Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 8835–8839, November 1990
Cell Biology

Molecular cloning of the large subunit of transforming growth factor type β masking protein and expression of the mRNA in various rat tissues

(primary structure/binding protein/epidermal growth factor-like domain/cysteine-rich internal repeat)

TAKASHI TSUJI*, FUMIO OKADA†, KAZUO YAMAGUCHI†, AND TOSHIKAZU NAKAMURA*‡

*Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan; and †Tokyo Research Laboratories, Kyowa Hakko Kogyo Co., Ltd., Machida, Tokyo 194, Japan

Communicated by Takashi Sugimura, August 16, 1990 (received for review May 14, 1990)

ABSTRACT **Masking protein (MP), which neutralizes the activity of transforming growth factor type β1 (TGF-β1), is composed of a dimeric N-terminal part of a TGF-β1 precursor of M_r 39,000 and an unknown large subunit of M_r 105,000–120,000. The deduced primary structure of the MP large subunit was elucidated by determining the nucleotide sequence of its cDNA. The cDNA encodes a prepro-precursor of 1712 amino acid residues with a calculated M_r of 186,596. The mature large subunit seems to be derived proteolytically from a prepro-precursor and the calculated M_r is 91,606. The precursor has seven N-linked glycosylation sites and an unusual structure containing 18 epidermal growth factor-like domains and four cysteine-rich internal repeats. The large subunit mRNA is synthesized in parallel with the expression of TGF-β1 mRNA in various rat tissues.**

Transforming growth factor type β (TGF-β) is well recognized as a multifunctional regulator for cell growth and differentiation (1, 2). TGF-β was first identified by its ability to induce anchorage-independent growth of normal fibroblasts in soft agar, like malignant cells (3). TGF-β is a ubiquitous protein in mammals and is synthesized by various normal and transformed cells. TGF-β also is thought to be a potent regulator of wound healing, the immune response, and bone remodeling *in vivo*.

TGF-β1 is also known to be synthesized and secreted in a latent form from a wide variety of normal cells, including platelets (4–6). This latency is due to its formation of a complex with a high molecular weight binding protein, which has been named masking protein (MP) because of its function in reversibly masking TGF-β1 activity (6). Under physiological conditions, these components formed a high molecular weight complex of M_r ≈400,000 linked by noncovalent bonds. The latent TGF-β1 complex is activated *in vitro* by treatment with acid, urea, and heat (7, 8). MP may be important in regulation of the extracellular activity of TGF-β1 at local sites of various tissues during many biological phenomena.

Previously, we reported the purification and structural analysis of the latent TGF-β1 complex from rat platelets (7). MP is a glycoprotein of M_r 180,000, which is composed of one dimeric small subunit of M_r 39,000 and a single molecule of the large subunit of M_r 105,000–120,000 linked by disulfide bonds (7, 8). The N-terminal amino acid sequence of the MP small subunit was identical to the N-terminal part of the TGF-β precursor lacking a signal sequence (8, 9). Here we report the deduced primary structure of the MP large subunit from the nucleotide sequence of the cDNA and expression of the mRNA in various rat tissues.§

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Amino acid sequences of four peptide fragments derived from the MP large subunit of rat

Peptide	Sequence	Location in the MP large subunit
P-1	FSEQQR	898–903
P-2	GYSPTPDHRH	1059–1068
P-3	GFVPAGESSYETGGENYK	1397–1414
P-4	EGTYYDPVK	1445–1453

The numbers and positions of peptides are same as in Fig. 1.

MATERIALS AND METHODS

Materials. Materials used for the purification of a latent TGF-β1 were as described (7). The cDNA synthesis system, the cDNA cloning system, the multiprime labeling system, and the nylon membrane filter Hybond-N were from Amersham. Lambda ZAP II vector was purchased from Stratagene. Radioisotopic materials were from Amersham. Sequenase was from United States Biochemical.

Purification and Sequencing of the MP Large Subunit. The latent TGF-β1 complex was purified from rat platelets as described (7). To isolate the MP large subunit, the purified latent TGF-β1 complex was reduced and subjected to SDS/PAGE and then transferred to a poly(vinylidene difluoride) membrane (10). The isolated large subunit on the membrane was subjected to cyanogen bromide cleavage followed by lysyl endopeptidase digestion. Four peptide fragments of the MP large subunit were separated by reverse-phase HPLC with a gradient of 0–80% acetonitrile in 0.1% trifluoroacetic acid. Their amino acid sequences were determined with an Applied Biosystems model 470A gas-phase sequencer equipped with a 120A phenylthiohydantoin-derivatized amino acid analyzer.

Synthesis and Labeling of Oligonucleotides. Synthetic long oligodeoxynucleotide probes encoding the following peptide sequences were synthesized on an Applied Biosystems model 381A DNA synthesizer. These probes, containing inosine, are 5′-TTITAITTITCICCICCIGTITCITAIIIIIIIT-CICCIGCIGGIACIAAICC-3′ (peptide P-3) and 5′-T/CTTIACIGGA/GTCA/GTAA/GTAIGTICCT/CTC-3′ (peptide P-4). These synthetic oligonucleotide probes were radiolabeled by 5′-terminal phosphorylation using T4 polynucleotide kinase and [γ-^{32}P]ATP (11).

Construction and Screening of cDNA Libraries and DNA Sequencing. The cDNA was synthesized using oligo(dT)$_{12-18}$ as a primer from poly(A)$^+$ RNA of rat megakaryocytes, which were purified from bone marrow of the long bones of

Abbreviations: TGF-β1, transforming growth factor type β1; MP, masking protein; EGF, epidermal growth factor.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M55431).

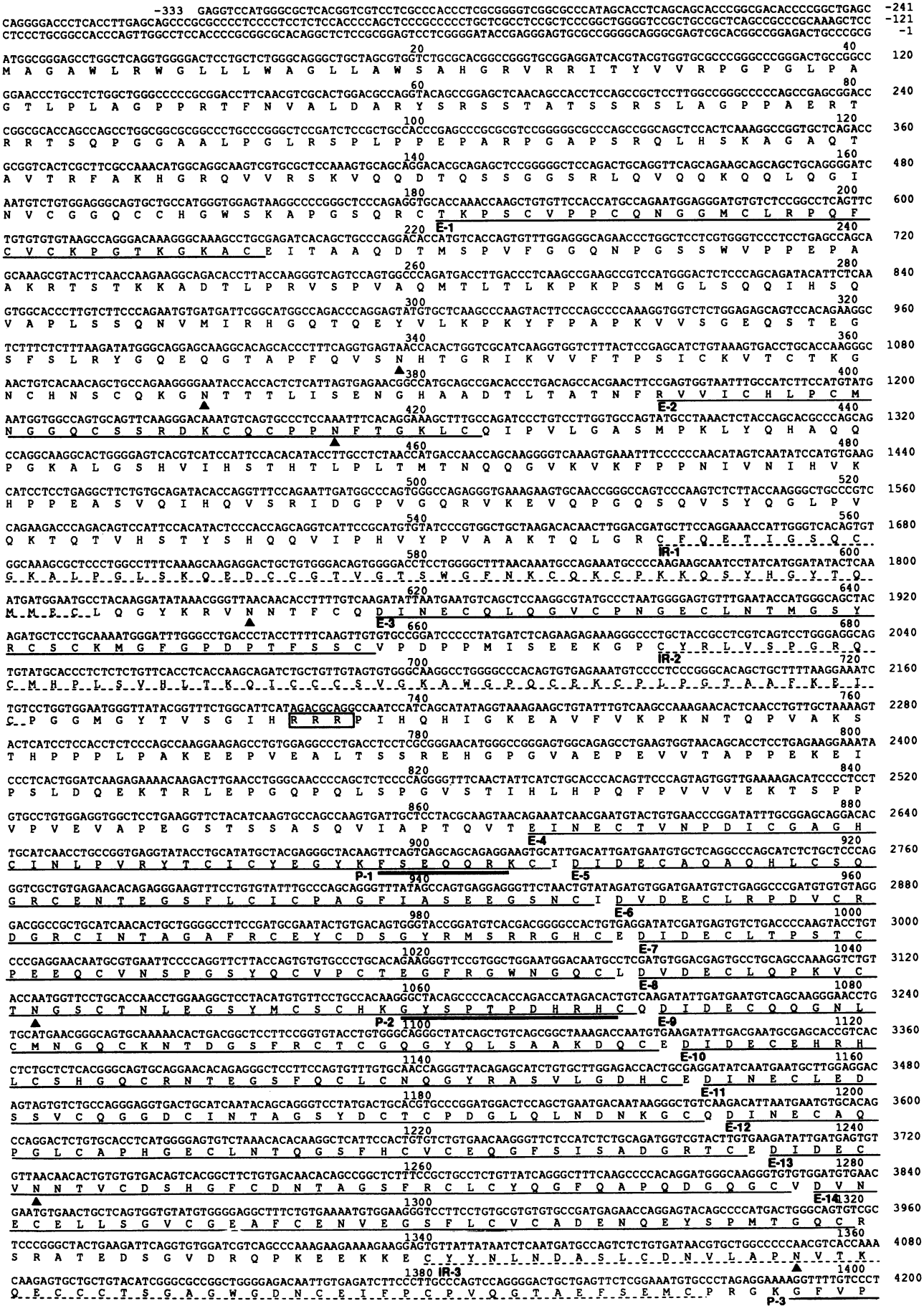

FIG. 1. (*Figure continues on the opposite page.*)

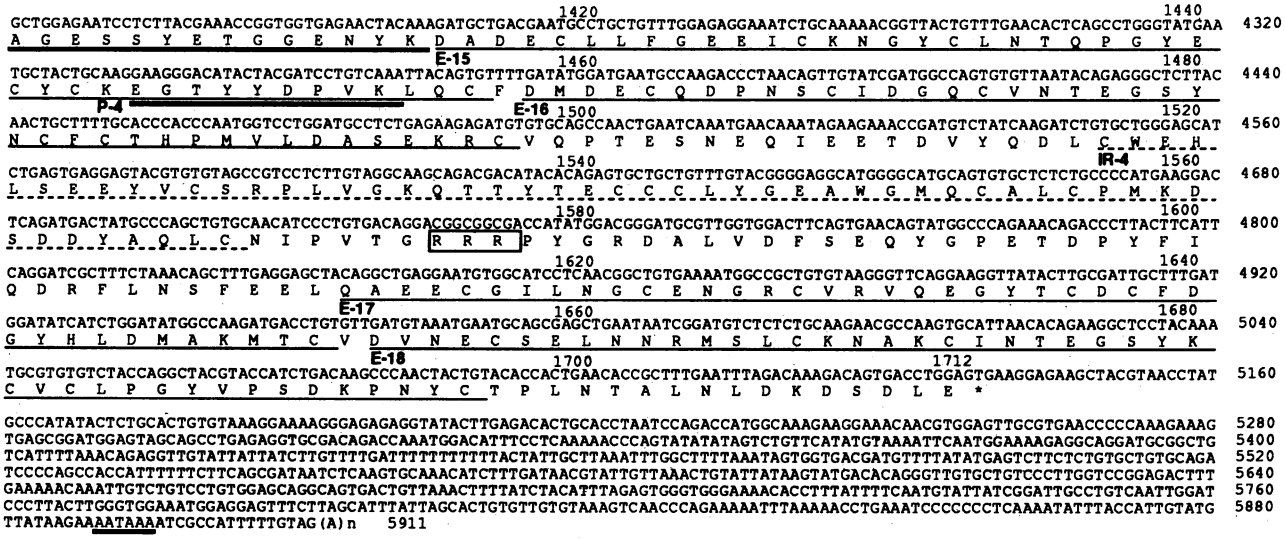

FIG. 1. Nucleotide sequence and deduced amino acid sequence of the MP large subunit cDNA of rat. Nucleotides are numbered on the right. The translated amino acid sequence is shown beneath the corresponding nucleotide sequence with the number above it. Chemically determined peptide sequences are underlined with a heavy bar. The putative processing sites are boxed. The EGF-like repeats are underlined. Dashed underlines indicate internal repeats. The potential sites for asparagine-linked glycosylation sites are indicated with arrowheads.

500 rats by density-gradient centrifugation (12). The cDNA library was constructed in the Lambda ZAP II vector. Plaques of the library were screened with synthetic oligonucleotide and conditions for hybridization have been described (13). We obtained five positive clones and the largest cDNA insert was 5 kilobases (kb), designated as RM5. The rat kidney cDNA library was constructed from rat kidney poly(A)$^+$ RNA using an oligo(dT)$_{12-18}$ and a random hexamer as primers in Lambda ZAP II as before. Recombinants (5×10^5 plaque-forming units) were screened with the ^{32}P-labeled 5-kb cDNA insert of RM5. Positive clones were plaque-purified and were allowed *in vivo* excision and recircularization of pBluescript from Lambda ZAP II. The cDNA inserts of these clones were sized and characterized. Three clones were selected and sequenced by generating progressive deletions and by the dideoxynucleotide chain-termination method (14) with Sequenase.

Northern Blot Analysis. Poly(A)$^+$ RNA (2 μg per lane) was electrophoresed into a 1% agarose gel containing 0.66 M formaldehyde (15) and blotted onto a Hybond-N nylon membrane filter. The filter was hybridized with the ^{32}P-labeled probes. The filter was exposed on x-ray film for 16 hr at −80°C using an intensifying screen.

RESULTS AND DISCUSSION

cDNA Cloning and Nucleotide Sequence of the MP Large Subunit. The latent TGF-β1 complex was purified to homogeneity from rat platelets as described (7). The yield was 1.4 mg from platelets of 2500 rats, corresponding to a recovery of 8%. To determine the partial amino acid sequence of the MP large subunit, the large subunit was isolated from the purified latent TGF-β1 complex by SDS/PAGE and was cleaved with cyanogen bromide followed by lysyl endopeptidase. We isolated four peptide fragments of the MP large subunit after digestion and determined the amino acid sequences of the peptides (Table 1). These amino acid sequences had no homology to any known proteins, suggesting that the MP large subunit is a previously unknown protein.

Amino acid sequences of peptides P-3 and P-4 were used to design synthetic oligonucleotide probes. Five positive clones were isolated from a cDNA library (2.5×10^5 plaque-forming units) constructed with poly(A)$^+$ RNA of rat megakaryocytes and one of them had the 5-kb insert (clone RM5). RM5 was sequenced and predicted the translation of a 1400-amino acid protein. All four partial amino acid sequences determined chemically were encoded by the 5-kb cDNA (RM5). Northern blot analysis showed that the large subunit mRNA was detectable as two species of about 5.3 and 6.2 kb in various rat tissues (see Fig. 5). Next we attempted to isolate the full-sized cDNA for the 6.2-kb mRNA, because the major population of the large subunit mRNA is 6.2 kb and the human large subunit mRNA detected in placenta was only 6.2 kb. Northern hybridization also showed that rat kidney contains a relatively high level of the

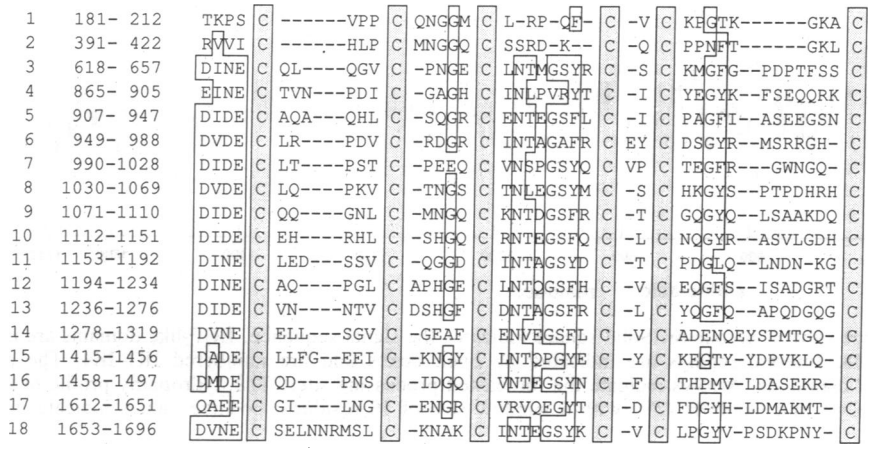

FIG. 2. Alignment of the amino acid sequence of the EGF-like domains in the MP large subunit precursor. Amino acid residues are indicated by the single letter code. Dashes indicate a gap introduced for maximal alignment. Residues are boxed when a majority of the residues are the same or substitutions are conservative in equivalent positions. The cysteine residues are shaded. The number of residues is the same as in Fig. 1.

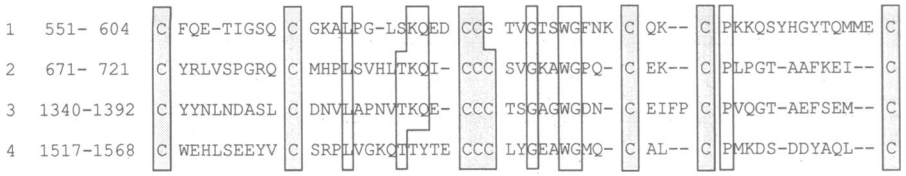

FIG. 3. Alignment of the amino acid sequence of the cysteine-rich internal repeats in the MP large subunit precursor. Dashes, boxes, and shading are the same as in Fig. 2.

large subunit mRNA (see Fig. 5). Therefore, to obtain the full-length cDNA containing an entire open reading frame for the MP large subunit, oligo(dT)$_{12-18}$ or a random-primed cDNA library using kidney poly(A)$^+$ RNA was constructed and screened with the ^{32}P-labeled cDNA insert of RM5 as a probe. Eleven positive clones were obtained and characterized in detail. These overlapping clones cover 6244 nucleotides containing the entire coding region. The complete cDNA and translated protein sequence of the MP large subunit of rat are shown in Fig. 1. The nucleotide sequence surrounding the initiation codon agrees reasonably well with the consensus sequence (A/GCCATGG) (16). The sequence comprises a 5′ untranslated region of 333 nucleotides, a single open reading frame of 5136 nucleotides, and a 3′ untranslated region of 775 nucleotides. The polyadenylation signal AATAAA (positions 6223–6228) is found 16 nucleotides upstream of the poly(A)$^+$ tract.

Deduced Primary Structure of the MP Large Subunit. The large subunit cDNA encodes a polypeptide of 1712 amino acid residues. The initiation methionine is followed by the stretch of 20 amino acid residues that are hydrophobic and closely resemble a signal sequence. The putative signal peptidase cleavage site exists between the serine and alanine amino acid positions 20 and 21, judging from the formulating rules for prediction of the cleavage site (17). The big difference between the apparent M_r of 105,000–120,000 of the MP large subunit determined by SDS/PAGE and the calculated M_r of 186,596 could be accounted for by posttranslational proteolytic processing. The N-terminal amino acid sequence of the MP large subunit was not determined by protein microsequencing. However, the large subunit precursor contains two pairs of Arg-Arg-Arg residues at both the N-terminal side (positions 734–736) and the C-terminal side (positions 1575–1577), which could also undergo posttranslational cleavage and give rise to the mature large subunit having 838 amino acid residues. Also, determination of the hydropathy profile predicts that these basic residues are located in a hydrophilic region, which would make it accessible to trypsin-like proteases (see Fig. 4). The calculated M_r of the mature large subunit is 91,606. The precursor contains seven potential N-glycosylation sites and three of them are located in the mature large subunit. Thus, the differences between the molecular weight determined by SDS/PAGE and that calculated from the deduced amino acid sequence may be due to the glycosylation of the mature large subunit. This is also supported by a finding that the size obtained after chemical deglycosylation is close to that predicted from the cDNA sequence (data not shown).

The large subunit precursor contains two types of cysteine-rich repeats. One is the epidermal growth factor (EGF)-like domain with six cysteines, and another is the internal repeat with eight cysteines. An EGF-like domain has been found in many other proteins; cell-surface receptors for low density lipoprotein (18) and insulin-like growth factor II (19), coagulation factors (20), thrombospondin (21), extracellular matrix proteins such as laminin (22) and tenascin (23), and *Drosophila* notch (24). The precursor contains 18 EGF-like domains, which are composed of 32–44 amino acid residues and 13 of them are localized in the mature large subunit. The alignment of these domains is shown in Fig. 2. EGF-like domains 1 and 2 are slightly different from the others. In addition, the EGF-like domains 2, 5, and 14 are distinguished from the others by having a remarkable hydrophobic region, whereas the others are hydrophilic (see Fig. 4). The precursor also contains four internal repeats comprising 51–54 amino acid residues and two of the repeats are localized in the mature large subunit. The alignment of these repeats is shown in Fig. 3. These repeats contain eight cysteine residues, except for internal repeat 1 where there are three successive cysteines. These internal repeats represent an additional type of cysteine-rich repeat. The internal repeats are predicted to form a three-dimensional structure, such as the EGF-like domain, by disulfide bonds. However, there are no data on the location of the disulfide bridges. A cysteine residue in these internal repeats seems to participate in the disulfide bridges on the N-terminal part of the TGF-β precursor (small dimeric subunit) because no free cysteine residue exists in the region other than the cysteine-rich repeats and all cysteines in the EGF-like

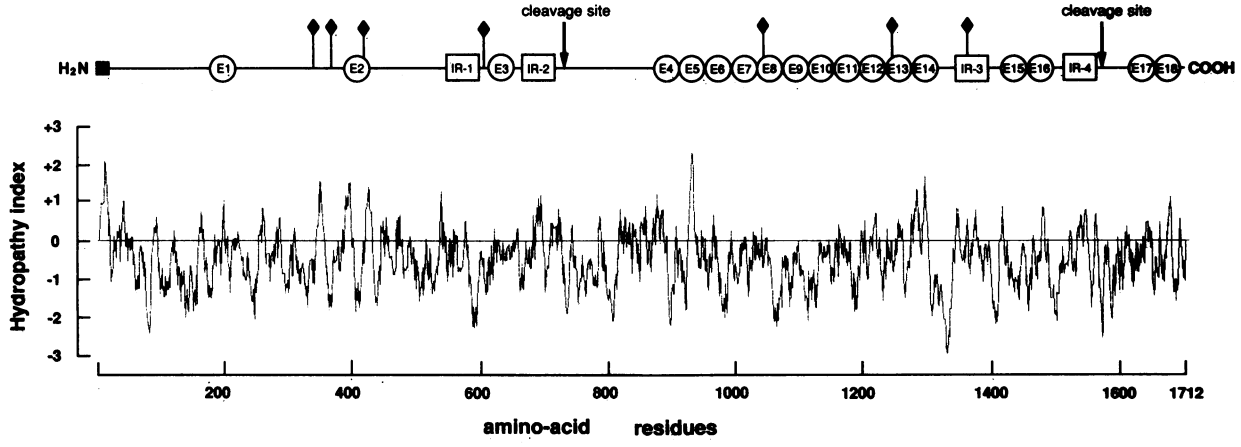

FIG. 4. Putative structure of the MP large subunit precursor. Solid square indicates the signal sequence. EGF-like domains are indicated by open circles and are numbered E1–E18. Internal repeats are indicated by open boxes and are numbered IR1–IR4. The potential N-glycosylation sites are marked with solid diamonds. Arrows indicate the forecasted processing sites. The hydropathy profile of the large subunit precursor is from the method of Kyte and Doolittle (25), with a window size of 10 residues; positive values indicate increasing hydrophobicity.

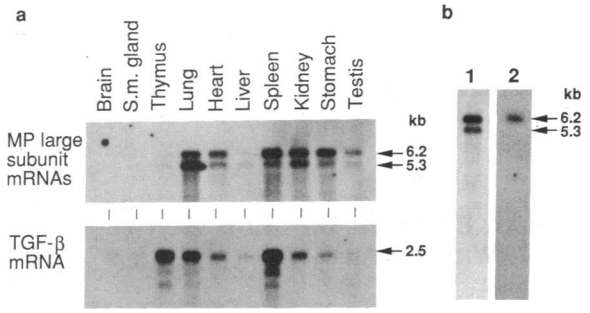

FIG. 5. Northern blot analysis of the MP large subunit and TGF-β1 mRNA. (*a*) Expression of the large subunit and TGF-β1 mRNAs in various normal rat tissues. The following tissue sources of poly(A)$^+$ RNA were used: brain, submandibular gland (S.m. gland), thymus, lung, heart, liver, spleen, kidney, stomach, and testis. Arrows indicate sizes of the major transcripts for the MP large subunit (6.2 and 5.3 kb) and TGF-β (2.5 kb). The large subunit mRNAs were detected by the cDNA insert of the large subunit clone RM5. TGF-β1 mRNA was detected by the *Eco*RI/*Bgl* I cDNA fragment of rat TGF-β1. (*b*) Characterization of two species of large subunit mRNAs. Poly(A)$^+$ RNA from kidney was analyzed with the cDNA insert of the large subunit clone RM5 (lane 1) and the *Eco*RI/*Sac* I 519-base 5′-terminal cDNA fragment of the MP large subunit (lane 2).

domains are expected to be involved in the intrachain disulfide linkage. The predicted structure of the large subunit precursor is shown with a hydropathy profile in Fig. 4.

Expression of the MP Large Subunit mRNA in Rat Tissues. We next examined the tissue distribution of the MP large subunit mRNA in rat (Fig. 5). The large subunit mRNA, which was represented by two species of 5.3 and 6.2 kb, was detectable in various tissues. The 6.2-kb species had a higher relative intensity than the 5.3-kb species in rat tissues except for lung. The 6.2-kb species was only detectable when a 0.5-kb cDNA fragment from 5′ termini was used as a probe for Northern hybridization (Fig. 5*b*). Therefore, the difference in size is attributed to the missing 5′ sequence. There is a good correlation between the relative mRNA levels of the large subunit and TGF-β1 in various tissues except for thymus. This result suggests that the large subunit mRNA is synthesized in parallel with the expression of TGF-β1 mRNA in rat tissues. On the contrary, an exceptional case was found in the thymus. This suggests that, in thymus, the regulation of TGF-β1 activity may not be due to MP.

The latent TGF-β1 complex may be activated by dissociation of the complex to the active TGF-β1, wherever and whenever TGF-β1 should act during many physiological phenomena. It is still unknown, however, how this latent TGF-β1 is activated *in vivo*. One possibility is that a protease or specific activation factor that is produced at a local site may lead to release of TGF-β1 from the latent complex through limited proteolysis of the MP components or by stimulating dissociation. However, neither such protease nor activation factor has been found. Another possibility is that TGF-β1 is released when the MP components interact with the cell-surface protein or an extracellular matrix. In fact, some studies in which a coculture technique was used suggested that the activation mechanisms of the latent TGF-β complex are necessary for direct interaction between the complex and a cell-surface molecule (26).

A recent report describing the expression technique of recombinant TGF-β1 suggested that the N-terminal part of the TGF-β1 precursor, a small dimeric subunit of MP, is mainly involved in latency of TGF-β1 (27). On the other hand, more recently, we examined whether the MP large subunit purified from human platelets has the binding and masking activities of TGF-β1. However, it was found to have no activity for binding and masking of TGF-β1. From these findings, the MP large subunit may play an important role in the activation mechanisms of the latent TGF-β1 rather than in association with TGF-β1. It is interesting to speculate that the large subunit may interact directly with cell-surface or matrix proteins to cause a conformational change of MP, resulting in the dissociation of the complex so TGF-β1 can exert its effects at a local site. We think that EGF-like domains and internal repeats in the MP large subunit may act as a functional domain of the interactions to release TGF-β1 from the latent complex. Further studies on the elucidation of the role of those structures in MP components in the activation mechanisms of the latent TGF-β complex *in vivo* should provide very valuable information on liver regeneration, wound healing, inflammation, bone remodeling, and carcinogenesis.

Note Added in Proof. After submission of this work, a paper appeared (28) showing the molecular cloning of the cDNA, which is derived from 5.3-kb mRNA, encoding the large subunit of TGF-β masking protein. The deduced amino acid sequence is truncated 336 amino acid residues of the N-terminal part of our sequence (6.2-kb mRNA) and has a different signal sequence.

We thank Dr. Kosuke Tashiro of our laboratory for helpful suggestions during this study and Dr. Takashi Miyata of our department for amino acid sequence alignment. This work was supported by the Ministry of Education and Culture of Japan.

1. Sporn, M. B., Roberts, A. B., Wakefield, L. M. & de Crombrugghe, B. (1987) *J. Cell Biol.* **105,** 1039–1045.
2. Massagué, J. (1987) *Cell* **49,** 437–438.
3. Roberts, A. B., Anzano, M. A., Lamb, L. C., Smith, J. M. & Sporn, M. B. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5339–5343.
4. Pircher, R., Lawrence, D. A. & Jullien, P. (1984) *Cancer Res.* **44,** 5538–5543.
5. Pircher, R., Jullien, P. & Lawrence, D. A. (1986) *Biochem. Biophys. Res. Commun.* **136,** 30–37.
6. Nakamura, T., Kitazawa, T. & Ichihara, A. (1986) *Biochem. Biophys. Res. Commun.* **141,** 176–184.
7. Okada, F., Yamaguchi, K., Ichihara, A. & Nakamura, T. (1989) *J. Biochem.* **106,** 304–310.
8. Miyazono, K., Hellman, U., Wernstedt, C. & Heldin, C.-H. (1988) *J. Biol. Chem.* **263,** 6407–6415.
9. Okada, F., Yamaguchi, K., Ichihara, A. & Nakamura, T. (1989) *FEBS Lett.* **242,** 240–244.
10. Matsudaira, P. (1987) *J. Biol. Chem.* **262,** 10035–10038.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 125–126.
12. Berkow, R. L., Straneva, J. E., Bruno, E. D., Beyer, G. S., Burgess, J. S. & Hoffman, R. (1984) *J. Lab. Clin. Med.* **103,** 811–818.
13. Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K. & Shimizu, S. (1989) *Nature (London)* **342,** 440–443.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 202–203.
16. Kozak, M. (1989) *J. Cell Biol.* **108,** 229–241.
17. Heijne, G. V. (1983) *Eur. J. Biochem.* **133,** 17–21.
18. Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L. & Russell, D. W. (1984) *Cell* **39,** 27–38.
19. Morgan, D. O., Edman, J. C., Standring, D. N., Fried, V. A., Smith, M. C., Roth, R. A. & Rutter, W. J. (1987) *Nature (London)* **329,** 301–307.
20. Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W. & Kurachi, K. (1985) *Biochemistry* **24,** 3736–3750.
21. Lawler, J. & Hynes, R. O. (1986) *J. Cell Biol.* **103,** 1635–1648.
22. Sasaki, M., Kleinman, H. K., Huber, H., Deutzmann, R. & Yamada, Y. (1988) *J. Biol. Chem.* **263,** 16536–16544.
23. Pearson, C. A., Pearson, D., Shibahara, S., Hofsteenge, J. & Chiquet-Ehrismann, R. (1988) *EMBO J.* **7,** 2977–2981.
24. Wharton, K. A., Johansen, K. M., Xu, T. & Artavanis-Tsakonas, S. (1985) *Cell* **43,** 567–581.
25. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
26. Oreffo, R. O. C., Mundy, G. R., Seyedin, S. M. & Bonewald, L. F. (1989) *Biochem. Biophys. Res. Commun.* **158,** 817–823.
27. Gentry, L. E., Webb, N. R., Lim, G. J., Brunner, A. M., Ranchalis, J. E., Twardzik, D. R., Lioubin, M. N., Marquardt, H. & Purchio, A. F. (1987) *Mol. Cell. Biol.* **7,** 3418–3427.
28. Kanzaki, T., Olofsson, A., Moren, A., Wernstedt, C., Hellman, U., Miyazono, K., Claesson-Welsh, L. & Heldin, C.-H. (1990) *Cell* **61,** 1051–1061.