Biol. Chem., Vol. 387, pp. 31–40, January 2006 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BC.2006.006

Interaction of the BELL-like protein ATH1 with DNA: role of homeodomain residue 54 in specifying the different binding properties of BELL and KNOX proteins

Ivana L. Viola and Daniel H. Gonzalez

Cátedra de Biología Celular y Molecular, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, CC 242 Paraje El Pozo, 3000 Santa Fe, Argentina

*Corresponding author
e-mail: dhgonza@fbcb.unl.edu.ar

Abstract

We have studied the interaction of the BELL-like *Arabidopsis* homeodomain protein ATH1 with DNA. Analysis of oligonucleotides selected by the ATH1 homeodomain from a random mixture suggests that ATH1 preferentially binds the sequence TGACAGGT. Single nucleotide replacements at positions 2 or 3 of this sequence abolish binding, while changes at position 4 are more tolerated. Changes outside this core differentially affect binding, depending on the position. Hydroxyl radical footprinting and missing nucleoside experiments showed that ATH1 interacts with a 7-bp region of the strand carrying the GAC core. On the other strand, protection was observed over a 7-bp region, comprising one additional nucleotide complementary to T in position 1. A comparative analysis of the binding preferences of the homeodomains of ATH1 and STM (a KNOX homeodomain protein) indicated that they bind similar sequences, but with differences in affinity and specificity. The decreased affinity displayed by the ATH1 homeodomain correlates with the presence of valine (instead of lysine as in STM) at position 54. This difference also explains the decreased and increased selectivities, respectively, at positions 4 and 5. Our results point to an essential role of residue 54 in determining the different binding properties of BELL and KNOX homeodomains.

Keywords: binding-site selection; DNA-binding specificity; footprinting; plant homeodomain protein.

Introduction

Homeobox genes encode transcription factors usually involved in the regulation of developmental processes (Gehring, 1987). These genes contain a region coding for a homeodomain, a 60-amino-acid protein motif that folds into a characteristic three-helix structure that is able to specifically interact with DNA (Qian et al., 1989, 1994; Gehring et al., 1994a,b; Tsao et al., 1995). In plants, the first homeobox was identified in *Knotted1* (*kn1*), a maize gene involved in meristem development (Vollbrecht et al., 1995). Additional *kn1*-like genes (termed *knox* genes) have been isolated from maize and other monocot and dicot species (Chan et al., 1998), indicating that this class of genes constitutes a family present throughout the plant kingdom. A different class of plant homeobox genes was identified in *Arabidopsis thaliana* as a result of the *Bell1* mutation, which affects ovule development (Reiser et al., 1995). Another BELL-like protein, ATH1, is involved in a signal transduction pathway related to plant photomorphogenesis (Quadvlieg et al., 1995).

KNOX and BELL-like proteins belong to the TALE (three amino acid loop extension) superclass of homeodomain-containing proteins, since they have three extra amino acids within the loop connecting helices I and II (Bürglin, 1997). It has recently been shown that proteins from these two classes interact through domains present N-terminal to the homeodomain (Bellaoui et al., 2001; Müller et al., 2001; Smith et al., 2002; Chen et al., 2003; Bhatt et al., 2004). Regarding their interaction with DNA, binding-site selection and mutagenesis studies have indicated that KNOX proteins bind sequences containing a TGAC core (Krusell et al., 1997; Smith et al., 2002; Tioni et al., 2005). Similar studies with BELL-like proteins have not been reported, although a recent study demonstrated that the potato BELL-like protein StBEL5 binds to a TGAC-containing motif in the *ga20ox1* promoter (Chen et al., 2004).

In the present work, we investigated the interaction of the homeodomain of the *Arabidopsis* BELL-like protein ATH1 with DNA by a combination of binding-site selection and footprinting studies. Comparative analysis of its binding properties with those of the KNOX protein STM suggests that BELL homeodomains recognise a TGACAGGT sequence, with differences in affinity and specificity in comparison with KNOX homeodomains. The importance of residue 54 of the homeodomain in determining these different properties has also been investigated.

Results

DNA binding preferences of the ATH1 homeodomain

The homeodomain of the BELL-like *Arabidopsis* protein ATH1 was expressed in *Escherichia coli* as a fusion with *S. japonicum* glutathione *S*-transferase and purified by affinity chromatography. To identify DNA sequences that are specifically bound by ATH1, we applied the random oligonucleotide selection technique on a mixture of 4^{12} different oligonucleotides. Two independent selection experiments were performed. The first experiment was performed with an oligonucleotide containing the sequence TGACTGC, which resembles the consensus binding site obtained for ATH1 (see below), in one of its

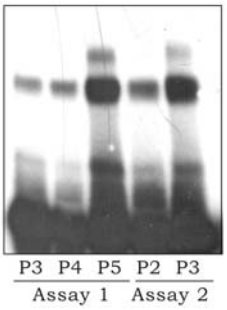


Figure 1 Binding of ATH1 to oligonucleotide populations obtained after selection.
Binding of ATH1 to oligonucleotide mixtures obtained after different rounds of selection in two different experiments was analysed by EMSA. P2–P5 indicate populations obtained after two–five rounds, respectively.

arms. Therefore, a second experiment was planned with a change in the arm sequence to TGGCTGC to avoid any effect of the binding of the protein to this sequence. Figure 1 shows an electrophoretic mobility shift assay (EMSA) in which the binding of ATH1 to different oligonucleotide mixtures obtained after three–five rounds of selection in each of the experiments was analysed. A higher proportion of bound DNA was progressively observed as the number of cycles increased, indicating that ATH1 is able to select specific DNA sequences from the mixture. In the second experiment, the increase in binding was evident at earlier stages. After seven rounds of selection, no further improvement in binding could be detected. The bound products were then cloned and 30 randomly picked clones were sequenced. Figure 2 shows a list of the sequences of the clones obtained in both experiments, arranged to maximise the alignment. Eight of the clones obtained in the second experiment contained the sequence TGAC repeated twice in tandem, suggesting that this sequence is important for ATH1 binding (Figure 2). Indeed, this motif was present in all clones analysed, except for one that contained TGAT. Clones with a single TGAC motif were then aligned to evaluate the presence of other conserved nucleotides besides TGAC. A consensus sequence TGACAGGT could be deduced from this alignment, with a highly conserved TGACA core at one side, followed by positions with defined nucleotides at a frequency of 50–60% (Figure 2). It is interesting that clones with two TGAC motifs appeared only in the second experiment. This may arise from the fact that the oligonucleotide used for the first experiment contains the sequence TGAC in one of its arms. A further analysis of the clones obtained in the first experiment indicates that the selected TGAC is present in the same strand in all of them, located in tandem with the TGAC present in the arm. We speculate that molecules with two tandem copies of TGAC motifs are efficiently bound by the ATH1 homeodomain, even if they do not contain the complete consensus sequence, maybe because they are able to bind two ATH1 molecules. Indeed, a second shifted band was observed with oligonucleotide mixtures in the later stages of selection (Figure 1). This band most likely represents DNA molecules bound to two ATH1 homeodomains. It is also noteworthy that the two copies of the TGAC motif selected in the second experiment are separated by three nucleotides in seven out of eight clones with this arrangement. It can be speculated that this may represent the optimal spacing for the binding of two ATH1 homeodomains. Spacings between the arm TGAC and the selected TGAC in the first experiment are, however, larger (three–seven nucleotides, mostly six). Since the random sequence used for selection contains 12 nucleotides, it seems more likely that three nucleotides is the minimal spacing required for the binding of two ATH1 molecules in tandem.

The consensus binding sequence obtained for ATH1 is similar to that selected by the KNOX proteins Hooded and Knotted1 (Krusell et al., 1997; Smith et al., 2002), indicating that these proteins may bind DNA by similar mechanisms. To perform a comparative analysis of the binding properties of BELL and KNOX homeodomains, we also expressed the homeodomain of the *Arabidopsis* KNOX protein STM (Long et al., 1996) and analysed its binding to oligonucleotide mixtures selected by ATH1 (data not shown). An increase in binding efficiency was observed upon progression of selection, suggesting that STM and ATH1 indeed prefer similar sequences.

To validate the results of the selection experiment, the relevance of different positions within the consensus sequence for ATH1 or STM binding was analysed by EMSA using synthetic oligonucleotides containing changes at single positions within the TGAC core (Figure 3). Neither ATH1 nor STM were able to bind to oligonucleotides with modifications within the GA dinucleotide of the TGAC core (Figure 3A, BS2 and BS3), while changes at the first position (T) had a reduced effect on binding (Figure 3A, BS1). Changes at the fourth position behaved differently, depending on whether A or G is introduced instead of C. While none of the proteins is able to bind an oligonucleotide containing TGAG, ATH1 shows reasonable binding to TGAA, which is recognised less efficiently by STM (Figure 3B, BS4A and BS4G). Changes at positions located 5′ to the TGAC core had no effect on binding, consistent with the fact that no selection of specific nucleotides was observed at these positions (Figure 3B, BS-1). Binding of ATH1 to these oligonucleotides was rather low compared to observations made with oligonucleotides obtained from the selected clones (see below). Low binding is likely due to two different factors. First, these oligonucleotides do not contain the complete consensus sequence (i.e., they contain A at positions 7 and 8). Second, ATH1 binding to synthetic oligonucleotides is less efficient, independent of the sequence (data not shown), probably because they are shorter in length, suggesting that ATH1 prefers larger DNA fragments for binding.

The importance of positions 3′ to the TGAC core was analysed using oligonucleotides prepared from the selected clones. Figure 4 shows that deviations from the consensus, with the sole exception of the introduction of A instead of T at the eighth position (BSS8), produce a decrease in ATH1 binding, indicating that the deduced consensus truly represents the preferred target site. In a similar way, only BSS8 competes as efficiently as an oligonucleotide containing the selected sequence (BSS) for binding, while oligonucleotides with changes at other

Consensus	C	G	G	T	G	A	C	A	G	G	T	G
T	2	4	3	19	–	–	1	1	6	2	10	4
G	3	7	7	–	19	–	–	2	11	12	3	8
A	5	3	4	–	–	19	–	16	2	4	6	5
C	9	5	5	–	–	–	18	–	–	1	–	1
%	50	37	37	100	100	100	95	84	56	63	53	44
Position	-3	-2	-1	1	2	3	4	5	6	7	8	9

Figure 2 Compilation of the sequences of 30 random clones obtained after cloning the oligonucleotide populations selected by ATH1.

The sequences of the central 12-bp portions of the different clones are indicated in uppercase letters. Arm sequences are indicated in lowercase letters. Clone numbers (arbitrary) are indicated to the left; E1 or E2 indicate clones from the first or second experiment, respectively, and the name of the DNA fragments employed in the gel-shift analyses is underlined. The sequences have been positioned to maximise the alignment. In all cases, the strand that carries the consensus TGAC is shown. Nucleotides that match the consensus are shown in bold. Below, a Table indicating the nucleotide frequencies at each position, together with the derived consensus sequence, is shown. This Table has been constructed using clones that contain a single TGAC sequence to show the consensus outside this core.

positions are less effective (Figure 4, upper right panel). Furthermore, simultaneous changes at positions 6, 7 and 8 of the consensus sequence significantly affect ATH1 binding (Figure 4, BSS6–8). On the other hand, an oligonucleotide containing two tandem copies of the TGAC core separated by three nucleotides was bound with similar efficiency as the consensus (Figure 4, BSSD). This explains the fact that clones of this type also appeared as a result of selection.

When comparing the effect of single substitutions on ATH1 and STM binding, some differences are evident. Changes at positions 5 and 6 (TGACAG) produce a much smaller effect on STM than on ATH1 binding (Figure 4, BSS6 and BSS5,6). This is especially evident at position 5, since STM binds an oligonucleotide containing TGACC or TGACT almost as well as the consensus TGACA, while these changes severely affect ATH1 binding (Figure 3, BS5,6 and Figure 4, BSS5,6). G at position 7 seems to be particularly important for binding by both proteins (Figure 4, BSS7), while the change of T for A at position 8, which does not affect ATH1 binding, reduces complex formation by STM (Figure 4, BSS8). Competition experiments also demonstrate the existence of differences in the binding preferences of ATH1 and STM. For example, BSS5,6 competes as efficiently as BSS6 and BSS8 for binding to STM (Figure 4, lower right panel), but is a poor competitor for binding to ATH1 (Figure 4, upper right panel). It can be concluded that although both proteins recognise similar sequences, subtle differences exist in their interaction with DNA.

Analysis of ATH1 and STM binding to DNA by hydroxyl radical footprinting and interference assays

The binding of ATH1 to its target site was also analysed by footprinting experiments using hydroxyl radical attack of DNA. For this purpose, an oligonucleotide containing the consensus binding site, previously labelled specifically at one of its 3′-ends, was incubated with ATH1 and subjected to hydroxyl radical attack. Free and bound DNA were separated by EMSA, recovered from the gel and analysed by denaturing polyacrylamide gel electrophoresis. Figure 5A shows the footprinting patterns of the top (i.e., the one containing the TGAC motif) and bottom strands, obtained after labelling the oligonucleotide at different 3′-ends. Comparison of the respective cleavage patterns indicates that ATH1 protects seven nucleotides from the top strand. The protected area includes

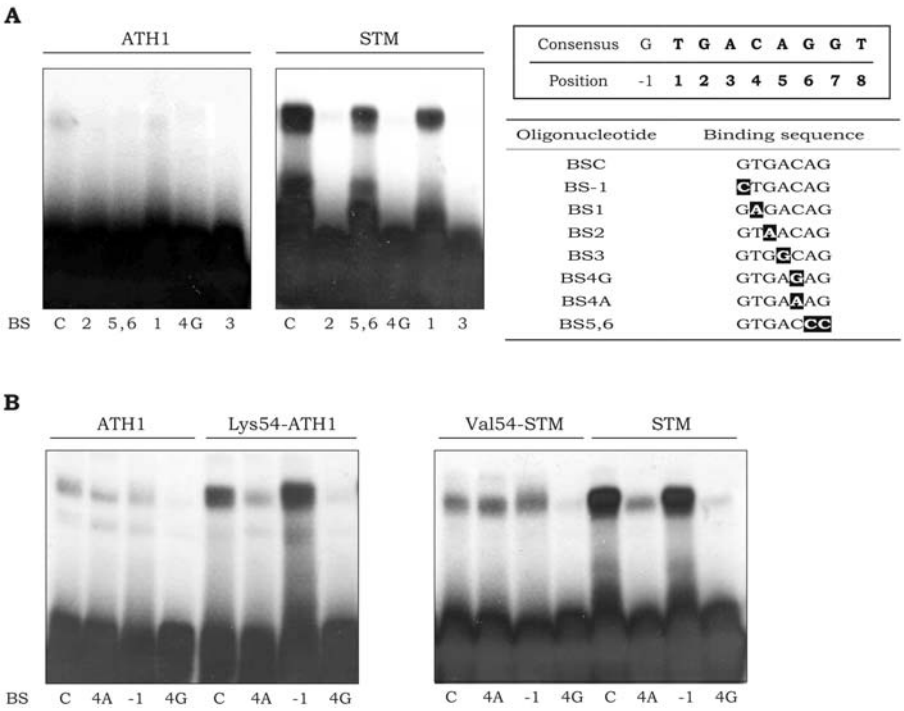


Figure 3 Binding of ATH1 and STM to oligonucleotides containing variants of the core consensus sequence.
(A) ATH1 and STM were analysed by EMSA for binding to 24-mer synthetic double-stranded oligonucleotides with changes within and outside the TGAC core. The Table to the right describes the sequences of the different oligonucleotides (BSC to BS5,6) that were used in each lane. (B) Binding of mutants at position 54 of the homeodomain to different oligonucleotides. The oligonucleotide sequences are those indicated in (A).

GAC and four adjacent nucleotides towards the 3′-end. Protection is higher within GACA and progressively decreases towards the 3′-end, so that the seventh protected position (T) shows only moderate, but reproducible protection (Figure 5A). On the bottom strand, the protected region covers nucleotides GTCA, complementary to TGAC, but the first two positions are much less protected than the rest (Figure 5A). It is noteworthy that a second protected region is observed in the bottom strand. This region contains the sequence GACAAT, which comprises the most important nucleotides of the core. We speculate that this protection arises from a second molecule of ATH1 that binds to this sequence. We were unable to observe the corresponding footprint in the top strand, since this sequence is located near the 3′-end of this strand and signals are too weak in this region. When the footprinting experiment was performed with an oligonucleotide lacking this sequence and containing a TGACAGAA core, the protected region in the bottom strand was more extended, covering nucleotides complementary to TGACAGA (Figure 5B).

Additional information about the nucleotide positions that influence binding of ATH1 to DNA was obtained in missing nucleoside experiments. In these experiments, DNA was treated with hydroxyl radical-generating agents that produce a population of molecules with single cleavages along the phosphodiester backbone. This population was subjected to EMSA, from which the free and bound fractions were recovered. Molecules with cleavages at positions important for binding are then underrepresented in the bound fraction. Figure 5A shows a missing nucleoside experiment using the DNA fragment containing the consensus binding site previously labelled at one of its 3′-ends and treated with hydroxyl radical-generating agents. There is good correlation between the region protected by ATH1 and the nucleotide positions important for binding. This means that all nucleotides in the protected area establish contacts that contribute to binding efficiency to different degrees. Once again, two regions of interference were observed in the bottom strand. When an oligonucleotide lacking the second GACA motif was used, the interference pattern also extended towards the 3′-end, but nucleotides complementary to AG in TGACAGA showed low relative interference.

Footprinting and missing nucleoside experiments were also carried out with the STM homeodomain for comparison. Figure 6A shows the results obtained in interference experiments with an oligonucleotide containing the consensus sequence. For the top strand, the results are very similar to those observed with ATH1, except for the fact that the interference pattern of STM covers an additional nucleotide towards the 3′-end. In the bottom strand, interference is observed along an 8-bp stretch that covers the nucleotides complementary to GTGACAGG. With STM, no additional footprint was evident in the region of the bottom strand that contains a sequence similar to the core, as was observed with ATH1. This indicates that STM is able to efficiently discriminate among the two sequences under the conditions used for binding, while ATH1 binds to both. This difference may arise from the fact that higher protein concentrations had to

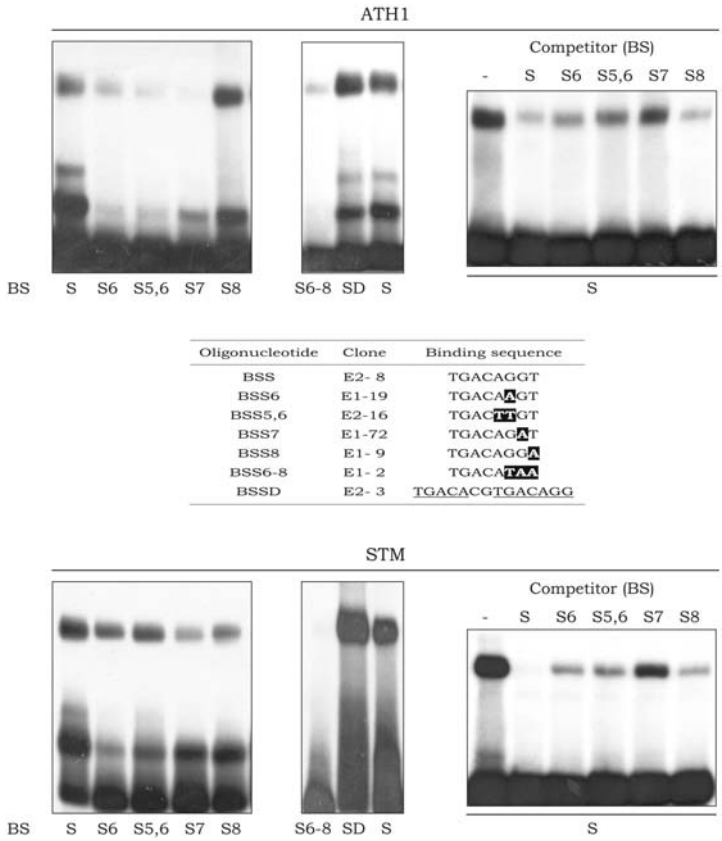


Figure 4 Binding of ATH1 and STM to oligonucleotides containing variations at positions 3′ of TGAC.
ATH1 and STM were analysed by EMSA for binding to 44-mer double-stranded oligonucleotides obtained after cleavage of the corresponding clones from the selection experiments with EcoRI and HindIII, and purification through polyacrylamide gels. The Table describes the relevant sequence of the different oligonucleotides (BSS to BSSD) and the clones they originate from. In the two rightmost panels, competition for binding to labelled BSS by a 50-fold excess of different unlabelled oligonucleotides (as described above the lanes of each gel) was analysed. All oligonucleotides were obtained by amplification from the corresponding clones with primers R1 and R2′, so that they do not contain the TGAC sequence in their arms.

be used in the case of ATH1 to obtain enough bound DNA to be analysed by this method, mainly because ATH1 binds less efficiently than STM to DNA.

Footprinting and missing-nucleoside data can be correlated using EMSA. The relaxed specificity at position 8 (TGACAGGT) displayed by ATH1 may be explained by the observation that this protein protects less efficiently nucleotides from both strands present at this position. Conversely, the tolerance to changes at the fourth nucleotide (TGAC) is not followed by reduced protection or interference. This may indicate that ATH1 contacts this position, but that the contacts are established with the sugar phosphate backbone rather than with specific bases. On the other hand, the reduced tolerance of STM to changes at position 8 is in agreement with the results obtained in footprinting and missing-nucleoside experiments, since this position is highly protected in both strands. Moreover, STM contacts eight nucleotides in the top strand, one more towards the 3′-end than ATH1.

To further analyse these observations, footprinting and missing-nucleoside experiments were performed with STM and an oligonucleotide containing the sequence TGACAGAA (Figure 6B). With this oligonucleotide, the protected region covers six nucleotides in the top strand TGACAGAA and eight nucleotides complementary to GTGACAGAA in the bottom strand. The decrease in the extension of DNA contacts compared with the consensus sequence is in agreement with the reduced affinity for the oligonucleotide BSS8 observed in EMSA. Taken together, these results demonstrate the importance of this position in DNA binding by STM.

Homeodomain position 54 is responsible for the different properties of the ATH1 and STM homeodomains

In an attempt to gain insight into the molecular basis of the different behaviour of the ATH1 and STM homeodomains, the amino acid residues that may be directly involved in DNA binding were compared, based on what is known for other homeodomains. Figure 7 shows an alignment of the homeodomain sequences of ATH1 and STM, together with those of representative members of the BELL and KNOX families. In agreement with the fact that both homeodomains recognise similar target sequences, it was observed that most putative important positions are conserved. An exception to this is residue 54, which is Lys in STM and Val in ATH1 (Figure 7, bottom). The role of residue 54 in determining the respective binding properties was then investigated by producing

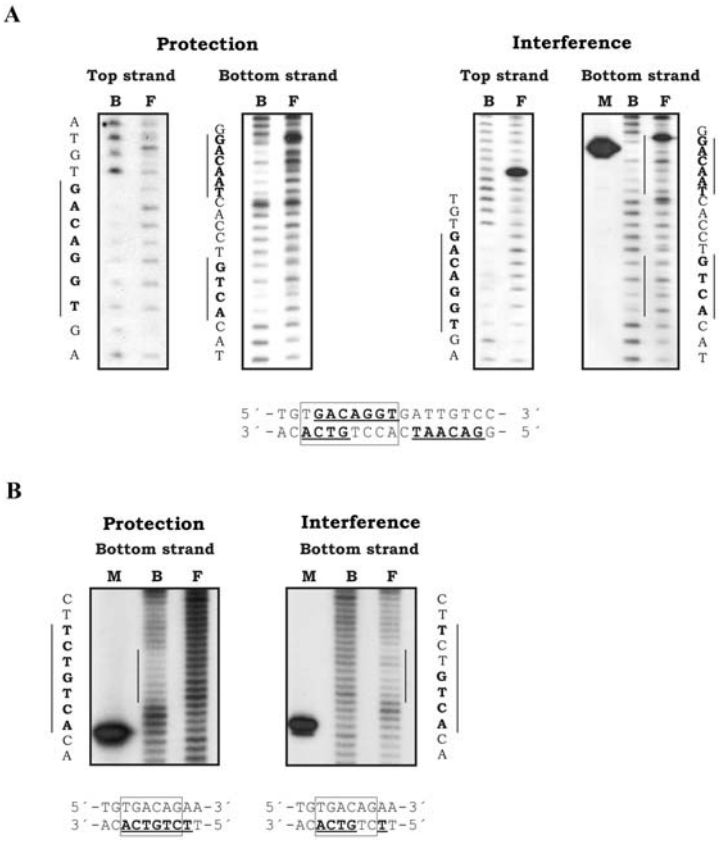


Figure 5 Hydroxyl radical footprinting and interference assays of the binding of ATH1 to DNA.
(A) An oligonucleotide containing the ATH1 consensus binding sequence previously labelled at the 3′-end of either strand (HindIII or EcoRI sites) was subjected to hydroxyl radical attack either after (protection) or before (interference) binding to ATH1. Free (F) and bound (B) DNA were separated and analysed. Lane 'M' indicates a 34-mer DNA fragment used to locate the position of the footprint. Letters beside each panel indicate the DNA sequence (5′-end in the upper part) of the corresponding strand in this region. In the lower part of the Figure, the sequence of the binding site is shown and the protected positions are indicated in bold and underlined. (B) Footprinting and interference experiments were also performed with the bottom strand of an oligonucleotide containing the non-consensus sequence TGACAGAA and lacking the sequence GACAAT. An aliquot of the same DNA digested with the restriction enzyme BglII was used to locate the position of the footprint (lane 'M'). In this case, an extended protected region comprising the consensus sequence is observed.

mutant homeodomains with exchanged amino acids at this position. The introduction of Val54 in the STM homeodomain produces a protein able to bind to TGAA, while Lys54-ATH1 shows a net preference for TGAC (Figure 3B, BSC and BS4A). Changes in the GA dinucleotide of TGAC affect binding by both proteins, as it does with their wild-type counterparts (Figure 8A, BS2 and BS3). Conversely, Lys54-ATH1 is less affected than the wild-type protein by changes at positions 5 and 6 (TGACAG), as evidenced by binding to BSS6 and BSS5,6 in Figure 8B. Val54-STM, in turn, displays relaxed specificity at position 4 and is more affected by changes at positions 5 and 6 (Figure 8A,B), as noted above for wild-type ATH1. Substitutions at position 54 also influence DNA binding affinity, since proteins with Lys54 displayed considerably higher affinity than their counterparts with Val. We conclude that residue 54 of the homeodomain is a main determinant of the different binding properties of ATH1 and STM, influencing both the affinity and specificity of DNA binding at positions 4, 5 and 6 of the consensus sequence. This assertion is not valid for the different binding behaviour observed upon changes at position 8 (TGACAGGT), since the mutants at position 54 behaved as the respective wild-type proteins when this nucleotide was modified (Figure 8B, BSS8).

Discussion

We analysed the interaction of the homeodomain of the BELL-like protein ATH1 with DNA by a combination of binding-site selection, footprinting and EMSA. The results obtained indicate that ATH1 binds the sequence TGACAGGT, establishing closer contacts with the underlined positions on both strands, and has additional interactions with nucleotides located towards the 3′-end of this core. ATH1 contains Asn, Ile, Asn, Val and Arg at positions 47, 50, 51, 54 and 55, respectively (Figure 7). The residues present at these positions are involved in establishing specific interactions with DNA in most homeodomains. This particular combination of amino acids is present only in BELL-like proteins, but some of them are also present in other homeodomains for which structural information is available. In particular, the structure of the Pbx1 and extradenticle homeodomains bound to DNA (Passner et al., 1999; LaRonde-LeBlanc and Wol-

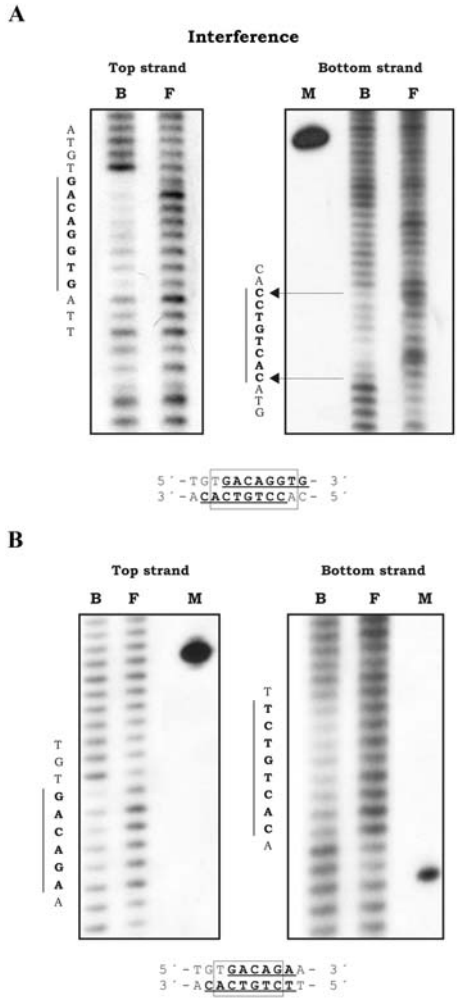


Figure 6 Hydroxyl radical interference assays of the binding of STM to DNA.
(A) An oligonucleotide containing the consensus binding site was labelled at the 3′-end of either strand (BamHI or XbaI sites) and subjected to hydroxyl radical attack before binding to STM. Free (F) and bound (B) DNA were separated and analysed. An aliquot of the DNA fragment labelled at XbaI and digested with the restriction enzyme HindIII was used to locate the position of the footprint (lane 'M'). Letters to the left of each panel indicate the DNA sequence (5′-end in the upper part) of the corresponding strand in this region. In the lower part of the figure, the sequence of the binding site is shown and the region of interference is indicated in bold and underlined. (B) An interference experiment was also performed with an oligonucleotide containing the non-consensus sequence TGACAGAA. An aliquot of the same DNA digested with the restriction enzyme BglII was used to locate the position of the footprint (lane 'M').

berger, 2003) may be relevant for comparison, since these proteins contain Asn47, Gly50, Asn51, Ile54 and Arg55 and bind the sequence TGAT. In the Pbx1-DNA complex, the first two positions are contacted by Arg55 through van der Waals' and hydrogen bond interactions, respectively, while Asn51 establishes hydrogen bonds with position 3 and van der Waals' interactions with position 4 (LaRonde-LeBlanc and Wolberger, 2003). We propose that similar interactions are established by Arg55 and Asn51 in both ATH1 and STM (Figure 9). Ile54 in Pbx1 makes contacts with the complementary strand, more precisely with the sugar moiety of the nucleotide adjacent to position 4.

In addition, we have recently proposed that the GAC nucleotides contained in TGAC interact with Arg55, Asn51 and Lys54 present in KNOX proteins (Tioni et al., 2005). Nucleotides 3′ to this core may be recognised by residues present at positions 50 and/or 54. It is noteworthy that BELL-like proteins contain similar amino acids at these positions, with the sole exception of Val54. Based on our analysis of the differential binding of ATH1 and the KNOX protein STM, as well as their respective mutants, we propose that Val54 of ATH1 may establish contacts with positions 5 and 6 (TGACAG), rather than with position 4, thus producing a relaxed specificity at this nucleotide position respective to homeodomains that contain Lys54 (Figure 9). Lys54 may be able to establish hydrogen bonds with a G complementary to C in GAC in STM and Lys54-ATH1. The establishment of these hydrogen bonds may also explain the increased affinity for DNA displayed by proteins with Lys54. The Val side-chain present in ATH and Val54-STM would only make less energetic van der Waals' or water-mediated contacts with DNA.

The putative DNA-contacting amino acids of ATH1 are also present in all described BELL-like proteins, indicating that they may all recognise identical or similar sequences. In addition, BELL-like proteins interact with KNOX proteins through domains present outside the homeodomain (Bellaoui et al., 2001; Müller et al., 2001). A recent study suggests that the complex formed by a KNOX and a BELL-like protein binds a sequence containing two tandem copies of a TGAC motif, located 1 bp apart, present in the *ga20ox* promoter (Chen et al., 2004). This would indicate that the third helices of BELL and KNOX homeodomains are located in an antiparallel orientation on opposite sites of the DNA molecule. Considering the binding characteristics described here, it can be speculated that the homeodomain that interacts with the first TGAC motif may reaccommodate some DNA contacts to allow binding of the other protein molecule. Notably, only the GA dinucleotide is essential for binding to the first TGAC motif, while the second one requires GAC (Chen et al., 2004), in coincidence with the requirements described here for a BELL and a KNOX protein, respectively.

In conclusion, although the DNA binding properties of ATH1, and probably all BELL-like proteins, show similarities with those presented by KNOX proteins, some differences in affinity and selectivity are also evident. These different properties, mainly attributable to the amino acid present at homeodomain position 54, are probably essential for the respective functions of these proteins *in vivo*, considering the conservation observed at this position within each protein family.

Materials and methods

Cloning, expression and purification of recombinant proteins

The ATH1 homeodomain coding sequence (amino acids 364–474) was amplified from clone RAFL09-24-O21 (Seki et al.,

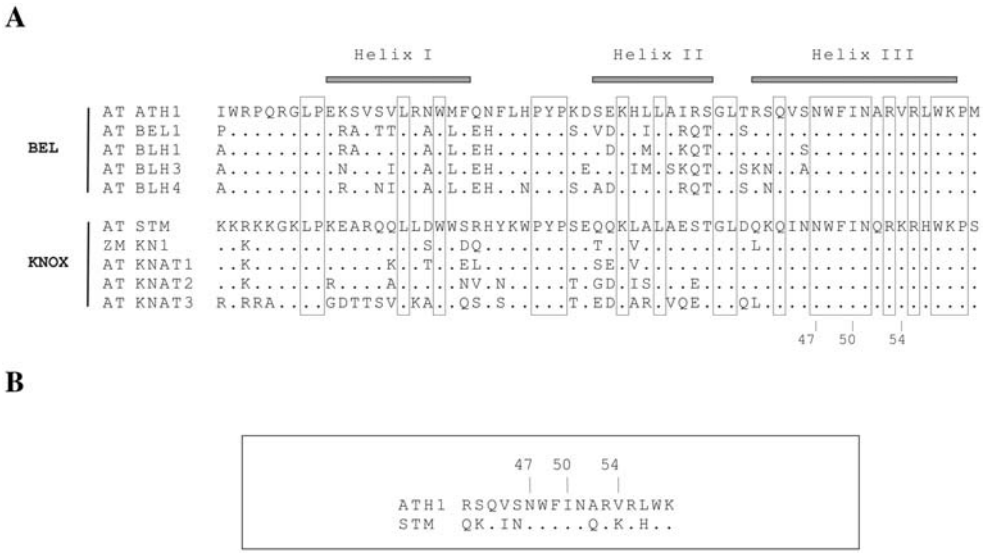
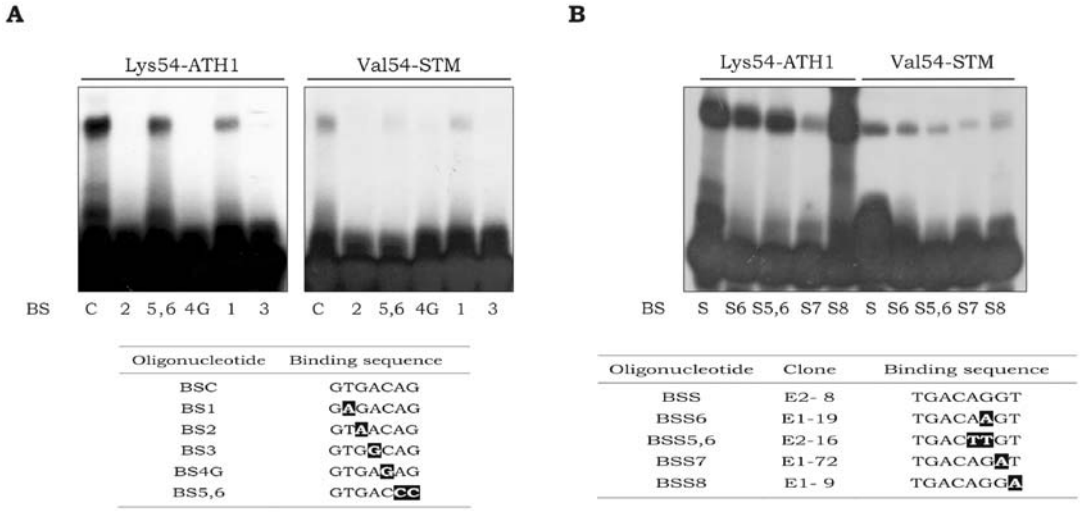


Figure 7 Sequences of the homeodomains of ATH1, STM and representative members of the BELL and KNOX families.
(A) Dots indicate conserved amino acids within each family with respect to ATH1 or STM. Residues conserved in both families are boxed. In panel (B) helix III residues present in ATH1 and STM are shown. Numbers indicate positions within the homeodomain.

2002) and inserted in frame into the BamHI and EcoRI sites of the expression vector pGEX-3X (Smith and Johnson, 1988). Amplifications were performed using oligonucleotides ATH11: 5′-GGC GGA TCC AGC AGC TGA AAC GAA AGA-3′ and ATH12: 5′-GGC GAA TTC CAA AGT TGG TAA ACA CAA T-3′. For the construction of Lys54-ATH1, oligonucleotides 5′-TAT AAA TGC TCG GAA AAG GCT ATG G-3′ and 5′-CCA TAG CCT TTTCCG AGC ATT TAT A-3′ (mutations underlined) were used in combination with primers ATH11 and ATH12 to amplify partially overlapping N- and C-terminal homeodomain fragments. The resulting products were mixed in buffer containing 50 mM Tris-HCl (pH 7.2), 10 mM $MgSO_4$, and 0.1 mM dithiothreitol (DTT), incubated at 95°C for 5 min, and annealed by allowing the solution to cool to 24°C in approximately 1 h. After this, 0.5 mM of each dNTP and 5 U of the Klenow fragment of *E. coli* DNA polymerase I were added, followed by incubation for 1 h at 37°C. A portion of this reaction was directly used to amplify the chimeric fragments with oligonucleotides ATH11 and ATH12.

The STM homeodomain coding sequence was amplified with primers STM1 (5′-CCG GGA TCC TAG GGA GCC TCA AGC AAG-3′) and STM2 (5′-GGC GAA TTC TCA AAG CAT GGT GGA GGA-3′) from clone RAFL09-36-A03 and inserted into pGEX-3X as described above. The K54V mutation was introduced using oligonucleotides 5′-TAA ACC ARA GGG TGC GGC AYT GGA-3′ and 5′-TCC ART GCC GCA CCC TYT GGT TTA-3′ following a similar strategy to that described above. All constructions were checked by DNA sequence analysis.

For expression, *E. coli* cells bearing the corresponding plasmids were grown and induced as previously described (Palena et al., 1998). Purification of the fusion products was carried out



Oligonucleotide	Binding sequence
BSC	GTGACAG
BS1	GAGACAG
BS2	GTAACAG
BS3	GTGCCAG
BS4G	GTGACAG
BS5,6	GTGACCC

Oligonucleotide	Clone	Binding sequence
BSS	E2- 8	TGACAGGT
BSS6	E1-19	TGACAAGT
BSS5,6	E2-16	TGACTTGT
BSS7	E1-72	TGACAGAT
BSS8	E1- 9	TGACAGGA

Figure 8 Binding of ATH1 and STM mutants at position 54 of the homeodomain to oligonucleotides containing different binding sequences.
Lys54-ATH1 and Val54-STM were analysed by EMSA for binding to oligonucleotides containing variants of the consensus sequence. (A) Binding to 24-mer synthetic double-stranded oligonucleotides with changes within and outside the TGAC core. The sequences of the different oligonucleotides (BSC to BS5,6) are shown below. (B) Binding to 44-mer double-stranded oligonucleotides obtained after cleavage of the corresponding clones from the selection experiments. The relevant sequences of the different oligonucleotides (BSS to BSS8) and the clones they originate from are shown below. All oligonucleotides were obtained by amplification from the corresponding clones with primers R1 and R2′, so that they do not contain the TGAC sequence in their arms.

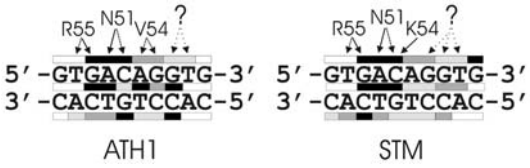


Figure 9 Schematic model of the interaction of ATH1 and STM with DNA.
Both strands of the binding sequence recognised by these proteins are shown. Bars between both strands show the importance of each base pair for binding, as deduced from the use of different binding sites. The upper and lower bars represent the importance of nucleotides in the top and bottom strands, respectively, for binding, as deduced from interference experiments. The intensity of grey indicates the relative importance of each position. Above the sequences, the homeodomain residues that presumably interact with each position are shown. Solid and interrupted arrows indicate hydrogen bond and van der Waals' interactions, respectively.

essentially as described by Smith and Johnson (1988), with modifications described by Palena et al. (1998).

DNA-binding assays

For EMSA, aliquots of purified proteins were incubated with double-stranded DNA (0.3–0.6 ng, 30 000 cpm labelled with [α-^{32}P]dATP by filling in the 3′-ends using the Klenow fragment of DNA polymerase) generated by hybridisation of the complementary synthetic oligonucleotides 5′-AAT TCA GAT CTT GTG ACA GAA GAG-3′ and 5′-GAT CCT CTT CTG TCA CAA GAT CTG-3′, or derivatives with modifications within the binding sequence as described in the text. Alternatively, oligonucleotides amplified from clones obtained after the random selection technique were used. Binding reactions (20 μl) containing 20 mM HEPES (pH 7.5), 50 mM KCl, 2 mM $MgCl_2$, 0.5 mM EDTA, 1.0 mM DTT, 0.5% Triton X-100, 22 ng/μl BSA, 50 ng/μl poly(dI-dC), and 10% glycerol were incubated for 20 min at room temperature, supplemented with 2.5% Ficoll and immediately loaded onto a running gel (5% acrylamide, 0.08% bis-acrylamide in 0.5× TBE plus 2.5% glycerol; 1× TBE is 90 mM Tris-borate, pH 8.3, 2 mM EDTA). The gel was run in 0.5× TBE at 30 mA for 1.5 h and dried prior to autoradiography.

Binding-site selection

To select DNA molecules specifically bound by ATH1, the random oligonucleotide selection technique (Oliphant et al., 1989) was applied, using procedures described by Blackwell and Weintraub (1990). A labelled 52-mer double-stranded oligonucleotide containing a 12-bp central core with random sequences (5′-GAT GAA GCT TCC TGG ACA ATN_{12} GCA GTC ACT GAA GAA TTC TG-3′) was incubated with purified protein as described above. Bound DNA molecules were separated by EMSA and eluted from gel slices with 0.5 ml of 0.5 M ammonium acetate, 10 mM $MgCl_2$, 1 mM EDTA, and 0.1% (w/v) SDS. The selected DNA molecules were amplified using oligonucleotides R1 (5′-GAT GAA GCT TCC TGG ACA AT-3′) and R2 (5′-CAG AAT TCT TCA GTG ACT GC-3′). Amplification reactions were performed as follows: 30 cycles of 1 min at 94°C, 1 min at 53°C, and 1 min at 72°C. After purification through polyacrylamide gels, the amplified molecules were subjected to new cycles of binding, elution and amplification. Enrichment in sequences bound specifically by ATH1 was monitored by binding and competition analysis in EMSA. After seven rounds of selection, the population of oligonucleotides was cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, USA). Randomly picked clones were sequenced. In a second experiment, the arm of the oligonucleotide containing the random core was modified using primer R2′ (5′-CAG AAT TCT ACA GTG GCT GC-3′) instead of R2 for amplification.

Oligonucleotides from selected clones to be used in EMSA were obtained by amplification with primers R1 and R2′, followed by cleavage with EcoRI and HindIII and filling in with Klenow fragment and [α-^{32}P]dATP.

Footprinting analysis

For the analysis of hydroxyl radical footprinting patterns, a double-stranded oligonucleotide containing the ATH1 binding site with BamHI and EcoRI compatible cohesive ends was cloned into similar sites of pBluescript SK$^-$. From this clone, DNA fragments were obtained by PCR using reverse and universal primers, followed by cleavage with BamHI and EcoRI (in assays with ATH1) or with BamHI and XbaI (in assays with STM). The fragments were labelled at one of their 3′-ends by incubation with the Klenow fragment of DNA polymerase and [α-^{32}P]dATP prior to cleavage with the second enzyme, and were subsequently purified by non-denaturing polyacrylamide gel electrophoresis. Binding of either ATH1 or STM to these oligonucleotides (200 000 cpm) was performed as described for EMSA in 15 μl of 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM β-mercaptoethanol, 0.1 mM EDTA, 22 ng/μl BSA, 10 ng/μl poly(dI-dC) and 800 ng of ATH1 or 200 ng of STM. After binding, bound and free oligonucleotides were subjected to hydroxyl radical cleavage by the addition of 10.5 μl of 6.6 mM sodium ascorbate, 0.66 mM EDTA (pH 8.0), 0.33 mM $(NH_4)_2Fe(SO_4)_2$ and 0.2% H_2O_2, and separated by EMSA (Dixon et al., 1991). The corresponding fractions were excised from the gel, eluted and analysed on denaturing polyacrylamide gels. Footprinting assays with the oligonucleotide containing the sequence TGACAGAA were performed as previously described (Tioni et al., 2005).

Missing nucleoside experiments

For analysis of the nucleosides required for ATH1 or STM binding, labelled oligonucleotides containing the respective binding sites were obtained from clones in pBluescript SK$^-$ as described above and subjected to hydroxyl radical cleavage (Dixon et al., 1991). Binding of the proteins to the treated oligonucleotide (200 000 cpm) and separation of the free and bound fractions by EMSA were performed as described. These fractions were excised from the gel, eluted and analysed on a denaturing polyacrylamide gel.

Miscellaneous methods

Total protein was measured as described by Sedmak and Grossberg (1977). For quantitative analysis, radioactive bands were cut from exposed gels and measured by scintillation counting.

Acknowledgements

We gratefully acknowledge Dr. Raquel Chan for helpful comments and for critically reading the manuscript. We also acknowledge the RIKEN BRC Experimental Plant Division (Japan) for sending us full-length cDNA clones. This work was supported by grants from CONICET, ANPCyT and Universidad Nacional del Litoral (Argentina). DHG is a member of CONICET; ILV is a fellow of the same institution.

References

Bellaoui, M., Pidkowich, M.S., Samach, A., Kushalappa, K., Kohalmi, S.E., Modrusan, Z., Crosby, W.L., and Haughn, G.W. (2001). The *Arabidopsis* BELL1 and KNOX TALE

homeodomain proteins interact through a domain conserved between plants and animals. Plant Cell *13*, 2455–2470.

Bhatt, A.M., Etchells, J.P., Canales, C., Lagodienko, A., and Dickinson, H. (2004). VAAMANA-a BEL1-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in *Arabidopsis*. Gene *328*, 103–111.

Blackwell, T.K. and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science *250*, 1104–1110.

Bürglin, T.R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res. *25*, 4173–4180.

Chan, R.L., Gago, G.M., Palena, C.M., and Gonzalez, D.H. (1998). Homeoboxes in plant development. Biochim. Biophys. Acta *1442*, 1–19.

Chen, H., Rosin, F.M., Prat, S., and Hannapel, D.J. (2003). Interacting transcription factors from the TALE superclass regulate tuber formation. Plant Physiol. *132*, 1391–1404.

Chen, H., Banerjee, A.K., and Hannapel, D.J. (2004). The tandem complex of BEL and KNOX partners is required for transcriptional repression of *ga20ox1*. Plant J. *38*, 276–284.

Dixon, W.J., Hayes, J.J., Levin, J.R., Weidner, M.F., Dombroski, B.A., and Tullius, T.D. (1991). Hydroxyl radical footprinting. Methods Enzymol. *208*, 380–413.

Gehring, W.J. (1987). Homeo boxes in the study of development. Science *236*, 1245–1252.

Gehring, W.J., Affolter, M., and Bürglin, T. (1994a). Homeodomain proteins. Annu. Rev. Biochem. *63*, 487–526.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K. (1994b). Homeodomain-DNA recognition. Cell *78*, 211–223.

Krusell, L., Rasmussen, I., and Gausing, K. (1997). DNA binding sites recognised *in vitro* by a knotted class I homeodomain protein encoded by the *hooded* gene, *k*, in barley (*Hordeum vulgare*). FEBS Lett. *408*, 25–29.

LaRonde-LeBlanc, N.A. and Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes Dev. *17*, 2060–2072.

Long, J.A., Moan, E.I., Medford, J.I., and Barton, M.K. (1996). A member of the knotted class of homeodomain proteins encoded by the STM gene of *Arabidopsis*. Nature *378*, 66–69.

Müller, J., Wang, Y., Franzen, R., Santi, L., Salamini, F., and Rohde, W. (2001). *In vitro* interactions between barley TALE homeodomain proteins suggest a role for protein-protein associations in the regulation of *Knox* gene function. Plant J. *27*, 13–23.

Oliphant, A.R., Brandl, C.J., and Struhl, K. (1989). Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. Mol. Cell. Biol. *9*, 2944–2949.

Palena, C.M., Gonzalez, D.H., Guelman, S., and Chan, R.L. (1998). Expression of sunflower homeodomain containing proteins in *Escherichia coli*: purification and functional studies. Protein Expr. Purif. *13*, 97–103.

Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S., and Aggarwal, A.K. (1999) Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature *397*, 714–719.

Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J., and Wüthrich, K. (1989). The structure of the Antennapedia homeodomain determined by MNR spectroscopy in solution: comparison with prokaryotic repressors. Cell *59*, 573–580.

Qian, Y.Q., Furukubo-Tokunaga, K., Müller, M., Resendez-Perez, D., Gehring, W.J., and Wüthrich, K. (1994). Nuclear magnetic resonance solution structure of the fushi tarazu homeodomain from *Drosophila* and comparison with the Antennapedia homeodomain. J. Mol. Biol. *238*, 333–345.

Quaedvlieg, N., Dockx, J., Rook, F., Weisbeek, P., and Smeekens, S. (1995). The homeobox gene *ATH1* of *Arabidopsis* is derepressed in the photomorphogenic mutants *cop1* and *det1*. Plant Cell *7*, 117–129.

Reiser, L., Modrusan, Z., Margossian, L., Samach, A., Ohad, N., Haughn, G.W., and Fischer, R.L. (1995). The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the *Arabidopsis* ovule primordium. Cell *83*, 735–742.

Sedmak, J. and Grossberg, S. (1977). A rapid, sensitive, and versatile assay for protein using Coomassie brilliant blue G-250. Anal. Biochem. *79*, 544–552.

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., et al. (2002). Functional annotation of a full-length *Arabidopsis* cDNA collection. Science *296*, 141–145.

Smith, H.M.S., Boschke, I., and Hake, S. (2002). Selective interaction of plant homeodomain proteins mediates high DNA-binding affinity. Proc. Natl. Acad. Sci. USA *99*, 9579–9584.

Smith, D.B. and Johnson, K.S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. Gene *67*, 31–40.

Tioni, M.F., Viola, I.L., Chan, R.L., and Gonzalez, D.H. (2005). Site-directed mutagenesis and footprinting analysis of the interaction of the sunflower KNOX protein HAKN1 with DNA. FEBS J. *272*, 190–202.

Tsao, D.H.H., Gruschus, J.M., Wang, L., Nirenberg, M., and Ferretti, J.A. (1995). The three-dimensional solution structure of the NK-2 homeodomain from *Drosophila*. J. Mol. Biol. *251*, 297–307.

Vollbrecht, E., Veit, B., Sinha, N., and Hake, S. (1991). The developmental gene *Knotted-1* is a member of a maize homeobox gene family. Nature *350*, 241–243.

Received July 27, 2005; accepted September 23, 2005