# A new type of transforming growth factor- $\beta$ , TGF- $\beta$ 3

Rik Derynck, Patricia B.Lindquist, Angela Lee, Duanzhi Wen, Joseph Tamm, Jeannette L.Graycar, Lucy Rhee, Anthony J.Mason, Duncan A.Miller<sup>1</sup>, Robert J.Coffey<sup>1</sup>, Harold L.Moses<sup>1</sup> and Ellson Y.Chen

Departments of Developmental Biology and Molecular Biology, Genentech, Inc., 460 Pt San Bruno Boulevard, South San Francisco, CA 94080 and <sup>1</sup>Department of Cell Biology, Vanderbilt University Medical Center, Nashville, TN 37232, USA

Communicated by B.Hogan

A new type of TGF- $\beta$ , TGF- $\beta$ 3, has been identified by cDNA characterization. The amino acid sequence of mature TGF- $\beta$ 3 and its precursor has been derived from porcine and human cDNA sequences. The human TGF- $\beta$ 3 gene is spread over seven exons as in the case of the TGF- $\beta$ 1 gene. Comparison with TGF- $\beta$ 1 and - $\beta$ 2 indicates a strong conservation of the mature sequences, but a relaxed homology in the precursor segments. TGF- $\beta$ 3 mRNA is mainly expressed in cell lines from mesenchymal origin, suggesting a biological role different from the other TGF- $\beta$ 3.

Key words: transforming growth factor- $\beta$ /proliferation/ growth inhibition/growth factor

# Introduction

Transforming growth factors- $\beta$  (TGFs- $\beta$ ) are polypeptides that influence the proliferation and differentiation of many cells types (Massagué, 1987; Sporn et al., 1986, 1987). Two distinct homodimeric TGF- $\beta$  polypeptides have been identified. TGF- $\beta$ 1 was the first polypeptide characterized and is present in high concentrations in platelets (Childs et al., 1982; Assoian et al., 1983). The mature form contains two identical chains of 112 amino acids that are each synthesized as the C-terminal part of a 390 amino acid precursor (Derynck et al., 1985, 1986). TGF- $\beta$ 2 has been recently identified in bone extracts (Seyedin et al., 1987), porcine blood platelets (Cheifetz et al., 1987) and the medium of some cell lines (de Martin et al., 1987; Marquardt et al., 1987; Hanks et al., 1988). The amino acid sequences of these two mature TGF- $\beta$  polypeptides are equal in length and are ~70% similar (Cheifetz et al., 1987; de Martin et al., 1987; Marquardt et al., 1987; Seyedin et al., 1987; Hanks et al., 1988; Madisen et al., 1988), but the precursor moieties have a higher degree of structural divergence (de Martin et al., 1987; Hanks et al., 1988; Madisen et al., 1988). Relatively few comparative data on the biological activities of TGF- $\beta$ 1 and - $\beta$ 2 have been published. It appears that both species are about equally potent in some activities such as inhibition of cell proliferation and adipogenic differentiation (Cheifetz et al., 1987). However, there are some striking differences in the activities of both TGF- $\beta$  species as shown by their differential effect on multipotential hematopoietic progenitor cells (Ohta *et al.*, 1987) and on mesoderm induction in *Xenopus laevis* embryos (Rosa *et al.*, 1988). We report here the existence of a third type of TGF- $\beta$ and propose the name TGF- $\beta$ 3. The structure of mature TGF- $\beta$ 3 and its precursor has been deduced from cDNAs isolated from porcine and human cDNA libraries. The TGF- $\beta$ 3 gene is spread over seven exons, similarly to the TGF- $\beta$ 1 gene. Comparison with the precursor sequences of TGFs- $\beta$ 1 and - $\beta$ 2 indicates a strong conservation of the mature sequences but a relaxed homology in the precursor segments. Expression of TGF- $\beta$ 3 mRNA is more restricted than in the case of TGF- $\beta$ 1.

# **Results and discussion**

# Characterization of porcine and human TGF- $\beta$ 3 precursor cDNAs

A cDNA library in the  $\lambda$ gt10 vector was constructed using mRNA derived from a porcine ovary, induced for superovulation (Mason *et al.*, 1985). A 1050 bp human TGF- $\beta$ 1 cDNA insert from  $\lambda\beta$ C1, that comprises the entire sequence for mature TGF- $\beta$ 1 and most of the precursor sequence (Derynck et al., 1985), was hybridized to the phage plaques under low stringency conditions. Approximately 300-600 of the 600 000 plaques displayed a hybridization signal of variable intensity. Some of these plaques presumably corresponded to the porcine TGF- $\beta$ 1 cDNA (Derynck and Rhee, 1987) since they also hybridized under high stringency conditions. Sequence analysis of five cDNA inserts revealed that four of these contained a segment of very high G-C content, yet lacked similarity at the amino acid sequence level with the TGF- $\beta$ 1 sequence. A low stringency hybridization of these cDNAs with the TGF- $\beta$ 1 cDNA probe is presumably due to the high G-C content of the TGF- $\beta$ 1 cDNA sequence (Derynck et al., 1985; Derynck and Rhee, 1987). The fifth cDNA analyzed contained a sequence of  $\sim$  300 bp with structural similarity at the amino acid level with mature TGF- $\beta$ 1. This cDNA was hybridized to the same ovarian cDNA library ( $\sim 1.5 \times 10^6$  phages) under high stringency conditions. The longest two of the 25 isolated cDNAs were analyzed by nucleotide sequence analysis. Using a porcine cDNA as a hybridization probe, several libraries derived from human cell sources were screened. Three hybridizing cDNAs, H-4, G3-7 and  $\beta$ 3-2000, were isolated from an ovary, an A172 glioblastoma and a placenta cDNA library respectively. The combined nucleotide sequence analyses resulted in the cDNA and derived amino acid sequences for the human homologue. The cDNA and deduced amino acid sequences for the human and porcine cDNAs are shown in Figure 1.

The amino acid sequences predicted from the human and porcine cDNA sequences are 410 and 409 amino acids long respectively and have a C-terminal sequence that resembles the previously established sequences for mature TGFs- $\beta$ . The

MET HIS LEU GLN ARG ALA LEU VAL VAL LEU ATG CAC TTG CAA AGG GCT CTG GTG GTC CTG 201 TOTOTTOCTO TOCAGECOTT ECCETECCOC TEECCTOTT TOCCAGETCA CACATEAAE VAL GLU ALA ILE ARG GLV GLN ILE LEU SER LYS LEU ARG LEU THR SER PRO PRO GLU PRO THR VAL MET THR HIS VAL PRO 371 6T6 6AA 6CC ATT AG6 6GA CAG ATC TT6 AGC AAG CTC AGG CTC ACC AGC CCC IG AG CCA ACG GTG AT6 ACC CAC GTC CCC GLU ASH THR GLU SER GLU TYR TYR ALA LYS GLU ILE HIS LYS PHE ASP HET ILE GLH GLY LEU ALA GLU HIS ASH 533 GAA AAC ACC GAG TCG GAA TAC TAT GCC AAA GAA ATC CAT AAA TTC GAC ATG ATC CAG GGG CTG GCG GAG CAC AAC ALA VAL CV& PRO LVS GLV ILE THR SER LVS VAL PHE ARG PHE ASH VAL SER SER VAL GLU LVS ASH ARG THR ASH LEU PHE 614 GCT GTC TGC CCT AAA GGA ATT ACC TCC AAG GTT TTC CGC TTC AAT GTG GTG GAG AAA AAT AGA ACC AAC CTA TTC ARG ALA GLU PHE ARG VAL LEU ARG VAL PHO ASH PHO SER SER LYS ARG ASH GLU GLM ARG ILE GLU LEU PHE GLM ILE LEU 695 CGA GCA GAA TITE CGG GITE TTG CGG GIG CCC AAC CCC AGE TCT AAG CGG ASH GAG CAG AGG ATG GAG CTC TITE CAG ATC CTI GLN ARG PRO ASP GLU HIS ILE ALA LYS GLN ARG TYR ILE GLY GLY LYS ASH LEU PRO THR ARG GLY THR ALA GLU TRP LEU SER 776 CGG CCA GAT GAG CAC ATT GCC AAA CAG CGC TAT ATC GGT GGC AAG AAT CTG CCC ACA CGG GGC ACT GCC GAG TGG CTG AC CVB HIS THE PHE GLW PEO ASH GLV ASP ILE LEU GLU ASH ILE HIS GLU VAL MET GLU ILE LYS PHE LYS GLU VAL ASP ASH 938 TGT CAC ACC TTT CAG CCC AAT GGA GAT ATC CTG GAA AAC ATT CAC GAG GTG ATG GAA ATC AAA TTC AAA GGC GTG GAC AAT GLU ASP ASP HIS GLY ARG GLY ASP LEU GLY ARG LEU LYS LYS GLW LYS ASP HIS HIS ASN PRO HIS LEU ILE LEU MET MET 1019 GAG GAT GAC CAT GGC CGT GGA GAT CTG GGG CGC CTC AAG AAG CAG AAG GAT CAC CAC AAC CCT CAT CTA ATC CTC ATG ATG ASP 230 LEU ALA + + + + - 300 LLE PRO PRO HIS ARG LEU ASP ASH PRO GLY GLH GLY GLY ARG LYS LYS ARG ALA LEU ASP THR ASH TYR CYS PHE ARG 1100 ATT CCC CCA CAC CGG CTC GAC AAC CCG GGC CAG GGG GGT CAG AGG AAG AAG CGG GCT TIG GAC ACC AAT TAC TGC TTC CGC ASM LEU GLU GLU ASM EVE CVE VAL ARG PRO LEU TYR ILE ASP PHE ARG GLM ASP LEU GLY TRP LYS TRP VAL HIS GLU PRO 1181 AAC TIG GAG GAG AAC 1GC 1GT GTG CGC CCC CIC TAC ATI GAC 11C CGA CAG GAT CTG GGC TGG AAG 1GG GTC CAT GAA CCT LYS GLY TYN TYN ALA ASH PHE EYNS SEN GLY PRO EYNS PRO TYN LEU ARG SEN ALA ASP THR TNR HIS SER THR VAL LEU GLY 1262 AAG GGC TAC TAT GCC AAC TTC TGC TCA GGC CCT TGC CCA TAC CTC CGC AGT GCA GAC ACA ACC CAC AGC ACG GTG GTA GGA LEU TYR ASM THR LEU ASM PRO GLU ALA SER ALA SER PRO CYS CYS VAL PRO GLW ASP LEU GLU PRO LEU TWR ILE LEU TYR 1343 CTG TAC AAC ACT CTG AAC CCT GAA GCA TCT GCC TEG CCT TGC TGC GTG CCC CAG GAC CTG GAG CCC CTG ACC ATC CTG TAC 1508 AGAGAGAGGG GAGAGAGAAC CACCACTGCC TGACTGCCCG CTCCTCGGGA AACACACAAG CAACAAACCT CACTGAGAGG CCTGGAGCCC ACAACCTTCG 1608 GCTCCGGGCA AATGGCTGAG ATGGAGGTTT CCTTTTGGAA CATTT CTTTGCTGGC TCTGAGAATC ACGGTGGTAA AGAAAGTGTG GGTTTGGTTA 1706 GAGGAAGGCT GAACTCTTCA GAACACACAG ACTITCTGTG ACGCAGACAG AGGGGATGGG GATAGAGGAA -AGGGATGGT AAGTIGAGAT GTIGTGTGGC 1805 ANTEGGATIT GEGETACECT ANAGEGAGAN GEANGEGEAG AGANTEGETE GETEAGEGEE AGACTEGANE ACACTTEAGA TETEAGETTE GATTTEETCA 1905 TIGCTGTACC ACATCIGCIC TAGGGAATCT GGATTATGTT ATACAAGGCA AGCATTITTT TITTTAAAGA CAGGTTACGA AGACAAAGTC CCAGAATTGT 2005 ATCICATACT -GTCTGGGAT TAAGGGCAAA TCTATTACTT TTGCAAACTG TCC--TCTAC ATCAATTAAC ATCGTGGGTC ACTACAGGGA GAAAATCCAG 2102 GICATGCAGT ICCIGGCCCCA ICAACIGIAT IGGGCCTITI GGATATGCIG AACGCAGAAG -AAAGGGGGG -AAATCAACC CICICCIGIC IGCCCICIGG 2200 BICCCICCTC ICACC ICIC CCICGAICAT ATTICCCCII GGACACTIGE TIAGACGCCI ICCAGGICAG GAIGCACATI ICIGGAITGI GETICCAIGC 

1 CCTGTTTAGA CAC<u>ATG</u>GACA ACAATCCCAG GECTACAAGG CACACAGTCC GCTTCTTCGT CCTAGGGTT CCTGGGACAGTC CTGAAGCTCT 6 6 A T 6 201 CGCAGTGCAG TGAGGTCC<u>ATG</u> CACCTTCTTG CCAAGCCTCA GTCTTTGGGGAAGG CCGCCTGGTT TTCCTCCCTC CTTCTGCACG TCTGCTGGGG

Fig. 1. Nucleotide sequence and deduced amino acid sequence of the human and porcine TGF- $\beta$ 3 precursor cDNAs. The human cDNA and amino acid sequences are shown in full. The porcine sequences differ only in the residues shown below the human nucleotide sequence or above the human amino acid sequence. Dashes indicate the absence of a residue. The 112 amino acid sequence of mature TGF- $\beta$ 3 (overlined) constitutes the C terminus of the porcine and is preceded by four basic residues (+). The precursor segment contains four overlined potential N-glycosylation sites. All cysteine residues are shaded. The AATAAA (porcine cDNA) and the related AGTAAA (human cDNA) sequence close to the 3' end of the cDNA and preceding the polyadenylation site are underlined.

2498 ATATTITIT & GEGCATECTE GATGATITCA TETTEGAA TATTETTET AGAACAGTAA AAGEETTATT CTAAGGTG T A T C G A G A G TA AAAAA

2399 GAAGCIGCAC AIGIGCCACA CAGIGACIIG GCCCCAGACG CAIAGACIGA GGTATAAAGA CAAGTAIGAA TATTACICIE AAAA-ICIII GTATAAATAA

C-terminal 112 amino acid sequence has ~80% similarity to the porcine (Derynck and Rhee, 1987) and human (Derynck *et al.*, 1985) TGF- $\beta$ 1 sequence and shares a similar

degree of homology with the sequence of TGF- $\beta$ 2 (Cheifetz *et al.*, 1987; de Martin *et al.*, 1987; Marquardt *et al.*, 1987; Hanks *et al.*, 1988; Madisen *et al.*, 1988). On the other hand,

A new type	of	transforming	growth	factor- $\beta$ ,	TGF-β	3
------------	----	--------------	--------	-------------------	-------	---

hTGF. <b>B</b> 3	20 40 MHLQRALVVLALLNFATVSLSLSTCTTLDFGHIKKKRVEAIRGQILSKLR
	********************
pTGF.β3	MHLQRALVVLALLNFATVSLSMSTCTTLDFDHIKRKRVEAIRGQILSKLR
	60 80 100
hTGF.β3	LTSPPEPTVMTHVPYQVLALYNSTRELLEEMHGEREEGCTOENTESEYYA
pTGF.β3	
P101.P3	LTSPPDPSMLANIPTQVLDLYNSTRELLEEVHGERGDDCTQENTESEYYA
hTGF. <b>B</b> 3	120 140
mor.p5	KEIHKFDMIQGLAEHNELAVCPKGITSKVFRFNVSSVEKNRTNLFRAEFR
pTGF. <b>β</b> 3	KEIYKFDMIQGLEEHNDLAVCPKGITSKIFRFNVSSVEKNETNLFRAEFR
	160 180 200
hTGF $.\beta$ 3	VLRVPNPSSKRNEQRIELFQILRPDEHIAKORYIGGKNLPTRGTAEWLSF
pTGF. <b>B</b> 3	*** ******* **************************
£	2 Ind in Sonnoby (11 b) wind bentany (11 bennep ingaarwest
htgf. <b>ß</b> 3	220 240 DVTDTVREWLLRRESNLGLEISIHCPCHTFQPNGDILENIHEVMEIKFKG
	**************************************
pTGF.β3	DVTDTVREWLLRRESNLGLEISIHCPCHTFQPNGDILENIQEVMEIKFKG
	260 280 300
htgf.β3	VDNEDDHGRGDLGRLKKQKDHHNPHLILMMIPPHRLDNPGQGGQRKKRAL
pTGF. <b>B</b> 3	**.*** *******************************
htgf.β3	320 340 DTNYCFRNLEENCCVRPLYIDFRQDLGWKWVHEPKGYYANFCSGPCPYLR
	******************
ptgf.β3	DTNYCFRNLEENCCVRPLYIDFRODLGWKWVHEPKGYYANFCSGPCPYLR
0.	360 380 400
htgf.β3	SADTTHSTVLGLYNTLNPEASASPCCVPQDLEPLTILYYVGRTPKVEQLS
pTGF.β3	SADTTHSSVLGLYNTLNPEASASPCCVPQDLEPLTILYYVGRTAKVEQLS
hTGF. $\beta$ 3	NMVVKSCKCS
pTGF. <b>β</b> 3	NMVVKSCKCS
P101.102	MANANOLICO

Fig. 2. Homology between the amino acid sequences of the human (h) and porcine (p) TGF- $\beta$ 3 precursors. The asterisks mark identical residues while a dot indicates a conservative replacement. The mature TGF- $\beta$ 3 sequences are boxed.

the sequence comparison indicates that this is a new species of TGF- $\beta$ , distinct from porcine and human TGF- $\beta$ 1 and - $\beta$ 2. We therefore propose the name TGF- $\beta$ 3.

The ATG we propose as the initiator codon for the porcine TGF- $\beta$ 3 precursor is located at nucleotide position 260 in the human sequence (Figure 1) and is in moderate agreement with the proposed consensus sequence for initiation codons. (Kozak, 1984). The second codon encodes a His residue which is also the case in the simian (Hanks et al., 1988) and human (de Martin et al., 1987; Madisen et al., 1988) TGF- $\beta$ 2 precursor sequences. In addition, the presumed initiator methionine is closely followed by a hydrophobic amino acid sequence that probably corresponds to the core of the signal peptide (Perlman and Halvorson, 1983; von Heyne, 1986) as in the case of TGF- $\beta$ 1 (Derynck et al., 1985, 1986; Derynck and Rhee, 1987) and TGF-\u00b32 (de Martin et al., 1987; Hanks et al., 1988; Madisen et al., 1988). However, it cannot be excluded that translation initiates six nucleotides upstream at an in-frame ATG which does not conform to the consensus sequence. Another ATG at position 118 is in the same reading frame as the porcine TGF- $\beta$ 3 precursor sequence, but is in a different frame in the human TGF- $\beta$ 3 cDNA sequence, indicating that a single nucleotide deletion took place in the porcine 5' untranslated cDNA sequence. Yet another ATG is present at position 14 but is followed by an in-frame stop codon (position 92 in the human sequence, Figure 1).

It is not known at what residue the signal peptide is cleaved from the rest of the TGF- $\beta$ 3 precursor. In the case of the TGF- $\beta$ 1 precursor, this cleavage precedes the Leu-Ser-Thr-Cys quadruplet at positions 30-33 (Miyazono *et al.*, 1988). This sequence can also be found in positions 22-25 in the

$\begin{array}{c} \mathtt{TGF-} \beta  \mathtt{1} \\ \mathtt{TGF-}  \beta  \mathtt{2} \\ \mathtt{TGF-}  \beta  \mathtt{3} \end{array}$	$\begin{array}{c} \texttt{M} \texttt{P} \texttt{P} \texttt{S} \texttt{G} \texttt{L} \texttt{R} \texttt{L} \underbrace{\texttt{L}}_{1} \underbrace{\texttt{L}}_{1} \texttt{L} \texttt{L} \texttt{L} \texttt{P} \underbrace{\texttt{L}}_{1} \texttt{L} \texttt{V} \underbrace{\texttt{L}}_{1} \texttt{V} \underbrace{\texttt{L}}_{1} \texttt{V} \underbrace{\texttt{L}}_{1} \texttt{V} \underbrace{\texttt{L}}_{1} \texttt{V} \underbrace{\texttt{P}}_{1} \underbrace{\texttt{A}}_{1} \underbrace{\texttt{G}}_{1} \\ \texttt{M} \texttt{H} \underbrace{\texttt{H}}_{2} \texttt{Y} \texttt{C} \underbrace{\texttt{V}}_{1} \underbrace{\texttt{L}}_{1} \texttt{K} \underbrace{\texttt{F}}_{1} \underbrace{\texttt{L}}_{1} \texttt{I} \underbrace{\texttt{I}}_{1} \texttt{L} \\ \texttt{M} \underbrace{\texttt{H}}_{2} \underbrace{\texttt{Y}}_{2} \texttt{C} \underbrace{\texttt{V}}_{1} \underbrace{\texttt{L}}_{2} \underbrace{\texttt{V}}_{1} \underbrace{\texttt{L}}_{1} \underbrace{\texttt{L}}_{1} \underbrace{\texttt{I}}_{1} \underbrace{\texttt{I}}_{1} \underbrace{\texttt{L}}_{1} \underbrace{\texttt{V}}_{1} \underbrace{\texttt{V}}_{2} \underbrace{\texttt{V}}_{2} \underbrace{\texttt{L}}_{2} \underbrace{\texttt{V}}_{2} \underbrace{\texttt{L}}_{2} \underbrace{\texttt{V}}_{2} \underbrace{\texttt{L}}_{2} \underbrace{\texttt{V}}_{2} \underbrace{\texttt{L}}_{2} \texttt{$
$\begin{array}{c} \mathtt{TGF-} & \beta \mathtt{1} \\ \mathtt{TGF-} & \beta \mathtt{2} \\ \mathtt{TGF-} & \beta \mathtt{3} \end{array}$	L S T C K T T D M E L V K R K R I E A I R G O I L S K L R L L S T C S T L D M D Q F M R K R I E A I R G O I L S K L K L L S T C T T L D F G H I K K K R V E A I R G Q I L S K L R L
$\begin{array}{c} \texttt{TGF-} \beta \texttt{1} \\ \texttt{TGF-} \beta \texttt{2} \\ \texttt{TGF-} \beta \texttt{3} \end{array}$	AS P P S Q GE V P P G P L P E A V L A L Y N S T R T S P P E D Y P E P E E V P P B V L S I Y N S T R T S P P E P I V M T H - V P Y Q V L A L Y N S T R
$\begin{array}{c} \mathtt{TGF-} \beta \mathtt{1} \\ \mathtt{TGF-} \beta \mathtt{2} \\ \mathtt{TGF-} \beta \mathtt{3} \end{array}$	$ \begin{array}{c} \hline D \\ \hline D \\ \hline L \\ L \\ L \\ E \\ E \\ L \\ L \\ E \\ \end{array} \begin{array}{c} - & - & - & - \\ - & - \\ - & - \\ - & - \\ R \\$
$\begin{array}{c} \mathtt{TGF-} \beta \mathtt{1} \\ \mathtt{TGF-} \beta \mathtt{2} \\ \mathtt{TGF-} \beta \mathtt{3} \end{array}$	K E V T R V L M V E T H N E I Y D K F K Q S T H S I Y K E V Y K I D M P P F F P S R N A T P P T F Y R P Y F B I V K E I H K F D M I Q G L A E H N E L A V C P K G I T S K V F
$\begin{array}{c} \mathtt{TGF-} \beta \mathtt{1} \\ \mathtt{TGF-} \beta \mathtt{2} \\ \mathtt{TGF-} \beta \mathtt{3} \end{array}$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
TGF-β1 TGF-β2 TGF-β3	LLRL-KLKVEQHVELYOILKSKDLTSP VERVPNSKRVPEQRIELYOILKSKDLTSP VLRVPNSKRNEQRIELYOILRPDEHIAK
$\begin{array}{c} \mathtt{TGF-} \beta \mathtt{1} \\ \mathtt{TGF-} \beta \mathtt{2} \\ \mathtt{TGF-} \beta \mathtt{3} \end{array}$	RYLSNRLLAPSDSPEWLSFDVTDAVHEW TQRYIDSKVVVKTRAEGEWLSFDVTDAVHEW - QRYIDGKNLPTRAEWLSFDVTDTVREW
$\begin{array}{c} \mathtt{TGF-} \beta \mathtt{1} \\ \mathtt{TGF-} \beta \mathtt{2} \\ \mathtt{TGF-} \beta \mathtt{3} \end{array}$	$ \begin{array}{c} 1 \\ \texttt{SR} \\ \texttt{G} \\ \texttt{G} \\ \texttt{G} \\ \texttt{G} \\ \texttt{G} \\ \texttt{E} \\ \texttt{E} \\ \texttt{I} \\ I$
$\begin{array}{c} \mathtt{TGF-} \beta \mathtt{1} \\ \mathtt{TGF-} \beta \mathtt{2} \\ \mathtt{TGF-} \beta \mathtt{3} \end{array}$	D S R D N T L Q V D I NGGF T T G R Z Z Z A T I NK S E E L E A R F A G I D G T S T Y T S G D Q K T I E N I HE V M E I K F K G V D N E D D H F Z Z Z Z
$\begin{array}{c} \mathtt{TGF} - \ \beta  \mathtt{1} \\ \mathtt{TGF} - \ \beta  \mathtt{2} \\ \mathtt{TGF} - \ \beta  \mathtt{3} \end{array}$	H G M N R P E L L L M A T P L E R A Q H LQ G M N R P E L L L M A T P L E R A L E S Q Q K S T R K K S G K T P H L L L M M L L P S K R L D N P G Q G K K K Q K - D H H N P H L T L M M K L P P H R L D N P G Q G
$\begin{array}{c} \mathtt{TGF-} \beta \mathtt{1} \\ \mathtt{TGF-} \beta \mathtt{2} \\ \mathtt{TGF-} \beta \mathtt{3} \end{array}$	SS - RHRRALDINYCFISSTEKNCCVRQLYID TNRBKKRALDAAYCFRNVQDNCCLRPLYID GQ - RKKRALDTNYCFRNLEENCCVRPLYID
$\begin{array}{c} \mathtt{TGF} = \begin{array}{c} \beta  \mathtt{1} \\ \mathtt{TGF} = \begin{array}{c} \beta  \mathtt{2} \\ \mathtt{TGF} = \begin{array}{c} \beta  \mathtt{3} \end{array} \end{array}$	FRKDLGWKWIHEPKGYYANFCLGPCPYLWS FKRDLGWKWIHEPKGYYANFCLGFCPYLWS FRQDLGWKWVHEPKGYYANFCCSGPCPYLRS
$\begin{array}{c} \mathtt{TGF} = \beta \mathtt{1} \\ \mathtt{TGF} = \beta \mathtt{2} \\ \mathtt{TGF} = \beta \mathtt{3} \end{array}$	LD T Q Y S K V L A L Y N Q H N F G A S A A P C C V P Q A L S D T Q H S R V L S L Y N T I N P E A S A S P C C V P Q D L A D T T H S T V L G L Y N T L N P E A S A S P C C V P Q D L
$\begin{array}{c} \mathtt{TGF-} \beta \mathtt{1} \\ \mathtt{TGF-} \beta \mathtt{2} \\ \mathtt{TGF-} \beta \mathtt{3} \end{array}$	E P L PI L Y Y V G R K P K V E Q L S N M I V E S C K C S E P L T I L Y Y G R K P K V E Q L S N M V V K S C K C S E P L T I L Y Y V G R T P K V E Q L S N M V V K S C K C S

**Fig. 3.** Polypeptide sequence similarity between the human TGF- $\beta$ 1, - $\beta$ 2 and - $\beta$ 3 precursors at the amino acid level. Dashes are introduced for maximal alignment. Identical amino acids are boxed. The heavy horizontal arrow indicates the start of the mature TGF- $\beta$  sequences. The cysteine residues are shaded, while the RGDL sequences are cross-hatched. The potential sites for N-glycosylation are underlined. The position of the introns in the TGF- $\beta$ 3 genes are indicated above the TGF- $\beta$ 1 or under the TGF- $\beta$ 3 sequence. The three dots represent the nucleotides of the codon for the corresponding amino acid and the arrow marks the insertion point of the intron.

TGF-β1	N-S	¥	Ψ	¥ ;;;		
TGF-β2	с <u>с</u> N- <mark>S</mark>	¥.	Ŷ	دون : 	<b>900000000</b>	22 22
TGF-β3	N-S	Ŷç.	, <u>μ</u> γ	сс Ш	ccc cc	

Fig. 4. Schematic representation of the three human TGF- $\beta$  precursors with the relative positions of the N-glycosylation sites and cysteines (C). The vertical arrows indicate the proposed cleavage sites following the signal peptide (S) and preceding the C-terminal mature TGF- $\beta$  sequence.

human TGF- $\beta$ 3 precursor (Figure 1). The TGF- $\beta$ 3 precursor sequence contains four potential N-glycosylation sites (Asn-X-Ser or -Thr; Winzler, 1973) and five cysteine residues.

The 112 amino acid TGF- $\beta$ 3 sequence is preceded by four basic residues, as in the case of TGF- $\beta$ 1 (Derynck *et al.*, 1985, 1986; Derynck and Rhee, 1987) and - $\beta$ 2 (de Martin

GTACTGCCCCACCTTTAGCTGGGCCAGCAACTGCCGGGCCCTGCTTCTCCCCACCTACTGGTGATCTTTTTTTT
TTTTCCATTCTCTTTTCTTTCTTTCAAGGCAAGGCAAGG
TETECCEATEGCEAAGEGGEGTTTGGEAATATEGAATATECHETETATTTATTTTACCTAAGGAAAAAETCEAGETCECHTECCAGTGCEATECCAGTGCEAT
GCCACCCCTCCCAGCCCTCTGCTTGCCCTGCCTGCCTGCC
AAGGCACAGTCCGCTTCTTCGTCCTCAGGGTTGCCAGGCTTCCTGGAAGTCCTGAAGCTCTCGCAGTGCAGTGCAGTTCATGCACCTTCTTGCCAAGC
CTCAGTCTTTGGGATCTGGGGAGGCCGCCTGGTTTTCCTCCCTC
10 Het His Lou Gin Ary Ala Lou Val Val Lou Ala Lou Lou Asn Pho Ala Thr Val Ser CTCTTCCCAGCTCACACATGAMG ATG CAC TTG CAA AGG GCT CTG GTG GTC CTG GTC CTG GTC TG AGC TTT GCC AGG GTC AGG
20 500 Lou Ser The Cys Thr Thr Lou Jap Pho Gly His Ile Lys Lys Lys Arg Val Glu Als Ile Arg Gly Gln Lou Ser Lou Ser Thr Cys Thr Thr Lou Jap Pho Gly His Ile Lys Lys Lys Arg Val Glu Als Ile Arg Gly Gln Cit Cit Cit Cit Cit Cit Cit Arg CA ROC TH Cit Cit Cit Arg Arg Arg Arg Arg Arg Arg Arg Cit Arg Gly Gln
Ile Leu Ser Lys Leu Arg Leu Thr Ser Pro Pro Glu Pro Thr Yel Mer Thr His Yel Pro Tyr Gin Yel Leu Ale ATC TTG MGC AMG CTC MGG CTC AGG CCC CCT GMG CCA ACG GTG ATG ACC CAC GTC CCC TAT CMG GTC CTG GCC
70 BO Lew Tyr Aen Ser Thr Arg Glu Lew Lew Glu Glu Net His Gly Glu Arg Glu Glu Gly Cys Thr Gln Glu Aen Thr CTT The Ame Age Ace Cog Gag Che Che Ghe Gag Gag Ang Gag Gag Gag Gag Gag Gag Gag Che Che Gam Ame Ace
100 Glu Ser Glu Tyr Tyr Ala Lye Glu Ile His Lye Phe Amp Met Ile Gln Gly Leu Ala Glu His A Gan TGG GAA HAT TAT GCC AMA GAA NT CAY AMA TYC GAC ATG ATC CAY GOOG CTC GOG GGA GA GAA argumentocasatto
GAG TOS GAA TÀC TÀT GOC AÀA GAA ATC CAT AÀA TTC GAÈ ATG ATC CAG GOS CTG GOS GAG CAC A gtaagtocaaatto
togotggggtgtotgototggagggtotgaactggagotgggagototgcagtggcagpogggaoggtggccaettggtgootgteogtt 120
120 an Glu Leu Ala Val Cys Fro Lys Gly ggcagcascatgttoctgestgtetogetoatgetgtgccctetgetttatetoetag AC GAA CTG GCT GTC TGC CCT AAA GGA
130 11e Thr Ser Lys Val Phe Arg Phe Asn Vel Ser Ser Val Glu Lys Asn Arg Thr Asn Leu Phe Arg Als Glu Phe ATT ACT TOC ANG GTT THC COR THC AAT GTG TOC TCA GTG GAG AAA AMT AGA ACC AAC CTA TTC COR GCG GAA TTC
150 Arg Val Lau Arg Val Fro Asn Fro Ser Ser Lys Arg Asn Glu Gln Arg 11e Glu Leu Phe Gln GG GFC TTG GGG GTC CCC AMC CTC TAG CGG AAT GAG GAC AGG ATC GAG GTC TTC CAG gtsactoctototog
COG GTC TTG COG GTG CCC AAC CCC AGC TCT ANG COG AAT GMG CMG AGG ATC GMG CTC TTC CAG gtasetoeteterag
agcagaaaccacacogaogggaaagctggttoctttgccatatcagggcaccactgggtgcagcgtttggcagaoctgggtttgaatoot
tt caggaaccactggtttootatogotooostcacotgtgststgscatsttstgggscaststgtgscatstgtscsgggtactcagggtactcastst
180 Ile Leu Arg Pro Amp Glu Him Ile Ale Lym Gln Arg Tyr lle Gly Gly Lym Aan Leu Pro Thr Arg goortgtag ATC CTT CGG CCA GAT GAG CAC ATT GCC AAA CAG CGG TAT ATC GGT GGC AAG BAT GTG CCC ACA CGG
200 Gly Thr Als Glu Trp Leu Her Phe App Val Thr App Thr Val Arg Glu Trp Leu Leu Arg Arg G GGC ACT GGC GAG TGG CTG TGC TTT GAT GTG ACT GGC TGG CTG GGC GGC GGC GGG GG GGG GG
GOC NOT GOC GAG THE CTG TOC TIT GAT GTC ACT GAC ACT GTG COT GAG THE CTG TTE AGA AGA & gtaggtggacoot
tcagataagcatttcagaatgaacctcaggtcocttagtcctocatgaaatggagggaagaggacagaattaagggagtcagagatctgggttcaaaccc
lu Ser tagttcotggtggtgggccacccttogccatgacacaccggctogcttttctgcactaattgtgtcttattttgcag AG TCC
220 Asn Lew Cly Lew Clu Lie Ser IIe His Cys Pro Cys His für Pho Cln Pro Asn Gly Asp IIe Lew Glu Asn II Amc TTA GGT CTA GAA ATC AGC ATT CAC TOT CCA TOT CAC ACC TTT CAG CCC AAT GGA GAT ATC CTG GAA AAC ATT
His Glu Val Met Glu Ile Lys Phe Lys G CAC GAG GTG ATG GAA ATC AAA TTC AAA G gleacanaatgaatgsglaggaggagggagggggggggggggggggg
gt graggagacort taccagacort caaggt correspand to ort ogragar
ly Val Amp Ann Glu Amp Ains ggaggcagcotocsagggstetesettottcaggagatgtcasttteetteettetteg GC GTG GAC AAT GAC GAT GAC CAT
240 Gly Ary Gly Arp Lea Uly Ary Lea Lys Lys Gin Lys Arp His His Ann Pro His Lea Ile Lea Het Het Ile Pro Get Coff God Art CTO GOG COC CTC ANA MAG CALE ANA GAT CAC CAC ANC CCT CAT CTA THE ARC CTC ATG AFT CCC
200 Pro His Arg Leu Asp Ann Pro Gly Gln Gly Gln Arg Lys Lys Arg Mis Leu Anp Thr Ann Tyr Cys Phe Ar Cra cur con cro cut Ale Con one cut one dong don cas and ale ale con one the are act and the for the for
210 g Ann Lew Glu Glu Ann Cyć Cys Val Krg Fro Lew Tyr Ile Kep Phe Krg Gln ttotstasastyttitjtjetjetag c Anc Tri Gade Gade Anc Toc Tor Tor Coc Coc Crt Tri Ant Gade The Coch CAG
Amp Law Gly Trp Lya Trp Val Hig Glu Fro Lya Gly Tyr Tyr Ala Amn Pha Cya Bar Gly Fro Cym Fro Tyr Law Gar etg gor tog ang tog gte car gan get ang goe the tat goe and the tre tog goe cet toe cen the cre
330 Ārg Šer Ala Amp Thr Thr Him Set Thr CCC AGT GCA GAC ACA ACC CAC ACC ACG gtatgangcaggetcatgeogtcatgangcactugggetgacogaccaggaccactigitaasaag
as to sol dandest to set of to set used as test extent of emericans as a
Val Low City Low Yr Man Thr gaactgegesettootgegtlocceggootgegestgegoottocsacctotcactgescatgtoottocceg OTG CFG GAL CFG TAC ANC ACT
10 10 10 10 10 10 10 10 10 10 10 10 10 1
400 Gly Arg Thr Pro Lys Val Glu Gln Lau Mor Ann Net Val Val Lys Ser Cys Lys Cys Ser 666 Ang Acc CCC Ana GTG GAG CAG CTC TCC ANC ATG GTG GTG ANG TCT TCT AAA TCT ANC TGA GACCCCCACGTOCGA
CARAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
OCTCCCCCCAAATCCCTCACATCCAACATCCCTTTTCCCAACATTCCTTCCTCC
AGGAAGGETGAACTETTERGAACACACAGACTTTETGTGREGERGAGGGAAGGGGATGGGGAAGGGATGGGTAAGTTGAGATGTTGT
TOGGATTTGGGCTACCCTALAGGGAAGGAAGGAAGGGAAGGAATGGCTGGGTCAGGCCAGACTGGAAGACACTTCAGATCTGAGGTTGGA
GCTGTACCACATCTGCTCTAGGGAATCTGGATTATGTTATACAAGGCAAGCATTTTTTTT
TOTATCTCA TACTOTCTOODA FTAADOOCAAA FCTA FTACTT TTOCAAACTOTCCTCTACA TCAACA TCAACA TCOTOODTCACTACAAODAAAAA TCCAG
gtcargengttoctggcoccatcanctgtattgggocttttggatatgctgancgchganggggggaaarcancoctctctctgtctgcoctctgggt
COCTOCTOCHOCTOCHORATCATATTTCCCCTTOGACACTTOGTTAGACGOCTTCCAGGTCAGGATGCACATTTCTGGATTGTGGTTCCATGCACC
CTT0000CATTAT000TTCTTCCCCCACTTCCCCTCCAAGACCCT010TTCATTT00T0TTCCT00AAGCAGTGCTACAACATGTQAGGCATTCC00GA
AGCTOCHCATGTOCCHCHCTGACTTGGCCCCAGACGCATAGACTGAGGTATAAAGGCAAGTATGAATATTACTCTCAAAATCTTTGTATAAATAA
TTTTT0000CATCCT00AT0ATTTCATCTTCT00AATATTGTTCTAGAACAGTAAAAOCCTTATTCTAAGGTOTALgLoLgeLageLaseLeLoLLC
est Leocot Leek Locat gt compost at ant se compy Leocosses at a gampoogsagest at gt cast coocese celt gat Legamese d co
oottassasyyyyyvaassoytyntessitoopootpotasyssetynastattynttytynytottoottottynagagtatattattottosgaga

et al., 1987; Hanks et al., 1988; Madisen et al., 1988). By analogy with the other two TGFs- $\beta$ , one can assume that proteolytic cleavage following this basic tetrapeptide generates the mature TGF- $\beta$ 3 homodimer.

Comparison of the porcine and human TGF- $\beta$ 3 precursor sequences reveals a 90% amino acid identity. The porcine sequence is one amino acid residue shorter due to a single deletion in the precursor segment (position 271, Figure 2). The sequence conservation between both mature TGF- $\beta$ 3 sequences is much stronger since only two conservative amino acid differences are present. The high degree of sequence conservation is also apparent at the cDNA level. However, it is remarkable that both the 5' and 3' untranslated sequences are also very highly conserved (Figure 1). This feature is rather uncommon among the known mRNA sequences and may reflect a relevant biological or regulatory role for these non-coding sequences.

# Sequence comparison of the three human TGF- $\beta$ precursors

The establishment of the sequence of this third type of TGF- $\beta$ and of its precursor allows a comparison of all three human TGF- $\beta$  precursor sequences (Figure 3). The sequence conservation of the mature TGFs- $\beta$  includes all nine cysteines which determine the disulfide bridge formation. TGF- $\beta$ 3 is ~80% similar to TGF- $\beta$ 1 and to TGF- $\beta$ 2, while TGF- $\beta$ 2 is 72% similar to TGF- $\beta$ 1. In contrast, the precursor sequences for the three TGFs- $\beta$  are remarkably dissimilar. Relatively large gaps have been introduced in the sequence in order to achieve maximal similarity. However, some structural features and sequences are conserved in all three precursors, presumably due to their biological significance. Both the TGF- $\beta$ 1 and TGF- $\beta$ 2 precursor contain three potential N-glycosylation sites, in contrast to the TGF- $\beta$ 3 precursor that has four sites (Figures 3 and 4). Two of these are found in all three precursors in corresponding positions. Also conserved in the TGF- $\beta$ 1 and TGF- $\beta$ 3 precursors is the tetrapeptide RGDL (residues 259-262 in Figure 1), which has been detected in several extracellular matrix proteins that are involved in interaction with the cells (Pierschbacher and Ruoslahti, 1984a,b; Ruoslahti and Pierschbacher, 1987). This tetrapeptide is absent in the TGF- $\beta$ 2 precursor. A major difference between the three precursor sequences is their number of cysteine residues. The TGF- $\beta$ 3 precursor segment contains five cysteines versus three in the corresponding TGF- $\beta$ 1 precursor sequence and six in the TGF- $\beta$ 2 precursor segment. Three of these are in corresponding positions in all three precursors (Figures 3 and 4). Recently, it has been shown that, following cleavage from the mature TGF- $\beta$ 1 sequence, the TGF- $\beta$ 1 precursor segment remains hydrogen-bonded with mature TGF-\u03b31 (Miyazono et al., 1988; Wakefield et al., 1988). The 'latent' or inactive TGF- $\beta$ 1 stored in platelets (Pircher et al., 1986) and presumably also the 'latent' TGF- $\beta$ 1 secreted by cells in culture may correspond to this complex. If TGF- $\beta$ 2 or TGF- $\beta$ 3 are also made in 'latent' form, it is possible that these inactive complexes are significantly different from each other and from 'latent'

Fig. 5. Partial nucleotide sequence of the human TGF- $\beta$ 3 gene. The asterisk marks the 5' most residue of the cDNA (Figure 1b). The incomplete intron sequences are in small letter type. The last nucleotide of the 3' untranslated region in capitals marks the presumed polyadenylation site. The mature TGF- $\beta$  sequence is overlined.

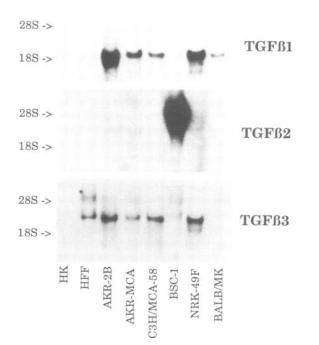


Fig. 6. Northern blot analysis of TGF- $\beta$ 1, TGF- $\beta$ 2 and TGF- $\beta$ 3 mRNA from cultured cells. The cell types are: HK, secondary cultures of human foreskin keratinocytes; HFF, secondary cultures of human foreskin dermal fibroblasts; the rodent fibroblastic cell lines AKR-2B, AKR-MCA, C3H/MCA58 and NRK49F; BALB/MK, a murine keratinocyte cell line and BSC-1, a monkey epithelial cell line.

Table I. Differential ex	pression of TGF	-β mRNAs in	tissues and
cultured cells	-		

	TGF-β1	TGF-β2	TGF-β3
Epithelial cell lines			
A431	+++	+	-
SW480	+++	-	
SW620	+	-	-
BSC-1	+	++++	-
BALB/MK	++	±	-
Mesenchymal cell lines			
HT-1080	+++	++	++
NRK-49F	+++	+	++
AKR-2B	+ +	±	++
AKR-MCA	+ +	±	+
C3H/MCA-58	++	±	+
Cell strains and tissues			
Human keratinocytes (HK)	+	±	±
Human foreskin fibroblasts (HFF)	+	±	+
Rat liver	++	-	_
Dog kidney	++	+	-

The cell lines examined and not illustrated in Figure 6 are derived from a human bronchio-alveolar carcinoma (A-431; Moses *et al.*, 1981), human colon carcinomas (SW480 and SW620; Coffey *et al.*, 1986) and human fibrosarcoma (HT1080; Moses *et al.*, 1981). Estimation of band intensities were made relative to positive controls, which were AKR-2B cells (TGF- $\beta$ 1 and TGF- $\beta$ 3) and BSC-1 cells (TGF- $\beta$ 2) and range from a weak (+) to an exceptionally strong (+++) signal. The rating of  $\pm$  indicates the presence of definite bands only after prolonged (2-week) exposures, but not with shorter (1-4 day) exposures. TGF- $\beta$ 1. This is suggested by the structural dissimilarity between the three TGF- $\beta$  precursor sequences. It is conceivable that several of the sequences conserved in all three precursors may play an important role in the inactivating interaction between the precursor segment and the mature TGF- $\beta$ . This could certainly be the case for the largest conserved continuous sequence in the precursor segment, located at positions 34-55 in the TGF- $\beta$ 3 sequence (Figures 1, 2 and 3). This sequence is rich in basic residues. The TGF- $\beta$ 3 stop codon in the cDNA sequence is followed by  $\sim 1080$  bases of 3' untranslated sequence (Figure 1). The porcine sequence ends with a poly-A tail which is preceded by the hexanucleotide AATAAA that presumably functions as the polyadenylation signal (Proudfoot and Brownlee, 1976). The stop codon of the TGF- $\beta$ 1 cDNA sequence is immediately followed by a very G-C rich sequence that may play a role in transcriptional or translational control (Derynck et al., 1985). A similar sequence also follows the stop codons for inhibin- $\beta_A$  and  $-\beta_B$ , which are structurally related to TGF- $\beta$  (Mason et al., 1985). Such a G-C rich sequence is absent in the 3' untranslated region of the TGF- $\beta$ 3 and - $\beta$ 2 precursor cDNAs.

# The human TGF- $\beta$ 3 gene: intron – exon structure

Using the porcine TGF- $\beta$ 3 cDNA as hybridization probe, we have isolated several recombinant genomic phage from a human genomic liver DNA library (Lawn et al., 1978). Detailed analysis and nucleotide sequence determination led to the characterization of the exons and intron-exon junctions in the TGF- $\beta$ 3 gene (Figure 5). As shown in Figures 3 and 5, the TGF- $\beta$ 3 gene contains seven coding exons, while there are no introns in the 3' untranslated region. It cannot be excluded that there may be one or more introns in the 5' untranslated region, which is incomplete in the cDNA. We have previously reported that the human TGF- $\beta$ 1 gene is also spread over seven exons (Derynck et al., 1987b). Comparison of the sequences of both the TGF- $\beta$ 1 and - $\beta$ 3 genes indicates that all intron-exon junctions are localized at exactly corresponding nucleotide positions, with the exception of the first intron (Figure 3). The location of this first intron-exon boundary in the coding sequences of the TGF- $\beta$ 1 and - $\beta$ 3 precursors differs by only three nucleotides. This striking conservation of the splice junctions stands in marked contrast to the relatively low degree of sequence similarity of the precursor segments. Areas with high sequence conservation are not encoded in separate exons, as is best illustrated in the fifth exon which contains a segment of the divergent precursor sequence and the beginning of the conserved mature TGF- $\beta$ . The conserved exon configuration indicates that the existence of genes for both TGF- $\beta$ 1 and - $\beta$ 3 (and also TGF- $\beta$ 2) is a result of an ancestral duplication of the entire gene. Recent chromosomal mapping studies (Brissenden et al., 1985; Barton et al., 1988) have localized the three TGF- $\beta$  genes to three different chromosomes. These data rule out a tandem duplication as has been suggested for many gene clusters. It is likely that the TGF- $\beta$  genes have been duplicated in very early times and that the structural similarities at the polypeptide level have been maintained by functional constraints.

### Synthesis of mRNA for the three TGF- $\beta$ precursors In order to evaluate possible cell sources for TGF- $\beta$ 3

better, Northern hybridizations were done using the human TGF- $\beta$ 1, - $\beta$ 2 and - $\beta$ 3 cDNAs as probe (Figure 6, Table I). Under the stringency conditions used, there was no cross hybridization of the different cDNA probes (data not shown). The major TGF- $\beta$ 1 mRNA species was ~2.5 kb long and could be found in all cell lines tested. The ubiquitous presence of TGF- $\beta$ 1 mRNA in cell lines is consistent with previously reported data (Derynck et al., 1985, 1987a). TGF- $\beta$ 2 transcripts were present in several but not all cell lines and had major transcript sizes of  $\sim 4$  and 6 kb. These mRNA sizes conform with the previously reported sizes of TGF- $\beta$ 2 mRNA (de Martin et al., 1987; Madisen et al., 1988). TGF- $\beta$ 3 is encoded by a single mRNA species of  $\sim$  3.2 kb, indicating that the cDNA sequences shown in Figure 1 are close to full length. TGF- $\beta$ 3 mRNA could be detected in several of the cell lines tested, mainly cell lines of mesenchymal origin (Table I). No TGF-\beta3 mRNA was detected in many cell lines from epithelial origin [Table I; also the cell lines KB, HBL100, MDA-MB 436 and T24 (for references see Derynck et al., 1987a)] and in the hematopoietic cell lines CEM and Raji (data not shown). In related studies, TGF- $\beta$ 3 mRNA has been detected in relatively high abundance (++) in freshly isolated rat testicular myoid and Sertoli cells (M.Skinner et al., unpublished). TGF- $\beta$ 3 mRNA was also present in freshly isolated bovine ovarian theca cells (+), but not in granulosa cells (M.Skinner et al., in preparation). Our data may reflect the possibility that tissues of mesodermal origin are the primary source for TGF- $\beta$ 3 synthesis in vivo. Many of the cell lines tested contain several species of TGF- $\beta$  mRNA.

We have thus established the existence and the structure of a third type of TGF- $\beta$  mRNA and have derived the amino acid sequence of the mature form and its precursor. TGF- $\beta$ 3 mRNA is synthesized by various cell lines, chiefly of mesenchymal origin. Considering the structural differences among these three types of TGF- $\beta$ , it will be important to explore their differential expression and their regulation in cell populations *in vivo*. Such evaluation will require specific tools, due to the high degree of sequence conservation of the mature TGFs- $\beta$  and it is unlikely that detection based on polyclonal antisera will discriminate between the different species. Additional studies will also be needed to evaluate the biological role of all three TGFs- $\beta$ in relation to each other, especially since the various cell sources may secrete several types of TGF- $\beta$ .

After submission of this manuscript, ten Dijke *et al.* (1988) reported the human TGF- $\beta$ 3 cDNA sequence. Their sequence is in agreement with ours, although they chose the ATG at position 254 (Figure 1) as initiator codon. There are several differences in the alignment of the sequence conservation between the three TGF- $\beta$  precursors by ten Dijke *et al.* (1988) and by us (Figure 3).

#### Materials and methods

### Isolation and characterization of cDNAs and gene fragments

About  $6 \times 10^5$  plaques of a  $\lambda$ gt10 based porcine ovary library were hybridized with the <sup>32</sup>P-labelled (Taylor *et al.*, 1976), 1050 bp long human TGF- $\beta$ 1 cDNA from  $\lambda\beta$ C1. The hybridization took place in 5 × SSC, 20% formamide, 50 mM sodium phosphate pH 6.8, 0.1% sodium pyrophosphate, 5 × Denhardt's solution, 50  $\mu$ g/ml salmon sperm DNA at 42°C for 15 h. The filters were washed at increasing stringency: 2 × SSC, 0.5 × SSC, 0.2 × SSC and 1 × SSC (all at 42°C). Autoradiography was performed following the washes at a given stringency. The human cDNAs were obtained by hybridizing a human ovary, an A172 glioblastoma and a human placenta library (~1 × 10<sup>6</sup> plaques each) with the porcine TGF- $\beta$ 3 cDNA. Hybridizations were at 42°C in 50% formamide, 5 × SSC, 50 mM sodium phosphate pH 6.8, 0.1% sodium pyrophosphate, 5 × Denhardt's solution, 50 µg/ml salmon sperm DNA and subsequent washes were in 0.2 × SSC at the same temperature. Using the same hybridization conditions and probes, 1.5 × 10<sup>6</sup> recombinant phage from a human genomic liver DNA library (Lawn *et al.*, 1978) were screened. This led to the isolation of 48 phage which were further characterized by hybridization to different TGF- $\beta$ 3 cDNA restriction fragments. Restriction fragments containing all human TGF- $\beta$ 3 precursor exons were derived from the two phage  $\lambda\beta$ 3-24 and  $\lambda\beta$ 3-5. The DNA fragments were subcloned into M13 phage derivatives (Messing *et al.*, 1981) and were sequenced using the dideoxy sequencing methods (Smith, 1980).

#### Northern hybridization

RNA was prepared from the following cell sources: HK, secondary cultures of human foreskin keratinocytes (Tucker et al., 1984); HFF, secondary cultures of human foreskin dermal fibroblasts (Tucker et al., 1984); AKR-2B, continuous line of nontransformed mouse embryo-derived fibroblastic cells (Moses et al., 1981); AKR-MCA, a chemically-transformed derivative of the AKR-2B cells (Moses et al., 1981); C3H/MCA-58, a chemicallytransformed derivative of the C3H/10T mouse embryo derived fibroblastic line (Moses et al., 1981); BSC-1, African green monkey kidney epithelial cell line (Hanks et al., 1988); NRK-49F, rat kidney-derived fibroblastic cell line (Assoian et al., 1983); and BALB/MK, a mouse skin keratinocyte cell line (Coffey et al., 1988). HT1080, a human fibrosarcoma cell line (Moses et al., 1981); A431, a human bronchio-alveolar carcinoma (Moses et al., 1981) and the human colon carcinomas SW480 and SW620 (Coffey et al., 1986). RNA was extracted as described (Schwab et al., 1983). Polyadenylated RNA (4 µg) was electrophoresed into a formaldehyde/1.2% agarose gel (Dobner et al., 1981) and blotted onto nitrocellulose (Thomas, 1980). The nitrocellulose filters were hybridized with <sup>32</sup>P-labelled (Taylor et al., 1976) cDNA probes in 50% formamide, 5 × SSC, 0.1% SDS,  $1\,\times$  Denhardt's, 250  $\mu g/ml$  salmon sperm DNA and 50  $\mu g/ml$  poly(A) at 43°C for 18-24 h. Washings were done in  $0.1 \times (TGF-\beta 1)$  or  $1.0 \times (TGF-\beta 2 \text{ and } TGF-\beta 3) \text{ SSC}, 0.1\% \text{ SDS}, \text{ and } 1 \text{ mM EDTA at } 43^{\circ}\text{C}.$ The TGF- $\beta$ 1 probe consisted of the 1050 bp cDNA of  $\lambda\beta$ C1 (Derynck et al., 1985). The 2.2 kb Eco RI fragment of human TGF-\20162 cDNA (Madisen et al., 1988) was used to probe for TGF- $\beta$ 2 mRNA. The human TGF- $\beta$ 3 cDNA, shown in Figure 1b but starting at position 712, was used in Northern hybridization for the TGF- $\beta$ 3 mRNA.

#### Acknowledgements

We thank Frances Hwang and Carol Collins for their assistance, Dr A.Ullrich for providing a cDNA library, and Oncogen Inc., Seattle, WA, for providing the TGF- $\beta$ 2 cDNA probe utilized for Northern blot analysis. This work was supported in part by PHS grant no. CA-42572 awarded by the National Cancer Institute, DHHS.

#### References

- Assoian, R.K., Komoriya, A., Meyers, C.A., Miller, D.M. and Sporn, M.B. (1983) J. Biol. Chem., 258, 7155-7160.
- Barton, D.E., Foellmer, B.E., Du, J., Tamm, J., Derynck, R. and Francke, U. (1988) Oncogene Research, 3, in press.
- Brissenden, J.E., Derynck, R. and Francke, U. (1985) Cancer Res., 45, 5593-5597.
- Cheifetz, S., Weatherbee, J.A., Tsang, M.L., Anderson, J.K., Mole, J.E., Lucas, R. and Massagué, J. (1987) Cell, 48, 409-415.
- Childs, C.B., Proper, J.A., Tucker, R.F. and Moses, H.L. (1982) Proc. Natl. Acad. Sci. USA, 79, 5312-5316.
- Coffey, R.J., Shipley, G.D. and Moses, H.L. (1986) Cancer Res., 46, 1164-1169.
- Coffey, R.J., Sipes, N.J., Bascom, C.C., Graves-Deal, R., Pennington, C.Y., Weissman, B.E. and Moses, H.L. (1988) Cancer Res., 48, 1596-1602.
- de Martin, R., Haendler, B., Hofer-Warbinek, R., Gangitsch, H., Wrann, M., Schlüsener, H., Seifert, J.M., Bodmer, S., Fontana, A. and Hofer, E. (1987) *EMBO J.*, 6, 3673-3677.
- Derynck, R. and Rhee, L. (1987) Nucleic Acids Res., 15, 3187.
- Derynck, R., Jarrett, J.A., Chen, E.Y., Eaton, D.H., Bell, J.R., Assoian, R.K., Roberts, A.B., Sporn, M.B. and Goeddel, D.V. (1985) Nature, 316, 701-705.
- Derynck, R., Jarrett, J.A., Chen, E.Y. and Goeddel, D.V. (1986) J. Biol. Chem., 261, 4377-4379.
- Derynck, R., Goeddel, D.V., Ullrich, A., Gutterman, J.V., William, R.D.,

Bringman, T.S. and Berger, W. (1987a) Cancer Res., 47, 707-712.

- Derynck, R., Rhee, L., Chen, E.Y. and Van Tilburg, A. (1987b) Nucleic Acids Res., 15, 3188-3189.
- Dobner, P.R., Kawasaki, E.S., Yu, L.Y. and Bancroft, F.C. (1981) Proc. Natl. Acad. Sci. USA, 78, 2230-2234.
- Hanks,S.K., Armour,R., Baldwin,J.H., Maldonado,F., Spiess,J. and Holley,R.W. (1988) Proc. Natl. Acad. Sci. USA, 85, 79-82.
- Kozak, M. (1984) Nucleic Acids Res., 12, 857-872.
- Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) *Cell*, 15, 1157-1174.
- Madisen, L., Webb, N.R., Rose, T.M., Marquardt, H., Ikeda, T., Twardzik, D., Seyedin, S. and Purchio, A.F. (1988) DNA, 7, 1-8.
- Marquardt, H., Lioubin, M.N. and Ikeda, T. (1987) J. Biol. Chem., 262, 12127-12131.
- Mason, A.J., Hayflick, J.S., Ling, N., Esch, F., Ueno, N., Ying, S.Y., Guillemin, R., Niall, H. and Seeburg, P.H. (1985) Nature, 318, 659-663.
- Massagué, J. (1987) Cell, 49, 437-438.
- Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res., 9, 309-321.
- Miyazono, K., Hellman, V., Wernstedt, C. and Heldin, C.H. (1988) J. Biol. Chem., 263, 6407-6415.
- Moses, H.L., Branum, E.B., Proper, J.A. and Robinson, R.A. (1981) Cancer Res., 41, 2842–2848.
- Ohta, M., Greenberger, J.S., Anklesaria, P., Bassols, A. and Massagué, J. (1987) Nature, 329, 539-541.
- Perlman, D. and Halvorson, H.O. (1983) J. Mol. Biol., 107, 391-409.
- Pierschbacher, M.D. and Ruoslahti, E. (1984a) Proc. Natl. Acad. Sci. USA, 81, 5985-5990.
- Pierschbacher, M.D. and Ruoslahti, E. (1984b) Nature, 309, 30-34.
- Pircher, R., Jullien, P. and Lawrence, D.A. (1986) Biochem. Biophys. Res. Commun., 136, 30-37.
- Proudfoot, N.J. and Brownlee, G.G. (1976) Nature, 263, 211-214.
- Rosa, F., Roberts, A.B., Danielpour, D., Dart, L.L., Sporn, M.B. and Dawid, I. (1988) Science, 239, 783-785.
- Ruoslahti, E. and Pierschbacher, M.D. (1987) Science, 238, 491-497.
- Schwab, M., Alitalo, K., Varmus, H.E. and Bishop, J.M. (1983) Nature, 303, 497-501.
- Seyedin, S.M., Segarini, P.R., Rosen, D.M., Thompson, A.Y., Bentz, H. and Graycar, J. (1987) J. Biol. Chem., 262, 1946-1949.
- Smith, A.J.H. (1980) Methods Enzymol., 65, 560-580.
- Sporn, M.B., Roberts, A.B., Wakefield, L.M. and Assoian, R.K. (1986) *Science*, 233, 532-534.
- Sporn, M.B., Roberts, A.B., Wakefield, L.M. and de Crombrugghe, B. (1987) J. Cell Biol., 105, 1039-1045.
- Taylor, J.M., Illmensee, R. and Summers, S. (1976) Biochim. Biophys. Acta, 442, 324–330.
- ten Dijke, P., Hansen, P., Iwata, K.K., Pieler, C. and Foulkes, J.G. (1988) Proc. Natl. Acad. Sci. USA, 85, 4715-4719.
- Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA, 77, 5201-5205.
- Tucker, R.F., Branum, E.L., Shipley, G.D., Ryan, R.J. and Moses, H.L. (1984) Proc. Natl. Acad. Sci. USA, 81, 6757-6761.
- von Heyne, G. (1986) Nucleic Acids Res., 14, 4683-4690.
- Wakefield, L.M., Smith, D.M., Flanders, K.C. and Sporn, M.B. (1988) J. Biol. Chem., 263, 7646-7654.
- Winzler, R.J. (1973) In Li, C.I. (ed.), Hormonal Proteins and Peptides. Academic Press, New York, pp. 1-15.

Received on June 16, 1988; revised on August 8, 1988