HMGB6 from *Arabidopsis thaliana* Specifies a Novel Type of Plant Chromosomal HMGB Protein†

Klaus D. Grasser,*,‡ Simon Grill,‡ Meg Duroux,‡ Dorte Launholt,‡ Malene S. Thomsen,‡ Birthe V. Nielsen,‡ Hanne K. Nielsen,‡ and Thomas Merkle§

Institute of Life Sciences, Aalborg University, Sohngaardsholmsvej 49, DK-9000 Aalborg, Denmark, and Institut für Biologie II, Albert-Ludwigs-Universität Freiburg, Schänzlestrasse 1, D-79104 Freiburg, Germany

Received October 29, 2003; Revised Manuscript Received November 25, 2003

ABSTRACT: The high-mobility group (HMG) proteins of the HMGB family are chromatin-associated proteins that act as architectural factors in various nucleoprotein structures, which regulate DNA-dependent processes such as transcription and recombination. Database analyses revealed that in addition to the previously identified HMGB1−HMGB5 proteins, the *Arabidopsis* genome encodes at least three other family members having the typical overall structure of a central HMG-box DNA binding domain, which is flanked by basic and acidic regions. These novel HMGB proteins display some structural differences, when compared to HMGB1−HMGB5. Therefore, a representative of the identified proteins, now termed HMGB6, was further analyzed. The HMGB6 protein of ∼27 kDa is the largest plant HMGB protein identified so far. This is essentially due to its unusually extended N-terminal domain of 109 amino acid residues. Subcellular localization experiments demonstrate that it is a nuclear protein. According to CD measurements, HMGB6 has an α-helical HMG-box domain. HMGB6 can bind DNA structure-specifically, and it is a substrate for the protein kinase CK2α. Because of these features, HMGB6, and presumably its relatives, can be considered members of the plant HMGB protein family. Hence, eight different chromosomal HMGB proteins are expressed in *Arabidopsis*, and they may serve specialized architectural functions assisting various DNA-dependent processes.

High-mobility group (HMG)[1] proteins represent a heterogeneous family of relatively abundant nonhistone proteins associated with eukaryotic chromatin (*1*). Proteins belonging to the subgroup of the HMGB proteins (*2*) act as architectural components in chromatin, facilitating the assembly of nucleoprotein complexes, which are involved, for instance, in the regulation of transcription and recombination (*1*, *3*, *4*). HMGB proteins contain one or two copies of a distinctive DNA-binding motif termed the HMG-box domain, whose three-dimensional fold is well-conserved, consisting essentially of three α-helices which are arranged in an L shape (*3*, *5*). The HMG-box domain mediates non-sequence-specific binding of the HMGB proteins to linear DNA, and high-affinity interactions with distorted DNA structures such as four-way junctions, minicircles, and *cis*-platinated DNA (*1*, *3*−*6*). In complexes with B-DNA, the concave surface of the HMG-box domain binds predominantly the minor groove of the DNA, bending the DNA by more than 90° (*3*, *5*).

Chromatin-associated HMGB proteins have been characterized from a number of plant species, revealing similarities as well as dissimilarities with their vertebrate, insect, and yeast counterparts (*7*−*13*). Typically, the plant HMGB proteins have a single HMG-box domain, which is flanked by a basic N-terminal domain and an acidic C-terminal domain. While the HMG-box domain of the various plant HMGB proteins is relatively conserved, the basic and acidic flanking regions are variable in length and sequence (*14*). According to Northern and Western blot analyses, HMGB proteins occur ubiquitously in the plant (*12*, *15*−*17*). Plant HMGB proteins bind linear DNA non-sequence-specifically with moderate affinity, but recognize specifically certain DNA structures, and they severely bend the DNA upon binding (*8*, *10*, *11*, *18*−*21*). Higher-plant nuclei contain a variety of HMGB proteins ranging from ∼13 to 20 kDa. Thus, five different HMGB proteins have been identified from both the monocot plant maize and the dicot plant *Arabidopsis* (*22*). The various maize HMGB proteins differ in their chromatin association (*23*), in their expression levels (*16*), in some of their DNA interactions (*8*), in their post-translational modification by protein phosphorylation (*24*), and in their interaction with transcription factor Dof2 (*25*). Therefore, they may be adapted to act as architectural factors in different nucleoprotein structures (*22*).

Here we have examined the entire complement of HMGB-type proteins encoded by the *Arabidopsis* genome, revealing that in addition to the five previously identified HMGB proteins (*19*), *Arabidopsis* encodes further candidates that may belong to the HMGB family. There are at least three genes encoding proteins, which display structural features

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

* To whom correspondence should be addressed. Telephone: +45 9635 9126. Fax: +45 9814 1808. E-mail: kdg@bio.auc.dk.

‡ Aalborg University.

§ Albert-Ludwigs-Universität Freiburg.

[1] Abbreviations: HMG, high-mobility group; EMSA, electrophoretic mobility shift assay; CD, circular dichroism; GFP, green fluorescent protein; UTR, untranslated region; SSRP1, structure-specific recognition protein 1.

10.1021/bi035931c CCC: $27.50
Published on Web 01/10/2004

typical of plant HMGB proteins. One member of this group, now termed HMGB6, was functionally characterized, proving that indeed it shares properties with members of the HMGB family.

EXPERIMENTAL PROCEDURES

Discovery and Analysis of Sequences Encoding HMG-Box Domains. The amino acid sequence (residues K33−Y106) of the *Arabidopsis* HMGB2 protein, which is a typical representative of plant HMGB proteins (*19*), was used to search the proteins encoded by the *Arabidopsis* genome (http://www.arabidopsis.org/) for sequences encoding HMG-box domains using BlastP (*26*). The 15 sequences most similar to the HMGB2 HMG-box sequence ($E < 0.1$) were inspected in more detail. Sequences were aligned pairwise (www.ebi.ac.uk/emboss/align/) and by multiple-sequence alignments (*27*), which were used to construct sequence similarity trees with ClustalW, revealing relationships among the analyzed sequences (www.ebi.ac.uk/clustalw/). Intron sequences were detected by comparison of cDNA and genomic DNA sequences.

Protein Production. The DNA sequences encoding full-length HMGB6(M1−Y241) and truncated HMGB6 versions (S110−N182 and S110−Y241) were amplified by PCR, using primers P1 (5′-AAGGATCCATGGCTGGACCATC-GACAACTT) and P2 (5′-AATTAAGCTTAGTAGTCATC-CAAAATCTCTTCT), P3 (5′-AAGGATCCTCTACTTCA-AACAAGCCCA) and P4 (5′-AATAAGCTTAATTACTTTC-CAGTGACTTGTTATA), and P3 and P2, respectively. PCRs were performed using *pfu* DNA polymerase (MBI Fermentas) and an *Arabidopsis* cDNA library as a template. The obtained PCR fragments were digested with *Bam*HI and *Hin*dIII and cloned into expression vector pQE9 (Qiagen), resulting in plasmids pQE9-HMGB6(M1−Y241), pQE9-HMGB6(S110−N182), and pQE9-HMGB6(S110−Y241), which were checked by DNA sequencing. Recombinant HMGB6 proteins were expressed as His_6-tagged proteins in *Escherichia coli* and purified by three-step column chromatography as described previously (*8*). Purified proteins were checked by SDS−PAGE and MALDI-TOF mass spectrometry.

Circular Dichroism (CD). CD measurements using 10 μM HMGB6(S110−N182) were performed using a Jasco J-715 instrument as described previously (*24*).

Electrophoretic Mobility Shift Assay (EMSA). HMGB6-(S110−N182) was incubated with supercoiled and *Eco*RI-linearized plasmid pUC19 (0.5 μg each) in binding buffer [10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, 0.05% bromophenol blue, and 0.05% xylene cyanol] for 10 min. Binding reactions were analyzed in 1% agarose gels in 0.5× TBE, and the DNA was stained with ethidium bromide.

Protein Phosphorylation by CK2. Full-length and truncated HMGB6 proteins were phosphorylated using purified recombinant maize protein kinase CK2α and [^{32}P]ATP, and proteins were analyzed by SDS−PAGE and phosphorimaging as described previously (*24*).

Transient Protoplast Transformation Assays with GFP Fusion Constructs. To construct a GFP fusion of HMGB6, the coding sequence of HMGB6 was amplified by PCR using primers P1 and P5 (5′-AATGGGCCCGTAGTCATCCAA-AATCTCTTCT) and an *Arabidopsis* cDNA library as a template. The obtained PCR fragment was digested with *Bam*HI and *Sma*I and cloned into *Bam*HI- and *Sma*I-digested plasmid p3′GFP, resulting in an in-frame fusion of the HMGB6 coding sequence with the GFP coding sequence at the 3′ end. The plasmid insertion was checked by DNA sequencing. The plasmid contains mGFP4(S_{65}T) under the control of the cauliflower mosaic virus 35S promoter and the nopaline synthase terminator in pUC19. Protoplasts were prepared from tobacco BY-2 cells (and *Arabidopsis* suspension cultured cells) grown in the dark, and transiently transformed with the HMGB6−GFP fusion construct (and control constructs) by PEG-mediated transformation as described previously (*28*). Excitation of GFP was performed with a standard UV light source and fluorescein isothiocyanate filters. For confocal laser scanning microscopy, samples were directly examined under oil with a 63× objective and a DM RBE TCS4D microscope (Leica) equipped with an argon−krypton laser (excitation at 488 nm, beam splitter at 510 nm, filter at 515 nm) using Leica Scanware. Analysis of the localization of the GFP fusion proteins was performed in three independent experiments, representing approximately 60−80 transformed protoplasts.

Table 1: HMG-Box Domain-Containing Sequences of *Arabidopsis*

protein[a]	length (amino acids)	mass	accession no.	AGI locus	expression[b]	HMGB[c]
HMGB1	178	20265	Y14071	At3g51880	cDNA, EST	+
HMGB2	144	15982	Y14072	At1g20693	cDNA, EST	+
HMGB3	141	15681	Y14073	At1g20696	cDNA, EST	+
HMGB4	138	15364	Y14074	At2g17560	cDNA, EST	+
HMGB5	125	14203	Y14075	At4g35570	cDNA, EST	+
HMGB6	241	26964	AY086023	At5g23420	cDNA, EST	+
	221	25761	NM120615	At5g05330	−	+
	151	17481	AY0846626	At2g34450	cDNA, EST	+
	149	16997	NM113794	At5g23405	cDNA, EST	+
SSRP1	646	71629	NM106260	At3g28730	cDNA, EST	−
	338	38757	NM112180	At1g76110	cDNA, EST	−
	319	36278	NM114221	At3g13350	cDNA, EST	−
	615	69247	NM117178	At3g43530	−	−
	446	52285	NM117178	At4g11080	cDNA	−
	456	53213	NM118511	At4g23800	cDNA	−

[a] Name of the protein, in cases where the (native and/or recombinant) protein has been characterized. [b] Evidence which proves that the gene is expressed. [c] A plus sign indicates whether the encoded protein has the overall structure that is typical of plant HMGB proteins by having a mass of <30 kDa and a central HMG-box domain, which is flanked by a basic N-terminal domain and an acidic C-terminal domain.

RESULTS

Discovery of Sequences Encoding Novel HMGB-Type Proteins. A BLAST search of the *Arabidopsis thaliana* database (http://www.arabidopsis.org/) using the sequence of the HMG-box DNA binding domain of the *Arabidopsis* HMGB2 protein (which is a typical representative of the plant HMGB family) identified 15 sequences containing putative HMG-box domain(s) (Table 1). Among the identified sequences, the previously characterized HMGB1−HMGB5 proteins (*19*) were found, as well as the SSRP1 protein that has been characterized from maize (*28*). To our knowledge, the other nine sequences have not previously been characterized. Four of these genes (At5g23420, At5g05330, At2g34450, and At5g23405) encode proteins (<30 kDa) that have a central HMG-box domain, which is flanked by a basic

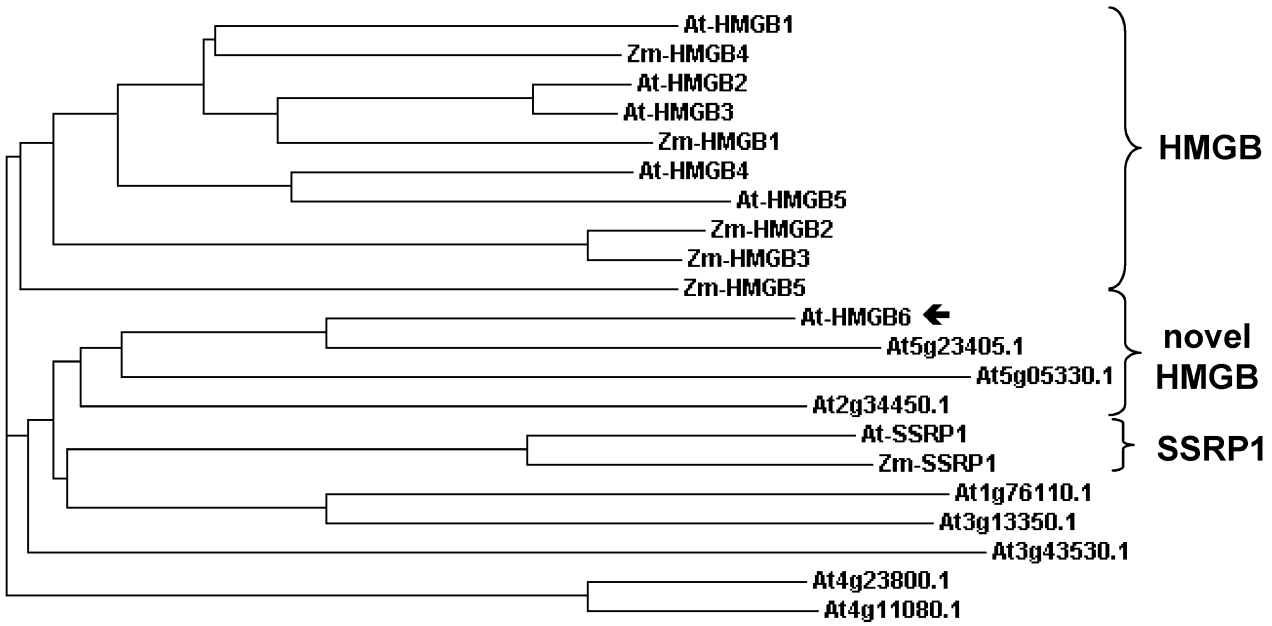

FIGURE 1: Amino acid sequence similarity of proteins containing HMG-box DNA binding domain(s) from *A. thaliana* (At) and *Zea mays* (Zm). The protein sequences were aligned by multiple-sequence alignment that was used to construct the tree (displayed as a phylogram). The top part contains all the *Arabidopsis* and maize HMGB proteins (HMGB1−HMGB5) that have been previously characterized. The bottom part contains the novel *Arabidopsis* HMGB proteins, the SSRP1 proteins, and other proteins containing putative HMG-box domains. The HMGB6 protein, which has been analyzed in more detail in this report, is denoted with an arrow.

N-terminal domain and an acidic C-terminal domain, and therefore match the overall structure typical of plant HMGB proteins (Table 1). The other proteins that were identified have putative HMG-box domain(s), but they do not share the characteristic overall structure of plant HMGB proteins, and have not been further analyzed. The sequences of the 15 identified *Arabidopsis* proteins were aligned (together with the characterized HMGB and SSRP1 proteins from maize), and an amino acid sequence similarity tree was constructed (Figure 1). Here, the previously characterized HMGB1−HMGB5 proteins from maize and *Arabidopsis* were found to cluster together, while the SSRP1 sequences and the other sequences derived from the *Arabidopsis* database form a second group. Within that group, the four above-mentioned HMGB-type sequences cluster in a group that we call novel HMGB proteins throughout this report (Figure 1). In addition to the amino acid sequence differences (relative to the previously characterized HMGB proteins), genes encoding the novel HMGB proteins do not contain an intron within the upstream untranslated region (UTR), whereas all the genes encoding the *Arabidopsis* HMGB1−HMGB5 proteins contain an intron in the 5′-UTR (data not shown). The locus At5g23420 encodes an HMGB-type protein of ~27 kDa and a calculated isoelectric point of ~4.5 (now termed HMGB6). HMGB6 is most similar to the protein encoded by the At5g23405 locus, its amino acid sequence being 37% identical over the entire lengths of sequence and 52% identical within the HMG-box domains. HMGB6 is the biggest plant HMGB protein identified to date, which is essentially due to its extended N-terminal domain of 109 amino acid residues (Figure 2). The relative position of the three introns within the region encoding the HMG-box domain is conserved between the *hmgb6* gene and the previously identified *Arabidopsis hmgb* genes, but the exon/intron structure differs outside the HMG-box region. Thus, the *hmgb6* gene has two extra exons accounting for the extensive N-terminal domain, while there is an intron lacking in the region encoding the acidic C-terminal domain (Figure 2). Comparison of the amino acid sequences of HMGB6 and of the HMGB1−HMGB5 proteins revealed that many critical residues within the HMG-box domain are conserved, such as the highly conserved tryptophan residue and the potential primary DNA intercalating residue (Phe123 in HMGB6), whereas others differ such as the potential secondary intercalating residue (Ala143 in HMGB6), which is a conserved valine residue in the *Arabidopsis* HMGB1−HMGB5 proteins (Figure 2). The residues of the HMG-box domain intercalating between base pairs of the DNA contribute significantly to DNA binding and bending (*29*−*34*). Because of these structural differences, HMGB6 was further analyzed, to determine whether the protein displays functional characteristics typical of plant chromosomal HMGB proteins.

HMGB6 Is Localized to the Cell Nucleus. Since HMGB6 contains this extensive N-terminal domain, we hypothesized that it might eventually contain a signal sequence targeting the protein to a destination other than the cell nucleus, as HMGB-type proteins have also been found in yeast and mammalian mitochondria (*14*). Therefore, we constructed a plasmid suitable for expression of an HMGB6−GFP fusion protein in plant protoplasts. In transient transformation assays performed with tobacco BY-2 suspension cell protoplasts, the expression of the HMGB6−GFP fusion protein was driven by the CaMV 35S promoter. Transformed protoplasts were analyzed by confocal laser scanning microscopy. Two constructs served as controls: GFP that is not enriched in the nucleus (Figure 3A) and GFP−NLS−CHS−NESmutated localized in the nucleus (*35*) (Figure 3B). The HMGB6−GFP fusion clearly accumulated in the nucleus (Figure 3C). Similar results were obtained using protoplasts derived from *Arabidopsis* suspension cultured cells (data not shown). Colocalization of the nucleus and green fluorescence is

```
HMGB6  [1]MAGPSTTSNAPKQRKRVEAETSSNTSTTLRRAKDGSAFALC▼EGCNKSVAVALISMHNCSLD

HMGB6   AKIRVNL▼EAQVVETQAEAKKKPAEK▼KKTTSDGPKPKRLKKTNDEKKSSSTSNKPKRPLTAFFIFM▼SDFRKTFKSEHN.GSLAKD▼AAKIGGEKWK
HMGB1  [1]MKTAKGKDKVKTTKEALKPVDDR▼KVGKRKAPAEKPTKRETRKEKKFTAKKDPNKPKRAPSAFFVFL▼EDFRVTFKKENPNVKAVSA▼VGKAGGQKWK
HMGB2          [1]MKGAKSKTETRSSK▼LSVTKKPAKGAGRGKAAAKDPNKPKRPASAFFVFM▼EDFRETFKKENPKNKSVAT▼VGKAAGDKWK
HMGB3          [1]MKGAKSKAETRSTK▼LSVTKKPAKGA...KGAAKDPNKPKRPSSAFFVFM▼EDFRVTYKEEHPKNKSVAA▼VGKAGGEKWK
HMGB4             [1]MKGGESKAGATSTDQR▼LKTRGRKAGKKTKKDPNQPKRPPSAFFVFL▼EDFRKEFNLANPNNKSVAT▼VGKAAGARWK
HMGB5             [1]MKDNQTEVESRSTDDR▼LKVRGNKVGKKTKDPNRPKKPPSPFFVFL▼DDFRKEFNLANPDNKSVGN▼VGRAAGKKWK

HMGB6   .LTEE▼EKKVYLDKAAELKAEYNKSLESNDADEE▼EEDEEKQSDDVDDAEEKQVDDDDEVEEKEVENTDDDKKEAEGKEEEEEEEILDDY[241]  aa241
HMGB1  SMSQA▼EKAPYEEKAAKRKAEYEKQMDAYNKNL▼EEGSDESEKSRSEINDEDEASGE▼EELLEKEAAGDDEEEEEEEDDDDDDDEEED[178]  aa178
HMGB2  SLSDS▼EKAPYVAKAEKRKVEYEKNIKAYNKKL▼EEGPKEDEESDKSVSEVNDEDDAEDGSEE▼EEDDD[144]  aa144
HMGB3  SLSDS▼EKAPYVAKADKRKVEYEKNMKAYNKKL▼EEGPKEDEESDKSVSEVNDEDDAEDGSEE▼EEDDD[141]  aa141
HMGB4  SMTDE▼DKAPYVAKAESRKTEYIKNVQQYNLKL▼ASGTNREEDDSDKSKSEVDEAVSEE▼EAEDDD[138]  aa138
HMGB5  TMTEE▼ERAPFVAKSQSKKTEYAVTMQQYNMEL▼ANGNKTTGDDEKQEKAADD[125]  aa125
```

FIGURE 2: Amino acid sequence alignment of HMGB6 and the previously characterized *Arabidopsis* HMGB1−HMGB5 proteins. The N-terminal and C-terminal amino acid positions of the protein sequences are indicated. Note the extensive N-terminal region unique to HMGB6 in the first line of the alignment. The central HMG-box domain of the proteins is highlighted in black, and the relative positions of introns in the *hmgb* genes are represented with triangles (▼) in the amino acid sequences. Potential primary and secondary DNA intercalating residues [deduced by analogy from the determined structures of other HMGB proteins (*3*)], which are conserved Phe and Val residues in the *Arabidopsis* HMGB1−HMGB5 proteins, are in bold.

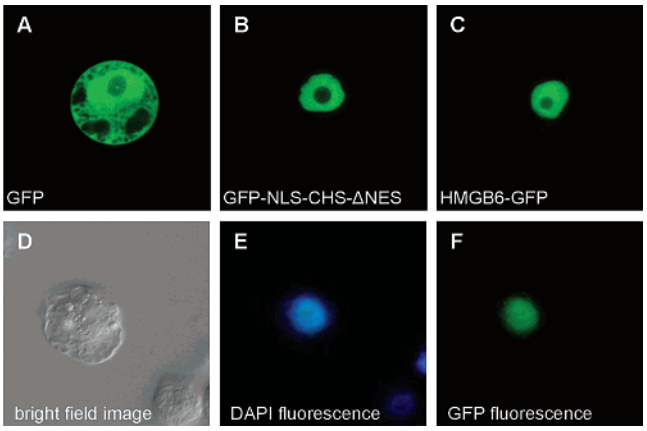

FIGURE 3: HMGB6 is localized in the cell nucleus. Tobacco BY-2 protoplasts were transiently transformed with plasmids, resulting in the expression of GFP, which does not accumulate in the nucleus and therefore can be detected in cytoplasm and nucleus (A), or expressing GFP fused to chalcone synthase (CHS) containing the nuclear localization signal (NLS) of the SV40 large T-antigen as the only functional targeting signal (*35*), and therefore localizes to the nucleus (B). The HMGB6−GFP fusion is clearly localized in the cell nucleus (C). The subcellular localization of the green fluorescence was assessed by confocal laser scanning microscopy. The colocalization of the nucleus and green fluorescence is demonstrated by analysis of a BY-2 protoplast expressing the GFP−SSRP1 fusion protein localized in the nucleus (*28*). Bright field image using Normarski optics (D), DAPI fluorescence (E), and GFP fluorescence (of GFP−SSRP1, F) are shown for the same protoplast. All images are at the same scale.

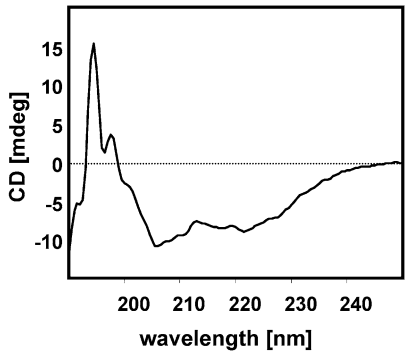

FIGURE 4: HMG-box domain of HMGB6 with an α-helical structure. A CD spectrum was recorded for the HMGB6(S110−N178) protein in the wavelength range of 190−250 nm, resulting in a spectrum indicating α-helical structure.

demonstrated by comparing the bright field image, DAPI fluorescence, and nuclear localization of the GFP−SSRP1 protein (*28*) (Figure 3D−F). Hence, typical of chromosomal HMGB proteins, *Arabidopsis* HMGB6 is most likely a nuclear protein.

HMGB6 Has Structural and Functional Features Characteristic of HMGB Proteins. To test structural and functional properties of the *Arabidopsis* HMGB6 protein, we expressed full-length (M1−Y241) and truncated (S110−N182 and S110−Y241) versions of HMGB6 in *E. coli* and purified the recombinant proteins by three-step column chromatography. Full-length HMGB6 proved to be very sensitive to proteolytic degradation (relative to other plant HMGB proteins we have analyzed; see also below). HMGB6(S110−N182) represents the individual HMG-box domain, and HMGB6(S110−Y241) comprises the HMG-box domain and the acidic C-terminal tail, lacking the basic N-terminal domain. Both truncated proteins could be produced efficiently, suggesting that mainly the unusual N-terminal domain of HMGB6 accounts for the sensitivity of the full-length protein to proteolysis. The individual HMG-box DNA binding domain was examined by CD spectroscopy. A CD wavelength spectrum was recorded in the range of 190−250 nm. The spectrum displays a shape typical of an α-helical protein, having the characteristic negative peaks around 208 and 222 nm (Figure 4). Accordingly, the HMG-box domain of HMGB6 has an α-helical structure consistent with the published structures of a number of HMG-box domains (*3*, *5*).

Using EMSAs, the functionality of the HMG-box domain of HMGB6 in DNA binding was tested. Increasing amounts of HMGB6(S110−N182) were incubated with a mixture of supercoiled and linearized plasmid DNA, and protein binding to DNA was examined by agarose gel electrophoresis of the samples. As is evident from the reduced electrophoretic mobility of the bands corresponding to the unbound DNA

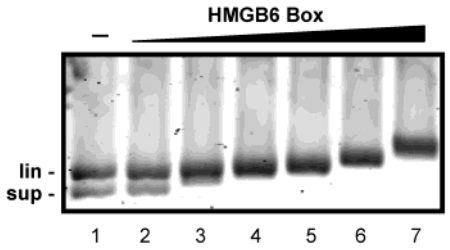

FIGURE 5: HMG-box domain of HMGB6 which structure-specifically binds to supercoiled DNA. A mixture of the supercoiled and linearized pUC19 plasmid was incubated in the absence (lane 1) or presence of increasing concentrations of the HMGB6(S110−N178) protein (0.1, 0.25, 0.5, 0.75, 1, and 1.5 μM, in lanes 2−7, respectively). The DNA binding reactions were analyzed by electrophoresis in agarose gels. The electrophoretic migration positions of unbound supercoiled (sup) and linearized (lin) pUC19 are indicated.

upon protein binding, HMGB6(S110−N182) interacts with both types of DNA (Figure 5). The supercoiled plasmid is bound at lower protein concentrations (starting from 0.25 μM, lane 3) than the linearized plasmid (starting from 1 μM, lane 6). Therefore, the HMG-box domain of HMGB6 preferentially binds the supercoiled DNA. The DNA interaction (binding to the supercoiled pUC19 starting from 0.25 μM and the ~4-fold preference for the supercoiled form over the linear plasmid) resembles the DNA binding properties of the maize HMGB proteins (*8*). The recognition of DNA structures such as in supercoiled DNA is another characteristic feature of HMG-box DNA binding domains (*14*).

As the *Arabidopsis* HMGB1, HMGB2/3, and HMGB4 proteins are phosphorylated by protein kinase CK2α at several amino acid positions (*36*), the question of whether HMGB6 is also a substrate for this enzyme was analyzed. Full-length HMGB6 and the truncated versions were examined in phosphorylation assays with recombinant maize CK2α and [^{32}P]ATP as the phosphate donor (Figure 6). Full-length HMGB6 is phosphorylated by CK2α, but not as efficiently as the *Arabidopsis* HMGB1 protein, which served as a positive control for the phosphorylation reaction (lanes 1 and 2). In this experiment (lane 2), in addition to full-length HMGB6 (denoted with an asterisk), two predominant degradation products are visible (mentioned above). The protein consisting of the HMG-box domain and the acidic tail (HMGB6 ΔN-ter, S110−Y241) is a substrate for the protein kinase, whereas the individual HMG-box domain (S110−N182) is not phosphorylated (lanes 3 and 4). Therefore, as with maize HMGB1 and HMGB2/3 (*24*), residue(s) within the acidic C-terminal tail are phosphorylated in HMGB6 by protein kinase CK2α. According to the results of MALDI-TOF mass spectrometry performed on HMGB6-(S110−Y241) phosphorylated by CK2α in the presence of unlabeled ATP, a single residue is phosphorylated (data not shown). The most likely phosphoacceptor site is Thr220 (*37*, *38*), which is compatible with phosphorylation site predictions based on the amino acid sequence of HMGB6 using the PhosphoBase 2.0 software (*39*, *40*).

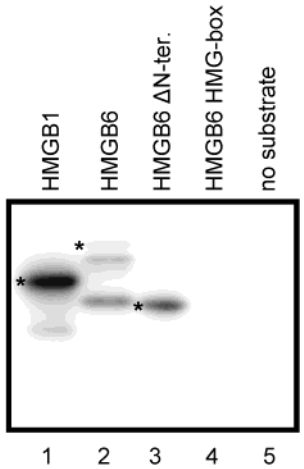

FIGURE 6: HMGB6 is phosphorylated by protein kinase CK2α within its acidic C-terminal domain. Full-length HMGB6(M1−Y241), the N-terminally truncated protein HMGB6(S110−Y241), and the individual HMG-box domain HMGB6(S110−N178) were incubated with recombinant maize CK2α in the presence of [γ-^{32}P]-ATP as the phosphate donor (lanes 2−4, respectively). *Arabidopsis* HMGB1 (lane 1) served as a positive control for the phosphorylation reaction (*36*), and the phosphorylation reaction mixture with no substrate protein added was the negative control (lane 5). Phosphorylation reactions were analyzed by SDS−PAGE and scanning of the gels using a Phosphorimager. The phosphorylated substrate proteins are denoted with asterisks, but some degradation products (which are predominant with full-length HMGB6) can be seen as well (see the text).

DISCUSSION

HMGB proteins have been structurally and functionally characterized from various mono- and dicot plants (*7*−*13*, *17*, *18*). Taking advantage of their solubility in 2% trichloroacetic acid, Moehs et al. (*41*) have isolated HMG proteins from *Arabidopsis* leaves, but their exact identity has not been determined. More recently, five HMGB proteins encoded by *Arabidopsis* cDNAs were found to bind DNA structure-specifically (*19*), a characteristic feature of chromosomal HMGB proteins (*3*, *5*, *14*). By searching the *Arabidopsis* genome database for sequences encoding HMG-box DNA binding domains, we discovered four novel *hmgb* genes, and according to EST/cDNA data (http://www.arabidopsis.org/), three of them are expressed (Table 1). Therefore, together with the five previously characterized HMGB proteins (*19*), eight different HMGB proteins are present in *Arabidopsis*. Since at least five different HMGB proteins are expressed simultaneously in maize (*16*), plants in general seem to have more HMGB variants than other eukaryotes that express a smaller number of different HMGB proteins (*3*, *42*). The majority of the maize and *Arabidopsis* HMGB1−HMGB5 proteins display also a comparatively low degree of amino acid sequence conservation, especially outside the HMG-box domain (*14*). Moreover, plant HMGB proteins are differentially phosphorylated by protein kinase CK2α within their acidic C-terminal domains, increasing the number of plant HMGB variants (*24*, *36*). The phosphorylation enhances the interaction of the acidic tail with the basic N-terminal domain of maize HMGB1 (M. S. Thomsen and K. D. Grasser, unpublished results), modulating functional properties such as the DNA binding selectivity and protein−protein interactions (*24*, *25*). The reason for the variability of chromatin-associated HMGB proteins in plants is unclear, but recent results indicate that there are functional differences between the various HMGB proteins as well as their differently phosphorylated variants (*8*, *23*−*25*, *43*). Therefore, the HMGB variants may be adapted to act as architectural factors in specific nucleoprotein structures.

Here we have characterized the *Arabidopsis* HMGB6 protein as being representative of the novel plant HMGB proteins, which share the typical overall structure of the previously characterized plant HMGB proteins (central HMG-box domain, flanked by basic N-terminal and acidic C-terminal domains). However, they differ in structural details of their genes (e.g., lacking the 5′-UTR intron) from the previously characterized genes encoding HMGB1−HMGB5. Because of its unusually extensive N-terminal domain, HMGB6 is the largest plant HMGB protein identified so far (Table 1). The protein localizes to the cell nucleus, and its HMG-box domain has an α-helical structure. Moreover, HMGB6 binds structure-specifically to DNA, and it can be phosphorylated by protein kinase CK2α. These features indicate that HMGB6 is a member of the plant chromosomal HMGB protein family (*22*). Nevertheless, HMGB6, and the other novel HMGB proteins, display some differences in their amino acid sequences relative to the HMGB1−HMGB5 proteins. Thus, they have residues other than the valine residue conserved among the HMGB1−HMGB5 proteins at the position of the potential secondary DNA intercalating residue. Mutational and structural analyses have demonstrated that the two intercalating residues play a critical role in determining the DNA interactions of the HMGB proteins (*29*−*34*). Another example for differences in the primary structure of the novel HMGB proteins and the *Arabidopsis* HMGB1−HMGB5 proteins is the conserved proline residue at the N-terminal end of the HMG-box domains of the HMGB1−HMGB5 proteins, which is a serine residue (S112) in HMGB6. The proline residue is important for directing the basic N-terminal domain of the yeast HMGB-type NHP6A protein into the major groove of the DNA, while the HMG-box domain is positioned in the minor groove (*31*, *44*). Therefore, it is possible that the novel *Arabidopsis* HMGB proteins, and possibly related proteins of other species, display somewhat different DNA interactions than the HMGB1−HMGB5 proteins, which needs to be addressed by detailed DNA binding studies in the future. The discovery of the novel *Arabidopsis* HMGB proteins in this study, typified by the HMGB6 protein, emphasizes the structural variability of plant chromosomal HMGB proteins. On the basis of previous functional studies, it is likely that the different HMGB variants have specialized functions for assisting in the formation of different nucleoprotein complexes involved in the regulation of various DNA-dependent processes (*22*).

REFERENCES

1. Bustin, M. (1999) *Mol. Cell. Biol. 19*, 5237−5246.
2. Bustin, M. (2001) *Trends Biochem. Sci. 26*, 152−153.
3. Thomas, J. O., and Travers, A. A. (2001) *Trends Biochem. Sci. 26*, 167−174.
4. Agresti, A., and Bianchi, M. E. (2003) *Curr. Opin. Genet. Dev. 13*, 170−178.
5. Travers, A. (2000) *Curr. Opin. Struct. Biol. 10*, 102−109.
6. Zlatanova, J., and van Holde, K. (1998) *FASEB J. 12*, 421−431.
7. Pedersen, T. J., Arwood, L. J., Spiker, S., Guiltinan, M. J., and Thompson, W. F. (1991) *Plant Mol. Biol. 16*, 95−104.
8. Ritt, C., Grimm, R., Fernández, S., Alonso, J. C., and Grasser, K. D. (1998) *Plant J. 14*, 623−631.
9. Spiker, S. (1984) *J. Biol. Chem. 259*, 12007−12013.
10. Webster, C. I., Packman, L. C., Pwee, K.-H., and Gray, J. C. (1997) *Plant J. 11*, 703−715.
11. Wu, Q., Zhang, W., Pwee, K.-H., and Kumar, P. P. (2003) *Arch. Biochem. Biophys. 411*, 105−111.
12. Yamamoto, S., and Minamikawa, T. (1998) *Biochim. Biophys. Acta 1396*, 47−50.
13. Zhang, W., Wu, Q., Pwee, K.-H., and Kini, R. M. (2003) *Arch. Biochem. Biophys. 409*, 357−366.
14. Grasser, K. D. (1998) *Trends Plant Sci. 3*, 260−265.
15. O'Neill, S. D., and Zheng, C. C. (1998) *Plant Mol. Biol. 37*, 235−241.
16. Stemmer, C., Grimm, R., and Grasser, K. D. (1999) *Plant Mol. Biol. 41*, 351−361.
17. Wu, Q., Zhang, W., Pwee, K.-H., and Kumar, P. P. (2003) *Gene 312*, 103−109.
18. Grasser, K. D., Krech, A. B., and Feix, G. (1994) *Plant J. 6*, 351−358.
19. Stemmer, C., Ritt, C., Igloi, G. L., Grimm, R., and Grasser, K. D. (1997) *Eur. J. Biochem. 250*, 646−652.
20. Webster, C. I., Cooper, M. A., Packman, L. C., Williams, D. H., and Gray, J. C. (2000) *Nucleic Acids Res. 28*, 1618−1624.
21. Webster, C. I., Packman, L. C., and Gray, J. C. (2001) *Eur. J. Biochem. 268*, 3154−3162.
22. Grasser, K. D. (2003) *Plant Mol. Biol. 53*, 281−295.
23. Lichota, J., and Grasser, K. D. (2001) *Biochemistry 40*, 7860−7867.
24. Stemmer, C., Schwander, A., Bauw, G., Fojan, P., and Grasser, K. D. (2002) *J. Biol. Chem. 277*, 1092−1098.
25. Krohn, N. M., Yanagisawa, S., and Grasser, K. D. (2002) *J. Biol. Chem. 277*, 32438−32444.
26. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
27. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.
28. Röttgers, K., Krohn, N. M., Lichota, J., Stemmer, C., Merkle, T., and Grasser, K. D. (2000) *Plant J. 23*, 395−405.
29. He, Q., Ohndorf, U.-M., and Lippard, S. J. (2000) *Biochemistry 39*, 14426−14435.
30. Klass, J., Murphy, F. V., Fouts, S., Serenil, M., Changela, A., Siple, J., and Churchill, M. E. A. (2003) *Nucleic Acids Res. 31*, 2852−2864.
31. Masse, J. E., Wong, B., Yen, Y.-M., Allain, F. H. T., Johnson, R. C., and Feignon, J. (2002) *J. Mol. Biol. 323*, 263−284.
32. Murphy, F. V., Sweet, R. M., and Churchill, M. E. A. (1999) *EMBO J. 18*, 6610−6618.
33. Ohndorf, U.-M., Rould, M. A., He, Q., Pabo, C. O., and Lippard, S. J. (1999) *Nature 399*, 708−712.
34. Stros, M., and Muselíková, E. (2000) *J. Biol. Chem. 275*, 35699−35707.
35. Haasen, D., Köhler, C., Neuhaus, G., and Merkle, T. (1999) *Plant J. 20*, 695−706.
36. Stemmer, C., Leeming, D. J., Franssen, L., Grimm, R., and Grasser, K. D. (2003) *Biochemistry 42*, 3503−3508.
37. Litchfield, D. W. (2003) *Biochem. J. 369*, 1−15.
38. Meggio, F., and Pinna, L. A. (2003) *FASEB J. 17*, 349−368.
39. Blom, N., Gammeltoft, S., and Brunak, S. (1999) *J. Mol. Biol. 294*, 1351−1362.
40. Kreegipuu, A., Blom, N., and Brunak, S. (1999) *Nucleic Acids Res. 27*, 237−239.
41. Moehs, C. P., McElwain, E. F., and Spiker, S. (1988) *Plant Mol. Biol. 11*, 507−515.
42. Baxevanis, A. D., and Landsman, D. (1995) *Nucleic Acids Res. 23*, 1604−1613.
43. Stemmer, C., Fernández, S., Lopez, G., Alonso, J. C., and Grasser, K. D. (2002) *Biochemistry 41*, 7763−7770.
44. Yen, Y.-M., Wong, B., and Johnson, R. C. (1998) *J. Biol. Chem. 273*, 4424−4435.

BI035931C