Isolation and sequence of the murine Fgf6 cDNA

Vincent Ollendorff, Olivier Rosnet, Irène Marics, Daniel Birnbaum, Odile Delapeyrière

Biochimie (1992) 74, 1035–1038

Short communication

Isolation and sequence of the murine *Fgf6* cDNA*

V Ollendorff, O Rosnet, I Marics, D Birnbaum, O deLapeyrière**

Laboratory of Molecular Oncology, U119 INSERM, 27 Bd Leï Roure, 13009 Marseille, France

(Received 16 September 1992; accepted 24 September 1992)

Summary – We have studied the structure of the murine *Fgf6* gene encoding a fibroblast growth factor with the purpose of looking for putative regulatory sequences in the 5' and 3' non-coding regions. The *Fgf6* cDNA contains a very long 3' untranslated portion of 4015 nucleotides.

fibroblast growth factor / oncogene / nucleotide sequence / regulatory elements

Introduction

The fibroblast growth factor (FGF) family comprises to date seven members that are involved in several biological processes [1]. We have characterized the gene encoding the sixth member of this family in man [2–4] and in mouse [5]. The transcripts of the *FGF* genes in general, and of *Fgf6* in particular, comprise a long 3' tail corresponding to an untranslated region. The role of this sequence, present in human as well as in mouse remains unknown. However, several pieces of evidence indicate that this type of sequence may play some role. Additional open reading frames (ORFs) have been found in 5' or 3' regions of *FGF* genes [6, 7]. More importantly, Curatola and Basilico [8] have established the presence of regulatory elements with enhancer activity in the 3' non-coding region of the closely related *Fgf4/K–Fgf* gene. This enhancer binds nuclear factors [9]. For the purpose of studying the structure and possible regulatory role of the 5' and 3' untranslated regions of *Fgf6*, we cloned and sequenced the murine *Fgf6* cDNA.

Materials and methods

cDNA cloning and screening

A cDNA library from mRNA of 15.5-day mouse embryo was constructed in λ gt10. The cDNA synthesis, addition of *Eco*RI linkers and ligation to λ gt10 arms were performed using Amersham synthesis and cloning kits as recommended by the manufacturer. The ligated material was packaged using Gigapack Gold extracts (Stratagene) and the phages were plated using NM514 as a bacterial host strain. This library was screened with genomic (*Eco*RV-*Nae*I fragment) or cDNA (*Pst*I-*Nco*I fragment) mouse *Fgf6* probes (both previously described in [5]).

Isolation of cDNA clones by PCR amplification

An aliquot (2 µg) of total RNA from 15.5-day mouse embryo was used to generate a first strand cDNA. The RNA was heated to 90°C for 5 min and the cDNA was synthetized by incubating 10 min at room temperature and 1 h at 42°C in a 20-µl final volume reaction containing 100 pmol of random hexamers, 1 mM of each dNTPs, 200 U of Mo.MLV-reverse transcriptase (BRL), 1 $U\mu m^{-1}$ Rnasin (Promega) and 1 x PCR buffer. After heat inactivation at 95°C for 5 min, half of the cDNA reaction mix was subjected to PCR amplification in a final volume of 50 µl of 1 x PCR buffer containing 25 pmol of each oligonucleotide primer and 1 U of *Taq* polymerase (Cetus). The same PCR buffer (50 mM KCl, 20 mM Tris-HCl pH 8.4, 1.5 mM $MgCl_2$, 1 mg ml^{-1} BSA) was used in both reverse transcriptase and PCR reaction. After an initial denaturation step at 94°C for 3 min, a 'touchdown' PCR [10] was carried out, in which the annealing temperature was progressively decreased 1°C every cycle from 65°C to 55°C. At that last temperature 30 cycles were carried out. Cycles were 2 min at annealing temperature, 3 min at 72°C, and 1 min at 94°C, except for the last cycle where the extension time was 10 min. The PCR products were then blunt-ended using the Klenow polymerase and subcloned into a *Sma*I-digested Bluescript (Stratagene) plasmid.

cDNA sequencing

Nucleotide sequencing was performed by the dideoxy-chain termination method using single-stranded and/or double-stranded Bluescript (Stratagene) templates and the modified T7 polymerase (Sequenase) protocol (US Biochemicals).

*The nucleotide sequence data presented in this paper have been submitted to the EMBL-Genbank Data Libraries under the accession numbers M92415 (550 first nucleotides, of genomic origin) and M92416 (the rest of the sequence).

**Correspondence and reprints.

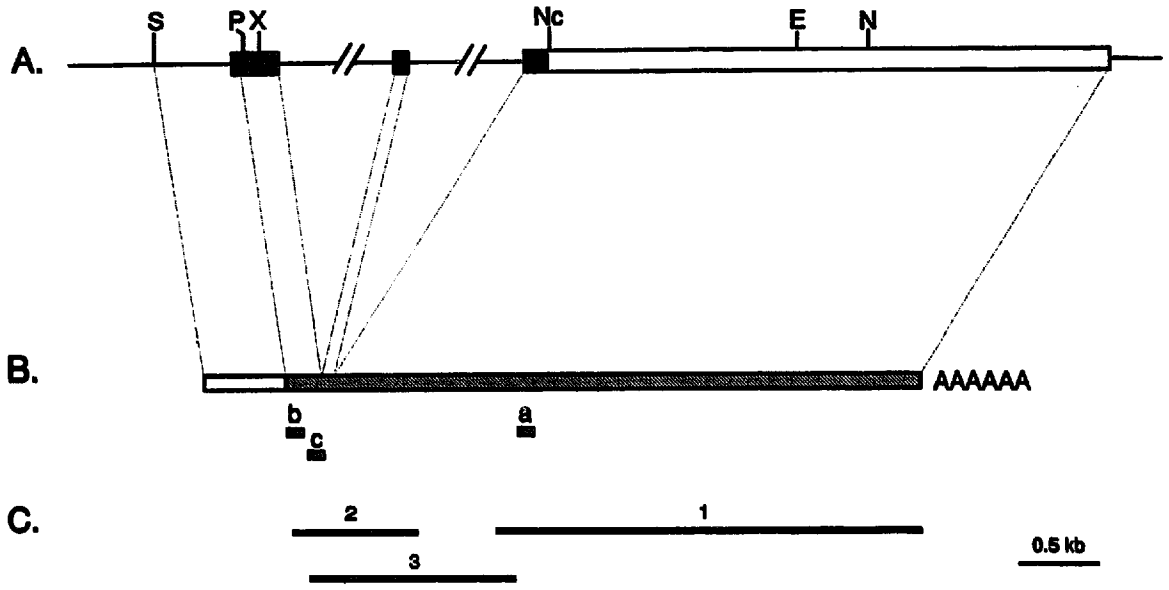

Fig 1. Cloning of *Fgf6* coding and untranslated regions. Most of the coding regions (black boxes) and the totality of the 3' untranslated region (open box) of the *Fgf6* gene (**A**) were obtained by cDNA cloning (**B**, dashed box). 5' Sequences (**B**, open box) were derived from genomic cloning. Restriction enzyme sites mentioned in the text are shown; abbreviations are as follows: E, *Eco*RV; N, *Nae*I; Nc, *Nco*I; P, *Pst*I; S, *Sph*I; X, *Xho*I. The inserts of the two cDNA clones (1 and 2) and the product of the PCR cloning (3) discussed in the text, are shown (**C**). The positions of the oligonucleotide primers (a, b, and c) are shown above. The scale is the same for the three panels.

Results and discussion

In figure 1 the clones that were isolated from a mouse embryo cDNA library are described. A first cDNA clone (clone 1) contained 2.8 kb of 3' untranslated sequence. A second cDNA clone (clone 2) contained part of the 3' region, the whole of the coding sequences of exons 3 and 2, and extended 5' into the biggest part of exon 1. In between these two cDNAs, a portion of the 3' sequence missing in the cDNA clones was obtained by cloning of a polymerase chain reaction (PCR) product obtained after two PCR steps. The first step used a 3' primer (a) designed in the 5' end of the first clone and a 5' primer (b) chosen in the first exon. These primers span two different exons. Their nucleotide sequences were 5'-GGATTGTCTTGCAGGTCAGG-3' and 5'-AACACACGAGGAGAACCCCT-3', respectively. The second step was performed in the same conditions except for the use of a 5'-AGAGTGCTCTCTTCATTGCC-3'- nested 5' primer (c) designed in the second exon. The combination of the cDNA clones and PCR product resulted in a 4.6-kb cDNA, missing the first 54 nucleotides of *Fgf6* coding sequence. Further screening of either the original or newly made cDNA libraries did not allow to isolate cDNA clones extended fully to the 5' end, presumably because of strong secondary structures. Further upstream sequences corresponding to the 5' coding end of the *Fgf6* gene were therefore obtained from genomic clones. The genomic sequence was extended even more 5' to include putative regulatory elements. Portions of previously isolated cosmid clones from the *Fgf6* locus were subcloned. One such fragment, named XX [5], corresponded to the 5' portion of the *Fgf6* gene. A *Sph*I–*Xho*I subfragment (fig 1A) derived from XX was sequenced.

As in human [3], a 624-nucleotide-long ORF is able to code for a protein of 208 residues (fig 2). However, there are three in-frame ATG codons and two shorter proteins could possibly be made. The 5' region contains neither CAAT nor TATA boxes. Putative regulatory elements are located about 300 bp upstream of the ATG (fig 2); they include a decanucleotide repeat AGGGTGGTGA, and the palindromic TGACGTCA binding site for the CREB element found in the 5' part of cAMP regulated genes [11]. The 3' untranslated region, starting after the stop codon at position 1121, is 4015 nucleotides long. The polyadenylation site is preceded by an unusual poly(A) signal sequence at position 5122. A poly(A) stretch starting at position 1686 may represent the remnant of an archaic polyadenylation sequence. It is

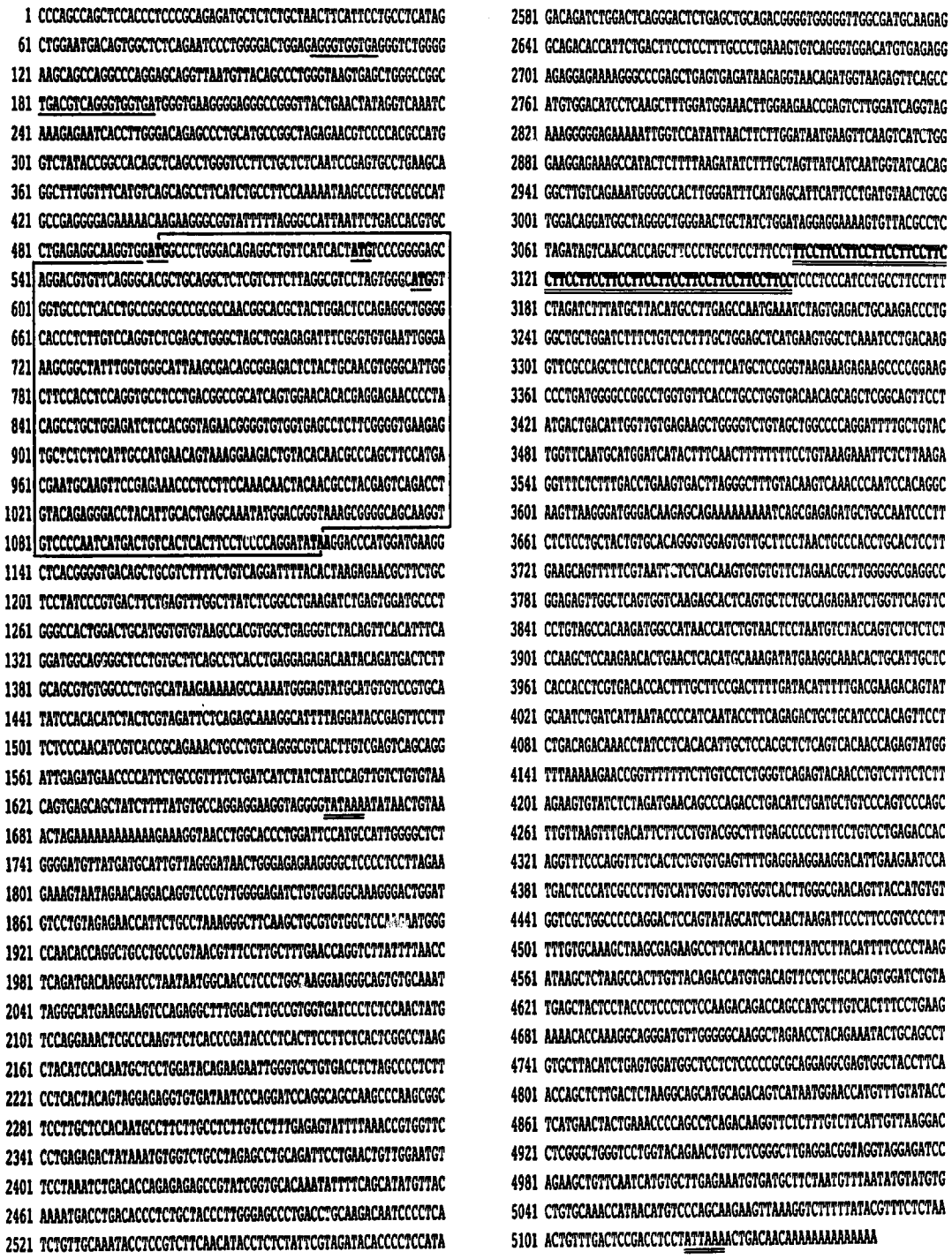

Fig 2. Nucleotide sequence of the mouse *Fgf6* coding and untranslated regions. The nucleotide sequence is a composite of genomic sequences (1 to 550) and cDNA sequence (551 to the end). The coding sequence is boxed. Special features are as follows: three ATG codons (bold and underlined), the stop codon (bold), putative 5' regulatory elements (underlined), a variant polyadenylation signal sequence (doubly underlined) and a putative alternative polyadenylation signal sequence (doubly underlined, position 1664). Doubly underlined and in bold is a block of 15 tetranucleotide repeats.

preceded by a putative polyadenylation signal sequence that may be used (shorter *Fgf6* transcripts of about 1.6 kb are indeed sometimes observed in some tissues). A block of 15 tetranucleotide repeats is present from position 3099 to 3159. These repeats identify a putative microsatellite and could be used to reveal a polymorphism in mouse, and maybe also in human where they could be conserved [12].

Indeed, in a more general way, comparison of the 5' and 3' non-coding sequences described here with the corresponding human sequences could yield information about important conserved features of the 5' and 3' regions. Furthermore, the cloning and sequencing of the 3' untranslated region of *Fgf6* will help in deriving specific *Fgf6* probes useful in other studies such as *in situ* hybridization analysis of *Fgf6* expression. Indeed, the 5' fragment of clone 1 obtained by *Eco*RV digestion (named dream) used as template for antisense riboprobes gives a good signal in hybridization of mouse embryo sections (deLapeyrière *et al*, in preparation).

Acknowledgments

This work was supported by Inserm and grants from ARC and AFM. We are grateful to F Coulier for helpful advice and J Planche for technical assistance. VO is the recipient of a fellowship from the 'Ligue Nationale Contre le Cancer'.

References

1 Gospodarowicz D (1991) Biological activities of fibroblast growth factors. *Ann NY Acad Sci* 638, 1–8

2 Marics I, Adelaide J, Raybaud F, Mattei MG, Coulier F, Planche J, deLapeyrière O, Birnbaum D (1989) Characterization of the Hst-related FGF-6 gene, a new member of the fibroblast growth factor gene family. *Oncogene* 4, 335–340

3 Coulier F, Batoz M, Marics I, deLapeyrière O, Birnbaum D (1991) Putative structure of the *FGF6* gene product and role of the signal peptide. *Oncogene* 6, 1437–1444

4 Coulier F, Ollendorff V, Marics I, Rosnet O, Batoz M, Planche J, Marchetto S, Pebusque MJ, deLapeyrière O, Birnbaum D (1991) The *FGF6* gene within the *FGF* multigene family. *Ann NY Acad Sci* 638, 53–61

5 deLapeyrière O, Rosnet O, Benharroch D, Raybaud F, Marchetto S, Planche J, Galland F, Mattei MG, Copeland N, Jenkins N, Coulier F, Birnbaum D (1990) Structure, chromosome mapping and expression of the murine *Fgf*-6 gene. *Oncogene* 5, 823–831

6 Taira M, Yoshida T, Miyagawa K, Sakamoto H, Terada M, Sugimura T (1987) cDNA sequence of a human transforming gene *hst* and identification of the coding sequence required for transforming activity. *Proc Natl Acad Sci USA* 84, 2980–2984

7 Zhan X, Bates B, Hu X, Goldfarb M (1988) The human FGF-5 gene encodes a novel protein related to fibroblast growth factors. *Mol Cell Biol* 8, 3487–3495

8 Curatola AM, Basilico C (1990) Expression of the K-*fgf* proto-oncogene is controlled by 3' regulatory elements which are specific for embryonal carcinoma cells. *Mol Cell Biol* 10, 2475–2484

9 Schoorlemmer J, Kruijer W (1991) Octamer-dependent regulation of the kFGF gene in embryonal carcinoma and embryonic stem cells. *Mech Dev* 36, 75–86

10 Don R, Cox P, Wainwright B, Baker K, Mattick J (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* 19, 4008

11 Hyman SE, Comb M, Lin YS, Pearlberg J, Green MR, Goodman HM (1988) A common *trans*-acting factor is involved in transcriptional regulation of neurotransmitter genes by cyclic AMP. *Mol Cell Biol* 8, 4225–4233

12 Moore S, Sargeant L, King T, Mattick J, Georges M, Hetzel D (1991) The conservation of dinucleotide microsatellites among mammalian genomes alllows the use of heterologous PCR primer pairs in closely related species. *Genomics* 10, 654–660