Characterization of a BMP Responsive Element in the Tlx2 promoter

by

Soo-yeon Sammy Kim

A thesis submitted in conformity with the requirements for the degree of
Master of Science
Graduate Department of Molecular and Medical Genetics
University of Toronto

© Copyright by Soo-yeon Sammy Kim (1999).
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni desextraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-46093-2
Characterization of a BMP Responsive Element in the TLx2 promoter

By Soo-yeon Sammy Kim

Master's of Science
Graduate Department of Molecular and Medical Genetics,
University of Toronto
1999

ABSTRACT

Due to the broad range of processes controlled by Bone Morphogenetic Proteins (BMPs), there is an intense interest in elucidating the molecular mechanisms that regulate diversity and specificity of the biological responses induced by these factors. We recently showed that TLx2, a homeobox gene essential for gastrulation in mouse, is a target gene for BMP signalling. In order to determine the mechanisms involved in regulation of TLx2 transcriptional activation I have undertaken analysis of the 5' flanking region of the TLx2 for BMP-dependent regulatory elements. I have identified a single 191 base pair BMP responsive element (BRE) in the upstream region of the TLx2 gene, which confers BMP responsiveness to a heterologous promoter. BMP dependent induction of BRE in the TLx2 promoter is only observed in P19 mouse embryonic carcinoma cell line and in early embryos, indicating that cell type specific nuclear factors may be required for such response. Electrophoretic Mobility Shift Assay (EMSA) and in vitro
footprinting analysis with bacterially expressed Smad proteins showed that the BRE contains two separate binding sites for Smad1, Smad3, and Smad4. These Smad binding sites partially overlap with two regions that are protected by P19 nuclear extracts. Further detailed analysis of the BRE showed that Smad binding sites are required for BMP-dependent induction of the BRE. In addition, induction was dependent on an adjacent AT-rich site that contains a canonical homeodomain binding site. Furthermore, the residues in the Smad binding sites that are essential for direct binding of Smad1 and Smad4 proteins were determined. Taken together, these results suggest a model in which the interaction of Smads with cell type specific nuclear target protein(s) mediates BMP dependent induction of tissue specific genes, such as Tlx2.
ACKNOWLEDGEMENTS

First, I would like to take this opportunity to thank my supervisor, Dr. Jeff Wrana for his continual guidance, encouragement, and having me in his lab. His balanced criticisms and perspectives greatly improved my ability to think critically and make a sound scientific judgement. This work was supported by a scholarship from the Natural Science and Engineering Research Council of Canada. I am also deeply thankful to my supervisory committee, Drs. C.C. Hui, James Ellis, and Tony Pawson, and to Dr. Liliana Attisano for their valuable and helpful comments, advice, and encouragement. I would like to thank all the past and present members of the Wrana and Attisano lab, Dylan Pannell and Ted Chiang for their insightful discussion, advice, and friendship, and also for making the long working hours much enjoyable. The expert guidance of Pamela Hoodless who first brought this project to life, is greatly appreciated.

I am continually thankful to all my friends, especially Joshua for continuous prayer, encouragement, and for loving my soul. To my wonderful Mom (Lulu) and Dad (Puma), and my brothers, Won and Simon, and my lovely grandmothers, eternal thanks for your endless love, support, and strength, you are with me always.

Most importantly, to God who made all things possible.
Table of Contents

ABSTRACT ii
ACKNOWLEDGEMENTS iv
TABLE OF CONTENTS iv
LIST OF FIGURES AND TABLES v

1. INTRODUCTION
   1.1 TGFβ receptors and downstream signalling components, Smads 1
   1.2 Activation of Smads 5
   1.3 Nuclear functions of Smad complexes 5
   1.4 Identification of the Tlx2 homeobox gene as a downstream target of BMP signalling 8
   1.5 Rationale and objective for this study 10

2. MATERIALS AND METHODS
   2.1 Cell cultures 11
   2.2 Reporter gene construction 11
   2.3 Transient transfection and reporter assay 13
   2.4 Production and purification of Smad fusion proteins 13
   2.5 Electrophoretic Mobility Shift Assay 14
   2.6 DNaseI footprinting analysis 15

3. RESULTS
   3.1 Identification of a BMP Responsive Element (BRE) in the Tlx2 promoter 17
   3.2 Cell type and ligand specific induction of the BRE by BMP2 19
   3.3 Identification of two protected regions in the BRE 20
   3.4 Induction of BRE requires three regions including homeodomain protein binding site 25
   3.5 Identification of the Smad binding elements required for induction of the BRE 32
   3.6 Detailed functional analysis of the SB#2 site in the BRE 39

4. DISCUSSION
   4.1 Tlx2 is a direct target gene for BMP/Smad1 signalling pathway 51
   4.2 Characterization of BRE in the Tlx2 promoter 52
   4.3 Smad-dependent regulation of Tlx2 promoter 53
   4.4 Functional significance of AT-rich region 55
   4.5 Negative regulation of BMP-mediated Tlx2 induction by Nkx2.5 58
   4.6 Model for BMP/Smad1-dependent regulation of the BRE 60
   4.7 Conclusion 63
   4.8 Future directions 64

5. REFERENCES 68
List of Figures and Tables

Figures

<table>
<thead>
<tr>
<th>Figures</th>
<th>Description</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Different functional domains in Smad proteins</td>
<td>3</td>
</tr>
<tr>
<td>Figure 2</td>
<td>A general model for BMP/Smad signalling</td>
<td>6</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Identification of a BMP Responsive Element in the Tlx2 promoter</td>
<td>18</td>
</tr>
<tr>
<td>Figure 4A</td>
<td>Cell type specific induction of Tlx2 by BMP2</td>
<td>21</td>
</tr>
<tr>
<td>Figure 4B</td>
<td>BRE responds specifically to BMP</td>
<td>21</td>
</tr>
<tr>
<td>Figure 5A</td>
<td>In vitro DNaseI footprint analysis with P19 nuclear extract</td>
<td>24</td>
</tr>
<tr>
<td>Figure 5B</td>
<td>Comparison between in vitro DNaseI footprint analysis of P19 &amp; HepG2</td>
<td>24</td>
</tr>
<tr>
<td>Figure 6</td>
<td>BMP dependent induction of BRE requires three regions, Region I, II, and AT-rich site</td>
<td>27</td>
</tr>
<tr>
<td>Figure 7A</td>
<td>Detailed functional analysis of the AT-rich region</td>
<td>31</td>
</tr>
<tr>
<td>Figure 7B</td>
<td>Sequence comparison among Tlx2 and consensus binding sites for S8 and Nkx2.5</td>
<td>31</td>
</tr>
<tr>
<td>Figure 8A</td>
<td>Nkx2.5 inhibits BMP-mediated induction of BRE in a dose dependent manner</td>
<td>33</td>
</tr>
<tr>
<td>Figure 8B</td>
<td>Mutation in Nkx2.5 binding site abolish BMP-dependent Tlx2 transcriptional activity</td>
<td>33</td>
</tr>
<tr>
<td>Figure 9</td>
<td>In vitro DNaseI footprint analysis with GST-Smads on Tlx2 promoter</td>
<td>35</td>
</tr>
<tr>
<td>Figure 10A</td>
<td>EMSA analysis with GST-Smads using FPA probe</td>
<td>38</td>
</tr>
<tr>
<td>Figure 10B</td>
<td>EMSA analysis with GST-Smads using FPB probe</td>
<td>38</td>
</tr>
<tr>
<td>Figure 11</td>
<td>10-bp sequence similarity between SB#1 and SB#2 sites</td>
<td>40</td>
</tr>
<tr>
<td>Figure 12A</td>
<td>Identification of the SB#2b as an essential sequence for BMP dependent signalling</td>
<td>43</td>
</tr>
<tr>
<td>Figure 12B</td>
<td>Detailed linker-scanning mutation analysis of SB#2b</td>
<td>43</td>
</tr>
<tr>
<td>Figure 13A</td>
<td>6-bp (GGCGGC) in SB#2b is the target sequence for Smad4 binding</td>
<td>46</td>
</tr>
<tr>
<td>Figure 13B</td>
<td>Two AGGG sequences in SB#2 are essential for Smad1 binding, but not for Smad4 binding</td>
<td>46</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Nucleotide sequence of the 191-bp BRE</td>
<td>49</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Possible mechanisms for Smad/transcription factor coregulation of the Tlx2 promoter</td>
<td>62</td>
</tr>
</tbody>
</table>

Tables

<table>
<thead>
<tr>
<th>Tables</th>
<th>Description</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Subgroups of Smads and their conservation in vertebrates, Drosophila, and C.elegans</td>
<td>3</td>
</tr>
<tr>
<td>Table 2</td>
<td>Molecular components of TGFβ superfamily signal transduction</td>
<td>4</td>
</tr>
<tr>
<td>Table 3</td>
<td>Summary of signalling and binding activities and profiles on competition for all the mutant and wild type constructs</td>
<td>48</td>
</tr>
</tbody>
</table>
INTRODUCTION

The Transforming Growth Factor β (TGFβ) superfamily is comprised of a large group of structurally related polypeptide growth factors, each capable of regulating a wide range of biological processes including cell proliferation, differentiation, cell fate determination, morphogenesis, and apoptosis (Wrana, 1998). Work over the past few years has led to the elucidation of a unique signal transduction network for these factors (reviewed by Attisano and Wrana, 1998; Heldin et al., 1997; Kretzschmar and Massagué, 1998; Massagué, 1998). This network involves a family of transmembrane receptor serine/threonine kinases, and their cytoplasmic substrates, the Smad proteins which, upon activation of the pathway, move into the nucleus where they activate target gene expression in association with DNA binding partners. Distinct repertoires of receptors, Smad proteins, and DNA binding partners seemingly underlie the multifunctional nature of TGFβ biology. BMPs, the largest and most heterogeneous group within the TGFβ superfamily, are critical regulators of a wide range of developmental processes; BMPs control the development and maintenance of a multitude of diverse cell types (Hogan, 1996). The intracellular mechanisms by which BMP signalling is transduced into multiple developmental fate is therefore a current focus of interest for understanding how BMPs function during early embryogenesis.

1.1 TGFβ receptors and downstream signalling components, Smads
Members of the TGFβ superfamily signal through a family of transmembrane ser/thr kinases, the type I and type II receptors, which form heteromeric complexes upon ligand binding (Wrana et al., 1994). Within this complex, the type II transphosphorylates and activates the type I receptor kinase, thereby allowing propagation of the signal to a downstream family of signal transduction molecules called Smads. Smads were initially identified in Drosophila melanogaster as the product of the gene, Mothers against dpp (Mad) (Raftery et al., 1995; Sekelsky et al., 1995). At least ten distinct Smads (Smad1 through Smad10) have been identified in vertebrates. Sequence comparisons within members of this family indicate the presence of three domains; a conserved amino-terminal MH1 (Mad Homology 1) domain, a highly conserved carboxy-terminal MH2 domain, and a central, non-conserved, linker region (Figure 1) (reviewed by Attisano and Wrana, 1998). On the basis of structural and functional criteria, the Smad family can be divided into three subgroups (Table 1). The first group, receptor-regulated Smads (R-Smads), is comprised of those Smads that are direct substrates of activated type I receptors. The second group, the common Smads, includes Smads that are not direct substrates of the receptors but participate in signalling by associating with the R-Smads. The third group, inhibitory (or anti-) Smads, consists of proteins that block signalling by acting as intracellular antagonists of the TGFβ family-induced Smad signalling pathways (reviewed by Attisano and Wrana, 1998). Although the three categories mentioned above provide a general framework for how Smads work, the definition and number of the categories may change as our understanding of Smad progresses. Table 2 provides a summary of relatively well characterized sets of components of TGFβ superfamily signalling pathways.
Figure 1

Different functional domains in Smad proteins.
The highly conserved MH1 and MH2 domains are separated by a non-conserved linker region. The model depicts the known functions of the domains, with the functions of the basal state at top and the functions of the activated state at bottom. The site of regulatory phosphorylation events, the C-terminal SSXS motif, present in receptor-regulated Smads is indicated. (Model adapted from Heldin, et al. 1997)

<table>
<thead>
<tr>
<th>Receptor-regulated</th>
<th>Common</th>
<th>Inhibitory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smad1</td>
<td>Smad4</td>
<td>Smad6</td>
</tr>
<tr>
<td>Smad5</td>
<td>Smad4</td>
<td>Smad7</td>
</tr>
<tr>
<td>Vertebrates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smad8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGFβ/Activin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smad2</td>
<td>Medea</td>
<td>DAD</td>
</tr>
<tr>
<td>Smad3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drosophila</td>
<td>Mad</td>
<td>Medea</td>
</tr>
<tr>
<td>C.elegans</td>
<td>Sma2</td>
<td>Sma4</td>
</tr>
<tr>
<td></td>
<td>Sma3</td>
<td>def3</td>
</tr>
</tbody>
</table>

Table 1

Subgroups of Smads and their conservation in Vertebrates, Drosophila and C.elegans. Members of the Smad family can be subdivided into three groups, receptor-regulated, common and inhibitory (or anti-) Smads, with distinct functional roles in TGFβ family signalling. (Table adapted from Attisano and Wrana, 1998)
<table>
<thead>
<tr>
<th>Ligands</th>
<th>Receptors</th>
<th>Receptor regulated Smads</th>
<th>Common Smads</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>typII</td>
<td>typI</td>
<td></td>
</tr>
<tr>
<td>TGFβ</td>
<td>TβRII</td>
<td>TβRI</td>
<td>Smad2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Smad3</td>
</tr>
<tr>
<td>Activin</td>
<td>ActRII</td>
<td>ActRIIB</td>
<td>Smad2</td>
</tr>
<tr>
<td></td>
<td>ActRIIB</td>
<td></td>
<td>Smad3</td>
</tr>
<tr>
<td>BMP2/4</td>
<td>BMPRII</td>
<td>ALK3</td>
<td>Smad1</td>
</tr>
<tr>
<td></td>
<td>ActRII</td>
<td>ALK6</td>
<td>Smad5</td>
</tr>
<tr>
<td></td>
<td>ActRIIB</td>
<td></td>
<td>Smad8</td>
</tr>
<tr>
<td>Dpp</td>
<td>punt</td>
<td>Tkv</td>
<td>Mad</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sax</td>
<td>Medea</td>
</tr>
</tbody>
</table>

**Table 2** Molecular components of TGFβ superfamily signal transduction. Listed are examples of components of relatively well-characterized TGFβ superfamily signalling pathways.
1.2 Activation of Smads

In response to ligand-induced assembly of a heteromeric receptor complex, R-Smad proteins interact specifically with type I receptors and are rapidly phosphorylated by the activated kinase on the last two serines of a conserved C-terminal "SSXS" motif (Abdollah et al., 1997; Macías-Silva et al., 1996) (Figure 1). The interaction between R-Smads and the receptor provides specificity in this pathway. Thus, members of the BMP subfamily signal through Smad1, Smad5, and Smad8 (Chen et al., 1997b; Hoodless et al., 1996; Kretzschmar et al., 1997), while TGFβ and activin, another major subgroup of the TGFβ superfamily, signal through Smad2 and Smad3 (Eppert et al., 1996; Liu et al., 1997b; Nakao et al., 1997b; Yingling et al., 1996). Following phosphorylation of these R-Smads (i.e. Smad1, 2, 3, 5, and 8), heteromeric complexes are assembled with the common mediator Smad, Smad4 (Chen et al., 1997b; Lagna et al., 1996; Nakao et al., 1997a; Wu et al., 1997), and these complexes subsequently translocate into the nucleus (Baker and Harland, 1996; Hoodless et al., 1996; Liu et al., 1996; Nakao et al., 1997a; Zhang et al., 1997) where they are predicted to play an important role in regulating target gene expression (Figure 2).

1.3 Nuclear functions of Smad complexes

The role of Smad proteins as downstream effectors of signalling by receptors for the TGFβ superfamily has revealed new insights into how these factors regulate target gene expression. Smads have been shown to act as transcriptional activators; the C-
Figure 2. A general model for BMP/Smad signalling.

In this model, a hypothetical nuclear factor associates with activated Smad complexes and activates a BMP-specific target gene, such as Tlx2. Mutations in the receptors or in Smad1 that block each phosphorylation event are indicated.
terminal MH2 domain of Smad1 and Smad4 have transactivation activity when fused to a heterologous DNA binding domain (Liu et al., 1996). In addition, Smad proteins directly bind DNA in cooperation with other DNA binding proteins, such as FAST1 and FAST2 (Chen et al., 1996; Labbé et al., 1998). Smad2 and Smad4 together with FAST1, a winged helix DNA binding factor, associate to form the activin response factor (ARF) that binds the activin response element (ARE) of the *Xenopus Mix.2* promoter (Chen et al., 1996; Chen et al., 1997a). Evidence suggests that Smad2 interacts directly with FAST1 and recruits Smad4 into the complex where Smad4 functions to activate transcription (Liu et al., 1997a). At the *goosecoid* promoter, Smad2/4 complex cooperate with FAST2, the mammalian FAST1 homologue, to form a higher order DNA binding complex in response to TGFβ or activin signalling (Labbé et al., 1998). In this case it was demonstrated that transcriptional activation and DNA binding required a Smad4 DNA binding site that is present adjacent to the FAST2 binding site. In an analogous way, the Smad3/4 and transcriptional cofactor complex interacts with responsive elements in the PAI-1 (Feng et al., 1998; Hua et al., 1998), and collagenase (Zhang et al., 1998) promoters in response to TGFβ.

In addition to targeting promoters via transcription factors like FAST1 and FAST2, Smads have an intrinsic, site-specific DNA-binding activity. This activity was first demonstrated for the MH1 domain of *Drosophila* Mad, which binds a site in a Dpp-responsive element within the *vestigial* (*vg*) quadrant enhancer *in vitro* (Kim et al., 1997); the isolated MH1 domain binds more potently than the full length protein, suggesting that the Linker/MH2 regions may negatively regulate DNA-binding activity. Similarly,
Medea, the Smad4 homologue in Drosophila, interacts with GC-rich sequences in the tinman enhancer in Drosophila through its MH1 domain (Xu et al., 1998).

Direct DNA binding activities have also been demonstrated for Smad3 and Smad4 with TGFβ-inducible elements in the PAI-1 (Dennler et al., 1998), JunB (Jonk et al., 1998), and collagenase (Yingling et al., 1997) promoters. Using a PCR-based selection procedure, an optimal DNA sequence for Smad3 and Smad4 binding has been identified. This Smad binding element (SBE) is an 8-bp palindromic DNA sequence (GTCTAGAC) with two copies of GTCT (Zawel et al., 1998). Tandem repeats of this sequence confer TGFβ-dependent transcriptional activation, indicating that direct Smad binding to DNA can be sufficient for transcriptional regulation by TGFβ (Dennler et al., 1998; Zawel et al., 1998). The crystal structure of the MH1 domain of Smad3 bound to SBE revealed that a base-specific DNA interaction is provided by a minimal Smad box, GTCT (Shi et al., 1998). These studies suggest that direct DNA binding of Smad3 and Smad4 may play a role in transcriptional activation by TGFβ.

1.4 Identification of the Tlx2 homeobox gene as a downstream target if BMP signalling

Homeobox genes encode an evolutionarily conserved superfamily of transcriptional factors that play vital roles in various aspects of development (Krumlauf, 1994). In Drosophila and Xenopus some homeobox genes that function in early development are targeted by BMP signalling pathways (Huang et al., 1995; Ladher et al., 1996; Mead et al., 1996; Onichtchouk et al., 1996). Tang et al. (1998) recently identified
a mammalian target gene of the BMP signalling pathway called, \textit{Tlx2} (Tang et al., 1998). \textit{Tlx2} is a mouse homologue of the human T cell leukemia homeobox oncogene, Hox11, and is required during early development for extension of the normal primitive streak and for mesoderm formation during gastrulation (Tang et al., 1998). \textit{Tlx2} expression is first detected in the mouse embryo at E7.0 throughout the epiblast (Tang et al., 1998). By E7.5, expression becomes restricted to a region flanking the posterior primitive streak where BMP4 and other TGFβ-related factors are expressed (Winnier et al., 1995). In later embryos, \textit{Tlx2} is expressed specifically in a neuronal subset of neural crest derived-tissues in developing peripheral nervous system (Wen et al., 1994). The expression pattern of \textit{Tlx2} overlaps that of BMP (i.e. posterior primitive streak of E6.5-7.5 mouse embryos) and the phenotype of targeted mutations in BMP type I receptor (ALK3) and BMP4 display some similarities (e.g. no primitive streak, defects in mesoderm induction) to that of targeted mutation in \textit{Tlx2} in the early mouse embryo. These observations led Tang and colleagues to examine the ability of BMPs to induce expression of \textit{Tlx2}. Incubation of mouse embryos at E6.5, which is just prior to the time of \textit{in vivo} initiation of \textit{Tlx2} induction, with exogenous BMP2 resulted in induction of \textit{Tlx2} mRNA within three hours (Tang et al., 1998). Tang and colleagues also demonstrated that the \textit{Tlx2} promoter is responsive to BMP2 signalling in embryonic carcinoma (EC) cells and is induced by constitutively active forms of the BMP type I receptors, ALK3 or ALK6. Furthermore, increasing Smad1 expression potentiated BMP2-inducing activity and this was dependent on phosphorylation of Smad on the C-terminal SSXS motif. Thus, \textit{Tlx2} defines a specific target gene of a BMP/Smad1 signalling pathway.
1.5 Rationale and objectives for this study

BMPs are critical regulators of a wide range of developmental processes. However, a mechanistic understanding of how BMP signalling regulates these processes is poorly defined. Furthermore, little is known of how BMPs regulate cell type-specific target genes. Our identification of a mammalian target gene of BMP signalling, Tlx2, that functions during gastrulation provides an ideal opportunity to investigate BMP-specific signalling during early mammalian development. Thus, it is the overall goal of my study to gain a molecular understanding of how BMPs regulate diverse developmental processes during early mammalian development by defining a region in the Tlx2 promoter that is a target for transcriptional regulation in response to embryonic BMP signals.
MATERIALS AND METHODS

2.1 Cell cultures

P19 cells were maintained in α-minimal essential media (α-MEM, Gibco-BRL) supplemented with 2.5% calf serum and 7.5% fetal calf serum (Hyclone). HepG2 cells and Mv1Lu cells were maintained in minimal essential media (MEM) with 10% fetal calf serum and nonessential amino acids (Gibco-BRL).

2.2 Reporter gene construction

The 1.6-kb genomic fragment containing the Tlx2 cDNA translational start site at its 3' end was subcloned into the BglII and NcoI sites of the luciferase reporter vector, pGL3 basic (Promega), to create 1.6kbTlx2. To create 1.0kbTlx2, the 1.6kbTlx2 construct was digested with Asp718 and StuI, blunt-ended with Klenow enzyme, and religated to remove 600 bases from the 5' end of the Tlx2 promoter. The 600-bp fragment removed was cloned into the Asp718 and SmaI sites of the pGL3 promoter vector to create 0.6kbTlx2. The deletion panel PCR#1 to #2-8 was generated by PCR using the following primers with designed for ligations into Asp718 and BglII sites: for PCR#1, CTA GCA AAA TAG GCT GTC CC (5' primer), and CCG AGA TCT CGC CTC CGA GCC GGT GAG ACT G (3' primer); for PCR#2, CCG GGT ACC GGG GTC TTC TGG CCA GGA CG (5' primer), and CCG AGA TCT GGA GCG GGG AGG AGA AAC TGA CC (3' primer); for PCR#3, CCG GGT ACC GTG GCG GCA GGG GCG TGT CTT CG (5' primer), and CTT TAT GTT TTT GGC GTC TTC C (3' primer); for #2-1, CCG GGT ACC GGG GTC TTC TGG CCA GGA CG (5' primer), and CCG AGA TCT
GCC ATA ATC GGG TTA GCT AGG (3' primer); for #2-2, CCG GGT ACC GAA GGT AAT GTA ATT CCG (5' primer), and CCG AGA TCT GGA GCG GGG AGA AAG TGA CC (3' primer); for #2-3, CCG GGT ACC GAA GGT AAT GTA ATT CCG (5' primer), and CCG AGA TCT GCC ATA ATC GGG TTA GCT AGG (3' primer); for #2-4, CCG GGT ACC TCG GAG GCG CCA GGG CC (5' primer), and CCG AGA TCT CCT GCC CGG TGC AGG CC (3' primer); for #2-5, CCG GGT ACC TCG GAG GCG CCA GGG CC (5' primer), and CCG AGA TCT GCC ATA ATC GGG TTA GCT AGG (3' primer); for #2-6, CCG GGT ACC GAA GGT AAT GTA ATT CCG (5' primer), and CCG AGA TCT CCT GCC CGG TGC AGG CC (3' primer); for #2-7, CCG GGT ACC TCG GAG GCG CCA GGG CC (5' primer), and CCG AGA TCT GGA GCG GGG AGA AAG TGA CC (3' primer); for #2-8, CCG GGT ACC GGG GTC TTC TGG CCA GGA CG (5' primer), and CCG AGA TCT CCT GCC CGG TGC AGG CC (3' primer).

The 5' deletion panel ΔFPA/AT/SB-lux was created by PCR based strategy, using primers with 5' overhang (BglIII site) designed for ligations. Three internal deletion derivatives, ΔAT/FPB (14)-lux, ΔAT-lux, and ΔFPB (14)-lux, were generated by replacing each region to be removed with lacZ DNA sequence of corresponding size. Four scanning mutations spanning 35-bp AT-rich region (AT mut#1, 2, 3, and 4) were incorporated into the wild type promoter construct by PCR using primers carrying 8-bp random mutations. For the series of constructs with random mutations in SB sites, sense and anti-sense primers with desired mutations were used in PCR based method. The sense strands of the oligonucleotides had the following sequences (mutated sequences are underlined): SB mut#1, CCG GGT ACC TCG GAA ATC ATC AAT CCC CCC CCA GGG CC; SB mut#1-1, CCG GGT ACC TCG GAG GCG CCC AAT CCC CCC CCA.
GGG CC; SB mut#2, GAT TAT GGA ATC CTG CAC CGG GCC AAT TAA TCA TCA ATG CGT GTC TTC GG; SB mut#2a, CGA TTA TGG AAT CCT GCA CCG GGC; SB mut#2b, GGG CAG GGT AAT CAT CAA TGC GTG TCT TCG G; SB mut#2b-1, TGC ACC GGG CCA ATT GGC GGC AGG GG; SB mut#2b-2, GGG CAG GGT AAT GGC AGG GGC GTG TCT TCG G; SB mut#2b-3, GGG CAG GGT GGC CAT AGG GGC GTG TCT TCG G; and SB mut#2b-4, TGG CGG CCA ATG CGT GTC TTC GG. All PCR-generated constructs were sequenced to monitor for PCR-associated nucleotide incorporation errors.

2.3 Transient transfection and reporter assay

All the cell lines used in the study were transiently transfected using a calcium phosphate-DNA precipitation method that has been described (Hoodless et al., 1996). Cells were seeded at 20% confluency in 24-well plates and transfected overnight with 0.1μg reporter DNA per well. Sixteen hours after transfection, cells were incubated in the presence or absence of 10nM BMP2, and BMP-induced luciferase activity was assayed after 18-20 hours. Cells were lysed and luciferase gene expression in the lysates was measured by integrating total light emission over 20 seconds using a Berthold luminometer (Lumat LB 9501). All the transfections were normalized for transfection efficiency by co-transfecting 0.02 μg of a cytomegalovirus β-galactosidase expression vector (pCMV-βgal) per well (Sambrook et al., 1989).

2.4 Production and purification of Smad fusion proteins
GST-fusion constructs of Smad1, Smad2, Smad3, Smad4, Smad1MH1 (amino acids 1-147), Smad2 MH1 (amino acids 1-151), Smad3 MH1 (amino acids 1-144), and Smad4 MH1 (amino acids 1-151) were prepared in pGEX-4T1. The purification was carried out using Glutathione Sepharose 4B (Pharmacia Biotech) according to the manufacturer’s protocol. Expression of each fusion protein was monitored by SDS-PAGE and the yield was estimated by the Bio-Rad protein assay using bovine serum albumin as a reference standard.

2.5 Electrophoretic Mobility Shift Assay (EMSA)

To prepare nuclear extracts, P19 cells were plated in 15 cm plates and grown to about 80% confluency. Cells were then incubated in the presence or the absence of 10 nM BMP2 for 2.5 hours. Cells were washed twice in cold phosphate-buffered saline and resuspended in a hypotonic buffer containing 10mM Hapes-KOH (pH7.9), 1.5mM MgCl₂, 10mM KCl, 0.5mM DTT, and 0.2mM PMSF. Swollen cells were then homogenized and nuclei were pelleted and incubated in high salt nuclear extraction buffer containing 25% glycerol, 20mM Hapes-KOH (pH7.9), 1.5mM MgCl₂, 600mM KCl, 0.2mM PMSF, 0.5mM DTT, and 0.2mM EDTA for 30 minutes on ice. Following extraction, the nuclei were removed by centrifugation, the nuclear extract (supernatant) was dialyzed into a moderate salt solution (100mM KCl), and any precipitated protein was removed by centrifugation. Protein concentrations were determined by using the Bio-Rad assay. [$\alpha^{-32}$P] dCTP Klenow-labeled double stranded oligonucleotides had the following sequences (only top strands are shown): FPA (5'-TCG GAG GCG CCA GGG CCC-3'); FPB (5'-CCG GGC AGG GTG GCG GCA GGG GGG TGT CTT CGG CTT
TGG GGC CCG CCC GGC AGG GGT CCC ATC CCG GGG AGG GAC CTG GAG-3'); GS-SB#2a WT (5'-ACC CGA TTA TGG CGG C-3'); GS-SB#2b WT (5'-GCA CCG GGC AGG GTG GCG GCA GGG GCG TGT CT-3'); GS-SB#2b-1 mut (5'-GCA CCG GGC CAA TTG GCG GCA GGG GCG TGT CT-3'); GS-SB#2b-2 mut (5'-GCA CCG GGC AGG GTG GCC ATA GGG GCG TGT CT-3'); GS-SB#2b-3 mut (5'-GCA CCG GGC AGG GTG GCC ATA GGG GCG TGT CT-3'); GS-SB#2b-4 mut (5'-GCA CCG GGC AGG GTG GCC AAT GCG TGT CT-3'); SB#2(40) WT (5'-GGC GGC CTG CAC CGG GCA GGG TGG CGG CAG GGG CGT GTC T-3', mutated nucleotides are underlined). EMSA was performed as described (Buratowski and Chodosh, 1997) with the following modifications. One to three μg of nuclear extracts or 0.4-1.0 μg of bacterial fusion proteins were incubated for 10 min on ice in 20μL reaction volume containing 20 mM Hepes-KOH (pH7.9), 20% glycerol, 2 mM EDTA, 2 mM DTT, 7.25 mM MgCl₂, 100 ng of salmon sperm DNA, and 2 μg of poly dI-dC. Binding was allowed to proceed for 30 min at room temperature in the presence of 0.01-0.02 pmol or 20,000 cpm of [32P]-labeled DNA prior to non-denaturing electrophoresis. DNA-protein complexes were resolved on a 0.5X TBE, 5% polyacrylamide gel (40:1 acrylamide:bis-acrylamide) at 200 volts for 3 hours at 4°C. The gels were subsequently dried, and autoradiography was performed.

2.6 DNaseI footprinting analysis

To prepare end-labeled DNA probe for in vitro DNaseI footprinting, 3 pmol of 1.6kbTlx2 construct was digested with Asp718, resulting in a 3' recessed end. In order to generate end-labeled DNA with high specific radioactivity, Klenow "fill-in" reaction of
the 3'-recessed end using 30 μCi of [α-32P] dCTP was performed. To obtain a probe labeled only at one end and on one strand, the labeled construct was digested with a second enzyme, HindIII, and purified using agarose gel electrophoresis followed by autoradiography, elution, and ethanol precipitation. DNaseI footprinting assays were performed as described (Buratowski and Chodosh, 1997) with the following modifications. Increasing amount of nuclear extract or GST fusion Smads were incubated in binding buffer containing 20 mM Hepes-KOH (pH 7.9), 2 mM DTT, 2 mM MgCl₂, 100 mM KCl, 20% glycerol, 0.5 mM PMSF, and 2 μg poly dI-dC for 3 min at room temperature to allow for non-specific binding. Approximately, 5,000 dpm of asymmetrically [32P]-labeled probe was added to each footprinting reaction and incubated at room temperature for 30 min. The protein-DNA equilibrium mixtures were then subjected to DNaseI-catalyzed hydrolysis for 2 min at room temperature. Each reaction was stopped by adding 25 μL of 2X stop solution (1.2 M NaCl, 0.2% SDS, 20 mM EDTA, 200 μg/mL, and yeast tRNA). The denatured hydrolysis products were separated by electrophoresis on 6.5% denaturing DNA sequencing gel to visualize binding sites. A chemical sequencing reaction (GA ladder) was run in parallel to determine precise residues that mediate interaction of the protein with DNA.
RESULTS

3.1 Identification of a BMP Responsive Element (BRE) in the Tlx2 promoter

In order to understand the mechanisms through which BMP regulates Tlx2 expression during embryogenesis, I determined the regions of the Tlx2 promoter responsible for the transcriptional activation by BMP. To this end, I have undertaken an analysis of the Tlx2 promoter to identify elements that might mediate BMP-dependent induction of the gene. To facilitate this analysis I used a mouse embryonic carcinoma (EC) cell line, P19, which expresses functional receptors for both BMPs and activins (Hoodless and Hemmati-Brivanlou, 1997), and can model numerous inductive events and cell fate changes that occur during early development (Vidricaire et al., 1994). It was previously shown that the Tlx2 promoter was active in P19 cells and responded to BMP treatment (Tang et al., 1998). To functionally define important cis-DNA elements in the Tlx2 promoter, a 1.6 kb flanking region of the gene was initially divided into two fragments (1.0 kb and 0.6kbTlx2), and subcloned into a vector containing a minimal promoter and a luciferase reporter gene. Luciferase expression was assayed following transient transfections of these constructs into P19 cells. As shown in Figure 3, removal of the 0.6 kb (-1.6 to -1.0 kb from the ATG start codon) sequence from the 5' end of the region completely abrogated functional promoter activity of the Tlx2 5' flanking region in response to BMP, indicating the presence of a BMP response element in this region. To examine this 0.6 kb region in detail, three overlapping constructs (PCR #1, #2, and #3) spanning the region were generated by PCR, transfected into P19 cell, and assayed for the
Figure 3. Identification of a BMP Responsive Element in the Tlx2 Promoter.

Schematic diagrams summarizing the 1.6 kb region in Tlx2 promoter and the DNA fragments constructed for analysis in the luciferase assay shown in the right panel. Analysis of luciferase expression driven by various constructs of the Tlx2 promoter. Corresponding constructs diagrammed in the left panel were transfected into P19 cells and transfectants were left untreated (open bars) or were treated with 10 nM BMP2 for 16 hours (closed bars). Luciferase activity was normalized to β-galactosidase activity and is expressed relative to the basal SV40 promotor-luciferase gene. The error bars represent the average deviation (S.D.) between three separate transfections in representative experiments. Where error bars are not evident, S.D. is below the figure resolution. Nucleotide numbering is relative to the first nucleotide of the initiation methionine codon (ATG). The minimal sequence (191 bp) necessary for BMP transcriptional response is shown (BRE).
ability to induce BMP-dependent luciferase expression. The PCR construct #2 that spans a 280-bp (base pair) region in the middle retained BMP responsiveness while the other two constructs (PCR#1 and #3) completely lost the ability to respond to BMP (Figure 3). A further series of deletion constructs were designed to examine this 280-bp region in a greater detail. Each deletion construct was transiently transfected into P19 cells and tested for ligand dependent luciferase activity. One construct containing 191-bp, #2-7, retained the ability to respond to BMP, suggesting the presence of positive BMP regulatory element(s) in this region (Figure 3). These studies revealed that in P19 cells, the BMP2-mediated induction of the Tlx2 promoter requires the presence of a 191-bp element located 1.073 kb upstream of the transcription initiation site (Iitsuka et al., 1999). This sequence, termed BRE (for BMP Responsive Element), confers BMP responsiveness to a heterologous promoter. Moreover, short deletions from either end of this fragment abolished BMP-dependent induction (data not shown), thus defining this region as a minimal sequence required for BMP dependent induction in embryonic cells.

3.2 Cell type and ligand specific induction of the BRE by BMP2

To investigate the cell type specificity of Tlx2 regulation, I tested the BMP responsiveness of the BRE in several different cell lines, including HepG2, a human hepatoma cell line, Mv1Lu, a mink lung epithelial cell line, and mouse embryonic fibroblasts (MEFs). BRE(191)-lux (construct #2-7 in Figure 3) was transiently transfected into each cell line with or without various components of the BMP signalling pathway, BMP type I (Alk6), type II (ActRIIB) receptors, and Smad1. Following
transfection, cells were incubated in the presence or absence of BMP2 for 16 hours, lysed, and assayed for luciferase expression. In contrast to the strong stimulation of the BRE observed in P19 cells, no induction of the element was detected in either HepG2, Mv1Lu, or MEFs (data not shown) even when all the known components of the BMP signalling pathway were overexpressed in the cells (Figure 4A). These studies indicate that P19 cells express cell type specific factor(s) required for BMP-dependent induction of the BRE. This further suggests that the Tlx2 promoter is not likely to be regulated solely by Smad binding to DNA as has been suggested for Mad regulation of the $vg$ quadrant enhancer in *Drosophila* (Kim et al., 1997), but requires additional factor(s).

To test the specificity of BRE regulation with respect to other TGFB superfamily members, I transfected P19 cells with the BRE reporter construct and treated the transfected cells with BMP2, activin, or TGFB. Of the three ligands tested, only BMP2 activated the BRE reporter construct (Figure 4B). Consistent with this, a constitutively activated version of BMP type I receptor (Alk6*), but not that of TGFB (TβRI*) or activin type I receptor (ActRIB*), induced BRE-mediated transcription in P19 cells (Figure 4B). Thus, as was previously reported for the entire Tlx2 flanking region (Tang et al., 1998), the BRE responds specifically to BMP, but not to TGFB or activin signalling. Taken together, these results indicate that the Tlx2 promoter only responds to BMPs and its induction occurs in a cell type specific manner.

3.3 Identification of two protected regions in the BRE
Figure 4

A Cell type specific induction of Tlx2 by BMP2
P19 or HepG2 cells were transiently transfected with the BRE(191)-lux and various components of the BMP signalling pathway. Following transfection, cells were incubated for 16 hours with or without 10 nM BMP2. Normalized luciferase activities are shown as the mean ± S.D. of triplicates. In contrast to the strong stimulation of the BRE observed in P19 cells, no induction was detected in HepG2 cells even when all the known components (i.e. receptors and Smad1) of the BMP signalling pathway were overexpressed.

B BRE responds specifically to BMP.
P19 cells were transiently transfected with the BRE(191)-lux with or without various receptor combinations. The cells were incubated for 16 hours in the presence or absence of 10 nM BMP2, 80pM TGFβ, or 2nM activin. Normalized luciferase activities are shown as the mean ± S.D. of triplicates. Adding TGFβ or activin and constitutively activated versions of TGFβ and activin type I receptors (TβRI* and ActRIB*, respectively) failed to induce BRE mediated transcription in P19 cells.
In an attempt to better define how the BMP-dependent induction of BRE is regulated, I characterized BRE binding proteins by *in vitro* DNaseI footprinting analysis using nuclear extracts from P19 cells. For this, the DNA probe containing the BRE was \[^{32}\text{P}]\)-labeled at one end and incubated with nuclear extracts from BMP2 treated or untreated P19 cells. The mixtures were then subjected to a partial digestion with DNaseI. Both BMP2 treated and untreated extracts protected a 46 base pair region (FPB) and a 15 base pair region (FPA) in the BRE (Figure 5A). By comparing footprints from BMP2-treated and untreated cells, I did not detect any patterns that were dependent on BMP stimulation of the cells. However, these footprints coincided with regions that are required for BMP-dependent transcription; signalling constructs missing either of these regions lost ability to signal upon addition of BMP2 (see construct #2-2 (missing FPA) and #2-4 (missing FPB) in Figure 3). Interestingly, the same regions in the BRE were protected when incubated with nuclear extracts from HepG2, the cell line in which the BRE remains non-responsive to BMP (Figure 5B). Together, these data suggest that FPA and FPB represent binding sites for the ubiquitous DNA binding proteins involved in general transcription initiation. However, I predict that they can not be the sole determinant for the cell type specific induction of BRE in P19 cells based on my observation of P19 specific induction of BRE.

Next, I examined whether a BMP-induced DNA-binding complex is formed with either FPA or FPB sequences using EMSA. Nuclear extracts from BMP2-treated or untreated P19 cells were incubated with \[^{32}\text{P}]\)-labeled FPA or FPB. Consistent with the
Figure 5. *In vitro* DNasel footprint analysis with P19 nuclear extract defined two protected regions within the BRE.

(A) A Tlx2 promoter fragment containing the BRE was assymetrically labeled and incubated with increasing amounts of nuclear extracts (wedge) made from either untreated (lanes 1-4) or BMP2-treated (lanes 5-8) P19 cells. The samples were subjected to a partial digestion with DNasel followed by denaturing electrophoresis and autoradiography. Two protected regions (FPA and FPB), determined from a Maxam-Gilbert G sequencing reaction carried out in parallel (lane 9), are indicated.

(B) Nuclear extracts from P19 and HepG2 cells protected the same regions in the BRE. The probe used in Fig.2A was incubated with increasing amounts of nuclear extracts (wedge) made from either untreated or BMP2-treated P19 (lanes 2-9) or HepG2 (lanes 10-17). The same regions (FPA and FPB) were protected by both nuclear extracts.
Figure 5
footprint analysis, nuclear extracts from both untreated and BMP2 treated P19 cells contain a constitutive binding activity that complexes with the FPA or FPB containing probe (data not shown). To assess whether Smads form part of nuclear complexes that bind to FPA and FPB sequences within the BRE, an EMSA supershift assay was performed using antibodies directed against Smad1 and Smad4. Neither Smad1 nor Smad4 antibodies produced supershifted bands with FPA or FPB probe even after overexpressing Smad1 and Smad4 in the P19 cells (data not shown). The inability to supershift the constitutive complex suggests that Smads are not components of this DNA binding activity. Failure to detect BMP-dependent complexes by EMSA analyses may be due to the possibility that the nuclear BMP target protein(s) of interest is present at a low level. The constitutive binding activity may thus mask the presence of a BMP-induced DNA binding complex or a supershifted band that actually exists. Alternatively, it is also possible that a BMP-induced DNA-binding complex may not be stable under the conditions of our EMSA.

3.4 Induction of BRE requires three regions including homeodomain protein binding site

To determine more precisely the DNA sequences mediating BMP-dependent induction of Tlx2 promoter, I carried out further analysis of internal sequences in the BRE required for BMP-dependent induction. For this purpose the BRE was divided into three different regions based on the footprint analysis using P19 nuclear extracts: Region
Figure 6. BMP-dependent induction of BRE requires three regions, Region I, II, and AT-rich site.

Activity profiles of BRE(191)-lux reporter 5' deletion and internal substitution constructs in P19 cells. The constructs were transiently transfected into P19 cells and luciferase activity in the presence (closed bars) or absence (open bars) of BMP2 was measured. Luciferase activity was normalized to β-galactosidase activity and is expressed as the mean ± S.D. of triplicates from a representative experiment. Two regions (FPA and FPB) protected by P19 nuclear extracts are indicated by yellow boxes. The blue boxes represent Smad binding regions as determined by DNaseI footprinting analysis with GST-Smads (see Figure 9). The numbering represents base pairs relative to the transcription initiation site.
Figure 6
I (45-bp) which contains the FPA sequences, **Region II** (111-bp) that contains the FPB region, and an 'AT-rich' region (35-bp) separating region I and region II (Figure 6).

In order to determine whether these three regions are required for BMP-dependent induction of the BRE, I deleted each region and tested each construct for BMP-dependent signalling activity in P19 cells. Deletion in the region I (ΔFPA-lux) resulted in a significant reduction in promoter activity, indicating that this region is required for BRE induction (Figure 6). This result confirms our previous data since deletion in 5'end of the BRE abolished the BMP-dependent response (Figure 3). To see if region II is essential for induction of the BRE, first 40-bp of the region II, including 14-bp of the FPB sequence, was replaced with a neutral DNA sequence of the same size to keep the spacing constant (ΔFPB(14)-lux). Deletion of this sequence produced a complete reduction in activity of the Tlx2 promoter (Figure 6). From this data, I concluded that at least the first 40-bp of the region II are absolutely required for BMP-dependent induction of the BRE.

Homeodomain proteins recognize the consensus DNA sequence ATTA (Desplan et al., 1988; Hoey and Levine, 1988; Mannervik, 1999). Sequence analysis revealed that the AT-rich region of the BRE contains three copies of homeodomain protein binding sequence, ATTA (see Figure 14). Interestingly, in the entire 1.6 kb Tlx2 promoter region, only five ATTA sequences are found of which three are present in this 35-bp 'AT-rich' region in the BRE. In order to determine whether the intervening AT-rich region is essential for BMP-dependent activation of the BRE, it was deleted and its effect on signalling was tested. An internal deletion in the BRE was prepared by PCR, replacing the AT-rich sequence with a neutral DNA sequence of corresponding size in order to
keep the spacing between the region I and II constant (Figure 6). As shown in Figure 6, replacing the AT-rich sequence (ΔAT-lux) alone completely abrogated BMP-dependent signalling. This result indicates that the AT-rich site is critically important but not sufficient for BMP-dependent induction of the BRE (the construct #2-3 in Figure 3 represents the AT-rich region alone).

Having identified that the AT-rich region is essential for BMP-dependent induction of the BRE, I investigated the functional consequences of the mutations in this region. In order to determine their involvement in BMP2-induced transcriptional activation, I systematically mutated the region by performing 8 base pair scanning mutagenesis (Figure 7A). This mutagenesis generated mutations in each ATTA sequence independently (see constructs AT mut#1, 2, and 4 in Figure 7A). As shown in Figure 7A, constructs with mutations in the first two ATTA sequences (AT mut#1, and AT mut#2) retained the ability to be induced by BMP2, albeit at reduced levels. However, AT mut#3 and the construct with mutations in the third ATTA sequence (AT mut#4) had completely lost the ability to be activated by BMP2 (Figure 7A). Thus the third ATTA, a core binding sequence for homeodomain proteins, plays a crucial role in induction of the BRE by BMP2.

In order to search for candidate proteins with potential binding sites in this AT-rich region, I conducted a transcriptional factor database search. I found that the AT-rich region contains consensus binding sequences for both S8 and Nkx2.5 proteins. S8 is a member of a subfamily of paired-related homeobox transcription factors (de Jong et al., 1993). An Eleven-bp sequence containing the third ATTA core resembles the S8 consensus binding site (Figure 7B). Interestingly, a consensus binding site for Nkx2.5
Figure 7. Detailed functional analysis of the AT-rich region.

(A) Eight-bp scanning mutagenesis was performed on the AT-rich region and signalling activities of each mutant. Thirty five-bp element containing three copies of the homeodomain binding core sequences, ATTA, is indicated by solid box. Scanning mutants introduced an 8-bp (7-bp in AT mut#3) substitution indicated in lower case letters. The location of each mutation is indicated by striped boxes. P19 cells were transiently transfected with each mutant construct. Luciferase activity from cells incubated with (closed bars) or without (open bars) BMP2 is plotted as in Figure 3.

(B) Comparison between the sequence in AT-rich region essential for BRE induction by BMP2 and consensus binding sites for the homeodomain proteins, S8 and Nkx2.5. ATTA core sequences in each site are indicated in boldface type. The location of this motif in all sequences is indicated by a large box.
Figure 7
protein was found on the complementary strand of the same region (Figure 7B). Nkx2.5 is a mouse cardiac-specific homeobox gene, and is a vertebrate homologue of Drosophila Nk-2 homeobox gene, tinman (Chen and Schwartz, 1995). To determine if Nkx2.5 protein is involved in BMP-mediated induction of BRE, I overexpressed Nkx2.5 in P19 cells transfected with the reporter plasmid (BRE(191)-lux). As shown in Figure 8A, expression of Nkx2.5 blocked BMP-dependent induction of BRE in a dose-dependent manner. This indicates that Nkx2.5 is a potential negative regulator of BMP-dependent Tlx2 expression. In order to further examine this negative regulation by Nkx2.5 I introduced random point mutations of the Nkx2.5 binding site in the AT-rich region, expecting these mutations to relieve Nkx2.5-mediated repression of Tlx2 transcriptional activity. Surprisingly, however, the mutation totally abolished BMP-dependent induction of the BRE in P19 cells (Figure 8B).

3.5 Identification of the Smad binding elements required for induction of the BRE

Since previous studies have indicated that Smads may function at target promoters by directly contacting DNA (Dennler et al., 1998; Kim et al., 1997; Yingling et al., 1997; Zawel et al., 1998), I considered the possibility that the BRE contains Smad binding sites. To investigate this, I carried out in vitro DNaseI footprint analysis using bacterially expressed Smad1, Smad2, Smad3, and Smad4. Although Smad2 did not protect the Tlx2 promoter fragment (lanes 29-32), Smad1, 3, and 4 protected 13-bp and 17-bp sequences partially overlapping with the FPA and FPB regions, respectively (lanes 10-13, and 25-28) (Figure 9). Several DNaseI-hypersensitive residues were apparent
A Nkx2.5 inhibits BMP-mediated induction of BRE in a dose dependent manner. P19 cells were transiently transfected with the reporter construct, BRE(191)-lux, and an increasing amount of Nkx2.5 constructs (0.005-3.0 μg). Relative luciferase activity from cells incubated with (closed bars) or without (open bars) BMP2 is plotted as in Figure 3.

B Mutations in Nkx2.5 binding site abolish BMP dependent Tlx2 transcriptional activity. Wild type reporter construct, BRE(191)-lux, or mutant reporter with point mutations in Nkx2.5 binding site was transiently transfected into P19 cells. Luciferase activity in the presence (close bars) or absence (open bars) of BMP2 was measured and plotted as in Figure 3.

Figure 8
Figure 9. *In vitro* DNasel footprint analysis with GST-Smads on *Tlx2* promoter.

Increasing concentrations of bacterially expressed MH1 domains of Smad1 (S1) (lanes 6-9), Smad2 (S2) (lanes 3-5), Smad3 (S3) (lanes 14-17), Smad4 (S4) (lanes 21-24), full length S1 (lanes 10-13), full length S2 (lanes 29-32), full length S4 (lanes 25-28), or the GST control (lanes 2, 18-20) were incubated with the labeled *Tlx2* promoter fragment (see Fig. 2A), and subjected to DNasel digestion. The Smad1, Smad3, and Smad4 protected regions are shown. Two regions (FPA & FPB), protected in the presence of P19 nuclear extracts, are indicated. Neither the MH1 domain (lanes 3-5) nor full length version (see lanes 29-32) of Smad2 yielded any detectable protection on this promoter region. Hypersensitive base residues are indicated by arrows.
(arrows in Figure 9). These hypersensitive residues, a "hallmark" of protection, are often found adjacent to regions protected by DNA-bound proteins that "bend" DNA, rendering the flanking residues "hypersensitive" to DNaseI attack. Four hypersensitive residues were apparent with Smad4 protection while only two of the four residues were detectable with protection by Smad1 and Smad3 proteins (Figure 9). Since DNA binding activity of receptor regulated Smads appears to require release from intramolecular auto-inhibitory activity between the MH1 and MH2 domains (Hata et al., 1997; Kim et al., 1997; Zawel et al., 1998), I also examined the ability of MH1 domain of Smads alone to bind the BRE. I observed that the MH1 domains of Smad1, Smad3, and Smad4 protected the same regions that were bound by their full length counterparts (lanes 6-9, 14-17, and 21-24), whereas the MH1 domain of Smad2 (lanes 3-5) did not yield any detectable protection (Figure 9).

EMSAs using bacterially produced Smads and [32P]-labeled probes containing FPA (Figure 10A) or FPB (Figure 10B) revealed that both the MH1 and full length versions of Smad1 and Smad4 and the MH1 domain of Smad3 were able to bind both probes, albeit at differential levels. However, full length Smad3 was not able to bind to either probe (Figure 10). The specificity of the interaction between Smad proteins and FPA or FPB probe in a complex was determined by competition experiments using unlabeled probes. The presence of excess amounts of specific competitors (unlabeled FPA or FPB) reduced the formation of the radiolabeled FPA or FPB-Smad complex, whereas the presence of a non-specific competitor (unlabeled Activin Responsive Element) did not affect the DNA-protein interaction (data not shown). However, more than 100-fold molar excess of unlabeled FPA or FPB competitors were required to reduce
Figure 10. EMSA analysis with GST-Smads using FPA or FPB probe.

(A) Electrophoretic Mobility Shift Assays (EMSAs) using FPA probe and GST-Smads. Bacterially expressed GST control (lane2), full length or MH1 domain versions of S1 (lanes3-7), S2 (lanes 8,9), S3 (lanes 10-13), and S4 (lanes 14-18) were subjected to gel shift assays using [\textsuperscript{32}P] labeled probe (40-bp) containing the FPA region. Protein/DNA binding complexes were resolved on a 5% non-denaturing polyacrylamide gel and visualized by autoradiography.

(B) Electrophoretic Mobility Shift Assays (EMSAs) using FPB probe and GST-Smads. Increasing amounts of bacterially expressed MH1 domains of Smad1 (lanes 3-5), Smad2 (lanes 10,11), Smad3 (lanes 12,13), Smad4 (lanes 16, 17), or full length versions of Smad1 (lanes 6-9), Smad3 (lanes 14, 15), Smad4 (lanes, 18,19), and GST control (lane 2) were subjected to gel shift assays as above, using [\textsuperscript{32}P] labeled probe (81-bp) containing the FPB region. The complexes were resolved on a 5% non-denaturing polyacrylamide gel and visualized by autoradiography.
the formation of Smad-DNA complex by 50%, indicating a low affinity interaction of Smads with Smad binding sites flanking FPA or FPB. This is consistent with previous observations that interaction of Smads with DNA occurs with relatively low affinity (Labbé et al., 1998; Shi et al., 1998; Zawel et al., 1998). Consistent with the footprint analysis, neither the MH1 domain nor full length version of Smad2 binds the DNA.

Taken together, DNaseI footprint analysis and EMSAs with bacterially expressed Smads defined two Smad binding regions in the BRE; a sequence partially overlapping FPA (designated as SB#1 for Smad Binding site#1), and a region partially overlapping FPB (SB#2) (see Figure 6). Sequence comparison between two SB sites revealed 10-bp sequence similarity (only one base pair difference). To determine the significance of such a repeat in SB sites, the GGCG\text{cAGGG} sequence in each site was mutated independently and its effect on BMP-dependent signalling was tested. As shown in Figure 11, mutation in all ten residues resulted in a marked reduction in activation of BRE in response to BMP in P19 cell (SBmut#1 and SBmut#2 in Figure 11). Moreover, mutation of the last four bases in either SB site was sufficient to abolish BMP-dependent transcriptional stimulation (SB mut #1-1 and #2b-4 in Figure 11). Taken together, these results show that SB sites are indispensable for activation of the BRE, and indicate that direct binding of Smad proteins is absolutely required for activation of Tlx2 promoter in response to BMP.

3.6 Detailed functional analysis of the SB#2 site in the BRE.
Figure 11

Sequence similarity between SB#1 and SB#2 sites.
The GGCG/ CAGGG sequence in each site was independently mutated and its effect on BMP signalling was assessed by luciferase signalling assay.
Since identical mutations in SB#1 and SB#2 site resulted in abrogation of BMP-dependent induction of the BRE (Figure 11), I decided to focus on the SB#2 site for further detailed analysis. SB#2 is composed of a 30-bp region containing two distinct Smad-protected elements, SB#2a and SB#2b. The first element, SB#2a, consists of protected residues, CGG, and is separated by 12-bp from the second Smad-protected element. This 15-bp second element, SB#2b, contains four protected residues (AGGG) and a DNaseI hypersensitive "T" residue just upstream of the 10-bp element common in both SB#1 and SB#2 (See Figure 14 on page 49). In order to determine the requirement of each element in BRE activation, luciferase constructs containing mutations in each individual element (construct SB mut#2a and mut#2b in Figure 12A) were generated. The construct SB mut#2a retained ability to respond to BMP2 at a level similar to that of the wild type promoter construct, BRE(191)-lux, indicating that this Smad binding site is not essential for BMP signalling. In contrast, SB#2b completely lost the ability to be induced by BMP2 (Figure 12A). From this study, I concluded that 15-bp SB#2b is essential for BMP-dependent induction of the BRE. Having identified that SB#2b sequence is essential for BMP-induced transcriptional activation of the BRE, I examined the functional consequences of scanning mutations in the region. As shown in Figure 12B, mutation in the first four base pairs (SB mut#2b-1) reduced ability to be activated by BMP2; it was induced by 2.0-2.5 fold consistently upon BMP2 treatment. However, all the other scanning mutant constructs (SB mut#2b-2, #2b-3, and #2b-4) completely abrogated BMP-induced activation of the BRE. This data suggests that the last ten base residues (GGCGGCAGGG) in SB#2b are necessary for activation of the BRE. The first
Figure 12. SB#2 site is required for efficient induction of the Tlx2 promoter and direct Smad4 binding.

(A) Identification of a Smad-binding sequence as essential for BMP dependent signalling. Solid boxes represent Smad binding sites containing mutated sequences. P19 cells were transiently transfected with the indicated constructs. Relative luciferase activity from cells incubated with (closed bars) or without (open bars) BMP2 is plotted in the right panel, as in Figure 3.

(B) Detailed linker-scanning mutation analysis of Smad binding sequences. All the mutants are identical to the wild type BRE(191)-lux with the exception of the underlined bases in lowercase letters, shown for each mutant construct. Ten-bp sequences present in both SB#1 and SB#2 are highlighted and a hypersensitive residue (T) within SB#2b is indicated by asterisk (*). The constructs were transiently transfected into P19 cells and BMP-dependent luciferase activity was measured and normalized to β-galactosidase activity.
Figure 12
four base pairs in SB#2b may also play an important role in BRE induction since mutation in this region decreased the ability of the BRE to respond BMP2.

Having identified the SB sequence essential for BMP-dependent induction of the BRE, I examined which residues in SB#2 are directly involved in Smad binding. For this purpose, I carried out EMSA using bacterially expressed Smads and [32P]-labeled DNA probes encompassing individual groups in SB#2 (SB#2a and #2b). Interestingly, the probe containing SB#2a alone (GS-SB#2a WT) was incapable of binding any of the GST-Smad fusion proteins suggesting that it is a weak Smad binding site (Figure 13A, lanes 1,2). This is consistent with signalling data indicating that the site is not required for a functional BRE. In contrast, the SB#2b (GS-SB#2b WT) had the ability to bind both MH1 and full length versions of GST-Smad4 at a level similar to that of the wild type SB#2 probe, GS-SB#2(40) WT (compare lanes 4,5 to lanes 19,20 in Figure 13A). I next designed four separate scanning mutants to encompass the entire SB#2b probe in search of the specific sequences that are essential for Smad4 binding. As shown in Figure 13A, mutant #2b-2 and #2b-3 (GS-SB#2b-2 and GS-SB#2b-3, respectively) eliminated DNA binding by Smad4 (lanes 10,11 and 13,14), while mutations in SB#2b-1 or in SB #2b-4 alone had no effect on the binding of either MH1 or full length Smad4 (lanes 7,8 and 16,17). Thus the region necessary for Smad4 binding lies within the bases (GGCGGC) mutated in scanning mutants, SB#2b-2 and SB#2b-3. Notably, the mutation that completely abolished BRE activation by BMP2 (SB#2b-4 in Figure 12B) had no effect on binding of Smad4 in EMSA (lanes 16 and 17 in Figure 13A). Taken together with the signalling data, this result suggests that while all 10 base pairs (GGCGGCAGGG) in SB#2b are critically important for BMP dependent signalling, only
Figure 13. Determination of SB#2 residues involved in Smad binding.

(A) A 6-bp (GGCGGC) binding site is the target sequence for Smad4 binding. Top panel; wild-type and mutant probes used in EMSA analysis. Lowercase letters indicate mutations introduced into the wild-type nucleotide sequence. The nucleotide residues protected by purified GST-Smads are indicated by solid circles (*), and hypersensitive residues are indicated by asterisks (*). Bottom panel; EMSA was performed with the $[^{32}P]$ labeled probes in top panel and bacterially expressed GST control (lanes 1, 3, 6, 9, 12, 15, and 18), MH1 domains (lanes 2, 4, 7, 10, 13, 16, and 19) and full-length versions (lanes 5, 8, 11, 14, 17, and 20) of Smad4 protein.

(B) Two AGGG sequences are essential for Smad1 binding, but not for Smad4 binding. Increasing amount of GST-Smad1 MH1 domain (lanes 1-4, 6-9, and 11-14) or GST-Smad2 MH1 (lanes 5, 10, and 15) was incubated with $[^{32}P]$ labeled wild-type (GS-SB#2b WT) or mutant (GS-SB#2b-1 mut, GS-SB#2b-4 mut) probe. Protein/DNA binding complexes were visualized by autoradiography.
Figure 13
six residues (GGCGGC) are required for direct Smad4 binding.

By DNaseI footprinting, I observed that the Smad1 MH1 Domain bound to a region in the BRE that coincided with that of Smad4 (Figure 9). Thus, I considered the possibility that Smad1 bound to the same sequences protected by Smad4. Interestingly, I did not observe any Smad1-DNA complex; neither MH1 nor full length Smad1 was able to bind any of the SB#2 probes tested (data not shown). One possibility is that Smad1 binding is not stable under this particular EMSA condition. Therefore I examined the possibility of Smad1 MH1 binding to SB#2b site by assessing whether increasing amounts of Smad1 MH1 can compete with Smad4 for binding the same region. Surprisingly, Smad1 MH1 could compete Smad4 MH1 for binding only to the wild type probe (lanes 1-4 in Figure 13B). Addition of increasing amount of Smad1 MH1 did not affect Smad4 binding to the mutant probes carrying point mutations in the first or last four base pairs (both AGGG) in SB#2b (lanes 6-9 and 11-14 in Figure 13B). Adding equivalent amount of Smad2 MH1 did not alter the binding activity of Smad4 protein, indicating that the binding competition is specific for Smad1 protein (lane 5 in Figure 13B). This result suggests that the first and last four base pairs (AGGG) of the SB#2b are essential for Smad1 binding, but not for Smad4 binding. Signalling and binding activities and profiles on competition for all the mutant and wild type SB#2 constructs are summarized in Table 3.
<table>
<thead>
<tr>
<th>Constructs</th>
<th>Smad4 MH1 binding</th>
<th>Signalling activity</th>
<th>Competition by Smad1MH1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB#2 (40) WT</td>
<td>++++</td>
<td>++++</td>
<td>N/A</td>
</tr>
<tr>
<td>SB#2a WT</td>
<td>-</td>
<td>++++</td>
<td>N/A</td>
</tr>
<tr>
<td>SB#2a mut</td>
<td>N/A</td>
<td>++++</td>
<td>N/A</td>
</tr>
<tr>
<td>SB#2b Wt</td>
<td>++++</td>
<td>++++</td>
<td>+</td>
</tr>
<tr>
<td>SB#2b-1 mut</td>
<td>++++</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>SB#2b-2 mut</td>
<td>-</td>
<td>-</td>
<td>N/A</td>
</tr>
<tr>
<td>SB#2b-3 mut</td>
<td>-</td>
<td>-</td>
<td>N/A</td>
</tr>
<tr>
<td>SB#2b-4 mut</td>
<td>++++</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 3**

Summary of signalling and binding activities and profiles on competition for all the mutant and wild type SB#2 constructs studied. The activity levels of mutants are relative to that of the wild type (set as ++++).
**Figure 14**

**Nucleotide sequence of the 191-bp BRE.** Region I, II and AT-rich regions are indicated. Two Smad binding sites partially overlapping FPA and FPB are boxed, and 10-bp sequence present in both SB sites are highlighted. Three copies of the homeodomain binding core sequence, ATTA, are indicated in boldface type. Consensus binding sites for S8 and Nkx2.5 family members are shown. Hypersensitive nucleotide residues determined from DNAseI footprinting analysis are indicated by vertical arrows.
DISCUSSION

Until recently, very little was known about how TGFβ superfamily signals are transduced intracellularly. This situation has changed rapidly over the past few years, however, with the discovery and characterization of the Smad family of gene products as transducers of TGFβ superfamily signals. Although our understanding of how Smads mediate cell type-specific transcriptional regulation is still incomplete, characterization of these signal transducers has begun to provide clues as to how the TGFβ superfamily of factors regulates complex patterns of cell specification. In mammals, BMPs play prominent roles in the development of nearly all organs and tissues and are critical for establishing the embryonic body plan. Despite the multitude of processes controlled by BMPs, the mechanistic understanding of how BMP signalling regulates diverse developmental processes is poorly defined. Furthermore, little is known of how BMPs regulate cell type specific target genes. Tlx2, a homeobox gene essential for gastrulation in mouse, is a target for BMP signalling. In order to understand the mechanisms involved in regulation of BMP-induced Tlx2 transcriptional activation, I have undertaken an analysis of the Tlx2 promoter. I have identified a 191-bp BRE from the promoter of the Tlx2 gene, which confers cell type-specific BMP responsiveness to a heterologous promoter. I have demonstrated that the presence of two separate 'GC-rich' Smad binding sites and an intervening 'AT-rich' region with multiple copies of homeodomain binding core sequence, ATTA, in the BRE is required for BMP-dependent induction of the Tlx2. I have also determined the residues in the Smad binding site that are essential for direct binding of Smad1 and Smad4 proteins. Taken together, these results suggest a model in
which the interaction of Smads with cell type-specific nuclear target protein(s) mediates BMP-dependent induction of tissue specific genes, such as Tlx2.

4.1 Tlx2 is a direct target gene for BMP/Smad1 signalling pathway

BMPs play a central role in vertebrate gastrulation (Hogan, 1996; Hoodless and Wrana, 1997), but how these signals regulate this developmental process is not yet clear. Previously, Tang and colleagues identified Tlx2 as a critical target gene for BMP signalling during mammalian gastrulation. Tlx2 is expressed in the posterior primitive streak of E7.0-7.5 embryos in a pattern coincident with that of BMP4 (Tang et al., 1998). Furthermore, activation of the BMP signalling pathway rapidly induces Tlx2 expression in mouse embryos and regulates induction of the Tlx2 promoter in P19 EC cells (Tang et al., 1998). In this study, I identify Smad binding sites within the BMP responsive region in the Tlx2 promoter and demonstrate that these Smad binding sites are essential for BMP-dependent induction of the Tlx2 promoter, and for direct Smad1 and Smad4 binding. The presence of Smad binding sites in the BMP responsive region of the Tlx2 promoter, along with its rapid induction (within three hours) in whole embryos in response to exogenous BMP2 (Tang et al., 1998) strongly indicates that Tlx2 represents a direct target of a BMP/Smad1 signalling pathway. Interestingly, Ladher and colleagues identified a Xenopus homeobox gene, Xom, as an immediate early gene responsive to BMP4 signalling (Ladher et al., 1996). Like Tlx2, Xom belongs to the Hox11 subfamily, suggesting the existence of an evolutionarily conserved BMP/Tlx2 (Xom) signalling pathway essential for vertebrate development.
4.2 Characterization of BRE in the Tlx2 promoter

Our identification of a mammalian target gene of BMP signalling that functions during gastrulation provides an ideal opportunity to investigate BMP-specific signalling during early mammalian development. Through detailed functional analysis of the 5′ flanking region of the Tlx2 promoter, I have identified a 191-bp element (BRE) required for BMP-activated transcription and have demonstrated that this sequence can confer BMP inducibility to a minimal basal promoter. BMP-dependent induction of BRE in the Tlx2 promoter is only observed in P19 cells and in early embryos, indicating that cell type specific nuclear factors may be required for such responses. However, it should be noted that the possibility of transcriptional repressor molecules being expressed in non-responsive cell lines, such as HepG2 and Mv1Lu, can not be ruled out. BMP signalling is initiated when the ligand induces heteromeric complex formation between its type II receptors and the type I receptors, ALK3 or ALK6. This results in phosphorylation of Smad1 and the highly related Smad5 by type I receptors followed by association with Smad4 and nuclear translocation (Hoodless et al., 1996; Kretzschmar et al., 1997; Lagna et al., 1996; Liu et al., 1996). An analogous pathway has been defined for Smad2 and Smad3 in response to TGFβ/activin signalling (Lagna et al., 1996; Macías-Silva et al., 1996; Zhang et al., 1997). Importantly, these R-Smads appear to function specifically in either in BMP or TGF/activin pathways (Wrana and Attisano, 1996). For instance, in Xenopus, overexpression of Smad1 or Smad2 induces ventral or dorsal mesoderm, respectively (Baker and Harland, 1996; Graff et al., 1996; Thomsen, 1996). My observation on BMP-dependent regulation of the BRE is consistent with these models;
the BRE is induced by activated ALK3 or ALK6 and is unresponsive to TGFβ/activin signalling.

Analysis of DNaseI footprints obtained in the presence of P19 nuclear extract revealed two protected regions, FPA and FBP, in the BRE. FPA is only weakly protected by P19 nuclear extracts when compared to FPB, possibly due to lower inherent affinities of the binding protein for the site. By comparing footprints from BMP2-treated and untreated cells, no patterns that were dependent on BMP stimulation of the cells were detected. One plausible explanation is that these protected regions represent binding sites for basal transcriptional machinery with high DNA binding affinity and protection by P19 specific nuclear BMP responsive factor(s) is not detectable due to its low inherent binding affinity. Alternatively, it is possible that the nuclear target(s) may be pre-bound to the BRE and that activated Smads which have translocated to the nucleus may associate to form a transactivating complex. This type of constitutive binding activity has been described for FAST1 binding to the ARE on the Mix.2 promoter in Xenopus (Chen et al., 1996; Chen et al., 1997a). The latter case is unlikely to occur since the same regions (FPA and FPB) were found protected by nuclear extracts from HepG2 cells, a BMP non-responsive cell line that is not expected to have a BMP nuclear target protein(s).

4.3 Smad-dependent regulation of Tlx2 promoter

Previous studies have demonstrated the ability of Smads to directly target promoter/enhancer regulatory sites (Chen et al., 1996; Dennler et al., 1998; Kim et al.,
have identified two separate 'GC-rich' sites (SB#1 and SB#2) that are targets for Smad1 and Smad4 proteins. Sequence analysis revealed that these sites are highly similar (SB#1: GGC GCC AGG G, SB#2: AGG GTG GCG GCA GGG). Point mutations in either site lead to abrogation of both Smad4 binding and transcriptional activation in response to BMP, demonstrating functional significance of Smad binding sites in Tlx2 regulation. This finding directly links BMP-mediated DNA binding activity of Smad4 with its ability to transactivate the promoter. However, mutation in the first and last four base pairs (AGGG) in SB#2 did not alter Smad4 binding activity, while abolishing BMP responsiveness of the promoter. This result suggests that Smad4 binding alone is not the only requirement for transcriptional activation of Tlx2, uncoupling the DNA binding function of Smad4 from its ability to activate the promoter. In fact, these sequences (AGGG) in SB#2 seem to be involved in Smad1 binding which explains why induction of the BRE is abolished when they are mutated. This suggests that both Smad1 and Smad4 binding are essential for efficient induction of the BRE. Comparison of the Smad binding sites (SB#1 and SB#2) with other recently described Smad4 binding sequences (Dennler et al., 1998; Yingling et al., 1997; Zawel et al., 1998) revealed few if any similarities, although they bear some resemblance to a GC-rich Smad4 binding site in the mouse goosecoid promoter (Labbé et al., 1998). Interestingly, however, it shares high sequence similarity with the binding site in the vg promoter for the Drosophila Smad1 homologue, Mad (Kim et al., 1997). Mutating the SB#2 site to consensus Mad binding sequence did not affect either Smad1 or Smad4 binding activity or signalling activity of the BRE even when Mad was overexpressed in P19 cells (data not shown).
DNaseI footprinting and EMSA analysis suggest that Smad3 also binds SB#1 and SB#2. Several studies demonstrated that Smad3 and Smad4 proteins recognize the same DNA sequences in response to TGFβ and activate transcription (Johnson et al., 1999; Labbé et al., 1998; Liberati et al., 1999; Zawel et al., 1998). Thus, it is likely that the Smad3 binding to the Smad4 binding sites in the BRE is functionally insignificant, and merely due to similar preferences for binding site since overexpression of Smad3 has no effect on BMP-mediated induction of the BRE (data not shown). Interestingly, the main difference between Smad2 and Smad3 protein is a unique insert found in the MH1 domain of Smad2. This insert, controlled by alternative splicing, disrupts DNA binding of Smad2 protein (Yagi et al., 1999). Consistent with this finding, neither the MH1 domain nor full length version of Smad2 binds any of the SB probes tested in this study. Direct DNA binding of Smad3 and Smad4 proteins to regulatory sequences in native TGFβ/activin response elements have been demonstrated through a number of studies (reviewed in Derynck et al., 1998). However, direct binding of Smad1 protein to an endogenous BMP response element in a native promoter has not been reported. Therefore, this study represents the first report demonstrating direct binding of Smad1 to BMP response element in an endogenous promoter.

4.4 Functional significance of the AT-rich region

The ubiquitous expression of Smads1, 2 and 4 during early embryogenesis (Graff et al., 1996; Lagna et al., 1996; Waldrip et al., 1998; Yang et al., 1998) and their ability to mediate a broad range of responses during embryogenesis raises the question of how cell
type specific responses to Smads are generated. Observations that Smads have an intrinsic, site-specific DNA-binding activity provide a molecular handle on the problem of how different promoters are regulated by Smads in different cell types. A major part of the solution to this problem may lie in other cell type specific factors binding to additional regulatory elements in the same promoter to generate specific responses. Numerous examples of how this might work are provided by examination of TGFβ/activin/Dpp responsive promoters that are shown to depend on Smad/Mad binding sites and nearby cofactor binding sites (Halder et al., 1998; Hua et al., 1998; Labbé et al., 1998; Szüts et al., 1998; Zhang et al., 1998). For example, activation of the mouse goosecoid promoter by Smad2/4 and the cofactor FAST2 has been shown to be dependent upon a FAST2 binding site and nearby GC-rich Smad4 binding site (Labbé et al., 1998). However, little is known about the downstream DNA binding protein(s) involved in the BMP signalling pathway. Recently it has been shown that Smad1 interaction with Hoxc-8, a homeodomain transcription factor, dislodges Hoxc-8 from its DNA binding element in the osteopontin promoter, resulting in the induction of gene expression (Shi et al., 1999). In this case, however, a Smad1/Smad4 binding site(s) was not identified within the osteopontin promoter.

In the present study, I have demonstrated that an AT-rich sequence intervening two Smad binding sites (SB#1, SB#2) in the BRE is required for BMP-dependent activation of Tlx2. Interestingly, the region contains three copies of the homeodomain binding core sequence, ATTA. Many homeodomain binding sites identified to date are characterized by the presence of an ATTA motif (reviewed in Mannervik, 1999); the ATTA sequence has been shown to be essential for DNA binding of ATTA-preferring
homeodomains, such as *Antennapedia* (*Antp*)-type and * engrailed* (*en*)-type homeodomains (Kissinger et al., 1990; Otting et al., 1990). Through scanning mutagenesis over the entire AT-rich region, I have shown that only one of the three ATTA core sequences (closest to SB#2) is critically important for BMP-dependent transcriptional activation. This suggests that BMP-mediated cell type specific regulation of the *Tlx2* promoter involves cooperation between Smads and a homeodomain containing transcription factor(s) that binds to a nearby regulatory element with ATTA core sequence. For the past few years, many investigators have demonstrated that several homeodomain proteins play a role in downstream events in BMP signalling (Ladher et al., 1996; Mead et al., 1996; Suzuki et al., 1997; Tang et al., 1998). Since the interaction of Smads with DNA occurs with a relatively low affinity (this study; Labbé et al., 1998; Zawel et al., 1998), I predict that BMP target homeodomain protein(s) is required to recruit heteromeric complexes of Smad1 and Smad4 to the BRE. There have been a number of well-documented studies demonstrating cooperative interactions between Smads and other transcription factors, driving ligand-induced transcription from responsive promoters (Chen et al., 1996; Hua et al., 1998; Labbé et al., 1998; Shi et al., 1999; Zhang et al., 1998). Based on these studies and my analysis of the *Tlx2* promoter, I propose that Smad DNA binding functions cooperatively with high affinity DNA-binding partner(s) (e.g. homeodomain proteins) to activate specific target genes, like *Tlx2*. It should be noted, however, that this AT-rich putative homeodomain binding site was not protected by P19 nuclear extract under the conditions used for DNaseI footprinting analysis. Lack of protection may be due to the possibility that the BMP target homeodomain protein of interest is present at a low endogenous level.
4.5 Negative regulation of BMP-mediated Tlx2 induction by Nkx2.5

A transcription factor data base search performed with the AT-rich sequence located a binding site for both Nkx2.5 and S8 to the same site (i.e. on either strand of the same region) within the AT-rich region. S8 is a mouse homeobox gene, expressed during embryogenesis predominantly in mesenchyme. It encodes a homeodomain which resembles that of the Drosophila segmentation gene, paired (prd) and forms a subfamily called paired-related homeobox transcription factors. A recent study identified a DNA element in the regulatory region of the cell adhesion molecule, L1 gene required for establishing the neural pattern of L1 expression (Meech et al., 1999). Interestingly, this element, called HPD, contains binding motifs for both homeodomain and paired domain proteins, and responds to BMP signals during development (Meech et al., 1999). An ATTA sequence within the core of the HPD was required for binding to the homeodomain protein Barx2, while a separate paired domain recognition motif, found in the overlapping region, was necessary for binding to the Pax6 protein (Meech et al., 1999). They demonstrated that the HPD and the proteins to which it binds are key components in the induction of L1 expression by BMPs. Regardless of the precise mechanism for regulation, this type of organization where binding sites for both homeodomain and paired domain proteins coexist in the regulatory region of BMP-responsive promoters like those of Tlx2 and L1, may represent common regulatory targets for BMP-dependent regulation of gene expression.

Nkx2.5 is a mouse cardiac-specific homeobox gene which plays a key role in the heart development (Chen and Schwartz, 1995). This gene is of particular interest since it
is a mouse homologue of *Drosophila* NK-2 class of homeobox gene, *tinman*, which is required for insect heart development. *tinman* is activated in a broad, mesoderm-specific pattern by the transcription factor Twist early in *Drosophila* development (Bodmer et al., 1990). During later mesodermal patterning, *tinman* expression is restricted to the dorsal mesoderm; this later expression is dependent on Dpp signalling (Frasch, 1995). Dpp induces *tinman* in the mesoderm, but not in the ectoderm, in which Dpp is also clearly active in patterning. Mad and Medea bind to the *tinman* promoter *in vitro*, but these binding sites are not sufficient for induction *in vivo*. The basis for the mesoderm specific action of Dpp/Mad/Medea sites on *tinman* regulation appears to be the *tinman* protein itself, which binds at sites adjacent to the Mad/Medea sites and synergizes with Mad/Medea to activate *tinman* transcription (Xu et al., 1998). In order to determine the possibility that Nkx2.5 represents the BMP responsive factors mediating activation of the BRE, I overexpressed Nkx2.5 in P19 cells. However, expression of Nkx2.5 repressed BMP-dependent induction of BRE in a dose-dependent manner. Interestingly, point mutations in the Nkx2.5 binding site totally abolished BMP-dependent induction of the BRE rather than just relieving Nkx2.5 mediated repression of *Tlx2* transcriptional activity. This data suggests that a BMP responsive transcriptional activator protein also binds the Nkx2.5 binding site or a partially overlapping site that was destroyed by the introduced mutation. It is currently unclear what mechanism underlies Nkx2.5-mediated negative regulation of the *Tlx2* promoter. Whether Nkx2.5 homeodomain protein represents an endogenous negative regulator specific to the native *Tlx2* promoter remains to be determined.
4.6 Model for BMP/Smad-dependent regulation of the BRE

Based on previous studies on the mechanism of Smad-mediated transcriptional activation and my analysis of the BRE in the Tlx2 promoter, I propose a model in which a cell type specific DNA binding partner(s) (e.g. homeodomain protein) recruits heteromeric complexes of BMP-specific receptor regulated Smad (e.g. Smad1, 5, or 8) and common Smad (Smad4) to the BRE, presumably through its interaction with BMP-specific Smads. This is likely to promote low affinity interaction between the Smad4 (and possibly Smad1) and Smad binding sites adjacent to the putative homeodomain protein binding site in the AT-rich region of the BRE. This model is consistent with a general model for transcriptional activation by Smads where Smad responsive promoters have a dual DNA sequence requirement (Derynck et al., 1998). One sequence (e.g. AT-rich region) confers the specificity to bind the specific transcription factors that cooperate with the Smad complex. Another adjacent sequence (e.g. SB#2) is required for direct Smad binding and confers Smad selectivity to the first sequence. Thus, only a subset of promoter sequences that bind these cooperating transcription factors are targeted by Smad signalling and this selectivity is provided by flanking or partially overlapping Smad binding sequences.

The cooperativity of Smads with their partner transcription factors is likely a consequence of ligand-induced associations. It is worthwhile to consider a couple of possible modes of such interaction. Firstly, ligand regulated protein-protein interactions may drive association of Smads with other DNA-binding proteins to form a multi-protein complex that coordinately binds multiple target sites within a promoter (Figure 15A).

60
Secondly, Smads and additional transcription factors recognize target sites separately, but protein-protein interactions stabilize binding to the target site, thereby enhancing the specificity or affinity of promoter recognition (Figure 15B). Based on my observations, I predict that the latter case is more probable since bacterially produced Smad proteins bind to the BRE in the absence of the BRF. However, the first case can not be ruled out since it is also plausible that, although Smads bind to the BRE on their own, the presence of BRF(s) may enhance affinity and specificity of binding to activate transcription. In order to test this model, it will be necessary in future experiments to identify the BMP responsive nuclear target protein(s). The details of how Smads function in the case of BMP-dependent induction of the Tlx2 promoter remain to be worked out. Regardless of the precise mechanism for Smad dependent regulation, the intrinsic transcriptional activation function of Smads, their ability to directly target promoter regulatory sites and their ability to interact specifically with other transcription factors, provide a framework for investigating their role in the regulation of gene expression.
Figure 15
Possible mechanisms for Smad/transcription factor coregulation of the BRE. Shown is a schematic of variations on coregulation of transcription at promoter element by Smad and BMP responsive factor, BRF. Double arrows indicate ligand-induced protein-protein interaction between Smad complex and BRF.
4.7 CONCLUSION

BMPs are a large family of secreted polypeptide growth and differentiation factors that are critical regulators of a wide range of developmental processes. Although they are used repeatedly in various contexts throughout development, a mechanistic understanding of how BMP signalling regulates developmental processes is poorly defined. Our identification of a mammalian target gene of BMP signalling, Tlx2, that functions during gastrulation provides an ideal opportunity to investigate BMP-specific signalling during early mammalian development. I have identified a 191-bp BMP responsive element (BRE) in the Tlx2 promoter. BMP dependent induction of BRE in the Tlx2 promoter is only observed in P19 cells and in early embryos, indicating that cell type specific nuclear factors may be required for such response. Detailed analyses of the BRE have revealed that induction requires Smad binding sites and an adjacent element that contains a canonical homeodomain binding site. Our data suggest that cell type specific regulation of the Tlx2 promoter involves cooperation between Smads and homeodomain containing transcription factor(s). Identification of the factor(s) that binds to the homeodomain binding site and cooperates with Smads to mediate BMP-dependent induction of the BRE will invariably lead to a better understanding of how cell type specific responses to BMPs are achieved.
4.8 FUTURE DIRECTIONS

Significant contributions to our understanding of gene regulation have thus far come from the identification of the promoter elements involved in regulation, and the discovery of the transcription apparatus and the factors that regulate it. Future progress concerning BMP-dependent induction of the tissue specific gene, Tlx2, must certainly focus on identifying the cell type-specific factor(s) that mediate BMP-dependent induction of BRE in the Tlx2 promoter.

4.8.1 Identification of the nuclear factor(s) mediating BMP-dependent induction of the BRE

In this study, I demonstrated that a 35-bp AT-rich region, next to a Smad binding site, is required for BMP-dependent activation of Tlx2. This region containing a homeodomain protein binding site represents an ideal sequence for screening the BMP responsive factors (BRFs). One approach to identify the factor(s) that bind to the target AT-rich sequence is the yeast one hybrid screen (Li and Herskowitz, 1993). In this assay, multiple tandem copies of AT-rich region are linked to a minimal promoter designed to drive expression of reporter gene, such as HIS3 and lacZ. These reporter constructs are integrated into the yeast genome to create a reporter strain. To screen a library for gene(s) encoding BRF(s), the reporter strain can be transformed with an activation domain (AD) fusion cDNA library. Importantly, using a non-signalling mutant AT-rich sequence, a mutant type reporter construct, otherwise identical to original wild type reporter should be prepared to eliminate potential false positive clones. The yeast one
hybrid screen allows for obtaining the genes encoding the DNA binding protein of interest immediately after a library screening, however, there are number of potential problems associated with the method. Ability of AD/fusion proteins to interact with the AT-rich sequence will be interfered if the fusion proteins are not stably expressed in yeast cells, or if they can not be localized to the yeast nucleus. Moreover, if the inserted AT-rich sequence interacts with endogenous yeast transcriptional activator, or if it does not require trans-acting factors to activate the HIS3 or lacZ reporter, then very high background growth will result. In such cases, it may be necessary to redesign the target element and construct new reporter strains. Southwestern screening will serve as an alternative approach to a one hybrid screen, however, this method has certain intrinsic limitations as well. Fusion proteins encoded by the library must be capable of folding correctly in bacteria (and in yeast, in the case of one hybrid screen), and of maintaining their structure on a nitrocellulose membrane.

A more serious caveat to both one hybrid and southwestern screening approaches would be if Smad1 and/or Smad4 function to promote binding of the BRF to the BRE. In this case, neither approach would be successful. One strategy to circumvent this problem would be to repeat the screen using yeast or bacteria cells co-expressing Smad1 and Smad4. This may promote binding of the appropriate factor to the BRE and allow activation of the reporter genes. In addition, if high affinity binding requires formation of heteromeric protein complexes, then neither approach may be sensitive enough to allow a single component of such complexes to be identified. Moreover, any post-translational modifications necessary for efficient binding will generally not occur in bacteria. To
address such problems, it will be necessary to purify the BRE-binding complex using an affinity column purification.

4.8.2 Functional characterization of the BRF

Isolation of cDNA a clone(s), by one of three methods discussed, encoding a putative BRF(s) provides means of obtaining large amounts of the BRF for functional studies. This is achieved by subcloning the cDNA downstream of a prokaryotic promoter in a bacterial expression vector. The protein produced in this way should be tested to determine if it has similar activity to the endogenous protein, being capable of binding to AT-rich region in footprinting or EMSA and of activating the BMP dependent induction of the BRE in luciferase signalling assays.

4.8.3 Determining interaction between Smads and the BRF

As mentioned earlier, BMPs act primarily through Smad1, Smad5, and Smad8. In order to examine whether the BRF physically interacts with BMP specific Smads in a BMP2-dependent manner, it is necessary to carry out co-immunoprecipitation using lysates from COS-1 cells, co-transfected with BMP2 type I and type II receptors and various combinations of epitope tagged Smad1 (also Smad5 and Smad8), Smad4, and the BRF that have been either treated with BMP2 or untreated. To address the specificity of the interaction between BRF and TGFβ/activin specific Smads, the same type of experiment should be performed with Smad2 and Smad3.

In addition, it is worthwhile to investigate whether BRF binds directly to the BRE and whether it supports formation of a DNA binding complex with Smad1 and Smad4 in
a BMP2-dependent manner. Nuclear extracts from BMP2-treated or untreated P19 cells, co-transfected with various combinations of tagged Smads and BRF should be tested for DNA binding activity by EMSA. Once a BMP2-dependent DNA binding complex is identified, antibodies against the epitope tags present on Smads and BRF should be used to induce supershifts of the complex.

**4.8.4 Further characterization of BRF**

To identify regions of the BRF necessary for binding to the BRE, for forming BMP2-dependent DNA binding complex, and for association with Smad1 and Smad4, epitope-tagged deletion constructs of BRF should be generated and tested for such activities by *in vitro* DNaseI footprinting, EMSA, EMSA supershift, and co-immunoprecipitation analyses. Furthermore, it will be necessary to test if overexpression of BRF is sufficient to restore BMP2-mediated induction of the BRE in normally non-responsive cell lines, such as HepG2 and Mv1Lu cells. If such is the case, it would suggest that regulated expression of the BRF is critical in controlling specific responses to BMPs. Further characterization of BRF also includes examining its expression pattern in various tissues by Northern blot analysis and whole mount *in situ* hybridization.

In summary, the proteins that mediate cell type specific responses to BMPs are currently unknown and the proposed future work should lead to the identification of such targets. By identifying the nuclear factors (BRFs) that interact with this region of the *Tlx2* promoter, we can then begin to understand how
the specificity and diversity of biological responses is achieved by the TGFβ superfamily of ligands during mammalian development.
REFERENCES


MADR2 maps to 18q21 and encodes a TGFβ MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell*, **86**, 543-552.


