The EMBO Journal vol.10 no.7 pp.1787-1791, 1991

A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif

Ida Ruberti, Giovanna Sessa, Sabrina Lucchetti[1] and Giorgio Morelli[1,2]

Centro di studio per gli Acidi Nucleici, c/o Dipartimento di Genetica e Biologia Molecolare, Università di Roma La Sapienza, P.le Aldo Moro 5 00185, Rome and [1]Unità di Nutrizione Sperimentale, Istituto Nazionale della Nutrizione, Via Ardeatina 546 00178 Rome, Italy

[2]Corresponding author

Communicated by W.J.Gehring

The homeobox, a 183 bp DNA sequence element, was originally identified as a region of sequence similarity between many *Drosophila* homeotic genes. The homeobox codes for a DNA-binding motif known as the homeodomain. Homeobox genes have been found in many animal species, including sea urchins, nematodes, frogs, mice and humans. To isolate homeobox-containing sequences from the plant *Arabidopsis thaliana*, a cDNA library was screened with a highly degenerate oligonucleotide corresponding to a conserved eight amino acid sequence from the helix-3 region of the homeodomain. Using this strategy two cDNA clones sharing homeobox-related sequences were identified. Interestingly, both of the cDNAs also contain a second element that potentially codes for a leucine zipper motif which is located immediately 3′ to the homeobox. The close proximity of these two domains suggests that the homeodomain-leucine zipper motif could, via dimerization of the leucine zippers, recognize dyad-symmetrical DNA sequences.

Key words: A.thaliana/HD-Zip motif/homeodomain/leucine zipper

Introduction

The homeobox (HB), a 183 bp DNA sequence element, was originally identified as a region of sequence similarity shared by several genes involved in the control of *Drosophila* development (Gehring, 1987). The HB sequence encodes a 61 amino acid sequence, known as the homeodomain (HD). Several lines of evidence indicated that HD proteins are transcription factors in which the homeodomain is responsible for sequence-specific recognition of DNA (Affolter *et al.*, 1990; Hayashi and Scott, 1990, and references therein). HB genes have now been identified in many animal species, including sea urchins, nematodes, frogs, mice and humans, via a low-stringency hybridization with HB-containing probes (Scott *et al.*, 1989). Related sequences have also been found in yeast (Shepard *et al.*, 1984) and in the fungus *Ustilago maydis* (Schulz *et al.*, 1990).

Surprisingly, the homeodomain has not yet been detected in plants, despite the fact that a substantial fraction of plant regulatory factors contains structural motifs such as helix-loop-helix (Ludwig and Wessler, 1990), leucine zipper (Hartings *et al.*, 1989; Katagiri *et al.*, 1989; Singh *et al.*, 1990) or zinc finger motifs (Lam *et al.*, 1990), homologous to those described in yeast and animal transcription factors (Glover, 1989; Johnson and McKnight, 1989).

In a direct approach to identify sequences coding for HD proteins from the plant *Arabidopsis thaliana*, we took advantage of a strategy which has been successfully used to isolate several HB genes from the nematode *Caenorhabditis elegans* (Burglin *et al.*, 1989). In this way two HB-containing sequences were isolated from an *A.thaliana* cDNA library. DNA sequence analysis revealed that the two cDNA clones are also related by virtue of a second conserved element that potentially codes for a leucine zipper motif (Landschulz *et al.*, 1988). Furthermore, both sequences exhibit an identical alignment between the homeodomain and the leucine zipper motif, with the former immediately N-terminal to the latter. These observations led us to propose the existence of a novel class of plant proteins constructed from a combination of the homeodomain (DNA contacting region) and the leucine zipper (dimerization motif), each of which was previously described, but only in separate classes of DNA-binding proteins (Glover, 1989; Johnson and McKnight, 1989).

Results

Isolation of cDNA encoding homeodomain-containing proteins

To isolate sequences coding for HD proteins from *A.thaliana*, we took advantage of the strategy developed by Burglin *et al.* (1989). They designed a degenerate oligonucleotide, HB-1 (Figure 1A), corresponding to a conserved eight amino acid sequence from the helix-3 region of the homeodomain (Burglin, 1988), able to detect a large set of HB genes. Using conditions resulting in strong hybridization to DNA fragments with two or fewer mismatches (Burglin *et al.*, 1989), more than 15 bands of varying intensities could be distinguished on an *A.thaliana* genomic Southern blot probed with HB-1 (Figure 1B).

We screened 300 000 clones from an *A.thaliana* cDNA λgt11 library (Clontech) with HB-1 and selected four independent λ clones. Two of these cDNA clones, designated Athb-1A and Athb-2A (Athb for *A.thaliana* homeobox), were further characterized.

Nucleotide and predicted amino acid sequences of Athb-1

The homeodomain is a 61-residue amino acid sequence which is highly conserved among the known HB genes (Scott *et al.*, 1989). In all of them, four amino acid residues located within helices 3 and 4 are invariant and eight residues distributed throughout the 61-residue sequence are highly conserved (Gehring *et al.*, 1990). Sequence analysis of the Athb-1A cDNA clone revealed a 722 bp insert with a poly(dA) stretch located at one end (Figure 2A). The predicted amino acid sequence contained all the invariant

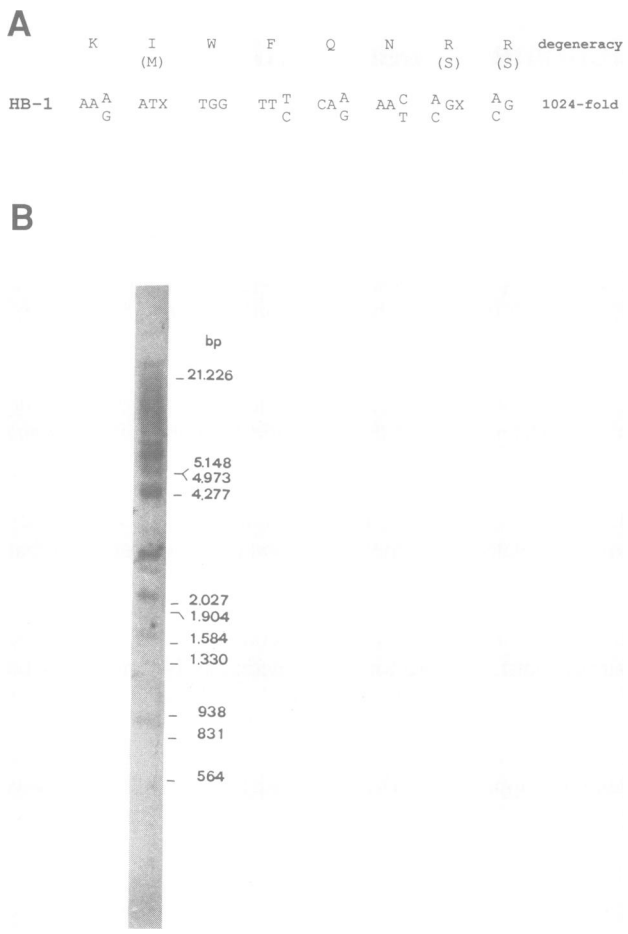


Fig. 1. Southern blot of *A.thaliana* DNA hybridized with HB-1. (**A**) Sequence of the degenerate oligonucleotide HB-1 (Burglin *et al.*, 1989) and its predicted translation. The single letter code is used to designate amino acid residues. Amino acids in brackets have not been found at that position in the homeodomain, but were included as a byproduct of the degenerate codons. (**B**) Southern blot of *A.thaliana* genomic DNA (6 μg) digested with *Eco*RI and probed with HB-1. The numbers indicate the size, in bp, of DNA mol. wt markers.

amino acid residues of the homeodomain. Using the Athb-1A insert as a probe, two additional overlapping clones (Athb-1B and Athb-1C) were subsequently isolated by screening another 300 000 plaques of the *A.thaliana* cDNA library. Their location relative to Athb-1A was determined by DNA sequencing. Athb-1C cDNA is 1302 bp long and overlaps completely the Athb-1A clone and most of the Athb-1B clone (Figure 2A). Analysis of the predicted amino acid sequence of Athb-1C cDNA (Figure 2B) revealed an 816 bp open reading frame starting with an ATG at position 331 and ending with a stop codon at nucleotide 1147. Athb-1C encodes for a protein of 272 amino acids (aas) in length, with a calculated molecular mass of 31 kd (Figure 2B). The putative polypeptide is highly hydrophilic and has several interesting features. It contains two acidic regions (aspartic or glutamic acid), one in the N-terminal part of the polypeptide, between aas 39 and 63 and the other in the C-terminal part, between aas 210 and 243 (Figure 2B). Clusters of aspartate or glutamate residues have been described in many transcription factors, and there is strong evidence that acidic regions function in the activation of RNA polymerase II transcription initiation (Ptashne, 1988). A region of the Athb-1 protein from aa 64 to 124 (Figure 2B) contains the invariant as well as most of the highly conserved amino acid residues of the homeodomain, as shown by the alignment with the *Antp* homeodomain (Gehring *et al.*, 1990) (Figure 4 and see below). Finally, exactly at the end of the homeodomain, there are a series of leucine residues separated by seven aa residues (aas 125–160 in Figure 2B). A similar configuration of leucine residues in the transcription factor C/EBP has been suggested to form an amphipathic α-helical structure, where the leucine residues are arranged along one side of the helix. Two such helices are thought to interact by a 'leucine zipper' mechanism generating a dimeric protein complex (Landschulz *et al.*, 1988).

Interestingly, DNA sequence analysis also revealed that the 330 bp-long 5′-untranslated leader sequence of Athb-1C contains two short open reading frames, one of 28 amino acids in-frame with the ATG at position 331 and one of 8 amino acids out of frame. Short open reading frames have been described in the 5′-untranslated leader sequence of plant regulatory factor mRNAs (Hartings *et al.*, 1989; Singh *et al.*, 1990) and found to be responsible for translational regulation of GCN4 expression in yeast (Hinnebusch, 1984 and 1988). The 153 bp-long 3′-untranslated region contains a putative poly(A) addition sequence, AATAAT, 63 bp upstream of the polydA sequence (Figure 2B).

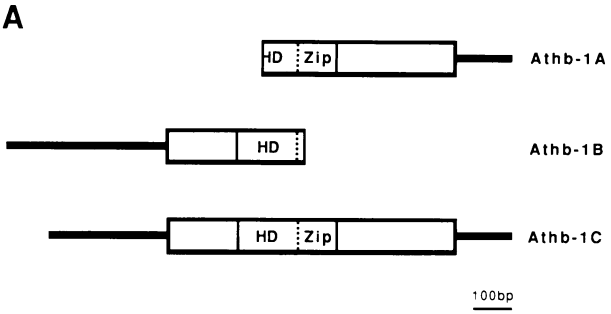
Northern blot analysis of *A.thaliana* leaf poly(A)$^+$ RNA revealed that Athb-1 mRNA appears to be a single species of ~1450 bases (Figure 3). Thus, the 1302 bp Athb-1C clone represents most of the Athb-1 mRNA sequence.

Comparison of the predicted amino acid sequence of the Athb-1 and Athb-2 homeodomain regions

DNA sequence analysis of the region flanking the probe-hybridizing region of the Athb-2A cDNA clone also revealed the presence of a complete homeodomain (Figure 4), which shows a 52% homology (32 out of 61 amino acid residues) to the Athb-1 homeodomain.

The homeodomains of Athb-1 and Athb-2 diverge from previously characterized HD sequences: at most they share only 41% homology (25 out of 61 amino acid residues) with any of the previously identified genes (Scott *et al.*, 1989). However, Athb-1 and Athb-2 homeodomains contain the four invariant aa residues and seven out of the eight highly conserved aa residues of the homeodomain consensus sequence (Gehring *et al.*, 1990). The invariant residues are tryptophan, phenylalanine, asparagine and arginine at positions 49, 50, 52 and 54, respectively. The highly conserved aas are located in positions 13 (glutamine), 17 (leucine), 21 (phenylalanine), 41 (leucine), 46 (valine), 56 (arginine) and 58 (lysine). Furthermore, Athb-1 and Athb-2 contain, at position 6, a conservative substitution (lysine versus arginine) (Figure 4).

Athb-1 and Athb-2 are also related by virtue of a second conserved sequence element that codes for a potential leucine zipper motif (Figure 5). The leucine zipper is a special case of the coiled-coil motif with a series of leucine residues spaced by exactly seven aa residues. The start of each heptad repeat is designated by the letter **a**, such that the hydrophobic residues occur in positions **a** and **d**, with leucine residues usually found at position **d** (Glover, 1989). This distribution of residues permits the theoretical formation of an amphipathic α-helix, which can pair by interaction of the aligned hydrophobic residues (Landschulz *et al.*, 1988).



B

```
Athb-1C  TGT CTC TCT CTC TGG CTC TCT TTC TCT TAA CGT GAT CAT AAC GTG ATT
                  9          18          27          36          45

CGA AAA TTG GAT ATA GAT AGG TTT CTT GTT GGA TCT TGA TCC CTC TGG AAA AGG
        57          66          75          84          93         102

AGG GGA GAA TAG CAG TTC ATG ATG GGA TTT TGT ATC TGC CCG TTG GAG TCA CCT
       111         120         129         138         147         156

GCG AGA TTA CTA TGG AGT ACA AGC TTC TTC CGC CAT AAG ATC ATG ATC TTC TAA
       165         174         183         192         201         210

TCC TTC CTA CTT CTT CCC ATC TTT TTA ATC ATC TTC TCG CTA TCT CTG CTT CCT
       219         228         237         246         255         264

CTT TCT CTC TGT TTC CTC TTT CTC AGA ACT CAG AAG TAG TTG TTG TTT TAT TTC
       273         282         291         300         309         318

          1 - Met Glu Ser Asn Ser Phe Phe Phe Asp Pro Ser Ala Ser His - 14
TGT TGA TCA AAA ATG GAA TCC AAT TCG TTT TTC TTC GAT CCA TCT GCT TCA CAC
       327         336         345         354         363         372

Gly Asn Ser Met Phe Phe Leu Gly Asn Leu Asn Pro Val Val Gln Gly Gly Gly - 32
GGC AAC AGC ATG TTC TTC CTT GGG AAT CTC AAT CCC GTC GTC CAA GGA GGA GGA
       381         390         399         408         417         426

Ala Arg Ser Met Met Asn Met Glu Glu Thr Ser Lys Arg Arg Pro Phe Phe Ser - 50
GCA AGA TCG ATG ATG AAC ATG GAG GAA ACT TCG AAG CGA AGG CCC TTC TTT AGC
       435         444         453         462         471         480

Ser Pro Glu Asp Leu Tyr Asp Asp Asp Phe Tyr Asp Asp Gln Leu Pro Glu Lys - 68
TCC CCT GAG GAT CTC TAC GAC GAT GAC TTT TAC GAC GAC CAG TTG CCT GAA AAG
       489         498         507         516         525         534

Lys Arg Arg Leu Thr Thr Glu Gln Val His Leu Leu Glu Lys Ser Phe Glu Thr - 86
AAG CGT CGC CTC ACT ACC GAA CAA GTG CAT CTG CTG GAG AAA AGC TTC GAG ACA
       543         552         561         570         579         588

Glu Asn Lys Leu Glu Pro Glu Arg Lys Thr Gln Leu Ala Lys Lys Leu Gly Leu -104
GAG AAC AAG CTA GAG CCT GAA CGC AAG ACT CAG CTT GCC AAG AAG CTT GGT CTA
       597         606         615         624         633         642

Gln Pro Arg Gln Val Ala Val Trp Phe Gln Asn Arg Arg Ala Arg Trp Lys Thr -122
CAG CCA AGG CAA GTG GCT GTC TGG TTT CAG AAT CGC CGA GCT CGT TGG AAA ACA
       651         660         669         678         687         696

Lys Gln Leu Glu Arg Asp Tyr Asp Leu Leu Lys Ser Thr Tyr Asp Gln Leu Leu -140
AAA CAG CTT GAG AGA GAC TAC GAT CTT CTC AAG TCC ACT TAC GAC CAA CTT CTT
       705         714         723         732         741         750

Ser Asn Tyr Asp Ser Ile Val Met Asp Asn Asp Lys Leu Arg Ser Glu Val Thr -158
TCT AAC TAC GAC TCC ATC GTC ATG GAC AAC GAT AAG CTC AGA TCC GAG GTT ACT
       759         768         777         786         795         804

Ser Leu Thr Glu Lys Leu Gln Gly Lys Gln Glu Thr Ala Asn Glu Pro Pro Gly -176
TCC CTG ACC GAA AAG CTT CAG GGC AAA CAA GAG ACA GCT AAT GAA CCA CCT GGT
       813         822         831         840         849         858

Gln Val Pro Glu Pro Asn Gln Leu Asp Pro Val Tyr Ile Asn Ala Ala Ala Ile -194
CAA GTG CCC GAA CCA AAC CAA CTT GAT CCG GTT TAC ATT AAT GCG GCA GCA ATC
       867         876         885         894         903         912

Lys Thr Glu Asp Arg Leu Ser Ser Gly Ser Val Gly Ser Ala Val Leu Asp Asp -212
AAA ACC GAG GAC CGG TTA AGT TCA GGG AGC GTT GGG AGC GCG GTA CTA GAC GAC
       921         930         939         948         957         966

Asp Ala Pro Gln Leu Leu Asp Ser Cys Asp Ser Tyr Phe Pro Ser Ile Val Pro -230
GAC GCA CCT CAA CTA CTA GAC AGC TGT GAC TCT TAC TTC CCA AGC ATC GTA CCC
       975         984         993        1002        1011        1020

Ile Gln Asp Asn Ser Asn Ala Ser Asp His Asp Asn Asp Arg Ser Cys Phe Ala -248
ATC CAA GAC AAC AGC AAC GCC AGT GAT CAT GAC AAT GAC CGG AGC TGT TTC GCC
      1029        1038        1047        1056        1065        1074

Asp Val Phe Val Pro Thr Thr Ser Pro Ser His Asp His His Gly Glu Ser Leu -266
GAC GTC TTT GTG CCC ACC ACT TCA CCG TCG CAC GAT CAT CAC GGT GAA TCA TTG
      1083        1092        1101        1110        1119        1128

Ala Phe Trp Gly Trp Pro TER
GCT TTC TGG GGA TGG CCT TAG AAA ACC ACT CTG ATA ATA AAT GTG TGT TTA TTT
      1137        1146        1155        1164        1173        1182

AAG TTC AAG AGT CAT CTT CTT GTT GTT TCC ATG TTG ACG ATA ATT GTT GAC TCG
      1191        1200        1209        1218        1227        1236

TGG AAT AAT TCC GCT GTT CAA CGG TAT TTT TAT CAG TTG CAT TAT ATG CTT TTA
      1245        1254        1263        1272        1281        1290

TGA AAA AAA AAA
      1299
```

Athb-1 and Athb-2 zipper motifs are formed by 6 and 5 heptad repeats, respectively. Athb-1 contains an isoleucine at position $\mathbf{d}_4$ and Athb-2 contains a threonine residue at position $\mathbf{d}_1$. It should be noted that Athb-1 has the amino acid tyrosine repeated three times at position $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$, always followed by an aspartate residue. Consistent with the hypothesis of α-helix formation, no helix-disrupting proline or glycine residues are present in these heptad repeats. In addition, like the first example of a leucine zipper protein, C/EBP (Landschulz *et al.*, 1988), the potential α-helical region in Athb-1 and Athb-2 possesses a high density of paired charged residues that could stabilize the structure. Interestingly, in both the Athb proteins, the position of the $\mathbf{a}_1$ hydrophobic residue corresponds to the 59th position of the homeodomain, which is usually occupied by a positively charged amino acid (Scott *et al.*, 1989). We aligned the homeodomain-leucine zipper regions with representative leucine zipper proteins, which show the contiguous basic region-leucine zipper (b-ZIP) architecture (Vinson *et al.*, 1989). This revealed that the putative recognition helices of the Athb-1 and Athb-2 homeodomains are located, with respect to the leucine zipper motif, in a position equivalent to that occupied by the DNA contacting region (the basic region) in the b-ZIP proteins (Figure 5).

Discussion

We report the first isolation of homeobox-containing sequences from the plant *A.thaliana*. A cDNA library was screened with the degenerate oligonucleotide HB-1 (Burglin *et al.*, 1989), which corresponds to a conserved amino acid sequence from the helix-3 region of the homeodomain (Burglin, 1988), and two clones, termed Athb-1 and Athb-2, were selected.

A domain of both Athb-1 and Athb-2 has homology to the homeodomain sequence first identified in developmental regulatory factors of *Drosophila* (Gehring, 1987) and subsequently found in vertebrates and worms (Scott *et al.*, 1989). Recently, homeodomain proteins have been grouped into different families of related sequences (Scott *et al.*, 1989). Sequence comparisons of the putative Athb-1 and Athb-2 products with a representative member of each homeodomain family revealed that they cannot be assigned to any of the previously identified classes. Although Athb-1 and Athb-2 have to be included in the group of the unclassified homeodomains (Scott *et al.*, 1989), they exhibit a strong homology with the helix-3 region and the highly conserved residues of most of the HD sequences (Gehring *et al.*, 1990). The conservation of these residues suggests that Athb-1 and Athb-2 homeodomains may adopt three-dimensional structures similar to those of others (Qian *et al.*, 1989; Kissinger *et al.*, 1990). Nuclear magnetic resonance studies of the *Antp*

Fig. 2. DNA and amino acid sequence of Athb-1. (**A**) Schematic representation of the Athb-1 cDNA clones. The location of the three cDNA isolates relative to each other is shown. The boxed region of Athb-1C denotes the open reading frame starting from the ATG at nucleotide 331 to a TAG at nucleotide 1147. The location of the homeodomain (HD) and the leucine zipper (Zip) of the predicted Athb-1 protein is indicated. (**B**) The entire nucleotide sequence and the deduced amino acid sequence of the Athb-1C cDNA. Chequered bars delimit the glutamate- and aspartate-rich regions. The amino acid sequence enclosed by an open bar indicates the homeodomain. Periodic leucines and a single isoleucine are boxed.

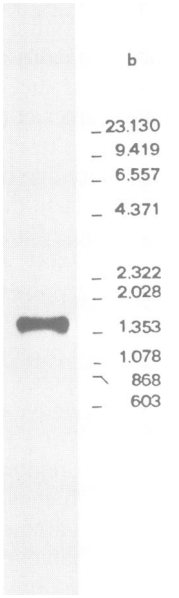


Fig. 3. Northern blot of Athb-1 mRNA. A 485 bp *Hin*dIII−*Eco*RI 3′ end fragment of Athb-1C was randomly labeled with ^{32}P and used to probe 1 μg of poly(A)$^+$ RNA from leaves of 6 week old *Arabidopsis* plants. The numbers indicate the size, in bases, of denaturated DNA molecular weight markers.

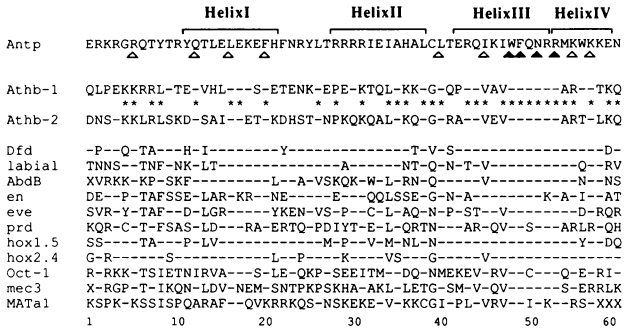


Fig. 4. Amino acid comparison of Athb-1 and Athb-2 with other homeodomains. Single letter code designates amino acid residues. The amino acid sequence of the *Antp* homeodomain and the location of the four helices are indicated. The invariant amino acids are marked by black triangles, the highly conserved ones by open triangles (Gehring *et al.*, 1990). Dashes indicate identity of the putative Athb-1 and Athb-2 homeodomains and of selected members of each HB family (Scott *et al.*, 1989) to the *Antp* homeodomain. Identity between the two Athb homeodomains are indicated by asterisks. *Antp* (McGinnis *et al.*, 1984a), *Dfd* (Laughon *et al.*, 1985), *labial* (Hoey *et al.*, 1986), *AbdB* (Regulski *et al.*, 1985), *en* (Poole *et al.*, 1985), *eve* (Harding *et al.*, 1986), *prd* (Frigerio *et al.*, 1986), hox 1.5 (McGinnis *et al.*, 1984b), hox 2.4 (Hart *et al.*, 1987), Oct-1 (Scheidereit *et al.*, 1988), *mec3* (Carrasco *et al.*, 1984), MATa1 (Nasmyth *et al.*, 1981).

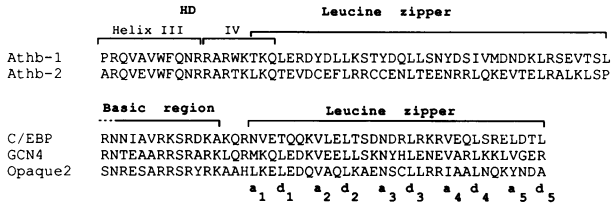


Fig. 5. Amino acid sequence comparison of the Athb-1 and Athb-2 leucine zipper motifs with selected b-Zip proteins. Athb-1 and Athb-2 sequences, designated according to the single letter code, were aligned to the leucine repeat of the b-Zip proteins. The first residue of each heptad repeat is designed by the letter **a**, with leucine residues at position **d**. C/EBP (Landschulz *et al.*, 1988), GCN4 (Hinnebusch, 1984), *opaque 2* (Hartings *et al.*, 1989).

homeodomain (Qian *et al.*, 1989; Otting *et al.*, 1990) and an X-ray crystallographic study of the *engrailed* (*en*) homeodomain (Kissinger *et al.*, 1990) demonstrated that they contain a helix−turn−helix motif. The presence of this motif in the homeodomain was previously suggested on the basis of sequence comparisons between HD proteins and prokaryotic regulatory proteins (Laughon and Scott, 1984).

Crystallographic studies revealed that the monomeric form of the *en* homeodomain bound to its target DNA can provide only a modest amount of sequence specificity. Due to the paucity of sequence-specific contacts additional factors might be necessary to attain the requisite specificity (Kissinger *et al.*, 1990). Some cases where the DNA binding specificity of homeodomains is enhanced either by sequences outside the domain (Hoey *et al.*, 1988) or by combinatorial interaction between a HD protein and other transcription factors (Keleher *et al.*, 1988) or by dimerization of HD proteins (Miller *et al.*, 1985; Nicosia *et al.*, 1990) have been described.

A potential dimerization domain is encoded by Athb-1 and Athb-2 in the form of a leucine zipper motif. These Athb-1 and Athb-2 motifs are formed by 6 and 5 heptad repeats, respectively. Each repeat contains a leucine at position **d**, with the exception of an isoleucine ($\mathbf{d_4}$) in Athb-1 and a threonine ($\mathbf{d_1}$) in Athb-2 (Figure 5). A threonine in position $\mathbf{d_1}$ has been found in the leucine zipper of C/EBP (Landschulz *et al.*, 1988). Furthermore, mutagenic analysis of the GCN4 leucine zipper has shown that replacement of the first leucine ($\mathbf{d_1}$) with a threonine, or the fourth leucine ($\mathbf{d_4}$) with an isoleucine results in functional dimerization domains (Hu *et al.*, 1990).

Alignment of the Athb homeodomain-leucine zipper regions with b-ZIP proteins revealed that the putative recognition helices of the Athb-1 and Athb-2 homeodomains are located, with respect to the leucine zipper motif, in a position equivalent to that occupied by the DNA contacting region in the b-ZIP proteins. Therefore we propose that proteins with a contiguous homeodomain-leucine zipper architecture should be referred to as HD-Zip proteins.

It has been shown that the helix-3 of the homeodomain has the ability to recognize specific DNA sequences (Affolter *et al.*, 1990; Hayashi and Scott, 1990) and by inference, the putative recognition helix of the Athb proteins may have the same function in the plant cell. In addition, the Athb leucine zipper motifs have all the structural features to result in functional dimerization domains. The exact spatial register between the homeodomain and the leucine zipper motif in the two Athb proteins is similar to that observed between the DNA binding and the dimerization domains in the b-ZIP proteins (Vinson *et al.*, 1989). Therefore, in analogy with the model proposed for the b-Zip proteins (Vinson *et al.*, 1989), we suggest that the HD-ZIP proteins might use the dimerization domain to closely juxtapose a pair of DNA contacting surfaces, each of which fits into half of a dyad-symmetric recognition sequence.

Materials and methods

Isolation of cDNA encoding homeodomain-containing proteins

An *A.thaliana* cDNA library in λgt11 (Clontech) was screened with the oligonucleotide HB-1 (Burglin *et al.*, 1989). The oligonucleotide was 5′ end-labelled with T4 kinase according to Ausubel *et al.* (1987) using 3000 Ci mmol^{-1} [γ-^{32}P]ATP (Amersham). Filters were prehybridized in 6× SSC, 5× Denhardt's solution, 0.05% Na pyrophosphate, 1% SDS and 100 mg ml^{-1} herring sperm DNA for 6 h at 42°C. Approximately 5 × 10^6 c.p.m. (5 ng) of labelled oligonucleotide per ml of hybridization solution were hybridized for 24 h at 42°C. Filters were washed exactly as described by Burglin *et al.* (1989).

DNA sequencing

HB-1 hybridizing *Eco*RI fragments were subcloned into Bluescript SKII (Stratagene) or pUC9 using standard techniques (Maniatis *et al.*, 1982). Sequencing was carried out on double-stranded DNA using Sequenase 2 (USB) according to manufacturer's instructions. Initial sequences were obtained using either universal and reverse primers or the degenerate oligonucleotide (HB-1). Sequences were then extended by synthesizing specific primers. Oligonucleotides were synthesized on a Beckman DNA-SM synthesizer and purified according to manufacturer's instructions.

Southern blot analysis

Restriction digests, gel electrophoresis and Southern transfer onto Hybond-N membrane (Amersham) were done as described (Maniatis *et al.*, 1982). Hybridization with HB-1 probe was as described above.

Northern blot analysis

Northern blot analysis of *A.thaliana* RNA was performed essentially as described by Thomas (1980), except that the probe was prepared by the random priming method (Feinberg and Vogelstein, 1983).

Computer analysis

Comparative sequence analysis were performed with programs from the University of Wisconsin Genetics Computer Group (Devereux *et al.*, 1984).

Acknowledgements

We wish to thank A.Gasch for gift of poly (A)$^+$ RNA; W.J.Gehring for helpful suggestions; E.Boncinelli, N.H.Chua, G.Coruzzi and C.Desplan for discussion and comments; H.P.Saluz, A.Wallace and K.Wiebauer for critical reading of this manuscript. We would also like to thank Angelo Di Francesco, Roberto Gargamelli and Maurizio Di Felice for skilled technical assistance. G.S. is supported by a fellowship of the Fondazione Sigma-Tau. This work was supported, in part, by the Fondazione Istituto Pasteur-Fondazione Cenci Bolognetti, Università di Roma La Sapienza (I.R.), by Progetto Finalizzato Biotecnologie e Biostrumentazione, C.N.R. (I.R.) and by Piano Nazionale Svilluppo di Tecnologie Avanzate Applicate alle Piante, Ministero Agricoltura e Foreste (G.M.).

References

Affolter,M., Schier,A. and Gehring,W.J. (1990) *Current Opin. Cell Biol.*, **2**, 485–495.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.C., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, New York.

Burglin,T.R. (1988) *Cell*, **53**, 339–340.

Burglin,T.R., Finney,M., Coulson,A. and Ruvkun,G. (1989) *Nature*, **341**, 239–243.

Carrasco,A.E., McGinnis,W., Gehring,W.J. and De Robertis,E.M. (1984) *Cell*, **37**, 409–414.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.

Frigerio,G., Burri,M., Bopp,D., Baumgartner,S. and Noll,M. (1986) *Cell*, **47**, 735–746.

Gehring,W.J. (1987) *Science*, **236**, 1245–1252.

Gehring,W.J., Muller,M., Affolter,M., Percival-Smith,A., Billeter,M., Qian,Y.Q., Otting,G. and Wuthrich,K. (1990) *Trends Genet.*, **6**, 323–329.

Glover,C.V.C., III (1989) *Plant Mol. Biol. Reporter*, **7** (3), 183–208.

Harding,K., Rushlow,C., Doyle,H.J., Hoey,T. and Levine,M. (1986) *Science*, **233**, 953–959.

Hart,C.P., Fainsod,A. and Ruddle,F.H. (1987) *Genomics*, **1**, 182–195.

Hartings,H., Maddaloni,M., Lazzaroni,N., Di Fonzo,N., Motto,M., Salamini,F. and Thompson,R. (1989) *EMBO J.*, **8**, 2795–2801.

Hayashi,S. and Scott,M.P. (1990) *Cell*, **63**, 883–894.

Hinnebusch,A.G. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6442–6446.

Hinnebusch,A.G. (1988) *Microbiol. Rev.*, **52**, 248–273.

Hoey,T., Doyle,H.J., Harding,K., Wedeen,C. and Levine,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4809–4813.

Hoey,T., Warrior,R., Manak,J. and Levine,M. (1988) *Mol. Cell. Biol.*, **8**, 4598–4607.

Hu,J.C., O'Shea,E.K., Kim,P.S. and Sauer,R.T. (1990) *Science*, **250**, 1400–1403.

Johnson,P.F. and McKnight,S.L. (1989) *Annu. Rev. Biochem.*, **58**, 799–839.

Katagiri,F., Lam,E. and Chua,N.H. (1989) *Nature*, **340**, 727–730.

Keleher,C.A., Goutte,C. and Johnson,A.D. (1988) *Cell*, **53**, 927–936.

Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.

Lam,E., Kano-Murakami,Y., Gilmartin,P., Niner,B. and Chua,N.H. (1990) *Plant Cell*, **2**, 857–866.

Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.

Laughon,A. and Scott,M.P. (1984) *Nature*, **310**, 25–31.

Laughon,A., Carroll,S.B., Storfer,F.A., Riley,P.D. and Scott,M.P. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 253–262.

Ludwig,S.R. and Wessler,S.R. (1990) *Cell*, **62**, 849–851.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

McGinnis,W., Levine,M., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984a) *Nature*, **308**, 478–483.

McGinnis,W., Hart,C.P., Gehring,W.J. and Ruddle,F.H. (1984b) *Cell*, **38**, 675–680.

Miller,A.M., MacKay,V.L. and Nasmyth,K.A. (1985) *Nature*, **314**, 598–603.

Nasmyth,K.A., Tatchell,K., Hall,B.D., Astell,C. and Smith,M. (1981) *Nature*, **289**, 244–250.

Nicosia,A., Monaci,P., Tomei,L., De Francesco,R., Nuzzo,M., Stunnenberg,H. and Cortese,R. (1990) *Cell*, **61**, 1225–1236.

Otting,G., Qian,Y.Q., Billeter,M., Muller,M., Affolter,M., Gehring,W.J. and Wuthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.

Poole,S.J., Kauvar,L., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.

Ptashne,M. (1988) *Nature*, **335**, 683–689.

Qian,Y.Q., Billeter,M., Otting,G., Muller,M., Gehring,W.J. and Wuthrich,K. (1989) *Cell*, **59**, 573–580.

Regulski,M., Harding,K., Kostriken,R., Karch,F., Levine,M. and McGinnis,W. (1985) *Cell*, **43**, 71–80.

Scheidereit,C., Cromlish,J.A., Gerster,T., Kawakami,K., Balmaceda,C.-G., Currie,R.A. and Roeder,R.G. (1988) *Nature*, **336**, 551–558.

Schulz,B., Banuett,F., Dahl,M., Schlesinger,R., Schafer,W., Martin,T., Herskowitz,I. and Kahmann,R. (1990) *Cell*, **60**, 295–306.

Scott,M.P., Tamkun,J.W. and Hartzell,G.W., III (1989) *BBA Rev. Cancer*, **989**, 25–48.

Shepard,J.C., McGinnis,W., Carrasco,A.E., De Robertis,E.M. and Gehring,W.J. (1984) *Nature*, **310**, 70–71.

Singh,K., Dennis,E.S., Ellis,J.G., Llewellyn,D.J., Tokuhisa,J.G., Wahleithner,J.A. and Peacock,W.J. (1990) *Plant Cell*, **2**, 891–903.

Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5205.

Vinson,C.R., Sigler,P.B. and McKnight,S.L. (1989) *Science*, **246**, 911–916.

Received on March 6, 1991

Note added in proof

While this article was in press, Vollbrecht *et al.*, (1991) *Nature*, **350**, 241–243, reported that the *knotted-1* gene of maize contains a homeobox related sequence. The sequence data reported here have been submitted to the EMBL/GenBank/DDBJ databases and are available under the accession number X58821.