Interaction of the PHD-Finger Homeodomain Protein HAT3.1 from *Arabidopsis thaliana* with DNA. Specific DNA Binding by a Homeodomain with Histidine at Position 51†

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

Received December 28, 2006; Revised Manuscript Received April 3, 2007

ABSTRACT: HAT3.1 is a member of the PHD-finger homeodomain protein family. The HAT3.1 homeodomain is highly divergent in sequence even at positions that are almost invariable among homeodomains. In this work, we have applied the random oligonucleotide selection technique to investigate if the HAT3.1 homeodomain is able to recognize specific DNA sequences. Analysis of the selected molecules followed by hydroxyl radical footprinting experiments and yeast one-hybrid assays indicated that HAT3.1 shows a preference for the sequence T(A/G)(A/C)ACCA, different from those bound by other homeodomains. Binding was dependent on homeodomain residues located at positions 47, 50, 51, and 54, the same positions that usually participate in DNA binding in most homeodomains. The study of the interaction of mutants at these positions with DNA carrying nucleotide changes at specific sites suggested that H51 and K50 most likely interact with nucleotides 2 to 4 and 5 to 6, respectively, while W54 would establish contacts with position 4. The presence of H51 and W54 represents an innovation among homeodomain structures. The fact that the HAT3.1 homeodomain is able to interact with specific DNA sequences is evidence of the inherent plasticity of the homeodomain as a DNA binding unit.

The homeodomain (HD[1]) is a 60 amino acid motif found in a vast group of eukaryotic transcription factors. Its name comes from the study of homeotic mutations in *Drosophila*, which produce important alterations as a result of the development of tissues or organs in wrong places. HD-containing proteins have now been found in almost every eukaryotic organism that was searched for them, and a common theme seems to be that they participate in the regulation of developmental processes (*1*−*3*).

From X-ray diffraction and NMR studies, a general model for the structure of the HD has emerged (*4*−*6*). Briefly, it is conformed by three α-helices connected by a loop (helices I and II) and a turn (helices II and III). Helices I and II are antiparallel, and helix III is perpendicular to the other two. The structure is maintained by hydrophobic contacts between conserved residues from the different helices. The first nine residues are disordered and constitute the N-terminal arm. The fact that helices II and III form a structure that resembles the helix−turn−helix motif found in prokaryotic transcription factors suggested for the first time that the HD may function in DNA binding (*7*). Protein−DNA interaction studies have shown that this is indeed the case. Specific contacts with DNA are established by residues located within helix III and in the N-terminal arm (*8*−*11*). Nonspecific contacts are also established by residues in helices II and III and in the loop. A majority of HDs bind the sequence TAATNN, where the first two residues are contacted by the N-terminal arm, while the third and fourth are recognized by the almost invariant N51 and highly conserved isoleucine or valine at position 47, respectively (*11*). The next two positions are more variable and are contacted by residues 50 and 54, which are considerably less conserved among different HDs (*12*, *13*).

Plant genomes contain a large number of genes potentially encoding HDs. The count is almost 100 for *Arabidopsis* (*14*). Plant HDs can be divided into different families according to sequence conservation and the presence of additional motifs besides the HD (*15*). One of these families, termed PHD-finger HD, contains proteins in which the HD is associated with a zinc finger-like motif (PHD-finger, for *p*lant *h*omeo*d*omain finger; ref *16*). Although it was discovered in plants, the PHD-finger has also been found in other eukaryotic organisms, but not associated with HDs (*17*). It has recently been shown that the PHD-finger functions in the recognition of chromatin, binding to lysine 4 methylated histone H3 (*18*, *19*).

The *Arabidopsis* genome encodes two PHD-finger HD proteins, named HAT3.1 (the first protein in which the PHD-finger was recognized; ref *16*) and PRHA (*20*). Proteins from the same family have been also recognized in other plants, from which the ZmHox family of maize has been studied (*21*, *22*). It is noteworthy that the HDs of different PHD-finger HD proteins differ markedly in sequence (most notably in residues that putatively interact with DNA), which is an unusual feature. A pertinent question is then if the HD of

† Supported by grants from CONICET, ANPCyT, and Universidad Nacional del Litoral. D.H.G. is a member of CONICET; I.L.V. is a fellow of the same Institution.

* To whom correspondence should be addressed. Phone/fax: 54-342-4575219. E-mail: dhgonza@fbcb.unl.edu.ar.

[1] Abbreviations: BSA, bovine serum albumin; DTT, dithiothreitol; GST, glutathione *S*-transferase; HD, homeodomain.

10.1021/bi602657e CCC: $37.00
Published on Web 05/26/2007

these proteins functions in DNA binding and recognizes specific DNA sequences. At least for one of the proteins (ZmHox2a), which contains two HDs, it has been shown that this is the case (*23*). Conversely, for HAT3.1 it was previously not possible to show specific DNA binding (*16*).

To determine if the HD of HAT3.1 is able to recognize specific DNA sequences, we have applied the random oligonucleotide selection technique. We have found that the HAT3.1 HD is able to specifically bind sequences with the core TRMACCA. The analysis of HDs with site-specific mutations showed that binding is dependent on residues located at positions that are usually important in other HDs, indicating that, in spite of sequence divergence, this HD uses conserved mechanisms for DNA recognition. This fact suggests that HDs possess a high degree of functional plasticity even at positions that are highly conserved, making them attractive candidates for further evolutionary changes of eukaryotic transcription networks.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of Recombinant Proteins. The HAT3.1 HD coding sequence (amino acids 542 to 660; locus number At3g19510) was amplified from clone RAFL16-64-G01 (*24*) and inserted in frame into the *Bam*HI and *Eco*RI sites of the expression vector pGEX-3X (*25*). Amplifications were performed using oligonucleotides HAT31-1, 5′-ATGGTGGTTCTGGTGAAAT-3′, and HAT31-2, 5′-TGATTGAATCTAAAGTCTAC-3′. For the construction of the H51N mutant, oligonucleotides 5′-CTGGTTTAAGAATAGGCGTTGG-3′ and 5′-CCAACGCCTATTCTTAAACCAG-3′ were used in combination with primers HAT31-1 and HAT31-2 to amplify partially overlapping N-terminal and C-terminal HD fragments. The resulting products were mixed in buffer containing 50 mM Tris-HCl (pH 7.2), 10 mM $MgSO_4$, and 0.1 mM DTT, incubated at 95 °C during 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 units of the Klenow fragment of *Escherichia coli* DNA polymerase I were added, and incubation was followed for 1 h at 37 °C. A portion of this reaction was directly used to amplify the chimeric fragments with oligonucleotides HAT31-1 and HAT31-2. Mutants H51A, K50Q, K50A, N47A, N47V, and W54A were produced in a similar way but using primers 5′-CTGGTTTAAGGCTAGGCGTTGG-3′, 5′-CCAACGCCTAGCCTTAAACCAG-3′, 5′-TAACTGGTTTCAGCATAGGCGT-3′, ACGCCTATGCTGAAACCAGTTA-3′,5′-TAACTGGTTTGCACATAGGCGTTG-3′, 5′-CAACGCCTATGTGCAAACCAGTTA-3′, 5′-CAAGTTAATGCCTGGTTTAAG-3′,5′-CTTAAACCAGGCATTAACTTG-3′, 5′-CAAGTTAATGTCTGGTTTAAG-3′, 5′-CTTAAACCAGACATTAACTTG-3′, 5′-GCATAGGCGTGCGTCGATAAAC-3′, and 5′-GTTTATCGACGCACGCCTATGC-3′, respectively. All constructions were checked by DNA sequence analysis.

For expression, *E. coli* cells bearing the corresponding plasmids were grown and induced as described previously (*26*). Purification of the fusion products was carried out essentially as described by Smith and Johnson (*25*), with modifications described by Palena et al. (*26*).

DNA-Binding Assays. For electrophoretic mobility shift assays, aliquots of purified proteins were incubated with double stranded DNA (0.3−0.6 ng, 30000 cpm, labeled with [α-^{32}P]dATP by filling in the 3′-ends using the Klenow fragment of DNA polymerase) generated by hybridization of the complementary synthetic oligonucleotides 5′-AATTCGCTGCACCATCTAAACCATTG-3′ and 5′-GATCCAATGGTTTAGATGGTGCAGCG-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, 0.5 μg of 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% bisacrylamide 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. DNA binding assays were performed with the proteins fused to GST. Controls made with proteins obtained after cleavage with factor Xa indicated that the GST moiety does not affect the behavior of the recombinant proteins.

Binding Site Selection. To select DNA molecules specifically bound by HAT3.1, the random oligonucleotide selection technique (*27*) was applied, using procedures described by Blackwell and Weintraub (*28*). A labeled 52-mer double-stranded oligonucleotide containing a 12-bp central core with random sequences (5′-GATGAAGCTTCCTGGACAATN$_{12}$-GCAGTCACTGAAGAATTCTG-3′) was incubated with purified protein as described above. Bound DNA molecules were separated by electrophoretic mobility shift assays 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′-GATGAAGCTTCCTGGACAAT-3′) and R2 (5′-CAGAATTCTTCAGTGACTGC-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 HAT3.1 was monitored by binding and competition analysis in electrophoretic mobility shift assays. After 12 rounds of selection, the population of oligonucleotides was cloned into the pCR 2.1-TOPO vector (Invitrogen).

Oligonucleotides from selected clones to be used in electrophoretic mobility shift assays were obtained by amplification with primers R1 and R2 followed by cleavage with *Eco*RI and filling in with Klenow fragment and [α-^{32}P]-dATP.

Missing Nucleoside Experiments. For the analysis of the nucleosides required for HAT3.1 binding, a double-stranded oligonucleotide containing the HAT3.1 binding site with *Bam*HI and *Eco*RI 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 *Hin*dIII and *Xba*I. The fragments were labeled in 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 nondenaturing polyacrylamide gel

electrophoresis. The labeled oligonucleotide (15 μL) was then 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 (*29*). Binding of HAT3.1 to this oligonucleotide (200000 cpm) and separation of free and bound forms by electrophoretic mobility shift assays were performed as described above. The corresponding fractions were excised from the gel, eluted, and analyzed on denaturing polyacrylamide gels.

One-Hybrid Analysis in Yeast. For the analysis of HAT3.1 binding to the selected target sequence *in vivo* in yeast, a 1379-bp cDNA fragment (encoding amino acids 344 to 660) was cloned in frame with the coding region of the GAL4 activation domain in the *Eco*RI and *Xho*I sites of plasmid pGADT7 (Clontech). To obtain a yeast strain carrying the HAT3.1 binding sequence inserted into the genome, three tandem copies of the corresponding oligonucleotide were cloned in the *Bam*HI−*Eco*RI sites of vector pHIS3NX (*30*) in front of the HIS3 reporter gene preceded by its own minimal promoter. The entire cassette was then transferred to the *Not*I−*Xba*I sites of the integrative yeast plasmid pINT1, which confers resistance to the antibiotic G418. An *Nco*I−*Sac*I fragment of the clone in pINT1 was introduced into the *PDC6* locus of the yeast genome as described (*30*). The presence of the fragment of interest in the genome of cells resistant to G418 was analyzed by PCR with specific oligonucleotides. Alternatively, the HAT3.1 binding site was placed in front of the *lacZ* reporter gene contained in plasmid pLacZi (Clontech). Plasmid linearized in its *Nco*I site was introduced into the *URA3* locus of the yeast AW303 strain.

Plasmids or DNA fragments were introduced into yeast using the standard lithium acetate transformation method (*31*). Transformed cells were streaked into supplemented minimal medium lacking histidine to evaluate the activation of the HIS3 reporter gene by the HAT3.1−ADGAL4 fusion. As controls, yeast strains carrying only a minimal promoter−HIS3 fusion or a fragment with a sequence not bound by HAT3.1 *in vitro* were used. β-Galactosidase activity in cells carrying the *lacZ* reporter was measured as described in ref *32*.

Miscellaneous Methods. Total protein was measured as described by Sedmak and Grossberg (*33*). For quantitative analysis, radioactive bands were cut from exposed gels and measured by scintillation counting.

RESULTS

The HAT3.1 HD Is Able To Select Specific DNA Sequences from a Population. The HAT3.1 HD was expressed in *E. coli* as a fusion protein with GST. The expressed protein was purified by affinity chromatography and used to select DNA sequences from a population of 4^{12} different molecules. Separation of bound and free DNA was performed using electrophoretic mobility shift assays. Bound molecules from each selection cycle were amplified and used as targets of new selection rounds. Figure 1 shows that upon progression of the selection experiment a higher proportion of bound DNA was observed, suggesting that the HAT3.1 HD is indeed able to specifically interact with DNA. A constant increase in binding was observed up to cycle 9, after which only slight differences were evident. Competition experiments using unlabeled DNA from different rounds of selection indicated, however, that the populations obtained in cycles 11 and 12 were bound with the highest affinity (not shown). Accordingly, molecules from cycle 12 were cloned and individual clones picked at random were sequenced.

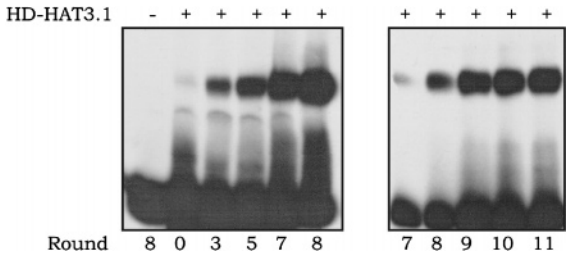


FIGURE 1: Binding of HAT3.1 to oligonucleotide populations obtained after selection. Binding of HAT3.1 (left panel, 200 ng; right panel, 20 ng) to oligonucleotide mixtures obtained after different rounds of selection was analyzed by an electrophoretic mobility shift assay. Numbers below gel images indicate the number of rounds of selection applied in each case.

The Consensus Binding Site of the HAT3.1 HD Does Not Contain a TAAT Core. A list of the sequences of the variable parts of the different clones is shown in uppercase letters in Figure 2, together with adjacent fixed positions in lowercase. An alignment of the variable part of the clones yields the sequence ACCATCTAAACC (underlined positions present in at least 27 out of 28 sequenced clones) present always in the same strand respective to the nonvariable arms. This possibly indicates that arm sequences influence the selection procedure, perhaps because they form part of the preferred sequence. The presence of an important part of the HAT3.1 binding site in the arms is nevertheless unlikely since the protein binds very poorly to the initial population of molecules which contain identical arm sequences (Figure 1). If the analysis extends to arm residues, it becomes evident that the selected molecules contain two tandem copies of the sequence TRMACCA (R = A or G; M = A or C). It can be speculated that these sequences represent single binding sites for one HAT3.1 HD, with a clear preference for A over C at the third position (TRAACCA) when this nucleotide was not fixed. None of these sequences contains the TAAT core recognized by many HDs. To our knowledge, similar sequences were not reported as binding sites for other HDs either. This sequence is also different from the ones recognized by the HDs of the maize PHD-finger protein ZmHox2a (TCCT and GATC, ref *23*). Accordingly, it seems that the HAT3.1 HD has evolved a new binding specificity.

HAT3.1 Binds the Selected Sequence in Vivo. Binding of HAT3.1 to the selected sequence was also tested *in vivo* in a one-hybrid assay in yeast. For this purpose, three copies of an oligonucleotide carrying the HAT3.1 binding site were cloned in tandem in front of the HIS3 reporter gene fused to its own minimal promoter and introduced into the yeast genome. This strain was able to grow in the absence of histidine when transformed with a construct expressing HAT3.1 fused to the GAL4 activation domain, but not when vector expressing only the activation domain was introduced (Figure 3A), suggesting that HAT3.1 is able to bind to the promoter and stimulate the expression of the HIS3 gene. The specificity of this interaction is reflected by the fact that when the HAT3.1 binding site was omitted or replaced by a different sequence not bound by HAT3.1 *in vitro*, growth in the absence of histidine was not recovered (Figure 3A).

C1	gaattctacagtggctgc	**ACCA**	AG	**T**	G	**AACC**	attgtccaggaagcttcatc
C2	gaattctacagtggctgc	**ACCA**	GG	**T**	G	**AACC**	attgtccaggaagcttcatc
C3	gaattctacagtggctgc	**ACCA**	TT	**T**	G	**AACC**	attgtccaggaagcttcatc
C4	gaattctacagtggctgc	**ACCA**	TC	**TAAACC**			attgtccaggaagcttcatc
C7	gaattctacagtggctgc	**ACCA**	CC	**TAAACC**			attgtccaggaagcttcatc
C8	gaattctacagtggctgc	**ACCA**	TC	**TAAACC**			attgtccaggaagcttcatc
C9	gaattctacagtggctgc	**ACCA**	TC	**TAAACC**			attgtccaggaagcttcatc
C12	gaattctacagtggctgc	**ACCA**	GC	**T**	G	**AACC**	attgtccaggaagcttcatc
C13	gaattctacagtggctgc	**ACCA**	TC	**TAAACC**			attgtccaggaagcttcatc
C14	gaattctacagtggctgc	**ACCA**	AG	**T**	G	**AACC**	attgtccaggaagcttcatc
C15	gaattctacagtggctgc	**ACCA**	CC	**TA**	C	**ACC**	attgtccaggaagcttcatc
C16	gaattctacagtggctgc	**ACCA**	GC	**TAAACC**			attgtccaggaagcttcatc
C17	gaattctacagtggctgc	**ACC**	GCC	**TAAACC**			attgtccaggaagcttcatc
C18	gaattctacagtggctgc	**ACCA**	AT	**T**	G	**AACC**	attgtccaggaagcttcatc
C20	gaattctacagtggctgc	**ACCA**	CT	**TAAACC**			attgtccaggaagcttcatc
C21	gaattctacagtggctgc	**ACCA**	AG	**T**	G	**AACC**	attgtccaggaagcttcatc
C22	gaattctacagtggctgc	**ACCA**	AG	**T**	G	**AACC**	attgtccaggaagcttcatc
C49	gaattctacagtggctgc	**ACCA**	TG	**T**	G	**AACC**	attgtccaggaagcttcatc
C50	gaattctacagtggctgc	**ACCA**	TC	**TAAACC**			attgtccaggaagcttcatc
C51	gaattctacagtggctgc	**ACCA**	GG	**T**	G	**AACC**	attgtccaggaagcttcatc
C52	gaattctacagtggctgc	**ACCA**	CG	**TAAACC**			attgtccaggaagcttcatc
C56	gaattctacagtggctgc	**ACCA**	TC	**TAAACC**			attgtccaggaagcttcatc
C57	gaattctacagtggctgc	**ACCA**	AC	**TAAACC**			attgtccaggaagcttcatc
C58	gaattctacagtggctgc	**ACCA**	TC	**TAAACC**			attgtccaggaagcttcatc
C59	gaattctacagtggctgc	**ACCA**	TG	**TAAACC**			attgtccaggaagcttcatc
C60	gaattctacagtggctgc	**ACCA**	AC	**TAAACC**			attgtccaggaagcttcatc
C61	gaattctacagtggctgc	**ACCA**	AT	**TAAACC**			attgtccaggaagcttcatc
C62	gaattctacagtggctgc	**ACCA**	TC	**TAAACC**			attgtccaggaagcttcatc

CONSENSUS	Binding site 1								Binding site 2							
	T	G	C	A	C	C	A	(T/A)	C	T	A	A	A	C	C	A
A				**28**	-	-	**27**	**8**	-	-	**18**	**27**	**28**	-	-	
G				-	-	-	1	4	9	-	10	-	-	-	-	
C				-	**28**	**28**	-	5	**15**	-	-	1	-	**28**	**28**	
T				-	-	-	-	**11**	4	**28**	-	-	-	-	-	
%				100	100	100	96	(39/29)	54	100	64	96	100	100	100	

FIGURE 2: Compilation of the sequences of 28 random clones obtained after cloning the oligonucleotide populations selected by HAT3.1. 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. 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.

In addition, a K50A mutant of HAT3.1 that is unable to efficiently bind DNA (see below) was also unable to recover growth in the absence of histidine (Figure 3A). Binding of HAT3.1 to its target site was also analyzed in yeast cells carrying a fusion of this site to the *lacZ* gene. As shown in Figure 3B, cells that express HAT3.1 fused to the GAL4 activation domain show a 10-fold increase in β-galactosidase activity respective to cells that express the GAL4 activation domain alone or the K50A mutant of HAT3.1. These results indicate that HAT3.1 specifically binds the sequence identified in the selection experiments *in vivo*.

Nucleosides from the Selected Region Are Required for HAT3.1 Binding. A more detailed picture of the interaction of HAT3.1 with DNA was obtained using missing nucleoside experiments. In these experiments, the DNA-binding protein is incubated with oligonucleotides that have been previously subjected to hydroxyl radical attack to generate a population of DNA molecules with single nucleosides removed at different positions of each strand. Nucleosides important for protein binding can then be monitored, since molecules with gaps at these positions are underrepresented in the bound DNA fraction. Figure 4A shows the patterns obtained after analyzing the populations of molecules that either were bound by HAT3.1 or remained free. The pattern obtained when the top strand (the one containing the ACCA sequences) was labeled indicates that nucleosides comprising the TRMACCA sequences are required for binding, while the TC dinucleotide located between them and arm sequences are not important. A closer examination shows positions relatively more important (TGCACCA in the 5′ region, TAAACCA in the 3′ region). In the bottom strand, a more uniform pattern is evident since regions complementary to both RMACCA sequences are required (Figure 4A). In the 3′ region, an A complementary to T in TAAACCA also shows interference. The observed patterns reinforce the view that the HAT3.1 binding site obtained after selection is a tandem duplication of the sequence TRMACCA.

Nucleotides at Invariant Positions Are Important for HAT3.1 Binding. The fact that two similar 7-bp sequences are present in the selected molecules and protected by HAT3.1 binding may indicate that HAT3.1 recognizes DNA as a dimer. To analyze the effect of mutations in each of the repeated sequences, we mutated the respective CC to TA. As shown in Figure 5, mutation of CC in the 5′ region (oligonucleotide BS-8) still allowed considerable binding, while a very faint shifted band was observed when the 3′ CC (BS-9) was mutated. Accordingly, we speculate that the 3′ region is preferentially bound by HAT3.1, which is not

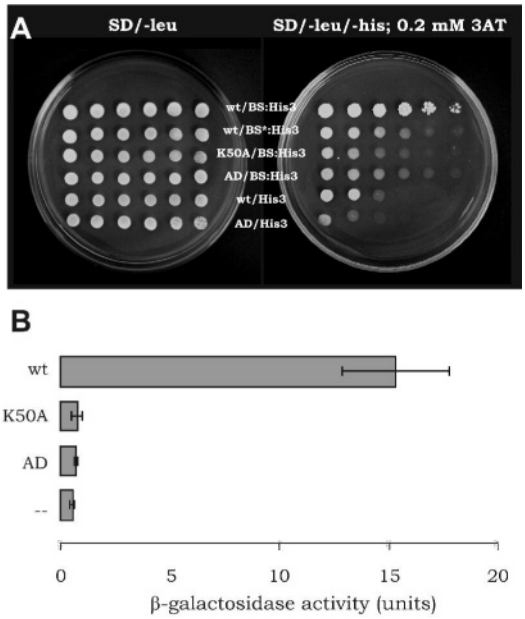


FIGURE 3: Binding of HAT3.1 to its target sequence *in vivo* in yeast. (A) Different dilutions of an overnight culture of yeast strains carrying a minimal promoter−HIS3 fusion either preceded by the HAT3.1 target sequence (BS:His3), a sequence bound by the ATH1 HD (BS*:His3; ref *37*), or alone (His3), were spotted onto plates containing minimal medium supplemented as necessary in the presence and absence of histidine (left and right panels, respectively). The strains had been previously transformed with constructs expressing a portion of HAT3.1 or a K50A mutant fused to the GAL4 activation domain (wt or K50A, respectively) or with the pGADT7 plasmid without insert (AD, expressing only the GAL4 activation domain). (B) *β*-Galactosidase activity of yeast cells carrying a fusion of the HAT3.1 target sequence to the *lacZ* gene. Cells were transformed with different plasmids expressing GAL4 activation domain fusions as described in (A). The mean activity value (±SD) of three independent measurements is shown.

unexpected considering the fact that it is the one that has been almost completely selected, while the 5′ repeat is formed by 3 invariable arm residues of the oligonucleotide used for selection.

Since only one shifted band was observed under all circumstances, it can be concluded that either HAT3.1 binds as a monomer to either one of the sequences, or it binds as a dimer to both. Since molecules with only the TAAACCA site are efficiently bound by HAT3.1, it can be concluded that the interaction of only one HD with DNA is enough for specific binding. This is supported by the fact that only two shifted bands were observed when a mixture of HAT3.1 proteins of two different sizes (i.e., a GST fusion and the corresponding moiety obtained after factor Xa treatment) were used (data not shown). The presence of mixed dimers would have produced a shifted band of intermediate mobility.

Since the 3′ region is preferentially bound by the HAT3.1 HD, we focused on this region to analyze the importance of different nucleotides for binding. This was performed using oligonucleotides with point mutations at specific positions as probes (Figure 5). The main effect on binding was observed when any of the C of the ACCA sequence were changed (oligonucleotides BS-5 and BS-6). The effect of these point mutations was similar to the one observed when both C were modified (BS-9). Mutations of T and A at positions 1 and 4 (BS-1 and BS-4) produced a less but still important decrease in binding, while the change of A at position 7 (BS-7) was almost ineffective (Figure 5). For positions 2 and 3 (AA, not absolutely conserved), oligonucleotides with individual differences at these positions from clones obtained after the binding site selection experiment were used. These oligonucleotides were bound almost as efficiently as the oligonucleotide containing AA (not shown).

Helix III Residues Are Required for HAT3.1 Binding to DNA. Assuming that TAAACCA (TGGTTTA in the complementary strand) is the best binding site for HAT3.1, the question arises on how the HAT3.1 HD interacts with this sequence. A comparison of the amino acid sequence of the HAT3.1 HD with the Antennapedia HD (usually taken as a model) indicates that only 14 positions (23%) are identical (Figure 6). Among these, 11 positions are considered almost invariable or highly conserved in all HDs, including L16, L38, V45, W48, and F49, which conform the hydrophobic core involved in maintaining the tertiary structure. The presence of M40 instead of leucine can be considered as a conservative substitution. Important nonconserved positions include those of the N-terminal arm and H51, which are known to directly participate in DNA binding. The N-terminal arm of HAT3.1 does not contain any of the residues known to interact with DNA in other HDs, namely, R3, R5, R7, or Y/F8 (*3*), making it unlikely that HAT3.1 uses this portion of the HD for binding. From this, it can be concluded that the HAT3.1 HD most likely folds like a *bona fide* HD, but protein−DNA interactions may differ from those established by the rest of HDs.

Site-directed mutagenesis of helix III residues located at four HD positions usually involved in establishing specific contacts with DNA (positions 47, 50, 51, and 54) was used to investigate the role of these amino acids in the HAT3.1 HD. Inclusion of alanine at either position 50, 51, or 54 produced proteins with very low affinity for the target sequence (Figure 7), suggesting that residues at these positions establish important contacts with DNA. The presence of alanine instead of asparagine at position 47 produced a smaller but still significant decrease in binding. Residues 47, 50, and 51 were also changed to those most frequently present at these positions in other HDs. Introduction of V47 or Q50 almost completely abolished binding (Figure 7). The effect of V47 may be due to steric hindrance due to the introduction of a hydrophobic moiety in addition to the loss of contacts that may be established by N47. In the case of Q50, it is more likely that the contacts established by K50 are lost with this replacement, since binding is similar to the one observed with alanine at this position. The inclusion of asparagine at position 51, in turn, had only a minor effect (Figure 7). It can be speculated that H51 participates in the interaction with DNA in a way that is similar to that observed for N51 in other HDs, although the presence of histidine seems to be optimal in the context of the HAT3.1 HD.

We have also constructed HDs with double mutations (N47V/W54A and K50Q/W54A) and the triple mutant N47V/H51N/W54A. Binding of these HDs to the HAT3.1 binding site was negligible (Figure 7 and results not shown).

The Protonation State of H51 Determines the Efficiency of HAT3.1 Binding to DNA. Histidines in proteins change their protonation state at near neutral pH (pK_a about 6.0−7.0). It was then analyzed if binding of the HAT3.1 HD is pH-dependent. Figure 8 shows that the amount of bound

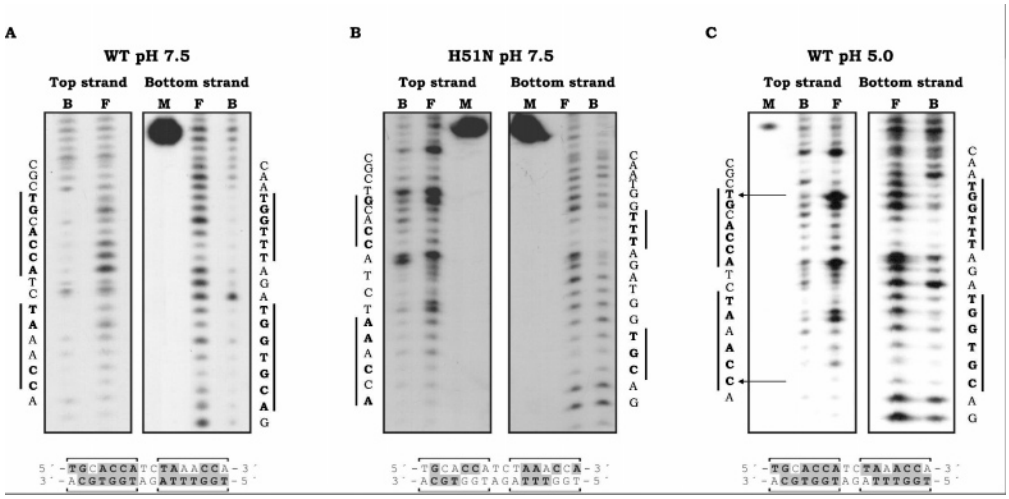
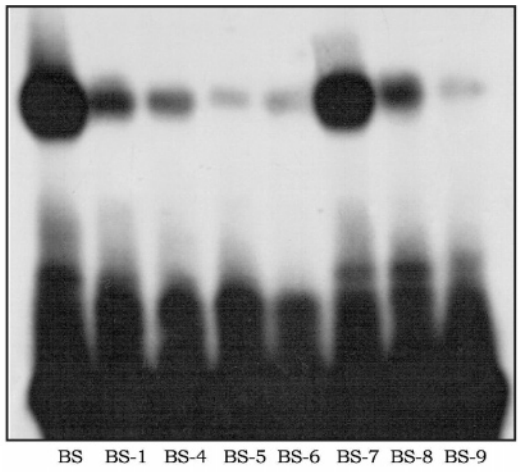


FIGURE 4: Hydroxyl radical interference assays of the binding of HAT3.1 to DNA. An oligonucleotide containing the HAT3.1 consensus binding sequence previously labeled in the 3′-end of either strand (*Hin*dIII or *Xba*I sites) was subjected to hydroxyl radical attack before binding to HAT3.1 (A) or H51N-HAT3.1 (B) at pH 7.5. Free (F) and bound (B) DNA were separated and analyzed. Lanes M indicate a 42-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, the sequence of the binding site is shown and the positions that interfere with binding are shown in bold and shadowed. (C) A similar experiment using HAT3.1 at pH 5.0.



OLIGONUCLEOTIDE	SEQUENCE
BS	TGCACCATCTAAACCA
BS-1	TGCACCATCGAAACCA
BS-4	TGCACCATCTAATCCA
BS-5	TGCACCATCTAAAGCA
BS-6	TGCACCATCTAAACGA
BS-7	TGCACCATCTAAACCT
BS-8	TGCATAATCTAAACCA
BS-9	TGCACCATCTAAATAA

FIGURE 5: Binding of HAT3.1 to oligonucleotides containing variants of the consensus sequence. Binding of HAT3.1 to 26-mer synthetic double-stranded oligonucleotides with single changes in the TAAACCA sequence (BS-1 to BS-7) was analyzed by an electrophoretic mobility shift assay. Oligonucleotides with CC to TA changes within each of the ACCA sequences (BS-8 and BS-9) were also tested. The table describes the relevant sequences of the different oligonucleotides that were used.

DNA increased as the pH of the binding reaction was lowered from 8.2 to 4.9, with an inflection point between 6.0 and 6.5, most likely reflecting that protonation of a histidine improves the interaction with DNA. In addition, replacement of H51 with asparagine produced a protein nonresponsive to pH changes (Figure 8). It can then be speculated that the protonation state of H51 modulates the DNA binding efficiency of the HAT3.1 HD.

In known HD structures, N51 establishes a bidentate H-bond with A at the third position of TAAT or with an A present in other recognition sequences (*11*). Since asparagine can partially replace H51, it can be postulated that H51 may also bind an A located within the sequence TAAACCA. In principle, histidine could make H-bonds with Nδ and Nϵ, as donors if they are protonated. Interaction with adenine may be through N7 as acceptor, so that only one H-bond could be established. The remaining N of H51 may either make specific contacts with another nucleotide of the recognition sequence or interact through its positive charge with the phosphate backbone. It can also be envisaged that H51 may interact with another amino acid residue thus favoring the interaction of this residue with DNA or stabilizing the HD structure.

We have studied the interaction of H51N-HAT3.1 with DNA using missing nucleoside experiments. We noticed that A4 in the top strand of the recognition sequence TRMACCA shows less relative interference than the CC dinucleotide located 3′ to it and the adjacent A or C in the sequence TRMACCA (Figure 4B). With the wild-type protein at pH 7.5, A4 shows similar interference as the preceding A in TAAACCA and stronger interference than C in TGCACCA (Figure 4A). This may indicate that a contact at A4 is lost upon replacement of H51 by asparagine. In addition, A3 in TAAACCA shows higher relative interference for binding to the H51N mutant.

Examination of the interference pattern obtained after binding of the wild-type protein at pH 5.0 (Figure 4C) indicates that the relative importance of A4 for binding is increased at this pH. We postulate that H51 makes contacts with A4 of the recognition sequence and that these contacts are stabilized upon protonation of this residue, thus explaining the preference of histidine over asparagine at position 51 within the context of the HAT3.1 HD.

HAT3.1 HD Mutants Have Different Preferences for Modified Target Sequences. Some of the site-directed

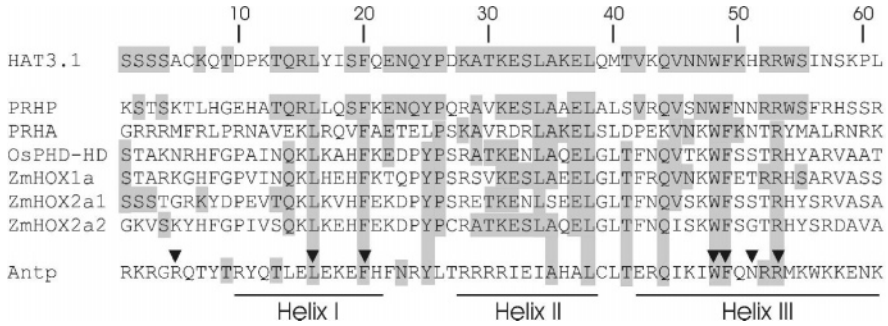


FIGURE 6: Alignment of the HD sequences of HAT3.1 and other HD proteins. The sequence of the HAT3.1 HD was compared with those of other plant PHD-finger HD proteins (parsley PRHP, accession number L21975; *Arabidopsis* PRHA, accession number NM_119140; rice OsPHD-HD, accession number NM_001063749; maize ZmHOX1a and ZmHOX2a, accession numbers X67561 and X89760, respectively) or with the Antennapedia HD. Shaded residues are those present in HAT3.1 and at least one of the remaining HDs. Numbers at the top indicate residue position within the HD. The bars below the sequences indicate the position of the three helices. Arrowheads indicate positions considered invariant among HDs.

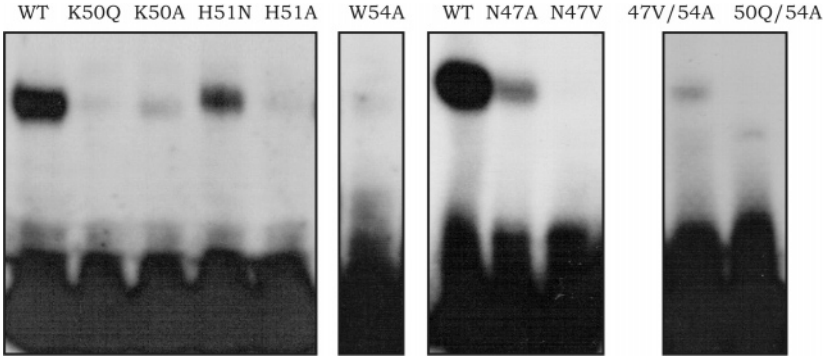


FIGURE 7: Binding of HAT3.1 HD mutants at positions 47, 50, 51, and 54 to DNA. HAT3.1 mutants with different changes at HD positions 47, 50, 51, or 54, as indicated, were analyzed by electrophoretic mobility shift assays for binding to a 26-mer oligonucleotide containing the sequence TCTAAACCATT.

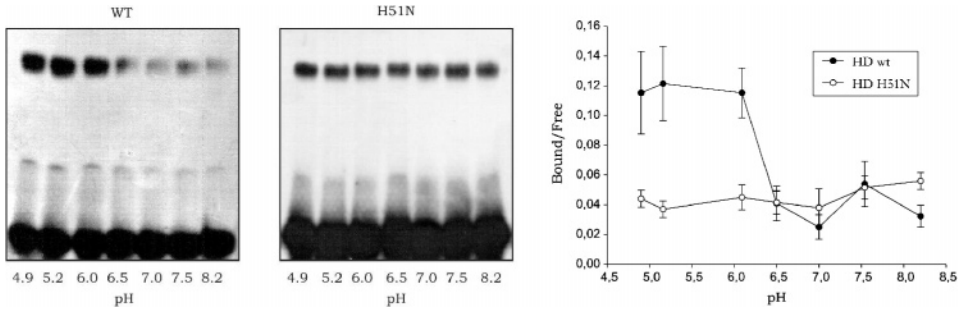


FIGURE 8: Effect of pH on the interaction of the HAT3.1 HD with DNA. Binding of the HAT3.1 HD or its H51N mutant to an oligonucleotide containing the consensus binding site was performed in buffers containing a mixture of HEPES/MOPS (0.2 M each) brought to different pH (indicated below the figure) using NaOH. The right panel shows a quantitative analysis of binding at different pH obtained after scintillation counting of the radioactivity present in the respective bands. Bars represent SD of three different experiments.

mutants at positions 47, 50, 51, and 54 showed, in addition to reduced binding efficiency, changes in their relative preferences for the different binding sequences respective to the wild-type HAT3.1 HD (Figure 9). H51N-HAT3.1 was relatively less affected by an A to T change at position 4 of the recognition sequence (oligonucleotide BS-4), that introduces the TAAT core bound by most HDs (Figure 9). This is an additional indication that H51 establishes closer contacts with A4 than N51. In addition, H51N-HAT3.1 binding was almost unaffected by a CC to TA change in the 5′ part of the sequence (BS-8; Figure 9). Since a similar change in the 3′ region (BS-9) does produce a significant effect, this may reflect that the mutant has a higher preference for the second binding site (i.e., for AA instead of GC at positions 2 and 3). It can be assumed, then, that N51 establishes preferential contacts with A located at positions 2 or 3, while H51 seems to interact mainly with A at position 4.

Binding of the K50A mutant was not affected by a C to G change at position 5 (BS-5) or CC to TA changes at either part of the repeated binding unit (BS-8 and BS-9; Figure 9). This observation can be explained assuming that K50 interacts with one or both members of the CC dinucleotide, while A50 does not. This lack of interaction explains the significant decrease in binding to the nonmutated sequence and the small effect observed when positions 5 and 6 are changed.

For the W54A mutant, no effect on binding was observed with a change at position 4 (BS-4; Figure 9). As discussed above, this may reflect that W54 interacts with A at this position. A somewhat unexpected behavior was evident for H51A- and N47V-HAT3.1 since they showed a significant

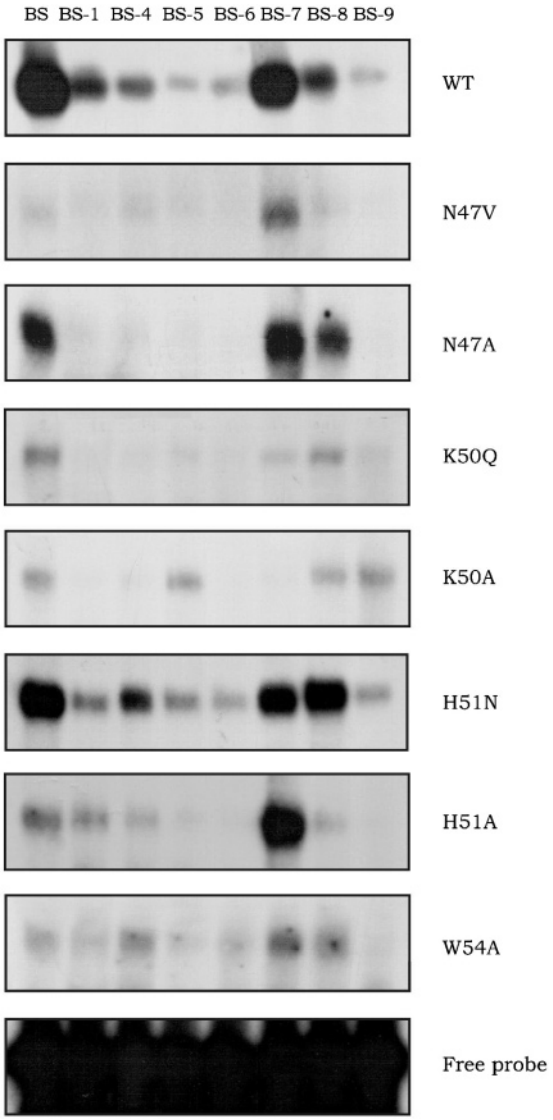


FIGURE 9: Binding of HAT3.1 mutants to oligonucleotides containing different binding sequences. Proteins with different point mutations at defined positions of the HAT3.1 HD were analyzed by electrophoretic mobility shift assays for binding to oligonucleotides containing variants of the consensus sequence. Only the part of each gel showing the shifted band is shown. The amount of free probe is shown only for the nonmutated (WT) protein for comparison. The amount of protein used in each case was as follows: WT, K50A, and H51N, 20 ng; N47A and K50Q, 40 ng; H51A and W54A, 50 ng; N47V, 60 ng. The sequences of the different oligonucleotides are described in Figure 5.

preference for BS-7 when compared to the nonmutated target site. Increased binding most likely reflects the establishment of new contacts with the modified nucleotide position. While V47 may be itself involved in these contacts, it is unlikely that the side chain of A51 can establish specific contacts with DNA. It can be speculated that upon removal of histidine a new interaction is formed between another HD residue and T7 of the modified target sequence.

The double and triple mutants of HAT3.1, in turn, did not show significant interaction with any of the sequences that were tested. It is noteworthy that the triple mutant N47V/H51N/W54A has the same residues at positions 47 to 54 as the PITX2 HD, which efficiently binds the sequence TAATCC (*34*).

DISCUSSION

Arabidopsis HAT3.1 was discovered during a screening of an expression library with a promoter fragment of the *cab-E* gene (*16*). However, once expressed, the protein was unable to show specific binding to DNA, displaying affinity for any DNA sequence larger than 50 bp (*16*). Since the HAT3.1 HD is highly divergent in sequence, the question remained open if it was able to function as an independent and specific DNA binding unit. In this report we show that the HAT3.1 HD selects specific sequences from a random population of DNA molecules. This analysis suggests that HAT3.1 binds to sequences of the type TRMACCA with a preference for A at the second and third positions. In combination with expression data from plants with changes in HAT3.1 expression levels, the selected sequence will be useful to identify HAT3.1 target genes. In fact, the *cab-E* promoter region used to screen the expression library contains the sequence TACACCA (*16*), suggesting that this portion was responsible for the identification of a HAT3.1 cDNA clone. Even if the selected consensus contains two tandem copies of this sequence, HAT3.1 is able to interact with an oligonucleotide containing one TAAACCA, indicating that a single HD is capable of efficient DNA binding. Although binding as a dimer cannot be ruled out, we found no evidence of dimer formation by the HAT3.1 HD. The fact that two tandem copies of the same sequence were obtained most likely reflects that an increase in the number of specific binding sites per molecule produces an increase in binding efficiency.

Most of the conserved residues found in the HAT3.1 HD correspond to those that conform the hydrophobic core that maintains the tertiary structure of the three-helix module. Accordingly, its recognition mechanisms must follow similar rules as those observed in other HDs. This is supported by the fact that residues at positions 47, 50, 51, and 54, involved in DNA binding in other HDs (*3*, *11*), are also important in HAT3.1. The study of the interaction of wild-type and mutagenized HAT3.1 HDs with the selected and modified target sequences suggested that K50, H51, and W54 interact with specific nucleotides as depicted in Figure 10A. Central to this discussion is the interaction of K50 with nucleotides 5 and 6, since lysine is present at this position in other HDs. K50 in the bicoid HD is responsible for the specificity of binding to TAATCC, as opposed to TAATTG, contacting the GG dinucleotide complementary to CC (*35*; Figure 10B). An interaction of K50 in the HAT3.1 HD with nucleotides at positions 5 and 6 of its target site, which are also C, can be proposed from our experiments. This assignment allows the direct comparison of the sequence TAATCC with the HAT3.1 target sequence TAAACC (Figure 10). In bicoid, N51 establishes contacts with the AA dinucleotide of TAATCC, while R54 binds to nucleotides complementary to T4 and C5 (*35*). This fits well with results from HAT3.1 mutants showing that N51 influences binding to positions 2 and 3, while residue 54 is involved in determining the fourth position. The importance of W54 in the HAT3.1 HD is highlighted by the fact that its mutation to alanine almost completely abolishes binding, even if other HDs contain alanine at this position (*8*). As an example, the PITX2 HD contains V47, K50, N51, and A54 and binds to TAATCC. In this case, hydrophobic contacts with the fourth position

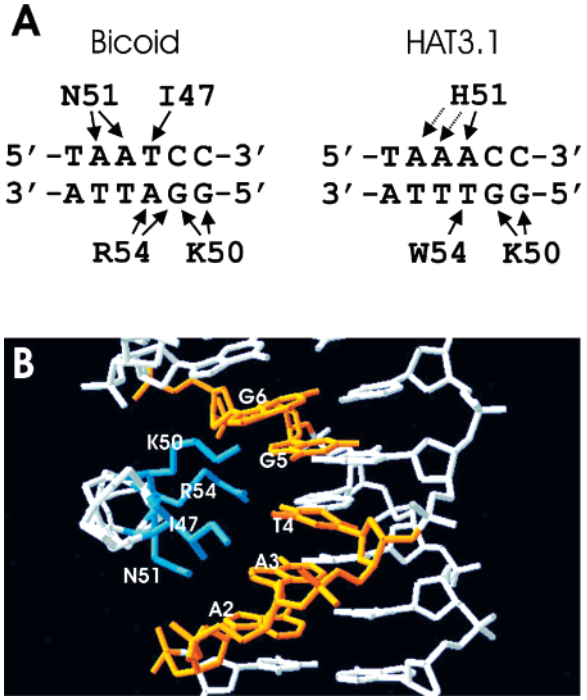


FIGURE 10: A model showing the interaction of the HAT3.1 HD with DNA. (A) Contacts established between helix III residues of the bicoid HD and specific bases of its target DNA, determined from the corresponding solution structure (*35*), are shown by solid arrows in the left. In the right part, putative contacts of the HAT3.1 HD with the sequence TAAACC, deduced from the experiments described here, are shown. Solid arrows indicate contacts deduced from our experiments. Broken arrows indicate other possible contacts that may be established by H51. (B) A model of the interaction of helix III of the bicoid HD with DNA. The side chains of residues 47, 50, 51, and 54 are shown in blue. Specific nucleotides from the TAATCC target sequence are shown in orange. G5 and G6 are complementary to the CC dinucleotide.

(T) are established by V47 (*34*). A similar interaction could be observed in bicoid and many other HDs (*8*, *11*, *35*). Although mutation of N47 to alanine produced a significant decrease in binding, it is unlikely that N47 in HAT3.1 can establish similar contacts as V47. Indeed, introduction of V47 in the HAT3.1 HD produced a protein with very low affinity for both TAAACC and TAATCC (BS and BS-4, respectively). Consequently, it can be proposed that W54 may be one of the determinants of binding to TAAA instead of TAAT, as most HDs do, and that its presence would preclude the interaction with TAAT even in the presence of V47. Binding to the half-sequence TAAATG instead of TAATTG has been ascribed to combinatorial interactions of residues at positions 47 and 54 (phenylalanine and threonine, respectively) in HD-Zip IV family proteins (*36*). For HAT3.1, a more complex situation can be envisaged, however, since the double-mutant N47V/W54A and the triple mutant N47V/H51N/W54A were unable to bind DNA containing either TAAA or TAAT (not shown).

The HAT3.1 HD represents an extreme case of divergence among HD structures. The introduction of W54 and H51 represents an innovation since, to our knowledge, these residues are not present in other HDs at these positions. While the presence of tryptophan at the more variable position 54 likely originated a new binding specificity (in fact, positions 50 and 54 are usually regarded as the main source of change in binding site preferences; ref *12*), the presence of H51 is more intriguing, since asparagine is almost invariably present at this position (*3*). An interesting question is why histidine is not present in other HDs if it can efficiently replace asparagine. It is likely that the presence of histidine either requires compensatory changes in other parts of the molecule or represents an advantage only in the context of the HAT3.1 HD. For example, histidine may interact with the side chain(s) of other amino acid(s), thus stabilizing the HD structure. This combinatorial effect is reflected by the fact that the introduction of residues that are usually present in other HDs (V47, Q50, A54), in several combinations, caused a significant decrease in binding efficiency. Nevertheless, the example of HAT3.1 indicates that the HD has a considerable degree of plasticity and that it represents a useful source of evolutionary changes in protein−DNA interaction mechanisms.

ACKNOWLEDGMENT

We acknowledge the RIKEN BRC Experimental Plant Division (Japan) for sending us the HAT3.1 full-length cDNA clone.

REFERENCES

1. Gehring, W. J. (1987) Homeoboxes in the study of development, *Science 236*, 1245−1252.
2. Levine, M., and Hoey, T. (1988) Homeobox proteins as sequence-specific transcription factors, *Cell 55*, 537−540.
3. Gehring, W. J., Affolter, M., and Bürglin, T. (1994) Homeodomain proteins, *Annu. Rev. Biochem. 63*, 487−526.
4. Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., and Wüthrich, K. (1989) The structure of the *Antennapedia homeodomain determined by* NMR spectroscopy in solution: comparison with prokaryotic repressors, *Cell 59*, 573−580.
5. Qian, Y. K., 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.
6. 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.
7. Otting, G., Qian, Y. Q., Müller, M., Affolter, M., Gehring, W. J., and Wüthrich, K. (1988) Secondary structure determination for the *Antennapedia homeodomain by* nuclear magnetic resonance and evidence for a helix-turn-helix motif, *EMBO J. 7*, 4305−4309.
8. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions, *Cell 63*, 579−590.
9. Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J., and Wüthrich, K. (1990) Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution, *EMBO J. 9*, 3085−3092.
10. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions, *Cell 67*, 517−528.
11. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K. (1994) Homeodomain-DNA recognition, *Cell 78*, 211−223.
12. Laughon, A. (1991) DNA binding specificity of homeodomains, *Biochemistry 30*, 11357−11367.
13. Wilson, D. S., Sheng, G., Jun, S., and Desplan, C. (1996) Conservation and diversification in homeodomain-DNA interactions: a comparative genetic analysis, *Proc. Natl. Acad. Sci. U.S.A. 93*, 6886−6891.

14. Palaniswamy, S. K., James, S., Sun, H., Lamb, R. S., Davuluri, R. V., and Grotewold, E. (2006) AGRIS and AtRegNet, a platform to link cis-regulatory elements and transcription factors into regulatory networks, *Plant Physiol. 140*, 818−829.
15. Chan, R. L., Gago, G. M., Palena, C. M., and Gonzalez, D. H. (1998) Homeoboxes in plant development, *Biochim. Biophys. Acta 1442*, 1−19.
16. Schindler, U., Beckmann, H., and Cashmore, A. R. (1993) HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region, *Plant J. 4*, 137−150.
17. Bienz, M. (2006) The PHD finger, a nuclear protein-interaction domain, *Trends Biochem. Sci. 31*, 35−40.
18. Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF, *Nature 442*, 91−95.
19. Mellor, J. (2006) It takes a PHD to read the histone code, *Cell 126*, 22−24.
20. Plesch, G., Stormann, K., Torres, J. T., Walden, R., and Somssich, I. E. (1997) Developmental and auxin-induced expression of the Arabidopsis prha homeobox gene, *Plant J. 12*, 635−647.
21. Klinge, B., Uberlacker, B., Korfhage, C., and Werr, W. (1996) ZmHox: a novel class of maize homeobox genes, *Plant Mol. Biol. 30*, 439−453.
22. Uberlacker, B., Klinge, B., and Werr, W. (1996) Ectopic expression of the maize homeobox genes ZmHox1a or ZmHox1b causes pleiotropic alterations in the vegetative and floral development of transgenic tobacco, *Plant Cell 8*, 349−362.
23. Kirch, T., Bitter, S., Kisters-Woike, B., and Werr, W. (1998) The two homeodomains of the ZmHox2a gene from maize originated as an internal gene duplication and have evolved different target site specificities, *Nucleic Acids Res. 26*, 4714−4720.
24. Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M., Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., and Shinozaki, K. (2002) Functional annotation of a full-length Arabidopsis cDNA collection, *Science 296*, 141−145.
25. 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.
26. 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 Expression Purif.* **13**, 97−103.
27. 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.
28. 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.
29. 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.
30. Meijer, A. H., Ouwekerk, P. B. F., and Hoge, H. C. (1998) Vectors for transcription factor cloning and target site identification by means of genetic selection in yeast, *Yeast 14*, 1407−1416.
31. Gietz, D., St. Jean, A., Woods, R. A., and Schiestl, R. H. (1992) Improved method for high efficiency transformation of intact yeast cells, *Nucleic Acids Res. 20*, 1425.
32. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1987) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York.
33. 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.
34. Chaney, B. A., Clark-Baldwin, K., Dave, V., Ma, J., and Rance, M. (2005) Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome, *Biochemistry 44*, 7497−7511.
35. Baird-Titus, J. M., Clark-Baldwin, K., Dave, V., Caperelli, C. A., Ma, J., and Rance, M. (2006) The solution structure of the native K50 bicoid homeodomain bound to the consensus TAATCC DNA-binding site, *J. Mol. Biol. 356*, 1137−1151.35.
36. Tron, A. E., Bertoncini, C. W., Palena, C. M., Chan, R. L., and Gonzalez, D. H. (2001) Combinatorial interactions of two amino acids with a single base pair define target site specificity in plant dimeric homeodomain proteins, *Nucleic Acids Res. 29*, 4866−4872.
37. Viola, I. L., and Gonzalez, D. H. (2006) Interaction of the BELL-like protein ATH1 with DNA. A role of homeodomain residue 54 in specifying the different binding properties of BELL and KNOX proteins, *Biol. Chem. 387*, 31−40.

BI602657E