Front. Agric. China 2008, 2(4): 493–497
DOI 10.1007/s11703-008-0078-3

RESEARCH ARTICLE

Xiaohui ZHANG, Shangzhong XU, Xue GAO, Hongyan REN, Jinbao CHEN

Cloning and bioinformatics analysis of cDNA encoding cattle *Smad*4 gene

© Higher Education Press and Springer-Verlag 2008

Abstract The cDNA of cattle *Smad*4 gene was cloned by RT-PCR, 3′ RACE and 5′ RACE and got a 3503-bp full-long cDNA sequence. The cloned cattle *Smad*4 cDNA sequence had been send to GenBank and got an accession number: DQ494856. Cattle *Smad*4 gene consists of 12 exons and codes 553 amino acids. Cattle *Smad*4 cDNA shares 99%, 96%, 95%, 91% and 91% similarity in nucleic acid sequences, and 99%, 98%, 98%, 99% and 98% similarity in amino acid sequences with sheep, pig, human, rat and mouse, respectively. *Smad*4 cDNA was found in the testes, pancreas, liver, small intestine, ovary, lymph, cardiac muscle, skeleton muscle and thymus gland, which indicated that *Smad*4 was broadly expressed in cattle.

Keywords *Smad*4, RT-PCR, bioinformatics, tissue expression analysis

1 Introduction

TGF (Transforming Growth Factor) plays an important role in controlling the cell's proliferation, differentiation and apoptosis. The signal transmission of TGF-β from cell membrane to nucleus needs all kinds of proteins and the most widely researched proteins were Smads. Nine kinds of Smads have been found in mammals and these Smads could be categorized into three groups according to structure and function. The first group was named R-Smads (Receptor Activated Smads) including Smad1, 2, 3, 5, 8 and 9 which could band with the Ser/Thr domain of TGF/BMP/Action type I receptor (Massague, 1998). The second group was named Co-Smads (Commom Mediator Smads) and the Smad4 was the only Co-Smads in mammals which could band with all the R-Smads and form heterodimers. The Smad4 protein lacked the SSXS domain in C-terminate compared with other Smads. This specific structure induced Smad4 protein can not be phosphorylated and banded with TGF type I receptor. Smad4 had a praline rich region, named Smad4 activation domain (SAD), which was indispensable in Smad4 protein activation. The third group, named I-Smads (Inhibitory Smads), included Smad6 and Smad7 which could inhibit the TGF signal transmission. The Smad6 protein mainly inhibited the BMP signal transmission and the Smad7 could inhibit the BMP/TGF-ß signal transmission (Heldin et al., 1997).

R-Smads could be phosphorylated by activated type I receptor and then banded with Smad4 protein to form a heterodimer. The heterodimer could move into the nucleus and band a specific region of genome DNA to control the expression of specific genes. Some heterodimers, such as Smad3-Smad4 could band DNA directly (Wong et al., 1999) while other heterodimers, such as Smad1-Smad4 and Smad2-Smad4 need specific transcription factors to help them band DNA (Chen et al., 1996; Germain et al., 2000; Hata et al., 2000).

About 50% *Smad*4 gene in pancreatic cancer and colon carcinoma patients had deletion mutations (Hahn et al., 1996; Thiagalingam et al., 1996; Yanagisawa et al., 1999). In recent years, insert/deletion mutations were found in *Smad*4 gene in human seminoma patients (Bouras et al., 2000). So *Smad*4 gene was considered as a tumor-suppressing gene at first. *Smad*4 had an important effect on male reproduction and mainly affected the spermatogenesis and maturation by controlling the interstitial cell (liuhuiwen, 1995). The expression level of *Smad*4 was high in transgenic mice testes tissues and these transgenic mice cannot reproduce for serious defects in testes, but the level of FSH in the blood was not significantly different with the control group, indicating that the *Smad*4 affected the spermatogenesis by paracrine ways. The synthesis of gonadal hormone was also interfered in *Smad*4 transgenic mice, which indicated the level of Smad4 protein had an important effect on TGF-ß signal transmission (Anita et al., 2002).

Translated from *Acta Veterinaria et Zootechnica Sinica*, 2007, 38(4): 321–325 [译自: 畜牧兽医学报]

Xiaohui ZHANG, Shangzhong XU (✉), Xue GAO, Hongyan REN, Jinbao CHEN
Institution of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100094, China
E-mail: Zhangxiaohui78@126.com

The *Smad*4 cDNA of human, rat, mouse, sheep and pig had been cloned but the cattle *Smad*4 cDNA had not been cloned. In this study, the full-long cDNA of cattle *Smad*4 gene was cloned by RT-PCR and RACE-PCR, the cDNA sequence was analyzed by bioinformatics. The tissue expression was analyzed by RT-PCR in multiple tissues.

2 Materials and methods

2.1 Materials

Samples including Luxi cattle testes, pancreas, liver, small intestine, ovary, lymph, cardiac muscle, skeleton muscle and thymus gland tissues were collected and frozen quickly in liquid nitrogen, then stored at −80°C until analysis. PMD-19T plasmid was purchased from Takara Corporation. Bacterium JM110 was stored in our laboratory. TRIzol Reagent was purchased from GIBCOBRL Corporation, RNA PCR kit ver3.0 was purchased from Takara Corporation, SMART RACE was purchased from Clontech Corporation, DNA gel extraction kit was purchased from the Tiangen Biotech (Beijing) Corporation and *Taq* DNA polymerase was purchased from Newprobe Corporation.

2.2 Methods

2.2.1 Total RNA isolation

Total RNA was extracted from different tissues by using TRIzol reagent and dissolved in DEPG treated water. RNA was tested by agarose gel electrophoresis and stored at −70°C until analysis.

2.2.2 Primer design

Thirty-five cattle *Smad*4 gene ESTs sequences were founded by the BLAST program using human *Smad*4 cDNA (NM005359) as core sequence. These ESTs sequences were spliced by SeqMan program, getting a 3208-bp consensus sequence. Three pairs of primers were designed according to this consensus sequence by Primer 5.0 software. The first pair of primers (C1) had a forward: 5′-CGTTAACTGTTGTTTTTCGCTG-3′ and a reverse: 5′-AAGGGTCTACATATCCATCAAC-3′ and the predicted product length was 1162 bp. The second pair of primers (C2) had a forward: 5′-TGGTGTTCCAT TGCTTACTTC-3′ and a reverse: 5′-AACGATACTCC-TTTGTCTGGT-3′ and the predicted product length was 1392 bp. The third pair of primers (C3) had a forward: 5′-ATTGTGCCATGTGGGTGAGTTA-3′ and a reverse: 5′-GCCTGTCATTTAGTAGAAGGTGTC-3′ and the predicted product length was 943 bp. There was at least 100 bp that overlapped between PCR products. The 3′RACE primer was 5′-TACCAGACAA AGGAGTAT-CGTT-3′ and the 5′RACE primer was 5′-GCTCTTTT-GGCGAATGTTTCACTCTC-3′. The last pair of primers was used to study the tissue expression with a forward: 5′-TACCAGACAAAGGA GTATCGTT-3′ and a reverse: 5′-CGGTAAGATATTAGATGGAGTGAC-3′, with a predicted product length of 209 bp.

2.2.3 Reverse transcriptase polymerase chain reaction

The first strand of cDNA was synthetized by RNA PCR Kit ver3.0 according to instruction of the kit (TaKaRa). The PCR reaction mixture contained 1 μL cDNA template, 40 $\mu mol \cdot L^{-1}$ of each dNTP, 0.2 $\mu mol \cdot L^{-1}$ of each primer, 0.6 $mmol \cdot L^{-1}$ of $MgCl_2$, and 0.6 U of *Taq* polymerase in 20 μL final volume. The conditions for PCR amplification were as follows: at 95°C for 3 min; followed by 36 cycles at 94°C for 30 s, anneal for 30 s, 72°C for 1 min, with final extension at 72°C for 10 min. The products of RT-PCR were tested by 1.5% agarose gel electrophoresis.

2.2.4 3′RACE and 5′RACE

RNA PCR Kit ver3.0 was used in reverse transcription reaction for 3′RACE according to instruction of the kit (TaKaRa). The PCR reaction mixture contained 1 μL cDNA template, 40 $\mu mol \cdot L^{-1}$ of each dNTP, 0.2 $\mu mol \cdot L^{-1}$ of M13 primer M4 and 3′RACE primers, 0.6 $mmol \cdot L^{-1}$ of $MgCl_2$, and 0.6 U of *Taq* polymerase in 20 μL final volume. The conditions for 3′RACE PCR amplification were as follows: at 95°C for 3 min; followed by 6 cycles at 94°C for 30 s, 61°C for 30 s, 72°C for 2 min; followed by 6 cycles of 94°C for 30 s, 59°C for 30 s, 72°C for 2 min; then 30 cycles at 94°C for 30 s, 57°C for 30 s, 72°C for 2 min with final extension at 72°C for 10 min. The product of 3′RACE PCR was tested by 1.5% agarose gel electrophoresis.

SMART RACE Kit was used in reverse transcription reaction for 5′RACE according to instruction of the kit (Clontech). The PCR reaction mixture contained 1 μL cDNA template, 40 $\mu mol \cdot L^{-1}$ of each dNTP, 0.6 $\mu mol \cdot L^{-1}$ of UPM primer and 0.2 $\mu mol \cdot L^{-1}$ of 5′RACE primer, 0.6 $mmol \cdot L^{-1}$ of $MgCl_2$, and 0.6 U of *Taq* polymerase in 20 μL final volume. The conditions for 5′RACE PCR amplification were as follows: at 95°C for 3 min; followed by 6 cycles at 94°C for 30 s, 72°C for 3 min; then another 6 cycles at 94°C for 30 s, 70°C for 30 s, 72°C for 3 min; then 30 cycles at 94°C for 30 s, 68°C for 30 s, 72°C for 3 min with the final extension at 72°C for 10 min. The product of 5′RACE PCR was tested by 1.5% agarose gel electrophoresis.

2.2.5 Sequencing

PCR products were purified by DNA gel extraction kit and connected with PMD-19T plasmid then transformed the JM110 bacterial. The bacterial was selected by

Blue-White Color Screening and bacterium was used for sequencing.

2.2.6 Bioinformatics analysis

ORF was analyzed by the online ORFFinder tool in NCBI. The cattle *Smad*4 gene cDNA was blasted with the newest cattle genome database to locate the cattle *Smad*4 gene in genome. The exons and introns were determined by GT/AG ruler. The isoelectric point and molecular weight were predicted by protparam tool. The subcellular localization and motif analysis were conducted by using the PSORT and Scanprosite tool in Expasy.

3 Results

3.1 Sequencing and RT-PCR

Figure 1 shows the results of RT-PCR by 1.5% agarose gel electrophoresis. The bands were consistent with the predicted PCR production length. The 2661 bp cattle *Smad*4 cDNA was obtained by splicing three cDNA fragments using SeqMan program after sequencing.

3.2 Results of 3′RACE and 5′RACE

Six hundred and forty-one bp 3′UTR sequence was obtained by 3′RACE PCR and 466 bp 5′UTR sequence was obtained by 5′RACE PCR. Figure 2 shows the results of 3′RACE and 5′RACE by 1.5% agarose gel electrophoresis.

3.3 Results of bioinformatics analysis

The cattle *Smad*4 gene consisted of 12 exons and 11 introns by comparing the cDNA sequence with genome using BLAST program.

Cattle *Smad*4 gene encoded 553 amino acid residues. The predicted molecular weight of cattle Smad4 protein was 60542.2 U and the isoelectric point was 6.50. This indicated that the Smad4 protein may be a weak acidic protein. No signal peptide and transmembrane domain were found. The matured Smad4 protein was likely located in cytoplasm and nucleus by subcellular localization analysis.

The cattle Smad4 protein had two conservative domains. The first one was named MH1 domain and consisted of 124 amino acid residues (18–142). The second was named MH2 domain and consisted of 165 amino acid residues (324–489). An alanine rich region was found in MH2 domain (452–489). Three glycosylation sites, four casein kinase II phosphorylation sites, seven N-myristoylation sites and three Protein kinase C phosphorylation sites were found in Smad4 protein.

3.4 Phylogenetic tress analysis of Smad4 protein

Cattle *Smad*4 cDNA shared 99%, 96%, 95%, 91% and 91% similarity in nucleic acid sequences and 99%, 98%, 98%, 99% and 98% similarity in amino acid sequences with sheep, pig, human, rat and mouse, respectively. This indicated *Smad*4 gene was very conservative in molecular evolution.

The phylogenetic tress was marked by BioEdit software (Fig. 3) and cattle had more genetic relationship with sheep in Fig. 3.

3.5 Tissue expression analysis of cattle *Smad*4 gene

*Smad*4 cDNA was found in testes, pancreas, liver, small intestine, ovary, lymph, cardiac muscle, skeleton muscle and thymus gland, which indicated *Smad*4 was broad expressed in cattle (Fig. 4). This was consistent with the fact that all the TGF-ß superfamily members need *Smad*4 for signal transmission.

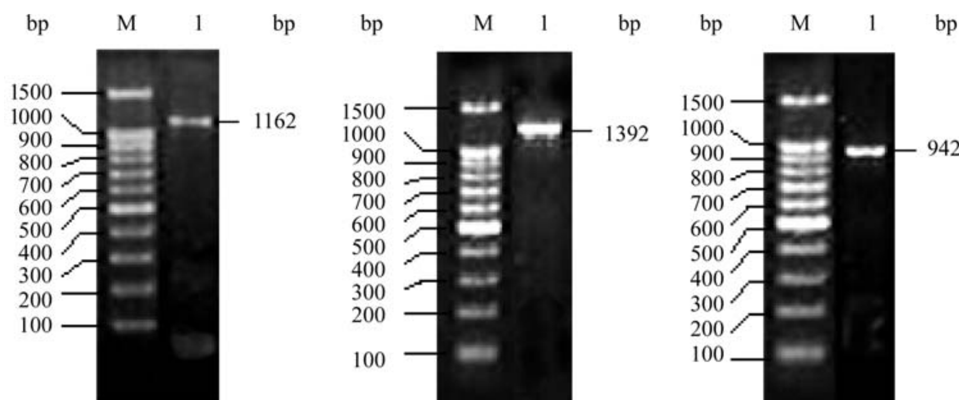

Fig. 1 Analysis of cattle *Smad*4 gene RT-PCR product by gel electrophoresis
Note: M represents 100 bp Plus DNA Ladder Marker; 1 represents product of RT-PCR.

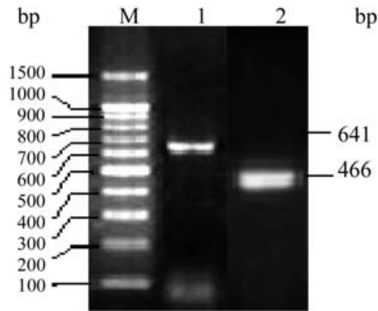

Fig. 2 Analysis of cattle *Smad*4 gene 3′ RACE and 5′ RACE product by gel electrophoresis
Note: M represents 100 bp Plus DNA Ladder Marker; 1 represents product of 3′ RACE and 2 represents product of 5′ RACE.

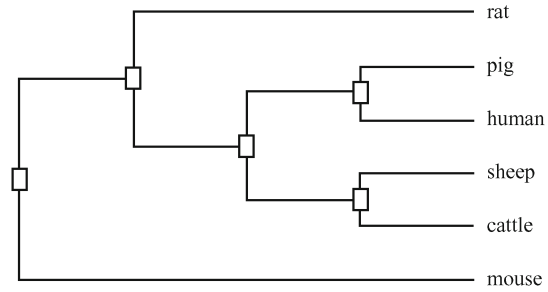

Fig. 3 Phylogenetic tree of Smad4 protein amino acids sequences

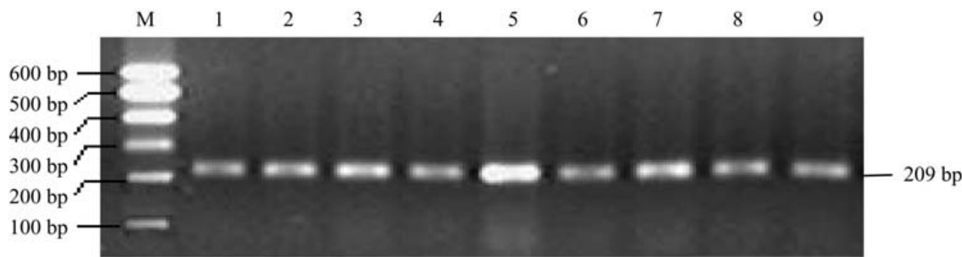

Fig. 4 The agarose gel electrophoresis analysis of *Smad*4 gene expression in various cattle tissues by RT-PCR
Note: M represents 100 bp Plus DNA Ladder marker; 1–9 represent testes, pancreas, liver, small intestine, ovary, lymph, cardiac muscle, skeleton muscle and thymus gland, respectively.

4 Discussion

Smad4 protein has three specific regions including the N-terminal and C-terminal that are highly conservative and the variable link region with rich alanine. The N-terminal, named MH1 domain, consists of about 130 amino acid residues (Grishin, 2001). The MH2 domain consisted of about 200 amino acid residues (Shi, 2001). The MH1 domain and MH2 domain could inhibit each other in base state.

The *Smad*4 gene had different transcript variants that may induce the variable length of link region of Smad4 protein. In recent years, two transcript variants of *Smad*4 gene were found in *Xenopus laevis* (Howell et al., 1999; Masuyama et al., 1999). The length of link regions were different in these transcript variants but they all could band with R-Smads and were similar in transcriptional activation. Six kinds of *Smad*4 transcript variants were found in humans and the length of link regions was also different but they all could bind with Smad2 and Smad3 to form composites with the help of transcription factor Fast-1. The activation of Smads protein needed exon5, exon 6 and exon 7 and the different transcript variants had different subcellular localizations (Christophe et al., 2000).

*Smad*4 was broadly expressed in the ovary, especially in follicles. Rat *Smad*4 had a specific spatiotemporal expression model with the development of the ovary. During the early stage of follicular development, *Smad*4 was mainly expressed in primary follicles and primordial follicles but the interstitial cells had very low Smad4 protein. With the development of the ovary, the expression of *Smad*4 increased in interstitial cells. The expression of *Smad*4 decreased in oocytes but increased in theca cells and had significant change in granulosa cells after sexual maturity. This indicates that *Smad*4 plays an important role in controlling follicular development (Miaozhu, 2005). *Smad*4 was expressed in the testes of 3 d, 7 d and 14 d and mature rats by immunohistochemistry (Hu et al., 2001). Other studies have also found Smad4 protein in the mammary gland and kidney. This accorded with the fact that all the signal transmission of TGF-β members needs Smad4 (Luukko, 2001).

References

Anita N, Signe K, Teresa K (2002). *Smad*4 overexpression causes germ cell ablation and leydig cell hyperplasia in transgenic Mice. American Journal of Pathology, 161: 342–356

Bouras M, Tabone E, Bertholon J, Sommer P, Bouvier R, Droz J P, Benahmed M (2000). A novel *Smad*4 gene mutation in seminoma germ cell tumors. Cancer Res, 60: 922–928

Chen X, Rubock M J, Whitman M (1996). A transcriptional partner for MAD proteins in TGF-ß signalling. Nature, 383: 691–696

Christophe E, Pierreux F J, Caroline S H (2000). Transforming Growth Factor ß independent shuttling of *Smad*4 between the cytoplasm and nucleus. Molecular and Cellular Biology, 20(23): 9041–9054

Germain S, Howell M, Esslemont G M, Hill C S (2000). Homeodomain and winged–helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. Genes Dev, 14(4): 435–451

Grishin N V (2001). Mh1 domain of Smad is a degraded homing endonuclease. Mo1 Biol, 307: 31–37

Hahn S A, Schutte M, Hoque A T (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21. Science, 271: 350–353

Hata A, Seoane J, Lagna G, Montalvo E, Hemmati-Brivanlou A, Massagué J (2000). OAZ uses distinct DNA-and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways. Cell, 100(2): 229–240

Heldin C H, Miyazono K, ten Dijke P (1997). TGF–ß signaling from cell membrane to nucleus through SMAD proteins. Nature, 390(6659): 465–471

Howell M, Itoh F, Pierreux C E, Valgeirsdóttir S, Itoh S, ten Dijke P, Hill C S (1999). Xenopus *Smad*4b is the co-Smadcomponent of developmentally- regulated transcription factor complexes responsible for induction of early mesodermal genes. Dev Biol, 214: 354–369

Hu J, Zhang Y Q, Liu X P (2001). Location studies of smad4 protein in the rat testis during postnatal development. Acta Biologiae Experimentalis Sinica, 47(5): 558–564 (in Chinese)

Liu H W, Li C, Jin L H, Liu Z Y, Li X J (1995). Ultrastructural studies and stereological analysis in the development of leydig cells of rat testis. Acta Anatomica Sinica, 16(4): 431–435 (in Chinese)

Luukko K, Ylikorkala A, Makela T P (2001). Developmentally regulated expression of Smad3, Smad4, Smad6, and Smad7 involved in TGF-beta signaling. Mech Dev, 101: 209–212

Massague J (1998). TGF-ß signal transduction. Annu Rev Biochem, 67: 753–791

Masuyama N, Hanafusa H, Kusakabe M (1999). Identification of two Smad4 proteins in Xenopus. Their common and distinct properties. J Biol Chem, 274: 12163–12170

Miao Z L, Wang Z N, Cheng L Q, Zhang Y (2005). Expression of *Smad*4 during rat ovarian development. Journal of First Military Medical University, 25(2): 127–131

Shi Y (2001). Structural insights on Smad function in TGF-ß signaling. Bioessays, 23(3): 223–232

Thiagalingam S, Lengauer C, Leach F S, Schutte M, Hahn S A, Overhauser J, Willson J K, Markowitz S, Hamilton S R, Kern S E, Kinzler K W, Vogelstein B (1996). Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. Nat Genet, 13(3): 343–346

Wong C E, Rougier J P, Frederick M B (1999). Smad3-Smad4 and AP-1 complexes synergize in transcriptional activation of the c-Jun promoter by transforming growth factor ß. Mol Cell Biol, 19: 1821–1830

Yanagisawa J, Yanagi Y, Masuhiro Y, Suzawa M, Watanabe M, Kashiwagi K, Toriyabe T, Kawabata M, Miyazono K, Kato S (1999). Convergence of transforming growth factor-ß and vitamin D signaling pathways on SMAD transcriptional coactivators. Science, 283: 1317–1321