

## **Structural Basis of Intracellular TGF- $\beta$ Signaling: Receptors and Smads**

Apirat Chaikuad

Structural Genomics Consortium, University of Oxford, Oxford OX3 7DQ, United Kingdom

Tel: +44 1865 617576; Fax: +44 1865 617575; Email: [apirat.chaikuad@sgc.ox.ac.uk](mailto:apirat.chaikuad@sgc.ox.ac.uk)

Alex N. Bullock\*

Structural Genomics Consortium, University of Oxford, Oxford OX3 7DQ, United Kingdom

Tel: +44 1865 617754; Fax: +44 1865 617575; Email: [alex.bullock@sgc.ox.ac.uk](mailto:alex.bullock@sgc.ox.ac.uk)

\*Corresponding author

Running title: Structural basis of intracellular signaling

Stimulation of the TGF- $\beta$  family receptors activates an intracellular phosphorylation-dependent signaling cascade that culminates in Smad transcriptional activation and turnover. Structural studies have identified a number of allosteric mechanisms that control the localisation, conformation and oligomeric state of the receptors and Smads. Such mechanisms dictate the ordered binding of substrate and adaptor proteins that determine the directionality of the signaling process. Activation of the pathway has been illustrated by the various structures of the R-Smads with SARA, Smad4 and YAP, respectively, while mechanisms of downregulation have been elucidated by the structural complexes of FKBP12, Ski, Smurf1. Interesting parallels have emerged between the R-Smads and the FHA and IRF-associated domains, as well as the Hippo pathway. However, important questions remain as to the mechanism of Smad-independent signaling.

## **INTRODUCTION**

Extracellular binding of TGF- $\beta$  superfamily ligands to heteromeric complexes of the type I and type II receptors is expected to bring the receptor intracellular domains into close proximity allowing for the intracellular activation of the receptors and in turn the receptor-associated Smads (R-Smads). The type I receptors require specific transphosphorylation by the type II receptors before they are activated to bind and phosphorylate their R-Smad substrates (Wrana et al. 1994). While structural studies of the receptor extracellular domains have defined the precise assembly of several ternary ligand-receptor complexes, similar models for the arrangement of the intracellular domains are lacking. Such studies are complicated by the dynamic nature of phosphorylation-dependent signaling pathways, which are determined by transient protein-protein interactions that are generally of lower affinity than those of receptor extracellular domains. Furthermore, these interactions are tightly regulated by a number of adaptor domains. Nonetheless, progress in the mapping of many new intracellular interactions

has facilitated a wealth of structural information that adds significantly to our understanding of the regulatory mechanisms of TGF- $\beta$  signaling.

## **RECEPTOR KINASE ACTIVATION**

### **Structures of the type II receptor kinase domain**

The human kinome comprises over 500 unique family members (Manning et al. 2002). Within this phylogenetic tree, the type I and type II receptors are both classified as members of the tyrosine kinase-like (TKL) family. As this name suggests, their catalytic domains share considerable sequence and structural similarity with the tyrosine kinases, although specific sites determining substrate selection are conserved with other serine/threonine kinases (Huse et al. 1999; Krupa et al. 2004). Consequently, TGF- $\beta$  family receptors are described as dual-specificity kinases since they can weakly phosphorylate tyrosine residues as well as the preferred serine and threonine residues (Lawler et al. 1997; Lee et al. 2007). Comparisons across a diverse range of kinase structures have revealed that tyrosine kinases have a wider substrate pocket that allows for the larger tyrosine side chain to be optimally positioned for receipt of the terminal  $\gamma$ -phosphate of ATP (Krupa et al. 2004). This is partly established by the presence of a bulky tryptophan residue at a conserved position (e.g. c-ABL Trp405) in the activation loop of tyrosine kinases that is typically a smaller side chain in serine/threonine kinases (Krupa et al. 2004).

To date, crystal structures are available for three of the five type II receptor kinase domains, including BMPRII, ActRIIA and ActRIIB (Han et al. 2007; Chaikuad et al. 2012). All three structures show the common kinase domain fold, which has a bilobal architecture (Figures 1A and 1B). The N-terminal lobe comprises a five-stranded  $\beta$ -sheet and a single  $\alpha$ -helix ( $\alpha$ C),

which forms a key regulatory element in many protein kinases. The C-terminal lobe is largely  $\alpha$ -helical, but also includes the catalytic loop harbouring critical catalytic residues, such as the His-Arg-Asp (HRD) motif, as well as the activation loop, which contributes to the substrate-binding pocket and is often a site of activating phosphorylation in other protein kinase families, but not within the TGF- $\beta$  receptor family. The activation loop starts with an Asp-Phe/Leu-Gly (DFG or DLG) motif, which contributes to the binding of the co-substrate ATP and a magnesium (or manganese) ion. The available structures of the type II receptors reveal a conserved active conformation of the kinase domain that is consistent with the constitutive activity shown *in vitro* (Han et al. 2007; Chaikuad et al. 2012). Importantly, the N and C-terminal kinase lobes are stably bound together to create a functional ATP-binding pocket at their interface. In the N-lobe, the  $\beta$ 1- $\beta$ 2 hairpin forms a glycine-rich loop, also known as the phosphate-binding loop (P-loop), which contacts the phosphate moieties in ATP, whereas the hinge region linking the N-lobe to the C-lobe establishes two hydrogen bonds to the adenosine.

### **Structural basis for activation of the type I receptor**

The type I receptors have evolved an efficient control mechanism to prevent their inappropriate activation in the absence of TGF- $\beta$  ligand. Their intracellular domains are distinguished by a regulatory glycine-serine rich (GS) domain that is located in the juxtamembrane region immediately N-terminal to their kinase domain (Figure 1A). Transphosphorylation by the type II receptor switches the GS domain from a binding site for the inhibitor FKBP12 to a binding site for the R-Smad substrate (Huse et al. 2001). FKBP12 (FK506-binding protein 1A (12 kDa)) is a small  $\alpha/\beta$  fold protein best known for its binding to the immunosuppressant drugs rapamycin and FK506 (Michnick et al. 1991).

The structural basis for this activation switch has been established by crystal structures of the type I TGF- $\beta$  receptor (T $\beta$ R-I or ALK5) solved either in the presence and absence of FKBP12 (Figure 1C) (Huse et al. 1999; Huse et al. 2001). In both structures, the GS domain adopts a helix-loop-helix motif that folds across the top of the kinase N-lobe  $\beta$ 4 strand. Serine and threonine residues targeted by the type II receptor are found in the central GS loop sequence 185-TTSGSGSGLP-194. Under quiescent conditions, FKBP12 sits atop the GS domain and shields the GS loop from the type II receptor. Its binding is centred on the downstream helix  $\alpha$ GS2. Two residues from this helix, Leu195 and Leu196, are inserted into the macrolide-binding pocket of FKBP12, and therefore compete for binding with rapamycin (Huse et al. 1999).

FKBP12 binding forces the GS loop to insert into the kinase domain where it forms an inhibitory wedge between the N-lobe  $\beta$ -sheet and the  $\alpha$ C helix. This distorts the kinase N-lobe by pushing out the C-terminal end of the  $\alpha$ C helix, whereas the N-terminus is swung into the ATP pocket (Figure 1C inset box). As a result, the kinase domain adopts an inactive conformation in which the catalytic salt bridge between T $\beta$ R-I Asp245 ( $\alpha$ C helix) and Lys232 ( $\beta$ 3-strand) is broken. The inhibitory complex is stabilized by a number of arginine side chain interactions that are strictly conserved in the type I receptors, but divergent in the type II receptors. For example, the  $\alpha$ GS2 helix is tethered to the kinase domain by Arg203, whereas the associated GS loop is held by Arg255, which is located just C-terminal to the  $\alpha$ C helix. The binding sites for ATP and substrate are additionally blocked by a salt bridge between Arg372 (activation loop) and the DLG motif (Asp351).

Significantly, the equivalent residues in the type I BMP receptor ALK2 (ACVR1) are found mutated in the skeletal malformation disorder fibrodysplasia ossificans progressiva (Shore et al.

2006; Kaplan et al. 2009), as well as the childhood cancer diffuse intrinsic pontine glioma (Buczkowicz et al. 2014; Fontebasso et al. 2014; Taylor et al. 2014; Wu et al. 2014). The structure of the ALK2-FKBP12 complex shows that these residues participate in similar inhibitory interactions to the equivalent T $\beta$ R-I complex (Chaikuad et al. 2012). Thus, upon mutation these bonds are broken and the receptor complex is mildly activated (Fukuda et al. 2008; Fukuda et al. 2009; Shen et al. 2009; Song et al. 2010; van Dinther et al. 2010; Chaikuad et al. 2012). Remarkably, the mutant ALK2 receptor also displays a novel gain of function where it signals in response to activin A, as well as the expected BMP ligands (Hatsell et al. 2015; Hino et al. 2015).

The crystal structure of the T $\beta$ R-I intracellular domain solved in the absence of FKBP12 shows that the inhibitory interactions between the GS loop and kinase domain are relieved (Huse et al. 1999). This frees the GS loop for exposure to the type II receptor and concomitantly allows the  $\alpha$ C helix to adopt the correct conformation for catalysis. Without phosphorylation, the type I receptor lacks significant activity against the R-Smads (Wrana et al. 1994; Huse et al. 2001). GS domain phosphorylation inhibits FKBP12 binding, but significantly enhances the binding of R-Smads and therefore provides an effective switch for receptor activation. Unfortunately, there are currently no structures describing precisely how the type II receptor interacts with the GS domain to complete this phosphorylation. Some general insights can be drawn from other kinase families which bind to their substrates across the top of the kinase C-lobe and align the phosphorylatable residue with the  $\gamma$ -phosphate of ATP (Endicott et al. 2012).

## **RECEPTOR-SMAD INTERACTION AND SMAD ACTIVATION**

All R-Smads share a similar domain architecture comprising an N-terminal MH1 domain, a central proline-rich linker and a C-terminal MH2 domain. Their activation is dependent on the

type I receptors which phosphorylate a common Ser-X-Ser motif present at the extreme C-terminus of the MH2 domain (Abdollah et al. 1997; Souchelnytskyi et al. 1997). Generally, the TGF- $\beta$  and activin-specific receptors phosphorylate Smad2 and Smad3, whereas the BMP-specific receptors phosphorylate Smad1, Smad5 and Smad8.

### **Structural basis for R-Smad recruitment by SARA**

R-Smad proteins have been shown to interact with a number of proteins associated with the endosome, which appear to help direct the R-Smads to the receptors at the cell membrane as well as through the early endosome (Tsukazaki et al. 1998; Miura et al. 2000; Chen et al. 2007; Shi et al. 2007). Co-crystal structures have been solved for the MH2 domains of Smad2 and Smad3 in complex with one of these proteins, Smad anchor for receptor activation (SARA; Figure 2A) (Wu et al. 2000; Qin et al. 2002). SARA is a large protein targeted to the membrane by a central FYVE domain that confers binding to phosphatidylinositol 3-phosphate (PtdIns3P). A small Smad-binding domain (SBD) has been identified immediately C-terminal to the FYVE domain with specificity for Smad2 and Smad3, but not the BMP-type R-Smads (Tsukazaki et al. 1998).

The Smad MH2 domain contains three main structural elements (Figure 2A). At one end of the structure the Smad N-terminus forms to a large  $\beta$ -sandwich that shares its fold with the FHA (Forkhead-associated) and IAD (IRF-associated domain) domains (Durocher et al. 2000; Qin et al. 2003). At the centre of the structure, a three-helix bundle is established by an insertion and a single C-terminal  $\alpha$ -helix. Finally, in the R-Smads, the domain structure is completed by a large C-terminal tail that extends to the C-terminal Ser-X-Ser motif (Wu et al. 2001).

At different stages of R-Smad activation there are changes to both the structure and oligomeric state of the MH2 domain that contribute to the directionality of the signaling process (Wu et al. 2000; Chacko et al. 2001; Wu et al. 2001; Qin et al. 2002; Chacko et al. 2004). Complexes of Smad2 or Smad3 with the SBD of SARA display a heterodimeric structure (Wu et al. 2000; Qin et al. 2002). In these structures, the SBD adopts an extended conformation that folds across the  $\beta$ -sandwich of the R-Smad and along a hydrophobic groove running down the length of the three-helix bundle (Figure 2A). Here, the C-terminal SBD region makes additional interactions with the R-Smad N-terminus, which is induced to form part of a two-stranded antiparallel  $\beta$ -sheet. No interactions are formed with the R-Smad C-terminal region, including the Ser-X-Ser motif which is disordered in these structures.

### **Insights into R-Smad interaction with the type I receptor**

In their purified forms, the unphosphorylated MH2 domains exist predominantly as monomers, but they also display a propensity to trimerize (Kawabata et al. 1998; Chacko et al. 2001; Correia et al. 2001; Moustakas and Heldin 2002). The binding of SARA stabilizes the 'monomeric' form of Smad2 and Smad3 and therefore helps to present the MH2 domain to the type I receptor in an accessible state for phosphorylation of the free C-terminal Ser-X-Ser motif (Qin et al. 2002).

In the absence of structural information for the receptor-Smad complex, models for the Smad interaction have been proposed based on the structural similarity of the FHA domain (Durocher et al. 2000; Huse et al. 2001). The FHA domain is a well-studied recognition module for phospho-threonine (pThr)-containing peptides. Co-structures of the FHA domain have revealed a phosphopeptide binding site at one end of the core  $\beta$ -sandwich (Durocher et al. 2000). Interestingly, this site corresponds to a basic patch on the Smad MH2 domain. By comparison,



the basic patch on the R-Smads is expected to form a recognition site for the phosphorylated GS loop, providing an explanation for the specific binding of the phosphorylated receptor (Figure 2B). Phosphorylation therefore switches the GS region from a binding site for the inhibitory protein FKBP12 into a recruitment site for the substrate R-Smad (Huse et al. 2001). Phosphoamino acid analysis has suggested that each of the serine and threonine residues located within the GS loop sequence 185-TTSGSGSGLP-194 of T $\beta$ R-I can be phosphorylated (Souchelnytskyi et al. 1996). However, scanning mutagenesis has failed to identify a single residue that strictly requires phosphorylation suggesting some redundancy within these Gly-Ser repeats (Wieser et al. 1995). Interestingly, a Thr204Asp mutant of T $\beta$ R-I showed constitutive signaling even in the absence of T $\beta$ R-II and TGF- $\beta$  (Wieser et al. 1995). This residue in the  $\alpha$ GS2 helix is not a recognized phosphorylation site and similar activation is observed in BMP receptors such as ALK2 with the equivalent Gln207Asp mutation at the same position (Macias-Silva et al. 1998). Thus, this site may also contribute to the R-Smad interaction.

The specificity of the receptor-Smad interaction has also been investigated by protein chimeras and site-directed mutagenesis (Feng and Derynck 1997; Chen et al. 1998; Lo et al. 1998; Persson et al. 1998; Chen and Massague 1999). These studies have identified a further cluster of four residues within the  $\beta$ 4- $\beta$ 5 loop (L45 loop) of the type I receptor kinase N-lobe that appear to determine Smad selectivity. Exchange of these residues between BMP and TGF- $\beta$  receptors reverses their respective specificity towards the R-Smads. Moreover, two complementary residues are found in the L3 loop of the R-Smads, suggesting a potential direct interaction between the L45 loop of the receptor and the L3 loop of the R-Smads. Furthermore, the concave face of the MH2 domain appears a good match both in shape and size for the convex face of the receptor kinase domain (Figure 2B). While the precise interaction surface remains to be determined by high resolution structural studies, the binding of the globular MH2 domain has the

potential to stabilize an active conformation of the receptor kinase. Finally, the C-terminal Ser-X-Ser motif of the R-Smad must also be engaged by the substrate pocket within the kinase C-lobe for phosphorylation to complete. Thus, the GS domain and N-terminal kinase lobe are predicted to form a docking site for the R-Smad MH2 domain that correctly positions the phosphorylatable Ser-X-Ser motif adjacent to the ATP-binding pocket. Further structural studies are required to define these precise interactions.

### **Dissociation of the phosphorylated R-Smad**

Following phosphorylation, the R-Smad has a high propensity to trimerize which favours its dissociation from the receptor and SARA (Kawabata et al. 1998). A trimeric subunit arrangement is revealed by the crystal structure of the phosphorylated MH2 domain of Smad2 (Wu et al. 2001). In this structure, contacts between adjacent subunits are mediated by the interaction of helices H4 and H5 of the three-helix bundle with the helix H1 and  $\beta$ 4 strand of the neighbouring subunit. The phosphorylated Ser-X-Ser tail also makes contact with the basic patch of the neighbouring subunit, including the L3 loop and  $\beta$ 8 strand. These interactions within the Smad trimer are therefore mutually exclusive with those involving the receptor GS domain and L45 loop.

The phosphorylated MH2 domain also undergoes concerted conformational changes that perturb the SARA-binding pocket at the outside periphery of the trimer. Firstly, the N-terminal  $\beta$ -strand of Smad2 is repositioned from its SARA interaction to participate in the new trimer interface (Wu et al. 2001). Secondly, the packing within the MH2 domain between the three-helix bundle and  $\beta$ -sheet is subtly changed to promote the trimeric arrangement of the different Smad subunits. Finally, the conformation of the L3 loop is also altered through its binding of the phosphorylated Ser-X-Ser motif.

## **HETEROMERIC SMAD ASSEMBLY**

Activation of the R-Smads is continued by their assembly with the Co-Smad, Smad4 (also known as DPC4), and their translocation to the nucleus for transactivation of Smad-target genes (Lagna et al. 1996). Biochemical analyses suggest that the heteromeric Smad complex is more stable than the Smad homotrimer and comprises two phosphorylated R-Smad subunits and one Smad4 subunit (Kawabata et al. 1998; Chacko et al. 2001; Correia et al. 2001).

### **Structural features of Smad4**

Overall, Smad4 shows a somewhat similar domain arrangement to the R-Smads, but lacks the C-terminal Ser-X-Ser extension. Crystal structures of the Smad4 MH2 domain also reveal a similar folding topology to the R-Smads (Figure 3A) (Shi et al. 1997; Qin et al. 1999). However, a specific 35-residue insertion is identified that greatly lengthens the H3 helix and generates an additional  $\beta$ -strand that folds back to form an antiparallel interaction with the shortened C-terminal tail. The transcriptional activity of Smad4 also depends on a Smad-activation domain (SAD) consisting of a 45-residue proline-rich segment located N-terminal of the MH2 domain (Qin et al. 1999).

### **Structural basis of heteromeric Smad assembly**

Crystal structures have been determined for both the heterotrimeric Smad2-Smad4 and Smad3-Smad4 complexes (Figure 3B) (Chacko et al. 2004). As expected, their assemblies are highly similar to those of the homotrimeric R-Smads, except that the inclusion of Smad4 creates three non-identical binding interfaces. A number of the interface residues are conserved between Smad4 and the R-Smads, allowing for two different R-Smads to be incorporated into the same heterotrimer. Indeed, this appears to be one mechanism to increase the available diversity of

the TGF- $\beta$  pathway. For example, the simultaneous activation of Smad2/3 and Smad1/5 can generate mixed TGF- $\beta$  and BMP-type R-Smad complexes leading to altered promoter binding and transcriptional activation (Daly et al. 2008).

### **Structural parallels between R-Smads and IRF-family proteins**

IRF-family transcription factors are known for their role in the innate immune response where they function downstream from pattern recognition receptors, such as the RIG-I-like receptors and Toll-like receptors. Crystal structures of IRF3 (Qin et al. 2003; Takahashi et al. 2003; Qin et al. 2005; Takahashi et al. 2010) and IRF5 (Chen et al. 2008) have revealed striking structural conservation between their transactivation domains and the MH2 domain of the R-Smads, including a core  $\beta$ -sandwich that is decorated at one end by an  $\alpha$ -helical bundle and a C-terminal extension. This observation suggests a shared evolutionary origin of the R-Smads and IRFs. Moreover, these structures have enabled a consensus view of IRF activation that shows many parallels with the R-Smads. For example, the IRFs are similarly activated by receptor-dependent C-terminal phosphorylations that facilitate IRF oligomerization, nuclear translocation and assembly with co-activators, such as p300 and CBP (Qin et al. 2005; Takahashi et al. 2010). Remarkably, activated IRF3 can also bind directly to Smad3 and inhibit its activation by T $\beta$ R-I or block its assembly with Smad4 and additional co-activators in the nucleus (Xu et al. 2014). Further analyses are necessary to unravel the structural basis of this cross talk.

The precise molecular mechanisms that determine IRF and R-Smad activation also display notable differences. Uniquely, the IRF proteins contain an additional  $\alpha$ -helix at their C-terminus that folds back onto the helical bundle motif to form an autoinhibitory interaction (Qin et al. 2003). Phosphorylations within the intervening C-terminal extension break this interaction and allow the C-terminus to instead mediate IRF dimerization, which is therefore distinct from the

trimerization of Smads (Chen et al. 2008; Takahashi et al. 2010). Furthermore, this switch exposes a hydrophobic pocket in the helical bundle motif that mediates further interaction with the co-activators p300/CBP (Qin et al. 2005). While the structure of IRF3 has been solved in complex with CBP (Qin et al. 2005), there are currently no available structures of the equivalent R-Smad complex. It will be interesting in future to see how these complexes compare.

## **TRANSCRIPTIONAL ROLES OF SMADS**

### **Smad-DNA Interaction**

Interactions with DNA are mediated by the N-terminal MH1 domains of the R-Smads and Smad4. The precise binding mode was first revealed by the crystal structure of the MH1 domain of Smad3 which was solved in complex with a canonical Smad-binding element (SBE) including the core palindromic sequence GTCTAGAC (Shi et al. 1998; Chai et al. 2003). The MH1 domain adopts an  $\alpha/\beta$  fold stabilized by a zinc atom that is coordinated by three cysteine residues and a histidine (Figure 4A). Two MH1 domains bind on opposite faces of the DNA to each half site (or Smad box) of the SBE using a  $\beta$ -hairpin structure that inserts into the major groove. Here, specific hydrogen bonding interactions with the bases are formed by the three conserved residues Arg74, Gln76, and Lys81. Additional interactions with the phosphate backbone are made by both the  $\beta$ -hairpin motif and the large  $\alpha 2$  helix.

Similar complex structures have been solved subsequently for the MH1 domains of Smad1 (BabuRajendran et al. 2010) and Smad4 (Baburajendran et al. 2011). While both proteins display the same overall fold and specific base contacts, subtle differences are observed that likely contribute to the DNA-binding specificity of the different Smads. The MH1 domain of

Smad1 displays a more open conformation in which the  $\alpha$ 1 helix is dissociated to form a domain-swapped dimer. As a result, a subtle shift in the packing disrupts interactions between the  $\alpha$ 2 helix and the phosphate backbone (Figure 4B). Some flexibility in the  $\alpha$ 1 site is suggested by the fact that the protein is monomeric in solution (BabuRajendran et al. 2010) and can also form 1:1 complexes with the SBE at low concentrations (Baburajendran et al. 2011). By contrast, the MH1 domain of Smad4 forms exclusively 2:1 complexes with the SBE yet displays no evidence of dimeric protein-protein contacts across the DNA duplex (Baburajendran et al. 2011). Sequence alignments suggest that the remaining Smad family members will also yield similar structures, except for Smad2 which fails to bind DNA due to a 30-residue insertion prior to the critical  $\beta$ -hairpin motif (Dennler et al. 1999; Yagi et al. 1999).

Unfortunately, there are no crystal structures of a heterotrimeric Smad complex bound to DNA. Tandem repeats of the SBE have been shown to enhance TGF- $\beta$  signaling suggestive of a cooperative binding mode (Zawel et al. 1998). Molecular dynamics simulations suggest that the interaction of Smad4 with the DNA facilitates the cooperative recruitment of associated R-Smads (Wang et al. 2013). The presence of mixed Smad complexes, as well as the potential inclusion of Smad2, is expected to confer different affinities, and therefore specificity, for sites with variable spacings or orientations of the SBE motifs. The precise transcriptional response will additionally dependent on the combinatorial assembly of other transcription factors on the promoter elements of Smad-target genes. An increasing number of transcription factors have been shown to interact directly with Smads. While there are currently no available structural models, some of these proteins are known to compete directly for the hydrophobic pocket of the Smad MH2 domain that is bound by the proline-rich Smad-binding domain of SARA (Randall et al. 2002).

### **Smad interactions with the co-repressor Ski**

Transactivation by Smads is additionally regulated by a number of transcriptional co-activators and co-repressors (Ross and Hill 2008; Massague 2012). Among the best characterized are Ski and its homologue SnoN, which recruit the co-repressors N-CoR and mSin3A to repress TGF- $\beta$  signaling (Deheuninck and Luo 2009). Both proteins contain an N-terminal DHD (Dachshund homology domain) and a Smad4-binding domain. Structures of the DHD domain of Ski and SnoN show a conserved  $\alpha/\beta$  fold with some features reminiscent of the forkhead/winged-helix family of DNA-binding proteins (Wilson et al. 2004; Nyman et al. 2010). However, neither protein is thought to bind directly to DNA. Instead, the DHD may function as a protein-protein interaction domain to enable co-repressor recruitment (Deheuninck and Luo 2009).

A crystal structure has also been solved of the Smad4-binding domain of Ski in complex with the MH2 domain of Smad4 (Figure 3C) (Wu et al. 2002). The Ski domain adopts the SAND domain fold, named after the nuclear proteins Sp100, AIRE-1, NucP41/75, and DEAF-1. A twisted five-stranded antiparallel  $\beta$ -sheet packs on the concave face with three short  $\alpha$ -helices and a large C-terminal  $\alpha$ -helix. Unusually, one side of the  $\beta$ -sheet also contains a Cys<sub>2</sub>-His<sub>2</sub>-type zinc-binding motif that appears critical for folding. This motif extends into a large interaction loop (I loop) that binds the Smad4 MH2 domain, rather than forming the typical DNA interaction (Wu et al. 2002). When mapped onto the alternative structure of the Smad2-Smad4 complex, the Ski subunit is positioned on the outer edge of the Smad trimer, where its  $\beta$ -sheet would buttress against the neighbouring Smad2 tail. The model predicts a severe clash between Ski and Smad2 Ser418 to Arg420. Indeed, the purified Ski fragment is sufficient to disrupt the Smad2-Smad4 interaction *in vitro*, suggesting an additional mechanism for Ski-mediated transcriptional repression (Wu et al. 2002).

This conclusion has been challenged by a number of experiments using Ski fragments in cells. Such analyses are complicated by the presence of a separate region in Ski that can bind instead to Smad2 or Smad3 (Ueki and Hayman 2003). Nonetheless, a Ski mutant containing only the functional Smad4-binding site could still bind simultaneously to Smad4 and the R-Smads (Ueki and Hayman 2003; Takeda et al. 2004). Moreover, a pull down study has revealed the accumulation of this inactive complex at target promoter elements (Suzuki et al. 2004). Finally, Ski has also been reported to bind to the receptor T $\beta$ R-I and to trap non-functional T $\beta$ R-I-R-Smad-Smad4 complexes in the cytoplasm (Ferrand et al. 2010). Thus, further structural studies are needed to unravel the complex nature of these interactions. In particular, the large I-loop of Ski is somewhat separated from the globular core and may provide some limited flexibility to adapt to different protein complexes.

### **Smad linker interactions**

Following their translocation to the nucleus, Smads are subject to multiple phosphorylations within their central linker region that determine Smad transcriptional activation and turnover (Massague 2012; Sudol 2012; Gaarenstroom and Hill 2014). These posttranslational modifications are read by modular WW domain-containing proteins, including transcriptional effectors, inhibitory Smads and E3 ubiquitin ligases (Chong et al. 2006; Alarcon et al. 2009; Chong et al. 2010; Aragon et al. 2011; Aragon et al. 2012). The WW domain, named after two conserved tryptophans, is a small domain of approximately 40 residues that typically occurs in a tandem repeat manner with up to four sequential copies (Bork and Sudol 1994; Kato et al. 2004). Despite their small size, WW domains are folded into a stable, three-stranded  $\beta$ -sheet. Proline-rich peptide sequences are bound on one face of the  $\beta$ -sheet stabilized by stacking interactions with the conserved C-terminal tryptophan of the WW domain.



Central to this regulation are the MAP kinases and the cyclin-dependent protein kinases (CDKs) that selectively phosphorylate conserved Ser/Thr-Pro motifs. In particular, CDK8 and CDK9 have been shown to promote Smad transcriptional activation. These kinases can phosphorylate Smad1 and Smad3 to establish binding sites for the WW domains of the Hippo pathway effector YAP and the prolyl isomerase PIN1, respectively (Alarcon et al. 2009; Aragon et al. 2011). Subsequent linker phosphorylation by GSK3 then directs Smad1 and Smad3 to the WW domains of the E3 ligases, Smurf1 and Nedd4L, respectively (Figure 5).

Structural and biophysical studies have shown that YAP binds to the Smad1 linker region in a bipartite manner (Aragon et al. 2011). The first WW domain binds to a phosphorylated Pro-X-pSer-Pro motif, whereas the second WW domain binds to an unphosphorylated Pro-Pro-X-Tyr motif slightly downstream (Figure 5). Notably, the two peptide classes bind to WW domains in opposite orientations. The Smad1 phosphopeptide adopts a rather linear conformation that positions the C-terminal proline (Pro207) against YAP Trp199. Phospho-Ser206 is also within hydrogen bonding distance of YAP Thr182, Tyr188 and Gln186 (Aragon et al. 2011). By comparison, the Smad1 region containing the Pro-Pro-X-Tyr motif binds the second WW domain in a bent conformation that positions the N-terminal Pro224 against YAP Trp258.

Interestingly, the first WW domain of YAP can also revert to binding a Pro-Pro-X-Tyr motif, as demonstrated by the inhibitory protein Smad7 (Aragon et al. 2012). This alternative interaction is facilitated by the local rearrangement of the YAP side chains Tyr188 and Trp199. While the overall binding affinity is reduced, it is compensated for by high local concentrations of Smad7 in the nucleus (Aragon et al. 2012).

Upon further phosphorylation, the binding of Smad1 is switched to the HECT-type E3 ligase Smurf1 for degradation by the proteasome. The first WW domain of Smurf1 binds preferentially

to a Smad1 peptide doubly phosphorylated on Ser210 and Ser214 (Aragon et al. 2011). Ser214 falls within a consensus Pro-X-pSer-Pro motif, while the additional binding of phospho-Ser210 is enabled by the extra interaction of Smurf1 Arg243 (Figure 5). Similar structural mechanisms control the fate of multiply phosphorylated Smad2 and Smad3, which switch their binding from PIN1 to the E3 ligase Nedd4L (Kuratomi et al. 2005; Gao et al. 2009; Aragon et al. 2011).

## **CONCLUDING REMARKS**

An array of new structures have defined the ordered phosphorylation-dependent activation of the TGF- $\beta$  pathway from the cell membrane to the nucleus. Recent work has also identified other additional regulatory mechanisms that await structural analysis, including novel posttranslational modifications, such as arginine methylation (Inamitsu et al. 2006; Xu et al. 2013) and sumoylation (Lin et al. 2003; Kang et al. 2008). Challenges also remain to elucidate the structures of protein complexes fundamental to TGF- $\beta$  signaling. Most important in this regard, are the interactions of the GS domain with the type II receptor and the R-Smad. Similarly, there are no MH2 domain structures of the inhibitory Smads, Smad6 and Smad7, which can bind constitutively to the type I receptors and promote recruitment of Smurf E3 ligases for their degradation. Future studies promise to reveal yet more of the elegant structural solutions that have evolved to coordinate TGF- $\beta$  signaling strength and duration as well as the cross-talk with other cellular processes.

## **Acknowledgements**

The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no.

115766], Janssen, Merck & Co., Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome Trust [092809/Z/10/Z].

## References

- Abdollah S, Macias-Silva M, Tsukazaki T, Hayashi H, Attisano L, Wrana JL. 1997. TbetarI phosphorylation of Smad2 on Ser465 and Ser467 is required for Smad2-Smad4 complex formation and signaling. *J Biol Chem* **272**: 27678-27685.
- Alarcon C, Zaromytidou AI, Xi Q, Gao S, Yu J, Fujisawa S, Barlas A, Miller AN, Manova-Todorova K, Macias MJ et al. 2009. Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. *Cell* **139**: 757-769.
- Aragon E, Goerner N, Xi Q, Gomes T, Gao S, Massague J, Macias MJ. 2012. Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF-beta Pathways. *Structure* **20**: 1726-1736.
- Aragon E, Goerner N, Zaromytidou AI, Xi Q, Escobedo A, Massague J, Macias MJ. 2011. A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes Dev* **25**: 1275-1288.
- Baburajendran N, Jauch R, Tan CY, Narasimhan K, Kolatkar PR. 2011. Structural basis for the cooperative DNA recognition by Smad4 MH1 dimers. *Nucleic Acids Res* **39**: 8213-8222.
- BabuRajendran N, Palasingam P, Narasimhan K, Sun W, Prabhakar S, Jauch R, Kolatkar PR. 2010. Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors. *Nucleic Acids Res* **38**: 3477-3488.
- Bork P, Sudol M. 1994. The WW domain: a signalling site in dystrophin? *Trends Biochem Sci* **19**: 531-533.
- Buczkowicz P, Hoeman C, Rakopoulos P, Pajovic S, Letourneau L, Dzamba M, Morrison A, Lewis P, Bouffet E, Bartels U et al. 2014. Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations. *Nat Genet* **46**: 451-456.
- Chacko BM, Qin B, Correia JJ, Lam SS, de Caestecker MP, Lin K. 2001. The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization. *Nat Struct Biol* **8**: 248-253.
- Chacko BM, Qin BY, Tiwari A, Shi G, Lam S, Hayward LJ, De Caestecker M, Lin K. 2004. Structural basis of heteromeric smad protein assembly in TGF-beta signaling. *Mol Cell* **15**: 813-823.
- Chai J, Wu JW, Yan N, Massague J, Pavletich NP, Shi Y. 2003. Features of a Smad3 MH1-DNA complex. Roles of water and zinc in DNA binding. *J Biol Chem* **278**: 20327-20331.
- Chaikuad A, Alfano I, Kerr G, Sanvitale CE, Boergermann JH, Triffitt JT, von Delft F, Knapp S, Knaus P, Bullock AN. 2012. Structure of the bone morphogenetic protein receptor ALK2 and implications for fibrodysplasia ossificans progressiva. *J Biol Chem* **287**: 36990-36998.
- Chen W, Lam SS, Srinath H, Jiang Z, Correia JJ, Schiffer CA, Fitzgerald KA, Lin K, Royer WE, Jr. 2008. Insights into interferon regulatory factor activation from the crystal structure of dimeric IRF5. *Nat Struct Mol Biol* **15**: 1213-1220.
- Chen YG, Hata A, Lo RS, Wotton D, Shi Y, Pavletich N, Massague J. 1998. Determinants of specificity in TGF-beta signal transduction. *Genes Dev* **12**: 2144-2152.
- Chen YG, Massague J. 1999. Smad1 recognition and activation by the ALK1 group of transforming growth factor-beta family receptors. *J Biol Chem* **274**: 3672-3677.

- Chen YG, Wang Z, Ma J, Zhang L, Lu Z. 2007. Endofin, a FYVE domain protein, interacts with Smad4 and facilitates transforming growth factor-beta signaling. *J Biol Chem* **282**: 9688-9695.
- Chong PA, Lin H, Wrana JL, Forman-Kay JD. 2006. An expanded WW domain recognition motif revealed by the interaction between Smad7 and the E3 ubiquitin ligase Smurf2. *J Biol Chem* **281**: 17069-17075.
- . 2010. Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity. *Proc Natl Acad Sci U S A* **107**: 18404-18409.
- Correia JJ, Chacko BM, Lam SS, Lin K. 2001. Sedimentation studies reveal a direct role of phosphorylation in Smad3:Smad4 homo- and hetero-trimerization. *Biochemistry* **40**: 1473-1482.
- Daly AC, Randall RA, Hill CS. 2008. Transforming growth factor beta-induced Smad1/5 phosphorylation in epithelial cells is mediated by novel receptor complexes and is essential for anchorage-independent growth. *Mol Cell Biol* **28**: 6889-6902.
- Deheuninck J, Luo K. 2009. Ski and SnoN, potent negative regulators of TGF-beta signaling. *Cell Res* **19**: 47-57.
- Dennler S, Huet S, Gauthier JM. 1999. A short amino-acid sequence in MH1 domain is responsible for functional differences between Smad2 and Smad3. *Oncogene* **18**: 1643-1648.
- Durocher D, Taylor IA, Sarbassova D, Haire LF, Westcott SL, Jackson SP, Smerdon SJ, Yaffe MB. 2000. The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol Cell* **6**: 1169-1182.
- Endicott JA, Noble ME, Johnson LN. 2012. The structural basis for control of eukaryotic protein kinases. *Annu Rev Biochem* **81**: 587-613.
- Feng XH, Derynck R. 1997. A kinase subdomain of transforming growth factor-beta (TGF-beta) type I receptor determines the TGF-beta intracellular signaling specificity. *EMBO J* **16**: 3912-3923.
- Ferrand N, Atfi A, Prunier C. 2010. The oncoprotein c-ski functions as a direct antagonist of the transforming growth factor-{beta} type I receptor. *Cancer Res* **70**: 8457-8466.
- Fontebasso AM, Papillon-Cavanagh S, Schwartzentruber J, Nikbakht H, Gerges N, Fiset PO, Bechet D, Fauray D, De Jay N, Ramkissoon LA et al. 2014. Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma. *Nat Genet* **46**: 462-466.
- Fukuda T, Kanomata K, Nojima J, Kokabu S, Akita M, Ikebuchi K, Jimi E, Komori T, Maruki Y, Matsuoka M et al. 2008. A unique mutation of ALK2, G356D, found in a patient with fibrodysplasia ossificans progressiva is a moderately activated BMP type I receptor. *Biochem Biophys Res Commun* **377**: 905-909.
- Fukuda T, Kohda M, Kanomata K, Nojima J, Nakamura A, Kamizono J, Noguchi Y, Iwakiri K, Kondo T, Kurose J et al. 2009. Constitutively activated ALK2 and increased SMAD1/5 cooperatively induce bone morphogenetic protein signaling in fibrodysplasia ossificans progressiva. *J Biol Chem* **284**: 7149-7156.
- Gaarenstroom T, Hill CS. 2014. TGF-beta signaling to chromatin: How Smads regulate transcription during self-renewal and differentiation. *Semin Cell Dev Biol* **32C**: 107-118.
- Gao S, Alarcon C, Sapkota G, Rahman S, Chen PY, Goerner N, Macias MJ, Erdjument-Bromage H, Tempst P, Massague J. 2009. Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling. *Mol Cell* **36**: 457-468.
- Han S, Loulakis P, Griffor M, Xie Z. 2007. Crystal structure of activin receptor type IIB kinase domain from human at 2.0 Angstrom resolution. *Protein Sci* **16**: 2272-2277.
- Hatsell SJ, Idone V, Wolken DM, Huang L, Kim HJ, Wang L, Wen X, Nannuru KC, Jimenez J, Xie L et al. 2015. ACVR1R206H receptor mutation causes fibrodysplasia ossificans progressiva by imparting responsiveness to activin A. *Sci Transl Med* **7**: 303ra137.

- Hino K, Ikeya M, Horigome K, Matsumoto Y, Ebise H, Nishio M, Sekiguchi K, Shibata M, Nagata S, Matsuda S et al. 2015. Neofunction of ACVR1 in fibrodysplasia ossificans progressiva. *Proc Natl Acad Sci U S A* **112**: 15438-15443.
- Huse M, Chen YG, Massague J, Kuriyan J. 1999. Crystal structure of the cytoplasmic domain of the type I TGF beta receptor in complex with FKBP12. *Cell* **96**: 425-436.
- Huse M, Muir TW, Xu L, Chen YG, Kuriyan J, Massague J. 2001. The TGF beta receptor activation process: an inhibitor- to substrate-binding switch. *Mol Cell* **8**: 671-682.
- Inamitsu M, Itoh S, Hellman U, Ten Dijke P, Kato M. 2006. Methylation of Smad6 by protein arginine N-methyltransferase 1. *FEBS Lett* **580**: 6603-6611.
- Kang JS, Saunier EF, Akhurst RJ, Derynck R. 2008. The type I TGF-beta receptor is covalently modified and regulated by sumoylation. *Nat Cell Biol* **10**: 654-664.
- Kaplan FS, Xu M, Seemann P, Connor JM, Glaser DL, Carroll L, Delai P, Fastnacht-Urban E, Forman SJ, Gillissen-Kaesbach G et al. 2009. Classic and atypical fibrodysplasia ossificans progressiva (FOP) phenotypes are caused by mutations in the bone morphogenetic protein (BMP) type I receptor ACVR1. *Hum Mutat* **30**: 379-390.
- Kato Y, Nagata K, Takahashi M, Lian L, Herrero JJ, Sudol M, Tanokura M. 2004. Common mechanism of ligand recognition by group II/III WW domains: redefining their functional classification. *J Biol Chem* **279**: 31833-31841.
- Kawabata M, Inoue H, Hanyu A, Imamura T, Miyazono K. 1998. Smad proteins exist as monomers in vivo and undergo homo- and hetero-oligomerization upon activation by serine/threonine kinase receptors. *EMBO J* **17**: 4056-4065.
- Krupa A, Preethi G, Srinivasan N. 2004. Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases. *J Mol Biol* **339**: 1025-1039.
- Kuratomi G, Komuro A, Goto K, Shinozaki M, Miyazawa K, Miyazono K, Imamura T. 2005. NEDD4-2 (neural precursor cell expressed, developmentally down-regulated 4-2) negatively regulates TGF-beta (transforming growth factor-beta) signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. *Biochem J* **386**: 461-470.
- Lagna G, Hata A, Hemmati-Brivanlou A, Massague J. 1996. Partnership between DPC4 and SMAD proteins in TGF-beta signalling pathways. *Nature* **383**: 832-836.
- Lawler S, Feng XH, Chen RH, Maruoka EM, Turck CW, Griswold-Prenner I, Derynck R. 1997. The type II transforming growth factor-beta receptor autophosphorylates not only on serine and threonine but also on tyrosine residues. *J Biol Chem* **272**: 14850-14859.
- Lee MK, Pardoux C, Hall MC, Lee PS, Warburton D, Qing J, Smith SM, Derynck R. 2007. TGF-beta activates Erk MAP kinase signalling through direct phosphorylation of ShcA. *EMBO J* **26**: 3957-3967.
- Lin X, Liang M, Liang YY, Brunicardi FC, Melchior F, Feng XH. 2003. Activation of transforming growth factor-beta signaling by SUMO-1 modification of tumor suppressor Smad4/DPC4. *J Biol Chem* **278**: 18714-18719.
- Lo RS, Chen YG, Shi Y, Pavletich NP, Massague J. 1998. The L3 loop: a structural motif determining specific interactions between SMAD proteins and TGF-beta receptors. *EMBO J* **17**: 996-1005.
- Macias-Silva M, Hoodless PA, Tang SJ, Buchwald M, Wrana JL. 1998. Specific activation of Smad1 signaling pathways by the BMP7 type I receptor, ALK2. *J Biol Chem* **273**: 25628-25636.
- Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002. The protein kinase complement of the human genome. *Science* **298**: 1912-1934.
- Massague J. 2012. TGFbeta signalling in context. *Nat Rev Mol Cell Biol* **13**: 616-630.
- Michnick SW, Rosen MK, Wandless TJ, Karplus M, Schreiber SL. 1991. Solution structure of FKBP, a rotamase enzyme and receptor for FK506 and rapamycin. *Science* **252**: 836-839.

- Miura S, Takeshita T, Asao H, Kimura Y, Murata K, Sasaki Y, Hanai JI, Beppu H, Tsukazaki T, Wrana JL et al. 2000. Hgs (Hrs), a FYVE domain protein, is involved in Smad signaling through cooperation with SARA. *Mol Cell Biol* **20**: 9346-9355.
- Moustakas A, Heldin CH. 2002. From mono- to oligo-Smads: the heart of the matter in TGF-beta signal transduction. *Genes Dev* **16**: 1867-1871.
- Nyman T, Tresaugues L, Welin M, Lehtio L, Flodin S, Persson C, Johansson I, Hammarstrom M, Nordlund P. 2010. The crystal structure of the Dachshund domain of human SnoN reveals flexibility in the putative protein interaction surface. *PLoS One* **5**: e12907.
- Persson U, Izumi H, Souchelnytskyi S, Itoh S, Grimsby S, Engstrom U, Heldin CH, Funa K, ten Dijke P. 1998. The L45 loop in type I receptors for TGF-beta family members is a critical determinant in specifying Smad isoform activation. *FEBS Lett* **434**: 83-87.
- Qin B, Lam SS, Lin K. 1999. Crystal structure of a transcriptionally active Smad4 fragment. *Structure* **7**: 1493-1503.
- Qin BY, Lam SS, Correia JJ, Lin K. 2002. Smad3 allostery links TGF-beta receptor kinase activation to transcriptional control. *Genes Dev* **16**: 1950-1963.
- Qin BY, Liu C, Lam SS, Srinath H, Delston R, Correia JJ, Derynck R, Lin K. 2003. Crystal structure of IRF-3 reveals mechanism of autoinhibition and virus-induced phosphoactivation. *Nat Struct Biol* **10**: 913-921.
- Qin BY, Liu C, Srinath H, Lam SS, Correia JJ, Derynck R, Lin K. 2005. Crystal structure of IRF-3 in complex with CBP. *Structure* **13**: 1269-1277.
- Randall RA, Germain S, Inman GJ, Bates PA, Hill CS. 2002. Different Smad2 partners bind a common hydrophobic pocket in Smad2 via a defined proline-rich motif. *EMBO J* **21**: 145-156.
- Ross S, Hill CS. 2008. How the Smads regulate transcription. *Int J Biochem Cell Biol* **40**: 383-408.
- Shen Q, Little SC, Xu M, Haupt J, Ast C, Katagiri T, Mundlos S, Seemann P, Kaplan FS, Mullins MC et al. 2009. The fibrodysplasia ossificans progressiva R206H ACVR1 mutation activates BMP-independent chondrogenesis and zebrafish embryo ventralization. *J Clin Invest* **119**: 3462-3472.
- Shi W, Chang C, Nie S, Xie S, Wan M, Cao X. 2007. Endofin acts as a Smad anchor for receptor activation in BMP signaling. *J Cell Sci* **120**: 1216-1224.
- Shi Y, Hata A, Lo RS, Massague J, Pavletich NP. 1997. A structural basis for mutational inactivation of the tumour suppressor Smad4. *Nature* **388**: 87-93.
- Shi Y, Wang YF, Jayaraman L, Yang H, Massague J, Pavletich NP. 1998. Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell* **94**: 585-594.
- Shore EM, Xu M, Feldman GJ, Fenstermacher DA, Cho TJ, Choi IH, Connor JM, Delai P, Glaser DL, LeMerrer M et al. 2006. A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat Genet* **38**: 525-527.
- Song GA, Kim HJ, Woo KM, Baek JH, Kim GS, Choi JY, Ryoo HM. 2010. Molecular consequences of the ACVR1(R206H) mutation of fibrodysplasia ossificans progressiva. *J Biol Chem* **285**: 22542-22553.
- Souchelnytskyi S, Tamaki K, Engstrom U, Wernstedt C, ten Dijke P, Heldin CH. 1997. Phosphorylation of Ser465 and Ser467 in the C terminus of Smad2 mediates interaction with Smad4 and is required for transforming growth factor-beta signaling. *J Biol Chem* **272**: 28107-28115.
- Souchelnytskyi S, ten Dijke P, Miyazono K, Heldin CH. 1996. Phosphorylation of Ser165 in TGF-beta type I receptor modulates TGF-beta1-induced cellular responses. *EMBO J* **15**: 6231-6240.
- Sudol M. 2012. WW domains in the heart of Smad regulation. *Structure* **20**: 1619-1620.
- Suzuki H, Yagi K, Kondo M, Kato M, Miyazono K, Miyazawa K. 2004. c-Ski inhibits the TGF-beta signaling pathway through stabilization of inactive Smad complexes on Smad-binding elements. *Oncogene* **23**: 5068-5076.

- Takahashi K, Horiuchi M, Fujii K, Nakamura S, Noda NN, Yoneyama M, Fujita T, Inagaki F. 2010. Ser386 phosphorylation of transcription factor IRF-3 induces dimerization and association with CBP/p300 without overall conformational change. *Genes Cells* **15**: 901-910.
- Takahashi K, Suzuki NN, Horiuchi M, Mori M, Suhara W, Okabe Y, Fukuhara Y, Terasawa H, Akira S, Fujita T et al. 2003. X-ray crystal structure of IRF-3 and its functional implications. *Nat Struct Biol* **10**: 922-927.
- Takeda M, Mizuide M, Oka M, Watabe T, Inoue H, Suzuki H, Fujita T, Imamura T, Miyazono K, Miyazawa K. 2004. Interaction with Smad4 is indispensable for suppression of BMP signaling by c-Ski. *Mol Biol Cell* **15**: 963-972.
- Taylor KR, Mackay A, Truffaux N, Butterfield YS, Morozova O, Philippe C, Castel D, Grasso CS, Vinci M, Carvalho D et al. 2014. Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma. *Nat Genet* **46**: 457-461.
- Tsukazaki T, Chiang TA, Davison AF, Attisano L, Wrana JL. 1998. SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor. *Cell* **95**: 779-791.
- Ueki N, Hayman MJ. 2003. Direct interaction of Ski with either Smad3 or Smad4 is necessary and sufficient for Ski-mediated repression of transforming growth factor-beta signaling. *J Biol Chem* **278**: 32489-32492.
- van Dinther M, Visser N, de Gorter DJ, Doorn J, Goumans MJ, de Boer J, ten Dijke P. 2010. ALK2 R206H mutation linked to fibrodysplasia ossificans progressiva confers constitutive activity to the BMP type I receptor and sensitizes mesenchymal cells to BMP-induced osteoblast differentiation and bone formation. *J Bone Miner Res* **25**: 1208-1215.
- Wang G, Li C, Wang Y, Chen G. 2013. Cooperative assembly of Co-Smad4 MH1 with R-Smad1/3 MH1 on DNA: a molecular dynamics simulation study. *PLoS One* **8**: e53841.
- Wieser R, Wrana JL, Massague J. 1995. GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex. *EMBO J* **14**: 2199-2208.
- Wilson JJ, Malakhova M, Zhang R, Joachimiak A, Hegde RS. 2004. Crystal structure of the dachshund homology domain of human SKI. *Structure* **12**: 785-792.
- Wrana JL, Attisano L, Wieser R, Ventura F, Massague J. 1994. Mechanism of activation of the TGF-beta receptor. *Nature* **370**: 341-347.
- Wu G, Chen YG, Ozdamar B, Gyuricza CA, Chong PA, Wrana JL, Massague J, Shi Y. 2000. Structural basis of Smad2 recognition by the Smad anchor for receptor activation. *Science* **287**: 92-97.
- Wu G, Diaz AK, Paugh BS, Rankin SL, Ju B, Li Y, Zhu X, Qu C, Chen X, Zhang J et al. 2014. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat Genet* **46**: 444-450.
- Wu JW, Hu M, Chai J, Seoane J, Huse M, Li C, Rigotti DJ, Kyin S, Muir TW, Fairman R et al. 2001. Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. *Mol Cell* **8**: 1277-1289.
- Wu JW, Krawitz AR, Chai J, Li W, Zhang F, Luo K, Shi Y. 2002. Structural mechanism of Smad4 recognition by the nuclear oncoprotein Ski: insights on Ski-mediated repression of TGF-beta signaling. *Cell* **111**: 357-367.
- Xu J, Wang AH, Oses-Prieto J, Makhijani K, Katsuno Y, Pei M, Yan L, Zheng YG, Burlingame A, Bruckner K et al. 2013. Arginine Methylation Initiates BMP-Induced Smad Signaling. *Mol Cell* **51**: 5-19.
- Xu P, Bailey-Bucktrout S, Xi Y, Xu D, Du D, Zhang Q, Xiang W, Liu J, Melton A, Sheppard D et al. 2014. Innate antiviral host defense attenuates TGF-beta function through IRF3-mediated suppression of Smad signaling. *Mol Cell* **56**: 723-737.
- Yagi K, Goto D, Hamamoto T, Takenoshita S, Kato M, Miyazono K. 1999. Alternatively spliced variant of Smad2 lacking exon 3. Comparison with wild-type Smad2 and Smad3. *J Biol Chem* **274**: 703-709.

Zawel L, Dai JL, Buckhaults P, Zhou S, Kinzler KW, Vogelstein B, Kern SE. 1998. Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell* 1: 611-617.

### Figure legends

**Figure 1.** Structures of the receptor kinase domains. (A) Domain organisation of the type I and type II receptors. (B) Structure of the kinase domain of the type II receptor ActRIIB. Secondary structural elements are labelled as well as the kinase activation loop (A-loop, magenta) and the hinge region, which connects the kinase N and C-lobes. (C) Regulation of T $\beta$ R-I by FKBP12. (Left panel) Structure of the T $\beta$ R-I GS (yellow) and kinase (brown) domains in complex with the inhibitory protein FKBP12 (grey). (Right panel) Structure of the T $\beta$ R-I GS and kinase domains in the absence of FKBP12. Inset boxes highlight the different conformational changes between the two structures. FKBP12 binding pushes the GS loop against the kinase domain  $\alpha$ C helix (upper box) causing it to angle into the front of the ATP pocket (lower box).

**Figure 2.** R-Smad interactions with SARA and the type I receptor. (A) Structure of the Smad2 MH2 domain in complex with the Smad-binding domain (SBD) of SARA. The MH2 domain consists of a  $\beta$ -sandwich capped by the H1 helix (lilac), a three helix bundle (magenta) and a flexible tail containing a C-terminal Ser-X-Ser motif (SSVS in Smad2; dashed line indicates that this region was disordered in this structure). The basic patch (lilac) and L3 loop (brown) implicated in interactions with the type I receptor are also highlighted. (B) Model for R-Smad interaction with the type I receptor.

**Figure 3.** Heteromeric Smad assembly. (A) Structure of the Smad4 MH2 domain. The N-terminal Smad activation domain (SAD) is coloured yellow. The three helix bundle (magenta) contains a 35-residue insertion. (B) Structure of the Smad2-Smad4 heterotrimer. Phosphorylated serines in the Smad2 Ser-X-Ser motif are displayed in spacefill representation



and bind residues from the basic patch. (C) Structure of the Smad4-Ski complex. The  $\beta$ -sheet and I-loop of Ski are coloured yellow and red, respectively. The bound zinc ion in Ski is coloured blue.

**Figure 4.** Smad MH1 domain interactions with DNA. (A) DNA co-structure of the Smad3 MH1. Two MH1 domains bind on opposite faces of the DNA to each half site (or Smad box) of the Smad-binding element (SBE) using a  $\beta$ -hairpin structure that inserts into the major groove. (B) DNA co-structure of the Smad1 MH1 domain. The MH1 domain of Smad1 displays a more open conformation; the  $\alpha$ 1 helix is dissociated causing a subtle shift in the position of the  $\alpha$ 2 helix that disrupts some interactions with the phosphate backbone.

**Figure 5.** Smad linker interactions determining transcriptional activation and turnover. (A) Schematic representation of the Smad linker and its binding to different WW domain proteins upon sequential phosphorylation by CDK8/9 and GSK3, respectively. Structures are shown for Smad1 binding to the first and second WW domains of (B) YAP and (C) Smurf1, respectively.

Figure 1

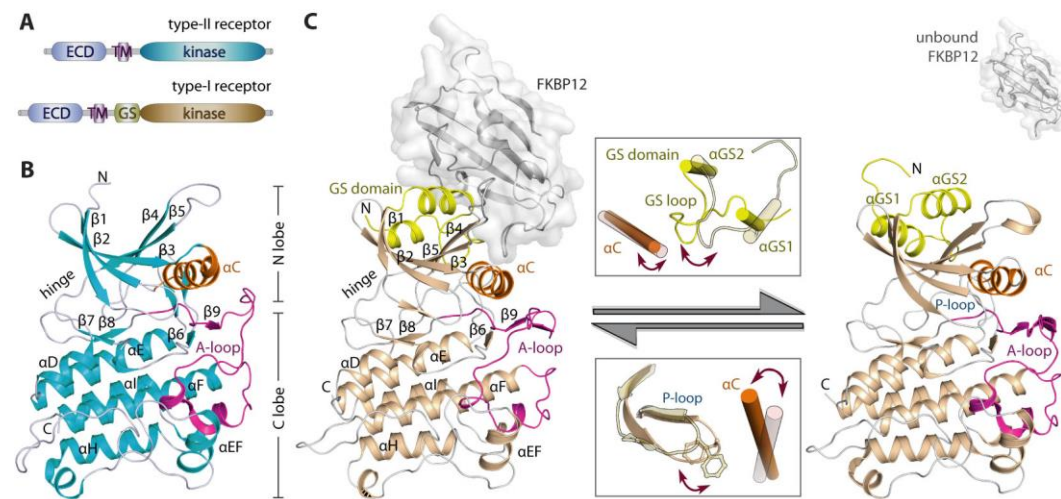


Figure 2

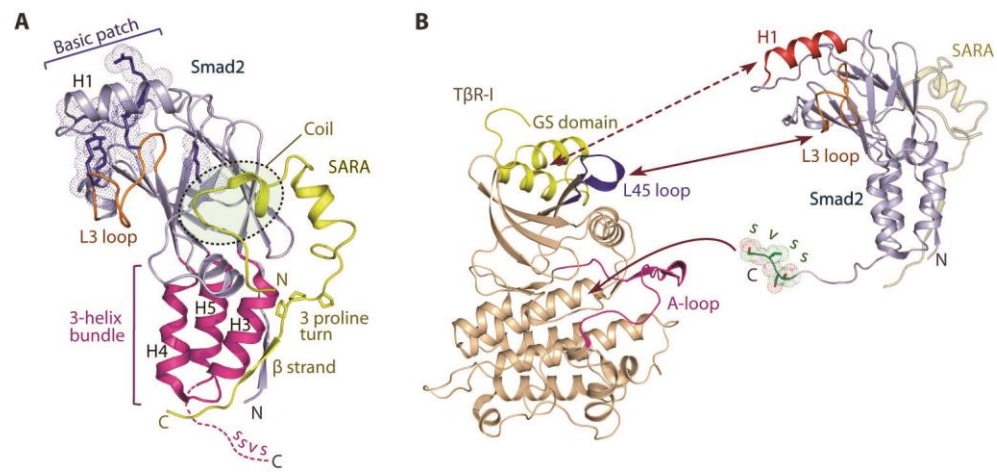


Figure 3

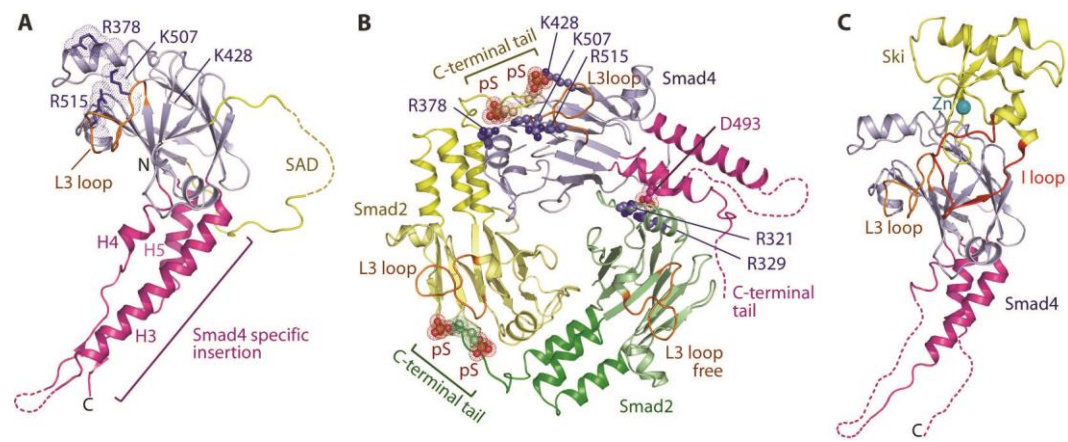


Figure 4

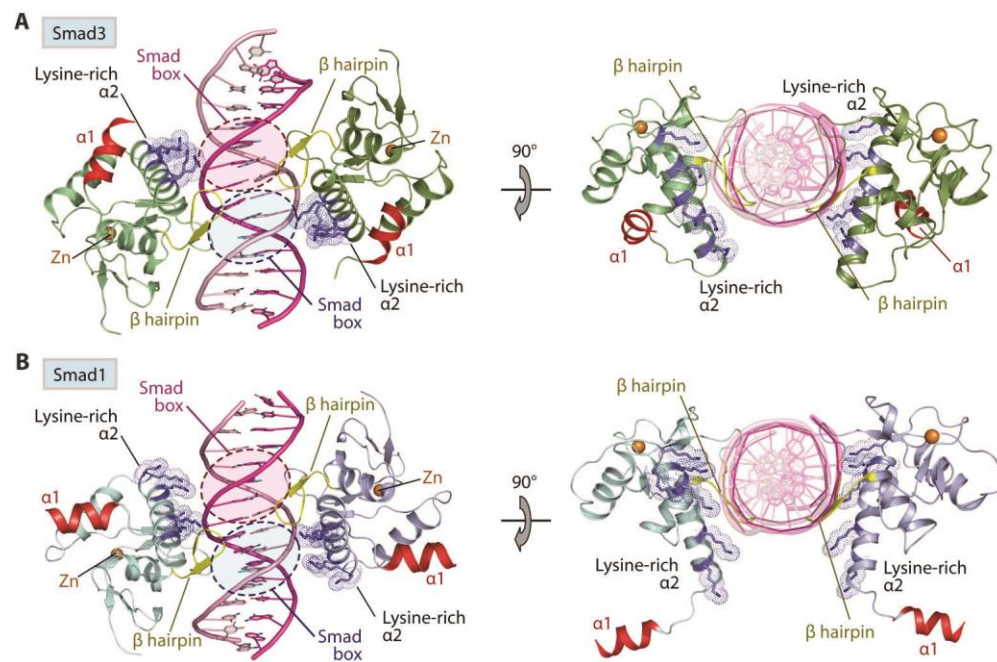


Figure 5

