Determinants in the Sequence Specific Binding of Two Plant Transcription Factors, CBF1 and NtERF2, to the DRE and GCC Motifs†

Dongyun Hao,‡ Kazuhiko Yamasaki,‡ Akinori Sarai,§ and Masaru Ohme-Takagi*,‡

Gene Discovery Research Centre, National Institute of Advanced Industrial Science & Technology (AIST), Central 6, Tsukuba 305-8566, Japan, and Tsukuba Life Science Centre, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba 305-0074, Japan

Received November 26, 2001; Revised Manuscript Received January 23, 2002

ABSTRACT: *Arabidopsis* ERF proteins such as DREB1, DREB2, and CBF1 bind to the dehydration-responsive element (DRE), which has the sequence TACCGACAT. Mutation analyses reveal that a central 5 bp CCGAC core of the DRE is the minimal sequence motif (designated as the DRE motif in this paper), to which the ERF domain fragment of CBF1 (CBF1-F) binds specifically with a binding K_d at the nanomolar level. In contrast, the ERF domain fragment of the tobacco ERF2 (NtERF2-F) does not interact with the DRE motif, but restrictedly recognizes the sequence containing a minimal 6 bp GCCGCC motif (designated as the GCC motif in this paper). However, CBF1-F binds to the GCC motif with a binding activity similar to its binding activity for the DRE motif. These in vitro binding variations were further demonstrated through reporter cotransformation assays, suggesting that the DRE and GCC motifs are two similar sequence motifs sharing a common core region of CCGNC with a discriminating guanine base at the 5′-end of the GCC motif. Binding analyses with the mutated ERF domain show that such a unique binding of NtERF2-F to the GCC motif can be altered by the substitution of A14 with valine in β-strand 2 of its ERF domain, the mutant NtERF2-F, ERFav, acquiring a binding to the DRE motif with a K_d comparable to that for CBF1-F binding to the DRE motif. This demonstrates that A14 is an important determinant of the NtERF2-F binding specificity. A possible mechanism of the binding specificity determination is discussed.

Ethylene-responsive element (ERE) binding factors (ERFs)[1] are the products of a large multigene family and are involved in many critical processes of plant development through gene regulation and in mediating signal transduction pathways in response to environmental stress (see a review in ref *1*). The ERE is an 11 bp sequence motif (TAAGAGCCGCC) (*2*−*4*), termed the GCC box, which was shown to be the target of the ERFs (*5*). ERFs were first identified in tobacco, and originally designated as ERE binding proteins (EREBPs) (*5*), but were later renamed ERE binding factors (ERFs) (*6*). Since then, an increasing number of ERFs from various plants have been reported which are divergent except for a highly conserved DNA binding domain, the so-called ERF domain, of 58 or 59 amino acids. ERF proteins have not been found in animals, yeast, or fungi, indicating that this is a group of plant specific proteins. Bioinformatics analysis of the *Arabidopsis* genome has identified some 124 genes encoding the ERFs [in which ERFs are termed the EREBP subfamily under the AP2/EREBP family (*7*)].

Arabidopsis ethylene-responsive element binding factor 1 (AtERF1) is so far the only ERF in this superfamily whose ERF domain structure has been determined (*8*). The ERF domain of AtERF1 consists of a three-stranded (β1−3) antiparallel β-sheet, and an α-helix packed almost parallel to the β-sheet. Unlike other DNA-binding proteins that interact with DNA through amino acids of an α-helix, the ERF domain of AtERF1 makes contact specifically with 9 bp of the AGCCGCC-containing double helix in the major groove through seven amino acids of mainly arginines and tryptophans (R6, R8, W10, E16, R18, R26, and W28) in the β-sheet (see Figure 1). It is interesting that, except W10, these amino acids are found to be completely conserved in the ERF domains of all reported ERFs (data not shown). W10 is also highly conserved, but two notable exceptions are the ERF domain of CBF1 and DREB1. The degree of similarity of the ERF domains varies from 60 to >90%, implying that they may fold into a similar conformation and recognize a similar cis-acting motif in the promoter region of their target genes. Our previous studies have revealed that tobacco ERF2 and ERF4 (formerly EREBP2 and EREBP4, respectively), the homologue ERFs of AtERF1, exhibit specific binding to a 6 bp GCCGCC core of the GCC box (designated as the GCC motif in this paper) with a K_d at the picomolar level (*9*). G6, G9, and C11 (see Figure 1 for position numbering) are found to be essential for ERF binding, since mutation at these bases completely abolishes

† D.H. was a recipient of an AIST fellowship from the former Agency of Industrial Science and Technology (AIST), MITI, Japan, and is the current recipient of an STA fellowship from the Science and Technology Agency of Japan.

* To whom correspondence should be addressed. Phone: (+81) 298 616500. Fax: (+81) 298 616505. E-mail: m-takagi@aist.go.jp.

‡ National Institute of Advanced Industrial Science & Technology.

§ The Institute of Physical and Chemical Research.

[1] Abbreviations: CRT/DRE, C-repeat/dehydration-responsive element; CBF1, CRT/DRE-binding factor 1; ERF, ethylene-responsive element binding factor; NtERF2, *Nicotiana tabacum* ERF2 (formerly EREBP2); DBD, DNA binding domain; EMSA, electrophoresis mobility shift assay; SDS−PAGE, SDS−polyacrylamide gel electrophoresis; CaMV 35S, cauliflower mosaic virus 35S protein gene.

10.1021/bi015979v CCC: $22.00
Published on Web 03/07/2002

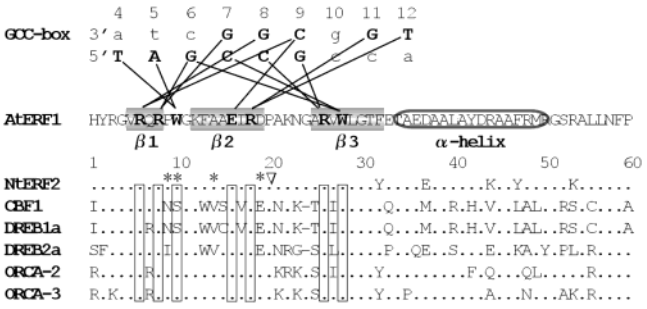

FIGURE 1: Sequence alignment of selected ERF domains and the amino acid−base contacts between the ERF domain of AtERF1 and the GCC motif. The contacting bases and amino acids are in bold, and the intermolecular contacts are shown by lines (interpreted according to the data of ref *8*). The α-helix and *β*-sheets are marked with a patch box and shaded bars, respectively. The identical amino acids are shown with dots, and the seven conserved base-contacting residues are enclosed by a vertical box. The amino acids with asterisks or a triangle symbol in the NtERF2 sequence are those residues of which mutations were made and the position at which the chimeric domain exchange mutations were made, respectively.

binding by NtERF2 and NtERF4 (*9*). Other bases within the GCC motif show moderate effects on single-base substitutions varying from ERF to ERF. The revealed residue−DNA contacts are consistent with the findings of the NMR solution structure of the AtERF1−GCC box complex (*8*).

Some other ERF proteins have been reported to bind to different sequence motifs. *Arabidopsis* C-repeat/dehydration-responsive element binding factor 1 (CBF1), for instance, possesses an ERF domain that binds to a cis-acting element, with the sequence TGGCCGAC, in the promoter region of the *cor15a* gene, and functions as a transcriptional activator leading to cold regulation activities in plants (*10*−*13*). Two counterpart proteins of CBF1, the cold/dehydration-responsive element (DRE) binding proteins, DREB1a and DREB2a, were also shown to bind to a similar 9 bp cis-acting element with the sequence TACCGACAT, and named DRE (*14*). A core motif, CCGAC, is apparently conserved between these two genes, and also among other cold- or dehydration-inducible genes (*15*−*21*), including the cold/dehydration-independent gene *Str* (*22*, *23*).

A number of ERFs have been shown to be capable of binding to the GCC box (see a review in ref *1*), although only two ERFs from tobacco have been shown to recognize restrictedly the 6 bp GCCGCC motif (*9*). Extensive genetic and molecular studies on many ERF proteins have provided insight into the regulatory characteristics and diverse functions of this transcription factor family. However, little information has been obtained on the biochemical and structural mechanism of ERF recognition in general. In this work, the ERF domain of CBF1 was characterized in detail and its DNA binding consensus refined. The in vitro binding results that were obtained are further confirmed by in vivo reporter cotransformation assays. Finally, amino acid mutagenesis of the ERF domain was employed in the binding studies in an effort to explore further the possible determinants of binding variation.

EXPERIMENTAL PROCEDURES

Preparation of the ERF Domains and Their Mutants. The cDNA for CBF1 was isolated via the polymerase chain reaction (PCR) methodology from an *Arabidopsis* cDNA library. The coding region of the ERF domain fragment from CBF1 (amino acids 37−142) was prepared via the PCR methodology, and the expressed fragment is designated in this paper as CBF1-F, which contains 10 and 38 amino acids in the N- and C-terminal flanking regions, respectively. The ERF domain of tobacco NtERF2 with 10 and 8 amino acid flanking regions, respectively (NtERF2-F), was prepared as described previously (*9*). Point mutations in NtERF2-F were introduced by PCR-based site-directed mutagenesis. The chimeric ERF domain mutants of ERF2-F and CBF1-F, ENCC (the ERF2-F N-terminal half and the CBF1-F C-terminal half) and CNEC (the CBF1-F N-terminal half and the ERF2-F C-terminal half), respectively, were produced by fusing the corresponding half-domain coding regions that were synthesized by PCR (see Figure 1 for the mutation junction). CNEC was cloned into the pET3C plasmid (Novagen), while ENCC was cloned into the pET16b plasmid (Novagen); both were expressed in *Escherichia coli* and purified as described previously (*9*). Protein concentrations were determined using the bicinchoninic acid protein assay kit (Pierce), and further confirmed by the method of Gill and von Hippel (*24*).

Quantitative Binding Assays and Titration Analyses. Two 16 bp fragments, DREwt (5′-ATAC**TACCGACAT**GAG-3′) and GCCwt (5′-CA**TAAGAGCCGCC**ACT-3′), from the promoter region of the *Arabidopsis rd29A* gene (*14*) and the tobacco *Gln2* gene (*5*), respectively, were prepared together with their mutants by chemically synthesizing both strands. The DRE and GCC box within the fragments are highlighted, underlined, and numbered from the first nucleotide of the 5′-terminus (see Table 1 and Figure 1 for numbering). After annealing had been carried out, double-stranded DNA fragments of the wild type and mutants were labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase, and purified just before the binding experiments. Quantitative electrophoresis mobility shift binding assays (EMSAs) were performed as described previously (*9*). The binding dissociation constant (K_d) for dissociation of protein from DNA was determined by binding titration analysis, in which a nonlinear regression was performed using Sigma Plot version 7.0 (SPSS Science) using the equation

$$y = a + \frac{b}{1 + K_d/x}$$

where y is the percent binding, x is the protein concentration, and a and b are constants of minimal and maximal y, respectively. A synthesized 16 bp oligonucleotide that shows no context of binding site was tested for nonspecific binding, of which the K_d was determined to be ≥ 1 μM.

Transient Expression Assays. The reporter gene constructs were prepared by fusing a 4-fold multimerized fragment containing the GCC motif, DRE or their mutants, and a minimal 35S promoter (base pairs −46 to 8) of cauliflower mosaic virus (CaMV) to the luciferase (LUC) gene. The effector plasmid, *35S-NtERF2*, was described previously (*25*), and *35S-CBF1* was constructed by fusing the coding region of the full-length *CBF1* gene to a CaMV 35S promoter. The tobacco mosaic virus (TMV) Ω sequence was inserted between either the *NtERF2* or *CBF1* coding region and a CaMV 35S promoter to enhance the translation efficiency. Luciferase assays were performed using the Dual-

Table 1: Truncation and Mutational Binding Assays for Binding of CBF1-F to DRE and the GCC Box[a]

Fragment	*cis*-acting element-containing sequences	K_d
	1 5 10 15	
DREwt	5′-ATAC**TACCGACAT**GAG-3′	2.10 ± 2.0 nM
5′M1	5′-C**TACCGACAT**GAGTTC-3′	2.85 ± 0.8 nM
5′M2	5′-**TACCGACAT**GAGTTCC-3′	6.72 ± 0.6 nM
3′M1	5′-ATATAC**TACCGACAT**G-3′	5.84 ± 1.5 nM
3′M2	5′-GATATAC**TACCGACAT**-3′	6.19 ± 1.3 nM
Mm	5′-tatg**TACCGACAT**cac-3′	5.07 ± 1.1 nM
DREt5a	5′-ATAC**aACCGACAT**GAG-3′	5.39 ± 0.9 nM
DREa6t	5′-ATAC**TtCCGACAT**GAG-3′	9.05 ± 0.7 nM
DREc7t	5′-ATAC**TAtCGACAT**GAG-3′	40.3 ± 14 nM
DREa12t	5′-ATAC**TACCGACtT**GAG-3′	3.93 ± 1.6 nM
DREt13a	5′-ATAC**TACCGACAa**GAG-3′	4.36 ± 0.8 nM
DREc11t	5′-ATAC**TACCGAtAT**GAG-3′	50.5 ± 12 nM
GCCwt	5′-CA**TAAGAGCCGCC**ACT-3′	5.63 ± 0.6 nM
GCCg9t	5′-CA**TAAGAGCCtCC**ACT-3′	0.20 ± 0.04 μM
GCCc10a	5′-CA**TAAGAGCCGaC**ACT-3′	5.51 ± 0.3 nM
GCCg6a-c10a	5′-CA**TAAGAaCCGaC**ACT-3′	5.32 ± 0.7 nM

[a] A 16 bp wild-type DRE sequence (DREwt) or GCC box sequence (GCCwt) contains the DRE or GCC box, respectively, the sequences of which are highlighted and underlined. Base position numbering is the same as that shown in Figure 1 for the GCC box. Quantitative EMSA was carried out using either the DREwt or the GCCwt, and their single-base substitution mutants (lowercase) interacted with CBF1-F at a concentration gradient from 10 μM to 10 pM. The K_d values are represented as the means ± standard deviations of triplicate determinations.

Luciferase Reporter Assay System (Promega, Madison, WI), and the luminescence signal detected by a luminescence reader (BLR-201; Aloka, Tokyo, Japan). In cotransformation assays, the reporter and effector constructs, in a ratio of 1:2, were coated on gold particles and bombarded onto actively growing *Arabidopsis* leaves as described previously (*26*). As an experimental control, a pUC119 plasmid was used instead of the effector, and cotransformed together with the reporter gene. After bombardment, the leaves were incubated in sterilized 50% Murashige & Skoog basic plant culture medium at 22 °C for 6 h before quantification for luciferase activity. The detected luciferase activity from a minimum of three independent replicates is represented as the mean ± standard deviation of relative activity, which was normalized by dividing the value by the control.

RESULTS

Minimal DRE Motif for CBF1 Binding. Quantitative EMSA demonstrated that CBF1-F binds to the DREwt fragment with a binding dissociation constant (K_d) of 2.1 ± 2.0 nM, which is a typical value for monomeric binding. Binding titration analysis (see Figure 2) revealed that it is a single-site interaction, since the nonlinear regression using the single-site binding equation matched the best to the experimental data, while the double-site binding equation was shown not to match (data not shown).

To verify the minimal DRE element that is required for stable binding by CBF1, the 16 bp wild-type DRE-containing fragment (DREwt) was tested for binding by the ERF domain fragment of CBF1 (CBF1-F). Initial assays were conducted

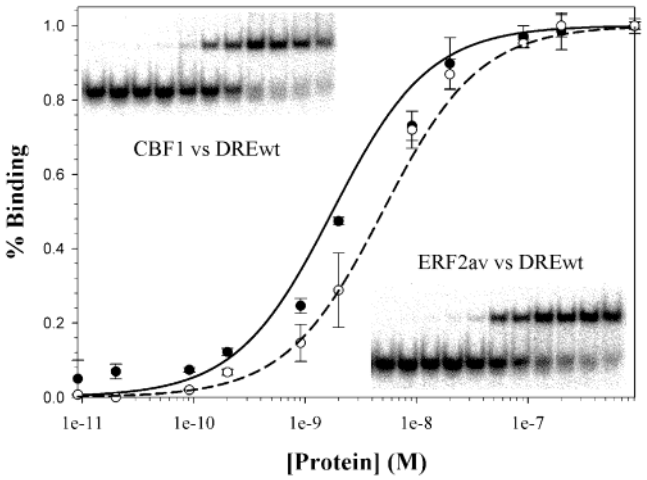

FIGURE 2: Quantitative EMSA and binding titration analyses of CBF1-F and NtERF2av binding to DREwt. Data for the binding titration of CBF1 to DREwt and NtERF2av to DREwt are shown by solid and empty dots, respectively, of each sampling point, which were extracted from quantitative EMSA shown alongside each titration curve. Nonlinear regressions were performed to give the solid and dashed curves for binding of CBF1 and NtERF2av, respectively, using the single-site equation presented in Experimental Procedures. All data are represented as means ± standard deviations of triplicate determinations.

in which mutations and truncations were made at the flanking regions of the 9 bp DRE (i.e., TACCGACAT), keeping the length of 16 bp as presented in Table 1. The flanking region of the 16 bp DREwt may be shortened up to 4 and 3 bp at the 5′- and 3′-ends, respectively, as judged from the K_ds obtained with the 5M1, 5M2, 3M1, and 3M2 fragments, after taking into account statistical variation, without affecting the binding by CBF1-F. This was further verified by the assay of binding of CBF1-F to a mutant DREwt of which both flanking regions of the 9 bp DRE were replaced with their complementary bases (Mm). This suggests that the central 9 bp DRE satisfies the specific binding of CBF1-F.

Subsequently, the minimal motif of DRE was identified using single-base substitutions within the 9 bp DRE of the 16 bp DREwt. This was due to the consideration of possible deformation at both ends of the double helix as a consequence of further truncation of the DREwt toward the central region. Table 1 shows that the first thymine from the 5′-end of the DRE (T5) can be replaced with an adenine (DREt5a) without affecting binding activity. Substitution of the next adjacent A6 with thymine (DREa6t) decreases binding activity slightly, while further mutation of the next C7 to thymine (DREc7t) causes a 10-fold decline in binding activity, suggesting that the adenine at position 6 (A6) is likely to be the 5′-end border of the DRE motif. At the 3′-terminus, substitutions of A12 and T13 (DREt13a and DREa12t, respectively) did not affect the binding by CBF1-F in comparison with the binding to the DREwt. However, mutation at C11 (DREc11t) decreases binding activity by more than 10-fold. These results suggest that the 5 bp core region of DRE, CCGAC, is the minimal sequence motif that ensures stable binding by CBF1-F, and was thus designated as the DRE motif in this paper.

CBF1 Binds in Vitro to both the DRE Motif and the GCC Motif with Similar Binding Activities. Due to the similarity of the DRE motif to the GCC motif, it is rational to predict

that the DRE motif may be closely related to the GCC motif, and thus, the binding of CBF1-F to the GCC motif fragment (GCCwt) and its mutants was studied. The results of binding titration analyses presented in Table 1 show that CBF1-F binds to GCCwt with a K_d of 5.6 ± 0.6 nM, which is close to the K_d for binding of CBF1-F to DREwt. When the guanine at position 9 (G9) of GCCwt was substituted with thymine (GCCg9t), the binding activity decreased approximately 200-fold. However, replacement of C10 (GCCc10a), or both C10 and G6 with adenine (GCCg6a-c10a), does not affect the binding by CBF1-F. In fact, the GCCg6a-c10a mutant of GCCwt resembles a DRE motif, suggesting that CBF1-F does not require a G at position 6 to conduct typical monomeric binding. This is apparently different from the binding of NtERF2 to the GCC motif of which G6 is essential for stable binding (*9*).

CBF1 Activates the Transcription of a Reporter Gene in Vivo through Binding to both the DRE Motif and the GCC Box. To further investigate if CBF1 is capable of regulating transcription in plants through binding to either the DRE motif or the GCC motif that has been observed in vitro, reporter−effector cotransformation assays were performed. The effector plasmid possessing the coding region of the full-length *CBF1* or *NtERF2* gene, driven by a CaMV 35S promoter, was coexpressed in *Arabidopsis* leaves, together with the reporter plasmid containing either the DRE motif or the GCC motif, or their mutants, fused to the upstream regulatory region of a luciferase gene (*LUC*). Figure 3B shows that when the DRE motif-carrying reporter plasmid (DREwt) is coexpressed with the *CBF1* effector plasmid, luciferase activity increases by ~5-fold over the control that contains no effector plasmid. The *CBF1* effector plasmid was also found to be capable of activating the transcription of *LUC* that was fused to the GCC motif (GCCwt), with a greater than 3-fold increase in luciferase activity over the control. Moreover, a greater than 6-fold increase in luciferase activity was detected when the *CBF1* effector was coexpressed with a reporter gene carrying the mutant GCC motif, GCCt, G8 of which was replaced with T. However, no luciferase activity was detected when the *NtERF2* effector was coexpressed with the reporter genes containing either the DRE motif (DREwt) or GCCt. Neither *CBF1* nor the *NtERF2* effector activates the *LUC* gene fused to DREt (i.e., CCTAAC), while the *NtERF2* effector activates *LUC* fused to the GCC motif (GCCwt), resulting in an increase in luciferase activity of more than 5-fold over the control, a level similar to the activation of *LUC* fused to DREwt by *CBF1*. These results agree with the in vitro binding observations in this and previous work (*9*), indicating that CBF1 recognizes a DRE motif shorter than the 6 bp GCC motif to activate gene transcription. Comparison of the binding motifs of the four binding sequences [DREwt, DREt, GCCwt, and GCCt (see Figure 3A)] has indicated that G6 of the GCC motif is a prerequisite component that is recognized specifically by NtERF2 for transcription activation, but not by CBF1.

Site-Induced Mutagenesis Alters the Binding Property of NtERF2. Although they share a high degree of similarity in their ERF domains, CBF1-F binds to a 5 bp DRE motif while NtERF2-F binds to a 6 bp GCC motif. To explore the possible amino acid(s) in the ERF domain that determines the binding divergence, NtERF2-F was adopted for site-directed mutagenesis studies because its binding to the GCC motif has been extensively characterized (*9*). Since six of the seven amino acids in the β-sheet of its ERF domain that directly contact DNA are conserved among all ERF proteins, the determining amino acid(s) apparently lies among those divergent residue(s) in the ERF domain. From studies on the structure of the AtERF1−GCC box sequence complex (*8*), and sequence alignment of the CBF1-F and NtERF2-F ERF domains, a few candidate amino acids were selected (see Figure 4) in which their side chain arrangements exhibit potential to affect the local and/or globular conformation. Two of those divergent amino acids are W10 and P9 of NtERF2-F that are replaced with S10 and N9 in CBF1-F, respectively (see Figure 1). Table 2 shows that substitution of P9 with asparagine in NtERF2-F (ERFpn) increases slightly the binding activity to GCCwt, while the change of

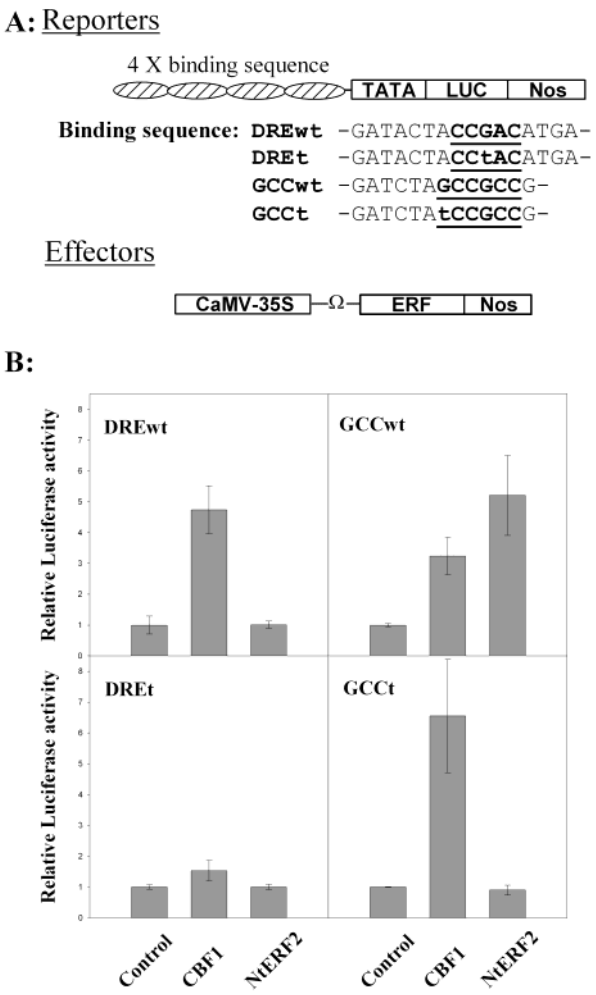

FIGURE 3: CBF1 and ERF2 activation of the transcription of the luciferase reporter gene driven by the DRE motif, the GCC motif, and their mutants. (A) Schematic diagram of the reporter and effector plasmids used in a transient assay. The *LUC* reporter gene contains a tetramer sequence of a cis-acting binding motif, DREwt, DREt, GCCwt, or GCCt, the sequences of which are highlighted and underlined. The effector was constructed with a full-length *NtERF2* or *CBF1* cDNA that is controlled under the CaMV 35S promoter upstream of a translation enhancer (Ω) from tobacco mosaic virus. (B) Transactivation of the reporter gene induced by those effectors. The control (Ctrl) in the transient assays is the same as the experiments without the addition of effector. Transactivation results are shown as relative luciferase activity per control. All data are represented as means ± standard deviations of triplicate determinations.

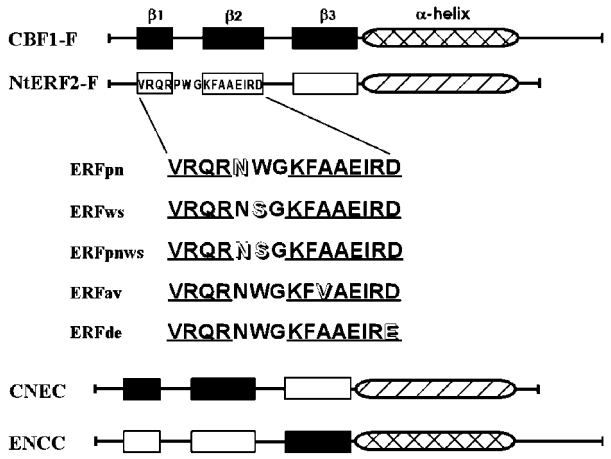

FIGURE 4: Schematic diagram of the truncated CBF1-F and NtERF2-F fragments, and their mutants, used in this study. The results of binding of these proteins to the DRE motif, the GCC motif, or their mutants are shown in Table 2.

Table 2: Binding of CBF1-F, NtERF2-F, and Their Mutants to the DRE and GCC Motifs[a]

ERF protein	cis-acting element-containing sequence	K_d
NtERF2-F		>1.00 μM
CBF1-F		1.46 ± 1.0 nM
ERFpn		>1.00 μM
ERFws	DREwt	>1.00 μM
ERFpn.ws	5′-ATACTA**CCGAC**ATGAG-3′	>1.00 μM
ERFav		4.98 ± 2.3 nM
ERFde		0.25 ± 0.1 μM
CNEC		4.01 ± 1.4 nM
ENCC		1.65 ± 1.2 nM
NtERF2-F		0.11 ± 0.3 nM
CBF1-F		5.63 ± 0.6 nM
ERFpn		62.6 ± 21.0 pM
ERFws	GCCwt	38.5 ± 17.0 nM
ERFpn.ws	5′-CATAAGA**GCCGCC**ACT-3′	6.03 ± 1.3 nM
ERFav		0.45 ± 0.3 nM
ERFde		4.57 ± 1.1 nM
CNEC		24.3 ± 19.0 nM
ENCC		1.36 ± 0.9 nM

[a] The details of the CBF1-F and NtERF2-F proteins and their mutants are presented in Figure 3. Quantitative EMSA was performed using the DREwt or GCCwt with a protein concentration gradient from 10 μM to 10 pM (or lower for NtERF2 binding to the GCCwt). The K_d values are represented as means ± standard deviations of triplicate determinations.

W10 to serine (ERFws), or both W10 and P9 to asparagine and serine (ERFpnws), causes a 100- or 10-fold decrease in binding activity to GCCwt, respectively. However, none of these mutations can bind to the DREwt. These results suggest that W10 and P9 are unlikely to contribute to the alteration of the binding specificity of NtERF2.

The other divergent residues are at positions 14 and 19 of β-strand 2, where A14 and D19 of NtERF2-F are substituted by valine and glutamic acid, respectively, in CBF1-F. The A14V mutant of NtERF2-F (ERF2av) appears not to affect the binding activity to GCCwt, but acquires a new capability for binding to DREwt with a binding K_d comparable to that of CBF1-F binding to DREwt (Table 2). Titration analysis shows that the binding of NtERF2av to DREwt matches the single-site binding kinetics like that of CBF1-1 binding to DREwt (see Figure 2), suggesting that the binding is specific. The substitution of D19 with E in NtERF2-F (ERFde) causes a 10-fold lower binding activity to GCCwt compared with that of NtERF2-F, but also yields a very weak binding to DREwt with a K_d of 0.1 μM. These results indicate that A14 in the β-sheet of the ERF domain is likely to be an important determinant of NtERF2 in switching from GCC motif binding to DRE motif binding.

Chimeric Mutation of Inter-ERF Domains. To further verify if the determinants of binding specificity are region-dependent, two chimeric domain exchange mutants, CNEC and ENCC, were prepared as shown in Figure 4. The CNEC includes the first and second β-strands of CBF1-F and the third β-strand and α-helix of NtERF2-F, while ENCC has the first and the second β-strands of NtERF2-F and the third β-strand and the α-helix of CBF1-F. Table 2 shows that both chimeric ERF proteins bind to either DREwt or GCCwt with different binding activities. CNEC binds to DREwt with a K_d close to that of CBF1-F binding to DREwt, but binds to GCCwt with a 10- or 100-fold lower activity than that of CBF1-F or NtERF2-F binding to the GCCwt, respectively. ENCC shows similar binding activity to either GCCwt or DREwt, but its binding activity is ∼10-fold lower than that of NtERF2-F. It is interesting that ENCC binds to DREwt as strongly as CBF1-F, although it is not surprising considering that ENCC has inherited A14 and D19, the residues that have been shown to effectively change the binding specificity of NtERF2, from the ERF domain of NtERF2-F. Thus, these results indicate that the binding specificity may also be altered by domain exchange mutations, although it has not been shown whether the chimeric ERF proteins fold into a local conformation similar to that of their wild-type precursors.

DISCUSSION

In vitro analyses with truncation mutations and base substitution binding assays (see Table 1) have identified a 5 bp DRE motif, CCGAC, as the minimal sequence motif necessary for recognition by CBF1, and is also likely to be used by other transcription factors such as DREBs (*14*). There are, so far, two DRE variants that have been reported, TACCGACAT from the *cor78* and *rd29A* genes (*10*, *12*) and TGGCCGAC from the *cor15a* gene (*11*), in *Arabidopsis*. These DRE variants all share a common DRE motif that has also been found in the promoter region of the cold-inducible gene *bn115* in oilseed rape (*20*). A recent full-length cDNA microarray analysis has shown that 11 target genes of DREB1a, a close homologue of CBF1, contain the DRE or DRE-related CCGAC motif in their promoter regions (*27*). These results are consistent with our findings.

The truncated ERF domain fragment of CBF1 used in this study, CBF1-F, interacts with a 16 bp DRE-containing sequence with a binding K_d of ∼2.10 nM. This is a typical value for K_d found in the monomer protein, or subunits of the complex protein, and DNA interactions (*28*), but the binding is several 10-fold weaker than the binding of NtERF2-F to the GCC box (*9*). Both in vitro and in vivo binding analyses have confirmed that G6 of the GCC motif is a prerequisite for NtERF2 to bind and to activate gene transcription with a uniquely high in vitro binding activity compared with previously reported cooperative DNA binding studies (*28*). Interestingly, however, CBF1-F interacts with the sequence containing not only the GCC motif but also its

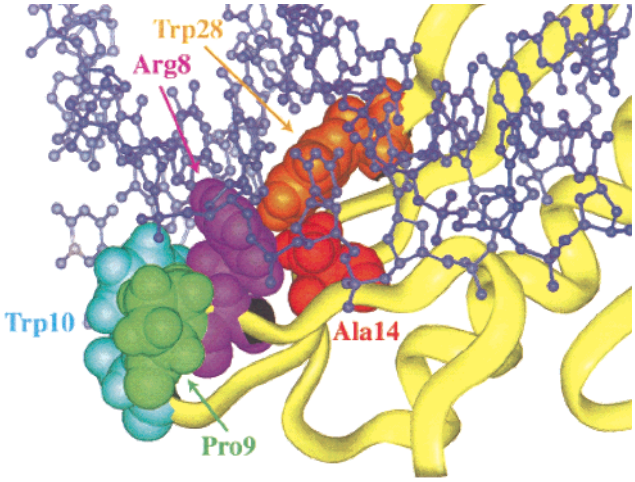

FIGURE 5: Intermolecular interface around the mutational positions observed in the tertiary structure of the AtERF1−GCC box complex (PDB entry 1gcc). A part of the DNA, the protein backbone, and amino acid residues of interest are shown in the ball-and-stick, ribbon, and van der Waals surface models, respectively. The figure was prepared using the program Insight II (MSI).

mutants, GCCc10a (GCCGaC) and GCCg6a-c10a (aCCGaC), with a binding activity similar to that of binding to the DRE motif (see Table 2). Indeed, CBF1 activates transcription of a reporter gene driven by either of these two binding motifs in *Arabidopsis* leaves. It is apparent that the DRE motif and GCC motif are two closely related sequence motifs that mediate gene transcription regulation upon ERF binding.

The question of what makes such a binding specificity between CBF1 and NtERF2 so different has been raised, considering that ERF binding proteins share a high degree of similarity in their ERF domains. From the ERF domain−GCC motif binding profile shown in Figure 1, there are two residues, R8 of β1 and W28 of β3, which directly communicate with G6, the apparently discriminating base for the binding of NtERF2 from CBF1, of the GCC motif. However, these residues are found to be completely conserved among all ERF proteins, suggesting that the determinants are those divergent residues in the ERF domain between CBF1 and NtERF2 whose side chains are close to R8 and/or W28. P9 and W10 are two residues of this kind that contact R8 through their side chains (Figure 5). However, the substitution of P9 or W10 with N or S, respectively, does not enable NtERF2-F to bind to the DRE motif (see Table 2 and Figure 4). The slight increase in the GCC motif binding activity caused by the P9N mutation of NtERF2-F (ERFpn) might be caused by an increase in backbone flexibility that may favor access of a base. Thus, the side chain interaction of P9 or W10 to R8 is unlikely to alter the binding specificity.

Through sequence comparison, Liu et al. (*14*) suggested two other residues, A14 and D19 in strand β2 of the DREB ERF domain, which might be important in the determination of ERF binding specificity. Coincidentally, our mutational analysis demonstrates that substitution of A14 in NtERF2-F with the valine of CBF1 (ERFav) does produce binding to the DRE motif with a K_d comparable to the binding of CBF1-F to the DRE motif (Figure 2 and Table 2). The titration analysis reveals a single-site pattern of binding of ERFav to the DRE motif, which is similar to that of CBF1-F binding to DREwt (Figure 2). The solution structure of the ERF domain of AtERF1, a close homologue of NtERF2, shows that A14 itself does not contact DNA, but shows significant van der Waals contacts to both R8 and W28, the side chains of which make direct contact with G6 of the GCC box (*8*; Figures 1 and 5). A possible explanation is that the A14V mutation introduces two methyl groups that impose a shift in the three-dimensional positions of the R8 and W28 side chains and, consequently, result in a significant rearrangement of the protein−DNA interface. In addition, replacement of D19 of NtERF2-F with the counterpart E of CBF1 can also produce a binding to the DRE motif, although the binding activity is very weak. These observations imply that A14 is an important determinant residue at least for some tobacco ERF proteins such as NtERF2 and NtERF4 whose GCC motif binding patterns are very similar (*9*). However, other residues in the ERF domain may also contribute to the binding determination. Our chimeric mutation analyses appear to support this assumption in that both the CNEC and ENCC proteins exhibit binding activities to the DRE motif similar to that of their precursor CBF1-F, although their affinity for binding to the GCC motif is reduced. Also, it has been recently reported that two ERF proteins, ORCA2 and ORCA3 from periwinkle, bind to a sequence containing TAGACCGCC (*22*, *23*), which does not appear to contain the 5′-discriminating guanine, yet interestingly, both proteins contain A14 and D19 instead of V14 and E19 in their ERF domains.

Current findings reveal that CBF1 and NtERF2 are two divergent ERF proteins exhibiting not only unique binding specificity but also different affinities of binding to their target sites. Although CBF1 shares a relatively low level of sequence similarity in its ERF domain in comparison with other ERF proteins (unpublished data), its ERF domain has been predicted to have a conformation similar to that of AtERF1 (*8*), the homologue of NtERF2. However, CBF1 interacts with its 5 bp DRE motif with a typical binding activity of a monomeric interaction, while NtERF2 binds to its 6 bp target GCC motif with an exceptionally high binding affinity. The DRE motif and GCC box are two similar sequence motifs sharing a common core region of CCGNC with a discriminating guanine (G6) at the 5′-end of the GCC motif. The recognition requirements and binding magnitudes of different ERF proteins are likely to be related to the differential regulation of their target genes. These results, therefore, warrant further investigation into the differential recognition requirements among the different members of this ERF superfamily.

ACKNOWLEDGMENT

We are grateful to Dr. Paul Reay (AIST) for the critical reading and discussion of the manuscript.

REFERENCES

1. Riechman, J. L., and Meyerowitz, E. M. (1998) *J. Biol. Chem. 279*, 633−646.
2. Ohme-Takagi, M., and Shinshi, H. (1990) *Plant Mol. Biol. 15*, 941−946.
3. Eyal, Y., Meller, Y., Lev-Yadum, S., and Fluhr, R. (1993) *Plant J. 4*, 225−234.
4. Hart, C. M., Nagy, F., and Meins, F., Jr. (1993) *Plant Mol. Biol. 21*, 121−131.
5. Ohme-Takagi, M., and Shinshi, H. (1995) *Plant Cell 7*, 173−182.

6. Suzuki, K., Suzuki, N., Ohme-Takagi, M., and Shinshi, H. (1998) *Plant J. 15*, 657−665.
7. Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B. K., and Yu, G.-L. (2000) *Science 290*, 2105−2110.
8. Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998) *EMBO J. 17*, 5484−5496.
9. Hao, D., Ohme-Takagi, M., and Sarai, A. (1998) *J. Biol. Chem. 273*, 26857−26861.
10. Yamaguchi-Shinozaki, K., and Shinozaki, K. (1994) *Plant Cell 6*, 251−264.
11. Maker, S. S., Wilhelm, K. S., and Thomashoe, M. F. (1994) *Plant Mol. Biol. 24*, 701−713.
12. Stockinger, E. J., Gilmour, S. J., and Thomashow, M. F. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1035−1040.
13. Jaglo-Ottosen, R. K., Gilmour, S. J., Zarka, D. G., Schabenberger, O., and Thomashow, M. F. (1998) *Science 280*, 104−106.
14. Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Setsuko, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998) *Plant Cell 10*, 1391−1406.
15. Nordin, K., Heino, P., and Palva, E. T. (1991) *Genes Dev. 10*, 3018−3027.
16. Kurkela, S., and Borg-Franck, M. (1992) *Plant Mol. Biol. 19*, 689−692.
17. Horvath, D. P., McLarney, B. K., and Thomashow, M. F. (1993) *Plant Physiol. 103*, 1047−1053.
18. Yamaguchi-Shinozaki, K., and Shinozaki, K. (1993) *Mol. Gen. Genet. 238*, 17−25.
19. Wang, H., Datla, R., Georges, F., Loewen, M., and Cuter, A. J. (1995) *Plant Mol. Biol. 28*, 605−617.
20. Jiang, C., Iu, B., and Singh, J. (1996) *Plant Mol. Biol. 30*, 679−684.
21. Iwasaki, T., Kiyosue, T., Yamaguchi- Shinozaki, K., and Shinozaki, K. (1997) *Plant Physiol. 115*, 1287−1289.
22. Menke, F. L. H., Champion, A., Kijine, J. W., and Memelink, J. (1999) *EMBO J. 18*, 4455−4463.
23. van de Fits, L., and Memelink, J. (2000) *Science 289*, 295−297.
24. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
25. Ohta, M., Ohme-Takagi, M., and Shinshi, H. (2000) *Plant J. 22*, 29−38.
26. Fujimoto, S. Y., Ohta, M., Usui, A., Shinshi, H., and Ohme-Takagi, M. (2000) *Plant Cell 12*, 393−404.
27. Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y., and Shinozaki, K. (2001) *Plant Cell 13*, 61−72.
28. Prabakaran, P., An, J., Gromiha, M., Selvaraj, S., Uedaira, H., Kono, H., and Sarai, A. (2001) *Bioinformatics* (in press).

BI015979V