THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 272, No. 40, Issue of October 3, pp. 25176–25183, 1997

Printed in U.S.A.

Genomic Organization of Human and Mouse Genes for Vascular Endothelial Growth Factor C*

(Received for publication, June 2, 1997, and in revised form, July 24, 1997)

Dmitri Chilov‡§, Eola Kukk‡¶, Suvi Taira, Michael Jeltsch, Jaakko Kaukonen, Aarno Palotie, Vladimir Joukov, and Kari Alitalo‖

From the Molecular/Cancer Biology Laboratory, Haartman Institute, PL21 (Haartmaninkatu 3), University of Helsinki, Helsinki 00014, Finland

We report here the cloning and characterization of human and mouse genes for vascular endothelial growth factor C (VEGF-C), a newly isolated member of the vascular endothelial growth factor/platelet-derived growth factor (VEGF/PDGF) family. Both VEGF-C genes comprise over 40 kilobase pairs of genomic DNA and consist of seven exons, all containing coding sequences. The VEGF homology domain of VEGF-C is encoded by exons 3 and 4. Exons 5 and 7 encode cysteine-rich motifs of the type $C_6C_{10}CRC$, and exon 6 encodes additional $C_{10}CXCXC$ motifs typical of a silk protein. A putative alternatively spliced rare RNA form lacking exon 4 was identified in human fibrosarcoma cells, and a major transcription start site was located in the human VEGF-C gene 523 base pairs upstream of the translation initiation codon. The upstream promoter sequences contain conserved putative binding sites for Sp-1, AP-2, and NF-κB transcription factors but no TATA box, and they show promoter activity when transfected into cells. The VEGF-C gene structure is thus assembled from exons encoding propeptides and distinct cysteine-rich domains in addition to the VEGF homology domain, and it shows both similarities and distinct differences in comparison with other members of the VEGF/PDGF gene family.

The process of growth and development of new blood vessels from preexisting ones is termed angiogenesis. Angiogenesis plays a critical role in providing growing tissues with oxygen and nutrients and also occurs in some pathological conditions, including tumor growth and metastasis, inflammatory lesions, wound healing, and endocrine diseases (1). Specific growth factors for the vascular endothelium, namely vascular endothelial growth factor (VEGF),[1] VEGF-B, VEGF-C, and placenta growth factor (PlGF) form a subfamily of the platelet-derived growth factor family (2–5). Certain other angiogenic growth factors may also act indirectly on endothelial cells via the induction of VEGF secretion in adjacent cells (4, 6). Besides angiogenesis, VEGF is implicated in vasculogenesis or formation of blood vessels *de novo* by *in situ* differentiation from mesodermal precursor cells (7). Both vasculogenesis and angiogenesis appear to be further regulated by inhibitory signals from cells and pericellular matrices (8, 9).

In situ hybridization studies have revealed that the three VEGF receptors VEGFR-1 (Flt-1), VEGFR-2 (Flk-1/KDR), and VEGFR-3 (Flt4) have specific, although partly overlapping, expression patterns in developing fetal tissues (10). VEGFR-3 expression has been localized in early embryonic blood vessels, especially in regions where the sprouting of developing lymphatic vessels takes place during the embryonic period (11). Its continued expression primarily in the lymphatic endothelium of adult tissues suggests that it functions as a regulator of these vessels. Notably, the expression of VEGF-C is often found in mesenchymal tissues surrounding the developing lymphatic vessels (12), and overexpression of VEGF-C induces lymphatic hyperplasia in transgenic mice (13).

Amino acid sequences of the PDGF/VEGF growth factor family are distinguished by eight conserved cysteine residues. In the case of PDGF, the cysteine residues have been shown to be essential for the correct folding and dimerization of the protein (14, 15), although activity was preserved when the interchain disulfide bonds were mutated (16). The VEGF-C dimers consist of polypeptides of 29- and 31-kDa, generated by proteolytic cleavage of the precursor between residues 227 and 228, and 21 kDa, produced via further N-terminal cleavages (17–19). In its C terminus, the VEGF-C precursor contains a unique domain of cysteine rich sequences resembling those of a silk protein. Other VEGF/PDGF family members typically have short C-terminal stretches of basic amino acid residues, which are encoded by alternatively spliced mRNAs and mediate VEGF binding to heparan sulfate proteoglycans (20).

The isolation and characterization of the VEGF gene promoter has been reported from human, rat, and mouse species (21–23). The human promoter was shown to contain multiple binding sites for Sp-1, AP-1, and AP-2 transcription factors and hypoxia-regulated elements (21–23). This promoter is TATA-less and responds to various effectors, including hypoxia, tumor necrosis factor-α, and basic fibroblast growth factor (24). A 1.2-kb 5′-flanking region of the mouse VEGF gene contains basal promoter activity (21, 23). However, although the promoter also has a hypoxia-regulated element, VEGF gene induction by hypoxia is also effected via mRNA stabilization (25).

We report here the cloning and characterization of the human and mouse VEGF-C genes. The genomic structure of the VEGF-C gene has been analyzed, and one alternatively spliced human mRNA isoform is described. We show that human genomic fragments containing the transcription initiation site are sufficient for basal promoter activity in transfected cells.

* This study was supported by the Finnish Cancer Organizations, the Finnish Academy, the Sigfrid Juselius Foundation, the Center for International Mobility (CIMO), the University of Helsinki, and the State Technology Development Center. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ The first two authors contributed equally.

§ Supported by a grant from Ida Montini Foundation.

¶ Supported by a grant from Maud Kuistila and Ida Montini Foundations.

‖ To whom correspondence should be addressed. Tel.: 358-9-1912-6434; Fax: 358-9-1912-6448.

[1] The abbreviations used are: VEGF, vascular endothelial growth factor; PlGF, placenta growth factor; PDGF, platelet-derived growth factor; VEGFR, vascular endothelial growth factor receptor; bp, base pair(s); kb, kilobase pair(s); PCR, polymerase chain reaction; RT-PCR, reverse transcriptase-polymerase chain reaction.

MATERIALS AND METHODS

Isolation of Human and Mouse Genomic Clones—A human genomic library in bacteriophage EMBL-3 λ (CLONTECH) was screened using a PCR fragment corresponding to nucleotides 632–746 of human VEGF-C cDNA (GenBank™ accession number X94216). One positive clone termed λ3 was obtained, and the insert was subcloned as a 14-kb *Xho*I fragment into the pBluescript II SK(±) vector (Stratagene). Restriction endonuclease digestion and hybridization analysis indicated that clone λ3 contains exons 2 and 3. The screening was continued using a fragment from the 5′-noncoding region of the cDNA (bp 1–95). Of two positive clones, clone λ5 contained exon 1 and the putative promoter region. To isolate additional clones encompassing further exons, purified DNA from a P1 plasmid clone 7760 (26) was used. *Eco*RI subclones of the P1 insert DNA containing human VEGF-C cDNA homologous sequences were isolated, and three different fragments were obtained, which contained the remaining exons.

About 10^6 plaques of bacteriophage λ FIX II library (Stratagene, La Jolla) containing 129SV mouse genomic DNA were screened with VEGF-C cDNA probes. Five cDNA fragments of human and mouse origin were used: 1) a PCR-amplified fragment corresponding to bp 495–1660 of human cDNA sequence (GenBank™ accession number X94216), 2) the full-length mouse cDNA (accession number U73620), 3) a fragment of a truncated mouse cDNA clone corresponding to bp 300–655, 4), a similar fragment corresponding to bp 498–655, and 5) a *Bst*XI-*Eco*RI fragment of mouse cDNA (bp 1370–1816). Lambda inserts varying from 10 to 15 kb in size were subcloned into pBluescript SK+ (Stratagene, La Jolla, CA). Computer analysis of the promoter region was performed with the MatInspector program (23).

Analysis of mRNA Expression—Polyadenylated RNA was isolated from NS-ø plasmacytoma, PYS-1 parietal yolk sac tumor (a kind gift from Dr. Eero Lehtonen), C_2C_{12} myoblast, WEHI-3 myelomonocyte, HeLa human cervical adenocarcinoma, and LLC Lewis lung carcinoma cell lines, and, as a positive control, from the human lung carcinoma HT-1080 cell line. The cells were starved overnight in media containing 0.2% fetal calf serum and then stimulated either with 10% of fetal calf serum or with 5 ng/ml 12-*O*-tetradecanoylphorbol-13-acetate for the indicated periods of time. Polyadenylated RNA was isolated, and 5 μg of each sample was electrophoresed in a 0.8% agarose gel containing formaldehyde and blotted onto Hybond-N (Amersham Corp.) filters. Blots were subsequently hybridized with radiolabeled probes specific for VEGF-C exons 4 (bp 713–836) or 2–4 (bp 335–836) or for the the full-length mouse cDNA and washed in stringent conditions. Mouse β-actin probe (CLONTECH) was used as a control.

RT-PCR Analysis of Splice Variants—A human HT-1080 fibrosarcoma cell cDNA library in λ gt11 vector (CLONTECH, HL1048b) was screened using the 153-bp human VEGF-C cDNA fragment described previously (18). Nine positive clones were picked and analyzed by PCR amplification using primers 1 (5′-CACGGTTATGCAAGCAAAG-3′) and 2 (5′-AACACAGTTTTCCATAATAG-3′), flanking bp 494–1661 (X94216). PCR was carried out at an annealing temperature of 55 °C for 25 cycles. Five out of nine clones were of the expected length of 1147 bp, and one was shorter. Both types of fragments were cloned into the pCR™II vector (Invitrogen) and analyzed by sequencing. The human cDNAs and cDNA libraries analyzed by PCR were made from placenta (Invitrogen A900–11), fetal lung (CLONTECH HL3022s), and heart (CLONTECH PT1156–2). The poly(A)$^+$ RNA was analyzed by RT-PCR using oligonucleotides 3 (5′-ATAGATGTGGGGAAGGAGTTT-3′) and 4 (5′-CATAAAATCTTCCTGAGCC-3′) (bp 765–1143) or 5 (5′-TCCTTC-CACCATGCACTTGC-3′) and 6 (5′-CAGAAAACCAGTCTT-3′) (bp 350–1806). Reverse transcription employed the avian myeloblastosis virus enzyme and oligo(dT) priming and was carried out at 42 °C for 1 h. PCR was carried out at an annealing temperature of 56 °C for 30 cycles.

To study potential splicing variants of mouse VEGF-C, the RT-PCR amplification employed 2 μg of the RNA, avian myeloblastosis virus polymerase, and oligo(dT) for priming. 2 μl of the cDNA mix was used for PCR amplification of the coding region of VEGF-C cDNA. 30 cycles of amplification were carried out using sense primer 7 (5′-CCCCA-GCCTGCGCCAGCCA-3′), antisense primer 8 (5′-GTGACTGACT-GAAAACTGGTATG-3′) and Dynazyme™ polymerase (Finnzymes) under the following conditions: 95 °C for 60 s, 62 °C for 60 s, and 72 °C for 75 s.

RNase Protection—RNA antisense probes of 122 and 256 bp were generated either from *Not*I-linearized plasmid containing a 90-bp *Sac*II fragment corresponding to nucleotides −16 to 74 or from a *Ban*I-linearized plasmid containing a 166-bp *Ban*I-*Rsr*II fragment corresponding to nucleotides 45–211 of the human VEGF-C promoter, using T3 polymerase and [γ-^{32}P]UTP. 3 μg of poly(A)$^+$ RNA was incubated with labeled probe at 55 °C overnight. Unhybridized RNA was digested with RNase A (10 units/ml) and T1 (1 mg/ml) at 30 °C for 1 h. The RNases were inactivated by proteinase K digestion at 37 °C for 15 min, and the samples were analyzed in 8% sequencing gels.

Analysis of Promoter Activity—Restriction fragments of genomic DNA containing 5′ portions of the first exon were cloned into the polylinker of the pGL3 reporter vector (Promega) and confirmed by sequencing. About 10 μg of the individual constructs in combination with 2 μg of pSV2-β-galactosidase control plasmid (ATCC) were transfected into HeLa or HT-1080 cells using the calcium phosphate-mediated transfection method. Two days after transfection, the cells were harvested and analyzed for luciferase activity. For serum stimulation experiments, the cells were transfected as described above, and 24 h after transfection they were changed into medium containing 0.5% bovine serum albumin. About 24 h later, the cells were stimulated for 4 h with 10% fetal calf serum and assayed for luciferase activity.

Fluorescence in Situ Hybridization—Mouse fetal fibroblast culture was established according to Ref. 29, and metaphase spreads were made using standard procedures. The cosmid probes specific for the mouse VEGF-C gene (exon 3) and Aga subclones A and B (28) were labeled with biotin-11-dUTP (Sigma) and digoxigenin-11-dUTP (Boehringer Mannheim) by nick translation. The fluorescence *in situ* hybridization was performed and analyzed as described previously (29, 30).

RESULTS

Preliminary Characterization of the Human and Mouse VEGF-C Genes—Two DNA clones covering exons 1, 2, and 3 of the human VEGF-C gene and several clones covering exons 1–7 of the mouse VEGF-C gene were isolated from human or mouse genomic λ DNA libraries using VEGF-C cDNA fragments as probes. Three additional genomic fragments containing exons 4–7 of the human gene were subcloned from a P1 plasmid clone (26). To determine the genomic organization, the clones were mapped using restriction endonuclease cleavage, Southern blotting, and hybridization analysis. The coding regions and exon-intron junctions were partially sequenced. The resulting exon-intron organization of the human and mouse VEGF-C genes are shown in Fig. 1*A*. As can be seen from the figure, both the human and mouse VEGF-C genes consist of seven exons, which all contain coding sequences. In Fig. 1*B*, the known structures of other genes of the VEGF family are shown for comparison. The sequences of all VEGF-C exon-intron boundaries conform to the consensus splicing signals (Fig. 2). The length of the intron between exons 5 and 6 was determined by nucleotide sequencing to be 301 bp in the human gene and 350 bp in the mouse gene. The length of the intron between exons 2 and 3 of the human gene was determined by restriction mapping and Southern hybridization to be approximately 1.6 kb. The length of other introns both in the human and mouse genes varied from 4.5 to over 10 kb (Fig. 2).

Analysis of the distribution of the various motifs encoded by both the human and mouse VEGF-C genes shows that the signal sequence and the first residues of the N-terminal propeptide are encoded by exon 1 (Fig. 1). The second exon encodes the carboxyl-terminal part of the N-terminal peptide and the amino terminus of the VEGF homology domain. The region encoding the 21-kDa form of VEGF-C (19) after proteolytic processing of both the N and C termini is marked *black* in Fig. 1*A*. The most conserved sequences of the VEGF homology domain are distributed in exons 3 (containing 6 conserved Cys residues) and 4 (containing 2 Cys residues). Exon 4 encodes the sequence for the major proteolytic cleavage site of the VEGF-C precursor between the VEGF homology domain and the C-terminal propeptide (19). Notably, in contrast with the other VEGF genes, exon 4 contains a 24-amino acid C-terminal extension after the last cysteine residues conserved in the VEGF homology domain.

Although there is no significant homology in the more C-terminal sequences, exons 5 and 7 each contain a C_6C_{10}CRC motif, resembling that found in exon 7 of VEGF, and exon 6

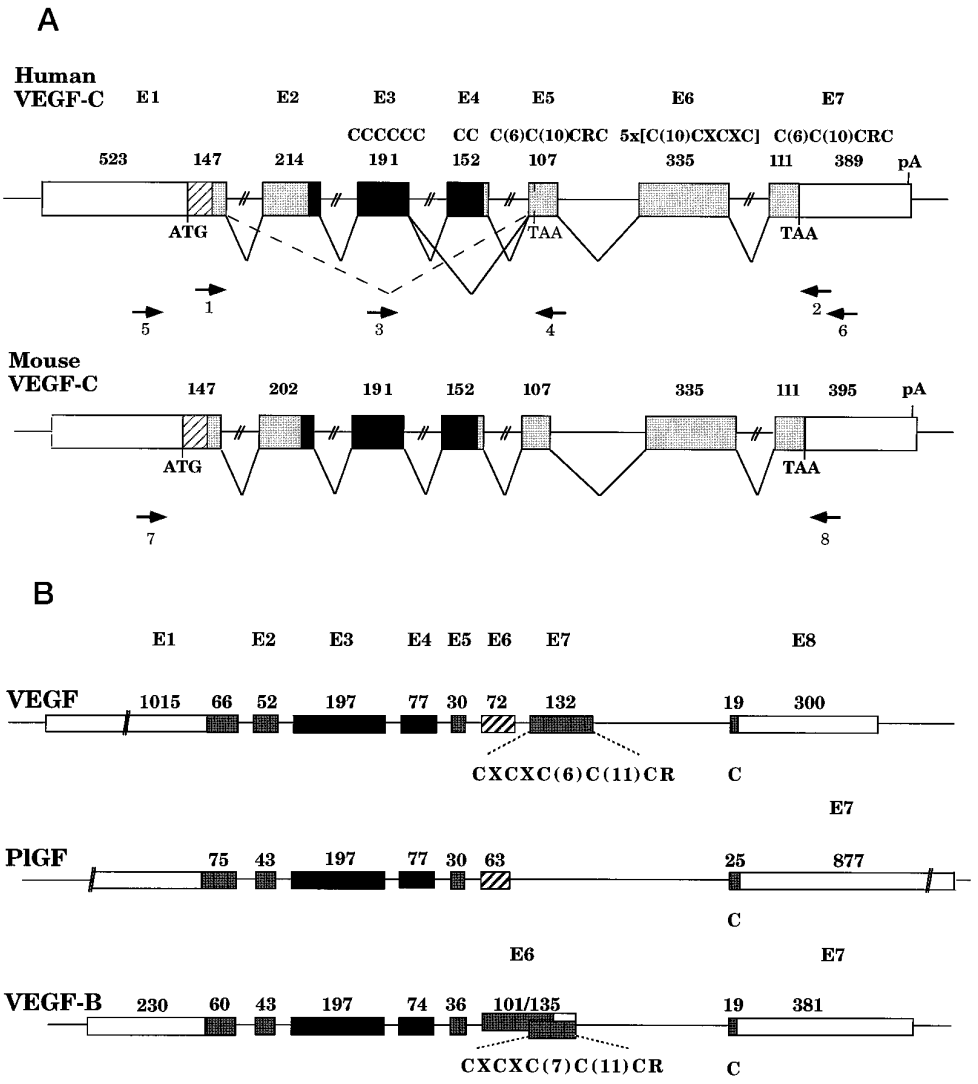

FIG. 1. **Exon-intron organization of human and mouse VEGF-C genes and other members of the VEGF family.** *A*, the exons are shown as *boxes* and their lengths in bp are indicated. *White boxes*, noncoding portions; *hatched boxes*, signal sequences; *gray boxes*: sequences encoding C- and N-terminal propeptides that are removed during successive steps of proteolytic processing; *black boxes*, sequences encoding the fully processed VEGF-C polypeptide. The cysteine motifs encoded by the different exons are indicated *above* the *boxes*. The translational start (ATG) and stop (TAA) codons are marked in *boldface type*, as is the polyadenylation signal (pA). The TAA stop codon in the 5th exon refers to the alternatively spliced mRNA form lacking exon 4, which is also marked. The putative mRNA form lacking exons 2–4 is indicated with the *broken line*. Primers used for the RT-PCR are marked with the *numbered arrows*; their sequences are detailed under "Materials and Methods." *B*, the structures of the VEGF, VEGF-B, and PlGF genes are shown for comparison. *Black boxes* indicate the cooling region, and the *striped box* indicates the exon encoding part of the heparin-binding region. The *bars* are drawn to scale. The structures of the VEGF, PlGF, and VEGF-B genes are from Refs. 22, 49, 50, and 52.

encodes an almost complete 5-fold repeated $C_{10}CXCXC$ motif, *i.e.* the Balbiani ring-3 protein-homologous domain (18). No sequences corresponding to exons 5 and 6 of VEGF and PlGF can be identified in VEGF-C. This is striking, because basic residues encoded by exons 6 of VEGF and PlGF contain parts of strong heparin-binding motifs. The length of 3′-noncoding region of the VEGF-C gene is 389 bp in the human and 395 bp in the mouse gene, followed by a polyadenylation signal. The 3′ region of mouse VEGF-C mRNA has three AUUUA motifs implicated in the rapid turnover of many cytokine mRNAs, including VEGF mRNA (31); however, only one of these sequences is conserved between the human and mouse mRNAs.

Chromosomal Localization of the Mouse VEGF-C Gene—The localization of the mouse VEGF-C gene was determined by fluorescent *in situ* hybridization. Initial analysis suggested localization to chromosome 8. As an internal control, we therefore performed cohybridization with a probe specific for the mouse aspartylglucosaminidase gene, previously mapped to chromosome 8 B3 (28). As analyzed from 22 metaphases, VEGF-C was localized telomeric of the aspartylglucosaminidase gene in chromosome 8, subband B3 (Fig. 3).

Identification of VEGF-C Splicing Variants—We reported earlier on the hybridization of the VEGF-C cDNA with a major 2.4-kb mRNA and with a 200–400-bp shorter minor mRNA species in human tissues (18). Because only one polyadenylation signal was found 367 bp downstream of the translational stop codon of the human VEGF-C gene, the length of the VEGF-C mRNA was 2267 bp plus the polyadenylation sequence, which is approximately the length of the 2.4 transcript. To clarify the origin of the smaller RNA, we PCR-amplified human cDNAs from placenta, fetal lung and heart, and the HT-1080 cell line, using oligonucleotides 1 and 2 indicated in Fig. 1*A*, and screened the PCR products in agarose gels for their size. PCR fragments resulting from amplification of five isolated VEGF-C cDNA clones from the HT-1080 fibrosarcoma cDNA library were of the expected length. The sixth clone

Exon length	Donor site	Intron length	Acceptor site

HUMAN

```
...........G...E...A...T(49).....................................A...Y...A...S.
E1.670.bp..GGC.GAG.GCC.ACGgtaggtctgcgt......>10.kb....tttctttgacagGCT.TAT.GCA.AGC
...........E...I...L...K(120)....................................S...I...D...N.
E2.214.bp..GAG.ATC.TTG.AAA.Agtaagtatggg.....1.6.kb.....atgacttgacagGT.ATT.GAT.AAT
...........L...S...K...T(184)....................................L...F...E...I.
E3.191.bp..CTC.AGC.AAG.ACGgtgggtattgt.......~10.kb...cccttctttgtagTTA.TTT.GAA.ATT
...........T...L...P...Q(235)....................................C...Q...A...A.
E4.152.bp..ACA.CTA.CCA.CAgtgagtatgaattaaa...~10.kb...ttcttccaaagG.TGT.CAG.GCA.GCG
...........A...G...D...(270).....................................D...S...T...D.
E5.107.bp..GCT.GGA.GAT.Ggtagcagaatg.........301.bp....ctatttgtctagAC.TCA.ACA.GAT
...........Q...T...C...S(382).....................................C...Y...R...R.
E6.334.bp..CAA.ACA.TGC.AGgtaagagatcc........~10.kb...tgttctcctagC.TGT.TAC.AGA.CGG
............Q...M...S(419)Stop..............polyA
E7.(501).bp..CAA.ATG.AGC.TAA.GATTGTACTGTT..TTAAAATAAAATGAAAATTGTATTAT
```

MOUSE

```
...........G...E...V...K(49).....................................A...F...E...G.
E1.........GGC.GAG.GTC.AAGgtaggtgcaagg....>10.kb...attgtctttgacag.GCT.TTT.TGA.AGG
...........E...I...L...K(116)....................................S...I...D...N.
E2.201.bp..GAG.ATC.CTG.AAA.Agtaagtag........4.5.kb....tgtgactcgacagGT.ATT.GAT.AAT
...........L...S...K...T(180)....................................L...F...E...I.
E3.191.bp..CTC.AGC.AAG.ACGgtaggtat..........9.kb.....ttgtcccttgtagTTG.TTT.GAA.ATT
...........T...L...P...Q(231)....................................C...Q...A...A.
E4.152.bp..ACA.TTA.CCA.CAgtgagtatg.........10.kb...gtctccccaaaagG.TGT.CAG.GCA.GCT
...........N...V...E...D(266).....................................D...S...T...N.
E5.107.bp..AAT.GTT.GAA.GAT.Ggtaagtaaaa.....350.bp.............tctagAC.TCA.ACC.AAT
...........Q...T...C...S(378).....................................C...Y...R...R.
E6.334.bp..CAA.ACA.TGC.AGgtaaggagtgt.........6.kb....ttttcccctagT.TGT.TAC.AGA.AGA
...........H...L...N(415)Stop.............polyA
E7.507.bp..CAT.CTG.AAC.TAA.GATCATACC...TTAAAATAAAATGAAAATTGTATTATAA
```

FIG. 2. **Exon-intron junctions of human and mouse VEGF-C genes.** Nucleotide sequences of the exon-intron borders are shown; nucleotides present in the cDNAs are indicated as *uppercase letters*. The corresponding codons and amino acid residues encoded at the junctions as well as exon and estimated intron lengths are indicated. Nucleotide numbering is according to the published human and mouse VEGF-C cDNA sequences (12, 18).

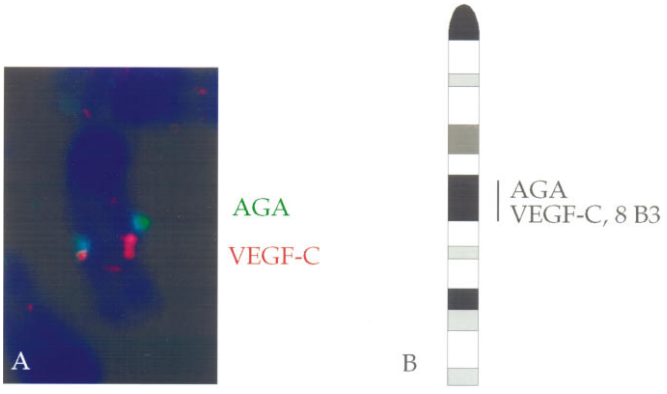

FIG. 3. **Chromosomal localization of the mouse VEGF-C gene.** *A*, a digital image of fluorescent *in situ* hybridization on chromosome 8 visualizing the specific location and order of the VEGF-C (*red*) and aspartylglucosaminidase (*green*) genes in subband B3. *B*, an idiogram of mouse chromosome 8 is also shown, illustrating the G-banding of the chromosome.

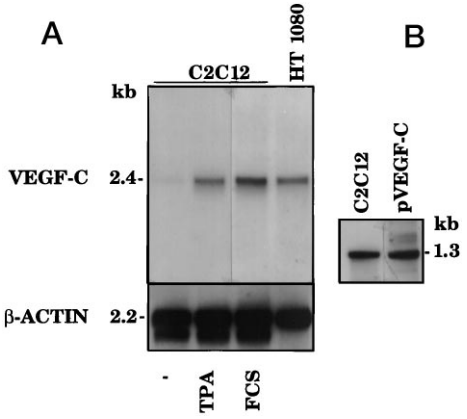

FIG. 4. **Northern blotting and RT-PCR analysis of mouse VEGF-C mRNA.** *A*, a Northern blot showing analysis of RNA from unstimulated, fetal calf serum and tetradecanoyl-13-phorbol acetate-stimulated C_2C_{12} cells. For comparison, RNA from the human HT-1080 cells was used (see Ref. 18). Mouse β-actin probe was used as a loading control. *B*, analysis of the VEGF-C open reading frame by RT-PCR. The indicated RNA samples were processed as described under "Materials and Methods," electrophoresed in 1% agarose, blotted, and hybridized to a cDNA probe covering the whole open reading frame.

yielded a somewhat shorter product, with a deletion of 152 bp, consisting of nucleotides 906–1058 of the published cDNA sequence. The missing nucleotides corresponded to exon 4 of the human gene (Figs. 1*A* and 2). To determine whether the shorter RNA represented an alternatively spliced version, we analyzed RNA from PC-3 and HT-1080 cells by RT-PCR using primers 3 and 4 and primers 5 and 6 indicated in Fig. 1*A*. The amplified region contained either exons 3–5 or the whole coding region of the human VEGF-C cDNA. The PCR products were analyzed by Southern blotting and hybridization with the full-length VEGF-C cDNA. However, only one major full-length form of VEGF-C could be amplified (data not shown).

To determine whether alternative splicing occurs in the mouse VEGF-C gene, mouse cell lines positive for VEGF-C were identified by Northern blotting and hybridization and analyzed further by RT-PCR. From the tested cell lines, only the myoblast cell line C_2C_{12} and the Lewis lung carcinoma cell line LLC appeared to contain a single 2.4-kb VEGF-C mRNA in Northern blotting. As in several human cells (32), mouse VEGF-C mRNA was increased 5-fold by fetal calf serum and 12-*O*-tetradecanoylphorbol-13-acetate treatment of the expressing cells, but the mobility of the mRNA was not changed (Fig. 4). The RT-PCR analysis of the coding region of the previously published mouse cDNA (12) revealed only one amplification product, suggesting that most of the mouse VEGF-C mRNA is composed of only one isoform.

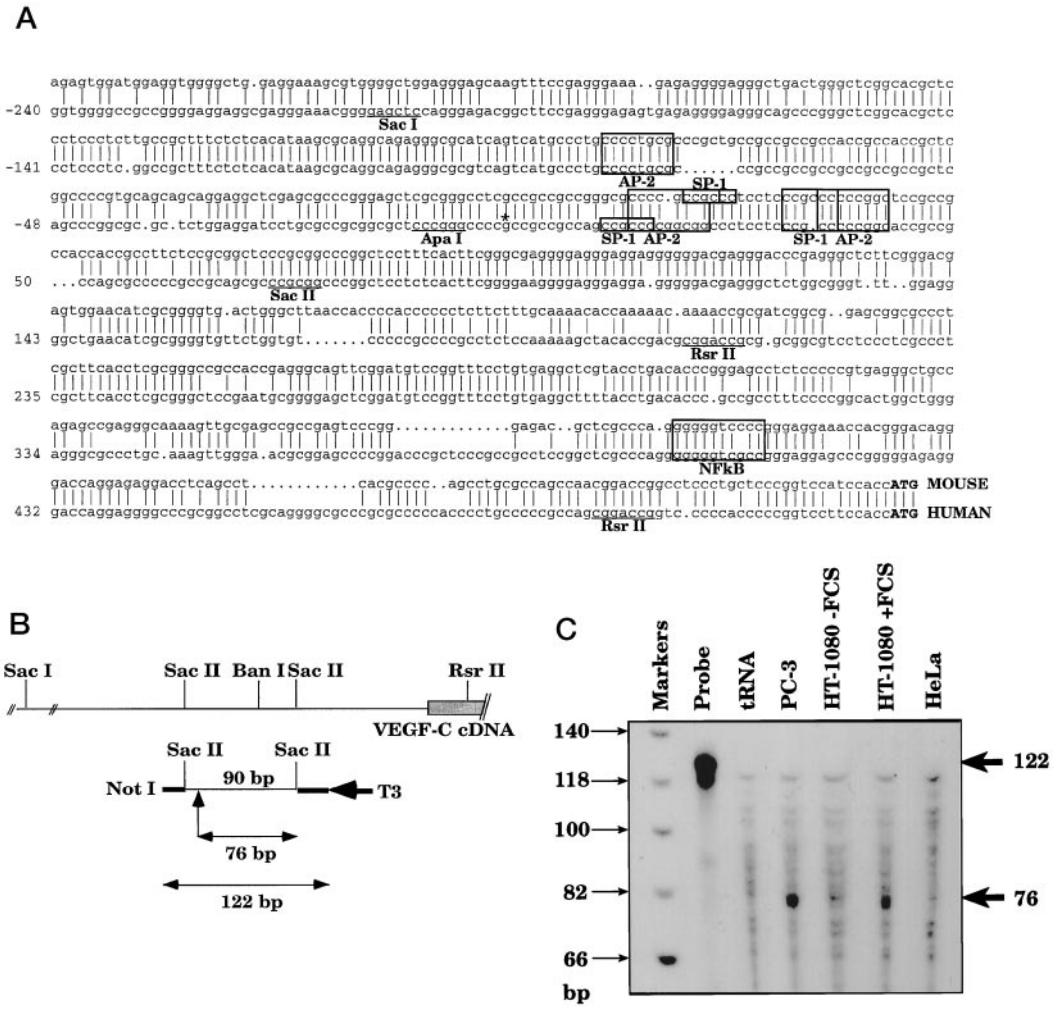

FIG. 5. **Comparison of the upstream regions of the human and mouse VEGF-C genes and mapping of the human VEGF-C mRNA start site.** *A*, alignment of human and mouse VEGF-C gene promoter regions. Putative transcription factor binding sites conserved in both species are indicated. The promoter region was analyzed using the MatInspector program (23). The transcription start site is indicated with an *asterisk*. Consensus binding sites for the relevant transcription factors are *boxed*: Sp-1, GGGCGG (51); AP-2, 5′-GSSWGSCC-3′ (43); NF-κB, 5′-GGGRN-NYYCC-3′ (27). Relevant restriction endonuclease cleavage sites are also shown. The translation initiation codon is marked in *boldface type*. The GenBank™ accessions numbers are AF020393 (human) and AF020392 (mouse). *B*, schematic structure of the promoter region and the cRNA probe generated from the *Sac*II fragment using the T3 RNA polymerase. The 5′-end of the VEGF-C cDNA (X94216) is shown as the *shaded box* on the *right*. The *vertical arrow* shows the mRNA start site. *C*, poly(A)$^+$ RNA from PC-3 cells, starved or serum-stimulated HT-1080 cells, and HeLa cells was isolated and hybridized with the 122-bp antisense riboprobe shown in *A*. Bands corresponding to the probe and the protected fragment of 76 bp (see *A*) are indicated with *arrows* on the *right*.

Identification of the VEGF-C Promoter Region and the mRNA Start Site—Restriction mapping of the human VEGF-C clone λ5 using a combination of single and double digestions and Southern hybridization showed that it covers approximately a 6-kb region upstream of ATG, exon 1 and at least 5 kb of intron 1. An 8-kb *Xho*I-*Eco*RI fragment containing exon 1 and 5.4- and 2-kb stretches of 5′- and 3′-flanking sequences, respectively, was subcloned and further analyzed. The nucleotide sequences of approximately 0.7 kb of the human and mouse VEGF-C promoters and 5′-untranslated regions are shown in Fig. 5*A*. Computer analysis was performed to find potential binding sites for transcription factors. The human and mouse promoter sequences were compared, and conserved sites observed in both species are marked in Fig. 5*A*. They include putative binding sites for Sp-1, AP-2, and NF-κB transcription factors.

The location of the human VEGF-C mRNA start site was established by RNase protection assay. Hybridization of an antisense riboprobe generated from a 166-bp *Ban*I-*Rsr*II genomic fragment with poly(A)$^+$ RNA from PC-3 cells expressing high levels of VEGF-C resulted in the protection of the full-length probe (Fig. 5*B* and data not shown). The use of a further upstream 90-bp *Sac*II fragment resulted in a protected band of 76 bp (Fig. 5*C*). This indicates that a major VEGF-C mRNA start site (marked *+1* in the figure) is located 523 bp upstream of the ATG translation initiation codon. A similar result was obtained when using poly(A)$^+$ RNA from serum stimulated HT-1080 cells, whereas an approximately 3-fold weaker RNA signal was obtained from serum-starved HT-1080 cells or HeLa cells.

Analysis of VEGF-C Promoter Activity—To identify DNA elements important for the basal expression of VEGF-C in transfected cells, a set of luciferase reporter plasmids containing serial 5′-deletions of the human VEGF-C promoter region was constructed. The promoter-reporter constructs were co-transfected into HeLa cells, which gave a high frequency of positive cells in control transfections with the pSV2-β-galactosidase control plasmid. The luciferase activity was normalized to that of the pGL3 luciferase control vector driven by the SV40 promoter/enhancer. As can be seen from Fig. 6, the 5.4-kb *Xho*I-*Rsr*II fragment gave 9-fold higher activity in comparison with a promoterless vector. Deletion of a 5′ *Xho*I-*Hin*dIII frag-

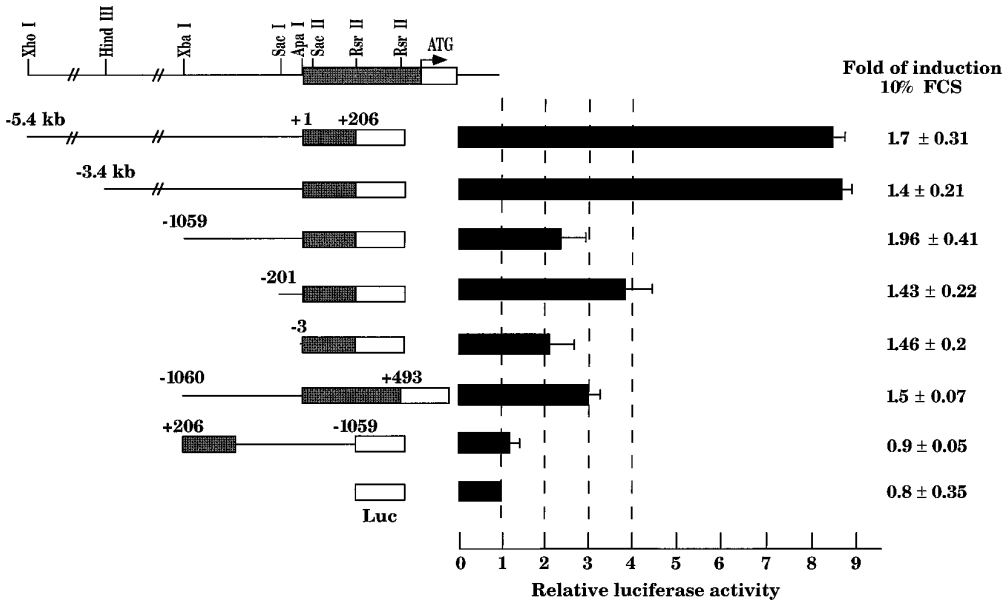

FIG. 6. **Schematic structure of the VEGF-C promoter-reporter constructs and their activities in transfected HeLa cells.** Constructs were generated by linking putative VEGF-C promoter fragments to the luciferase reporter gene in the pGL3 vector (Promega) and introduced into HeLa cells by the calcium phosphate-mediated transfection method. The luciferase activities obtained were compared with the activity of a promoterless pGL3basic construct to obtain a measure of relative promoter activity. For analysis of the stimulation of promoter activity by serum, transfected HeLa cells were starved for 24 h followed by serum stimulation for 4 h. The ratio of serum-stimulated/starved luciferase activities is given. All luciferase values are mean values from at least three experiments, normalized for transfection efficiency using the pSV2-β-galactosidase plasmid.

ment of 2 kb did not change the activity. The activity of the 1.1-kb *Xba*I-Rsr II fragment was approximately twice that of the pGL3 basic vector, while the activity of the same fragment in the reverse orientation was at background level. Further deletion of the 5′ *Xba*I-*Sac*I fragment caused an increase of the promoter activity, suggesting the presence of silencer elements in the region from −1059 to −201. The *Sac*II-*Rsr*II (+74 to +206) fragment yielded only background activity, which was consistent with the fact that the mRNA initiation site was not present in this construct (data not shown).

To determine whether further sequences in the first exon of the human VEGF-C gene are important for basal expression, an Rsr II fragment spanning nucleotides +206 to +493 was subcloned between the *Xba*I-*Rsr*II fragment and the luciferase reporter gene. Indeed, the construct obtained showed a 50% increase in activity in comparison with the *Xba*I-*Rsr*II fragment.

Effect of Serum Stimulation on VEGF-C mRNA Expression—Because the increase of VEGF-C mRNA obtained after serum stimulation may have been due to mRNA stabilization or an increased transcription rate, we sought to elucidate whether serum stimulates VEGF-C promoter activity. VEGF-C promoter constructs were transfected into HeLa cells, and the transfected cells were subsequently stimulated with serum as described under "Materials and Methods." As can be seen from the data listed in the right-hand part of Fig. 6, the 1.1-kb *Xba*I-*Rsr*II promoter construct yielded a 2-fold increase of activity upon serum stimulation, whereas the same fragment in the reverse orientation showed no response. The other VEGF-C promoter constructs also showed up-regulation, ranging from 1.4- to 1.6-fold. However, no significant increase was seen in the activity of the control vector. The *Xho*I-*Rsr*II promoter construct gave only a 1.7-fold induction upon serum stimulation upon transfection into HT1080 cells (data not shown). For comparison, the increase of the VEGF-C steady-state mRNA after 4 h of serum stimulation was about 1.7-fold in HeLa cells and about 3-fold in HT-1080 cells.

DISCUSSION

VEGF-C is the fourth member of the VEGF subfamily of PDGF-related growth factors and the largest VEGF gene described so far. The genomic structures of the previously described members, VEGF, VEGF-B, and PlGF show a high degree of similarity between each other in exon-intron junctions and some similarities of alternative splicing. In the case of VEGF-C, the exon structure resembles that of the other VEGF family members, particularly in exons 3 and 4 encoding the core growth factor domain, which contains the eight cysteine residues and the signature sequence P*X*CV*XXX*RC*X*GCC conserved in all PDGF/VEGF family members. The most variable parts of these genes are the regions encoded by the sixth and seventh exons, which are alternatively spliced in all other members except VEGF-C. A short fifth exon is present in all splicing variants of the other genes, but VEGF-C appears to lack such an exon. Instead, there is a C-terminal extension in the sequences encoded by exon 4. Also, sequence corresponding to exon 6, which encodes part of a heparin binding domain present in $VEGF_{145}$, $VEGF_{165}$, $VEGF_{189}$, and $PlGF_{152}$ is missing in VEGF-C (22, 33).

All VEGFs have a cysteine-rich C terminus. In the case of VEGF-C, the cysteine-rich domain is extended in a repeated motif and divided into three exons. Exons 5–7 show no significant homology with VEGF, although the C_6C_{10}CRC cysteine motifs, which are encoded by exons 5 and 7 of VEGF-C may be related to the C*X*C*X*C_6C_{11}CRC pattern distributed in exons 7 and 8 of VEGF (22) and 6A and 7 of VEGF-B (34, 35). The distribution of functional domains in the VEGF-C gene resembles to some extent that of the PDGF-B gene, where the N- and C-terminal propeptides are also cleaved off after polypeptide biosynthesis (36).

The most striking differences between VEGF-C and the other VEGFs concern the potential alternative splicing variants. We have previously shown hybridization of mRNA species of 2.4 and 2.0–2.2 kb with the VEGF-C probe. Lee *et al.* (37) reported two VEGF-C cDNA clones containing deletions of 152

or 557 nucleotides corresponding to exon 4 or exons 2–4, respectively. These variant cDNAs were considered to be the result of alternative splicing. We report here a similar, alternatively spliced version of VEGF-C lacking exon 4. Deletion of exon 4 would result in a premature stop codon caused by a frameshift after the alternative splice acceptor of exon 5. This protein would not contain the C-terminal cleavage site (amino acids 227 and 228) of the precursor and would lack 2 of the 8 conserved Cys residues of the VEGF homology domain (18). Indeed, a recombinant protein of the expected size was produced from a baculovirus vector containing the variant cDNA, but in preliminary experiments, we have been unable to show that this protein is capable of stimulating the autophosphorylation of VEGFR-3.[2] Studies of VEGF crystal structure have shown that the corresponding region contains the receptor binding sites and essential cysteine residues involved in inter- and intrachain disulfide bonding (38). All isoforms of the VEGFs described earlier contain this part of the protein. Furthermore, RT-PCR analysis suggested that the PC-3 and HT-1080 cells as well as several tissues express only one major form of VEGF-C. Therefore, the alternatively spliced mRNA detected in the HT-1080 cell library may represent a rare transcript that is unlikely to have significant biological relevance in the cells and tissues analyzed. A similar conclusion was made from RT-PCR analysis of the mouse VEGF-C coding region.

The transcription of the human VEGF-C gene starts 523 bp before the first coding base; thus, the mRNA has a long untranslated 5′-leader. As for the VEGF gene (22), the VEGF-C gene promoter lacks consensus TATA and CCAAT sequences. Instead, it has putative binding sites for Sp-1, a ubiquitous nuclear protein that can initiate transcription from TATA-less genes (39) and a long GC-rich 5′-untranslated region, which is typical for several growth factor genes and is postulated to have a role in translational regulation (22, 23, 28, 40). In addition, sequences upstream of the VEGF-C translation start site were found to contain consensus binding sites for the AP-2 transcription factor. This suggests that the cAMP-dependent protein kinase and TGF-α as activators of the AP-2 transcription factor can regulate VEGF-C transcription (41–43). Besides, we have shown here that the VEGF-C promoter activity is approximately 2-fold induced by serum stimulation. The corresponding mRNA increase was 2–5-fold in different cells, although some of this increase could be attributed to mRNA stabilization (32). Thus, important serum-responsive elements may be located outside of the promoter fragments used in the present work, for example in the first intron of this large gene.

Potential binding site for the NF-κB transcription factor important for the lymphoid and immune systems was also found. NF-κB is associated with rapid response activation mechanisms and is known to be induced by a variety of inflammatory stimuli, being a mediator of tissue-specific gene expression (44, 45). The presence of NF-κB sites suggests that NF-κB may be implicated in the induction of the VEGF-C mRNA by interleukin-1 and tumor necrosis factor-α.[3]

Unlike the VEGF gene, the VEGF-C gene promoter does not contain putative binding sites for hypoxia-regulated factors. Numerous studies have shown that another major control point for the hypoxic induction of the VEGF gene is the regulation of the steady-state level of mRNA (21, 46, 47). The VEGF mRNA stability is considered to be determined by the presence of protein-binding sequence motifs in its 3′-untranslated region (5, 25). The 3′-untranslated region of the VEGF-C gene also contains a motif of this type, but at least our preliminary analyses indicate that the mRNA steady-state levels are not increased by hypoxia. Thus, besides the significant differences in its exon structure and encoded protein domains, the VEGF-C gene also has interesting differences in its regulation and function in comparison with the VEGF prototype of this gene family (13, 32). Interestingly, if gaps introduced into the amino acid sequences for the purposes of alignment are ignored for calculations, the newly cloned FIGF/VEGF-D is about 48% identical with human VEGF-C and about 31% identical with human VEGF (48, 52). On the basis of their multiple amino acid sequence alignments, VEGF-C and VEGF-D are likely to have similar structural and functional properties and gene structures. Perhaps the PDGF-A/PDGF-B and VEGF-C/VEGF-D genes have diverged from VEGF/VEGF-B and from PlGF earlier than from each other during evolution.

Acknowledgments—We thank Mari Helanterä and Tapio Tainola for excellent technical assistance.

REFERENCES

1. Folkman, J. (1995) *Nature Medicine* **1,** 27–31
2. Ferrara, N., and Davis-Smyth, T. (1997) *Endocr. Rev.* **18,** 4–25
3. Neufeld, G., Tessler, S., Gitay-Goren, H., Cohen, T., and Levi, B.-Z. (1994) *Prog. Growth Factor Res.* **5,** 89–97
4. Dvorak, H. F., Brown, L. F., Detmar, M., and Dvorak, A. M. (1995) *Am. J. Pathol.* **146,** 1029–1039
5. Klagsburn, M., and D'Amore, P. A. (1996) *Cytokine Growth Factor Rev.* **7,** 259–270
6. Pertovaara, L., Kaipainen, A., Mustonen, T., Orpana, A., Ferrara, N., Saksela, O., and Alitalo, K. (1994) *J. Biol. Chem.* **269,** 6271–6274
7. Risau, W., Sariola, H., Zerwes, H.-G., Sasse, J., Ekblom, P., Kemler, R., and Doetschman, T. (1988) *Development* **102,** 471–478
8. Folkman, J., and Shing, Y. (1992) *J. Biol. Chem.* **267,** 10931–10934
9. Hanahan, D., and Folkman, J. (1996) *Cell* **86,** 353–364
10. Kaipainen, A., Korhonen, J., Pajusola, K., Aprelikova, O., Persico, M. G., Terman, B. I., and Alitalo, K. (1993) *J. Exp. Med.* **178,** 2077–2088
11. Kaipainen, A., Korhonen, J., Mustonen, T., van Hinsbergh, V. W. M., Fang, G.-H., Dumont, D., Breitman, M., and Alitalo, K. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 3566–3570
12. Kukk, E., Lymboussaki, A., Taira, S., Kaipainen, A., Jeltsch, M., Joukov, V., and Alitalo, K. (1996) *Development* **122,** 3829–3837
13. Jeltsch, M., Kaipainen, A., Joukov, V., Meng, X., Lakso, M., Rauvala, H., Swartz, M. D. F., Jain, R., and Alitalo, K. (1997) *Science* **276,** 1423–1425
14. Andersson, M., Östman, A., Bäckström, G., Hellman, U., George-Nascimento, C., Westermark, B., and Heldin, C.-H. (1992) *J. Biol. Chem.* **267,** 11260–11266
15. Oefner, C., DArcy, A., Winkler, F. K., Eggimann, B., and Hosang, M. (1992) *EMBO J.* **11,** 3921–3926
16. Kenney, W. C., Haniu, M., Herman, A. C., Arakawa, T., Costigan, V. J., Lary, J., Yphantis, D. A., and Thomason, A. R. (1994) *J. Biol. Chem.* **269,** 12351–12359
17. Hu, J.-S., Hastings, G., Cherry, S., Gentz, R., Ruben, S., and Coleman, T. (1997) *FASEB J.* **11,** 498–504
18. Joukov, V., Pajusola, K., Kaipainen, A., Chilov, D., Lahtien, I., Kukk, E., Saksela, O., Kalkkinen, N., and Alitalo, K. (1996) *EMBO J.* **15,** 290–298
19. Joukov, V., Sorsa, T., Kumar, V., Jeltsch, M., Claesson-Welsh, L., Cao, Y., Saksela, O., Kalkkinen, N., and Alitalo, K. (1997) *EMBO J.* **16,** 3898–3911
20. Soker, S., Fidder, H., Neufeld, G., and Klagsbrun, M. (1996) *J. Biol. Chem.* **271,** 5761–5767
21. Levy, A. P., Levy, N. S., Wegner, S., and Goldberg, M. A. (1995) *J. Biol. Chem.* **270,** 13333–13340
22. Tischer, E., Mitchell, R., Hartman, T., Silva, M., Gospodarowicz, D., Fiddes, J. C., and Abraham, J. A. (1991) *J. Biol. Chem.* **266,** 11947–11954
23. Shima, D. T., Kuroki, M., Deutsch, U., Ng, Y.-S., Adamis, A. P., and D'Amore, P. A. (1996) *J. Biol. Chem.* **271,** 3877–3883
24. Ryuto, M., Ono, M., Izumi, H., Yoshida, S., Weich, H. A., Kohno, K., and Kuwano, M. (1996) *J. Biol. Chem.* **271,** 28220–28228
25. Levy, A. P., Levy, N. S., and Goldberg, M. A. (1996) *J. Biol. Chem.* **271,** 25492–25497
26. Paavonen, K., Horelli-Kuitunen, N., Chilov, D., Kukk, E., Pennanen, S., Kallioniemi, O., Pajusola, K., Olofsson, B., Eriksson, U., Joukov, V., Palotie, A., and Alitalo, K. (1996) *Circulation* **93,** 1079–1082
27. Sha, W., Liou, H.-C., Tuomanen, E., and Baltimore, D. (1995) *Cell* **80,** 321–330
28. Bonthron, D., and Orkin, S. H. (1988) *Eur J. Biochem* **171,** 51–57
29. Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W., and Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 2156–2160
30. Pinkel, D., Straume, T., and Gray, J. W. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 2934–2938
31. Chen, C.-Y., and Shyu, A.-B. (1994) *Mol. Cell. Biol.* **14,** 8471–8482
32. Enholm, B., Paavonen, K., Ristimaki, A., Kumar, V., Gunji, Y., Klefstrom, J., Kivinen, L., Laiho, M., Olofsson, B., Joukov, V., Eriksson, U., and Alitalo, K. (1997) *Oncogene* **14,** 2475–2483
33. Poltorak, Z., Cohen, T., Sivan, R., Kandelis, Y., Spira, G., Vlodavsky, I.,

[2] D. Chilov, E. Kukk, S. Taira, M. Jeltsch, J. Kaukonen, A. Palotie, V. Joukov, and Kari Alitalo, unpublished data.

[3] A. Ristimaki, D. Chilov, E. Kukk, S. Taira, M. Jeltsch, J. Kaukonen, A. Palotie, V. Joukov, and Kari Alitalo, unpublished data.

Keshet, E., and Neufeld, G. (1997) *J. Biol. Chem.* **272,** 7151–7158
34. Lagercrantz, J., Larsson, C., Grimmond, S., Fredriksson, M., Weber, G., and Piehl, F. (1996) *Biochem. Biophys. Res. Commun.* **220,** 147–152
35. Olofsson, B., Pajusola, K., Kaipainen, A., Euler, G., Joukov, V., Saksela, O., Orpana, A., Pettersson, R., Alitalo, K., and Eriksson, U. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 2576–2581
36. Heldin, C.-H., and Westermark, B. (1989) *Trends Genet.* **5,** 108–111
37. Lee, J., Gray, A., Yuan, J., Luoh, S.-M., Avraham, H., and Wood, W. I. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 1988–1992
38. Muller, Y., Li, B., Christinger, H., Wels, J., Cunningham, B., and Devos, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 7192–7197
39. Pugh, B. F., and Tjian, R. (1990) *Cell* **61,** 1187–1197
40. Rao, C. D., Igarashi, H., Chiu, I.-M., Robbins, K., and Aaronson, S. A. (1986) *Proc. Natl. Acad. Sci.* **83,** 2392–2396
41. Imagawa, M., Chiu, R., and Karin, M. (1987) *Cell* **51,** 251–260
42. Gille, J., Swerlick, R., and Caughman, S. (1997) *EMBO J.* **16,** 750–759
43. Mitchell, P. J., Wang, C., and Tjian, R. (1987) *Cell* **50,** 847–861
44. Baeuerle, P. A., and Baltimore, D. (1996) *Cell* **87,** 13–20
45. Lenardo, M. J., and Baltimore, D. (1989) *Cell* **58,** 227–229
46. Shweiki, D., Itin, A., Soffer, D., and Keshet, E. (1992) *Nature* **359,** 843–845
47. Levy, A. P., Levy, N. S., and Goldberg, M. A. (1996) *J. Biol. Chem.* **271,** 2746–2753
48. Orlandini, M., Marconcini, L., Ferruzzi, R., and Oliviero, S. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 11675–11680
49. Olofsson, B., Pajusola, K., G., von Euler, G., Chilov, D., Alitalo, K., and Eriksson, U. (1996) *J. Biol. Chem.* **271,** 19310–19317
50. Hauser, S., and Weich, H. A. (1993) *Growth Factors* **9,** 259–68
51. Faisst, S., and Meyer, S. (1992) *Nucleic Acids Res.* **20,** 3–26
52. Maglione, D., Guerriore, V., Viglietto, G., Ferraro, M. G., Aprelikova, O., Alitalo, K., Vecchio, S. D., Lie, K.-J., Chou, J. Y., and Persico, G. (1993) *Oncogene* **8,** 925–931
53. Yamada, Y., Nezu, J., Shimane, M., and Hirata, Y. (1997) *Genomics* **42,** 483–488