Specificity of Minor-Groove and Major-Groove Interactions in a Homeodomain−DNA Complex†

Sarah E. Ades and Robert T. Sauer*

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

Received July 14, 1995; Revised Manuscript Received August 30, 1995®

ABSTRACT: To assess the importance of minor-groove and major-groove interactions in homeodomain−DNA recognition, the binding properties of variants of the altered-specificity engrailed homeodomain, containing Lys50, and its DNA site TAATCC were determined. This homeodomain contacts bases in the minor groove of the DNA using Arg3 and Arg5 from its N-terminal arm and contacts bases in the major groove of the DNA using Ile47, Lys50, and Asn51 from its third α-helix. Mutation of Arg3 or Ile47 to alanine reduces binding affinity 10−20-fold while mutation of Arg5, Asn51, or Lys50 to alanine reduces binding affinity >100-fold, indicating that both minor-groove and major-groove interactions contribute to the overall binding energy. Binding site selections and affinity measurements show that the homeodomain can also discriminate among different base pairs in the minor groove and the major groove. However, the interactions between Lys50 of the recognition helix and the major-groove edges of base pairs 5 and 6 are more specific than interactions mediated by Arg3 and Arg5 in the N-terminal arm and the minor-groove edges of base pairs 1 and 2.

Specificity is one of the hallmarks of the binding of transcription factors to their DNA recognition sites. Interactions mediated by the protein must be capable of providing both high-affinity binding to the proper site and permitting discrimination against closely related DNA sites. The structures of a large number of protein−DNA complexes have now been solved, providing a detailed molecular view of the interactions with the DNA bases and sugar−phosphate backbone that stabilize the correct protein−DNA complex. In these complexes, the majority of interactions with the DNA bases are localized to the major groove. There are fewer examples of interactions with bases in the minor groove, and less is known about the importance of these interactions. Modeling studies have suggested that minor-groove interactions may be less specific than major-groove interactions because there are fewer ways to uniquely distinguish among the hydrogen-bond acceptors and donors on the edges of bases in the minor groove (Seeman et al., 1976).

The homeodomain provides a simple model system in which to study the relative contributions of minor-groove and major-groove interactions to the specificity and stability of a protein−DNA complex. In each of the known homeodomain−DNA structures, residues from α-helix 3 mediate a set of major-groove contacts, while residues from an extended N-terminal arm, which is unstructured in the absence of DNA, mediate a set minor-groove contacts (Kissinger et al., 1990; Otting et al., 1990; Woldberger et al., 1991; Billeter et al., 1993; Klemm et al., 1994). The overall architecture of the engrailed homeodomain−DNA complex is illustrated in Figure 1. In the cocrystal structure, residues from the N-terminal arm of engrailed contact the minor-groove edges of first two bases of the core sequence TAAT (Kissinger et al., 1990). These two base pairs are not contacted in the major groove, and yet are strongly conserved in binding site selections *in vitro* (Ades & Sauer, 1994). Moreover, the optimal binding sites of many homeodomains contain the same TAAT core sequence (Müller et al., 1988; Ekker et al., 1991, 1992; Florence et al., 1991; Regulski et al., 1991; Catron et al., 1993). These findings suggest that homeodomains can discriminate among potential binding sites on the basis of minor-groove interactions but do not provide a quantitative analysis of the specificity afforded by these interactions. Here, we use binding site selections and mutagenesis of a homeodomain and its DNA site to gain a better understanding of the role of minor-groove

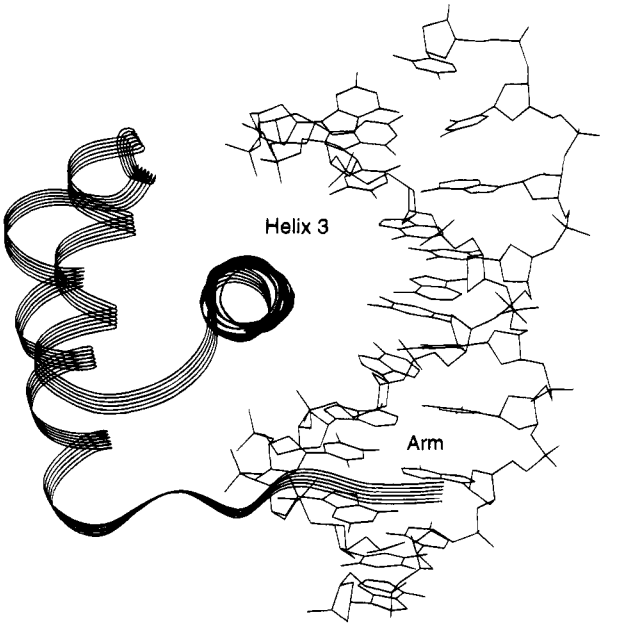

FIGURE 1: Molecular graphics representation of the engrailed homeodomain bound to DNA (Kissinger et al., 1990). The polypeptide backbone of the protein is represented by a ribbon. The view is down α-helix 3 which lies in the major groove. The N-terminal arm of the protein lies in the minor groove.

† Supported by NIH Grant AI-16892.
® Abstract published in *Advance ACS Abstracts,* October 15, 1995.

0006-2960/95/0434-14601$09.00/0

and major-groove interactions in determining binding energy and specificity. The wild-type engrailed homeodomain can bind in two symmetric orientations to its preferred site, TAATTA, which complicates mutational studies. In the studies presented here, we use an altered-specificity mutant of the engrailed homeodomain containing lysine at position 50 which binds in a unique orientation to the DNA site TAATCC (Ades & Sauer, 1994). We refer to this protein as the altered-specificity homeodomain. The cocrystal structure of the altered-specificity homeodomain complexed with the TAATCC site has recently been determined (Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T., & Pabo, C. O., manuscript in preparation).

MATERIALS AND METHODS

Oligonucleotides. Oligonucleotides were synthesized on an Applied Biosystems Model 381A DNA synthesizer and gel-purified by standard methods. The oligonucleotide 5′-cgcagtg**TAATCC**cctcgac-3′ and its complement, with an additional 5′ overhang for end-filling purposes, were synthesized for binding studies. The altered-specificity binding site is indicated in boldface type, and all binding site mutations studied were in this background. The sequence of the oligonucleotide used for binding site selections, N_2, is 5′-ccgcaggcaactcgagcttacgtcg**NNATCC**gctgcagtcatgctc-tccgtct-3′ (where N refers to an equimolar mixture of A, T, G, and C). Primers to N_2 were synthesized for PCR and second-strand synthesis.

Site-Directed Mutagenesis. All proteins used in this work are derivatives of the altered-specificity mutant of the engrailed homeodomain which contains Lys50 in place of the wild-type Gln50 residue (Ades & Sauer, 1994). Mutants encoding the RA3, RA5, IA47, and NA51 substitutions were constructed by cloning synthetic oligonucleotide cassettes encoding the substitutions between the appropriate restriction sites of plasmid pSEA100-QK50, which expresses the altered-specificity protein from the T7 promoter. The altered-specificity homeodomain and a variant containing Ala50 (KA50) were available from a previous study (Ades & Sauer, 1994).

Expression and Purification of Proteins. Proteins were purified from *Escherichia coli* strain BL21(DE3)/pLysS transformed with the appropriate derivatives of pSEA100-QK50 essentially as described (Ades & Sauer, 1994). Cells were grown in 400 mL of LB broth supplemented with 150 μg/mL ampicillin, and transcription from the T7 promoter was induced by the addition of IPTG to 0.4 mM. After cells were harvested by centrifugation, the cell pellet was resuspended in 15 mL of lysis buffer [100 mM Tris-HCl (pH 8.0), 200 mM KCl, 1 mM EDTA, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 2 mM NaN_3, and 50% glycerol], 10 μL of a fresh 100 mM solution of phenylmethanesulfonyl fluoride in ethanol was added to inhibit proteolysis, and cells were lysed by sonication. Nucleic acids were precipitated by the addition of 0.5% poly(ethylenimine), and proteins were precipitated from the resulting supernatant by the addition of solid ammonium sulfate to 95% saturation. The ammonium sulfate pellet was resuspended in column buffer [25 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, and 1.4 mM 2-mercaptoethanol] plus 100 mM NaCl, dialyzed extensively against the same buffer, and loaded onto a 5 mL DEAE Sephacel column. The flow-through fraction and first column volume of wash from the DEAE column were collected and loaded directly onto a 10 mL Affi-Gel Blue column (Bio-Rad; 100−200 mesh) equilibrated in column buffer plus 100 mM NaCl. The protein was eluted with successive washes of column buffer containing increasing concentrations of NaCl. The altered-specificity homeodomain and IA47 and NA51 variants eluted in the 0.7−0.8 M NaCl washes while the RA3 and RA5 variants eluted slightly earlier, in the 0.6−0.7 M NaCl washes, consistent with the removal of a positively charged arginine side chain. Each of these proteins binds tightly to the Affi-Gel Blue resin and is ~99% pure upon elution as judged by Coomassie Blue staining of Tris−tricine polyacrylamide gels (Schagger & von Jagow, 1987). The fractions containing pure protein were dialyzed into column buffer plus 100 mM NaCl, concentrated by ultrafiltration, and stored at 4 °C.

Circular dichroism experiments were conducted to monitor the folding and stability of the homeodomains. Spectra of samples containing protein at 3 μM in 50 mM potassium phosphate (pH 7.0) and 100 mM KCl were obtained by averaging five scans, each collected at 20 °C in 1 nM steps with a 1 s averaging time. The thermal stabilities of the proteins were determined by measuring the ellipticity at 222 nM at 1 °C intervals from 15 to 90 °C with a 1 min equilibration time and 30 s averaging time. The CD spectra (data not shown) and the thermal stabilities of the purified mutants were very similar to those of the altered-specificity homeodomain, suggesting that none of the mutations affects the overall fold or stability of the homeodomain. Fitting the denaturation curves using nonlinear least-squares methods yields the following values: altered-specificity homeodomain, t_m = 53.8 °C, ΔH = 34.3 kcal/mol; RA3, t_m = 55.2 °C, ΔH = 33.3 kcal/mol; RA5, t_m = 54.9 °C, ΔH = 32.4 kcal/mol; IA47, t_m = 56.6 °C, ΔH = 33.4 kcal/mol; KA50, t_m = 56.8 °C, ΔH = 39.0 kcal/mol; and NA51, t_m = 50.1 °C, ΔH = 25.6 kcal/mol. All of the proteins are greater than 99% folded at 20 °C, the temperature at which DNA binding affinities were measured.

Equilibrium and Kinetic Assays of DNA Binding. When necessary, double-stranded binding site oligonucleotides used in gel mobility shift assays were labeled by end-filling in a reaction containing 1 pmol of DNA in sequenase reaction buffer, 1 unit of sequenase v2.0 (United States Biochemicals), and 30 μCi of [α-^{32}P]dATP (6000 Ci/mmol) for 30−60 min at room temperature. The reactions were extracted with phenol/chloroform (1:1), and unincorporated nucleotides were removed using a G-25 Sephadex quick-spin column (Boehringer Mannheim). Equilibrium and kinetic constants were determined using gel mobility shift assays performed at 20 °C in binding buffer containing 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 50 mM NaCl, 0.02% NP-40, 50 μg/mL bovine serum albumin, and 5% glycerol.

Equilibrium gel mobility shift assays were conducted as previously described (Ades & Sauer, 1994). Briefly, varying concentrations of protein were incubated with radiolabeled DNA fragments (≤5 pM) for 2 h in a 50 μL reaction, and then 30 μL was loaded onto 0.5× TBE, 10% polyacrylamide gels (prerun for >30 min at 300 V) running at 300 V. The voltage was reduced to 150 V after the samples had entered the gel. Tracking dyes were loaded in the outer lanes of the

Table 1: Equilibrium and Kinetic DNA Binding Constants for Alanine Substitution Mutants

amino acid substitution[a]	DNA contact	K_d (M)	rel affinity ($K_{d\text{-mut}}/K_{d\text{-wt}}$)[b]	rel k_{off} ($k_{off\text{-mut}}/k_{off\text{-wt}}$)[b]
RA3 (arm)	minor groove position 2	$8.8 (\pm 4.3) \times 10^{-11}$	10	20
RA5 (arm)	minor groove position 1	nd[d]	>100	nd
IA47 (helix 3)	major groove position 4	$1.9 (\pm 0.7) \times 10^{-10}$	20	17
KA50 (helix 3)	major groove positions 5 and 6	nd[c]	>100	nd
NA51 (helix 3)	major groove position 3	nd	>100	nd

[a] As noted in the text, all mutations are in the altered-specificity (Lys50) background of the engrailed homeodomain. [b] $K_{d\text{-wt}} = 8.9 (\pm 4.0) \times 10^{-12}$ M and $k_{off\text{-wt}} = 0.003$ s^{-1}, the equilibrium and kinetic constants for altered-specificity homeodomain binding to the TAATCC site at 20 °C in binding buffer. [c] The affinity of the KA50 mutant for the TAATCC binding site was measured in a previous study using a different method and found to be $3.4 (\pm 2.6) \times 10^{-9}$ M (Ades & Sauer, 1994). [d] nd: It was not possible to determine the equilibrium and kinetic constants for these proteins because they do not give a stable gel shift.

gel and were not included in the samples. After electrophoresis, gels were dried and exposed to film at −70 °C with an intensifying screen. Binding assays were quantified by scanning densitometry, and the loss of the free band was used to determine the fraction of bound DNA. Equilibrium dissociation constants were determined by linear regression using the Scatchard equation. Three or more gel mobility shift assays were conducted for each binding constant determined.

Equilibrium constants for dissociation of the altered-specificity homeodomain protein from variant binding sites were determined using a competition gel mobility shift assay. Sufficient protein was added to bind 80−90% of a radiolabeled TAATCC fragment at a concentration of 1 pM. Aliquots of this mixture were added to tubes containing 0.02−250 nM competitor DNAs containing the variant binding sites. After equilibration, samples were loaded onto gels and electrophoresed as described above. The free bands were quantified and equilibrium dissociation constants for the competitor DNA fragments were calculated as described in Ades and Sauer (1994). Again, three or more assays were conducted for each binding constant determined.

Dissociation rates were measured by assaying the increase in free radiolabeled DNA as a function of time after the addition of unlabeled competitor DNA. Sufficient protein was equilibrated with radiolabeled binding site oligonucleotides to bind 80−90% of the DNA. An excess of unlabeled competitor DNA was added, and aliquots were loaded at the appropriate times onto a 0.5× TBE, 10% polyacrylamide gel running at 300 V. Gels were electrophoresed and processed as described above. Dissociation rate constants were determined by fitting the data to a first-order rate equation.

Binding Site Selections. The base preferences of the altered-specificity homeodomain and the RA3 and RA5 variants at the first two positions of the DNA binding site were determined in selections using the N_2 oligonucleotide which contains the sequence NNATCC (where N represents an equal mixture of A, T, G, and C). The starting pool of DNA was generated by annealing a primer to N_2 and extending with sequenase v2.0 in the presence of unlabeled nucleotides and a small amount of [α-^{32}P]dATP. In the first round of selection, protein at several different concentrations, from 0.001 nM to 1 μM depending on the variant, was incubated with roughly 0.5 nM randomized DNA in a 50 μL reaction. Bound DNA was separated from free DNA using a gel mobility shift assay as described above. The bound DNA was eluted from dried gels from the lane containing the lowest concentration of protein for which a bound band was visible: 0.1 nM for the altered-specificity homeodomain, 1 nM for the RA3 variant, and 100 nM for the RA5 variant. The eluted DNA was then amplified by the polymerase chain reaction using a ^{32}P end-labeled primer and subjected to three more rounds of selection and amplification. In these latter rounds of selection with the altered-specificity homeodomain and RA3 variant, an excess of DNA over protein was used in the binding reactions, 0.05−0.07 nM altered-specificity homeodomain and 0.3−0.6 nM RA3 protein were equilibrated with roughly 1−2 nM amplified DNA. In the remaining rounds of selection with the RA5 variant, roughly equimolar quantities of protein and DNA (~10 nM) were used in binding reactions. After four rounds of selection and amplification, the selected pools of binding sites were cloned between the *Xho*I and *Pst*I sites of pBluescript/KS+ (Stratagene), and individual clones were sequenced.

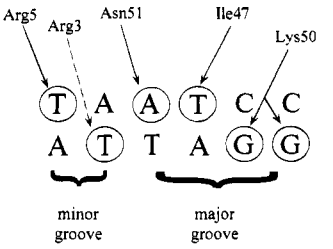

FIGURE 2: Contacts between the altered-specificity homeodomain and DNA. Solid lines indicate contacts from the cocrystal structure (Tucker-Kellogg et al., in preparation). Dashed lines indicate structurally plausible contacts inferred from the mutational studies in this paper. Contacts with the major-groove and minor-groove edges of base pairs are indicated.

RESULTS

Contributions of Minor-Groove and Major-Groove Contacts Probed by Alanine Mutations. To determine the contribution of side chain−base contacts to the overall DNA binding energy of the altered-specificity homeodomain, we constructed alanine substitution mutants for each residue involved in a base contact in the cocrystal structure (Figure 2; Kissinger et al., 1990; Tucker-Kellogg et al., in preparation). The five mutant proteins (RA3, RA5, IA47, KA50, and NA51) were purified, and the binding of each variant to the TAATCC site was probed by gel mobility shift assays. As shown in Table 1, the affinities of the RA3 and IA47 proteins for the TAATCC site were reduced by roughly 10−20-fold and the dissociation rates of the protein−DNA complexes were increased by roughly 20-fold. The RA5, KA50, and NA51 mutations reduced binding to the point where stable, quantifiable gel shifts with the TAATCC site

Table 2: Tabulation of Data from Binding Site Selections Using NNATCC[a]

	altered-specificity homeodomain		RA3 mutant		RA5 mutant		unselected	
base	position 1	position 2	position 1	position 2	position 1	position 2	position 1	position 2
G	1	2	2	9	8	9	10	7
A	5	28	13	22	5	8	7	5
T	25	1	21	4	7	2	8	11
C	0	0	0	1	0	1	5	7
total	31	31	36	36	20	20	30	30

[a] The recovery of individual bases at positions 1 and 2 in sites after binding site selections using the altered-specificity homeodomain, the RA3 variant of the altered-specificity homeodomain, and the RA5 variant of the altered-specificity homeodomain is tabulated. The last column shows the recovery of individual bases at positions 1 and 2 in sites from the starting pool of oligonucleotides which were not subjected to binding site selection. The totals refer to the number of binding sites sequenced.

were not observed. On the basis of the faint gel shifts that were observed, it appears that the affinities of these mutant proteins for the TAATCC site are reduced by at least 2 orders of magnitude. The affinity of the KA50 mutant for the TAATCC site was measured in a previous study by a competition method and found to be reduced by roughly 400-fold (Ades & Sauer, 1994).

Binding Site Selections To Probe the Base Preferences at Positions 1 and 2. On the basis of the cocrystal structures (Kissinger et al., 1990; Tucker-Kellogg et al., in preparation), the minor-groove edges of the first two base pairs of the TAATCC binding site are expected to be contacted by Arg3 and Arg5 from the homeodomain's N-terminal arm (Figures 1 and 2). To assess the base preferences at these sites of minor-groove interactions, binding site selections were conducted with the altered-specificity homeodomain, the RA3 mutant, and the RA5 mutant using a population of binding sites in which the first two positions were randomized (NNATCC). Although the RA5 protein does not give a shifted band that is sufficiently stable for quantification, it does give a faint, shifted band at high concentrations of protein. After four rounds of selection and amplification, the pools of DNA enriched for tightly binding sequences were cloned and sequenced. The results are shown in Table 2. The altered-specificity homeodomain shows a marked preference for the expected bases, T (80%) at position 1 and A (90%) at position 2. The binding site preferences of the RA3 mutant are broader. At the first position, T (58%) is still the preferred base, but there is a secondary preference for A (36%). At the second position, A (60%) is the preferred base with weaker preferences for G (25%) and T (11%). For the RA5 mutant, the significant preferences seem to be against C at position 1 and against C and T at position 2. It is also notable that C:G base pairs were rarely recovered at either position in any of the three selections, even though sequencing of randomized but unselected oligonucleotides showed that C:G base pairs were present at reasonable frequencies in the starting pool (Table 2).

Affinity for Binding Sites with Substitutions at Positions 1 and 2. To evaluate the ability of the altered-specificity homeodomain to discriminate among binding sites in a more quantitative fashion, equilibrium dissociation constants for a set of binding sites containing natural base pair substitutions at positions 1 and 2 were determined (Table 3). The affinities for binding sites with C:I and I:C substitutions were also measured to help distinguish between minor-groove and major-groove effects. Inosine lacks the exocyclic N2 amino group of guanine, and thus a C:I base pair resembles a T:A base pair in the minor groove and a C:G base pair in the major groove (Figure 3). The equilibrium dissociation

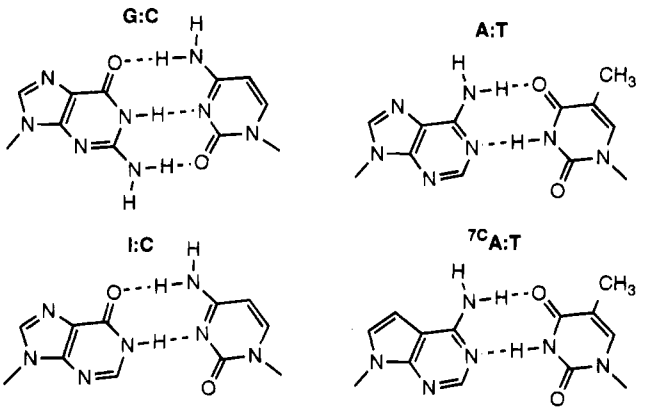

FIGURE 3: Base pairs used for binding site substitutions in these studies. ^{7C}A is N7-deazaadenine.

constants for each of the altered sites were determined using an assay in which an unlabeled variant site competed for binding of the altered-specificity homeodomain to a labeled TAATCC site.

(A) Position 1. Compared to the preferred TAATCC site, the altered-specificity homeodomain shows modestly reduced affinity for each of the position 1 variants tested, including those with I:C or C:I (Table 3). The largest loss of affinity, about 6-fold, occurs when the wild-type T:A base pair is replaced by a C:G base pair. The C:I substitution, which differs from the C:G base pair only in the minor groove, reduces affinity 3−4-fold. Sites bearing the A:T, G:C, or I:C transversion substitutions also have affinities reduced by approximately 3−4-fold. In the cocrystal structure, the side chain of Arg5 contacts the base pair at position 1 (Figure 2). It is important to note that the reductions in affinity caused by the base-substitution mutations at position 1 are small when compared with the greater than 100-fold reduction caused by the RA5 mutation.

(B) Position 2. The affinities of the altered-specificity homeodomain for sites with substitutions at position 2 (TAATCC) fall into two classes (Table 3). Transition mutations (A:T to G:C or I:C) have little effect on affinity (reduced 2−3-fold for G:C; unchanged for I:C). Transversion mutations (A:T to T:A, C:G, or C:I) have larger effects ranging from 7- to 28-fold. Surprisingly, affinity is reduced 7-fold for the T:A mutation but reduced 21-fold for the C:I mutation. Since both base pairs have identical functional groups in the minor groove, it seems likely that interactions mediated by the major-groove edge of the base pairs must be responsible for the observed difference. Because DNA sites with a purine on the top strand at position 2 have the highest affinities for the altered-specificity homeodomain, we reasoned that the N7 position of the purine might be

Table 3: Equilibrium DNA Binding Constants to Altered Sites

binding site	altered-specificity homeodomain			RA3 mutant		
	K_d (pM)	rel affinity[a]	$\Delta\Delta G$ (kcal/mol)	K_d (pM)	rel affinity[a]	$\Delta\Delta G$ (kcal/mol)
T A A T C C A T T A G G	8.9 (±4)	1.0	-	88 (±43)	1.0	0.0
C - - - - - - G - - - - - -	56 (±29)	6.3	1.1			
C - - - - - - I - - - - - -	31 (±15)	3.5	0.7			
A - - - - - - T - - - - - -	39 (±15)	4.4	0.9	150 (±50)	1.7	0.3
G - - - - - - C - - - - - -	31 (±15)	3.5	0.7			
I - - - - - - C - - - - - -	25 (±12)	2.8	0.6			
- G - - - - - - C - - - - -	22 (±9)	2.5	0.5			
- I - - - - - - C - - - - -	13 (±6)	1.5	0.2			
- T - - - - - - A - - - - -	60 (±18)	6.7	1.1	250 (±100)	2.8	0.6
- C - - - - - - G - - - - -	250 (±70)	28.1	1.9	1100 (±500)	12.5	1.5
- C - - - - - - I - - - - -	190 (±90)	21.3	1.8	1600 (±500)	18.2	1.7
- ^{7C}A - - - - - - T - - - - -	45 (±14)	5.1	0.9			
C C - - - - - G G - - - - -	570 (±240)	64.0	2.4			
A T - - - - - T A - - - - -	60 (±23)	6.7	1.1	170 (±60)	1.9	0.4
- - ^{7C}A - - - - - T - - -	960 (±340)	107.9	2.7			
- - - - T - - - - - A -	120 (±43)	13.5	1.5			
- - - - A - - - - - T -	92 (±46)	10.3	1.4			
- - - - G - - - - - C -	120 (±49)	13.5	1.5			
- - - - - T - - - - - A	83 (±39)	9.3	1.3			
- - - - - A - - - - - T	120 (±60)	13.5	1.5			
- - - - - G - - - - - C	180 (±50)	20.2	1.7			
- - - - T A - - - - A T	320 (±160)	36.0	2.1			

[a] Relative affinity ($K_{d\text{-site}}/K_{d\text{-TAATCC}}$).

important for binding. To test this idea, we determined the affinity for a binding site with a ^{7C}A:T base pair (where ^{7C}A represents N7-deazaadenine; see Figure 3) at position 2. The affinity for this ^{7C}A:T site was reduced 5-fold, consistent with the idea that the N7 position in the major groove does influence binding affinity in some fashion. To provide a comparison, we also synthesized a DNA site with a ^{7C}A:T base pair replacing the normal A:T base pair at position 3. The side chain of Asn51 forms a bidentate hydrogen bond to the N7 and N6 positions of this adenine in the protein–DNA complex. When adenine 3 is changed to N7-deazaadenine, the affinity of the altered-specificity homeodomain is reduced approximately 100-fold (Table 3).

Binding of RA3 to Sites Altered at Position 2. In principle, the RA3 mutation should remove interactions with the minor-groove edge of base pair 2 but should not directly affect any major-groove interactions. Hence, the results described above suggest that the RA3 mutant should retain some sensitivity to position 2 alterations to the extent that the effects of these alterations are mediated through the major groove. To test this, the affinity of the RA3 mutant was determined for several position 2 mutants (Table 3). In general, the relative affinity of RA3 for each mutant DNA site is reduced compared with the relative affinity of the parent protein for that site. However, the RA3 protein still binds more strongly to the TAATCC site than to any of the position 2 variant sites. These two findings are consistent with the idea that the base pair at position 2 affects affinity both via minor-groove interactions mediated by Arg3 and through major-groove interactions.

Effects of Base Pair Substitutions at Positions 5 and 6. As mentioned above, most homeodomain proteins, including

engrailed, bind to sites containing the conserved core sequence TAAT. The differential specificity of homeodomains, however, is frequently determined by the identity of residue 50 in the protein and the identities of positions 5 and 6 in the DNA site (Hanes & Brent, 1989, 1991; Treisman et al., 1989; Percival-Smith et al., 1990). To evaluate the contributions of base pairs 5 and 6 to the affinity and specificity of binding by the altered-specificity homeodomain, we measured the affinity of the protein for binding sites with all natural, single base pair substitutions at these positions (Table 3). At position 5, each base pair substitution reduces binding 10−13-fold. At position 6, each substitution reduces binding 9−20-fold.

Interactions between DNA Positions. In the binding site selections, specificity is broadened at both the first and second positions when either Arg3 (which is thought to contact position 2) or Arg5 (which contacts position 1) is changed to alanine (Table 2). This finding suggests that interactions with the first two base pairs of the binding site may be coupled. If these interactions are coupled, then the effects of mutations at positions 1 and 2 of the DNA site should not be additive in terms of binding energies. To test this, the affinities of the altered-specificity homeodomain for the ATATCC and CCATCC sites were measured (Table 3). Binding to the ATATCC site is reduced by about 1.1 (±0.3) kcal/mol, whereas a 2.0 (±0.5) kcal/mol reduction would be expected if the effects of each mutation were independent. Binding to the CCATCC site is reduced by 2.4 (±0.4) kcal/mol, a value within error of the 3.0 (±0.5) kcal/mol reduction expected on the basis of independent mutant effects. Hence, interactions of the homeodomain with base pairs 1 and 2 seem to be energetically coupled for binding to some sites but not others.

The preferred binding site of the wild-type engrailed homeodomain (TAATTA) differs from the preferred site for the altered-specificity protein (TAATCC) at both positions 5 and 6. The affinity of the altered-specificity protein for the TAATTA site is reduced 2.1 (±0.4) kcal/mol (Table 3; Ades & Sauer, 1994) compared to the preferred site, whereas a reduction of 3.0 (±0.5) kcal/mol would be expected if the base-substitution effects were independent. This result suggests a small energetic coupling between interactions at base pairs 5 and 6.

DISCUSSION

Cocrystal structures of protein−DNA complexes provide a three-dimensional map of molecular interactions, while biochemical studies provide a way to address the importance of interactions to binding affinity and specificity. Structures have been solved for both the engrailed homeodomain/TAATTA complex (Kissinger et al., 1990) and the engrailed altered-specificity/TAATCC complex (Tucker-Kellogg et al., in preparation), providing a basis for interpreting the functional studies presented here. In particular, we have probed the contribution to affinity and specificity of two parts of the altered-specificity complex: interactions between the protein's N-terminal arm and the minor groove and interactions of the lysine at position 50 with bases in the major groove.

Several issues need to be considered in evaluating the results presented here. First, in considering the effects of mutations at base positions 1 and 2 with those at 5 and 6, minor-groove interactions are being compared to major-groove interactions, but interactions from a flexible region of protein are also being compared with those from a relatively rigid unit of secondary structure. Second, in addition to perturbing the expected base contacts, a mutation may perturb backbone or base contacts at other positions via effects on the overall DNA or protein structure. Finally, a base change introduces new functional groups which may permit new interactions with the protein. Structures of each mutant complex would be needed to know with certainty whether significant conformational changes occur or new contacts are made. Nevertheless, comparing the observed functional effects of mutations with the simplest expectations based upon the known protein−DNA structures is still worthwhile. In cases where this fails to provide a satisfactory explanation, more complex mechanisms probably contribute to the observed effects, and structural studies are indicated.

Three side chains of the altered-specificity homeodomain (Ile47, Lys50, Asn51) make base contacts in the major groove, and two side chains (Arg3, Arg5) make base contacts in the minor groove. As measured by the effects of alanine substitution mutations, each of these side chains contributes to binding affinity, albeit at different levels. Mutations of Arg3 or Ile47 reduce affinity modestly (1.3−1.7 kcal/mol), while mutations of Arg5, Asn51, or Lys50 have larger effects (>2.7 kcal/mol). At a general level, these results show that the overall energetic contributions of the minor-groove interactions made by the flexible N-terminal arm are comparable to those of the major-groove interactions made by the recognition helix.

To compare the contributions to binding specificity of the base pairs at positions 1, 2, 5, and 6 of the binding site, we calculated the specificity index (I_{spec}) as defined by Stormo et al. (1991). For each position, I_{spec} is calculated from the relative affinities of the three mutant sites with natural base substitutions (Table 3) and ranges from 0 bits of information (no specificity) to 2 bits of information (maximum specificity). For the altered-specificity homeodomain, the results are as follows: position 1, $I_{spec} = 0.41$; position 2, $I_{spec} = 0.63$; position 5, $I_{spec} = 0.97$; and position 6, $I_{spec} = 1.02$. By this measure, base pairs 5 and 6 of the binding site have a higher information content, *i.e.*, greater specificity, than base pairs 1 and 2. Hence, the major-groove interactions mediated by residue 50 of the recognition helix are more specific than those formed in the minor groove by the N-terminal arm.

Several general points are worth noting with respect to specificity and affinity. First, the minor-groove interactions made by the flexible N-terminal arm do contribute to binding specificity, even if the effect is modest. The modeling studies of Seeman et al. (1976) correctly suggested that minor-groove interactions would have lower specificity than major-groove interactions but also indicated that proteins would not be able to differentiate between T:A and A:T base pairs in the minor groove. However, the altered-specificity homeodomain differentiates between T:A and A:T at position 1 as well as it differentiates between other base substitutions. Finally, there is no simple correlation between the affinities suggested for particular interactions by the alanine mutations and the specificities inferred for these contacts. For example, Arg5 and Lys50 contribute approximately equally to affinity, but the interactions mediated by Arg5 show significantly lower specificity than those mediated by Lys50. We assume

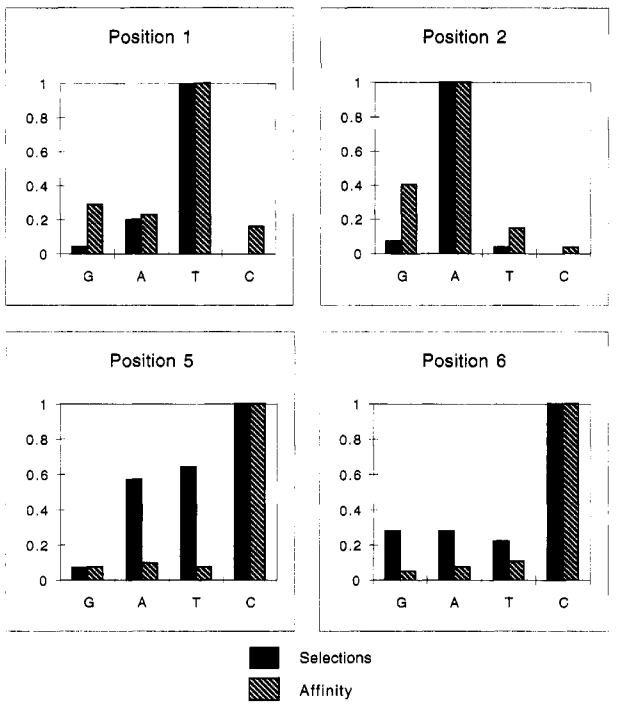

FIGURE 4: Comparison of specificity inferred from binding site selections (solid bars) and affinity measurements (striped bars) at positions 1, 2, 5, and 6. All measurements are normalized to the preferred base at that position: (frequency of base)/(frequency of preferred base) for binding site selections and $K_{\text{d-preferred site}}/K_{\text{d-mutant site}}$ for affinity measurements. Binding site selection experiments for positions 5 and 6 are from Ades and Sauer (1994).

that this occurs because Arg5, to a greater extent than Lys50, is able to make alternative contacts with either the mutant bases or the sugar−phosphate backbone of the mutant DNA. Both the arginine and lysine side chains should have comparable flexibility, but contacts from the N-terminal arm are presumably more easily rearranged than those from the recognition α-helix.

Binding site selection experiments provide an additional probe of binding specificity which can be compared to results from affinity measurements. Selections for positions 1 and 2 were performed here, and selections for positions 5 and 6 were described previously (Ades & Sauer, 1994). As shown in Figure 4, although both methods identify the same preferred bases (T:A at position 1, A:T at position 2, C:G at positions 5 and 6), the binding site selections can overestimate or underestimate the degree of specificity. This is not surprising. First, because many rounds of site selection and amplification are performed, there is no reason that the results should be strictly proportional to thermodynamic stability. Second, the binding site selections were performed following randomization of several base pairs. If there are cooperative interactions between base positions (as appears to be the case both for positions 1 and 2 and positions 5 and 6; see below), this will affect the selections in a manner not mirrored by single-site affinity studies. Although binding site selections do not provide direct thermodynamic information about specificity, they do provide a rapid means of determining rough base preferences which can then be examined more systematically.

The structural and mutational analyses of the interaction between Arg3 and base pair 2 provide examples of some of the complexities that can emerge in such studies. In the cocrystal structures, Arg3 of the wild-type engrailed homeodomain is positioned to contact the minor-groove face of base pair 2 (Kissinger et al., 1990) but Arg3 in the altered-specificity homeodomain is poorly ordered (Tucker-Kellogg et al., in preparation). Nevertheless, our mutational studies indicate that Arg3 does contribute to binding affinity of the altered-specificity homeodomain and show that interactions mediated by base pair 2 contribute to binding specificity. However, transition mutations at this position have a much smaller effect than transversion mutations, indicating that purines are favored on the sense strand of the binding site by the altered-specificity homeodomain. Several lines of evidence suggest that this effect is mediated, at least in part, through the major groove of the DNA. First, substitution of the preferred A:T base pair with C:I reduces binding to a greater extent than with T:A, even though both C:I and T:A have similar functional groups in the minor groove. Second, substitution of A:T with ^{7C}A:T also reduces affinity, even though this substitution only affects the major groove. Third, the RA3 mutant, which should no longer interact with the minor groove at position 2, is still sensitive to base substitutions at this position. In the altered-specificity complex, there are no contacts between side chains and base pair 2 in the major groove; there is an ordered water close to the N7 of adenine 2, but this does not appear to mediate a protein contact with the DNA (Tucker-Kellogg et al., in preparation). We cannot exclude the possibility that the latter interaction contributes directly to specificity, but it seems more likely that transversions at position 2 affect the structure of the DNA to some extent and thereby perturb other contacts in the complex.

The nonadditivity of some mutational effects suggests that several interactions between the homeodomain and DNA are energetically coupled. Coupling is most simply explained by entropic considerations when interactions help to stabilize each other and is commonly observed when the mutations probe functional groups that are close or interact directly in the structure (Wells, 1990). At positions 1 and 2, both the results of binding site selections and the nonadditivity of mutations (for the ATATCC site, in particular) suggest linkage. The Arg3 and Arg5 side chains do not appear to interact with each other in the cocrystal structure, but interactions of these amino acids with the DNA could serve to fix the position of the otherwise flexible arm and help position the second amino acid for its contact. At positions 5 and 6, linkage can be explained in a simple fashion since the ϵ-amino group of Lys50 is positioned to form hydrogen bonds with both bases.

The general conclusions presented here from experiments with the altered-specificity engrailed homeodomain are likely to apply to the wild-type (Gln50) engrailed homeodomain and to other homeodomains which use arginine at positions 3 and 5 of the N-terminal arm to contact the minor groove. Binding data available for such homeodomains indicate that they can distinguish among bases in the minor groove and that the contacts mediated by the N-terminal arm are important for binding affinity and specificity (Percival-Smith et al., 1990; Ekker et al., 1991, 1992, 1994; Florence et al., 1991). However, it would not be surprising if there were subtle differences in the ways in which different homeodomains utilize the N-terminal arm for DNA recognition.

ACKNOWLEDGMENT

We thank Kristen Chambers, Carl Pabo, Mark Rould, and Lisa Tucker-Kellogg for helpful discussions, advice, and communication of unpublished structural information.

REFERENCES

Ades, S. E., & Sauer, R. T. (1994) *Biochemistry 33*, 9187−9194.
Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1993) *J. Mol. Biol. 234*, 1084−1097.
Catron, K. M., Iler, N., & Abate, C. (1993) *Mol. Cell. Biol. 13*, 2354−2365.
Ekker, S. C., Young, K. E., von Kessler, D. P., & Beachy, P. A. (1991) *EMBO J. 10*, 1179−1186.
Ekker, S. C., von Kessler, D. P., & Beachy, P. A. (1992) *EMBO J. 11*, 4059−4072.
Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E., & Beachy, P. A. (1994) *EMBO J. 13*, 3551−3560.
Florence, B., Handrow, R., & Laughon, A. (1991) *Mol. Cell. Biol. 11*, 3613−3623.
Hanes, S. D., & Brent, R. (1989) *Cell 57*, 1275−1283.
Hanes, S. D., & Brent, R. (1991) *Science 251*, 426−430.
Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579−590.
Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., & Pabo, C. O. (1994) *Cell 77*, 21−32.
Müller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K., & Gehring, W. J. (1988) *EMBO J. 7*, 4299−4304.
Percival-Smith, A., Müller, M., Affolter, M., & Gehring, W. J. (1990) *EMBO J. 9*, 3967−3974.
Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J., & Wüthrich, K. (1990) *EMBO J. 9*, 3085−3092.
Regulski, M., Dessain, S., McGinnis, N., & McGinnis, W. (1991) *Genes Dev. 5*, 278−286.
Schagger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368−379.
Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804−808.
Stormo, G. D., & Yoshioka, M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5699−5703.
Treisman, J., Gönczy, P., Vashishtha, M., Harris, E., & Desplan, C. (1989) *Cell 59*, 553−562.
Wells, J. A. (1990) *Biochemistry 29*, 8509−8517.
Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., & Pabo, C. O. (1991) *Cell 67*, 517−528.

BI951615P