doi:10.1006/jmbi.2001.4563 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2001) 308, 39–47

JMB

Positively Charged Residues at the N-terminal Arm of the Homeodomain are Required for Efficient DNA Binding by Homeodomain-leucine Zipper Proteins

Claudia M. Palena, Adriana E. Tron, Carlos W. Bertoncini Daniel H. Gonzalez and Raquel L. Chan*

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

Plant homeodomain-leucine zipper proteins, unlike most animal homeodomains, bind DNA efficiently only as dimers. In the present work, we report that the deletion of the homeodomain N-terminal arm (first nine residues) of the homeodomain-leucine zipper protein Hahb-4 dramatically affects its DNA-binding affinity, causing a 70-fold increase in dissociation constant. The addition of the N-terminal arm of *Drosophila* Antennapedia to the truncated form restores the DNA-binding affinity of dimers to values similar to those of the native form. However, the Antennapedia N-terminal arm is not able to confer increased binding affinity to monomers of Hahb-4 lacking the leucine zipper motif, indicating that the inefficient binding of monomers must be due to structural differences in other parts of the molecule. The construction of proteins with modifications at residues 5 to 7 of the homeodomain suggests strongly that positively charged amino acids at these positions play essential roles in determining the DNA-binding affinity. However, the effect of mutations at positions 6 and 7 can be counteracted by introducing a stretch of positively charged residues at positions 1 to 3 of the homeodomain. Sequence comparisons indicate that all homeodomain-leucine zipper proteins might use contacts of the N-terminal arm with DNA for efficient binding. The occurrence of a homeodomain with a DNA-interacting N-terminal arm must then be an ancient acquisition in evolution, earlier than the separation of lines leading to metazoa, fungi and plants.

Keywords: DNA-binding; Hd-Zip protein; homeodomain; leucine zipper; N-terminal arm

*Corresponding author

Introduction

The homeodomain (HD) is a 61 amino acid residue protein motif found in eukaryotic transcription factors generally involved in the regulation of developmental processes.[1-3] It folds into a characteristic three-helix structure that interacts specifically with DNA.[36] Helices II and III form a structure that resembles the helix-turn-helix motif found in many prokaryotic transcription factors. Helix III (the recognition helix) fits into the major groove of DNA, making extensive contacts with specific bases and the sugar-phosphate backbone.[7-10]. Helices I and II are roughly perpendicular to helix III. These two helices are antiparallel and pack against helix III, forming a hydrophobic core that determines the overall structure of the HD. Only two residues in helix II and one residue in the loop between helices I and II make contacts with DNA. These contacts are thought to be important in fixing the recognition helix in the correct position in DNA. C-terminal to the three-helix structure, a fourth helix has been described for some HDs,[4,11] whilst an extended helix III is present in others.[7,9] The N-terminal portion of the HD (first nine amino acid residues) forms a disordered arm that protrudes from helix I and orients towards DNA, making several specific contacts within the minor groove.[7-10]

Abbreviations used: HD, homeodomain; Hd-Zip, homeodomain-leucine zipper; GST, glutathione *S*-transferase; Antp, Antennapedia.

E-mail address of the corresponding author: rchan@fbcb.unl.edu.ar

0022-2836/01/010039–9 $35.00/0

A group of about seven amino acid residues is conserved in almost all HDs recognized so far. Four of these are hydrophobic amino acid residues that participate in maintaining the structure of the three-helix core, two are base-contacting amino acid residues from helix III, and the remaining one is Arg5, a DNA-contacting residue from the N-terminal arm.[3] In spite of the resemblance in structure between the HD and the helix-turn-helix motif, a striking difference is that HDs usually bind DNA as monomers with high affinity.[12,13] This fact may be explained by the presence of extended contacts along the recognition helix and, especially, by the stabilizing effect of contacts made by the N-terminal arm. Accordingly, the deletion of this arm results in a decrease in affinity of approximately 100-fold.[14,15]

HDs are present in almost every eukaryotic organism that has been investigated. In plants, several families of HD proteins have been described.[16] One of these families, named homeodomain-leucine zipper (Hd-Zip), comprises proteins with a typical leucine zipper motif adjacent to the C-terminal end of the HD.[17,18] As expected, these proteins bind DNA as dimers.[19] The removal of the leucine zipper, or the introduction of extra amino acid residues between the HD and the zipper, significantly reduces binding affinity, indicating that the leucine zipper is responsible for the correct positioning of the HD for efficient binding.[19] The analysis of binding at different protein concentrations suggests that dimer formation is a prerequisite for DNA binding.[20] One of the questions that immediately arise upon these observations is that of why the HD of Hd-Zip proteins is unable to bind DNA as a monomer, whilst other HDs (notably those animal HDs most related in sequence) do. An analysis of residues important for DNA binding indicates that two residues that have been implicated in binding, namely Arg3 and Arg5, are not present in most Hd-Zip proteins. This has prompted some authors to conclude that the inefficient binding of monomers would be the result of the lack of interactions between the N-terminal arm and DNA.

We have recently analyzed the interaction of the Hd-Zip protein Hahb-4 with DNA, which represents the most divergent member of the Hd-Zip I subclass reported up to now.[20] Since this protein, unlike other Hd-Zip proteins, contains an Arg at position 5 of the HD, we became interested in investigating the role of the N-terminal arm in DNA binding. We show here that, although Arg5 can be mutated to Lys without any significant effect, the presence of the N-terminal arm is required for efficient binding by dimers. We also show that binding efficiency is related directly with the presence of positively charged amino acid residues within this segment. Accordingly, the HD of Hd-Zip proteins resembles other HDs in this respect, and the lack of efficient binding by monomers may be the consequence of more subtle differences in structure.

Results

Deletion of the N-terminal arm of the HD affects Hahb-4 binding to DNA

The deletion of the first seven amino acid residues of the HD of the Hd-Zip protein Hahb-4 had a profound effect on DNA-binding efficiency (Figure 1(a)). A detailed analysis of the binding of native Hahb-4 and its truncated form (ΔN-Hahb-4) to oligonucleotides bearing different variations of the preferred sequence 5′-CAAT(A/T)ATTG-3′ suggested that these proteins had the same binding preferences (not shown). This result indicates that the overall structure of the HD is not altered in the truncated protein, and that the reduced affinity probably arises from the fact that some contacts between the protein and DNA have been lost. Binding analysis at different protein concentrations yielded an overall dissociation constant (K_{12}) of 9.03×10^{-13} M^2 (Figure 2(a) and (c)). Assuming that dimer formation is not affected in the mutant, this represents a 70-fold reduction in binding affinity of the dimer for DNA, an effect similar to that observed with ftz (130-fold) and Msx (>30-fold) for the binding of monomers with truncated N-terminal arms.[14,15]

Since the experiments were performed with proteins fused to glutathione-*S*-transferase (GST) in their N-terminal part, we have considered the possibility of steric hindrance due to the presence of a folded domain immediately adjacent to the HD as the cause of inefficient binding. UV-photocrosslinking studies indicated, however, that two molecules of the truncated protein are able to interact with one DNA molecule possessing a single pseudopalindromic DNA-binding site, just as native Hahb-4 does (not shown). In addition, the removal of the GST moiety did not reconstitute the binding efficiency of the truncated protein (not shown). Indeed, a further decrease in affinity, relative to the form fused to GST, was evident. This result indicates that the N-terminal arm does not function solely as a flexible linker between the HD and the rest of the protein. It seems more likely that positive interactions with DNA are lost upon removal of these seven amino acid residues.

The N-terminal arm of Antennapedia restores the DNA-binding affinity to the truncated Hahb-4 HD

If the N-terminal arm of Hahb-4 is required for efficient DNA binding, we reasoned that the addition of a functional N-terminal arm from another HD could restore the DNA-binding affinity to the truncated protein. We then constructed a chimeric protein that contained the first 11 residues of the *Drosophila* Antennapedia (Antp) HD, flanked by Hahb-4 sequences (Antp-Hahb-4). This protein bound DNA with an affinity similar to that of native Hahb-4 (Figure 1(b)). From the analysis of the binding of Antp-Hahb-4 to DNA at different

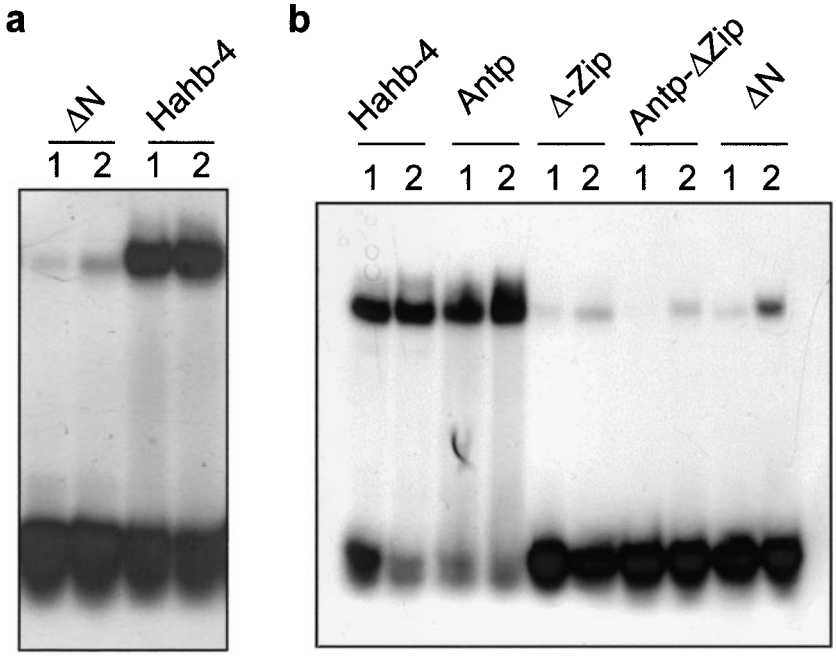

Figure 1. (a) Deletion of the N-terminal arm of the HD affects the DNA-binding affinity of Hahb-4. Two different amounts (50 and 100 ng in lanes 1 and 2, respectively) of either Hahb-4 or ΔN-Hahb-4 (ΔN) were analyzed by an electrophoretic mobility shift assay for binding to a 24-mer synthetic double-stranded oligonucleotide containing the preferred sequence CAATAATTG. (b) The N-terminal arm of the Antp HD restores the DNA-binding affinity of truncated Hahb-4 dimers. Binding of different forms of Hahb-4 with modifications at the N-terminal arm of the HD or leucine zipper deletions were analyzed for DNA-binding as described. Antp, Antp-Hahb-4; Δ-Zip, Hahb-4 with the complete leucine zipper removed; Antp-ΔZip, Antp-Hahb-4 with the complete leucine zipper removed; ΔN, ΔN-Hahb-4. Lanes 1 and 2 contain 50 and 100 ng of protein, respectively.

protein concentrations a K_{12} of 1.28×10^{-14} M^2 was calculated (Figure 2(b) and (c)). This value closely resembles the K_{12} of native Hahb-4 (1.31×10^{-14} M^2), indicating that the lost interactions have been restored. In this context, a pertinent question is if the presence of the N-terminal arm of Antp, which binds DNA as a monomer, confers the ability of efficient binding to Hahb-4 monomers. The form of the binding curve shown in Figure 2(c) suggests, however, that either dimer formation is a prerequisite for efficient binding, or cooperative interactions exist between two monomers binding to adjacent sequences. To analyze the requirement of dimer formation for binding, we have deleted the leucine zipper portion to the Antp-Hahb-4 fusion. This truncated protein showed a considerably lower affinity, equivalent to that of native Hahb-4 without the leucine zipper motif (Figure 1(b)), and its binding behavior suggested that it binds DNA as a monomer. Since the DNA-binding site used for these studies contains two overlapping monomer binding sites, we have employed UV-photocrosslinking to determine if two monomers of Antp-Hahb-4 without the leucine zipper motif are able to occupy both sites at the same time. The appearance of a band with the mobility of dimers (85 kDa) in a proportion similar to that observed for native Hahb-4 suggests that two monomer molecules are able to bind to a single pseudopalindromic oligonucleotide containing two adjacent binding sites (Figure 3). This rules out steric interference between both monomers as the cause of low-affinity binding to DNA and reinforces the idea that dimer formation through the leucine zipper is required by Antp-Hahb-4 for efficient binding. We then conclude that the N-terminal arm of Antp reconstitutes high DNA-binding affinity to ΔN-Hahb-4 dimers, but is not able to transform an inefficient DNA-binding monomer into an efficient one.

Positive charges in the N-terminal arm are required for efficient binding

Alignment of the sequences immediately adjacent to helix I of the HD of the different proteins analyzed showed almost no conservation in amino acid residues present at defined positions in the N-terminal arm that could be correlated with the efficiency of binding, with the sole exception of Arg at position 1, present in Hahb-4 and Antp-Hahb-4, and absent from ΔN-Hahb-4 (Figure 4). Indeed, positions 3, 4, and 5 (Glu, Gly and Arg, respectively) are conserved between Hahb-4 and ΔN-Hahb-4, which show very different affinities. This would suggest that position 5 is not important *per se* for DNA binding, in accordance with the fact that mutation of Arg5 to Lys did not affect the binding properties of Hahb-4 (results not shown).

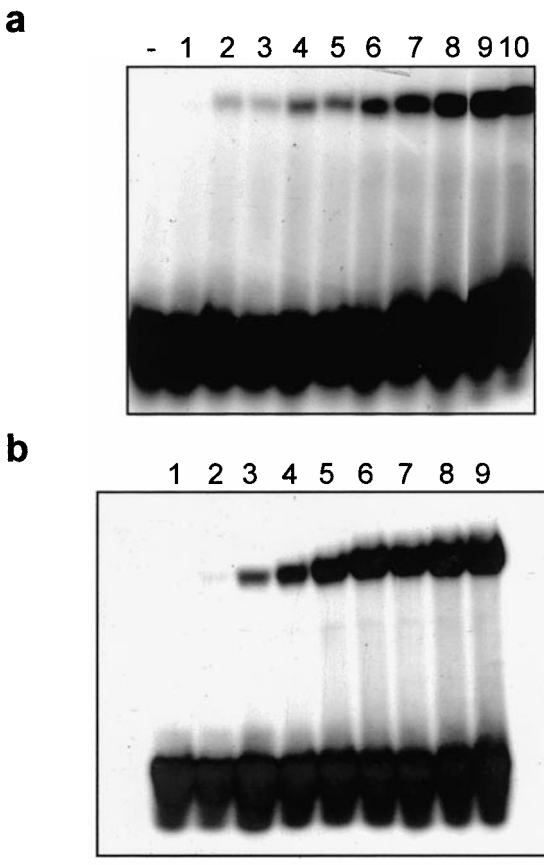

Figure 2. Binding curves of different amounts of ΔN-Hahb-4 and Antp-Hahb-4 to DNA. Different amounts of ΔN-Hahb-4 ((a) 37-365 nM), or Antp-Hahb-4 ((b) 10-104 nM) were analyzed for DNA-binding as described. (c) Quantitative analysis of the ratio of bound to free DNA as a function of total protein concentration for Hahb-4 (squares), Antp-Hahb-4 (circles), and ΔN-Hahb-4 (triangles).

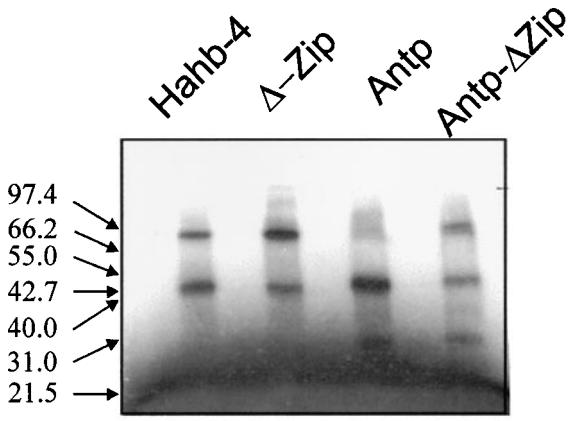

Figure 3. Photocrosslinking of different forms of Hahb-4 to DNA. Amounts of different proteins that produced similar ratios of bound to free DNA in electrophoretic mobility shift assays were photocrosslinked to labeled DNA and analyzed as described (see Materials and Methods). Δ-Zip, Hahb-4 with the complete leucine zipper removed; Antp, Antp-Hahb-4; Antp-ΔZip, Antp-Hahb-4 with the complete leucine zipper removed. As control, labeled DNA was incubated without protein under similar conditions and loaded in intermediate lanes. The bands at 42 kDa and 85 kDa correspond to one and two protein monomers covalently bound to DNA, respectively. The band at 30 kDa represents a truncated product, presumably present due to photochemical cleavage of the protein-DNA adduct.

A close examination of the number of charged amino acid residues along the N-terminal arm of the different proteins suggests that the number of positive charges may be an important factor for N-terminal arm function. Values of +3 and +4 are obtained for Hahb-4 and Antp-Hahb-4, whilst ΔN-Hahb-4 has one positively charged and one negatively charged amino acid residue (Figure 4). To analyze the influence of charges in the N-terminal arm, and to rule out any possible effects of the distance between the HD and adjacent sequences belonging to GST, we have constructed Hahb-4 derivatives with either Ala or Glu at positions 5, 6, and 7 of the HD (Figure 4). The affinity of these proteins was compared with a protein that contained Arg, Lys, and Arg at these positions, respectively, as in Hahb-4. In these proteins, all sequences N-terminal to the HD, including the first two amino acid residues of the N-terminal arm, have been removed and are then replaced by GST sequences. As shown in Figure 5, the positively charged derivative (N-Arg-Hahb-4) showed the same affinity as native Hahb-4. The protein with Ala in the N-terminal arm showed an affinity comparable to that of the N-terminal truncated form. For the protein with Glu, no binding could be observed. These results clearly indicate that positive charges at the N-terminal end of the HD are essential for efficient DNA binding.

We have then analyzed the effect of single mutations within this segment. Figure 6 shows that the introduction of Ala at either position 5, 6 or 7 has a small effect on DNA binding. On the other hand, a double mutant with Ala at positions 6 and 7 shows a significant reduction in affinity, while

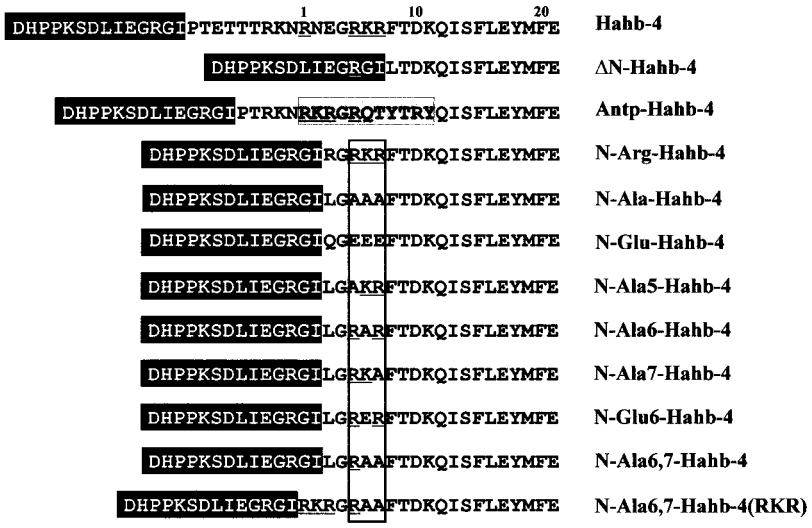

Figure 4. Sequences at the N-terminal arm of the HD in different forms of Hahb-4. The N-terminal arm and adjacent sequences of Hahb-4 are displayed together with sequences of the different proteins used in this study. Residues are numbered from the beginning of the HD. Alignment was performed taking helix I (beginning at residue 10) as reference. The black and gray squares denote GST and Antp sequences, respectively. The white square highlights the amino acid residues at positions 5 to 7 that were mutated. Positive charges present in the N-terminal arm region are underlined.

the introduction of a single Glu at position 6 completely abolishes binding. These results suggest that there is an additive effect of positive charges within this segment on DNA binding.

Based on this, it is noteworthy that Antp-Hahb-4, which binds DNA efficiently, contains neutral residues at positions 6 and 7. Then, how does the Antp N-terminal arm participate in DNA binding within the context of the Hahb-4 HD? Figure 6 shows that a protein with Ala at positions 6 and 7 binds DNA efficiently if the sequence Arg-Lys-Arg (as in Antp) is introduced at positions 1 to 3 of the Hahb-4 HD. This points to the existence of flexible interactions between positively charged residues of the HD N-terminal arm and DNA. The N-terminal arm may adopt different conformations to accommodate these residues for efficient interaction.

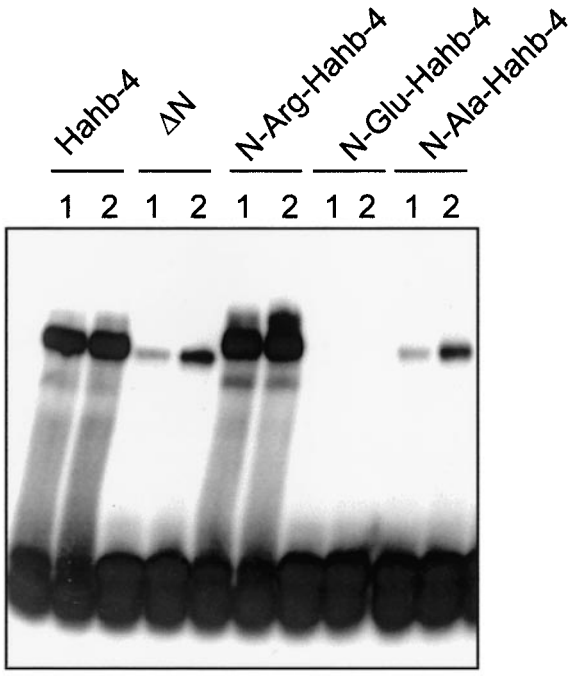

Figure 5. Influence of positive charges at the N-terminal arm on Hahb-4 binding to DNA. Binding of different forms of Hahb-4 with modifications at positions 5 to 7 of the N-terminal arm of the HD were analyzed for DNA-binding as described. ΔN, ΔN-Hahb-4. Lanes 1 and 2 contain 50 ng and 100 ng of protein, respectively.

Discussion

Current models of DNA binding by Hd-Zip proteins[19] are based on comparisons of conserved amino acid residues also present in animal HDs, the structure and DNA-binding properties of which have been studied by X-ray crystallography[7,9] or NMR.[4,8] These models have been validated by experimental evidence showing that Hd-Zip proteins bind a 9-bp pseudopalindromic DNA sequence of the type 5′-CAATNATTG-3′,[19] which can be decomposed into two antiparallel, partially overlapping 5′-TNATTG-3′ sequences, similar to those bound by most animal HDs (Figure 7). The third and fourth positions within this sequence are recognized by Asn51 (almost invariant) and Ile/Val47 (highly conserved) in animal HDs, while the fifth and sixth positions are thought to interact with amino acid residues at positions 50 and 54 of the HD.[3,10,21] Hd-Zip proteins contain Asn at position 51, Ile/Val at position 47, Gln (as in Antp and engrailed) at position 50, and Ala (as in engrailed, Met in Antp) at position 54, suggesting a similar mode of interaction. The first two nucleotides of the 5′-TNATTG-3′ sequence interact with amino acid residues present in the N-terminal arm in animal HDs, namely Arg3 and Arg5 in Antp and engrailed. Residue 3 is either Asn or Glu in Hd-Zip I proteins, and Lys or Arg in Hd-Zip II proteins; position 5 is either Gln, Lys, or Arg.[22] This lack of conservation seems to indicate a less relevant role for the N-terminal arm in DNA binding. In addition, position 1 in the binding sequence is coincident with position 3 of the other

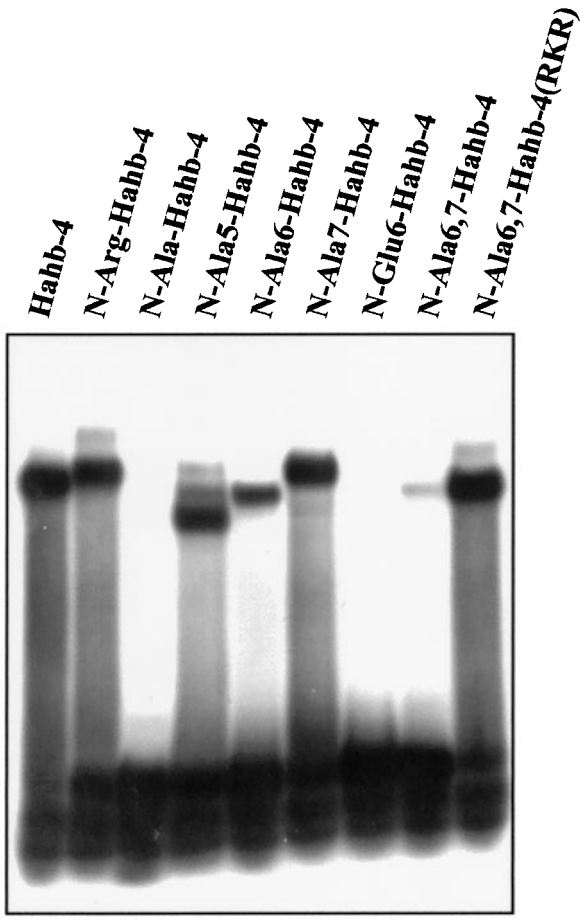

Figure 6. Influence of amino acid residue changes at different positions within the N-terminal arm of the Hahb-4 HD on DNA-binding. Binding of different forms of Hahb-4 with single, double, or triple mutations at positions 5 to 7 of the HD were analyzed for DNA-binding as described. See Figure 4 for details of the corresponding sequences.

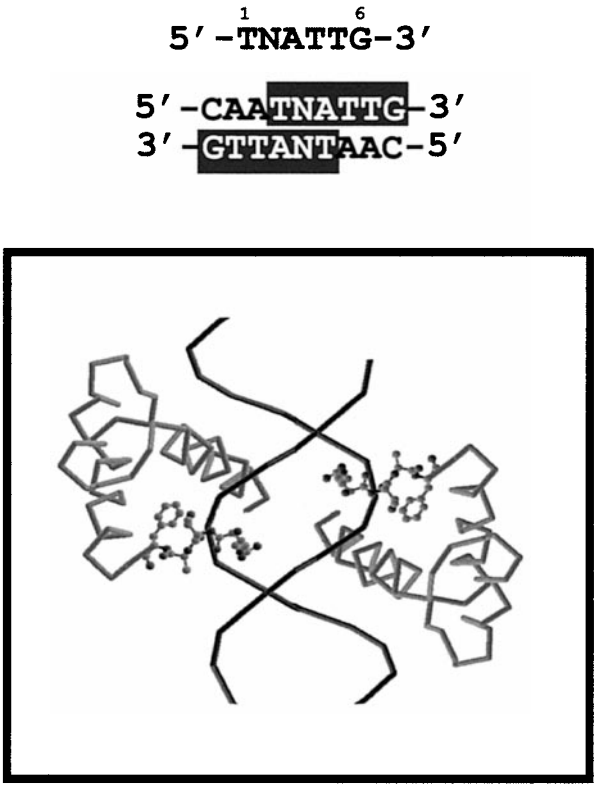

Figure 7. The binding of two HDs to the pseudopalindromic DNA sequence 5′-CAAT(N)ATTG-3′. The upper part of the Figure shows the consensus sequence bound by most animal HDs (in which Ala at position 2 has been replaced by Asn to accommodate it to sequences bound by Hd-Zip proteins). Below, the arrangement of this sequence within the pseudopalindromic site recognized by Hd-Zip proteins is highlighted by a gray square. The lower part shows the putative binding of two HDs to the highlighted sequences. To obtain this picture, the crystal structure of the engrailed HD bound to DNA[7] has been rotated 180° around an axis perpendicular to the direction of the second base-pair (the central base-pair of the pseudopalindromic site) and contained within the plane formed by this base-pair, and the resulting images were superimposed. Residues of the N-terminal arm of the HD are depicted as sticks and balls. Images were generated using the program RasMol.

binding site defined on the complementary strand, which means that interactions at this position could be established by Asn51 of the second monomer. This leaves position 2 (the central position of the psudopalindromic binding sequence) as the only position for which DNA-contacting amino acid residues cannot be ascribed from comparisons with monomeric HDs. Sessa *et al.*[19] have compared the recognition helices of Hd-Zip and b-Zip proteins, and proposed that the central position could be in contact with Arg55 in Hd-Zip proteins. In fact, substitution of this residue with Ala dramatically reduced binding affinity, and certain evidence suggests that its positioning may be related, in part, with sequence-specificity at the central position.[23] In principle, then, there would be no requirement for the N-terminal arm for DNA binding, although the number of nucleotides contacted by each monomer would be considerably reduced with respect to other HDs. This fact could explain the inefficient binding of Hd-Zip protein monomers.[20]

A close examination of sequences at the N-terminal arm of different Hd-Zip proteins, however, reveals a high degree of conservation, suggesting that this part of the HD may play an important role.[22] Our results indicate that, indeed, the removal of the N-terminal arm has a profound effect on binding affinity. The magnitude of this effect is similar to that of those observed for the truncation of the HDs of ftz and Msx.[14,15] Moreover, binding affinity could be restored by the addition of the Antp N-terminal arm, suggesting the existence of similar roles for both protein

segments. It is noteworthy that, although the N-terminal arm of Antp restored the binding affinity of dimers, it did not produce tight-binding monomers. It is likely that this reflects the fact that poor binding of monomers of Hd-Zip proteins is, at least in part, a consequence of structural differences outside the N-terminal arm. It can be speculated that some contacts between the HD and DNA that aid in the positioning of the recognition helix have been lost in Hd-Zip proteins, where this role is undertaken by the leucine zipper motif. For example, Tyr25 in the loop between helices I and II, and Arg31 in helix II, which are present in more than 80 % of the HDs and contact DNA in Antp, engrailed, and MATα2 HDs,[3,7-9] are not present in Hd-Zip proteins.

Replacement of N-terminal arm residues 5 to 7 indicated strongly that positively charged residues are important for DNA binding by Hahb-4. As a rule, despite small differences among particular positions, single mutations to Ala produce only slight effects. A strong reduction in affinity, however, is obtained after the introduction of a negative charge or two neutral amino acid residues. Hence, there seems to be an additive effect of charges at the N-terminal arm on binding efficiency rather than a positional effect. This fact may apply to the segment comprising amino acids 5 to 7, and to the complete N-terminal arm. Accordingly, a mutant with Ala at positions 6 and 7 can bind DNA efficiently if a segment of positively charged residues is included at the beginning of the N-terminal arm. It seems then that there is some degree of flexibility in the interaction. This correlates with the fact that this part of the HD is disordered in the solution structures of HDs. Flexibility is reflected also by the fact that, in MATα2, DNA contacts are made by an Arg residue at position 7,[9] while Arg2, rather than Arg3, contacts DNA in the Oct-1 and Prd (Gln50) HDs.[24,25] In addition, changes in N-terminal arm residues 6 to 8 produce changes in DNA-binding specificity in Ubx, Abd-B, TTF-1, and Antp HDs.[26,27]

A comparison of N-terminal arm sequences in Hd-Zip proteins shows the existence of a net charge of +3 (exceptionally +2) in most Hd-Zip I proteins, and +4 in Hd-Zip II proteins. In general, the consensus sequences XX(E/N)(K/N)X(R/K)R(L/F)X and XX(R/K)KKLRL(S/T) can be derived for Hd-Zip I and II proteins, respectively.[22] This sequence conservation and our results showing the essential role of positive charges in DNA binding by Hahb-4 argue strongly in favor of a similar role of the N-terminal arm in all Hd-Zip proteins. In this respect, then, the HD of Hd-Zip proteins appears to be related to many animal and fungal HDs, although the type of interactions established by this segment with DNA varies considerably among different HD proteins. Figure 7 shows a model of the putative arrangement of the two HDs of an Hd-Zip protein dimer, deduced from the crystal structure of the engrailed HD monomer bound to DNA.[7] Although this model is speculative, the fact that most DNA-contacting amino acid residues have conserved functions suggests that it may be a good approximation. It can be observed that the respective N-terminal arms are located along the minor groove close to each other, but do not superimpose. Hence, it becomes possible that residues in the N-terminal arm of both HDs make contacts with DNA. Alternatively, this flexible part may have different orientations in the two monomers, so that only one of the N-terminal arms binds to DNA. Detailed structural studies are required to establish the validity of these assumptions. Our results imply that changes in the interaction of the N-terminal arm with DNA (brought about, for example, by the binding of additional factors) could be used to modulate the affinity of Hd-Zip proteins for DNA.

The fact that the N-terminal arm of Hd-Zip proteins participates in DNA binding is interesting from an evolutionary point of view. The question of how the present diversity in HD-protein structures has arisen is important for understanding the evolution of developmental processes at the molecular level. For example, did the present HDs derive from an ancestor molecule with the typical three-helix structure that was transformed into a high-affinity binding module by the addition of either a DNA-binding N-terminal arm or a leucine zipper dimerization motif? Or is the N-terminal arm an earlier acquisition, maybe inherent to ancient HD function? Our results suggest that a DNA interacting N-terminal arm was already present in HDs of the last common ancestor of angiosperms and metazoa. This fact is concordant with phylogenetic trees that place the HD of Hd-Zip proteins in a subgroup with Antp, leaving outside another HD that clearly utilizes the N-terminal arm for DNA binding, such as MATα2.[28,29] The acquisition of extra domains may have taken place later in evolution. These events, together with changes within the flexible N-terminal arm, were undoubtedly important factors in producing the present structural diversity of HD-containing proteins.

Materials and Methods

Cloning, expression, and purification of recombinant proteins

Hahb-4 coding sequences were amplified and cloned in-frame into the *Bam*HI and *Eco*RI sites of the expression vector pGEX-3X[30] as described.[31] Amplifications were performed on a clone of the entire Hahb-4 coding region in pUC119 as template, using the following 5′ oligonucleotides: 5′-GGCGGATCCCAACAGAAACAACCACCAGG-3′ (for native Hahb-4), 5′-GGGGGATCCTTACCGACAAACAAATAAG-3′ (for ΔN-Hahb-4), 5′-CCGGGATCCGTGGGCGGAAACGATTTACC-3′ (for N-Arg-Hahb-4), 5′-CCGGGATCCTGGGCGCGGCGGCGTTTACCGACAAACAAATA-3′ (for N-Ala-Hahb-4), and 5′-CCGGGATCCAGGGCGAAGAAGAGTTTACCGACAAACAAATA-3′ (for N-Glu-Hahb-4). N-Ala5, N-Ala6, N-Ala7, and N-Glu6-Hahb-4 were constructed in a similar way, using

N-Arg-Hahb-4 as template and the 5′ oligonucleotides 5′-CCGGGATCCTTGGGGCGAAACGATTTAC-3′, 5′-CCGGGATCCTTGGGCGGGCACGATTTACC-3′, 5′-CCGGGATCCTTGGGCGGAAAGCATTTACCGAC-3′, and 5′-CCGGGATCCTTGGGCGGGAACGATTTAC-3′, respectively. For the construction of N-Ala6,7-Hahb-4 and N-Ala6,7-Hahb-4(RKR), N-Ala-Hahb-4 as template, and 5′-GGCGGATCCTGGGCCGGGCGGCGTTTAC-3′ and 5′-GGCGGATCCGAAAGCGGGGCCGGGCGGCGTTTAC-3′, as 5′ oligonucleotides, were used. As 3′ oligonucleotides, either 5′-GGCGAATTCACATTTCTCAGCACCTCC-3′ or 5′-GCGGAATTCGCGCGTTATACTCTTGC-3′, for the amplification of regions encoding the Hd-Zip domain, or the HD alone, respectively, were used.

For the construction of the Antp-Hahb-4 fusion (carrying the N-terminal arm of Antp fused to the rest of the Hahb-4 Hd-Zip domain) two complementary oligonucleotides (5′-GATCCCCACCAGGAAGAACCGGAAACGAGGTCGACAAACGTATACCCGCTACCA-3′ and 5′-GATCTGGTAGCGGGTATACGTTTGTCGACCTCGTTTCCGGTTCTTCCTGGTGGG-3′) which generate single-stranded 5′-GATC-3′ ends were cloned in the correct orientation into the *Bam*HI and *Bgl*II sites of an Arg5 → Lys mutant described previously[20]. In this construct, the coding region for the first 11 amino acid residues of the HD of Hahb-4 are replaced for those of the Antp HD. All constructions were checked by DNA sequence analysis.

For expression, *Escherichia coli* cells bearing the corresponding plasmids were grown and induced as described.[31] Purification and cleavage of the fusion products were carried out essentially as described by Smith & Johnson,[30] with modifications described by Palena *et al*.[31]

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) generated by hybridization of the complementary oligonucleotides 5′-AATTCAGATCTCAATAATTGAGAG-3′ and 5′-GATCCTCTCAATTATTGAGATCTG-3′, and labeled with [α-^{32}P]dATP by filling-in the 3′-ends using the Klenow fragment of *E. coli* DNA polymerase I. 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 dithiothreitol, 0.5 % (v/v) Triton X-100, 1(g poly(dI-dC), and 10 % (w/v) glycerol, were incubated for 20 minutes at room temperature, supplemented with 2.5 % (w/v) Ficoll and immediately loaded onto a running gel (5 % (w/v) acrylamide, 0.08 % (w/v) bis-acrylamide in 0.5 × TBE plus 2.5 % (w/v) glycerol; TBE is 90 mM Tris-borate (pH 8.3), 2 mM EDTA). The gel was run in 0.5 × TBE at 20 mA for two hours and dried prior to autoradiography. For quantitative analysis, poly(dI-dC) was omitted.

For photocrosslinking experiments, binding reactions performed under identical conditions were incubated on ice during 20 minutes below a germicide UV lamp ($\lambda = 253.7$ nm) at a distance of 5 cm, and then loaded onto an SDS/12 % polyacrylamide gel. The gel was dried and autoradiographed, and the M_r of the respective bands were calculated by their migration relative to commercial standards. In these experiments, amounts of the different proteins that produced similar free-to-bound ratios in electrophoretic mobility shift assays were used.

Miscellaneous methods

Total protein was measured as described by Sedmak & Grossberg.[32] For quantitative analyses, radioactive bands were cut from exposed gels and measured by scintillation counting. Data handling and curve fitting were performed using Sigma plot software. Overall dissociation constants (K_{12}) of the dimer-DNA complexes into monomers and free DNA were calculated using the equation $K_{12} = P_2 \times D/P_2D$, according to the binding sequence: $2\ P \rightarrow P_2$, and $P_2 + D \rightarrow P_2D$, where P, P_2, D, and P_2D represent protein monomers and dimers, and free and bound DNA, respectively.[20]

Acknowledgements

This work was supported by grants from CONICET, ANPCyT and Fundacion Antorchas (Argentina). R.L.C. and D.H.G. are members of CONICET, C.M.P. and A.E.T. are fellows of the same Institution. C.W.B. is an undergraduate fellow of Universidad Nacional del Litoral.

References

1. Gehring, W. J. (1987). Homeo boxes in the study of development. *Science,* **236**, 1245-1252.
2. Levine, M. & Hoey, T. (1988). Homeobox proteins as sequence-specific transcription factors. *Cell,* **55**, 537-540.
3. Gehring, W. J., Affolter, M. & Bürglin, T. (1994a). Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487-526.
4. Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1989). The structure of the *Antennapedia* homeodomain determined by MNR 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. & 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. & Ferretti, J. A. (1995). The three dimensional solution structure of the NK-2 homeodomain from *Drosophila*. *J. Mol. Biol.* **251**, 297-307.
7. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell,* **63**, 579-590.
8. Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & 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.
9. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. (1991). Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell,* **67**, 517-528.
10. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D.,

Affolter, M., Otting, G. & Wüthrich, K. (1994b). Homeodomain-DNA recognition. *Cell,* **78**, 211-223.
11. Otting, G., Qian, Y. Q., Müller, M., Affolter, M., Gehring, W. J. & 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.
12. Affolter, M., Percival-Smith, A., Müller, M., Leupin, W. & Gehring, W. J. (1990). DNA binding properties of the purified Antennapedia homeodomain. *Proc. Natl Acad. Sci. USA,* **87**, 4093-4097.
13. Florence, B., Handrow, R. & Laughon, A. (1991). DNA-binding specificity of the fushi tarazu homeodomain. *Mol. Cell. Biol.* **11**, 3613-3623.
14. Percival-Smith, A., Müller, M., Affolter, M. & Gehring, W. J. (1990). The interaction with DNA of wild-type and mutant fushi tarazu homeodomains. *EMBO J.* **9**, 3967-3974.
15. Shang, Z., Isaac, V. E., Li, H., Patel, L., Catron, K. M., Curran, T., Montelione, G. T. & Abate, C. (1994). Design of a ''minimAl'' homeodomain: the N-terminal arm modulates DNA binding affinity and stabilizes homeodomain structure. *Proc. Natl Acad. Sci. USA,* **91**, 8373-8377.
16. Chan, R. L., Gago, G. M., Palena, C. M. & Gonzalez, D. H. (1998). Homeoboxes in plant development. *Biochim. Biophys. Acta*, **1442**, 1-19.
17. Ruberti, I., Sessa, G., Lucchetti, S. & Morelli, G. (1991). A novel class of proteins containing a homeodomain with a closely linked leucine zipper motif. *EMBO J.* **10**, 1787-1791.
18. Schena, M. & Davis, R. W. (1992). HD-Zip proteins: members of an *Arabidopsis* homeodomain protein superfamily. *Proc. Natl Acad. Sci. USA,* **89**, 3894-3898.
19. Sessa, G., Morelli, G. & Ruberti, I. (1993). The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. *EMBO J.* **12**, 3507-3517.
20. Palena, C. M., Gonzalez, D. H. & Chan, R. L. (1999). A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein Hahb-4 with DNA. *Biochem J.* **341**, 81-87.
21. Laughon, A. (1991). DNA binding specificity of homeodomains. *Biochemistry,* **30**, 11357-11367.
22. Gonzalez, D. H., Valle, E. M., Gago, G. M. & Chan, R. L. (1997). Interaction between proteins containing homeodomains associated to leucine zippers from sunflower. *Biochim. Biophys. Acta,* **1351**, 137-149.
23. Sessa, G., Morelli, G. & Ruberti, I. (1997). DNA-binding specificity on the homeodomain-leucine zipper domain. *J. Mol. Biol.* **274**, 303-309.
24. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. & Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell,* **77**, 21-32.
25. Wilson, D. S., Guenther, B., Desplan, C. & Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell,* **82**, 709-719.
26. Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E. & Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *EMBO J.* **13**, 3551-3560.
27. Damante, G., Pellizzari, L., Esposito, G., Fogolari, F., Viglino, P., Fabbro, D., Tell, G., Formisano, S. & Di Lauro, R. (1996). A molecular code dictates sequence-specific DNA recognition by homeodomains. *EMBO J.* **15**, 4992-5000.
28. Bharathan, G., Janssen, B. J., Kellog, E. A. & Sinha, N. (1997). Did homeodomain proteins duplicate before the origin of angiosperms, fungi and metazoa? *Proc. Natl Acad. Sci. USA,* **94**, 13749-13753.
29. Bürglin, T. R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX; Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucl. Acids Res.* **25**, 4173-4180.
30. Smith, D. B. & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene,* **67**, 31-40.
31. Palena, C. M., Gonzalez, D. H., Guelman, S. & Chan, R. L. (1998). Expression of sunflower homeodomain containing proteins in *Escherichia coli*: purification and functional studies. *Protein Expr. Purif.* **13**, 97-103.
32. Sedmak, J. & Grossberg, S. (1977). A rapid, sensitive, and versatile assay for protein using Coomassie brilliant blue G-250. *Anal. Biochem.* **79**, 544-552.

Edited by M. Yaniv

(Received 5 September 2000; received in revised form 19 February 2001; accepted 19 February 2001)