This is to certify that the
dissertation entitled

CHARACTERIZATION OF THE ARABIDOPSIS THALIANA CBF1 TRANSCRIPTION FACTOR: FUNCTIONAL ROLE OF TWO EVOLUTIONARILY CONSERVED SIGNATURE SEQUENCES

presented by

Donatella Canella

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Cell and Molecular Biology

Major Professor's Signature

11/26/07

Date

CHARACTERIZATION OF THE ARABIDOPSIS THALIANA CBF1 TRANSCRIPTION FACTOR: FUNCTIONAL ROLE OF TWO EVOLUTIONARILY CONSERVED SIGNATURE SEQUENCES

By

Donatella Canella

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Program in Cell and Molecular Biology

2007

ABSTRACT

CHARACTERIZATION OF THE ARABIDOPSIS THALIANA CBF1 TRANSCRIPTION FACTOR:
FUNCTIONAL ROLE OF TWO EVOLUTIONARILY CONSERVED SIGNATURE SEQUENCES

By

Donatella Canella

Over the course of evolution plants have adapted to the environment by developing ways to cope with different biotic and abiotic stresses. Among the abiotic stresses, low temperatures represent a major limiting factor for the growth and development of plants. The *CRT/DRE*-Binding Factors (CBFs) are transcriptional activators that are rapidly activated in response to low temperature and in turn induce the expression of a battery of cold-regulated (*COR*) genes to increase plant freezing tolerance. Arabidopsis plants overexpressing *CBF1*, *CBF2* or *CBF3* are constitutively freezing tolerant, indicating that these regulators are master regulators of cold adaptation. CBF proteins belong to the APETALA2/Ethylene-Response Binding Protein (AP2/EREBP) family, which includes 145 members in Arabidopsis. Proteins in this family share high similarity within their AP2/EREBP DNA-binding domain. A unique feature of the CBF proteins is that they contain two conserved sequences flanking the DNA-binding domain. These sequences, represented by the consensus motifs PKK/RPAGRxKFxETRHP and DSAWR, are also found in CBF-like proteins from evolutionarily diverse plant species, suggesting that these "signature" sequences play a role in CBF activity. Overexpression of wild type *CBF1* in Arabidopsis results in constitutive *COR* gene expression. Transgenic lines overexpressing *CBF1* carrying

alanine substitutions in the signature sequences showed reduced or no *COR* gene expression, indicating that these sequences play an important role in CBF1 activity. Analysis of protein levels revealed that alanine mutations throughout the DSAWR motif affect protein accumulation *in planta*, and could explain the lower *COR* gene accumulation in those plants. On the contrary, mutations in the PKKPAGR motif did not affect protein steady state levels; instead they impaired the ability of CBF1 to bind its cognate *CRT/DRE* element from the *COR* gene promoters. The most pronounced effect was observed when two conserved Arg and Phe residues were substituted with Lys and Ala, respectively; these substitutions were sufficient to abrogate DNA binding, indicating an essential role of those residues and potentially a base-specific recognition. Altogether these results indicated that DNA binding activity in the AP2/EREBP family extends beyond the canonical DNA-binding domain previously described to include the N-flanking PKKPAGR region. Based on these observations and secondary structure prediction studies, we developed a computational model describing CBF1 bound to the DNA. According to this model, CBF1 binds the DNA major groove through a three-stranded beta sheet, as described for other AP2/EREBP proteins. In addition, a helical stretch within the PKKPAGR motif makes essential interactions with the DNA minor groove in close proximity to a conserved thymine that is a specificity determinant in the *CRT/DRE* element bound by the CBF proteins. Additional investigations will elucidate whether residues within the PKKPAGR motif represent specificity switches for the recognition of the *CRT/DRE* promoter element by CBF proteins and whether a similar mechanism has been conserved in other protein of the AP2/EREBP family.

ACKNOWLEDGMENTS

I came to the United States to work in the Plant Research Laboratory as a visiting scholar. The plan was to stay one year. Several years later, I prepare to leave with a doctoral degree. Being here has been a great experience, at the professional and personal level, and I am very grateful to the many people that have accompanied me in this amazing journey.

First of all, I want to thank my mentor, Mike Thomashow, for his great guidance and support. I am very grateful for the freedom, trust and encouragement that he has given me all throughout my PhD. I feel privileged to have met him and very fortunate to have worked in his lab. I have learned a lot from our meetings and discussions; these times will always stay with me wherever I'll go.

I would like to thank my committee members: Sheng Yang He, Lee Kroos, Ken Keegstra, and Steve Triezenberg. Thank you for your time and your constructive criticism. Your advice over the years has been critical in the completion of my thesis work and has been an important contribution to my growth as a scientist.

I want to thank Leslie Kuhn, with whom I have collaborated for the last few years of my PhD. It has been a great pleasure to work together; I have greatly benefited from your scientific experience, and your encouraging attitude had made me a more confident scientist.

Thank you to all the present and past members of the Thomashow lab: Sarah Gilmour, Keenan Amundsen, Marcela Carvallo, Diane Constan, Daniel Cook, Colleen Doherty,

Malia Dong, Carri Duncan, Sarah Fowler, Chin-Mei Lee, Michael Mikkelsen, Susan Myers, Kanchan Pavangadkar, Ritu Sharma, Bonnie St.John, Heather Van Buskirk, Jonathan Vogel, Vandana Yadav, Dan Zarka and Xin Zhang. I can't believe I won't be coming to the lab next week. I will miss you.

A very big thank you to all my friends. It's been wonderful to have you in my life. Thanks to each and all of you for being with me during all this time and for making my life so much richer and happier. It's amazing to think of how many little everyday things we have shared; all these times are such a big part of who I am today; there's no doubt in my heart I could not have come so far without you.

A special thank you to Alberto, who has been by my side with his love and support. Thank you for never doubting that I could make it. Your strength, patience and constant presence have been a wonderful gift that will stay with me forever.

A warm thank you to Kathryn and Edward Boucher, my American family. You have been my home away from home. Thank you for the many suppers and for our chats after supper. Your positive and strong spirit has touched me deeply; I will never forget it.

My family has always been there for me. Thank you mom, dad, Robi and Cira. You have been my rock with your endless love, constant presence and encouragement. Thank you for believing in me and being proud of me, no matter what. Thank you very much.

TABLE OF CONTENTS

LIST OF TABLES

Images in this dissertation are presented in color.

LIST OF FIGURES

CHAPTER ONE

LITERATURE REVIEW

Plant growth and development can be greatly affected by low temperatures. Plants differ in their ability to cope with low temperatures. Plants native to tropical regions, such as tomato and rice, are typically very sensitive to low temperatures and will suffer chilling injuries when temperatures drop to the range of 0-12°C. In contrast, plants originating from temperate regions, such as the model plant *Arabidopsis thaliana*, rye and canola, are generally more resistant, and are not only chilling tolerant but also freezing tolerant (Sakai and Larcher, 1987). The ability of plants to tolerate freezing can be significantly enhanced by pre-exposure to low non-freezing temperatures, an adaptive process called cold acclimation. For instance, upon cold acclimation wheat can increase freezing tolerance from -5°C to -20°C (Thomashow, 1998).

The events occurring inside a plant upon cold acclimation reflect a complex network of changes that the plant mounts against freeze-induced damage (Levitt, 1980; Thomashow, 1999). These include changes at the physiological, biochemical and transcriptional level. In this chapter, I will summarize the advances in the field of cold acclimation, with specific focus on the CBF family of transcriptional regulators and their role in the CBF cold response pathway in the model plant *Arabidopsis*.

Freeze-induced cell damage and role of cold acclimation.

The negative effects of freezing temperatures on cell survival have been long known (Levitt, 1980). The plasma membrane is the primary target of freeze-induced cell

damage. This injury is mainly caused by freeze-induced dehydration, which occurs in the cell upon exposure to sub-zero temperatures. The first event following freezing inside the cell is ice formation in the extracellular space; this is due to the presence of ice nucleating agents and lower solute concentration in the extracellular space relative to the intracellular space. This process can cause membrane damage due to adhesion of membranes. However, the most severe damage results from freeze-induced dehydration. When ice crystals form at sub-zero temperatures, the chemical potential in the outer space drops, and causes an outward movement of intracellular water to the outer space. The water loss can be as severe as 90% at approximately -10°C (Thomashow, 1999). The type of injury at the membrane site varies depending on the freezing temperatures: freeze-thaw cycles between -2°C to -4°C can cause expansion-induced cell lysis; a temperature range of -4°C to -10°C typically causes phase transition of bilayer lipids from lamellar to hexagonal II; in the most severe cases (below -10°C), fracture jump lesions will arise (Steponkus and Webb, 1992).

Membrane damage induced by freezing temperatures is greatly reduced upon cold acclimation. The protective role of cold acclimation has been under scrutiny for many years, and includes a plethora of changes at the biochemical, physiological and transcriptional levels. Studies by several investigators have established that one of the roles of cold acclimation is to prevent cellular damage by stabilizing membranes. Changes in membrane lipid composition are among the first and best documented events that underline cold acclimation (Thomashow, 1999). Additional metabolic changes also accompany cold acclimation, including the accumulation of small cryoprotective

molecules such as soluble sugars and proline (Rudolph and Crowe, 1985; Strauss and Hauser, 1986; Carpenter and Crowe, 1988).

Changes in gene expression underline the cold acclimation response.

In more recent years, changes in gene expression have been correlated to adaptation to low temperatures. The first direct evidence that cold acclimation was accompanied by changes in gene expression was presented by Guy and colleagues (1985), who reported that differential pools of translatable mRNAs accumulate in non-acclimated and acclimated spinach leaves (Guy *et al.*, 1985). This discovery was followed by the identification of numerous cold-responsive genes in other plant species, including *Brassica napus* (Johnson-Flanagan and Singh, 1987), alfalfa (Mohapatra *et al.*, 1987) and *Arabidopsis* (Gilmour *et al.*, 1988; Kurkela *et al.*, 1988). Those genes were given names such as low temperature-induced genes (*LTI*), cold-regulated genes (*COR*), *KIN* (cold-induced), or *RD* (responsive to dehydration) (reviewed in Thomashow, 1999). Analysis of their expression patterns revealed a positive correlation between their transcript accumulation in response to low temperatures and the level of freezing tolerance achieved by plants (Guy and Haskell, 1987; Mohapatra *et al.*, 1987). Despite their highly divergent sequences, the proteins encoded by these genes share a set of properties that suggest that they might share a common mechanism of action in protecting the cell against freezing. Indeed, some of those properties have been previously described for cryoprotectant molecules, and include high hydrophilicity, presence of amphipathic helices, and solubility in aqueous buffers at high temperatures (Volger and Heber, 1975).

The observation that changes in gene expression accompany cold acclimation suggested that such changes might represent the initial signal regulating the events required to achieve freezing tolerance. However, overexpression of *COR15a* in *Arabidopsis* had little effect on overall plant freezing tolerance, despite its significant role in protecting the chloroplast envelope (Steponkus *et al.*, 1998). This observation indicated that individual *COR* genes were not sufficient to trigger the whole set of changes induced by cold acclimation. It was soon hypothesized that an upstream regulator of gene expression was responsible for the activation of a whole set of *COR* genes. Consistent with this idea, characterization of a battery of *COR* genes from *Arabidopsis* suggested that they might be coordinatively regulated in response to low temperatures. Consistent with this idea, *COR6.6*, *COR15a*, *COR47*, and *COR78* expression followed similar kinetics; they were upregulated as early as four hours after a treatment at 4°C, reached a peak at about 24 hours and remained upregulated as long as the plants were kept under inducing conditions (Hajela *et al.*, 1990) (Horvath *et al.*, 1993). Furthermore, analysis of *COR* gene promoters unveiled the presence of a common *cis*-regulatory element, named C-repeat/Drought Responsive Element *CRT/DRE* (Baker *et al.*, 1994; Yamaguchi-Shinozaki and Shinozaki, 1994). Similar induction patterns and the presence of a conserved element in their promoters suggested that the *COR* genes might be regulated by a common factor.

The CBF cold-response pathway in *Arabidopsis thaliana*.

Discovery of the CBF pathway.

A major breakthrough in the field of cold acclimation was the discovery of the transcriptional regulator *CRT/DRE* Binding Factor1 (CBF1) (Stockinger *et al.*, 1997) and its key role in regulating cold acclimation in *Arabidopsis*. CBF1 was identified in a yeast-one-hybrid screen in which the *CRT/DRE* promoter element was used as bait. CBF1 could specifically bind this *cis*-element *in vitro* and induce the expression of a *COR15a::lacZ* reporter gene in yeast (Stockinger *et al.*, 1997). *Arabidopsis* plants overexpressing *CBF1* constitutively expressed the whole battery of *COR* genes and displayed constitutive freezing tolerance similar to freezing tolerance of cold-acclimated non-transgenic plants, indicating that this transcription factor is a master regulator of the cold acclimation response in *Arabidopsis* (Jaglo-Ottosen *et al.*, 1998). The discovery of CBF1 and its role in cold acclimation was significant because it provided the first evidence that genetic manipulation was indeed possible to generate more freezing tolerant plants, whereas years of breeding attempts had resulted in only modest improvement of freezing tolerance. Furthermore, it revealed that the whole set of changes necessary to induce freezing tolerance could be induced by overexpressing a single gene (Thomashow, 1999).

Shortly after, it was discovered that *Arabidopsis* has three *CBF* genes, *CBF1*, *CBF2*, and *CBF3* – also known as *DREB1B*, *DREB1C*, and *DREB1A*, respectively - arranged in tandem array on chromosome IV (Gilmour *et al.*, 1998; Liu *et al.*, 1998; Gilmour *et al.*, 2000). They follow a similar induction pattern in response to low temperatures, whereby they are induced within 10-15 minutes of cold treatment, peak at 2-3 hours, and return to a basal expression level by 24 hours of cold treatment. Overexpression of *CBF1*, *CBF2*, or *CBF3* in *Arabidopsis* rendered plants constitutively

freezing tolerant (Jaglo-Ottosen *et al.*, 1998; Liu *et al.*, 1998; Gilmour *et al.*, 2000; Gilmour *et al.*, 2004) and induced constitutive expression of the same pool of genes, indicating that the three transcriptional activators have matching transcriptional activities. Three additional *CBF* homologs were identified more recently. *CBF4/DREB1D* (Haake *et al.*, 2002; Sakuma *et al.*, 2002), was described as a regulator of drought response in *Arabidopsis*, in that its expression was up-regulated by drought stress, but not by low temperature. Its overexpression in *Arabidopsis* resulted in constitutive accumulation of *CRT/DRE*-containing genes, and conferred to plants both drought and freezing tolerance, similarly to what observed for overexpression of the other *CBF*s (Haake *et al.*, 2002). This result is not surprising, given the physiological overlap between dehydration and freezing stress. *DDF1/DREB1F* and *DDF2/DREB1E* (Sakuma *et al.*, 2002; Magome *et al.*, 2004) also show high sequence homology to the CBF family, but little is known on their role in response to abiotic stresses. *DDF1/DREB1F* expression is induced upon salinity stress (Sakuma *et al.*, 2002; Magome *et al.*, 2004), similarly to what happens to CBF1-3, suggesting that these genes may display overlapping activity. The role of *DDF1/DREB1F* and *DDF2/DREB1E* in low temperature response, however, has not been elucidated.

Predominant role of CBF in configuring low temperature responses in Arabidopsis.

Technological advances and the availability of whole genome sequences have made it possible to conduct high-throughput experiments to identify the molecular changes at the transcriptional and post-transcriptional level occurring in response to low temperature.

Microarray technology has proven a powerful tool for surveying global gene expression in *Arabidopsis* in response to cold treatment. Transcriptome profiling studies using cDNA microarrays and Affymetrix GeneChips have been very useful to identify novel genes, determine their kinetics of induction, and help delineate the complex network of cascades that are configured. These analyses have revealed that hundreds of genes are regulated during cold acclimation in *Arabidopsis* (Seki *et al.*, 2001; Fowler and Thomashow, 2002; Kreps *et al.*, 2002; Seki *et al.*, 2002; Gilmour *et al.*, 2004; Maruyama *et al.*, 2004; Vogel *et al.*, 2005). Among the up-regulated genes are several transcription factors, some of which are induced in parallel with CBF. An important conclusion from these studies has been that additional cold response pathways exist besides the *CBF* pathway (Seki *et al.*, 2001; Fowler and Thomashow, 2002; Kreps *et al.*, 2002; Seki *et al.*, 2002). This is consistent with previous reports of *Arabidopsis* mutants that are constitutively more freezing tolerant than wild type plants despite the fact that *COR* genes are not induced. One example of a CBF-independent response is represented by the *eskimo1 Arabidopsis* mutant (Xin and Browse, 1998). *eskimo1* plants display constitutive freezing tolerance and yet do not show constitutive expression of the *CBF* regulon (Xin and Browse, 1998; Xin *et al.*, 2007). Indeed, transcriptome profiling analysis showed an overlap of about 12% in gene transcripts between the *eskimo1* mutants and *CBF2*-overexpressing plants (Xin *et al.*, 2007). Interestingly, *eskimo1* plants are not drought tolerant, which makes them a valuable tool for understanding the players and mechanisms that are specific for the cold response in *Arabidopsis*. Similarly to the *eskimo1* mutation, *ada2b* mutant plants in *Arabidopsis* are constitutively more freezing tolerant than wild type plants, and yet they do not show constitutive *COR* gene

expression (Vlachonasios *et al.*, 2003); therefore ADA2b is implicated in a CBF-independent pathway.

Despite the presence of additional cold response pathways, the CBF family plays a predominant role in configuring the changes observed when plants respond to cold. Comparison between genes that are cold- and *CBF*-responsive has helped define the *CBF* regulon as the pool of approximately 100 genes that are regulated by both low temperature and CBF (Fowler and Thomashow, 2002; Vogel *et al.*, 2005). Eighty five genes in the *CBF* regulon are up-regulated and account for almost 30% of the transcripts that accumulate in response to low temperatures. In addition, members of the *CBF* regulon are among the most highly induced upon cold treatment (Vogel *et al.*, 2005). For instance, 84% of the cold-responsive genes that are up-regulated more than 15 fold and 50% of the genes that are up-regulated between 5-10 fold belong to the *CBF* regulon.

Metabolite profiling in *Arabidopsis* has enabled the analysis of changes occurring at low temperature to changes in small organic compounds associated with low temperature response. The role of CBF in mediating metabolic changes that occur upon cold exposure was monitored by metabolomic profiling of wild type *Arabidopsis* plants and plants overexpressing *CBF3*. Major rearrangements were detected during cold acclimation, with 75% of the metabolites analyzed showing changes. Of those changes, the majority (79%) could be attributed to *CBF3* overexpression (Cook *et al.*, 2004). Therefore, the *CBF* pathway plays a major role in configuring the changes that are needed for plants to adapt to low temperatures, both at the transcriptional and post-transcriptional level.

Upstream events in the regulation of the *CBF* pathway.

Much of the research effort in the field of cold acclimation has focused on elucidating the upstream events regulating the induction of *CBF* genes. Reporter gene analyses and mutant screens have proven very valuable to identify novel genes that play important roles during cold acclimation.

Promoter analysis of *CBF2* unveiled a minimal 125-bp promoter region that is cold responsive (Zarka *et al.*, 2003). Within this promoter fragment, two regions named ICEr1 and ICEr2 (induction of *CBF* expression region 1 and 2) cooperated in directing cold regulation of *CBF* expression (Zarka *et al.*, 2003). ICEr1 contains a *cis*-acting element which is recognized by ICE1 (Inducer of *CBF* Expression 1). ICE1 was isolated through a mutant screen of plants that were defective in *CBF3* cold-induction; however, *CBF1* or *CBF2* expression was not affected, suggesting that one of the other 139 bHLH proteins present in *Arabidopsis* might be involved in the specific recognition of this element at low temperatures (Chinnusamy *et al.*, 2003). Interestingly, overexpression of ICE1 is not sufficient to confer constitutive freezing tolerance, but a cold stimulus is required. Presumably, post-translational mechanisms or the presence of a co-activator must occur to activate ICE1 (Chinnusamy *et al.*, 2003). More recently, it was discovered that a transcriptional regulator member of the MYB family, MYB15, can bind to the promoter of *CBF1*, *CBF2*, and *CBF3*, and physically interacts with ICE1 *in vitro* (Agarwal *et al.*, 2006). MYB15 overexpression negatively regulates *CBF* expression in the cold and is most evident in the early stages of induction (approximately 3 hours). However, MYB15 overexpression does not significantly influence *COR* gene expression (Agarwal *et al.*, 2006). The significance of these results remains unclear at the present.

Recently, Vogel *et al.* (2005) have characterized ZAT12, a zinc-finger transcription factor that appears to be coordinately regulated with *CBF2* in response to low temperature. Interestingly, *ZAT12* overexpression in *Arabidopsis* has a negative effect on cold-regulated induction of the CBF genes, whereas in the *zat12* knock-out lines *CBF* expression is induced. Presumably this repressive activity is mediated by an EAR-like motif present in ZAT12 (Ohta *et al.*, 2001). At the same time, *ZAT12* induces the expression of a *ZAT12* regulon, which partly overlaps with the *CBF* regulon, suggesting that CBF and ZAT12 might cooperate to induce expression of cold responsive genes (Vogel *et al.*, 2005).

It is evident from the recent advances that the regulation *CBF* expression during cold stress is complex and involves many factors. Identification of additional components and a better understanding of their mechanisms of action will help gain a more detailed picture of how transcriptional networks are regulated in response to low temperatures.

Conservation of the *CBF* pathway in other plant species.

A major interest in the field of cold acclimation has been to determine whether the *CBF* pathway has been conserved in other plant species. This is of particular interest for agronomical application to crop plants. CBF-like genes have been identified in a wide variety of plant species, including both chilling sensitive and freezing tolerant plants. A central question is to determine whether these represent functional homologs of the *Arabidopsis* CBFs, and whether any differences in the CBF pathway of these plants can explain different tolerance to low temperatures.

Among freezing tolerant plants, *CBF*-like genes have been identified in *Brassica napus* (Jaglo *et al.*, 2001), poplar (Benedict *et al.*, 2006), barley (Skinner *et al.*, 2005), maize (Qin *et al.*, 2004) and others. Some of these CBFs have been shown to be functional homologs of the *Arabidopsis* CBFs based on their sequence similarity, transcriptional activity, and the ability to enhance freezing and drought tolerance when overexpressed. Furthermore, a number of *CRT/DRE*-containing genes present in those plants and their transcripts can be constitutively accumulated when CBF is overexpressed (Choi *et al.*, 1999; Dal Bosco *et al.*, 2003).

Chilling sensitive plants also contain CBF-like genes. Some of these genes represent functional homologs of the *Arabidopsis CBF*s, since they are cold-responsive and can induce constitutive expression of a *CBF* regulon when overexpressed in *Arabidopsis*, and results in increased tolerance to freezing and dorught. Some examples include *CBF*s from tomato (Hsieh *et al.*, 2002; Zhang *et al.*, 2004) and rice (Dubouzet *et al.*, 2003). The tomato CBF locus includes three CBF genes (Zhang *et al.*, 2004). *LeCBF1* is cold-responsive and can induce constitutive expression of a *CBF* regulon and impart freezing tolerance when overexpressed in *Arabidopsis*. Tomato plants overexpressing *LeCBF1* or *AtCBF1* can also activate a *CBF* regulon. However, microarray analysis of a quarter of the tomato genome indicates that, despite the fact that similar classes of proteins are induced in the two plants, transgenic tomato plants can only slightly increase their chilling tolerance (Zhang *et al.*, 2004). Current efforts are in place to determine whether the inability of chilling sensitive plants to show a functional CBF response can be attributed to smaller and consequently less diverse CBF regulons or additional defects that limit CBF function.

Mechanisms of transcriptional activation in the CBF family of proteins.

The CBF family of proteins.

CBF proteins are members of the AP2/EREBP family of transcription factors (Riechmann and Meyerowitz, 1998). Members of this family are widespread among plants, but a few examples have been recently found in other organisms such as bacteria and bacterial viruses (Magnani *et al.*, 2004; Wuitschick *et al.*, 2004; Balaji *et al.*, 2005). AP2/EREBP proteins play a variety of roles in plant growth and development as well as in biotic and abiotic stress responses (Riechmann and Meyerowitz, 1998). A common feature of these proteins is the presence of a conserved AP2/EREBP DNA-binding domain. The *Arabidopsis* AP2/EREBP family includes 145 predicted proteins sharing high homology within this domain. Based on the sequence similarity within the DNA-binding domain, Sakuma *et al.* (2002) have divided the *Arabidopsis* AP2/EREBP family into 5 different groups: the DREBs (56 members); the ERFs (65 members); the AP2s (14 members); the RAVs (6 members); and a fifth group (4 members). The DREB group includes the CBF family. CBF proteins are characterized by an N-terminal domain including 32 amino acids of unknown function, the AP2/EREBP DNA-binding domain, and a C-terminal domain resembling acidic activation domains from other transcriptional activators such as the herpesvirus VP16, the mammalian p53 protein and RelA (Triezenberg, 1995). A distinctive feature of the CBF family in *Arabidopsis* is the presence of two conserved signature sequences flanking the DNA-binding domain, defined by the consensus PKKP/RAGRxKFxETRHP and DSAWR (Jaglo *et al.*, 2001). Strikingly, these sequences have been highly conserved throughout evolution, and they can be found in very diverse plant species (Jaglo *et al.*, 2001). This observation suggests

that they might play an important role in mediating CBF transcriptional activity. The functional characterization of these motifs will be presented in this dissertation.

DNA-binding properties of CBF proteins and other AP2/EREBP family members.

The DNA-binding activity of the AP2/EREBP proteins was first described by Ohme-Takagi and Shinshi (1995). They reported that a 59-amino acid region that was highly conserved among four ethylene-binding proteins was required for specific binding to their target DNA, called the GCC box (Hart *et al.*, 1993; Ohme-Takagi and Shinshi, 1995; Sato *et al.*, 1996). Shortly after, additional transcription factors with the same conserved domain were identified, and the conserved region named AP2/EREBP DNA-binding domain (Riechmann and Meyerowitz, 1998). One of the most significant contributions towards characterizing the binding activity of the AP2/EREBP proteins has been the NMR analysis of this domain from AtERF1 bound to its target *cis*-regulatory element. This domain is defined by a three stranded β-sheet followed by a C-terminal helix; the β-sheet is packed against the major groove of the DNA and makes specific contacts through seven amino acids, mainly arginines and tryptophans (Allen *et al.*, 1998). The degree of similarity among the *Arabidopsis* AP2/EREBP proteins ranges from approximately 60 to 90% and six of the seven amino acids involved in direct DNA contact in ERF1 are conserved in the AP2/EREBP proteins (Hao *et al.*, 2002). The high degree of conservation implies that these domains fold similarly and might make similar DNA contacts.

Very little information is available on the DNA binding preferences within the AP2/EREBP family. One of the reasons is that the target *cis*-acting elements have been

identified only for a few AP2/EREBP proteins, including ABI4, ORCA, CBFs, DREB2s, and ERFs (Stockinger *et al.*, 1997; Allen *et al.*, 1998; Liu *et al.*, 1998; van der Fits and Memelink, 2001; Acevedo-Hernandez *et al.*, 2005). Most of the current knowledge on the binding properties of the AP2/EREBP family has been derived from the analysis of the ERF and CBF proteins. The *cis*-acting elements targeted by the CBF and ERF proteins have long been known and are represented by the *CRT/DRE* and *GCC* box promoter elements, respectively (Baker *et al.*, 1994; Yamaguchi-Shinozaki and Shinozaki, 1994; Ohme-Takagi and Shinshi, 1995). The specificity determinants in these two *cis*-acting elements are defined by the two core sequences A/GCCGACNT and AGCCGCC for the CBF and ERF family, respectively (Hao *et al.*, 1998; Hao *et al.*, 2002; Maruyama *et al.*, 2004). Much remains to be understood in terms of specificity switches within proteins of the two subfamilies. Most of the work has focused on elucidating the role of amino acids within the AP2/EREBP domain that differ between ERFs and CBFs. One of the most important findings is that an alanine residue that is highly conserved in the ERF subfamily can impart specific sequence recognition to proteins in this subgroup. In fact, when alanine is substituted with the corresponding valine residue from the CBF proteins, the mutated ERF alters its binding specificity and can recognize the *CRT/DRE* promoter element (Hao *et al.*, 2002). However, there is no evidence indicating the corresponding mutation in the AP2/EREBP domain of CBF can modify its binding preference. Sakuma *et al.* (2002) have proposed that two conserved amino acids within the AP2/EREBP domain, which differ between ERFs and CBFs, function as specificity determinants. The authors showed that swapping of the residues between the two groups of proteins could abrogate binding to their respective cis-acting elements; however, there is no direct

evidence that residue swapping can drive CBF proteins to the ERF promoter. Based on the current studies, it is unclear whether additional residues are essential for DNA binding within the AP2/EREBP family. Hao *et al.* (1998) observed that short stretches of amino acids flanking the DNA-binding domain of several ERF proteins are important for binding to the GCC box. Since the flanking regions of ERF proteins share little or no similarity, it was suggested that the role of these residues is to stabilize the protein-DNA complex (Hao *et al.*, 1998).

Characterization of CBF1 trans-activating properties.

Despite the fact that CBF1 was identified a decade ago, not much is known on the mechanism whereby CBF proteins activate *COR* gene expression. The C-terminal domain of CBF ($CBF1_{116-213}$) functions as the activation domain in both yeast and Arabidopsis (Wang *et al.*, 2005). Within this domain are several hydrophobic clusters; each of them provides some functional redundancy, presumably to maintain efficient activation of the CBF regulon under cold stress conditions (Wang *et al.*, 2005).

More recently, investigators have explored the role of chromatin-modifying factors in facilitating CBF trans-activating properties. The CBF1 activation domain is acidic in nature, similar to that of other transcription factors such as VP16 and Gcn4 (Hope and Struhl, 1986; Seipel *et al.*, 1994). These proteins stimulate transcription in part by recruitment of chromatin-modifying complexes such as Ada or SAGA (Kuo *et al.*, 1998; Ikeda *et al.*, 1999). The transcriptional adaptors ADA2 and ADA3 and the histone acetyltransferase GCN5 are part of these large multi-protein complexes (Grant *et al.*, 1997). They were shown to be essential for CBF1 transcriptional activity in yeast

(Stockinger *et al.*, 2001). More recently, it was discovered that CBF1 can also interact *in vitro* with AtGCN5 and AtADA2b, the *Arabidopsis* homologues of GCN5 and ADA2 from yeast (Mao *et al.*, 2006). Contrary to what expected, these interaction occur through the DNA-binding domain of CBF1, and not through the activation domain. Furthermore, they are not specific to CBF1 but appear to be conserved within the AP2/EREBP family, as a similar interaction occurs with another member of the family, TINY (Wilson *et al.*, 1996; Mao *et al.*, 2006). To better understand the role of AtGCN5 and AtADA2b *in vivo*, T-DNA insertion mutations for the two genes were analyzed. Disruption of *AtGCN5* and *AtADA2b* affects cold-regulated induction of the *CBF* regulon in *Arabidopsis*, by reducing the level of *COR* gene expression and delaying the kinetics of induction (Vlachonasios *et al.*, 2003). Taken together, the emerging evidence is that these two adaptor proteins might be recruited by CBF1 to mediate *COR* gene induction.

In the effort to identify components of the cold acclimation response in *Arabidopsis*, a set of mutants sensitive to freezing was isolated (McKown *et al.*, 1996). In one of those mutants, *sfr6*, cold-induced activation of *CRT/DRE*-containing genes was affected in one of the steps that follow *CBF1* induction (Knight *et al.*, 1999). *CBF1* transcript still accumulated in response to low temperatures, yet *COR* gene expression was compromised (Knight *et al.*, 1999; Boyce *et al.*, 2003). The nature of the *sfr6* mutation is still unknown, and thus the contribution of this factor to the *CBF* pathway has not been elucidated. Given that the CBF transcript is unaffected in *sfr6* plants, the role of the SFR6 protein is post-transcriptional. Among other explanations, it is reasonable that SFR6 might act as a co-factor of CBF.

Additional post-transcriptional events regulating CBF activity.

One way to modulate the activity of transcription factors is by actively controlling their cellular localization (Tran and Wente, 2006). Whether nuclear shuttling of CBF1 is dependent on a nuclear localization signal (NLS), and whether this is a low temperature-regulated process is not clear. It is expected that some CBF1 protein can accumulate in the nucleus without low temperature treatment, since *Arabidopsis* lines overexpressing *CBF1*, *CBF2* or *CBF3* can accumulate high levels of *COR* gene transcripts compared to the non transgenic plants in the absence of a cold stimulus (Jaglo-Ottosen *et al.*, 1998; Gilmour *et al.*, 2000). On the other hand, when *CBF*-overexpressing plants are cold treated, *COR* gene expression increases (Gilmour *et al.*, 2000). Among other reasons, this may be due to more efficient nuclear transport. Recent studies focused on the subcellular localization of several members of the AP2/EREBP family have revealed that nuclear targeting within this family can be either constitutive or stimulus-dependent. For instance, localization studies on AINTEGUMENTA (Klucher *et al.*, 1996) have indicated that nuclear targeting of this AP2/EREBP protein is constitutive and depends on the presence of specific basic residues (Krizek and Sulli, 2006). On the contrary, the cytokinin response factors CRFs accumulate in the nucleus in response to cytokinin treatment (Rashotte *et al.*, 2006). Moreover, a high-throughput localization study of Arabidopsis proteins has identified an AP2/EREBP protein (id. At2g22880) that can be found both in the nucleus and in the cytoplasm and four that are constitutively nuclear localized (id. At1g13260, At2g39250, At2g33710) under standard growth conditions (Koroleva *et al.*, 2005). Altogether these studies indicate that nuclear localization of proteins in the AP2/EREBP family is regulated differently, and a case by

case study is required to understand the mechanisms dictating nuclear import of each protein.

Based on its similarity to known nuclear localization motifs found in other plant and non-plant transcription factors, the PKKPAGR motif has been proposed to be the nuclear targeting signal of CBF1 (Stockinger *et al.*, 1997). Whether this motif is required and/or sufficient for nuclear import of CBF proteins remains to be determined.

Conclusion

Since the discovery of CBF1 in *Arabidopsis*, several lines of evidence have suggested that additional cold response pathways are present. While it is important to define these alternative pathways, furthering our knowledge of the CBF action mechanisms could better our understanding of how these novel pathways are activated and promote freezing tolerance in plants. The relevance of the CBF pathway has been emphasized over the recent years by large-scale analyses, such as transcriptome and metabolome profiling, and by numerous reports showing that this pathway is conserved across diverse plants species. Understanding how CBF proteins are activated and regulate the changes that lead to freezing tolerance will help define some of the basic mechanisms that have been conserved not only the CBF pathway but also in novel cold-responsive pathways, and it will ultimately help establish how plants sense and respond to low temperatures.

To contribute to this effort, in dissertation I will discuss the functional characterization of the *Arabidopsis* CBF1 protein, by describing the functional properties of two conserved domains in the *CBF* family of transcriptional activators.

LITERATURE CITED

Acevedo-Hernandez, G.J., Leon, P., and Herrera-Estrella, L.R. (2005). Sugar and ABA responsiveness of a minimal RBCS light-responsive unit is mediated by direct binding of ABI4. The Plant Journal **43,** 506-519.

Agarwal, M., Hao, Y., Kapoor, A., Dong, C.-H., Fujii, H., Zheng, X., and Zhu, J.-K. (2006). A R2R3 type MYB transcription factor is involved in the cold regulation of *CBF* genes and in acquired freezing tolerance. J. Biol. Chem. **281,** 37636-37645.

Allen, M., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998). A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. EMBO J **17,** 5484-5496.

Baker, S., Wilhelm, K., and Thomashow, M. (1994). The 5'-region of *Arabidopsis thaliana* cor15a has *cis*-acting elements that confer cold-, drought- and ABA-regulated gene expression. Plant Mol Biol **24,** 701-713.

Balaji, S., Babu, M., Iyer, L., and Aravind, L. (2005). Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucl. Acids Res. **33,** 3994-4006.

Benedict, C., Skinner, J.S., Meng, R., Chang, Y., Bhalerao, R., Huner, N.P.A., Finn, C.E., Chen, T.H.H., and Hurry, V. (2006). The CBF1-dependent low temperature signalling pathway, regulon and increase in freeze tolerance are conserved in Populus spp. Plant, Cell & Environment **29,** 1259-1272.

Boyce, J., Knight, H., Deyholos, M., Openshaw, M., Galbraith, D., Warren, G., and Knight, M. (2003). The *sfr6* mutant of *Arabidopsis* is defective in transcriptional activation via CBF/DREB1 and DREB2 and shows sensitivity to osmotic stress. The Plant Journal **34,** 395-406.

Carpenter, J.F., and Crowe, J.H. (1988). The mechanism of cryoprotection of proteins by solutes. Cryobiology **25,** 244.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B.-h., Hong, X., Agarwal, M., and Zhu, J.-K. (2003). ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev. **17,** 1043-1054.

Choi, D., Zhu, B., and Close, T. (1999). The barley (Hordeum vulgare L.) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv Dicktoo. Theor. Appl. Genet. **98,** 1234–1247.

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M.F. (2004). A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. PNAS **101,** 15243-15248.

Dal Bosco, C., Busconi, M., Govoni, C., Baldi, P., Stanca, A.M., Crosatti, C., Bassi, R., and Cattivelli, L. (2003). *cor* gene expression in barley mutants affected in chloroplast development and photosynthetic electron transport. Plant Physiol **131,** 793–802.

Dubouzet, J.G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E.G., Miura, S., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. The Plant Journal **33,** 751-763.

Fowler, S., and Thomashow, M.F. (2002). *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell **14,** 1675-1690.

Gilmour, S.J., Hajela, R.K., and Thomashow, M.F. (1988). Cold Acclimation in *Arabidopsis thaliana*. Plant Physiol. **87,** 745-750.

Gilmour, S.J., Fowler, S.G., and Thomashow, M.F. (2004). *Arabidopsis* transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. Plant Mol Biol **54,** 768-781.

Gilmour, S.J., Sebolt, A.M., Salazar, M.P., Everard, J.D., and Thomashow, M.F. (2000). Overexpression of the *Arabidopsis* CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant Physiol. **124,** 1854-1865.

Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M., and Thomashow, M.F. (1998). Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. The Plant Journal **16,** 433-442.

Grant, P.A., Duggan, L., Cote, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., Berger, S.L., and Workman, J.L. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev. **11,** 1640-1650.

Guy, C.L., and Haskell, D. (1987). Induction of freezing tolerance in spinach is associated with the synthesis of cold acclimation induced proteins. Plant Physiol. **84,** 872-878.

Guy, C.L., Niemi, K.J., and Brambl, R. (1985). Altered gene expression during cold acclimation of spinach. PNAS **82,** 3673-3677.

Haake, V., Cook, D., Riechmann, J.L., Pineda, O., Thomashow, M.F., and Zhang, J.Z. (2002). Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiol. **130,** 639-648.

Hajela, R.K., Horvath, D.P., Gilmour, S.J., and Thomashow, M.F. (1990). Molecular cloning and expression of cor (Cold-Regulated) genes in *Arabidopsis thaliana.* Plant Physiol. **93,** 1246-1252.

Hao, D., Ohme-Takagi, M., and Sarai, A. (1998). Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J Biol Chem **273,** 26857-26861.

Hao, D., Yamasaki, K., Sarai, A., and Ohme-Takagi, M. (2002). Determinants in the sequence specific binding of two plant transcription factors, CBF1 and NtERF2, to the DRE and GCC motifs. Biochemistry **41,** 4202-4208.

Hart, C., Nagy, F., and Meins, F. (1993). A 61 bp enhancer element of the tobacco beta-1,3-glucanase B gene interacts with one or more regulated nuclear proteins. Plant Mol Biol **21,** 121-131.

Hope, I.A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of Yeast. Cell **46,** 885.

Horvath, D.P., McLarney, B.K., and Thomashow, M.F. (1993). Regulation of *Arabidopsis thaliana* L. (Heyn) cor78 in response to low temperature. Plant Physiol. **103,** 1047-1053.

Hsieh, T.-H., Lee, J.-T., Yang, P.-T., Chiu, L.-H., Charng, Y.-y., Wang, Y.-C., and Chan, M.-T. (2002). Heterology expression of the *Arabidopsis* C-Repeat/Dehydration Response Element Binding Factor 1 gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. Plant Physiol. **129,** 1086-1094.

Ikeda, K., Steger, D.J., Eberharter, A., and Workman, J.L. (1999). Activationd domain-specific and general transcription stimulation by native histone acetyltransferase complexes. Mol. Cell. Biol. **19,** 855-863.

Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O., and Thomashow, M.F. (1998). *Arabidopsis* CBF1 overexpression induces *COR* genes and enhances freezing tolerance. Science **280,** 104-106.

Jaglo, K.R., Kleff, S., Amundsen, K.L., Zhang, X., Haake, V., Zhang, J.Z., Deits, T., and Thomashow, M.F. (2001). Components of the *Arabidopsis* C-Repeat/Dehydration-Responsive Element Binding Factor cold-response pathway are conserved in *Brassica napus* and other plant species. Plant Physiol. **127,** 910-917.

Johnson-Flanagan, A.M., and Singh, J. (1987). Alteration of gene expression during the induction of freezing tolerance in *Brassica napus* suspension cultures. Plant Physiol. **85,** 699-705.

Klucher, K.M., Chow, H., Reiser, L., and Fischer, R.L. (1996). The AINTEGUMENTA gene of *Arabidopsis* required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. Plant Cell **8,** 137-153.

Knight, H., Veale, E.L., Warren, G.J., and Knight, M.R. (1999). The *sfr6* mutation in *Arabidopsis* suppresses low-temperature induction of genes dependent on the *CRT/DRE* sequence motif. Plant Cell **11,** 875-886.

Koroleva, O., Tomlinson, M., Leader, D., Shaw, P., and Doonan, J. (2005). High-throughput protein localization in *Arabidopsis* using *Agrobacterium*-mediated transient expression of GFP-ORF fusions. Plant J **41,** 162-174.

Kreps, J.A., Wu, Y., Chang, H.-S., Zhu, T., Wang, X., and Harper, J.F. (2002). Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. Plant Physiol. **130,** 2129-2141.

Krizek, B., and Sulli, C. (2006). Mapping sequences required for nuclear localization and the transcriptional activation function of the *Arabidopsis* protein AINTEGUMENTA. Planta **224,** 612-621.

Kuo, M.-H., Zhou, J., Jambeck, P., Churchill, M.E.A., and Allis, C.D. (1998). Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. Genes Dev. **12,** 627-639.

Kurkela, S., Franck, M., Heino, P., Lang, V., and Palva, E. (1988). Cold induced gene expresssion in *Arabidopsis thaliana* L. Plant Cell Reports **7,** 495-498.

Levitt, J. (1980). Responses of plants to environmental stresses. (New York: Academic Press).

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell **10,** 1391-1406.

Magnani, E., Sjolander, K., and Hake, S. (2004). From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. Plant Cell **16,** 2265-2277.

Magome, H., Yamaguchi, S., Hanada, A., Kamiya, Y., and Oda, K. (2004). *dwarf* and *delayed-flowering 1*, a novel *Arabidopsis* mutant deficient in gibberellin

biosynthesis because of overexpression of a putative AP2 transcription factor. The Plant Journal **37,** 720-729.

Mao, Y., Pavangadkar, K., Thomashow, M., and Triezenberg, S. (2006). Physical and functional interactions of Arabidopsis ADA2 transcriptional coactivator proteins with the acetyltransferase GCN5 and with the cold-induced transcription factor CBF1. Biochimica et Biophysica Acta **1759,** 69-79.

Maruyama, K., Sakuma, Y., Kasuga, M., Ito, Y., Seki, M., Goda, H., Shimada, Y., Yoshida, S., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. The Plant Journal **38,** 982-993.

McKown, R., Kuroki, G., and Warren, G. (1996). Cold response of *Arabidopsis* mutants impaired in freezing tolerance. Journal of Experimental Botany **47,** 1919-1925.

Mohapatra, S.S., Poole, R.J., and Dhindsa, R.S. (1987). Changes in protein patterns and translatable messenger RNA populations during cold acclimation of alfalfa. Plant Physiol. **84,** 1172-1176.

Ohme-Takagi, M., and Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell **7,** 173-182.

Ohta, M., Matsui, K., Hiratsu, K., Shinshi, H., and Ohme-Takagi, M. (2001). Repression domains of Class II ERF transcriptional repressors share an essential motif for active repression. Plant Cell **13,** 1959-1968.

Qin, F., Sakuma, Y., Li, J., Liu, Q., Li, Y.-Q., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea mays* L. Plant Cell Physiol. **45,** 1042-1052.

Rashotte, A.M., Mason, M.G., Hutchison, C.E., Ferreira, F.J., Schaller, G.E., and Kieber, J.J. (2006). A subset of *Arabidopsis* AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. PNAS **103,** 11081-11085.

Riechmann, J.L., and Meyerowitz, E.M. (1998). The AP2/EREBP family of plant transcription factors. Biol Chem **379,** 633-646.

Rudolph, A.S., and Crowe, J.H. (1985). Membrane stabilization during freezing: The role of two natural cryoprotectants, trehalose and proline. Cryobiology **22,** 367.

Sakai, A., and Larcher, W. (1987). Frost survival of plants: responses and adaptation to freezing stress. (Berlin: Springer-Verlag).

Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun **290,** 998-1009.

Sato, F., Kitajima, S., and Koyama, T. (1996). Ethylene-induced gene expression of osmotin-like protein, a neutral isoform of tobacco PR-5, is mediated by the AGCCGCC eft-sequence. Plant Cell Physiol. **37,** 249-255.

Seipel, K., Georgiev, O., and Schaffner, W. (1994). A minimal transcription activation domain consisting of a specific array of aspartic acid and leucine residues. Biol Chem Hoppe Seyler **375,** 463-470.

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y., and Shinozaki, K. (2001). Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a Full-Length cDNA microarray. Plant Cell **13,** 61-72.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K. (2002). Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. The Plant Journal **31,** 279-292.

Skinner, J., von Zitzewitz, J., Szucs, P., Marquez-Cedillo, L., Filichkin, T., Amundsen, K., Stockinger, E., Thomashow, M., Chen, T., and PM., H. (2005). Structural, functional, and phylogenetic characterization of a large CBF gene family in barley. Plant Mol Biol **59,** 533-551.

Steponkus, P., and Webb, M. (1992). Freeze-induced dehydration and membrane destabilization in plants. (Berlin: Springer-Verlag).

Steponkus, P.L., Uemura, M., Joseph, R.A., Gilmour, S.J., and Thomashow, M.F. (1998). Mode of action of the COR15a gene on the freezing tolerance of *Arabidopsis thaliana*. PNAS **95,** 14570-14575.

Stockinger, E.J., Gilmour, S.J., and Thomashow, M.F. (1997). *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the *C-repeat/DRE*, a *cis*-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. PNAS **94,** 1035-1040.

Stockinger, E.J., Mao, Y., Regier, M.K., Triezenberg, S.J., and Thomashow, M.F. (2001). Transcriptional adaptor and histone acetyltransferase proteins in Arabidopsis and their interactions with CBF1, a transcriptional activator involved in cold-regulated gene expression. Nucl. Acids Res. **29,** 1524-1533.

Strauss, G., and Hauser, H. (1986). Stabilization of lipid bilayer vesicles by sucrose during freezing. PNAS **83,** 2422-2426.

Thomashow, M.F. (1998). Role of cold-responsive genes in plant freezing tolerance. Plant Physiol. **118,** 1-8.

Thomashow, M.F. (1999). Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology **50,** 571-599.

Tran, E.J., and Wente, S.R. (2006). Dynamic Nuclear Pore Complexes: Life on the Edge. Cell **125,** 1041.

Triezenberg, S.J. (1995). Structure and function of transcriptional activation domains. Current Opinion in Genetics & Development **5,** 190.

van der Fits, L., and Memelink, J. (2001). The jasmonate-inducible AP2/ERF-domain transcription factor ORCA3 activates gene expression via interaction with a jasmonate-responsive promoter element. The Plant Journal **25,** 43-53.

Vlachonasios, K.E., Thomashow, M.F., and Triezenberg, S.J. (2003). Disruption mutations of ADA2b and GCN5 transcriptional adaptor genes dramatically affect *Arabidopsis* growth, development, and gene expression. Plant Cell **15,** 626-638.

Vogel, J.T., Zarka, D.G., Van Buskirk, H.A., Fowler, S.G., and Thomashow, M.F. (2005). Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. The Plant Journal **41,** 195-211.

Volger, H., and Heber, U. (1975). Cryoprotective leaf proteins. Biochim Biophys Acta **412,** 335-349.

Wang, Z., Triezenberg, S., Thomashow, M., and EJ., S. (2005). Multiple hydrophobic motifs in *Arabidopsis* CBF1 COOH-terminus provide functional redundancy in trans-activation. Plant Mol Biol **58,** 543-559.

Wilson, K., Long, D., Swinburne, J., and Coupland, G. (1996). A dissociation insertion causes a semidominant mutation that increases expression of TINY, an *Arabidopsis* gene related to APETALA2. Plant Cell **8,** 659-671.

Wuitschick, J.D., Lindstrom, P.R., Meyer, A.E., and Karrer, K.M. (2004). Homing endonucleases encoded by germ line-limited genes in *Tetrahymena thermophila* have APETALA2 DNA binding domains. Eukaryotic Cell **3,** 685-694.

Xin, Z., and Browse, J. (1998). *eskimo1* mutants of *Arabidopsis* are constitutively freezing-tolerant. PNAS **95,** 7799-7804.

CHAPTER TWO

FUNCTIONAL ROLE OF THE CONSERVED PKKPAGR AND DSAWR MOTIFS IN CBF1 ACTIVITY

INTRODUCTION

CBF proteins are members of the AP2/EREBP family of transcription factors in *Arabidopsis* (Riechmann and Meyerowitz, 1998). This multi-gene family is highly conserved among plants and is characterized by the presence of the AP2/EREBP DNA-binding domain (Jofuku *et al.*, 1994; Ohme-Takagi and Shinshi, 1995). NMR studies on AtERF1 (Allen *et al.*, 1998) revealed that this DNA-binding domain is composed of a three-strand β-sheet structure followed by an α-helix. The β-sheet, at the N-terminus, is the fold which mediates DNA recognition. High identity within the DNA-binding domain and conservation of the critical amino acids required for protein-DNA interaction suggests that proteins in the AP2/EREBP family likely adopt a similar 3D structure.

Within the AP2/EREBP family in *Arabidopsis*, CBF1/DREB1B, CBF2/DREB1C, CBF3/DREB1A, CBF4/DREB1D, DREB1E/DDF1 and DREB1F/DDF2 (Stockinger *et al.*, 1997; Gilmour *et al.*, 1998; Liu *et al.*, 1998; Haake *et al.*, 2002; Sakuma *et al.*, 2002; Magome *et al.*, 2004) define a small sub-family that is characterized by the presence of two short polypeptide sequences that flank the AP2/EREBP domain: PKKP/RAGRxKFxETRHP (abbreviated as PKKPAGR) and DSAWR, at the N-and C-terminus, respectively (Figure 2.1), designated the "signature" sequences (Jaglo *et al.*, 2001).

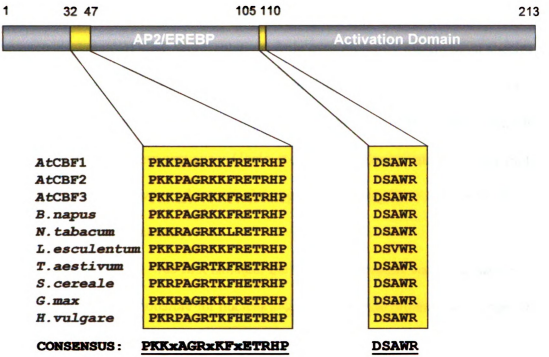


Figure 2.1. PKKPAGR and DSAWR signature sequences in the CBF family of plant transcription factors.
Top panel. Schematic diagram of CBF1 protein. Yellow boxes represent PKKPAGR and DSAWR motifs. The grey box at the C-terminus represents the activation domain. The AP2/EREBP DNA-binding domain is indicated.
Bottom panel. Sequence alignment of AtCBF1-3 and CBF-like proteins. The signature sequences are included in the yellow boxes. The consensus sequence is indicated below. At, *Arabidopsis thaliana*; B. napus, *Brassica napus*; N. tabacum, *Nicotiana tabacum*; S. cereale, *Secale cereale*; T. aestivum, *Triticum aestivum*; G. max, *Glycine max*; H. vulgare, *Hordeum vulgare*.

Orthologues of the *CBF* genes are present in a wide variety of plant species. A sequence alignment of CBF representatives from evolutionarily distant plant species showed that most of the identity among the *Arabidopsis* CBF family and CBF-like proteins from other species is due to a high degree of identity between their AP2/EREBP domains and the signature sequences (Jaglo *et al.*, 2001). Conservation of these motifs across evolutionarily diverse plants species suggests that they have an important functional role in CBF activity.

When AtCBF1 was first isolated, the PKKPAGR motif was proposed to represent the signal responsible for nuclear localization of the protein (Stockinger *et al.*, 1997). This observation was based on the similarity of this sequence to known nuclear localization signals from other plant proteins, such as the early induced Aux/IAA genes (Abel *et al.*, 1994), the photomorphogenic repressor protein COP1 (von Arnim and Deng, 1994) and the heat shock factors, HSF (Lyck *et al.*, 1997). In addition, localization studies of the member of the AP2/EREBP family AINTEGUMENTA (Klucher *et al.*, 1996) have indicated that nuclear targeting can depend on the presence of specific basic residues (Krizek and Sulli, 2006).

Whether nuclear shuttling of CBF1 is dependent on the PKKPAGR motif, and whether this is a low temperature-regulated process is not clear. Recent studies focused on the subcellular localization of several members of the AP2/EREBP family have revealed that nuclear targeting within this family can be either constitutive or stimulus-dependent (Klucher *et al.*, 1996; Rashotte *et al.*, 2006); therefore it is not possible to draw a general conclusion based on studies on other AP2/EREBP family members. It is expected that some CBF protein can accumulate in the nucleus in the absence of a cold

stimulus, as *Arabidopsis* lines overexpressing *CBF1*, *CBF2* or *CBF3* display constitutive accumulation of *COR* transcripts compared to the non-transgenic plants (Jaglo-Ottosen *et al.*, 1998; Gilmour *et al.*, 2000). On the other hand, when *CBF*-overexpressing plants are cold-treated, *COR* gene expression increases (Gilmour *et al.*, 2000), suggesting that nuclear transport may be more efficient at low temperatures.

For several families of DNA-binding proteins, regions flanking the DNA-binding domain can be an essential extension important not only for binding affinity but for binding specificity as well (Crane-Robinson *et al.*, 2006). Based on their proximity to the AP2/EREBP domain, a reasonable hypothesis is that the PKKPAGR and DSAWR motifs participate in binding to the *CRT/DRE* promoter element. In addition, the N-terminal PKKPAGR region of this motif is rich in basic residues, bearing positive charges that may favor a DNA-protein interaction. Finally, deletion studies of ERF1 have shown that the presence of 10 amino acids immediately upstream and 8 amino acids downstream of the AP2/EREBP domain can greatly stabilize DNA binding (Hao *et al.*, 1998). Based on these observations, we hypothesized that the two signature motifs might play a role in CBF binding to its cognate *CRT/DRE* promoter element.

The main goal of the experiments described in this chapter was to investigate the roles of the PKKPAGR and DSAWR motifs in CBF function. The question was addressed by taking a mutational approach. *Arabidopsis* transgenic lines were generated by overexpressing *CBF1* transgenes harboring specific mutations in the PKKPAGR and DSAWR motifs. As a hallmark of CBF1 activity, *COR* gene expression was tested. The results demonstrate that both conserved motifs are required for CBF1 function, since *Arabidopsis* plants overexpressing a *CBF1* transgene carrying mutations in the signature

sequences show greatly diminished *COR* gene expression compared to plants overexpressing the wild type CBF1 transgene. The effect of the mutations cannot be explained by reduced protein levels, as indicated by western blot analysis of plants overexpressing the *6xMyc:CBF1* transgenes. Contrary to what was originally hypothesized, the PKKPAGR motif alone is not required for nuclear localization of a CBF1:GFP:GUS chimera. Electrophoretic-mobility gel shift experiments showed that the PKKPAGR and DSAWR motifs play a role in mediating the recognition of the *CRT/DRE* *cis*-acting element present in the *COR* gene promoters. Extensive mutational analysis within the PKKPAGR motif combined with secondary structure prediction methods suggested that certain residues in the predicted helical region represented by the RKKFRET sequence are involved in direct protein-DNA interaction. A computational model of CBF1 bound to its target *CRT/DRE* promoter element has led to the identification of specific residues that are essential to DNA binding and might mediate specific DNA recognition.

RESULTS

A MUTATIONAL APPROACH TO INVESTIGATE THE FUNCTIONAL IMPORTANCE OF THE SIGNATURE SEQUENCES IN CBF1 FUNCTION.

To investigate the importance of the signature sequences in CBF function, *Arabidopsis* transgenic lines were generated by overexpressing *CBF1* transgenes harboring specific mutations in the PKKPAGR and DSAWR motifs. *AtCBF1* was selected as a representative model for the *Arabidopsis* CBF family, as it has been shown that overexpression of *AtCBF1, AtCBF2,* or *AtCBF3* results in activation of the same *CBF* regulon of genes, indicating that the three proteins have overlapping function (Gilmour *et al.*, 2004). Alanine scanning was used to introduce stretches of three alanines in place of the original amino acids within the PKKPAGR motif, which resulted in 5 mutant ORFs named *M1-M5* (Figure 2.2). In addition, the entire PKKPAGR motif was deleted, and the resulting *ORF* named *ΔPKK*. The wild type and mutagenized versions of *CBF1* driven by the constitutive 35*S* Cauliflower Mosaic Virus (35*S* CaMV) promoter were transformed into *Arabidopsis* and are referred hereafter as *pkkpagr* (*ΔPKK, M1-M5*) and *dsawr* lines. For each construct, fifteen to twenty independent homozygous lines were selected, and five to seven representative lines were chosen for further characterization.

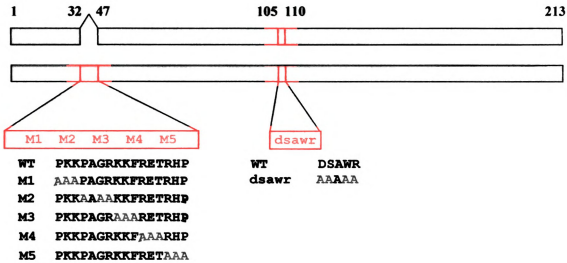


Figure 2.2. Site-directed mutagenesis was used to generate CBF1 ORFs containing specific mutations or deletions of the signature sequences.
Top. Schematic of full-length CBF1 and CBF1 lacking the PKKPAGR motif (*ΔPKK*).
Bottom. Description of specific mutations within the signature sequences.
WT: wild type sequence for the two signature sequences.
M1-M5, dsawr: Amino acids in red indicate where the wild type CBF1 sequence has been substituted with alanines.

Overexpression of *AtCBF1, AtCBF2,* or *AtCBF3* leads to constitutive expression of the *CBF* regulon, a group of about 85 genes containing the *cis*-acting *CRT/DRE* promoter element (Vogel *et al.*, 2005). As a result, *CBF*-overexpressing plants are constitutively freezing-tolerant. However, overexpression of wild type CBF1 results in transgenic plants that are smaller in stature and delayed in flowering compared to the wild type plants or transgenic plants harboring the empty vector. In addition, higher amounts of CBF transcripts correlate with smaller plants that are delayed in flowering (Haake *et al.*, 2002; Gilmour SJ *et al.*, 2004; Gilmour *et al.*, 2004).

Transgenic lines overexpressing CBF1-*pkkpagr* at levels similar to or higher than plants overexpressing wild type CBF1 were selected by Northern blot analysis (Figure 2.3 C) and their growth phenotype was compared (representative examples are shown in Figure 2.3 A).

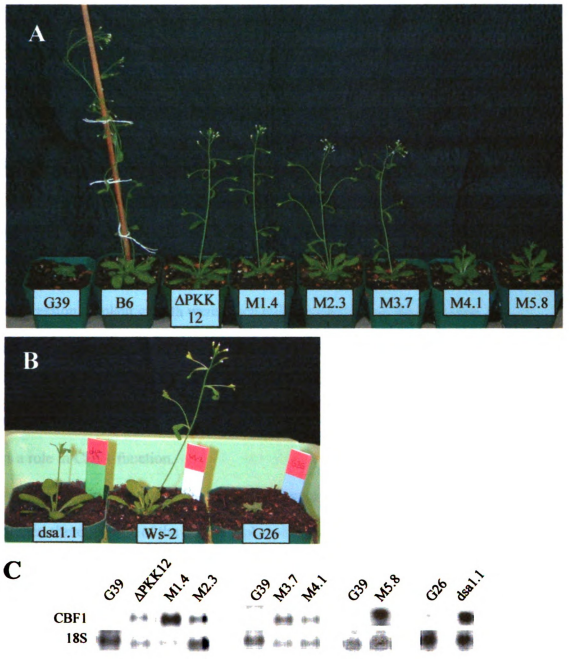


Figure 2.3. *Arabidopsis pkkpagr* and *dsawr* lines.
Seedlings harboring the *35S::CBF1* construct (G26 and G39), CBF1 mutants (*M1-5, ΔPKK, and dsawr*) or the empty vector (*B6*) were grown on plates for 14 days before being transferred to soil. Pictures were taken 30 days after transplanting into soil.
A. *pkkpagr* lines.
B. Left, *dsawr* transgenic line.
C. Northern blot analysis of different 35S::CBF1 Arabidopsis plants. Total RNA extracted from two-week old seedlings was tested for the expression of CBF1. RNA levels of the 18S ribosomal RNA were used as a normalization control.

Overall, overexpression of *CBF1* carrying alanine mutations in the PKKPAGR motif resulted in plants showing different degrees of growth retardation, as compared to the control line B6, carrying the empty vector. However, the growth retardation observed in the *pkkpagr* lines was always less severe than that observed in the WT CBF1 over-expressing plants, when similar transcript levels were compared (compare plant growth and CBF1 expression levels in Figures 2.3A and C). These observations suggested that the PKKPAGR motif is required for CBF1 function.

As was observed in the *pkkpagr* lines, the negative effects of CBF1-*dsawr* overexpression on plant growth were much milder than those in the wild type overexpressor (Figure 2.3B). It must be noted that *CBF1* transcript levels in the *dsawr* plants were similar to or higher than in the plants overexpressing wild type CBF1 (representative example shown in Figure 2.3 C), suggesting that the DSAWR motif also plays a role in CBF1 function.

In summary, overexpression of *CBF1* transgenes mutated in the two signature sequences altered of one of the hallmark phenotypes of CBF-overexpressing plants, growth retardation, providing a first suggestion that these two conserved motifs play a role in CBF1 activity.

FUNCTIONAL ROLE OF THE PKKPAGR MOTIF IN CBF1 ACTIVITY.

Mutations in the PKKPAGR motif affect *COR* gene activation.

To assess the effect of alanine mutations on CBF1 activity, wild type and *pkkpagr* lines were analyzed for *COR* gene expression by northern blot analysis. Overexpression of *AtCBF1* leads to constitutive expression of the *CBF* regulon, including *COR6.6*, *COR15A*, *COR47*, and *COR78* (Jaglo-Ottosen *et al.*, 1998). If the PKKPAGR motif is required for CBF1 activity, we would expect that the ability of CBF1 to activate these target genes would be impaired or absent in transgenic plants expressing CBF proteins with mutations in this motif. This possibility was tested by determining expression levels of target *COR* genes in transgenic plants overexpressing wild type and mutant versions of the PKKPAGR sequence.

Total RNA was analyzed for the expression of *CBF1* and *COR* genes in independent *pkkpagr* lines, as well as in transgenic lines overexpressing the wild type *CBF1* or harboring the empty vector. The experiment was carried out by growing *Arabidopsis* seedlings at control temperature (22°C). At this temperature the endogenous *CBF* and *COR* gene transcripts are barely detectable by northern blotting (Figure 2.4A and B, see B6 line, which is transformed with the empty vector). In contrast, transgenic lines overexpressing wild type *CBF1* are able to activate constitutive expression of the *COR* genes (Jaglo-Ottosen *et al.*, 1998; Figure 2.4A and B, see all WT lines).

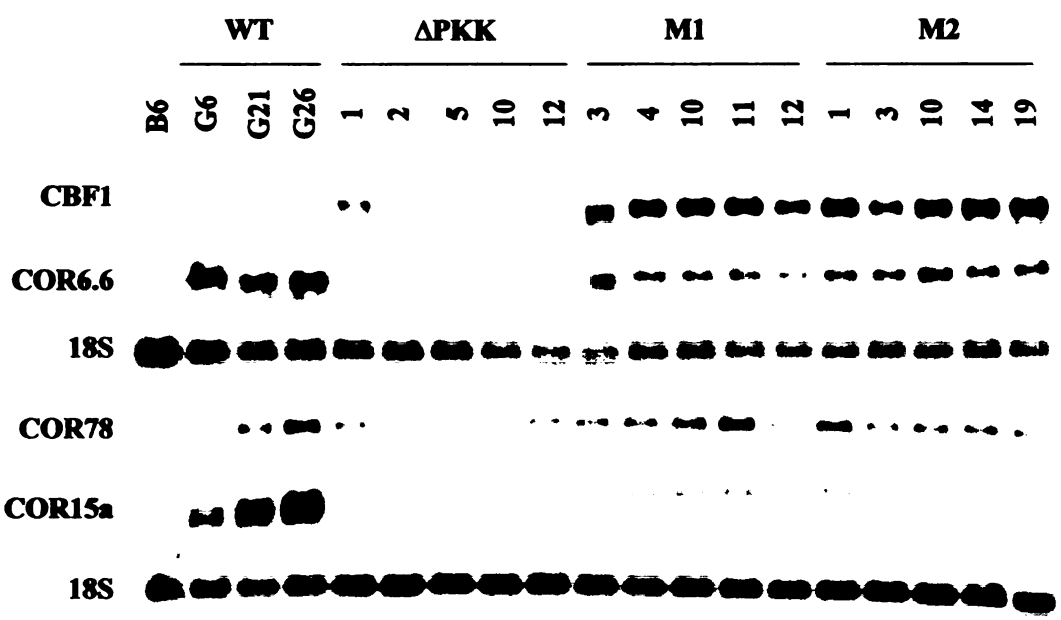


Figure 2.4 A. Analysis of *COR* gene induction in *ΔPKK*, *M1* and *M2 pkkpagr* lines by Northern blot analysis. Total RNA was extracted from 2-week old seedlings and tested for the expression of *CBF1*, *COR6.6*, *COR78*, and *COR15a*. 18S ribosomal RNA levels are shown as loading control.
B6, transgenic plants carrying an empty expression vector; WT, wild type CBF1 overexpressing plants; ΔPKK, M1 and M2, transgenic lines overexpressing CBF1 lacking the entire PKKPAGR region or carrying triple alanine mutations, respectively. Numbers above lanes represent independent transgenic lines.

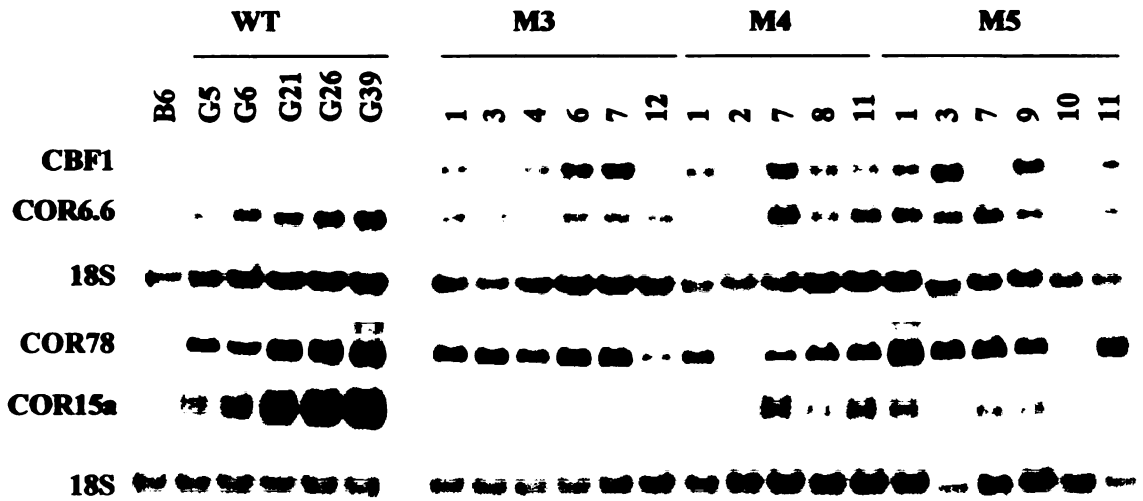
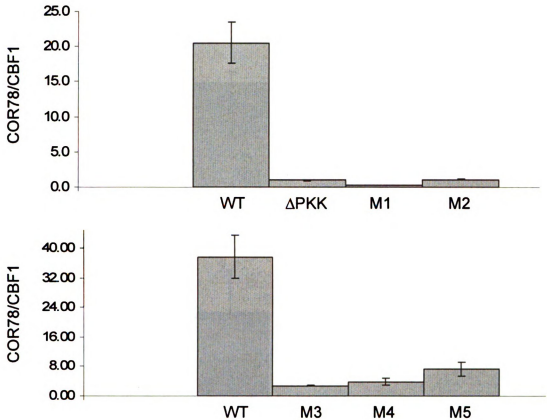


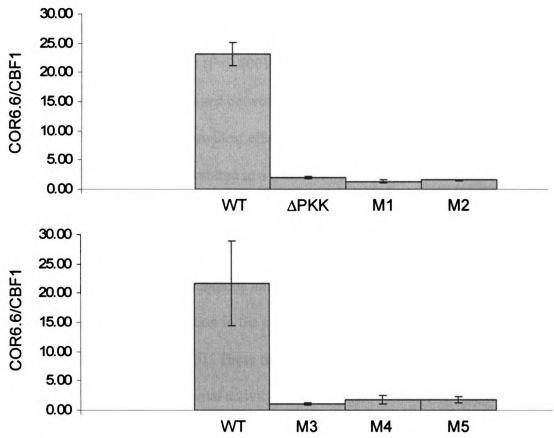
Figure 2.4 B. Analysis of *COR* gene induction in *M3*, *M4* and *M5 pkkpagr* lines by Northern blot analysis. Total RNA was extracted from 2-week old seedlings and tested for the expression of *CBF1*, *COR6.6*, *COR78*, and *COR15a*. 18S ribosomal RNA levels are shown as a loading control.
B6, transgenic plants carrying an empty expression vector; WT, wild type CBF1 overexpressing plants; M3-M5, transgenic lines overexpressing CBF1 carrying triple alanine mutations. Numbers above lanes represent independent transgenic lines.

The *pkkpagr* lines showed higher *COR* gene induction than the control line *B6*, suggesting that the CBF1 mutated proteins had retained some ability to activate transcription (Figures 2.4 A and B, compare *pkkpagr* lines M1-M5, and ΔPKK with B6). However, this function was affected by alanine substitutions in the PKKPAGR motif. For instance, *COR15a* induction was greatly reduced in all the *pkkpagr* lines as compared to the transgenic plants overexpressing wild type CBF1 transgene (Figures 2.4 A and B, M1-M5 lines), and almost undetectable in the lines overexpressing CBF1 transgene lacking the whole PKKPAGR motif (Figures 2.4 A, ΔPKK lines). Induction of COR6.6 and COR78 appeared to be also affected, in that higher levels of mutated *CBF1* transcript were needed to reach comparable *COR* transcript levels (Figures 2.4 A and B, compare *COR6.6* and *COR78* transcript levels and corresponding *CBF1* transcript levels in WT and *pkkpagr* lines M1-M5). To determine whether mutations in the PKKPAGR motif had a significant effect on *COR6.6* and *COR78* induction, we compared the *COR/CBF1* transcript ratios in the transgenic lines overexpressing wild type and mutated CBF1 by ANOVA (Figures 2.5A and B).



Differences of Least Square Means - COR78/CBF1 ratios							
WT versus ΔPKK, M1, M2				WT versus M3, M4, M5			
Genotype comparison	Estimate	t value	P value	Genotype comparison	Estimate	t value	P value
WT vs. ΔPKK	-1.2876	-16.62	<.0001	WT vs. M3	-1.1174	-9.14	<.0001
WT vs. M1	-1.7986	-23.22	<.0001	WT vs. M4	-1.0185	-7.98	<.0001
WT vs. M2	-1.2879	-16.62	<.0001	WT vs. M5	-0.7671	-6.28	<.0001
ΔPKK vs. M1	-0.5109	-6.59	0.9968	M3 vs. M4	-0.0989	-0.81	0.4288
ΔPKK vs. M2	-0.0003	0.00	<.0001	M3 vs. M5	-0.0350	-3.01	0.0076
M1 vs. M2	-0.5106	-6.59	<.0001	M4 vs. M5	-0.2514	-2.06	0.0054

Figure 2.5A. Analysis of variance of *COR78/CBF1* transcript ratios in transgenic plants overexpressing wild type or mutated CBF1 transgenes.
Top. Bar graphs represent the average of *COR78/CBF1* ratios for 5-6 independent transformants for different transgenic lines. WT, M1-M5, ΔPKK, transgenic lines overexpressing CBF1 wild type, its mutated versions or lacking the PKKPAGR motif. Bottom. Differences of least square means of COR78/CBF1 transcript ratios for transgenic lines overexpressing CBF1 wild type or its mutated versions.



Differences of Least Square Means - COR6.6/CBF1 ratios							
WT versus ΔPKK, M1, M2				WT versus M3, M4, M5			
Genotype comparison	Estimate	t value	P value	Genotype comparison	Estimate	t value	P value
WT vs. ΔPKK	-1.0578	-14.37	<.0001	WT vs. M3	-1.2853	-7.04	<.0001
WT vs. M1	-1.2605	-17.12	<.0001	WT vs. M4	-1.1368	-5.96	<.0001
WT vs. M2	-1.1783	-16.01	<.0001	WT vs. M5	-1.1158	-6.11	<.0001
ΔPKK vs. M1	-0.2026	-2.75	0.0142	M3 vs. M4	-0.1485	-0.81	0.4268
ΔPKK vs. M2	-0.1204	-1.64	0.1213	M3 vs. M5	-0.1695	-0.97	0.3433
M1 vs. M2	-0.0821	-1.12	0.2807	M3 vs. M4	-0.0210	-0.11	0.9097

Figure 2.5B. Analysis of variance of *COR6.6/CBF1* transcript ratios in transgenic plants overexpressing wild type or mutated CBF1 transgenes. Log_{10} values were used for this analysis. Top. Bar graphs represent the average of *COR6.6/CBF1* ratios for 5-6 independent transformants for different transgenic lines. WT, M1-M5, ΔPKK, transgenic lines overexpressing CBF1 wild type, its mutated versions or lacking the PKKPAGR motif.

Bottom. Differences of least square means of *COR6.6/CBF1* transcript ratios for transgenic lines overexpressing CBF1 wild type or its mutated versions.

In all cases, the difference in *COR78/CBF1* and *COR6.6/CBF1* transcript ratios between plants overexpressing the native *CBF1* (control lines) and any of its mutated version was highly significant ($P<0.0001$) according to ANOVA analysis. For instance, *COR78/CBF1* ratio values ranged between 20.5 and 37 in the control lines, whereas M5 mutation, which showed the smallest effect on *COR78* accumulation, (Figure 2.4B, compare WT and M5 lines) resulted in a reduction of *COR78/CBF1* ratio of 7.4 folds (Figure 2.5A). Similarly, a *COR6.6/CBF1* ratio of approximately 23 folds in the control lines ranged between 1-2 folds in the *pkkpagr* transgenic lines (Figure 2.5B).

In summary, all the mutations introduced in the PKKPAGR motif caused a reduction in *COR* gene induction in the *pkkpagr* lines compared to the lines overexpressing wild type CBF1. These results indicate that an intact PKKPAGR motif is required for CBF1 transcriptional activity.

Reduced *COR* gene induction in the *pkkpagr* lines is not caused by lower protein levels.

The reduced *COR* gene expression observed by northern blot analysis could be due to different reasons. One would be that the mutations generated within the PKKPAGR region reduced CBF1 stability; i.e. for a given transcript level, less mutated protein accumulated compared to the wild type protein. To test this possibility, we analyzed CBF1 protein levels in *Arabidopsis* protein extracts of wild type and transgenic plants.

Antibodies raised against full-length CBF1 or specific peptides have failed to detect CBF1 protein in either cold-acclimated wild type or 35*S*::*CBF1* transgenic *Arabidopsis* lines (Jaglo-Ottosen PhD thesis; Canella, unpublished). Therefore, to conduct protein detection

experiments by western blot analysis, we generated *Arabidopsis* plants overexpressing the *6xMyc:CBF1* transgene under control of the 35*S* CaMV promoter (Figure 2.6A). This particular tag was chosen because it has been successfully used to detect different proteins in plant extracts by western blot analysis (DeWitt *et al.*, 1996; Rivas *et al.*, 2002).

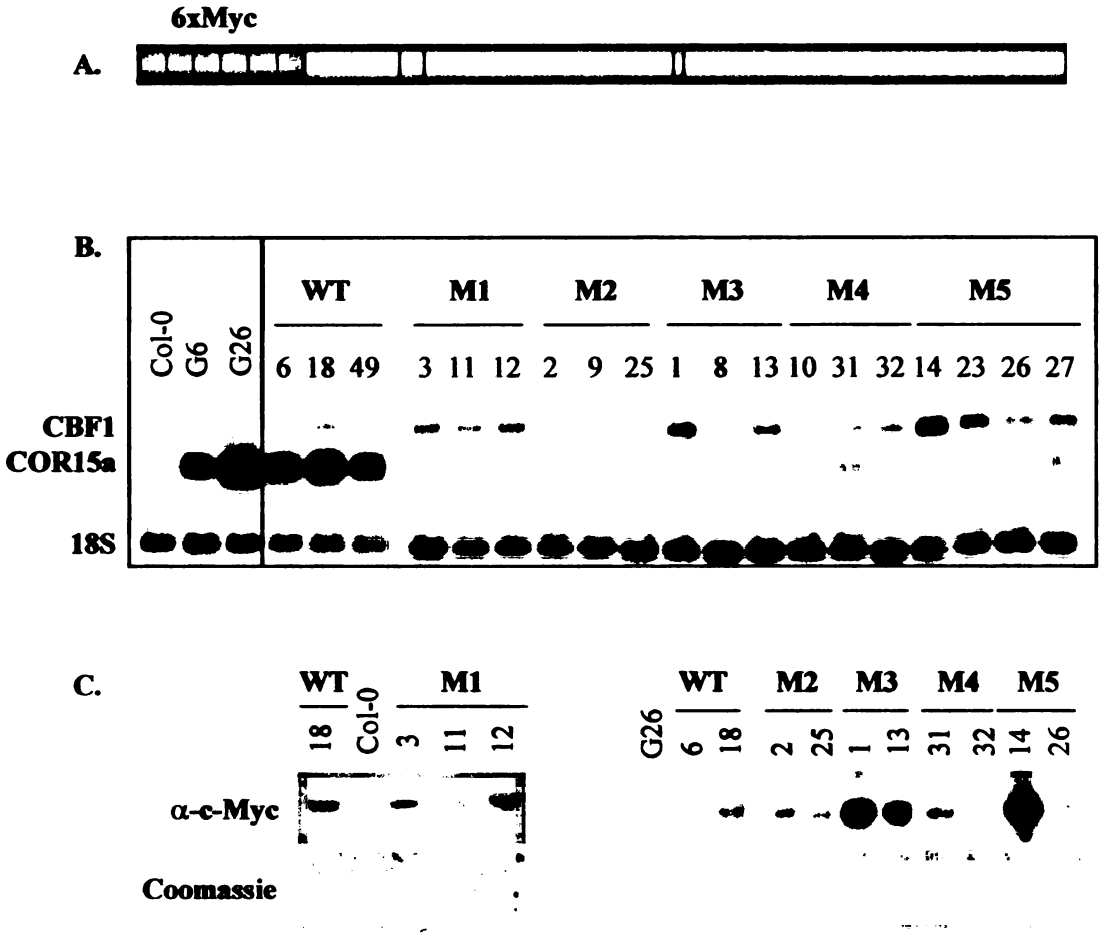


Figure 2.6. Comparison of CBF1 transcript and protein levels in *Arabidopsis* plants overexpressing 6xMyc:CBF1 transgene.
A. Schematic of 6xMyc:CBF1 construct. The N-terminal 6xMyc tag is indicated by the dark grey boxes; yellow boxes represent the two signature sequences flanking the DNA-binding domain.
B. Northern blot analysis of 35S::6xMyc:CBF1 transgenic lines. Total RNA was analyzed for CBF1 and *COR15a* expression. Transcript amounts were normalized using the 18S ribosomal loading control.
C. Western blotting showing the presence of 6xMyc:CBF1 protein in the transgenic plants (detected with a monoclonal anti-c-Myc antibody).
Col-0, non-transgenic plants; G6 and G26, *Arabidopsis* plants overexpressing wild type CBF1 without 6xMyc tag; WT and M1-M5, transgenic plants overexpressing a 6xMyc tag fused to a wild type (WT) or mutated (M1-5) CBF1 transgene. Protein loading is indicated by the Coomassie-stained membrane.

Several independent lines per construct were analyzed for the expression levels of *CBF1* and *COR* genes. Northern blot analysis indicated that plants overexpressing a wild type CBF1 transgene fused to a 6xMyc tag could accumulate *COR15a* levels comparable to plants overexpressing the wild type CBF1 transgene (Figure 2.6B). In contrast, overexpression of *6xMyc:CBF1* transgenes carrying mutations in the PKKPAGR motif resulted in a dramatic reduction of *COR15a* accumulation. These data indicate that addition of a tag to the N-terminus of CBF1 doesn't significantly affect the activity of CBF1 wild type or CBF1 mutated in the PKKPAGR motif.

To determine the levels of the CBF1 wild type of mutated protein *in planta*, total protein extracts from two representative transgenic lines were chosen for western blot analysis. Comparison of the northern and western blotting results indicated a good agreement between CBF1 transcript and protein levels. For instance, the two independent lines for the PKKPAGR mutation M2 (line #2 and line #25) show *CBF1* transcript levels similar to the control line WT18 (Figure 2.6B). Consistently, the protein levels are very similar (Figure 2.6C, right panel). Moreover, in transgenic lines that could accumulate high levels of mutated CBF1 protein compared to the wild type protein, *COR* gene induction was still much lower than in the wild type overexpressing lines (compare M3.1 and M5.14 in Figure 2.6B and 265C, right panel). Therefore, we can conclude that the reduced *COR* gene induction observed in the *pkkpagr* lines cannot be attributed to reduced CBF1 protein accumulation.

Role of the PKKPAGR motif in nuclear targeting of CBF1.

As discussed above, one possible function of the PKKPAGR motif could be that of mediating CBF targeting to the nucleus. To test this hypothesis, we generated *Arabidopsis* plants overexpressing translational fusions of CBF1 to GFP:GUS. As a control, *Arabidopsis* plants overexpressing GFP:GUS alone and Nuclear Inclusion protein a (NIa):GFP:GUS (Carrington *et al.*, 1991) were generated. GFP:GUS and NIa:GFP:GUS are known for being mainly localized to the cytoplasm and the nucleus, respectively (Grebenok *et al.*, 1997). In addition to the full-length CBF1 ORF, a construct lacking the PKKPAGR motif (CBF1ΔPKK) was designed to assess whether the nuclear localization of CBF1 requires this region of the protein. To determine whether the PKKPAGR motif is sufficient to drive nuclear entry of CBF1, the N-terminal fragment of CBF1 containing the first 32 residues and an intact PKKPAGR motif (PKK:GFP:GUS) was generated (Figure 2.7A). The localization of the proteins was monitored by laser confocal microscopy.

Independent transgenic lines were selected by northern blot analysis and analyzed by laser confocal microscopy. For the construct harboring the full-length CBF1 transgene fused to GFP:GUS, it was not possible to isolate any transgenic plants, despite three independent transformation events by the floral dip method (Clough and Bent, 1998). One possible explanation for this is that those plants produced a highly stable CBF1 protein which is extremely toxic to plants. For the remaining constructs, northern blot analysis of two representative lines confirmed the expression of the transgenes (Figure 2.7B, upper panel - see GFP probe).

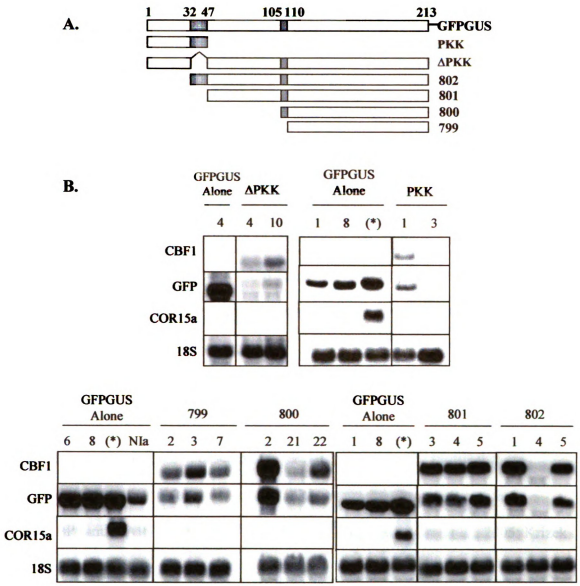


Figure 2.7. Northern blot analysis of *Arabidopsis* seedlings overexpressing different CBF1:GFP:GUS constructs.
A. Schematic of the different CBF1:GFP:GUS constructs. CBF1(aa1-213), full-length protein, ΔPKK, a construct in which the PKKPAGR region has been deleted; PKK, first 32 amino acids including the PKKPAGR motif. All the other constructs represent 5' deletions of CBF1 ORF. Dark grey boxes represent the two signature sequences.
B. Northern blot analysis of transgenic lines overexpressing constructs indicated in the top panel. Total RNA (10 ug) from 12-day-old seedlings was tested for *CBF1*, *GFP*, and *COR15a* transcript levels. Numbers above the lanes indicate independent lines for each given construct. The asterisk represent samples from cold-treated (4°C, 10 hours) *Arabidopsis* seedlings overexpressing GFP:GUS alone.

To determine the cellular localization of different CBF1:GFP:GUS chimeras, six-to-seven day old *Arabidopsis* seedlings were grown under controlled (22°C) conditions and root tips were analyzed by laser confocal microscopy (Figure 2.8). Analysis of the control lines confirmed that GFP:GUS and NIa:GFP:GUS were mainly localized to the cytoplasm and nucleus, respectively. The CBF1ΔPKK:GFP:GUS chimera could direct the cytoplasmically localized GFP:GUS to the nucleus, indicating that nuclear localization of the protein is not affected by the deletion of the PKKPAGR motif alone. In addition, a fusion of the 47 N-terminal residues of CBF1 to GFP:GUS could not direct the chimera into the nucleus, indicating that the presence of the PKKPAGR motif in this context is not sufficient to promote nuclear entry of CBF1:GFP:GUS.

That the PKKPAGR is neither required nor sufficient for nuclear localization of CBF1 was a surprising finding, based on the resemblance to similar motifs that are functional NLSs in other plant transcription factors, including the member of the AP2/EREBP family AINTEGUMENTA (Abel *et al.*, 1994; Koroleva *et al.*, 2005; Krizek and Sulli, 2006; Rashotte *et al.*, 2006).

In summary, localization studies showed that, under the experimental conditions presented, the lack of the PKKPAGR motif does not prevent CBF1 from entering the nuclear compartment, and therefore decreased transcriptional activity of the protein cannot be explained by failure to localize to the nucleus.

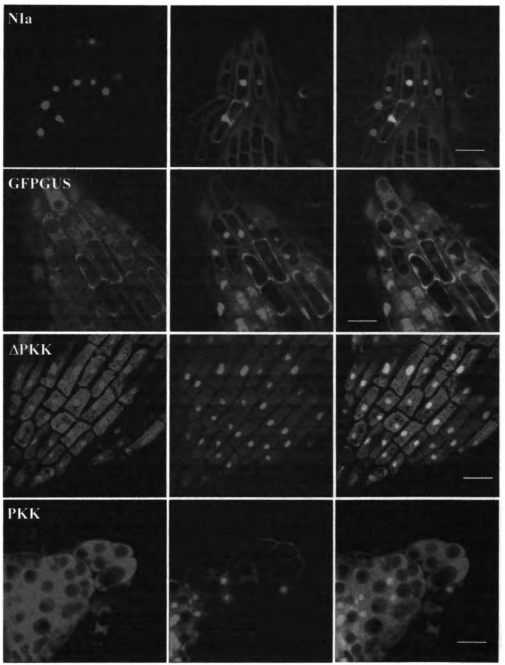


Figure 2.8. Localization studies of CBF1:GFP:GUS chimeras.
Laser confocal images of root tips of *Arabidopsis* plants overexpressing various CBF1:GFP:GUS chimeras.
Left, GFP fluorescence (green); center, Propidium Iodide fluorescence (red); right, merged images (yellow color indicates overlap of red and green fluorescence). All constructs described in Figure 2.6; NIa, Nuclear Inclusion protein A, a nuclear marker.

In addition, these results indicate that CBF1 can gain access to the nuclear compartment through an active mechanism, since it can facilitate the nuclear import of GFP:GUS (approximately 100KDa), whose size is well above the size cut-off for free diffusion into the nucleus (Kaffman and O'Shea, 1999). For a facilitated transport across the nuclear envelope, proteins need a recognition sequence for interaction with nuclear receptors or directly with the nucleoporins that are an integral part of the Nuclear Pore Complex (NPC).

Identifying regions in CBF1 involved in nuclear transport.

Several reports have indicated that it is not unusual to find multiple karyophilic signals in nuclear proteins, leading to the suggestion that they could provide some redundancy (Sudbeck and Scherer, 1997; Tang *et al.*, 1997; Walther *et al.*, 2005). For instance, different NLS could provide independent mechanisms to target the protein to the nucleus in reponse to different stimuli. If PKKPAGR represented one of multiple NLSs in CBF1, then deletion of the PKKPAGR motif alone might not eliminate import into the nucleus to a detectable level; however, simultaneous deletion of an additional motif would affect this process, as it has been observed in the case of the thyroid transcription factor-1 and HMG-containing proteins (Sudbeck and Scherer, 1997; Christophe-Hobertus *et al.*, 1999). As a first approach to determine whether PKKPAGR is one of multiple signals in CBF1 that are required for transport across the nuclear membrane, 5' deletions of CBF1 ORF fused to GFP:GUS were generated and transformed into *Arabidopsis* plants for localization studies.

5' deletions of *CBF1* ORF were designed by analysis of the protein sequence. No obvious nuclear localization motif could be identified in CBF1 based on the current understanding of NLSs. In fact, the most commonly described nuclear localization signal rich in basic residues, does not represent the consensus motif for nuclear localization, and there have been numerous examples of nuclear import through novel targeting signals (Poon and Jans, 2005). Consequently, the following 5' deletions were chosen based on motif conservation or on the presence of known domains in CBF proteins (Figure 2.7A): deletion of residues 1-32 up to the PKKPAGR motif (construct 802), deletion up to the C-terminal end of the PKKPAGR motif (construct 801), deletion of the DNA-binding domain as a whole based on the presence of positive charges throughout the domain (construct 800), and deletion of five additional residues that eliminated the DSAWR motif (construct 799), based on its conservation in the CBF family of proteins. *Arabidopsis* plants overexpressing each of these constructs were generated, and overexpression of the transgene was confirmed by northern blot analysis for two to four representative lines per construct (Figure 2.7B).

To determine the subcellular localization of the different CBF1:GFP:GUS chimeras, fluorescence images were captured by laser confocal microscopy by analysis of the root tips of 6-7-day-old *Arabidopsis* seedlings grown at 22°C. Deletion of amino acids 1-32 (construct 802), and of amino acids 1-47 (construct 801) did not appear to significantly affect nuclear localization of CBF1:GFP:GUS, confirming that the PKKPAGR motif is not essential to this function (Figure 2.9). However, when the AP2/EREBP DNA-binding domain (construct 800) or the DNA-binding domain and the DSAWR motif (construct 799) were deleted, there was a detectable change in the cellular

distribution of the proteins, whereby GFP fluorescence could be detected not only in the nucleus but throughout the cell. These observations indicated that the protein could no longer be actively imported into, or retained in the nucleus.

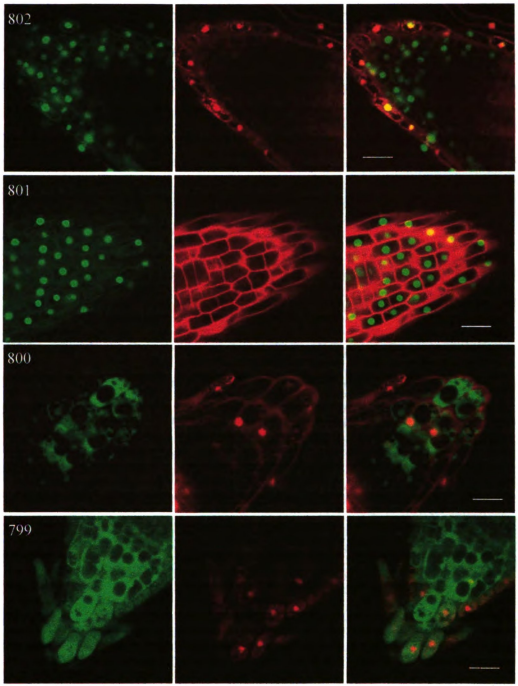


Figure 2.9. Localization studies of 5' deletions of CBF1:GFP:GUS chimeras.
Laser confocal images of root tips of *Arabidopsis* plants overexpressing various CBF1:GFP:GUS chimeras.
Left, GFP fluorescence (green); center, Propidium Iodide fluorescence (red); right, merged images (yellow color indicates overlap of red and green fluorescence). All constructs described in Figure 2.6.

To verify that the fluorescence detected *in planta* indeed corresponded to the full-length proteins expressed, and was not the result of unpredicted protein truncation or degradation, total protein extracts were analyzed by western blot analysis. In all the samples tested, it was possible to detect specific signals in which the observed molecular weight corresponded approximately to the expected size for that construct, indicating that the fluorescence detected by confocal imaging resulted from the localization of the indicated proteins (Figure 2.10).

Taken together, these results indicate that efficient nuclear localization of CBF1:GFP:GUS can occur under warm conditions when the PKKPAGR region has been deleted. The exact nature of the nuclear localization signal for CBF1 has yet to be determined. The analysis presented here suggests that the AP2/EREBP DNA-binding domain contains the signal required for efficient nuclear targeting and/or retention of CBF1. Additional analysis will be required to identify the specific region(s) or residues responsible for this process.

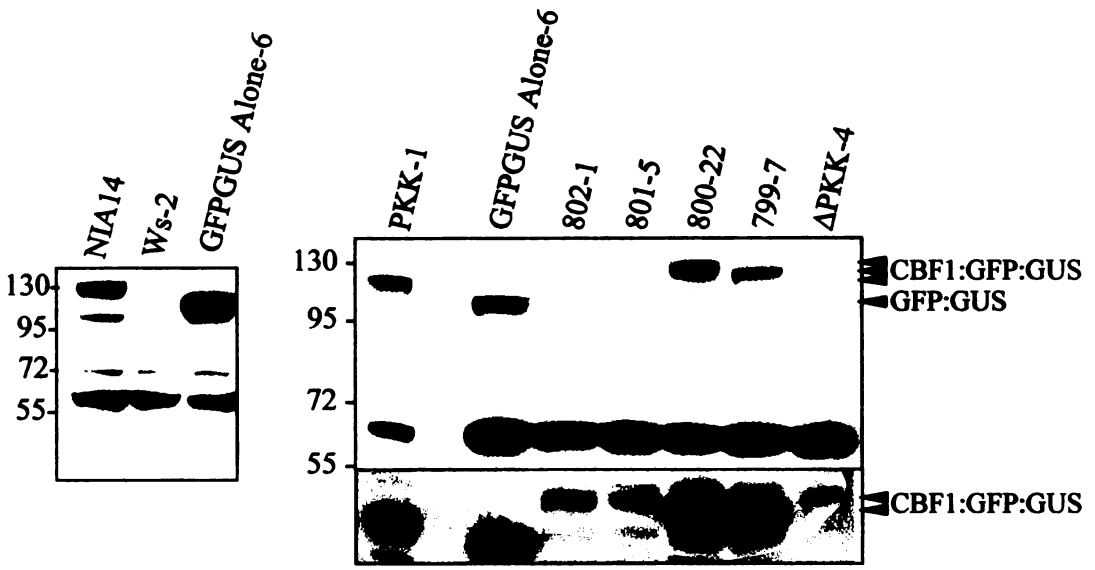


Figure 2.10. Western blot analysis of total protein extracts from *Arabidopsis* seedlings overexpressing different CBF1:GFP:GUS constructs or GFP:GUS alone.
Total protein extracts were prepared from 12-day-old seedlings grown at 22°C.
Monoclonal antibody raised against GFP was used for protein detection of 70 ug of total protein extracts.
Top and bottom blots shown on the right represent shorter and longer exposures of the same film, respectively.
The two bands detected at 55 and 72 KDa represent cross-reacting bands, as they can be detected in the protein sample from the non-transgenic plants represented by Arabidopsis Ws-2.

Role of the PKKPAGR motif in DNA binding.

We hypothesized that the PKKPAGR motif might play a role in CBF1 DNA binding activity based on the proximity of this motif to the AP2/EREBP domain and the presence of positively charged residues that might provide a favorable electrostatic interaction with the negatively charged DNA. This hypothesis was tested *in vitro* using electrophoretic-mobility shift assays to measure the binding activity of wild type and mutated CBF1 proteins.

To conduct this analysis, recombinant full-length CBF1 proteins harboring the same alanine mutations within the PKKPAGR region previously described, were fused to a histidine tag (6xHis-CBF1$_{FL}$) for expression and purification from *E. coli* (Figure 2.11A). 6xHis-CBF1$_{FL}$ proteins were tested in a binding reaction in which 1 ng of ^{32}P-radiolabeled DNA probe containing the *CRT/DRE* binding site from the *COR15a* promoter was mixed with 200 ng of each recombinant protein (Figure 2.11B). Equal protein loading was tested by analyzing diluted protein aliquots directly taken from the binding reaction by western blotting, using a α-CBF1 antibody (Figure 2.11C). As previously reported (Stockinger *et al.*, 1997), wild type CBF1 could cause a band shift of a 20-bp fragment containing a monomer of the *CRT/DRE* element. The shift was specifically abolished in the presence of unlabeled competitor DNA. The same unlabeled DNA fragment carrying point mutations was unable to compete for binding, indicating that CBF1 protein specifically binds the *CRT/DRE cis*-acting element. Unlike the wild type protein, none of the mutant proteins were able to efficiently bind and cause a shift of the ^{32}P-labeled probe.

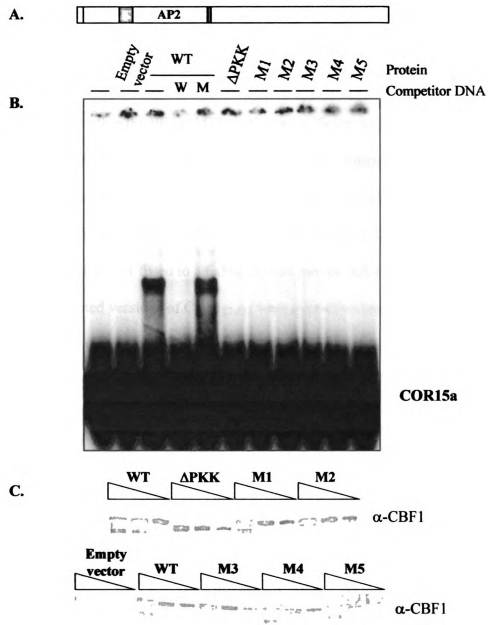


Figure 2.11. Triple alanine mutations abrogate the binding activity of a full-length 6xHis-CBF1 protein *in vitro*.
A. Schematic illustration of 6xHis-CBF1 constructs. Dark grey boxes, signature sequences; light grey box, 6xHis tag; AP2, AP2/EREBP DNA-binding domain.
B. EMSA. 200 ng of each recombinant protein were used in a 12 ul binding reaction containing 0.5 ng of radiolabeled *CRT/DRE* element from *COR15a* promoter. Where indicated, 100 ng of unlabeled competitor DNA were added. W and M, wild type and mutated competitor DNA.
C. Recombinant 6xHis-CBF1 proteins were analyzed by Western blot analysis using an antibody raised against full-length CBF1. Three different protein amounts (50, 25 and 12.5 ng) were tested.

These data showed that the presence of an intact PKKPAGR sequence is crucial to direct CBF1 to a *CRT/DRE*-containing element *in vitro*. The failure to bind could result from improper folding, loss of side chain interaction, or both.

To test whether the PKKPAGR motif was sufficient to impart DNA binding activity to the AP2/EREBP domain in the context of this domain alone, a new set of constructs was designed, in which only the CBF1 region including the DBD and the two flanking signature sequences ($CBF1_{27\text{-}112}$) was used (Figure 2.12A). Since purification of soluble full-length CBF1 fused to a 6xHis tag had proven difficult, wild type and PKKPAGR mutated versions of $CBF1_{27\text{-}112}$ were over-expressed as a translational fusion to Maltose Binding Protein (MBP), which has been shown to increase protein solubility when over-expressed in *E.coli* (Guana *et al.*, 1987).

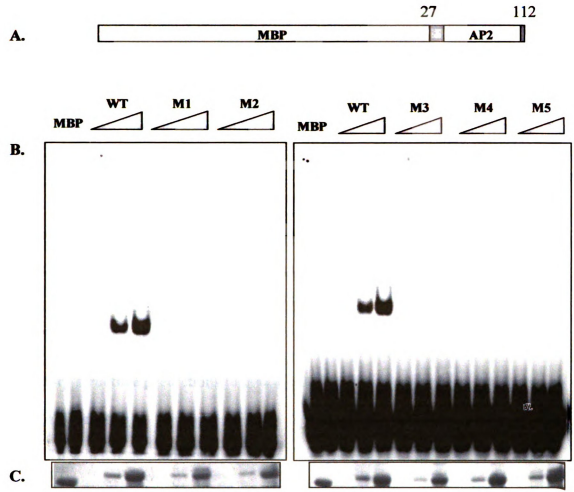


Figure 2.12. Triple alanine mutations abrogate the binding activity of MBP-CBF_{27-112} proteins *in vitro*.
A. Schematic illustration of MBP-CBF1 constructs. Dark grey boxes,signature sequences; the MBP tag and the AP2/EREBP DNA-binding domain are indicated.
B. Electro mobility shift assays. Increasing amounts (0.2, 2.0 and 15 ug) of each recombinant protein were used in a 25 ul binding reaction in the presence of 0.5 ng of radiolabeled *CRT/DRE* element from *COR15a* promoter.
C. Western blot analysis of MBP alone and the different MBP-CBF1 proteins, reflecting the increasing amounts of the recombinant proteins used in the electrophoretic-mobility shift assay (anti-MBP antibody).

As previously observed in the context of the full-length wild type protein, MBP-$CBF1_{27\text{-}112}$ was able to cause a specific shift of the labeled *CRT/DRE* probe (shown later, Figure 2.24), indicating that wild type $CBF1_{27\text{-}112}$ was sufficient for recognition of the *CRT/DRE* promoter element. On the contrary, mutations in the PKKPAGR region caused a dramatic reduction in binding affinity, as shown in the titration experiment of Figure 2.12B.

In summary, a CBF1 fragment composed of the AP2/EREBP DNA-binding domain and the two flanking signature sequences is sufficient to recognize a target *CRT/DRE* promoter element. This activity is lost in CBF1 protein when the PKKPAGR signature sequence is mutated.

Secondary structure predictions to investigate the structural preferences of the PKKPAGR motif.

To gain a better understanding of how mutations within the PKKPAGR region might affect DNA binding by CBF1, it is important to know the structural preferences and spatial orientation of the native region relative to the target DNA. This information will help to discriminate whether the contribution of PKKPAGR to binding is indirect, by providing the proper folding required to fit on the *COR* gene promoters, direct, due to the presence of essential side chain interactions that make DNA contacts, or both. To obtain insights into the structural propensity of the PKKPAGR motif, we conducted secondary structure prediction studies.

The search was carried out using a suite of web-based prediction servers that are rated among the most accurate according to the independent evaluator EVA (Eyrich *et*

al., 2001). They are PredictProtein (Rost and Liu, 2003), Psipred (McGuffin *et al.*, 2000), SAM-T02 (Karplus *et al.*, 2003), and SABLE (Adamczak *et al.*, 2005). CBF1 residues 1-60 were used as input, which includes the N-terminus of the protein, the PKKPAGR motif and the first 12 residues of the β-sheet DNA-binding domain. This sequence length represents the longest CBF1 fragment that restricts the automatic identification and alignment of sequence homologs to close relatives of CBF1, but not other AP2/EREBP proteins that do not contain the PKKPAGR motif. The secondary structure predictions generated by the individual methods were compared to identify consistent predictions across different methods (Table 2.1).

Table 2.1. Secondary structure predictions of the PKKPAGR motif.

	PROFsec[a]		SAM-T02[a]		SABLE[a]		PSI PRED[a]		Consensus[b]
P	C[c]	+++[d]	C	++	C	+++	C	+++	C
K	C	+++	C	++	C	+++	C	+++	C
K	C	+++	C	+++	C	+++	C	+++	C
P	C	+++	C	++	C	+++	C	++	C
A	C	++ [d]	T[c]	++	C	+++	C	+	C
G	C	++	T	++	C	+++	C	+	C
R	C	++	H[c]	0	C	+++	C	+	C
K		0[d]	H	+[d]	C	++	C	+	c/h
K		0	H	+	C	+		0	c/h
F		0	H	++	C	++	H	+	c/h
R		0	H	++	C	++	H	+	c/h
E		0	C	++	C	++	H	+	c/h
T		0	C	++	C	+++		0	c
R	C	++	C	++	C	+++	C	++	C
H	C	+++	C	++	C	+++	C	+++	C
P	C	++	C	+++	C	++	C	++	C

a PROFsec (Rost and Liu, 2003), SAM-T02 (Karplus *et al.*, 2003), SABLE (Adamczak *et al.*, 2005), and PSIPRED (McGuffin *et al.*, 2000) are the prediction methods used.
b Agreement between different prediction methods. Capital letters listed in the consensus column indicate agreement between the predictions, and lower case letters indicate differences.
c C, random coil conformation; H, helicity, T, turn.
d The number of pluses refers to the strength of prediction; one, two and three pluses indicate low, moderate and high confidence, respectively. Zero indicates that no prediction could be made for those specific residues.
Individual residues of the PKKPAGR motif are indicated in the first column.

Overall, we observed a high confidence in predicting the secondary structure for two clusters in the PKKPAGR motif, represented by residues PKKPAGR and RHP. Based on the agreement between different methods, those two clusters are likely to be in a coiled or irregular secondary structure (Table 2.1, very right column). On the contrary, the confidence of prediction dropped for the central region represented by residues KKFRET, with some helicity predicted by SAM-T02 and PsiPred, whereas SABLE and JPred predicted coil, and PROFsec returned no prediction. Therefore, no secondary structure could be assigned to the KKFRET residues within the PKKPAGR motif.

Since the above methods were principally developed to predict secondary structure over full-length protein sequences, a local predictor - Sequery - was used for further analysis of the KKFRET region (Collawn *et al.*, 1990; Craig *et al.*, 1998). Sequery predicts the secondary structures of local motifs (typically 4-6 residues) by identifying similar sequences in the Protein Data Bank, a database of experimentally determined 3D structures. Sequery matches the query tetrapeptide to sequences in the database and analyzes how often they occur in a particular secondary structure. To avoid statistical bias due to the over-representation of structures from the same super-families, a low-homology subset of protein chains with $\leq$25% sequence identity was generated. Tetrapeptides within the N-flanking region of CBF1 were designed by progressively sliding one residue at a time to cover the whole PKKPAGR region, resulting in 14 motif searches, identified as exact matches (Table 2.2, left box). To ensure a significant number of matches, we generated substitution patterns based on conservative substitutions and the consensus motif described for the PKKPAGR motif (Jaglo *et al.*, 2001). These tetrapeptide substitutions were classified as close matches (Table 2.2, very left column).

Table 2.2. Sequery analysis of the 3D structures observed for tetrapeptide sequences in the PKKPAGR region of CBF1.

Exact Matches[a]						Close Matches[a]					
Query[b]	H (%)[e]	E (%)[e]	T (%)[e]	I (%)[e]	Hits[f]	Query[c]	H (%)	E (%)	T (%)	I (%)	Hits
CPKK	N/A	N/A	N/A	N/A	0	(CS)(PS)(KR)(KR)	30	0	13	57	23
PKKP	0	0	0	100	2	(PS)(KR)(KR)(PS)	0	3	23	74	39
KKPA	0	0	0	100	4	(KR)(KR)(PS)A	31	2	10	57	42
KPAG	14	0	0	86	7	(KR)(PS)AG	21	0	9	70	34
PAGR	0	0	22	78	9	(PS)AGR	14	0	19	67	21
AGRK	14	14	0	72	7	AGR(KR)	25	5	20	50	20
GRKK	38	0	0	62	8	GR(KR)(KR)	25	5	12	58	64
RKKF	56	0	22	22	9	R(KR)(KR)F	51	6	8	35	49
KKFR	75	0	0	25	4	(KR)(KR)F(KR)	56	12	6	26	34
KFRE	100	0	0	0	6	(KR)F(KR)(ED)	56	2	7	35	73
FRET	100	0	0	0	3	F(KR)(ED)(TS)	47	3	21	29	38
RETR	33	17	17	33	6	(KR)(ED)(TS)(KR)	30	7	12	51	67
ETRH	33	0	0	67	3	(ED)(TS)(RK)H	36	0	36	28	14
TRHP	0	0	0	100	2	(TS)(RK)(HRK)P	0	0	32	68	38

Triple alanine substitutions[a]					
Query[d]	H (%)	E (%)	T (%)	I (%)	Hits
RKKF → RAAA	83	0	0	14	22
KKFR → AAAR	86	3	0	14	29
FRET → FAAA	64	5	5	26	22
TRHP → TAAA	72	14	7	7	29
RHPI → AAAI	88	8	4	0	24

a,b,c,d Secondary structure predictions obtained by using tetrapeptide motifs covering the whole PKKPAGR region. Those motifs represent either the exact sequence (exact matches box, b), conservative substitutions based on the consensus motif described for the PKKPAGR (Jaglo *et al.*, 2001) (close matches box, c), or alanine substitutions described in this study (triple alanine substitutions box, d).
e H, E, T, and I indicate helical, extended, turn and irregular structures, respectively.
f Number of matches retrieved by Sequery. Each tetrapeptide motif was used to search a non-redundant database of low-homology (<25% pair-wise sequence identity) 3D structures from the Protein Data Bank (Wang and Dunbrack, 2003).
The area of the table highlighted in grey indicates tetrapeptides showing strong helical propensity according to Sequery results.

Sequery predictions for the PKKPAGR region were generally consistent with the secondary structure prediction results obtained with the methods described above and illustrated in Table 2.1. In fact, in both cases the PKKPAG and RHP clusters were predicted to be coiled or irregular in structure (compare very right column of Table 2.1 with the two left columns of Table 2.2). In addition, Sequery results revealed a strong tendency for a helical structure in the RKKFRET region: approximately 40-50% of the close matches occurred in helices. The trends were even stronger when the exact matches were used as input, with values ranging from 50 to 100%. To assess possible structural perturbations by replacement with three consecutive alanines in the RKKFRET region, Sequery analysis for these mutations was also conducted (lower section of Table 2.2, see prediction for helical structure). Each of these substituted sequences was present in helical conformation 64-88% of the time, which was expected based on the helical-promoting effect of poly-alanines (Chou and Fasman, 1978; Zhang *et al.*, 1991; Chakrabartty *et al.*, 1994). This result suggested that triple alanine substitutions within the RKKFRET region caused little perturbation of the native structure.

In summary, secondary structure prediction studies suggested that two clusters in the PKKPAGR motif, represented by PKKPAG and RHP, are likely to be irregular in nature. Based on the helix-inducing propensity of poly-alanines (Chou and Fasman, 1978; O'Neil and DeGrado, 1990; Chakrabartty *et al.*, 1994), it is likely that alanine substitutions in these regions will affect the native structure. Therefore, it is not possible to discriminate whether residues in this two regions are essential for protein folding or are directly involved in interaction with the DNA, based on these substitutions. On the

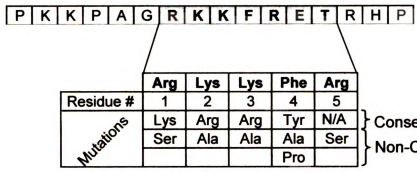
contrary, the central RKKFRET region - rich in basic residues - showed strong helical preference according to the predictor Sequery, suggesting that triple alanine mutations in this region do not affect protein folding. Based on these observations and the notion that patches of positively charged groups are energetically more stable at the protein surface (Kim *et al.*, 2005), we propose loss of binding due to mutations in the RKKFRET region of CBF1 are likely due to a loss of side chain-DNA interactions required to bind the *CRT/DRE* element.

Specific amino acids within the PKKPAGR motif have important roles in binding to the *CRT/DRE* promoter element.

Electrophoretic-mobility shift assays and secondary structure predictions suggested that residues within the RKKFRET motif might participate to DNA binding by contributing important side chain interactions with the target *CRT/DRE* promoter element. To test the contribution of individual residues in the predicted helical motif, RKKFRET, we tested the role of specific amino acids by mutational analysis. The main focus of our investigation was to assess the relevance of the positively charged side chains in the DNA recognition by CBF1, since they are usually enriched in DNA-binding domains (Nadassy *et al.*, 1999; Luscombe and Thornton, 2002). These residues could favor the formation of a stable protein-DNA complex through electrostatic interactions with the negatively charged phosphates of the DNA backbone, or participate in direct recognition of DNA bases. In addition, we were interested in testing the role of the conserved phenylalanine. This residue is not generally enriched at the protein-DNA interface; however, when it is present, it can form hydrophobic contacts in complexes

with bent DNA, like in the case of the TATA-binding proteins [see (Kim *et al.*, 1993) for reference].

Mutations of the residues mentioned above were designed either to preserve the side chain chemistry (conservative mutations) or to alter it (non-conservative mutations) (Figure 2.13A, upper panel). In all cases but one, one additional constraint was applied: the mutations should be compatible with the predicted helical structure of the RKKFRET region, to exclude local protein unfolding in case reduced binding activity was observed for those mutant proteins. The Phe4→Pro mutation was designed to test the effect of a helix-breaking residue on the stability of the protein-DNA complex. Based on these criteria, a reciprocal Arg↔Lys substitution and Phe→Tyr were chosen as conservative mutations; Arg→Ser, Lys→Ala, and Phe→Ala represented non-conservative substitutions. Preserving the positive charge, in the case of Arg and Lys, and the aromatic ring in the case of Phe→Tyr, may result in a negligible effect on DNA binding; on the contrary, the non-conservative mutations would directly test the importance of the positive charge, in the case of Arg and Lys, or the contribution of the aromatic ring, in the case of Phe. Recombinant proteins carrying point mutations were tested by gel mobility shift assays in the presence of *CRT/DRE* promoter elements from *COR15a*, *COR6.6*, and *COR78* (Figure 2.13A, B, C).



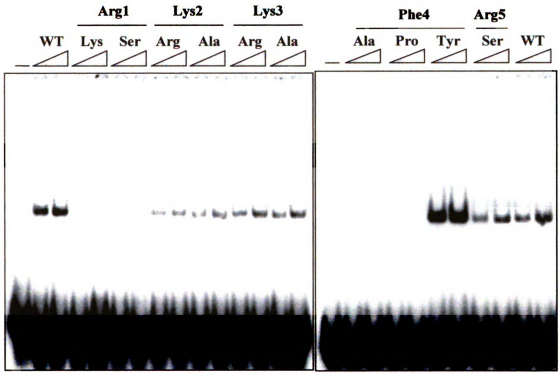


Figure 2.13 A. Effect of point mutations in the predicted RKKFRET helical region of CBF1 on binding activity of the protein to the *COR78* gene promoter.
Top panel. Substitution table indicating the point mutations designed within the RKKFRET region of CBF1. Conservative and non-conservative mutations are indicated. H and C below the PKKPAGR motif indicate the predicted helical and coiled structures, respectively.
Bottom panel. Gel mobility shift assays indicating the binding activity of different MBP:CBF1$_{27\text{-}112}$ proteins. DNA binding reactions were carried out in a 15 ul reaction containing increasing amounts (300 and 600 ng) of each recombinant protein in the presence of 0.5 ng of radiolabeled *CRT/DRE*-containing probe from *COR78* promoter.

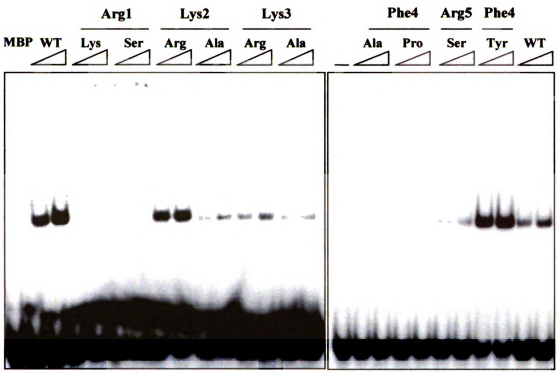


Figure 2.13 B. Effect of point mutations in the predicted RKKFRET helical region of CBF1 on binding activity of the protein to the *COR15a* gene promoter.
Gel mobility shift assays indicating the binding activity of different MBP:$CBF1_{27-112}$ proteins. DNA binding reactions were carried out in a 15 ul reaction containing increasing amounts (300 and 600 ng) of each recombinant protein in the presence of 0.5 ng of radiolabeled *CRT/DRE*-containing probe from *COR15a* promoter.

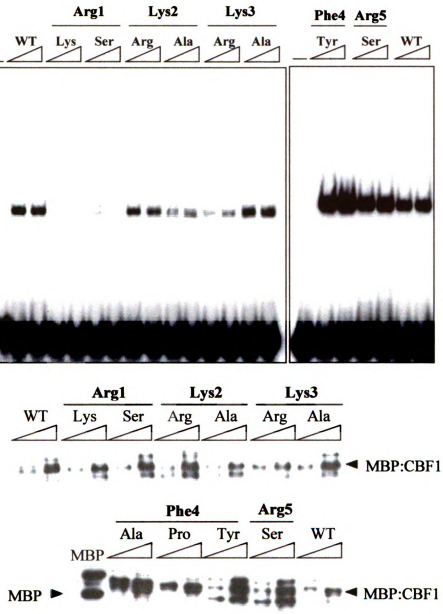


Figure 2.13 C. Effect of point mutations in the predicted RKKFRET helical region of CBF1 on binding activity of the protein to the *COR6.6* gene promoter.
Top panel. Gel mobility shift assays indicating the binding activity of different MBP:CBF1$_{27\text{-}112}$ proteins. DNA binding reactions were carried out in a 15 ul reaction containing increasing amounts (300 and 600 ng) of each recombinant protein in the presence of 0.5 ng of radiolabeled *CRT/DRE*-containing probe from *COR78* promoter. Bottom panel. Western blot analysis (α-MBP antibody) showing protein loading. Twenty-five and 50 ng of each recombinant protein were analyzed.

The binding results observed at the three different promoters showed that most of the residues tested play an important role in DNA binding by MBP:CBF1$_{27\text{-}112}$. The most striking observation was that the two mutations in Arg1 (Arg→Lys and Arg→Ser), and two of the three mutations on Phe4 (Phe→Ala and Phe→Pro) caused almost complete loss of DNA binding to the three promoters tested. The loss of binding by substitution of Arg1 with Lys, of similar side-chain length and charge, was significant and suggested that the interaction of this residue with DNA is possibly occurring through the DNA base. Loss of binding for the Phe4→Ala substitution suggested that the aromatic ring in this residue is essential for the interaction with the DNA. Pro is a less conservative mutation that is likely to interrupt helicity, and therefore loss of DNA binding was expected. The conservative Phe→Tyr mutation did not cause a significant change in binding activity; in fact, this mutated protein displayed an increase in DNA binding affinity compared to the wild type protein. This result suggests that the aromatic ring contributes favorable interactions with DNA bases; in addition, the terminal hydroxyl group, with its H-bonding donating accepting capacity, could add H-bonds between the side chain and either the DNA bases of with part of the backbone. Overall, substitutions at Lys2 and Lys3 showed a moderate effect on DNA binding, and less effect when the mutant was electrostatically conservative (Lys→Arg) than when the side chain was truncated (Lys→Ala). These data suggested that the contribution of those residues to DNA binding is less important, and they could likely be involved in salt bridging to the phosphate backbone of the DNA. Finally, Arg5→Ser substitution showed no significant effect on DNA binding, indicating that this residue is not important for this function.

Interestingly, we could observe some differences in the binding affinity of the mutant proteins to different promoters. For instance, Lys→Arg mutation in Lys2 affected binding to the *CRT/DRE* from the *COR78* promoter more strongly than the *COR15a* promoter (compare Figures 2.13A and B), and mutant CBF1 proteins carrying the Lys→Ala mutation displayed different binding affinities to the three promoters. It must be noted, however, that binding to the *COR78* and the *COR6.6* promoters was tested only one time. Additional technical replicates will elucidate whether these results are reproducible.

Taken altogether, these data indicate that specific amino acids within the PKKPAGR motif play a significant role in CBF1 binding to its target *CRT/DRE* promoter element. Based on the moderate loss of DNA binding for substitutions at Lys2 and Lys3, it is likely that these residues are participating in electrostatic interactions at the protein-DNA interface through their positively charged side chains. The more dramatic effects observed for substitutions at Arg1 and Phe4 suggest that these amino acids might provide side chain interactions that are base-specific.

Side chain conservation in AP2/EREBP-like proteins that use their flanking sequences to stabilize DNA binding.

An important role of amino acid sequences flanking DNA-binding domains in DNA recognition has been previously reported for a few families of proteins, such as the homeodomain and high-mobility group families (Lnenicek-Allen *et al.*, 1996; Wolberger, 1996; Dragan *et al.*, 2004). The importance of specific amino acids in the PKKPAGR motif in DNA binding, made us wonder whether this phenomenon is conserved in DNA-

binding proteins that, like the AP2/EREBP proteins, use their β-sheets to interact with their target DNA. We addressed this question by searching the Protein Data Bank for structures similar to the three β-stranded domain described for AtERF1 (Allen *et al.*, 1998). Any consensus in this region would further support the idea that these residues sit close to the DNA, and participate directly to the binding activity of the protein, possibly by a common mechanism.

The atomic coordinates for the NMR structure of the AP2/EREBP domain of AtERF1 [PDB id. 1gcc (Allen *et al.*, 1998)] were used as a query to search the comprehensive Protein Data Bank (http://rcsb.org), which includes 3D structures obtained from crystallography and NMR studies. The method of choice was Dali (Holm and Sander, 1993), a well-established website for protein structure comparison. The server returned a list of 26 hits that shared significant structural similarity (Z-score >2.0) with 1gcc, including ERF1 itself (Table 2.3). Six of the other 25 structures did not contain any structural information on the sequences flanking the β-sheet at its N-terminus, and were discarded (data not shown). Of the remaining hits, the analysis was focused on nucleic acid-binding proteins, represented by ten structures.

Table 2.3. Summary of Dali analysis showing significant structural matches of nucleic acid binding proteins with the AP2/EREBP DNA binding domain of AtERF1.

PDB id.[a]	*Z-score*[b]	*% Ide*[c]	*Nucleic acid-binding proteins containing a β-sheet*	*Nucleic Acid bound in structure*[d]	*Flanking sequence*[e]
2gcc	11.8	100	*AtERF1* fragment	√	N/A
2bn8	4.2	16	Cell division activator (*E. coli*)	-	9L
1kjk	3.9	10	Viral protein – integrase	√	7L
1z1b	3.8	9	DNA integrase mutant	√	2L
2bb8	3.4	10	Tn916 integrase	√	9L
2d35	2.8	13	Cell division activator (*E. coli*)	√	7L
1rth	2.8	15	HIV-1 reverse transcriptase	√	16L
1mkm	2.8	8	*T. maritima* 0065, IclR family	-	2L12H
1bhi	2.5	8	DNA-binding protein cre-bp1	-	N/A
1di2	2.3	9	*X. laevis* dsRNA binding protein A	√	1L12H3L
1ytb	2.2	6	*S. cerevisiae* TATA-box binding protein	√	H/L

2gcc (Allen *et al.*, 1998); 2bn8 (Chen *et al.*, 2005); 1kjk (Wojciak *et al.*, 2002); 1z1b (Biswas *et al.*, 2005); 2bb8 (Connolly *et al.*, 1998); 2d35 (unpublished; http://www.rcsb.org/pdb/explore/explore.do?structureId=2D35); 1rth (Ren *et al.*, 1995); 1mkm (Zhang *et al.*, 2002); 1bhi (Nagadoi *et al.*, 1999); 1di2 (Ryter and Schultz, 1998); 1ytb (Kim *et al.*, 1993).

a Protein Data Bank identifier.

b Degree of significant structural similarity (Z-score >2.0), according to Dali Cα matching.

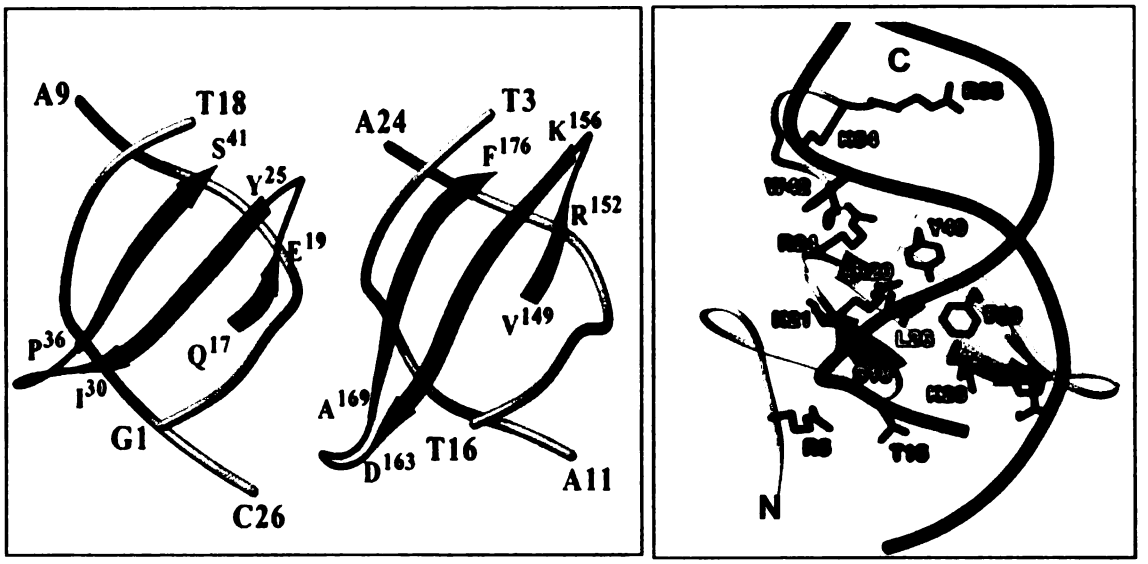
c percentage of sequence identity over positions of structural homology.

d check marks under "Nucleic Acid bound in structure" indicate that the 3D structure of the protein is in complex with either DNA or RNA.

e secondary structure of sequences N-terminal to their β-sheets, where available. H=helix; L=loop; N/A=structure of this region of the protein is not present or not defined in the crystal or NMR structure.

The similarity between three of them, represented by PDB identifiers 1kjk and 1z1b (representing the same protein), and 2bb8 and the AP2/EREBP domain of ERF1 had previously been reported (Connolly *et al.*, 1998; Wojciak *et al.*, 2002; Biswas *et al.*, 2005; Chen *et al.*, 2005). An extensive comparative analysis of the DNA-bound structure of ERF1 and Tn916 integrase (Wojciak *et al.*, 1999) has previously indicated significant similarities in the mode of DNA recognition by these β-sheet containing proteins (Connolly *et al.*, 2000). This recognition involves a similar orientation of the β-sheet relative to major groove of the target DNA, and conserved protein-DNA contacts within the AP2/EREBP-like structure (Figure 2.14).

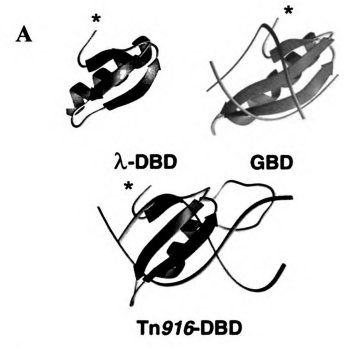
We wanted to extend our analysis to the N-flanking regions, in search of protein-DNA complexes in which both the β-sheet and the N-flanking motif were involved in DNA-binding. Among the AP2/EREBP-like structures described in Table 2.3, lambda integrase and Tn916 integrase displayed this binding mode. NMR studies showed that the secondary structure of the N-flanking regions of these proteins differs from the helical structure predicted for the PKKPAGR region RKKFRET. This N-terminal region is represented by a loop in Tn916 integrase and by an unstructured segment in lambda integrase. However, despite the structural differences, some compelling similarities were found when we analyzed the amino acid composition of these motifs (Figure 2.15A, B).



Connolly et al., 2000

Figure 2.14. Three-dimensional structures of proteins that use three-stranded β-sheets to recognize the major groove of DNA.
Left. The three-stranded β-sheet complexes of the DNA binding domain from Tn916 integrase (left) and AtERF1 (right). The Figure emphasizes the conserved orientation of the sheet structures relative to the DNA. The conserved β-sheets and the DNA backbone are dark and light grey, respectively.
Right. DNA-binding domain of Tn916 in complex with DNA. Labeled residues were mutated and the DNA binding activity of the mutated protein was tested by electrophoretic mobility shift assay. Arg5, in the N-flanking region of Tn916, contributes to DNA binding affinity. DNA and protein backbones are shown as tubes, and the three-stranded β-sheet is represented by ribbons. N and C, N and C termini of the DNA-binding domain including the N-flanking region

.



B

```
                    +++ +- +
CBF1   (1gcc)  PKKPAGRKKFRETRHPI----YRGVRQR
λ-Int  (1kjk)      MGRRRSHERRDLPP---NYRNNGY
Tn916  (1tn9)      MEKRR--DNRGRILKTGESQRKDG
                     ::: .: *
```

Figure 2.15. Comparison of β-sheet-flanking sequences experimentally shown to be important in DNA binding.
A. Schematic of the NMR derived structure of the Integrases Tn916 and Lambda (λ-DBD and Tn916 DBD, respectively) and ERF1 (GBD) as described by Wojciak *et al.* (1999). The NMR structures of ERF1 and Tn916 were solved in complex with the DNA. The antiparallel β-sheet is blue, the C-terminal helix is red. Asterisks N-flanking motifs.
B. Alignment of N-flanking regions from Lambda and Tn916 integrases and ERF1. Alignment is based on optimizing charged residue correspondence. Underlined residues are known to be important for binding. (*), exact conservation of residues; (:), strong conservation; (.), some degree of conservation in the residue chemistry. (+) and (–), positively and negatively charged residues, respectively. Area highlighted in grey, residues at the 5' end of the beta sheets. 1gcc (Allen et al., 1998); 1kjk (Wojciak et al., 2002); 1b69 (Connolly et al., 2000).

.

Not only are the N-terminal residues proximal to or directly contacting the minor groove of the DNA in the NMR structure (Figure 2.15A), mutational analysis of residues in the N-flanking regions of these two proteins indicated that one or more positively charged arginines play important roles in DNA binding. For instance, mutation of an arginine residue in the N-flanking sequence of Tn916 integrase causes a significant loss in DNA-binding affinity (Connolly *et al.*, 2000), whereas deletion of two arginines in the β-sheet flanking region of lambda integrase could abrogate DNA-binding (Wojciak *et al.*, 2002). When we analyzed the sequence composition of those motifs we observed that the central core of those N-flanking regions includes a cluster of positively charged residues, and one negatively charged residue.

In summary, we could identify a subgroup of β-sheet DNA-binding proteins in which one or more residues present in the N-flanking region provide a significant contribution to DNA-binding affinity. The presence of a positively charged cluster, their proximity to the DNA and their contribution to DNA binding, is consistent with our findings and supports the idea that the PKKPAGR motif provides a similar contribution at the *CRT/DRE* promoter site.

A working model for DNA binding by CBF1.

Extensive mutagenesis and structural analysis of the PKKPAGR motif and comparison to AP2/EREBP-like proteins that use their N-flanking regions to contact the DNA, prompted us to describe CBF1 interaction with the DNA by computational modeling. The experimental results obtained by point mutations within the RKKFRET predicted helix drove the model design. The constraints used in modeling the orientation of the helix relative to CBF and the DNA were the following: (1) The C-terminus of the RKKFRET helix had to lie within two residues of the N-terminus of the β-stranded region of CBF; (2) Arg1 and Phe4 residues, critical to DNA binding, were placed in close proximity to the DNA, given that mutations to these positions have drastic effects on DNA binding. (3) The helix should pack well against the DNA and CBF1. The other features described below were derived from the placement of the RKKFRET helix to meet these three constraints. The NMR structure of DNA bound to the CBF1 homolog ERF1 (PDB accession 1gcc), was used as template to model the interaction between the DNA minor groove and the RKKFRET N-terminal flanking helix (Figure 2.16A). Figure 2.16B illustrates the working model of CBF1 bound to the DNA.

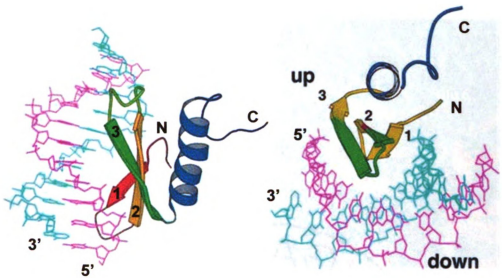
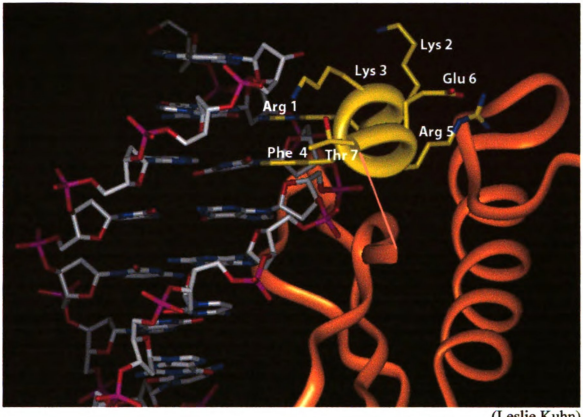


Figure 2.16A. NMR structure of the AP2/EREBP DNA-binding domain of ERF1 bound to its cognate DNA (Allen *et al.*, 1998).
Left and right figures represent two different views of the DNA binding domain in complex with its cognate promoter element. The three-stranded beta sheet is represented by ribbons numbered 1-3 that contact the DNA by reaching into its major groove. N and C termini of the AP2/EREBP domain are indicated by N and C, respectively. 5' and 3' end of the double-stranded DNA are indicated.



(Leslie Kuhn)

Figure 2.16B. Working model describing DNA-binding by CBF1. The NMR structure of DNA bound to the CBF homolog ERF1 (PDB id. 1gcc), was used as the basis to model the interaction between the minor groove of DNA and the RKKFRET N-terminal flanking helix (yellow ribbon and side chains) relative to the main chain of the β-stranded DNA-binding domain of CBF1 (orange ribbon). The DNA is shown in tubes, with carbons in white, nitrogens in blue, oxygens in red, and phosphorus in magenta. In the RKKFRET helix, Arg 1 and Phe 4 residues bind within the DNA minor groove, where Arg1 forms hydrogen bonds and Phe4 in involved in stacking interactions. The main-chain structure of the β-strand domain of CBF was modeled with SwissModel. Two additional residues, Arg and His, bridge between the C-terminus of the KKFRET and the N-terminus of the beta-stranded DNA-binding domain, as indicated by the orange line. This connection likely involves a slight reorientation of the first four residues of the beta-stranded region (orange ribbon), as observed in the structures of two DNA integrases (PDB id. 1b69 and 1z1b), which also use a β-stranded domain to bind DNA, with an N-terminal flanking region that enters the minor groove.

The AP2/EREBP domain and the RKKFRET region (orange and yellow, respectively) are placed against the DNA major and minor groove, respectively. The AP2/EREBP domain (shown in orange vertically oriented), is placed into the major groove, and its interaction with DNA forms a smooth pocket where the RKKFRET helix is packed. The proximity of this region to the DNA is also derived by the physical constraint of a short two-residue connection to the three-stranded β-sheet. Mutational analysis indicated that Arg1 and Phe4 are critically important for DNA binding. Arg1→Lys abrogated DNA binding, despite conserving side-chain length and charge. For the Phe mutations, Ala is helix-propensive and hydrophobic but less bulky than Phe, whereas Pro is a less conservative mutation and is likely to interrupt helicity. The loss of DNA binding upon these single-site mutations suggests that these residues interact intimately with the DNA, and both are positioned to bind within the minor groove. Phe→Tyr mutation didn't affect DNA binding. This is consistent with the phenyl moiety of Tyr maintaining favorable aromatic interactions with adjacent DNA bases, while its terminal hydroxyl group can both donate and accept hydrogen bonds either with bases in the groove of the DNA backbone. The phenyl moiety of Phe4 (or Tyr4) can make aromatic ring stacking interactions with adjacent bases in the groove, which have been reported when Phe interactions with DNA occur (Luscombe and Thornton, 2002; Morávek *et al.*, 2002) and this is likely to distort the local DNA conformation (Morávek *et al.*, 2002). Substitutions in the other residues of the RKKFRET motif resulted in a more moderate effect on DNA binding. Accordingly, Lys2 is shown salt bridging to the phosphate backbone and Lys3 is hydrogen bonding to a backbone ribose residue, consistent with mutations to these residues having moderate (non-base-specific) effects

on DNA binding, and less effect when the mutant is electrostatically conservative (Lys→Arg) than when the side chain is truncated (Lys→Ala). Arg5 and Glu6 are positioned such that they can form a salt bridge that stabilizes the helix, on the opposite side of the DNA. This is also consistent with Arg5→Ser mutation having little effect on DNA binding, and with the Ser mutation also being able to form a stabilizing intra-helix hydrogen bond with Glu6.

Based on the experimental constraints imposed on the model, the helical RKKFRET region can be placed close to the 3'end of the *CRT/DRE* motif - (G/A)CCGACNT. According to the model, Arg1 can make hydrogen-bond and pi-cation interactions with bases in the minor groove, corresponding to the region in which bases B11-B13 (from chain B) pair with C14-C16 (from chain C) in the 1gcc structure (Figure 2.17). This B11-B13/C14-C16 sequence follows the conserved consensus GCCGCC, representing the core sequence bound by ERF1 via its β-sheet. Interestingly, the corresponding region from the *CRT/DRE* promoter element contains a conserved thymine that has been shown to be required for specific binding of CBFs to *CRT/DRE*-containing promoter elements (Maruyama *et al.*, 2004; Sakuma *et al.*, 2006). The proximity of the RKKFRET region to this DNA base suggests that either Arg1 or Phe4 could direct specific recognition of the DNA by CBF proteins.

CBF cognate DNA (COR15A)	5'-	T	G	G	**C**	**C**	**G**	**A**	**C**	C	**T**	G -	3'
Residue # in Chain B of 1gcc		3	4	5	6	7	8	9	10	11	12	13	
ERF DNA in PDB 1gcc	5'-	T	**A**	**G**	**C**	**C**	**G**	**C**	**C**	A	G	C	3'
ERF DNA in PDB 1gcc	3'-	A	**T**	**C**	**G**	**G**	**C**	**G**	**G**	T	C	G	5'
Residue # in Chain C of 1gcc		24	23	22	21	20	19	18	17	16	15	14	
CBF cognate DNA (COR15A)	3'-	A	C	C	**G**	**G**	**C**	**T**	**G**	G	**A**	C	5'

Figure 2.17. Nucleotide sequences of the *cis*-acting elements in CBF and ERF proteins. The sequences shown for ERF DNA refer to the coding and complementary strand of DNA used in the NMR studies by Allen *et al.* (1998), in which the structure of the DNA-binding domain was solved in complex with the DNA. The conserved bases are shown in bold.

In summary, by integrating secondary structure predictions and experimental analysis, it was possible to describe CBF1 interaction with the DNA by computational modeling. We propose that the interactions of N-terminal flanking residues with the minor groove involve bases beyond those bound by the β-sheet, some of which might confer specificity for *COR* gene promoters relative to *ERE*.

FUNCTIONAL ROLE OF THE DSAWR MOTIF IN CBF1 ACTIVITY.

Mutations in the DSAWR sequence affect *COR* gene activation.

To determine whether the DSAWR motif plays a role in CBF1 function, *Arabidopsis* plants overexpressing *dsawr* mutant or wild type CBF1 transgenes were generated. Morphological analysis of those plants provided a first indication that residues within the conserved DSAWR motif might play a role in CBF1 function (Figure 2.3).

The role of the DSAWR motif in CBF1 activity was assessed by northern blot analysis as described for the PKKPAGR motif. Representative results are shown in Figure 2.18.

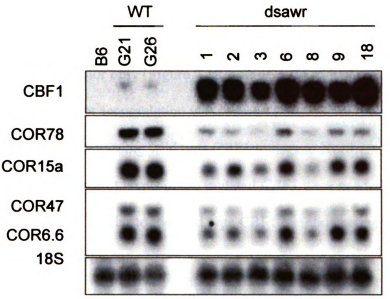


Figure 2.18. Accumulation of *CBF* and *COR* transcripts in *Arabidopsis* plants overexpressing wild type or *dsawr* CBF1.
Total RNA was isolated from warm-grown seedlings (22°C), and 5 ug were analyzed by Northern blotting. Northern blot analyses were performed using probes for *CBF1, COR6.6, COR15A, COR47, and COR78* transcripts.
B6, transgenic lines harboring the empty vector; WT, transgenic lines overexpressing wild type CBF1; dsawr, transgenic lines overexpressing CBF1 transgene carrying alanine substitutions in the DSAWR region. Numbers above lanes indicate independent transgenic lines.

Similarly to what was observed for the *pkkpagr* lines (Figures 2.4 and 2.5), Arabidopsis plants overexpressing *CBF1-dsawr* transgene still retained some ability to activate transcription at *COR* gene promoters, compared to the control plants (B6) harboring the empty vector only (Figure 2.18, compare B6 to independent *dsawr* lines). However, *COR* transcript accumulation was reduced when *dsawr* lines were compared to Arabidopsis lines overexpressing wild type CBF1 transgene Northern blot analysis showed that higher transcript levels of the mutated *CBF1* were needed to achieve *COR* transcript accumulation similar to the wild type CBF1 overexpressing lines (Figure 2.18, compare WT and *dsawr* lines).

Altogether these data indicated that alanine substitutions throughout the DSAWR motif of CBF1 affected its transcriptional activity.

CBF1 protein detection in *dsawr* lines.

To determine whether the *dsawr* mutation affected CBF1 protein accumulation, it was necessary to measure CBF1 protein levels *in planta*. The experiment was carried out by generating transgenic plants overexpressing wild type and mutated *CBF1* transgenes fused to a Myc tag. Protein levels were measured by western blot analysis, as described for the *pkkpagr* lines. CBF1 gene and protein expression levels are shown in Figure 2.19.

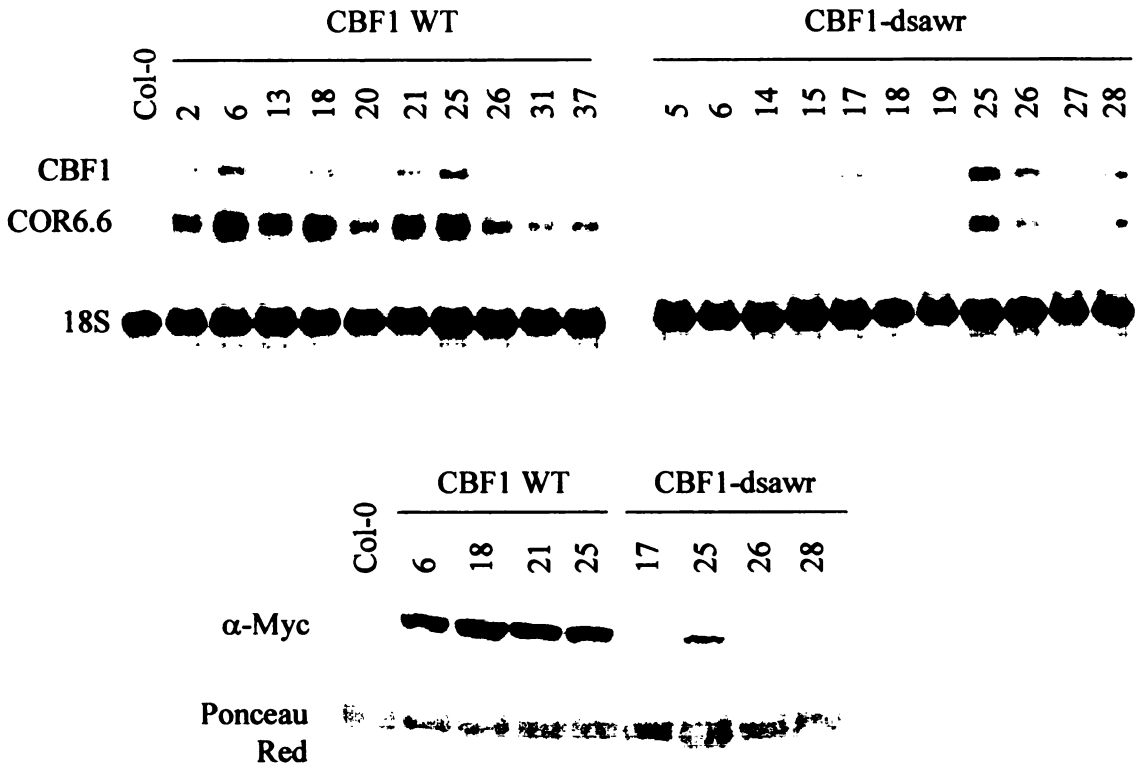


Figure 2.19. Northern and western blot analysis of *dsawr* plants and plants overexpressing wild type CBF1.
Top. Northern blot analysis of 35S::6xMyc:CBF1 transgenic lines. Total RNA was analyzed for CBF1 and *COR6.6*. Transcript amounts were normalized using the 18S loading control.
Bottom. Western blot analysis showing the presence of 6xMyc:CBF1 protein in the transgenic plants (detected with a monoclonal anti-Myc antibody).
WT and *dsawr*, transgenic plants overexpressing a 6xMyc tag fused to a wild type (WT) or mutated (*dsawr*) CBF1 transgene; Col-0, non-transgenic plants.

Several independent lines were analyzed for the expression of *CBF1* and *COR6.6*. Representative lines overexpressing similar levels of My-CBF1 were chosen for protein detection from plant extracts. Western blot analysis clearly indicated that *dsawr* plants were affected in CBF1 accumulation when compared to Arabidopsis plants overexpressing wild type CBF1 transgene at similar level (Figure 2.19, compare).

Taken together, these results indicate that the inability of the *dsawr* lines to induce *COR* gene expression can be explained, at least in part, by reduced protein accumulation.

Role of DSAWR in DNA binding activity of CBF1.

Based on its proximity to the AP2/EREBP DNA-binding domain, and by analogy to the PKKPAGR sequence, we postulated that the DSAWR sequence might play a role in *CRT/DRE* binding by CBF1. Gel mobility shift assays were conducted to compare the binding activity of the wild type and mutated protein.

Experiments were conducted as previously described for the PKKPAGR mutants. Binding reactions were set up by incubating wild type of mutated CBF1 proteins with $^{\gamma 32}$P-radiolabeled probe representing *CRT/DRE* elements from *COR6.6*, *COR15a*, and *COR78* promoters. In all cases, alanine substitutions greatly reduced the binding activity of CBF1, indicating that the DSAWR motif contributes to recognition (Figure 2.20).

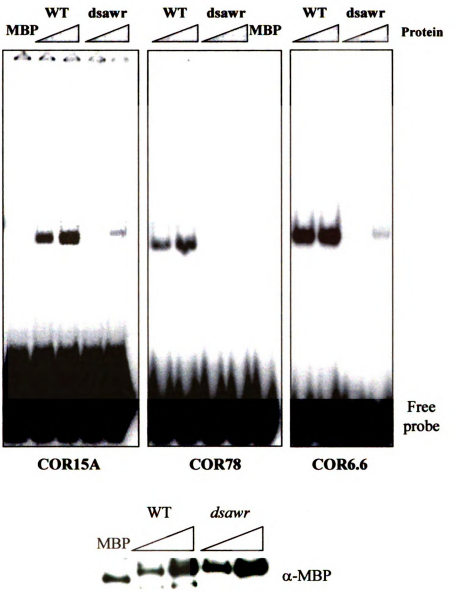
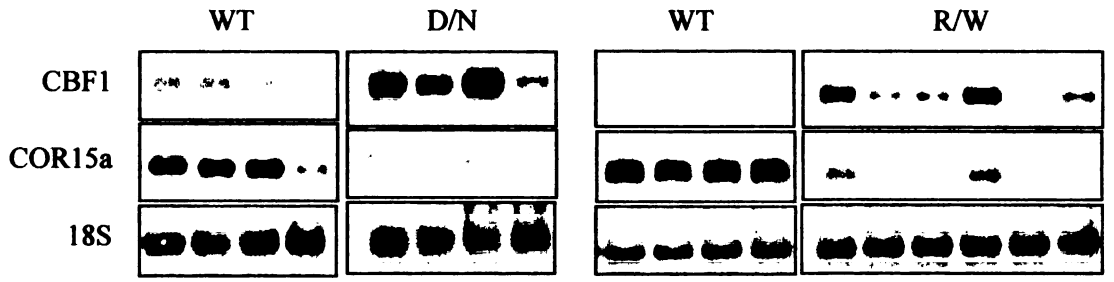


Figure 2.20. DNA binding activity of an MBP-CBF1$_{27\text{-}112}$ protein carrying alanine substitutions in the DSAWR region.
Upper panel. Gel shift mobility assay in which binding of MBP:CBF1$_{27\text{-}112}$ proteins to *COR6.6*, *COR15a*, and *COR78* is shown. 0.5 and 1.0 ug of protein were used in a 15.0 ul reaction in the presence of 0.5 ng of radiolabeled *CRT/DRE*-containing probes from *COR6.6*, *COR15a*, and *COR78* promoters.
Lower panel. Western blot analysis showing protein loading. 25 and 50 ng of each recombinant protein were analyzed.

Point mutations within the DSAWR region of CBF2 reveal the importance of an Asp residue in *CRT/DRE* recognition.

To better understand the contribution of the DSAWR region to DNA binding by the CBF proteins, analyzing the effect of point mutations can be informative. In a mutant suppressor screen designed to identify regulators of the CBF pathway that act through the *CBF* regulon, several mutations in the *CBF2* transgene were identified (Gilmour, unpublished). Two mutations fell in the DSAWR region of CBF2, Asp→Asn and Arg→Trp. Overexpression of those mutated transgenes in *Arabidopsis* produced plants that showed reduced accumulation of *COR* transcript compared to the plants overexpressing the wild type gene (Figure 2.21).



Sarah Gilmour, unpublished

Figure 2.21. Northern blot analysis of transgenic plants overexpressing wild type and mutated CBF2 transgenes.
Total RNA was isolated from two-week old warm-grown (22°C) seedlings. Northern blot analysis was performed using probes for CBF1, COR15a, and 18S ribosomal RNA. WT, RNA samples from independent lines overexpressing wild type CBF1; D/N and R/W, transgenic plants overexpressing a mutated CBF2 transgene harboring an Asp to Asn or Arg to Trp mutation, respectively.

Transcriptome profiling by Gilmour and colleagues (2004) revealed that CBF1, CBF2, and CBF3 display overlapping functions when overexpressed in *Arabidopsis*. In addition, the DNA-binding domains of CBF1 and CBF2 only differ in two residues, neither of which is involved in direct contact with DNA based on the AP2/EREBP structure described for AtERF1, so presumably the two DNA-binding domains can be considered equivalent. To determine the effect of Asp→Asn and Arg→Trp mutations on CBF2 binding affinity, MBP:$CBF2_{27\text{-}112}$ fusions were subcloned in *E. coli* and the purified protein extracts tested in a gel mobility shift assay. As observed for CBF1, CBF2 could specifically bind a *CRT/DRE*-containing probe (Figure 2.22). The Asp→Asn mutation, conservative in structure but with no charge, resulted in a significant loss of DNA binding, but retained its specific sequence recognition, as shown by the competition assay. Surprisingly, the Arg→Trp substitution, despite the significant change in side chain structure, did not alter CBF2 binding affinity.

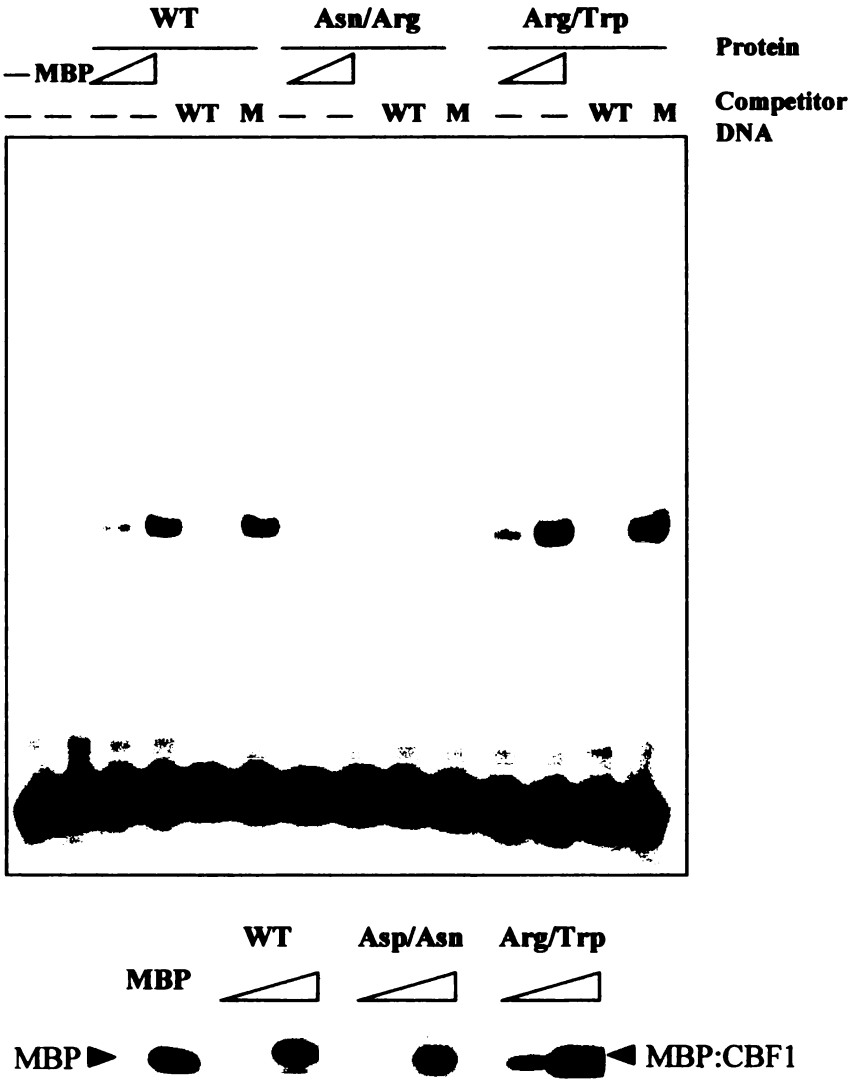


Figure 2.22. DNA binding activity of MBP:CBF2$_{27\text{-}112}$ proteins carrying point mutations in the DSAWR motif.
Upper panel. Gel shift assay showing different DNA binding activity. Increasing amounts (150 and 450 ng) of wild type and mutated CBF2 proteins were tested in a 15 ul binding reaction containing 0.5 ng of radiolabeled *CRT/DRE*-containing probe from *COR15a* promoter. WT and M indicate a wild type and mutated unlabeled DNA (100 ng each), respectively used for the competition experiments indicated.
Lower panel. Western blotting showing protein loading (15 and 45 ng of each protein were analyzed). α-MBP was used for protein detection.

Taken together, these observations suggest that the reduced ability to induce *COR* gene expression observed for the *dsawr* plants could be explained, in part, by inefficient binding of CBF1 to *CRT/DRE*-containing promoters. Based on the effect on protein accunulation observed when multiple alanines are introduced into the DSAWR motif, reduced ability to induce *COR* gene expression also appears to be due to reduced accumulation of the mutant protein.

DISCUSSION and FUTURE DIRECTIONS

The main goal of this study was to determine whether two conserved signature sequences in the CBF family of proteins are required for CBF1 transcriptional activity, and, if so, to elucidate their role. A mutational approach was taken to generate transgenic plants overexpressing wild type and mutated versions of the two motifs.

Induction of several members of the *CBF* regulon - *COR6.6*, *COR15A*, and *COR78* - was tested by northern blot analysis to evaluate the relevance of the signature sequences on CBF1 activity. In all cases tested, reduced *COR/CBF1* transcript ratios in the *pkkpagr* lines compared to the transgenic plants overexpressing wild type CBF1 indicated that the PKKPAGR motif is required for CBF1 activity. Interestingly, we noticed a difference on the effect of these mutations at different *COR* gene promoters. For instance, *COR15a* induction was dramatically affected by all mutations tested compared to *COR6.6* and *COR78* induction. One possible explanation is that PKKPAGR confers higher binding affinity to different *COR* promoters, *COR15a* in this case. Once this motif has been mutated, the DNA binding function is lost or reduced and the effect will be more pronounced at the most highly induced genes. Alternatively, induction of *COR15a*-like genes might represent a sub-group of the CBF regulon that requires a CBF1-interacting factor that acts through the PKKPAGR motif. When the PKKPAGR motif is mutated, the interaction is lost and CBF1 ability to recruit the transcriptional machinery will be greatly impaired. Yeast-two-hybrid screen by using the PKKPAGR motif as bait could help identify such factor. Overexpression of CBF1 in T-DNA knock-

out lines for that gene might show that a subset of *COR15a*-like genes is affected, similarly to what has been observed in the *pkkpagr* lines.

Several hypotheses were postulated to assign a functional role to the PKKPAGR motif. Western blot analysis of transgenic plants overexpressing *6xMyc:CBF1* constructs ruled out the possibility that the PKKPAGR motif lowers CBF1 protein steady-state levels. On the contrary, substitution of the KKF residues with alanines resulted in higher CBF1 protein accumulation (Figure 2.5, compare CBF1 transcript and protein levels in WT and M3). Transcription factors tend to be very unstable proteins and it has been extensively reported that their activity and their turn-over can be tightly correlated (Tansey, 2001). Their degradation is usually mediated by the proteolytic activity of the proteasome, which recognizes a poly-ubiquitin chain attached to a lysine residue of the target protein. Increased CBF1 protein levels and reduced transcriptional activity when KKF residues are converted to alanines are consistent with a role of the ubiquitin pathway in modulating CBF1 activity. The obvious question is whether the difference in protein levels in wild type and mutated CBF1 proteins can be attributed to this pathway. To address this question it will be necessary to determine whether CBF1 is ubiquitylated in vivo, and whether the protein can accumulate to higher levels in the presence of specific proteasome inhibitors. If this is the case, the following step will be the identification of the ubiquitin ligase that is responsible for the addition of the ubiquitin chain that signals CBF1 degradation to the proteasome. Ultimately, the goal is to extend this analysis to wild type Arabidopsis plants and study the effect of low temperature on the regulation of CBF1 protein stability.

The PKKPAGR region, rich in basic residues, represents a potential NLS, and resembles NLSs previously identified in other plant transcription factors, including the member of the AP2/EREBP family AINTEGUMENTA (Krizek and Sulli, 2006). We tested whether the PKKPAGR motif played a role in the transport of CBF1 across the nuclear envelope by nuclear localization studies of *Arabidopsis* plants overexpressing translational fusions of CBF1 to GFP:GUS. Laser confocal microscopy of *Arabidopsis* plants overexpressing CBF1ΔPKK:GFPGUS revealed that this chimera is present in the nucleus of root tip cells, indicating that the PKKPAGR motif is not required for nuclear targeting of CBF1; on the contrary, localization studies of 5' deletions of CBF1 fused to GFP:GUS revealed that in the absence of its N-terminal domain (residues 1-104) CBF1:GFP:GUS loses its nuclear localization, thus mapping the nuclear targeting signal to the N-terminal of the protein which includes the DNA-binding domain. These data correlate with the fact that $CBF1_{1\text{-}115}$–$VP16_{AD}$ can activate *COR* gene induction in stable *Arabidopsis* lines, and therefore must contain a nuclear targeting motif (Wang *et al.*, 2005). The involvement of the DNA-binding domain in nuclear targeting is not a new finding; it has been calculated that for 79% of DNA-binding proteins in which both NLS and DNA-binding domain have been characterized, the two functions map to the same protein domain and sometime the very same residues (LaCasse and Lefebvre, 1995). Some examples include the plant-specific SBP-domain proteins (Birkenbihl *et al.*, 2005), the yeast transcription factor GAL4 (Silver *et al.*, 1984; Carey *et al.*, 1989; Birkenbihl *et al.*, 2005), and the transcription factors from the high-mobility group family (Sudbeck and Scherer, 1997; Prieve *et al.*, 1998). Due to lack of a consensus signal for nuclear import, we could not identify any potential NLS in this region. However, the

AP2/EREBP domain contains several positively charged residues, mainly concentrated at its N-terminus, next to the PKKPAGR motif. It is possible that both the PKKPAGR region and this portion of the DNA-binding domain must be deleted to eliminate transport across the nuclear membrane.

Based on its proximity to the DNA-binding domain, we hypothesized that the PKKPAGR motif might be helping CBF1 to bind to the *CRT/DRE* element, and found that triple alanine mutations within this region abrogated DNA-binding. Secondary structure predictions suggested that several residues in the PKKPAGR motif are likely to be coiled, while a seven-residue motif – RKKFRET – showed strong helical propensity. Point mutations within the RKKFRET motif revealed that specific amino acids are crucial for CBF1 DNA binding activity. Point mutations at either Arg1 or Phe4 (Arg→Lys and Arg→Ser and Phe→Ala) caused the most dramatic loss in DNA-binding, whereas point mutations at Lys2 and Lys3 (Lys→Arg and Lys→Ala) had a minor effect and Arg5→Ser substitution showed no significant effect. The CRT/DRE element is specifically recognized both by CBF and DREB2A proteins (Liu *et al.*, 1998). A conserved thymine in the core sequence of this cis-acting element represents a specificity determinant for CBF1-dependent DNA binding activity (Sakuma *et al.*, 2006). Based on the NMR structure available for ERF1 and the high homology among AP2/EREBP proteins, it was proposed that residues within this domain are involved in the specific binding to the CCGAC core of the CRT/DRE element (Allen et al., 1998 - verify); however the conserved thymine is located downstream of these DNA bases and therefore is unlikely to be contacted by residues within the AP2/EREBP domain. We integrated these observations with experimental data, obtained from mutational analysis and

secondary structure predictions, in a computational model that illustrates how CBF1 interacts with the DNA major groove through its β-sheet and with the DNA minor groove via the RKKFRET motif. In this model, the helical region of the PKKPAGR motif reaches into the DNA minor groove with both Arg1 and Phe4 oriented in close proximity to the conserved thymine. Arginines are frequently associated to base specific recognition, when present at the protein-DNA interface; this interaction occurs mainly via hydrogen bonding with a guanine, in fewer cases with thymine (Luscombe *et al.*, 2001; Luscombe and Thornton, 2002). Phenylalanines are not frequently found; however, when present they can contribute specific DNA recognition by providing stacking interactions with adjacent bases in the minor groove, preferentially at a thymine-adenine pair (Luscombe and Thornton, 2002; Morávek *et al.*, 2002). Lys2 and Lys3, are placed in the vicinity of the DNA phosphates to provide favorable electrostatic interactions, whereas Arg5 does not participate to the protein-DNA complex. Overall, the experimental data can be nicely fit into this working model and explain how the PKKPAGR motif provides side chain interactions that are important or essential for CBF1 DNA binding activity. One important question that remains to be answered is whether Arg1 and/or Phe4 indeed represent specificity determinants and whether they are involved in direct recognition of the conserved thymine found in the CRT/DRE promoter element. NMR studies of CBF1 in complex with its cognate DNA will help shed some light on the nature of those interactions.

A role for N-flanking regions in DNA-binding affinity has been recognized in homeodomain proteins (Wolberger, 1996). In this family, binding specificity is driven by α-helices mainly via hydrophobic interactions in the major groove of the target DNA,

whereas the N-terminal tail makes contact with the DNA minor groove (Wolberger, 1996). DNA-binding by β-sheets has been extensively characterized at the structural and biochemical level in several DNA-binding proteins (Tateno *et al.*, 1998; Connolly *et al.*, 1998). However, most of the research has focused on defining the conserved structural features important for DNA binding within the β-sheet only. In contrast, very little is known about the role of their N-flanking sequences. Analysis of N-flanking sequences from other β-sheet-containing DNA-binding proteins revealed that these regions tend to be enriched in basic residues (Figure 2.14). Moreover, one or more residues in these regions are crucial for binding to the cognate DNA. These observations together with the results obtained with our mutational analysis suggest that the formation of protein-DNA complexes in the β-sheet family extends beyond the AP2/EREBP-like structure to include residues from the N-flanking region that provide important side chain interactions at the protein-DNA interface.

Along these lines, identification of the binding determinants in the PKKPAGR motif could provide some insights into the mode of DNA binding in the AP2/EREBP family. To date, most of the characterization has been limited to the binding activity of the ERF and CBF proteins. It has been proposed that the specificity determinants in these two subfamilies lie within the AP2/EREBP DNA-binding domain (Hao *et al.*, 1998; Sakuma *et al.*, 2002). However, high sequence similarity and the conservation of the residues that make direct contact with the DNA in these proteins suggests that the specificity switch may instead be found outside of this domain. Studies on several ERF proteins indicated that 10 amino acid residues upstream of the AP2/EREBP domain were essential for DNA binding (Hao *et al.*, 1998). Since proteins in the ERF group do not

share significant similarity in the region immediately upstream of the AP2/EREBP domain, the authors concluded that these residues do not contribute to specific binding, but instead their function is to stabilize the protein-DNA complex (Hao *et al.*, 1998). Interestingly, ERF proteins contain a cluster of positively charged residues flanking the AP2/EREBP domain at its N-terminus. It would be of interest to test the role of those basic amino acid residues in DNA binding by the ERF proteins. A large-scale comparison of the N-flanking sequences from members of this family combined with mutational analysis might reveal the presence of similar amino acids in other proteins within the AP2/EREBP family. In addition, sequence comparisons within different subgroups of the AP2/EREBP family could indicate whether there are any conserved residues that are specific for different subfamilies. If present, those residues would represent potential specificity determinants.

Similarly to what had been observed for the *pkkpagr* lines, the growth retardation in the *dsawr* plants was not as severe as in for wild type CBF1-overexpressing plants. Consistently, *COR* gene induction was affected by this mutation. When overexpressed in *Arabidopsis*, *dsawr* CBF1 proteins do not accumulate as well as wild type CBF1 proteins. Inability to accumulate this protein *in planta* provides an explanation for reduced *COR* gene expression in the *dsawr* lines compared to the wild type *CBF1*-overexpressing lines. It is not clear at the moment whether this result is solely due to protein instability. Alanine substitutions within the DSAWR motif impaired CBF1 DNA binding activity. Analysis of two point mutation in the DSAWR motif – Asp→Asn and Arg→Trp - suggest that at least one residue (Asp) might play an important role in DNA binding. In fact the Asp→Asn mutation, conservative in size though not in charge, may suggest that

no major change in protein folding occurred, and therefore a side chain interaction with the partner DNA might have been lost. On the contrary, mutation of Arg→Trp did not show any effect on DNA-binding affinity. Interestingly, overexpression of a *CBF2* transgene carrying that mutation in *Arabidopsis* has a dampening effect on *COR* gene expression as compared with overexpression of wild type CBF2 (Figure 2.23). Whether protein stability is affected in these plants is currently unknown. An intriguing possibility is that this residue is critical for protein-protein interaction that is required for transcriptional activation at the *COR* gene promoters.

MATERIAL AND METHODS

Mutagenesis of the PKKPAGR and DSAWR motifs.

Alanine scanning mutagenesis.

Site-directed mutagenesis (Li and Wilkinson, 1997) was performed to convert the original amino acids in the signature sequences of CBF1 to stretches of three alanines. Primers were designed to introduce the desired mutations to the PKKPAGR motif of CBF1 protein using the QuikChange mutagenesis kit (Stratagene), and the protocol was carried out according to the manufacturer instructions. A NotI restriction site was included in these primers for the screening of plasmids containing the mutated ORFs. The template DNA used was the full-length CBF1 ORF inserted into pBS/SK$^-$.The primers used for the PKKPAGR mutants are the following: Mutant1 (PKK), fwd (MT640): 5'gcc acg agt tgt gcg gcc gca ccg gcg ggc cgt3'; rev. (MT641) 5'acg gcc cgc cgg tgc ggc cgc aca act cgt ggc3'; Mutant2 – (PAGR), fwd. (MT642): 5'cga gtt gtc cga aga aag cgg ccg ccg cta aga agt ttc gtg aga c3'; rev. (MT643): 5'gtc tca cga aac ttc tta gcg gcg gcc gct ttc ttc gga caa ctt g3'; Mutant3 – (KKF), fwd. (MT644): 5'ccg gcg ggc cgt gcg gcc gct cgt gag act cgt3'; rev. (MT645): 5'acg agt ctc acg agc ggc cgc acg gcc cgc cgg3'; Mutant4 (RET), fwd. (MT646): 5'gcg ggc cgt aag aag ttt gcg gcc gct cgt cac cca att tac ag3'; rev (MT647): 5'ctg taa att ggg tga cga gcg gcc gca aac ttc tta cgg ccg cg3'; Mutant5 (RHP), fwd. (MT686): 5'aag aag ttt cgt gag act gcg gcc gca att tac aga gga gtt cgt3'; rev (MT687): 5'acg aac tcc tct gta aat tgc ggc cgc agt ctc acg aaa ctt ctt3'. Primers used for DSAWR mutant: fwd. (MT817) 5'gtc tca act tcg ctg ccg cgg ccg cgg cgc tac gaa tcc cgg ag3'; rev. (MT818): 5'ctc cgg gat tcg tag cgc cgc ggc cgc ggc agc gaa gtt gag ac3'.

The ΔPKK mutant, that lacks the entire PKKPAGR region, was made using a modified protocol based on the QuikChange method (Wang and Malcolm, 1999). To overcome the tendency of the perfectly complementary mutagenic primers to anneal to each other rather than to the target sequence, a two-stage PCR was performed, running two separate single-primer reactions before the final PCR-amplification. Primers: fwd. (MT665) 5'taa atg tct ggt caa gca gtt tct ttg agc cg3'; rev. (MT666): 5'tgt tga gca ccg gtt gca gcc tgt tat tag ag3'. All constructs were verified by DNA sequencing.

Point mutations in the RKKFRET region of CBF1.

The point mutations in the RKKFRET motif were designed by using environment-specific substitution tables (Overington *et al.*, 1992), which allowed to choose substitutions compatible with the helical structure. To introduce point mutations, the QuikChange mutagenesis protocol was followed as described in the Strategene manual. The following primers were used: Arg1→Lys, fwd: 5'ccg aag aaa ccg gcg ggc aag aag aag ttt cgt gag act cg3'; rev: 5'cga gtc tca cga aac ttc ttc ttg ccc gcc ggt tct tcg3'; Arg1→Ser, fwd: 5'cga aga acc ggc ggg ctc gaa gaa gtt tcg tga gac tcg3'; rev: 5'cga gtc tca cga aac ttc ttc gag ccc gcc ggt ttc ttc g3'; Lys2→Arg, fwd: 5'ccg gcg ggc aga aga aag ttt cgt gag act cg3'; rev: 5'cga gtc tca cga aac ttt ctt ctg ccc gcc gg3'; Lys2→Ala, fwd: 5'ccg gcg ggc aga gcg aag ttt cgt gag act cgt cac3'; rev: 5'gtg acg agt ctc acg aaa ctt cgc tct gcc cgc cgg3'; Lys3→Arg, fwd:5'ccg gcg ggc aga aag aga ttt cgt gag act cgt cac c3'; rev:5'ggt gac gag tct cac gaa atc tct ttc tgc ccg cc gg3'; Lys3→Ala, fwd: 5'ccg gcg ggc aga aag gcg ttt cgt gag act cgt cac c3'; rev: 5'ggt gac gag tct cac gaa acg cct ttc tgc ccg ccg g3'; Phe4→Ala, fwd: 5'ggg ccg taa gaa ggc tcg aga gac tcg tca ccc3'; rev: 5'ggg tga cga gtc tct cga gcc ttc tta cgg ccc3'; Phe4→Pro, fwd: 5'gcg ggc cgt aag aag cct cga gag act cgt cac cc3'; rev: 5'ggg tga cga gtc

tct cga ggc ttc tta cgg ccc gc3'; Phe4→Tyr, fwd: 5'ccg gcg ggc aga aag aag tac cgt gag act cgt c3'; rev: 5'gac gag tct cac ggt act tct ttc tgc ccg ccg g3'; Arg5→Ser, fwd: 5'ccg gcg ggc aga aag aag ttt agt gag act cgt cac c3'; rev: 5'ggt gac gag tct cac taa act tct ttc tgc ccg ccg g3'. For the screening of the clones harboring the desired mutation, the primers were designed to disrupt a pre-existing Sau96I restriction site in all mutant ORFs, except for Phe4→Ala and Phe4→Pro mutations, where a new XhoI restriction site was inserted.

Preparation of constructions for plant transformation.

Plant expression vector harboring wild type and mutated CBF1.

BglII restriction ends were added by PCR in order to subclone the ORF of wild type or mutated *CBF1* into pGA643 (An *et al.*, 1988) which harbors the constitutive 35S CaMV promoters. Primers BglII fwd: 5'gaa gat cta tga act cat ttt cag ctt ttt ctg3'; BglII rev: 5'gaa gat ctc tcg ttt cta caa caa taa aat aaa3'. All constructs were verified by DNA sequencing.

Plant expression vector harboring 6xMyc:CBF1 ORFs.

Fusional translation of wild type and mutated CBF1 to 6xMyc, were generated by subcloning a SmaI/SacI CBF1 ORF into the binary vector pKVB24 (Vlachonasios *et al.*, 2003) containing a 6xMyc tag, under the control of CaMV 35*S* promoter. The translational fusion results in 6xMyc at the N-terminus of CBF1. Primers used to add SmaI/SacI ends by PCR amplification were: MT708, fwd: 5'agc ccg ggg atg aac tca tt3' MT709, rev.: 5'cag agc tct tac taa ctc ca3'.

Plant expression vector harboring CBF1:GFP:GUS.

Wild type CBF1 and CBF1ΔPKK ORFs were pEZT-CL(GUS). The vector was engineered by subcloning a BamHI GUS insert into the pEZT-CL plant expression vector (Schena *et al.*, 1991), resulting in an in frame fusion to GFP under the control of the CaMV 35S promoter. The eGFP gene in pEZT is based on mGFP4 (Haseloff *et al.*, 1997) and contains additional mutations (S65T, Y66H) to increase intrinsic GFP fluorescence (Cormack *et al.*, 1996). BamHI restriction ends were added to GUS by PCR amplification using the following primers: fwd. (MT794): 5'cag gat ccg cat cga taa gct tga att cac c3'; rev. (MT795): 5'aga gga tcc cca att ccc gag gct gta3'. Full-length and 5' deletions of CBF1 ORF were subcloned into the XhoI restriction site of pEZT-CL(GUS). In frame fusions of CBF1 to GFP:GUS were generated by inserting an XhoI fragment into the binary vector. A PCR approach was used to add the proper restriction ends, as follows: full-length CBF1 XhoI 5' end (MT704): 5'gcc tcg aga tga act cat ttt cag3'; XhoI 3' end (MT825): 5'ccc tcg agg cgt aac tcc aaa gcg3'For the 5' ORF deletions the following forward primers were used: construct 802 (MT802): 5'acc tcg aga tgc cga aga aac cgg cgg gcc g3'; construct 801 (MT801): 5'acc tcg aga tga ttt aca gag gag ttc g3'; construct 800 (MT800): 5'acc tcg aga tgg act cgg ctt ggc ggc tac g3'; construct 799 (MT799): 5'acc tcg aga tgc tac gaa tcc cgg agt caa c3'. The XhoI insert for NIa was PCR-amplified from the yeast plasmid pAVA367, containing the in frame fusion Nia-GFP (Schena *et al.*, 1991), with the following primers: fwd. (MT815): 5'acc tcg aga tgg gga aga aga atc aga agc aca agc taa aga tga g3'; rev. (MT826): 5'ccc tcg agt gac ctg tca atg gat cca c3'.

Plant transformation.

All the transgenic lines used in this study were generated by the floral dip method described by Clough and Bent (1998). pEZT-CL(GUS) or pKVB24 constructs were

transformed into the *A. tumefaciens* strain LBA4404, whereas pGA643 constructs were transformed into strain GV3101. Generation and selection of transgenic lines was performed as follows. *Arabidopsis* plants (T0) were grown in soil to bolting and then transformed by *A. tumefaciens*-mediated transformation using the floral dip method (Clough and Bent, 1998). Seeds from T0 plants were screened for antibiotic or herbicide resistance (600uM Basta; 50ug/ml Kanamycin). Resistant T1 seedlings were grown to maturity to produce T2 seeds. T2 seeds were plated onto selective media containing antibiotic or herbicide and screened for a 3:1 survival ratio, indicating that there was an insertion in a single locus in the T1 genome. Homozygous lines were selected by growing plants on selective media and screened for 100% survival.

Growth conditions and Northern blot analysis of transgenic *Arabidopsis* plants.

Seeds were sterilized and stratified for 3 days at 4°C, and then plated on Gamborg's plates supplemented with B5 nutrients and 23.21gr/liter of sucrose (Caisson Laboratories, Inc., Rexburg, ID, USA) and 0.8% phytagar (Life Technologies Inc., Gaithersburg, MD, USA). Seedlings were grown under continuous cool-white fluorescent light ($100\mu mol \cdot m^{-2} \cdot s^{-1}$) at 22°C for two weeks. Tissue samples were harvested in liquid Nitrogen and total RNA extracted using either the RNeasy Plant Miniprep kit (QIAGEN) or the Trizol reagent (Life Technologies, Gaithersburg, MD). Five to ten micrograms of total RNA were fractionated in 1% formaldehyde gels and transferred onto nitrocellulose membranes as described (Sambrook *et al.*, 1989). Membranes were hybridized in Church buffer (1% BSA, 1 mM EDTA, 0.5 M NaPO4 pH 7.2, 7% SDS) (Church and Walter, 1984) at 65°C, overnight. Blots were probed with a ^{32}P-labelled fragments prepared using

the Random Primers DNA Labeling System (Invitrogen), according to the manufacturer instructions. Probes for all the transcripts were obtained by using the full-length ORF as a probe. After hybridization, membranes were washed three times in .1XSSC, 0.1%SDS at 55°C, for 15 minutes. mRNA levels in different samples were normalized by comparing them to the levels of 18S rRNA determined from the same blots. Following the washes, the membranes were exposed to a phosphorimager screen; the screen was later scanned and then quantified using a QuantityOne software from BioRad (Hercules, CA, USA).

Analysis of variance (ANOVA) of *COR/CBF1* transcript ratios.

ANOVA tests were carried out using to determine the effect of mutations in the PKKPAGR motif on COR78/CBF1 mRNA ratios. mRNA values for CBF1 and different *COR* genes were obtained by exposing radioactive membranes to a phosphorimager screen. The screen was later scanned and the photointensity of different bands was measured using a QuantityOne software from BioRad (Hercules, CA, USA). Variables were log-transformed to meet the normality assumptions of statistical analyses. Each analysis was conducted by averaging ratios from two technical replicates. Each group analyzed was represented by five or six independent transgenic lines. Statistical difference in COR/CBF1 ratios of transgenic plants overexpressing a wild type CBF1 transgene or CBF1 transgenes mutated in the signature sequences was analyzed by analysis of variance (ANOVA) using SAS Proc Mixed procedures (SAS Institute, Cary, NC), version 9.1. To test for significant difference between COR/CBF1 ratios between transgenic plants overexpressing wild type CBF1 or the mutated versions described

earlier we estimated least-square means for each genotype and compared them to the least-square means of the control plants overexpressing wild type CBF1. These estimates were used to calculate the t-values and the statistical significance at $P<0.0001$ for all the transgenic lines analyzed.

Subcloning of *CBF1* mutant ORFs into bacterial expression vectors for expression and purification of 6xHis:CBF1 and MBP:CBF1 proteins.

Expression and purification of 6xHis:CBF1 proteins.

All *CBF1* ORFs described earlier were introduced into the pET28a$^+$ expression vector (Novagen/EMD, San Diego, CA) under the control of the bacteriophage IPTG-inducible T7 promoter. The constructs were generated by NcoI/XhoI cleavage of the pSJG6 construct (ref.), in which CBF1 is fused to 6His-T7 tag. The NcoI/XhoI fragments were ligated onto pET28a$^+$. Sequences were confirmed for all constructs and plasmids were transformed into BL21-CodonPlus(DE3)-RIL (Stratagene, La Jolla, CA) *E. coli* strain for optimal expression of the recombinant proteins (Kleber-Janke and Becker, 2000). Protein expression and purification was started by inoculation of individual colonies containing plasmid DNA in 3mL of Luria-Bertani broth containing 50ug/ml Kanamycin and 40ug/ml Chloramphenicol (Sigma-Aldrich, St.Louis, MO) and grown at 37°C, overnight (250 rpm). 50ul of that culture were inoculated into 500ml of the same broth to an $O.D_{600}$ of approximately 0.9 before addition of 1mM IPTG (Research Organics, Cleveland, OH) to induce protein expression for 4 hours, at 37°C. Cells were harvested by centrifugation, and resuspended in 10ml of cell lysis buffer (50mM $NaH_2PO_4 \cdot H_2O$, 300mM NaCl, 10mM Imidazole, pH8.0). Cells were lysed by sonication (BRANSON Sonifier150, Pittsburgh, PA) in the presence of

protease inhibitors (Complete EDTA-free protease inhibitor cocktail tablets, Roche, Mannheim, Germany). The soluble protein fraction was separated by centrifugation, at 4°C. The supernatant fraction containing 6xHis-T7-CBF1 proteins was loaded over a nickel column equilibrated with wash buffer (50mM $NaH_2PO_4 \cdot H_2O$, 300mM NaCl, 20mM Imidazole, pH8.0). After three washes in wash buffer, 6xHis-T7-CBF1 proteins were eluted by washing with elution buffer (50mM $NaH_2PO_4 \cdot H_2O$, 300mM NaCl, 250mM Imidazole, pH8.0). Protein concentration was determined using the Bradford dye binding assay (Bio-Rad, Hercules, CA), with BSA used as standard.

Expression and purification of MBP:CBF1$_{27-112}$ proteins

A 258bp XmnI/XbaI *CBF1* fragment (aa27-112) was cloned into pMAL-c2x vector (New England BioLabs, Beverly, MA), downstream of the *malE* gene, which encodes Maltose Binding Protein (MBP) under the control of the strong P_{tac} promoter. this resulted in the translational fusion of CBF1 to the C-terminus of MBP. The primers used for wild type CBF1 and all the PKKPAGR mutants were: MT908(fwd): 5'ATA TTT TCT AGA TTC GTA GCC3', and MT909(rev.): 5'CCA TGG AAG GAT TTC GGC CAC GAG TTG T3'. For the *dsawr* CBF1 mutant, MT910 was used as the reverse primer: 5'ATA TTT TCT AGA TTC GTA GCG C3'. Individual transformed cells were grown at 37°C in 3ml Rich Medium (10g tryptone, 5g yeast extract 5g NaCl, 2g glucose, per liter) supplemented with 100ug/ml Ampicillin to an O.D. of 0.5. Protein expression was induced by addition of 1mM IPTG to the bacterial suspension, and let grow for 3 additional hours. Cells were harvested by centrifugation (6,000 rpm, 10min., 4°C), and lysed by sonication, as described above. The soluble protein fraction was separated by centrifugation at 13,000rpm, 30 min., at 4°C. The supernatant containing the fusion

proteins was loaded over an amylase column for protein purification by affinity. The amylase column was equilibrated in column buffer (20mM Tris-HCl, 200mM NaCl, 1mM EDTA). MBP-CBF1 proteins were eluted in column buffer containing 10mM maltose. Protein concentration was determined using the Bradford dye binding assay (Bio-Rad, Hercules, CA).

Protein isolation and immunoblot analyses.

Immunodetection of CBF1 protein from E.coli extracts.

Western Blotting and protein detection were carried out by first separating the proteins on a 10% bis-acrylamide tricine gel. Proteins were transferred in Towbin buffer (25mM2 Tris, 192mM Glycine, 15% MeOH, 0.02% SDS) to a PVDF membrane (Immobilon-P, Millipore corporations, Bedford, MA) using by electrophoretic transfer (Genie®, Idea Scientific, Minneapolis, MN), for 45min. at 1Amp, 30 volts.
The membrane was blocked in PBS-Tween (per liter: 80g NaCl, 2gr KCl, 14,4gr Na_2HPO_4, 2.4gr KH_2PO_4, pH7.2, 0.1% (v/v) Tween 20) and 5% non-fat dried milk for 1h at RT with gentle rocking. Incubation with the primary antibody was conducted for 1h at room temperature in PBS-T and 5% non-fat dried milk for 1h at RT. Antibody dilutions were: 1/7,500 for α-CBF1 (F1, IgG purified CBF1 antiserum raised against the full-length CBF1 protein fused to a 6xHis tag); 1/10,000 for α-MBP (New England Biolabs, Beverly, MA). The presence of the protein on the membrane was assayed by chemiluminescence detection (Amersham Biosciences, Piscataway, NJ).

Immunodetection of CBF1 protein from plant extracts.

Total protein extracts were prepared from two-week old *Arabidopsis* seedlings by grinding frozen tissue in protein extraction buffer (20 mM Tris-HCl, pH 8.0, 50mM NaCl, 5 mM EDTA, 0.05 % SDS) supplemented with protease inhibitor cocktail from Roche. Protein concentration was determined using the Bradford reagent (Bio-Rad, Hercules, CA, USA) with BSA as the standard. Equal amounts of proteins were subjected to SDS-polyacrylamide gel electrophoresis (PAGE) (4-20% gradient gel). After electrophoresis, the proteins were transferred to polyvinylidene difluoride (PVDF) membranes by electrotransfer. Membranes were blocked in 5% nonfat milk powder in Tris-buffered saline (TBS) - 0.1% Tween-20 for 1 h at room temperature, and then incubated overnight at 4°C with specific antibodies. Antibody dilutions were: 1/5,000 for the α-Myc monoclonal antibody (Roche-Mannheim, USA), and 1/10,000 for the α-GFP monoclonal antibody (Pierce Biotechnology, Rockford, IL). The next day, membranes were washed three times and incubated in horseradish peroxidase-coupled anti-mouse IgG antibody [1/10,000 dilution of the rabbit anti-mouse IgG (Pierce Biotechnology, Inc., Rockford, IL)] for 1 h at room temperature. Immunoreactive bands were visualized by an enhanced chemiluminesence assay following the manufacturer's inst2ructions.

Electrophoretic-Mobility Shift Assays (EMSA).

24-bp fragments from *COR15a*, *COR78* and *COR6.6* promoters containing a *CRT/DRE* element were prepared by synthesizing both strands. The oligomers were suspended in 1× STE buffer (10mM Tris, pH8.0; 10mM NaCl; 1mM EDTA), and the complementary strands were annealed. The resulting double strand oligonucleotides were labeled with [γ-^{32}P]ATP (NEN Life Science Products) by T4 polynucleotide kinase (New

England Biolabs, Beverly, MA) and purified through a Sephadex G-50 column. Mutant and wild type sequences were prepared following the same procedure. The DNA oligomers used for the binding assays are the following: *COR15a*: 5'ATT TCA TGG CCGAC CTG CTT TTT3'; *COR78*: 5'AAT ATA CTA CCGAC ATG AGT TCT 3'; *COR6.6*: 5'AAA AAG CTA CCGAC ATA AGC CAA3'. A mutant version of *COR15a* in which the entire C-repeat, CCGAC, had been altered was used for all the competition assays: 5' ATTTCATGGtatgtCTGCTTTTT3'. The binding of 6xHis:CBF1 or MBP:$CBF1_{27\text{-}112}$ to the *CRT/DRE*-containing DNA probes was tested in a total volume of 12.0 ul as follows: 0.5 ng of a ^{32}P-labeled DNA probe encoding the *CRT/DRE* binding site was mixed with a gradient concentration of each recombinant protein (100 - 300 ng of 6xHis:CBF1 and 0.3 – 15 ug for MBP:$CBF1_{27\text{-}112}$) and incubated in binding buffer (20 mM Tris-HCl, pH 8.0; 100 ug/ml BSA; 30mM KCl; 5mM $MgCl_2$; 4% Glycerol; 1mM DTT) in the presence/absence of 100 ng unlabeled competitor DNA. After incubation at room temperature for 20 min., samples were loaded into nondenaturing polyacrylamide gels (4% wt/vol), and fractionated by electrophoresis at 150V, for 3 hours at 4°C. The gels were dried at 80°C, 30 min., and exposed to a phosphorimager screen. The screen was later scanned and the photointensity of complexed DNA was measured using a QuantityOne software from BioRad (Hercules, CA, USA).

Fluorescence imaging of *Arabidopsis* root tips overexpressing *CBF1:GFP:GUS* transgenes.

For cellular localization of CBF1:GFP:GUS proteins, *Arabidopsis* seeds were sterilized, stratified for 3 days at 4°C, and plated on Gamborg agar plates supplemented

with sucrose as described above. The Petri dishes were oriented vertically and the seedlings grown for 6-7 days before being imaged. Laser confocal images were collected using an upright LSM Zeiss 510 META microscope equipped with a 40X oil immersion objective. To visualize the nuclei, *Arabidopsis* seedlings were incubated with a solution of 50 μg/ml Propidium Iodide (PI) and 5 μg/ml RNase, in the dark for 30 min – 1 hr. Samples were rinsed with distilled water a couple of times and mounted in tap water. GFP fluorescence images were obtained using Argon ion laser excitation of 488 nm with a 505/530 nm bandpass filter. PI fluorescence images were collected using an excitation line of 543 nm with a 560 nm longpass filter. Postacquisition image processing was done with the LSM 5 Image Browser (Zeiss) and Adobe Photoshop 5.0 software (Mountain View, CA).

LITERATURE CITED

Abel, S., Oeller, P., and Theologis, A. (1994). Early auxin-induced genes encode short-lived nuclear proteins. PNAS **91,** 326-330.

Adamczak, Porollo A, and J, M. (2005). Combining prediction of secondary structure and solvent accessibility in proteins. Proteins: Structure, Function, and Bioinformatics **59,** 467-475.

Allen, M., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998). A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. EMBO J **17,** 5484-5496.

An, G., Ebert, P., Mitra, A., and Ha, S. (1988). Binary vectors. (Dordrecht: Kluwer Academic Publishers.).

Birkenbihl, R., Jach, G., Saedler, H., and Huijser, P. (2005). Functional dissection of the plant-specific SBP-domain: overlap of the DNA binding and nuclear localization domain. JMB **352,** 585-596.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by [lambda] integrase. Nature **435,** 1059.

Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F., and Ptashne, M. (1989). An amino-terminal fragment of GAL4 binds DNA as a dimer. Journal of Molecular Biology **209,** 423.

Carrington, J.C., Freed, D.D., and Leinicke, A.J. (1991). Bipartite signal sequence mediates nuclear translocation of the plant potyviral Nla protein. Plant Cell **3,** 953-962.

Chakrabartty, A., Kortemme, T., and Baldwin, R.L. (1994). Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci **3,** 843-852.

Chen, H.A., Simpson, P., Huyton, T., Roper, D., and Matthews, S. (2005). Solution structure and interactions of the *Escherichia coli* Cell Division Activator protein CedA. Biochemistry **44,** 6738-6744.

Chou, P.Y., and Fasman, G.D. (1978). Empirical predictions of protein conformation. Annual Review of Biochemistry **47,** 251-276.

Christophe-Hobertus, C., Duquesne, V., Pichon, B., Roger, P.P., and Christophe, D. (1999). Critical residues of the homeodomain involved in contacting DNA bases also specify the nuclear accumulation of thyroid transcription factor-1. European Journal of Biochemistry **265,** 491-497.

Church, G., and Walter, G. (1984). Genomic sequencing. PNAS **81,** 1991-1995.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana.* The Plant Journal **16,** 735-743.

Collawn, J.F., Stangel, M., Kuhn, L.A., Esekogwu, V., Jing, S.Q., Trowbridge, I.S., and Tainer, J.A. (1990). Transferrin receptor internalization sequence YXRF implicates a tight turn as the structural recognition motif for endocytosis. Cell **63,** 1061-1072.

Connolly, K.M., Wojciak, J.M., and Clubb, R.T. (1998). Site-specific DNA binding using a variation of the double stranded RNA binding motif. Nat Struct Biol **5,** 546-550.

Connolly, K.M., Ilangovan, U., Wojciak, J.M., Iwahara, M., and Clubb, R.T. (2000). Major groove recognition by three-stranded [beta]-sheets: affinity determinants and conserved structural features. Journal of Molecular Biology **300,** 841-856.

Cormack, B.P., Valdivia, R.H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). Gene Flourescent Proteins and Applications **173,** 33-38.

Craig, L., Sanschagrin, P.C., Rozek, A., Lackie, S., Kuhn, L.A., and Scott, J.K. (1998). The role of structure in antibody cross-reactivity between peptides and folded proteins. Journal of Molecular Biology **281,** 183-201.

Crane-Robinson, C., Dragan, A.I., and Privalov, P.L. (2006). The extended arms of DNA-binding domains: a tale of tails. Trends in Biochemical Sciences **31,** 547.

DeWitt, N.D., Hong, B., Sussman, M.R., and Harper, J.F. (1996). Targeting of two *Arabidopsis* H+-ATPase isoforms to the plasma membrane. Plant Physiol. **112,** 833-844.

Dragan, A., Read, C., Makeyeva, E., Milgotina, E., Churchill, M., Crane-Robinson, C., and Privalov, P. (2004). DNA binding and bending by HMG boxes: energetic determinants of specificity. Journal of Molecular Biology **343,** 371.

Eyrich, V.A., Marti-Renom, M.A., Przybylski, D., Madhusudhan, M.S., Fiser, A., Pazos, F., Valencia, A., Sali, A., and Rost, B. (2001). EVA: continuous automatic evaluation of protein structure prediction servers. Bioinformatics **17,** 1242-1243.

Gilmour, S.J., Fowler, S.G., and Thomashow, M.F. (2004). *Arabidopsis* transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. Plant Mol Biol **54,** 768-781.

Gilmour, S.J., Sebolt, A.M., Salazar, M.P., Everard, J.D., and Thomashow, M.F. (2000). Overexpression of the *Arabidopsis* CBF3 transcriptional activator mimics

multiple biochemical changes associated with cold acclimation. Plant Physiol. **124,** 1854-1865.

Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M., and Thomashow, M.F. (1998). Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. The Plant Journal **16,** 433-442.

Gilmour SJ, Fowler SG, and MF, T. (2004). Arabidopsis transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. Plant Mol Biol **54,** 768-781.

Grebenok, R., Pierson, E., Lambert, G., Gong, F., Afonso, C., Haldeman-Cahill, R., Carrington, J., and Galbraith, D. (1997). Green-fluorescent protein fusions for efficient characterization of nuclear targeting. Plant J **11,** 573-586.

Guana, C., Lib, P., Riggsa, P., and Inouyeb, H. (1987). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. Gene **67,** 21-30.

Haake, V., Cook, D., Riechmann, J.L., Pineda, O., Thomashow, M.F., and Zhang, J.Z. (2002). Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiol. **130,** 639-648.

Hao, D., Ohme-Takagi, M., and Sarai, A. (1998). Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J Biol Chem **273,** 26857-26861.

Haseloff, J., Siemering, K.R., Prasher, D.C., and Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. PNAS **94,** 2122-2127.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J Mol. Biol. **233,** 123-138.

Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O., and Thomashow, M.F. (1998). *Arabidopsis* CBF1 overexpression induces *COR* genes and enhances freezing tolerance. Science **280,** 104-106.

Jaglo, K.R., Kleff, S., Amundsen, K.L., Zhang, X., Haake, V., Zhang, J.Z., Deits, T., and Thomashow, M.F. (2001). Components of the *Arabidopsis* C-Repeat/Dehydration-Responsive Element Binding Factor cold-response pathway are conserved in *Brassica napus* and other plant species. Plant Physiol. **127,** 910-917.

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

Kaffman, A., and O'Shea, E. (1999). Regulation of nuclear localization: a key to a door. Annual Review of Cell and Developmental Biology **15,** 291-339.

Karplus, K., Draper, J., Casper, J., Mandel-Gutfreund, Y., Diekhans, M., and Hughey, R. (2003). Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. Proteins: Structure, Function, and Genetics **53,** 491-496.

Kim, J., Mao, J., and Gunner, M. (2005). Are acidic and basic groups in buried proteins predicted to be ionized? Journal of Molecular Biology **348,** 1283.

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. Nature **365,** 512.

Kleber-Janke, T., and Becker, W.-M. (2000). Use of modified BL21(DE3) *Escherichia coli* cells for high-level expression of recombinant peanut allergens affected by poor codon usage. Protein Expression and Purification **19,** 419.

Klucher, K.M., Chow, H., Reiser, L., and Fischer, R.L. (1996). The AINTEGUMENTA gene of *Arabidopsis* required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. Plant Cell **8,** 137-153.

Koroleva, O., Tomlinson, M., Leader, D., Shaw, P., and Doonan, J. (2005). High-throughput protein localization in *Arabidopsis* using *Agrobacterium*-mediated transient expression of GFP-ORF fusions. Plant J **41,** 162-174.

Krizek, B., and Sulli, C. (2006). Mapping sequences required for nuclear localization and the transcriptional activation function of the *Arabidopsis* protein AINTEGUMENTA. Planta **224,** 612-621.

LaCasse, E., and Lefebvre, Y. (1995). Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. Nucleic Acids Res. **23,** 1647-1656.

Li, S., and Wilkinson, M. (1997). Site-directed mutagenesis: a two-step method using PCR and DpnI. Biotechniques **23,** 588-590.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell **10,** 1391-1406.

Lnenicek-Allen, M., Read, C., and Crane-Robinson, C. (1996). The DNA bend angle and binding affinity of an HMG box increased by the presence of short terminal arms. Nucl. Acids Res. **24,** 1047-1051.

Luscombe, N., and Thornton, J. (2002). Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. Journal of Molecular Biology **320**, 991.

Luscombe, N., Laskowski, R., and Thornton, J. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucl. Acids Res. **29**, 2860-2874.

Lyck, R., Harmening, U., Hohfeld, I., Treuter E, Scharf KD, and L, N. (1997). Intracellular distribution and identification of the nuclear localization signals of two plant heat-stress transcription factors. Planta **202**, 117-125.

Magome, H., Yamaguchi, S., Hanada, A., Kamiya, Y., and Oda, K. (2004). *dwarf* and *delayed-flowering 1*, a novel *Arabidopsis* mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. The Plant Journal **37**, 720-729.

Maruyama, K., Sakuma, Y., Kasuga, M., Ito, Y., Seki, M., Goda, H., Shimada, Y., Yoshida, S., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. The Plant Journal **38**, 982-993.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics **16**, 404-405.

Morávek, Z., Neidle, S., and Schneider, B. (2002). Protein and drug interactions in the minor groove of DNA. **30**, 1182-1191.

Nadassy, K., Wodak, S., and Janin, J. (1999). Structural features of protein-nucleic acid recognition sites. Biochemistry **38**, 1999-2017.

Nagadoi, A., Nakazawa, K., Uda, H., Okuno, K., Maekawa, T., Ishii, S., and Nishimura, Y. (1999). Solution structure of the transactivation domain of ATF-2 comprising a zinc finger-like subdomain and a flexible subdomain. J.Mol.Biol. **287**, 593-607.

O'Neil, K., and DeGrado, W. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. **250**, 646-651.

Ohme-Takagi, M., and Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell **7**, 173-182.

Overington, J., Donnelly, D., Johnson, M.S., Sali, A., and Blundell, T.L. (1992). Environment-specific amino acid substitution tables: Tertiary templates and prediction of protein folds. Protein Sci **1**, 216-226.

Poon, I.K.H., and Jans, D.A. (2005). Regulation of nuclear transport: central role in development and transformation? Traffic **6,** 173-186.

Prieve, M.G., Guttridge, K.L., Munguia, J., and Waterman, M.L. (1998). Differential importin-alpha recognition and nuclear transport by nuclear localization signals within the High-Mobility-Group DNA binding domains of Lymphoid Enhancer Factor 1 and T-Cell Factor 1. Mol. Cell. Biol. **18,** 4819-4832.

Rashotte, A.M., Mason, M.G., Hutchison, C.E., Ferreira, F.J., Schaller, G.E., and Kieber, J.J. (2006). A subset of *Arabidopsis* AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. PNAS **103,** 11081-11085.

Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., and Stuart, D. (1995). High resolution structures of HIV-1 RT from four RT-inhibitor complexes. Nat Struct Biol. **2,** 293-302.

Riechmann, J.L., and Meyerowitz, E.M. (1998). The AP2/EREBP family of plant transcription factors. Biol Chem **379,** 633-646.

Rivas, S., Romeis, T., and Jones, J.D.G. (2002). The Cf-9 disease resistance protein is present in an ~420-Kilodalton heteromultimeric membrane-associated complex at one molecule per complex. Plant Cell **14,** 689-702.

Rost, B., and Liu, J. (2003). The PredictProtein server. Nucl. Acids Res. **31,** 3300-3304.

Ryter, J.M., and Schultz, S.C. (1998). Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. Embo J **17,** 7505-7513.

Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun **290,** 998-1009.

Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2006). Functional analysis of an Arabidopsis transcription factor, DREB2A, involved in drought-responsive gene expression. Plant Cell **18,** 1292-1309.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.).

Schena, M., Picard, D., and Yamamoto, K. (1991). Vectors for constitutive and inducible gene expression in yeast. Methods Enzymology **194.**

Silver, P.A., Keegan, L.P., and Ptashne, M. (1984). Amino terminus of the yeast GAL4 gene product is sufficient for nuclear localization. PNAS **81,** 5951-5955.

Stockinger, E.J., Gilmour, S.J., and Thomashow, M.F. (1997). *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the *C-repeat/DRE*, a *cis*-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. PNAS **94,** 1035-1040.

Sudbeck, P., and Scherer, G. (1997). Two independent nuclear localization signals are present in the DNA-binding High-mobility Group domains of SRY and SOX9. J. Biol. Chem. **272,** 27848-27852.

Tang, Y., Ramakrishnan, C., Thomas, J., and DeFranco, D.B. (1997). A role for HDJ-2/HSDJ in correcting subnuclear trafficking, transactivation, and transrepression defects of a glucocorticoid receptor zinc finger mutant. Mol Biol Cell **8,** 795-809.

Tansey, W.P. (2001). Transcriptional activation: risky business. Genes Dev. **15,** 1045-1050.

Vlachonasios, K.E., Thomashow, M.F., and Triezenberg, S.J. (2003). Disruption mutations of ADA2b and GCN5 transcriptional adaptor genes dramatically affect *Arabidopsis* growth, development, and gene expression. Plant Cell **15,** 626-638.

Vogel, J.T., Zarka, D.G., Van Buskirk, H.A., Fowler, S.G., and Thomashow, M.F. (2005). Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. The Plant Journal **41,** 195-211.

von Arnim, A.G., and Deng, X.-W. (1994). Light inactivation of *Arabidopsis* photomorphogenic repressor COP1 involves a cell-specific regulation of its nucleocytoplasmic partitioning. Cell **79,** 1035-1045.

Walther, R.F., Atlas, E., Carrigan, A., Rouleau, Y., Edgecombe, A., Visentin, L., Lamprecht, C., Addicks, G.C., Hache, R.J.G., and Lefebvre, Y.A. (2005). A Serine/Threonine-rich motif is one of three nuclear localization signals that determine unidirectional transport of the mineralocorticoid receptor to the nucleus. J. Biol. Chem. **280,** 17549-17561.

Wang, W., and Malcolm, B. (1999). Two-Stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. BioTechniques **26,** 680-682.

Wang, Z., Triezenberg, S., Thomashow, M., and EJ., S. (2005). Multiple hydrophobic motifs in *Arabidopsis* CBF1 COOH-terminus provide functional redundancy in trans-activation. Plant Mol Biol **58,** 543-559.

Wojciak, J., Sarkar, D., Landy, A., and Clubb, R. (2002). Arm-site binding by lambda -integrase: solution structure and functional characterization of its amino-terminal domain. PNAS **99,** 3434-3439.

Wojciak, J.M., Connolly, K.M., and Clubb, R.T. (1999). NMR structure of the Tn916 integrase-DNA complex **6,** 366-373.

Wolberger, C. (1996). Homeodomain interactions. Current Opinion in Structural Biology. **6,** 62-68.

Zhang, R.-g., Kim, Y., Skarina, T., Beasley, S., Laskowski, R., Arrowsmith, C., Edwards, A., Joachimiak, A., and Savchenko, A. (2002). Crystal structure of *Thermotoga maritima* 0065, a member of the IclR transcriptional factor family. J. Biol. Chem. **277,** 19183-19190.

Zhang, X., Baase, W., and Matthews, B. (1991). Toward a simplification of the protein folding problem: a stabilizing polyalanine alpha-helix engineered in T4 lysozyme. Biochemistry **30,** 2012-2017.