# Differences in Water Release with DNA Binding by Ultrabithorax and Deformed Homeodomains†

Likun Li and Kathleen S. Matthews*

*Department of Biochemistry & Cell Biology, Rice University, Houston, Texas 77251*

*Received December 3, 1996; Revised Manuscript Received March 27, 1997*⊗

ABSTRACT: The amino acid sequences of the homeodomains (HD) within the Ultrabithorax (Ubx) and Deformed (Dfd) proteins from *Drosophila melanogaster* are highly conserved despite distinct genetic regulatory functions for these proteins in embryonic development. We reported recently that Ubx-HD binding to a single target site displayed significantly increased affinity and greater salt concentration dependence at lower pH; in contrast, Dfd-HD did not show pH dependence in its DNA binding properties [Li, L., *et al.* (1996) *Biochemistry 35,* 9832−9839]. We demonstrate in this study that water activity differentially affects Ubx-HD and Dfd-HD DNA binding affinity. The sensitivity of the protein−DNA binding constant to osmotic pressures generated by neutral solutes was measured, and the formation of the Ubx-HD−DNA complex is associated with significantly greater water release than that of the Dfd-HD−DNA complex. No influence of pH on water release was detected for either HD. Experiments with chimeric Ubx−Dfd homeodomains demonstrated that the C-terminal region of the Ubx-HD is the primary determinant for the greater water release associated with DNA binding for this protein. DNA sequences do not exert a significant effect on the magnitude of water release associated with protein−DNA binding for Ubx-HD and the chimeric HD, UDU.

Homeoproteins, found widely in insects and vertebrates, direct body plan organization during embryonic development by regulating the expression of target genes both spatially and temporally (Akam, 1987; Duncan, 1987; Scott *et al.*, 1989; Kaufman *et al.*, 1990; McGinnis & Krumlauf, 1992; Gehring *at al*., 1994a,b). The homeodomain, an evolutionarily conserved sequence found in homeoproteins, contains a fold similar to the helix−turn−helix motif in prokaryotic DNA binding proteins and mediates sequence-specific DNA recognition (McGinnis *et al.*, 1984a,b; Scott & Weiner, 1984; Scott *et al.*, 1989; Hayashi & Scott, 1990). Despite significant efforts to understand gene regulation by homeoproteins, the mechanism by which specificity of DNA recognition is achieved remains poorly defined. The central difficulty arises from the ability of diverse homeodomains to bind the same target sequence with only slightly differing affinities (Desplan *et al.*, 1988; Hoey & Levine, 1988; Ekker *et al.*, 1991, 1992, 1994).

Ultrabithorax (Ubx)[1] and Deformed (Dfd) are two homeoproteins from *Drosophila melanogaster*. Ubx specifies the identities of parasegments 5 and 6, comprising the posterior thorax and part of the first abdominal segment, while Dfd specifies mandibular and maxillary segment identities (Akam, 1987; Duncan, 1987; Kaufman *et al.*, 1990; McGinnis & Krumlauf, 1992). Despite these distinct functions, homeodomains from these two proteins display only small differences in sequence-dependent DNA binding *in vitro* (Ekker *et al.*, 1992, 1994), although distinct DNA contacts were revealed by hydroxyl radical footprinting and missing nucleoside experiments (Draganescu *et al.*, 1995). To explain how homeodomain proteins with marginal differences in DNA binding specificity might perform their distinct function *in vivo*, a multisite integrative model of homeotic protein action has been proposed on the basis of the occurrence of multiple clusters of individual binding site sequences and cooperative DNA binding by Ubx protein (Beachy *et al.*, 1993). According to this model, small differences in binding of homeotic proteins to individual sites can be summed to yield large overall differences in binding to multiple sites. Furthermore, cooperative DNA binding and increased sequence selectivity of homeodomain proteins might be generated through association with protein cofactors, illustrated by the Hox−Pbx interaction [Chang *et al.*, 1996; reviewed in Mann and Chan (1996)]. Recently, we have found that changes in the physical environment can significantly and selectively enhance Ubx-HD−DNA binding *in vitro*, suggesting other potential mechanisms for generating site discrimination *in vivo* (Li *et al.*, 1996). In this study, we have explored the role of water in DNA binding by both Ubx and Dfd homeodomains.

Bound water is involved in multiple types of macromolecular interactions, including ligand binding (Colombo *et al.*, 1992; Rand *et al.*, 1993; Sidorova & Rau, 1995), electron transfer (Kornblatt & Hui Bon Hoa, 1990), and protein−protein interactions (Kornblatt *et al.*, 1993). Typically, protein and DNA in solution are surrounded by a "hydration shell", which excludes many solutes. When association between protein and DNA takes place, a portion of water bound on each of their surfaces is displaced by the partner and released into the bulk solution. The role of water in mediating specificity of protein−DNA interactions has been

† This work was supported by grants to K.S.M. from the NIH (GM22441) and the Robert A. Welch Foundation (C-576).

* Corresponding author. Phone: 713-527-4871. Fax: 713-285-5154. E-mail: ksm@bioc.rice.edu.

⊗ Abstract published in *Advance ACS Abstracts,* June 1, 1997.

[1] Abbreviations: Ubx, Ultrabithorax; Dfd, Deformed; HD, homeodomain; TEG, triethylene glycol; TBE, Tris−boric acid−ethylenediaminetetraacetic acid (EDTA); DTT, dithiothreitol; BSA, bovine serum albumin; TBB, Tris binding buffer; PBB, phosphate binding buffer.

S0006-2960(96)02966-2 CCC: $14.00 

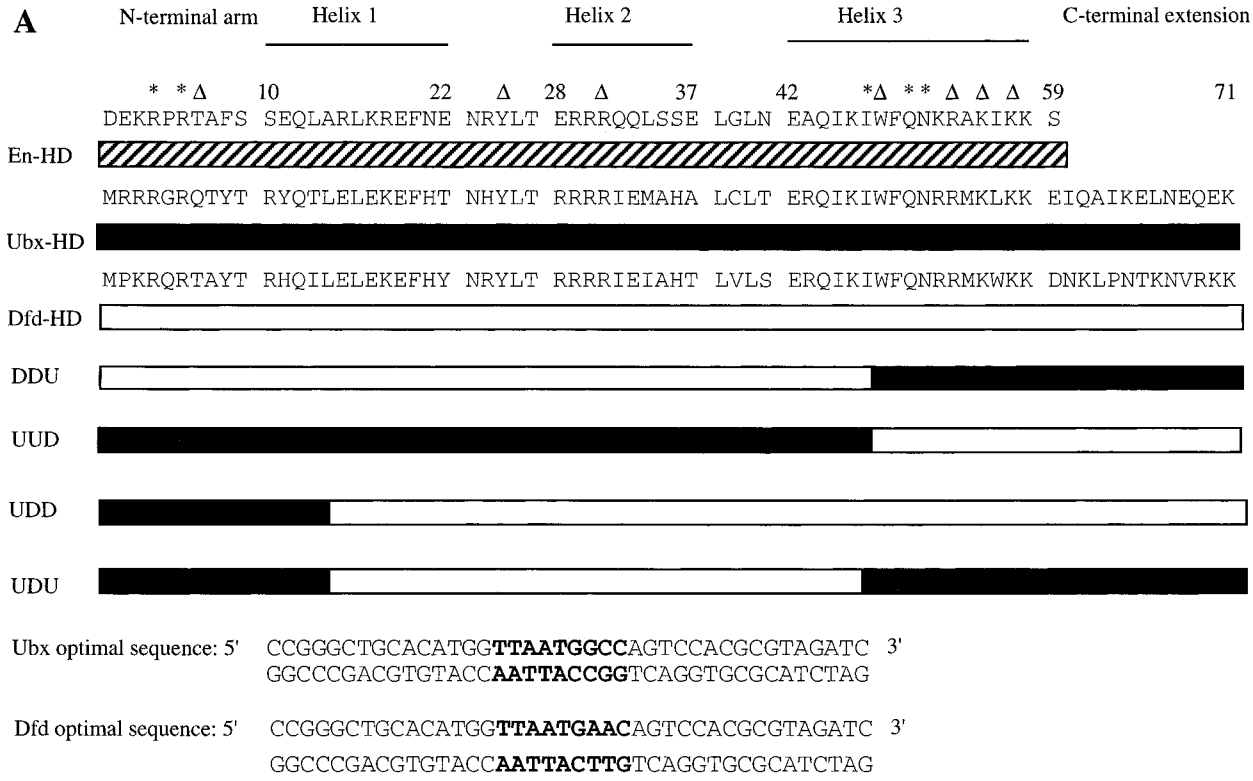


FIGURE 1: (A) Sequences of Engrailed-HD (En-HD), Ultrabithorax-HD (Ubx-HD), Deformed-HD (Dfd-HD), and four chimeric proteins derived from Ubx- and Dfd-HDs. The location of the three helices is based on the cocrystal structure of the En-HD−DNA complex described by Kissinger *et al.* (1990). The asterisks (*) correspond to base contacts and triangles (△) to phosphate contacts assigned in the En-HD−DNA cocrystal. The amino acid sequences of the Ubx-HD and Dfd-HD are shown as filled and open bars, respectively. The regions of Ubx and Dfd sequences contained in each of the four chimeras are shown schematically, and these are abbreviated as DDU, UUD, UDD, and UDU, to indicate the segments corresponding to the parent-HD. All homeodomains were constructed and purified as described in Ekker *et al.* (1991, 1992). Concentration of purified homeodomains was determined using molar absorptivity at 280 nm, and the active concentration was determined by a DNA binding activity assay under stoichiometric conditions. (B) DNA sequences containing the Ubx optimal binding site or Dfd optimal binding site. The binding site sequences in each DNA are highlighted.

illuminated by structural studies of multiple protein−DNA complexes, including homeodomain−DNA complexes (Lawson & Carey, 1993; Otwinowski *et al.*, 1988; Chuprina *et al.*, 1993; Luisi *et al.*, 1991; Gewirth & Sigler, 1995; Feng *et al.*, 1994; Steitz, 1990; Pabo & Sauer, 1992; Qian *et al.*, 1989, 1994; Otting *et al.*, 1990; Billeter *et al.*, 1993, 1996; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Billeter, 1995; Li *et al.*, 1995; Morton & Ladbury, 1996). These studies have shown that water molecules may both compete for and mediate hydrogen bonds between these macromolecules.

In the Antennapedia homeodomain−DNA complex, for example, molecular dynamics simulations have indicated that Ile 47, Gln 50, and Asn 51 as well as other residues may each be involved in multiple direct hydrogen bond interactions with different base pairs and may participate in a number of water-mediated hydrogen bonds with bases (Billeter, 1996). Obviously, such interactions do not occur simultaneously, illuminating the dynamic nature of the binding interface and suggesting an active role for water in complex formation and maintenance. Presumably, a greater number of water molecules will remain in a protein−DNA interface if water-mediated hydrogen bonds dominate over direct hydrogen bonds between protein and DNA. While NMR in aqueous solutions and X-ray crystal diffraction studies provide information on the lifetime and positions of tightly bound hydration water molecules, complementary information may be derived from osmotic stress techniques, in particular to detect additional weakly bound waters that are difficult to discern by other methods (Rand, 1992). The osmotic stress strategy is simple, generally applicable, and sensitive (Rand, 1992; Parsegian *et al.*, 1995).

The number of water molecules released by the formation of a complex between protein and DNA can be estimated by examining the effects of altering water activity on the binding affinity using neutral solutes to vary the osmotic strength (Parsegian *et al.*, 1986, 1995; Rand, 1992). With variation of osmotic stress, water participation in molecular recognition of specific DNA sites has been demonstrated for numerous proteins, including restriction endonucleases *Eco*RI, *Eco*RV, *Bam*HI, and *Pvu*II (Robinson & Sligar, 1993, 1994, 1995a; Sidorova & Rau, 1996), *gal* repressor (Garner & Rau, 1995), and Hin recombinase (Robinson & Sligar, 1996). Using this approach, we show in the present study that Ubx-HD−DNA binding involves greater water release than does Dfd-HD−DNA binding. Such a disparity in osmotic stress-related behavior for closely related proteins has been observed previously for DNA binding by restriction enzymes (Robinson & Sligar, 1995a), oxygen binding by hemoglobin and myoglobin (Colombo *et al.*, 1992), and protein−protein complex formation of cytochrome *c*−cytochrome $b_5$ and cytochrome *c*−cytochrome *c* oxidase (Kornblatt *et al.*, 1993).

## MATERIALS AND METHODS

*Construction and Purification of Homeodomains.* Ubx and Dfd homeodomains and four chimeric homeodomains derived from these proteins, DDU, UUD, UDD, and UDU (see Figure 1A for structures), were constructed and purified

as described in Ekker *et al.* (1991, 1992). Concentration of purified homeodomains was determined using the molar absorptivity for aromatic amino acids at 280 nm (Wetlaufer, 1962), calculated on the basis of the amino acid composition of each homeodomain (Ekker *et al.*, 1992). The values in units of $M^{-1}$ $cm^{-1}$ used were as follows: Ubx-HD, 9860; Dfd-HD, 15 460; DDU, 9860; UUD, 15 460; UDD, 14 040; and UDU, 8440. The active concentrations of the homeodomains were determined by DNA binding activity assays under stoichiometric conditions. The percentage of activity measured for these proteins was as follows: Ubx-HD, ∼100%; Dfd-HD, ∼50%; DDU, ∼100%; UUD, ∼70%; UDD, ∼40%; and UDU, ∼50%. The calculated active concentration based on the percent of activity was used in the experimental analyses.

*Oligonucleotide Sequences for DNA Binding.* Two synthetic oligonucleotides of 40 bp were used for DNA binding measurements (see Figure 1B for sequences). The first contained the Ubx optimal DNA binding site sequence, TTAATGGCC, and the second contained the Dfd optimal DNA binding site sequence, TTAATGAAC (Ekker *et al.*, 1992). The differences in the two binding site sequences are indicated by underlining. Equal molar amounts of synthesized single-stranded oligonucleotides (top and bottom strands) were annealed, purified by electrophoresis if necessary, and then end-labeled with [$\gamma$-$^{32}$P]ATP. The labeled oligonucleotide was separated from free [$\gamma$-$^{32}$P]ATP using a NICK column (Pharmacia) or a QIA-quick-spin column (QIAGEN).

*Equilibrium DNA Binding Measurements.* Binding reactions were performed by incubating homeodomain polypeptides (concentrations ranging from $2 \times 10^{-12}$ to $1 \times 10^{-8}$ M) with constant 40 bp oligonucleotides ($5 \times 10^{-12}$ M) in reaction mixtures containing 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 50 $\mu$g/mL BSA, and 1 mM DTT, with various concentrations of neutral solutes. The buffer contained no glycerol or tracking dyes to avoid problems with osmotic pressure or the influence of dyes on the protein−DNA interaction. Freshly thawed homeodomain polypeptide was used for each DNA binding experiment. Electrophoresis of retardation gels was performed as described previously (Li *et al.*, 1996). Briefly, gels which contain 8% polyacrylamide (19:1 acrylamide:bis-acrylamide), 0.5× TBE (0.045 M Tris−borate and 0.001 M EDTA), and 3% glycerol were preelectrophoresed to a constant current at 100 V with recirculation of 0.5× TBE buffer. Aliquots (15 $\mu$L) of each sample were loaded onto the gels running at 300 V. To avoid sample diffusion and protein−DNA complex dissociation, especially for samples without addition of neutral solutes, samples were loaded slowly onto the bottom of wells with microcapillary gel tips. Once tracking dyes (loaded side by side with experimental samples in a separate lane) had separated from each other, the voltage was reduced to 150 V, and the gel was electrophoresed at room temperature for ∼1.5 h.

Dried gels were exposed to a Fuji phosphor imaging plate overnight, and homeodomain−DNA complexes were quantified by scanning and analyzing the image generated using a Fuji BAS1000 Imaging Analyzer (Fuji Photo Film Co., Ltd., Japan) (Li & Matthews, 1995). Each set of titrations included a sample with only free DNA as a background for quantitative analysis of protein−DNA complexes in other samples. Although the contrast for prints of the gel images can be varied by adjusting the brightness and/or contrast, these changes do not affect the data values recorded (*i.e.*, contrast is independent of data valuation in the Fuji Imaging system). For determination of the equilibrium binding constant, a simultaneous fit for multiple sets ($n \geq 4$) of DNA binding data was carried out using the program Igor Version 2.0 to generate fits to the binding equation

$$R = [\text{protein}]/(K_d + [\text{protein}]) \quad (1)$$

where $R$ is the normalized apparent saturation, $K_d$ is the equilibrium dissociation constant, and [protein] is the protein concentration. Both minimum and maximum values of $R$ were floated in the fitting process. Bound DNA was employed for analysis of binding isotherms, although disappearance of free DNA and appearance of bound DNA yielded similar results.

*Osmotic Stress Strategy.* We have employed a series of neutral solute molecules, which are presumably excluded from the surface water that surrounds protein and DNA surfaces. These solutes lower the water activity outside the hydration shell and thereby favor water removal. This effect in turn may alter the binding affinity observed for a specific protein−DNA interaction. In an analysis similar to conventional determinations of thermodynamic linkage (Wyman & Gill, 1990), *e.g.*, for the number of ions involved in protein−DNA binding (Record *et al.*, 1976, 1991), the number of water molecules released can be quantified by the equation

$$d[\ln(K_a)]/d[\text{solute}] = -\Delta n_w/55.6 \quad (2)$$

where $K_a$ is the equilibrium association constant, [solute] is the osmolal concentration of the neutral solute employed, $\Delta n_w$ is the change in the number of water molecules in the hydration shell as a result of protein−DNA association, and 55.6 is the number of moles of water in 1 L (Parsegian *et al.*, 1995). To propagate errors from fits of the binding data to generate errors for the slopes to determine $\Delta n_w$, fits were performed using the NonLin program (Johnson & Frasier, 1985; Johnson & Faunt, 1992), and confidence intervals for the fits are reported at one standard deviation.

Four chemically diversified neutral solutes, triethylene glycol (TEG), sucrose, betaine, and glycerol, were selected to vary the water activity. Osmotic pressure data for sucrose and glycerol were obtained from the *Handbook of Chemistry and Physics* (Weast, 1985). The osmolal concentrations for TEG and betaine ($P$) were derived from molal concentrations ($m$) according to data measured with a Wescor model 5100C vapor pressure osmometer and provided by D. C. Rau (personal communication; Garner & Rau, 1995; also see http://aqueous.labs.brocku.ca/osfile.html):

$$P_{TEG} = 1.0263m + 0.0291m^2 \quad (3)$$

$$P_{betaine} = 0.0021 + 0.9941m + 0.1248m^2 \quad (4)$$

## RESULTS

*Ubx-HD−DNA Binding Involves Greater Water Release Than Dfd-HD−DNA Binding.* Binding of Ubx-HD to its optimal binding sequence, TTAATGGCC, was significantly enhanced by the presence of a neutral solute, sucrose (Figure 2A−C). The effects of TEG, betaine, and glycerol were found to be qualitatively similar to that of sucrose, despite

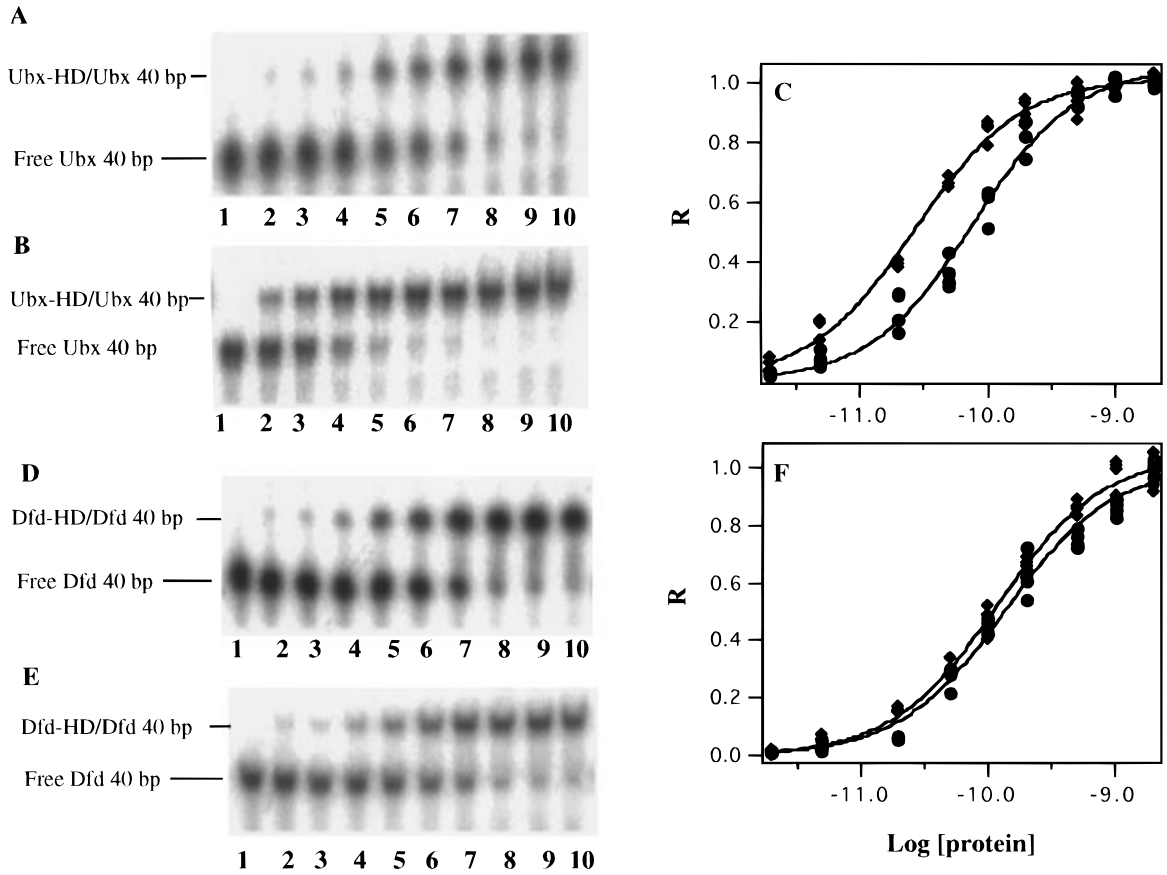


FIGURE 2: Effect of the neutral solute sucrose on Ubx-HD and Dfd-HD binding to their optimal DNA binding sequences. Binding reactions were conducted by incubating homeodomain polypeptides with $5 \times 10^{-12}$ M 40 bp oligonucleotide in reaction mixtures containing 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 50 $\mu$g/mL BSA, and 1 mM DTT with or without 1.5 M sucrose. To determine the equilibrium binding constant, DNA binding data from multiple experiments were analyzed using the program Igor version 2.0 to generate fits to the binding equation $R = [\text{protein}]/(K_d + [\text{protein}])$, where $R$ is the normalized apparent saturation, $K_d$ is the equilibrium dissociation constant, and [protein] is the protein concentration. The maximum and minimum values of $R$ were floated in the fitting process. Lines in panels C and F are the fits to the data shown. (A−C) Ubx-HD−Ubx 40 bp DNA: (A) a single titration without sucrose in the reaction mixture and (B) with 1.5 M sucrose in the reaction mixture and (C) data and fits for multiple titrations under the conditions shown in panels A and B. (D−F) Dfd-HD−Dfd 40 bp DNA: (D) a single titration without sucrose in the reaction mixture and (E) with 1.5 M sucrose in the reaction mixture and (F) data and fits for multiple titrations under the conditions shown in panels D and E. Protein concentrations from lanes 1 to 10 in panels A, B, D, and E are as follows: 0, $2 \times 10^{-12}$, $5 \times 10^{-12}$, $2 \times 10^{-11}$, $5 \times 10^{-11}$, $1 \times 10^{-10}$, $2 \times 10^{-10}$, $5 \times 10^{-10}$, $1 \times 10^{-9}$, and $2 \times 10^{-9}$ M. In panels C and F, filled circles indicate experiments without sucrose in the reaction mixtures and filled diamonds represent experiments with 1.5 M sucrose in the reaction mixtures.

the distinct chemical natures of these four compounds. To quantify the effect of neutral solutes on the binding of Ubx-HD to this DNA, the equilibrium association constant ($K_a$) was measured for this series of solutes at varying concentrations. The data generated were analyzed by plotting ln $K_a$ *vs* $[\text{solute}]_{\text{osmolal}}$ (eq 2). Figure 3 shows that TEG, sucrose, betaine, and glycerol have quantitatively similar effects on the binding of Ubx-HD to its optimal binding sequence, with release of ~25 molecules of water upon complex formation (Table 1).

In contrast, only very minor effects of the neutral solute sucrose were observed on the binding of Dfd-HD to its optimal DNA sequence, TTAATGAAC (Figures 2D−F and 3). No qualitative or quantitative differences were noted among the series of four solutes utilized. For Dfd-HD binding, complex formation is accompanied by the release of ~5 water molecules (Table 1), 5-fold lower than observed for the closely related Ubx-HD.

*The C-Terminal Region of Ubx-HD Is Responsible for the Enhancement of Water Release on DNA Binding.* To determine which subregion(s) in Ubx-HD and Dfd-HD might be responsible for the significant difference in water release associated with the binding of these two homeodomains to DNA, we have examined four chimeric homeodomains derived from Ubx- and Dfd-HDs (see Figure 1A for structures). Similar to that of Ubx-HD, DNA binding affinity increased with addition of the neutral solute TEG for two of these chimeric homeodomains, DDU and UDU (Figure 4C,F). Both of these chimeric proteins contain the C-terminal region of the Ubx-HD. In contrast to DDU and UDU, no significant solute-dependent increase in binding affinity was found for UUD or UDD (Figure 4D,E). Similar results were observed in experiments using glycerol as a neutral solute (data not shown). These data demonstrated that the C-terminal region of the Ubx-HD is the primary determinant for differential water release associated with DNA binding between the Ubx-HD and Dfd-HD.

*DNA Binding Site Sequences Do Not Exert Significant Effect on the Magnitude of Water Release Associated with Ubx-HD−DNA Binding.* To explore whether the DNA binding site sequences affect water release associated with binding of Ubx-HD, we have compared the binding of Ubx-HD to its optimal binding sequence, TTAATGGCC, with its binding to the Dfd optimal sequence, TTAATGAAC.

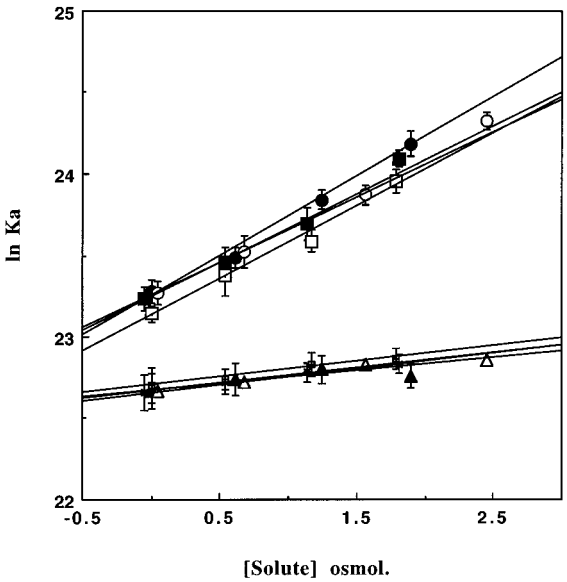


FIGURE 3: Effects of neutral solutes on equilibrium association constants ($K_a$) of Ubx-HD and Dfd-HD binding to DNA. The conditions for binding reactions for determining equilibrium association constants were identical to those indicated in Figure 2. Triethylene glycol (TEG), sucrose, betaine, and glycerol in different concentrations were used to vary the osmolality. DNA binding data from multiple titrations were simultaneously fit using Igor program version 2.0, and the number of water molecules released can be quantified by the equation $d[\ln(K_a)/d[\text{solute}] = -\Delta n_w/55.6$, where $K_a$ is the equilibrium association constant, [solute] is the osmolal concentration of the neutral solute employed, $\Delta n_w$ is the change in the number of water molecules in the hydration shell as a result of protein−DNA association, and 55.6 is the number of moles of water in 1 L (Parsegian *et al.*, 1995): (●) Ubx-HD−TEG, (○) Ubx-HD−sucrose, (■) Ubx-HD−betaine, (□) Ubx-HD−glycerol, (▲) Dfd-HD−TEG, (△) Dfd-HD−sucrose, (*) Dfd-HD−betaine, (□ with +) Dfd-HD−glycerol. The resulting $\Delta n_w$ values are summarized in Table 1.

Table 1: Water Release upon Formation of Homeodomain−DNA Complexes[a]

| homeodomain−DNA | solute | $\Delta n_w$ |
|---|---|---|
| Ubx-HD−Ubx 40 bp | TEG | 27 ± 3 |
| Ubx-HD−Ubx 40 bp | sucrose | 23 ± 3 |
| Ubx-HD−Ubx 40 bp | betaine | 22 ± 4 |
| Ubx-HD−Ubx 40 bp | glycerol | 25 ± 3 |
| Dfd-HD−Dfd 40 bp | TEG | 5 ± 2 |
| Dfd-HD−Dfd 40 bp | sucrose | 5 ± 1 |
| Dfd-HD−Dfd 40 bp | betaine | 5 ± 2 |
| Dfd-HD−Dfd 40 bp | glycerol | 5 ± 1 |
| Ubx-HD−Dfd 40 bp | TEG | 28 ± 5 |
| UDU-HD−Ubx 40 bp | TEG | 30 ± 4 |
| UDU-HD−Dfd 40 bp | TEG | 30 ± 5 |

[a] The number of water molecules released was quantified as described in Materials and Methods and the legend to Figure 3. To propagate errors carried from fits to the binding data and to generate errors for the slopes, fits were performed using the NonLin program (Johnson & Frasier, 1985; Johnson & Faunt, 1992), and confidence intervals for the fits are reported at one standard deviation.

Neutral solutes increased the binding affinity of Ubx-HD for both DNA binding sequences examined to a similar degree (Figure 5A). Analogous results were obtained for the binding of UDU to Ubx and Dfd optimal binding sequences (Figure 5B). These data suggest that protein structure is the primary determinant for the effect of water activity on protein−DNA binding and that, at least for the two DNAs examined, base sequence does not influence the magnitude of the observed water release.

*No Influence of pH on Water Release Was Indicated for either Ubx- or Dfd-HD.* We reported recently that Ubx-HD binding to a single optimal target site displayed significantly increased affinity at lower pH, while the binding of Dfd-HD to DNA was unaffected by pH (Li *et al.*, 1996). We therefore attempted to determine whether the pH effect correlates with water release associated with Ubx-HD binding to DNA. Our experiments showed that solute effects on Ubx-HD binding to DNA were similar at pH 7.5 and 6.0, and addition of solutes did not increase the binding of the Dfd-HD to DNA at either pH (data not shown).

## DISCUSSION

*Role of Water Activity in Homeodomain−DNA Binding.* Water activity exerts a significant influence on multiple cellular functions, from membrane assembly to protein−protein interactions to protein−DNA association [Rand & Parsegian, 1989; Parsegian & Zimmerberg, 1993; Preisler *et al.*, 1995; reviewed in Rand (1992), Leikin *et al.* (1993), Robinson and Sligar (1995b), and Parsegian *et al.* (1986, 1995)]. The binding of hexokinase to glucose is accompanied by the release of ~65 water molecules, and the specific binding of *gal* repressor to its operator sequence is accompanied by the release of ~100−180 water molecules (Garner & Rau, 1995). The results from the present study demonstrate that the binding of Ubx-HD to DNA is associated with significantly greater water release (~25 water molecules) than the binding of Dfd-HD to DNA (~5 water molecules).

In various complexes for which protein−DNA structures have been completed, extensive networks of hydrogen bonds and salt bridges have been found between amino acid side chains and the bases and sugar−phosphate backbone of double-stranded DNA (Steitz, 1990; Pabo & Sauer, 1992; Qian *et al.*, 1989, 1994; Otting *et al.*, 1990; Billeter *et al.*, 1993, 1996; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Billeter, 1995; Li *et al.*, 1995; Morton & Ladbury, 1996). Although water release that accompanies protein−DNA complex formation is presumably a general phenomenon, a direct correlation between water activity and structural complementarity of protein and DNA has been explored only recently [reviews in Leikin *et al.* (1993), Robinson and Sligar (1995b), Parsegian *et al.* (1995), and Morton and Ladbury (1996)]. For example, water molecules appear to participate in DNA recognition by restriction endonucleases (Robinson & Sligar, 1993, 1994, 1995a; Sidorova *et al.*, 1995; Sidorova & Rau, 1996), Hin recombinase (Robinson & Sligar, 1996), the drug netropsin and its analogue (Sidorova & Rau, 1995; Sidorova *et al.*, 1995), and the Antp homeodomain (Billeter *et al.*, 1996).

A coupling has been found between local folding, water release, and site-specific binding of DNA in numerous protein−DNA complexes, including homeodomain proteins (Ha *et al.*, 1989; Spolar & Record, 1994; Morton & Ladbury, 1996; Tsao *et al.*, 1994, 1995; Cox *et al.*, 1995; Carra & Privalov, 1997). Furthermore, the number of water molecules that appear to participate in macromolecular interaction can be quite large. While a small number of tightly bound waters have been identified in crystallographic and NMR studies of protein−DNA complexes, significantly less information is available on weakly bound waters that have low occupancies and short half-lives. Molecular dynamics

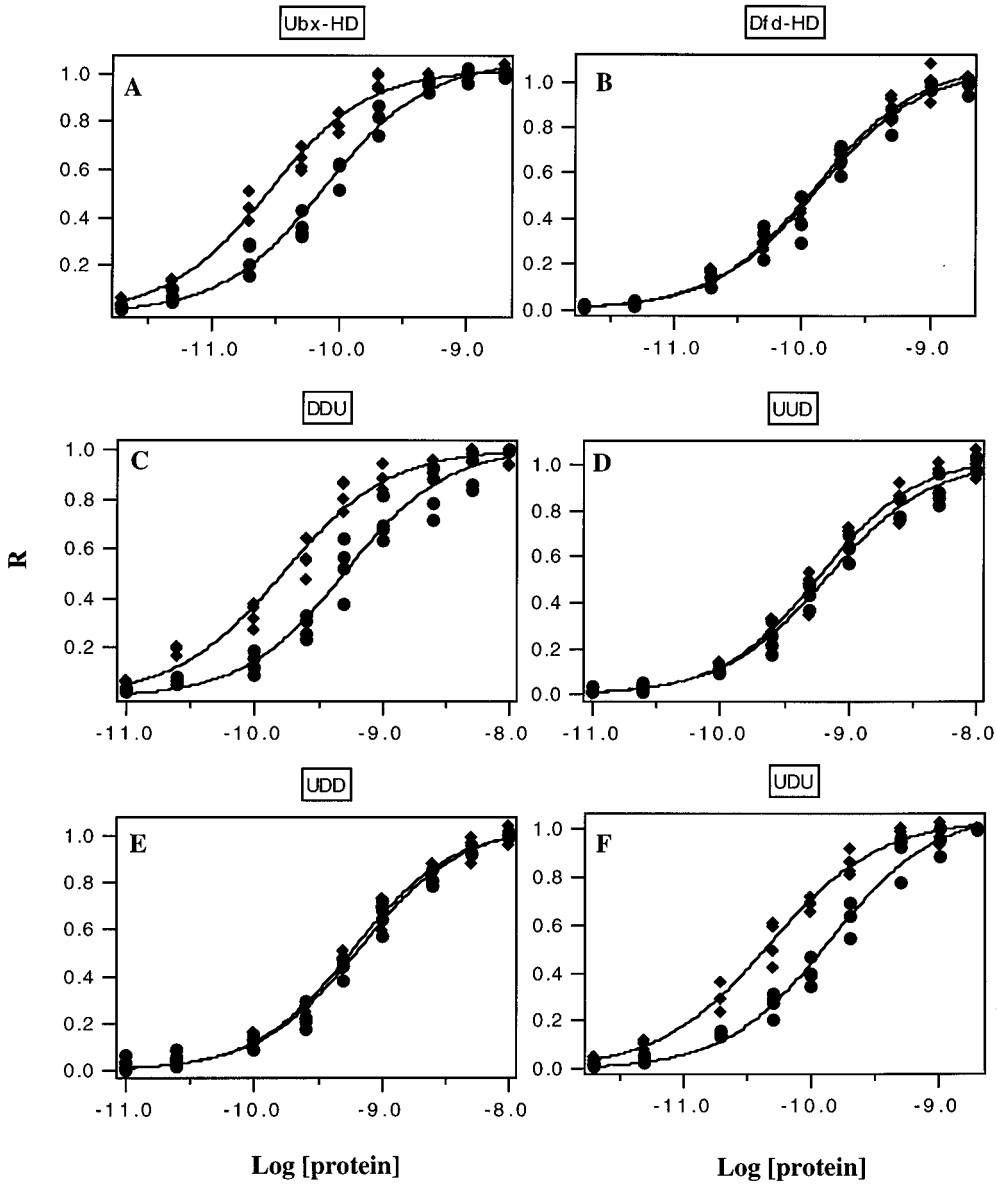


FIGURE 4: Comparison of the effect of TEG on DNA binding by chimeric homeodomains derived from Ubx- and Dfd-HDs. The Ubx optimal DNA binding sequence was used for binding reactions of Ubx-HD and chimeric homeodomains, while the Dfd optimal sequence was used for binding reactions of Dfd-HD. The identification of homeodomain protein used in the individual experiments is indicated on the top of each panel. Filled circles indicate experiments in the reaction mixtures without TEG, and filled diamonds specify experiments in the reaction mixtures with 1.5 M TEG. The results demonstrate that the C-terminal region of Ubx-HD is the primary determinant for the greater water release associated with DNA binding for this protein.

analyses indicate that residues in the Antennapedia homeodomain−DNA complex may participate in multiple direct hydrogen bonds and be involved in a number of distinct water-mediated hydrogen bonds with bases, all with different occupancies and different half-lives (Billeter, 1996). These results suggest that water molecules may participate in a dynamic fashion, with varying degrees of retention, in protein−DNA interactions.

*Protein Structure Appears To Be the Major Determinant for the Differential Water Release with Ubx-HD and Dfd-HD Binding to DNA.* The difference between Ubx-HD and Dfd-HD water release upon the formation of their respective protein−DNA complexes must arise from their structural differences, although these two homeodomains share a high degree of sequence similarity. The homeodomain−DNA structures determined by X-ray crystallography or NMR reveal that the homeodomains examined have a similar molecular fold and form specific interactions with DNA in a conserved manner. The N-terminal region contacts bases in the minor groove, and residues in helix 3 contact bases in the major groove (Qian *et al.*, 1989, 1994; Otting *et al.*, 1990; Billeter *et al.*, 1993; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Billeter, 1995; Li *et al.*, 1995). Residues contacting the DNA backbone were found to be concentrated in the helix 3, although additional points of contact are found throughout the entire homeodomain (Qian *et al.*, 1989, 1994; Otting *et al.*, 1990; Billeter *et al.*, 1993; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Li *et al.*, 1995; also see Figure 2A).

Our studies of chimeric homeodomains indicate that the C-terminal region of Ubx-HD contributes significantly to the high level of water release associated with the formation of the Ubx-HD−DNA complex. Ubx and Dfd homeodomains have near amino acid sequence identity in helix 3; the only major exception is L56 in Ubx-HD and W56 in Dfd-HD (see Figure 1A for comparison). Although no critical contact

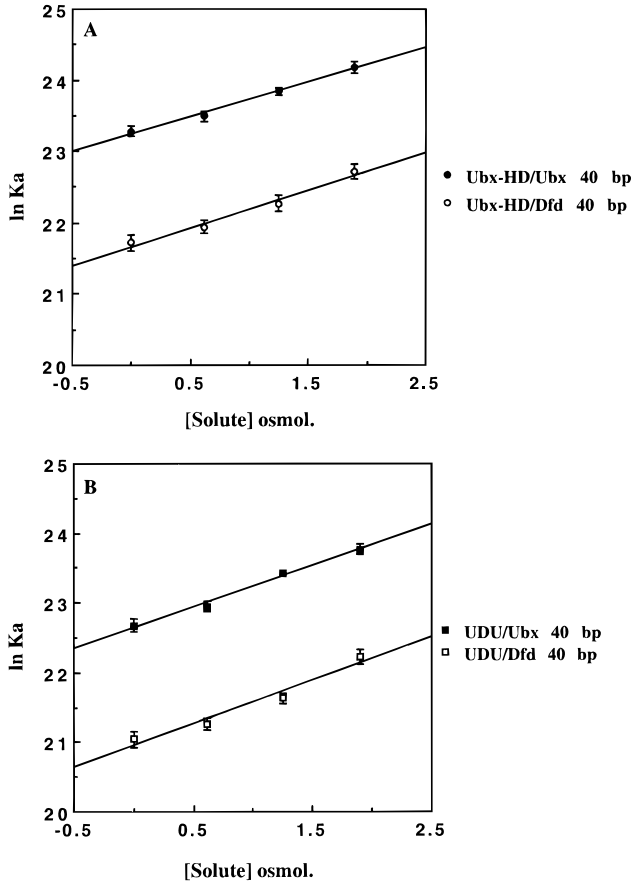


FIGURE 5: Effects of DNA binding sequences on water release by protein−DNA complex formation. Ubx− and Dfd−optimal 40 bp oligonucleotides were used to monitor effects of different DNA sequences on the osmotic pressure sensitivity of protein−DNA binding. The conditions for binding reactions were the same as those indicated in Figure 2. The resulting DNA binding data were analyzed as described in Materials and Methods and in the legends to Figures 2 and 3 and Table 1. The differences observed for the binding to Ubx 40 bp and Dfd 40 bp DNAs are not significant within 67% confidence limits. (A) (●) Ubx-HD−Ubx 40 bp DNA and (○) Ubx-HD−Dfd 40 bp DNA. (B) (●) UDU-chimeric HD−Ubx 40 bp DNA and (□) UDU-chimeric HD−Dfd 40 bp DNA.

of this side chain with either a base or phosphate in the DNA backbone was indicated in structural studies for homologous homeodomain proteins (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Li *et al.*, 1995; Otting *et al.*, 1990; Billeter *et al.*, 1993), a bulky hydrophobic side chain at this position is important to packing of the core of the protein and the elongation of helix 3 observed in some proteins upon DNA binding (Qian *et al.*, 1994; Tsao *et al.*, 1994, 1995). Moreover, the adjacent residue at position 57 (often Lys) makes contact with the phosphate backbone in some structures (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Billeter *et al.*, 1993). The extension of helix 3 upon complex formation of the homeodomain with its DNA target sequence has been demonstrated for the NK-2 homeodomain, the Oct-1 POU homeodomain, and the MATα2 homeodomain (Tsao *et al.*, 1994, 1995; Cox *et al.*, 1995; Carra & Privalov, 1997).

We demonstrate that helix 3 with its C-terminal extension contains residues that contribute to the increased water release upon the interaction of Ubx-HD and DNA that may arise either from coupled binding and folding or from other mechanisms. For example, the C-terminal extension of Ubx-HD is more apolar and more electrostatically neutral than that for Dfd-HD, and this difference may have a significant impact on the associated water release for Ubx-HD. A greater degree of water exclusion may occur in the interface of Ubx-HD and DNA. In a similar vein, the more hydrophilic character of the C-terminal extension of Dfd-HD may result in greater water retention in the protein−DNA interface, resulting in less water being excluded from the binding surface. Whether these waters participate directly in hydrogen bond interactions in the complex is not discernible from the data. A further possibility is that similar amounts of water may be released from the protein−DNA interface of each homeodomain, but water binds elsewhere upon the formation of the complex. This water binding might occur as a consequence of local unfolding of Dfd-HD when it binds to DNA, as has been observed in binding of a subset of restriction enzymes to DNA (Robinson & Sligar, 1995).

Previous studies reported that specific protein−DNA binding is associated with significant water release, whereas weaker, nonspecific protein−DNA associations are unaffected by solute addition (Ha *et al.*, 1992; Garner & Rau, 1995; Sidorova & Rau, 1996). In this instance, however, both Ubx-HD and Dfd-HD binding to DNA are sequence-specific; nonetheless, the extent of water release upon complex formation differs greatly. The origin of this difference for such similar proteins remains unclear, and conclusions regarding the detailed mechanism by which differential water release is generated for these related homeodomains will require further study. However, striking differences in the response to neutral solutes have also been found in other structurally similar sets of proteins. For example, the cleavage specificity of *Eco*RI, *Bam*HI, and *Pvu*II was dramatically reduced at elevated osmotic pressure, while the cleavage specificity for *Eco*RV was unaffected (Robinson & Sligar, 1993, 1994, 1995a). Approximately 60 water molecules are associated with hemoglobin binding to oxygen, but myoglobin binding to oxygen does not show this effect (Colombo *et al.*, 1992). Complex formation of cytochrome *a*−cytochrome $b_5$ is enhanced by glycerol; however, association of cytochrome *c*−cytochrome *c* oxidase is inhibited by glycerol (Kornblatt *et al.*, 1993). These observations, in concert with the results reported here, demonstrate that effects of water activity on protein−ligand binding may not be similar even within the same structural families of proteins. Subtle differences in protein structures, as indicated for the homeodomain family (see below), may generate a significant differential response to the physical environment.

*Differential Responses to the Physical Environment Provide Potential Site Discrimination.* The data presented in this study demonstrate that quite small differences in Ubx-HD and Dfd-HD structures yield significant differences in the response of their DNA binding to the presence of neutral solutes. As indicated previously, Ubx-HD, Dfd-HD, and other homeodomains from both insect and mammalian sources share a high degree of sequence identity (Scott & Weiner, 1984; Scott *et al.*, 1989; Hayashi & Scott, 1990). For the homeodomain proteins for which structures have been determined, a similar molecular fold has been observed (Gehring *et al.*, 1994a,b), although subtle differences in structure have been identified [*e.g.*, Qian *et al.* (1994), Tsao *et al.* (1994, 1995), Billeter (1995), and Draganescu *et al.* (1995)]. Furthermore, these proteins recognize similar DNA binding sequences (Ekker *et al.*, 1991, 1992, 1994). An

important question that remains largely unanswered is how the individual homeodomain proteins, which exhibit only low-level discrimination in site recognition *in vitro*, perform distinct biological functions *in vivo*. Individual homeodomains have particular preferences in sequences flanking the TAAT core, which account for their small differences in binding specificity and affinity (Scott *et al.*, 1989; Kuziora & McGinnis, 1989; Hayashi & Scott, 1990; Catron *et al.*, 1993; Gehring *et al.*, 1994a,b; Ekker *et al.*, 1991, 1992, 1994; Pomerantz & Sharp, 1994). The optimal binding sites for Ubx-HD and Dfd-HD differ by two bases at positions 7 and 8 in the 9 bp binding site sequences (TTAATGGCC for Ubx-HD and TTAATGAAC for Dfd-HD) (Ekker *et al.*, 1992). Different flanking sequences, however, generate differences of only a few orders of magnitude in both the dissociation rate constant and half-lives of protein−DNA complexes (Ekker *et al.*, 1992, 1994).

Recent examination of the Ubx- and Dfd-homeodomains by hydroxyl radical footprinting and missing nucleoside experiments revealed striking differences in contacts with DNA between these two homeodomains using the optimal DNA sequences for either of these proteins (Draganescu *et al.*, 1995). These observations suggest that subtle differences in protein structures (including the C-terminal extension region) can result in different side chain−nucleoside interactions in the protein−DNA complexes and presumably differential participation of water in the binding interface. Nonetheless, the low level of discrimination in site recognition (within 10-fold in binding affinity) is still surprising given the apparently discrete biological functions of the homeodomain proteins. Two mechanisms by which specific target site recognition could occur *in vivo*, even with minimal differential for an individual binding site, are by homomeric cooperative DNA binding to multiple sites (Beachy *et al.*, 1993) or by heteromeric cooperative DNA binding involving partners that amplify specificity, such as observed for Hox-Pbx−Extradenticle binding [Chang *et al.*, 1996; reviewed in Wilson and Desplan (1995) and Mann and Chan (1996)]. An interaction between Extradenticle and Ultrabithorax proteins has been observed using the yeast two-hybrid system (Johnson *et al.*, 1995), and this association appears to increase DNA binding specificity of the Ultrabithorax protein (Chan *et al.*, 1994; van Dijk & Murre, 1994; Johnson *et al.*, 1995; Chan & Mann, 1996).

We have established that marginal binding affinity differences derived from small differences in protein structures and DNA sequences can be amplified by differential responses of protein−DNA interactions to the physical environment. In the case of Ubx-HD, a lower pH can generate a 24-fold higher binding affinity to its optimal binding sequence compared to that with a neutral pH. In contrast, pH has no significant effect on DNA binding affinity of Dfd-HD (Li *et al.*, 1996). In this study, the differences in DNA binding between Ubx-HD and Dfd-HD have been further expanded. Neutral solutes that exist normally in living cells and protect them from freezing or osmotic shock (Timasheff, 1993) elicit distinct effects on DNA binding for Ubx-HD and Dfd-HD. The binding of Ubx-HD to DNA can be enhanced by increasing the osmolality of the bulk solution, while the binding of Dfd-HD to DNA is much less sensitive to the same changes in osmotic conditions. Significant discrimination in site recognition may be achieved by integrating binding affinity differences derived from small differences in protein structures and DNA sequences with those generated by homo- and heteromeric protein interactions and differential responses of protein−DNA interactions to the physical environment.

## ACKNOWLEDGMENT

We especially thank Dr. Philip Beachy and Doris von Kessler for providing the purified homeodomain proteins used in this study and for comments on the manuscript. We thank Drs. Clifford Robinson, Stephen Sligar, and Donald Rau for communication of their results and suggestions for the manuscript and Drs. Liskin Swint-Kruse and Ross Reedstrom for suggestions and assistance on analysis of data. We also appreciate feedback, discussion, and comments on the manuscript from Dr. D. Wayne Bolen and laboratory members Diane Wycuff, Liskin Swint-Kruse, Catharine Falcon, and Nicole Magnasco. The facilities of the Keck Center for Computational Biology were used in this study.

## REFERENCES

Akam, M. (1987) *Development 101*, 1−22.
Beachy, P. A., Varkey, J., Young, K. E., von Kessler, D. P., Sun, B. I., & Ekker, S. C. (1993) *Mol. Cell. Biol. 13*, 6941−6956.
Billeter, M. (1995) *Prog. Nucl. Magn. Reson. Spectrosc. 27*, 635−645.
Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W., & Wüthrich, K. (1993) *J. Mol. Biol. 234*, 1084−1097.
Billeter, M., Güntert, P., Luginbühl, P., & Wüthrich, K. (1996) *Cell 85*, 1057−1065.
Carra, J. H., & Privalov, P. L. (1997) *Biochemistry 36*, 526−535.
Catron, K. M., Iler, N., & Abate, C. (1993) *Mol. Cell. Biol. 13*, 2354−2365.
Chan, S. K., & Mann, R. S. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5223−5228.
Chan, S. K., Jaffe, L., Capovilla, M., Botas, J., & Mann, R. S. (1994) *Cell 78*, 603−615.
Chang, C.-P., Brocchieri, L., Shen, W.-F., Largman, C., & Cleary, M. L. (1996) *Mol. Cell. Biol. 16*, 1734−1745.
Chuprina, V. P., Rullmann, J. A. C., Lamerichs, R. M. J. N., van Boom, J. H., Boelens, R., & Kaptein, R. (1993) *J. Mol. Biol. 234*, 446−462.
Colombo, M. F., Rau, D. C., & Parsegian, V. A. (1992) *Science 256*, 655−659.
Cox, M., van Tilborg, P. J., de Laat, W., Boelens, R., van Leeuwen, H. C., van der Vliet, P. C., & Kaptein, R. (1995) *J. Biomol. NMR 6*, 23−32.
Desplan, C., Theis, J., & O'Farrell, P. H. (1988) *Cell 54*, 1081−1090.
Draganescu, A., Levin, J. R., & Tullius, T. D. (1995) *J. Mol. Biol. 250*, 595−608.
Duncan, I. (1987) *Annu. Rev. Genet. 21*, 285−319.
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.
Feng, J.-A., Johnson, R. C., & Dickerson, R. E. (1994) *Science 263*, 348−355.
Garner, M. M., & Rau, D. C. (1995) *EMBO J. 14*, 1257−1263.
Gehring, W. J., Affolter, M., & Bürglin, T. (1994a) *Annu. Rev. Biochem. 63*, 487−526.
Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., & Wüthrich, K. (1994b) *Cell 78*, 211−223.
Gewirth, D. T., & Sigler, P. W. (1995) *Struct. Biol. 2*, 386−394.
Ha, J.-H., Spolar, R. S., & Record, M. T., Jr. (1989) *J. Mol. Biol. 209*, 801−816.

Ha, J.-H., Capp, M. W., Hohenwalter, M. D., Baskerville, M., & Record, M. T., Jr. (1992) *J. Mol. Biol. 228*, 252−264.
Hayashi, S., & Scott, M. P. (1990) *Cell 63*, 883−894.
Hoey, T., & Levine, M. (1988) *Nature 332*, 858−861.
Johnson, F. B., Parker, E., & Krasnow, M. A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 739−743.
Johnson, M. L., & Frasier, S. G. (1985) *Methods Enzymol. 117*, 301−342.
Johnson, M. L., & Faunt, L. M. (1992) *Methods Enzymol. 210*, 1−37.
Kaufman, T. C., Seeger, M. A., & Olsen, G. (1990) *Adv. Genet. 27*, 309−362.
Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579−590.
Kornblatt, J. A., & Hui Bon Hoa, G. (1990) *Biochemistry 29*, 9370−9376.
Kornblatt, J. A., Kornblatt, M. J., Hui Bon Hoa, G., & Mauk, A. G. (1993) *Biophys. J. 65*, 1059−1065.
Kuziora, M. A., & McGinnis, W. (1989) *Cell 59*, 563−571.
Lawson, C. L., & Carey, J. (1993) *Nature 366*, 178−182.
Leikin, S., Parsegian, V. A., Rau, D. C., & Rand, R. P. (1993) *Annu. Rev. Phys. Chem. 44*, 369−395.
Li, L., & Matthews, K. S. (1995) *J. Biol. Chem. 270*, 10640−10649.
Li, L., von Kessler, D., Beachy, P. A., & Matthews, K. S. (1996) *Biochemistry 35*, 9832−9839.
Li, T., Stark, M. R., Johnson, A. D., & Wolberger, C. (1995) *Science 270*, 262−269.
Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., & Sigler, P. B. (1991) *Nature 352*, 497−505.
Mann, R. S., & Chan, S.-K. (1996) *Trends Genet. 12*, 258−262.
McGinnis, W., & Krumlauf, R. (1992) *Cell 68*, 283−302.
McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., & Gehring, W. J. (1984a) *Cell 37*, 403−408.
McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., & Gehring, W. J. (1984b) *Nature 308*, 428−433.
Morton, C. J., & Ladbury, J. E. (1996) *Protein Sci. 5*, 2115−2118.
Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J., & Wüthrich, K. (1990) *EMBO J. 9*, 3085−3092.
Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321−329.
Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
Parsegian, V. A., & Zimmerberg, J. (1993) in *Thermodynamics of Membrane Receptors and Channels* (Jackson, M. B., Ed.) pp 389−405, CRC Press, Boca Raton, FL.
Parsegian, V. A., Rand, R. P., Fuller, N. L., & Rau, D. C. (1986) *Methods Enzymol. 127*, 400−416.
Parsegian, V. A., Rand, R. P., & Rau, D. C. (1995) *Methods Enzymol. 259*, 43−95.
Pomerantz, J. L., & Sharp, P. A. (1994) *Biochemistry 33*, 10851−10858.
Preisler, R. S., Chen, H. H., Colombo, M. F., Choe, Y., Short, B. J., Jr., & Rau, D. C. (1995) *Biochemistry 34*, 14400−14407.
Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1989) *Cell 59*, 573−580.
Qian, Y. Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Müller, M., Gehring, W. J., & Wüthrich, K. (1994) *J. Mol. Biol. 238*, 333−345.
Rand, R. P. (1992) *Science 256*, 618.
Rand, R. P., & Parsegian, V. A. (1989) *Biochim. Biophys. Acta 988*, 351−376.
Rand, R. P., Fuller, N. L., Butko, P., Francis, G., & Nicholls, P. (1993) *Biochemistry 32*, 5925−5929.
Record, M. T., Jr., Lohman, T. M., & deHaseth, P. (1976) *J. Mol. Biol. 107*, 145−158.
Record, M. T., Jr., Ha, J.-H., & Fisher, M. A. (1991) *Methods Enzymol. 208*, 291−343.
Robinson, C. R., & Sligar, S. G. (1993) *J. Mol. Biol. 234*, 302−306.
Robinson, C. R., & Sligar, S. G. (1994) *Biochemistry 33*, 3787−3793.
Robinson, C. R., & Sligar, S. G. (1995a) *Proc. Natl. Acad. Sci. U.S.A. 92*, 3444−3448.
Robinson, C. R., & Sligar, S. G. (1995b) *Methods Enzymol. 259*, 395−428.
Robinson, C. R., & Sligar, S. G. (1996) *Protein Sci. 5*, 2119−2124.
Scott, M. P., & Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4115−4119.
Scott, M. P., Tamkun, J. W., & Hartzell, G. W., III (1989) *Biochim. Biophys. Acta 989*, 25−48.
Sidorova, N. Y., & Rau, D. C. (1995) *Biopolymers 35*, 377−384.
Sidorova, N. Y., & Rau, D. C. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 12272−12277.
Sidorova, N. Y., Gazoni, P., & Rau, D. C. (1995) *J. Biomol. Struct. Dyn. 13*, 367−385.
Spolar, R. S., & Record, M. T., Jr. (1994) *Science 263*, 777−784.
Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205−280.
Timasheff, S. N. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 67−97.
Tsao, D. H. H., Gruschus, J. M., Wang, L.-H., Nirenberg, M., & Ferretti, J. A. (1994) *Biochemistry 33*, 15053−15060.
Tsao, D. H. H., Gruschus, J. M., Wang, L.-H., Nirenberg, M., & Ferretti, J. A. (1995) *J. Mol. Biol. 251*, 297−307.
van Dijk, M. A., & Murre, C. (1994) *Cell 78*, 617−624.
Weast, R. C., Ed. (1985) *Handbook of Chemistry and Physics*, 67th ed., The Chemical Rubber Co., Cleveland, OH.
Wetlaufer, D. B. (1962) *Adv. Protein Chem. 17*, 303−390.
Wilson, D. S., & Desplan, C. (1995) *Curr. Biol. 5*, 32−34.
Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., & Pabo, C. O. (1991) *Cell 67*, 517−528.
Wyman, J., & Gill, S. J. (1990) *Binding and Linkage. Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA.

BI962966M