Volume 15 Number 18 1987 Nucleic Acids Research

Properties of a CCAAT box-binding protein

R.A.M.Hooft van Huijsduijnen, J.Bollekens, A.Dorn, C.Benoist and D.Mathis

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg cedex, France

Received July 21, 1987; Accepted August 20, 1987

ABSTRACT

NF-Y is a sequence-specific DNA-binding protein that interacts with the conserved Y motif of the major histocompatibility complex class II gene, E_α. Since it is actually a CCAAT box-binding protein, NF-Y also attaches to other promoters bearing CCAAT sequences; yet, it is neither of the previously described transcription factors, CBP or CTF/NF-1. In this report, we document the cell-type distribution and various biochemical properties of NF-Y. The most important findings are that this protein is ubiquitously distributed, that it is probably a metallo-protein, that it has a protease-resistant DNA-binding domain and that the NF-Y/E_α-oligo complex seems extremely large (>200kD). These data should prove useful in comparisons of NF-Y with other sequence-specific DNA-binding proteins; they have already provided new insights into NF-Y's structure.

INTRODUCTION

Eukaryotic promoters seem to be modular - each composed of a particular set of short DNA sequence motifs (for reviews, see 1 and 2). The various motifs act as binding sites for regulatory proteins, and it is probably the interaction of these proteins with their target sequences and with each other that confers promoter specificity (for reviews, see 3 and 4). To understand this interplay, it will first be necessary to identify and characterize all the relevant proteins.

The detection of sequence-specific DNA-binding proteins has been rendered almost commonplace by recent technological developments. "Gel retardation" and "footprinting" assays, in particular, have permitted a rapid accummulation of data (5,6) ; careful collation of these data has revealed unexpected subtleties and complexities. For example, a given DNA sequence motif may be the target of two or three different proteins (eg. 7-9) ; conversely, a particular protein may recognize several distinct but related binding sites (eg. 9-11). One transcription factor has even been shown to bind to two apparently unrelated promoter sequences (12). Most disconcertingly, independently isolated factors have turned out to be

the same protein even though they were purified on the basis of separate in vitro activities, and despite the fact that they recognized motifs whose similarity was not obvious on first inspection (13,14).

Thus, it will be important to extensively characterize any newly detected DNA-binding activity, comparing it with already described proteins, before proceeding with laborious purification procedures. Criteria for comparison should include properties of the binding site, as revealed by DNAse I protection, methylation interference or site-specific mutagenesis experiments. In addition, some attempt should be made to characterize the protein biochemically. Distinctive features might include : sensitivity to various disruptants (temperature, urea, salt, chelators), protease sensitivity and apparent molecular weight. These properties should also serve in verifying that any purified factor is indeed the entity that was detected in the crude extract, a verification necessitated by the observation that more than one protein can bind to the same DNA sequence motif.

Herein, we report the cell-type distribution and various biochemical properties of NF-Y, a protein that binds specifically to the promoter of a major histocompatibility complex (MHC) class II gene, E_α (9,15). These data have provided new insights into the structure of NF-Y. We anticipate that they will also prove useful in comparisons of NF-Y with other sequence-specific DNA-binding proteins.

MATERIALS AND METHODS

Oligonucleotides :

Double-stranded oligonucleotides (oligos) were prepared as described previously (9,15). Briefly, complementary strands were synthesized on discs or on a Gene Assembler (Pharmacia), one strand was [^{32}P] 5'-endlabelled, the two complementary strands were annealed, and the double-stranded oligos were gel-purified. The oligos used in this paper were :

E_α 39mer	5' GTCTGAAACATTTTTCTGATTGGTTAAAAGTTGAGTGCT 3'
	3' CAGACTTTGTAAAAAGACTAACCAATTTTCAACTCACGA 5'
E_α 22mer	5' ATTTTTCTGATTGGTTAAAAGT 3'
	3' TAAAAAGACTAACCAATTTTCA 5'
TK 23mer	5' GCGTCTTGTCATTGGCGAATTCG 3'
	3' CGCAGAACAGTAACCGCTTAAGC 5'

Gel Retardation and Methylation Interference Assays :

Nuclear extracts were prepared and gel retardation assays performed as detailed previously (9,15). In short, a 1 M salt extract of nuclei was incubated with the [^{32}P]-labelled test oligo and various amounts of poly[d(I-C)] competitor, the free DNA and DNA:protein complexes were separated on a polyacrylamide gel, and the resulting bands were revealed by autoradiography.

Procedures involved in methylation interference mapping have also been described (9). Briefly, [^{32}P]-labelled double-stranded oligo was partially methylated and then incubated with nuclear extract in the presence of poly[d(I-C)]. Free and protein-bound DNA were separated, extracted, cleaved with piperidine (A+G) and electrophoresed on a sequencing gel.

Protease treatment :

Binding reactions were set up as usual (9,15). After a 20 min incubation on ice, various proteases (Proteinase K, Merck; Subtilisin A, Serva; Bromelain, Biochemical Research; Trypsin Type I, Sigma) were added. The reaction mixtures were incubated for a further 10 min at room temperature and then loaded onto a 5% polyacrylamide gel and processed in the standard fashion (9,15).

Heat treatment :

Nuclei were extracted as usual with 4 volumes of 1M NaCl - 0.5mM $MgCl_2$ - 0.2mM EDTA - 20mM Hepes (pH7.9) - 25% glycerol -0.5mM dithiothreitol, containing a cocktail of protease inhibitors (9). Undiluted crude nuclear extract was dispensed into 500 µl Eppendorf tubes (2 µg protein/µl ; 3 µl/tube). For each point on the temperature curve, an aliquot was incubated in a water bath for 5 min. Samples were kept on ice before and after heating, such that all were in the thawed state for the same length of time. One µl of each aliquot was used in a standard gel retardation assay.

Treatment with chelators :

The various chelators were made 10 mM in binding buffer. 1,10-(ortho-) phenanthroline (OP ; Sigma) was dissolved in a small volume of hot water, the solution was further diluted and the pH adjusted to 7.1 and lastly, the other components of the binding buffer were added. Lomofungin (Upjohn Co.) was dissolved as described (16). The binding reactions were conducted in binding buffer + chelator diluted to various degrees with binding buffer.

Pore-gradient gel electrophoresis :

Linear 4-10% polyacrylamide gels were poured using a gradient maker. The buffer was 0.5 X TBE [1 x TBE = 89mM Tris-base (pH8.5) - 89mM Boric Acid - 2.5mm EDTA] ; the 10% polyacrylamide solution also contained 3% sucrose. The gels were run 150 V at 4° for 20-24 hrs. Under these conditions, the tracking bromophenol blue migrates off the gel in about 5 hrs. After the run, the gel was fixed in 3.5% perchloric acid, stained with Coomassie Brilliant Blue G-250 as described previously (17), dried and exposed. The molecular weight markers were thyroglobulin (669 kD), β-galactosidase (520 kD), phosphorylase a (380 kD), catalase (232 kD), bovine serum albumin (BSA ; 68 kD and multimers) and ovalbumin (43 kD).

Glycerol gradient centrifugation :

Before centrifugation, 100 µl ($4x10^6$cpm) of crude nuclear extract from ^{35}S-methionine labelled LMTK cells was concentrated and fractionated by ammonium sulfate precipitation. Saturated ammonium sulfate was added to the extract to a final concentration of 16% saturation, the mixture was incubated 1 h on ice, and the precipitated proteins pelleted by centrifugation (15 min in an Eppendorf centrifuge, 11,500 rpm, 4°C). The supernatant was again mixed with a saturated solution of ammonium sulfate - this time to a final concentration of 30% saturation - kept on ice for 1 h, and centrifuged at 11,500 rpm. This second pellet was suspended in 200 µl of 25 mM Tris-HCl (pH 7.75) - 1 mM dithiothreitol- 75 mM NaCl - 5% glycerol and layered onto a linear 10 to 30% glycerol gradient (4 ml) prepared in 50 mM Tris-HCl (pH 7.75) - 2 mM dithiothreitol - 150 mM NaCl - 0.4 mg/ml BSA (18). A parallel glycerol gradient was loaded with a mixture of protein standards : phosphorylase b (94 kD), ovalbumin (43 kD), carbonic anhydrase- (30 kD), soybean trypsin inhibitor (20 kD), α-lactalbumin (14 kD). After centrifugation in a Beckman SW60 rotor at 48,000 rpm for 16 h (0°C), the gradients were collected from the bottom of the tube. 19 fractions (200 µl) were assayed for DNA-binding activity (to the E_α 39 mer) and were also analyzed by SDS-polyacrylamide gel electrophoresis (19). The gels were treated with 2,5-diphenyloxazole (PPO), dried and autoradiographed. The markers on the parallel gradient were visualized by Serva blue staining after electrophoresis on a 10% SDS-polyacrylamide gel.

RESULTS AND DISCUSSION

NF-Y is a sequence-specific DNA-binding protein that interacts with the promoter of the E_α gene, a murine MHC class II locus (9,15). It was first identified by the gel retardation assay as binding to the Y motif —

$T^{T}_{G}CTGATTGG^{T}_{C}T^{AA}_{CC}$— a sequence conserved amongst all MHC class II genes, and required for the efficient and accurate initiation of transcription on E_{α} in transgenic mice (15). Definitive evidence of this localization has been provided by deletion mutagenesis, DNase I footprinting, oligonucleotide binding, methylation interference and site-specific mutagenesis experiments (9,15). NF-Y is in fact a CCAAT box-binding protein and as such binds to some, but not all, promoters bearing CCAAT sequences : human α-globin, the murine sarcoma virus LTR, rat albumin (9 and unpublished results). NF-Y is not, however, either of the previously characterized CCAAT box-binding proteins, CBP and CTF/NF-1 (9).

The NF-Y binding site has already been extensively analyzed and compared with that of other previously identified DNA-binding proteins (9). We have now sought to more completely characterize NF-Y by studying its cell-type distribution and various biochemical properties.

Cell-type distribution

E_{α} is subject to elaborate regulation, being transcribed primarily in macrophages and other antigen presenting cells, B cells, and thymic epithelial and dendritic cells. In the B lineage, this gene is regulated according to the state of differentiation : it is expressed in mature B cells but not in immature preB precursors nor in terminally differentiated plasma cells. E_{α} expression is induced in macrophages by γ-interferon and in B cells by interleukin-4 (for review see 20).

The cell-type distribution of NF-Y was assessed by the gel retardation assay : a ^{32}P-labelled E_{α} 39mer was incubated with poly [d(I-C)] and extracts of nuclei from various cell types, the free and protein-complexed DNAs were separated on a polyacrylamide gel, and the bands were revealed by autoradiography. As illustrated in Figure 1A, all cells so far examined exhibit (NF-Y)-like activity. Included are mouse cells of the lymphoid lineage, a macrophage line, a fibroblast, an epithelial cell and a teratocarcinoma (see Figure legend for details). Included are cells which express the E_{α} gene (M12, A20, P388-D1) and cells which do not (70Z3, BW5147, LMTK, MCA and T-113). (NF-Y)-like activity is not significantly increased by γ interferon treatment of a macrophage line (P388 D1) even though E_{α} expression is induced 20-50 fold by this treatment. NF-Y, or a similar protein, is also detected in a human B lymphoblastoid line, Raji.

With some of the extracts (A20, BW5147, P388-D1), quite prominent lower bands are observed in Figure 1A, but these probably reflect non-specific DNA:protein interactions and/or protein degradation. Their presence is variable for any given cell-type, depending on the particular prepa-

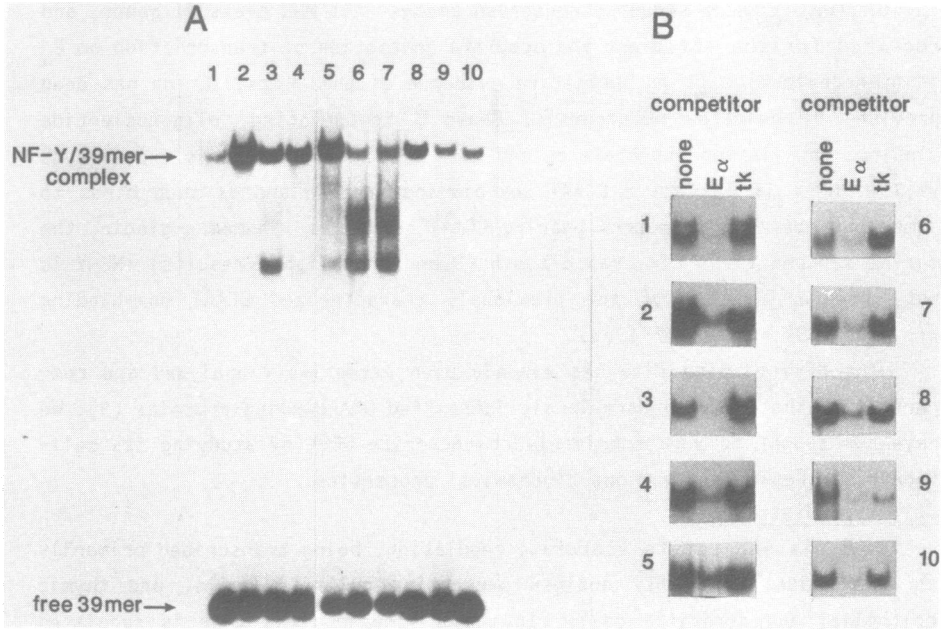

FIGURE 1: **A. Cell-type distribution of NF-Y.**
A ^{32}P-labelled E_α 39mer (4000cpm) was mixed with 150 ng of poly [d(I-C)] and extracts of nuclei from various cell lines. After incubation, the reaction mixtures were loaded onto a 5% polyacrylamide gel and electrophoresed. Bands were revealed by autoradiography. The cell lines were : 1, Raji - a human B lymphoblastoid line ; 2, M12 - a B lymphoma line ; 3, A20 - a B lymphoma line ; 4, 70Z3 - a pre B cell lymphoma line ; 5, BW5147 - a thymoma line ; 6 and 7, P388-D1 - a macrophage-like line ; 8, LMTK - a fibroblast ; 9, MCA - an epithelial cell and 10, T-113 - a teratocarcinoma (21). P388 D1 was treated (7) or not (6) with 50 units/ml of γ interferon (Boerhinger Ingleheim) for 24 hrs.

B. Cross-competition experiments using extracts from various cell types.

Binding reactions were set up as above, or were supplemented with 200 fmoles of unlabelled E_α 22mer or TK 23mer as competitor (See Materials and Methods for sequences.) The cell lines are numbered as in A.

ration ; they are seen with negative control DNAs such as a pBR322 fragment and an E_α 39 mer that has random sequence replacing the Y-box (15, data not shown).

Other cell-lines that have been tested and shown to exhibit (NF-Y)-like activity include : TA3, a B lymphoma-B cell hybrid ; CH27, CH37 and WEHI 231, all B lymphoma lines ; X-63, a plasmacytoma ; NIH 3T3, a fibroblast ; three macrophage hybridomas (+/- γ interferon) ; and WEHI 3, a cell

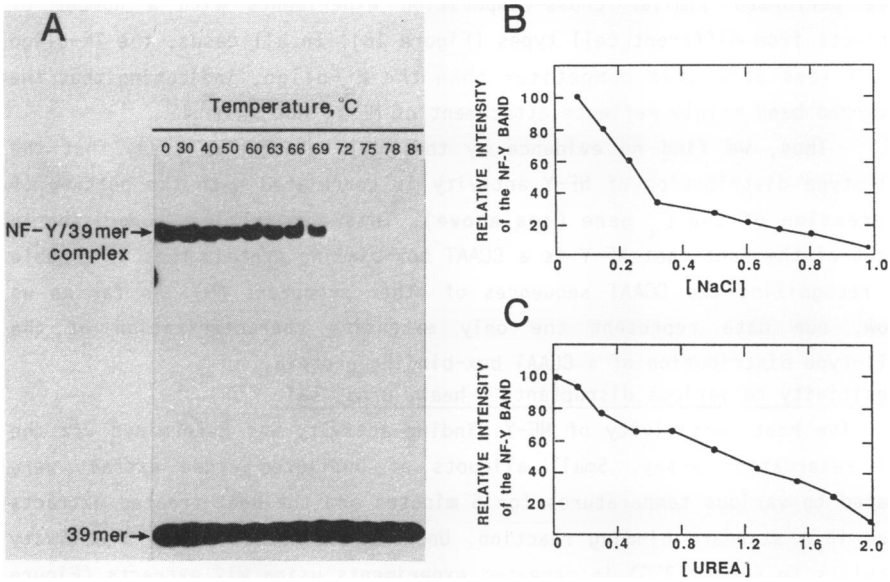

FIGURE 2: **A. Heat-sensitivity of NF-Y activity.**
Small aliquots of M12 nuclear extract were incubated 5 min at the indicated temperature, cooled on ice and assayed for binding activity by the gel retardation assay.
B and C. Salt and urea sensitivity of NF-Y activity.
Reactions were set up as usual, except that the binding buffer was supplemented with various amounts of NaCl (B) or urea (C). The reaction mix was electrophoresed, the gel autoradiographed, and the intensity of the bands representing the NF-Y/E_α-oligo complex quantitated by densitometry. The 100% value indicates the amount of complex formed under standard conditions, ie 80mM NaCl, no urea.

of the macrophage-monocyte lineage. In addition, nuclear extracts from fresh mouse tissue - liver, spleen, thymus - and even from completely undifferentiated teratocarcinoma tissue (T-103 ; see 21) proved positive (data not shown).

A similar band was observed with all of the extracts, but one can not assume that this band always reflects binding of the same protein. In fact, using the same assay conditions, we have detected another protein, NF-Y* that binds to the CCAAT box of the herpes simplex thymidine kinase gene (tk) to form a complex that comigrates with the NF-Y/E_α-oligo complex (9). In competition experiments (using an M12 extract), unlabelled TK-oligo competes ten times less efficiently than unlabelled E_α-oligo for NF-Y binding to ^{32}P-labelled E_α-oligo ; conversely, E_α-oligo does not compete as efficiently as TK-oligo for NF-Y* binding to the ^{32}P-labelled TK-oligo. We

have performed similar cross-competition experiments with a number of extracts from different cell types (Figure 1B). In all cases, the TK-oligo was a less effective competitor than the E_α-oligo, indicating that the retarded band mainly reflects attachment of NF-Y, not NF-Y*.

Thus, we find no evidence by this gel retardation assay that the cell-type distribution of NF-Y activity is correlated with the pattern of expression of the E_α gene (see above). This appears less surprising in light of the fact that NF-Y is a CCAAT box-binding protein that is capable of recognizing the CCAAT sequences of other promoters (9). As far as we know, our data represent the only extensive characterization of the cell-type distribution of a CCAAT box-binding protein.

Sensitivity to various disruptants : heat, urea, salt

The heat sensitivity of NF-Y binding activity was determined via the gel retardation assay. Small aliquots of undiluted crude extract were heated to various temperatures for 5 minutes and the heat-treated extracts used in a standard binding reaction. Under these conditions NF-Y activity persists to 69° ($\pm$ 3°C) in repeated experiments using M12 extracts (Figure 2A). The abruptness of the active/inactive transition is quite striking.

Diluting the extract in binding buffer before heating markedly reduces the heat stability. Salt, divalent cations, extraneous DNA, and other protein could all affect this property so that when comparing the temperature sensitivity of different DNA-binding proteins, rigorous care must be taken to standardize conditions.

NF-Y binding is quite resistant to both salt and urea. There is less than a four-fold reduction in the amount of NF-Y/E_α-oligo complex formed at an NaCl concentration of 0.5 M or at a urea concentration of 1.8 M, as compared with the standard conditions of 80 mM NaCl, no urea. (Figure 2 B and C).

Sensitivity to various chelators – evidence that NF-Y is a metalloprotein

Many of the proteins that catalyze reactions involving nucleic acids are metallo-enzymes or require metal ions for activity (for review, see 22). Attempts to stimulate NF-Y binding to the E_α 39mer by adding various amounts of Ca^{2+}, Mg^{2+}, Mn^{2+} or Zn^{2+} salts were unsuccessful, implying that NF-Y either does not need these cofactors or that the trace amounts present in binding buffer or the nuclear extracts suffice for optimum activity. In order to determine whether metal ions are required for NF-Y activity, we supplemented the binding buffer with increasing amounts of various metal-chelating compounds. As can be seen in Fig 3, ortho-phenanthroline (OP) strongly inhibits the binding reaction when present at mM concentrations.

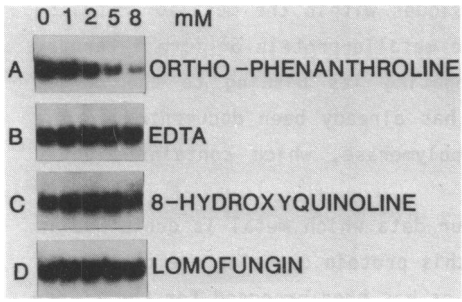

FIGURE 3:
Effect of metal chelators on NF-Y. Standard reactions were set up with the E_α 39mer, poly[d(I-C)], and M12 nuclear extract, except that the binding buffer was supplemented with 0-8 mM ortho-phenanthroline (A), EDTA (B), 8-hydroxyquinoline (C), or lomofungin (D). The reaction mix was electrophoresed and bands were revealed by autoradiography. Only the bands representing the NF-Y/E_α-oligo complex are shown.

We found the same degree of inhibition when OP was added to the reaction mixture at various times (up to 30 min) before the oligo probe, and when it was added last. In contrast, the other metal chelators EDTA, 8-hydroxy-quinoline and lomofungin (16) do not affect NF-Y binding in the same concentration range (Figure 3).

OP is an efficient chelator of Zn^{2+}, Mn^{2+}, Cu^{2+} and $Fe^{2,3+}$ ions, and has been frequently used at mM concentrations to demonstrate the presence of Zn^{2+} in the catalytic center of Zn metallo-enzymes, such as reverse transcriptase, terminal deoxynucleotidyltransferase, DNA and RNA polymera-ses, various aminoacyl-tRNA transferases and elongation factor I (see 23 and references therein). We established that the metal-binding properties of OP are really responsible for its inhibitory effect on NF-Y by conduct-ing the binding reaction in the simultanuous presence of OP and a 1.5-fold molar excess of Zn^{2+} ; under these conditions the OP inhibition is partially relieved (not shown). Because metal/chelator complexes can exhibit nuclease activity (24), we evaluated the integrity of the oligonu-cleotide probe after incubation in OP-containing buffer by running it on a denaturing polyacrylamide gel ; no deleterious effect was observed (not shown).

Since the other chelators used in this study have affinities for metal ions that overlap those of OP (25,26), we do not believe that OP simply acts by complexing free metal ions required for the binding reaction. Rather, we hypothesize that OP by virtue of its specific chemical structure is able to interact with metal ions as they are complexed internally within NF-Y. Perhaps the unique configuration of heterocyclic nitrogen atoms in OP, as opposed to EDTA, lomofungin and 8-hydroxyquinoline which chelate by

means of hydroxyl groups, enables the former to interact with metal ions complexed by cysteine and histidine residues within the metallo-proteins. OP may either leach the ions out of the metallo-protein or form a ternary complex with it, in either case abrogating its binding to the target sequence. The mechanism we hypothesize has already been documented in the case of OP inhibition of E.coli RNA-polymerase, which contains tightly associated Zn^{2+} ions (27).

Although we cannot conclude from our data which metal is contained in NF-Y, it is tempting to speculate that this protein contains one or several Zn^{2+}-complexing, DNA-binding "fingers" as has been proposed for the transcription factor TF IIIA (28), which is also OP-sensitive (29). The observation that metal-binding motifs are found in an increasing number of factors involved in transcriptional control [eg proteins involved in developmental (30) and hormone-regulated events (31)], suggests that "finger" proteins may be widespread amongst sequence-specific DNA-binding proteins.

Protease sensitivity - evidence for a DNA-binding domain

Many sequence-specific DNA-binding proteins have a structurally distinct DNA-binding domain. These include prokaryotic (32-35), yeast (36-38), and higher eukaryotic (39-42) transcription factors. The DNA-binding domain was traditionally identified as a protease-resistant "core" polypeptide, but more recently, it has been demonstrated by deletion mutagenesis and X-ray crystallography experiments.

In order to determine whether NF-Y has a protease-resistant DNA-binding domain, we incubated the E_α 39mer with poly [d(I-C)] and crude nuclear extract under standard conditions, treated the binding mixture with increasing amounts of protease and ran the digested material on a polyacrylamide gel. As illustrated in figure 4A, treatment of the NF-Y/E_α-oligo complex with progressively higher concentrations of proteinase K reveals a resistant DNA-bound "core". The transition from the intact complex to the more rapidly migrating form is quite abrupt, occuring over a three- to ten-fold concentration range. The limit digest is remarkably stable, persisting over a wide range of proteinase K concentrations.

The similar intensity of the bands from digested and undigested material suggests that the resistant "core" does actually derive from the NF-Y/E_α-oligo complex. To establish this point, the proteolysis experiment was conducted using methylated DNA. After electrophoresis of the binding mixture, the DNAs in the NF-Y band and in the protease-resistent "core" band were extracted, cleaved and displayed on a sequencing gel. The methy-

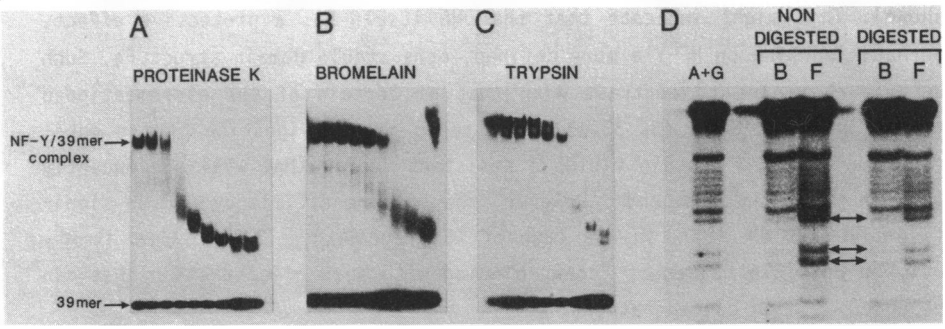

FIGURE 4: **A-C. Protease sensitivity of the NF-Y/E_α-oligo complex.** Reactions were set up as usual with the E_α 39mer, 400ng poly [d(I-C)] and LMTK nuclear extract. The reaction mix was incubated at 0°C for 20 min and then proteinase K (A), bromelain (B), or trypsin (C) was added. Proteinase K and trypsin were at 0, 0.3, 1, 3, 10, 30, 100, 300 or 1000ng ; bromelain was at 0, 0.1, 0.3, 1, 5, 10, 30, 100 or 300ng.

D. Methylation interference mapping of the intact NF-Y/E_α-oligo complex and the protease-resistant "core" complex. The coding strand of the E_α 39mer was 5' end-labelled and annealed with its complement. The double-stranded oligo was partially methylated and then introduced into the standard binding reaction along with poly[d(I-C)] and M12 nuclear extract. Half the material was further incubated with 32ng of proteinase K ; the other half was left undigested. The reaction mixtures were run on a polyacrylamide gel. Free DNA, DNA from the intact NF-Y/E_α-oligo complex, and DNA from the protease-resistant "core" complex was extracted, concentrated, cleaved by piperidine and run on a sequencing gel. B = protein-bound DNA ; F = free DNA ; the "A+G" lane shows the cleavage pattern of the E_α 39mer treated with piperidine (A+G specific cleavage). The arrows indicate bases whose methylation inhibits NF-Y binding.

lation interference patterns shown in Figure 4D are identical, indicating that the lower band is indeed derived from the upper.

When the binding mixture is treated with bromelain or subtilisin, other endopeptidases of broad specificity, a limit digest similar to that seen with proteinase K can be observed (Figure 4B and data not shown). Trypsin, a serine protease of narrow specificity, also generates the resistant "core" complex (Figure 4C). Digestion of the NF-Y/E_α-oligo complex with trypsin at intermediate concentrations gives a clearer picture of the partial digestion products than does treatment with the other three proteases.

If the nuclear extract is treated with protease before conducting the binding reaction, we observe no complex formation on the E_α/39mer (data not

shown). This might indicate that the DNA itself has a protective effect, perhaps imposing on NF-Y a more defined, more stable domain structure. Such behaviour appears to contrast with that of certain of the aforementioned transcription factors (32,33,41) ; in these cases, digestion of the purified, uncomplexed protein yields a resistant "core" that will subsequently bind to the target sequence. However, the binding efficiency can be significently reduced (100x in the case of λC1 repressor - 32) so that it may only be possible to detect "core" binding with a purified protein that can be added at high concentrations.

The very similar patterns obtained with the different proteases (Figure 4A-C) suggest that there is a protease "hot spot(s)" somewhere on NF-Y. This might correspond to a flexible "hinge" between the domain involved in DNA-binding and another perhaps responsible for protein:protein interactions. Such a structure is known to exist in certain prokaryotic transcription factors, eg the λC1 repressor and lex A (32,33). It has been hypothesized that the "hinge" plays an important role in the regulation of transcription by these two factors, in particular as a target for proteases in the SOS reaction (43). Alternatively, NF-Y could consist of several different polypeptides; the limit digest would then represent the protease resistant DNA-binding chain.

Molecular Weight Determination

Three methods have been employed to estimate the molecular weight of NF-Y : pore-gradient gel electrophoresis, Ferguson plot analysis and glycerol gradient centrifugation.

- Pore-gradient gel electrophoresis.

This technique involves electrophoresing the proteins on non-denaturing polyacrylamide gels of graded porosity ; each protein migrates until it reaches its "pore size limit", which is size-dependent (44). Such a method has been used successfully to predict the molecular weight of TUF, a yeast transcription factor (37).

The NF-Y/E_α-oligo complex was formed under standard conditions, and then loaded onto a polyacrylamide gradient gel. Since preliminary experiments indicated that the complex is probably very large (making this approach especially suitable - 44) we chose a linear 4-10% polyacrylamide gel in 0.5 x TBE. After fixation, staining and drying of the gel the position of the complex was located by autoradiography (Figure 5A, lane 1). By comparing its migration with the migration of proteins of known size, including the multimeric BSA "ladder" which forms spontaneously in TBE (44), the molecular weight of the NF-Y/E_α -oligo complex was estimated as

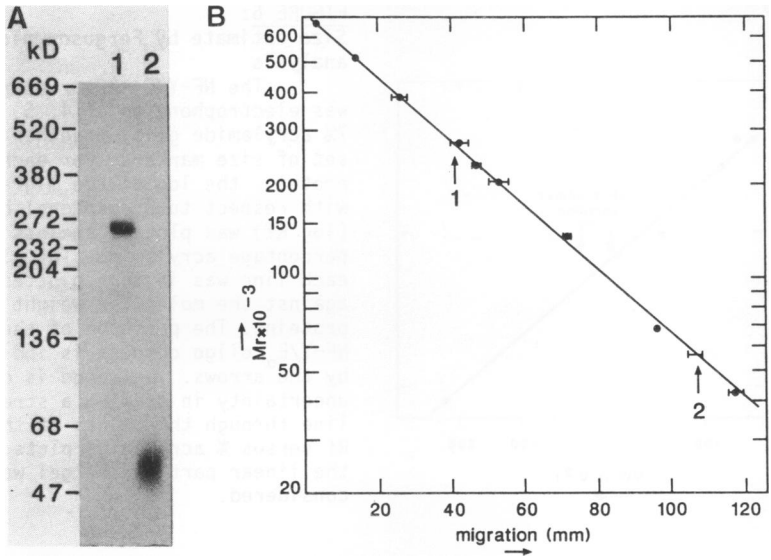

FIGURE 5: **Size estimate by pore-gradient gel electrophoresis**
A. The polyacrylamide gradient gel. (1) A binding reaction was set up as usual with the ^{32}P-labelled E_α 39mer, poly[d(I-C)], and M12 nuclear extract. After incubation, the mix was run on a 4-10% polyacrylamide gel. (2) The same, except that before loading the gel, the reaction mix was treated with 300ng proteinase K. To the right of the gel are indicated the positions of marker proteins (see Materials and Methods for details) run on the same gel.
B. Plot of log molecular weight versus migration. The line derives from plotting values for the marker proteins indicated in panel A. (1) The intact NF-Y/E_α-oligo complex. (2) The protease-resistant core.

about 270 kD (Figure 5). The contribution of the labelled E_α 39mer is probably minor because the complex formed with an E_α 22mer migrates only slightly faster.

- Ferguson plot analysis.

A method purported to surmount the influence of charge on the migration of proteins is Ferguson plot analysis. Proteins are electrophoresed on several gels of differing polyacrylamide concentration. For a given protein, the logarithm of relative electrophoretic mobility (log Rf.) is plotted against the percentage acrylamide ; the slope of the resulting line is size-dependent (45). NF-Y was bound to the labelled E_α 39mer under standard conditions. The reaction mix and a set of marker proteins were electrophoresed on gels containing 4, 5, 6 and 7% acrylamide. For each marker protein, the log Rf was plotted against the acrylamide concentration ; the slope of each line was then plotted against the marker protein's

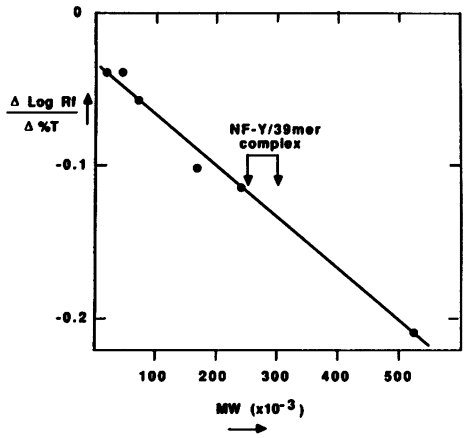

FIGURE 6:
Size estimate by Ferguson plot analysis
The NF-Y/E_α 39mer complex was electrophoresed on 4, 5, 6 and 7% acrylamide gels, together with a set of size markers. For each protein, the log of the migration with respect to bromophenol blue (log Rf) was plotted against the percentage acrylamide. The slope of each line was in turn plotted against the molecular weight of the proteins. The position of the NF-Y/E_α-oligo complex is indicated by the arrows. The range is due to uncertainty in drawing a straight line through the points of the log Rf versus % acrylamide plots. Only the linear part of the gel was considered.

molecular weight (Figure 6). The position of the point representing the NF-Y/E_α-oligo complex on this standard curve prompts us to estimate its molecular weight as 250-300 kD. Again, the E_α 39mer should make only a minor contribution to this value.

- Glycerol gradient centrifugation.

Finally, the molecular weight of NF-Y was estimated from its sedimentation rate through a glycerol gradient. ^{35}S-methionine labelled nuclear extract was concentrated and then loaded onto a 10-30 % gradient. Fractions were collected and an aliquot of each was assayed for NF-Y activity using the standard gel retardation assay. Another aliquot was loaded onto an SDS-polyacrylamide gel to give some idea of the range of molecular weights included in each fraction. A set of marker proteins was run on a parallel gradient and aliquots of fractions from this gradient were also electrophoresed on an SDS-polyacrylamide gel. As figure 7A illustrates, NF-Y was located primarily in fractions 12-15, peaking in fraction 13. The 43 kD molecular weight marker (ovalbumin) peaked in fractions 15 and 16 and the 94 kD marker (phosphorylase b) in fractions 8 and 9, suggesting that NF-Y is in the 50-70 kD size range. The profile of ^{35}S-labelled proteins from fractions 12-15 is consistent with this value (Figure 7B).

This result seems to contradict the molecular weight estimates derived from the other two methods, both of which give a value >250 kD. There are several possible explanations for this discrepancy :

1) An assumption made in all the size determinations is that NF-Y has

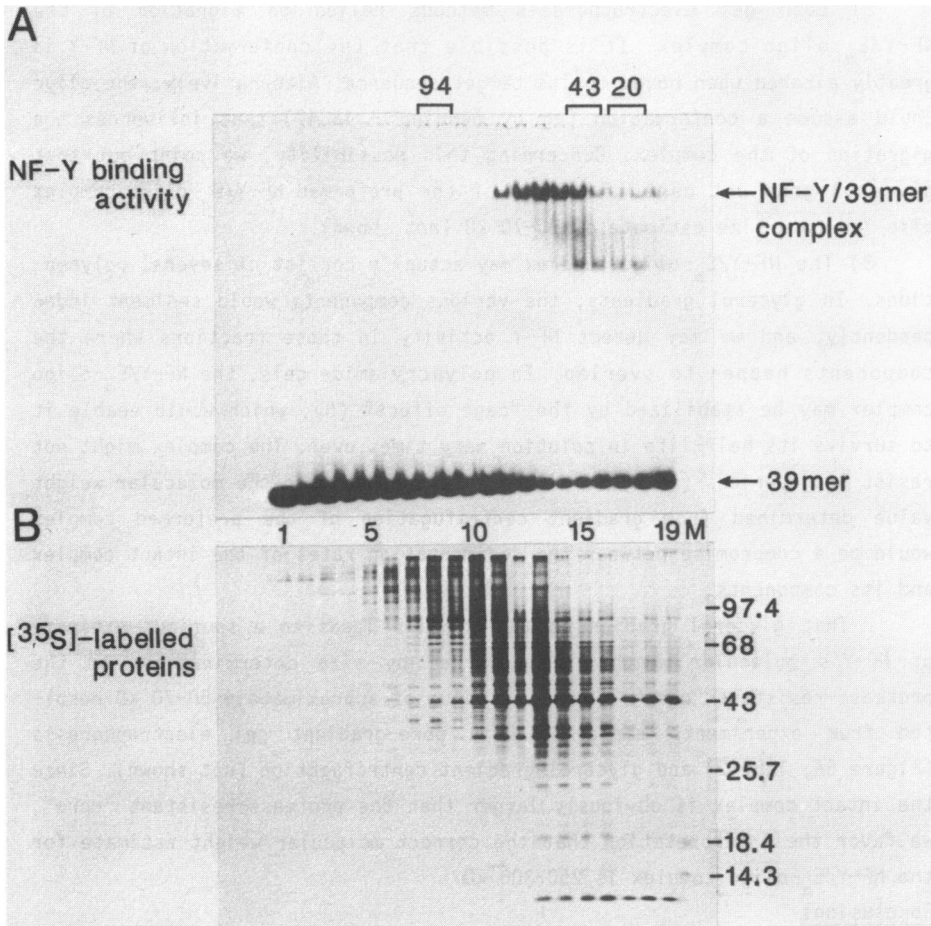

FIGURE 7 : **Size estimate by glycerol gradient centrifugation**
An ^{35}S-methionine labelled nuclear extract from LMTK cells was centrifuged on a glycerol gradient. 19 fractions were collected and aliquots of each (A) tested for NF-Y activity in a standard gel retardation assay using the E_{α} 39mer or (B) run on an SDS-polyacrylamide gel. Above panel A are indicated the locations of size markers run on a parallel glycerol gradient (see Materials and Methods for details). Lane M of panel B shows the positions of ^{14}C-labelled size markers (BRL) ; adjacent to lane M are the relevant molecular weights.

the same conformation as the proteins used as size markers (which are globular) ; any deviation from this ideal could influence differentially the values derived using the different methods. We do not favor this explanation because of the magnitude of the discrepancy.

2) Both gel electrophoresis methods relied on migration of the NF-Y/E_α oligo complex. It is possible that the conformation of NF-Y is greatly altered when bound to its target sequence. Alternatively, the oligo could assume a conformation (eg by bending - 46,47) that influences the migration of the complex. Concerning this possibility, we point out that glycerol gradient centrifugation of the preformed NF-Y/E_α-oligo complex also led to a size estimate of 50-70 kD (not shown).

3) The NF-Y/E_α-oligo complex may actually consist of several polypeptides. In glycerol gradients, the various components would sediment independently, and we may detect NF-Y activity in those fractions where the components happen to overlap. In polyacrylamide gels, the NF-Y/E_α-oligo complex may be stabilized by the "cage effect" (5), which would enable it to survive its half-life in solution many times over. The complex might not resist long periods (16 h) in dilute solution, so that the molecular weight value determined from gradient centrifugation of the preformed complex would be a compromise between the sedimentation rates of the intact complex and its components.

That glycerol gradient centrifugation leads to a spurious estimate of NF-Y's molecular weight is supported by size determinations on the protease-resistant "core" complex. A value of approximately 50-70 kD resulted from experiments employing both pore-gradient gel electrophoresis (Figure 5A, lane 2) and glycerol gradient centrifugation (not shown). Since the intact complex is obviously larger than the protease-resistant "core", we favor the interpretation that the correct molecular weight estimate for the NF-Y/E_α-oligo complex is 250-300 kD.

Conclusions

We have further characterized the NF-Y protein in order to provide a basis for comparing it with previously described and to-be-described DNA-binding proteins. The most striking properties revealed by this set of experiments are that :

- NF-Y has a ubiquitous cell-type distribution.
- NF-Y is probably a metallo-protein.
- NF-Y has a protease-resistant DNA-binding domain.
- The NF-Y/E_α-oligo complex seems very large, and may be composed of multiple proteins.

These properties have provided new insights into the structure of NF-Y. In addition, they should prove useful in comparative studies and in attempts to verify the identify of purified preparations of NF-Y.

ACKNOWLEDGEMENTS

We would like to thank Denis Duboule for cell lines, Adrian Staub and Frank Ruffenach for oligonucleotides, and Philippe Gerber for expert technical assistance. This work was supported by institutional grants from the INSERM and CNRS and by a grant to D.M. and C.B. from the Association pour la Recherche Contre le Cancer (A.R.C.). R.H. received postdoctoral fellowships from the A.R.C. and EMBO, and A.D. from the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Serfling, E., Jasin, M., and Schaffner, W. (1985). Trends in Genetics 1, 224-230.
2. Voss. S.D., Schlokat, U., and Gruss, P. (1986). TIBS 287-290.
3. Dynan, W.S., and Tjian, R. (1985). Nature 316, 774-778.
4. Maniatis, T., Goodbourn, S., Fischer, J.A. (1987). Science 236, 1237-1245.
5. Galas, D.J., and Schmitz, A. (1978). Nucleic Acids Res. 5, 3157-3168.
6. Fried, M., and Crothers, D.M. (1981). Nucleic Acids Res. 9, 6505-6525
7. Davidson, I., Fromental, C., Augereau, P., Wildeman, A., Zenke, M., and Chambon, P. (1986). Nature 323, 544-548.
8. Pfeifer, K., Arcangioli, B., and Guarente, L. (1987). Cell 49, 9-18.
9. Dorn, A., Bollekens, J., Staub, A. Benoist, C., and Mathis, D. (1987). Cell, in press.
10. Bohmann, D., Keller, W., Dale, T., Schöler, H.R., Tebb, G., and Mattaj, I.W. (1987). Nature 325, 268-272.
11. Kadonaga, J.T., Jones, K.A., and Tjian, R. (1986). TIBS 11, 20-24.
12. Pfeifer, K., Prezant, T., and Guarente, L. (1987). Cell 49, 19-27.
13. Johnson, P.F., Landschulz, W.H., Graves, B.J., and McKnight, S.L. (1987). Genes and Development 1, 133-146.
14. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987). Cell 48, 79-89.
15. Dorn, A., Durand, B., Marfing, C., Le Meur, M., Benoist, C., Mathis, D. (1987). Proc. Natl. Acad. Sci. USA, in press.
16. Falchuk, K.H., Mazus, B., Ulpino, L., and Vallee, B.L. (1976). Biochemistry 15, 4468-4475.
17. Hooft van Huijsduijnen, R.A.M., Alblas, S.W., De Rijk, R.H., and Bol, J.F. (1986). J. Gen. Virol. 67, 2135-2143.
18. Dorn, A., Brauer, D., Otto, B., Fanning, E., and Knippers, R. (1982). Eur. J. Biochem. 128, 53-62.
19. Laemmli, U.K. (1970). Nature 227, 680-685.
20. Flavell, R.A., Allen, H., Huber, B., Wake, C., and Widera, G. (1985). Immunol. Rev. 84, 29-50.
21. Duboule, D., Baron, A., Mähl, P., and Galliot, B. (1986). EMBO J. 5, 1973-1980.
22. Berg. J.M. (1986). Science 232, 485-487.
23. Posorske, L.H., Cohn, M., Yanagisawa, N., and Auld, D.S. (1979). Bioch. Biophys. Acta 576, 128-133.
24. Hertzberg, R.P., and Dervan, P.B. (1982). J. Am. Chem. Soc. 104, 313-315.
25. Sillen, L.G., and Martell, A.E. (1964). The Chemical Society, London, Special Publ. 17, 664-677.
26. Pavletich, K., Kuo, S-C., and Lampen, J.O. (1974). Bioch. Biophys. Res. Comm. 60, 942-950.
27. Scrutton, M.C., Wu, C.W., and Goldthwait, D.A. (1971). Proc. Natl. Acad. Sci. USA 68, 2497-2501.

28. Miller, J., McLachlan, A.D., and Klug, A. (1985). EMBO J. 4, 1609-1614.
29. Hanas, J.S., Hazuda, D.J., Bogenhagen, D.F., Wu, F.Y.H., Wu, C.W. (1983). J. Biol. Chem. 258, 14120-14125.
30. Schuh, R., Aicher, W., Gaul, U., Coté, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schröder, C., Kemler, R., and Jäckle, H. (1986). Cell 47, 1025-1032.
31. Green, S., and Chambon, P. (1987). Nature 325, 75-78.
32. Sauer, R.T., Pabo, C.O., Meyer, B.J., Ptashne, M., and Backman, K. (1979). Nature 279, 396-400.
33. Little, J.W., and Hill, S.A. (1985). Proc. Natl. Acad. Sci. USA 82, 2301-2305.
34. Anderson, J.E., Ptashne, M., and Harrison, S.C. (1987). Nature 326, 846-852.
35. Zhang, R-G., Joachimiak, A., Lawson, C.L., Schevitz, R.W., Otwinowski, Z., and Sigler, P.B. (1987). Nature 327, 591-597.
36. Keegan, L., Gill, G., Ptashne, M. (1986). Science 231, 699-704.
37. Huet, J., and Sentenac, A. (1987). Proc. Natl. Acad. Sci. USA 84, 3648-3652.
38. Marzouki, N., Camier, S., Ruet, A., Moenne, A., and Sentenac, A. (1986). Nature 323, 176-178.
39. Smith, D.R., Jackson, I.J., and Brown D.D. (1984). Cell, 37, 645-652.
40. Miller, J., McLachlan, A.D., and Klug, A. (1985). EMBO J. 4, 1605-1614.
41. Briggs, M.R., Kadonaga, J.T., Bell, S.P., Tjian, R. (1986). Science 243, 47-52.
42. Green, S., and Chambon, P. (1986). Nature 324, 615-617.
43. Little, J.W., and Mount, D.W. (1982). Cell 29, 11-22.
44. Andersson, L-O., Borg, H., and Mikaelsson, M. (1972). FEBS 20, 199-202.
45. Gel Electrophoresis of Proteins: A practical approach (1981). Eds (Hames, B.D., Rickwood, D.) IRL Press, 12-18.
46. Marini, J.C., Levene, S.D., Crothers, D.M., and Englund, P.T. (1982). Proc. Natl. Acad. Sci. USA 79, 7664-7668.
47. Koo, H.S., Wu, H.M., and Crothers, D.M. (1986). Nature 320, 501-505.