Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 366 (2006) 128 – 144
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Genomic structure, organization and promoter analysis of the human F11R/F11 receptor/junctional adhesion molecule-1/JAM-A

T. Sobocki [a,*], M.B. Sobocka [b], A. Babinska [c], Y.H. Ehrlich [d], P. Banerjee [a], E. Kornecki [c]

[a] *Program in Neuroscience and Department of Chemistry, College of Staten Island, CUNY, 2800 Victory Blvd., 6S-326, Staten Island, NY, 10314, USA*
[b] *Department of Medicine, SUNY Downstate Medical Center, 450 Clarkson Ave., Brooklyn, NY, 11203, USA*
[c] *Department of Anatomy and Cell Biology, SUNY Downstate Medical Center, 450 Clarkson Ave., Brooklyn, NY, 11203, USA*
[d] *Program in Neuroscience and Department of Biology, College of Staten Island, CUNY 2800 Staten Island, NY, 10314, USA*

Received 3 June 2005; received in revised form 11 August 2005; accepted 25 August 2005
Available online 5 December 2005
Received by A.J. van Wijnen

Abstract

The F11-receptor (F11R) (a.k.a. JAM-1, JAM-A, CD321) is a cell adhesion molecule of the immunoglobulin superfamily involved in platelet adhesion, secretion and aggregation. In addition, the F11R plays a critical role in the function of endothelial cells and in platelet adhesion to inflamed endothelium. In the present study, we used partial sequences of the human F11R gene, F11R cDNAs, and information in unannotated human genome databases, to delineate the F11R gene. We found that the F11R gene is composed of 13 exons (E1a, 1b, 1c, E1–E10) encoding two groups of mRNAs differing in length and sequence at their 5′ UTRs, referred to as type 1 and type 2 messages. Type 1 cDNAs are shorter at the 5′ end and contain a region not found within type 2 messages. Type 1 mRNAs are present in endothelial cells (EC), platelets, white blood cells and in the cell lines CMK, HeLa, K562, HOG and A549, while type 2 messages are limited to EC. Type 1 messages contain exons E1–E10 whereas type 2 messages usually contain exons E1a, 1c, part of E1 and E2–E10. The translation start site is localized in the 3′ end of E1, common for both type 1 and type 2 messages. Expression of these messages is regulated by two alternative promoters, P1 and P2. P1 is a TATA-less promoter containing an initiator element, multiple transcription start sites, several GC and CCAAT boxes, and GATA, NF-κB and ets consensus sequences. The cloned P1 drives efficient expression of the luciferase reporter gene. A high level of similarity between human P1 and its rat and mouse counterparts was observed. Promoter P2, located upstream of P1, contains a TATA box, GC boxes, a CCAAT box and GATA and ets consensus sequences. 3′ RACE provided evidence for variability in the 3′ UTR due to the presence of two polyadenylation signals. The finding of multiple regulatory sites in the promoters supplements the biochemical evidence that the F11R has several different roles in the functional repertoire of endothelial cells, platelets and other cells. In particular, the presence of NF-κB provides additional evidence to the significance of the F11R function in the initiation of inflammatory thrombosis.

Keywords: F11 receptor; F11R; JAM; JAM-A; Cell adhesion molecule; Immunoglobulin superfamily; Platelets; Endothelium; CD321

Abbreviations: CAM, cell adhesion molecule; CASK, Ca2+/calmodulin-dependent membrane-associated kinase; CPBP, core promoter binding protein; CRE, cAMP response element; E1a, 1b, …, exon 1a, 1b, …; CIP, calf intestinal phosphatase; CNS, central nervous system; DPE, downstream core promoter element; EC, endothelial cells; EKLF, erythroid Krüpel-like factor; F11R, F11 receptor; FCS, fetal calf serum; FGF, fibroblast growth factor; gDNA, genomic DNA; HGP, Human Genome Project; HOG, human oligodendroglioma; HUVEC, human umbilical vein endothelial cells; I1, 2, …, intron 1, 2, …; Ig, immunoglobulin; Inr, initiator; IRF, interferon regulatory factor; JAM-1, junctional adhesion molecule-1; JAM-A, junctional adhesion molecule-A; M.Ab.F11, monoclonal antibody F11; NF1, nuclear factor 1; NF-AT, nuclear factor of activated T-cells; NF-Y, nuclear factor Y; Oct, octamer-binding factor; PKC, protein kinase C; RACE, rapid amplification of cDNA ends; SP1, stimulating protein 1; TAP, tobacco acid pyrophosphatase; TGFβ, transforming growth factor β; TNF-α, tumor necrosis factor α; WBC, white blood cells

* Corresponding author. Tel.: +1 718 982 3993; fax: +1 718 982 3944.
E-mail address: tomasz_sobocki@mindspring.com (T. Sobocki).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.08.025

1. Introduction

The F11 receptor was discovered and characterized originally as the protein target of a stimulatory monoclonal antibody M.Ab.F11, which induces platelet activation and aggregation (Kornecki et al., 1990). Two forms of the F11R protein, 32 and 35 kDa, were identified, differing in degree of *N*-glycosylation (Naik et al., 1995). Subsequently, a cDNA clones providing the complete amino acid sequence of the human and mouse F11R proteins were obtained from platelets and epithelial cells (Martin-Padura et al., 1998; Ozaki et al., 1999; Sobocka et al., 2000; Naik et al., 2001). The F11R is a type I transmembrane protein containing two extracellular immunoglobulin folds and has been classified as an Ig-CAM. We have reported that human platelets express on their surface 8067±1307 F11R molecules/platelet (Kornecki et al., 1990).

The F11R is involved in the regulation of platelet aggregation and adhesion to the inflamed endothelium. The N-terminus of F11R and first Ig domain of F11R play critical roles in these processes (Babinska et al., 2002a,b). The F11R becomes phosphorylated in platelets activated by the physiological agonists thrombin and collagen, and M.Ab.F11 stimulatory antibody itself (Sobocka et al., 2000, 2004). Platelet activation by M.Ab.F11 involves cross-linking of F11R to FcγRII, followed by platelet shape changes as a result of cytoskeletal reorganization, filopodia formation, secretion of the granular contents and platelet aggregation (Kornecki et al., 1990; Naik et al., 1995; Sobocka et al., 2000). The biochemical events associated with these changes include activation and translocation of specific PKC isozymes (Wang et al., 1995), an increase in intracellular free-ionized calcium levels, activation of phosphatidylinositide 3-kinase, and phosphorylation of pleckstrin and myosin light chain (Kornecki et al., 1990; Sobocka et al., 2000, 2004).

In addition to its role in platelets, the same protein was also studied extensively in the endothelium. Immunofluorescence studies revealed F11R as a component of tight junctions in different types of epithelial cells (Martin-Padura et al., 1998; Williams et al., 1999; Gupta et al., 2000; Liang et al., 2000; Liu et al., 2000). In epithelium, F11R colocalizes with the tight junction components ZO-1 (Bazzoni et al., 2000; Ebnet et al., 2000) and calcium/calmodulin-dependent serine protein kinase, CASK (Martinez-Estrada et al., 2001). Association with the tight junction components is mediated by the PDZ type II-binding motif found to be present in the C-terminal region of F11R. Expression of F11R in HUVEC cells is regulated by the proinflammatory cytokines IFN-γ and TNF-α. Treatment of HUVEC with these cytokines caused a decrease of F11R in intercellular junctions with concomitant increase on the cell surface without change in the total number of expressed protein (Ozaki et al., 1999). The F11R has been identified as a potential target receptor for the T3 reovirus, which spreads to the CNS and causes lethal encephalitis (Barton et al., 2001). In cultured endothelial cells, reovirus infection induced apoptosis by activation of the nuclear factor NF-κB. Both NF-κB activation and apoptosis were found to be F11R-dependent.

Although all reported F11R cDNA sequences appear to encode an identical protein (except *AF154005.1*, Liu et al., 2000, which encodes a shorter protein of 259 amino acids), there are variations in lengths and nucleotide sequences of their 5′ untranslated regions (5′ UTRs). The longest 5′ UTRs are present in four cDNAs: *AF154005.1*, *AF111713* (Ozaki et al., 1999), *AF191495* (Gupta et al., 2000) and *AF172398.1* (Naik et al., 2001). Northern blot analysis of F11R, using as a probe the full coding region of the message (Williams et al., 1999; Naik et al., 2001), provided evidence for two F11R mRNAs of 2.0–2.4 and 4.4 kb. It suggests that the 4.4 kb message contains a 5′ and/or a 3′ UTR sequence that is longer than cDNA clones identified thus far.

The study presented here addresses the issue of variability among the F11R mRNAs, proposes the organization of the F11R gene with localization of the F11R promoters and presents detailed analysis of the 3′ UTR demonstrating the presence of two major F11R species.

2. Materials and methods

2.1. Reagents

The GeneRacer™ Kit for full-length, RNA ligase-mediated rapid amplification of 5′ and 3′ cDNA ends (RLM-RACE), the TOPO-TA™ Cloning Kit and Thermozyme™ thermostable DNA polymerase, the THERMOSCRIPT™ RT-System, the Platinum® Taq High Fidelity thermostable DNA polymerase, Agarose 1000™ and a dNTPs mixture were purchased from Invitrogen Corporation (Carlsbad, CA). The QIAprep Miniprep plasmid DNA isolation kit, QIAquick PCR purification kit, QIAquick gel extraction kit, the Effectene™ Transfection Reagent, RNeasy Mini kit including QIAShredder™ columns and on-column DNase digestion kit were purchased from QIAGEN (Valencia, CA). The pGL3-Basic (no promoter) and pGL3-Promoter (SV40 promoter) firefly luciferase reporter plasmids, pRL-TK vector, Luciferase Assay System and the Wizard® Genomic DNA Purification Kit were purchased from Promega (Madison, WI). Restriction endonuclease *EcoRI* was purchased from New England Biolab (Beverly, MA) and *KpnI* and *NheI* from Promega. All other reagents were obtained from Sigma (St. Louis, MO).

2.2. PCR primer design

Primer design for PCR of genomic DNA or cDNA was conducted by use of the rawprimer program (http://alces.med.umn.edu/rawprimer.html). All custom primers were synthesized by Invitrogen and their sequences are listed in Fig. 1B. Primers (oligo dT_{20} and random hexamers) used for reverse transcription were included in the THERMOSCRIPT™ RT-PCR System. Primers used for amplification of the 5′ and 3′ cDNA ends: GeneRacer™ 5′ [forward] Primer (GeneRacer 5′F), GeneRacer™ 5′ Nested [forward] Primer (GeneRacer 5′NF) and GeneRacer™ 3′ reverse were included in the GeneRacer™ kit.

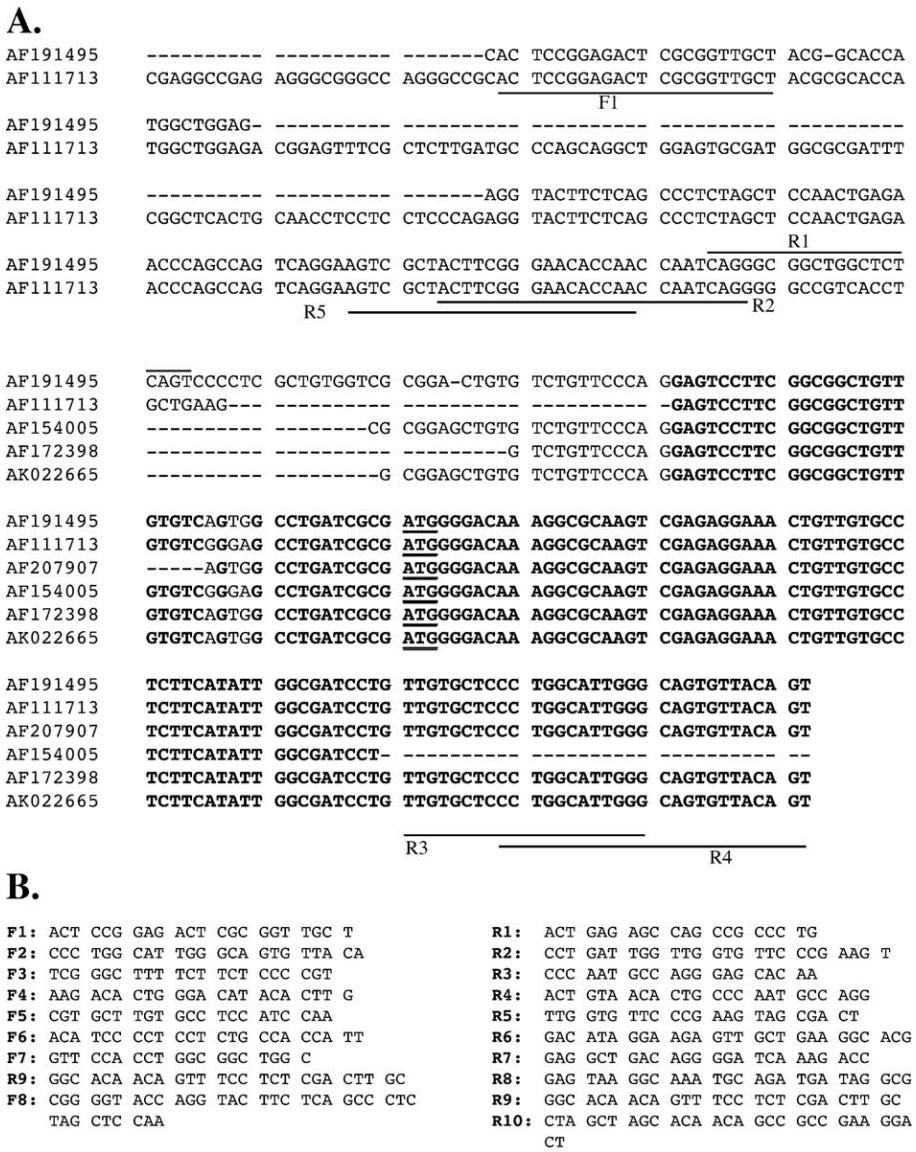


Fig. 1. 5′ ends of F11R cDNAs and primers used in genomic and cDNA studies. Panel (A) 5′ ends of selected F11R cDNAs. In bold are shown nucleotides, which are identical in all cDNAs. The translation start site is underlined. Included are primers selected for 5′ RACE (R2–R5) and primers used for identification of the type 2 mRNAs (F1/R1). Panel (B) Sequences of designed primers. 'F' in the primers' name indicates a forward primer, 'R' a reverse primer. Primers F2–F6, R3, R4 and R6–R8 were designed based on the sequences of F11R cDNAs see Panel (A). F1, R2, R5 on the sequence of the clone *AF111718*; R1-clone *AF191495*. The sequences of primers F7 and R9 are designed based on the 5′ RACE clone C3 (Fig. 2). For details on primers F8 and R10 refer to Section 2.9 and Fig. 7A.

2.3. Cells and cell lines

Human lung carcinoma, A549 and HeLa cells were obtained from Dr. John Lewis (Anatomy and Cell Biology, SUNY Downstate, Brooklyn, NY) and grown in DMEM media plus 10% FCS. CMK, human megakaryoblastic cells, obtained from Dr. Hava Avraham (Hematology and Oncology Research Laboratory, New England Deaconess Hospital of Boston, MA) were maintained in RPMI 1640 medium with 10% FBS. Two batches of HUVEC cells (passage 3–6) were obtained from Dr. O. Batuman, (Medicine, SUNY Downstate) (HUVEC-1) and from Dr. William A. Muller, Cornell Medical School, New York, NY (HUVEC-2). HUVEC-1 were grown in 10% FCS with ECGS (endothelial cell growth supplement, Invitrogen), whereas HUVEC-2 were grown in 10% FCS without any growth supplements. K562 human myelogenous leukemia cells were obtained from Dr. Allen Norin, (Anatomy and Cell Biology, SUNY Downstate). White Blood Cells (WBC) were freshly isolated from human blood. In brief, whole blood, obtained from drug-free consenting healthy adult donors, was collected into 1/10 volume of the anti-coagulant sodium citrate (100 mM, pH 6.5), plus 140 mM glucose. The protocol for the preparation of WBC and isolation of gDNA from these cells was followed according to procedures provided in the Wizard® Genomic DNA Purification Kit (Promega). Steps 1–6 of section III B of the manufacturer's manual were followed for the isolation of WBC, and subsequently utilized for RNA isolation; steps 1–16 of this protocol were followed for the isolation of gDNA from these cells (Section 2.6).

2.4. Preparation of RNA

RNA from cells was isolated using the RNeasy kit and subjected to oncolumn DNase digestion for 30 min at 23 °C. The quality of the isolated RNA was verified by formaldehyde/formamide agarose gel electrophoresis.

2.5. Mapping 5′ and 3′ ends of the F11R cDNA using 5′ RACE and 3′ RACE

RACE procedures were performed using the GeneRacer™ kit according to the manufacturer's protocol. RNA prepared from K562, CMK, HeLa, HUVEC and WBC was dephosphorylated using CIP and treated with TAP. After ligation of the GeneRacer™ RNA oligo, reverse transcription was carried out using Avian Myeloblastosis Virus Reverse Transcriptase and GeneRacer™ Oligo dT Primer.

2.5.1. Identification of the type 1 mRNAs (5′ RACE)

Primary amplification was carried out using the GeneRacer 5′F primer and cDNA specific primers R4 (clones C1–C8) or R7 (clones C9 and C10). Primary products were obtained using Thermozyme DNA polymerase. PCR conditions (primary amplification, touchdown PCR) were: hot start 94 °C/2 min; 5 cycles of 94 °C/30 s, 70 °C/30 s and 72 °C/4 min; 7 cycles of 94 °C/30 s, 68 °C/30 s and 72 °C/4 min; 20 cycles of 94 °C/30 s, 65 °C/30 s and 72 °C/4 min; and a final extension 72 °C/5 min. Secondary PCR amplification was carried out using 1 μl of 25× diluted primary product and two nested primers: GeneRacer 5′ NF and either primers R3 (clones C6–C8) or R6 (clones C9, C10). For secondary PCR amplification the conditions were: 94 °C/1 min 45 s; 20 cycles of 94 °C/20 s, 65 °C/20 s and 68 °C/3 min; and a final extension at 68 °C/10 min. Products of PCR were analyzed by agarose gel electrophoresis (3% Agarose 1000™). Secondary PCR products from HeLa, WBC, K562 and HUVEC cells were cloned as follows: after electrophoresis, the major PCR products shown in Fig. 2A were excised, extracted from the gel using QIAquick Gel extraction kit and cloned into the pCR4-TOPO TA cloning vector. All clones were sequenced.

2.5.2. Ubiquitous expression of the type 1 message

Total RNA was isolated from all cells, cDNAs were obtained using the Thermoscript® RT kit and treated with RNAseH. For PCR, 1/10th of the cDNA obtained in one RT reaction was used. PCR amplification was carried out using F7 and R9 primers (Fig. 1B) and products shown in Fig. 3A. PCR conditions were: 94 °C/1 min 45 s; 36 cycles of 94 °C/20 s, 65 °C/20 s and 68 °C/35 s; and a final extension at 68 °C/1 min. PCR products were analyzed by 3% Agarose 1000™ gel electrophoresis.

2.5.3. Identification of the type 2 message

RNA was isolated from all cells as detailed (Section 2.4.) and cDNA was obtained using the Thermoscript® RT kit (Section 2.5.2). Primers used for PCR were F1 and R9 and conditions were 94 °C/1 min 45 s; 36 cycles of 94 °C/20 s, 65 °C/20 s and 68 °C/1 min; and a final extension at 68 °C/1 min. PCR products were analyzed by 3% Agarose 1000 gel electrophoresis (Fig. 3B, lane 1).

2.5.4. Type 2 mRNAs (5′ RACE)

PCR conditions (primary amplification, touchdown PCR) were: hot start 94 °C/1 min 45 s; 5 cycles of 94 °C/30 s, 67 °C/30 s and 68 °C/4 min; 10 cycles of 94 °C/30 s, 65 °C/30 s and 68 °C/4 min; 20 cycles of 94 °C/30 s, 62 °C/30 s and 68 °C/4 min; and a final extension at 68 °C/5 min. Primers used were GeneRacer 5′F and R2. Secondary amplification utilized 1 μl of 25× diluted primary amplification products in a 50 μl reaction. The PCR conditions were: 94 °C/1 min 45 s; 20 cycles of 94 °C/30 s, 62 °C/20 s and 68 °C/4 min; and a final extension at 68 °C/10 min. Primers used were GeneRacer 5′NF and R5. PCR products were analyzed by 3% Agarose 1000 gel electrophoresis. PCR products of ~160 and ~260 bp (Fig. 3B, lane 2) were cloned and sequenced.

2.5.5. 3′ RACE

Primers used were F5 and GeneRacer™ 3′ reverse. PCR conditions (touchdown PCR) were: hot start 94 °C/1 min 45 s; 5 cycles of 94 °C/30 s, and 70 °C/6 min 30 s; 10 cycles of 94 °C/30 s and 68 °C/6 min 30 s; 20 cycles of 94 °C/30 s, 65 °C/30 s and 68 °C/6 min; followed by a final extension at 68 °C/10 min. PCR products were analyzed by 0.8% agarose gel electrophoresis.

2.6. PCR amplification of genomic DNA

Genomic DNA isolated from human WBC was prepared using the Wizard® Genomic DNA Purification Kit. The quality of the DNA was tested by agarose gel electrophoresis. The average size of genomic DNA was 50–100 kb. PCR amplification of genomic DNA (10 or 100 ng) utilized Platinum® High Fidelity Taq polymerase and the following primer pairs: F4/R6, F3/R6, F2/R7, F2/R6, F6/R8 and F5/R8. For the first four primer pairs, the PCR conditions were: 94 °C/1 min; 5 cycles of 94 °C/15 s, 65 °C/30 s and 68 °C/12 min; 30 cycles of 94 °C/15 s, 60 °C/30 s and 68 °C/12 min; and a final extension at 68 °C/12 min. For the last two primer pairs, the PCR conditions were: 94 °C/30 s; 5 cycles of 94 °C/15 s, 65 °C/30 s and 68 °C/12 min; 30 cycles of 94 °C/15 s, 58 °C/30 s and 68 °C/12 min; and a final extension at 68 °C/12 min. PCR products were analyzed by 1% agarose gel electrophoresis.

2.7. DNA sequencing

Sequencing of the clones was performed by ACGT Inc. (Northbrook, IL). Each clone was sequenced in both directions using universal forward and reverse primers. In addition, clones G14 and G15 were sequenced using internal primers: GAA TGG GTA TGG GAC ACC (sense) and ACT AGG CTG GCT GTA AAT CAC (antisense).

2.8. Identification of transcription factor binding sites

The search was performed using the MatInspector (Quandt et al., 1995). In addition to the computer search, the sequences

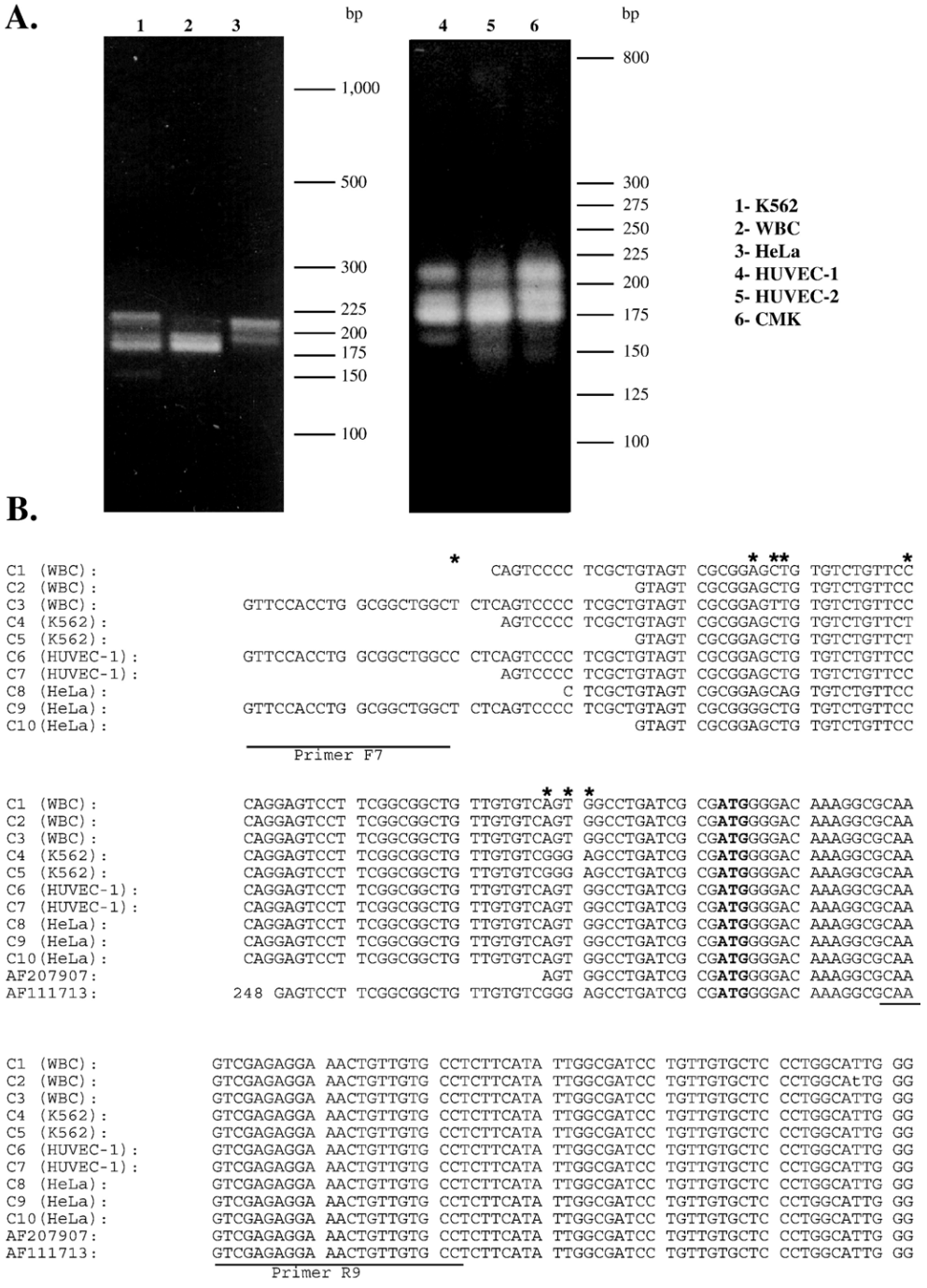
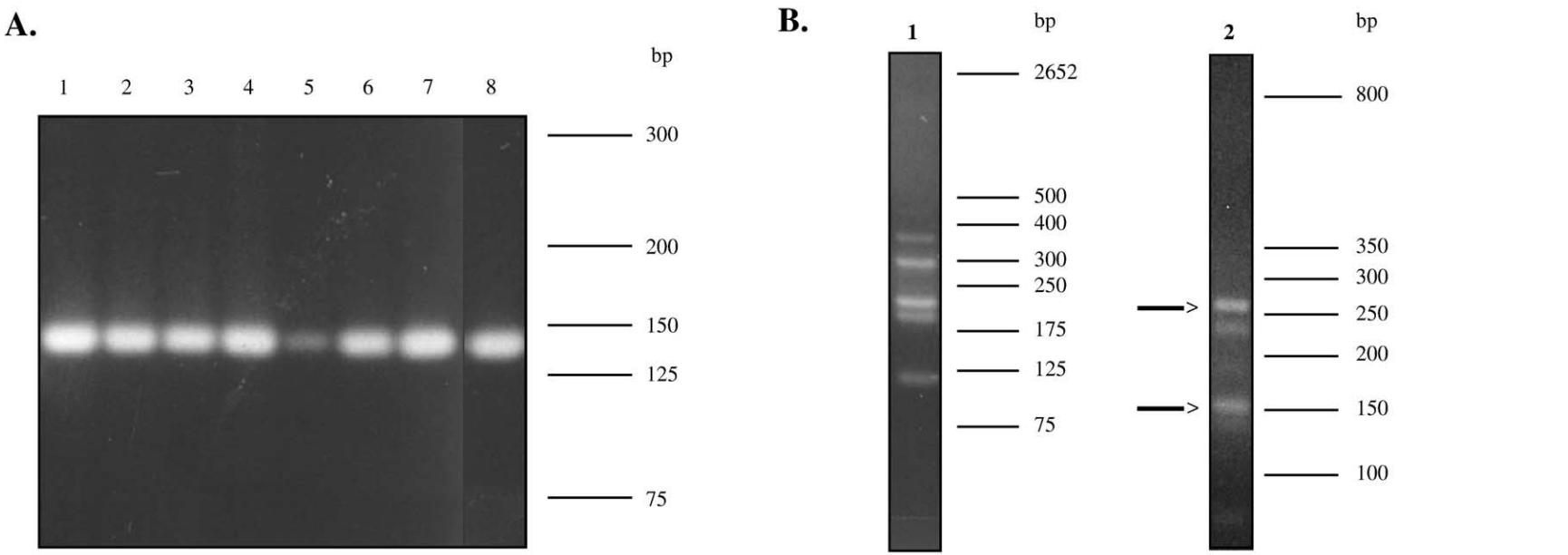


Fig. 2. 5′ RACE amplification of the type 1 F11R cDNAs. Panel (A): Amplification of cDNAs from selected cells using 5′ RACE procedure (see Section 2.5.1). Lanes 1–3; 5′ RACE products from K562, WBC and HeLa. Lanes 4–6: 5′ RACE products from HUVEC-1, HUVEC-2 and CMK cells. Panel (B): Analysis of the representative 5′ RACE clones from HUVEC-1, K562, HeLa and WBC. Two F11R cDNAs, *AF111713* and *AF207907*, are included for comparison. Translation start site (ATG) is in bold letters. Primers F7 and R9 (underlined) were designed based on sequence of the longest clones and used for amplification of the type 1 transcript shown in Fig. 3. The asterisk (*) indicates position with difference in sequence.

were analyzed using optimized weight matrices defining TATA-box, CCAAT and GC-boxes and for the presence of an initiator and core DPE elements (Bucher, 1990; Burke et al., 1998).

2.9. Transcriptional activity of promoter P1

The PCR amplification was carried out directly from genomic DNA using Platinum® Taq High Fidelity DNA polymerase and primers containing restriction sites at their 5′ ends: *KpnI* in the forward primer (F8) and *NheI* in the reverse (R10). Analysis of the PCR products by agarose gel electrophoresis revealed a single product of the expected size (not shown). The amplified PCR product was cloned into pGL3 Basic reporter vector. Sequencing of the promoter insert on both strands confirmed its identity. The construct F11R/P1 was transiently transfected into cells using the Effectene™ Reagent. Transient



C.

```
      C11  ---------- -----GGGCC AGGGCCGCAC TCCGGAGACT CGCGGTTGCT ACGCGCACCA TGGCTGGAG- ---------- ---------- ---------- ---------- ----------  54
      C12  ---------- -----GGGCC AGGGCCGCAC TCCGGAGACT CGCGGTTGCT ACGCGCACCA TGGCTGGAG- ---------- ---------- ---------- ---------- ----------  54
      C13  ---------- ---------- ------GCAC TCCGGAGACT CGCGGTTGCT ACGCGCACCA TGGCTGGAGC GCCCACGGTC TCGCTTCCTG AACTCCGTTC ACTCCTAGCC TCCGGACGGG  94
AF111713  CGAGGCCGAG AGGGCGGGCC AGGGCCGCAC TCCGGAGACT CGCGGTTGCT ACGCGCACCA TGGCTGGAG- ---------- ---------- ---------- ---------- ----------  69

      C11  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------  54
      C12  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------  54
      C13  CCCGGCTCTT CGACGTGCGC TCTCGCGAGG AGGCGGCAGC TGGGACCATC CCAGGGGCGC TCAACATCCC GG-------- ---------- ---------- ---------- ---------- 166
AF111713  ---------- ---------- ---------- ---------- ---------- ---------- ---------- --ACGGAGTT TCGCTCTTGA TGCCCAGCAG GCTGGAGTGC GATGGCGCGA 117

      C11  ---------- ---------- ---------- AGGCACTTCT CAGCCCTCTA GCTCCAACTG AGAACCCAGC CAGTCAGGAA GTCGCTACTT CGGGAACACC AA 126
      C12  ---------- ---------- ---------- AGGTACTTCT CAGCCCTCTA GCTCCAACTG AGAACCCAGC CAGTCAGGAA GTCGCTACTT CGGGAACACC AA 126
      C13  ---------- ---------- ---------- AGGTACTTCT CAGCCCTCTA GCTCCAACTG AGAACCCAGC CAGTCAGGAA GTCGCTACTT CGGGAACACC AA 238
AF111713  TTTCGGCTCA CTGCAACCTC CTCCTCCCAG AGGTACTTCT CAGCCCTCTA GCTCCAACTG AGAACCCAGC CAGTCAGGAA GTCGCTACTT CGGGAACACC AA 219
```

Fig. 3. Expression of type 1 and 2 mRNAs. Panel (A): Expression of type 1 message in vascular primary cells and cell lines. Lanes 1–8 indicate the cells and cell lines used: lane 1—HUVEC-2, lane 2—A549, lane 3—HOG, lane 4—CMK, lane 5—WBC, lane 6—HeLa, lane 7—K562 and lane 8—HUVEC—1. PCR amplification was carried out as described in Section 2.5.2. Panel (B): Identification of the type 2 messages in HUVEC-2. Lane 1: Expression of type 2 messages in HUVEC-2. PCR was carried out on cDNA with primers F1 and R9, as described in Section 2.5.3. Lane 2: 5′ RACE products obtained using Gene Racer 5′F/R2 and Gene Racer 5′NF/R5 primers (see Section 2.5.4). Products of ~270 and ~160 bp (see arrows) were cloned. Panel (C): Analysis of the type 2 clones from HUVEC-2 cells. PCR products (Panel B) of ~160 bp (clones C11 and C12) and ~270 bp band (clone C13) were cloned, sequenced and aligned to the 5′ end of the cDNA clone *AF111713*. Clone C13 contains exon 1a′ (in bold) of 123 bp (see Section 3.4).

transfections with F11R/P1 promoter constructs in pGL3Basic (firefly luciferase) and pRL-TK (*Renilla* luciferase) into HOG and EA.hy926 cells were carried out using Effectene™ reagent. Cotransfection mixtures (in 24-well plates) contained: promoter construct (195 ng for EA.hy926 and 200 ng for HOG), pRL-TK (5 ng for EA.hy926 and 2.5 ng for HOG), DNA condensation buffer (60 μl), Enhancer (1.6 μl), Effectene™ (5 μl) and DMEM —10% FBS—1% Penicillin–Streptomycin (350 μl). The cotransfection mixture was added to cells overlayered with 350 μl of DMEM—10% FBS—1% Penicillin–Streptomycin. The medium plus cotransfection mixes were subsequently removed after either 4 (EA.hy926) or 24 h (HOG) and the cells were overlayered with DMEM—10% FBS—1% Penicillin–Streptomycin (1 ml). Cells were harvested 48 h after transfection, and luciferase activities were assayed using a dual luciferase assay kit (Promega) as specified by the manufacturer. Luminescence from either firefly or Renilla luciferase was measured using a TD20/20 Luminometer (Turner Design, Sunnyvale, CA). Results were expressed as a ratio of Firefly to Renilla luminescence.

3. Results

3.1. Analysis of the 5′ UTR in human F11R mRNAs

Fig. 1A compares six human F11R cDNA sequences. The longest 5′ ends are displayed by two cDNAs, *AF111713* and *AF191495*, that have sequences that are not identical. *AF111713* sequence has a longer 5′ end, contains 78 bp insert and differs from the *AF191495* sequence from base 229. Other cDNAs are shorter and may represent incomplete stretches at their 5′ end cDNAs.

The 5′ RACE was used to investigate the transcriptional start site of the F11R mRNA in cells that express F11R: K562, HeLa, CMK, WBC and endothelial cells of two batches HUVEC-1 and -2. Reverse primers for both primary and secondary amplification (R4 or R7 and R3 or R6, respectively) were designed based on the sequence common for cDNAs (Fig. 1B). The PCR products obtained are shown in Fig. 2A; their sizes range from ~180 to 220 bp.

3.2. Sequence analysis of the 5′ RACE clones

Fig. 2B provides the full sequences for ten 5′ RACE clones labeled C1 to C10 and partial sequences of two F11R cDNAs, *AF207907* and *AF111713*. All of the clones align with *AF207907* but their 5′ ends are much longer. Three clones, C3, C6 and C9 have the longest 5′ UTR and begin exactly at the same position, 87 bp upstream of the 5′ end of *AF207907*. Other sequences are shorter: clones C4 and C7 contain additional 63 bp in the 5′ terminus and clones C2, C5 and C10 are 52 bp longer than *AF207907*. Finally, clones C1 and C8 begin 72 and 58 bp upstream of F11R cDNA. Sequence differences found at six positions (indicated with asterisks in Fig. 2B) most likely represent polymorphism. Expression of messages containing the longest 5′ UTR sequence were found in all cells (shown in Fig. 3A), although WBC (lane 5) expressed lower levels of the transcript.

As shown in Fig. 2B the sequences of the clones are identical to *AF111713* from nucleotide 248, however their 5′ ends are different.

Comparison of sequences for C1–C10 clones with *AF111713*, as shown in Fig. 2B, indicates that 5′ RACE products obtained here were not truncated versions of *AF111713*. Our methods used the GeneRacer™ procedure, yielding 5′ RACE products that could have originated only from full length mRNA. Thus, 5′ RACE procedure carried out on 5 cell lines produced 5′ ends, which were 52 to 87 bases longer than *AF207907*. In no instance was a product obtained with a 5′ end of the full length and sequence of *AF111713*. The 5′ end of *AF207907* was obtained using an earlier 5′ RACE procedure (as detailed in Sobocka et al., 2000), a procedure, which at that time did not eliminate truncated mRNAs, and mRNAs that lacked a cap structure. Accordingly, our results provided evidence that cDNAs with shorter 5′-UTRs, as those sequenced here, are not truncated versions of messages with long 5′ UTRs and suggest that the expression of short and long messages is under the control of different promoters. Promoter P1 (P1) was termed to identify transcripts with short 5′ UTR, called type 1 mRNAs. Longer transcripts of the type reported for *AF111713* were termed type 2 mRNAs, and regulated by the Promoter P2 (P2).

3.3. Identification of type 2 messages

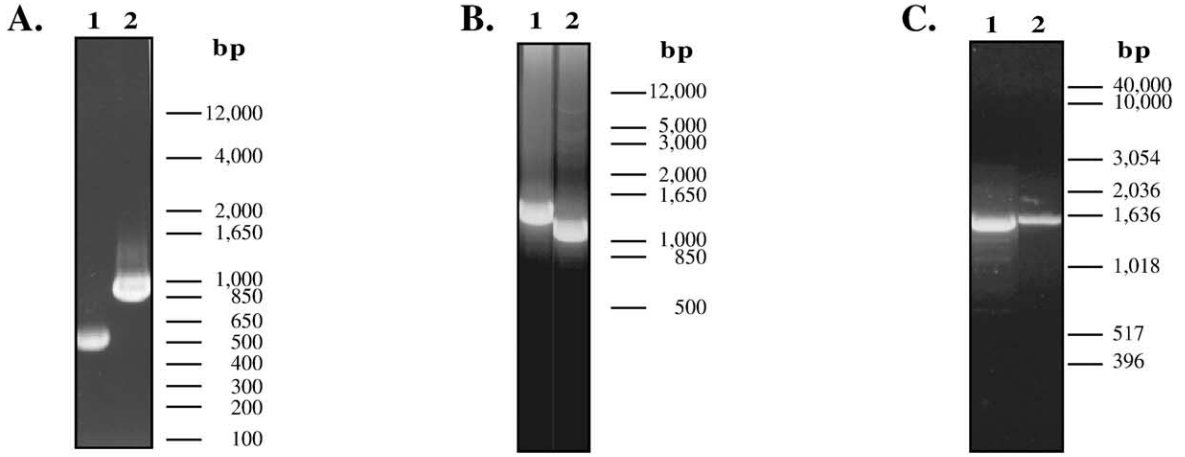
Type 2 mRNAs (expected size ~400 bp) were not observed among the results depicted in Fig. 2, although amplification of these messages by 5′ RACE was expected, using R3 and R4 primers specific for both types of messages. To determine whether these cells do not express type 2 mRNA or whether the level of type 2 mRNAs is lower than type 1 mRNAs, primers were designed based on the type 2 cDNAs: *AF111713* and *AF191495*. PCR using these primers would be expected to amplify only type 2 messages. Of all the cells examined, only HUVEC-2 was found to express type 2 messages. As shown in Fig. 3B, lane 1, amplification of HUVEC-2 cDNA using primers F1 and R9 produced products ranging from 115 to 380 bp, which may represent alternatively spliced forms of the type 2 message.

Although *AF111713* and *AF191495* appear as type 2 cDNAs, there are significant differences identified in their 5′ UTRs sequences (Fig. 1A). The 5′ end of *AF191495* has two stretches (bases 1–42 and 43–123) identical to *AF111713*. Bases 124–174 of *AF191495* were not found within *AF111713*, although this region was found in clone C3 as bases 12–63. No products were obtained with primers F1 and R1 designed to detect *AF191495*.

3.4. 5′ RACE analysis of type 2 messages

At least three PCR products with size of approximately 160, 230 and 260 bp were detected in HUVEC-2 (Fig. 3B, lane 2). Two major products (160 and 260 bp, see arrows) were excised from the gel, eluted, cloned and sequenced. The alignment of the sequences of clones (C11, C12 and C13) with *AF111713* is shown in Fig. 3C. Each clone contains a 5′ sequence identical to *AF111713*, although they are shorter than *AF111713* and did

not contain a sequence corresponding to nucleotides 70–147. In addition, clone C13 contained 123 bp (bp 44–166) not present in *AF111713* but found in messages encoding the KAT protein, unrelated to the F11R (Wenzel et al., 2003). These 123 bp align with a putative exon (see Fig. 9, exon *1a′*) located in the F11R gene downstream to the first exon, E1a. Alignment of the



D.

```
cons      CCCTGGCATT GGGCAGTGTT ACAGTGCACT CTTCTGAACC TGAAGTCAGA ATTCCTGAGA ATAATCgtga gttgggaggg gccatggagg gtgtgagggt gagcagttgc cggccgcctg gggatctaga
AF207907  CCCTGGCATT GGGCAGTGTT ACAGTGCACT CTTCTGAACC TGAAGTCAGA ATTCCTGAGA ATAATC  (148)

cons      gagcacccag cccagcctgc agtttggggc tgttccccat ctcgtgtatt tgctgctgct tcttccttcg gctcattttg accctgtctg cagCTGTGAA GTTGTCCTGT GCCTACTCGG GCTTTTCTTC
AF207907                                                                                                    (149)  CTGTGAA GTTGTCCTGT GCCTACTCGG GCTTTTCTTC

cons      TCCCCGTGTG GAGTGGAAGT TTGACCAAGG AGACACCACC AGACTCGTTT GCTATAATAA CAAGATCACA Ggtgagttgc ttctcctcct tcctgattgc ctaggttgtg gagggattat catgcctgga
AF207907  TCCCCGTGTG GAGTGGAAGT TTGACCAAGG AGACACCACC AGACTCGTTT GCTATAATAA CAAGATCACA G  (256)

cons      ttctacgatc cagactttgt gcccttatcc cagggcacac cctggggctg aagtcctcat ggccttccta ccccccttttc tgggggaaga ggagagcact tgcgtacctt ggatgccccc ttcatcagcc

cons      tcccttaact ctccagttca ctctgtcctt gctcttgcct cctcttgtgg tagCTTCCTA TGAAGACCGG GTGACCTTCT TGCCAACTGG TATCACCTTC AAGTCCGTGA CACGGGAAGA CACTGGGACA
AF207907                                                (257)    CTTCCTA TGAGGACCGG GTGACCTTCT TGCCAACTGG TATCACCTTC AAGTCCGTGA CACGGGAAGA CACTGGGACA

cons      TACACTTGTA TGGTCTCTGA GGAAGGCGGC AACAGCTATG GGGAGGTCCA GGTCAAGCTC ATCGTGCTTG gtacgtgccc catgtcttct gggtgggcct ggagttagtt acttctcata gcaggctctg
AF207907  TACACTTGTA TGGTCTCTGA GGAAGGCGGC AACAGCTATG GGGAGGTCAA GGTCAAGCTC ATCGTGCTTG (403)

cons      ctgcactttg gagctctttg atggtgagaa tacccacagg tgggctattc agagtcctgg agctgccaga gaggaagatg agcccctggt taggtggcag ccaggggcat ggtgagagta ggaggcttga

cons      gctgagtagt tgggggtacc attgtagttg gttacttaga ctcaagcagc ccccaccctg tactcagcac cttctgtctt tctcttccac agTGCCTCCA TCCAAGCCTA CAGTTAACAT CCCCTCCTCT
AF207907                                                                                                    (404)  TGCCTCCA TCCAAGCCTA CAGTTAACAT CCCCTCCTCT

cons      GCCACCATTG GGAACCGGGC AGTGCTGACA TGCTCAGAAC AAGATGGTTC CCCACCTTCT GAATACACCT GGTTCAAAGA TGGGATAGTG ATGCCTACGA ATCCCAAAAG CACCCGTGCC TTCAGCAACT
AF207907  GCCACCATTG GGAACCGGGC AGTGCTGACA TGCTCAGAAC AAGATGGTTC CCCACCTTCT GAATACACCT GGTTCAAAGA TGGGATAGTG ATGCCTACGA ATCCCAAAAG CACCCGTGCC TTCAGCAACT

cons      CTTCCTATGT CCTGAATCCC ACAACAGGAG AGCTGgtatg tatggggtgg tggtgaaaga tgtgtggggg atgcagttat agaaccccaa aaggtgggag gaagaacagg caacagggtg aacttggaac
AF207907  CTTCCTATGT CCTGAATCCC ACAACAGGAG AGCTG  (606)

cons      tggggtcaag agacattaaa aaatgtctat tttgcttctt cagcaggaag ccaatttgtt ttccttcctc agGTCTTTGA TCCCCTGTCA GCCTCTGATA CTGGAGAATA CAGCTGTGAG GCACGGAATG
AF207907                                                                                                (607)   GTCTTTGA TCCCCTGTCA GCCTCTGATA CTGGAGAATA CAGCTGTGAG GCACGGAATG

cons      GGTATGGGAC ACCCATGACT TCAAATGCTG TGCGCATGGA AGCTGgtgaa aattaaggct tgagcccaga cctggggtgg ggattggggt ctgcttttca tcctgccatt ggtgatactg taggtaccgt
AF207907  GGTATGGGAC ACCCATGACT TCAAATGCTG TGCGCATGGA AGCTG  (709)

cons      gagtgagtat ccccgtgcct gacttcttat ttgtgtctcg cagTGGAGCG GAATGTGGGG GTCATCGTGG CAGCCGTCCT TGTAACCCTG ATTCTCCTGG GAATCTTGGT TTTTGGCATC TGGTTTGCCT
AF207907                                      (710) TGGAGCG GAATGTGGGG GTCATCGTGG CAGCCGTCCT TGTAACCCTG ATTCTCCTGG GAATCTTGGT TTTTGGCATC TGGTTTGCCT

cons      ATAGCCGAGG CCACTTTGAC Agtaagtatc tgcccccaga ggctctcctt tgtactgccc ccatcccagg gcctggcatg ggtgtcatct gagtactgac cctgatactg tgctcatgtg tgtgggtgtt
AF207907  ATAGCCGAGG CCACTTTGAC A    (817)

cons      gtcccctggg gaacactgat cacccatcta acaccaagag ctgggggcct gcccctccat cttcccaagc catgctgtct tctgtcctgc tgacactcat gatcccattt cttctctttc agGAACAAAG
AF207907                                                                                                                                (818)   GAACAAAG

cons      AAAGGgtgag tgaggtgctg tcctggggtt ctccaagttt gagagcatgg atgcatgtgg tttgaagctg aagtggacct aggggaatgg gttgaaggca gaagcaacca gtttggaggg aaggcatttg
AF207907  AAAGG  (830)

cons      gatatccagc cctttctctg tggcctcggc cctgggtctg tcctgttacc ccacccatac ctgtctgctg cgcactctgt gcttctgtag cattctcgct tctggccttt aaagttggca aggggaggtt

cons      aataagcacc taggtggctg agtgtctctg tcttctggct tgttcacagG ACTTCGAGTA AGAAGGTGAT TTACAGCCAG CCTAGTGCCC GAAGTGAAgt gagtatgcct cccctgggca ggggtgggcc
AF207907                                                (831) G ACTTCGAGTA AGAAGGTGAT TTACAGCCAG CCTAGTGCCC GAAGTGAA  (879)

cons      acctggggct ggagtgagga gatttctagc ccatgtttat gtgttttggg gattatcagc taaagaagac tattgagagt tgactggatg ttttatttct gcagGGAGAA TTCAAACAGA CCTCGTCATT
AF207907                                                                                                                (880) GGAGAA TTCAAACAGA CCTCGTCATT

cons      CCTGGTGTGA GCCTGGTCGG CTCACCGCCT ATCATCTGCA TTTGCCTTAC TC
AF207907  CCTGGTGTGA GCCTGGTCGG CTCACCGCCT ATCATCTGCA TTTGCCTTAC TC (957)
```

Fig. 4. Amplification of a fragment of the F11R gene from human genomic DNA using primers spanning bases 83–957 in *AF207907* F11R cDNA. Panels (A)–(C) Genomic DNA from WBC was amplified using the following primer pairs: Panel (A) lane 1, primers F4/R6; lane 2, F3/R6; Panel (B) lane 1, F2/R7; lane 2, F2/R6; and Panel (C) lane 1, F6/R8; lane 2, F5/R8. The PCR products shown in Panel (B) (lane 1) and Panel (C) (lane 2) were cloned and sequenced (clones G14–G17, Fig. 5, Panel A). Panel (D) Consensus sequence (cons) resulting from the alignment of the genomic clones G14–G17. The exons (in upper case) and introns (in lower case) are identified using the cDNA sequence *AF207907*.

sequenced clones allowed us to identify transcription start sites for type 2 mRNAs.

3.5. Initial cloning of genomic DNA

At the time this investigation was initiated, no information was available on the F11R gene and therefore our first attempts focused on obtaining partial genomic sequence of this gene. Primers for the amplification of genomic DNA were selected based on the F11R cDNA (Fig. 1B). Panels A–C, in Fig. 4, show that single PCR products were obtained with each primer combination. The PCR products obtained with primers F5/R8 (clones G14 and G15, Panel C, lane 2) and with F2/R7 primers (clones G16 and G17, Panel B, lane 1) were sequenced. The derived genomic consensus sequences were aligned with the F11R cDNA (*AF207907*). As shown in Fig. 4D, eight exons were identified with sequences identical to counterparts at the mRNA level. Correspondingly, 8 introns were identified with the terminal 5′ and 3′ sequences matching the typical GT-AG pattern.

3.6. Organization of the F11R gene

A BLAST search of the genomic sequences of clones G14, G15, G16 and G17 against unannotated sequences (database 'htgs', Unfinished High Throughput Genomic Sequences: phases 0, 1 and 2) revealed >99.9% identity to a region within contigs of human chromosome 1, that of clone RP11-381D2 with accession number *AL162592* (ver. 8, GI: 8977662) and clone RP11-312J18 (accession number *AL354714* ver. 5, GI: 8977674 and ver. 7, GI: 10277938). Moreover, the sizes of the PCR products could be explained by analysis of the sequence of these two clones RP11-381D2 and RP11-312J18. Using the primer pair F4/R6, the calculated MW was 544 bp whereas the observed size was ~500 bp (Fig. 4A, lane 1). Likewise, using F3/R6, the calculated size of 934 vs. observed ~900 bp (Fig. 4A, lane 2); F2/R7, calculated size of 1396 bp vs. observed ~1.3–1.4 kb (Fig. 4B, lane 1), F2/R6, calculated 1181 bp vs. observed ~1.1–1.2 kb (Fig. 4B, lane 2), F6/R8, calculated 1496 bp vs. observed ~1.5–1.6 kb (Fig. 4C, lane 1) and F5/R8, calculated 1528 bp vs. observed ~1.5–1.6 kb (Fig. 4C, lane 2).

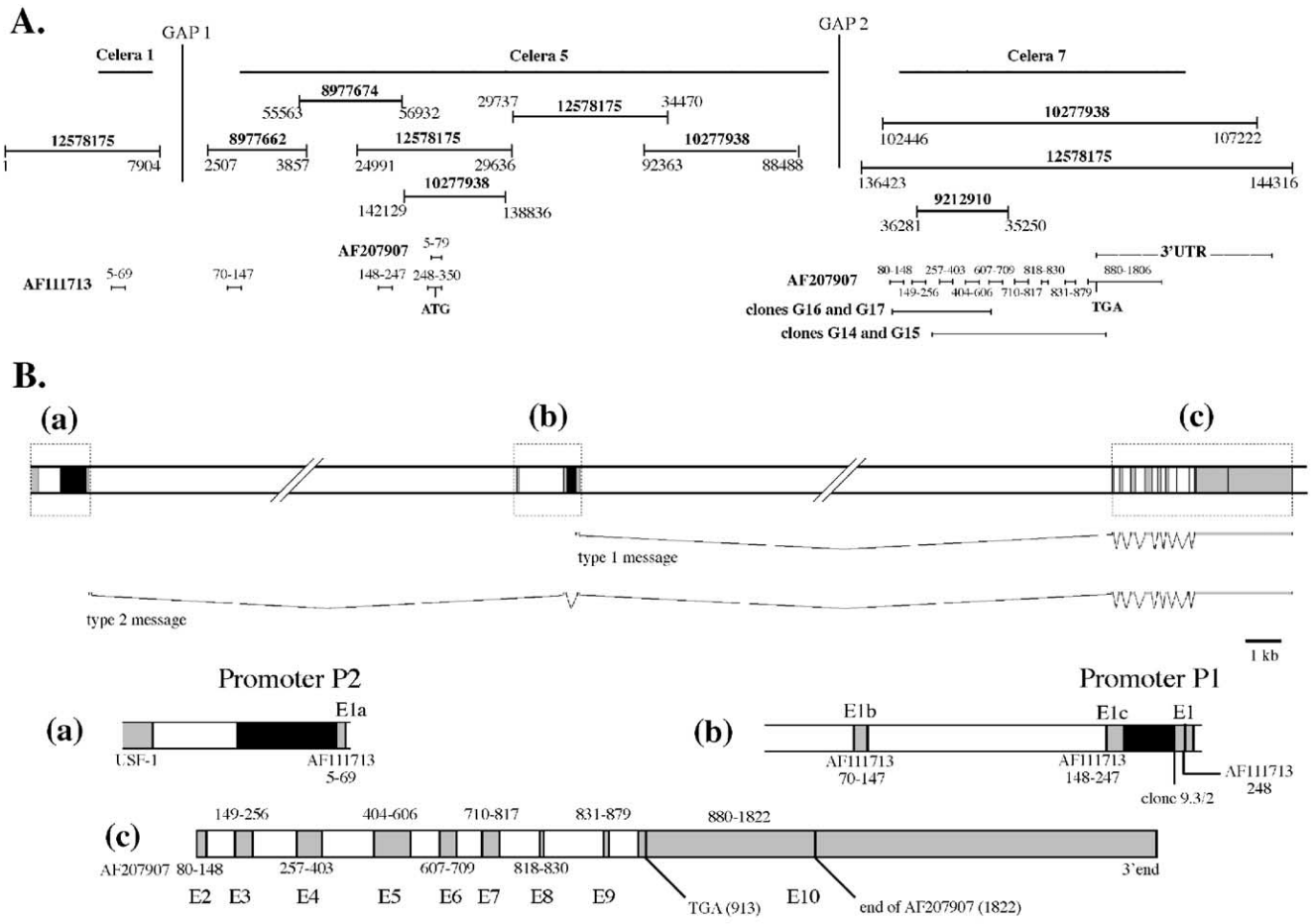


Fig. 5. Assembly and organization of the F11R gene sequence. Panel (A): Summary of the assembly of the F11R gene performed by the alignment of selected contigs from clones from the human genome unannotated database. The number above each contig is the GI number. The number below is a contig number. The parts of the cDNAs *AF207907* and *AF111713* which correspond to the exons in the F11R gene and which are found within selected contigs are shown below the contigs. Panel (B): Summary of the structural features of the F11R gene. The exons are in grey and introns in white. The two putative promoters P1 and P2 are in black. Transcripts of the type 1 and type 2 are outlined below. The squared regions (a), (b) and (c), are further enlarged to show details (three bottom panels). Exons are numbered according to *AF207907* or *AF111713*. The gaps in the sequence are denoted by '//'. E1a, E1b, E1c, E1–E10: exons 1a, 1b, 1c, 1–10.

BLAST searches using the F11R cDNA sequences allowed us to identify the additional exons and introns in the clone described above RP11-18E13 (*AL162255*, ver. 5, GI: 9212910), as well as RP11-381D2 (*AL162592*, ver. 12, GI: 12578175, AL162592, ver. 8, GI: 8977662) and clone RP11- 312J18 (*AL354714*, ver. 7, GI: 10277938).

We compared the genomic sequences with a proposed gene structure for F11R (ID: hCG20857) in Celera databases (Venter et al., 2001). The Celera F11R gene is composed of 7 fragments, among which fragments 1, 5 and 7 align with HGP contigs (Fig. 5). We also found that the *AF111713* cDNA sequence (bases 5–69) is identical to a sequence included in the 3′ end of the USF-1 gene (GI: 4586911; bases 9655 to 9719, Fukamizu A, published only in Database 1999). Moreover, alignment between the USF-1 gene and contig 1 of AL162592/GI: 12578175 revealed that bases 4982–9724 of USF-1 are 99.9% identical to bases 1664–6409 of GI: 12578175. This is in agreement with the information provided by the Human Genome Project showing that the USF-1 gene is located immediately upstream of the proposed F11R gene (as of May 2004) (IHGSC, 2001).

Based on the information in cDNA sequences and gDNA sequences, the contigs in the above described versions of clones RP11-381D2, RP11-18E13 and RP11-312J18 were rearranged to produce the structure of the F11R gene as outlined in Fig. 5A. In the alignment detailed there, there are two unclosed positions, GAP 1 and GAP 2. GAP 1 is located between GI: 12578175, contig 1 and GI: 8977662, contig 2, and GAP 2 is found within intron 1. PCR amplification of the F11R gene through intron 1 revealed a ≥20 kb product (not shown). Currently, sizes of these introns according to HGP (May 2004) are 16 and 19.7 kb.

The proposed exon/intron organization of the gene is outlined in Fig. 5B and detailed in Table 1. The major distinction between type 1 and type 2 mRNAs is found in their 5′ UTRs. In type 1 messages, the 5′ UTR is shorter and encoded by a single exon (E1), which also contains the translation start site. In type 2 mRNA, the 5′ UTR is encoded by at least 3 exons (E1a, E1c and part of E1). Exon 1 encodes the 5′ UTR and most of the signal sequence. Exon 2 encodes the remaining part of the signal sequence and the N-terminus of the mature F11R protein (strand A of the first Ig fold). Exons E3 and E4 encode most of the first Ig fold. The second Ig fold is encoded by two exons, E5 and E6. The transmembrane domain is encoded by exon 7, whereas three exons, E8–E10, encode the cytoplasmic tail.

3.7. Alternative polyadenylation signal usage results in varying length of the F11R mRNA 3′ UTRs

Northern blots for F11R mRNAs revealed several bands, with the largest of about 4 kb (Williams et al., 1999; Naik et al., 2001). There is also considerable discrepancy in the 3′ ends of the published cDNAs and ESTs. Two cDNAs, *AF207907* and *AF111713*, have 3′ UTRs of approximately the same length. The human lung cDNA clone FLJ23012 (accession number *AK026665*, Kawakami et al., unpublished) is identical to F11R cDNA (*AF207907*) except for an additional 47 bp at the 5′ UTR and a 3′ UTR longer by 394 bp. The longest 3′ UTRs was found in cDNA *AF172398.2* from stomach (Naik et

Table 1
Intron–exon organization of the F11R gene

Exon/intron	Sequence	Size (bp)	cDNA (AF207907)	Features
Exon 1a	GCCGAGA……GCTGGAG	65	5–69 [a]	5′ UTR
Intron 1a	**GT**ACCTG……TTTTG**AG**	>1995		
Exon 1b	ACGGAGT……CTCCCAG	78	70–147 [a]	5′ UTR
Intron 1b	**GT**ACAAG……TCCTT**AG**	1262		
Exon 1c	AGGTACT……GCTGAAG	100	148–247 [a]	5′ UTR
Intron 1c	**GT**GCGGA……TTCTC**AG**	275		
Exon 1	GAGTCCT……CTGTTGT	103/75	5–79	5′ UTR, leader
Intron 1	**GT**AAGTT……TCCTT**AG**	>10,591		
Exon 2	GCTCCCT……AATAATC	69	80–148	Leader, N-term/Ig-1
Intron 2	**GT**GAGTT……TCTGC**AG**	157		
Exon 3	CTGTGAA……ATCACAG	108	149–256	Ig-1
Intron 3	**GT**GAGTT……GTGGT**AG**	242		
Exon 4	CTTCCTA……GTGCTTG	147	257–403	Ig-1
Intron 4	**GT**ACGTG……TCCAC**AG**	282		
Exon 5	TGCCTCC……AGAGCTG	203	404–606	Ig-2
Intron 5	**GT**ATGTA……TCCTC**AG**	167		
Exon 6	GTCTTTG……GAAGCTG	103	607–709	Ig-2
Intron 6	**GT**GAGAG……CTCGC**AG**	128		
Exon 7	TGGAGCG……TTTGACA	108	710–817	TM
Intron 7	**GT**AAGTA……CTTTC**AG**	231		
Exon 8	GAACAAAGAAAGG	13	818–830	Cytoplasmic tail
Intron 8	**GT**GAGTG……TTCAC**AG**	304		
Exon 9	GACTTCG……AAGTGAA	49	831–879	Cytoplasmic tail
Intron 9	**GT**GAGTA……TCTGC**AG**	136		
Exon 10	GGAGAAT……	>1172/>2678 [b]	880–1805 [c]	Cytoplasmic tail, 3′ UTR

[a] These sequences are not present in *AF207907*. Therefore we included corresponding sequences found in *AF111713*.
[b] Two polyadenylation signals present and thus longer and shorter types of 3′ UTR's (Fig. 6).
[c] *AF207907* appears to contain an incomplete 3′ end.

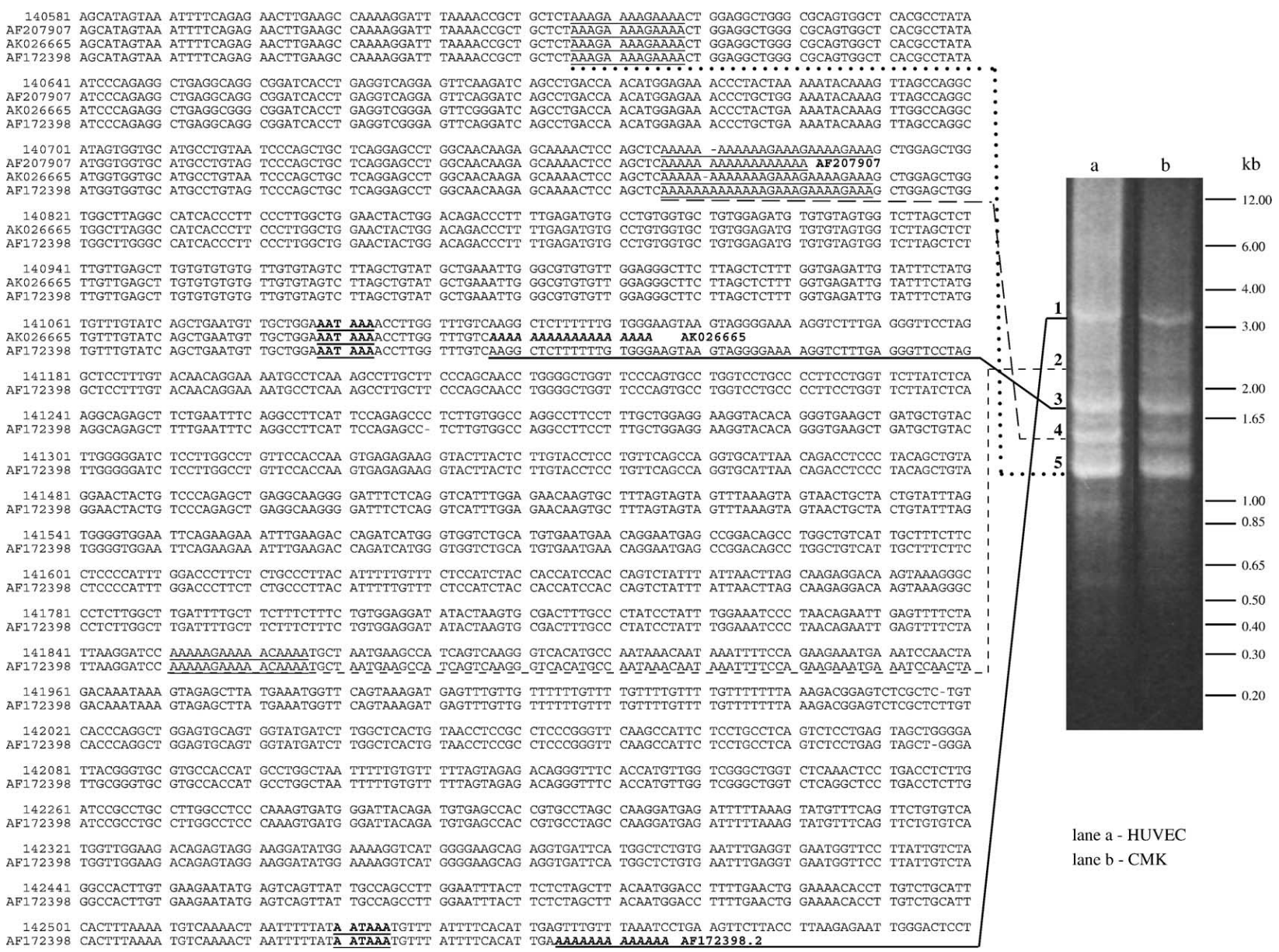


Fig. 6. 3′ RACE amplification of the F11R cDNA. Left Panel: Comparison of the 3′ end of the F11R gene and F11R cDNAs with varying 3′ ends. The top sequence, bases 140581 to 142680, is part of chromosome 1 (GI12578175, Fig. 5A). The sequences aligned to it represent parts of the F11R cDNAs: *AF207907*, *AK026665* and *AF172398.2*. PolyA stretches, which are part of the gene are underlined. Polyadenylation signals are underlined and in bold. PolyA stretches added posttranscriptionally are in bold italics. Right Panel: 3′ RACE amplification of the F11R message using F5 and GeneRacer™ 3′ reverse revealed five products in HUVEC and CMK cells (lanes a and b). The sizes of the bands were identical and estimated to be 3.3, 2.4, 1.8, 1.4 and 1.2 kb (numbered 1–5). The five lines originating from the bands visible in the gel and pointing to the sequence indicate the beginning of the A-rich stretches or polyA tails added posttranscriptionally for each PCR product obtained.

al., 2001). However the first 85 bases of AF172398.2 show no similarity to any sequences found in the human genome as determined by BLAST Search against human genome or by HGP browser of April 2003.

We utilized 3′ RACE to determine the full length of the F11R message at the 3′ end. Amplification using primers F5 (Fig. 1B) and GeneRacer™ 3′ reverse revealed five products in HUVEC and CMK cells (Fig. 6, lanes a and b, respectively). The sizes of

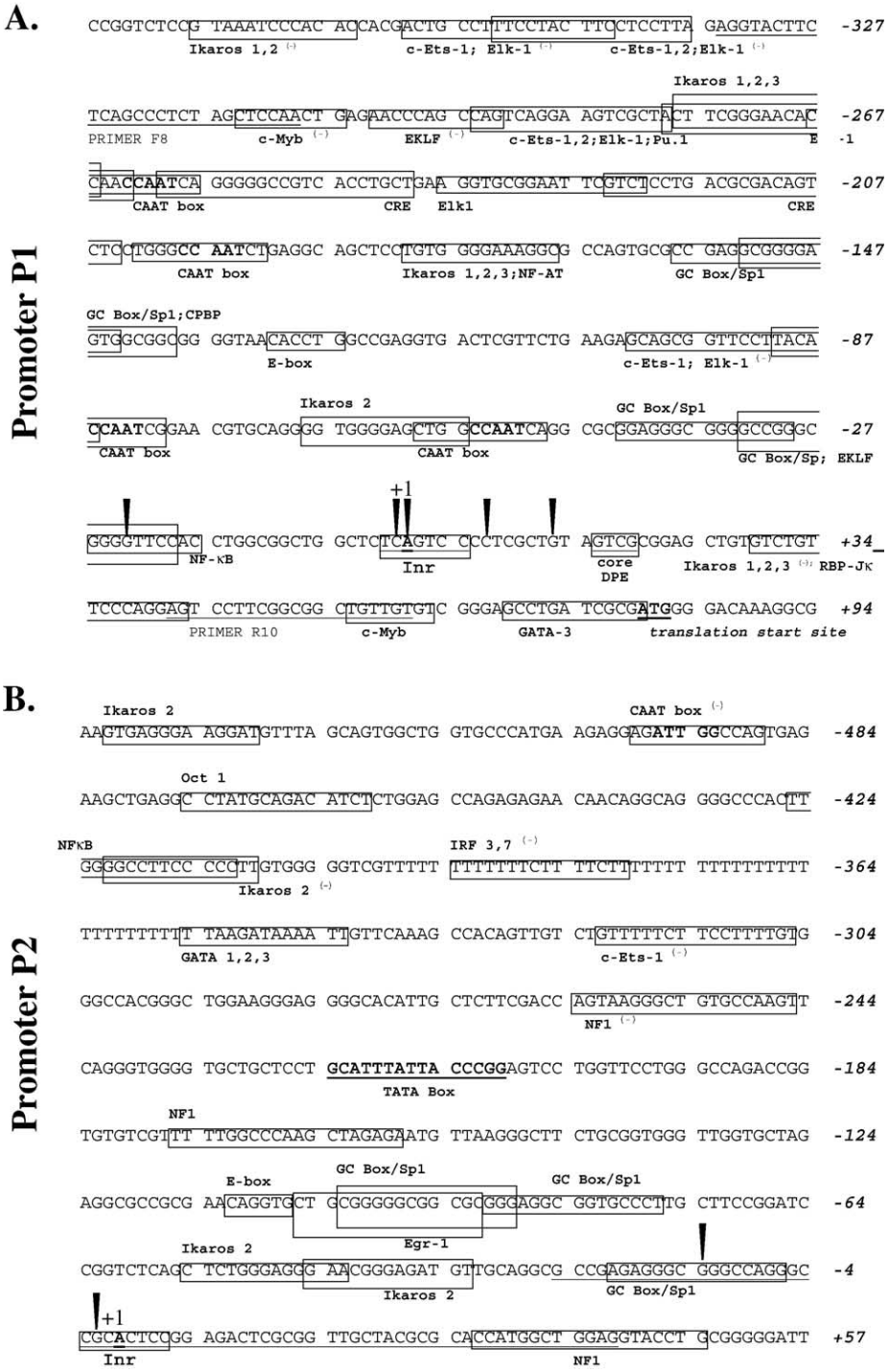


Fig. 7. Analysis of the Promoters P1 and P2. Panel (A): Nucleotide sequence of the promoter P1. The sequence corresponds to bases 28081–25560 of GI12578175 (Fig. 5A). Included are the sequences of primers (F8 and R10) used for PCR amplification of the promoter. Panel (B): Nucleotide sequence of the Promoter P2 and the first exon the F11R gene. The sequence corresponds to bases 5821–6420 of GI12578175 (Fig. 5A). The TATA box at position −223/−209 is in bold and underlined. Underlined (at bases −24/+41) is the region corresponding to the first exon and bases 5–69 of AF111713. Panels (A) and (B): Putative transcription factor binding sites are boxed. Multiple transcription start sites identified by 5′ RACE are denoted by arrowheads. The numbering on the right (bold, italics) is relative to the transcription start site which is defined as base ‘A’ (in bold) in the initiator (Inr).

the bands were identical in both cell types and estimated as 3.3, 2.4, 1.8, 1.4 and 1.2 kb (numbered 1–5, respectively). Two putative polyadenylation AATAAA signals (Fig. 6, underlined and bold) and three A-rich stretches (underlined) were found in the corresponding genomic sequence. This genomic sequence is enriched in repetitive elements including Alu Sp, Alu Sg, Alu Sc, L1MA9, MIR, Alu Sq/x and L1P. The presence of the five 3′ RACE products can be explained as follows: (a) the largest 3.2–3.3 kb band resulted from amplification of the RT product from the full length message, (b) the 1.8 kb band represents a variant of the F11R mRNA with a shorter 3′ end, (c) the 1.2, 1.4 and 2.2 kb bands may have been amplified from an RT product in which the oligo dT primer bound to the internal polyA stretch. Accordingly, we have estimated that full length mRNAs with long 3′ UTRs (either type 1 or 2) are about 3.8–3.9 kb in size, whereas the shorter 3′ UTRs are 2.3–2.4 kb, sizes similar to those identified by Northern blotting. Type 1 mRNAs can have longer or shorter 3′ UTRs.

3.8. Potential transcription factor binding sites in promoters P1 and P2

The 5′ RACE results provide evidence for the existence of two promoters, P1 and P2, regulating transcription of type 1

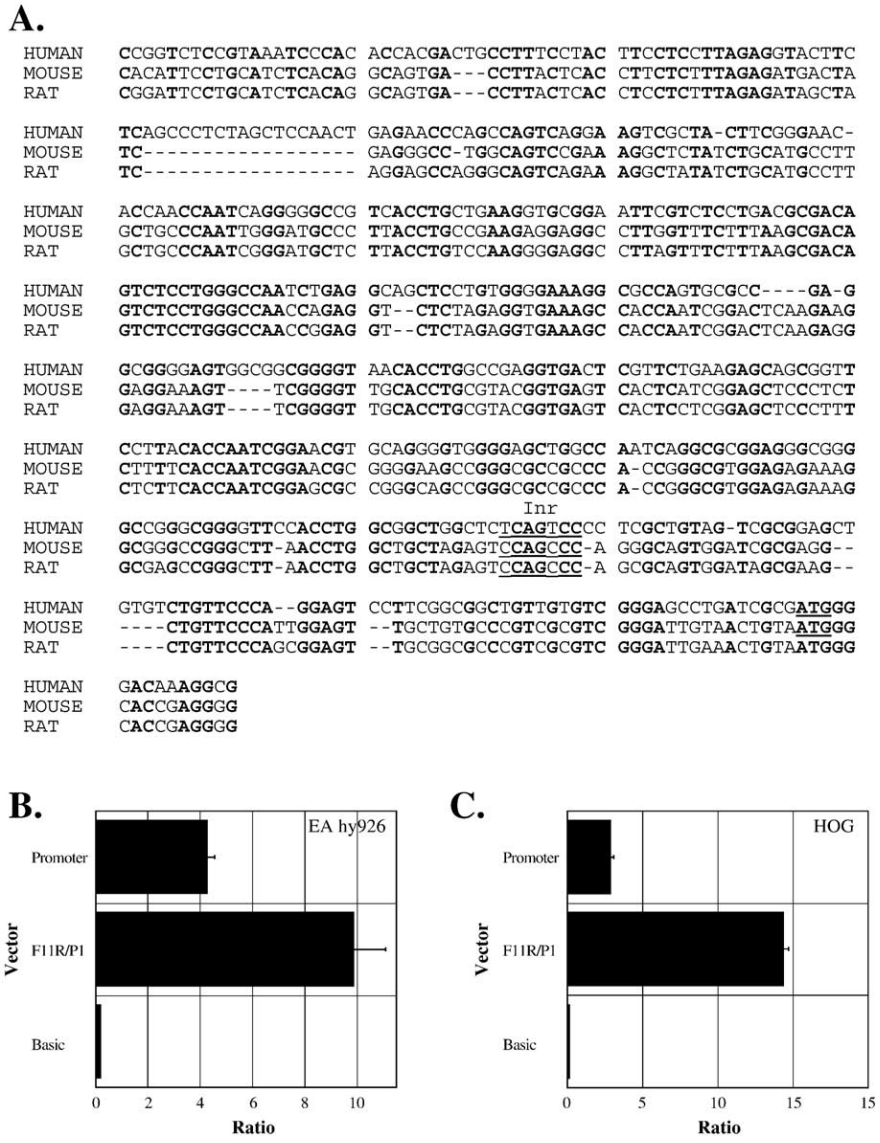


Fig. 8. Activity of the Promoter P1 and comparison of the Promoter P1 sequences in human, mouse and rat. Panel (A): Comparative analysis of the human, mouse and rat Promoter P1 sequences. Sequences homologous to P1 were found on the mouse chromosome 1 bases 171508437–171508887 (genome browser on mouse update of May 2004) and on the rat chromosome 13 bases 87558437–87558888 (genome browser on rat update of Jun 2003). Translation start sites (ATG) of human and mouse F11R are underlined. The similarity between species is as follows: human vs. rat ~57%, human vs. mouse ~56%, rat vs. mouse ~92% (258 identities). Panels (B) and (C): Activity of the F11R promoter P1 (F11R/P1) in human endothelial (EA.hy926) and HOG cells. Luciferase activities of the P1 promoter constructs in pGL3Basic and pRL-TK were measured in both cell lines as described in Section 2.9. 'Basic' is a basal activity of the vector. The data are compared to the activity of the pGL3-Promoter vector (a positive control). Assays were performed using a dual luciferase assay system and results are presented as a ratio of firefly to *Renilla* luminescence.

and type 2 mRNAs of F11R, respectively. Analysis for the potential transcription factor binding sites revealed numerous sites located within promoter regions P1 and P2. As shown in Fig. 7A, P1 was identified as a TATA-less promoter containing an initiator element (Inr). The bolded letter 'A' within the initiator element has been designated as a transcription start site (+1). The results of 5′ RACE indicated multiple transcription start sites commonly found in TATA-less promoters. Three 5′ RACE clones were initiated at position −23, one clone at −1, two clones at +1, one at +7 and three at position +13. Within promoter P1, sequence 'GTCG' at position +16 matches the core DPE consensus sequence G(A/T)CG with sequence requirements for the extended consensus sequence (A/G)G(A/T)CGTG almost perfectly met (Bucher, 1990; Burke et al., 1998). Four GC boxes were identified in promoter P1 at positions: −158/−144, −153/−139, −43/−29 and −33/−19. In addition to GC boxes, another promoterproximal element frequently found in many promoters is a CAAT box. We have identified four CCAAT consensus sequences at positions: −267/−257, −202/−192, −90/−80 and −59/−49. In addition, analysis of the promoter sequence revealed numerous potential transcription factor binding sites for GATA (+69/+80), NF-κB (−26/−17), c-Myb (−314/−306 and +56/+63), ets family transcription factors (−360/−340, −353/−337, −295/−279, −279/−263, −237/−221 and −102/−86) and for Ikaros (−377/−365, −278/−266, −180/−168, −68/−56 and +29/+43). Furthermore, two CRE sites (−260/−240 and −228/−204), one NF-AT site (−180/−168), one RBP-Jκ site (+29/+43), one EKLF site (−303/−293) and one E-box (−131/−126) were also localized in Promoter P1.

Promoter P2 (Fig. 7, Panel B) contains an initiator element, similar to Promoter P1, and the bold letter 'A' within the initiator sequence has been designated as a transcription start site (+1). Two major transcription start sites were identified by 5′ RACE, indicated by the arrowheads at positions −13 and −2. P2 does not contain a TATA box at an expected location, but we have identified a TATA box-like sequence about 200 bases from both the Inr element. Promoter P2 was found to contain three GC boxes at positions −102/−89, −90/−76 and −20/−6, four transcription factor Ikaros binding sites at −541/−529, −421/−409, −54/−42 and −44/−32, three NF-1 sites at −263/−245, −175/−157 and +30/+48, and binding sites for GATA at −354/−342, NF-κB at −425/−411, interferon regulatory factors IRF 3 and 7 (−393/−379), Oct-1 (−474/−460) and Egr-1 (−105/−91). In addition, one CCAAT box in reverse ATTGG orientation was located at position −496/−492.

A comparison of the human P1 sequence with the corresponding mouse and rat sequences is shown in Fig. 8A. Highly conserved regions were found within the promoter, which included the location of two CAAT boxes and the initiator sequence.

3.9. Transcriptional activity of promoter P1

To obtain experimental evidence for the transcriptional activity of the predicted promoter P1, we selected a region in the genomic DNA upstream to and containing the transcriptional start sites, as defined by the results obtained from 5′ RACE (Fig. 7). As shown in Fig. 8, Panels B and C, the F11R/P1 showed strong transcriptional activity in both human endothelial cells (EA.hy926) and human glioma HOG cells, respectively.

3.10. Identification of upstream ORFs in the type 2 F11R mRNAs

Although type 1 and type 2 messages encode identical proteins, the efficiency of their translation may be different. In addition to the known translation start site, the type 2 mRNAs contain the strong upstream AUG codon and hence two open reading frames: the upstream ORF and the main ORF (Fig. 9). Because the presence of the strong upstream AUG generally inhibits the translation of the main ORF (Kozak, 2002), the possibility exists that the type 2 mRNAs may not be translated or that their translation is under tight control.

4. Discussion

The structure of the F11R gene was assembled by combining the experimental data obtained in our laboratories and reported here with information obtained from available databases. Each of the two Ig folds in the F11R protein was found to be encoded predominantly by 2 exons, the first by bases 149–256 and 257–403, and the second by bases 404–606 and 607–709 of the F11R cDNA (*AF207907*). The use of two or more exons per Ig fold has been shown to occur in

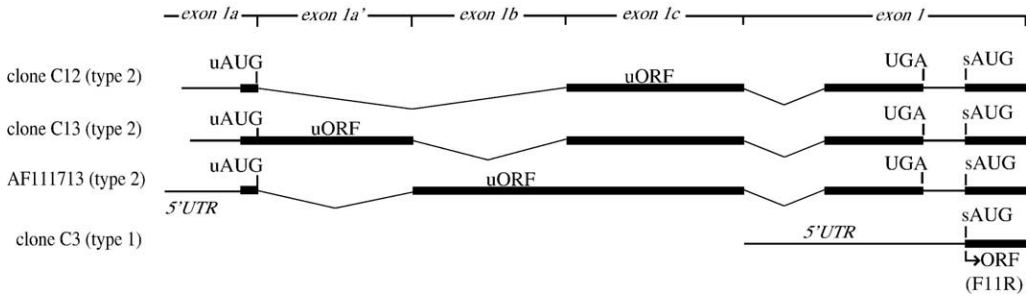


Fig. 9. Identification of upstream ORFs in the type 2 F11R mRNAs. Schematic diagrams of type 2 and type 1 F11R mRNAs. Shown are: uAUG, upstream translation start codon; uORF, upstream ORF; UGA, stop codon; ORF (F11R), the main ORF encoding the F11R and sAUG, translation start codon. Above diagrams are indicated corresponding exons in the F11R gene. Sequence corresponding to E1a′ was found only in clone C13, therefore this exon is not included in Table 1.

immunoglobulin superfamily members, such as the neural cell recognition molecule F11 (Plagge and Brummendorf, 1997) and the melanoma-associated glycoprotein, MUC18 (Sers et al., 1993). However, in general, each Ig fold of most Ig superfamily members is encoded by a single exon, including PECAM-1 (Newman et al., 1990), ICAM-2 (Voraberger et al., 1991) and glycoprotein VI (Ezumi et al., 2000).

Similar to F11R, genes for several other CAMs (see Plagge and Brummendorf, 1997) have one or more long introns at the 5′ end. It has been proposed that the large introns in these CAM genes contain regulatory enhancer- and silencer-like elements, and such appear to be the case for introns 1a and 1 of the F11R gene.

Our results of 5′ RACE combined with database sequences provide evidence for F11R mRNAs with alternate 5′ UTRs and two promoters, P1 and P2. Type 1 mRNA is the major transcript produced from the F11R gene by the P1 promoter in all of the cells studied. The type 2 mRNA is present at levels significantly lower than type 1, and found only in endothelial cells. Type 2 mRNAs have published sequences (Ozaki et al., 1999; Gupta et al., 2000). All of the type 2 mRNAs share the 5′ end. However, the internal sequences of their 5′ UTRs show differences (Fig. 3C). These variations can be explained by alternative splicing of the type 2 messages. At present, the mechanism by which the expression of type 2 F11R mRNAs is regulated as well as its significance is unknown, however it may be related to developmental regulation. Three alternative promoters and 11 splice variants of mRNA, which are developmentally regulated, have been found for the Ig superfamily member F3/contactin (De Benedictis et al., 2001).

Results obtained with 3′ RACE combined with information on the F11R gene structure provided an explanation for the existence of mRNAs differing at their 3′ ends. Only two cDNA's for F11R appear to be derived from 3′ full length messages, *AK026665* (Kawakami et al., unpublished) and *AF172398.2*. Together with 5′ RACE results, the predicted size of either type 1 or 2 with shorter or longer 3′ UTRs are in agreement with Northern blot analyses (Williams et al., 1999; Naik et al., 2001) which detected two major messages of 2–2.5 and ~4 kb. In contrast, cDNAs *AF207907* and *AF111713* appear to have been reverse-transcribed by the binding of oligo dT primer to mRNA at the internal polyA stretch.

Analysis of type 1 and type 2 F11R mRNAs indicated the presence of two promoters P1 and P2. The P2 promoter contains both a TATA box and an Inr element however the TATA box is located at a distance of 208 bases away from both the Inr element and the transcription start sites. A similar distant location was noted in the promoter for platelet glycoprotein V in mouse and rat (Ravanat et al., 1997).

The P1 promoter is a typical TATA-less, Inr containing promoter. The results of 5′ RACE indicate multiple transcription start sites, which have also been reported in TATA-less promoters of other members of IgSF such as PECAM-1 and ICAM-2, as well as promoters of genes other than IgCAMs. Within P1, four GC boxes were identified which might be binding sites for Sp1 or Sp1-related transcription factors depending on the cell type. Single or multiple GC boxes have been found in many TATA-less promoters and the presence of an Inr-containing promoter was demonstrated to compensate for the lack of a TATA box. Similarly, mutations in GC boxes in a TATA-less ICAM-2 promoter decreased the activity of the promoter (Cowan et al., 1998).

The promoter P1 contains a downstream core promoter element (DPE), which is not present in mouse P1 (Figs. 7A and 8A) but has been proposed to be functionally analogous to a TATA box. The DPE is usually located at about +30 (typically from +28 to +34) relative to the transcription start site and the spacing between the Inr and DPE was reported to be important for transcription and TFIID binding, since any changes in spacing between the DPE and Inr result in a decrease in transcriptional activity. Because the DPE in P1 is located closer to major transcription start sites, it is difficult to hypothesize at this time whether the DPE plays any role in the function of the P1 of F11R.

Three transcription factor binding sites, GATA, ets and NF-κB may be of particular interest in the regulation of F11R. GATA transcription factors have been identified in megakaryocytic lineages, multipotential progenitor cell lines and endothelial cells. A critical role for GATA-1 in megakaryopoiesis has been demonstrated. GATA-1 deficiency in megakaryocytes and platelets resulted in thrombocytopenia with an increased number of megakaryocytes characterized by marked ultrastructural abnormalities (Vyas et al., 1999). GATA binding elements were predicted in the TATA-less promoter for PECAM-1 (Almendro et al., 1996; Gumina et al., 1997). An electrophoretic mobility shift assay suggested the binding of GATA-2 protein to the GATA element identified at position −24 of PECAM-1, which is similar to the location of the predicted GATA-1 at position −43 in P1 (Gumina et al., 1997). Aird et al. (1994) have demonstrated that similarly situated (with respect to the distance from the TSS), the GATA motif served as a weak surrogate for the binding of TFIID in the TATA less promoters in rat platelet factor IV, mouse erythropoietin and chicken β globin genes. Furthermore, GATA-1, Sp1 and Ets-binding sites in the promoter of GPVI, a collagen receptor, were found to be critical for the expression of GPVI in megakaryocytes (Holmes et al., 2002; Furihata and Kunicki, 2002). And indeed, such a situation may regulate the expression of the F11R.

Ets transcription factor family members are expressed in endothelial cells during angiogenesis (Lelievre et al., 2001). Expression of Ets-1, one of the members of the family, is upregulated in endothelial cells in response to angiogenic factors such as TNF-α, FGF 2 and TGF β. The F11R has been reported to play a key role in FGF-2 signaling and the present finding of Ets in F11R gene may be related to its proposed role in angiogenesis (Naik et al., 2003).

Two potential NF-κB binding sites were identified in the PECAM-1 promoter with its regulation by the cytokine TNF-α (Gumina et al., 1997; Botella et al., 2000). The TNF-α-response element was also found in the ICAM-2 promoter (McLaughlin et al., 1999). Expression of F11R in HUVEC

cells was downregulated by single cytokine treatment, although upregulation of F11R occurred following a combined cytokine treatment (Gupta et al., 2000). We have found that the adhesion of platelets to cytokine-treated HUVEC was mediated via F11R, indicating an important role for F11R in the initial steps leading to inflammatory thrombosis (Babinska et al., 2002b). Further studies designed to determine the role of NF-κB and other transcription factor binding sites present in the promoters of the F11R gene both in normal physiology and in the pathophysiological development of inflammatory thrombosis and atherosclerosis, are currently in progress.

References

Aird, W.C., Parvin, J.D., Sharp, P.A., Rosenberg, R.D., 1994. The interaction of GATA-binding proteins and basal transcription factors with GATA box-containing core promoters. A model of tissue-specific gene expression. J. Biol. Chem. 269, 883–889.

Almendro, N., et al., 1996. Cloning of the human endothelial cell adhesion molecule-1 promoter and its tissue-specific expression. Structural and functional characterization. J. Immunol. 157, 5411–5421.

Babinska, A., et al., 2002a. Two regions of human platelet receptor F11R critical for platelet aggregation, potentiation and adhesion. Thromb. Haemost. 87, 712–721.

Babinska, A., et al., 2002b. F11-receptor (F11R/JAM) mediates platelet adhesion to endothelial cells: role in inflammatory thrombosis. Thromb. Haemost. 88, 843–850.

Barton, E.S., et al., 2001. Junction adhesion molecule is a receptor for reovirus. Cell 104, 441–451.

Bazzoni, G., Martinez-Estrada, O.M., Orsenigo, F., Cordenonsi, M., Citi, S., Dejana, E., 2000. Interaction of junctional adhesion molecule with the tight junction components ZO-1, cingulin, and occludin. J. Biol. Chem. 275, 20520–20526.

Botella, L.M., et al., 2000. Identification of a functional NF-κB site in the platelet endothelial cell adhesion molecule-1 promoter. J. Immunol. 164, 1372–1378.

Bucher, P., 1990. Weight matrix description of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J. Mol. Biol. 212, 563–578.

Burke, T.W., Willy, P.J., Kutach, A.K., Butler, J.E.F., Kadonaga, J.T., 1998. The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harbor Symposia on Quantitative Biology, volume LXIII. Cold Spring Harbor Laboratory Press, pp. 75–82.

Cowan, P.J., et al., 1998. The human ICAM-2 promoter is endothelial cell-specific in vitro and in vivo and contains critical Sp1 and GATA binding sites. J. Biol. Chem. 273, 11737–11744.

De Benedictis, L., et al., 2001. Alternative promoters drive the expression of the gene encoding the mouse axonal glycoprotein F3/contactin. Mol. Brain Res. 95, 55–74.

Ebnet, K., Schulz, C.U., Meyer zu Brickwedde, M.K., Pendl, G.G., Vestveber, D., 2000. Junctional adhesion molecule interacts with the PDZ domain-containing proteins AF-6 and ZO-1. J. Biol. Chem. 275, 27979–27988.

Ezumi, Y., Uchiyama, T., Takayama, H., 2000. Molecular cloning, genomic structure, chromosomal localization, and alternative splice forms of the platelet collagen receptor glycoprotein VI. Biochem. Biophys. Res. Commun. 277, 27–36.

Furihata, K., Kunicki, T.J., 2002. Characterization of human glycoprotein VI gene 5′ regulatory and promoter regions. Arterioscler. Thromb. Vasc. Biol. 22, 1733–1739.

Gumina, R.J., Kirschbaum, N.E., Piotrowski, K., Newman, P.J., 1997. Characterization of the human platelet/endothelial cell adhesion molecule 1 promoter: identification of a GATA2 binding element required for optimal transcriptional activity. Blood 89, 1260–1269.

Gupta, S.K., Pillarisetti, K., Ohlstein, E.H., 2000. Platelet agonist F11 receptor is a member of the immunoglobulin superfamily and identical with junctional adhesion molecule (JAM): regulation of expression in human endothelial cells and macrophages. IUBMB Life 50, 51–56.

Holmes, M.L., Bartle, N., Eisbacher, M., Chong, B.H., 2002. Cloning and analysis of the thrombopoietin-induced megakaryocyte-specific glycoprotein VI promoter and its regulation by GATA-1, Fli-1 and SP1. J. Biol. Chem. 277, 48333–48341.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Kornecki, E., Walkowiak, B., Naik, U.P., Ehrlich, Y.H., 1990. Activation of human platelets by a stimulatory monoclonal antibody. J. Biol. Chem. 265, 10042–10048.

Kozak, M., 2002. Pushing the limits of the scanning mechanism for initiation of translation. Gene 299, 1–34.

Lelievre, E., Lionneton, F., Soncin, F., Vandenbunder, B., 2001. The ets family contains transcriptional activators and repressors involved in angiogenesis. Int. J. Biochem. Cell Biol. 33, 391–407.

Liang, T.W., et al., 2000. Characterization of huJAM: evidence for involvement in cell–cell contact and tight junction regulation. Am. J. Physiol., Cell Physiol. 279, C1733–C1743.

Liu, Y., et al., 2000. Human junction adhesion molecule regulates tight junction resealing in epithelia. J. Cell Sci. 113, 2363–2374.

Martin-Padura, I., et al., 1998. Junctional adhesion molecule, a novel member of the immunoglobulin superfamily that distributes at intercellular junctions and modulates monocyte transmigration. J. Cell Biol. 142, 117–127.

Martinez-Estrada, O.M., Villa, A., Breviario, F., Orsenigo, F., Dejana, E., Bazzoni, G., 2001. Association of junctional adhesion molecule with calcium/calmodulin-dependent serine protein kinase (CASK/LIN-2) in human epithelial Caco-2 cells. J. Biol. Chem. 276, 9291–9296.

McLaughlin, F., Ludbrook, V.J., Kola, I., Campbell, C.J., Randi, A.M., 1999. Characterization of the tumor necrosis factor (TNF)-α response element in the human ICAM-2 promoter. J. Cell Sci. 112, 4695–4703.

Naik, U.P., Ehrlich, Y.H., Kornecki, E., 1995. Mechanisms of platelet activation by a stimulatory monoclonal antibody: cross-linking of a novel platelet receptor for M.Ab.F11 with the FcγRII receptor. Biochem. J. 310, 155–162.

Naik, U.P., Naik, M.U., Eckfeld, K., Martin-DeLeon, P., Spychala, J., 2001. Characterization and chromosomal localization of JAM-1, a platelet receptor for a stimulatory monoclonal antibody. J. Cell Sci. 114, 539–547.

Naik, M.U., Mousa, S.A., Parkos, C.A., Naik, U.P., 2003. Signaling through JAM-1 and v3 is required for the angiogenic action of FGF: dissociation of the JAM-1 and v3 complex. Blood 102, 2108–2114.

Newman, P.J., et al., 1990. PECAM-1 (CD31) cloning and relation to adhesion molecules of the immunoglobulin gene superfamily. Science 247, 1219–1222.

Ozaki, H., et al., 1999. Cutting edge: combined treatment of TNF-α and IFN-γ causes redistribution of junctional adhesion molecule in human endothelial cells. J. Immunol. 163, 553–557.

Plagge, A., Brummendorf, T., 1997. The gene of the neural cell adhesion molecule F11: conserved exon–intron arrangement in genes of neural members of the immunoglobulin superfamily. Gene 192, 215–225.

Quandt, K., Frech, K., Karas, H., Wingender, E., Werner, E., 1995. MatInd and MatInspector: a new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 23, 4878–4884.

Ravanat, C., et al., 1997. Gene cloning of rat and mouse platelet glycoprotein V: identification of megakaryocyte-specific promoters and demonstration of functional thrombin cleavage. Blood 89, 3253–3262.

Sers, C., Kirsch, K., Rothbacher, U., Riethmuller, G., Johnson, J.P., 1993. Genomic organization of the melanoma-associated glycoprotein MUC18: implications for the evolution of the immunoglobulin domains. Proc. Natl. Acad. Sci. U. S. A. 90, 8514–8518.

Sobocka, M.B., et al., 2000. Cloning of the human platelet F11 receptor: a cell adhesion molecule member of the Ig superfamily involved in platelet aggregation. Blood 95, 2600–2609.

Sobocka, M.B., et al., 2004. Signaling pathways of the F11 receptor (F11R; a.k.a. JAM-1, JAM-A) in human platelets: F11R dimerization, phosphorylation and complex formation with the integrin GPIIIa. J. Recept. Signal Transduct. Res. 24, 85–105.

Venter, J.C., et al., 2001. The sequence of the human genome. Science 291, 1304–1350.

Voraberger, G., Schaffer, R., Stratowa, C., 1991. Cloning of the human gene for the intercellular adhesion molecule-1 and analysis of its 5′-regulatory region. J. Immunol. 147, 2777–2786.

Vyas, P., Ault, K., Jackson, C.W., Orkin, S.H., Shivdasani, R.A., 1999. Consequences of GATA-1 deficiency in megakaryocytes and platelets. Blood 93, 2867–2875.

Wang, F., Naik, U.P., Ehrlich, Y.H., Osada, S.-I., Kornecki, E., 1995. Stimulatory antibody-induced activation and selective translocation of protein kinase C isoenzymes in human platelets. Biochem. J. 311, 401–406.

Wenzel, K., et al., 2003. Identification and characterization of KAT, a novel gene preferentially expressed in several human cancer cell lines. Biol. Chem. 384, 763–775.

Williams, L.A., Martin-Padura, I., Dejana, E., Hogg, H., Simmons, D.L., 1999. Identification and characterization of human junctional adhesion molecule (JAM). Mol. Immunol. 36, 1175–1188.