Engineering the Loop Region of a Homeodomain-Leucine Zipper Protein Promotes Efficient Binding to a Monomeric DNA Binding Site†

Adriana E. Tron, Elina Welchen, 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 August 12, 2004; Revised Manuscript Received September 27, 2004

ABSTRACT: Plant homeodomain-leucine zipper (HD-Zip) proteins, unlike many animal homeodomains (HDs), are unable to bind DNA as monomers. To investigate the molecular basis of their different behavior, we have constructed chimeras between the HD of the sunflower HD-Zip protein Hahb-4 and that of *Drosophila* engrailed (EN). Analysis of the interaction of these proteins with the pseudopalindromic Hahb-4 binding site and the monomeric EN binding site suggests that the loop located between helix I and helix II (amino acids 21−28) of EN is enough to confer efficient DNA binding activity to the Hahb-4 HD. Accordingly, the combined mutation of residues 24 and 25 of Hahb-4 to those present in EN (S24R/R25Y) originated an HD able to interact with the EN binding site, while single mutations were ineffective. We have also determined that a protein with the leucine zipper and helix III of Hahb-4 fused to the rest of the EN HD binds to the Hahb-4 pseudopalindomic binding site with increased affinity and shows extended contacts with DNA respective to Hahb-4. We conclude that the loop located between helix I and helix II of the HD must be regarded as one of the segments that contribute to the present-day diversity in the properties of different HDs.

The homeodomain (HD)[1] is a 61-amino acid 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 (*3*−*6*). 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 connected by a disordered loop and are roughly perpendicular to helix III. The helices pack against each other, 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*), while an extended helix III is present in others (*7*, *9*). The N-terminal portion of the HD (the first nine amino acids) forms a disordered arm that protrudes from helix I and orients toward DNA, making several specific contacts within the minor groove (*7*−*10*).

Despite the resemblance in structure between the HD and the helix−turn−helix motif, a striking difference is that many HDs bind DNA as monomers with high affinity (*12*, *13*). This fact has been explained by the presence of extended contacts along the recognition helix and, specially, by the stabilizing effect of contacts made by the N-terminal arm.

HDs are present in almost every eukaryotic organism that has been investigated. In plants, several families of HD proteins have been described (*14*). One of these families, named HD-Zip, comprises proteins with a typical leucine zipper motif adjacent to the C-terminal end of the HD (*15*, *16*). As expected, these proteins bind DNA only as dimers, recognizing a pseudopalindromic DNA sequence composed of two half-sites, each one similar to sequences bound by many animal HDs (*17*−*19*). Either mutations in any of the half-sites, the removal of the leucine zipper, or the introduction of extra amino acids between the HD and the zipper significantly reduce binding affinity, indicating that both monomers must be correctly positioned for efficient binding (*17*, *18*). Redox changes in cysteines present at or near the leucine zipper also influence binding capacity in a group of HD-Zip proteins (*20*). The analysis of binding at different protein concentrations suggests that dimer formation is a prerequisite for DNA binding (*18*). 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, while other HDs (notably those animal HDs most related in sequence) do. We have previously shown that differences within the respective N-terminal arms do not explain this behavior, since this segment is also required for efficient binding by dimers of the sunflower HD-Zip protein Hahb-4 and the N-terminal arm of the Antennapedia HD (which binds DNA as a monomer) cannot confer efficient binding to Hahb-4 monomers (*21*). This suggests that the

† Supported by grants from CONICET, ANPCyT, and Universidad Nacional del Litoral. D.H.G. is a member of CONICET; A.E.T. and E.W. are fellows of the same institution.

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

[1] Abbreviations: EN, engrailed; HD, homeodomain.

10.1021/bi048254a CCC: $27.50
Published on Web 11/19/2004

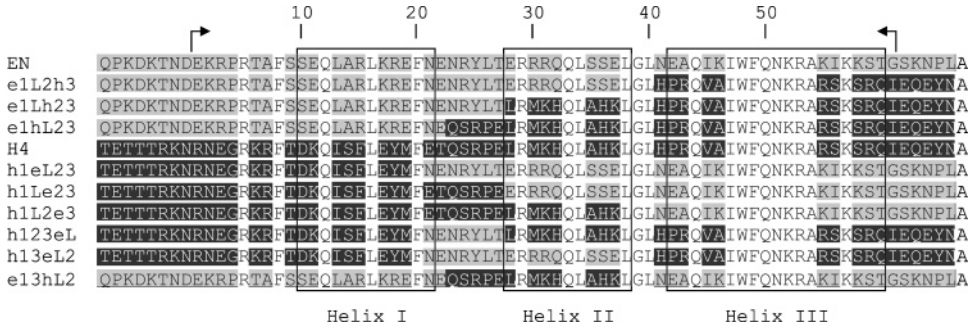

FIGURE 1: Sequence of EN/Hahb-4 chimeras. The amino acid sequences of the different recombinant proteins used in this study are shown. The arrows indicate the location of the HD and numbers indicate amino acid positions starting from the N-terminus of the HD. Hahb-4 and EN amino acids are within black and gray boxes, respectively, while those shared by both HDs are not boxed. Proteins were named as follows: 1, 2, 3, and L stand for helices I, II. and III and the loop. They are written after e or h if the corresponding protein contains this portion from either EN or Hahb-4, respectively. As an example, e1L2h3 names a protein that contains helix I, the loop, and helix II from EN fused to Hahb-4 helix III.

lack of efficient binding by Hahb-4 monomers may be the consequence of differences in other parts of the HD.

In the present work, we have constructed a series of chimeric HDs composed of segments of Hahb-4 and *Drosophila* EN, which binds DNA as a monomer (*7*). The analysis of these chimeras and of single and double mutants suggests that residues within the loop located between helix I and helix II, particularly those at positions 24 and 25, are responsible for the different DNA binding properties of these HDs.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of Recombinant Proteins. Hahb-4 and EN coding sequences were amplified and cloned in frame into the *Bam*HI and *Eco*RI sites of the expression vector pGEX-3X (*22*). The construction of clones that express the Hahb-4 HD-Zip domain (H-Zip) or the HD alone (H4) has been described previously (*23*). To express the EN HD (amino acids −8 to 67, considering +1 the first HD amino acid), amplifications were performed using oligonucleotides ENNH (5′-CCGGGATCCAACAGCCAAAGGACAGA-3′) and ENCO (5′-GGAGAATTCGTGCCAGCGGATTTTTGG-3′) and an EN cDNA clone as template.

Fragments encoding chimeric proteins were synthesized using primers containing Hahb-4 sequences fused to EN sequences to amplify partially complementary fragments that were then hybridized to reconstruct the complete chimeric HD. The sequences of the different proteins that were used in this study are shown in Figure 1. Primers EN1 (5′-TCGCGACTGCTCGTTGAACTCCCGCTT-3′), EN1L (5′-CATCCTTAAGGTCAGATAGCGATTCTC-3′), or EN12 (5′-AGGATGAAGGCCCAACTCGCTGCTCAG-3′) (Hahb-4 sequences underlined) were used together with primer ENNH to amplify EN HD N-terminal sequences comprising amino acids −8/22, −8/27, or −8/39. These products were hybridized with fragments of Hahb-4 comprising amino acids 23/67, 28/67, or 40/67 amplified using primers H4L23 (5′-TTCAACGAG CAGTCGAGACCCGAGTTA-3′), H423 (5′-TATCTGACC TTAAGGATGAAACACCAG-3′), or H43 (5′-GAGTTGGGC CTTCATCCTCGTCAAGTG-3′) together with primer deltaZip (5′-GCGGAATTCGCGCGTTATACTCTTGC-3′). The 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 reach 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 fragment encoding proteins e1hL23, e1Lh23, or e1L2h3 using primers ENNH and deltaZip. In a similar way, Hahb-4 HD N-terminal fragments amplified with primers H41 (5′-ATTCTCGTTAAACATGTACTCTAGGAA-3′), H41L (5′-TCTCCGCTCCTCGGGTCTAGACTGTGT-3′), or H41L2 (5′-CGCCTCGTTAAGCCCGAGTTTATGTGC-3′) together with primer H4NH (5′-GGCGGATCCCAACAGAAACAACCACCAGG-3′) were hybridized with EN HD C-terminal fragments amplified with ENL23 (5′-TACATGTTTAACGAGAATCGCTATCTG-3′), EN23 (5′- AGACCCGAGGAGCGGAGACGCCAGCAG-3′), or EN3 (5′-CTCGGGCTTAACGAGGCGCAGATCAAG-3′) together with primer ENCO. The resulting products were amplified with primers H4NH and ENCO, thus producing fragments encoding h1eL23, h1Le23, or h1L2e3. Sequences encoding proteins h123eL and h13eL2 were constructed by hybridizing Hahb-4 fragments amplified with H4NH and H41 with fragments amplified with ENL23 and deltaZip using either e1Lh23 or e1L2h3 as templates. A sequence encoding protein e13hL2 was constructed by hybridizing an EN fragment amplified with ENNH and EN1 with a fragment amplified with H4L23 and ENCO using h1L2e3 as template. Sequences encoding proteins with amino acids 24, 25, 24/25, or 25/26 of EN within the Hahb-4 HD were constructed by hybridizing fragments amplified with H4NH and either 24R (5′-CTCGGGTCTGCGCTGTGTCTC-3′), 25R (5′-TAACTCGGGGTACGACTGTGT-3′), 2425R (5′-TAACTCGGGGTAGCGCTGTGTCTC-3′), or 2526R (5′-CCTTAACTCCAGGTACGACTGTGT-3′) with fragments amplified with deltaZip and either 24F (5′-GAGACACAGCGCAGACCCGAG-3′), 25F (5′-ACACAGTCGTACCCCGAGTTA-3′), 2425F (5′-GAGACACAGCGCTACCCCGAGTTA-3′), or 2526F (5′-ACACAGTCGTACCTGGAGTTAAGG-3′), using Hahb-4 as template. All constructions were checked by DNA sequence analysis.

For expression, *E. coli* cells bearing the corresponding plasmids were grown and induced as described previously (*23*). Purification by affinity chromatography was carried out essentially as described by Smith and Johnson (*22*), with modifications described by Palena et al. (*23*). When necessary, purified proteins were subjected to cleavage by factor

Xa as described (*22*). Purified proteins (>95% as judged by Coomassie Brilliant Blue staining of denaturing polyacrylamide gels) were used for the assays. Protein amounts were measured as described by Sedmak and Grossberg (*24*) and verified by inspection of the corresponding bands in polyacrylamide gels.

DNA-Binding Assays. For electrophoretic mobility shift assays, aliquots of purified proteins were incubated with double-stranded DNA (0.3−0.6 ng, 30 000 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 oligonucleotides 5′-AATTCAGATCT-CAATAATTGAGAG-3′ and 5′-GATCCTCTCAATTAT-TGAGATCTG-3′ (binding site for Hahb-4 underlined) or 5′-AATTCTTGGATGTAATTACCGACTC-3′ and 5′-TC-GAGAGTCGGTAATTACATCCAG-3′ (binding site for EN underlined). 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, 1 μg poly(dI-dC), 10% glycerol, and 22 ng/μL BSA were incubated for 20 min at room temperature, supplemented with 2.5% Ficoll, and immediately loaded onto a running gel (5% acrylamide, 0.08% bis-acrylamide in 0.5× TBE plus 2.5% glycerol; 1× TBE is 90 mM Tris-borate, pH 8.3, 2 mM EDTA). The gel was run in 0.5× TBE at 30 mA for 1.5 h and dried prior to autoradiography. When comparing binding to different oligonucleotides, equal amounts of DNA (as judged by absorbance at 260 nm before labeling and by scintillation counting after labeling) were used. The assays shown are representative of at least three experiments that yielded essentially the same results. For quantitative analysis, radioactive bands were cut from exposed gels and measured by scintillation counting.

Missing Nucleoside Experiments. For the analysis of the nucleosides required for E-Zip and H-Zip binding, a double-stranded oligonucleotide containing the Hahb-4 binding site (described above) was cloned into the *Bam*HI and *Eco*RI sites of pBluescript SK$^-$. From this clone, a 48-bp fragment was obtained and labeled in one of its 3′ ends. This was accomplished by PCR using reverse and universal primers, followed by cleavage with either *Hin*dIII or *Xba*I (from the pBluescript polylinker), incubation with the Klenow fragment of DNA polymerase and [α-^{32}P]dATP, cleavage with the other enzyme, and purification by nondenaturing polyacrylamide gel electrophoresis. The labeled oligonucleotide was subjected to hydroxyl radical cleavage as described by Dixon et al. (*25*). Binding of either H-Zip or E-Zip to the treated oligonucleotide (200 000 cpm) and separation of the free and bound fractions by electrophoretic mobility shift assay was performed as described. These fractions were excised from the gel, eluted, and analyzed on a denaturing polyacrylamide gel (*25*).

RESULTS

Hahb-4/EN Chimeric HDs Bind to a Monomeric DNA Binding Site. To understand the molecular basis of the different DNA binding behavior of Hahb-4 and EN HD monomers, we have constructed a set of chimeras between these HDs (Figure 1). The corresponding proteins were expressed in *E. coli* as fusions with glutathion *S*-transferase and purified by affinity chromatography. Since we have previously shown that the deletion of the leucine zipper prevents efficient DNA binding by the Hahb-4 HD, we have initially looked at the DNA binding capacity of a protein containing helix III of Hahb-4 (amino acids 41−67) within the context of the EN HD. As shown in Figure 2, this protein, named e1L2h3, is able to bind to the pseudopalindromic Hahb-4 recognition site (H4BS), even though it lacks the leucine zipper, with similar efficiency as the Hahb-4 HD-Zip domain (H-Zip). Almost no binding by the Hahb-4 HD alone (H4) was detected under these conditions. The complex formed by e1L2h3 displayed higher mobility than the one formed by H-Zip. Since H-Zip binds to DNA as a dimer, this may arise from the fact that only one molecule of e1L2h3 is bound to the Hahb-4 recognition site or simply because H-Zip is a larger protein. Different mobilities due to electrostatic or conformational differences cannot be ruled out either.

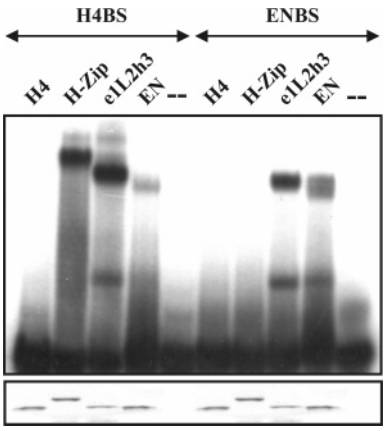

FIGURE 2: A protein with helix III of Hahb-4 fused to the rest of the EN HD (e1L2h3) binds efficiently to a monomeric DNA binding site. The binding of different proteins (200 ng) to oligonucleotides containing the Hahb-4 or the EN binding site (H4BS and ENBS, respectively) was analyzed by electrophoretic mobility shift assay. H-Zip and H4 contain the Hahb-4 HD with or without the leucine zipper, respectively. In the lanes labeled - -, no protein was added. Below, a Coomassie Brilliant Blue stained polyacrylamide gel with protein amounts equivalent to those used in the electrophoretic mobility shift assay is shown.

e1L2h3 also binds the EN monomeric recognition site (ENBS), which is not recognized by either the Hahb-4 HD-Zip domain or the HD alone (Figure 2), suggesting that, indeed, the chimeric protein binds efficiently to a half-site of the pseudopalindromic Hahb-4 recognition site. This result suggests that helix III of Hahb-4 is fully functional, even in the absence of the leucine zipper, at least within the context of the EN HD.

Chimeric proteins with increasing C-terminal portions of the Hahb-4 HD recognize the EN binding site as far as they retain the EN loop (Figure 3). When only the EN N-terminal arm and helix I were left (protein e1hL23), no binding could be detected (Figure 3). The results also indicate that residues within helix II of EN determine binding efficiency, since protein e1L2h3 binds better than e1Lh23.

Chimeras with different N-terminal portions of the Hahb-4 HD were also analyzed. Protein h1eL23, which contains the N-terminal arm and helix I of Hahb-4, binds DNA almost as efficiently as the EN HD (Figure 3). We have previously shown that the Antennapedia HD N-terminal arm, which has similar functions as the EN N-terminal arm, is unable to

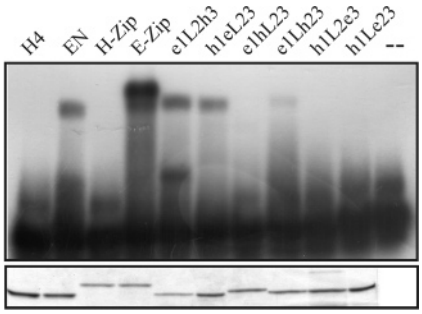

FIGURE 3: DNA-binding properties of HDs with different N-terminal and C-terminal portions of Hahb-4. The binding of different proteins (200 ng) to an oligonucleotide containing the EN binding site was analyzed by electrophoretic mobility shift assay. H-Zip and H4 contain the Hahb-4 HD with or without the leucine zipper, respectively. E-Zip contains the Hahb-4 helix III and leucine zipper fused to the rest of the EN HD. In the lane labeled - -, no protein was added. Below, a Coomassie Brilliant Blue stained polyacrylamide gel with protein amounts equivalent to those used in the electrophoretic mobility shift assay is shown.

confer efficient DNA binding to the Hahb-4 HD and that the N-terminal arm of Hahb-4 is necessary for high-affinity binding by Hahb-4 dimers (*21*). The results with protein h1eL23 also indicate that the N-terminal arm of Hahb-4 can functionally replace the EN HD N-terminal arm. On the other hand, if the loop or the loop plus helix II of Hahb-4 are added (proteins h1Le23 and h1L2e3, respectively), binding capacity is completely abolished (Figure 3). The main difference between the EN and Hahb-4 HDs, regarding their capacity to bind a monomeric target site, seems then to be located within the loop located between helix I and helix II.

In this experiment, a protein composed of helix III plus the leucine zipper of Hahb-4 fused to the rest of the EN HD (E-Zip) was also analyzed. It was observed that the presence of the leucine zipper produces a significant increase in binding capacity to the monomeric EN binding site (compare E-Zip with e1L2h3 in Figure 3). The importance of EN sequences located N-terminal to helix III is also evident with this protein, since the Hahb-4 HD-Zip domain (H-Zip) does not bind to the EN binding site.

The Loop and Adjacent Amino Acids of the EN HD Are Enough To Confer Efficient DNA Binding to the Hahb-4 HD. The importance of the loop and helix II in determining the efficiency of DNA binding by the Hahb-4 HD was directly assessed by replacing these portions of the protein with the corresponding regions of EN. Figure 4A shows DNA binding by different amounts of proteins h123eL, which is the Hahb-4 HD with the EN loop, and h13eL2, which contains the loop and helix II of EN. This analysis indicates that the EN loop is enough to confer binding to the Hahb-4 HD, although with reduced affinity with respect to the complete EN HD. The addition of the EN helix II to h123eL produces a HD with binding affinity similar to that of the EN HD. The opposite replacement, that is, the inclusion of the Hahb-4 loop and helix II within the EN HD, completely abolishes binding to DNA (not shown).

To avoid any effect of the glutathione *S*-transferase moiety present in the recombinant proteins, we have also analyzed DNA binding after factor Xa cleavage. As shown in Figure 4B, the results obtained were essentially the same as those observed with the fusion proteins. While the Hahb-4 HD, either with or without the leucine zipper (H-Zip and H4,

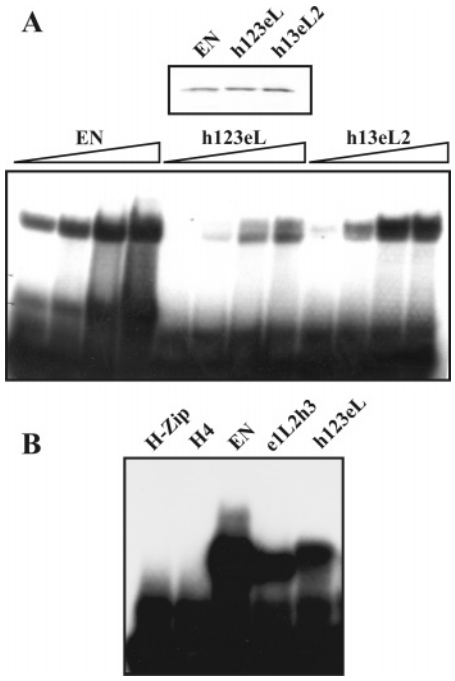

FIGURE 4: The EN loop confers efficient binding to the Hahb-4 HD. (A) Different amounts (20, 50, 100, and 200 ng) of fusion proteins EN, h123eL (containing the EN loop) and h13eL2 (containing the loop and helix II of EN) were analyzed for binding to an oligonucleotide containing the EN binding site by electrophoretic mobility shift assay. Above the electrophoretic mobility shift assay, a Coomassie Brilliant Blue stained polyacrylamide gel with equivalent protein amounts is shown as a control. (B) Electrophoretic mobility shift assay using proteins digested with factor Xa. Equivalent amounts of the different proteins were analyzed for binding to the EN binding site. H-Zip and H4 contain the Hahb-4 HD with or without the leucine zipper, respectively.

respectively), was unable to recognize the EN binding site, inclusion of the EN loop (protein h123EL) allowed significant binding.

Analysis of the EN crystal structure (*7*) indicates that loop residues 24−26 are oriented toward the DNA and helix III of the HD. Particularly, Tyr25 interacts with the phosphate backbone in the EN and Antennapedia structures (*7*, *8*). We then constructed single mutants at positions 24 and 25 and double mutants at positions 24/25 and 25/26, replacing residues in Hahb-4 with those of EN. The results obtained, shown in Figure 5, indicate that the combined mutation of Ser24 into Arg and Arg25 into Tyr originates an HD able to interact with the EN binding site. Changes at positions 25 and 26 did not produce any effect (Figure 5), while a single mutant at position 24 showed only a small amount of binding, barely observable in this particular figure, but confirmed along several experiments. We conclude that residues present at positions 24 and 25 of the HD are main determinants of DNA binding efficiency.

A Chimeric HD-Zip Domain Containing EN Sequences Displays Increased Affinity for a Pseudopalindromic DNA Binding Site. In HD-Zip proteins, poor DNA binding by monomers may be a requisite for efficient DNA binding by dimers linked through a leucine zipper (i.e., the spatial orientation required for efficient binding of two HDs to the corresponding half-sites contained within a pseudopalindromic recognition site may be restrictive for efficient monomer binding). In fact, it has been reported that the binding of head-to-head fushi tarazu HD monomers to adjacent binding

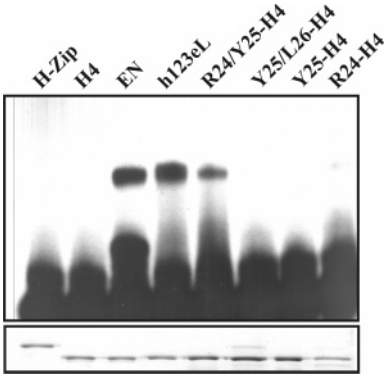

FIGURE 5: EN loop residues Arg24 and Tyr25 improve DNA binding activity of the Hahb-4 HD. The binding of the different proteins (100 ng) to oligonucleotides containing the EN binding site was analyzed by an electrophoretic mobility shift assay. H-Zip and H4 contain the Hahb-4 HD with or without the leucine zipper, respectively. h123eL is the Hahb-4 HD with the EN loop. The other proteins are single or double mutants of the Hahb-4 HD, as indicated. Below, a Coomassie Brilliant Blue stained polyacrylamide gel with protein amounts equivalent to those used in the electrophoretic mobility shift assay is shown.

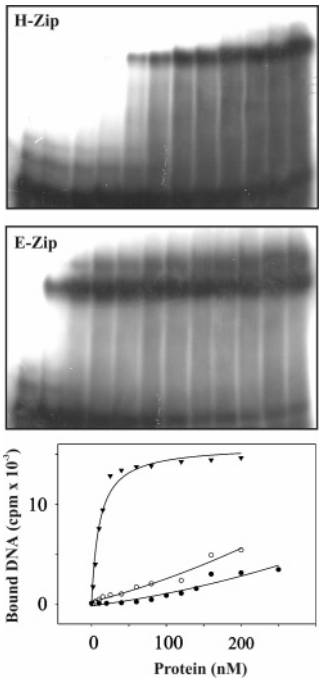

FIGURE 6: A protein with helix III and the leucine zipper of Hahb-4 fused to the rest of the EN HD (E-Zip) binds DNA with high affinity. The binding of different amounts of proteins E-Zip (0, 2, 5, 10, 15, 25, 40, 60, 80, 120, 160, and 200 ng) and H-Zip (0, 10, 20, 40, 60, 80, 100, 120, 140, 160, 200, and 250 ng) to an oligonucleotide containing the Hahb-4 binding site was analyzed by electrophoretic mobility shift assays. The lower panel shows the relationship of bound DNA with protein concentration for H-Zip (closed circles) and E-Zip (lower band, triangles; upper band, open circles). The data were adjusted to a single rectangular hyperbola (triangles) or to a quadratic equation (circles).

sites is severely reduced when these sites are less than 4 bp apart (*13*). To analyze this, we constructed an HD-Zip protein with helix III and the leucine zipper of Hahb-4 and the rest of the EN HD. As shown above, this protein (E-Zip) binds efficiently to the EN monomeric binding site. Figure 6 shows the binding of E-Zip at different concentrations to the Hahb-4 pseudopalindromic DNA recognition site. A similar analysis was made for the Hahb-4 HD-Zip domain (H-Zip). Upon comparing these results, it becomes evident that E-Zip, the chimeric protein, shows enhanced affinity for DNA. This most likely arises from the fact that monomers are capable of DNA binding in E-Zip but inactive in H-Zip. Two bands of different mobility are observed for the protein−DNA complexes formed by E-Zip. We speculate that these bands correspond to monomers and dimers bound to DNA. In fact, the binding curve of the putative monomers of E-Zip (lower band) to DNA shows a hyperbolic response, indicative of a 1:1 interaction between protein and DNA (Figure 6). The other binding curves show a quadratic response to total protein concentration, which is indicative of a process of dimer formation before binding, as we have shown before (*18*). The results clearly indicate that two active monomers can bind efficiently to DNA when linked by an adjacent leucine zipper domain. Then, the lack of efficient DNA binding displayed by plant HD-Zip monomers, rather than a requisite for the formation of active dimers, seems to constitute a means of regulating DNA binding and the formation of different protein complexes by changes in protein concentration.

We then asked if the higher affinity displayed by E-Zip was a consequence of an increase in the number of contacts with DNA respective to Hahb-4. To analyze this, we conducted missing nucleoside experiments on DNA subjected to hydroxyl radical cleavage. The results, shown in Figure 7, indicate that E-Zip makes extensive contacts with both strands of the pseudopalindromic Hahb-4 DNA binding site, including nucleotides adjacent to the 9-bp core. In comparison, Hahb-4 contacts the 3′ region of each strand, but not the 5′ region, including the first two nucleotides of the core. This different behavior may explain why one of the proteins is able to bind DNA as a monomer while the other is not.

DISCUSSION

HD-Zip proteins, though unique to plants, combine motifs found in a great number of transcription factors found in most eukaryotic organisms. The spatial relationship between the HD and the leucine zipper is similar to that observed between the basic DNA-binding domain and the zipper in b-Zip proteins (*17*). This led to the speculation that the dimerization motif plays a central role in the correct orientation of the recognition helix (helix III) along the major groove of DNA. The resemblance to b-Zip transcription factors also extends to the fact that HD-Zip proteins are unable to bind DNA as monomers, unlike several animal HDs (*7*, *8*, *17*, *18*). Given the high degree of sequence conservation observed in the entire HD, we became interested in determining which differences between these HDs are responsible for their different behavior. We have previously shown that the N-terminal arm of Antennapedia, a major determinant of DNA binding affinity, cannot confer efficient DNA binding to HD-Zip monomers and that positively charged residues at the N-terminal arm of the HD are required for efficient DNA binding by HD-Zip dimers (*21*). These results indicated that the different behavior may originate in differences in other part(s) of the HD.

We have chosen the EN HD, which has been the subject of detailed structural studies (*7*), as a framework to investigate the structural similarities and differences with the HD of the HD-Zip protein Hahb-4. These HDs are 35% identical

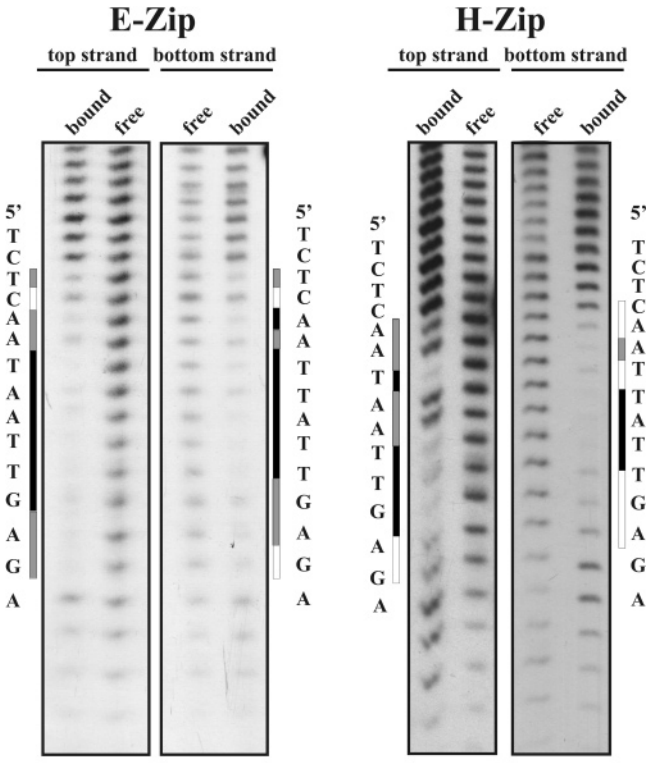

FIGURE 7: E-Zip shows extended DNA contacts with respect to Hahb-4. A missing nucleoside experiment for E-Zip and H-Zip binding to the pseudopalindromic Hahb-4 DNA binding site is shown. An oligonucleotide containing the Hahb-4 binding site was specifically labeled in one of its strands, subjected to hydroxyl radical attack, and incubated with either E-Zip or H-Zip. Bound and free fractions were separated and analyzed on a denaturing polyacrylamide gel. Beside each autoradiography, a schematic representation of the results (as boxes of different gray intensities according to the strength of interference) is shown, together with the sequence of each strand.

in sequence and the similarity raises to 45% when conserved amino acids are considered. The analysis of the DNA-binding preferences of a set of nine chimeric EN/Hahb-4 HDs plus the two original ones led us to conclude that the loop between helix I and helix II is a major determinant of the different DNA binding efficiencies displayed by the two HDs. Upon analyzing the EN HD−DNA crystal structure (*7*), it can be observed that loop residue Tyr25 makes a contact with a phosphate group of DNA [and also with the sugar moiety in Antennapedia (*8*)]. This residue is conserved in most HDs that bind DNA efficiently as monomers and is considered important for the correct positioning of helix III (*3*). Hahb-4 contains an Arg at this position. Among the HDs for which structures in complex with DNA have been determined, none contains Arg25. A theoretical model of the Hahb-4 HD bound to DNA using the program Swiss-Model (*26*) available in the ExPASy web server with the EN−DNA complex as template indicated that Arg25 may be unable to make similar contacts as those established by Tyr25 (not shown). Other HD-Zip proteins contain Thr or Lys at this position (*27*). According to our results, however, the inclusion of Tyr at position 25 alone is unable to promote Hahb-4 binding to a monomeric binding site, since the additional inclusion of Arg at position 24 is needed for this purpose. The exact role of this residue is unclear at present, but Arg24 is also conserved among HDs that bind DNA as monomers. The importance of loop amino acids 24 and 25 for efficient binding has also been noted in studies with the Mata1 HD, which binds DNA only as a partner of the Matalpha2 HD (*28*, *29*). When residues Gln24 and Ser25 of the Mata1 HD were mutated to Arg and Tyr, respectively, a significant increase in DNA binding affinity was observed (*30*). These changes produced an increased stability of the HD and an enlargement of helix III, the final residues of which are unstructured in wild-type Mata1. These studies point to an additional role of the loop in stabilizing the folding of the recognition helix. In fact, in the EN crystal structure, Tyr25 points to the C-terminal end of helix III and may be able to make hydrophobic contacts with the side chain of Arg53. A model structure of the Hahb-4 HD suggests that this contact is lost in HD-Zip proteins (not shown).

Thus, an important role of the loop may be to fix in the correct orientation the DNA contacting amino acids for efficient binding. The loss of these contacts may also explain the poor affinity for DNA displayed by the HD of HD-Zip proteins in the absence of a leucine zipper. The presence of additional amino acids (i.e.: those that compose the leucine zipper) at the C-terminus of the HD does not seem to compensate for this, but formation of the coiled-coil in dimers may be relevant to stabilize the C-terminal portion of the recognition helix. Nevertheless, as shown by the missing nucleoside experiments, the number of contacts made by HD-Zip dimers seems to be lower than those established by a couple of EN monomers linked through a leucine zipper, and this difference is not attributable to helix III. Conformational changes at the loop may also be used to modulate DNA binding capacity of HDs, a role that has also been proposed for the N-terminal arm (*31*).

The lack of a functional loop may be inherent to HD-Zip protein function. Since only dimers are active, DNA binding by these proteins increases cooperatively with protein concentration. It is well-established that several HD-Zip protein genes are significantly induced by environmental conditions (*32*−*35*). In this way, sharp responses may be obtained by modulating the amount of HD-Zip proteins within plant cells.

From an evolutionary point of view, the loop and the N-terminal arm of the HD may be later acquisitions added to the basic three helix module. Animal HOX proteins and yeast Matalpha2 use the N-terminal arm and the loop for efficient DNA binding (*3*, *7*−*9*, *12*). HD-Zip proteins seem to use the N-terminal arm for efficient binding as dimers (presumably through interaction of positive charges with the phosphate backbone) but not (unlike HOX proteins) as a determinant of specificity (*21*). As demonstrated in this study, the N-terminal arm plus helix I of Hahb-4 can functionally replace this segment of EN without any significant loss in DNA-binding capacity. The main difference between HOX and HD-Zip proteins is then constituted by the presence of a functional loop in the animal proteins. HD-Zip proteins, instead, have incorporated a leucine zipper that promotes dimer formation and efficient binding simply by increasing the number of contacts with DNA. Other HD proteins have increased their DNA binding efficiency through the ability to interact with other (notably HD) proteins (*36*, *37*). Many of these proteins belong to the TALE superclass and contain a three amino acid extension within the loop (*38*). Accordingly, the loop must be regarded as one of the segments that

contribute to the present-day diversity in the properties of different HDs.

ACKNOWLEDGMENT

We gratefully acknowledge Dr. Thomas Kornberg, University of California, San Francisco, for sending us an engrailed cDNA clone. We also acknowledge Dr. Raquel Chan for helpful comments and for critically reading the manuscript.

REFERENCES

1. Gehring, W. J. (1987) Homeo boxes 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. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Affolter, M., Percival-Smith, A., Müller, M., Leupin, W., and Gehring, W. J. (1990) DNA binding properties of the purified Antennapedia homeodomain, *Proc. Natl. Acad. Sci. U.S.A. 87*, 4093−4097.
13. Florence, B., Handrow, R., and Laughon, A. (1991) DNA-binding specificity of the fushi tarazu homeodomain, *Mol. Cell. Biol. 11*, 3613−3623.
14. Chan, R. L., Gago, G. M., Palena, C. M., and Gonzalez, D. H. (1998) Homeoboxes in plant development, *Biochim. Biophys. Acta 1442*, 1−19.
15. Ruberti, I., Sessa, G., Lucchetti, S., and Morelli, G. (1991) A novel class of proteins containing a homeodomain with a closely linked leucine zipper motif, *EMBO J. 10*, 1787−1791.
16. Schena, M., and Davis, R. W. (1992) HD-Zip proteins: Members of an *Arabidopsis* homeodomain protein superfamily, *Proc. Natl. Acad. Sci. U.S.A.* 89, 3894−3898.
17. Sessa, G., Morelli, G., and Ruberti, I. (1993) The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities, *EMBO J. 12*, 3507−3517.
18. Palena, C. M., Gonzalez, D. H., and 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.
19. 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.
20. Tron, A. E., Bertoncini, C. W., Chan, R. L., and Gonzalez, D. H. (2002) Redox regulation of plant homeodomain transcription factors, *J. Biol. Chem. 277*, 34800−34807.
21. Palena, C. M., Tron, A. E., Bertoncini, C. W., Gonzalez, D. H., and Chan, R. L. (2001) Positively charged residues at the N-terminal arm of the homeodomain are required for efficient DNA binding by homeodomain-leucine zipper proteins, *J. Mol. Biol. 308*, 39−47.
22. 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.
23. Palena, C. M., Gonzalez, D. H., Guelman, S., and Chan, R. L. (1998) Expression of sunflower homeodomain containing proteins in *Escherichia coli*: Purification and functional studies, *Protein Expr. Purif. 13*, 97−103.
24. 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.
25. 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.
26. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res. 31*, 3381−3385.
27. Gonzalez, D. H., Valle, E. M., Gago, G. M., and Chan, R. L. (1997) Interaction between proteins containing homeodomains associated to leucine zippers from sunflower, *Biochim. Biophys. Acta 1351*, 137−149.
28. Goutte, C., and Johnson, A. D. (1993) Yeast a1 and alpha2 homeodomain proteins form a DNA-binding activity with properties distinct from those of either protein, *J. Mol. Biol. 233*, 359−371.
29. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA, *Science 270*, 262−269.
30. Hart, B., Mathias, J. R., Ott, D., McNaughton, L., Anderson, J. S., Vershon, A. K., and Baxter, S. M. (2002) Engineered improvements in DNA-binding function of the MATa1 homeodomain reveal structural changes involved in combinatorial control, *J. Mol. Biol. 316*, 247−256.
31. Phelan, M. L., and Featherstone, M. S. (1997) Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX:PBX heterodimers, *J. Biol. Chem. 272*, 8635−8643.
32. Carabelli, M., Sessa, G., Baima, S., Morelli, G., and Ruberti, I. (1993) The *Arabidopsis Athb-2* and *-4* genes are strongly induced by far-red-rich light, *Plant J. 4*, 469−479.
33. Söderman, E., Mattsson, J., and Engström, P. (1996) The *Arabidopsis* homeobox gene *ATHB-7* is induced by water deficit and by abscisic acid, *Plant J. 10*, 375−381.
34. Lee, Y. H., and Chun, J. Y. (1998) A new homeodomain-leucine zipper gene from *Arabidopsis thaliana* induced by water stress and abscisic acid treatment, *Plant Mol. Biol. 37*, 377−384.
35. Gago, G. M., Almoguera, C., Jordano, J., Gonzalez, D. H., and Chan, R. L. (2002)*Hahb-4*, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower, *Plant Cell Environ. 25*, 633−640.
36. Phelan, M. L., Rambaldi, I., and Featherstone, M. S. (1995) Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif, *Mol. Cell. Biol. 15*, 3989−3997.
37. Knoepfler, P. S., Calvo, K. R., Chen, H., Antonarakis, S. E., and Kamps, M. P. (1997) Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1, *Proc. Natl. Acad. Sci. U.S.A. 94*, 14553−14558.
38. Bürglin, T. R. (1997) Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals, *Nucleic Acids. Res. 25*, 4173−4180.

BI048254A