Species-specific cis-Regulatory Elements in the 3′-Untranslated Region Direct Alternative Polyadenylation of Bone Morphogenetic Protein 2 mRNA*\[S\]

Received for publication, June 26, 2008, and in revised form, July 30, 2008. Published, JBC Papers in Press, August 14, 2008, DOI 10.1074/jbc.M804895200

Donglin Liu*, David T. Fritz, Melissa B. Rogers, and Aaron J. Shatkin From the Center for Advanced Biotechnology and Medicine, Piscataway, New Jersey 08854 and Department of Biochemistry and Molecular Biology, UMDNJ-New Jersey Medical School, Newark, New Jersey 07101

BMP2 (bone morphogenetic protein 2) is a multifunctional member of the transforming growth factor-β family of growth factors. Disruption of BMP2 signaling results in developmental defects, cancers, and other diseases. BMP2 mRNAs are alternatively polyadenylated, resulting in mRNAs with distinct 3′-untranslated regions. The longer mRNA contains additional putative binding sites for post-transcriptional regulatory factors, including micro-RNAs. We combined functional assays with computational analyses of emerging genome data to define site-specific polyadenylation determinants. In all mouse and human cell lines tested, shorter mRNAs resulting from using the first polyadenylation signal (PA1) were more abundant than mRNAs from the second signal (PA2). However, the PA1/PA2 usage ratios were 2–3-fold higher in human than in mouse cells. Expression of human BMP2 constructs in mouse cells and mouse constructs in human cells showed that cis-regulatory elements direct species-specific 3′ processing of BMP2 transcripts. A 72-nucleotide region downstream of PA2 in the mouse sequence contains two novel cis-acting elements previously hypothesized to regulate polyadenylation in a bioinformatics analysis. Mutations that humanized the mouse-specific elements lowered the affinity for cleavage stimulation factor CstF64 and significantly weakened the PA2 signal relative to the PA1 signal. Thus, we have experimentally defined for the first time cis-regulatory elements that control a species-specific difference in the 3′-end processing of BMP2 and potentially of other genes.

BMP2, a founding member of the bone morphogenetic protein family, plays critical roles in cardiac and skeletal development and numerous other processes (1–8). Understanding the mechanisms that regulate BMP2 synthesis is important, because abnormal BMP2 expression and signaling activity have been linked to many pathological conditions, including cancers (9–12), osteoporosis (13), osteoarthritis (14), and birth defects (15, 16).

Consistent with its irreplaceable roles in development and postnatal physiology, BMP2 is evolutionarily conserved at the level of function as well as sequence (17, 18). For example, the DPP (Drosophila decapentaplegic) protein, which is 71% identical to human BMP2, is functionally interchangeable with mammalian bone morphogenetic proteins in a mammalian bone induction assay (19). Unusually well conserved sequences also occur in BMP2 noncoding regions. Indeed, the 3′-UTRs of human and dog BMP2 transcripts include a region of 1080 nt that is 83% identical to the mouse transcript (18). The striking conservation of the 3′-UTRs of human, chimpanzee, several rodents, dog, deer, and cow implies vital post-transcriptional regulatory functions.

Regulation of gene expression through the 3′-UTR can include 3′-end processing, subcellular targeting, translational control, and differential mRNA stability (20–26). It has been suggested that the 3′-UTR is “a molecular hotspot for pathology” because of its major impact on tissue-specific gene regulation that, if disrupted, leads to developmental defects and life-threatening diseases (27). Because alternative cleavage and polyadenylation is an important level of BMP2 gene regulation (18), we analyzed the mechanisms that control BMP2 poly(A) signal selection.

Processing at the 3′-end is an obligatory step in the maturation of eukaryotic pre-mRNAs, with transcripts first losing a 3′-terminal fragment by endonucleolytic cleavage and then receiving a poly(A) tail in a tightly coupled reaction (20, 28–30). Many essential “core” elements and auxiliary regulatory elements have been identified using empirical and bioinformatics approaches (see Refs. 31–33 and references therein). A consensus hexamer, AAUAAA or a close variant, 10–35 nt upstream of the actual cleavage/polyadenylation site, is the core upstream element, also termed the polyadenylation signal. Another important determinant of 3′ processing is a (U/G)U-rich sequence downstream of the cleavage site, also known as the core downstream element (31, 34). The upstream and down-
stream core elements bind the essential polyadenylation factors, cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF), respectively (20). In addition, auxiliary elements upstream of the AAUAAA sequence and downstream of the (U/G)U-rich sequence have been identified as cis-acting regulatory elements that influence polyadenylation efficiency and mRNA maturation (28–30).

Alternative polyadenylation of the BMP2 3′-UTR yields two mRNAs differing only in their 3′-ends. Furthermore, the conserved region spanning the two poly(A) signals induces reporter gene expression in the mouse F9 embryonal carcinoma cell model of Bmp2 expression (18). Because poly(A) signal selection will include or exclude additional regulatory elements, such as micro-RNA and stability sequences, that may influence the function and metabolic fate of the mRNA, alternative polyadenylation may play a pivotal role in BMP2 expression.

Interestingly, the relative abundance of long and short transcripts differed significantly in mouse and human cells and suggests that BMP2 poly(A) signal selection is species- or cell type-specific (18).

To investigate cis- and trans-regulatory mechanisms that influence 3′-end formation, we examined the poly(A) patterns of BMP2 mRNAs in different cell types of mouse and human origin. Alternative polyadenylation of BMP2 mRNA occurs in all of the cell lines we tested. Although both human and mouse cell lines have some preference for the promoter-proximal polyadenylation (PA1) signal, the ratio of short to long transcripts differs significantly in mouse and human cells and influence the function and metabolic fate of the mRNA, alternative polyadenylation may play a pivotal role in BMP2 expression.

EXPERIMENTAL PROCEDURES

Cell Lines—The POS-1 murine osteosarcoma cell line (kindly provided by Yoshiyasu Nakamura, Kanagawa Cancer Center, Japan) was cultured in RPMI 1640 (Invitrogen, CA) medium supplemented with 10% fetal bovine serum (FBS; Valley Biomedical, VA). After reaching 70–80% confluence, cells were treated for 6 h with 100 ng/ml human sRANKL (PeproTech) to induce Bmp2 expression (35). Mouse F9 embryonal carcinoma cells were maintained and induced to differentiate as described previously (18). Mouse embryonic fibroblasts and NIH3T3 and human HeLa cervical cancer and embryonic kidney 293T cells were maintained in Dulbecco’s modified Eagle’s medium (Invitrogen) with 10% FBS. Human non-small cell lung carcinoma A549 cells were maintained in Dulbecco’s modified Eagle’s medium with 5% FBS, 1% penicillin/streptomycin, and 1% l-glutamine (Invitrogen), and the prostate cancer cell line CWR-22R was maintained in RPMI 1640 supplemented with 10% FBS, 1% penicillin/streptomycin, and 1% l-glutamine. The adult human normal prostate epithelial cell line MLC SV40 was maintained in keratinocyte-serum-free medium with 5 ng/ml recombinant epidermal growth factor, 25 ng/ml bovine pituitary extract (Invitrogen), and 1% penicillin/streptomycin. The human benign prostate epithelial cell line BPH-1 was cultured in RPMI 1640 with 10% FBS and 1% penicillin/streptomycin. All cultures were maintained in a humidified atmosphere of 5% CO₂ at 37 °C except for F9 cells, which were in 10% CO₂.

Plasmids—Unless otherwise indicated, nucleotide positions are given with respect to the mouse or human BMP2 stop codon (TAG). All restriction and modifying enzymes were from New England Biolabs (Ipswich, MA) except for T4 DNA ligase from Invitrogen. All constructs were confirmed by sequencing.

Templates for species-specific riboprobes were made by inserting an EcoRI site in the human sequence and a Stul site in both mouse and human sequences, excising a mouse Bmp2 Stul-EcoRI fragment (nt 627–1288) spanning the two poly(A) signals from pCBS-mB2, and inserting it into pBluescript II KS to generate pBSmB2-PA. Similarly, a corresponding human Bmp2 Stul-EcoRI fragment (nt 621–1300) was inserted into pBluescript II KS to generate pBShb2-PA. To detect poly(A) signal selection in transcripts expressed from chimeric constructs described below, the corresponding chimeric fragments were inserted into pBluescript II KS to generate plasmids for chimeric probe synthesis. All these plasmids were linearized with XhoI and transcribed with T7 RNA polymerase to produce antisense probes.

To test in vivo cleavage and polyadenylation, the pCBS (36) was used as parental plasmid. This vector contains a multiple cloning site downstream of the cytomegalovirus promoter and intron 1 of the rabbit β-globin gene accompanied by the splice donor and acceptor sites. To avoid interference, the bovine growth hormone poly(A) sequence (324 nt) was deleted from the vector.

For transfection assays, a PvuII-BglIII fragment containing the mouse Bmp2 3′-UTR and downstream flanking region (nt 86–1871) was excised from pGLB2-3′-UTR (18) and inserted into the Small–HindIII sites of vector pCBS to create pCBS-mB2 (mouse). A similar human BMP2 sequence (nt 117–1758) was generated by ligating HindIII-BamHI and BamHI-NsiI fragments containing nt 117–728 and nt 729–1758, respectively. These were from expressed sequence tag (EST) clone AA902384 (Invitrogen) and pmHBMP2-68 (a gift from Eileen M. Shore, University of Pennsylvania). Insertion of these ligated fragments into the HindIII and NsiI sites of pCBS created pCBS-hB2 (human). To make mouse-human chimeric constructs pCBSB2-Chim1 to -8, an EcoRI site was inserted at the human nt 1298 position of pCBS-hB2. The mouse sequence in pCBS-mB2 has a similarly positioned EcoRI site at nt 1281. An NsiI site was inserted at all other positions of both mouse and human sequences as indicated below so that various mouse sequences could be replaced with human sequences and vice versa. Specifically, the mouse regions nt 1019–1281, nt 1148–1281, nt 1148–1220, and nt 1220–1281 were exchanged with the human regions nt 1029–1298, nt 1151–1298, nt 1151–1233, and nt 1233–1298, respectively, in constructs pCBSB2-Chim1 and -5, pCBSB2-Chim2 and -6, pCBSB2-Chim3 and -7, and pCBSB2-Chim4 and -8 (Fig. 5A).

To prepare RNA probes for electrophoretic mobility shift assays (EMSA), plasmids containing the PA2 region (Fig. 1) were constructed by ligating the HindIII–EcoRI fragments of mouse wild type or mut6 (both nt 1100–1242) or human wild type (nt 1109–1255) PCR products into pGEM-4 vector. These plasmids were linearized with EcoRI and transcribed with T7 RNA polymerase to produce sense probes.
cis-Regulation of BMP2 mRNA Polyadenylation

Rapid Site-directed Mutagenesis Using Mutagenic Primer-directed DNA Replication—To map and define potential cis-acting elements that affect efficient cleavage and polyadenylation of BMP2 pre-mRNAs, site-directed mutants were prepared using the Stratagene QuikChange® II XL site-directed mutagenesis kit according to the manufacturer’s manual. In brief, mutagenic primer-directed amplification reactions were carried out at 95 °C for 1 min for initial denaturation and then 18 cycles of 95 °C for 50 s, 60 °C for 50 s, 68 °C for 6 min 20 s, followed by an extra extension of 68 °C for 7 min. To remove the remaining template after amplification, a DpnI digestion was performed for 1 h at 37 °C. The PCR product was then transformed into Escherichia coli XL10-Gold® ultracompotent cells. Positive clones were isolated and sequenced to verify mutagenesis and fidelity.

Transfection—Cells were seeded in 60-mm dishes ∼24 h before transfection. When cells reached 85–90% confluence, they were transfected as follows using Lipofectamine™ 2000 as delivery reagent. Plasmid DNA (8 μg) and Lipofectamine™ 2000 (20 μl) were each diluted into 0.5 ml of Opti-MEM® I medium without serum. After 5 min, the two solutions were combined, mixed gently, incubated at room temperature for 20 min, and added to a dish containing cells and medium. The pCB5 empty vector was used as negative control. After 24 h, cells were harvested, and total RNA was extracted immediately from the cell pellet.

Total RNA Isolation—Total RNA was extracted from cells by the TRIzol method (Invitrogen) according to the manufacturer’s instructions. In brief, cell pellets were lysed in TRIzol reagent (1 ml/10 cm² surface area of the plate) for 5 min, and chloroform (one-fifth volume) was added to the homogenate. The phases were separated by centrifugation at 12.000 × g for 15 min. RNA was precipitated from the aqueous phase by adding isopropl alcohol (one-half the volume of the TRIzol reagent) and washed with 75% ethanol. The total RNA was dissolved in ultrapure distilled water (Invitrogen). We also obtained RNAs from the human medulloblastoma cell line 2056, 3078, and 2061 from Dr. Cory Abate-Sheren (Center for Advanced Biotechnology and Medicine).

In Vitro Transcription of Strand-specific Riboprobes and RNase Protection Assays (RPAs)—Strand-specific, [α-32P]UTP-labeled riboprobes were synthesized in the presence of 50 μCi of 800 Ci/mmole [α-32P]UTP (PerkinElmer Life Sciences) using T7 polymerase (Promega) according to the supplier’s protocol. RNAs were extracted with phenol/chloroform and purified by a G-50 Sephadex Quick Spin column (Roche Applied Science). RPAs were performed using the RPA III kit (Ambion Inc.) according to the manufacturer’s manual. Briefly, total RNA and radiolabeled RNA probe were co-precipitated with ethanol, resuspended in Hybridization Buffer III, denatured at 90–95 °C for 3–4 min, and hybridized overnight at 42 °C. After 30 min of RNase A/T1 digestion at 37 °C, reactions were inactivated with RNase Inactivation Solution III. Subsequently, RNAs were precipitated, dissolved in 7 μl of gel loading buffer, and analyzed by PAGE in denaturing 5% gels (8 μ mole; 37.5:1, acrylamide/bisacrylamide). Protected RNAs were visualized by autoradiography and quantified using GE Healthcare PhosphorImager and ImageQuant software.

Electrophoretic Mobility Shift Assays—EMSAs were carried out as described previously (38) with modifications. In brief, [α-32P]UTP-labeled RNA probes (12.5 ng, 0.6–1 × 10⁶ cpm) were incubated with increasing levels (0–225 ng) of recombinant human CstF64 (1–285) protein in 12.5 μl of binding buffer containing 8 mM HEPES-NaOH (pH 7.9), 40 mM NaCl, 2 mM EDTA, 0.2 mM dithiothreitol, 0.2 mM phenylmethylsulfonyl fluoride, and 8% (v/v) glycerol at 30 °C for 10 min. The reaction mixtures were loaded on prerun (1 h) 5% non-denaturing polyacrylamide gels, electrophoresed, dried, and exposed to x-ray film or a PhosphorImager plate for quantification.

Statistical Analyses—Results were corrected for UMP content of the probes and are expressed ± S.D. Analyses were performed by three-sample two-tailed Student’s t test. p values of <0.05 were considered significant.

RESULTS

Conservation of cis-Regulatory Elements between Mammals and Chicken—Alternative polyadenylation in mouse and human cells resulted in BMP2 mRNAs with short (mouse ∼870 nt; human ∼880 nt) or long (mouse ∼1185 nt; human ∼1188 nt) 3′-UTRs (18). We previously demonstrated that the proximal (PA1) and distal (PA2) poly(A) regions were conserved among four orders of mammals (18). However, because poly(A) signal selection requires specific sequences downstream of the cleavage and polyadenylation site, our analysis was limited to those species for which genomic sequence was available. Because evolutionary conservation of sequence indicates important functional constraints and enables prioritization of sequences for experimental analysis, we aligned BMP2 poly(A) regions from newly completed genomes for other species, most importantly chicken. Remarkably, using the MultiPipMaker global sequence alignment program (available on the World Wide Web), we discovered that the entire chicken 3′-UTR and flanking sequence, including the poly(A) region, is extensively conserved with mammals (Fig. 1) (18). Indeed, two AAUAAA hexamers (the CPSF-binding element used in 53–59% of human and mammalian genes, respectively (32)) and (U/G)U-rich elements in similar relative positions are clearly apparent in chickens and suggest that alternative polyadenylation of BMP2 arose prior to separation of the mammalian and avian lineages ∼310 million years ago (39).

Although the CPSF and CstF binding sites are the most important core determinants for cleavage and polyadenylation, additional upstream and downstream elements contribute to this process in a combinatorial manner. Because such cis-regulatory elements would play an essential role in controlling BMP2 polyadenylation and thus BMP2 mRNA synthesis, we used a bioinformatics approach to map putative regulatory motifs. The program PolyA_SVM analyzes 15 cis-elements grouped as follows: auxiliary upstream sequence elements (−100 to −41 nt); core upstream sequence elements (−40 to −1 nt; i.e. CPSF binding sites); core downstream sequence elements (+1 to +40 nt; i.e. CstF binding sites); and auxiliary downstream sequence elements (+41 to +100 nt) (40). The similarity of sequences to known elements and their relative...
locations are scored to provide an estimate of site strength. The PolyA_SVM cut-off score represents the inverse probability of poly(A) signal use (the lower the number, the stronger the predicted signal). Based on the previously established mouse and human poly(A) signals, the chicken sequence was predicted to have two strong poly(A) signals (supplemental Fig. 1). As described below, we have characterized the function of these putative regulatory elements.

**Alternative Polyadenylation of Endogenous BMP2 mRNAs in Mouse and Human Cell Types**—Interestingly, the relative scores for PA1 and PA2 within each species differed. Specifically, the mouse and chicken signals had similar scores (mouse, 1.187 and 1.530; chicken, 1.224 and 2.529). Although PA1 was predicted to be stronger in all species, the human scores diverged more (2.949 and 7.680). This suggested hypotheses that PA1 would be the dominant, most frequently used site in mouse and rat, while PA2 would be more frequently used in human and chimpanzee, and possibly other species as well.

**FIGURE 1. Sequence alignment of BMP2 mRNA poly(A) regions.** Two canonical AAUAAA polyadenylation signals, PAS1 and PAS2 (shaded), and downstream G/U- and U-rich sequences (underlined) are conserved among four mammalian orders and chicken. Three cis-acting elements were experimentally defined (boxed) by in vivo polyadenylation. The PA2 region of mouse and human sequences are marked (dashed underline) as the templates for RNA probes in EMSA. Numbers are relative to the mouse stop codon. Nucleotides identical to mouse are indicated by a dot, and gaps are shown by a dash. Known cleavage and polyadenylation sites are indicated by a caret. Sequence accession numbers are as follows: mouse, Mus musculus strain C57Bl/6, NT_039207; rat, Rattus norvegicus, NW_047658; human, Homo sapiens, NT_011387; chimpanzee, Pan troglodytes, NW_001230484; monkey, Macaca mulatta, NW_001095147; cow, Bos taurus, NW_928644; dog, Canus familiaris, NW_876277; chicken, Gallus gallus, NW_001471677.
cis-Regulation of BMP2 mRNA Polyadenylation

**TABLE 1**

Survey of BMP2 3′-UTR EST polyadenylation

| Source                  | PA1 | PA2 | Accession number |
|-------------------------|-----|-----|------------------|
| **Mouse**               |     |     |                  |
| Urinary bladder         | 1   |     | BB124190         |
| Corpora quadrigeinina   | 1   |     | BB299912         |
| Skin                    | 2   |     | AV231557         |
| Embryo                  | 1   |     | AV259587         |
| Liver tumor             | 2   |     | AV305394         |
| Spleen                  | 1   |     | BB717842         |
| Mammary gland           | 1   |     | BB719777         |
| Medulla oblongata       | 2   |     | AV333263         |
| Cecum                   | 1   |     | AV374010         |
| Cornea enriched for endothelium and stroma | 1 |     | DV553298         |
| Decidual tissue         |     | 1   | D18317           |
| Diencephalon            |     | 1   | BB805888         |
| Trophoblast             |     | 1   | BQ030901         |
| Total                   | 9   | 7   |                  |
| **Human**               |     |     |                  |
| Fibroblast W1-38        | 1   |     | F13604           |
| Fetal heart             | 2   |     | AA011062         |
| Pooled organs           | 1   |     | BX101090         |
| Colon tumor             | 1   |     | BF438972         |
| Left pelvis             |     | 1   | BQ003843         |
| Glioblastoma            | 2   |     | CB055215         |
| Fibrosarcoma            | 1   |     | BQ020545         |
| Eye lens                | 1   |     | BM670141         |
| Anaplastic oligodendrogloma | 1 |     | AI686047         |
| Placenta                | 1   |     | BX424085         |
| Stomach colon           | 1   |     | AI783838         |
| Pancreatic islet        | 1   | 1   | AI707471         |
| HeLa S3 line            | 1   | 1   | C75049           |
| Lung                    | 1   |     | RP386830         |
| Total                   | 16  | 3   |                  |

All species and that PA2 would be particularly weak in human cells. To test the use of these predicted BMP2 poly(A) signals in a variety of mouse and human cell types, we surveyed the current EST data bases for mouse and human BMP2 mRNA 3′-UTRs that reflected utilization of either the promoter proximal or distal BMP2 poly(A) signals (Table 1). According to the UniGene transcript end assignment, nine mouse ESTs end at the PA1 site, and seven end at the PA2 site. In contrast, 16 human ESTs end at PA1, and three end at PA2. Thus, as predicted by the bioinformatics analysis, human mRNAs from many cell types end at PA1 more frequently than PA2.

The bioinformatics prediction and relative abundance of PA1-derived ESTs from human cells mirrored the relatively greater abundance of shorter mRNAs that we observed previously in human HeLa cells (28). We also assessed the relative abundance of the two forms of BMP2 mRNAs in normal and malignant mouse and human cell lines using RPAs. Mouse- and human-specific antisense probes spanning the two predicted cleavage and polyadenylation sites were protected by RNAs corresponding to PA1 or PA2 (Fig. 2, A and B), showing that both poly(A) signals were utilized in each of the cell lines. As predicted by the higher PolyA_SVM score (supplemental Fig. 1) and the EST distribution (Table 1), all cell lines preferentially produced PA1 transcripts (corrected for UMP content). However, whereas the ratios of PA1 to PA2 transcripts averaged 2.38 ± 0.34 in mouse cell lines (Fig. 2C, left), the PA1/PA2 ratios in all human cell lines averaged 4.11 ± 0.93 (Fig. 2C, right). Thus, all tested human cell types significantly underexpressed the longer mRNAs ending at PA2 relative to those ending at PA1.

The bioinformatics analysis and EST and RNase protection results raised two questions. First, what is the basis for the more frequent use of PA1 versus PA2? This question is relevant to the mechanisms that control alternative polyadenylation. Second, what makes PA2 particularly weak in human cells relative to mouse cells? This question addresses the subtle regulatory differences that control the expression of vital proteins in different species. We have begun to address these questions using reporter gene and protein binding assays.

**BMP2 mRNA Cleavage Site Choice Is Directed by 3′-UTR and Downstream Sequences**—The differential polyadenylation patterns in mouse versus human cells may have resulted from species-specific cis-regulatory elements in the DNA or RNA molecules, from species-specific trans-factor differences, or from both. Identifying the mechanism(s) controlling the use of PA1 or PA2 in the BMP2 context would help elucidate the important regulatory process of alternative polyadenylation. Furthermore, because the BMP2 poly(A) regions are exceedingly well conserved, the relevant cis-elements must be restricted to the small fraction of sequence that is not conserved between the mouse and human genes. To compare the relative importance of species-specific cis-elements and trans-factors, we subcloned fragments with ~1.8 kb of mouse or human BMP2 3′-UTR and downstream sequence into the poly(A) reporter vector pCβS (Fig. 3A). We then transfected the mouse construct, pCβS-m2B, into mouse embryonic fibroblasts and the human construct, pCβS-h2B, into human HeLa cells. Both mouse and human constructs retained the polyadenylation patterns of endogenous BMP2 mRNAs (Fig. 2, B and C) with PA1/PA2 ratios of 2-fold in mouse cells and more than 3-fold in human cells (reporter gene data not shown). We then transfected the mouse construct, pCβS-m2B, into three human cell lines and the human construct, pCβS-h2B, into three mouse cell lines (Fig. 3B). The mouse Bmp2 construct in all human cell lines yielded RNAs with PA1/PA2 usage ratios (Fig. 3C) resembling those of endogenous mouse Bmp2 mRNAs (i.e. PA1 transcripts were no more than 2.5 times as abundant as PA2 transcripts). In contrast, the human BMP2 construct in all mouse cell lines produced RNAs with PA1/PA2 usage ratios resembling those of the endogenous human mRNAs (i.e. PA1 transcripts were 4–5 times as abundant as PA2 transcripts). These results indicate that cis-acting elements in the BMP2 gene sequence, rather than human specific trans-factors, account for the relatively greater abundance of human PA1 BMP2 mRNAs.

**A Motif That Contributes to the Greater Strength of the Promoter-proximal Polyadenylation Signal (PA1) in Mouse and Human Cells**—We previously observed that transcripts ending at the PA1 signal were at least 2-fold more abundant than the longer RNAs ending at PA2 in HeLa cells (28). Furthermore, RNase protection assays of RNA from several additional mouse and human cell lines and analysis of ESTs (Fig. 2 and Table 1) replicated this finding.

A UUUUUU motif present between the promoter proximal AAUAAA hexamer and the PA1 cleavage and polyadenylation site
is conserved among all mammalian genes (Fig. 1). Bioinformatics studies have indicated that U-rich elements are overrepresented in strong poly(A) signals (31). To test if this element contributes to preferential polyadenylation at PA1, we mutated the UUUCUU motif to CCGCUU. Consistent with the hypothesis that UUU-CUU promotes polyadenylation, mouse transcripts ending at PA1 were reduced by 33% relative to those ending at PA2. Likewise, human transcripts ending at PA1 were reduced by 22% (Fig. 4). This experiment is one of the first empirical tests of an element identified by purely bioinformatics approaches.

To map more precisely potential cis-regulatory elements, we exchanged several smaller sequences (Fig. 5, constructs 2–4 and 6–8). The dramatic skewing of poly(A) signal choice to PA1 was achieved by replacing only 72 nt (mouse nt 1148–1220) immediately downstream of the mouse PA2 consensus hexamer AAUAAA with the corresponding human sequence (nt 1151–1233, construct 3; Fig. 5, B and C, lane 3). In contrast, construct 4, made by replacing an adjacent downstream region (mouse nt 1220–1281) with human sequence (nt 1233–1298), generated a polyadenylation pattern approaching that observed with the intact mouse sequence (Fig. 5, B and C, lane 4, compared with Figs. 2 and 3). In other words, replacement of mouse nt 1148–1220 with human nt 1151–1233 significantly crippled the mouse PA2 signal.

Within the human context, the entire mouse PA2 region or mouse nt 1148–1281 significantly promoted the production of PA2 transcripts at the expense of PA1 transcripts (Fig. 5, B and C, lanes 5 and 6). However, neither the minimal mouse region (nt 1148–1220), whose replacement with the human sequence skewed poly(A) site choice to PA1 (Fig. 5B, lane 3), nor its adjacent downstream region (nt 1220–1281) alone enhanced PA2 use relative to PA1 (Fig. 5B, lanes 7 and 8). Thus, the flanking elements required for recognition of polyadenylation signals often reside within 100 nt of the cleavage/polyadenylation site (31, 32), we replaced the mouse PA2 region (nt 1019–1281) with the human PA2 region (Fig. 5, construct 1). The intact mouse BMP2 sequence efficiently generated transcripts resulting from both signals in 293T cells, yielding a modest PA1/PA2 mRNA ratio of 2.34 ± 0.03 (Fig. 3C). In striking contrast, chimeric PA2 transcripts were nearly absent, although the PA1 site was used efficiently (Fig. 5B, lane 1). Conversely, replacing the human PA2 region (nt 1029–1298) with the mouse PA2 region (Fig. 5A, construct 5) generated transcripts resulting predominantly from PA2 (PA1/PA2 = 0.46 ± 0.02; Fig. 5, B and C, lane 5). Thus, the mouse PA2 region contains specific regulatory elements that enhance Bmp2 mRNA cleavage and polyadenylation activity relative to the PA1 site. In contrast, the human PA2 region is particularly weak within the mouse context.

To map more precisely potential cis-regulatory elements, we exchanged several smaller sequences (Fig. 5, constructs 2–4 and 6–8). The dramatic skewing of poly(A) signal choice to PA1 was achieved by replacing only 72 nt (mouse nt 1148–1220) immediately downstream of the mouse PA2 consensus hexamer AAUAAA with the corresponding human sequence (nt 1151–1233, construct 3; Fig. 5, B and C, lane 3). In contrast, construct 4, made by replacing an adjacent downstream region (mouse nt 1220–1281) with human sequence (nt 1233–1298), generated a polyadenylation pattern approaching that observed with the intact mouse sequence (Fig. 5B, lane 4, compared with Figs. 2 and 3). In other words, replacement of mouse nt 1148–1220 with human nt 1151–1233 significantly crippled the mouse PA2 signal.

Within the human context, the entire mouse PA2 region or mouse nt 1148–1281 significantly promoted the production of PA2 transcripts at the expense of PA1 transcripts (Fig. 5, B and C, lanes 5 and 6). However, neither the minimal mouse region (nt 1148–1220), whose replacement with the human sequence skewed poly(A) site choice to PA1 (Fig. 5B, lane 3), nor its adjacent downstream region (nt 1220–1281) alone enhanced PA2 use relative to PA1 (Fig. 5B, lanes 7 and 8). Thus, the flanking

![Figure 2: Identification of mouse and human BMP2 RNA cleavage and polyadenylation sites by RNase protection assays.](image-url)
human sequence was sufficient to generate the typical human PA1/PA2 ratio of 4–5.

The sequence alignment (Fig. 1) revealed that the essential regulatory region (nt 1148–1220) downstream of the mouse distal polyadenylation signal (PA2) is ~70% identical to the corresponding human region. We converted seven sites in the mouse pC/H9252S-mB2 poly(A) reporter construct to the human sequence (Fig. 6A) and transfected these mutated constructs into 293T cells (Fig. 6, B and C). A four-nucleotide conversion (mutation 6; a combination of mutations 5 and 7) significantly weakened the PA2 signal relative to the PA1 signal (Fig. 6, B and C, lane 6). These humanizing mutations altered two sequences, GUGUCU and CUGUGU, which have been previously annotated as potential core downstream elements in a bioinformatics survey of strong poly(A) regions (31). Our results indicate that these two elements stimulate cleavage and polyadenylation at the mouse PA2 site.

Mutation 6 humanized two putative core downstream elements, GUGUCU and CUGUGU (31). Because both motifs have GU-rich characteristics associated with binding sites for the 64-kDa subunit of CstF (CstF-64) (41), we postulated that increased CstF-64 affinity accounts for the higher efficiency of the distal mouse site. To test this hypothesis, we used EMSAs to compare the affinity of purified recombinant human CstF-64-(1–285), including the RNA binding domain (amino acids 17–96) and a loop region (38), for three RNAs containing the mouse distal polyadenylation signal with or without mutation 6 (Fig. 6A) or the human distal polyadenylation signal. Although both mouse and human RNAs bound CstF64-(1–285), 50% of the mouse RNA was bound at half the protein concentration relative to the human RNA (Fig. 7). Thus, the affinity of the mouse pA2 site is 2-fold greater than the human pA2 site for CstF64-(1–285). Furthermore, mutation 6, which humanized the two core downstream elements and reduced use of the pA2 site within the mouse context (compare Fig. 6C with Fig. 2C), bound the recombinant CstF64 more weakly than either of the wild type RNAs. The results indicate that species-specific changes in CstF64 binding influence polyadenylation site use in different species.

**DISCUSSION**

Polyadenylation is an essential step in the maturation of all eukaryotic cellular mRNAs with the exception of histone mRNAs (42, 43). Alternative polyadenylation, which occurs in the majority of mRNAs, generates mature transcripts with different 3′-ends and is an important mechanism for regulating differential gene expression (21, 30). Because precisely regulat-
ing the level of BMP2 is essential for normal development and physiology, we mapped conserved and nonconserved regulatory elements within the BMP2 3’-UTR using computational and biochemical approaches.

The 3’-terminal regions of mammalian and chick BMP2 genes have two sets of canonical AAUAAA hexamers and downstream GU/U rich elements. The two predicted poly(A) signals functioned with at least a 2-fold preferential use of the PA1 signal in all mouse and human cells examined (Fig. 2 and Table 1). We now provide mutational data indicating that a UUUCUU motif between the promoter proximal AAUAAA polyadenylation signal and cleavage site partially contributes to preferential polyadenylation at PA1 (Fig. 4). This element is conserved among all mammalian genes (Fig. 1). Furthermore, bioinformatics studies have indicated that U-rich elements are overrepresented in strong poly(A) signals (31). This result validates a putative cis-regulatory element predicted solely from bioinformatics studies.

In human cells, the PA1 signal was used 3.2–5.4 times more frequently than PA2 (Fig. 2C). In contrast, the PA1/PA2 ratio was close to 2.4 in all mouse cell lines tested. This species-specific difference proved to be predominantly directed by cis-regulatory sequences in the 3’-UTR rather than by trans-acting factors in the different cell types (Fig. 3). Chimeric constructs containing human PA2 sequences in place of the corresponding mouse sequences severely impaired PA2 selection without impacting usage of the PA1 cleavage/polyadenylation site (Fig. 5, constructs 1–3). By deletion and point mutation analyses, we determined that a short region downstream of the PA2 consensus hexamer was responsible for the relatively efficient usage of this site in the mouse gene (Figs. 5 and 6). Since this downstream region is not in the mRNA, its effect on PA2 usage and the relative amounts of long and short Bmp2 mRNAs is unlikely to be due to differential stability. Also consistent with cleavage site selection rather than stability, PA1/PA2 ratios were increased by only 15% in RNAs (kindly provided by Dr. M. Gorospe) from HeLa cells in which HuR, a ubiquitously expressed RNA-binding protein that selectively binds and stabilizes ARE-
containing mRNAs, was >90% decreased by small interfering RNA treatment (data not shown).

Two elements in the PA2 downstream region, GUGUCU and CUGUGU, were previously suggested to be core downstream elements by bioinformatics analyses (31). Our mutational and biochemical analyses demonstrated for the first time that these elements are functional elements involved in species-specific CstF64 binding (Fig. 7) and polyadenylation site choice (Fig. 6). Experimental evidence of this type is necessary to assess the relevance of subtle differences between putative cis-regulatory elements identified by bioinformatics approaches.

In contrast to the human PA2 signal, the mouse PA2 signal functioned very efficiently in the human context (Fig. 5, constructs 5 and 6). Furthermore, the presence of the mouse PA2 signal or a downstream sequence (nt 1151–1298) weakened the human PA1 site. However, the essential mouse sequence containing mRNAs, was >90% decreased by small interfering RNA treatment (data not shown).

Two elements in the PA2 downstream region, GUGUCU and CUGUGU, were previously suggested to be core downstream elements by bioinformatics analyses (31). Our mutational and biochemical analyses demonstrated for the first time that these elements are functional elements involved in species-specific CstF64 binding (Fig. 7) and polyadenylation site choice (Fig. 6). Experimental evidence of this type is necessary to assess the relevance of subtle differences between putative cis-regulatory elements identified by bioinformatics approaches.

In contrast to the human PA2 signal, the mouse PA2 signal functioned very efficiently in the human context (Fig. 5, constructs 5 and 6). Furthermore, the presence of the mouse PA2 signal or a downstream sequence (nt 1151–1298) weakened the human PA1 site. However, the essential mouse sequence containing mRNAs, was >90% decreased by small interfering RNA treatment (data not shown).

Two elements in the PA2 downstream region, GUGUCU and CUGUGU, were previously suggested to be core downstream elements by bioinformatics analyses (31). Our mutational and biochemical analyses demonstrated for the first time that these elements are functional elements involved in species-specific CstF64 binding (Fig. 7) and polyadenylation site choice (Fig. 6). Experimental evidence of this type is necessary to assess the relevance of subtle differences between putative cis-regulatory elements identified by bioinformatics approaches.

In contrast to the human PA2 signal, the mouse PA2 signal functioned very efficiently in the human context (Fig. 5, constructs 5 and 6). Furthermore, the presence of the mouse PA2 signal or a downstream sequence (nt 1151–1298) weakened the human PA1 site. However, the essential mouse sequence containing mRNAs, was >90% decreased by small interfering RNA treatment (data not shown).

Two elements in the PA2 downstream region, GUGUCU and CUGUGU, were previously suggested to be core downstream elements by bioinformatics analyses (31). Our mutational and biochemical analyses demonstrated for the first time that these elements are functional elements involved in species-specific CstF64 binding (Fig. 7) and polyadenylation site choice (Fig. 6). Experimental evidence of this type is necessary to assess the relevance of subtle differences between putative cis-regulatory elements identified by bioinformatics approaches.

In contrast to the human PA2 signal, the mouse PA2 signal functioned very efficiently in the human context (Fig. 5, constructs 5 and 6). Furthermore, the presence of the mouse PA2 signal or a downstream sequence (nt 1151–1298) weakened the human PA1 site. However, the essential mouse sequence containing mRNAs, was >90% decreased by small interfering RNA treatment (data not shown).

Two elements in the PA2 downstream region, GUGUCU and CUGUGU, were previously suggested to be core downstream elements by bioinformatics analyses (31). Our mutational and biochemical analyses demonstrated for the first time that these elements are functional elements involved in species-specific CstF64 binding (Fig. 7) and polyadenylation site choice (Fig. 6). Experimental evidence of this type is necessary to assess the relevance of subtle differences between putative cis-regulatory elements identified by bioinformatics approaches.

In contrast to the human PA2 signal, the mouse PA2 signal functioned very efficiently in the human context (Fig. 5, constructs 5 and 6). Furthermore, the presence of the mouse PA2 signal or a downstream sequence (nt 1151–1298) weakened the human PA1 site. However, the essential mouse sequence containing mRNAs, was >90% decreased by small interfering RNA treatment (data not shown).

Two elements in the PA2 downstream region, GUGUCU and CUGUGU, were previously suggested to be core downstream elements by bioinformatics analyses (31). Our mutational and biochemical analyses demonstrated for the first time that these elements are functional elements involved in species-specific CstF64 binding (Fig. 7) and polyadenylation site choice (Fig. 6). Experimental evidence of this type is necessary to assess the relevance of subtle differences between putative cis-regulatory elements identified by bioinformatics approaches.

In contrast to the human PA2 signal, the mouse PA2 signal functioned very efficiently in the human context (Fig. 5, constructs 5 and 6). Furthermore, the presence of the mouse PA2 signal or a downstream sequence (nt 1151–1298) weakened the human PA1 site. However, the essential mouse sequence containing mRNAs, was >90% decreased by small interfering RNA treatment (data not shown).

Two elements in the PA2 downstream region, GUGUCU and CUGUGU, were previously suggested to be core downstream elements by bioinformatics analyses (31). Our mutational and biochemical analyses demonstrated for the first time that these elements are functional elements involved in species-specific CstF64 binding (Fig. 7) and polyadenylation site choice (Fig. 6). Experimental evidence of this type is necessary to assess the relevance of subtle differences between putative cis-regulatory elements identified by bioinformatics approaches.

In contrast to the human PA2 signal, the mouse PA2 signal functioned very efficiently in the human context (Fig. 5, constructs 5 and 6). Furthermore, the presence of the mouse PA2 signal or a downstream sequence (nt 1151–1298) weakened the human PA1 site. However, the essential mouse sequence containing mRNAs, was >90% decreased by small interfering RNA treatment (data not shown).

Two elements in the PA2 downstream region, GUGUCU and CUGUGU, were previously suggested to be core downstream elements by bioinformatics analyses (31). Our mutational and biochemical analyses demonstrated for the first time that these elements are functional elements involved in species-specific CstF64 binding (Fig. 7) and polyadenylation site choice (Fig. 6). Experimental evidence of this type is necessary to assess the relevance of subtle differences between putative cis-regulatory elements identified by bioinformatics approaches.
cis-Regulation of BMP2 mRNA Polyadenylation

between nt positions 1148 and 1220 failed to have much effect on the PA1 site or to change the site usage ratio in the context of human sequence (Fig. 5, construct 7), although the equivalent human sequence crippled the mouse PA2 signal (Fig. 5, construct 3). Combinatorial gene regulation is an overriding feature of post-transcriptional regulation (21, 30, 44, 45). Our results demonstrate that species-specific cis-regulatory elements within the BMP2 3′-UTR control the site and efficiency of mRNA 3′-end processing in different animals. A central tenet of biology is that conserved organismal features, such as the nearly identical BMP2 protein sequences in mammals, reflect vital functions that tolerate limited variation. Conversely, subtle changes in the synthesis of these essential morphogenetic factors, as described here, must have permitted the diverse morphologies observed in the animal kingdom to evolve (46, 47).

In summary, we have mapped regulatory elements that contribute to the complex mechanisms controlling alternative polyadenylation of the BMP2 transcript. Alternative polyadenylation leading to mRNAs with distinct 3′-UTRs, along with usage of distinct promoters (48, 49) and other regulatory mechanisms (15, 17, 18, 50, 51), are important pieces of the mechanisms controlling alternative polyadenylation of the BMP2 transcript. Alternative polyadenylation within the BMP2 3′-UTR control the site and efficiency of mRNA 3′-end processing in different animals. A central tenet of biology is that conserved organismal features, such as the nearly identical BMP2 protein sequences in mammals, reflect vital functions that tolerate limited variation. Conversely, subtle changes in the synthesis of these essential morphogenetic factors, as described here, must have permitted the diverse morphologies observed in the animal kingdom to evolve (46, 47).

Acknowledgments—We thank Drs. Carol S. Lutz and Bin Tian for helpful comments; Dr. Liang Tong and Yun Bai for purified CstF64 (1–285); Dr. Chun Chu for assistance; and Drs. Cory Abate-Shen, Myriam Gorospe, Rudi Pullmann, and James M. Olson for RNAs.

REFERENCES

1. Chen, D., Zhao, M., and Mundy, G. R. (2004) Growth Factors 22, 233–241
2. Hogan, B. L. (1996) Genes Dev. 10, 1580–1594
3. Reddi, A. H. (2001) J. Bone Joint Surg. Am. 83, Suppl. 1, 1–6
4. Urist, M. R. (1965) Science 150, 893–899
5. Wozney, J. M., Van Der Stricht, A., Duchscher, J., Chipman, G. W., Mundy, G. R., and Cafferty, J. A. (1988) Science 242, 1528–1534
6. Chen, D., Zhao, M., Harris, S. E., and Mi, Z. (2004) Front. Biosci. 9, 349–358
7. Riley, E. H., Lane, J. M., Urist, M. R., Lyons, K. M., and Lieberman, J. R. (1996) Clin. Orthop. Relat. Res. 324, 39–46
8. Shimakawa, S., Moore, R. K., Otsuka, F., and Erickson, G. F. (2004) Endocr. Rev. 25, 72–101
9. Hardwick, J. C., Van Den Brink, G. R., Bleumling, S. A., Ballester, I., Van Der Brande, J. M., Keller, J. J., Offerhaus, G. J., Van Deventer, S. J., and Sutherland, R. L. (2004) J. Bone Joint Surg. Am. 86, 2517–2521
10. Nishimura, N., Kuroda, M., and Shimizu, N. (2004) Int. J. Oncol. 24, 533–539