The FGF6 gene within the FGF multigene family

François Coulier, Vincent Ollendorff, Irène Marics, Olivier Rosnet, Michèle Batoz, Jacqueline Planche, Sylvie Marchetto, Marie-Josephe Pébusque, Odile Delapeyrière, Daniel Birnbaum

The *FGF6* Gene within the *FGF* Multigene Family[a]

FRANÇOIS COULIER, VINCENT OLLENDORFF,
IRÈNE MARICS, OLIVIER ROSNET, MICHÈLE BATOZ,
JACQUELINE PLANCHE, SYLVIE MARCHETTO,
MARIE-JOSÈPHE PEBUSQUE, ODILE deLAPEYRIERE,
AND DANIEL BIRNBAUM

U. 119 INSERM
27 Boulevard Leï Roure
13009, Marseille, France

Cell growth and differentiation are controlled at multiple levels through various pathways. External signals and their specific receptors represent important protagonists of this control. Among the signaling molecules are peptide regulatory factors including numerous neuropeptides, cytokines, and growth factors, often grouped in families. Growth factors and their receptors are involved in various processes including cell proliferation, cell differentiation, cell-cell interactions and development.

The FGF (historically meaning fibroblast growth factor) family presently comprises seven related members (hereafter designated FGF1 to FGF7) but the actual size of the family remains unknown. Like most of the signaling molecules, they have various known properties as well as suspected functions in basic processes as diverse as cell proliferation, angiogenesis, tissue differentiation and regeneration, and embryogenesis.[1] Due to these activities, FGFs may also play an important role in carcinogenesis. This involvement in carcinogenesis is strongly suspected but is still, in early 1991, a matter of discussion. The FGFs can behave as oncogene products or potent angiogenic and mitogenic factors in *in vitro* cell cultures and in animal models.[2-4] Moreover, amplifications of some of the *FGF* (*INT2* and *HST/FGFK*) and *FGF* receptor (*FLG* and *BEK*) genes are observed in human carcinomas.[5-8] In cancer, each member of the couple growth factor–growth factor receptor represents a particular potential therapeutical target aimed at blocking the specific interaction. Therefore, it is paramount to characterize every member of the FGF and FGFR families.

The sixth member of the *FGF* family, the *FGF6* gene, was isolated a couple of years ago[9] by virtue of its sequence similarities with *FGF4* (*HST/FGFK*). The human[10] and murine[10] genes have been characterized, mapped on their respective chromosome, and their expression and transforming capacity studied. These results are the object of this report.

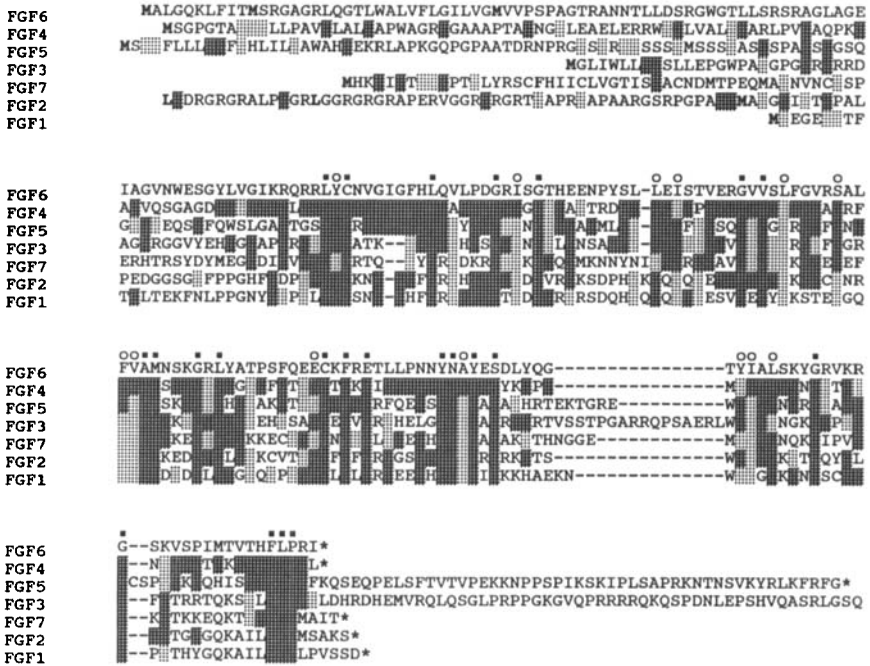## STRUCTURE OF THE *FGF6* GENE AND ENCODED PROTEINS

The *FGF6* gene is localized on chromosome 12, band p12, in the human and on chromosome 6, region F3-G1, in the mouse. It is composed of three coding exons separated by large introns. Human and mouse deduced amino acid sequences are shown in FIGURE 1. There is 93% amino acid sequence identity between the human and mouse gene products. At the nucleotide level (not shown), the calculated rate of nonsynonymous substitutions is 0.17×10^{-9} *per* site *per* year (assuming a time of divergence between human and rodent genes of 80×10^6 years ago). This is well below average[11] and indicates a strong conservation.

The localization of the splice junctions is similar to that of the other members of the family. There are four stop codons located immediately upstream of the beginning of the coding sequence, followed by three in-frame ATG codons. With the first ATG starts an open reading frame of 624 nucleotides, encoding a putative protein of 208 amino acids. A potential signal peptide, constituted of a hydrophobic region of 25 amino acids, starts shortly after the methionine residue corresponding to the second ATG, extends from position 16 (considering the first ATG as position 1) to position 40. Two potential proteolytic cleavage sites are located towards the end of the signal peptide. *In vitro* translation experiments (not shown) seem to indicate that it is the first site that is used. A potential asparagine-linked glycosylation site is immediately adjacent, at position 45. This site is functional in *in vitro* translation experiments (not shown).

The comparison of the deduced amino acid sequence of the human *FGF6* gene product with those of the other FGFs (FIGURE 2) shows that the closest relative is the *FGF4* gene product, as was expected considering the cross-reactivity of the respective molecular probes. FGFs have sequence similarities in a "core" portion coded by the 3' half of exon 1, exon 2, and exon 3, but differ greatly in the portion coded by the 5' half of exon 1. In addition, some FGF proteins present specific insertions and extensions. Paralogous FGF6 proteins have more similarities than FGF6 and FGF4 proteins of the same species; this is especially true when considering the portion coded by the 5' part of exon 1. It is possible that exon 1 corresponds to a merger between two ancestral exons, created as a result of the loss of an intron. The structural features of the *FGF6* gene product are summarized in FIGURE 3.

```
                          ████████████████████████        •
Human    MALGQKLFITMSRGAGRLQGTLWALVFLGILVGMVVPSPAGTRANNTLLDSRGWGTLLSR   60
Mouse    =====r=============v====Q======v=============a===G=============

                                              o
Human    SRAGLAGEIAGVNWESGYLVGIKRQRRLYCNVGIGFHLQVLPDGRISGTHEENPYSLLEI  120
Mouse    ==========g=========================================P=============

                                                        o
Human    STVERGVVSLFGVRSALFVAMNSKGRLYATPSFQEECKFRETLLPNNYNAYESDLYQGTY  180
Mouse    =============k====i=========t====Hd=====================R===

Human    IALSKYGRVKRGSKVSPIMTVTHFLPRI*                                 208
Mouse    ===========================*
```

FIGURE 1. Conservation between human and mouse FGF6 deduced amino acid sequences. The human FGF6 amino acid sequence is shown using the single-letter code. The mouse corresponding sequence is shown beneath using an = symbol for identical amino acids. Different but conserved amino acids are in lowercase letters. Specific features are indicated as follows: signal peptide, black bar; three potential methionine initiation codons, bold and underlined; *N*-glycosylation site and cysteine residues, bold, underlined with a symbol. The overall amino acid identity is 93.3% (97.6% when considering conservative changes).

```
FGF6   MALGQKLFITMSRGAGRLQGTLWALVFLGILVGMVVPSPAGTRANNTLLDSRGWGTLLSRSRAGLAGE
FGF4   MSGPGTA    LLPAV LAL APWAGR GAAAPTA NG LEAELERRW LVAL ARLPV AQPK
FGF5   MS  FLLL  F HLIL AWAH EKRLAPKGQPGPAATDRNPRG S R   SSS MSSS AS SPA S GSQ
FGF3                                              MGLIWLL SLLEPGWPA GPG R RRD
FGF7             MHK I T   PT LYRSCFHIICLVGTIS ACNDMTPEQMA NVNC SP
FGF2   L DRGRGRALP GRLGGRGRGRAPERVGGR RGRT APR APAARGSRPGPA  MA G I T PAL
FGF1                                               M EGE   TF
```

```
                *O*          *       * O *                O O      * * O      O
FGF6   IAGVNWESGYLVGIKRQRRLYCNVGIGFHLQVLPDGRISGTHEENPYSL-LEISTVERGVVSLFGVRSAL
FGF4   A VQSGAGD            A        G   A TRD-        F         A RF
FGF5   G   EQS FQWSLGA TGS   R        Y     N  A MI -     F SQ   G R F N
FGF3   AG RGGVYEH G AP R    ATK--     H S   N  I NSA  -       V    R F GR
FGF7   ERHTRSYDYMEG DI V    RTQ-- Y R DKR   K Q MKNNYNI   R  AV    K E EF
FGF2   PEDGGSG FPPGHF DP    KN  F R H     D VR KSDPH K Q Q E      K C NR
FGF1   T LTEKFNLPPGNY P L   SN -HF R     T D R RSDQH Q Q   ESV E Y KSTE GQ
```

```
       OO**    *    *     O*  *  *      **O*  *               OO O    *
FGF6   FVAMNSKGRLYATPSFQEECKFRETLLPNNYNAYESDLYQG----------------TYIALSKYGRVKR
FGF4   S     G  F T   T K I            YK P-----------------M    N  T
FGF5   SK    H AK T    RFQE S      A A HRTEKTGRE----------W    N R A
FGF3   K     EH SA E V R HELG      A R RTVSSTPGARRQPSAERLW     NGK P
FGF7   KE    KKEC    N L  E H      A AK THNGGE-----------M    NQK IPV
FGF2   KED   L KCVT  F F R GS      R RK TS---------------W    K T QY L
FGF1   D D L  G Q P  L L R EE H    I KKHAEKN-----------W    G K N SC
```

```
       *                 ***
FGF6   G--SKVSPIMTVTHFLPRI*
FGF4   --N     T K       L*
FGF5   CSP  K QHIS        FKQSEQPELSFTVTVPEKKNPPSPIKSKIPLSAPRKNTNSVKYRLKFRFG*
FGF3   --F TRRTQKS L       LDHRDHEMVRQLQSGLPRPPGKGVQPRRRQKQSPDNLEPSHVQASRLGSQ
FGF7   --K TKKEQKT         MAIT*
FGF2   --  TG GQKAIL       MSAKS*
FGF1   --P THYGQKAIL       LPVSSD*
```

FGF3   LEASAH*

FIGURE 2. Comparison between seven human FGF amino acid sequences. Sequences are aligned, using the single-letter code, to allow comparison with the FGF6 amino acid sequence. Identical amino acids are shown in dark shaded area and conserved amino acids in light shaded area. Initiation methionines are in bold, stops in asterisks. Black squares indicate seven identical residues; and circles, seven identical and conserved residues. FGFs are designated 1 to 7. The correspondence with commonly used names is as follows: FGF1 = aFGF or FGFA or HBGF1; FGF2 = bFGF or FGFB or HBGF2; FGF3 = INT2; FGF4 = HST or HSTF1 or K-FGF or FGFK; FGF7 = KGF. A 5' extension for INT2 has been demonstrated in mouse and probably also exists in human but was not represented. Sequences were taken from References 10 and 27–32.

The structure of the *FGF* genes is very similar in human and mouse. The overall phylogenetic conservation is however difficult to evaluate. A member of the *FGF* gene family has been cloned in frog.[12] We have looked for evidence of such a conservation in lower species using zoo blot hybridizations (FIGURE 4). A murine Fgf6 probe hybridizes well with the human *FGF6* gene as well as with the human *FGF4* (*HST/FGFK*) gene [lighter bands of 6 kilobases (kb) and 2.7 kb]. It lights up clear bands in chicken and frog DNAs. Bands are also present in the *Drosophila* lane. Except for human, it is not known presently whether these cross-reacting bands correspond to homologous genes.

In view of the high degree of sequence similarities existing between them, we can readily surmise that the *FGF* genes are homologues. A phylogenetic tree based on sequence data is shown in FIGURE 5A. *FGF* genes have probably evolved by

successive duplications but the evolution may have followed complex schemes at certain stages. Among many possibilities, two such schemes could be hypothesized. In the first, a recent duplication is at the origin of *FGF3* (*INT2*) and *FGF4* (*HST/FGFK*), both located on chromosomal band 11q13.3, separated by less than 40 kb.[13] Then, *FGF3* evolved separately and rapidly diverged from *FGF4,* which remained very similar to *FGF6.* The second scheme hypothesizes a putative gene similar to *FGF3* on human chromosome 12. In FIGURE 5, we have tentatively depicted these hypotheses (note that proteins and not gene sequences were used in this figure).

TRANSFORMING CAPACITY OF CLONED *FGF6* GENE

The normal *FGF6* gene has a transforming capacity comparable to that of *FGF4.*[14-16] This was tested by bioassays involving the transfection of human and murine cloned genes into cultured murine NIH 3T3 fibroblasts followed by either a tumorigenicity assay or a focus assay. Clones of the normal *FGF6* gene were either genomic cosmid clones or cDNA clones inserted into mammalian expression vectors. It is noteworthy that the sequence encoding the signal peptide is necessary for the transforming activity of the transfected *FGF6* gene (Coulier *et al.,* in press). Supernatants from mass culture of the transformed cells and proteins from *in vitro* translated *FGF6* gene are mitogenic for fibroblasts (FIGURE 6). Moreover, *FGF6-*transformed cells are tumorigenic when injected into nude mice.

Despite this intrinsic transforming capacity, the *FGF6* gene has never been detected so far in transfection assays, in contrast to *FGF4* which has been frequently identified in these assays.[17,18] Furthermore, the gene has never been found altered in human tumors. The intrinsic oncogenic capacity of the *FGF* genes must be repressed *in vivo.* However, the mechanisms of this repression are not known.
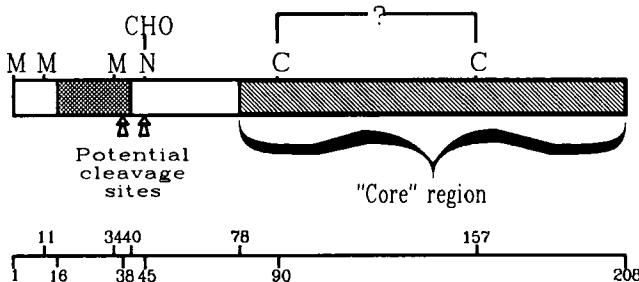
FIGURE 3. Structure of the FGF6 protein. The structure of the FGF6 protein was predicted from cDNA sequence and *in vitro* translation experiments. A long open reading frame starting with three in-frame ATG codons is able to code for a protein of 208 residues. Only the two first ATG codons seem to be used for translation initiation. A stretch of 25 hydrophobic residues, indicated by a hatched box (a bar in FIGURE 1), may act as a signal peptide. Two potential cleavage sites are indicated. A unique N-linked glycosylation site, which appears to be used *in vitro,* is indicated by N. Two cysteine residues, conserved among the seven FGFs described to date, are indicated; they may be involved in a disulfide bridge. A "core" region (hatched) represents a region having sequence similarities in all members. The corresponding amino acid positions are indicated below.
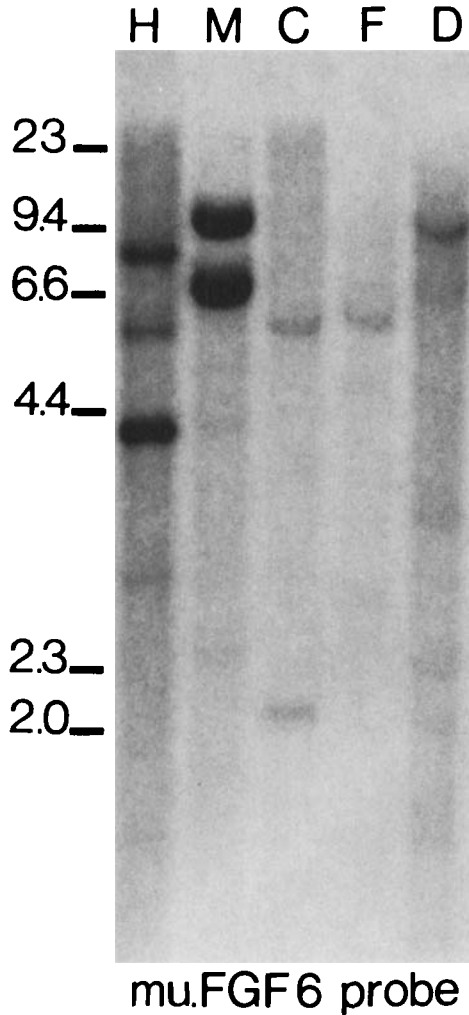
FIGURE 4. Conservation of the *FGF6* gene. Southern "zoo blot" hybridization was carried out at 42°C, in 35% formamide, 5 × SSC, 5 × Denhardt's, on 10 μg of DNA from various species as follows: H, human; M, mouse; C, chicken; F, frog (*Xenopus laevis*); D, fly (*Drosophila melanogaster*). The probe used (probe PN) was a murine *Fgf6* cDNA spanning the three coding exons.

EXPRESSION OF THE *FGF6* GENE IN ADULT AND EMBRYONIC TISSUES

FGF genes are expressed in a wide variety of cells, in the adult as well as in the embryo. *FGF* genes present various patterns of expression in the developing embryo, where they could play an important physiological role as inducers of cell determination and migration.[12,19]

Fgf6 expression was analyzed by northern blot hybridization in the adult and in

the embryonic mouse.[10] In adult tissues the *Fgf6* gene is expressed in skeletal muscle and heart, but not in smooth muscle. An *Fgf6* transcript was also detected in adult testis. Use of northern blot hybridization with sense and antisense RNA probes showed that this testis-specific expression corresponds to an antisense transcript from the *Fgf6* locus (FIGURE 7). Presently, nothing is known of its structure and role.
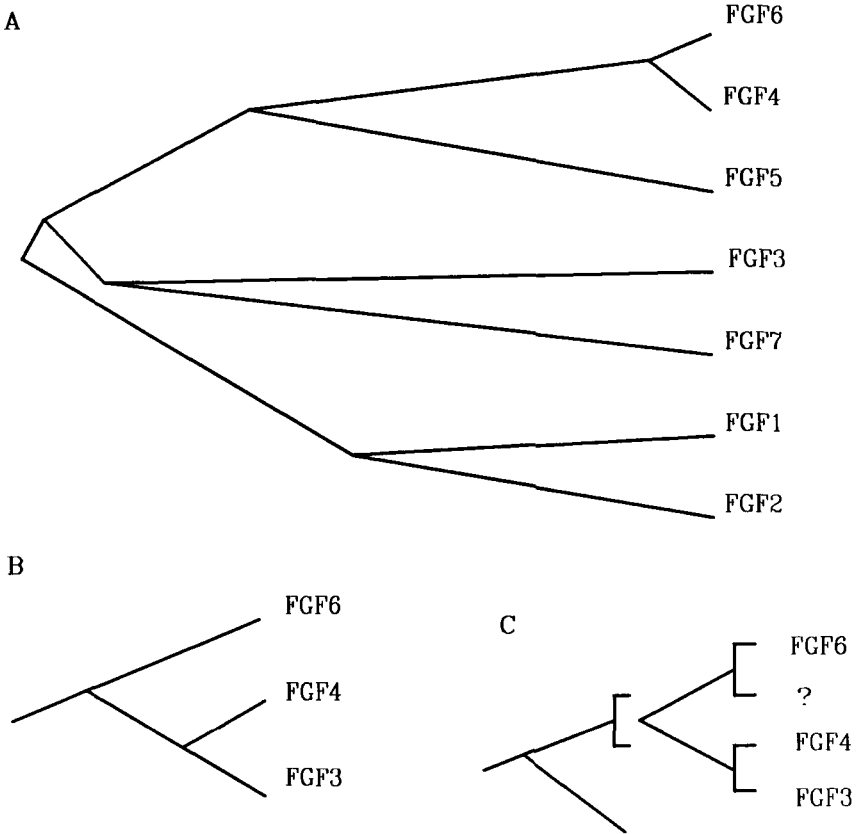

FIGURE 5. Phylogenetic FGF protein trees. **A:** Unrooted phylogenetic tree inferred from the amino acid sequences of the FGFs. The operational units (OU) present at the external nodes are the seven FGF proteins with names used in FIGURE 2. The tree shows one topology only (out of 954 theoretically possible given 7 OU; Reference 11). Branches are scaled, i.e., their lengths are proportional to the number of substitutions. **B** and **C:** portions of unrooted trees inferred from sequence and chromosomal localization data, with focus on three OU; *FGF3* and *FGF4* are clustered on chromosomal band 11q13,[13] while *FGF6* lies on band 12pter.[10] In these particular trees, a specific rapid divergence of FGF3 (B) or a cluster duplication (C) is hypothesized.

The presence of an antisense transcript originating from an *FGF* locus has already been noted in the case of *FGF2* in *Xenopus laevis*.[20]

In mouse embryos, *Fgf6* is expressed during middle and late gestation with a peak at day 15. The transcript has a size of 4.8 kb. As a rule, the expression level is low, necessitating use of poly(A)⁺ RNA to be detected.
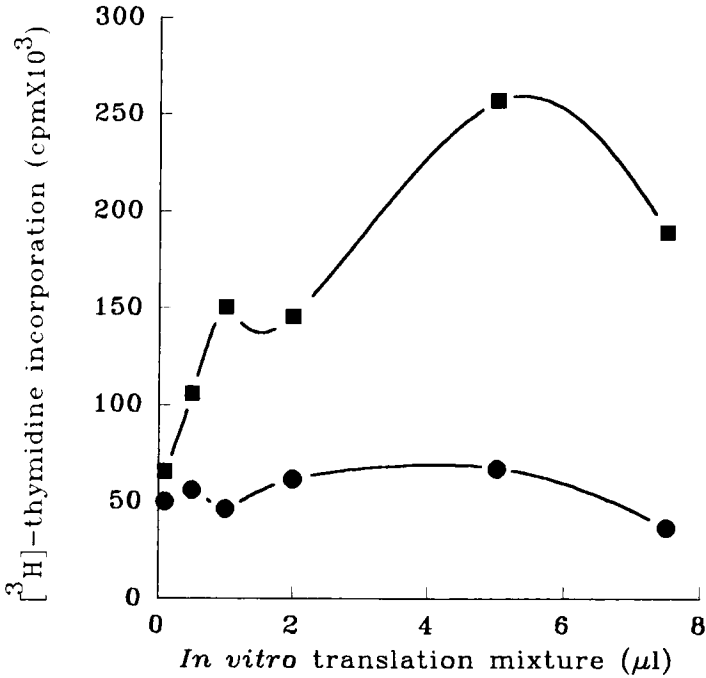
FIGURE 6. Mitogenic activity of *in vitro* translated FGF6 peptides. RNAs transcribed *in vitro* were used in a rabbit reticulocyte lysate system, in the presence of canine pancreatic microsomal membranes, to translate FGF6 peptides. Increasing amounts of *in vitro* translation mixtures were tested for stimulation of [³H]thymidine incorporation of Balb/c 3T3 cells. Samples tested included translation products from antisense (circles) or sense (squares) transcripts.
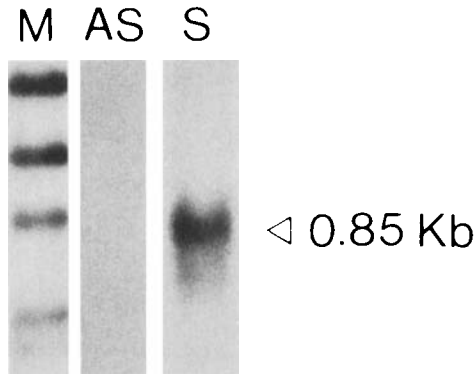

FIGURE 7. Detection of an antisense transcript from the *FGF6* locus by northern blot hybridization. Poly (A⁺) RNA was isolated from 2-month-old mouse testis and hybridized with [³²P]CTP-labeled RNA, transcribed either in the *Fgf6* sense orientation by T7 (lane S) or in the antisense orientation by T3 (lane AS) polymerase from the appropriately linearized cDNA clone p173. This clone represents the mouse *Fgf6* gene first exon. Markers on the left represent Hae III DNA fragments from phage PhiX 174. The sense probe reveals an antisense transcript of 0.85 kilobase pair (kbp) (arrowhead).

CONCLUSION

When the number of the known FGFs suddenly rose from two to seven, it became important to define their specific role and how they assume it. Characterization of the FGF receptors could have helped in resolving this question. However, the analyses of ligand-receptor interactions reveal a peculiar pattern characterized by an ambiguity in the interacting FGF-FGFR couples; for the time being this contributes to darkening the picture a little more. How the cells discriminate between the various FGFs remains unknown. Furthermore, nothing is known about possible interactions between FGFs themselves. Is there such a thing as a coordinate regulation of *FGF* gene expression, at least during certain periods of development, under certain circumstances, or in certain areas? Study of the *FGF6* gene has so far only introduced an element of complexity into these problems. It has shown that some members of the family can be more closely related than the average. It has confirmed that some FGF genes do present an intrinsic transforming activity even if their role in carcinogenesis—especially of those having this capacity—is far from being established. Finally, the restricted pattern of expression of the gene points to a specific activity of FGF6 in the embryo and in cardiac and skeletal muscle cells. Regulating muscular activity or development appears to be an important function of the FGFs, and members of the family seem to exert a control on myoblasts.[21-26] The precise role of FGF6 with respect to muscle function will have to be defined.

ACKNOWLEDGMENTS

REFERENCES

1. GOLDFARB, M. 1990. Cell Growth Diff. **1:** 439–445.
2. DICKSON, C., R. SMITH, S. BROOKES & G. PETERS. 1984. Cell **37:** 529–536.
3. MULLER, W., F. LEE, C. DICKSON, G. PETERS, P. PATTENGALE & P. LEDER. 1990. EMBO J. **9:** 907–913.
4. HALABAN, R., R. LANGDON, N. BIRCHALL, C. CUONO, A. BAIRD, G. SCOTT, G. MOELLMAN & J. McGUIRE. 1988. J. Cell. Biol. **107:** 1611–1619.
5. LIDEREAU, R., R. CALLAHAN, C. DICKSON, G. PETERS, C. ESCOT & I. U. ALI. 1988. Oncogene Res. **2:** 285–291.
6. FANTL, V., M. RICHARDS, R. SMITH, G. LAMMIE, G. JOHNSTONE, D. ALLEN, W. GREGORY, G. PETERS, C. DICKSON & D. BARNES. 1990. Eur. J. Cancer **26:** 423–429.
7. TSUDA, T., E. TAHARA, G. KAJIYAMA, H. SAKAMOTO, M. TERADA & T. SUGIMURA. 1989. Cancer Res. **49:** 5505–5508.
8. ADNANE, J., P. GAUDRAY, C. DIONNE, G. CRUMLEY, M. JAYE, J. SCHLESSINGER, P. JEANTEUR, D. BIRNBAUM & C. THEILLET. 1991. Oncogene **6:** 659–663.
9. MARICS, I., J. ADELAIDE, F. RAYBAUD, M.-G. MATTEI, F. COULIER, J. PLANCHE, O. DELAPEYRIERE & D. BIRNBAUM. 1989. Oncogene **4:** 335–340.
10. DELAPEYRIERE, O., O. ROSNET, D. BENHARROCH, F. RAYBAUD, S. MARCHETTO, J. PLANCHE, F. GALLAND, M. G. MATTEI, N. G. COPELAND, N. A. JENKINS, F. COULIER & D. BIRNBAUM. 1990. Oncogene **5:** 823–831.
11. LI, W-H. & D. GRAUR. 1991. Fundamentals in Molecular Evolution. Sinauer Assoc. Sunderland, Mass.
12. KIMELMAN, D., J. ABRAHAM, T. HAAPARANTA, T. PALISI & M. KIRSCHNER. 1988. Science **242:** 1053–1056.

13. NGUYEN, C., D. ROUX, M. G. MATTEI, O. DELAPEYRIERE, M. GOLDFARB, D. BIRNBAUM & B. JORDAN. 1988. Oncogene **3**: 703–708.
14. SAKAMOTO, H., T. YOSHIDA, M. NAKAKUKI, H. ODAGIRI, K. MIYAGAWA, T. SUGIMURA & M. TERADA. 1988. Biochem. Biophys. Res. Commun. **151**: 965–972.
15. MIYAGAWA, K., H. SAKAMOTO, T. YOSHIDA, Y. YAMASHITA, Y. MITSUI, M. FURUSAWA, S. MAEDA, F. TAKAKU, T. SUGIMURA & M. TERADA. 1988. Oncogene **3**: 383–389.
16. DELLI BOVI, P., A. M. CURATOLA, K. NEWMAN, Y. SATO, D. MOSCATELLI, R. HEWICK, D. RIFKIN & C. BASILICO. 1988. Mol. Cell. Biol. **8**: 2933–2941.
17. SAKAMOTO, H., M. MORI, M. TAIRA, T. YOSHIDA, S. MATSUKAWA, K. SHIMIZU, M. SEKIGUCHI, M. TERADA & T. SUGIMURA. 1986. Proc. Nat. Acad. Sci. USA **83**: 3997–4001.
18. DELLI BOVI, P. & C. BASILICO. 1987. Proc. Nat. Acad. Sci. USA **84**: 5660–5664.
19. WILKINSON, D., G. PETERS, C. DICKSON & A. MCMAHON. 1988. EMBO J. **7**: 691–695.
20. VOLK, R., M. KOSTER, A. POTING, L. HARTMANN & W. KNOCHEL. 1989. EMBO J. **8**: 2983–2988.
21. GOSPODAROWICZ, D., J. WESEMAN, J. MORAN & J. LINDSTROM. 1976. J. Cell. Biol. **70**: 395–405.
22. ALLEN, R., M. DODSON & L. LUITEN. 1984. Exp. Cell. Res. **152**: 154–160.
23. LATHROP, B., K. THOMAS & L. GLASER. 1985. J. Cell. Biol. **101**: 2194–2198.
24. SEED, J. & S. HAUSCHKA. 1988. Dev. Biol. **128**: 40–49.
25. VAIDYA, T., S. RHODES, E. TAPAROWSKY & S. KONIECZNY. 1989. Mol. Cell. Biol. **9**: 3576–3579.
26. ALTERIO, J., Y. COURTOIS, J. ROBELIN, D. BECHET & I. MARTELLY. 1990. Biochem. Biophys. Res. Commun. **166**: 1205–1212.
27. JAYE, M., R. HOWK, W. BURGESS, G. RICCA, I. M. CHIU, M. RAVERA, S. O'BRIEN, W. MODI, T. MACIAG & W. DROHAN. 1986. Science **233**: 541–544.
28. PRATS, H., M. KAGHAD, A. C. PRATS, M. KLAGSBRUN, J. M. LELIAS, P. LIAUZUN, P. CHALON, J. P. TAUBER, F. AMALRIC, J. SMITH & D. CAPUT. 1989. Proc. Nat. Acad. Sci. USA. **86**: 1836–1840.
29. BROOKES, S., R. SMITH, G. CASEY, C. DICKSON & G. PETERS. 1989. Oncogene **4**: 429–436.
30. TAIRA, M., T. YOSHIDA, K. MIYAGAWA, H. SAKAMOTO, M. TERADA & T. SUGIMURA. 1987. Proc. Nat. Acad. Sci. USA. **84**: 2980–2984.
31. ZHAN, X., B. BATES, X. HU & M. GOLDFARB. 1988. Mol. Cell. Biol. **8**: 3487–3495.
32. FINCH, P., J. RUBIN, T. MIKI, D. RON & S. AARONSON. 1989. Science **245**: 752–755.