

# Features of a Smad3 MH1-DNA Complex

ROLES OF WATER AND ZINC IN DNA BINDING\*

Received for publication, March 28, 2003, and in revised form, April 9, 2003  
Published, JBC Papers in Press, April 9, 2003, DOI 10.1074/jbc.C300134200

Jijie Chai‡, Jia-Wei Wu‡, Nieng Yan‡, Joan Massagué§, Nikola P. Pavletich¶, and Yigong Shi‡¶

From the ‡Department of Molecular Biology, Princeton University, Lewis Thomas Laboratory, Princeton, New Jersey 08544 and the §Cell Biology Program, ¶Cellular Biochemistry and Biophysics Program, Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

**The Smad family of proteins mediates transforming growth factor- $\beta$  signaling from cell membrane to the nucleus. In the nucleus, Smads serve as transcription factors by directly binding to specific DNA sequences and regulating the expression of ligand-response genes. A previous structural analysis, at 2.8-Å resolution, revealed a novel DNA-binding mode for the Smad MH1 domain but did not allow accurate assignment of the fine features of protein-DNA interactions. The crystal structure of a Smad3 MH1 domain bound to a palindromic DNA sequence, determined at 2.4-Å resolution, reveals a surprisingly important role for water molecules. The asymmetric placement of the DNA-binding motif (a conserved 11-residue  $\beta$ -hairpin) in the major groove of DNA is buttressed by seven well ordered water molecules. These water molecules make specific hydrogen bonds to the DNA bases, the DNA phosphate backbones, and several critical Smad3 residues. In addition, the MH1 domain is found to contain a bound zinc atom using four invariant residues among Smad proteins, three cysteines and one histidine. Removal of the zinc atom results in compromised DNA binding activity. These results define the Smad MH1 domain as a zinc-coordinating module that exhibits unique DNA binding properties.**

TGF- $\beta$ <sup>1</sup> signaling plays an important role in the development and homeostasis of all metazoans (1, 2). TGF- $\beta$  signaling from plasma membrane to the nucleus is mediated by the Smad family of proteins (1, 3–7). The receptor-regulated Smads (R-Smads) include Smad1, -2, -3, -5, and -8, each of which is involved in an agonist-specific signaling pathway. The Co-mediator Smad (Co-Smad), namely Smad4 in vertebrate, is an obligate partner for all R-Smads and participates in signaling by diverse TGF- $\beta$  family members. Upon ligand activation, a specific R-Smad is phosphorylated by a corresponding receptor Ser/Thr kinase. The phosphorylated R-Smad is translocated

into the nucleus and forms a functional signaling complex with Smad4 to modulate expression of the ligand-responsive genes. Essential to their function, both R-Smad and Smad4 exhibit sequence-specific DNA binding activity.

The Smad proteins, about 400–500 amino acids in length, consist of highly conserved MH1 and MH2 domains. The MH2 domain is responsible for the formation of homomeric as well as heteromeric Smad complexes, whereas the MH1 domain exhibits sequence-specific DNA binding activity and negatively regulates the function of the MH2 domain. The DNA binding activity of the MH1 domain is central to Smad-mediated function (8, 9).

Investigation on transcriptional activation by Smad proteins has revealed a complex pattern involving multiple proteins and diverse response promoter elements (10–12). Using a PCR-based random-oligonucleotide selection process, an 8-bp palindromic DNA sequence, 5'-GTCTAGAT-3', was identified as the Smad-binding element, or SBE (13). Extensive DNA mutagenesis verified that this SBE indeed represents the optimal DNA-binding sequence for Smad3 and Smad4.

The crystal structure of a Smad3 MH1 domain bound to a 16-bp DNA containing the SBE sequence was determined at 2.8-Å resolution (14). The Smad MH1 domain appeared to adopt a unique fold, with little resemblance to other known structures. In the MH1 domain, a protruding 11-residue  $\beta$ -hairpin was identified as a novel DNA-binding motif, which makes specific contacts to the DNA bases and the backbone phosphate groups in the major groove (14). However, due to the relatively low resolution, the solvent molecules could not be explicitly considered for specific DNA contacts. Consequently, in the close vicinity of the base-contacting residues in the major groove, there was some electron density that could not be accounted for.

To reveal the fine features of DNA binding by the Smad MH1 domain, we generated better crystals of a Smad3 MH1-DNA complex and solved its structure at 2.4-Å resolution (Table I). In the final refined model, water molecules were found to play an extremely important role, mediating a number of significant contacts between residues in MH1 and specific bases and phosphates in the SBE DNA. Compared with the low resolution structure, 17 more hydrogen bonds are now in place to buttress the specific recognition of the SBE by the MH1 domain. We also identified a bound zinc atom in the MH1 domain. Removal of zinc using EDTA leads to a compromised DNA binding activity. These results define the Smad MH1 domain as a zinc-coordinating module that exhibits unique DNA binding properties.

## EXPERIMENTAL PROCEDURES

*Protein and DNA Preparation*—Recombinant Smad3 MH1, corresponding to residues 1–145, was overexpressed and purified as described (14) except that 100  $\mu$ M zinc chloride was included throughout

\* This work was supported by National Institutes of Health Grant CA82171 (to Y. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1OZJ) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¶ To whom correspondence and requests for materials should be addressed: Dept. of Molecular Biology, Princeton University, Princeton, NJ 08544. Tel.: 609-258-6071; Fax: 609-258-6730; E-mail: yshi@molbio.princeton.edu.

<sup>1</sup> The abbreviations used are: TGF, transforming growth factor; SBE, Smad-binding element; r.m.s.d., root mean square deviation.

purification. The 16-bp DNA (SBE) was synthesized as described previously (14).

**Crystallization and Data Collection**—The preparation and crystallization of a Smad3 MH1-DNA complex were similar to those described earlier (14) except the inclusion of 5 mM spermine in the well buffer. This additive resulted in a significant improvement for the size as well as resistance to radiation damage of the crystals. The crystals are in the monoclinic space group  $P2_1$ , with unit cell dimensions  $a = 45.5 \text{ \AA}$ ,  $b = 60.2 \text{ \AA}$ ,  $c = 71.4 \text{ \AA}$ ,  $\beta = 102.0$  degrees, and contain one complex in the asymmetric unit. Diffraction data were collected using an R-Axis-IV imaging plate detector mounted on a Rigaku 200HB generator at room temperature. The data were processed using Denzo and Scalepack (15).

**Refinement**—Both the unit cell dimensions and space group are the same as those reported earlier (14). Thus the atomic coordinates of the 2.8 Å structure of a Smad3 MH1-SBE complex were directly subjected to refinement using CNS (16) after removal of all water molecules. The refined model was checked against the calculated electron density using the program O (17). The electron density for both the zinc atoms and the well ordered water molecules was excellent after preliminary refinement. A total of 47 water molecules were built into the atomic model over five cycles of refinement, each involving positional as well as individual  $B$  factor refinement. The final refined model contains two molecules of Smad3 MH1 (residues 7–132 and 9–132, respectively), a 15-bp DNA duplex, 47 well ordered water molecules, and two zinc atoms. Residues at the N and C termini have no electron density in the maps, and we presume that these regions are disordered in the crystals.

**Electrophoretic Mobility Shift Assay**—DNA-binding assays were performed as described previously (14). The SBE sequences used were 5'-GTATGCTCT AGACTGAA-3'.

## RESULTS

The Smad MH1 domain contains a novel DNA-binding motif, an 11-residue  $\beta$ -hairpin embedded asymmetrically in the major groove of DNA. The crystals used in the previous study were of poor quality and small size; the native data set was derived from a total of seven small crystals as the crystals were extremely susceptible to radiation damage by the X-rays (14). Consequently, due to the relatively low resolution of 2.8 Å, the previously reported structure does not allow an accurate understanding of the fine features of interactions between DNA and the MH1 domain. There was also some electron density in the MH1 domain that could not be explicitly accounted for. To resolve these ambiguities, we tried to improve the diffraction limit by generating larger crystals. We screened a large number of additives and precipitants and were able to significantly improve crystals of the Smad3 MH1 domain (residues 1–145) bound to a 16-bp SBE sequence using 5 mM spermine as an optimal additive. The structure, determined on the basis of a single crystal, has been refined to a resolution of 2.4 Å (Table I and Fig. 1).

The new structure allowed visualization of three additional DNA bases and four more residues in the MH1 domain. As anticipated, the overall structure is similar to that reported earlier (14). Each of the two MH1 molecules binds identically and independently to the major groove of a 4-bp Smad box, 5'-GTCT-3' or 5'-AGAC-3', making hydrogen bond contacts to the bases and to the phosphodiester backbones of the DNA (Fig. 1). The Smad3 MH1 domain adopts a compact globular fold, with four  $\alpha$ -helices, six short  $\beta$ -strands, and five loops. The MH1 domain employs a  $\beta$ -hairpin, formed by strands B2 and B3, to contact DNA in the major groove. The  $\beta$ -hairpin, protruding outward from the globular MH1 core, is rigidly held in place by networks of hydrogen bonds both within itself and between itself and DNA (Figs. 1 and 2).

Despite an overall similarity, the new structure exhibits two significant and prominent features that are absent in the reported structure (14). A number of well ordered water molecules make extensive networks of hydrogen bonds to both residues in the  $\beta$ -hairpin and the DNA bases and phosphate backbones as well as among each other (Figs. 1B and 2). For each  $\beta$ -hairpin, the associated water molecules make hydrogen

TABLE I

Data collection and statistics from the crystallographic analysis

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

Data set	Native
Resolution (Å)	99.0–2.4
Total observations	68,907
Unique observations	14,360
Average $I/\sigma$ (outer shell)	33.7 (5.3)
Data redundancy (-fold)	4.8
Data coverage (outer shell)	95.9% (82.5%)
$R_{\text{sym}}$ (outer shell)	0.051 (0.166)
Refinement	
Resolution range (Å)	15–2.4
Number of reflections ( $I > 2 \sigma$ )	13,219
$R_{\text{working}}/R_{\text{free}}$	21.2%/27.0%
Number of total atoms	2726
Number of waters	47
Number of zinc atoms	2
r.m.s.d. bond length (Å)	0.005
r.m.s.d. bond angles (degree)	1.265
Average $B$ factor	All atoms Protein atoms All water molecules Water (bound to DNA)
	55.85 56.20 44.56 34.98

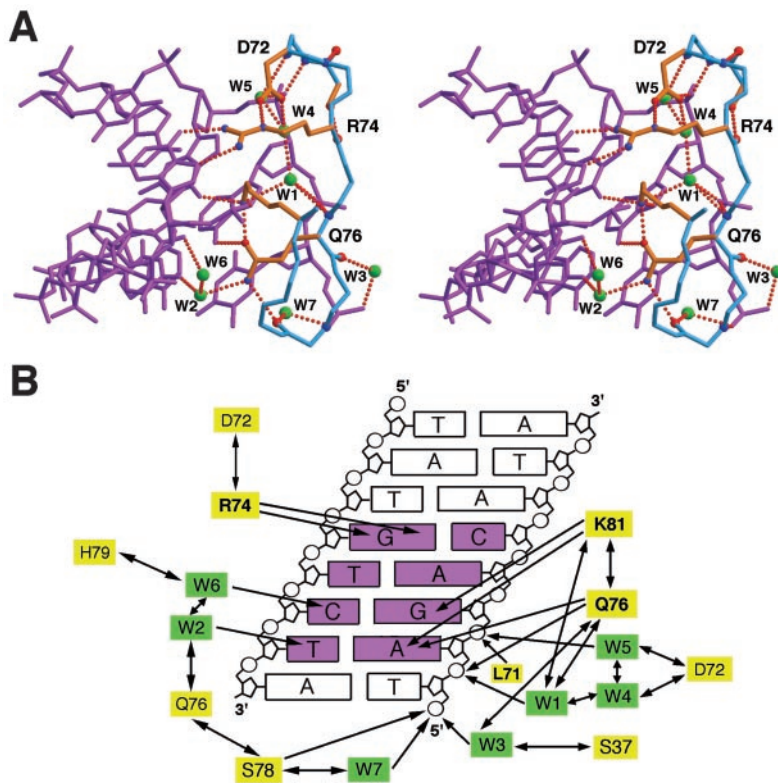
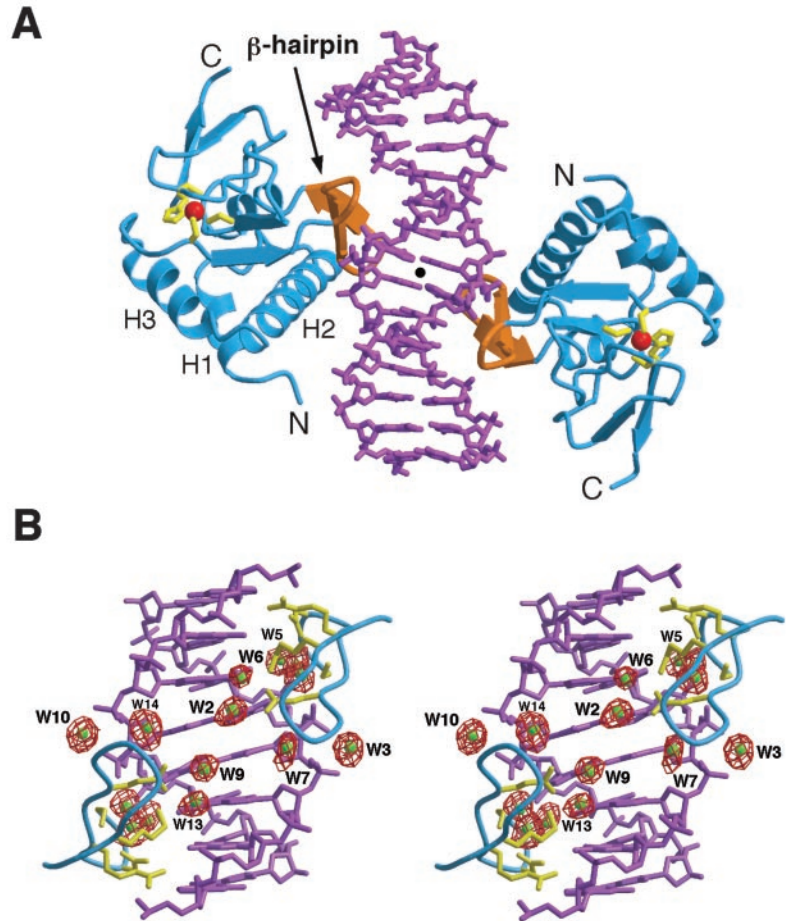
bonds to both strands of the DNA (Fig. 2). Supporting a critical role for these water molecules, the average temperature factor ( $B$  factor) is only 34.98 for the 14 water molecules that are bound to the  $\beta$ -hairpin and DNA (Table I and Fig. 1B). This value is lower than the average  $B$  factor for all solvent molecules, 44.56, or for all protein atoms, 56.20.

These 14 solvent molecules, symmetrically located on the palindromic SBE, make a total of 28 hydrogen bonds to stabilize DNA recognition by the two Smad MH1 domains, with seven water molecules involved in DNA binding by one MH1 domain (Figs. 1B and 2). Among these seven water molecules, w6 and w2 make direct hydrogen bonds to the third and fourth bases of the 5'-GTCT-3' sequence, respectively, whereas w1, w3, w5, and w7 hydrogen bond to the phosphate groups of the DNA backbone (Fig. 2 and Table II). The last water molecule w4 connects with w1, w5, and the side chain carboxylate of residue Asp<sup>72</sup> through hydrogen bonds (Fig. 2 and Table II).

Although the overall DNA-binding mode in this structure is very similar to that reported in the 2.8-Å structure, there are significant differences. The unambiguous and important role of solvent molecules in DNA-binding was not characterized in the 2.8-Å structure. In addition, some residues in the  $\beta$ -hairpin that specifically recognize DNA bases did not appear to have an ideal conformation. For example, the distance between the O $\epsilon$ 1 atom of Gln<sup>76</sup> and the N6 atom of the base A8' was 3.35 Å (14); in this 2.4 Å structure, the distance is refined to 2.98 Å (Table II), which is better suited for a hydrogen bond. More importantly, one additional base-specific hydrogen bond was identified between the N $\zeta$  atom of Lys<sup>81</sup> and the O6 atom of the base G7' (Fig. 2 and Table II).

Another interesting structural feature is that the Smad3 MH1 domain coordinates a bound zinc atom, using three cysteine residues, Cys<sup>64</sup>, Cys<sup>109</sup>, and Cys<sup>121</sup>, and one histidine, His<sup>126</sup>. All four residues are invariant among members of the Smad family as well as their homologues in *Drosophila* and *Caenorhabditis elegans*, suggesting the conserved nature of zinc binding (Fig. 3A). This zinc atom is deeply buried in the MH1 structure; even the four coordinating residues are com-

**FIG. 1. Structural features of the Smad3 MH1 domain bound to SBE.** *A*, overall structure of the Smad3 MH1 domain bound to SBE. The palindromic DNA and the MH1 domain are colored purple and cyan, respectively. The DNA-binding motif is highlighted in orange. The bound zinc atom is shown in red, and its coordinating residues are colored yellow. *B*, a stereo view of the ordered water molecules between the DNA bases and the  $\beta$ -hairpin. The electron density “omit” map on the water molecules, shown in red and labeled, was calculated by omitting these solvents and contoured at  $3\sigma$ . The  $\beta$ -hairpin is shown in cyan, and DNA-binding residues are highlighted in yellow. This figure, Fig. 2A, and Fig. 3B were prepared using MOLSCRIPT (28).



**FIG. 2. Specific DNA contacts mediated by water molecules.** *A*, a stereo view of the specific contacts between the  $\beta$ -hairpin and the DNA. Hydrogen bonds are represented by red dashed lines. Water molecules are shown as green spheres. *B*, a summary of the specific DNA recognition by the  $\beta$ -hairpin and water molecules. Hydrogen bonds are represented by arrows.

pletely solvent inaccessible (Fig. 3B). These observations demonstrate that zinc plays a structural role here and is likely bound with a high affinity.

To assess the role of zinc in the function of the MH1 domain, we used the chelating agent EDTA to remove the zinc ion in the Smad3 MH1 domain and examined the DNA-binding ability of



In a recent sequence and structure-based analysis, the MH1 domain was found to exhibit some homology to the zinc-binding protein I-PpoI, a member of the diverse His-Me finger endonuclease family (27). Since the zinc-coordinating residues in I-PpoI are conserved in the MH1 domain, it was postulated that the Smad MH domain might also contain a zinc atom (27). This prediction is now confirmed by our 2.4-Å resolution structure of the Smad3 MH1 bound to DNA, which shows a zinc atom coordinated by three cysteine and one histidine residues. Interestingly, the previous electron density map at 2.8-Å resolution did not contain significant density at the position of the zinc atom (14). More importantly, chemical analysis of the Smad3 MH1 domain used in the previous study did not reveal the presence of the zinc atoms (data not shown). Nevertheless, removing zinc has a negative impact on the DNA-binding ability of the Smad3 MH1 domain. This finding has practical implications for the study of Smad-DNA interactions *in vitro*. At present, most published gel shift experiments on Smad proteins involve the use of chelating agents such as EDTA. As shown in this study (Fig. 3C), this could complicate the interpretation of results.

## REFERENCES

- Massague, J. (1998) *Annu. Rev. Biochem.* **67**, 753–791
- Moustakas, A., Pardali, K., Gaal, A., and Heldin, C.-H. (2002) *Immunol. Lett.* **82**, 85–91
- Heldin, C.-H., Miyazono, K., and ten Dijke, P. (1997) *Nature* **390**, 465–471
- Wrana, J. L., and Attisano, L. (2000) *Cytokine Growth Factor Rev.* **11**, 5–13
- Miyazono, K., ten Dijke, P., and Heldin, C. H. (2000) *Adv. Immunol.* **75**, 115–157
- ten Dijke, P., Goumans, M.-J., Itoh, F., and Itoh, S. (2002) *J. Cell. Physiol.* **191**, 1–16
- Shi, Y. (2001) *Bioessays* **23**, 223–232
- Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S., and Gauthier, J.-M. (1998) *EMBO J.* **17**, 3091–3100
- Yingling, J. M., Datto, M. B., Wong, C., Frederick, J. P., Liberati, N. T., and Wang, X.-F. (1997) *Mol. Cell. Biol.* **17**, 7019–7028
- Massague, J., and Wotton, D. (2000) *EMBO J.* **19**, 1745–1754
- Attisano, L., and Wrana, J. L. (2000) *Curr. Opin. Cell Biol.* **12**, 235–243
- ten Dijke, P., Miyazono, K., and Heldin, C.-H. (2000) *Trends Biochem. Sci.* **25**, 64–70
- Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B., and Kern, S. E. (1998) *Mol. Cell* **1**, 611–617
- Shi, Y., Wang, Y.-F., Jayaraman, L., Yang, H., Massague, J., and Pavletich, N. P. (1998) *Cell* **94**, 585–594
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- Terwilliger, T. C., and Berendzen, J. (1996) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **52**, 749–757
- Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119
- Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A., and Sigler, P. B. (1994) *Nature* **368**, 469–473
- Joachimiak, A., Haran, T. E., and Sigler, P. B. (1994) *EMBO J.* **13**, 367–372
- Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) *Nature* **335**, 321–329
- Xiao, Z., Liu, X., Henis, Y. I., and Lodish, H. F. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7853–7858
- Xiao, Z., Watson, N., Rodriguez, C., and Lodish, H. F. (2001) *J. Biol. Chem.* **276**, 39404–39410
- Choo, Y., and Klug, A. (1997) *Curr. Opin. Struct. Biol.* **7**, 117–125
- Johnson, K., Kirkpatrick, H., Comer, A., Hoffmann, F. M., and Laughon, A. (1999) *J. Biol. Chem.* **274**, 20709–20716
- Ishida, W., Hamamoto, T., Kusanagi, K., Yagi, K., Kawabata, M., Takehara, K., Sampath, T. K., Kato, M., and Miyazono, K. (2000) *J. Biol. Chem.* **275**, 6075–6079
- Labbe, E., Silvestri, C., Hoodless, P. A., Wrana, J. L., and Attisano, L. (1998) *Mol. Cell* **2**, 109–120
- Grishin, N. V. (2001) *J. Mol. Biol.* **307**, 31–37
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950