GST-R$_{411-610}$ was able to pull down the 35S-labeled Gal4$^{AD}$-RIF-1C, while the GST control could not (Fig 4.4).

4.3.3 RIF-1C expression profile in maize.

In order to determine the expression pattern of RIF-1C in maize, we searched the available EST data and found that RIF-1C is expressed in a wide range of tissues and developmental stages as is shown in Table 4.1.

4.3.4 RIF-1C localizes to the nucleus.

R has been shown to have three nuclear localization signals (NLSs) distributed along the protein so that one is placed at the N-terminal MIR region, another within the bHLH region, and one within the C-terminal dimerization domain (Fig. 4.1A). Previously, it was shown that the NLS in the bHLH is necessary for nuclear localization but not sufficient as it needs the presence of at least one of the other two NLSs [144]. A later report showed that elimination of the C-terminal NLS resulted in less efficient nuclear localization [160]. Since RIF-1C does not have an NLS that can be detected by computational methods, we wanted to verify whether it is capable to localize to the nucleus.

In order to see the localization pattern of RIF-1C, we made a C-terminal GFP fusion construct driven by the 35S promoter in a binary vector (see Materials and methods). We infiltrated $N. benthamiana$ leaves with a culture of Agrobacterium strain
GV101 transformed with the RIF-1C-GFP construct and then examined the fluorescence pattern in the epidermal cells. As it is obvious in Figure 4.6C, RIF-1C is restricted to the nucleus, and it forms a dotted pattern, which differs dramatically from the free GFP control (Fig. 4.6A), and resembles the pattern obtained with R-GFP (Fig. 4.6B).

Given the resemblance between the patterns obtained with the R-GFP and RIF-1C-GFP fusion proteins, we determined the localization pattern of the R mutant that lacks the bHLH. As it is shown in figure 4.6D, R^{AbHLH}-GFP is able to localize exclusively to the nucleus, and it also stays out of the nucleolus, but the pattern is changed. The clearly defined dots have almost disappeared, giving rise to a more diffused pattern. This result suggests that the pattern observed with R-GFP is dependent on its interaction with RIF-1C.

4.3.5 The plant ENT domain proteins.

BLAST analysis of the RIF-1C protein sequence revealed that it has homologs in rice and *Arabidopsis* for which no function has yet been assigned (Fig. 4.7A). In addition we found a number of proteins in *Arabidopsis* and rice that have a domain structure similar to that of RIF-1C, including the eight proteins previously identified in *Arabidopsis* by the authors that defined the ENT domain [157]. In *Arabidopsis*, six proteins exhibit the ENT-Agenet structure while in rice we were able to find three proteins that have this structure (Fig 4.7B). Other proteins containing an ENT domain exhibit variations on this structural theme: two proteins in *Arabidopsis* and one in rice have an Agenet-Agenet-ENT structure, one rice protein has two ENT domains in tandem.
located at the N-terminus and no Agenet domain, while both *Arabidopsis* and rice have one protein with an ENT domain but not an Agenet domain (Fig. 4.7B).

It is also worth noting that there are at least four genes in *Arabidopsis* that exhibit high levels of identity to RIF-1C, of which At5g13020 is the one that is most similar. At least two of these genes are found in duplicated regions of the *Arabidopsis* genome (At3g12140 and At5g06780) according to the TIGR *Arabidopsis* segmental genome duplications database.

The number of proteins in plants with an ENT domain is significant as in animals no protein other than EMSY exhibits an ENT domain [157]. Another interesting fact is that the majority of the plant ENT domain proteins also have an Agenet domain which belongs to the same super-family as the chromo domains which are found in EMSY-interacting proteins.

### 4.3.6 *Arabidopsis* has an RIF-1C homolog.

As shown in figure 4.7, *Arabidopsis* expresses a protein that shares high identity to RIF-1C, which we denominate ACK1. Although there are high levels of identity in the ENT (72%) and the Agenet (70%) domains, there is also a high level of identity in the rest of the protein (42%).

According to the microarray data available in the Genevestigator database (https://www.genevestigator.ethz.ch/at/), the expression pattern of ACK1 is very broad across organs and developmental stages, with highest expression in the inflorescence,
seed, shoot apex and senescent leaf, and during bolting (21-25 days; stage 5), silique
growth (36-45 days; stage 6.5), and silique maturation (45-50 days; stage 9), respectively
[161] (Fig. 4.8).

In yeast two-hybrid assays we determined that the ENT domain of ACK1 is also
able to homodimerize (Fig 4.9-1). In addition, we found that it can heterodimerize with a
library clone that corresponds to At3g12140, which also contains an ENT domain, in a
small yeast two hybrid screen using the ACK1\(^{\text{ENT}}\) fused to the Gal4\(^{\text{DBD}}\) (Fig. 4.9-4).
Furthermore, the expression pattern of At3g12140, correlates well with that of ACK1
(Fig. 4.8).

4.3.7 Phenotype of AtAck1 mutants.

In order to understand the function of ACK1 we decided to knock out the gene in
two ways: by using the TILLING method [162], and by RNAi. From the Arabidopsis
TILLING project we recovered several mutant alleles for the ACK1 gene, and we
analyzed three of these alleles (Fig 4.10A): \(ack1-1\) which has a mutation in the ENT
domain (D74N), \(ack1-2\) is a truncated allele as it has an early stop codon at position 150
(Q150*), and \(ack1-3\) which has a mutation in the Agenet domain (D233N) (Fig 4.10A).
When compared to the wild type background plants, \(ack1-2\) and \(ack1-3\) exhibit delayed
bolting (Fig 4.10B). All three alleles display smaller flowers than the wild type control
(Fig. 4.10C) and a significant frequency of deformation and abortion of seeds (Fig
4.10D). Due to the latter phenotype of these mutations, and since the gene that we first
isolated came from maize, a species native of the Americas, we named this gene Achikee
(Ack1) after a Quechua mythical character that killed her own daughter. The flower and
developmental delay phenotypes exhibit full penetrance, while the seed phenotype is less penetrant. Homozygous mutant plants for all three alleles have a high frequency of abortion (Fig 4.11C), but not all the seeds are aborted, in fact, some look just like wild type. This explains how the allele has been propagated.

As shown in figure 4.11A, homozygous mutant plants for all three alleles exhibit a reduction in the silique length. Since the length of the silique is proportional to the number of fertilized ovules, I also examined the silique content (Fig. 4.11B). The mutants do show a slight reduction in the number of seeds per cm of silique. Additionally, I found that even though the phenotype is observed in all three alleles, the frequency of deformation and abortion of seeds was highest in ack1-1 (Fig. 4.11C).

To determine if these mutations were dominant or recessive, I crossed ack1-1, ack1-2, or ack1-3 homozygous mutant plants with wild type plants. The crosses were done in both directions so that it could be ascertained if these phenotypes resulted from a maternal effect caused by these mutations (Table 4.2). If the mutations were recessive I was expecting to generate heterozygous plants that would be indistinguishable from wild type. We would also expect no seed deformation or abortion in the silique resulting from the cross. However, contrary to my expectations, we did recover a small number of deformed and/or aborted seeds in the F1 seed population with ack1-1 and ack1-2 (Table 4.2). In spite of the fact that the aborted seeds were seen more frequently in a mutant x wild type cross, they were also recovered from wild type x mutant crosses, thus ruling out a maternal effect. This indicates then that these mutations are semi-dominant for the weakly penetrant seed phenotype.
The developmental and flower phenotypes of \textit{ack1-2} and \textit{ack1-3}, and the \textit{ack1-3} seed phenotype, were fully complemented in the heterozygous plants. Thus \textit{ack1-3} is a recessive allele.

The RNAi approach yielded six verified transgenic lines, two of which we chose to do RT-PCR, along with an empty vector control. The results indicate that the gene was not totally knocked out in these lines as shown in figure 4.13A. It also showed us that the RNAi was specific for \textit{ACK1} in these lines, as the expression of At3g12140 (the gene that showed the highest identity at the nucleotide level in the region targeted for RNAi), was not affected.

Line RNAi6 exhibited significantly decreased silique length (Fig. 4.12A), and reduced seed content in the silique (Fig. 4.12B), indicating that in this line the seed phenotype is more severe, as it seems to result in the total absence of seed development after fertilization. With the exception of line RNAi9, most of the lines showed less frequency of deformed and/or aborted seeds than the TILLING mutants (Fig. 4.12C). Line RNAi9 exhibits deformed seeds in addition to a reduction in silique content. In both alleles however, the phenotypes are not fully penetrant. These results prompted me to check the expression of \textit{ACK1} in these lines by RT-PCR. As shown in figure 4.13B, the expression of \textit{ACK1} is lower in line RNAi6 consistent with its more severe phenotype. The levels of \textit{ACK1} expression in RNAi9 are virtually the same as in RNAi12, but unlike in RNAi12, the RNAi was not specific for \textit{ACK1}, as indicated by the reduction in the expression of At3g12140. Since I had identified At3g12140 as one of the positive clones in the yeast two-hybrid screen with the \textit{ACK1}\text{\textsuperscript{ENT}} (Fig. 4.9), it is possible that the
phenotype of RNAi9 is due to the reduction in expression of both ACK1 and At3g12140, suggesting that a heterodimer of these two proteins is functionally relevant in planta. In line RNAi6 the levels of At3g12140 expression were about the same as in RNAi9, which could also contribute to the severe phenotype. However, since I have not checked the expression of the other genes in this clade (Fig 4.7B), I cannot rule out the possibility that these effects are due to reduction in expression of those genes in lines RNAi6 and RNAi9. Interestingly in these lines there is no delay in bolting, and the flowers seem to be the same size as the empty vector line.

The fact that a modest reduction in the expression of ACK1 results in a visible phenotype, and given that the range in severity of the phenotype in transgenic lines recovered is dependent on positional effects, I reasoned that a transgenic line with a complete knock out in expression would result in a lethal phenotype. These seeds would probably be aborted, so I examined the T₀ seeds under the microscope and found that there were a significant number of aborted seeds. While empty vector control seeds exhibit an abortion or abnormal seed phenotype at a frequency of 0.32% (±0.28), the RNAi T₀ seeds had a significantly higher frequency of 1.48% (±0.5). This could mean that in cases where the gene is knocked down with high efficiency, the seed is aborted, and it does not germinate. Therefore, by plating them in selective medium, transgenic plants with moderate to low knock out efficiencies were selected. It should be noted that no obvious pigmentation, seed mucilage, seed coat morphology, or root hair phenotypes were observed in any of the Arabidopsis mutants (not shown). These are phenotypes associated with the function of R-like bHLH transcription factors in Arabidopsis.
4.4 Discussion.

We have shown that a maize ENT domain protein, RIF-1C, interacts with the bHLH region of the maize R protein. RIF-1C interaction is lost when the bHLH is mutated as in the $D12$ allele previously described [65]. We have also determined that the deletion of the bHLH as well as the $D12$ mutation have a negative effect in the accumulation of anthocyanins in BMS cells (Fig 4.1), suggesting that the bHLH is important in a chromatin context. Consistent with this idea, preliminary data from our lab indicates that the promoters of these genes are acetylated only when they are transcriptionally active (Morohashi and Grotewold, unpublished; Fig 4.2). Thus, one function of the bHLH of R could be to recruit the factors responsible for the changes in the chromatin modifications that are necessary for activation.

It is conceivable that RIF-1C is responsible for recruiting a histone acetyl transferase that modifies the chromatin in the promoter regions of the $A1$ and other anthocyanin biosynthetic genes allowing the rest of the transcriptional complex to assemble and activate. Alternatively, given the fact that the Agenet domain belongs to a family of domains that recognize methylated substrates (histones, DNA, or RNA) [151], it is possible that RIF-1C recognizes methylated H3K4, which is a mark of epigenetically active chromatin, and subsequently recruits R to the active promoter.

Since its discovery, EMSY has been described as a transcriptional repressor of BRCA2, and the fact that it interacts with known transcriptional repressors (i.e. HP1β and BS69), supports this role for EMSY [157]. It is possible that RIF-1C is a repressor that
acts via whatever mechanism the EMSY-HP1 and/or EMSY-BS69 employ. However, given the results that indicate that the bHLH of R is necessary for function, perhaps because of its interaction with RIF-1C, the R-RIF-1C complex would have to repress the expression of a repressor of the pathway. In fact, *Arabidopsis kyp2* mutants, which are defective in DNA methylation, have little or no accumulation of anthocyanins under inductive conditions (Hernandez and Grotewold unpublished). This would be consistent with a repressor of the pathway which is in turn repressed by methylation of its promoter. It is possible that the R-RIF-1C complex is involved in a chromatin silencing strategy of this repressor. Therefore, a mechanism that would prevent the R-RIF-1C complex from silencing the anthocyanin biosynthetic genes would then have to be invoked. One possibility is that at those promoters the interaction of the bHLH with RIF-2C and/or RIF-3C prevents the formation of the R-RIF-1C complex.

The ENT domain was first discovered in the human protein EMSY. In animals, no protein, other than EMSY, has been found to contain an ENT domain, but in plants there are about 6-9 ENT containing proteins (Fig 4.7B). This is one more example of gene families that have undergone a dramatic expansion in the plant kingdom [163]. More interesting is the fact that most ENT-proteins in plants also contain an Agenet domain, which belongs to the Royal super-family of domains [151]. This is significant because HP1β and BS69, two proteins known to interact with EMSY [157], contain chromo and PWWP domains respectively, and these domains also belong to the Royal super-family. This suggests that in plants a single protein has the two functions that are achieved by the interaction of two distinct proteins in animals.
The phylogenetic analysis of the ENT proteins in plants indicates the possible orthology between RIF-1C and ACK1. However, the phenotypes obtained in the \textit{ack1} mutants do not include any that would relate this protein to the known pathways regulated by GL3, EGL3, and TT8, which are the \textit{Arabidopsis} R-orthologs [64, 61, 87]. This could be an indication that the functions related to these known pathways are also regulated by a redundant ENT gene, making it necessary to knock out the function of the second gene in order to see these phenotypes. Alternatively, ACK1 may have different functions in maize and \textit{Arabidopsis}, although they may be acting by similar mechanisms.

Mutations in \textit{ACK1} are lethal, but the alleles so far recovered indicate that these mutations are not fully penetrant. The fact that the flower and developmental phenotypes behave like recessive full penetrant alleles indicates that during seed development ACK1 may have a different role than in the rest of the plant, or that the process it regulates occurs with different frequency than in other tissues and organs. One possibility to explain the random nature of this phenomenon could be related to a function of this gene in a process such as DNA repair of double strand breaks (DSB), which is also a process that requires chromatin remodeling. In fact it has been shown that events like homologous recombination, which involve DSB are not uniformly occurring in all plant organs [164], and it is developmentally regulated [165]. In humans, EMSY interacts with BRCA2, a protein that has a role in DSB repair (reviewed in [166]). Furthermore, EMSY co-localizes in nuclear foci with phosphorylated H2AX, a marker of DSB DNA damage [167, 168, 157]. RIF-1C also localizes to nuclear foci (Fig 4.6C), which could be an indication of DSB caused by the integration of the binary plasmid into the \textit{N}.
benthamiana genome. This could imply that R is also involved in DSB repair given its interaction with RIF-1C and its nuclear localization pattern (Fig 4.6B). A co-localization study is required to clarify if the nuclear patterns of R and RIF-1C coincide, but it is possible that this is a new function of R, which is probably redundant in maize, thus explaining the lack of a phenotype related to DNA repair in maize R mutants.

Arabidopsis has two BRCA2 orthologs, which are essential for the repair of DSB during homologous recombination in meiosis [169]. Moreover, both AtBRCA2 proteins interact with AtRad51, AtDmc1, and two AtDss1 isoforms which are known to have key roles in homologous recombination mechanisms [170]. We have not tested if AtBRCA2 directly interacts with ACK1, however if they interact, it is probably not through the ENT domain as the TAD domain present in the human BRCA2, which mediates the interaction with the EMSY ENT domain, is not present in AtBRCA2. A more likely scenario would be that ACK1 interacts with an AtBRCA2-interacting protein.

In addition to the domains already described, it seems like ACK1 also harbors a sequence that is reminiscent of a PHD finger domain albeit the typical cysteine residues found in PHD fingers are missing in the ACK1 proteins (Fig. 4.14). This region of the ACK1 protein is extremely well conserved across species and among the Arabidopsis proteins of similar structural organization (Fig. 4.7A). The PHD finger proteins have been implicated to be members of chromatin remodeling complexes where they interact with proteins belonging to the Royal super-family [171]. Moreover, many Agenet domain proteins in Arabidopsis also contain a PHD domain [151]. More recently, PHD domains have been shown to bind phosphoinositides and it has been suggested that this signaling
molecule is involved in the regulation of the chromatin remodeling machinery [172], [173]. In fact, the *Arabidopsis* protein ATX1, which we used to discover the PHD-like domain in ACK1, has been shown to be involved in the methylation of H3K4, H3K9, and H3K27 [174]. This provides another possible link between ACK1 and chromatin remodeling.

All the data we have gathered for RIF-1C and ACK1 point to these proteins involvement in epigenetic control. It is possible that they play a role in transcription in some plant species and a role in DNA repair in other species; however, it is more likely that they are involved in both processes, using similar mechanisms, in all plant species, with spatial and temporal variations. It has been well documented that DNA repair and transcriptional regulation use similar mechanisms to deal with chromatin structure, via a ‘histone code’ [175, 176, 149]). The fact that plants have several of these ENT proteins, may indicate that their genomes are subject to a higher frequency of DSB damage which could be caused by events such as transposon insertions and excisions, which involve DSB DNA repair. Many experiments need to be done to confirm the function of the RIF-1C and ACK1 proteins, but the evidence so far supports the possibility that they are a convergent point for epigenetic transcriptional control and DNA repair.
Figure 4.1: Effect of mutations in the bHLH domain of R in transient expression assays.

A. Domain structure of the R protein. The nuclear localization signals are indicated. B. Percent fold activation of C1 and R compared to C1 and R$^{\Delta bHLH}$, and C1 and R$^{\text{LcD12}}$ on a luciferase reporter constructs driven by the A1 promoter (pA1Luc). Constructs of C1, R, R$^{\Delta bHLH}$, or R$^{\text{LcD12}}$ were driven by the constitutive CaMV 35S promoter and co-bombarded in BMS maize cells. Assays were done as described in figure 3.2. C. BMS cells bombarded with the overexpression constructs indicated. After 48 hours anthocyanins start accumulating if the pathway is activated. D. Quantification of pigment accumulating cells. Cells accumulating anthocyanins after co-bombardment of the constructs indicated and a GUS construct driven by the ubiquitin promoter, were counted. Proteins were then extracted from the total cell sample and assayed for GUS activity to normalize the data. The graph indicates the ratio of number of red cells to GUS activity.
Figure 4.1
Figure 4.2: Chromatin immunoprecipitation analysis of the promoters of the maize anthocyanin biosynthetic genes.

BMS cells (unpigmented), BMS cells expressing R and C1 from a constitutive promoter (35S), thus pigmented (A), and leaf tissues from green (b1, Pl1), and pigmented (B-l, B-Peru, Pl1) plants (B) were crosslinked and proteins were extracted under non-denaturing conditions. The immunoprecipitations were done with antibodies against histone 3 (H3) as a positive control, acetylated (Ac) and dimethylated (Me) lysine 9 in histone 3, and a non specific IgG (Ig) or no antibody (-) were used as negative controls. PCR reactions were done with uncrosslinked immunoprecipitates using primers to amplify fragments between 200 and 500 bp from the dehydroflavonol reductase (A1). The actin gene was used as normalizing control. A positive control for the PCR reactions was done with input total extract (In). These data were generated by Dr. Kengo Morohashi who has kindly allowed me to include them in this document.
Figure 4.3: Yeast two-hybrid analysis of RIF-1C.

A. Interaction of Gal4<sup>AD</sup>-RIF-1C library construct with the C-terminal region of R including the bHLH (R<sup>411-610</sup>), the C-terminal region of R excluding the bHLH (R<sup>362-610</sup>), and the R bHLH domain alone (R<sup>411-464</sup>). All R constructs are fused to the Gal4<sup>DBD</sup> at the N-terminus in pBD-GAL4, which is the plasmid used in the empty vector control. B. Loss of interaction with RIF-1C when the bHLH of R contains the D12 allele sequence (R<sup>LcD12411-464</sup>). C. Homodimerization of the ENT domain of RIF-1C (RIF-1C<sup>ENT</sup>) with the full length protein and itself. Empty vector corresponds to pBD-GAL4 if listed on the left, or pAD-GAL4 if listed on the right. For the yeast strain used and the selections applied see figure 3.4.
Figure 4.3
Figure 4.4: Interaction of RIF-1C and the C-terminus of R \textit{in vitro}.

Autoradiogram and stained gel of an SDS PAGE of a GST pull-down using GST-R^{411-610} (including the bHLH) as bait (lane 2) with an \textit{in vitro} transcribed and translated Gal4^{AD}-RIF-1C. Gal4^{AD}-RIF-1C was radiolabelled with \[^{35}\text{S}\] Methionine and shown in the input lane (lane 1). GST alone was used as a negative control (lane 3).
Figure 4.5: Domain structure and dissection of RIF-1C.

The ENT and Agenet domains are indicated and the amino acid numbers that limit each domain are shown. A diagram of the deletion constructs for yeast-two hybrid experiments is indicated below. Each fragment indicated was cloned in to pAD-GAL4 and pBD-GAL4 and tested with R\(^{411-610}\), R\(^{458-464}\), and RIF-1C\(^{ENT}\).
Figure 4.6: Nuclear localization of RIF-1C.

Bright field and fluorescent micrographs of Agro-infiltrated *N. benthamiana* leaves transiently expressing free GFP (A), or C-terminal GFP fusions to R (B), RIF-1C (C), or R^ΔbHLH (D). The bar corresponds to 10 μm.
Figure 4.7: ENT domain proteins in plants.

A. ClustalW alignment of RIF-1C and its homologs in rice (OsRIF-1C, P0670E08.30) and *Arabidopsis* (ACK1, AT5G12030). The ENT domain is highlighted in pink, and the Agenet domain in green. B. Phylogenetic tree of RIF-1C and the ENT proteins in *Arabidopsis* and rice. The human EMSY protein was also included. The domain structure of the plant proteins is shown.
Figure 4.8: Expression of ACK1 and closely related genes in *Arabidopsis* organs and developmental stages.

Data obtained from the Genevestigator database (https://www.genevestigator.ethz.ch/) [161]. The color scale of the heat maps is shown.
Figure 4.9: Yeast two-hybrid analysis of the ENT domain of ACK1.

Homodimerization of the ENT domain of ACK1 (ACK1$^{\text{ENT}}$), and its interaction with a library construct of Gal4$^{\text{AD}}$At3g12140. Empty vector corresponds to pBD-GAL4 if listed on the left, or pAD-GAL4 if listed on the right. Strain and selection as in figure 3.4.
Figure 4.10: Mutant alleles of ACK1.

A. Diagram of ACK1 with the mutations present in each mutant allele (ack1-1, ack1-2, and ack1-3) indicated. B. Plant growth and development phenotype. C. Flower size phenotype. D. Seed morphology phenotype, including SEM micrographs.
Figure 4.10
Figure 4.11: Silique and seed analysis of ACK1 mutants.

A. Silique length of wild type and mutant plants. B. Silique content expressed as the ratio between the number of seeds and the silique length. C. Percent of deformed seeds (aborted and malformed) per silique. Pink bars indicate average for siliques taken from a group of plants, red bars indicate the average for siliques in one plant.
Figure 4.12: Silique and seed analysis of ACK1 RNAi transgenic lines.

A. Silique length of transgenic plants expressing empty vector and the RNAi construct driven by the constitutive CaMV 35S promoter. B. Silique content expressed as the ratio between the number of seeds and the silique length. C. Percent of deformed seeds (aborted and malformed) per silique.
Figure 4.12
Figure 4.13: Analysis of ACK1 expression in RNAi lines by real time PCR.

The expression of At3g12140 was used as a control to check the silencing specificity of the RNAi construct. Signal was normalized with the expression of actin. Lines RNAi3 and RNAi12 were analyzed in a separate experiment (A) from lines RNAi6 and RNAi9 (B).
Figure 4.14: Protein alignment of the PHD domain of ATX1 and ACK1.

Identical residues are labeled in red and similar residues in blue. The underlined residues in ACK1 correspond to the Agenet domain.
<table>
<thead>
<tr>
<th>Genebank Accession Number</th>
<th>Tissue</th>
<th>Cultivar</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI372386</td>
<td>Glume ear: 2 weeks post-pollination</td>
<td>W23-TGA</td>
</tr>
<tr>
<td>AW065539</td>
<td>Root: 3-4 days old</td>
<td>W23</td>
</tr>
<tr>
<td>AW313403</td>
<td>Tassel: length from 0.1 to 2.5 cm</td>
<td>OH43</td>
</tr>
<tr>
<td>BG517340</td>
<td>Seedling - leaf and stem, including leaf base: 2 week old</td>
<td>B73</td>
</tr>
<tr>
<td>BG842427</td>
<td>Seedling and silk</td>
<td>B73</td>
</tr>
<tr>
<td>BM499965</td>
<td>BMS cells in culture: mixed logarithmic and stationary growth phases</td>
<td>BMS (Black Mexican Sweet)</td>
</tr>
<tr>
<td>BQ547713</td>
<td>Inflorescence meristem - floral organ primordia: 0.5 cm to 2 cm</td>
<td>OH43</td>
</tr>
<tr>
<td>CB886008</td>
<td>Ear: 2 mm</td>
<td>B73</td>
</tr>
<tr>
<td>CD440725</td>
<td>Endosperm: 7-23DAP</td>
<td>W22</td>
</tr>
<tr>
<td>CD447698</td>
<td>Endosperm: 4-6 DAP</td>
<td>W22</td>
</tr>
<tr>
<td>CD977022</td>
<td>Pericarp</td>
<td>F2</td>
</tr>
<tr>
<td>CD999520</td>
<td>Ear leaf</td>
<td>F2</td>
</tr>
<tr>
<td>CF005214</td>
<td>Sheath</td>
<td>F2</td>
</tr>
<tr>
<td>CF349269</td>
<td>Embryo sac cDNA</td>
<td>A188</td>
</tr>
<tr>
<td>DN222658</td>
<td>Vegetative Shoot Apical Meristem (SAM) and leaf primordia staged P1-P4</td>
<td>B73</td>
</tr>
</tbody>
</table>

Table 4.1: RIF-1C expression profile according to available EST data.
<table>
<thead>
<tr>
<th>Cross</th>
<th>Silique length (cm)</th>
<th>Total No of seeds</th>
<th>No of abnormal seeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT x ack1-1</td>
<td>1.0</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>n.d.</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>ack1-1 x WT</td>
<td>0.3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>WT x ack1-2</td>
<td>0.6</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>n.d.</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Col x ack1-2</td>
<td>1.1</td>
<td>33</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>ack1-2 x WT</td>
<td>1.1</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>WT x ack1-3</td>
<td>0.4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>ack1-3 x WT</td>
<td>0.3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>16</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.2: Results of crosses between homozygous wild type background plants (WT) or wild type Columbia plants (Col), and homozygous mutant ack1-1, ack1-2, ack1-3 plants.
CHAPTER 5

FINAL DISCUSSION

In this study the maize flavonoid pathway has been used as a system to investigate the details of combinatorial transcriptional regulation in plants. This pathway has been well characterized and many mutants in both structural and regulatory genes are available. In addition, it provided an ideal system to address the question of biological specificity, as two very closely related R2R3-MYB domain genes regulate distinct branches of the pathway. We were able to use genetic, molecular and biochemical tools to dissect out some intricacies of the regulation of the pathway.

The three main conclusions of this study were: a) The specificity of the R2R3-MYB protein C1 relies on its interaction with bHLH R, and it is because of this interaction that C1, but not P1, can activate the anthocyanin branch of the pathway; b) the bHLH R contributes to the activation of the target genes by making additional direct or indirect DNA contacts with the promoter (R-enhanced activity), and by relieving C1 from a plant inhibitory effect that prevents it from binding and activating its target genes (R-dependent activity); c) the bHLH domain of R seems to be involved in a epigenetic strategy of regulation of the pathway, and it interacts with at least one factor harboring domains that implicates it in chromatin remodeling activities.
5.1 The MYB-bHLH interaction and the implications for identifying other regulatory complexes.

We identified four residues in the R3 of C1 that are necessary for the interaction with R, and when transferred to P1, the regulator of the phlobaphene branch of the pathway, the resulting mutant protein was able to interact with R. An additional two changes in R3 were necessary to create a protein (P1*) that is able to activate the genes in the anthocyanin branch of the pathway [130]. We found that these residues are very well conserved in all the C1 orthologs identified to date. In fact, a recent survey of interactions between *Arabidopsis* MYB proteins and bHLH proteins containing a MYB interaction region (MIR), allowed the authors to identify a consensus amino acid signature in the MYB R3 responsible for the interaction with the MIR: \( \text{D/E-L-X}^2\text{-R/K-X}^3\text{-L-X}^6\text{-L-X}^3\text{-R} \) [177]. This motif includes three of the four residues we identified to be important for the interaction in C1, and two we deemed important for activation of transcription (L-X\(^2\)-R-X\(^2\)-R-L-X\(^6\)-L-X\(^2\)-G-R). The discrepancies between our work and the *Arabidopsis* experiments include residues 16 and 19 (underlined in the sequence above; numbering according to Zimmermann et al. 2004 [177]) which when mutated in PAP1 (the C1 ortholog in *Arabidopsis*) did not have a major effect in either interaction with EGL3 (one *Arabidopsis* R ortholog) nor activation of the DFR promoter, and residue 33 (also underlined) which was shown to affect interaction with EGL3 as well as DFR activation. These results can be explained if one considers that there may be small differences between EGL3 and R that would account for the interaction dependencies on the above mentioned residues. Indeed, the PAP1 point mutants were only tested with EGL3;
perhaps these mutants lost the ability to interact with GL3, TT8, R or other unknown MIR bHLH, or alternatively, gained new interactions. It is possible that this consensus includes residues that mediate interaction to MIR domains in general, and residues that determine MIR interaction specificity. Thus, the residues involved in the specificity of the interaction, as opposed to those necessary for the interaction, were not distinguished in our study. In addition, the nature of our assays were different as we were working within the context of the P1 MYB domain, which already shared some of the residues in the consensus signature. Finally, it should be noted that considering these residues singly is possibly wrong as the six residues we changed to generate P1* seem to be co-evolving (Braun and Grotewold unpublished data). The fact that some residues in the consensus are conserved in non-interacting MYB proteins, including residues that were deemed essential for stability of the interaction with MIR bHLHs (12, 29, and 33; underlined above in the motif identified by Zimmerman et al.) supports the fact that the entire combination has to be considered and not each residue separately. For example, changes in position 19 which is highly variable within MIR interacting MYB proteins are always accompanied by a change in position 16 to lysine (see Fig 4 in [177]).

The finding of this consensus is important as predictions can now be made about interactions with R-like bHLH proteins as was shown for AtMYB82 and AtMYB4 [177]. However, this can be taken a step further if one considers the entire motif and classifies MYB domains based on this consensus, including groups that are not closely related to C1 but which include at least one member that has been shown to regulate genes together with a MIR bHLH, such as MYB group 17 (to which ATR1 belongs), and MYB group 20.
(which includes AtMYB2). Making predictions about the regulatory complexes in which the MIR bHLH proteins (groups IIId-IIIif [56]) participate is now possible, and will allow us to start testing other MYB-bHLH-WD40 networks to gain more insight into the regulatory network not only in Arabidopsis but in other plants as well.

5.2 The R-enhanced and the R-dependent activities may be conserved in all plant species.

It is clear that the protein-protein interactions that allow the C1/R/PAC1 complex to assemble have been conserved in all plant species that have been examined so far. We have determined that the interaction with R is key for the regulatory specificity of C1, but we also have determined that a second component in determining the specificity of the regulatory complex is the presence of cis-DNA elements in the promoters of the target genes. The ARE element has been identified in all the anthocyanin structural genes in maize [131, 132, 140, 178], and has been shown to play an important role in the regulation by C1 and R in vivo [133]. We have shown that this element is important in the R-enhanced activity and we proposed that R recruits a factor that binds the ARE, although it is not clear if direct R binding would also be necessary [119]. An ARE sequence has been identified in Gerbera hybrida and transient expression assays using reporters with mutations in this DNA cis-element, showed that this element is important for the activation of the GDFR gene [142]. The conservation of the ARE in other species indicates that the factors binding this element are probably conserved as well, which is to be expected given that every other factor involved in the regulation of this pathway has
been conserved in all the species so far investigated. This also implies that the R-enhanced activity is a feature its orthologs in other species. The ARE sequence has not been identified in the Arabidopsis flavonoid biosynthetic genes yet, which could be indicating that in this instance, the R-like factors lost their ability to recruit the ARE binding factor allowing for mutations to accumulate, or that a different kind of protein, with a different binding site is recruited. However, in Arabidopsis other cis-DNA elements that have been identified in the early genes of the pathway such as CHS, CFI, and F3H, are included in a 40 bp long conserved non-coding sequence present in the CHS promoters of Arabidopsis, maize, and rice (Hernandez and Grotewold unpublished data). The cis-DNA elements are a MYB binding site, a CG box and an ACE (ACGT-containing element) which are proposed to be bound by a MYB domain transcription factor (similar to C1 or P1), a bHLH (R-like protein), and possibly a bZIP factor respectively [179]. This indicates that the conservation of the regulation of this pathway is not limited to protein-protein interactions, but also includes protein-DNA interactions.

The dependency of C1 on R is also a characteristic shared by all the C1 homologs identified in other plant species. We proposed that the R relieves C1 from a plant specific inhibitory effect [119]. We have recently gathered some initial GFP localization data that strongly suggests that, at least in tobacco, C1 does not depend on R for nuclear localization. The new data also argues against a problem with protein stability (Feller and Grotewold unpublished data). Thus, the most likely cause for the dependency of C1 on R is the presence of an inhibitory protein that interferes with C1 activity. As mentioned in chapter 2, known candidates for this inhibitor are IN1, an R-like bHLH, or a MYB
protein similar to MYBx from petunia [145], CPC and TRY from Arabidopsis [146], or FaMYB1 from strawberry [52]. So, it is possible that the need for having an inhibitory mechanism is a characteristic that has been conserved in all plant species, but the mechanism of inhibition may be different in different species.

5.3 **R-interacting factors.**

We have proposed that R achieves its functions with the help of other factors it recruits to the regulatory complex. This is supported by the multiple domains that comprise the R protein. One such factor is PAC1, a WD repeat protein, which interacts with the acidic region of R (Oh and Grotewold unpublished data) and has been shown to participate in the regulation of the anthocyanin pathway [73]. Homologs in petunia [68] and Arabidopsis [71] have also been shown to be necessary for pigment biosynthesis. In order to identify other factors in this complex, we have done yeast-two hybrid screens and found a number of proteins that interact with R. These include a coiled coil protein that interacts with the MIR and we have designated R interacting factor 1N (RIF-1N), RIF-1C which is the ENT protein described in chapter 4, that interacts with the bHLH domain of R, and two bHLH proteins, RIF-2C and RIF-3C, which interact with the C-terminal region of R including the bHLH. In addition, we have found that the last 85 amino acids of R comprise an ACT-like domain that mediates homodimerization of R and we have evidence that this domain plays a role in the regulation of the anthocyanin pathway, but it does not affect the ARE dependent activity (Feller et al. 2006. Manuscript submitted). Whether all of these factors play a role in the known functions of R is still
unknown, but a tempting hypothesis is that either RIF-2C, or RIF-3C, or both could be mediating the R-enhanced activity through the ARE. RIF-3C also has an ACT-like domain in the very C-terminus and the interaction with R requires both the bHLH and the ACT-like domain, as suggested by yeast-two hybrid data (Feller and Grotewold, unpublished data), making this a candidate worth pursuing.

5.4 Are RIF-1C and R involved in gene silencing?

The function of RIF-1C is not clear yet, but the data we have gathered provides some evidence for the participation of the R bHLH, the region of R with which RIF-1C interacts, in chromatin-level transcriptional control mechanisms. The presence of domains related to chromatin remodeling in RIF-1C are consistent with this functions. We also have evidence that indicates that the promoters of the genes of this pathway undergo changes in chromatin remodeling when the genes are active (Morohashi and Grotewold, unpublished data). Recently it has been reported that in mutants of the Arabidopsis Heterochromatin Protein 1 (LHP1), a component of both heterochromatin and euchromatin [180], [181], the accumulation of anthocyanins as well as syanopoyl esters were decreased after induction by either jasmonic acid treatment or heat treatment [182]. This is consistent with our observation that a mutation in the Arabidopsis KRYPTONITE gene (kyp2) causes a decrease in the accumulation of anthocyanins under inductive conditions (Hernandez and Grotewold, unpublished data). KRYPTONITE is an Arabidopsis histone 3 (H3) methyltransferase that specifically acts on the Lys 9 residue of H3, a modification that is recognized and bound by LHP1, which in turn recruits the
DNA methyltransferase CHROMOMETHYLASE 3 (CMT3) [183]. This would mean that when the cell has deficiencies in this gene silencing mechanism, the anthocyanin pathway is shut down or operates at very low levels. One explanation that would be consistent with our data is that R and RIF-1C control the expression of an inhibitor of the pathway by silencing its promoter through DNA methylation. This mechanism would be especially important when the plant needs to synthesize large amounts of anthocyanins under inductive conditions. One could even speculate that the inhibitor that makes C1 dependent on R may be the target of negative regulation by a putative R-RIF-1C complex. Since the human protein EMSY is able to interact with HP1β, it is conceivable that RIF-1C may be able to recruit LHP1 to target promoters. However, preliminary yeast-two hybrid data carried out in collaboration with the laboratory of Valerie Gaudin provides no support for this hypothesis that this is not the case (Hernandez and Grotewold unpublished data). Alternatively, RIF-1C may be able to directly recruit CMT3 thus sharing some functional redundancy with LHP1. The Arabidopsis homolog of RIF-1C, ACK1, may be acting through similar mechanisms in a different set of target genes, or it may include the genes of the anthocyanin pathway but because they are also targets of a redundant ACK1-like protein, our mutants have no anthocyanin accumulation phenotype.

5.5 Final remarks.

Significant progress has been made toward understanding of the mechanism by which the C1-R complex regulates the anthocyanin pathway in maize, or how the RIFs
and PAC1 contribute to the regulatory complex, yet much remains to be learned. However, while PAC1 and these putative new players have given us new leads to follow, P1* and the mutations in the promoters of the structural genes have provided us with excellent tools to understand and test regulatory mechanisms. Due to the high degree of conservation of the regulatory and structural genes of the pathway, the new players we identify in the regulation of this pathway in maize will undoubtedly allow us to discover new layers of regulatory specificity that are the result of this combinatorial regulation of the anthocyanin pathway in plants.
LIST OF REFERENCES


83. Sagasser M, Lu G-H, Hahlbrock K, Weisshaar B: **A. thaliana TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins.** *Genes and Development* 2002, 16:138-149.


115. Hegvold AB, Gabrielsen OS: **The importance of the linker connecting the repeats of the c-Myb oncoprotein may be due to a positioning function.** *Nuc. Acids Res.* 1996, **24**:3990-3995.


117. Szymanski DB, Jilk RA, Pollock SM, Marks D: **Control of GL2 expression in Arabidopsis leaves and trichomes.** *Development* 1998, **125**:1161-1171.


163. Shiu SH, Shih MC, Li WH: **Transcription factor families have much higher expansion rates in plants than in animals.** Plant Physiol. 2005, **139**:18-26.

164. Boyko A, Filkowski J, Hudson D, Kovalchuk I: **Homologous recombination in plants is organ specific.** Mutat Res 2006, **25**.


166. Shamoo Y: **Structural insights into BRCA2 function.** Curr Opin Struct Biol. 2003, **13**:206-211.


