Structural and Functional Organization of a Porcine Gene Coding for Nuclear Factor I[†,‡]

Michael Meisterernst, Lars Rogge, Renate Foeckler, Marina Karaghiosoff, and Ernst-L. Winnacker*

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

Received June 12, 1989; Revised Manuscript Received July 20, 1989

ABSTRACT: This paper describes the structure of a 70-kb porcine gene for nuclear factor I, including its promoter region, comprising a total of 11 exons. Different mRNAs that we have isolated as cDNAs from both porcine liver and human HeLa cells presumably are generated from this gene by differential splicing events. One cDNA species from porcine liver that lacks exon 9 carries coding information for a protein of 439 amino acids. The in vitro translated protein displays all the properties of an NFI-like protein with high affinity toward the sequence element TGG(N)$_6$GCCAA, as shown by gel shift analysis, and no or little affinity toward CCAAT box containing sequences. Cotranslation experiments with full-length and truncated variants of the protein demonstrate that it binds as a dimer to its cognate DNA recognition sequence. Its DNA-binding domain which is retained in all cDNA clones was mapped by deletion analysis to the 250 N-terminal amino acids of the protein. No structural homologies are observed between this protein and other known DNA-binding proteins; instead, the protein contains a novel α-helical sequence motif consisting of several lysine residues spaced at intervals of seven amino acids which we have termed the "lysine helix". The C-terminal portion of the protein derived from full-length cDNAs encodes a short amino acid sequence which is identical with the heptapeptide repeat CT7 observed in the C-terminal domain of the largest subunits of yeast and mouse RNA polymerase II. This region is removed by differential splicing in some of the NFI/CTF cDNAs and thus may be of functional significance.

The 5′-CCAAT-3′ sequence motif serves as a cis-regulatory element both in eukaryotic class II gene transcription and in adenovirus DNA replication. Deletions or point mutations affect basal levels of transcription, for example, from the human α-globin, the mouse β-globin, the RSV-LTR, the α1(I) and α2(I) collagen promoters, the human *hsp*70, and the HSV thymidine kinase promoter (Myers et al., 1986; Jones et al., 1985, 1987; Maity et al., 1988). CCAAT box binding proteins have been shown to bind to a variety of other promoters, e.g., the sea urchin histone H2B promoter, the MHC class II E_α promoter, the human γ-globin promoter, the mouse α1 globin gene promoter, the adenovirus ML promoter, the untranslated leader region of the HIV-LTR, and the yeast CYC1 gene promoter (Cohen et al., 1986; Barberis et al., 1987; Dorn et al., 1987; Chodosh et al., 1988; Jones et al., 1988; Superti-Furga et al., 1988).

The CCAAT sequence motif frequently occurs as part of the inverted repeat TGG(N)$_6$GCCAA, often referred to as the NFI binding site, which has been found within the 5′-flanking regions of liver-specific genes, e.g., the chicken lysozyme gene, the hamster HMG-CoA reductase gene, and the albumin- and the retinol-binding protein gene (Borgmeyer et al., 1984; Gil et al., 1988; Colantuoni et al., 1987; Lichsteiner et al., 1987; Raymondjean et al., 1988), and within the origin of adenovirus DNA replication. Protein binding to the NFI site in the adenovirus terminal repetition has been shown to stimulate the initiation reaction of adenovirus DNA replication (Nagata et al., 1982). The protein involved has been proposed to be identical with the CCAAT binding factor stimulating transcription from the α-globin promoter and is thus referred to as the NFI/CTF protein (Jones et al., 1987). Detailed binding studies have led us to question these conclusions and, instead, to postulate the existence of a family of NFI-like proteins (Meisterernst et al., 1988a).

At least three different CCAAT-binding protein factors [designated *C*CAAT-binding *t*ranscription *f*actors of CTFs by Santoro et al. (1988)] have been characterized recently as cDNA clones. Among these are TGG(N)$_6$GCCAA-binding proteins from rat and hamster liver (Paonessa et al., 1988; Gil et al., 1988), a CCAAT-binding protein from rat liver, referred to as C/EBP (Landschulz et al., 1988a), and a family of CCAAT-binding proteins from human HeLa cells (Santoro et al., 1988). In addition, a portion of the gene for a porcine liver derived NFI-like factor has been characterized by Meisterernst et al. (1988b). While factor C/EBP is certainly distinct from the others on the basis of its cDNA sequence, all other cDNA clones appear to be related but not identical. The observed identity is extremely patchy with extended regions of homology in the 5′-terminal third of the molecules, followed by regions of only marginal homology toward the 3′ terminus. These variations have been attributed to differential RNA splicing events in the case of the HeLa cell derived cDNAs and to the presence of different genes in the hamster liver derived clones (Gil et al., 1988). In order to contribute to a solution of this puzzle, we have continued to characterize the porcine liver derived NFI gene, the structure of which is presented in this paper. We will also discuss the structures of additional and novel cDNA clones obtained both from porcine liver and from human HeLa cells. In addition, we will present a variety of functional studies concerning the DNA binding site of the NFI/CTF protein including experiments on the stoichiometry of DNA binding. Finally, the presence of a novel structural element within the NFI/CTF DNA

[†] This work was supported by a grant to E.-L.W. from the Deutsche Forschungsgemeinschaft (Forschergruppe Fa 138 3/1).

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02875.

* To whom correspondence should be addressed.

0006-2960/89/0428-8191$01.50/0

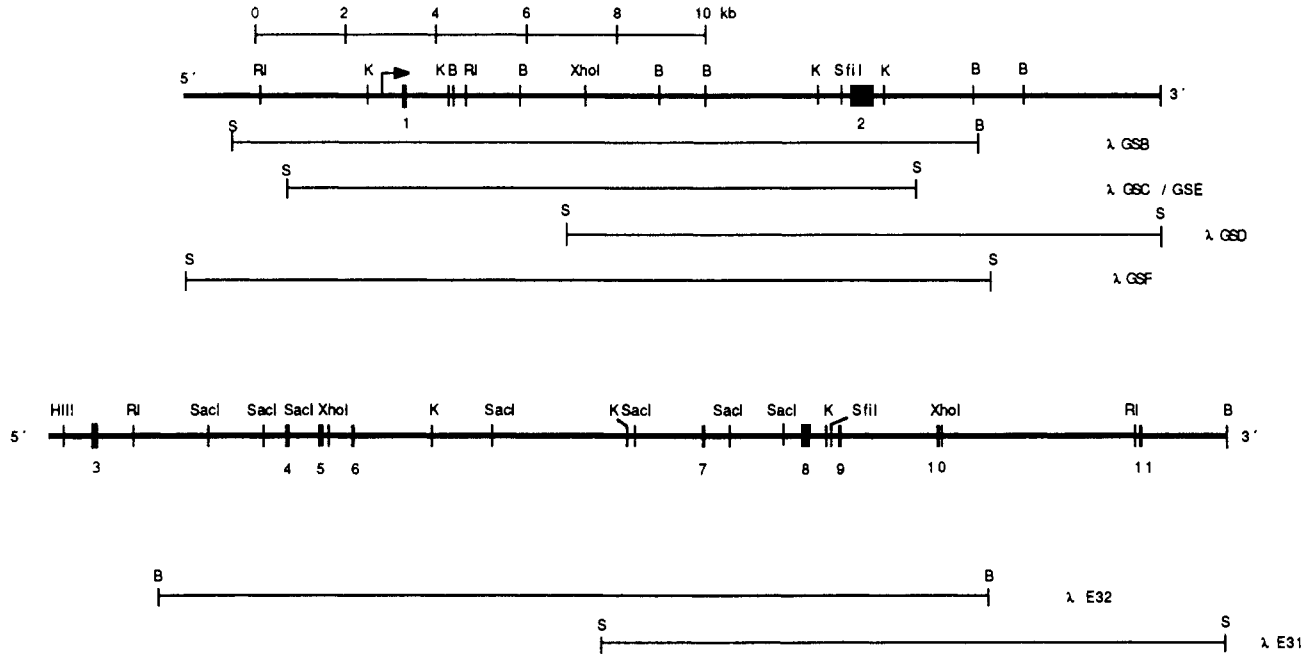

FIGURE 1: Structure and restriction map of the porcine NFI/CTF gene (B = *Bam*HI, HIII = *Hin*dIII, K = *Kpn*I, RI = *Eco*RI, S = *Sal*I). The upper and the lower parts represent 5′ and 3′ regions of the gene, respectively. They are separated from each other by at least 30 kb and are represented by EMBL3A clones, as indicated. Exons are marked by numbers beginning with exon 1 at the 5′ terminus of the gene. The major transcriptional start site is marked with an arrow.

binding domain, a lysine repeat or "lysine helix", will be discussed.

Materials and Methods

Purification of NFI. NFI has been purified to homogeneity from porcine liver by conventional column chromatography and DNA affinity chromatography as described previously (Meisterernst et al., 1988a). NFI from HeLa cells has been purified by a procedure essentially analogous to porcine NFI purification. Purified HeLa cell derived NFI preparations consisted of mixtures of DNA-binding proteins, as previously observed by Jones et al. (1987).

Isolation of the NFI/CTF Gene and Corresponding Porcine and Human cDNA Clones. The isolation of genomic porcine NFI clones from an EMBL3A library of porcine genomic DNA has been described previously (Meisterernst et al., 1988b). cDNA clones were obtained by screening commercially available λgt11 porcine liver and HeLa cell derived cDNA libraries (Clontech, PL 1001 B and HL 1008) with a 423-bp *Pst*I/*Sac*I fragment covering 141 amino acids from exon 2. Out of a total of 1.4 million clones in each library, 20 porcine cDNA and 16 human cdNA clones were obtained and sequenced. The porcine cDNA clone (NFI/p-CTF2) containing sequences from exon 2 to exon 8 as well as exons 10 and 11 was used as a probe to isolate the carboxy-terminal part of the porcine NFI gene. Exon 9 was detected by sequencing part of the intron between exons 8 and 10. The entire gene has been subcloned in pUC vectors. Coding regions as well as 300–1000 bp of the surrounding introns were sequenced by a modified T7 polymerase protocol according to the directions of the manufacturer (U.S. Biochemical Corp.).

Southern Blot Analyses. Southern analysis was performed by a standard protocol (Maniatis et al., 1982). The "Church" protocol was used to perform DNA carrier free Southern blot analysis (Church & Gilbert, 1984). DNA fragments including cDNA probes were labeled by the random hexamer primer method according to the protocol provided by the manufacturer (Boehringer, Mannheim).

In Vitro Synthesis of RNA and Expression of NFI. A synthetic oligonucleotide of the sequence AAGCTTCGCATGCCTGCAGCCGCCATG displaying a functional eukaryotic translational start site (Kozak, 1984) was fused with the first codon of clone NFI/p-CTF2, thus creating an *Nco*I site and an artificial start codon in front of exon 2. The resulting construct was cloned into an *Eco*RI/*Hin*dIII-linearized pGEM-3Z (Promega) downstream of the SP6 promoter. Capped transcripts were synthesized in vitro from this expression clone and from various 3′ deletions in the presence of 500 μM m^7GpppG according to Pelletier and Sonnenberg (1985). RNA was treated with RNase-free DNase, phenolized, purified on Sephadex G-50 columns, and precipitated twice with 1 volume of 2-propanol. In vitro translations were carried out in rabbit reticulocyte lysates (Promega) at 30 °C for 2 h. Protein mixtures were analyzed for quality and size on SDS–polyacrylamide gels and for DNA-binding activities on native polyacrylamide gels.

Gel Electrophoresis on Native Polyacrylamide Gels. One to three microliters out of 50 μL of a reticulocyte extract was incubated in buffer A (25 mM Hepes/KOH, pH 7.8, 150 mM KCl, 5 mM EDTA, 1 mM DTT, and 10% glycerol), 1 μg of poly(dI/dC), and 30 fmol of the unlabeled DNA fragment F146 for 20 min at 20 °C and subsequently loaded onto a 12% native polyacrylamide gel. The polyacrylamide gel was prepared as described previously (Meisterernst et al., 1988a). F146 is a 146-bp DNA fragment derived from the left-hand terminus of adenovirus type 5 DNA and thus includes the entire origin region with an intact NFI binding site. In protein competition assays, the DNA was incubated for 10 min at 0 °C with affinity-purified NFI from porcine liver prior to addition of the reticulocyte extract containing ^{35}S-labeled NFI or truncated NFI protein. In DNA competition assays, DNA fragment F146 and a 10- or 25-fold molar excess of competitor DNA were added at the same time.

RNase Protection and Primer Extension Assays. For RNase protection assays, appropriate DNA fragments were cloned into vector pGEM-3Z/4Z. Following linearization, [α-^{32}P]CTP-labeled run-off transcripts were synthesized with

----	Exon1	**GT**ACG**G**CCCCCCGCCTGGCCC
TGACCTCTGGCTTCTCGCGCCCGC**AG**	Exon2	**GT**GAG**G**CGGGGCGTGCACGT
----	Exon3	----
TGACCCCCATCCTCCTACTCTTCC**AG**	Exon4	**GT**AAG**T**GTCCTCACCCACAC
ACTCTCCTCTCTGTTGCCCTCCAC**AG**	Exon5	**GT**ATG**T**GCCCAAAGCCAAAC
ATTGGCTTCCTCCCCTCGTCCCTT**AG**	Exon6	**GT**GGG**G**CTGGCGGCGCGGCG
CCTTGACCCCATGGCCCTCCCCAC**AG**	Exon7	**GT**AAG**C**ACTGAGGGCCCCAG
GCGGACTAACCGCCCCCTCTCCAC**AG**	Exon8	**GT**GAG**T**TTTGGGGGGCACGA
CCACCTGTCTCTCTGTTTCCCCTC**AG**	Exon9	**GT**AAG**C**ACTCAGGGAAGTGG
GCTCTCGGTCTCTCTCCTCCCCGC**AG**	Exon10	**GT**AGG**G**GGCCAAGAGGCGGC
TCTCTCTCTTTCTCTGGGCCACCC**AG**	Exon11	----

FIGURE 2: DNA sequences of splice sites of the NFI/CTF gene. Each line represents the 3′ splice site of an intron followed by the appropriate exon and the 5′ splice site of the subsequent intron. Conserved bases are shown in boldface letters.

SP6 RNA polymerase. One nanogram of radiolabeled probe was hybridized overnight to 100 μg of total RNA or 10 μg of poly(A) RNA at 48 °C in a volume of 30 μL in a buffer containing 80% formamide, 400 mM NaCl, 40 mM Pipes-HCl, pH 6.7, and 1 mM EDTA. RNase digestions were performed at 30 °C for 1.5 h in 350 μL of a buffer containing 10 mM Tris-HCl, pH 7.5, 5 mM EDTA, 300 mM NaCl, 40 μg/mL RNase A, and 2 μg/mL RNase T_1. Reaction mixtures were phenol extracted, precipitated, and analyzed on 5% polyacrylamide–8.3 M urea sequencing gels. Primer extension analyses were performed according to a protocol of Ausubel et al. (1987). The hybridization buffer contained 400 mM NaCl, 40 mM Pipes-HCl, pH 6.7, and 1 mM EDTA.

RESULTS

Structure of a Porcine NFI Gene Including Its Transcriptional Start Site. Our initial screen of an EMBL3A library of genomic porcine DNA with oligonucleotides as hybridization probes identified an EMBL3A clone containing exon 2 of the NFI gene (Meisterernst et al., 1988b). A 423 bp long *Pst*I/*Sac*I fragment covering parts of this exon was used to rescreen the library for additional clones. As indicated in Figure 1, these clones characterize 11 exons which were identified through a comparison with the porcine liver derived cDNA sequence NFI/p-CTF2 (Figure 6). The two regions of the gene represented by exons 1 and 2 and exons 3–11, respectively, have not yet been linked by overlapping clones. However, Southern blot analysis indicates that the distance between exons 2 and 3 amounts to approximately 35 kb. We thus estimate the NFI gene to have a size of approximately 70 kb. Sequences of 5′ and 3′ splice sites are depicted in Figure 2 and display the characteristic consensus motifs of splice junctions.

Exon 1 has been identified on the basis of its presence, together with an untranslated leader, in one of our cDNA clones, h-CTF4, from human HeLa cells (Figure 7). By using a DNA fragment derived from clone h-CTF4, it has been possible to isolate an EMBL3A clone from a porcine genomic library and to obtain the porcine sequence for exon 1 depicted in Figures 4 and 6. Its sequence is identical with that of the human sequence in clone h-CTF4.

Transcriptional starts sites of the porcine NFI/CTF gene were identified both by RNase protection and by primer extension analyses (Figure 3). RNase protection assays were performed with a labeled, 1.8-kb antisense riboprobe obtained as an SP6 run-off transcript from a 1.8-kb *Kpn*I fragment (positions −795 to +1000; Figure 3C) cloned into vector pGEM-3Z. When hybridized to total porcine liver RNA, two fragments of lengths of approximately 465 and 409 bp were

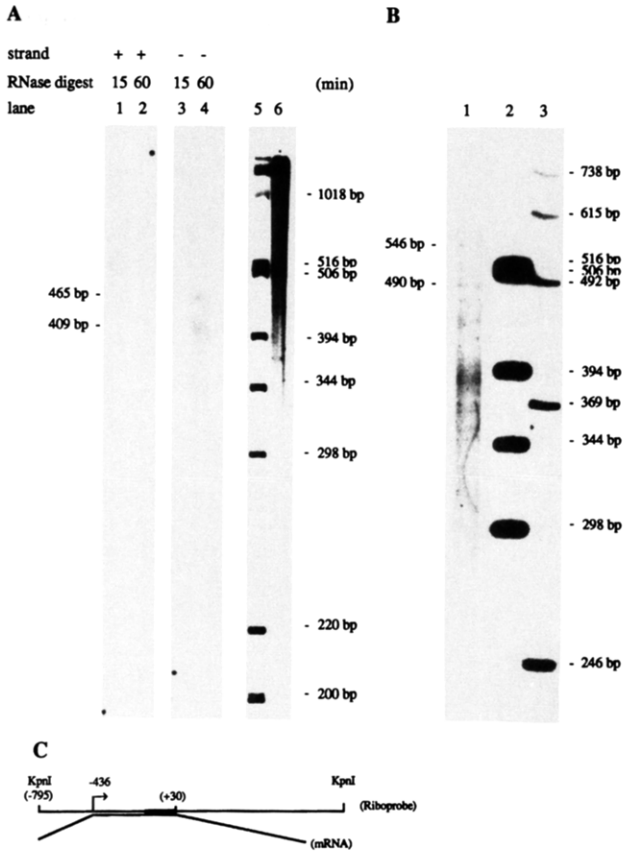

FIGURE 3: Identification of transcriptional start sites by RNase protection and primer extension assays. (A) Gel electrophoretic analysis of reaction products obtained with a 1.8-kb antisense riboprobe covering the promoter region of the NFI/CTF gene. Lanes 1 and 2: Control reaction with a sense riboprobe. Lanes 3 and 4: RNase-resistant reaction products from a hybridization of the antisense probe against total RNA from porcine liver. The two protected bands of lengths of 465 and 409 nucleotides, respectively, are indicated. Lane 5: DNA molecular weight marker (in bp). Lane 6: Untreated antisense run-off transcript. (B) Primer extension analysis. Lane 1: Electrophoretic analysis of extension products obtained with a 39 bp long primer derived from positions 73–112 within exon 2 (numbering system of Figure 6). The two products of lengths of 490 and 566 nucleotides are indicated. Lanes 2 and 3: DNA molecular weight markers. (C) A schematic representation of the experiment performed in (A).

protected from RNase digestion (Figure 3A). For the primer extension analysis, a 39 bp long antisense oligonucleotide covering positions +73 to +112 of the cDNA (numbering system of Figure 6) was used to prime DNA synthesis from total porcine liver RNA. Electrophoresis revealed the presence of two bands (Figure 3B) of lengths of 490 and 546 nucleotides. The two start sites around positions −427 and −380 (Figure 4) determined from two different techniques thus turned out to be identical.

The region upstream and around these two transcriptional start sites contains a variety of cis-sequence elements typical of class II gene promoters (Figure 4). These include four Sp1 sites (at positions −464, −536, −540, and −631) and two putative AP2 sites at positions −415 (CCCCATGC) and −467 (GCCCAGGG) which deviate from the consensus CCCCAGGC at one and two positions, respectively. The sequence TAATAATAA (position −444, Figure 4) shares homology with the consensus TATA box sequence. However, the abundance of start sites seems to indicate that it is not fulfilling its typical positioning function in this case. The first 355 bp of the promoter region display a G+C content of 68%, which is typical of constitutively expressed genes. The presence

```
 KpnI                                    SmaI
GGTACCGTGG AACCCCTTCT CCTGCCTGGC ACCCCGGGTC CTACATCTTT AAGGGGAAAG   -732

ACGCGCGCTC GCGCAGGGCC TCTGCACACG TCACTTCTTC ATTCTTTCCC TCCCCCCCCC   -672

CCTTGGCTCC TTTCTCCAAC TTCCCGGCCA GCTCGGGGCG GGTTGCGGGG GGGGGGGTGT   -612

CTTCAGAATC CACCTCCGGC CCCAAAATGA ATGTCCCGCA ACAGCCCACC CCTCCCCACT   -552

CGACTAGGGC GGGCGGTCCA GGGGGACCCG CTGTAGTTGG GGTGCGCGCG GCCCGTTGGT   -492
                                    TaqI                      ***
CCCCGTCTGG TGCCCCGGCC CAGGGCGGCT CGAAATTTGT AATAATAAGC CATGATGGGG   -432
                                                             ***
GTTCAGGAGG CATGGGGAGG GGGGAGCCGC TTTCCATCCC CTCTCGCGCT CTTTTTTCCC   -372

TCGTTTCTTT GAAAGTTGGA TGTTGAGAAG TGGGAGGTTC GGGGTGGGAG GGGAGAAAAA   -312
                                                            DdeI
TTGGGGGAGG GAGGGAGGAA GGAGGTTGCC AAAGGGAGGA GAGACAGAGA CGCTGAGAGA   -252

GACAGAGACA GAGGAGGAGA GAGGGAGGAG GGACAAGNAA GNAAGGGAGA GAGACAGAAA   -192

GAGGGAGAGA GAGAGGGAGT CGGGCGCAGG AGAGAGGGAG AGCGCGGGAG GGAGGAGGGA   -132
                                    NaeI
GGGAGACTGA GGGAGGAGGC GGCGAGGAGC GCGCCGGCCG GNCGGGGGGG GGGGGGGTGG    -72
                                                NaeI
TTTGGAAAAA TGACTCAGTA AGTTCAGCGC GCCCGCTNCG GCCGGCCCTG CGCCTCCCGC    -12
     SmaI  +1                                +30
CGCGCCCCGG GATGTATTCG TCCCCGCTCT GCCTGACCCA GGTACGCCCC CCGCCTGGCC     49

CCTGCCCGCG CCCCCTGCGC CCCAGCCCCG CGTCCCCCAT AGCCGGGCCC CGGAGTTTGG    109

GAGACAGAGG CGGGACGAAA AAGGCGCGCG TCCCTCCCGT GGCGGCGGCC GGAATGCCCC    169

CGCCACGGGC CTGGGTCCGG TGGGGGATGG GCTGGGCCCG GACACCCCCC CCCCCGCAAC    229

CCGGG
```

FIGURE 4: DNA sequences around the transcriptional start site of the porcine NFI/CTF gene. The sequence starts at a *Kpn*I site marking the 5′ end of a 1.8-kb *Kpn*I fragment located between 7 and 9 kb upstream of exon 2 (cf. Figure 1). Exon 1 is doubly underlined. Position +1 is the A residue in the start codon of exon 1. The two transcriptional start sites around positions −426 and −380 are marked by asterisks. Putative recognition sites for DNA binding factors, as discussed in the text, are marked in boldface type. Inverted repeats are underlined with a dotted line and restriction sites with a single line.

of 11 bp long contiguous stretches of either G or C residues indicates a potential for the formation of cruciform structures in this region. Work is in progress for studying the functional properties of these particular sequence elements.

Number of NFI/CTF Genes. Southern blot analyses have been performed to determine the total number of genes coding for NFI-like proteins. By use of a 423 bp long *Pst*I/*Sac*I probe derived from exon 2, at least two bands were detected in *Bam*HI, *Pst*I, and *Kpn*I digests of porcine genomic DNA (Figure 5). The ones of higher intensity in each of the digests are expected from the known structure of the NFI gene (Figure 1). The bands of lower intensity, e.g., the 0.6-kb band from the *Pst*I digest, were detected even under conditions of high stringency (tRNA as carrier). When a 3.3-kb genomic probe covering exons 4–6 was used, only the bands expected for the NFI/CTF gene could be detected under these conditions (not shown). We thus conclude that at least one additional NFI/CTF gene of considerable homology in the second exon but lesser homology in the more 3′-terminal exons must be present in porcine genomic DNA.

Structures of NFI cDNAs. cDNAs were isolated both from porcine liver and from HeLa cell cDNA libraries. They were designated as either p-CTFn (p for porcine) or h-CTFn (h for human). One clone from porcine liver and three from HeLa cells were characterized in detail. The single clone from porcine liver, NFI/p-CTF2, has a coding sequence of 1317 bp corresponding to 439 amino acids (Figure 6). The poly(A) part of the sequence cannot represent the actual 3′ terminus of the mRNA since these are considerably longer (see Discussion). Instead, the poly(A) stretch is derived from a genomic, A-rich stretch of DNA which is present in the 3′ untranslated region immediately following the 3′ end of the cDNA and which is apparently used as oligo(dT) primer binding site in reverse transcription.

The single porcine clone NFI/p-CTF2 lacks entire exon 1 (10 amino acids) and initiates with the first amino acid, an aspartic acid residue, of exon 2. A comparison with the three human CTF isolates CTF1, CTF2, and CTF3 of Santoro et al. (1988) demonstrates that it is identical in its exon composition with clone h-CTF2, although the latter lacks the first three amino acids of exon 2. The other two cDNA clones identified by Santoro et al. (1988) are distinguished from h-CTF2 or NFI/p-CTF2 by either the lack (h-CTF3) or the addition (h-CTF1) of protein sequences. These lacking or

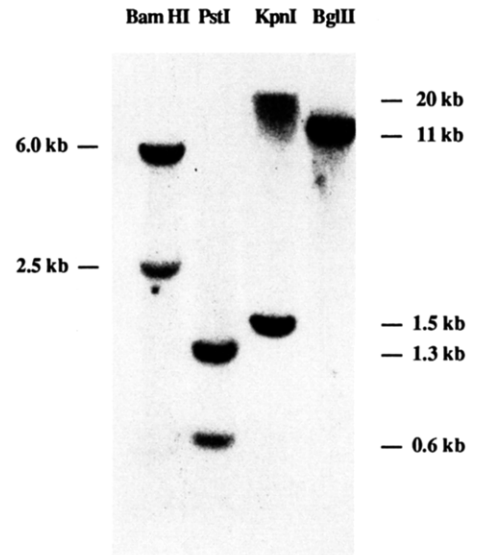

FIGURE 5: Southern blot analysis for the porcine NFI/CTF gene family. A 423 bp long *PstI*/*SacI* fragment covering parts of exon 2 (cf. Figure 6) was hybridized under stringent conditions with yeast tRNA as carrier against digests of porcine genomic DNA. The stronger signals (6 kb for *Bam*HI, 1.3 kb for *Pst*I, 1.5 kb for *Kpn*I, and 10 kb for *Bgl*II) are consistent with the known structure of the NFI/CTF gene (cf. Figure 1), while the weaker signals refer to a related but different gene.

additional stretches are represented exactly by exons 3 and 9, respectively, of the porcine NFI gene.

The deletion of exon 9 results in a change of reading frame in p-CTF2 or h-CTF2 as compared to that of h-CTF1 such that NFI/p-CTF2 terminates within 16 amino acids of exon 10 while h-CTF1 continues through exon 10 with a different reading frame until it reaches a UAA termination codon in exon 11.

The actual sequence differences between the human and the porcine NFI/CTF2s are minimal as the degree of identity amounts to 93.5% at the nucleotide and 98.4% at the protein level. This corresponds to a total number of nucleotide and amino acid differences of 85 (out of 1317) and 7 (out of 439), respectively.

We have been able to isolate three additional cDNA species from HeLa cell derived cDNA libraries, h-CTF4, h-CTF5 and h-CTF6 (Figure 7). h-CTF4 contains 42 bp of the untranslated leader, exons 1 and 2, and three amino acids of exon 3. The adjacent reading frame of a length of 10 amino acids has not yet been identified as a porcine genomic DNA sequence. It thus may represent an unrecognized, additional exon of the NFI gene. An *Eco*RI/*Pst*I DNA fragment derived from the untranslated leader of h-CTF4 and extending into exon 2 was used as a probe to identify a porcine genomic clone with the corresponding region of the porcine NFI gene. It is this sequence from which we identified the first exon and the leader of the gene depicted in Figures 1 and 6.

Clone h-CTF5 lacks a stretch of DNA covered exactly by exons 9 and 10 as well as a portion of 73 amino acids (out of 177) of exon 2. Finally, clone h-CTF6 is distinguished from the others by an entirely different 5′ end which is spliced to a reading frame beginning at codon 88 of NFI gene exon 2. The corresponding genomic sequence has not yet been identified.

In analyzing the protein sequence of NFI/p-CTF2, we could not detect any of the known sequence elements characteristic of DNA-binding proteins, i.e., zinc(II) fingers, helix–turn–helix motifs, leucine zippers, and/or glutamic acid rich stretches (Landschulz et al., 1988b; O'Shea et al., 1989; Courey & Tjian, 1988). However, a new sequence motif has been discovered that represents a stretch of arginine and lysine residues spaced at intervals of seven amino acids eventually terminating with a proline residue. The repeat begins with an arginine residue at position 37 of the cDNA (Figure 6) and continues via four lysine residues at positions 44, 51, 58, and 65. The proline residue at position 66 which interrupts this structure is only present in the human sequence such that it may well continue in the porcine sequence up to lysine 79. Since these elements are present in a region predicted by two algorithms (Chou & Fasman, 1974; Garnier et al., 1978) to have a high α-helix probability, we have named this element the "lysine helix" (Figure 8). Experiments are in progress for analyzing the functional significance of this observation.

Another sequence motif has been found within the coding region of the CTF family of proteins near the C-terminus at the junction of exons 9 and 10 (amino acids 469–476). It is characterized by a stretch of nine hydrophilic amino acids (SPTSPSYSP; Figure 6) which is identical in sequence with the canonical CT7 repeat of yeast and mouse RNA polymerases II (Sigler, 1988).

Functional Studies of NFI/p-CTF2. We have expressed the NFI/p-CTF2 protein in an in vitro translation system derived from rabbit reticulocytes. Porcine cDNA clone NFI/p-CTF2 with an appropriate eukaryotic translational start site (see Materials and Methods) but lacking the 10 amino acids of exon 1 was cloned into vector pGEM-3Z in order to produce sufficient amounts of mRNA in a reaction catalyzed by the SP6 polymerase. Following in vitro translation in the presence of [^{35}S]methionine, a labeled protein was formed that bound to a 146 bp long, unlabeled DNA fragment covering the origin of adenovirus type 5 DNA replication and carrying an NFI binding site, as shown by gel retention assays for an *Nde*I-derived truncated protein (Figure 9A). The specificity of this interaction was demonstrated by competition with a 5- or 25-fold molar excess of smaller, 35 bp long synthetic DNA fragments carrying either the same NFI binding site or the CCAAT box regions from the human α-globin or the HSV-tk promoter. As shown in Figure 9A, specific binding is demonstrated by a shift to faster moving DNA/protein complexes. This shift is complete in the case of the DNA fragment containing the NFI site while only a partial shift is observed with the α-globin and no shift with the HSV-tk fragment or the mutated NFI fragment. This behavior is indistinguishable from that of purified NFI protein prepared by DNA affinity chromatography from porcine liver (Meisterernst et al., 1988a; Thalmeier et al., 1989), indicating not only that NFI/p-CTF2 is indeed the analogue of NFI but also that the N-terminal 10 amino acids of the protein do not contribute to DNA binding and DNA specificity. Recent experiments with a full-length clone have confirmed these conclusions.

In order to locate the DNA binding domain, we analyzed the behavior of deletion constructs lacking various C-terminal portions of NFI/p-CTF2. These were obtained by linearization of the SP6 expression vector with the enzymes *Sac*I, *Bst*XI, and *Nde*I, respectively. These sites are marked in Figure 6. All translation assays led to the formation of the expected truncated proteins which were subsequently analyzed by gel retention in the presence of an oligonucleotide carrying the appropriate binding site. While the *Nde*I-derived run-off transcript, leading to a size reduction to 247 amino acids, was still fully active (Figure 9B, lanes 4–6), a 165 amino acid long truncated protein formed after *Sac*I digestion did not lead to an observable gel shift (not shown). An intermediate activity was observed with the 199 amino acid long derivative obtained

```
CTF1 h                                                1|2
CTF1 p   Met Tyr Ser Ser Pro Leu Cys Leu Thr Gln Asp Glu Phe His Pro Phe    16
CTF1 p   ATG TAT TCG TCC CCG CTC TGC CTG ACC CAG GAT GAG TTC CAC CCA TTC    48
CTF1 h                                    AT                   G
                                                                PstI
 Ile Glu Ala Leu Leu Pro His Val Arg Ala Phe Ala Tyr Thr Trp Phe Asn Leu    34
 ATC GAG GCG CTG CTG CCT CAC GTC CGC GCC TTC GCC TAC ACC TGG TTC AAC CTG    102
           C

 Gln Ala Arg Lys Arg Lys Tyr Phe Lys Lys His Glu Lys Arg Met Ser Lys Asp    52
 CAG GCG CGG AAG CGC AAG TAC TTC AAG AAG CAT GAG AAG CGG ATG TCA AAA GAT    156
                                       C                   G   G   C
                                                       Pro
 Glu Glu Arg Ala Val Lys Asp Glu Leu Leu Gly Glu Lys Ala Glu Val Lys Gln    70
 GAG GAG CGC GCG GTG AAG GAC GAG CTG CTG GGC GAG AAG GCC GAG GTC AAG CAG    210
           T                                       C

 Lys Trp Ala Ser Arg Leu Leu Ala Lys Leu Arg Lys Asp Ile Arg Pro Glu Cys    88
 AAG TGG GCG TCG CGG CTG CTG GCC AAG CTG CGC AAG GAC ATC CGG CCC GAA TGC    264
                                                             G
                         Ser
 Arg Glu Asp Phe Val Leu Ala Ile Thr Gly Lys Lys Ala Pro Gly Cys Val Leu    106
 CGC GAG GAC TTT GTG CTG GCC ATC ACC GGC AAG AAG GCG CCA GGC TGC GTG CTC    318
               C         AG                            G

 Ser Asn Pro Asp Gln Lys Gly Lys Met Arg Arg Ile Asp Cys Leu Arg Gln Ala    124
 TCC AAC CCC GAC CAG AAG GGC AAG ATG CGC CGC ATC GAC TGC CTG CGC CAG GCC    372
                                    G                   C   G       G

 Asp Lys Val Trp Arg Leu Asp Leu Val Met Val Ile Leu Phe Lys Gly Ile Pro    142
 GAC AAG GTG TGG CGG CTG GAC CTG GTC ATG GTC ATC CTC TTC AAG GGC ATC CCG    426
                                                G

 Leu Glu Ser Thr Asp Gly Glu Arg Leu Val Lys Ala Ala Gln Cys Gly His Pro    160
 CTG GAG AGC ACC GAC GGC GAG CGC CTG GTC AAG GCG GCA CAG TGC GGC CAC CCG    480
                                          T   G         T
                                                   SacI
 Val Leu Cys Val Gln Pro His His Ile Gly Val Ala Val Lys Glu Leu Asp Leu    178
 GTG CTC TGC GTG CAG CCA CAC CAC ATT GGG GTG GCG GTC AAG GAG CTC GAT CTC    534
   C   G              G               C       C                 G   C
                                    2|3                        SacII
 Tyr Leu Ala Tyr Phe Val Arg Glu Arg Asp Ala Glu Gln Ser Gly Ser Pro Arg    196
 TAC CTG GCC TAC TTC GTG CGC GAG CGA GAT GCA GAG CAG AGC GGC AGT CCG CGG    588
                          T                     A               C
 Thr                                          BstXI        3|4
 Ala Gly Met Gly Ser Asp Gln Glu Asp Ser Lys Pro Ile Thr Leu Asp Thr Thr    214
 GCA GGG ATG GGC TCC GAC CAG GAG GAC AGC AAG CCC ATC ACA CTG GAC ACC ACC    642
 A               T                                   G           G
                                                                SacI
 Asp Phe Gln Glu Ser Phe Val Thr Ser Gly Val Phe Ser Val Thr Glu Leu Ile    232
 GAC TTC CAG GAG AGC TTC GTC ACC TCT GGT GTG TTC AGT GTC ACT GAG CTC ATC    696
                    T           C   C           C
                4|5
 Gln Val Ser Arg Thr Pro Val Val Thr Gly Thr Gly Pro Asn Phe Ser Leu Gly    250
 CAA GTG TCC CGG ACA CCT GTG GTG ACT GGA ACA GGA CCC AAC TTC TCC CTG GGG    750
                     C
                     Leu NdeI                                   Leu
 Glu Leu Gln Gly His Met Ala Tyr Asp Leu Asn Pro Ala Ser Thr Gly Met Arg    268
 GAG CTT CAG GGG CAC ATG GCA TAT GAC CTA AAT CCA GCC AGC ACT GGC ATG AGA    804
       G             C         C       G   C                   C C
                                    5|6
 Arg Thr Leu Pro Ser Thr Ser Ser Ser Gly Ser Lys Arg His Lys Ser Gly Ser    286
 AGA ACG CTA CCC AGC ACT TCC TCC AGC GGG AGC AAG CGG CAC AAA TCG GGC TCG    858
           G           C           T

 Met Glu Glu Asp Val Asp Thr Ser Pro Gly Gly Asp Tyr Tyr Thr Ser Pro Ser    304
 ATG GAG GAA GAC GTG GAC ACG AGC CCC GGC GGC GAT TAC TAC ACC TCG CCC AGT    912
                                 T                       T           C
                     Ser                                    6|7
 Ser Pro Thr Ser Ser Asn Arg Asn Trp Thr Glu Asp Met Glu Gly Gly Ile Ser    322
 TCT CCC ACG AGT AGC AAC CGC AAC TGG ACG GAG GAC ATG GAA GGA GGC ATC TCA    966
   G                 G                                                G
```

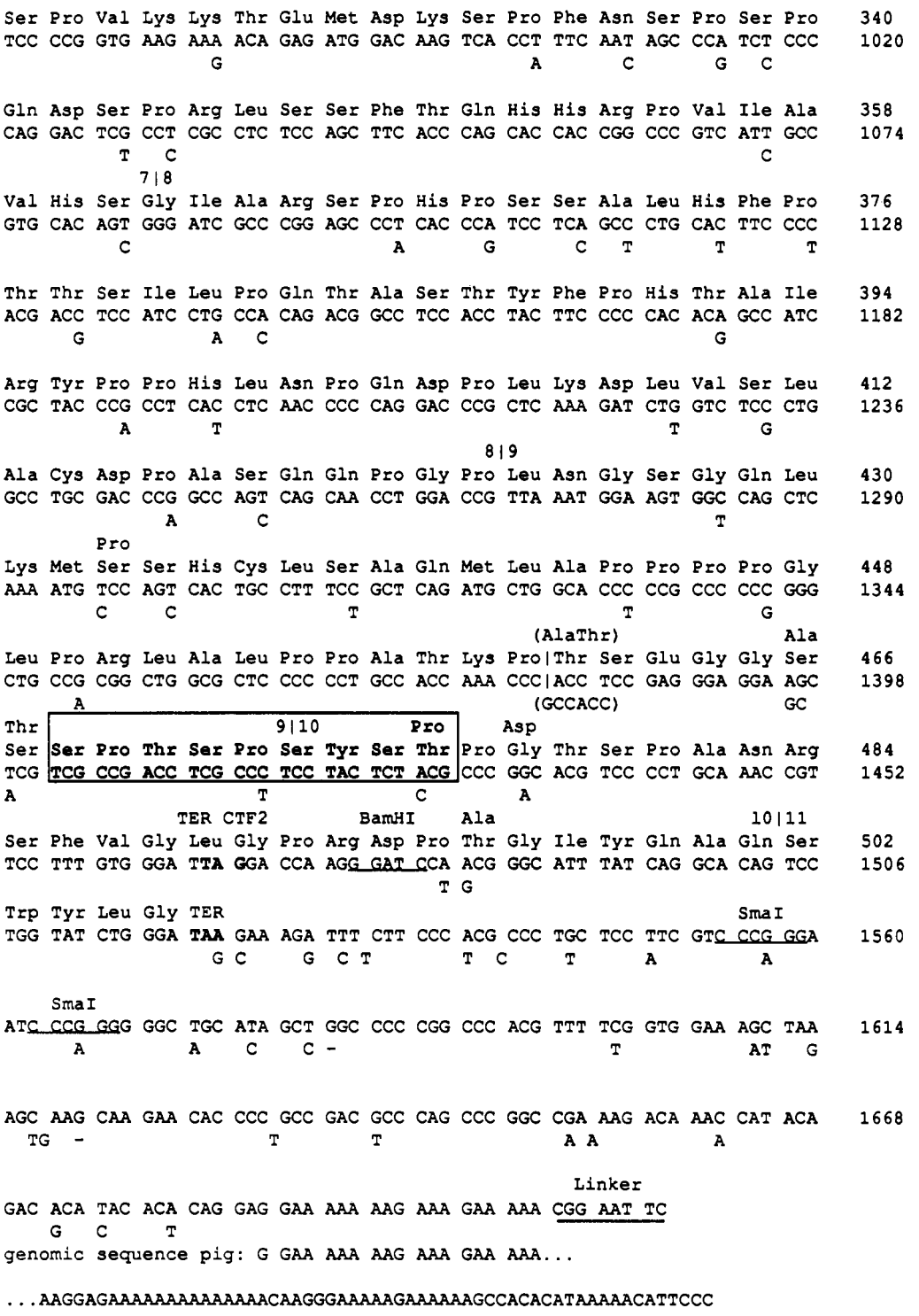

FIGURE 6: Sequence comparison of porcine (p) and human (h) CTF1. Exon/intron boundaries are indicated by vertical lines. The marked restriction sites were used in the preparation of deletion constructs. A black background indicates the positions of arginine and lysine residues present at seven amino acid intervals and identifying the "lysine helix". The CT7n box characteristic for carboxy-terminal regions of yeast and mouse RNA polymerases II is boxed and marked by boldface type. Clone CTF2 lacks exon 9. A concomitant change in reading frame leads to its termination at codon 16 (TAG) of exon 10, indicated as TER CTF2 and marked with boldface letters. A short stretch of porcine genomic DNA sequences which is apparently used as the oligo(dT) primer binding site in reverse transcription is displayed underneath the 3'-terminal base pairs of the cDNA clone.

by linearization with *Bst*XI. The DNA binding domain of NFI/p-CTF2 therefore is located within amino acids 11–257 of the protein molecule.

In order to analyze the stoichiometry of the NFI/DNA interaction, we performed cotranslation experiments with run-off transcripts derived from full-length or *Nde*I-truncated clones. In gel shift experiments with a 146 bp long DNA fragment carrying an NFI binding site, the ^{35}S-labeled protein permitted autoradiographic detection of three different complexes, namely, the two complexes observed with authentic full-length or truncated protein (marked as the homodimers AA and BB, respectively) and a third complex (marked as AB) of intermediate mobility (Figure 9B, lanes 7–9). Again, the specificity of the interactions was demonstrated by gel shifts to higher mobility of all three complexes in the presence of a 10-fold molar excess of a smaller, competing DNA fragment

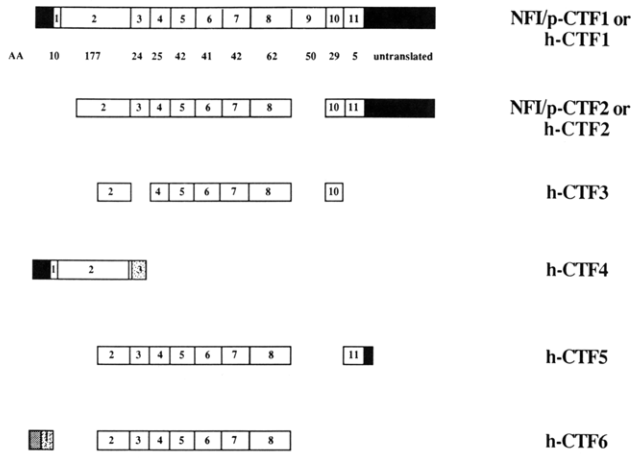

FIGURE 7: Schematic structure of CTF/NFI cDNA clones. Clones h-CTF1 to h-CTF3 from human HeLa cells were described by Santoro et al. (1988). Clones NFI/p-CTF2 from porcine liver and clones h-CTF4 to h-CTF6 from human HeLa cells were identified by us. Exons are numbered within the appropriate open rectangles, as indicated. Numbers underneath the structure of h-CTF1 represents the number of amino acids in the respective exons. The structure of h-CTF1 was supplemented with the 5′ untranslated region and the first exon found in h-CTF4. h-CTF4 contains the 5′ untranslated region as well as exons 1 and 2 followed by three amino acids of exon 3 linked to a short stretch of ten amino acids of unknown origin (dotted box). h-CTF5 lacks exons 9 and 10. h-CTF6 displays an entirely different 5′ untranslated region and first exon (stippled box) spliced to base 263 of exon 2. The changes in reading frames occuring through the various splice processes are described in the text.

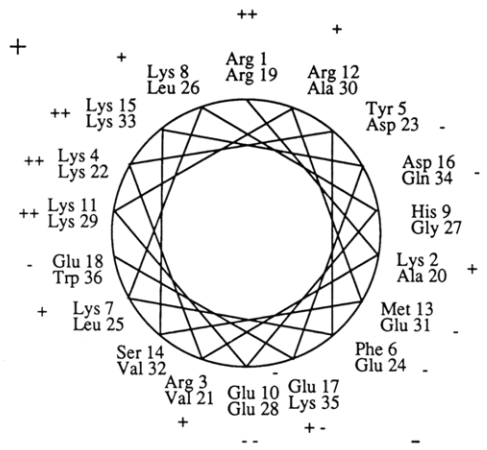

FIGURE 8: Putative NFI lysine helix. The α-helix presentation assumes a length of 3.6 amino acids per helical turn. The arginine residue designated as 1 refers to codon 37 of clone NFI/p-CTF2 (Figure 6). Charged amino acids are marked by (+) and (−) signs. A cluster of several positive charges on *one* side of the helical wheel is indicated by a large (+) sign.

(marked as AA*, BB*, or AB*, respectively). The simplest explanation for the occurrence of an intermediate band in this experiment would be that NFI/p-CTF2 binds as a dimer to its cognate DNA binding site. No intermediate band was found when the two in vitro translation products were only mixed after separate translation experiments, indicating the kinetic stability of the dimer complexes.

Upon SDS–polyacrylamide gel electrophoresis, the 247 amino acid long *Nde*I-derived protein displays about the same size of ~30 kDa as the protein purified from porcine liver by DNA affinity chromatography (not shown). This indicates that the purification from porcine liver does not yield an intact protein but rather a protease-resistant fragment carrying the DNA-binding domain.

Discussion

CCAAT binding factors are able to contribute both to basal and to enhancing activities required for, and involved in, class II gene transcription as well as to the initiation of adenovirus DNA replication. The molecular relationships between these different activities have not been elucidated. In an effort to unravel this problem, this paper describes the structure of a porcine gene which was originally isolated on the basis of peptide sequence information derived from one of the CCAAT-binding factors, adenovirus DNA replication initiation factor NFI.

The porcine NFI/CTF gene consists of at least 11 exons which are distributed along at least 70 kb of genomic DNA. DNA sequence data account for 478 out of a total of 501 codons defined by various cDNA clones. From the complete identity of porcine cDNA and of porcine genomic DNA sequences as well as from Southern blot analyses, we conclude that the two separate portions of the NFI/CTF gene representing exons 1 and 2 and exons 3–10, respectively, are part of a single transcription unit. The stretch of DNA representing the 24 codons of exon 3 has not yet been sequenced but was shown by Southern blot analysis to be located 3–5 kb upstream of exon 4.

In addition to the one gene described in detail, we have identified at least one other NFI-related gene in porcine genomic DNA. The reduced intensity of the autoradiographic bands in Southern blots (Figure 5) indicates that its sequence in the area covered by the probe is related but not identical. The existence of at least one additional CTF gene had also been inferred from a recent characterization of cDNA clones isolated from a hamster liver derived cDNA library (Gil et al., 1988). Although these hamster clones share an almost complete homology within the N-terminal 200 amino acids, they are distinguished from the human, rat, and porcine CTF cDNA clones by highly divergent 3′-terminal sequences and thus may well be derived from the second NFI/CTF gene identified in the present study by Southern blotting. This NFI/CTF-related gene has not been cloned and characterized as yet from porcine DNA. However, we have recently cloned and sequenced parts of such a NFI/CTF-related gene from human genomic DNA. Its second exon displays about 90% homology to the corresponding sequence in the human cDNA clones h-CTF (Santoro et al., 1988) while its pattern of base substitutions toward the 3′ terminus indicates a relationship to the hamster cDNA NFI/X mentioned above (Gil et al., 1988). Both human and porcine genomic DNAs thus contain at least two NFI genes which are highly homologous around their 5′ portions but which divert considerably in sequence toward their respective 3′ termini.

Low-stringency hybridization has revealed the existence of a number of additional genes, the significance of which however remains unclear. Since some of our human CTF cDNA clones, i.e., h-CTF4 and h-CTF6, contain coding regions that are not present in the NFI/CTF gene, we assume that these were derived from the related genes. This supports the hypothesis that it is a family of genes that encodes the CCAAT-binding family of proteins.

Three cDNAs compatible with the structure of the NFI/CTF gene and designated CTF1, CTF2, and CTF3 (Santoro et al., 1988) have been isolated previously from HeLa cells. This paper describes the structure of a CTF2-like cDNA from porcine liver as well as the structures of three novel cDNAs from HeLa cells. The various cDNAs are distinguished from each other by the presence, or absence, of sequences which correspond exactly to DNA stretches defined by exon 3, 9, or

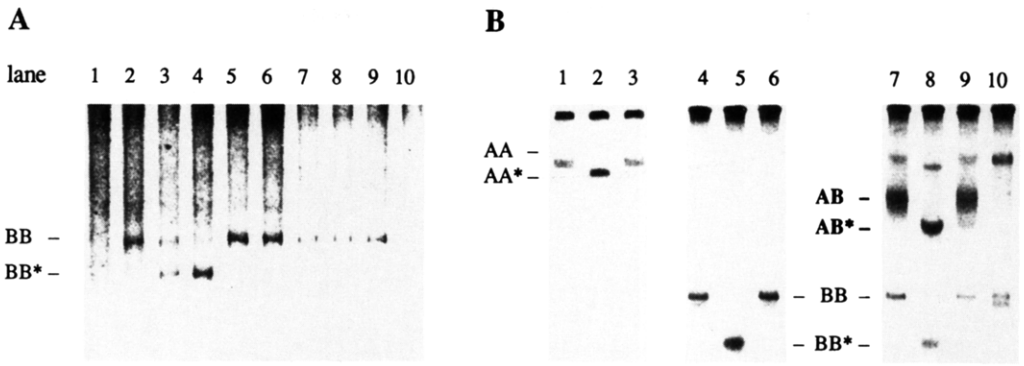

FIGURE 9: Functional analysis of the DNA-binding domain of NFI and the stoichiometry of the NFI/DNA interaction. Intact NFI/p-CTF2 protein and various carboxy-terminal deletions were synthesized in a rabbit reticulocyte in vitro translation system with SP6 run-off transcripts derived from cloned NFI/p-CTF2 or appropriate deletions as templates. In parts A and B, [^{35}S]methionine-labeled full-length or truncated proteins were incubated with an unlabeled 146 bp long DNA fragment (F 146) from the left-hand terminus of adenovirus type 5 DNA and analyzed in gel retention assays. Dimeric full-length and truncated complexes formed in the presence of fragment F146 are marked AA and BB, respectively. Complexes formed after competition with a 25-fold molar excess of a 32 bp long fragment carrying the adenovirus type 5 NFI site are marked AA* and BB*. (A) An *Nde*I-derived truncated protein was analyzed for binding specificity by competition of the 146-bp DNA fragment with 5- and 25-fold excesses of 32 bp long synthetic DNA fragments carrying various CCAAT box sequences (see text). Competition and thus specificity is demonstrated by a faster moving band indicative of the protein/DNA complex with the smaller, competing fragment. Lane 1, no DNA; lane 2; no competitor DNA; lane 3 and 4, Ad5-NFI oligo as competitor; lanes 5 and 6, HSV-tk CCAAT as competitor; lanes 7 and 8, α-globin CCAAT as competitor; lanes 9 and 10, Ad5-NFI mutant IV1/2 (with TGGN$_6$GCCAA mutated to TGGN$_6$GAACC). The sequence of competitor oligonucleotides has been described (Meisterernst et al., 1988a). (B) Comparison of the mobility of DNA/protein complexes formed with different, ^{35}S-labeled proteins and unlabeled DNA fragment F 146. Lane 1, complex AA with full-length NFI/p-CTF2; lane 2, complex AA*; lane 3, complex AA after competition with the α-globin CCAAT sequence containing oligo; lanes 4–6, competition experiments as in lanes 1–3 but for the truncated protein; lanes 7–9, cotranslation experiments with full-length and truncated mRNAs. An intermediate band interpreted to represent the heterodimer is marked as AB or AB* in the competition experiments. Competition experiments are in the same order as above. Lane 10, analysis of a mixture of full-length and truncated proteins after separate translation experiments. No heterodimeric band is visible.

10, respectively, in porcine genomic DNA. h-CTF1, for example, carries an additional stretch of 50 amino acids as compared to h-CTF2, which is encoded by exon 9. Similarly, h-CTF3 and h-CTF5 lack stretches of 24 amino acids and of 79 amino acids coded for by exons 9 and 10, respectively. These observations are strongly indicative for the occurrence of alternate splicing during processing of the NFI/CTF mRNA precursor as was indeed postulated by Santoro et al. (1988). Strictly speaking, however, this conclusion is only correct for the porcine mRNA species, i.e., for clone NFI/p-CTF2, since in principle exon/intron boundaries may well be different for the porcine and the human gene. The fact however that, for example, sequence deletions in clones h-CTF2 and h-CTF3 coincide exactly with exons in the porcine gene strongly argues for an identical exon/intron organization of the human NFI/CTF mRNA precursor.

In the case of the exon 3 deletion, as observed in clone h-CTF3, the splicing process leaves an intact translational reading frame, while exon 9 deletion (in p-CTF2 and h-CTF-2) results in a (+1) shift of the reading frame. The reading frame of CTF2 thus terminates within 16 amino acids (out of a total of 50) of exon 10. A simultaneous deletion of exons 9 and 10 as observed in clone h-CTF5 retains the CTF1 reading frame which, in turn, includes the 50 amino acids of exon 10 and the 5 amino acids of exon 11. The possible functional significance of the predicted protein sequence differences of the various cDNA clones is under study.

The postulated differential splicing process mainly affects exons 3, 9, and 10. The question thus arises whether these exons contain sequences of functional relevance which might distinguish the different proteins formed from the different cDNAs. We have identified such a sequence of possible functional significance around the junction of exons 9 and 10. It is a stretch of nine hydrophilic amino acids (SPTSPSYSP) which is completely homologous to the consensus sequence of the canonical CT7 repeat observed in the carboxy-terminal regions of the largest subunits of yeast and mouse RNA polymerase II. This sequence which occurs in multiple repeats in these protein molecules has been suggested to represent a target site of cellular trans-acting factors (Sigler, 1988) and may well serve a similar function in CTF1. Interestingly enough, this site covers the boundary of exons 9 and 10 and is thus entirely deleted in NFI/p-CTF2 due to the lack of exon 9 as well as a shift in reading frame. Experiments are under way to study whether and how this sequence element may affect and distinguish CTF1 and CTF2 function.

The only cDNA that we were able to isolate from porcine liver cDNA libraries is a species encoding a CTF2-like protein, while different cDNAs were obtained from HeLa cell cDNA libraries. In agreement with these observations, Northern blots from total porcine liver RNA yielded a major band of 4.5-kb length with some minor signals at 6.5 and 8.0 kb, while a similar analysis with total RNA from HeLa cells identified three major RNA species of lengths of 4.5, 6.5, and 8.0 kb (Meisterernst et al., 1988b). The observed heterogeneity of CTF proteins in HeLa cells (Jones et al., 1987) and the presence of only a single protein in porcine liver thus is readily explained, although we cannot exclude the possibility that this diversity of mRNAs may be caused by the presence of different genes rather than by cell-type specificity.

Functional analyses of the porcine NFI cDNA clone p-CTF2 were performed in rabbit reticulocyte extracts. By using in vitro synthesized mRNA species of different lengths, we have shown that a protein comprising amino acids 11–257 displays full DNA binding activity. Since it was shown by Santoro et al. (1988) that h-CTF3 which lacks exon 3 binds as well to DNA as h-CTF2, amino acids 189–212 appear unnecessary for DNA binding. Nevertheless we feel that at least parts of exon 4 may contribute to the DNA-binding domain since C-terminal deletions which reduce the size of the protein to approximately 200 amino acids are only partially active in DNA binding.

Within the large DNA-binding domain, starting with amino acid 37, we noted a novel structural element consisting of

several lysine and arginine residues spaced at seven amino acid intervals which we termed the "lysine-helix". While lysine-rich stretches are, in general, typical and common for DNA binding proteins, e.g., histones, a computer search has failed to identify another DNA-binding protein with such a regular repeat structure. This structure extends over nine α-helical turns in a region which is predicted to display α-helicity. Presented on a helical wheel, it leads to a cluster of basic residues which may contact backbone phosphates of the DNA. Whether this cluster contributes to, and plays a role in, specific DNA recognition remains to be seen.

By performing in vitro cotranslation experiments with full-length and truncated templates, we could clearly demonstrate that NFI/p-CTF2 binds as a dimer to its DNA binding site. The dimerization domain however has yet to be identified.

ACKNOWLEDGMENTS

We are grateful to Dr. R. Mertz for preparing oligonucleotides and to Dr. H. Ibelgaufts for helping with the preparation of the manuscript. Claudia Mosch has kindly provided her data on the structure of a human NFI/CTF gene. We thank Matthias Müller and Bertram Brenig for many helpful discussions and for supplying us with magnificent preparations of porcine genomic DNA.

REFERENCES

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1987) *Current Protocols in Molecular Biology*, Chapter 4.8.1, Wiley, New York.

Barberis, A., Superti-Furga, G., & Busslinger, M. (1987) *Cell 50*, 347–359.

Borgmeyer, U., Nowock, J., & Sippel, A. (1984) *Nucleic Acids Res. 12*, 4295–4311.

Chodosh, L. A., Baldwin, A. S., Carthew, R. W., & Sharp, P. A. (1988) *Cell 53*, 11–24.

Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 222–245.

Church, G., & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1991–1995.

Cohen, R. B., Sheffery, M., & Kim, C. G. (1986) *Mol. Cell. Biol. 6*, 821–832.

Colantuoni, V., Pirozzi, A., Blance, C., & Cortese, R. (1987) *EMBO J. 6*, 631–636.

Courey, A. J., & Tjian, R. (1988) *Cell 55*, 887–898.

Dorn, A., Bollekens, J., Staub, A. Benoist, Ch., & Mathis, D. (1987) *Cell 50*, 863–872.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *Mol. Biol. 120*, 97–120.

Gil, G., Smith, J. R., Goldstein, J. L., Slaughter, C. A., Orth, K., Brown, M. S., & Osborne, T. F. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8963–8967.

Jones, K. A., Yamamoto, K. R., & Tjian, R. (1985) *Cell 42*, 559–572.

Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J., & Tjian, R. (1987) *Cell 48*, 79–89.

Jones, K. A., Luciw, P. A., & Duchange, N. (1988) *Genes Dev. 2*, 1101–1114.

Kozak, M. (1984) *Nucleic Acids Res. 12*, 857–872.

Landschulz, W. H., Johnson, P. H., Adashi, E. Y., Graves, B. J., & McKnight, S. L. (1988a) *Genes Dev. 2*, 786–800.

Landschulz, W. H., Johnson, P. H., & McKnight, S. L. (1988b) *Science 240*, 1759–1764.

Lichsteiner, S., Wuarin, J., & Schibler, U. (1987) *Cell 51*, 963–973.

Maity, S. N., Golumbek, P. T., Karsenty, G., & de Crombrugghe, B. (1988) *Science 241*, 582–585.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Meisterernst, M., Gander, I., Rogge, L., & Winnacker, E. L. (1988a) *Nucleic Acids. Res. 18*, 4419–4435.

Meisterernst, M., Rogge, L., Donath, C., Gander, I., Lottspeich, F., Mertz, R., Dobner, Th., Foeckler, R., Stelzer, G., & Winnacker, E. L. (1988b) *FEBS Lett. 236*, 27–37.

Myers, R. M., Tilly, K., & Maniatis, T. (1986) *Science 232*, 613–618.

O'Shea, E. K., Rutkowski, R., & Kim, P. S. (1989) *Science 243*, 538–543.

Paonessa, G., Gounari, R., Frank, R., & Cortese, R. (1988) *EMBO J. 7*, 3115–3123.

Pelletier, J., & Sonnenberg, N. (1985) *Cell 40*, 515–526.

Raymondjean, M., Cereghini, S., & Yaniv, M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 757–761.

Santoro, C., Mermod, N., Andrews, P. C., & Tjian, R. (1988) *Nature 334*, 218–224.

Sigler, P. B. (1988) *Nature 333*, 210–212.

Superti-Furga, G., Barberis, A., Schaffner, G., & Busslinger, M. (1988) *EMBO J. 7*, 3099–3107.

Thalmeier, K., Synovzik, M., Mertz, R., Winnacker, E. L., & Lipp, M. (1989) *Genes Dev. 3*, 527–536.