

A Novel DNA-Binding Motif, Hallmark of a New Family of Plant Transcription Factors^{1,2}

José L. Carrasco, Gema Ancillo³, María José Castelló, and Pablo Vera*

Instituto de Biología Molecular y Celular de Plantas, Universidad Politécnica de Valencia, Consejo Superior de Investigaciones Científicas, 46022 Valencia, Spain

A novel family of plant-specific transcription factors is described. They are structurally related to DBP1 (for DNA-binding protein phosphatase 1), a new transcription factor recently characterized in tobacco (*Nicotiana tabacum*), which exhibits both sequence-specific DNA-binding and protein phosphatase activity (Carrasco et al., 2003). The C-terminal part of DBP factors shows high sequence similarity to protein phosphatases of the 2C class (PP2C; Smith and Walker, 1996), and recombinant tobacco DBP1 is indeed an active protein phosphatase (Carrasco et al., 2003). Within the N-terminal region of these factors, with no significant homology to known protein sequences, we have identified and functionally characterized a novel and highly conserved motif involved in sequence-specific DNA binding. Protein phosphorylation/dephosphorylation enables cells to rapidly and reversibly modulate transcription factor function, and thereby gene expression, in response to signaling stimuli (Whitmarsh and Davis, 2000; Kobor and Greenblatt, 2002). This modulation is accomplished by the coordinated activity of protein kinases and protein phosphatases. Originally considered to act merely by reversing the effects of protein kinases, protein phosphatases are now recognized to fulfill essential regulatory functions in signaling pathways (Luan, 2003). In the last years, biochemical and genetic studies have identified PP2Cs as negative modulators of stress-responsive signaling pathways in animals, yeast, and also in plants (Maeda et al., 1994; Sheen, 1996; Gaits et al., 1997; Meskiene et al., 1998; Takekawa et al., 1998; Gosti et al., 1999; Meskiene et al., 2003). The identified

DBP1 protein from tobacco plants was found to participate in the transcriptional regulation of the expression of the *CEVII* gene (Carrasco et al., 2003), a defense-related gene that is induced in susceptible plants as a consequence of viral infection (Mayda et al., 2000). Plant PP2Cs are characterized by the presence of specific, sequence-unrelated N-terminal extensions of undefined function (Rodríguez, 1998). Thus, specific functional features of DBP1, like its DNA-binding capacity, likely rest on this region, although sequence analysis did not result in the identification of any obvious, previously characterized DNA-binding motif. Because of the unique combination of protein phosphatase activity and sequence-specific DNA binding within the same protein molecule, the identification of tobacco DBP1 unveiled a new mechanism of transcriptional control remaining to be fully elucidated.

With the aim of further characterizing this novel class of transcriptional regulators, we have investigated the function of the distinctive N-terminal region of DBP1 on the hypothesis that it would likely be involved in DNA-protein interactions. By using TBLASTN (Altschul and Gish, 1996), with the BLOSUM62 matrix and a reference E-value of 0.01, we compared the N-terminal region of DBP1 with sequences in the National Center for Biotechnology Information, The Arabidopsis Information Resource, and The Institute for Genomic Research databases. Significant hits were found only in plants and corresponded to proteins also bearing downstream PP2C domains. Sequence alignment of the N-terminal region of the identified DBP-like factors (Fig. 1A) denotes the existence of a conserved domain, comprising a motif almost invariant in all the aligned sequences from both monocot and dicot plant species. This highly conserved motif, F(L,V)PX(L,V,I)RSGXWSDIG, which we refer to as DNC motif (for DBP N-terminal core), might represent a structural element critical for DBP function. By contrast, sequences within the very N terminus are more divergent, although keep significant similarity within taxa, and could confer functional specificity. Full-length deduced amino acid sequences could be retrieved from only some of the identified cDNAs, the rest corresponding to expressed sequence tags. With the full-length sequences, we generated phylogenetic trees using Phylip programs based on the Neighbor-Joining method, and the Fitch-Margoliash and least-squares methods (Fig. 1B).

¹ This work was supported by the Spanish Ministry of Science and Technology (grant to P.V.).

² We wish to dedicate this article to the loved and esteemed memory of Prof. Dr. Julio López-Gorgé, Department of Plant Biochemistry, Cell and Molecular Biology, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, Granada, Spain, who died June 7, 2004, at the age of 69.

³ Present address: Instituto Valenciano de Investigaciones Agrarias, Carretera Moncada-Náquera, km. 4.5, 46113 Moncada (Valencia), Spain.

* Corresponding author; e-mail vera@ibmcp.upv.es; fax 34-963877859.

www.plantphysiol.org/cgi/doi/10.1104/pp.104.056002.

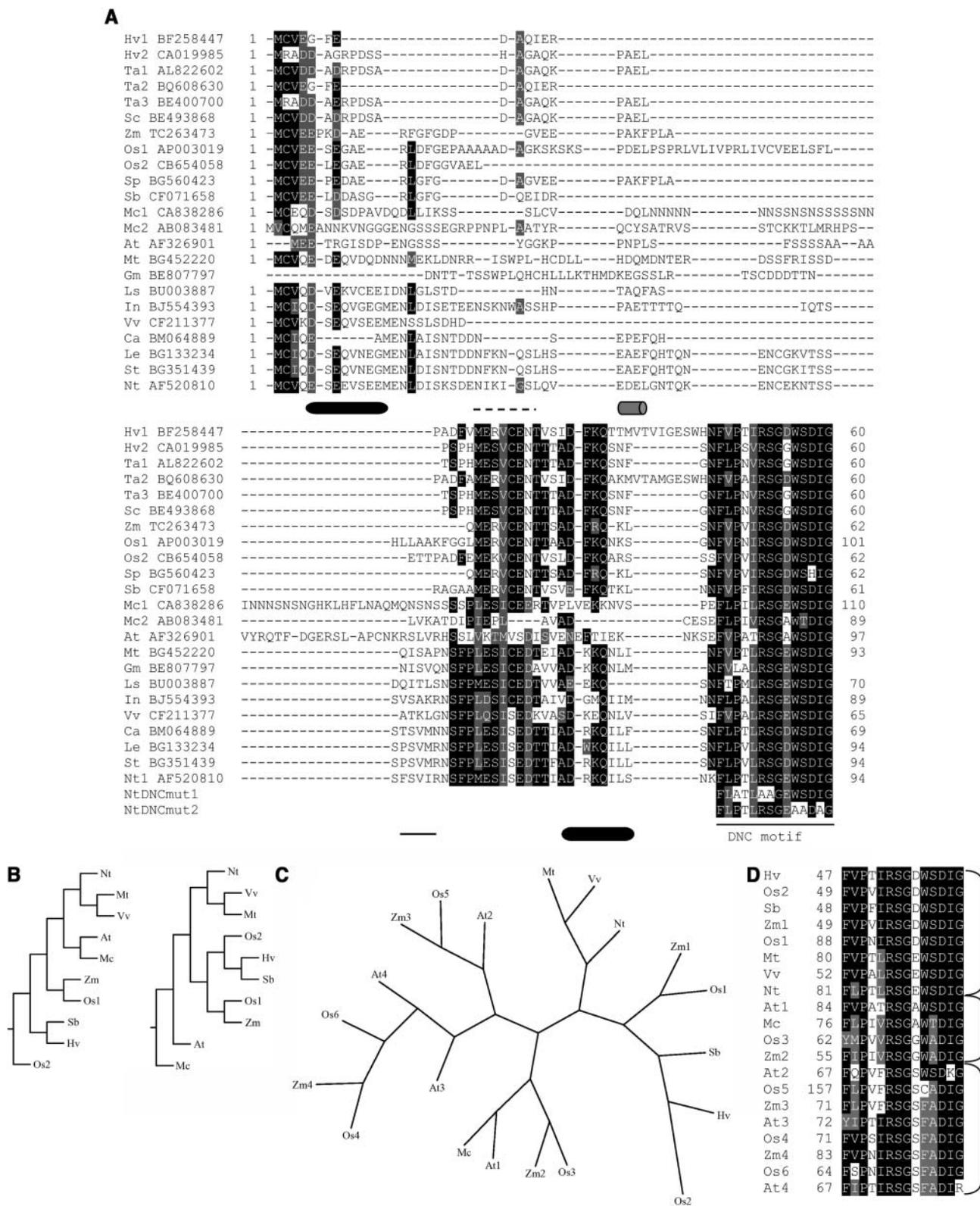


Figure 1. A, Multiple sequence alignment of the N-terminal regions of the identified DBP-like factors. The alignment was generated with T-coffee (<http://igs-server.cnrs-mrs.fr/Tcoffee/>; Notredame et al., 2000) and viewed with Boxshade (http://www.ch.embnet.org/software/BOX_form.html). Black boxes denote sequence identity, whereas gray boxes refer to conservative changes. Dashes indicate gaps included for a better alignment. Species abbreviations: Hv, *Hordeum vulgare*; Ta, *Triticum aestivum*; Sc, *Secale cereale*; Zm, maize; Os, rice; Sp, *Sorghum propinquum*; Sb, *Sorghum bicolor*; Mc, *Mesembryanthemum crystallinum*; At, Arabidopsis; Mt, *Medicago truncatula*; Gm, *Glycine max*; Ls, *Lactuca sativa*; In, *Ipomoea nil*; Vv, *Vitis vinifera*;

Common to the generated trees, the two full-length DBP sequences available for rice (*Oryza sativa*) are placed in separate branches and closer to sequences from other monocot species. This result could be indicative of the existence of different classes of DBP factors and supportive of a functional diversification of this novel family of plant transcriptional regulators. In this respect, reducing the stringency of the homology search revealed the existence of additional full-length sequences encoding DBP1-related PP2Cs in *Arabidopsis* (*Arabidopsis thaliana*), maize (*Zea mays*), and rice, all endowed with a less conserved DNC motif. As shown in Figure 1C, with the exception of Zm2 and Os3, which cluster close to Mc and At1, the rest of the newly identified sequences are grouped into a separate branch in a phylogenetic tree generated with Quicktree (Howe et al., 2002) and may represent a different subclass of DBP-related proteins. Remarkably, these phylogenetic relationships, which were defined on the basis of the full-length deduced amino acid sequences, also reflect sequence conservation within the DNC motif (Fig. 1D). According to this criterion, the new DBP-like sequences also fall in a different group, with Zm2 and Os3 showing, along with Mc and At1, an intermediate degree of conservation when compared to the originally identified sequences (Fig. 1D). Some of the amino acid residues exchanged in the new sequences were strictly conserved in the first group of DBP-related factors and shown to be involved in DNA binding to the 3a4 probe by both tobacco and *Arabidopsis* DBP1. Whether these additional DBP variants show the same sequence specificity or even retain the capacity of binding DNA remain to be elucidated.

Sequence homology in the N-terminal region among the identified DBP factors suggests functional conservation of the DNA-binding activity inherent to tobacco DBP1. To validate this hypothesis, the *Arabidopsis* DBP1 homolog was expressed in *Escherichia coli* fused to a C-terminal hexa-His tag (AtDBP1-his₆). The *Arabidopsis* protein is distantly related to tobacco DBP1, with sequence similarity in the N-terminal region almost restricted to the DNC motif (Fig. 1A) and separate clustering in phylogenetic trees (Fig. 1B). As shown in Figure 2, the purified AtDBP1-his₆ recombinant protein was able to bind to the 3a4 probe in electrophoretic mobility shift assays (EMSA) with similar affinity as

tobacco DBP1, the latter also expressed and purified as a C-terminal fusion to a hexa-His tag (DBP1-his₆). The 3a4 probe derives from the promoter of the *CEVI-1* gene (Mayda et al., 2000) and is specifically recognized in vitro by DBP1 (Carrasco et al., 2003). To analyze the implication of the N-terminal region of tobacco DBP1 in sequence-specific DNA binding, a fusion protein consisting of the maltose-binding protein (MBP) and the N-terminal region of DBP1 (MBP-DBP1Nt) was expressed in *E. coli*. The purified recombinant protein was able to bind to the radiolabeled 3a4 DNA probe in vitro, whereas MBP alone was not (Fig. 3A, lanes 2 and 3). In addition, binding to the 3a4 probe mediated by the N-terminal region of DBP1 showed the same sequence specificity as the full-length protein; thus, the 3a4mut double-stranded oligonucleotide, which is not recognized by DBP1 (Carrasco et al., 2003), was not bound by MBP-DBP1Nt (Fig. 3A, lane 6). Therefore, sequence-specific DNA binding relies on amino acid residues in the N-terminal region of DBP1. Moreover, the N-terminal region of *Arabidopsis* AtDBP1 also retained DNA-binding activity when expressed in *E. coli* as an MBP fusion protein (MBP-AtDBP1Nt), and showed the same sequence specificity as full-length tobacco DBP1 and DBP1-Nt (Fig. 3A, lanes 4 and 7).

The results described above strongly suggest that the highly conserved DNC motif of DBP factors is involved in DNA binding. If this is true, mutations in conserved positions within the DNC motif should interfere with its ability to interact with DNA. Therefore, two mutated versions of the N-terminal region of tobacco DBP1 were expressed in *E. coli* and purified as translational fusions to MBP (MBP-DNCmut1 and MBP-DNCmut2). The introduced amino acid changes in each mutated protein are shown in Figure 1A. They were all Ala substitutions affecting highly conserved positions within the DNC motif. Both MBP-DNCmut1 and MBP-DNCmut2 purified recombinant proteins were analyzed in EMSA for their ability to bind to the 3a4 probe. As shown in Figure 3B, mutations introduced in the DNC motif significantly reduce the capacity of the N-terminal region of DBP1 to interact with its target DNA. These results indicate that sequence integrity of the DNC motif is critical for interaction with DNA, and provide conclusive evidence for the direct implication of this motif in DNA binding.

Figure 1. (Continued.)

Ca, *Capsicum annuum*; Le, *Lycopersicon esculentum*; St, *Solanum tuberosum*; and Nt, tobacco. The GenBank accession number of each sequence is given. The consensus secondary structure prediction for the DBP1 N-terminal region provided by the Network Protein Sequence @nalysis server (NPS@; <http://npsa-pbil.ibcp.fr>) using default parameters (Combet et al., 2000) is indicated below the amino acid sequence; α -helices and β -strands are shown as cylinders and lines, respectively. Gray cylinder and dashed line denote uncertain prediction. The amino acid sequence of the DNC motif in the mutated DBP1 proteins is included at the bottom of the alignment. B, Phylogenetic trees generated using Phylip programs (<http://bioweb.pasteur.fr/seqanal/phylogeny/phylyip-uk.html>). Full-length amino acid sequences were aligned using T-coffee (<http://igs-server.cnrs-mrs.fr/Tcoffee/>; Notredame et al., 2000). Left, Phylogenetic tree provided by Quicktree, a program for the rapid reconstruction of phylogenies by the Neighbor-Joining method (Howe et al., 2002). Right, Phylogenetic tree obtained with Fitch, a program that uses the Fitch-Margoliash and some related least-squares criteria. C, Unrooted phylogenetic tree generated with Quicktree (Howe et al., 2002), including full-length sequences containing less conserved DNC motifs. D, Multiple sequence alignment of the DNC motifs of the DBP-related sequences used in C.

DISCUSSION

In the last few years, direct links between protein phosphorylation and function of DNA-bound transcription regulatory complexes have been reported. Accordingly, ERK5, a mouse mitogen-activated protein kinase, contributes a potent transcription activation domain after recruitment to the promoter of specific target genes by a MEF2 family transcription factor and functions as a coactivator in the transcriptional induction of genes involved in immature T cell apoptosis (Kasler et al., 2000). Similarly, Hog1, another mitogen-activated protein kinase, has been found to be part of transcription activating complexes and essential for transcriptional induction of some stress-responsive genes in yeast. Hog1 is recruited by the transcription

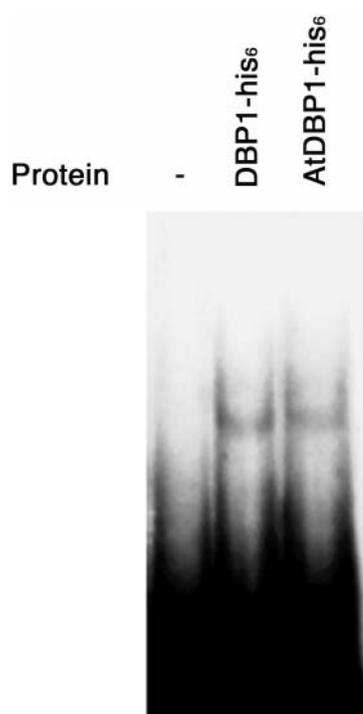


Figure 2. Assay of in vitro DNA binding by full-length DBP1 and AtDBP1. Recombinant full-length DBP proteins were expressed in *E. coli* bearing a C-terminal hexa-His tag, purified by nickel affinity chromatography according to the manufacturer's instructions (Qiagen, Valencia, CA), and assayed for their ability to bind to the double-stranded 3a4 oligonucleotide probe as described (Carrasco et al., 2003). Briefly, complementary oligonucleotides leaving 5' protruding ends were annealed and radioactively labeled using the Klenow fragment of *E. coli* DNA polymerase I. The sequences of the oligonucleotides used to build the double-stranded 3a4 probe are the following, with the nucleotides modified in the mutated 3a4mut probe shown in bold: 3a4-plus, 5'-TCGACTAGGCGGCT**TAAT**ATTTGCCT-TTGCTCCCTC-3'; and 3a4-minus, 5'-TCGAGAGGGAGACA**AAAG**-CAAATATTAGCCGCCTAG-3'. Purified proteins (1 μ M) were incubated for 10 min at room temperature with the radiolabeled DNA probe (0.3 nM) in binding buffer [20 mM HEPES-KOH, pH 7.6, 4 mM KCl, 0.1 mM EDTA, 1 mM dithiothreitol, 10% (v/v) glycerol], prior to separation on a native 6% (w/v) polyacrylamide gel in 0.5 \times Tris-borate/EDTA buffer.

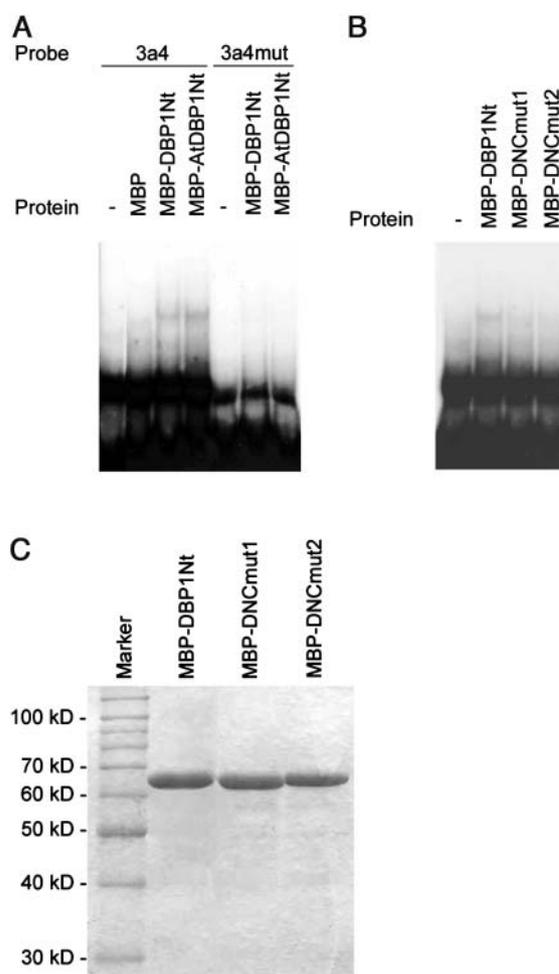


Figure 3. A, DNA binding in vitro by the N-terminal regions of DBP1 and AtDBP1. Recombinant MBP fusion proteins were expressed in *E. coli* and purified by amylose affinity chromatography, according to the manufacturer's instructions (New England Biolabs, Beverly, MA). Electrophoretic mobility shift assays were performed as described above. The sequences of the oligonucleotides used to generate the double-stranded mutated probe 3a4mut are the following, with changed nucleotides shown in bold: 3a4mut-plus, 5'-TCGACTA-GGCGGC**AGGAG**TTTGGGAGAGTCTCCCTC-3'; and 3a4mut-minus, 5'-TCGAGAGGGAGACT**CTCC**AAACTCCTGCCGCCTAG-3'. Both 3a4 and 3a4mut probes were radioactively labeled as described above. Labeling was quantified by liquid scintillation and found to be almost identical in both cases. The same amount of each labeled probe was used in EMSA. B, Effect of mutations in the DNC motif on DNA binding by the N-terminal region of DBP1. Mutations in MBP-DNCmut1 and MBP-DNCmut2 were introduced using the megaprimer method (Datta, 1995). Mutant proteins were expressed in *E. coli* and purified as described above. Purified recombinant proteins were incubated at a final concentration of 1 μ M with a radiolabeled 3a4 DNA probe (0.3 nM) for 10 min at room temperature in binding buffer [20 mM HEPES-KOH, pH 7.6, 4 mM KCl, 0.1 mM EDTA, 1 mM dithiothreitol, 10% (v/v) glycerol], prior to electrophoresis on a native 6% (w/v) polyacrylamide gel in 0.5 \times Tris-borate/EDTA buffer. C, SDS-PAGE of purified recombinant proteins used in EMSA.

factor Hot1 to the promoter of these stress-inducible genes (Alepuz et al., 2001), where, in turn, it mediates recruitment of a histone deacetylase activity and the RNA polymerase II holoenzyme (Alepuz et al., 2003; de Nadal et al., 2004). Another example is eyes absent, a nuclear protein Tyr phosphatase implicated in mammalian organogenesis, which acts as a coactivator through an interaction with homeodomain-containing proteins (Li et al., 2003). DBP1 represents one step forward in this direction since it is a protein phosphatase directly targeted to specific gene promoters by an intrinsic sequence-specific DNA-binding activity, without the need for a recruiting protein partner. We have now identified the DNC motif as involved in this specific DNA binding. Furthermore, we have described a large family of putative new transcription factors related to tobacco DBP1. These factors, apparently unique to plants, are all endowed with a PP2C catalytic domain and characterized by an N-terminal region with the herein identified and highly conserved DNC DNA-binding motif, also functional in the distantly related Arabidopsis DBP1 homolog AtDBP1, as the most salient signature of the family. Sequence conservation of this motif throughout the plant kingdom suggests that DBP factors might bind similar sequences, although a contribution to sequence specificity by additional residues cannot be excluded. The identification of target genes of DBP factors and interacting protein partners is an exciting challenge for the near future, and should help establish the mechanistic basis of the regulatory function of this novel class of transcription factors and the signaling pathways they are involved in.

ACKNOWLEDGMENT

We thank Dr. Peter Tompa (Institute of Enzymology, Budapest) for his invaluable advice and for critical reading of the manuscript.

Received November 3, 2004; November 26, 2004; December 2, 2004.

LITERATURE CITED

- Alepuz PM, de Nadal E, Zapater M, Ammerer G, Posas F (2003) Osmo-stress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. *EMBO J* **22**: 2433–2442
- Alepuz PM, Jovanovic A, Reiser V, Ammerer G (2001) Stress-induced MAP kinase Hog1 is part of transcription activation complexes. *Mol Cell* **7**: 767–777
- Altschul SF, Gish W (1996) Local alignment statistics. *Methods Enzymol* **266**: 460–480
- Carrasco JL, Ancillo G, Mayda E, Vera P (2003) A novel transcription factor involved in plant defense endowed with protein phosphatase activity. *EMBO J* **22**: 3376–3384
- Combet C, Blanchet C, Geourjon C, Deléage G (2000) NPS@: Network Protein Sequence Analysis. *Trends Biochem Sci* **25**: 147–150
- Datta AK (1995) Efficient amplification using “megaprimer” by asymmetric polymerase chain reaction. *Nucleic Acids Res* **23**: 4530–4531
- de Nadal E, Zapater M, Alepuz PM, Sumoy L, Mas G, Posas F (2004) The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmo-responsive genes. *Nature* **427**: 370–374
- Gaits F, Shiozaki K, Russell P (1997) Protein phosphatase 2C acts independently of stress-activated kinase cascade to regulate the stress response in fission yeast. *J Biol Chem* **272**: 17873–17879
- Gosti F, Beaudoin N, Serizet C, Webb AAR, Vartanian N, Giraudat J (1999) ABI1 protein phosphatase 2C is a negative regulator of abscisic acid signaling. *Plant Cell* **11**: 1897–1909
- Howe K, Bateman A, Durbin R (2002) QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics* **18**: 1546–1547
- Kasler HG, Victoria J, Duramad O, Winoto A (2000) ERK5 is a novel type of mitogen-activated protein kinase containing a transcriptional activation domain. *Mol Cell Biol* **20**: 8382–8389
- Kobor MS, Greenblatt J (2002) Regulation of transcription elongation by phosphorylation. *Biochim Biophys Acta* **1577**: 261–275
- Li X, Ohgi KA, Zhang J, Kronen A, Bush KT, Glass CK, Nigam SK, Aggarwal AK, Maas R, Rose DW, et al (2003) Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. *Nature* **426**: 247–254
- Luan S (2003) Protein phosphatases in plants. *Annu Rev Plant Physiol Plant Mol Biol* **54**: 63–92
- Maeda T, Wurgler-Murphy SM, Saito H (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**: 242–245
- Mayda E, Marqués C, Conejero V, Vera P (2000) Expression of a pathogen-induced gene can be mimicked by auxin insensitivity. *Mol Plant Microbe Interact* **13**: 23–31
- Meskiene I, Baudouin E, Schweighofer A, Liwosz A, Jonak C, Rodríguez PL, Jelinek H, Hirt H (2003) Stress-induced protein phosphatase 2C is a negative regulator of a mitogen-activated protein kinase. *J Biol Chem* **278**: 18945–18952
- Meskiene I, Bögre L, Glaser W, Balog J, Brandstötter M, Zwerger K, Ammerer G, Hirt H (1998) MP2C, a plant protein phosphatase 2C, functions as a negative regulator of mitogen-activated protein kinase pathways in yeast and plants. *Proc Natl Acad Sci USA* **95**: 1938–1943
- Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**: 205–217
- Rodríguez PL (1998) Protein phosphatase 2C (PP2C) function in higher plants. *Plant Mol Biol* **38**: 919–927
- Sheen J (1996) Ca²⁺-dependent protein kinases and stress signal transduction in plants. *Science* **274**: 1900–1902
- Smith RD, Walker JC (1996) Plant protein phosphatases. *Annu Rev Plant Physiol Plant Mol Biol* **47**: 101–125
- Takekawa M, Maeda T, Saito H (1998) Protein phosphatase 2C α inhibits the human stress-responsive p38 and JNK MAPK pathways. *EMBO J* **17**: 4744–4752
- Whitmarsh AJ, Davis RJ (2000) Regulation of transcription factor function by phosphorylation. *Cell Mol Life Sci* **57**: 1172–1183