

Structural Basis of Smad2 Recognition by the Smad Anchor for Receptor Activation

Geng Wu,^{1*} Ye-Guang Chen,^{2*} Barish Ozdamar,³
Cassie A. Gyuricza,¹ P. Andrew Chong,³ Jeffrey L. Wrana,³
Joan Massagué,² Yigong Shi^{1†}

The Smad proteins mediate transforming growth factor- β (TGF β) signaling from the transmembrane serine-threonine receptor kinases to the nucleus. The Smad anchor for receptor activation (SARA) recruits Smad2 to the TGF β receptors for phosphorylation. The crystal structure of a Smad2 MH2 domain in complex with the Smad-binding domain (SBD) of SARA has been determined at 2.2 angstrom resolution. SARA SBD, in an extended conformation comprising a rigid coil, an α helix, and a β strand, interacts with the β sheet and the three-helix bundle of Smad2. Recognition between the SARA rigid coil and the Smad2 β sheet is essential for specificity, whereas interactions between the SARA β strand and the Smad2 three-helix bundle contribute significantly to binding affinity. Comparison of the structures between Smad2 and a co-mediator Smad suggests a model for how receptor-regulated Smads are recognized by the type I receptors.

TGF β signaling plays a central role in regulating cellular responses such as growth, differentiation, and cell fate specification (1). TGF β signaling from the membrane to the nucleus is mediated by the Smad family of proteins, which contains at least eight members in vertebrates (2, 3). Two Smad proteins, Smad2 and Smad4, are tumor suppressors in humans (2, 3).

The Smad proteins are divided into three functional classes: (i) the co-mediator Smads (co-Smads), Smad4 and Smad10, participate in signaling by diverse TGF β family members; (ii) the receptor-regulated Smads (R-Smads), including Smad1, Smad2, Smad3, Smad5, and Smad8, are each involved in a specific signaling pathway; and (iii) the antagonistic Smads, including Smad6 and Smad7, negatively regulate these pathways (2).

To initiate a particular TGF β response, a specific TGF β ligand binds to a specific pair of transmembrane Ser-Thr receptor kinases, the type I and type II receptors, and this activates the Ser-Thr kinase in the cytoplasmic domain of the type I receptor (2, 4). The signal is then propagated by type I receptor-mediated phosphorylation of specific R-Smads. The R-Smads, Smad2 and 3, are recruited to the TGF β recep-

tors by SARA. This process appears to involve direct interactions between SARA and Smad2, SARA and TGF β receptors, and Smad2 and TGF β receptors (5). SARA does not interact with either Smad1 or Smad5 (5), which share ~80% sequence identity with Smad2. These interactions are important in regulating specific Smad phosphorylation. The phosphorylated R-Smad hetero-oligomerizes with the co-Smad, Smad4, translocates into the nucleus and associates with sequence-specific DNA binding proteins, resulting in the positive or negative regulation of agonist-responsive genes (2).

The Smad proteins are conserved across species, with homology mainly in the NH₂-terminal MH1 domain and the COOH-terminal MH2 domain. Most tumor-derived mutations map to the MH2 domain that is involved in receptor recognition, interaction with transcription factors, and homo- and hetero-oligomerization among Smads (2). The MH2 domain of Smad2 or Smad3 interacts with an 85-residue Smad-binding domain (SBD) in SARA (5). The MH1 domain exhibits sequence-specific DNA binding activity and negatively regulates the functions of the MH2 domain.

To examine the molecular basis of signaling specificity in receptor activation, we have determined the crystal structure of a Smad2 MH2 domain bound to a SARA SBD. In the crystals, SARA SBD exhibits an extended conformation, folding on top of Smad2 and shielding the hydrophobic residues on the surface of Smad2. Analyses of the SARA-Smad2 interface and the surface features of Smad2 give insight into Smad2 interactions with SARA and how R-Smads are recognized by the type I receptors.

Characterization of Smad2-SARA inter-

action. To facilitate structural studies of a Smad2-SARA complex, we took a biochemical and cell biological approach to identify the minimal domains required for complex formation. Co-transfection of Smad2 and SARA followed by co-immunoprecipitation and protein immunoblot analysis (6) indicated that the Smad2 MH2 domain (residues 270–467), derived from a structure-based comparison (7), exhibited a weak interaction with SARA (Fig. 1A). Including nine additional NH₂-terminal residues (residues 261–467) restored efficient interaction (Fig. 1A). To define the molecular determinants of binding, we examined Smad2 mutants for their ability to interact with SARA. The interaction does not require either the L3 loop or the H1 or H2 helices (Fig. 1B) (7), which interact with receptors at the membrane and transcription factors in the nucleus (8, 9). A single missense mutation in the β strand B9, Asn³⁸¹ \rightarrow Ser³⁸¹ (N381S) (10), nearly abolished Smad2 recognition by SARA (Fig. 1A). This residue is subtype-specific among Smads, being an Asn in Smad2 and Smad3 but replaced by a Ser in Smad1, Smad5, and Smad8. Other subtype-specific mutations, such as Q294S and R337H, had little or no effect on Smad2-SARA recognition (Fig. 1A).

Limited proteolysis on a Smad2-SARA complex was used to identify a 57-residue (665–721) trypsin-resistant fragment in SARA as sufficient for interaction with Smad2 (11). Proteolytic treatment of this isolated fragment destroyed the polypeptide and failed to generate any detectable structural cores (11). Thus, SARA SBD might adopt a flexible conformation that is stabilized by binding to Smad2. We generated a series of point mutations on SARA SBD and evaluated their ability to interact with Smad2 in an *in vitro* binding assay (12). Mutation of any of the four residues in SARA, Tyr⁶⁸⁰, Cys⁶⁸¹, Pro⁶⁸⁶, and Val⁷⁰³ significantly diminished complex formation between SARA and Smad2 (Fig. 1C). In contrast, point mutations such as Q688A and Q691A had no detectable effect on binding (Fig. 1C).

Overall structure of the complex. To gain additional insight into the specificity in SARA-Smad2 interactions, we determined the 2.2 Å crystal structure of a Smad2 MH2 domain (residues 253–462) in complex with a SARA SBD (residues 665–721) (Table 1) (13). In the complex, SARA SBD adopts an extended conformation that consists sequentially of a rigid coil, an amphipathic α helix, a proline-rich turn, and a β strand (Fig. 2A). SARA SBD does not have a hydrophobic core, and its secondary structural elements do not interact with each other. Instead, the extended structure of SARA SBD covers a large surface area on Smad2 and interacts with multiple surface structural motifs on Smad2 that are as far apart as 40 Å. For example, the COOH-terminal β strand of SARA SBD in-

¹Department of Molecular Biology, Princeton University, Lewis Thomas Laboratory, Princeton, NJ 08544, USA.

²Cell Biology Program, Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA. ³Program in Molecular Biology and Cancer, Department of Medical Genetics and Microbiology, University of Toronto, Room 1075, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada M5G 1X5.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: yshi@molbio.princeton.edu

RESEARCH ARTICLES

teracts with the NH₂-terminal strand B1' in Smad2, forming an antiparallel β sheet. This β sheet packs against a hydrophobic surface formed by helices H3 and H5 on Smad2 (Fig. 2A).

Binding between SARA and Smad2 results in close packing of hydrophobic residues and buries 2573 Å² of surface area. The interfaces involving the rigid coil and the β strand in SARA contribute approximately 900 Å² (35%) and 1150 Å² (45%) buried surface areas, respectively. The four residues shown biochemically to affect binding (Fig. 1C) contribute to large buried surface areas upon complex formation (Fig. 2B).

The Smad2 MH2 domain contains a central β sandwich, with a three-helix bundle (H3, H4, and H5) and a single strand (B1') on one end, and a loop-helix region (L1, L2, L3, and H1) on the other end (Fig. 2A). The overall structural architecture is similar to that of Smad4 MH2, with a root-mean-square-deviation (rmsd) of 1.27 Å for 174 aligned Cα atoms (Fig. 2C). The similarity between these two structures is higher in the central β sandwich, with an rmsd of 0.55 Å for 129 aligned Cα atoms. Two differences between the MH2 domains of Smad2 and Smad4 are likely to play important functional roles in signaling. The β strand B1' is protease-resistant and present in Smad2 MH2, but in Smad4 MH2 this region is protease-sensitive and likely unstructured (7). This β strand comprises the NH₂-terminal nine residues that are critical for interaction with SARA (Fig. 1A). The orientation of the three-helix bundle relative to the central β sandwich is also different between Smad2 and Smad4. In Smad4, the large flexible linker between helices H3 and H4 is disordered; in Smad2, this portion is replaced with a well-ordered four-residue short linker. Consequently, the three-helix bundle in Smad2 appears to be more compact and is at a wider angle relative to the planes of the β sandwich.

Smad2-SARA interactions. The Smad2-SARA interactions involve three contact surfaces from Smad2 and three corresponding secondary structural elements in SARA: the rigid coil, the amphipathic α helix, and the β strand (Fig. 3A). The hydrophobic surfaces on the Smad2 MH2 domain form two shallow grooves and one deep groove; the hydrophobic side chains of SARA SBD pack closely against these three grooves (Fig. 3A).

The rigid coil (residues 671–682) binds to one of the shallow surface grooves that is created by residue side chains on strands B8/B9, and the short loop connecting helix H2 and strand B8 (Fig. 3B). In this region of SARA, the interacting residues that include Pro⁶⁷², Pro⁶⁷⁴, Pro⁶⁷⁷, Tyr⁶⁸⁰, and Cys⁶⁸¹ form an extensive network of van der Waals contacts with the surface residues in Smad2. On Smad2, the interacting residues are Tyr³⁶⁶, Trp³⁶⁸, Thr³⁷², Cys³⁷⁴, Ile³⁷⁶, Pro³⁷⁷,

and Leu³⁸². In addition to the van der Waals contacts, this part of the interface contains five intermolecular H bonds (Fig. 3B) that include two backbone-to-backbone contacts between Ser⁶⁸² on SARA and Lys³⁸³ on Smad2 and three backbone-to-side-chain

contacts. In particular, the carbonyl O atom of Pro⁶⁷² accepts a H bond from the side chain of subtype-specific Trp³⁶⁸ on Smad2, whereas Tyr⁶⁸⁰ in SARA makes a H bond to the backbone amide of Lys³⁷⁵ on Smad2 (Fig. 3B).

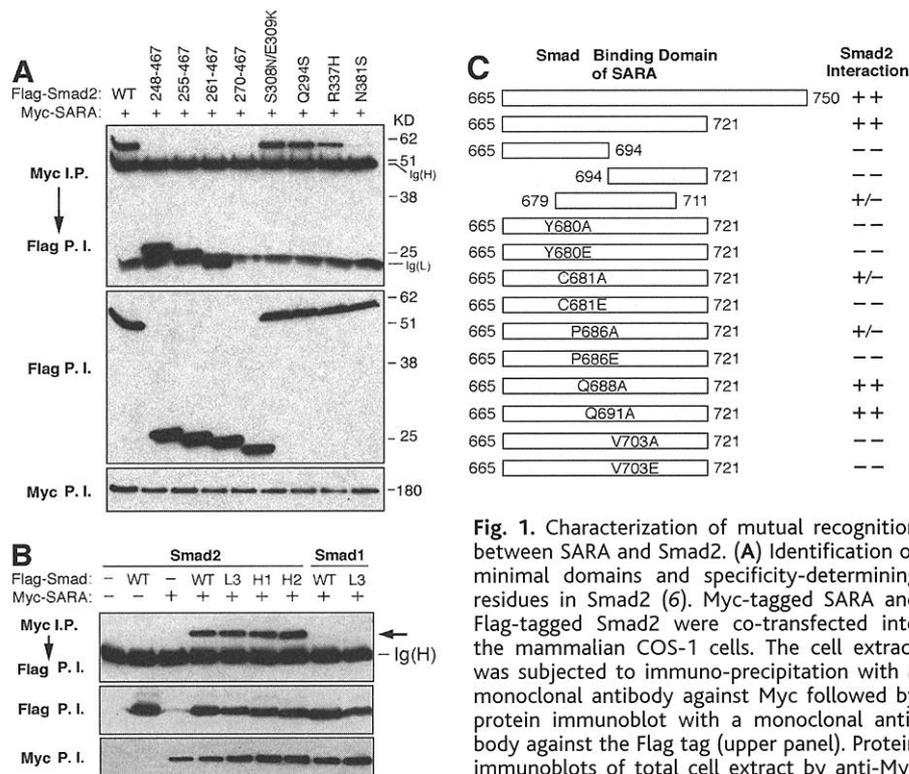


Fig. 1. Characterization of mutual recognition between SARA and Smad2. (A) Identification of minimal domains and specificity-determining residues in Smad2 (6). Myc-tagged SARA and Flag-tagged Smad2 were co-transfected into the mammalian COS-1 cells. The cell extract was subjected to immuno-precipitation with a monoclonal antibody against Myc followed by protein immunoblot with a monoclonal antibody against the Flag tag (upper panel). Protein immunoblots of total cell extract by anti-Myc and anti-Flag, respectively, demonstrated similar expression levels for SARA and for various constructs of Smad2 (middle and lower panels). I.P., immunoprecipitate; P.I., protein immunoblot. (B) Subtype-conserved surface motifs L3 loop and helices H1 and H2 are not involved in interactions with SARA. In lanes 5, 6, and 7, these motifs in Smad2 are individually replaced with those from Smad1. In lane 8, the L3 loop in Smad1 was substituted with that from Smad2. (C) Identification of minimal domains and key residues in SARA for interaction with Smad2. The SARA fragments were individually expressed as a GST-fusion protein and were purified to homogeneity through a glutathione sepharose column (12). The interaction between these proteins and Smad2 was then investigated as described (12).

Table 1. Data collection and statistics from the crystallographic analysis.

Data set	Native 1 (RAXIS II)	Native 2 (X25, NSLS)
	<i>Data collection</i>	
Resolution (Å)	99.0–3.0	99.0–2.2
Total observations	89,878	218,815
Unique observations	12,533	31,296
Data coverage (outer shell)	97.8% (96.2%)	99.1% (97.6%)
R_{sym} (outer shell)*	0.066 (0.203)	0.039 (0.167)
	<i>Refinement</i>	
Resolution range (Å)	20.0–2.2	
Number of reflections ($I > 2\sigma$)	29,613	
$R_{\text{working}}/R_{\text{free}}\ddagger$	21.8%/27.6%	
Number of atoms	3888	
Number of water molecules	243	
rmsd bond length (Å)‡	0.006	
rmsd bond angles (degrees)‡	1.343	
rmsd B factors‡	3.375	

* $R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \bar{I}_h| / \sum_h \sum_i I_{hi}$, where I_{hi} is the mean intensity of the i observations of symmetry related reflections of h . $\ddagger R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$, where $F_{\text{obs}} = F_p$ and F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with 5% of the reflections). \ddagger rmsd in bond lengths and angles are the deviations from ideal values, and the rmsd in B factors is calculated between bonded atoms.

The rigid coil has low-temperature factors in the crystals and an unusually compact conformation that is facilitated by three proline residues at positions 672, 674, and 677 (Fig. 3B). This rigid structure is strengthened by a network of nine intramolecular H bonds (Fig. 3B, right panel). In this network, Asn⁶⁷³ and Asn⁶⁷⁶ each mediate four H bond contacts.

In the second part of the Smad2-SARA interface, an amphipathic α helix from SARA binds parallel to the hydrophobic strands B5 and B6 on Smad2 (Fig. 3C). A sharp turn between the rigid coil and the helix is stabilized by intermolecular H bonds from the side chain

of Asn³⁸¹ on Smad2 to the backbone carbonyl group of Ile⁶⁸⁴ and to the side chain of Gln⁶⁸⁹ on SARA. In addition, the side chains of Ser⁶⁸² and Gln⁶⁸⁹ on SARA make two intramolecular H bonds to the backbone amide and carbonyl groups of Ile⁶⁸⁴, respectively (Fig. 3C). The importance of these interactions is highlighted by the observation that the single point mutation N381S in Smad2 leads to disruption of the Smad2-SARA interaction (Fig. 1A). In contrast to the extensive van der Waals interactions at the interface involving the rigid coil, only two residues from SARA, Pro⁶⁸⁶ and Ala⁶⁹⁰, contact the hydrophobic side chains of Tyr³³⁹, Ile³⁴¹, and Phe³⁴⁶ on Smad2 (Fig. 3C).

In the third part of the Smad2-SARA interface, the COOH-terminal β strand from SARA SBD forms an antiparallel β sheet with the NH₂-terminal strand B1' on Smad2, through five backbone-to-backbone H bonds (Fig. 3D). This chimeric β sheet covers an exposed hydrophobic surface on helices H3 and H5 of Smad2. The side chains of Pro⁷⁰¹, Val⁷⁰³, and Val⁷⁰⁵ from SARA are involved in multiple van der Waals contacts with the residues Ala³⁹¹, Leu³⁹⁴, and Val³⁹⁸ on helix H3; Trp⁴⁴⁸, Gln⁴⁵⁵, and Met⁴⁵⁶ on helix H5; and Tyr²⁶⁸ and Val²⁶⁶ on strand B1'.

Specificity of Smad2-SARA interaction. Despite 80% sequence identity between the MH2 domains of Smad1 and Smad2, SARA only binds the TGF β /activin-regulated R-Smads, Smad2 or Smad3, but not the BMP-2/BMP-4-activated Smad1 or Smad5 (5). Our structural analysis of the Smad2-SARA complex provides a molecular basis for this specificity (Fig. 3 and Fig. 4A).

Five Smad2 residues that interact with SARA are subtype-conserved in Smad3, but are replaced by other residues in Smad1, Smad5, or Smad8 (Fig. 4A). Two of these residues, Asn³⁸¹ and Trp³⁶⁸, participate in H bonds (Fig. 3B) and likely constitute the critical determinants of specificity in recognition of Smad2. The remaining three residues, Tyr³⁶⁶, Ile³⁴¹, and Phe³⁴⁶, stabilize the complex through hydrophobic interactions. Together, these residues define the subtype-specific recognition of Smad2 or Smad3 by SARA. Their role in signaling is illustrated by the detection in colon cancer of the relatively conservative mutation, F346V in Smad2, that likely leads to reduced van der Waals interactions between Smad2 and SARA (14).

None of the five specificity-determining residues on Smad2 interacts with the β strand in SARA, although the β strand contributes 45% to the buried surface area. Thus, the recognition specificity is determined primarily from the rigid coil and α helix of SARA interacting with the strands B8, B9, B5, and B6 of Smad2, whereas significant binding affinity is provided from the interface between the β strand of SARA and the three-helix bundle of Smad2.

To demonstrate the specificity in TGF β signaling, we compared cellular localization patterns of SARA and either wild-type or the N381S mutant Smad2 using confocal microscopy (15). In cells expressing wild-type or mutant (N381S) Smad2 alone, the protein exhibited a diffuse staining pattern throughout the cytosolic compartment (Fig. 4B, top panels). In contrast, SARA alone displayed a characteristic punctate staining pattern (Fig. 4B, top panels). The localization of the wild-type Smad2 in the presence of SARA exhibited a dramatic change to a punctate staining pattern (Fig. 4B, middle panels). However, the localization of the mutant Smad2

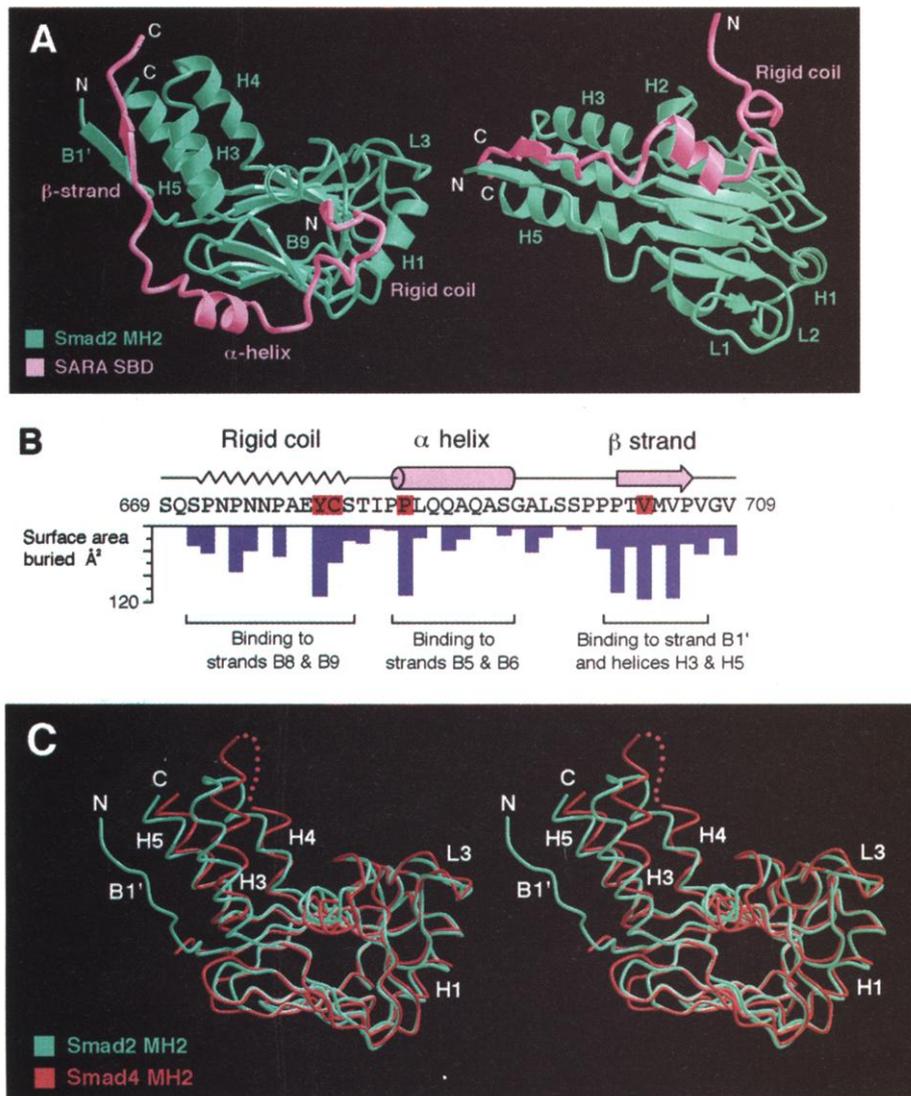


Fig. 2. Overall structure of the Smad2 MH2 domain in complex with a SARA SBD. **(A)** The schematic representation on the right panel is related to the one on the left by a 90° rotation along the horizontal axis. Smad2 and SARA are shown in green and pink, respectively. The secondary structural elements in SARA and some prominent features in Smad2 are labeled and color-coded. **(B)** Sequence of the SARA SBD showing its secondary structural elements. The bar graph below sequence shows the buried surface area per SARA residue upon complex formation. The residues that are targeted by inactivating mutations are highlighted in red (Fig. 1C). **(C)** Superimposition of the structures of the MH2 domains in Smad2 (green) and Smad4 (red), shown in stereo view. The disordered loop between helices H3 and H4 in Smad4 is indicated by a red dotted line. This figure was prepared with MOLSCRIPT (24).

(N381S) was not influenced by the presence of SARA, demonstrating that the mutant Smad2 has lost its ability to co-localize with the wild-type SARA (Fig. 4B, lower panels), consistent with the loss of binding in co-immunoprecipitation assays (Fig. 1A). To evaluate the impact of this mutation on TGF β signaling, we performed a luciferase reporter assay with a TGF β -responsive *Mix.2* promoter (Fig. 4C) (16, 17). The mutant Smad2 failed to synergize with SARA to potentiate the TGF β response whereas the wild-type Smad2 did synergize (Fig. 4C).

Model of recognition between Smads and TGF β receptors. Despite having generally similar structures, the MH2 domains of Smad2 and Smad4 have different surface features. A direct comparison of the electrostatic potential reveals the presence of a highly positively charged groove next to the L3 loop on Smad2, but not on Smad4 (Fig. 5A). This basic surface contains residues that are conserved in R-Smads but not in co-Smads, suggesting that this region might be important for receptor binding. Analysis of the type I TGF β receptor (T β RI) cytoplasmic domain reveals a complementary pattern on its surface. Specifically, the L45 loop, which specifies interactions with Smad2 (8, 18), is located immediately adjacent to the flexible GS region (19), which is phosphorylated upon ligand binding (20, 21, 22) and becomes very acidic. Thus, the phosphorylated GS region on the type I receptors might interact with a highly basic surface groove on R-Smads to provide binding affinity, whereas the L45 loop on the type I receptors recognizes the L3 loop of specific R-Smads to provide specificity (Fig. 5B).

This model is consistent with biological and biochemical evidence. All pathway-restricted Smads share five basic residues in the basic groove, four of which are invariant. The variable basic residue, located on the L3 loop, is Arg in Smad2 and Smad3 (Arg⁴²⁷ in Smad2) and His in Smad1, Smad5, and Smad8 (His 425). This Arg and another subtype-specific residue on the L3 loop, Thr in Smad2 and Smad3 (Thr 430 in Smad2) and Asp in Smad1, Smad5, and Smad8 (Asp 428), are the specificity-determinants in the Smad-receptor interactions (8, 9). The position of these two residues next to the basic surface in R-Smads is consistent with the topological arrangement between the L45 loop and the GS region in the type I receptors.

If interactions in the GS region contribute to binding affinity, they should be relatively non-specific and tolerant of mutations. Consistent with this prediction, single or double point mutations of the Thr-Ser residues in the GS region had little or no effect on the signaling activities of the type I TGF β or activin receptors, respectively (20, 21). Several mutations, such as T204D or T204E in T β RI (20) and T206D in a type I activin receptor (21), result in constitu-

tively active kinases. These mutations affect a Thr residue at the COOH-terminal end of the GS-region. On the basis of our model, the mutation from Thr to a negatively charged Asp or Glu could favor the initial apposition between T β RI and Smad2, which should facilitate subsequent binding and phosphorylation.

This model also places the COOH-terminal portion of SARA close to T β RI, consistent with reported interactions (5). After phosphorylation, the COOH-terminal SSXS motif becomes highly acidic and could compete with T β RI for binding to the highly basic surface patch next to the L3 loop, which would destabilize the Smad2-receptor complex (5).

The details of recognition between the receptors and Smads will only be revealed by a

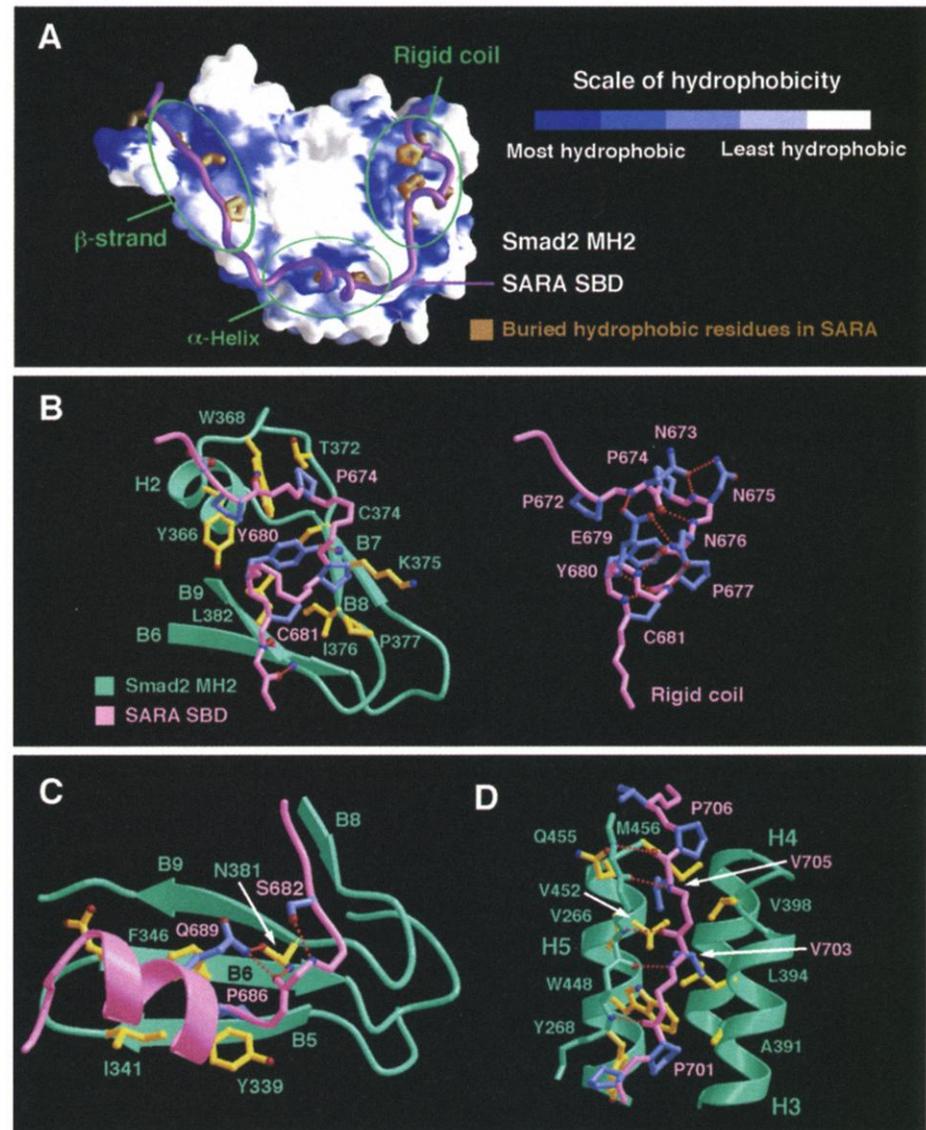


Fig. 3. Schematic representation of the interactions between Smad2 and SARA. (A) The interactions are predominantly hydrophobic in nature. The surface of Smad2 MH2 domain is represented by degrees of hydrophobicity. The α backbone of SARA SBD is shown in pink and the buried hydrophobic residues are highlighted in orange. This figure was prepared with GRASP (25). (B) A closeup view of the interactions between the rigid coil of SARA and the strands B8 and B9 of Smad2. Smad2 and SARA are colored green and pink, respectively. The interacting side chains are shown in yellow for Smad2 and in purple for SARA. The O and N atoms are shown as red and blue balls, respectively. The left panel shows the interface, whereas the right panel shows the conformation of the rigid coil by itself. Aside from extensive van der Waals interactions at the interface, there are a total of five intermolecular H bonds. These include: Ser⁶⁷¹ O γ to Tyr³⁶⁶ carbonyl, Pro⁶⁷² carbonyl to Trp³⁶⁸ N ϵ 1, Tyr⁶⁸⁰ O η to Lys³⁷⁵ amide, Ser⁶⁸² amide to Asn³⁸¹ carbonyl, and Ser⁶⁸² carbonyl to Asn³⁸¹ amide. (C) A closeup view of the interactions between the α helix of SARA and strands B5 and B6 of Smad2. Color coding scheme is identical to (B). (D) A closeup view of the interactions between the β strand of SARA and the three-helix bundle and strand B1' of Smad2.

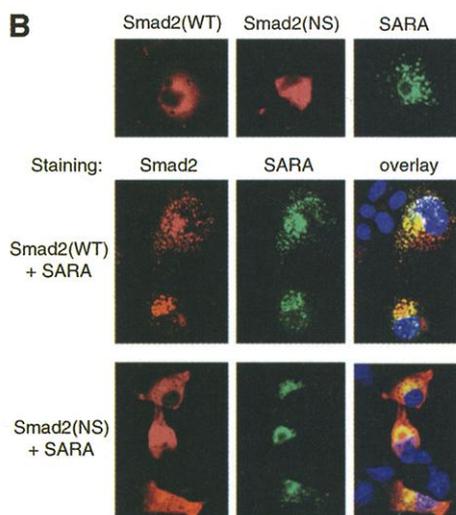
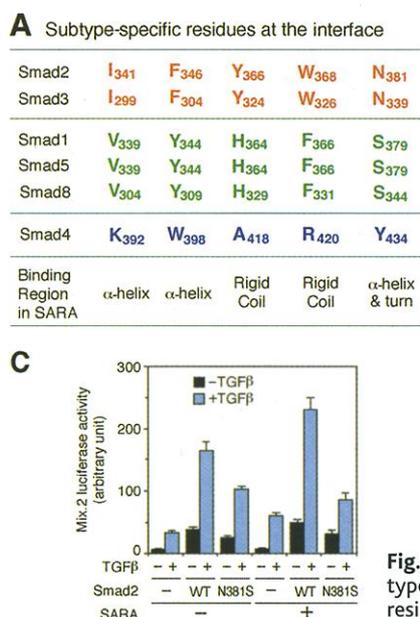


Fig. 4. Specificity of Smad2-SARA interactions. (A) Subtype-specific residues at the Smad2-SARA interface. The residues that are conserved and specific for Smad2 and Smad3 are highlighted in red, whereas those for Smad1, Smad5, and Smad8 are shown in green. All these residues are involved in interactions with SARA. (B) Immunofluorescence assays showing the loss of co-localization between SARA and the mutant Smad2 (N381S) (15). After transfection with Flag-Smad2 and Myc-SARA, COS-1 cells were immunostained with primary mouse anti-Flag (M2) and rabbit anti-Myc and secondary rhodamine-conjugated goat anti-mouse and FITC-conjugated goat anti-rabbit antibodies, and examined with a confocal microscope. To illustrate the cellular compartments, the nuclei were stained with DAPI in the two overlay panels. (C) SARA potentiates the activity of wild-type but not the mutant Smad2 (N381S) in a *Mix.2* ARE-luciferase reporter assay (17). HepG2 cells were co-transfected with a *Mix.2* ARE-luciferase reporter, FAST-2, with or without SARA and Smad2 constructs, as indicated. Cells were treated with TGFβ and the luciferase activity was determined. The experiment was performed in triplicate.

structure of the complex. For example, the TβRI kinase domain in the crystal structure was in an inactive form (19). Phosphorylation of the GS region might induce additional conformational changes in the receptor that further facilitate interactions with Smad2.

Other perspectives. The binding of SARA-SBD to Smad2 resembles p27 binding to the Cyclin A-CDK2 complex (23). For example, p27 also displays an extended structure and contains sequentially a rigid coil, an α helix, and a β strand (23). Both peptides interact mostly through van der Waals interactions; both contain a rigid coil contacting a shallow hydrophobic surface groove; both use an α helix to connect to another major interaction interface; and both use a β strand to pair up with an existing strand from the bound partner. The analogy can be extended to some details. For example, there is a conserved Trp residue at the surface groove of Cyclin 2 for interaction with the rigid coil from p27 (23); this feature is also present in the SARA-Smad2 complex (Fig. 3B).

It is likely that the structure of Smad2 MH2 by itself includes two antiparallel β strands at the NH₂-terminus, one of which is displaced by SARA binding. This hypothesis is supported by several biochemical observations. First, a Smad2 MH2 domain beginning at residue 253 is not soluble by itself but

becomes highly soluble when co-expressed with SARA SBD, suggesting that the β strand from SARA shields and rescues the exposed hydrophobic surface on helices H3 and H5. Second, a Smad2 MH2 domain beginning at residue 241 is soluble by itself, suggesting that residues 241–253 might play the same role as the β strand in SARA. Third, a Smad2 MH2 domain beginning at residue 269 is not soluble either by itself or when co-expressed with SARA SBD, presumably because of the loss of strand B1' (residues 263–268), which eliminates the possibility of forming a β sheet with SARA SBD.

Our biochemical and structural analyses reveal a molecular basis for the specific recognition of Smad2 by SARA, provide a plausible model for receptor-Smad interaction, and serve as a framework for further elucidation of Smad functions in TGFβ signaling.

References and Notes

1. A. B. Roberts and M. B. Sporn, in *Peptide Growth Factors and Their Receptors*, M. B. Sporn and A. B. Roberts, Eds. (Springer-Verlag, Heidelberg, Germany, 1990), pp. 419–472.
2. J. Massagué, *Annu. Rev. Biochem.* **67**, 753 (1998).
3. C.-H. Heldin, K. Miyazono, P. ten Dijke, *Nature* **390**, 465 (1997).
4. J. L. Wrana, L. Attisano, R. Wieser, F. Ventura, J. Massagué, *Nature* **370**, 341 (1994).
5. T. Tsukazaki, T. A. Chiang, A. F. Davison, L. Attisano, J. L. Wrana, *Cell* **95**, 779 (1998).
6. COS1 cells were transfected by lipofectAMINE (GIBCO)

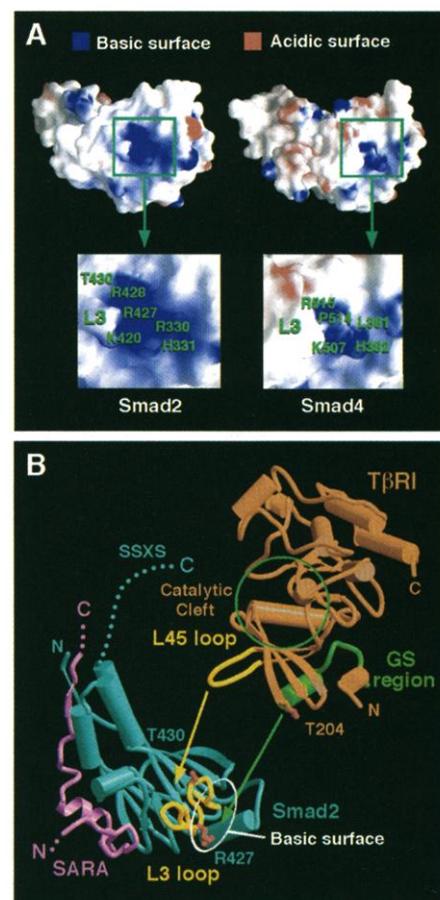


Fig. 5. Identification of a model for interactions of R-Smads with the type I receptors. (A) Comparison of surface electrostatic potentials on the MH2 domains of Smad2 and Smad4. The highly basic surface patch is conserved in all R-Smads but absent in co-Smads, Smad4. The surfaces with electrostatic potentials $< -10 k_B T$ are colored red, whereas those $> +10 k_B T$ are represented in blue (k_B and T are the Boltzmann constant and temperature, respectively). (B) A proposed model for Smad2-TβRI interaction. TβRI (19), shown in gold, interacts with Smad2, shown in green, through two interfaces, with the phosphorylated GS region binding to the highly basic surface patch to provide affinity and the L45 loop recognizing L3 loop to provide specificity.

with Flag-Smad and Myc-SARA constructs. The cell lysates were immunoprecipitated with anti-Myc 9E10 (Santa Cruz Biotechnology), and the immunoprecipitates were analyzed by protein immunoblotting with anti-Flag M2 (Sigma). Protein expression was confirmed by protein immunoblotting of total cell lysates with antibodies to Flag (Smads) or myc (SARA).

7. Y. Shi, A. Hata, R. S. Lo, J. Massagué, N. P. Pavletich, *Nature* **388**, 87 (1997).
8. Y.-G. Chen *et al.*, *Genes Dev.* **12**, 2144 (1998).
9. R. S. Lo, Y.-G. Chen, Y. Shi, N. P. Pavletich, J. Massagué, *EMBO J.* **17**, 996 (1998).
10. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. X indicates any residue.
11. Individually purified Smad2 MH2 domain (residues 241–467) and SARA SBD (residues 665–750) were mixed in a 1:1 molar ratio. After limited proteolysis by trypsin or by subtilisin, the resulting complex was

- visualized on 15% denaturing polyacrylamide gels using tricine buffer. All stable fragments larger than 2 kD were identified through NH₂-terminal sequencing and mass spectrometric analyses.
12. Mutant constructs were generated using standard polymerase chain reaction–based cloning strategy, and the identities of individual clones were verified through double-strand plasmid sequencing. The Smad-binding fragment from SARA (residues 665–721) was overexpressed in *Escherichia coli* strain BL21(DE3) as a glutathione S-transferase (GST)–fusion protein using a pGEX-2T vector (Pharmacia) and was purified by a glutathione sepharose 4B affinity column. The MH2 domain from Smad2 (residues 241–467) was overexpressed in a pET3d vector (Novagen). The soluble fraction of Smad2 MH2 in the *E. coli* lysate was purified by cation-exchange chromatography (SP-sepharose; Pharmacia) and gel-filtration chromatography (Superdex-75 column; Pharmacia). Equimolar amounts of GST-SARA SBD and Smad2 MH2 domain were mixed and incubated in 25 mM NaMES (pH 6.0), 50 mM NaCl, and 2 mM dithiothreitol (DTT). The complex was then passed through a cation-exchange column (SP-sepharose; Pharmacia), to which Smad2 MH2 domain binds avidly. GST-SARA SBD in isolation does not bind this column. Smad2 binding to the column has no effect on interaction with SARA. The bound complex was eluted from this column with 1 M NaCl and visualized on 15% SDS–polyacrylamide gel electrophoresis.
 13. Proteins of the Smad2 MH2 domain and SARA SBD were individually purified and mixed in a 1:1 molar ratio. The final complex was concentrated and purified through gel-filtration chromatography (Superdex-75 column; Pharmacia). The concentration of the complex is ~20 mg/ml. Heavily twinned crystal clusters were grown at 4°C by the hanging-drop vapor-diffusion method by mixing the SARA-Smad2 protein complex with an equal volume of reservoir solution containing 100 mM Tris buffer (pH 8.5), 10% Dioxane (v/v), 2.0 M ammonium sulfate, and 10 mM DTT. Streak-seeding followed by three rounds of macro-seeding eventually generated crystals suitable for x-ray diffraction. The crystals, with a typical size of 0.1 mm by 0.1 mm by 0.4 mm, are in the trigonal space group P3121, with unit cell dimensions $a = b = 138.5 \text{ \AA}$, $c = 55.9 \text{ \AA}$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$, and contain two complexes in the asymmetric unit. Initial diffraction data were collected using an R-AXISIIIC imaging plate detector mounted on a Rigaku 200HB generator. High-resolution data sets were collected at beamline X25 at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory. All data sets were collected under freezing conditions; crystals were equilibrated in a cryoprotectant buffer containing 100 mM Tris buffer (pH 8.5), 10% Dioxane (v/v), 2.0 M ammonium sulfate, and 20% glycerol, and were flash frozen under a –170°C nitrogen stream. The structure was primarily determined by molecular replacement using the software AMoRe [J. Navaza, *J. Acta Crystallogr. A* **50**, 157 (1994)]. The atomic coordinates of Smad4 MH2 were used for a rotational search against a 15–3.5 Å data set. The top 50 solutions from the rotational search were individually used for a subsequent translational search, which yielded one solution with a correlation factor of 20.8 and an *R*-factor of 52.4%. This solution was used to locate the second complex in the crystals. Together, these two solutions gave a combined correlation factor of 33.5 and an *R*-factor of 44%. This model was examined with the program O [T. A. Jones *et al.*, *Acta Crystallogr. A* **47**, 110 (1991)], and the Smad4 side chains were replaced with those of Smad2. Refinement by simulated annealing with the program X-PLOR (A. T. Brünger, Yale University), against a 3.0 Å native data set decreased the *R* factor and *R* free to 35% and 42%, respectively. Refinement against 2.2 Å resolution data allowed progressive identification of the SARA fragment. The final refined model contains two complexes of Smad2 (residues 263–456) and SARA (residues 669–709), and 243 water molecules. The NH₂- and COOH-terminal residues in Smad2 have no electron density, and we presume that these regions are disordered in the crystals. The two complexes in one asymmetric unit are similar with a rmsd of 0.79 Å for all aligned C α atoms. We only report here one representative complex.
 14. M. Miyaki *et al.*, *Oncogene* **18**, 3098 (1999).
 15. Immunofluorescence was performed as described (9). Rhodamine-conjugated goat anti-mouse and fluorescein isothiocyanate (FITC)–conjugated goat anti-rabbit antibodies were purchased from Jackson ImmunoResearch Laboratories.
 16. X. Chen, M. J. Rubock, M. Whitman, *Nature* **383**, 691 (1996).
 17. Using a calcium phosphate–DNA precipitation method, HepG2 cells were transfected with the *Mix.2* ARE-luciferase reporter (16), FAST-2 [F. Liu *et al.*, *Mol. Cell. Biol.* **19**, 424 (1999)], with or without SARA and Smad2 constructs. Cells were treated with TGF β 2 for 20 hours, and luciferase activity was determined.
 18. X.-H. Feng and R. Derynck, *EMBO J.* **16**, 3912 (1997).
 19. M. Huse, Y.-G. Chen, J. Massagué, J. Kuriyan, *Cell* **96**, 425 (1999).
 20. R. Wieser, J. L. Wrana, J. Massagué, *EMBO J.* **14**, 2199 (1995).
 21. S. Willis, C. M. Zimmerman, L. Li, L. S. Mathews, *Mol. Endocrinol.* **10**, 367 (1996).
 22. X.-H. Feng, E. H. Filvaroff, R. Derynck, *J. Biol. Chem.* **270**, 24237 (1995).
 23. A. A. Russo, P. D. Jeffrey, A. K. Patten, J. Massagué, N. P. Pavletich, *Nature* **382**, 325 (1996).
 24. P. J. Klaulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
 25. A. Nicholls, K. A. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 281 (1991).
 26. We thank H. Lewis for help with the X25 beamline at NSLS, S. Kyin for peptide sequencing and mass spectroscopic analysis, F. Hughson for critically reading the manuscript, L.-A. Swaby for technical support, M. Taga for help in early stages of this project, and E. Steckman for secretarial assistance. Supported by NIH grant CA85171 (Y.S.), Howard Hughes Medical Institutes (J.M.), Medical Research Council of Canada (J.L.W.), the Searle Foundation and the Rita Allen Foundation (Y.S.). Coordinates have been deposited with the Protein Data Bank (accession number 1DEV).

19 October 1999; accepted 19 November 1999

Generating Solitons by Phase Engineering of a Bose-Einstein Condensate

J. Denschlag,¹ J. E. Simsarian,¹ D. L. Feder,^{1,2} Charles W. Clark,¹ L. A. Collins,³ J. Cubizolles,^{1,4} L. Deng,¹ E. W. Hagley,¹ K. Helmerson,¹ W. P. Reinhardt,^{1,5} S. L. Rolston,¹ B. I. Schneider,⁶ W. D. Phillips¹

Quantum phase engineering is demonstrated with two techniques that allow the spatial phase distribution of a Bose-Einstein condensate (BEC) to be written and read out. A quantum state was designed and produced by optically imprinting a phase pattern onto a BEC of sodium atoms, and matter-wave interferometry with spatially resolved imaging was used to analyze the resultant phase distribution. An appropriate phase imprint created solitons, the first experimental realization of this nonlinear phenomenon in a BEC. The subsequent evolution of these excitations was investigated both experimentally and theoretically.

Ultimate control over a physical system can be achieved by precisely manipulating its quantum mechanical wave function, which fully characterizes its state. A BEC of a dilute gas (1) is particularly well suited for such manipulations because of its directly observable wave function: It has many identical atoms in the same quantum state, and it is large enough to be optically imaged.

We demonstrate two optical techniques to prepare and measure the phase of a BEC wave function. A chosen pattern of laser light

imaged onto a condensate shapes its phase almost arbitrarily in two dimensions (2–4). Matter-wave interferometry (5) using optically induced Bragg diffraction (6, 7) is then used to analyze the spatial phase distribution by direct imaging (8). These methods are applied in experimental studies of a phenomenon in nonlinear atom optics (9), the propagation of solitons [solitary waves (10)] in a BEC. Three-dimensional (3D) numerical calculations agree well with and substantiate the experimental observations of soliton generation and propagation. Both reveal the rich dynamics of this nonlinear system, such as the formation of multiple solitons.

Theoretical background. Solitons are stable, localized waves that propagate in a nonlinear medium without spreading. They appear in diverse contexts of science and engineering, such as the dynamics of waves in shallow water (11), transport along DNA and other macromolecules (12), and fiber

¹National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA. ²University of Oxford, Parks Road, Oxford OX1 3PU, UK. ³Theoretical Division, Mail Stop B212, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ⁴Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris, France. ⁵Department of Chemistry, University of Washington, Seattle, WA 98195, USA. ⁶Physics Division, National Science Foundation, Arlington, VA 22230, USA.