In the TTF-1 homeodomain the contribution of several amino acids to DNA recognition depends on the bound sequence

Dora Fabbro, Gianluca Tell, Antonio Leonardi, Lucia Pellizzari, Carlo Pucillo, Renata Lonigro, Silvestro Formisano and Giuseppe Damante*

Dipartimento di Scienze e Tecnologie and 1 Dipartimento di Biologia e Patologia Cellulare e Molecolare ‘L. Califano’, Napoli, Italy

Received July 4, 1996; Accepted July 15, 1996

ABSTRACT

The thyroid transcription factor-1 homeodomain (TTF-1HD) shows a peculiar DNA binding specificity, preferentially recognizing sequences containing the 5′-CAAG-3′ core motif. Most other homeodomains instead recognize sites containing the 5′-TAAT-3′ core motif. Here, we show that TTF-1HD efficiently recognizes another sequence, called D1, devoid of the 5′-CAAG-3′ core motif. Different experimental approaches indicate that TTF-1HD contacts the D1 sequence in a manner which is different to that used to interact with sequences containing the 5′-CAAG-3′ core motif. The binding activities of mutants of TTF-1HD display with the D1 sequence or with the sequence containing the 5′-CAAG-3′ core motif indicate that the role of several DNA-contacting amino acids is different. In particular, during recognition of the D1 sequence, backbone-interacting amino acids not relevant in binding to sequences containing the 5′-CAAG-3′ core motif play an important role. In the TTF-1HD, therefore, the contribution of several amino acids to DNA recognition depends on the bound sequence. These data indicate that although a common bonding network exists in all of the HD/DNA complexes, peculiarities important for DNA recognition may occur in single cases.

INTRODUCTION

The homeodomain (HD) represents the DNA binding domain of a large number of transcription factors controlling cell fate decisions (1,2). Different HDs show a similar structure consisting of three helical regions (I, II and III) folded into a tight globular structure (3). Helix I is preceded by an N-terminal arm and separated by a loop from helix II, which forms, with helix III, a helix–turn–helix motif (H-T-H). The latter has previously been described for several prokaryotic gene-regulatory proteins (4). However, differently from prokaryotic H-T-H motifs, in which dimerization is required for high affinity DNA binding, most HD bind with high affinity to DNA as monomers (5). Only for the paired (prd) class of HDs has a cooperative dimerization been observed (6–7). The DNA binding mode of HDs is highly conserved (3,8–9). Helix III (also called the recognition helix) lies in the major groove of the DNA, where it establishes specific contacts to bases. Additional specific contacts to bases are made by the N-terminal arm (in the minor groove). The loop between helices I and II interacts with the DNA backbone.

The amino acids which control the DNA binding specificity of HDs are located mostly in the recognition helix and in the N-terminal arm (10–21). However, the relative importance of each residue appears to be different, depending on the HD. For example, only a functional comparison of fushi tarazu (ftz) and muscle cell homeobox (msh) HDs has revealed that the amino acids at positions 28 and 43 could be relevant in the control of DNA binding specificity (18). These observations could suggest that HD–DNA interactions, although based on a common bonding network, could differ in some details sometimes important for DNA recognition. These differences may occur even when the same protein interacts with different DNA sequences. In fact, it has been demonstrated that the binding specificity of the prdHD is controlled by amino acids at the N-terminus of the recognition helix only in interaction with particular sequences (14,22).

The HD of thyroid transcription factor-1 (TTF-1HD) has a structure similar to other HDs (23) but shows a peculiar DNA binding specificity. In fact, TTF-1HD preferentially recognizes sequences having the 5′-CAAG-3′ core motif and, in contrast to other HDs, it binds only with low affinity sequences containing the 5′-TAAT-3′ core motif (17). In this study we show that TTF-1HD is able to specifically recognize a sequence devoid of the 5′-CAAG-3′ core motif. By means of different experimental approaches we demonstrate that TTF-1HD interacts with this sequence by using a binding mode which is different to that used to recognize sequences containing the 5′-CAAG-3′ core. Backbone-interacting amino acids not relevant in binding to sequences containing the 5′-CAAG-3′ core motif play an important role in recognizing the sequence devoid of the 5′-CAAG-3′ core motif.
MATERIALS AND METHODS

Preparation of proteins and oligonucleotides

Construction of the bacterial expression vectors pT7.7 TTF-1HD and pT7.7 TTF-1HD(K50), pT7.7 TTF-1HD(6NA) and M2HD has been described elsewhere (17,24). Mutants TTF-1HD(R28) and M2HD(R28) have been constructed by the method of Ho et al. (25) and cloned in the vector pT7.7. Proteins were expressed using the BL21 (DE3) Escherichia coli strain (26). The expressed proteins were purified essentially as described (5) with two chromatographic steps using Biorex and Mono S ion exchange matrices. The purity of the proteins was checked by SDS–PAGE and was ≥99%. The sequences of the oligonucleotides used in gel retardation and methylation interference assays were as follows (top strand): C, 5′-CACGTGCCAGCAGTCGTTTCTTGTA-3′; C Ant1, 5′-CAGTGAGTGGGCTGAACTAGTC-3′; D1, 5′-ACGATGAGTTTGCTGACTGTCAT-3′; D133mer, 5′-CGAAGGAGTTTGCTGACTGTCAT-3′; D1, 5′-ACGATGAGTTTGCTGACTGTCAT-3′; D133mer, 5′-CGAAGGAGTTTGCTGACTGTCAT-3′.

Oligonucleotides were labeled at the 5′-end using polynucleotide kinase and [γ-32P]ATP and annealed with the complementary strand.

Gel retardation and methylation interference assays

Gel retardation assays were performed by incubating protein and DNA in a buffer containing 20 mM Tris–HCl, pH 7.6, 75 mM KCl, 0.25 mg/ml bovine serum albumin (BSA), 5 mM dithiothreitol (DTT), 5 µg/ml calf thymus DNA, 10% glycerol for 30 min at room temperature. Calf thymus DNA was omitted in gel retardation assays to measure equilibrium dissociation constants (Kd). Protein-bound and free DNA were separated on native 7% polyacrylamide gels run in 0.5× TBE for 2 h at 4°C. Gels were dried, exposed to X-ray films and the bands quantitated by densitometric scanning of the autoradiogram using a LKB laser densitometer. Kd values were calculated as in Damante et al. (27). The DNA binding activities of HDs shown in Figure 5 were determined by calculating the ratio of protein-bound and free DNA signals using the same amount of protein. The values were then normalized to the value obtained for the complex TTF-1HD/C sequence, arbitrarily considered as 100.

Methylation interference experiments were done according to Zannini et al. (28), using as probes dimethylsulfate-treated D133mer oligonucleotides. Protein-bound and free DNAs were separated by preparative PAGE and recovered as described (29).

After the chemical cleavage the products were separated on 20% denaturing polyacrylamide gels and visualized by autoradiography.

Circular dichroism analysis

Spectra were recorded in a 2 mm cell at 10 µM DNA duplex and 10 µM protein concentrations in 5 mM NaH2PO4, 5 mM MgCl2, 50 mM NaClO4, pH 7.5, at 25°C. Data were collected on a Jasco J-600 spectropolarimeter (Jasco Inc., Easton, MD). The averages of 10 scans were baseline-corrected with a spectrum of buffer alone and were smoothed using software provided by Jasco. Difference spectra were calculated as described (30,31). The α-helical content was calculated according to Menendez-Arias et al. (32).

RESULTS AND DISCUSSION

TTF-1HD is able to recognize a sequence devoid of the 5′-CAAG-3′ core motif

The D1 sequence has been identified by site selection experiments starting from rat genomic DNA (33; Fabbro et al., unpublished data). A measure of the equilibrium dissociation constants (Kd) of the TTF-1HD/D1, TTF-1HD/C and TTF-1HD/C Ant1 complexes was performed (Fig. 1). The C sequence is a high affinity TTF-1HD binding site containing the 5′-CAAG-3′ core motif. The C Ant1 sequence is derived from the C sequence by changing the 5′-CAAG-3′ motif to 5′-TAAT-3′. As such, C Ant1 is a low affinity binding site for TTF-1HD (22). The apparent Kd for the TTF-1HD/D1 complex is similar to that measured for the TTF-1HD/C complex (0.34 × 10⁻⁹ and 0.22 × 10⁻⁹ M respectively) and much lower than that observed for the TTF-1HD/C Ant1 complex (0.38 × 10⁻⁸ M). These data demonstrate that the D1 sequence, though not containing a 5′-CAAG-3′ core motif (see D1 sequence in Fig. 3B), is a high affinity binding site for TTF-1HD.

TTF-1HD interacts with the D1 sequence in a manner which is different to that used with the C sequence

The lack of significant homology between the C and D1 oligonucleotides suggests that TTF-1HD could recognize these sequences by using different binding modes. This possibility is also suggested by the different migration rates observed for the TTF-1HD/C and the TTF-1HD/D1 complexes in gel retardation assays. Figure 2 shows that although the free 22mer D1 oligonucleotide migrates faster than the free 24mer C oligonucleotide, the TTF-1/D1 complex migrates slower than the TTF-1/C complex. These results are compatible with different structural conformations of the protein/DNA complexes resulting in different hindrances during gel migration. It is important to note that the complexes established by TTF-1HD with other oligonucleotides containing the 5′-CAAG-3′ motif but having
Figure 2. Different migration rates of TTF-1HD/C and TTF-1HD/D1 complexes. F, free DNA; B, protein-bound DNA. Note that the TTF-1HD/D1 complex migrates more slowly than the TTF-1/C complex, although the free 22mer D1 migrates faster than the free 24mer C.

either a different length (18mer or 14mer) or base mutations show a migration rate equal to that observed for the TTF-1HD/C complex (data not shown). We can exclude the possibility that the slow migration rate of the TTF-1HD/D1 complex is due to dimerization because when two TTF-1HD molecules bind to the same DNA molecule the ternary complex show a much slower migration rate (data not shown).

The contacts established by TTF-1HD with the C sequence have been previously evaluated by methylation interference (17). In order to align and compare the C and D1 sequences with respect to TTF-1HD binding, we performed a methylation interference experiment on the TTF-1HD/D1 complex. The results are shown in Figure 3A, together with the contacts previously mapped on the C sequence (Fig. 3B) (17). The C and D1 sequences do not show significant homologies in the contact areas. Moreover, the pattern of contacts observed on the D1 sequence is different to that observed on the C sequence. These data further indicate that TTF-1HD could recognize the D1 sequence by using a different binding mode relative to that utilized to interact with the C sequence.

Structural studies on the NK2HD have revealed that, upon interaction, DNA is able to increase the α-helical content of the protein (34–35). This conformational change occurs at the level of the C-terminal moiety of the recognition helix. The primary sequence of the NK2HD is very similar to that of TTF-1HD (36), suggesting that a similar phenomenon could occur during the TTF-1HD–DNA interaction. Therefore, eventual differences in the C and D1 sequences with respect to DNA-induced conformational changes could further support the existence of different modes of interaction. Circular dichroism (CD) spectroscopy was used to tackle such an issue. The CD spectrum of the peptide (Fig. 4A and B, continuous lines) shows the typical minima at 208 and 222 nm of an α-helix conformation. In the absence of DNA the predominant secondary structure was α-helix (34%). Interaction with the D1 sequence has no effect on the helical content of TTF-1HD (Fig. 4A). In contrast to D1, the C sequence induces an increase (Fig. 4B, dotted line) in the helical content of TTF-1HD (signal at 222 nm). The α-helical content increases from 34 to 45% upon interaction with the C sequence. The lack of changes in the region of the spectrum dominated by signals from the oligonucleotides (245–310 nm; data not shown) indicates that the changes observed below 245 nm result from changes in peptide rather than oligonucleotide structure (always B-form DNA).

The contribution of several amino acids to the binding affinity is different between the TTF-1HD/C and the TTF-1HD/D1 complexes

The data described above indicate that TTF-1HD interacts with the D1 sequence in a manner which is different to that used to interact with the C sequence. Based on these data we wanted to address the question of whether critical amino acids for specificity in the TTF-1HD/C interaction would play a similar role in the TTF-1/D1 interaction. Figure 5 shows the binding activity of several TTF-1HD mutants with the C, C\text{Ant}1 and D1 sequences.
Mutant TTF-1HD(6NA) lacks the N-terminal arm, which is known to contribute to the interaction when the protein binds the C site (17). The binding activity of TTF-1HD(6NA) is low with all sequences, indicating that the N-terminal arm is essential to establish an efficient interaction with either the C sequence or with the D1 sequence.

In mutant M2HD the TTF-1HD N-terminus of the recognition helix has been mutagenized. Pro42, Thr43 and Val45 have been changed to Glu, Arg and Ile respectively (24). In the C context, mutant M2HD recognizes the 5’-CAAG-3’ but not the 5’-TAAT-3’ motif and, therefore, behaves as the wild-type TTF-1HD. However, M2HD shows very much reduced binding activity with the D1 sequence, indicating that the N-terminal region of the recognition helix (containing Pro42, Thr43 and Val45) plays a significant role in recognition of this sequence. One of the amino acids controlling TTF-1HD binding specificity is Glu50 (17). When Glu50 is changed to Lys [mutant TTF-1HD(K50)], binding activity to the sequence 5’-CAAGGG-3’ is reduced, while binding activity to the sequence 5’-CAAGGGCC-3’ is increased (17). The binding activity of mutant TTF-1HD(K50) to the C sequence is 5-fold reduced compared with the wild-type protein. However, TTF-1HD(K50) binds to the D1 sequence as well as wild-type TTF-1HD.

In the context of the C sequence, the amino acid located at position 54 plays a role in binding specificity. In fact, when the wild-type Tyr54 is changed to Met, binding activity to the C sequence is reduced while, in contrast, binding activity to the C Ant1 sequence is increased (21). TTF-1HD(M54) binds D1 as well as wild-type TTF-1HD, indicating that, in the context of the D1 sequence, the side chain of the amino acid at position 54 is less important than in the context of the C sequence.

Amino acids at positions 50 and 54 of HDs contact bases in the major groove (3). The binding activities of both TTF-1HD(K50) and TTF-1HD(M54) indicate that in the TTF-1HD/D1 interaction the contacts to bases established by amino acids at position 50 and 54 play a role much less important with respect to that played in the TTF-1HD/C interaction. In contrast, the binding activity of M2HD indicates that in the context of D1 sequence amino acids at the N-terminus of the recognition helix appear to be important. In MatTA2/DNA and Antp/DNA complexes amino acids at positions 42 and 43 respectively establish contacts with the sugar–phosphate backbone (9, 37). Therefore, the different binding activities shown by TTF-1HD and M2HD with the C and D1 sequences would indicate a significant role of backbone contact(s) in the TTF-1HD/D1 interaction but not in the TTF-1HD/C interaction.

The amino acid at position 43 plays a role in the differential DNA binding properties observed for ftzHD and mshHD, in combination with the amino acid at position 28 (18). Interestingly, TTF-1HD possesses Ala and Thr respectively at positions 28 and 43. This configuration is very similar to that of mshHD (Ile28 and Thr43). However, M2HD possesses an Arg at position 43, as does the ftzHD. These observations induced us to test the relevance of the amino acid at position 28 of TTF-1HD. In place of the naturally occurring Ala28, we introduced Arg, which instead is present at position 28 of ftzHD. Arg28 was introduced into both TTF-1HD and M2HD, giving rise to TTF-1HD(R28) and M2HD(R28) respectively. Figure 5 shows that the binding activity of TTF-1HD(R28) is only moderately reduced compared with that of TTF-1HD with the C sequence. In contrast, TTF-1HD(R28) interacts much less efficiently than TTF-1HD with the D1 sequence. These results indicate that the amino acid at position 28 plays an important role in DNA recognition only in the context of the D1 sequence, and not in that of the C sequence. The binding activity of mutant M2HD(R28) with the C sequence is reduced to 40%, compared with TTF-1HD, but to 5% with the D1 sequence. In the Antp/DNA complex Arg28 establishes an electrostatic interaction with the phosphate backbone (37). Therefore, our results indicate that backbone interactions (established by amino acids at positions 28 and 43) play an important role in recognition of the D1 sequence, but only a marginal role in recognition of the C sequence. In contrast, results obtained with mutants TTF-1HD(K50) and TTF-1HD(M54) indicate that base contacts established by amino acids at positions 50 and 54 play a role C recognition, but not in D1 recognition. These data are summarized in Figure 6.

Structural studies are required to clarify the reasons accounting for the different effects of TTF-1HD mutations on the binding efficiency to the C and D1 sequences. Nevertheless, our data could be explained essentially by a different use of the recognition helix. In fact, the binding activity of both mutants where contact amino acids of helix III have been changed (at positions 50 and
Figure 5. DNA binding activity of TTF-1HD mutants with the C, CAn1, and D1 sequences. DNA binding activity was evaluated by gel retardation assay as described in Materials and Methods. Bars indicate the mean value (± SD) of at least three different experiments.

Figure 6. Schematization of the differences existing between the TTF-1/C and the TTF-1/D1 complexes. DNA is shown in green, the structure of the HD is shown in black. α-Helices are indicated by cylinders. Red arrows indicate positions (50 and 54) in which mutations are important for interaction with the C but not D1 sequence. Blue arrows indicate positions (N-terminal of helix III and position 28) in which mutations are important for interaction with the D1 but not C sequence.

54) would indicate that the recognition helix is used in a different manner with the D1 sequence with respect to the C sequence. Moreover, in view of the results obtained for the NK2HD (34–35), our CD data would further indicate a different use of the recognition helix between the C and D1 sequences. Mutations of backbone-contacting amino acids would be relevant only for interaction with the D1 sequence because, in this case, these mutations would change the position of the recognition helix in such a way as to reduce the efficiency of its interaction with DNA. In contrast, mutations of backbone-contacting amino acids would not interfere with the interaction efficiency of the recognition helix with the C sequence. A direct negative effect of mutations of backbone-contacting amino acids on the TTF-1HD–D1 interaction is much less likely because the arginines at position 28 or 43 should be prone to establish electrostatic interactions with the phosphates of the DNA backbone.

CONCLUSION

Our data demonstrate that TTF-1HD is able to recognize with high affinity a sequence devoid of the 5′-CAAG-3′ core motif. In such a way, the DNA binding specificity of this protein appears to be wider than previously appreciated. Results obtained for other HD-containing proteins, using either in vivo or in vitro binding assays, indicate that other HDs are also able to recognize a broad spectrum of DNA sequences (38–39). Therefore, the ability to recognize different DNA sequences appears to be a common feature of HDs.

The binding mode of TTF-1HD and, in turn, the importance of contact amino acids may change in recognizing different DNA sequences. The biological implications of this phenomenon are evident: the nature of a contact amino acid would be relevant for the recognition of only a subset of controlled genes. Such a possibility is compatible with results obtained in vivo with the ftz protein (40). In fact, by using Drosophila mutants defective in the ftz gene, it has been clearly demonstrated that when a ftz mutation in which the Gln 50 of ftz is changed to Lys is introduced, the mutant protein is able to provide an efficient rescue of para-segments 8 and 14, but a much poorer rescue of other structures. In view of our findings, the Lys50-containing ftz mutant would rescue only the activity of genes controlled by DNA sequences for which the nature of the amino acid at position 50 of ftz protein
is not relevant for an efficient interaction. The versatility of HDs in recognizing DNA sequences is one of the most important features to explain the success of these proteins during evolution (41).

Although several studies have revealed that the binding mode of HDs is conserved overall, the present data demonstrate that in each HD/DNA complex peculiarities may occur and play a role in DNA recognition. To fully understand the molecular bases of the biological activity of HD-containing proteins, it might be of relevance to reveal these peculiarities.

ACKNOWLEDGEMENTS

The first two authors (D.F. and G.T.) have contributed equally to this work. This research is funded by a grant to G.D. from the Associazione Italiana per la Ricerca sul Cancro (AIRC). D.F. is also supported by a AIRC fellowship. We gratefully acknowledge R. Di Lauro for helpful discussions, L. Minichiello for suggestions in writing the manuscript and L. Cattarossi for computational support.

REFERENCES