J. Biochem. 122, 795-801 (1997)

Genomic Organization of the Rat Nuclear Factor I-A Gene[1]

Mingxu Xu, Shigehiro Osada, Masayoshi Imagawa, and Tsutomu Nishihara[2]

Laboratory of Environmental Biochemistry, School of Pharmaceutical Sciences, Osaka University, 1-6 Yamada-Oka, Suita, Osaka 565

Received for publication, May 29, 1997

The nuclear factor 1 (NF1) protein family functions as a cellular transcription factor as well as an adenovirus DNA replication factor. This family consists of four subtypes, NFI-A, NFI-B, NFI-C, and NFI-X, each encoded by a different gene. Each subtype possesses different isoforms generated by alternative splicing. To date, only a porcine NFI-C gene has been cloned, and the gene structures of the other NF1 proteins have not yet been identified. We recently isolated four kinds of NFI-A cDNA clones from the rat liver. To gain additional insight into the structure of NFI-A, we isolated the rat NFI-A gene. This gene is composed of 11 exons spanning over 70 kb. All of the exon/intron boundaries are consistent with the GT/AG rule, and consensus sequences surrounding the splice boundaries are also found. The 5′-flanking region lacks a canonical TATA box, but contains several GC-box and AP2 binding sites. A 5′-rapid amplification of cDNA end analysis indicated that the transcription of the NFI-A gene is initiated at multiple sites. We also found conservation in the genomic structure between the rat NFI-A and the porcine NFI-C, suggesting that duplication of an ancestral gene occurred rather recently to produce the NFI-A and NFI-C genes.

Key words: exon/intron, gene cloning, genomic structure, nuclear factor 1, promoter.

Nuclear factor 1 (NF1) was originally described as a DNA binding protein purified from HeLa cells that stimulates the replication of adenovirus DNA *in vitro* (*1, 2*). It was subsequently demonstrated that NF1 also activates transcription for the promoters of the human α-globin (*3*), human papillomavirus type 16 (*4*), and mouse myelin basic protein genes (*5*), and inactivates the transcription of the rat glutathione transferase P gene (*6*).

In vertebrates, the NF1 family of proteins consists of four subtypes, NFI-A, NFI-B, NFI-C, and NFI-X, each of which is encoded by a different gene (*7, 8*). In addition, each subtype possesses different isoforms generated by alternative splicing (*5-7, 9*). All of the NF1 polypeptides identified thus far display a conserved amino-terminal domain and can be distinguished through their divergent carboxy-terminal regions. The amino-terminal parts of NF1 proteins contain the domain for DNA-binding, dimerization, and replication functions, while the carboxy-terminal domain is involved in the regulation of transcription (*10, 11*).

We recently identified four kinds of NFI-A cDNA clones (NFI-A1, NFI-A2, NFI-A3, and NFI-A4) from rat liver (*6*). NFI-A1 had no deleted region compared with NFI-A2, NFI-A3, and NFI-A4. It is unclear which exons are deleted in the latter three splicing isoforms, since only a porcine NFI-C gene has been cloned, and the gene structures of other NF1 proteins are not yet known (*12*).

To gain additional insight into the structure of NFI-A as well as to approach the question of how NFI-A gene expression is regulated, it is necessary to determine the rat NFI-A gene structure. Here we present the complete genomic organization of the rat NFI-A gene, including the complete exon/intron boundaries and the 5′-flanking region. The results of our investigation revealed a gene comprised of 11 exons and spanning more than 70 kb of genomic DNA. The putative promoter was located within a G+C-rich region which lacks a canonical TATA box and contains several Sp1 binding sites. Our 5′-rapid amplification of cDNA end (RACE) assay revealed multiple transcription initiation sites, a typical characteristic of a TATA-less gene promoter as found in housekeeping genes.

MATERIALS AND METHODS

Screening of Genomic Library and Sequence Analysis—An EMBL3 SP6/T7 genomic DNA library from adult Sprague-Dawley male rat liver (Clontech Lab., Palo Alto, CA, USA) was probed with the rat NFI-A cDNA fragments generated by restriction enzyme digestion or polymerase chain reaction (PCR). To isolate the entire rat NFI-A gene, 3.3×10^6 recombinant phage clones were screened. The probes were labeled using a *Bca*BEST Labeling Kit (TaKaRa, Kyoto) and [α-^{32}P]dCTP (Amersham, Buckinghamshire, England). Positive clones were purified, and the DNA fragment from each clone was isolated. The DNA fragments including exons were identified by Southern blot analysis and cloned into pBluescript vector. The nucleotide sequence was determined by an automated DNA sequencer DSQ 1000 (Shimadzu, Kyoto) and the dideoxy method

[1] This work was supported in part by grants from the Ministry of Education, Science, Sports and Culture of Japan and from Nissan Science Foundation. The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GeneBank nucleotide sequence databases under the accession number AB005129.

[2] To whom correspondence should be addressed. Tel: +81-6-879-8240, Fax: +81-6-879-8244, E-mail: nisihara@phs.osaka-u.ac.jp

Abbreviations: LA PCR, long and accurate polymerase chain reaction; NF1, nuclear factor 1; RACE, rapid amplification of cDNA end.

using ^{32}P (*13*).

Long and Accurate PCR (LA PCR)—LA PCR was employed to determine the intron sizes. The reaction was performed on 1 ng of phage DNA or 250 ng of genomic DNA from rat liver in 50 μl containing 1×LA PCR buffer, 2.5 μM $MgCl_2$, 400 μM dNTP, 0.2 μM each primer, and 2.5 units of LA Taq DNA polymerase (TaKaRa). The PCR was run for 30 cycles as follows: denaturing at 94°C for 1 min;

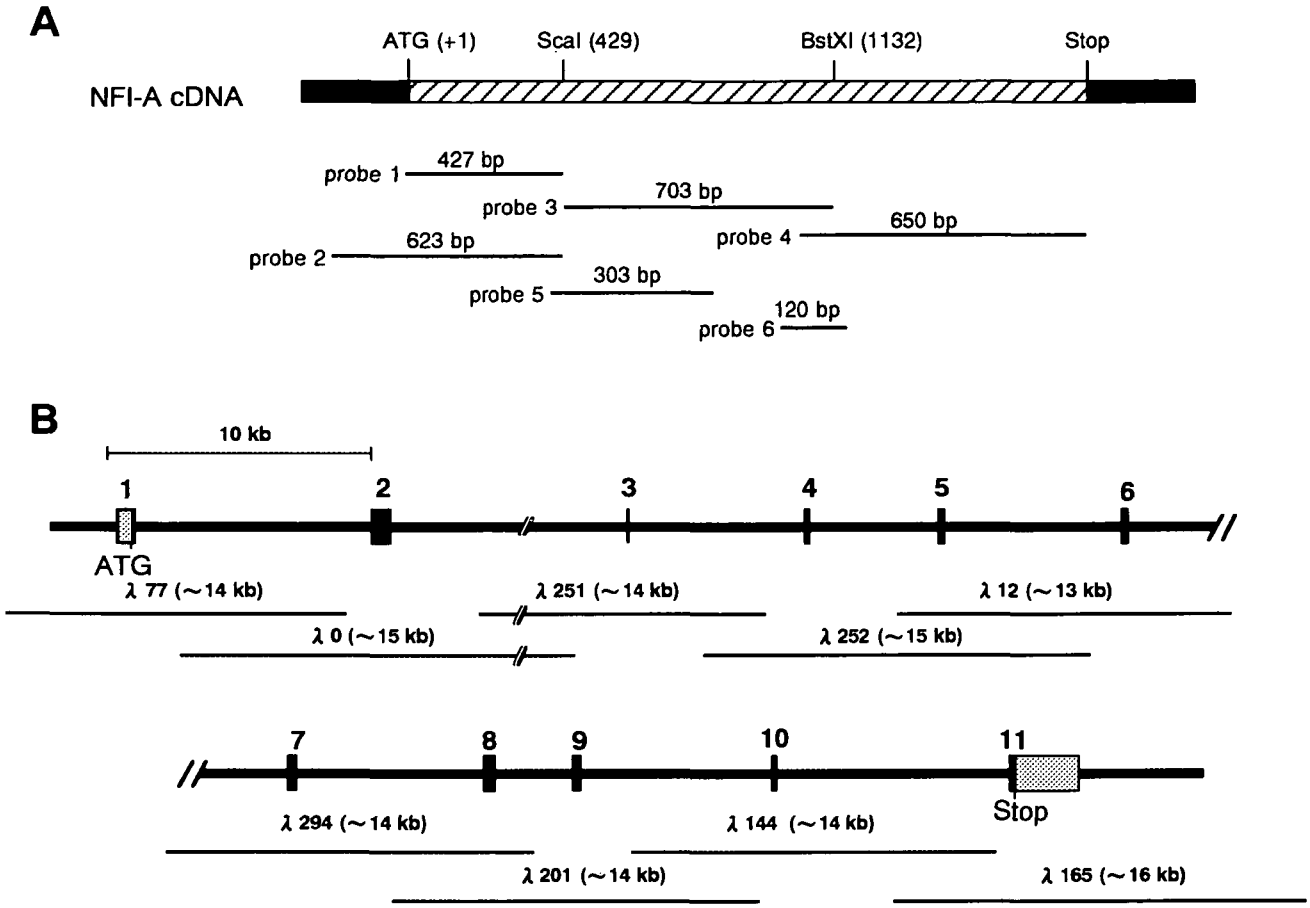


Fig. 1. **Structure and organization of the gene encoding rat NFI-A.** (A) Schematic representation of the probes used for the screening of the rat NFI-A gene. The upper panel shows a diagram of the cDNA isolated previously (*6*). The slanting-striped box represents the open reading frame and the closed boxes indicate 5′- and 3′-untranslated regions, respectively. The lower panel shows the positions of the probes for screening which were generated by the restriction enzyme digestion of the NFI-A cDNA (probes 1, 2, and 3) or PCR (probes 4, 5, and 6). (B) Schematic representation of the rat NFI-A genomic structure. The boxes show the exons with the relative size and the positions in the rat NFI-A gene. Numerals above the boxes indicate exon numbers. The DNA fragments isolated from nine overlapping phage clones are represented in the lower part of the figure. The checkered boxes show the 5′- and 3′-untranslated regions.

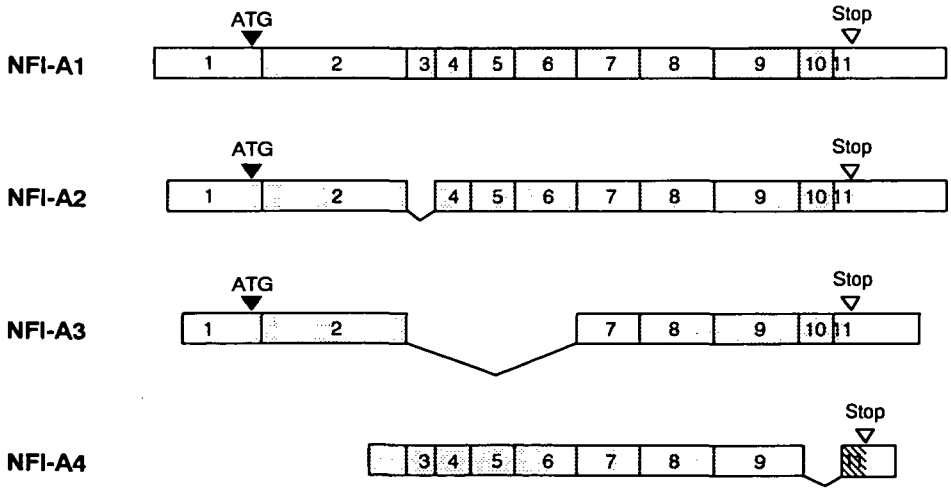


Fig. 2. **Schematic representation of the rat NFI-A cDNA isoforms.** Numerals indicate the exons. The checkered boxes show exons, and the open boxes indicate 5′- and 3′-untranslated regions. The frame-shifted region in NFI-A4 is marked by a slanting-striped box. The 5′-terminal portion of NFI-A4 has not been isolated yet. Filled and open triangles represent the start and stop points of translation, respectively.

annealing at 59-62°C (depending on the primer length) for 1 min; and extension at 72°C for 2 or 4 min (depending on the product length). The products were applied to 0.7% agarose gels, and their sizes were estimated by comparison

TABLE I. **Oligonucleotides used for determination of NFI-A gene structure.**

	Sequence (5→3)	GC%	Position[a]
S-91	CCGAGTTGGAAATGTGAACGCA	50	−155−−33
S-81	AGACAGAGCGGAGAATACAT	45	1-20
S-72	TCGGGTCGGATATCTTTCCG	55	237-257
S-88	GTCATGGTGATCTTGTT	41	394-410
S-92	AGTCAATCTGAAAGTCCCAGCC	50	564-586
S-93	TCTGGCTGGTCCTTAATGTCAG	50	599-620
S-94	ACATTTGGGCTTCCAGGACAGC	54	627-648
S-95	TGACACTCTTACTAGCTCAGTC	45	675-696
S-63	AGGAACCGGCCCCAATTTTT	45	714-733
S-96	TGGTAGAGGATGTGCTGGGTAG	54	795-817
S-89	GGTGAAGAACCATTTTACACAGGCC	48	865-889
S-97	GGCTCTACTTCATGCCATCCGC	59	923-944
S-98	ATGCCATCTCCAACCACTCTGA	50	949-970
S-99	TCCTGTTATGACAGGTCGGTGA	50	1047-1068
S-101	CATCGACTCTCCACTTTCCAACGTC	52	1094-1118
S-102	GGCAGACAAGTTGGACAAACTCCTTC	50	1188-1213
S-103	GTTCCTCAATCCCAATGGAAGCAGT	48	1245-1269
S-104	AGGCTGCACCGCCTTCTGTTGATGTG	57	1380-1405
S-105	CGCAAACCGATTCGTCAGTGTTGGA	52	1446-1470
S-106	GTCTGTTGAGGGATATTTACAAAGC	40	1484-1508
S-107	GGTACCTGGGATAAAAGTTGCAGCG	52	1513-1541
S- 64	AGGATGGTGGGACGCTGCAA	60	1534-1553
S-108	GTTGAGGTTGCGTCCATGTTACAGA	48	1583-1607

[a]Position +1 is the A residue of the translation start codon.

TABLE II. **Nucleotide sequences of the exon/intron boundaries in the rat NFI-A gene.**[a]

No.	Exon size (bp)	5′ splice donor		Intron size (kb)	3′ splice acceptor	
1	320	ACCCAG	gtaact	~ 9.0	ccctag	GATGAG
2	532	CAGCAG	gtaagt	~20	tttcag	GATTCA
3	66	AAAATG	gtaagt	~ 7.0	ccacag	GACATT
4	75	CACAAA	gtaagt	~ 5.5	ctctag	CACCAA
5	118	TACCAG	gtaatt	~ 6.6	tttcag	CTCTAG
6	128	AGCCAG	gtgagc	~ 8.0	tgccag	GCATGC
7	129	CGACAG	gtagtt	~ 6.0	ttgtag	CAAGCC
8	179	CTCAAT	gtaagg	~ 2.3	ttgcag	CCCAAT
9	166	CACCGA	gtgagt	~ 6.5	tcacag	CCTACT
10	91	GACACA	gtgggc	~ 8.0	ttccag	GTCCTG
11	>150	CTGGGA	—			

[a]The nucleotide sequence of each exon/intron boundary and the sizes of the exon and the intron are shown. Exon sequences are in capital letters; intron sequences are in lower case letters.

TABLE III. **Frequency of nucleotides at the exon/intron boundaries in the rat NFI-A gene.**[a]

Donor frequencies

	Exon				Intron					
	+4	+3	+2	+1	−1	−2	−3	−4	−5	−6
G	1	0	1	6	10	0	3	2	7	1
A	2	4	7	3	0	0	7	8	0	0
T	0	0	1	1	0	10	0	0	2	7
C	7	6	1	0	0	0	0	0	1	2
NFI-A	C	A/C	A	G	G	T	A	A	G	T
Consensus	N	A/C	A	G	G	T	R	A	G	T

Acceptor frequencies

	Intron																Exon			
	−17	−16	−15	−14	−13	−12	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	+1	+2	+3	+4
G	1	2	0	1	0	0	2	1	2	0	0	0	1	2	0	0	10	5	0	0	2
A	0	1	2	0	0	0	1	2	0	0	0	0	0	2	0	10	0	0	5	2	3
T	9	6	4	7	6	8	5	5	6	6	8	7	6	2	3	0	0	0	2	3	3
C	0	1	4	2	4	2	2	1	2	4	2	3	3	4	7	0	0	5	3	5	2
NFI-A	T	T	Y	Y	Y	T	Y	T	Y	Y	Y	Y	Y	N	Y	A	G	G/C	N	N	N
Consensus	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	A	G	G	N	N	N

[a]The frequencies of the different nucleotides at the exon/intron boundaries of the rat NFI-A gene are compared with the consensus sequence (*20*). Splice junctions are between −1 and +1. N, any nucleotide; R, purine; Y, pyrimidine.

with the size markers of λ DNA and pBluescript digested with *Hin*dIII and *Hpa*II, respectively.

Analysis of the Transcription Start Site—The transcription start site was determined by 5′-RACE PCR according to the manufacturer's instructions (Clontech, Palo Alto, California, USA). Five micrograms of total RNA extracted from rat liver was used as the substrate for the first-strand cDNA synthesis. Second-strand synthesis was performed and then an adapter was attached. The PCR was then performed using the adapter-specific primer 1 (AP1) and oligonucleotide S-123 as described below.

AP1: 5′-CCATCCTAATACGACTCACTATAGGGC-3′
S-123: 5′-CGCTGAACTTTAACCCCGCCTAGAGTTT-CC-3′

Plasmid Construction—The fragment containing −597 to +130 was inserted into the *Xho*I site and the *Hin*dIII site in a promoter-less luciferase vector, PGV-B (Toyo Ink Mfg., Tokyo) according to the standard protocol (*14*). The constructs used here were checked by sequencing according to the dideoxy method using denatured plasmid templates (*13*).

Cell Culture, Transfection, and Luciferase Assay—HeLa cells were cultured in minimal essential medium (MEM), containing 10% fetal bovine serum (FBS). The cells were transfected by the calcium phosphate co-precipitation technique described by Chen and Okayama (*15*) using 5.5 μg of reporter plasmid DNA and 0.5 μg of pRSVGAL, a eukaryotic expression vector that contains the *Escherichia coli* β-galactosidase structural gene controlled by the Rous sarcoma virus long-terminal repeat, as an internal control. The cells were harvested after a 40-h incubation following the transfection, and the luciferase activities were determined with Pikka Gene (Toyo Ink). The transfection experiment was performed three times using two different preparations of DNA, and the results are presented as mean values for the luciferase activity relative to β-galactosidase activity with standard deviations. The activity of β-galactosidase was assayed as described (*16*).

RNA Isolation and Northern Blot Analysis—Total RNA from the rat brain was extracted using TRIzol (Gibco BRL Life Technologies, Grand Island, NY, USA). For the Northern blot analysis, 30 μg of total RNA was electrophoresed on a 1.0% agarose gel containing 2% formaldehyde, and then transferred to a nitrocellulose filter (Schleicher & Schuell, Dassel, Germany). The upstream regions of rat NFI-A genomic DNA were used as a probe for the hybridization.

RESULTS AND DISCUSSION

Isolation of the NFI-A Genomic Clones—Our initial screening of an EMBL3 SP/T7 library of rat genomic DNA using a 439-bp-long fragment of the *Eco*RI/*Sca*I fragment as a hybridization probe (which corresponds to nucleotide −6/+433 of the NFI-A cDNA sequence from the ATG translation start codon) (Fig. 1A) identified one positive clone, λ0. This clone, however, only contained exon 2 of the NFI-A gene based on the result of sequencing. To isolate the entire rat NFI-A gene, five other DNA fragments corresponding to different regions of NFI-A cDNA (Fig. 1A) were employed, and about 300 positive clones were obtained. According to the results of PCR analyses using oligonucleotide primers which were based on estimated exon sequences, eight clones were further analyzed (Fig. 1B). The genomic structure was investigated by subcloning phage DNA into pBluescript. The exon/intron boundaries were sequenced by using the exon-specific oligonucleotide primers shown in Table I. The intron sizes were estimated

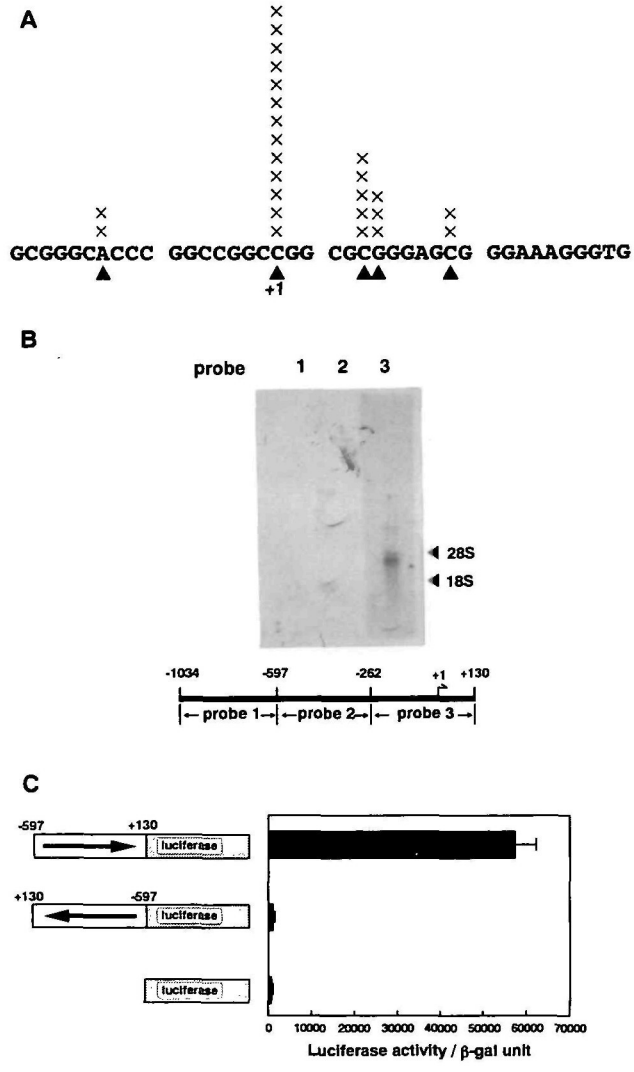


Fig. 3. **Mapping of the transcription start sites of the NFI-A gene.** (A) The result of sequencing of 5′-RACE PCR products. The 5′ ends of 25 clones are shown by X on the NFI-A gene sequence. Since 13 of the 25 clones were localized at the C residue at 293 bp upstream from the translation start codon, this nucleotide was designated the transcription start site of the NFI-A gene. (B) Northern blotting analysis of NFI-A mRNA. Three probes which covered the transcription start site and a further upstream region are shown in the lower panel. The first and second probes covered −1034 to −598 and −597 to −263 of the NFI-A gene, respectively (lanes 1 and 2). The third probe covered −262 to +130 of this gene (lane 3), which contains the assigned transcription start site. (C) Promoter activity of the 5′-flanking region of the NFI-A gene. The left and right panels show the schematic structures of the constructs and the results of the luciferase assay, respectively. The fragment containing −597 to +130 was subcloned into PGV-B, a promoter-less luciferase vector, in normal orientation as well as reverse orientation. These plasmids were transfected into HeLa cells and the luciferase activities were determined. The data are the mean values for the luciferase activity relative to β-galactosidase activity with standard deviations.

by combining the results of the LA PCR and Southern blotting.

According to the NFI-A cDNA and the genomic DNA sequences as well as the results of the LA PCR and Southern blot analyses, we concluded that the entire NFI-A gene had been isolated. The comparison of the genomic DNA and the NFI-A cDNA suggested that the rat NFI-A gene consists of 11 exons spanning approximately 70 kb (Fig. 1B). The exon of the NFI-A gene varied in size from 66 to 532 bp. Exon 2 is the largest exon (532 bp), corresponding to a part of the NFI-A DNA binding domain. The average length of the remaining exons is 135 bp, similar to the average length of eukaryotic exons of 140 bp (*17*). The introns range in size from 2.8 to 20 kb. All of the exon/intron boundary sequences are consistent with the GT/AG rule (Table II) (*18, 19*). The consensus sequences surrounding the splice boundaries in RNA polymerase II transcribed genes are also found to be in agreement with the exon/intron boundary of the NFI-A gene (Table III). Such an extreme conservation strongly suggests that a functional splice junction must possess a good match for the consensus, and that a component of the splicing mechanism binds at consensus sequences in the 5′ to 3′ direction until an AG is found (*20*).

Structure of Rat NFI-A cDNA—We have isolated four kinds of NFI-A cDNA clones from rat liver as described previously (*6*). They were designated NFI-A1, NFI-A2, NFI-A3, and NFI-A4 (Fig. 2). Our comparison of the genomic DNA of rat NFI-A with those of these NFI-A cDNA clones revealed that NFI-A1 had no deleted region, consisting of all of the exons. NFI-A2 and NFI-A3, however, lacked exon 3, and exons 3 to 6, respectively, leaving the protein coding frame unchanged. In contrast, NFI-A4 lacked exon 10, leading to a frame shift. Since the 5′-

A

```
  GGATCCAA ACAAGGTACC CTTCTAGCGT GTGTCTGGGG AACAGGGGAC TCAAATGTTA GCCCATGTCT AGTCCTGGG CTCCCATTTGG  -949
                                                                                  NF-IL6
GGATACAAAC ATATCCCTTT TCTAAGGGGG TAACTCTTTC ACGGTTTTCC TTCTCAGTTA TACAGAAAAA ATGTCCCAGC TAAACCACTG  -859
                                                                                  NF-IL6
TTGTTTCTGA CGTGACAGAA CTTTTTTGGG CGTTTGTGTT CAAACCCACC AGCTCCTGAG TTCTGGAAGA GAAGAGCGGT CGCTCACGAC  -769
                                                                                  NF-IL6
TCTCCGCCTT CCTTTTTCCA GATTCCCCCC CCCCCCCGCA ATCTGGGGAC GCTTTGGGGT GATCTTACTA GTCCCCCGCA CCCAGATTCT  -679
                          AP2      AP2
GACTGCTATG GGGCCGTAGC TGTAAAGTTA GTAGGCTTTT GTGCATATTA CCTCTGCGGA CGACACTATT CCCACAGAAA GCTTCGCAGG  -589
                                                                                  Oct
AGAGATTACA ATGCTTTGAG CTTTACCGCA TTTCAGAGGA AGGGGGGAAA AGTTACCTAG AAGGTCTGAT TTTTACAAGA AATTTGCATA  -499
Oct
CATGCAAATG TACCGCGCGC CCTTCTCTGA GGTTCCTAGG GAGCCCACCC GGCGGCCCAC GGGGGTGCGG GGCGCCCGGA CAAAGTTGCC  -409
                                    AP2      AP2
CCAATCGTCT ACCGACGTCC GCGGAGGGCT GCGGCGCGGG TGGGGGGATG GGGAGGAGGA AAAGTAGTTT TGTTTGCTTA AGCACATCCT  -319
CCAAT                                                                             HNF-5
ATGGCAGCCT ATAGCCGCCC GGAGGACACG GTGAAAACCG CTCCCCCCGC GTTCCCGGGA CCGCGCGCGA GGGGAGGGGG CGGGGACCGA  -229
           Sp1                                      AP2                                  Sp1
GGGAGGCAGC GACCGCGGAG CAAGCGACGG GCGGGCGGGC GTGCAGACTG GGGACCGGTG TCAGCGCCGC AGCCGCCCCT CCCCCACCCG  -139
                          Sp1                                                Sp1      AP2
GCTCCCCCCA CCCCACCCCA CCCCATCCCC CCCACCCCCA CCCCCCGCGG CGGCGGCGCG AGCGGGCGGC GGCCGTGCGG TGCGGTGCAG  -49
AP2       AP2       AP2       AP2 +1                                    Sp1
AACGCGGCGC AGGCGGGCGG CGGGCAGCTC GGCGGGCACC CGGCCGGCCG GCGCGGGAGC GGGAAAGGGT GCGCTATGCC TTTAACGCCC  +42
          Sp1                                ▲
GCGTACAGTA GACAGTATAG TGGAGTGTAG GGAAACTCTA GGCGGGGTTA AAGTTCAGCG CGTGGAGCGG CAAGAGCGCT GGCTGCTGCA  +132
GTTGAGCCGA GTTGGAAATG TGAACGCAAG AAGCAGGCTT GATTTTTTTT TCCTCCCCCC TCTCCTAGCT CTCTCCTCCT CTCTCTCCCT  +222
CTCGCACCCA CACTCACGCA CACCTCCAGC CCGCACACAG ACGTGCACGC ACCCCCCGCC GCCCGGCAGT TATG                    +296
```

B

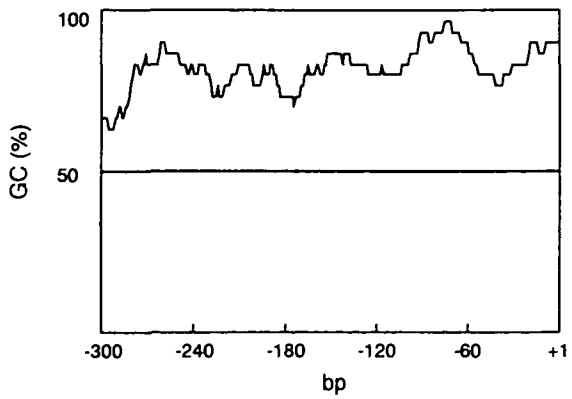


Fig. 4. **Nucleotide sequence and GC content of the upstream region of the rat NFI-A gene.** (A) Nucleotide sequence of the promoter region with putative *cis*-acting regulatory elements including six Sp1-binding sites (double underlined) and several other elements (underlined). The major transcription start site and the translation start codon are shown as +1 and in boldface, respectively. (B) Percentage distribution of G+C along the upstream region.

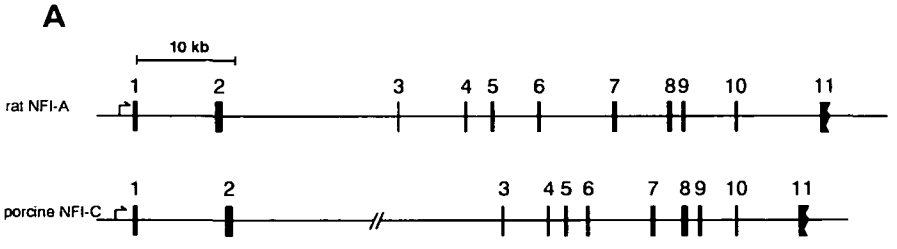
terminal end of this cDNA has not been isolated yet, it is unclear whether NFI-A4 has the entire exons 1 and 2 as found in the other three kinds of NFI-A. The question thus arose as to whether these exons contain sequences of functional relevance which might distinguish the different proteins formed from the different cDNAs. However, luciferase assay of the carboxy terminal of the NFI-A splicing variants which were fused to the DNA-binding domain of GAL4 (1-147) revealed that the carboxy-terminals of all of the NFI-As could function as repressors of transcription in a DNA binding site-dependent manner, and that the differences in repression activities were indistinguishable among the NFI-As (*6*). DNA binding assays have revealed that the amino-terminal region, including exons 1-4, of the NF1 proteins is involved in the DNA-binding activity (*21, 22*). Since the rat NFI-A splicing variants NFI-A2 and NFI-A3 lack a part of this region, they may possess lower DNA-binding activity than that of the full-length NFI-A1.

Identification of Transcription Start Site—We performed a 5′-RACE PCR analysis to determine the transcription start sites, because our attempts to do so by primer extension and S1 nuclease protection analyses were unsuccessful. The cDNA was synthesized using total RNA from rat liver. The oligonucleotides S-123 located at position 221 bp upstream from the ATG translation start codon and AP1 were used for the PCR amplification of the cDNA as primers. The amplified fragments (about 150 bp) were subcloned into pBluescript vector. Twenty-five independent clones were sequenced to determine the 5′ ends of the products. We found that the 5′ ends of 13 of the 25 clones were localized at 293 bp upstream of the cytosine residue from the ATG translation start codon, and those of an additional 12 clones surrounded this nucleotide (Fig. 3A). We therefore concluded that the former nucleotide might be the main transcription start site (+1). However, additional transcription start sites might exist nearby, because multiple 5′ ends of the 5′-RACE PCR products were determined.

The transcription start sites estimated above were confirmed by Northern blotting analysis using three gemonic fragments corresponding to the positions −1034 bp/−598 bp, −597 bp/−263 bp, and −262 bp/+130 bp upstream from that of the estimated transcription start site, respectively (Fig. 3B). If the estimation of the transcription start site is correct, only the third probe would hybridize to the NFI-A mRNA. As anticipated, only probe 3 containing the transcription start site hybridized with the rat brain RNA, in which NFI-A mRNA is expressed at a high level.

Next, the results of the 5′-RACE PCR were confirmed by the luciferase assay of the promoter activity of the 5′-flanking region. In general, the −400−−300 bp fragments from the transcription start site were used to determine the promoter activity (*23*). Hence, the fragment corresponding to positions −597 bp/+130 bp from that of the estimated transcription start site was used in our experiment. This fragment was joined to a promoter-less luciferase gene in normal orientation as well as reverse orientation, and then transfection analysis was performed with HeLa cells, in which NFI-A mRNA was expressed at a high level. As shown in Fig. 3C, the region comprising −597 to +130 bp in normal orientation exhibited a high level of promoter activity compared with that of the promoter-less luciferase vector. In contrast, the promoter activity was almost undetectable with the construct in the reverse orientation. This result indicates that we have cloned a functional promoter for the rat NFI-A gene. The Northern blotting analysis and luciferase assay results further supported the validity of the determination of the NFI-A gene transcription start sites by the 5′-RACE PCR.

Sequence Analysis of the 5′-Flanking Region—A 2.5 kb *Bam*HI fragment from λ77 known to contain the first exon and a portion of the 5′-flanking region was sequenced to characterize the potential regulatory sequences involved in NFI-A gene expression. Figure 4A shows the sequence of the 5′-flanking region 1.3 kb upstream from the ATG translation start codon. The NFI-A gene promoter lacks a canonical TATA-like element in close proximity to the transcription start sites and exhibits a high content of GC



B

	exon											
	1	2	3	4	5	6	7	8	9	10	11	(1-11)
rat NFI-A (aa*)	9	177	22	25	39	43	43	60	54	32	5	(509)
porcine NFI-C (aa*)	10	177	24	25	41	42	42	62	49	29	5	(506)
homology (%)	90	91	39	77	54	22	50	50	41	74	100	(61)

*number of amino acid residues

Fig. 5. **Comparison of the structures of the rat NFI-A and porcine NFI-C genes.** (A) Schematic representation of the gene structures of rat NFI-A and porcine NFI-C. The closed boxes show the exons with the relative sizes and the positions in each gene. The numerals above the boxes indicate the number of the exons. The major transcription start sites are marked by arrows. (B) Comparison of the length of exons and amino acid sequences between the two genes. The number of amino acid residues in the respective exons and the percent of homology between each exon of the two genes are indicated.

(more than 80%) in the 300 bp immediately upstream of the transcription start site (Fig. 4B). Six GC-boxes, which are known to guide initiation in some TATA-less promoters, are present at −304 (inverted), −241, −200, −156 (inverted), −75, and −34 bp from the transcription start site. These features have been described in genes that encode oncoproteins, growth factor, signal transduction components, and transcription factors, as well as many housekeeping proteins, and are usually associated with multiple transcription start sites (*24*).

In addition, several other potential *cis*-acting elements appear in the promoter region, such as multiple AP2-like recognition sites, a CCAAT sequence, and octameric sequences. A functional role for any of these sequences in the modulation of the NFI-A gene expression remains to be experimentally tested.

Finally, we compared the structure and the organization of the rat NFI-A gene with those of the porcine NFI-C gene reported by Meisterernst *et al.* (Fig. 5) (*12*). These two genes showed differences in the sequences of the 5′-flanking region and *cis*-acting elements in the promoter region (data not shown). These results suggest that the regulatory mechanisms controlling the rat NFI-A gene are different from that for the porcine NFI-C gene. However, these two genes share a similar exon/intron organization. Each gene was found to consist of 11 exons, and all of the exons were of similar sizes. In addition, the genes have a high level of similarity of the amino-terminal DNA binding domain (exon 1 and exon 2), and possess similar exon/intron boundary sequences. This conservation in the genomic structure and organization between the two genes suggests that duplication of an ancestral gene occurred rather recently to produce the NFI-A and NFI-C genes. It is not clear why these two genes have low levels of homology in exons 3, 6, and 9 (Fig. 5B). Since NFI-A2 and NFI-A3 lack exon 3, and exons 3 to 6, respectively, of the NFI-A cDNA, exons 3 and 6 might be involved in some specific function of the transcriptional regulation. It was reported that exon 9 was deleted in some isoforms of NFI-C (*9, 12*). Therefore it is possible that exon 9 is also involved in the different activities of the *trans*-activation/repression of these *trans*-acting factors.

REFERENCES

1. Nagata, K., Guggenheimer, R.A., Enomoto, T., Lichy, J.H., and Hurwitz, J. (1982) Adenovirus DNA replication *in vitro*: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. *Proc. Natl. Acad. Sci. USA* **79**, 6438-6442
2. Nagata, K., Guggenheimer, R.A., and Hurwitz, J. (1983) Specific binding of cellular DNA replication protein to the origin of replication of adenovirus DNA. *Proc. Natl. Acad. Sci. USA* **80**, 6177-6181
3. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987) A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* **48**, 79-89
4. Apt, D., Chong, T., Liu, Y., and Bernard, H.U. (1993) Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16. *J. Virol.* **67**, 4455-4463
5. Inoue, T., Tamura, T., Furuichi, T., and Mikoshiba, K. (1990) Isolation of complementary DNAs encoding a cerebellum-enriched nuclear factor I family that activates transcription from the mouse myelin basic protein promoter. *J. Biol. Chem.* **265**, 16065-16070
6. Osada, S., Daimon, S., Ikeda, T., Nishihara, T., Yano, K., Yamasaki, M., and Imagawa, M. (1997) Nuclear factor 1 family proteins bind to the silencer element in the rat glutathione transferase P gene. *J. Biochem.* **121**, 355-363
7. Kruse, U. and Sipple, A.E. (1994) The genes transcription factor nuclear factor I give rise to corresponding splice variants between vertebrate species. *J. Mol. Biol.* **238**, 860-865
8. Qian, F., Kruse, U., Lichter, P., and Sipple, A.E. (1995) Chromosomal localization of the four genes (NFIA, B, C, and X) for the human transcription factor nuclear factor I by FISH. *Genomics* **28**, 66-73
9. Santoro, C., Mermod, N., Andrews, P.C., and Tjian, R. (1988) A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature* **334**, 218-224
10. Gounari, F., Francesco, R.D., Schmitt, J., van der Vliet, P.C., Cortese, R., and Stunnenberg, H. (1990) Amino-terminal domain of NF1 binds to DNA as a dimer and activates adenovirus DNA replication. *EMBO J.* **9**, 559-566
11. Mermod, N., O'Neill, E.A., Kelly, T.J., and Tjian, R. (1989) The proline-rich transcriptional activator of CTF/NF1 is distinct from the replication and DNA binding domain. *Cell* **58**, 741-753
12. Meisterernst, M., Rogge, L., Foeckler, R., Karaghiosoff, M., and Winnacker, E.L. (1989) Structural and functional organization of a porcine gene coding for nuclear factor I. *Biochemistry* **28**, 8191-8200
13. Hattori, M. and Sakaki, Y. (1986) Dideoxy sequencing method using denatured plasmid templates. *Anal. Biochem.* **152**, 232-238
14. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
15. Chen, C. and Okayama, H. (1987) High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7**, 2745-2752
16. Herbomel, P., Bourachot, B., and Yaniv, M. (1984) Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. *Cell* **39**, 653-662
17. Naora, H. and Deacon, N.J. (1982) Relationship between the total size of exons and introns in protein-coding genes of higher eukaryotes. *Proc. Natl. Acad. Sci. USA* **79**, 6196-6200
18. Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349-383
19. Schmitt, J., Mielke, R., and Schrewe, H. (1995) Genomic organization of a mouse type I activin receptor. *Biochem. Biophys. Res. Commun.* **213**, 211-217
20. Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459-472
21. Novak, A., Goyal, N., and Gronostajski, R.M. (1992) Four conserved cysteine residues are required for the DNA binding activity of nuclear factor I. *J. Biol. Chem.* **267**, 12986-12990
22. Dekker, J., van Oosterhout, J.A.W.M., and van der Vliet, P.C. (1996) Two regions within the DNA binding domain of nuclear factor I interact with DNA and stimulate adenovirus DNA replication independently. *Mol. Cell. Biol.* **16**, 4073-4080
23. Mitchell, P.J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378
24. Kozak, M. (1992) Regulation of translation in eukaryotic systems. *Annu. Rev. Cell. Biol.* **8**, 197-225