Interaction of Smad Complexes with Tripartite DNA-binding Sites*

(Received for publication, March 29, 1999, and in revised form, May 6, 1999)

Kirby Johnson‡, Heidi Kirkpatrick‡, Allen Comer‡, F. Michael Hoffmann§, and Allen Laughon‡‡

From the ‡Laboratory of Genetics, University of Wisconsin-Madison, Madison, Wisconsin 53711, and the §Ophidian Pharmaceuticals, Madison, Wisconsin 53711, and the ¶McArdle Laboratory, University of Wisconsin-Madison, Madison, Wisconsin 53706

The Smad family of transcription factors function as effectors of transforming growth factor-β signaling pathways. Smads form heteromultimers capable of contacting DNA through the amino-terminal MH1 domain. The MH1 domains of Smad3 and Smad4 have been shown to bind to the sequence 5′-GTCT-3′. Here we show that Smad3 and Smad4 complexes can contact three abutting GTCT sequences and that arrays of such sites elevate reporter expression relative to arrays of binding sites containing only two GTCTs. Smad3/4 complexes bound synergistically to probes containing two of the four possible arrangements of three GTCT sequences and showed a correlated ability to synergistically activate transcription through these sites. Purified Smad3 and Smad4 were both able to contact three abutting GTCT sequences and reporter experiments indicated that either protein could mediate contact with all three GTCTs. In contrast, the Smad4 MH1 domain was essential for reporter activation in combination with Smad1. Together, these results show that Smad complexes are flexible in their ability to interact with abutting GTCT triplets. In contrast, Smads have high affinity for only one orientation of abutting GTCT pairs. Functional Smad-binding sites within several native response elements contain degenerate GTCT triplets, suggesting that trimeric Smad-DNA interaction may be relevant in vivo.

The Smad family of transcription factors function as the effectors of signal transduction in response to cytokines of the transforming growth factor-β (TGF-β) superfamily. Through specific associations with pathway restricted receptor complexes, different Smads selectively mediate signaling for certain TGF-β family members; Smad2 and Smad3 permit signaling by TGF-β and activin ligands (1–4) while Smad1, Smad5, and Smad8 are involved in BMP signal transduction (5–9). Phosphorylation of these receptor-associated Smads (r-Smads) permits assembly with Smad4, a common mediator of signaling for TGF-β pathways (10–15). The Smad complexes translocate to the nucleus where they regulate target genes through interactions with DNA and with other nuclear factors (15–18).

Smads contain two conserved domains, MH1 and MH2, separated by a poorly conserved linker. The carboxyl-terminal MH2 domain functions as a transcriptional activator (8, 19) and mediates interactions with receptors (20–22), other Smads (14, 23), and cofactors (11, 24, 25). The amino-terminal MH1 domain inhibits MH2 transactivation (8, 23) and functions as a sequence-specific DNA-binding domain (26–30). Evidence that direct Smad-DNA contact plays a role in transcriptional activation has been reported for Drosophila Dpp response elements controlling expression of vestigial (vg), Ultrabithorax (Ubx), and tinman (tin) (26, 31–33) and for TGF-β/activin response elements within the collagense, PAI-1, JunB, COL7A1, Mix.2, and goosecoid promoters (11, 15, 27, 29, 34–37). Binding site selection experiments identified the inverted repeat 5′-GTCTA-GAC-3′ as an optimal site for GST-Smad3MH1 and GST-Smad4MH1 fusion proteins (28). Structural analysis of the Smad3 MH1 domain bound to this site showed that each GTCT is bound by a single MH1 domain (30). This GTCT motif, termed the “Smad box,” resembles GnCn repeats present in functional Smad-binding sites of native response elements. The occurrence of GnCn sequences in abutting pairs or triplets within native response elements suggests that individual Smad complexes, which appear to exist either as trimers or hexamers (38, 39), may bind adjacent Smad boxes by contact with two or more MH1 domains.

Here we examine how Smad box number and arrangement affect the ability of Smad complexes to bind DNA and activate transcription. The results show that individual Smad complexes are capable of contacting up to three abutting Smad boxes. We also found that contact with three Smad boxes can stabilize binding and enhance transcriptional activation relative to contact with two Smad boxes. While Smads interact efficiently with only one orientation of two Smad boxes, there is surprising flexibility in the ability to interact with various orientations of three Smad boxes. Furthermore, a subset of triplet Smad box arrangements that resemble sequences of some natural Dpp and TGF-β response elements exhibit a high degree of synergism in response to heteromeric Smad complexes. We also show that Smad1/4 complexes, although similar to Smad3/4 complexes in their ability to recognize Smad box sites, are more dependent on Smad4 for DNA contact. Together, these results show how Smad box number and orientation affects direct Smad-DNA contact and provides a framework for investigating the role of direct Smad contact with native response elements.

EXPERIMENTAL PROCEDURES

Plasmid Construction—Reporter elements were introduced into the haploCNSpeR plasmid (40) for expression of β-galactosidase. To build the reporter constructs, pairs of oligos (Operon) were annealed and ligated into the KpnI and NotI restriction sites as three-copy inserts. Each copy contained two arrays of Smad boxes separated by 11 nucleotide pairs. Each Smad box is comprised of the sequence GTCT.

The single copy sequences of each insert are as follows: RL, ACTTGCTTTA-GACTTGGATCTTCTTAGCTTAGCTTGG; RR, ACTTGCTCTAGACTTGGATCTTCTTAGCTTAGCTTGG; LR, ACTTGCTTTA-GACTTGGATCTTCTTAGCTTGG; RLR, ACTTGCTTTA-GACTTGGATCTTCTTAGCTTGG.

This paper is available on line at http://www.jbc.org

* This work was supported by grants from the Cremer scholarship fund (to K. J.) and the National Science Foundation (to A. L.). This is Laboratory of Genetics Paper No. 3540. 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.

‡ To whom correspondence should be addressed. Tel.: 608-262-2456; Fax: 608-262-2976; E-mail: alaunchon@facstaff.wisc.edu.

§ The abbreviations used are: TGF, transforming growth factor; GST, glutathione S-transferase.
AGACGTCATTTGG; RRR, ACTGGCTCTCGTCATTTGGAAATCTAGGTGTATCCTATGTTAGCTTGG; RLL, ACTGCTCAAAGTGTTAATCTCTTATGCTCTAGCTTGG; LRR, ACTTACGAAGCTCTCTCCATTCTATGCTCTAGCTTGG; LmR, ACTGCTTTAGAGCTCTTGG; RlR, ACTTACGAAGCTCTCTCTACAGCTCTAGCTTGG; RML, ACTGCTTACAGTAGCTTGAATCTCATAGACTTGG.

To constitutively express proteins in Drosophila S2 cells we used the ActSC promoter plasmid pPacPL. Smad cDNAs used to generate effector plasmids were generously provided by Y. Zhang, R. Derynck, and J. Massague. To generate Smad3NLc, the Smad3 coding sequence, possessed by a BamHI site immediately adjacent to the initiating AUG, was cloned into the BamHI and KpnI sites of pPaC. The construct Smad3LC is identical to Smad3NLc except 142 amino acids were removed from the MH1 domain and a new BamHI site and ATG were introduced at the EagI site using a pair of oligos. The sequence between the BamHI and EagI sites is 5′-GGATCCATGTCGTCCCCGGCCGGCGG-3′. FLAG-Smad4NLc was built by 1) cloning FLAG-Smad4NLc from pCMV5 into the HindIII and BamHI sites of pBluescript KS+ and then 2) transfer of the resulting HindIII (blunt)-NotI fragment into the EcoRV and NotI sites of pPaC. Removal of the Smad4 DNA-binding domain for construction of FLAG-Smad4LC was by an internal deletion of 573 nucleotides between two XcmI sites within FLAG-Smad4NLc. The 3′ overhangs generated by XcmI cleavage were removed with Klenow prior to religation to preserve the reading frame. The Smad1 coding sequence was introduced into the BamHI and KpnI sites of pPaC to generate Smad1NLc. For the construct Smad1LC, 147 amino acids were removed from the amino terminus and a new ATG was introduced by polymerase chain reaction. Primers for polymerase chain reaction were TTTTCGTCATGTCGTCCCCGGCCGGCGGCGG (5′ EcoRI) and TCAGACGTCTTTGACTTTGCG (3′ KpnI). Activated versions of the ActRR, ActRII(TD), and Thickveins, Tkv(QD) receptors were kindly provided by J. Massague and Y. Chen, respectively. Each was expressed in SF9 cells as His-tagged fusion proteins using a baculovirus expression system. Smad3 was coexpressed with the Alk5 receptor possessing the activating mutation. The Smads were purified to near homogeneity using Talon metal-affinity resin (CLONTECH).

Gel Shifts—Smad3 and Smad4 were individually expressed in SF9 cells as His-tagged fusion proteins using a baculovirus expression system. Smad3 was coexpressed with the Alk5 receptor possessing the activating mutation. The Smads were purified to near homogeneity using Talon metal-affinity resin (CLONTECH).

FIG. 1. Structure of Native Smad-binding Sites—Fig. 1A shows that Smad-binding sites within native TGFB-activin response elements of PAI-1, collagenase, COL7A1, Mix2, goosecoid, and JunB contain abutting perfect (solid arrows) and degenerate (dotted arrows) Smad box sequences in an RRR direct repeat arrangement. Dpp box arrangements primarily as direct repeats. We refer to this direct repeat arrangement as right-right (RR), as shown in Fig. 1B. Among the characterized TGFB-activin targets, the right-left (RL) arrangement corresponding to the GTCTAGACCT sequence identified by binding site selection occurs only once in the goosecoid element, but occurs five times in a Dpp response element of tinman, including one example of a mutation to a non-Smad site. The goosecoid and JunB elements contain examples of three abutting Smad box/degenerate box sequences in an RRR direct repeat arrangement. COL7A1 contains an RR arrangement with an inverted third degenerate box, an arrangement we refer to as left-right-right (LRR) that occurs five times in a Dpp response element of tinman, including one example of a mutation to a non-Smad site. The goosecoid and JunB elements contain examples of three abutting Smad box/degenerate box sequences in an RRR direct repeat arrangement. COL7A1 contains an RR arrangement with an inverted third degenerate box, an arrangement we refer to as left-right-right (LRR). The LRR arrangement also occurs in Mad/Meda-bind sites of the vestigial, Ubx, and tinman Dpp response ele-

20710 Smad-DNA Interaction
FIG. 2. Responsiveness of two-Smad box reporters to Smads depends on Smad box arrangement. Drosophila S2 cells were transfected with reporter constructs and combinations of expression vectors for Smad3 and FLAG-Smad4. A, transfections done in the absence of ActRIB(TD). B, transfections done in the presence of ActRIB(TD). Smad boxes of the RL+1 reporter are separated by a single base pair. For the comparison of RL to RL+1 reporters, both Smad3 and FLAG-Smad4 were included in the transfections (inset). β-Galactosidase activity was measured by colorimetric assay as the rate of color change over time. Results shown represent the average β-galactosidase activity from transfections done in triplicate.

FIG. 3. Synergy between Smad3 and Smad4 exhibited by some three-Smad box reporters. A, two- and three-Smad box reporters were co-transfected with an ActRIB(TD) expression vector and combinations of expression vectors for Smad3 and FLAG-Smad4. B, binding to outside Smad boxes is insufficient for activation of three-Smad box reporters. The sequence of the middle Smad box of the reporters, RLMR, RLmL, and LRmR, was changed from GTCT to ATCT. Reporters were co-transfected with expression constructs for Smad3, FLAG-Smad4, and ActRIB(TD). Results shown represent the average β-galactosidase activity from transfections done in triplicate.

not phosphorylated by receptors. Increasing the spacing between the RL Smad boxes by a single base pair abolished activation (Fig. 2B inset, compare RL to RL+1). In contrast to RL, the RR and LR reporters responded synergistically to co-expression of Smad3 and Smad4 in the presence of ActRIB(TD), although the overall levels of induction were 8- and 20-fold lower than for RL, respectively. The observed failure of RR and LR reporters to respond to Smad3 alone is consistent with previous reports that without Smad4, Smad3 has little effect on the PAI-1, collagenase I, or JunB promoters (15, 29, 34), each of which contains functional RR sites (Fig. 1). Together, these results indicate that the level of Smad-dependent transcriptional activation depends on both the orientation and spacing of Smad boxes. Furthermore, the RR and LR orientations of Smad boxes require both Smad3 and Smad4 for maximal transcriptional activation, while RL is able to respond to Smad3 alone.

All Orientations of Three Abutting Smad Boxes Mediate High Levels of Activation by Smad3 and Smad4—As discussed above, potentially tripartite Smad-binding sites exist in both TGF-β/activin and Dpp response elements. To determine whether interaction with a third Smad box would enhance activation, we constructed a set of reporters in which each binding site consisted of three Smad boxes. Each of the four possible arrangements of three Smad boxes was tested with ActRIB(TD) in the presence or absence of co-transfected Smads (Fig. 3A). Addition of a third Smad box to RR, either as RRR or LRR, elevated reporter expression in response to Smad3 plus...
Smad4 in comparison to the levels detected for RR and LR (e.g. compare Fig. 3A, bar 4, LR, RR, and LRR). For RRR and LRR, coexpression of both proteins produced responses 2- and 5-fold greater than the additive response of Smad3 and Smad4 expressed separately (e.g. compare Fig. 3A, LRR bar 4 to bars 2 and 3). Reporter activation was also elevated by addition of a third Smad box to RL, either as RLR or RLL, although the effect was less dramatic than for RRR and LRR. Like RL, RLR and RLL had a less than additive (synergistic) response to the coexpression of both Smad3 and Smad4 (in Fig. 3A compare bar 4 with bars 2 and 3 for RLR and RLL reporters).

While these data suggest that Smad3/4 complexes are capable of recognizing triple-Smad box sites, it was also possible that in each case the observed levels of reporter activity resulted from interaction with only the two outside Smad boxes. To determine whether the middle Smad box was dispensable for activation we generated a set of reporter constructs in which the sequence of the middle Smad box was changed from GTCT to ATCT (Fig. 3B). The Smad3 MH1 domain has been shown to make specific contacts to the guanine at the first position in the Smad box (30) and mutation to adenine has been shown to disrupt binding of GST-Smad3MH1 or GST-Smad4MH1 in the context of an RL pair (28). We found that the mutated reporter constructs were not induced by co-transfection with Smad3 and Smad4 regardless of the orientation of the outside Smad boxes (Fig. 3B).

Activated Smad3/4 Complexes Bind Cooperatively to Two Arrangements of Triple Smad Boxes—To assess the effects of binding site organization on Smad DNA binding affinity, gel shift assays were performed using whole cell extracts from S2 cells co-transfected with Smad3, FLAG-Smad4, or both Smads. ActRIB/TD was included in each transfection. Unlike the reporters, which had six Smad-binding sites, each gel shift probe contained only a single Smad-binding site consisting of two or three Smad boxes. The RR and LR probes were not bound by Smad3, Smad4, or both proteins in the same extract (Fig. 4A, lanes 2–4 and 6–8). In contrast, the RRR and LRR probes gave rise to weak gel-shift bands with Smad3 or Smad4 alone (Fig. 4A, lanes 16, 17, 20, and 21), and a stronger novel band when both proteins were coexpressed (asterisks in Fig. 4A, lanes 18 and 22). This result provides direct evidence that native Smad3/4 complexes can bind cooperatively to three tandem Smad boxes, and suggests that the correlated effects on reporter activation described above are due in part to differences in DNA binding affinity.

Lighter exposure in the bottom half of Fig. 4A shows that the novel gel-shift band is also specific to Smad3/4 extracts in lanes 12, 26, and 30 that contain RL, RLR, and RRR probes, respectively. This novel band was eliminated by incubation with antibody against Smad3 or against the Flag tag at the amino terminus of Smad4 (Fig. 4A, lanes 13 and 14), evidence that both proteins were contained in this complex. Anti-Flag gave a stronger super-shifted band than anti-Smad3 (compare bands marked “S” in Fig. 4A, lanes 13 and 14), suggesting that anti-Smad3 might have acted primarily to disrupt formation of the Smad3/4 complex. Consistent with this, we failed to detect a super-supershift when both antibodies were included in the binding reaction (data not shown).

Although the stoichiometry of the Smad3/4 gel shift complex is unknown, the ability of the complex to bind RRR and LRR probes but not RR or LR probes suggests it consists of at least three subunits (compare lanes 18 and 22 with lanes 4 and 8). The slower mobility Smad4 band was weaker in the shift of the RRR probe (lane 18) suggesting that RRR has a higher affinity for the Smad3/4 complex than for Smad4 alone, consistent with the synergistic response to coexpression of Smad3 and Smad4 shown in Fig. 3A. A faint band from the control extract migrated at the same position as the Smad3 band and may represent binding by endogenous Drosophila Smads (compare lanes 9 and 10, for example).

The binding activity of Smad3 in whole cell extracts was approximately 50-fold lower than what was observed for Smad4. Western blot analysis (Fig. 4B) showed that this reflected a difference in binding activity since Smad3 protein was only about 2-fold lower in concentration than the Smad4 protein in these extracts. Relative to Smad4, the low DNA binding activity of Smad3 in whole cell extracts might be explained by incomplete activation by co-transfected ActRIB/TD since the DNA binding activity of the Smad3 MH1 domain has been shown to be inhibited in non-activated full-length Smad3 generated by bacterial expression (29). Alternatively, the high DNA binding affinity of baculovirus-derived Smad3 (see below) may have been an artifact of its modified sequence (i.e. it was tagged at the NH2 terminus with hexahistidine) or of abnormal post-translational modification resulting from overexpression.

It is not clear why Smad3/4 complexes migrate at a faster rate than complexes with only Smad3 or Smad4 but this ob-
Smad-DNA Interaction

Fig. 5. Smad3 and Smad4 DNA-binding domains contribute redundantly to target recognition. Reporters were transfected with ActRIB(TD) and combinations of constructs for expression of full-length (NLC) and MH1 deleted (LC) Smads. For the Smad3LC construct 142 amino acids were removed from the amino terminus while FLAG-Smad4LC was generated by an internal deletion of 190 amino acids by removal of the sequence between two XcmI sites. Results shown represent the average β-galactosidase activity from transfections done in triplicate.

Figure 5: Graph showing beta-galactosidase activity for various Smad-DNA interactions.

Cell extracts at concentrations just as high as the full-length proteins. Furthermore, co-transfection of Smad3LC dramatically stimulated the ability of full-length Smad4 to activate transcription (compare Fig. 5, LRR bar 4 versus bar 3 versus bar 7). This potent stimulation is likely to have resulted from increased nuclear translocation and derepression of the MH2 trans-activation function (1, 8, 12). The inability of Smad3LC to activate transcription without coexpressed Smad4 provides additional evidence that reporter activation in response to full-length Smad3 (i.e. without Smad4) results from Smad3-DNA contact. Co-transfection of Smad4LC with Smad3NLC stimulated transcription of reporters containing RL inverted repeat sites (RL and RLR) to levels similar to what was observed for full-length proteins (compare bar 8 with bar 6 for RL and RLR reporters). Thus, for RL and RLR, it appears that the Smad3 MH1 domain is sufficient for DNA contact. Conversely, co-expression of Smad4LC reduced the response of the RR, RRR, LR, and LRR reporters (Fig. 5, compare bar 8 with bar 6 for these reporters), indicating that these arrangements favor DNA contact by Smad4. Nonetheless, the combination of Smad4LC and Smad3NLC still resulted in levels of induction that were slightly higher than those observed for Smad3 alone (compare bar 8 with bar 3), possibly reflecting a weak contribution of the Smad4 MH2 domain toward stabilization of Smad complexes or to transcriptional activation. The failure of Smad4LC to enhance the activity of Smad3NLC is consistent with the ability of activated r-Smads to translocate to the nucleus independently of Smad4(12), but also fits with evidence presented here that homomeric complexes of Smad3 are capable of activating transcription, particularly through recognition of RL-binding sites.

The failure of Smad3LC and Smad4LC to activate reporters demonstrates that DNA contact is mediated by MH1 domains of the exogenously expressed human Smads in this experimental system. The dramatic enhancement of reporter activation by co-transfection of Smad3LC with Smad4NLC confirms that Smad3LC is stable and capable of activation when complexed with Smad4. While the MH1 domain of either Smad3 or Smad4 is sufficient for recognition of any of the reporters, the RR, RRR, and LRR sites exhibit a stronger dependence on the Smad4 MH1 than do the RL, RLR, and RLL sites. This suggests that Smad4 may bind or activate RR, RRR, and LRR more efficiently than Smad3.

A gel shift assay using whole cell extracts of transfected S2 cells was used to determine whether Smad4LC forms a DNA-bound complex in combination with Smad3 (Fig. 6). Smad4LC alone failed to form a complex with the RL DNA probe (lane 5). However, Smad4LC combined with Smad3 to form a complex intermediate in mobility between a complex containing Smad3 alone and one containing Smad3 plus full-length Smad4 (lane 9, 3'4'LC arrow, compare with bands in lanes 2 and 7). This Smad3-Smad4LC complex was super-shifted to a slower mobility band by inclusion of anti-Smad4 antibody in the binding reaction (compare lane 10 with lane 9 of Fig. 6). It is not clear why the Smad3/LC complex migrates faster than complexes containing Smad3 or Smad4 alone, or why removal of the Smad4 MH1 domain decreases the mobility of this complex.

Smad1 Prefers Binding Sites with Three Smad Boxes—Smad1 mediates BMP-specific biological responses that differ from the TGF-β/activin responses regulated by Smad2 and Smad3 (4). Nevertheless, Smad1 β-hairpin DNA contact residues are identical to those of Smad3, and the MH1 domains of both proteins bind to a GTCT probe with similar affinity (30). Similarity in Smad1 and Smad3 DNA binding specificity is also suggested by the responsiveness of a multimerized JunB promoter sequence, 4X(CAGACAGT), to both TGF-β and BMP2.
(34), although DNA contact may be mediated by Smad4, as discussed below.

We tested the ability of Smad1 to activate the Smad box reporters in the presence or absence of Smad4. An activated form of the Drosophila BMP Type 1 receptor homolog Thickveins, Tkv(QD), was included in transfections to promote phosphorylation of Smad1. Overall, the results obtained from coexpression of Smad1 with Smad4 were similar to the results from coexpression of Smad3 with Smad4 (Fig. 7A). In both instances there was a preference for binding sites possessing the RL arrangement and a synergistic activation of the RRR and LRR reporters. The LRR reporter exhibited an 8-fold synergism when co-transfected with Smad1 and Smad4 (compare bar 4 with bars 2 and 3 for LRR) and also yielded the highest levels of activity with a 50-fold induction over the no-Smad control. In the absence of Tkv(QD), reporter expression was dramatically reduced for all reporters (Fig. 7B and data not shown).

We also examined the requirement for DNA binding by Smad1 using an MH1-deleted construct, Smad1LC. When full-length Smad1 was co-transfected with Smad4LC and the RL reporter, a 15-fold reduction in activity was observed compared with transfections done with full-length Smad1 and Smad4 (Fig. 7C). This level of activation was lower than the level observed for Smad1 alone (Fig. 7A, RL bar 2), indicating that Smad4LC inhibited Smad1, possibly by forming complexes in which Smad4 fails to provide DNA contact. Conversely, co-transfection of Smad1 LC with full-length Smad4 resulted in a 6-fold elevation above the level of reporter expression obtained using both full-length proteins. Thus, while Smad1 is capable of activating through RL sites without Smad4, removal of its MH1 domain dramatically enhances its synergy with Smad4, possibly because an inhibitory MH1-MH2 interaction has been eliminated (23).

The ability of Smad1 to bind DNA was tested by gel shift using whole cell extracts from S2 cells. Figure 8 shows that Smad1 bound the RL probe weakly when expressed alone (lane 1). Like Smad3, coexpression of Smad1 with Smad4 resulted in a novel, faster migrating band (asterisk in lane 3) that was super-shifted by antibodies specific for either Smad1 or Smad4 (S, T, and data not shown). This level of activation was lower than the level of Smad box number and arrangement are important but not on that of Smad3, an indication that the Smad4 MH1 domain is the more effective partner at contacting RR, RR, and LRR sites. Like Smad3, Smad1 activates the RR, RRR, and LRR reporters synergistically in combination with Smad4. However, unlike Smad3, Smad1 appears to be inhibited by Smad4LC, perhaps an indication that such complexes are un-
able to contact DNA through Smad1. Thus, while Smad1 and Smad3 appear to have very similar preferences for Smad box organization, oligomerization with Smad4 appears to severely diminish the ability of Smad1 to contribute to DNA contact.

Although the crystal structure of the Smad4 MH2 domain reveals that it forms a trimer (38), the oligomerization state of the native Smad DNA binding complex has not been completely resolved. One model proposes that Smad homotrimers interact to form heterohexamers (38) while another model favors formation of heterotrimers of Smad4 and receptor associated Smads (39). Consistent with either model is genetic evidence for nonredundant function of Sma2, Sma3, and Sma4 in regulating C. elegans body size (44), and evidence that Smad2, Smad3, and Smad4 can form a single, functional complex (45). Here we demonstrate that Smad complexes can synergistically interact in vivo and in vitro with three GTCT sequences in RRR and LRR sites, indicating that these complexes contain a minimum of three subunits. Taken alone, this finding is consistent with either the heterohexamer or heterotrimer models. However, this synergism is retained in vivo when the Sma3 MH1 domain is deleted, indicating that RRR and LRR can be activated by Smad3Lc-Smad4 complexes containing at least three Smad4 subunits. This observation is consistent with the heterohexamer model but not with the trimer model. An alternative explanation is that Smad3Lc stimulates reporter activation solely by increasing the nuclear localization of Smad4, and that once inside the nucleus Smad4 can activate independently of Smad3. This explanation is consistent with the apparent ability of homomeric Smad3 and Smad4 complexes to contact all three GTCT sites of RRR and LRR probes in gel shift experiments.

How do the effects of Smad-binding site organization on Smad binding affinity and synergism relate to the sequences of native TGF-β, activin, BMP, or Dpp response elements? As shown in Fig. 1, the majority of functional Smad-binding sites within TGF-β/activin response elements appear to be composed of RR arrangements of GTCT or GTCT-like sequences, while the Dpp response elements of vg and Ubx contain RL, LR, and LRR arrangements. Our results show that, while less responsive than RL, RLL, or RLR to elevated levels of Smad3, Smad4, or Smad1, the RR, RRR, and LRR configurations exhibit a higher degree of synergism in response to coexpression of Smad4 with Smad3 or with Smad1. Thus RR, RRR, and LRR sites may perform some special role in responding specifically to the signaling-dependent formation of Smad3/4 or Smad1/4 complexes.

Smad-cofactor interactions are known to play an important role in the activation of some Smad targets (11, 15, 24, 36, 46). A single strong RL Smad-binding site is necessary but insufficient for activation when the cofactor FAST-1 is bound to a nearby site (28, 47). Activation of the mouse goosecoid promoter by Smad2/4 and the cofactor FAST-2 has been shown to be dependent upon a FAST-2-binding site and nearby sites for Smad4 (36). The two regions footprinted by Smad4 contain GC-rich RRR, RR, and RL sites. Other TGF-β/activin/Dpp responsive promoters have been shown to depend on non-RL Smad-binding sites and nearby cofactor-binding sites (15, 32, 48, 49). It remains to be determined how Smad-binding site arrangement affects regulation within the context of these natural elements. In addition to the inherent differences in Smad binding affinity and synergism described here, Smad-binding site organization may be important for the topology of Smad-cofactor interactions. Cofactor interactions with the MH1 domain, e.g. the c-Jun-Smad3 MH1 interaction (15), are likely to be especially sensitive to MH1 orientation imposed by contact with DNA.

While it is remarkable that Smads are able to bind DNA through contact with three consecutive Smad boxes, there is precedent for such a trimeric interaction with DNA. Trimers of the heat shock factor are capable of stable contact with either two or three abutting repeats of the 5-base pair monomer-binding site for this inducible activator (50, 51). A difference between the two is that HSF has similar affinity for pairs of binding sites that are inverted in either orientation, whereas Smads interact efficiently only with the RL configuration. A flexible linker also connects the POU domain and homeodomain of POU transcription factors, resulting in an analogous degeneracy in binding site recognition that allows POU protein function in a variety of distinct binding site and cofactor contexts (52). While the present study shows that Smads are capable of flexible interaction with tripartite binding sites, it will be important to determine whether, like POU proteins, such flexibility is significant in native regulatory contexts.

Acknowledgments—We thank Y. Chen, Y. Zhang, R. Derynck, and J. Massague for plasmids, K. Rashka for assistance with cell culture, and T. Lockie, J. Wolff, and the Weisman Center Vector Laboratory for providing transfection reagent. We also thank G. Halder for technical advice, S. Callaci for help with improvement of the gel shift assay, and Sean Carroll for suggestions that improved this manuscript.

REFERENCES

1. Baker, J. C., and Harland, R. M. (1996) Genes Dev. 10, 1880–1889
2. Zhang, Y., Feng, X., Wu, K., and Derynck, R. (1996) Nature 383, 168–172
3. Eppert, K., Scheer, S. W., Ozcelik, H., Preece, R., Hoodless, P., Kim, H., Yeui, L. C., Bapat, B., Gallinger, S., Andrusiuk, I. L., Thomsen, G. H., Wrrana, J. L., and Attisano, L. (1996) Cell 86, 543–552
4. Graff, J. M., Bansal, A., and Melton, D. A. (1996) Cell 85, 479–487
5. Hoodless, P. A., Huey, T., Abdallah, S., Stapleton, M., O’Connor, M. B., Attisano, L., and Wrana, J. L. (1996) Cell 85, 489–500
6. Nakayama, T., Snyder, M. A., Grewal, S. S., Tsuneizumi, K., Tabata, T., and Christian, J. L. (1998) Development 125, 857–867
7. Tamaki, K., Souchelnytskyi, S., Itoh, S., Nakao, A., Sampath, K., Heldin, C., and Tendikij, P. (1998) J. Cell. Biol. 177, 355–363
8. Liu, F., Hata, A., Dodd, J., Carcione, J., Harland, R. M., and Massague, J. (1996) Nature 381, 620–623
9. Thomsen, G. H. (1996) Development 122, 2359–2366
10. Lagurska, G., Hata, A., Hemmati-Brivanlou, A., and Massague, J. (1996) Nature 383, 832–836
11. Liu, F., Pouponnot, C., and Massague, J. (1997) Genes Dev. 11, 3157–3167
12. Zhang, Y., Musci, T., and Derynck, R. (1997) Curr. Biol. 7, 270–276
13. Candia, A. F., Watabe, T., Hawley, S. B., Onichtchouk, D., Zhang, Y., Derynck, J. C.
Smad-DNA Interaction

R., Niehrs, C., and Cho, K. Y. (1997) Development 124, 4467–4480
14. Wu, R. Y., Zhang, Y., Feng, X. H., and Derynck, R. (1997) Mol. Cell. Biol. 17, 2521–2528
15. Zhang, Y., Feng, X., and Derynck, R. (1998) Nature 394, 909–913
16. Massague, J. (1998) Annu. Rev. Biochem. 67, 753–791
17. Padgett, R. W., Das, P., and Krishna, S. (1998) Bioessays 20, 382–390
18. Whitman, M. (1996) Genes Dev. 10, 2445–2462
19. Meersseman, G., Verschueren, K., Nelles, L., Blumenstock, C., Kraft, H., Wuytens, Remacle, J., Kozak, C. A., Tylzanowski, P., Niehrs, C., and Huybreck, D. (1997) Mech. Dev. 61, 127–140
20. Kretsing, M., Liu, F., Hata, A., Doody, J., and Massague, J. (1997) Genes Dev. 11, 984–995
21. Tsukazaki, T., Chiang, T., Davison, A., Attisano, L., and Wrana, J. (1998) Cell 95, 779–791
22. Chen, Y. G., Hata, A., Lo, R. S., Wotton, D., Shi, Y., Pavletich, N., and Massague, J. (1998) Genes Dev. 12, 2144–2152
23. Shi, Y., Hata, A., Lo, R. S., Massague, J., and Pavletich, N. P. (1997) Nature 388, 87–93
24. Feng, X. H., Zhang, Y., Wu, R. Y., and Derynck, R. (1998) Genes Dev. 12, 2153–2163
25. Kim, J., Johnson, K., Chen, H. J., Carroll, S., and Laughon, A. (1997) Nature 388, 304–308
26. 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
27. Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S. B., Kinzler, K. W., and Vogelstein, B. (1998) Mol. Cell 1, 611–617
28. Henninger, J. M., and Whitman, M. (1996) Nature 383, 691–696
29. Zhou, S., Zawel, L., Meng, C., Kinzler, K. W., and Vogelstein, B. (1998) Mol. Cell 12, 121–127