The Chicken *c-fos* Gene: Cloning and Nucleotide Sequence Analysis

KOSAKU T. FUJIWARA,^{1†} KOUICHI ASHIDA,¹ HIROSHI NISHINA,¹ HIDEO IBA,¹ NOBUYUKI MIYAJIMA,²
MAKOTO NISHIZAWA,³ AND SADAANKI KAWAI^{3*}

*Department of Biophysics and Biochemistry, Faculty of Science, University of Tokyo, Hongo Bunkyo-ku, Tokyo 113,¹ and
Department of Oncology² and Department of Tumor Virus Research,³ The Institute of Medical Science, University of
Tokyo, Shirokanedai, Minato-ku, Tokyo 108, Japan*

Received 24 June 1987/Accepted 31 August 1987

Using a DNA probe from an avian transforming virus, NK24, that contains the *fos* gene and that was newly isolated and characterized (M. Nishizawa, N. Goto, and S. Kawai, *J. Virol.* 61:3733-3740, 1987), we cloned and sequenced the chicken *c-fos* gene. Results showed that this gene, like the mouse or human *c-fos* gene, contains four exons encoding a predicted gene product of 367 amino acids which is about 79% homologous to p55^{c-fos} (mouse), excluding several small insertions or gaps. The predicted gene product, however, contains two nonhomologous regions, as suggested by sequence analysis of the NK24 genome. About half the identical amino acids of these gene products are encoded by different codons, and the guanine-plus-cytosine content in the third letter of each codon used in the entire chicken *c-fos* coding region is high (93%). The 5'-flanking region of the TATA box contains a sequence that is homologous to a putative transcriptional regulatory sequence in the mouse or human *c-fos* gene. The 3'-noncoding region of chicken *c-fos* is strikingly homologous to that of mouse or human *c-fos* and covers a 67-base-pair adenine-plus-thymine-rich stretch which was previously reported to be essential for an inhibitory effect on the expression of the mouse *c-fos* gene.

Recent studies have shown that in many different cell types, the expression of a proto-oncogene, *c-fos*, is promptly and transiently activated upon induction of growth or differentiation of the cells and that this proto-oncogene may play an important role in cellular proliferation or differentiation (2, 10, 11, 16, 18, 19). The *c-fos* proto-oncogene was originally isolated from mouse genomic DNA (3, 27) as a cellular counterpart of the *v-fos* gene (5, 7) carried by FBJ murine osteosarcoma virus, which causes osteosarcomas in mice. The *c-fos* gene was shown to have transforming activity (14, 15) when it was put under the transcriptional control of the long terminal repeat after the removal of its 3'-noncoding region, including a 67-base-pair (bp) adenine-plus-thymine (A + T)-rich sequence.

With a *v-fos* probe from FBJ murine osteosarcoma virus genome DNA, the *c-fos* genes in several species were screened and the human and mink genomes were shown to have highly homologous sequences (5). The human *c-fos* gene was further cloned and characterized by nucleotide sequence analysis (28) and shown to encode a gene product, p55^{c-fos} (human), that was highly homologous to p55^{c-fos} (mouse) (4, 6). In the chicken genome, it was difficult to detect DNA that could hybridize to a *v-fos* probe from the FBJ murine osteosarcoma virus genome under stringent reannealing conditions (3; unpublished results). Nucleotide sequence analysis of a *v-fos* gene that had been transduced into a newly isolated avian transforming virus, NK24, suggested that the chicken *c-fos* gene may contain some regions that are fairly divergent from the mouse *c-fos* gene (20). Here we report the molecular cloning of the chicken *c-fos* gene by use of a *v-fos* probe from NK24 genome DNA (20) and

describe its gene structure deduced by nucleotide sequence analysis.

Chicken chromosomal DNA was partially digested with *Sau3AI* and cloned into the *Bam*HI sites of lambda Charon 35. Approximately 10⁶ lambda clones from the chicken genomic library were screened with a 1,047-bp *v-fos*-specific probe (20) from the NK24 genome by a standard procedure for stringent hybridization, and 17 positive clones were isolated. Physical mapping of the isolates confirmed that they were all derived from a single locus in the chicken chromosome, and two clones, 2 and 7, were studied further (Fig. 1). Using the same DNA probe, we subjected these cloned DNAs to hybridization experiments to identify restriction fragments that contained exon sequences. Two positive fragments, 2.2-kilobase *Sac*I-*Sac*I, and 1.4-kilobase *Bam*HI-*Bgl*II (Fig. 1), were subcloned into pUC plasmids in both directions. From each subcloned plasmid, a series of deletion derivatives was generated, cloned by the method of Henikoff (12) with slight modifications, and used as templates for dideoxy nucleotide sequencing (22). Closing of the sequence gap was accomplished by direct sequencing of subclones that covered the gap region. For determination of the guanine-plus-cytosine (G+C)-rich regions, deoxy-7-deazaguanosine triphosphate (17) was used instead of dGTP as a substrate for DNA polymerase. Regions with high G+C contents (about 70%) were also sequenced by the method of Maxam and Gilbert (13).

The complete nucleotide sequence of the *c-fos* chicken gene is shown in Fig. 2. It contains a TATA box (nucleotide 991) and AATAAA, a putative signal for polyadenylation (nucleotide 3679). The chicken *c-fos* gene locus is highly homologous with the *fos* coding sequence of the NK24 genome, confirming that the oncogene in the NK24 genome is derived from this locus of the chicken chromosome. The highly homologous sequence in this locus, however, is separated into four discontinuous regions (7, 249, 108, and

* Corresponding author.

† Present address: Department of Tumor Virus Research, The Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108, Japan.

FIG. 1. Restriction endonuclease map of the chicken *c-fos* locus. The 2.2-kilobase (kb) *SacI-SacI* and 1.4-kilobase *BamHI-BglII* fragments that were subjected to sequence analysis are indicated by arrows. Black boxes show the coding sequences of the four exons in the chicken *c-fos* gene.

606 nucleotides) by three interrupting sequences. A comparison of the nucleotide sequences around the ends of these three interrupting sequences with the consensus sequences for RNA splice donor and acceptor sites (24) showed that the sequences around the 5' ends of the interrupting regions matched the consensus sequence of the donor site and that the sequences around the 3' ends matched the consensus sequence of the acceptor site (Table 1). These results indicate that these three interrupting sequences are introns and suggest that these four discontinuous regions correspond to exons or parts of exons of the chicken *c-fos* gene. Since the mouse *c-fos* gene consists of four exons, we compared these discontinuous regions with the coding regions of the mouse *c-fos* exons. The discontinuous region composed of the stretch of 7 bp which has a sequence identical to that of the 5' end of the *fos*-specific sequence in NK24 is located upstream of the three other discontinuous regions and is homologous with the 3' end of mouse *c-fos* exon 1 (five of seven residues are identical), and the region upstream of the 7-bp stretch is homologous with the entire coding region of mouse *c-fos* exon 1, which suggests that the homologous region that includes the discontinuous region of the 7-bp stretch represents the first exon of the chicken *c-fos* gene. Furthermore, exon 2, exon 3, and the coding region of exon 4 of mouse *c-fos* are homologous with each of the three other discontinuous regions in chicken DNA. From these results, the structure of the chicken *c-fos* gene was deduced (Fig. 2). The chicken *c-fos* gene encodes a protein of 367 amino acids.

TABLE 1. Predicted splicing junctions in the chicken *c-fos* gene^a

Intron	Donor site	Acceptor site
1	AG / GTGAG	GTCCCGCAG /
2	AG / GTGAG	TGTCCCGCAG /
3	CG / GTGAG	CTTCCCGCAG /

^a Nucleotide sequences of the chicken *c-fos* gene at the points of discontinuity with the *v-fos* region of the NK24 genome (indicated by shills) are shown. The consensus sequences for the donor sites (5' end of intron) and acceptor sites (3' end of intron) for splicing (24) were AG/GTNAG and YNYYNCAG/, respectively. Nucleotides that did not agree with the consensus sequence are underlined.

The predicted chicken *c-fos* protein would be initiated at nucleotide 1154, the first ATG codon from the TATA box. This protein is expected to be acidic and hydrophilic and to contain high proportions of alanine, serine, and proline residues (49, 45, and 40 residues, respectively).

A comparison of the nucleotide sequences of the chicken *c-fos* gene and the *v-fos* gene of NK24 indicated that the oncogene product of NK24 contains a truncated version of chicken *c-fos* that lacks the initial 45 amino acids and is fused to the *gag* portion of avian leukosis virus at its N terminus. The nucleotide difference between the *v-fos* and *c-fos* coding regions caused no amino acid change. A comparison of the chicken *c-fos* sequence around the presumed recombination junction (TCAACT) and the sequence of the *gag* portion of the Prague C strain of Rous sarcoma virus (RSV) (23) revealed a homologous sequence of six nucleotides (TTAAAT), suggesting that recombination at the 5' site of the transduced *fos* sequence occurred between these homologous sequences (Fig. 2). The 3' terminus of the chicken *c-fos* sequence carried by NK24 is found 111 nucleotides downstream from the termination codon of the *fos* gene. Following the *c-fos* terminus, NK24 has a sequence of 55 nucleotides derived from the *pol* gene. The DNA sequence around the 3' junction of the chicken *c-fos* gene is found to share an identical stretch of 9 bp with an internal sequence of the *pol* gene (23), indicating that the 3' end of the chicken *c-fos* gene is transduced into the retrovirus genome by a homologous recombination.

For comparison of the chicken and mouse *c-fos* gene products, the deduced amino acid sequences were aligned for maximum amino acid homology. Excluding gaps and insertions of a few amino acids, a direct comparison of a sequence of 362 amino acids was possible, and 285 amino acids (79%) were found to be common to both gene products. In the middle part of the chicken *c-fos* coding sequence (from the end of exon 2 to the beginning of exon 4), a long amino acid sequence (from residues 116 to 204) is highly homologous to that of mouse *c-fos*. On the other hand, the latter part of exon 2 and the middle part of exon 4 encode rather nonhomologous amino acid sequences, which correspond to regions 2 and 4 of the NK24 *v-fos* gene shown in

SacI BamHI SmaI
 GAGCTCAAGAGCGGGAGAGGCTGCCAGCAGGATCCCAGCCTGCCCGGTGACGCTCCTGCTTGGAGACTGCGCTGCGGGCTCTGTTTGT 90

PstI
 TGCCGTAAACACGCTGCCGTGCAGCTTCCCTGTTACTATCTATAGCCGCGATCTCTTGGCTCCCGTGCAGGCGAGGCTGCGGTCTGGC 180

CGCCGGCTCTGCTGCCCGAGCCAGGGCTGCCACATCCCGCGGTGGGTGGTGTCTCCGGGCTTCGCCGCGACGGGGATCAAAACAA 270

AACCCAACTGGAGAAAGGACGGGCTCGCCCTCGCCCTCCCGCGCGCCCCCGCGGTGCCCGGTCCCGACGCCCGGGGGCTCTC 360

SmaI MluI
 CGGAGGGCCACGCGGCCCGGGCGCCCGCACTGCTTCCGGCCCTCCGTGCGGTGACCCCGCGGGCGGACGCGTCCCGCGCGCA 450

MluI
 GCCGCGACGGCCCGCACACGCGGCACCCCGCGCGAGGGACGAGGGCGCGCGCCCGCGCGCCCGCGCGCCCGCGCGTGGCCGC 540

GCCCGGCCCGTTAGTACGGAACCCCGCGGGCCGCTCGGGGCGGGGGTGC CGGCGCGGGGGCGCTACGCGGGGGGGGGGGG 630

CTGGGGCGGCCCGTCCACCGCTCCCGTCCCGCGCGCACTTGGCCGCGAGCCTGGAGCGCTTCCCGTCACTACGCCCTCGCACAG 720

EcoRV
 GATGTATCCCATATAAGGATATCTGCGTCAGTGGGTTCCTCGAGCCCGCGGTACCACTCCCGAGGGGGGGGAAACCGGAACCGCTC 810

CCCCGCCCGCGCGCTCTCCCGCCCGCGGGCCCGTTGTTCCCGCGCCCGCGGCTCGGCAGGGCGGGCGGGCGGGCGGGCGGG 900

CGGGACGAGACGACGCGGGCAGAGAGGCCCTCCCGGGCGGGCGTGGTGCCTCCACGTGACGTAGGCGATTACGCGGGCGCAGGG 990

TATA box
 TATAAGGGGGCGGCGCGAGCGATACTCCAGCGCAGAGCAGCTCCACAGCGGAGAGAGACGCGGCGCGGACCGGACACAGGG 1080

ACGCAGCCGGCCCTCTTCGCTTCTCCCGCAGGGCGCTCCGCTAGCACCGCACGGCCCCGCAACAGCCGACATGATGTACAGGGCTT 1170
 MetMetTyrGlnGlyPh

PstI
 CGCTGGGGAGTACGAGGCGCCTCTCCCGCTGCAGCAGCGCTTCCCGCGCGGGGACAGCCTACCTACTACCGTCCCGGCGGACTC 1260
 eAlaGlyGluTyrGluAlaProSerSerArgCysSerSerAlaSerProAlaGlyAspSerLeuThrTyrTyrProSerProAlaAspSe

1st intron KpnI
 CTCTCCAGCATGGGTTCCTCCCTCAACTCGCAGGTGAGGGCCGTTGCGCGGCCCGCGGGGCTCGGTGTTGGGGTACCGGGGAGGAG 1350
 rPheSerSerMetGlySerProValAsnSerGln

ATGGGCGCGGGGGCGGGGGCGGTATCGGGCCGGTCCGGGCGGTGCGGGACCCGAGGGGCGGGTCCGGCTTCAGCGGGGCCGGGG 1440

CAGCGTGGGAGCGGCCGCGCGCGCGGGGGCGGGCGGGGGAGCCGGGTCCGGGCTGCGCGTTGCGGCGGTGTGTAAAGCGGCT 1530

TCATGTATAAACGCGAGTTTCATTGAGGAGACTCCGGAGCGGCCCTGCGTCAGCGCGGACGTACAGATATTTATATCGGGCCGCTCTC 1620

SmaI
 GTGCGGAGCGCGCTGGCGCGCGGCCCGGGGCGGGGAGCGGCCGAGGGGAGTGCTGACCGCAGCCCCCTCACCGCTCC 1710

CCCCCACGCTGTGTCCCGCAGGATTTCTGCACCGACCTGGCCGTGTCAGCGCCAACCTCGTGCCACCGTGACGGCCATCTCCACCAG 1800
 AspPheCysThrAspLeuAlaValSerSerAlaAsnPheValProThrValThrAlaIleSerThrSe

PstI
 TCCCGACCTGCAGTGGTGGTGACGCCACCTCATCTCCTCGGTGCGCCCTCCAGAACCGGGGACCCCTACGGCGTTCCGGCGCC 1890
 rProAspLeuGlnTrpLeuValGlnProThrLeuIleSerSerValAlaProSerGlnAspArgGlyHisProTyrGlyValProAlaPr

C
 CGCCCTCCCGCCGCTATTCCCGCCCGCGGTGTTGAAGGCGCGGGCGCGGGGCGAGCATCGGACGAAGGGGCAAAGTCAGCA 1980
 oAlaProProAlaAlaTyrSerArgProAlaValLeuLysAlaProGlyGlyArgGlyGlnSerIleGlyArgArgGlyLysValGluG1

2nd intron
 GGTGAGCGGGGCTGGGGGAGGAAGGGGGGGTTG²GGGGGGCGGAGGGGTGGGAGCGGGGCGGGGCGGGGCGGGGCGGGGCTCTG 2070
 n

PvuII BamHI
 ACGGACGCGGTGACCGCCCTCTGTGTCGCGAGCTGTCCCGGAGGAGGAGGAAAAGAGGAGGATCCGCGGGAGAGGAACAAGATGGCAG 2160
 LeuSerProGluGluGluLysArgArgIleArgArgGluArgAsnLysMetAlaA

SacI PstI 3rd intron SmaIBamHISmaI
 CGGCCAAGTGCCGCAACCGGCGGGAGCTACCGACACTCTGCAGGCGGTGAGTGTGCCCGGGATCCCGGGGCTGCGGGCGGGG 2250
 laAlaLysCysArgAsnArgArgArgGluLeuThrAspThrLeuGlnAla

PvuII

ACCCCGTGGGTAGAGCGGTGAGGTTGGCCGCGGTGACGGCCGCTCTGCTCTTCGCGAGGAGACGGACCGTGGAGGAGGA 2340
GluThrAspGlnLeuGluGluGlu

PstI

GAAGTCCGCTCTGCAGGCGGAGATAGCCAACCTGCTGAAGGAGAAGGAGAAGCTGGAGTTTCATCTGGCGGCGCACCGGCCCGCTGCAA 2430
uLysSerAlaLeuGlnAlaGluIleAlaAsnLeuLeuLysGluLysGluLysLeuGluPheIleLeuAlaAlaHisArgProAlaCysLy

StuI

GATGCCCGAGGAGCTGGCTTCTCCGAGGAGCTGGCGGCCGCCACCGCACTGGACCTGGGCGCACCCAGCCCGCGCGGCGGAGGAGGC 2520
sMetProGluGluLeuArgPheSerGluGluLeuAlaAlaAlaThrAlaLeuAspLeuGlyAlaProSerProAlaAlaAlaGluGluAl

StuI

CTTCGCCCTGCCGCTAATGACCGAGGCGCGCGCGCGCTGCCGCCAAGGAGCGGAGCGGCGGCGCTGGAGCTGAAGGCCGAGCCCTT 2610
aPheAlaLeuProLeuMetThrGluAlaProProAlaValProProLysGluProSerGlySerGlyLeuGluLeuLysAlaGluProPh

StuI

CGACGAGCTGCTTTTCTCCGCGGGGCGCGGAGGCTCCCGCTCGGTGCCTGACATGGACCTGCCCGGAGCTCTCTCTTCTACGCGTTC 2700
eAspGluLeuLeuPheSerAlaGlyProArgGluAlaSerArgSerValProAspMetAspLeuProGlyAlaSerSerPheTyrAlaSe

MluI

BstEII

GGACTGGGAGCGCTGGGCGCGGAGCGCGGCGGAGCTGGAGCCCTCTGCACCCCGTGGTGACCTGCACCCCGTGCCTAGCACCTA 2790
rAspTrpGluProLeuGlyAlaGlySerGlyGlyGluLeuGluProLeuCysThrProValValThrCysThrProCysProSerThrTy

PvuII

CACCTCCACCTTCGTCTTACCTACCCGAGGCGGACGCTTCCCGAGCTGGCGCGCTGGCACCCGGAAGGCGAGCAGCAGCAACGAGCC 2880
rThrSerThrPheValPheThrTyrProGluAlaAspAlaPheProSerCysAlaAlaAlaHisArgLysGlySerSerSerAsnGluPr

C

ApaI

CTCGTCCGACTCCCTCAGCTCCCGACCCCTGCTGGCTTGTGAGGGGCTCGCCCTGCGCTGACTGACCTGCCGGGCCCCCTCCCTGCCC 2970
oSerSerAspSerLeuSerProThrLeuLeuAlaLeuTER

SmaI

ACGGCCCCACGACTCGCCACGCCCCCGGGGCTCCCGACCTGGGAGGGCGCTGCTGCTCCACCCCCCTGTCCAGCCGGGGCC 3060

SmaI

CGGGCCCTGGTGGAGTGCTGGGCCATACCTCCTCCAGAGATGTAGCAACCGCATGGAGTTTGTCTTGTCCCCAACGGCCCATCTGTG 3150

SmaI

AGAGCTGGTAGTCTGTAGCATGTCCACATGGCTGGGTGGTGACTCCCCCTCCTTAGTATCACTAGCATTAACTAATTAATCCCTCG 3240

DraI

PvuII

GTTTTAAATGATTGAATTAACCTGGTGTGGGTATCTCCAACCTGTATCTAGTGCAGCTGATTAACAATAACTACTGTGTTCTCGGCAA 3330

TATCGTGTCTGTGACTTAGCAACGACCATCTTACGTGGGGGGAAGAGACTCTATTTTATTTCTAGTAGGTAGATGAATAGCTAT 3420

ClaI

ATCCATGTACTGTAGTTCTACATCGATGTTTCATTGTAACCTTACTGATCATGTCATTGTTGAGGTGGTCTGAATGTTCTGACATAAGTTTT 3510

CCATGAAAACGTTTTTATTGTGTTTTTAATTTATTTATTAAGATGGATTCTCAGATATTTATTTTATTTTATTTTCTACCTTGA 3600

polyA sig.

GGTCTTTTTTGACATGTGGAAGTGAATTGGATGAACTTAAGCATTGTTTGTCTTATTATTGTTTCAGAGACATTGTCATAAAAGCATT 3690

TAAGTTGAGTGTGGAGCTGCTTGTCTTACCTTTTGCATTACTCCCTCTGACCCCAAAGGGGCTCCCTGAACCTGCTGTGGGGTTGGG 3780

BglII

CTTTACTGGCATCAGACTGCTGTGAGCTGCCCTAAGCGTAGGGTGGCTTGGGGTGGGAAAACCATTAAGTGCAGATCT 3858

FIG. 2. Nucleotide sequence (3,858 nucleotides) of the chicken *c-fos* gene. A TATA box and a poly(A) signal (polyA sig.) are indicated by boldface underlining. The 5'-noncoding region which is highly homologous to that of the mouse *c-fos* gene and the A + T-rich stretch of 67 bp (12) are indicated by lightface underlining, and both regions are compared with the corresponding regions of the mouse and human *c-fos* genes in Fig. 4. Direct repeats composed of 19 nucleotides are indicated by boldface arrows, and the boxes in the repeats represent GC boxes. The amino acids encoded by each exon are shown below the nucleotide sequence. Possible sequences involved in homologous recombination in the process of NK24 generation are boxed. The 5' terminus (▽) and the 3' terminus (▼) of the NK24-specific nucleotide sequence, which is defined in the accompanying paper (20), are also indicated. The nucleotides in the chicken *c-fos* gene that differ from the corresponding nucleotides in the *fes*-derived region in NK24 genome DNA are indicated above the sequence.

FIG. 3. Comparison of chicken, mouse, and human *c-fos* gene products. Amino acids are indicated by the single-letter code, and the numbers for the amino acid residues of the chicken *c-fos* gene product are shown. Identical amino acids in at least two of these three gene products are boxed, and the three arrowheads indicate the positions of introns present at the nucleotide level. Identical amino acid residues in the chicken and mouse *c-fos* gene products that are encoded by different codons are shown by small black inverted triangles above the amino acid residues of the chicken *c-fos* gene. Dashes indicate the absence of the corresponding amino acid.

Fig. 6 in the accompanying paper (20). Much higher homology, however, was previously reported between the human and mouse *c-fos* gene products (28). There are no gaps or insertions between them, and only 24 amino acid exchanges can be detected. It is noteworthy that most of the amino acid changes between mouse and human *c-fos* gene products are located in the nonhomologous region of the mouse and chicken *c-fos* gene products.

About half of the identical amino acids in the chicken and mouse *c-fos* gene products are encoded by different codons (Fig. 3), and the G + C content of the third letter of each codon used in the entire coding sequence is significantly higher in chicken *c-fos* (93%) than in mouse *c-fos* (68%) or human *c-fos* (71%). This significant difference in the nucleotide levels seems to explain why the *v-fos* probe from FBJ murine osteosarcoma virus DNA did not hybridize to the chicken *c-fos* gene under stringent conditions (5; unpublished data). As reported for several other genes in which the third position of codons in the coding regions have high G + C contents (1), the *c-fos* gene is surrounded by G + C-rich introns and G + C-rich flanking regions.

The expression of the mouse and human *c-fos* genes in a wide variety of cell types is promptly and transiently induced by several reagents, such as 12-*O*-tetradecanoylphorbol-13-

acetate, serum, growth factors, and Ca^{2+} , and there are many reports of studies of the transcriptional regulatory sequences by analysis of the chromosome structure with DNase I, the transient expression method, the gel retardation method, the footprinting method, and other methods (7, 8, 21, 25, 26). The regulatory sequence identified by several methods includes a common region located 275 ± 25 bp upstream from the TATA box in both the mouse and human *c-fos* genes (21). A search for homologous sequences in the 5'-noncoding regions of the chicken and mouse *c-fos* genes was made, and a fairly homologous region was detected (Fig. 4A). This region included the transcriptional regulatory sequence found in the mouse or human *c-fos* gene. Between this region and the TATA box, the chicken *c-fos* gene has two direct repeats (19 bp with a 2-bp mismatch) which contain a consensus sequence for a GC box (9).

The sequence homology found between the latter part of the 3'-untranslated region of chicken *c-fos* and that of mouse *c-fos* is rather striking. A 67-bp A + T-rich stretch which is reported to function as an inhibitory sequence for the expression of the mouse *c-fos* gene (14) is included in these 3'-noncoding regions. This sequence is identical to that of the human gene, except for the deletion of 1 bp, and only 1 nucleotide exchange can be detected between the 67-bp

A)

chick; CTGGGGCGGGCCCGTCCACCGTCCCGTCCCGCGCCGCAC-----TTGGCCGCAG 682 (-309)
 mouse; CCCC G G -- GT CC T T CCTCAGAG T 550 (-323)
 human; CCCC A G G - C -- A T CCTCGGTG T 374 (-327)

chick; CCTG-GAGCGCTTCCCGTCAGTCACGCCC-CC-TCGCACAGGATGTATCCCATATAAGGA 739 (-252)
 mouse; G C TG A C T T T TA --- T 607 (-266)
 human; C C AG A C T - C TA --- T 430 (-271)

chick; TATCTGCGTCAGTGGGTTCGCCGAGCCCGG-CCGTACCACTCCGCGAGGGGGGGGAAAC 798 (-193)
 mouse; C CA T AC G T C GTTG T - - ACC TCT 665 (-208)
 human; C CA T AC G TTT C GTAG C T - - A --CC TC 486 (-219)

B)

chick; GACCCATCTTACGTGGGGGGGAAAGAGACTCTATTTTATTTCTAGTAGGTAGATG--AA 3413 (490)
 mouse; A A G C AATAT AAATAAG A TAGG C C A AATC 3995 (521)
 human; A ATG C AATAT ATACTAAG A TACG C G A AATA 3839 (510)

chick; TAGCTATATCCATGTACTGTAGTTCTACATCG-----ATGTTCAATTGTAACTTTACTGA 3467 (544)
 mouse; C T G ---TCA C TG 4052 (578)
 human; T T T AACATCA TG 3899 (570)

chick; TCATGCATTGTTGAGGTGGTCTGAATGTTCTGACAT-A--AGTTTCCATGAAAACGTTT 3524 (601)
 mouse; C T AC 4112 (638)
 human; T AC - 3958 (629)

chick; TTATTGTGTTTTTAATTTATTTATTAAGATGGATTCTCAGATATTATATTTTATTTTA 3584 (661)
 mouse; C 4172 (698)
 human; 4018 (689)

chick; TTTTTTCTACCTTGAGGTCTTTTTTGACATGTGAAAGTGAATTGGATG--AAA-CTT 3641 (718)
 mouse; C -- C A AA TT 4230 (756)
 human; -- A AA T- 4075 (746)

chick; AAGCATTGTTTGCTTATTATTGTTCTAGAGACATTGTCAATAAAAGCATTTAAGTTGAGTG 3701 (778)
 mouse; --- CAG A 4287 (813)
 human; --- CA A 4132 (803)

chick; CTGGAGCTGTCCTTGCTTTACCTTTTGCAATTAC 3734 (811)
 mouse; GACCA CT T GCTC CTITA CT G TT 4320 (846)
 human; GACCAACC TG GCTC TTT A CTGGAAGT 4165 (836)

FIG. 4. (A) Sequence homology in transcriptional regulatory sequences located in the 5'-flanking regions of the *c-fos* genes. Nucleotide numbers from TATA boxes are shown in parentheses on the right. (B) Sequence homology in the 3'-noncoding regions of the *c-fos* genes. Nucleotide numbers from the termination codons are shown in parentheses on the right. A 67-bp A + T-rich sequence is indicated by arrows above the sequence. The nucleotides of the mouse and human *c-fos* genes are numbered as described in reference 29. The nucleotides of the chicken *c-fos* gene are numbered as shown in Fig. 2. polyA sig., poly(A) signal (underlined sequence). Dashes indicate the absence of the corresponding nucleotide.

regions of the chicken and mouse sequences (Fig. 4B). The evolutionary conservation as well as the fact that this region is absent in the NK24 genome (20) suggest that this region is important for regulation of the expression of the *c-fos* gene.

We thank Y. Okada for helpful discussions and encouragement throughout this work and T. Meshi for help in the computer analysis.

This work was supported in part by a grant-in-aid for scientific research from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

1. Aota, S., and T. Ikemura. 1986. Diversity in G+C content at the third position of codons in vertebrate genes and its cause. *Nucleic Acids Res.* 14:6345-6355.
2. Cochran, B. H., J. Zullo, I. M. Verma, and C. D. Stiles. 1984. Expression of the *c-fos* gene is stimulated by platelet-derived growth factor. *Science* 226:1080-1082.
3. Curran, T., W. P. MacConnell, F. van Straaten, and I. M. Verma. 1983. Structure of the FBJ murine osteosarcoma virus genome: molecular cloning of its associated helper virus and the cellular homolog of the *v-fos* gene from mouse and human cells. *Mol. Cell. Biol.* 3:914-921.
4. Curran, T., A. D. Miller, L. Zokas, and I. M. Verma. 1984. Viral and cellular *c-fos* proteins: a comparative analysis. *Cell* 36:259-268.
5. Curran, T., G. Peters, C. Van Beveren, N. M. Teich, and I. M. Verma. 1982. FBJ murine osteosarcoma virus: identification and molecular cloning of biologically active proviral DNA. *J. Virol.* 44:674-682.
6. Curran, T., and N. M. Teich. 1982. Candidate product of the FBJ murine osteosarcoma virus oncogene: characterization of a 55,000-dalton phosphoprotein. *J. Virol.* 42:114-122.
7. Darby, M. K., R. E. Herrera, H.-P. Vosberg, and A. Nordheim. 1986. DNA topoisomerase II cleaves at specific sites in the 5' flanking region of *c-fos* proto-oncogene *in vitro*. *EMBO J.* 5:2257-2265.
8. Deschamps, J., F. Meijlink, and I. M. Verma. 1985. Identification of a transcriptional enhancer element upstream from the proto-oncogene *fos*. *Science* 230:1174-1177.
9. Dynan, W. S., and R. Tjian. 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA binding proteins. *Nature (London)* 316:774-778.
10. Greenberg, M. E., L. A. Greene, and E. B. Ziff. 1985. Nerve growth factor and epidermal growth factor induce rapid transient changes in proto-oncogene transcription in PC12 cells. *J. Biol. Chem.* 260:14101-14110.
11. Greenberg, M. E., and E. B. Ziff. 1984. Stimulation of 3T3 cells induces transcription of the *c-fos* proto-oncogene. *Nature (Lon-*

- don) 311:433–438.
12. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**: 351–359.
 13. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499–560.
 14. Meijlink, F., T. Curran, A. D. Miller, and I. M. Verma. 1985. Removal of a 67-base-pair sequence in the noncoding region of protooncogene *fos* converts it to a transforming gene. *Proc. Natl. Acad. Sci. USA* **82**:4987–4991.
 15. Miller, A. D., T. Curran, and I. M. Verma. 1984. *c-fos* protein can induce cellular transformation: a novel mechanism of activation of a cellular oncogene. *Cell* **36**:51–60.
 16. Mitchell, R. L., L. Zokas, R. D. Schreiber, and I. M. Verma. 1985. Rapid induction of the expression of proto-oncogene *fos* during human monocyte differentiation. *Cell* **40**:209–217.
 17. Mizusawa, S., S. Nishimura, and F. Seela. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.* **14**:1319–1324.
 18. Morgan, J. I., and T. Curran. 1986. Role of ion flux in the control of *c-fos* expression. *Nature (London)* **322**:552–555.
 19. Müller, R., R. Bravo, J. Burkhardt, and T. Curran. 1984. Induction of *c-fos* gene and protein by growth factors precedes activation of *c-myc*. *Nature (London)* **312**:716–720.
 20. Nishizawa, M., N. Goto, and S. Kawai. 1987. An avian transforming retrovirus isolated from a nephroblastoma that carries the *fos* gene as the oncogene. *J. Virol.* **61**:3733–3740.
 21. Renz, M., M. Neuberg, C. Kurz, R. Bravo, and R. Müller. 1985. Regulation of *c-fos* transcription in mouse fibroblasts: identification of DNAase I-hypersensitive sites and regulatory upstream sequences. *EMBO J.* **4**:3711–3716.
 22. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
 23. Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. *Cell* **32**:853–869.
 24. Sharp, P. A. 1981. Speculations on RNA splicing. *Cell* **23**: 643–646.
 25. Treisman, R. 1985. Transient accumulation of *c-fos* RNA following serum stimulation requires a conserved 5' element and *c-fos* 3' sequences. *Cell* **42**:889–902.
 26. Treisman, R. 1986. Identification of a protein-binding site that mediates transcriptional response of the *c-fos* gene to serum factors. *Cell* **46**:567–574.
 27. Van Beveren, C., F. van Straaten, T. Curran, R. Müller, and I. M. Verma. 1983. Analysis of FBJ-MuSV provirus and *c-fos* (mouse) gene reveals that viral and cellular *fos* gene products have different carboxy termini. *Cell* **32**:1241–1255.
 28. van Straaten, F., R. Müller, T. Curran, C. Van Beveren, and I. M. Verma. 1983. Complete nucleotide sequence of a human *c-onc* gene: deduced amino acid sequence of the human *c-fos* protein. *Proc. Natl. Acad. Sci. USA* **80**:3183–3187.
 29. Weiss, R., N. Teich, H. Varmus, and J. Coffin. 1985. RNA tumor viruses. *Molecular biology of tumor viruses*, 2nd ed., part 2, p. 884–899. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.