Human transforming growth factor-β complementary DNA sequence and expression in normal and transformed cells

Rik Derynck, Julie A. Jarrett, Ellson Y. Chen, Dennis H. Eaton, John R. Bell*, Richard K. Assoian†, Anita B. Roberts†, Michael B. Sporn† & David V. Goeddel

Departments of Molecular Biology and * Protein Biochemistry, Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

† Laboratory of Chemoprevention, National Cancer Institute, Bethesda, Maryland 20205, USA

The partial amino-acid sequence of purified human transforming growth factor-β (TGF-β) was used to identify a series of cDNA clones encoding the protein. The cDNA sequence indicates that the 112-amino acid monomeric form of the natural TGF-β homodimer is derived proteolytically from a much longer precursor polypeptide which may be secreted. TGF-β messenger RNA is synthesized in various normal and transformed cells.

Two different peptides which can induce the reversible phenotypic transformation of mammalian cells in culture are termed transforming growth factor (TGF)[1] α and β. TGF-α is synthesized by various transformed cell lines[2,3] and competes with epidermal growth factor (EGF) for binding to the same cell-surface receptor[2]. A 50-amino-acid-long TGF-α species has been purified and shown to share sequence homology with EGF[4]. This TGF-α is initially synthesized as part of a 160-amino-acid-long precursor molecule which undergoes NH_2- and C-terminal proteolytic processing to yield the mature peptide[5,6].

TGF-β has been isolated from tumour and normal cells and tissues[7], including kidney[8], placenta[9] and blood platelets[10,11]. The relatively high levels of TGF-β present in platelets, which also contain platelet-derived growth factor (PDGF), have led to the suggestion that TGF-β plays a role in wound healing[12]. It is a homodimer of relative molecular mass (M_r) ~25,000 (ref. 11) which recognizes a cell surface receptor distinct from the EGF receptor[13,14]. Treatment of NRK fibroblasts with TGF-β does, however, result in an increase in the number of membrane receptors for EGF[15], which may explain the ability of TGF-β to greatly potentiate transforming activity of EGF and TGF-α on these cells[7]. Moreover, TGF-β alone can induce AKR-2B fibroblasts to form colonies in soft agar[16]. In addition to stimulating cell proliferation, TGF-β has recently been demonstrated to inhibit the anchorage-dependent growth of a variety of human cancer cell lines[17] TGF-β may be identical or very similar to a growth inhibitor isolated from African green monkey (BSC-1) cells[18] and has many properties in common with the tumour inhibitory factor TIF-1 (ref. 19). Whether TGF-β acts to stimulate or inhibit the growth of a particular cell type seems to depend on many variables, such as the physiological condition of the cell and the presence of additional growth factors.

Here we report the amino-acid sequence of human TGF-β determined from direct protein sequencing and cDNA cloning. The nucleotide sequence of the TGF-β cDNA reveals that the 112-amino-acid-long TGF-β monomer is initially synthesized as part of a precursor polypeptide. Northern hybridization analysis demonstrates that the TGF-β mRNA is present in normal cells and in many types of tumour cells.

Purification and sequence analysis

A purification scheme has been developed by Assoian *et al.*[11] which yields TGF-β from human blood platelets, a major site of TGF-β storage. This procedure was modified and used to obtain human TGF-β with a purity of >95% for subsequent sequence analysis (Fig. 1 legend). In agreement with previous work[11] the non-reduced TGF-β migrated as an M_r 25,000 protein in a SDS polyacrylamide gel, whereas reduction with β-mercaptoethanol converted it into an M_r 12,500 species. This suggests that TGF-β consists of two M_r 12,000 polypeptide chains linked by intermolecular disulphide bridges. The purified TGF-β was reduced, alkylated and subjected to amino-terminal sequence analysis. The amino-acid sequences of several peptides obtained after clostripain digestion of TGF-β were also determined. These analyses yielded a 60-residue long contiguous sequence starting from the amino terminus (Fig. 1*a*) as well as several shorter peptide sequences (data not shown). The NH_2-terminus of the human TGF-β is identical to the previously determined NH_2-terminal sequence of bovine TGF-β[8]. The fact that no heterogeneity was seen during the NH_2-terminal sequence analysis, and that all TGF-β sequences obtained could be accounted for by an M_r 12,500 polypeptide, provide additional evidence for the homodimeric nature of this protein.

Isolation of TGF-β cDNAs

The approach we followed for the identification of the nucleotide sequence encoding TGF-β was similar to the strategy used previously for TGF-α[5]. Long oligonucleotides designed on the basis of the partial protein sequence were used as hybridization probes for the identification of a TGF-β exon in a human genomic DNA library, which was then used as a probe for the isolation of TGF-β cDNAs.

Two 44-mer deoxyoligonucleotides, βLP1 and βLP2, complementary to sequences encoding amino acids 3–17 and 30–44, respectively, were chemically synthesized (Fig. 1*a*). The choice of these nucleotide sequences was based on previously described considerations[5]. In addition, 16 14-mers were synthesized which are complementary to all possible codons for amino acids 13–16 (Fig. 1*a*). A human genomic DNA library[20] was screened under low stringency hybridization conditions using ^{32}P-labelled βLP1 as probe. One of the 58 hybridizing phage, βλ58, also hybridized with the 44-mer βLP2 and with the pool of 14-mers. Nucleotide sequence analysis of this clone revealed the presence of a TGF-β exon encoding amino acids 10–60 of TGF-β. Only about half of the sequence corresponding to the βBP1 oligonucleotide is present. Nevertheless, the probe hybridized because of the correct assignment of 23 of the 24 nucleotides in the portion of the probe corresponding to the exon sequence (data not shown).

To obtain the entire TGF-β coding sequence, this exon was used as a probe to screen a λgt10-based cDNA library derived from human term placenta mRNA. The screening of ~750,000 oligo(dT) primed cDNA clones resulted in the isolation of one TGF-β cDNA (λβC1) of ~1,050 base pairs (bp). The previously determined partial TGF-β sequence established the reading frame and revealed the sequence encoding the complete TGF-β polypeptide, followed by a stop codon 20 bp from the 3′ end of the cDNA.

Northern hybridization[21,22] using the λβC1 cDNA insert as probe showed that TGF-β mRNA of ~2.5 kilobases (kb) is present in the A172 glioblastoma and HT1080 fibrosarcoma cell

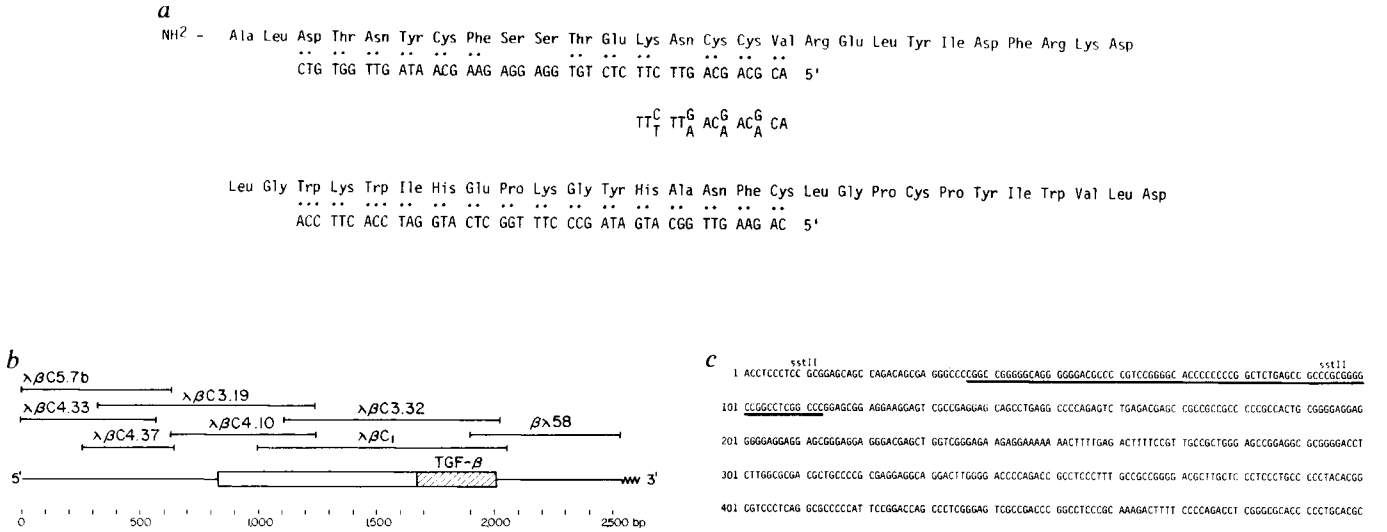

Fig. 1 Nucleotide sequence and deduced amino-acid sequence of TGF-β cDNA. *a*, Peptide sequence obtained from sequencing of the NH_2-terminus and of clostripain digestion products of TGF-β. The deoxyoligonucleotides which were used as hybridization probes for the detection of the human TGF-β cDNA are shown and are complementary to the coding strand. The nucleotides marked with a dot are residues without codon ambiguities. *b*, Schematic diagram of the TGF-β mRNA with the coding sequence boxed. The 112 amino-acid-long TGF-β (dashed) is encoded by the 3′ end of the coding sequence. The sequenced cDNA inserts of $\lambda\beta$C1, 3.19, 3.32, 4.10, 4.33, 4.37 and 5.7 bases and the genomic DNA sequence for the 3′ untranslated region are aligned above the diagram. *c*, Nucleotide sequence and deduced amino-acid sequence of the TGF-β cDNA, determined from the several overlapping cDNAs and the genomic 3′ region. The 5′ terminal region which could be folded into stable hairpin loops is underlined. TGF-β cDNA encodes a 391-amino-acid-long precursor, of which the C-terminal 112 amino acids (boxed) encode the mature TGF-β. The sequence of hydrophobic amino acids which could represent the hydrophobic core of an NH_2-terminal signal peptide is overlined. An overlined Arg-Arg dipeptide precedes the proteolytic cleavage site. The three potential *N*-glycosylation sites in the TGF-β precursor are overlined. The stop codon is followed by the underlined G-C rich sequence and a downstream TATA-like sequence. The AATAAA polyadenylation signal and TTCAGGCC sequence preceding the putative polyadenylation site are also underlined. Several restriction endonuclease recognition sites are indicated.

Methods. TGF-β was purified from human platelets essentially as described previously[11] with an additional chromatography step over a C18 (Synchropak) HPLC column in 0.1% trifluoroacetic acid eluted with a 20–50% acetontrile gradient. TGF-β was dialysed into 8 M urea and reduced by incubation in 0.1 M Tris-HCl (*p*H 8.5), 10 mM dithiothreitol, 8 M urea. Subsequent alkylation took place in the presence of 50 mM iodoacetate at room temperature in the dark. A fraction of the reduced and alkylated TGF-β was digested with clostripain[41] and the peptides generated were purified by reverse-phase HPLC chromatography. Sequence determination took place using either an extensively modified Beckman 890C spinning cup sequencer[42] or a vapour phase sequencer as described by Hewick *et al.*[43]. (Applied Biosystems model 470A), with amino-acid derivative identification by reversed-phase HPLC on a Rainin Microsorb C-8 column. The oligonucleotides were synthesized using the phosphotriester[44] or the phosphite[45] method. The cDNA was prepared[46] from polyadenylated mRNA by priming with poly(dT)$_{12-18}$ or the deoxyoligonucleotide ACACGGGTT-CAGGTAC (complementary to nucleotides 1,271–1,287). The double-stranded cDNA was treated with nuclease S_1 (Miles Laboratories) followed by *E. coli* DNA polymerase I Klenow fragment (Boehringer Mannheim) and subcloned into *Eco*RI-cleaved λgt10 as described[47], except that asymmetric *Eco*RI linkers[48] were used, thus avoiding the need for the *Eco*RI methylase treatment. The recombinant phase were plated on *E. coli* C600 Hfl (ref. 47) and replica plated onto nitrocellulose filters[49]. These were hybridized with ^{32}P-labelled[50] specific restriction fragments at 42°C in 50% formamide, 5×SSC, 50 mM sodium phosphate *p*H 6.8, 0.1% sodium pyrophosphate, 5×Denhardt's, 50 μg ml^{-1} salmon-sperm DNA and washed in 0.2×SSC, 0.1% SDS at the same temperature. Low-stringency hybridization conditions[51] were used for ^{32}P-labelled synthetic deoxyoligonucleotides. The nucleotide sequence of the TGF-β cDNA restriction fragments was determined by the dideoxyoligonucleotide chain termination method[52] after subcloning into M13 phage derivatives[53]. λBC1 was isolated from a human placenta cDNA library using the genomic exon as a probe. The $\lambda\beta$C1 *Eco*RI cDNA insert was used in turn to screen the A172 glioblastoma cDNA library (kindly provided by A. Ullrich) leading to the isolation of $\lambda\beta$C3.19. Screening of a specifically primed HT1080 fibrosarcoma cDNA library with the ^{32}P-labelled *Kpn*I/*Kpn*I and the upstream *Eco*RI/*Kpn*I fragment of the $\lambda\beta$C3.19 cDNA insert yielded $\lambda\beta$C4.10, 4.33 and 4.37. Another similar library was screened with the $\lambda\beta$C4.33 insert and a synthetic 40-mer corresponding to nucleotides 1–40, leading to the isolation of $\lambda\beta$C5.7b.

lines. Extensive screening of cDNA libraries prepared from both these cell lines and of an additional placenta cDNA library yielded many different TGF-β cDNAs. Most of these seemed to contain sequence inversions generated during cDNA preparation ('Belgian' clones[23]). In addition, single base deletions and substitutions were found in some of the cDNAs. Because of the high incidence of cDNA artifacts, presumably resulting from specific features in the mRNA or cDNA structure, all parts of the sequence were confirmed by sequencing at least two different cDNAs. Alignment of different cDNA sequences (Fig. 1*b*) enabled us to establish 2 kb of sequence preceding the TGF-β termination codon (Fig. 1*c*).

None of the 70 TGF-β cDNAs isolated from different oligo(dT)-primed cDNA libraries contained more than a few nucleotides of 3′ untranslated sequences. As introns are rarely found in 3′ noncoding regions, the 3′ untranslated sequence was determined using cloned genomic DNA. DNA sequence analysis of phage βλ58 revealed the presence of an exon encoding the carboxy-terminal part of TGF-β, followed by the stop codon and the 3′ untranslated region (Fig. 1*c*). An AATAAA hexanucleotide sequence[24] was encountered 500 bp downstream from the termination codon, thus permitting an assignment of the putative polyadenylation site. The calculated size of TGF-β mRNA is in close agreement with the 2.5 kb length determined from the Northern hybridization experiments (Fig. 2).

Carboxy-terminus of TGF-β

The cDNA sequence establishes that TGF-β is synthesized as the C-terminal segment of a precursor polypeptide. As several proteins undergo post-translational removal of C-terminal sequences, we determined the actual C-terminal amino-acid sequence of TGF-β. The cDNA sequence predicted the presence of a single methionine near the C-terminus in TGF-β. The protein was reduced, alkylated and cleaved with cyanogen bromide (CNBr). Sequence analysis of the mixture of peptides resulting from the CNBr digestion revealed the C-terminal sequence predicted from the cDNA (data not shown), although it cannot be excluded that a fraction of the molecule had undergone C-terminal proteolysis.

Unmodified TGF-β was also treated with CNBr. Cleavage at the methionine residue resulted in the complete loss of biological activity, showing that at least part of this C-terminal octapeptide is needed for biological activity (data not shown).

mRNA in different cell types

The availability of a TGF-β cDNA allowed us to screen various cells for TGF-β mRNA (Fig. 2). TGF-β mRNA was detectable in all human tumour cell lines including Wilms Tumour, A172 (glioblastoma) and the carcinoma cell lines T24 (bladder carcinoma), A431 (Squamous epidermoid carcinoma), MCF-7 (mammary carcinoma) and KB (nasopharyngeal carcinoma). HT1080, a fibrosarcoma-derived cell line, which we had chosen as a source of mRNA for the cDNA cloning, contained relatively high levels of TGF-β mRNA. TGF-β mRNA was not only present in cell lines derived from solid tumours of meso-, endo- and ectoblastic origin, but was also detectable in tumour cell lines of hematopoietic origin, for example Daudi (Burkitt lymphoma B lymphoblast), Raji (Burkitt lymphoma B lymphoblast) and Molt-4 (T-cell leukaemia). The presence of TGF-β mRNA is not restricted to tumour cells, as it is clearly detectable in placenta and peripheral blood lymphocyte (PBL) mRNA. Strikingly, the level of TGF-β mRNA is significantly elevated after mitogenic stimulation of PBLs. TGF-β mRNA was not detectable in human liver, yet was present in the HEP-G2 hepatoma cell line. In all cases, the TGF-β mRNA migrated as a species with an apparent length of ~2.5 kb.

Discussion

Analysis of several overlapping cDNAs and gene fragments has led to the determination of a continuous sequence of 2,439 bp corresponding to the TGF-β precursor mRNA. An initiator ATG is located 842 nucleotides from the 5′ end and establishes

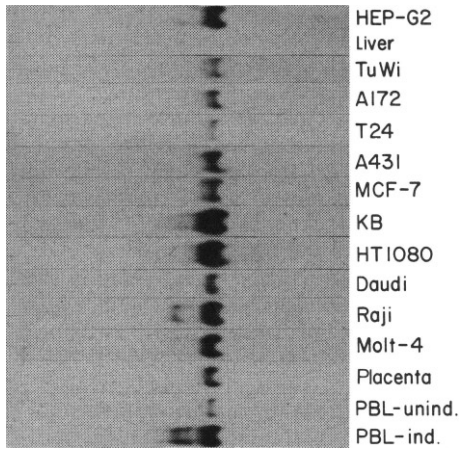

Fig. 2 Northern hybridization of mRNA from different sources with a TGF-β cDNA as probe. The cell source of the polyadenylated RNA is shown above each lane. The following cell lines were used: the hepatoma HEP-G2, Wilms tumour TuWi, glioblastoma A172, bladder carcinoma T24, squamous epidermoid carcinoma A431, mammary carcinoma MCF-7, nasopharyngeal carcinoma KB, fibrosarcoma HT1080, Burkitt lymphoma B lymphoblasts Daudi and Raji and T lymphoblast Molt-4. The cells growing in the monolayer were collected at confluency. The peripheral blood lymphocytes were prepared and mitogen-induced with staphylococcal enterotoxin B and phorbol myristate as described previously[54]. RNA was prepared 24 h after induction. Comparison with the position of the 28S and 18S ribosomal RNA on the gel suggests a length of ~2.5 kb for the TGF-β mRNA. 4 μg of polyadenylated mRNA was electrophoresed into a formaldehyde 1.2% agarose gel[21] and blotted onto nitrocellulose filters[22]. The ^{32}P-labelled[50] *Eco*RI cDNA insert of λβC1 was used as a probe under high stringency conditions (see Fig. 1 legend).

a coding sequence of 1,173 nucleotides, thus encoding a 391 amino-acid long polypeptide (Fig. 1*b*, *c*). Several areas in the cDNA sequence have an exceptionally high G-C content. In particular, the first 450 bp at the 5′-terminus has regions with >80% G-C content. The location of these G-C rich regions coincides with the areas in which the many cDNA cloning artifacts occurred and where partial length cDNAs were obtained. It is likely that the first ATG triplet located at position 842 (Fig. 1*c*) is the initiation codon as it is preceded in the same reading frame by several stop codons and it could initiate translation into a continuous sequence of a large polypeptide which comprises the TGF-β monomer. However, the sequence context at this ATG is not in good agreement with the proposed initiation codon consensus sequence G/ACCATGG[25]. Furthermore, this ATG is localized in a very G-C rich area. A second ATG, found in the same reading frame at position 954, is in much better agreement with the initiation codon consensus sequence and is localized within a 40-nucleotide region of relatively low G-C content (~50%). This ATG may be more accessible to the ribosome, especially as it is localized within a large G-C region. Although it is more likely that translation starts at the first ATG, it is conceivable that the second ATG could also be used to initiate translation of a smaller TGF-β precursor.

The 5′ untranslated region of the TGF-β mRNA is at least 841 nucleotides long and contains a 61-nucleotide-long sequence (positions 192–252) consisting almost exclusively of purines. The biological relevance of the exceptionally long 5′ untranslated region of high G-C content is unknown, but it is similar to the structural organization of the mRNAs for c-*myc*[26] and insulin-like growth factor-2 (ref. 27). However, there is no striking sequence homology between these sequences. The first 120 bp of the TGF-β cDNA can theoretically be folded into hairpin loop structures with a calculated stability of −92 kcal (data not shown). It is tempting to suggest that the long 5′ untranslated sequence and these hairpin loop structures could play a role in mRNA stability or in the regulation of transcription.

The stop codon at residue 2,016 is immediately followed by a remarkable, G-C-rich sequence of 75 nucleotides (Fig. 1*c*). This sequence consists of multiple repeats of CCGCC and ends with GGGGGC. The peculiar nature of this sequence is probably responsible for the fact that the 3′ untranslated end of the mRNA could not be cloned as a cDNA sequence, perhaps because of the inability of the *Escherichia coli* DNA polymerase I to use this sequence as a template for the second strand cDNA synthesis. Similar repeat sequences have been found in the promoter regions of the genes for human dihydrophfolate reductase[28], human adenosine deaminase[29] and Herpesvirus thymidine kinase[30]. In the latter case, McKnight *et al.*[30] have shown that these structural elements are important in transcription efficiency. In addition, it has been shown that the promoter specific transcriptional factor Sp1 binds to such sequences in the simian virus 40 early promoter region and in a related monkey promoter[31]. In all these cases the G-C rich repeats are followed closely by the Goldberg-Hogness TATA sequence. In the case of TGF-β, however, these sequences are located in the 3′ untranslated region of the gene, but, interestingly, are also followed by a TATA-like sequence. The hexanucleotide AATAAA, ~500 nucleotides downstream from the stop codon, probably functions as the TGF-β polyadenylation signal as this would agree with the size of TGF-β mRNA estimated from Northern hybridizations. Benoist *et al.*[32] have proposed a consensus sequence TTCACTGC which immediately precedes the poly(A) tail. A similar sequence, TTCAGgCC, follows the AATAAA sequence in the TGF-β mRNA, providing further support for the assignment of the polyadenylation site at position 2,539 (Fig. 1*c*).

The TGF-β cDNA encodes a polypeptide of 391 amino acids, assuming that the first ATG is used for translational initiation. Comparison with the experimentally determined NH_2-terminus of TGF-β shows that the C-terminal 112 amino acids constitute the TGF-β sequence. The TGF-β monomer is cleaved from the precursor at the Arg-Arg dipeptide immediately preceding the TGF-β NH_2-terminus. Similar proteolytic cleavage sites have been found in several other polypeptide precursor sequences[33,34]. Determination of the hydrophobicity profile[35] predicts that this Arg-Arg sequence is located within a hydrophilic region which would make it accessible to a trypsin-like protease. Whereas cleavage of the precursor gives rise to the mature TGF-β monomer, it is not known if other biologically active peptides are also released. The TGF-β precursor contains several pairs of basic residues (Fig. 1*c*) which could also undergo post-translational cleavage and give rise to separate polypeptide entities. However, the TGF-β polypeptide itself contains two Arg-Lys dipeptides which apparently are not cleaved. The TGF-β precursor contains three potential *N*-glycosylation sites[36], none of which are localized within the TGF-β polypeptide (Fig. 1*c*).

The 112 amino-acid sequence of TGF-β contains nine cysteines, whereas the rest of the precursor contains only three (positions 33, 224 and 226; Fig. 1*c*). There are no data either on the location of the disulphide bridges or on which cysteines are involved in the interchain disulphide linkage of the TGF-β dimer. Previous studies have shown that reduction of the TGF-β dimer of M_r 25,000 generates a single band of M_r 12,500 on denaturing polyacrylamide gel[11]. Sequence analysis of the TGF-β amino-terminus and of the TGF-β peptides obtained after clostripain digestion strongly suggest that the TGF-β dimer consists of two identical polypeptides. This homodimeric nature is also supported by the presence of only a single hybridizing DNA fragment on Southern hybridization of human genomic DNA with a TGF-β exon probe (data not shown). Chou-Fasman analysis of the secondary structure shows[37] that the TGF-β polypeptide contains extensive β-sheet with little α-helicity. The region immediately preceding the basic dipeptide cleavage site is probably α-helical.

TGF-β is stored in relatively large quantities in platelets and is also found in association with mammalian cells. It is also released into the culture medium of these cells and, therefore, is probably a secreted protein. The deduced polypeptide sequence of the 391-amino-acid-long TGF-β precursor shows a contiguous sequence of 16 hydrophobic residues (positions 8-23, Fig. 1*c*). This sequence may constitute the hydrophobic core[38] of an NH_2-terminal signal peptide involved in protein secretion. It is not known at which residue this cleavage by the signal peptidase could occur, but comparison with other known signal peptides shows an average signal peptide length of 20-30 amino acids[38]. The mature TGF-β would subsequently be released from the precursor as a result of the cleavage by a dibasic peptidase, localized at the external side of the membrane[39]. Translational initiation at the second ATG (position 954) would result in a 354-amino-acid-long precursor devoid of the putative signal peptide.

Most growth factors seem to be made only by specific cell types. PDGF is present in or released by platelets, endothelial cells, smooth muscle cells and various tumour cells. TGF-α is made by tumour cells of non-haematopoietic origin but has not been demonstrated to be synthesized by any normal adult cells. Previous studies have shown that TGF-β can be found in kidney[8], placenta[9] and platelets[10,11] and is released by several 'normal' and transformed fibroblast cell lines. We examined several types of normal and transformed cells from different embryonic origin for the synthesis of TGF-β mRNA. Our results show that all cell types except adult liver examined contain TGF-β mRNA. The amount of TGF-β mRNA seems to correlate with the degree of mitotic activity, which may explain our inability to detect TGF-β mRNA in adult liver tissue. Furthermore, TGF-β mRNA levels are higher in mitogen-induced lymphocytes than in unstimulated lymphocytes. This correlation also seems to hold for the higher levels of TGF-β expression in transformed cells compared with their non-transformed counterparts[40].

It is clear that much additional work is needed to elucidate the role of TGF-β during cell division and propagation and in the establishment and maintenance of the malignant phenotype. The knowledge of the structure of TGF-β and its precursor and the use of TGF-β cDNAs as molecular probes will certainly contribute to this understanding.

We thank Drs G. Nedwin, A. Ullrich, P. Seeburg and D. Capon for providing us with some cDNA libraries and mRNA samples, L. Rhee for help in DNA sequencing, L. Dart, N. Roche, D. Smith and J. Smith for assistance in the preparation of TGF-β and the Genentech DNA synthesis group for preparation of oligonucleotides. We also thank J. Arch for preparation of the manuscript.

Received 9 April; accepted 19 June 1985.

1. Roberts, A. B., Frolik, C. A., Anzano, M. A. & Sporn, M. B. *Fed. Proc.* **42**, 2621-2625 (1983).
2. Todaro, G. J., Fryling, C. & De Larco, J. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5258-5262 (1980).
3. Roberts, A. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3494-3498 (1980).
4. Marquardt, H., Hunkapiller, M. W., Hood, L. E. & Todaro, G. J. *Science* **223**, 1079-1082 (1984).
5. Derynck, R., Roberts, A. B., Winkler, M. E., Chen, E. Y. & Goeddel, D. V. *Cell* **38**, 287-297 (1984).
6. Lee, D. C., Rose, R. M., Webb, N. R. & Todaro, G. J. *Nature* **313**, 489-491 (1985).
7. Roberts, A. B., Anzano, M. A., Lamb, L. C., Smith, J. M. & Sporn, M. B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5339-5343 (1981).
8. Roberts, A. B. *et al. Biochemistry* **22**, 5692-5698 (1983).
9. Frolik, C. A., Dart, L. L., Meyers, C. A., Smith, D. M. & Sporn, M. B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3676-3680 (1983).
10. Childs, C. B., Proper, J. A., Tucker, R. F. & Moses, H. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5312-5316 (1982).
11. Assoian, R. K., Komoriya, A., Meyers, C. A., Miller, D. M. & Sporn, M. B. *J. biol. Chem.* **258**, 7155-7160 (1983).
12. Sporn, M. B. *et al. Science* **219**, 1329-1331 (1983).
13. Frolik, C. A., Wakefield, L. M., Smith, D. M. & Sporn, M. B. *J. biol. Chem.* **259**, 10995-11000 (1984).
14. Tucker, R. F., Branum, E. L., Shipley, G. D., Ryan, R. J. & Moses, H. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6757-6761 (1984).
15. Assoian, R. K., Frolik, C. A., Roberts, A. B., Miller, D. M. & Sporn, M. B. *Cell* **36**, 35-41 (1984).
16. Tucker, R. F., Volkenant, M. E., Branum, E. L. & Moses, H. L. *Cancer Res.* **43**, 1581-1586 (1983).
17. Roberts, A. B. *et al. proc. natn. Acad. Sci. U.S.A.* **82**, 119-123 (1985).
18. Tucker, R. F., Shipley, G. D., Moses, H. L. & Holley, R. W. *Science* **226**, 705-707 (1984).
19. Iwata, K., Fryling, C. M., Knott, W. B. & Todaro, G. J. *Cancer Res.* **45**, 2689-2697 (1985).
20. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157-1174 (1978).
21. Dobner, P. R., Kawasaki, E. S., Yu, L. Y. & Bancroft, F. C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2230-2234 (1981).

22. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77,** 5201-5205 (1980).
23. Volckaert, G., Tavernier, J., Derynck, R., Devos, R. & Fiers, W. *Gene* **15,** 215-223 (1981).
24. Proudfoot, N. J. & Brownlee, G. G. *Nature* **253,** 211-214 (1976).
25. Kozak, M. *Nucleic Acids Res.* **12,** 857-872 (1984).
26. Battey, J. *et al. Cell* **34,** 779-787 (1983).
27. Dull, T. J., Gray, A., Hayflick, J. S. & Ullrich, A. *Nature* **310,** 777-781 (1984).
28. McGrogan, M., Simonsen, C. C., Smouse, D. T., Farnham, P. J. & Schimke, R. T. *J. biol. Chem.* **260,** 2307-2314 (1985).
29. Valerio, D. *et al EMBO J.* **4,** 437-443 (1985).
30. McKnight, S. L., Kingsbury, R. C., Spence, A. & Smith, M. *Cell* **37,** 253-262 (1984).
31. Gidoni, D., Dynan, W. S. & Tjian, R. *Nature* **312,** 409-413 (1984).
32. Benoist, C., O'Hare, K., Braetnach, R. & Chambon, P. *Nucleic Acids Res.* **Vol?** 127-142 (1980).
33. Noda, M. *et al. Nature* **295,** 202-206 (1982).
34. Amara, S. G., Jonas, V., Rosenfeld, M., Ong, E. S. & Evans, R. M. *Nature* **298,** 240-244 (1982).
35. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157,** 105-132 (1982).
36. Winzler, R. J. in *Hormonal Proteins and Peptides* Vol. 1, (ed. Li, C. I.) 1-15 (Academic, New York, 1973).
37. Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120,** 97-120 (1978).
38. Perlman, D. & Halvorson, H. O. *J. molec. Biol.* **107,** 391-409 (1983).
39. Devault, A., Zollinger, M. & Crine, P. *J. biol. Chem.* **259,** 5146-5151 (1984).
40. Anzano, M. A. *et al. Molec. cell. Biology* **5,** 242-247 (1985).
41. Mitchell, W. M. *Meth. Enzym.* **47,** 165-170 (1977).
42. Rodriguez, H., Kohr, W. J. & Harkins, R. N. *Analyt. Biochem.* **140,** 538-547 (1984).
43. Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. J. *J. biol. Chem.* **256,** 7990-7997 (1981).
44. Crea, R. & Horn, T. *Nucleic Acids Res.* **8,** 2331-2348 (1980).
45. Beaucage, S. L. & Caruthers, M. *Tetrahedron Lett.* **22,** 1859-1869 (1981).
46. Wickens, M. P., Buell, G. N. & Schimke, R. T. *J. biol. Chem.* **253,** 2483-2495 (1978).
47. Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning Techniques, A Practical Approach* (Ed. Glover, D.) (IRL, Oxford, in the press).
48. Norris, K. E., Iserentant, D., Contreras, R. & Fiers, W. *Gene* **7,** 355-362 (1979).
49. Benton, W. D. & Davis, R. W. *Science* **196,** 180-182 (1977).
50. Taylor, J. M., Illmensee, R. & Summers, S. *Biocheim. biophys. Acta* **442,** 324-330 (1976).
51. Ullrich, A., Berman, C. H., Dull, T. J., Gray, A. & Lee, J. M. *EMBO J.* **3,** 361-364 (1984).
52. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5463-5467 (1977).
53. Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9,** 309-321 (1981).
54. Pennica, D. *et al. Nature* **312,** 724-729 (1984).

Evidence for Hα emission associated with the HI bridge connecting the Small and Large Magellanic Clouds

Michel Marcelin, Jacques Boulesteix & Yvon Georgelin

Observatoire de Marseille, 2 place Le Verrier,
13248 Marseille Cedex 04, France

The bridge of matter connecting the Small Magellanic Cloud (SMC) with the Large Magellanic Cloud (LMC) was first detected thanks to radio observations, and detailed maps of this H I bridge are now available[1]. Observations of the Magellanic system at Hα wavelength reported by Johnson *et al.*[2] in 1982 seem to show a diffuse Hα emission associated with the H I bridge. As they suggested, it is necessary to confirm this association by obtaining Hα profiles in the direction of this bridge. We did this in February 1984 with the help of a scanning Perot–Fabry interferometer and a photon-counting system at the 1.5-m European Space Observatory (ESO) telescope. The choice of the area to be observed was based on recent observations[3] with the Very Wide Field Camera (VWFC) on board Spacelab 1 (November 1983) showing an extension of the SMC towards the LMC at wavelengths of 1,930 and 1,650 Å. Detection of Hα emission would check that the ultraviolet emission was emanating from the bridge connecting the two galaxies by showing that the cloud of hot stars giving rise to this ultraviolet emission was ionizing part of the H I bridge.

The patchy diffuse area detected by Courtès *et al.* with the VWFC[3] is $\sim 1° \times 2°$ in size. Courtès suggested that we should obtain profiles of the Hα emission line where ultraviolet maxima had been found ($\alpha = 1$ h 50 min, $\delta = -74°17'$ and $\alpha = 2$ h 10 min, $\delta = -74°34'$).

The instrument used for our observations has been described elsewhere[4]. We used a simpler version of the CIGALE experiment at the 1.5-m ESO telescope. The focal reducer was a smaller one, which was operated manually and which reduced the $f:15$ aperture ratio of the telescope to $f:2.7$.

The field of view is ~7 arc min in diameter and a 5-cm diameter filter was used (this filter was centred at 6,567 Å with a full width at half maximum FWHM of 10 Å). The centre of the first selected field is at $\alpha = 1$ h 50 min and $\delta = -74°17'$ and falls on chart 58 V of the Hodge and Wright[5] photographic atlas of the SMC, approximately between reference points WG 5 and L114. The scanning Perot-Fabry interferometer used interference order 798 (at Hα), giving 8.2 Å, or 376 km s^{-1} for the free spectral range. The observed field has been scanned with 21 steps (each step, or channel, corresponding to a given spacing of the plates of the interferometer) ensuring complete coverage of the free spectral range. Figure 1 shows the real-time display of this observation on a TV monitor. Four channels are selected (number 1, 6, 11 and 16, clockwise from the upper left corner) and displayed on a 256 times 256 pixel screen while the full data (photon addresses in a 512 times 512 pixel grid for each of the 21 successive channels) are recorded on a magnetic tape. Each channel was exposed for 20 s in each scanning cycle of 21×20 s = 420 s. Then the cycle was repeated 25 times. Thus the total exposure time was 500 s for each of the 21 channels, with the seeing conditions fairly well averaged for each channel over the total exposure by the scanning mode used. The analysis of the data shows that the brighter interference rings on Fig. 1 are due to geo-coronal Hα emission (the innermost one on channel 1 of Fig. 1) and OH nightsky line at 6569 Å (the precise wavelength[6] of which is 6,568.72 Å ± 0.02 Å). The faintest ring (the one closest to the centre in channel 11 of Fig. 1) is due to faint Hα emission by the object observed in the ultraviolet by the VWFC on board Spacelab 1. The possibility that is was produced by any night-sky emission line has been ruled out because it was not detected on the second field which was observed during the same night ($\alpha = 2$ h 10 min, $\delta = -74°34'$) despite an exposure time three longer (but only unscanned this time). The Hα emission line profile that has been detected is broader and fainter than those of the night-sky emissions. The width of the line is as much as 60 km s^{-1} compared with ~20 km s^{-1} for the geo-coronal Hα and for OH (the resolution imposed by the Finesse of the Perot-Fabry interferometer). The detected emission is fairly uniform over all the observed field and we may estimate its intensity by comparison with the geo-coronal Hα emission line. Comparing interference rings of same diameter for the two lines, we calculated the ratio of the number of counts and applied a small correction (~15%) to take into account the fact that each line is transmitted through the interference filter with a slightly different transmission factor, because of its narrow bandpass. As a result we found an average intensity ratio Hα (object)/geo-coronal Hα = 0.4 ± 0.1. Similarly, we found OH(6,569 Å)/geo-coronal Hα = 0.8 ± 0.1. The problem with the intensity of the night-sky lines is that they are highly variable, depending on parameters such as the geographical location, date, the time of night, zenith angle, the phase of the solar cycle and so on. Noting the values given by Chamberlain[7], (5-20 Rayleighs for the geo-coronal Hα emission at the zenith) and if we consider 10 Rayleighs as a typical value, we conclude that the detected object had an intrinsic emission around 4 Rayleighs. Converting into more usual units this gives an intrinsic monochromatic brightness around 10^{-6} erg cm^{-2} s^{-1}, sr^{-1} or 10^{-9} J m^{-2} s^{-1} sr^{-1}. This result is in very good agreement with the Johnson *et al.*[2] observation of the SMC-LMC bridge; they found a maximum Hα brightness around $2\text{-}5 \times 10^{-9}$ J m^{-1} s^{-1} sr^{-1} in the bridge between the two clouds.

However, the most interesting result is the radial velocity measured from the Doppler shift of the Hα emission of this object. Comparing the diameters of the observed interference rings with calibration rings produced by a Hα lamp we first