University of Kentucky
UKnowledge

Theses and Dissertations--Plant and Soil Sciences

Plant and Soil Sciences

2013

DNA-BINDING SITE RECOGNITION BY bHLH AND MADS-DOMAIN TRANSCRIPTION FACTORS

Joshua R. Werkman
University of Kentucky, joshua.werkman@gmail.com

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Werkman, Joshua R., "DNA-BINDING SITE RECOGNITION BY bHLH AND MADS-DOMAIN TRANSCRIPTION FACTORS" (2013). *Theses and Dissertations--Plant and Soil Sciences*. 29.
https://uknowledge.uky.edu/pss_etds/29

This Doctoral Dissertation is brought to you for free and open access by the Plant and Soil Sciences at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Plant and Soil Sciences by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statements(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the non-exclusive license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Joshua R. Werkman, Student

Dr. Ling Yuan, Major Professor

Dr. Arthur G. Hunt, Director of Graduate Studies

DNA-BINDING SITE RECOGNITION BY bHLH AND MADS-DOMAIN TRANSCRIPTION FACTORS

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Agriculture
at the University of Kentucky

By
Joshua Robert Werkman

Lexington, Kentucky

Director: Dr. Ling Yuan, Associate Professor, Plant and Soil Sciences

Lexington, Kentucky

2013

Copyright © Joshua R. Werkman 2013

ABSTRACT OF DISSERTATION

DNA-BINDING SITE RECOGNITION
BY
bHLH AND MADS-DOMAIN TRANSCRIPTION FACTORS

Herewithin, two transcription factor (TF) regulatory complexes were investigated. A bHLH–MYB–WDR (BMW) DNA-binding complex from maize was the first complex to be studied. R, a maize bHLH involved in the activation of genes in the anthocyanin pathway, had been characterized to indirectly bind DNA despite the presence of a functional DNA-binding domain. Findings presented here reveal that this is only partially correct. Direct DNA-binding by R was found to be dependent upon two distinct dimerization domains that function as a switch. This switch-like mechanism allows R to be repurposed for the activation of promoters of differing *cis*-element structure.

The second regulatory complex studied was of the *Arabidopsis thaliana* MIKC-MADS TF family. For many TFs, DNA-binding site recognition is relatively straightforward and very sequence specific, while others exhibit relaxed sequence specificity. MADS-domain TFs are one family of TFs with a wider range of *cis*-element sequences. Though consensus *cis*-element sequences have been determined for various MADS-domains, correctly predicting and identifying biologically functional *cis*-elements has been a challenge. In order to study the influence of nucleobase associations within the *cis*-element, a DNA-Protein Interaction (DPI)-ELISA method was modified and optimized to screen a panel of specific probes. Screening of the SEP3 homodimer against a panel of sequential, palindromic probes revealed that nucleobases in position -1:+1 of the CArG-box influence binding strength between the MADS-domain and DNA. Additionally, the specificity of AGL15 towards CT-W6-AG forms was discovered to be determined by the functional groups present in the minor groove at position -4:+4 using inosine:cytosine (I:C) base pairs.

Finally, the FLC–SVP MADS-domain heterodimer, bound to a native *cis*-element, was modeled and binding simulated using molecular dynamics. In conjunction with simulations of AGL15 and SEP3 homodimers, a potential binding mechanism was identified for this unique heterodimer. DNA sequence recognition by the MADS-domain was found to occur asymmetrically. In the case of the FLC–SVP heterodimer, the direction of asymmetrical DNA-binding in heterodimers was found to be fixed. Furthermore, the molecular dynamics simulations provided insight towards understanding the results generated from previous DPI-ELISA experiments, which

should provide an improved means for predicting biologically significant CArG-boxes around genes.

KEYWORDS: DNA-BINDING SITE RECOGNITION, LEUCINE ZIPPER-LIKE DOMAIN, TRANSCRIPTIONAL SWITCH, DPI-ELISA/TF-EIA, MOLECULAR DYNAMICS

Joshua R. Werkman

Student's Signature

September 13, 2013

Date

DNA-BINDING SITE RECOGNITION BY bHLH AND MADS-DOMAIN TRANSCRIPTION FACTORS

By

Joshua Robert Werkman

Dr. Ling Yuan

Director of Dissertation

Dr. Arthur G. Hunt

Director of Graduate Studies

September 13, 2013

Date

In memory of my mother and in honor of my father
for the many sacrifices made for my brothers and I
and for unconditionally loving us.

ACKNOWLEDGEMENTS

First and foremost, I wish to gratefully thank Dr. Ling Yuan for his enduring patience, understanding, and help given to me over the years. During the past several years in his lab, he has provided an environment for independent learning, which allowed me to ask wild questions and learn through my failures. Through this, and despite butting heads at times, I believe that we grew both personally and in our knowledge and understanding of things scientific. I am glad to be able to call him my friend.

I want to thank Dr. Sitakanta 'Pintu' Pattanaik for the friendship and assistance he has provided through the years. In particular, he provided much needed support with protoplast assays and yeast two-hybrids, as well as help with some of the constructs. During the time that I worked with Dr. Barunava 'Barun' Patra, he has provided me insight into alternate research techniques and thinking outside the box. Barun had also assisted greatly me with the EMSAs. Additionally, I want thank all the members of Ling's lab since beginning grad school for the rich experiences we've had together. You have helped broaden my understanding of different cultures.

Dr. Sharyn Perry of the University of Kentucky has graciously provided me with insight into AGL15 protein production ─ thank-you.

Dr. Erich Grotewold of Ohio State University has my gratitude and thanks for the collaboration and many hours of tireless work between his lab and the Yuan lab.

Thomas Cheatham III provided instrumental guidance into constructing the code necessary to run my experiments. He has graciously made himself available to answer my questions in the areas that I did not have complete understanding.

Dr. Michael Sheetz was helpful to me when getting setup to use the Lipscomb HPC Cluster. Additionally, I would like to thank the UK Information Technology Department and Center for Computational Sciences for the use the Lipscomb HPC Cluster.

My gratitude to the Department of Plant and Soil Sciences and the University of Kentucky for the opportunity and funding to pursue my degree and my committee members for additional thoughts, insight, and patience.

Dr. James McConnell from the University of Guam for taking a young, inquisitive teenager under your wing and teaching me plant tissue culture, your mentorship will always be remembered. Wendy O'Donovan from Oglevee's, thank you for the opportunity to be part of the Oglevee family and providing the insight into the commercial aspects of the field. Dr. Charles Detwiler, professor and mentor in microbiology and life during my undergraduate years at Liberty University. Dr. Kelly Stefano-Cole at the University of Pittsburgh for the experience as a lab tech.

My deepest thanks go to the Smalle Lab for the use of GraphPad Prism. I will always remember fondly the many chats with Dr. Jasmina 'Mina' Kurepa. I learned so much through our years of friendship.

My mom, Mary Beth, and my dad, George, for teaching me to never let go of your dreams and providing me with the means to pursue those dreams. For teaching me how to grow plants and importance of family. My brothers, Jason and Justin, for their patience with me while growing up when I would talk forever about science.

Last but not least, my husband, Brandin Parrett. Without his help and support, his devotion and love, this journey would have been too much to bear. He has been there for me when my mother passed away and those long months and years after. He stays up into the middle of the night when a project goes into crunch time, even though I tell him to go to bed. There have been so many ways he has helped me through the struggles of this journey and I can only hope that I have and can provide him with the same help and support, devotion and love that he has shown me.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Review and Introduction

Many agronomically and ecologically valuable traits can be attributed to transcription factors (TFs) containing basic–helix–loop–helix (bHLH) domains or MADS domains. Partial functional redundancy within the group III*f* bHLHs in Arabidopsis allows them to regulate key enzymes in the flavonoid biosynthetic pathway, as well as, genes involved in trichome and root hair initiation and development. MADS-domain containing TFs most notable roles include the regulation of flowering time, flower differentiation, and fruit development.

What are transcription factors and their complexes?

TFs are proteins that typically bind unique DNA sequences or *cis*-elements in or near the promoter region of a gene through their DNA-binding domain (DBD) and consequently activate or repress transcription. *Cis*-elements are generally unique to a particular type of TF; however, they can be shared if, for example, two TF families share similar structure in their DBDs. With well over 2000 identified TFs and nearly 29,000 genes (protein coding and non-coding RNAs) [1] in *Arabidopsis thaliana*, a model plant related to broccoli and cabbage, most, if not all, TFs function in regulatory complexes and generate expansive regulatory networks. These TFs have been organized into 50 families, though many share similarly functioning domains [1].

The bHLH–MYB–WDR complex

Anthocyanin accumulation has been the target for the inquisitive naturalist (Figure 1.1). As a phenotype, the ease of detecting changes has contributed to its appeal and has made it the subject for both beauty and utility. In maize, the leaf color trait, *Lc*, is a trait in which the leaves of the plant turned red due to the accumulation of anthocyanins. The trait was isolated from a South American originating strain, Ecuador 1172, and was found to map near the *R* locus (RED1) [2-4]. In 1989, the gene at the *R* locus was identified as a bHLH-domain containing protein [5].

R protein sequence clusters with the group III*f* bHLHs identified in Arabidopsis by both sequence and function. These include AtMYC1 [6-8], AtGL3 (GLABRA 3) [9,10], AtEGL3 (ENHANCER of GLABRA 3) [11,12], and AtTT8 (TRANSPARENT TESTA 8)

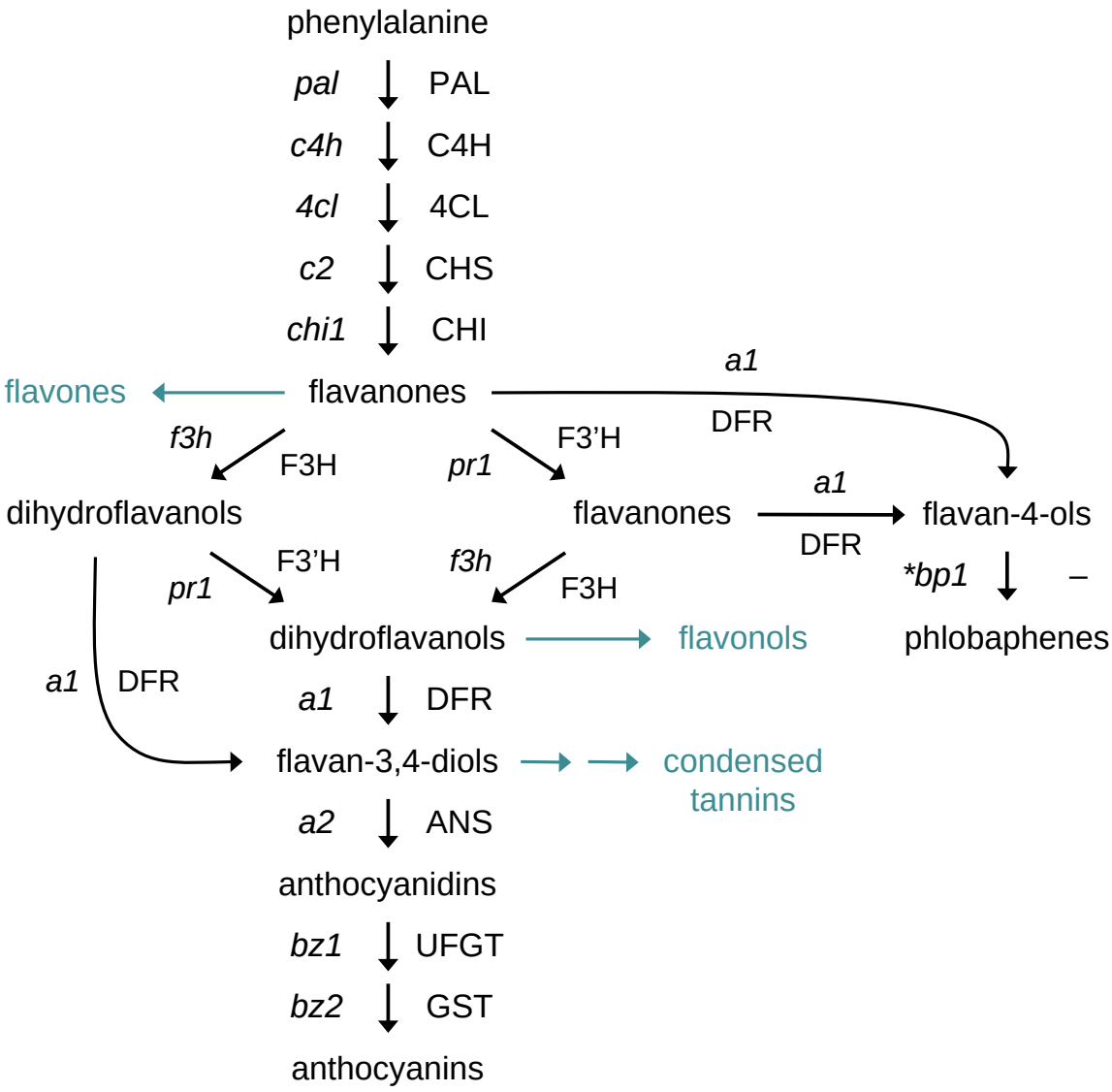


Figure 1.1 Anthocyanin biosynthetic pathway in maize. Beginning with phenylalanine to produce flavanones, a variety of products can be generated from the anthocyanin biosynthetic pathway. Abbreviations: phenylalanine ammonia lyase (*pal*, PAL); cinnamate 4-hydroxylase (*c4h*, C4H); 4-coumarate:CoA ligase (*4cl*, 4CL); chalcone synthase (*c2*, CHS); chalcone isomerase (*chi1*, CHI); flavanone 3-hydroxylase (*f3h*, F3H); flavonoid 3'-hydroxylase (*pr1*, F3'H); dihydroflavonol 4-reductase (*a1*, DFR); anthocyanidin synthase (*a2*, ANS); UDP-glucose:flavonoid O-glucosyltranserase (*bz1*, UFGT); glutathione S-transferase (*bz2*, GST); brown pericarp (*bp1*) has yet to be characterized.

[13,14]. Group III*f* bHLHs are multi-domain proteins comprised of a MYB-interacting region (MIR) [15-18], an acidic domain, a bHLH domain, and a C-terminal ACT (aspartate kinase, chorismate mutase and TyrA) domain [19] (Figure 1.2A). As

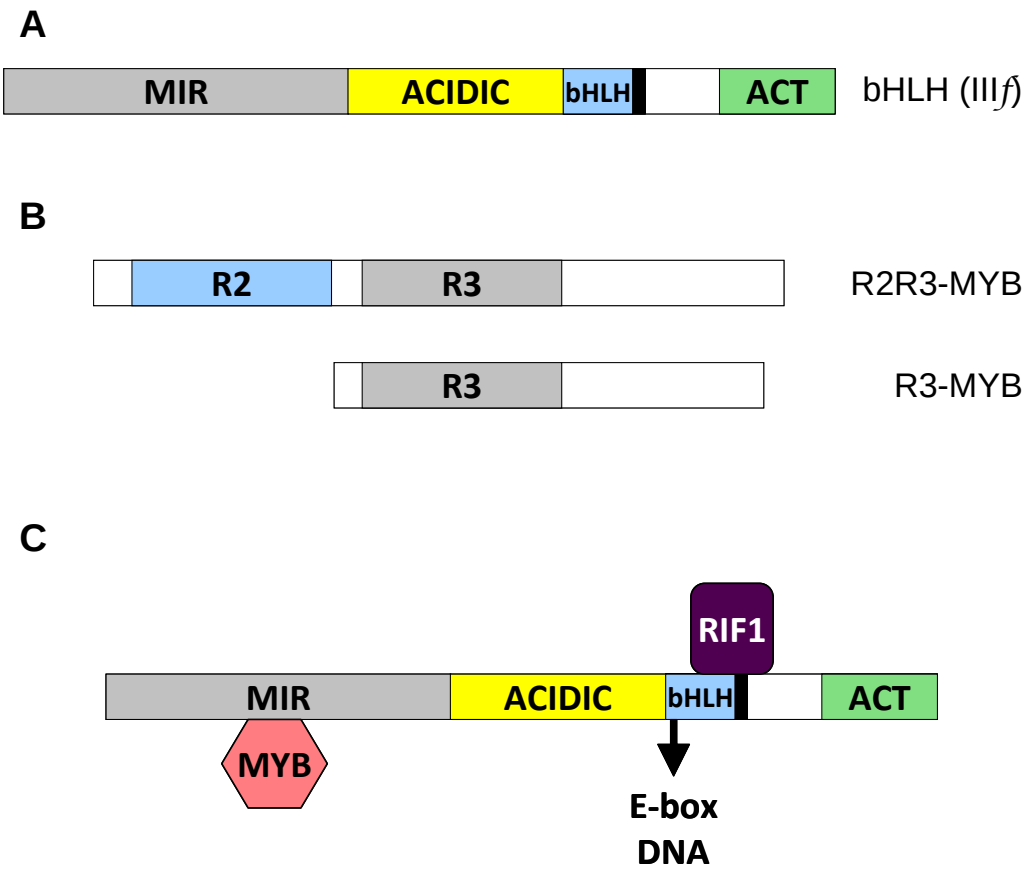


Figure 1.2 Schematic of group III*f* bHLH domain structure and anthocyanin biosynthetic pathway in maize. (A) Representation of the domain structure of Group III*f* bHLHs. (B) Differences between R2R3-MYBs and R3-MYBs. R2 is a DNA-binding domain and R3 is a protein interaction domain. (C) Relationship of interactions between maize R and partners. Please see text for details.

activators of gene expression, these bHLHs interact and bind R2R3- or R3-MYB TFs (naming derived from the myeloblastosis oncogene) (Figure 1.2B). The R3 domain is a protein–protein interaction domain that interacts with the MIR of group III*f* bHLHs, while the R2 domain is a DBD (Figure 1.2C) [20]. R3-MYBs function as transcriptional repressors by interacting with group III*f* bHLH proteins and not providing a DBD for stabilization of the protein complex with DNA. In maize, ZmC1 (COLORLESS 1) is the R2R3-MYB interacting partner for R. ZmP1 (PERICARP 1) is another R2R3-MYB involved with anthocyanin accumulation in maize; however, it can activate gene expression without a bHLH component [20-23].

A WD-repeat (WDR) domain protein is the third member of the complex. It does not bind DNA, but rather functions as a scaffold or adaptor between other proteins within the complex. WDRs are a diverse family of proteins that have roles in a wide range of cellular functions [24-29]. Many WDRs interact with kinases and/or chromatin [30], but how AtTTG1 (TRANSPARENT TESTA 1) [11,31-34] in Arabidopsis functions within the transcriptional complex is largely a mystery. It participates in trichome and root hair initiation, anthocyanin biosynthesis, and seed mucilage biosynthesis. Properties of its homolog ZmPAC1 (PALE ALEURONE COLOR 1) are even less known, though it appears that deletions of this protein only impact anthocyanin biosynthesis and not trichomes or root hairs [35]. Neither AtTTG1 nor ZmPAC1 interact with the MYB TFs, but only with the bHLH TFs.

In *Zea mays*, a fourth component was later identified, ZmRIF1 (R-INTERACTING FACTOR 1) [36,37]. It contains AGENET [38] and EMSY-like [39] domains, which are domains found in chromatin-associated proteins. Its discovery was not unexpected since most genomic DNA is a condensed-form (i.e. euchromatin). Euchromatin is predominately DNA wrapped around histones and condensed no further. This arrangement allows for another layer of regulatory control for gene expression and helps to keep genomic DNA organized within the nucleus. The R–RIF1 interaction, however, has only been found to be valid in maize thus far (Figure 1.2C) (unpublished data, E. Grotewold lab of Ohio State University).

DNA-binding in bHLH TFs

After early experiments failed to demonstrate that the bHLH domain in R could bind DNA, it was accepted that the domain was incapable of binding DNA despite the presence of a normal appearing bHLH domain. The typical bHLH domain binds E-boxes (CANNTG) or the canonical G-box (CACGTG). Comprised of two α-helices joined by a loop, the N-terminal portion of the first helix is comprised of many basic or negatively charged amino acid residues and interacts within the major groove of DNA (Figure 1.3). The rest of the first helix and the second helix are a protein–protein interaction domain, which allows bHLHs to form hetero- or homodimers. The ability of bHLHs to bind DNA is a function of a bHLH's ability to form dimers and without the ability to form dimers, this DNA–protein complex cannot be stabilized [40].

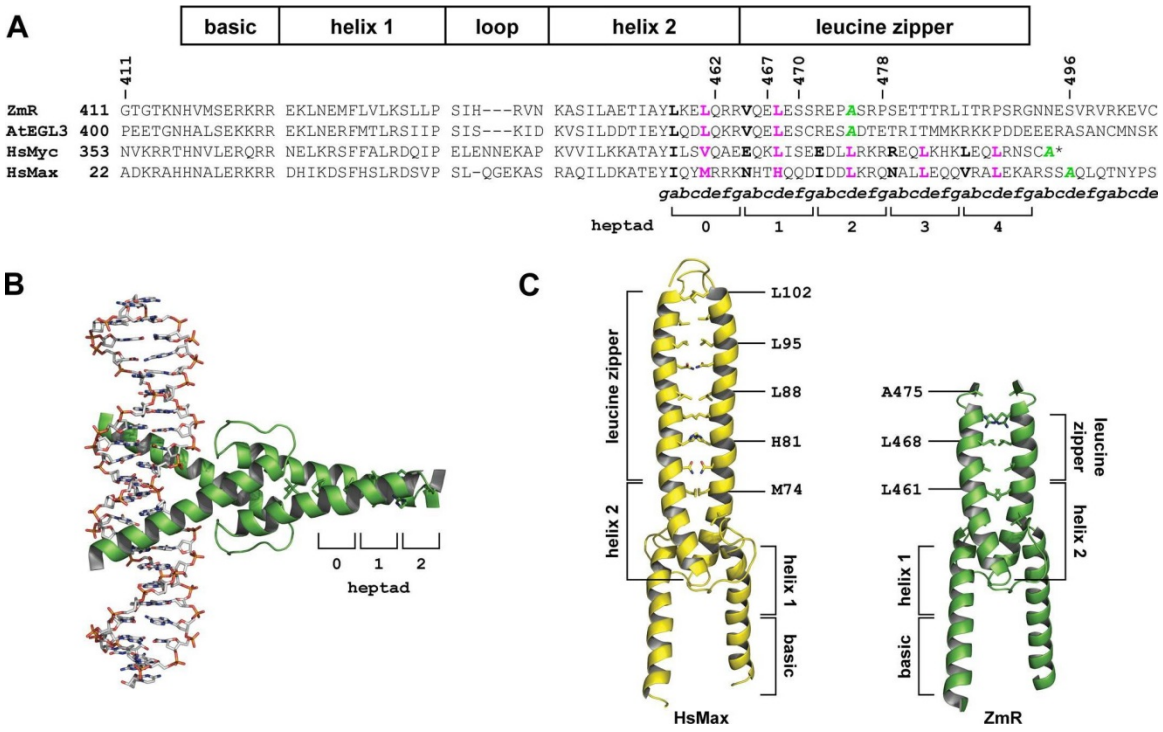


Figure 1.3 Basic-helix-loop-helix structure. (A) Partial alignment of group III*f* bHLHs, ZmR and AtEGL3, with bHLH-LZs found in humans. (B) A representation of bHLH–LZ interacting with DNA with the basic domains seated in the major groove of DNA. (C) Comparison of structural differences between HsMax and ZmR homodimers.

The origin to the belief that R is unable to bind DNA likely stems from an incomplete understanding of secondary structure. The coiled-coil is a feature of α-helices that allows them to interact with each other as in the case of the bHLH protein–protein interaction domain. The concept was proposed by Crick in 1952 [41]. When discovered in 1989, the bHLH domain of R was identified as a region of homology to L-*myc* and MyoD1 bHLHs in humans and spanned from amino acids 420 to 462 [5]. Later when studies were performed using individual domains, a full basic domain was included (411 to 462), but apparently no one seriously questioned whether position 462 was the appropriate place to make a C-terminal truncation (Figure 1.3A,C).

R forms transcriptional complexes on many promoters with two promoters in particular serving as models, the *ZmA1* and *ZmBz1* gene promoters (Figure 1.1). The *ZmA1* gene encodes a dihydroflavonol 4-reductase (DFR), while the *ZmBz1* (*Bronze1*) gene encodes an anthocyanidin 3-O-glucosyltransferase (UFGT).

MADS domain TFs

Within the context of TFs, when one thinks of a plant, its roots, its leaves, its flowers it can be easy to attribute a structure or phenotype to a particular locus. In Arabidopsis, for example, a loss-of-function mutation in the *AGAMOUS* (*AG*) locus gives rise to double flowers with repeating whorls of sepals and petals due to the loss of stamens and carpels, while the loss of *APETALA3* (*AP3*) or *PISTILLATA* (*PI*) results in flowers lacking sepals and petals [42]. Initially, it can be quite easy to think that *AG* is responsible for the formation of stamens and carpels or that *AtAP3* and *AtPI* are responsible for the formation of petals and stamens. However, these structures are merely an emergent property from a cascade of events regulated by *AtAG*, *AtAP3*, and *AtPI*.

With leaves being the default structure, sepals, petals, stamens, and carpels are all a modification of the leaf form. In the most evident observations, conversion of a leaf to a petal would include repression of chlorophyll biosynthetic genes and degradation of chlorophyll. For those plants whose flowers are pigmented, genes involved in the carotenoid and/or flavonoid biosynthetic pathways are upregulated and in the case of fragrant flowers, genes of the appropriate regulatory and biosynthetic pathways must also be upregulated. These changes, however, are almost superficial compared to the changes that must occur in order to alter form as the structure or organ to include changes in the cellular morphology by way of changes in cell wall deposition and organelle differentiation that result in the wildly different shapes and forms of flowers found in nature.

AtAG, *AtAP3*, and *AtPI* are members of the MADS TF family. MADS domain TFs comprise a family of DNA-binding proteins whose major role and function is in the regulation of animal and plant spatiotemporal development. This family of TFs derives its name from the various genes and proteins from which members of this family were originally discovered — MCM1 (minichromosome maintenance 1, *Saccharomyces cerevisiae*), AGAMOUS (*A. thaliana*), DEFICIENS (*Antirrhinum majus*, *AtAP3* homologue), and SRF (serum response factor, *Homo sapiens*). Few plant TF families possess the breadth of regulatory control as do MADS TFs and because of this MADS TFs could be viewed as master TFs.

Significance of plant MADS domains TFs

MADS domains of plant origin are without representation in structures determined by X-ray crystallography or solution NMR despite plants (Arabidopsis) possessing upwards of 30 times more MADS domain proteins than either animals (humans) or fungi (yeast) (see appendix for chart). Though important and tremendously helpful to our present day understanding of the MADS domain, past research involving motif swaps and targeted mutations within and between Type I (SRF-like) and Type II (MEF2-like) proteins only sampled a relatively small sample of the total available sequence-space utilized by MADS domains. Within the domain itself, plant MADS domains, as a whole, possess greater diversity of sequences – sequences that have a biological context and relevance.

Role of MADS domain TFs in mitigating impacts from climate change

Drastic changes to Earth's climate have been ongoing since its formation some 4.54 billion years ago [43]. Since the rise of *Homo sapiens*, our species has endured the hardships these great climatic fluctuations have created and yet, as a result, these changes helped to form who are as a species today. With an increasing world population, our food security, even our geopolitical stability, is at risk to the current changes in climate that disrupt our ability to produce food.

Nearly all of our food is derived in one way or another from plants. Whether it be in the form of fresh fruits and vegetables, grains for breads, or grains for livestock, flowering plants are the primary source of the world's food supply. Therefore, a full understanding of the flowering mechanism and its regulation is one of several steps towards increasing food security.

Over the past 50 some years (1956 to 2005), the global mean surface temperature has increased by 0.64°C ± 0.13°C [44]. However, this rise has not been steady and evenly distributed. Extreme weather patterns have been an expected outcome of global warming and include early springs and late freezes, long summers and late falls, heat waves, periods of extreme precipitation and drought [45,46]. These changes have the potential to considerably impact agricultural and ecological systems. Impacts include the obvious, such as drought, but also the non-obvious.

Flowering time is coordinated through the inputs of the autonomous, photoperiodic, thermosensory, and vernalization pathways. Early springs and late falls increase the number of degree days, which can speed-up or delay spring flowering in many perennials and trees [46-48]. Some of the greatest instabilities in temperatures have been found to occur during the spring and fall [49]. As a consequence of warmer springs, plants are breaking winter dormancy earlier and the new tender vegetation becomes highly susceptible to late freezes [50]. Control of winter dormancy in plants is tightly regulated in order to protect reproductive tissues from killing freezes. These late freezes can decimate fruit yields, which is damaging agronomically and ecologically for creatures dependent on fruits as a primary food source [50-52].

Not only is there a risk of freeze-damage to spring flowering plants, but also extreme high temperatures for summer flowering plants. Many plants cannot tolerate high temperatures or drought-stress during critical events such as pollen formation, pollination, and fruit development [53,54]. Dry weather and, in particular, soil moisture levels have been determined to be a significant factor of high temperature extremes as observed for the high temperatures in Texas in 2011 [45] and the high temperatures for the continental United States in the summer of 2012. Additionally, temperatures can also be differentiated into daytime and nighttime temperatures, of which, high nighttime temperatures appear to have to the greatest impact on plants.

Heat-induced, male-sterility in plants has been well-documented [55-61] and recently, it appears that pollen development is aborted as a result of auxin depletion due to the repression of auxin biosynthetic genes [53,62]. Through chromatin immunoprecipitation sequencing (ChIP–SEQ) using a dominant–repressor form of MADS TF, SEPALLATA3 (AtSEP3), 33 loci for auxin response and homeostasis were significantly overrepresented in heat-stressed Arabidopsis, indicating possible control via MADS domain TFs [63].

In cereals, there appears to be some debate as to the effects of high temperatures on the later steps of grain development. In general, high temperatures are thought to decrease yields. Yields, however, are determined by a number of factors such as grains per spike, grain size, grain composition, and so on. For example, in spring wheat, nighttime temperatures decreased overall yields, grains per spike, and grain size at

temperatures of 20°C (68°F) and above [64,65]. Complicating issues also include sowing times as these influence the eventual environmental conditions (photoperiod, temperature, and rainfall) for grain-filling [66]. MADS TFs of the SQUA-like clade (members include AtAP1, AtCAL, AtFUL) are thought to specify spikelet and/or floral meristem identity [67,68], which could have a direct influence on grains per spike.

The rest of what is known at the molecular biology level of grain-filling primarily comes from rice. A study [69] was conducted to determine whether low grain weights under high temperatures were due to decreased capacity of the developing grain to store assimilate (photosynthate) or the availability of assimilate. The developing grain can be thought of as a sink for assimilate and the grain dry matter increase rate as the rate at which assimilate is captured by the grain with capture of assimilate by the developing grain occurring during the grain-filling period. To increase the availability of assimilate, plots were thinned at the onset of the grain-filling period. Thinned plots were able to compensate in yield for the loss of spikelets due to thinning and reduced (accelerated [70]) grain-filling period induced by high temperatures. Results indicated that as temperatures increased during the grain-filling period, so did the grain dry matter increase rate, but was limited by the availability of assimilate.

Again, high night temperatures appear to be more harmful than high day temperatures. Cross-sectional analysis of rice grains under high temperatures revealed similar cell numbers within the endosperm; however, those exposed high night temperatures had reduced cell sizes compared to those exposed to high day temperatures [71]. Additionally, in both rice and wheat, high temperatures alter the protein to carbohydrate composition by a reduction in starch and type of starch accumulated [66,72,73].

An incredibly interesting piece of information was found in a transcriptome analysis study between heat susceptible and tolerant wheat varieties. Aside from the anticipated transcriptional changes for heat shock protein and heat shock factor genes, one particular probe ID stood out, Ta.3854.1.S1_at (Affymetrix GeneChip Wheat Genome Array). This probe ID represents a wheat MADS TF (*TaSEP2-B*/*TaAGL4*) with homology to Arabidopsis *SEPALLATA2* (*SEP2*) and is upregulated 15-fold in the heat susceptible variety over the heat tolerant variety [58]. Another SEPALLATA homolog, *TaMADS1*, was found to cause early flowering when overexpressed in Arabidopsis [74].

This could indicate that in the heat susceptible variety, the flowering program is advanced too quickly, limiting the ability of the wheat to form properly developed flowers as well as timing and ability of the wheat to form well-developed, full seeds.

Two other MADS TFs, OsMADS6 and OsMADS29, have been identified in rice that appear to have involvement in endosperm development and/or grain-filling. Defects in *OsMADS6* resulted in grains with seed coats and embryos, but had reductions in starch-filling and production of fully matured seed [75]. In the case of OsMADS29, it appears that pollination reinforced auxin signaling upregulates *OsMADS29* expression, which in turn, upregulates genes involved in further auxin signaling, programmed cell death, and starch biosynthesis. OsMADS29 appears to have greater importance than OsMADS6 to this process. If *OsMADS29* expression is reduced or inhibited into endosperm development, a reduction in nutrient transport gene expression is also observed. These events are the likely cause of the abnormal seed phenotype found in OsMADS29 mutants [76].

Origins and classification MADS TFs

The MADS domain proper is thought to have been derived from a motif of an ancestral prokaryote topoisomerase IIA subunit A (TOPOIIA-A) via a gene duplication event in a common ancestor of extant eukaryotes (Figure 1.4). A second gene duplication event in a 'most recent common ancestor', estimated to have occurred nearly 1.5 billion years ago, is understood to have given rise to the Type I (SRF-like) and Type II (MEF2-like) MADS domains [77]. Therefore, both lineages are represented in nearly all eukaryotes and this model also explains the absence of MADS domains in excavate eukaryotes (e.g. chromistans have only Type II) and prokaryotes (bacteria and archaea) [77].

Plants heavy reliance upon the MADS TF family is clearly evident when the number of MADS domain TF per organism is examined. While metazoans, such as *H. sapiens* and *Drosophila melanogaster* possess 5 and 2 MADS TFs, respectively, and fungi, such as *S. cerevisiae*, possess 4 MADS TFs, the Arabidopsis genome possesses an astonishing 105–111 MADS TFs [1,78,79]. Despite this disparity, most of what we know concerning determinants of DNA-binding specificity come from very limited sequence diversity found in *H. sapiens* and *S. cerevisiae*.

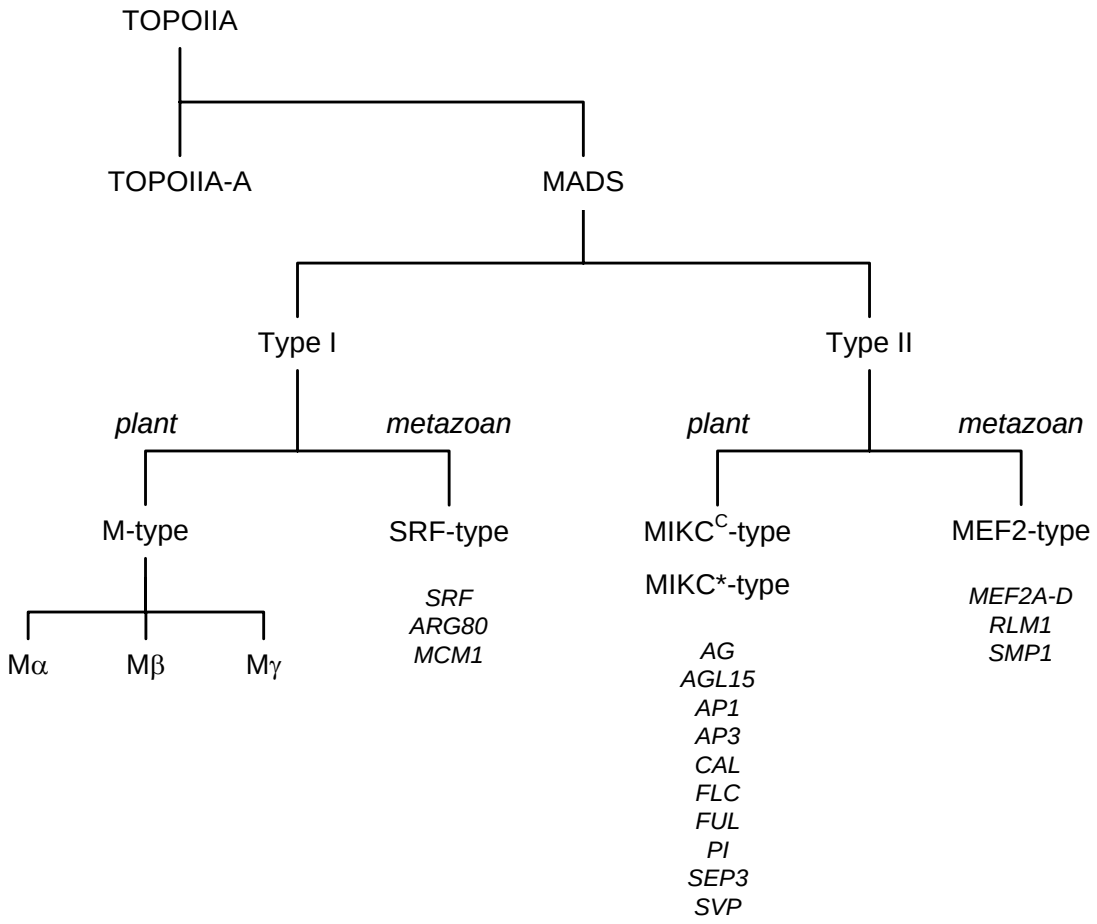


Figure 1.4 MADS domain origin, classifications, and relationships. The MADS domain is thought to have been derived from a gene duplication event involving topoisomerase IIA subunit A (TOPOIIA-A) leading to the formation of the MADS domain. Another gene duplication event in our 'most recent common ancestor' formed the basis of the Type I and II MADS TFs. Eventual diversification formed the many MADS TFs present today [77].

In animal and fungal lineages, the two types are well-known by whether they contain a SRF-like or MEF2-like MADS domain. Here, animal and fungal Type I MADS TFs (SRF-type MADS TFs) possess a SAM-domain (SRF, ARG80, MCM1) [80] following the MADS domain, while animal and fungal Type II MADS TFs (MEF2-type MADS TFs) possess a MEF2-domain (myocyte enhancer factor 2). It should be stressed that these are not the only domains that accompany MADS TFs in animals and fungi. Many MADS domain containing TFs typically range between 500 and 700 amino acids with several other domains present. In the case of human blood flukes, *Schistosoma sp.*, MEF2-type MADS TFs have been found with lengths of over 700 amino acids (CAX69766 and CCD78322).

Major groups and clades

Based upon phylogenetic analysis (and function [81]), the MADS TF family in the plant lineage has been separated into various groups and clades [82-85].

Plant Type I MADS TFs (M-type MADS TFs) have been separated into three clades: Mα, Mβ, and Mγ [78,85,86] (Figure 1.4). Though thought to be derived of the SRF lineage, M-type MADS TFs lack the SAM-domain characteristic of Type I MADS TFs in animals and fungi [87]. Within this group are there about 61 M-type MADS TFs in Arabidopsis [86], of which 20 are unlikely to be expressed [88] due to high rates of supersession/obsolescence for this group [89] leaving about 40 that are functional. As a group, their primary function appears centered upon developmental regulation of the female gametophyte and seed development [88,90-93] and because of this, much less is currently known about the M-type MADS TFs. Two notable M-type MADS TFs are AtAGL62 and AtAGL28. In the developing seed, AtAGL62 expression was found to abruptly decline before cellularization of the endosperm. If AtAGL62 expression was maintained, the endosperm did not cellularize [91]. However, in a role unexpected for this group (and with some disbelief [88]), the vegetatively and seed expressed AtAGL28 was found to cause precocious or early flowering via the autonomous pathway when overexpressed in Arabidopsis [94].

MIKC MADS TFs are of the Type II lineage and still possess a form of the MEF2 domain. This group possesses an Intervening domain adjacent to the MADS domain, followed by a Keratin-like coiled-coil domain, and finally, a variable C-terminal region [95] (Figure 1.5). With about 46 present in Arabidopsis, they fall into two groups, a large "classical" group termed MIKC^C [96] and a much smaller out group termed MIKC* (a.k.a. Mδ in earlier analysis/reports) [78,96-98]. MIKC* MADS TFs have qualities that distinguish them from the larger MIKC^C group, such as a longer I-domain and unique exon structure. Additionally, MIKC* MADS TFs have been found to predominantly serve critical roles in the gametophytic generation of plants [99] and in the case of flowering plants, pollen maturation, and pollen tube growth [100-102].

MIKC^C MADS TFs of angiosperms are comprised of about 15 clades, though not all are widely distributed. Those clades that are not widely distributed are likely from plants that have specialized characteristics and adaptations that those plants are known for.

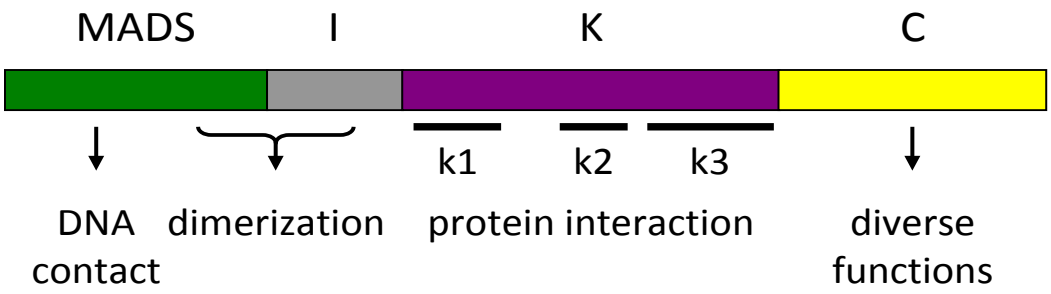


Figure 1.5 MIKC MADS domain structure. MIKC MADS are named by their unique domain organization. The MADS domain in located N-terminally and possesses a DNA-binding domain. Both the MADS and Intervening domains are responsible for protein-protein interactions leading to dimer formation. The K domain has three subdomains and is predominantly known for protein-protein interactions leading to tetramer formation in those that form tetramers. The C-terminal domain has widely varying sequence and can perform many diverse regulatory functions.

Clades that appear to not be widely distributed are AGL15-like, FLC-like, OsMADS32-like [103], and TM8-like [96]. The OsMADS32-like clade might be unique to only monocots [103].

For greater depth of the unique characteristics and roles of MADS TFs for individual divisions, the following reviews are available: Bryophyta (mosses) [97,104,105], Marchantiophyta (liverworts) [99], Lycopodiophyta (clubmosses) [106,107], Pteridophyta (ferns) [108,109], Coniferophyta (conifers) [110,111], and Ginkgophyta (Ginkgo) [112]. Additionally, reviews are available for green algae [113-115], monocots [116,117], and poplar [118].

Properties of the MIKC K domain

Protein–protein interaction studies have been conducted extensively for the plant MADS TFs. The most common methods being yeast two-hybrid interaction studies, followed by electrophoretic mobility shift assay (EMSA), and bimolecular fluorescence complementation (BiFC) [119-124]. Dimerization of MIKC MADS TFs is predominately determined by the K domain and to a lesser extent the MADS and Intervening domains. Within the K domain are three smaller α-helix subdomains: k1, k2, and k3 (Figure 1.5). Subdomains k1 and k2 tend to serve in dimer formation, while k3 tends to serves in tetramer formation for those that function in that manner [125].

Properties of the MIKC C-terminal domain

In MIKC MADS TFs, the C-terminal domain imparts some of the individual characteristics of each TF. Many of the motifs in the C-terminal domain are derived from a combination of gene duplication, subsequent mutation (frameshifts), functional diversification, and neo-functionalization [126,127]. For example, AtAP1 is a transcriptional activator as a result of two distinct transactivation domains present in the C-terminal domain [128].

Early studies showed that some phenotypes of floral Arabidopsis mutants could only be complemented if the C-terminal domain was not truncated [129], but in some cases, merely swapping C-terminal domains was sufficient in restoring the wild-type phenotype [130]. However, in some cases and for certain functions, such as for AtAP3 and AtPI, the C-terminal domain was dispensable [131] for functional specificity in petals and that this role was served through the MIK domains [132].

In tomato, domains responsible for protein–protein interactions between histone deacetylases (HDAC) and MIKC MADS have been discovered in the C-terminal domain through *in vitro* methods with a mammalian HDAC [133]. This is significant as several MADS TFs in plants have been demonstrated to function as both transcriptional activators and repressors. If demonstrated to be biologically functional in plants, this may explain the dual functionality of those MADS TFs [134].

Post-translational modification of MIKC MADS TFs

Post-translational modification has been well established for the animal and yeast MADS TFs. MEF2C's activity is modulated by phosphorylation [135] and MEF2D's activity is modulated by acetylation [134,136,137], while in yeast MCM1, activity is modulated through phosphorylation [138]. This understanding has been the basis for research regarding modifications to MADS TFs in plants.

Post-translational modification has not been well demonstrated for plant MADS TFs. In protein extracts from non-vernalized and *flc–20* null Arabidopsis, FLOWERING LOCUS C (FLC) was found to be predominantly post-translationally modified [139]. Variations in these populations were found to be dependent upon flowering time. Analysis suggested that these modifications were phosphorylated forms of FLC. Six sites and eight serines

were mutated to aspartic acid to mimic the phosphorylated form or to alanine to prevent formation of the phosphorylated form. In the phosphorylated mimic, flowering time was found to occur earlier, while the non-phosphorylated form resulted in plants with delayed flowering. Therefore, the active form of FLC appears to be the non-phosphorylated form. However, many follow-up questions remain.

Non-MADS TF interaction partners

Non-MADS interaction partners in plants have only recently come to light and therefore, knowledge about them is still relatively sparse compared to what is known for systems in animals and yeast. In animals and yeast, these interaction partners not only modulate a MADS' complex status between activation and repression, but also influences the DNA-binding site (DBS) selection as some *cis*-elements overlap each other. A possible explanation for this discrepancy between plants and metazoans could be due the sheer number of MIKC MADS TFs present in plants compared to animals and yeast. These are reviewed in references [120,140-142]. TCF protein (Elk1 or SAP-1) is required to engage SRF on non-optimal CArG sites [143,144] with the ets-motif being adjacent to the CArG box [145].

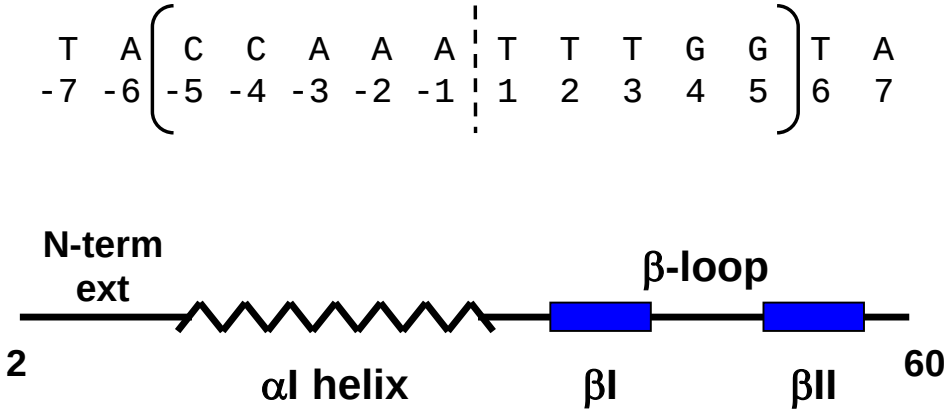


Figure 1.6 CArG box and the secondary structure of the MADS domain. CArG boxes are *cis*-elements targeted by MADS domain TFs (*top*). They are comprised of a 10-bp core, but binding is also significantly influenced by adjacent flanking bases. Due to how MADS domain TFs bind CArG boxes, they can be divided into half-sites and influence binding selection. The secondary structure of the MADS domain is comprised of a variable N-terminal extension, an alpha helix, and two beta-strands that form a beta-sheet with each other and its adjacent dimeric partner (*bottom*). The following Arabidopsis MIKC^{C} MADS TFs possess N-terminal sequence that extends upstream: AG, MAF5, SHP1, and SHP2.

DNA-binding site recognition

DNA, in its classical form, is comprised of just four nucleotides. Many proteins bind DNA and can be divided into those that bind DNA nonspecifically and specifically. Proteins that nonspecifically bind DNA include histones and DNA polymerases, while those that bind specific sequences of DNA include restriction endonucleases and TFs. Therefore, several questions have asked. How is a protein able to recognize specific DNA sequences, discriminate between similar sequences, and eventually bind DNA to form a stable complex? Why is this important and how can this information be applied?

Comprised of two antiparallel, sugar-phosphate backbones and four nitrogenous bases in the form of a double helix, deoxyribonucleic acid or DNA has a net negative charge located along the backbone of the polymer from the phosphates. This evenly distributed negative charge is the predominant force that responsible for nonspecific DNA–protein interactions. It is also utilized by proteins with sequence-specificities during the initial interaction or association between DNA and protein.

The binding of a protein to DNA can be characterized into several different steps: scanning, initial fitting, induced fitting, and stabilization. Protein movement or translocation along and throughout DNA occurs in a combination of four ways: 3D diffusion, sliding, hopping, and intersegmental transfer [146,147]. During translocation, potential binding sites are scanned. Once a potential binding site has been located, an initial test of the site's suitability occurs. If the initial test does not find the site suitable, the protein moves along. However, if the initial test finds the site suitable, then the initial test becomes an initial fit. After a successful initial fit, typically DNA bending is induced by the protein for an induced fit. As the DNA–protein complex forms more and more contacts, stabilization occurs. During stabilization, contacts made during the initial and induced fits may be released, while other more conserved contacts take their place.

A multitude of factors play a role in DNA–protein interactions

The overall structure or fine-structure of any particular fragment or stretch of DNA is dependent upon just two aspects that work in tandem. DNA structure, though not as malleable as protein or RNA, it by no means is a static and 'lifeless' molecule. The fine-structure of DNA is subject to the sequence or order of bases within the polymer and to the environment in which that polymer is located.

DNA structure is sensitive to hydration/dehydration, pH, and cation type and concentration, as well as biogenic cations found within the nucleus of cells. Biogenic cations present within the nucleus include multifunctional polyamines [148] such as putrescine, spermidine, and spermine [149-152]. Polyamines are able to modulate transitions of B-DNA to A- and Z-DNA [153-156]. With potassium (ionic radius of 152 pm for K^+ versus 116 pm for Na^+) as the dominant monovalent inorganic cation within the nucleus, it also becomes the dominant inorganic cation to neutralize the negative charge of DNA and also associate with the major and minor grooves of DNA [157]. In the case of divalent inorganic cations, such as magnesium (86 pm for Mg^{2+} versus 114 pm for Ca^{2+}), these tend to favor the major groove [158-160]. Competition between inorganic cations, biogenic cations, and water ultimately determines the degree of hydration DNA experiences and the conformation DNA takes [149,150,161-165]. Additionally, many other small molecules are known to interact within the major or minor grooves or even intercalate between base-steps, though these influence DNA structure in a manner similar to proteins [166].

The bases of DNA do not possess equal properties and therefore, should be expected to have contextual properties. Significant effort has been put forth to understand the intrinsic, sequence-dependent properties of DNA. Some macromolecular properties that are sequence-dependent include bending (protein independent) and DNA type (e.g. A-DNA [167], B-DNA, Z-DNA, etc.). For example, d(CG)n and d(CA/TG)n sequences favor the base-sugar conformations of an alternating *anti-syn* pattern, compared to all (or mostly) *anti-anti* conformation of B-DNA, resulting in the formation of Z-DNA [168,169]. Sequences with known preference for the B-DNA conformation, readily transition to the B-DNA conformation when started in an A-DNA conformation [170]. However, some B-DNA sequences are known to transition from B-DNA to A-DNA and are stabilized by specific ion associations [165,171]. This demonstrates some of the intricacies encountered when trying to understand the role of sequence in DNA structure.

Objectives

Chapter 2 –

Understanding the influence of a leucine zipper-like (LZL) domain, present in ZmR and other group III*f* bHLH TFs, leading to progress in the 'switch model', which allows ZmR to be repurposed for the activation of promoters of differing *cis*-element structure within the same biosynthetic pathway.

Chapter 3 –

Understanding changes within specific positions of CArG boxes and how that influences the ability of MADS TFs to bind. A DPI-ELISA method was modified and optimized for use with recombinant MADS TFs.

Chapter 4 –

Understanding differences in DNA-binding specificities of MADS TFs through molecular dynamics simulations and DPI-ELISA. AtFLC and AtSVP function as a heterodimer, individually possess different DNA-binding specificities, and have widely divergent residues from other MADS TF clades.

Copyright © Joshua R. Werkman 2013

CHAPTER 2

A shortened, yet critical, leucine zipper-like domain present in group IIIf bHLH transcription factors

Hypothesis

During truncation analysis of R and other group IIIf bHLH TFs, the function of the bHLH domain was thought to have been compromised based upon secondary structure analysis. Secondary structure prediction indicated an extended and continuous α-helix between residues 454 and 490, which is part of the bHLH interaction domain and has resemblance to bHLH–LZ domains. The exclusion of these residues may therefore be responsible for some of the inconsistencies observed in prior data. Restoration of these residues will thus improve the stability of the bHLH interaction domain and ultimately DNA binding, while mutation of restored residues would again disrupt the interactions.

Introduction

The *R* locus was one of the few loci originally found to be responsible for altering anthocyanin pigmentation in maize and came to be the basis of the *R*/*B* gene family. Members of the *R*/*B* gene family encode bHLH TFs and are currently classified as a group IIIf bHLH TF [172]. Members of group IIIf bHLH TFs are generally known to regulate anthocyanin biosynthesis or trichome and root hair initiation and development. In the case of Arabidopsis, significant of phenotypic overlap occurs between the bHLH TFs responsible for anthocyanin biosynthesis and trichome and root hair development.

Also known as *Lc*, leaf color, *R* is a heavily studied and important member of *R/B* gene family and is located about 2 cM from the *R* locus [3,5]. Other homologous *R*/*B* family genes in maize include *B-Peru*, *hopi*, *R-S*, and *Sn*. Unlike bHLH Myc-type TFs commonly seen in mammalian systems, R/B-type factors are much larger, containing several other additional domains such as an MYB-interaction domain [16,173], a transactivation domain [174], a bHLH domain [175], and an ACT domain [19] (Figure 1.2). Anthocyanin biosynthesis in maize is co-regulated by several loci comprised of bHLH (R/B) and R2R3-MYB (C1/Pl) transcription factors (TFs). R/B-type TFs require a MYB factor in order to activate target gene expression; while the MYB factor P1 (PERICARP 1) can bind DNA and activate target gene expression independently. The

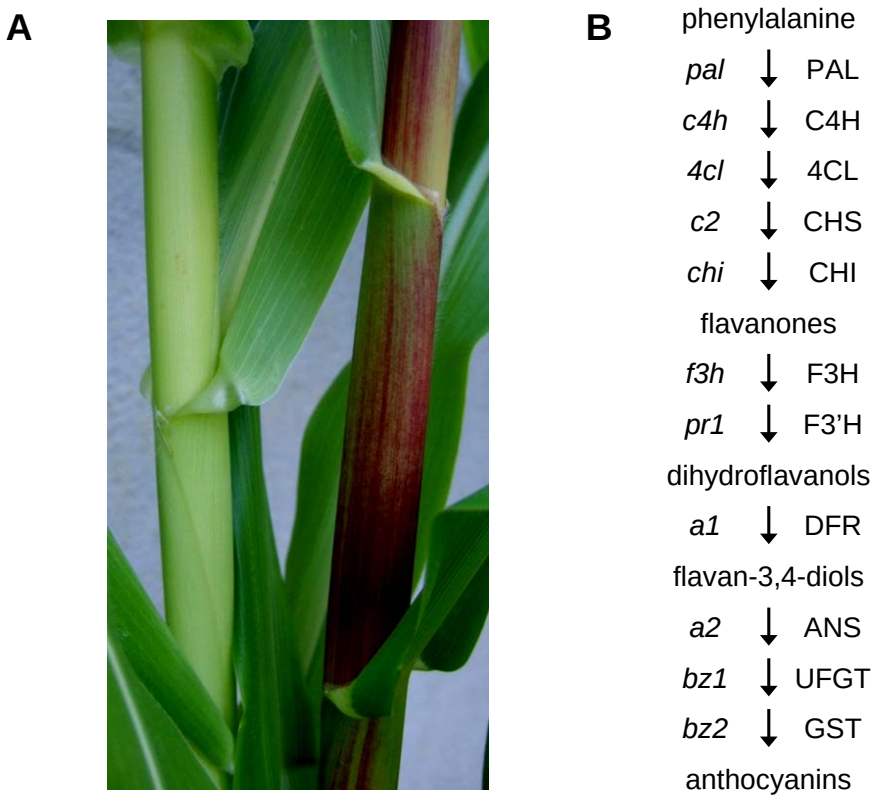


Figure 2.1 Maize stalks displaying differences in anthocyanin accumulation. (A) These maize stalks are likely similar to differences observed in mutants of the anthocyanin pathway. In the case of *R*, a transposon was inserted ahead of *R*, which resulted in activation of the gene and hyperaccumulation of anthocyanins (*right stalk*). (B) Brief pathway to anthocyanins from phenylalanine. Abbreviations: phenylalanine ammonia lyase (*pal*, PAL); cinnamate 4-hydroxylase (*c4h*, C4H); 4-coumarate:CoA ligase (*4cl*, 4CL); chalcone synthase (*c2*, CHS); chalcone isomerase (*chi1*, CHI); flavanone 3-hydroxylase (*f3h*, F3H); flavonoid 3'-hydroxylase (*pr1*, F3'H); dihydroflavonol 4-reductase (*a1*, DFR); anthocyanidin synthase (*a2*, ANS); UDP-glucose:flavonoid O-glucosyltranserase (*bz1*, UFGT); glutathione S-transferase (*bz2*, GST).

MYB factor C1 (COLORLESS 1), however, is dependent upon an R-type factor for gene expression [15,16,22,176-179]. Additional proteins, such as PALE ALEURONE COLOR 1 (ZmPAC1), a WD40 repeat (WDR) protein and implicated in protein–protein interactions, also have significant roles in regulating anthocyanin biosynthesis in maize by stabilizing the protein–DNA complex [35,180]. Together, these form the bHLH–MYB–WDR (BMW) transactivational complex.

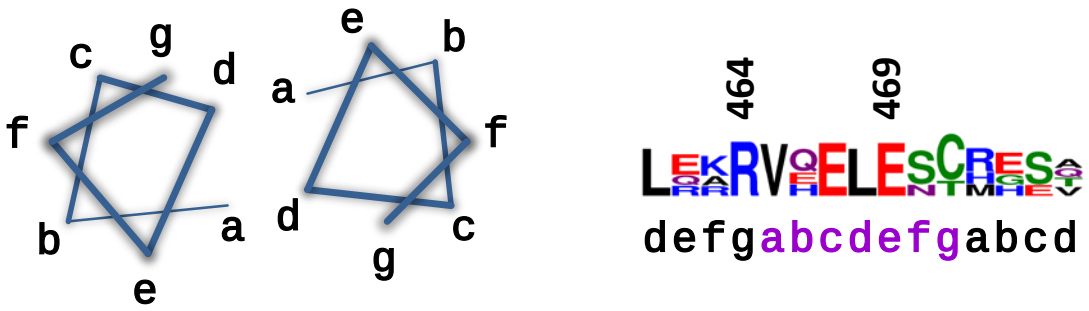


Figure 2.2 Labeling of residue positions in coiled-coils of leucine zippers and the sequence logo of *A. thaliana* group III*f* bHLHs at the LZL domain. LZs can be reduced to repeated heptads that defined by seven residues labeled *a* through *g*. In this view, one is looking down from the C-terminal end of the heptad. The previous heptad's *g* residue will typically participate in stabilization pair. Hydrophobic residues generally occupy positions *a* and *d*, while *g* and *e'* positions are generally occupied by charged residues. These charged residues assist in the selection of pairing of α-helices (*left*). The four group III*f* bHLH TFs of Arabidopsis (MYC1, GL3, EGL3, and TT8) were trimmed to the LZL region and processed through WebLogo3 to emphasize the conserved and complementary residues at positions 464-469 (relative to ZmR). This position corresponds to the equivalent of heptad 1 (*shown in purple*). Additionally, residues at 464 and 469 are highly conserved among all sequences available (*right*).

Group III*f* bHLH TFs have significant roles in regulating of flavonoid biosynthesis [172] and the basic region of the bHLH domain can recognize the core DNA sequence motif, the E-box (CANNTG), which is present in many dihydroflavonol 4-reductase (DFR) promoters [22,176]. More specifically, the canonical G-box (CACGTG), a subsequence of the E-box, is highly preferred. This interaction between the basic domain and the G-box is facilitated by the helix–loop–helix domain that follows. Despite the requirement of a MYB factor in order for R to bind DNA, R can bind DNA when individual domains are isolated [22,37,181]. The bHLH domain, $R^{411–462}$, has not been observed to bind DNA; however, when the bHLH domain is expanded slightly to include a region containing a shortened putative leucine zipper (LZ) that would be homologous to other bHLH–LZ domains, this extended bHLH domain, $R^{411–470}$, will bind DNA [37] (Figure 2.5). This lends to growing recognition of some group III*f* bHLH TFs containing a partially conserved bHLH–LZ domain [172].

Recognition of a shortened LZ immediately after the bHLH domain is not without controversy. In the identification of AtRAP1/AtMYC2/AtbHLH6, a drought-induced bHLH

TF in Arabidopsis, and currently classified as a group III*e* bHLH TF [172], the potential for a putative LZ immediately after the bHLH domain was mentioned [182]. R, Sn, and R-S were all described as containing a putative LZ domain, while B-Peru and AmDEL [183], an *Antirrhinum majus* homologue to R, did not [182]. In the disputed region, R and AtRAP1 have significant points of mention. Both have key amino acid residues for LZs; however, four putative heptad repeats can be identified in AtRAP1, while R has only one. Even more significantly for R, is the presence of three prolines within a span of 16 residues shortly after the mentioned heptad, which almost certainly limits the length of the any putative LZ. A separate classification of bHLH TFs refused to classify even AtRAP1 as containing a putative LZ based upon computer prediction of coiled-coils [184]

Classical LZs are parallel, amphipathic, coiled-coils comprised of α-helices that are four to five heptad repeats in length and each heptad having an inward facing leucine residue at position *d* [185,186]. Position *d* refers to one of seven positions, *g'abcdef*, that form a heptad (Figures 1.3A and 2.2) [187,188]. An ideal dimerization interface between two α-helices is thus created when positions *a* and *d* consist of hydrophobic residues and when positions *g* and *e'* consist of pairs of non-repulsive, charged residues. Branched amino acids, isoleucine and valine, are most favorable and common in position *a* [189] and highly charged residues are common in positions *g* and *e'* [185].

Electrostatic interactions of g↔e' pairs are another significant force in contributing to complex stability [190-200]. R heptad 1, positions g_0 and e_1 (Figures 1.3A and 2.2) revealed that heptad 1 contains an attractive, basic-acidic pair (R↔E), as do the vast majority of TFs similar to R, and it has been shown in LZ-containing proteins [201,202] that attractive g↔e' pairs assist in increasing complex stability. Mutations to repulsive, basic–basic or acidic–acidic residue pairs will likely destabilize complex stability and prevent DNA binding of truncated forms of R bHLH domains. It was therefore anticipated that the bHLH–LZL domain of R should behave in a similar manner.

Materials and Methods

Recombinant Protein Expression and Purification – Domain fragments were isolated by PCR from a *Zea mays* R cDNA template and inserted into pET41aTEV as either an N-terminal GST fusion protein or a C-terminal polyhistidine fusion protein. The plasmid, pET41aTEV, was a gift from Dr. Y.I. Chi of the University of Kentucky and was a

modified Novagen plasmid. It contains a TEV protease cleavage site between the GST coding region and the multiple cloning site. Verified constructs were transformed into *E. coli* BL21(DE3) cells. Cell cultures were grown to an OD_{600} of ~0.6 and then induced by adding IPTG to a final concentration of 1 mM. After induction for 3 h at 30°C, the cells were harvested by centrifugation and stored at –80°C until further use. Bacterial cells were lysed using 1X CelLytic B (Sigma) per the manufacturer's instructions. GST fusion proteins were bound to Glutathione Sepharose 4B columns (Amersham, NJ) and eluted with 50 mM Tris–HCl, pH 8.0, and 10 mM glutathione. The 6xHis fusion proteins were bound to HIS-Select Nickel Affinity Gel columns (Sigma) and eluted with 50 mM sodium phosphate, pH 8.0, 0.3 M NaCl, and 250 mM imidazole. Purified proteins were dialyzed against 1x DNA binding buffer (10 mM Tris–HCl, pH 7.5, 50 mM KCl, and 1 mM DTT).
Electrophoretic Mobility Shift Assay (EMSA) – Probes for EMSA were either the wild-type G-box motif from the Arabidopsis DFR promoter (5'-CGTTCCCCACGTGCTTCTC C-3') or the mutated G-box motif in which the core recognition sequence, CACGTG, was replaced with CAATTG or TGATAC. Complementary oligonucleotides, labeled with biotin at the 5'-end, were synthesized by Integrated DNA Technologies (Coralville, IA) and annealed to produce double-stranded probes. DNA binding reactions were carried out in 10 mM Tris–HCl, pH 7.5, 50 mM KCl, and 1 mM DTT. Purified proteins were incubated with 0.25–0.5 nM DNA probe on ice for 30 min in a final volume of 20 µL. DNA–protein complexes were resolved on 6% non-denaturing polyacrylamide gels and were detected with a chemiluminescent nucleic acid detection kit (Pierce, Rockford, IL).

Yeast One-hybrid and Two-hybrid Assays – In the yeast one-hybrid assays, an effector plasmid pAD–GAL4, containing $bHLH^{411–478}$ or $bHLH^{411–524}$, was transformed into yeast strain YM4271 containing a reporter plasmid, pLacZ–G–box or pLacZ–mG–box. Transformants were selected on SD medium lacking uracil and leucine. β-Galactosidase (β-gal) activity was assayed according to procedures described in the Yeast Protocols Handbook (Clontech, CA). Substrate used for the liquid culture assay was *O*-nitrophenyl-β-D-galactopyranoside (ONPG). One unit of β-galactosidase activity is defined as the hydrolysis of 1 µmol of ONPG to *O*-nitrophenol (yellow) and D-galactose per min per cell.

For yeast two-hybrid experiments, the plasmids, pAD–GAL4–2.1 and pBD–GAL4 Cam, containing truncated forms of *R*, $bHLH^{411–463}$, $bHLH^{411–470}$, $bHLH^{411–478}$, and $bHLH^{411–524}$,

were co-transformed into yeast strain AH109 (Clontech, CA). Transformants were selected on SD medium lacking leucine and tryptophan (SD –Leu –Trp (–LW)). Colonies from double selection plate were then screened for growth on triple selection SD medium lacking tryptophan, leucine, and histidine (SD –His –Leu –Trp (–HLW)) and quadruple selection SD medium lacking tryptophan, leucine, histidine, and adenine (SD –Ade –His –Leu –Trp (–AHLW)).

Predicted Structural Model of R bHLH Domain – A structural model of the bHLH domain of *Zea mays* R (NCBI Accession P13526) was generated by SWISS-MODEL 3.5 using the Alignment Interface [203-205]. Alignments to the *Homo sapiens* Max sequence (obtained from PDB structures 1an2 and 1hlo) were initially made with Clustal X 1.83 [206-209]. Alignments were then tailored in an educated manner to remove excessive

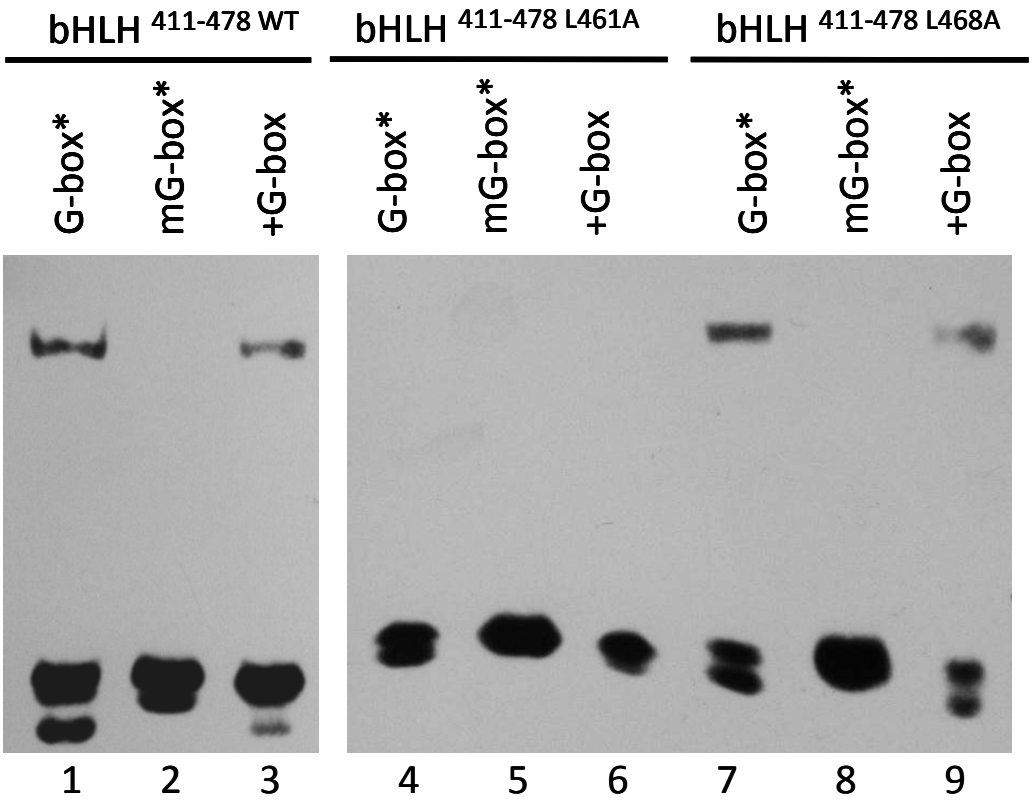


Figure 2.3 Significance of hydrophobic residues within the interface of the LZL demonstrated by EMSA. Mutation of leucines at positions 461 and 468 show that DNA-binding of the G-box by $R^{411–478}$ was disrupted in the L461A form (lane 4). DNA-binding, however, was not disrupted in the L468A form (lane 7). (G = G-box; mG = mutated G-box)

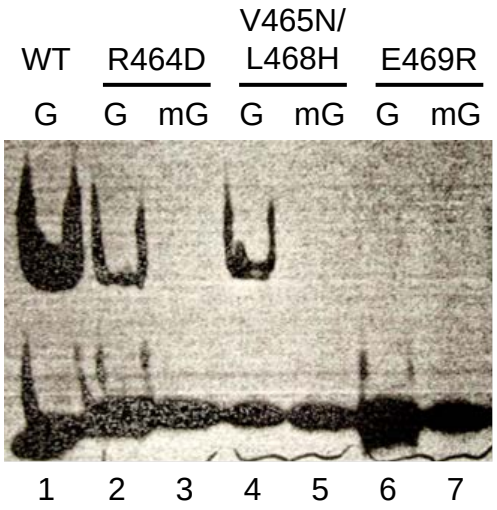


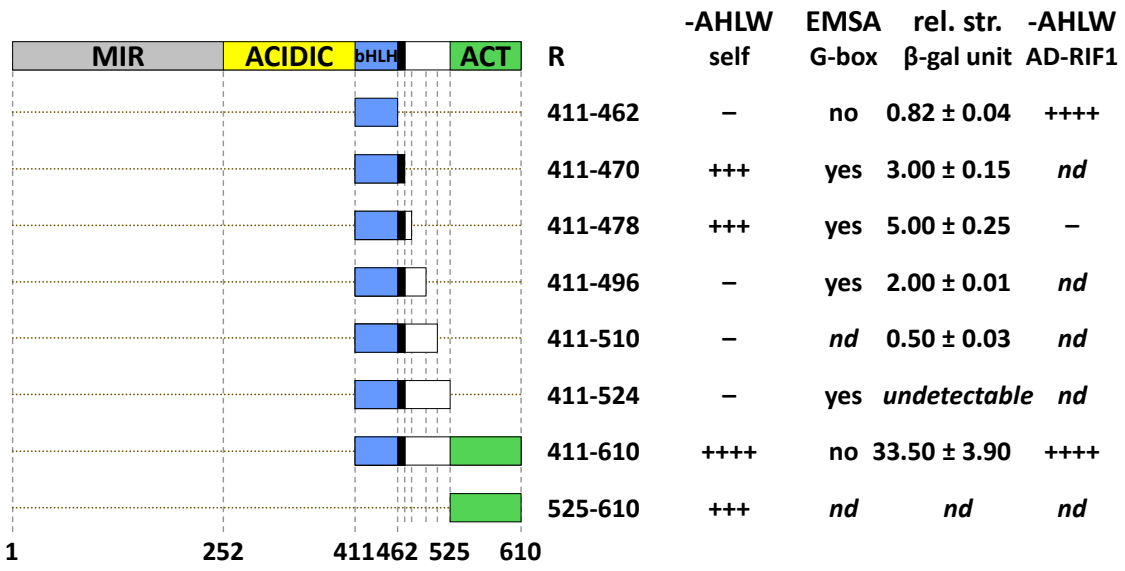
Figure 2.4 Significance of highly conserved residues found in the LZL of group III*f* bHLHs demonstrated by EMSA $R^{411–478}$.
Under the assumption that the putative LZL domain functions in a like manner, mutations were made to positions g_0 and e_1 that were predicted to disrupt the complementary electrostatic interactions thought to be participating in stabilization of the dimeric form. While the R464D mutation did not disrupt dimer formation and DNA-binding of the G-box (lane 2), the E469R mutation did disrupt DNA-binding (lane 6) thought to be due to the disruption of dimer formation. The double mutant, V465N/L468H, was thought to mimic the intermediate dimer formation found in the HsMax bHLH-LZ (lane 4). (G = G-box; mG = mutated G-box)

spacing. Since HsMax structures, 1an2 and 1hlo, covered different, but overlapping regions of the bHLH domain, R was modeled to both. The resulting models were overlaid with the HsMax structures and fused in DeepView 3.7 (SP5) [203].

Results

Impact of mutating the hydrophobic residues in the LZL

Two dominant features responsible for the stability of LZs are the presence hydrophobic residues in the heptad a_1 and d_1 positions and electrostatic interactions in the g_0 and e_1 positions. Knowing that $R^{411–470}$ and $R^{411–478}$ forms can both homodimerize and bind DNA, but not the $R^{411–462}$ form through a LZL motif, allowed for a simple test to determine if the hydrophobic residues contribute to the increased stability seen with the longer forms (Figure 2.5). Two single mutations were made in $R^{411–478}$ – L461A (position d_0) and L468A (position d_1). Both were then assayed for homodimerization in a yeast two-hybrid and for homodimerization and DNA binding in an EMSA. Yeast expressing L461A $R^{411–478}$ were unable to grow on either –His –Leu –Trp (–HLW) or –Ade –His –Leu –Trp (–AHLW), while yeast expressing L468A $R^{411–478}$ were not distinguishable from wild-type (Figure 2.6, sectors 2 and 5). In the EMSA, L461A $R^{411–478}$ did not bind and retard the G-box, while L468A $R^{411–478}$ was able to homodimerize and bind the G-box in agreement to the results of the yeast two-hybrid (Figure 2.3). These mixed results showed that homodimerization and DNA binding is dependent upon the compatible



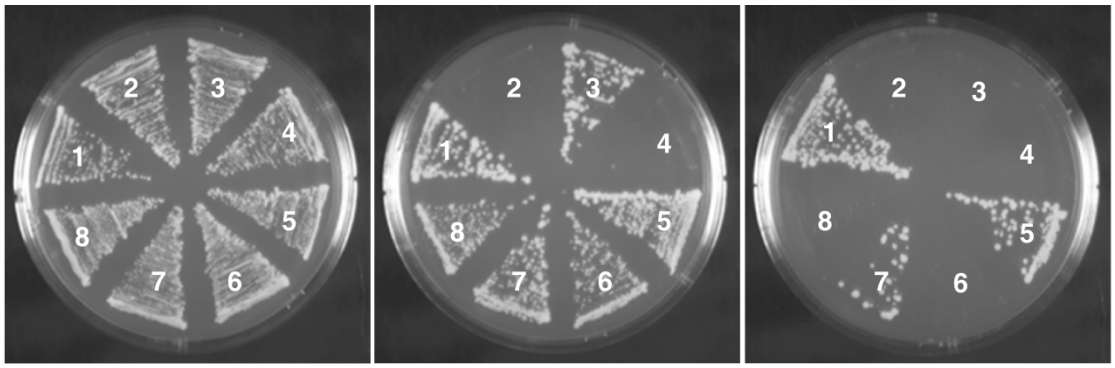
R	-AHLW self	EMSA G-box	rel. str. β-gal unit	-AHLW AD-RIF1
411-462	–	no	0.82 ± 0.04	++++
411-470	+++	yes	3.00 ± 0.15	*nd*
411-478	+++	yes	5.00 ± 0.25	–
411-496	–	yes	2.00 ± 0.01	*nd*
411-510	–	*nd*	0.50 ± 0.03	*nd*
411-524	–	yes	*undetectable*	*nd*
411-610	++++	no	33.50 ± 3.90	++++
525-610	+++	*nd*	*nd*	*nd*

Figure 2.5 Domain truncations relative to full-length TF and tested interactions. Many truncated forms of R have made and tested. These truncations have revealed that the bHLH-only ($R^{411-462}$) does not dimerize or bind DNA, while elongated forms ($R^{411-470}$ and $R^{411-478}$; based upon the current work) both dimerize and bind DNA, specifically the G-box. The ACT domain-only, $R^{525-610}$, dimerizes was the likely reason for increased dimerization in the $R^{411-610}$ form. However, with inclusion of the ACT domain, DNA-binding activity is lost. When interaction partner, RIF1, was tested for protein-protein interaction, interaction was lost in the $R^{411-478}$ form. RIF1 was determined to interact within the bHLH domain $R^{411-462}$. RIF1 appears to recognize the monomeric form of the bHLH and cannot bind if dimerization occurs at the bHLH. (–Ade –His –Leu –Trp (AHLW) drop-out medium in a yeast two-hybrid assay)

interaction to increase the stability of dimerization. Lack of disruption of dimerization and DNA binding by the L469A mutation is a possible indicator that forces, such as electrostatic interactions, were able to compensate for any imbalance due to the L469A mutation.

Dimerization of the R bHLH domain is dependent upon compatible electrostatic interactions in the LZL motif

The group III*f* bHLH TFs were proposed to only bind G-boxes and not the more general E-box. However in *Bronze1*, R recognizes a non-G-box form of the E-box. In Arabidopsis, the group III*f* bHLH TFs are EGL3, GL3, ATMYC1, and TT8, which, except for ATMYC1, are all involved in a similar regulatory mechanism for both anthocyanin biosynthesis and trichome and root hair initiation with some overlap with regard to



1. AD-$R^{411\text{-}478}$ + BD-$R^{411\text{-}478}$
2. AD-$R^{411\text{-}478\ L461A}$ + BD-$R^{411\text{-}478\ L461A}$
3. AD-$R^{411\text{-}478\ R464E}$ + BD-$R^{411\text{-}478\ R464E}$
4. AD-$R^{411\text{-}478\ R464D}$ + BD-$R^{411\text{-}478\ R464D}$
5. AD-$R^{411\text{-}478\ L468A}$ + BD-$R^{411\text{-}478\ L468A}$
6. AD-$R^{411\text{-}478\ E469R}$ + BD-$R^{411\text{-}478\ E469R}$
7. AD-$R^{411\text{-}478\ E469K}$ + BD-$R^{411\text{-}47\ 8E469K}$
8. AD-$R^{411\text{-}478\ V465N/L468H}$ + BD-$R^{411\text{-}478\ V465N/L468H}$

Figure 2.6 Yeast-two hybrid homodimerization analysis of mutations to the LZL region. Homodimerization of the bHLH domain is possible if additional residues are added. To determine the sensitivity of this interaction to perturbations in what is thought to be a LZ-like domain, mutations known to disturb coiled-coil interactions were made. All mutations tested, except for L468A and E469K, disrupted homodimerization under –AHLW selection. Mutations L461A and R464D severely disrupted homodimerization such that no growth occurred under –HLW selection. (–His –Leu –Trp (–HLW) and –Ade –His –Leu –Trp (–AHLW) drop-out media in a yeast two-hybrid assay)

formation of the BMW complex and subsequent downstream gene activation. EGL3, GL3, and TT8 are all able to heterodimerize. This, however, is not due to decreased stability of the homodimer, but rather due to the presence of the extremely conserved heptad 1 (Figure 1.3A and Appendix). Therefore, the differences in regulation functionality are due to differences in other domains and their interacting partners as well as spatiotemporal expression patterns.

Group III*f* bHLH TFs are notoriously difficult to work as full-length proteins and only up until recently [210], recombinant expression of this group had not been successful. Because of these difficulties and the multiple domains present, this group of TFs is regularly worked with as truncated proteins and individual domains. Some of the original work with this group had shown that the bHLH domain did not dimerize and did not bind DNA. This did not sit well with the known properties of this domain – that they bind DNA and the promoters known to be regulated by these TFs, have G-boxes. In bHLH–LZs,

helix 2 and the LZ are typically a single helix with the LZ being an extension of helix 2. Though this group did not the extend helix of bHLH–LZs, it did have a longer helical region that credited. Secondary structure prediction showed that these previous experiments had actually truncated the helix, of which helix 2 is a part of, too short.

Previous experiments have shown that extended helix displayed the expected properties. These properties included dimerization and binding DNA, exclusively the G-box. In bHLH–LZs, the LZ begins with heptad 0, located in helix 2 (Figures 1.3A and Appendix). Heptad 1 begins the defined LZ and heptads generally repeat themselves for a total of five heptads. Alignment with bHLH–LZ proteins revealed the presence of a second heptad in R that aligned with heptad 1 in bHLH–LZs. Unlike bHLH–LZs, R and other group IIIf bHLHs does not have additional heptad repeats. Further evidence that this region in R was likely to behave similarly to bHLH–LZs was that in bHLH–LZs, a conserved alanine was found at the fourth position (Figures 1.3A and Appendix) after the last full heptad and this conserved alanine was present in the group IIIf bHLHs.

A noted characteristic of LZs, in both bHLH–LZs and bZIPs, is that charged residues adjacent to the hydrophobic residues in positions *a* and *d*, are generally involved in electrostatic interactions or salt bridges, which increase or decrease the stability of the protein–protein interaction [185,191-193,197-200,202,211-222]. These electrostatic interactions (cutoff 4 angstroms) must be complementary for strong dimerization to occur. If they do not complement each other, then the two peptides will either only be able to weakly dimerize, of which the overall stability of the complex can be increased when binding DNA, or not dimerize at all. The inability or decreased ability to homodimerize among bHLH–LZ indicates that the particular bHLH–LZ could heterodimerize with one or more partners. Heterodimerization can be a means by which to increase the number of complexes available with a minimal number of peptides.

In order to determine whether heptads 0 and 1 in group IIIf bHLHs, functioned in ways similar to how the multiple heptad repeats function in bHLH–LZs, a series of experiments were conducted to determine if the introduction of antagonistic residues at positions *g* and *e* would disrupt the dimerization and DNA-binding abilities of this truncated domain in R. The residues present in R, at this motif, provide numerous possibilities for salt bridges; however, for these experiments, those that fit the classical positions, *g* and *e*,

for salt bridge formation in LZs were mutated. The goal was to determine how susceptible R dimerization and binding was when these electrostatic interactions were disrupted.

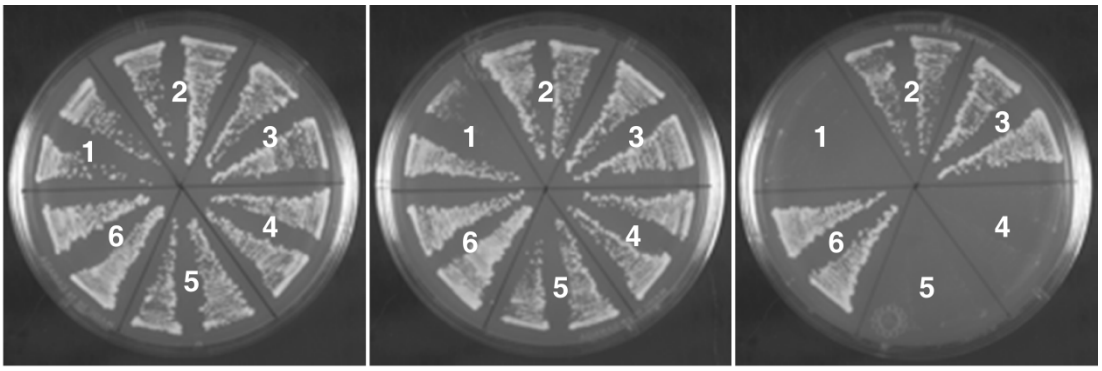
With very few exceptions, nearly all group III*f* bHLHs are comprised of a basic–basic–hydrophobic–neutral–acidic–hydrophobic–acidic motif. In R, it is RRVQELE (Figures 1.3A and Appendix). Single mutations to position g_0 were made, R464D and R464E, in $R^{411–478}$. In a yeast two-hybrid assay for homodimerization, the R464E form permitted yeast growth on –His –Leu –Trp (–HLW), but not on the more stringent –Ade –His –Leu –Trp (–AHLW) selection, while the R464D form prevented yeast growth on both drop out mediums (Figure 2.6, sectors 3 and 4). When the DNA-binding capabilities of R464D form towards the G-box were tested using EMSA (Figure 2.4), binding and sifting occurred, but at a much reduced degree compared to wild-type $R^{411–478}$ (R464E was not tested by EMSA). Thus, both the R464D and R464E mutations disrupted the electrostatic interactions between the helices in this region; reducing the stability and ability to homodimerize to that of the $R^{411–462}$ truncated form.

Additional single mutations were made to position e_1 in $R^{411–478}$ – E469R and E469K. When evaluated in a yeast two-hybrid assay, both forms permitted yeast growth on –His –Leu –Trp (–HLW) selection (Figure 2.6, sectors 6 and 7). On the –Ade –His –Leu –Trp (–AHLW) selection, yeast with the E469R form was unable to grow; however, unlike the mutations prior, moderate yeast growth occurred with the E469K form. With EMSA, the E469R form was unable to cause retardation of the G-box (Figure 2.4).

A special case – mimicking Max N78/H81 in R V465N/L468H

HsMax is a weakly interacting bHLH–LZ and therefore, does not readily homodimerize, but rather heterodimerizes. When N78V/H81L mutations are introduced into HsMax, the monomers are able to readily homodimerize [223]. The reverse of this experiment was repeated in R to determine as to what degree of stability these two positions provide in the homodimerization of R.

In the V465N/L468H $R^{411–478}$ mutant, yeast grew on –His –Leu –Trp (–HLW), but not the more stringent –Ade –His –Leu –Trp (–AHLW) plates in contrast to wild-type $R^{411–478}$ growing well on both (Figure 2.6, sector 8). This indicates that the V465N/L468H $R^{411–}$



1. AD-RIF1+BD-$R^{411\text{-}478}$
2. AD-RIF1+BD-$R^{411\text{-}610}$
3. AD-RIF1+BD-$R^{411\text{-}478\ L461A}$
4. AD-RIF1+BD-$R^{411\text{-}478\ R464D}$
5. AD-RIF1+BD-$R^{411\text{-}478\ V465N/L468H}$
6. AD-RIF1+BD-$R^{411\text{-}478\ E469R}$

Figure 2.7 Interaction of R and RIF1 via yeast-two hybrid analysis utilizing mutations in the LZL region. Disruption of interactions were apparent under quadruple selection. Sector 1 and 2 imitate the biological role since both are only truncated. Both $R^{411–478}$ and $R^{411–610}$ homodimerize without the presence of RIF1 (Figure 6 and [37]), however $R^{411–610}$ possesses two interaction domains of which the second, the ACT domain, has, at times, a stronger bond. With the homodimer stabilized at the ACT domains, the interaction between HLH-LZL domains is free to dissociate. RIF1 enters and interacts with the HLH domain(s) as seen in sector 2. Sectors 3 and 6 are similar in that these mutations inhibit homodimerization and generate free-form HLH monomers (Figure 6). With the introduction of RIF1, RIF1 is observed to interact with monomeric forms of R between residues 411 and 462 [37]. An interaction between $R^{411–478}$ R464D and RIF1 (sector 4) despite the R464D mutation's ability to disrupt homodimerization. (–His –Leu –Trp (–HLW) and –Ade –His –Leu –Trp (–AHLW) drop-out media in a yeast two-hybrid assay)

^{478} mutant destabilizes the homodimerization of R to make it weakly interacting as it does in the longer, full HsMax bHLH–LZ. When tested for binding a G-box in an EMSA using the same amount of protein and probe, V465N/L468H $R^{411–478}$ bound the G-box, but only with about half the intensity as that of wild-type R (Figure 2.4). Thus, R with a single heptad bHLH–LZL motif has a significant role in the maintaining the ability of R to homodimerize and that the regulatory mechanism of R is likely to not tolerant of mutations that reduce its capacity to homodimerize (Appendix).

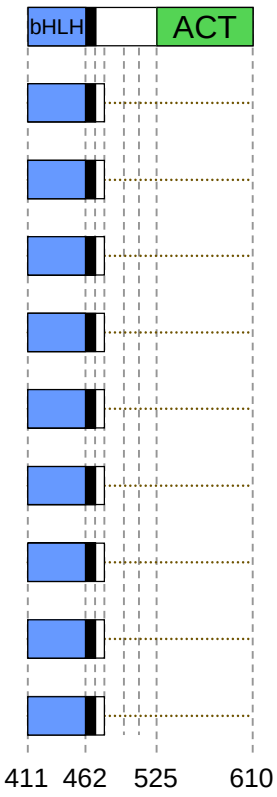
Impact of dimerization on the R^{bHLH}–RIF1 interaction

RIF1 interacts and binds the bHLH domain of monomeric $R^{411–462}$, but not dimeric $R^{411–478}$ or $R^{411–478}$ (Figures 2.5 and 2.7, sectors 1 and 2). However, with the inclusion of the ACT domain, homodimerization no longer occurs at the bHLH domain, but rather

through the ACT domain. Therefore, mutations that disrupt homodimer formation through the bHLH domain could facilitate RIF1 interaction. Mutations that fully disrupted bHLH homodimer formation, in all cases, and DNA binding were L461A and R464D (Figure 2.6, sectors 2 and 4). Those with a significant degree of disruption were R464E, E469R, and V465N/L468H (Figure 2.6, sectors 3, 6, and 8). These five forms were evaluated for interaction with RIF1 by a yeast two-hybrid assay. Since $R^{411–478}$ dimerizes in the yeast two-hybrid, as expected, it did not interact with RIF1 on –Ade –His –Leu –Trp (–AHLW) drop out medium (Figure 2.7, sector 1), while inclusion of the ACT domain in $R^{411–610}$ permitted interaction with RIF1 (Figure 2.7, sector 2). Additionally, R464D and V465N/L468H did not interact with RIF1 (Figure 2.7, sectors 4 and 5). Mutants that did interact with RIF1 were L461A and E469R (Figure 2.7, sectors 3 and 6). The interaction of RIF1 with R464D and not E469R was unexpected since R464D disrupts homodimerization, while E469R does not.

The bHLH–LZ TFs of the Myc/Max/Mad network are known for their latitude in partner swapping based upon these same hydrophobic and electrostatic interactions at the dimerization interface of their LZs [215,224-227]. Containing five heptads, Max has an intermediate stability to homodimerize in part due to deviations from optimal residues at heptad 1, positions *a* and *d*. Mutational analysis of these positions resulted in a marked increase in DNA binding and complex stability in the N78V/H81L Max double mutant, as well as, showing independent folding of helix 2 and the LZ [223]. Interestingly in R heptad 1, valine and leucine appear to occupy to same positions seen in the Max double mutant (Figure 1.3A) and the Max double mutant suggests that V465 and L468 in R could function in a similar manner to the LZ by increasing dimer and overall complex stability.

Other information supporting the presence of a coiled-coil, include consensus secondary structure predictions for AmDEL, R, and PfMyc-RP (*Perilla frutescens*). All three proteins show agreement for the presence of helical structure with R matching as to when the helix ends. R has a predicted helical structure to include heptad 1. AmDEL and PfMyc-RP are not as sharply demarcated, but the consensus predicts an end to helical structure in AmDEL after heptad 1 and midway into heptad 2 for PfMyc-RP. Interestingly for AtRAP1, the region corresponding to heptad 1 in R had some prediction results showing potential β-strand or non-helical structure as did HsSREBP1. Helical



R	-HLW self	-AHLW self	EMSA G-box	-AHLW AD-RIF1
411-478	++++	+++	yes	–
411-478 D12	*nd*	–	*nd*	–
411-478 L461A	–	–	no	++++
411-478 R464D	–	–	yes	–
411-478 R464E	+++	–	*nd*	*nd*
411-478 V465N,L468H	+++	–	yes	–
411-478 L468A	++++	+++	yes	*nd*
411-478 E469K	+++	+	*nd*	*nd*
411-478 E469R	+++	–	no	++++

Figure 2.8 Summary of results for mutations within the LZL domain of $R^{411-478}$. L461A and R464D mutations severely disrupted dimerization, but in the case of R464D, it was apparently not significant enough to disrupt cooperative DNA-binding. L468A and E469K mutations did greatly impact dimerization. L461A and E469R disrupted dimerization and DNA-binding and likely created obligate monomeric forms, which resulted in the ability to interact with RIF1. D12 is a naturally occurring deletion determined not to dimerize or interact with RIF1. (–His –Leu –Trp (–HLW) and –Ade –His –Leu –Trp (–AHLW) drop-out media in a yeast two-hybrid assay)

structure resumed in the heptad 2 equivalent and continued for another two heptads. This may have dissuaded the classification of AtRAP1 as a bHLH–LZ when coiled-coil prediction analysis was performed [184].

As to whether group III*f* TFs could merely possess an extend helix 2, this is unlikely due to the independent folding of helix 2 and the LZ in Max [223] and the highly optimized LZ nature of R heptad 1 consistent to the register of typical bHLH–LZ TFs. Thus, R heptad 1 and similar TFs should be recognized as containing a putative, though shortened, bHLH–LZL domain assisting as a 'leucine button' for a potential dimerization interface.

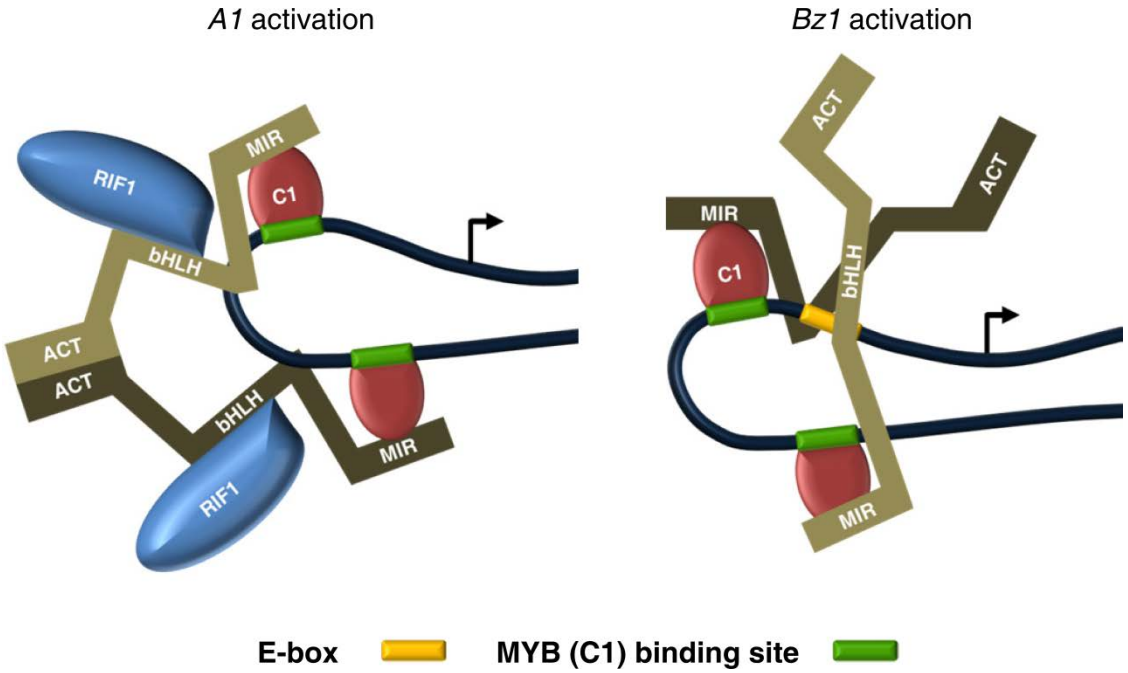


Figure 2.9 Switch model for promoter activation by R. (A) Representative of R interacting with the *A1* promoter (for DFR), which does not possess a G-box. The ACT domain dimerizes and RIF1 binds the helix-loop-helix domain of R. R is unable to directly bind DNA. It interacts with the promoter through the R2R3-MYB, C1, which binds DNA directly. (B) In the *Bz1* promoter (for UFGT) there is a G-box. R is able to directly bind with the promoter and activate it. (Adapted from [37])

Discussion

Model

The current model for promoter transactivation by R involves a bimodal or switch-like mechanism (Figure 2.9). Most importantly, this switch model allows for R to be repurposed for the activation of promoters with differing *cis*-element structure within the same biosynthetic pathway. The promoters of two late pathway genes are transactivated by R, *A1* and *Bz1*. While the promoter of *Bz1* possesses an E-box and no MYB-binding sites, the promoter of *A1* possesses MYB-binding sites, but no E-boxes. Despite this, both R and C1 are required for transactivation of both *A1* and *Bz1*. For *Bz1*, it is thought that R binds and transactivates in a more classical mechanism, in which the HLH–LZL domains homodimerize and the basic domains bind DNA at the E-box. Due to the stable interaction between the HLH–LZL interfaces, RIF1 is blocked from interacting. Additionally, the ACT domains remain separated. Whether this is due

to steric hindrance, posttranslational modification, or possibly even the binding of a small molecule ligand by the ACT domain is unknown.

For *A1*, the mechanism is slightly more complicated. Since no E-box is present, a stable bHLH–LZL–DNA complex is not formed. This exposes the HLH domain and permits RIF1 to interact. In this configuration, homodimerization occurs at the second dimerization interface, the ACT domain. Thus, R becomes a scaffold and is anchored to the *A1* promoter via interaction with C1 and the interaction with RIF1 and the nucleosome [37]. The order in which this complex is formed is not yet known.

Monomer versus dimer pathway for binding DNA

Cooperative binding is a well-documented phenomenon in bHLH–LZ and bZIP TFs [228,229]. DNA specificity and rate can be enhanced when binding occurs through the monomer pathway as opposed to the dimer pathway [230]. It has been reported that $HsMax^{22–113}$ fully dimerizes at concentrations above 10 μM and that when rapidly diluted to 0.6 μM, dissociation to monomers occurs, and redimerization only occurs slowly [219]. Using disulphide bond cross-linked $HsMax^{22–105}$, $HsMax^{22–105}$ was expected to show speedier DNA binding towards the E-box; however, preformed dimer was found to bind more slowly than monomeric $HsMax^{22–113}$ in experiments with and without nuclear extract [219]. The proposed mechanism for this difference is thought to be due to the interaction of nonspecific protein and/or DNA partners, which are entrapped and nonproductive [219,231,232].

Further work has demonstrated the in case of Max, Mad, and Myc dimer formation as protein dimerization was found to be the rate-limiting step. By following the monomer pathway, it is thought that the final complex could be formed faster [40]. Ecevit, O. et. al. go on to note that within the cellular environment, both monomeric and dimeric forms exist and compete for DNA. The pathway of DNA-binding taken would likely be a consequence of the ratio between monomeric and dimeric forms [40].

Could RIF1 be controlling which pathway R binds?

The inclusion of the LZL-domain in R assures a more stable dimerization interface at the HLH domain. This is likely critical for maintaining R homodimers that function as a bHLH and for those promoters that require functional R bHLH homodimer for DNA-binding.

Without the added affinity provided by the LZL-domain, it is likely that promoter transactivation would only occur through the RIF1 pathway and only those genes utilizing it.

Within the cellular environment, the presence of RIF1 would likely alter the availability of monomeric R and force R to bind promoters, like that of Bz1, through the dimeric pathway. With Max, Mad, and Myc, dimerization or the protein–protein interaction was found to be the rate limiting step during complex formation. Determining the binding kinetics between the HLH–LZL and ACT domain interfaces, as well as between the HLH and RIF1, would likely go a long way in understanding the bias of complex formation under the presence of RIF1.

The position of these two genes within the anthocyanin biosynthetic pathway may shed light upon the differences between the two different modes of regulation by R (Figure 2.1). Thought to localized to the surface of the endoplasmic reticulum in the cytoplasm, DFR begins the late anthocyanin biosynthetic pathway by the reduction of dihydroflavonols to flavan-3,4-diols. In turn, anthocyanidin synthase (ANS) oxidizes flavan-3,4-diols to anthocyanidins. Blockage of anthocyanidin transport to the vacuole can significantly decrease anthocyanidin accumulation and without transport into the vacuole, anthocyanidins are subject to oxidative degradation within the cytoplasm. This is what occurs with negative mutations to *Bronze1* (UFGT) and *Bronze2* (a transporter) in maize, which creates a bronze-like color. Glycosylation of anthocyanidins is the dominant modification that permits transport into the vacuole. It could be seen that DFR is a committed branch of the flavonoid biosynthetic pathway, deviating from flavanones (Figure 1.1), and tighter regulation of its transcription would be preferable. Genes downstream of *DFR*, such as *Bz1*, are obligated to be activated in order to clear potentially toxic metabolites from the cytoplasm and would thus likely have a lower threshold for activation of transcription.

Together with the knowledge of the preference for RIF1 to interact with monomeric rather than dimeric R at the bHLH domain and the highly conserved nature of the LZL motif within the group III*f* bHLHs would indicate that the reason for the degree of conservation in the LZL motif is due to a requirement for the dimerization of R at the bHLH domain during some point in time as it functions as a regulator of transcription. In

order to ensure that R can form a dimer at the bHLH domain, the LZL motif must have been selected to provide a sufficiently strong and redundant surface for dimerization. It would also seem to be the case that this mechanism is conserved throughout the group III*f* bHLHs as well.

Copyright © Joshua R. Werkman 2013

CHAPTER 3

An adapted ELISA-based method for determining binding specificity and kinetic data for MADS-domain transcription factors

Hypothesis

It is possible to develop measurements of the binding affinity with regards to specific, known DNA sequences that allows for a more quantitative assessment of DNA-binding site (DBS) recognition amongst MADS domains and complements bulk, nonspecific methods such as Chromatin ImmunoPrecipitation (ChIP)-based techniques or SELEX (Systematic Evolution of Ligands by EXponential enrichment).

Introduction

Identification of *cis*-elements with biological activity has been a major challenge with MADS box TFs and TFs in general. Putative CArG boxes can be readily identified by using a consensus sequence and simple string searching scripts. Advanced methods can be employed that make use of conserved pairs of *cis*-elements within promoters with similar expression profiles. This can be particularly problematic with CArG boxes that have upwards of 16 nucleotides when flanking sequence is included. By understanding the individual kinetics between unique CArG box sequences, one might be better able to understand the binding characteristics, requirements, and specificity of these interactions as compared to bulk methods such as those based upon ChIP or SELEX.

Three common approaches have been used when selecting *cis*-elements to study the binding specificity of TFs – native verified *cis*-elements from promoters, ChIP–seq, and SELEX. Each approach has its own pros and cons. Identifying native *cis*-elements from known promoters has been the classical method to identifying and understanding the biological significance of a *cis*-element and its corresponding TF(s). Typically in promoter analysis, the sequence of the promoter of interest is scanned for *cis*-elements using known consensus sequences from a database of *cis*-elements. Once the locations of putative *cis*-elements have been identified, truncations are made to promoter to combinatorially test select putative *cis*-elements by its ability, or lack of, to drive the expression of a reporter gene. Putative *cis*-elements can also be more

specifically tested by site-directed mutagenesis away from the consensus sequence of a particular *cis*-element in hopes of identifying biologically significant *cis*-elements from decoy binding sites [233] making promoter analysis a gold standard for the identification of relevant *cis*-elements.

Chromatin immunoprecipitation-based techniques include ChIP–on–Chip (ChIP–chip), ChIP–PCR/qPCR, and ChIP–sequencing (ChIP–seq). Each has specific advantages and disadvantages in the identification of *cis*-elements throughout an entire genome. ChIP–seq is an approach well-suited, though not required, to organisms with sequenced genomes and labs with high-throughput resources. Tissue is treated with a cross-linking agent to preserve TFs in their current interaction state with DNA. A lysate is made, the chromatin is sheared by sonication, and the resulting lysate is passed over an immunoaffinity column to enrich for chromatin fragments specifically bound by the TF. In cases where a strong, specific antibody is not available, the TF of interest can be tagged with a peptide sequence suitable for affinity purification, transformed into the organism, and subsequently, purify chromatin fragments over an appropriate affinity column. High-capacity sequencing is performed on the enriched, purified DNA. Multiple sequences are collected for each binding site genome-wide to allow for the mapping of binding sites and more representative consensus sequences as well as a host of additional bioinformatic analyses.

SELEX is a technique in which DNA probes are synthesized with a stretch of random sequence within the middle of the probe. Recombinant, tagged TF and random DNA probe is mixed and allowed to interact under suitable conditions. The TF is purified and DNA sequences that have the high affinities with the TF are co-purified. Co-purified DNA is then amplified by PCR and purified. Several rounds of binding and purification are repeated with those sequences possessing the highest binding affinities with the TF. Finally, probe from the later rounds are sequenced. Advantages to this technique include a wide range of sequences that the TF can interact within a more controlled environment.

Despite these powerful methods to identify which *cis*-elements that a particular TF binds, none offer a systematic analysis for the sequence specificity of a TF towards a *cis*-element. A systematic analysis for binding specificity involves beginning with a known

cis-element for which the TF binds with a high affinity and rotating through every base at every position in that *cis*-element. Frequently, canonical *cis*-elements are palindromic, especially those that are bound by dimeric TFs. For lengthy *cis*-elements, the sheer number of permutations for DNA probes could make a full analysis potentially cost prohibitive. E-boxes, for example, are six base pairs in length. To assay a six base pair element, permitting repetition without importance to order, would require the synthesis of nearly 84 probes. As mentioned before, CArG boxes are ten base pairs in length resulting in 286 permutations and with flanking sequence (14 bp), 680 permutations. Some sequences, of course, could be omitted due to duplications created between sense and antisense orientations. However, if only palindromic elements are initially considered, the number of permutations can be reduced considerably (Appendix). In the case of MADS domains and CArG boxes, each half of a palindromic element corresponds to a half-site (Figure 1.6). Palindromic cases are particularly beneficial in assuring an accurate description when measuring binding kinetics due to the behavior of half-sites and dimeric TFs.

$$number\ of\ sites = \frac{(n + r - 1)!}{(r)!\,(n - 1)!};$$

$$where\ n\ is\ the\ number\ of\ bases\ and\ r\ is\ the\ length\ of\ the\ element$$

The electrophoretic mobility shift assay (EMSA) has been the traditional means of examining the binding properties of MADS domain TFs. It is an adaption to the Western blot where proteins or DNA are separated by size in a polyacrylamide gel under the influence of an electric current (polyacrylamide gel electrophoresis; PAGE) followed the lateral transfer to a membrane and finally detected with a reporting system. In the EMSA, purified or unpurified TF is incubated with DNA probe in a binding reaction. The binding reaction is then separated by PAGE, transferred to a membrane, and UV cross-linked. It is also important to note that running conditions for EMSAs must be nondenaturing and that buffer conditions should not radically alter the structure and binding behavior from native structure and binding. Contemporary, non-radioactive EMSAs generally employ DNA probes end-labeled with biotin. The membrane is incubated with a streptavidin-HRP (horseradish peroxidase) conjugate and finally a chemiluminescent substrate to detect the location of biotin-labeled DNA. An interaction

between TF and DNA is indicated by a retardation or shift in size due to the formation of a larger protein–DNA complex.

Though a powerful and highly useful method, EMSA can be tedious, time-consuming, and susceptible to variation between multiple gels [234,235]. Using EMSA to examine binding kinetics greatly increases the number of gels required. Therefore, an alternative to EMSA was sought that would allow increased the throughput, ease of use, increased consistency, and improved signal measurement.

One widely adaptable method is the enzyme-linked immunosorbent assay (ELISA) (reviewed by R. M. Lequin [236]), which helped transform medicine since its invention in the 1960s by P. Perlmann, E. Engvall, A. Schuurs, and B. van Weemen. A relatively cheap and very sensitive nonradioactive form of the immunoassay, the general procedure of the ELISA involves the immobilization of antigen to an inert surface such as polystyrene. In this case, proteins from blood serum, including the antigen, nonspecially interact and associate with the irregular, hydrophobic surface on untreated polystyrene, a process known as adsorbance. The surface is washed with an appropriate buffer to remove unbound protein. To reduce signal background, another protein, such as casein from milk, is applied to the surface in excess as a blocking reagent and allowed to adsorb to any exposed surface as well as to any proteins that might nonspecifically interact with reagents used later in the assay. After removal of the blocking reagent solution, an antibody with specificity towards the antigen and that has also been chemically linked to an enzyme such as HRP is applied to the surface, which then binds to the immobilized antigen. Following removal of the antibody–enzyme conjugate and washing of the surface with buffer, a chromogenic substrate for the enzyme is applied, in which the enzyme acts upon producing an amplified signal that can measured with a spectrophotometer. Higher signals typically indicate higher levels of antigen. Common variations of the ELISA include: direct, indirect, sandwich, and competitive ELISA. It is well-suited for high-throughput, replicates, and statistical analysis and has been deconstructed and adapted for the study of nearly every type of interaction possible.

ELISA-based DNA–protein interaction assays have been described as early as 1994 and were used for the detection, quantification, and measurement of DNA–protein

binding kinetics [237,238]. However, it was not until 2001 that this adaption became mainstream for the study of the TF HsNFκB (nuclear factor kappaB), which is involved in the immune response and activated by inflammation [239] and represented a new target for the development of anti-inflammatory drugs. To detect the presence of NFκB in

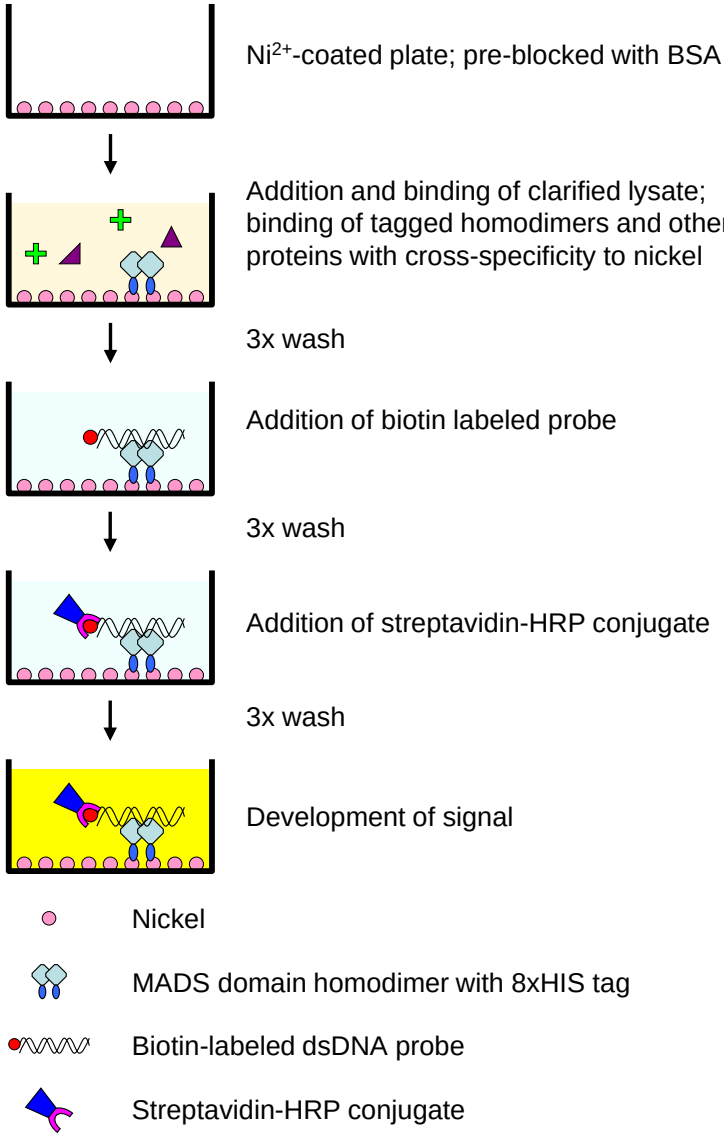


Figure 3.1 Flowchart for redesigned DPI-ELISA. Recombinant HIS-tagged protein is purified from processed lysate using a nickel coated plate. The plates are then washed and biotin labeled DNA probe is added to the wells of the plate and incubated. Following washing streptavidin-HRP conjugate is added to the wells and incubated. Finally after further washing, TMB substrate is added to the wells and the signal is allowed to develop.

whole cell lysates, biotinylated double-stranded DNA (dsDNA) oligonucleotide with the NFκB consensus binding site is immobilized to a streptavidin-coated microtiter plate. The wells are washed and cell lysates are added to the wells. Any active NFκB in the cell lysates interacts with its consensus binding site and binds. Another set of washes are performed to remove unbound proteins and rabbit anti-NFκB antibodies are then added to the wells and incubated. After washing, peroxidase-conjugated anti-rabbit antibodies are applied and incubated. Finally, NFκB is detected with the application of a chromogenic peroxidase substrate and the subsequent colorimetric reaction.

In plants, DNA–protein interaction ELISA (DPI-ELISA) or TF enzyme-linked immunoassay (TF-EIA) has been slow to be adopted in favor of the traditional EMSA. It has been used in the study of bZIPs and WRKYs TFs [240,241] using roughly the same method as described by Hibma and others [237]. Redesigning the method was sought in order to allow combined protein folding and purification (Figure 3.1). Briefly, C-terminally, HIS-tagged protein is expressed as inclusion bodies in *E. coli* BL21(DE3) cells, isolated inclusion bodies are solubilized with urea, solubilized inclusion bodies are diluted and applied to nickel-chelated coated microtiter plates. After immobilization of tagged protein to the plates, wells are washed to remove extraneous proteins. Biotin-labeled dsDNA probes, at varying nanomolar concentrations, are applied to the plate, followed by HRP conjugate. The plate is then colorimetrically developed with 3,3',5,5'-tetramethylbenzidine (TMB) substrate and evaluated in a spectrophotometer.

Materials and Methods

Preparation of Expression Constructs – Expression constructs were prepared using standard molecular biology techniques. Gene fragments were amplified by PCR based upon desired domain combinations from *Arabidopsis thaliana* cDNA clones purchased from the Arabidopsis Stock Center (AGL15 (PYAT5G13790), FLC (U89754), SEP3 (U90075), and SVP (U89996)). DNA fragments were inserted into the NdeI and XhoI sites of pET41 for the generation of C-terminally polyhistidine tagged fusion protein without a GST tag. The final arrangement chosen was in the form of NcoI–MADS$^{1–60}$–HindIII–SEP3^{IK}–SpeI–XhoI–8xHIS–Stop or NcoI–MIK– SpeI–XhoI–8xHIS–Stop. Site-directed mutagenesis was used to remove a native HindIII site in the K domain of SEP3. Alternative MADS domains can be swapped in as NcoI–HindIII fragments and the K

domain provides bulk and an extended surface for homodimerization. Additionally, alternative C-terminal tags can be added such as a possible internal control. Verified constructs were transformed into *E. coli* BL21(DE3) cells.

Preparation of Recombinant Protein – Protein expression was performed overnight with 100 mL ZYM-5052 auto-induction medium with 100 µg/mL kanamycin in 500 mL Erlenmeyer flasks at 37°C, 300 rpm [242]. Cells were pelleted and resuspended in 1X CelLytic B (Sigma B7435), 10 µL/mL protease inhibitor cocktail (Sigma P8849), 1.5 U/mL benzonase (Sigma E1014), and 10 µL/mL lysozyme (10 mg/mL) (Sigma L3790). Lysis was performed osmotically with the addition of five volumes water. Inclusion bodies were pelleted by centrifugation, washed with water, resuspended in storage buffer (50 mM Tris HCl, pH 8.0, 100 mM KCl, 50% (v/v) glycerol, 1 mM DTT (fresh)), and stored at –20°C. Total protein concentration of the recombinant protein stock in storage buffer was measured from serial dilutions of solubilized inclusion bodies in 6 M urea using a microtiter plate formatted Bradford assay (Bio-Rad 500-0006) against a BSA (Bio-Rad 500-0007) standard curve also in 6 M urea.

Preparation of Biotinylated dsDNA Probes – DNA oligomers were purchased from Integrated DNA Technologies (Coralville, IA). The 5'-biotinylated sense strand was in the form of 5'-CCGGGTTTTACTATATATAGTAAAAGGGCC-3' while the antisense strand was without modification. Sequence flanking the CArG box (underlined) was based upon the HABS isolated for AGL15 [243] and a stretch of GC-rich sequence was used for a clamp for either end. Oligomers were reconstituted to 100 mM with Tris–EDTA buffer pH 8.0. Equimolar amounts of each oligomer were combined in a buffer of 75 mM Tris–HCl, 150 mM NaCl, 15 mM $MgCl_2$, 1.5 mM dithiothreitol, pH 7.9 (1.5X NEBuffer #3) for a final individual oligomer concentration of 10 mM. Diluted oligomers were annealed by heating to 95°C in a heat block for 5 min, followed by moving the heat block to a 37°C incubator to anneal. The heat block was moved to a bench at room temperature to finish cooling before placing on ice or storing in the freezer at –20°C.

Generalized Assay – A calculated amount of purified inclusion bodies for the DPI-ELISA at hand are removed from the protein stock solubilized on-demand in 6 M urea made the same day. With exception for 8 M urea, all solutions containing urea were kept in an ice bath and at no point were any solutions with urea permitted to rise above 25°C. After 30

minutes at 6 M urea, the concentration was incrementally reduced by dilution to a total protein concentration of 200 ng/mL and 0.5 M urea.

HIS-tagging and nickel plates were selected for the following reasons: “in position” purification, multiple use of the small protein tag, adsorption could block the DNA-binding domain (DBD) of the dimer and by using this method the DBD is oriented away from the surface. Nickel-chelate coated plates (96-well) (Pierce 15142 or 15442 (0.24 µg/well)) are washed once with wash buffer (10 mM HEPES, 50 mM NaCl, 50 µM EDTA, 0.005% Tween 20, 5% (v/v) glycerol, pH 7.2, 0.25% BSA) and 200 µL diluted, refolded protein is added to each well except for specific control wells. Plates are covered and incubated for 2 hours at room temperature with shaking on a Thermo Scientific microtiter plate shaker (Model 4625-Q) at setting 5. Protein solution is then removed by pipette and wells are washed 3 times with wash buffer. Buffer conditions for binding and washing were found to be radically different than for IMAC purification of HIS-tagged proteins and were modeled after buffer conditions for surface plasmon resonance (SPR) NTA sensor chips [244]. An added benefit is that only minimal optimization will likely be needed if binding kinetics were to be analyzed using SPR.

Previously prepared biotin-labeled dsDNA probes, 30 bp in length, are diluted in deep-well microtiter plates (1 mL) at varying nanomolar concentrations for measurement of binding kinetics or a single concentration (90 nM) for screening probes. Diluted probes are then applied to the wells accounting for appropriate controls. Plates are covered and incubated for 1 hour at room temperature with shaking on a microtiter plate shaker. Probe solution is then removed by pipette and wells are washed 3 times with wash buffer.

NeutrAvidin-horseradish peroxidase (NA-HRP) conjugate (Pierce 31001) is diluted to 2.5 µg/mL and sterile filtered to remove precipitated clumps, which is critical for assay reproducibility. Plates are covered and incubated for 30 minutes at room temperature with shaking on a microtiter plate shaker. NA-HRP conjugate is removed by pipette and wells are washed 4 times with wash buffer. Plates are developed with room temperature 1-Step Ultra TMB substrate (Pierce 34028) for 5 minutes at room temperature with shaking. Peroxidase reaction in stopped with 2 M sulfuric acid. The developed plate is read at 450 nm in a DTX 880 plate reader.

Results

Recombinant protein production

Early in the production of MADS TFs, it was found that expression of MI domains AGL15 (10.1 kDa) resulted in an abrupt halt in cell growth in *E. coli* BL21(DE3) cells upon addition of IPTG and no identifiable protein. Standard methods to resolve this issue were tried including: pLysS, pLysSRARE2, temperature, rare-codon plasmids (Figure 3.2). Expression of only the MI domain was chosen in order to focus solely on the DNA-binding specificities of the MADS domain and to minimize effects from other domains. Later, personal communication with Dr. Sharyn Perry, University of Kentucky, indicated that recombinant expression of AGL15 with the presence of the MADS domains was notoriously difficult and that these forms would be located in inclusion bodies, while truncated forms without the MADS domain were soluble. Furthermore, $SEP1^{1–70}$ (AGL2) was reportedly to be weakly expressed and difficult to work with when compared to $SEP1^{1–77}$ [245]. Realizing additional domains would be necessary to successfully produce recombinant MADS proteins; a truncated form, $AGL15^{MIK}$, (196 aa, 22.6 kDa) and the full-length $AGL15^{MIKC}$ (30.5 kDa) were also expressed in *E. coli*. Both longer forms formed inclusion bodies and did not severely stunt the growth of *E. coli* (Figure 2.3). The truncated MIK form was selected for production as expression was slightly better than for the full-length and additional, unknown bands were consistently present with SDS-PAGE with the full-length protein.

Development and optimization of DPI-ELISA

Initially, attempts were made using nickel-chelated plates and a phosphate buffer system, which resulted in real differences between probes though at very low signal intensities. Improvements and optimizations would be required to make this a viable method. Aspects examined included changing the chelated metal ions, testing various blocking agents, modification of the buffer system, and titration of the components.

Copper versus nickel plates

As a variation of the direct ELISA, which is known to be less sensitive than indirect or sandwich ELISAs, an increase in the amount of HIS-tagged protein bound by the chelate

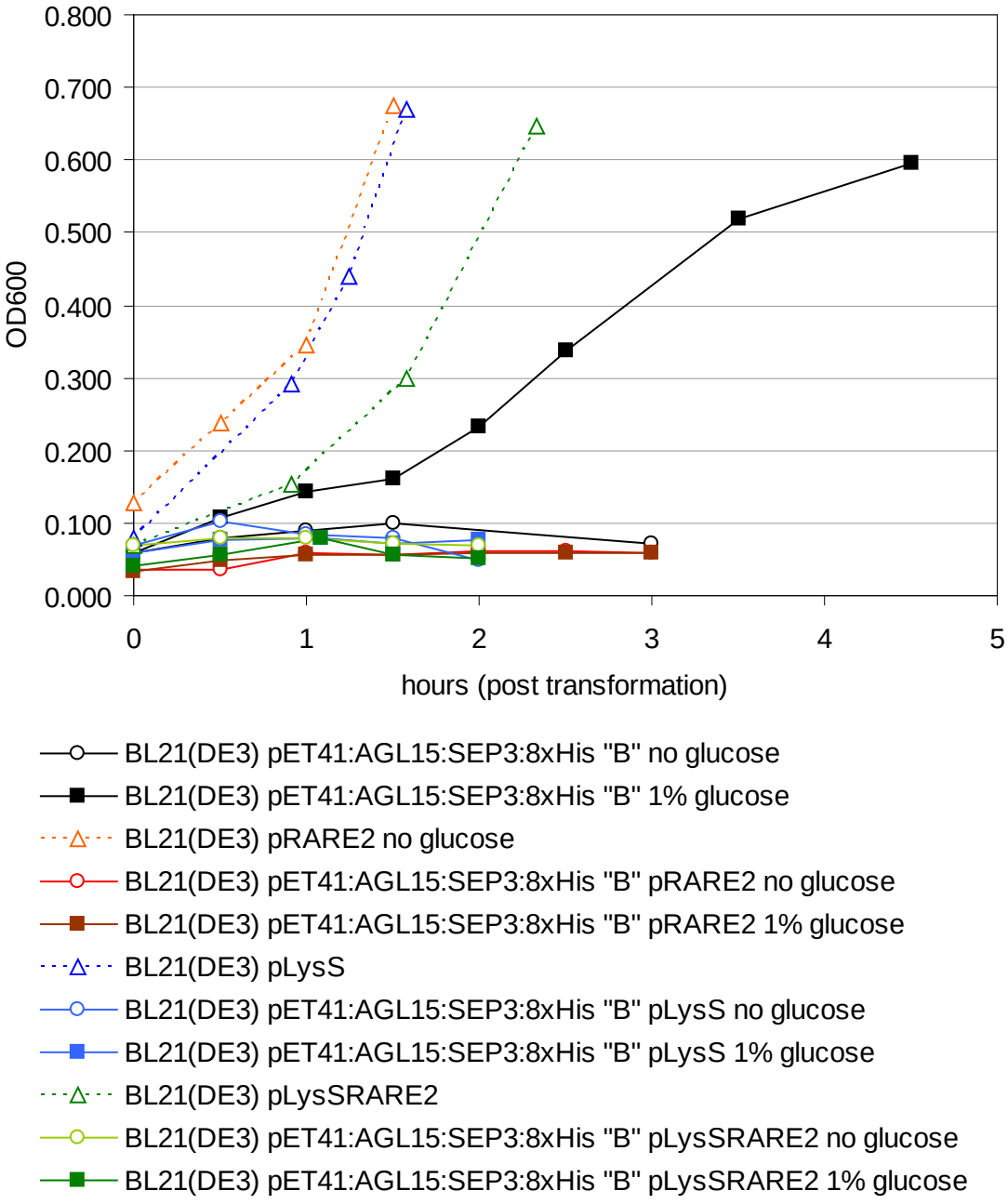


Figure 3.2 Growth characteristics of BL21(DE3) cells producing recombinant AGL15 protein. Only one set of conditions permitted cell growth while containing the vector. These conditions used BL21(DE3) transformed with pET41 and 1% glucose. When additional helper plasmids were included cell growth was inhibited. To detect cytotoxicity, plasmids were transformed into BL21(DE3) cells and their growth monitored following transformation.

metal ion could hypothetically increase the signal and sensitivity of the assay. Plates chosen were Pierce brand from Thermo Fisher Scientific. The means by which metal atoms are chelated to the polystyrene plates is proprietary, but were assumed to be nitrilotriacetic acid (NTA)-based. NTA-based anchors can accept a variety of metal atoms, such as cobalt, copper, or nickel. Each metal has unique properties and affinities for polyhistidine tags and proteins in general.

Of the three mentioned, cobalt and copper have greater capacity for protein binding, but with lower specificity towards polyhistidine tags. Nickel, on the other hand, has the greatest specificity, but with lower capacity. Assuming Pierce's nickel- and copper-

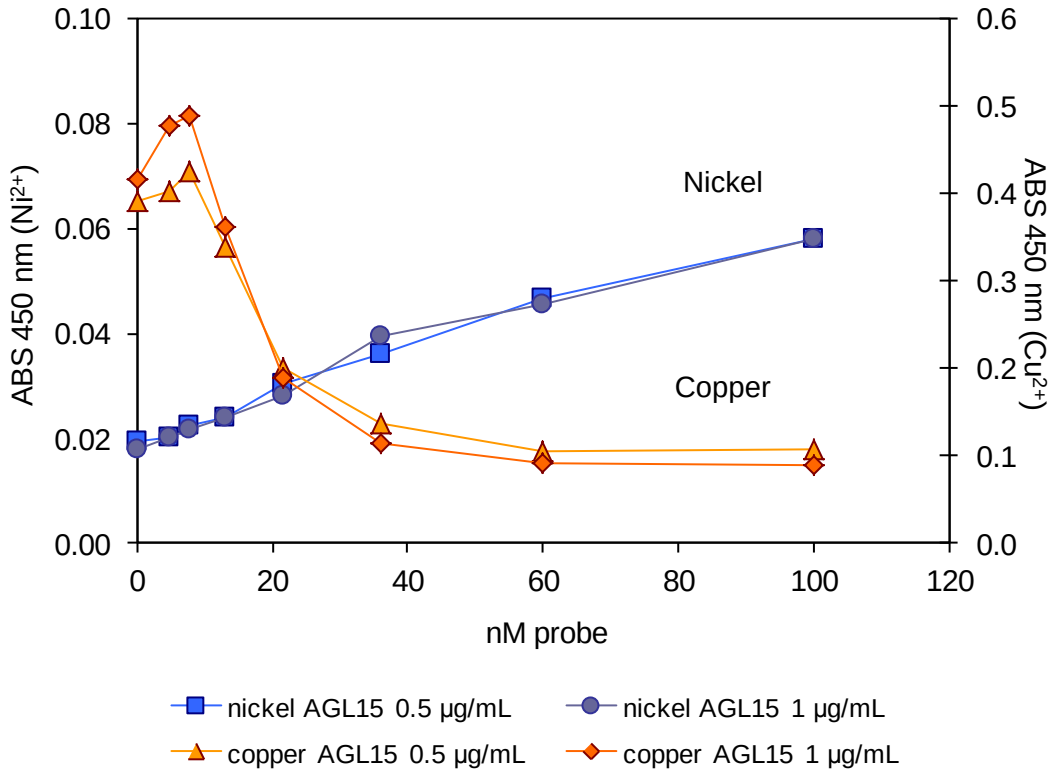


Figure 3.3 Selection of metal ion for "in position" IMAC purification and DPI-ELISA. Two concentrations of re-solubilized AGL15 inclusion bodies were applied to copper or nickel coated plates followed by the addition of probe. The copper coated plates resulted in a much higher overall signal at 450 nm though the signal was inverse with respect to probe concentration. The nickel coated plates had a much lower overall signal, however the signal was linear and positively correlated to probe concentration.

coated plates are coated with the same density of metal atoms, these differences are apparent by Pierce's reported binding capacities. Pierce nickel-coated plates (15142) have a binding capacity of 8.9 pmoles of a 27 kDa HIS-tagged fluorescent protein per well, while their high-binding capacity copper-coated plates (15143) have a binding capacity of 36.5 pmoles. Thus, the copper coated plates have the potential to increase the signal approximately four-fold. However, the reduced specificity of copper could also reduce the signal-to-noise ratio [246].

In a comparison experiment between nickel and copper coated plates, the copper plates did outperform in overall measured signal intensity (Figure 3.3). However, in the copper plates there was a non-linear, inverse relationship between the signal and concentration of probe, while the nickel plates showed the complete opposite. Data for the nickel plate is indicative of nonspecific binding between the probe and the chelated nickel and not AGL15. Neither outcome fit the expected hyperbolic model for one-site binding of an AGL15 homodimer and a single CArG box. Conditions for the assay included metal-chelate binding with resolubilized AGL15 at 0.5 and 1.0 µg/mL total protein, phosphate-based buffer, biotin-labeled probe, and 5 µg/mL NA-HRP.

Avidin is a notoriously "sticky" protein that will bind to proteins and surfaces in general and is a major source for background in ELISAs. NeutrAvidin has been chemically modified to reduce such nonspecific interactions. For both metals, plates are coated with AGL15, followed by the addition of biotin-labeled probe, and then NA-HRP. In the copper plate, biotin-labeled probe likely nonspecifically bound the chelated copper and suggested in the nickel plate. However, the nonspecific interaction between NA-HRP and copper appeared to be much stronger than between NA-HRP and nickel.

A follow-up experiment without AGL15 or labeled probe was performed. Serially diluted sheared salmon sperm DNA was applied to part of a copper plate, while only buffer was applied to another part. Following washing, 5 µg/mL NA-HRP was applied to the part of the plate with DNA and serially diluted NA-HRP was applied to the part of the plate with only buffer. NA-HRP was found to nonspecifically bind chelated copper (Figure 3.4B). Since unlabeled DNA was applied prior to the addition of NA-HRP and possessing a high-affinity for copper, it was very effective in blocking the nonspecific binding of NA-HRP. This effect was not observed on the nickel plate, which would suggest that the

affinity of DNA for copper is large and to a degree specific. Additionally, NA-HRP has significant affinity for copper, though not so large as to be able to displace DNA, while NA-HRP showed negligible affinity for nickel and permitting the detection of nonspecific

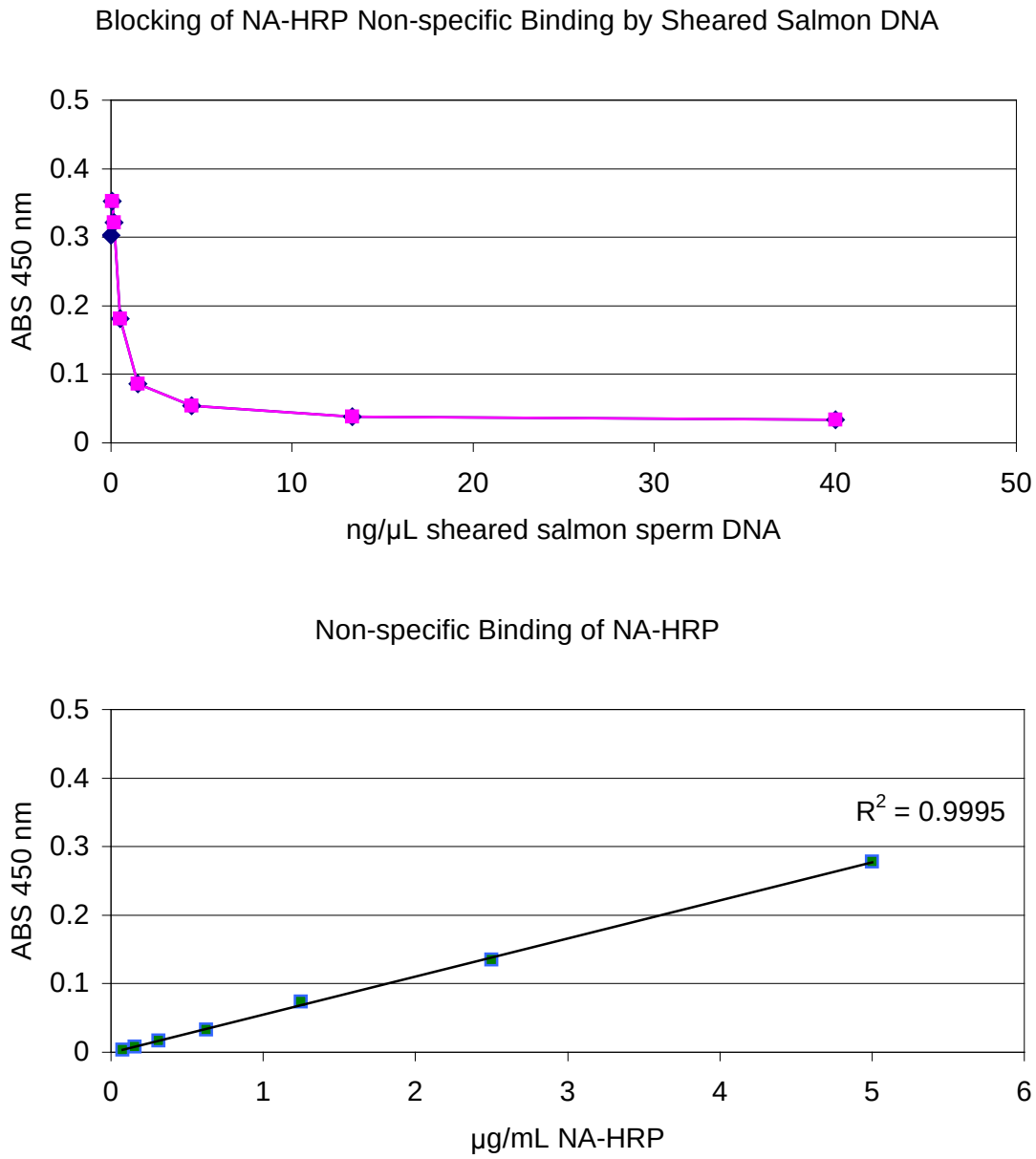


Figure 3.4 (A) Blocking of NA-HRP nonspecific binding by sheared salmon sperm DNA. Sheared salmon sperm DNA drastically reduced all signal from the assay and omitted because of reduction in signal without significant improvements in noise reduction. **(B) Nonspecific binding of NA-HRP to nickel coated plate with adsorbed AGL15.** The addition of NA-HRP was linearly correlated to concentration and indicated nonspecific binding.

binding by the biotin-labeled DNA towards the chelated nickel.

One might expect that since the probe is labeled, NA-HRP should be able to specifically bind the biotin-label. Two factors come into play. Since NA-HRP was nonspecifically interacting with copper and the probe was applied prior to the addition of NA-HRP, there was a reduction in the availability of unbound copper in which NA-HRP can bind. The second factor is that despite the probe being biotinylated, the interaction between avidin and the biotin was sterically hindered. Since the probe was nonspecifically interacting with the copper, conceivably along the entire length of the probe, it was likely that the avidin was unable to maneuver around the biotin. Furthermore, if AGL15 has not been displaced, it too was likely blocking NA-HRP from accessing the biotin as a tree canopy shades the forest floor from sunlight.

Copper plates were therefore deemed to likely be more problematic than nickel plates. In copper plates, blocking would require significant optimization for both the nonspecific interaction of NA-HRP and probe with no guarantee that AGL15 is not undergoing significant displacement.

Buffer system and protein concentration

Buffer guidelines for performing IMAC protein purification recommend using phosphate based buffers and against the use of buffers with secondary or tertiary amines. Secondary and tertiary amines are known to reduce Ni^{2+}; rendering the ion unable to coordinate histidines present in polyhistidine tags. Furthermore, amine containing buffers can also participate in coordination with the metal ion.

ELISA and surface plasmon resonance (SPR) use methodologies that are very similar to each other. In the case of IMAC, NTA sensor chips and protocols have been worked out for SPR. These protocols provided invaluable insight on the behavior of the nickel-polyhistidine interaction.

A critical aspect of a strong interaction that can often be overlooked are the rebinding effects. Even in an ELISA, if excessive quantities of protein are used to coat a plate, the resulting calculations will be skewed from the true values. In SPR, it was found that the binding stability is improved at lower concentrations of HIS-tagged protein [244].

Continuous rebinding was the critical for stable binding of the lower affinity polyhistidine–nickel interaction and continuous rebinding was only made possible if free nickel sites were available. In order to provide free nickel sites, the amount of protein applied must be low. Therefore, the amount of AGL15 applied was also evaluated.

Buffer condition effects on the binding of HIS-tagged proteins to NTA sensor chips have been described by L. Nieba [244]. Certain considerations can be made that can improve the affinity of this interaction. Buffer components that have the greatest reduction in binding include increased pH and salt concentration. Compared with TRIS and

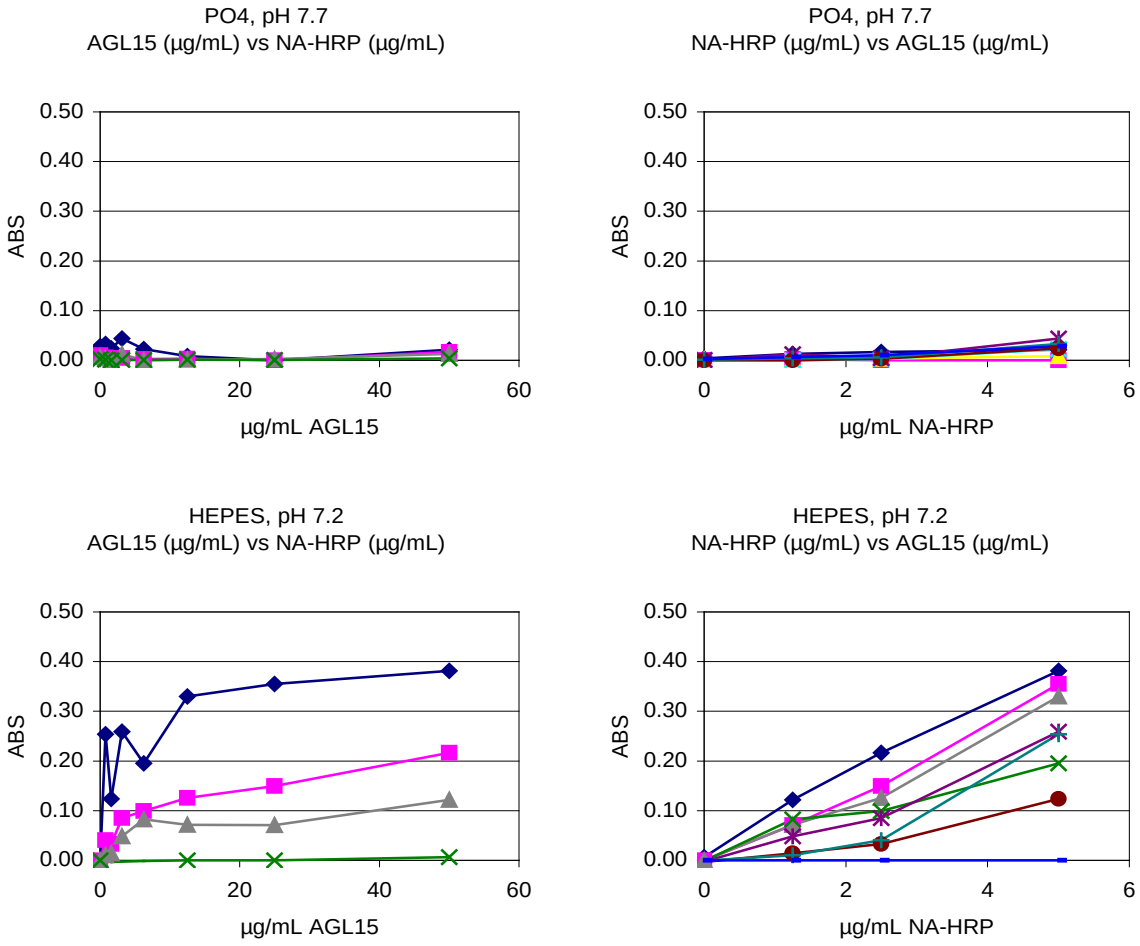


Figure 3.5 Buffer conditions influence binding. Phosphate buffer resulted in low signal which is likely caused by known interactions with the chelated nickel ion. HEPES buffer however, permitted greater stabilization of the immobilized complex resulting in higher signal yields and lower signal-to-noise ratios. AGL15 showed a hyperbolic curve profile upon immobilization, while NA-HRP showed a linear profile.

phosphate buffers, HEPES has been reported provide better detection sensitivity [247]. EDTA is added to sequester trace metal ion impurities that can bind to the polyhistidine tag and reduce immobilization to the plate [244]; of course, too much EDTA will strip nickel ions from the plate. Two buffers were selected for a side-by-side comparison, a phosphate buffer composed of 10 mM sodium phosphate, 50 mM potassium chloride, 50 µM EDTA, 0.005% Tween 20, 10% (v/v) glycerol, pH 7.8 and a HEPES buffer composed of 10 mM HEPES, 50 sodium chloride, 50 µM EDTA, 0.005% Tween 20, 10% (v/v) glycerol, pH 7.2 (Figure 3.5). Assay conditions included nickel coated plates and all washes and components diluted in the respecting buffers. Furthermore, it was also suspected that NA-HRP was also nonspecifically binding AGL15 and that this property could be later used to assess the binding of HIS-tagged AGL15 homodimer to the chelated nickel. Thus, no probe was used.

In the phosphate buffer system, the signal measured was exceedingly low indicating that either the nonspecific binding of NA-HRP to AGL15 was weakened or the capture of HIS-tagged AGL15 by the chelated nickel was poor (Figure 3.5A,B). However, the HEPES buffer system revealed several important pieces of information. NA-HRP was observed to nonspecifically bind AGL15 and the nickel-coated plate (Figure 3.5D). Without AGL15, the signal increases linearly as the concentration of NA-HRP increases. As the concentration of AGL15 applied increases, the signal increases further. When the amount of NA-HRP applied was held constant, a hyperbolic saturation curve was observed within the range of AGL15 concentrations applied. In the case of AGL15 homodimer, the multimeric complex has two octahistidine tags and therefore, technically two dissociation constants. At 5 µg/mL NA-HRP, the apparent dissociation constant was found to be approximately 79 nM and approximately 65 nM at 2.5 µg/mL NA-HRP for an average apparent total K_D of 72 nM (assuming no free monomers were present). Reported apparent K_D values widely vary [244,248-250]. The apparent K_D for a free hexahistidine polypeptide is 14 nM [248], while the apparent K_D for a hexahistidine-streptavidin is 121 nM [244].

Blocking agent

At this point, it has been established that NA-HRP nonspecifically binds both AGL15 and the plate and to a degree, A/T rich DNA probes also nonspecifically bind chelated nickel. Proteinaceous blocking agents are commonly used to uniformly block sites that would

otherwise be occupied by components such as avidin. Both bovine serum albumin (BSA) and non-fat dry milk (NFDM) were tested for effectiveness in blocking nonspecific binding of NA-HRP to AGL15 (Figure 3.6).

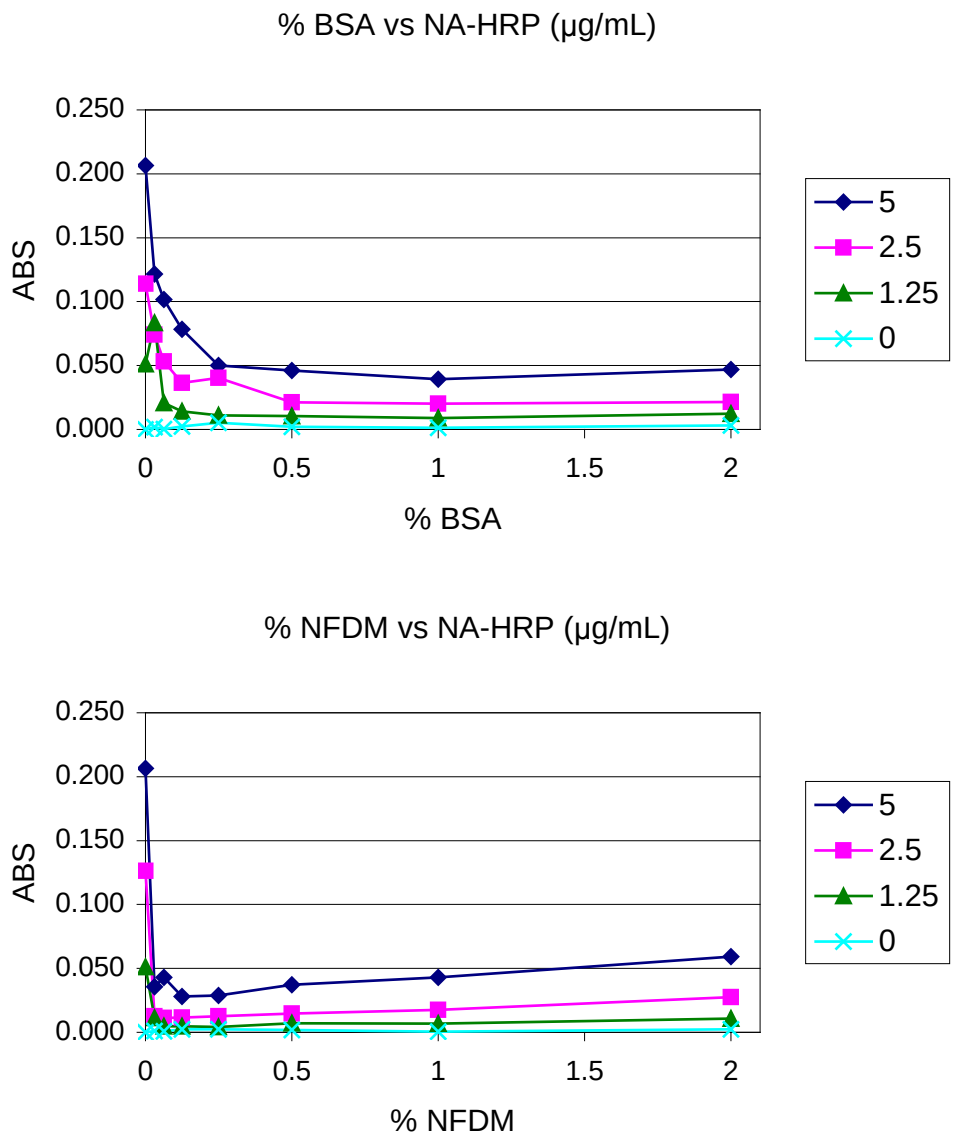


Figure 3.6 Selection of a blocking agent towards NA-HRP. Both bovine serum albumin (BSA) and nonfat dry milk (NFDM) were tested as blocking agents against NA-HRP. While NFDM controlled background signal it severely reduced detection of the signal. BSA also strongly reduced signal at higher concentrations, but concentrations below 0.25%, nonspecific binding of NA-HRP was reduced.

Building upon previous findings, 3 µg/mL AGL15 (~70 nM dimers) was chosen as concentration for coating the nickel plate. Following binding, serial dilutions of BSA or NFDM in buffer were applied to the plate starting at 2%. Blocking was permitted for 30 minutes. Upon removal the blocking agent, various concentrations of NA-HRP were applied. No probe was used.

BSA performed as expected (Figure 3.6) in reducing the nonspecific binding by NA-HRP with increasing percentages of BSA. This was not the case for NFDM as the signal was barely detectable across all percentages. NFDM is known to contain endogenous biotin and it was assumed this free biotin was likely occupying the binding sites in the avidin of NA-HRP. Subsequently, untethered NA-HRP would be washed away. Both BSA and NFDM were effective in reducing background signal. BSA was predictable and selected as the blocking agent.

HRP positive control

In order to monitor direct binding of HIS-tagged proteins to the metal coated microtiter plates during assay development, a HIS-tagged HRP control was desired. Unfortunately, this product was not commercially available at this time. A literature search revealed that a synthetic version of recombinant HRP (rHRP) has been made before in *E. coli*, but that it is poorly expressed [251]. Furthermore, it requires heme as a cofactor, which is very limited within *E. coli*.

HRP has been a favorite for studying enzyme mechanics and engineering improved variants. HRP1A6 is one such variant. HRP1A6 is an evolved variant with a N255D mutation that has a 14-fold higher than wild-type at ~110 µg/L of culture. The mutation is believed not to increase the enzyme's activity, but rather increase the amount of correctly folded protein [252]. HRP1A6 was kindly provided by Dr. Frances Arnold, California Institute of Technology, in the plasmid, pETpelBHRP1A6–KAN. This plasmid contained an unused 6xHIS-tag. A small adaptor was designed in which the stop codon was removed and ligated into the plasmid. The plasmid was transferred to BL21(DE3) and rHRP1A6:6xH expressed under IPTG induction. Cells were lysed with 1X CelLytic B (Sigma) followed by clarification of the lysate in a microcentrifuge. For proper activity, hemin chloride (FW 652.0, H-5533, Sigma) in DMSO (1 mg/mL) was added to the extract in order to supply the heme prosthetic group. Dilutions were made to titrate the

extract. Binding to nickel coated microtiter plates occurred over an hour followed by four washes in wash buffer. TMB substrate was used to detect the presence of bound HRP. A negative control lysate was also included.

Validation via EMSA and determination of the apparent dissociation constant for AGL15

Validation of the AGL15^{MIK} binding capabilities was performed using EMSA (Figure 3.7) and the AGL15 high-affinity binding site (HABS) (CTATATATAG) [243]. AGL15^{MIK} plus probe yielded a shift towards a larger molecular mass, indicting formation of a protein–DNA complex as expected. Following successful validation and optimization of assay components, the apparent dissociation constant for AGL15 was measured.

The apparent dissociation constant between AGL15 and the AGL15 high-affinity binding site was measured using DPI-ELISA (Figure 3.7). Probe concentration ranged from 0.54–90 nM. Data was fitted with GraphPad Prism 5.0 using the model for one-site, total

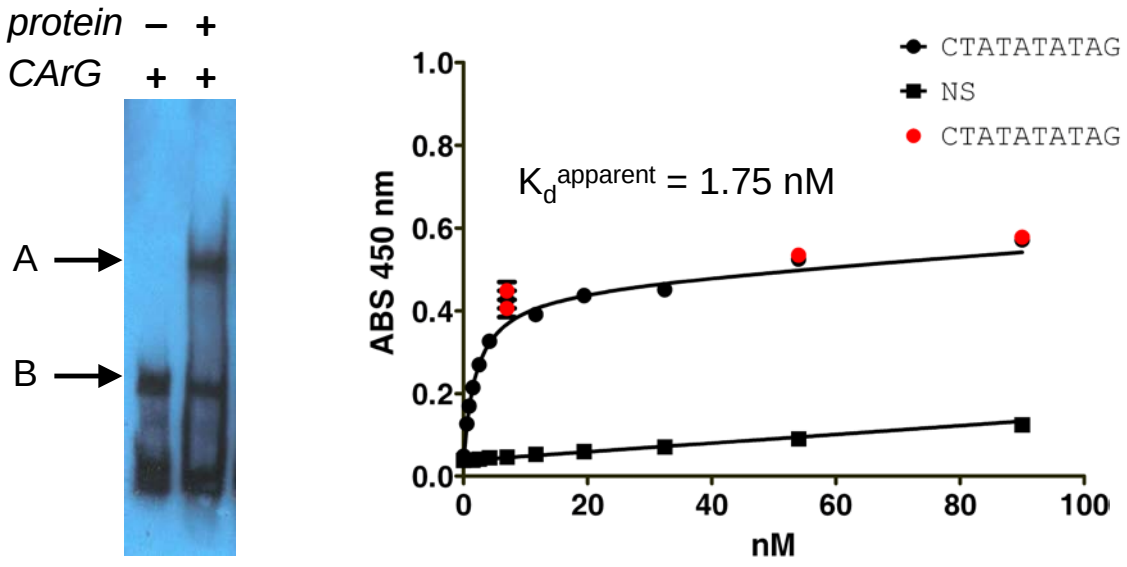


Figure 3.7 Validation of recombinant AGL15 by EMSA and determination of the apparent dissociation constant for AGL15 and HABS. EMSA with re-solubilized AGL15 incubated with or without the AGL15 HABS probe. In the presence of protein, a second band of a higher molecular weight, "A", was detected upon probing for biotin labeled probe indicating a shift in binding from unbound DNA probe, "B". Using the generalized assay, the apparent K_d for AGL15 and HABS was determined to be 1.75 nM using the one-site model (fit total and nonspecific binding).

binding and the calculated apparent dissociation constant between AGL15 and the AGL15 high-affinity binding site was determined to be K_d^{app} = 1.75 nM (B_{max} = 0.3993 ABS@450nm). This value indicates a slow rate of dissociation of AGL15 from the high-affinity probe at equilibrium and therefore, a high-affinity of AGL15 towards this specific probe.

Screening SEP3 against a panel of palindromic probes for relative binding

DPI-ELISA is particularly amendable towards screening probes with specific and distinct sequences. This allows for the possibility of rationally screening many different sequences for their interaction with a particular MADS domain. Measurement of the relative binding affinities of SEP3 towards a panel of palindromic DNA sequences (Appendix) was evaluated using the adapted DPI-ELISA. Both CCWWW and CTWWW variants yielded fold changes above 10 compared to the nonspecific probe (TTCGCGCGTT) with exception to CCTAT 6.5 (Figure 3.8).

Relative changes in signal intensity, correspondent to relative affinity, agreed with previous reports of specificity with regard to variations made at position -4:4 using the palindromic AGL15 HABS (CTATATATAG) as the reference sequence. Variants with pyrimidines (C/T) ($\bar{x}$=27.0) in position -4:4 had fold differences that were on average 12-fold higher than variants with purines (A/G) ($\bar{x}$=2.2) in the same position (Figure 3.9).

Variations at position -3:3 within the AGL15 HABS reference sequence and a CC form (CCATATATGG) did not display a pyrimidine versus purine distinction, consistent with the consensus sequence (Figure 3.9). C(C/T)(A/T)TA variants ($\bar{x}$=20.5) had fold differences on average 4-fold higher than C(A/G)(C/G)TA variants ($\bar{x}$=5.1), while within the C(C/T)(A/T)TA subgroup, C(C/T)ATA variants ($\bar{x}$=27.0) had fold differences on average 1.9-fold higher than C(C/T)TTA variants ($\bar{x}$=14.0). Notably, the CCGTA variant signal intensity was 8.2-fold above the nonspecific probe, indicating that the CCG motif yields some favorability towards a stabile SEP3–DNA complex. In contrast, the CTCTA variant had a signal intensity of 2.2-fold above the nonspecific probe. Overall, CC variants (σ=7.79) appear more moderate and have less spread than CT variants (σ=13.08) with regard to this data set.

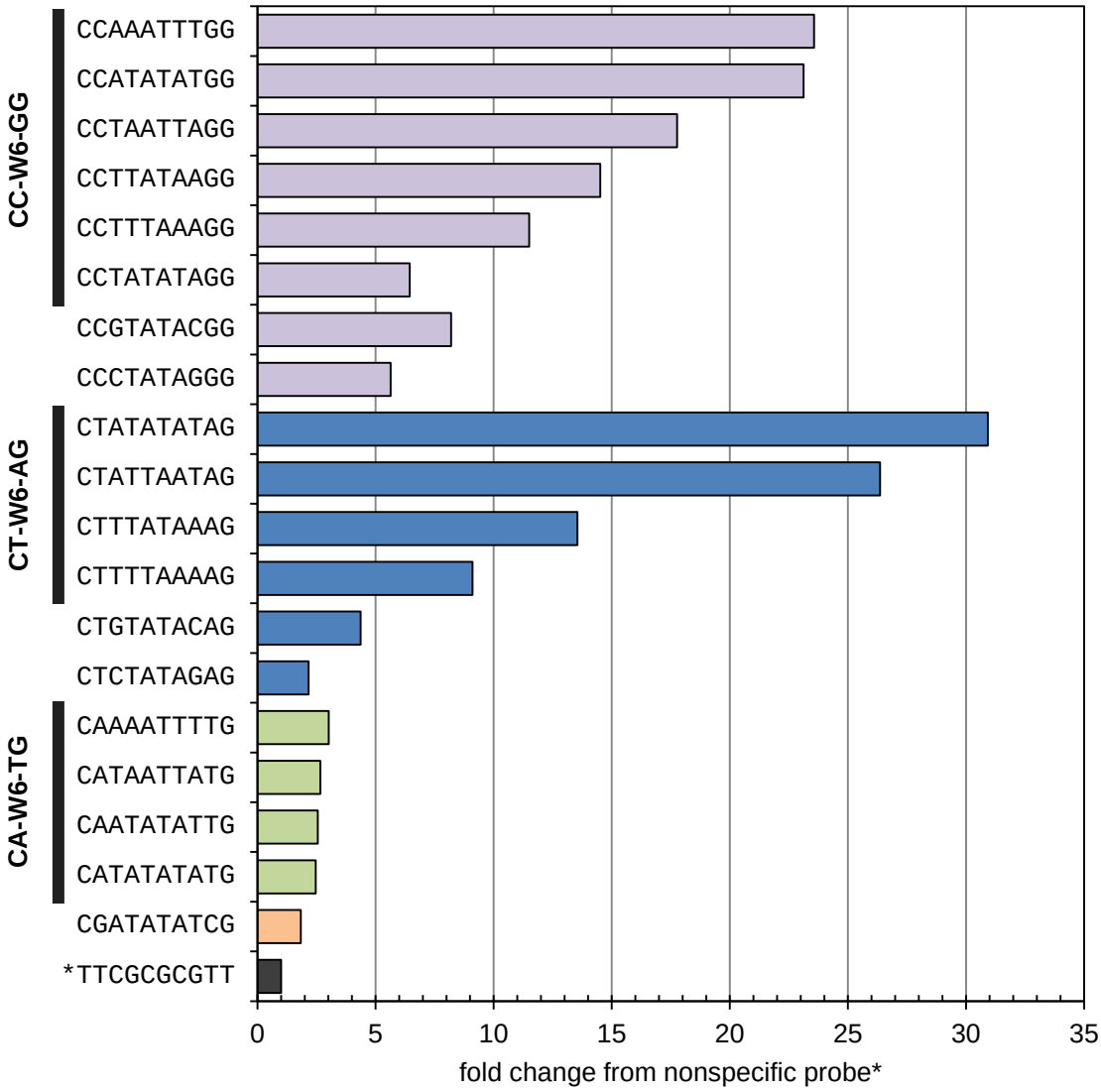


Figure 3.8 Relative binding of SEP3^{MIK} towards a panel of palindromic probes via modified DPI-ELISA. Using the optimized DPI-ELISA method, SEP3^{MIK} was allowed to reach equilibrium with a panel of palindromic probes. Absorbances were measured at 450 nm. Background was subtracted from the data set and then each probe was divided by the value obtained for the nonspecific control (TTCGCGCGTT) to obtain the fold-change. Probes are grouped by sequence characteristics (CC, CT, CA, and CG).

As positions -4:4 and -3:3 are adjacent to one another, binding affinity tends to display some interdependency between combinations of bases. Despite the dominant effects in affinity from these positions, position -1:1 also has a role in stabilizing binding separate that of positions -4:4 and -3:3 (later examined in Chapter 4). A trend observed within the panel of probes evaluated revealed that position -1:1 influenced the binding affinity

between the probe and SEP3 (Figure 3.9). This trend was observed among C(C/T/A)WWW variants. Those probes with an AT ($\bar{x}$=19.2) at position -1:1 consistently yielded higher fold differences above the nonspecific probe with SEP3 than a TA ($\bar{x}$=13.4). A thymidine at position -2:2 appears to compensate for the lower affinity found with a TA at position -1:1 within the CCT subgroup (CCTAT (6.5) < CCTTT,CCTTA,CCTAA ($\bar{x}$=14.6)).

Position -4:4

CCATA	23.1
CTATA	30.9
CAATA	2.6
CGATA	1.8

Position -3:3

CCCTA	5.6
CCTTA	14.5
CCATA	23.1
CCGTA	8.2

CTCTA	2.2
CTTTA	13.5
CTATA	30.9
CTGTA	4.4

Position -1:1

CCTAA	17.8	CCTAT	6.5
CCTTA	14.5	CCTTT	11.5
CTTTA	13.5	CTTTT	9.1
CTATA	30.9	CTATT	26.4
CATAA	2.7	CATAT	2.5

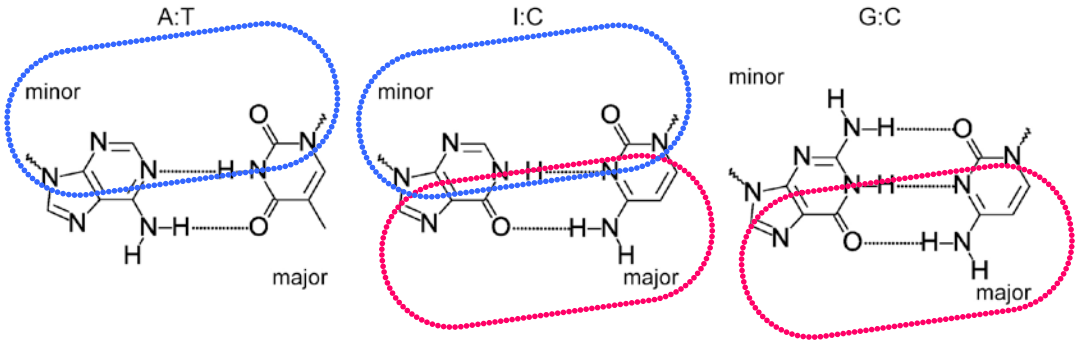
Figure 3.9 Fold-differences for SEP3^{MIK} and series selected palindromic probes. Fold-differences determined for select palindromic probes based upon nucleotide rotations at specific positions. Data obtained from DPI-ELISA.

An I:C base pair to mimic the minor groove of an A:T base pair at positions -4:4

Binding of MADS domains to CArG boxes is particularly sensitive to bases located in positions -5:-4 and 4:5 of the CArG box. As a result, consensus sequences appear in two forms, CC-W6-GG and C-W8-G (more preferably, CT-W6-AG). In either case, both are variants of the form CY-W6-RG. However, some MADS TFs, such as AGL15 and FLC, possess greater specificity towards the C-W8-G form than the CC-W6-GG form. Those that bind both forms, such as SEP3, could be said to possess relaxed specificity.

A distinguishing characteristic between CC-W6-GG and CT-W6-AG forms, is the availability of a free C2 keto group of thymine in the minor groove of the A:T pairs at positions -4:4 in the CT-W6-AG form (Figure 3.10A). The availability of the C2 keto group of thymine as a hydrogen bond donor results from the absence of a C2 amino

A



B

CArG-box	major groove seen as	minor groove seen as
5′-C**C**AAATTT**I**G-3′	5′-C**C**AAATTT**G**G-3′	5′-C**T**AAATTT**A**G-3′
3′-G**I**TTTAAA**C**C-5′	3′-GGTTTAAA**C**C-5′	3′-GATTTAAA**T**C-5′
5′-C**C**ATATAT**I**G-3′	5′-C**C**ATATAT**G**G-3′	5′-C**T**ATATAT**A**G-3′
3′-G**I**TATATA**C**C-5′	3′-G**G**TATATA**C**C-5′	3′-G**A**TATATA**T**C-5′

C

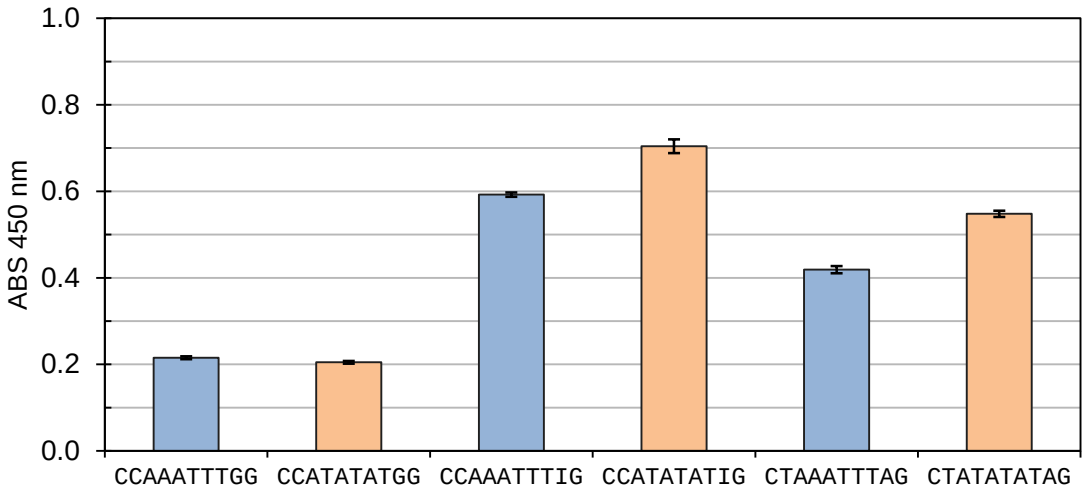


Figure 3.10 Mimicking an A:T base pair with an I:C base pair at positions -4:4. (A) View of major and minor grooves of A:T, I:C, and G:C base pairs respectively. The blue highlights how the minor groove of the I:C base pair resembles the minor groove of the A:T base pair. The red highlights how the major groove of the I:C base pair remains similar to the G:C base pair. (B) Selected CArG box sequences (left), when substituted with inosine resemble the respective major and minor grooves of sequences of which AGL15 has higher and lower affinity for. (C) Results of the optimized DPI-ELISA with AGL15 and inosine-containing probes and their respective mimic sequences for comparison.

group in the corresponding adenine. In the CC-W6-GG form, cytosine pairs with guanidine and utilizes all available hydrogen bond acceptors and donors of cytosine. Therefore, it is likely that MADS TFs, such as AGL15 and FLC, do not intrinsically possess the appropriate amino acids in the appropriate locations to stabilize an

interaction with a CC-W6-GG form *cis*-element. Thus, for a MADS TF, such as AGL15 or FLC, to form a stabile complex, it is likely that a hydrogen bond donor must be available for the MADS domain interact with at position -4:4. For MADS TFs, such as SEP3, which have relaxed specificity, it is likely that this group of MADS TFs do not require the presence of an available C2 keto group at position -4:4 in order to form a stabile protein–DNA complex. However, it is likely that some other mechanism compensates for the stability that the C2 keto group provides.

Using modified and uncommon nucleobases, the target(s) for binding specificity can be determined with greater clarity than with typical single base pair substitutions. With modified and uncommon nucleobases, binding characteristics can be deconstructed and reduced to either the major or minor groove and even to particular functional groups. Inosine is an uncommon, yet natural, purine. Resembling guanidine, it lacks the C2 amino group. When paired with cytosine, the cytosine C2 keto group is made available as in an A:T base pair. Thus, the inosine:cytosine (I:C) base pair mimics the functional group profile of the minor groove of the A:T base pair while retaining the functional group profile of the major groove of the G:C base pair (Figure 3.10A).

AGL15 has higher affinity towards CT-W6-AG forms, such as CTAAATTTAG and CTATATATAG, and has lower affinity for CCAAATTTGG and CCATATATGG sequences (Figure 3.10C). Therefore, it was hypothesized that if the guanidines at positions -4:4 of CCAAATTTGG and CCATATATGG were substituted with inosines, AGL15 would bind with higher affinity. The substituted probes would resemble the minor groove of CTAAATTTAG and CTATATATAG, respectively, while the major grooves would resemble CCAAATTTGG and CCATATATGG, respectively (Figure 3.10B). Binding with high-affinity toward the substituted probes would be akin to a gain-of-function.

The binding capability of AGL15 to the six probe sequences (above) was evaluated using DPI-ELISA (Figure 3.10C). Both CCAAATTTGG and CCATATATGG probes interacted with lower affinity towards AGL15, while CTAAATTTAG and CTATATATAG probes interacted with high-affinity. Probes CCAAATTTIG and CCATATATIG interacted with higher affinity than for the CT-W6-AG form probes. The ability of AGL15 to interact with high-affinity with the inosine substituted CC-W6-GG form probes indicates that

AGL15 is interacting within the minor groove of the probes at positions -4:4 in a stabilizing manner and not the major groove (considering that the major groove was unmodified). Thus, the binding specificity of AGL15 was found to be oriented toward the functional group profile of the minor groove of an A:T pair at positions -4:4 rather than those in the major groove.

Discussion

These findings demonstrate that DPI-ELISA possesses the ability to resolve the changes in binding of MADS TFs towards small differences between CArG box *cis*-elements and has the capacity for high-throughput analysis.

Assay development and optimization

A microtiter, ELISA-based assay was selected for its ability to handle multiple concentrations and samples easily and efficiently. Additionally, several other groups have had success with similarly based assays. Previous DPI-ELISA designs first bind biotin-labeled DNA probe to the matrix. Here, I modified earlier DPI-ELISA designs to accommodate “in position” protein purification at the initial step (Figure 3.1). Though protein purification has become relatively straight-forward, determining the concentration of correctly folded and active protein, especially from resolubilized inclusion bodies, can be a challenge. Swapping the order of which component is bound first to the matrix, reduces the requirement for knowing the concentration of correctly folded protein applied.

AGL15 as a positive control

The apparent dissociation constant provides a means to understand the role of both large and subtle differences conveyed by changes in the DNA sequence of a binding site. If subtle, the changes in binding affinity cannot be readily assessed through quantitative methods such as EMSA. To qualitatively measure the influences made by seemingly small changes in a *cis*-element or protein that change the protein–DNA interaction, the affinity of that interaction must be determined. Measurement of the apparent dissociation constant, using the optimized DPI-ELISA, between AGL15 and the AGL15 HABS yielded a value (K_d^{app} = 1.75 nM) within the expected range for macromolecular high-affinity interactions (Figure 3.7). An additional improvement over

EMSA is that the optimized DPI-ELISA method can be completed in one day with prepared reagents.

SEP3 binds similar CArG boxes differently

Besides successfully screening of SEP3 against a large sample of uniquely, identifiable probes, some notable observations were made. With the AGL15 HABS (CTATATATAG) as a reference sequence, the CCGTATACGG probe sequence had a signal intensity 8.2-fold higher than of the nonspecific probe sequence (Figure 3.8). Though nowhere near the 30.9-fold difference for AGL15 HABS, it is still nearly 3-fold higher than highest fold change of any CAWWW variants evaluated. Therefore, the binding specificity SEP3 is likely more fluid than given credit. Overall, SEP3 binding to CC-W6-GG variants appeared to have less spread than CT-W6-AG variants, which have lower low values and higher high values.

SEP3 serves a unique role within the MADS TF regulatory network as a common component in a multitude of combinations with other MADS TFs [253]. Therefore, it stands to reason that the binding specificity of SEP3 is relaxed in comparison to MADS TFs involved in limited or specific roles. Several studies have been performed to deduce the consensus sequence specifically for SEP3 through methods such as SELEX and ChIP with the Arabidopsis genome [63,254]. Those methods provided a basis for understanding the overall binding of SEP3 complexes as well as the location of binding within the Arabidopsis genome. However, because SEP3 serves as a common component in many MADS TFs combinations, methods such as ChIP cannot isolate the specific binding characteristics of lone SEP3 [254]. Additionally, even with SELEX, nuances imparted by particular sequence combinations cannot be directly detected and measured in order to determine whether a particular sequence combination would be preferred over another.

An example of subtle sequence combinations with dramatic differences in SEP3 binding was observed between CCTAATTAGG and CCTATATAGG probes. SEP3 had a signal intensity of 17.8-fold with the CCTAATTAGG probe compared to a signal intensity of 6.5-fold with the CCTATATAGG probe (Figure 3.9). Whether or not sequence combinations with low or moderate affinity serve a biological function any different than those with high

affinity is uncertain at this time, but without the ability to classify sequence combinations by affinity, an answer to biological significance will likely remain unanswerable.

AGL15 binding stability with its c*is*-element is a result of the functional group profile of the minor groove at positions -4:4

Compared with the relaxed specificity observed for SEP3, the AGL15 binding preference is for the CT-W6-AG form (Figures 3.8 and 3.10A). Substitution of CC-W6-GG probes with inosine in place of guanidine, at positions -4:4, results in probes with minor grooves that resemble the CT-W6-AG form, while the major groove remains unchanged (Figure 3.10B). Using the optimized DPI-ELISA, AGL15 was observed to bind CC-W6-IG and CT-W6-AG probes with higher signal intensity than the CC-W6-GG probes (Figure 3.10C). This indicated a gain-of-function as a result of changing the minor groove's functional group profile from that of a G:C base pair to one of an A:T base pair (Figure 3.10A). It is likely that SEP3 possess residues at other locations that allow it to compensate for the absence of binding properties afforded by the functional group profile of the minor groove of an A:T base pair. Since AGL15 likely does not possess the compensating residues, it can only be stabilized by the functional group profile of the minor groove an A:T base pair at position -4:4 and thus restricting it to only binding CT-W6-AG forms.

Context and motivation

A significant issue has been the inability to accurately predict biologically active *cis*-elements. For example, when an accession line in wheat (*Triticum monococcum* G2528, allele *Vrn-A^{m}1a*) with a strong vernalization requirement was linked to the VERNALIZATION 1 (*VRN1*) allele, it was hypothesized that a deletion encompassing an identified CArG box located within the promoter of *VRN1* could be responsible for the observed phenotype. Analysis of other accessions with alterations to this region revealed that this CArG box was not essential for vernalization [255].

Furthermore, work performed in tobacco (*Nicotiana tabacum*) by Drs. Yanhong Bai and Sitakanta Pattanaik, University of Kentucky, showed that when several CArG boxes in either the *DFR* or *AN1* (homologue of ZmR [256]) promoters were destroyed in combination, no changes in anthocyanin accumulation or transcript levels were observed (unpublished data). It had been anticipated that since activation of both Nt*DFR* and

Nt*AN1* in tobacco is developmentally regulated, it was likely that MADS TFs would be involved in the regulation of those genes due to the role of MADS TFs in flower timing and specification. Since the changes in gene expression were not observed, it was concluded that these CArG boxes are not essential to the regulation of either Nt*DFR* or Nt*AN1*.

One hypothesis has been proposed for these frustrating dead-ends is that despite the lack of direct biological activity, there may still be a biological function for multiple CArG box sites. It has been proposed that the multitude of *cis*-elements without direct biological function could serve as decoy sites to sequester transcriptional activators, thereby inhibiting gene expression [233]. Clearly, a better understanding of the specifics of protein–DNA interactions for MADS TFs is still needed. TFs are active over a range of affinities, but at what point does this affinity diminish, rendering the TF inactive or unable to bind? The goal of this study was to develop a scalable, sensitive, and simple assay in which to measure the relative affinities between TFs and DNA with very small, subtle variations.

The format was altered from that used for the detection and study of HsNFκB [239] and later plant bZIP and WRKY TFs [240,241]. The premise behind the format for the NFκB DPI-ELISA was to monitor NFκB activation, as only active NFκB binds DNA [239]. Attachment of a single DNA probe to the substrate allowed for “in position” purification of free (active) NFκB from cell lysates followed by simple detection of NFκB with an antibody as in the classical ELISA. This allowed for detection and quantification of active NFκB within tissue and ultimately the measurement of inflammation.

Conclusion

As stated previously, the goal was to measure the relative affinities of TF–DNA interactions with very small, subtle changes to either the TF or DNA. This involved the synthesis of many synthetic DNA probes, which have a higher degree of purity and consistency than from resolubilized and refolded, recombinant proteins. Since a greater degree of control was available from the DNA probes, the DNA probe was selected to be varied. With regard to the TF, because the starting material was recombinant protein, several other factors were considered. Our recombinant protein was a fusion protein

with a polyhistidine tag. It was also formed as inclusion bodies and as a single band by SDS-PAGE. By utilizing IMAC technology, our particular recombinant proteins were able to be resolubilized and refolded, then purified “in position”. Since binding affinity is not affected by protein concentration, knowing the exact concentration of protein was not necessary and allowed one to vary DNA concentration, which is much simpler determine and control. An advantage of using the protein as the component with an unknown concentration is that for determining kinetics, the concentration only needs to be consistent. In conclusion, DPI-ELISA fills a void in the analysis of DNA:TF binding interactions. Here, it permitted the determination of the apparent dissociation constant for AGL15 and several probes. It also permitted the broad evaluation of affinities between SEP3 and a panel of DNA probes.

Copyright © Joshua R. Werkman 2013

CHAPTER 4

Modeling DNA recognition of the FLC–SVP heterodimer via molecular dynamics

Hypothesis

Within the MADS domain family, the amino acid residues involved in DNA-binding have been identified and these residues are highly conserved. These residues have been identified through static X-ray crystallography and mutation analysis predominantly in MADS domains of human (five for *Homo sapiens*) and yeast (four for *Saccharomyces cerevisiae*) origin. Despite the highly conserved nature of these residues, MADS domains show sequence specificities that cannot be explained by these residues, especially considering the limited number of MADS TFs in these organisms compared to the numbers found in plants (107 in *Arabidopsis thaliana*). MIKC MADS TFs of angiosperms can be separated into about 15 clades based upon sequence. It is the variable residues adjacent to the highly conserved DNA-binding residues that provide a means for sequence-specific recognition of *cis*-elements. The FLC and StMADS11 (SVP) clades are two clades with divergent adjacent residues. These two clades display different DNA-binding specificities, can form heterodimers between clades, and function within the same regulatory pathways. Molecular dynamics, computationally intensive simulation of matter at an atomic level, provides a way to access the dynamic nature of DNA recognition. These simulations can be used to address the behavior and role of those variable residues.

Introduction

Most of what is known concerning the binding determinants of MADS domain TFs has been performed with proteins of mammalian and fungal origin. Within the genomes of plants are a disproportionate number of MADS domain TFs compared to animals and fungi and where our basis of understanding lies. Plant MADS domains possess a greater diversity of sequence, which has implications for the understanding of both protein sequences that provide redundant DNA recognition and novel DNA recognition as well as the potential to the better understand the aspects and mechanisms of DNA binding in the MADS domain. By utilizing the sequence repertoire of plant MADS domain TFs, a large pool of biologically relevant sequences, we might be able to

discover different routes to DNA-binding site (DBS) recognition within the MADS domain and a better understanding of noncanonical site selection.

FLOWERING LOCUS C (FLC) and SHORT VEGETATIVE PHASE (SVP) represent MADS domain TFs from two different clades within Arabidopsis and are involved in the repression of flowering. While neither FLC nor SVP homodimerize, they do form a heterodimer (Figure 4.1) [257]. Additionally, SVP is the only known MADS domain containing interaction partner for FLC [119]. However, SVP, unlike FLC, is not limited and was shown to interact with nine other MIKC MADS TFs via yeast two-hybrid assays [119]. In Arabidopsis, the FLC clade is comprised of an additional five genes (*MAFS1-5*, *MADS AFFECTING FLOWERING*), while the StMADS11 clade is comprised of *AGAMOUS-LIKE 24* (*AGL24*) and *SVP*. Despite the inability of SVP to homodimerize and AGL24 and SVP to heterodimerize, AGL24 can form homodimers [119,258].

Repression of flowering is essential to the life cycle of plants. While repression of flowering is controlled by a set fundamental factors and mechanisms, evolutionary lineage and pressure from specific climates and environments has altered these regulatory pathways by the addition or subtraction of flowering timing mechanisms. The cue to flower is determined by the integration and final consensus of one or more of the following input pathways: autonomous, photoperiod, and vernalization/dormancy. Vernalization, in winter annuals, and dormancy, in perennials and woody plants, is the block to flowering lifted by cold temperature.

FLC	SVP
SVP	AGL6
	AGL14
	AGL15
	AGL21
	AP1
	CAL1
	FLC
	SEP1
	SEP3
	SOC1

Figure 4.1 Interaction partners of FLC and SVP. FLC has only one identified MIKC MADS dimerization partner, SVP and does not homodimerize. SVP has ten identified MIKC MADS dimerization partners and also does not homodimerize. Other members of the FLC clade are notably absent with regard towards dimerization with SVP.

FLC clade description

Like most other MADS TFs in Arabidopsis, homologs of the FLC clade were anticipated to be discovered in most plants. However as genomic and transcriptomic sequencing projects began releasing data, it was soon realized that the FLC clade was not as wide spread as originally expected. The FLC clade has been found to be predominately limited to the Brassicaceae family, which includes *Arabidopsis* and *Brassica*. Occurrences outside Order Brassicales have only been recently discovered due to expanded genome sequencing efforts. These occurrences have been found in the Asteraceae family of Order Asterales. FLC-like homologs have been identified in the Asteraceae family in *Artemisia annua*, *Barnadesia spinosa*, *Cichorium intybus*, *C. endivia*, *Lactuca sativa* and *L. serriola* with only those in *C. intybus* having been characterized [259].

FLC is thought to have arisen recently during the evolution of plants [96]. Since FLC is a critical component in vernalization and the repression of flowering in plants, such as Arabidopsis, it has been proposed for plants that do not undergo vernalization, no longer maintained its' functionality and have either repurposed it (OsMADS32 [260]) or let it become a pseudogene. Only in plants that have undergone genomic sequencing can it be determined if the FLC clade has dwindled into pseudogene status. There also exists the possibility of convergent evolution towards the FLC clade in Brassicaceae and Asteraceae sourced from a gene duplication event.

FLC expression is heavily dependent upon its chromatin state, which is beneficial for long term regulatory control. For example, the chromatin state is modified during gametogenesis and embryogenesis [261,262] and then at the onset of floral transitioning. Expression of FLC is reactivated during reproduction by the assistance of FRIGIDA, a scaffold protein involved in the recruitment of chromatin modification factors [263]. Repression of FLC occurs through both the autonomous and vernalization pathways and therefore, can be redundantly repressed.

StMADS11 (SVP) clade description

The StMADS11 clade is relatively ubiquitous among plants. The twist is that multiple SVP-like copies arise. This is possibly due to the adaption of the niches that each species has evolved for and to compensate for the absence of the FLC clade.

Understandably, floral transitioning and initiation requires the integration of several different conditions [264-266]. FLOWERING LOCUS T (FT) and LEAFY (LFY) [264] are two significant activators of floral transition. FT accepts photoperiod pathway signals through upregulation by CONSTANS (CO) [267]. Repression of FT expression occurs through the FLC–SVP heterodimeric complex.

In the case of the vernalization pathway, VERNALIZATION INSENSITIVE 3 (VIN3) assists with histone deacetylases (HDACs) to inactivate FLC expression [268-270]. Only upon long periods of cold exposure is VIN3 expression induced; though VIN3 alone is not sufficient for vernalization and suppression of FLC expression [271]. Both AGL24 (StMADS11 clade) and AGL19 have been demonstrated to promote flowering by vernalization through an FLC-independent pathway, but are still thought to be regulated by VIN3 [272,273].

SVP expression, on the other hand, relies heavily on cues from the circadian clock [274] and the thermosensory pathway [275]. Deletion mutants of SVP show early flowering [276], while overexpression of SVP yields late-flowering plants [275,277]. The thermosensory capabilities are not redundant in AGL24 (StMADS clade). Inputs from the thermosensory pathway appear to enter downstream of FLC [275].

Transcriptional activator(s) for both FLC and SVP are unknown at this time. It is known, however, that chromatin remodeling is significant for the expression of FLC [268-270]. For SVP, there is a possibility of constitutive gene activation through minimal type promoter elements, which would then be interrupted by floral specific transcriptional repression and further chromatin remodeling. When expressed, SVP transcript levels remain relatively constant between day and night [274,278]. This is in contrast to the diurnal pattern observed for SVP protein in which it more abundant at dawn and morning and decreasing by evening and night under long days [274].

DNA bending

Secondary structure prediction for full-length HsMEF2A, once beyond the MADS and MEF2 domains, is predicted to be relatively disordered. Whereas in MIKC MADS TFs, the entire protein is shown to possess a good deal of defined secondary structure as opposed to long stretches of undefined disordered sequence (data not shown). Based

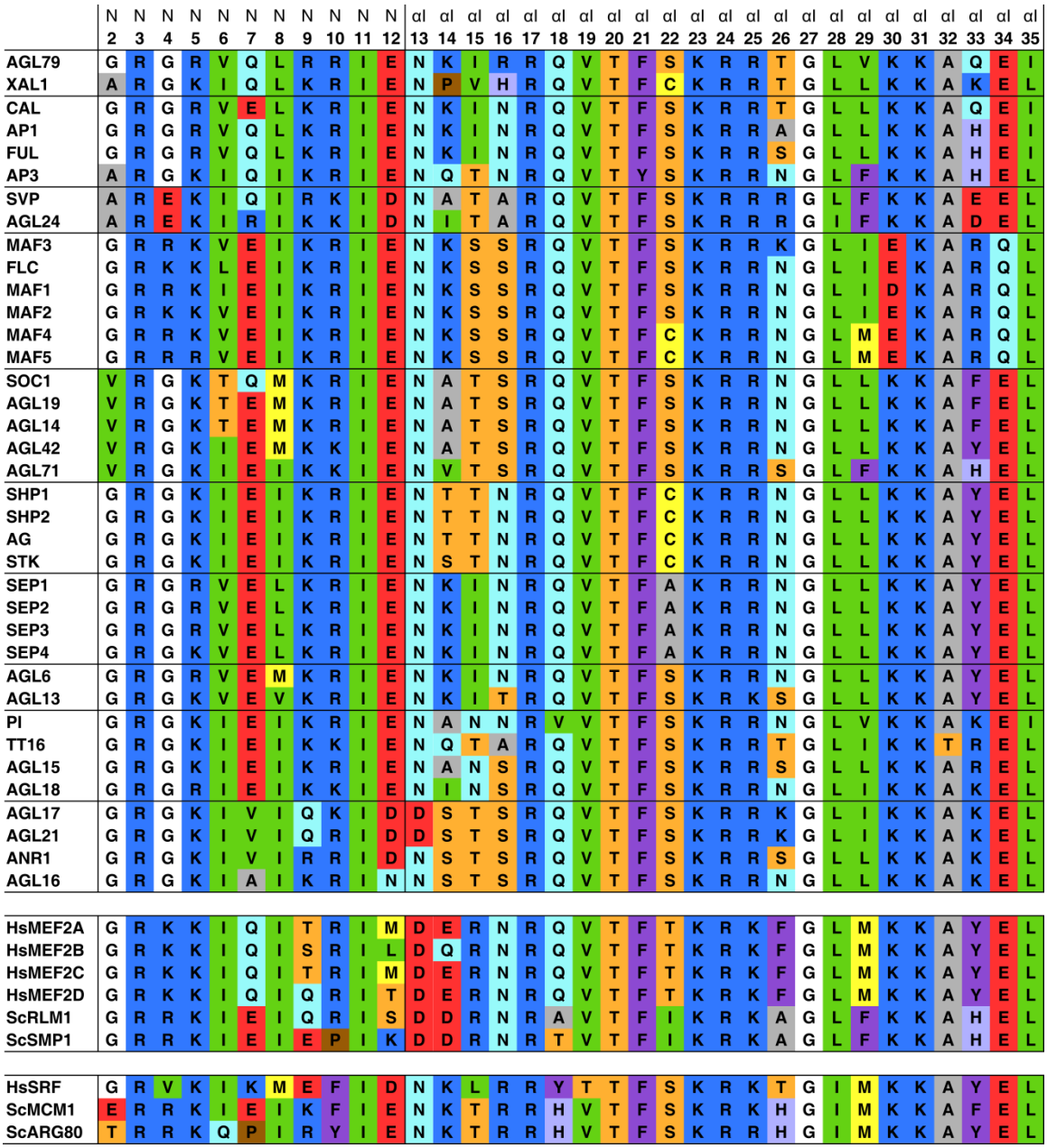


Figure 4.2 Alignment of MADS domain for residues predominately involved in DNA binding. The majority of MIKC MADS domains from Arabidopsis are represented. They are grouped based upon similarities observed at positions containing variable residues and upon clade. Representatives from Type I and II MADS domains from *Homo sapiens* and *Saccharomyces cerevisiae* are also included as a reference. Number positions are based upon AGL15. (N-terminal domain (N), αl helix (αl))

upon what is known for the organization of quaternary structure, MIKC MADS TFs are unlikely to possess this turn back towards the DNA if in fact MEF2A in reality does this as MEF2A, itself, is truncated in the crystal structures [279]. In the case of MIKC MADS

TFs, those involved in the ABC model [280] are thought to form tetramers [281,282]. However, in other complexes such as those for AGL15 homodimer and the FLC–SVP repressor complex that are known not to require upwards of four separate MADS TFs, it is likely that these may only have a dimeric quaternary structure.

Interaction

For MIKC MADS TFs, DNA binding to single CArG boxes is thought to proceed via the dimerization pathway, in which dimers are formed prior to DNA binding [219,281]. Kinetic data for AGL15 does not show cooperative binding as the data is hyperbolic and not sigmoidal (Chapter 3, Figure 3.7). In quartet formation and DNA looping, when two CArG boxes are present, cooperative and phase-dependent DNA binding occurs. Dimer formation occurs prior to DNA binding and looping. DNA binding is represented by K_{d1} for the first dimer-CArG box interaction and K_{d2} for the second interaction. Following this, if DNA looping is to occur, then a K_{d3} exists for the protein–protein interaction between the pair of dimers [281,282].

Materials and Methods

Molecular Dynamics – Computations were performed on the Lipscomb High Performance Computing Cluster (2010-2012) built by Dell, Inc. in McVey Hall on the campus of the University of Kentucky. The basic cluster is comprised of 376 nodes (4512 cores) assembled from Intel Xeon X5650 Westmere 2.66 GHz CPUs with two sockets per node and six cores per socket for a total of twelve cores per node and 36 GB RAM per node. It operates on Red Hat 4.1.2-46 Linux 2.6.18-164.e15. Interconnect is provided through a Mellanox Quad Data Rate Infiniband switch. This provides nearly 40 teraflops of computing power and draws about 180 KW during peak usage. In 2012, four GPU nodes with 36 GB of RAM were added to the cluster comprised four NVidia M2070 GPUs per node.

Protein sequences were obtained from translated Arabidopsis Biological Resource Center cDNA clones for AGL15 (PYAT5G13790; NM_121382), FLC (U89754; BT030637), SEP3 (U90075; BT033157), and SVP (U89996; BT033098). Threaded models of the MADS domain, residues 2–60, were generated using Swiss-Model Automatic Modeling (matching to PDB structure ID 1N6J). Coordinate files for linear B-DNA probes were generated with AmberTools Nucleic Acid Builder (NAB). Protein and

DNA structures were combined and oriented with DeepView and LEaP with the MADS dimer sitting just above the CArG box. Alignments were performed to previous finalized orientations, so that all cases simulated were started in as similar positions as possible. Again using AmberTools LEaP, ff10, parmbsc0 [283], and the Joung ion set [284,285] force field parameters were loaded onto the system. The system was neutralized with potassium counter ions, followed by raising the salt concentration to 200 mM KCl, and hydrated with an 11 Å deep truncated octahedron shell of TIP3P waters. Salt ions were randomized around the DNA, the MADS dimer, and the counter ions with waters by a distance of 5 Å and an overlap of 3.5 Å using AmberTools PTRAJ to disperse them throughout the solvent shell. A copy of the topology file was made and edited to link each strand of DNA and each peptide all into a single molecule to prevent compression of those molecules during minimization, heating, and equilibration of the system. On average, systems were comprised of around 113,000 atoms, of which nearly 109,000 were TIP3P waters.

Minimization, heating, and equilibration (MHE) was performed on CPUs using AMBER 11.0 pmemd.MPI. *Initial minimization* – Systems were initially minimized through 500 steps of steepest descent and 500 steps of conjugated gradient with a non-bonded cutoff of 9 Å, maintaining constant volume, and 25 kcal/mol restraints on the protein and DNA. Steps are in 2 femtosecond increments. *Heating* – Maintaining the restraints and constant volume, the system was initiated with the "weak-coupling ensemble" at 100°K (–173°C) and heated to 300°K (26.9°C) over the course of 5,000 steps followed by 45,000 steps of molecular dynamics with shaking at 300°K for a time of 100 ps. *Equilibration* – A series of five alternating minimizations and equilibrations were then performed starting 5 kcal/mol restraints on the protein and DNA and dropping 1 kcal/mol after each alternation. Settings for each minimization were the same as the initial minimization except for subsequent reductions in restraint force. Each equilibration used the previous minimization's restraint force, a constant temperature of 300°K using the "weak-coupling ensemble," a constant pressure of 1 atm, a coupling constant of 0.2 with shaking, and 25,000 steps for a time of 50 ps. After the equilibration at 1 kcal/mol, the last restrained equilibration was performed at 0.5 kcal/mol, under the previous conditions, while reading the previous velocities. Once the overall system was equilibrated, the original topology file was used; thereby, unlinking the strands of DNA

and peptides. A brief 50 ps equilibration without restraints and a coupling constant of 5.0 readied the system for production.

Production runs were performed on single nodes of four GPUs at a time using AMBER 11.0 pmemd.CUDA.MPI in 1 ns increments at about 8.15 ns/day. Molecular dynamics were conducted at a constant temperature of 300°K using the "weak-coupling ensemble," a constant pressure of 1 atm, a coupling constant of 5.0 with shaking, and a nonbonded cutoff of 9 Å. Translational and rotational center-of-mass motion was removed every 5,000 steps. To keep output files manageable, coordinates were wrapped and outputted in binary NETCDF format. Coordinates were written every 1,000 steps, while the restart file was written every 10,000 steps. Depending upon the system, runs were taken to either 100 or 200 ns.

Preparation of Expression Constructs – Expression constructs were prepared using standard molecular biology techniques. Gene fragments were amplified by PCR, based upon desired domain combinations from *Arabidopsis thaliana* cDNA clones purchased from the Arabidopsis Biological Resource Center (AGL15 (PYAT5G13790), FLC (U89754), SEP3 (U90075), and SVP (U89996)). DNA fragments were inserted into the NdeI and XhoI sites of pET41 for the generation of C-terminally polyhistidine tagged fusion protein without a GST tag. The final arrangement chosen was in the form of NcoI–$MADS^{1–60}$–HindIII–$SEP3^{IK}$–SpeI–XhoI–8xHIS–Stop or NcoI–MIK–SpeI–XhoI–8xHIS–Stop. Site-directed mutagenesis was used to remove a native HindIII site in the K domain of SEP3. Alternative MADS domains can be swapped in as NcoI–HindIII fragments and the K domain provides bulk and an extended surface for homodimerization. Verified constructs were transformed into *E. coli* BL21(DE3) cells.

Preparation of Recombinant Protein – Protein expression was performed overnight with 100 mL ZYM-5052 auto-induction medium with 100 µg/mL kanamycin in 500 mL Erlenmeyer flasks at 37°C, 300 rpm [242]. Cells were pelleted and resuspended in 1X CelLytic B (Sigma B7435), 10 µL/mL protease inhibitor cocktail (Sigma P8849), 1.5 U/mL benzonase (Sigma E1014), and 10 µL/mL lysozyme (10 mg/mL) (Sigma L3790). Lysis was performed osmotically with the addition of five volumes water. Inclusion bodies were pelleted by centrifugation, washed with water, resuspended in storage buffer (50 mM Tris–HCl, pH 8.0, 100 mM KCl, 50% (v/v) glycerol, 1 mM DTT (fresh)),

and stored at –20°C. Total protein concentration of the recombinant protein stock in storage buffer was measured from serial dilutions of solubilized inclusion bodies in 6 M urea using a microtiter plate formatted Bradford assay (Bio-Rad 500-0006) against a BSA (Bio-Rad 500-0007) standard curve also in 6 M urea.

Preparation of Biotinylated dsDNA Probes – DNA oligomers were synthesized by Integrated DNA Technologies (Coralville, IA). The 5'-biotinylated sense strand was in the form of 5'-CCGGGTTTTACTATATATAGTAAAAGGGCC-3' while the antisense strand was without modification. Sequence flanking the CArG box (underlined) was based upon the HABS isolated for AGL15 and a stretch of GC-rich sequence was used for a clamp for either end. Oligomers were reconstituted to 100 mM with Tris–EDTA buffer pH 8.0. Equimolar amounts of each oligomer were combined in a buffer of 75 mM Tris–HCl, 150 mM NaCl, 15 mM $MgCl_2$, 1.5 mM dithiothreitol, pH 7.9 (1.5X NEBuffer #3) for a final individual oligomer concentration of 10 mM. Diluted oligomers were annealed by heating to 95°C in a heat block for 5 min, followed by moving the heat block to a 37°C incubator to anneal. The heat block was moved to a bench at room temperature to finish cooling before placing on ice or storing in the freezer at –20°C.

DPI-ELISA Assay – A calculated amount of purified inclusion bodies for the DPI-ELISA at hand was removed from the protein stock solubilized on-demand in 6 M urea made the same day. With exception for 8 M urea, all solutions containing urea were kept in an ice bath and at no point were any solutions with urea permitted to rise above 25°C. After 30 minutes at 6 M urea, the concentration was incrementally reduced by dilution to a total protein concentration of 200 ng/mL and 0.5 M urea.

Nickel-chelate coated microtiter plates (96-well) (Pierce 15142 or 15442 (0.24 µg/well)) were washed once with wash buffer (10 mM HEPES, 50 mM NaCl, 50 µM EDTA, 0.005% Tween 20, 5% (v/v) glycerol, pH 7.2, 0.25% BSA) and 200 µL diluted, refolded protein was added to each well except for specific control wells. Plates were covered and incubated for 2 hours at room temperature with shaking on a Thermo Scientific microtiter plate shaker (Model 4625-Q) at setting 5 for "in position" purification. Protein solution was then removed by pipette and wells were washed three times with wash buffer. Buffer conditions for binding and washing were based upon buffer conditions for surface plasmon resonance (SPR) NTA sensor chips [244].

Previously prepared biotin-labeled dsDNA probes, 30 bp in length, were diluted in deep-well microtiter plates (1 mL) at varying nanomolar concentrations for measurement of binding kinetics or a single concentration (90 nM) for screening probes. Diluted probes were then applied to the wells accounting for appropriate controls. Plates were covered and incubated for 1 hour at room temperature with shaking on the microtiter plate shaker at setting 5. Probe solution was then removed by pipette and wells were washed three times with wash buffer.

NeutrAvidin-horseradish peroxidase (NA-HRP) conjugate (Pierce 31001) was diluted to 2.5 µg/mL and sterile filtered to remove precipitated clumps, which was critical for assay reproducibility. Plates were covered and incubated for 30 minutes at room temperature with shaking on a microtiter plate shaker at setting 5. NA-HRP conjugate was removed by pipette and wells were washed four times with wash buffer. Plates were developed with room temperature 1-Step Ultra TMB substrate (Pierce 34028) for 5 minutes at room temperature with shaking. The peroxidase reaction was stopped with 2 M sulfuric acid. The developed plate was read at 450 nm in a Beckman Coulter DTX 880 plate reader.

Results

Molecular dynamics

Four systems were generated and simulated using molecular dynamics. The first system is the well-characterized AGL15 homodimer with its HABS (CTATATATAG), a CT-W6-AG form CArG box. Additionally, the SEP3 homodimer binding was simulated with CCAAATTTGG, a CC-W6-GG form CArG box) as well as the AGL15 HABS. The fourth system simulated was of the FLC–SVP heterodimer binding to a known binding-site (CTATTTTTGG) from the *SOC1* promoter [257]. In all cases, the fluctuations in RMSD (Å) of the protein dimers were small compared with those of the DNA fragment (Figures 4.9–4.12). Results are described and interpreted based upon the simulation timescales unless otherwise noted. Stability of the complex can be inferred by a reduction in large movements as well as what is generally known concerning the final confirmation of MADS domain:DNA complexes. Generally, Type II MADS TFs (Chapter 1, Figure 1.4) bend DNA upon binding [286-293]. Therefore, the DNA bending, as determined by the distance between phosphates (P6 and P36) of the DNA fragment, can indicate increased stability of the complex (Figures 4.3 and 4.8B). The initial distance between

A

AGL15-CTATATATAG and SEP3-CTATATATAG

								-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
DC5	C	G	G	G	T	T	T	T	A	C	T	A	T	A	T	A	T	A	G	T	A	A	A	A	G	G	G	C	DC3
DG3	G	C	C	C	A	A	A	A	T	G	A	T	A	T	A	T	A	T	C	A	T	T	T	T	C	C	C	G	DG5
60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31

SEP3-CCAAATTTGG

								-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
DC5	C	G	G	G	T	T	T	T	A	C	C	A	A	A	T	T	T	G	G	T	A	A	A	A	G	G	G	C	DC3
DG3	G	C	C	C	A	A	A	A	T	G	G	T	T	T	A	A	A	C	C	A	T	T	T	T	C	C	C	G	DG5
60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31

FLC-SVP-CTATTTTTGG

								-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
DC5	C	G	G	G	T	T	T	T	A	C	T	A	T	T	T	T	T	G	G	T	A	A	A	A	G	G	G	C	DC3
DG3	G	C	C	C	A	A	A	A	T	G	A	T	A	A	A	A	A	C	C	A	T	T	T	T	C	C	C	G	DG5
60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31

B

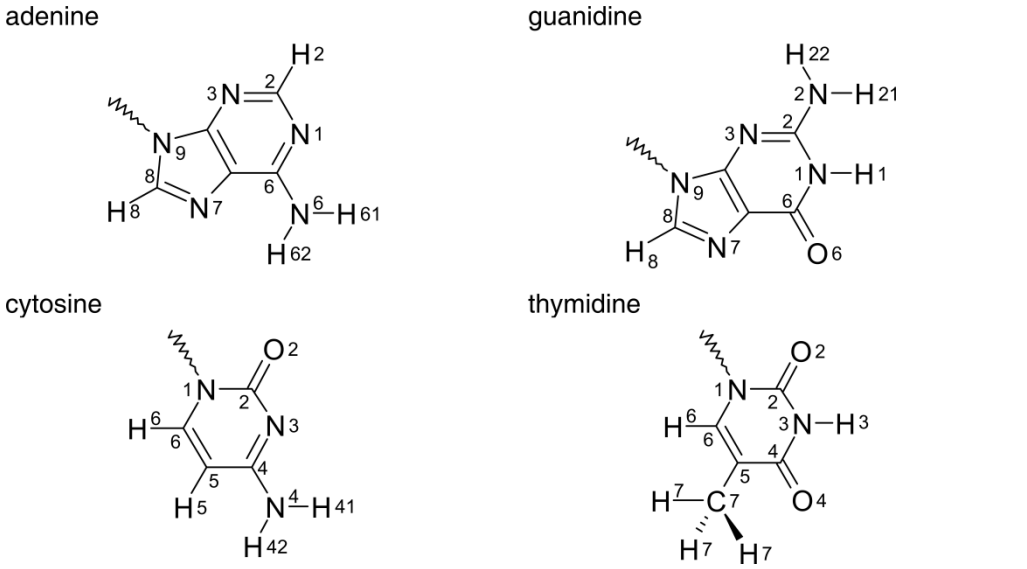


Figure 4.3 Numbering of nucleobase positions in simulated DNA fragments and atoms within individual nucleobases. The numbering of the positions 1–60 is the numbering used by the coordinate file to uniquely identify each nucleotide, while the numbering of the positions -7:7 is used for referencing positions within the CArG box. Positions 6 and 36 are highlighted yellow to indicate the positions of phosphates used in measuring distance (an indicator of DNA bending). (A) AGL15 high-affinity binding site used in both the AGL15 and SEP3 homodimer simulations, a palindromic CC-W6-GG binding site used with SEP3 homodimer, and a known binding-site of the FLC–SVP heterodimer identified from the *SOC1* promoter. It is a mixed form CArG box. DC5, DC3, DG5, and DG3 are representations of nucleotides with end-cap hydrogens. (B) Naming of atoms within the four nucleobases used for the simulations [294].

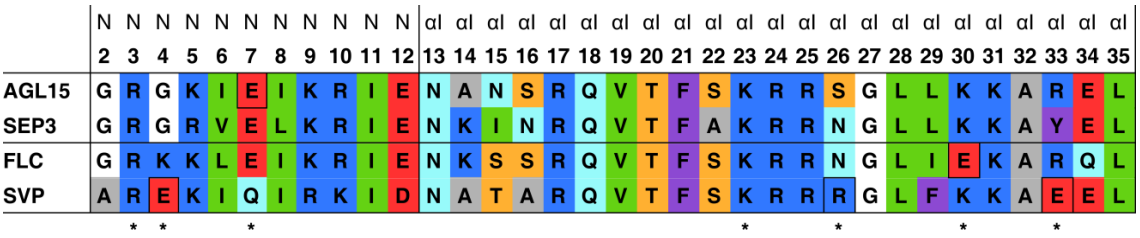


Figure 4.4 Focused alignment of MADS domain for residues predominately involved in DNA binding. Asterisks indicate positions found to be unique or important. Positions 4 and 33 in SVP (and AGL24) is unique with the presence of oppositely charged residues compared with the consensus, which includes animal and yeast proteins. In FLC, position 30 also displays an oppositely charged residue from the rest. Positions 3 and 23 are highly conserved and involved in direct DNA contact. (N-terminal domain (N), αI helix (αI))

P6 and P36 for all DNA fragments was measured to be 58.0 Å. For a graphical description of the parameters used in Figures 4.9–4.12, please see Figure 4.8.

AGL15 homodimer binding of CTATATATAG

Extreme changes in RMSD (>5 Å) for the AGL15 homodimer and DNA alone tapered off within the 30 ns of the simulation (Figure 4.9, panel A). Convergence was not reached until 85 ns after the start of the simulation (Figure 4.9, panel C and D). Bending of the DNA fragment occurred within 30 ns (Figure 4.9, panel C). The homodimer interacted with the DNA fragment by first "falling" to one side (similar to Figure 4.5D), situating the αI helix (AA 14–39) into the major groove of the probe and over positions -5:-4 of the CArG box (Figures 4.3, 4.8A,C, and 4.9, panel D, green). As both the homodimer and the DNA fragment are chemically symmetrical, the direction of this interaction is likely not significant. With an initial distance of 58.0 Å, the distance between P6 and P36 approached 35 Å within 30 ns and generally remained under 40 Å for the rest of the simulation (Figure 4.9, panel C). The distances between LYS23 (in the αI helix) of Chain A and B, to their corresponding binding positions with the DNA fragment (positions -6,-5,-4:+4,+5,+6) reveal that despite the symmetrical nature of the complex, AGL15 homodimer did not symmetrically bind the DNA fragment (Figure 4.9, panel D). The highly conserved ARG3 was used as a representative to track the behavior of the N-terminal domain in each chain. According to hydrogen bond analysis (via ptraj), the ARG3, Chain A formed hydrogen bonds with nucleobase 12 (T) nearly 100% of the time and nucleobase 13 (A) 6.87% of the time (Appendix). ARG3, Chain B

behaved differently by interacting with phosphate backbone as well as intramolecularly with GLU7.

SEP3 homodimer binding of CCAAATTTGG and CTATATATAG

Two sequences, representing both CC-W6-GG and CT-W6-AG forms, were selected for simulation with SEP3 because of the dual-binding capabilities of SEP3. In the SEP3-CCAAATTTGG case, the RMSD of the DNA fragment quickly moderated to around 5 Å and remained consistent for the remainder of the simulation (Figure 4.10, panel A). The αI helices (AA 14–39) for both SEP3 chains showed little change with the RMSD remaining under 1 Å (Figure 4.10, panel B). The stability in the positions of each αI helix was observed in greater detail by calculating the distance between the conserved LYS23 of both chains to their corresponding binding positions with the DNA fragment (positions -6,-5,-4:+4,+5,+6 of CCAAATTTGG) (Figure 4.10, panel D). LYS23, Chain A approached no closer than 10 Å, while LYS23, Chain B came within range of hydrogen bond formation (<3 Å) (Figure 4.10, panel D). As water-mediated hydrogen bonds were not calculated, the degree of coordination between LYS23, Chain A and the DNA fragment is uncertain. The SEP3 N-terminal domains in the SEP3–CCAAATTTGG simulation case interact within the minor groove of the DNA fragment (Appendix). The N-terminal domain of Chain A was observed interacting outside of the CArG box (ARG3 with nucleobases 38 and 39), while the N-terminal domain of Chain B was observed interacting with the central nucleobases (ARG3 with nucleobases 17, 44, and 45) (Appendix). With an initial distance of 58.0 Å between P6 and P36, the distance did not vary greatly from this initial distance (Figure 4.10, panel C) and indicated that protein-induced bending did not occur despite previously mentioned indicators of stability (Figure 4.10, panel A, B, and D).

In the SEP3–CTATATATAG simulation case, the RMSD of the complex did not stabilize as observed with the other simulations (Figure 4.11, panel A). However, the RMSD of both αI helices (AA 14–39) remained around 1 Å (Figure 4.11, panel B). The distance between P6 and P36 was erratic throughout the simulation and did not indicate DNA bending (Figure 4.11, panel C). Examination of the distances between LYS23 and the anticipated DNA interaction positions within the major groove, revealed a stark difference in the movements of the αI helices (Figure 4.11, panel D). LYS23 of the αI helix of SEP3, Chain B appears to have strongly interacted with nucleobases 19, 20, and

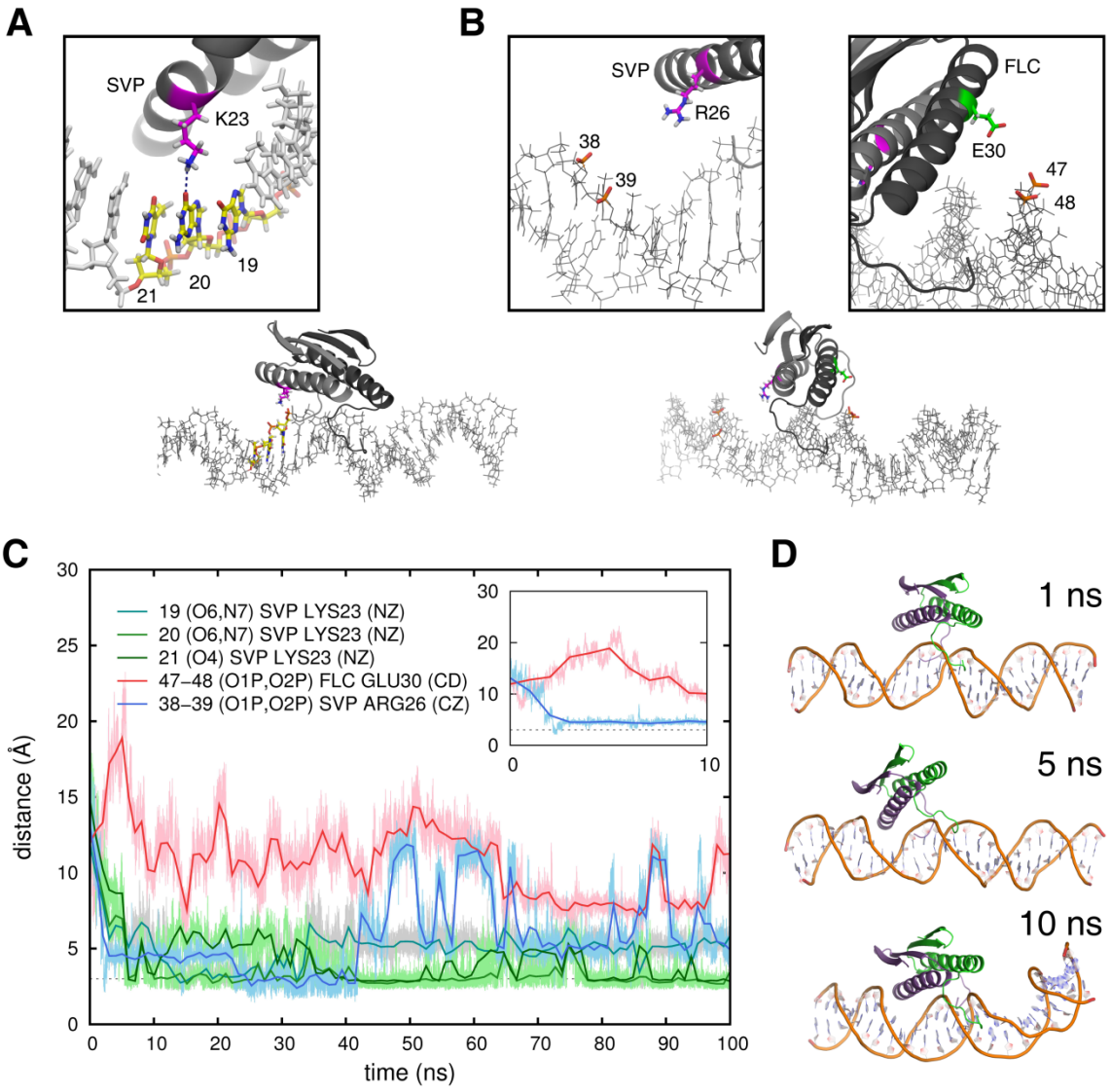


Figure 4.5 Key residues and movements within the FLC–SVP DNA complex. (A) SVP LYS23 (K23) from the αI helix situated in the major groove over nucleobases 19–21 (refer to Figure 4.3). (B) Interactions were observed between the phosphate backbone and specific residues. SVP ARG26 (R26) and FLC GLU30 (E30) were of particular interest. O1P and O2P atoms of nucleotide positions 38, 39, 47, and 48 are identified in relation to these specific residues and only the phosphate and oxygens are highlighted. The center of mass between O1P and O2P of 38–39 and 47–48 was used as a point to indicate the phosphate backbone, while atoms CZ of SVP ARG26 and CD of FLC GLU30 were used as points for the respective residues. (C) Plot of distances between identified points. The center-of-mass between nucleobase 19, 20, or 21 atoms O6 and N7 or O4 were used as points of measurement between lysine atom, N7, indicating how far the αI helix enters into the major groove. The inset graph shows the earliest changes to occur within the first 10 ns. A possible attractive shift between the positively charged SVP LYS26 and the negatively charged phosphate backbone occurred concurrently with a likely repulsion of the negatively charged FLC GLU30 and negatively charged

phosphate backbone. (D) These electrostatic interactions are thought to be a significant cause of the initial movements and initial recognition of the binding site. The asymmetric interaction biases the αI helix of SVP to be the first of the αI helices to interact with the major groove; thus, setting the initial preference for base pair recognition with SVP over FLC. (SVP is shown in purple and FLC is shown in green.)

21 by retaining a close distance. Chain A became very distant with regard to the anticipated DNA interactions (nucleobases 49, 50, and 51). Upon investigation of some of the final coordinate files generated, it appeared that the SEP3 homodimer was dissociating from the DNA fragment by "rolling off the end" of the fragment.

FLC–SVP heterodimer binding of a known *cis*-element, CTATTTTTGG

The flanking sequence of the simulated DNA fragment used for this simulation remained the same as for the previous simulations. The core (10 bp) sequence (CTATTTTTGG) was swapped to that of a characterized and known FLC–SVP binding-site in the *SOC1* promoter (tatgCTATTTTGGtcttt) [257]. It is a mixed form CArG box comprised of a CT-W6-AG form half-site (CTATT) and a CC-W6-GG form half-site (CCAAA). Other known binding-sites for FLC–SVP are also found as mixed forms, e.g. tttaCTTATTTTGGttttt (again from the *SOC1* promoter) [295]. Based upon DPI-ELISA (Figure 4.13), the αI helix of FLC was situated over the CTATT half-site and the αI helix of SVP was situated over the CCAAA half-site.

Within the first few nanoseconds of the simulation, the distance between the phosphate backbone of the DNA fragment (nucleotides 38–39) and SVP ARG26 sharply dropped (Figure 4.5C, inset, blue line), allowing SVP ARG26 to hydrogen bond with the DNA backbone. In contrast with AGL15 (SER26) and FLC and SEP3 (ASP26), basic residues at this position are largely unique to the StMADS clade (Figures 4.2 and 4.6). Meanwhile, the distance between FLC GLU30 and the phosphate backbone of the DNA fragment (nucleotides 47–48) increases (Figure 4.5C, inset, red line). It is thought these two events are coordinated with the like charge between FLC GLU30 and the phosphate backbone serving to directionally repel the αI helix of FLC away from the DNA fragment at its current location (Figure 4.5D and 4.6). It appears that this reinforces and directs the SVP αI helix into the major groove of the half-site of the CArG box (CCAAA). Acidic residues at position 30 are unique to the FLC clade (Figure 4.2).

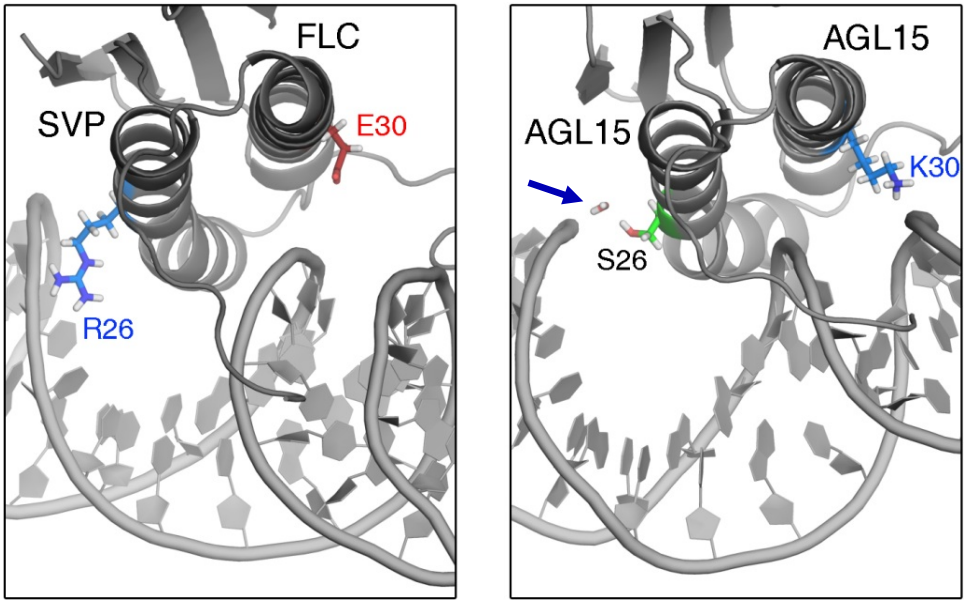


Figure 4.6 Comparative view of the FLC–SVP heterodimer versus the AGL15 homodimer in the early phase of DNA-binding site selection. Arginines at position 26 are unique to just a handful of the MADS domains, while a negatively charged residue at position 30 is even rarer (Figure 4.2). In the case of AGL15 homodimer, the highly conserved LYS30 (K30) is attracted to the phosphate backbone. This appears to be in contrast to what was observed in the FLC–SVP DNA complex. Additionally, AGL15 lacks a strong, positively charged residue at position 26 and may reduce the selectiveness observed by SVP in the FLC–SVP DNA complex. However, a water-mediated hydrogen bond appears to partially compensate (arrow). In an evolutionary perspective, SVP was likely selected for a longer and positively charged residue to 'lock-in' a particular interaction mechanism. In MADS domains, such as AGL15, this interaction mechanism may regulated through phosphorylation resulting in longer, positively charged side-chains.

The status quo was disrupted by the interaction of K^+ ions with SVP ARG26, releasing the hydrogen bonded SVP ARG26 from the backbone (Figure 4.7, panel B, 50 ns) Once released, SVP ARG26 was free to hydrogen bond with SVP GLU33 and only in the StMADS clade are acidic residues found at this position (Figure 4.2). Until the 50 ns mark (Figure 4.12), all parameters measured appear stable.

The disruption of SVP ARG26 from the DNA backbone by K^+ ions coincides with several other changes. Though already relatively stable (SVP LYS23, Figure 4.12, panel D), the SVP αI helix appears to further stabilize with the interaction of SVP ARG26 with K^+ ions with a reduction in the RMSD for SVP αI helix (AA 14–39) (Figure 4.12, panel B, 50 ns).

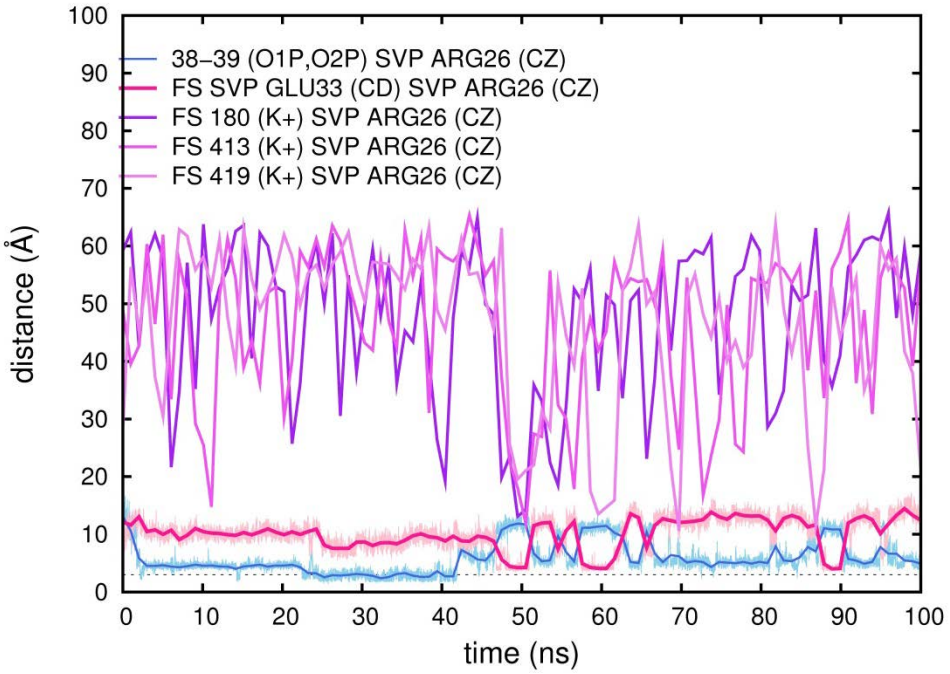


Figure 4.7 Disturbances of interactions by K^+ ions leads the way toward rearrangements within the FLC–SVP DNA complex. SVP GLU33 is adjacent to SVP ARG26 along the αI helix. Throughout the first 40 ns, the interaction between the SVP ARG26 and the phosphate backbone remained stable (also in Figure 4.5C). A disturbance, possibly initiated by K^+ ion movements (purple lines indicate individual ions) around the system or bending of the probe, appeared to permit SVP ARG26 to move slightly away from the phosphate backbone (blue line). Once this occurred, the distance between SVP ARG26 and SVP GLU33 closed (red line), releasing SVP ARG26 from the phosphate backbone and a likely energy minima. Rearrangement of the complex progressed after escaping this minimum.

There was also an increase in the RMSD for FLC's N-terminal domain (AA 2–8), followed by a drop and stabilization in the RMSD for FLC's N-terminal domain (Figure 4.12, panel B, 40–60 ns). Following these observations, a reduction in the distance between P6 and P36 occurred (Figure 4.12, panel C, 60 ns), which is indicative of DNA bending as observed in the AGL15–CTATATATAG case (Figure 4.9, panel C, 20 ns).

As the bend in the DNA fragment was stabilized, FLC LYS23, in the αI helix, moved closer into the major groove of the opposite half-site (CTATT) (Figure 4.12, panel D, 60 ns). This series of events culminate into an apparent convergence about 80 ns after the

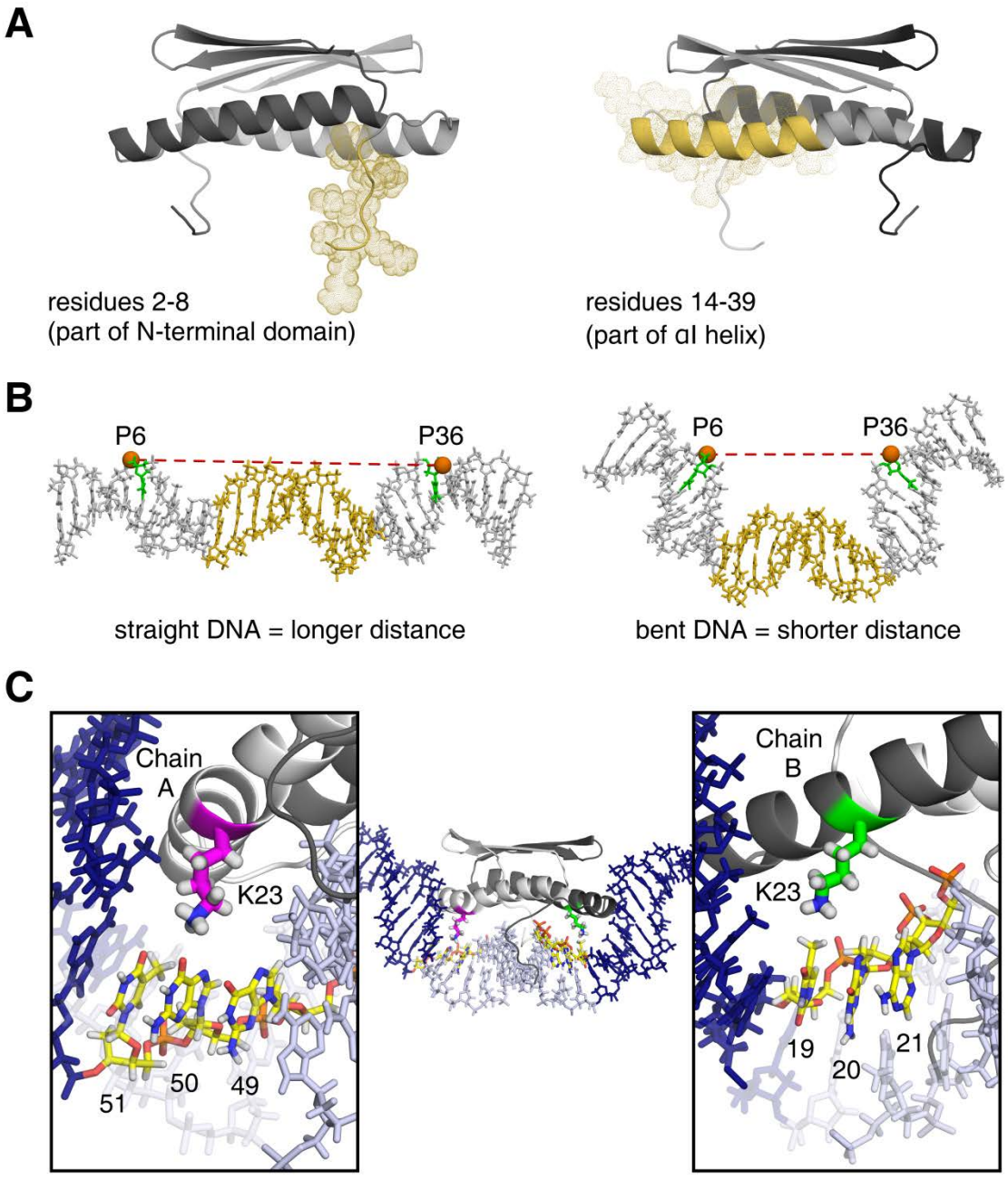


Figure 4.8 Location reference for subsequent figures. (A) The highlighted regions (yellow) correspond to residues used to determine the RMSD values for the N-terminal domain and αI helix of the second panel. The structure is of a dimer and only one of the chains has been highlighted. (B) Orange spheres represent phosphates for nucleotides 6 and 36 (represented in green). The numbering system is listed in Figure 4.3A and is based upon the residue numbering system used for the MD simulations. The central region (yellow) highlights the location of the CArG box. (C) This structural model is representative of a MADS TF–DNA complex. Several key positions are highlighted to

show the relationships plotted in the bottom panel of Figures 4.9–4.12. The conserved K23 (Figure 4.2) is highlighted in either purple (Chain A) or green (Chain B). Nucleotide positions 19, 20, and 21 and 49, 50, and 51 are highlighted in yellow and the expanded views show their positions within the major groove in relation to K23. Figure 4.3B can be used to identify the specific locations of the atoms used in the measurements. For K23, the NZ atom corresponds to the nitrogen at the end of the side chain. The central region (light blue) highlights the CArG box.

start of the simulation. The RMSD of the DNA fragment stabilized below 5 Å (Figure 4.12, panel A) and the distance between P6 and P36 dropped and stabilized around 45 Å (Figure 4.12, panel C). No parts of the αI helices or their end-caps entered into the major groove or otherwise interacted with the bases.

Hydrogen bond analysis (via ptraj) on position 3 for ARG3 of both FLC and SVP revealed that FLC ARG3 interacted with nucleobase 12 (T) nearly 90% of the time, nucleobase 50 (G) for a minimum of 10%, and nucleobase 11 (C) for 7% (Appendix). SVP ARG3 interacted with nucleobase 17 (T) nearly 75% of the time, nucleobase 18 (T) 11%, and nucleobase 45 (A) nearly 45%. Furthermore, SVP ARG3 hydrogen bonded to the adjacent glutamic acid, SVP GLU4, at a minimum 45% of the time. Acidic residues at position 4 are unique to the StMADS clade (Figure 4.2).

Figure 4.9 AGL15 homodimer binding of CTATATATAG.

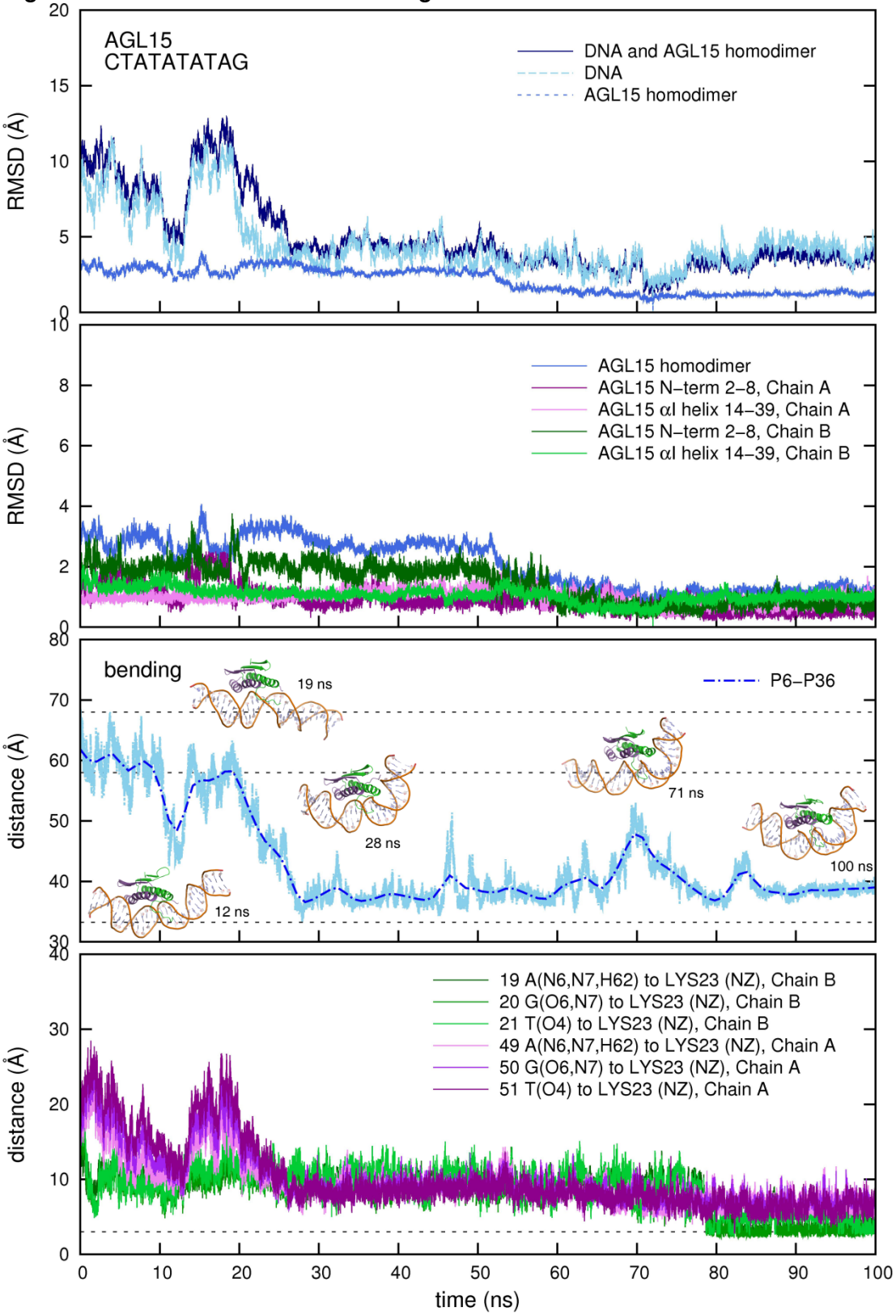


Figure 4.10 SEP3 homodimer binding of CCAAATTTGG.

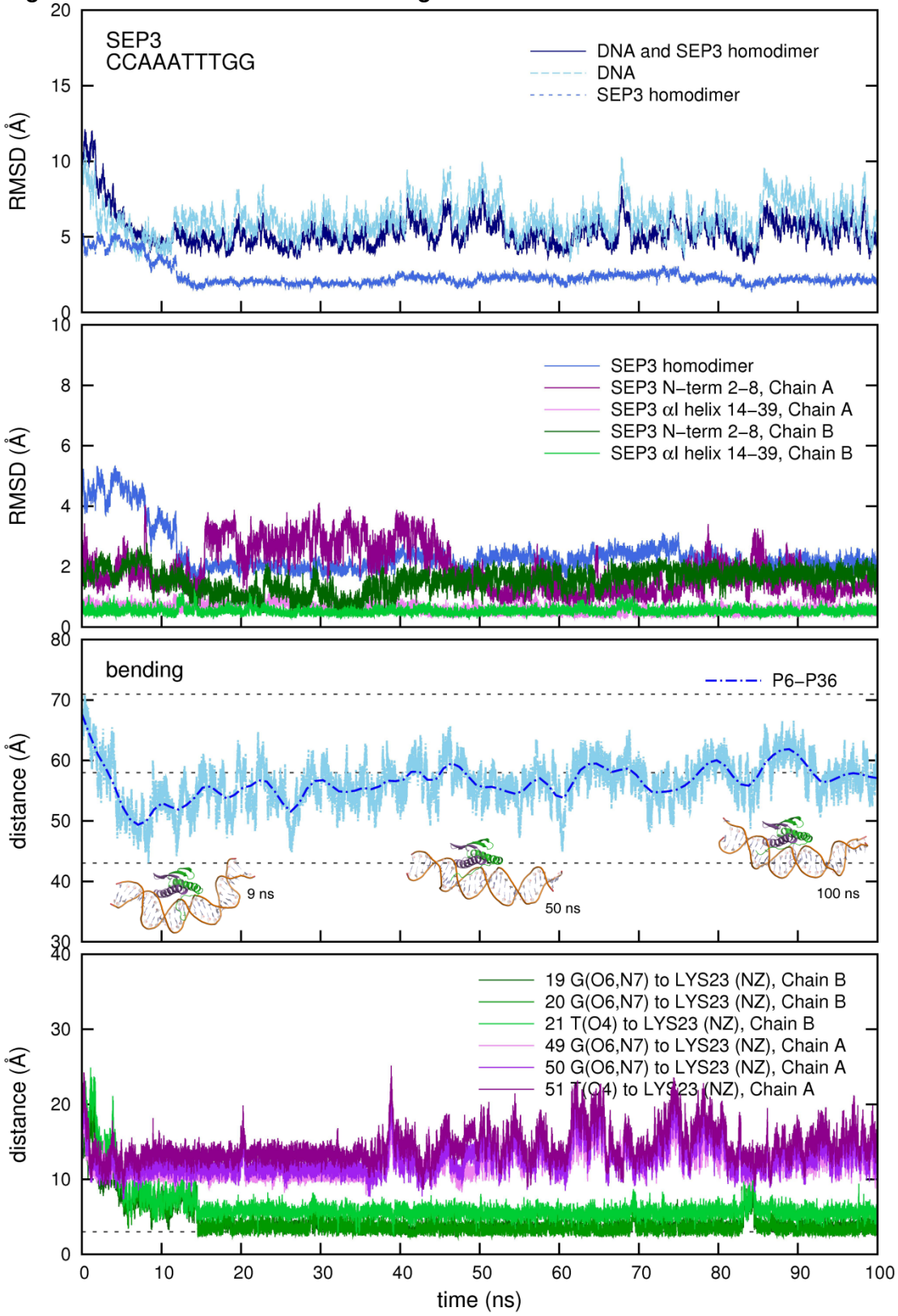


Figure 4.11 SEP3 homodimer binding of CTATATATAG.

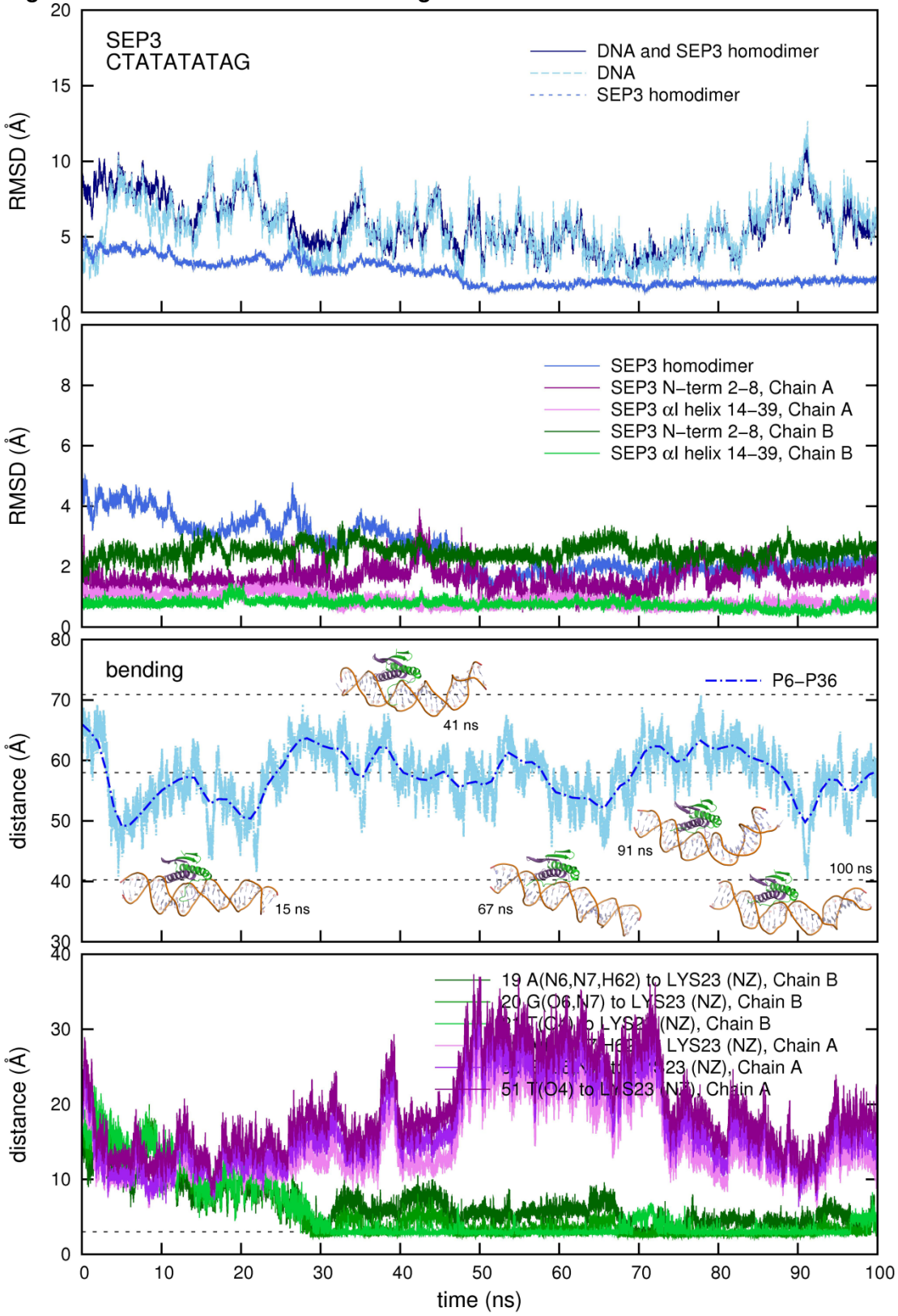
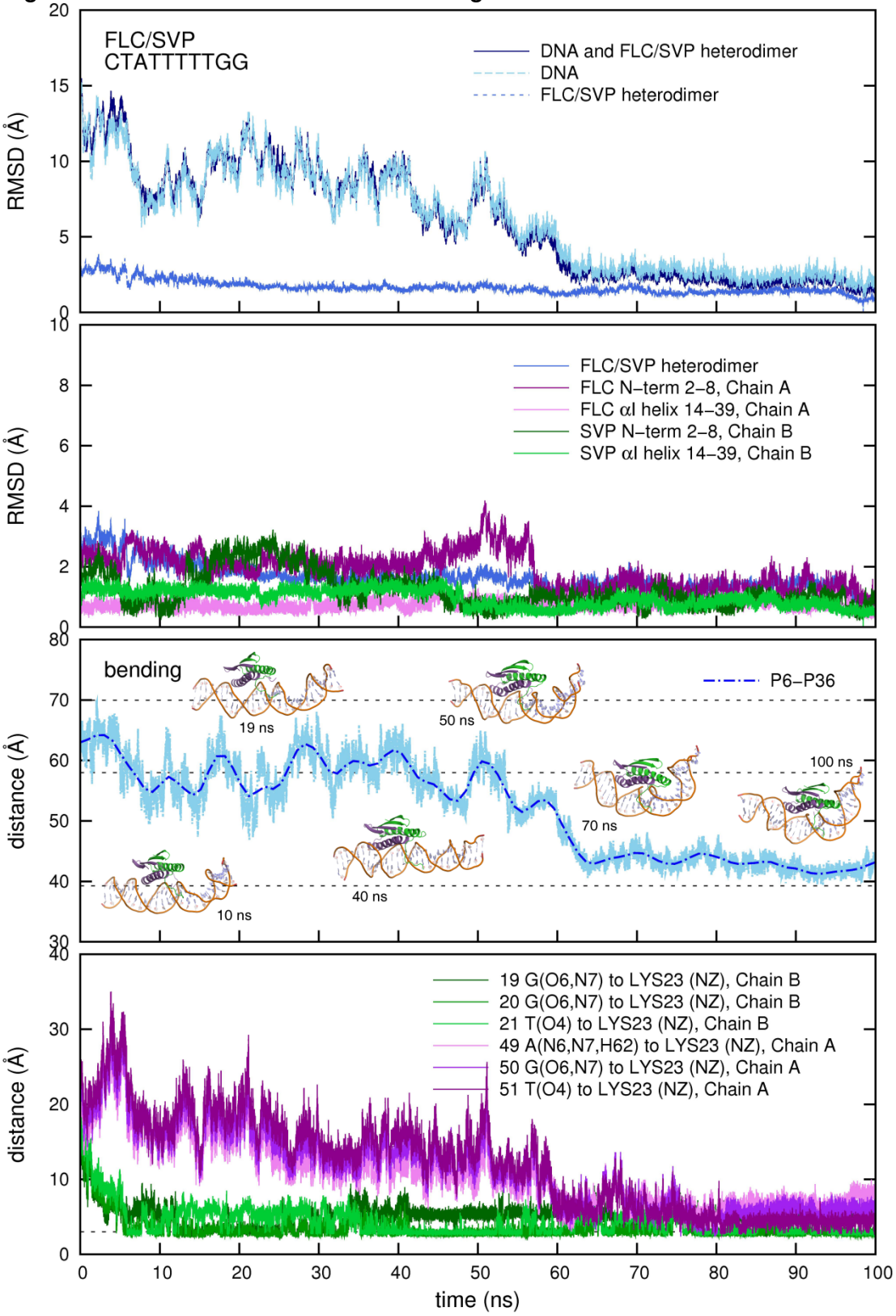


Figure 4.12 FLC–SVP heterodimer binding of CTATTTTTGG.



Comparison of binding for artificial FLC and SVP homodimers by DPI-ELISA

Recombinant, chimeric protein FLC and SVP was produced ($FLC^{1–60}$–$SEP3^{IK}$–8xHIS and $SVP^{1–60}$–$SEP3^{IK}$–8xHIS). Since FLC and SVP do not homodimerize, swapping the major dimerization domains (I and K) with those from SEP3 permitted the formation of artificial FLC and SVP homodimers. This allowed characterization of the binding specificities by DPI-ELISA of each MADS domain in an isolated manner.

$FLC^{1–60}$–$SEP3^{IK}$–8xHIS bound CTATATATAG (AGL15 HABS) and CTATTAATAG sequences at greater than 5-fold above the nonspecific probe, while $SVP^{1–60}$–$SEP3^{IK}$–8xHIS bound CTATATATAG (AGL15 HABS), CCAAATTTGG, CCATATATGG, and CCTAATTAGG sequences to about the same degree (Figure 4.13). These results are in agreement with reports concerning two characterized binding sites for FLC from the *SOC1* promoter and correspond to the following half-sites: CTATT, CTTAT, and CCAAA [257,295]. In general, it appears that the FLC MADS domain has a preference for the CT-W6-GG form CArG boxes and SVP has a preference for CC-W6-GG form CArG

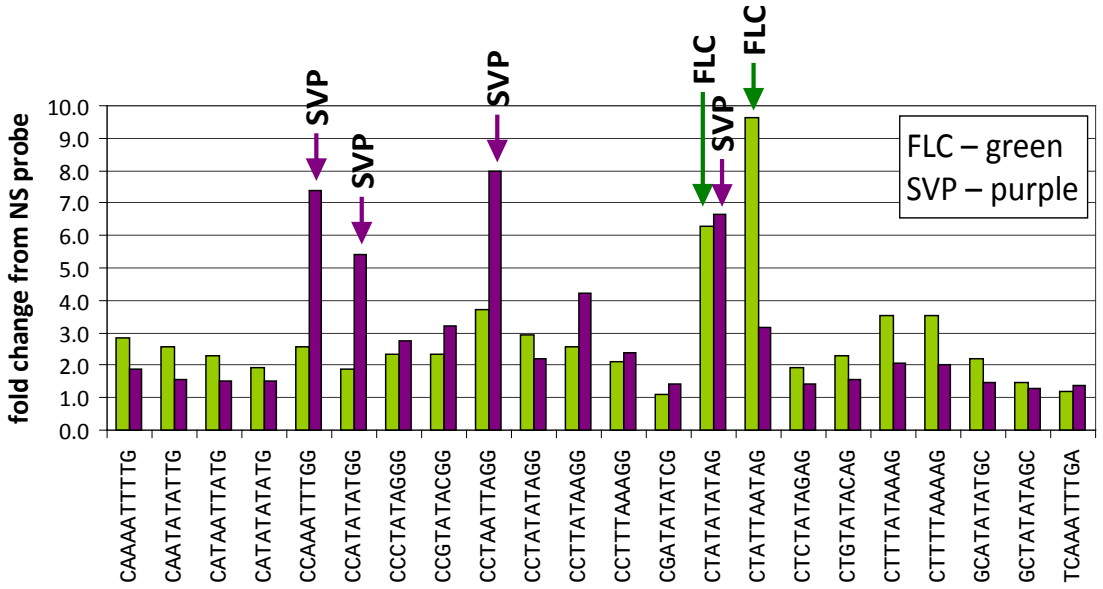


Figure 4.13 DPI-ELISA screening of DNA binding selectivity in artificial FLC and SVP homodimers. FLC was shown to be very selective towards CT-W6-AG form half-sites, while SVP showed a preference towards CC-W6-GG forms and the AGL15 HABS, a CT-W6-AG form. One characterized native DNA-binding site for the FLC-SVP heterodimer is a mixed form of CArG box (CTATTTTTGG), comprised of a CT-W6-AG form half-site (CTATT) and a CC-W6-GG form half-site (CCAAA). The DPI-ELISA suggests that FLC binds the CTATT half-site and SVP binds the CCAAA half-site.

boxes. The exception was with SVP and the AGL15 HABS. It is uncertain as to whether there properties unique to this sequence that permit it to be broadly recognized by MADS domains in general.

Discussion

Investigation of the FLC–SVP heterodimer and supporting simulation cases did indeed reveal roles and mechanisms for DNA recognition that are both unique to the FLC–SVP heterodimer, yet insightful to MADS domain DNA recognition in general.

Function/role of the αI Helix

The αI helix interacts and recognizes functional groups of the major groove. Only one αI helix from one monomer can initially interact within the major groove, which also corresponds to a half-site. Though the highly conserved LYS23 is a dominant interacting residue and the only residue from the αI helix to directly hydrogen bond to nucleobases of the major groove (Appendix), it is the surrounding residues that specify how the protein first interacts with DNA at the major groove of a potential binding site.

Residues in positions 26 and 30 (Figures 4.2 and 4.6) can adjust which helix will be the first to interact with the nucleobases within the major groove. It is probable that other nearby residues can perform this role; however, these are the only positions with an observed role. In the FLC–SVP–CTATTTTTGG case, SVP αI helix acted as a fulcrum, while FLC GLU30 appeared to provide some repulsion of FLC αI helix away from the DNA backbone and permitting SVP ARG26 to latch onto the DNA backbone (O1P 39 and O2P 39) via hydrogen bonding (Figure 4.5C,D). All adjacent ARG25 residues of the corresponding chain also participate in latching onto the DNA backbone similarly to SVP ARG26, but the ARG26 of SVP likely reinforces this latching in conjunction with FLC GLU30.

In the AGL15–CTATATATG case, which contains matching peptide chains, AGL15 LYS30, of either chain, would not provide a repulsive force. AGL15 SER26, Chain A formed a water-mediated hydrogen bond to the DNA backbone in the same position as SVP ARG26 (Figure 4.6). There is the possibility that phosphorylation of AGL15 SER26 could an additional mechanism of enhancing DNA-binding, making SER26 mimic SVP ARG26. Thus, residues of the αI helix aid in directing which αI helix is first to interact

with the major groove. Only after the direction is set, does the conserved LYS23 migrate to the nucleobases and hydrogen bond predominantly with guanidine (O6 20) (-5:5) (Figure 4.9, panel D).

Function/role of the N-terminal domain

While the αI helix interacts with major groove, the N-terminal domain interacts throughout the minor groove. In the AGL15–CTATATATG and FLC–SVP–CTATTTTTGG cases, the conserved ARG3 of the appropriate chain, hydrogen bonded to the O2 of thymidine at nucleobase 12 (position -4) (Appendix). This, however, is not specific since cytosine also possesses an O2 in the same location in the minor groove.

Of greater interest was the observation that the predominant positions for hydrogen bonding to nucleobases within the CArG box included 17, 18, 44, and 45 (Appendix). Nucleobases 17 and 44 represent position +2. When $SEP3^{MIK}$ was screened against a panel of palindromic probes by DPI-ELISA (Chapter 3, Figures 3.8 and 3.9), position -1:1 displayed a small preference for an AT over a TA center (Appendix). It appears that the asymmetrical preferences observed around position -2:2 can be attributed to the N-terminal domain. Additionally, AGL15 ARG3 was observed to interact with AGL15 GLU7 and SVP ARG3 was observed to interact with SVP GLU4. It is possible that these nearby residues modulate the availability of ARG3 to directly hydrogen bond to nucleobases of the minor groove, enhancing the effect seen here.

MADS N-terminal domain as a disordered tail

Disordered tails and regions have been implicated in a Monkey Bar-like DNA search mechanism [296,297], in which a disordered tail with an overall positive charge density is electrostatically attracted to the negatively charged DNA backbone. When two of these tails are present, one can remain attached while the other is free to search for nearby DNA backbone whether it is adjacent or intersegmental. Once secure, the first positively disordered tail is free to dissociate. The process is repeated allowing the protein to migrate amongst the DNA. Additionally, disordered tails and regions also have the potential to function as tuners of affinity in protein–DNA interactions [298].

The N-terminal domains of MADS domains have an overall positive charge density – consistent with the requirements of the Monkey Bar mechanism. AGAMOUS (AG),

MADS-AFFECTING FLOWERING 5 (MAF5), and SHATTERPROOF 1 and 2 (SHP1/2) possess an extended N-terminal domain. Therefore, it would not be surprising if these extended N-terminal domains modulated the affinity of the MADS domain and DNA or even the binding specificity.

A dynamic mechanism for DNA recognition by MIKC MADS domain TFs

Of the four simulations, only ARG3 of the N-terminal domain and LYS23 of the αI helix directly hydrogen bonded to the nucleobases of the DNA fragment (Appendix). This corresponds to results from previous studies. However, this adds to the question of how these MADS domains display sequence specificity. Here, it is proposed that a dynamic mechanism directs sequence specificity.

Since the N-terminal domain is the first subdomain of the MADS domain to direct sequence specificity as it is likely to be the first subdomain to interact with DNA based upon the Monkey Bar DNA search mechanism. Once an N-terminal domain interacts with a sequence with a low dissociation constant, the αI helices and remaining N-terminal domain interact nonspecifically with the DNA backbone. The αI helices and remaining N-terminal domain "test" the underlying sequence. If the sequence is incompatible, the dissociation constant will be high and the DNA–protein interaction will not stabilize. If the sequence is compatible, one αI helix will migrate into the major groove dictated by residues present at positions 26 and 30 of the peptide chains. As the underlying sequence is "tested" and the DNA–protein complex stabilizes, the complex undergoes an organizational shift with a further increase in stability and in the case of AGL15 and FLC–SVP, strong bending of the DNA. During the organizational shift, the conserved LYS23 of the αI helix, hydrogen bonds to positions within the major groove. This later configuration permits the greatest stability for the DNA–protein complex, but also hides what was the originally perceived DNA-binding sequence.

Biological implications

With the proposed mechanism for DNA recognition by MIKC MADS TFs, perturbations made early could have a significant impact on DNA specificity. Perturbations would include other DNA-binding proteins or adaptor proteins, as well as the influence of ions, though their influence is likely to be minimal in a consistent environment. These additional proteins would likely influence which αI helix is first to interact with the major

groove and thus orchestrate DNA recognition. This has been observed and characterized for animal and fungal MADS TFs [299], but only on a very limited basis for plant MADS TFs due to limited number of proteins identified that interact with plant MADS TFs. With HsMEF2A, DNA is only minimally bent [290,291], while HsSRF requires a TCF protein in order to bind non-optimal CArG boxes [143,144].

Another aspect which may possess regulatory functionality from binding to specificity is the N-terminal domain. In peach, for example, the FLC clade is absent, yet members of the StMADS (SVP) clade are present in multiple copies [300,301]. Though unlike Arabidopsis, peach SVP-like MADS TFs possess extended N-terminal domains. Therefore the question arises as to whether the extended N-terminal domain compensates for the absence of FLC or if there is another purpose for it.

MADS domains bind CT-W6-AG forms by default

Within Type I and II MADS TFs, there are both MADS domains that preferentially bind CC-W6-GG forms and others that preferentially bind CT-W6-AG forms. SEP3 and SVP bind both, though both still possess a preference for CC-W6-GG forms (Figures 3.8 and 4.13). Therefore, a question that arises is whether CTA-type MADS TFs can be converted to CCA-types and vice versa. This question was examined previously with MEF2A (RSRFC4) and SRF ($p67^{SRF}$) [302,303]. The METcore SRF MADS domain can bind both CC-W6-GG and CT-W6-AG forms, while MEF2A is only able to bind CT-W6-AG forms.

When chimeric proteins were generated with METcore SRF as the backbone, some variants lost the ability to bind CT-W6-AG forms [302]. Position 14, relative to MEF2A and most MIKC MADS TFs, in the METcore SRF was found to be susceptible. A K14E mutation in METcore SRF destroyed the ability of core SRF to bind CC-W6-GG forms. However, when attempts were made to isolate residues that would yield a MADS domain that would obligately bind CC-W6-GG forms, no such combination could be found [302,303]. With regard to the MIKC MADS TFs examined here, neither AGL15 nor FLC possess a glutamic acid at position 14 (Figure 4.2), yet both are to have preference for CT-W6-AG forms. Though not addressed, data presented here suggests that all MADS TFs bind CT-W6-AG forms by default and that binding may be expanded to include CC-W6-GG forms. Additionally, one concern that was brought up at the time

was that residues identified in shifting DNA-binding specificities did not appear to be involved in direct base pair recognition [302].

When examining the movements of the N-terminal domain from the MD simulations, it was observed that ARG3, a highly conserved residue, hydrogen bonded with nucleobases in relatively common positions along the CArG box (Appendix). For both AGL15 and FLC, ARG3 interacted with a range of nucleobases at positions 11–13 and 50. For SVP, the location of ARG3 is similar to that of SEP3, which interact with a range of nucleobases at positions 17–19, 44, and 45. In these two examples, the interaction is directed toward the -4:4 position. Additionally, the other chain of SEP3 interacted nucleobases 38 and 39.

The N-terminal domain is likely to support DNA searches via the Monkey Bar mechanism and without the N-terminal domain, initial DNA-binding would likely be negligible. Furthermore, the N-terminal domain appears to possess specificity for CT-W6-AG forms. Therefore, MADS TFs are most likely by default, binders of CT-W6-AG forms. In the K14E SRF mutant [302], the ability for SRF to interact with C:G pairs at positions -4:4 in CC-W6-GG forms was likely destroyed, leaving the N-terminal domain to dominate stabilization. This means that it is not probable to find a MADS domain that obligately binds CC-W6-GG forms.

Conclusion

The current knowledge gap surrounding MADS domain DNA recognition involves determination of what imperfections to the canonical CArG box sequence are permitted during DNA recognition by a MADS domain. Several highly conserved residues are present within the MADS domain and have had their roles identified DNA sequence recognition. However, these studies were done utilizing a limited and under-representative set of MADS domains (Appendix). Furthermore, by the inherent nature of X-ray crystallography, structural studies have been limited to a single static end-state that gave little information outside the highly conserved residues (Appendix). Molecular dynamics brought a means to examine the events leading up to the static end-state retrieved by X-ray crystallography. Additionally, examination of MADS domain clades with under-represented sequence characteristics provided insight to a broader set of biologically-relevant MADS domain sequences. What was found was that the N-terminal

plays a significant role in specifying bases within the central portions of the CArG box. Also, asymmetric DNA recognition by the αI helices and the N-terminal domains is a likely cause for the acceptance of imperfections to the canonical CArG box sequence. Later experiments may be able to ascertain patterns between imperfections permitted to the canonical CArG box and specific non-conserved residues in the MADS domain family.

Copyright © Joshua R. Werkman 2013

CHAPTER 5

Future Directions

The following are questions that arose from my research that would be interesting to pursue as a postdoc or an independent researcher.

Is there a pocket in the ACT domain of group III*f* bHLH TFs and do small molecules regulate dimer formation?

Two particular experiments that would be intriguing to conduct would be to model the ACT domain of R and attempt to determine the basis for dimerization of the domain. It is known that for some ACT domains, small molecules play a role in dimerization [19,37]. Modeling the ACT domain of R would allow one to determine if a pocket is present before or after dimerization as well as the size and any residues that could be used to stabilize a putative ligand metabolite. Knowing this information would be useful for site-directed mutagenesis studies both *in vitro* and *in vivo*, which may yield an undiscovered feed-back regulatory mechanism through allosteric conformational changes at the ACT domain. Since ACT-like domains appear to be present in the entire group III*f* bHLH TFs, this information would be valuable for not only to further understanding anthocyanin biosynthetic pathway, but also in trichome and root initiation and development. Ultimately if a pocket were identified in the ACT domain, identification of putative small molecule ligands would be critical to identification of a putative feed-back regulatory mechanism. Therefore the main goal of modeling the R ACT domain would be to perform computational small molecule ligand discovery similar to what is performed for drug discovery. This study would be conducted with software such as AutoDock, DOCK (UCSF), Hex, SwissDock, or other similarly available ligand docking software with a small molecule library interface.

Does a DNA loop form in the bHLH–MYB–DNA activation complex and is DNA binding dependent upon the correct phasing of *cis*-elements?

The second experiment attempts to answer questions concerning the overall arrangement and confirmation of the transcription factor activation complex. As mentioned before in the proposed model for R binding DNA, it appears that R is unique among the group III*f* bHLH TFs that regulate anthocyanin biosynthesis in how it interacts

with DNA. The others, however, are thought to bind G-box *cis*-elements in the respective promoters. In the promoters of *DFR* genes in *Nicotiana tabacum* and other Solanaceous species, such as *Petunia* sp., the presence of multiple G-boxes that can be aligned indicate the conservation of these G-boxes through evolution is likely a result of the maintenance of a similar regulatory mechanism. An experiment that would have been interesting to conduct would be to determine if the phase of these conserved *cis*-elements (G-boxes) in relation to each is critical for binding and activation.

DNA should be thought of as a tube rather than a string and the position of *cis*-elements in a promoter can be critical [281,282]. For example, if the distance between two *cis*-elements bound by a bHLH and MYB, respectively, is increased. The major groove of one *cis*-element will rotate to the opposite side of the DNA strand in relationship to the other. There is the likelihood that the bHLH–MYB complex will not be flexible enough to efficiently reach and bind both *cis*-elements. This scenario is more relevant to shorter distances between *cis*-elements due to the ability of larger DNA loops to overcome this. Thus, an experiment where the length between two *cis*-elements in a representative *DFR* promoter is increased and decreased to come in and out of phase would expect to observe a reporter signal to increase and decrease, respectively. This would provide insight to the overall arrangement and confirmation of the transcription factor activation complex and putative DNA loop. This could also provide insight into whether putative *cis*-elements would be blocked in ways similar to the *lac* promoter for *E. coli*. Furthermore, if phasing is critical to activation of this type of TF complex, this information would be useful for identifying biologically significant *cis*-elements for those promoters utilizing it.

Improved tag for DPI-ELISA

One regret of the current DPI-ELISA is the associated cost of the nickel-coated microtiter plates. The nickel–HIS-tag system was selected to minimize the introduction of new variables. However, this required the use of specially treated microtiter plates that were significantly more expensive than those used for a typical ELISA. There is a small push to reduce costs for ELISAs in general. One of these systems involves the use of a small, hydrophilic polystyrene tag to bind directly to the microtiter plate [304-309]. Some of the polystyrene tags identified through biopanning had affinities of adsorption that where 10 times higher than for wild-type GST [309]. A concern of the Ni^{2+}NTA HIS tag

system is the lower affinity of the HIS-tag towards Ni^{2+} and the use of a system, such as the polystyrene tag, could also alleviate this concern.

Map interactions and resolve positional redundancy between CArG boxes and plant MADS domain TFs using inosine and other non-standard bases

A set of probes similar to the inosine probes used in Chapter 3 used be designed and synthesized. These would be used to probe the binding of MADS domains to specific positions and functional groups using DPI-ELISA and possibly surface plasmon resonance (SPR). Representatives from each major clade and/or sequence subclade would be evaluated for binding specificity. Performing an assay in this manner will provide a consistent background and permit the data recovered from each clade to be more easily compared between clades. This should provide a good map for the understanding of the function of the unique sequence differences observed between major clades and/or sequence subclades. Additionally, using non-standard bases in particular combinations should be able resolve if and how redundancy is distributed within the CArG box.

Are the unique combinations of residues found in each MADS domain clade necessary and sufficient for the observed binding specificities?

Site-directed mutagenesis and swapping would be performed on select residues in MADS domain clades, such as FLC and SVP. Select positions would be centered on positions 7, 13–16, and 30–34 (see Figure 4.2). Additionally, mutations that would mimic a phosphorylated residue could be performed at positions 6, 14–16, 22, and 26. The rational for this approach is the presence of residues that mimic phosphorylation (arginine and lysine) are already present at position 26 in AGL24, SVP, AGL17, and AGL21, for example. Using SEP3 as a neutral or nonspecific MADS domain due its broad affinity for various CArG box sequences, mutations would also be made into SEP3 to attempt the conversion of SEP3 specificity to that observed in AGL15, FLC, or SVP. Course-grained modeling would be used to permit the testing of the behavior of site-directed mutagenesis candidates. It would allow for a reiterative approach and validation of both wet lab and simulation data.

What is the significance of extended N-terminal domains?

Extended N-terminal domains are found in MADS domains like AGAMOUS (AG), MADS-AFFECTING FLOWERING 5 (MAF5), and SHATTERPROOF 1 and 2 (SHP1/2). Additionally, extended N-terminal domains are also found in the SVP-like MADS TFs of peach, which does not possess any members of the FLC clade. Expanding the analysis of extended N-terminal domain function would include an extensive search and classification of MADS domains that include an extended N-terminal extension, but without any further N-terminal domains. The binding presences of representatives would be determined by DPI-ELISA. Studies of their binding characteristics and kinetics would be determined through truncations of the extended portion of the N-terminal domain, as well as swapping experiments. It would also be interesting to know if the extended N-terminal domain plays a role in DNA sequence selectivity in the flanking sequence of CArG boxes. Course-grained modeling would also be used to try to understand behave of the N-terminal domain at longer time frames.

Did the FLC clade evolve into presence after the StMADS11 clade? If so, how did the StMADS11 clade adapt?

Since the FLC clade has limited presence, it likely formed after the StMADS11 clade. Understanding these adaptations could help to better understand how SVP-like MADS TFs work in general. Another question concerning the presence of FLC-like clade members in distantly related plants is whether or not their common ancestor had a FLC-like clade member. If it did, then why was it lost? If it did not, then did the FLC-like MADS TFs from distantly related plants evolve by convergent evolution? If convergent evolution is at work, then this could indicate the binding specificity of the FLC clade is limited to a unique set of residues. Questions such as this have been raised previously, but answers still wait [116].

Can affinity and kinetics parameters be computationally modeled and then used to predict whether or not a MADS domain will bind a specific *cis*-element?

Using a combination of MD simulations and wet lab techniques such as EMSA, DPI-ELISA, SPR, circular dichroism, and/or isothermal titration calorimetry, a rank order would compiled. The goal would be two-fold. The first goal would be to attempt to discover what elements can be used to predict affinity and kinetics between specific MADS domains and specific DNA sequences. The second goal would occur

simultaneously and would be to reduce the complexity and multitude of factors involved into discrete elements useful for predicting affinity and binding kinetics. The importance of this second goal is to reduce the amount of resources required to solve the first goal. The tools developed here should be applicable to other discrete TFs and hopefully in a broader sense with multi-factor TF regulatory complexes.

Copyright © Joshua R. Werkman 2013

APPENDIX

Abbreviations:

4CL – 4-coumarate:CoA ligase

a1 – dihydroflavonol 4-reductase, *Zea mays*

a2 – anthocyanidin synthase, *Zea mays*

ACT – aspartate kinase, chorismate mutase and TyrA

AD – activating domain

AGL – agamous-like

AHLW, HLW, LW – Drop-out medium lacking adenine, histidine, leucine, and tryptophan

AMBER – Assisted Model Building with Energy Refinement

ANS – anthocyanidin synthase

app – apparent

BD – binding domain

β-gal – β-galactosidase

bHLH – basic helix loop helix

bHLH–LZ – basic helix loop helix leucine zipper

bHLH–LZL – basic helix loop helix leucine zipper-like

BiFC – bimolecular fluorescence complementation

BMW – bHLH–MYB–WDR

bp1 – BROWN PERICARP 1, *Zea mays*

BSA – bovine serum albumin

bz1 – UDP-glucose:flavonoid O-glucosyltranserase, *Zea mays*

bz2 – glutathione S-transferase, *Zea mays*

bZIP – basic leucine zipper

C1 – COLORLESS 1, *Zea mays*

c2 – chalcone synthase, *Zea mays*

C4H – cinnamate 4-hydroxylase

cDNA – complementary DNA

CHI – chalcone isomerase

ChIP – chromatin immunoprecipitation

ChIP–SEQ – chromatin immunoprecipitation with sequencing

CHS – chalcone synthase

COM – center of mass

CPU – central processing unit

DBD – DNA-binding domain

DBS – DNA-binding site

DFR – dihydroflavonol 4-reductase

DMSO – dimethyl sulfoxide

DPI-ELISA – DNA-protein interaction enzyme-linked immunosorbent assay

dsDNA – double-stranded DNA

DTT – dithiothreitol

EDTA – ethylenediaminetetraacetic acid

ELISA – enzyme-linked immunosorbent assay

EMSA – electrophoretic mobility shift assay

EMSY – named after Luke Hughes-Davies sister, because “SISTER” appears in the first line of the protein sequence

F3’H – flavonoid 3’-hydroxylase

F3H – flavanone 3-hydroxylase

FLC – FLOWERING LOCUS C, *Arabidopsis thaliana*

fs – femtosecond

GB – gigabyte

GPU – graphics processing unit

GST – glutathione S-transferase

HABS – high-affinity binding site

HDAC – histone deacetylase

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPC – high performance computing

HRP – horseradish peroxidase

IMAC – immobilized metal ion affinity chromatography

IPTG – isopropyl β-D-1-thiogalactopyranoside

K_d – dissociation constant

kDa – kiloDalton

Lc – LEAF COLOR, *Zea mays*

LZL – leucine zipper-like

MADS – MCM1, AGAMOUS, DEFICIENS, and SRF

MAF – MADS AFFECTING FLOWERING

MCM1 – MINICHROMOSOME MAINTENANCE 1, *Saccharomyces cerevisiae*

MD – molecular dynamics

MEF2 – MYOCYTE ENHANCER FACTOR 2, *Homo sapiens*

MHE – minimization, heating, and equilibration

MIKC – MADS, Keratin-like coiled-coil, Intervening, and C-terminal domains

$MIKC^C$ – MIKC classical

MIR – MYB interacting region

MYB – myeloblastosis oncogene family of transcription factors

NA-HRP – NeutrAvidin-horseradish peroxidase conjugate

NFκB – nuclear factor kappaB

NFDM – non-fat dry milk

NMR – nuclear magnetic resonance

ns – nanosecond

NTA – nitrilotriacetic acid

ONPG – O-nitrophenyl-β-D-galactopyranoside (ONPG)

P1 – PERICARP 1, *Zea mays*

PAC1 – PALE ALEURONE COLOR 1, *Zea mays*

PAGE – polyacrylamide gel electrophoresis

PAL – phenylalanine ammonia lyase

pm – picometer

pr1 – flavonoid 3'-hydroxylase, *Zea mays*

ps – picosecond

R – RED1 locus in maize

RIF1 – R-INTERACTING FACTOR 1, *Zea mays*

RMSD – root-mean-square deviation

SAM – SRF, ARG80, and MCM1

SDS-PAGE – sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SELEX – systematic evolution of ligands by exponential enrichment

SOC1 – SUPPRESSOR OF CONSTANS OVEREXPRESSION 1, *Arabidopsis thaliana*

SPR – surface plasmon resonance

SREBP1 – STEROL REGULATORY ELEMENT BINDING PROTEIN 1, *Homo sapiens*

SRF – SERUM RESPONSE FACTOR, *Homo sapiens*

StMADS – *Solanum tuberosum* MADS

SVP – SHORT VEGETATIVE PHASE, *Arabidopsis thaliana*

TEV – tobacco etch virus

TF – transcription factor

TF-EIA – transcription factor enzyme-linked immunoassay

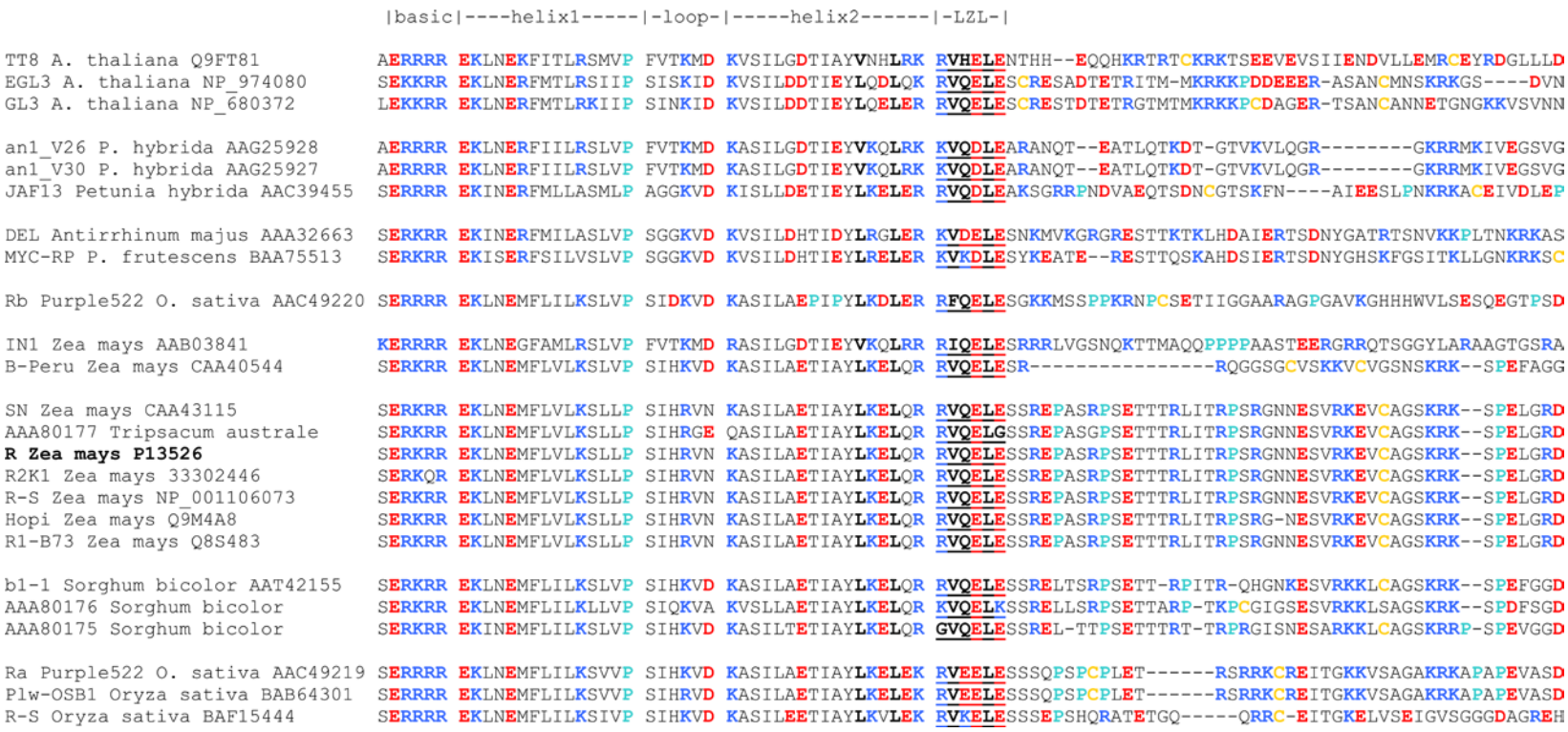
TIP3P – transferable intermolecular potential 3 point

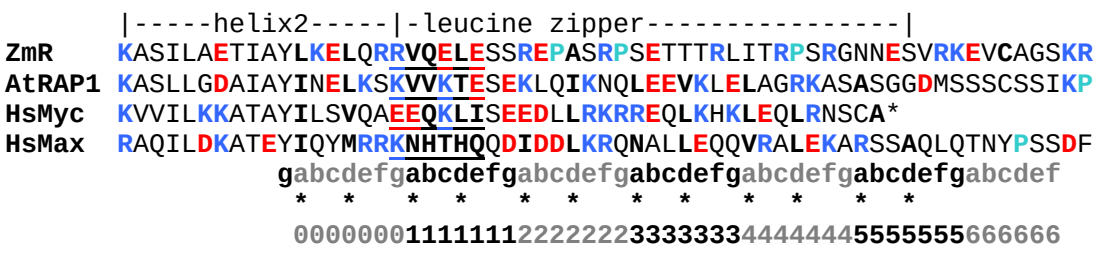
TMB – 3,3',5,5'-tetramethylbenzidine

TOPOIIA-A – topoisomerase IIA, subunit A

UFGT – UDP-glucose:flavonoid O-glucosyltranserase

WDR – WD-repeat (tryptophan-aspartate repeat); specifically WD40


|basic|----helix1-----|-loop-|-----helix2------|-LZL-|
TT8 A. thaliana Q9FT81 AERRRR EKLNEKFITLRSMVP FVTKMD KVSILGDTIAYVNHLRK RVHELENTHH--EQQHKRTRTCKRKTSEEVEVSIIENDVLLEMRCEYRDGLLLD
EGL3 A. thaliana NP_974080 SEKKRR EKLNERFMTLRSIIP SISKID KVSILDDTIEYLQDLQK RVQELESCRESADTETRITM-MKRKKPDDEEER-ASANCMNSKRKGS----DVN
GL3 A. thaliana NP_680372 LEKKRR EKLNERFMTLRKIIP SINKID KVSILDDTIEYLQELER RVQELESCRESTDTETRGTMTMKRKKPCDAGER-TSANCANNETGNGKKVSVNN
an1_V26 P. hybrida AAG25928 AERRRR EKLNERFIILRSLVP FVTKMD KASILGDTIEYVKQLRK KVQDLEARANQT--EATLQTKDT-GTVKVLQGR--------GKRRMKIVEGSVG
an1_V30 P. hybrida AAG25927 AERRRR EKLNERFIILRSLVP FVTKMD KASILGDTIEYVKQLRK KVQDLEARANQT--EATLQTKDT-GTVKVLQGR--------GKRRMKIVEGSVG
JAF13 Petunia hybrida AAC39455 SERRRR EKINERFMLLASMLP AGGKVD KISLLDETIEYLKELER RVQDLEAKSGRRPNDVAEQTSDNCGTSKFN----AIEESLPNKRKACEIVDLEP
DEL Antirrhinum majus AAA32663 SERKRR EKINERFMILASLVP SGGKVD KVSILDHTIDYLRGLER KVDELESNKMVKGRGRESTTKTKLHDAIERTSDNYGATRTSNVKKPLTNKRKAS
MYC-RP P. frutescens BAA75513 SERKRR EKISERFSILVSLVP SGGKVD KVSILDHTIEYLRELER KVKDLESYKEATE--RESTTQSKAHDSIERTSDNYGHSKFGSITKLLGNKRKSC
Rb Purple522 O. sativa AAC49220 SERRRR EKLNEMFLILKSLVP SIDKVD KASILAEPIPYLKDLER RFQELESGKKMSSPPKRNPCSETIIGGAARAGPGAVKGHHHWVLSESQEGTPSD
IN1 Zea mays AAB03841 KERRRR EKLNEGFAMLRSLVP FVTKMD RASILGDTIEYVKQLRR RIQELESRRRLVGSNQKTTMAQQPPPPAASTEERGRRQTSGGYLARAAGTGSRA
B-Peru Zea mays CAA40544 SERKRR EKLNEMFLVLKSLVP SIHKVD KASILAETIAYLKELQR RVQELESR----------------RQGGSGCVSKKVCVGSNSKRK--SPEFAGG
SN Zea mays CAA43115 SERKRR EKLNEMFLVLKSLLP SIHRVN KASILAETIAYLKELQR RVQELESSREPASRPSETTTRLITRPSRGNNESVRKEVCAGSKRK--SPELGRD
AAA80177 Tripsacum australe SERKRR EKLNEMFLVLKSLLP SIHRGE QASILAETIAYLKELQR RVQELGSSREPASGPSETTTRLITRPSRGNNESVRKEVCAGSKRK--SPELGRD
R Zea mays P13526 SERKRR EKLNEMFLVLKSLLP SIHRVN KASILAETIAYLKELQR RVQELESSREPASRPSETTTRLITRPSRGNNESVRKEVCAGSKRK--SPELGRD
R2K1 Zea mays 33302446 SERKQR EKLNEMFLVLKSLLP SIHRVN KASILAETIAYLKELQR RVQELESSREPASRPSETTTRLITRPSRGNNESVRKEVCAGSKRK--SPELGRD
R-S Zea mays NP_001106073 SERKRR EKLNEMFLVLKSLLP SIHRVN KASILAETIAYLKELQR RVQELESSREPASRPSETTTRLITRPSRGNNESVRKEVCAGSKRK--SPELGRD
Hopi Zea mays Q9M4A8 SERKRR EKLNEMFLVLKSLLP SIHRVN KASILAETIAYLKELQR RVQELESSREPASRPSETTTRLITRPSRG-NESVRKEVCAGSKRK--SPELGRD
R1-B73 Zea mays Q8S483 SERKRR EKLNEMFLVLKSLLP SIHRVN KASILAETIAYLKELQR RVQELESSREPASRPSETTTRLITRPSRGNNESVRKEVCAGSKRK--SPELGRD
b1-1 Sorghum bicolor AAT42155 SERKRR EKLNEMFLILKSLVP SIHKVD KASILAETIAYLKELQR RVQELESSRELTSRPSETT-RPITR-QHGNKESVRKKLCAGSKRK--SPEFGGD
AAA80176 Sorghum bicolor SERKRR EKLNEMFLILKLLVP SIQKVA KVSLLAETIAYLKELQR KVQELKSSRELLSRPSETTARP-TKPCGIGSESVRKKLSAGSKRK--SPDFSGD
AAA80175 Sorghum bicolor SERKRR EKINEMFLILKSLVP SIHKVD KASILTETIAYLKELQR GVQELESSREL-TTPSETTTRT-TRPRGISNESARKKLCAGSKRRP-SPEVGGD
Ra Purple522 O. sativa AAC49219 SERRRR EKLNEMFLILKSVVP SIHKVD KASILAETIAYLKELEK RVEELESSSQPSPCPLET------RSRRKCREITGKKVSAGAKRKAPAPEVASD
Plw-OSB1 Oryza sativa BAB64301 SERRRR EKLNEMFLILKSVVP SIHRVD KASILAETIAYLKELEK RVEELESSSQPSPCPLET------RSRRKCREITGKKVSAGAKRKAPAPEVASD
R-S Oryza sativa BAF15444 SERRRR EKLNEMFLILKSIVP SIHKVD KASILEETIAYLKVLEK RVKELESSSEPSHQRATETGQ-----QRRC-EITGKELVSEIGVSGGGDAGREH


|-----helix2-----|-leucine zipper----------------|
ZmR KASILAETIAYLKELQRRVQELESSREPASRPSETTTRLITRPSRGNNESVRKEVCAGSKR
AtRAP1 KASLLGDAIAYINELKSKVVKTESEKLQIKNQLEEVKLELAGRKASASGGDMSSSCSSIKP
HsMyc KVVILKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLRNSCA*
HsMax RAQILDKATEYIQYMRRKNHTHQQDIDDLKRQNALLEQQVRALEKARSSAQLQTNYPSSDF
gabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdef
0000000111111122222223333333444444455555556666666

-5	-4	-3	-2	-1	+1	+2	+3	+4	+5
C	**T**	**A**	**T**	**A**	**T**	**A**	**T**	**A**	**G**
A	T	A	T	A	T	A	T	A	T
G	T	A	T	A	T	A	T	A	C
T	T	A	T	A	T	A	T	A	A
C	A	A	T	A	T	A	T	T	G
C	C	A	T	A	T	A	T	G	G
C	G	A	T	A	T	A	T	C	G
C	T	C	T	A	T	A	G	A	G
C	T	G	T	A	T	A	C	A	G
C	T	T	T	A	T	A	A	A	G
C	T	A	A	A	T	T	T	A	G
C	T	A	C	A	T	G	T	A	G
C	T	A	G	A	T	C	T	A	G
C	T	A	T	C	G	A	T	A	G
C	T	A	T	G	C	A	T	A	G
C	T	A	T	T	A	A	T	A	G

-5	-4	-3	-2	-1	+1	+2	+3	+4	+5
C	**C**	**A**	**T**	**A**	**T**	**A**	**T**	**G**	**G**
A	C	A	T	A	T	A	T	G	T
G	C	A	T	A	T	A	T	G	C
T	C	A	T	A	T	A	T	G	A
C	A	A	T	A	T	A	T	T	G
C	G	A	T	A	T	A	T	C	G
C	T	A	T	A	T	A	T	A	G
C	C	C	T	A	T	A	G	G	G
C	C	G	T	A	T	A	C	G	G
C	C	T	T	A	T	A	A	G	G
C	C	A	A	A	T	T	T	G	G
C	C	A	C	A	T	G	T	G	G
C	C	A	G	A	T	C	T	G	G
C	C	A	T	C	G	A	T	G	G
C	C	A	T	G	C	A	T	G	G
C	C	A	T	T	A	A	T	G	G

-5	-4	-3	-2	-1	+1	+2	+3	+4	+5
C	**A**	**T**	**A**	**T**	**A**	**T**	**A**	**T**	**G**
A	A	T	A	T	A	T	A	T	T
G	A	T	A	T	A	T	A	T	C
T	A	T	A	T	A	T	A	T	A
C	C	T	A	T	A	T	A	G	G
C	G	T	A	T	A	T	A	C	G
C	T	T	A	T	A	T	A	A	G
C	A	A	A	T	A	T	T	T	G
C	A	C	A	T	A	T	G	T	G
C	A	G	A	T	A	T	C	T	G
C	A	T	C	T	A	G	A	T	G
C	A	T	G	T	A	C	A	T	G
C	A	T	T	T	A	A	A	T	G
C	A	T	A	A	T	T	A	T	G
C	A	T	A	C	G	T	A	T	G
C	A	T	A	G	C	T	A	T	G

-5	-4	-3	-2	-1	+1	+2	+3	+4	+5
C	**C**	**T**	**A**	**T**	**A**	**T**	**A**	**G**	**G**
A	C	T	A	T	A	T	A	G	T
G	C	T	A	T	A	T	A	G	C
T	C	T	A	T	A	T	A	G	A
C	A	T	A	T	A	T	A	T	G
C	G	T	A	T	A	T	A	C	G
C	T	T	A	T	A	T	A	A	G
C	C	A	A	T	A	T	T	G	G
C	C	C	A	T	A	T	G	G	G
C	C	G	A	T	A	T	C	G	G
C	C	T	C	T	A	G	A	G	G
C	C	T	G	T	A	C	A	G	G
C	C	T	T	T	A	A	A	G	G
C	C	T	A	A	T	T	A	G	G
C	C	T	A	C	G	T	A	G	G
C	C	T	A	G	C	T	A	G	G

cis-element	protein	ACCEPTOR chain	residue	peptide residue number	pdb residue number	atom	atom (H)	DONOR base/ residue	position	pdb residue number	atom	% occupied	percent occupancy over course of the trajectory file (each character or space represents 2 ns of time)
CTATATATAG	AGL15	A	ARG	3	62	NH2	HH21	T	-4	12	O4'	16.63	|.-o .-o-.-...----.-...--. . .-. ... . .-.... .-|
CTATATATAG	AGL15	A	ARG	3	62	NH2	HH21	T	-4	12	O2	97.90	|@@@@@@@@@@@@@@@@*@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@|
CTATATATAG	AGL15	A	ARG	3	62	NE	HE	T	-4	12	O2	96.47	|@@@@@*@@@@@@@@@@@@@@@@@@@@**@**@*@*@@@*@**@@@@@**@@|
CTATATATAG	AGL15	A	ARG	3	62	NE	HE	A	-3	13	O4'	6.87	|. . . ..- - - . . .. |
CTATATATAG	AGL15	B	ARG	3	121	NH1	HH12	T	-8	8	O1P	7.50	| - .*xxxx|
CTATATATAG	AGL15	B	ARG	3	121	NH2	HH22	T	-8	8	O1P	5.93	| -o .o-xox|
CTATATATAG	AGL15	B	ARG	3	121	NH1	HH12	T	-7	9	O1P	5.67	| -....o - o. .o |
CTATATATAG	AGL15	B	ARG	3	121	NH1	HH12	T	-7	9	O2P	17.33	| *@x*-.@xo. . o. .. oo. |
CTATATATAG	AGL15	B	ARG	3	121	NH2	HH22	T	-7	9	O2P	15.30	| .x*xx..*xx- o. .o |
CTATATATAG	AGL15	B	ARG	3	121	NH1	HH12	GLU	7	125	OE1	9.30	| ..ooox*o- |
CTATATATAG	AGL15	B	ARG	3	121	NH2	HH22	GLU	7	125	OE1	17.30	| -@*@@**x@@. |
CTATATATAG	AGL15	B	ARG	3	121	NH1	HH12	GLU	7	125	OE2	17.30	| -@*@@@*o@@. |
CTATATATAG	AGL15	B	ARG	3	121	NH2	HH22	GLU	7	125	OE2	15.10	| .*xxo*x@x*- |

Dumping schematic of time series after each h-bond, key follows:

```
   |          .       -       o       x       *       @    |
      0-5%   5-20%  20-40%  40-60% 60-80% 80-95% 95-100% occupancy
```

"|" (or pipe) represents the beginning or end of the time course

cis-element	protein	ACCEPTOR chain	residue	peptide residue number	pdb residue number	atom	atom (H)	DONOR base/ residue	position	pdb residue number	atom	% occupied	percent occupancy over course of the trajectory file (each character or space represents 2 ns of time)
CCAAATTTGG	SEP3	A	ARG	3	62	NH1	HH12	T	+8	38	O2	35.43	| .@@@x -xx@xoxooxxooxo-o*x.** |
CCAAATTTGG	SEP3	A	ARG	3	62	NH2	HH22	T	+8	38	O2	18.93	|. .ooo. .--o.----oo-.o-.-oo -o |
CCAAATTTGG	SEP3	A	ARG	3	62	NH2	HH21	T	+7	39	O2	25.00	| . -**@**x**@@**@@|
CCAAATTTGG	SEP3	B	ARG	3	121	NH1	HH12	T	+2	17	O2	39.73	|@xoxxoox@*xooo-xxxxxxxx-xooooo -x. |
CCAAATTTGG	SEP3	B	ARG	3	121	NH2	HH22	T	+2	17	O2	14.97	|-oo--.-...-..-.--oo--.. -... . .-. |
CCAAATTTGG	SEP3	B	ARG	3	121	NH2	HH22	A	+2	44	N3	5.67	| ***. |
CCAAATTTGG	SEP3	B	ARG	3	121	NH1	HH12	A	+1	45	N3	26.50	|--o-oxx- .-xx-x-o-.ooo-*ooooxo |
CCAAATTTGG	SEP3	B	ARG	3	121	NH2	HH22	A	+1	45	N3	16.37	| -o*-o- o-- o. --oo-.o.o- . . |

cis-element	protein	ACCEPTOR chain	residue	peptide residue number	pdb residue number	atom	atom (H)	DONOR base/ residue	position	pdb residue number	atom	% occupied	percent occupancy over course of the trajectory file (each character or space represents 2 ns of time)
CTATATATAG	SEP3	A	ARG	3	62	NH2	HH21	T	+8	38	O2	10.67	|x--..--oo-..- . --. |
CTATATATAG	SEP3	A	ARG	3	62	NH2	HH22	T	+8	38	O2	6.33	| . . -..-. -- ...-- .. . |
CTATATATAG	SEP3	B	ARG	3	121	NH1	HH11	T	+3	18	O2	95.93	|@@*@@@@@@@@@@@@@@@@@@@@@@@@****@@@@@@*****@*x@@*@*@|
CTATATATAG	SEP3	B	ARG	3	121	NH1	HH12	A	+4	19	N3	56.27	|xooxxox--oxox*xoxxxoxx.oo-.ooooxxx-oxxxxoo- x-xxo|
CTATATATAG	SEP3	B	ARG	3	121	NH1	HH11	T	+2	44	O2	9.37	| . . . . . ---.---o-.-.--|

Dumping schematic of time series after each h-bond, key follows:

	.	-	o	x	*	@	
0-5%	5-20%	20-40%	40-60%	60-80%	80-95%	95-100%	occupancy

"|" (or pipe) represents the beginning or end of the time course

cis-element	protein	ACCEPTOR						DONOR				% occupied	percent occupancy over course of the trajectory file (each character or space represents 2 ns of time)
		chain	residue	peptide residue number	pdb residue number	atom	atom (H)	base/ residue	position	pdb residue number	atom		
CTATTTTTGG	FLC	A	ARG	3	62	NH2	HH22	C	-5	11	O2	7.07	| .-......- ..... .- -.. - . |
CTATTTTTGG	FLC	A	ARG	3	62	NH2	HH22	T	-4	12	O4'	45.83	| oo-o-oxxoooxoxoooooox-ooxooooxxxoxxoooxxo|
CTATTTTTGG	FLC	A	ARG	3	62	NH1	HH12	T	-4	12	O2	89.60	| -@@*@@@@@@@@@@@@@@@@@@@@@x@@@@@@@@@@@@@@@@@@@@|
CTATTTTTGG	FLC	A	ARG	3	62	NH2	HH22	T	-4	12	O2	74.97	| .o--ox***************x@@**@****@***x******@*xx|
CTATTTTTGG	FLC	A	ARG	3	62	NH2	HH22	G	-5	50	N3	10.57	|@@x*..*x |
CTATTTTTGG	FLC	A	ARG	3	62	NH1	HH12	G	-5	50	N3	9.00	|@x@*o - |
CTATTTTTGG	SVP	B	ARG	3	121	NH1	HH11	T	+2	17	O2	74.53	|*x*xxxx*-o-xoxooooooo@**@@**@***xxx**x*****@*x*xx*|
CTATTTTTGG	SVP	B	ARG	3	121	NH1	HH12	T	+3	18	O2	11.37	|. ... . -- . ..- .--. . ---.... .. .-.|
CTATTTTTGG	SVP	B	ARG	3	121	NH1	HH11	A	+1	45	N3	49.83	|.o--o--o*x*oxx*xx*x*x--o--------oxxoooo-oxo--xoo-.|
CTATTTTTGG	SVP	B	ARG	3	121	NE	HE	GLU	4	122	OE1	34.20	| .o-..-o@@@@@@x ooo@ o. -@@@o@- |
CTATTTTTGG	SVP	B	ARG	3	121	NH2	HH21	GLU	4	122	OE1	43.57	| . .o@@@o*@*x*x**x .ooo@ - o@- .-@@@o@- |
CTATTTTTGG	SVP	B	ARG	3	121	NE	HE	GLU	4	122	OE2	37.07	|**@**@@*@@@@@xo-.o--. - o@- .- - |
CTATTTTTGG	SVP	B	ARG	3	121	NH2	HH21	GLU	4	122	OE2	45.43	|@@@*@@***x****@@@@x@x - o. oo- .-. o |

Dumping schematic of time series after each h-bond, key follows:

```
   |          .       -       o       x       *       @    |
      0-5%   5-20%  20-40%  40-60% 60-80% 80-95% 95-100% occupancy
```

"|" (or pipe) represents the beginning or end of the time course

AGL15

										-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
DC5	C	G	G	G	T	T	T	T	A	C	T	A	T	A	T	A	T	A	G	T	A	A	A	A	G	G	G	C	DC3
DG3	G	C	C	C	A	A	A	A	T	G	A	T	A	T	A	T	A	T	C	A	T	T	T	T	C	C	C	G	DG5
60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
										-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7						

SEP3

										-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
DC5	C	G	G	G	T	T	T	T	A	C	C	A	A	A	T	T	T	G	G	T	A	A	A	A	G	G	G	C	DC3
DG3	G	C	C	C	A	A	A	A	T	G	G	T	T	T	A	A	A	C	C	A	T	T	T	T	C	C	C	G	DG5
60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
										-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7						

SEP3

										-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
DC5	C	G	G	G	T	T	T	T	A	C	T	A	T	A	T	A	T	A	G	T	A	A	A	A	G	G	G	C	DC3
DG3	G	C	C	C	A	A	A	A	T	G	A	T	A	T	A	T	A	T	C	A	T	T	T	T	C	C	C	G	DG5
60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
										-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7						

FLC-SVP

										-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
DC5	C	G	G	G	T	T	T	T	A	C	T	A	T	T	T	T	T	G	G	T	A	A	A	A	G	G	G	C	DC3
DG3	G	C	C	C	A	A	A	A	T	G	A	T	A	A	A	A	A	C	C	A	T	T	T	T	C	C	C	G	DG5
60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
										-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7						

Chain A

Chain B

Structure ID	MADS	DNA	Other molecules	Type	Organism	Ref.
1hbx	HsSRF	26mer	Hs ETS-domain protein ELK-4	I	human	[310]
1k6o	HsSRF	23mer	Hs ETS-domain protein ELK-4	I	human	[145]
1srs	HsSRF	19mer	none	I	human	[311]
1mnm	ScMCM1	26mer	ScMATalpha	I	yeast	[289]
1c7u	HsMEF2A	20mer	none	II	human	[286]
1egw	HsMEF2A	17mer	none	II	human	[312]
3kov	HsMEF2A	13mer	none	II	human	[279]
3mu6	HsMEF2A	17mer	(E)-N'-(2-aminophenyl)-N-phenyl-oct-3-enediamide	II	human	[313]
3p57	HsMEF2A	15mer	Histone acetyltransferase p300 TAZ Hs Zn 2+	II	human	[137]
1n6j	HsMEF2B	14mer	Hs Calcineurin-binding protein Cabin 1	II	human	[314]
1tqe	HsMEF2B	17mer	Mus musculus Histone deacetylase 9	II	human	[299]

REFERENCES

1. ftp://ftp.arabidopsis.org/: 2010.
2. Bray RA: **A plant color factor linked to the *R* locus**. *Maize Genet. Coop. News Letter* 1964, **38**:134.
3. Dooner HK, Kermicle JL: **Displaced and tandem duplications in the long arm of chromosome 10 in maize**. *Genetics* 1976, **82**:309-322.
4. Dooner HK: **Identification of an R-locus region that controls the tissue specificity of anthocyanin formation in maize**. *Genetics* 1979, **93**:703-710.
5. Ludwig SR, Habera LF, Dellaporta SL, Wessler SR: **Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region**. *Proc Natl Acad Sci U S A* 1989, **86**:7092-7096.
6. Urao T, Yamaguchi-Shinozaki K, Mitsukawa N, Shibata D, Shinozaki K: **Molecular cloning and characterization of a gene that encodes a MYC-related protein in Arabidopsis**. *Plant Mol Biol* 1996, **32**:571-576.
7. Zhao H, Wang X, Zhu D, Cui S, Li X, Cao Y, Ma L: **A single amino acid substitution in IIIf subfamily of basic helix-loop-helix transcription factor AtMYC1 leads to trichome and root hair patterning defects by abolishing its interaction with partner proteins in Arabidopsis**. *J Biol Chem* 2012, **287**:14109-14121.
8. Symonds VV, Hatlestad G, Lloyd AM: **Natural allelic variation defines a role for ATMYC1: trichome cell fate determination**. *PLoS Genet* 2011, **7**:e1002069.
9. Ramsay NA, Walker AR, Mooney M, Gray JC: **Two basic-helix-loop-helix genes (MYC-146 and GL3) from Arabidopsis can activate anthocyanin biosynthesis in a white-flowered Matthiola incana mutant**. *Plant Mol Biol* 2003, **52**:679-688.
10. Payne CT, Zhang F, Lloyd AM: **GL3 encodes a bHLH protein that regulates trichome development in arabidopsis through interaction with GL1 and TTG1**. *Genetics* 2000, **156**:1349-1362.
11. Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A: **A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis**. *Development* 2003, **130**:4859-4869.
12. Feyissa DN, Lovdal T, Olsen KM, Slimestad R, Lillo C: **The endogenous GL3, but not EGL3, gene is necessary for anthocyanin accumulation as induced by nitrogen depletion in Arabidopsis rosette stage leaves**. *Planta* 2009, **230**:747-754.
13. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L: **TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana**. *Plant J* 2004, **39**:366-380.
14. Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L: **The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques**. *Plant Cell* 2000, **12**:1863-1878.
15. Hernandez JM, Heine GF, Irani NG, Feller A, Kim MG, Matulnik T, Chandler VL, Grotewold E: **Different mechanisms participate in the R-dependent activity of the R2R3 MYB transcription factor C1**. *J Biol Chem* 2004, **279**:48205-48213.
16. Goff SA, Cone KC, Chandler VL: **Functional analysis of the transcriptional activator encoded by the maize B gene: evidence for a direct functional**

interaction between two classes of regulatory proteins. *Genes Dev* 1992, **6**:864-875.
17. Goff SA, Klein TM, Roth BA, Fromm ME, Cone KC, Radicella JP, Chandler VL: **Transactivation of anthocyanin biosynthetic genes following transfer of B regulatory genes into maize tissues**. *EMBO J* 1990, **9**:2517-2522.
18. Smolen GA, Pawlowski L, Wilensky SE, Bender J: **Dominant alleles of the basic helix-loop-helix transcription factor ATR2 activate stress-responsive genes in Arabidopsis**. *Genetics* 2002, **161**:1235-1246.
19. Feller A, Hernandez JM, Grotewold E: **An ACT-like domain participates in the dimerization of several plant basic-helix-loop-helix transcription factors**. *J Biol Chem* 2006, **281**:28964-28974.
20. Grotewold E, Sainz MB, Tagliani L, Hernandez JM, Bowen B, Chandler VL: **Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R**. *Proc Natl Acad Sci U S A* 2000, **97**:13579-13584.
21. Grotewold E, Drummond BJ, Bowen B, Peterson T: **The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset**. *Cell* 1994, **76**:543-553.
22. Sainz MB, Grotewold E, Chandler VL: **Evidence for direct activation of an anthocyanin promoter by the maize C1 protein and comparison of DNA binding by related Myb domain proteins**. *Plant Cell* 1997, **9**:611-625.
23. Dias AP, Braun EL, McMullen MD, Grotewold E: **Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication**. *Plant Physiol* 2003, **131**:610-620.
24. Shen Z, Sathyan KM, Geng Y, Zheng R, Chakraborty A, Freeman B, Wang F, Prasanth KV, Prasanth SG: **A WD-repeat protein stabilizes ORC binding to chromatin**. *Mol Cell* 2010, **40**:99-111.
25. Valeyev NV, Downing AK, Sondek J, Deane C: **Electrostatic and functional analysis of the seven-bladed WD beta-propellers**. *Evol Bioinform Online* 2008, **4**:203-216.
26. van Nocker S, Ludwig P: **The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function**. *BMC Genomics* 2003, **4**:50.
27. Hoecker U, Quail PH: **The phytochrome A-specific signaling intermediate SPA1 interacts directly with COP1, a constitutive repressor of light signaling in Arabidopsis**. *J Biol Chem* 2001, **276**:38173-38178.
28. Smith TF, Gaitatzes C, Saxena K, Neer EJ: **The WD repeat: a common architecture for diverse functions**. *Trends Biochem Sci* 1999, **24**:181-185.
29. de Vetten N, Quattrocchio F, Mol J, Koes R: **The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals**. *Genes Dev* 1997, **11**:1422-1434.
30. Craig EA: **Eukaryotic chaperonins: lubricating the folding of WD-repeat proteins**. *Curr Biol* 2003, **13**:R904-905.
31. Appelhagen I, Lu GH, Huep G, Schmelzer E, Weisshaar B, Sagasser M: **TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of Arabidopsis thaliana seeds**. *Plant J* 2011, **67**:406-419.
32. Maes L, Inze D, Goossens A: **Functional specialization of the TRANSPARENT TESTA GLABRA1 network allows differential hormonal control of laminal**

and marginal trichome initiation in Arabidopsis rosette leaves. *Plant Physiol* 2008, **148**:1453-1464.
33. Bouyer D, Geier F, Kragler F, Schnittger A, Pesch M, Wester K, Balkunde R, Timmer J, Fleck C, Hulskamp M: **Two-dimensional patterning by a trapping/depletion mechanism: the role of TTG1 and GL3 in Arabidopsis trichome formation**. *PLoS Biol* 2008, **6**:e141.
34. Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC: **The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein**. *Plant Cell* 1999, **11**:1337-1350.
35. Carey CC, Strahle JT, Selinger DA, Chandler VL: **Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in Arabidopsis thaliana**. *Plant Cell* 2004, **16**:450-464.
36. Hernandez JM, Feller A, Morohashi K, Frame K, Grotewold E: **The basic helix loop helix domain of maize R links transcriptional regulation and histone modifications by recruitment of an EMSY-related factor**. *Proc Natl Acad Sci U S A* 2007, **104**:17222-17227.
37. Kong Q, Pattanaik S, Feller A, Werkman JR, Chai C, Wang Y, Grotewold E, Yuan L: **Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R**. *Proc Natl Acad Sci U S A* 2012, **109**:E2091-2097.
38. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP: **The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains**. *Trends Biochem Sci* 2003, **28**:69-74.
39. Hughes-Davies L, Huntsman D, Ruas M, Fuks F, Bye J, Chin SF, Milner J, Brown LA, Hsu F, Gilks B, et al.: **EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer**. *Cell* 2003, **115**:523-535.
40. Ecevit O, Khan MA, Goss DJ: **Kinetic analysis of the interaction of b/HLH/Z transcription factors Myc, Max, and Mad with cognate DNA**. *Biochemistry* 2010, **49**:2627-2635.
41. Crick FH: **Is alpha-keratin a coiled coil?** *Nature* 1952, **170**:882-883.
42. Bowman JL, Smyth DR, Meyerowitz EM: **Genes directing flower development in Arabidopsis**. *Plant Cell* 1989, **1**:37-52.
43. Dalrymple GB: **The age of the Earth in the twentieth century: a problem (mostly) solved**. *Geological Society, London, Special Publications* 2001, **190**:205-221.
44. Trenberth KE, Jones PD, Ambenje P, Bojariu R, Easterling D, Klein Tank A, Parker D, Rahimzadeh F, Renwick JA, Rusticucci M, et al.: **Observations: Surface and Atmospheric Climate Change**. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by Solomon SD, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL: Cambridge University Press; 2007.
45. Mueller B, Seneviratne SI: **Hot days induced by precipitation deficits at the global scale**. *Proc Natl Acad Sci U S A* 2012, **109**:12398-12403.
46. Menzel A, Seifert H, Estrella N: **Effects of recent warm and cold spells on European plant phenology**. *Int J Biometeorol* 2011, **55**:921-932.

47. Wolkovich EM, Cook BI, Allen JM, Crimmins TM, Betancourt JL, Travers SE, Pau S, Regetz J, Davies TJ, Kraft NJ, et al.: **Warming experiments underpredict plant phenological responses to climate change**. *Nature* 2012, **485**:494-497.
48. Cook BI, Wolkovich EM, Parmesan C: **Divergent responses to spring and winter warming drive community level flowering trends**. *Proc Natl Acad Sci U S A* 2012, **109**:9000-9005.
49. Beaubien E, Hamann A: **Spring Flowering Response to Climate Change between 1936 and 2006 in Alberta, Canada**. *BioScience* 2011, **61**:514-524.
50. Gu L, Hanson PJ, Post WM, Kaiser DP, Yang B, Nemani R, Pallardy SG, Meyers T: **The 2007 Eastern US Spring Freeze: Increased Cold Damage in a Warming World?** *BioScience* 2008, **58**:253-262.
51. Saino N, Ambrosini R, Rubolini D, von Hardenberg J, Provenzale A, Huppop K, Huppop O, Lehikoinen A, Lehikoinen E, Rainio K, et al.: **Climate warming, ecological mismatch at arrival and population decline in migratory birds**. *Proc Biol Sci* 2011, **278**:835-842.
52. Hufkens K, Friedl MA, Keenan TF, Sonnentag O, Bailey A, O'Keefe J, Richardson AD: **Ecological impacts of a widespread frost event following early spring leaf-out**. *Global Change Biology* 2012, **18**:2365-2377.
53. Sakata T, Oshino T, Miura S, Tomabechi M, Tsunaga Y, Higashitani N, Miyazawa Y, Takahashi H, Watanabe M, Higashitani A: **Auxins reverse plant male sterility caused by high temperatures**. *Proc Natl Acad Sci U S A* 2010, **107**:8569-8574.
54. Madan P, Jagadish SV, Craufurd PQ, Fitzgerald M, Lafarge T, Wheeler TR: **Effect of elevated CO2 and high temperature on seed-set and grain quality of rice**. *J Exp Bot* 2012, **63**:3843-3852.
55. Bita C, Zenoni S, Vriezen W, Mariani C, Pezzotti M, Gerats T: **Temperature stress differentially modulates transcription in meiotic anthers of heat-tolerant and heat-sensitive tomato plants**. *BMC Genomics* 2011, **12**:1-18.
56. Hedhly A: **Sensitivity of flowering plant gametophytes to temperature fluctuations**. *Environmental and Experimental Botany* 2011, **74**:9-16.
57. Chauhan H, Khurana N, Tyagi A, Khurana J, Khurana P: **Identification and characterization of high temperature stress responsive genes in bread wheat (*Triticum aestivum* L.) and their regulation at various stages of development**. *Plant Mol Biol* 2011, **75**:35-51.
58. Qin D, Wu H, Peng H, Yao Y, Ni Z, Li Z, Zhou C, Sun Q: **Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using Wheat Genome Array**. *BMC Genomics* 2008, **9**:1-19.
59. Abiko M, Akibayashi K, Sakata T, Kimura M, Kihara M, Itoh K, Asamizu E, Sato S, Takahashi H, Higashitani A: **High-temperature induction of male sterility during barley (*Hordeum vulgare* L.) anther development is mediated by transcriptional inhibition**. *Sex Plant Reprod* 2005, **18**:91-100.
60. Endo M, Tsuchiya T, Hamada K, Kawamura S, Yano K, Ohshima M, Higashitani A, Watanabe M, Kawagishi-Kobayashi M: **High temperatures cause male sterility in rice plants with transcriptional alterations during pollen development**. *Plant Cell Physiol* 2009, **50**:1911-1922.
61. Kim SY, Hong CB, Lee I: **Heat Shock Stress Causes Stage-specific Male Sterility in *Arabidopsis thaliana***. *J Plant Res* 2001, **114**:301-307.
62. Oshino T, Miura S, Kikuchi S, Hamada K, Yano K, Watanabe M, Higashitani A: **Auxin depletion in barley plants under high-temperature conditions**

represses DNA proliferation in organelles and nuclei via transcriptional alterations. *Plant Cell Environ* 2011, **34**:284-290.
63. Kaufmann K, Muino JM, Jauregui R, Airoldi CA, Smaczniak C, Krajewski P, Angenent GC: **Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower**. *PLoS Biol* 2009, **7**:e1000090.
64. Prasad PVV, Bukovnik U, Fritz AK, Pisipati SR, Ristic Z: **Impact of Nighttime Temperature on Physiology and Growth of Spring Wheat [electronic resource]**. *Crop science* 2008, **48**:2372-2380.
65. Lobell DB, Ortiz-Monasterio JI: **Impacts Of Day Versus Night Temperatures On Spring Wheat Yields**. *Agron. J.* 2007, **99**:469-477.
66. Singh S, Gupta AK, Kaur N: **Influence of drought and sowing time on protein composition, antinutrients, and mineral contents of wheat**. *ScientificWorldJournal* 2012, **2012**:485751.
67. Danilevskaya ON, Meng X, Selinger DA, Deschamps S, Hermon P, Vansant G, Gupta R, Ananiev EV, Muszynski MG: **Involvement of the MADS-box gene ZMM4 in floral induction and inflorescence development in maize**. *Plant Physiol* 2008, **147**:2054-2069.
68. Preston JC, Kellogg EA: **Conservation and divergence of APETALA1/FRUITFULL-like gene function in grasses: evidence from gene expression analyses**. *Plant J* 2007, **52**:69-81.
69. Kobata T, Uemuki N: **High Temperatures During The Grain-filling Period Do Not Reduce The Potential Grain Dry Matter Increase Of Rice**. *Agron. J.* 2004, **96**:406-414.
70. Semenov MA, Halford NG: **Identifying target traits and molecular mechanisms for wheat breeding under a changing climate**. *J Exp Bot* 2009, **60**:2791-2804.
71. Morita S, Yonemaru J, Takanashi J: **Grain growth and endosperm cell size under high night temperatures in rice (Oryza sativa L.)**. *Ann Bot* 2005, **95**:695-701.
72. Fois S, Schlichting L, Marchylo B, Dexter J, Motzo R, Giunta F: **Environmental conditions affect semolina quality in durum wheat (Triticum turgidum ssp. durum L.) cultivars with different gluten strength and gluten protein composition**. *J Sci Food Agric* 2011, **91**:2664-2673.
73. Lin CJ, Li CY, Lin SK, Yang FH, Huang JJ, Liu YH, Lur HS: **Influence of high temperature during grain filling on the accumulation of storage proteins and grain quality in rice (Oryza sativa L.)**. *J Agric Food Chem* 2010, **58**:10545-10552.
74. Zhao XY, Cheng ZJ, Zhang XS: **Overexpression of TaMADS1, a SEPALLATA-like gene in wheat, causes early flowering and the abnormal development of floral organs in Arabidopsis**. *Planta* 2006, **223**:698-707.
75. Zhang J, Nallamilli BR, Mujahid H, Peng Z: **OsMADS6 plays an essential role in endosperm nutrient accumulation and is subject to epigenetic regulation in rice (Oryza sativa)**. *Plant J* 2010, **64**:604-617.
76. Yin LL, Xue HW: **The MADS29 transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development**. *Plant Cell* 2012, **24**:1049-1065.
77. Gramzow L, Ritz MS, Theißen G: **On the origin of MADS-domain transcription factors**. *Trends Genet* 2010, **26**:149-153.
78. Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, et al.: **Molecular and phylogenetic analyses**

of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell* 2003, **15**:1538-1551.
79. Messenguy F, Dubois E: **Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development**. *Gene* 2003, **316**:1-21.
80. Shore P, Sharrocks AD: **The MADS-box family of transcription factors**. *Eur J Biochem* 1995, **229**:1-13.
81. Nam J, dePamphilis CW, Ma H, Nei M: **Antiquity and evolution of the MADS-box gene family controlling flower development in plants**. *Mol Biol Evol* 2003, **20**:1435-1447.
82. Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF: **MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes**. *Plant J* 2000, **24**:457-466.
83. Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF: **Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family**. *Genetics* 1995, **140**:345-356.
84. Theißen G, Kim JT, Saedler H: **Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes**. *J Mol Evol* 1996, **43**:484-516.
85. Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M: **Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses**. *Mol Biol Evol* 2003, **20**:1963-1977.
86. De Bodt S, Raes J, Florquin K, Rombauts S, Rouze P, Theißen G, Van de Peer Y: **Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants**. *J Mol Evol* 2003, **56**:573-586.
87. Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martinez-Castilla L, Yanofsky MF: **An ancestral MADS-box gene duplication occurred before the divergence of plants and animals**. *Proc Natl Acad Sci U S A* 2000, **97**:5328-5333.
88. Bemer M, Heijmans K, Airoldi C, Davies B, Angenent GC: **An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis**. *Plant Physiol* 2010, **154**:287-300.
89. Nam J, Kim J, Lee S, An G, Ma H, Nei M: **Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms**. *Proc Natl Acad Sci U S A* 2004, **101**:1910-1915.
90. Bemer M, Wolters-Arts M, Grossniklaus U, Angenent GC: **The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules**. *Plant Cell* 2008, **20**:2088-2101.
91. Kang IH, Steffen JG, Portereiko MF, Lloyd A, Drews GN: **The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis**. *Plant Cell* 2008, **20**:635-647.
92. Steffen JG, Kang IH, Portereiko MF, Lloyd A, Drews GN: **AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis**. *Plant Physiol* 2008, **148**:259-268.
93. Portereiko MF, Lloyd A, Steffen JG, Punwani JA, Otsuga D, Drews GN: **AGL80 is required for central cell and endosperm development in Arabidopsis**. *Plant Cell* 2006, **18**:1862-1872.

94. Yoo SK, Lee JS, Ahn JH: **Overexpression of AGAMOUS-LIKE 28 (AGL28) promotes flowering by upregulating expression of floral promoters within the autonomous pathway**. *Biochem Biophys Res Commun* 2006, **348**:929-936.
95. Kaufmann K, Melzer R, Theißen G: **MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants**. *Gene* 2005, **347**:183-198.
96. Becker A, Theißen G: **The major clades of MADS-box genes and their role in the development and evolution of flowering plants**. *Mol Phylogenet Evol* 2003, **29**:464-489.
97. Henschel K, Kofuji R, Hasebe M, Saedler H, Munster T, Theißen G: **Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens**. *Mol Biol Evol* 2002, **19**:801-814.
98. Riese M, Faigl W, Quodt V, Verelst W, Matthes A, Saedler H, Munster T: **Isolation and characterization of new MIKC*-Type MADS-box genes from the moss Physcomitrella patens**. *Plant Biol (Stuttg)* 2005, **7**:307-314.
99. Zobell O, Faigl W, Saedler H, Munster T: **MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants**. *Mol Biol Evol* 2010, **27**:1201-1211.
100. Adamczyk BJ, Fernandez DE: **MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis**. *Plant Physiol* 2009, **149**:1713-1723.
101. Verelst W, Saedler H, Munster T: **MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters**. *Plant Physiol* 2007, **143**:447-460.
102. Verelst W, Twell D, de Folter S, Immink R, Saedler H, Munster T: **MADS-complexes regulate transcriptome dynamics during pollen maturation**. *Genome Biol* 2007, **8**:R249.
103. Zhao T, Ni Z, Dai Y, Yao Y, Nie X, Sun Q: **Characterization and expression of 42 MADS-box genes in wheat (Triticum aestivum L.)**. *Mol Genet Genomics* 2006, **276**:334-350.
104. Krogan NT, Ashton NW: **Ancestry of plant MADS-box genes revealed by bryophyte (Physcomitrella patens) homologues**. *New Phytologist* 2000, **147**:505-517.
105. Theißen G, Münster T, Henschel K: **Why don't mosses flower?** *New Phytologist* 2001, **150**:1-5.
106. Svensson ME, Engström P: **Closely related MADS-box genes in club moss (Lycopodium) show broad expression patterns and are structurally similar to, but phylogenetically distinct from, typical seed plant MADS-box genes**. *New Phytologist* 2002, **154**:439-450.
107. Svensson ME, Johannesson H, Engstrom P: **The LAMB1 gene from the clubmoss, Lycopodium annotinum, is a divergent MADS-box gene, expressed specifically in sporogenic structures**. *Gene* 2000, **253**:31-43.
108. Himi S, Sano R, Nishiyama T, Tanahashi T, Kato M, Ueda K, Hasebe M: **Evolution of MADS-box gene induction by FLO/LFY genes**. *J Mol Evol* 2001, **53**:387-393.
109. Munster T, Pahnke J, Di Rosa A, Kim JT, Martin W, Saedler H, Theißen G: **Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants**. *Proc Natl Acad Sci U S A* 1997, **94**:2415-2420.

110. Englund M, Carlsbecker A, Engstrom P, Vergara-Silva F: **Morphological "primary homology" and expression of AG-subfamily MADS-box genes in pines, podocarps, and yews**. *Evol Dev* 2011, **13**:171-181.
111. Lovisetto A, Guzzo F, Tadiello A, Toffali K, Favretto A, Casadoro G: **Molecular analyses of MADS-box genes trace back to Gymnosperms the invention of fleshy fruits**. *Mol Biol Evol* 2012, **29**:409-419.
112. Jager M, Hassanin A, Manuel M, Le Guyader H, Deutsch J: **MADS-box genes in Ginkgo biloba and the evolution of the AGAMOUS family**. *Mol Biol Evol* 2003, **20**:842-854.
113. Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Munster T, Theißen G, Nozaki H, Ito M: **Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes**. *Proc Natl Acad Sci U S A* 2005, **102**:2436-2441.
114. Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynie S, Cooke R, et al.: **Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features**. *Proc Natl Acad Sci U S A* 2006, **103**:11647-11652.
115. Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, et al.: **The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation**. *Proc Natl Acad Sci U S A* 2007, **104**:7705-7710.
116. Ciaffi M, Paolacci AR, Tanzarella OA, Porceddu E: **Molecular aspects of flower development in grasses**. *Sex Plant Reprod* 2011, **24**:247-282.
117. Mondragon-Palomino M, Hiese L, Harter A, Koch MA, Theißen G: **Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses**. *BMC Evol Biol* 2009, **9**:81.
118. Leseberg CH, Li A, Kang H, Duvall M, Mao L: **Genome-wide analysis of the MADS-box gene family in Populus trichocarpa**. *Gene* 2006, **378**:84-94.
119. de Folter S, Immink RG, Kieffer M, Parenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, et al.: **Comprehensive interaction map of the Arabidopsis MADS Box transcription factors**. *Plant Cell* 2005, **17**:1424-1433.
120. Smaczniak C, Immink RG, Muino JM, Blanvillain R, Busscher M, Busscher-Lange J, Dinh QD, Liu S, Westphal AH, Boeren S, et al.: **Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development**. *Proc Natl Acad Sci U S A* 2012, **109**:1560-1565.
121. Leseberg CH, Eissler CL, Wang X, Johns MA, Duvall MR, Mao L: **Interaction study of MADS-domain proteins in tomato**. *J Exp Bot* 2008, **59**:2253-2265.
122. Ripoll JJ, Rodriguez-Cazorla E, Gonzalez-Reig S, Andujar A, Alonso-Cantabrana H, Perez-Amador MA, Carbonell J, Martinez-Laborda A, Vera A: **Antagonistic interactions between Arabidopsis K-homology domain genes uncover PEPPER as a positive regulator of the central floral repressor FLOWERING LOCUS C**. *Dev Biol* 2009, **333**:251-262.
123. Wang YQ, Melzer R, Theißen G: **Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm Gnetum gnemon provide a clue to the evolutionary origin of 'floral quartets'**. *Plant J* 2010, **64**:177-190.
124. Consortium AIM: **Evidence for network evolution in an Arabidopsis interactome map**. *Science* 2011, **333**:601-607.

125. Yang Y, Jack T: **Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins**. *Plant Mol Biol* 2004, **55**:45-59.
126. Vandenbussche M, Theißen G, Van de Peer Y, Gerats T: **Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations**. *Nucleic Acids Res* 2003, **31**:4401-4409.
127. Kramer EM, Su HJ, Wu CC, Hu JM: **A simplified explanation for the frameshift mutation that created a novel C-terminal motif in the APETALA3 gene lineage**. *BMC Evol Biol* 2006, **6**:30.
128. Cho S, Jang S, Chae S, Chung KM, Moon YH, An G, Jang SK: **Analysis of the C-terminal region of Arabidopsis thaliana APETALA1 as a transcription activation domain**. *Plant Mol Biol* 1999, **40**:419-429.
129. Mizukami Y, Huang H, Tudor M, Hu Y, Ma H: **Functional domains of the floral regulator AGAMOUS: characterization of the DNA binding domain and analysis of dominant negative mutations**. *Plant Cell* 1996, **8**:831-845.
130. Krizek BA, Meyerowitz EM: **Mapping the protein regions responsible for the functional specificities of the Arabidopsis MADS domain organ-identity proteins**. *Proc Natl Acad Sci U S A* 1996, **93**:4063-4070.
131. Piwarzyk E, Yang Y, Jack T: **Conserved C-terminal motifs of the Arabidopsis proteins APETALA3 and PISTILLATA are dispensable for floral organ identity function**. *Plant Physiol* 2007, **145**:1495-1505.
132. Su K, Zhao S, Shan H, Kong H, Lu W, Theißen G, Chen Z, Meng Z: **The MIK region rather than the C-terminal domain of AP3-like class B floral homeotic proteins determines functional specificity in the development and evolution of petals**. *New Phytol* 2008, **178**:544-558.
133. Gaffe J, Lemercier C, Alcaraz JP, Kuntz M: **Identification of three tomato flower and fruit MADS-box proteins with a putative histone deacetylase binding domain**. *Gene* 2011, **471**:19-26.
134. Gregoire S, Xiao L, Nie J, Zhang X, Xu M, Li J, Wong J, Seto E, Yang XJ: **Histone deacetylase 3 interacts with and deacetylates myocyte enhancer factor 2**. *Mol Cell Biol* 2007, **27**:1280-1295.
135. Molkentin JD, Li L, Olson EN: **Phosphorylation of the MADS-Box transcription factor MEF2C enhances its DNA binding activity**. *J Biol Chem* 1996, **271**:17199-17204.
136. Youn HD, Chatila TA, Liu JO: **Integration of calcineurin and MEF2 signals by the coactivator p300 during T-cell apoptosis**. *EMBO J* 2000, **19**:4323-4331.
137. He J, Ye J, Cai Y, Riquelme C, Liu JO, Liu X, Han A, Chen L: **Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly**. *Nucleic Acids Res* 2011, **39**:4464-4474.
138. Kuo MH, Nadeau ET, Grayhack EJ: **Multiple phosphorylated forms of the Saccharomyces cerevisiae Mcm1 protein include an isoform induced in response to high salt concentrations**. *Mol Cell Biol* 1997, **17**:819-832.
139. Robertson M, Helliwell CA, Dennis ES: **Post-translational modifications of the endogenous and transgenic FLC protein in Arabidopsis thaliana**. *Plant Cell Physiol* 2008, **49**:1859-1866.
140. Sridhar VV, Surendrarao A, Liu Z: **APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development**. *Development* 2006, **133**:3159-3166.

141. Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, Gorla MS, Mantovani R, Colombo L, Kater MM: **Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB**. *J Biol Chem* 2002, **277**:26429-26435.
142. Gamboa A, Paez-Valencia J, Acevedo GF, Vazquez-Moreno L, Alvarez-Buylla RE: **Floral transcription factor AGAMOUS interacts in vitro with a leucine-rich repeat and an acid phosphatase protein complex**. *Biochem Biophys Res Commun* 2001, **288**:1018-1026.
143. Latinkic BV, Zeremski M, Lau LF: **Elk-1 can recruit SRF to form a ternary complex upon the serum response element**. *Nucleic Acids Res* 1996, **24**:1345-1351.
144. Williams GT, Lau LF: **Activation of the inducible orphan receptor gene nur77 by serum growth factors: dissociation of immediate-early and delayed-early responses**. *Mol Cell Biol* 1993, **13**:6124-6136.
145. Mo Y, Ho W, Johnston K, Marmorstein R: **Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex**. *J Mol Biol* 2001, **314**:495-506.
146. Givaty O, Levy Y: **Protein sliding along DNA: dynamics and structural characterization**. *J Mol Biol* 2009, **385**:1087-1097.
147. Berg OG, Winter RB, von Hippel PH: **Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory**. *Biochemistry* 1981, **20**:6929-6948.
148. Imai A, Matsuyama T, Hanzawa Y, Akiyama T, Tamaoki M, Saji H, Shirano Y, Kato T, Hayashi H, Shibata D, et al.: **Spermidine synthase genes are essential for survival of Arabidopsis**. *Plant Physiol* 2004, **135**:1565-1573.
149. Lindemose S, Nielsen PE, Mollegaard NE: **Polyamines preferentially interact with bent adenine tracts in double-stranded DNA**. *Nucleic Acids Res* 2005, **33**:1790-1803.
150. Korolev N, Lyubartsev AP, Laaksonen A, Nordenskiold L: **Molecular dynamics simulation study of oriented polyamine- and Na-DNA: sequence specific interactions and effects on DNA structure**. *Biopolymers* 2004, **73**:542-555.
151. Panagiotidis CA, Artandi S, Calame K, Silverstein SJ: **Polyamines alter sequence-specific DNA-protein interactions**. *Nucleic Acids Res* 1995, **23**:1800-1809.
152. Von Hippel PH, McGhee JD: **DNA-protein interactions**. *Annu Rev Biochem* 1972, **41**:231-300.
153. Garriga P, Garcia-Quintana D, Sagi J, Manyosa J: **An A-form of poly(amino2dA-dT).poly(amino2dA-dT) induced by polyamines**. *Biochemistry* 1993, **32**:1067-1071.
154. Haworth IS, Rodger A, Richards WG: **A molecular dynamics simulation of a polyamine-induced conformational change of DNA. A possible mechanism for the B to Z transition**. *J Biomol Struct Dyn* 1992, **10**:195-211.
155. Thomas TJ, Gunnia UB, Thomas T: **Polyamine-induced B-DNA to Z-DNA conformational transition of a plasmid DNA with (dG-dC)n insert**. *J Biol Chem* 1991, **266**:6137-6141.
156. Minyat EE, Ivanov VI, Kritzyn AM, Minchenkova LE, Schyolkina AK: **Spermine and spermidine-induced B to A transition of DNA in solution**. *J Mol Biol* 1979, **128**:397-409.
157. Shen X, Gu B, Che SA, Zhang FS: **Solvent effects on the conformation of DNA dodecamer segment: a simulation study**. *J Chem Phys* 2011, **135**:034509.

158. Li W, Nordenskiold L, Mu Y: **Sequence-specific Mg2+-DNA interactions: a molecular dynamics simulation study**. *J Phys Chem B* 2011, **115**:14713-14720.
159. Gavryushov S: **Electrostatics of B-DNA in NaCl and CaCl2 solutions: ion size, interionic correlation, and solvent dielectric saturation effects**. *J Phys Chem B* 2008, **112**:8955-8965.
160. Martin AM, Sam MD, Reich NO, Perona JJ: **Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease**. *Nat Struct Biol* 1999, **6**:269-277.
161. Howerton SB, Sines CC, VanDerveer D, Williams LD: **Locating monovalent cations in the grooves of B-DNA**. *Biochemistry* 2001, **40**:10023-10031.
162. Jagoda-Cwiklik B, Vacha R, Lund M, Srebro M, Jungwirth P: **Ion pairing as a possible clue for discriminating between sodium and potassium in biological and other complex environments**. *J Phys Chem B* 2007, **111**:14077-14079.
163. Cheng Y, Korolev N, Nordenskiold L: **Similarities and differences in interaction of K+ and Na+ with condensed ordered DNA. A molecular dynamics computer simulation study**. *Nucleic Acids Res* 2006, **34**:686-696.
164. Varnai P, Zakrzewska K: **DNA and its counterions: a molecular dynamics study**. *Nucleic Acids Res* 2004, **32**:4269-4280.
165. Robinson H, Wang AH: **Neomycin, spermine and hexaamminecobalt (III) share common structural motifs in converting B- to A-DNA**. *Nucleic Acids Res* 1996, **24**:676-682.
166. Hamilton PL, Arya DP: **Natural product DNA major groove binders**. *Nat Prod Rep* 2012, **29**:134-143.
167. Basham B, Schroth GP, Ho PS: **An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA**. *Proc Natl Acad Sci U S A* 1995, **92**:6464-6468.
168. Rich A, Nordheim A, Wang AH: **The chemistry and biology of left-handed Z-DNA**. *Annu Rev Biochem* 1984, **53**:791-846.
169. Wang AH, Gessner RV, van der Marel GA, van Boom JH, Rich A: **Crystal structure of Z-DNA without an alternating purine-pyrimidine sequence**. *Proc Natl Acad Sci U S A* 1985, **82**:3611-3615.
170. Cheatham TE, 3rd, Kollman PA: **Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution**. *J Mol Biol* 1996, **259**:434-444.
171. Cheatham TE, 3rd, Kollman PA: **Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]2 in the presence of hexaamminecobalt(III)**. *Structure* 1997, **5**:1297-1311.
172. Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC: **The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity**. *Mol Biol Evol* 2003, **20**:735-747.
173. Gong ZZ, Yamazaki M, Saito K: **Critical role of alanine-161 in Delila protein involved in regulation of anthocyanin pigmentation for transcriptional activation in yeast**. *Plant Biotech* 2000, **17**:309-314.
174. Gong ZZ, Yamagishi E, Yamazaki M, Saito K: **A constitutively expressed Myc-like gene involved in anthocyanin biosynthesis from Perilla frutescens: molecular characterization, heterologous expression in transgenic plants and transactivation in yeast cells**. *Plant Mol Biol* 1999, **41**:33-44.

175. Goodrich J, Carpenter R, Coen ES: **A common gene regulates pigmentation pattern in diverse plant species**. *Cell* 1992, **68**:955-964.
176. Tuerck JA, Fromm ME: **Elements of the maize A1 promoter required for transactivation by the anthocyanin B/C1 or phlobaphene P regulatory genes**. *Plant Cell* 1994, **6**:1655-1663.
177. Bodeau JP, Walbot V: **Regulated transcription of the maize Bronze-2 promoter in electroporated protoplasts requires the C1 and R gene products**. *Mol Gen Genet* 1992, **233**:379-387.
178. Roth BA, Goff SA, Klein TM, Fromm ME: **C1- and R-dependent expression of the maize Bz1 gene requires sequences with homology to mammalian myb and myc binding sites**. *Plant Cell* 1991, **3**:317-325.
179. Reddy AR, Britsch L, Salamini F, Saedler H, Rohde W: **The A1 (anthocyanin-1) locus in Zea mays encodes dihydroquercetin reductase**. *Plant Science* 1987, **52**:7-13.
180. Selinger DA, Chandler VL: **A mutation in the pale aleurone color1 gene identifies a novel regulator of the maize anthocyanin pathway**. *Plant Cell* 1999, **11**:5-14.
181. de Pater S, Pham K, Memelink J, Kijne J: **RAP-1 is an Arabidopsis MYC-like R protein homologue, that binds to G-box sequence motifs**. *Plant Mol Biol* 1997, **34**:169-174.
182. Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K: **Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression**. *Plant Cell* 1997, **9**:1859-1868.
183. Martin C, Prescott A, Mackay S, Bartlett J, Vrijlandt E: **Control of anthocyanin biosynthesis in flowers of Antirrhinum majus**. *Plant J* 1991, **1**:37-49.
184. Toledo-Ortiz G, Huq E, Quail PH: **The Arabidopsis basic/helix-loop-helix transcription factor family**. *Plant Cell* 2003, **15**:1749-1770.
185. Krylov D, Barchi J, Vinson C: **Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids**. *J Mol Biol* 1998, **279**:959-972.
186. Landschulz WH, Johnson PF, McKnight SL: **The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite**. *Science* 1989, **243**:1681-1688.
187. Hodges RS, Sodek J, Smillie LB, Jurasek L: **Tropomyosin: Amino Acid Sequence and Coiled-Coil Structure**. *Cold Spring Harb Symp Quant Biol* 1973, **37**:299-310.
188. McLachlan AD, Stewart M: **Tropomyosin coiled-coil interactions: evidence for an unstaggered structure**. *J Mol Biol* 1975, **98**:293-304.
189. Wagschal K, Tripet B, Lavigne P, Mant C, Hodges RS: **The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins**. *Protein Sci* 1999, **8**:2312-2329.
190. Kohn WD, Kay CM, Hodges RS: **Protein destabilization by electrostatic repulsions in the two-stranded alpha-helical coiled-coil/leucine zipper**. *Protein Sci* 1995, **4**:237-250.
191. O'Shea EK, Lumb KJ, Kim PS: **Peptide 'Velcro': design of a heterodimeric coiled coil**. *Curr Biol* 1993, **3**:658-667.
192. Schuermann M, Hunter JB, Hennig G, Muller R: **Non-leucine residues in the leucine repeats of Fos and Jun contribute to the stability and determine the specificity of dimerization**. *Nucleic Acids Res* 1991, **19**:739-746.

193. Spek EJ, Bui AH, Lu M, Kallenbach NR: **Surface salt bridges stabilize the GCN4 leucine zipper**. *Protein Sci* 1998, **7**:2431-2437.
194. John M, Briand JP, Granger-Schnarr M, Schnarr M: **Two pairs of oppositely charged amino acids from Jun and Fos confer heterodimerization to GCN4 leucine zipper**. *J Biol Chem* 1994, **269**:16247-16253.
195. Thompson KS, Vinson CR, Freire E: **Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4**. *Biochemistry* 1993, **32**:5491-5496.
196. Steinmetz MO, Jelesarov I, Matousek WM, Honnappa S, Jahnke W, Missimer JH, Frank S, Alexandrescu AT, Kammerer RA: **Molecular basis of coiled-coil formation**. *Proc Natl Acad Sci U S A* 2007, **104**:7062-7067.
197. Matousek WM, Ciani B, Fitch CA, Garcia-Moreno B, Kammerer RA, Alexandrescu AT: **Electrostatic contributions to the stability of the GCN4 leucine zipper structure**. *J Mol Biol* 2007, **374**:206-219.
198. Marti DN, Bosshard HR: **Electrostatic interactions in leucine zippers: thermodynamic analysis of the contributions of Glu and His residues and the effect of mutating salt bridges**. *J Mol Biol* 2003, **330**:621-637.
199. Phelan P, Gorfe AA, Jelesarov I, Marti DN, Warwicker J, Bosshard HR: **Salt bridges destabilize a leucine zipper designed for maximized ion pairing between helices**. *Biochemistry* 2002, **41**:2998-3008.
200. Krystek SR, Jr., Bruccoleri RE, Novotny J: **Stabilities of leucine zipper dimers estimated by an empirical free energy method**. *Int J Pept Protein Res* 1991, **38**:229-236.
201. Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, Bonovich M: **Classification of human B-ZIP proteins based on dimerization properties**. *Mol Cell Biol* 2002, **22**:6321-6335.
202. Vinson CR, Hai T, Boyd SM: **Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design**. *Genes Dev* 1993, **7**:1047-1058.
203. Guex N, Peitsch MC: **SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling**. *Electrophoresis* 1997, **18**:2714-2723.
204. Kopp J, Schwede T: **The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models**. *Nucleic Acids Res* 2004, **32**:D230-234.
205. Schwede T, Kopp J, Guex N, Peitsch MC: **SWISS-MODEL: An automated protein homology-modeling server**. *Nucleic Acids Res* 2003, **31**:3381-3385.
206. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank**. *Nucleic Acids Res* 2000, **28**:235-242.
207. Brownlie P, Ceska T, Lamers M, Romier C, Stier G, Teo H, Suck D: **The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control**. *Structure* 1997, **5**:509-520.
208. Ferre-D'Amare AR, Prendergast GC, Ziff EB, Burley SK: **Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain**. *Nature* 1993, **363**:38-45.
209. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22**:4673-4680.

210. Patra B, Pattanaik S, Yuan L: **Ubiquitin protein ligase 3 mediates the proteasomal degradation of GLABROUS 3 and ENHANCER OF GLABROUS 3, regulators of trichome development and flavonoid biosynthesis in Arabidopsis**. *Plant J* 2013.
211. Liu Y, Chapagain PP, Gerstman BS: **Stabilization of native and non-native structures by salt bridges in a lattice model of the GCN4 leucine dimer**. *J Phys Chem B* 2010, **114**:796-803.
212. Kammerer RA, Jaravine VA, Frank S, Schulthess T, Landwehr R, Lustig A, Garcia-Echeverria C, Alexandrescu AT, Engel J, Steinmetz MO: **An intrahelical salt bridge within the trigger site stabilizes the GCN4 leucine zipper**. *J Biol Chem* 2001, **276**:13685-13688.
213. Durr E, Jelesarov I, Bosshard HR: **Extremely fast folding of a very stable leucine zipper with a strengthened hydrophobic core and lacking electrostatic interactions between helices**. *Biochemistry* 1999, **38**:870-880.
214. Lumb KJ, Kim PS: **Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper**. *Science* 1995, **268**:436-439.
215. McDuff FO, Naud JF, Montagne M, Sauve S, Lavigne P: **The Max homodimeric b-HLH-LZ significantly interferes with the specific heterodimerization between the c-Myc and Max b-HLH-LZ in absence of DNA: a quantitative analysis**. *J Mol Recognit* 2009, **22**:261-269.
216. Rishi V, Gal J, Krylov D, Fridriksson J, Boysen MS, Mandrup S, Vinson C: **SREBP-1 dimerization specificity maps to both the helix-loop-helix and leucine zipper domains: use of a dominant negative**. *J Biol Chem* 2004, **279**:11863-11874.
217. Deppmann CD, Acharya A, Rishi V, Wobbes B, Smeekens S, Taparowsky EJ, Vinson C: **Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs**. *Nucleic Acids Res* 2004, **32**:3435-3445.
218. Gorfe AA, Ferrara P, Caflisch A, Marti DN, Bosshard HR, Jelesarov I: **Calculation of protein ionization equilibria with conformational sampling: pK(a) of a model leucine zipper, GCN4 and barnase**. *Proteins* 2002, **46**:41-60.
219. Kohler JJ, Metallo SJ, Schneider TL, Schepartz A: **DNA specificity enhanced by sequential binding of protein monomers**. *Proc Natl Acad Sci U S A* 1999, **96**:11735-11739.
220. Hendsch ZS, Tidor B: **Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding**. *Protein Sci* 1999, **8**:1381-1392.
221. Zeng X, Zhu H, Lashuel HA, Hu JC: **Oligomerization properties of GCN4 leucine zipper e and g position mutants**. *Protein Sci* 1997, **6**:2218-2226.
222. Baxevanis AD, Vinson CR: **Interactions of coiled coils in transcription factors: where is the specificity?** *Curr Opin Genet Dev* 1993, **3**:278-285.
223. Jean-Francois N, Frederic G, Raymund W, Benoit C, Lavigne P: **Improving the thermodynamic stability of the leucine zipper of max increases the stability of its b-HLH-LZ:E-box complex**. *J Mol Biol* 2003, **326**:1577-1595.
224. Hu J, Banerjee A, Goss DJ: **Assembly of b/HLH/z proteins c-Myc, Max, and Mad1 with cognate DNA: importance of protein-protein and protein-DNA interactions**. *Biochemistry* 2005, **44**:11855-11863.

225. Hurlin PJ, Queva C, Eisenman RN: **Mnt, a novel Max-interacting protein is coexpressed with Myc in proliferating cells and mediates repression at Myc binding sites**. *Genes Dev* 1997, **11**:44-58.
226. Muhle-Goll C, Nilges M, Pastore A: **The leucine zippers of the HLH-LZ proteins Max and c-Myc preferentially form heterodimers**. *Biochemistry* 1995, **34**:13554-13564.
227. Ayer DE, Kretzner L, Eisenman RN: **Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity**. *Cell* 1993, **72**:211-222.
228. Metallo SJ, Schepartz A: **Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA**. *Nat Struct Biol* 1997, **4**:115-117.
229. Wendt H, Thomas RM, Ellenberger T: **DNA-mediated folding and assembly of MyoD-E47 heterodimers**. *J Biol Chem* 1998, **273**:5735-5743.
230. Kohler JJ, Schepartz A: **Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization**. *Biochemistry* 2001, **40**:130-142.
231. Kohler JJ, Schepartz A: **Effects of nucleic acids and polyanions on dimer formation and DNA binding by bZIP and bHLHZip transcription factors**. *Bioorg Med Chem* 2001, **9**:2435-2443.
232. Cranz S, Berger C, Baici A, Jelesarov I, Bosshard HR: **Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site**. *Biochemistry* 2004, **43**:718-727.
233. Lee TH, Maheshri N: **A regulatory role for repeated decoy transcription factor binding sites in target gene expression**. *Mol Syst Biol* 2012, **8**:576.
234. Fried M, Crothers DM: **Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis**. *Nucleic Acids Res* 1981, **9**:6505-6525.
235. Garner MM, Revzin A: **A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system**. *Nucleic Acids Res* 1981, **9**:3047-3060.
236. Lequin RM: **Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA)**. *Clin Chem* 2005, **51**:2415-2418.
237. Hibma MH, Ely SJ, Crawford L: **A non-radioactive assay for the detection and quantitation of a DNA binding protein**. *Nucleic Acids Res* 1994, **22**:3806-3807.
238. Benotmane AM, Hoylaerts MF, Collen D, Belayew A: **Nonisotopic quantitative analysis of protein-DNA interactions at equilibrium**. *Anal Biochem* 1997, **250**:181-185.
239. Renard P, Ernest I, Houbion A, Art M, Le Calvez H, Raes M, Remacle J: **Development of a sensitive multi-well colorimetric assay for active NFkappaB**. *Nucleic Acids Res* 2001, **29**:E21.
240. Alonso R, Onate-Sanchez L, Weltmeier F, Ehlert A, Diaz I, Dietrich K, Vicente-Carbajosa J, Droge-Laser W: **A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation**. *Plant Cell* 2009, **21**:1747-1761.
241. Brand LH, Kirchler T, Hummel S, Chaban C, Wanke D: **DPI-ELISA: a fast and versatile method to specify the binding of plant transcription factors to DNA in vitro**. *Plant Methods* 2010, **6**:25.
242. Studier FW: **Protein production by auto-induction in high density shaking cultures**. *Protein Expr Purif* 2005, **41**:207-234.

243. Tang W, Perry SE: **Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study**. *J Biol Chem* 2003, **278**:28154-28159.
244. Nieba L, Nieba-Axmann SE, Persson A, Hamalainen M, Edebratt F, Hansson A, Lidholm J, Magnusson K, Karlsson AF, Pluckthun A: **BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip**. *Anal Biochem* 1997, **252**:217-228.
245. Huang H, Tudor M, Su T, Zhang Y, Hu Y, Ma H: **DNA binding properties of two Arabidopsis MADS domain proteins: binding consensus and dimer formation**. *Plant Cell* 1996, **8**:81-94.
246. Bolanos-Garcia VM, Davies OR: **Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography**. *Biochim Biophys Acta* 2006, **1760**:1304-1313.
247. O'Shannessy DJ, O'Donnell KC, Martin J, Brigham-Burke M: **Detection and quantitation of hexa-histidine-tagged recombinant proteins on western blots and by a surface plasmon resonance biosensor technique**. *Anal Biochem* 1995, **229**:119-124.
248. Knecht S, Ricklin D, Eberle AN, Ernst B: **Oligohis-tags: mechanisms of binding to Ni2+-NTA surfaces**. *J Mol Recognit* 2009, **22**:270-279.
249. Khan F, He M, Taussig MJ: **Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on ni-nitrilotriacetic acid surfaces**. *Anal Chem* 2006, **78**:3072-3079.
250. Paborsky LR, Dunn KE, Gibbs CS, Dougherty JP: **A nickel chelate microtiter plate assay for six histidine-containing proteins**. *Anal Biochem* 1996, **234**:60-65.
251. Smith AT, Santama N, Dacey S, Edwards M, Bray RC, Thorneley RN, Burke JF: **Expression of a synthetic gene for horseradish peroxidase C in Escherichia coli and folding and activation of the recombinant enzyme with Ca2+ and heme**. *J Biol Chem* 1990, **265**:13335-13343.
252. Lin Z, Thorsen T, Arnold FH: **Functional expression of horseradish peroxidase in E. coli by directed evolution**. *Biotechnol Prog* 1999, **15**:467-471.
253. Immink RG, Tonaco IA, de Folter S, Shchennikova A, van Dijk AD, Busscher-Lange J, Borst JW, Angenent GC: **SEPALLATA3: the 'glue' for MADS box transcription factor complex formation**. *Genome Biol* 2009, **10**:R24.
254. Gregis V, Andres F, Sessa A, Guerra RF, Simonini S, Mateos JL, Torti S, Zambelli F, Prazzoli GM, Bjerkan KN, et al.: **Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis**. *Genome Biol* 2013, **14**:R56.
255. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J: **Positional cloning of the wheat vernalization gene VRN1**. *Proc Natl Acad Sci U S A* 2003, **100**:6263-6268.
256. Bai Y, Pattanaik S, Patra B, Werkman JR, Xie CH, Yuan L: **Flavonoid-related basic helix-loop-helix regulators, NtAn1a and NtAn1b, of tobacco have originated from two ancestors and are functionally active**. *Planta* 2011, **234**:363-375.
257. Li D, Liu C, Shen L, Wu Y, Chen H, Robertson M, Helliwell CA, Ito T, Meyerowitz E, Yu H: **A repressor complex governs the integration of flowering signals in Arabidopsis**. *Dev Cell* 2008, **15**:110-120.

258. Gregis V, Sessa A, Colombo L, Kater MM: **AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower development in Arabidopsis**. *Plant Cell* 2006, **18**:1373-1382.
259. Locascio A, Lucchin M, Varotto S: **Characterization of a MADS FLOWERING LOCUS C-LIKE (MFL) sequence in Cichorium intybus: a comparative study of CiMFL and AtFLC reveals homologies and divergences in gene function**. *New Phytol* 2009, **182**:630-643.
260. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S: **MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress**. *BMC Genomics* 2007, **8**:242.
261. Yun H, Hyun Y, Kang MJ, Noh YS, Noh B, Choi Y: **Identification of regulators required for the reactivation of FLOWERING LOCUS C during Arabidopsis reproduction**. *Planta* 2011, **234**:1237-1250.
262. Choi J, Hyun Y, Kang MJ, In Yun H, Yun JY, Lister C, Dean C, Amasino RM, Noh B, Noh YS, et al.: **Resetting and regulation of Flowering Locus C expression during Arabidopsis reproductive development**. *Plant J* 2009, **57**:918-931.
263. Choi K, Kim J, Hwang HJ, Kim S, Park C, Kim SY, Lee I: **The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors**. *Plant Cell* 2011, **23**:289-303.
264. Benlloch R, Kim MC, Sayou C, Thevenon E, Parcy F, Nilsson O: **Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of APETALA1**. *Plant J* 2011, **67**:1094-1102.
265. Lee J, Lee I: **Regulation and function of SOC1, a flowering pathway integrator**. *J Exp Bot* 2010, **61**:2247-2254.
266. Horvath D: **Common mechanisms regulate flowering and dormancy**. *Plant Science* 2009, **177**:523-531.
267. Valverde F: **CONSTANS and the evolutionary origin of photoperiodic timing of flowering**. *J Exp Bot* 2011, **62**:2453-2463.
268. Zografos BR, Sung S: **Vernalization-mediated chromatin changes**. *J Exp Bot* 2012.
269. Kim DH, Sung S: **Coordination of the vernalization response through a VIN3 and FLC gene family regulatory network in Arabidopsis**. *Plant Cell* 2013, **25**:454-469.
270. Satake A, Iwasa Y: **A stochastic model of chromatin modification: cell population coding of winter memory in plants**. *J Theor Biol* 2012, **302**:6-17.
271. Sung S, Amasino RM: **Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3**. *Nature* 2004, **427**:159-164.
272. Schonrock N, Bouveret R, Leroy O, Borghi L, Kohler C, Gruissem W, Hennig L: **Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway**. *Genes Dev* 2006, **20**:1667-1678.
273. Michaels SD, Ditta G, Gustafson-Brown C, Pelaz S, Yanofsky M, Amasino RM: **AGL24 acts as a promoter of flowering in Arabidopsis and is positively regulated by vernalization**. *Plant J* 2003, **33**:867-874.
274. Fujiwara S, Oda A, Yoshida R, Niinuma K, Miyata K, Tomozoe Y, Tajima T, Nakagawa M, Hayashi K, Coupland G, et al.: **Circadian clock proteins LHY and CCA1 regulate SVP protein accumulation to control flowering in Arabidopsis**. *Plant Cell* 2008, **20**:2960-2971.

275. Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH: **Role of SVP in the control of flowering time by ambient temperature in Arabidopsis**. *Genes Dev* 2007, **21**:397-402.
276. Hartmann U, Hohmann S, Nettesheim K, Wisman E, Saedler H, Huijser P: **Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis**. *Plant J* 2000, **21**:351-360.
277. Scortecci K, Michaels SD, Amasino RM: **Genetic interactions between FLM and other flowering-time genes in Arabidopsis thaliana**. *Plant Mol Biol* 2003, **52**:915-922.
278. Mizoguchi T, Yoshida R: **Punctual coordination: switching on and off for flowering during a day**. *Plant Signal Behav* 2009, **4**:113-115.
279. Wu Y, Dey R, Han A, Jayathilaka N, Philips M, Ye J, Chen L: **Structure of the MADS-box/MEF2 domain of MEF2A bound to DNA and its implication for myocardin recruitment**. *J Mol Biol* 2010, **397**:520-533.
280. Bowman JL, Smyth DR, Meyerowitz EM: **The ABC model of flower development: then and now**. *Development* 2012, **139**:4095-4098.
281. Melzer R, Verelst W, Theißen G: **The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro**. *Nucleic Acids Res* 2009, **37**:144-157.
282. Melzer R, Theißen G: **Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins**. *Nucleic Acids Res* 2009, **37**:2723-2736.
283. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, 3rd, Laughton CA, Orozco M: **Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers**. *Biophys J* 2007, **92**:3817-3829.
284. Joung IS, Cheatham TE, 3rd: **Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters**. *J Phys Chem B* 2009, **113**:13279-13290.
285. Joung IS, Cheatham TE, 3rd: **Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations**. *J Phys Chem B* 2008, **112**:9020-9041.
286. Huang K, Louis JM, Donaldson L, Lim FL, Sharrocks AD, Clore GM: **Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors**. *EMBO J* 2000, **19**:2615-2628.
287. Acton TB, Mead J, Steiner AM, Vershon AK: **Scanning mutagenesis of Mcm1: residues required for DNA binding, DNA bending, and transcriptional activation by a MADS-box protein**. *Mol Cell Biol* 2000, **20**:1-11.
288. West AG, Sharrocks AD: **MADS-box transcription factors adopt alternative mechanisms for bending DNA**. *J Mol Biol* 1999, **286**:1311-1323.
289. Tan S, Richmond TJ: **Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex**. *Nature* 1998, **391**:660-666.
290. West AG, Shore P, Sharrocks AD: **DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending**. *Mol Cell Biol* 1997, **17**:2876-2887.
291. West AG, Sharrocks AD: **The role of DNA-bending in MADS-box transcription factor function**. *Biochem Soc Trans* 1997, **25**:S639.

292. Acton TB, Zhong H, Vershon AK: **DNA-binding specificity of Mcm1: operator mutations that alter DNA-bending and transcriptional activities by a MADS box protein**. *Mol Cell Biol* 1997, **17**:1881-1889.
293. Riechmann JL, Wang M, Meyerowitz EM: **DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS**. *Nucleic Acids Res* 1996, **24**:3134-3141.
294. Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wuthrich K: **Recommendations for the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy**. *Eur J Biochem* 1998, **256**:1-15.
295. Hepworth SR, Valverde F, Ravenscroft D, Mouradov A, Coupland G: **Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs**. *EMBO J* 2002, **21**:4327-4337.
296. Vuzman D, Levy Y: **DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail**. *Proc Natl Acad Sci U S A* 2010, **107**:21004-21009.
297. Vuzman D, Azia A, Levy Y: **Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails**. *J Mol Biol* 2010, **396**:674-684.
298. Vuzman D, Levy Y: **Intrinsically disordered regions as affinity tuners in protein-DNA interactions**. *Mol Biosyst* 2012, **8**:47-57.
299. Han A, He J, Wu Y, Liu JO, Chen L: **Mechanism of recruitment of class II histone deacetylases by myocyte enhancer factor-2**. *J Mol Biol* 2005, **345**:91-102.
300. Wu RM, Walton EF, Richardson AC, Wood M, Hellens RP, Varkonyi-Gasic E: **Conservation and divergence of four kiwifruit SVP-like MADS-box genes suggest distinct roles in kiwifruit bud dormancy and flowering**. *J Exp Bot* 2012, **63**:797-807.
301. Jimenez S, Lawton-Rauh AL, Reighard GL, Abbott AG, Bielenberg DG: **Phylogenetic analysis and molecular evolution of the dormancy associated MADS-box genes from peach**. *BMC Plant Biol* 2009, **9**:81.
302. Sharrocks AD, von Hesler F, Shaw PE: **The identification of elements determining the different DNA binding specificities of the MADS box proteins p67SRF and RSRFC4**. *Nucleic Acids Res* 1993, **21**:215-221.
303. Sharrocks AD, Gille H, Shaw PE: **Identification of amino acids essential for DNA binding and dimerization in p67SRF: implications for a novel DNA-binding motif**. *Mol Cell Biol* 1993, **13**:123-132.
304. Kogot JM, Sarkes DA, Val-Addo I, Pellegrino PM, Stratis-Cullum DN: **Increased affinity and solubility of peptides used for direct peptide ELISA on polystyrene surfaces through fusion with a polystyrene-binding peptide tag**. *Biotechniques* 2012, **52**:95-102.
305. Kumada Y, Shiritani Y, Hamasaki K, Ohse T, Kishimoto M: **High biological activity of a recombinant protein immobilized onto polystyrene**. *Biotechnol J* 2009, **4**:1178-1189.
306. Kumada Y, Hamasaki K, Shiritani Y, Nakagawa A, Kuroki D, Ohse T, Choi DH, Katakura Y, Kishimoto M: **Direct immobilization of functional single-chain variable fragment antibodies (scFvs) onto a polystyrene plate by genetic fusion of a polystyrene-binding peptide (PS-tag)**. *Anal Bioanal Chem* 2009, **395**:759-765.

307. Kumada Y, Zhao C, Ishimura R, Imanaka H, Imamura K, Nakanishi K: **Protein-protein interaction analysis using an affinity peptide tag and hydrophilic polystyrene plate**. *J Biotechnol* 2007, **128**:354-361.
308. Kumada Y, Katoh S, Imanaka H, Imamura K, Nakanishi K: **Development of a one-step ELISA method using an affinity peptide tag specific to a hydrophilic polystyrene surface**. *J Biotechnol* 2007, **127**:288-299.
309. Kumada Y, Tokunaga Y, Imanaka H, Imamura K, Sakiyama T, Katoh S, Nakanishi K: **Screening and characterization of affinity peptide tags specific to polystyrene supports for the orientated immobilization of proteins**. *Biotechnol Prog* 2006, **22**:401-405.
310. Hassler M, Richmond TJ: **The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex**. *EMBO J* 2001, **20**:3018-3028.
311. Pellegrini L, Tan S, Richmond TJ: **Structure of serum response factor core bound to DNA**. *Nature* 1995, **376**:490-498.
312. Santelli E, Richmond TJ: **Crystal structure of MEF2A core bound to DNA at 1.5 A resolution**. *J Mol Biol* 2000, **297**:437-449.
313. Jayathilaka N, Han A, Gaffney KJ, Dey R, Jarusiewicz JA, Noridomi K, Philips MA, Lei X, He J, Ye J, et al.: **Inhibition of the function of class IIa HDACs by blocking their interaction with MEF2**. *Nucleic Acids Res* 2012, **40**:5378-5388.
314. Han A, Pan F, Stroud JC, Youn HD, Liu JO, Chen L: **Sequence-specific recruitment of transcriptional co-repressor Cabin1 by myocyte enhancer factor-2**. *Nature* 2003, **422**:730-734.

Copyright © Joshua R. Werkman 2013

Vita

Joshua R. Werkman

Education

B.S., Biology, Liberty University, Lynchburg, VA 2002

Professional Positions

Graduate Research Assistant, University of Kentucky, College of Agriculture, Department of Plant and Soil Sciences, Lexington, KY 2004–2013

Research Technician III, University of Pittsburgh, College of Medicine, Department of Medicine, Division of Infectious Diseases, Pittsburgh, PA 2002–2004

Undergraduate Teacher Assistant, Liberty University, College of Arts and Sciences, Department of Biology and Chemistry, Lynchburg, VA 1998–2001

Greenhouse/Laboratory Assistant, Oglevee Ltd., Connellsville, PA 1997–2000

Journal Publications:

Kong, Q., S. Pattanaik, S., A. Feller, **J. R. Werkman**, C. Chai, Y. Wang, E. Grotewold and L. Yuan (2012). “A regulatory switch enforced by bHLH and ACT domain-mediated dimerizations of the maize transcription factor R.” Proc Natl Acad Sci U S A **109**(30): E2091–E2097.

Bai, Y., S. Pattanaik, B. Patra, **J. R. Werkman**, C. H. Xie and L. Yuan (2011). "Flavonoid-related basic helix–loop–helix regulators, NtAn1a and NtAn1b, of tobacco have originated from two ancestors and are functionally active." Planta **234**(2): 363–375.

Pattanaik, S., Q. Kong, D. Zaitlin, **J. R. Werkman**, C. H. Xie, B. Patra and L. Yuan (2010). "Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco." Planta **231**(5): 1061–1076.

Fan, Z., **J. R. Werkman** and L. Yuan (2009). "Engineering of a multifunctional hemicellulase." Biotechnol Lett **31**(5): 751–757.

Book Chapters:

Werkman, J. R., S. Pattanaik and L. Yuan (2011). "Directed evolution through DNA shuffling for the improvement and understanding of genes and promoters." Methods Mol Biol **754**: 325–342.

Pattanaik, S., **J. R. Werkman** and L. Yuan (2011). "Bimolecular fluorescence complementation as a tool to study interactions of regulatory proteins in plant protoplasts." Methods Mol Biol **754**: 185–193.

Pattanaik, S., **J. R. Werkman**, Q. Kong and L. Yuan (2010). "Site-directed mutagenesis and saturation mutagenesis for the functional study of transcription factors involved in plant secondary metabolite biosynthesis." Methods Mol Biol **643**: 47–57.

Posters:

Bai, Y., S. Pattanaik, B. Patra, **J. R. Werkman**, C. H. Xie, and L. Yuan (2010). "Flavonoid-specific basic helix–loop–helix regulators NtAn1a and NtAn1b of tobacco are originated from its two ancestors and functionally active." American Society of Plant Biologists - Plant Biology 2010.

Pattanaik, S., Q. Kong, D. Zaitlin, **J. R. Werkman**, C. H. Xie, and L. Yuan (2009). "The R2R3 MYB Transcription Factor, NtAn2, Is a Key Regulator of Anthocyanin Biosynthesis in Tobacco (Nicotiana tabacum L.)." American Society of Plant Biologists - Plant Biology 2009.

Cole K. S. , J. D. Steckbeck, I. Orlov, K. A. Miller, J. Bruno, **J. R. Werkman**, N. D. Polansky, G. Lin, J. A. Hoxie, and R. C. Montelaro (2004). “Qualitative properties of antibody binding properties associated with neutralization of SIV.” Eleventh Conference on Human Retroviruses and Opportunistic Infections 2004, San Francisco, CA.

Cole K. S., **J. R. Werkman**, N. D. Polansky, J. D. Steckbeck, K. A. Miller, I. Orlov, J. Bruno, G. Lin, J. A. Hoxie, and R. C. Montelaro (2003). “Defining mechanisms of antibody-mediated neutralization of SIV.” XXI Annual Symposium on Nonhuman Primate Models for AIDS 2003, Seattle, WA.

Cole K. S., J. D. Steckbeck, K. A. Miller, I. Orlov, J. Bruno, **J. R. Werkman**, N. D. Polansky, B. A. Puffer, R. W. Doms, and R. C. Montelaro (2003). “Defining mechanisms of antibody-mediated neutralization of SIV.” Tenth Conference on Human Retroviruses and Opportunistic Infections 2003, Boston, MA.