Available online at www.sciencedirect.com

SCIENCE DIRECT

Developmental Biology 281 (2005) 210 – 226

DEVELOPMENTAL BIOLOGY

www.elsevier.com/locate/ydbio

Genomes & Developmental Control

Phylogenetic footprinting and genome scanning identify vertebrate BMP response elements and new target genes

Andreas von Bubnoff[a,1], Daniel A. Peiffer[a,1], Ira L. Blitz[a], Tadayoshi Hayata[a], Souichi Ogata[a], Qiandong Zeng[b,2], Matthew Trunnell[b], Ken W.Y. Cho[a,*]

[a]*Department of Developmental and Cell Biology, Developmental Biology Center, University of California, Irvine, CA 92697, USA*
[b]*Genedata Inc., 601 Gateway Boulevard Suite, South San Francisco, CA 94080, USA*

Received for publication 10 December 2004, revised 8 February 2005, accepted 16 February 2005
Available online 19 March 2005

Abstract

The complex gene regulatory networks governed by growth factor signaling are still poorly understood. In order to accelerate the rate of progress in uncovering these networks, we explored the usefulness of interspecies sequence comparison (phylogenetic footprinting) to identify conserved growth factor response elements. The promoter regions of two direct target genes of Bone Morphogenetic Protein (BMP) signaling in *Xenopus*, *Xvent2* and *XId3*, were compared with the corresponding human and/or mouse counterparts to identify conserved sequences. A comparison between the *Xenopus* and human *Vent2* promoter sequences revealed a highly conserved 21 bp sequence that overlaps the previously reported *Xvent2* BMP response element (BRE). Reporter gene assays using *Xenopus* animal pole ectodermal explants (animal caps) revealed that this conserved 21 bp BRE is both necessary and sufficient for BMP responsiveness. We combine the same phylogenetic footprinting approach with luciferase assays to identify a highly conserved 49 bp BMP responsive region in the *Xenopus Id3* promoter. GFP reporters containing multimers of either the *Xvent2* or *XId3* BREs appear to recapitulate endogenous BMP signaling activity in transgenic *Xenopus* embryos. Comparison of the *Xvent2* and the *XId3* BRE revealed core sequence features that are both necessary and sufficient for BMP responsiveness: a Smad binding element (SBE) and a GC-rich element resembling an OAZ binding site. Based on these findings, we have implemented genome scanning to identify over 100 additional putative target genes containing 2 or more BRE-like sequences which are conserved between human and mouse. RT-PCR and in situ analyses revealed that this in silico approach can effectively be used to identify potential BMP target genes.

Keywords: *VENTX2*; *Id3*; *Xenopus*; TGFβ; BMP; Phylogenetic footprinting; Genome scanning

Introduction

As biology enters the post-genome sequencing era, vast amounts of genome-based information are becoming available, enabling comparisons between whole genomes of different species to identify conserved sequences (phylogenetic footprinting). *Xenopus* is a well-suited system for the use of phylogenetic footprinting for several reasons. First, since its evolutionary distance from mammals such as human and mouse is large (350 million years) (Muller et al., 2002), sequence conservation between *Xenopus* and human or mouse is likely to yield biologically meaningful information. Second, the sequencing of the *Xenopus tropicalis* genome based on 7.5× coverage has been completed (http://genome.jgi-psf.org/Xentr3/Xentr3.home.html). Third, while phylogenetic footprinting is a valuable tool to identify candidate sequences that may represent important *cis*-acting regulatory elements, it is essential to experimentally verify the biological importance of putative regulatory elements identified by this approach, preferably in vivo. In this regard, *Xenopus* embryos are an experimentally amenable system for the study of gene

* Corresponding author. Fax: +1 949 824 9395.
E-mail address: kwcho@uci.edu (K.W.Y. Cho).
[1] Contributed equally to this work.
[2] Present address: The Broad Institute, Massachusetts Institute of Technology, NE125-208, Cambridge, MA 02141-2023, USA.

0012-1606/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2005.02.014

regulation, as promoter activity can be easily examined in transgenic embryos (Gottgens et al., 2000; Kroll and Amaya, 1996; Offield et al., 2000).

In this study, we combine phylogenetic footprinting and reporter gene assays to identify conserved Bone Morphogenetic Protein (BMP) responsive elements located within the promoters of two direct BMP target genes. BMPs are members of the TGFβ super family of secreted polypeptide growth and differentiation factors. During *Xenopus* embryonic development, *BMP2*, *4*, and *7* are expressed dynamically, sometimes in overlapping regions, but at other times are expressed in different regions, suggesting that BMPs have overlapping yet distinct roles in development. For instance, *BMP2*, *4*, and *7* are all expressed in overlapping regions of the embryonic mesoderm and ectoderm to regulate dorsal–ventral patterning of the mesoderm and specification of epidermis (Dale et al., 1992; Fainsod et al., 1994; Jones et al., 1992; Suzuki et al., 1994). On the other hand, *BMP7* is expressed in the organizer, while neither *BMP2* nor *BMP4* are expressed in this region (Hawley et al., 1995).

A number of direct BMP target genes have been identified in recent years by treating cells or embryonic tissues with BMP in the presence of the protein synthesis inhibitor cycloheximide (CHX), among them *Xvent2* and *Xvent2B*, *msx1*, *msx2*, *Id1*, *Id2*, *Id3*, *GATA2*, *Dlx5*, *Tob*, and *FGFR2* (Friedle and Knochel, 2002; Hollnagel et al., 1999; Hussein et al., 2003; Korchynskyi and ten Dijke, 2002; Ladher et al., 1996; Miyama et al., 1999; Peiffer et al., 2005; Rastegar et al., 1999; Yoshida et al., 2000). The promoters of some of these genes, such as *Xvent2*, *Xvent2B*, and *Id1*, were shown to have a Smad binding element (SBE; [C]AGAC or its reverse complement, GTCT[G]) crucial for BMP responsiveness (Hata et al., 2000; Henningfeld et al., 2000). Interestingly, while the SBE is thought to represent a binding site for both R-Smads and the Co-Smad (Smad4) (Shi and Massague, 2003; Shi et al., 1998), this sequence has been shown to preferentially bind Smad4 in the context of the *Xvent2B* and *Id1* promoters (Henningfeld et al., 2000; Korchynskyi and ten Dijke, 2002; Lopez-Rovira et al., 2002).

While reporter genes with multimerized SBEs can respond to BMP signaling, Smad binding to DNA alone is generally believed to be too weak for Smads to function alone as effective and highly specific DNA binding proteins in vivo (Kusanagi et al., 2000; Shi et al., 1998). Thus, additional DNA-binding partners are thought to be required for efficient DNA binding and specific selection of BMP target genes. While numerous such cofactors have been identified for activin/TGFβ responsive genes (among others FAST1, mixer/milk, c-jun/c-fos, TFE3, and WBSCR11) (Chen et al., 1996; Germain et al., 2000; Hua et al., 1999; Ring et al., 2002; Zhang et al., 1998), only few candidate factors have been identified for BMP target genes, one of which is the zinc finger protein OAZ (Hata et al., 2000). OAZ is thought to be involved in activating the *Xvent2* gene upon BMP stimulation and has been suggested to bind to the sequence TGGAGC. Both this sequence and the SBE are required for BMP responsiveness of the *Xvent2* gene (Hata et al., 2000). Other studies show that sequences other than SBEs are important for BMP responsiveness, for example GCAT in the *Xvent2B* promoter (Henningfeld et al., 2000), and GC-rich sequences in the *Id1* and *Smad6* promoters (Ishida et al., 2000; Korchynskyi and ten Dijke, 2002; Lopez-Rovira et al., 2002). However, both the GCAT and the GC-rich sequences can bind Smad proteins (Henningfeld et al., 2000; Ishida et al., 2000; Korchynskyi and ten Dijke, 2002; Lopez-Rovira et al., 2002). Thus, it is still unclear as to whether factors other than Smads are involved in the activation of these BMP target genes. It should be kept in mind that much of the currently available evidence supporting the role of these sequence motifs in vertebrate BMP signaling derive from overexpression studies, while there is little in vivo evidence to suggest that these BMP response elements (BREs) actually respond to endogenous levels of BMP signaling in developing tissues with the exception of a recent study examining the BMP4 synexpression group in *Xenopus* (Karaulanov et al., 2004).

In the current study, we compared the promoter sequences between *Xenopus laevis* and human *Vent2* to identify a 21 bp core region of the *Xvent2* BRE that is both necessary and sufficient for BMP responsiveness. Using a similar approach, a highly conserved BRE in the *Xenopus Id3* promoter was also identified. Further alignments of the *Xvent2* BRE and the *XId3* BRE revealed a conserved core BRE consisting of an SBE and an OAZ-binding-site-like sequence. This core sequence is similar to that recently identified in *Drosophila* (Pyrowolakis et al., 2004). Importantly, this core region is still sufficient for BMP responsiveness in vivo as transgenic embryos harboring BRE-GFP reporters drove GFP expression in areas of endogenous BMP activity in living embryos. Next, implementing an in silico approach, we have used the consensus BRE sequence motif to scan the conserved regions of both the human and mouse genomes to uncover potential BMP responsive genes. Though 1104 genes contain only one copy of the BRE, 138 genes contain 2 or more potential BRE sites including several known BMP target genes. Moreover, we identified 26 orthologues of these genes in *Xenopus* using reciprocal BLAST searches and analyzed the response of these genes toward BMP2 by in situ and RT-PCR analyses. Nearly 50% of these genes are induced directly by BMP2 in dissociated animal cap tissues, suggesting the usefulness of a genome scan approach together with a phylogenetic footprinting method.

Materials and methods

Bioinformatic tools and genomic sequences

Alignments in Figs. 1 and 3 used the Family Relations software (http://family.caltech.edu) (Brown et al., 2002) with a 20 bp window sliding in 1 bp increments. Under these conditions, any similarity above 70% is highly unlikely to be