 Disorder in a Target for the Smad2 Mad Homology 2 Domain and Its Implications for Binding and Specificity*

P. Andrew Chong‡§§, Barish Oztanam§, Jeffrey L. Wranas§§**, and Julie D. Forman-Kay‡ ‡‡§§

From the §Department of Structural Biology and Biochemistry, Hospital for Sick Children, Toronto M5G 1X8, §Department of Medical Genetics and Microbiology, University of Toronto, Toronto M5S 1A8, and ¶¶Department of Biochemistry, University of Toronto, Toronto M5S 1A8, Canada

The Smad2 Mad homology 2 (MH2) domain binds to a diverse group of proteins which do not share a common sequence motif. We have used NMR to investigate the structure of one of these interacting proteins, the Smad binding domain (SBD) of Smad anchor for receptor activation (SARA). Our results indicate that the unbound SBD is highly disordered and forms no stable secondary or tertiary structures. Additionally we have used fluorescence binding studies to study the interaction between the MH2 domain and SBD and find that no region of the SBD dominates the interaction between the MH2 and the SBD. Our results are consistent with a series of hydrophobic patches on the MH2 that are able to recognize disordered regions of proteins. These findings elucidate a mechanism by which a single domain (MH2) can specifically recognize a diverse set of proteins which are unrelated by sequence, lead to a clearer picture of how MH2 domains function in the transforming growth factor-β-signaling pathway and suggest possible mechanisms for controlling interactions with MH2 domains.

The TGF-β1 superfamily of cytokines plays an essential role in a variety of cellular responses including differentiation, cell fate specification, and growth inhibition (1–4). Dysregulation of TGF-β signaling has been associated with many diseases, such as human cancers, fibrosis, hereditary hemorrhagic telangiectasia (2, 5), and Marfan syndrome (6). TGF-β signal transduction is mediated intracellularly by the Smad proteins, including the R-Smads (receptor-regulated Smads), the I-Smads (inhibitor Smads), and the Co-Smad (common Smad). Ligand binding to the TGF-β receptor complex or a homologous complex leads to receptor phosphorylation of the R-Smads (7, 8). Subsequently, the phosphorylated R-Smads can interact with the Co-Smad, Smad4, and accumulate in the nucleus (9, 10). In the nucleus the Smad proteins bind to transcription factors and promoter regions and play a role in the transcription of various genes (11, 12).

The R-Smads and Smad4 consist of a conserved N-terminal Mad homology 1 (MH1) domain and a conserved C-terminal MH2 domain joined by a poorly conserved linker region. The MH1 domain is known to function in binding DNA (13). The MH2 domain appears to exhibit many different roles and binds to many different proteins. Although many modular protein domains involved in signal transduction, such as SH3, SH2, WW, PTB, and PDZ domains, interact with a recognizable sequence motif (14), MH2 domains appear to be able to bind to a wide variety of ligands that do not share a single common motif. For example the MH2 domain of Smad2 binds a plethora of non-homologous proteins including SARA (15), the TGF-β receptors, FoxH1 (16), Mixer (17), TGIF (18, 19), CBP (20), AML1 (21, 22), Ski (23, 24), and SIP1 (25). These interacting partners represent a diverse group of protein types, encompassing receptors, membrane anchoring proteins, and transcription factors. Furthermore, they have no region of sequence or structural similarity in common. A key question is how the MH2 domain functions in such a versatile manner, i.e. what properties allow the MH2 domain to maintain specificity despite being able to recognize a diverse group of proteins.

To examine the unique properties of the MH2, we have focused on the interaction between Smad2 and SARA because it is the best characterized. SARA is an anchoring protein that specifically binds to Smad2 and 3, transporting them to the receptor complex and increasing the efficiency of phosphorylation by the receptors (15). Crystallographic studies of Smad2 and Smad3 MH2 bound to the 60-residue SBD of SARA have provided evidence for an extensive hydrophobic interaction surface on the MH2 (26, 27). The bound SBD wraps around the MH2, forming an extended structure comprised of a proline-rich rigid coil region, an α-helix, and a β-strand. These three structural elements bind extensive hydrophobic grooves on the MH2 covering ~2600 Å².

The extended structure of the SBD in complex with MH2 suggests that the unbound SBD is disordered in solution or that it undergoes an unfolding transition before interaction with the MH2. Although the prevailing paradigm had been that proteins must form folded structures to function, the relevance of protein disorder in biological systems has been recognized recently (28). Using a neural network program to identify disordered regions on the basis of protein primary sequence, 35–51% of eukaryotic proteins are estimated to have disordered regions of 40 residues or more (29). Disordered protein regions are thought to play roles in the regulation of
many protein-protein and protein-DNA interactions. In many cases disordered or partially ordered proteins undergo folding transitions during target recognition. Examples include GCN4 (30), LEF1 (31), and bacteriophage A protein N (32). In another case, folding induced by phosphorylation is thought to control the interaction between the kinase-inducible activation domain of CREB and the KIX domain of CREB-binding protein (33). Disordered regions have been postulated to function in thermodynamic fine-tuning, plasticity, and control of protein interactions (28).

To investigate the structural requirements of a MH2-interacting partner, we have examined the structure of the unbound SBD of SARA, compared it to the bound form, and probed the binding of MH2 and SBD. NMR analysis indicates that the unbound SBD is highly disordered, allowing the SBD to make extensive contacts with the MH2. Individual contacts provide low binding affinity; however, together these interactions provide a substantial free energy of interaction. No region of the SBD appears to make a dominant contribution to the interaction between the SBD and MH2. These results indicate that the disorder of the SBD facilitates an interaction with the MH2 and suggest a binding mechanism that enables the MH2 to recognize a diverse group of disordered regions.

EXPERIMENTAL PROCEDURES

Disorder Prediction—Disorder predictions were made using the neural network program PONDR (Predictor Of Native Disordered Regions; Molecular Kinetics), which predicts probable disordered regions of proteins based on their primary sequence (34, 35). Specifically we used the VLXT predictor (www.pondr.com/background.html).

Protein and Peptide Preparation—Initially SBD was purified using a non-denaturing purification protocol. SARA-SBD (amino acids 663–751) was amplified using PCR and cloned into pPSETB, a His-tag vector. The N-terminal and XhoI restriction sites were not included in this construct, as these residues are not involved confirmed by sequencing. The last 30 residues of the previous construct were amplified using PCR, cloned into pGEX 4T1 (Amersham Biosciences) using the BamH1 and XhoI restriction sites, and purified on glutathione-Sepharose. SBD was expressed in E. coli and purified on a phenyl-Sepharose column (Amersham Biosciences) and eluted from the column using 50 mM NaPO4, pH 7.4, 5% ethanol. The protein was concentrated using macroSep centrifugal concentrators (Pall Filtron). Protein quantification was carried out using the method of Gill and von Hippel (36).

Peptides used in binding studies correspond to the SBD rigid region (SSQSPNPNNPAPYEY, α-helix (STIPL4QQAASGL), and β-strand (SSPPPTTVMPVGY)) regions were synthesized using standard Fmoc (N-(9-fluorenylmethoxycarbonyl) chemistry and purified using a C4 HPLC column (Phenomenex). Peptide identity was confirmed by mass spectrometry. NMR spectra were processed and analyzed using nmrPipe/nmrDraw (37), NMRView (38), and PIPP (39) software. HNCO (40), HNCACB (41), HBCBCA(CONH) (42), HACAN (43), and COC-TOSY (44) experiments were used to assign the 1H, 15N, and 13C NMR resonances. The collected fractions corresponding to proline isomers was completed by traditional backbone assignment methods (47). The presence of a cis peptide bond was confirmed by the proline carbon side-chain resonance frequencies (48, 49). Proline isomer ratios were measured by comparing well resolved peak volumes in the HBCBCA(CONH) and HSQC using PIPP. We did not account for possible differences in magnetization transfer rates in these experiments. The hydrodynamic radius of SBD was measured using pulsed field gradient diffusion experiments (50, 51). The diffusion rate of the SBD was measured and compared with an internal standard, dioxane, a molecule of known size, to determine the hydrodynamic radius of the SBD.

Fluorescence Binding Studies—Fluorescence binding experiments were carried out at 16°C using an AVIV Ratio spectrophotometer model ATP105 and a Microlab 500 Series automated titrator. Smad2 MH2 protein was extensively dialyzed into a filtered and argon-purged buffer containing 50 mM NaPO4, pH 7.4, 100 mM NaCl, and 5% ethanol. Fluorescence levels were measured as the bound ligand was added. The dissociation constant, F, is the fluorescence, and F_max is the fluorescence maximum.

GST Pull-down Assays—GST fusions with truncated and mutant SBDs were amplified using PCR, cloned into pGEK 4T1 (Amersham Biosciences) using the BamH1 and XhoI restriction sites and transformed into BL21(DE3) E. coli DH5α (Amersham Biosciences) using the BamH1 and XhoI restriction sites, and transformed into BL21(DE3) E. coli DH5α (Amersham Biosciences). The Magic Plasmid containing 15N RNA genes for codons that are rare in E. coli. After induction with 1 mM isopropyl-β-D-thiogalactoside, the protein was expressed for 16 h at 25°C, harvested, and purified immediately. Cells were lysed in 25 mM MES, pH 6.0, 5 mM dithiothreitol with Complete EDTA-free protease inhibitors (Roche Applied Science) and purified on Mono S fast protein liquid chromatography column (Amersham Biosciences) using a NaCl gradient. The collected fraction was then purified on a phenyl-Sepharose column (Amersham Biosciences) and eluted from the column using 50 mM NaPO4, pH 7.4, 5% ethanol. The protein was concentrated using macroSep centrifugal concentrators (Pall Filtron). Protein quantification was carried out using the method of Gill and von Hippel (36).

Fluorescence binding experiments were carried out at 16°C using an AVIV Ratio spectrophotometer model ATP105 and a MicroLab 500 Series automated titrator. Smad2 MH2 protein was extensively dialyzed into a filtered and argon-purged buffer containing 50 mM NaPO4, pH 7.4, 100 mM NaCl, and 5% ethanol. Fluorescence levels were measured as the bound ligand was added. The dissociation constant, F, is the fluorescence, and F_max is the fluorescence maximum.

GST Pull-down Assays—GST fusions with truncated and mutant SBDs were amplified using PCR, cloned into pGEK 4T1 (Amersham Biosciences) using the BamH1 and XhoI restriction sites, and transformed into BL21(DE3) E. coli DH5α (Amersham Biosciences) using the BamH1 and XhoI restriction sites, and transformed into BL21(DE3) E. coli DH5α (Amersham Biosciences). The Magic Plasmid containing 15N RNA genes for codons that are rare in E. coli. After induction with 1 mM isopropyl-β-D-thiogalactoside, the protein was expressed for 16 h at 25°C, harvested, and purified immediately. Cells were lysed in 25 mM MES, pH 6.0, 5 mM dithiothreitol with Complete EDTA-free protease inhibitors (Roche Applied Science) and purified on Mono S fast protein liquid chromatography column (Amersham Biosciences) using a NaCl gradient. The collected fraction was then purified on a phenyl-Sepharose column (Amersham Biosciences) and eluted from the column using 50 mM NaPO4, pH 7.4, 5% ethanol. The protein was concentrated using macroSep centrifugal concentrators (Pall Filtron). Protein quantification was carried out using the method of Gill and von Hippel (36).

Disorder Prediction—Disorder predictions were made using the neural network program PONDR (Predictor Of Native Disordered Regions; Molecular Kinetics), which predicts probable disordered regions of proteins based on their primary sequence (34, 35). Specifically we used the VLXT predictor (www.pondr.com/background.html).

Protein and Peptide Preparation—Initially SBD was purified using a non-denaturing purification protocol. SARA-SBD (amino acids 663–751) was amplified using PCR and cloned into pPSETB, a His-tag vector. The N-terminal and XhoI restriction sites were not included in this construct, as these residues are not involved confirmed by sequencing. The last 30 residues of the previous construct were amplified using PCR, cloned into pGEX 4T1 (Amersham Biosciences) using the BamH1 and XhoI restriction sites, and purified on glutathione-Sepharose. SBD was expressed in E. coli and purified on a phenyl-Sepharose column (Amersham Biosciences) and eluted from the column using 50 mM NaPO4, pH 7.4, 5% ethanol. The protein was concentrated using macroSep centrifugal concentrators (Pall Filtron). Protein quantification was carried out using the method of Gill and von Hippel (36).

Peptides used in binding studies correspond to the SBD rigid region (SSQSPNPNNPAPYEY, α-helix (STIPL4QQAASGL), and β-strand (SSPPPTTVMPVGY)) regions were synthesized using standard Fmoc (N-(9-fluorenylmethoxycarbonyl) chemistry and purified using a C4 HPLC column (Phenomenex). Peptide identity was confirmed by mass spectrometry. NMR spectra were processed and analyzed using nmrPipe/nmrDraw (37), NMRView (38), and PIPP (39) software. HNCO (40), HNCACB (41), HBCBCA(CONH) (42), HACAN (43), and COC-TOSY (44) experiments were used to assign the 1H, 15N, and 13C NMR resonances. The collected fractions corresponding to proline isomers was completed by traditional backbone assignment methods (47). The presence of a cis peptide bond was confirmed by the proline carbon side-chain resonance frequencies (48, 49). Proline isomer ratios were measured by comparing well resolved peak volumes in the HBCBCA(CONH) and HSQC using PIPP. We did not account for possible differences in magnetization transfer rates in these experiments. The hydrodynamic radius of SBD was measured using pulsed field gradient diffusion experiments (50, 51). The diffusion rate of the SBD was measured and compared with an internal standard, dioxane, a molecule of known size, to determine the hydrodynamic radius of the SBD.

Fluorescence Binding Studies—Fluorescence binding experiments were carried out at 16°C using an AVIV Ratio spectrophotometer model ATP105 and a MicroLab 500 Series automated titrator. Smad2 MH2 protein was extensively dialyzed into a filtered and argon-purged buffer containing 50 mM NaPO4, pH 7.4, 100 mM NaCl, and 5% ethanol. Fluorescence levels were measured as the bound ligand was added. The dissociation constant, F, is the fluorescence, and F_max is the fluorescence maximum.

GST Pull-down Assays—GST fusions with truncated and mutant SBDs were amplified using PCR, cloned into pGEK 4T1 (Amersham Biosciences) using the BamH1 and XhoI restriction sites, and transformed into BL21(DE3) E. coli DH5α (Amersham Biosciences) using the BamH1 and XhoI restriction sites, and transformed into BL21(DE3) E. coli DH5α (Amersham Biosciences). The Magic Plasmid containing 15N RNA genes for codons that are rare in E. coli. After induction with 1 mM isopropyl-β-D-thiogalactoside, the protein was expressed for 16 h at 25°C, harvested, and purified immediately. Cells were lysed in 25 mM MES, pH 6.0, 5 mM dithiothreitol with Complete EDTA-free protease inhibitors (Roche Applied Science) and purified on Mono S fast protein liquid chromatography column (Amersham Biosciences) using a NaCl gradient. The collected fraction was then purified on a phenyl-Sepharose column (Amersham Biosciences) and eluted from the column using 50 mM NaPO4, pH 7.4, 5% ethanol. The protein was concentrated using macroSep centrifugal concentrators (Pall Filtron). Protein quantification was carried out using the method of Gill and von Hippel (36).
Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 25 mM NaF, 10 mM sodium pyrophosphate, 1 mM phenylmethylsulfonyl fluoride, 1 mM Na3VO4, pepstatin (0.01 mg/ml), leupeptin (0.01 mg/ml), antipain (0.01 mg/ml), benzamidine hydrochloride (0.1 mg/ml), soybean trypsin inhibitor (0.1 mg/ml), and 0.5% Triton X-100. After clearing by centrifugation the lysate was incubated with glutathione-Sepharose 4B Resin (Amerham Biosciences) for 1 h, after which the resin was washed with buffer containing 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, and 0.1% Triton. The purified proteins were subsequently incubated with 293T cell lysates expressing FLAG-tagged Smad2 for 1 h then washed with a buffer solution of 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 0.1% Triton. Protein complexes were released from GST resin by boiling in SDS sample buffer.

**RESULTS**

**Analysis of MH2-interacting Regions**—Because many modular binding domains bind to a recognizable sequence motif, we attempted to identify a consensus target motif for the Smad2 MH2 domain. To this end we tried to align the regions that have been shown to be necessary for interaction with Smad2 MH2 from seven different proteins. Because of the unrelated nature of the sequences, no meaningful alignment was possible; inspection of these regions (Fig. 1a) reveals that, although it is possible to identify similarities between pairs of sequences, there is no universal MH2 interaction motif.

One previously identified pairwise similarity is the Smad interaction motif (SIM) (52), shown in red (Fig. 1). The SIM has the sequence PNxxxxxahxxxIPPh (where a is an acidic residue, h is a hydrophobic residue, and x is any residue) and is shared by SARA and a subset of the Mixer family of transcription factors and partially shared by FoxH1. The Mixer SIM, like the SARA SBD, is sufficient for an interaction with the MH2 domains of Smad2 and 3 and likely occupies some of the same binding sites since they compete for binding to MH2. Several of these residues including Pro-672, Tyr-680, Pro-686, and Leu-
687 of SARA were shown to contact the surface of Smad2 MH2 (Fig. 1b). However, Asn-673, Glu-679, Ile-684, and Pro-685 do not contact the MH2 surface. Significantly, Mixer residues Asn-293 and Pro-305, which align with Asn-673 and Pro-685 in SARA, are required for a high affinity interaction with MH2, as mutations at these positions significantly decrease the ability of Mixer to interact with the MH2 (52). These observations indicate that the SIM does not fit the definition of a sequence recognition motif per se, since not all of the residues bind to the MH2. One possibility is that the Mixer SIM and SARA SIM bind in different conformations. Alternatively, they bind in the same conformation, and those residues that do not contact the MH2 surface yet are required for a high affinity interaction are involved in maintaining structural features of the SIM, which make it amenable to interacting with the MH2.

**SARA SBD Is Natively Disordered**—Because the affinity of an interaction between two proteins is dependent on the difference in energy between the bound state and the free state, we decided to investigate the unbound SBD to increase our understanding of the interaction between the MH2 and SARA SBD. Previous studies have shown that the SBD interacts with the MH2 in an extended conformation and does not form its own hydrophobic core. Therefore, the SBD must either undergo an unfolding transition before binding or exist in solution as a disordered region. An analysis of the sequence of the SBD using the program PONDR was used to assess whether the SBD has a propensity to be natively disordered (35). PONDR is a neural network program trained on a data base of disordered proteins and used to predict disordered regions of proteins based on their primary sequence. Disordered regions are correctly identified with 79% accuracy for stretches of 21 amino acids or longer. PONDR gives a convincing, uninterrupted prediction of disorder for the entire SBD (Fig. 1c). Consistent with this, the SBD sequence is of low complexity and is enriched in amino acid types thought to be disorder-promoting including nine Pro, seven Ser, four Ala, and four Gln residues and has a paucity of residues thought to be order-promoting such as Trp, Phe, Ile, and Leu (53).

PONDR was also used to predict the propensity of other Smad2-interacting regions to be disordered (Fig. 1c). Although both Mixer and FoxH1 are expected to bind to the MH2 in an extended conformation similar to the SARA interaction (52), neither Mixer nor FoxH1 SIMs are predicted to have significant lengths of disorder by PONDR. Although the SIMs might still be disordered, these results raise the possibility that they undergo an unfolding transition before binding to the MH2. Significantly, three of the other MH2-interacting regions of the partners TGIF, CBP, and AML1 are predicted to have disordered regions of 20 residues in length or more in the region that interacts with the MH2. Thus, the sequence comparison data and PONDR predictions suggest that the MH2 may have some propensity for binding to natively disordered or unfolded regions of interacting proteins.

NMR is an extremely useful tool for investigating disordered states. Its uses include but are not restricted to determining whether a protein is disordered, probing for elements of secondary structure, detection of long range interactions, and measuring hydrodynamic properties. We have used NMR approaches to determine experimentally whether SARA SBD is disordered and to further probe its structural properties. In disordered or unfolded proteins the amide protons are mostly solvent-exposed and experience a similar chemical environment, unlike amide protons in folded proteins. As a result they resonate in a narrow range between 8.0 and 8.8 ppm. The proton-nitrogen \(^{1}H,^{15}N\) HSQC spectrum of SARA SBD revealed a narrow range of amide proton chemical shifts characteristic of disordered proteins (Fig. 2). Similar results were seen when a longer SBD comprising residues 683–751, which was purified under non-denaturing conditions, was used. Additionally, different buffers and temperatures were used to record spectra of this construct as described under "Experimental Procedures." These data provide strong evidence that the unbound SBD is disordered in solution.

In the crystal structure of the SBD bound to Smad2 MH2, the extended SBD forms three secondary structure elements, a proline-rich rigid coil region, an α-helix, and a β-strand. To determine whether these secondary structure elements are present in the free SBD, we analyzed the Ca and Cβ chemical shifts (Fig. 3). Ca chemical shifts more than 0.7 ppm above random coil values and Cβ chemical shifts more than 0.7 ppm below random coil are indicative of α-helices (54). The opposite holds true for β-strands. Our results, which have been corrected for nearest neighbor effects of alanine and proline (55), demonstrate a complete lack of α-helix in the free SBD, indicating that the α-helix must undergo a folding transition prior to binding to the MH2. In contrast, Ca and Cβ chemical shifts in the region that forms a β-strand when bound to the MH2 are consistent with β-strand values. This region probably samples many different conformations but appears to have some propensity for \(\phi,\psi\) angles characteristic of a β-strand. The Ca and Cβ chemical shifts are, thus, consistent with an extended structure.

Unlike chemical shift data, which reflect the population-weighted average of the ensemble of structures in the disordered state, NOEs provide a probe for specific interactions that may not be present in the majority of conformers. The NOE is a through space transfer of magnetization, which is population-dependent and distance-dependent to the negative sixth power. Consequently the presence of an NOE between two protons indicates that these atoms are close in space (less than 6 Å apart) for a period of time. We used protein deuterated at all positions other than the amide positions, which are exchangeable, to perform NOESY experiments with long NOE transfer times. Deuteration reduces magnetic relaxation rates, allowing one to increase the length of time for NOE transfer and permitting detection of weak NOEs from interactions that may be transient or long range. This methodology has successfully been used to detect interactions in disordered states (56–58). In the case of the SBD we detected NOEs between consecutive residues and residues that were separated by one (i, i + 2) or two (i, i + 3) residues (Fig. 4). Weak NOEs between residues that are separated by two residues are found in the regions of the SBD immediately after the region that is α-helical in the bound state. This suggests that an α-helix or a turn may be transiently formed in this region. We were unable to detect any long range NOEs, indicating an absence of significantly populated secondary and tertiary structures in the free state of the SBD. A comparison of the intensities of the two sequential NOEs, \(\alpha N\) and NN, using a \(^{1}H,^{15}N,^{13}C\)-labeled sample showed that NOEs between amide protons were much weaker than NOEs between the amide proton and the Ca proton of the previous residue. This provides additional evidence for \(\phi,\psi\) angles consistent with an extended structure and a lack of persistent α-helical structure.

To further support our conclusion that the SBD does not have a compact structure, we measured the hydrodynamic radius of the unbound SBD using pulsed field gradient diffusion experiments (50, 51). To determine this we compared diffusion rates for SBD and dioxane, a molecule of known size. We obtained a hydrodynamic radius of 23.6 ± 0.7 Å for the 60-residue SBD. The correlation between hydrodynamic radius (\(R_h\)) and polypeptide chain length (\(N\)) for polypeptides...
denatured in 6 M guanidinium chloride has been reported as $R_h = (2.21 \pm 1.07)N^{0.57} = 0.02$ Å (51), which would give $22.8 \pm 11.0$ Å for a 60-residue protein. For folded polypeptides the hydrodynamic radius is dependent on the shape of the molecule but generally fits the empirical equation $R_h = (4.75 \pm 1.11)N^{0.29} = 0.02$ Å or $15.6 \pm 3.6$ Å (51). Thus, the measured hydrodynamic radius of the SBD is significantly larger than that expected for a folded protein and above the norm for a polypeptide dissolved in 6 M guanidinium chloride. This corroborates the chemical shift and NOE data suggesting that the ensemble of SBD conformations includes many highly extended structures.

Proline Isomer Ratios Indicate Some Conformational Restriction—The peptide bond preceding a proline residue can exist in either the cis or trans conformation. The trans form of the bond is generally slightly lower in energy and, thus, is the preferred orientation. The percentage of peptide bonds in the cis conformation can vary between 6.0 and 37.7% in disordered peptides depending on the identity of the residue preceding the proline (59). In folded proteins the percentage of bonds in the cis conformation ranges from 1.8 to 12.4% because a single conformation is usually adopted that restricts the bond to one or the other isomer. Similarly, in disordered proteins conformational restriction may result in a reduced population of peptide bonds in the cis conformation. Comparison of the populations of cis and trans for the peptide bonds that precede prolines in the SBD with the values seen in random coil peptides suggests a limited form of conformational restriction (Table I). Both the cis and trans conformations can be accessed, but the population in the cis conformation is reduced in a number of cases. In particular, prolines 674 and 677 in the rigid coil region and prolines 701, 706, 713, and 720 in the C-terminal portion of the SBD show conformational restriction. Nonetheless, all of the prolines that we observed were able to access both the cis and trans conformation, confirming a lack of stable structure.

No Region of the SBD Makes a Dominant Contribution to the Interaction with Smad2 MH2—To determine whether a specific region of the SBD dominates the interaction with the MH2, we measured the affinities of peptides corresponding to the proline-rich rigid coil region, the $\alpha$-helix, and the $\beta$-strand
Results provide no evidence for significantly populated secondary or tertiary structures. NOEs were measured from a $^1$H,$^2$N HSQC-NOESY-HSQC of a $^1$H,$^2$N,$^13$C sample of SBD with a mixing time of 600 ms at 800 MHz (A) and a NOESY-HSQC of a $^1$H,$^2$N,$^13$C sample of SBD with a mixing time of 150 ms at 500 MHz (B). Dashed lines represent ambiguous NOE data. No NOEs were observed between amide protons separated by more than three residues in the primary sequence.

### Table I

| Residue | % cis | Context | % cis random coil |
|---------|-------|---------|------------------|
| Pro-672 | 11.1 ± 1.3 | Ser-Pro | 10.3 |
| Pro-674 | 5.1 ± 0.5 | Asn-Pro | 11.6 |
| Pro-677 | 7.0 ± 0.5 | Asn-Pro | 11.6 |
| Pro-699 | 13.5 ± 0.4 | Ser-Pro | 10.3 |
| Pro-701 | 2.5 ± 0.6 | Pro-Pro | 6.0 |
| Pro-706 | 6.1 ± 0.8 | Val-Pro | 10.4 |
| Pro-713 | 6.6 ± 0.2 | His-Pro | 16.5 at pH 3.5 |
| Pro-720 | 7.0 ± 0.6 | Gln-Pro | 11.5 |

### Table II

| Peptide | Residues | $K_d$ (nM) |
|---------|----------|------------|
| Full SBD | 664–721 | (2.4 ± 0.3) $\times 10^{-7}$ |
| Rigid coil | 688–681 | (6.2 ± 0.4) $\times 10^{-4}$ |
| $\alpha$-Helix | 682–696 | (2.0 ± 0.2) $\times 10^{-3}$ |
| $\beta$-Strand | 697–709 | (9.5 ± 0.3) $\times 10^{-4}$ |

### Table III

| Peptide | Sequence | Interaction |
|---------|----------|-------------|
| SBD | SQSQPRPNSQAICTPPLQOOAQASGLSSPPTMVM | + |
| $\Delta$-rigid coil | PASYCSTPPLQOOAQASGLSSPPTMVM | - |
| $\Delta$-helix A | SQSPNPNSAQICTPPLQOOAQASGLSSPPTMVM | - |
| $\Delta$-helix B | SQSPNPNSAQICTPPLQOOAQASGLSSPPTMVM | - |
| $\Delta$-strand | SQSPNPNSAQICTPPLQOOAQASGLSSPPTMVM | - |
| SPA | SQSPNPNSAQICTPPLQOOAQASGLSSPPTMVM | - |
| dSARA-eln | SQSPNPNSAQICTPPLQOOAQASGLSSPPTMVM | + |
| dSARA-eln A | SQSPNPNSAQICTPPLQOOAQASGLSSPPTMVM | + |

### Fig. 4

NOE data for the SARA SBD. Results provide no evidence for significantly populated secondary or tertiary structures. NOEs were measured from a $^1$H,$^2$N HSQC-NOESY-HSQC of a $^1$H,$^2$N,$^13$C sample of SBD with a mixing time of 600 ms at 800 MHz (A) and a NOESY-HSQC of a $^1$H,$^2$N,$^13$C sample of SBD with a mixing time of 150 ms at 500 MHz (B). Dashed lines represent ambiguous NOE data. No NOEs were observed between amide protons separated by more than three residues in the primary sequence.

### Fig. 5

GST pull-down analysis of Smad2-SBD interaction. Purified GST-SBD and mutated SBD were incubated with 293T cell lysates expressing FLAG-tagged Smad2. After washing, protein complexes were released from the GST resin by boiling in SDS sample buffer and separated by SDS-PAGE. Bound FLAG-Smad2 was visualized with anti-FLAG antibody. Representative experiments for the data in Table III are shown. The amount of Smad2 used in each pull-down is shown in the “Smad2 input” lane.
Although the interaction of MH2 with $\Delta\alpha$-helix A and $\Delta\alpha$-helix B was undetectable, we noted that the $\alpha$-helical region of Drosophila SARA is significantly different from its mammalian orthologs. Therefore, we examined additional $\alpha$-helical mutants. dsSARA-chimA and dsSARA-chimB are chimeras in which the human $\alpha$-helical region was substituted with the $\alpha$-helical region from the Drosophila homologue. Because this region is shorter in the Drosophila protein, we made a long (A) and short (B) construct. In both cases the interaction was detectable. This suggests that there is considerable flexibility in the sequences that can interact with the corresponding MH2 surface. To confirm this we made a third construct, $\Delta\alpha$-helix C, in which we substituted a portion of an $\alpha$-helix taken from Aspergillus oryzae ribonuclease T1 (60). We selected this sequence because it is known to form an independently stable $\alpha$-helix. $\Delta\alpha$-Helix C also made a detectable interaction with Smad2 MH2. Altogether, these data suggest that although several of the residues in the wild type $\alpha$-helix do interact with the MH2 (the four residues on the opposite face of the helix do not interact), the interaction is primarily hydrophobic and can presumably be satisfied by a number of different residues. Furthermore, removal of three residues in dsSARA-chimB did not abrogate the interaction with the MH2, indicating that there is sufficient "slack" in the wild type SBD to accommodate removal of these residues or, alternatively, that dsSARA does not form an $\alpha$-helix. As a consequence of the large size of the interaction interface between MH2 and SBD and the lack of inherent structure in the SBD, there appears to be flexibility in the exact primary sequence of the SBD required for binding, at least in the $\alpha$-helical region. The different sequences of other MH2 partners suggest that Smad2 has an ability to accommodate divergent sequences.

**DISCUSSION**

We have shown here that the free SBD is disordered in solution and has very little propensity for any stable secondary structure. NMR-based NOE, chemical shift, and pulsed field gradient diffusion measurements all confirm the disordered and extended nature of the SBD conformational ensemble. Additionally we have found that the rigid coil, $\alpha$-helical and $\beta$-strand regions of the SBD all make significant contributions to the free energy of binding based on fluorescence binding and GST pull-down experiments. Importantly, no region appears to make a dominant contribution.

The MH2 domain of Smad2 recognizes a wide variety of protein ligands that do not share a common sequence motif. Our demonstration that the unbound SBD is disordered and that no region of the SBD energetically dominates the interaction with MH2 suggests that the MH2 domain, with its large binding area, is amenable to interacting with a diverse group of disordered hydrophobic ligands. X-ray crystallography studies have shown that the SBD of SARA binds to the MH2 domain of Smad2 in an extended conformation stretching around the MH2 and contacting a large primarily hydrophobic surface area on the MH2. Protein regions that adopt an extended conformation are potentially able to interact with a larger portion of this surface area than a folded protein containing a similar number of residues. This may give the MH2 a preference for binding disordered regions of proteins. PONDR predictions that CBP, TGF, and AML1 have propensity for disorder and data implying that FoxH1 and Mixer also bind to Smad2 MH2 in an extended conformation support the idea that the MH2 has a large binding site with affinity for multiple disordered ligands.

The idea that a protein can have a surface that is a preferred site for binding multiple ligands is not unique to this study of Smad2 MH2. Delano et al. (61) conducted a study in which they examined protein and peptide interactions with the Fc fragment of human immunoglobulin G (61). Screening a library of random peptides for those that bound to the Fc fragment led to the identification of two peptides, both of which were found to bind the hinge region of the Fc fragment. Additionally, the hinge region interacts with four other proteins that are unrelated by structure or sequence. The hinge region of the Fc fragment was characterized as being an accessible hydrophobic surface area with few sites for polar interactions, which impose geometric constraints. Delano et al. (61) propose that these properties make the hinge region a preferred binding site for a wide variety of different hydrophobic ligands. We propose here that the MH2 domain has similar preferred binding sites across the portion of its surface that recognizes the SBD. These hydrophobic binding patches are able to recognize a wide variety of targets, which do not share a common sequence motif, provided that they have a threshold of appropriately spaced hydrophobic residues. The spacing of hydrophobic residues in the ligand with respect to the spacing of hydrophobic patches on the MH2 may be as important as the specific identity of the residue, thus conferring considerable plasticity on MH2 domain interactions. Consistent with this, we have found that no region of the SBD makes a dominant contribution to the interaction. Rather, the combination of several weak interactions leads to a high affinity interaction. This is also supported by the diverse sequences that bind to the MH2.

The free energy of interaction is the difference between the free energy of the complex and the unbound components, emphasizing the importance of the structures of the free components. Because the SBD binds to the MH2 domain in an extended conformation, it must proceed through an extended or disordered state before binding. If the SBD were natively folded it would have to proceed from a compact folded structure through an extended disordered state before binding the MH2. Thus, whether the SBD is natively folded or disordered, there is a disorder to order transition and a considerable loss of configurational entropy upon binding. However, a natively folded SBD would be less favorable because of the additional energy contribution associated with unfolding the SBD. Thus, the disordered state of the free SBD allows it to interact with a large surface area on the MH2 domain with no energetic penalty associated with an unfolding event.

Further analysis is required to determine whether these principles are also applicable to other MH2 interactions. Certainly the interaction between disordered proteins and the large hydrophobic surfaces on the MH2 surface can account for the exceptionally diverse binding capabilities of the MH2. Disordered ligands are flexible and can potentially access a larger proportion of the surface area of an interacting protein than can a structured protein. There is also more plasticity in the arrangement and identity of particular residues in the sequence of a disordered protein. Coupled with the loose requirement for appropriately spaced hydrophobic residues, this could translate into a large subset of amino acid sequences which can be recognized by the MH2. Future work will involve developing an algorithm to design MH2 binding partners based on this model. Our model also suggests how the affinity of interactions with the MH2 may be controlled. First, the affinity of a particular ligand for the MH2 may depend on the proportion of the MH2 binding areas that are occupied.

Downloaded from http://www.jbc.org/ by guest on July 24, 2018
binding sites on the MH2. This is relevant in the identification of the MH2 binding region in an interacting partner, as the minimum sequence required for a detectable interaction is not necessarily the full MH2 binding region. A second implication of our studies is that interaction with the MH2 may be controlled by the structure of the ligand. Not all of the MH2 ligands are predicted to have extensive natively disordered regions, and thus, an interaction with the MH2 may require an unfolding event, which could represent a layer of control for MH2 interactions. The model also emphasizes that specific residues that do not physically interact with the MH2 may be important because they function in maintaining the structure or disorder of the ligand.

It is noteworthy that SARA binds with high affinity to Smad2 and -3 but not to Smad1, -4, -6, or -7. Despite Smad2 MH2 being able to bind a diverse group of ligands, specificity is maintained. There are proteins that bind to Smad2 but not Smad1 and vice versa. Further work will be required to demonstrate whether MH2 domains from Smad1, -4, -6, or -7 have a similar mechanism of target recognition. It is clear that the complexity of responses to the relatively simple Smad pathway is governed by a large number of interacting proteins. The unique mechanism employed by Smad2 MH2 domains may facilitate exquisite control of binding affinity with a diverse set of targets.

Acknowledgments—We thank Drs. Ranjith Muhandiram and Daiven Yang for help in setting up NMR experiments, Jun Liu for use of the calorimeter, and Alan Davidson for use of the fluorimeter. We thank Dr. Dan Garrett for the program PIPP, which was useful in analyzing our data. Access to PONDR was provided by Molecular Kinetics (P. O. mkinetics@Box 2475 CS, Pullman, WA 99165-2475; E-mail: turbonet.com), under license from the Washington State University Research Foundation (WSU) Foundation. PONDR is copyright © 1999 by the WSU Research Foundation, all rights reserved.

REFERENCES

1. Massague, J. (1998) Annu. Rev. Biochem. 67, 753–791
2. Massague, J., Blain, S. W., and Lo, R. S. (2000) Cell 103, 295–309
3. Whitman, M. (1998) Cytokine Growth Factor Rev. 10, 30–40
4. Weis, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) Nature 347, 575–578
5. Love, J. J. (1999) Biophysical Characterization of the HMGR-1 Box Domain of the Lymphoid Enhancer Binding Factor-1. Ph.D. thesis, University of California, San Diego
6. Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E., and Greenblatt, J. (1998) Mol. Cell 1, 265–275
7. Redakrishnan, I., Perez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) J. Cell Biol. 139, 741–752
8. Romero, P., Obradovic, Z., Kissing, C. R., Vilafranca, J. E., and Dunker, A. K. (1997) Proc. IEEE International Conference on Neural Networks 1, 90–95
9. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Genome Inform. 10, 30–40
10. Gill, S. C., and van Hnippe, H. P. (1999) Anal. Biochem. 276, 319–326
11. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) J. Biomol. NMR 4, 277–295
12. Johnson, B. A., and Blevins, R. A. (1994) J. Biomol. NMR 4, 603–614
13. Garrett, D. S., Powers, R., Groenborn, A. M., and Clere, G. M. (1991) J. Magn. Reson. 93, 214–220
14. Keller, X., Yu, G. Y., and Yamazaki, T. (1994) J. Magn. Reson. A 109, 129–133
15. Wittekind, M., and Mueller, L. (1993) J. Magn. Reson. B 101, 201–205
16. Gresestek, S., and Bax, A. (1992) J. Am. Chem. Soc. 114, 6291–6293
17. Kanesis, Y., Donaldson, L., Muhandiram, D. R., Rotin, D., Forman-Kay, J. D., and Kay, L. E. (2000) J. Biomol. NMR 16, 253–259
18. Logan, T. M., Olejniczak, E. T., Xu, X., and Fesik, S. W. (1993) J. Biomol. NMR 3, 225–231
19. Zhang, O., Forman-Kay, J. D., Shortle, D., and Kay, L. E. (1997) J. Biomol. NMR 9, 181–200
20. Ikura, M., Bax, A., Clere, G. M., and Groenborn, A. M. (1990) J. Am. Chem. Soc. 112, 9020–9022
21. Leopold, M. F., Urbauer, J. L., and Wand, A. J. (1994) Mol. Biotechnol. 2, 61–93
22. Derman, D. E., and Bovier, F. A. (1973) J. Org. Chem. 38, 2379–2383
23. Grathwohl, C., and Wuthrich, K. (1976) Biopolymers 15, 2025–2041
24. Gibbs, S. J., and Johnson, C. S., Jr. (1991) J. Magn. Reson. 93, 395–402
25. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Ham, L. J. (1999) Biochemistry 38, 16424–16431
26. Randall, R. A., Germain, S., Inman, G. J., Bates, P. A., and Hill, C. S. (2002) EMBO J. 21, 145–156
27. Romero, P., Obradovic, Z., Rani, M., Dunker, A. K., and Forman-Kay, J. D. (2002) Proteins 42, 38–48
28. Wishart, D. S., and Sykes, B. D. (1994) J. Biomol. NMR 4, 171–180
29. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) J. Biomol. NMR 5, 67–81
30. Mok, Y. K., Kay, C. M., Kay, L. E., and Forman-Kay, J. D. (1999) J. Mol. Biol. 289, 619–638
31. Crowhurst, K., and Forman-Kay, J. D. (2003) Biochemistry 42, 8867–8895
32. Zhang, O., Kay, L. E., Shortle, D., and Forman-Kay, J. D. (1997) J. Mol. Biol. 272, 9–20
33. Reimar, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) J. Mol. Biol. 279, 449–460
34. Myers, K. J., Smith, J. S., Pace, C. N., and Scholtz, J. M. (1996) J. Mol. Biol. 263, 390–395
35. delano, W., Ultsch, M. H., de Vos, A. M., and Wells, J. A. (2000) Science 287, 1279–1283

40714 Interaction between Smad2 MH2 and SBD
Disorder in a Target for the Smad2 Mad Homology 2 Domain and Its Implications for Binding and Specificity

P. Andrew Chong, Barish Ozdamar, Jeffrey L. Wrana and Julie D. Forman-Kay

J. Biol. Chem. 2004, 279:40707-40714.
doi: 10.1074/jbc.M404375200 originally published online July 1, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M404375200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 60 references, 16 of which can be accessed free at http://www.jbc.org/content/279/39/40707.full.html#ref-list-1