Nucleic Acids Research, 1994, Vol. 22, No. 15 3075–3083

Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain

G.Damante*, D.Fabbro, L.Pellizzari, D.Civitareale[1], S.Guazzi[1], M.Polycarpou-Schwartz[1], S.Cauci, F.Quadrifoglio, S.Formisano and R.Di Lauro

Dipartimento di Scienze e Tecnologie Biomediche, Universita di Udine, Via Gervasutta 48, Udine, Italy and [1]European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg, Germany

Received April 29, 1994; Revised and Accepted June 20, 1994

ABSTRACT

The molecular basis for the DNA binding specificity of the thyroid transcription factor 1 homeodomain (TTF-1HD) has been investigated. Methylation and ethylation interference experiments show that the TTF-1HD alone recapitulates the DNA binding properties of the entire protein. Studies carried out with mutant derivatives of TTF-1HD indicate a precise correspondence of some of its amino acid residues with specific bases in its binding site, allowing a crude orientation of the TTF-1HD within the protein – DNA complex. TTF-1HD shows an overall geometry of interaction with DNA similar to that previously observed for Antennapedia class HDs, even though the binding specificities of these two types of HDs are distinct. We demonstrate that the crucial difference between the binding sites of Antennapedia class and TTF-1 HDs is in the motifs 5′-TAAT-3′, recognized by Antennapedia, and 5′-CAAG-3′, preferentially bound by TTF-1. Furthermore, the binding of wild type and mutants TTF-1 HD to oligonucleotides containing either 5′-TAAT-3′ or 5′-CAAG-3′ indicate that only in the presence of the latter motif the Gln_{50} in TTF-1 HD is utilized for DNA recognition. Since the Gln at position 50 is an essential determinant for DNA binding specificity for several other HDs that bind to 5′-TAAT-3′ containing sequences, we suggest that utilization by different HDs of key residues may depend on the sequence context and probably follows a precise hierarchy of contacts.

INTRODUCTION

The homeodomain (HD), which was first identified as a protein sequence encoded by a highly conserved segment of DNA (the homeobox), is present in several eukaryotic nuclear proteins and is capable of sequence specific DNA binding. HD-containing proteins play important roles in regulating developmental programs and appear to function as transcriptional regulators (1–2). In the majority of the cases examined the DNA-binding specificity of HD-containing proteins resides exclusively on the HD itself. In some cases however, other domains, including POU domains, leucine zippers, LIM domains and zinc-fingers, are able to cooperate with the HD in determining the DNA-binding specificity of the entire protein (3–6). Recently, the three-dimensional structure of the HDs present in the *Antennapedia*, *Engrailed* and *MAT-α2* gene products, as well as the complexes with their cognate DNA sequences, have been obtained by NMR spectroscopy and crystallographic studies (7–9). These structures have demonstrated that HDs contain three α-helices identified, from the amino-terminus, as helix I, II and III. Helices II and III form the helix-turn-helix (H-T-H) motif which is also found in many prokaryotic transcriptional regulatory proteins. However, both structural and functional studies revealed that HDs and prokaryotic H-T-H motifs significantly differ in the mode of contacting their target DNA sequences since: i) HDs appear to bind target sequences as monomers, rather than dimers (2, 10); ii) in the recognition helix (helix III of HDs, corresponding to the second helix in the H-T-H motif) residues important in establishing the DNA-binding specificity are located at the carboxyl-terminus in the case of HDs, at the amino-terminus for prokariotic repressors (11–13); iii) amino acids of the HD outside the H-T-H motif contact DNA (8–9, 14) and seem to be important in determining the DNA-binding specificity (15).

The molecular basis by which HDs are able to recognize specific DNA sequences have been extensively investigated. Helix III has been recognized as very important in determining the DNA-binding specificity of HDs. When the HD is complexed with DNA this helix fits into the major groove and some of its amino acids make contacts with base pairs of the cognate DNA sequence. Among these amino acids, one (Asn_{51}) forms a set of hydrogen bonds with a very much conserved adenine in the binding site (7–9). Since Asn_{51} is present in all HDs isolated up to now it is conceivable that it contributes mostly to the correct positioning of the helix on the major groove and not to the DNA-binding specificity. Residues 47, 50 and 54 are much less conserved and their variability presumably contributes to the DNA-binding specificity. Functional studies have revealed that residue 50 is a very critical amino acid for controlling specificity (11–14). Its identity has been shown to affect base preferences in both *in vitro* and *in vivo* assays. Moreover, there are several examples where also regions of the HD outside the recognition

*To whom correspondence should be addressed

helix contribute to the DNA-binding specificity (15–17). In several cases the N-terminal sequence of the HD was discovered to be critical for the *in vivo* target selection (18–20). It is not clear however, whether residues of the amino-terminal arm important for the *in vivo* target selection, operate through modification of the DNA-binding specificity or through other mechanisms (See ref. 21).

The thyroid transcription factor 1 (TTF-1) (22) contains a HD (TTF-1HD) whose DNA-binding specificity is to a large extent dictated by sequences outside of its recognition helix (15). In this paper the DNA contacts made by the entire TTF-1 and by its HD on a high affinity binding site have been determined. The data indicate that TTF-1 binding to DNA is only due to the HD moiety. Furthermore, we demonstrated that the geometry of contacts made by TTF-1HD is very similar to that of Antennapedia class HDs, indicating that the different DNA-binding specificities of these two types of HDs do not reflect a substantially different mode of binding. The relative binding affinity of TTF-1 and Antennapedia HDs for mutant oligonucleotides indicate that the critical difference between the recognition sites lies in the motifs 5′-CAAG-3′, recognized by TTF-1, and 5′-TAAT-3′, bound by Antp. Finally, we provide evidence suggesting that the usage of Gln_{50} by the TTF-1 HD is subordinate to contacts made with the 5′-CAAG-3′ core, since this residue does not seem to play any role when a 5′-TAAT-3′ motif is present.

MATERIALS AND METHODS

Solutions and buffers

Solution I: 10 mM Hepes pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM EGTA, 0.35 M sucrose, 0.5 mM DTT, 0.5 mM PMSF. Solution II: 10 mM Hepes pH 7.9, 400 mM NaCl, 1.5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT, 0.5 mM PMSF, 5% glycerol. Solution D: 20 mM Hepes pH 7.9, 0.1 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT, 20 % glycerol. Solution E is solution D with 20 mM Tris–HCl pH 8.5 substituting for the Hepes. Solution F is solution E containing 0.02% NP-40. The KCl molarity in solutions D, E and F is indicated by subscripts. Buffer Q: 20 mM Tris–HCl pH 7.4, 0.4 M NaCl, 1.0 mM EDTA, 5mM DTT, 10% glycerol, 1mM PMSF. Buffer E: 20 mM Tris–HCl pH 7.4, 0.8 M NaCl, 1.0 mM EDTA, 5mM DTT, 10% glycerol, 1mM PMSF.

Oligodeoxynucleotide synthesis and purification

Oligodeoxynucleotides were synthesized on a automated DNA synthesizer Model 380B (Applied Biosystem) according to standard procedures and purified by FPLC on MonoQ column (Pharmacia) utilizing an ammonium bicarbonate gradient. The purity of oligonucleotides was controlled on a 20% acrylamide/7M urea gel electrophoresis.

Preparation of cell and calf thyroid nuclear extract and purification of TTF-1

Calf thyroids were homogenized in three volumes of Solution I with a Polytron (5×7″ bursts) and filtered through two layers of gauze. NP-40 was then added to 0.4% and the nuclei pelleted by centrifugation for 10' at 4000 rpm in the JS 4.2 rotor of the J6 Beckman centrifuge. The supernatant was decanted and the nuclei resuspended in Solution II (5×the pellet volume). The conductivity was adjusted to the equivalent of 0.4 M NaCl and, after stirring for 30′, the suspension was centrifuged for 50′ at 35000 rpm (in the rotor 45Ti.) (the supernatant of this step will be referred to as Crude Nuclear Extract). Solid ammonium sulfate was added to the Crude Nuclear Extract (0.33 g/ml) and the precipitate was collected by centrifugation at 10000 rpm (in the Sorvall GSA rotor). The protein pellet was dissolved in Solution $D_{0.3}$ (1/10th of the crude extract volume) and dialyzed against 5 liters of the same solution for 6–12 hours. The dialyzed extract was frozen in liquid nitrogen and stored at −80°C. Nuclear extracts were loaded on a monoQ column (Pharmacia) (60 ml) connected in series with a similarly sized column of Blue Sepharose (Pharmacia). Both columns were equilibrated in Solution $D_{0.3}$. After loading, the two columns were washed with 300 ml of solution $D_{0.3}$. The Blue-Sepharose column was then disconnected from the MonoQ, washed with an additional 200 ml aliquot of Solution $D_{0.3}$ and eluted with Solution E_1. The protein peak eluting with the E_1 solution was collected, dialyzed against buffer $F_{0.1}$ and loaded on a 30 ml DNA Sepharose column (Pharmacia) (1.5×17 cm). The column was washed with 120 ml $F_{0.1}$ and eluted with two subsequent linear salt gradients in Solution F. The first gradient (150 ml) was from 0.1 to 0.3 M KCl and the second (150ml) from 0.3 to 1.5 M KCl. Fractions (4ml) were assayed for TTF-1 activity by a gel-retardation assay using as a substrate the oligonucleotide C (see below). Active fractions were pooled, diluted with buffer F_0 to a KCl concentration of 0.2 M, supplemented with poly d(I-C) at 2.5 μg/ml and loaded on a 10 ml oligo C affinity column (1×14 cm), equilibrated in Solution $F_{0.2}$. The column was washed with 10 ml of Solution $F_{0.2}$ and eluted with a 50 ml linear KCl gradient (0.2 to 1.5 M in solution F). The active fractions were pooled together and purified again through the affinity chromatography step, omitting in this case the poly d(I-C). The active fractions after the affinity step were pooled, dialyzed against $F_{0.2}$, frozen in liquid nitrogen and stored at −80°C.

Preparation of nuclear extracts from FRTL-5 cells was done as previously described (23).

Construction and expression of TTF-1HD mutants

PT7.7HD, was obtained cloning the DNA sequence encoding the TTF-1HD into the plasmid pT7.7 (24). This construct contains the T7 RNA polymerase promoter and its structure is described in ref. 15. The Antennapedia HD expression vector (the plasmid pAop2, described in ref. 25) was a gift of M.Müller. The TTF-1HD(6NΔ) is a deletion mutant of TTF-1HD lacking the first 6 residues at the amino-terminus of TTF-1HD (amino acid positions in the HD are numbered according to Ref. 26). TTF-1HD (K_{50}) is a mutant where the glutamine at position 50 was changed with a lysine. Both mutants were constructed using polymerase chain reaction according to the procedure of Ho *et al.* (27). The sequences of oligonucleotides used for construction of TTF-1HD(6NΔ) and TTF-1HD (K_{50}), respectively, were:

5′-CCCGGGATATGCTCTTCTCCCAGGCGCAGGTG-3′ and
5′-GTGAAGATCTGGTTCAAGAACCACCGCTAC-3′.

After cloning the DNA into the expression vector pT7.7, mutations were verified by nucleotide sequencing. TTF-1HD, TTF-1HD(6NΔ) and TTF-1HD(K_{50}) were expressed in BL21 (DE3) cells (24). After growth and production of recombinant proteins, bacteria were pelleted and stored at −80°C. In order to extract the recombinant proteins, bacterial pellets were resuspended in a ice-cold Buffer Q and sonicated on ice. The resulting material was centrifuged (17,500 g, 4°C, 20′), the

supernatant was collected and stored at −80°C. Homeodomains were then partially purified using Econo-Pac S cartridges (Bio-Rad). Crude bacterial extracts (in Buffer Q) were loaded on the column, washed with 6 column volumes of Buffer A and eluted with Buffer E. The concentration of the partially purified HDs was determined by oligonucleotide saturation assay. A gel-retardation assay was performed without poly d(I-C) and increasing amount of oligonucleotides (from 0.3 to 50 nM), then the values of protein-bound and free DNA obtained were used in a Scatchard plot analysis (28).

Gel-retardation assay and interference studies

Double-stranded oligodeoxynucleotides, labelled at the 5′ end terminal with ^{32}P, were used as probes in the gel-retardation assays. The C site is a 24mer whose top strand sequence is 5′-CACTGCCCAGTCAAGTGTTCTTGA-3′. The BS2 site is a 18 mer whose top strand sequence is 5′-GAGAAAAAGCCATTAGAG-3′. Mutants of the C and BS2 described in the Results section were based on wild-type oligodeoxynucleotides thus, C and BS2 mutants were 24mer and 18mer respectively. Gel-retardation assays were performed as described (15) but the KCl concentration was raised to 150 mM and poly d(I-C) was included to a final concentration of 50 $\mu g/\mu l$. Quantitation of the autoradiographic signals was performed by densitometric scanning of the autoradiograms using a LKB laser densitometer. The dissociation constants of protein/DNA complexes shown in Fig. 2 were calculated by Scatchard plot analysis (28). DNA-binding activities shown in Figs. 5, 6 and 7, were determined by measuring the protein-bound/free DNA ratio in gel-retardation assays. In all of these experiments proteins and oligonucleotides were used at the concentration of 5×10^{-7}M (unless otherwise noted) and 10^{-8}M, respectively. Results are always expressed as a fraction of the protein-bound/free DNA ratio found for the wild-type situation which in each figure has been set arbitrarily to 1. Methylation interference experiments were done according to Ref. 29, using as probes dimethylsulphate-treated oligonucleotides. Ethylation interference experiments were done using ^{32}P-labelled oligo C as probe treated with ethyl-nitrosourea and subsequently cleaved at modified phosphates as described (30). In both methylation and ethylation interference studies, protein-bound and free DNAs were separated by preparative PAGE and recovered as described (30). After the chemical cleavage the products were separated on 20% denaturing polyacrilamide gels and visualized by autoradiography.

RESULTS

The TTF-1HD contacts DNA in a manner very similar to the native TTF-1 protein

To characterize in detail the interaction of TTF-1 with its recognition site, methylation and ethylation interference experiments were carried out on the oligonucleotide C (see Fig. 1d), derived from the TATA proximal TTF-1 binding site of the rat thyroglobulin promoter (23). These experiments were performed using as a source of TTF-1 protein: i) crude nuclear extracts prepared from the thyroid cell line FRTL-5 (FRTL-5 NE); these extracts contain full size TTF-1, as judged by molecular weight determination from western blot analysis (data not shown); ii) bovine TTF-1, purified from calf thyroid glands (Pur.bTTF-1, see Materials and Methods) and iii) purified, recombinant, TTF-1HD; (see Materials and Methods). As shown in Fig. 1 (panels a and b), the methylation interference pattern obtained with TTF-1HD was identical to that obtained with purified bTTF-1 or FRTL-5 nuclear extracts. Also the phosphate groups of the DNA backbone are contacted in a similar manner by all three TTF-1 preparations (Fig. 1c), even though TTF-1HD seems to be more efficient in protecting an additional phosphate residue in the bottom strand of the recognition site. A summary of methylation and ethylation interference data is presented in Fig. 1d. For the remainder of this paper the nucleotides recognized by TTF-1 will be referred to using the numbering system illustrated in Fig. 1d.

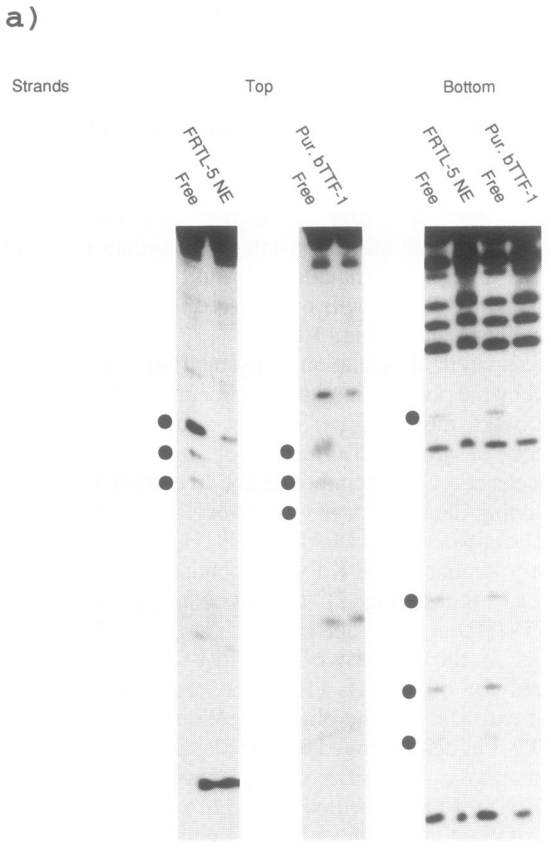
The apparent dissociation constant (Kd) of the complexes formed by the oligonucleotide C with either TTF-1HD or native TTF-1 were measured performing a binding reaction in the presence of poly d(I-C) and of increasing amounts of oligonucleotide C. Protein-bound and free DNA values, obtained by scanning densitometry, were used for a Scatchard analysis (28). The apparent Kd value for TTF-1HD/C complex (2.6 nM), was very similar to the Kd value obtained for native TTF-1/C complex (3.0 nM) (Fig. 2). The Kd obtained in these conditions for TTF-1HD is about one order of magnitude higher than that obtained in the absence of poly d(I-C) (31), as previously observed for other HDs (10). The complex seen incubating nuclear extracts with the oligonucleotide C is solely due to full size TTF-1 as demonstrated by: the identical mobility, in the gel-retardation assay, of the DNA complex with FRTL-5NE and with the full size TTF-1 purified from an overexpressing bacterial strain; its reactivity to an anti-TTF-1 antibody (data not shown).

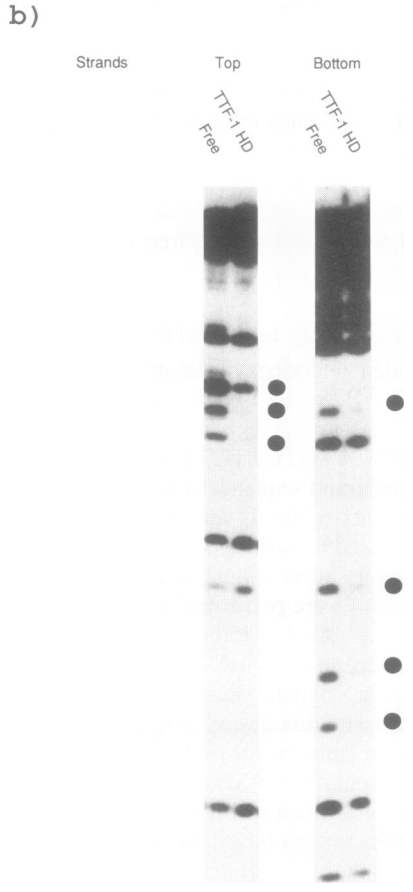
Taken together, interference and affinity data demonstrate that TTF-1HD recapitulates, quantitatively and qualitatively, the entire DNA binding properties of the native, full-length TTF-1 protein.

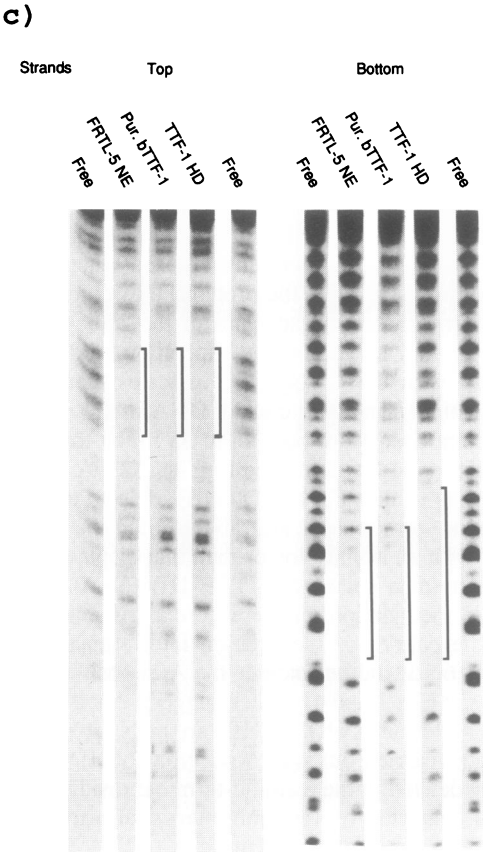
The amino-terminal arm of TTF-1 HD contacts DNA

To obtain insights on the amino acid residues of TTF-1HD that are likely to establish contacts with specific base pairs of the oligonucleotide C, the behaviour of mutant derivatives of the TTF-1HD was studied.

The amino-terminal arm of the Antennapedia (7), engrailed (8) and MATα2 (9) HDs is unstructured when the HD is not complexed with DNA. In the protein/DNA complex the amino-terminal arm contacts base pairs in the minor groove of the recognition site. An homology modelling study of TTF-1HD complexed with the C sequence indicates that the TTF-1HD amino-terminal arm could contact DNA in the minor groove in an area comprised from base 1a to 3a in the top strand and 6b to 8b in the bottom strand (See Fig. 4) (32). A deletion mutant of TTF-1HD, called TTF-1HD(6NΔ), deleted of 6 amino acids at the amino- terminus, is still able to bind the oligonucleotide C, albeit with an affinity lower than that displayed by the wild-type protein (data not shown). TTF-1HD(6NΔ) could therefore be used in methylation interference experiments. As shown in Fig. 3, the pattern of methylation interference obtained with TTF-1HD(6NΔ) did not include the A in position 8b (indicated by arrows in Fig. 3 and 4), which suggests that the 6 amino acids at the amino-terminus of TTF-1HD are in its proximity. A similar experiment, carried out with wild type ftz HD and its mutant derivative deleted of the 6 amino-terminal residues, showed that the bases contacted by the amino-terminal arm of ftz HD are found on the 3′ side of the motif 5′-ATTA-3′, which is an essential sequence for the binding of all Antp type HDs (ref. 14,







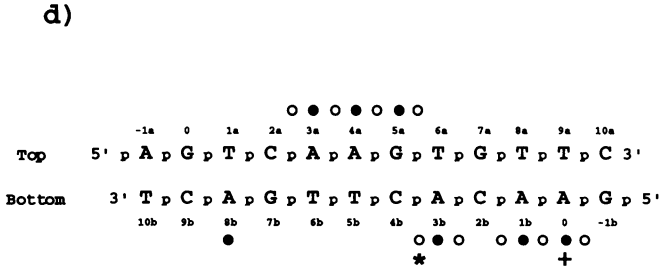


Figure 1. FRTL-5 NE, Pur bTTF-1 and TTF-1HD contact the site C of the Tg promoter in a very similar manner.(**a**) Base-pair contacts of FRTL-5 extracts (FRTL-5 NE) and purified TTF-1 (Pur. bTTF-1) on the oligonucleotide C of the Tg promoter. Dots indicate the nucleotides where methylation interferes with binding. (**b**) Base-pair contacts of the TTF-1HD on the oligonucleotide C of the Tg promoter. Dots indicate the nucleotides where methylation interferes with binding. The stronger appearance of the interference at base G5a with FRTL-5 NE and Pur.bTTF-1 compared to that with TTF-1HD (compare panels a and b) is due to overexposure of the autoradiogram of panel b. Scanning densitometry of low-exposure autoradiograms gave no difference between the bound/free ratios of the three interferences at base G5a. (**c**) Contacts of FRTL-5 extracts (FRTL-5 NE), purified TTF-1 (Pur. bTTF-1) and TTF-1HD with the phosphate groups of the DNA backbone of the C sequence. Brackets indicate the areas where ethylation interferes with binding. (**d**) Summary of the methylation and ethylation interference patterns on the C site of the Tg promoter obtained by FRTL-5 NE, Pur. b TTF-1 and TTF-1HD. Black and white dots indicate base-pair and phosphate interactions respectively. * indicates a phosphate contact present in TTF-1HD/C complex but lacking in FRTL-5/C and Pur. bTTF-1/C complexes. + indicates an adenine contact stronger in TTF-1HD/C complex.

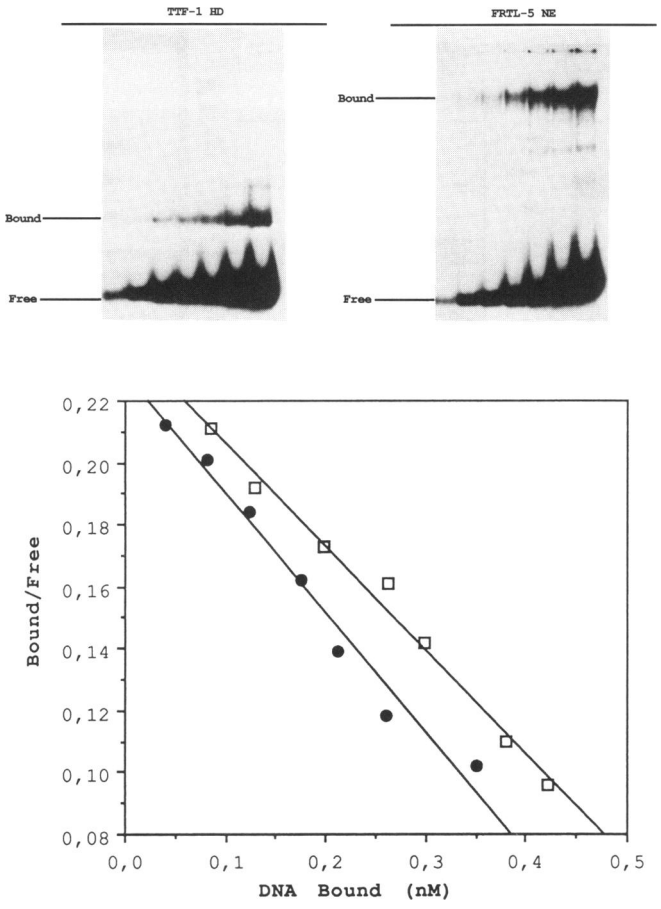


Figure 2. Binding affinity of FRTL-5 NE and TTF-1HD for the C site of the Tg promoter. The different proteins were obtained as described in Materials and Methods. Gel-retardation assays were run using different amount of ^{32}P-labelled oligonucleotide C. Bands representing free and protein-bound forms of DNA were quantitated by scanning densitometry of the autoradiograms and the obtained values used for Scatchard analysis. Top and bottom panels show respectively the autoradiograms of the gels and the Scatchard plots of the values obtained by densitometric scanning (black dots, TTF-1HD; white squares, FRTL-5NE). The slope of the plots indicates the dissociation constants (Kd) that are 3.0 and 2.6 nM for TTF-1HD and FRTL-5 NE respectively. The experiment shown is representative of several. The variability of Kd values obtained in different experiments is no more than 30% of the values obtained in the experiment shown in this figure.

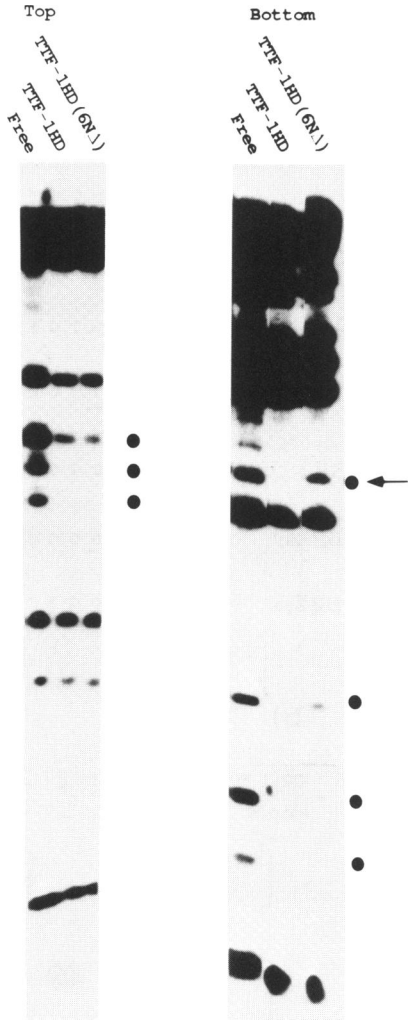


Figure 3. Comparison of base pair contacts of TTF-1HD and TTF-1HD(6NΔ) with the site C of Tg promoter. Dots indicate positions that interfere with binding. The arrow indicate the contact lost in TTF-1HD(6NΔ) compared to the wild-type protein.

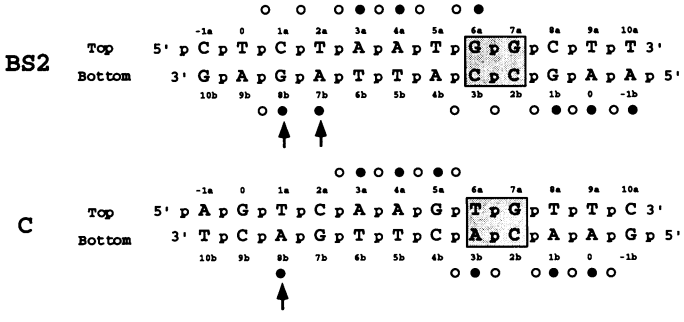


Figure 4. The overall geometry of binding is conserved between TTF-1 and ftz HDs. The sequence of both top and bottom strands of C and BS2 sites are shown. Black and white dots indicate contacts with bases and phosphates respectively of ftz (on BS2 site) and TTF-1HD (on C site). The arrows below both sites indicate base contacts that are abolished in protein/DNA complexes of mutant HDs lacking the first 6 amino acids at the amino-terminus (see Fig. 3). The shadowed box in both sites indicate the dinucleotides (GG in the top strand of BS2, TG in the top strand of C) contacted by Gln_{50} of ftz and TTF-1 HD (see Fig.6). Data of ftzHD and of its mutants lacking the first 6 amino acids or having a lysine in place of glutamine at position 50 were taken from reference 14.

indicated by arrows in Fig. 4). In the case of the TTF-1 binding site, the motif 5′-CTTG-3′ is found in a similar position with respect to the contacts made by the TTF-1 amino-terminal arm (Fig. 4). For the sake of an easier comparison of our data with those obtained with other HDs, for the remainder of this paper the binding sites of TTF-1 and of other HDs will be compared on the strand complementary to the one just examined, containing the motif 5′-TAAT-3′ and 5′-CAAG-3′ for Antp type HDs and TTF-1HD, respectively.

The behaviour of the TTF-1HD(6NΔ) mutant described here and that of the correspondent mutant of ftz HD (14) allows an alignment of the respective binding sites centered on the two adenines in the middle of the core sequences (5′-TAAT-3′ for BS2 and 5′-CAAG-3′ for C, Fig. 5). With this alignment, 9 out of 12 phosphate contacts and 5 out of 9 base-pair contacts of

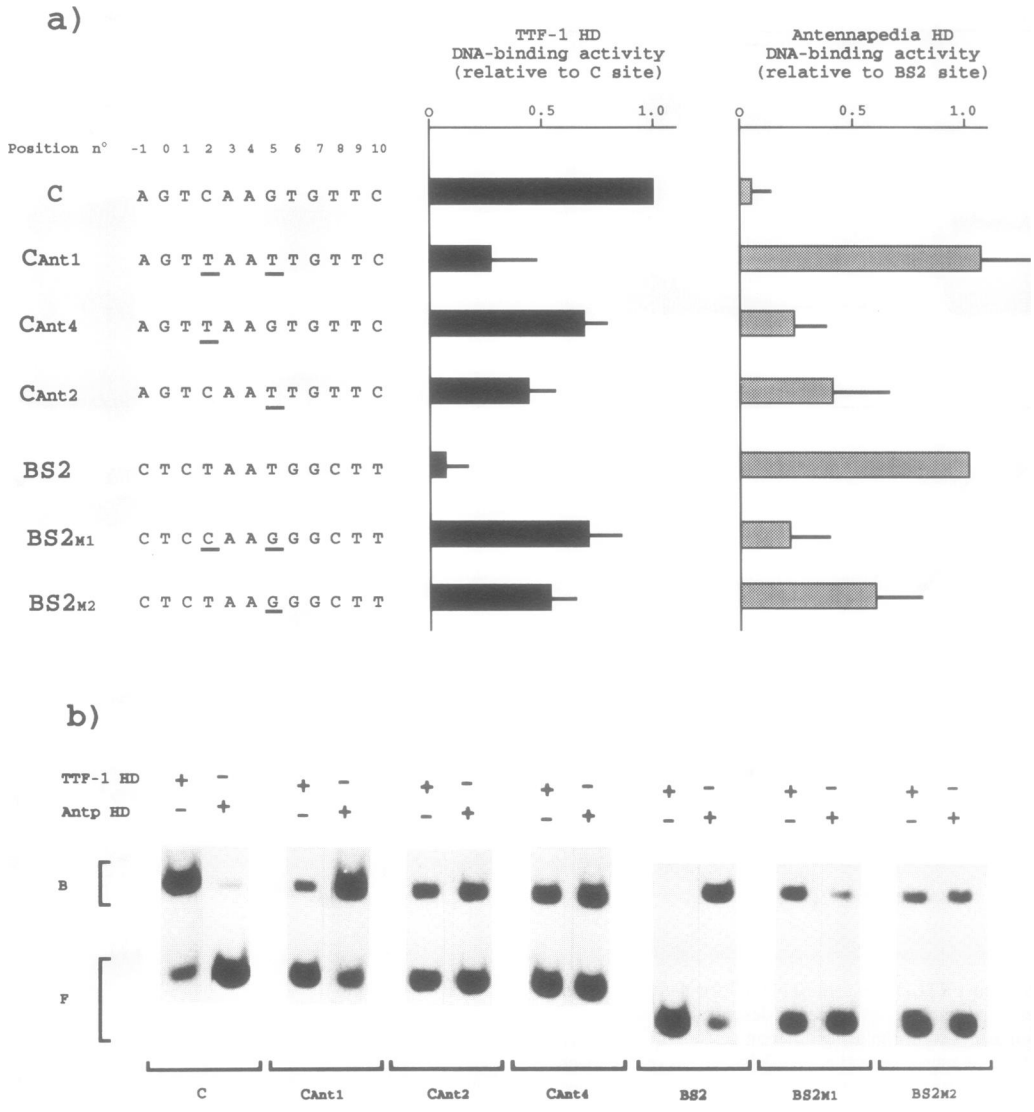


Figure 5. Binding activity of mutants of C and BS2 sequences to TTF-1 and Antennapedia HDs. (**a**) Underlined bases indicated mutations respect to C and BS2 wild-type sequences. DNA-binding activity was determined by gel-retardation assay and scanning densitometry of autoradiographic signals. Values are expressed as binding activity relative to wild-type C and BS2 sequences with TTF-1HD and Antennapedia HD respectively. Each bar represent the mean value (± SD) of at least 3 different experiments. The Student's t test was used to appreciate whether the binding activities with mutant sites were significantly different to those with wild-type sites. For TTF-1HD binding activities all differences were significant. For Antp HD binding activities the same phenomenon was observed except that the binding activity with C_{Ant1} was not significantly different to the binding activity with BS2. (**b**) Autoradiograms showing exemplicative experiments. B, protein/bound DNA; F, free DNA.

the Antp HD/BS2 complex are similarly arranged in the TTF-1HD/C complex.

Definition of the differences of sequences preferentially recognized by TTF-1 and *Antennapedia* HDs

To verify whether the critical difference in sequences specifically recognized by TTF-1 and by Antp class HDs is in the motifs 5′-CAAG-3′ and 5′-TAAT-3′, respectively, mutants of either C or BS2 sequences were constructed and their binding affinities for TTF-1 and Antennapedia HDs were measured by gel-retardation assay (Fig. 5). Mutations in the oligonucleotide C that change the 5′-CAAG-3′ motif to 5′-TAAG-3′ (C_{Ant4}), 5′-CAAT-3′ (C_{Ant2}) or 5′-TAAT-3′ (C_{Ant1}) show an affinity for TTF-1HD lower than that of the wild type oligonucleotide C, confirming the relevance of the bases on either side of the central AA. TTF-1HD seems to be particularly sensitive to the base present in position 5, since a mutation in this base only (C_{Ant2}) causes a significant decrease in binding activity. Conversely, binding of Antennapedia HD to mutated C sequences increases in the C_{Ant} mutants, the sequence in C_{Ant1} being the one displaying the highest binding activity. This result was in part expected, since C_{Ant1} contains the motif 5′-TAAT-3′ which is known to be recognized by the Antp class HDs. Nonetheless, this result is of great relevance since it demonstrates that the bases flanking the conserved AA dinucleotide are those that allow the Antp HD to discriminate between oligonucleotides C and BS2 and that there is no other sequence in the C oligonucleotide which is incompatible for the binding of this HD. Results obtained with

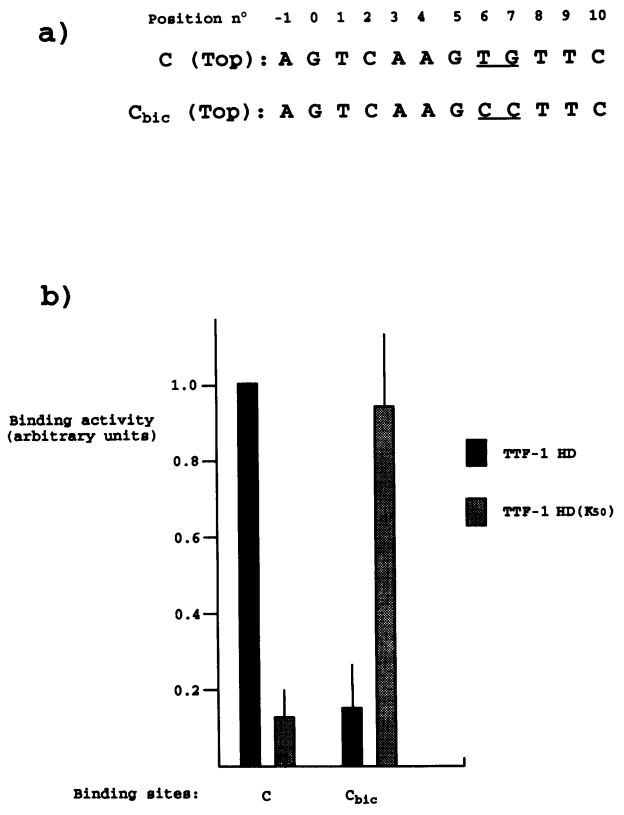


Figure 6. Amino acid at position 50 of TTF-1HD contributes to the DNA-binding specificity. (**a**) Sequences of C and C_{bic} sites. The bases that are different are underlined. (**b**) Comparison of DNA-binding specificity of TTF-1HD and TTF-1HD (K_{50}). DNA-binding activity of each protein was measured by gel-retardation assay and quantitation of autoradiographic signals by scanning densitometry. Values are expressed af fraction of TTF-1HD binding with C sequence. Each bar represent the mean value (±SD) of four experiments.

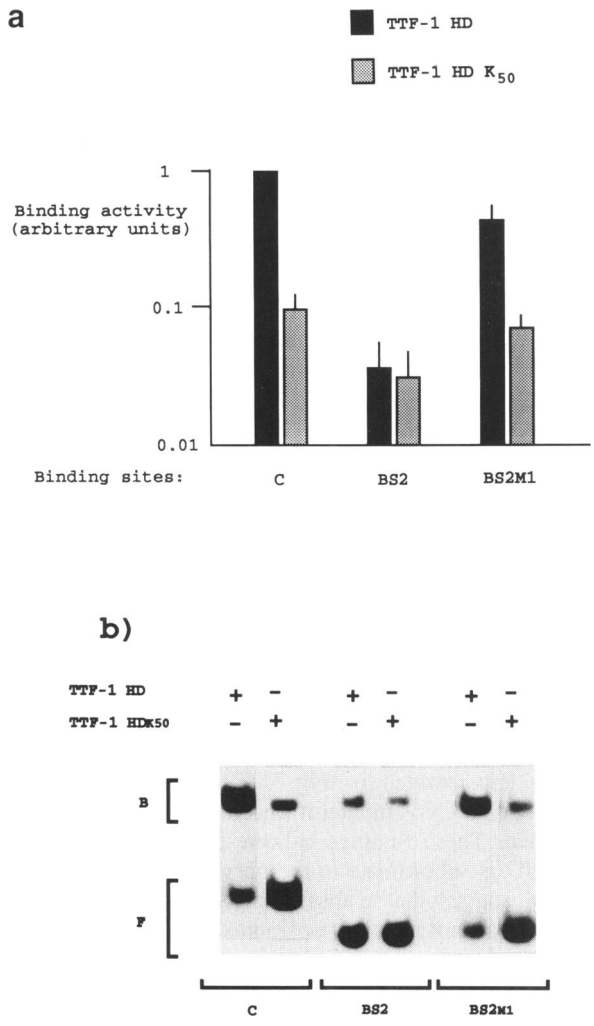


Figure 7. DNA-binding activity of TTF-1HD and TTF-1HD (K_{50}) with C, BS2 and $BS2_{M1}$ sites. The DNA-binding activity was determined by gel-retardation assay using a final concentration of 10^{-6}M of protein. (**a**) Each bar represent the mean value (±SD) of 3 experiments. In this panel a logarithmic scale is used for the binding activity axis in order to better show the binding activities with BS2 site. (**b**) Autoradiograms showing exemplicative experiments. B, protein/bound DNA; F, free DNA.

mutants of the BS2 oligonucleotide confirm the importance of positions on either side of the conserved AA sequence. In fact, $BS2_{M1}$, which differs from BS2 only for the presence of 5′-C-AAG-3′ replacing 5′-TAAT-3′, shows a decreased binding activity for Antp HD. Conversely, the binding activity of $BS2_{M1}$ for TTF-1HD is about 5 times higher than that of BS2. Also in the BS2 context, position 5 seems to be very important for TTF-1HD binding as $BS2_{M2}$, where only T at position 5 has been changed with G, is recognized by TTF-1HD much better than BS2 and similarly to $BS2_{M1}$, where both base pairs at position 5 and 8 were changed. Thus, position 5 plays a major role in dictating binding for either TTF-1HD (that prefers G in this position) or Antp HD (whose preference is T). Antp HD appears more sensitive than TTF-1 to the base in position 2.

The usage of Gln_{50} by TTF-1HD is subordinate to the presence of a 5′-CAAG-3′ motif

To further support the similarity of the DNA-binding geometry between TTF-1 and Antp class HDs, a TTF-1HD mutant, with a lysine substituting for the glutamine residue in position 50 (TTF-1HDK) was constructed. It has been extensively demonstrated that the amino acid residue in position 50 of homeodomains plays an important role in DNA-binding specificity (11–14). Base pairs contacted by residue 50 of ftz and bicoid HDs are located on the 3′ side of the motif 5′-TAAT-3′ (13, 14, 24). For instance, a Gln to Lys mutation at position 50 in the ftz HD causes a change in the recognition sequence from 5′-TAATGG-3′ to 5′-TAATCC-3′(14). Similarly, TTF-1HDK_{50} has a DNA binding specificity different from the wild type HD since it binds efficiently to a mutated C oligonucleotide (C_{bic}) where the sequence 5′-CAAGTG-3′ was changed to 5′-CAAGCC-3′. Conversely, the wild type TTF-1HD shows an affinity for C_{bic} much lower than for the oligonucleotide C (Fig. 6). These data indicate that the residue in position 50 of the TTF-1HD makes contacts within the dinucleotide 3′ to the motif 5′-CAAG-3′. This finding, compared to those obtained using ftz, bicoid and their respective mutant HDs (13–14) definitively demonstrates that TTF-1HD, although possessing a peculiar DNA-binding specificity, recognizes its

cognate sequence using a geometry of interaction similar to that of other HDs.

These data and previous results obtained with ftz HD (14) demonstrate that Gln_{50} is important in DNA recognition for both HDs. Furthermore, bases involved in Gln_{50} contacts are within the dinucleotides GG in BS2 and TG in C (Fig. 4). The similar orientation of ftz and TTF-1HD within their respective DNA–protein complex predicts that Gln_{50} of either HD should be able to make contacts within the appropriate dinucleotide in both C and BS2. A simple test of this hypothesis was performed by measuring the affinity of TTF-1HD and of its mutant derivative TTF-1HD (K_{50}) for BS2. If Gln_{50} of TTF-1HD was making a contact with the dinucleotide GG, similarly to the Gln_{50} of *ftz*HD, TTF-1HD(K_{50}) should show for BS2 an even lower affinity. Surprisingly, TTF-1HD and TTF-1HD (K_{50}) show a similar affinity for the oligonucleotide BS2 (Fig. 7). This result suggests that Gln_{50} in TTF-1HD does not recognize the GG dinucleotide in the sequence 5′-TAATGG-3′, even though it is able to contact the dinucleotide TG in the context of the TTF-1 recognition sequence (5′-CAAGTG-3′). This effect does not depend on the nature of the sequence that the Gln_{50} should contact (GG in BS2 and TG in C), as demonstrated by the behaviour of TTF-1HD and its Lys_{50} mutant on the oligonucleotide BS2M1. As detailed above, BS2M1 contains a TTF-1 motif 5′-CAAG-3′ in the context of the BS2 sequence, hence it contains the dinucleotide GG as potential target for Gln_{50}. As demonstrated in Fig.7, TTF-1HD binds well to BS2M1, while its Lys_{50} mutant is unable to recognize efficiently this sequence. These findings indicate that the usage of Gln_{50} by TTF-1HD is subordinate to the presence of the motif 5′-CAAG-3′ and suggest that in absence of this motif the amino acid at position 50 cannot properly positioned in order to play a role for sequence-specific recognition.

DISCUSSION

Methylation and ethylation interference data presented in this paper define the contacts that TTF-1 establishes with the bases and the phosphate backbone of its target sequence. Such contacts do not significantly vary when bacterially expressed TTF-1HD is used in place of the entire native protein, thus confirming that the HD is the only part of TTF-1 contacting the DNA. This indication is further supported by the finding that native TTF-1 and its isolated HD have a very similar affinity for the oligonucleotide C.

TTF-1HD makes specific contacts with DNA in an area of 10 base-pairs. The length of the contacted area is very similar to that contacted by the Antennapedia class HDs (14–29). Deletion of the first 6 amino acids at the amino-terminus of TTF-1HD determines a loss of a contact at one end of the bound area. A very similar phenotype was detected when the same deletion was performed in the ftz HD (14). In addition, the same point mutation (lysine in place of glutamine at position 50) elicits similar phenotypic changes both in TTF-1 and Antennapedia class HDs. These findings indicate that between TTF-1 and Antennapedia class HDs the overall geometry of the DNA binding is conserved and are consistent with the recent demonstration that the structure of TTF-1HD is very similar to that of other HDs (33). Nevertheless, we find that in TTF-1HD the usage of the amino acid at position 50 depends on the DNA sequence context. The nature of the amino acid residue at position 50 determines the different DNA-binding specificity existing between Antennapedia class HDs (glutamine), Bicoid HD (lysine) and Paired HD (serine) (11–13). This residue interacts within the two bases 3′ to the core 5′-TAAT-3′ motif (13–26). In the sequence recognized by the TTF-1HD, the bases 3′ to the core 5′-CAAG-3′ motif are TG, the same present in several Antennapedia class specific sequences (26, 34). The mutant TTF-1HD (K_{50}) shows a DNA-binding specificity different from that of the wild-type protein and, similarly to a homologous mutant of ftz HD (14), recognizes better a sequence in which CC replaces TG at position 6 and 7, adjacent to the 3′ side of the motif 5′-CAAG-3′. Surprisingly, when a 5′-TAAT-3′ replaces the core sequence 5′-CAAG-3′, the nature of the residue in position 50 of the TTF-1HD becomes irrelevant. This observation is in sharp contrast to the finding that several HDs, binding to 5′-TAAT-3′ containing sequences, use the residue in position 50 as an essential determinant for DNA binding specificity (11–14). The finding that Gln_{50} in TTF-1HD plays a role for sequence discrimination only when the 5′-CAAG-3′ core is present, suggests that Gln_{50} either interacts with bases within this sequence or, at least in the context of TTF-1HD, is allowed to bind DNA only if a previous contact, by an as yet unidentified side chain, has been established within the core 5′-CAAG-3′ sequence. Only after this contact, the Gln_{50} can be properly positioned in order to have a significant role in sequence-specific recognition. Taking into account the base preference at the dinucleotide 3′ to the 5′-TAAT-3′ motif of Ubx HD, it has been proposed that Gln_{50} may interact with DNA using alternative modes (35). In addition, NMR studies of the Antp/DNA complex indicate that Gln_{50} could contact into the major groove several bases using a fluctuating network of weak-bonding interactions (36). Very recently an 'induced fit' model has been proposed to explain the large negative heat capacity change in the site-specific protein/DNA recognition (37). In this model protein and DNA do not have rigid but flexible and adaptable interacting surfaces. Following this model, the 5′-CAAG-3′, but not the 5′-TAAT-3′, motif could induce a conformational change in TTF-1HD which is necessary for the proper positioning of Gln_{50}.

TTF-1HD is unable to efficiently recognize the general HD core target sequence 5′-TAAT-3′. Rather, TTF-1HD binds efficiently to sequences containing 5′-CAAG-3′ as a core. Such binding specificity is very likely conserved also in other HDs of the NK family (38). In fact, no NK HDs are present in clones obtained after screening of a *Drosophila* embryo expression library with an oligonucleotide containing the 5′-TAAT-3′ core sequence, although many other HD-containing clones were identified (39). The base pair just 3′ to the conserved AA plays the most important role for the different sequence recognition of Antennapedia class and TTF-1HDs. Methylation interference data indicate that TTF-1HD contacts N7 of the guanine. When, in place of a purine, a pyrimidine is present (T for G in this case), the imidazole N7 is substituted by the C5 atom which is bound to a nonpolar hydrogen atom. Since N7 of purines can accept a hydrogen bond while pyrimidine C5 group cannot and, in addition, it protrudes much more in the major groove, a protein residue able to bind a purine will be not allowed to recognize a pyrimidine at that position for both steric and electrostatic reasons (40). Which amino acid of TTF-1HD contacts the G 3′ to the conserved AA and thus contributes to the peculiar DNA-binding specificity of this protein? There are several reasons to believe that this amino acid is the tyrosine at position 54. First, a tyrosine at this position is peculiar and conserved in all members of the NK class of HDs (26, 41). Second, Antennapedia class

HDs have a conserved methionine at this position (26). Third, structural studies of Antennapedia/DNA and MATα2/DNA complexes demonstrate that residue at position 54 contacts DNA (7–10). Fourth, a modelling study shows that Tyr_{54} of TTF-1HD is in an excellent position to contact DNA into the major groove very close to the 5′-CAAG-3′ motif (30). It is tempting to speculate that the residue in position 54 plays an important role in DNA recognition not only because it provides an important contact within the core 5′-CAAG-3′ motif, but also because it controls whether the residue in position 50 is allowed to contact DNA. Both functional and structural studies are needed in order to prove our hypothesis and to conclusively build up a model explaining the different DNA-binding specificity existing between TTF-1 and Antennapedia class HDs.

ACKNOWLEDGEMENTS

We thank Dr Alfonso Colombatti for the use of the radioisotope facility at the Centro di Riferimento Oncologico of Aviano, Italy. We gratefully acknowledge Sig. Carlo Lo Cascio for the technical assistance during the art work for figures preparation. This work was partially supported by the Progetto Finalizzato Ingegneria Genetica of the Consiglio Nazionale delle Ricerche, Rome, Italy and by the Associazione Italiana Ricerca sul Cancro (AIRC).

REFERENCES

1. Gehring, W.J. (1987) Science 236, 1245–1252.
2. Scott, M.P., Tamkun, J.W., Hartzell, G.W. (1989) Bioch. Bioph. Acta 989, 25–48.
3. Ingraham, H.A., Flynn, S.E., Voss, J.W., Albert, V.R., Kapiloff M.S., Wilson, L., Rosenfeld, M.G. (1990) Cell 61, 1021–1033.
4. Sessa, G., Morelli, G., Ruberti, I. (1993) EMBO J. 9, 3507–3517.
5. Freyd, G., Kim, S.K., Horvitz, H.R. (1990) Nature 344, 876–879.
6. Fortini, M.E., Lai, Z., Rubin, G.M. (1991) Mech. Dev. 34, 113–122.
7. Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J., Wuthrick, K. (1991) EMBO J. 9, 3085–3092.
8. Kissinger, C.R., Liu, B., Martin-Blanco, G., Kornberg, T.B., Pabo, C.O. (1990) Cell 63, 579–590.
9. Wolberger, C., Vershon, A.K., Beishan, L., Johnson, A.D., Pabo, C.O. (1991) Cell 67, 764–772.
10. Florence, B., Handrow, R., Laughon, A. (1991) Mol. Cell. Biol. 11, 3613–3623.
11. Treisman, J., Gonczy, P., Vashinshtha, M., Harris, E., Desplan, C. (1989) Cell 59, 553–562.
12. Hanes, S.D., Brent, R. (1989) Cell 57, 1275–1283.
13. Hanes, S.D., Brent, R. (1991) Science 251, 426–430.
14. Percival-Smith, A., Muller, M., Affolter, M., Gehring, W.J. (1990) EMBO J. 9, 3967–3974.
15. Damante, G., Di Lauro, R., (1991) Proc. Natl. Acad. Sci. (USA) 88, 5388–5392.
16. Furukubo-Tokunaga, K., Muller, M., Affolter, M., Pick, L., Kloter, U., Gehring, W.J. (1992), Genes and Dev. 6, 1082–1096.
17. Ekker, S.C., von Kessler, D.P., Beachy, P.A. (1992) EMBO J. 11, 4059–4072.
18. Zeng, W., Andrew, D.J., Mathies, L.D., Horner, M.A., Scott, M.P. (1993) Development 118, 339–352.
19. Lin, L., McGinnis, W. (1992) Genes and Dev. 6, 1071–1081.
20. Chan, S.K., Mann, R.S. (1993) Genes and Dev. 7, 796–911.
21. Kornberg, T.B. (1993) J. Biol. Chem. 268, 26813–26816.
22. Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M.G., Di Lauro, R. (1990) EMBO J. 9, 3631–3639.
23. Civitareale, D., Lonigro, R., Sinclair, A.J., Di Lauro, R., (1989) EMBO J. 8, 2537–2542.
24. Studier, F.W., Moffat, B.A., (1986) J. Mol. Biol. 189, 113–130.
25. Muller, M., Affolter, M., Leupin, W., Otting, G., Wuthrich, K., Gehring, W.J. (1988). EMBO J. 7, 4299–4304.
26. Laughon A., (1991) Biochemistry 30, 11357–11372.
27. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R. (1989) Gene 77, 51–60.
28. Scatchard, G. (1949) Ann. N.Y. Acad. Sci. 51, 660–672.
29. Affolter, M., Percival-Smith, A., Muller M., Leupin, W., Gehring, W.J. (1990) Proc. Natl. Acad. Sci. (USA) 87, 4093–4097.
30. Siebenlist, U., Gilbert, W. (1980) Proc. Natl. Acad. Sci. (USA) 77, 122–126.
31. Damante, G, Tell, G, Formisano, S., Fabbro, D., Pellizzari, L., Di Lauro, R. (1993) Bioch. Bioph. Res. Comm.197, 632–638.
32. Fogolari, F., Esposito, G., Viglino, P., Damante, G. Pastore, A. (1993) Prot. Eng. 6, 513–519.
33. Viglino, P., Fogolari, F., Formisano, S., Bortolotti, N., Damante, G., Di Lauro, R., Esposito, G. (1993) FEBS Lett. 336, 397–402.
34. Pick, L., Schier, A., Affolter, M., Schmidt-Glenewinkel, T., Gehring, W.J. (1990) Genes Dev 4, 1224–1239.
35. Ekker, S.C., Young, K.E., von Kessler, D.P., Beachy, P.A. (1991) EMBO J. 5, 1179–1186.
36. Billeter, M., Qian, Y.Q., Otting, G., Muller, M., Gehring, W.J., Wuthrich K. (1993) J. Mol. Biol. 234, 1084–1097.
37. Spolar, R.S., Record, M.T. (1994) Science 263, 777–784.
38. Kim, Y., Niremberg,M., (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 7716–7720.
39. Kalionis, B., O'Farrel, P. (1993) Mech. Dev. 43, 57–70.
40. Seeman, N.C., Rosemberg, J.M., Rich, A. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 804–808.
41. Price, M., Lazzaro, D., Pohl, T., Mattei, M.G., Ruther, U., Olivo, C-J., Duboule, D., Di Lauro, R. (1992) Neuron 8, 241–255.