A Novel Family of Plant DNA-Binding Proteins Containing both HMG-Box and AT-Rich Interaction Domains[†]

Frederik T. Hansen,[‡] Claus K. Madsen,[‡] Anne Mette Nordland,[‡] Marion Grasser,[‡] Thomas Merkle,[§] and Klaus D. Grasser*[,‡]

Department of Life Sciences, Aalborg University, Sohngaardsholmsvej 49, DK-9000 Aalborg, Denmark, and Genome Research, Faculty of Biology, University of Bielefeld, Universitätsstrasse 25, D-33594 Bielefeld, Germany

Received September 17, 2008; Revised Manuscript Received October 21, 2008

ABSTRACT: The A/T-rich interaction domain (ARID) and the HMG-box domain represent DNA-interaction modules that are found in sequence-specific as well as nonsequence-specific DNA-binding proteins. Both domains are found in a variety of DNA-interacting proteins in a wide range of eukaryotic organisms. Proteins that contain both an ARID and an HMG-box domain, here termed ARID-HMG proteins, appear to be specific for plants. This protein family is conserved in higher plants (both mono- and dicot plants) as well as lower plants such as the moss *Physcomitrella*. Since ARID-HMG proteins have not been studied experimentally, we have examined here two family members from *Arabidopsis*. The genes encoding ARID-HMG1 and ARID-HMG2 are widely expressed in *Arabidopsis* but at different levels. Subcellular localization experiments studying ARID-HMG1 and ARID-HMG2 fused to GFP by fluorescence microscopy show that both proteins localize primarily to cell nuclei. Analyses of the DNA-binding properties using electrophoretic mobility shift assays revealed that mediated by the HMG-box domain, ARID-HMG1 binds structure specifically to DNA minicircles. Mediated by the ARID, the protein binds preferentially to A/T-rich DNA, when compared with G/C-rich DNA. Therefore, both DNA-binding domains contribute to the DNA interactions of ARID-HMG1. Accordingly, the protein combines DNA-binding properties characteristic of ARID and HMG-box proteins.

The HMG box is a protein domain that initially was found to be characteristic of the chromatin-associated high mobility group (HMG)[1] proteins of the HMGB family (*1, 2*). Meanwhile, the HMG-box domain has been identified in a wide variety of eukaryotic DNA-binding proteins including transcription factors and subunits of chromatin remodelling complexes (*3*−*5*). The HMG-box domain of ∼75 amino acid residues consists mainly of three α-helices that form an L-shaped molecule. Mediated by the concave face of this domain HMG-box proteins interact primarily with the minor groove of the DNA helix. There are HMG-box domains that interact with DNA sequence specifically (e.g., in transcription factors) or nonsequence specifically (e.g., in HMGB proteins). Both types of HMG-box domains bind with high affinity certain DNA structures such as DNA minicircles, four-way junctions, and cis-platin modified DNA (*6*−*8*). HMG-box proteins typically have DNA-bending activity and as architectural proteins are involved in the formation of higher order nucleoprotein complexes regulating DNA-dependent processes including transcription, recombination, and DNA repair (*6*−*11*).

The AT-rich interaction domain (ARID) is a DNA-binding module that was first identified in the mouse B-cell transcription factor Bright and the *Drosophila* Dead ringer protein (*12, 13*). More recently, ARIDs have been found in a variety of eukaryotic transcription factors that regulate cell proliferation, differentiation, and development. The ARID consensus sequence spans ∼100 amino acid residues that are organized in a modified helix−turn−helix fold (*14*−*16*). ARIDs bind DNA through a novel mechanism involving major groove immobilization of a large loop that connects the helices of the helix−turn−helix motif and through concomitant structural rearrangements adding stabilizing contacts from a β-hairpin (*17*−*20*). The first characterized ARID proteins were found to bind preferentially to A/T-rich sequences, prompting the name of the domain. Analysis of protein/DNA interactions of other ARIDs by various methods revealed that ARIDs of different proteins bind DNA nonsequence specifically (*18, 19, 21, 22*).

In plants, HMG-box proteins occur in four families: HMGB proteins, SSRP1, proteins containing three HMG-box domains, and proteins containing an HMG-box domain and an ARID (*5*). While the first two families have been studied, there are no experimental data on members of the latter two families. Thus, the relatively abundant chromosomal HMGB proteins are widely expressed in plants (*23*−*26*). They bind DNA nonsequence specifically but recognize DNA structures and structurally flexible DNA sites (*27*−*31*).

[†] This research was supported by a grant from the Danish Research Council to K.D.G.

* To whom correspondence should be addressed. Tel: +45 9940 9126. Fax: +45 9814 1808. E-mail: kdg@bio.aau.dk.

[‡] Aalborg University.

[§] University of Bielefeld.

[1] Abbreviations: ARID, A/T-rich interaction domain; HMG, high mobility group; SSRP1, structure-specific recognition protein 1; EMSA, electrophoretic mobility shift assay; GFP, green fluorescent protein; NLS, nuclear localization signal; GST, glutathione *S*-transferase; CD, circular dichroism.

10.1021/bi801772k CCC: $40.75
Published on Web 11/17/2008

The HMG-box protein SSRP1 is a subunit of the FACT complex with histone-chaperone activity assisting transcription. Maize SSRP1 interacts with DNA nonsequence specifically and binds DNA structure selectively (*32, 33*).

Plant genomes encode various ARID proteins, and in most cases the ARID occurs in combination with various other protein domains including Myb, PHD, and HMG-box domains (*34, 35*). So far only a single plant ARID protein has been studied experimentally. SIP1 from *Lotus japonicus* was identified very recently, and it contains an ARID (but no HMG-box domain) binding AT-rich elements of the *NIN* gene promoter. It has been suggested that SIP1 plays a role in nodule formation related to the rhizobium-legume plant symbiosis (*35*). The proteins containing both an ARID and an HMG-box domain appear to be specific for plants (*5, 34*), and they have not been analyzed experimentally. Therefore, we have biochemically examined here members of this protein family from *Arabidopsis thaliana*.

EXPERIMENTAL PROCEDURES

Plasmid Constructions. The sequence encoding ARID-HMG1 was amplified by PCR using DeepVent DNA polymerase (NEB), primers P1 and P2 (see Supporting Information Table S1 for primer sequences), and an *Arabidopsis* cDNA library as template. The PCR fragment was digested with *Bam*HI/*Hin*dIII at terminal restriction enzyme recognition sites (introduced through the primers) and cloned into the *Bam*HI/*Hin*dIII-digested plasmid pQE9 (Qiagen), giving pQE9-ARID-HMG1. Similarly, the sequence encoding ARID-HMG2 was amplified by PCR using primers P3 and P4 and cloned into *Bam*HI/*Hin*dIII-digested plasmid pQE9, giving pQE9-ARID-HMG2. The sequence encoding the HMG-box domain of ARID-HMG1 was also amplified by PCR using primers P5 and P6 and cloned into *Bam*HI/*Hin*dIII-digested plasmid pQE9 (Qiagen), giving pQE9-ARID-HMG1box. Plasmid pQE9-ARID-HMG1 was consecutively treated with *Hin*dIII, Klenow enzyme, and *Bam*HI, and the sequence encoding ARID-HMG1 was cloned into *Bam*HI/*Sma*I-digested pGEX-2T, giving pGEX-2T-ARID-HMG1. Plasmids pQE9-ARID-HMG1 and pQE9-ARID-HMG2 were consecutively treated with *Hin*dIII, Klenow enzyme, and *Eco*RI, and the sequences encoding ARID-HMG1 and ARID-HMG2 were cloned into *Eco*RI/*Sma*I-digested p5′GFP (*36*), giving p5′GFP-ARID-HMG1 and p5′GFP-ARID-HMG2. All plasmid constructions were checked by DNA sequencing.

Protein Production and Purification. Full-length ARID-HMG1 fused to glutathione *S*-transferase (GST) was expressed in *Escherichia coli* using the plasmid pGEX-2T-ARID-HMG1 and purified by affinity chromatography using glutathione−Sepharose 4B (Amersham) as previously described (*37*), except that the protein was eluted with 10 mM reduced glutathione in 50 mM Tris-HCl, pH 8.0. 6 × His-tagged HMG-box domains of maize HMGB1 and *Arabidopsis* ARID-HMG1 were expressed in *E. coli* using the plasmids pQE9 cm-HMGa(G35-Y109) (*29*) and pQE9-ARID-HMG1box and purified by three-step column chromatography (metal-chelate, ion exchange, hydrophobic interaction) as previously described (*29*). Purified recombinant proteins were checked for purity by SDS−PAGE and MALDI-TOF mass spectrometry.

Circular Dichroism (CD). CD measurements using 15 μM ARID-HMG1box were performed using a Jasco J-810 instrument as described previously (*38*).

rtPCR. A. thaliana ecotype Col-0 was grown under long-day conditions as previously described (*39*). Total RNA was extracted from ∼100 mg of frozen plant tissue using the TRIzol method (Invitrogen) and treated with DNase (MBI Fermentas). Reverse transcription was performed using 1 μg of RNA and Revert Aid H minus M-MuLV reverse transcriptase (MBI Fermentas). The obtained cDNA was used for PCR analyses using *Taq* DNA polymerase (Amplicon) and oligonucleotide primers P7−P12 specific for *ARID-HMG1*, *ARID-HMG2*, and 16S RNA (see Supporting Information Table S1 for primer sequences).

Transient Protoplast Transformation Assays with GFP Fusion Constructs. Protoplasts were prepared from dark-grown tobacco BY-2 cells and transiently transformed with plasmids encoding GFP fusions (and control constructs) by PEG-mediated transformation as described previously (*36*). Excitation of GFP was performed with a standard UV light source and fluorescein isothiocyanate (FITC) filters. For confocal laser scanning microscopy, samples were directly examined under oil with a 63× objective and a DM RE TCS4D microscope (Leica) equipped with an argon−krypton laser (excitation 488 nm, beam splitter 500 nm, detection 500−560 nm for GFP using Leica Scanware). Analysis of the localization of the GFP fusion proteins was performed in three independent experiments, representing approximately 60−80 transformed protoplasts.

Electrophoretic Mobility Shift Assays (EMSAs). Structure-specific DNA binding of recombinant proteins was examined by EMSAs using a mixture of linear and circularized 78-bp *Kpn*I/*Xba*I DNA fragment originating from pBluescript (Stratagene). The circularized form of the fragment was produced by intramolecular ligation of the ^{32}P-labeled fragment (after creating blunt ends with the Klenow fragment of DNA polymerase I) using T4 DNA ligase in the presence of recombinant maize HMGB1(M1-K123) as described previously (*29*). Proteins were incubated with the DNA in a final volume of 20 μL of binding buffer (10 mM Tris/HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, 0.05% bromophenol blue, 0.05% xylene cyanol) for 15 min. To test protein binding to A/T- and G/C-rich DNA, 63-bp DNA fragments of the *Tub1* promoter region were amplified by PCR from genomic *Physcomitrella* DNA and ^{32}P-labeled as previously described (*40*). Protein binding reactions with the A/T- and G/C-rich fragments were performed as described above. All binding reactions were analyzed in 5% polyacrylamide gels in 1 × TBE, and the DNA was visualized by phosphorimaging using a Typhoon 8600 instrument (Amersham Biosciences).

RESULTS AND DISCUSSION

Identification of Proteins Containing an ARID and an HMG-Box Domain. To get an overview about the plant proteins that contain both an ARID and an HMG-box domain, we searched the *Arabidopsis* database (http://www.arabidopsis.org/) and the Plant Chromatin Database (http://www.chromdb.org/). Our survey yielded a number of amino acid sequences containing both types of (putative) DNA-binding domains (Table 1). Thus, the genome of the

Table 1: Plant ARID-HMG Proteins

protein[a]	length (aa)	mass (Da)	locus	species
ARID-HMG1	338	38774	At1g76110	*Arabidopsis*
ARID-HMG2	319	36296	At3g13350	*Arabidopsis*
HMGB11	337	38049	At1g55650	*Arabidopsis*
HMGB15	448	50004	At1g04880	*Arabidopsis*
HMGB903	329	37753		poplar
HMGB904	316	36037		poplar
HMGB908	389	43836		poplar
HMGB909	463	51452		poplar
HMGB702	467	51629	Os02g27060	rice
HMGB709	306	35019	Os09g37250	rice
HMGB115	442	49704		maize
HMGB1511	753	81975		*Physcomitrella*

[a] Except for the ARID-HMG1 and ARID-HMG2 proteins studied here, the nomenclature of the Plant Chromatin Database (http://www.chromdb.org/) was adopted.

dicot plant *Arabidopsis* codes for four sequences, while the genome of the monocot plant rice encodes two proteins of this type. In line with earlier observations (*5, 34*), proteins containing both an ARID and an HMG-box domain, termed here ARID-HMG, appear to be specific for plants, because we did not find related sequences in other organisms, when searching GenBank. In addition to higher plants, we identified a related protein that is encoded in the genome of the moss *Physcomitrella patens* (Table 1). Therefore, ARID-HMG proteins occur in higher and lower plant species. The overall structure of ARID-HMG proteins (~35−52 kDa, except the larger *Physcomitrella* protein) is relatively conserved among the proteins of different species. They contain an ARID in the N-terminal part, while the HMG box resides within the C-terminal part of the proteins (cf. Figure 1A). The amino acid sequences of the ARID-HMG proteins from *Arabidopsis*, poplar, rice, maize, and *Physcomitrella* were used for construction of a neighbor-joining tree that showed that the proteins group into subfamilies (Figure 1B). Thus, Zm-HMGB115 and Os-HMGB702 form for instance a monocot-specific branch, while the branches around ARID-HMGB2, At-HMGB11, and Pt-HMGB904 or At-HMGB15, Pt-HMGB908, and Pt-HMGB909 are dicot-specific. The group around ARID-HMG1 appears to occur in both mono- and dicot plants. A multiple sequence alignment of the amino acid sequences revealed that the ARIDs (45%−96% amino acid sequence identity) and HMG-box domains (34%−96% amino acid sequence identity) are well conserved among the ARID-HMG proteins (Figure 2). Outside the two domains, only a short region of ~25 residues situated between the two domains is well conserved among the different ARID-HMG proteins. Within the ARID five residues found to be characteristic of animal ARIDs (*14, 15*) are well conserved among the ARIDs of plant ARID-HMG proteins (Figure 2A). Similarly, the sequence around the primary DNA-intercalating residue and the invariant tryptophan residue of the HMG-box domains (*8, 11*) is conserved in the HMG-box domains of ARID-HMG proteins (Figure 2B), while the sequence around the position of the secondary DNA-intercalating residue is less conserved.

The ARID-HMG Proteins Are Widely Expressed in Arabidopsis and Localize Primarily to the Nucleus. Of the four *Arabidopsis* genes encoding ARID-HMG proteins, according to the *Arabidopsis* database, currently there is no evidence (EST, cDNA) for expression for the gene At1g55650, while At1g04880 appears to be expressed at very low levels.

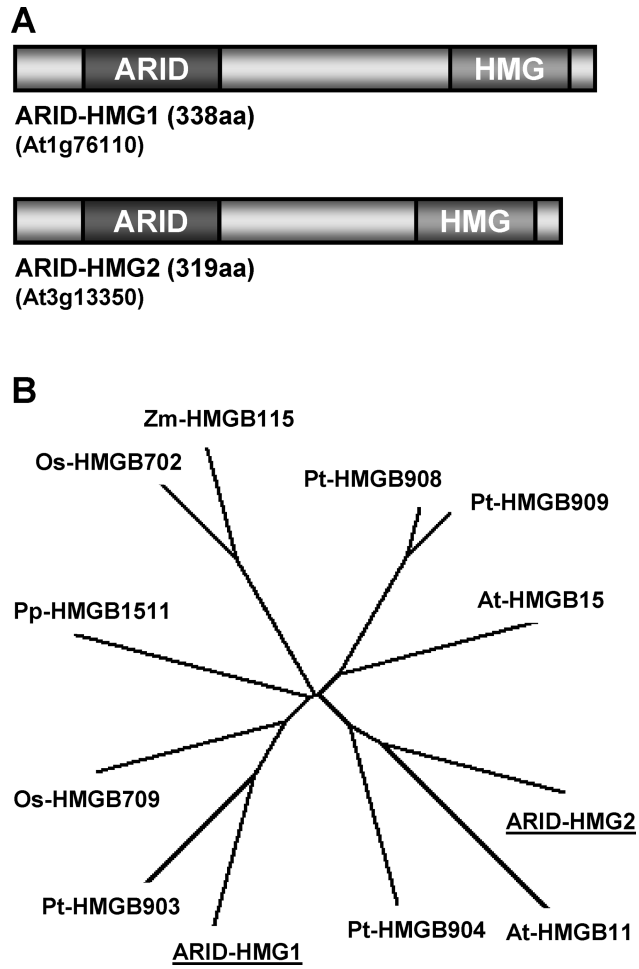

FIGURE 1: Overall structure of ARID-HMG proteins and amino acid sequence similarity between different family members. (A) Schematic representation of the overall structure of *Arabidopsis* ARID-HMG1 (encoded by locus At1g76110) and ARID-HMG2 (encoded by locus At3g13350). The proteins of this plant-specific family contain two DNA-binding domains, an N-terminal ARID, and a C-terminal HMG-box domain. (B) Amino acid sequence similarity of ARID-HMG proteins from different sources. The protein sequences (cf. Table 1) were aligned by multiple sequence alignment (CLUSTALW) and used to construct a neighbor-joining tree (http://align.genome.jp/sit-bin/clustalw), representing the degree of sequence similarity between the proteins.

Therefore, we have chosen the 38.8 kDa ARID-HMG1 (At1g76110) and the 36.3 kDa ARID-HMG2 (At3g13350) (cf. Figure 1A) for experimental analysis. To examine the expression of the two genes encoding these proteins in *Arabidopsis*, RNA was isolated from different tissues, and after reverse transcription it was used for estimating relative transcript levels by PCR. Using *ARID-HMG1*-specific primers the amplification product was obtained from all RNA samples, but with very different efficiency (Figure 3). Thus, the transcript level is clearly higher in leaves, flower buds, and seedlings compared to roots, cultured cells, and seeds. By contrast, the *ARID-HMG2*-specific PCR product is detected similarly with all RNA samples, indicating that this gene is expressed at comparable levels in all tissues, similar to the 16S RNA control. The rtPCR experiment shows that the two *ARID-HMG* genes are widely expressed in *Arabidopsis* and that the expression levels of *ARID-HMG1* differ between different tissues. In animals, there are ARID-containing proteins that display a broad expression pattern, while other family members are expressed in a tissue-specific

A

```
                                    ▼          ▼          ▼          ▼                    ▼
ARID-HMG1    FWDTLRRFHSIMSTKFMIPVIGGKELDLHVLYVEVTRRGGYEKVVVEKKWREVGGVFRFSATTTSASFVLRKHYLNLLFHYE-125
Pt-HMGB903   FWDTLRRFHFVMGTKFMIPVIGGKELDLQVLYVETTNRGGYDKVVAEKKWREVGSVFCFSATTTSASFVLKKHYFSLLYHYE-116
Os-HMGB709   FMDTLRRFHSLMGTKFMIPVIGGKEMDLHALYVEVTSRGGLAKVMEERKWREVMARFSFPATTTSASYVLRRYYLSLLHHYE-105
Os-HMGB702   FRRALEGLHAQMGTKLKVPIIGGKDLDLHQLFKEVTSRGGIDKVKSDNRWREVTASFIFPATATNASFMLKKYYMSLLYHFE-162
Zm-HMGB115   FRAALEGLHAHMNTRLKVPIIGGKDLDLHQLYKEVTSRGGIDKLKAENRWREVTASFIFPATATNASFMLKKYYMSLLYHFE-146
Pp-HMGB1511  FADALNKFHTILGTRITIPKLGGNDLDLHLLYVEVTSRGGLEQVIKDRKWKEVINCFKFPENTTSASYILRKYYVGLLYYFE-105
Pt-HMGB908   FMETLEKFHAAMGTKFMIPIIAGKELNLHRLFVEVTSRGGIEKIIREKRWKEVTSVFNFPSTATNASFVLRKYYGSLLQHYE-122
Pt-HMGB909   FMETLEKFHAAMGTKFMIPIIGGKELNLHRLFVEVTSRGGIEKIIRERRWKEVTGVFNFPSTATNASFVLRKYYGSLLQHYE-122
At-HMGB15    FMTSLERLHSLLGTKFMVPIIGGRDLDLHKLFVEVTSRGGINKILNERRWKEVTATFVFPPTATNASYVLRKYYFSLLNNYE-116
ARID-HMG2    FWEKLRAFLGLTSKTLKVPTVGGNTLDLHRLFIEVTSRGGIERVVKDRKWKEVIGAFSFPTTITSASFVLRKYYLKFLFQLE-127
At-HMGB11    FWEMLRDFHESSDKKFKIPIVGGKSLDLHRLFNEVTSRGGLEKVIKDRRCKEVIDAFNFKTTITNSAFVLRKSYLKMLFEFE-121
Pt-HMGB904   FWEKLKAFHQSFGTKFMVPTVGGKALDLHHLFVEVTSRGGIEKVITDRKWKEVITAFNFPTTITSASFVLRKHYLSLLYHFE-129
```

B

```
                                    ▼              (▼)            ▼
ARID-HMG1    EDPNYP-KPNRSGYNFFFAEKHCKLKSLYPNKEREFTKLIGESWSNLSTEERMVYQDIGLKDKERYQRELNEY-321
Pt-HMGB903   GDPSYP-KPNRSGYNFFFAEKHYKLKSLYPNREREFTKMIGQSWSSLSAEERMVYQNIGLKDKERYKRELKEY-306
Os-HMGB709   RDPAQP-RPNRSAYNFFFKEKHPELKATHPHREREYSRMIGDAWNRLAADDKMVYYRHSAEDKERYKREMQEY-293
Os-HMGB702   TDPNHP-KPNRSGYNFFFQDQHRKLKPEYPGQDRLISKMIGERWNNLGPEDKAVYQEKGVEDKARYQRQLALY-351
Zm-HMGB115   TDPRHP-KPNRSGYNFFFQDQHRMLKPQCPGQDRLISKMIGERWNNLSPEDKAVYQERGVQDKERYRTQLAAY-337
Pp-HMGB1511  KDPNAP-RPHRTGYNFFFAEQRARLKALHPEKDKELSRMIGDAWNNLTEEEKTLYQDRGVQDKERYKTELREY-307
Pt-HMGB908   RDPAHP-KPNRSGYNFFFSEQHARLKPLYPGKDREISRMIGELWNKINDSQKAVYQEKALRDKERYKIEMEGY-318
Pt-HMGB909   RDPAHP-KPNRSGYNFFFAEQHARLKPLYPGKDREISRMIGELWNKINDSQKAVYQDKALKDKERYKIEMEGY-331
At-HMGB15    RDPDHP-KPNRSGYNFFFAEQHARLKPLHPGKDRDISRMIGELWNKLNEDEKLIYQGKAMEDKERYRTEMEDY-329
ARID-HMG2    VDTQKP-KCHRSGYNFFFAEQYARLKPEYHGQERSITKKIGHMWSNLTESEKQVYQDKGVKDVERYRIEMLEY-304
At-HMGB11    GDSHKPPKRQRTGYNFFVAEQSVRIKAENAGQKVSSPKNFGNMWTNLSESDRKVYYEKSREDGKRYKMEILQY-281
Pt-HMGB904   RDPSQP-KSNRSGYNFFFTEHYAQLKPLHHGQEKVISKKIGLLWNNLTEAEKQVYQEKGLRDKEKYRNEMLEY-310
```

FIGURE 2: Alignment of ARIDs and HMG-box domains of plant ARID-HMG proteins. (A) Multiple amino acid sequence alignment of the ARIDs of ARID-HMG proteins from *Arabidopsis thaliana* (At), *Populus trichocarpa* (Pt), *Oryza sativa* (Os), *Zea mays* (Zm), and *Physcomitrella patens* (Pp). The alignment includes the two proteins studied here, ARID-HMG1 and ARID-HMG2. For the other proteins of the alignment the nomenclature of the Plant Chromatin Database (http://www.chromdb.org/) has been adopted. The five critical residues identified in animal ARIDs (*15*) are indicated by ▼. The amino acid sequence position of the C-terminal ARID residue within the protein sequence is given. (B) Amino acid sequence alignment of the HMG-box domains of the same proteins as in (A). The invariant tryptophan residue and the position of the primary DNA intercalating residue (*8*) conserved among the sequences are indicated by ▼, while the less conserved position around the secondary DNA intercalating residue is indicated by (▼).

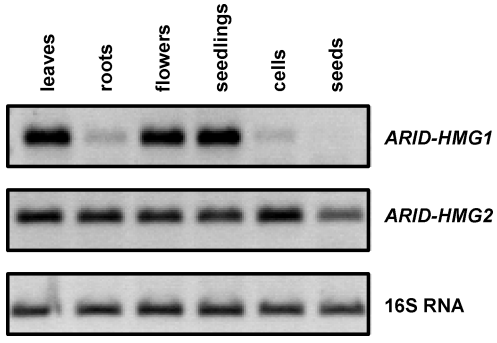

FIGURE 3: *ARID-HMG1* and *ARID-HMG2* are widely expressed in *Arabidopsis*. rtPCR analysis of total RNA, isolated from leaves, roots, and flower buds of 4-week-old plants, from green parts of 10-day-old seedlings, from seeds harvested 15 days after flowering (daf), and from suspension cultured cells. PCR was performed with primers specific for *ARID-HMG1* resulting in a 120-bp band (top panel), for *ARID-HMG2* resulting in a 120-bp band (middle panel), and for 16S RNA resulting in a 130-bp band (bottom panel).

manner (*15*). Similarly, some animal HMG-box proteins are broadly expressed, while others have a more restricted expression pattern (*3, 5*). In plants, HMG-box proteins of the HMGB family (*23, 24*) and of the SSRP1 family (*33, 41*) are widely expressed.

According to predictions using the software pSORT (http://psort.ims.u-tokyo.ac.jp/form.html) ARID-HMG1 and ARID-HMG2 are nuclear proteins, but predictions of the subcellular localization of proteins currently are not very accurate (*42*). Therefore, we have examined the subcellular localization of the two *Arabidopsis* proteins. We have constructed plasmids suitable for expression of N-terminal GFP fusion proteins in plant protoplasts. In transient transformation assays performed with tobacco BY-2 suspension cell protoplasts, the expression of the GFP fusion proteins was driven by the CaMV 35S promoter. Transformed protoplasts were analyzed by confocal laser scanning microscopy (Figure 4). We have used constructs that served as controls in this experiment: GFP that is found in both the cytosol and the nucleus (Figure 4G) and GFP-NLS-CHS-NES(mutated) that accumulates in the nucleus due to a strong functional basic nuclear localization signal (NLS) (Figure 4E,F) (*36*). GFP-ARID-HMG1 and GFP-ARID-HMG2 show a prominent localization in the nucleus (Figure 4A−D). However, they did not completely accumulate in the nuclear compartment like the control GFP-NLS-CHS-NES(mutated) or GFP-HMGB1 (Figure 4H) (*39*). With both GFP-ARID-HMG1 and GFP-ARID-HMG2, GFP fluorescence was also observed in the cytosol that is much weaker than that in the nucleus but clearly detectable (Figure 4C,D). Part of this finding could be due to the overexpression of the GFP fusion proteins in the protoplasts. On the other hand, control GFP fusion proteins containing strong NLSs did not show this effect in the same system (Figure 4E,H). Alternatively, the weak cytosolic signal of GFP-ARID-HMG1 and GFP-ARID-HMG2 may indicate the presence of relatively weak NLSs in these proteins and/or their association with as yet unknown partner(s) in the cytosol. The occurrence of speckles in the cytosol (Figure 4A−D) may favor the latter view.

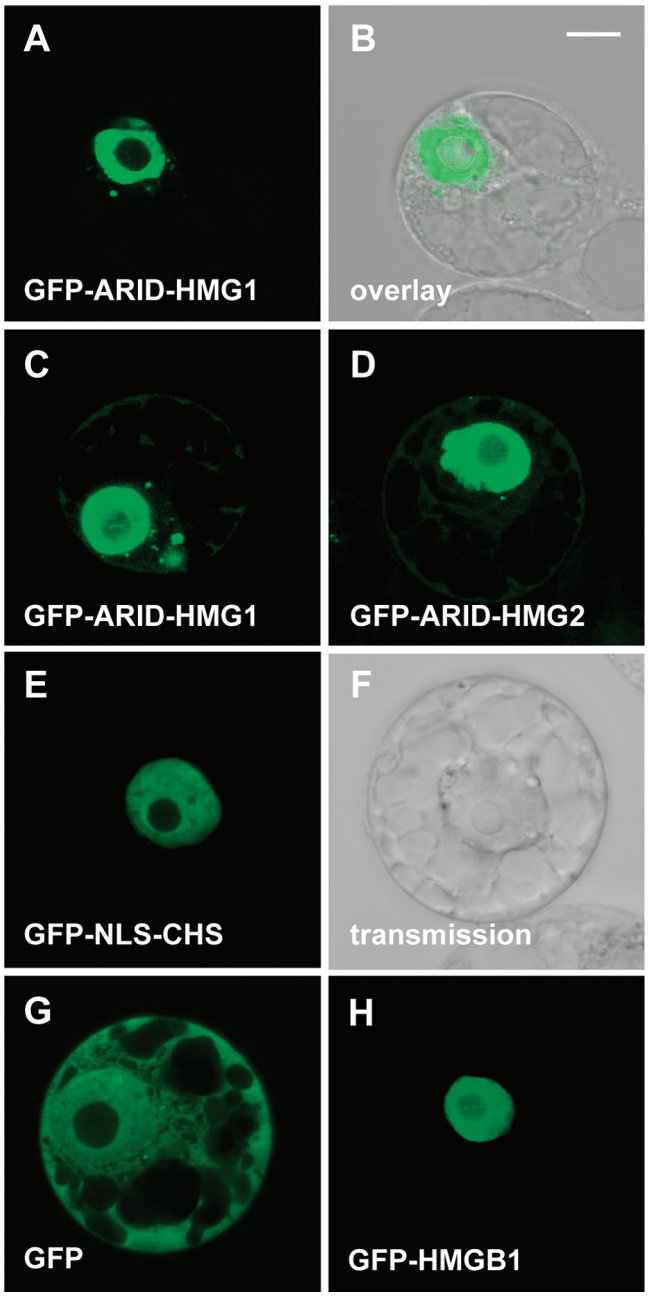

FIGURE 4: GFP-ARID-HMG1 and GFP-ARID-HMG2 show a predominant nuclear localization. Protoplasts prepared from BY-2 tobacco cell suspension cultures were transformed with plasmids driving the expression of the indicated GFP fusion proteins. (A−C) GFP-ARID-HMG1; (B) overlay of (A) with the corresponding transmission image; (D) GFP-ARID-HMG2; (E) GFP-NLS-CHS-NES(mutated) as a control that shows nuclear accumulation (*36*); (F) transmission image of the same protoplast shown in (E); (G) GFP alone that localizes to the cytosol and to the nucleus; (H) GFP-HMGB1 localizing to the nucleus (*39*). Confocal laser scanning images are shown. The size bar equals 10 μm.

Production of Recombinant ARID-HMG1. To examine the DNA-binding properties of ARID-HMG1, we set out to produce in *E. coli* the full-length protein and individual domains. Despite testing a variety of conditions (different *E. coli* strains, growth conditions, etc.) we were unable to produce soluble 6 × His-tagged ARID-HMG1. However, we succeeded in producing ARID-HMG1 fused to gluthatione *S*-transferase (GST), albeit the expression level was relatively low (compared to other proteins we produce in our laboratory). Despite testing various conditions, we were unable to

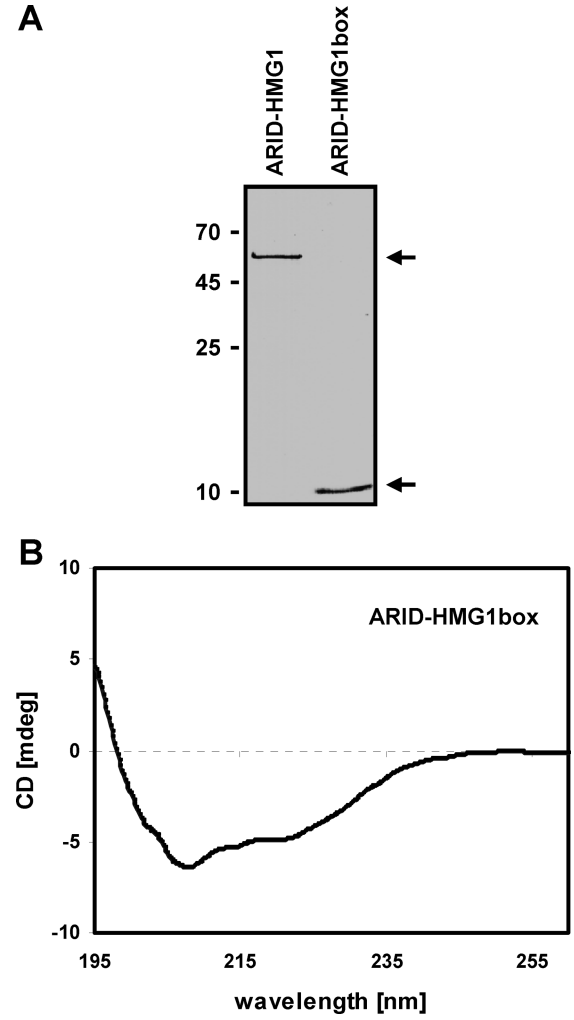

FIGURE 5: SDS−PAGE and CD spectroscopy of recombinant ARID-HMG1 proteins. (A) Purified recombinant ARID-HMG1 (fused to GST) and 6 × His-tagged ARID-HMG1box were analyzed by SDS−PAGE using a 18% polyacrylamide gel and Coomassie staining. (B) CD spectra were recorded for the ARID-HMG1box in the wavelength range of 195−260 nm, resulting in spectra indicating α-helical structure.

produce the individual ARID in *E. coli*, whereas the individual HMG-box domain could be expressed efficiently as 6 × His-tagged protein. The GST-ARID-HMG1 fusion was purified by glutathione−Sepharose affinity chromatography, while the HMG-box domain was purified by three-step column chromatography. SDS−PAGE analysis of the purified proteins revealed that both GST-ARID-HMG1and the HMG-box domain are essentially pure (Figure 5A). Using CD spectroscopy, we have examined the HMG-box domain of ARID-HMG1. CD wavelength scans were recorded in the range of 195−260 nm (Figure 5B). The spectrum displays a shape typical of α-helical proteins, having the characteristic local minima around 208 and 222 nm. As seen with HMG-box domains of plant HMGB proteins (*29, 43*), the CD measurements indicate that the HMG-box domain of ARID-HMG1 has an α-helical structure consistent with the solved structures of a number of HMG-box domains (*8*).

DNA Interactions of ARID-HMG1. It is well documented that plant HMGB proteins and SSRP1, mediated by their HMG-box domains, bind DNA nonsequence specifically, but they interact with high affinity with certain DNA structures, including DNA minicircles and four-way junc-

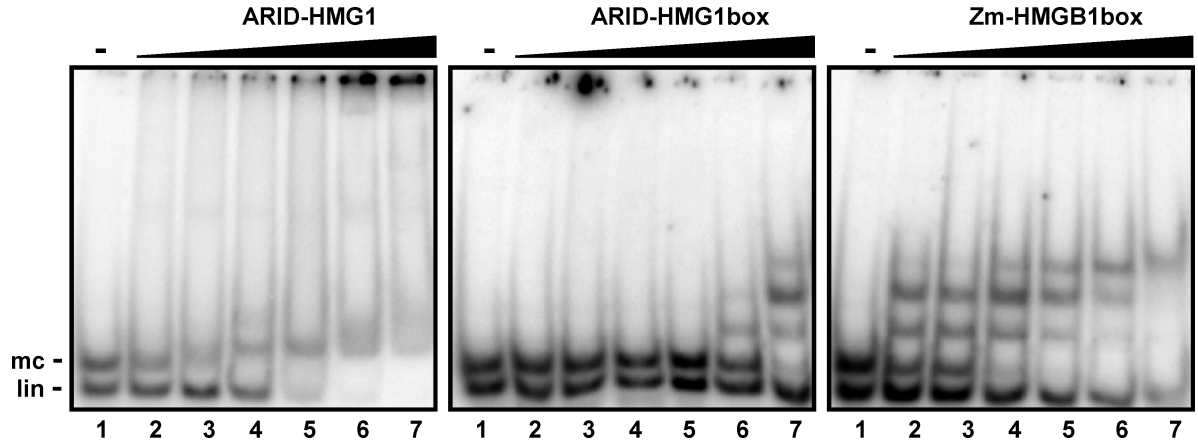

FIGURE 6: ARID-HMG1 binds structure specifically to DNA minicircles. A mixture of linear and circularized ^{32}P-labeled 78-bp DNA fragment was incubated in the absence (lanes 1) or in the presence of increasing concentrations of ARID-HMG1 (left panel) and of ARID-HMG1box (middle panel) (25, 50, 100, 200, 400, 800 nM; lanes 2−7, respectively, for ARID-HMG1 and ARID-HMG1box) and of Zm-HMGB1box (right panel) (10, 25, 50, 100, 200, 400 nM; lanes 2−7, respectively). The DNA-binding reactions were analyzed by electrophoresis in native polyacrylamide gels, and the dried gels were scanned using a phosphorimager. The electrophoretic migration position of unbound linear (lin) and minicircle (mc) DNA are indicated.

tions (*27, 29−31, 33*). Purified full-length ARID-HMG1 and the individual HMG-box domain were examined for structure-specific DNA binding using electrophoretic mobility shift assays (EMSAs). For comparison we have used the HMG-box domain of the well-characterized maize HMGB1 protein (Zm-HMGB1box) (*29, 37*). Increasing amounts of the three proteins were incubated with a mixture of linear and circularized 78-bp DNA fragment, and DNA-binding was examined by native polyacrylamide gel electrophoresis of the samples. With increasing protein input protein/DNA complexes with reduced electrophoretic mobility appear, demonstrating that the proteins bind to DNA (Figure 6). The three proteins bind structure specifically to the minicircle DNA, since compared to the minicircle DNA, binding to linear DNA requires higher protein concentrations. This is seen very clearly with the reference protein Zm-HMGB1box, which from a concentration of ~10 nM binds the DNA minicircle, while binding to the linear fragment requires a protein concentration of ~400 nM (right panel, compare lanes 2 and 7). Compared to the maize HMGB1box, the affinity for DNA of the ARID-HMG1box is lower. Thus, ARID-HMG1-box binds the minicircle from a concentration of ~400 nM, and no interaction with the linear fragment is seen at 800 nM (middle panel, lanes 6 and 7). However, the two HMG-box domains form the same type of complexes with the minicircle DNA (compare lane 7, middle panel, and lane 4, right panel). Therefore, the DNA-binding properties of the ARID-HMG1box resemble those of the HMG-box domain of the *Arabidopsis* HMGB-type protein At2g34450 (rather than those of Zm-HMGB1box), since this protein also binds DNA minicircles structure specifically, but it has a generally reduced affinity for DNA, when compared to the Zm-HMGB1box (*28*). As seen by the decreasing amount of the unbound DNA band, full-length ARID-HMG1 binds the minicircle from a concentration of ~25 nM and the linear fragment from a concentration of ~200 nM (left panel, lanes 2 and 5). In contrast to the HMG-box domains, with full-length ARID-HMG1 no distinct complexes are formed, but the protein-bound DNA rather appears to accumulate in the wells of the gel. A similar effect was seen with the full-length maize SSRP1 protein but not with the individual SSRP1box (*33*). The EMSA experiment shows that both ARID-HMG1box and full-length ARID-HMG1 bind DNA structure specifically, and it appears that this property is mediated by the HMG-box domain. The significantly higher affinity for DNA of full-length ARID-HMG1 relative to the individual ARID-HMG1box most likely is due to the presence of the ARID domain in the full-length protein, generally enhancing the affinity for DNA (both linear and circularized). Therefore, our experiments indicate that both the ARID and the HMG-box domain contribute to the DNA binding of ARID-HMG1.

Some mammalian ARID-containing proteins interact preferentially with A/T-rich DNA sites, while other members of the ARID family do not share this property (*18, 19, 21, 22, 44*). Therefore, recognition of A/T sites is not an intrinsic feature of the domain. We have compared DNA binding of full-length ARID-HMG1 to A/T-rich and G/C-rich DNA. Increasing amounts of the protein were incubated with 63-bp DNA fragments, and DNA-binding was examined by native polyacrylamide gel electrophoresis of the samples (Figure 7). With the A/T-rich DNA fragment (containing several A/T stretches) binding was observed from a concentration of 200 nM, as evident from the appearance of a distinct protein/DNA complex (lane 3, top panel). At higher protein concentrations, additional complexes of lower electrophoretic mobility appeared. With the G/C-rich fragment (containing no A/T stretch longer than 4 bp) protein binding was observed only from a concentration of ~800 nM (lane 6, bottom panel), indicating that ARID-HMG1 displays a higher affinity for A/T-rich DNA. Moreover, distinct complexes like those formed with the A/T-rich fragment were not seen with the G/C-rich fragment. Since no binding to the A/T- and G/C-rich fragments was observed with the individual HMG-box domain of ARID-HMG1 (up to 2 μM protein concentration, data not shown), most likely the binding of ARID-HMG1 to the linear DNA fragments can be attributed to ARID−DNA interactions. This is consistent with the EMSA using the 78-bp linear fragment and minicircle (Figure 6), where binding to the linear fragment is only observed for ARID-HMG1 but not for the individual ARID-HMG1box. In summary, our DNA-binding experiments show that, like SIP1 (the only plant ARID protein characterized so far (*35*)), ARID-HMG1 belongs the group of ARID proteins that bind

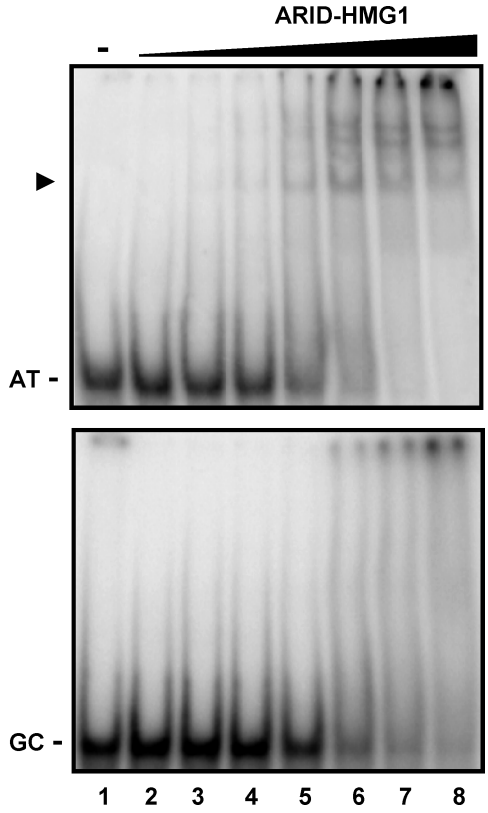

FIGURE 7: ARID-HMG1 binds preferentially to A/T-rich DNA. ^{32}P-labeled A/T- (top panel) or G/C-rich (bottom panel) 63-bp DNA fragments were incubated in the absence (lanes 1) or in the presence of increasing concentrations of the ARID-HMG1 (100 nM, 200 nM, 400 nM, 600 nM, 800 nM, 1.2 μM, 1.8 μM; lanes 2−8, respectively). The DNA-binding reactions were analyzed by native PAGE and phosphorimaging. An arrowhead points at a distinct complex that forms at low protein concentrations with the A/T-rich fragment but not with the G/C-rich fragment. The electrophoretic migration positions of the unbound A/T-rich (AT) and G/C-rich (GC) DNA fragments are indicated.

preferentially A/T-rich DNA. Mediated by the HMG-box domain ARID-HMG1 can bind DNA structure specifically. Accordingly, both the ARID and the HMG-box domain contribute to the DNA interactions of ARID-HMG1, giving the protein characteristics of ARID and HMG-box proteins.

CONCLUSION

The ARID-HMG proteins containing ARID and HMG-box domains represent a conserved plant family of DNA-binding proteins. In the *Arabidopsis* ARID-HMG1 protein, both domains contribute to the DNA interactions of the protein. Accordingly, our biochemical analyses have shown that ARID-HMG1 can bind DNA structure specifically and that it displays a preference for A/T-rich DNA when binding to linear DNA. Proteins from various eukaryotic organisms containing ARID or HMG-box domains are implicated in a variety of functions primarily associated with the regulation of transcription. In the databases, based on the presence of the two DNA-binding motifs the plant HMG-ARID proteins currently are annotated as transcriptional regulators, although there is no experimental evidence for that. However, in view of the emerging evidence for a transcriptional role of other proteins that contain ARID or HMG-box domains, it is likely that the plant ARID-HMG proteins as transcription factors or as chromatin modulators are involved in transcriptional regulation. Future experiments, for instance studying mutants in *ARID-HMG* genes in model systems such as *Arabidopsis*, rice, or *Physcomitrella* will address the role that this novel family of DNA-binding proteins plays in plants.

SUPPORTING INFORMATION AVAILABLE

A table listing the sequences of all the oligonucleotide primers used in this study for PCRs related to molecular cloning and rtPCR (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Bianchi, M. E., Falciola, L., Ferrari, S., and Lilley, D. M. (1992) The DNA binding site of HMG1 protein is composed of two similar segments (HMG boxes), both of which have counterparts in other eukaryotic regulatory proteins. *EMBO J. 11*, 1055–1063.
2. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D., and Thomas, J. O. (1993) Structure of the HMG box motif in the B-domain of HMG1. *EMBO J. 12*, 1311–1319.
3. Baxevanis, A. D., and Landsman, D. (1995) The HMG-1 box protein family: classification and functional relationships. *Nucleic Acids Res. 23*, 1604–1613.
4. Read, C. M., Cary, P. D., Crane-Robinson, C., Discroll, P. C., Carrillo, M. O. M., and Norman, D. G. (1995) *The structure of the HMG box and its interactions with DNA* (Eckstein, F., and Lilley, D. M. J., Eds.) pp 222−250, Springer, Berlin.
5. Stros, M., Launholt, D., and Grasser, K. D. (2007) The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell. Mol. Life Sci. 64*, 2590–2606.
6. Bustin, M. (1999) Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol. Cell. Biol. 19*, 5237–5246.
7. Grosschedl, R., Giese, K., and Pagel, J. (1994) HMG domain proteins: architectural elements in nucleoprotein structures. *Trends Genet. 10*, 94–100.
8. Thomas, J. O., and Travers, A. A. (2001) HMG1 and 2, and related architectural DNA-binding proteins. *Trends Biochem. Sci. 26*, 167–174.
9. Agresti, A., and Bianchi, M. E. (2003) HMGB proteins and gene expression. *Curr. Opin. Genet. Dev. 13*, 170–178.
10. Bianchi, M. E., and Agresti, A. (2005) HMG proteins: dynamic players in gene regulation and differentiation. *Curr. Opin. Genet. Dev. 15*, 496–506.
11. Grasser, K. D., Launholt, D., and Grasser, M. (2007) High mobility group proteins of the plant HMGB family: dynamic chromatin modulators. *Biochim. Biophys. Acta 1769*, 346–357.
12. Gregory, S. L., Kortschak, R. D., Klionis, B., and Saint, R. (1996) Characterization of the Dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA-binding proteins. *Mol. Cell. Biol. 16*, 792–799.
13. Herrscher, R. F., Kaplan, M. H., Lelsz, D. L., Das, C., Scheuermann, R., and Tucker, P. W. (1995) The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev. 15*, 3067–3082.
14. Kortschak, R. D., Tucker, P. W., and Saint, R. (2000) ARID proteins come in from the desert. *Trends Biochem. Sci. 25*, 294–299.
15. Wilsker, D., Patsialou, A., Dellas, P. B., and Moran, E. (2002) ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ. 13*, 95–106.
16. Wilsker, D., Probst, L., Wain, H. M., Maltais, L., Tucker, P. W., and Moran, E. (2005) Nomenclature of the ARID family of DNA-binding proteins. *Genomics 86*, 242–251.
17. Iwahara, J., and Clubb, R. T. (1999) Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). *EMBO J. 18*, 6084–6094.
18. Iwahara, J., Iwahara, M., Daughdrill, D. W., Ford, J., and Clubb, R. T. (2002) The structure of the dead ringer-DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. *EMBO J. 21*, 1197–1209.
19. Tu, S., Teng, Y. C., Yuan, C., Wu, Y. T., Chan, M. Y., Cheng, A. N., Lin, P. H., Juan, L. J., and Tsai, M. D. (2008) The ARID

domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. *Nat. Struct. Mol. Biol. 15*, 419–421.
20. Yuan, Y. C., Whitson, R. H., Liu, Q., Itakura, K., and Chen, Y. (1998) A novel DNA-binding motif shares structural homology to DNA replication and repair nucleases and polymerases. *Nat. Struct. Biol. 5*, 959–864.
21. Patsialou, A., Wilsker, D., and Moran, E. (2005) DNA-binding properties of ARID family proteins. *Nucleic Acids Res. 33*, 66–80.
22. Wilsker, D., Patsialou, A., Zumbrun, S. D., Suhkmann, K., Chen, Y., Dallas, P. B., and Moran, E. (2004) The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes. *Nucleic Acids Res. 32*, 1345–1353.
23. Launholt, D., Grønlund, J. T., Nielsen, H. K., and Grasser, K. D. (2007) Overlapping expression patterns among the genes encoding *Arabidopsis* chromosomal high mobility group (HMG) proteins. *FEBS Lett. 581*, 1114–1118.
24. O'Neill, S. D., and Zheng, C. C. (1998) Abundance of mRNAs encoding HMG1/HMG2 class high-mobility-group DNA-binding proteins are differentially regulated in cotyledons of *Pharbitis nil*. *Plant Mol. Biol. 37*, 235–241.
25. Wu, Q., Zhang, W., Pwee, K.-H., and Kumar, P. P. (2003) Cloning and characterization of rice HMGB1 gene. *Gene 312*, 103–109.
26. Yamamoto, S., and Minamikawa, T. (1998) The isolation and charactersiation of a cDNA encoding a high mobility group protein HMG-1 from *Canavalia gladiata* D.C. *Biochim. Biophys. Acta 1396*, 47–50.
27. Grasser, K. D., Krech, A. B., and Feix, G. (1994) The maize chromosomal HMGa protein recognises structural features of DNA and increases DNA flexibility. *Plant J. 6*, 351–358.
28. Grasser, M., Lentz, A., Lichota, J., Merkle, T., and Grasser, K. D. (2006) The *Arabidopsis* genome encodes structurally and functionally diverse HMGB-type proteins. *J. Mol. Biol. 358*, 654–664.
29. Ritt, C., Grimm, R., Fernández, S., Alonso, J. C., and Grasser, K. D. (1998) Basic and acidic regions flanking the HMG domain of maize HMGa modulate the interactions with DNA and the self-association of the protein. *Biochemistry 37*, 2673–2681.
30. Webster, C. I., Packman, L. C., Pwee, K.-H., and Gray, J. C. (1997) High mobility group proteins HMG-1 and HMGI/Y bind to a positive regulatory region of the pea plastocyanin gene. *Plant J. 11*, 703–715.
31. Wu, Q., Zhang, W., Pwee, K.-H., and Kumar, P. P. (2003) Rice HMGB1 protein recognizes DNA structures and bends DNA efficiently. *Arch. Biochem. Biophys. 411*, 105–111.
32. Krohn, N. M., Stemmer, C., Fojan, P., Grimm, R., and Grasser, K. D. (2003) Protein kinase CK2 phosphorylates the HMG domain protein SSRP1 inducing the recognition of UV-damaged DNA. *J. Biol. Chem. 278*, 12710–12715.
33. Röttgers, K., Krohn, N. M., Lichota, J., Stemmer, C., Merkle, T., and Grasser, K. D. (2000) DNA-interactions and nuclear localisation of the chromosomal HMG domain protein SSRP1 from maize. *Plant J. 23*, 395–405.
34. Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, C., Sherman, B. K., and Yu, G. (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science 290*, 2105–2110.
35. Zhu, H., Chen, T., Zhu, M., Fang, Q., Kang, H., Hong, Z., and Zhang, Z. (2008) A novel ARID DNA-binding protein interacts with SymRK and is expressed during early nodule development in *Lotus japonicus*. *Plant Physiol. 148*, 337–347.
36. Haasen, D., Köhler, C., Neuhaus, G., and Merkle, T. (1999) Nuclear export of proteins in plants: AtXPO1 is the export receptor for leucine-rich nuclear export signals in *Arabidopsis thaliana*. *Plant J. 20*, 695–706.
37. Thomsen, M. S., Franssen, L., Launholt, D., Fojan, P., and Grasser, K. D. (2004) Interactions of the basic N-terminal and the acidic C-terminal domains of the maize chromosomal HMGB1 protein. *Biochemistry 43*, 8029–8037.
38. Stemmer, C., Schwander, A., Bauw, G., Fojan, P., and Grasser, K. D. (2002) Protein kinase CK2 differentially phosphorylates maize chromosomal high mobility group B (HMGB) proteins modulating their stability and DNA interactions. *J. Biol. Chem. 277*, 1092–1098.
39. Launholt, D., Merkle, T., Houben, A., Schulz, A., and Grasser, K. D. (2006) *Arabidopsis* chromatin-associated HMGA and HMGB use different nuclear targeting signals and display highly dynamic localization within the nucleus. *Plant Cell 18*, 2904–2918.
40. Lyngaard, C., Stemmer, C., Stensballe, A., Graf, M., Gorr, G., Decker, E., and Grasser, K. D. (2008) *Physcomitrella* HMGA-type proteins display structural differences compared to their higher plant counterparts. *Biochem. Biophys. Res. Commun. 374*, 653–657.
41. Duroux, M., Houben, A., Ruzicka, K., Friml, J., and Grasser, K. D. (2004) The chromatin remodelling complex FACT associates with actively transcribed regions of the *Arabidopsis* genome. *Plant J. 40*, 660–671.
42. Koroleva, O. A., Tomlinson, M. L., Leader, D., Shaw, P., and Doonan, J. H. (2005) High-throughput protein localization in *Arabidopsis* using *Agrobacterium*-mediated transient expression of GFP-ORF fusions. *Plant J. 41*, 162–174.
43. Grasser, K. D., Grill, S., Duroux, M., Launholt, D., Thomsen, M. S., Nielsen, B. V., Nielsen, H. K., and Merkle, T. (2004) HMGB6 from *Arabidopsis thaliana* specifies a novel type of plant chromosomal HMGB protein. *Biochemistry 43*, 1309–1314.
44. Kim, S., Zhang, Z., Upcharch, S., Isern, N., and Chen, Y. (2004) Structure and DNA-binding sites of the SWI1 AT-rich interaction domain (ARID) suggest determinants for sequence-specific DNA recognition. *J. Biol. Chem. 279*, 16670–16676.

BI801772K