Volume 236, number 1, 27–32 FEB 06185 August 1988

Isolation and characterization of the porcine nuclear factor I (NFI) gene

Michael Meisterernst, Lars Rogge, Cornelia Donath, Irene Gander, Friedrich Lottspeich*, Ronald Mertz*, Thomas Dobner, Renate Föckler, Gertraud Stelzer and Ernst-L. Winnacker

Institut für Biochemie der Universität München, Karlstrasse 23, D-8000 München 2 and **Genzentrum der Universität München am Max-Planck Institut für Biochemie, D-8033 Martinsried bei München, FRG*

Received 15 June 1988

This study describes the isolation of a major portion of the gene for nuclear factor I (NFI) including its 5'-flanking region with transcriptional start sites. We screened a porcine liver, genomic DNA library in phage EMBL3A with synthetic oligonucleotides derived from tryptic and cyanogen-bromide peptide sequences obtained from purified NFI protein. The NFI gene is present as a single copy in porcine DNA.

DNA replication; Adenovirus DNA; Nuclear factor I gene; CCAAT-binding factor

1. INTRODUCTION

Nuclear factor I (NFI) was originally identified as a factor required for the efficient initiation of adenovirus type 2/5 DNA replication [1]. Subsequently it was shown to mediate its effect by binding to a dyad-symmetry binding site within the 102 bp long inverted terminal repetition of the viral DNA [2]. Additional binding sites have been found within the genomes of other DNA viruses [3,4] and in cellular, chromosomal DNA [5–7]. It was pointed out by Jones et al. [8] that one half site of the consensus sequence TGG(A/C)N_5CCAA represents the CCAAT box, identified as a promoter element in a variety of eukaryotic promoters [9,10]. This observation suggests a similarity or even an identity of NFI with the CCAAT-element binding transcription factor. However, the situation appears considerably more complex. Chodosh et al. [11] have shown recently that at least three different CCAAT-binding proteins with different, non-overlapping specificities to various CCAAT elements can be isolated from HeLa cell nuclear extracts. In addition, Meisterernst et al. [12] have demonstrated that a highly purified NFI preparation from porcine liver does not bind with significant equilibrium binding constants to CCAAT elements from various eukaryotic promoter elements. Therefore it remains unclear whether there are a number of distinct CCAAT proteins derived from different genes, or whether they represent a family of closely related, but functionally slightly different, proteins derived from a single transcriptional unit by differential splicing events and/or different post-translational modifications.

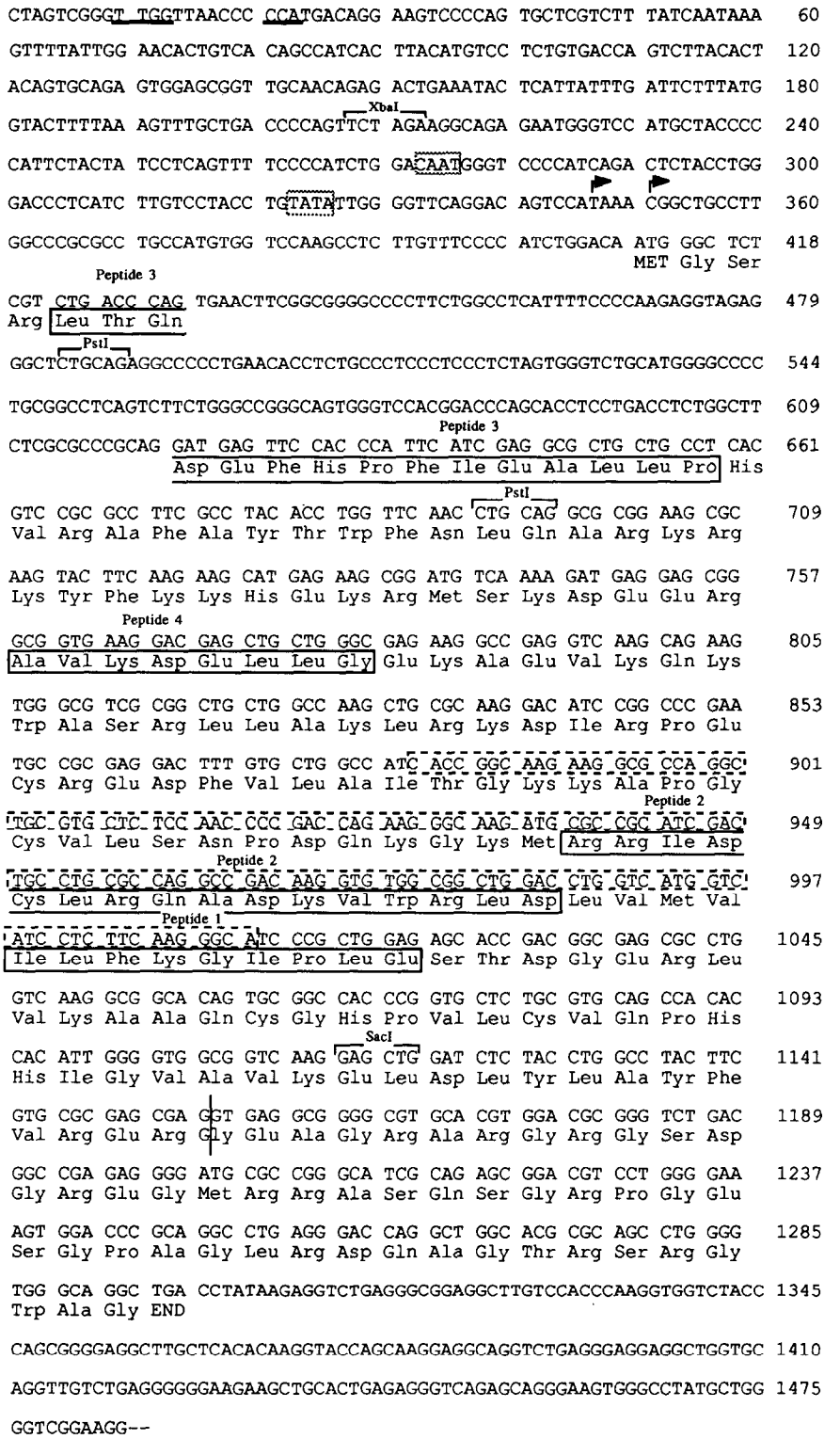
As a prerequisite for further functional studies of NFI we cloned the corresponding gene.

Correspondence address: M. Meisterernst, Institut für Biochemie der Universität München, Karlstrasse 23, D-8000 München 2, FRG

2. MATERIALS AND METHODS

NFI was purified from freshly prepared porcine liver nuclear extracts [12,13]. Approximately 20 μg (500 pmol) of a homogeneous NFI preparation was lyophilised and treated in 70% formic acid with a 50-fold molar excess of cyanogen bromide (0.5 mg/ml) in the dark at room temperature for 1 h. Trypsin digestion of the cyanogen bromide-cleaved NFI was performed in 0.1 M ammonium hydrogen carbonate at a protein concentration of 30 ng/μl after addition of 1/40 of the probe weight of trypsin (0.5 ng). After digestion at 37°C for 18 h the peptide mixture was acidified to 0.1% trifluoroacetic acid and separated by HPLC reversed-phase chromatography

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50


XbaI
Peptide 3
PstI
Peptide 3
PstI
Peptide 4
Peptide 2
Peptide 2
Peptide 1
SacI

on LiChrospher columns at 0.5 ml/min in a gradient of 0–60% acetonitrile in 0.1% trifluoroacetic acid. Sequences were obtained by automated Edman degradation on an Applied Biosystems model 470A sequencer as described [14].

Phage from an EMBL3A library of porcine liver genomic DNA were plated at a density of 50000 plaques per 15 cm petri dish. After transfer onto nitrocellulose filters (Schleicher and Schüll BA85) they were marked and prepared for hybridization as described [15]. The following three probes were used:

Probe 1: 5′-AT$^{T}_{C}$CTGTT$^{T}_{C}$AAGGG$^{C}_{G}$AT$^{T}_{C}$CC$^{C}_{G}$CTGGA-3′

5′-AT$^{TT}_{CC}$TGTT$^{T}_{C}$AA$^{A}_{G}$GGCAT$^{T}_{C}$CCC$^{T}_{C}$TGGA-3′

The corresponding CNBr-peptide (P1) sequence was: ILFKGIPLE.

Probe 2: 5′-GTCCCAGGCGCCACACCTTGTCAGCCTGC-

CGCAGGCAGTCAATGCGCCG-3′

The corresponding CNBr-peptide (P2) sequence was: RRIDCLRQADKVWRLD.

Probe 3: 5′-GC$^{T}_{C}$TCAAT$^{A}_{G}$AANGG$^{A}_{G}$TG$^{A}_{G}$AA$^{C}_{T}$TCATC-3′

5′-GC$^{T}_{C}$TCTAT$^{A}_{G}$AANGG$^{A}_{G}$TG$^{A}_{G}$AA$^{C}_{T}$TCATC-3′

The corresponding tryptic peptide (P3) sequence was: LTQDEFHPFIEALLP.

Screening with mixed probes 1 and 3 (1–2 pmol each of ^{32}P-labelled oligonucleotide) was performed according to [15] at a temperature of 42°C; long probe 2 was hybridized in 1 × SSC, 1 × Denhardt's solution and 100 μg/ml tRNA (yeast) at 42°C. Washing was performed for 10 min at room temperature, 1 × 10 min at 50°C and 1 × 10 min at 60°C in 6 × SSC for probes 1 and 3, and in 1 × SSC for long probe 2.

Double-stranded DNA sequencing was performed as described [15]. Procedures for the isolation of total RNA and of mRNA were taken from [15]. Southern blots were performed according to [15] and Northern blots according to [16]. Transcript mapping by an S_1 nuclease protection assay was performed according to [17].

3. RESULTS

NFI was purified from porcine liver crude nuclear extracts by various chromatography steps including DEAE-cellulose, MonoQ anion-exchange chromatography and sequence-specific DNA-affinity chromatography [12]. After enrichment to a factor of approx. 100000, the protein preparation migrated in SDS-polyacrylamide gel electrophoresis as a single molecular species with a molecular mass of 36 kDa [12]. Since the amino-terminus of the protein was blocked, it was subjected to CNBr and/or trypsin digestion in order to obtain partial peptide sequences by sequential Edman degradation. Synthetic oligonucleotides were derived either as mixed or as long probes; in the latter case they were optimized according to human codon usage [18]. Three probes (see section 2) were employed to screen a porcine genomic library in phage EMBL3A. Clones which lit up simultaneously with all these probes were enriched and rescreened. Out of 1.2×10^6 clones screened a total of six were identified which hybridized to all three probes. All clones share only a number of restriction fragments (not shown) and thus appear to be independent isolates of the same genomic sequence.

The DNA sequence of a 3.7 kb *Bgl*II fragment was determined from a set of *Hpa*II subclones in pUC12. A selected portion (1500 bp) of this sequence is shown in fig.1. It contains the sequence of the three probes used for the original screening and a sequence obtained from an additional tryptic peptide (P4; AVKDELL). All these sequences are present in a single long reading frame extending from positions 623 to 1301, corresponding to 224 codons. The three N-terminal amino acids of peptide P3 however could not be detected in this reading frame. Instead, they were found together with a preceding arginine residue (expected from the tryptic origin of this peptide) in a reading frame extending between positions 350 and 430. The only Met codon in this reading frame is situated at position 410, defining an N-terminal sequence of seven amino acids. The reading frame in this region is therefore interrupted by an intron of 192 bp between positions 431 to 622. This intron itself is defined by a tryptic peptide which overlaps the two splice sites.

The remote possibility of the four N-terminal amino acids to occur elsewhere in the genome, i.e. further upstream in this genomic DNA sequence, was ruled out by S_1 nuclease protection assays (fig.2) which place the transcriptional start site of the gene at positions 347 and 351, i.e. immediately upstream of the expected Met codon discussed above.

Fig.1. Nucleotide sequence from the porcine NFI gene. Arrows point to the location of transcriptional start sites. Boxes mark the four peptides obtained from NFI by tryptic and/or cyanogen-bromide digestion. The vertical bar indicates the putative splice site between exons 2 and 3. Dotted boxes enclose putative CAAT or TATA boxes, the dashed box a stretch of DNA between positions 880 and 1013 with 67% identity to the chicken histone 2B gene. A potential NFI site in the 5′-flanking region is underlined.

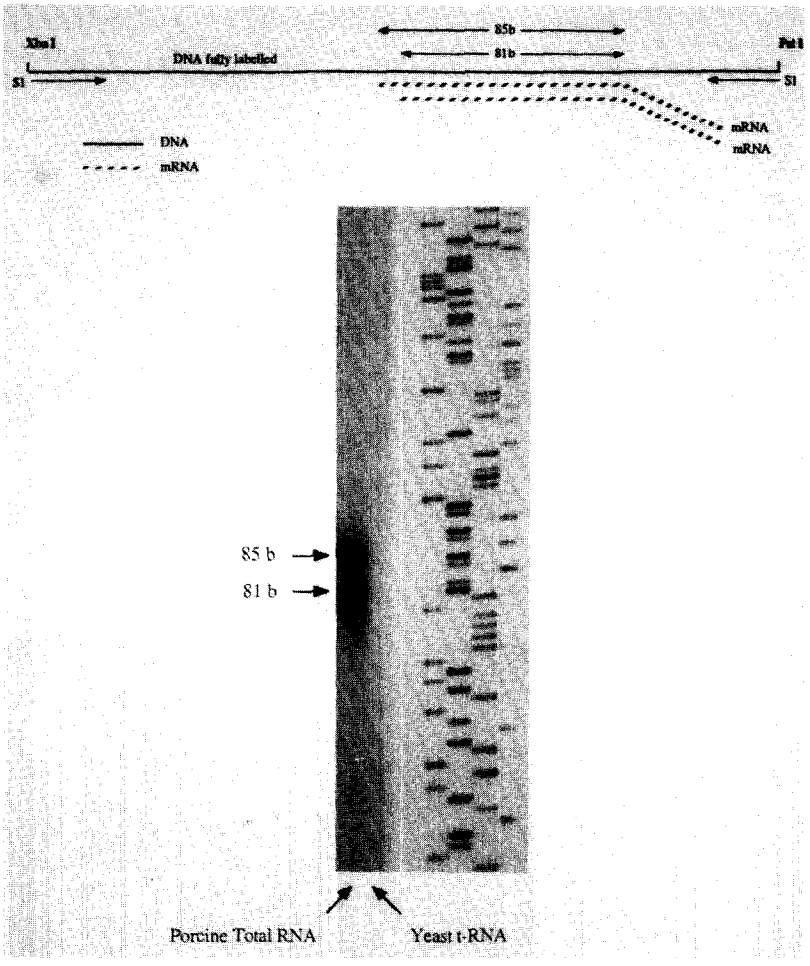


Fig.2. Mapping of transcriptional start sites through the S_1 nuclease protection assay. A schematic representation of the experiment in which a fully labelled, 280 bp *XbaI/PstI* fragment from the expected 5′-flanking region was hybridized to cytoplasmic RNA from porcine liver is shown in the upper part. The lower part displays an autoradiograph of the polyacrylamide gel which separated the products remaining after S_1 nuclease digestion. Size markers are the four reaction products from a sequencing reaction run on the same gel. Product sizes are 81 and 85 bp, respectively.

The sequence described in this communication covers a total of 231 amino acids and thus cannot represent the entire protein (36 kDa). We thus propose the existence of at least one additional intron and exon region. Considering the known consensus splice region sequence (GTA/GAGT; [16]) as well as the fact that the following sequence stretch contains multiple stop codons in all three reading frames we assume at present that the 5′-splice site for this intron is located at position 1154 (fig.1; vertical bar). This leaves a total of 184 amino acids identified so far for the NFI protein. We are currently searching our EMBL3A library for the lacking 3′-portion of the NFI gene which should comprise at least an additional 150 codons.

A 1.3 kb *PstI* fragment covering most of exon 2 and extending into intron 2 and a 279 bp *HpaII* fragment from within exon 2 were used as probes in Southern and Northern blot analyses, respectively. As shown in the autoradiograph of the Southern blot (fig.3), we detected only one single fragment in porcine DNA after digestion with several restriction enzymes which have no cleavage site in the probe DNA sequence. In Northern blots

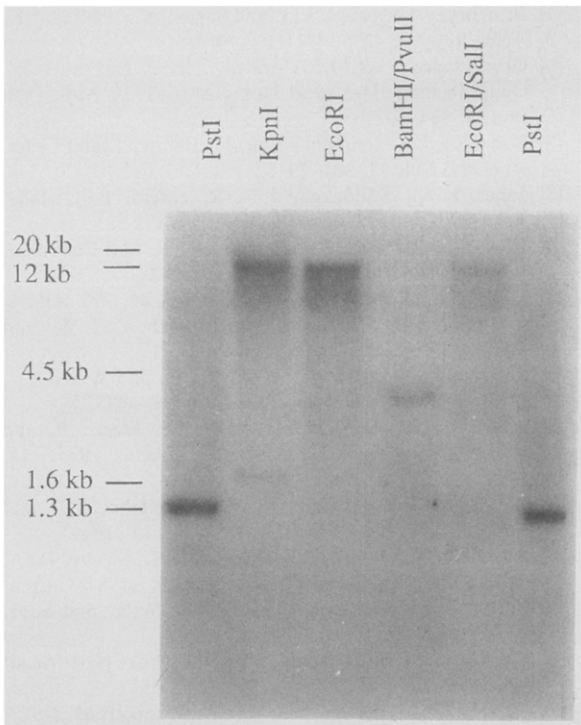


Fig.3. Genomic Southern blot analysis of the porcine NFI gene. The probe was a 1.3 kb *Pst*I fragment covering parts of exon 2 and intron 2.

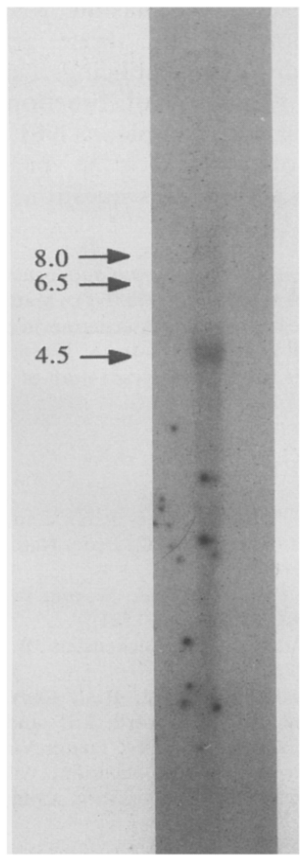


Fig.4. Northern blot analysis of cytoplasmic poly(A) containing RNA from porcine liver. In addition to a stronger band of 4.5 kb, two minor bands at 6.8 and 8 kb are clearly visible. The probe was a 279 bp *Hpa*II fragment extending from position 565 to position 844.

of porcine liver poly(A) selected cytoplasmic RNA (fig.4), a strong band of 4.5 and two minor bands of 6.8 and 8 kb could be detected.

4. DISCUSSION

This paper describes the DNA sequence of a genomic clone from porcine liver DNA which contains a significant portion of the porcine NFI gene. This conclusion is mainly confirmed by the fact that the sequences of four cyanogen-bromide and tryptic peptides obtained from a homogeneous and active NFI preparation could be identified in a single, long reading frame covering more than half of the expected protein sequence. As confirmed by S_1 nuclease protection assays, our clone also contains the transcriptional start site and 5′-flanking regions of the NFI gene.

The following points should be raised. Firstly, the putative promoter region not only contains the expected TATA and CCAAT-box elements but, in addition, reasonably strong NFI binding sites (cf. fig.1). It remains to be seen, whether these are involved in transcriptional regulation of this gene, i.e. in autoregulation. Secondly, the larger of the two exons contains a stretch of DNA between positions 880 and 1013 which displays a 65% degree of identity with the chicken histone H2B gene [19]. Since this homology does not extend to the protein level its significance remains unknown. Thirdly, we want to emphasize that the 5′-splice site of the first intron, which is defined by an overlapping cyanogen-bromide peptide, contains the dinucleotide TG rather than GT as described for consensus 5′-splice sites [20]. The significance of this variation is not known. Finally, we have identified at least three mRNA species in Northern blots of

total porcine liver cytoplasmic RNA, indicating that expression of this single gene might be regulated by differential splicing. This in turn may result in the formation of functionally different proteins acting, for example, as NFI or as CCAAT binding factors. Work is in progress in our laboratory to address this question.

Acknowledgements: This study was supported by the Deutsche Forschungsgemeinschaft (Fa 138/3-1). Matthias Müller and Peter Groitl have been extremely generous in providing porcine DNA and valuable technical advice. We are grateful to Dr Ibelgaufts for his help in the preparation of this manuscript.

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, vol.4, pp.525–542, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[5] Borgmeyer, U., Nowock, J. and Sippel, A. (1984) Nucleic Acids Res. 12, 4295–4311.
[6] Gronostajski, R.M., Adhya, S., Nagata, K., Guggenheimer, R.A. and Hurwitz, J. (1985) Mol. Cell. Biol. 5, 964–971.
[7] Siebenlist, U., Henninghausen, L., Battey, J. and Leder, P. (1984) Cell 37, 381–391.
[8] Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) Cell 48, 79–89.
[9] Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127–142.
[10] McKnight, S. and Tjian, R. (1986) Cell 46, 795–805.
[11] Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) Cell 53, 11–24.
[12] Meisterernst, M., Gander, I., Rogge, L. and Winnacker, E.L. (1988) Nucleic Acids Res. 16, 4419–4435.
[13] Schneider, R., Gander, I., Muller, U., Mertz, R. and Winnacker, E.L. (1986) Nucleic Acids Res. 14, 1303–1317.
[14] Eckerskorn, C., Mewes, W., Goretzki, H. and Lottspeich, F. (1988) Eur. J. Biochem., in press.
[15] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G. and Struhl, K. (1987) Current Protocols in Molecular Biology, John Wiley and Sons, New York.
[16] Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.
[17] Calzone, F.J., Britten, R.J. and Davidson, E.M. (1987) Methods Enzymol. 152, 611–632.
[18] Lathe, R. (1985) J. Mol. Biol. 183, 1–12.
[19] Harvey, R.P., Robins, A.J. and Wells, J.R.E. (1982) Nucleic Acids Res. 10, 7851–7863.
[20] Mount, S.M. (1982) Nucleic Acids Res. 10, 459–472.