

Computational analysis of prodomain cysteines in human TGF- β proteins reveals frequent loss of disulfide-dependent regulation in tumors

Samantha M. Daly, Ashley Peraza, Stuart J. Newfeld  *

School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA

*Corresponding author: School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA. Email: newfeld@asu.edu

Abstract

The functionally diverse members of the human Transforming Growth Factor- β (TGF- β) family are tightly regulated. TGF- β regulation includes 2 disulfide-dependent mechanisms—dimerization and partner protein binding. The specific cysteines participating in these regulatory mechanisms are known in just 3 of the 33 human TGF- β proteins. Human prodomain alignments revealed that 24 TGF- β prodomains contain conserved cysteines in 2 highly exposed locations. There are 3 in the region of the β 8 helix that mediates dimerization near the prodomain carboxy terminus. There are 2 in the Association region that mediates partner protein binding near the prodomain amino terminus. The alignments predict the specific cysteines contributing to disulfide-dependent regulation of 72% of human TGF- β proteins. Database mining then identified 9 conserved prodomain cysteine mutations and their disease phenotypes in 7 TGF- β proteins. Three common adenoma phenotypes for prodomain cysteine mutations suggested 7 new regulatory heterodimer pairs. Two common adenoma phenotypes for prodomain and binding partner cysteine mutations revealed 17 new regulatory interactions. Overall, the analysis of human TGF- β prodomains suggests a significantly expanded scope of disulfide-dependent regulation by heterodimerization and partner protein binding; regulation that is often lost in tumors.

Keywords: adenocarcinoma; LTBP; Fibrillin; Genome Data Commons

Introduction

Secreted signaling proteins in the Transforming Growth Factor- β (TGF- β) family modulate a vast array of cellular processes in metazoan animals from sponges to humans (Hinck *et al.* 2016). In humans, developmentally defective TGF- β signaling, either loss or gain of function causes birth defects such as hereditary hemorrhagic telangiectasia or fibrodysplasia ossificans progressiva (Haupt *et al.* 2019; Balachandar *et al.* 2022). Later in life, gain of TGF- β signaling facilitates fibrosis (Meng *et al.* 2016) while the loss of TGF- β anti-mitotic signals leads to tumors (Derynck *et al.* 2021). The severe mutant phenotypes produced by flawed TGF- β signaling dictate that these proteins be tightly regulated. We employ a computational approach exploiting amino acid conservation and common mutant phenotypes to suggest new tumor-associated regulatory interactions in the TGF- β family.

One mechanism of regulation is intrinsic to TGF- β proteins. All family members are encoded as 3 domains. At the amino terminus there is a signal sequence that is removed prior to secretion. Downstream is a roughly 250 residue prodomain that is also cleaved prior to secretion. Unlike the signal sequence, after cleavage, the prodomain remains associated with the carboxy-terminal ligand domain. The ligand domain is roughly 110 residues and contains a stereotypical pattern of 6–9 cysteines. Proper folding and dimerization of the ligand domain are facilitated by

disulfide bonds between cysteines in the prodomains of 2 TGF- β monomers. Cysteines contributing to dimerization have only been identified in 3 human TGF- β proteins.

An extrinsic mechanism of regulation via binding partners has been shown for TGF β 1. Regulation is implemented via disulfide bonds between a single prodomain cysteine in 2 TGF β 1 monomers and a pair of cysteines in LTBP1 (Latent TGF- β -Binding Protein; Saharinen and Keski-Oja 2000) or LRC32 (GARP; Wang *et al.* 2012). TGF β 1 proteins are secreted as multisubunit latent complexes containing 2 disulfide-linked monomers each with a distinct disulfide bond to LTBP1 (Rifkin *et al.* 2022). Latent complexes are unable to bind receptors and limit the possibility of erroneous TGF β 1 signaling. The specific cysteines participating in TGF- β binding have only been identified in these 2 partners out of 10 proteins with the potential to form disulfide bonds with TGF- β family members.

Human TGF- β family members are divided into 3 subfamilies. The TGF- β subfamily has 8 members. We maintain the following naming hierarchy: TGF- β family is all 33 proteins, TGF- β subfamily is 8 proteins, individual proteins such as TGF β 1 use their formal name and the group of siblings is TGF β 1-3. The Activin subfamily has 8 members. The BMP subfamily has 17 members. Subfamilies are distinguishable by their amino acid sequences (Wisotzkey and Newfeld 2020) and representative homodimers have distinct structures.

Received: August 01, 2022. Accepted: September 29, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

TGFB1 has a closed-ring conformation. The prodomain $\beta 8$ region contains 2 cysteines (Cys223/225) that facilitate homodimerization via an interchain bond between 2 monomers (Shi et al. 2011). INHBA in the Activin subfamily (also called ActivinA) has 2 cysteines (Cys244/247) in a prodomain loop between $\alpha 6$ and $\beta 8$ that form an intrachain bond. INHBA homodimers display a cross-arm conformation (Wang et al. 2016). BMP9 (now formally GDF2) has 2 prodomain cysteines (Cys156/237) in a loop between $\beta 7$ and $\beta 9'$ that form an intrachain bond. BMP9 homodimers exhibit a widely open conformation (Mi et al. 2015).

To better understand the function of TGF- β family prodomeins, we previously generated alignments and trees of this domain from mouse, fly, and nematode proteins (Wisotzkey and Newfeld 2020). That data suggested that heterodimers were more common than currently appreciated. Here, we applied the same methods in an analysis of the 33 human TGF- β family members. We identified conserved prodomain cysteines, retrieved mutations in these cysteines from public databases, and utilized common mutant phenotypes to generate hypotheses for disulfide-based regulation via heterodimers or binding partners.

Human prodomain alignments revealed that 24 TGF- β prodomeins contain conserved cysteines in 2 highly exposed locations. There are 3 in region of the $\beta 8$ helix that mediates dimerization near the prodomain carboxy terminus. There are 2 in the Association (Assn) region that mediates partner protein binding near the prodomain amino terminus. The alignments predict the specific cysteines contributing to disulfide-dependent regulation of 72% of human TGF- β proteins. Database mining then identified 9 conserved prodomain cysteine mutations and their disease phenotypes in 7 TGF- β proteins. Three common adenoma phenotypes for prodomain cysteine mutations suggested 7 new regulatory heterodimer pairs. Two common adenoma phenotypes for prodomain and binding partner cysteine mutations revealed 17 new regulatory interactions. Overall, the analysis of human TGF- β prodomeins suggests a significantly expanded scope of disulfide-dependent regulation by heterodimerization and partner protein binding; regulation that is often lost in tumors.

Materials and methods

Prodomain sequences and alignments

The newest version of the longest isoform of each human TGF- β protein was identified via BLASTp (blast.ncbi.nlm.nih.gov; Gish and States 1993) employing the mouse sequence as a query (Wisotzkey and Newfeld 2020). Human GDNF was retained as a comparator for consistency with previous results. We utilized the UniProt database (www.uniprot.org; UniProt Consortium 2021) to identify the signal sequence at the amino terminus of the prodomain. The signal sequence was removed from the prodomain sequence before alignment. The mouse cleavage site that separates the prodomain and ligand was employed to identify the human cleavage site at the carboxy terminus of the prodomain. The ligand was removed from the prodomain sequence before alignment. The full set of trimmed human prodomain sequences was aligned with default settings in Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo; Sievers et al. 2020). Employing Clustal allowed us to maintain the continuity of underlying assumptions with our prior analysis of model organism TGF- β prodomeins. BoxShade 3.3.1 (ftp://www.isrec.isb-sib.ch/pub/software/unix/boxshade/3.3.1/; written by Kay Hofmann and Michael Baron) was utilized to annotate the alignment employing a cutoff for shading of 20% amino acid similarity and a biochemical

definition of similar amino acids (Wisotzkey and Newfeld 2020). This was followed by manual alignment of structural features according to an alignment in Mi et al. (2015) to minimize the impact of length variability on the amino and carboxy termini.

Prodomain and binding partner cysteine mutations and disease phenotypes

For all prodomeins and partner proteins, we mined 3 databases of deleterious disease mutations: GeneCards (www.genecards.org; Safran et al. 2022), MalaCards (www.malacards.org; Rappaport et al. 2017), and the National Cancer Institute Genome Data Commons (portal.GDC.cancer.gov; Grossman et al. 2016). Within these databases, manual scanning of each protein's complete profile led to the capture of every mutation that impacted cysteine. These include: (1) loss-of-function mutations "from a cysteine" to another residue and (2) gain-of-function mutations "to a cysteine" from another residue. We recorded each mutation's location and their phenotype if any. The majority of phenotypes were from GDC, since many disease phenotypes in the other databases were not connected to a specific residue. Common phenotypes for conserved cysteine mutations in 2 prodomeins were identified. All cysteine loss-of-function mutations in each of the 10 potential prodomain-binding proteins were examined for phenotypes matching mutations in TGF- β prodomain conserved cysteines.

Results

Identification of conserved cysteines in the prodomain of human TGF- β proteins

Building on the approach of Wisotzkey and Newfeld (2020), we downloaded the longest isoform for each of the 33 human TGF- β sequences plus GDNF (Supplementary Table 1), trimmed each at the signal sequence and ligand cleavage site (Supplementary Table 2), and then created a complete prodomain alignment. To neutralize the impact of length variability at the termini, we implemented structure-based refinements to create alignments of the carboxy and amino termini (Figs. 1 and 2).

The alignments revealed 5 distinct positions that contain conserved prodomain cysteines in 24 human TGF- β family members (Table 1). These include the human homologs of 13 mouse proteins identified previously (Wisotzkey and Newfeld 2020) plus 11 newly identified proteins. To codify specific conserved cysteine positions across proteins, we created a naming convention. Cys@1 is the first conserved cysteine in either the $\beta 8$ or Assn region. For example, TGFB2 Cys254 is $\beta 8$ Cys@1. Cys@3 is the middle conserved cysteine (third residue in the region). For example, TGFB2 Cys256 is $\beta 8$ Cys@3. Cys@4 is the last conserved cysteine (fourth residue in the region). For example, TGFB2 Cys257 is $\beta 8$ Cys@4.

The conserved cysteine positions were then visualized in the 3 TGF- β subfamily prodomain structures (Fig. 3). The number of cysteines shown for each subfamily is a compilation and not every member has all of them. In all 3 structures, conserved cysteines are located in 2 highly exposed locations. The $\beta 8$ region near the ligand cleavage site is part of a protruding loop in all subfamilies. The loop is composed of 2 β -sheets ($\beta 8$ and $\beta 9$) in the crystal structure of TGFB1. The conserved cysteine loop in the other 2 subfamilies does not contain β -sheets but still sits at the surface. In the $\beta 8$ region, partners could be other TGF- β family members to form heterodimers. The Assn region near the signal sequence cleavage site also extends beyond the core of the

Subfamily	---	β6---	---	α4---	---	β7---	---	β8---	---	β9---	
BMP15	BMP	snaawkemditqlvqqrfrwnnkgkhrilrlr	---	fm ^c qq	qkdsqgle	-----					219
GDF9	BMP	khkwiqidvtsllqplvasnkrsihmsin	---	ft ^c mk	dqlehpsagn	-----					251
NODAL	BMP	slgsmvlevtrplskwl-khpgalekqmsr	---	vage ^c wp	r	-----					183
GDF5	BMP	gsgwevfidiwklfrnf--kn	---	saql ^c le	leawe	-----					327
GDF7	BMP	gqrweafdvadamrrhrreppp	---	raf ^c ll	lravagp	-----					250
GDF6	BMP	pagwevfvdwqglrh	---	qp ^c wkql ^c le	lraawgeldageaeearargpqqp	-----					255
BMP8a	BMP	degwlvldvtaa	---	sd ^c wl	lkrhkdlglrlyve-tedghsvd	-----					229
BMP8b	BMP	degwlvldvtaa	---	sd ^c wl	lkrhkdlglrlyve-tedghsvd	-----					229
BMP6	BMP	eegwlefditat	---	snlwvvtpqhnm	glqlsvv	trdgvhvh--praa	-----				344
BMP7	BMP	eegwlvfditat	---	snhwvvnprhnl	glqlsve	tldggsin--pkla	-----				262
BMP5	BMP	dvgwlvfditvt	---	snhwvinpqnnl	glql ^c ae	tgdgrsin--vk	sa-----				286
BMP3	Act	imswlskditqllrkakeneefligfnits	---	kgr	---	qlpkrrl	-----				233
GDF10	Act	rglwqakdispivkaarrdgelllsaqlds	---	eer	---	pgvprp	-----				237
MSTN	Act	tgiwqsidvktvlqnwlkqpesnlgieika	---	ldenghd	lavl	---	-----				240
GDF11	Act	sghwqsidfkqvlhswfrqpqsngwieina	---	fdpsgt	lavl	---	-----				272
TGFB1	TGF	spewlsfdvtgvvrqwlsggeiegfrls	---	h ^c s ^c dsr	d--nt-lqvdingfttgrrgdla	---	-----				248
TGFB2	TGF	egewlsfdvtdavhewlhkdrnlgfkis	---	h ^c p ^c ctf	vpsnnyiipnkseelearfagid	---	-----				282
TGFB3	TGF	taewlsfdvtdtvrewllrresnlgleisi	---	h ^c p ^c htf	qpngdilenihevmei-kfkgvd	---	-----				254
INHA	TGF	pphwavlhlatsalsllth	---	pvlvlll	r ^c pl ^c tc	sar	-----				208
GDF15	TGF	srswdvtrpl	---	rrqlsla	---	---	-----				148
GDF1	BMP	gaawarnasw	---	prslrlalalrpr	apaa ^c ar	laeaslllv	-----				238
GDF3	BMP	kdwndnprknfglfleilvkedrdsqvnfg	---	pedt ^c ar	lrclshasllvvtlnpdqchpsr	---	-----				247
INHBA	Act	kstwhvfpvsssiqrllldqgkssldvri	---	ae ^c q ^c qe	sgasl	---	-----				254
INHBB	Act	rsgwhtfplteaigalfergerrlnldv	---	qd ^c s ^c qe	lavvp	---	-----				262
INHBC	Act	asgwhqlplgpeaga	---	acsqghl	tlelvtl--egqvaqss	---	-----				209
INHBE	Act	nlgwhtltlpssglrgeksgvlk	---	lql--	d ^c rpleg	ns--t	-----				200
GDF2	BMP	etlevssavkrwvrsdstksknklevtves	---	hrkg ^c dt	ldis	---	-----				243
BMP10	BMP	nsewetfdvtdairrwqksqstthq	---	lev	hieskhd	ea-edass	-----				236
LEFTY1	TGF	esgwkafdvteavnfwqqlsrprqp1llqv	---	svqrehl	gplsgah	---	-----				216
LEFTY2	TGF	esgwkafdvteavnfwqqlsrprqp1llqv	---	svqrehl	gplsgah	---	-----				216
AMH	TGF	sglaltlqprgedsrlstarlqallfgddh	---	r ^c cftr	--mtpallllprsepaplpahgql	---	-----				266
GDNF	Out	---	---	---	---	---	-----				108
BMP2	BMP	asrwesfdvtpavmrwtagghanhgvvev	---	ahleekq	gv--skrh	---	-----				243
BMP4	BMP	vtrwetfdvsvpavlrwtreqpnvylaiev	---	thlhqtr	th--qgqh	---	-----				251

Fig. 1. Prodomain β8 region structurally guided refined alignment. Alignment of 60 residues begins upstream of β6 and ends downstream of β9. Noted at the top are structural features derived from the alignment of TGFB1 and BMP9 shown in [Mi et al. \(2015\)](#). The β8 helix is further identified by a blue boundary. Note BMP9 is shown here employing its synonym GDF2. Any inconsistency with the carboxy terminus of the prodomain alignments in [Wisotzkey and Newfeld \(2020\)](#) is due to the presence of distinct species and the absence of structural information. Numbering is accurate and indicates the last residue on each line. One, two, or three red cysteines in the β8 region are conserved in 21 of the 33 human TGF-β family proteins. A cysteine conservation summary is in [Table 1](#) and a cysteine mutation summary is in [Table 2](#).

structure in all subfamilies. The exposed locations suggest that their conserved cysteines participate in bonds with external partners. In the Assn region, partners could be one of 10 potential binding proteins in 4 families: LTBP, Fibrillin (FBN), Leucine Repeat Containing (LRC) or E-Selectin (SELE).

Conserved cysteines in the β8 region of 21 human TGF-β proteins

We observed that all cysteine positions in the β8 region correspond to demonstrated functional cysteines ([Table 1](#)). One pair (TGFB1 Cys223/225; [Shi et al. 2011](#)) that we refer to β8Cys@1 and β8Cys@3 forms an interchain bond with their counterparts in a homodimer. A second functional cysteine pair (INHBA Cys244/247; [Wang et al. 2016](#)) that we refer to β8Cys@1 and β8Cys@4 forms an intrachain bond in a homodimer.

TGFB2, INHA, INHBA and INHBB also share cysteines β8Cys@1 and β8Cys@4. TGFB2 and INHA were predicted to form heterodimers in our prior study ([Wisotzkey and Newfeld 2020](#)) and both of these are known to heterodimerize with their siblings ([Cheifetz et al. 1988; Walton et al. 2009](#)). All other proteins with β8Cys@4 are BMP subfamily members. The structure of BMP9 homodimers showed that β8Cys@4 forms an intrachain bond (Cys237 in GDF2; synonym for BMP9).

Overall, there are 13 BMP, 4 Activin, and 4 TGF-β subfamily members with at least 1 conserved cysteine in β8. All conserved cysteine positions in the β8 region correspond to cysteines that participate in disulfide bonds in their respective subfamily structures. The discovery of cysteine conservation in the β8 region of 21 TGF-β proteins broadens the potential impact of disulfide-dependent regulation by heterodimerization from 10% to 64% of proteins.

Subfamily	ATG signal sequence	--Assn----	-----α1----->	
BMP15	BMP	mvllsilrilflcelvlf-----m	ehraqmaegg	qssiallaeaptlp 43
GDF9	BMP	marpnkflwfcfcfawlcfpislg-----s	qasggeaqse	lesgampwsllqhi 49
NODAL	BMP	mhahclpflhwallqagaatvat-----a	lllrtrgqps	psplaymlslyrdp 51
GDF5	BMP	mrlpklltflwylawldlefictvlg-----a	pdlgqrpggt	rpglakaeakerpp 52
GDF7	BMP	mdlsaaaalclwllsacrp-----r	dgleaaaavl	raagagpvrsppgg 44
GDF6	BMP	mdtprvllsavflisflwldlp-----f	qqasisssss	saclgstkgmrsk 47
BMP8a	BMP	maarpplwllgltlcalg-----gggpglrp	ppgcpqrrlg	arerrdvqreilav 51
BMP8b	BMP	mtalpgplwllglalcalg-----gggpglrp	ppgcpqrrlg	arerrdvqreilav 51
BMP6	BMP	mpglgrraqwlcwwwgllc-----	-scgppplr	pplpaaaaaaggq 42
BMP7	BMP	mhvrsraaaphsfvalwaplflrsala-----d	fsldnevhss	fihrllrsqwrrem 54
BMP5	BMP	mhltvflkigivgflwscwvlgvaykgglg-----d	nhvhssfiyr	rlrnherreiqrqi 55
BMP3	Act	magasrllflwlgcfcvslagg-----e	rpkppfpelr	kavpgdrtagggpd 47
GDF10	Act	mahvpartspgpgqllllllplfllllrdvag--s	hrapawsalp	aaadglggdrdlqr 58
MSTN	Act	mqklqlcvyiyflmlivagpvdlnenseqkenvek	eglcnactwr	qntkssrieaikiq 59
GDF11	Act	<-aegpaaaaaaaaaaaaagvggerssrpapsvape	pdgcpvcvwr	qhsrelrlesiks 83
TGFB1	TGF	mppsgrllrlplllplwllvltprpaag-----	lstcktidme	lvkrkrieairgqi 53
TGFB2	TGF	mhycvlsaflihlvtvals-----	lstcstldmd	qfmrkrieairgqi 44
TGFB3	TGF	mkmhlqralvvlallnfatvsls-----	lstcttldfg	hikkkrveairgqi 47
INHBA	TGF	mvhlhllflflltpqggh-----	--scqglela	relvlakvralfld 39
GDF15	TGF	mpgqelrtvngsqmllvllvswlphgga-----l	slaeasrasf	pgpselhsedsrfr 54
GDF1	BMP	mppppqgpcghhlllllalllpslptr-----//	lqpchveelg	vagnivrhlpdrgall 11
GDF3	BMP	mlrflpdlafsfllilalqgavqf-----//	rdlcyykelg	vrgnvlrflpdqgf 85
INHBA	Act	mpllwlrqfllascwiivrs-----sptpgseghsa	apdcpscala	alpkdvpnsqpmv 56
INHBB	Act	<--aacllllaagwlgpeawgsptpppppgspggs	qdtctscggf	rrpeelgrvdgdf 76
INHBC	Act	msslllaflllapttvat-----pra	ggqcpacggp	tlelesqrellldl 45
INHBE	Act	mrlpdvqlwlvllwalvra-----qgt	gsvcpscggs	klapqaeralvlel 46
GDF2	BMP	mcpgalwvalpllsllagslqg-----k	plqswgrgsa	ggnaahsplgvpgg 47
BMP10	BMP	mgslvltlcalfcclaaylvsg-----s	pimnleqspl	eedmslfgdvfseq 46
LEFTY1	TGF	mqplwlcwalwvlpaspaga-----l	tgeollgsll	rqlqlkevptldra 46
LEFTY2	TGF	mwplwlcwalwvlpagpgaa-----l	teeollgsll	rqlqlsevpvldra 46
AMH	TGF	mrdlpltslalvlsalgalgte-----//	epclvalgg	dsngsssplrvvga 69
GDNF	Out	mqslpnsngaaagrdfkmk-----lwdv	vavclvlht	asafplpagkrppe 47
BMP2	BMP	mvagtrcllalllpqvlvggaag-----l	vpelgrrkfa	aassgrpssqpsde 48
BMP4	BMP	mpgnrmlmvvllcqvlvg-----k	vaeiqghggr	rsgqshellrdrfea 44

Fig. 2. Prodomain Assn region structurally guided refined alignment. Alignment of 60 residues begins at Met1 for all proteins except GDF11 and INHBB due to long sequences between the signal sequence and Assn region. The ATG, signal sequence, and structural features derived from the alignment of TGFB1 and BMP9 (Hinck et al. 2016) are noted at the top. The Assn region is further identified by a blue boundary. Note that any inconsistency with the amino terminus of the alignments in Wisotzkey and Newfeld (2020) is due to the presence of distinct species and the absence of structural information. Numbering is accurate and indicates the last residue on each line. One or two red cysteines in the Assn region are conserved in 16 of the 33 human TGF- β proteins plus GDNF. A cysteine conservation summary is in Table 1 and a cysteine mutation summary is in Table 2.

Conserved cysteines in the Assn region of 16 human TGF- β proteins

There are more conserved cysteine positions in the Assn region than known functional cysteines. The Assn region has 2 conserved cysteine positions (Table 1), but the only known functional cysteine position is AssnCys@1 (TGFB1 Cys33). Each monomer of a TGFB1 homodimer forms a disulfide linkage via this cysteine to either LTBP1 Cys1359 or Cys1384. TGFB1 also employs Cys33 to bind to LTBP3 and LTBP4. Conserved cysteines in this position in TGFB2 (Cys24) and TGFB3 (Cys27) are presumed to serve the same function, though this has not formally been shown. The fact that 16 proteins have an AssnCys@1 suggests that partner protein disulfide regulation extends to members of all 3 subfamilies.

The second Assn region-conserved cysteine (AssnCys@4) is present only in 6 Activin subfamily proteins. These are the 4 INHBs plus MSTN and GDF11. The latter 2 are the only proteins

with 2 Assn region cysteines and no β 8 region cysteines. With the 2 Assn cysteines showing identical spacing to β 8Cys@1 and β 8Cys@4 and the Assn region also highly exposed, 1 new hypothesis is that the 2 Assn region cysteines facilitate an Activin subfamily-specific mechanism for heterodimerization.

Overall, there are 5 BMP, 6 Activin and 5 TGF- β subfamily members with at least 1 conserved cysteine in this region. All 16 sequences contain AssnCys@1 with 6 also displaying AssnCys@4. The highly exposed nature of the Assn region together with the role of TGFB1 AssnCys@1 in binding LTBP1/LRC32 suggests that a disulfide-dependent binding partner regulates as many as 48% of human TGF- β proteins.

Common phenotype of prodomain region cysteine mutations indicates new heterodimers

To assign specific interactions to individual cysteines, we mined 3 databases of disease-associated mutations. We applied the

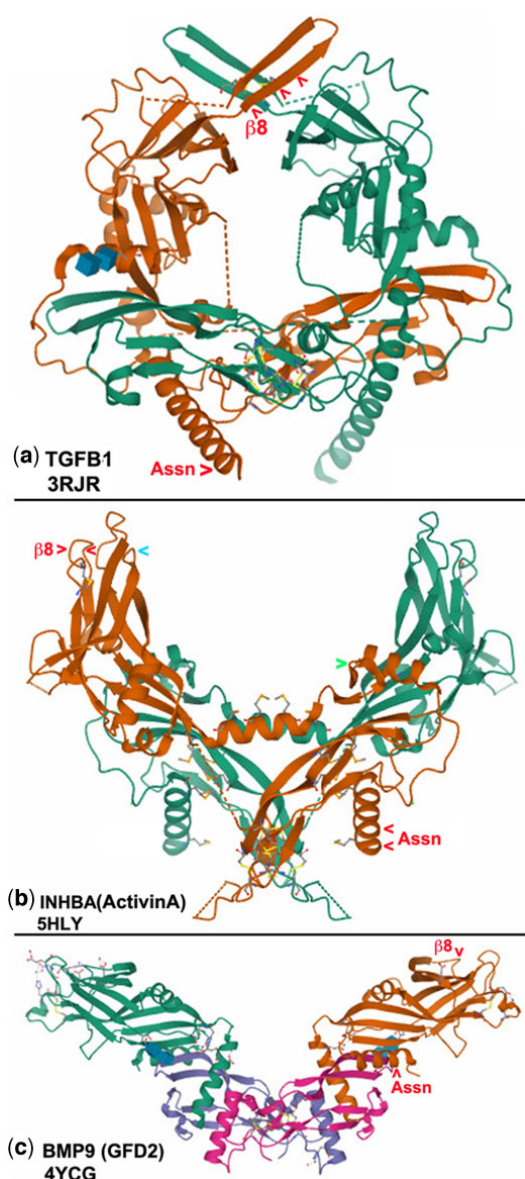


Fig. 3. Conserved prodomain cysteines in 2 exposed structural regions. a) TGFB1. Crystal structure of the porcine prodomain dimer with monomers in green and brown. Red arrowheads indicate 4 conserved cysteines in the brown monomer seen in various combinations in 5 of 8 TGF- β subfamily members. The Assn region has 1 cysteine exposed near the amino terminus. The $\beta 8$ region has 3 cysteines exposed near the carboxy-terminal cleavage site. The dimer has a closed-ring structure. Image created in Mol* (Sehna et al. 2021; Burley et al. 2021) (www.rcsb.org/3d-view/3RJR). b) INHBA (ActivinA). Crystal structure of the human prodomain dimer with monomers in green and brown. Red arrowheads indicate 4 conserved cysteines in the brown monomer seen in various combinations in 6 of 8 Activin subfamily members. The Assn and $\beta 8$ regions each have 2 cysteines in exposed locations. The blue arrowhead indicates IHNBB Arg223Cys a gain-of-function mutation found in endometrial tumors. The green arrowhead indicates IHNBB Ser154Cys a gain of function mutation found in plasma cell tumors. The dimer has an open-arm conformation (www.rcsb.org/3d-view/5HLY). c) BMP9 (GDF2). Crystal structure of the mouse complete dimer with the prodomains in green and brown. Red arrowheads indicate 2 conserved cysteines in the brown monomer seen in various combinations in 13 of 17 BMP subfamily members. The Assn and $\beta 8$ regions each have 1 cysteine in exposed locations. The dimer has a widely open conformation (www.rcsb.org/3d-view/4YCG).

concept originated by Beadle and Tatum (1941) for the dissection of biochemical pathways: common mutant phenotypes result from a lost biochemical interaction that would normally achieve a common function.

The phenotypes of 9 conserved prodomain cysteines that display a loss-of-function mutation are shown (Table 2). Common phenotypes in different proteins would broaden the impact of regulation by heterodimerization in the TGF- β family. In the table, mutations are color-coded and numbered to display potential within subfamily and across subfamily heterodimer pairs. Eight cysteine mutations are in the $\beta 8$ region and 1 is in the Assn region.

There are 3 mutations in BMP subfamily members. BMP15 $\beta 8$ Cys@4 has Cys209Gly found in lung squamous cell neoplasms. GDF1 $\beta 8$ Cys@4 Cys227-Del is an in-frame deletion of 145 residues causing inherited right atrial isomerism, a heart defect. For GDF6, there are 2 colon adenocarcinoma mutations, 1 in $\beta 8$ Cys4 (Cys230Arg) and 1 in the ligand (Cys419Tyr). Furthermore, a GDF5 ligand mutation Arg438Cys is also associated with colon adenocarcinomas. When a ligand cysteine mutation generates the same phenotype as a conserved prodomain cysteine mutation, it provides additional confidence that the prodomain cysteine is essential. The common phenotype of GDF5 and GDF6 suggests heterodimerization.

Six mutations are in Activin and TGF- β subfamily members. INHBA $\beta 8$ Cys@1 has Cys244Tyr found in lung adenocarcinomas. Note that BMP15 and INHBA cancers originate in distinct cell types (squamous cells vs epithelial cells, respectively), limiting the relevance of these mutations to understand heterodimerization. TGFB1 $\beta 8$ Cys@1 has 3 mutations Cys223Arg/Gly/Ser and TGFB1 $\beta 8$ Cys@3 has 2 Cys225Arg/Tyr. All 5 can cause Camurati-Engelmann disease, an autosomal dominant disease with skeletal hyperplasia.

INHBE AssnCys@4 has Cys29Tyr that is associated with plasma cell tumors. INHBB has a prodomain gain-of-function mutation that is also associated with plasma cell tumors. IHNBB Ser154Cys maps to an exposed loop between $\beta 1$ and $\beta 2$ that is topologically near to the Assn region in the INHBA structure (Fig. 3b, green arrowhead right). A new cysteine in close proximity to the INHBB pair of Assn cysteines likely interferes with their function. If INHBE and INHBB formed a heterodimer or were regulated by a common binding partner, then loss or gain of cysteine near the Assn region could be disruptive and result in tumors.

TGFB2 $\beta 8$ Cys@3 has Cys256Stop identified in endometrial adenocarcinomas. The same tissue and tumor type are seen with the Cys291Trp in the ligand of INHA that shares $\beta 8$ Cys@1 and $\beta 8$ Cys@4 with TGFB2. Endometrial adenomas were again seen with Arg223Cys in the prodomain of INHBB that shares $\beta 8$ Cys@1 and $\beta 8$ Cys@4 with TGFB2 and INHA. Arg223Cys maps to an exposed loop between $\beta 6$ and $\beta 7$ that is topologically near to $\beta 8$ in the INHBA structure (Fig. 3b, blue arrowhead left). A new cysteine in close proximity to the pair of INHBB $\beta 8$ region cysteines likely interferes with their function. The common phenotype for INHA and INHBB that is known to heterodimerize increases confidence in identifying heterodimer partners via cysteine mutations and common disease phenotypes. The fact that TGFB2 shares 2 $\beta 8$ cysteines and an endometrial adenoma phenotype with INHA and INHBB implies that TGFB2 can heterodimerize with either to prevent endometrial tumor formation.

TGFB2 $\beta 8$ Cys@4 has 2 mutations. First, Cys257Phe is associated with Holt-Oram syndrome. This is an autosomal dominant disease with a proximate cause of nonfunctional TBX5 (Boogerd et al. 2010). It is characterized by skeletal abnormalities and heart

Table 1. Twenty-four human TGF- β proteins with conserved prodomain cysteines.

Protein and subfamily	Region, Position ^a and Residue #				
	AssnCys@1	AssnCys@4	β 8Cys@1	β 8Cys@3	β 8Cys@4
BMP15 BMP subfamily					209
GDF9 BMP subfamily					239
NODAL BMP subfamily					180
GDF5 BMP subfamily					310
GDF6 BMP subfamily					230
GDF7 BMP subfamily					231
BMP5 BMP subfamily					272
GDF2 (BMP9) BMP subfamily					237 ^e
BMP6 BMP subfamily	22				
BMP8a BMP subfamily	31				205
BMP8b BMP subfamily	31				205
GDF1 BMP subfamily	91				227
GDF3 BMP subfamily	71				232
MSTN Act subfamily	39	42			
GDF11 Act subfamily	62	65			
INHBA Act subfamily	35	38	244 ^c		247 ^c
INHBB Act subfamily	55	58	252		255
INHBC Act subfamily	26	29	190		
INHBE Act subfamily	26	29	192		
AMH TGF subfamily	55		241		
INHA TGF subfamily	19		200		203
TGFB1 TGF subfamily	33 ^b		223 ^d	225 ^d	
TGFB2 TGF subfamily	24		254	256	257
TGFB3 TGF subfamily	27		227	229	
a. Cys@1 is the first in the respective region with Cys@3 and Cys@4 the 3rd & 4th residue downstream.					
b. TGFB1 Cys33 binds LTBP1 (Saharinen and Keski-Oja 2000).					
c. INHBA Cys244/Cys247 form an intrachain bond (Wang et al. 2016).					
d. TGFB1 Cys223/Cys225 form an interchain bond (Shi et al. 2011).					
e. BMP9 (GDF2) Cys237 participates in an intrachain bond (Mi et al. 2015)					

defects. A TGFB2 ligand mutation Cys378Tyr is also found in Holt-Oram.

The TGFB2 mutation Cys257Stop in β 8Cys@4 was noted in Loeys–Dietz syndrome. This is an autosomal dominant disease with systemic effects on connective tissue and blood vessels. A TGFB2 ligand mutation Cys439Ser was also noted in Loeys–Dietz, as was a ligand mutation in TGFB3 Cys409Tyr. The common phenotype for these mutations suggests heterodimerization. Demonstrated heterodimers of TGFB2 and TGFB3 (Cheifetz et al. 1988) serve to validate this hypothesis. What is new is that TGFB2 β 8Cys4 shows the phenotype, thus suggesting a β 8-based mechanism for TGFB2 and TGFB3 heterodimer formation.

Cysteine mutations in TGFB2, INHA, and INHBB associated with endometrial adenocarcinoma plus cysteine mutations in TGFB2 and TGFB3 with Loeys–Dietz syndrome lead to the hypothesis that TGFB2 is capable of heterodimerization with numerous partners. Further evidence for TGFB2 versatility is that 2 TGFB2 ligand mutations Cys246Tyr and Cys407Ser share a lung

adenocarcinoma phenotype with a mutation in β 8Cys@1 of INHBA Cys244Tyr. Taken together, TGFB2 cysteine mutations have a common phenotype with cysteine mutations in 4 TGF- β family members INHA, INHBA, INHBB, and TGFB3.

Overall, phenotypic analyses of 9 prodomain-conserved cysteine disease mutations in 7 proteins suggested 7 new heterodimers. Four heterodimers are within a subfamily (1 BMP, 1 Activin, and 2 TGF- β) and 3 are across subfamilies (all are Activin with TGF- β). Six heterodimers are associated with tumors, with 5 of these adenomas. We then identified common disease phenotypes shared by conserved prodomain cysteine mutations and partner-binding protein cysteine mutations.

Common phenotype of Assn region cysteine mutations suggest new regulation

We analyzed the single Assn region mutation separately from the β 8 region mutations. A common phenotype for the Assn region cysteine mutation and a cysteine mutation in one of the 10

Table 2. Common disease phenotypes of cysteine mutations suggest 3 new across subfamily heterodimers and 4 new within subfamily heterodimers.

Colors and numbered pairs show three across subfamily heterodimers	Cysteine	Mutation	Disease Phenotype Colors and numbered pairs show four within subfamily heterodimers
BMP15 BMP subfamily	$\beta 8\text{Cys}@4$	Cys209Gly	lung squamous cell neoplasms
GDF1 BMP subfamily	$\beta 8\text{Cys}@4$	Cys227-Del	inherited right atrial isomerism
GDF6 BMP subfamily	$\beta 8\text{Cys}@4$	Cys230Arg	colon adenocarcinomas 1
	Ligand	Cys419Tyr	colon adenocarcinoma
GDF5 BMP subfamily	Ligand	Arg438Cys	colon adenocarcinoma 1
INHBA Act subfamily 1	$\beta 8\text{Cys}@1$	Cys244Tyr	lung adenocarcinomas
INHBB Act subfamily 2, 3	prodomain	Arg223Cys	endometrial adenocarcinomas
	prodomain	Ser154Cys	plasma cell tumors 2
INHBE Act subfamily	AssnCys@4	Cys29Tyr	plasma cell tumors 2
INHA TGF subfamily 2	Ligand	Cys291Trp	endometrial adenocarcinomas 3
TGFB1 TGF subfamily	$\beta 8\text{Cys}@1$	Cys223Ser	Camurati-Engelmann Disease
	$\beta 8\text{Cys}@3$	Cys225Arg	
TGFB2 TGF subfamily 3	$\beta 8\text{Cys}@3$	Cys256Stop	endometrial adenocarcinomas 3
	$\beta 8\text{Cys}@4$	Cys257Phe	Holt-Oram Syndrome
		Cys257Stop	Loeys-Dietz Syndrome 4
TGFB2 TGF subfamily 1	Ligand	Cys246Tyr Cys407Ser	lung adenocarcinomas
TGFB3 TGF subfamily	Ligand	Cys409Tyr	Loeys-Dietz Syndrome 4

potential TGF- β -binding proteins would suggest a regulatory interaction that broadens the impact of disulfide-dependent partner protein binding. INHBE Cys29Tyr in AssnCys@4 is found in plasma cell tumors from the hematopoietic and phagocytic systems. This suggests that undetected binding partners can regulate Activin subfamily members.

We identified all cysteine mutations showing a plasma cell tumor phenotype in each potential binding partner (LTBP1-4, FBN1-3, LRC32/33, and SELE). We found 11 cysteine mutations in 4 proteins that generate this phenotype (Table 3). There is 1 plasma cell tumor cysteine mutation in LTBP1. Cys1022Tyr is in EGF-like repeat-7, roughly 330 residues upstream of TGF- β -binding domain-3 that binds TGFB1. The common phenotype with INHBE suggests an interaction between LTBP1 and INHBE via AssnCys@4. The location of LTBP1 Cys1022 outside of TGF- β -binding domain-3 is logical since INHBA has its Assn cysteines in a topologically distinct location from TGFB1 (Fig. 3b).

There are 2 plasma cell tumor cysteine mutations in FBN1. One is Cys1431Tyr in EGF-like repeat-24 and the other Cys1687Phe is in EGF-like repeat-28. The common phenotype suggests an interaction between FBN1 and INHBE via AssnCys@4. Another FBN1 mutation Cys1431Trp is associated with Loeys-Dietz syndrome, a phenotype generated by prodomain cysteine mutations in TGFB2 and TGFB3. The common Loeys-Dietz syndrome phenotype for cysteine mutations in FBN1, TGFB2, and TGFB3 supports their interaction via conserved prodomain cysteines.

There are 5 plasma cell tumor cysteine mutations in FBN2. One is in EGF-like repeat-16 and 2 are in EGF-like repeat-22

including Cys1406. Mutations in Cys1406 have distinct associations: Cys1406Phe is found in plasma cell tumors and Cys1406Ser is associated with colon adenocarcinoma. The mutations Cys1579Gly and Cys1608Tyr are in TGF- β -binding domain-6. Their spacing of 27 amino acids is similar to the spacing of the TGFB1-binding cysteines in LTBP1 (25 residues). An alignment of FBN2 TGF- β -binding domain-6 and LTBP1 TGF- β -binding domain-3 (Fig. 4a) shows that both FBN2 mutant cysteines are spaced similarly to those in LTBP1. Spacing similarity supports the hypothesis that these cysteines in FBN2 TGF- β -binding domain-6 are each capable of binding a monomer of INHBE AssnCys@4.

There are 3 plasma cell tumor cysteine mutations in FBN3. They are in EGF-like repeat-2, EGF-like repeat-20 and in TGF- β binding domain-6. An alignment of TGF- β -binding domain-6 from FBN3 with TGF- β -binding domain-6 in FBN2 (Fig. 4a) shows that Cys1519Arg impacts the same cluster of 3 cysteines as FBN2 Cys1608Tyr mutation and yields the same phenotype. Phenotypic correspondence provides additional evidence in support of the hypothesis that FBN family members bind to Activin subfamily Assn region cysteines.

Independent support for the FBN-Activin-binding hypothesis derives from an NMR study of FBN1 TGF- β -binding domain-6 and the adjacent downstream EGF-like repeat (Yuan et al. 2002). The Marfan syndrome mutation N2144S in the FBN1 EGF-like repeat led to a conformational shift impacting the 2 nearest intrachain disulfide bonds. The loss of those bonds altered the protein's local topology in both directions including TGF- β -binding domain-6. The authors conclude that lost interactions of TGF- β -binding domain-6 contribute to the disease.

Table 3. Common disease phenotype for an INHBE Assn region cysteine mutation suggests 4 new binding partner interactions associated with plasma cell tumors.

Protein	Cysteine	Mutation	Disease Phenotype Colors match Table 2
INHBE Act subfamily	AssnCys@4	Cys29Tyr	plasma cell tumors
LTBP1	EGF-like-7	Cys1022Tyr	plasma cell tumors
FBN1	EGF-like-24	Cys1431Tyr	plasma cell tumors
		Cys1431Trp	Loeys-Dietz Syndrome ^a
	EGF-like-28	Cys1687Phe	plasma cell tumors
FBN2	EGF-like-16	Cys1131Stop	plasma cell tumors
	EGF-like-22	Cys1378Ser Cys1406Phe ^b	plasma cell tumors
	TGF- β bind-6	Cys1579Gly Cys1608Tyr	plasma cell tumors
FBN3	EGF-like-2	Cys252Tyr	plasma cell tumors
	EGF-like-20	Cys1349Ser	plasma cell tumors
	TGF- β bind-6	Cys1519Arg	plasma cell tumors
a. This phenotype matches mutations in TGFB2 and TGFB3 shown in Table 2.			
b. Mutation FBN2 Cys1406Ser yields colon tumors as shown in Table 4.			

Overall, the common phenotype analysis of an Assn region cysteine mutation in INHBE suggested 4 new binding partners. Loss of each regulatory interaction results in plasma cell tumors suggesting that regulation serves to prevent an INHBE promitotic signal from impacting the hematopoietic and/or phagocytic systems.

Common phenotype of β 8 region cysteine mutations suggests additional new regulation

Analysis of β 8 region cysteines began with GDF6 β 8Cys@4 (Cys230Arg) that is found in colon adenocarcinomas (Table 4). Five partner proteins display cysteine mutations found in this tumor. In LTBP1, 2 plasma cell mutations in TGF- β -binding domain-1 Cys559Tyr and Cys594Trp are 35 residues apart affecting the first and sixth cysteines. In LTBP1 TGF- β -binding domain-3, the cysteines that bind TGFB1 are the second and sixth. Perhaps the extra space is necessary to accommodate the widely open structure of BMP proteins. Each of the 3 FBN proteins has at least 3 cysteine mutations linked to colon adenocarcinomas. In FBN3, Cys1406Ser occurs in the same cysteine where Cys1406Phe was found in plasma cell tumors. Distinct missense mutations in a single FBN3 cysteine in tumors linked to BMP and Activin proteins indicate that FBN3 may have multiple partners.

A second β 8 region mutation TGFB2 β 8Cys@3 (Cys256Stop) is found in endometrial adenocarcinomas. Eight partner proteins display cysteine mutations in this tumor. Four of these have a single mutation LTBP1, LTBP3, LRC32, and SELE. The Cys342Arg mutation in LRC32 is in the same repeat (LRR-12) as Cys350 that binds to TGFB1. These are the only 2 cysteines in this repeat. There are 2 mutant cysteines each in FBN2 and FBN3 and 4 mutant cysteines in LTBP2 and FBN1.

Overall, the common phenotype analysis of β 8 region cysteine mutations in GDF6 and TGFB2 suggested 13 new binding partners. Loss of each regulatory interaction results in

adenocarcinomas suggesting that regulation serves to prevent a GDF6 or TGFB2 promitotic signal from impacting the colon or endometrium, respectively.

Taken together, the common phenotype analysis of Assn and β 8 region-conserved cysteine mutations (1 protein from each subfamily) suggested 17 new regulatory interactions with 8 binding partners. Sixteen of these interactions are associated with tumors. Four of the partners (LTBP1 and FBN1-3) demonstrate a common disease phenotype with each mutant TGF- β protein suggesting the ability to bind to members of all 3 subfamilies. As all regulatory partners act to restrict signaling, common mutant phenotypes in tumors suggest that the absence of partner binding (due to a prodomain or a partner cysteine mutation) contributes to TGF- β family members' well-known ability to serve tumors as promitotic signals.

LTBP1 repeat that binds TGFB1 aligns well with repeats in FBN proteins

To investigate the basis for the common phenotypes noted above for FBN1-3 cysteine mutations, we created alignments with LTBP1 of FBN1-3 TGF- β -binding domain-6 (Fig. 4a) and an EGF-like repeat (Fig. 4b). Two important features of TGF- β -binding domain-3 in LTBP1 that shape its ability to bind TGFB1 Cys33 are: (1) a 2 residue (FP) insertion between cysteines 6 and 7 and (2) a suite of 5 acidic residues (D/E) that form a docking site (Lack et al. 2003; Chen et al. 2005). The fact that FBN proteins do not bind TGFB1 is attributed to the lack of the FP insertion in any of their TGF- β -binding domains (Rifkin et al. 2022). Data in Table 3 address the question "would the absence of the insertion prevent Activin subfamily binding?" The common plasma cell tumor phenotype for a cysteine mutation in INHBE and 3 mutations in TGF- β -binding domain-6 of FBN2 and FBN3 suggest it does not.

Similarly, a different 5 D/E residue docking site would be expected for Activin subfamily binding by an FBN TGF- β -binding domain.

Table 4. Common disease phenotypes for GDF6 and TGFB2 $\beta 8$ region cysteine mutations suggest 13 new binding partner interactions associated with tumors.

Protein and subfamily	Location	Mutation	Disease phenotype is adenocarcinoma Colors match Table 2
GDF6 BMP subfamily	$\beta 8$ Cys@4	Cys230Arg	colon
LTBP1	TGF- β bind-1	Cys559Tyr Cys594Trp	colon
Fibrillin1	EGF-like-2 EGF-like-11 EGF-like-21 EGF-like-30	Cys119Gly Cys763Phe Cys1307Tyr Cys1847Tyr	colon
Fibrillin2	EGF-like-22 EGF-like-31 EGF-like-35	Cys1406Ser ^a Cys1903Tyr Cys2072Arg	colon
Fibrillin3	EGF-like-6 EGF-like-11 EGF-like-14	Cys557Tyr Cys885Tyr Cys1096Phe	colon
SELE	Sushi-3	Cys304Arg	colon
TGFB2 TGF subfamily	$\beta 8$ Cys@3	Cys256Stop	endometrial
LTBP1	EGF-like-2	Cys413Phe	endometrial
LTBP2	EGF-like-3 EGF-like-7 TGF- β bind-4 EGF-like-20	Cys648Tyr Cys985Tyr Cys1611Tyr Cys1788Arg	endometrial
LTBP3	EGF-like-13	Cys1282Tyr	endometrial
Fibrillin1	EGF-like-13 EGF-like-19 EGF-like-35 EGF-like-37	Cys821Tyr Cys1201Tyr Cys2053Tyr Cys2190Arg	endometrial
Fibrillin2	EGF-like-4 EGF-like-23	Cys280Trp Cys1412Trp	endometrial
Fibrillin3	EGF-like-35 EGF-like-36	Cys2181Phe Cys2225Arg	endometrial
LRC32	LRR-12	Cys342Arg	endometrial
SELE	Sushi-6	Cys506Tyr	endometrial

a. Mutation FBN2 Cys1406Phe yields plasma cell tumors as shown in Table 3.

Different docking sites are needed to accommodate the cross-arm structure of Activin proteins vs the closed-ring of TGF- β proteins. A comparison of FBN TGF- β -binding domain-6 D/E content to LTBP1 TGF- β -binding domain-3 reveals that 2 of the D/E residues that form 1 edge of the docking site in LTBP1 TGF- β -binding domain-3 are conserved in all 3 FBN sequences (D12, E42; [Chen et al. 2005](#)). Three other D/E residues are not (E5, D17, D39). Instead there are 3 D/E residues aligned in all FBN sequences that are spaced similarly. The 5 D/E docking sites in FBN TGF- β -binding domain-6 with a slightly different shape for Activin proteins seem likely.

Overall, the mutant phenotype analysis expands the scope of TGF- β family partner interactions to include EGF-like repeats in LTBP1 and the 3 FBN proteins. Eight cysteine mutations in EGF-like repeats in LTBP1 and in the 3 FBN proteins are associated with tumors. Two cysteine mutations in EGF-like repeat-22 of FBN2 show spacing of 28 residues similar to the spacing of LTBP1 TGFB1-binding cysteines. Finally, 6 acidic residues are present in EGF-like repeats of

LTBP1 and the FBN proteins that could form a docking site for an Activin protein.

Discussion

Frequent regulatory interactions with partner proteins seem likely given the number of cysteines in each potential partner protein. The secreted LTBP and FBN families each contain multiple 8-Cys repeat domains. An 8-Cys repeat in LTBP1 contains the 2 cysteines that each bind to Cys33 in a monomer of TGFB1 ([Saharinen and Keski-Oja 2000](#)). As a result, 8-Cys repeats in the 4 LTBP and 3 FBN proteins are also called TGF- β -binding domains. Proteins in these 2 families also contain a large number of 6-Cys EGF-like repeats. For example, LTBP1 has 16 and FBN1 has 47 of them. While EGF-like repeats are common, 8-Cys repeats are not found in any other protein. To date no TGF- β partner for LTBP2 is known and only TGFB1-3 have been tested for binding to LTBP1-4.



Fig. 4. Conserved cysteine mutations and acidic residues are aligned in LTBP1 and FBN TGF- β binding and EGF-like domains. a) Top row is the LTBP1 TGF- β -binding domain-3 (TB3) domain with its 8 cysteines in red. The 2 cysteines that bind TGFB1 are indicated above the sequence. The sequence of TGF- β -binding domain-6 (TB6) in each of the 3 FBN family members is aligned. Three cysteine mutations in FBN2 and FBN3 that are found in plasma cell tumors are shown below their sequence. Note the expected gap in FBN proteins opposite the LTBP1 FP insertion that is unique to TGFB1 binding. Acidic residues (D/E) in the docking site for TGFB1 are in green (D12 and E42) if they are conserved in FBN sequences and in purple (E5, D17, and D39) if they are not. Three D/E residues present in all FBN sequences but that are absent in LTBP1 are in blue. An additional conserved E in all 4 sequences with no known function is in orange. b) Top row is the LTBP1 EGF-like repeat-7 with its 6 cysteines in red. One EGF-like repeat with a cysteine mutation from each FBN family member is aligned. A cysteine mutation in LTBP1 and 7 in FBN proteins that are associated with plasma cell tumors are indicated. Three cysteine mutations present in other FBN EGF-like repeats are indicated by their repeat numbers (#28, #16, and #2). Five of the 6 cysteines in a canonical EGF-like repeat are mutated in plasma cell tumors. Three D/E residues immediately upstream of the first cysteine of the repeat and 3 D/E residues immediately downstream of the last are in blue. All 6 are present in LTBP1 and the 3 FBN sequences.

Studies of FBN1 binding to TGFB1 have been negative, dampening enthusiasm for this family (Yuan et al. 2002). As a result, to our knowledge, no other TGF- β family members have been tested for binding to any FBN protein. Each FBN protein has 8 TGF- β -binding domains. Alignments indicate that TGF- β -binding domains in FBN family members share features with TGF- β -binding domain-3 in LTBP1 that binds TGFB1 Cys33. There are 105 potential partnerships between the 15 TGF- β proteins that are not TGFB1 but have AssnCys@1 and the 7 TGF- β -binding domain proteins (4 LTBP plus 3 FBN proteins). The probability that a subset of these interactions is real seems high.

The cell surface transmembrane proteins LRC32 (also known as GARP) and LRC33 are also TGFB1-binding partners. The extracellular portion of LRC32 binds to TGFB1 Cys33 in the same 2:1 ratio as LTBP1 (Liénart et al. 2018). In contrast to the proximity of the 2 cysteines in LTBP1 (Cys1539/1584, 25 residues), the 2 cysteines in LRC32 are far apart (Cys211/350, 139 residues). TGFB1 binding to LRC33 has been reported but the relevant cysteines are not identified (Ma et al. 2019). These proteins have roughly 20 Leucine Rich Repeats containing 1 or 2 cysteines each. It again seems likely that additional TGF- β proteins can bind LRC32 and LRC33.

The ER transmembrane protein E-Selectin-ligand (SELE) plays a different role. Rather than maintaining latency in the secreted TGFB1 dimer like the others, this binding partner is a negative regulator of TGFB1 and TGFB2 secretion (Yang et al. 2010). Neither the cysteine in SELE nor the cysteine in TGFB1 and TGFB2 is known. SELE contains 6 Sushi domains each with 6 cysteines. The potential for a disulfide bond between SELE and the 14 TGF- β family proteins with AssnCys@1 that are not TGFB1 and TGFB2 is considerable.

Nevertheless, we recognize that not all the tumor phenotypes provided by the NCI Cancer Genome Commons precisely describe a specific cell type. While colon, endometrial, and lung adenocarcinoma describe only epithelial cell tumors, the phenotype plasma cell tumors covers many cell types. We also recognize that

heterodimerization and partner binding require expression of both proteins in the same cell. To date, for most human TGF- β family proteins, detailed expression data to confirm or deny our hypotheses are lacking. Thus, we urge our colleagues to test our hypothesis for new heterodimers and new binding partner interactions.

In conclusion, for over 20 years, our understanding of TGF- β signaling has benefited from computational approaches. A recent advance in our ability to align TGF- β family prodomains across species facilitated this study of human TGF- β prodomain cysteines. New human prodomain alignments revealed that 24 TGF- β prodomains contain conserved cysteines in 2 highly exposed locations. There are 3 in the region of the β 8 helix that mediates dimerization near the prodomain carboxy terminus. There are 2 in the Assn region that mediates partner protein binding near the prodomain amino terminus. The alignments predict the specific cysteines contributing to disulfide-dependent regulation of 72% of human TGF- β proteins. Database mining then identified 9 conserved prodomain cysteine mutations and their disease phenotypes in 7 TGF- β proteins. Three common adenoma phenotypes for prodomain cysteine mutations suggested 7 new regulatory heterodimer pairs. Two common adenoma phenotypes for prodomain and binding partner cysteine mutations revealed 17 new regulatory interactions. Overall, the analysis of human TGF- β prodomains suggests a significantly expanded scope of disulfide-dependent regulation by heterodimerization and partner protein binding; regulation that is often lost in tumors.

Data availability

The authors affirm that all data necessary for confirming the conclusions of the article are present in the article, figures, tables, and supplemental information.

Supplemental material is available at G3 online.

Acknowledgments

This article is dedicated to my dear friend and collaborator Robert Wisotzkey. His insight and ingenuity in sequence analysis will be sorely missed.

Funding

SJN is supported by National Institutes of Health OD024794 and OD028242. SMD was supported by aResearch Supplement to Promote Diversity in Health-Related Research to SJN (OD024794-S1).

Conflicts of interest

None declared.

Literature cited

- Balachandar S, Graves TJ, Shimonty A, Kerr K, Kilner J, Xiao S, Slade R, Sroya M, Alikian M, Curetean E, et al.; Genomics England Research Consortium. Identification and validation of a novel pathogenic variant in GDF2 (BMP9) responsible for hereditary hemorrhagic telangiectasia and pulmonary arteriovenous malformations. *Am J Med Genet A*. 2022;188(3):959–964.
- Beadle GW, Tatum EL. Genetic control of biochemical reactions in neurospora. *Proc Natl Acad Sci USA*. 1941;27(11):499–506.
- Boogerd CJJ, Dooijes D, Ilgun A, Mathijssen IB, Hordijk R, Van De Laar IMBH, Rump P, Veenstra-Knol HE, Moorman AFM, Barnett P, et al. Functional analysis of novel TBX5 T-box mutations associated with Holt-Oram syndrome. *Cardiovasc Res*. 2010;88(1):130–139.
- Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichtlow GV, Christie CH, Dalenberg K, Di Costanzo L, Duarte JM, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research*. 2021;49(D1):D437–D451.
- Cheifetz S, Bassols A, Stanley K, Ohta M, Greenberger J, Massagué J. Heterodimeric TGF- β : biological properties and interaction with three types of cell surface receptors. *J Biol Chem*. 1988;263(22):10783–10789.
- Chen Y, Ali T, Todorovic V, O'leary JM, Kristina Downing A, Rifkin DB. Amino acid requirements for formation of the TGF- β -Latent TGF- β binding protein complexes. *J Mol Biol*. 2005;345(1):175–186.
- Derynck R, Turley SJ, Akhurst RJ. TGF β biology in cancer progression and immunotherapy. *Nat Rev Clin Oncol*. 2021;18(1):9–34.
- Gish W, States DJ. Identification of protein coding regions by database similarity search. *Nat Genet*. 1993;3(3):266–272.
- Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, Staudt LM. Toward a shared vision for cancer genomic data. *N Engl J Med*. 2016;375(12):1109–1112.
- Haupt J, Stanley A, McLeod CM, Cosgrove BD, Culbert AL, Wang L, Mourikioti F, Mauck RL, Shore EM. ACVR1R206H FOP mutation alters mechanosensing and tissue stiffness during heterotopic ossification. *Mol Biol Cell*. 2019;30(1):17–29.
- Hinck AP, Mueller TD, Springer TA. Structural biology and evolution of the TGF- β family. *Cold Spring Harb Perspect Biol*. 2016;8(12):a022103.
- Lack J, O'Leary JM, Knott V, Yuan X, Rifkin DB, Handford PA, Downing AK. Solution structure of the third TB domain from LTBP1 provides insight into assembly of the large latent complex that sequesters latent TGF- β . *J Mol Biol*. 2003;334(2):281–291.
- Liénart S, Merceron R, Vanderaa C, Lambert F, Colau D, Stockis J, van der Woning B, De Haard H, Saunders M, Coulie PG, et al. Structural basis of latent TGF β 1 presentation and activation by GARP on human regulatory T cells. *Science*. 2018;362(6417):952–956.
- Ma W, Qin Y, Chapuy B, Lu C. LRC33 is a novel binding and potential regulating protein of TGF β 1 in human acute myeloid leukemia cells. *PLoS One*. 2019;14(10):e0213482.
- Meng XM, Nikolic-Paterson D, Lan HY. TGF- β : the master regulator of fibrosis. *Nat Rev Nephrol*. 2016;12(6):325–338.
- Mi L-Z, Brown CT, Gao Y, Tian Y, Le VQ, Walz T, Springer TA. Structure of BMP9 pro-complex. *Proc Natl Acad Sci USA*. 2015;112(12):3710–3715.
- Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, Gershoni M, Morrey CP, Safran M, Lancet D, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res*. 2017;45(D1):d877–D887.
- Rifkin D, Sachan N, Singh K, Sauber E, Tellides G, Ramirez F. The role of LTBP1 in TGF- β signaling. *Dev. Dyn*. 2022;251(1):75–84.
- Saharinen J, Keski-Oja J. Specific sequence motif of 8-Cys repeats of TGF- β binding proteins LTBP1 creates a hydrophobic interaction surface for binding of small latent TGF- β . *Mol Biol Cell*. 2000;11(8):2691–2704.
- Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koca J, Rose AS. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res*. 2021;49(W1):W431–W437.
- Safran M, Rosen N, Twik M, BarShir R, Iny-Stein T, Dahary D, Fishilevich D, Lancet D. The GeneCards suite. In: Abugessaisa I, Kasukawa T, editors. *Practical Guide to Life Science Databases*. Singapore: Springer; 2022. p. 27–56.
- Shi M, Zhu J, Wang R, Chen X, Mi L, Walz T, Springer TA. Latent TGF- β structure and activation. *Nature*. 2011;474(7351):343–349.
- Sievers F, Barton GJ, Higgins DG. Multiple sequence alignment. In: AD Baxevanis, GD Bader, DS Wishart, editors. *Bioinformatics*. 4th ed. USA: John Wiley & Sons; 2020. p. 227–250.
- Uniprot Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D884–D891.
- Walton KL, Makanji Y, Wilce MC, Chan KL, Robertson DM, Harrison CA. A common biosynthetic pathway governs the dimerization and secretion of Inhibin and related TGF- β ligands. *J Biol Chem*. 2009;284(14):9311–9320.
- Wang R, Zhu J, Dong X, Shi M, Lu C, Springer TA. GARP regulates the bioavailability and activation of TGF- β . *Mol Biol Cell*. 2012;23(6):1129–1139.
- Wang X, Fischer G, Hyvonen M. Structure and activation of pro-ActivinA. *Nat Commun*. 2016;7:12052.
- Wisotzkey RG, Newfeld SJ. TGF- β prodomain alignments reveal unexpected cysteine conservation consistent with phylogenetic predictions of cross-subfamily heterodimerization. *Genetics*. 2020;214(2):447–465.
- Yang T, Mendoza-Londono R, Lu H, Tao J, Li K, Keller B, Jiang MM, Shah R, Chen Y, Bertin TK, et al. E-selectin ligand-1 regulates growth plate homeostasis in mice by inhibiting the intracellular processing and secretion of mature TGF- β . *J Clin Invest*. 2010;120(7):2474–2485.
- Yuan X, Werner JM, Lack J, Knott V, Handford PA, Campbell ID, Downing AK. Effects of the N2144S mutation on backbone dynamics of a TB-cbEGF domain pair from human fibrillin-1. *J Mol Biol*. 2002;316(1):113–125.

Communicating editor: E. Betran

Supplemental Material: Two Tables

Table S1. Human TGF- β superfamily sequences, subfamilies and accession numbers.

Subfamily	Name	Accession	Description (synonym)
Activin	BMP3	NP_001192.4	bone morphogenetic protein 3 preproprotein
Activin	GDF10	NP_004953.1	growth/differentiation factor 10 preproprotein
Activin	GDF11	NP_005802.1	growth/differentiation factor 11 preproprotein (syn. BMP11)
Activin	INHBA	NP_002183.1	inhibin beta A chain preproprotein
Activin	INHBB	NP_002184.2	inhibin beta B chain preproprotein
Activin	INHBC	NP_005529.1	inhibin beta C chain preproprotein
Activin	INHBE	NP_113667.1	inhibin beta E chain preproprotein
Activin	MSTN	NP_005250.1	growth/differentiation factor 8 preproprotein (synonym GDF8)
8 Activin			
BMP	BMP2	NP_001191.1	bone morphogenetic protein 2 preproprotein
BMP	BMP4	NP_001193.2	bone morphogenetic protein 4 isoform a preproprotein
BMP	BMP5	NP_066551.1	bone morphogenetic protein 5 isoform 1 preproprotein
BMP	BMP6	NP_001709.1	bone morphogenetic protein 6 preproprotein
BMP	BMP7	NP_001710.1	bone morphogenetic protein 7 preproprotein
BMP	BMP8a	NP_861525.2	bone morphogenetic protein 8A preproprotein
BMP	BMP8b	AAP74560.1	bone morphogenetic protein 8B
BMP	BMP10	NP_055297.1	bone morphogenetic protein 10 preproprotein
BMP	BMP15	NP_005439.2	bone morphogenetic protein 15 preproprotein (synonym GDF9b)
BMP	GDF1	NP_001483.3	embryonic growth/differentiation factor 1 precursor
BMP	GDF2	NP_057288.1	growth/differentiation factor 2 preproprotein (synonym BMP9)
BMP	GDF3	NP_065685.1	growth/differentiation factor 3 preproprotein
BMP	GDF5	NP_000548.2	growth / differentiation factor 5 preproprotein
BMP	GDF6	NP_001001557.1	growth/differentiation factor 6 preproprotein
BMP	GDF7	NP_878248.2	growth/differentiation factor 7 preproprotein
BMP	GDF9	NP_005251.1	growth/differentiation factor 9 preproprotein
BMP	NODAL	NP_060525.3	nodal homolog isoform 1 preproprotein
17 BMP			
TGF- β	AMH	NP_000470.3	mullerian inhibiting factor preproprotein (synonym MIS)
TGF- β	GDF15	NP_004855.2	growth/differentiation factor 15 preproprotein
TGF- β	INHA	NP_002182.1	inhibin alpha chain isoform 1 preproprotein
TGF- β	LEFTY1	NP_066277.1	left-right determination factor 1 preproprotein
TGF- β	LEFTY2	NP_003231.2	left-right determination factor 2 isoform 1 preproprotein
TGF- β	TGFB1	NP_000651.3	transforming growth factor beta-1 preproprotein
TGF- β	TGFB2	NP_001129071.1	transforming growth factor beta-2 isoform 1 precursor
TGF- β	TGFB3	NP_001316868.1	transforming growth factor beta-3 isoform 1 preproprotein
8 TGF- β			

33 total TGF- β superfamily sequences

Outgroup GDNF NP_001177397.1 glial cell line-derived neurotrophic factor isoform 3

Table S2. Human TGF- β superfamily signal sequences and ligand cleavage sites.

Subfamily	Name	Signal		Cleavage	
		Sequence	Residues ^a	Prodomain side	Ligand side
Activin	BMP3	1 - 22	ALSIERR RKKR	//	STGVLLPLQN
Activin	GDF10	1 - 23	LKPRPGR KDR	//	RKKGQEVFMA
Activin	GDF11	1 - 24	VLENTKRS R	//	NLGLDCDEHS
Activin	INHBA	1 - 20	SEDPH RRRR	//	RGLECDGKVN
Activin	INHBB	1 - 28	RLGDSR HRIR	//	KRGLECDGRT
Activin	INHBC	1 - 18	RVGGKH QIHR	//	RGIDCQGGSR
Activin	INHBE	1 - 19	EPGAGR ARRR	//	TPTCEPETPL
Activin	MSTN	1 - 23	VTDTPK RSRR	//	DFGLDCDEHS
BMP	BMP2	1 - 23	GHPLHK REKR	//	QAKHKQRKRL
BMP	BMP4	1 - 19	ALTRRR RAKR	//	SPKHHSQRAR
BMP	BMP5	1 - 30	ASEVLL RSVR	//	AANKRKNQNR
BMP	BMP6	1 - 20	VSEVHV RTTR	//	SASSRRRQQS
BMP	BMP7	1 - 29	ATEVHF RSIR	//	STGSKQRSQN
BMP	BMP8a	1 - 19	ASPSIR TPR	//	AVRPLRRRQP
BMP	BMP8b	1 - 19	SPIRTP RAVR	//	PLRRRQPKKS
BMP	BMP10	1 - 21	IYDSTAR IRR	//	NAKGNKYCKRT
BMP	BMP15	1 - 18	ERESLL RRTR	//	QADGISA EVT
BMP	GDF1	1 - 29	GPGGAC RARR	//	DAEPVLGGGP
BMP	GDF2	1 - 22	AGSTLA RRKR	//	SAGAGSHCQK
BMP	GDF3	1 - 24	DQCHPS RKRR	//	AAIPVPKLSC
BMP	GDF5	1 - 27	EYLFSQ RRKR	//	RAPLATRQ GK
BMP	GDF6	1 - 22	KRHGKK SRLR	//	CSKKPLHVNF
BMP	GDF7	1 - 19	AGRGHG RGR	//	SRCSRKPLHV
BMP	GDF9	1 - 24	GRSSH RHRR	//	GQETVSSELK
BMP	NODAL	1 - 26	SWEWGK RHRR	//	HHLPDRS QLC
TGF- β	AMH	1 - 24	DPRGPG RAQR	//	SAGATAADGP
TGF- β	GDF15	1 - 29	LRPQA ARGRR	//	RARARNGDHC
TGF- β	INHA	1 - 18	PPSGGE RARR	//	STPLMSWPWS
TGF- β	LEFTY1	1 - 21	LASGAH KLVR	//	FASQGAPAGL
TGF- β	LEFTY2	1 - 21	LASGAH KLVR	//	FASQGAPAGL
TGF- β	TGFB1	1 - 29	QHLQSS RHRR	//	ALDTNYCFSS
TGF- β	TGFB2	1 - 20	SQQTNR RKKR	//	ALDAAYCFRN
TGF- β	TGFB3	1 - 23	PGQGGQ RKKR	//	ALDTNYCFRN
Outgroup	GDNF	1 - 20	QMAVL PRRER	//	NRQAAAANPE

a. Source: www.uniprot.org

b. Source: Wistozkey and Newfeld; Genetics 2020