A quantitative analysis of nuclear factor I/DNA interactions

Michael Meisterernst, Irene Gander, Lars Rogge and Ernst-L. Winnacker

Institut für Biochemie, Universität München, Karlstrasse 23, D-8000 München 2, FRG

Received February 12, 1988; Revised and Accepted April 12, 1988

ABSTRACT

Nuclear factor I (NFI) was purified to homogeneity from porcine liver by DNA-affinity chromatography and displays a single band with a molecular weight of 36 kDa in SDS-polyacrylamide gels. The purified protein was used to determine absolute equilibrium binding constants by gel retardation techniques for a variety of DNA fragments with genuine or mutated NFI binding sites and a number of DNA fragments derived from various eukaryotic promoters carrying the CCAAT-box as a half-site for NFI binding. We present a model which allows prediction of the functional significance of mutated NFI binding-sites from sequence data. The data suggest that the single molecular species of NFI from porcine liver may not be able to recognize and activate the -CCAAT- promoter element *in vivo* without additional interactions, *e.g.* with other proteins.

INTRODUCTION

Nuclear Factor I (NFI) is a sequence-specific DNA binding protein which stimulates the initiation of adenovirus DNA replication *in vitro*. It was originally detected in Hela cell nuclear extracts (1) and subsequently shown (2) to mediate its effect *via* a small (16 bp) recognition sequence located within the inverted terminal repetition (ITR) at the termini of the linear adenovirus type 5 DNA genome (Fig. 1). This sequence is also present in promoter and replication origins of other viruses (3, 4) and in cellular, chromosomal DNAs (5-7). The consensus sequence TGG (A/C) N₅ GCCAA was derived from a comparison of known NFI recognition sites, from mutagenesis studies (2, 7, 8, 9, 10) and from an analysis of randomly synthesized binding sites (11, 12). A portion of this sequence matches the sequence -CCAAT- of a eukaryotic promoter element which is required for transcription of a variety of cellular and viral genes (13, 14).

NFI has been purified from Hela cells and from porcine liver by a combination of conventional purification steps as well as by recognition site affinity chromatography (2, 10, 15, 16) or preparative gel retardation (17). Similar purification protocols have been described for the purification of nuclear proteins which bind selectively to the sequence CCAAT (18, 19, 20). While human HeLa cells (18) contain a group of polypeptides, designated CTF (CCAAT binding transcription factors), with molecular weights between 52-66 kDa, the nuclear protein from rat liver (19) is a heat-stable 20 kDa protein termed EBP20 (enhancer binding protein 20). It is

1 20 40 60
 .
 CATCATCAATAATATACAGTTAGCAAAAAATGGCGCCTTTGT**TTGGCTTTGTCCA**ACTGT
 80 100 120
 TTTTGGCCCGAGTTGGGTTTCGTTTTCCCGGAATGACGTGTGAAAAGGGTCTGGGGCT
 140
TTTGGCACTGTGCCAACTGTGTTGTG-----

1 20 40 60
CATCATCAATAATATACCTTTATTTGGATTGAAGCCAAATATGATAATGAGGGGGTGGAGT
80 100
TTGTACGTGGCGCGGGGCGTGGGAACGGGGCGGGTGACGTAG-----

thought to be identical to a rat liver nuclear protein termed CBP (a CCAAT-binding protein) described previously (21).

In order to resolve these discrepancies we undertook to purify NFI from porcine liver to homogeneity and analyzed quantitatively its binding properties *in vitro* to DNAs carrying typical NFI and CCAAT-binding sites. Since porcine liver contains comparatively high concentrations of NFI as a single polypeptide, this source is an excellent model system for a quantitative study of eukaryotic protein/DNA interactions.

Buffers

Buffers used in the experiments described in this communication are composed as follows: **Buffer A:** 25 mM Hepes/KOH, pH 7.5, 1 mM EDTA, 1mM DTT, 100 mM NaCl and 10% glycerol. **Buffer B:** 25 mM Hepes/KOH, pH 7.5, 150 mM NaCl, 1mM EDTA, 1mM DTT and 10 % glycerol. **Buffer C:** 25 mM Hepes/KOH, pH 7.5, 5mM EDTA, 1mM DTT, 100 mM NaCl

	Oligonucleotide	K [M ⁻¹]
	<div>31</div> <div>↓</div>	
A1/B1	-CCTTATTT TGG ATTGAAG GCCAA TATGATAATGAGG- -GGAATAAAA ACCTAACTTCGGTT AATACTATTACTCC-	2 × 10 ¹⁰
L1/2	--GGGCTttt tg gCACTGT gccaa CTGTGtTgtga--- --CCCGaaa acc GTGAC acggtt GACACaACact---	n.d.
E1/2	-GGGGCTttt tg gCACTGT gccaa CTGTGtTgtgagg- -CCCCGaaa acc GTGAC acggtt GACACaACactcc-	2 × 10 ¹¹
EW1/2	-ccttTttt tg gattgaag gccaa tatgataatgagg- -ggaaAaaa accta actt cgg ttatactattactcc-	4 × 10 ¹⁰
TT1/2	-ccttatttt tg gaGCgCG gccaa tatgataatgagg- -ggaataaaa acct CGcGC cgg ttatactattactcc-	2 × 10 ¹⁰
G1/2	-ccttatttt tg gattgaag Gcaa tatgataatgagg- -ggaataaaa accta actt cG ttatactattactcc-	2 × 10 ⁸
T1/2	-ccttatttt tg gattgaag ccT atgataatgagg- -ggaataaaa accta actt cggA tatactattactcc-	1 × 10 ⁹
GT1/2	-ccttatttt tg gattgaag GcT atgataatgagg- -ggaataaaa accta actt cG Atatactattactcc-	1 × 10 ⁸
G3/4	-ccttatttt tg Cattgaag gccaa tatgataatgagg- -ggaataaaa acG taactt cgg ttatactattactcc-	6 × 10 ⁸
IV1/2	-ccttatttt tg gattgaag AACC tatgataatgagg- -ggaataaaa accta actt cTTGG atactattactcc-	1 × 10 ⁷

Table I: DNA sequences of synthetic DNA fragments used in this communication. Bold-face letters represent the NFI consensus sequence. Capital letters indicate sequence differences from the standard Ad5 sequence which is displayed in fragment A1/B1. Lower case letters are bases identical to A1/B1. Fragment L1/2 was not used in binding studies but only in the preparation of the DNA-affinity column according to (22).

or 100 mM KCl, 20 % glycerol, and 0.001 % NP-40. 0.5 M STE: 0.5 M NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0. Occasionally, 1 mM PMSF (Boehringer, Mannheim) or 10 mM sodium-metabisulfite were added to the buffers.

Preparation of nuclear extracts

Nuclei were prepared from three porcine livers as described previously (10). About 2 × 10¹¹ nuclei from a typical preparation were extracted at 4 °C in the presence of 200 mM NaCl, 25 mM Hepes/KOH, pH 7.5, 5 mM DTT, 4 mM CaCl₂, and 200 mM sucrose in a total volume of 300 ml. The resulting extract was dialyzed against buffer A for 15 hours at 4 °C. This extract was used immediately in further purification steps.

Purification of nuclear factor I

(i) Assay, DEAE-Cellulose and MonoQ Chromatography

The activity of NFI in extracts and column eluates during the purification was monitored by gel retention assays (10) with ^{32}P -labelled synthetic DNA fragments (39 bp) carrying an NFI-binding site from adenovirus type 5 (Table I, fragment A1/B1).

Dialyzed crude extracts were applied to a DEAE-cellulose column equilibrated with buffer A. The flow-through fraction was subjected to an FPLC preparative MonoQ column (HR 16/10, Pharmacia) equilibrated with buffer A. Proteins were eluted with a 120 ml gradient of NaCl in buffer A (from 100 to 1000 mM NaCl) in 2-ml fractions. Active fractions which eluted between 150 and 300 mM were pooled and adjusted to 300 mM NaCl/0.001% NP40.

(ii) DNA-affinity chromatography

A 32 bp DNA fragment (fragment L1/2 in Table I) carrying the symmetrical NFI recognition sequence from mouse adenovirus FL was linked to Sepharose as described elsewhere (22). Protein extracts from DEAE-cellulose chromatography or NFI-enriched preparations from MonoQ-columns were loaded onto 5 ml of a DNA-affinity column in buffer C, containing 300 mM NaCl and 0.001% NP40. In a salt gradient established between 300 mM and 3 M potassium chloride, NFI eluted between 1.5 and 2.5 M potassium chloride. Active fractions were dialyzed against buffer C, 150 mM NaCl and could be stored at 4 °C without significant loss of NFI activity for several months. However, all studies described in this communication were performed with freshly prepared material. The entire purification procedure, from slaughter house to affinity-purified NFI, takes five days. The enrichment was between 25.000 to 50.000-fold.

Preparation of oligonucleotides

Oligonucleotides were synthesized chemically on an automatic Applied Biosystems Synthesizer Type 300. They were purified by gel electrophoresis under denaturing conditions (8 M urea/15 % PAA), eluted with water, and precipitated twice as described (10). Concentrations were determined by UV spectroscopy. Equimolar amounts of two complementary oligonucleotides were mixed, heated to 37 °C, cooled down to 20 °C for 1 hour in 10 mM Tris/0,5 mM EDTA, pH 7,4 and end-labelled either by the polynucleotide kinase or the Klenow reaction, following the protocols provided by the manufacturers (Boehringer, Mannheim; New England Biolabs). Oligonucleotides were purified after the labelling reaction on a DE-52 ion exchange column. After washing with 5 x 0.5 ml 0.2 M STE, pH 8.0 the material was eluted with 3 x 0,33 ml 0.5 M STE, pH 8.0. Yields in this final purification step usually exceeded 97 %.

In order to measure unspecific binding of NFI, a 38 nucleotides long oligonucleotide was synthesized chemically which consisted of 26 random deoxyribonucleotides, followed by 12 bases derived from positions 39 to 51 from the adenovirus type 5 inverted terminal repetition (see Fig. 1 for positions). A 16 base long primer was synthesized complementary to these 12 bases with four additional bases as overhanging terminus. A two-fold excess of the primer was hybridized in 100 mM Tris HCl pH 7.4 against the template oligonucleotide at 65 °C and cooled

Figure 2: Gel retention assay of NFI with a 146 bp long DNA fragment (F 146) carrying an NFI binding site (Ad5 ori). The assay was performed at constant DNA concentration (0.1 fmole/ μ l) in the presence of increasing concentrations of NFI (0.02, 0.1 and 1.60 fmole per μ l in lanes 1, 2 and 3, respectively).

down during 30 min to 16 °C and stored at -20 °C. The primer extension reaction was performed in the presence of Klenow polymerase and 32 P-labelled dCTP under conditions described by the manufacturer (Boehringer, Mannheim).

SDS gel electrophoresis

SDS polyacrylamide gels were performed as described previously (34) and analyzed by silver staining (35).

Gel retention assay

NFI preparations were incubated with 32 P-labelled DNA fragments (synthetic or otherwise) for 15 min at RT in buffer B. An aliquot was loaded onto an 11 % polyacrylamide gel prepared in 0.375 M Tris-HCl, pH 8.8. The running buffer was Tris-glycine (0.05 M/0.384 M). After electrophoresis for about seven hours at 50 V/15 mA, the gels were fixed in 40 % MeOH/10 % acetic acid for 60 minutes and dried under vacuum. Gels were analyzed by autoradiography and/or scintillation counting on a Berthold Scanner or by scintillation counting of cut-out bands. Fixation did not influence the amount of radioactivity by more than 10%. Complex formation

Figure 3: Scatchard plot of NFI/DNA interactions. The data were obtained from the experiment displayed in Fig. 7 and were analyzed according to equation 3.

between DNA and NFI was not influenced by glycerol at concentrations of up to 30%, NP-40 of up to 0.01% or Mg^{2+} , EDTA and DTT in concentrations of up to 10 mM.

Determination of equilibrium binding constants

NFI from porcine liver reacts with DNA carrying an NFI binding site from the Ad5 ITR (inverted terminal repetition) in a bimolecular reaction. This conclusion is based upon the observation that even in the presence of NFI excess only a single molecular species of a DNA/protein complex is formed in appropriate gel retention assays (Fig. 2). Thus, the equilibrium binding constant K is given by equation (1):

$$K = \frac{[c]}{[DNA][NFI]} \quad (1)$$

which can be rearranged to (2),

$$K = \frac{[c]}{([DNA]^0 - [c])([NFI]^0 - [c])} \quad (2)$$

where $[c]$ symbolizes the complex concentration at equilibrium and the subscript 0 indicates initial concentrations. The initial concentration of DNA, $[DNA]^0$, can be measured spectrophotometrically. $[NFI]^0$ is measured in the presence of a 10 to 100-fold excess of DNA and thus determined from the amount of DNA shifted into the complex position in a gel retention assay (see Fig. 7, lane 8). The value for $[DNA]^0$ must be corrected for the amount of "active" DNA present in a given DNA preparation. This fraction is 100% when restriction enzyme fragments are used and often is less than 70 % for synthetic DNA (see text). The concentration $[c]$ of the DNA/protein complex at equilibrium can be determined from the amount of radioactivity in the shifted position and the known specific activity of the DNA. Under conditions of low binding constants ($< 10^7 M^{-1}$) we observed single and clearly identifiable bands with short oligonucleotide fragments and multiple bands when larger DNA fragments were used (e.g. Fig. 9). There is no detectable dissociation of the protein/DNA complexes down to binding constants of $10^7 M^{-1}$.

For a Scatchard plot analysis, equation (1) has to be rearranged into (3):

$$\frac{[c]}{[DNA]} = K [NFI]^0 - K [c] \quad (3)$$

The plot of $[c]/[DNA]$ versus $[c]$ resulted in a straight line (Fig. 3) and thus a single value for the slope K , when proper corrections were made for the "active" concentration of DNA (see above). There were no differences between K values obtained from the slope of the Scatchard plot or K values calculated directly from equation (2).

Figure 4: SDS polyacrylamide gel electrophoresis of purified NFI. The 10 % gel was stained with silver. Lane A displays a mixture of molecular weight marker proteins (numbers are kDa), as indicated; lane B represents a sample of ~100 ng of DNA-affinity chromatography purified NFI.

Practical difficulties arise for the selection of proper amounts of DNA and protein in the case of very high ($K > 10^{11} \text{ M}^{-1}$) or very low K values ($K < 10^7 \text{ M}^{-1}$). The former case requires high specific activities of the DNA (or long exposure times), the latter large amounts of protein.

K can also be determined accurately from equation (4) in cases when $[\text{DNA}]^0 \ll [\text{NFI}]^0$.

$$K [\text{NFI}]^0 = \frac{[c]}{[\text{DNA}]} \quad (4)$$

This is particularly true for low binding constants ($K < 10^8 \text{ M}^{-1}$). Thus, equation (4) proved not only valuable but absolutely necessary for determinations of K for the binding of NFI to unspecific DNA, since in our case of a random sequence oligonucleotide fragment, the value of $[\text{DNA}]^0$ (and thus of the specific activity of the DNA) could not be determined easily due to the variable yields of the Klenow reaction.

K values obtained from gel retention analyses were in good agreement with data derived from nitrocellulose filter binding studies. However, apart from the many advantages discussed below (RESULTS), the gel retention method is at least 10 times more accurate than the filter binding method.

DNase footprinting.

Protein fractions were incubated with about 10 fmol of DNA as described under Gel retention assay. After 15 minutes, the reaction mixture was diluted 1:5 with a solution containing 25 mM Hepes/KOH pH 7.5, 5 mM CaCl_2 and 10 mM MgCl_2 . DNaseI dissolved in the same

Table II: Purification of NFI from porcine liver

Procedure	Total Protein [mg]	Total Activity [nmol]	Specific Activity [pmol/mg]	Purification factor	Yield [%]
Porcine liver (1.5kg)					
Nuclear extract	9800	5,7	0,6	1	100
DEAE cellulose	2600	4,3	1,7	3	75
FPLC Mono Q	296	3,2	11	21	56
DNA affinity column	0,1	2,8	28.000	52.500	49

buffer was added immediately to a final concentration of 0.02 to 0.05 ng/ μ l. The mixture was then incubated for one minute. The reaction was terminated by the addition of hot phenol (65 °C) saturated with 25 mM EDTA. DNA was purified and analyzed as described (18).

In vitro replication assay.

The *in vitro* initiation reaction was performed as described previously (23) with modifications reported (2).

Nitrocellulose filter binding assay

Nitrocellulose-filter binding assays were performed exactly as described in (12).

RESULTS

Purification of Nuclear Factor I (NFI)

NFI was purified from nuclei of porcine liver to homogeneity through three chromatographic steps, DEAE-cellulose, MonoQ-, and recognition site affinity chromatography. This last step employed a sepharose column containing covalently bound double-stranded oligonucleotides of a length of 32 bp, carrying a binding site for NFI (fragment L1/2 in Table I)(22). This binding site displays perfect dyad symmetry and is derived from mouse adenovirus FL DNA, which carries this sequence towards its right end terminus outside of the ITR (Fig. 1) (24). NFI from porcine liver eluted from the DNA affinity column at a salt concentration of 1.5 - 2.5 M KCl. The total purification achieved in these steps was 25.000 to 50.000 - fold (Table II).

Characterization of Nuclear Factor I

Upon SDS-polyacrylamide gelelectrophoresis, the affinity-purified factor shows a single band with a molecular weight of 36 kDa (Fig. 4). In DNase I protection experiments the observed footprints were undistinguishable from those observed with crude extracts and various column fractions (Fig. 5). Initiation of adenovirus DNA replication *in vitro* was tested with cytoplasmic and nuclear extracts from adenovirus-infected Hela cells. The observed increase (Fig. 6) in the transfer of [α - 32 P] dCTP to the precursor of the terminal protein, an indicator for

Figure 5: DNaseI footprint analyses of porcine liver NFI preparations. The analysis was performed on a 146 bp restriction fragment (F146) derived from the left-hand terminus of Ad5 DNA containing the origin of DNA replication with NFI and NFIII binding sites. Lanes 1,5,6 and 9 represent footprints performed in the absence of NFI. Lanes 2, 3 and 4 contain 1, 2 and 10 ng (in a volume of 50 μ l) of affinity purified NFI, lanes 7 and 8 show footprint analysis with 10 ng of partially purified NFI. The positions on the adenovirus ITR are indicated.

the initiation of adenovirus DNA replication, clearly confirmed the activity of the purified porcine nuclear protein. The stimulation (3-5 fold) of the initiation reaction by the affinity purified porcine factor was less under our conditions than reported for the human factor (20 fold). However, the stimulation was independent of whether porcine NFI was added before or after the addition of the HeLa cell nuclear extract (containing the virus-coded DNA polymerase/terminal protein complex in addition to some human NFI) indicating that porcine NFI both binds to the DNA and stimulates the reaction. Porcine liver NFI thus has all the properties expected of an NFI-like protein.

DNA-binding characteristics of porcine liver NFI

As an additional characteristic of porcine liver NFI we studied its binding to a variety of

Figure 6: *In vitro* assay for adenovirus DNA replication with NFI from porcine liver. Part A: Stimulation of the initiation reaction with a partially purified NFI preparation; Lane 1: Ad5 infected cytoplasmic extracts alone; Lane 2: plus NFI. Part B: Complementation of infected cytoplasmic extracts (Lane 1) with: affinity column flow through (Lane 2), purified NFI (Lane 3), flow through plus purified NFI (Lane 4) and with infected HeLa cell nuclear extract (Lane 5). Part C: Initiation of DNA replication with 2 μ l (Lane 1) and 8 μ l (Lane 2) nuclear extract from infected HeLa cells. Lane 3: Replication assay with 8 μ l infected nuclear extract from infected HeLa cells after preincubation of the DNA template with 400 fmol affinity purified NFI from porcine liver. The protein concentration in the nuclear extract was 5 mg/ml.

DNA-binding sites by the gel retention assay. Equilibrium binding constants were determined by measuring the amount of complex formation between NFI and DNA as a function of increasing DNA concentrations (Fig. 7). The calculations, as outlined in [Material and Methods](#) require the determination of three parameters: i) of the total concentration of "active" NFI; ii) of "active" DNA and iii) of the respective DNA-protein complex concentrations. The concentration of NFI is identical to the amount of complex formed in the presence of an excess of DNA. The concentration of "active" DNA is calculated as the difference between the amount of DNA transferred into the complex and free DNA, when measured in the presence of an excess of protein. While restriction fragments were always fully "active", the activity of chemically synthesized oligonucleotides fragments varied between 60 - 80 % (Fig. 7, lane 1).

Equilibrium binding constants were obtained from a titration of active NFI against increasing concentrations of DNA at 25 °C and 150 mM NaCl, pH 7.5. In general, K was determined from three to five measurements of [c] (Fig. 7, lanes 3 to 7) performed at increasing concentrations of [DNA] 0 . In Fig. 7, lane 7, the complex concentration [c] was similar to that in lane 8 which measured [NFI] 0 in the presence of approximately a 100-fold excess of DNA. This value in lane 7 was thus regarded as too close to saturation and not used in calculating K.

Figure 7: Determination of equilibrium binding constants of the interaction of NFI with oligonucleotide fragments carrying NFI binding sites. The fragment used in this experiment was EW1/2 (Table I) with a binding site derived from Ad5 DNA. The concentrations are given in $[M] \times 10^{11}$, binding constants in $1/[M] \times 10^{-10}$. Lanes 1 and 8 are controls used for the determination of "active" DNA (lane 1; protein excess) and for the initial concentration of NFI (lane 8; DNA excess).

Equilibrium binding constants are summarized in Fig. 8 as well as in Tables I and III. The highest constants ($2 \times 10^{11} \text{ M}^{-1}$) were observed for binding to a fully symmetrical NFI binding site as present close to the terminus, but outside of, the ITR of mouse adenovirus FL (*cf.* Fig. 1). A random sequence oligonucleotide shows a binding constant of $5 \times 10^6 \text{ M}^{-1}$.

Base substitutions **outside** of the consensus sequence TGG(A/C)N₅GCCAA had little or no effect. Transversions **within** the consensus-binding site invariably led to significant reductions of at least two orders of magnitude. The closer these are located to the central G/C base pair at position 31 the stronger is their influence on the binding constant. Thus, there is a difference of a factor of 5 in the binding constant of fragment G1/2 versus T1/2. Since the limit of error in these experiments is only approximately 100%, the observed differences are significant. The most dramatic effect was observed for the case of a replacement of the entire half-site -CCAA- by -AACC- (fragment IV1/2). An exchange of the central pentanucleotide -TTGAA- by -GCGCG- is without influence on the binding constant (fragment TT1/2 in Table I). This result is certainly in agreement with the observations by de Vries et al. (33) that phosphates from these central base pairs, when ethylated, display only weak, if any, interference with NFI binding.

Figure 8: Summary of binding constants for DNA fragments with full or with half-sites for NFI. The presentation is half-logarithmic. Bars are labelled as shown in Table II. The sequence in the lower portion of the Figure represents the sequence of the NFI binding site from Ad5 DNA. Sequences above, connected with arrows, indicate mutations with the arrows pointing to the appropriate bars.

For oligonucleotide fragments containing the CCAAT-box site from transcriptional start sites of various eukaryotic genes, e.g. the human α -globin gene, the human *c-myc* gene, the HIV-LTR enhancer, and the HSV-*tk* gene, we observed equilibrium binding constants between 10^7 to $2 \times 10^8 M^{-1}$ (Table III). These values are two to four orders of magnitude lower than binding constants observed for binding sites with the full, dyad-symmetry.

We also studied the binding of NFI to a 226 bp, CCAAT-box containing DNA fragment derived from the human *hsp70* promoter as well as to a linker-scan mutant (25) displayed in Table III (*hsp70**). As shown in Fig. 9, both the wildtype and the mutant fragment were retarded in gel retardation assays. However, binding constants could only be estimated because of the presence of multiple bands. These are comparable to those observed for the binding of NFI to the HSV-*tk* CCAAT fragment and thus display only little specificity. The estimated values of about $10^7 M^{-1}$ approach those observed for unspecific binding of NFI to DNA although we consistently observed a preference for the *hsp70*-CCAAT box containing fragment as compared to the linker-scan mutant of at least a factor of 2.

		K	Sym
Consensus	NN TGG NNNNNN GCCA NNNN	$> 10^{10}$	3/5
h-hsp70	TT CCCTT CTGAG GCCA TCAC	$\sim 10^7$	0/5
h-hsp70*	TT CCCTT CTGAG GCCCG TCGA	$< 10^7$	0/3
HIV-LTR	GCT TGGG ACTT TCCAG GGAG	$\sim 10^8$	3/3
HSV-tk	TG TTCGA ATT GCCA ATGAC	1×10^7	1/5
α -globin (1)	GCT TCCG CGCCAG GCCA ATGAG	?	1/5
α -globin (2)	CC GGG CTCCGC GCCAG AACG	2×10^8	2/4
c-myc	TT CTTTT TCCC GCCA GCCT	2×10^7	0/5
random	NNNNNNNNNNNNNNNNNN	5×10^6	-

Table III: Sequence homologies between various promoter CCAAT boxes and the NFI consensus sequence. Sequences in positions comparable to both motifs of the NFI consensus sequence are shown in bold-face letters. Equilibrium binding constants of purified NFI from porcine liver to the corresponding DNA fragments are indicated (in M^{-1}) as well as the extent of identity (Sym) to the respective motifs in the NFI consensus. The sequence hsp70* indicates a linker-scan mutation (25). Binding constants for the hsp70 sequences are only estimates (see text).

DISCUSSION

We describe a purification procedure for porcine NFI which is based on classical purification steps followed by DNA-affinity chromatography. The result is a homogeneous protein preparation with a molecular weight of 36 kDa. This was independent of the presence of PMSF or meta-bisulfite in the various buffers when added as a precaution against proteolysis. In contrast, the protein preparation purified by similar procedures from human Hela cells invariably represents a mixture of proteins with apparent MWs varying between 52 and 56 kDa (16) and even 160 kDa (26). These discrepancies remain unresolved and may reflect the different species and cell types/organs (human *versus* porcine; tissue culture cells *versus* liver).

The polypeptide characterized in this communication displays all the properties expected from an NFI preparation, *i.e.* binding to the NFI binding site from Ad2-DNA, and stimulation *in vitro* of the initiation of adenovirus DNA replication.

Equilibrium binding constants for the interaction of DNA binding proteins have been determined traditionally by employing nitrocellulose filter binding assays (27, 28). Assuming a simple bimolecular reaction, a value of $5 \times 10^{10} M^{-1}$ was obtained (16) for the equilibrium binding constant of the interaction between NFI and its Ad2-DNA binding site. The use of gel retention assays was originally introduced by Fried and Crothers (29) and has recently been employed (30) to study the thermodynamics of interaction between transcription factor MLTF and its binding site within the adenovirus major late promoter. It has several distinct advantages which make it superior to the filter-binding assay (29, 31). The gel retention assay is particularly

Figure 9: Gel retardation assay of a promoter fragment from the human hsp70 gene with purified NFI from porcine liver. Lanes 1 to 4 represent analyses with a 226 bp DNA fragment from positions -130 to 95 of the hsp70 gene carrying the -CCAAT-site at -62 to -67. Lanes 2 and 4 were performed with the wt promoter, lanes 1 and 3 with a fragment of identical size carrying a linker-scan mutation in the CCAAT-box (OSS 64-69) which changes the sequence CCAAT into CCCGT. Lane 1 to 4 contain each about 20 fmol DNA, 200 ng poly d(IC) and 800 fmol NFI (Lane 1, 2), 1600 fmol (Lane 3, 4) in a volume of 20 μ l. The wildtype and mutant sequences surrounding the CCAAT sites are shown in Table III.

useful for the study of low-abundance proteins because it can be used at a level of sensitivity of much less than 1 % of the total activity in an extract, as compared to 5 to 10% in the filter-binding assay. Since the method does not only select on the basis of binding specificity but also on a characteristic mobility shift, gel retention does include neither degradation products nor unspecific DNA-binding proteins in its measurements as long as these lead to different mobility shifts. Finally and most important, the gel shift assay can be used accurately for measuring low equilibrium binding constants (10^7 M^{-1}), i.e. under conditions where the addition of carrier DNA already results in considerable competition. This, however, requires small fragment sizes (30 - 40 bp) since at binding constants close to those observed for unspecific binding, associations between the protein and larger DNA fragments give rise to multiple bands (e.g. Fig. 9).

Our equilibrium binding constants for NFI from porcine liver vary between $2 \times 10^{10} \text{ M}^{-1}$ (fragment A1/B1) and $2 \times 10^{11} \text{ M}^{-1}$ (fragment E1/E2) for interactions with the binding sites from Ad2 DNA and AdFL DNA, respectively. Within the experimental error these values are independent of DNA fragment lengths, i.e. a 39 bp long synthetic double-stranded DNA fragment and a restriction fragment of larger size will give the same results. The observed values for NFI equilibrium binding constants compare favourably with other DNA-binding proteins, e.g. cro-repressor (32), the *E. coli* CAP protein (29) and the adenovirus major late transcription factor MLTF (30).

We were surprised to find that, even under conditions of protein excess, a certain fraction of the DNA never entered the DNA/protein complex. When synthetic DNA fragments were used, "resistant" DNA frequently amounted to up to 50% of the input DNA under equilibrium conditions. This was almost never observed with native restriction fragments. This was not a question of single- versus double-stranded DNA. Displaying overhanging termini, fragments were always labelled using the Klenow fragment of DNA polymerase I such that only double-

stranded DNA could be detected. One reason for the observed impairment of complex formation may be chemical modifications, *i.e.* protecting groups remaining on the fragment after their synthesis.

One central question with respect to the function of NFI relates to its role as a -CCAAT- motif binding transcription factor. This role was suggested by the fact that the canonical CCAAT sequence matches part of the NFI consensus sequence and that a protein (CTF-1) purified from affinity columns carrying the α -globin or Ha-*ras* -CCAAT- sites was indistinguishable from NFI (18).

We have not been able to purify a specific CCAAT-binding factor from porcine liver by CCAAT-motif DNA affinity chromatography using an HSV-*tk* CCAAT-derived oligonucleotide although we have seen small amounts of an EBP20-like activity (19). Instead, we have only obtained the NFI activity described in this report with binding constants to CCAAT-binding sites which are at least two orders of magnitude below the values observed for canonical NFI binding sites. These low values raise questions as to the biological significance of these interactions.

For once, they are far below those observed previously (18) when a difference of only a factor of 5 was obtained between adenovirus type 2 NFI-sites and the CCAAT-binding sites from the human α -1 globin gene. By purifying NFI activities from HeLa cells, we are currently trying to clarify this issue. Preliminary data indicate that the situation in HeLa cells is indeed considerably more complex than in the porcine system.

Most disconcertingly, we have identified an additional protein from porcine liver nuclear extracts which, while displaying little or no sequence-specific DNA binding capacity, strongly binds to double-stranded DNA with estimated binding constants of $>10^9 \text{ M}^{-1}$. These values thus exceed those for the interaction of the NFI-like protein from porcine liver with canonical -CCAAT- sites by at least a factor of 10. Our preliminary data indicate that the amount of this protein in nuclear extracts equals that of NFI. This protein could therefore interfere with NFI binding to low-affinity sites.

Finally, regarding the low-affinity binding sites, the question arises why the equilibrium binding constants for these various sites vary by factors of 20 and more. After comparing the various sequences we would like to propose that the magnitude of a binding constant depends on the degree of identity of a potential recognition site with **both** motifs of the NFI consensus (marked in bold letters in Table III), *i.e.* the TGG **and** the GCCAA motif. This proposition is in agreement with the contact point analysis described in (33) for high-affinity binding interactions. For example, in the HSV-*tk* site, the extent of identity is 1 out of 3 (TCC instead of TGG) and 5 out of 5 (GCCAA); in the *hsp70* site, it is 0 and 5, and, accordingly, the binding constants are close to values observed for unspecific binding. However, this rule would not fit the case of the α -1 globin gene. With an extent of identity of only 1 and 5 it should display a binding constant similar to that of the *tk*- gene; instead, it is 20 times higher. If, however, we look at a sequence

removed by only 4 bp upstream from the expected binding site, we recognize two motifs in the α -1 globin promoter with degrees of identity of 2 and 4 (see α -globin 2 in Table III). This higher extent of symmetry thus would explain the higher binding constant but would also lead to the conclusion that NFI does **not** bind exactly to the CCAAT box in the α -globin promoter. It remains to be seen whether we can confirm this conclusion experimentally.

In the case of the HIV-LTR (*tat* enhancer region) the extent of identity is 3 and 3 thus predicting a higher binding constant for NFI than in the case of the HSV-*tk* promoter. This could indeed be demonstrated (Table III) by obtaining a value of $2 \times 10^8 \text{ M}^{-1}$ and a clear footprint (protected from positions -95 to -75, not shown) for this particular sequence.

In addition, we analyzed mutations within the NFI consensus sequence. Changing the first G residue, for example, in the GCCAA box to a T (as it is found in the NFI binding site within the ITR of mouse AdFL, see Fig. 1) reduced the affinity to a factor of approximately five (data not shown), while the replacement of the first A by a T-residue resulted in a reduction of the binding constant by a factor of 20. The binding is even more sensitive to an exchange of the first C-residue in the GCCAA box by G, leading to a reduction of K by a factor of 100 (see Fig. 8 and Table I).

Finally, the high affinity to E1/2 and to the mutation from TGG(N)₅AGCCAA to TGG(N)₅TGCCAA (see Figure 8) seems to indicate that the half site TGCCAA is recognized better than AGCCAA.

In closing, we have to stress that the functional role of NFI, apart from its role as a binding factor which stimulates adenovirus DNA replication, remains obscure. One prerequisite for the solution of this problem as well as of the question concerning the discrepancies between the human and porcine systems may be the cloning of the appropriate cDNAs. We have recently compared the aminoacid sequences of various tryptic- and cyanogen bromide fragments from our highly purified porcine liver NFI preparation with sequences from cloned CTF cDNAs (R. Tjian, unpublished). These two proteins, CTF-1 and NFI, indeed share considerable stretches of their protein sequence, while other regions are totally unrelated. It thus appears that NFI and CTF-1 are members of a closely related family of proteins which have different functions in DNA replication and transcription. These questions are currently under study.

ACKNOWLEDGEMENTS

We are grateful to Renate Föckler for her splendid technical assistance. The hsp70 derived promoter fragments were kindly provided by Dr. R. Kingston, Boston, while we owe the *c-myc* derived fragments to Dr. M. Lipp and K. Thalmeier. This work was supported by the Deutsche Forschungsgemeinschaft (Fa-138/3-1).

REFERENCES

1. Nagata, K., Guggenheimer, R.A., Enomoto, T., Lichy, J.H., and Hurwitz, J. (1982) Proc. Natl. Acad. Sci. USA **79**, 6438-6442.
2. Leegwater, P.A., van Driel, W., and van der Vliet, P.C. (1985) EMBO J. **4**, 1515-1521.

3. Henninghausen, L., and Fleckenstein, B. (1986) *EMBO J.* **5**, 1367-1371.
4. Rawlins, D.R., Rosenfeld, P.J., Kelly, T.J., Milman, G.R., Jeang, K.T., Hayward, S.D., and Hayward, G.S. (1986) In *Cancer Cells-DNA Tumor Viruses: Control of Gene Expression and Replication*, Volume 4 (Cold Spring Harbor, New York, 11724: Cold Spring Harbor Laboratory), pp. 525-542.
5. Siebenlist, U., Henninghausen, L., Battey, J., and Leder, P. (1984) *Cell* **37**, 381-391.
6. Borgmeyer, U., Nowock, J., and Sippel, A. (1984) *Nucl. Acids Res.* **12**, 4295-4311.
7. Gronostajski, R.M., Adhya, S., Nagata, K., Guggenheimer, R.A., and Hurwitz, J. (1985) *Mol. Cell. Biol.* **5**, 964-971.
8. de Vries, E., van Driel, W., Tromp, M., van Boom, J., van der Vliet, P.C. (1985) *Nucl. Acids Res.* **13**, 4935-4950.
9. Wides, R.J., Challberg, M.D., Rawlins, D.R., and Kelly, T.J. (1987) *Mol. Cell. Biol.* **7**, 864-874.
10. Schneider, R., Gander, I., Muller, U., Mertz, R., and Winnacker, E. L. (1986) *Nucl. Acids Res.* **14**, 1303-1317.
11. Gronostajski, R.M. (1986) *Nucl. Acids. Res.* **14**, 9117-9232.
12. Gronostajski, R.M. (1987) *Nucl. Acids. Res.* **15**, 5545-555.
13. Jones, K.A., Yamamoto, K.R., and Tjian, R. (1985) *Cell* **42**, 559-572..
14. McKnight, S.L., Kingsbury, R.C., Spence, A., and Smith, M. (1984) *Cell* **37**, 253-263.
15. Nagata, K., Guggenheimer, R.A., and Hurwitz, J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6177-6181.
16. Rosenfeld, P.J., and Kelly, T.J. (1986) *J. Biol. Chem.* **261**, 1398-1306.
17. Gander, I., and Winnacker, E. L. (1988), submitted.
18. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987) *Cell* **48**, 79-89.
19. Johnson, P.F., Landschutz, W.H., Graves, B.J., and McKnight, S.L. (1987) *Genes and Development* **1**, 133-146.
20. Cohen, R.B., Sheffrey, M., and Kim, C.G. (1986) *Mol. Cell. Biol.* **6**, 821-832.
21. Graves, B.J., Johnson, P.F., and McKnight, S.L. (1986) *Cell* **44**, 565-576.
22. Kadonaga, J.T., and Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5889-5893.
23. Lally, C., Dörper, Th., Gröger, W., Antoine, G., and Winnacker, E. L. (1984) *The EMBO J.* **3**, 33-337.
24. Antoine, G., Alestrom, P., Schilling, R., Pettersson, U., and Winnacker, E. L. (1982) *The EMBO J.* **1**, 453-459.
25. Greene, J.M., Larin, Z., Taylor, I.C.A., Prentice, H., Gwinn, K., and Kingston, R.E. (1987) *Mol. Cell. Biol.* **7**, 3245-3655.
26. Diffley, J.F.X., and Stillman, B. (1986) *Mol. Cell. Biol.* **6**, 1363-1373.
27. Riggs, A.D., and Bourgeois, S. (1968) *J. Mol. Biol.* **34**, 361-368.
28. Riggs, A.D., Bourgeois, S., and Cohn, M. (1970) *J. Mol. Biol.* **53**, 401-417.
29. Fried, M.G., and Crothers, D.M. (1984) *J. Mol. Biol.* **172**, 241-262.
30. Chodosh, L.A., Carthew, R.W., and Sharp, P.A. (1986) *Mol. Cell. Biol.* **6**, 4723-4733.
31. Rosenfeld, P.J., O'Neill, E.A., Wides, R.J., and Kelly, T.J. (1987) *Mol. Cell. Biol.* **7**, 875-886.
32. Takeda, Y., Folkmanis, A., and Echols, H. (1977) *J. Biol. Chem.* **252**, 6177-6183.
33. de Vries, E., van Driel, W., van den Heuvel, S.J.L., and van der Vliet, P. (1987) *EMBO J.* **6**, 161-168.
34. Laemmli, U.K. (1970) *Nature* **227**, 680-685.
35. Oakley, B.R., Kirsch, D.R., and Morris, N.R. (1980) *Anal. Biochem.* **105**, 361-363.