The EMBO Journal vol.14 no.24 pp.6280–6291, 1995

Structure of the Even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity

Joel A.Hirsch and Aneel K.Aggarwal

Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA

even-skipped **is a homeobox gene important in controlling segment patterning in the embryonic fruit fly. Its homeobox encodes a DNA binding domain which binds with similar affinities to two DNA consensus sequences, one AT-rich, the other GC-rich. We describe a crystallographic analysis of the Even-skipped homeodomain complexed to an AT-rich oligonucleotide at 2.0 Å resolution. The structure reveals a novel arrangement of two homeodomains bound to one 10 bp DNA sequence in a tandem fashion. This arrangement suggests a mechanism for the homeoproteins' regulatory specificity. In addition, the functionally important residue Gln50 is observed in multiple conformations making direct and water-mediated hydrogen bonds with the DNA bases.**

Keywords: crystal structure/*even-skipped*/homeodomain/protein–DNA interactions

Introduction

even-skipped (*eve*) is a member of the pair-rule class of segmentation genes in *Drosophila melanogaster*, required for the establishment of segment patterning in the developing embryo (Nusslein-Volhard and Wieschaus, 1980). The gene encodes a protein (Eve) that functions as a repressor and possibly as an activator of transcription (Biggin and Tjian, 1989; Jiang *et al.*, 1991; Manoukian and Krause, 1992; Han and Manley, 1993; TenHarmsel *et al.*, 1993). The DNA binding activity of Eve is primarily due to a conserved 60 amino acid region known as the homeodomain (HD), located near the N-terminal end of the protein (Hoey and Levine, 1988; Hoey *et al.*, 1988). HDs have now been found in a variety of developmentally important proteins isolated from organisms ranging from yeast to man (Gehring *et al.*, 1994).

HD sequences can be grouped into several classes based on various sequence criteria (Bürglin, 1994). The Eve class of HDs has now at least 10 members. Amongst invertebrates, homologs of the Eve HD have been found in grasshopper (Patel *et al.*, 1992) and sponge (Miles and Miller, 1992). Vertebrate homologs have been found for zebrafish (Joly *et al.*, 1993), mice (Bastian and Gruss, 1990) and humans (D'Esposito *et al.*, 1991; Faiella *et al.*, 1991). The Eve HD has been extraordinarily conserved over evolution, with only two differences out of a possible 60 from fly to humans. In all cases, the Eve homologs appear to play an important role in development. More specifically, Eve participates in neuronal development in several different species (Doe *et al.*, 1988; Bastian and Gruss, 1990; Patel *et al.*, 1992).

The Eve HD has been shown to bind with approximately equal affinity to two categories of DNA consensus sequences, derived from the *engrailed* and *eve* promoters: an AT-rich site (10 bp) and a GC-rich site (9 bp) (Hoey and Levine, 1988; Hoey *et al.*, 1988). The AT-rich sequence (TCAATTAAAT) is bound by many different HD proteins *in vitro* (Desplan *et al.*, 1988; Hoey and Levine, 1988). Crystallographic analyses of Engrailed (En) (Kissinger *et al.*, 1990) and Matα2 (Wolberger *et al.*, 1991) HDs and NMR spectroscopic analysis of the Antennapedia (Antp) HD (Otting *et al.*, 1990) complexed to AT-rich oligonucleotides have helped to define the DNA core sequences important for binding: ATTA for En and Antp and GTAA for Matα2. We report here the three-dimensional structure of the Eve HD complexed to two repeats of its AT-rich consensus sequence at 2.0 Å resolution.

Results

Crystallography

Initially we expected that one HD would bind to one 10 bp consensus sequence. *In vitro* experiments with HDs implied an 8–9 bp preference (Ekker *et al.*, 1991). Also, DNase I footprinting protected an ~10 bp length (Hoey and Levine, 1988; Hoey *et al.*, 1988), while other biochemical studies had shown that HDs bind as monomers (Affolter *et al.*, 1990; Florence *et al.*, 1991). Thus it was a surprise when we discovered that two HDs are bound by one 10 bp consensus sequence in a tandem fashion. HD I binds to one face of the DNA and HD II binds to the opposite face, each with their recognition helices fitting properly in the major groove (Figures 1 and 2). This novel

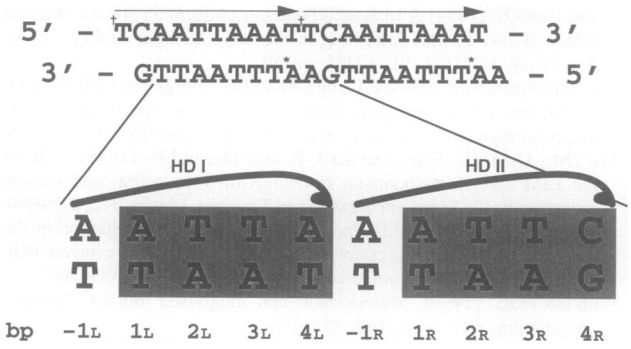

Fig. 1. At the top is shown the DNA oligomer used for co-crystallization. Arrows over the sequence indicate the consensus sequence (Hoey and Levine, 1988; Hoey *et al.*, 1988). † and * denote thymines substituted by iodouracils in derivatives Iodo1 and Iodo2 respectively (see Table I). The bottom duplex is a magnification, showing the tandem binding arrangement of HDs. Due to translational mixing of the DNA in the crystals both HDs bind to the averaged site AATTA/C (see text). The ATTA/C core sequences have been shaded for emphasis.

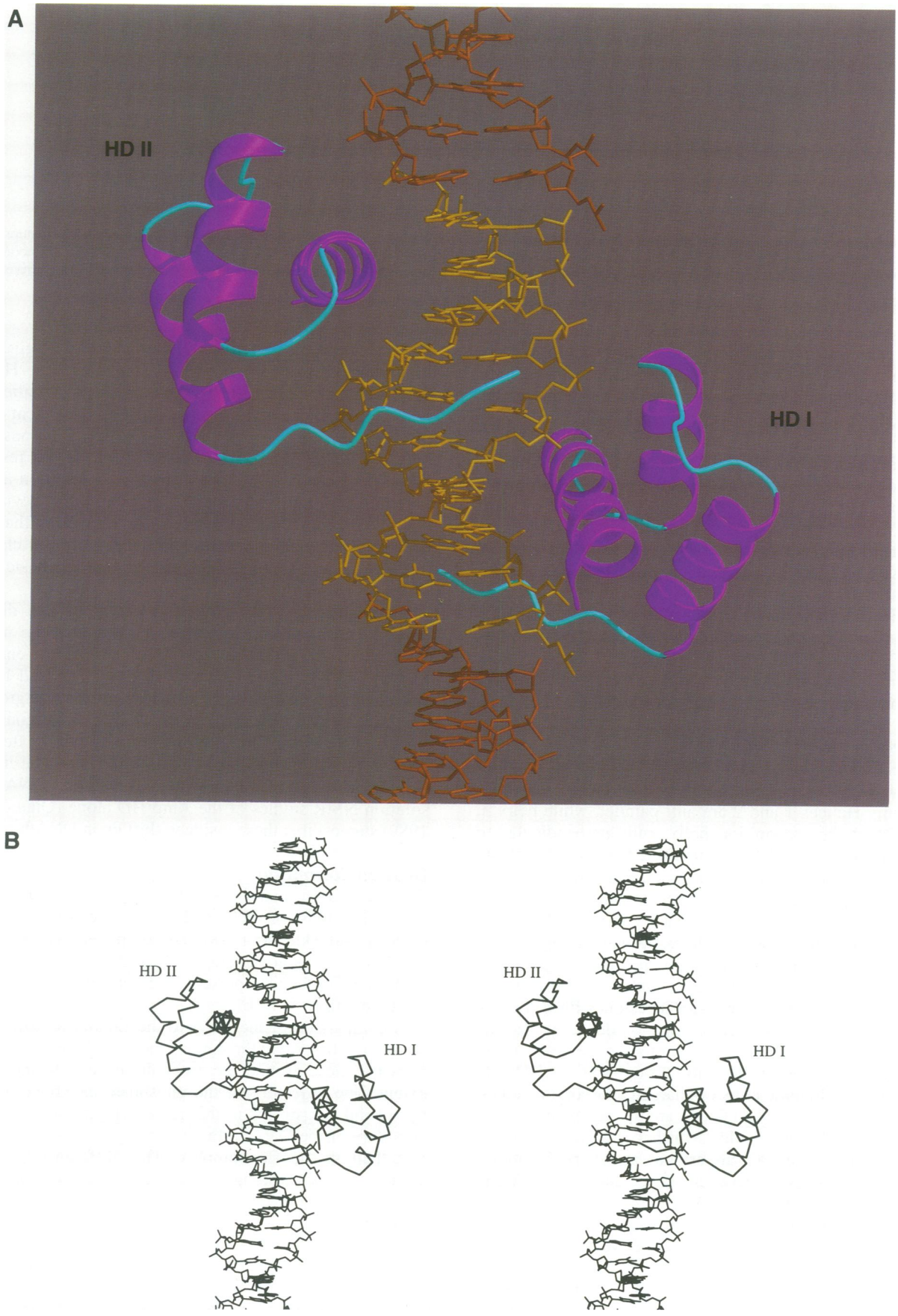

Fig. 2. Eve HD binds in a tandem fashion on opposite faces of the DNA. The recognition helix fits into the major groove and the flexible N-terminal arm lies in the minor groove. (**A**) A schematic representation of the complex. The DNA (colored yellow) is the 10 bp duplex shown at the bottom of Figure 1. Drawn with Raster3D (Merritt and Murphy, 1994). (**B**) A stereo representation. Drawn with MOLSCRIPT (Kraulis, 1991).

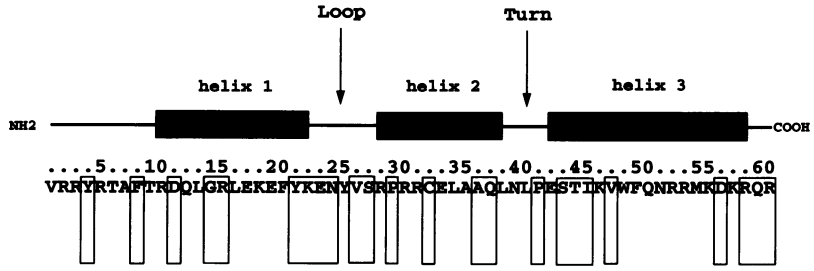

Fig. 3. Schematic diagram of the Eve HD polypeptide sequence. The black boxes over the sequence indicate α-helices. The secondary structure assignment was performed with the algorithm of Kabsch and Sander (1983) as implemented in PROCHECK (Laskowski *et al.*, 1993). Amino acids which are enclosed in boxes are those conserved within the Eve class of HDs and represent differences from the consensus HD sequence arrived at by Bürglin (1994).

configuration vividly illustrates the modular nature of HD–DNA recognition. The configuration bears resemblance to the POU domain–DNA complex structure (Klemm *et al.*, 1994); both result in helix–turn–helix (HTH) binding on opposite faces of the DNA.

Closer inspection of the 10 bp consensus sequence reveals the basis for the binding of two HDs. The sequence contains two directly repeated subsites, AATTA and AATTC, that differ by only 1 bp. In Figure 1 we have shown HD I and HD II as interacting with subsites AATTA and ATTAC. This is for convenience, because, due to translational mixing of the DNA in our crystals (see Material and methods), the sequence bound by both HDs is in essence the same and an average of the two subsites (AATTA/C). (Unlike the DNA, the two HDs remain unaveraged, as evidenced by distinct electron densities for residue Gln50; see Materials and methods.)

Protein conformation

The conformation of both HDs is, as expected, tri-α-helical with an N-terminal arm that lacks secondary structure. Helices 1 and 2 are anti-parallel, while helix 3, also called the recognition helix, runs perpendicular to them. Helices 1 and 2 are connected by a loop of six residues which are in an extended conformation with alternating residues either exposed or buried in the core of the domain. Helices 2 and 3 are connected by a turn and represent the helix–turn–helix motif (see Figure 3).

The hydrophobic core of the polypeptide is a parallelepiped circumscribed by helices 1 and 2 on one side and helix 3 on the other. The non-polar residues Phe8, Leu16, Phe20, Val26, Leu34, Leu38, Leu40, Ile45, Trp48 and Phe49 play the primary role in defining the core. These residues are amongst the most conserved in HDs. In addition, the aliphatic arms of charged and polar residues, such as Arg31, Arg52 and Glu42 are buried and contribute to the core. The basic fold is further stabilized by intrahelical and interhelical salt bridges at several locations. Specifically, ion pairs between Arg15 and Glu19, Arg30 and Glu33 and Arg52 and Asp56 represent intrahelical interactions. Interhelical salt bridges are found between Glu17 and Arg52 (helices 1 and 3), Glu19 and Arg30/Arg31 (helices 1 and 2) and Arg28 and Glu42 (helices 2 and 3). Interestingly, the structure of the En HD shows some residues in common with Eve that are ion paired in the same way (Clarke *et al.*, 1994). Furthermore, many of the Eve residues involved in salt bridge formation are conserved across HD classes.

A superposition with the DNA-bound En HD gives an r.m.s. deviation of 0.9 Å for all Cα atoms. Eve HD is 43% identical with the En HD, but has been classified by several sequence criteria to represent an independent class of HDs (Scott *et al.*, 1989; Bürglin, 1994). Superposition with the Matα2 HD (25% identity), beginning at residue 9 and not including the three amino acid insertion at the C-terminus of helix 1 in Matα2, gives an r.m.s. deviation of 1.2 Å. From these superpositions it is clear that the conserved differences between families and even classes represent more subtle distinctions beyond the basic fold of this domain.

The trajectory of the N-terminal arm in Eve differs from the En structure, as it appears to follow the minor groove versus reaching across it. Such a trajectory enables Eve to hug the minor groove, with the long and bulky side chains that comprise the arm inserting into the groove. The arm is also positioned to make backbone contacts. Residues preceding Thr6 presumably adopt their extended conformation only upon association with DNA. The crystal structures of the En HD with and without DNA and solution NMR studies of the Antp HD alone (Qian *et al.*, 1989) suggest that these residues shift upon DNA binding.

DNA conformation

The DNA conformation is in the canonical B form, with an average twist of 35.9° and an average rise of 3.41 Å (Lavery and Sklenar, 1988). The propeller twist in some AT base pairs is considerable (maximum of –25.2°), but in the range of twists present in other studies containing AT base pairs.

A comparison of the DNA to the En and Matα2 DNA structures shows that the major groove in the Eve complex does not close around the recognition helix to the same extent. The trajectory of the phosphate backbone differs by several Ångstroms in this region (Figure 4). Thus the curvature of the DNA in the Eve complex is substantially less than in the En complex. This difference may be caused by the binding of two Eve HDs on opposite DNA faces, requiring the DNA to adopt a 'straighter' conformation.

Protein–DNA interface

Protein–DNA contacts arise primarily from the recognition helix (helix 3), which projects across the major groove of the DNA, and the N-terminal arm, which winds along the minor groove. The structures of En and Antp HD complexes have defined an ATTA (or TAAT on the comple-

Fig. 4. A stereo plot of the En DNA (black) and the Eve DNA (red) phosphate backbones. The complexes were superimposed by matching the Cα atoms from the respective HDs. The view is along the recognition helix axis. As discussed in the text, the En DNA closes around the recognition helix to a greater extent than the Eve DNA.

mentary strand) core sequence important for these interactions. Eve HD I and HD II interact with a matching ATTA/C core sequence (labeled bp 1–4 in Figure 5). Electron density of the interface between the recognition helix and the major groove is well defined (Figure 6).

Val47, Asn51, Met54. Contacts that are in common with the previous structures include hydrophobic interactions between Val47 and the methyl group of thymine of bp 1 and bidentate hydrogen bonds between Asn51 and the adenine of bp 2. These contacts explain the very strong conservation of Ile/Val47 and the absolute conservation of Asn51 in all HDs (Bürglin, 1994). We also observe in both Eve HDs Met54 making van der Waals contacts to the N7 atom of the adenine of bp 1. In En, Met54 is replaced by an alanine residue, consequently no such interaction was seen in the crystal structure. The Antp structure, however, indicates a similar contact as Eve.

N-terminal arm. Base pairs 3 and 4 of the core sequence are contacted in the minor groove by residues Arg3, Tyr4 and Arg5 from the N-terminal arm. Arg5 is highly conserved amongst different HDs and in previous structures it interacted with thymine at position 4. In our structure Arg5 occupies a similar position, with the potential to hydrogen bond to either O2 of thymine or N3 of the guanine of the averaged AT/CG base pair. Tyr4 is unique to the Eve class of HDs (En contains a proline and Antp contains a glycine) and its interaction with an adenine at position 3 may be one of the factors that leads to a different path for the N-terminal arm in Eve. One of the consequences of this movement of the N-terminal arm is that Arg3 is pushed up towards the adenine of bp 2. This interaction contrasts with the En structure, in which Arg3 reaches over to the thymine at position 3, and it differs dramatically from the Antp structure, in which Arg3 reaches all the way down to the phosphate group of the guanine at position 5. Thus, even though Arg3 is highly conserved between different HDs, it can interact with DNA differently depending upon the trajectory of the N-terminal arm. Refined atomic *B* factors for these residues show high relative values, indicating that the polypeptide chain in this region is highly mobile.

Gln50. The amino acid at position 50 in HDs is considered to be the primary residue responsible for the selection of different HD DNA binding sequences (Hanes and Brent, 1989, 1991; Treisman *et al.*, 1989). The residue has been shown to confer specificity for base pairs preceding the ATTA core sequence. For instance, replacement of Lys50 in the Bicoid HD by a glutamine residue results in a change of specificity from GGATTA to CAATTA. Yet despite its demonstrated importance, the structural role of residue 50 has remained somewhat ambiguous: in the En structure Gln50 is seen to form van der Waals contacts with the thymine of an AT base pair at position –2; in the Antp structure Gln50 is close to a cytosine of a GC base pair at position –1 and the thymine of an AT base pair at position 1. From our structure it appears that part of the reason for the ambiguity is the inherent plasticity of Gln50. We observe three different conformations for Gln50 (Figure 7). In HD I, Gln50 forms direct hydrogen bonds with the AT base pair at position 1, donating a hydrogen bond to O4 of thymine and accepting a hydrogen bond from N6 of adenine (Figure 8A). In addition, the electron density suggests an alternate conformation in which the side chain places the Oε atom within hydrogen bonding distance of N6 of the adenine of the AT base pair at position –1. In HD II a third conformation for Gln50 is observed, in which the residue forms a water-mediated bond to O4 of thymine (bp 1) and a possible van der Waals interaction with the base pair at position –2 (Figure 8B). The side chain of Gln50 is fixed in position by a hydrogen bond to Lys46. This conformation of Gln50 strongly resembles the En HD conformer, but no water molecules were modeled in the En structure due to its lower resolution.

Protein–DNA backbone. Contacts to the DNA backbone arise primarily from the N-terminal arm, the loop between helices 1 and 2 of the HD and the recognition helix. Residues Thr6 and Lys57 are close (<3.1 Å) to the DNA backbone and well positioned to make direct hydrogen bonds, while Phe8, Tyr25 and Trp48 are further away (~3.8 Å) and probably involved in van der Waals and/or electrostatic interactions with the backbone. The positively charged residues

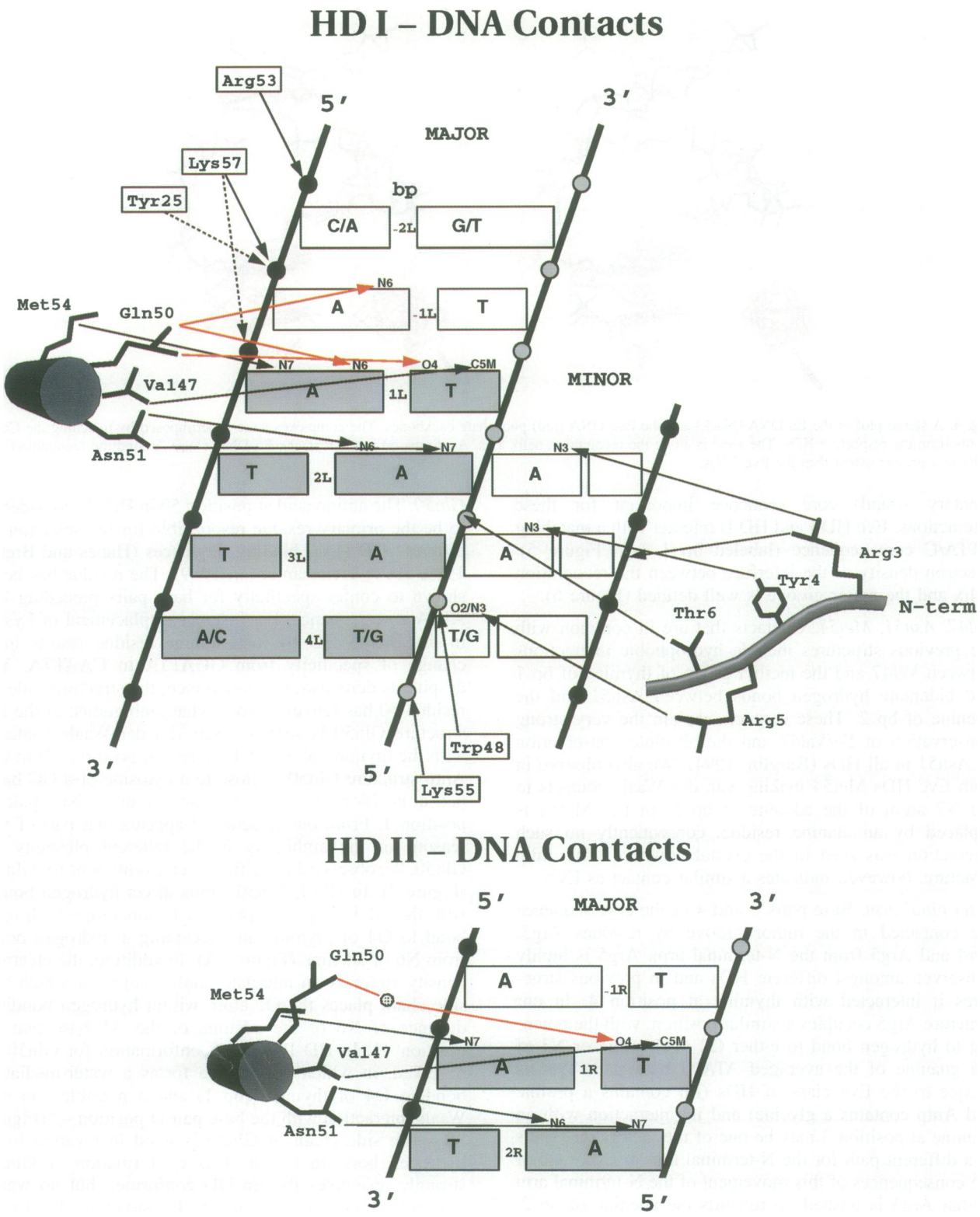

Fig. 5. (**A**) Schematic diagram of the contacts between Eve HD I and DNA (bp –2L to 4L). Contacts occur in both the major and minor grooves and with the phosphate backbone. The ATTA/C core sequence is shown shaded. The DNA atoms labeled are those involved in contacts with the protein. Gln50 contacts are drawn in red. Side chains not depicted are enclosed in boxes. (**B**) Schematic diagram of the contacts between Eve HD II recognition helix and DNA. Contacts in the minor groove are identical to Eve HD I.

Arg28, Arg31, Lys46, Arg53 and Lys55 are at distances ranging from 3.8 to 7 Å from the phosphate groups and presumably contribute to binding through electrostatic interactions. The strength of these electrostatic interactions will depend on both distance and local shape of the protein–DNA interface (Honig and Nicholls, 1995). Most of these residues (especially Arg53) are highly conserved across HD classes, underscoring their functional importance.

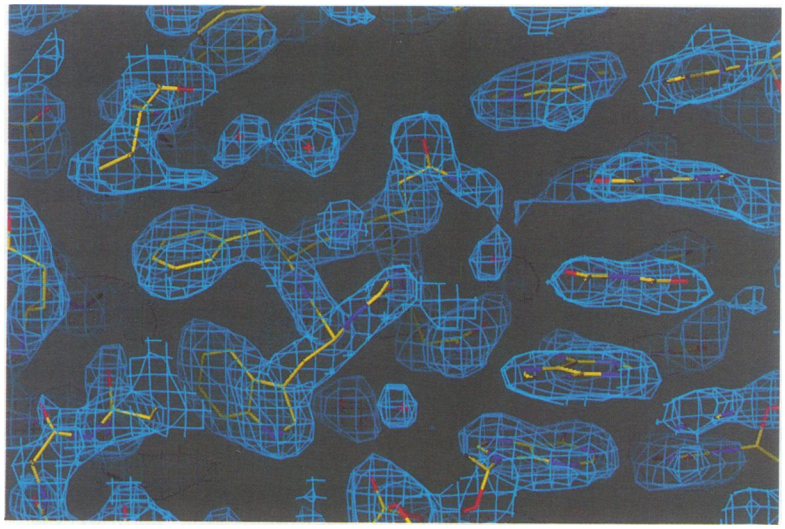

Fig. 6. A portion of the refined electron density from a $2F_0$–F_c map at 2.0 Å, contoured at the 1σ level, showing the protein–DNA interface of HD II. The view is along the recognition helix. Water molecules are shown as red crosses.

Surface area. Burial of solvent-accessible surface area is thought to be a major driving force in macromolecular association (Ha *et al.*, 1989). Calculations of the change in accessible surface area upon binding of the Eve HD to its cognate site (Nicholls *et al.*, 1991) give a value of 1528 Å^2. This value is relatively small compared with surface areas for other DNA binding proteins. The restriction enzymes have changes of ~4000 Å^2 and many of the transcription factors have changes of ~2500 Å^2 (M.Newman and A.Aggarwal, unpublished results). It is surprising that such high affinity binding (HDs bind with K_d of ~10^{-9}–10^{-10} M; Affolter *et al.*, 1990; Ekker *et al.*, 1991; Ades and Sauer, 1994) can be achieved with a relatively small change in accessible surface area, especially in view of the other DNA binding proteins.

Solvent

We observe several solvent molecules at the protein–DNA interface. Similarly, NMR experiments and molecular dynamic simulations (Billeter *et al.*, 1993; Qian *et al.*, 1993) have indicated that there are water molecules at the protein–DNA interface of the Antp complex. The NMR experiments could detect the presence of water in the interface but due to fast on–off rates were unable to locate the positions of them, except for a water molecule buried near the hydrophobic residue Ile47 of Antp (Qian *et al.*, 1993). We also observe molecular water in the same region (in Eve, Ile47 is replaced by Val). Specifically, two water molecules are positioned below Val47 and form a network of hydrogen bonds between the phosphate backbone and the carbonyl oxygen of Thr44, helping to position the recognition helix. Another pair of water molecules forms a network connecting Arg53 and the carbonyl oxygen of Val26. In addition to the water molecule mediating contacts between Gln50 and DNA (described above), we also detect solvent molecules extending from the carbonyl oxygen of residue 50, filling the space between the polypeptide, the phosphate backbone and the bases of the major groove. This water molecule may substitute for the tighter grip of the major groove around the recognition helix in the En and Matα2 structures.

Discussion

Protein conformation

The HD is an extraordinarily simple example of a module that binds to DNA with relatively high affinity. The HD appears to be a modification and simplification of the DNA binding domains of bacterial repressors that contain five or six helices (reviewed in Harrison and Aggarwal, 1990). Others, like λ cro protein, have β-strand elements. The HD, in contrast, is a minimalist version, with only three helices (plus an N-terminal arm) that provides complete and autonomous functionality in the sense that they bind as monomers.

The hydrophobic core of the Eve HD includes residues from all three helices. These residues are highly conserved across the >300 HD sequences now known (Bürglin, 1994). Specifically, Trp48 is invariant, while Leu16 (99%), Phe20 (96%) and Phe49 (99%) are highly conserved. More intriguing than the hydrophobic core is the conservation of residues on the surface of the Eve HD. Eve belongs to a distinct class of HDs, characterized by many conserved residues beyond those conserved across classes. It is remarkable that there are only two amino acid differences between the Eve homolog HDs of flies and humans (covering ~600 million years of evolution). What could be the reason for this conservation? Most of these Eve class residues map to the surface of the HD, and in general are bunched together forming patches (see Figure 9). Three patches lie: (i) in the area of the amino half of helix 1; (ii) in the C-terminal half of helix 1 and the beginning of the loop; (iii) near the end of helix 3. It is possible that these surface patches encode as yet uncharacterized functions that have been conserved over evolution. Such functions may be interactions with cofactors or recognition of DNA sequences other than the AT-rich site. One implication of this hypothesis is that Eve cofactors might also have been conserved across species.

DNA recognition

The manner of AT-rich DNA binding is a case of direct read out. Helix 3 extends across the major groove and

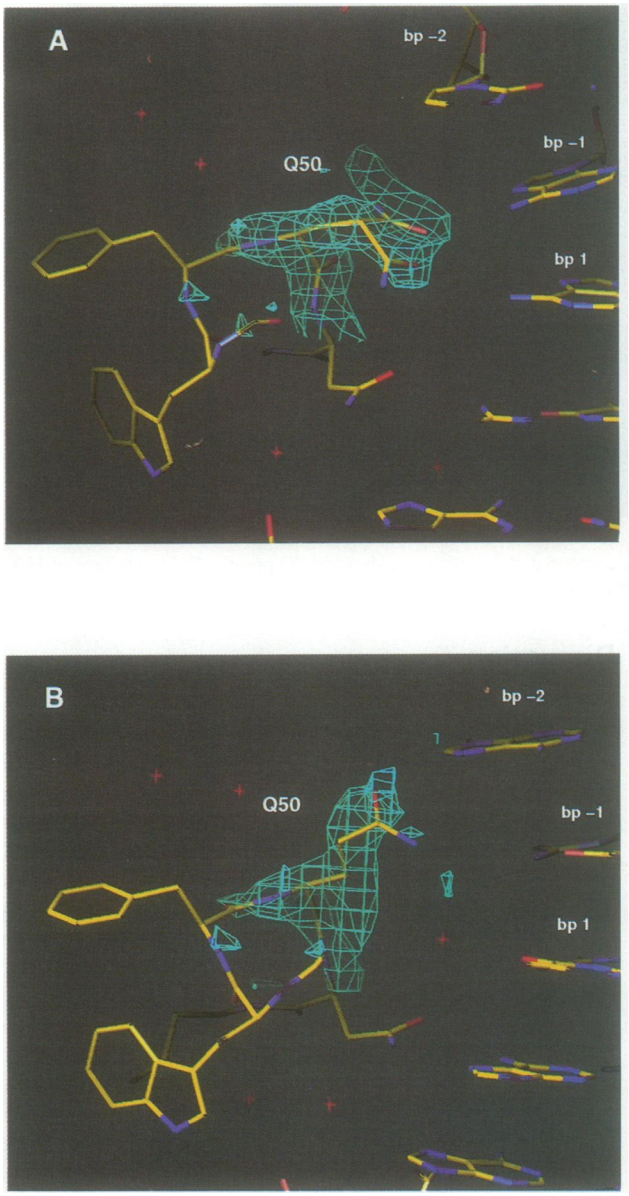

Fig. 7. A $2F_0$–F_c simulated anealing omit map of Gln50 calculated by the standard method of first heating the molecule to 1000 K to remove model bias (Brunger, 1992). The map is contoured at the 1σ level. (**A**) The conformations of Gln50 from HD I. (**B**) The conformation found in HD II.

residues Val47, Asn51 and Met54 interact by way of hydrogen bonds and van der Waals interactions with the functional groups of bp 1 and 2. The functionally important residue Gln50 interacts in multiple ways with base pairs all the way from 1 to –2. A question arises as to why the residue is so flexible. We suggest that the plasticity of Gln50 may be due to its less than ideal distance from DNA base pairs. Even the hydrogen bonds that we observe may be weak, causing the residue to sample other possibilities, including water-mediated interactions. These transient interactions may reduce the overall specificity of Gln50, and explain why an alanine substitution results in only a 2.4-fold loss of binding (Ades and Sauer, 1994). On the other hand, a lysine residue at position 50, with its longer reach, could interact uniquely with GC base pairs at positions –1 and –2 and confer high affinity for the Bicoid sequence GGATTA (Hanes and Brent, 1991; Treisman *et al.*, 1989).

Concomitant with these major groove interactions the N-terminal arm grasps the DNA in the minor groove, contacting bases and sugar moieties of bp 2–4. The trajectory of the N-terminal arm differs from that in the En and Antp structures. This difference may be due to Tyr4 and its interaction with the adenine of bp 3. Tyr4 is unique to the Eve class of HDs (En contains a proline and Antp contains a glycine). Due to variabilty in the N-terminal paths, the highly conserved residue Arg3 interacts with DNA differently in the three structures. This is consistent with studies showing that replacement of N-terminal residues with residues from another HD result in slightly different DNA binding specificities (Ekker *et al.*, 1992; Gehring *et al.*, 1994). Judging from our structure, the identity of residue 4 may be particularly important in determining the path of the N-terminal arm and the consequent DNA binding preferences.

NMR experiments and simulations (Billeter *et al.*, 1993; Qian *et al.*, 1993) have indicated that there are water molecules at the protein–DNA interface of the Antp complex. Our structure locates some of this solvent. We observe at least five well-defined water molecules in the major groove (*B* factors <30 Å^2). They appear to play a dual role, facilitating specific side chain–base read out (Gln50) and non-specific, energetic contributions to binding by filling gaps between the recognition helix and the major groove. Water-mediated recognition of DNA base pairs has been seen in several crystal structures, including the *trp* repressor–operator complex (Luisi and Sigler, 1990) and the *Bam*HI–DNA complex (Newman *et al.*, 1995).

Eve mutants

The structure described here may be used to rationalize some of the mutational data on the Eve protein. Molecular analysis of mutant alleles, isolated through chemical or radiation mutagenesis, uncovered two alleles which had point mutations in the HD (Frasch *et al.*, 1988). The two mutants, IIR59 and ID19, change Thr6 to isoleucine and Arg52 to histidine respectively. IIR59 is a hypomorphic allele, indicating only a partial loss of function, which is correlated with a significant decrease in DNA binding affinity (Hoey, 1989). In the light of the structure one can see why affinity would be affected. Thr6, which is part of the N-terminal arm, makes a hydrogen bond to the phosphate of bp 3 (Figure 5A). Replacing it with isoleucine would remove the hydrogen bond and possibly create steric interference in the minor groove. This substitution is similar but not as drastic as truncation of the N-terminal arm by Gehring and co-workers (Qian *et al.*, 1994). They found that such a HD had a lower affinity for the DNA by about an order of magnitude.

The ID19 mutation is a temperature-sensitive allele. The Arg→His substitution can be seen in the context of our structure to destablize the HD. The aliphatic arm of Arg52 plays a role in delimiting the hydrophobic core. In addition, the residue salt bridges to Glu17 to provide a second type of structural brace. Thus a histidine at this position might disrupt the stabilizing role of Arg52. Interestingly, while Arg52 is fairly well conserved (87.5%; including lysine, 94%), nine HDs contain histidine at this

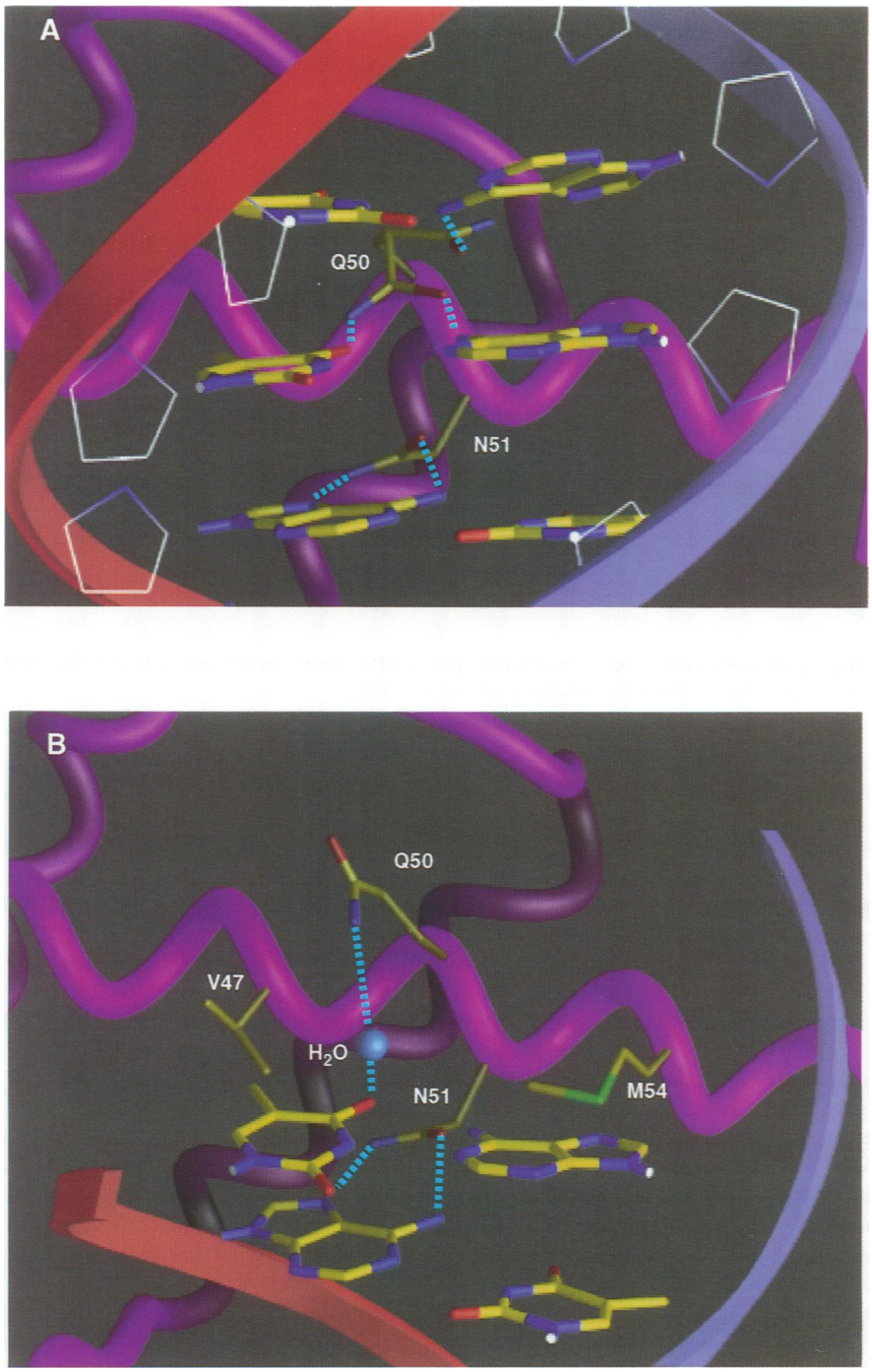

Fig. 8. Interactions in the major groove. (**A**) A view perpendicular to the recognition helix of Eve HD I, showing Asn51 and Gln50 making direct hydrogen bonds (blue dotted lines) with DNA base pairs. Gln50 displays two conformers. In the down position it hydrogen bonds with bp 1. In the lateral position it hydrogen bonds with an adenine of bp −1. Val47 and Met54 are not shown for clarity. (**B**) A view perpendicular to the recognition helix of HD II, showing its interaction with DNA (only bp 1 and 2 are shown). Residues Val47, Gln50, Asn51 and Met54 have been labeled. The water-mediated interaction between the thymine of bp 1 and Gln50 is depicted. The temperature factor of the bridging water molecule is 18 $Å^2$. Drawn with GRASP (Nicholls *et al.*, 1991).

position. These HDs are more divergent, but it is not clear why they would not be as thermally labile as ID19.

Biological specificity

A crucial question with respect to HDs is how they achieve their biological specificity. Homeoproteins have overlapping DNA binding specificities and yet in cells, tissues or parasegments which express several homeoproteins they display remarkable biological specificity. Several hypotheses have been forwarded. One proposal is that relatively small differences in DNA binding may account for the functional specificity *in vivo* (Dessain *et al.*, 1992; Ekker *et al.*, 1992). These studies have shown that very closely related HDs, from the same class or even family, have differing affinities for different sequences. In this regard, the Eve structure provides new details of how such binding differences can arise, especially due to residues 50 and 54 of the recognition helix and the path

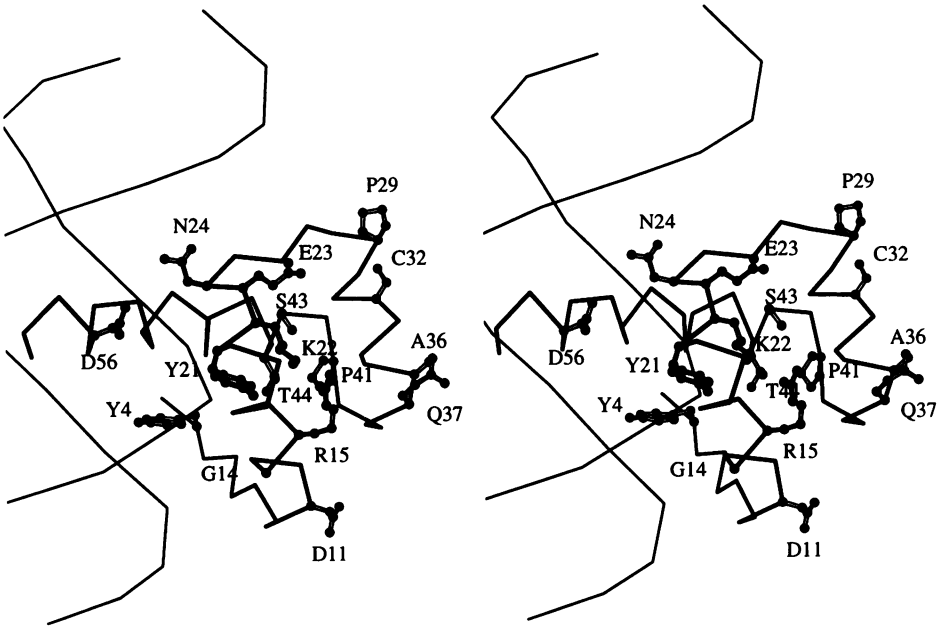

Fig. 9. Stereo Cα plot of the Eve HD with residues unique to the Eve class of HDs depicted in ball and stick. Only the residues which are non-conservative substitutions of the consensus HD sequence (Bürglin, 1994) are shown.

of the N-terminal arm in the minor groove, which may be altered by the identity of residue 4.

A second mechanism postulates the existence of cofactors (possibly also containing HDs) that bind to adjacent DNA sites to accentuate or modulate specificity. A well-studied example of this model is the yeast Matα2/Mata1/MCM system (Goutte and Johnson, 1994; Phillips *et al.*, 1994). In that case, the individual proteins alone have relatively low specificity and affinity, but when they bind in a heterodimeric fashion, specificity and affinity are increased considerably. Our crystallographic observation that two HDs can potentially bind alongside on a DNA site as short as 10 bp without any steric clashes provides a structural rationale for such a mechanism. For instance, the homeoproteins Mec-3 and Unc-86 from *Caenorhabditis elegans* cooperatively bind to promoter sequences of the *mec-3* gene (Xue *et al.*, 1993). More recently it has been shown in *Drosophila* that the homeoproteins Ubx and Exd interact cooperatively and their DNA binding sites overlap over a stretch of just 12 bp (Chan *et al.*, 1994). Based on our model, both HD proteins could be easily accommodated on opposite faces of DNA, with regions within and immediately outside the HD favorably disposed for protein–protein interactions. As seen in Figure 2, the N-terminal arm and polypeptide that continues from it are well positioned to interact with the end of helix 1, the loop between helices 1 and 2 and the end of helix 3 of the adjacent HD. Strikingly, this area is exactly one of the patches that is known to be critical in the Ubx–Exd interaction (residues 23, 25 and 57; Chan *et al.*, 1994). In addition, examination of an Eve HD from the adjacent unit cell, i.e. translated along the DNA axis by one turn and on the same face, shows that the beginning of helix 1 (Arg10) is within 4 Å of the loop of the adjacent protein. Therefore, it is tempting to speculate that these surfaces that seem so well disposed for protein–protein interactions are indeed utilized in such a fashion.

Our structure also suggests that the monomeric nature of HDs, combined with their ability to bind tandemly or inversely on the opposite or the same face of DNA, could be a mechanism to generate different configurational permutations with the same HD. Some permutations may potentiate transcription, others may attenuate it. Multiple binding sites for homeoproteins have been found in several promoters. Beachy and co-workers have found that a dozen or more Ubx molecules can bind to sites in the Ubx promoter (Beachy *et al.*, 1993). Moreover, both Ubx and Eve proteins produce DNase I footprints of 40 and 80 bp over this region (Biggin and Tjian, 1989; Johnson and Krasnow, 1992; TenHarmsel *et al.*, 1993). The sequence of this region is rich in repeats of TAA or TAATCG, but neither the configuration nor the exact stoichiometry of protein binding is known. Experiments with the Antp promoter (Appel and Sakonju, 1993) show sites where ATTA cores are as closely spaced as in our crystal structure. Finally, recent *in vivo* experiments with Bicoid homeoprotein revealed not only a requirement for multiple sites to acheive activation, but also specific configurations, such as widely spaced sites positioned on opposite faces of the DNA (2.5 turns of DNA; Hanes *et al.*, 1994). This configurational sensitivity is very much in line with a model that envisions configurational variability as conferring specificity. Such a mechanism operates not by increasing DNA binding specificity, but rather by modifying and varying the HD protein's interactions with the rest of the transcriptional machinery.

Materials and methods

Crystals

The purification, crystallization and preliminary diffraction analysis of the Eve HD–DNA complex have been described previously (Hirsch and Aggarwal, 1995). The crystals belong to space group $P2_1$ with unit cell dimensions a = 34.06, b = 61.61, c =39.99 Å, β = 90.0°. The crystals

Table I. Crystallographic data

	Native (PF[a])	Native (BNL[a])	Iodo1[b]	Iodo2[b]
Number of crystals	1	2	5	2
d_{min} (Å)	3.0	2.0	3.0	3.0
Measured reflections	5801	20 689	6863	4259
Unique reflections	2865	8613	3085	2584
Completeness (%)[c]	85.5 (64.8)	77.1 (52.4)	92.7 (90.2)	76.8 (78.7)
R_{merge}[c,d]	0.104 (0.29)	0.076 (0.19)	0.087 (0.10)	0.074 (0.14)
R_{iso}[e]			0.10	0.11

[a]PF, Photon Factory; BNL, Brookhaven National Laboratory.
[b]See Figure 1 for the positions of iodine substitutions.
[c]Figures in parentheses are for the highest resolution shell.
[d]$R_{merge} = \Sigma|I_{obs} - <I>|/\Sigma<I>$.
[e]$R_{iso} = \Sigma|I_{ph} - I_p|/\Sigma I_p$, where I_{ph} is Iodo1 or Iodo2 and I_p is native (BNL).

are pseudo-orthorhombic and could be (and were initially) assigned to space group $P2_12_12$ (see below). The space group and unit cell parameters were assigned from precession photographs which showed *mmm* symmetry and systematic absences along the *a* and *b* axes. There was an ambiguity regarding the length of the *a* axis; whether it should be 34.06 (10 bp) or 68.12 Å (20 bp), since the oligonucleotide used for crystallization was a 20mer. Layer lines along the *a* axis gave no hint of a 68.12 Å spacing, despite the fact that the synthetic oligomer lacked a 5′-phosphate group, thus making the two direct repeats not chemically identical. The fact that we do not observe any measurable intensities at 68.12 Å spacing along *a*, even in the 2 Å measured data (see below), suggests that the missing phosphate group between one unit cell and the next has little consequential effect on the binding of HDs. This is not surprising, since the missing phosphate is well away from DNA recognition residues such as Gln50 and Asn51 (12–18 Å).

Data collection

Data were measured using synchrotron radiation at the Photon Factory (PF) (beamline BL6A2; Sakabe, 1991) and the Brookhaven National Laboratory (BNL) (beamline X4A). PF data were measured first, from a single crystal (10°C) at 3 Å resolution (Table I). The data were recorded on phosphorimaging plates (Fuji) using the Weissenberg geometry (10° oscillations) and processed by the WEIS software package (Higashi, 1989). The BNL data were measured from two crystals (4°C), using phosphorimaging plates and a standard oscillation geometry. The data extended to 2 Å resolution (Table I). These and all subsequent data sets were processed using DENZO (Otwinowski, 1993) and merged using ROTAVATA/AGROVATA (Collaborative Computational Project, 1994). Heavy atom derivatives were prepared by substituting iodouracil for thymidine residues on the DNA. These crystals proved to be highly radiation sensitive, in proportion to the number of iodines incorporated. Data sets were measured at BNL for two derivatives, iodo1 and iodo2, incorporating two iodines in the 20 bp DNA duplex (Table I).

Structure determination

Since the PF data became available first, it was used to start the structure determination by molecular replacement (MR) in space group $P2_12_12$. Initially we expected that one HD would bind to one 10 bp consensus sequence. Using a model from the En HD–DNA complex (one HD bound to 10 bp) we calculated a cross rotation function followed by Patterson correlation refinement with X-PLOR (Brunger, 1992). The Patterson correlation refinement clearly singled out one orientation. Translation function searches (using the programs X-PLOR and TFFC; Collaborative Computational Project, 1994) gave a clear top solution, but the molecular packing showed that the DNA symmetry mates were overlaid on each other. Alternate translation function solutions were pursued, but without success. We re-evaluated our search model and reasoned that the top solution could essentially be correct if there were two HDs bound to one 10 bp DNA consensus sequence, related by a screw symmetry along the DNA axis. The screw symmetry could be treated as either crystallographic (space group $P2_12_12$) or non-crystallographic (space group $P2_1$). The binding of two HDs was consistent with the 10 bp consensus sequence, which on closer inspection revealed an imperfect direct repeat (AATTA and AATTC) within it (Figure 1). Because the direct repeat was not exact we pursued the structure analysis in space group $P2_1$, in case the two HDs bound differently to their 5 bp subsites.

The new model gave a top MR solution (R = 48.6% for data from

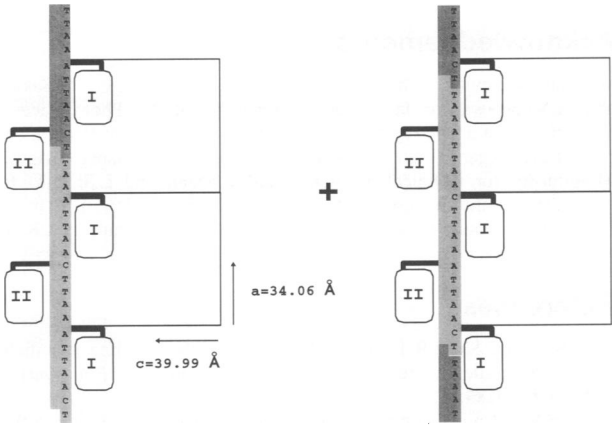

Fig. 10. A packing diagram of the crystal showing the translational mix of the DNA. Two unit cells are depicted with *a* and *c* axes labeled and the *b* axis coming out of the plane of the paper. HDs I and II are labeled. The DNA is the 20 bp oligonucleotide used in co-crystallization (c.f. Figure 1), shown spanning two unit cells and depicted as a cylinder.

15 to 4 Å) without any steric clashes. Rigid body refinement, breaking the model into increasingly smaller units, resulted in an R factor of 40.2%. Electron density maps were inspected and found to be of good quality. At this stage high resolution BNL data from native and iodinated crystals became available. Difference Fourier syntheses utilizing model phases and isomorphous differences clearly showed iodine peaks (>6 σ) at the appropriate positions; establishing the validity of our model. However, instead of one peak for each derivative, as expected, two peaks were found, separated by 5 bp on the same DNA strand (ratios 1:1.1 for iodo1 and 1:1.3 for iodo2). We conclude from this that there is translational mixing of the DNA in this crystal form, similar to the rotational mixing of DNA found in other protein–DNA (Wolberger *et al.*, 1991; Ferre-D'Amare *et al.*, 1993; Schwabe *et al.*, 1995) and DNA crystals (DiGabriele *et al.*, 1989). Thus the two subsites (AATTA and AATTC) comprising the 10 bp consensus sequence are averaged in our structure. The net effect is that the DNA subsite 'under' both HDs is essentially the same in which the last base pair is an average of AT and CG base pairs. A packing model illustrating this is shown in Figure 10. It might be expected that the two HDs would also be averaged, but they appear to maintain distinct conformations, as evidenced by distinct electron densities for residue Gln50 (Figure 7).

The structure determination was continued by combining iodine phases with model phases to compute combined maps for consultation in model building. The model derived from the PF data was subjected to a round of simulated annealing against the BNL data set, which brought the R factor down to 28.6%. The model was rebuilt using O (Jones *et al.*, 1991) and FRODO (Jones, 1985) and refined, this time along with temperature factors. Using $2F_0–F_c$ and $F_0–F_c$ maps, 68 water molecules were added and the model refined using conjugate Powell methods in X-PLOR. The model was checked continually by omit maps calculated by deleting portions of the model and doing 30 cycles of positional refinement to minimize bias. In addition, simulated annealing omit maps

were calculated and used to confirm the model in certain portions. A striking feature to emerge from these omit maps was the different conformations of residue Gln50 in the two HDs. From this we conclude that, unlike the DNA, the two HDs remain unaveraged (Figure 7). The asymmetry of Gln50 could be due to subtle differences in the electrostatics of crystal packing environments, but it is not obvious from the structure. It is not uncommon to observe asymmetric conformations for flexible residues in crystal structures. For example, in the structure of the NF-κB homodimer bound to a palindromic DNA site (Ghosh *et al.*, 1995), residues Lys241, Lys272 and Arg305 adopt asymmetric conformations despite perfect two-fold symmetry of the DNA site. The present *R* factor stands at 22.9% and R_{free} = 31.6% with good stereochemistry, as determined by the program PROCHECK (Laskowski *et al.*, 1993). [The *R* factors were calculated for data within 8–2.0 Å resolution and $F_0 > 2\sigma(F_0)$.] The r.m.s. deviations from ideality for bond lengths and bond angles are 0.016 Å and 2.56° respectively. The averages of the final refined B-factors for HDI, HDII and the DNA are 26.9, 27.9 and 25.6 Å^2, respectively.

Acknowledgements

We thank M.Levine for help in the early stages of this project, N.Sakabe and A.Nakagawa for facilitating experiments at the Photo Factory, W.Hendrickson and C.Ogata for facilitating experiments at BNL, members of the Aggarwal laboratory for data collection assistance, Grigoriy Mogilnitsky for technical assistance and R.Mann and L.Shapiro for comments on the manuscript. Coordinates are being deposited in the Brookhaven Protein Data Bank. Supported by an NIH grant to A.K.A.

References

Ades,S.E. and Sauer,R.T. (1994) Differential DNA-binding specificity of the Engrailed homeodomain: the role of residue 50. *Biochemistry*, **33**, 9187–9184.

Affolter,M., Percival-Smith,A., Muller,M., Leupin,W. and Gehring,W.J. (1990) DNA binding properties of the purified Antennapedia homeodomain. *Proc. Natl Acad. Sci. USA*, **87**, 4093–4097.

Appel,B. and Sakonju,S. (1993) Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in *Drosophila* embryos. *EMBO J.*, **12**, 1099–1109.

Bastian,H. and Gruss,P. (1990) A murine *even-skipped* homologue, *Evx 1*, is expressed during early embryogenesis and neurogenesis in a biphasic manner. *EMBO J.*, **9**, 1839–1852.

Beachy,P.A. *et al.* (1993) Cooperative binding of an *Ultrabithorax* homeodomain protein to nearby and distant DNA sites. *Mol. Cell. Biol.*, **13**, 6941–6956.

Biggin,M.D. and Tjian,R. (1989) A purified *Drosophila* homeodomain protein represses transcription *in vitro*. *Cell*, **58**, 433–440.

Billeter,M. *et al.* (1993) Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain–DNA complex. *J. Mol. Biol.*, **234**, 1084–1093.

Brunger,A.T. (1992) *X-PLOR, Version 3.1, a System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.

Bürglin,T. (1994) A comprehensive classification of homeobox genes. In Duboule,D. (ed.), *Guidebook to the Homeobox Genes*. Oxford University Press, Oxford, UK, pp. 27–71.

Chan,S.K., Jaffe,L., Capovilla,M., Botas,J. and Mann,R.S. (1994) The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell*, **78**, 603–615.

Clarke,N.D., Kissinger,C.R., Desjarlais,J., Gilliland,G.L. and Pabo,C.O. (1994) Structural studies of the engrailed homeodomain. *Protein Sci.*, **3**, 1779–1787.

Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

D'Esposito,M. *et al.* (1991) EVX2, a human homeobox gene homologous to the *even-skipped* segmentation gene, is localized at the 5′ end of HOX4 locus on chromosome 2. *Genomics*, **10**, 43–50.

Desplan,C., Theis,J. and O'Farrell,P.H. (1988) The sequence specificity of homeodomain–DNA interaction. *Cell*, **54**, 1081–1090.

Dessain,S., Gross,C.T., Kuziora,M.A. and McGinnis,W. (1992) Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. *EMBO J.*, **11**, 991–1002.

DiGabriele,A.D., Sanderson,M.R. and Steitz,T.A. (1989) Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natl Acad. Sci. USA*, **86**, 1816–1820.

Doe,C.Q., Smouse,D. and Goodman,C.S. (1988) Control of neuronal fate by the *Drosophila* segmentation gene *even-skipped*. *Nature*, **333**, 376–378.

Ekker,S.C., Young,K.E., von Kessler,D.P. and Beachy,P.A. (1991) Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. *EMBO J.*, **10**, 1179–1186.

Ekker,S.C., von Kessler,D.P. and Beachy,P.A. (1992) Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J.*, **11**, 4059–4072.

Faiella,A. *et al.* (1991) Isolation and mapping of *EVX1*, a human homeobox gene homologous to *even-skipped*, localized at the 5′ end of HOX1 locus on chromosome 7. *Nucleic Acids Res.*, **19**, 6541–6545.

Ferre-D'Amare,A.R., Prendergast,G.C., Ziff,E.B. and Burley,S.K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature*, **363**, 38–45.

Florence,B., Handrow,R. and Laughon,A. (1991) DNA-binding specificity of the fushi tarazu homeodomain. *Mol. Cell. Biol.*, **11**, 3613–3623.

Frasch,M., Warrior,R., Tugwood,J. and Levine,M. (1988) Molecular analysis of *even-skipped* mutants in *Drosophila* development. *Genes Dev.*, **2**, 1824–1838.

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

Ghosh,G., Van Duyne,G., Ghosh,S. and Sigler,P.B. (1995) Structure of NF-κB homodimer bound to a κB site. *Nature*, **373**, 303–310.

Goutte,C. and Johnson,A.D. (1994) Recognition of a DNA operator by a dimer composed of two different homeodomain proteins. *EMBO J.*, **13**, 1434–1442.

Ha,J.H., Spolar,R.S. and Record,M.T. (1989) Role of the hydrophobic effect in stability of site-specific protein–DNA complexes. *J. Mol. Biol.*, **209**, 801–816.

Han,K. and Manley,J.L. (1993) Transcriptional repression by the *Drosophila* even-skipped protein: definition of minimal repression domain. *Genes Dev.*, **7**, 491–503.

Hanes,S.D. and Brent,R. (1989) DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell*, **57**, 1275–1283.

Hanes,S.D. and Brent,R. (1991) A genetic model for interaction of the homeodomain recognition helix with DNA. *Science*, **251**, 426–430.

Hanes,S.D., Riddihough,G., Ish-Horowicz,D. and Brent,R. (1994) Specific DNA recognition and intersite spacing are critical for action of the Bicoid morphogen. *Mol. Cell. Biol.*, **14**, 3364–3375.

Harrison,S.C. and Aggarwal,A.K. (1990) DNA recognition by proteins with the helix–turn–helix motif. *Annu. Rev. Biochem.*, **59**, 933–969.

Higashi,T. (1989) The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J. Appl. Crystallogr.*, **22**, 9–18.

Hoey,T.C. (1989) *DNA Binding Properties of Drosophila Homeo Box Proteins*. Ph.D. thesis. Columbia University, New York.

Hirsch,J.A. and Aggarwal,A.K. (1995) Purification, crystallization, and preliminary X-ray diffraction analysis of Even-skipped homeodomain complexed to DNA. *Proteins*, **21**, 268–271.

Hoey,T. and Levine,M. (1988) Divergent homeo box proteins recognize similar DNA sequences in *Drosophila*. *Nature*, **332**, 858–861.

Hoey,T., Warrior,R., Manak,J. and Levine,M. (1988) DNA-binding activities of the *Drosophila melanogaster* even-skipped protein are mediated by its homeo domain and influenced by protein context. *Mol. Cell. Biol.*, **8**, 4598–4607.

Honig,B. and Nicholls,A. (1995) Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.

Jiang,J., Hoey,T. and Levine,M. (1991) Autoregulation of a segmentation gene in *Drosophila*: combinatorial interaction of the *even-skipped* homeo box protein with a distal enhancer element. *Genes Dev.*, **5**, 265–277.

Johnson,F.B. and Krasnow,M.A. (1992) Differential regulation of transcription preinitiation complex assembly by activator and repressor homeo domain proteins. *Genes Dev.*, **6**, 2177–2189.

Joly,J.S., Joly,C., Schulte-Merker,S., Boulekbache,H. and Condamine,H. (1993) The ventral and posterior expression of the zebrafish homeobox gene *eve1* is perturbed in dorsalized and mutant embryos. *Development*, **119**, 1261–1275.

Jones,T.A. (1985) Interactive computer graphics: FRODO. *Methods Enzymol.*, **115**, 157–171.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved

methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

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 Å resolution: a framework for understanding homeodomain–DNA interactions. *Cell*, **63**, 579–590.

Klemm,J.D., Rould,M.A., Aurora,R., Herr,W. and 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.

Kraulis,P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lavery,R. and Sklenar,H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, **6**, 63–91.

Luisi,B.F. and Sigler,P.B. (1990) The stereochemistry and biochemistry of the trp repressor–operator complex. *Biochim. Biophys. Acta*, **1048**, 113–126.

Manoukian,A.S. and Krause,H.M. (1992) Concentration-dependent activities of the *even-skipped* protein in *Drosophila* embryos. *Genes Dev.*, **6**, 1740–1751.

Merritt,E.A. and Murphy,M.E.P. (1994) Raster3D Version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr.*, **D50**, 869–873.

Miles,A. and Miller,D.J. (1992) Genomes of diploblastic organisms contain homeoboxes: sequence of eveC, an even-skipped homologue from the cnidarian *Acropora formosa*. *Proc. R. Soc. Lond. Sect. B*, **248**, 159–61.

Newman,M., Strzelecka,T., Dorner,L., Schildkraut,I. and Aggarwal,A.K. (1995) Structure of *Bam*HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656–663.

Restriction endonuclease *Bam*HI partially folds and unfolds on DNA binding: crystal structure of *Bam*HI–DNA complex at 2.2 A resolution. *Science*, in press.

Nicholls,A., Sharp,K. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.

Nusslein-Volhard,C. and Wieschaus,E. (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature*, **287**, 795–801.

Otting,G. *et al.* (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.

Otwinowski,Z. (1993) Oscillation data reduction program. In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *Data Collection and Processing*. SERC Daresbury Laboratory, Daresbury, UK, pp. 56–62.

Patel,N.H., Ball,E.E. and Goodman,C.S. (1992) Changing role of *even-skipped* during the evolution of insect pattern. *Nature*, **357**, 339–342.

Phillips,C.L., Stark,M.R., Johnson,A.D. and Dahlquist,F.W. (1994) Heterodimerization of the yeast homeodomain transcriptional regulators α2 and a1 induces an interfacial helix in α2. *Biochemistry*, **33**, 9294–9302.

Qian,Y.Q. *et al.* (1989) The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573–580.

Qian,Y.Q., Otting,G. and Wüthrich,K. (1993) NMR detection of hydration water in the intermolecular interface of a protein–DNA complex. *J. Am. Chem. Soc.*, **115**, 1189–1190.

Qian,Y.Q., Resendez-Perez,D., Gehring,W.J. and Wüthrich,K. (1994) The des(1–6)Antennapedia homeodomain: comparison of the NMR solution structure and the DNA-binding affinity with the intact Antennapedia homeodomain. *Proc. Natl Acad. Sci. USA*, **91**, 4091–4095.

Sakabe,N. (1991) X-ray diffraction collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nucl. Instrum. Methods Phys. Res.*, **A303**, 448–463.

Schwabe,J.W.R., Chapman,L. and Rhodes,D. (1995) The oestrogen receptor recognises an imperfectly palindromic response element through an alternative side-chain conformation. *Structure*, **3**, 201–213.

Scott,M.P., Tamkun,J.W. and Hartzell,G.W. (1989) The structure and function of the homeodomain. *Biochim. Biophys. Acta*, **989**, 25–48.

TenHarmsel,A., Austin,R.J., Savenelli,N. and Biggin,M.D. (1993) Cooperative binding at a distance by even-skipped protein correlates with repression and suggests a mechanism of silencing. *Mol. Cell. Biol.*, **13**, 2742–2752.

Treisman,J., Gonczy,P., Vashishtha,M., Harris,E. and Desplan,C. (1989) A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell*, **59**, 553–562.

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.

Xue,D., Tu,Y. and Chalfie,M. (1993) Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3. *Science*, **261**, 1324–1328.

Received on July 31, 1995; revised on September 15, 1995