Cell, Vol. 63, 579–590, November 2, 1990, Copyright © 1990 by Cell Press

Crystal Structure of an engrailed Homeodomain–DNA Complex at 2.8 Å Resolution: A Framework for Understanding Homeodomain–DNA Interactions

Charles R. Kissinger,* Beishan Liu,* Enrique Martin-Blanco,† Thomas B. Kornberg,† and Carl O. Pabo*

* Department of Molecular Biology and Genetics and Howard Hughes Medical Institute
Johns Hopkins University School of Medicine
Baltimore, Maryland 21205

† Department of Biochemistry and Biophysics
University of California
San Francisco, California 94143

Summary

The crystal structure of a complex containing the engrailed homeodomain and a duplex DNA site has been determined at 2.8 Å resolution and refined to a crystallographic R factor of 24.4%. In this complex, two separate regions of the 61 amino acid polypeptide contact a TAAT subsite. An N-terminal arm fits into the minor groove, and the side chains of Arg-3 and Arg-5 make contacts near the 5′ end of this "core consensus" binding site. An α helix fits into the major groove, and the side chains of Ile-47 and Asn-51 contact base pairs near the 3′ end of the TAAT site. This "recognition helix" is part of a structurally conserved helix-turn-helix unit, but these helices are longer than the corresponding helices in the λ repressor, and the relationship between the helix-turn-helix unit and the DNA is significantly different.

Introduction

The homeodomain is a DNA binding motif that plays a central role in eukaryotic gene regulation (Scott et al., 1989). It was first discovered in a set of Drosophila proteins that regulate development, but it is now clear that the homeodomain occurs in a large family of proteins that regulate transcription in many higher organisms. Comparison of homeodomains from different genes and different organisms shows that their amino acid sequences are highly conserved (Scott et al., 1989). Although careful sequence comparisons allow the homeodomains to be grouped into subfamilies, it seems likely that all the homeodomains will have similar three-dimensional structures and use generally similar modes of DNA recognition. The structure of the Antennapedia homeodomain, a prototypical member of one of the largest sequence subfamilies, was recently determined by two-dimensional nuclear magnetic resonance (Qian et al., 1989). As anticipated from sequence comparisons (Laughon and Scott, 1984; Shepard et al., 1984), this homeodomain contains a helix-turn-helix (HTH) motif very similar to the HTH motif present in a number of prokaryotic repressors (Pabo and Sauer, 1984). Although high resolution crystal structures have been reported for several repressor–operator complexes (Otwinowski et al., 1988; Jordan and Pabo, 1988; Aggarwal et al., 1988), there have not been any structural data available about homeodomain–DNA complexes. Intensive genetic and biochemical studies have elucidated some features of the protein–DNA interactions, but many puzzling questions remain about the specificity of homeodomain–DNA interactions, about the role of conserved residues in complex formation, and about the overall contribution of the homeodomain to site-specific binding and gene regulation.

We recently crystallized a 61 amino acid peptide that contains the homeodomain from the engrailed protein of Drosophila (a prototypic member of another major homeodomain subfamily) and also grew cocrystals with a 21 bp duplex DNA site (Liu et al., 1990; Figure 1). Using a strategy developed when cocrystallizing the λ repressor–operator complex (Jordan et al., 1985), we had tested a series of different DNA fragments with the engrailed homeodomain. It is not yet known which binding site(s) is functional in vivo, but excellent crystals were obtained with the DNA fragment shown in Figure 1, and gel mobility shift experiments confirmed that the engrailed homeodomain binds tightly to this site (K_D = 1–2 × 10^{-9} M in a buffer containing 100 mM KCl and 25 mM HEPES at pH 7.6). In this paper, we report the structure of this homeodomain–DNA complex and discuss the implications for our understanding of protein–DNA recognition and gene regulation.

Results and Discussion

Overall Arrangement of the Homeodomain–DNA Complex

In the crystal, two copies of the homeodomain bind to the 21 bp duplex. One binds near the center of the DNA fragment, and the other binds near the end (Figure 2). Superimposing the refined structures revealed that the conformations of the two protein monomers and the contacts they make with the DNA are nearly identical.

The engrailed homeodomain contains three α helices and an extended N-terminal arm. The structure of this motif provides the basis for understanding homeodomain–DNA interactions (Figure 3). Helix 1 (residues 10–22) and helix 2 (residues 28–37) pack against each other in an antiparallel arrangement. In the complex, each of these helices spans the major groove and is roughly perpendicular to the local direction of the DNA backbone. However, both of these helices are too far from the DNA to make many contacts. Helix 3 (residues 42–58) is roughly perpendicular to the first two helices. The exposed, hydrophilic face of helix 3 fits directly into the major groove, and side chains on this helix make extensive contacts with the DNA. The hydrophobic face of helix 3 packs against helices 1 and 2 to form the interior of the protein. In the engrailed cocrystal structure, we see no evidence for the kink (Qian et al., 1989) that has caused Wüthrich and his colleagues to describe residues 53–59 of the isolated Antennapedia homeodomain as a distinct helical segment

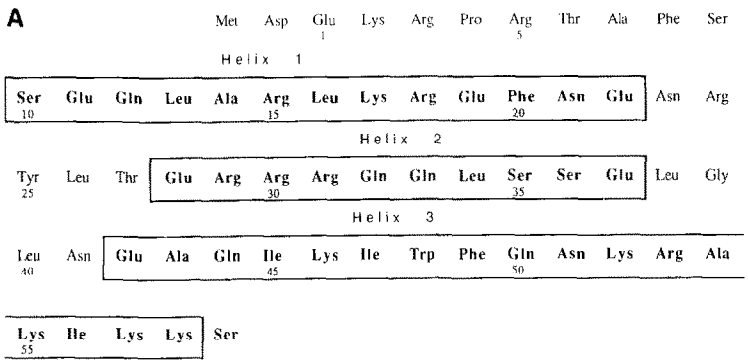

Figure 1. Sequences of engrailed Homeodomain and Binding Site Used for Cocrystallization

(A) Sequence of the engrailed homeodomain. The fragment used for cocrystallization includes 60 amino acids from the Drosophila engrailed protein (Kornberg, 1981; Poole et al., 1985), and the cloning procedure adds a methionine at the N-terminal end of this peptide. Boxes mark the three α helices observed in the cocrystal. The numbering scheme corresponds to that used by Wüthrich and his colleagues in describing the structure of the Antennapedia homeodomain (Qian et al., 1989).

(B) DNA sequence used for cocrystallization. In the crystal, the overhanging 5′ ends pair with those of neighboring duplexes to form a pseudocontinuous double helix. As discussed in the text, two copies of the homeodomain bind to this site. One protein makes its primary contacts near the TAAT site that includes base pairs 11–14 (shown in boldface). The other homeodomain binds at the end of the DNA (where neighboring duplexes overlap in the crystal). This homeodomain makes critical contacts in a region that includes a TAAA site on the lower strand (i.e., base pairs 1–3 and 21, shown in outline).

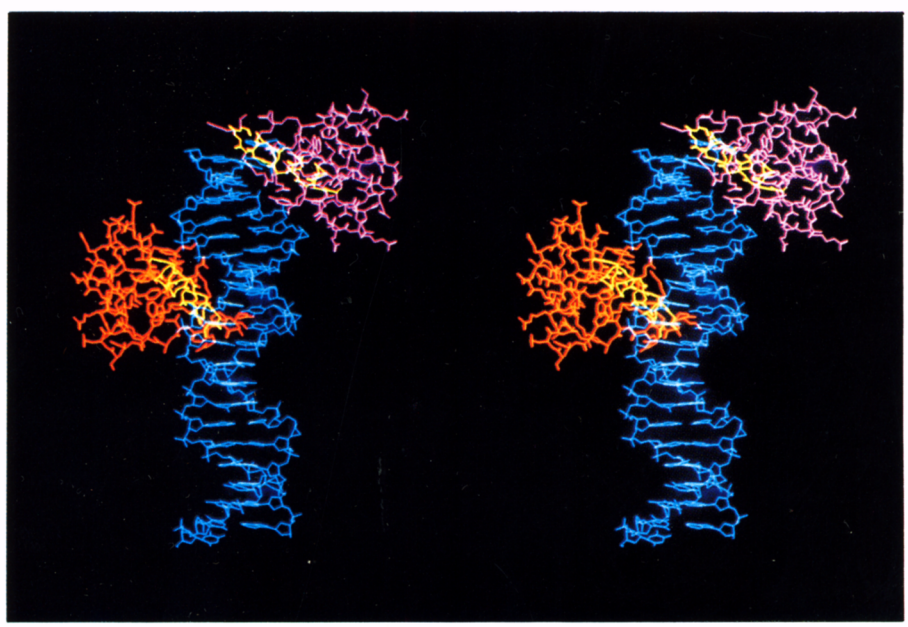

Figure 2. Two Homeodomains Bind to the 21 bp Duplex

One copy of the homeodomain (shown in orange) binds tightly to a site near the center of the DNA duplex. This is a higher affinity binding site, and the paper focuses on this complex. Another copy of the homeodomain (shown in purple) binds to a weaker site at the end of the duplex, and this homeodomain makes additional contacts with a contiguous duplex in the crystal. The overall arrangements of the two proteins with respect to the DNA are very similar, and nearly all of the protein–DNA contacts are identical. Helix 3 of each protein (highlighted in yellow) fits directly into the major groove of the double-helical DNA (shown in blue).

(helix IV). It is possible that this region changes conformation upon DNA binding, and crystallographic refinement of the isolated engrailed protein is now in progress (N. Clarke, C. R. K., B. L., and C. O. P., unpublished data).

The first few residues of the homeodomain appear to be disordered in the crystal, but residues 3–9 form an extended N-terminal "arm" that fits into the minor groove and supplements the contacts made by helix 3. Helices 1 and 2 are connected by a relatively open loop (residues 23–27), and helices 2 and 3 are connected by a somewhat shorter "turn" (residues 38–41). The overall arrangement of the homeodomain reveals a simple, functional design for DNA recognition.

There are no contacts between the two homeodomains seen in the crystal, and they appear to bind as independent monomers. Binding experiments using duplexes with the isolated subsites (see Experimental Procedures) indicate that the engrailed homeodomain binds to the central sequence with a K_D of 1–2 × 10^{-9} M and binds to the terminal sequence with a K_D of 10^{-7} M. (Control experiments with a λ operator site show no binding; any interaction must be at least an order of magnitude weaker.) The

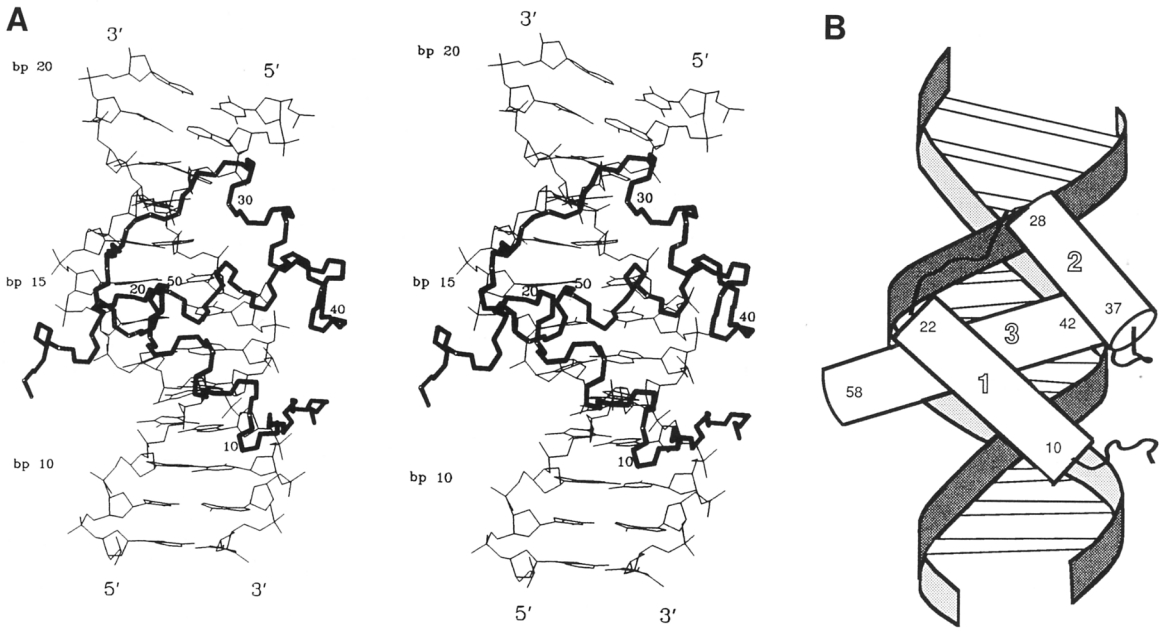

Figure 3. Overview of the Homeodomain–DNA Complex

(A) Stereo diagram showing how the α helices and the N-terminal arm are arranged in the homeodomain–DNA complex. To make it easier to see the overall relationship, this diagram shows only backbone atoms for the protein (N, C_α, and C). Every tenth protein residue is numbered. The DNA shown here includes base pair 8 (bottom) through base pair 20 (top). Base pairs 10, 15, and 20 are labeled, along with the 3′ and 5′ termini of each DNA strand.

(B) Sketch summarizing the relationship of the α helices and the N-terminal arm with respect to the double-helical DNA. Cylinders are used to show the position of the α helices, ribbons are used to show the sugar-phosphate backbone of the DNA, and bars indicate base pairs.

two complexes seen in the crystal happen to be related by an approximate two-fold symmetry axis, but this does not appear to have any real significance for the engrailed–DNA interactions. Although the N-terminal arms have slightly different conformations, most of the contacts are the same in the two complexes. Our discussion will focus on the central site, since engrailed binds to this site with a higher affinity. Binding to the terminal sequence is inherently weaker, and interactions may also be perturbed because the protein is binding at the end of the duplex. (This homeodomain actually contacts two DNA duplexes, since the DNA fragments have overlapping 5′ ends and stack to form a pseudocontinuous helix in the crystal.)

Helix 3: Invariant Residues and Critical Contacts with the DNA

Helix 3 fits directly into the major groove and makes extensive contacts with the bases and with the sugar-phosphate backbone. It is interesting to note the critical roles played by Trp-48, Phe-49, Asn-51, and Arg-53. These residues occur in every one of the higher eukaryotic homeodomains compiled by Scott et al. (1989), and the crystal structure shows that these invariant residues occur in the section of helix 3 closest to the major groove (Figure 4). Trp-48 and Phe-49 form a key part of the hydrophobic core. They must play a major role in stabilizing the folded

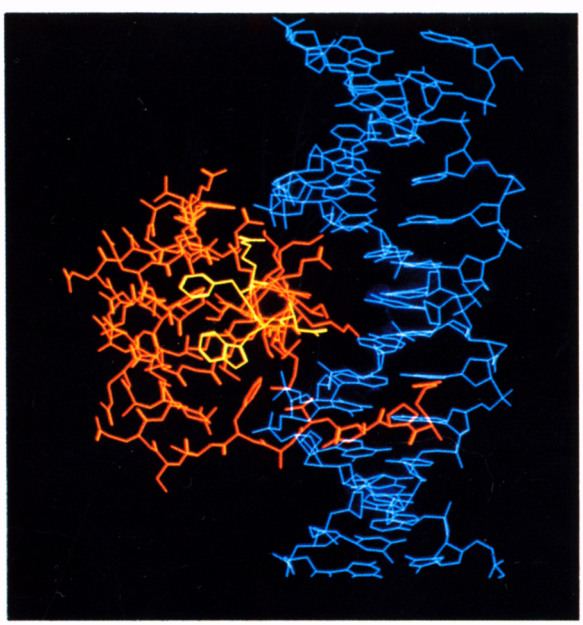

Figure 4. Photograph Highlighting the Position of Invariant Residues: Trp-48, Phe-49, Asn-51, and Arg-53

This view is looking down the axis of helix 3. The homeodomain is shown in orange with invariant residues highlighted in yellow. The DNA is shown in blue.

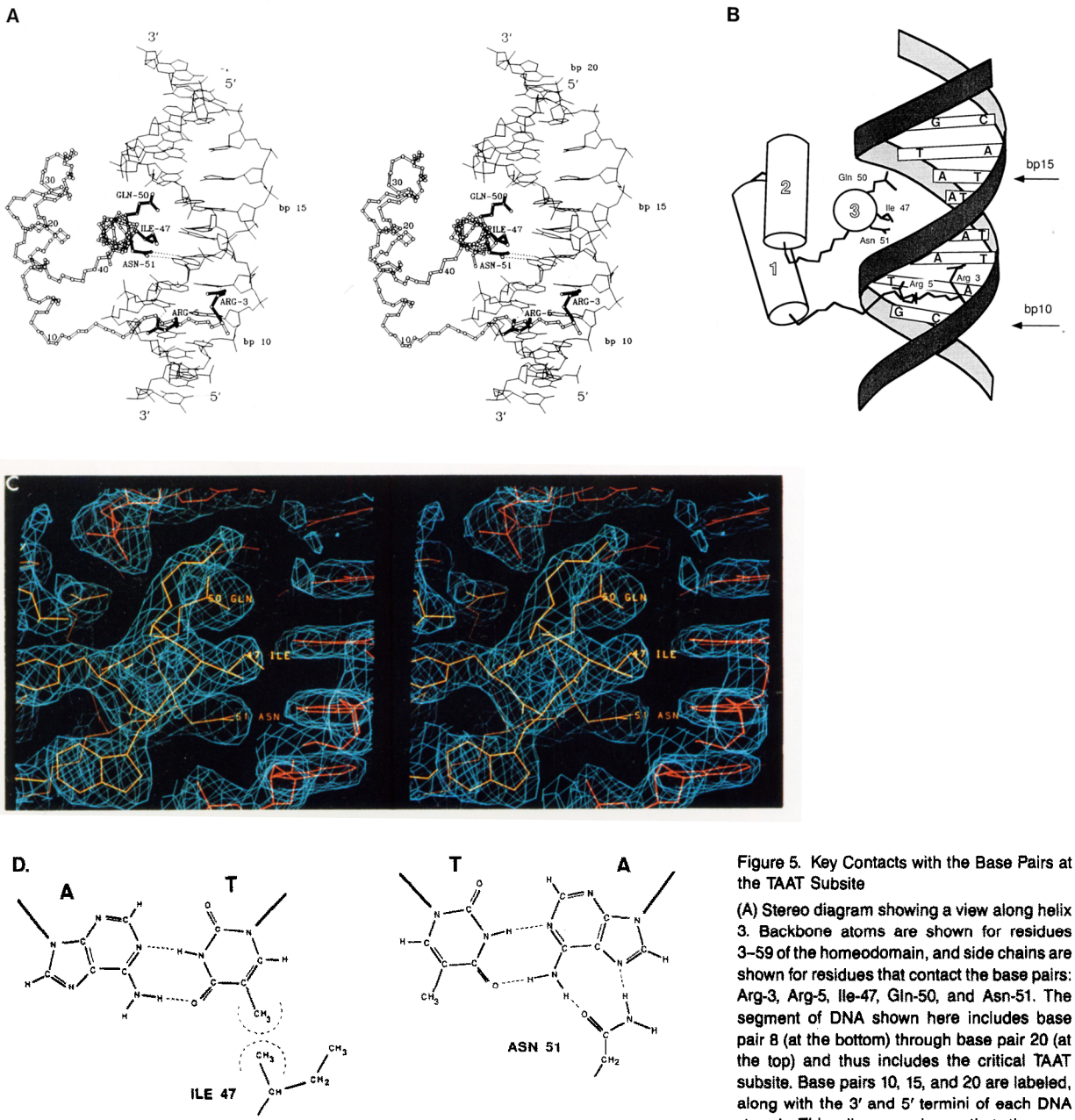

Figure 5. Key Contacts with the Base Pairs at the TAAT Subsite

(A) Stereo diagram showing a view along helix 3. Backbone atoms are shown for residues 3–59 of the homeodomain, and side chains are shown for residues that contact the base pairs: Arg-3, Arg-5, Ile-47, Gln-50, and Asn-51. The segment of DNA shown here includes base pair 8 (at the bottom) through base pair 20 (at the top) and thus includes the critical TAAT subsite. Base pairs 10, 15, and 20 are labeled, along with the 3′ and 5′ termini of each DNA strand. This diagram shows that the arm makes minor-groove contacts near the 5′ end of the TAAT subsite and that helix 3 makes major-groove contacts near the 3′ end of the TAAT subsite.

(B) Sketch summarizing the critical contacts.

(C) Stereo view of the calculated electron density from a $2F_o - F_c$ map in the vicinity of helix 3. The protein is shown in yellow, and the side chains of Ile-47, Gln-50, and Asn-51 are labeled. The DNA is shown in red and the electron density is shown in blue. The electron density is contoured at a level of one rms deviation above the average density.

(D) Sketch showing the base contacts made by Ile-47 and Asn-51.

structure and in controlling how helix 1 packs against helix 3. (This will be critical for DNA recognition because it affects the spatial relationship between contacts made by the N-terminal arm and contacts made by helix 3.)

The invariant hydrophilic residues—Asn-51 and Arg-53—make critical contacts with the DNA. Asn-51 makes a pair of hydrogen bonds with the adenine at base pair 13, donating a hydrogen bond to the N7 position and accepting a hydrogen bond from the N6 position (Figure 5). Arg-53 hydrogen bonds with two phosphate groups on the other strand of the DNA (Figures 6 and 7).

Several of the neighboring residues in helix 3 make critical contacts with the DNA and are conserved within subsets of the homeodomain proteins. The side chains of Ile-47 and Gln-50 appear to be especially important for DNA recognition. Ile-47 provides a sequence-specific interaction by making hydrophobic contacts with the methyl group of the thymine at base pair 14. Valine, which occurs

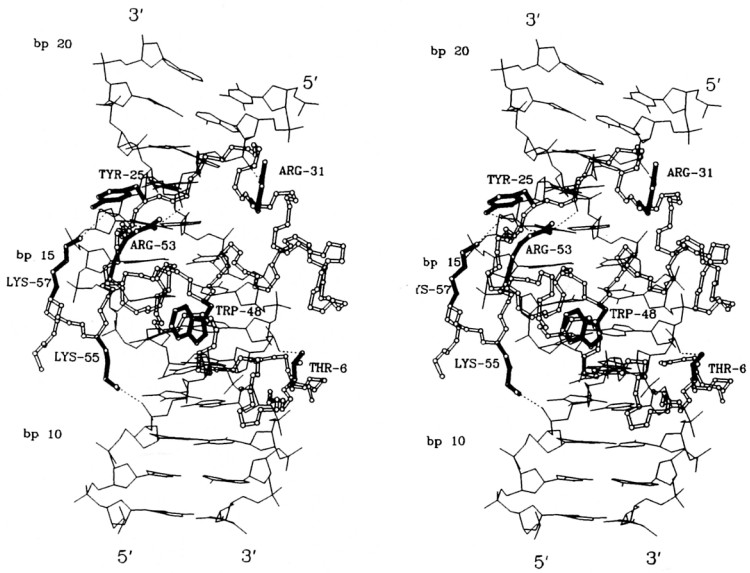

Figure 6. Contacts with the Phosphates

Stereo diagram showing side chains that contact the phosphodiester oxygens. The DNA segment is the same as that in Figures 3A and 5A. The view is roughly perpendicular to helix 3 and thus is similar to the view in Figure 3. Critical residues include Thr-6, Tyr-25, Arg-31, Trp-48, Arg-53, Lys-55, and Lys-57. These contacts also are summarized in Figure 7.

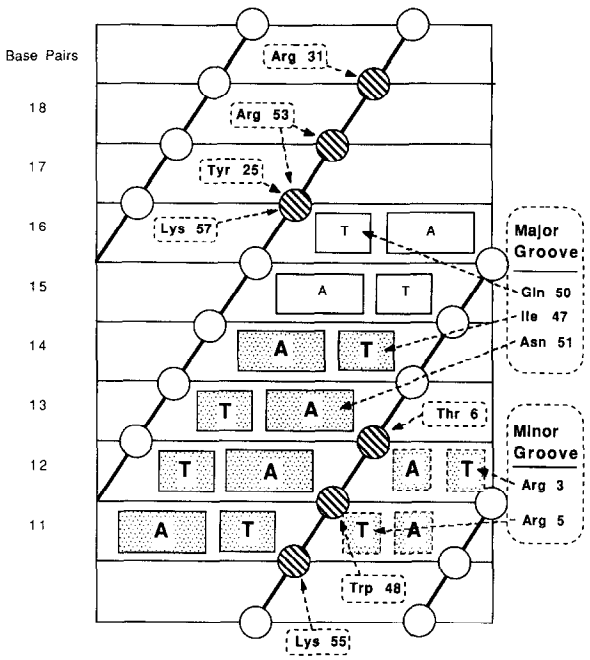

Figure 7. Sketch Summarizing All the Contacts Made by the Homeodomain

The DNA is represented as a cylindrical projection, and the shading emphasizes the TAAT subsite. Phosphates are represented with circles; hatched circles show phosphates that are contacted by the homeodomain.

at this position in other homeodomains, would be able to make a similar hydrophobic contact. The side chain of Gln-50 projects into the major groove and clearly is in a position to make sequence-specific contacts. In the current structure, Gln-50 makes van der Waals contacts with the methyl group of the thymine at base pair 16. A modest rotation of the side chain would allow it to hydrogen bond with the O4 of the thymine at base pair 14, but the -NH_2 of the Gln side chain and the O4 of the thymine are about 4 Å apart in the current structure. It also appears that small changes in the DNA conformation could allow Gln-50 to hydrogen bond with the adenine at base pair 15. (The DNA geometry is discussed in a later section.) Residues on helix 3 also make an extensive set of contacts with the sugar-phosphate backbone (Figure 6); these contacts are discussed in a later section.

Contacts Made by the N-Terminal Arm

The crystal structure shows that the N-terminal arm binds to the minor groove of the DNA and reveals that conserved residues from the arm contact base pairs adjacent to the ones contacted by helix 3. The poor electron density for residues 1 and 2 of the engrailed homeodomain indicates that this region is disordered in the crystal. However, some backbone density can be seen for residue 2, and residues 3–5 form a well-defined region of extended chain that fits into the minor groove. The side chain of Arg-5, the most highly conserved residue in this portion of the homeodomain, reaches directly into the minor groove and hydrogen bonds with the O2 of the thymine at base pair 11. The electron density for Arg-3 is not nearly as well defined as the density for other side chains. However, it appears that the side chain of Arg-3 hydrogen bonds with the O2 of the thymine at base pair 12 and/or hydrogen bonds with the sugar oxygen from the adenosine at base pair 13.

Further studies are needed to understand what degree of sequence specificity is provided by these minor-groove contacts. Clearly these contacts explain a preference among homeodomains for AT-rich sites, since these have hydrogen bond acceptors at appropriate positions in the minor groove. Presumably a GC or a CG base pair would interfere with binding since the arginine side chain would have unfavorable steric or electrostatic interactions with the -NH_2 of the guanine. However, it is not clear whether these contacts could distinguish an AT from a TA base pair, since the N3 of adenine and the O2 of thymine occupy similar positions in the minor groove (Seeman et al., 1976).

Contacts with the DNA Backbone

The homeodomain makes an extensive set of contacts with the sugar-phosphate backbone, and we presume that these are critical for DNA binding. The full set of backbone contacts is shown in Figure 6 and summarized (along with the base contacts) in Figure 7. Most of these contacts are clustered in two regions flanking the sites where helix 3 makes contacts in the major groove. (In the sequence, each of these regions is offset in the 5′ direction along the DNA backbone; in the three-dimensional structure, these contacts surround the regions where helix 3 contacts specific base pairs.) The engrailed homeodomain makes one set of contacts with phosphates just "above" the region where helix 3 contacts bases in the major groove (Figures 6 and 7). At the edge of the site, Arg-31 contacts the phosphate that lies between the adenine at base pair 19 and the guanine at base pair 18. Proceeding in the 5′ to 3′ direction along this strand, we see that Arg-53 contacts the next phosphate. The phosphate after this appears especially critical and has contacts from Arg-53, Tyr-25, and Lys-57. (It is interesting that helix 1 does not make any contacts with the DNA and that Tyr-25 and Arg-31 provide the only DNA contacts from the loop or from helix 2.)

On the other strand, there are several contacts with phosphates just "below" the region where helix 3 fits into the major groove (Figures 6 and 7). Lys-55 contacts a phosphodiester oxygen from the thymidine at base pair 11. Proceeding in a 5′ to 3′ direction, we see that Trp-48 is very close to the next phosphate. The ring nitrogen is not in a favorable position for hydrogen bonding to the DNA, but the partial positive charge on the edge of the aromatic ring (Burley and Petsko, 1985) may provide a favorable electrostatic interaction with the phosphodiester oxygen. Thr-6 makes side chain and main chain hydrogen bonds to a phosphodiester oxygen of the adenosine at base pair 13.

Table 1. Local Helical Parameters for the DNA Site

Position	Base Pair	Twist (Degrees) Helical	Twist (Degrees) Propeller	Tilt (Degrees)
2	T·A		15.8	4.1
		35.3		
3	T·A		9.2	0.8
		32.1		
4	T·A		15.3	5.5
		33.1		
5	G·C		10.4	2.2
		37.4		
6	C·G		6.9	1.5
		27.9		
7	C·G		5.5	3.9
		44.7		
8	A·T		6.7	6.1
		29.5		
9	T·A		10.0	4.5
		40.6		
10	G·C		12.6	2.9
		28.1		
11	T·A		5.6	6.1
		39.8		
12	A·T		18.3	5.4
		35.1		
13	A·T		21.5	6.2
		30.4		
14	T·A		15.2	5.4
		32.2		
15	T·A		12.9	3.9
		34.5		
16	A·T		14.6	4.6
		31.0		
17	C·G		21.8	7.7
		40.6		
18	C·G		17.4	16.6
		31.9		
19	T·A		12.9	18.8
		33.6		
20	A·T		13.8	21.7
		32.6		
21	A·T		20.2	18.7

DNA Conformation

The DNA duplex in the crystal is a relatively straight segment of B-DNA. The average helical twist of 34.2° (Table 1) corresponds to 10.53 bp per turn. This is very close to the average expected for B-DNA (Wang, 1979), and this suggests that protein binding and crystallization have not resulted in any large overall distortion of the DNA structure. As observed in single-crystal studies of B-DNA (Dickerson and Drew, 1981), the base pairs have significant propeller twist (average = 13.3°). The individual helical twists range from 28.1° to 44.7°, and the individual propeller twists range from 5.4° to 21.7° (Table 1). Base pairs at one end of the duplex (base pairs 13–14) have a significant tilt. The change in tilt occurs near base pairs 17 and 18, and the minor groove is unusually wide in this region.

Superimposing B-DNA on the complex confirms that there are no drastic distortions in the binding site, but it is clear that the major groove is several angstroms wider than normal in the region where helix 3 binds. Most of the changes seem to occur in the DNA strand that contacts the C-terminal portion of helix 3. The slightly lower helical twists between base pairs 14 and 15 and between base pairs 13 and 14 may contribute to the widening of the major groove. The base pairs also have a significant tilt in this region, and this may affect the groove width.

TAAT Subsite Allows Alignment with Other Binding Sites

A deeper understanding of homeodomain–DNA interactions requires that we integrate the structural data with results from genetic and biochemical studies. To proceed with any detailed comparison, we need to align the binding site used in the crystal with binding sites used in other studies. Obtaining the correct alignment is complicated by the fact that individual homeodomain proteins can recognize a variety of different binding sites. However, the subsequence TAAT occurs in most homeodomain binding sites (Scott et al., 1989), and recent experiments with Ultrabithorax have emphasized the importance of a TAAT core (S. C. Ekker, K. E. Young, and P. A. Beachy, submitted). Aligning this subsite should provide the best prospect for correlating our structural data with results from other studies.

The most plausible alignment uses the TAAT subsite

that includes base pairs 11–14 on the upper strand. This is satisfying from a structural perspective, since the engrailed homeodomain contacts each of these base pairs. Another TAAT sequence occurs at base pairs 16–13 on the lower strand, but the engrailed homeodomain makes fewer contacts with this TAAT sequence and it seems less plausible that this could constitute the "core consensus" binding site. Other data confirm that the TAAT site at base pairs 11–14 is the appropriate one to use when aligning sequences:

First, this alignment is consistent with the model that S. D. Hanes and R. Brent (submitted) derived from an elegant genetic analysis of the homeodomain–DNA contacts. Their data allowed them to infer that the ninth residue of the recognition helix (i.e., Gln-50 of engrailed) makes critical contacts just to the 3′ side of the TAAT subsite. Our alignment is fully consistent with their model, since Gln-50 contacts base pair 16. Their data exclude the alternative assumption (that base pairs 16–13 provide the critical TAAT) since this would leave Gln-50 near the 5′ end of the TAAT subsite.

Second, ethylation interference experiments have been done with an Antennapedia binding site (Affolter et al., 1990) that contains a TAAT sequence. Our preferred alignment matches every phosphate contact seen in the engrailed cocrystal structure with a phosphate contact inferred from these ethylation interference experiments. (It seems safe to assume that the Antennapedia and engrailed homeodomains make generally similar contacts with the DNA backbones, and studies of the bacterial repressors have demonstrated that there is a very close correlation between the contacts seen in a cocrystal and contacts inferred from ethylation interference experiments [Johnson, 1980; Jordan and Pabo, 1988].)

Third, footprinting experiments using fragments of Oct-1 and complexes of Oct-1 with other proteins show that the Oct-1 homeodomain binds to a TAAT subsite and indicate that helix 2 is on the 3′ side of this TAAT site (T. M. Kristie and P. A. Sharp, submitted).

Comparison with Biochemical and Genetic Data about Homeodomain–DNA Contacts

The structure reported here is consistent with a vast body of data about homeodomain–DNA interactions. One obvious and satisfying aspect of the structure is the important role played by the most highly conserved residues of the homeodomain. Our structure shows that each of the invariant residues plays a critical role in folding and/or recognition. The structure also helps us understand why genetic experiments have pinpointed residue 50 as a critical residue for controlling specificity of the homeodomain–DNA interactions (Hanes and Brent, 1989; Treisman et al., 1989; S. D. Hanes and R. Brent, submitted). This side chain points directly into the major groove and clearly is in an excellent position to contribute to the specificity of binding and recognition. In the current complex, it forms a van der Waals contact with the methyl group of a thymine, but we were surprised that Gln-50 does not hydrogen bond to the adenine at base pair 15 or to some other position in the major groove. Studies are in progress to see whether Gln-50 can hydrogen bond to a base under other circumstances. For example, it is possible that binding of the neighboring homeodomain distorts the DNA enough to prevent Gln-50 from making its preferred contact. The neighboring homeodomain contacts the phosphates between bases 16 and 19 on the upper strand. The high propeller twist at base pair 17, the high helical twist between base pairs 17 and 18, and the unusual tilt angles in this region (Table 1) all suggest that the DNA is somewhat distorted in this region. To address this issue, we have recently grown cocrystals in which the upper strand has been synthesized as two separate segments. After annealing, this leaves a "nick" between nucleotides 16 and 17 on the upper strand, and this may prevent the binding of one homeodomain from distorting the binding site of the other. In vitro selection schemes are also being used to find the optimal sequence for engrailed binding (E. M.-B. and T. K., unpublished data) and these experiments should help determine whether Gln-50 might prefer other bases at positions 15 and 16.

Residue 50 plays an important role in distinguishing one homeodomain binding site from another (Hanes and Brent, 1989; Treisman et al., 1989), but the structure makes it clear that Ile-47 and Asn-51 also play very important roles in recognition. The genetic experiments focused on differences in amino acid sequence that were responsible for differences in specificity. Asn-51 is invariant, and Ile or Val (which could make a similar sequence-specific contact) is almost always present at position 47. Contacts made by these residues will have a central role in recognizing the TAAT subsite and in distinguishing this from nonspecific DNA.

The structure reported here readily explains the other highly conserved residues and regions of the homeodomain. A number of conserved residues—for example, Ile-16 and Phe-20—help to form the hydrophobic interior in engrailed and Antennapedia (Qian et al., 1989). The crystal structure of the complex also reveals important contacts made by highly conserved residues (such as Arg-3 and Arg-5) near the N-terminal end of the homeodomain and by highly conserved residues near the C-terminal end of the homeodomain (Figures 5, 6, and 7).

The Homeodomain Uses the HTH Motif in a Novel Way

Sequence comparisons had suggested (Laughon and Scott, 1984; Shepard et al., 1984) and nuclear magnetic resonance studies of the Antennapedia homeodomain had confirmed (Qian et al., 1989) that the homeodomain contains an HTH motif that is structurally similar to that observed in the prokaryotic repressors (Pabo and Sauer, 1984). The engrailed structure also confirms that these HTH motifs are quite similar. The C_αs for residues 33–52 of λ repressor can be superimposed on the C_αs for residues 31–50 of the engrailed homeodomain with a root-mean-square (rms) distance of only 0.84 Å between corresponding atoms (Figure 8A). This rms distance is only slightly larger than the distances typically obtained when superimposing the HTH units of two prokaryotic repressors. Some of the aligned residues in the two proteins

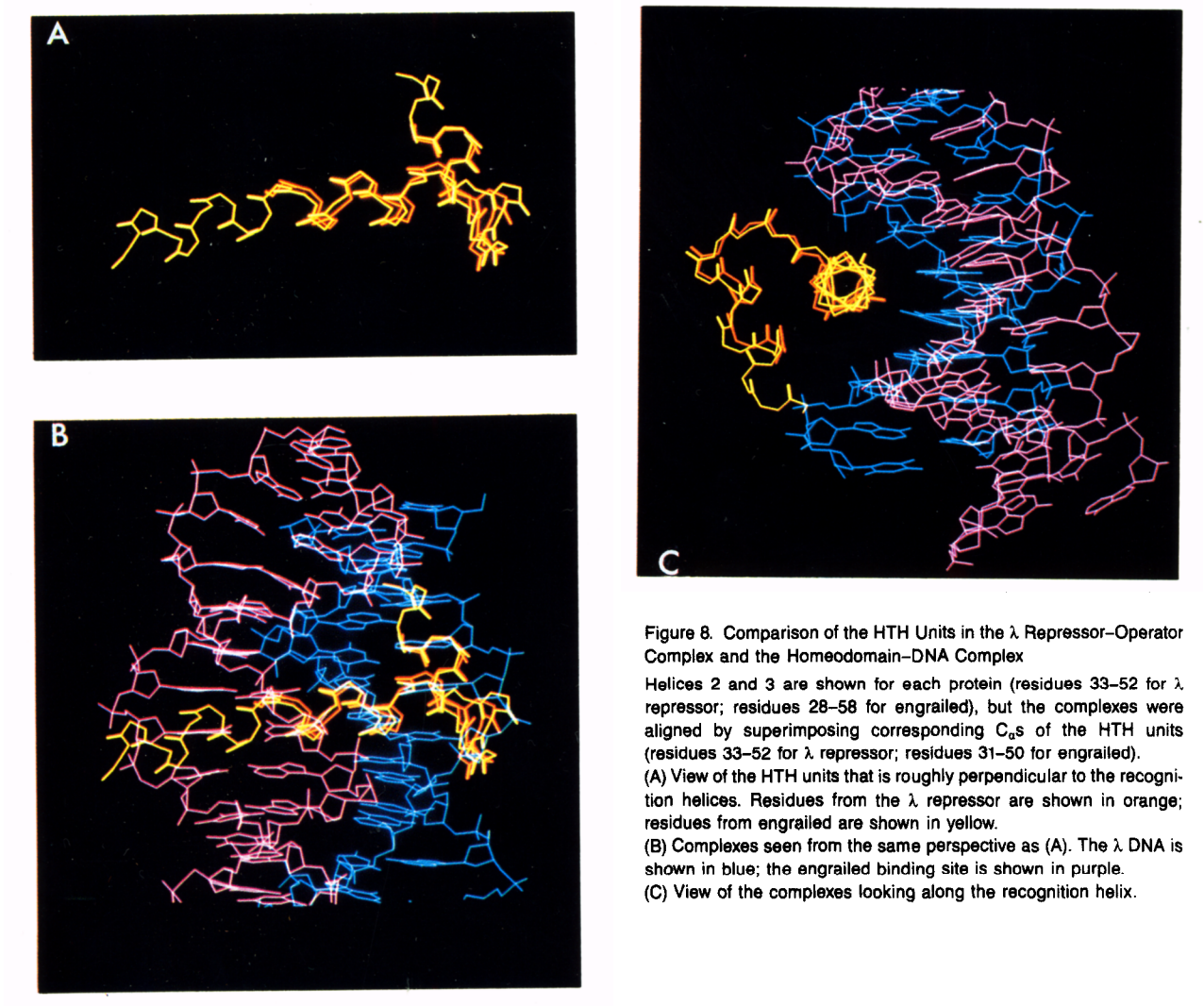

Figure 8. Comparison of the HTH Units in the λ Repressor–Operator Complex and the Homeodomain–DNA Complex

Helices 2 and 3 are shown for each protein (residues 33–52 for λ repressor; residues 28–58 for engrailed), but the complexes were aligned by superimposing corresponding C_αs of the HTH units (residues 33–52 for λ repressor; residues 31–50 for engrailed).

(A) View of the HTH units that is roughly perpendicular to the recognition helices. Residues from the λ repressor are shown in orange; residues from engrailed are shown in yellow.

(B) Complexes seen from the same perspective as (A). The λ DNA is shown in blue; the engrailed binding site is shown in purple.

(C) View of the complexes looking along the recognition helix.

(e.g., Ile-45 of engrailed and Val-47 of λ repressor) play similar roles in stabilizing the HTH units, and these are precisely the positions that are most diagnostic in sequence searches for the HTH unit.

In spite of the similar backbone structures for the HTH units, the engrailed cocrystal structure reveals that the HTH units are used in significantly different ways in the homeodomain–DNA and repressor–operator complexes. Superimposing the two HTH units allows one to establish a common frame of reference for the two complexes, and it is clear that the DNA duplexes have very different positions in this reference frame (Figure 8B). If we take the arrangement seen in the λ repressor–operator complex as a starting point, it appears that the DNA in the homeodomain complex has been shifted toward the C-terminal end of the second helix in the HTH unit. These differences also are apparent if we compare the positions of residues that make critical contacts with the base pairs. Residues near the N-terminal end of the recognition helix make critical contacts in the λ repressor–operator complex. The critical residues in engrailed are near the center of an extended recognition helix.

A side view reveals other differences between the two arrangements (Figure 8C). In the λ repressor–operator complex, the first helix of the HTH unit (helix 2) fits partway into the major groove, and the N-terminal end of this helix contacts the sugar-phosphate backbone of the DNA. In contrast, the first helix of the homeodomain's HTH unit (helix 2) lies above the major groove. Relative to the framework provided by the superimposed HTH units, it appears that the DNA for the homeodomain has been rotated away from the N-terminal end of this helix. Although the position of helix 2 is rather different in the λ and in the engrailed complexes, it clearly plays related roles in the two complexes. Since it packs against helix 3, we expect that it will stabilize folding of the recognition helix. It also seems to serve as an "outrigger" that will prevent helix 3 from rocking in the major groove. (Moving in one direction would cause helix 2 to collide with the sugar-phosphate backbone; moving in the other direction would break the hydrogen bonds that helix 2 makes with the DNA backbone.) It is interesting that Arg-31, which provides a phosphate contact in the engrailed complex, is aligned with Gln-33, which provides a corresponding contact in the λ complex.

Even though helix 2 of engrailed is longer and has a different orientation with respect to the DNA, a structurally analogous residue contacts the DNA backbone.

Implications for Understanding Protein–DNA Interactions

These differences in the arrangement of the HTH unit appear dramatic and certainly help us understand why the homeodomain proteins constitute a distinct subfamily of HTH proteins. However, the prokaryotic and eukaryotic complexes are similar in many fundamental ways. Although the homeodomain–DNA complex reveals a number of new and interesting features, it supports most of the fundamental ideas that have been developed from structural studies of the prokaryotic HTH proteins. In this section we try to place the homeodomain–DNA complex in a broader perspective. We discuss the role of α helices in recognition and discuss the respective roles of contacts with the base pairs and contacts with the sugar-phosphate backbone.

Fundamentally, the HTH unit of the prokaryotic repressors provides a way of positioning an α helix in the major groove, and side chains on this helix play a major role in site-specific binding (Jordan and Pabo, 1988; Aggarwal et al., 1988). (The Trp repressor–operator complex does not fit this paradigm [Otwinowski et al., 1988], but recent experiments suggest that the Trp repressor may have been crystallized with a nonspecific binding site [Staacke et al., 1990].) Despite significant differences in the position and orientation of the helices, it is clear that the recognition helix of the homeodomain has a fundamentally similar role. In fact, there was no reason to suppose that the position of the recognition helix needed to be conserved in all HTH proteins. There are no strict constraints on this relationship because there is no strict relationship between the periodicity of an α helix and the periodicity of B-DNA, and at an atomic level the shape and appearance of an α helix will be dominated by the amino acid sequence of the helix. Since the sequence will change from one protein to the next, the recognition helix will look significantly different and may pack against the DNA in a significantly different way.

Structural studies of the prokaryotic repressors also made it clear that a single α helix would only be able to contact a few base pairs. This constraint has a relatively simple geometric basis: since the major groove has the shape of a "helical saddle," a straight α-helical segment can only contact a few adjacent base pairs. Even though the recognition helix of the homeodomain is much longer, we see a similar constraint in this complex. A dramatic kink—sufficient to generate independent helical segments—would be needed to keep a longer helical region in contact with the major groove. Although the solution structure of the Antennapedia homeodomain had suggested that there was a small kink (of about 30°) after residue 52, the corresponding region of engrailed forms a continuous helix in the homeodomain–DNA complex. In any case, modeling experiments show that a small kink—like that reported for the Antennapedia homeodomain—would not allow significant additional contacts in the major groove.

Structural studies of prokaryotic DNA binding proteins indicate that there is no "code" for protein–DNA interactions (i.e., that there are no rigid rules determining which amino acids can contact which base pairs). The structure of the engrailed–DNA complex supports the notion that recognition depends on the detailed structure of the protein–DNA interface, but we note that the contacts between Asn-51 and adenine are similar to the Gln–adenine contacts seen in the λ and 434 complexes (Jordan and Pabo, 1988; Aggarwal et al., 1988). This type of contact—which had been proposed by Seeman et al. (1976)—may be an especially favorable hydrogen bonding arrangement and may play a significant role in site-specific recognition. Comparisons of the λ and 434 complexes (Pabo et al., 1990) had also emphasized that amino acids tend to make similar contacts when they occur at similar positions within the HTH unit, and it will be interesting to see whether similar relationships hold among the set of homeodomain proteins. It may be possible, within the context of a highly conserved structure, to predict the bases that will be preferred by particular amino acids at particular positions of the HTH unit.

The homeodomain makes an extensive set of contacts with the DNA backbone, and we presume that these play an important role in binding and recognition. As in the λ repressor–operator complex, these contacts occur just on the 5′ sides of the region where the α helix contacts the base pairs, and in three dimensions these contacts are on the closest edges of the major groove. Although these backbone contacts could have some role in "indirect readout" of sequence information, it seems unlikely since this region is relatively uniform B-DNA. It seems simpler to imagine that their primary role is to provide a set of "fiducial marks" that help to align the homeodomain as it approaches the DNA. Such contacts would help to control the position and orientation of the recognition helix and thus would serve to enhance the specificity of complex formation.

Another interesting aspect of the homeodomain–DNA interactions involves the minor-groove contacts that are made by the N-terminal arm. Although Arg-43 of the 434 repressor fits partway into the minor groove (Aggarwal et al., 1988), the engrailed structure is the first complex to show how several side chains from an extended chain can make base-specific contacts in the minor groove. The engrailed contacts may be very similar to the contacts made by the Hin recombinase, which has been studied by incorporating EDTA–Fe at particular amino acid residues and then mapping DNA cleavage patterns (Sluka et al., 1990). Like engrailed, the N-terminal portion of this DNA binding domain contains an Arg-Pro-Arg sequence, and high resolution chemical footprinting experiments show that this N-terminal arm binds in the minor groove.

The λ repressor also has an extended N-terminal arm that contributes to site-specific binding, and some of the critical residues have identical positions when the two protein sequences are aligned. However, the repressor's arm binds in the major groove (Jordan and Pabo, 1988). Helix 1 has a very different orientation in the two proteins, and this seems to determine which groove will be closest to

the N-terminal arm. Future studies will be needed to understand the precise role of contacts that engrailed makes in the minor groove, and it will be important to see how this N-terminal region of the homeodomain is constrained when it is present in the context of the intact protein. It also will be interesting to determine how often similar interactions occur in other proteins and whether regions of extended polypeptide chain might make similar contacts in the major groove of A-DNA (which is too narrow to accommodate an α helix).

General Principles of Homeodomain–DNA Interactions

The structure of the engrailed–DNA complex suggests some general principles about specificity in homeodomain–DNA interactions. It appears that the highly conserved residues on helix 3 and a few residues on the N-terminal arm form a sort of core recognition unit that is responsible for many of the contacts with the TAAT subsite. However, there will be a vast number of TAAT sequences in the genome, and these contacts, even when supplemented by sequence-specific contacts from residue 50, would presumably not be sufficient to provide for the differential regulation of gene expression. Specificity could be enhanced in many ways:

First, since the homeodomain occupies both the major and minor grooves, it may not be able to bind to DNA within nucleosomes. This could drastically simplify the problem of finding the appropriate binding sites.

Second, other mechanisms could be involved in particular cases. Cooperative binding of homeodomains to neighboring sites, binding in conjunction with other proteins, and dimerization of homeodomain proteins may increase the effective specificity of binding in particular systems.

Third, it also is interesting that both the N-terminus and C-terminus of the homeodomain are close to the DNA and that particular subsets of the homeodomain proteins often have a cluster of conserved residues immediately preceding or immediately following the standard 60 amino acid homeodomain. Some of these conserved clusters of residues may serve to modulate site-specific binding. The cocrystal structure makes it clear that these neighboring regions would be in an excellent position to directly contact neighboring bases on the DNA or to influence the structure or orientation of the N-terminal arm and the C-terminal helix. Finally, they also could provide "attachment" or "targeting" sites for other proteins that would bind to neighboring regions and modulate the specificity and/or affinity of DNA binding. At this stage, little information is available about the precise role of neighboring regions. However, the POU domain is a particularly striking example of a nearby conserved region that occurs in a number of homeodomain proteins (Herr et al., 1988), and recent experiments suggest that the POU domain recognizes a neighboring subsite on the DNA (Ingraham et al., 1990; Phillip Sharp, personal communication).

Although our model is somewhat speculative and schematic at this stage, it provides a clear connection between the structure of the complex and the known biological roles of the homeodomain proteins. Differential or modulated recognition may be particularly important for the subtle regulatory controls involved in differentiation and development. These structural data also provide a simple picture of how the family of homeodomain proteins may have diverged. Recognition could be based on a set of contacts with a core consensus sequence, and modulating the interactions would generate new specificities and therefore new regulatory activities.

These ideas provide a framework for thinking about family–subfamily relationships in proteins containing homeodomains. It clearly will be necessary to get structural information about other homeodomain–DNA complexes (particularly about the intact proteins) and to use the structural data to design more incisive experiments about the roles that particular residues or regions of the protein play in sequence-specific binding.

Conclusions

This study reveals the following basic structural features of homeodomain–DNA interactions:

The homeodomain makes contacts in both the major and minor grooves, and the critical contacts are centered on a conserved TAAT subsite that biochemical studies have highlighted as the most important part of the homeodomain binding site (S. C. Ekker, K. E. Young, and P. A. Beachy, submitted).

The HTH unit plays a significant role in recognition, but the helices in the homeodomain are longer than the corresponding helices of λ repressor and the orientation of the HTH unit with respect to the DNA is rather different. In the homeodomain, the first helix of the HTH unit lies entirely above the major groove. The second helix of the HTH unit (the recognition helix) fits directly into the major groove, but the critical contacts occur in a region that would correspond to the C-terminal end of the canonical (20 residue) HTH unit. However, the recognition helix is much longer in the homeodomain–DNA complex and residues that contact the bases are near the center of this extended helix.

Each of the invariant residues (defined by comparing sets of homeodomain sequences) plays a central role in folding and/or recognition. Trp-48 and Phe-49 form a central part of the hydrophobic core, and Trp-48 may have favorable electrostatic interactions with a phosphodiester oxygen. Asn-51 makes a pair of hydrogen bonds with an adenine at the third position of the TAAT subsite (TAAT). Arg-53 hydrogen bonds with a pair of phosphate groups on the DNA backbone.

Two other residues contact bases in the major groove. Ile-47 makes hydrophobic contacts with a thymine at the fourth position of the TAAT subsite (TAAT). Gln-50 projects directly into the major groove. In the current structure, it makes van der Waals contacts with a thymine methyl group, but small motions would allow Gln-50 to interact with several different positions near the 3′ side of the TAAT subsite (TAATNN).

Residues near the N-terminal end of the homeodomain make minor-groove contacts near the 5′ end of the conserved TAAT subsite. Arg-5 hydrogen bonds to the thy-

Table 2. Statistics for Data and Derivatives

Item	Native	IdU_{16}	IdU_1	IdU_{1+16}
Resolution (Å)	2.8	2.7	3.8	2.8
Measured reflections	21,130	26,820	12,900	14,403
Unique reflections	9,072	10,052	3,849	8,969
R_{sym}[a]	4.26	4.34	3.34	3.23
Mean isomorphous difference[b]		0.21	0.11	0.20
Phasing power[c]		1.71	1.32	2.04
Cullis R factor[d]		0.60	0.65	0.56

Designations for the derivative data sets indicate the base pair(s) at which 5-iodouracil was substituted for thymine in the DNA used for crystallization.
[a] $\Sigma_h\Sigma_i|I_{h,i} - I_h|/\Sigma_h\Sigma_i I_{h,i}$, where I_h is the mean intensity of the i observations of reflection h.
[b] $\Sigma|F_{PH} - F_P|/\Sigma F_{PH}$, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively.
[c] $[(F_{H(calc)}{}^2/|(F_{PH(obs)} - F_{PH(calc)})^2]^{1/2}$.
[d] $\Sigma||F_{der} \pm F_{nat}| - F_{H(calc)}|/\Sigma|F_{der} - F_{nat}|$ for centric reflections, where $F_{H(calc)}$ is the calculated heavy atom structure factor.

mine at the first position (TAAT). Arg-3 appears to hydrogen bond to the thymine of the second base pair (TAAT), although electron density for this side chain is not as clear as for the other contacts.

There are extensive contacts with the sugar-phosphate backbone, and many of these contacts are made by residues in the N-terminal arm and by residues in helix 3.

Helix 1 does not make any direct contacts with the DNA. Only one residue from the following loop and only one residue from helix 2 actually contact the DNA. The primary role of these helices is to help stabilize the folded structure and to help fix the relative orientation of the N-terminal arm and helix 3.

The limited number of base contacts in the complex suggests that the isolated homeodomain can provide only a modest amount of sequence specificity. Cooperative binding, with other homeodomains and/or with other regulatory proteins, may serve to enhance specificity. However, since the N-terminal and C-terminal regions of the homeodomain are near the DNA, it also is possible that neighboring regions of the intact proteins could modulate the affinity or specificity and thus allow differential regulation of gene expression.

Experimental Procedures

The cocrystals were grown from a solution that contained equimolar amounts of the engrailed homeodomain and duplex DNA in a buffer containing 30 mM Bis-Tris–HCl. When the crystallization drops were set up, the pH was raised to 8.0–9.0 by the addition of ammonium hydroxide; crystals grew as the ammonium hydroxide diffused out and the pH returned to 6.7. Precession photographs revealed that the cocrystals form in space group C2 with a = 131.2 Å, b = 45.5 Å, c = 72.9 Å, and β = 119.0° (Liu et al., 1990). The crystals diffract to 2.5 Å in all directions, but the current data set is weak beyond 2.7 or 2.8 Å.

Native diffraction data were collected on a Xentronics area detector (Table 2), and isomorphous derivatives were obtained by preparing duplex DNA that had 5-iodouracil substituted for thymine at specific positions. The first derivative had iodouracil substituted at base pair 16 on the lower strand, and the second derivative had iodouracil substituted on the 5′ end of the upper strand. A third derivative had iodouracil at both positions. Difference Pattersons revealed that there was one DNA duplex in the asymmetric unit, and the doubly substituted derivative allowed us to determine the relative y coordinates for the iodine sites.

Heavy-atom parameters were first refined using the program REFINE from the CCP4 (1979) package (S.E.R.C. [U.K.] Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography, distributed from Daresbury Laboratory, Warrington, UK) and these data were used to phase an initial MIR map. This first map was relatively noisy, but the phases were improved by refining the heavy atom parameters against solvent-flattened phases (Rould et al., 1989). This is a cyclic process involving the following: B.-C. Wang's protocol for iterative phase improvement (Wang, 1985); refinement, using the program PHARE, of the heavy-atom parameters against the Wang phases; and calculation of new MIR phases. After several cycles, inspection of the Wang map revealed left-handed α helices, so we inverted the coordinates of the heavy atom sites and proceeded with several more cycles of flattening and refinement. Phases for the last MIR map had a mean figure of merit of 0.67 for data from 10–3.5 Å resolution. After a final round of solvent flattening, this map gave excellent density for the protein, and the map immediately revealed that there were two copies of the homeodomain bound to the 21 bp DNA fragment. We assumed that the structure of the engrailed homeodomain would be quite similar to that of the Antennapedia homeodomain, and we were able to rapidly fit the protein density by using a model extracted from published stereo photographs of the Antennapedia homeodomain (Qian et al., 1989; Rossmann and Argos, 1980).

The electron density for the DNA was less clearly defined, but it was readily fit by using a model of uniform B-DNA and matching the refined iodine positions with the methyl groups of the appropriate thymines. The program X-PLOR was used for refinement of this initial model (Brünger et al., 1987; Brünger, 1990). The rigid-body refinement option was used for an initial adjustment of the overall position and orientation of the DNA and two protein molecules. The first cycle of simulated annealing gave R = 0.30 for data from 10–3.2 Å, but extensive rebuilding of the DNA was required. Subsequent cycles of rebuilding and refinement gave the current model, which has R = 0.244 for data from 10–2.8 Å resolution using a single overall temperature factor of 17.2 $Å^2$ and without any water molecules included in the model. This model has good stereochemistry. The rms deviation for bond lengths is 0.019 Å, and the rms deviation for bond angles is 3.7°. This model fits the MIR map extremely well. It also has been confirmed with a procedure developed by Axel Brünger (personal communication) that involves systematically deleting short segments of the structure, using simulated annealing to minimize model bias from the phases, and examining $2F_o - F_c$ maps that span the deleted region.

Because structural analysis revealed that two homeodomains bind to the duplex used for cocrystallization, we synthesized two other duplexes so that we could estimate the intrinsic affinity of the engrailed homeodomain for each of these subsites. To eliminate binding at the end, we synthesized a DNA duplex with the sequence CCATGTAATTACCTGG (and its complement). Gel mobility shift experiments (in a buffer containing 100 mM KCl and 25 mM HEPES at pH 7.6) revealed that engrailed bound to this site with a K_D of 1–2 × 10^{-9} M. To estimate the intrinsic affinity for the sequences that overlap the ends of the 21 bp duplex, we permuted the sequence of the original duplex and changed some of the flanking bases. We used a site with the sequence CCGCCTAATTTTGCCA (and its complement), and gel mobility shift

experiments indicated that this site had a K_D of 1×10^{-7} M. (Base pairs 4–16 of this site correspond to base pairs 17–21 and 1–8 of the duplex used for crystallization.)

Acknowledgments

This project was supported by grants from the National Institutes of Health to C. O. P. (GM31471) and to T. B. K. (GM24635). We thank Anatoli Collector and Cynthia Wendling for synthesizing the DNA, Upul Obeysekare and Neil Clarke for helping to build a model of the engrailed structure from stereo photographs, and Mark Rould and Tom Steitz for their help with refinement of the heavy-atom parameters against solvent-flattened phases. We thank Phil Beachy, Roger Brent, Axel Brünger, Neil Clarke, Phil Sharp, and Cynthia Wolberger for helpful discussions. We thank Roger Brent, Steven Hanes, and Phil Sharp for communicating unpublished data that was central for deriving the correct alignment of the TAAT subsites. We gratefully acknowledge the National Cancer Institute for use of the Cray during molecular dynamics refinement. B. L., C. R. K., and C. O. P. are in the Howard Hughes Medical Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received August 31, 1990; revised September 24, 1990.

References

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.

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science *242*, 899–907.

Brünger, A. T. (1990). X-PLOR v2.1 Manual (New Haven, Connecticut: Yale University Press).

Brünger, A. T., Kuriyan, J., and Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. Science *235*, 458–460.

Burley, S. K., and Petsko, G. A. (1985). Aromatic–aromatic interactions: a mechanism of protein structure stabilization. Science *229*, 23–28.

Dickerson, R. E., and Drew, H. R. (1981). Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. J. Mol. Biol. *149*, 761–786.

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.

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G., and Horvitz, H. R. (1988). The POU domain: a large conserved region in the mammalian *pit*-1, *oct*-1, *oct*-2, and *Caenorhabditis elegans unc-86* gene products. Genes Dev. *2*, 1513–1516.

Ingraham, H. A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M. S., Wilson, L., and Rosenfeld, M. G. (1990). The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1–Pit-1 interactions. Cell *61*, 1021–1033.

Johnson, A. (1980). Mechanism of action of λ-cro protein. Ph.D. thesis, Harvard University, Cambridge, Massachusetts.

Jordan, S. R., and Pabo, C. O. (1988). Structure of the lambda complex at 2.5 Å resolution: details of the repressor–operator interactions. Science *242*, 893–899.

Jordan, S. R., Whitcombe, T. V., Berg, J. M., and Pabo, C. O. (1985). Systematic variation in DNA length yields highly ordered repressor–operator cocrystals. Science *230*, 1383–1385.

Kornberg, T. (1981). *engrailed*: a gene controlling compartment and segment formation in *Drosophila*. Proc. Natl. Acad. Sci. USA *78*, 1095–1099.

Laughon, A., and Scott, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature *310*, 25–31.

Liu, B., Kissinger, C., Pabo, C. O., Martin-Blanco, E., and Kornberg, T. B. (1990). Crystallization and preliminary diffraction studies of the engrailed homeodomain and of an engrailed homeodomain–DNA complex. Biochem. Biophys. Res. Commun. *171*, 257–259.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. Nature *335*, 321–329.

Pabo, C. O., and Sauer, R. T. (1984). Protein–DNA recognition. Annu. Rev. Biochem. *53*, 293–321.

Pabo, C. O., Aggarwal, A. K., Jordan, S. R., Beamer, L. J., Obeysekare, U. R., and Harrison, S. C. (1990). Conserved residues make similar contacts in two repressor–operator complexes. Science *247*, 1210–1213.

Poole, S. J., Kauvar, L. M., Drees, B., and Kornberg, T. (1985). The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell *40*, 37–43.

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.

Rossmann, M. G., and Argos, P. (1980). Three-dimensional coordinates from stereodiagrams of molecular structures. Acta Crystallogr. *B36*, 819–823.

Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. Science *246*, 1135–1142.

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

Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. USA *73*, 804–808.

Shepard, J. C., McGinnis, W., Carrasco, A. E., De Robertis, E. M., and Gehring, W. J. (1984). Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins. Nature *310*, 70–71.

Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I., and Dervan, P. B. (1990). Importance of minor-groove contacts for recognition of DNA by the binding domain of Hin recombinase. Biochemistry *29*, 6551–6561.

Staacke, D., Walter, B., Kisters-Woike, B., Wilcken-Bergmann, B., and Müller-Hill, B. (1990). How Trp repressor binds to its operator. EMBO J. *9*, 1963–1967.

Treisman, J., Gönczy, 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.

Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Meth. Enzymol. *115*, 90–112.

Wang, J. C. (1979). Helical repeat of DNA in solution. Proc. Natl. Acad. Sci. USA *76*, 200–203.

Availability of Coordinates

Coordinates are being deposited with the Brookhaven Data Bank. While these are being processed by Brookhaven and prepared for distribution, interested scientists may obtain a set of coordinates either by sending an appropriate BITNET address to us at PABO@JHUIGF or by sending a ½" tape with a mailing envelope and sufficient return postage.

Note Added in Proof

A preliminary NMR analysis of the Antp–DNA complex has been published: 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. Twelve intermolecular NOEs from six residues were used to dock the homeodomain against a model of B-DNA; each of these NOEs is consistent (using a cutoff of about 5 Å for the H–H distances) with the detailed model from our crystallographic study.