Structural Organization and Chromosomal Assignment of the Gene Encoding the Human Heparin-Binding Epidermal Growth Factor-like Growth Factor/Diphtheria Toxin Receptor†

Zhou Fen,‡ Mandeep S. Dhadly,§ Masao Yoshizumi,‡,∥ Robert J. Hilkert,§ Thomas Quertermous,⊥ Roger L. Eddy,# Thomas B. Shows,# and Mu-En Lee*,‡,∥,○

Cardiovascular Biology Laboratory, Harvard School of Public Health, Cardiovascular Division, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, Massachusetts General Hospital, Boston, Massachusetts 02114, Division of Cardiology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, and Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14163

Received February 11, 1993; Revised Manuscript Received May 6, 1993

ABSTRACT: Heparin-binding epidermal growth factor-like growth factor (HB-EGF) is a recently identified, potent smooth muscle cell mitogen of macrophage origin. It is expressed in a highly regulated fashion in vascular endothelial and smooth muscle cells, indicating a potentially important role for this gene in atherosclerosis. In addition, the HB-EGF precursor has recently been found to function as a receptor for diphtheria toxin. Using an HB-EGF cDNA probe, we cloned the human gene encoding HB-EGF. The HB-EGF gene contains six exons and five intervening sequences spanning 14 kb of DNA. By primer extension and S1 nuclease analysis, we located a major transcription start site (corresponding to an A residue) 14 bp beyond the 5′end of the HB-EGF cDNA. There were no TATAAA or CCAAT consensus sequences upstream of the transcription start site. The density of primer extension bands generated by RNA from endothelial cells treated with tumor necrosis factor-α (TNF-α) was 10 times higher than that of bands generated by the control, indicating that TNF-α increased the level of HB-EGF mRNA. Using transient reporter gene transfection experiments, we show that 2.0 kb of HB-EGF 5′-flanking sequence has promoter activity in bovine aortic endothelial cells. By analysis of DNA isolated from human-mouse somatic hybrid cell lines, we assign the HB-EGF gene to chromosome 5. By functional study, chromosome 5 has been associated with diphtheria toxin susceptibility.

Heparin-binding epidermal growth factor-like growth factor (HB-EGF) is a 22-kilodalton protein originally purified from the conditioned medium of macrophage-like U-937 cells (Higashiyama et al., 1991). The mature peptide contains an epidermal growth factor (EGF) domain characterized by three disulfide loops and is considered a member of the EGF family (Carpenter & Cohen, 1990; Carpenter & Wahl, 1990). In addition to the EGF domain, prepro-HB-EGF contains a signal peptide, a propeptide, a putative heparin-binding domain, a hydrophobic transmembrane domain, and a cytoplasmic domain (Higashiyama et al., 1992). HB-EGF binds EGF receptors and is mitogenic for fibroblasts and vascular smooth muscle cells but not for endothelial cells (Higashiyama et al., 1991). Although the mechanism is unknown, HB-EGF is a more potent smooth muscle cell mitogen than EGF. The potency of HB-EGF is similar to that of platelet-derived growth factor (Higashiyama et al., 1991).

We have demonstrated that HB-EGF is transcribed in human vascular endothelial cells (Yoshizumi et al., 1992), rat aortic smooth muscle cells (Temizer et al., 1992), and MOLT-4 cells (a T cell line) but not in other cells such as JEG-3, A-172, or HEL 92.1.7 (choriocarcinoma, glioblastoma, and erythroleukemia lines, respectively) (M. Yoshizumi and M.-E. Lee, unpublished results). Furthermore, expression of HB-EGF is highly regulated by cytokines in vascular endothelial cells (Yoshizumi et al., 1992) and by phorbol esters and angiotensin II in rat aortic smooth muscle cells (Temizer et al., 1992). Thus, HB-EGF, which was originally purified from macrophages, also appears to be transcribed and regulated in cultured cells commonly associated with atherosclerotic lesions (Libby & Hansson, 1991; Munro & Cotran, 1988; Ross, 1986; Schwartz et al., 1990). The fact that these cells transcribe the gene for this potent smooth muscle mitogen suggests that HB-EGF may have an important role in the proliferation of smooth muscle cells during atherogenesis (Yoshizumi et al., 1992).

The cDNA encoding the diphtheria toxin receptor was cloned recently (Naglich et al., 1992), and the integral membrane protein it was predicted to encode is identical to the HB-EGF precursor. Thus, HB-EGF not only appears to function as a potent mitogen, but also serves as the receptor for a bacterial toxin. Although diphtheria toxin inhibits protein synthesis in eukaryotic cells (Eidels et al., 1983), not all cells are equally susceptible to it. Several lines of evidence suggest that differences in susceptibility are due to the number of functional cell surface receptors, which are essential for the toxin's entry into cells. Knowledge about the regulation of HB-EGF gene expression is fundamental to determining the differential sensitivity of cells to diphtheria toxin, as well as the role of this peptide in vascular disease.

† Supported by a grant from the Bristol-Myers Squibb Pharmaceutical Research Institute. This work was done during the tenure of a Clinician Scientist Award and Grant-in-Aid (both to M.-E.L.) from the American Heart Association (with funds contributed in part by the AHA, Massachusetts Affiliate).

* To whom correspondence should be addressed at Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115. Telephone: (617) 432-4994. Fax: (617) 432-4098.

‡ Harvard School of Public Health.

§ Massachusetts General Hospital.

∥ Harvard Medical School.

⊥ Vanderbilt University School of Medicine.

Roswell Park Memorial Institute.

○ Brigham and Women's Hospital.

0006-2960/93/0432-7932$04.00/0

In studying the cellular mechanisms regulating expression of the HB-EGF gene, it is important to understand its structural and promoter organization. To our knowledge, this is the first report of the isolation of the HB-EGF gene and the characterization of its genomic organization. We also delineated its transcription start site by primer extension and S1 nuclease protection assay, demonstrated the promoter activity of its 5′-flanking sequences, and, because of the potential involvement of HB-EGF in human disease, determined the chromosomal localization of the gene by somatic cell hybridization.

EXPERIMENTAL PROCEDURES

Cloning the HB-EGF Gene. A 0.8-kilobase (kb) *Eco*RI–*Kpn*I restriction fragment prepared from a human HB-EGF cDNA (Yoshizumi et al., 1992) was used to screen a cosmid library of human placenta DNA in the vector pWE15 (Stratagene, La Jolla, CA). The fragment was radiolabeled with DNA polymerase 1 in the presence of random-sequence hexanucleotides and [α-^{32}P]dCTP (Sambrook et al., 1989). The cosmid library was plated, and nitrocellulose filters were prepared as described (Sambrook et al., 1989). The filters were hybridized at 42 °C with the radiolabeled probe and washed as described at 55 °C. Hybridizing cosmid clones were then purified, and DNA was prepared according to standard procedures (Sambrook et al., 1989).

Sequencing the HB-EGF Gene. Restriction fragments derived from the cosmid clone containing the HB-EGF gene were ligated into the plasmid vector pBS (Stratagene) or pUC 18 (Pharmacia LKB Nuclear, Gaithersburg, MD). Nucleotide sequence was determined by the dideoxy chain termination method, with T7 DNA polymerase and alkaline-denatured double-stranded plasmid DNA templates (Sanger et al., 1977).

Cell Culture and mRNA Isolation. Human umbilical vein endothelial cells (HUVEC) were obtained from Clonetics Corp. (San Diego, CA) and grown in EGM-UV medium (Clonetics) containing 2% fetal calf serum. HUVEC were passaged every 4–6 days; cells from passages 3–5 were used in the experiments. HUVEC were grown to confluence and stimulated for 70 min with 3 ng/mL recombinant human tumor necrosis factor-α (TNF-α), a gift from Knoll Pharmaceuticals (Whippany, NJ). Bovine aortic endothelial cells (BAEC) were isolated and cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (Sigma Chemical, St. Louis, MO) as described (Yoshizumi et al., 1992). JEG-3 and U-937 cells obtained from the American Type Culture Collection were cultured as suggested by the supplier. The U-937 cells were stimulated with 120 nM 12-*O*-tetradecanoylphorbol 13-acetate (TPA, Sigma Chemical) for 2–48 h. Cellular RNA from HUVEC and U-937 cells (treated and untreated) was prepared by guanidinium isothiocyanate extraction and centrifugation through cesium chloride (Sambrook et al., 1989).

Primer Extension Analysis. A synthetic oligonucleotide primer (5′TTCTTGAGTGTCTTGTCTTGCTCACTCAGCCC 3′) complementary to the 5′ end of HB-EGF cDNA (Higashiyama et al., 1991) was end-labeled with [γ-^{32}P]ATP [with T4 polynucleotide kinase (New England Biolabs, Beverly, MA)] to a specific activity of 6×10^6 cpm/ρmol. About 250 fmol of labeled primer was hybridized to 20 μg of each RNA sample at 50 °C for 2 h in 20 μL of annealing buffer containing 20 mM Tris-HCl, pH 7.5, 250 mM NaCl, and 1 mM EDTA. The reverse transcription reaction was carried out at 42 °C for 1.5 h in 150 μL of reaction mixture containing 0.5 mM each of dideoxynucleotide, 3 mM $MgCl_2$, 75 mM KCl, 50 mM Tris-HCl (pH 7.5), 10 mM dithiothreitol, 0.5 unit/mL RNasin (Promega Biotec, Madison, WI), and 200 units of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Gaithersburg, MD). Extension products were analyzed by electrophoresis with an 8% sequencing gel.

S1 Nuclease Protection Analysis. A 0.8-kb *Pvu*II–*Not*I fragment of the HB-EGF gene (Figure 5) was end-labeled and digested with *Pst*I to generate a 383-base pair (bp) probe (−313 to +70). The labeled probe (6×10^4 cpm) was hybridized to 20 μg of each RNA sample in 20 μL of hybridization buffer containing 40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (1,4-piperazinediethanesulfonic acid), 1 mM EDTA, 0.4 M NaCl, and 80% formamide at 55 °C for 14 h. The DNA–RNA hybrid was mixed with 300 μL of nuclease S1 mapping buffer (0.28 M NaCl, 0.05 M sodium acetate, pH 4.4, 4.5 mM $ZnSO_4$, and 20 μg/mL glycogen). The mixture was then incubated with 500 units of S1 nuclease (Promega Biotec) at 37 °C for 30 min. The reaction was terminated with 80 μL of stop buffer containing 4 M ammonium acetate, 50 mM EDTA (pH 8.0), and 50 μg/mL glycogen. The final reaction products were analyzed with an 8% sequencing gel.

Transfection and Chloramphenicol Acetyltransferase Assay. The plasmids pCAT-basic and pCAT-control contained the prokaryotic chloramphenicol acetyltransferase (CAT) gene (Promega). pCAT-basic had no promoter, and pCAT-control was driven by the SV40 promoter and enhancer. The plasmid pSVβGAL (Promega) contained the β-galactosidase gene driven by the SV40 promoter and enhancer. A DNA fragment containing 2.0 kb of 5′-flanking region from the HB-EGF gene and 60 bp of its first exon was cloned into the pCAT-basic vector to generate the plasmid HB-EGF CAT.

BAEC were transfected with 20 μg of the appropriate CAT construct plasmid DNA by the calcium phosphate method as described (Lee et al., 1990). To correct for variability in transfection efficiency, 10 μg of pSVβGAL plasmid DNA was cotransfected in all experiments. Cell extract was prepared 48 h after transfection by 3 cycles of freezing and thawing. The CAT assay was performed with a modified two-phase fluor diffusion method (Lee et al., 1990). A reaction mixture containing [^{14}C]acetyl coenzyme A (NEC-313L, 4.0 mCi/mmol; Du Pont) and 1 M Tris-HCl (pH 7.8) was added to the cell extract to achieve a final concentration of 1.0 mM chloramphenicol, 0.1 mM [^{14}C]acetyl coenzyme A, and 100 mM Tris-HCl. The reaction mixture was gently overlaid with 5 mL of a water-immiscible scintillation fluor (Econofluor, Du Pont) and then incubated at 37 °C for 2 h. Scintillation assays were performed in a Beckman counter. β-Galactosidase activity was assayed as described (Lee et al., 1990). The ratio of CAT activity to β-galactosidase activity in each sample served as a measure of normalized CAT activity. Each construct was transfected at least 3 times, and each transfection was done in duplicate. Data for each construct were expressed as the mean ± standard error of the mean (SEM). Comparisons of normalized CAT activity among constructs were performed by applying a factorial analysis of variance followed by Fisher's least significant difference test. Statistical significance was accepted at $p < 0.05$.

Chromosomal Mapping. Southern blots were prepared as described (Naylor et al., 1983) from *Hin*dIII-digested DNA from 31 human–mouse somatic hybrid cell lines, as well as parental human and parental mouse cell lines. The hybrids were derived from 18 unrelated human and 4 mouse cell lines (Shows et al., 1982, 1984; Shows, 1983) and characterized by karyotypic analysis and with mapped enzyme markers and mapped DNA probes (Shows et al., 1982, 1984; Shows, 1983). Blots were hybridized with ^{32}P-labeled DNA probes prepared

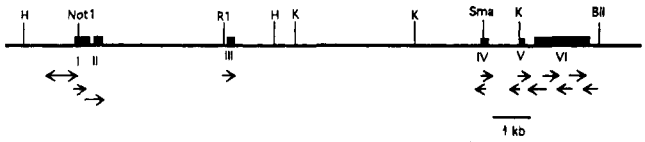

FIGURE 1: Structural organization of the gene coding for human HB-EGF. Partial restriction map of a portion of the cosmid insert containing the HB-EGF gene. H, *Hin*dIII; N, *Not*I; R, *Eco*RI; K, *Kpn*I; S, *Sma*I; B, *Bgl*II. Arrows indicate the direction of sequencing. Filled boxes and roman numerals indicate exons.

from HB-EGF cDNA, washed at high stringency, and autoradiographed as described (Naylor et al., 1983).

RESULTS

Isolation and Characterization of the Human HB-EGF Gene. A 0.8-kb *Eco*RI–*Kpn*I HB-EGF cDNA fragment was used to screen a genomic library prepared from human placenta DNA in a cosmid vector. One hybridizing clone was examined by restriction enzyme DNA mapping (Figure 1), and the restriction fragments were subcloned into plasmid vectors for nucleotide sequence analysis. This cosmid clone contained the entire HB-EGF gene and 8 kb of 5′ flanking sequence.

The human HB-EGF gene is organized into six exons spanning approximately 14 kb (Figure 1). The exons range from 91 bp to 1452 bp long, whereas the introns range from 197 bp to 6.6 kb long (Figure 2). Excluding noncoding exon VI, the average length of the HB-EGF exons, 182 bp, is consistent with the length of exons in general, 137 bp (Hawkins, 1988). Also, nucleotide sequences flanking the HB-EGF gene exons have features typical of many eukaryotic genes. For example, the five introns are bounded by consensus 5′-GT and 3′-AG splicing sequences (Breathnach & Chambon, 1981).

Exon I contains the 5′-untranslated region and sequences encoding the first 16 amino acids of prepro-HB-EGF (including most, if not all, of the signal peptide). Exon II includes sequences encoding amino acids 17–74 of prepro-HB-EGF, probably corresponding to the HB-EGF propeptide (Higashiyama et al., 1992). Exon III encodes the putative heparin-binding domain (Higashiyama et al., 1992) and the first two disulfide loops of the mature HB-EGF peptide (amino acids 75–133). Exon IV encodes the third disulfide loop and the hydrophobic transmembrane domain (amino acids 134–185).

Exon V encodes the cytoplasmic domain of the HB-EGF precursor (amino acids 186–208), and exon VI contains the 3′-untranslated sequences. At the 3′ end of the HB-EGF gene, exon VI contains a consensus polyadenylation signal (Figure 2). Poly(A) addition occurs 17 bp downstream from this AATAAA motif (Higashiyama et al., 1991). The 3′-untranslated portions of the HB-EGF cDNA and gene also contain several ATTTA motifs (Higashiyama et al., 1991), which are common to RNA species with short half-lives (Malter, 1989).

Identification of the Transcription Start Site. The transcription start site of the HB-EGF gene was defined by primer extension and S1 nuclease protection. Primer extension analysis was performed on total RNA from U-937 cells treated with phorbol ester and on total RNA from HUVEC treated with or without TNF-α. Primer extension analysis revealed a primary transcription start site located 14 bp beyond the 5′ end of the cDNA (marked by an arrow, Figure 3) and corresponding to an A nucleotide (Figure 4). The 5′CA3′ nucleotide pair at this site is the most common eukaryotic transcription start site (Bucher & Trifonov, 1986). The analysis also revealed a minor band corresponding to a minor transcription start site (Figure 4). The start site of the HB-EGF gene was also confirmed by using an end-labeled HB-EGF fragment (nucleotides −313 to +70) in an S1 nuclease protection assay (Figure 5). Protected fragments were observed when the probe was incubated with total RNA prepared from TPA-treated U-937 or TNF-α-treated HUVEC but not when the probe was incubated with tRNA or total RNA from JEG-3 (a choriocarcinoma cell line that does not express HB-EGF). The transcription start sites predicted by the two major protected fragments were consistent with the start sites identified by primer extension.

U-937 and HUVEC RNA both generated visible bands on primer extension analysis (Figure 4), which is consistent with the presence of HB-EGF mRNA in both cell types (Higashiyama et al., 1991; Yoshizumi et al., 1992). The density of the bands generated by RNA isolated from HUVEC treated with TNF-α was 10 times higher than that generated by RNA from untreated HUVEC (Figure 4). This difference is consistent with our previous observation that TNF-α increased the steady-state level of HB-EGF mRNA 12-fold (Yoshizumi et al., 1992).

Identification of Potential Cis-Acting Sequences. The 5′-flanking sequences of the HB-EGF gene are characterized by islands rich in G and C and the absence of TATAA and CCAAT boxes (Figure 3). These characteristics are very similar to those of the gene encoding another member of the EGF family, transforming growth factor-α (TGF-α) (Jakobovits et al., 1988). A CTGGGA hexanucleotide is located at position −701 (Figure 3). This sequence is thought to mediate induction of acute-phase reactant genes (Adrian et al., 1986). In the HB-EGF promoter are three GATA consensus sequences, at positions −584, −489, and −466. These sequences have been shown to bind GATA-1, GATA-2, and GATA-3 (Orkin, 1990; Lee et al., 1991b; Ho et al., 1991). An atypical AP1 consensus sequence, TCAGTCA, is located at position −324 of the HB-EGF promoter. Finally, the HB-EGF promoter contains one AP2 site (at position −148) and three Sp1 sites (at positions −167, −72, and −62). Sp1 sites have been shown to be important to the function of the TGF-α promoter (Chen et al., 1992; Shin et al., 1992).

Promoter Activity of the 5′-Flanking Sequence of the HB-EGF Gene. To determine whether the sequences flanking the 5′ end of the HB-EGF gene have promoter activity, we transfected into BAEC the plasmid HB-EGF CAT, which contained 2.0 kb of HB-EGF 5′-flanking sequence linked to the CAT gene, as well as two control plasmids (pCAT-basic and pCAT-control). Background activity in our CAT assay measured about 2500–3000 cpm. The normalized CAT activities of pCAT-basic, HB-EGF CAT, and pCAT-control were 4015 ± 1331, 424 792 ± 119 791*, and 837 231 ± 269 869* cpm, respectively (asterisk indicates activity significantly greater than that of pCAT-basic, $p < 0.05$). Thus, HB-EGF CAT induced a level of CAT activity more than 100-fold higher than that induced by promoterless pCAT-basic. This difference indicates that the 2.0-kb 5′-flanking region of the HB-EGF gene contains promoter activity.

Chromosomal Assignment of the Human HB-EGF Gene. DNA isolated from human–mouse somatic hybrid cell lines and their parental cells was examined for the presence or absence of the human HB-EGF gene by Southern blot analysis (Naylor et al., 1983). Human and mouse DNA restriction fragments hybridized to the HB-EGF cDNA in easily distinguishable patterns. In *Hin*dIII-digested DNA, two human bands (11.2 and 6.8 kb) were noted whereas a single, faint murine band (14 kb) was noted (data not shown). Southern blots from 31 hybrids were scored for the presence or absence of human HB-EGF sequences (Table I). Concordant hybrids either retained or lost the human bands

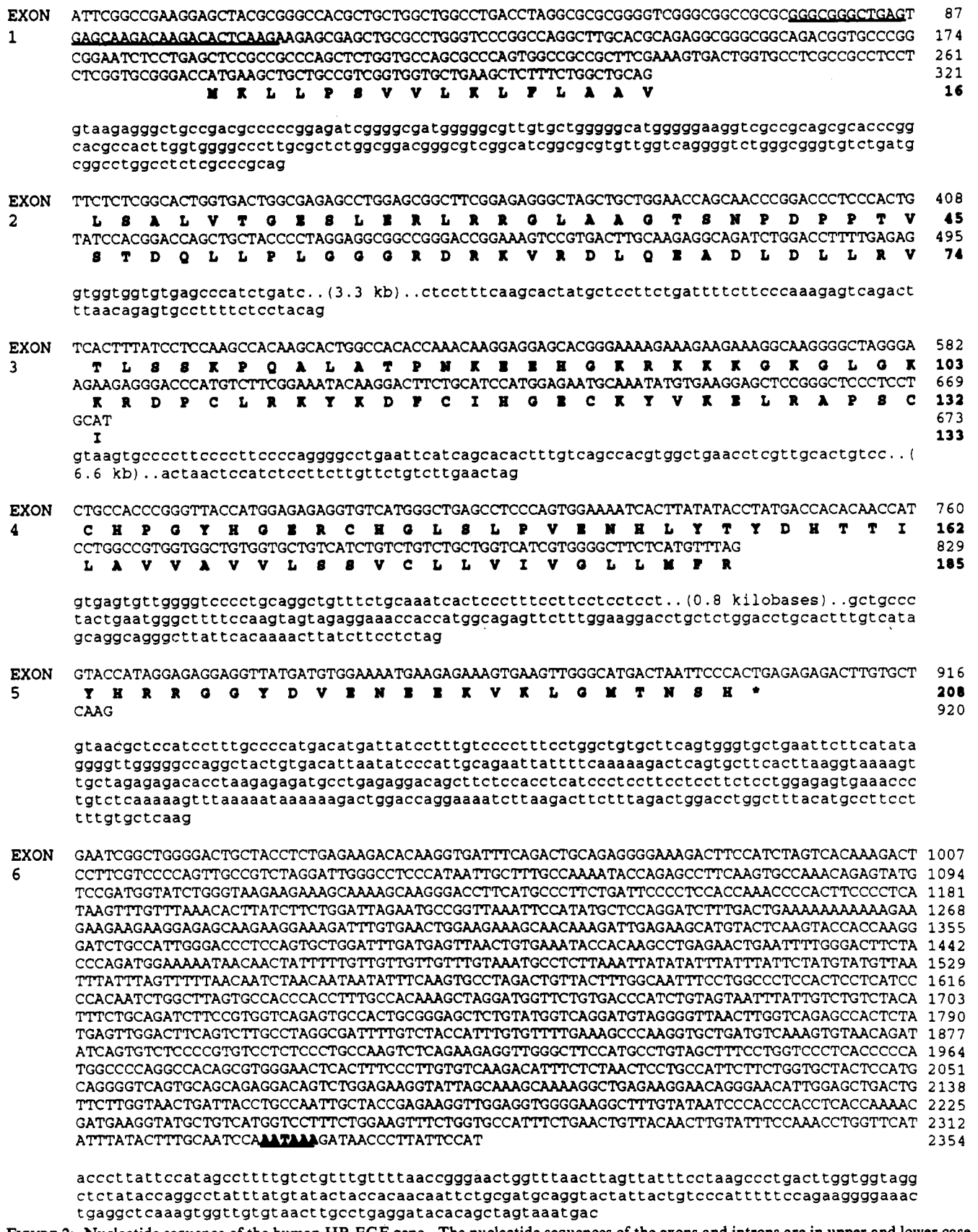

FIGURE 2: Nucleotide sequence of the human HB-EGF gene. The nucleotide sequences of the exons and introns are in upper and lower case letters, respectively. The deduced peptide sequences are capitalized boldface letters. Nucleotide and peptide sequences (in boldface) are numbered on the right, and the lengths of unsequenced intron are in parentheses. In exon I, the sequences used in the design of the oligonucleotide for primer extension are underlined. The polyadenylation signal of exon VI is in capitalized boldface underlined. GenBank Accession Numbers: exon I, L17028; exon II, L17029; exon III, L17030; exon IV, L17031; exon V, L17032.

together with a specific human chromosome (Shows et al., 1982, 1984; Shows, 1983). Discordant hybrids either retained human bands but not a specific chromosome or vice versa. Percent discordance (Table I) indicates the degree of discordant segregation for a marker and a chromosome. Assignment of the HB-EGF gene to chromosome 5 was based on a 0% discordance. [The genes for EGF, transforming growth factor-α, and amphiregulin have been assigned to chromosomes

```
gataagcttgatctgggtcctatgggagggtggctcatttttttatgttgaggaaaaaggtggcagtcaaagg  -765

caaggaaaagaataaggcttgtttcacaacctgtgtctctccaatctcagaagcctgggaattacaacccat  -684

agttcaaaaatcctaatgtcccctgagggaagatgggaggaagccactccagagactaagagagtactaact  -605
                GATA
tcctttgcttttgctggattgatactaagtgcctgcaacttcaactctggttatctcgggtttatgagaacg  -526
                          GATA                    GATA
gctggcattcacgtgtctaatttttaaatgatagatgaattaattattatttgatacttgccgtgttggtag  -447

gccaaaaggtgggcttctgaaacttagagggtgacagggatggagacccctggtatggggcttagatgaagc  -368
                                AP1
aagaggccaggcccggaacagaactcagctcagcagtcagtcacaaggctgcagagtgacaggcgtttgggg  -289

aaaggtaggaaccgctctccgccacctgccggtcagtagccctacccacaatcaggccgcggctgcctcggc  -210
                          Sp1                   AP2
ctggtcccaaaaattgtgaacgccgccctctcctcccgccagtcctccgcgcgcccgcgggtcgggggct  -130
                                                Sp1       Sp1
gcctgccggcgccgcgagccgggcgagcttcccagcacgggaggagggaggggcggcgccgggcgggcgg  -52
                                                →         ‡
agcttccgcgcctcctggagccttattcgcgccgcgagccgggcagcgctcattcggccgaaggagctacgc  +21
```

FIGURE 3: HB-EGF promoter sequences. Nucleotide sequences are numbered on the right. Potential cis-acting sequences are underlined. The major transcription start is marked by an arrow, and the end of the cDNA (Higashiyama et al., 1991) is marked by a double dagger. The GenBank Accession Number for the promoter sequence is L17033.

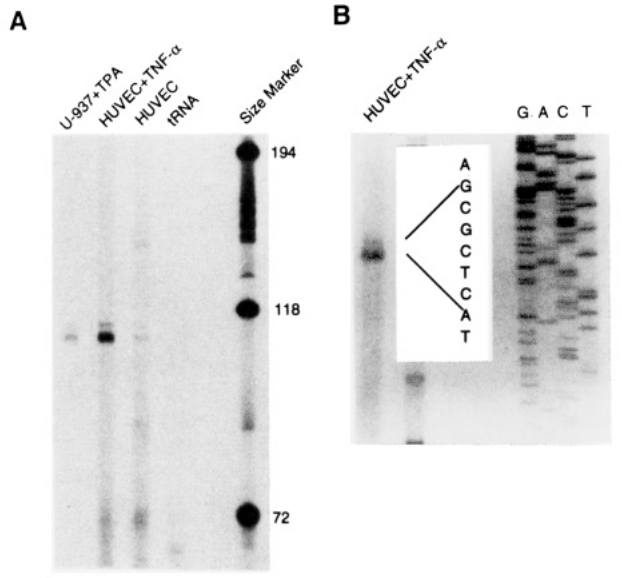

FIGURE 4: Primer extension analysis of the HB-EGF gene transcription start site. (A) The oligonucleotide highlighted in Figure 2 (exon I) was hybridized to 20 μg of total RNA prepared from U-937 cells treated with 12-*O*-tetradecanoylphorbol 13-acetate (U-937+TPA) for 2 h, untreated human umbilical vein endothelial cells (HUVEC), HUVEC treated with tumor necrosis factor-α (HUVEC + TNF-α), and tRNA. Extension products were analyzed on an 8% sequencing gel. The size marker (in bp) was prepared by radiolabeling ϕX174 RF DNA digested with *Hae*III. (B) A Sanger sequencing reaction, with the same oligonucleotide used as sequencing primer, was run alongside the HUVEC + TPA primer extension reaction. Sequences of the HB-EGF gene corresponding to the primer extension products are shown.

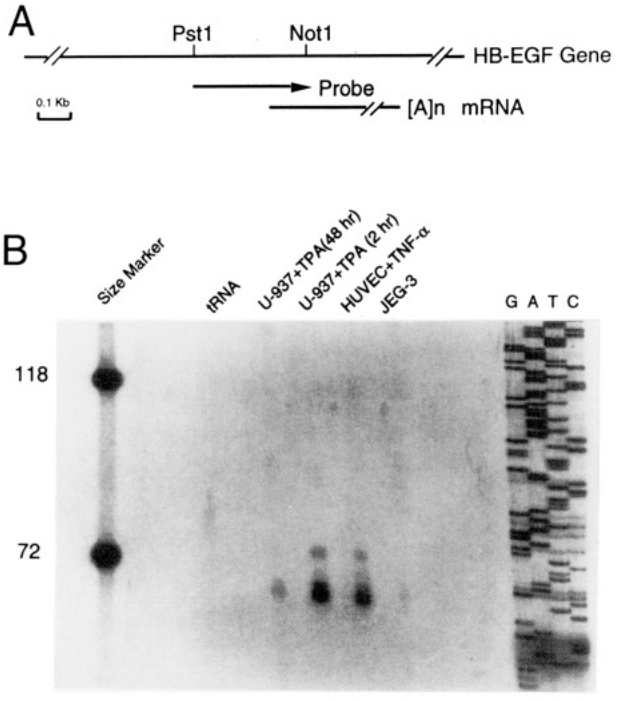

FIGURE 5: S1 nuclease protection assay. (A) Strategy for mapping the 5′ end of the HB-EGF gene by S1 nuclease analysis. (B) S1 nuclease protection mapping of the HB-EGF gene start site. Total RNA from U-937 cells treated with 12-*O*-tetradecanoylphorbol 13-acetate (TPA) for 48 and 2 h, human umbilical vein endothelial cells (HUVEC) treated with tumor necrosis factor-α (TNF-α) for 70 min, JEG-3 cells, and tRNA was incubated with a 383-bp ^{32}P-labeled probe spanning the immediate 5′ region of the HB-EGF gene. The resulting DNA–RNA hybrids were incubated with S1 nuclease, and the protected fragments were analyzed on an 8% sequencing gel. A Sanger sequencing reaction was run alongside to estimate the size of the protected fragment. The size marker (in bp) was the same as used in Figure 4. Arrows indicate the major protected bands.

3, 2, and 4, respectively (Brissenden et al., 1984, 1985; Plowman et al., 1990).]

DISCUSSION

Exon Organization of the EGF-like Domain of the HB-EGF Gene Is Identical to That of Other Members of the EGF Family. The genes encoding three other members of the EGF family (EGF, TGF-α, and amphiregulin) have been isolated and characterized (Brissenden et al., 1984, 1985; Plowman et al., 1990). Like TGF-α and amphiregulin, the human HB-EGF gene is divided into 6 exons (whereas the human EGF gene contains 24 exons). The HB-EGF gene spans 14 kb, a size similar to that of the gene for amphiregulin (10.2 kb) but much smaller than that of the genes for EGF (110 kb) and TGF-α (85 kb). In spite of these differences in size, the exon organization of the EGF-like domains of the four genes is completely conserved. For HB-EGF, the mature peptide is encoded by exons III and IV. For EGF, TGF-α, and amphiregulin, the mature peptides are similarly encoded by two exons (Bell et al., 1986; Blasband et al., 1990; Plowman et al., 1990). Exons III and IV interrupt the EGF domain of the HB-EGF peptide at Ile^{59}, the same point at which mature

Table I: Segregation of an HB-EGF DNA Probe with Human Chromosomes in *Hin*dIII-Digested Human–Mouse Cell Hybrid DNA[a]

#	HYBRID	HB-EGF	Human Chromosomes 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Translocations
1298	20L-4	-	-	-	-	-	-	+	+	+	-	+	t	-	-	-	-	-	+	-	-	+	+	+	-	11p-
1285	20L-9	+	t	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	1/17
1287	20L-17	-	t	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	t	-	-	-	-	-	-	17/1
1277	20L-37	+	t	-	+	+	+	+	+	+	-	+	-	+	-	-	-	-	+	+	-	+	-	-	+	1/17
1293	31L-6	-	-	-	+	-	-	+	-	-	-	+	-	+	+	+	-	-	+	+	-	-	+	-	-	
1243	35R-14	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+	-	-	-	-	+	+	+	
1238	55R-16	+	+	+	+	t	+	+	+	+	-	+	t	+	+	-	+	+	-	-	+	-	+	+	+	11/4, 7q-
1246	55R-33	+	+	+	+	+	+	+	+	+	+	-	t	+	+	+	-	+	+	-	-	-	+	+	+	11/4
48	ATR-13	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	-	-	-	t	5/X
1211	DUA-3BSAGA	-	-	+	-	-	-	-	+	+	-	-	-	-	-	+	+	-	-	+	-	-	-	-	-	-
186	DUM-13	+	+	+	+	-	+	+	+	-	-	+	+	+	-	+	t	+	+	+	+	+	+	+	t	X/15, 15/X
1185	GAR-1	+	-	-	+	-	+	-	-	+	-	+	-	+	+	-	+	+	-	-	-	+	-	-	+	
1178	JSR-17S	+	+	+	+	+	+	-	t	+	t	+	+	+	+	+	+	+	+	+	-	+	+	+	-	7/9
653	JWR-22H	-	t	t	-	+	-	+	-	+	-	+	+	+	+	+	+	-	+	+	-	+	+	-	-	2/1
187	JWR-26C	+	t	+	+	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	-	+	+	-	+	1/2
40	NSL-5	-	+	-	-	-	-	-	-	+	t	+	-	+	-	+	-	+	t	+	-	+	-	-	-	17/9, 12q+
1146	NSL-9	+	-	-	-	-	+	-	-	+	t	+	-	+	+	+	+	+	+	+	-	+	+	+	-	17/9
192	NSL-16	+	-	-	+	+	+	-	+	+	t	+	-	+	-	+	+	+	+	+	-	+	+	-	-	17/9
42	REW-11	-	-	-	-	+	-	-	+	-	-	-	+	+	+	-	-	+	-	-	-	+	+	+	+	
184	REX-11BSAgB	-	-	-	+	-	-	-	-	-	-	+	-	-	-	+	+	-	-	+	-	-	-	-	-	
1162	RSR-3	-	-	-	-	+	-	-	+	-	-	+	+	-	-	+	-	+	+	-	-	-	+	-	+	
665	SIR-11	-	-	+	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	+	-	+	
643	TSL-1	-	-	+	+	+	-	-	-	-	-	+	+	-	+	-	-	+	+	+	-	+	+	-	-	
53	TSL-2	+	-	+	t	-	+	+	-	-	-	+	-	+	-	-	-	-	t	+	-	+	+	-	+	17/3, 3/17
395	VTL-6	-	-	+	-	-	-	+	+	+	-	+	+	-	-	-	+	-	+	-	+	+	+	+	-	
407	VTL-17	+	-	-	-	-	+	-	+	-	-	+	+	-	+	+	-	-	+	-	-	+	+	-	-	
1098	VTL-23	+	-	-	-	-	+	-	+	+	-	+	+	+	-	+	-	-	+	+	-	-	+	+	-	
559	W12	+	-	+	-	-	+	-	-	-	-	-	+	-	-	+	-	-	+	-	+	+	+	+	-	11p-
794	WIL-8Y	-	-	-	+	-	-	+	+	+	-	+	+	-	-	+	-	-	+	+	-	+	+	-	+	
25	WIL-15	-	-	+	+	+	-	+	+	-	-	+	+	+	+	+	+	-	+	+	-	+	+	-	+	
1224	XTR-22	+	-	+	-	+	+	+	-	+	-	+	+	-	+	-	+	-	-	+	+	+	+	+	+	
	Chromosome		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	
	Concordant #	(+/+)	6	10	11	8	17	10	11	12	3	15	9	14	10	11	10	11	12	11	6	12	14	10	9	
	of Hybrids	(-/-)	11	8	9	9	14	8	7	8	12	4	6	9	7	6	10	10	3	7	13	6	4	11	9	
	Discordant #	(+/-)	8	7	5	8	0	7	5	5	11	2	6	3	7	6	6	6	4	6	11	5	3	7	6	
	of Hybrids	(-/+)	1	5	5	5	0	6	7	6	1	10	7	5	7	8	4	4	9	7	1	8	10	3	5	
	% Discordancy		35	40	33	43	0	42	40	35	44	39	46	26	45	45	33	32	46	42	39	42	42	32	38	

[a] The 31 hybrids involve 18 unrelated human cell lines and 4 mouse cell lines. Hybrids were characterized by karyotypic analysis and by mapped enzyme markers. The HB-EGF DNA probe was hybridized to Southern blots containing *Hin*dIII-digested DNA from the human–mouse hybrids. A "t" indicates a chromosome translation with no intact chromosome present. Scoring was determined by the presence (+) or absence (-) of human bands in the hybrids on the blots. A 0% discordance indicates a matched segregation of the DNA probe with a chromosome.

EGF, TGF-α, and amphiregulin are interrupted by two exons.

Other proteins such as the low-density lipoprotein receptor, tissue-type plasminogen activator, urokinase, human coagulation factors IX, X, and XII, and endothelial-leukocyte adhesion molecule 1 also contain an EGF-like domain (Carpenter & Cohen, 1990; Carpenter & Wahl, 1990). However, none of these EGF-like domains maintains the exact spacing of all six cysteines, and none has been shown to bind the EGF receptor. Furthermore, the EGF-like domain in these proteins is contained within a single exon (Carpenter & Cohen, 1990; Carpenter & Wahl, 1990; Collins et al., 1991; Plowman et al., 1990). These differences would suggest distinct origins for the two types of EGF-like domain. On the basis of its ability to bind the EGF receptor (Higashiyama et al., 1991) and its characteristic two-exon EGF-like domain (present study), we conclude that HB-EGF is the fourth member of the EGF family of growth modulators present in the human genome.

The HB-EGF 5′-Flanking Region Functions as a Promoter in Endothelial Cells. The 2.0-kb HB-EGF 5′-flanking sequence directed a significant level of CAT activity when transfected into BAEC, an observation consistent with our previous finding that vascular endothelial cells transcribe the HB-EGF gene (Yoshizumi et al., 1992). It is also noteworthy that the promoters of both HB-EGF and endothelin-1 (a potent vasoconstrictor expressed at a high level in vascular endothelial cells) contain AP-1 and GATA consensus sequences (Lee et al., 1990, 1991b) (Figure 3). Both the AP-1 site and the GATA site are necessary to direct high-level expression of the endothelin-1 gene in endothelial cells. The HB-EGF AP-1

site differs from the classic AP-1 site TGAG/CTCA (Curran & Franza, 1988; Risse et al., 1989) by only one base pair. It also happens that the endothelin-1 AP-1 site TGACTAA, at position −109 of the gene (Lee et al., 1991a), differs from the classic AP-1 site by only one base pair. According to the study by Risse et al. (1989) on the effect of AP-1 point mutations on Fos and Jun binding affinity, neither the HB-EGF nor the endothelin-1 AP-1 sequence should bind the Fos and Jun proteins. However, we have shown that the endothelin-1 AP-1 site binds Fos and Jun, allows endothelin-1 gene regulation by these *trans*-acting factors, and is essential for expression of the gene in vascular endothelial cells in culture (Lee et al., 1991a). It will be useful to determine whether the HB-EGF AP-1 and GATA sites also have essential roles in the transcriptional regulation of the HB-EGF gene in vascular endothelial cells (Yoshizumi et al., 1992).

Both the HB-EGF Gene and Susceptibility to Diphtheria Toxin Map to Chromosome 5. A recent report of the cloning of the diphtheria toxin receptor indicates that it is identical to the HB-EGF precursor (Naglich et al., 1992). Human cells are 10 000 times more sensitive than mouse cells to diphtheria toxin in vitro because of the presence of more functional surface receptors on the human cells. However, mouse–human hybrid cells in which chromosome 5 is present are as sensitive to the toxin as are human cells, whereas hybrids without chromosome 5 are as resistant to it as mouse cells (Creagan et al., 1975). On the basis of experiments in mouse–human hybrids, diphtheria toxin susceptibility has been assigned to the q23 region of chromosome 5 (Creagan et al., 1975; Hayes et al., 1987). This assignment is consistent with our assignment of the HB-EGF gene to chromosome 5 (Table I), and it suggests that the HB-EGF gene may be located in the q23 region of this chromosome.

In summary, we have cloned the human gene encoding HB-EGF (and the diphtheria toxin receptor as well), defined the gene's transcription start site, demonstrated the promoter activity of its 5′ flanking sequences, and mapped the gene to chromosome 5. The availability of the HB-EGF gene along with its 5′-flanking sequences will allow us to study the molecular mechanisms regulating its expression, the role of the HB-EGF peptide in vascular disease, and the differential sensitivity of cells to diphtheria toxin.

ACKNOWLEDGMENT

We thank Dr. Edgar Haber for his continued enthusiasm and support of this work. We thank Dr. H. S. Sul for her helpful suggestions as well as critical review of the manuscript and Mr. Thomas J. McVarish for his editorial assistance.

REFERENCES

Adrian, G. S., Korinek, B. W., Bowman, B. H., & Yang, F. (1986) *Gene 49*, 167–175.
Bell, G. I., Fong, N. M., Stempien, M. M., Wormsted, M. A., Caput, D., Ku, L., Urdea, M. S., Rall, L. B., & Sanchez-Pescador, R. (1986) *Nucleic Acids Res. 14*, 8427–8446.
Blasband, A. J., Rogers, K. T., Chen, X., Azizkhan, J. C., & Lee, D. C. (1990) *Mol. Cell. Biol. 10*, 2111–2121.
Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.
Brissenden, J. E., Ullrich, A., & Francke, U. (1984) *Nature 310*, 781–784.
Brissenden, J. E., Derynck, R., & Francke, U. (1985) *Cancer Res. 45*, 5593–5597.
Bucher, P., & Trifonov, E. N. (1986) *Nucleic Acids Res. 14*, 10009–10026.
Carpenter, G., & Cohen, S. (1990) *J. Biol. Chem. 265*, 7709–7712.
Carpenter, G., & Wahl, M. I. (1990) in *Peptide Growth Factors and Their Receptors* (Sporn, M. B., & Roberts, A. B., Eds.) Vol. 1, pp 69–171, Springer-Verlag, Berlin and Heidelberg.
Chen, X., Azizkhan, J. C., Lee, D. C. (1992) *Oncogene 7*, 1805–1815.
Collins, T., Williams, A., Johnston, G. I., Kim, J., Eddy, R., Shows, T., Gimbrone, M. A. J., & Bevilacqua, M. P. (1991) *J. Biol. Chem. 266*, 2466–2473.
Creagan, R. P., Chen, S., & Ruddle, F. H. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2237–2241.
Curran, T., & Franza, B. R. J. (1988) *Cell 55*, 395–397.
Dynan, W. S., & Tjian, R. (1983) *Cell 32*, 669–680.
Dynan, W. S., Sazer, S., Tjian, R., & Schimke, R. T. (1986) *Nature 319*, 246–248.
Eidels, L., Proia, R., & Hart, D. A. (1983) *Microbiol. Rev. 47*, 596–620.
Hawkins, J. D. (1988) *Nucleic Acids Res. 16*, 9893–9908.
Hayes, H., Kaneda, Y., Uchida, T., & Okada, Y. (1987) *Chromosoma 96*, 26–32.
Higashiyama, S., Abraham, J. A., Miller, J., Fiddes, J. C., & Klagsbrun, M. (1991) *Science 251*, 936–939.
Higashiyama, S., Lau, K., Besner, G. E., Abraham, J. A., & Klagsbrun, M. (1992) *J. Biol. Chem. 267*, 6205–6212.
Ho, I.-C., Vorhees, P., Marin, N., Oakley, B. K., Tsai, S.-F., Orkin, S. H., & Leiden, J. M. (1991) *EMBO J. 10*, 1187–1192.
Imagawa, M., Chiu, R., & Karin, M. (1987) *Cell 51*, 251–260.
Jakobovits, E. B., Schlokat, U., Vannice, J. L., Derynck, R., & Levinson, A. D. (1988) *Mol. Cell. Biol. 8*, 5549–5554.
Lee, M. E., Bloch, K. D., Clifford, J. A., & Quertermous, T. (1990) *J. Biol. Chem. 265*, 10446–10450.
Lee, M. E., Dhadly, M. S., Temizer, D. H., Clifford, J. A., Yoshizumi, M., & Quertermous, T. (1991a) *J. Biol. Chem. 266*, 19034–19039.
Lee, M. E., Temizer, D. H., Clifford, J. A., & Quertermous, T. (1991b) *J. Biol. Chem. 266*, 16188–16192.
Libby, P., & Hansson, G. K. (1991) *Lab. Invest. 64*, 5–15.
Malter, J. S. (1989) *Science 246*, 664–666.
Munro, J. M., & Cotran, R. S. (1988) *Lab. Invest. 58*, 249–261.
Naglich, J. G., Metherall, J. E., Russell, D. W., & Eidels, L. (1992) *Cell 69*, 1051–1061.
Naylor, S. L., Sakagushi, A. Y., Shows, T. B., Law, M. L., Goeddel, D. V., & Gray, P. W. (1983) *J. Exp. Med. 157*, 1020–1027.
Orkin, S. H. (1990) *Cell 63*, 665–672.
Plowman, G. D., Green, J. M., McDonald, V. L., Neubauer, M. G., Disteche, C. M., Todaro, G. J., & Shoyab, M. (1990) *Mol. Cell. Biol. 10*, 1969–1981.
Risse, G., Jooss, K., Neuberg, M., Brüller, H.-J., & Müller, R. (1989) *EMBO J. 8*, 3825–3832.
Ross, R. (1986) *N. Engl. J. Med. 314*, 488–500.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory Press, Plainview, NY.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schwartz, S. M., Heimark, R. L., & Majesky, M. W. (1990) *Physiol. Rev. 70*, 1177–1209.
Shin, T. H., Paterson, A. J., Grant, J. H., Meluch, A. A., & Kudlow, J. E. (1992) *Mol. Cell. Biol. 12*, 3998–4006.
Shows, T. B. (1983). *Isoenzymes: Curr. Top. Biol. Med. Res. 10*, 323–339.
Shows, T. B., Sakaguchi, A. Y., & Naylor, S. L. (1982) *Adv. Hum. Genet. 12*, 341–452.
Shows, T. B., Eddy, R., Haley, L., Byers, M., Henry, M., Fujita, T., Matsui, H., & Taniguchi, T. (1984) *Somat. Cell Mol. Genet. 10*, 315–318.
Temizer, D. H., Yoshizumi, M., Perrella, M. A., Susanni, E. E., Quertermous, T., & Lee, M. (1992) *J. Biol. Chem. 267*, 24892–24896.
Yoshizumi, M., Kourembanas, S., Temizer, D. H., Cambria, R. P., Quertermous, T., & Lee, M. E. (1992) *J. Biol. Chem. 26*, 9467–9469.