

UC Irvine

UC Irvine Previously Published Works

Title

The Twisted gastrulation family of proteins, together with the IGFBP and CCN families, comprise the TIC superfamily of cysteine rich secreted factors

Permalink

<https://escholarship.org/uc/item/3wm151vg>

Journal

Molecular Pathology, 54(5)

ISSN

1366-8714

Authors

Vilmos, P
Gaudenz, K
Hegedus, Z
[et al.](#)

Publication Date

2001-10-01

DOI

10.1136/mp.54.5.317

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

The Twisted gastrulation family of proteins, together with the IGFBP and CCN families, comprise the TIC superfamily of cysteine rich secreted factors

P Vilmos, K Gaudenz, Z Hegedus, J L Marsh

Abstract

Aims—To analyse the similarities between the Twisted gastrulation (TSG) proteins known to date; in addition, to determine phylogenetic relations among the TSG proteins, and between the TSGs and other protein families—the CCN (for example, CCN2 (CTGF), CCN1 (CYR61), and CCN3 (NOV)) and IGFBP (insulin-like growth factor binding protein) families.

Methods—TBLASTN and FASTA3 were used to identify new tsg genes and relatives of the TSG family. The sequences were aligned with ClustalW. The predictions of sites for signal peptide cleavage, post-translational modifications, and putative protein domains were carried out with software available at various databases. Unrooted phylogenetic trees were calculated using the UPGMA method.

Results—Several tsg genes from vertebrates and invertebrates were compared. Alignment of protein sequences revealed a highly conserved family of TSG proteins present in both vertebrates and invertebrates, whereas the slightly less well conserved IGFBP and CCN proteins are apparently present only in vertebrates. The TSG proteins display strong homology among themselves and they are composed of a putative signal peptide at the N-terminus followed by a cysteine rich (CR) region, a conserved domain devoid of cysteines, a variable midregion, and a C-terminal CR region. The most striking similarity between the TSGs and the IGFBP and CCN proteins occurs in the N-terminal conserved cysteine rich domain and the characteristic 5' cysteine rich domain(s), spacer region, and 3' cysteine rich domain structure.

Conclusion—The family of highly conserved TSG proteins, together with the IGFBP and CCN families, constitute an emerging multigene superfamily of secreted cysteine rich factors. The TSG branch of the superfamily appears to predate the others because it is present in all species examined, whereas the CCN and IGFBP genes are found only in vertebrates.

(J Clin Pathol: Mol Pathol 2001;54:317–323)

For several years, the Twisted gastrulation (TSG) protein of *Drosophila melanogaster* stood alone with no close relatives in other species.¹ The recent identification of genes related to tsg reveals that the drosophila tsg gene is not an evolutionary oddity of insects but represents the first example of an emerging family of proteins that is structurally and functionally conserved in both vertebrates and invertebrates.^{2–8} Here, we examine the evolutionary relations among the members of this family and place the TSG family in the growing superfamily of secreted cysteine rich factors, which include the insulin-like growth factor binding protein (IGFBP) and CCN families of proteins.

tsg was initially identified as one of the seven dorsal group genes that pattern the fly embryo.^{9–14} We now know that early dorsal–ventral cell fate decisions in fly, frog, and fish embryos are regulated by members of the bone morphogenetic protein (BMP) family of growth and differentiation factors. Activity of BMP ligands is regulated in the extracellular space by several secreted factors including TSG. By itself, TSG exhibits little or no activity, but together with Short gastrulation (Sog)¹⁵ or the vertebrate homologue Chordin (Chd),¹⁶ TSG potentiates the inhibitory action of Sog/Chd to create a complex that is a powerful antagonist of BMP activity.^{2–4 6 7} The metalloprotease BMP1 protein in vertebrates¹⁷ and its fly homologue, Tollid (Tld),¹⁸ act on the Sog–TSG–BMP complex to cleave Sog and thereby regulate ligand activity by releasing it from this complex.^{19 20} Thus, TSG can act together with Sog and Tld to modulate BMP signalling.^{2–8}

In our study, we describe the growing family of tsg and tsg related genes to show that the TSG proteins constitute an emerging branch of a growing superfamily of secreted, cysteine rich factors in vertebrates. The two other branches of this superfamily are the family of IGFBPs²¹ and the CCN family of growth factors²² (named after the cysteine rich (CCN1; CYR61), the connective tissue growth factor (CCN2; CTGF), and the nephroblastoma overexpressed (CCN3; NOV) proteins). Here, we refer to the TIC superfamily, which is so named after the subfamilies: TSG, IGFBP, CCN. The members of this superfamily are all secreted, cysteine rich factors that share a common conserved N-terminal, cysteine rich domain and a similar overall topology by exhibiting a characteristic conserved 5' cysteine

Department of
Developmental and
Cell Biology,
University of
California Irvine,
Irvine, CA 92697, USA
P Vilmos
K Gaudenz
J L Marsh

Biological Research
Center of the
Hungarian Academy of
Sciences, Molecular
Modeling Laboratory,
Szeged 6701, Hungary
Z Hegedus

Correspondence to:
Dr Marsh
jmarsh@uci.edu

Accepted for publication
1 May 2001

Keywords: twisted gastrulation; insulin-like growth factor binding protein; CCN; tsg

Table 1 List of TSG family members known to date

Name	Species	Genbank accession number	Ref.	Signal*
HsTSG	<i>Homo sapiens</i> (human)	AF196834	2	25/26
MmTSG	<i>Mus musculus</i> (mouse)	MMU297390	2	24/25
GgTSG	<i>Gallus gallus</i> (chicken)	AF255731	3	26/27
DrTGS	<i>Danio rerio</i> (zebrafish)	DRE297393	2	24/25
DrTSG2	<i>Danio rerio</i> (zebrafish)	AF267692	2,3	26/27
XITSG	<i>Xenopus laevis</i> (frog)	AF245221	6	25/26
DmTSG	<i>Drosophila melanogaster</i> (fruit fly)	U09808	1	21/22
DmTSG2	<i>Drosophila melanogaster</i> (fruit fly)	CG12410	2	26/27
DmTSG3	<i>Drosophila melanogaster</i> (fruit fly)	CG11582	NCBI	–
DhTSG	<i>Drosophila hydei</i> (fly)	Unpublished		26/27
DnTSG	<i>Drosophila navajoa</i> (fly)	Unpublished		17/18
DvTSG	<i>Drosophila virilis</i> (fly)	Unpublished		16/17
BmTSG	<i>Bombyx mori</i> (silkworm)	AU006229	NCBI	20/21

*Amino acid positions of predicted signal peptide cleavage sites.

The unpublished data are from JC Becerra, 1998, PhD Thesis, University of California Irvine, USA.

rich domain(s), a non-conserved spacer, and a 3' cysteine rich domain.

Materials and methods

TBLASTN at NCBI (<http://www.ncbi.nlm.nih.gov>), Flybase (<http://flybase.bio.indiana.edu>), and FASTA^{3,23} at the European Bioinformatics Institute (<http://www.ebi.ac.uk/fasta3/>) were used to identify new tsg genes in addition to those already physically described. The sequences of the 13 TSG-like proteins were aligned using ClustalW²⁴ of MacVector 7.0. The predictions of sites for signal peptide cleavage,²⁵ post-translational modification, and putative protein domains (profilescan) were carried out with software available at the Technical University of Denmark (<http://www.cbs.dtu.dk/services/SignalP/>), the Baylor College of Medicine search launcher (<http://dot.imgen.bcm.tmc.edu/>), and the EXPASY proteomics server of the Swiss Institute of Bioinformatics (<http://expasy.cbr.nrc.ca/>). TBLASTN and FASTA³ were used to identify proteins with domains or motifs similar to TSG. Unrooted phylogenetic trees were calculated with the UPGMA and neighbour joining methods of MacVector 7.0, and an optimised alignment (Multiclustal, and ClustalW) was used to build a tree with the Fitch-Margoliash method of Felsenstein's phylogenetic package (PHYLIP, version 3.57c).

Results

THE TSG FAMILY

The first tsg gene was described in *Drosophila*^{14, 26} and was thought to play a role in BMP signalling.^{1, 27–29} Recently, TSG homologues have also been identified in several vertebrate species (table 1). Cross phyla expression experiments have demonstrated that TSGs from species as different as humans, mice, fish, frogs, and flies are functionally equivalent and able to synergise with Sog to antagonise BMP activity.² The human tsg gene maps to 18p11.2–3, and may be included in the minimum critical region for Holoprosencephaly 4 (HPE4).² The mouse gene maps to the syntenic region 17E1.3–E2, whereas the zebrafish tsg is located at linkage group 24–74,5, which is also syntenic to the human locus, further supporting the view that all three genes are probably functional orthologues.² In

Drosophila, TSG is required for proper establishment of the dorsal–ventral (D/V) axis. This role is shared with *Xenopus* TSG, which is expressed on the ventral side of the embryo during gastrulation and exhibits ventralising activity by regulating the Chd–BMP pathway,^{3–5} and with zebrafish where loss of TSG leads to dorsalised embryos. Thus, Tsgs from diverse species are required for D/V patterning in the embryo and/or share a common biochemical activity in being able to interact with Sog to antagonise BMP activity.

Our searches using TBLASTN of all uncharacterised sequences filed in NCBI and Flybase, and FASTA versus the Swissprot Database revealed the human homologue of tsg, the tsg gene in *Bombyx mori* (silkworm), and a second (tsg2) and third (tsg3) tsg gene in *D. melanogaster*. Unicellular organisms do not possess BMP-like genes and therefore it was not surprising that extensive searches of microbial genomes of 88 species did not reveal proteins with significant homology to TSG. In addition, database searches of the genome of *Caenorhabditis elegans* revealed no putative tsg homologue, although there are BMP homologues in worms.

The alignment of the protein sequences (fig 1) showed that TSGs display strong homology, with approximately 50% of 202 amino acid residues matching in 12 species, and with higher similarity in the conserved domains—for example, 76% in the CR1 domain. All TSG proteins have a similar topology, with a putative signal peptide at their N-terminus, followed by a highly conserved cysteine rich region (N-term CR or domain I) and a conserved domain devoid of cysteines (domain II) (fig 1). A variable region connects domain II to a cysteine rich C-terminal conserved domain (C-term CR or domain III). Domain II and the variable region constitute the midregion of the protein that can serve as a hinge between the two CR domains. In insects, domain III is followed by a C-terminal tail of about 40 amino acids. The putative TSG3 protein from *D. melanogaster* is composed of domain II and CR2 only. Its map position places it very close to the shrew (srw) gene (Flybase ID: FBgn0003508), another member of the dorsal group genes^{28, 30} thought to be involved in BMP signalling.³¹ Table 1 summarises the predicted cleavage sites of the signal peptides.

Interestingly, where known, the genomic organisation of the tsg genes closely parallels the groupings suggested by sequence similarity. For example, the first *Drosophila* tsg gene is encoded by a short intronless transcription unit. However, the human and *Drosophila hydei* tsg genes and the *D. melanogaster* tsg2 gene have three putative introns (fig 1). Unlike the CCN proteins, where conserved structural domains are encoded by separate exons,³² the protein domains of the tsg genes are not defined by intron–exon boundaries. However, the second intron in the human gene and the first introns in flies are adjacent to the C-terminus of the N-terminal CR domain. Moreover, they are at the same position in all three genes, suggesting that this intron appeared before the divergence of invertebrates and vertebrates. Similarly, the

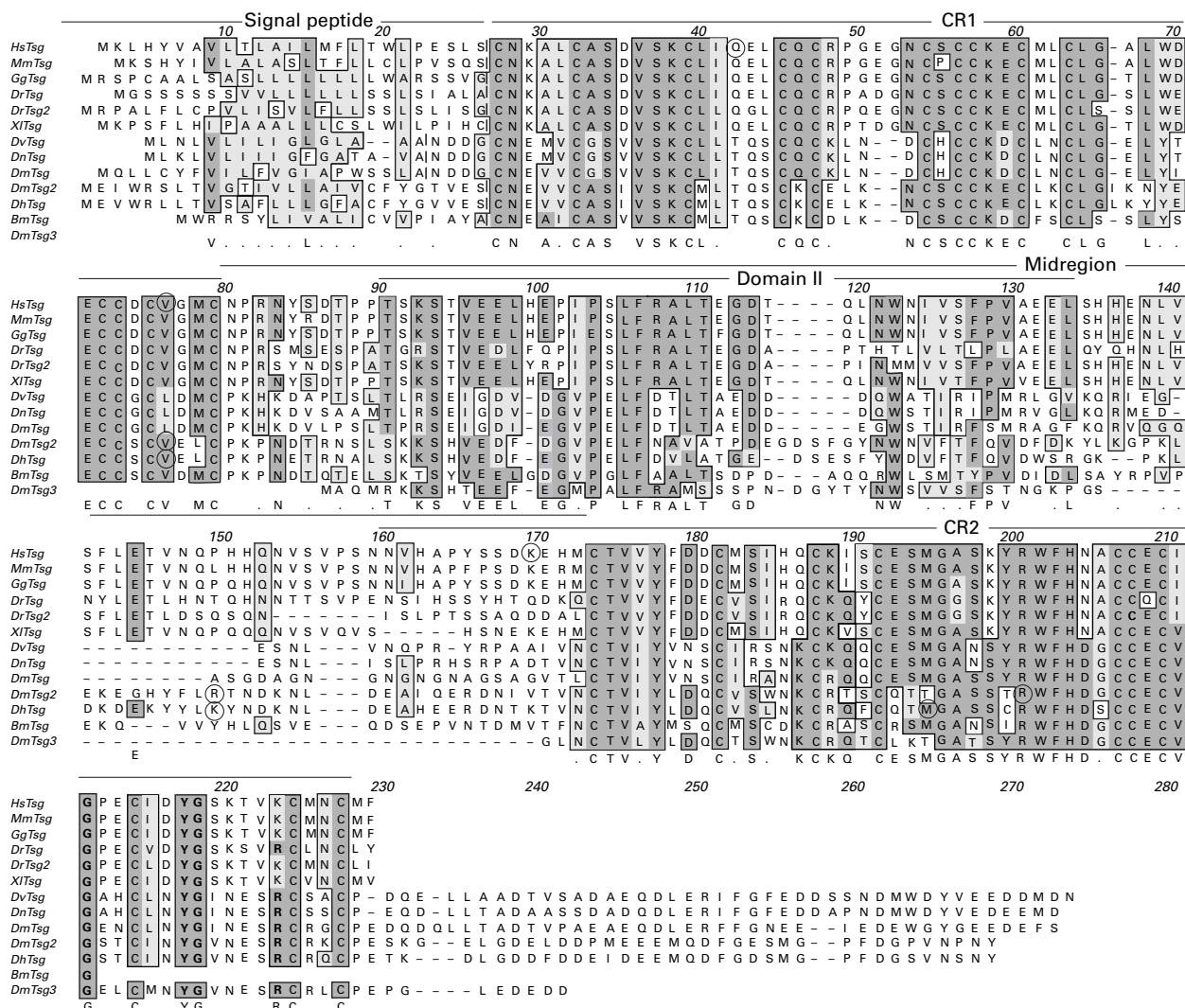


Figure 1 Multiple sequence alignment of Twisted gastrulation (TSG) proteins. Identical amino acids are darkly shaded and similarities lightly shaded. Gaps are represented by a dash (-). A consensus sequence is listed below the alignment. Amino acids at the 3' end of exons are circled, cleaved sites of signal peptides are indicated by a vertical line. See table 1 for abbreviations.

integration of the second intron into the fly genes happened in the common ancestor of *D melanogaster* and *D hydei*, some 40–60 million years ago.³³ The genomic structure of the fish genes is not yet available, but it will be interesting to compare these because the protein sequence suggests independent duplication events in fish and flies.

We have searched for proteins with similarity to the individual domains or motifs of TSGs. At present, only the IGFbps and CCN family members share their general topology with TSG.^{1 21} The cysteine rich domains also bear a weak resemblance to the CR domains of Sog/Chd and the von Willebrand factor.³⁴ In contrast, no homologue of domain II was found in the protein databases.

The similarity between the TSG proteins and the IGFBP and CCN proteins is mainly the result of the arrangement of cysteines in the conserved N-terminal cysteine rich region (fig 2). The N-terminal domain is consistent in size, ranging from 70 to 93 amino acids in all three families. The conserved cysteine pattern in this

region of the TSGs, is: C-x(4)-C-x(6)-C-x(5)-C-x-C-x(4,6)-C-x-CC-x(2)-C-x(7,8)-CC-x-C-x(3)-C. The domain contains a conserved local motif (xC-x-CC-x(2)-C-x(3)-LG) (fig 2), a portion of which is shared with the IGFBP and CCN proteins, namely: GCGCC-x(2)-C. The GCGCC-x(2)-C motif may be important for the interaction with IGFs.¹⁸ In the C-terminal CR domain of the CCN proteins, there is an unusual arrangement of six cysteines linked to form a “cysteine knot” conformation.³⁵ The C-terminal cysteine knot (CTCK) motif is: C-C-x(13)-C-x(2)-[GN]-x(12)-C-x-C-x(2,4)-C. This structural motif is believed to be involved in disulphide linked dimerisation. The CR2 domain in the TSG proteins has a related pattern: C-C-x-C-x(4)-C-x(9)-C-x(2)-C, raising the possibility of dimerisation.

All protein sequences and the consensus sequence were examined for possible conserved modification sites. Putative N-glycosylation sites occur at the sequence NCS near the N-terminus (at approximately amino acid 53) in all vertebrate TSGs and the *D melanogaster* and *D hydei*

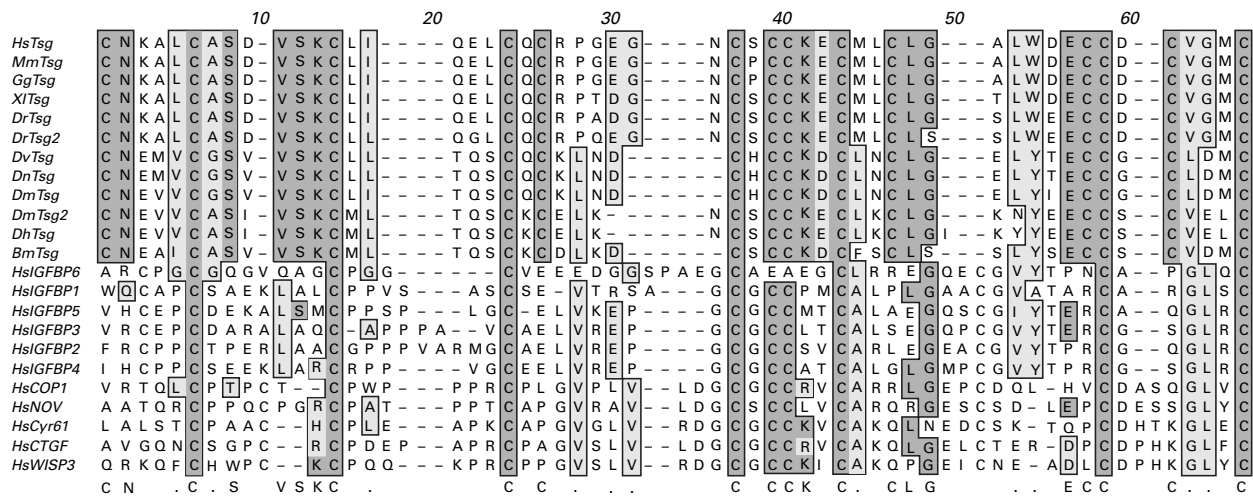


Figure 2 ClustalW alignment of the N-terminal cysteine rich domains from Twisted gastrulation (TSG) and human representatives of the insulin-like growth factor binding proteins (IGFBPs) and CCN proteins. Identical amino acids are darkly shaded and similarities lightly shaded. Gaps are represented by a dash (-). A consensus sequence is listed below the alignment. A conserved local motif is underlined below the consensus sequence. See table 1 for abbreviations.

tsg2 genes. A second possible N-linked glycosylation site occurs near the C-terminus (NES, positions 199–202 in *D melanogaster*) of all the insect proteins. Sequences in the cysteine free midregion of IGFBP-1, IGFBP-3, and IGFBP-5 have been identified as targets of phosphorylation by protein kinase C (PKC) and casein kinase II (CKII).^{36–38} Putative PKC and CKII phosphorylation sites occur at similar positions in both insect and vertebrate TSGs. A conserved putative PKC phosphorylation site (ST)X(RK) occurs in all TSG proteins in domain II (for example, TSK at position 90–93 in humans). Two putative CKII target sites [(ST)XX(DE)] occur nearby. All TSG proteins have a putative CKII site at the end of domain II (for example, TEGD at position 110–113 in human TSG) and all except *D melanogaster*, *D virilis*, and *D navajoa* TSGs have a second putative site at 93–96 (for example, STVE in human TSG). By comparison, all of the IGFBPs have putative PKC and CKII sites downstream of the first cysteine rich domain (exceptions are human IGFBP-2 which has neither site, human IGFBP-1, which has just a CKII site, and human IGFBP-6, which has just the PKC site). The CCN family shows only sporadic examples of putative sites for either of these effectors. A limited number of other putative sites are shared by subsets of the TSGs. For example, all TSGs have a second putative PKC site in the conserved C terminal CR2 domain, but the location is slightly different in vertebrates than in insects. In all the insects, a site occurs at 198—for example, STR in *D melanogaster* TSG, whereas in all vertebrates, a site occurs at position 221—for example, TVK in human TSG. It remains to be tested whether these conserved sites are real targets of modification.

THE IGFBP FAMILY

IGFBPs are cysteine rich proteins with conserved N-domains and C-domains. They appear to be more functionally diverse than the TSGs, having both IGF dependent and independent activities.²¹ They exhibit growth inhibitory effects by competitively binding

IGFs, thereby preventing them binding to the IGF receptor.³⁹ Conversely, they slowly release IGF for receptor interactions, thus protecting the receptor from downregulation by high IGF exposure.^{40–41} Tyr60 in the N-terminal CR domain is one of the contact points with IGF,⁴² and the GCGCC-xx-C motif is thought to be important for the interaction with IGFs.⁴³ The variable midregion contains sites for glycosylation,⁴⁴ proteolysis by IGFBP proteases,⁴⁵ and phosphorylation by PKC and CKII.³⁶ It has been suggested that phosphorylation might influence the interaction with the extracellular matrix but its exact role is unclear. The C-terminal CR domain was found to be essential for IGF binding,⁴⁶ and contains the integrin interaction motif (RGD) in IGFBP-1 and IGFBP-2.⁴⁶ Similar RGD sequences do not appear in the TSG proteins. IGFBP-3 and IGFBP-5 also possess a nuclear localisation signal⁴⁷ and indeed, upon binding, IGFBP-3 can be found in the nucleus.⁴⁸ Heparin binding motifs are also found in the C-terminal domain of IGFBP-3, IGFBP-5, and IGFBP-6⁴⁹; in addition, the drosophila TSG can bind heparin.²⁷ The cysteines in the C-terminal domain of IGFBP-2 form intradomain disulfide bonds.⁵⁰

In addition to its IGF dependent actions, IGFBP-3 inhibits growth in an IGF independent fashion, and breast cancer cell surface proteins (BCCSPs) of 20, 26, and 50 kDa were identified that can bind IGFBP-3 in a dose dependent manner.⁵¹ The midregion of the IGFBP-3 molecule was identified as the site for binding to these BCCSPs.⁵² Moreover, affinity crosslinking studies have shown that a 400 kDa receptor protein (type V receptor) has high affinity for both transforming growth factor β and IGFBPs;⁵³ and IGFBP-3 was shown to inhibit cell growth by interacting with the type V receptor in a Smad independent manner.⁵³ The binding is mediated by the motif W/RXXD, which is located at the end of a solvent accessible loop and is shared with IGFBPs 3–6. Interestingly, all the insect TSGs have a

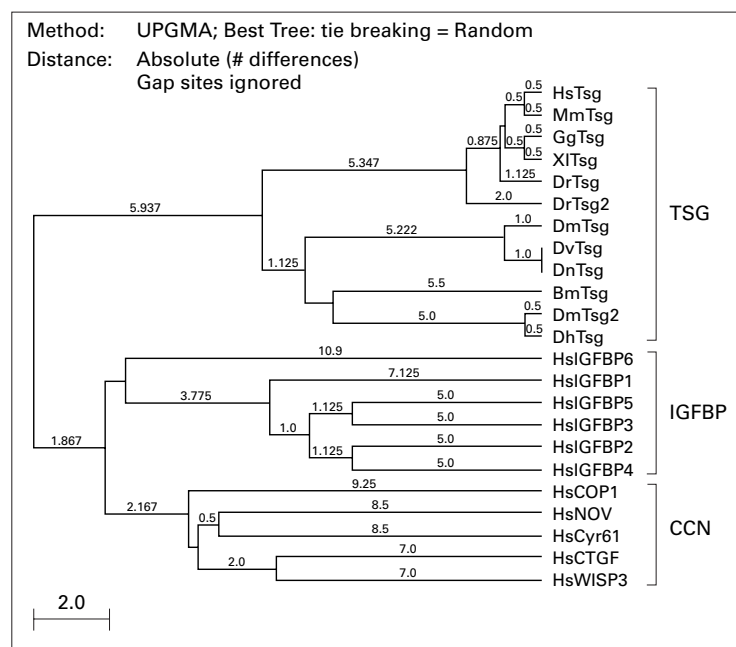


Figure 3 Unrooted dendrogram showing the phylogenetic relations between the N-terminal cysteine rich domains of the TIC (Twisted gastrulation (TSG), insulin-like growth factor binding protein (IGFBP), CCN) superfamily. Numbers indicate the total number of differences between sequences. See table 1 for abbreviations.

similar motif (WFHD) at a similar location in their CR2 domain, although the vertebrate proteins have N in place of D. Recent studies demonstrate that the actions of IGFBP-3 and IGFBP-5 can be modulated by IGFBP proteases, which bind to the heparin binding domain of IGFBP proteins.⁵⁴

THE CCN FAMILY

The CCN proteins are secreted, extracellular matrix associated proteins. Of the three branches of the TIC superfamily, this family shows the greatest diversity of biological function. They have been shown to be involved in angiogenesis, chondrogenesis, wound healing, and many diseases by regulating such cellular processes as differentiation, adhesion, migration, and mitogenesis.²² Despite the myriad of biological activities ascribed to this family, much less is known about the biochemical mechanism(s) of action of these factors. Whether they act through their own receptors or whether they modulate other factor activities in the extracellular space is not known. The proteins are organised into discrete and conserved structural domains, each encoded by a separate exon.³² In vitro studies suggest that the central variable region is highly susceptible to enzymatic digestion.⁵⁵ For example, when cleavage occurs between domains III and IV, the C-terminal fragment of CCN1 (CTGF) was shown to bind heparin and have mitogenic activity.⁵⁶ In addition, a natural N-truncated isoform of human CCN3 (NOV) was shown to bind fibulin 1C, a protein of the extracellular matrix that interacts with several other regulators of cell adhesion, with a higher affinity than the full length protein, indicating that truncation of the protein might be crucial for the biological activity.⁵⁷ Recently, a protein of 240 kDa that specifically binds CCN1

(CTGF) was described in a human chondrocytic cell line,⁵⁸ whereas Liu *et al* identified a 221 kDa receptor protein that was phosphorylated by CCN3 (NOV).⁵⁹ To date, no common candidates for extracellular binding by CCN family members have emerged such as are seen with the TSGs and IGFbps.

PHYLOGENETIC ANALYSIS

Our sequence analysis shows that the TSG proteins constitute a novel family of extracellular factors. The relations between TSG family members in unrooted phylogenetic trees calculated using the UPGMA method (fig 3), the distance algorithm (neighbour joining, not shown), and the Fitch-Margoliash method (not shown) were equivalent, and trees calculated with each of the domains individually also reflected similar relations (not shown).

The high overall homology between the TSG proteins and the finding that the protein domains are not encoded by separate exons support the view that the domains of the TSG proteins have evolved together and are not the result of exon swapping during evolution. However, the similarity of the N-terminal CR domains of TSG, IGFBP, and CCN proteins and the fact that they are encoded by a single exon in IGFbps and CCNs, and that an intron is adjacent to the C-terminus of the domain in some TSGs, supports the view²¹ that exon shuffling was probably involved in the dissemination of this domain.

When compared among themselves, the vertebrate *tsg* genes clearly form a distinct branch well separated from the insect genes (fig 3). Because invertebrates probably possess two copies of the *tsg* gene and our searches did not uncover a second *tsg* gene in humans, it is possible either that the second *tsg* gene was deleted from vertebrates, with the exception of the fishes, or that the second human *tsg* will be revealed when the completed genome is available. The *tsg2* gene in *D melanogaster* is more closely related to the *D hydei* and *bombyx* *tsgs* than to its paralogue in *D melanogaster*, indicating that the duplication event occurred before the divergence of flies and moths.

THE TIC SUPERFAMILY

The TSG family is clearly related to but distinct from an emerging superfamily of secreted, cysteine rich growth factors in vertebrates. The family of IGFbps and the CCN family of growth factors belong to this superfamily (fig 3). All members of this superfamily are secreted, cysteine rich factors that share a common conserved N-terminal cysteine rich domain and similar topology by exhibiting the characteristic conserved 5' CR domain, a non-conserved, cysteine free mid-region, and a 3' CR domain. Among the TSGs, the N-terminal CR1 domains are 76% identical, whereas they share about 30% similarity with the CCNs and IGFbps, with slightly lower similarity (25%) to IGFBP-6, which is more diverged from the other IGFbps. The phylogenetic analysis of the TIC proteins indicates that the evolution of the superfamily

forms three lineages: IGFbps found in vertebrates only, CCN proteins also found in vertebrates, and TSG proteins isolated so far from vertebrates and insects. The fly genome contains a single gene (CG13735) that gives a significant match to CCN1 (CTGF) in a blast search. However, it shows no similarity to the N-terminal cysteine rich domain typical of CCNs and only an interrupted match to parts of the C-terminal domain of CCNs, and is not included in this analysis.

It has been suggested that the CCNs are simply divergent IGFbps.¹⁸ Grotendorst *et al* have argued that the CCNs and IGFbps represent two distinct families of proteins and that the CCN family should not be considered part of the IGFBP family.^{60–61} Our analysis presented here strongly supports the view that the CCNs, IGFbps, and TSGs are separate families of a superfamily.

Discussion

The TSG proteins are highly conserved, cysteine rich, secreted factors and all appear capable of affecting BMP signalling. To date, 13 members of the TSG family have been found in nine different species from insects to humans. They constitute a novel family of extracellular factors that may play important roles in cell–cell communication. Alignment of the proteins indicates that all have a putative signal peptide at their N-terminus, followed by a highly conserved domain, a variable hinge region containing a conserved subdomain, and a highly conserved C-terminal domain. The N-terminal, cysteine rich domain is similar to the N-terminal domains of the IGFBP and CCN proteins, the midregion has putative sites for phosphorylation by PKC and CKII that are shared with the IGFbps, but not apparently the CCNs, whereas the C-terminal CR domain might be involved in dimerisation and heparin binding.

Interestingly, the TSG proteins only are found in insects, with no evidence for genes encoding IGFBP-like or CCN-like proteins being found in the complete genome of *D melanogaster*. Nevertheless, the fly genome harbours at least four genes that encode proteins distinctly similar to the IGFs (unpublished) and flies are known to have insulin responsive receptors.^{62–66} Apparently, the functional role of IGFbps and CCNs emerged after the divergence of vertebrates and invertebrates, suggesting that the TSG-like family may predate the other two branches of the TIC superfamily.

The TSG family, together with the IGFBP and CCN proteins comprise the emerging TIC multigene superfamily of growth factors. The TSG family represents the most conserved branch of the superfamily. This structural conservation is reflected in functional conservation—all the TSGs seem capable of modulating BMP signalling and all those tested have shown an ability to interact with Sog/Chd. However, TSGs are not present everywhere that BMP signalling is active, and they may play a role only in BMP signalling events that

require spatially regulated threshold responses.^{2–30} In contrast, the IGFbps can interact with IGFs and also with receptors independent of the IGFs.^{52–53} The CCNs are the most structurally diverse set of proteins in the superfamily and they exhibit a myriad of biological functions and, to date, no clearly common biochemical mechanism of action. Structural studies that permit the analysis of the subdomains of this intriguing family of secreted factors will greatly aid our ability to understand the relations between these effectors.

This study was supported by NIH grants HD36049 and HD36081 to JLM, and by the Biological Research Center of the Hungarian Academy of Sciences to ZH. We thank Dr R Bush for assistance with the phylogenetic programs and members of the Marsh laboratory—O Marcu, N Cros, and R Sousa-Neves—for comments on the manuscript.

- Mason ED, Konrad KD, Webb CD, *et al*. Dorsal midline fate in drosophila embryos requires Twisted gastrulation, a gene encoding a secreted protein related to human connective tissue growth factor. *Genes Dev* 1994;**8**:1489–501.
- Ross JJ, Shimmi O, Vilmos P, *et al*. Twisted gastrulation is a conserved extracellular BMP antagonist. *Nature*, 2001;**410**:479–83.
- Scott IC, Blitz IR, Pappano WN, *et al*. Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signaling. *Nature* 2001;**410**:475–8.
- Chang C, Holtzman DA, Chau S, *et al*. Twisted gastrulation functions as a BMP antagonist. *Nature* 2001;**410**:483–7.
- Oelgeschläger M, Larrain J, Geissert D, *et al*. The evolutionary conserved BMP-binding protein Twisted gastrulation promotes BMP signaling. *Nature* 2000;**405**:757–63.
- Harland RM. Developmental biology. A twist on embryonic signaling. *Nature* 2001;**410**:423–4.
- Ray RP, Wharton KA. Twisted perspective. New insights into extracellular modulation of BMP signaling during development. *Cell* 2001;**104**:801–4.
- De Robertis EM, Larrain J, Oelgeschläger M, *et al*. The establishment of Spemann's organizer and patterning of the vertebrate embryo. *Nat Rev Genet* 2000;**1**:171–81.
- Irish VF, Gelbart WM. The decapentaplegic gene is required for dorsal–ventral patterning of the drosophila embryo. *Genes Dev* 1987;**1**:868–79.
- Jurgens G, Wieschaus E, Nüsslein-Volhard C, *et al*. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* II. Zygotic loci on the third chromosome. *Roux's Archives of Developmental Biology* 1984;**193**:283–95.
- Nüsslein-Volhard C, Wieschaus E, Kluding H. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. I. Zygotic loci on the second chromosome. *Roux's Archives of Developmental Biology* 1984;**193**:267–82.
- Wakimoto BT, Turner FR, Kaufman TC. Defects in embryogenesis in mutants associated with the Antennapedia gene complex of *Drosophila melanogaster*. *Dev Biol* 1984;**102**:147–72.
- Wieschaus E, Nüsslein-Volhard C, Jurgens G. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* III. Zygotic loci on the X-chromosome and fourth chromosome. *Roux's Archives of Developmental Biology* 1984;**193**:296–307.
- Zusman S, Wieschaus E. Requirements for zygotic gene activity during gastrulation in *Drosophila melanogaster*. *Dev Biol* 1985;**111**:359–71.
- Francois V, Solloway M, O'Neill JW, *et al*. Dorsal–ventral patterning of the drosophila embryo depends on a putative negative growth factor encoded by the short gastrulation gene. *Genes Dev* 1994;**8**:2602–16.
- Dale L, Jones CM. BMP signaling in early xenopus development. *Bioessays* 1999;**21**:751–60.
- Sarras MP, Jr. BMP-1 and the astacin family of metalloproteinases: a potential link between the extracellular matrix, growth factors and pattern formation. *Bioessays* 1996;**18**:439–42.
- Finelli A, Bossie CA, Xie T, *et al*. Mutational analysis of the drosophila tollid gene, a human BMP-1 homolog. *Development* 1994;**120**:861–70.
- Marques G, Musacchio M, Shimell MJ, *et al*. Production of a Dpp activity gradient in the early drosophila embryo through the opposing actions of the Sog and Tld proteins. *Cell* 1997;**91**:417–26.
- Mullins MC. Holy tollido: tollid cleaves Sog/Chordin to free DPP/BMPs. *Trends Genet* 1998;**14**:127–9.
- Hwa V, Oh Y, Rosenfeld RG. The insulin-like growth factor-binding protein (IGFBP) superfamily. *Endocr Rev* 1999;**20**:761–87.
- Brigstock DR. The connective tissue growth factor/cysteine-rich 61/nephroblastoma overexpressed (CCN) family. *Endocr Rev* 1999;**20**:189–206.
- Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 1988;**85**:2444–8.

- 24 Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 1988;73:237–44.
- 25 Nielsen H, Engelbrecht J, Brunak S, et al. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997;10:1–6.
- 26 Mason E, Marsh JL. Changes in the pattern of Twisted gastrulation gene expression among drosophila species. *J Mol Evol* 1998;46:180–7.
- 27 Mason ED, Williams S, Grotendorst GR, et al. Combinatorial signaling by Twisted gastrulation and Decapentaplegic. *Mech Dev* 1997;64:61–75.
- 28 Arora K, Nusslein-Volhard C. Altered mitotic domains reveal fate map changes in drosophila embryos mutant for zygotic dorsoventral patterning genes. *Development* 1992;114:1003–24.
- 29 Ferguson EL, Anderson KV. Localized enhancement and repression of the activity of the TGF- β family member, decapentaplegic, is necessary for dorsal–ventral pattern formation in the drosophila embryo. *Development* 1992;114:583–97.
- 30 Rusch J, Levine M. Threshold responses to the dorsal regulatory gradient and the subdivision of primary tissue territories in the drosophila embryo. *Curr Opin Genet Dev* 1996;6:416–23.
- 31 Nicholls RE, Gelbart WM. Identification of chromosomal regions involved in decapentaplegic function in drosophila. *Genetics* 1998;149:203–15.
- 32 Lau LF, Lam SC-T. The CCN family of angiogenic regulators: the integrin connection. *Exp Cell Res* 1999;248:44–57.
- 33 Beverley SM, Wilson AC. Molecular evolution in drosophila and the higher Diptera. II. A time scale for evolution. *J Mol Evol* 1984;21:1–13.
- 34 Kweon Y, Srinivasan S, Shimmi O, et al. Processing of the drosophila Sog protein creates a novel BMP inhibitory activity. *Development* 2000;127:2143–54.
- 35 Bork P. The modular architecture of a new family of growth regulators related to connective tissue growth factor. *FEBS Lett* 1993;327:125–30.
- 36 Frost RA, Tseng L. Insulin-like growth factor binding protein-1 is phosphorylated by cultured human endometrial cells and multiple protein kinases in vitro. *J Biol Chem* 1991;266:18082–8.
- 37 Hoock WG, Mukku VR. Identification of the major sites of phosphorylation in IGF binding protein-3. *J Cell Biochem* 1994;56:262–73.
- 38 Coverley JA, Baxter RC. Regulation of insulin-like growth factor (IGF) binding protein-3 phosphorylation by IGF-I. *Endocrinology* 1995;136:5778–81.
- 39 Ferry RJ, Katz LE, Grimberg A, et al. Cellular actions of insulin-like growth factor-binding proteins. *Horm Metab Res* 1999;31:192–202.
- 40 Koistinen R, Ikonen P, Selenius P, et al. Insulin-like growth factor binding protein-1 inhibits binding of IGF-I on fetal skin fibroblasts but stimulates their DNA synthesis. *Biochem Biophys Res Commun* 1990;173:408–15.
- 41 Conover CA, Powell DR. Insulin-like growth factor (IGF)-binding protein-3 blocks IGF-I-induced receptor down-regulation and cell desensitization in cultured bovine fibroblasts. *Endocrinology* 1991;129:710–16.
- 42 Hobba GD, Lothgren A, Holmberg E, et al. Alanine screening mutagenesis established tyrosine 60 of bovine insulin-like growth factor binding protein-2 as a determinant of insulin-like growth factor binding. *J Biol Chem* 1998;273:19691–8.
- 43 Kiefer MC, Masiarz FR, Bauer DM, et al. Identification and molecular cloning of two new 30-kDa insulin-like growth factor binding proteins isolated from adult human serum. *J Biol Chem* 1991;266:9043–9.
- 44 Zapf J, Born W, Chang JY, et al. Isolation and NH₂-terminal amino acid sequences of rat serum carrier proteins for insulin-like growth factors. *Biochem Biophys Res Commun* 1988;156:1187–94.
- 45 Cohen P, Peehl D, Grave H, et al. Biological effects of prostate specific antigen (PSA) as an IGF binding protein-3 (IGFBP-3) protease. *J Endocrinol* 1994;142:407–15.
- 46 Jones JI, Gockerman A, Busby WH, Jr, et al. Insulin-like growth factor binding protein 1 stimulates cell migration and binds to the 51 integrin receptor by means of its Arg-Gly-Asp sequence. *Proc Natl Acad Sci U S A* 1993;90:10553–7.
- 47 Radulescu RT. Nuclear localization signal in insulin-like growth factor-binding protein type 3. *Trends Biochem Sci* 1994;19:278.
- 48 Wraight CJ, Liepe IJ, White PJ, et al. Intracellular localization of insulin-like growth factor binding protein-3 (IGFBP-3) during cell division in human keratinocytes. *J Invest Dermatol* 1998;111:239–42.
- 49 Fowlkes J, Serra D. Characterization of glycosaminoglycan-binding domains present in insulin-like growth factor-binding protein-3. *J Biol Chem* 1996;271:14676–9.
- 50 Forbes BE, Turner D, Hodge SJ, et al. Localization of an insulin-like growth factor (IGF) binding site of bovine IGF binding protein-2 using disulfide mapping and deletion mutation analysis of the C-terminal domain. *J Biol Chem* 1998;273:4647–52.
- 51 Oh Y, Müller HL, Lamson G, et al. Insulin-like growth factor (IGF)-independent action of IGF-binding protein-3 in Hs578T human breast cancer cells. Cell surface binding and growth inhibition. *J Biol Chem* 1993;268:14964–71.
- 52 Yamanaka Y, Fowlkes JL, Wilson EM, et al. Characterization of insulin-like growth factor binding protein-3 (IGFBP-3) binding to human breast cancer cells: kinetics of IGFBP-3 binding and identification of receptor binding domain on the IGFBP-3 molecule. *Endocrinology* 1999;140:1319–28.
- 53 Leal SM, Huang SS, Huang JS. Interactions of high affinity insulin-like growth factor binding proteins with the type V transforming growth factor-beta receptor in mink lung epithelial cells. *J Biol Chem* 1999;274:6711–17.
- 54 Durham SK, Suwanichkul A, Hayes JD, et al. The heparin-binding domain of insulin-like growth factor binding protein (IGFBP)-3 increases susceptibility of IGFBP-3 to proteolysis. *Horm Metab Res* 1999;31:216–25.
- 55 Kireeva ML, Mo F-E, Yang GP, et al. Cyr61, product of a growth factor-inducible immediate early gene, promotes cell proliferation, migration, and adhesion. *Mol Cell Biol* 1996;16:1326–34.
- 56 Brigstock DR, Steffen CL, Kim GY, et al. Purification and characterization of novel heparin-binding growth factors in uterine secretory fluids. Identification as heparin-regulated Mr 10,000 forms of connective tissue growth factor. *J Biol Chem* 1997;272:20275–82.
- 57 Perbal B, Martinierie C, Sainson R, et al. The C-terminal domain of the regulatory protein NOVH is sufficient to promote interaction with fibulin 1C: a clue for role of NOVH in cell-adhesion signaling. *Proc Natl Acad Sci U S A* 1999;96:869–74.
- 58 Nishida T, Nakanishi T, Shimo T, et al. Demonstration of receptors specific for connective tissue growth factor on a human chondrocytic cell line (HCS-2/8). *Biochem Biophys Res Commun* 1998;247:905–9.
- 59 Liu Q, Huang SS, Huang JS. Function of the type V transforming growth factor beta receptor in transforming growth factor beta-induced growth inhibition of mink lung epithelial cells. *J Biol Chem* 1997;272:18891–5.
- 60 Grotendorst GR, Lau LF, Perbal B. CCN proteins are distinct from and should not be considered members of the insulin-like growth factor binding protein superfamily. *Endocrinology* 2000;141:2254–6.
- 61 Grotendorst GR, Lau LF, Perbal B. CCN proteins are distinct from, and should not be considered members of the insulin-like growth factor-binding protein superfamily. *J Clin Endocrinol Metab* 2001;86:944–5.
- 62 Gorczyca M, Augart C, Budnik V. Insulin-like receptor and insulin-like peptide are localized at neuromuscular junctions in drosophila. *J Neurosci* 1993;13:3692–704.
- 63 Fernandez R, Tabarini D, Azpiazu N, et al. The drosophila insulin receptor homolog: a gene essential for embryonic development encodes two receptor isoforms with different signaling potential. *EMBO J* 1995;14:3373–84.
- 64 Pimentel B, de la Rosa EJ, de Pablo F. Insulin acts as an embryonic growth factor for drosophila neural cells. *Biochem Biophys Res Commun* 1996;226:855–61.
- 65 Chen C, Jack J, Garofalo RS. The drosophila insulin receptor is required for normal growth. *Endocrinology* 1996;137:846–56.
- 66 Petruzzelli L, Herrera R, Garcia-Arenas R, et al. Acquisition of insulin-dependent protein tyrosine kinase activity during drosophila embryogenesis. *J Biol Chem* 1985;260:16072–5.