A comparative analysis of Smad-responsive motifs identifies multiple regulatory inputs for TGF-β transcriptional activation

Received for publication, June 21, 2019, and in revised form, August 29, 2019 Published, Papers in Press, September 3, 2019, DOI 10.1074/jbc.RA119.009877

Yuka Itoh1, Daizo Koinuma3, Chiho Omata1, Tomohiro Ogami5, Mitsuyoshi Motizuki8, So-ichi Yaguchi9, Takuma Itoh1, Kunio Miyake1, Shuichi Tsutsumi11, Hiroyuki Aburatani*, Masao Saitoh8,**, Kohei Miyazono5, and † Keiji Miyazawa11

From the 1Department of Biochemistry, 3Research Training Program for Undergraduates, 5Department of Social Medicine, and 8,11Center for Medical Education and Science, Graduate School of Medicine, University of Yamanashi, Yamanashi 409-3898, Japan, the 2Department of Molecular Pathology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan, and the 8,11Genome Science Division, Research Center for Advanced Science and Technology (RCAST), The University of Tokyo, Tokyo 153-8904, Japan

Edited by Joel M. Gottesfeld

Smad proteins are transcriptional regulators activated by TGF-β. They are known to bind to two distinct Smad-responsive motifs, namely the Smad-binding element (SBE) (5′-GTCTAGAC-3′) and CAGA motifs (5′-AGCCAGACA-3′ or 5′-TGCTCAGA-3′). However, the mechanisms by which these motifs promote Smad activity are not fully elucidated. In this study, we performed DNA CASTing, binding assays, ChIP sequencing, and quantitative RT-PCR to dissect the details of Smad binding and function of the SBE and CAGA motifs. We observed a preference for Smad3 to bind CAGA motifs and Smad4 to bind SBE, and that either one SBE or a triple-CAGA motif forms a cis-acting functional half-unit for Smad-dependent transcription activation; combining two half-units allows efficient activation. Unexpectedly, the extent of Smad binding did not directly correlate with the abilities of Smad-binding sequences to induce gene expression. We found that Smad proteins are more tolerant of single bp mutations in the context of the CAGA motifs, with any mutation in the SBE disrupting function. CAGA and CAGA-like motifs but not SBE are widely distributed among stimulus-dependent Smad2/3-binding sites in normal murine mammary gland epithelial cells, and the number of CAGA and CAGA-like motifs correlates with fold-induction of target gene expression by TGF-β. These data, demonstrating Smad responsiveness can be tuned by both sequence and number of repeats, provide a compelling explanation for why CAGA motifs are predominantly used for Smad-dependent transcription activation in vivo.

Transforming growth factor-β (TGF-β)2 is a pleiotropic cytokine that plays crucial roles in various biological processes through regulation of growth, motility, and differentiation of target cells. TGF-β signals are mediated through two types of transmembrane receptors, type I and type II. Upon ligand binding, they form a heterotetrameric complex, in which constitutively active type II receptor phosphorylates type I receptor and activates it. The activated type I receptor transmits signals through the Smad signaling pathway and non-Smad signaling pathways (1, 2). In the Smad signaling pathway, the activated type I receptor phosphorylates cytoplasmic effector molecules called Smad2 and Smad3. Smad2 and Smad3 subsequently form a heteromeric complex with Smad4 that is translocated into the nucleus, where they positively or negatively regulate target gene expression through association with transcription regulatory genomic regions, together with other transcription factors as well as coactivators and corepressors (3).

Several DNA elements have been reported to interact with Smad proteins (4). Zawel et al. (5) analyzed the DNA-binding specificity of bacterially expressed Smad3 and Smad4 using a cyclic amplification of selected target (CASTing) method. The consensus Smad-binding sequence was determined to be an 8-bp palindromic sequence, 5′-GTCTAGAC-3′ (Smad-binding element, SBE). Structural analysis indicated that the MH1 domains of Smad3 and Smad4 interact with the major groove of a 4-bp sequence (GTCT/AGAC) within the motif through a protruding β-hairpin structure (6–8). Dennler et al. (9) and Song et al. (10) independently analyzed cis-regulatory elements in the human plasminogen activator inhibitor-1 (PAI-1) promoter and identified TGF-β-responsive elements that function in vivo.

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grants 16H05150 (to K. Miyazawa) and 18K06626 (to Y. I., JSPS Core-to-Core Program “Cooperative International Framework in TGF-β Family Signaling,” the Fukuoka Trust for Medicinal Research (to K. Miyazawa), the Mitsubishi Science Foundation (to K. Miyazawa), and the Terumo Foundation for Life Science and Arts (to K. Miyazawa). The authors declare that they have no conflicts of interest with the contents of this article.

This article contains Tables S1–S7 and Figs. S1-S6.

Raw data of CASTing analysis have been deposited in the DDBJ BioProject database under BioProject accession number DRA007794.

Table S1: The ChIP-seq data reported in this paper have been submitted to the Gene Expression Omnibus (GEO) database under GEO accession number GSE121254.

1 To whom correspondence should be addressed: Dept. of Biochemistry, Graduate School of Medicine, University of Yamanashi, Yamanashi 409-3898, Japan. Tel.: 81-55-273-6784; Fax: 81-55-273-6784; E-mail: kmiyazawa@yamanashi.ac.jp.

2 The abbreviations used are: TGF-β, transforming growth factor-β; SBE, Smad-binding element; DNAP, DNA affinity precipitation; EMSA, electrophoretic mobility shift assay; NMuMG, normal murine mammary gland epithelial; CASTing, cyclic amplification of selected target; CA, constitutively active; PAI-1, plasminogen activator inhibitor type 1.
β-responsive sequences as 5’-AG(C/A)CAGACA-3’ (CAGA motifs) or 5’-AGACAGGTTGT-3’ (9, 10). Jonk et al. (11) examined Smad-binding elements in the junB promoter and concluded that 5’-CAGACA-3’ is responsible for Smad3/Smad4-dependent transcription activation. Badis et al. (12) identified 5’-CCAGACA-3’ as a motif that binds to mouse Smad3 using a protein binding microarray. Through these studies, the sequence 5’-GTCT-3’ and its complementary sequence 5’-AGAC-3’ have been regarded as typical Smad-binding motifs. ChIP-sequencing (ChIP-seq) approaches essentially supported this idea, although SBE was not often found in the regulatory regions of target genes (13, 14). In addition, the activator protein-1–binding motif also interacts with Smad3 (15). Quite recently, 5-bp GC motifs, 5’-GGC(GC)/(CG)-3’, were shown to be high-affinity Smad3/4-binding motifs (16). However, the distinct properties of these Smad-binding sequences are not well understood.

The activated Smad complex usually requires DNA-binding transcription factors, also known as Smad cofactors, to successfully regulate target genes because the affinity of Smad proteins to the Smad-binding sequences is thought to not be sufficiently high. Alternatively, Smad proteins alone can activate transcription when multiple Smad-binding sequences are present in tandem (reviewed in Ref. 17). Jonk et al. (11) reported that four copies of the 5’-CAGACA-3’ sequence are sufficient to induce transactivation but that two copies are not. Johnson et al. (18) examined the ability of abutting SBE half-sites to enhance transcription. They found that a reporter construct with three copies of 5’-GTCT-3’ has a higher enhancer activity than one with two copies. Some of these studies, however, only utilized assay systems in which Smad3 and Smad4 are overexpressed. Therefore, they may not reflect the situation of endogenous Smad proteins. At present, the minimal requirements for endogenous Smad-dependent transcription activation remain to be determined.

Here, we compared the functional properties of the SBE and CAGA motifs and examined requirements for Smad-dependent transcription activation. We found that either one SBE or a triple CAGA motif acts as a cis-acting functional half-unit for transcription activation; they are, by themselves, inactive but the combination of two half-units enables gene expression. Importantly, the abilities of Smad-binding sequences to induce transcription do not reflect their abilities to interact with Smad proteins. In an electrophoretic motility shift assay, active and inactive DNA sequences yielded band shifts of differing mobilities. We also found that several CAGA-like motifs with single-base replacement are partially active in inducing gene expression, whereas only the complete SBE is active. The number of CAGA and CAGA-like motifs in Smad2/3-binding sites surrounding TGF-β target genes, detected using ChIP-seq analysis of normal murine mammary gland epithelial (NMuMG) cells, correlated well with the fold-induction of target genes after TGF-β stimulation. The functional properties of SBE and CAGA motifs revealed in this study help explain why CAGA motifs are predominantly used to fine-tune Smad-dependent induction of endogenous target genes.

### Properties of Smad-binding motifs

| Smad-binding motifs | TGF-β preference |
|---------------------|------------------|
| SBE                 | Smad3 > Smad4    |
| CAGA                | Smad3 < Smad4    |

### Results

**DNA-binding properties of Smad proteins overexpressed in mammalian cells**

The DNA-binding specificity of Smad3 and Smad4 was previously analyzed by a cyclic amplification of selected target (CASTing) method using bacterially expressed Smad proteins (5). The consensus Smad-binding sequence was found to be an 8-bp palindromic sequence, 5’-GTCTAGAC-3’ (SBE). We applied this strategy to Smad proteins expressed in mammalian cells with an input of TGF-β signaling. In addition, we used a pool with 32-bp degenerated sequences instead of a pool with 20-bp degenerated sequences that were used to select Smad-binding sequences in the previous report (5).

HEK293T cells were transfected with FLAG-tagged Smad3 and constitutively active TGF-β type I receptor (CA-TβRI) or FLAG-tagged Smad4 alone. Thirty hours later, cells were harvested. A nuclear extract was prepared from the harvested cells for CASTing analysis using anti-FLAG antibody. After five cycles of selection, selected DNA fragments were sequenced with a next-generation sequencer. The top 25 concentrated sequences are shown in Table S1. In addition to one or two copies of SBE, a part of the CAGA motifs (5’-CCAGACA-3’ and its reverse complement 5’-TGCTCTGG-3’) was principally concentrated. Thus, we obtained results distinct from those of CASTing analysis using bacterially expressed Smad proteins (5). Notably, recently reported 5-bp GC motifs (16) were less frequently observed.

Table 1 summarizes the occurrence of Smad-binding motifs, SBE, CAGA, and 5-bp GC in Smad3, as well as Smad4-binding sequences. Although both Smad3 and Smad4 interact with SBE and CAGA motifs, Smad3 prefers to bind CAGA motifs, whereas Smad4 prefers SBE. Occurrence ratios of SBE to CAGA motifs were around 1:1.5 for Smad3 and 5:1 for Smad4. We also noticed that Smad3 and Smad4 have different preferences for bases surrounding SBE (Table 2). Both Smad3 and Smad4 prefer 5’-T-GTCTAGAC-A-3’, but Smad3 exhibits a more pronounced preference than Smad4 (Table 2, blue shadowed). In addition, Smad4 prefers 5’-C-GTCTAGAC-A-3’ or its complementary 5’-T-GTCTAGAC-G-3’ (Table 2, orange shadowed). As for bases surrounding the CAGA complementary motif, Smad3 prefers 5’-G-TGCTCTGG-T-3’, but Smad4 does not exhibit a clear preference (data not shown).

### Designing a new series of reporters containing SBE or CAGA motifs

We went on to compare the properties of each element in TGF-β–induced transcription activation, using a new series of
luciferase reporters with either SBE or CAGA motifs. We designed a reporter with three tandem repeats of either SBE or the CAGA motif (Fig. 1A), taking the following three points into account: 1) locations of the common core motifs, 5′-GTCT-3′ or 5′-AGAC-3′, were kept constant; 2) preferences of Smad proteins for bases surrounding SBE (Table 2) were considered; and 3) base replacement between constructs was made minimal. These constructs were designated as SSS-Luc, CCC-Luc, and CcCcCc-Luc, respectively.

SSS-Luc exhibited higher activity than 4× SBE-Luc (5) but lower than (CAGA)12-MLP-Luc (9) in NMuMG cells, as measured using fold-induction by TGF-β (Fig. 1B). In particular, three tandem repeats of the CAGA motifs (5′-AGCCAGACA-3′, CCC-Luc; 5′-TGCTGGCTGCT-3′, CcCcCc-Luc) exhibited only marginal activity (Fig. 1B). Similar results were obtained using an immortalized human keratinocyte cell line, HaCaT (data not shown).

A DNA probe with three tandem repeats of the CAGA motif (“3×CAGA” probe) has been shown to interact efficiently with the Smad complex and has been widely used for DNA affinity precipitation (DNAP) assays (19). Because the spacing between the CAGA motifs in the 3×CAGA probe is different from that in CCC-Luc, we also constructed a reporter containing the 3×CAGA sequence, namely 3×CAGA-Luc (Fig. 1A). However, this construct failed to exhibit activity in response to TGF-β stimulation (Fig. 1B). Thus, 3×CAGA can bind to Smad proteins, but it is inactive as a cis-regulatory element.

The abilities of DNA sequences to cis-activate transcription do not reflect their binding affinities to Smad proteins

We synthesized a biotinylated SSS sequence and investigated its binding to Smad proteins. NMuMG cells were stimulated with TGF-β for 1 h before being harvested. Cell lysates were subjected to DNAP assay. SSS precipitated Smad3 as well as Smad4 from cells stimulated with TGF-β (Fig. 2A), which was outcompeted by an excess of nonlabeled SSS but not by its mutant (SmSmSm)( Sm;5′-CTGATCAG-3′) (Fig. 2B). The DNA probe precipitated only a small fraction of phosphorylated Smad2, but a significant portion of phosphorylated Smad3 (Fig. S1, “unbound and input”), suggesting that it interacts efficiently with the Smad complex composed of phospho-Smad3 and Smad4. Other probes, 3×CAGA and CCC, also precipitated Smad proteins (Fig. 2A), although they lack transcription induction activity.

Intriguingly, Smad proteins bound to biotinylated SSS were outcompeted by 3×CAGA, and those bound to biotinylated

### Table 2
Distinct preferences of Smad3 and Smad4 for bases surrounding SBE

| Sequences       | occurrence among total SBE reads (%) | Smad3 | Smad4 |
|-----------------|--------------------------------------|-------|-------|
| 5′-GTCTAGAC-A   | 9.2                                  | 11.7  |
| 5′-GTCTAGAC-G   | 0.5                                  | 3.6   |
| 5′-GTCTAGAC-C   | 0.2                                  | 0.4   |
| 5′-GTCTAGAC-T   | 0.5                                  | 2.1   |
| 5′-GTCTAGAC-A   | 1.0                                  | 1.5   |
| 5′-GTCTAGAC-G   | 0.1                                  | 0.9   |
| 5′-GTCTAGAC-C   | 0.0                                  | 0.1   |
| 5′-GTCTAGAC-T   | 0.2                                  | 0.4   |
| C- GTCTAGAC-A   | 6.4                                  | 15.3  |
| C- GTCTAGAC-G   | 0.6                                  | 4.9   |
| C- GTCTAGAC-C   | 0.1                                  | 0.9   |
| C- GTCTAGAC-T   | 0.6                                  | 3.9   |
| T- GTCTAGAC-A   | 61.6                                 | 2.4   |
| T- GTCTAGAC-G   | 7.1                                  | 16.2  |
| T- GTCTAGAC-C   | 1.2                                  | 1.5   |
| T- GTCTAGAC-T   | 10.4                                 | 12.6  |
$3\times$CAGA were outcompeted by SSS (Fig. 2C). We also examined serial precipitation using SSS and $3\times$CAGA probes. After depleting Smad proteins that interact with either SSS or the $3\times$CAGA probe, the other probes precipitated barely anymore Smad proteins (Fig. 2D), suggesting that both probes interact with mostly overlapping populations of Smad proteins.
Properties of Smad-binding motifs

We found that both the active cis-regulatory element SSS and the inactive 3′×CAGA or CCC interact in a similar fashion with the Smad complex in DNAP assay. When the Smad complex interacts with the SSS sequence, it activates gene expression, but when it interacts with 3′×CAGA, CCC, or CcCcCc sequence, it fails to activate gene expression. Smad-bound DNA sequences thus affect the ability of the Smad complex to transactivate gene expression.

Distinct properties of SBE and CAGA in relationship to transcription activation

Although three tandem repeats of the CAGA motif (3′×CAGA, CCC, or CcCcCc, referred to henceforth as the triple-CAGA motif) were insufficient for TGF-β–induced transcription activation, a 9 or 12 repeat of the CAGA motif is widely used as a highly sensitive reporter to monitor TGF-β signaling (9). Next, we asked how many repeats of CAGA motif or SBE are required for efficient transcription activation. We found that at least five to six CAGA repeats are required for greater than 10-fold activation, whereas two SBEs are enough (Fig. 3A). One SBE is inactive probably because it only has weak affinity for the Smad complex as shown by DNAP assay (Fig. 3B), whereas the triple-CAGA motif is inactive although it interacts efficiently with Smad proteins. Furthermore, more than four tandem repeats of SBE appeared to suppress transcription activation (Fig. 3C and Fig. S2). Although the underlying mechanism remains to be elucidated, similar effects of binding site number on transcription activity have been reported for some yeast transcription factors (20). Similar results were obtained using HaCaT and a human lung adeno-carcinoma cell line, A549 (data not shown). Therefore, the CAGA motif is additive in nature but SBE has an optimum number of repeats when acting as a cis-regulatory element.

Next, we designed a series of composite reporter constructs composed of two SBEs and single-CAGA motifs (C or Cc). Substituting SBE with C or Cc affected reporter activities depending on the relative positions and orientations (Fig. 3D). Substituting 3′-SBE with Cc (SSCc-Luc) enhanced the reporter activity, whereas doing the same with 5′-SBE (CcSS-Luc) severely compromised it. Substitution with C resulted in weak to moderate inhibition, whereas C is preferred in the 5′-position. Intriguingly, sequences surrounding the core motifs, 5′-GTCT-3′ or 5′-AGAC-3′, play a substantial role in transcription activation because mutation of either Cc (5′-TGCTGGCT-3′) to Cc3m (5′-TGCTTTAGAC-3′) or C (5′-AGCCAGACA-3′) to C3m (5′-CTGAAGACA-3′) halved the activity level (Fig. 3E). Thus, the presence of the core motifs, 5′-GTCT-3′ or 5′-AGAC-3′, alone is not enough. Instead, the core motifs are enhanced by adjacent sequences in C and Cc.

Furthermore, we constructed reporters containing either one SBE and two C or Cc motifs. We found that most of these reporters exhibited weak activities (Fig. S3). We also observed a preference for Cc in the 3′-position and for C in the 5′-position in these constructs.

These findings indicate that SBEs and CAGA motifs have distinct properties with respect to transcription activation. We noticed that two SBEs are comparable in their potency to two tandem triple-CAGA motifs in this type of reporter constructs, whereas both one SBE and a triple-CAGA motif fail to induce transcription (Fig. 3A). This observation suggests that both a single SBE and a triple-CAGA motif can act as half-units for transcription activation.

A single SBE is interchangeable with a triple-CAGA motif in Smad-dependent transcription activation

Next, we determined whether a single SBE in SS-Luc can be replaced by a triple-CAGA motif. We either substituted 3′-SBE with one to three Cc motifs, or 5′-SBE with one to three C motifs, whereas taking positional effects into account. As we anticipated, one SBE can be successfully replaced by three tandem Cc or C elements (Fig. 4A). Reporter activity was maximal when the orientation of all three individual CAGA motifs within a triple-CAGA motif is the same (Fig. 4B). At least two consecutive motifs must be oriented in the same direction for efficient transcription activation.

The artificial reporters described above lack binding sites for other transcription factors and are possibly driven by Smad transcription factors only. Endogenous Smad proteins, however, often have to cooperate with DNA-binding transcription factors known as Smad cofactors, when they transactivate endogenous target genes (reviewed in Ref. 21). Nakano et al. (22) previously reported the construction of an artificial reporter composed of three tandem CAGA motifs and a TCF7L2-binding element (TTE). Because three tandem repeats of either the CAGA motif or TTE alone failed to induce transcription, the reporter’s activity appears to depend on the cooperation between Smad proteins and TCF7L2. We modified the reporter construct for further analysis, which is henceforth denoted as CCCT-Luc. We truncated three nucleotides from the 5′ end of the original version, and verified that both elements are required for transcription activation (Fig. 4C). We reduced the number of CAGA motifs in the CCCT-Luc construct. CCT-Luc exhibited moderate activity, whereas CT-Luc was almost inactive. Next, we replaced the CCC sequence with one SBE (ST-Luc) and determined its activity. As we predicted, ST-Luc exhibited strong transcription activation, similarly to the CCCT-Luc (Fig. 4C). Thus, a single SBE appears to be as
potent as a triple-CAGA motif in this reporter system. Similar results were obtained using HaCaT cells (data not shown).

Our hypothesis that a single SBE and a triple-CAGA motif are equally adept at inducing transcription was verified by two independent reporter systems, with or without a Smad cofactor-binding site. These findings indicate that both a single SBE and a triple-CAGA motif form a cis-acting half-unit required for transcription activation. By themselves, they are inactive,
but they can cooperate with either Smad-binding or Smad cofactor-binding sequences. Interestingly, a double-CAGA motif appears to be a partially active half-unit. However, both the single SBE and triple-CAGA motifs have distinct Smad-binding properties; the former is unable to bind to Smad proteins by itself, whereas the latter can interact efficiently with Smads on its own.

**Transcriptionally active and inactive Smad-binding sequences are bound by distinct Smad complexes**

The next question we posed concerned the crucial difference between active and inactive Smad-binding sequences. To determine the nature of this difference, we measured the binding of Smad proteins to both active and inactive sequences using an electrophoretic mobility shift assay (EMSA) with fluorescent probes, inactive S, 1×CAGA, 3×CAGA, and active SSS.

We prepared and used nuclear extracts from 293T cells over-expressing full-length Smad3 and Smad4 with CA-TβRI, and extracts from mock-transfected cells as a control (Fig. 5A). Inactive 1×CAGA and 3×CAGA as well as active SSS yielded shifted bands. By contrast, S did not yield shifted bands although it did shift weakly when either bacterially expressed Smad3 or Smad4 MH1 domain was used (Fig. 5B). Importantly, compared with inactive 1×CAGA and 3×CAGA, active SSS yielded a slowly migrating shifted band, which appears to represent a DNA probe bound to transcriptionally active Smad complex. Furthermore, we verified the compositions of Smad proteins in the shifted bands by incubating the gel-shift mixture with either anti-Smad2/3 or anti-Smad4 antibodies (Fig. 5B). Incubation with normal IgG did not affect the SSS-shifted band. Incubation with either anti-Smad2/3 or anti-Smad4 resulted in a supershift of the SSS-shifted band, indicating that the SSS-shifted band is composed of Smad3 and Smad4. Similar results were obtained when we used 3×CAGA as a probe. Notably, the SSS-shifted band was outcompeted with nonlabeled 3×CAGA probe and vice versa (Fig. 5C). These results are consistent with those from DNAP assays and again suggest that both probes interact with mostly overlapping populations of Smad proteins. Thus, Smad complexes are likely to be converted to the transcriptionally active form upon interaction with SSS but not with 3×CAGA.

**Usage of CAGA and CAGA-like motifs in TGF-β–induced transcription activation of endogenous target genes**

Our final aim was to determine how the properties of SBE and CAGA are reflected in transcriptional regulation of endogenous TGF-β target genes. Therefore, we examined the occurrence of SBE as well as CAGA motifs in Smad2/3-binding sites surrounding TGF-β target genes. NMuMG cells were stimulated with TGF-β for 1.5 h, harvested, and subjected to determination of Smad2/3-binding sites by ChIP-seq (GSE121254). We detected 5,847 peaks throughout the genome.

The CAGA motif was reported as a 9-bp element, 5′-AG(C/A)CAGACA-3′ (or its complementary 5′-TGTAAG(G/T)CT-3′). However, the effects of base replacement have not been investigated thus far. We utilized SccCcCc-Luc as a platform and replaced one of the bases with others and measured corresponding reporter activities (Table 3). The last two positions affected the activity less, whereas the other positions significantly affected the activity. In addition to 5′-TGTAAGG-3′ (Cc) and 5′-TGTAAGTGT-3′ (Cc-GTGT) that are originally reported (9), we selected 5′-TGTAAGTAG-3′ (Cc-G6A), 5′-TGA-CCTGG-3′ (Cc-T3A), and 5′-CGTAAGGTC-3′ (Cc-T1C), as active CAGA-like motifs (with activities more than one-third of that of the original motif). Similarly, the base requirement in SBE was determined using SS-Luc as a platform. However, any base substitutions resulted in a drastic drop in activities down to less than 10% (Table S2), indicating that the base requirement in
SBE is strict. Therefore, we only examined the occurrence of the complete SBE sequence. Among 5,847 Smad2/3-binding sites, 3,818 had at least one SBE, CAGA, or CAGA-like motifs in 1,000-bp length genomic sequences flanking the middle position of the Smad2/3-binding sites. The incidence of each element is summarized in Table S3.

SBE was not frequently observed in Smad2/3-binding sites, whereas the CAGA and CAGA-like motifs, especially Cc-G7T, as well as 5-bp GC motifs were commonly found.

We next examined the roles of these elements in regulating target gene expression. We previously performed DNA microarray analysis of TGF-β target genes in NMuMG cells (23) (GSE46405). Seventy-nine genes were significantly expressed (log₂ intensity >7) and induced by TGF-β stimulation (1 h) more than 1.6-fold (Table S4). Among these, 67 genes had a total of 239 Smad2/3-binding sites located within 50 kb upstream of the annotated transcription start site or downstream of the transcription termination site. We selected 15 representative target genes with various numbers of Smad-binding motifs, and determined their expression after TGF-β stimulation for 1.5, 3, and 8 h (Fig. 6A and Fig. S5). We observed a positive correlation (R around 0.7) between the number of CAGA and CAGA-like motifs and fold-induction after TGF-β stimulation (3 or 8 h), which correlated less well with the number of CAGA motifs only, and not with that of 5-bp GC motifs (Fig. 6B and C). Thus, CAGA and CAGA-like motifs appear to play a central role in Smad-dependent transcription activation in vivo, and SBE is only rarely utilized although Smad proteins have a high intrinsic affinity for this sequence motif.

Discussion

TGF-β signaling is principally transmitted by Smad2 and Smad3, which are phosphorylated by TGF-β type I receptor.

### Table 3

Effects of single-base pair substitution in the CAGA complementary motif on reporter activities

Relative luciferase reporter activities of SccCccCc-Luc constructs with single-base-pair substitution in the CAGA complementary motifs were measured in NMuMG cells after TGF-β stimulation for 18 h. Averages of three or four independent determinations are shown. Blue-shaded cells denote activity of the original motif, and orange-shaded cells denote activities more than one-third of that of the original motif.

| CAGA complementary motif | T | G | T | C | T | G | G/T | C | T |
|-------------------------|---|---|---|---|---|---|-----|---|---|
| A                       | 0.08 | 0.02 | 0.35 | 0.05 | 0.02 | 0.77 | 0.23 | 0.53 | 0.50 |
| G                       | 0.05 | 1.00 | 0.05 | 0.08 | 0.04 | 1.00 | 1.00 | 1.01 | 0.56 |
| C                       | 0.34 | 0.03 | 0.11 | 1.00 | 0.28 | 0.10 | 0.18 | 1.00 | 0.49 |
| T                       | 1.00 | 0.02 | 1.00 | 0.07 | 1.00 | 0.16 | 0.35 | 0.59 | 1.00 |

### Figure 5

Detection of Smad complexes formed on transcriptionally active or inactive cis-regulatory elements in electrophoretic motility shift assay. A, 293T cells were transiently transfected with constructs encoding FLAG-Smad3, FLAG-Smad4, or CA-TβRI-HA. Thirty hours later, nuclear proteins were extracted and subjected to EMSA using the indicated Alexa 633-labeled probes. B, supershift assay. EMSA was performed with the indicated Alexa 633-labeled probes in the presence of normal IgG, anti-Smad2/3, or anti-Smad4 antibodies. C, cross-competitive binding assay. EMSA was performed with the indicated Alexa 633-labeled probes in the presence of 60-fold excess of each unlabeled probe. Asterisks indicate nonspecific bands that were also observed in mock samples.
Properties of Smad-binding motifs

|   | TGF-β (h) | Serpine1 | Snai1 | Pmea1 | Skil | Ctgf | Bmf | Ski | Klf10 | Sox4 | Gadd45g | Smad7 | Trib1 | Junb | Ski1 | Smad6 |
|---|-----------|----------|-------|-------|------|------|-----|-----|------|------|---------|-------|-------|------|------|-------|
| TGF-β (3 h) | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
| R=0.69 | p=0.005 | | | | | | | | | | | | | | | |
| TGF-β (8 h) | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
| R=0.70 | p=0.004 | | | | | | | | | | | | | | | |

Figure 6. Fold-induction of TGF-β target genes is correlated with occurrence of CAGA and CAGA-like motifs in Smad2/3-binding sites located within 50 kb from the genes. A, heat map of the TGF-β-induced gene expression. NMuMG cells were stimulated with TGF-β for 1.5, 3, or 8 h and expression of the selected target genes was determined by quantitative RT-PCR. The genes were arranged in descending order at 3 h. B and C, correlation analysis between occurrence of CAGA, CAGA/CAGA-like motifs, or 5-bp GC motifs and fold-induction of target genes by TGF-β at 3 h (B) and 8 h (C). All p values were two sided.

Smad2 and Smad3 phosphorylation leads to formation of a trimeric complex with Smad4 (co-Smad) to cooperatively regulate target gene expression in the nucleus. The DNA-binding affinity of the activated Smad complex is not high enough to support gene expression by itself, and additional DNA-binding transcription factors (Smad cofactors) are usually required for stable association of the Smad complex with the regulatory regions of target genes. Alternatively, when multiple Smad-binding sequences are present, Smad proteins do not require Smad cofactors (17).

Thus far, several DNA motifs have been known as Smad-binding sequences. Two prominent examples of such motifs are 5′-GTCTAGAC-3′ (SBE) and 5′-AG(C/A)CAGACA-3′ or its complementary 5′-TGTCG(T/G)CT-3′ (CAGA motifs). These sequences share 5′-GTCT-3′ or its complementary 5′-AGAC-3′ motifs. However, functional properties of these two motifs have not been examined in depth. In addition, the reason why CAGA motifs are preferentially utilized in regulation of endogenous gene expression (Ref. 13 and this study) has remained to be clarified.

Smad-binding to a triple-CAGA motif is not sufficient to induce gene expression

Because Smad2 and Smad3 have the ability to trans-activate gene expression when fused to the Gal4 DNA-binding domain (24), it appeared likely that binding to the target gene promoters would be sufficient to activate gene expression. Johnson et al. (18) also reported that binding of Smad proteins to cis-regulatory elements is well correlated with transactivation. By contrast, Yingling et al. (25) described a discrepancy between Smad-binding and Smad-dependent transcription activation and discussed that Smad-binding sites may function only in the appropriate promoter context. In addition, Labbé et al. (26) found that both Smad2 and Smad3 form a stable complex with Smad4 and FoxH1 on the goosecoid promoter; however, the same group also found that only the complex containing Smad2 is active.

In this study, we observed that a reporter construct containing 3XCAGA failed to respond to TGF-β stimulation although 3XCAGA interacts efficiently with Smad proteins (Figs. 1B and 2A). By contrast, a single SBE failed to interact with Smad and to induce gene expression. Only two tandem SBE repeats are sufficient for both Smad-binding and transcription activation (Fig. 3, A and B). Zawel et al. (5) previously reported that “2×SBE” fails to induce transcription in response to TGF-β stimulation. Consistently, we found that the Smad-binding affinity of 2×SBE is much weaker than that of SS (Fig. S6). A 2×SBE has four bases (5′-GGCA-3′) between two SBEs, whereas SS has three bases (5′-AAT-3′) (Table S6). Both the number and sequence of nucleotides between two SBEs appear to affect their activities. Therefore, there is a discrepancy between DNA-binding affinity and transcription activation in the case of CAGA-containing sequences.

We also investigated the minimum requirement for the CAGA-motif–dependent transcription activation and found that at least five or six repeats are required for efficient activation. Song et al. (10) examined the potency of tandem repeats of sequence 5′-AGACAGGTTGTT-3′, which is derived from the human PAI-1 promoter region (−720/−708). Four repeats of the sequence induced reporter activity 4-fold, whereas six repeats induced reporter activity 40-fold following TGF-β stimulation. Although the sequence is different from the typical CAGA motif, the core 5′-AGAC-3′ motif is shared. The sequence therefore appears to behave similarly to the CAGA motif.
From the results described above, we noticed that both a single SBE and the triple-CAGA motif behave similarly in transcription activation, although the latter can interact with Smad proteins, whereas the former fails to do so (Fig. 7A). Because neither is sufficient to mediate Smad-dependent transcription by itself, we propose that a single SBE and the triple-CAGA motif are transcriptionally inactive half-units that can induce Smad-dependent transcription activation when they are combined with another half-unit or possibly, a Smad cofactor site(s).

The next important question concerns the critical differences between Smad complex formed on inactive and active cis-regulatory elements. Active SSS and inactive 3×CAGA sequences exhibited similar behavior in DNAP assays (Fig. 2, C and D), but they behaved differently in EMSA (Fig. 5A). Notably, both bands were outcompeted by either SSS or 3×CAGA, indicating that DNA sequences apparently do not distinguish between active and inactive complexes. There are two possibilities to explain the difference between active and inactive complexes.

One possibility is that Smad complexes with the same composition bind to both 3×CAGA and SSS, but only those that bind to the latter can adopt the specific conformation that allows them to recruit additional component(s) required for efficient transcription activation (Fig. 7B). It was previously reported that p300, a histone acetyltransferase, and SMARCA4/Brg1, a component of the SWI/SNF chromatin remodeling complex, must be recruited to the Smad complex to modify chromatin templates, a requirement for the Smad complex to assemble basic transcription

**Properties of Smad-binding motifs**

From the results described above, we noticed that both a single SBE and the triple-CAGA motif behave similarly in transcription activation, although the latter can interact with Smad proteins, whereas the former fails to do so (Fig. 7A). Because neither is sufficient to mediate Smad-dependent transcription by itself, we propose that a single SBE and the triple-CAGA motif are transcriptionally inactive half-units that can induce Smad-dependent transcription activation when they are combined with another half-unit or possibly, a Smad cofactor site(s).

The next important question concerns the critical differences between Smad complex formed on inactive and active cis-regulatory elements. Active SSS and inactive 3×CAGA sequences exhibited similar behavior in DNAP assays (Fig. 2, C and D), but they behaved differently in EMSA (Fig. 5A). Notably, both bands were outcompeted by either SSS or 3×CAGA, indicating that DNA sequences apparently do not distinguish between active and inactive complexes. There are two possibilities to explain the difference between active and inactive complexes.

One possibility is that Smad complexes with the same composition bind to both 3×CAGA and SSS, but only those that bind to the latter can adopt the specific conformation that allows them to recruit additional component(s) required for efficient transcription activation (Fig. 7B). It was previously reported that p300, a histone acetyltransferase, and SMARCA4/Brg1, a component of the SWI/SNF chromatin remodeling complex, must be recruited to the Smad complex to modify chromatin templates, a requirement for the Smad complex to assemble basic transcription
Properties of Smad-binding motifs

machinery (27). Thus far, p300 is known to interact with the Smad complex bound to the inactive 3×CAGA sequence (19). We also confirmed it for ourselves (data not shown). Other indispensable components may be excluded from the Smad complex formed on inactive sequences.

Another possible explanation is that binding of two Smad complexes is required for transcription activation, whereas binding of one Smad complex is not enough; the activated Smad complex may be a trans-acting functional half-unit (Fig. 7C). This model is consistent with our finding that both a single SBE and the triple-CAGA motif are cis-acting half-units and that at least two of them are required for transcription activation. The model is also reminiscent of Smad-cofactor-dependent transcription activation (Fig. 7D). In addition, transcription factors generally function in a combinatorial manner to regulate gene expression (28) and p300 is activated by transcription factor dimerization (29). All these arguments support the second model. Currently, there is no direct experimental evidence to support the second model. Further investigation is required to solve this important question.

Distinct functional properties of CAGA and SBE

Both Smad3 and Smad4 interact well with SBE (5′-GTCTTAGAC-3′) although Smad3 prefers CAGA motifs. Nevertheless, CAGA motifs, but not SBE, are widely used for Smad-dependent transcription activation. The reason underlying the preference for the CAGA motif remains to be elucidated. However, the distinct functional properties of SBE and CAGA that we uncovered in this study may help explain this preference for the CAGA motif.

First, DNA-binding appears to be sufficient for SBE-dependent transcription activation, whereas additional steps after DNA-binding is required for CAGA-dependent transcription activation, suggesting that the latter system has more points for functional regulation. Second, the effect of SBE on transcription activation is additive for up to three SBE units but suppressive beyond that number. By contrast, CAGA motifs have an additive effect probably up to 9–12 units. Third, sequence requirement for SBE to behave as a cis-regulatory element is very stringent. SBE thus has an “all-or-none” property. By contrast, CAGA motifs are more flexible and some base replacements are permitted without drastic loss of the activity. This may be because SBE is composed only of the core sequences, 5′-GTCT-3′ and its complementary 5′-AGAC-3′. It is noteworthy that requirement of the core 5′-GTCT-3′ was strict even in CAGA-like motifs (Table 3), except for the second position that was shown to be less involved in the contact with Smad proteins (6–8). Because multiple CAGA motifs exhibit additive effects and because partially active CAGA-like motifs also exist, combining CAGA and CAGA-like motifs would enable the regulation of target gene expression to be fine-tuned.

CAGA and CAGA-like motifs in regulating endogenous gene expression

We noticed that some of the Smad2/3-binding sites lack known Smad-binding elements in 1,000-bp length genomic sequences flanking the middle position. It is possible that Smad2 and Smad3 associate with these sites either indirectly through other DNA-binding transcription factors, as previously reported for Mixer, GATA-3, and Sox9 (30–32), or directly through yet unidentified Smad-binding sequences.

The motif that we observed most frequently in Smad2/3-binding sites surrounding these early target genes (239 sites) was Cc-G7T (134), followed by the CAGA motif (119) (Table S4). Although the Cc-G7T motif exhibits lower potency to activate transcription, it appears to be widely used as a regulatory element. These CAGA/CAGA-like motifs are clustered in some cases, but dispersed or scattered in others. In artificial reporter constructs, the location and orientation of the CAGA motif considerably affect transcriptional activities (Figs. 3D and 4B), probably because of local constraints in the context of the luciferase constructs. It remains unclear how the arrangement of Smad-binding elements in natural promoters affects gene expression. However, within the genomic context, the location and direction of CAGA-like motifs may affect transcriptional activities to a lesser degree because they are placed with sufficient spaces, which enables favorable configuration of each element through looping of DNA.

Notably, the numbers of CAGA/CAGA-like motifs are significantly correlated with fold-induction of target genes at 3 or 8 h, but not 1.5 h, after stimulation. Some of the Smad2/3-binding sites do not appear to regulate transcription during the early phase after stimulation. This can be explained by our finding that Smad binding to DNA via CAGA/CAGA-like motifs is not sufficient for transcription activation. At the later phase, Smad complexes regulate gene expression possibly in cooperation with Smad cofactor(s) that are induced by TGF-β, which is known as the “self-enabling” gene response (33). Alternatively, Smad proteins bound to genomic DNA induce chromatin remodeling in cooperation with SMARCA4/Brg1, which enables expression of some target genes (34). Requirement for such a sequence of events is consistent with our previous observation that Smad-binding early after TGF-β stimulation is significantly correlated with later changes in the induction of target genes (13).

Conclusion

Recently there are a number of studies on cellular responses induced by TGF-β using ChIP-seq analysis. However, molecular events leading to transcription activation including chromatin modifications still remain elusive. In this study, utilizing luciferase reporter model systems, we found distinct functional properties of two representative Smad-binding sequences, SBE and CAGA motifs, which explain why CAGA/CAGA-like motifs are widely used for Smad-dependent transcriptional regulation in vivo. In addition, we presented basic data for the minimal requirements that two half-units together enable successful Smad-dependent transcription activation. Our study has thus provided valuable information to facilitate a better understanding of how Smad proteins are activated to induce the expression of genes that play crucial roles in development and maintenance of the architecture of multicellular organisms.
Experimental procedures

Cell culture

HEK293T and NMuMG cells were obtained from the American Type Culture Collection and maintained in Dulbecco’s modified Eagle’s medium containing 10% fetal bovine serum (FBS), 50 units/ml of penicillin, and 50 μg/ml of streptomycin. For culturing NMuMG cells, the culture medium was supplemented with insulin (10 μg/ml).

Plasmids, antibodies, and ligands

FLAG-tagged Smad3 and Smad4 as well as CA-TβRI receptor were described previously (35, 36). Antibodies used in this study are as follows: anti-FLAG M2 Magnetic Beads (M8823, Sigma); anti-Smad2/3 (610843, BD Bioscience); anti-Smad4 (B-8) (sc-7966, Santa Cruz Biotechnology); anti-phospho-Smad2 (134D4, Cell Signaling Technology); anti-phospho-Smad3 (9520, Cell Signaling Technology); and anti–α-tubulin (DM1A) (T9026, Sigma). Human recombinant TGF-β1 was obtained from R&D Systems.

Cyclic amplification of selected target (CASTing) method

Determination of Smad-binding DNA sequences was performed as described previously (37) with modifications. An oligonucleotide library containing 32 random nucleotides flanked by sequences for PCR amplification (5′-ACTGTACA-CTGCGTCGTTC-(N32)-CTGCTCGAGCGTTCGGATCC-3′) was synthesized and used for screening. HEK293T cells were transiently transfected with FLAG-tagged Smad3 and CA-TβRI or Smad4 alone using PEI Max transfection reagent (Polysciences). Thirty hours later, cells were harvested and a nuclear extract was prepared by NucBuster Protein Extraction Kit (Millipore), followed by CASTing analysis using anti-FLAG M2 magnetic beads. A hundred micrograms of protein was used in each round of selection. Taq polymerase (Takara Bio) was used for DNA amplification. After five rounds of selection, amplified DNA fragments were subjected to sequencing using Ion Torrent and Ion Plus Fragment Library Kit (Life Technologies). Collected sequences were clustered using Aptamer Technologies. Probe sequences are listed in Table S6.

Luciferase assay

Luciferase assay was performed as described previously (36) using TGF-β–responsive reporters, (CAGA)₁₂-MLP-Luc (9), 4×SBE-Luc (5), and other constructs. In brief, NMuMG cells were transfected with the indicated luciferase reporter constructs and pRL-TK vector using PEI Max transfection reagent (Polysciences), stimulated with TGF-β (1 ng/ml) for 18 h, and harvested. Luciferase activities in the lysates were measured using the dual-luciferase reporter system (Promega) and a luminometer (Spectra Max L, Molecular Devices). Values were normalized to those measured for Renilla luciferase under the control of the thymidine kinase promoter. Activities are shown as fold-induction by TGF-β. All the reporter constructs used in this study have the pGL4-MLP backbone (pGL4 vector containing the TATA box and the initiator sequence of the adenovirus major late promoter). Oligonucleotides for artificial reporter constructs were synthesized with flanking KpnI and XhoI sites, and inserted into pGL4-MLP. Sequences for artificial reporter constructs are listed in Table S5.

DNAP

NMuMG cells were stimulated with TGF-β (1 ng/ml) for 1 h and harvested. Cell lysates were prepared with a buffer containing 1% Nonidet P-40, 20 mM Tris–HCl (pH 7.4), 150 mM NaCl, 5 mM EDTA, protease inhibitor mixture (Nacalai Tesque), and phosphatase inhibitor mixture (Nacalai Tesque). DNAP assay was performed as described previously (19) with modifications. Dynabeads M280 streptavidin (10 μl, Life Technologies) were incubated with biotinylated DNAs (30 pmol) for 15 min, followed by washing. Lysates (350 μg protein) were added to DNA-bound Dynabeads in the presence of salmon sperm DNA (80 μg/ml) and incubated overnight at 4 °C. For competition experiments, lysates were preincubated with competitor oligonucleotides for 15 min. Dynabeads were later washed and bound proteins were separated by SDS-PAGE and detected by immunoblotting (36). Sequences for biotinylated probes are listed in Table S6.

EMSA

Alexa 633-labeled probes were purchased from Life Technologies. Probe sequences are listed in Table S6. Nuclear extracts were prepared from 293T cells transfected with expression plasmids for FLAG-Smad3, FLAG-Smad4, and CA-TβRI as described previously (38). Nuclear extracts (40 μg protein) were incubated with probes (4 pmol) for 30 min on ice. Complexes were then resolved on a 4.5% polyacrylamide gel and analyzed by ImageQuant LAS 4000 (GE Healthcare). For supershift assay, antibodies (0.8 μg) and nuclear extracts were preincubated for 1 h on ice.

ChIP-sequencing

Chromatin isolation, sonication, and immunoprecipitation using antibody against Smad2/3 were performed as described previously (13). NMuMG cells were stimulated with TGF-β (1 ng/ml) for 1.5 h in the presence of 10% FBS and harvested. High-throughput sequencing of the ChIP fragments was performed using Illumina Genome Analyzer IIX (Illumina). Unfiltered 36-bp sequence reads were aligned against the mouse reference genome (mm9) using Bowtie2. Peaks were called using CisGenome version 2 (two-sample analysis with an FDR cutoff of 0.05) (39). For SBE and CAGA/CAGA-like motif mapping, 1,000-bp length genomic sequences flanking the middle position of the Smad2/3–binding sites were obtained using CisGenome. Raw ChIP-seq and peak-called data are available at the Gene Expression Omnibus (GSE121254).

Quantitative RT-PCR

Total RNA was extracted using ISOGEN (NIPPON GENE). First-strand cDNA synthesis was performed using PrimeScript 1st Strand cDNA Synthesis Kit and random hexamers (Takara Bio). Quantitative RT-PCR was performed using the StepOne Plus Real-Time PCR System and Power SYBR Green Master Mix (Thermo Fisher Scientific). Mouse Gapdh was used for normalization. Primer sequences are shown in Table S7.
Properties of Smad-binding motifs

Motif enrichment analysis
Motif enrichment analysis (40) was performed using MEME Suite online platform (http://meme-suite.org/index.html) (41).3

Author contributions—Y. I. and K. Miyazawa conceptualization; Y. I. and K. Miyazawa funding acquisition; Y. I., D. K., C. O., T. O., M. M., T. I., and K. Miyake investigation; Y. I. and K. Miyazawa writing—original draft; D. K., T. O., K. Miyake, S. T., H. A., and K. Miyazono data curation; D. K., M. S., and K. Miyazawa writing—review and editing; S.-i. Y. resources; D. K., M. S., and K. Miyazono project administration.

References
1. Hata, A., and Chen, Y. G. (2016) TGF-β signaling from receptors to Smads. Cold Spring Harb. Perspect. Biol. 8, a022061 CrossRef Medline
2. Zhang, Y. E. (2017) Non-Smad signaling pathways of the TGF-β family. Cold Spring Harb. Perspect. Biol. 9, a022129 CrossRef Medline
3. Shi, Y., and Massague, J. (2003) Mechanisms of TGF-β signaling from cell membrane to the nucleus. Cell 113, 685–700 CrossRef Medline
4. Derynick, R., Zhang, Y., and Feng, X. H. (1998) Smads: transcriptional activators of TGF-β responses. Cell 95, 737–740 CrossRef Medline
5. Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B., and Kern, S. E. (1998) Human Smad3 and Smad4 are sequence-specific transcription activators. Mol. Cell 1, 611–617 CrossRef Medline
6. Shi, Y., Wang, Y. F., Jayaraman, L., Yang, H., Massague, J., and Pavletich, N. P. (1998) Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-β signaling. Cell 94, 585–594 CrossRef Medline
7. Chai, J., Wu, J. W., Yan, N., Massagué, J., Pavletich, N. P., and Shi, Y. (2003) Features of a Smads3MH1–DNA complex: roles of water and zinc in DNA binding. J. Biol. Chem. 278, 20327–20331 CrossRef Medline
8. Baburajendran, N., Jauch, R., Tan, C. Y., Narasimhan, K., and Kolatkar, P. R. (2011) Structural basis for the cooperative DNA recognition by Smad4-MH1 dimers. Nucleic Acids Res. 39, 8213–8222 CrossRef Medline
9. Dohner, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S., and Gauthier, J. M. (1998) Direct binding of Smad3 and Smad4 to critical TGFβ-inducible elements in the promoter of human plasminogen activator inhibitor–type 1 gene. EMBO J. 17, 3091–3100 CrossRef Medline
10. Song, C. Z., Siok, T. E., and Gelehrter, T. D. (1998) Smad4/DPC4 and Smad3 mediate transforming growth factor-β (TGF-β) signaling through direct binding to a novel TGF-β-responsive element in the human plasminogen activator inhibitor-1 promoter. J. Biol. Chem. 273, 29287–29290 CrossRef Medline
11. 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-β responsive activator. J. Biol. Chem. 273, 21145–21152 CrossRef Medline
12. Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C. F., Coburn, D., Newburger, D. E., Morris, Q., Hughes, T. R., and Bulyk, M. L. (2009) Diversity and complexity in DNA recognition by transcription factors. Science 324, 1720–1723 CrossRef Medline
13. Koinuma, D., Tsutsuami, S., Kaminura, N., Taniguchi, H., Miyazawa, K., Sunamura, M., Imamura, T., Miyazono, K., and Abaratuni, H. (2009) Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TAFP2A in transforming growth factor β signaling. Mol. Cell Biol. 29, 172–186 CrossRef Medline
14. Ding, N., Yu, R. T., Subramaniam, N., Sherman, M. H., Wilson, C., Rao, R., Leblanc, M., Coulter, S., He, M., Scott, C., Lau, S. L., Atkins, A. R., Barish, G. D., Gunton, J. E., Liddle, C., Downes, M., and Evans, R. M. (2013) A vitamin D receptor/SMAD genomic circuit gates hepatic fibrotic response. Cell 153, 601–613 CrossRef Medline
15. Zhang, Y., Feng, X. H., and Derynick, R. (1998) Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-β-induced transcription. Nature 394, 909–913 CrossRef Medline
16. Martin-Malpartida, P., Batet, M., Kaczmarsza, Z., Freier, R., Gomes, T., Aragón, E., Zou, Y., Wang, Q., Xi, Q., Ruiz, L., Vea, A., Márquez, I. A., Massagué, J., and Macias, M. J. (2017) Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors. Nat. Commun. 8, 2070 CrossRef Medline
17. Ross, S., and Hill, C. S. (2008) How the Smads regulate transcription. Int. J. Biochem. Cell Biol. 40, 383–408 CrossRef Medline
18. Johnson, K., Kirkpatrick, H., Comer, A., Hoffmann, F. M., and Laughon, A. (1999) Interaction of Smad complexes with tripartite DNA-binding sites. J. Biol. Chem. 274, 20709–20716 CrossRef Medline
19. Nishihara, A., Hanai, J., Immamura, T., Miyazono, K., and Kawabata, M. (1999) E1A inhibits transforming growth factor-β signaling through binding to Smad proteins. J. Biol. Chem. 274, 28716–28723 CrossRef Medline
20. Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat. Biotechnol. 30, 521–530 CrossRef Medline
21. Hill, C. S. (2006) Transcriptional Control by the Smads. Cold Spring Harb. Perspect. Biol. 8, a022079 CrossRef Medline
22. Nakano, N., Itoh, S., Watanabe, Y., Maeyama, K., Itoh, F., and Kato, M. (2010) Requirement of TCF7L2 for TGF-β-dependent transcriptional activation of the TMEPAI gene. J. Biol. Chem. 285, 38023–38033 CrossRef Medline
23. Motizuki, M., Isegaya, K., Miyake, K., Ikushima, H., Kubota, T., Miyazono, K., Saitoh, M., and Miyazawa, K. (2013) Oligodendrocyte transcription factor 1 (Olig1) is a Smad cofactor involved in cell motility induced by transforming growth factor-β. J. Biol. Chem. 288, 18911–18922 CrossRef Medline
24. Feng, X. H., Zhang, Y., Wu, R. Y., and Derynick, R. (1998) The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are co-activators for smad3 in TGF-β-induced transcriptional activation. Genes Dev. 12, 2153–2163 CrossRef Medline
25. Yingling, J. M., Datto, M. B., Wong, C., Frederick, J. P., Liberati, N. T., and Wang, X. F. (1997) Tumor suppressor Smad4 is a transforming growth factor-β inducible DNA binding protein. Mol. Cell Biol. 17, 7019–7028 CrossRef Medline
26. Labbé, É., Silvestri, C., Hoodless, P. A., Wrana, J. L., and Attisano, L. (1998) Smad2 and Smad3 positively and negatively regulate TGF-β-dependent transcription through the forkhead DNA-binding protein FAST2. Mol. Cell 2, 109–120 CrossRef Medline
27. Ross, S., Cheung, E., Petraits, T. G., Howell, M., Kraus, W. L., and Hill, C. S. (2006) Smads orchestrate specific histone modifications and chromatin remodeling to activate transcription. EMBO J. 25, 4490–4502 CrossRef Medline
28. Reiter, F., Wimmererother, S., and Stark, A. (2017) Combinatorial function of transcription factors and cofactors. Curr. Opin. Genet. Dev. 43, 73–81 CrossRef Medline
29. Ortega, E., Rengachari, S., Ibrahim, Z., Hoghoughi, N., Gaucher, J., Holehouse, A. S., Khochbin, S., and Panne, D. (2018) Transcription factor dimerization activates the p300 acetyltransferase. Nature 562, 538–544 CrossRef Medline
30. Blokzijl, A., ten Dijke, P., and Ibáñez, C. F. (2002) Physical and functional interaction between GATA-3 and Smad3 allows TGF-β regulation of GATA target genes. Curr. Biol. 12, 35–45 CrossRef Medline
31. Furumatsu, K., Tsuda, M., Taniguchi, N., Tajima, Y., and Asahara, H. (2005) Smad3 induces chondrogenesis through the activation of SOX9 via CREB-binding protein/p300 recruitment. J. Biol. Chem. 280, 8343–8350 CrossRef Medline
32. Germain, S., Howell, M., Esslemont, G. M., and Hill, C. S. (2000) Homeodomain and winged-helix transcription factors recruit activated Smads
to distinct promoter elements via a common Smad interaction motif. *Genes Dev.* **14**, 435–451 [CrossRef Medline](https://doi.org/10.1101/gad.6722203)

33. Kang, Y., Chen, C. R., and Massague´, J. (2003) A self-enabling TGFβ response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. *Mol. Cell* **11**, 915–926 [CrossRef Medline](https://doi.org/10.1016/S1097-2765(03)00305-9)

34. Coda, D. M., Gaarenstroom, T., East, P., Patel, H., Miller, D. S., Lobley, A., Matthews, N., Stewart, A., and Hill, C. S. (2017) Distinct modes of SMAD2 chromatin binding and remodeling shape the transcriptional response to NODAL/Activin signaling. *Elife* **6**, e22474 [CrossRef Medline](https://doi.org/10.7554/eLife.22474)

35. Inoue, Y., Itoh, Y., Abe, K., Okamoto, T., Daitoku, H., Fukamizu, A., Onozaki, K., and Hayashi, H. (2007) Smad3 is acetylated by p300/CBP to regulate its transactivation activity. *Oncogene* **26**, 500–508 [CrossRef Medline](https://doi.org/10.1038/sj.onc.1210190)

36. Nagano, Y., Mavrakis, K. J., Lee, K. L., Fujii, T., Koinuma, D., Sase, H., Yuki, K., Isogaya, K., Saitoh, M., Imamura, T., Episkopou, V., Miyazono, K., and Miyazawa, K. (2007) Arkadia induces degradation of SnoN and c-Ski to enhance transforming growth factor-β signaling. *J. Biol. Chem.* **282**, 20492–20501 [CrossRef Medline](https://doi.org/10.1074/jbc.M607387200)

37. Oliphant, A. R., Brandl, C. J., and Struhl, K. (1989) Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mol. Cell Biol.* **9**, 2944–2949 [CrossRef Medline](https://doi.org/10.1128/MCB.9.11.2944-2949.1989)

38. Itoh, Y., Hayashi, H., Miyazawa, K., Kojima, S., Akahoshi, T., and Onozaki, K. (2007) 17β-Estradiol induces IL-1α gene expression in rheumatoid fibroblast-like synovial cells through estrogen receptor α (ERα) and augmentation of transcriptional activity of Sp1 by dissociating histone deacetylase 2 from ERα. *J. Immunol.* **178**, 3059–3066 [CrossRef Medline](https://doi.org/10.4049/jimmunol.178.5.3059)

39. Ji, H., Jiang, H., Ma, W., and Wong, W. H. (2011) Using CisGenome to analyze ChIP-chip and ChIP-seq data. *Curr. Protoc. Bioinformatics Chapter 2*, Unit 2.13 [CrossRef Medline](https://doi.org/10.1002/0471142727.bio21087)

40. McLeay, R. C., and Bailey, T. L. (2010) Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 [CrossRef Medline](https://doi.org/10.1186/1471-2105-11-165)

41. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 [CrossRef Medline](https://doi.org/10.1093/nar/gkp335)