Topology of the Thyroid Transcription Factor 1 Homeodomain−DNA Complex†

Andrea Scaloni,‡,§ Maria Monti,‡ Renato Acquaviva,|| Gianluca Tell,⊥ Giuseppe Damante,⊥ Silvestro Formisano,|| and Piero Pucci*,‡,@

Centro Internazionale di Servizi di Spettrometria di Massa, CNR-Università di Napoli "Federico II", via Pansini 5, 80131 Napoli, Italy, IABBAM, CNR, 80147 Napoli, Italy, Centro Endocrinologia e Oncologia Sperimentale, CNR, Napoli, Italy, Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, 33100 Udine, Italy, and CEINGE Biotecnologie Avanzate, via Pansini 5, 80131 Napoli, Italy

Received June 1, 1998; Revised Manuscript Received September 10, 1998

ABSTRACT: The topology of the thyroid transcription factor 1 homeodomain (TTF-1HD)−DNA complex was investigated by a strategy which combines limited proteolysis and selective chemical modification experiments with mass spectrometry methodologies. When limited proteolysis digestions were carried out with the protein in the absence or presence of its target oligonucleotide, differential peptide maps were obtained from which the amino acid residues involved in the interaction could be inferred. Similarly, selective acetylation of lysine residues in both the isolated and the complexed homeodomain allowed us to identify the amino acids protected by the interaction with DNA. Surface topology analysis of isolated TTF-1HD performed at neutral pH was in good agreement with the three-dimensional structure of the molecule as determined by NMR studies under acidic conditions. Minor differences were detected in the C-terminal region of the protein which, contrary to NMR data, showed no accessibility to proteases. Analysis of the complex provided an experimental validation of the model proposed on the basis of the homology with the homeodomain structures described so far. An increased accessibility of the C-terminal region was observed following the interaction, suggesting its displacement from the protein core by the oligonucleotide molecule. Comparative experiments with DNA fragments differing in sequence and binding capabilities highlighted structural differences among the complexes, mainly located in the N-terminal region of the homeodomain, thus accounting for their different dissociation constants.

Selective gene expression in eukaryotic cells is achieved by the action of transcriptional regulatory proteins which function by interacting with sequence-specific DNA elements in the transcriptional control regions of their respective target genes. Interactions mediated by proteins must be capable of both providing high-affinity binding to the proper site and permitting discrimination against closely related DNA sites. The structures of a large number of protein−DNA complexes have been solved by X-ray crystallography or NMR, providing a detailed description of the protein regions that interact with the DNA bases and the sugar−phosphate backbone stabilizing the complex (*1*−*6*). These studies showed that a number of different structural motifs have evolved for the specific recognition of DNA, such as zinc finger, leucine zipper, helix−loop−helix, and homeodomain. These structural elements are usually part of multiprotein machineries that control gene expression (*7*−*9*).

However, structural studies may suffer from difficulties in the crystallization of the complexes or in their limited solubility in aqueous media. Recently, the surface accessibility of isolated proteins has been probed by selective chemical modifications of specific residues (*10*, *11*), enzymatic hydrolysis (*12*, *13*), and H−D exchange analysis (*14*). In protein complexes, the polypeptide regions occurring at the molecular interface, usually accessible in the isolated molecules, become protected following the interaction. On the basis of these observations, a series of experimental approaches which combine limited proteolysis and selective chemical modification experiments with mass spectrometric analysis have been employed to identify accessible amino acid residues. Since the contact regions of the proteins in the complex are protected against hydrolysis and labeling reactions, when the experiments are carried out with the isolated proteins and the complex, differential peptide maps can be generated from which the definition of the interface zones can be extrapolated (*15*, *16*).

Thyroid transcription factor 1 (TTF-1)[1] is a 378-residue protein responsible for transcriptional activation of genes expressed in follicular thyroid cells and lung epithelial cells (*17*, *18*). TTF-1 binds DNA by a 61-amino acid homeodomain (HD) which, contrary to other HDs, specifically recognizes oligonucleotide sequences containing the 5′-

† This work was supported in part by MURST, Progetto di Ricerca di Interesse Nazionale (PRIN) "Biologia Strutturale".

* Corresponding author: Centro Internazionale di Servizi di Spettrometria di Massa, CNR-Universita' "Federico II", via Pansini 5, 80131 Napoli, Italy. Telephone: 39-81-5465136. Fax: 39-81-5465484. E-mail: pucci@unina.it.

‡ CNR-Università di Napoli "Federico II".

§ IABBAM, CNR.

|| Centro Endocrinologia e Oncologia Sperimentale, CNR.

⊥ Università di Udine.

@ CEINGE Biotecnologie Avanzate.

[1] Abbreviations: CNBr, cyanogen bromide; ES, electrospray; HD, homeodomain; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; PTH, phenylthiohydantoin; RP-HPLC, reversed phase high-performance liquid chromatography; TFA, trifluoroacetic acid; TTF-1, thyroid transcription factor 1.

10.1021/bi981300k CCC: $18.00
Published on Web 12/12/1998

CAAG-3′ core motif (*19*−*21*); most homeodomains, in fact, bind DNA sequences containing the segment 5′-TAAT-3′. Then a structural model of the TTF-1HD−DNA complex could be relevant to understanding how distinct HDs can recognize different target DNA sequences. The three-dimensional structure of recombinant TTF-1HD has been solved by NMR analysis, whereas no structural data are available for its complex with DNA (*22*). Recently, a molecular model describing the possible interaction of TTF-1HD with its target oligonucleotide has been proposed on the basis of the structures of the HD−DNA complexes reported so far (*23*).

This paper describes the surface topology of the TTF-1HD−DNA complex determined by limited proteolysis and selective acetylation experiments combined with mass spectrometric analyses. The results reported here suggest that at physiological pH the structure of isolated TTF-1HD displays some differences with respect to that obtained by NMR at acidic pH, mainly located in the C-terminal region. Moreover, the structural analysis demonstrated that the homeodomain interacts with its target oligonucleotide essentially according to the proposed model. Finally, comparative experiments carried out with different oligonucleotides endowed with different binding capabilities highlighted the structural differences that account for their relative binding affinities.

EXPERIMENTAL PROCEDURES

Materials. The polypeptide encompassing the TTF-1HD (68 amino acids corresponding to residues 160−226 of rat TTF-1 with a methionyl residue at the N terminus, referred to as Met1) was obtained by means of overexpression in *Escherichia coli* strain BL21, by the procedure described by Guazzi et al. (*17*). The crude material was purified by ion-exchange chromatography following the same protocol as reported for the *Antennapedia* homeodomain (*24*). The overall final yield was 0.5−1 mg of pure polypeptide/L of culture. The oligonucleotides 5′-CAGTCAAGTGTTC-3′ (C), 5′-GAACAGATATCTGG-3′ (Cm), and 5′-ACGATGAGTG-GCTCATAAATCG-3′ (D) were synthesized by the oligonucleotide synthesis service at the Stazione Zoologica "A. Dohrn", Naples, and further purified. Trypsin, elastase, chymotrypsin, subtilisin, thermolysin, and acetic anhydride were purchased from Sigma; endoproteinase Glu-C was purchased from Boheringer. Solvents were HPLC-grade from Carlo Erba.

Limited Proteolysis Experiments. Limited proteolysis experiments were conducted by treating the isolated proteins or the complex separately with trypsin, chymotrypsin, elastase, endoproteinase Glu-C, thermolysin, and subtilisin. Oligonucleotides were annealed in 20 mM Tris-HCl and 75 mM KCl (pH 7.5). TTF-1HD and each oligonucleotide were mixed in a molar ratio of 1:1.2 and incubated at room temperature for 20 min prior to the protease addition. Enzymatic digestions were all performed in 20 mM Tris-HCl and 75 mM KCl (pH 7.5) at 25 °C by using an *E*:*S* ratio ranging from 1:10 to 1:10000 (subtilisin, elastase, and endoproteinase Glu-C), from 1:10 to 1:70000 (chymotrypsin), and from 1:500 to 1:50000 (trypsin). The extent of digestion was monitored on a time course basis by sampling the reaction mixture at different interval times ranging from 10 min to 3 h. To precipitate the oligonucleotide, the digested proteins samples (100 μL) were treated with the same volume of 10 M ammonium acetate and 1 mL of cold ethanol and stored at −20 °C for 24 h; then the sample was centrifuged at 13 000 rpm for 20 min, and the supernatant was removed from DNA pellets and then lyophilized. Peptide mixtures were fractionated by RP-HPLC on a Vydac C18 column (250 mm × 2.1 mm, 5 μm, 300 Å pore size); peptides were eluted by means of a linear gradient of 5 to 50% acetonitrile in 0.1% TFA over the course of 50 min, at flow rate of 0.2 mL/min, and elution was monitored at 220 and 280 nm. Individual fractions were collected and identified by ESMS.

Chemical Modification Experiments. Selective acetylation of lysine residues was performed by adding a 1−10-fold molar excess of acetic anhydride to isolated or complexed TTF-1HD, in 20 mM Tris-HCl and 75 mM KCl (pH 7.5) (final volume of 100 μL), for 10 min at 25 °C. The oligonucleotide was precipitated as reported above; the reaction mixtures were directly injected onto the RP-HPLC system and eluted by means of a linear gradient of 20 to 65% acetonitrile in 0.1% TFA over the course of 35 min at a flow rate of 0.2 mL/min, and individual fractions were collected and identified by ESMS.

Acetylated TTF-1HD samples were hydrolyzed with cyanogen bromide in 70% TFA at room temperature for 18 h in the dark. The peptide mixtures were fractionated by RP-HPLC as reported above; individual fractions were collected and identified by mass spectrometry or amino acid sequence analysis.

Mass Spectrometry and Amino Acid Sequence Analyses. Protein samples or proteolytic fragments were analyzed by ESMS using a BIO-Q triple-quadrupole mass spectrometer equipped with an electrospray ion source (Micromass, Manchester, U.K.) or by MALDIMS using a Voyager DE MALDI-TOF mass spectrometer (Perseptive Biosystem) as previously described (*16*, *25*).

The amino acid sequence of the modified peptides derived from the acetylation of TTF-1HD was determined by using a Perkin-Elmer-Applied Biosystems 477A pulsed-liquid protein sequencer equipped with a Perkin-Elmer-Applied Biosystems 120A HPLC apparatus for PTH-amino acid identification as previously reported (*11*).

RESULTS

Characterization of the Homeodomain. The recombinant homeodomain from thyroid transcription factor 1 (TTF-1HD) was analyzed by HPLC and characterized by ESMS, showing the occurrence of two equimolar species that displayed mass values of 8443.6 ± 0.7 and 8427.2 ± 0.8 Da, respectively. The first component corresponded to the expected HD molecule (theoretical mass value of 8443.8 Da), whereas the second species was shown to contain a Phe residue at position 26 instead of the normally occurring Tyr ($\Delta m = -16$ Da). Since this substitution should not alter the three-dimensional structure of TTF-1HD or its binding capability, all the experiments were performed with the mixture of the two components.

Limited proteolysis digestions were carried out on a time course basis on two different samples, isolated TTF-1HD and the complex with its target oligonucleotide, according to the strategy previously described (*13*, *16*). When present, DNA was removed from the peptide mixtures by selective

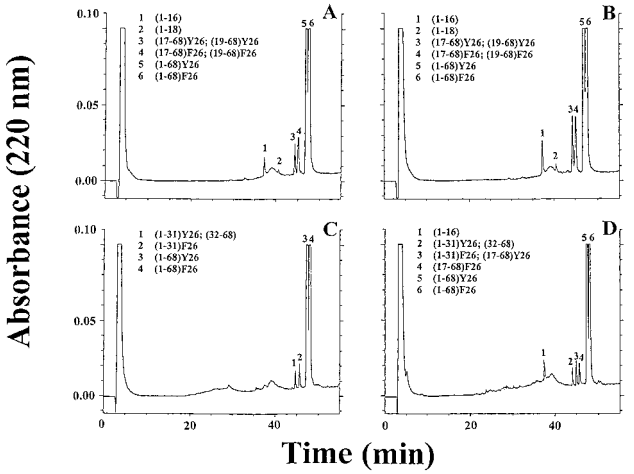


FIGURE 1: HPLC time course analysis of the endoprotease Glu-C digests of isolated TTF-1HD (*E*:*S* ratio of 1:10000) after incubation for 15 (A) and 30 min (B) and the TTF-1HD−C complex (*E*:*S* ratio of 1:10) after incubation for 15 (C) and 30 min (D). Peptides identified by ESMS are indicated. Tyr26-containing fragments are labeled with Y26; those with Phe at this position are marked with F26.

precipitation with ammonium acetate/cold ethanol (*26*); under this condition, a quantitative recovery of the peptides was obtained.

Following the formation of the complex, the presence of the oligonucleotide shields the TTF-1HD residues located within the interface region. Therefore, differential proteolytic patterns were obtained depending on whether the digestion was carried out with the isolated homeodomain or with the complex. The interpretation of the results was then based on the concept of "protection", as discussed previously (*11*, *13*, *15*, *16*). Once the appropriate conditions were established, the comparison of the differential proteolytic maps led to the identification of the cleavage sites accessible in the complex, those masked by the oligonucleotide and those exposed as a result of conformational changes occurring upon the interaction of the two components.

Limited Proteolysis Experiments. Panels A and B of Figure 1 show the time course analysis of the Glu-C digestion of TTF-1HD carried out by using an *E*:*S* ratio of 1:10000. Under these conditions, TTF-1HD remained mostly undigested, showing that its native conformation is susceptible to proteolysis at few, specific, and very likely flexible sites. After 15 min, only two peptide pairs (residues 1−16 and 17−68, and residues 1−18 and 19−68) were detected (Figure 1A), and no further fragments were released after 30 min (Figure 1B). As previously reported (*13*, *16*), the definition of the primary cleavage sites was inferred by the identification of the complementary peptides released by a single proteolytic event occurring on the intact protein. According to these criteria, uncomplexed TTF-1HD was preferentially cleaved at Glu16 and Glu18.

Similar experiments were performed following the interaction of the homeodomain with C, a synthetic oligonucleotide containing the 5′-CAAG-3′ motif. When this complex was analyzed under the same conditions, no proteolytic fragments were released even after incubation for 4 h, demonstrating a tighter conformation of the complex. Partial digestion was observed only by increasing the *E*:*S* ratio; panels C and D of Figure 1 show the time course analysis

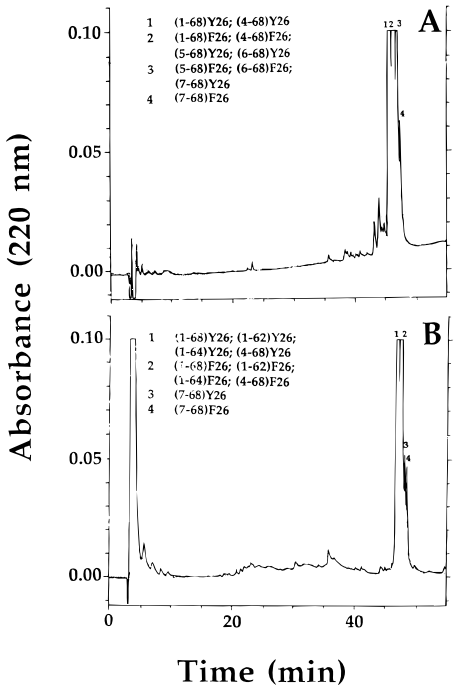


FIGURE 2: HPLC analysis of the tryptic digests of isolated TTF-1HD (*E*:*S* ratio of 1:35000) (A) and the TTF-1HD−C complex (*E*:*S* ratio of 1:1000) (B) after incubation for 3 and 1 h, respectively. Peptides identified by ESMS are indicated.

of the Glu-C hydrolysis of the complex, when an *E*:*S* ratio of 1:10 was used. A different proteolytic pattern was obtained with preferential cleavage sites occurring at Glu31 and Glu16 as demonstrated by the identification of the complementary peptides of residues 1−31 and 32−68, released at early stages of digestion (Figure 1C), and residues 1−16 and 17−68, observed after 30 min (Figure 1D). These results clearly demonstrate that, following complex formation, Glu18 is no longer accessible to endopeptidase Glu-C whereas Glu16 is still exposed to the protease action. Furthermore, a new preferential hydrolysis site was identified at position 31.

The different accessibility of the basic residues in isolated TTF-1HD and in the complex with oligonucleotide C was probed by tryptic hydrolysis. Figure 2A shows the HPLC profile obtained during a time course experiment on the isolated HD after digestion for 3 h, using an *E*:*S* ratio of 1:35000. Mass spectral analysis of the proteolytic fragments released from the isolated protein led to the identification of the peptides of residues 4−68, 5−68, 6−68, and 7−68 generated by preferential proteolytic events at Arg3, Lys4, Arg5, and Arg6. These results strongly suggested that the N-terminal segment of TTF-1HD is very flexible and highly exposed to the solvent.

A significative hydrolysis of the TTF-1HD−C complex could only be observed when an *E*:*S* ratio of 1:1000 was used, confirming the tightness of the complex. Figure 2B shows the HPLC profile obtained after digestion for 1 h. Examination of the proteolytic digest immediately showed the occurrence of a pronounced masking effect generated by the interaction; preferential cleavage sites within the N-terminal region of the homeodomain, in fact, could only be identified at Arg3 and Arg6, whereas Lys4 and Arg5 were not recognized by the protease, even after longer periods of hydrolysis. In addition, the appearance of the peptides of residues 1−62 and 1−64 indicated Lys62 and Lys64 are

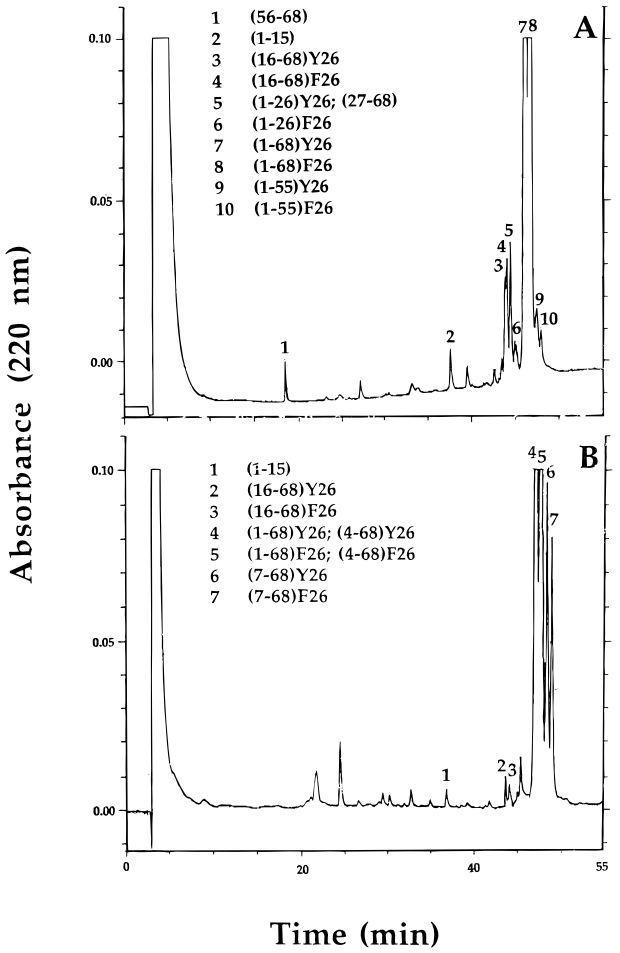


FIGURE 3: HPLC analysis of the chymotryptic digests of TTF-1HD (*E*:*S* ratio of 1:70000) (A) and the TTF-1HD−C complex (*E*:*S* ratio of 1:10) (B) after incubation for 180 and 15 min, respectively. Peptides identified by ESMS are indicated.

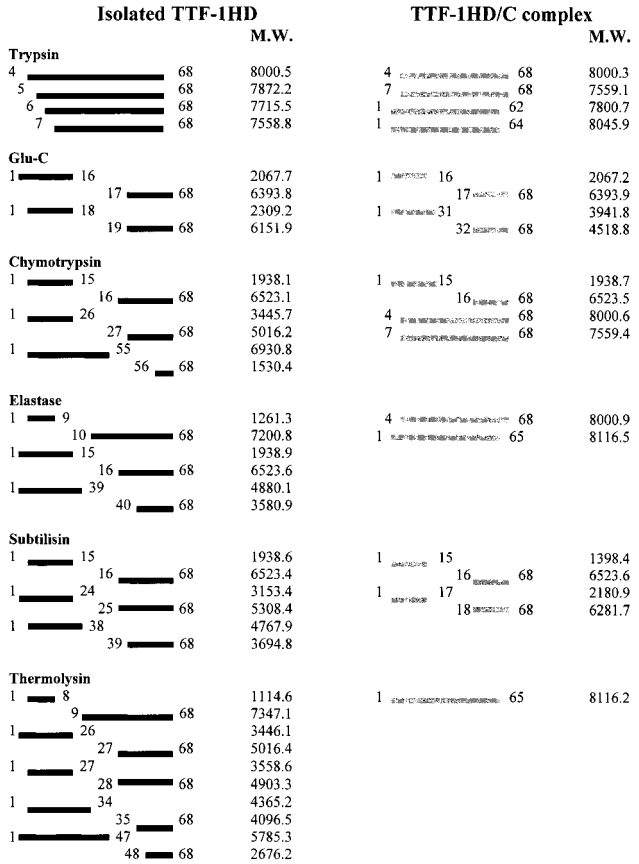
additional cleavage sites in the C-terminal region of the homeodomain. The accessibility of these residues, not previously recognized by trypsin in the isolated protein, strongly suggests a conformational change of this region within the complex structure.

Limited proteolysis experiments with specific proteases provided clear evidence for the protection effect exerted by the complex toward the homeodomain. However, they gave a limited amount of information due to the narrow specificity of the enzymes which decreased the possible number of cleavage sites. Further digestions were then carried out with broader-specificity proteases such as chymotrypsin, elastase, subtilisin, and thermolysin following the same experimental and interpretation procedures described above.

As an example, Figure 3 shows the HPLC chromatograms of the chymotrypsin digests of the isolated TTF-1HD (Figure 3A) and the TTF-1HD−C complex (Figure 3B), using 1:70000 and 1:10 *E*:*S* ratios, after 180 and 15 min, respectively. Limited proteolysis of the isolated homeodomain generated only three complementary peptide pairs, residues 1−15 and 16−68, 1−26 and 27−68, and 1−55 and 56−68, which identified Tyr15, Tyr/Phe26, and Tyr55 as the primary cleavage sites.

Following interaction with the oligonucleotide (Figure 3B), most of the accessible sites in the isolated protein were

Scheme 1: Diagram Indicating the Cleavage Site Fragments and Their Measured Mass Values in Isolated TTF-1HD (Black Lines) and in the Complex (Gray Lines)



masked as demonstrated by the disappearance of the respective proteolytic fragments in the chromatographic profile. Mass analysis led to the identification of the primary cleavage sites at Arg3 and Arg6 through recognition of the fragments of residues 4−68 and 7−68, respectively. Only a small amount of the peptide pair of residues 1−15 and 16−68 was detected, indicating the presence of a secondary hydrolysis site at Tyr15. It should be noted that usually chymotrypsin does not display specific proteolytic activity toward Arg residues; indeed, both Arg3 and Arg6 were not cleaved by this enzyme in isolated HD. Interestingly, the same two Arg residues had been recognized by trypsin in the TTF-1HD−C complex. As a whole, these data indicate that the N-terminal region of HD underwent a structural change following complex formation leading to a complete exposition of the two Arg residues in a flexible conformation. Moreover, the HD residues Tyr/Phe26 and Tyr55 seem to be either located within the interface region or involved in a conformational or rigidity change of the polypeptide chain, and hence, they are shielded from the protease action.

The overall results of the limited proteolysis experiments are summarized in Scheme 1 and Table 1, from which a number of considerations can be drawn. Preferential cleavage sites in isolated HD gathered into specific regions of the protein, the most exposed segment being the N-terminal portion spanning residues 3−9. Further accessible areas were located within the first and second helixes (residues 15−18 and 34−39, respectively) and the short loop connecting the two helices encompassing residues 24−27. Moreover, Lys47 and Tyr55 located within the recognition helix were specif-

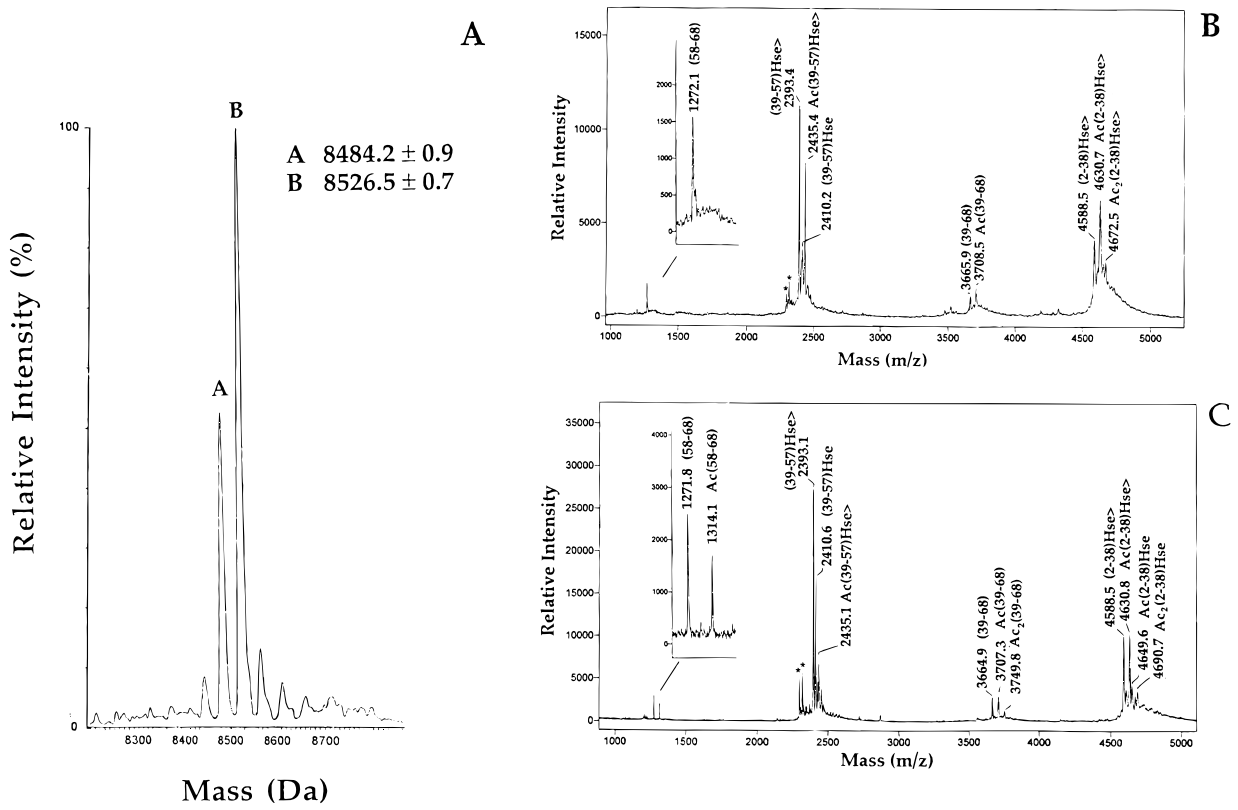


FIGURE 4: (A) ESMS analysis of acetylated TTF-1HD following HPLC purification. Mass values of the protein components are indicated. The major species corresponds to the diacetylated form of TTF-1HD, whereas the minor component represents the monomodified protein. (B) MALDI/MS analysis of the CNBr digest of the diacetylated TTF-1HD sample. (C) MALDI/MS analysis of the CNBr digest of the dimodified TTF-1HD from the acetylation of the C complex. Mass values and the corresponding peptides are indicated. Asterisks indicate double-charged ions.

Table 1: Preferential Cleavage Sites in Isolated TTF-1HD and Its Complex with Oligonucleotide C As Determined by Limited Proteolysis Experiments

	isolated TTF-1HD	complexed TTF-1HD
Glu-C	Glu16 and Glu18	Glu16 and Glu31
trypsin	Arg3, Lys4, Arg5, and Arg6	Arg3, Arg6, Lys62, and Lys64
chymotrypsin	Tyr15, Tyr/Phe26, and Tyr55	Tyr15, Arg3, and Arg6
elastase	Phe9, Tyr15, and Ile39	Arg3 and Ala65
subtilisin	Tyr15, Gln24, and Met38	Tyr15 and Leu17
thermolysin	Leu8, Tyr/Phe26, Leu27, His34, and Lys47	Ala65

ically identified. Interestingly, no cleavage sites were detected in the C-terminal region.

Following the formation of the complex, a marked protection effect exerted by the oligonucleotide was observed as demonstrated by the large decrease in the number of accessible sites, especially in the N-terminal region, the second helix, the loop of residues 24−27, and the recognition helix. Moreover, a significant number of new cleavage sites were detected in the C-terminal region of TTF-1HD.

Chemical Modification Experiments. The topology of the TTF-1HD regions involved in the interaction with oligonucleotide C was also studied by differential chemical modifications at the lysine residues (*10*, *11*). Comparative acetylation experiments on the homeodomain were then conducted in the presence or in the absence of oligonucleotide C. Appropriate experimental conditions were defined to maximize the stability of the protein conformation and to limit the extent of the modification to the desired number of amino acid residues, according to the strategy previously described (*11*).

Preliminary investigations with isolated HD using a 1:1 acetic anhydride:homeodomain molar ratio yielded a mixture of unmodified and monomodified TTF-1HD (in a 1:1 ratio). Unsuccessful Edman degradation attempts with the monoacetylated product and MALDIMS analysis of its CNBr hydrolysate, where only unmodified peptide lactones of residues 2−38, 39−57, and 58−68 were recovered, clearly demonstrated that acetylation only occurred at the N terminus of the polypeptide chain. Under the same experimental conditions, no acetylation of the TTF-1HD−C complex was observed.

Acetylation of the isolated protein using a 2:1 reagent: homeodomain molar ratio resulted in the production of mono- and dimodified TTF-1HD forms, as assessed by ESMS analysis of the HPLC-fractionated products (Figure 4A). It should be emphasized that any modified TTF-1HD component may consist of a population of isomers carrying the same number of acetyl groups but located at different sites. To identify the lysine residues acetylated in addition to the amino-terminal group, the modified homeodomain was hydrolyzed with CNBr and the resulting peptide mixture analyzed by MALDIMS (Figure 4B). Besides unmodified fragments, the MALDIMS analysis showed the presence of the peptide lactones Ac(2−38), Ac_2(2−38), and Ac(39−57), as demonstrated by the increase of 42 and 84 Da, respectively, of the corresponding mass signals. The peptide mixture was purified by RP-HPLC, and the acetylated peptides were collected and further characterized. Location

of the *N*-ϵ-Ac-Lys within the peptide of residues 2−38 was not as straightforward since this peptide contains three lysine residues at positions 4, 22, and 25. MALDIMS analysis of the endoproteinase Lys-C digest of the acetylated peptide showed the presence of fragments Ac(2−22), Ac(5−25), and Ac(23−38), each containing a single putative modification site. Therefore, this analysis demonstrates that acetylation specifically occurred at Lys4, Lys22, and Lys25. On the contrary, identification of the *N*-ϵ-acetyllysine within the peptide Ac(39−57) was directly obtained by amino acid sequence analysis. When analyzed by automated Edman degradation, the purified fragment, containing Lys47 and Lys56, both candidates for modification, showed the presence of *N*-ϵ-Ac-Lys at position 47; no trace of modified Lys56 was observed. Under these experimental conditions, no acetylation was detected at Lys56, Lys58, Lys62, and Lys64.

When the TTF-1HD−C complex was examined, a 10:1 reagent:homeodomain molar ratio was needed to observe the mono- and dimodified forms of the protein. The presence of the acetylated N-terminus was determined as described above. Furthermore, direct MALDIMS analysis of the unfractionated CNBr hydrolysate (Figure 4C) showed that acetylation mainly occurred within the peptides of residues 2−38 and 58−68; a minor extent of modification was also detected in the peptide of residues 39−57. The CNBr peptide mixture was purified by RP-HPLC, and the acetylated peptides were collected and further characterized. MALDIMS analysis of the peptide Ac(2−38) following endoproteinase Lys-C digestion showed the presence of a signal at *m*/*z* 2752.6, corresponding to the peptide Ac(5−25); this result demonstrated that acetylation only occurred at Lys22. Identification of the *N*-ϵ-acetyllysine within the peptide Ac(58−68) was directly obtained by amino acid sequence analysis. The purified fragment, containing Lys58, Lys62, and Lys64, showed the presence of *N*-ϵ-Ac-Lys only at position 64; no trace of modified Lys58 and Lys62 was observed. Similarly, the small amount of the acetylated fragment of residues 39−57 observed in the digest identified *N*-ϵ-acetyllysine at position 56. No trace of modified Lys47 was observed.

TTF-1HD Complexes with Different Oligonucleotides. The same limited proteolysis−mass spectrometry approach was employed to detect structural differences among the complexes formed by TTF-1HD with three oligonucleotides differing in their sequences. Besides its target motif, in fact, this homeodomain was demonstrated to interact with DNA fragments lacking the specific recognition sequence. In particular, gel retardation experiments had shown that TTF-1HD can bind with different affinities to two other oligonucleotides, named D and Cm (*27*). Comparative limited proteolysis experiments were then carried out with the three complexes of TTF-1HD with oligonucleotides C, D, and Cm using thermolysin and chymotrypsin as proteolytic probes.

Figure 5 shows the HPLC analysis of the aliquots withdrawn after 30 min of thermolytic digestion of the complexes TTF-1HD−Cm (panel A), TTF-1HD−D (panel B), and TTF-1HD−C (panel C) using a 1:5 *E*:*S* ratio. A qualitative comparison of the chromatographic profiles for the three different complexes clearly indicates that Cm exhibited the lowest protection effect on the homeodomain toward thermolysin action. A complete digestion of the intact protein was, in fact, observed in this case when compared

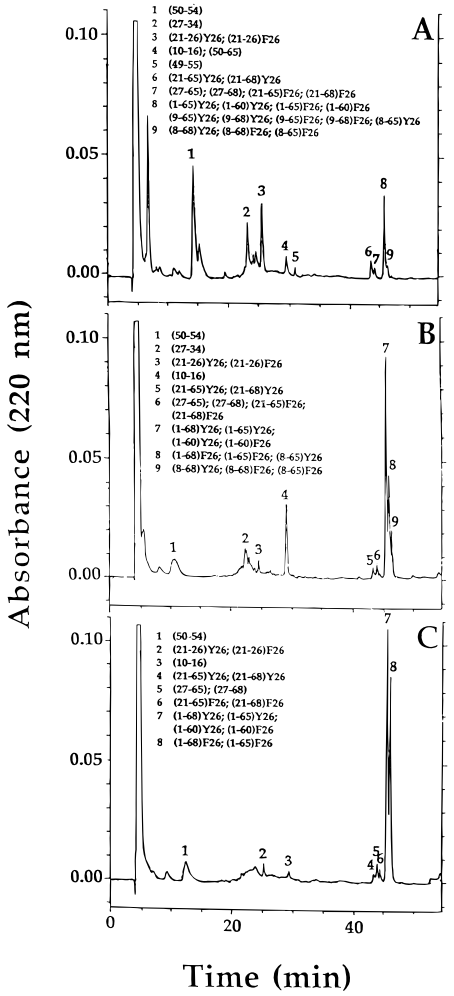


FIGURE 5: Comparative HPLC analysis of the thermolytic digests of the complexes TTF-1HD−Cm (A), TTF-1HD−D (B), and TTF-1HD−C (C) after incubation for 30 min (*E*:*S* ratio of 1:5). Peptides identified by ESMS are indicated.

to the other two complexes, as shown by the disappearance of the corresponding HPLC peak (Figure 5A). On the contrary, the highest recovery of intact HD was detected for the TTF-1HD−C complex (Figure 5C). The D oligonucleotide exhibited protection properties intermediate between those of Cm and C as shown by the amount of intact HD recovered. These results are in very good agreement with previous band shift experiments (*27*).

Besides the intact protein, the ESMS analysis of the individual fractions from the HD−C complex showed the occurrence of peptides of residues 1−60 and 1−65; the latter originated by cleavage at Ala65, as previously observed (see Scheme 1 and Table 1), whereas the former was generated by subdigestion of the longer fragment as demonstrated by time course analysis (data not shown). The same peptide fragments were detected in the analysis of the TTF-1HD−D complex (Figure 5B). However, the additional presence of peptides of residues 1−7, 8−65, and 8−68 demonstrated the occurrence of another hydrolysis site at Val7.

These results were confirmed by incubation of the three complexes with chymotrypsin (data not shown). Again, Cm exerted the lowest protection effect on the homeodomain, whereas the highest recovery of intact HD was observed with the TTF-1HD−C complex. Moreover, analysis of the frag-

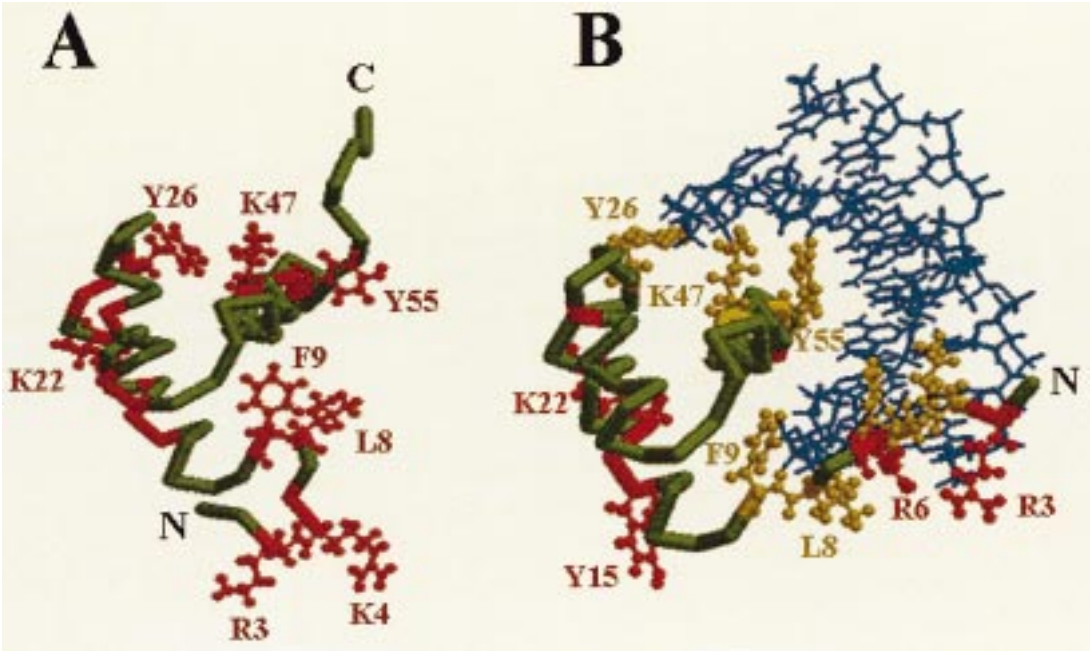


FIGURE 6: Three-dimensional structure of TTF-1HD as determined by Esposito et al. (*22*) (A) and the model proposed for its complex with oligonucleotide C (*23*) (B). Amino acid residues recognized by proteases or chemically modified in both cases are red. The side chains of residues exposed in the isolated protein but masked in the complex are yellow.

ment released from the HD−D complex by ESMS identified an additional hydrolysis site at Leu8.

These data clearly indicate that the three complexes possess slightly different structures; i.e., TTF-1HD binds the oligonucleotides in a slightly different way. These observations are in agreement with both electrophoretic experiments and the relative binding affinities measured for the oligonucleotides (*27*).

DISCUSSION

Limited proteolysis and chemical modification experiments were employed to study the surface topology of the TTF-1HD−DNA complex on the grounds that the interface regions of the protein that are exposed in the isolated molecule, and hence accessible to proteases and chemical reagents, become shielded following complex formation. When comparative experiments were carried out with both the isolated protein and the complex, differential peptide maps were obtained from which the interface regions could be inferred.

As the first stage, the surface accessibility of isolated TTF-1HD was investigated. Limited proteolysis results were essentially consistent with the homeodomain structure determined by NMR at pH 4 (*22*) and reported in Figure 6A, where the observed hydrolysis sites are red. Preferential cleavage sites were mainly grouped within four regions of the HD structure (Table 1 and Scheme 1), the N-terminal segment (Arg3, Lys4, Arg5, Arg6, Leu8, and Phe9), the first and the second helixes (Tyr15, Glu16, and Glu18 and His34, Met38, and Ile39, respectively), and the connecting loop (Gln24, Tyr/Phe26, and Leu27). Two specific residues (Lys47 and Tyr55) could also be identified in the recognition helix. These results confirmed that the N-terminal region of HD is very flexible and exposed, as it was cleaved at nearly any amino acid residue, in perfect agreement with the NMR structure depicting an N-terminal arm protruding from the protein core and endowed with a high degree of conformational freedom.

On the contrary, no accessible residues were identified at the C terminus of the homeodomain in contrast with the three-dimensional structure determined by NMR that described a disordered C-terminal segment fully exposed to the solvent. This discrepancy might depend on the different experimental conditions used, i.e., the different pH values (ΔpH = 3.5), possibly determining differential degrees of protonation at Lys62, Asp63, and Lys64. Potential salt bridges between the Asp63 carboxylate and the amino and guanidino groups of Lys56, Lys58, and Arg59 have, in fact, been described even at acidic pH (*22*). These interactions might become prominent at physiological pH due to favorable changes in the protonation state of these residues leading to an accommodation of the C terminus on the protein surface that prevented protease recognition of the peptide bonds within the region of residues 56−64.

To further investigate the accessibility of basic residues in isolated TTF-1HD, the protein was selectively acetylated under controlled conditions. Studies of the chemical reactivity of the eight Lys residues showed the high susceptibility of Lys4, Lys22, Lys25, and Lys47, indicating that these residues are easily accessible in the isolated homeodomain. These data are essentially in agreement with results from limited proteolysis experiments. As shown in Figure 6A, in fact, enzymatic hydrolyses and chemical modifications occurred at amino acid side chains exposed on the protein surface whereas the inner hydrophobic core became inaccessible.

However, Lys22 and Lys25 were not recognized by trypsin and could only be identified by chemical modification, yet both residues are located in the exposed loop connecting the first and second helixes. This result might be interpreted in terms of a local conformational rigidity of the polypeptide chain that prevented tryptic hydrolysis. It should be noted that such an apparent discrepancy had already been observed

in the analysis of Minibody, highlighting the complementarity of the limited proteolysis and/or chemical modification approaches (*11*). Lysine residues at positions 56, 58, 62, and 64 were never modified under the experimental conditions used, confirming the structured conformation of the C terminus.

A considerably different surface topology of TTF-1HD was observed as a result of the interaction with oligonucleotide C. The occurrence of a shielding effect exerted by the DNA molecule on the protein was clearly detected as indicated by the general lower accessibility of the complex with respect to the isolated protein. Proteolysis occurred only following a considerable increase in the protease concentration, suggesting a more structured and rigid conformation of TTF-1HD. This was also confirmed by the large decrease in the number of preferential cleavage sites observed in the complex as compared to those found in the isolated molecule (Table 1). Similarly, acetylation occurred at a reduced number of Lys residues and only by the addition of a 5-fold molar amount of acetic anhydride with respect to the isolated homeodomain.

Our results were instrumental in the verification of the molecular model proposed to describe the interaction of TTF-1HD with its target oligonucleotide sequence (5′-CAAG-3′) (Figure 6B) (*23*). According to the model, the recognition helix (residues 44−55) should lie into the major groove of the DNA, establishing specific contacts with the nucleotides (Ile48, Gln51, Asn52, and Tyr55) and with the sugar−phosphate backbone (Gln45, Lys47, Trp49, and His53) (*21*, *23*). Both limited proteolysis and acetylation experiments showed that two residues in this helix, Lys47 and Tyr55 that had been found exposed in the isolated protein, were totally protected following interaction with oligonucleotide C, affording an experimental validation of the proposed model in this region.

Following complex formation, the N-terminal region of HD could only be cleaved at Arg3 and Arg6, whereas Lys4, Arg5, Leu8, and Phe9 were significantly shielded by oligonucleotide C (see Figure 6B). Moreover, Lys4, selectively acetylated in isolated HD, was completely protected even toward chemical modification in the complex. These results clearly showed that following interaction with DNA, the N-terminal arm of TTF-1HD undergoes a conformational change, intercalating into the minor groove of the double helix. In addition, experimental data demonstrated that specific interactions in the complex involve Lys4 and Arg5 and not Arg6 as predicted from molecular modeling studies.

The amino acid residues occurring in the loop connecting helix 1 and helix 2, Gln24, Lys25, Tyr26, and Leu27, selectively hydrolyzed or acetylated in the isolated protein, were completely masked following interaction with the oligonucleotide. This observation is in good agreement with data on other HD−DNA complexes, where Tyr26 interacts directly with DNA (*28*, *29*). The remaining three residues became inaccessible probably because of the steric hindrance and the conformational rigidity of this region caused by the interaction of the aromatic residue with the oligonucleotide molecule. On the contrary, Lys22 was selectively acetylated both in the absence and in the presence of the oligonucleotide, thus confirming its location on the molecular surface just outside the interacting loop (Figure 6B).

DNA binding seems also to have a local influence on the molecular conformation of the homeodomain. Glu31, involved in the N-capping motif with Ser28 and Leu17 in the isolated protein (*22*), in fact, became accessible to proteases in the complex. Most important, a general structural rearrangement of the C terminus was observed. This region was not included in the TTF-1HD−DNA model reported in Figure 6B, and no prediction was available. According to our results, this region became completely exposed to enzyme action following DNA interaction. Proteolytic cleavages were, in fact, detected at Lys62, Lys64, and Ala65 with acetylation specifically occurring at Lys64. These data indicate that the C-terminal portion of HD is displaced from the protein surface by the interaction with oligonucleotide C, assuming a disordered and flexible conformation. These findings suggest a direct effect of DNA binding on the molecular conformation of the entire TTF-1 protein. A number of recent reports support the hypothesis that conformational changes upon binding DNA may be the rule rather than the exception, in DNA-binding proteins (*30*, and reference therein). In this respect, the displacement of the TTF-1HD C-terminal tail might affect the C-terminal domain of the protein, causing this region to assume a biologically relevant conformation. Future studies will be needed to determine whether this phenomenon might be directly associated with the biological functions of the transcription factor or due to the peculiar structure of the recombinant homeodomain that was studied.

The overall results presented here essentially confirmed the predicted model for the interaction of TTF-1HD with its target oligonucleotide, although some refinements are needed. Recently, the three-dimensional structure of vnd/NK-2 HD, a homeodomain that in a manner similar to that of TTF-1HD specifically recognizes the sequence 5′-CAAGTG-3′, has been solved resulting in a model similar to that proposed for TTF-1HD complex (*31*).

TTF-1HD was shown to interact with other DNA fragments lacking the target sequence (5′-CAAG-3′), such as oligonucleotides D and Cm, although these complexes were characterized by different equilibrium dissociation constants as compared to those of the HD−C complex (*27*). When comparative limited proteolysis experiments were carried out with the three complexes, an excellent agreement between structural data and thermodynamic parameters was observed. In particular, oligonucleotide Cm was immediately shown to exert a very poor protection effect on the HD structure, in agreement with the high value of the dissociation constant observed for this molecule. When oligonucleotides C and D were used, comparable results were obtained for the two HD−DNA complexes. However, a detailed examination of the data showed that, following the addition of oligonucleotide C, a strong protection effect at Leu8 and Phe9 was obtained. These residues have been described to play a significant role in sequence-specific interaction at the 5′-position of the recognition motif (*21*). When the same experiments were performed with oligonucleotide D, an essentially similar proteolytic pattern was observed with the exception that Val7 and Leu8 were recognized by proteases.

These results clearly showed that TTF-1HD interacts with oligonucleotide D by adopting the same general fold described for the HD−C complex but displaying some differences in the N-terminal arm. In the HD−D complex,

this segment intercalates into the minor groove of DNA but it is not able to make the appropriate specific contacts observed for the HD−C complex. These data again demonstrate the influence of residues located in the N-terminal region of the homeodomain in contributing to the sequence binding specificity.

In this respect, HD-containing proteins greatly differ from prokaryotic transcriptional regulators. These proteins, in fact, bind to DNA as dimers, recognizing base pairs through amino acids of the helix−turn−helix motif only. Because of this mode of DNA binding, prokaryotic transcriptional regulators show a remarkable specificity of sequence recognition. In contrast, both in vitro and in vivo experiments demonstrate that HD-containing proteins may efficiently bind to different sequences and, therefore, could control the expression of a large number of target genes (*32*). Accordingly, to explain the different biological effects of HD-containing proteins, a "widespread binding" model has been recently proposed (*33*). The alternative modes of using the N-terminal arm observed in this study might constitute the molecular mechanism for providing the HD structure with the necessary binding flexibility for a widespread DNA-binding mode.

ACKNOWLEDGMENT

We thank Prof. Gennaro Marino for helpful discussion and Dr. Gennaro Esposito for providing the molecular coordinates of the TTF-1HD model.

REFERENCES

1. Pabo, C., and Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
2. Gehring, W. J., Affolter, M., and Burglin, T. (1994) *Annu. Rev. Biochem. 63*, 487−526.
3. Wolberger, C. (1996) *Curr. Opin. Struct. Biol. 6*, 62−68.
4. Chytil, M., and Verdine, G. L. (1996) *Curr. Opin. Struct. Biol. 6*, 91−100.
5. Burley, S., Xie, X., Clark, K., and Shu, F. (1997) *Curr. Opin. Struct. Biol. 7*, 94−102.
6. Choo, Y., and Klug, A. (1997) *Curr. Opin. Struct. Biol. 7*, 117−125.
7. Chen, L., Glover, J. N. M., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) *Nature 392*, 42−48.
8. Batchelor, A., Piper, D. E., de la Brousse, F. C., McKnight, S. L., and Wolberger, C. (1998) *Science 279*, 1037−1041.
9. Tan, S., and Richmond, T. J. (1998) *Nature 391*, 660−666.
10. Glocker, M. O., Borchers, C., Fiedler, W., Suckau, D., and Przybylski, M. (1994) *Bioconjugate Chem. 5*, 583−590.
11. Zappacosta, F., Ingallinella, P., Scaloni, A., Pessi, A., Bianchi, E., Sollazzo, M., Tramontano, A., Marino, G., and Pucci, P. (1997) *Protein Sci. 6*, 1901−1909.
12. Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) *Folding Des. 266*, 223−230.
13. Zappacosta, F., Pessi, A., Bianchi, E., Venturini, S., Sollazzo, M., Tramontano, A., Marino, G., and Pucci, P. (1996) *Protein Sci. 5*, 802−813.
14. Zhang, Z., and Smith, D. L. (1993) *Protein Sci. 2*, 522−531.
15. Cohen, S. L., Ferrè-D'amare, A. R., Burley, S. K., and Chait, B. T. (1995) *Protein Sci. 4*, 1088−1099.
16. Scaloni, A., Miraglia, N., Orrù, S., Amodeo, P., Motta, A., Marino, G., and Pucci, P. (1998) *J. Mol. Biol. 277*, 945−958.
17. Guazzi, S., Price, M., Damante, G., Mattei, M., and Di Lauro, R. (1990) *EMBO J. 9*, 3631−3639.
18. Damante, G., and Di Lauro, R. (1994) *Biochim. Biophys. Acta 1218*, 255−266.
19. Damante, G., and Di Lauro, R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5388−5392.
20. Damante, G., Fabbro, D., Pellizzari, L., Civitareale, D., Guazzi, S., Polycarpou-Schwartz, M., Cauci, S., Quadrifoglio, F., Formisano, S., and Di Lauro, R. (1994) *Nucleic Acids Res. 22*, 3075−3083.
21. Damante, G., Pellizzari, L., Esposito, G., Fogolari, F., Viglino, P., Fabbro, D., Tell, G., Formisano, S., and Di Lauro, R. (1994) *EMBO J. 15*, 4992−5000.
22. Esposito, G., Fogolari, F., Damante, G., Formisano, S., Tell, G., Leonardi, A., Di Lauro, R., and Viglino, P. (1996) *Eur. J. Biochem. 241*, 101−113.
23. Fogolari, F., Esposito, G., Viglino, P., Damante, G., and Pastore, A. (1993) *Protein Eng. 6*, 513−519.
24. Muller, M., Affolter, M., Leupin, W., Otting, G., Wuthrich, K., and Gehring, W. J. (1988) *EMBO J. 7*, 4299−4304.
25. Tell, G., Scaloni, A., Pellizzari, L., Formisano, S., Pucillo, C., and Damante, G. (1998) *J. Biol. Chem.* (in press).
26. Greig, M., Klopchin, P., Griffey, R., and Hughes, J. (1995) HP Application Note MS 95-3, Hewlett-Packard, Palo Alto, CA.
27. Fabbro, D., Tell, G., Leonardi, A., Pellizzari, L., Pucillo, C., Lonigro, R., Formisano, S., and Damante, G. (1996) *Nucleic Acids Res. 24*, 3283−3288.
28. Gehring, W. J., Qiu Qian, Y., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994) *Cell 78*, 211−223.
29. Qiu Qian, Y., Furukubo-Tokunaga, K., Resendez-Perez, D., Muller, M., Gehring, W. J., and Wuthrich, K. (1994) *J. Mol. Biol. 238*, 333−345.
30. Lefstin, J. A., and Yamamoto, K. R. (1998) *Nature 392*, 885−888.
31. Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1997) *Biochemistry 36*, 5372−5380.
32. Walter, J., Dever, C., and Biggin, M. D. (1994) *Genes Dev. 8*, 1678−1692.
33. Biggin, M. D., and McGinnis, W. (1997) *Development 124*, 4425−4433.

BI981300K