

A novel *Arabidopsis* DNA binding protein contains the conserved motif of HMG-box proteins

Kazuko Yamaguchi-Shinozaki and Kazuo Shinozaki*

Laboratory of Plant Molecular Biology, Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan

Received November 2, 1992; Accepted November 19, 1992

DDJB accession no. D13491

We have isolated cDNA clones from a lambda gt11 library of mRNA prepared from rosette plants of *Arabidopsis thaliana* to identify DNA binding proteins that bind to the promoter region of the desiccation-responsive *rd29A* gene (1, 2, 3). We used a 40 bp conserved sequence (3) in the promoter regions of *rd29A* and *rd29B* genes, which may contain enhancer-like sequence(s), as a probe for the screening. We isolated two independent cDNA clones that bind to the 40-bp sequence. One of the cDNAs encodes a putative protein of 643 amino acids (aa) which contains a highly conserved sequence of high mobility group (HMG) protein which is referred to as ATHMG. The molecular weight of the putative ATHMG protein is 71,547. Yeast nonhistone chromosomal protein NHP6A (4) of 93 aa has the highest similarity with the carboxyterminal region of the ATHMG protein although it differs in size, as shown in Figure 1A. The ATHMG protein contains the HMG box in the carboxyterminal region. In plants, two soybean cDNAs, SB11A (5) and SB16A (6), and one maize cDNA (7) have been reported to contain a HMG-protein like domain (Figure 1B). However, their putative proteins are only 152, 153 and 157 aa long, respectively. Fewer sequence similarities were observed between the HMG-box domains in ATHMG and other plant HMG-like proteins (Figure 1B). The ATHMG protein contains consensus motifs for the phosphorylation site of casein kinase II (Figure 1A) which is located in the upstream region of the HMG box. An acidic region is located upstream from the basic HMG box, while other HMG proteins contain the acidic regions in the carboxyterminal region. The remaining 500-aa amino terminal region of the ATHMG protein does not contain motifs or regions which are homologous with the reported proteins in the protein data base. The function of this long amino terminal region remains to be elucidated.

ACKNOWLEDGEMENTS

This work was supported in part by the Special Coordination Fund of the Science and Technology Agency of the Japanese Government and a Grant-In-Aid from the Ministry of Education, Science and Culture of Japan to K.S. K.Y.-S. was supported by a fellowship from the Science and Technology Agency of Japan.

REFERENCES

1. Yamaguchi-Shinozaki, K., Koizumi, M., Urao, S. and Shinozaki, K. (1992) *Plant Cell Physiol.* **33**, 217–224.
2. Yamaguchi-Shinozaki, K. and Shinozaki, K. (1992) *Mol. Gen. Genet.* in press.

3. Yamaguchi-Shinozaki, K. and Shinozaki, K. (1992) *Plant Physiol.* in press.
4. Kolodrubetz, D. and Burgum, A. (1990) *J. Biol. Chem.* **265**, 3234–3239.
5. Laux, T., Seurinck, J. and Goldberg, R.B. (1991) *Nucleic Acids Res.* **19**, 4768.
6. Laux, T. and Goldberg, R.B. (1991) *Nucleic Acids Res.* **19**, 4769.
7. Grasser, K. and Feix, G. (1991) *Nucleic Acids Res.* **19**, 2573–2577.

A

```

AT  MADGHFLNNISLSGRGGKPNGLLKINSGGQWKKQGGGKAVEVDRSDIVSV  50
    SWTAKVTKSNQLGVKTKDGLYYKFGVFRDQDVPSSLSSFFQSSYGKTPDEKQ  100
    LSVSGRNWGEVDLHGNTLTFLVGSKQAFEVSLADVSTQLQKGNVDVLEF  150
    MLMILLVLMRKTPLMEISFHIPNSNTQFVGDENRPPSQVFNDTIVAMADV  200
    SPGVEDAUVTFESIAILLTPRGRYNVELHLSFLRLQEQANDFKIQYSSVVR  250
    LFLLPKSNQPHTFVVISLDPPIRKGGQTMYPHIVMQFETDTVVESELSISD  300
    ELMNTKFKDKLERSYKGLIHEVFTTVLRWLSGAKITKPGKFRSSQDGFV  350
    KSSLKAEDGVLYPLEKGFFFLPKPPTLILHDEIDYVEFERHAAGGANMHY  400
    FDLLIRLKTDEHLEFRNIQRNEYHNLTYTFISSKGLKIMNLGGAGTADGVA  450
    AVLGDNDDDDAVDPHLTRIRNQAADSEDEDEDVFMGEDDDGGSPDSSG  500
    GDDSDASEGGVGEIKEKSIKKEPKKEASSKGLPPKRKTVAADEGSSKRK  550
                                     *:::
6A  MVTPREPKKR  10'
    KPKKKKDPNAPKRAMSGFMFFSQMERDNIKKEHPGIAFGEVKGVLGDKCV  600
    ::::::::::::::::::::*:::***: * ::::*::::***** **
    TTRKKKDPNAPKRALSAYMFFANENRDIRSENPDITFGQVGGKLGKWK  60'
    KCLLMIKSHMKPRLQVDKQRYKDEISDYKPNQPMNVDSGNDSSG  643
    * : : : * : : : * : : : * : : : * : : : * : : : * : : :
    ALTPEEKQPYEAKAQADKKRYESEKELYNATLA  93'
  
```

B

Arabidopsis	HMG	<u>KKK</u> DENAPKRAMSGFMEE
yeast	NH6PA	<u>KKK</u> DENAPKRALSAYMEE
human	HMG1	<u>KKK</u> DENAPKRPPSAFFLE
maize	HMG	<u>AGK</u> DENKPKRAPSAFFVF
soybean	SB11	<u>AAK</u> DENKPKRPPSAFFVF
soybean	16A	<u>PPK</u> DPNAPKPSPKAKATP

Figure 1. (A) Comparison of an amino acid sequence of the putative ATHMG (AT) protein with that of yeast nonhistone nuclear protein NH6A (6A). Numbers and dashed numbers indicate amino acids of ATHMG and NH6A proteins, respectively. A possible phosphorylation site by casein kinase II is underlined. (B) Comparison of the conserved regions of HMG proteins. Identical amino acid residues are shadowed.

* To whom correspondence should be addressed