Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 6886–6891, July 1996
Biochemistry

Conservation and diversification in homeodomain–DNA interactions: A comparative genetic analysis

(transcription factor/DNA binding/structure/evolution)

DAVID S. WILSON*, GUOJUN SHENG, SUSIE JUN, AND CLAUDE DESPLAN

Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10021

Communicated by James E. Darnell, Jr., The Rockefeller University, New York, NY, March 26, 1996 (received for review August 23, 1995)

ABSTRACT **Nearly all metazoan homeodomains (HDs) possess DNA binding targets that are related by the presence of a TAAT sequence. We use an *in vitro* genetic DNA binding site selection assay to refine our understanding of the amino acid determinants for the recognition of the TAAT site. Superimposed upon the conserved ability of metazoan HDs to recognize a TAAT core is a difference in their preference for the bases that lie immediately 3′ to it. Amino acid position 50 of the HD has been shown to discriminate among these base pairs, and structural studies have suggested that water-mediated hydrogen bonds and van der Waals contacts underlie for this ability. Here, we show that each of six amino acids tested at position 50 can confer a distinct DNA binding specificity.**

The homeodomain (HD) is one of the most common DNA binding motifs in eukaryotic transcriptional regulators (1). Structural studies have shown that the HD is composed of three α-helices and a flexible N-terminal arm (2). The third helix of the HD inserts itself into the major groove of the DNA and is, therefore, called the recognition helix (3, 4). In addition, the N-terminal arm of the HD inserts into the minor groove of the DNA, making several contacts with the bases (4, 5).

Despite the vast number of different biological functions of homeoproteins, nearly all homeodomains from metazoans bind to sequences containing a TAAT core motif (6–8). The anatomical descriptions of HD–DNA interactions derived from structural studies suggest a mechanism by which the conserved TAAT motif is recognized by HDs (3, 4, 9–11). In this report, we analyze the optimal DNA binding specificities of several HDs. By combining these results with previously reported DNA site selection experiments, we derive an amino acid consensus sequence for the recognition of the TAAT motif. This consensus is similar to but distinct from those derived previously using different assumptions or more limited data sets. Our consensus includes five amino acid positions that interact with the base pairs of the TAAT motif. Two of these amino acid positions, which map to the N-terminal arm, can form alternative interaction with the same base pair.

Although most metazoan HDs share a preference for the TAAT motif, they can differ from each other in their preference for the bases immediately 3′ to this core (8, 12–15). This preference is determined, at least in part, by the identity of amino acid position 50. Because HD sequences have been identified that possess at least 10 different amino acids at position 50, we investigate whether multiple DNA binding specificities can be conferred by changing this position to a variety of amino acid side chains.

MATERIALS AND METHODS

Glutathione *S*-transferase–HD fusion proteins were created by introducing *Bam*HI and *Eco*RI sites into various homeobox sequences by PCR and cloning them into the pGEX-2T vector (Pharmacia). The fusion protein with the different HDs resulted in the following number of amino acids upstream and downstream, respectively, of the HD: Ftz, 4, 10; Prd HD mutants, 17, 4; Otd, 10, 4; Bk27, 18, 5. Fusion proteins were expressed and purified as described (16).

The random sequence library was prepared by synthesizing the sequence CATGAATTCTCCTATACTGAGTTCATGA-TN_{18}TGATATCGAACTGTATCGATGAATTCCAC and primers that anneal to the first or last 20 bases. The binding site selection was performed as described (17). The random library was mixed with the following amounts of glutathione *S*-transferase–HD fusion protein bound to glutathione-coated beads: Bcd, Bk27, and Prd mutants, 10 ng; Otd, 2 ng; Ftz, 50 ng. DNA bound to the beads was then amplified by PCR. This constituted one "generation" of selection. This amplified DNA was then mixed with the same protein in the second generation. The total number of generations differs between proteins (see *Results*). After the last PCR amplification step, the product is cloned and sequenced.

RESULTS

To compare the DNA binding specificities of a diverse range of metazoan HDs, we determined the optimal DNA binding site consensus (18) for HDs from different classes, using multiple generations of binding selections from a random DNA seqeunce library.

Selection by Diverse Homeodomains of TAAT-Containing Sequences. Five generations of selection with the *Drosophila* Fushi tarazu (Ftz) HD led to the sequencing of 21 oligonucleotides, all of which could be aligned to a single consensus. The consensus sequence, shown in Fig. 1*A*, is **TAATG**. The consensus matches the proposed Ftz binding sites identified through extensive foot-printing (19, 20) and gel shifting (14) of target promoters and conforms to the predicted optimal binding sites determined by scanning mutagenesis (21) of proposed target sites. Sites very similar to these have been shown to mediate *in vivo ftz* activity (22).

The *Drosophila* Bicoid (Bcd) HD selection consisted of five generations. Twenty-eight selected library members were sequenced. All except 4 could be aligned easily, as shown in Fig. 1*B*. The consensus sequence is **TAATCC**. This corresponds well with the consensus sequence (**CTAATCCC**) defined from sites in the hunchback promoter that are believed to mediate the activation of this gene by *bcd in vivo* (20). The site, like that of Ftz, is composed of the nearly universal TAAT motif, but the bases 3′ to the TAAT differ. This difference has been shown to be due to the presence of a lysine, rather than a glutamine, at position 50. The lysine specifies the CC sequence immediately 3′ to the **TAAT** motif (8, 12–15).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: HD, homeodomain.

*Present address: Massachusetts General Hospital, Department of Molecular Biology, Boston, MA 02114.

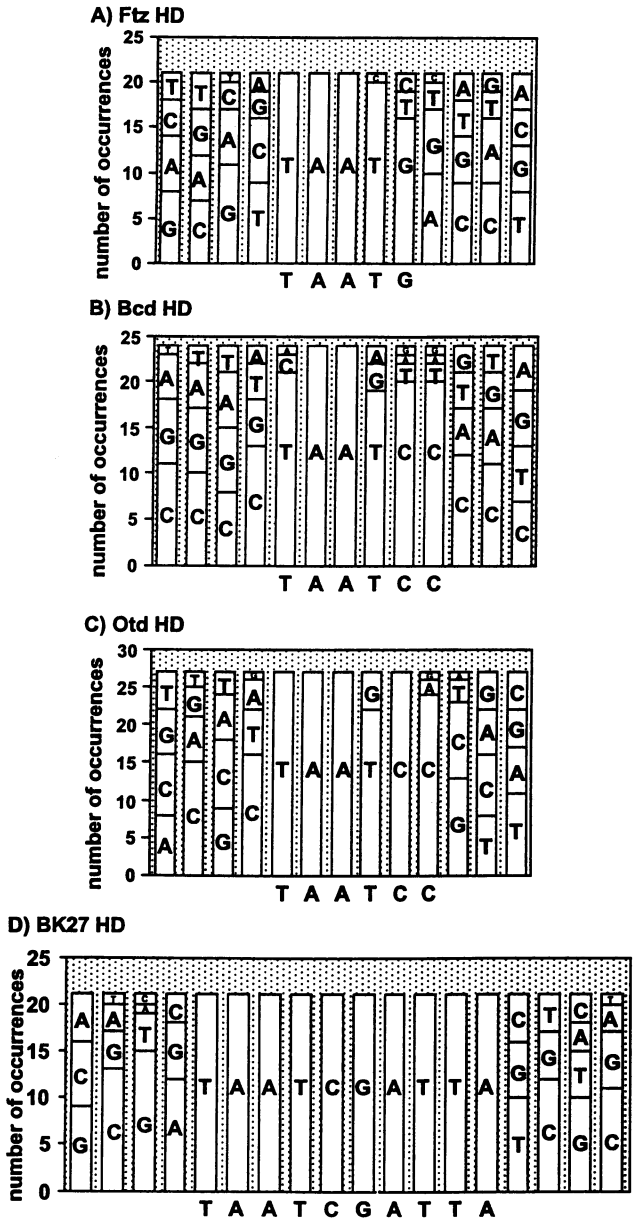

FIG. 1. Optimal DNA binding specificity of the Ftz (*A*), Bcd (*B*), Otd (*C*), and Bk27 (*D*) HDs. The height of each bar represents the number of occurrences of each position along the consensus.

The *Drosophila* Orthodenticle (Otd) HD also has a lysine at position 50. After five generations of selection with this HD, 27 oligonucleotides were sequenced. All of them could easily be aligned and the consensogram is shown in Fig. 1*C*. The optimal sequence of this HD is very similar to that of Bcd, **TAATCC**. There are no known target sites for Otd. However, the Otd protein can activate transcription by binding to the site **TAATCC** upstream of a minimal promoter in cell culture (E. Ronchi and C.D., unpublished results). We have shown (17) that Lys-50 HDs in the Prd class (of which Otd is a member) can cooperatively dimerize on the sequence **TAATCCGATTA**, called P3K. However, among Prd class HDs, those with Lys-50 possess the weakest cooperativity (25-fold). Thus, five generations of selection may not have been sufficient to isolate the palindromic site. For the purposes of this paper, the relevant result is the selection of the TAAT sequence.

HDs in the Prd class that possess a serine at position 50 bind optimally to a sequence that is also palindromic, but has only 2 bp between the inverted TAAT sites ("P2," **TAATYGATTA**). Prd, for instance, binds slightly better to this site than it does to the P3 sequence (17). We determined the optimal binding specificity of the *Drosophila* Bk27 HD, which is in the Prd class and possesses a serine at position 50 (23). After nine generations of selection, 21 oligonucleotides were sequenced, and they all contained exact matches to the sequence **TAATCGATTA**, and moderate-flanking symmetrical preferences are also observed (Fig. 1*D*). No known target genes have been identified for BK27. However, considering the general agreement between the sites selected by the Ftz and Bcd HDs (Fig. 1 *A* and *B*) and their biological target sites, the Bk27 optimal binding consensus should prove valuable in identifying relevant DNA binding sites in candidate target gene promoters of Bk27.

Discrimination Between TAAT-Target Sites by Position 50. The fact that many diverse HDs optimally recognize TAAT-containing target sites suggests that the origin of the functional differences between them will result from DNA contacts outside of this motif. Because lysine and glutamine at position 50 have been shown to confer different binding specificities for the base pairs immediately 3′ to the TAAT sequence, we investigated whether this position was capable of conferring further specificities by undergoing further mutation. We therefore examined the effect of six different amino acids at position 50 of a single HD, that from the Prd protein. We substituted the serine at position 50 (Ser-50) with four amino acids that are present in other HDs, histidine (His-50, present in the Cut HD), isoleucine (Ile-50, present in the a1 HD), lysine (Lys-50, present in the Bcd and Otd HDs, see above), and glutamine (Gln-50, present in most HDs), and with one amino acid, asparagine (Asn-50), that has not been found in any other HDs but is stereochemically related to glutamine.

After selection by the Prd His-50 HD for seven generations, we sequenced 34 oligonucleotides, 32 of which could be aligned unambiguously. A sequence resembling P3 was selected (**TAATTAGRTTA**, Fig. 2*A*). However, the sequence differs in two respects from the P3 sequence selected by the Prd (Ser-50) HD. (*i*) One of the TAAT sequences has been compromised to TAAY. (*ii*) There is a different specificity at the central base pairs, as TAG rather than YNR is selected. Note that the sequence is inherently asymmetrical; that is, the presence of a thymidine at position 5 of the consensus is nearly always correlated with a guanosine at position 7, for instance. Selection with Prd Ile-50 for eight generations yielded yet another specificity. All 28 of the selected oligonucleotides could be aligned according to the consensus shown in Fig. 2*B*. Note that while one of the TAAT motifs is preserved, the other is altered to TAAC. As in the case of the Prd His-50 selection, the central-most five positions in the binding site are asymmetric, and the asymmetrical positions are, again, strongly correlated with each other. Seven generations of selection with Prd Asn-50 resulted in the unambiguous alignment of 34/37 sequenced random library members. Again, a site resembling P3 is selected. However, the base pairs between the two TAAT motifs are altered to yet another specificity. The optimal DNA binding selections with Prd Ser-50, Gln-50, and Lys-50 HDs have been reported (17). The selections on Prd HD variants are summarized in Table 1.

DISCUSSION

Determinants for Recognition of the Quasi-Universal TAAT Motif. Early biochemical investigations of HD–DNA interaction showed that most metazoan HDs can recognize TAAT-containing DNA sequences (for review, see refs. 6–8). In principle, this could have reflected an inherent similarity in the DNA binding specificity of all these HDs, with an implied structural conservation at the protein–DNA interface. Alternatively, these HDs could each have had unique DNA binding specificities but could have possessed the tendency to cross react with each other's preferred DNA binding sites. To

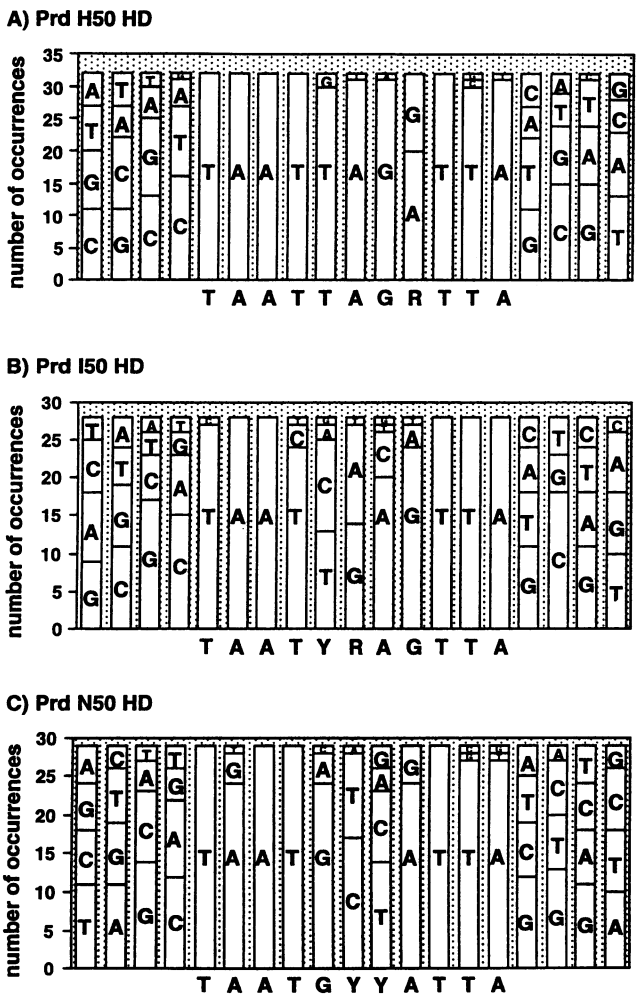

FIG. 2. Optimal DNA binding specificity of the Prd HD with histidine (*A*), isoleucine (*B*), or asparagine (*C*) at position 50. The height of each bar represents the number of occurrences of each position along the consensus.

distinguish between these two possibilities, we and others have determined the optimal DNA binding specificities for several HDs. We herein consolidate the data from the selections described above and elsewhere to define a "TAAT-binding consensus" of amino acids. A similar consensus, based on the same principle and a less diverse set of HDs, has been proposed previously (24), and other approaches toward identifying such amino acids have also been reported (1, 4, 6, 23). Fig. 3*A* shows an alignment of HDs that optimally recognize TAAT-containing sequences. This represents a diverse set of metazoan HDs from several different classes. Our assumption is that only those amino acids that are involved in folding of the HD, binding to the DNA backbone, or recognizing the 4 bp of the TAAT motif, will be conserved among this set of HD sequences. Below the list of sequences in Fig. 3*A* is the amino acid conservation.

Table 1. Binding sequences for HDs with various residues 50

Residue 50	Optimal binding sequence
S	**TAAT** Y G **ATTA**
Q	**TAAT** Y N R **ATTA**
K	**TAAT** C C G **ATTA**
H	**TAAT** T A G **RTTA**
I	**TAAT** Y R A **GTTA**
N	**TAAT** G Y Y **ATTA**

Of the 60 amino acids of the HD, 22 are common to HDs that optimally recognize the TAAT core site, thus defining the TAAT-binding consensus of amino acids. The structural basis for the conservation of these amino acids can be deciphered by examining the crystal and NMR structures of HD–DNA complexes. To date, the structures of several such complexes have been determined—Engrailed (En) (4), Antennapedia (Antp) (25), the $\alpha 2/\mathbf{a}1$ heterodimer (5, 11), Oct-1 (9), and Prd Gln-50 (10). All of these except $\alpha 2/\mathbf{a}1$ have been shown to bind optimally to TAAT-containing sequences (Fig. 3*A*). Based on these crystal and NMR studies, the amino acids conserved in the TAAT-binding consensus can be assigned to three different functions: contributing to the hydrophobic core, contributing to DNA backbone recognition, and contributing to the specific recognition of the bases making up the TAAT subsite. Some residues are important for more than one of these reasons.

The consensus includes 12 positions that make up the hydrophobic core (positions 8, 16, 17, 20, 26, 31, 34, 40, 45, 48, 49, and 52) and 8 positions that are involved in making DNA phosphate contacts (positions 3, 25, 31, 48, 53, 55, 57, and 58) (Fig. 3*A*). Residues that contact the bases making up the TAAT motif are in the N-terminal arm (positions 2, 3, and 5) and the recognition helix (positions 47 and 51). The base-specific contacts made by the TAAT-binding consensus residues are shown in Fig. 3*B*. We refer to base-pair positions in the **TAAT**NNN motif as numbering 1–7, 5′ to 3′, where the TAAT-containing strand is referred to as the "top strand." Residues in the N-terminal arm make sequence-specific contacts in the DNA minor groove. All crystallographic structures of metazoan HDs on TAAT sites show that the thymidine (O2) of the first base pair receives a hydrogen bond from Arg-5. The second base pair (O2 of thymidine on bottom strand) of the TAAT motif receives a hydrogen bond from the side chain of Arg-3 in the En structure. In both the Oct-1 HD and the Prd Gln-50 HD structures, however, an arginine at position 2, rather than 3, achieves this contact, demonstrating that there are alternative ways of making similar sequence-specific interactions in this extended protein region (Fig. 4). From the TAAT-binding consensus alignment (Fig. 3*A*), it appears that the specification of this second base pair can be achieved by either an arginine or lysine residue at either position 2 or 3 of the HD. Positions 2 and/or 3 may also utilize water-mediated hydrogen bonds to specify bases at positions 3 and 4 (Fig. 3*B*). It is unclear how any of these contacts in the minor groove consistently differentiate T·A versus A·T base pairs, since the spatial positioning of hydrogen bond acceptors (the O2 of thymidine and the N3 of adenosine) in the two cases is very similar.

This consensus, like those proposed previously (1, 4, 6, 23, 24), suggests that positions 47 and 51 in the recognition helix are also important for interaction with the TAAT motif. All HD–DNA structures solved to date show a hydrophobic interaction between the side chain of position 47 (isoleucine or valine) and the methyl group of thymidine 4. There are two exceptions in Fig. 3*A* to the presence of isoleucine or valine at position 47, and Pomerantz *et al.* (26) have shown that several structurally diverse side chains at this position allow for the specification of bp 4 as a thymidine. Several other side chains at position 47 may be capable of making similar hydrophobic/van der Waals contacts with the thymidine at position 4. Alternatively, amino acid residue at other positions may specify the identity of this base pair through water-mediated contacts (Fig. 3*B*). All of the HD–DNA crystal structures show a bidentate hydrogen bond between Asn-51 and the adenosine at position 3, and Pomerantz *et al.* (26) provided evidence that Asn-51 cannot be mutated to any other residue without severe reduction in binding affinity.

One prediction of the above comparative analysis is that most animal HDs possess recognition sequences containing a TAAT motif, as suggested by the fact that the majority of known metazoan HDs have all of the residues that make up the

A

	helix I			helix II			helix III		Reference:
	1	11	21	31		41	51		
OTD	QRRERTTFTR	AQLDVLEALF	GKTRYPDIFM	REEVALKINL		PESRVQVWFK	NRRAKCRQQ		this study
Paired	QRRCRTTFSA	SQLDELERAF	ERTQYPDIYT	REELAQRTNL		TEARIQVWFS	NRRARLRKQ		17
Pax-6	LQRNRTSFTA	QQIEELEKEF	ERTHYPDVFA	RERLAQKIDL		PEARIQVWFS	NRRAKWRRE		29
S8	QRRNRTTFNS	SQLQALERVF	ERTHYPDAFV	REELARRVNL		SEARVQVWFQ	NRRAKFRRN		30
Bk27	RRRNRTTFSP	EQLEELEKEF	DKSHYPCVST	RERLSSRTSL		SEARVQVWFS	NRRAKWRRE		this study
Ftz	SKRTRQTYTR	YQTLELEKEF	HFNRYITRRR	RIDIANALSL		SERQIKIWFQ	NRRMKSKKD		this study
Ubx	RRRGRQTYTR	YQTLELEKEF	HTNHYLTRRR	RIEMAHALCL		TERQIKIWFQ	NRRMKLKKE		31
Dfd	PKRQRTAYTR	HQILELEKEF	HYNRYLTRRR	RIEIAHTLVL		SERQIKIWFQ	NRRMKWKKD		24
Antp	RKRGRQTYTR	YQTLELEKEF	HFNRYLTRRR	RIEIAHALCL		TERQIKIWFQ	NRRMKWKKE		32
Oct-1	RRKKRTSIET	NIRVALEKSF	LENQKPTSEE	ITMIADQLNM		EKEVIRVWFC	NRRQKEKRI		33
Bicoid	PRRTRTTFTS	SQIAELEQHF	LQGNRYLAPR	LADLSAKLAL		GTAQVKIWFK	NRRRRHKIQ		this study
mClox	LKKPRVVLAP	EEKEALKRAY	QQKPYPSPKT	IEELATQLNL		KTSTVINWFH	NYRSRIRRE		34
Isl-1/2	TTRVRTVLNE	KQLHTLRTCY	AANPRPDALM	KEQLVEMTGL		SPRVIRVWFQ	NKRCKDKKR		35, 36
Prh	RKGGQVRFSN	EQTIELEKKF	ETQKYLSPPE	RKRLAKLLQL		SERQVKTWFQ	NRRAKWRRL		37
En	EKRPRTAFSS	EQLARLKREF	NENRYLTERR	RQQLSSELGL		NEAQIKIWFQ	NKRAKIKKS		38
TAAT-Binding Consensus	**BB R** O	LZ f	YO B	B O	O	I **I**WF V **V**	**N**BR B BB		

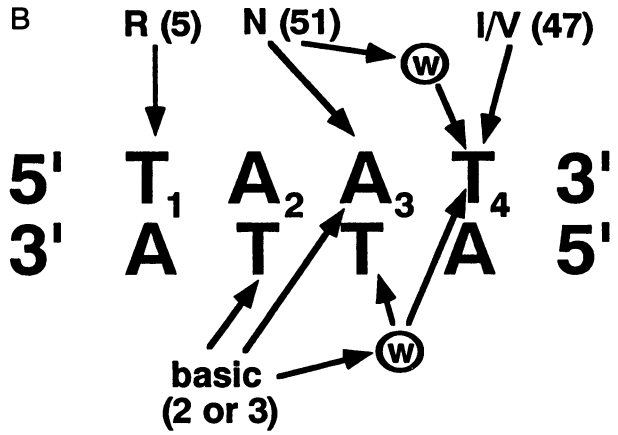

FIG. 3. (*A*) Alignment of sequences of HDs that have been shown to optimally recognize TAAT-containing sequences. Below the list of HDs is the "TAAT-binding consensus of amino acids." Residues in the consensus that make DNA phosphate contacts are underlined and those that contact bases making up the TAAT motif are in boldface type. All remaining positions of this consensus contribute to the hydrophobic core. Positions 31 and 48, in addition to contacting DNA phosphates, also contribute to the hydrophobic core. B, arginine or lysine; O, hydrophobic; Z, charged; f, phenylalanine or tyrosine; i, isoleucine or leucine. (*B*) Base-specific contacts between amino acids represented in the TAAT-binding consensus and base pairs making up the TAAT motif. Water molecules are represented by an encircled w. The water-mediated contacts have only been observed in the high-resolution crystal structure of the Prd Gln-50 HD–TAAT complex but are assumed to be of general importance since the positioning of the side chains and bases is very similar in other known HD–TAAT complexes investigated crystallographically.

TAAT-binding consensus. This generalization was also drawn by Kalionis *et al.* (23), who screened a *Drosophila* expression library for proteins that could interact with a TAAT-containing oligonucleotide. HD clones representing several different HD classes were thus obtained, showing that a diverse range of metazoan HD proteins possess related DNA binding specificities. The authors suggested that the origin of this relatedness lies in the complex interactions between HD proteins in genetic networks, analogous to the genetic interactions between λ Repressor and Cro proteins, which must bind to similar sequences to function together.

Changing the Specificity of Binding to Different TAAT-Containing Sites. Although most HDs recognize a TAAT core sequence, they have evolved different DNA binding specificities by altering amino acid position 50, which can interact with bp 5 and 6 (for review, see ref. 8). We have shown that each of six amino acids tested confers a different DNA binding specificity (Table 1). The HD appears to have utilized the potential for this position to confer different DNA binding preferences, as judged by its polymorphism throughout evolution (alanine, cysteine, glutamate, glycine, histidine, isoleucine, lysine, glutamine, arginine, and serine occur at this position in different HDs), and by its conservation between homologous HD sequences from different organisms (1).

How can position 50 influence the DNA binding specificity of HDs? In the five HD–DNA interfaces that have been investigated by x-ray crystallography, no direct hydrogen bonds have been observed between this residue and the bases. In the En (Gln-50), Prd (Gln-50), Oct-1 (Cys-50), α2 (Ser-50), and **a**1 (Ile-50) structures, however, position 50 is observed to make van der Waals contacts with bases at positions 5, 6, and/or 7. The high-resolution structures of the Prd (Gln-50) and α2 (Ser-50) HDs show water-mediated contacts between position 50 and bp 4 and/or 5. Such water-mediated contacts were also identified in the NMR solution studies of the Antp (Gln-50) HD–DNA complex (25, 27). Thus, these structural studies suggest that, depending on its identity, position 50 can be involved in specifying positions 4–7, which is identical to the

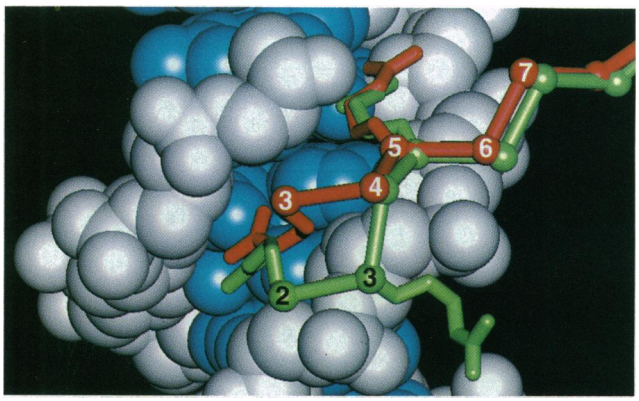

FIG. 4. Alternative means of recognizing bp 2. The crystal structure of the Prd Gln-50 HD bound to a TAAT-containing oligonucleotide is shown, with the DNA base pairs in blue and the DNA backbone in white (space-filling model). The "top strand" (which contains the TAAT sequence) is on the right, with the 5′-to-3′ direction running from the top to the bottom. The green stick figure represents the backbone of the Prd Gln-50 HD from residues 2 to 8, with C-α positions shown as spheres that are connected by sticks. In addition, all nonhydrogen atoms are displayed as stick figures for arginine side chains of residues 2, 3, and 5. In red is shown the N-terminal arm of the En HD after least squares superimposition of the TAAT motifs of the En HD–DNA structure and the Prd Gln-50 HD–DNA structure. C-α positions from amino acids 4–58 are nearly superimposable for the two structures. Both the En and Prd HDs show very similar contacts between the guanido group of Arg-5 and the O2 atom (behind and slightly to the left of the Prd Gln-50 HD Arg-5 guanido group) of the thymidine on the top strand at bp 1. However, the two HDs utilize different amino acid positions to contact bp 2. In the case of the En HD, the guanido group of the side chain of Arg-3 donates a hydrogen bond to the O2 atom (immediately to the left of the label 3 on the En HD) of the thymidine at bp 2 (bottom strand). In the case of Prd, however, an arginine at position 2, rather than 3, makes the analogous contact with bp 2. This results in the arginine at position 3 assuming a completely different conformation, where it makes a water-mediated contact with a phosphate. The conformation of the Oct-1 HD resembles that of the Prd HD.

range of base positions influenced by mutations at position 50 in the genetic studies reported herein. A model for the interaction of Lys-50 with its preferred base pairs at positions 5 and 6 has been offered (15).

An explanation for the ability of this single amino acid position to confer such a variety of distinct DNA binding preferences must lie in its particular structural relationship with the DNA bases: this region of the recognition helix is not in intimate contact with the bases of the major groove but separated from them by an appreciable distance. This relatively unrestrictive spatial relationship allows for a diverse range of side chains to be accommodated, each of which favors a particular sequence of DNA bases to which it can form relatively long range (i.e., van der Waals or water-mediated hydrogen bonding) contacts. This is in contrast to position 51, for example, which is universally conserved as an asparagine and forms more intimate, stereochemically precise contacts with the floor of the DNA major groove, which may account for the intolerance of this position to amino acid substitution (26).

The intriguing asymmetric nature of the optimal DNA binding sites for the HDs examined herein demonstrates that the details of the HD recognition helix–major groove interface are influenced by the HD bound to the adjacent TAAT half site. This fact can be appreciated by noting that Ile-50 of the **a**1 HD has been shown to interact with bp 7 of its DNA half site, and this would correspond to bp 5 of the second TAAT half site in the sequence selected by the Prd Ile-50 HD. Thus, the two half sites overlap. Furthermore, in the crystal structure of the Prd Gln-50 HD dimer bound to the **TAATYNRATTA** site, water-mediated contacts between the two HDs, in the vicinity of the recognition helix–major groove interface, have been directly observed. Such interactions between the two HDs at the DNA recognition interface could account for the influence of one HD on the binding preference of the other and for the resulting correlated asymmetry of the central positions in the optimal consensus sequences. Indeed, given such interactions (either direct or water mediated) between the two proteins near the center of the palindromic recognition sequence, there is no reason *a priori* why the optimal arrangement of side-chain conformations and DNA bases need be symmetrical. An asymmetrical binding preference has also been reported for the λ repressor dimer (28).

We thank Yan Gu for excellent technical assistance. Thanks to John Kuriyan and members of the Desplan lab for many important discussions. D.W. was supported by National Research Service Award Training Grant GM-07982-12.

1. Gehring, W. J., Affolter, M. & Bürglin, T. (1994) *Annu. Rev. Biochem.* **63,** 487–526.
2. Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J. & Wüthrich, K. (1989) *Cell* **59,** 573–580.
3. Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990) *EMBO J.* **9,** 3085–3092.
4. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990) *Cell* **63,** 579–590.
5. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. (1991) *Cell* **67,** 517–528.
6. Laughon, A. (1991) *Biochemistry* **30,** 11357–11367.
7. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G. & Wüthrich, K. (1994) *Cell* **78,** 211–223.
8. Treisman, J., Harris, E., Wilson, D. & Desplan, C. (1992) *BioEssays* **14,** 145–150.
9. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. & Pabo, C. O. (1994) *Cell* **77,** 21–32.
10. Wilson, D. S., Guenther, B., Desplan, C. & Kuriyan, J. (1995) *Cell* **82,** 709–719.
11. Li, T., Stark, M. R., Johnson, A. D. & Wolberger, C. (1995) *Science* **270,** 262–269.
12. Hanes, S. D. & Brent, R. (1989) *Cell* **57,** 1275–1283.
13. Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. & Desplan, C. (1989) *Cell* **59,** 553–562.
14. Percival-Smith, A., Muller, M., Affolter, M. & Gehring, W. J. (1990) *EMBO J.* **9,** 3967–3974.
15. Hanes, S. D. & Brent, R. (1991) *Science* **251,** 426–430.
16. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1991) *Current Protocols in Molecular Biology* (Wiley, New York).
17. Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. & Desplan, C. (1993) *Genes Dev.* **7,** 2120–2134.
18. Thiesen, H.-J. & Bach, C. (1990) *Nucleic Acids Res.* **18,** 3203–3209.
19. Pick, L., Schier, A., Affolter, M., Schmidt-Glenewinkel, T. & Gehring, W. J. (1990) *Genes Dev.* **4,** 1224–1239.
20. Nelson, H. B. & Laughon, A. (1990) *New Biol.* **2,** 171–178.
21. Florence, B., Handrow, R. & Laughon, A. (1991) *Mol. Cell. Biol.* **11,** 3613–3623.
22. Schier, A. F. & Gehring, W. J. (1992) *Nature (London)* **356,** 804–807.
23. Kalionis, B. & O'Farrell, P. H. (1993) *Mech. Dev.* **43,** 57–70.
24. Ekker, S. C., von Kessler, D. P. & Beachy, P. A. (1992) *EMBO J.* **11,** 4059–4072.
25. Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. & Wüthrich, K. (1993) *J. Mol. Biol.* **234,** 1084–1097.
26. Pomerantz, J. L. & Sharp, P. A. (1994) *Biochemistry* **33,** 10851–10858.
27. Qian, Y. Q., Otting, G. & Wüthrich, K. (1993) *J. Am. Chem. Soc.* **115,** 1189–1190.
28. Beamer, L. J. & Pabo, C. O. (1992) *J. Mol. Biol.* **227,** 177–196.
29. Czerny, T. & Busslinger, M. (1995) *Mol. Cell. Biol.* **15,** 2858–2871.

30. de Jong, R., van der Heijden, J. & Meijlink, F. (1993) *Nucleic Acids Res.* **21,** 4711–4720.
31. Ekker, S. C., Young, K. E., von Kessler, D. P. & Beachy, P. A. (1991) *EMBO J.* **10,** 1179–1186.
32. 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.
33. Verrijzer, C. P., Alkema, M. J., wan Weperen, W. W., van Leeuwen, H. C., Strating, M. J. J. & van der Vliet, P. C. (1992) *EMBO J.* **11,** 4993–5003.
34. Andrés, V., Chiara, M. D. & Mahdavi, V. (1994) *Genes Dev.* **8,** 245–257.
35. Sanchez-Garcia, I., Osada, H., Forster, A. & Rabbitts, T. H. (1993) *EMBO J.* **12,** 4243–4250.
36. Gong, Z. & Hew, C. L. (1994) *Biochemistry* **33,** 15149–15158.
37. Crompton, M. R., Bartlett, T. J., MacGregor, A. D., Manfioletti, G., Buratti, E., Giancotti, V. & Goodwin, G. H. (1992) *Nucleic Acids Res.* **20,** 5661–5667.
38. Ades, S. E. & Sauer, R. T. (1994) *Biochemistry* **33,** 9187–9194.