Evolution of the transforming growth factor-beta superfamily

Pergamon

Progress in Growth Factor Research, Vol. 5, p. 99–118, 1994

0955-2235/94 $24.00

EVOLUTION OF THE TRANSFORMING GROWTH FACTOR-BETA SUPERFAMILY

David W. Burt and Andrew S. Law

Department of Cellular and Molecular Biology
AFRC Roslin Institute
Roslin, Midlothian, EH25 9PS, U.K.

Transforming growth factor β1 (TGF-β1) is the prototype of an increasingly complex superfamily of growth and differentiation factors. To date, a total of 74 TGF-β-like sequences have been published, probably representing 23 distinct genes. These sequences were obtained from mammalian, avian, amphibian and insect species, thus emphasising the ancient nature of the TGF-β superfamily peptides. This article summarises current hypotheses concerning the evolutionary history of this protein superfamily, based on the molecular phylogeny of the published sequences. Comparison of the deduced amino acid sequences leads to the definition of five main groups within the superfamily (TGF-β, Bone Morphogenetic Proteins [BMP], Anti-Müllerian Hormone [AMH], Inhibin α [INHα] and GDF-9) and six subgroups within the BMPs (60A, Decapentaplegic [dpp], Vg1, BMP-3, Inhibin β [INHβ$_{A/B}$] and nodal). This classification predicts possible phylogenetic and functional relationships among these proteins.

Keywords: Transforming growth factor, bone morphogenetic protein, phylogenetic tree, molecular evolution, sequence comparisons.

INTRODUCTION

The Transforming Growth Factors-β (TGF-βs) are a family of multifunctional peptides that controls proliferation, differentiation and other functions in many cell types [1]. Mature TGF-β1, the prototype member of the family, is a disulphide-linked dimer of two identical 112 amino acid polypeptide chains. Each chain is derived by proteolytic cleavage from the C-terminus of a 390 amino acid precursor [1]. The glycosylated dimer is secreted as a complex with other proteins in a latent form that can be activated by heat, acid or protease treatment [1]. Soon after its discovery, a number of other proteins were found to share sequence homology with TGF-β1, either throughout the whole polypeptide sequence (**The TGF-β Family**) or restricted to the C-terminal region (**The TGF-β Superfamily**). The sequence homology observed between these various proteins and the conservation of the genes encoding the various

Acknowledgements—The assistance of Messrs E. Armstrong, R. K. Field and N. Russell in the preparation of the figures is gratefully acknowledged.

TGF-β-like molecules in species as diverse as insects and man strongly support the fundamental role of these molecules in regulating basic biological processes such as growth and development.

The number of proteins being assigned to the TGF-β superfamily on the basis of their sequence similarity continues to increase. Seventy-four proteins have been described to date, isolated from ten different species and falling into 23 distinct gene types. These are listed together with their alternative names, species of origin and reference in Fig. 1. It is beyond the scope of the article to present the biology of these growth factors in any detail. A brief overview is given and the reader is encouraged to consult the appropriate references for more details.

Five members of the immediate TGF-β family have been described; they are now known as TGF-β1 through to TGF-β5 [1]. However, the common practice of naming apparently novel unidentified proteins on the basis of their observed actions, coupled with the multifunctional nature of the TGF-β proteins, has resulted in a profusion of

Names	Species	References
Anti-Mullerian hormone (AMH), Mullerian inhibiting substance (MIS)	Hu, Bo, Mu, Ra	5, 46, 47, 48
Bone morphogenetic protein 2 (BMP-2, BMP-2A)	Hu, Bo, Mu, Xe	25, 49, 50, 51
Bone morphogenetic protein 3 (BMP-3, BMP-3A), Osteogenin	Hu, Bo	25
Bone morphogenetic protein 4 (BMP-4, BMP-2B)	Hu, Bo, Mu, Ra, Xe	25, 48, 51, 52, 53, 54, 55
Bone morphogenetic protein 5 (BMP-5), short ear (*se*)	Hu, Mu	23, 56
Bone morphogenetic protein 6 (BMP-6), Vg-related-1 (*Vgr-1*)	Hu, Mu, Ra	23, 57, 58
Bone morphogenetic protein 7 (BMP-7), Ostogenic protein 1 (OP-1)	Hu, Bo, Mu, Xe	23, 49, 51, 59, 60
Ostogenic protein 2 (OP-2)	Hu, Mu, Xe*	26
Decapentaplegic gene complex (*dpp*)	Dr	14
60A protein (*60A*)	Dr	17, 18
Dorsalin-1 (*dsl-1*)	Ck, Mu	16
Growth/Differentiation factor 1 (GDF-1)	Hu, Mu	19, 20
Growth/Differentiation factor 3 (GDF-3), Vg-related-2 (*Vgr-2*)	Mu	21, 22
Growth/Differentiation factor 9 (GDF-9)	Mu	21
Inhibin α subunit (INHα)	Hu, Bo, Po, Mu, Ra	61, 62, 63, 64, 65, 66, 67
Inhibin β_A subunit (INHβ_A), Activin β_A, XTC-MIF (mesoderm-inducing factor), Erythroid differentiation factor (EDF)	Hu, Bo, Po, Mu, Ra, Ck, Xe	11, 61, 62, 63, 66, 68, 69, 70, 71
Inhibin β_B subunit (INHβ_B), Activin β_B subunit	Hu, Po, Sh, Mu, Ra, Ck, Xe	61, 63, 66, 67, 70, 71, 72, 73, 74
Nodal	Mu	13
Transforming growth factor-β1 (TGF-β1, TGF-β4)	Hu, Ce, Bo, Po, Mu, Ra, Ck	40, 75, 76, 77, 78, 79, 80, 81, 82
Transforming growth factor-β2 (TGF-β2), Glioblastoma-derived T cell suppressor factor (G-TSF), BSC-1 cell growth inhibitor (GI), Polyergin	Hu, Ce, Po, Mu, Ck, Xe	41, 83, 84, 85, 86, 87, 88, 89
Transforming growth factor-β3 (TGF-β3)	Hu, Po, Mu, Ck	90, 91, 92, 93, 94
Transforming growth factor-β5 (TGF-β5)	Xe	95
Vegetal hemisphere protein 1 (*Vg1*)	Xe	12

FIGURE 1. The TGF-β superfamily, listing members whose sequence is known. Key to species: Hu: Human; Ce: African Green Monkey; Bo: Cow; Po: Pig; Sh: Sheep; Mu: Mouse; Ra: Rat; Ck: Chicken; Xe: *Xenopus laevis;* Dr: *Drosophila melanogaster.* *Xenopus laevis sequence AC JH0690 (PIR Database) shows moderate sequence homology to the mammalian OP-2 gene (56% over 106 residues). It is therefore included in the table under this heading. However, the sequence fragment also shows significant homology to other BMP-like proteins. Exactly which member of the gene family this sequence represents remains to be determined.

alternative names. In addition, many early studies utilised impure preparations. For example, Sarcoma Growth Factor (SGF) is a mixture of TGF-β and TGF-α, a structurally unrelated protein with similar activity to the TGF-βs in certain assay systems [2]. Buffalo rat liver cells produce differentiation inhibitor (DI), a potent inhibitor of myoblast differentiation, which is a mixture of several TGF-β isoforms [3]. Other names for TGF-β family members to be found in the literature include Cartilage-Inducing Factor-A (CIF-A) and CIF-B, which correspond to TGF-β1 and TGF-β2, respectively [4].

In the wider TGF-β Superfamily, Anti-Müllerian hormone, also known as Müllerian Inhibitory Factor (MIF) or Substance (MIS), is produced by the testes and is responsible for the regression of the female Müllerian ducts in the male embryo [5]. Paradoxically, it may also play a role in the development of the female reproductive system and in the adult female ovary [5, 6]. Other members of the superfamily involved in reproductive processes include the Inhibins, heterodimeric proteins consisting of a single inhibin α chain coupled to either an inhibin β_A (Inhibin A) or an inhibin β_B chain (Inhibin B), and Activins which are homo- or heterodimers of the inhibin β subunits. Inhibins and Activins are produced in the gonads and reportedly act to modulate the secretion of Follicle Stimulating Hormone (FSH) from the pituitary gland. They may also have intragonadal paracrine actions [7, 8, 9].

Activin has recently been found to have a wider range of biological activities, including a potent mesoderm-inducing activity in amphibian animal cap assays [10] and erythroid differentiation activity [11]. Indeed, roles in early development and pattern formation have been invoked for many members of the TGF-β superfamily. *Xenopus* vegetal hemisphere protein *Vg1* also reportedly induces the overlying animal pole to form mesodermal tissue, although there is some uncertainty as to the precise nature of this effect [12]. Mouse *nodal* encodes a molecule essential for mesoderm formation and subsequent organisation of axial structures in early development [13]. The *Drosophila* decapentaplegic protein (*dpp*) participates in the establishment of dorsal–ventral specification [14,15]. Chick *dorsalin-1 (dsl-1)* is the latest member of the TGF-β superfamily to be cloned and characterised and appears to regulate the differentiation of cell types along the dorso-ventral axis of the neural tube [16]. *Drosophila* 60A gene protein may participate in the development of the embryonic gut [17, 18].

The embryonic growth factor GDF-1 may mediate cell differentiation events during early development [19, 20]. Two further GDF-like sequences, *GDF-3* and *GDF-9*, have recently been isolated [21] using degenerate oligonocleotides. Their function is, as yet, unknown. However, a sequence identical to *GDF-9*, *Vgr-2* (Vg-related-2) has also been isolated independently by cross-hybridisation at low stringency with a *Xenopus Vg1* probe [22]. The expression pattern of *Vgr-2* suggests that it is involved in the complex process of bone formation in the embryo.

This action of members of the TGF-β superfamily in bone and cartilage formation is a recurrent theme. As mentioned previously, TGF-β1 and TGF-β2 were originally described by some authors as Cartilage-Inducing Factors. In addition, the osteogenic proteins BMP-2 through BMP-7 induce cartilage and bone formation [23, 24, 25]. The related protein OP-2 is expressed early in embryogenesis and may also have bone inducing activity [26]. However, multiple functions are suggested by the fact that all BMP-like molecules are expressed during early embryonic development, as well as later in life, in a range of tissue types [27].

HUMAN Locus	Location	MOUSE Locus	Location	References
TGFB1	19q13	Tgf-b1	7 (3.3)	96, 28
TGFB2	1q41	Tgf-b2	1 (86.5)	97, 98, 28
TGFB3	14q24	Tgf-b3	12 (36)	97, 24
AMH	19p13.3	Amh	10 (49.5)	99, 29
nd	nd	*Amh-rs1*	1 (10.4)	29
nd	nd	*Amh-rs2*	13 (50)	29
nd	nd	*Amh-rs3*	nd	29
nd	nd	*Amh-rs4*	12 (43)	29
nd	nd	*Amh-rs5*	nd	29
nd	nd	*Amh-rs6*	nd	29
nd	nd	*Amh-rs7*	15 (40)	29
nd	nd	*Amh-rs8*	nd	29
INHA	2q33-qter	Inha	1 (20.5)	100, 101
INHBA	7p15-p13	Inhba	13 (19)	100, 101, 102
INHBB	2cen-q13	Inhbb	12 (48-50)	100, MM
nd	nd	pInhbb	1 (30-34)	MM
BMP2	20p12	Bmp2a	2 (75)	103, 104, 28
BMP3	4p14-q21	Bmp3	5 (60)	103, 28
BMP4	14	Bmp2b-1	14 (24.4)	105, 28
nd	nd	Bmp4a	X (32.2)	28
BMP5	6	Bmp5, *se*	9 (42)	106, 56
BMP6	6	Vgr-1	13 (28)	106, 28
BMP7	20	Bmp7	nd	106
GDF3	nd	Vgr-2	6 (57.4)	22

FIGURE 2. Chromosomal location of genes of the TGF-β superfamily in man and mouse. Key: *Amh-rs* = Amh-related sequence; nd = not determined. MM = Marti Matzuk, personal communication. The figure in brackets in the mouse gene location column represents the distance of the gene from the centromere in centimorgans (cMs).

A number of TGF-β superfamily genes have been mapped onto the genomes of mouse and man and are listed in Fig. 2. Chromosomal assignments of TGF-β superfamily members indicate that these genes have become widely dispersed during their evolution [28]. It is likely that the entire multigene family evolved from a series of gene duplications and became separated by chromosomal translocations. Interestingly, a number of apparent, uncharacterised TGF-β-like genes were identified in these mapping studies. In addition to the expected BMP-4 locus on mouse chromosome 14, an X-linked *BMP-4a* gene was also discovered [28]. Whether this gene is present in other species remains to be determined. Eight unlinked, polymorphic AMH-related loci (*Amh-rs1* through *Amh-rs8*) have also been detected [29]. These may represent novel members of the TGF-β superfamily or artefactual sequences. Until these *Amh-rs* loci are cloned and characterised, this question must remain unanswered.

Many investigators have isolated, sequenced and compared TGF-β-like sequences. These comparisons have indicated homologies between polypeptides and defined conserved peptide domains or motifs. For some groups of proteins, homology is observed throughout the precursor protein (e.g. within the TGF-β or BMP subgroups) but taking the superfamily in its entirety, sequence homology is restricted to the C-terminal mature peptide. The degree and extent of sequence homology suggests phylogenetic relationships within the TGF-β superfamily. For distantly related genes the definition of homologous (orthologous) or duplicated (paralogous) genes is often difficult; a phylogenetic study can often solve this problem. In addition to defining its history, an evolutionary classification of the TGF-β superfamily into groups and subgroups can also suggest functional relationships. This is useful in an era when PCR can rapidly identify new members of a multigene family with relative ease. Consequently, new genes are often discovered purely on the basis of sequence homology and in the total absence of functional data. Evolutionary information from sequence data may therefore allow a prediction of the possible biochemical properties of these otherwise uncharacterised proteins.

THE EVOLUTION OF THE TGF-β SUPERFAMILY

The TGF-β Superfamily

To focus the discussion on the evolution of the multigene family rather than the evolution of the separate species studied, one distinct example of each TGF-β superfamily member (the human gene wherever possible) was selected for analysis. However, in some cases it was not possible to be absolutely certain that the genes included were heterologous genes. For example, it is impossible to determine whether the *Drosophila* genes *dpp* and *60A* and the *Xenopus TGF-β5* gene represent homologues of genes present in other species or are products of recent gene duplications due to the large evolutionary distance separating these species from the others involved in the study. A multiple alignment of the selected sequences was made using the PILEUP program [30] based on the C-terminal sequence conserved in all TGF-β superfamily members. The result is shown in Fig. 3. Proteins from the TGF-β superfamily are only active as homo- or heterodimers, with the two polypeptide chains being linked by a single disulphide bond. From X-ray crystallography studies of TGF-β2 it is known that all other cysteines are involved in intrachain disulphide bonds [31,32]. This structure would therefore not be expected to accept major insertions or deletions of amino acid residues. Consequently, the inclusion of gaps in the alignment was minimised by the imposition of large gap weights. The consensus sequence is displayed beneath the alignment. Fig. 4 plots graphically the relative degree of amino acid conservation at each site across the whole of the TGF-β superfamily. Together these two figures reveal the modular nature of the TGF-β superfamily proteins with the two most highly conserved domains lying at the N- and C-terminal regions of the mature proteins. These two domains contain the seven invariant residues conserved across the entire superfamily. A lesser, but still significant, region of homology lies in the centre. These three conserved domains have long been recognised [14]. The biological significance of these domains is unclear. However, there is experimental evidence that domains II and III are important in

```
                   *                 *      * * *                                                                                        * *
Hu  TGF-β1   CCVRQLYIDFRKDLGWK-WIHEPKGYHANFCLGPCPYIWSL-------DTQYSKVLALYNQHNPGASAA-PCC--VPQALEPLPIVYYVGRK-PKV-EQLSNMIVRSCKCS
Xe  TGF-β5   CCVKPLYINFRKDLGWK-WIHEPKGYEANYCLGNCPYIWSM-------DTQYSKVLSLYNQNNPGASIS-PCC--VPDVLEPLPIIYYVGRT-AKV-EQLSNMVVRSCNCS
Hu  TGF-β3   CCVRPLYIDFRQDLGWK-WVHEPKGYYANFCSGPCPYLRSA-------DTTHSTVLGLYNTLNPEASAS-PCC--VPQDLEPLTILYYVGRT-PKV-EQLSNMVVKSCKCS
Hu  TGF-β2   CCLRPLYIDFKRDLGWK-WIHEPKGYNANFCAGACPYLWSS-------DTQHSRVLSLYNTINPEASAS-PCC--VSQDLEPLTILYYIGKT-PKI-EQLSNMIVKSCKCS

Dr  60A      CQMQTLYIDFK-DLGWHDWIIAPEGYGAFYCSGECNFPLN----AHMNATNHAIVQTLVHLLEPKK-VPKPCC--APTRLGALPVLYHLNDE-NVNLKKYRNMIVKSCGCH
Hu  OP-2     CRRHELYVSFQ-DLGWLDWVIAPQGYSAYYCEGECSFPLD----SCMNATNHAILQSLVHLMKPNA-VPKACC--APTKLSATSVLYYDSSN-NVILRKHRNMVVKACGCH
Hu  BMP-7    CKKHELYVSFR-DLGWQDWIIAPEGYAAYYCEGECAFPLN----SYMNATNHAIVQTLVHFINPET-VPKPCC--APTQLNAISVLYFDDSS-NVILKKYRNMVVRACGCH
Hu  BMP-6    CRKHELYVSFQ-DLGWQDWIIAPKGYAANYCDGECSFPLN----AHMNATNHAIVQTLVHLMNPEY-VPKPCC--APTKLNAISVLYFDDNS-NVILKKYRNMVVRACGCH
Hu  BMP-5    CKKHELYVSFR-DLGWQDWIIAPEGYAAFYCDGECSFPLN----AHMNATNHAIVQTLVHLMFPDH-VPKPCC--APTKLNAISVLYFDDSS-NVILKKYRNMVVRSCGCH

Ck  dsl-1    CRRTSLHVNFK-EIGWDSWIIAPKDYEAFECKGGCFFPLT----DNVTPTKHAIVQTLVHLQNPKK-ASKACC--VPTKLDAISILYKDDAGVPTLIYNYEGMKVAECGCR
Dr  dpp      CRRHSLYVDFS-DVGWDDWIVAPLGYDAYYCHGKCPFPLA----DHFNSTNHAVVQTLVNNMNPGK-VPKACC--VPTQLDSVAMLYLNDQS-TVVLKNYQEMTVVGCGCR
Hu  BMP-2    CKRHPLYVDFS-DVGWNDWIVAPPGYHAFYCHGECPFPLA----DHLNSTNHAIVQTLVNSVNSK--IPKACC--VPTELSAISMLYLDENE-KVVLKNYQDMVVEGCGCR
Hu  BMP-4    CRRHSLYVDFS-DVGWNDWIVAPPGYQAFYCHGDCPFPLA----DHLNSTNHAIVQTLVNSVNSS--IPKACC--VPTELSAISMLYLDEYD-KVVLKNYQEMVVEGCGCR

Xe  Vg-1     CKKRHLYVEFK-DVGWQNWVIAPQGYMANYCYGECPYPLT----EILNGSNHAILQTLVHSIEPED-IPLPCC--VPTKMSPISMLFYDNND-NVVLRHYENMAVDECGCR
Hu  GDF-1    CRARRLYVSFR-EVGWHRWVIAPRGFLANYCQGQCALPVALSGSGGPPALNHAVLRALMHAAAPGA-ADLPCC--VPARLSPISVLFFDNSD-NVVLRQYEDMVVDECGCR
Mu  GDF-3    CHRHQLFINFQ-DLGWHKWVIAPKGFMANYCHGECPFSMT----TYLNSSNYAFMQALMHMADPK--VPKAVC--VPTKLSPISMLYQDSDK-NVILRHYEDMVVDECGCG

Hu  BMP-3    CARRYLKVDFA-DIGWSEWIISPKSFDAYYCSGACQFPMP----KSLKPSNHATIQSIVRAVGVVPGIPEPCC--VPEKMSSLSILFFDENK-NVVLKVYPNMTVESCACR

Hu  INHβA    CCKKQFFVSFK-DIGWNDWIIAPSGYHANYCEGECPSHIAGTSGSSLSFHSTVINHYRMRGHSPFANL-KSCC--VPTLLRPMSMLYYDDGQ-NIIKKDIQNMIVEECGCS
Hu  INHβB    CCRQQFFIDFR-LIGWNDWIIAPTGYYGNYCEGSCPAYLAGVPGSASSFHTAVVNQYRMRGLNP-GTV-NSCC--IPTKLSTMSMLYFDDEY-NIVKRDVPNMIVEECGCA

Mu  nodal    CRRVKFQVDFN-LIGWGSWIIYPKQYNAYRCEGECPNPV----GEEFHPTNHAYIQSLLKRYQPHR-VPSTCC--APVKTKPLSMLYVD-NG-RVLLEHHKDMIVEECGCL

Hu  AMH      CALRELSVDLRAERS----VLIPETYQANNCQGVCGWP---QSDRNPRYGNHVVLLLKMQARGAALARP-PCC--VPTAYAG-KLLISLSEE-RISAHHVPNMVATECGCR

Hu  INHα     CHRVALNISFQ-ELGWERWIVYPPSFIFHYCHGGCG--LHIPPNLSLPVPGAPPTPAQPYSLLPG---AQPCCAALPGTMRPLHVRTTSDGGYSFKYETVPNLLTQHCACI

Mu  GDF-9    CELHDFRLSFS-QLKWDNWIVAPHRYNPRYCKGDCPRAV----RHRYGSPVHTMVQNIIYEKLDPS-VPRPSC--VPGKYSPLSVLTIEPDG-SIAYKEYEDMIATRCTCR

Consensus    C+++.~y~-f+.-~gw..w~~-P.gy.~.yC.G-Cp.p~.........tn+~~~q.~~+...p....p+pcc..~pt+~...s.~y.-.....~~~+.y.nm~~..CgC+
                              DOMAIN I                                              DOMAIN II                      DOMAIN III
```

FIGURE 3. Multiple sequence alignment of TGF-β-like proteins beginning at the first invariant cysteine residue through to the C-terminus. For ease of comparison, the superfamily has been separated into nine subgroups (see Fig. 5). Comparison was by means of the PILEUP program [30] with parameter settings designed to minimise the introduction of gaps to the sequences (gap weight = 6.0, length weight = 0.1). Gaps are indicated in the figure by a dash (—). Amino acids conserved in all sequences are indicated with an asterisk above the figure. The consensus sequence is shown beneath. Within the consensus sequence, uppercase letters indicate amino acids conserved in all sequences. Lowercase letters indicate residues conserved in more than 50% of the sequences whilst a period (.) indicates no clear consensus. Conserved basic residues (R, K, H) are represented by a plus sign (+), acidic residues (D, E) by a minus sign (−) and a tilde (~) indicates strongly hydrophobic residues (A, V, L, I).

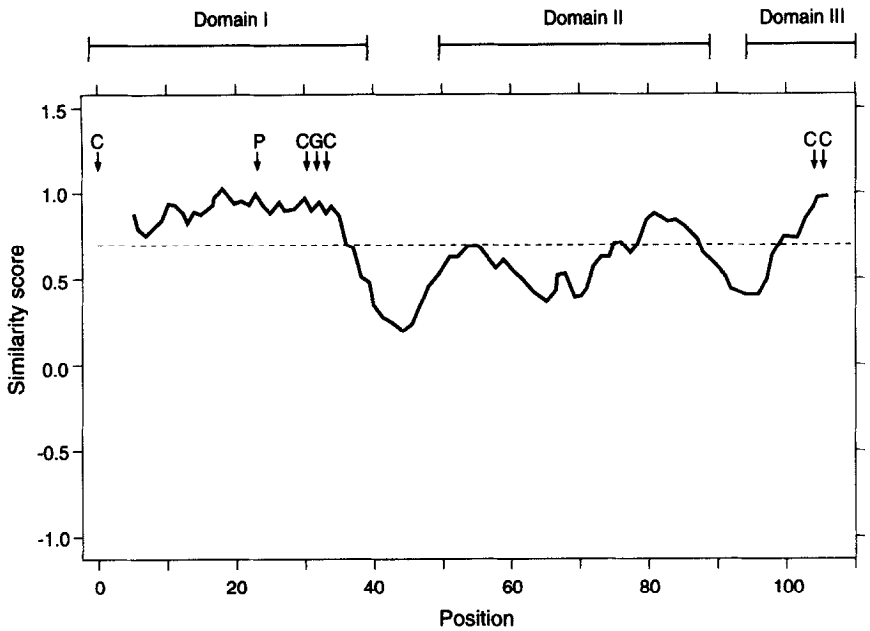

FIGURE 4. Degrees of amino acid conservation throughout the C-terminus of the 23 TGF-β superfamily proteins. The similarity plot was constructed with the PLOTSIMILARITY program [30] using the sequence alignment shown in Fig. 3. The horizontal axis represents the amino acid residue position relative to the first invariant cysteine. The vertical axis is the similarity score based on the 23 TGF-β-like sequences. Arrows indicate the sites where the amino acid residues are conserved in all proteins. See text for a discussion of the conserved domains (I, II, III).

TGF-β1 receptor binding [33]. The limited homology found in the central domain also overlaps with a region that distinguishes the specificities of the TGF-β1 and TGF-β2 proteins [34]. Within this region only 14 differences are found between TGF-β1 and TGF-β2, and presumably some or all of these residues specify this difference.

Figure 5 shows the relative degree of amino acid conservation between all the members of the TGF-β superfamily. This clearly shows the subfamily groupings, with the highest degree of conservation observed within the members of the TGF-β subfamily and the BMP groupings.

The TGF-β superfamily C-terminal sequences were used to generate a distance matrix (PROTDIST) using the Kimura 2-p correction for multiple substitutions and different transitions/transversion rates. PROTPARS, a maximum parsimony program, is often used to group together protein sequences. However, it has two major disadvantages. Firstly, PROTPARS does not correct for multiple substitutions. This is important for the analysis of distantly related genes. Secondly, the deduced tree topology is distorted by lineages with a greater than two-fold difference in rates of sequence divergence which is known to be the case for the TGF-β superfamily [35]. The distance methods FITCH and NEIGHBOR can correct for multiple substitutions and do not assume rate constancy across different lineages [36]. These methods were therefore used in this analysis.

Distance data derived from the PROTDIST program was used in the tree building program PITCH to find the single best phylogenetic tree. The result of this analysis is shown in Fig. 6. Since the length of the aligned region was short (111 amino acid

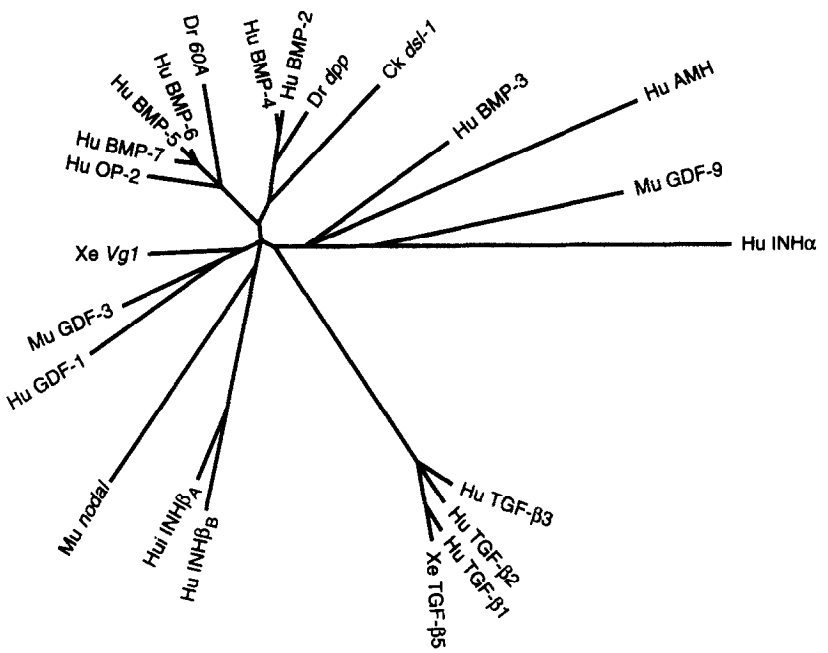

FIGURE 6. Phylogenetic tree obtained from distance matrix analysis of TGF-β superfamily C-terminal amino acid sequences. Branch lengths are drawn to scale. Units are units of distance as calculated in the distance matrix. 30,483 trees were examined with an average percentage standard deviation (APSD) of 7.404. The APSD provides an estimate of the error in branch length and position. The tree is drawn unrooted since it is impossible to determine which sequence represents the most ancient ancestral gene.

residues) and the presumed evolutionary distances large, care must be taken not to over-interpret the results obtained by this method. Consequently, a Bootstrap analysis was performed to assess the validity of the phylogenetic tree. A Bootstrap analysis is a method of repeatedly sampling the data under study to infer confidence limits for each branch of the tree and was introduced into phylogenetic studies by Felsenstein [37]. The result of the Bootstrap analysis is shown in Fig. 7. The tree described by this analysis groups the most closely related proteins together; the number at each bifurcation describes the number of times the grouping to the right of the branching point occurred as a percentage of the total number of analyses performed. However, this analysis is confounded by the large number of genes studied combined with the relative antiquity of the genes and the number of species included. A more meaningful analysis will only be possible when all the TGF-β superfamily gene sequences from a *single* species are described. However, it is clear that the divergence between the main branches illustrated in Fig. 6 is very ancient, occurring prior to the separation of insects and other arthropods.

Since the divergence rates are so variable between and within groups [35] it is not possible to use a program which assumes equal rates in all branches to derive a likely root to the phylogenetic tree. In addition, it is impossible to define an outgroup or distantly related sequence to root the TGF-β superfamily tree. The final gene tree is therefore unrooted, but serves to classify the family into groups and subgroups.

It is apparent from both methods of analysis that the TGF-β superfamily may be split into subgroupings. This is most obvious from Fig. 6. The immediate TGF-β

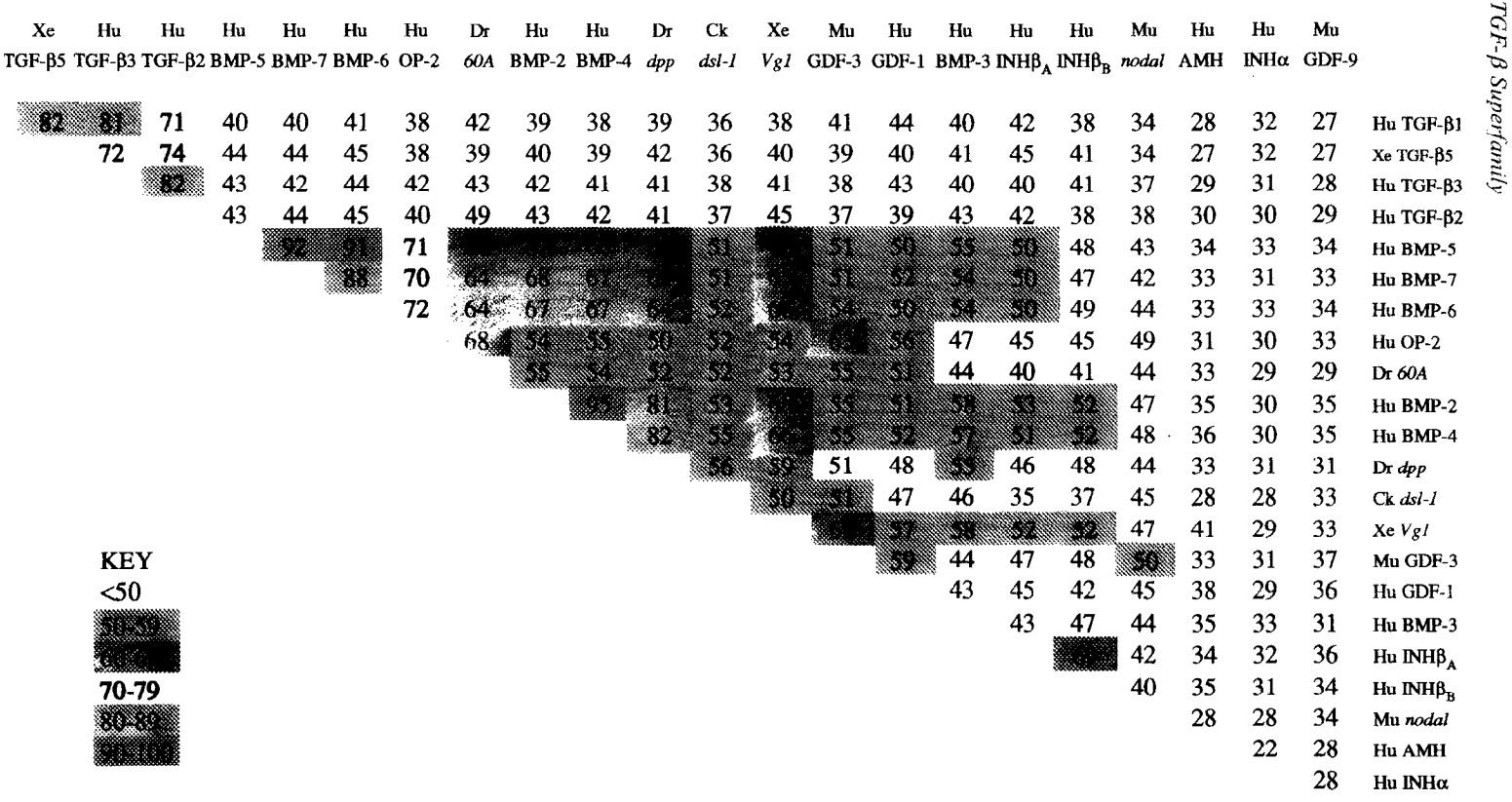

FIGURE 5. Amino acid homologies from pairwise comparisons between different members of the TGF-β superfamily. Numbers represent % amino acid identities between each pair calculated from the first invariant cysteine residue to the C-terminus, excluding gaps (see Fig. 3).

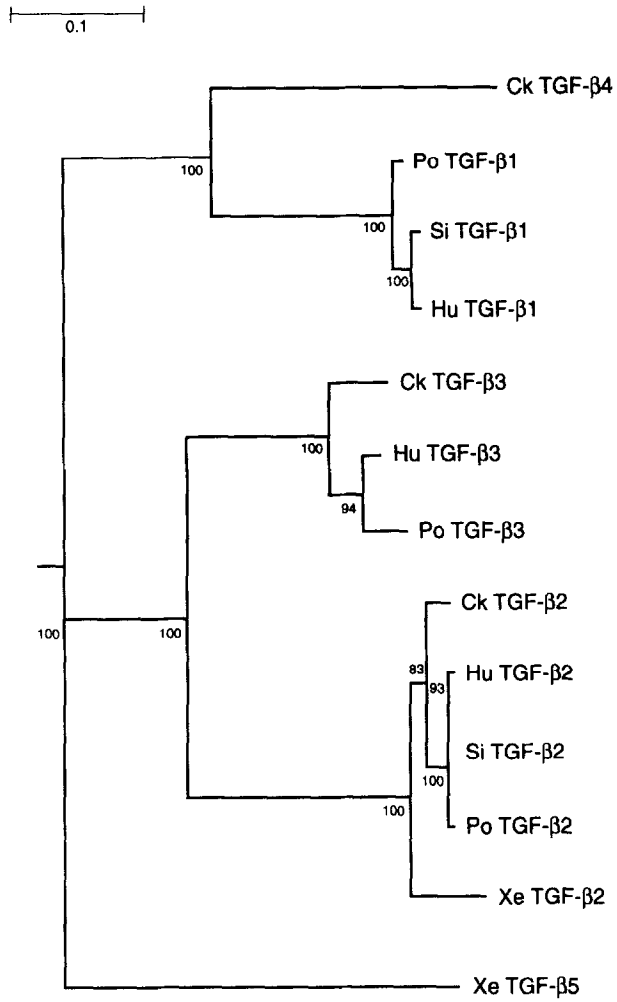

FIGURE 7. Majority-rule and strict consensus tree of TGF-β superfamily proteins derived from a bootstrapping analysis. The alignment shown in Fig. 3 was resampled using SEQBOOT (n = 100), distance matrices were calculated (PROTDIST), and a consensus tree was calculated using the program CONSENSE (PHYLIP 3.5). The number at each fork indicate the number of times the grouping to the right of that fork occurred as a percentage of the total number of data samplings. The branch lengths do not represent any scale. The tree has been arbitrarily rooted to *Xenopus laevis* TGF-β5.

sequences are clustered together as a subgroup, as are the BMP-related genes. The BMP-like genes may also be further divided into subgroupings with *BMP-5*, *-6* and *-7* more closely related to each other and, to a lesser extent, *60A* and *OP-2* than they are to *BMP-2* and *BMP-4*. *BMP-3*, whilst apparently separate from the main BMP grouping at first glance, represents a third sub-branch of this group presumably resulting from a more ancient duplication event. It is important to remember when viewing such trees that relative divergence is proportional to the length of branch between individual sequence names. Each node may be freely rotated provided that branch lengths are maintained.

TABLE 1. Pairwise comparisons of the propeptide and mature regions of the BMPs. The numbers represent % amino acid identity.

	BMP-2	BMP-4	BMP-5	BMP-7	BMP-6	OP-2	
BMP-2	-	**57**	27	29	29	27	
BMP-4	**86**	-	26	28	29	28	
BMP-5	58	56	-	**70**	**57**	**58**	**Pro-**
BMP-7	59	58	**79**	-	**75**	**57**	**region**
BMP-6	58	57	**81**	**66**	-	**56**	
OP-2	53	52	**61**	**62**	**62**	-	

Mature

BMP-related Genes

An extension of the comparison of mature peptide sequences to the propeptide region of the BMP-related proteins also verifies the tree topology. BMP-2 and BMP-4 share 86% sequence identity in the mature peptide and 57% in the proregion (Table 1). It would seem likely that BMP-2 and -4 are the products of a recent gene duplication which has occurred from a *dpp*-like gene after the divergence of insects and vertebrates 700 million years ago. Which of the two is the more ancient is impossible to determine (*c.f.* Bootstrap scores in Fig. 7). It is also clear that the chicken *dorsalin* gene is derived from this lineage, but again it is not possible to determine if this represents a gene homologous or heterologous to the *dpp* gene. *BMP-5*, *BMP-6*, *BMP-7* and *OP-2* also share a high level of sequence identity in their proregions. These genes are clearly derived from the same lineage as the *60A* gene. However, again it is impossible to determine the exact order of the duplications that led to these genes.

TGF-β1, TGF-β2 *and* TGF-β3 *Genes*

A more informative approach to determining the evolutionary history of the TGF-β family is to compare the DNA sequences that code for the precursor polypeptide sequence. The exact methodology chosen for such comparisons must be determined by the relative closeness of the genes under study. For closely related sequences, a comparison of the third base in each codon is very sensitive. Mutations in the third coding position are frequently 'silent' mutations and occur at a much higher frequency than at the first or second coding positions. Consequently, over a short period of time (in evolutionary terms), the number of mutations occurring in this position will be such that a meaningful statistical conclusion may be drawn. However, for more distantly related genes, this more rapid rate of mutation results in a near random variation at the third coding base and comparison of the first and second bases of each codon is recommended.

From a comparison of the synonymous components at the third base of the coding sequences of all the reported TGF-β sequences, it was suggested that the chicken *TGF-β4* gene was probably the avian homologue of the mammalian *TGF-β1* gene [38] (note: the sequence originally described as chicken *TGF-β1* [39] was the result of an erroneous sequencing of a porcine *TGF-β1* clone [40]). This was confirmed by Southern blotting of avian and mammalian genomic DNAs using either chicken TGF-β4 or human TGF-β1 probes. At first this seems inconsistent with the comparison of amino acid sequences. The TGF-β1 and TGF-β4 proteins show sequence identities of 45 and 82% in the pro- and mature regions, respectively. Usually, for any given TGF-β isoform, both these regions are highly conserved, typically, 80 and 95%, respectively. The lack of sequence identity in the proregion has frequently been cited as evidence that the five TGF-β proteins characterised so far (TGF-β1 through to TGF-β5) are the products of five distinct (heterologous) genes. However, the proposed TGF-β phylogeny suggests that the ancestral gene for TGF-β1 and TGF-β4 diverged almost 300 million years ago. This is approximately the time of the divergence of mammalian and avian ancestors. A common gene at this point would therefore have diverged sufficiently to create what might appear to be, at first sight, two distinct TGF-β proteins.

In that same study [38], it was suggested that the *TGF-β1* and *TGF-β3* genes shared a common ancestor approximately 300 million years ago, prior to the emergence of either the *TGF-β2* or *TGF-β5* genes. However, due to the stochastic nature of base substitutions at the third coding position and the long times of sequence divergence being considered (300–700 million years), it is possible that the position of the TGF-β genes in that proposed phylogeny may require some re-adjustment. The errors on the rates of divergence calculated are high and an examination of the first and second components may reveal a more significant relationship. Using essentially the same DNA alignment [38] but with the addition of the porcine TGF-β2 sequence [41], a distance matrix was calculated combining the information from coding bases 1 and 2. This was generated using DNADIST [36] with the Kimura 2-p model to correct for multiple base substitutions and a transition/transversion ratio of 2. The best tree was calculated using FITCH (local and global rearrangements of the tree, 10 jumbles) and is shown in Fig. 8. This tree suggests that *TGF-β1*, *TGF-β4* and *TGF-β5* may form a single group, and *TGF-β2* and *TGF-β3* another. This in turn suggests that the *Xenopus laevis TGF-β5* gene is the amphibian homologue of *TGF-β1*. This hypothesis is supported by the lack of any other TGF-β genes besides *TGF-β2* and *TGF-β5* in screens of *Xenopus laevis* TGF-β genes (Ali Hemmati-Brivanlou and Igor B. Dawid, personal communication). The lack of a *Xenopus laevis TGF-β3* gene implies that either the duplication that resulted in *TGF-β3* occurred after the divergence of amphibians or that the *TGF-β3* gene may have been deleted in this lineage. Only further studies of other amphibian species will resolve these questions.

In *Drosophila* the only *TGF-β*-like genes to be identified to date are *dpp* and *60A*, even using PCR methods [17, 18]. Primers designed specifically to amplify the TGF-βs, in preference to *dpp* or BMPs, did not permit recovery of any additional *Drosophila* sequences. Thus, if other TGF-β superfamily members exist in the *Drosophila* genome, they may require novel approaches for their isolation.

The isolation and comparison of the primary sequences of TGF-β genes from other species will provide further evidence to test the current models of TGF-β gene evolution.

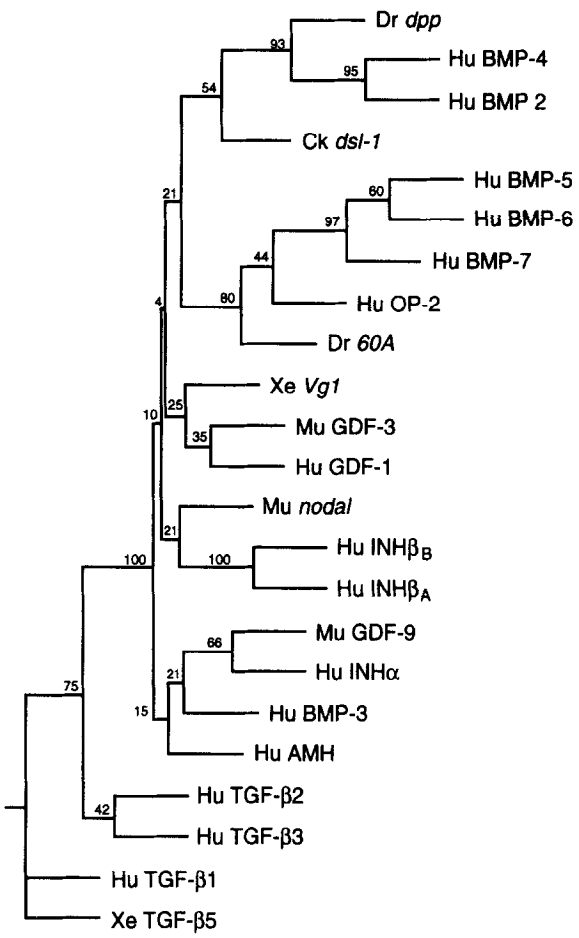

FIGURE 8. Phylogenetic tree obtained from distance matrix analysis of TGF-β nucleotide sequences, based on nucleotide substitutions at the first and second position in the aligned codons. The nucleotide alignment is identical to that previously published [38]. Branch lengths are drawn to scale in units of distance as calculated in the distance matrix. 4,265 trees were examined with an APSD of 2.605. This tree is drawn rooted to the *Xenopus laevis* TGF-β5 gene. The numbers at each fork indicate the percentage likelihood of the grouping to the right of the fork occurring in a bootstrapping analysis (n = 100 resamplings).

ANCIENT RELATIONSHIPS BETWEEN TGF-β SUPERFAMILY AND OTHER STRUCTURALLY RELATED MOLECULES

Evidence of ancient evolutionary relationships have previously been proposed on the basis of X-ray crystal structures, structural protein motifs, or conserved sequence motifs. Using the consensus TGF-β motif (I,V)xxPxx(Y,F)xxxxCxGxC defined in Fig. 3 and the program FINDPATTERNS [30], a search of protein sequence databases (SWISS-PROT release 25.0, PIR release 36.0; March 1993) was made. As expected, all the sequences listed in Fig. 1 were reported to match the pattern. No other matching sequences were found. Using an alternative approach, a search of nucleotide databases (GENBANK release 75.0, EMBL release 34.0; March 1993) with an ambiguous DNA sequence coding for this TGF-β4 motif was also performed.

Again all the previously identified TGF-β superfamily sequences were retrieved, along with two other sequence entries. The two extra matching sequences were Coxsackievirus A21 viral genome (EMBL Database AC D00538) and a *Homo sapiens* (D22S272) anonymous DNA segment (EMBL database AC Z16437) containing a (CA) repeat. Both lacked any homology outside of the TGF-β peptide motif. In addition, the human sequence is unlikely to represent an isolated exon of an unidentified TGF-β-like sequence since no classical splice donor/acceptor sites were found flanking the homology. These sequences are therefore likely to represent the false matches expected with such a database search.

Glutaredoxins (thiol-transferases)

Three regions of strong sequence conservation have been identified in the C-terminal domain of the mature TGF-β polypeptides (Figs 3 and 4). The most strongly conserved of these domains, the first, is very similar to a region found in animal glutaredoxins [42]. Out of 13 amino acid residues conserved in this region among TGF-βs, nine are also conserved in glutaredoxins. This pattern is thought to reflect a distant relationship and a shared functional specificity. It is unique to the TGF-β superfamily and the glutaredoxins and is not found elsewhere in the current sequence databases. The similarity between TGF-β superfamily polypeptides and glutaredoxins, particularly the presence of a potential dithiol active site in TGF-β, suggest a mechanism of action for ligand-induced receptor activation similar to that proposed for gonadotrophic hormones, in which dithiol–disulphide interchange reactions would be involved in the activation of the receptor [42].

pp63, Tyrosine Kinase Inhibitor

A phosphorylated *N*-glycoprotein secreted by rat hepatocytes, pp63, displays three regions of more than 70% DNA sequence identity with AMH, but fails to show any homology with other members of the TGF-β superfamily [43] or any other sequence in the protein databases. The phosphorylated form of this protein inhibits insulin receptor kinase and receptor autophosphorylation.

Nerve Growth Factor (NGF)

The recent crystallographic determination of the structure of human TGF-β2 [31, 32] uncovered an unusual protein fold. This is now thought to be very similar to a structural motif found in Nerve Growth Factor (NGF) [44, 45]. Comparison of the structures shows the topology of the four central β strands and the three pairs of disulphide bonds to be conserved, despite the low sequence homology. It remains to be seen if this is an example of convergent evolution or evidence of an ancient common ancestor.

CONCLUSION AND FUTURE DIRECTIONS

In summary, we have attempted to present models of TGF-β superfamily evolution and review the current evidence. There are many areas that remain to be explored.

For example, can we find evidence for more TGF-β-like molecules in *Drosophila* besides *dpp* and *60A*? If not, did all present day TGF-β-like sequences evolve from a *dpp/60A*-like gene? Do primitive species contain only a single TGF-β-like molecule? The current sequencing of the genomes of *Drosophila melanogaster, Caenorhabditis elegans, Saccharomyces cerevisiae* and man and searches with the TGF-β peptide motif may provide the final answer. These and many other questions need to be addressed in order that the evolution of such a diverse family of molecules can be understood. Whatever we find in these other species, it is very likely that during the next few years the description of new TGF-β-like sequences based on their homology to those already known is likely to progress much more rapidly than our knowledge about their function. At the very least, an evolutionary classification of sequence data may allow a prediction of the biochemical properties of these novel proteins.

REFERENCES

1. Roberts AB, Sporn MB. The transforming growth factor-βs. In: Sporn MB, Roberts AB, eds. *Peptide growth factors and their receptors, handbook of experimental pharmacology*. Heidelberg: Springer; 1990; 95: 419–472.
2. Massague J. Type β transforming growth factor from feline sarcoma virus-transformed rat cells. *J Biol Chem*. 1984; 259: 9756–9761.
3. Florini JR, Roberts AB, Ewton DZ, Falen SL, Flanders KC, Sporn MB. Transforming growth factor-β: a very potent inhibitor of myoblast differentiation, identical to the differentiation inhibitor secreted by buffalo rat liver cells. *J Biol Chem*. 1986; 261: 16509–16513.
4. Seyedin SM, Segarini PR, Rosen DM, Thompson AY, Bentz H, Graycar J. Cartilage-inducing factor-β is a unique protein structurally and functionally related to transforming growth factor-β. *J Biol Chem*. 1987; 262: 1946–1949.
5. Munsterberg A, Lovell-Badge R. Expression of the mouse anti-müllerian hormone gene suggests a role in both male and female sexual differentiation. *Development* 1991; 113: 613–624.
6. Lee M, Donahoe PK. Müllerian inhibiting substance: A gonadal hormone with multiple functions. *Endocr Rev*. 1992; 14: 152–164.
7. De Jong FH. Inhibin. *Physiol Rev*. 1988; 68: 555–607.
8. Mather JP, Woodruff TK, Krummen LA. Paracrine regulation of reproductive function by inhibin and activin. *Proc Soc Exp Biol Med*. 1992; 201: 1–15.
9. Ackland JF, Schwartz NB, Mayo KE, Dobson RE. Nonsteroidal signals originating in the gonads. *Physiol Rev*. 1992; 72: 731–787.
10. Ueno N, Nishimatsu S-I, Murakami K. Activin as a cell differentiation factor. *Prog Growth Factor Res*. 1990; 2: 113–124.
11. Murata M, Eto Y, Shibai H, Sakai M, Muramatsu M. Erythroid differentiation factor is encoded by the same mRNA as that of the inhibin β_A chain. *Proc Natl Acad Sci USA*. 1988; 85: 2434–2438.
12. Weeks DL, Melton DA. A maternal mRNA localised to the vegetal hemisphere in *Xenopus* eggs codes for a growth factor. *Cell* 1987; 51: 861–867.
13. Zhou X, Sasaki H, Lowe L, Hogan BLM, Kuehn MR. *Nodal* is a novel TGF-β-like gene expressed in the mouse node during gastrulation. *Nature (London)* 1993; 361: 543–547.
14. Padgett RW, St. Johnston RD, Gelbart WM. A transcript from a *Drosophila* pattern gene predicts a protein homologous to the transforming growth factor-β family. *Nature (London)* 1987; 325: 81–84.
15. Hoffmann FM. Transforming growth factor-β-related genes in *Drosophila* and vertebrate development. *Curr Opin Cell Biol*. 1991; 3: 947–952.
16. Basler K, Edlund T, Jessell TM, Yamada T. Control of cell pattern in the neural tube: regulation of cell differentiation by *dorsalin-1*, a novel TGFβ family member. *Cell* 1993; 73: 687–702.
17. Doctor JS, Jackson PD, Rashka KE, Visalli M, Hoffmann FM. Sequence, biochemical characterisation, and developmental expression of a new member of the TGF-β superfamily in *Drosophila melanogaster*. *Dev Biol*. 1992; 151: 491–505.

18. Wharton KA, Thomsen GH, Gelbart WM. *Drosophila 60A* gene, another transforming growth factor β family member, is closely related to human bone morphogenetic proteins. *Proc Natl Acad Sci USA.* 1991; 88: 9214–9218.
19. Lee S-J. Identification of a novel member (GDF-1) of the transforming growth factor-β superfamily. *Mol Endocrinol.* 1990; 4: 1034–1040.
20. Lee S-J. Expression of growth/differentiation factor-1 in the nervous system: conservation of a bicistronic structure. *Proc Natl Acad Sci USA.* 1991; 88: 4250–4254.
21. McPherron AC, Lee S-J. GDF-3 and GDF-9: Two new members of the transforming growth factor-β superfamily containing a novel pattern of cysteines. *J Biol Chem.* 1993; 268: 3444–3449.
22. Jones CM, Simon-Chazottes D, Guenet J-L, Hogan BLM. Isolation of *Vgr-2*, a novel member of the transforming growth factor-β-related gene family. *Mol Endocrinol.* 1992; 6: 1961–1968.
23. Celeste AJ, Iannazzi JA, Taylor RC, Hewick RM, Rosen V, Wang EA, Wozney JM. Identification of transforming growth factor β family members present in bone-inductive protein purified from bovine bone. *Proc Natl Acad Sci USA.* 1990; 87: 9843–9847.
24. Rosen V, Thies RS. The BMP proteins in bone formation and repair. *TIG.* 1992; 8: 97–102.
25. Wozney JM, Rosen V, Celeste AJ, Mitsock LM, Whitters MJ, Kriz RW, Hewick RM, Wang EA. Novel regulators of bone formation: molecular clones and activities. *Science* 1988; 242: 1528–1534.
26. Ozkaynak E, Schnegelsberg PNJ, Jin DF, Clifford GM, Warren FD, Drier EA, Oppermann H. Osteogenic Protein-2: A new member of the transforming growth factor-β superfamily expressed early in embryogenesis. *J Biol Chem.* 1992; 267: 25220–25227.
27. Wozney JM, Rosen V, Byrne M, Celeste AJ, Moutsatsos I, Wang EA. Growth factors influencing bone development. *J Cell Sci.*, Suppl. 1990; 13: 149–156.
28. Dickinson ME, Kobrin MS, Silan CM, Kingsley DM, Justice MJ, Miller DA, Ceci JD, Lock LF, Lee A, Buchberg AM, Siracusa LD, Lyons KM, Derynck R, Hogan BLM, Copeland NG, Jenkins NA. Chromosomal localisation of seven members of the murine TFG-β superfamily suggests close linkage to several morphogenetic mutant loci. *Genomics* 1990; 6: 505–520.
29. King TR, Lee BK, Behringer RR, Eicher EM. Mapping anti-müllerian hormone (*Amh*) and related sequences in the mouse: identification of a new region of homology between MMU10 and HSA19p. *Genomics* 1991; 11: 273–283.
30. Devereux J, Haeberli P, Smithies O. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 1984; 12: 387–395.
31. Daopin S, Piez KA, Ogawa Y, Davies DR. Crystal structure of transforming growth factor-β2: an unusual fold for the superfamily. *Science* 1992; 257: 369–373.
32. Schlunegger MP, Grutter MG. An unusual feature revealed by the crystal structure at 2.2Å resolution of human transforming growth factor-β2. *Nature (London)* 1992; 358: 430–434.
33. Flanders KC, Roberts AB, Ling N, Fleudelys NE, Sporn MB. Antibodies to peptide determinants in transforming growth factor β and their applications. *Biochemistry* 1988; 27: 739–746.
34. Qian SW, Burmester JK, Merwin JR, Madri JA, Sporn MB, Roberts AB. Identification of a structural domain that distinguishes the actions of the type 1 and 2 isoforms of transforming growth factor-β on endothelial cells. *Proc Natl Acad Sci USA.* 1992; 89: 6290–6294.
35. Burt DW. Evolutionary grouping of the transforming growth factor-β superfamily. *Biochem Biophys Res Commun.* 1992; 184: 590–595.
36. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.5c, Department of Genetics, University of Washington, Seattle, USA; 1993.
37. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; 39: 783–791.
38. Burt DW, Paton IR. Evolutionary origins of the transforming growth factor-β gene family. *DNA Cell Biol.* 1992; 11: 497–510.
39. Jakowlew SB, Dillard PJ, Sporn MB, Roberts AB. Nucleotide sequence of chicken transforming growth factor-beta 1 (TGF-β1). *Nucleic Acid Res.* 1988; 16: 8730.
40. Kondaiah P, van Obberghen-Schilling E, Ludwig RL, Dhar R, Sporn MB, Roberts AB. cDNA cloning of porcine transforming growth factor-β1 mRNAs. *J Biol Chem.* 1988; 18313–18317.
41. Zhou Y. Cloning and expression of porcine transforming growth factor beta 2 in immune cells. GenBank AC L08375. 1992; unpublished.
42. Guigo R, Smith TF. A common pattern between the TGF-β family and glutaredoxin. *Biochem J* 1991; 280: 833–834.
43. Auberger P, Falquerho L, Contreres JO, Pages G, Le Cam G, Rossi B, Le Cam A. Characterisation

of a natural inhibitor of the insulin receptor tyrosine kinase: cDNA cloning, purification and anti-mitogenic activity. *Cell* 1989; 58: 631–640.
44. Swindells MB. Structural similarity between transforming growth factor-β2 and nerve growth factor. *Science* 1992; 258: 1160–1161.
45. Daopin S, Cohen GH, Davies D. Structural similarity between transforming growth factor-β2 and nerve growth factor-response. *Science* 1992; 258: 1161–1162.
46. Cate RL, Mattaliano RJ, Hession C, Tizard R, Farber NM, Cheung A, Ninfa EG, Frey AZ, Gash DJ, Chow EP, Fisher RA, Bertonis JM, Torres G, Wallner BP, Ramachandran KL, Ragin RC, Manganaro TF, MacLaughlin DT, Donahoe PK. Isolation of the bovine and human genes for müllerian inhibiting substance and expression of the human gene in animals cells. *Cell* 1986; 45: 685–698.
47. Picard J-Y, Benarous R, Guerrier D, Josso N, Kahn A. Cloning and expression of cDNA for anti-müllerian hormone. *Proc Natl Acad Sci USA*. 1986; 83: 5464–5468.
48. Haqq C, Lee MM, Tizard R, Wysk M, DeMarinis J, Donahoe PK, Cates RL. Isolation of the rat gene for müllerian inhibiting substance. *Genomics* 1992; 12: 665–669.
49. Sampath TK, Coughlin JE, Whetstone RM, Banach D, Corbett C, Ridge RJ, Ozkaynak E, Oppermann H, Rueger DC. Bovine osteogenic protein is composed of dimers of OP-1 and BMP-2A, two members of the transforming growth factor-β superfamily. *J Biol Chem.* 1990; 13198–13205.
50. Dickinson ME. Bone morphogenetic protein homolog 2A-mouse. PIR database Accession: A34201.
51. Nishimatsu S, Suzuki A, Shoda A, Murakami K, Ueno N. Genes for bone morphogenetic proteins are differentially transcribed in early amphibian embryos. *Biochem Biophys Res Commun.* 1992; 186: 1487–1495.
52. Kurihara T, Kitamura K, Takaoka K, Nakazato H. Murine bone morphogenetic protein-4 gene: existence of multiple promoters and exons for 5′-untranslated region. EMBL database Accession: D14814.
53. Dickinson ME, Van Der Meer-De Jong R, Hogan BLM. Nucleotide sequence of the mouse bone morphogenetic protein-4 (BMP-4) cDNA. EMBL database accession: X56848.
54. Cheng D, Feng J, Feng M, Harris M, Mundy G, Harris S. Cloning and sequence of bone morphogenetic protein-4 from fetal rat calvarial cells. EMBL Database AC Z22607 1993; unpublished.
55. Koster M, Plessow S, Clement JH, Lorenz A, Tiedemann H, Knochel W. Bone morphogenetic protein-4 (BMP-4), a member of the TGF-β family, in early embryos of *Xenogus laevis*: analysis of mesoderm inducing activity. *Mech Dev.* 1991; 33: 191–200.
56. Kingsley DM, Bland AE, Grubber JM, Marker PC, Russell LB, Copeland NG, Jenkins NA. The mouse *short ear* skeletal morphogenesis lucus is associated with defects in a bone morphogenetic member of the TGF-β superfamily. *Cell* 1992; 71: 399–410.
57. Lyons K, Graycar JL, Lee A, Hashmi S, Lindquist PB, Chen EY, Hogan BLM, Derynck R. *Vgr-1*, a mammalian gene related to *Xenopus Vg-1*, is a member of the transforming growth factor-β gene superfamily. *Proc Natl Acad Sci USA*. 1989; 86: 4554–4558.
58. Sauermann U, Meyermann R, Schluesener HJ. Cloning of a novel TGF-β related cytokine, the *vgr*, from rat brain: cloning of and comparison to homologous human cytokines. *J Neurosci Res.* 1992; 33: 142–147.
59. Ozkaynak E, Rueger DC, Drier EA, Corbett C, Ridge RJ, Sampath TK, Oppermann H. OP-1 cDNA encodes an osteogenic protein in the TGF-β family. *EMBO J.* 1990; 9: 2085–2093.
60. Ozkaynak E, Schnegelsberg PNJ, Oppermann H. Murine osteogenic protein (OP-1): high levels of mRNA in kidney. *Biochem Biophys Res Commun.* 1991; 179: 116–123.
61. Mason AJ, Niall HD, Seeburg PH. Structure of two human ovarian inhibins. *Biochem Biophys Res Comm.* 1986; 135: 957–964.
62. Forage RG, Ring JM, Brown RW, McInerney BV, Cobon GS, Gregson RP, Robertson DM, Morgan FJ, Hearn MTW, Findlay JK, Wettenhall REH, Burger HG, de Kretser DM. Cloning and sequence analysis of cDNA species coding for the two subunits of inhibin from bovine follicular fluid. *Proc Natl Acad Sci USA*. 1986; 83: 3091–3095.
63. Mason AJ, Hayflick JS, Ling N, Esch FS, Ueno N, Ying S-Y, Guillemin R, Niall H, Seeburg PH. Complementary DNA sequences of ovarian follicular fluid inhibin show precursor structure and homology with transforming growth factor-β. *Nature (London)* 1985; 318: 659–663.
64. Albano RM. *M. musculus* mRNA for inhibin alpha-subunit. EMBL Database AC X69618 1992; unpublished.

65. Tone S, Katoh Y, Fujimoto H, Togashi S, Yanazawa M, Kato Y, Higashinakagawa T. Expression of inhibin alpha-subunit gene during mouse gametogenesis. *Differentiation* 1990; 44: 62–68.
66. Esch FS, Shimasaki S, Cooksey K, Mercado M, Mason AJ, Ying SY, Ueno N, Ling N. Complementary deoxyribonucleic acid (cDNA) cloning and DNA sequence analysis of rat ovarian inhibins. *Mol Endocrinol.* 1987; 1: 388–396.
67. Feng Z-M, Li Y-P, Chen C-LC. Analysis of the 5′-flanking regions of rat inhibin α- and β-B-subunit genes suggests two different regulatory mechanisms. *Mol Endocrinol.* 1989; 3: 1914–1925.
68. Albano RM, Godsave SF, Huylebroeck D, Van Nimmen K, Isaacs HV, Slack JMW, Smith JC. A mesoderm-inducing factor produced by WEHI-3 murine myelomonocytic leukemia cells is activin A. Development 1990; 110: 435–443.
69. Albano RM. *M. musculus* mRNA for inhibin beta-A subunit. EMBL database accession: X69619.
70. Mitrani E, Ziv T, Thomsen G, Shimoni Y, Melton DA, Bril A. Activin can induce the formation of axial structures and is expressed in the hypoblast of the chick. *Cell* 1990; 63: 495–501.
71. Thomsen G, Woolf T, Whitman M, Sokol S, Vaughan J, Vale W, Melton D. Activins are expressed early in *Xenopus* embryogenesis and can induce axial mesoderm and anterior structure. *Cell* 1990; 63: 485–493.
72. Mason AJ, Berkemeier LM, Schmelzer CH, Schwall RH. Activin B: precursor sequences, genomic structure and *in vitro* activities. *Mol Endocrinol.* 1989; 9: 1352–1358.
73. Rodgers RJ. Cloning of the inhibin/activin β_B subunit gene from the Booroola Merino sheep. *J Mol Endocrinol.* 1991; 6: 87–93.
74. Albano RM. *M. musculus* mRNA for inhibin beta-B subunit. EMBL database accession: X69620.
75. Derynck R, Jarrett JA, Chen EY, Eaton DH, Bell JR, Assoian RK, Roberts AB, Sporn MB, Goeddel DV. Human transforming growth factor-β complementary DNA sequence and expression in normal and transformed cells. *Nature (London)* 1985; 316: 701–705.
76. Sharples K, Plowman GD, Rose TM, Twardzik DR, Purchio AF. Cloning and sequence analysis of simian transforming growth factor-β cDNA. *DNA* 1987; 6: 239–244.
77. Van Obberghen-Schilling E, Kondaiah P, Ludwig RL, Sporn MB, Baker CC. Complementary deoxyribonucleic acid cloning of bovine transforming growth factor-β1. *Mol Endocrinol.* 1987; 1: 693–698.
78. Derynck R, Rhee L. Sequence of the porcine transforming growth factor-beta precursor. *Nucleic Acid Res.* 1987; 7: 3187.
79. Derynck R, Jarrett JA, Chen EY, Goeddel DV. The murine transforming growth factor-β precursor. *J Biol Chem.* 1986; 261: 4377–4379.
80. Qian SW, Kondaiah P, Roberts AB, Sporn MB. cDNA cloning by PCR of rat transforming growth factor-β1. *Nucleic Acids Res.* 1990; 18: 3059.
81. Burt DW, Jakowlew SB. Correction: A new interpretation of a chicken transforming growth factor-β4 complementary DNA. *Mol Endocrinol.* 1992; 6: 989–992.
82. Jakowlew SB, Dillard PJ, Sporn MB, Roberts AB. Complementary deoxyribonucleic acid cloning of a messenger ribonucleic acid encoding transforming growth factor-β4 from chicken embryo chondrocytes. *Mol Endocrinol.* 1988; 2: 1186–1195.
83. de Martin R, Haendler B, Hofer-Warbinek R, Gaugitsch H, Wrann M, Schlusener H, Seifert JM, Bodmer S, Fontana A, Hofer E. Complementary DNA for human glioblastoma-derived T cell suppressor factor, a novel member of the transforming growth factor-β gene family. *EMBO J.* 1987; 6: 3673–3677.
84. Madisen L, Webb NR, Rose TM, Marquardt H, Ikeda T, Twardzik D, Seyedin S, Purchio AF. Transforming growth factor-β2: cDNA cloning and sequence analysis. *DNA* 1988; 7: 1–8.
85. Hanks SK, Armour R, Baldwin JH, Malsonado F, Spiess J, Holley RW. Amino acid sequence of the BSC-1 cell growth inhibitor (polyergin) deduced from the nucleotide sequence of the cDNA. *Proc Natl Acad Sci USA.* 1988; 85: 79–82.
86. Miller DA, Lee A, Pelton RW, Chen EY, Moses HL, Derynck R. Murine transforming growth factor-β2 cDNA sequence and expression in adult tissues and embryos. *Mol Endocrinol.* 1989; 3: 1108–1114.
87. Burt DW, Paton IR. (1991). Molecular cloning, primary structure and evolution of the chicken transforming growth factor-β2 gene. *DNA Cell Biol.* 1991; 10: 723–734.
88. Jakowlew SB, Dillard PJ, Sporn MB, Roberts AB. Complementary deoxyribonucleic acid cloning of an mRNA encoding transforming growth factor-β2 from chicken embryo chondrocytes. *Growth Factors* 1990; 2: 123–133.

89. Rebbert ML, Bhatia-Dey N, Dawid IB. The sequence of TGF-β2 from *Xenopus laevis*. *Nucleic Acid Res.* 1990; 18: 2185.
90. Derynck R, Lindquist PB, Lee A, Wen D, Tamm J, Gracar JL, Rhee L, Mason AJ, Miller DA, Coffey RJ, Moses HL, Chen EY. A new type of transforming growth factor-β, TGF-β3. *EMBO J.* 1988; 7: 3737–3743.
91. ten Dijke P, Hansen P, Iwata KK, Pieler C, Foulkes JG. Identification of another member of the transforming growth factor-β gene family. *Proc Natl Acad Sci USA.* 1988; 85: 4715–4719.
92. Denhez F, Lafyatis R, Kondaiah P, Roberts AB, Sporn MB. Cloning by polymerase chain reaction of a new mouse TGF-β, mTGF-β3. *Growth Factors* 1990; 3: 139–146.
93. Miller DA, Lee A, Matsui Y, Chen EY, Moses HL, Derynck R. Complementary DNA cloning of the murine transforming growth factor-β3 (TGF-β3) precursor and the comparative expression of TGF-β3 and TGF-β1 messenger RNA in murine embryos and adult tissues. *Mol Endocrinol.* 1989; 3: 1926–1934.
94. Jakowlew SB, Dillard PJ, Kondaiah P, Sporn MB, Roberts AB. Complementary deoxyribonucleic acid cloning of a novel transforming growth factor-β messenger ribonucleic acid from chick embryo chondrocytes. *Mol Endocrinol.* 1988; 2: 747–755.
95. Kondaiah P, Sands MJ, Smith JM, Fields A, Roberts AB, Sporn MB, Melton DA. Identification of a novel transforming growth factor-β (TGF-β5) mRNA in *Xenopus laevis*. *J Biol Chem.* 1990; 265: 1089–1093.
96. Fujii D, Brissenden JE, Derynck R, Francke U. Transforming growth factor-β maps to human chromosome 19 long arm and to mouse chromosome 7. *Somatic Cell Mol Genet.* 1986; 12: 281–288.
97. Barton DE, Foellmer BE, Du J, Tamm J, Derynck R, Francke U. Chromosomal mapping of genes for transforming growth factors β2 and β3 in man and mouse: dispersion of TGF-β gene family. *Oncogene Res.* 1988; 3: 323–331.
98. Nishimura DY, Purchio AF, Murray JC. Linkage localisation of TGFB2 and the human homeobox gene HLX1 to chromosome 1q. *Genomics* 1993; 15: 357–364.
99. Cohen-Haguenauer O, Picard JY, Mattei MG, Serero S, Van Cong N, DeTand MF, Guerrier D, Horscayla MC, Josso N, Frezal J. Mapping of the gene for anti-müllerian hormone to the short arm of human chromosome 19. *Cytogenet Cell Genet.* 1987; 44: 2–6.
100. Barton DE., Yang-Feng TL, Mason AJ, Seeburg PH, Francke, U. Mapping of genes for inhibin subunits α, β_A, and β_B on human and mouse chromosomes and studies of *jsd* mice. *Genomics* 1989; 5: 91–99.
101. Barton DE, Yang-Feng TL, Mason AJ, Seeburg PH, Francke U. INHA, INHBA, and INHBB, the loci for the three subunits of inhibin, mapped in mouse and man. *Cytogent Cell Genet.* 1986; 46: 578.
102. Justice MJ, Silan CM, Seci JD, Buchberg AM, Copeland NG, Jenkins NA. A molecular genetic linkage map of mouse chromosome 13 anchored by the beige (*bg*) and satin (*sa*) loci. *Genomics* 1990; 6: 341–351.
103. Tabas JA, Zasloff M, Wasmuth JJ, Emanuel BS, Altherr MR, McPherson JD, Wozney JM, Kaplan FS. Bone morphogenetic protein: chromosomal localisation of human genes for BMP1, BMP2A and BMP3. *Genomics* 1991; 9: 283–289.
104. Rao VVNG, Loffer C, Wozney JM, Hansmann I. The gene for BMP2A is localised to human chromosome 20p12 by FISH. Submitted for publication (cited in Ref. 103).
105. Tabas JA, Hahn GV, Cohen RB, Seaunez HN, Modi WS, Wozney JM, Zasloff MA, Kaplan FS. Bone morphogenetic protein (BMP): Chromosomal assignment of the human gene for BMP4. *Clin Orthop Relat Res.* (in press).
106. Hahn GV, Cohen RB, Wozney JM, Levitz CL, Shore EM, Zasloff MA, Kaplan FS. A bone morphogenetic protein subfamily: chromosomal localisation of human genes for BMP5, BMP6 and BMP7. *Genomics* 1992; 14: 759–762.