Molecular cloning of mouse epiregulin, a novel epidermal growth factor-related protein, expressed in the early stage of development

Hitoshi Toyoda, Toshi Komurasaki*, Yoko Ikeda, Makoto Yoshimoto, Shigeo Morimoto

Molecular Biology Laboratory, Medicinal Research Laboratories, Taisho Pharmaceutical Co., Ltd., 1-403, Yoshino-cho, Ohmiya, Saitama 330, Japan

Received 6 November 1995; revised version received 20 November 1995

Abstract **A cDNA clone encoding a novel epidermal growth factor (EGF)-related growth regulator, epiregulin, was isolated from a cDNA library prepared from a mouse fibroblast-derived tumor cell line, NIH3T3/clone T7. The predicted amino acid sequence revealed that the purified epiregulin peptide of 46-amino acids was synthesized as an internal segment of a 162-amino acid putative transmembrane precursor. The structural organization was similar to that of TGF-α precursor among the members of the EGF family. Although epiregulin transcript was not detected in several adult normal tissues by Northern blot analysis, approximately 4.8-kb transcript was present in 7-day-old mouse embryo and then diminished to very low or undetectable levels. Our results suggest that epiregulin may play an important role in the regulation of epithelial cell growth during early development.**

Key words: Epiregulin; Epidermal growth factor family; cDNA cloning; Development

1. Introduction

Peptide growth factors are involved in various physiological processes including embryogenesis, immune and inflammatory responses, oncogenic transformation, neuronal survival, and hematopoiesis. They bind to specific membrane receptors to trigger a cascade of intracellular signals, resulting in specific cellular responses [1,2]. The elucidation of the mechanisms of ligand–receptor interactions, signaling pathways, and activation of transcriptional factors is useful for the understanding of the physiological role of these factors.

We have recently purified a novel epidermal growth factor (EGF)-related growth regulator, epiregulin, from a mouse fibroblast-derived tumor cell line, NIH3T3/clone T7 [3]. Epiregulin, along with EGF and transforming growth factor-α (TGF-α), is a small polypeptide of 46-amino acid residues that belongs to the EGF family. It inhibits the growth of several epithelial tumor cell lines but stimulates various types of other cells. The binding of epiregulin to the EGF receptors on epidermoid carcinoma A431 cells is weaker than that of EGF [3]. Nevertheless, it is more potent than EGF as a mitogen for rat primary hepatocytes and Balb/c 3T3 A31 fibroblasts [3].

In the present study, we report the isolation and the nucleotide sequence of cDNA encoding mouse epiregulin. We analyzed the expression of the epiregulin gene in several normal adult tissues and found that the expression was detected in the early stage of embryonic life, suggesting that the physiological significance of epiregulin is distinct from other members of the EGF family and that it is involved in the early development process as an EGF-related ligand.

2. Materials and methods

2.1. Amplification of epiregulin cDNA fragments for probes by PCR

Poly (A)$^+$ RNA of NIH3T3/ clone T7 cells was prepared with Fast Tract (Invitrogen, San Diego, CA) and oligo dT primed cDNA was synthesized using cDNA synthesis kit (Pharmacia, Uppsala, Sweden) according to the instructions provided by the manufacture. A cDNA library was constructed in λgt10 (Stratagene, La Jolla, CA) as a vector.

The following degenerated oligonucleotides were designed from the amino acid sequence of epiregulin and synthesized with an automated DNA synthesizer (Model 380B, Perkin-Elmer Instrument, Normalk, CT). P1, 5′-GTNCARATHACNAARTG-3′; P2, 5′-ARRAARAAR-TGYTCRCA-3′; P3, 5′-GTRTANCCNACYTCRCA-3′ (N = A, T, G, C; R = A, G; H = A, C, T; Y = C, T). P1 is a 5′-primer corresponding to amino acid residues 1–5. P2 and P3 are 3′-primers corresponding to 46–41, and 37–32, respectively (Fig. 1).

PCR was performed using GeneAmp PCR reagent kit with AmpliTaq DNA polymerase (Perkin-Elmer). Ten microliters of the phage suspension were boiled for 10 min and subjected to the first round PCR using P1 and P2 as primers under the following conditions for 30 cycles: 94°C for 1 min, 40°C for 2 min, and 72°C for 3 min. Ten microliters of the first round PCR product were then applied to a second round PCR using P1 and P3 as primers under the same conditions, and the nucleotide sequence of the PCR product (PR1) was determined as described below. P4 (5′-TAGTTCTGACATGGACGGCT-3′), P5 (5′-TGGACATGAGAGAGAAATTC-3′), P6 (5′-GAATTTCTCTCTC-ATGTCCA-3′), and P7 (5′-AGCCGTCCATGTCAGAACTA-3′) were designed on the basis of the nucleotide sequence of PR1. GF (5′-AGCAAGTTCAGCCTGGTTAA-3′) and GR (5′-ATGAGTATTTC-77CCAGGG-3′) are consensus to the sequence surrounding the *Eco*RI site of λgt10 and used as primers to amplify the suffers. Two more sets of nested PCR reactions were performed using the above primers as shown in Fig. 1. The annealing temperature was at 52°C for the first round PCR, while it was raised to 55°C for the second round PCR.

The resulting products (PR2 and PR3) were subcloned into the *Sma*I

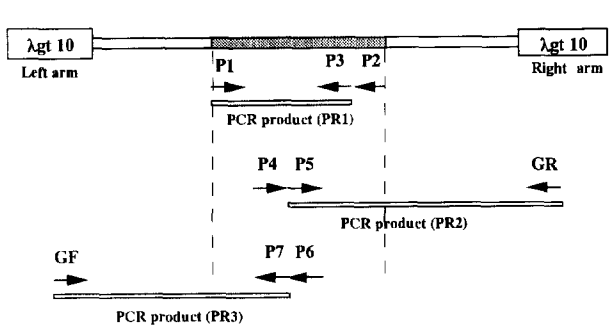

Fig. 1. Strategy for the amplification of epiregulin cDNA fragments by PCR. Right and left directed arrows indicate 5′- and 3′-primers, respectively. The sequences of oligonucleotide primers are described in section 2. The stippled box indicate the region of the amino acid sequence determined from purified epiregulin.

*Corresponding author. Fax: (81) (48) 652 7254.

Abbreviations: EGF, epidermal growth factor; TGF-α, transforming growth factor- α; PCR, polymerase chain reaction; HB-EGF, heparin binding EGF-like growth factor.

SSDI 0014-5793(95)01403-9

A

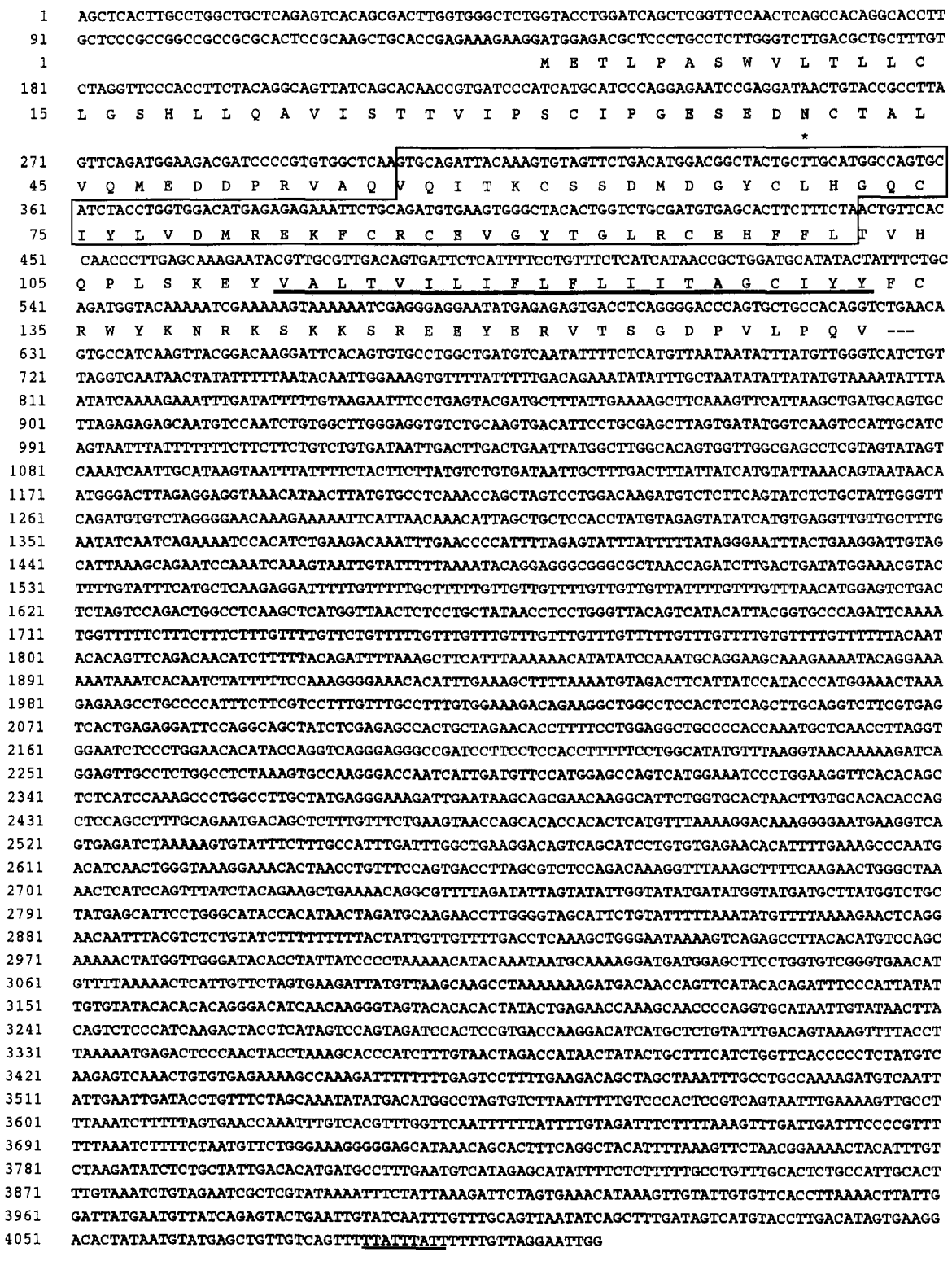

Fig. 2. Nucleotide sequence and deduced amino acid sequence of mouse epiregulin cDNA. (A) Nucleotide and deduced amino acid sequence. The boxed sequence indicates mature epiregulin. The bold underline indicates the putative transmembrane segment. A destabilizing signal (TTATTTATT) for cyotokine and proto-oncogene mRNA is underscored with a single line. The potential *N*-glycosylation site is denoted by an asterisk. Numbers on the left indicate nucleotide (above) and amino acid positions (bottom).

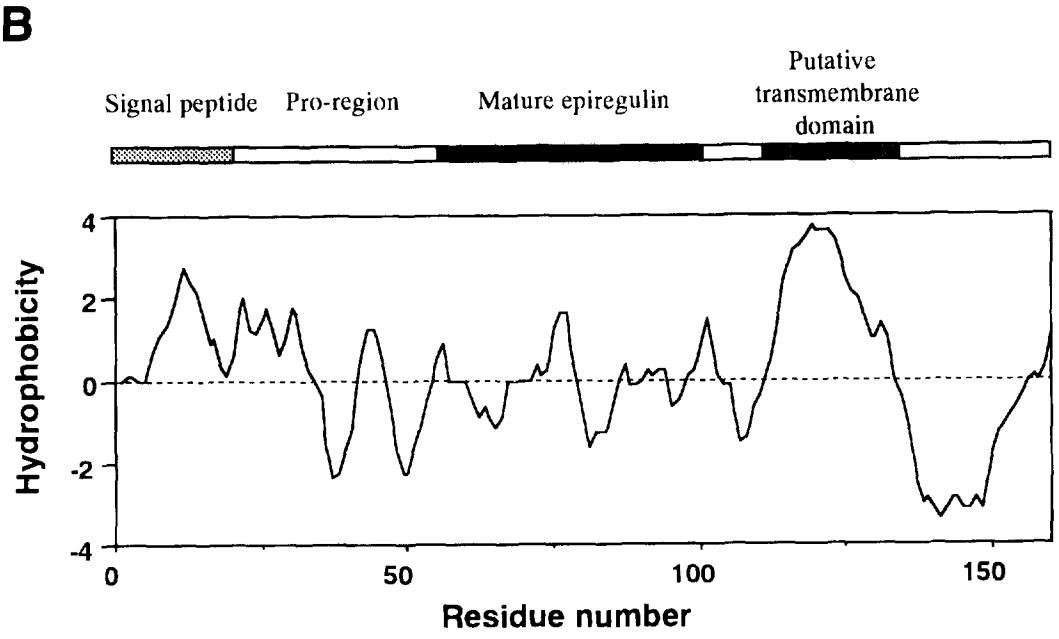

Fig. 2. (B) Hydropathy profile of the deduced amino acid sequence based on the algorithm of Kyte and Doolittle [26]. The nucleotide sequence will appear in the DDBJ/ENBL/GeneBank sequence data bank under the accession number D30782.

site of pUC18, and sequenced to confirm that each fragment encoded a part of epiregulin.

2.2. Screening of the cDNA library

Epiregulin cDNA fragments amplified by PCR (PR2 and PR3) were labeled with [α^{32}P] dCTP using multiprime labeling kit (Amersham, UK). Filter hybridization was carried out at 68°C for 16 hours in a buffer containing 5 × SSC, 5 × Denhardt's solution, 0.5% SDS, 100 mg/ml of sonicated heat-denatured salmon sperm DNA, and the ^{32}P-labeled probe. The blots were washed once in 2 × SSC for 30 min at 68°C, twice in 0.2 × SSC under the same conditions, and then Kodak X-Omat AR films were exposed to the filters at −70°C with intensifying screens. The clones containing the longest insert were isolated and the insert (mTG7a) was subcloned into the *Eco*RI site of pUC18 for DNA sequencing.

5′-terminal portion of the cDNA was amplified using RASE system (Gibco-BRL). The 5′-terminal portion was synthesized using P6 as a primer and an oligo-dC anchor sequence was added to the 5′-end of cDNA by terminal deoxynucleotidyl transferase, as the PCR template. The PCR reaction was performed at annealing temperature 55°C. P7 and a commercially supplied anchor primer were used as PCR primers. Amplified cDNA fragment of 340-bp contained the overlapped regions with mTG7a.

2.3. DNA sequencing

The nucleotide sequence was determined by the dideoxynucleotide chain terminator method [4] using the Sequenase DNA-sequencing kit (United States Biochemical Co., Cleveland, OH). The longest cDNA insert, mTG7a, subcloned into pUC as partially digested with exonuclease III from one end using Kilo-sequencing kit (Takara Shuzou, Co., Kyoto, Japan), and the resultant deletion derivatives were sequenced. Both strands of the cDNA were completely sequenced at least twice to confirm the sequence. At least three different clones from each PCR product were sequenced to determine the final consensus sequence.

2.4. Northern blot analysis

Mouse poly (A)$^-$ RNAs from several normal organs, mouse Multiple Tissue Northern (MTN) blot, and mouse embryo MTN blot filters were purchased from Clontech Laboratories, Inc. (Palo Alto, CA). Poly (A)$^+$ RNAs were separated through 1% agarose gel containing formaldehyde and transferred onto nylon membranes (Hybond N; Amersham). Northern blot hybridization was carried out using 763-bp *Hha*I–*Hin*dIII fragment of epiregulin cDNA as a probe which contained the open reading frame, under high stringent conditions recommended by the manufacture of the MTN blots. Two MTN blots of different lot number in each experiment were used for Northern blot hybridization.

3. Results and discussion

3.1. Isolation of cDNA encoding epiregulin

Fig. 1 illustrates the strategy used to amplify several parts of epiregulin cDNA by PCR as described above. The amino acid sequence predicted from the nucleotide sequence of PR1 (110 bp) was identical to that of epiregulin corresponding to amino acid residues 1–37. Comparison of the nucleotide sequences of PR2 (401 bp) and PR3 (310 bp) with that of PR1 showed that PR1 contained regions that overlapped with PR2 and PR3. By screening NIH3T3/clone T7 cDNA library using both PR2 and PR3 as probes, 14 positive clones were isolated from approximately 3.5 × 10^5 pharges. Among these, three clones contained an insert of approximately 4.1 kb, that gave bands of similar size upon cleavage with *Pst*I and *Acc*I, suggesting that they had the same reaction enzyme sites. The nucleotide sequence of one of these clones was determined. It appeared that the clone had covered nucleotides 43–4,105. Since the length of epiregulin transcript was estimated to be about 4.8 kb (Fig. 4), we cloned the proximal and the distal 5′-end of the cDNA using a similar nested PCR nucleotide and determined the nucleotide sequences 1–42.

3.2. Structure analysis of cDNA encoding epiregulin

Fig. 2A shows the nucleotide sequence of a 4,105 bp cDNA and the deduced amino acid sequence. The cDNA encoded single open reading frame for a protein of 162 amino acid residues flanked by a 140 bp 5′-untranslated region and a 3,476 bp 3′-untranslated region. Since the open reading frame contained the entire epiregulin sequence, it is likely that the protein is the precursor of epiregulin. Though the first ATG was not preceded by the optimal consensus sequence for translational initiation sites [5], it was thought to be a start site due to the

similarity of amino acid sequence to that of human homologue (H. Toyoda et al., unpublished data). The 3′-untranslated region was rich in A/T repeats and contained a destabilizing signals (TTATTTATT) for cytokines and proto-oncogene mRNAs [6,7], suggesting that epiregulin mRNA undergoes rapid turnover. Since this cDNA had no poly $(A)^+$ tail, it appeared to lack about 700 bp from the untranslated region. Hydropathy analysis revealed that the first ATG was followed by a hydrophobic region, which seemed to be the signal peptide (Fig. 2B). Although the precise processing site of the signal peptide was unclear, based on the criteria of von Heijne [8,9], amino acid residues 1–22 was most likely the signal peptide. In addition, a highly hydrophobic stretch was observed at residues 113–132, which was considered to be a putative transmembrane domain. A potential N-linked glycosylation site was found in the amino-terminal region of the epiregulin precursor at residue 40. These features are characteristic of the genes of proteins belonging to the EGF family, including EGF [10,11], TGF-α [12,13], amphiregulin [14], heparin-binding EGF-like growth factor (HB-EGF) [15], and betacellulin [16] (Fig. 3). The mature forms of amphiregulin, HB-EGF, and betacellulin carry relatively long N-terminal regions followed by EGF-like domains, whereas epiregulin as well as EGF and TGF-α do not. Among EGF family members, however, the length of not only the mature form but the pro-region, which precedes mature protein, varies considerably. For instance, mature TGF-α of 50 residues is flanked by the pro-region of 17 to 18 residues. The pro-domain of EGF spans 950 residues and contains mature form of 53 residues and nine EGF-like sequences. On the other hand, the length of mature amphiregulin and its pro-region are 84 and 81 residues, respectively. In the case of epiregulin, the mature and pro-region forms were 46 and 33 residues, respectively. In this respect, the structure of epiregulin precursor is very similar to that of TGF-α.

The purified epiregulin was an internal segment of the precursor. The members of the EGF family are all released as soluble forms by proteolytic processing from the ectodomain of their transmembrane precursors. It is likely that the mature epiregulin is cleaved out from the precursor. The cleavage of mature TGF-α of 50-amino acids from the precursor occurs between alanine and valine residues at both N- and C-termini [12,13]. The precursor form of epiregulin does not carry such

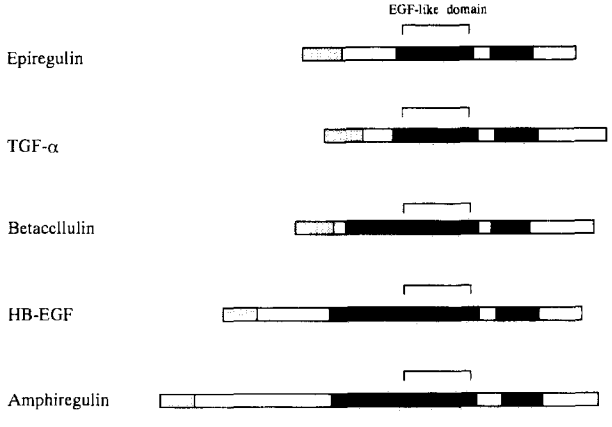

Fig. 3. Schematic representations of structural comparison of EGF-like growth factor precursors. The predicted structural organizations of their precursors are indicated as follows: signal peptide (stippled box), mature peptide (blacked box), and the transmembrane domain (cross-hatched box).

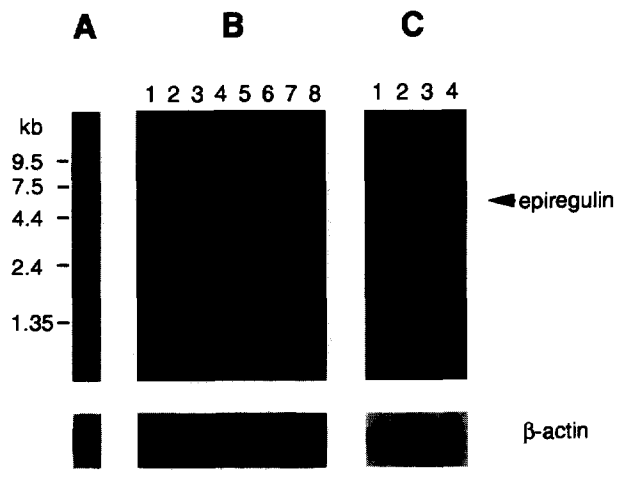

Fig. 4. Northern blot analysis of mouse epiregulin mRNA. Nylon membranes filters, onto which 2 μg of poly $(A)^+$ RNAs were transferred, were probed with 763-bp *Hha*I–*Hin*dIII fragment of epiregulin cDNA or β-actin. Mouse MTN blot filters were used in (B) and (C). (A) NIH3T3/clone T7. (B) Normal adult tissues. lane 1, heart; lane 2, brain; lane 3, spleen; lane 4, lung; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney; lane 8, testis. (C) Early embryos. lane 1, 7-day-old; lane 2, 11-day-old; lane 3, 15-day-old; lane 4, 17-day-old mouse embryos. The 7-day-old and 11-day-old embryos include the extra-embryonic membranes. The 15-day-old and 17-day-old embryos do not. The arrow head indicates the position of an approximately 4.8 kb epiregulin transcript.

sequences. Mature form of epiregulin, as well as those of amphiregulin [14,17], HB-EGF [15, 18], and betacellulin [16], are processed at different sequences of N- and C-termini. There are no consensus cleavage sites for the proteinase in the precursors of proteins that belong to the EGF family. Therefore, distinct proteinases or a proteinase with a broad specificity may process the precursor.

3.3. *Expression of epiregulin mRNA*

Northern blot analysis showed that an epiregulin transcript of approximately 4.8-kb was expressed at high levels in NIH3T3/clone T7 cells (Fig. 4A), while it was not detected in normal tissues of adult mice including the liver, lung, kidney, heart, brain, spleen, testis, smooth and skeletal muscles (Fig. 4B). This is in contrast to other members of the EGF family that are expressed ubiquitously in a variety of normal organs. This finding indicates that the expression profile of epiregulin is distinct from other members of the EGF family.

The expression of epiregulin during development of the mouse whole embryo is demonstrated in Fig. 4C. The expression of epiregulin was limited to 7-day-old embryo, and was very low or undetectable in 9- to 17-day-old embryos. Similar expression profiles during fetal development have been reported in those of TGF-α [19–22]. The highest expressions of mouse and rat TGF-α are observed on day 7 to 9 [19] and on day 8 to 9 [20], respectively, but diminish after that to undetectable levels.

Recently, mice homozygous for a disrupted EGF receptor gene have been generated [23–25]. The mutation resulted in peri-implantation, mid-gestational and postnatal mortality depending on the genetic background. Therefore, epiregulin may together with other EGF-related ligands play a role in the early stage of development. Further study for epiregulin localization

in the early development process will give us its functional significance.

Acknowledgements: We thank Dr. K. Hanada for his support to our research, and our colleagues in the Biotechnology Group of Taisho Pharmaceutical Co. Ltd. for their helpful discussions.

References

[1] Aaronson, S.A. (1991) Science 254, 1146–1152.
[2] Cross, M. and Dexter, M. (1991) Cell 64, 221–280.
[3] Toyoda, H., Komurasaki, T., Uchida, D., Takayama, Y., Isobe, T., Okuyama, T. and Hanada, K. (1995) J. Biol. Chem. 270, 7495–7500.
[4] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[5] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[6] Lagnado, C.A., Brown, C.Y. and Goodall, G.J. (1994) Mol. Cell. Biol. 14, 7984–7995.
[7] Zubiaga, A.M., Belasco, J.G. and Greenberg, M.E. (1995) Mol. Cell. Biol. 15, 2219–2230.
[8] von Heijne, G. (1983) Eur. J. Biochem. 13, 17–21.
[9] von Heijne, G. (1985) J. Mol. Biol. 184, 99–105.
[10] Gray, A., Dull, T.J. and Ullrich, A. (1983) Nature 303, 722–725.
[11] Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Sellby, M., Rutter, W.J. and Bell, G.I. (1983) Science 221, 236–240.
[12] Derynck, R., Roberts, A.B., Winkler, M.E., Chen, E.Y. and Goeddel, D.V. (1984) Cell 38, 287–297.
[13] Lee, D.C., Rose, T.M., Webb, N.R. and Todaro, G.J. (1985) Nature 313, 489–491.
[14] Plowman, G.D., Green, J.M., McDonald, Neubauer, M.G., Disteche, C.M., Todaro, G.J. and Shyoab, M. (1990) Mol. Cell. Biol. 10, 1969–1981.
[15] Higashiyama, S., Abraham, J.A., Miler, J., Fiddes, J.C. and Klagsbrun, M. (1991) Science 251, 936–939.
[16] Sing, Y., Christofori, G., Hanahan, D., Ono, Y., Sasada, R., Igarashi, K. and Folkman, J. (1993) 1604–1607.
[17] Johnson, G.R., Prigent, S.A., Gullick, W.J. and Stromberg, K. (1993) J. Biol Chem. 268, 18835–18843.
[18] Higashiyama, S., Lau, K., Besner, G.E., Abraham, J.A. and Klagsbrun, M. (1992) J. Biol Chem. 267, 6205–6212.
[19] Twardzik, D.R. (1985) Cancer Res. 45, 5413–5416.
[20] Lee, D.C., Rochford, R., Todaro, G.J. and Viillarreal, L.P. (1985) Mol. Cell. Biol. 5, 3644–3646.
[21] Han, V.K.M., Hunter III, E.S., Pratt, B.M., Zendegui, J.G. and Lee, D.C. (1987) Mol. Cell. Biol. 7, 2335–2343.
[22] Wilcox, J.N. and Derynck R. (1988) Mol. Cell. Biol. 8, 3415–3422.
[23] Thresdgill, D.W., Dlugosz, A.A., Hansen, L.A., Tennenbaum, T., lichti, U., Yee, D., LaMantia, C., Mourton, T., Herrup, K., Harris, R.C., Barnard, J.A., Yuspa, S.H., Coffey, R. and Maguson, T. (1995) Science 269, 230–234.
[24] Sibiia, M. and Wanger, E.F. (1995) Science 269, 234–238.
[25] Miettinen, P.J., Berger, J.E., Meneses, J., Phung, Y., Pedersen, R.A., Werb, Z. and Dery, R. (1995) Nature 376, 337341.
[26] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.