The two homeodomains of the ZmHox2a gene from maize originated as an internal gene duplication and have evolved different target site specificities

Thomas Kirch, Stephan Bitter, Brigitte Kisters-Woike1 and Wolfgang Werr*

Institut für Entwicklungsbiologie and 1Institut für Genetik, Universität zu Köln, Gyrhofstraße 17, 50923 Köln, Germany

Received June 11, 1998; Revised and Accepted August 26, 1998

ABSTRACT

The maize ZmHox2a gene encodes two homeodomains which originated by a 699 bp duplication within an ancestral precursor. The sequences of the two ZmHox2a homeodomains are highly diverged in the N-terminal arm, while residues in the helical part have mostly been conserved. We show here that both ZmHox2a homeodomains are functional DNA-binding motifs but exhibit different target site specificities. CASTing experiments reveal a TCCT motif recognized by HD1 but a GATC tetranucleotide as the recognition sequence of HD2. Mutation of the central nucleotides in both tetranucleotide core motifs abolishes DNA binding. A domain swap experiment indicates that target site specificity is achieved in a combinatorial manner by the contributions of the diverged N-terminal arms together with the slightly different recognition helices. Computer modelling suggests that K47 and H54 in the recognition helices preferentially contact the bases at the 3′-terminus of the tetranucleotide target sequences.

INTRODUCTION

The homeodomain encoded by the homeobox is a DNA-binding motif identified initially in developmental control genes of Drosophila. Shortly thereafter homeobox genes were found to be ubiquitous in the animal kingdom and also in yeast reviewed in 1,2. Identification of the Knotted-1 gene of maize as a homeobox gene (3) confirmed that this DNA-binding motif occurs in developmental regulators in plants. Due to the importance of its biological function, major efforts have been made to clarify the three-dimensional structure of the homeodomain and its mode of interaction with DNA targets. NMR studies and X-ray crystallography performed with the Antennapedia (4), Engrailed (5) and Mat2 (6) homeodomains showed that the 60 amino acid homeodomain consists of three helices: helix 1 (10–21) is linked by a loop to helix 2 (28–38) which is connected by a turn to helix 3 (residues 42–59). Due to their increased flexibility, the last residues of helix 3 (53–59) are sometimes considered as a separate helix 4. Helices 2 and 3 resemble the prokaryotic helix–turn–helix motif, with recognition helix 3 located in the major groove. Helix 1 is oriented antiparallel to helix 2 and enables contacts between the N-terminal (1–9) arm and bases in the minor groove which have been shown to contribute to target site specificity. A peculiar feature of metazoan homeodomains is the similarity between individual target sequences, with the TATC core motif being frequently conserved (7).

The number of known homeobox genes in plants is rapidly increasing and there may be as many in plants as in animal species, but target sequences of plant homeodomains are largely unknown. In maize, in addition to the Knotted-1 class of homeodomains (3,8), a second unrelated family of maize homeobox genes was identified in an expression screen with the 26 bp feedback control element of the Shrunken promoter (9). Four ZmHox (Zea mays homeobox) genes have been cloned and analysed at the DNA sequence and RNA expression levels. ZmHox1a/ZmHox1b and ZmHox2a/ZmHox2b represent two gene pairs that lie in larger duplicated chromosomal regions, indicating that the pairs originated by an ancient duplication of the maize genome (10,11). Expression of ZmHox genes is restricted to meristems, suggesting a function in plant development and growth (12).

The ZmHox2a and ZmHox2b genes encode large polypeptides of 1576 and 1539 residues respectively and are 83% identical in amino acid sequence (11). Unique to the two ZmHox2 gene products are two complete homeodomains, referred to as HD1 and HD2, which are part of a larger 699 bp DNA sequence duplication including three exons. The degree of DNA sequence identity in this duplication is 72%. The encoded polypeptide sequences are only 51% identical, while the homeodomains are 64% identical and 77% similar (11). This indicates conservation of a dual DNA-binding function after duplication within an ancestral gene. Two DNA-binding motifs are not uncommon in eukaryotic transcription factors. Different DNA-binding motifs are present in the POU or Paired families of homeobox genes, which are named for specific domains that contribute to target site specificities in these animal transcription factors (13,14). In the mouse zfh-1 gene five homeodomains are encoded in combination with two Zn fingers (15).

The highest degree of similarity between the ZmHox2a homeodomains and their animal counterparts is observed for the Engrailed (Eng) class (2), which is most pronounced at hydrophobic core positions. In ZmHox2a HD1 and HD2 residues L16, L34, L38, L40, I/V45, W48 and F49 are identical to the Drosophila Engrailed homeodomain (the Val→Ile exchange at position 45 is common in metazoan counterparts. Similarity is lower at
positions 6, 8, 25, 28, 31, 43, 44, 52, 53 and 55, which contribute contacts between the Eng HD and the sugar–phosphate backbone of the DNA target. However, none of the amino acids at positions providing direct base contacts in the Engrailed domain (3, 5, 47, 50, 51 and 54; comparison in Fig. 1b) is identical in the ZmHox2a homeodomains. Compared with animal homeodomains, uncommon or novel residues are frequent at these positions in the ZmHox domains (11). HD1 and HD2 differ from each other at positions 3 and 5 in the N-terminal arm and position 51 in the recognition helix. The latter position is intriguing concerning DNA sequence specificity because the highly conserved N51 in animal homeodomains contacts a central adenine in the TAAT target and is changed to either G51 or S51 in the ZmHox2a homeodomains. It may also be noteworthy that of the first nine residues comprising the N-terminal arm no single amino acid is identical between HD1 and HD2.

We describe here experiments that confirm that both ZmHox2a homeodomains are bona fide DNA-binding motifs with strikingly different target site specificities. Exchange of the N-terminal arms relaxes target site specificity, their divergence therefore accounts for a major part of sequence specificity.

**MATERIALS AND METHODS**

**Expression of the homeodomains**

DNA fragments encoding the ZmHox2a homeodomains were amplified by PCR using the primers AGTAGTAGCACAGGAC-

AGAAGCAACCTTGAATA, GCCAGGTCTCAAAATAT and AACAGCATCTTTGAATA (numbers 1–4 in Fig. 1a). PCR fragments were first inserted into the Smal site of pUC19 and subsequently directionally cloned into pGEX I using the BamHI and EcoRI polynuker linkers, resulting in plusmids pGEX-HD1 and pGEX-HD2. Creation of pGEX-HD1/2 followed the same strategy except that the PCR fragment was amplified using primers 1 and 4. To create HD1ex a HindIII/EcoRI double digestion was performed with pGEX-HD2 releasing the 3′-terminal sequences of the homeodomain, coding for the segment from helix 1 to the C-terminus of HD2. The corresponding HD1 sequences were amplified by PCR with the 5′ primer introducing the missing HindIII site and the 3′ primer overlapping the EcoRI site. After a HindIII/EcoRI double digestion of the resulting PCR fragment, the HD1 coding sequences were inserted into the pGEX-HD2 vector and verified by sequencing. The individual glutathione S-transferase (GST)–HD fusion proteins were overexpressed in Escherichia coli BL21(DE3) and isolated by batch affinity chromatography on glutathione–agarose beads. The purity of GST–HD fusions was confirmed by SDS–PAGE.

**Electrophoretic mobility shift assays (EMSA)**

All EMSA experiments were performed with synthetic oligonucleotides as indicated in Figure 2b. Upper and lower strand sequences were synthesized with a 5′-terminal AATT extension. After annealing, double-stranded oligonucleotides were labelled by Klenow fill-in reaction with [α-32P]ATP (30 µCi) in the presence of the other three nucleotides (50 µM each) and PAGE purified. Radioactively labelled DNA, GST fusion proteins and competitor DNA (sonicated calf thymus DNA or unlabelled oligonucleotide) were mixed and incubated for 10 min at room temperature in 4% glycerol, 1 mM EDTA, pH 7.5, 10 mM 2-mercaptoethanol, 10 mM Tris, pH 7.5. The samples were loaded on 5% polyacrylamide (80:1) gels (1 mm thick) and electrophoresed in 1 mM EDTA, 3.3 mM Na acetate, 6.7 mM Tris, pH 8.0, at 120 V at 4°C in a cold room. Gels were dried and autoradiographed on Kodak XAR5 film.

**Cyclic amplification of selected targets (CASTing)**

Homeodomain binding site selection was performed essentially as described by Pollock and Treisman (16). The oligonucleotides (sequences in Fig. 1c) contained a random 15 nt core flanked by conserved primer sequences for PCR amplification. In total three oligonucleotides were synthesized, the full-length oligonucleotide shown in Figure 1c and the upstream and the complementary downstream primers. Preparation of the double-stranded CASTing oligo was performed by annealing the full-length oligonucleotide with polynucleotide kinase 5′-border of the random core that was changed to TCG is underlined.
RESULTS

Expression of the ZmHox2a homeodomains and selection of DNA target sites

The sequences encoding HD1 and HD2 were amplified by PCR from the ZmHox2a cDNA clone and inserted into the pGEX 1 vector for expression as translational fusions with GST. The structure of the ZmHox2a protein and the primer positions relative to the homeodomains are shown in Figure 1a. In addition, primer combination 1 and 4 was used to amplify, clone and express HD1 and HD2 in one polypeptide, separated by the natural intervening amino acid sequences (GST–HD1/2). All GST constructs were verified by DNA sequencing and transformed into E.coli strain BL21 for protein expression. Fusion proteins were affinity purified on glutathione–agarose and analysed by SDS–PAGE. Both GST–HD1 and GST–HD2 migrated with the predicted molecular weight of 33 kDa, with GST–HD2 reproducibly exhibiting some degradation. Preparations of the GST–HD1/2 fusion (64 kDa band) were always contaminated with putative degradation products of higher mobility (data not shown). The amino acid sequences of HD1 and HD2 were compared with the Drosophila Eng homeodomain in Figure 1b; the hydrophobic core, backbone and base-contacting residues were identified based on the Eng X-ray data (5).

HD1 and HD2 target sites were selected by CASTing experiments (16). In this PCR-based technique recognition sites are enriched from a population of random oligonucleotides by multiple rounds of DNA binding in EMSA experiments, recovery of nucleoprotein complexes and reamplification of protein-bound oligonucleotides. We used a 15 bp random core flanked by 20 bp primer sequences containing EcoRI or PstI restriction sites for subsequent directional cloning as shown in Figure 1c. The sequence of the downstream primer was altered as results of the GST–HD1 selection experiment indicated that the AGG trinucleotide upstream of the random core contributed part of a HD1 binding site (Fig. 1c). The preparation of double-stranded targets, isolation of nucleoprotein complexes and reamplification of released targets are described in Materials and Methods.

HD1 selects TCCT while HD2 favours GATC tetranucleotides

Potential HD1 binding sites were enriched by six cycles of DNA binding, isolation of nucleoprotein complexes and reamplification of selected oligonucleotides. After the last reamplification step oligonucleotides were digested with EcoRI and PstI and sequenced after directional cloning. Of 71 independent oligonucleotides recovered, 48 contained one to three TCCT tetranucleotides. As 10 of these tetranucleotide motifs comprised the complementary AGG trinucleotide upstream of the random core (15′ primer sequence (1c), a modified oligonucleotide 2 in which the AGG was replaced by TCG (Fig. 1c) was used in a second CASTing experiment with HD1. Confirming the first result, the TCCT tetranucleotide motif was obtained in 12 of 24 sequences after five cycles. In conclusion, from a total of 95 sequences 60 contained at least one TCCT motif. Strikingly, many of the remaining 35 oligonucleotides (23 in experiment 1 and 12 in experiment 2) recovered had one or two CCT or TCC tetranucleotides in the random core, submotifs which may facilitate selection.

The same experimental routine was performed with GST–HD2 and the improved oligonucleotide 2. Enriched target sites were cloned and sequenced after seven cycles of DNA binding and reamplification. Of 68 oligonucleotides sequenced 39 contained a GATC tetranucleotide and six of these had two GATC subsequences within the random 15 nt core. A further 19 sequences contained a GATT motif, while only a single TCCT motif was found among the HD2-selected targets. This striking difference argues that the two populations truly reflect sequence preferences of HD1 and HD2. The 2-fold higher number of GATC versus GATT tetranucleotides bound by HD2 is taken as evidence that GATC is the preferred HD2 target site. For the consensus
sequence in Figure 2a nucleotide preferences at position –1, +1 and +2 outside the tetranucleotide motifs have also been considered. To avoid possible bias introduced by the conserved flanking primer sequences the effective numbers were reduced to 38 TCCT and 27 GATC tetranucleotides located entirely in the random core.

The TCCT or GATC tetranucleotides are essential for HD1 or HD2 binding

The interaction of HD1 or HD2 with oligonucleotides selected in CASTing experiments described above was initially tested in EMSA experiments with various of the cloned sequences released from the pUC polylinker sequences. But the comparison between HD1 and HD2 and individual target sequences was generally hampered by additional TCCT or GATC submotifs residing in the random core or the flanking polylinker sequences. We therefore took the nucleoprotein complexes detected with these target sequences only as an indication that the tetranucleotides are recognized within different sequence surroundings (data not shown). To conclusively compare and verify both the TCCT and the GATC motifs as HD1 or HD2 recognition sequences, we designed new synthetic oligonucleotides that minimized the contributions of flanking sequences. The sequence environment of the TCCT or GATC tetranucleotides was kept constant, except that sequence preferences at positions –1, +1 and +2 (Fig. 2a) were taken into account. The respective sequences are shown in Figure 2b. Mutated versions of the tetranucleotides carry two central nucleotide exchanges, TCCT→TaaT and GATC→GgcG (Fig. 2b). One reason for exchanging the central C residues in the TCCT motif to A residues was to test whether the core found in binding sites of animal genomes acts as a target site. All DNA binding experiments described below were performed with these defined oligonucleotides.

The EMSA experiments in Figure 3a and b combine the results of binding studies performed with the GST–HD1 fusion peptide and various oligonucleotides. The result in Figure 3a demonstrates that HD1 forms nucleoprotein complexes with the TCCT but not with the GATC or the mutated TaaT oligo. The four lane sets with each oligonucleotide contain increasing concentrations of sonificated calf thymus DNA. The slight decrease in the amount of TCCT–HD1 nucleoprotein across this range is explained by the higher levels of target sites in the competitor DNA. Discrimination against the TaaT motif by HD1 is further supported by the competition experiment in Figure 3b. Nucleoprotein complex formation between HD1 and the radioactively labelled TCCT oligonucleotide is competed out by a 300-fold molar excess of unlabelled TCCT and GATC oligonucleotides, respectively. (d) HD2 binding to radioactively labelled GATC, GgcG and TCCT oligonucleotides (lanes 1–4, 5–8 and 9–12). The lanes for each oligo from left to right contain a 200-, 300-, 500- and 1000-fold mass excess of calf thymus competitor DNA. Lane 13, labelled GATC oligonucleotide in the absence of protein. (d) Competition experiment with the radioactively labelled GATC oligo and HD2. Lane 1, nucleoprotein complexes formed in the absence of competing oligonucleotide; lanes 2 and 3, a 200-fold molar excess of unlabelled TCCT and TaaT oligonucleotides, respectively. (c) HD2 binding to radioactively labelled GATC, GgcG and TCCT (lanes 1–4, 5–8 and 9–12) oligonucleotides. The lanes for each oligo from left to right contain a 200-, 300-, 500- and 1000-fold mass excess of calf thymus competitor DNA. Lane 13, labelled GATC oligonucleotide in the absence of protein. The corresponding experiments for HD2 are shown in Figure 3c and d; the results are comparable with those for HD1. Neither the TCCT nor the GgcG oligo are targets for HD2, while the GATC oligonucleotide forms stable nucleoprotein complexes in the presence of increasing concentrations of non-specific competitor DNA (Fig. 3c). The mutated GgcG oligonucleotide also failed to compete with the labelled GATC oligonucleotide (Fig. 3d). Therefore HD2 binding is dependent on the integrity of the GATC core sequence. In summary, the EMSA experiments described here verify the selective affinity of HD1 and HD2 for the TCCT and GATC tetranucleotides, respectively, suggested by the CASTing experiments. Consequently, the two ZmHox2a homeodomains have evolved entirely different target site specificities.

Computer modelling

Both ZmHox2a homeodomains exhibit the highest degree of similarity to the Engrailed class of homeodomains, this being most pronounced at the hydrophobic core positions. Thus it was tempting to determine sites of potential base contacts of HD1 and HD2 with the respective oligonucleotide targets by computer modelling. Starting from the X-ray data for DNA–Eng HD co-crystals the TAAT target was changed to TCCT or GATC (flanking sequences in Fig. 2b), respectively, and the individual amino acid side chains were altered to the corresponding maize HD1 or HD2 sequences. On the assumption that the recognition helix of the maize homeodomains is arranged as in the DNA–Eng HD co-crystal these computer simulations suggest two crucial positions which may contribute to target site recognition (Fig. 4).
Firstly, H54 exclusively found in ZmHox homeodomains (2) presumably contacts the lower strand purine at the fourth position of both tetranucleotides and may establish dipole–dipole interactions to either the adenine or guanine residue. Secondly, the ε amino group of K47 in HD1 and HD2, which replaces the rather invariant I47 of metazoan homeodomains, invades the major groove between positions 3 and 4 of the TCCT motif or the GATC motif and may contact both pyrimidine residues in the upper strand. The modelling data therefore support enrichment of the 19 additional GATT tetranucleotides in the HD2 CASTing experiments, which may ultimately reflect a relaxed specificity of the K47 side chain. Unfortunately the modelling does not explain the target site specificity differences between HD1 and HD2. The side chains of both S51 and G51, the only amino acid substitutions in the recognition helix between HD1 and HD2 in a position involved in base contacts in the Eng HD, are too short to provide base contacts in the major groove. The same is true for S50, which replaces the prevalent Q50 (2) in both ZmHox2a homeodomains.

Divergence of the N-terminal arm contributes to target site selectivity

Not a single residue is identical between the two ZmHox2a homeodomains in the nine amino acids representing the flexible N-terminal arm. To test whether this sequence diversity mediates selectivity between the TCCT and GATC core motifs a domain swap experiment was performed. Using a natural HindIII restriction site (codons K15 and L16, Fig. 1b) in helix 1 of HD2, residues 1–14 of HD2 were transferred to the HD1 C-terminus (HD1ex). Besides the exchange of the first nine amino acid residues, HD1ex contains two additional exchanges, E11→H11 and T13→S13, but neither of these last positions has so far been shown to contribute DNA contacts (2). The resulting chimeric HD1ex homeodomain was expressed as a translational fusion with GST and tested for DNA binding in EMSA experiments (Fig. 5). Nucleoprotein complexes were formed with both the TCCT or GATC oligonucleotides and somewhat more weakly with the mutated TaaT oligonucleotide, but the GcgC tetranucleotide was hardly recognized at all. Thus by exchange of the N-terminal arm sequence selectivity has not been altered reciprocally from TCCT to GATC but relaxed to both motifs as well as other sequences, as indicated by nucleoprotein complexes formed with the TaaT oligonucleotide.

Individual target site specificities are maintained in the HD1–HD2 polypeptide

To test for combinatorial or synergistic effects between HD1 and HD2 a ZmHox2a–GST fusion polypeptide extending from the first residue of HD1 to the last amino acid of HD2 (Fig. 1a) was tested in EMSA experiments. The results with this GST–HD1/2 polypeptide are shown in Figure 6 and demonstrate binding to both the TCCT and GATC oligonucleotides. Competition experiments with an excess of unlabelled homologous or heterologous TCCT or GATC oligos, included in Figure 6, indicate that its target site is recognized by each homeodomain independently. Nucleoprotein complex formation with the radioactively labelled TCCT or GATC oligonucleotide is reduced by an excess of the homologous TCCT but not the GATC oligo. Vice versa, GATC binding is sensitive to addition of unlabelled GATC oligonucleotide but largely unaffected by an excess of the heterologous TCCT target. Although connected by an intervening peptide sequence HD1 and HD2 therefore interact relatively independently with target oligonucleotides in this experiment. However, proper folding and, consequently, correct spatial relationship of both ZmHox2a homeodomains may be dependent on the presence of additional N- and C-terminal ZmHox2a polypeptide sequences which were found to interfere with expression of soluble products in E.coli.
DISCUSSION

Dual DNA binding functions in the ZmHox2 gene products

We have shown here that ZmHox2a HD1 and HD2 are bona fide DNA-binding motifs with different target sites specificities (TCCT and GATC, respectively). All available sequence information indicates that the two homeodomains originated by duplication within an ancestral gene; consequently the target site specificity of both was originally identical. The greater conservation of the homeodomains relative to the duplicated flanking sequences indicates selection for the dual DNA-binding function. The immediate advantage of two identical DNA-binding motifs in one polypeptide is the increased probability of DNA contacts. In the original duplicated polypeptide, scanning DNA for new targets with the free homeodomain while bound to the DNA via the second motif might have allowed far more efficient selection of high affinity DNA-binding sites. Following the evolution of different target site specificities this simple possibility is now unlikely. An alternative idea is that cooperativity between HD1 and HD2 is needed to increase target site specificity. This situation resembles that of plant-specific leucine zipper homeodomain proteins, which have been shown to form DNA-binding homodimers (17). The combination of two homeodomains to increase DNA-binding specificity is thus not uncommon in plants. However, the real advantage of dimerization via a leucine zipper is not homo- but more likely heterodimerization, eventually combining DNA-binding modules with different target site specificities. Without exploiting the combinatorial possibilities of dimerization, a similar situation has arisen in the ZmHox2a gene product, where HD1 and HD2 have acquired different target site specificities. It should be emphasized here that CASTing experiments performed with GST–HD1/2 were unsuccessful when the 15 nt random core oligos were used. However, we can presently neither exclude inappropriate polypeptide folding nor the possibility that critical spatial requirements for adjacent target motifs are not fulfilled in the 15 bp random core of the CASTing oligonucleotides.

Cooperativity between both ZmHox2a homeodomains might result in a reduction in the number of potential binding sites in the maize genome, but a second alternative is that the sequence-specific differences in target recognition between HD1 and HD2 are used to transduce signals to different targets or genes. The ZmHox2a gene product is a large, 1576 amino acid polypeptide with multiple, highly conserved protein domains, including the homeodomains. Alternative DNA contacts contributed by either HD1 or HD2 could have been advantageous during evolution in increasing the number of potential downstream target genes. In the plant kingdom the rice GT-2 transcription factor contains two basic DNA-binding domains with different specificities and is involved in phytochrome-mediated signal transduction (18). For neither the GT-2 nor for the ZmHox2 gene products is it presently clear whether cooperativity of DNA-binding domains focusses a signal to fewer genes or increases the bandwidth of signalling. Target sites, however, need not reside in different genes; interactions with TCCT or GATC sequences in one gene by the ZmHox2a polypeptide may bring remote sequence elements together. Such a mechanism could be related to enhancer selection, e.g. positioning distal transcriptional enhancer elements in proximity to promoter sequences. It could be of interest relative to plant growth as expression of the ZmHox2 genes is confined to meristems and proliferating cells of the maize plant. Irrespective of what evolutionary constraints maintain two functional homeodomains in one polypeptide, it should be pointed out here that the mouse zfx-1 gene combines five homeodomains with two zinc fingers (15). It will therefore be interesting to determine what the multiplicity of DNA-binding motifs is used for in this animal gene.

Divergence of target site specificities

The two ZmHox2a homeodomains, as part of a larger sequence duplication, provide an excellent example of the divergence of sequence specificity. Of the 51 residues in the helical regions 39 are identical and we calculate 86% similarity due to isomorphic replacements. Not surprisingly, hydrophobic core residues essential for the homeodomain three-dimensional structure generally exhibit the highest conservation between homeodomains. HD1 and HD2 and Drosophila Eng HD are identical at positions 16, 34, 38, 40, 45, 48 and 49 and the S35→A35 and V45→I45 exchanges observed between HD1 and HD2 are not uncommon in metazoan homeodomains (2,5). None of these positions has so far been associated with a specificity change. The only other hydrophobic core position that differs between the ZmHox2a homeodomains is T13, which is rather frequent in various paralogous genes of the HOX clusters, versus S13, a rare but isomorphic replacement at this position in HD2. A second class of residues in the homeodomains contributes sugar–phosphate backbone contacts to the DNA double helix (open circles in Fig. 1). Of these positions 25, 28, 31, 43, 44, 42, 52, 53 and 55 are identical between HD1 and HD2, with Y25, Q44 and R53 being identical to Eng (5). Only positions 6 (R→Y) and 8 (Y→F) in the N-terminal arm differ between the two ZmHox2a homeodomains and consequently might account for differences in DNA recognition. The most crucial amino acids side chains for target site specificity are those that contribute base contacts (triangles in Fig. 1). Three exchanges exist here between HD1 and HD2; two are in the flexible N-terminal arm but the most intriguing is residue 51 (S→G) in the DNA recognition helix, occupied almost invariably by N51 in the majority of homeodomains, including Kn.1 and relatives from maize and Arabidopsis (2,8). The computer modelling data, however, argue that neither S51 nor G51 protrude deeply enough
into the major groove to provide base contacts that could account for the sequence specificity differences exhibited by HD1 and HD2.

Alternatively, packing of the homeodomains has been shown to be crucial for sequence recognition. The TTF-1 homeodomain is identical to the Ant homeodomain at all positions contacting the TAAT core motif but recognizes a CAAG motif (19) as changes in three-dimensional structure lead to qualitatively different contacts (2, 20, 21). A detailed analysis of the interactions of TTF-1 HD with the DNA target indicates a combinatorial mode of sequence recognition (22). The TTF-1 HD Y54 recognition helix is responsible for selection of the 3′ G residue, while minor groove contacts mediated by residues at positions 6–8 in the N-terminal arm contribute essential contacts to the 5′-end of the CAAG core element. Similarly, the modelling data for both ZmHox2a homeodomains suggest that residues H54 and K47 may both impinge on the 3′-end of the recognition motifs (Fig. 4) while the N-terminal surface of the recognition helices remains relatively remote from the DNA targets. A major contribution of the N-terminal arm sequences to target site specificity as in the TTF-1 HD is indicated by the domain swap experiment, with the 14 N-terminal residues of HD2 fused to the HD1 C-terminus (Fig. 5). Despite its undirectionality, this domain swap experiment proves that discrimination between the TCCT and GATC tetranucleotides by HD1 is dependent on interaction of the N-terminal arm with the corresponding C-terminal helices. In consequence, amino acid exchanges in the helices between HD1 and HD2 have subtle effects on DNA contacts, three-dimensional structure or other factors not evident in computer simulations which, however, contribute to target site specificity. The relaxation of binding specificity observed for three motifs, TCCT, GATC and TaaT, is presumably best explained by the assumption that the chimeric HD1ex has a novel sequence specificity.

Apart from the pronounced differences in target site specificity between HD1 and HD2 it should be emphasized that neither the TCCT nor the GATC tetranucleotide resembles the TAAT motif favoured by many homeodomains in animal proteins and some plant representatives (17). However, binding to the TaaT target was observed neither with HD1 (Fig. 3) nor HD2 (data not shown). This result is interesting with respect to the conserved positions of two small introns in all ZmHox homeoboxes that are shared with the ceh-5 and ceh-1,-2,-7,-9 and -12 gene products from Caenorhabditis elegans and the Drosophila lab, pb, Abd-B and Dil genes (11; W.Werr, in preparation). These intron positions differ from the location of the single intron in the Knoted class of homeobox genes (3,8) and therefore link the ZmHox genes more closely evolutionarily to these animal counterparts.

In summary, the data described here confirm both ZmHox2a homeodomains as functional DNA-binding motifs which acquired quite different target site specificities following a duplication. Divergence of the N-terminal arm and a few amino acid exchanges in the helices of HD1 and HD2 together are responsible for discrimination between the target tetranucleotides TCCT and GATC, respectively.

ACKNOWLEDGEMENTS

We thank Dr Paul Hardy for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through SFB 274.

REFERENCES