Published online 30 June 2011 Nucleic Acids Research, 2011, Vol. 39, No. 18 8213–8222
doi:10.1093/nar/gkr500

Structural basis for the cooperative DNA recognition by Smad4 MH1 dimers

Nithya Baburajendran[1,2], Ralf Jauch[1,*], Clara Yueh Zhen Tan[1], Kamesh Narasimhan[1,2] and Prasanna R. Kolatkar[1,2,*]

[1]Laboratory for Structural Biochemistry, Genome Institute of Singapore, Singapore-138672 and [2]Department of Biological Sciences, National University of Singapore, Singapore-117543

Received March 21, 2011; Revised and Accepted May 31, 2011

ABSTRACT

Smad proteins form multimeric complexes consisting of the 'common partner' Smad4 and receptor regulated R-Smads on clustered DNA binding sites. Deciphering how pathway specific Smad complexes multimerize on DNA to regulate gene expression is critical for a better understanding of the *cis*-regulatory logic of TGF-β and BMP signaling. To this end, we solved the crystal structure of the dimeric Smad4 MH1 domain bound to a palindromic Smad binding element. Surprisingly, the Smad4 MH1 forms a constitutive dimer on the SBE DNA without exhibiting any direct protein–protein interactions suggesting a DNA mediated indirect readout mechanism. However, the R-Smads Smad1, Smad2 and Smad3 homodimerize with substantially decreased efficiency despite pronounced structural similarities to Smad4. Therefore, intricate variations in the DNA structure induced by different Smads and/or variant energetic profiles likely contribute to their propensity to dimerize on DNA. Indeed, competitive binding assays revealed that the Smad4/R-Smad heterodimers predominate under equilibrium conditions while R-Smad homodimers are least favored. Together, we present the structural basis for DNA recognition by Smad4 and demonstrate that Smad4 constitutively homo- and heterodimerizes on DNA in contrast to its R-Smad partner proteins by a mechanism independent of direct protein contacts.

INTRODUCTION

Smads (homologs of Sma and MAD proteins in Drosophila) are transcription factor proteins that regulate gene expression in response to TGF-β signaling (1). Mammalian genomes encode eight different Smad family members (Smad 1–8) which are sub-divided into three functional classes (2). Smad1, Smad2, Smad3, Smad5 and Smad8 are activated via serine/threonine phosphorylation by the TGF-β class of receptors and were thus termed receptor regulated Smads (R-Smads). R-Smads are further subdivided based on the extracellular ligands they are responding to: Smads 2 and 3 are predominantly, but not exclusively, activated by TGF-β, activin and nodal signaling (henceforth termed TGF-β Smads) whereas the Smads 1, 5 and 8 are mostly activated by bone morphogenetic proteins (BMP) and anti-mullerin hormones (henceforth termed BMP-Smads) (3–6). Smad4/DPC4 (Deleted in Pancreatic Carcinoma locus 4) acts as common partner Smad (co-Smad) since it constitutes a critical component in both, the BMP and TGF-β mediated signaling pathways (7). Loss of function mutations in Smad4 are strongly correlated with the occurrence of colorectal and pancreatic cancers as well as the metastasis of prostate cancers underscoring the role of Smad4 as a tumor suppressor (8–10).

Smad4 and R-Smads span ~500 amino acids and share a common architecture consisting of two globular domains connected by linker of variable length and sequence. The N-terminal Mad Homology 1 domain (MH1 domain) is a highly conserved DNA binding domain present in R-Smads and Smad4 (11,12). The ubiquitous C-terminal MH2 domain is a protein–protein interaction module mediating Smad multimerization as well as transactivation. Upon ligand binding, TGF-β receptors phosphorylate a conserved Ser-X-Ser motif at the C-terminus of the MH2 domain of R-Smads. Crystallographic studies revealed that the phosphorylated R-Smad MH2 domains form homo as well as heterotrimers with Smad4 (13,14). By contrast, Smad4 is not subject to receptor phosphorylation and MH2 mediated homo-multimerization but rather functions as heteromerization partner for R-Smads. Since Smad4

*To whom correspondence should be addressed. Tel: +6568088102; Fax: +6568088305; Email: jauchr@gis.a-star.edu.sg
Correspondence may also be addressed to Prasanna R. Kolatkar. Tel: +6568088176; Fax: +6568088305; Email: kolatkarp@gis.a-star.edu.sg

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

contains a lysine rich nuclear localization signal (NLS) within its MH1 domain and a nuclear export signal (NES) in the linker region it acts as nuclear shuttle for activated R-Smads (15,16). After entering the nucleus, disparate Smad complexes, that is Smad4/BMP-Smad or Smad4/TGF-β-Smad multimers, are selectively recruited to specific genomic loci to regulate target genes that are likely earmarked by distinctive *cis*-regulatory sequences (17–20). Therefore, understanding the nature of these sequences and how they are recognized by particular Smad complexes is of tremendous value to shed light on TGF-β and BMP responses and ultimately predict gene expression programs from sequence (21). The problem of which DNA sequences are preferred by individual Smads as well as by Smad multimers has been tackled by a variety of methods (17,19,22–25). For example, the MH1 domain of Smad3 and Smad4 were found to specifically recognize a palindromic GTCTAGAC DNA motif [termed Smad binding element (SBE)] *in vitro* (25). On the contrary, Smad1 was reported to predominantly bind GC-rich sequences which often cluster with classical GTCT sequences (18). While the *in vitro* selected palindromic SBE is present in several TGF-β responsive promoters, GTCT-type and GC-rich elements are also found in a variety of alternative orientations and their relative preponderance remains elusive (17,18,23,25,26). Furthermore, it is unclear whether palindromic SBEs are preferentially targeted by R-Smads homodimers or by Smad4/R-Smad heterodimers (23,25,26). Since Smad4 is an essential cofactor for both TGF-β and BMP specific pathways, insights into its mechanism of DNA recognition, its preferential association with R-Smads on DNA and its preference for composite DNA motifs will shed light on how specificity is achieved in both TGF-β and BMP signaling at the level of gene regulation.

MATERIALS AND METHODS

Cloning and expression

The DNA binding MH1 domain of Smad4 (encoding amino acids 1–140, IMAGE ID: 6313280, BC046584) was amplified from full length mouse cDNA and transferred into expression vectors using Gateway BP and LR cloning technologies (22,27). The primer sequences containing an N-terminal tobacco etch virus (TEV) protease cleavage site are given in Supplementary Table S1. The resulting pDESTHis$_6$Thx-Tev-Smad4 MH1 expression construct was transformed into *Escherichia coli* (DE3) cells (Invitrogen) and grown at 37°C in Luria-Bertani (LB) broth containing 100 μg/ml ampicillin until an OD of 0.5 was reached. Protein expression was induced at 25°C with 0.2 mM isopropyl β–D-1 thiogalactopyranoside (IPTG). The cells were harvested by centrifugation after 5 h and stored at −80°C. Cells were thawed and resuspended in lysis buffer (50 mM Tris–HCl, pH 8.0, 300 mM NaCl) and disrupted by sonication. The His$_6$Thx-Smad4 MH1 fusion protein was purified by immobilized metal affinity chromatography and desalted into a buffer containing 10 mM Tris pH 8.0, 100 mM NaCl. The Smad4 MH1 was separated from the His$_6$Thx fusion tag by TEV protease cleavage, followed by heparin column purification. Finally, gel filtration was carried out using a S75 column and the pure Smad4 MH1 proteins were concentrated and stored in gel filtration buffer containing 10 mM Tris–HCl, pH 8.0, 100 mM NaCl, 2 mM TCEP. For the Smad4 MH1ΔN8 construct, 100 μM $CaCl_2$ was included in the gel filtration buffer. The Smad2 MH1 (encoding amino acids 1–183, IMAGE ID: 5066237) was cloned using Gateway BP and the amino acids encoded by exon 3 (79–108) were removed by PCR yielding the Smad2 MH1$^{\Delta E3}$ construct. The Smad2 MH1$^{\Delta E3}$ was transferred into the pDESTHis$_6$MBP expression vector and expressed and purified as described for the Smad1 MH1 (27).

Electrophoretic mobility shift assay

EMSAs were performed essentially as described in (28). In brief, Smad4 MH1 was serially diluted, mixed with 1 nM ds-Cy5 labeled DNA and 10 μl of the reaction mixture was loaded onto 12% native PAGE gels and electrophoresed using 1× Tris–Glycine buffer (25 mM Tris, pH 8.3, 192 mM Glycine). For the heterodimer assembly experiments, Smad4 MH1/SBE bound complex was incubated with serially diluted R-Smad MH1 proteins in EMSA buffer in a 15-μl reaction volume for 1 h at 4°C in the dark. An amount of 10 μl of the reaction mixture was then subject to native PAGE using 10 or 8% gels. The gel was run at 150 V for 45 min at 4°C and imaged using a typhoon phosphor imaging scanner (Amersham Biosciences).

Crystallization

PAGE-purified, deprotected single stranded palindromic SBE oligonucleotides (Sigma Proligo) were annealed by heating to 95°C for 5 min and gradually cooled to ambient temperature. The Smad4 MH1ΔN8 (spanning residues 9–140) and SBE DNA (5′ TGCAGTCTAGAC TGCA 3′) were mixed at a 2:1.2 ratio and incubated for 3–4 h on ice. Crystals were grown by mixing equal volumes of the protein/DNA complex and the reservoir buffer containing 200 mM $MgCl_2$, 100 mM Tris–HCl, pH 8.4, 30% PEG 4000 and spermine was directly added to the drop to a final concentration of 10 mM. Crystals grew overnight at 18°C using the sitting drop vapor diffusion technique. The crystals were cryoprotected by soaking in 15% glycerol for 10 min and flash frozen in liquid nitrogen.

Data collection, processing and structure solution

A 2.7 Å data set was collected at beamline X29 of the National Synchrotron Light Source using a 1.075 Å beam and the data set was integrated, scaled and merged using HKL2000 (Table 1) (29). A poly-alanine model derived from the Smad3 MH1 structure in complex with SBE DNA (1OZJ) was used for molecular replacement in PHASER integrated into PHENIX (30,31). The molecular replacement phases were improved using PARROT and the model was automatically built using BUCCANEER (32–34). The model was finalized manually in COOT using 2Fo-Fc and Fo-Fc maps (35). The refinement was carried

Table 1. Data collection and Refinement Statistics

Data collection	Smad4 MH1/SBE
Space group	C2
Cell dimensions (Å)	
a, *b*, *c*	178.36, 35.35, 139.54
β	93.83
Resolution (Å)	50–2.7 (2.8–2.7)[a]
R_{merge} (%)	7.7 (48.3)[a]
Completeness (%)	97.6 (100)[a]
I/σI	17.3 (3.0)[a]
Redundancy	6.9 (6.8)[a]
Refinement	
Resolution (Å)	24.9–2.7
No. reflections used	22 643
R_{work}/R_{free} (%)	21.2/27.2
No. atoms	
Protein/DNA	5392
Water	9
Zinc	4
Average isotropic/equivalent B-factors	
Macromolecule	68.4
Solvent	37.9
R.m.s deviations from ideal	
Bond lengths (Å)	0.004
Bond angles (°)	0.926
Ramachandran analysis	
Favored	94%
Additionally allowed	6%
Disallowed	0%

[a]Values for the highest resolution shell in parentheses.

out using PHENIX.REFINE applying NCS restraints on the equivalent protein chains and DNA strands. Translation/Libration/Screw (TLS) refinement was used during final stages of the refinement using each chain of protein and DNA as an individual group (36). PyMol was used for visualization and CURVES^{+} and 3DNA were used for analyzing the DNA topology (37–39).

RESULTS

Smad4 binds as a constitutive dimer to its palindromic SBE DNA element

In order to elucidate the DNA binding mechanism and the homodimeric assembly of the Smad4 MH1 on the palindromic SBE previously identified by SELEX (25), EMSAs were carried out (Figure 1A–C). The Smad4 MH1 binds as a constitutive dimer to the palindromic SBE but forms rather unstable monomeric complexes on DNA containing single GTCT motifs or GC-rich BRE elements (Figure 1A–C). Constitutive dimerization was not seen for equivalent constructs of R-Smads (22). Rather, Smad3 bound in an additive fashion and Smad1 showed a cooperative binding mode. Since Smad4 is also involved in BMP signaling and forms heteromeric complexes with Smad1, the DNA binding property of Smad4 MH1 to the BMP specific 'GC-BRE' element was investigated (Figure 1C). In contrast to Smad1 and Smad3 MH1 domains which both dimerize on the 'GC-BRE' compressed palindrome (22), Smad4 migrates in a poorly resolved monomeric band. Thus, the Smad4 MH1 appears to strongly prefer homodimeric association on palindromic SBE DNA as compared to elements containing single GTCT motifs and 'GC-BRE' type elements. Similarly, the Smad4 MH1 formed highly cooperative homodimers on the TGF-β responsive *JunB* and *OPN1* promoter elements where the GTCT element is arranged as a direct repeat or as a divergent palindromic element, respectively (Figure 1D and E) (23,40). These data suggest that R-Smads and their common partner Smad4 are set apart by qualitatively different DNA recognition mechanisms particularly when binding to palindromic SBEs.

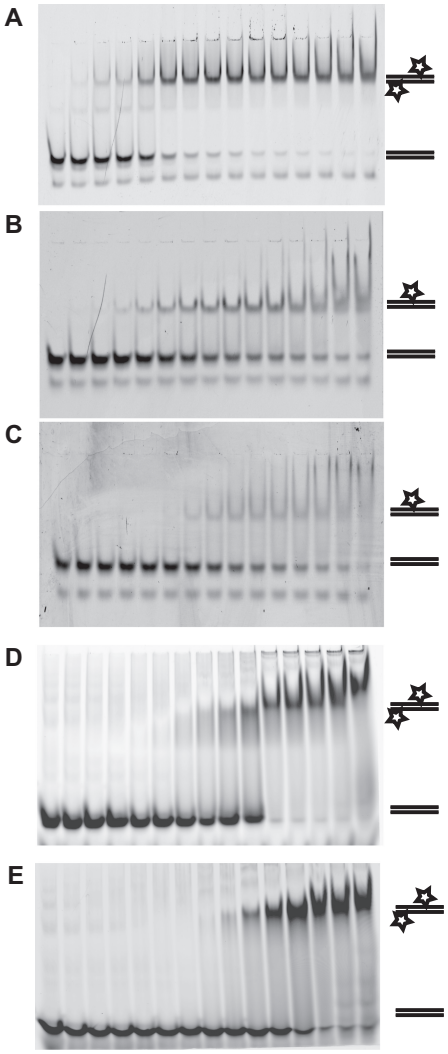

Figure 1. Smad4 forms a constitutive dimer on the palindromic SBE and poorly resolved monomers on other elements. Smad4 MH1 binding to (**A**) the palindromic SBE (TCAGTCTAGACATAC) (**B**) the single SBE (AGTATGTCTCAGATGA) (**C**) the 'GC-BRE' type element (CGCCTGGCGCCAGAGA) was analyzed by EMSA using 10% gels and 1 nM cy5 labeled DNA probes. Experiments were performed in triplicates. Smad4 MH1 binding to (**D**) the *JunB* promoter element with GTCT direct repeat (GACAGTCTGTCTGCC) and (**E**) the *OPN1* promoter element with divergent palindromic GTCT repeats separated by a 1-nt spacer (TGGAGACTGTCTGGA) were analyzed by EMSA using 10% gels and 500 nM cy5-labeled DNA probes. Protein concentrations increase in a 2-fold manner to 2000 nM from left to right.

The constitutive Smad4 dimer lacks protein–protein contacts

In order to understand the mechanism of DNA recognition by the constitutively dimeric Smad4 MH1 we sought to elucidate the crystal structure of a Smad4 MH1/SBE complex. A Smad4 MH1ΔN8/SBE complex lacking eight amino-terminal residues resulted in improved diffraction in the presence of $CaCl_2$ and spermine as additives and a 2.7 Å data set could be collected using synchrotron radiation (henceforth also termed Smad4 MH1; Table 1; Figure 2A, B, Supplementary Figure S1A and B). The asymmetric unit (ASU) contains two near-identical protein/DNA complexes each containing two Smad4 MH1 monomers and a double-stranded DNA (chain A and C in complex 1 and chain B and D in complex 2; Supplementary Figure S1C). The Smad4 MH1 monomers are globular in structure with four α-helices (H1–H5) and six short strands forming three β-sheets (β1–β6) (Figure 2B). Based on positive Fo-Fc density and the favorable placement of side chains of three Cys and one His residues, a zinc ion was modeled into the core of the globular domain (22,41). The two Smad4 MH1 monomers bind to the GTCT motif on opposite faces of the palindromic DNA and are structurally very similar with a Cα RMSD of 0.46 Å for 122 aligning residues (Figure 2B). Helix 1 of one of the Smad4 MH1 monomers (chain C) interacts with helix1 of another monomer (chain D) within the ASU bound to a different DNA leading to subtle conformational differences between molecules within a complex (Supplementary Figure S1C). Smad4 MH1 molecules not engaged in protein–protein contacts within the ASU (chain A and B) exhibit lower temperature factors and a better defined electron density than their partner molecules on opposite faces of the DNA and residues 10–138 are contained in the final model (Figure 2B and Supplementary Figure S1A). The electron densities for the functionally important residues involved in protein–DNA interactions in the β-hairpin region and elsewhere are well-defined (Supplementary Figure S1D). To our surprise, we could not observe direct protein–protein contacts between co-binding Smad4 MH1 proteins despite their constitutive dimerization in the presence of DNA. The closest proximity is seen for the tips of the β2/3 hairpin that are ~11 Å apart (Figures 2B and 3 A).

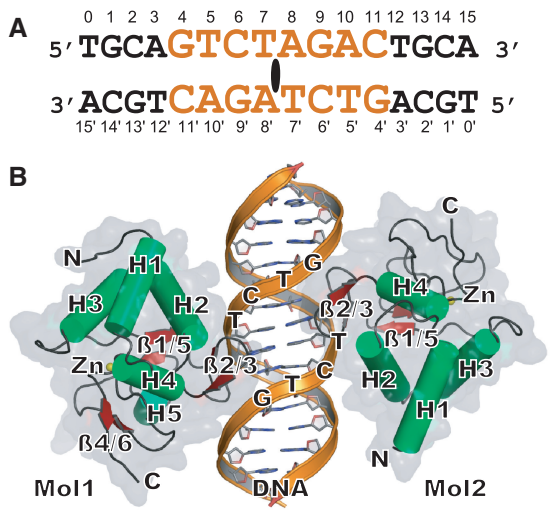

Figure 2. Overall structure of the Smad4 MH1/SBE complex. (**A**) Sequence of the 16-bp SBE palindromic DNA used for crystallization. The core SBE palindrome GTCTAGAC is shown in orange. (**B**) The overall structure of the dimeric Smad1 MH1 bound to SBE DNA shown as cartoon and semi-transparent van-der-Waals surface. α-helices are colored in green, β-sheets in red and loop regions in black.

DNA recognition by Smad4

The Smad4 MH1 recognizes the GTCT element through a β-hairpin motif formed by β2 and β3 strands (Figure 3A and B). The amino acids constituting the β2 and β3 strands that recognize specific nucleotides are identical among all the R-Smads and Smad4 (12,22). Consistently, the conserved Arg81, Gln83 and Lys88 make nucleotide specific contacts with G5′, A8 and G9, respectively (Figure 3A and B). However, the Ser-His dipeptide (Smad1 MH1 amino-acid numbering 78 and 79) within the connecting loop of β2 and β3 strands which is conserved in all R-Smads is replaced by Ala85 and Gly86 residues in the Smad4 MH1 and Ala85 makes phosphate contacts with the DNA. Although this dipeptide is not involved in nucleotide specific DNA recognition, we reasoned that it could affect the cooperative assembly of Smad proteins indirectly due to its proximity to DNA as well as to the juxtaposed protein. To test this, we replaced the Ala85Gly86 of Smad4 MH1 by SerHis found in R-Smads to generate the Smad4 $MH1^{AG/SH}$ protein. However, EMSA titrations revealed that Smad4 $MH1^{AG/SH}$ binds as constitutive homodimer to the palindromic SBE indistinguishable from the wild-type protein, indicating, that the Smad4 specific dipeptide does not contribute to its unique mode of DNA recognition (Supplementary Figure S1E). Furthermore, the Smad4 MH1ΔN8 construct and additional variants lacking 3, 10 or 12 N-terminal amino-acid residues did not diminish the formation of constitutive homodimers (Supplementary Figure S2A). This raises the question whether regions that contact the DNA non-specifically contribute to the cooperative assembly of Smad4. Indeed, several residues remote from the recognition β-hairpin in Smad4 MH1 are engaged in non-specific contacts with the DNA backbone (Figure 3A and B). For example, the Smad4 specific amino acids Arg38 and Lys106 make backbone contacts (Figure 3A and B). Other residues emanating from helix2 such as Ser42 and Lys45 also contact the phosphate backbone (Supplementary Table S2). It was previously shown that mutating Leu43, Lys46 and Lys50 of helix2 lead to drastic reduction in DNA binding affinity and transcriptional activation of Smad4 (42,43). We therefore hypothesized that the backbone contacts largely mediated by helix2 contribute to the binding affinity and, perhaps, complex assembly. In order to test this possibility, we produced mutant proteins where several of the Smad4 MH1 specific residues were replaced by their counterparts found in the Smad3 MH1. However, we found that the Smad4 $MH1^{K106S}$, Smad4

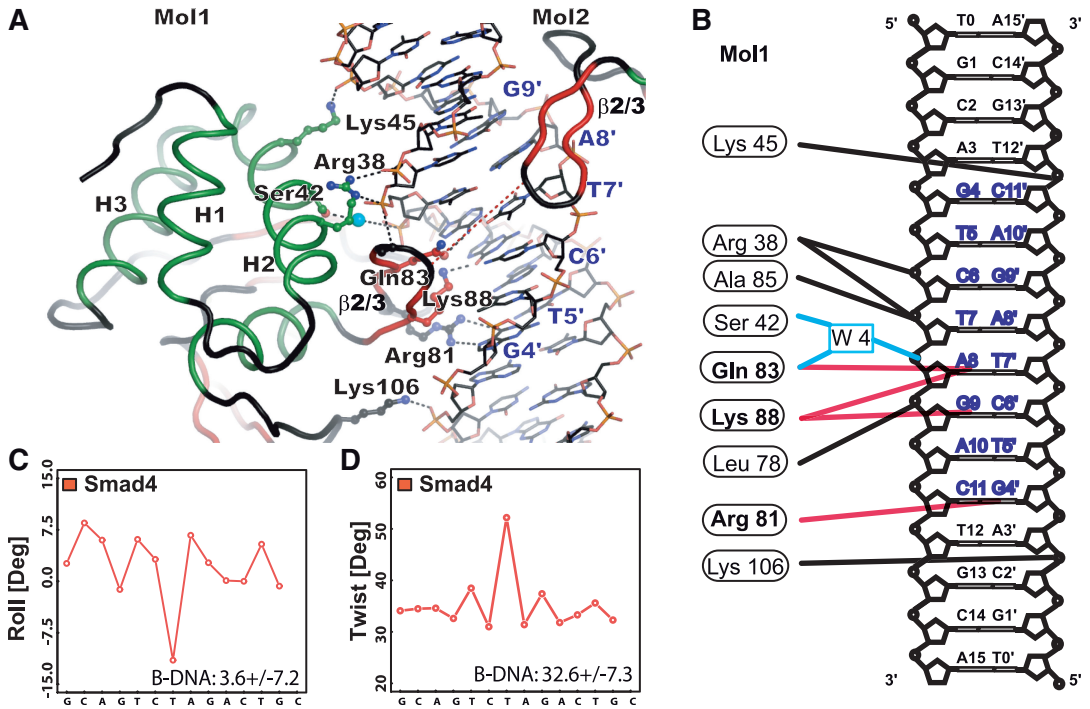

Figure 3. DNA recognition by Smad4. (**A**) Protein–DNA interactions of the Smad4 MH1. Arg81, Gln83 and Lys88 specifically interacting with A8, G9 and G4′ as well as residues engaged in backbone contacts are shown as ball-and-sticks. The water molecule mediating interactions with Ser42 is shown as a blue sphere. Phosphate interactions are indicated with black dashes. (**B**) Schematic drawing of Smad4 MH1 SBE DNA interaction. Amino acids engaging in specific DNA contacts (red lines) as well as phosphate backbone contacts (black lines) and water mediated contacts (blue line) are shown. Only contacts by molecule 1 are shown for clarity. Smad4 induces deviation from the B-DNA topology. (**C**) Roll and (**D**) twist angles determined using Curves^{+} plotted against the SBE DNA sequences highlighting the negative roll and the over-twisting at the palindromic center. Parameters for standard B-DNA are shown and (+/–) indicates standard deviations (38). The overall topologies at the palindromic centers are similar for Smad1, Smad3 and Smad4 bound DNA and details are given in Supplementary Figures S3 and S4.

MH1^{R38K} and Smad4 MH1$^{105HKN107/HSHH}$ mutants bound to SBE DNA as a constitutive homodimer indistinguishable from the wild-type Smad4 MH1 domain (Supplementary Figure S2B). Conversely, Smad3 MH1 mutants containing Smad4-like amino acids, Smad3 MH1^{K33R}, Smad3 MH1^{S100K}, Smad3 MH1$^{99HSHH102/HKN}$ showed additive binding to the SBE DNA reminiscent of the wild-type protein (Supplementary Figure S2B). Hence, the amino acids in the β-hairpin or the tested Smad4 specific amino acids involved in phosphate contacts are not sufficient to mediate cooperative binding of the Smad4 MH1 on DNA. Rather, the cooperation must be caused by a less obvious, indirect mechanism.

Smad4 induces a non-canonical DNA structure

It is becoming increasingly clear that both global and local shape of DNA influences the affinity, cooperativity and specificity of protein–DNA binding (44). Thus, we analyzed the topology of Smad4 MH1 bound DNA using Curves^{+} (38). The SBE DNA exhibits pronounced deviation from the canonical B-form conformation on the base pair and base pair step level (Figure 3C and D). In particular, the roll angle at the junction of inversion of the GTCT-AGAC palindrome decreases sharply to −11.5 leading to a slight compression of the minor groove (Figure 3C). Furthermore, the palindromic center is strongly overwound as reflected by a twist angle of 52.7° (Figure 3D). This junction also has an increased depth of the minor groove compared to the adjacent nucleotide basepairs (Figure 6, Supplementary Figures S3 and S4). These protein induced conformational changes in the DNA could affect secondary binding events and thus influence the cooperativity of Smad4 binding.

The Smad4 MH1 promotes heterodimerization with R-Smads

It has been shown that R-Smads and Smad4 form heteromeric complexes to enter the nucleus and regulate gene expression (14). Yet, it remains elusive how the MH1 domains of R-Smad/Smad4 complexes selectively heteromerize on differently configured GTCT repeat elements. To understand the role of the Smad4 MH1 in the R-Smad/Smad4 heteromeric complex formation, pre-incubated Smad4 MH1/SBE complexes were titrated with serially diluted Smad1, Smad2 and Smad3 MH1 proteins. Smad4/R-Smad heterodimers were formed in the presence of all the R-Smads (Figure 4A–C). Intriguingly, heterodimeric complexes appeared at concentrations when free DNA was still abundantly available and before R-Smad homodimers or R-Smad monomers were formed. Since heterodimer formation occurred

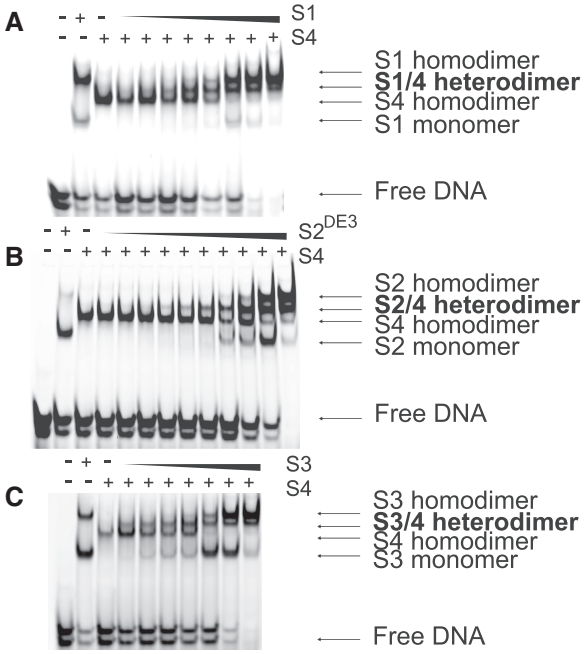

Figure 4. Heteromerization between Smad4 and R-Smads. An amount of 600 nM of Smad4 MH1 was pre-incubated with 2000 nM SBE DNA and 2-fold serial dilutions (right to left) of the (**A**) Smad1 MH1 (S1, 1600 nM in lane 1 and 2-fold serial dilution starting from 6400 nM), (**B**) the Smad2$^{\Delta E3}$ MH1 (S2, 320 nM in lane 1 and 2-fold serial dilution starting from 2560 nM) and (**C**) the Smad3 MH1 (S3, 800 nM in lane 1 and 2-fold serial dilution starting from 6400 nM) were added and the reactions were analyzed by EMSAs. The various DNA bound Smad complexes are marked. The titrations reveal that R-Smad/Smad4 heterodimers form at lower concentrations than R-Smad mono- or homodimers indicating that heterodimerization is the preferred binding mode on the palindromic SBE.

at the expense of pre-formed Smad4 homodimers, R-Smads effectively compete with otherwise constitutively homodimeric Smad4. As a consequence, on the palindromic SBE, R-Smad/Smad4 heterodimers represent the most favored multimeric state of Smad MH1 domains. This indicates that Smad4 can be seen as binding vehicle that cooperatively assembles with R-Smads to recruit and retain them on composite binding sites (Figure 4A–C).

What is the basis for Smad specific multimerization patterns?

The Smad4 MH1 structure presented here allows comparing the structures of representative members of the major Smad families in complex with palindromic SBE DNA. Importantly, the homodimerization pattern of the Smad family is fundamentally different: the TGF-β regulated Smad2$^{\Delta E3}$ and Smad3 bind additively, the BMP regulated Smad1 positively cooperates while the common partner Smad4 homodimerizes constitutively (Figure 5A–D). Yet, in the absence of DNA, Smad 1, 3 and 4 MH1 domains are monomeric [Figure 5E and (27)] indicating that the binding differences are due to variant modes of DNA recognition. Thermal melting analysis shows that the Smad4 MH1 is structurally stabilized when bound to DNA whereas the thermostablility of the Smad3 MH1 is not affected by DNA (Figure 5F). This suggests that the Smad4 MH1 undergoes a more substantial structural reorganization upon DNA binding than the Smad3 MH1. To explore structural differences between Smads, we superimposed the structures of Smad1, Smad3 and Smad4 and carefully inspected the DNA recognition interface (Figure 5G). As expected, the overall topology is very similar but some key differences were observed at the N-terminus encompassing helices 1 and 2. The most drastic rearrangement was observed for Smad1 that undergoes a crystallographic domain swap in this region (Figure 5G). However, Smad3 and Smad4 look very similar despite the drastic difference with regard to cooperative complex formation. A notable exception is the Smad4-specific Arg38 that engages in a tight backbone interaction (Figure 3A, B and Supplementary Figure S1D), whereas the Lys found in the corresponding positions in Smad1 and Smad3 point away from DNA. Further amino acids engaged in direct or indirect DNA contacts in Smad4 but not Smad3 comprise Ser42 and Lys106 (Figure 3A and B). However, introducing amino acids found in Smad3 at these positions leading to the mutant proteins Smad4 MH1^{K106S} and Smad4 MH1^{R38K} did not diminish the constitutive homodimer formation of Smad4 MH1 on SBE DNA (Supplementary Figure S2B). It has been shown that the DNA structure substantially affects protein–DNA binding through indirect readout mechanisms (12,22,44). We therefore carefully inspected the DNA shapes induced by the different Smad proteins. Smad4 exhibits the lowest overall bend (7.2°) when compared to Smad1 and Smad3 (14.5° or 19.8°, Figure 5H). On the base pair level, the typical B-form DNA conformation is modified in all three structures by the binding of two Smad MH1 domains (Figure 5H, Supplementary Figure S3 and S4). All three Smads overtwist and open base pairs at the palindromic center and exhibit a variety of altered base pair and base pair step parameters (Supplementary Figure S3 and S4). When inspecting the groove architectures we found a subtly stronger compression of the major and minor grooves within the right half of the palindrome for the Smad4/SBE when compared to Smad3/SBE and Smad1/SBE complexes (Figure 6A–D). Also, the oscillation of the major groove depth at the palindromic center is more pronounced for the Smad4 bound SBE (Figure 6A). By conducting Pearson's product moment correlation analysis we further established that a number of helical parameters including the minor groove width, rise, stretch, stagger and propeller are significantly different for SBE DNA bound by Smad1, Smad3 or Smad4 (Supplementary Table S3). While some of these differences may be due to alternative crystal packing, we expect most of them to be a consequence of protein binding. In particular intra base pair parameters at the center of the DNA element can hardly be caused by packing artifacts. It will be interesting to explore whether and how these subtle structural difference in DNA shape affect molecular recognition events and the complex assembly of Smad MH1 domains.

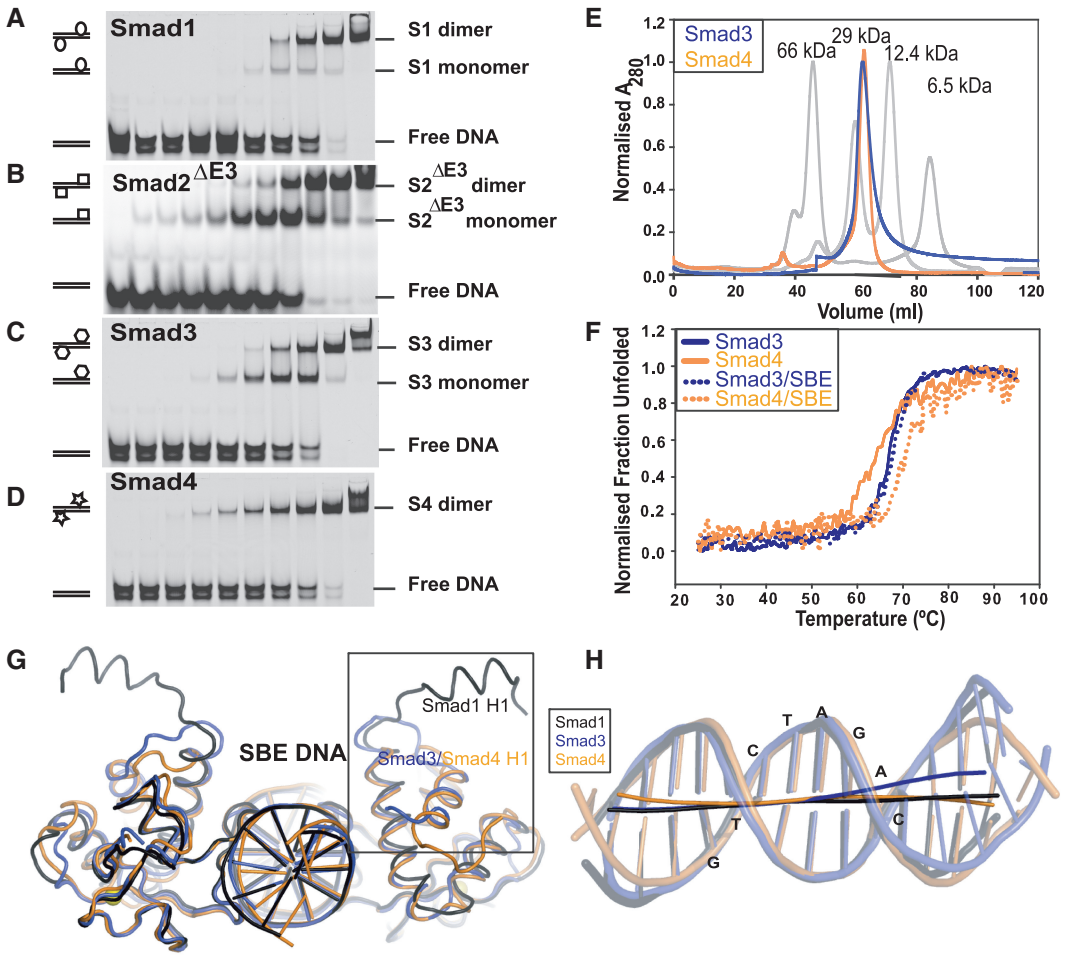

Figure 5. Comparison of Smad1, Smad3 and Smad4. EMSAs to compare the binding of (**A**) the Smad1 MH1, (**B**) the Smad2$^{\Delta E3}$ MH1, (**C**) the Smad3 MH1 and (**D**) the Smad4 MH1 to 2.5 μM SBE DNA reveal substantially different homodimerization patterns. (**E**) Gel filtration chromatograms showing that the Smad3 MH1 and the Smad4 MH1 elute with overlapping peaks. The chromatogram of a molecular weight standard is shown in grey (22,27). (**F**) Melting curves for the Smad3 and Smad4 MH1 in the absence and presence of DNA were recorded employing circular dichroism spectroscopy as described (22). The data indicate that the Smad4 MH1 is structurally stabilized upon DNA binding whereas the Smad3 MH1 is not. (**G**) Superposition of the MH1/DNA structures for Smad1, Smad3 and Smad4 emphasizing overall structural similarities with the notable exceptions of the 'open' conformation of helix 1 of Smad1 and an N-terminally shortened helix 2 seen for Smad4. (**H**) Overlay of the helical axes calculated with Curves^{+} and cartoons of the SBE DNAs bound to Smad1 (black), Smad3 (blue) and Smad4 MH1 (orange) illustrating differences of the overall curvature of the double helix.

DISCUSSION

Little is known how specificity is achieved in gene regulation and how transcription factors cooperate to selectively target genomic control regions (45,46). In TGF-β signaling, this can be achieved despite the short GTCT sequences commonly recognized by the DNA binding MH1 domain of Smads (11). Smads are thought to bind DNA as pre-formed complexes mediated by their MH2 domains but it is still debated whether they act as dimers or trimers (47). The variable recognition of differently configured GTCT motif in the form of direct, indirect or divergent repeats with varying spacers by distinct Smad complexes could increase the versatility and selectivity of Smad signaling and could set genes responding to TGF-β or BMP signaling apart. The ability of the MH1 domain to preferentially recognize such DNA binding sites is the key for a constructive complex assembly to occur. We therefore studied complex formation of MH1 domains from all major Smad families on DNA. By comparing the binding profile of R-Smads and Smad4 to the palindromic SBE we found substantially different cooperativity profiles with Smad4 homodimerizing in a constitutive fashion. Smad4 also binds in a constitutively homodimeric fashion on direct and divergent repeat elements derived from the promoters of the *JunB* and *OPN1* genes. Importantly, R-Smad/Co-Smad heterodimerization was found to constitute the preferred binding mode on the SBE DNA. The Smad4 MH1 therefore appears to strongly support homo as well as heterotypic dimerization and acts as a dimerization vehicle. Thus, it can be inferred that the MH1 domain plays an important role in the assembly of heteromeric R-Smad/Smad4 complex on TGF-β responsive GTCT repeat elements and is not merely required for nuclear shuttling of R-Smads. However, despite its strong cooperation with itself and other Smads, Smad4

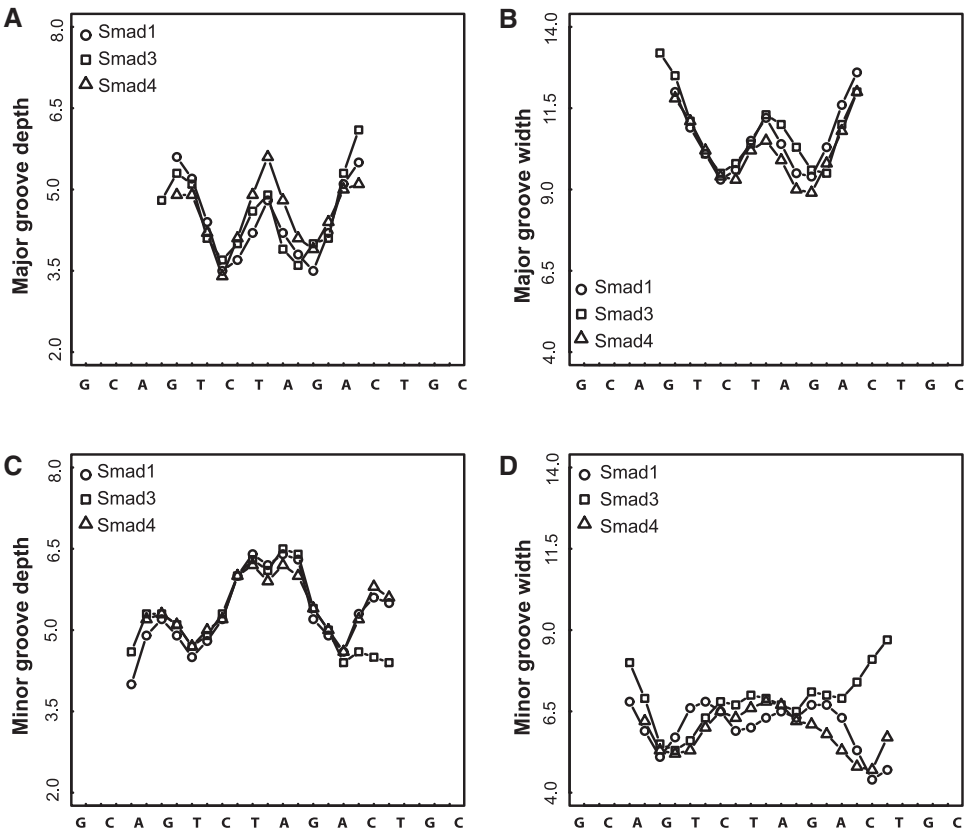

Figure 6. Major groove depth (**A**) and width (**B**) as well as minor groove depth (**C**) and width (**D**) were calculated using Curves^{+} for Smad1 MH1 (circles), Smad3 MH1 (squares) and Smad4 MH1 (triangles) bound SBE DNA. See Supplementary Table S3 for Pearson's product moment correlation analysis of helical parameters.

lacks direct protein–protein contacts in the MH1 domain and is structurally surprisingly similar to the non-cooperatively homodimerizing Smads. Thus, Smad4 most likely employs an indirect, DNA-mediated mode to facilitate the recruitment of other proteins. Apparently, the binding of the first Smad4 molecule drastically lowers the binding energy for the second molecule, leading to a macroscopically constitutive dimer formation. On the contrary, binding of the first Smad3 molecule leaves the second binding event unaffected. We envisage two possible inter-connected mechanisms underlying the DNA mediated cooperativity accompanying Smad4 binding: (i) an indirect–indirect readout mechanism and/or (ii) the removal of the entropic barrier by the first binding event facilitating the secondary binding. Generally defined, indirect readout refers to selective recognition of DNA shapes, that is DNA deviating from the B-form, such as groove architectures by DNA binding proteins (44). The basis for varying DNA shapes depends on its sequences and can be either pre-formed or reflect a propensity to be deformed upon protein binding. In the present study, we found a series of subtle conformational differences induced by different Smad protein (Figure 6A–D, Supplementary Figures S3 and S4, Supplementary Table S3). Yet, the DNA sequences are essentially identical for the palindromic SBE bound by Smad1, Smad3 and Smad4 excluding the possibility of disparate DNA shapes before association with proteins. Rather, individual Smad proteins appear to induce subtly different DNA shapes that affect the indirect readout and thus the binding affinity at the secondary binding site. We therefore refer to this mechanism as indirect-indirect readout involving an indirect readout that is allosterically triggered from a distance by a protein binding proximal to the secondary binding site.

Alternatively, the cooperativity differences could be due to Smad specific energetic profiles. The overall binding energy of protein–DNA interactions is a function of a multitude of components that also include changes in the vibrational/translational entropy and structural adaptations of the binding partners (48). Structural adaptations as well as entropic constraints typically counteract the binding and are overcompensated by energetically more favorable energy terms (48–51). It is therefore conceivable that in a multi-component system, the first binding event bears the bulk of the entropic cost and induces the majority of the structural adaptations which greatly reduces the energy barrier for the secondary binding event (52). The subtle structural differences of the DNA bound to Smad4 and Smad3 as well as the larger number of non-specific DNA contacts seen for Smad4 could lead to changes in the binding energy of the secondary binding event and therewith affect the cooperativity.

ACCESSION NUMBER

Coordinates and structure factors have been submitted to the protein databank with the accession number 3QSV.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Howard Robinson for data collection and processing. The authors thank Bob Robinson for providing access to X-ray diffraction equipment and crystallization robotics. The authors are also grateful to Calista Keow Leng Ng, Chloe Tan, Immanuel Kwok, Benson Leow and Siew Hua Choo for technical support. N.B. and R.J. designed research; N.B., C.T.Y.Z. and K.N. conducted experiments; N.B., R.J. and P.R.K. analyzed data; N.B. and R.J. wrote the paper; P.R.K. provided funding.

FUNDING

Diffraction data were collected at beamline X29 of the National Synchrotron Light Source (NSLS) supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences; Agency for Science, Technology and Research, Singapore. Funding for open access charge: Agency for Science, Technology and Research, Singapore.

Conflict of interest statement. None declared.

REFERENCES

1. Whitman,M. (1998) Smads and early developmental signaling by the TGFbeta superfamily. *Genes Dev.*, **12**, 2445–2462.
2. Massagué,J. (1998) TGF-beta signal transduction. *Annu. Rev. Biochem.*, **67**, 753–791.
3. Hoodless,P., Haerry,T., Abdollah,S., Stapleton,M., O'Connor,M., Attisano,L. and Wrana,J. (1996) MADR1, a MAD-related protein that functions in BMP2 signaling pathways. *Cell*, **85**, 489–500.
4. Kawai,S., Faucheu,C., Gallea,S., Spinella-Jaegle,S., Atfi,A., Baron,R. and Roman,S. (2000) Mouse smad8 phosphorylation downstream of BMP receptors ALK-2, ALK-3, and ALK-6 induces its association with Smad4 and transcriptional activity. *Biochem. Biophys. Res. Commun.*, **271**, 682–687.
5. Macías-Silva,M., Abdollah,S., Hoodless,P., Pirone,R., Attisano,L. and Wrana,J. (1996) MADR2 is a substrate of the TGFbeta receptor and its phosphorylation is required for nuclear accumulation and signaling. *Cell*, **87**, 1215–1224.
6. Zhang,Y., Feng,X., We,R. and Derynck,R. (1996) Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. *Nature*, **383**, 168–172.
7. Hahn,S.A., Schutte,M., Hoque,A.T., Moskaluk,C.A., da Costa,L.T., Rozenblum,E., Weinstein,C.L., Fischer,A., Yeo,C.J., Hruban,R.H. *et al.* (1996) DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science*, **271**, 350–353.
8. Ding,Z., Wu,C.J., Chu,G.C., Xiao,Y., Ho,D., Zhang,J., Perry,S.R., Labrot,E.S., Wu,X., Lis,R. *et al.* SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature*, **470**, 269–273.
9. Koyama,M., Ito,M., Nagai,H., Emi,M. and Moriyama,Y. (1999) Inactivation of both alleles of the DPC4/SMAD4 gene in advanced colorectal cancers: identification of seven novel somatic mutations in tumors from Japanese patients. *Mutat. Res.*, **406**, 71–77.
10. Sun,C., Yamato,T., Furukawa,T., Ohnishi,Y., Kijima,H. and Horii,A. (2001) Characterization of the mutations of the K-ras, p53, p16, and SMAD4 genes in 15 human pancreatic cancer cell lines. *Oncol. Rep.*, **8**, 89–92.
11. Fortuno,E.r., LeSueur,J. and Graff,J. (2001) The amino terminus of Smads permits transcriptional specificity. *Dev. Biol.*, **230**, 110–124.
12. Shi,Y., Wang,Y., Jayaraman,L., Yang,H., Massagué,J. and Pavletich,N. (1998) Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell*, **94**, 585–594.
13. Chacko,B., Qin,B., Tiwari,A., Shi,G., Lam,S., Hayward,L., De Caestecker,M. and Lin,K. (2004) Structural basis of heteromeric smad protein assembly in TGF-beta signaling. *Mol. Cell.*, **15**, 813–823.
14. Wu,J., Hu,M., Chai,J., Seoane,J., Huse,M., Li,C., Rigotti,D., Kyin,S., Muir,T., Fairman,R. *et al.* (2001) Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. *Mol. Cell*, **8**, 1277–1289.
15. Watanabe,M., Masuyama,N., Fukuda,M. and Nishida,E. (2000) Regulation of intracellular dynamics of Smad4 by its leucine-rich nuclear export signal. *EMBO Rep.*, **1**, 176–182.
16. Xiao,Z., Latek,R. and Lodish,H.F. (2003) An extended bipartite nuclear localization signal in Smad4 is required for its nuclear import and transcriptional activity. *Oncogene*, **22**, 1057–1069.
17. Dennler,S., Itoh,S., Vivien,D., ten Dijke,P., Huet,S. and Gauthier,J. (1998) Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. *EMBO J.*, **17**, 3091–3100.
18. Karaulanov,E., Knochel,W. and Niehrs,C. (2004) Transcriptional regulation of BMP4 synexpression in transgenic Xenopus. *EMBO J.*, **23**, 844–856.
19. Kusanagi,K., Inoue,H., Ishidou,Y., Mishima,H.K., Kawabata,M. and Miyazono,K. (2000) Characterization of a bone morphogenetic protein-responsive Smad-binding element. *Mol. Biol. Cell*, **11**, 555–565.
20. Stopa,M., Anhuf,D., Terstegen,L., Gatsios,P., Gressner,A.M. and Dooley,S. (2000) Participation of Smad2, Smad3, and Smad4 in transforming growth factor beta (TGF-beta)-induced activation of Smad7. THE TGF-beta response element of the promoter requires functional Smad binding element and E-box sequences for transcriptional regulation. *J. Biol. Chem.*, **275**, 29308–29317.
21. Fujibuchi,W. and Kanehisa,M. (1997) Prediction of gene expression specificity by promoter sequence patterns. *DNA Res.*, **4**, 81–90.
22. BabuRajendran,N., Palasingam,P., Narasimhan,K., Sun,W., Prabhakar,S., Jauch,R. and Kolatkar,P.R. (2010) Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors. *Nucleic Acids Res.*, **38**, 3477–3488.
23. Jonk,L., Itoh,S., Heldin,C., ten Dijke,P. and Kruijer,W. (1998) Identification and functional characterization of a Smad binding element (SBE) in the JunB promoter that acts as a transforming growth factor-beta, activin, and bone morphogenetic protein-inducible enhancer. *J. Biol. Chem.*, **273**, 21145–21152.
24. López-Rovira,T., Chalaux,E., Massagué,J., Rosa,J. and Ventura,F. (2002) Direct binding of Smad1 and Smad4 to two distinct motifs mediates bone morphogenetic protein-specific transcriptional activation of Id1 gene. *J. Biol. Chem.*, **277**, 3176–3185.
25. Zawel,L., Dai,J., Buckhaults,P., Zhou,S., Kinzler,K., Vogelstein,B. and Kern,S. (1998) Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol. Cell.*, **1**, 611–617.
26. Nagarajan,R., Zhang,J., Li,W. and Chen,Y. (1999) Regulation of Smad7 promoter by direct association with Smad3 and Smad4. *J. Biol. Chem.*, **274**, 33412–33418.
27. Baburajendran,N., Palasingam,P., Ng,C.K., Jauch,R. and Kolatkar,P.R. (2009) Crystal optimization and preliminary diffraction data analysis of the Smad1 MH1 domain bound to a

palindromic SBE DNA element. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **65**, 1105–1109.
28. Jauch,R., Ng,C.K., Saikatendu,K.S., Stevens,R.C. and Kolatkar,P.R. (2008) Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog. *J. Mol. Biol.*, **376**, 758–770.
29. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
30. Adams,P.D., Gopal,K., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Pai,R.K., Read,R.J., Romo,T.D. *et al.* (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron. Radiat.*, **11**, 53–55.
31. McCoy,A.J., Grosse-Kunstleve,R.W., Storoni,L.C. and Read,R.J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 458–464.
32. Zhang,K.Y., Cowtan,K. and Main,P. (1997) Combining constraints for electron-density modification. *Methods Enzymol.*, **277**, 53–64.
33. Collaborative Computational Project,N. (1994) The CCP4 Suite: programs for protein crystallography. *Acta Crystallographica*, **D 50**, 760–763.
34. Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 1002–1011.
35. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
36. Afonine,P.V., Grosse-Kunstleve,R.W. and Adams,P.D. (2005) A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 850–855.
37. DeLano,W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.
38. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
39. Lu,X.J. and Olson,W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.
40. Hullinger,T.G., Pan,Q., Viswanathan,H.L. and Somerman,M.J. (2001) TGFbeta and BMP-2 activation of the OPN promoter: roles of smad- and hox-binding elements. *Exp. Cell Res.*, **262**, 69–74.
41. Chai,J., Wu,J., Yan,N., Massagué,J., Pavletich,N. and Shi,Y. (2003) Features of a Smad3 MH1-DNA complex. Roles of water and zinc in DNA binding. *J. Biol. Chem.*, **278**, 20327–20331.
42. Kuang,C. and Chen,Y. (2004) Tumor-derived C-terminal mutations of Smad4 with decreased DNA binding activity and enhanced intramolecular interaction. *Oncogene*, **23**, 1021–1029.
43. Jones,J.B. and Kern,S.E. (2000) Functional mapping of the MH1 DNA-binding domain of DPC4/SMAD4. *Nucleic Acids Res.*, **28**, 2363–2368.
44. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
45. Jauch,R., Aksoy,I., Hutchins,A.P., Ng,C.K., Tian,X.F., Chen,J., Palasingam,P., Robson,P., Stanton,L.W. and Kolatkar,P.R. (2011) Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. *Stem Cells*, **29**, 940–951.
46. Mirny,L.A. (2011) Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl Acad. Sci. USA*, **107**, 22534–22539.
47. Massague,J., Seoane,J. and Wotton,D. (2005) Smad transcription factors. *Genes Dev.*, **19**, 2783–2810.
48. Becker,N.B., Wolff,L. and Everaers,R. (2006) Indirect readout: detection of optimized subsequences and calculation of relative binding affinities using different DNA elastic potentials. *Nucleic Acids Res.*, **34**, 5638–5649.
49. Jayaram,B., McConnell,K., Dixit,S.B., Das,A. and Beveridge,D.L. (2002) Free-energy component analysis of 40 protein-DNA complexes: a consensus view on the thermodynamics of binding at the molecular level. *J. Comput. Chem.*, **23**, 1–14.
50. Jen-Jacobson,L. (1997) Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers*, **44**, 153–180.
51. Koudelka,G.B., Mauro,S.A. and Ciubotaru,M. (2006) Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog. Nucleic Acid Res. Mol. Biol.*, **81**, 143–177.
52. Chen,Y. and Young,M.A. Structure of a thyroid hormone receptor DNA-binding domain homodimer bound to an inverted palindrome DNA response element. *Mol. Endocrinol.*, **24**, 1650–1664.