# Supplementary Data

# Structural basis for the Smad5 MH1 domain to recognize different DNA sequences

Nan Chai[1], Wan-Xin Li[2], Jue Wang[1], Zhi-Xin Wang[1], Shi-Ming Yang[2,*] and Jia-Wei Wu[1,3,*]

[1] MOE Key Laboratory of Protein Science, School of Life Sciences, Tsinghua University, Beijing 100084, China

[2] Department of Otolaryngology Head and Neck Surgery, Institute of Otolaryngology, Chinese PLA General Hospital, Beijing 100853, China

[3]Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China

* To whom correspondence should be addressed.

The supplementary data includes two supplementary tables and six supplementary figures.

**Supplementary Table 1. A brief summary of Smad binding sequences.**

| Gene[1] | *cis*-regulatory element | Position[2] | 5'-Sequence-3' [3] | Protein | Ref. |
|---|---|---|---|---|---|
| **SBE only** | | | | | |
| Hs *p21Cip1* | promoter | -1755 | AACAGACAGACAATGTCTAGTCTATT | Smad3, 4 | (1) |
| Mm *JunB* | promoter | -2813 | TTTCTCAGACAGTCTGTCTGCC | Smad3, 4 | (2) |
| Mm *Nkx2-5* | enhancer | -2778 | GCTGCCTGTCTCCA | Smad4 | (3) |
| Mm *Pparg* | promoter | -238 | GCGAGAC-$N_{32}$-GTCT-$N_{12}$-AGACAGC | Smad4 | (4) |
| Hs *ID1* | promoter | -903 | CGGTCT-$N_{13}$-AGACGCTGACACAGACCG | Smad1, 4 | (5) |
| **GC-rich only** | | | | | |
| Dm *Ubx* | promoter | -2743 | CAGCGCCGGCGCTT | Mad | (6-8) |
| Dm *pnr* | enhancer | 1513 | ACTGCTGGCGCCACCACT | Mad, Medea | (9) |
| Mm *Smad6* | promoter | -1906 | GCCGCGCCGGCTCCA | Smad1, 5, 4 | (10) |
| **GC & SBE** | | | | | |
| Dm *brk* | silencer | -8189 | ACTGGCGACATTCTGTCTGGT | Mad, Medea | (11,12) |
| Dm *dad* | enhancer | 2351 | GTGGCGCCATTCCGACG[4]A | Mad, Medea | (13) |
| Xl *ventx2* | promoter | -171 | GCAGACATGGTGGAGCCAGCT | Smad1, 4 | (14,15) |
| Xl *bambi* | enhancer | -2061 | GTCGTCTCAGTGGCGCCGG | Smad1, 4 | (16) |
| Mm *Id1* | promoter | -1104 | GCGCGGCGCCAGCCTGACA | Smad1, 5 | (17) |
| Mm *Id2* | promoter | -2781 | CCTGGCGCCAGAGAGTCTGCT | Smad1, 5, 4 | (18) |
| Hs *ID3* | promoter | -2962 | CTGGCGCCAGGCTGTCTGG | Smad1, 5, 4 | (19) |
| Mm *Smad7* | promoter | 18 | CCAGCCAGCCAGCCGGCGCCAC | Smad1, 4 | (20) |

[1] Hs: *Homo sapiens*, Mm: *Mus musculus*, Dm: *Drosophila melanogaster*, Xl: *Xenopus laevis.*

[2] The position of the Smad targeted sequence is derived from the references cited for each gene.

[3] The SBE and GC-rich sequences are highlighted in yellow and cyan, respectively.

[4] A non-canonical SBE site is underlined, which is a Medea binding site.

**Supplementary Table 2. Summary of composite DNA sequences used for crystallization.**

| DNA Name | Description | Sequence[1,2] | Crystal (Y/N) | Diffraction |
|---|---|---|---|---|
| Id2 | Mouse *Id2* DNA | CCTGGCGCCAGAGAGTCTGC | N | / |
| Bambi | Human *BAMBI* DNA | AGTGTCGTCTCGTTGGCGCCGGGTGC | N | / |
| Id3 | Human *ID3* DNA | AACCCTGGCGCCAGGCTGTCTGGGGC | N | / |
| Smad7 | Mouse *Smad7* DNA | CCAGCCAGCCAGCCGGCGCCAC | N | / |
| GCj | SBE+GC with no spacer | 5'ATCAGACGGCGCCAGAGA3' | Y | >10Å |
| GCRj | | 5'ATCAGACGCCGGCAGAGA3' | Y | 8Å |
| GCjt | SBE+GC with 1bp spacer | 5'ATCAGACTGGCGCCAGAGA3' | Y | >10Å |
| GCRjt | | 5'ATCAGACTGCCGGCAGAGA3' | Y | No |
| GCj4 | SBE+GC with long spacer | 5'ATCGTCTCTACTGGCGCCATA3' | Y | No |
| GCRj5 | | 5'ATCAGACCTACTGGCCGGCATA3' | Y | >10Å |
| GCj6 | | 5'ATCAGACCTACTGGGCGCCATA3' | Y | 8Å |
| GCj3 | SBE+GC+SBE | 5'ATCAGACGGCGCCGTCTATA3' | Y | 7Å |
| GCj2 | SBE+GC+SBE with1bp spacer | 5'ATCAGACTGGCGCCAGTCTATA3' | Y | >10Å |
| GCRj2 | | 5'ATCAGACTGCCGGCAGTCTATA3' | Y | 3.20Å |

[1] For space limitations, the complementary strands of DNA are not shown here.

[2] Canonical SBE and GC-rich sequences are highlighted as yellow and cyan, respectively.

## Supplementary Figure 1

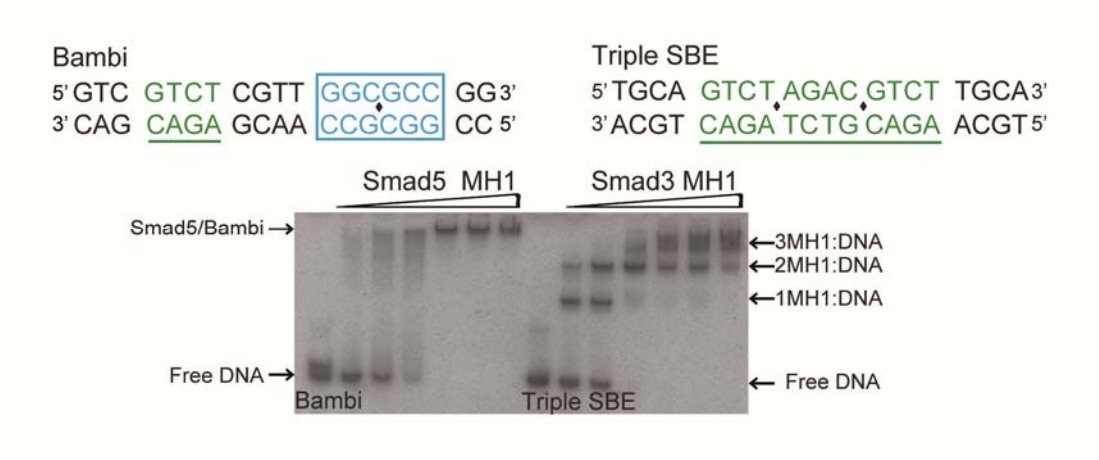


**Figure S1.** Smad5-MH1 binding to the *Bambi* enhancer element. The complexes of Smad3-MH1 binding to the Triple SBE DNA sequence (right panel of the gel) were used as molecular weight markers. The EMSA assays were done as that in Figure **1C**. Sequences used for EMSA are shown above the gel in which the SBE and GC-rich sites are displayed as that in Figure **1C**.

## Supplementary Figure 2

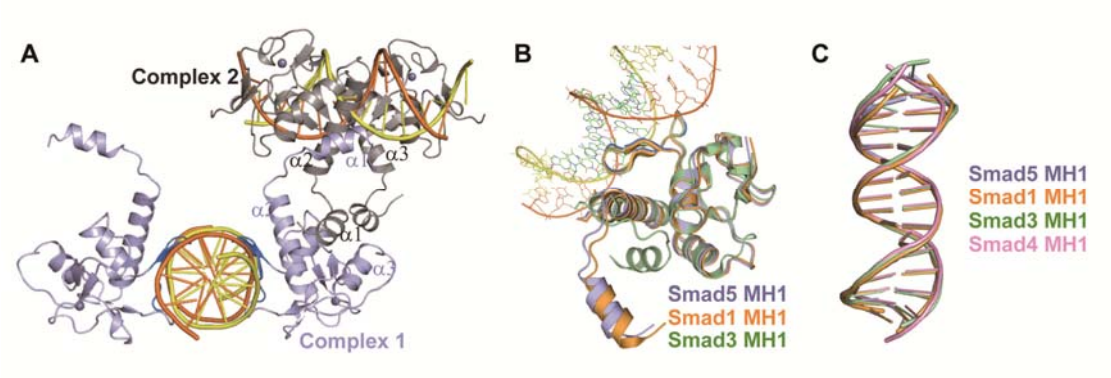


**Figure S2.** (**A**) The domain-swapped Smad5-MH1-SBE complexes formed by adjacent symmetry-related complexes. The colour scheme of Smad5-MH1 in complex 1 and the DNA duplexes in both complexes follows that in Figure **2A**, while the MH1 domains in complex 2 are coloured in grey. (**B**) Superposition of the MH1-SBE structures from the Smad3-MH1-SBE complex (PDB code: 1OZJ, coloured in green), the Smad1-MH1-SBE complex (PDB code: 3KMP, coloured in orange) and the Smad5-MH1-SBE complex (light blue). The displacements of helices α2 in both Smad5 and Smad1 lead to the loss of several phosphate contacts by the lysine residues. (**C**) Overlay of cartoons of the SBE DNAs bound to the MH1 domain of Smad1 (orange), Smad3 (green), Smad4 (pink) and Smad5 (light blue) illustrating the similar DNA shape.

## Supplementary Figure 3

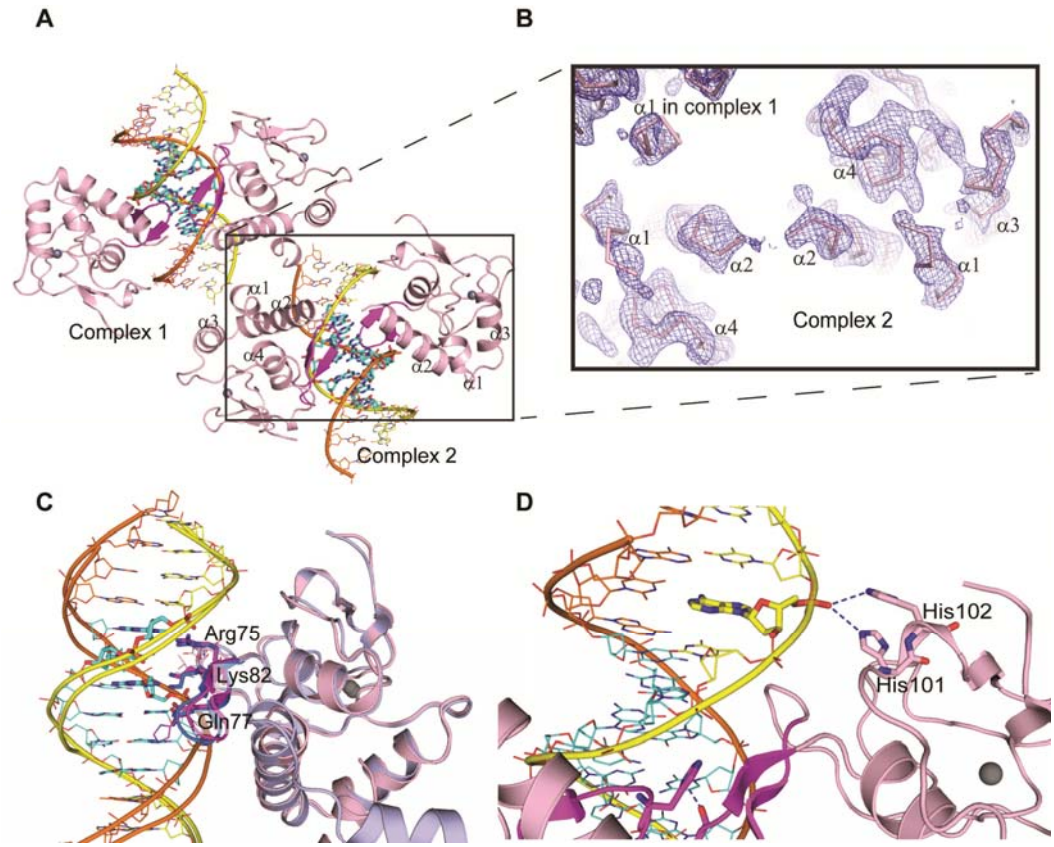


**Figure S3.** (**A**) The arrangement of two 2:1 Smad5-MH1-GC-BRE complexes in an asymmetric unit. The color scheme of Smad5-MH1 and DNA duplexes follows that in Figure 3A. (**B**) Close-up view of the 2*Fo-Fc* electron density of helix α1 in the Smad5-MH1-GC-BRE structure (contoured at 1.0 σ). (**C**) His101 and His102 in Smad5-MH1 contact the DNA phosphate of the flanking sequence around the central GC-rich site. (**D**) The 1:1 Smad5-MH1-SBE complex is superimposed with the 1:1 Smad5-MH1-GC-BRE complex, displaying the differences in the conformation of the three conserved β-hairpin residues.

## Supplementary Figure 4

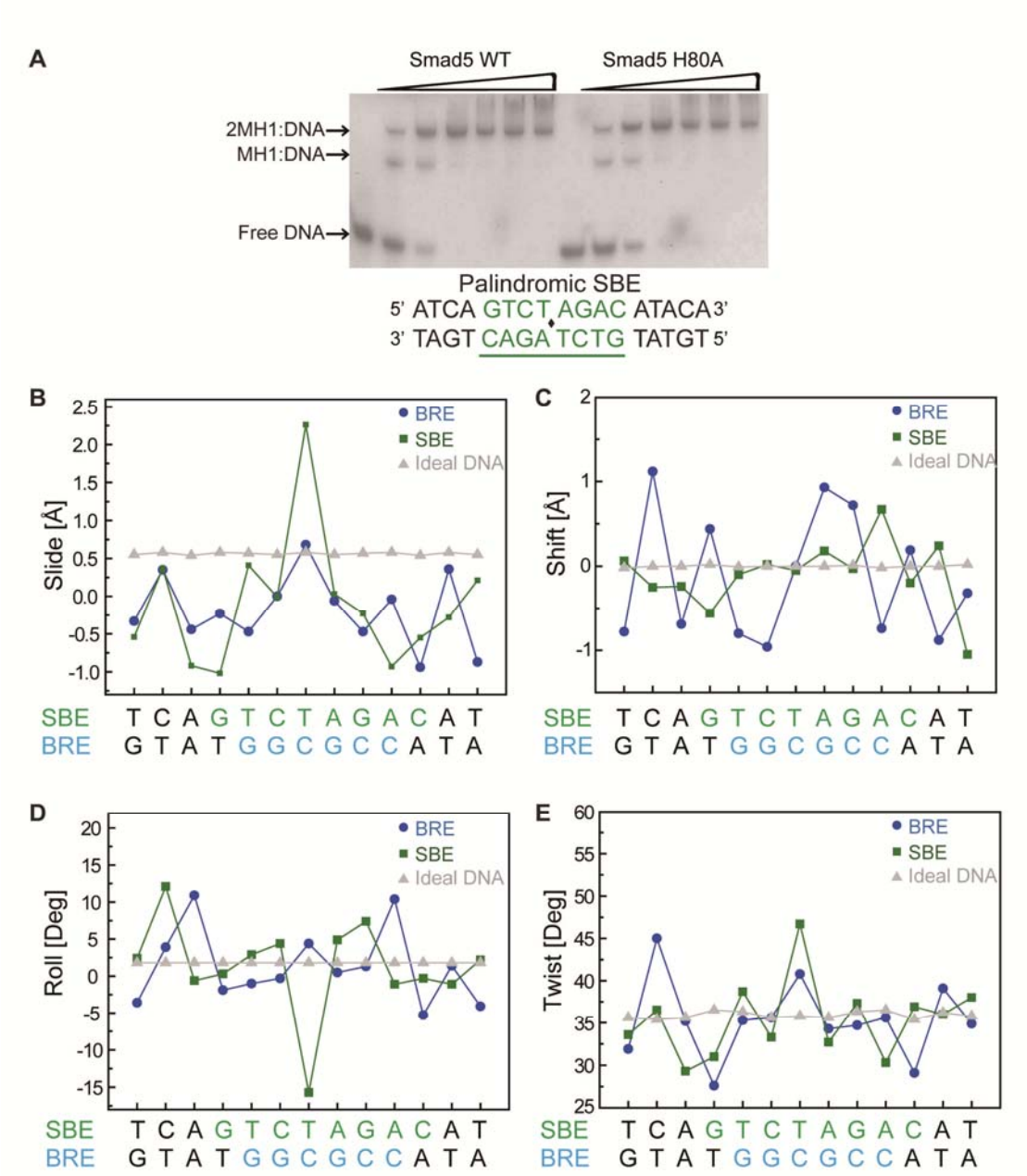


**Figure S4.** (**A**) EMSA of Smad5-MH1 wild type (left panel of the gel) and H80A mutant (right panel of the gel) bound to 1 μM of palindromic SBE DNA. Concentrations of the Smad5-MH1 proteins were 0, 0.5, 1, 2, 3, 4, 5 μM (from left to right in each panel of the gel). (**B, C, D, E**)The inter-bp step parameters of the SBE (green square), GC-BRE (blue circle) DNA bound with Smad5-MH1 and the ideal B DNA (grey triangle) calculated using Curves+. Representative parameters are listed, including (**B**) slide, (**C**) shift, (**D**) roll, (**E**) twist.

## Supplementary Figure 5

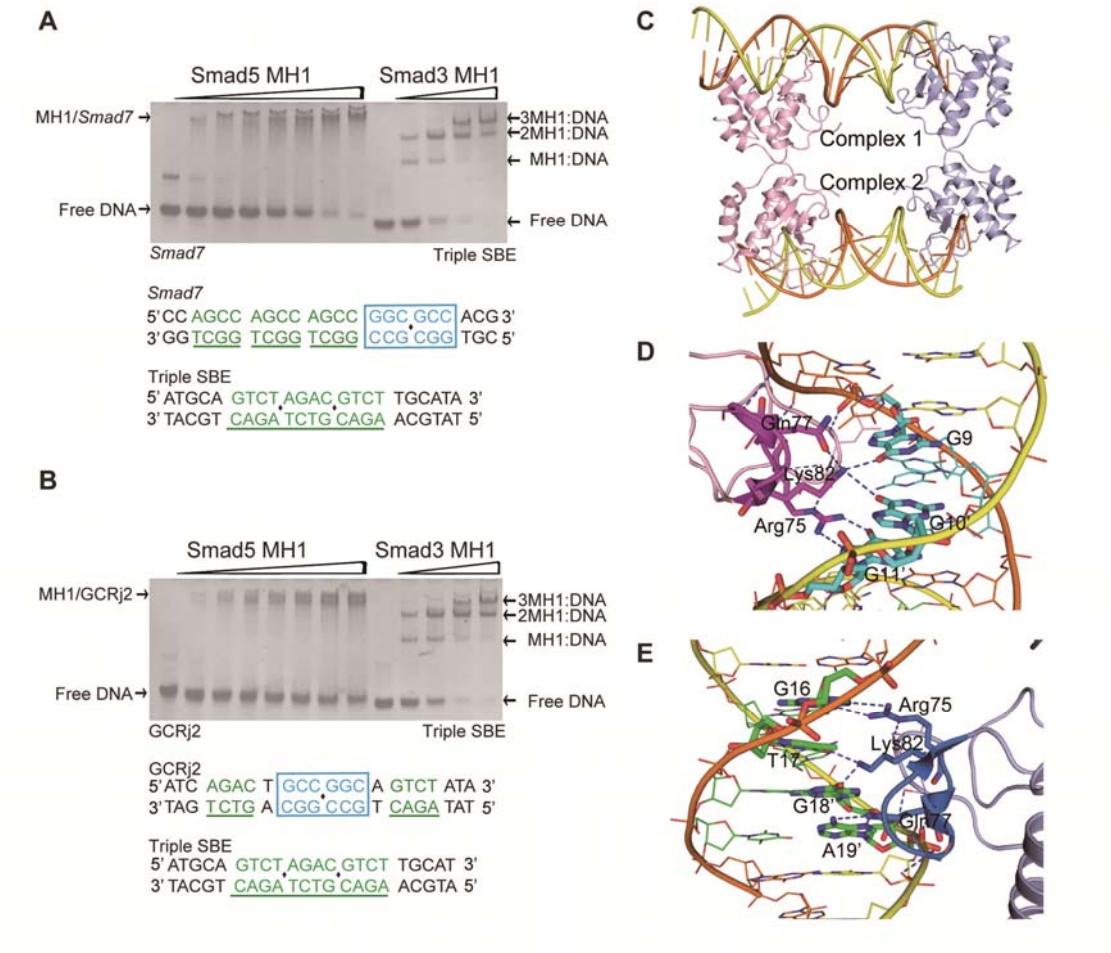


**Figure S5.** (**A**, **B**) Binding of Smad5-MH1 with a DNA sequence from mouse *Smad7* gene promoter (+18/+40) (**A**) and the GCRj2 DNA (**B**). The complexes of Smad3-MH1 binding to the Triple SBE DNA sequence (right panel of the gel) were used as molecular weight markers. The DNA concentration was 1 μM, while the protein concentrations of Smad5-MH1 were 0, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2 μM (from left to right in the left panel of each gel). The DNA sequences used for EMSA are shown as that in Figure 1C. (**C**) Formation of the 4:2 protein-DNA complex in an asymmetric unite of the Smad5-MH1-GCRj2 structure. The Smad5 MH1 proteins and dsDNA duplexes are coloured as that in Figure 5A. (**D, E**) Detailed illustrations of the interactions of Smad5-MH1 with (**D**) Site2 (a single GC-rich site) and (**E**) Site 4 (a single SBE site) in the GCRj2 DNA. The amino acids and DNA bases in the recognitions of Site 2 and Site 4 are shown as that in Figure **3C** and Figure **2C**, respectively.

## Supplementary Figure 6

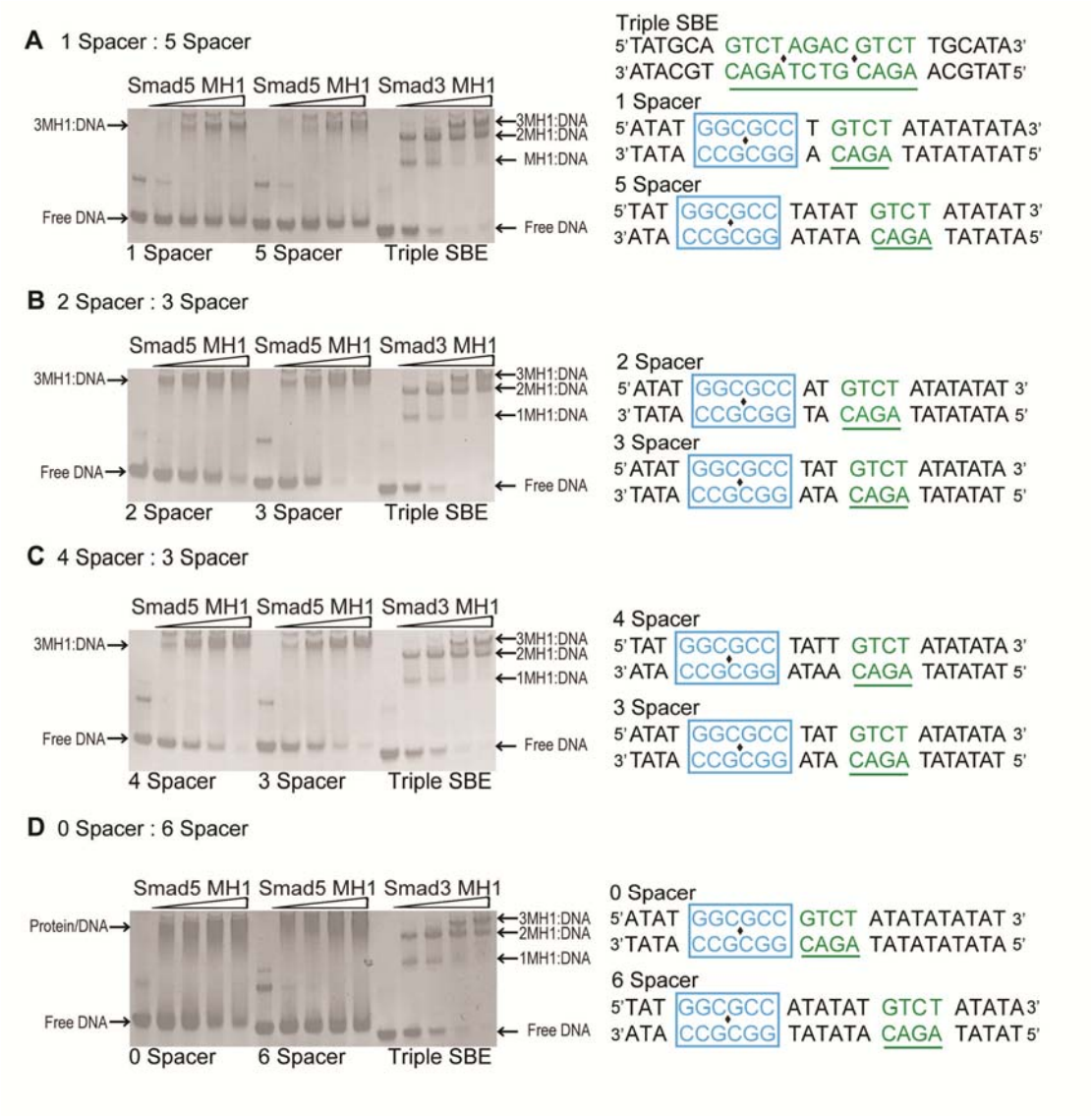


**Figure S6.** Representative electrophoretic mobility shift assays of the Smad5-MH1 proteins binding to the Id2 DNA mutants with different spacer-length between the bipartite elements. These different spacer-length includes (**A**) 1 and 5 bps (the native one), (**B**) 2 and 3 bps, (**C**) 4 and 3 bps, (**D**) 0 and 6 bps. The complexes of Smad3-MH1 binding to the Triple SBE DNA sequence (right panel of the gel) were used as molecular weight markers. The sequences used for EMSA are listed beside the gel. The protein concentrations used for Smad5-MH1 in each assays were 0, 0.6, 1.1, 1.6, 2.1 μM (from left to right in each panel of the gel). The free DNA and the 3MH1:DNA complex were quantified using Quantity One software (Bio-Rad). And the fraction of DNA bound at the protein concentration of 2.1 μM is plotted against the linker length, generating the bar graph in Figure **6B.**

**References:**


1. Seoane, J., Le, H.V., Shen, L., Anderson, S.A. and Massagué, J. (2004) Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell*, **117**, 211-223.
2. Jonk, L.J., Itoh, S., Heldin, C.H., 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.
3. Lien, C.L., McAnally, J., Richardson, J.A. and Olson, E.N. (2002) Cardiac-specific activity of an Nkx2-5 enhancer requires an evolutionarily conserved Smad binding site. *Dev. Biol.*, **244**, 257-266.
4. Jin, W., Takagi, T., Kanesashi, S.N., Kurahashi, T., Nomura, T., Harada, J. and Ishii, S. (2006) Schnurri-2 controls BMP-dependent adipogenesis via interaction with Smad proteins. *Dev. Cell*, **10**, 461-471.
5. López-Rovira, T., Chalaux, E., Massagué, J., Rosa, J.L. 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.
6. Saari, G. and Bienz, M. (1987) The structure of the Ultrabithorax promoter of Drosophila melanogaster. *EMBO J.*, **6**, 1775-1779.
7. Kim, J., Johnson, K., Chen, H.J., Carroll, S. and Laughon, A. (1997) Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. *Nature*, **388**, 304-308.
8. Szüts, D., Eresh, S. and Bienz, M. (1998) Functional intertwining of Dpp and EGFR signaling during Drosophila endoderm induction. *Genes Dev.*, **12**, 2022-2035.
9. Liang, H.L., Xu, M., Chuang, Y.C. and Rushlow, C. (2012) Response to the BMP gradient requires highly combinatorial inputs from multiple patterning systems in the Drosophila embryo. *Development*, **139**, 1956-1964.
10. Ishida, W., Hamamoto, T., Kusanagi, K., Yagi, K., Kawabata, M., Takehara, K., Sampath, T.K., Kato, M. and Miyazono, K. (2000) Smad6 is a Smad1/5-induced smad inhibitor. Characterization of bone morphogenetic protein-responsive element in the mouse Smad6 promoter. *J. Biol. Chem.*, **275**, 6075-6079.
11. Müller, B., Hartmann, B., Pyrowolakis, G., Affolter, M. and Basler, K. (2003) Conversion of an extracellular Dpp/BMP morphogen gradient into an inverse transcriptional gradient. *Cell*, **113**, 221-233.
12. Pyrowolakis, G., Hartmann, B., Müller, B., Basler, K. and Affolter, M. (2004) A simple molecular complex mediates widespread BMP-induced repression during Drosophila development. *Dev. Cell*, **7**, 229-240.
13. Weiss, A., Charbonnier, E., Ellertsdóttir, E., Tsirigos, A., Wolf, C., Schuh, R., Pyrowolakis, G. and Affolter, M. (2010) A conserved activation element in BMP signaling during Drosophila development. *Nat. Struct. Mol. Biol.*, **17**, 69-76.
14. von Bubnoff, A., Peiffer, D.A., Blitz, I.L., Hayata, T., Ogata, S., Zeng, Q., Trunnell, M. and Cho, K.W. (2005) Phylogenetic footprinting and genome scanning identify vertebrate BMP response elements and new target genes. *Dev. Biol.*, **281**, 210-226.

15. Yao, L.C., Blitz, I.L., Peiffer, D.A., Phin, S., Wang, Y., Ogata, S., Cho, K.W., Arora, K. and Warrior, R. (2006) Schnurri transcription factors from Drosophila and vertebrates can mediate Bmp signaling through a phylogenetically conserved mechanism. *Development*, **133**, 4025-4034.
16. Karaulanov, E., Knöchel, W. and Niehrs, C. (2004) Transcriptional regulation of BMP4 synexpression in transgenic Xenopus. *EMBO J.*, **23**, 844-856.
17. Korchynskyi, O. and ten Dijke, P. (2002) Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. *J. Biol. Chem.*, **277**, 4883-4891.
18. Nakahiro, T., Kurooka, H., Mori, K., Sano, K. and Yokota, Y. (2010) Identification of BMP-responsive elements in the mouse Id2 gene. *Biochem. Biophys. Res. Commun.*, **399**, 416-421.
19. Shepherd, T.G., Thériault, B.L. and Nachtigal, M.W. (2008) Autocrine BMP4 signalling regulates ID3 proto-oncogene expression in human ovarian cancer cells. *Gene*, **414**, 95-105.
20. Benchabane, H. and Wrana, J.L. (2003) GATA- and Smad1-dependent enhancers in the Smad7 gene differentially interpret bone morphogenetic protein concentrations. *Mol. Cell. Biol.*, **23**, 6646-6661.