Molecular Cell, Vol. 8, 1277–1289, December, 2001, Copyright ©2001 by Cell Press

Crystal Structure of a Phosphorylated Smad2: Recognition of Phosphoserine by the MH2 Domain and Insights on Smad Function in TGF-β Signaling

Jia-Wei Wu,[1] Min Hu,[1] Jijie Chai,[1] Joan Seoane,[2] Morgan Huse,[3] Carey Li,[3] Daniel J. Rigotti,[4] Saw Kyin,[1] Tom W. Muir,[3] Robert Fairman,[4] Joan Massagué,[2] and Yigong Shi[1,5]

[1] Department of Molecular Biology
Lewis Thomas Laboratory
Princeton University
Princeton, New Jersey 08544

[2] Cell Biology Program
Howard Hughes Medical Institute
Memorial Sloan-Kettering Cancer Center
New York, New York 10021

[3] Laboratory of Synthetic Protein Chemistry
Rockefeller University
1230 York Avenue
New York, New York 10021

[4] Department of Biology
370 Lancaster Avenue
Haverford College
Haverford, Pennsylvania 19041

Summary

Ligand-induced phosphorylation of the receptor-regulated Smads (R-Smads) is essential in the receptor Ser/Thr kinase-mediated TGF-β signaling. The crystal structure of a phosphorylated Smad2, at 1.8 Å resolution, reveals the formation of a homotrimer mediated by the C-terminal phosphoserine (pSer) residues. The pSer binding surface on the MH2 domain, frequently targeted for inactivation in cancers, is highly conserved among the Co- and R-Smads. This finding, together with mutagenesis data, pinpoints a functional interface between Smad2 and Smad4. In addition, the pSer binding surface on the MH2 domain coincides with the surface on R-Smads that is required for docking interactions with the serine-phosphorylated receptor kinases. These observations define a bifunctional role for the MH2 domain as a pSer-X-pSer binding module in receptor Ser/Thr kinase signaling pathways.

Introduction

Members of the TGF-β superfamily of cytokines regulate cell proliferation, recognition, differentiation, apoptosis, and specification of developmental fate, in species ranging from worms to mammals (Roberts and Sporn, 1990). TGF-β signaling from the cell membrane to the nucleus is mediated by receptor Ser/Thr kinases and the Smad family of proteins (Heldin et al., 1997; Massague, 1998; Wrana and Attisano, 2000; Shi, 2001).

Smad proteins are divided into three functional classes: the comediator Smads (co-Smads), namely Smad4, which participate in signaling by diverse TGF-β family members; the receptor-regulated Smads (R-Smads), including Smad1, -2, -3, -5, and -8, each of which is involved in a ligand-specific signaling pathway; and the inhibitory Smads (I-Smads), including Smad6 and -7, which negatively regulate these pathways. The Smad proteins are highly conserved in the N-terminal MH1 domain and the C-terminal MH2 domain.

TGF-β signaling is initiated by the binding of a specific cytokine to a pair of specific transmembrane receptors, leading to the phosphorylation of the GS region in the cytoplasmic Ser/Thr kinase domains (Massague and Chen, 2000; Wrana et al., 1994). Then, the activated kinase recruits a specific R-Smad and phosphorylates its C-terminal SSXS motif (Kretzschmar et al., 1997; Macias-Silva et al., 1996). The pSer motif triggers the formation of a heterooligomer between the R-Smad and the ubiquitous Smad4, which then translocates into the nucleus and regulates expression of ligand-responsive genes.

Signaling by receptor Ser/Thr kinases (RSKs) closely resembles that by receptor tyrosine kinases (RTKs) in many aspects. In both cases, the activation of receptors involves oligomerization and subsequent phosphorylation (Schlessinger, 2000). Phosphotyrosine (pTyr) residues in RTKs serve as a molecular switch for binding by SH2-containing downstream signaling proteins, whereas pSer residues in the GS region of RSKs are involved in the recruitment of R-Smads through interactions with the MH2 domains. Although the pTyr-SH2 interactions have been extensively characterized (Kuriyan and Cowburn, 1997), it is unclear how the MH2 domain recognizes pSer and whether it functions similar to the SH2 domain in RTK signaling.

Phosphorylation occurs in the last two Ser residues of the C-terminal SSXS motif in the R-Smads (Abdollah et al., 1997; Souchelnytskyi et al., 1997). Prior to phosphorylation, the C-terminal five residues are flexible and disordered in R-Smads (Wu et al., 2000). Although phosphorylation of R-Smads is indispensable in TGF-β signaling, how phosphorylation affects the conformation and the function of R-Smads remains completely unknown. Consequently, despite clear evidence implicating Smad4 and Smad2 as tumor suppressors, it has not been possible to directly assess the effect of tumor-derived missense mutations on the phosphorylation-dependent formation of heterooligomers between Co- and R-Smads.

Progress in understanding these fundamental issues by a structural approach has been hampered by the technical difficulty in generating the R-Smads with biologically relevant phosphorylation. Using an expressed protein ligation strategy (Cotton and Muir, 1999), we report here the successful creation of a recombinant Smad2 protein, with its C-terminal Ser465 and Ser467 homogeneously phosphorylated. Although the unphosphorylated Smad2 exists as a monomer, phosphorylation at its C terminus drives efficient formation of a homotrimer both in vitro and in vivo. We crystallized a phosphorylated Smad2 homotrimer and determined its three-dimensional structure at 1.8 Å resolution. In the crystals, the phosphorylated C terminus of one monomer adopts an extended conformation and reaches out to contact the L3/B8 loop-strand pocket of the adjacent monomer, with the two pSer residues playing an anchor-

[5] Correspondence: yshi@molbio.princeton.edu

ing role. Interestingly, the four amino acids that coordinate the pSer residues are invariant among Smad4 and all R-Smads; the corresponding residues in Smad4 are also targeted for tumorigenic missense mutations. These observations identify a highly conserved pSer binding motif (the L3/B8 loop-strand pocket) in the MH2 domain. In addition, the pSer binding motif coincides with the surface required for interactions with the phosphorylated GS region of the receptor kinases. This observation implicates the MH2 domain as a general pSer binding motif in the RSK-mediated signaling pathways and suggests a mutual exclusion mechanism for the dissociation of phosphorylated R-Smads from the receptors. Furthermore, our structural analysis, together with biochemical evidence, maps a functional interface between Smad2 and Smad4, implicating a Smad2-Smad4 heterodimer. This model is supported by extensive mutational analysis and provides a plausible explanation for the uneven distribution of cancer-derived mutations in Smads.

Results

Generation and Characterization of a Phosphorylated Smad2

The full-length Smad2 protein contains 467 amino acids, with the C-terminal five residues being Cys-Ser-Ser-Met-Ser. Previous biochemical characterization demonstrated that Ser465 and Ser467 are phosphorylated following ligand activation (Abdollah et al., 1997; Souchelnytskyi et al., 1997).

To generate the phosphorylated Smad2, we employed an expressed protein ligation strategy (Cotton and Muir, 1999). A C-terminally-truncated Smad2 (residues 1–462) was overexpressed in bacteria upstream of an intein and a chitin binding domain. The amide linkage between Smad2 and the intein is in equilibrium with a thioester bond involving the intein's active site cysteine (Figure 1A). The C-terminal 5-residue phosphopeptide (Cys-Ser-pSer-Met-pSer) was chemically synthesized. After purification, the peptide was ligated to Smad2 (1–462) under native conditions (Figure 1A). The completeness of the reaction was examined by SDS-PAGE (Figure 1B), and the identities of the final products were confirmed by mass spectroscopic analysis. Using this strategy, we were able to generate large quantities of full-length Smad2, Smad3, and their MH2 domains, each phosphorylated at the biologically relevant residues (Figure 1B and data not shown).

We first examined the biochemical properties of these phosphorylated R-Smads. Prior to phosphorylation, Smad2 exists mostly as a monomer, as judged by gel filtration analysis (Figures 1C and 1D). After phosphorylation, the elution volume for the full-length Smad2 corresponds to an apparent molecular mass of 170 kDa, suggesting the formation of a homotrimer (Figure 1D). Similarly, the phosphorylated MH2 domain (residues 241–467) of Smad2 also appears to form a homotrimer (Figure 1C). Sedimentation equilibrium analysis of this fragment, as performed at five rotor speeds using an analytical ultracentrifuge, provides a molecular weight of 76,000 $\pm$ 3,900 Da. This is very close to the expected molecular weight of the homotrimer (76,557 Da).

The homotrimerization is stable; at 2 μM, greater than 95% of the phosphorylated Smad2 remains in the trimeric state (Figure 1D). These results are consistent with sedimentation equilibrium analysis performed at a loading concentration of 5 μM. Phosphorylated Smad3 (residue 1–425) also forms a homotrimer with similar affinities (data not shown). Because the MH2 domain is both necessary and sufficient for homotrimerization and the MH2 domains of Smad2 and Smad3 are 98% identical in primary sequence, the formation of heterotrimers among phosphorylated Smad2 and Smad3 should be just as efficient as the homotrimers. Indeed, when phosphorylated Smad2 is incubated with the phosphorylated Smad3-MH2 domain, heterotrimers between these two were readily detected by gel filtration (data not shown).

These observations suggest that the endogenous Smad2 and Smad3 may also form homo- as well as heterotrimers upon TGF-β-mediated activation. To investigate this possibility, we took advantage of the difference in molecular mass between Smad2 ($\sim$53 kDa) and Smad3 ($\sim$48 kDa) and examined the possible formation of heterooligomers in three different cell lines (Figure 1E). Indeed, endogenous Smad2 and Smad3 form heterooligomers upon ligand activation, in agreement with our in vitro evidence.

These results indicate that R-Smads, such as Smad2 and Smad3, undergo phosphorylation-dependent homo- and heterotrimerization both in vitro and in vivo. These results were shown previously using overexpressed proteins (Kawabata et al., 1998). Two of the cell lines used, SW480 (human colon carcinoma) and MDA-468 (human breast carcinoma), lack Smad4 (Calonge and Massague, 1999; Schutte et al., 1996), indicating that the TGF-β-induced assembly of Smad2-Smad3 complexes does not require Smad4.

Overall Structure of a Smad2 Homotrimer

To reveal the structural basis of phosphorylation-induced homotrimerization, we crystallized a Smad2-MH2 fragment (residues 241–467) with Ser465 and Ser467 homogeneously phosphorylated. The structure was determined by molecular replacement and was refined at 1.8 Å resolution (Table 1).

In the crystals, the MH2 domain of Smad2 contains a central β sandwich, with a conserved three-helix bundle (H3, H4, and H5) on one end and a conserved loop-helix region (L1, L2, L3, and H1) on the other end (Figures 2A and 3). Compared to the unphosphorylated Smad2 (Wu et al., 2000), the N-terminal extension adopts a different conformation and is more than 12 Å away, whereas the previously flexible C terminus becomes well ordered after phosphorylation (Figure 2B).

The phosphorylated Smad2 forms a symmetric homotrimer, with both the phosphorylated C terminus and the N-terminal extension from one MH2 domain reaching out to interact with an adjacent domain (Figure 2C). On one hand, the negatively charged C terminus binds a positively charged surface pocket formed by the L3 loop and the B8 strand (Figures 2C and 2D), hereafter referred to as the "loop-strand pocket". Numerous hydrogen bonds and van der Waals contacts stabilize these two complementary surfaces, with the phosphorylated Ser465 and Ser467 playing an essential role. On the other hand, the N-terminal

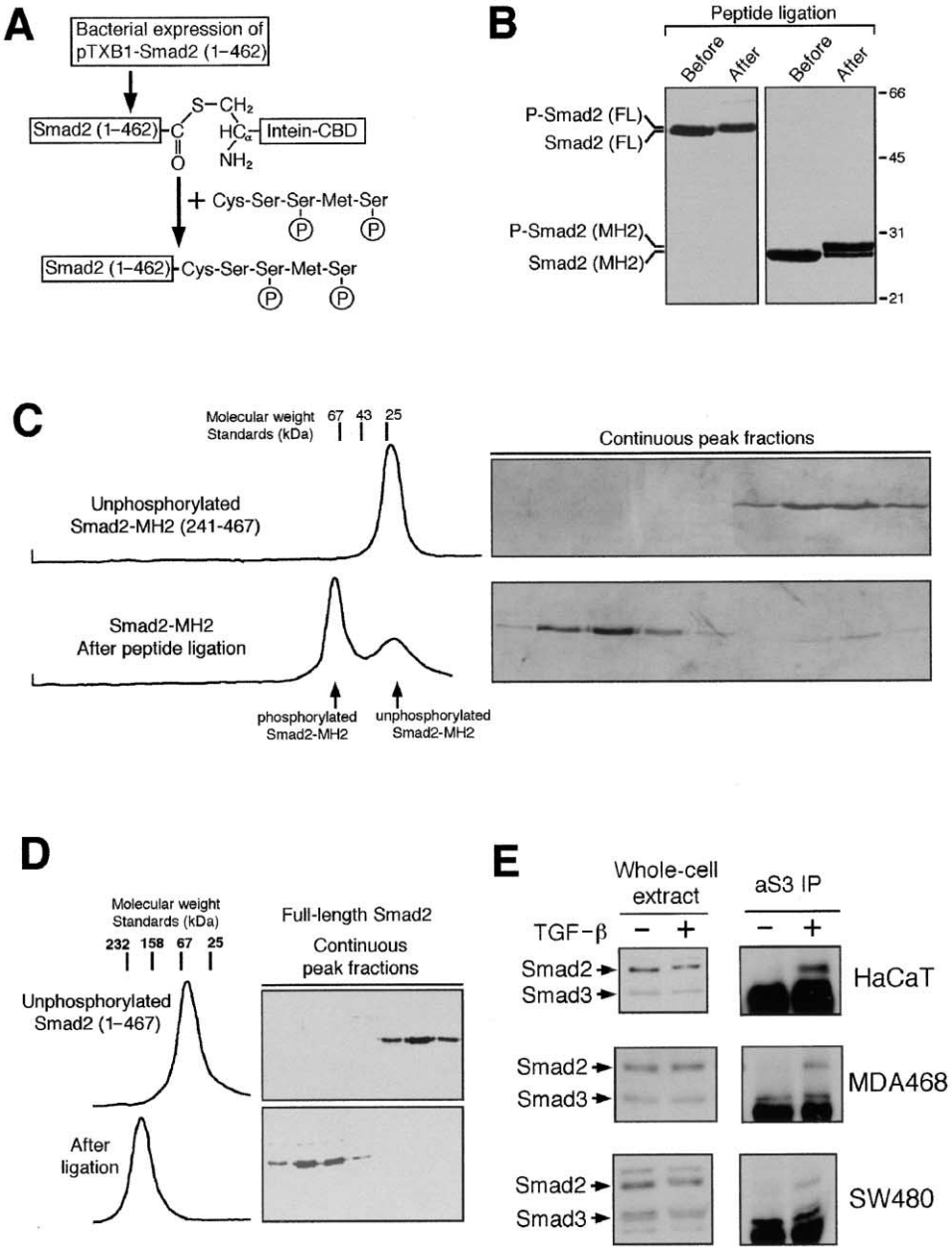

Figure 1. Phosphorylation in the C-Terminal SSXS Motif Drives Homotrimerization of R-Smads

(A) A schematic diagram of expressed protein ligation to generate phosphorylated Smad2. Recombinant Smad2 (1–462) was expressed and purified as an intein-fusion protein. The phosphorylated 5-residue peptide was chemically synthesized. Ligation was performed under native conditions to create a full-length wild-type phosphorylated Smad2. Phosphorylated Smad2-MH2 was similarly generated.

(B) SDS-PAGE showing Smad2 proteins before and after peptide ligation. The molecular weight of the ligated species is 674 Da higher than the unligated species.

(C) Gel filtration analysis of Smad2-MH2. Chromatographic traces of the gel filtration analysis are shown on the left. Before peptide ligation, Smad2-MH2 (241–462) behaves as a monomer on gel filtration (~25 kDa, upper panel). After peptide ligation, the phosphorylated Smad2-MH2 migrates as an apparent homotrimer (~74 kDa), while the unphosphorylated Smad2-MH2 remains as a monomer (lower panel).

(D) Gel filtration analysis of the full-length Smad2. Smad2 behaves as a monomer (~55 kDa) before ligation or as a homotrimer (~170 kDa) after ligation.

(E) Endogenous Smad2 and Smad3 form a heterooligomer in response to TGF-β treatment. Cell lysate was immunoprecipitated with an anti-Smad3 antibody (aS3), and the precipitated fraction was immunoblotted with an anti-Smad2-Smad3 antibody.

extension of one monomer also makes a few contacts to an acidic surface patch of the adjacent monomer, further strengthening the homotrimer (Figure 2D).

Formation of this homotrimer results in the burial of approximately 8300 Å^2 exposed surface area, with the phosphorylated C termini contributing more than 2600 Å^2 (Figure 3). In addition to the intermolecular interactions involving the N-terminal extension and the phosphory-

Table 1. Data Collection and Statistics from the Crystallographic Analysis

Data Set	Native 1 (RAXIS IV)	Native 2 (CHESS-F2)
Resolution (Å)	99.0–2.80	99.0–1.80
Total observations	66,199	122,780
Unique observations	6,169	22,718
Data coverage (outer shell)	99.5% (99.5%)	99.5% (99.2%)
R_{sym} (outer shell)	0.124 (0.440)	0.051 (0.180)
Refinement		
Resolution range (Å)	20–1.80	
Number of reflections ($I > \sigma$)	22,260	
Data coverage (outer shell)	97.7% (95.1%)	
$R_{working}/R_{free}$	21.5%/24.1%	
Number of atoms	1,757	
Number of waters	144	
Rmsd bond length (Å)	0.007	
Rmsd bond angles (°)	1.497	

$R_{sym} = \Sigma_h\Sigma_i|I_{h,i} - I_h|/\Sigma_h\Sigma_i I_{h,i}$, where I_h is the mean intensity of the i observations of symmetry-related reflections of h. $R = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$, where $F_{obs} = F_P$, and F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with 5% of the reflections). Rmsd in bond lengths and angles are the deviations from ideal values.

lated C terminus, a highly conserved protein-protein interface between two adjacent monomers appears to play an important role in the formation of the homotrimer (Figure 2C). This extensive protein-protein interface is highly similar to that observed in the homotrimeric structure of Smad4/DPC4 (Qin et al., 1999; Shi et al., 1997), with the loop-helix region of one monomer packing against the three-helix bundle of an adjacent monomer (Figure 2C). Nevertheless, Smad2 exists mostly as a monomer in the absence of phosphorylation in vivo (Jayaraman and Massague, 2000), indicating that this protein-protein interface alone is insufficient and the phosphorylation-induced conformational change provides the needed energy to drive the formation of homo- and heterooligomers.

A pSer Binding Surface on the MH2 Domain

The phosphorylated C terminus of R-Smads plays an essential role in the formation of homotrimers. In contrast to the disordered C-terminal 11 residues in the unphosphorylated Smad2 (Wu et al., 2000), phosphorylation on Ser465 and Ser467 triggers their interactions with an adjacent Smad2, stabilizing the well-defined conformation of this fragment (Figures 4A and 4B).

The intermolecular interactions between the phosphorylated C terminus and the loop-strand pocket are dominated by massive networks of hydrogen bonds, with the two phosphate groups as the nucleating centers (Figure 4C). Four residues, Lys375 on the β strand B8 and Lys420/Tyr426/Arg428 on the L3 loop, coordinate these hydrogen bonds (Figure 4C). Importantly, these four residues are invariant not only among all R-Smads but also for the Co-Smad (Smad4), suggesting critical roles in heterooligomerization as well.

On one side of the interface, Tyr426 and Lys375 donate three hydrogen bonds to the phosphate group of Ser465 (Figure 4C). These interactions are buttressed by several additional water-mediated hydrogen bonds and one intramolecular contact between the phosphate group and the backbone amide of Met466 (Figure 4C). On the other side, the phosphorylated Ser467 is coordinated by the other two invariant residues, Lys420 and Arg428. Lys420 also hydrogen bonds to the carbonyl group of Ser464 and the carboxylate of Ser467, whereas Arg428 and Arg427 make two additional hydrogen bonds to the carboxylate of Ser467 (Figure 4C). It is also noteworthy that a number of well-ordered water molecules bridge hydrogen bonds between the phosphorylated C terminus and the loop-strand pocket (Figure 4C).

The conformation of the C terminus is also significantly restrained by intramolecular contacts. Notably, the side chain of Ser464 accepts a hydrogen bond from the backbone amide of Ser467 while making two additional contacts to its phosphate (Figure 4C), consistent with its important role in Smad function (Abdollah et al., 1997; Kretzschmar et al., 1997). Intriguingly, in the conserved SSXS motifs in R-Smads, this Ser residue is not phosphorylated upon ligand activation. Our structural analysis reveals that phosphorylation of this Ser may disrupt binding by the phosphorylated C terminus, likely impeding TGF-β signaling.

In addition to hydrogen bonds, the extended C terminus also makes extensive van der Waals contacts, strengthening the binding of the phosphorylated C terminus and hence the formation of homotrimers (Figure 4D). For example, Cys463 is completely buried in a hydrophobic pocket formed by five residues (Met327, Thr328, His331, Val419, and Lys420) in the neighboring molecule (Figure 4D).

Protein-Protein Interface

In addition to the interface involving the phosphorylated C terminus, an extensive protein-protein interface between two adjacent monomers contributes significantly to the formation of a homotrimer (Figures 2 and 3). This interface is highly homologous to, but larger than, that observed in the homotrimeric structure of Smad4 (Qin et al., 1999; Shi et al., 1997).

Both hydrogen bonds and van der Waals contacts play important roles in this interface (Figure 5). A total of 15 inter- and 11 intramolecular hydrogen bonds constitute two networks. At the center of the interface, Arg310 makes a bifurcated hydrogen bond to Asp450 on helix H5 of the adjacent monomer. These contacts

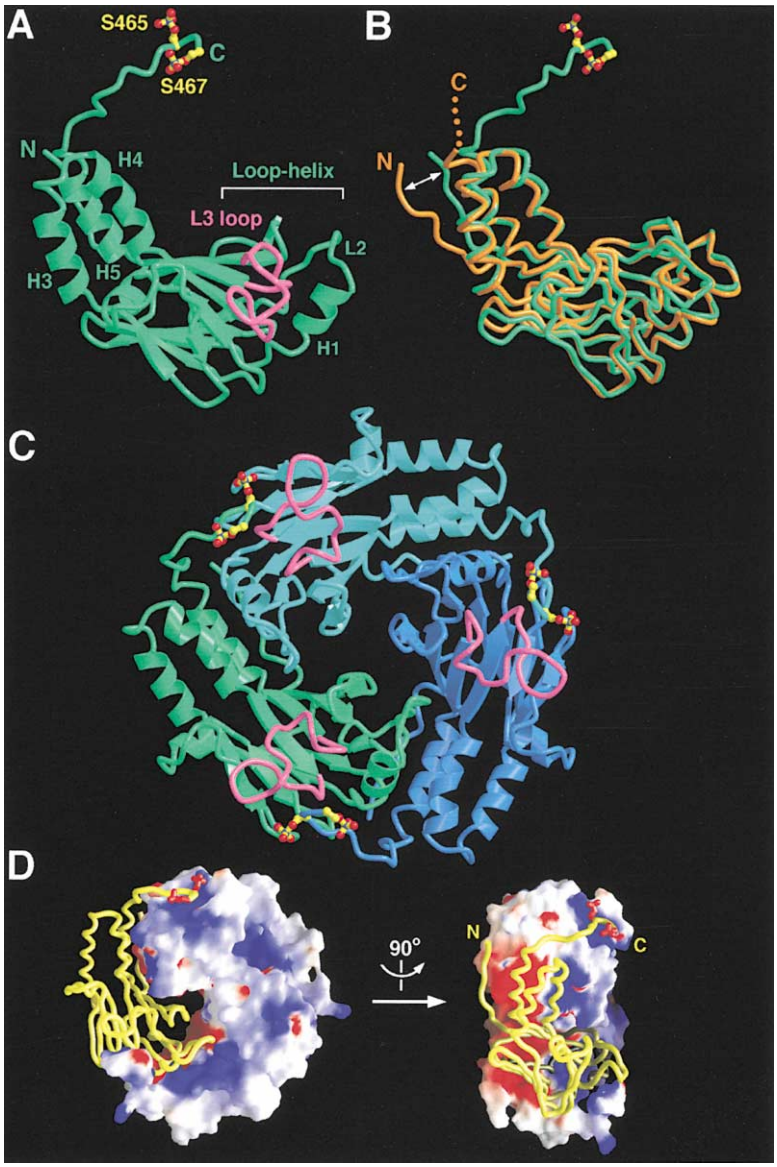

Figure 2. Overall Structure of the Phosphorylated Smad2

(A) A schematic representation of the phosphorylated Smad2-MH2 (green). The two pSer residues (S465 and S467) are highlighted, while the L3 loop is shown in pink.

(B) Structural comparison of the phosphorylated (green) and unphosphorylated (gold) Smad2-MH2 (PDB code 1DEV). The differences are apparent in two locations; the N-terminal extensions are more than 12 Å apart, and the flexible C terminus in the unphosphorylated Smad2 becomes rigid and ordered upon phosphorylation.

(C) An overall view of the trimeric structure of the phosphorylated Smad2-MH2. The three monomers are shown in green, cyan, and blue.

(D) A surface potential representation of the Smad2 homotrimer, with one monomer in yellow coil. The side chains of the two phosphorylated Ser residues are shown in red. Positively and negatively charged surfaces are colored blue and red, respectively. (A)–(C) and all other figures, except Figure 4B, were prepared using MOLSCRIPT (Klaulis, 1991). (D) and Figure 4B were prepared using GRASP (Nicholls et al., 1991).

are stabilized by two intramolecular contacts between Asp300 and Arg310 on one side and three additional interactions between Tyr268 and Asp450 and between Asp450 and the backbone amide groups of Thr303/Asp304 on the other side (Figure 5A). Asp300 further strengthens this network by accepting two intramolecular contacts from the backbone groups of Gly301/Phe302. At the periphery of this protein-protein interface, Asp304/Ser306 on the L1 loop form a second network of hydrogen bonds with Ser269/Gln447 of the adjacent monomer (Figure 5A). A high density of van der Waals interactions also help to stabilize this interface (Figure 5B). For example, Val319 on the L2 loop of one monomer makes multiple close contacts to four residues (Tyr280, Val431, Thr432, and Trp437) on the adjacent monomer.

A Functional Interface between Smad2 and Smad4

Several lines of evidence converge to identify a functional interface, involving the helix-bundle region of the phosphorylated Smad2 and the loop-helix region and the loop-strand pocket of Smad4 (Figure 5C). In the Smad2 homotrimer, the pSer-containing C terminus interacts with the loop-strand (L3-B8) region of the adjacent monomer (Figures 2 and 4). All important residues in the loop-strand pocket are invariant in Smad4, including the four key residues that coordinate the phosphate groups (Figure 3). This observation indicates that the phosphorylated C terminus of Smad2 can interact with Smad4 in an identical manner. Additional support comes from a comparison of the conserved protein-protein interfaces in the homotrimers of Smad2 and Smad4. On the basis of the proposed Smad2-Smad4 interface (Figure 5C), all Smad4 residues involved in hydrogen bond interactions are identical to the corresponding residues in the Smad2 homotrimer, including all key residues at the centers of the two hydrogen bond networks (Figures 2 and 5D).

Previous biochemical studies suggested the formation of a heterodimer between these proteins using the unphosphorylated Smad2 and Smad4 (Wu et al., 2001).

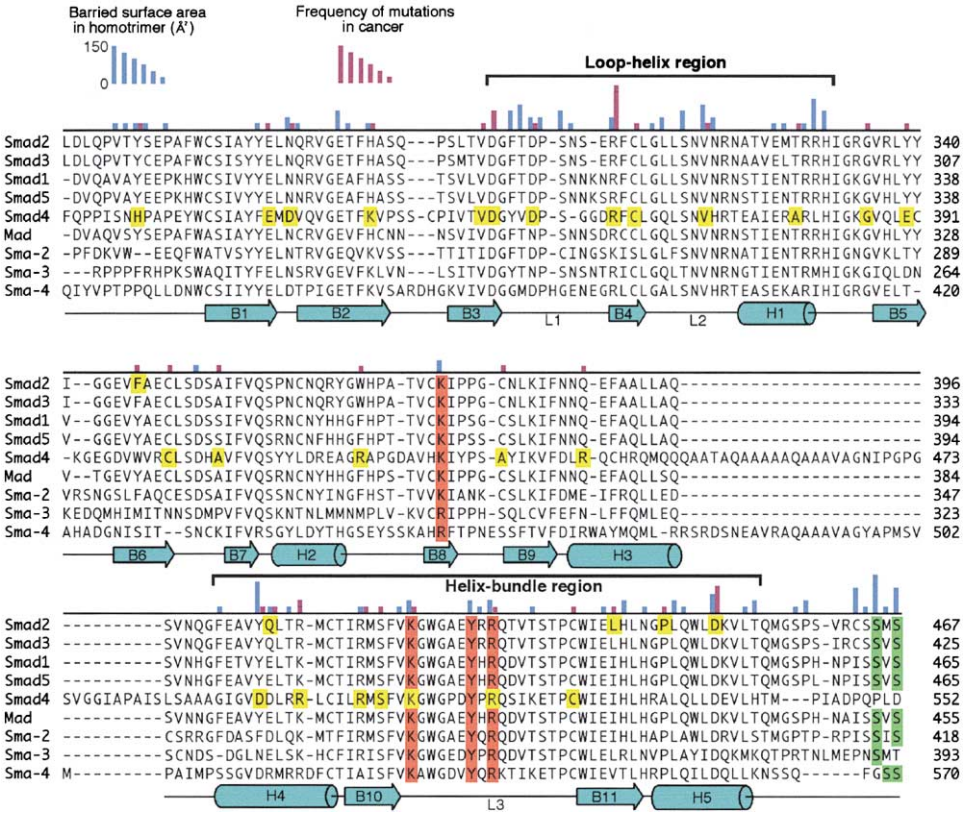

Figure 3. Sequence Alignment of the Smad Proteins and Their Homologs in *Drosophila* and *C. elegans*

The secondary structural elements are indicated below the alignment. The four residues that directly coordinate the phosphorylated Ser residues are highlighted in red, while the candidate residues for phosphorylation in the C termini are shown in green. By both criteria of barried surface area and frequency of tumorigenic mutations, the loop-helix region in Smad4 and the helix-bundle region in Smad2 are implicated in the formation of a heterooligomer. Residues targeted for missense mutation in cancer are shown in yellow.

Using the phosphorylated Smad2, we reexamined the nature of the Smad2-Smad4 complex. The elution volume of the isolated Smad4-MH2 (residues 251–552) corresponds to a molecular mass of 33 kDa, consistent with a monomer (Figure 5E). To ensure efficient formation of a heterocomplex, peptide ligation was performed in the presence of an equimolar amount of Smad4-MH2. After completion of the reaction, the complex was analyzed by gel filtration, which revealed a 1:1 heterocomplex between the phosphorylated Smad2-MH2 and Smad4-MH2 (Figure 5E). This suggests either a heterodimer or an equal mixture of 2:1 and 1:2 heterotrimers between Smad2 and Smad4. The migration of a Smad heterodimer on gel filtration is expected to be very similar to that of a trimer because they share an indistinguishable hydrodynamic radius.

Due to high binding affinities for a Smad2 homotrimer, it is unclear whether and how Smad4 can gain an edge to form a heterocomplex. To examine this scenario, we incubated Smad4-MH2 with the phosphorylated homotrimers of Smad2 and subjected the mixture to gel filtration analysis. Smad4-MH2 was able to disrupt the homotrimers of Smad2-MH2 in a time-dependent manner (Figure 5F). The observed kinetics suggest a slow off-rate for the dissociation of Smad2 homotrimers. To accelerate the formation of a Smad2-Smad4 heterocomplex, we introduced a missense mutation in Smad2, Lys375Ala, which is predicted to selectively weaken the formation of a Smad2 homotrimer. In this case, Smad4 readily formed a 1:1 heterocomplex with the phosphorylated Smad2 (Lys375Ala) (Figure 5F).

Mutational Analysis

To further investigate heterocomplex formation, we generated 11 and 17 missense mutations in the MH2 domains of Smad4 and Smad2, respectively (Table 2). Each of the Smad2 mutants contains phosphorylated Ser465 and Ser467. If Smad4 and the phosphorylated Smad2 formed a 2:1 or 1:2 heterotrimer using the conserved protein-protein interface, then both the loop-helix and the helix-bundle regions for both proteins would be involved in the heterotrimeric assembly. In this case, mutations affecting critical interface residues in either Smad2 or Smad4 would be expected to have negative effects. On the other hand, only mutations in the loop-helix region of Smad4 or the helix-bundle region of Smad2 would be expected to weaken a phosphorylation-dependent Smad2-Smad4 heterodimer (Figure 5C).

Two mutations affecting residues in the loop-helix region of Smad4, D351H and R361H, completely abolished the formation of the Smad2-Smad4 complex, while D537H in the helix-bundle region of Smad4 had no detectable effect (Table 2). In contrast, the corresponding mutations in Smad2, D300H and R310H, had no effect,

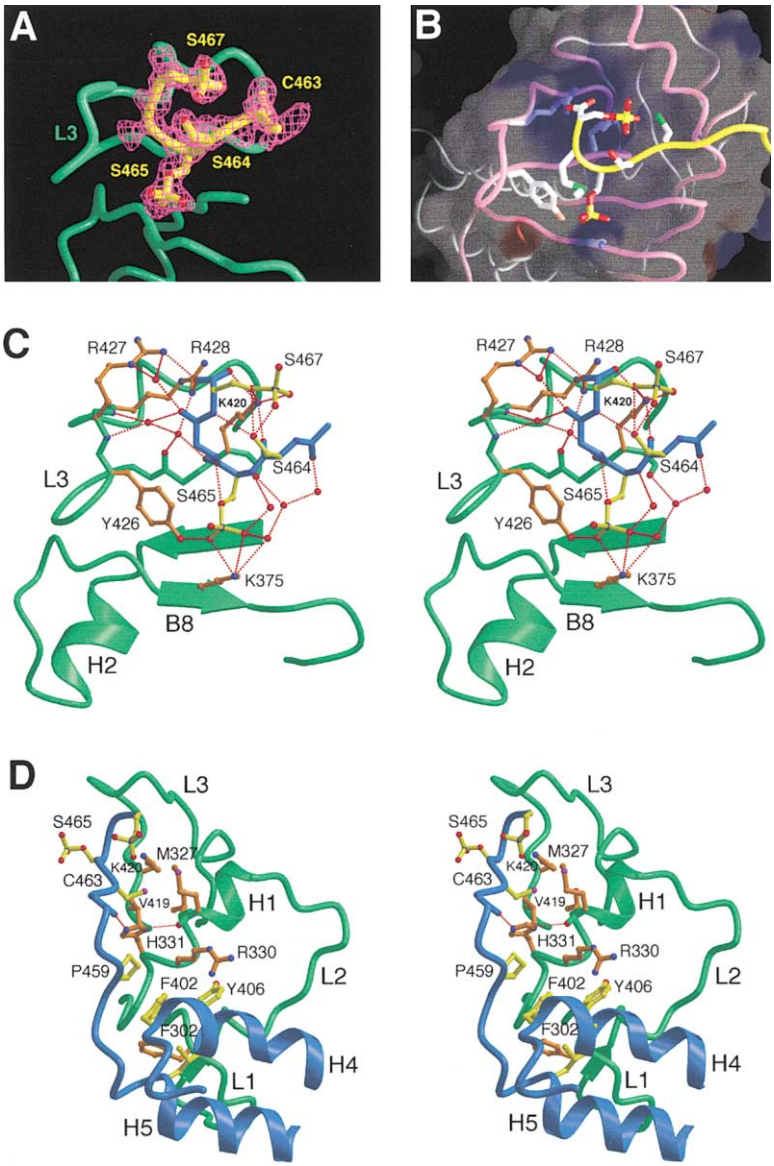

Figure 4. A Close-Up View of the Interactions between the Phosphorylated C Terminus from One Monomer and the Loop-Strand Pocket of the Adjacent Monomer

(A) An electron density map of the phosphorylated C terminus. The $2F_o$-F_c map (omit map), shown in pink, was contoured at 1.5σ and was calculated by simulated annealing using CNS (Brunger et al., 1998) with the omission of the C-terminal five residues. The backbone as well as the side chains of four residues are shown in yellow.

(B) An overall view of the interactions. The C terminus is shown as a yellow coil, while its binding partner is represented as a transparent surface with backbones in pink. The side chains of the last five residues in the C terminus (CSSMS) and the basic residues in the loop-strand pocket are shown.

(C) A stereo view of hydrogen bond networks. The two interacting monomers are shown in green and blue, respectively. Their side chains are colored gold and yellow. Hydrogen bonds among oxygen (red) and nitrogen (blue) atoms and water molecules (red) are indicated by red dashed lines.

(D) A stereo view of the van der Waals contacts between the phosphorylated C terminus from one monomer and the loop-strand pocket of the adjacent monomer. The coloring scheme is the same as in (C).

whereas D450H abrogated the formation of the Smad2-Smad4 complex (Table 2). These data support the notion that the phosphorylated Smad2 forms a heterodimer with Smad4 using the identified interface (Figure 5C).

Due to high affinity for homotrimerization, each of the single missense mutations in Smad2 retained a weakened ability to form a homotrimer (Table 2). These remaining tendencies to form a homotrimer are completely eliminated by the introduction of a second missense mutation (Table 2). None of these six double mutations affects the proposed interface between Smad2 and Smad4 (Figure 5C). Thus, despite failure to form a homotrimer, all six Smad2 double mutants retained complex formation with Smad4.

We also investigated the Smad2-Smad4 complex by sedimentation equilibrium analysis. The wild-type phospho-Smad2-Smad4 complex exists in equilibrium with the Smad2 homotrimers. However, double mutation (K375A, R310H) in phospho-Smad2 eliminated homotrimer formation (Table 2). Thus, we focused on the heterocomplex between this Smad2 variant and Smad4 (251–552). Data were collected at several concentrations and four speeds to assess the stoichiometry and stability of this complex. At a loading concentration of 17 μM for the complex, a molecular weight of 58,800 $\pm$ 4,500 Da is obtained, which is consistent with the formation of a stable heterodimer. The residuals from this single species analysis were random, indicating that there is no strong evidence for the presence of lower or higher molecular weight species. This heterodimer is stable over loading concentrations in the range of 2.5–17 μM. Fits to heterotrimeric models of 1:2 and 2:1 stoichiometries were significantly worse.

Although our gel filtration data cannot rule out the formation of an equal mixture of the 1:2 and 2:1 complexes between Smad2 and Smad4, sedimentation equilibrium data described above are inconsistent with this dynamic equilibrium model and provide additional support for a heterodimer model. Given that a proposed heterotrimer would involve nearly three times as much

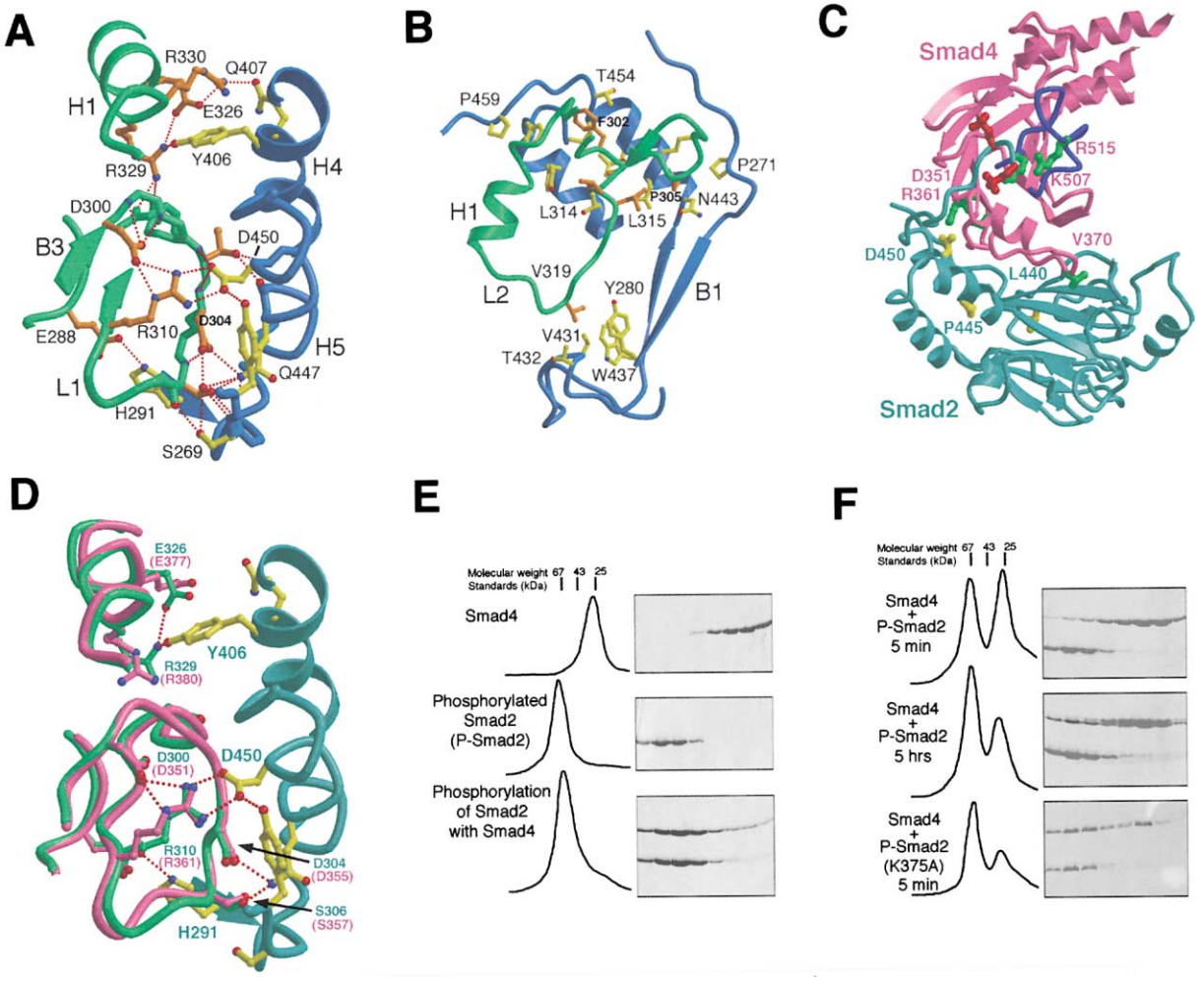

Figure 5. Identification of a Functional Smad2-Smad4 Hetero-Interface

(A) A close-up view of hydrogen bond networks at the protein-protein interface between two adjacent Smad2 molecules. The coloring scheme is the same as in Figure 4C.

(B) A close-up view of the van der Waals contacts between the loop-helix region of one monomer and the helix-bundle region of the adjacent monomer.

(C) A proposed interface between Smad2 (cyan) and Smad4 (pink). In this model, the phosphorylated C terminus of Smad2 interacts with the highly conserved loop-strand pocket of Smad4. Phosphorylated Ser465 and Ser467 are colored red. Residues targeted by tumor-derived missense mutations are shown in green and yellow for Smad4 and Smad2, respectively.

(D) A close-up view on the proposed and conserved Smad2-Smad4 interface. Two adjacent Smad2 molecules are colored cyan and green, respectively, while Smad4 is shown in pink. The side chains from Smad4 and the overlaying Smad2 are shown in pink and green, respectively. The side chains from the interacting Smad2 are colored yellow. These hydrogen bonds are invariantly preserved in the proposed hetero-interface of Sma2-Smad4.

(E) Phosphorylated Smad2 forms a 1:1 complex with Smad4. Chromatographs of the gel filtration analysis are shown on the left. Isolated Smad4-MH2 and free phosphorylated Smad2-MH2 behave as a monomer and a homotrimer, respectively (upper and middle panels). (F) Slow exchange between a preformed Smad2 homotrimer and a Smad2-Smad4 heterocomplex. In the upper and middle panels, excess Smad4 was incubated with the phosphorylated Smad2 (already trimers) for 5 min and 5 hr, respectively. Then the mixture was subjected to gel filtration analysis. Disruption of the trimeric Smad2 in favor of a 1:1 Smad2-Smad4 complex is slow but can be facilitated by introducing a point mutation into Smad2 (K375A) that selectively weakens the formation of the Smad2 homotrimers.

interface area as a heterodimer, it would be thermodynamically unfavorable to maintain a heterodimer if the heterotrimer model were correct. Yet, the principal species for a heterocomplex between Smad4 and Smad2 (K375A, R310H) is a heterodimer by sedimentation equilibrium analysis. These results strongly support the heterodimer model. Nevertheless, our data do not completely rule out the possibility that both heterodimers and heterotrimers can form.

One recent study suggested the existence of a heterotrimer between Smad4 and a pseudophosphorylated Smad3 using Glu to replace pSer (Chacko et al., 2001). Using the phosphorylated R-Smads, we are unable to verify this claim. Interestingly, although a heterocomplex between Smad4 and the phosphorylated Smad2 was interpreted to be a heterotrimer in one previous report, the molecular mass of this complex was appreciably smaller than that of a Smad2 homotrimer (Kawabata et

Table 2. Mutational Analysis of a Smad2-Smad4 Complex

Mutation in Phosphorylated Smad2	Formation of Homotrimer	Interaction with Smad4	Corresponding Mutation in Smad4	Interaction with Phosphorylated Smad2
Native	++	++	Native	++
Mutation in the Loop-Strand Pocket				
K375A	+/−	++	K428A	+/−
K420A	+/−	++	**K507A**	+/−
Y426A	+/−	++	Y513A	++
R428A	+/−	++	**R515A**	+/−
Double Mutation in the Loop-Strand Pocket and the Loop-Helix Region				
D300H/R310H	−−	++		
D300H/K375A	−−	++		
D300H/K420A	−−	++		
R310H/K375A	−−	++		
R310H/K420A	−−	++		
K375A/K420A	−−	++		
Mutation in the Loop-Helix Region				
D300H	+/−	++	**D351H**	−−
R310H	+/−	++	**R361H**	−−
D450H	+/−	−−	D537H	++
V319D	++	++	**V370D**	+−
W368H	++	++	**R420H**	++
N387P	++	++	**R441P**	++
Y406H	++	++	**D493H**	++

The identities of all wild-type and mutant proteins were confirmed by double-stranded plasmid sequencing and mass spectroscopic analysis. All Smad2 proteins used have Ser465 and Ser467 homogeneously phosphorylated. The original tumor-derived missense mutations in Smad4 and Smad2 are shown in bold face. Corresponding mutations in Smad2 and in Smad4 are listed in the same line. Formation of homo- and heterocomplexes was examined by gel filtration.

al., 1998). Thus, this observation is actually consistent with the presence of a heterodimer between Smad4 and the phosphorylated Smad2.

Tumor-Derived Mutations

Smad4, also known as deleted in pancreatic carcinoma locus 4 (Hahn et al., 1996), is inactivated in one half of pancreatic carcinomas (Schutte et al., 1996). Smad2 has also been identified as a tumor suppressor because of its mutations in colorectal and lung cancers (Eppert et al., 1996). The majority (40 out of 52) of tumor-derived missense mutations reported to date map to the MH2 domain (Figure 3). In particular, two residues in the loop-helix region of Smad4 (Asp351 and Arg361) are targeted with recurrent mutations in cancer, whereas missense mutations primarily affect residues in the helix-bundle region of Smad2 such as Asp450 (Figure 3). These uneven distributions coincide with the proposed Smad2-Smad4 interface involving the helix-bundle region of Smad2 and the loop-helix region of Smad4 (Figure 5C). Interestingly, Asp450 from Smad2 and Asp351 and Arg361 from Smad4 are proposed to form a network of hydrogen bonds in the heterocomplex (Figure 5D). Compromise of these interactions leads to abrogation of a heterocomplex between Smad2 and Smad4 (Table 2).

Another interesting example is the tumor-derived mutations of Lys507Gln and Arg515Gly in Smad4 (Miyaki et al., 1999). Lys507 and Arg515 in Smad4 correspond to Lys420 and Arg428 in Smad2, respectively, two essential residues coordinating pSer (Figure 4C). Thus, these two mutations in Smad4 are expected to negatively affect the formation of a heterocomplex. Supporting this analysis, mutation of these two residues to Ala were found to weaken the formation of a heterodimer between Smad2 and Smad4 (Table 2).

Discussion

Dissociation of the Phosphorylated Smad2 from Receptor Kinases

The phosphorylation of Smad2 is facilitated by physical interactions between the positively charged L3 loop region in Smad2 and the L45 loop and the GS region of the receptor (Chen et al., 1998b; Feng and Derynck, 1997; Huse et al., 2001; Lo et al., 1998; Wu et al., 2000) (Figure 6A). Ser residues in the GS region of the type I receptor are phosphorylated by the type II receptor kinase upon ligand activation (Wieser et al., 1995; Willis et al., 1996; Wrana et al., 1994), which increases its binding affinity for the basic surface in the L3 loop region (Huse et al., 2001). Interestingly, this L3 loop region coincides with the loop-strand pocket that interacts with the pSer-containing C terminus in the homotrimer (Figure 2). Thus, competition for binding to this surface between the pSer motif in the receptor kinase and the pSer-containing C terminus in Smad2 likely accounts for the dissociation of phosphorylated Smad2 from the receptor kinases (Figure 6A).

This model is strongly supported by mutational analysis on the SSXS motif. Our model predicts that loss of phosphorylation on Ser465 and/or Ser467 may lead to prolonged association between Smad2 and the receptor

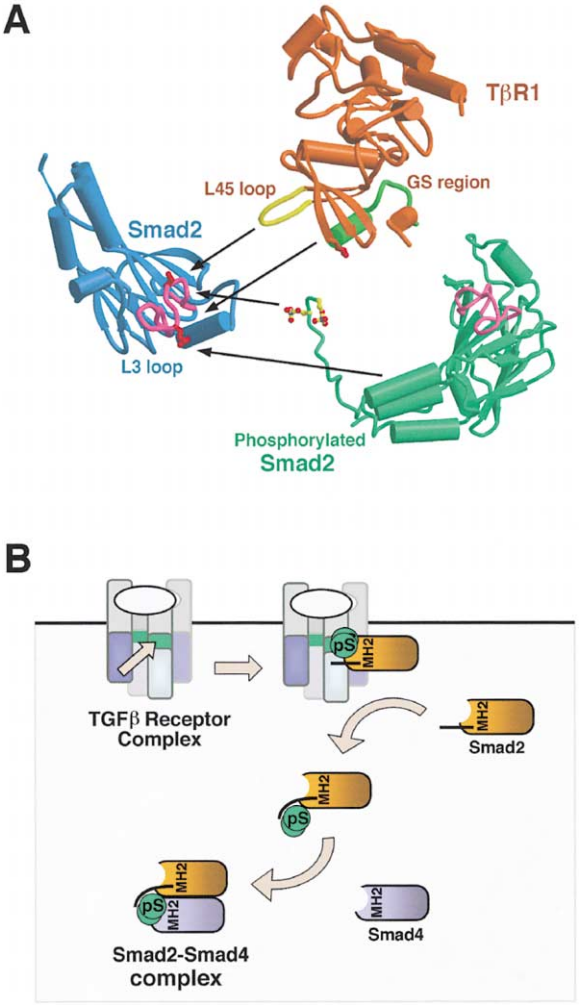

Figure 6. Implications for RSK-Mediated Signaling

(A) Proposed mechanisms of Smad2 dissociation from the receptor kinase (TβRI) (Huse et al., 1999) after phosphorylation. The positively charged loop-strand pocket on Smad2, which is responsible for binding the phosphorylated C terminus of another Smad2, coincides with the L3 loop region, which is involved in interactions with the L45 loop and the GS region of the receptor kinase. The mutual exclusion is proposed to lead to dissociation of phosphorylated Smad2 from the receptors.

(B) A schematic diagram of signal flow in the RSK-mediated signaling, highlighting the MH2 domain as the pSer binding motif.

kinases. In agreement, mutation of either Ser465 or Ser467 to Ala in Smad2 prevents its dissociation from the activated TβRI due to a stable interaction between the mutant Smad2 and the TGF-β receptor complex (Abdollah et al., 1997; Souchelnytskyi et al., 1997). This in turn leads to abrogation of complex formation with Smad4 and inhibition of TGF-β signaling (Abdollah et al., 1997; Souchelnytskyi et al., 1997). Moreover, a missense mutation in Smad2, Tyr426Ala, led to loss of interactions with the activated TGF-β receptors, demonstrating a direct role in receptor binding (Lo et al., 1998). This observation also suggests that Tyr426 may directly bind pSer in the GS region of the TGF-β receptor, as is the case in the Smad2 homotrimer (Figure 4).

The phosphorylated C terminus may come from another Smad2 molecule (Figure 6A) or from the same molecule that remains bound to the receptor. However, in the latter case, the C terminus will adopt a highly restrained conformation and will not be able to interact with the loop-strand pocket in the correct configuration. In fact, as the receptor complex contains two molecules of TβRI, which likely binds and phosphorylates two Smad2 molecules simultaneously, the phosphorylated Smad2 molecules might help each other to dissociate from the receptor complex.

Dissociation of Phosphorylated Smad2 from SARA

The cytosolic protein SARA specifically interacts with and facilitates the phosphorylation of Smad2 and Smad3 (Tsukazaki et al., 1998). Phosphorylation of Smad2, however, induces its dissociation from SARA with concomitant formation of a heterocomplex between Smad2 and Smad4 (Tsukazaki et al., 1998; Xu et al., 2000). Our structural analysis provides a plausible explanation for this observation. In the Smad2-SARA complex, 45% of the interface involves the C-terminal β strand of the bound SARA fragment, which forms a β sheet with the N-terminal B1′ strand of Smad2 (Wu et al., 2000). In the structure of the phosphorylated Smad2, however, the N-terminal extension moves more than 12 Å away from the position of the B1′ strand, thus destabilizing the SARA-Smad2 interface. Supporting this explanation, missense mutations in the C-terminal β strand of SARA led to complete abrogation of complex formation between SARA and Smad2 (Wu et al., 2000). In the cell, as SARA also interacts with the receptor complex, phosphorylation-driven dissociation of Smad2 from SARA may be one of the key events in releasing activated Smad2 from the receptor complex.

Since the phosphorylated R-Smads form stable homotrimers in vivo, how is the formation of a heterocomplex between Smad4 and Smad2 favored over that of a Smad2 homotrimer? As the phosphorylated Smad2 first appears upon receptor stimulation, it is exposed to a large pool of unphosphorylated Smad2, which is not effective at forming homotrimers, and Smad4, which is poised to form a stable heterocomplex with the phosphorylated Smad2. Thus, this dynamic equilibrium is likely to favor the formation of a Smad2-Smad4 heterocomplex over a Smad2 homotrimer.

MH2 Domain as the pSer Binding Motif

Our study identifies the MH2 domain of R-Smads and Co-Smad as the phosphoserine binding motif in the RSK-mediated TGF-β signaling. In response to ligand binding, the type II receptor kinase forms a complex with the type I receptor and phosphorylates its GS region. The pSer-X-pSer motifs in the GS region then serve as a molecular platform to recruit R-Smad through interactions with its MH2 domain and to facilitate its phosphorylation (Figure 6B). Once phosphorylated, the pSer-X-pSer motif at the C termini of R-Smad likely induces its dissociation from the receptor complex and mediates the formation of a heterocomplex through interactions with the conserved loop-strand region in the MH2 domain of Smad4 (Figure 6). Thus the MH2 domain defines a signaling motif as the pSer-X-pSer binding module, similar to the SH2 domain in the RTK-mediated signaling pathways.

In both pTyr-SH2 (Kuriyan and Cowburn, 1997) and pSer-MH2 complexes, the phosphate groups nucleate the interactions, with the negative charges of the phosphate groups neutralized by highly conserved Arg/Lys residues through hydrogen bonds (Figure 4). In response to RTK stimulation, effector proteins can be activated through at least two mechanisms, phosphorylation and conformational change, both of which depend on the pTyr-SH2 interactions (Schlessinger, 2000). These paradigms are precisely the same for the RSK-mediated TGF-β signaling. First, phosphorylation of the C-terminal Ser-X-Ser motif, which is required to activate the R-Smads, depends on the interactions between the MH2 domain and the pSer motif in the GS region of the receptor. Second, conformational changes triggered by phosphorylation drive the formation of a functional Smad heterocomplex, again depending on the pSer-MH2 interactions.

Phosphorylation is known to modulate protein-protein interactions in many ways. The regulation of the pTyr-SH2 and pSer-MH2 interactions exhibits similar strategies. For example, the intramolecular interaction between the phosphorylated Tyr527 and the SH2 domain leaves c-Src in an inactive closed state (Sicheri et al., 1997; Xu et al., 1997). However, this pTyr-SH2 interaction is structurally nonideal and relatively transient. The pTyr motif of RTKs can efficiently compete for binding to this SH2 domain, thus releasing the inhibited closed conformation into an open form (Sicheri and Kuriyan, 1997). In the pSer-MH2 interactions, the relatively transient MH2 binding from the pSer-X-pSer motif (GS region) of the receptor kinase is likely replaced by the strong binding from the phosphorylated C-terminal pSer-X-pSer motif, resulting in the dissociation of the R-Smad and complex formation with the Co-Smad. Another interesting example is the SH2-containing STAT proteins, which are recruited to the receptors by binding pTyr residues and are subsequently phosphorylated on their own Tyr residues by the Jak or receptor kinases (Darnell, 1997). The phosphorylated STATs dissociate from the receptors, form SH2-pTyr-mediated dimers, and are translocated into the nucleus where they bind specific DNA sequences and regulate gene transcription (Becker et al., 1998; Chen et al., 1998a; Darnell, 1997). Strikingly, following phosphorylation-driven heterooligomerization, Smads also function as direct DNA binding transcriptional regulators in the nucleus.

Although the pSer-MH2 interaction shares much similarity with the pTyr-SH2 binding interaction, the modular MH2 domain exhibits a more diverse array of functions. The MH2 domain is the effector domain in Smads, whereas the SH2 domain is a regulatory module for the effector proteins. Consequently, the RTK-SH2 interaction is fairly stable, whereas the RSK-MH2 binding interaction is relatively transient. Nonetheless, our structural analysis extends the molecular mimicry between the RTK- and the RSK-mediated signaling pathways to an atomic detail.

Experimental Procedures

Generation of Phosphorylated R-Smads In Vitro

A C-terminally-truncated Smad2 (1–462) was cloned into the pTXB1 bacterial expression vector (New England Biolab) upstream of the GyrA intein and a chitin binding domain. This construct was overexpressed in *E. coli* for 12 hr at room temperature. Cells were harvested and resuspended in buffer A containing 50 mM HEPES (pH 7.5) and 100 mM NaCl. After cell lysis, the soluble fraction was applied to chitin resin preequilibrated with buffer A. The amide linkage between Smad2 and the GyrA intein is in equilibrium with a thioester bond involving the intein's active site cysteine. The Smad2 moiety on the resin was cleaved by buffer A supplemented with 50 mM MESNA (2-mercaptoethanesulfonic acid, SIGMA), yielding a Smad2 molecule with the C-terminal thioester moiety required for chemical ligation. The cleaved Smad2 was further purified by anion exchange chromatography (Source-15Q, Pharmacia). The ligation reaction was performed at 4°C overnight by incubating the concentrated Smad2 with 1 mM of the chemically synthesized peptide (Cys-Ser-pSer-Met-pSer). Similar strategy was applied to generate the phosphorylated wild-type and mutant Smad2-MH2 (241–462), Smad3-MH2 (200–425), and the full-length Smad3 (1–425). The identities of the phosphorylated proteins were confirmed by mass spectroscopic analysis.

Site-Directed Mutagenesis and Protein Preparation

Point mutations were generated by standard PCR-based strategy, and the identities of individual clones were verified by sequencing. Recombinant wild-type and mutant Smad4-MH2 (251–552) were overexpressed as GST-fusion proteins using pGEX-2T (Pharmacia). The unphosphorylated wild-type and mutant Smad2-MH2 (241–462) was overexpressed in *Escherichia coli* strain BL21(DE3) using a pET-3d vector (Novagen). Protein purification was performed as described (Chai et al., 2001).

Gel Filtration Analysis

Size exclusion chromatography, using a Superdex-200 column (10/30, Pharmacia), was employed to examine the homo- as well as heterooligomerization for Smad4 and Smad2. The flow rate was 0.5 ml/min, and the buffer contained 25 mM Tris (pH 8.0), 150 mM NaCl, and 2 mM DTT. All fractions were collected at 0.5 ml each. Aliquots of relevant fractions were subjected to SDS-PAGE and visualized by Coomassie staining.

Immunoprecipitation and Immunoblotting

For protein immunoprecipitation, HaCaT cells were lysed in TNE buffer (10 mM Tris-HCl [pH 7.8], 150 mM NaCl, 1 mM EDTA, and 1% NP40) supplemented with protease and phosphatase inhibitors. Extracts were incubated with a polyclonal antibody against Smad3 (aS3, Zymed). Protein complexes were collected using protein A-sepharose and were detected by Western blotting using a polyclonal antibody against a conserved linker region between Smad2 and Smad3.

Crystallization and Data Collection

Crystals were grown by the hanging-drop vapor diffusion method by mixing the phosphorylated Smad2 (residues 241–467) (15 mg/ml) with an equal volume of reservoir solution containing 100 mM HEPES (pH 7.5), 10% 1,4-dioxane (v/v), and 750 mM K_2HPO_4/KH_2PO_4 (pH 7.0). Small crystals appeared after 1–4 days and were used for macroseeding to generate larger crystals. The crystals, with a typical dimension of $0.3 \times 0.3 \times 0.3$ mm^3, belong to the spacegroup I2$_1$3 with $a = b = c = 113.7$ Å. An initial diffraction data set at 2.8 Å resolution was collected using an R-AXIS-IV imaging plate detector. The final high-resolution native data set at 1.8 Å resolution was collected at CHESS beamline F2. For data collection, crystals were equilibrated in a cryoprotectant buffer containing reservoir buffer plus 20% glycerol (v/v) and were flash frozen in a cold nitrogen stream at –170°C. Both diffraction data sets were processed using Denzo and Scalepack (Otwinowski and Minor, 1997).

Structure Determination

The structure of the phosphorylated Smad2-MH2 domain was determined by molecular replacement, using AMORE (Navaza, 1994). The coordinates of Smad2 (PDB code 1DEV) were used for rotational research against all reflections between 15 and 3.0 Å in the 2.8 Å data set. The top 50 solutions from the rotational search were individually used for a subsequent translational search, which yielded a single promising solution. This model was examined with O (Jones et al.,

1991). Refinement by CNS (Brunger et al., 1998), against the 2.8 Å native data set, quickly decreased the R factor and allowed visualization of the phosphorylated C terminus. A model was built with O (Jones et al., 1991) and refined further against the 1.8 Å data set by simulated annealing. The final atomic model contains Smad2 residues 265–467 with Ser465 and Ser467 phosphorylated, and 144 ordered water molecules. The N-terminal 24 residues are disordered in the crystals.

Analytical Ultracentrifugation

Protein samples were prepared in 10 mM HEPES (pH 8.0), 150 mM NaCl, 3 mM DTT. All sedimentation equilibrium experiments were carried out at 4°C using a Beckman Optima XL-A analytical ultracentrifuge equipped with an An60 Ti rotor and using six-channel, 12-mm path length, charcoal-filled Epon centerpieces and quartz windows. Loading concentrations for the phosphorylated Smad2 MH2 were 78, 50, 25, 10, and 5 μM. Loading concentrations for the phospho-Smad2-Smad4 complex were 17, 10, 5, and 2.5 μM. Data were collected at four rotor speeds (8,000, 11,000, 14,000, and 17,000 rpm) and represent the average of 20 scans using a scan step size of 0.001 cm. Partial specific volumes and solution density were calculated using Sednterp (Laue et al., 1992). Data were analyzed using HID from the Analytical Ultracentrifugation Facility at the University of Connecticut (Storrs, CT).

Acknowledgments

We thank staff at the CHESS-F2 beamline for assistance, R. Flavell for technical support, and N. Hunt for administrative assistance. This research was supported by a National Science Foundation grant MCB-9817188 (R.F.), National Institutes of Health grants CA82171 (Y.S.) and CA34610 (J.M.), and the Howard Hughes Medical Institutes (J.M.). Y.S. is a Searle Scholar and a Rita Allen Scholar.

Received: August 27, 2001; revised October 10, 2001.

References

Abdollah, S., Macias-Silva, M., Tsukazaki, T., Hayashi, H., Attisano, L., and Wrana, J.L. (1997). TβRI phosphorylation of Smad2 on Ser465 and Ser467 is required for Smad2-Smad4 complex formation and signaling. J. Biol. Chem. *272*, 27678–27685.

Becker, S., Groner, B., and Muller, C.W. (1998). Three-dimensional structure of the Stat3β homodimer bound to DNA. Nature *394*, 145–151.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.

Calonge, M.J., and Massague, J. (1999). Smad4/DPC4 silencing and hyperactive Ras jointly disrupt transforming growth factor-β antiproliferative responses in colon cancer cells. J. Biol. Chem. *274*, 33637–33643.

Chacko, B.M., Qin, B., Correia, J.J., Lam, S.S., de Caestecker, M.P., and Lin, K. (2001). The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization. Nat. Struct. Biol. *8*, 248–253.

Chai, J., Shiozaki, E., Srinivasula, S.M., Wu, Q., Datta, P., Alnemri, E.S., and Shi, Y. (2001). Structural basis of caspase-7 inhibiton by XIAP. Cell *104*, 769–780.

Chen, X., Vinkemeier, U., Zhao, Y., Jeruzalmi, D., Darnell, J.E., and Kuriyan, J. (1998a). Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. Cell *93*, 827–839.

Chen, Y.-G., Hata, A., Lo, R.S., Wotton, D., Shi, Y., Pavletich, N., and Massasgue, J. (1998b). Determinants of specificity in the TGF-β signal transduction. Genes Dev. *12*, 2144–2152.

Cotton, G.J., and Muir, T.W. (1999). Peptide ligation and its application to protein engineering. Chem. Biol. *6*, R247–R256.

Darnell, J.E. (1997). STATs and gene regulation. Science *277*, 1630–1635.

Eppert, K., Scherer, S.W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, H., Tsui, L.-C., Bapat, B., Gallinger, S., Andrulis, I.L., et al. (1996). MADR2 maps to 18q21 and encodes a TGFβ-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. Cell *86*, 543–552.

Feng, X.-H., and Derynck, R. (1997). A kinase subdomain of transforming growth factor-β (TGF-β) type I receptor determines the TGF-β intracellular signaling specificity. EMBO J. *16*, 3912–3923.

Hahn, S.A., Schutte, M., Hoque, A.T.M.S., Moskaluk, C.A., da Costa, L.T., Rozenblum, E., Weinstein, C.L., Fischer, A., Yeo, C.J., Hruban, R.H., et al. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science *271*, 350–353.

Heldin, C.-H., Miyazomo, K., and ten Dijke, P. (1997). TGF-β signalling from cell membrane to nucleus through SMAD proteins. Nature *390*, 465–471.

Huse, M., Chen, Y.-G., Massague, J., and Kuriyan, J. (1999). Crystal structure of the cytoplasmic domain of the type I TGFβ receptor in complex with FKBP12. Cell *96*, 425–436.

Huse, M., Muir, T.W., Xu, L., Chen, Y.-G., Kuriyan, J., and Massague, J. (2001). The TGFβ receptor activation process: an inhibitor- to substrate-binding switch. Mol. Cell *8*, 671–682.

Jayaraman, L., and Massague, J. (2000). Distinct oligomeric states of SMAD proteins in the TGF-β pathway. J. Biol. Chem. *275*, 40710–40717.

Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110–119.

Kawabata, M., Inoue, H., Hanyu, A., Imamura, T., and 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.

Klaulis, P.J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Kretzschmar, M., Liu, F., Hata, A., Doody, J., and Massague, J. (1997). The TGF-β family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. Genes Dev. *11*, 984–995.

Kuriyan, J., and Cowburn, D. (1997). Modular peptide recognition domains in eukaryotic signaling. Annu. Rev. Biophys. Biomol. Struct. *26*, 259–288.

Laue, T., Shaw, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science, S. E. Harding, A. J. Rowe, and J. C. Horton, eds. (Cambridge, United Kingdom: The Royal Society of Chemistry), pp. 90–125.

Lo, R.S., Chen, Y.-G., Shi, Y., Pavletich, N.P., and Massague, J. (1998). The L3 loop: a structural motif determining specific interactions between SMAD proteins and TGF-β receptors. EMBO J. *17*, 996–1005.

Macias-Silva, M., Abdollah, S., Hoodless, P.A., Pirone, R., Attisano, L., and Wrana, L. (1996). MADR2 is a substrate of the TGFβ receptor and its phosphorylation is required for nuclear accumulation and signaling. Cell *87*, 1215–1224.

Massague, J. (1998). TGF-β signal transduction. Annu. Rev. Biochem. *67*, 753–791.

Massague, J., and Chen, Y.-G. (2000). Controlling TGF-β signaling. Genes Dev. *14*, 627–644.

Miyaki, M., Iijima, T., Konishi, M., Sakai, K., Ishii, A., Yasuno, M., Hishima, T., Koike, M., Shitara, N., Iwama, T., et al. (1999). Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. Oncogene *18*, 3098–3103.

Navaza, J. (1994). AMoRe and automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins Struct. Funct. Genet. *11*, 281–296.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Qin, B., Lam, S.S.W., and Lin, K. (1999). Crystal structure of a transcriptionally active Smad4 fragment. Structure *7*, 1493–1503.

Roberts, A.B., and Sporn, M.B. (1990). The transforming growth factor-βs. In Peptide Growth Factors and Their Receptors, M. B. Sporn, and A. B. Roberts, eds. (Heidelberg: Springer-Verlag), pp. 419–472.

Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. Cell *103*, 211–225.

Schutte, M., Hruban, R.H., Hedrick, L., Cho, K.R., Nadasdy, G.M., Weinstein, C.L., Bova, G.S., Isaacs, W.B., Cairns, P., Nawroz, H., et al. (1996). DPC4 gene in various tumor types. Cancer Res. *56*, 2527–2530.

Shi, Y., Hata, A., Lo, R.S., Massague, J., and Pavletich, N.P. (1997). A structural basis for mutational inactivation of the tumour suppressor Smad4. Nature *388*, 87–93.

Shi, Y. (2001). Structural insights on Smad function in TGFβ signaling. Bioessays *23*, 223–232.

Sicheri, F., and Kuriyan, J. (1997). Structures of Src-family tyrosine kinases. Curr. Opin. Struct. Biol. *7*, 777–785.

Sicheri, F., Moarefi, I., and Kuriyan, J. (1997). Crystal structure of the Src-family tyrosine kinase Hck. Nature *385*, 602–609.

Souchelnytskyi, S., Tamaki, K., Engstrom, U., Wernstedt, C., ten Dijke, P., and Heldin, C.H. (1997). Phosphorylation of Ser465 and Ser467 in the C terminus of Smad2 mediates interaction with Smad4 and is required for transforming growth factor-β signaling. J. Biol. Chem. *272*, 28107–28115.

Tsukazaki, T., Chiang, T.A., Davison, A.F., Attisano, L., and Wrana, J.L. (1998). SARA, a FYVE domain protein that recruits Smad2 to the TGFβ receptor. Cell *95*, 779–791.

Wieser, R., Wrana, J.L., and Massague, J. (1995). GS domain mutations that constitutively activate TβR-I, the downstream signaling component in the TGF-β receptor complex. EMBO J. *14*, 2199–2208.

Willis, S., Zimmerman, C.M., Li, L., and Mathews, L.S. (1996). Formation and activation by phosphorylation of activin receptor complexes. Mol. Endocrinol. *10*, 367–379.

Wrana, J.L., and Attisano, L. (2000). The Smad pathway. Cytokine Growth Factor Rev. *11*, 5–13.

Wrana, J.L., Attisano, L., Wieser, R., Ventura, F., and Massague, J. (1994). Mechanism of activation of the TGF-β receptor. Nature *370*, 341–347.

Wu, G., Chen, Y.-G., Ozdamar, B., Gyuricza, C.A., Chong, P.A., Wrana, J.L., Massague, J., and Shi, Y. (2000). Structural basis of Smad2 recognition by the Smad anchor for receptor activation. Science *287*, 92–97.

Wu, J.-W., Fairman, R., Penry, J., and Shi, Y. (2001). Formation of a stable heterodimer between Smad2 and Smad4. J. Biol. Chem. *276*, 20688–20694.

Xu, L., Chen, Y.-G., and Massague, J. (2000). Smad2 nuclear import function masked by SARA and unmasked by TGFβ-dependent phosphorylation. Nat. Cell Biol. *2*, 559–562.

Xu, W., Harrison, S.C., and Eck, M. (1997). Three-dimensional structure of the tyrosine kinase c-Src. Nature *385*, 595–602.

Accession Numbers

Atomic coordinates for the phosphorylated Smad2 have been deposited with the Protein Data Bank (assession number 1KHX).