MOLECULAR AND CELLULAR BIOLOGY, Sept. 1991, p. 4627–4634
0270-7306/91/094627-08$02.00/0

Vol. 11, No. 9

The Human Fibroblast Growth Factor Receptor Genes: a Common Structural Arrangement Underlies the Mechanisms for Generating Receptor Forms That Differ in Their Third Immunoglobulin Domain

DANIEL E. JOHNSON, JOHN LU, HELEN CHEN, SABINE WERNER, AND LEWIS T. WILLIAMS*

Howard Hughes Medical Institute, Program of Excellence in Molecular Biology, and Cardiovascular Research Institute, University of California, San Francisco, California 94143-0724

Received 25 March 1991/Accepted 17 June 1991

To determine the mechanisms by which multiple forms of fibroblast growth factor (FGF) receptors are generated, we have mapped the arrangement of exons and introns in the human FGF receptor 1 (FGFR 1) gene (*flg*). We found three alternative exons encoding a portion of the third immunoglobulin (Ig)-like domain of the receptor. One of these alternatives encodes a sequence that is part of a secreted form of FGFR 1. The other two encode sequences that are likely part of transmembrane forms of FGFR 1. One of these forms has not been previously reported in published cDNAs. Also, we have determined the structural organization of a portion of the human FGFR 2 gene (*bek*) and found a similar arrangement of alternative exons for the third Ig-like domain. The arrangement of these genes suggests that there are conserved mechanisms governing the expression of secreted FGF receptors as well as the expression of at least two distinct membrane-spanning forms of the FGF receptors. The diverse forms appear to be generated by alternative splicing of mRNA and selective use of polyadenylation signals.

Acidic fibroblast growth factor (aFGF) and basic fibroblast growth factor (bFGF) were initially characterized and purified on the basis of their mitogenic activities toward fibroblasts (7, 8). Subsequent studies have shown that FGFs stimulate a remarkably diverse group of biological responses (reviewed in references 3, 10, and 35), including proliferation of endothelial cells (9, 18), neurite outgrowth (24, 36, 37), survival of lesioned cholinergic neurons in vivo (1), and early steps of embryonic development (15, 16, 33).

In recent years, the complexity of the FGF system has become apparent with the discovery of at least seven genes that encode factors in the FGF family. The factors that have been identified include aFGF, bFGF, the product of the *int*-2 oncogene (23), the product of the *hst* oncogene (Kaposi sarcoma FGF) (4, 34), FGF-5 (39), FGF-6 (20), and keratinocyte growth factor (KGF) (29). There are also multiple genes that encode FGF receptors (5, 11, 12, 14, 17, 22, 25, 26). We have referred to the known FGF receptors as FGF receptors 1, 2, and 3 (FGFR 1 to 3).

The first identification of a full-length cDNA for an FGF receptor was based on experiments in which a functional FGF receptor was purified from chicken embryos (17). The cDNA encoding this purified protein was closely related to a previously published partial cDNA (30) of a *fms*-like gene (*flg*) that had been isolated on the basis of tyrosine kinase sequence homologies. It is now apparent that the FGF receptor and its related genes form a subclass of tyrosine kinase receptors that are different from the Fms/platelet-derived growth factor receptor group and other tyrosine kinase receptors. We refer to this first FGF receptor as FGFR 1. This gene is identical to the *flg* (human) (30) and *Cek*1 (chicken) genes (26). The chicken cDNA clones that were isolated for FGFR 1 were predicted to encode a protein that has three immunoglobulin (Ig)-like domains in the extracellular region, a single membrane-spanning segment, and a tyrosine kinase domain in the cytoplasmic region (Fig. 1 shows a schematic diagram of the receptor protein). In subsequent studies, human and mouse cDNAs encoding three-Ig-domain forms of FGFR 1 were isolated (5, 28). However, additional cDNA clones were found to encode FGFR 1 proteins containing only two Ig domains (13, 28) (Fig. 1). These two domains corresponded to the second and third Ig domains of the three-Ig-domain form of the receptor. Functional characterization of the cloned FGFR 1 proteins showed that both the three- and two-Ig-domain forms of the receptor can bind and become activated by either aFGF or bFGF (5, 13). These experiments demonstrated that (i) a single receptor protein can bind different members of the FGF family and (ii) the first Ig domain of the FGF receptor is not essential for binding aFGF and bFGF.

In addition to two- and three-Ig-domain forms of FGFR 1 that span the membrane, we isolated a cDNA encoding a secreted form of the receptor (13) (Fig. 1). The secreted form of the receptor contains two Ig-like domains and is identical to the membrane-spanning forms of the receptor except for the carboxy-terminal half of the third Ig domain. The sequence of this region of the secreted receptor lacks a hydrophobic membrane-spanning region and terminates at a point that would produce a secreted protein that lacks a cytoplasmic domain.

These findings have suggested that the expression of different forms of the FGF receptor is regulated by alternative splicing of mRNA. In this study, we have investigated the organization of exons and introns in the human FGFR 1 gene. This information has enabled us to determine the origin of sequences that are unique to mRNA transcripts encoding the secreted form of the receptor. Furthermore, we discovered an additional exon that encodes an alternative

* Corresponding author.

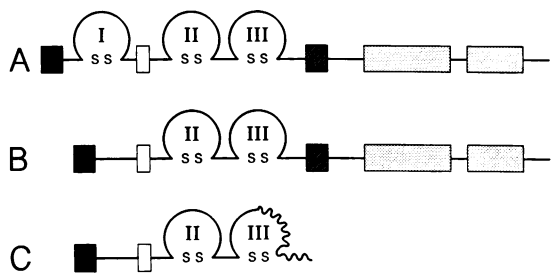

FIG. 1. Schematic representations of the three-Ig-domain (A) and the two-Ig-domain (B) forms of the membrane-spanning FGFR 1 and of the two-Ig-domain form of the secreted receptor. The following structural features are identified: hydrophobic leader sequence (solid squares), highly acidic domain (open squares), transmembrane domain (striped squares), kinase 1 and kinase 2 domains (stipled rectangles), and the divergent amino acid sequence specific for the secreted receptor form (wavy line).

sequence for the second half of the third Ig domain. Therefore, there are three possible coding sequences for this region of the receptor molecule.

Recent studies have shown that the product of the FGFR 2 gene (variously called bek [5], Cek3 [25], K-sam [11], TK14 [12], and KGF receptor [22]) also is capable of binding both aFGF and bFGF (5, 22). A naturally occurring variant of FGFR 2 (the KGF receptor [22]) that lacks some sequence in the extracellular region is activated by KGF. Overall, FGFR 2 is similar in structure and sequence to FGFR 1. As is the case for FGFR 1, both three- and two-Ig-domain forms of the FGFR 2 protein have been identified. We have determined the structural organization of the human FGFR 2 gene in the region encoding the third Ig domain. These studies indicate that the FGFR 2 gene is organized in a fashion nearly identical to that of the FGFR 1 gene. In both cases, multiple forms that differ in their third Ig domain are expressed. The expression of each unique form can be attributed to either alternative splicing or selective use of polyadenylation signals.

MATERIALS AND METHODS

Cell lines. Human foreskin fibroblast cells, human Ewing's sarcoma cells (ATCC HTB 166), human astrocytoma cells (grade IV; ATCC CRL 1718, CCF-STTG1), and human neuroblastoma cells (SH-EP) were grown in RPMI 1640 (GIBCO) containing 10% fetal calf serum. Human adrenal cortex adenocarcinoma cells (ATCC CCL 105) were grown in L-15 medium (GIBCO) containing 10% fetal calf serum. Human kidney endothelial cells were grown on gelatin-coated plates in Dulbecco modified Eagle medium (Irvine Scientific) supplemented with 10% fetal calf serum, 10 U of heparin per ml, and 30 μg of endothelial cell growth supplement (Collaborative Research Inc.) per ml. Rat L6 cells transfected with expression constructs containing cDNAs encoding either the human FGFR 1 secreted form or the previously published transmembrane form of FGFR 1 (13) were grown in Dulbecco modified Eagle medium containing 10% fetal calf serum.

Isolation of human genomic DNA. Human genomic DNA used as a template for polymerase chain reactions (PCR) was isolated from human foreskin fibroblast cells. High-molecular-weight DNA was isolated according to previously published protocols (19).

PCR. Human genomic DNA was amplified by using 750 ng of human genomic DNA as a template for PCR (31). Each reaction mix also contained 10 pmol of each primer, 200 μM each of the four deoxynucleoside triphosphates, and 1 U of *Taq* polymerase (Perkin Elmer Cetus) in 50 μl of 10 mM Tris hydrochloride (pH 8.3)–50 mM KCl–1.5 mM $MgCl_2$–100 ng of bovine serum albumin per ml. Reactions were carried out in an Ericomp twin block system. Thirty-one cycles, consisting of denaturation at 94°C for 50 s, annealing at 60 to 65°C for 1 min, and extension at 72°C for 3 min, were performed.

Using PCR, we amplified, subcloned, and sequenced 13 overlapping human genomic fragments, each of which contained at least one but not more than two introns. The primers were based on human FGFR 1 cDNA sequences (13). Listed below are the primer combinations, starting at the 5′ end of the coding region, that were used to amplify the 13 overlapping genomic fragments (fragment numbers are underlined). For each primer pair, the 5′ primer is listed first, followed by the 3′ primer. Indicated in parentheses is the exon number from which the primers were derived (Fig. 2). In some cases, primers were derived from intronic sequences. In these cases, the location of the intron from which the primer was derived is indicated. The primer pairs are as follows: 5′-GA(T/C)GACGTGCAG(A/T)(G/C)CATC AACTGGGTGCGTGATGG-3′ (exon 1) and 5′-CATCTTT TCTGGGGATGTCCAATATGGA-3′ (exon 3); 5′-CTCTTC

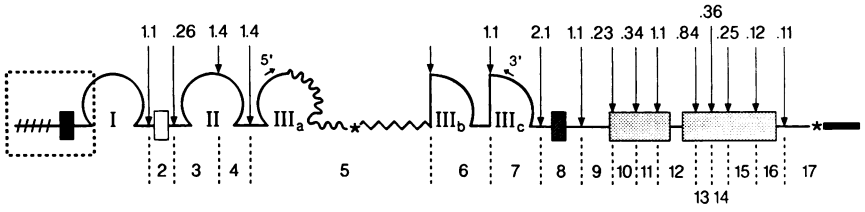

FIG. 2. Structure and organization of the human FGFR 1 gene, showing the arrangement of exons and introns. Arrows and vertical dashed lines indicate the positions of intron sequences. Numbers above the arrows indicate the size of the intron in kilobases. Numbers below the coding region and between the vertical dashed lines indicate the assigned exon numbers. The 5′ region of the receptor gene which remains uncharacterized is boxed. The size of the intron between exons 5 and 6 is not known because it is not clear where exon 5 ends. The approximate end of exon 6 was determined by comparison with the analogous exon from FGFR 2. The following structural features are identified: 5′ nontranslated sequence (cross-hatched line), hydrophobic leader sequence (solid square), highly acidic domain (open square), sequences encoding the divergent amino acid sequence which is specific for the secreted receptor (wavy line), transcribed sequences which become the 3′ nontranslated sequence unique to mRNA transcripts encoding the secreted receptor form (zigzag line), and transcribed sequences which become the 3′ nontranslated sequence specific for transcripts encoding membrane-spanning receptor forms (thick black line). The three alternative coding sequences for the second half of Ig domain III are indicated as IIIa, IIIb, and IIIc.

AGAGGAGAAAGAAACAGATA-3′ (exon 2) and 5′-CCG TACTCATTCTCCACAATGCA-3′ (exon 4); 5′-TATGC CACCTGGAGCATCATAATGGACTCTGTGGTGCCCTC TGAC-3′ (exon 4) and 5′-GACTGGCCCACGAAGACTG GTGCCAT (exon 5, derived from secreted form specific sequence); 5′-CAGACAACCTGCCTTATGTCCAGATCT-3′ (exon 5; labeled 5′ in Fig. 2) and 5′-CATCTCTTTGTCG GTGGTATTAACTCCA-3′ (exon 7; labeled 3′ in Fig. 2); 5′-CTTAGAGGTCTCATGTCCTGTGCTTGT-3′ (derived from sequence at the 3′ end of the intron directly preceding exon 7) and 5′-TCACTCCAGGTACAGGGGCGAGGT-3′ (exon 8); 5′-TGCTAACACCCTGTTCGCACT-3′ (derived from sequence at the 3′ end of the intron directly preceding exon 8) and 5′-GAGGCAGCTCCCAGCGA-3′ (exon 9); 5′-CTGACTCCAGTGCATCCAT-3′ (exon 9) and 5′-CCACCT GCCCAAAGCAGCCCTCTCCCAGGGGTTTGCCTAA-3′ (exon 10); 5′-CTTCACCAGCCCCAACTTATGCCACT-3′ (derived from sequence at the 3′ end of the intron directly preceding exon 10) and 5′-CAGCAGGTTGATGATA-3′ (exon 11); 5′-GCAACAGAGAAAGACTTGTCAGACCT-3′ (exon 11) and 5′-CTTGGAGGCATACTCCACGATGAC AT-3′ (exon 12); 5′-CAAACCTGCTCACCTGCTGCT-3′ (derived from sequence at the 3′ end of the intron directly preceding exon 12) and 5′-TCGATGTGGTGAATGTC-3′ (exon 13); 5′-CAGGTCTCGGTGTATGCACTTCT-3′ (exon 13) and 5′-CACTCTGGTGGGTGTAGAT-3′ (exon 14); 5′-GACTGCCTGAAGTGGAT-3′ (exon 14) and 5′-CATGTA CAGCTCGTTGGTGCAGTTA-3′ (exon 15); and 5′-CTG TGGGAGATCTTCACT-3′ (exon 15) and 5′-GAGAGGTGA GCTGAGTG-3′ (derived from 3′ nontranslated sequence associated with membrane-spanning receptor forms).

To characterize the region of the FGFR 2 gene encoding Ig domain III, we used PCR to generate three overlapping genomic fragments. Primers were designed on the basis of the published sequences of FGFR 2 cDNAs (5, 11). The primer combinations used to generate these three fragments were as follows: 5′-GAAAAGAACGGCAGTAAAT-3′ (exon 5) and 5′-CACTTCTGCATTGGAACTAT-3′ (exon 6); 5′-GACCATAGACAATGCTAAGA-3′ (derived from sequence at the 3′ end of the intron directly preceding exon 6) and 5′-GAGAACCTCAATCTCTTTGT-3′ (exon 7); and 5′-ATCATTCCTGTGTCGTCTA-3′ (derived from sequence at the 3′ end of the intron directly preceding exon 7) and 5′-TCATAGTCTGGGGAAGCTGTAATCT-3′ (exon 8).

Following the final extension of PCR reactions, 1 U of Klenow enzyme was added to each reaction mix, and extension was allowed to continue for 30 min at room temperature. Reaction products were then subcloned into the *Sma*I site of Bluescript KS (Stratagene), and subclones were sequenced (32) to determine the precise locations of intronic sequences.

Isolation of human RNA. Polyadenylated human RNA was isolated from cells by using a Fast Track kit (Invitrogen) as specified by the manufacturer. Total cellular RNA was isolated from cells by using guanidinium isothiocyanate as previously described (6).

Probes used for RNase protection assays. DNA probes specific for each of the three alternative coding sequences for the second half of Ig domain III (IIIa, IIIb, and IIIc) were subcloned into Bluescript KS. The secreted form-specific probe, IIIa, represents a 410-bp *Acc*I-*Sal*I fragment of the published secreted-form cDNA (13) (h5 cDNA). This fragment was subcloned into the *Acc*I site of Bluescript KS. The IIIb-specific probe was obtained by PCR amplification of a 140-bp fragment from the IIIb exon, using the primers 5′-CATTCGGGGATTA-3′ and 5′-TCTGGTGACAGTGAG CCA-3′. The IIIc-specific probe was obtained by PCR amplification of a 142-bp fragment from the IIIc exon, using the primers 5′-ACTGCTGGAGTTAATACCAC-3′ and 5′-TTCC AGAACGGTCAACCAT-3′. Both the IIIb and IIIc DNA probes were subcloned into the *Sma*I site of Bluescript KS.

RNase protection analyses. For RNase protection mapping of FGFR 1 transcripts, we used antisense RNA probes transcribed in vitro from the above-mentioned plasmid DNA templates. Before transcription, plasmid templates were linearized at unique sites in the Bluescript KS polylinker. Antisense RNAs were then transcribed by using T3 or T7 RNA polymerase and [^{32}P]rUTP (800 Ci/mmol; Amersham) as previously described (21). All three probes described above were transcribed to approximately the same specific activity (10^8 cpm/μg of linearized DNA template). Samples of polyadenylated RNA (1 μg) or total cellular RNA (20 μg) were hybridized with 10^5 cpm of the labeled antisense RNA for 12 h at 42°C. As a negative control, probes were hybridized with 20 μg of tRNA (*Escherichia coli*). Hybrids were digested with RNase A and RNase T_1 for 40 min at 30°C as described by Melton et al. (21). Protected fragments were separated on 5% polyacrylamide–8 M urea gels and analyzed by autoradiography.

RESULTS

Structural organization of the human FGFR 1 gene. Primers based on the sequence of the human FGFR 1 cDNA (13) were used to amplify regions of the gene, using human genomic DNA as a template for PCR (31). Using this approach, we amplified, subcloned, and sequenced 13 overlapping genomic fragments, each of which contained at least one but not more than two introns. The overlapping fragments covered the region extending from the amino-terminal half of Ig domain I to the 3′ nontranslated region associated with the membrane-spanning receptor forms (Fig. 2). Figure 2 indicates the locations and sizes of intronic sequences in the FGFR 1 coding region. We were unable to amplify any fragments with use of a primer derived from the 5′ nontranslated region or signal peptide sequence and a primer derived from Ig domain I, indicating that there may be a large intron(s) located in this region. For this reason, we started our numbering of exons with number 2 (Fig. 2). Figure 3 shows the nucleotide sequence of each intron-exon boundary. In general, the observed splice junction sequences were similar to published consensus sequences (2). However, in some cases the junction sequences differed from the consensus sequences (Fig. 3).

Genomic sequences involved in the generation of a secreted form of FGFR 1. Previously we have shown that at least two forms of FGFR 1 that differ in the third Ig domain are expressed (13). One of these forms represents a secreted extracellular domain that lacks transmembrane sequences, whereas the other form represents a membrane-spanning form of the receptor. The coding region of the secreted FGFR 1 form diverges completely from the sequence of the membrane-spanning forms at a position approximately halfway through Ig domain III (Fig. 1). Also, cDNAs encoding the secreted form of receptor contain approximately 0.95 kb of 3′ nontranslated sequence which is distinct from the 3′ nontranslated sequences of cDNAs encoding membrane-spanning receptors. When a primer derived from the amino-terminal half of Ig domain III (labeled 5′ in Fig. 2) and a primer derived from the carboxyl-terminal half of the membrane-spanning form of Ig domain III (labeled 3′) were used to amplify genomic DNA, a single 2.6-kb fragment was

Intron		Exon		Intron
		1....AATGTTTCAG N V S	/	GTTGGT
TCTTCTTCCCAT	/	AGATGCTCTC.....2....AACCGTATGC D A L N R M	/	GTGAGA
TTCCCCGGGCAG	/	CCGTAGCT.....3...GGCTACAAGGT P V A G Y K	/	ACGTGT
GCATTTCCCCTC	/	AGGTCCGTTAT....4....GATGTCGTGG V R Y D V V	/	GTAAGA
CCATCTTCCACA	/	GAGCGGTCC.....5 E R S		
TTGGTTTGTTCC	/	AGCATTCGGGG....6 H S G		
GTTGTCTCTTCT	/	AGACTGCTGGA....7....GTTCTGGAAG T A G V L E	/	GTACAC
TCGCACTGACTC	/	AGCCCTGGAA.....8....CAGGTAACAG A L E Q V T	/	AAAGTA
TCTTCCTTGCAG	/	GTGTCTGCT.....9...CCTCGGGACAG V S A P R D	/	GTAATA
GTTTCCCCCGAA	/	AGACTGGTC.....10....ATGTTGAAGT R L V M L K	/	GTAAGT
TGTTCTTTAAAG	/	CGGACGCA.....11.....CAGGATGGT S D A Q D G	/	GGGTGC
CCTGTGCCTGCA	/	GGTCCCTTG.....12.....TCCAAGAAG G P L S K K	/	GTGTGG
TCCTCCTTCCTC	/	AGTGCATACAC...13....ACAACCAACG C I H T T N	/	TGAGTG
TGCTGCTTTCAG	/	GGCCGACTG.....14.....AGTGATGTGTG G R L S D V	/	AGTTTG
GCCCATTTGCAG	/	GTGGTCTTTC....15...AACGAGCTGTA W S F N E L	/	AGCCCG
GTCCTGCCGCAG	/	GTACATGATG....16....TCCAACCAGG Y M M S N Q	/	TAAGGC
TCTCGCCCATCA	/	CAGGAGTAC.....17 Q E Y		

FIG. 3. Nucleotide junctions of introns and exons within the coding region of the human FGFR 1 gene. The first three and last three amino acids of each exon are translated to enable precise positioning relative to the published amino acid sequences of FGFR 1 (13). Since it is not clear where exons 5 and 6 end, the junction sequences at the ends of these exons are not shown. The exact 3′ end of exon 17 also is not known. For comparison, a consensus form for intron-exon boundaries is intron....$(Py)_8$NPyAG/G....exon....AG/GTPuAG....intron (2), where Py is pyrimidine and Pu is purine.

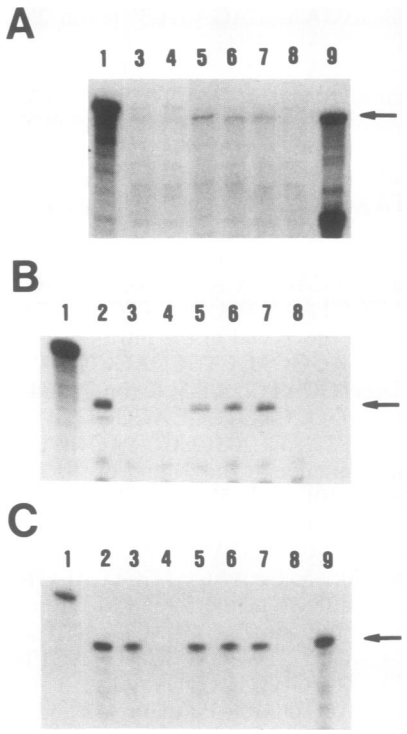

FIG. 4. Expression of FGFR 1 mRNA transcripts encoding different sequences for the second half of Ig domain III. Shown are the results of RNase protection assays of human mRNA samples using probes specific for the IIIa (secreted form) sequence (A), the IIIb sequence (B), and the IIIc sequence (C). The assays were performed as described in Materials and Methods. The predicted sizes of the protected fragments were 410, 140, and 142 bp for the IIIa, IIIb, and IIIc probes, respectively. Arrows indicate the positions of the protected fragments. In each panel, lane 1 contains an aliquot of undigested probe (10^3 cpm) and lane 8 represents a tRNA negative control. RNA samples from the following human cell lines were analyzed: kidney endothelial cells (lane 2), foreskin fibroblasts (lane 3), Ewing's sarcoma (lane 4), astrocytoma (lane 5), neuroblastoma (lane 6), and adrenal cortex (lane 7). The sample represented in lane 9 of panel A was 20 μg of total RNA from rat L6 cells transfected with a cDNA expression construct containing the IIIa exon. The sample represented in lane 9 of panel C was 20 μg of total RNA from rat L6 cells transfected with a cDNA expression construct containing the IIIc exon.

amplified. In this genomic fragment, sequences encoding the unique region of the secreted form of receptor are continuous with sequences encoding the amino-terminal half of Ig domain III and are followed by a stop codon and a 3′ nontranslated sequence previously found in the cDNA sequence of the secreted receptor (Fig. 2). Although the published sequences of the secreted form cDNAs do not contain a polyadenylation signal (27), a consensus polyadenylation signal (AATAAA) is found in the genomic sequence only 16 nucleotides downstream from the 3′ end of the published cDNA sequence for the secreted form (Fig. 6). It is possible that the published cDNA sequence for the secreted receptor form is truncated at the 3′ terminus and that the consensus polyadenylation signal found in the genomic sequence may be present in the mRNA transcripts for this form of receptor.

Approximately 1.4 kb downstream from the above-mentioned consensus polyadenylation signal, the sequences encoding the third Ig domain of the membrane-spanning form of the receptor are found (Fig. 2). This type of genomic arrangement suggests that the decision to generate mRNA transcripts encoding membrane-spanning forms of the receptor is mediated by alternative splicing. In contrast, the decision to generate transcripts encoding secreted forms of the receptor may be mediated by selective use of polyadenylation signals.

Identification of an additional extracellular exon for the third Ig domain. The existence of two distinct exons coding for the second half of Ig domain III raised the question of whether additional exons encoding this region of the receptor molecule were present in the FGFR 1 gene. To address this question, we sequenced the entire genomic region located between sequences encoding the end of Ig domain II and sequences encoding the transmembrane domain (from the beginning of exon 5 to the end of exon 8; Fig. 2). Interestingly, an open reading frame encoding sequence

elements characteristic of Ig-like domains (38) was discovered between sequences encoding the secreted form Ig domain III (exon 5) and sequences encoding the Ig domain III associated with the known membrane-spanning receptor (exon 7). In particular, this open reading frame encoded a cysteine residue in the form GlyGluTyrValCys. This sequence of amino acids resembled the consensus sequence surrounding the second cysteine residue in typical Ig-like domains: GlyXTyr(Ser/Thr)Cys (38).

To determine whether the newly identified open reading frame contained an authentic exon sequence, we amplified cDNA derived from human umbilical vein endothelial cell mRNA by using primers derived from this open reading frame and primers derived from elsewhere in the FGFR 1 cDNA sequence. The primers derived from elsewhere in the cDNA sequence were chosen so that amplification of genomic DNA would yield fragments containing intronic sequences. Thus, only amplification of authentic cDNA species would yield fragments without introns. Using a primer (5′ primer) derived from Ig domain II sequence and a primer (3′ primer) derived from the putative new exon sequence, we successfully amplified a single DNA fragment that contained no intervening sequences (data not shown). This verified that the open reading frame that we had identified contained an authentic exon sequence. This exon encoded another possible carboxy-terminal half of Ig domain III. In the genomic sequence, this exon is numbered 6 (Fig. 2 and 3 show the exon boundaries).

We were unable to amplify any cDNA fragments when we used a primer (5′ primer) derived from the new exon and a primer (3′ primer) derived from the cytoplasmic coding region sequence. However, by analogy with the FGFR 2 gene, it is likely that this newly discovered exon is part of mRNA transcripts encoding a membrane-spanning form of the receptor (see Discussion). Thus, there are at least three alternative exons that can code for the carboxy-terminal half of Ig domain III (labeled IIIa, IIIb, and IIIc in Fig. 2). Expression of the IIIa exon gives rise to a secreted form of the receptor. In contrast, expression of the IIIc exon (previously published sequence), and potentially the IIIb exon (newly discovered exon), gives rise to a membrane-spanning form of the receptor.

Expression of FGFR 1 mRNA transcripts encoding different sequences for the second half of Ig domain III. The discovery of three alternative exons encoding the second half of Ig domain III led us to examine the patterns of expression of these exons in a variety of human cell lines. To accomplish this, we performed RNase protection assays using probes specific for each of the three exons (see Materials and Methods for a description of the probes). Some of the cell lines tested (astrocytoma, neuroblastoma, and adrenal cortex) simultaneously expressed transcripts for each of the three different receptor forms (Fig. 4). Also, one cell line, foreskin fibroblasts, expressed only the IIIc exon. In summary, with the exception of foreskin fibroblasts cells, the patterns of expression of the three alternative exons are highly similar. It should be noted, however, that the levels of the different mRNA transcripts probably differ widely. This conclusion is based on the facts that the protection assay shown in Fig. 4A (IIIa; secreted form-specific probe) was exposed to film for 10 days, the protection assay shown in Fig. 4B (IIIb probe) was exposed for 2 days, and the protection assay shown in Fig. 4C (IIIc probe) was exposed for only 5 h.

Comparison of the three exons encoding the second half of Ig domain III. Figure 5A shows an alignment of the IIIa, IIIb, and IIIc amino acid sequences of FGFR 1. Interestingly, the two exons that appear to be part of membrane-spanning forms of FGFR 1 (IIIb and IIIc) are identical in 22 of 49 amino acids (45%). In contrast, the amino acid sequence of IIIa (secreted form) shows no apparent homology to the sequence of either IIIb or IIIc.

Recently, cDNAs encoding transmembrane forms of FGFR 2 that differ in the second half of Ig domain III have been reported. Two of these cDNAs (K-sam [11] and the KGF receptor [22]) encode a sequence that is highly similar to the IIIb sequence of FGFR 1. This sequence is shown in Fig. 5B and is labeled FGFR 2 IIIb. Since the FGFR 2 IIIb sequence is part of a membrane-spanning receptor, it seems highly likely that the FGFR 1 IIIb exon that we have discovered also is part of a transmembrane receptor form. The other two known FGFR 2 cDNAs (bek [5] and TK14 [12]) encode a sequence highly similar to the IIIc sequence of FGFR 1. This sequence is also shown in Fig. 5B and is labeled FGFR 2 IIIc.

Comparison of the IIIb sequences of FGFR 1 and FGFR 2 reveals that they are identical in 38 of 49 amino acids (78%; Fig. 5C). Similarly, the IIIc sequences of FGFR 1 and FGFR 2 are identical in 40 of 48 amino acids (83%; Fig. 5D). Thus, the FGFR 1 IIIb sequence is more closely related to the IIIb sequence of FGFR 2 than to the IIIc sequence of FGFR 1. Likewise, the FGFR 1 IIIc sequence is more closely related to the IIIc sequence of FGFR 2 than to the IIIb sequence of FGFR 1.

Structural organization of the human FGFR 2 gene in the region containing alternative Ig domain III exons. To determine whether the structural organization of the human FGFR 2 gene was similar to that of the FGFR 1 gene, we amplified overlapping genomic fragments of this gene in the region coding for Ig domain III (see Materials and Methods). Figure 6 compares the genomic organization in this region of the two FGF receptor genes. As shown, the linear arrangements of the IIIb and IIIc exons in the FGFR 1 and FGFR 2 genes are identical. The positions (Fig. 7) and sizes of intronic sequences are also similar.

In contrast with the FGFR 1 gene, a stop codon is found in the FGFR 2 genomic sequence only 4 amino acids downstream from sequences encoding the amino-terminal half of Ig domain III (compared with 79 amino acids downstream for human FGFR 1). Thus, a putative FGFR 2 IIIa exon (secreted form) would encode only four unique amino acids. Experiments are currently in progress to determine whether this form of the FGFR 2 protein is expressed.

DISCUSSION

In this study, we have investigated the structural organization of the human FGFR 1 gene in an effort to determine the underlying mechanisms that are responsible for generating diverse forms of the FGFR 1 protein. Our findings suggest that the diversity seen in the third Ig domain of the receptor can be explained by alternative splicing of mRNA or selective use of alternate polyadenylation signals. Also, in addition to exons that encode the reported third Ig domain sequence for the transmembrane and secreted forms, we found an additional extracellular exon which represents a third possible exon coding for the second half of Ig domain III. Finally, we have determined the structural organization of the human FGFR 2 gene in the region coding for Ig domain III and show that this gene is organized in a fashion nearly identical to that of FGFR 1. These findings suggest

A

```
FGFR 1:

IIIa    VIMAPVFVGQSTGKETTVSGAQVPVGRLSCPRMGSFLTLQAHTLHLSRDLATSPRTSNRGHKVEVSWEQRAAGMGGAGL*

IIIb    -HSGINSSDAE--VLTLFNVTEAQSGEYVCKVSNYIGEANQSAWLTVTRP
          | |  | |  || | ||     ||| |   | ||    ||||||
IIIc    -TAGVNTTDKEMEVLHLRNVSFEDAGEYTCLAGNSIGLSHHSAWLTVLE
```

B

```
FGFR 2:

IIIa    VRTS*

IIIb    HSGINSSNAE--VLALFNVTEADAGEYICKVSNYIGQANQSAWLTVLPKQQ
          | |    |  ||   |||  ||||| |   | ||    ||||||||
IIIc    AAGVNTTDKEIEVLYIRNVTFEDAGEYTCLAGNSIGISFHSAWLTVLP
```

C

```
FGFR 1 IIIb  HSGINSSDAEVLTLFNVTEAQSGEYVCKVSNYIGEANQSAWLTVTRP
             ||||||| |||| |||||||  ||| ||||||||  ||||||||
FGFR 2 IIIb  HSGINSSNAEVLALFNVTEADAGEYICKVSNYIGQANQSAWLTVLPKQQ
```

D

```
FGFR 1 IIIc  TAGVNTTDKEMEVLHLRNVSFEDAGEYTCLAGNSIGLSHHSAWLTVLE
             ||||||||| |||  ||| ||||||||||||||||| | ||||||||
FGFR 2 IIIc  AAGVNTTDKEIEVLYIRNVTFEDAGEYTCLAGNSIGISFHSAWLTVLP
```

FIG. 5. Comparison of the three exons encoding the second half of Ig domain III. (A) Amino acid sequences encoded by the FGFR 1 IIIa, IIIb, and IIIc exons; (B) amino acid sequences encoded by the FGFR 2 IIIb and IIIc exons. The short amino acid sequence of the putative FGFR 2 IIIa exon is also shown. In panels A and B, the amino acid sequences of IIIb and IIIc have been aligned to show the degree of amino acid identities between these two distinct domains. (C) Alignment of the FGFR 1 IIIb sequence with the FGFR 2 IIIb sequence; (D) alignment of the FGFR 1 IIIc sequence with the FGFR 2 IIIc sequence. For the FGFR 1 IIIa and FGFR 2 IIIa sequences, asterisks indicate the locations of termination codons. The precise carboxy-terminal end of the FGFR 1 IIIb sequence is unknown. The sequence shown for FGFR 1 IIIb represents the 5′ end of the downstream primer that was used to amplify cDNA encoding the FGFR 1 IIIb exon. The FGFR 2 IIIb sequence (11, 22) and the FGFR 2 IIIc sequence (5) were derived from previously published sequences.

that the generation of diversity in the third Ig domains of FGF receptors is functionally important.

We have designated the three alternative exons encoding the second half of Ig domain III as IIIa, IIIb, and IIIc. Expression of the IIIa exon leads to a secreted form of FGFR 1, whereas expression of the IIIc exon, and likely the IIIb exon, leads to a membrane-spanning form of the receptor. The amino acid sequences of IIIb and IIIc are identical at 22 of 49 amino acids. In contrast, the primary amino acid sequence of IIIa is not homologous to that of either IIIb or IIIc. RNase protection assays demonstrate that a variety of human cell lines simultaneously express mRNA transcripts encoding either the IIIa, IIIb, or IIIc domain.

The reasons for expressing multiple forms of FGF receptor that differ in the third Ig domain are not fully understood. It is possible that each of the alternative forms of receptor has a specific role and pattern of expression during embryonic development. It is clear that the observed diversity in the third Ig domain provides a means for expressing both secreted and membrane-spanning forms of the receptor. The secreted protein encoded by the secreted-form cDNA binds both aFGF and bFGF (5a) and may act as an extracellular reservoir of FGF or may block the action of FGF by preventing interaction of the factor with a membrane-spanning FGF receptor.

It is possible that each of the alternative exons for the third Ig domain of the membrane-spanning form confers a characteristic pattern of ligand binding such that some of the FGF ligands have higher relative affinities than others. There

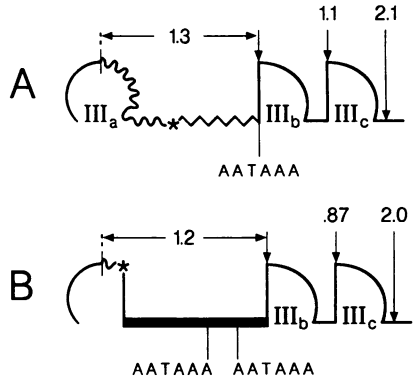

FIG. 6. Comparison of the structural organization of the human FGFR 1 (A) and FGFR 2 (B) genes in the regions encoding the third Ig domain of each receptor molecule. Arrows indicate the positions of intron sequences. Numbers above the arrows indicate the size of the intron in kilobases. Asterisks indicate the position of in-frame stop codons. In panel A, the wavy line represents sequences encoding the divergent amino acid sequence which is unique to the secreted receptor form; the zigzag line represents transcribed sequences which become the 3′ nontranslated sequence of secreted-form mRNA transcripts. In panel B, the thick black line represents genomic sequences found between the putative FGFR 2 IIIa coding region and sequences encoding the FGFR 2 IIIb exon. By analogy with the FGFR 1 gene, these sequences may represent 3′ nontranslated sequences that are part of secreted-form mRNA transcripts. The locations of consensus polyadenylation signals (AATAAA [27]) are shown.

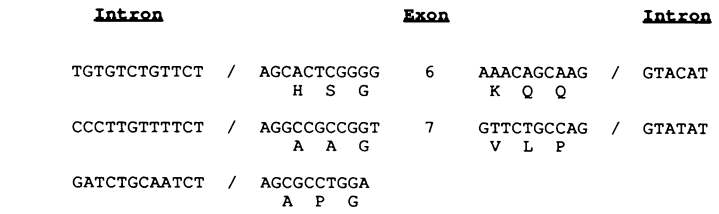

FIG. 7. Nucleotide junctions of introns and exons for the region of the FGFR 2 gene depicted in Fig. 6. The first three and last three amino acids of exons are translated to enable precise positioning relative to the published amino acid sequences of FGFR 2 (5, 11, 12, 22). We have numbered exons according to the corresponding exon numbers in the FGFR 1 gene (see Fig. 2). Since it is not clear where exon 5 ends, the junction sequence at the end of this exon is not shown.

may also be ligands that have not yet been discovered that bind to specific alternatively spliced forms of this receptor. In support of these ideas, the KGF receptor (22), an alternatively spliced variant of FGFR 2, exhibits ligand-binding specificities different from those of the bek form (5, 22) of FGFR 2. The KGF receptor lacks sequences encoding the highly acidic domain of the FGFR 2 extracellular region and contains sequences in the third Ig domain which correspond to the IIIb exon. The KGF receptor exhibits high-affinity binding for KGF and aFGF but only low-affinity binding for bFGF. The bek form of FGFR 2 contains sequences in the third Ig domain which correspond to the IIIc exon and also contains the highly acidic extracellular domain. In contrast to the KGF receptor, the bek protein exhibits high-affinity binding for both aFGF and bFGF.

Recently, several different membrane-spanning forms of the FGFR 2 protein have been reported. These proteins differ in sequence in the second half of the third Ig domain. By comparison with the exons determined for FGFR 1, it is clear that some of these proteins contain sequences which represent the IIIb form of FGFR 2 (K-sam [11] or KGF receptor [22]), while others contain sequences representing the IIIc form of FGFR 2 (bek [5] or TK14 [12]). On the basis of this information, we used PCR to determine the structure of the FGFR 2 gene in the region encoding Ig domain III. Our results show that the IIIb and IIIc exons of the FGFR 2 gene are arranged in the same linear order as they are in the FGFR 1 gene. Furthermore, the intronic sequences are located at nearly identical sites in the two genes. Recently, a third human FGF receptor gene has been identified (FGFR 3 [14]). The FGFR 3 protein, which may represent the human homolog of the chicken tyrosine kinase receptor Cek2 (25), is similar in structure and sequence to both FGFR 1 and FGFR 2. It will be interesting to determine whether the conserved structural arrangement of the FGFR 1 and FGFR 2 genes is also seen in the gene encoding FGFR 3.

The genomic arrangement of exons encoding the second half of the FGFR 1 Ig domain III provides clues regarding the possible mechanisms for generating either secreted or membrane-spanning forms of the receptor. Exons whose expression gives rise to membrane-spanning forms of the receptor (i.e., IIIb and IIIc) are separated from coding sequences of the receptor protein by intron sequences at both ends. Thus, expression of these exons, and hence membrane-spanning receptor forms, is probably mediated by alternative splicing. In contrast, genomic sequences encoding the secreted-form carboxyl terminus (IIIa) and the secreted-form 3′ nontranslated sequence are continuous in the genomic sequence with sequences encoding the amino-terminal half of Ig domain III. A consensus polyadenylation signal (AATAAA) is found in the genomic sequence only 16 nucleotides downstream from the end of the known secreted-form 3′ nontranslated sequence. It seems likely that this signal sequence may be present in full-length secreted-form mRNA transcripts. Thus, it is reasonable to propose that expression of the secreted FGFR 1 form is determined by selective use of polyadenylation signals. The conserved structural arrangement found in the FGFR 2 gene suggests that similar mechanisms may govern the expression of diverse forms of FGFR 2. Experiments are currently in progress to test these hypotheses.

ACKNOWLEDGMENTS

We are grateful to Marie Doherty for valuable assistance in the preparation of oligonucleotides. We thank J. Evan Sadler for providing a human umbilical vein endothelial cell cDNA library. We thank Kevin Peters and Mike Pazin for critical reading of the manuscript and insightful suggestions.

This work was supported by the National Institutes of Health Program of Excellence in Molecular Biology (HL-43821). D.E.J. was supported by American Heart Association fellowship 94-1219116. S.W. was supported by an Otto-Hahn fellowship of the Max-Planck Gesellschaft, Munich, Germany.

REFERENCES

1. **Anderson, K. J., D. Dam, S. Lee, and C. W. Cotman.** 1988. Basic fibroblast growth factor prevents death of lesioned cholinergic neurons in vivo. Nature (London) **332**:360–361.
2. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
3. **Burgess, W. H., and T. Maciag.** 1989. The heparin-binding (fibroblast) growth factor family of proteins. Annu. Rev. Biochem. **58**:575–606.
4. **Delli-Bovi, P., A. M. Curatola, F. G. Kern, A. Greco, M. Ittmann, and C. Basilico.** 1987. An oncogene isolated by transfection of Kaposi's sarcoma DNA encodes a growth factor that is a member of the FGF family. Cell **50**:729–737.
5. **Dionne, C. A., G. Crumley, F. Bellot, J. M. Kaplow, G. Searfoss, M. Ruta, W. H. Burgess, M. Jaye, and J. Schlessinger.** 1990. Cloning and expression of two distinct high-affinity receptors cross-reacting with acidic and basic fibroblast growth factors. EMBO J. **9**:2685–2692.

5a. **Duan, R., and L. T. Williams.** Unpublished data.

6. **Glisin, B., R. Crkvenjakov, and C. Byus.** 1974. Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry **13**: 2633–2637.
7. **Gospodarowicz, D.** 1974. Localisation of a fibroblast growth factor and its effect alone and with hydrocortisone on 3T3 cell growth. Nature (London) **249**:123–127.
8. **Gospodarowicz, D.** 1975. Purification of a fibroblast growth factor from bovine pituitary. J. Biol. Chem. **250**:2515–2520.
9. **Gospodarowicz, D., S. Massoglia, J. Cheng, G. M. Lui, and P. Bohlen.** 1985. Isolation of bovine pituitary fibroblast growth factor purified by fast protein liquid chromatography (FPLC). Partial chemical and biological characterization. J. Cell. Physiol. **122**:323–393.
10. **Gospodarowicz, D., G. Neufeld, and L. Schweigerer.** 1986. Molecular and biological characterization of fibroblast growth factor, an angiogenic factor which also controls the proliferation and differentiation of mesoderm and neuroectoderm derived cells. Cell Differ. **19**:1–17.
11. **Hattori, Y., H. Odagiri, H. Nakatani, K. Miyagawa, K. Naito, H. Sakamoto, O. Katoh, T. Yoshida, T. Sugimura, and M. Terada.** 1990. K-sam, an amplified gene in stomach cancer, is a member of the heparin-binding growth factor receptor genes. Proc. Natl. Acad. Sci. USA **87**:5983–5987.
12. **Houssaint, E., P. R. Blanquet, P. Chambon-Arnaud, M. C. Gesnel, A. Torriglia, Y. Courtois, and R. Breathnach.** 1990.

Related fibroblast growth factor receptor genes exist in the human genome. Proc. Natl. Acad. Sci. USA **87**:8180–8184.
13. **Johnson, D. E., P. L. Lee, J. Lu, and L. T. Williams.** 1990. Diverse forms of a receptor for acidic and basic fibroblast growth factors. Mol. Cell. Biol. **10**:4728–4736.
14. **Keegan, K., D. E. Johnson, L. T. Williams, and M. J. Hayman.** 1991. Isolation of an additional member of the fibroblast growth factor receptor family, FGFR-3. Proc. Natl. Acad. Sci. USA **88**:1095–1099.
15. **Kimelman, D., J. A. Abraham, T. Haaparanta, T. M. Palisi, and M. W. Kirschner.** 1988. The presence of fibroblast growth factor in the frog egg: its role as a natural mesoderm inducer. Science **242**:1053–1056.
16. **Kimelman, D., and M. Kirschner.** 1987. Synergistic induction of mesoderm by FGF and TGF-beta and the identification of an mRNA coding for FGF in the early Xenopus embryo. Cell **51**:869–877.
17. **Lee, P. L., D. E. Johnson, L. S. Cousens, V. A. Fried, and L. T. Williams.** 1989. Purification and complementary DNA cloning of a receptor for basic fibroblast growth factor. Science **245**:57–60.
18. **Maciag, T., T. Mehlman, R. Friesel, and A. B. Schreiber.** 1984. Heparin binds endothelial cell growth factor, the principal endothelial cell mitogen in bovine brain. Science **225**:932–934.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Marics, I., J. Adelaide, F. Raybaud, M. Mattei, F. Coulier, J. Planche, O. deLapeyriere, and D. Birnbaum.** 1989. Characterization of the HST-related FGF 6 gene, a new member of the fibroblast growth factor gene family. Oncogene **4**:335–340.
21. **Melton, D. A., P. A. Krieg, M. R. Regabliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
22. **Miki, T., T. P. Fleming, D. P. Bottaro, J. S. Rubin, D. Ron, and S. A. Aaronson.** 1991. Expression cDNA cloning of the KGF receptor by creation of a transforming autocrine loop. Science **251**:72–75.
23. **Moore, R., G. Casey, S. Brookes, M. Dixon, G. Peters, and C. Dickson.** 1986. Sequence, topography and protein coding potential of mouse int-2: a putative oncogene activated by mouse mammary tumour virus. EMBO J. **5**:919–924.
24. **Morrison, R. S., A. Sharma, J. de Vellis, and R. A. Bradshaw.** 1986. Basic fibroblast growth factor supports the survival of cerebral cortical neurons in primary culture. Proc. Natl. Acad. Sci. USA **83**:7537–7541.
25. **Pasquale, E. B.** 1990. A distinctive family of embryonic protein-tyrosine kinase receptors. Proc. Natl. Acad. Sci. USA **87**:5812–5816.
26. **Pasquale, E. B., and S. J. Singer.** 1989. Identification of a developmentally regulated protein-tyrosine kinase by using anti-phosphotyrosine antibodies to screen a cDNA expression library. Proc. Natl. Acad. Sci. USA **86**:5449–5453.
27. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
28. **Reid, H. H., A. F. Wilks, and O. Bernard.** 1990. Two forms of the basic fibroblast growth factor receptor-like mRNA are expressed in the developing mouse brain. Proc. Natl. Acad. Sci. USA **87**:1596–1600.
29. **Rubin, J. S., H. Osada, P. W. Finch, W. G. Taylor, S. Rudikoff, and S. A. Aaronson.** 1989. Purification and characterization of a newly identified growth factor specific for epithelial cells. Proc. Natl. Acad. Sci. USA **86**:802–806.
30. **Ruta, M., R. Howk, G. Ricca, W. Drohan, M. Zabelshansky, G. Laureys, D. E. Barton, U. Francke, J. Schlessinger, and D. Givol.** 1988. A novel protein tyrosine kinase gene whose expression is modulated during endothelial cell differentiation. Oncogene **3**:9–15.
31. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
33. **Slack, J. M. W., B. G. Darlington, J. K. Heath, and S. F. Godsave.** 1987. Mesoderm induction in early Xenopus embryos by heparin-binding growth factors. Nature (London) **326**:197–200.
34. **Taira, M., T. Yoshida, K. Miyagawa, H. Sakamoto, M. Terada, and T. Sugimura.** 1987. cDNA sequence of human transforming gene hst and identification of the coding sequence required for transforming activity. Proc. Natl. Acad. Sci. USA **84**:2980–2984.
35. **Thomas, K. A.** 1987. Fibroblast growth factors. FASEB J. **1**:434–440.
36. **Togari, A., G. Dickens, H. Kuzuya, and G. Guroff.** 1985. The effect of fibroblast growth factor on PC12 cells. J. Neurosci. **5**:307–316.
37. **Walicke, P., W. M. Cowan, N. Ueno, A. Baird, and R. Guillemin.** 1986. Fibroblast growth factor promotes survival of dissociated hippocampal neurons and enhances neurite extension. Proc. Natl. Acad. Sci. USA **83**:3012–3016.
38. **Williams, A. F., and N. A. Barclay.** 1988. The immunoglobulin superfamily-domains for cell surface recognition. Annu. Rev. Immunol. **6**:381–408.
39. **Zhan, X., B. Bates, X. Hu, and M. Goldfarb.** 1988. The human FGF-5 oncogene encodes a novel protein related to fibroblast growth factors. Mol. Cell. Biol. **8**:3487–3495.