GENOMICS 17, 408–415 (1993)

The Human Bone Sialoprotein Gene (IBSP): Genomic Localization and Characterization

JANET M. KERR,*,[1] LARRY W. FISHER,* JOHN D. TERMINE,*,[2] MARY G. WANG,† O. WESLEY MCBRIDE,† AND MARIAN F. YOUNG*

*Bone Research Branch, National Institute of Dental Research, and †National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

Received October 12, 1992; revised April 23, 1993

We have isolated and partially sequenced the human bone sialoprotein gene (IBSP). IBSP has been sublocalized by *in situ* hybridization to chromosome 4q28–q31 and is composed of six small exons (51 to 159 bp) and 1 large exon (~2.6 kb). The intron/exon junctions defined by sequence analysis are of class 0, retaining an intact coding triplet. Sequence analysis of the 5′ upstream region revealed a TATAA (nucleotides −30 to −25 from the transcriptional start point) and a CCAAT (nucleotides −56 to −52) box, both in the reverse orientation. Intron 1 contains interesting structural elements composed of polypyrimidine repeats followed by a poly(AC)$_n$ tract. Both types of structural elements have been detected in promoter regions of other genes and have been implicated in transcriptional regulation. Several differences between the previously published cDNA sequence (L. W. Fisher *et al.*, 1990, *J. Biol. Chem.* 265, 2347–2351) and our sequence have been identified, most of which are contained within the untranslated exon 1. Three base revisions in the coding region include a G to T (Gly to Val, amino acid 195), T to C (Val to Ala, amino acid 268), and T to A (Glu to Asp, amino acid 270). In conclusion, the genomic organization and potential regulatory elements of human IBSP have been elucidated. © 1993 Academic Press, Inc.

INTRODUCTION

Bone sialoprotein (BSP) is an ~70,000 M_r acidic glycoprotein that undergoes extensive post-translational modifications, including glycosylation (Fisher *et al.*, 1983) and sulfation (Ecarot-Charrier *et al.*, 1989; Midura *et al.*, 1990), and exists as a keratan sulfate proteoglycan in rabbit (Kinne and Fisher, 1987). BSP constitutes ~12% of the noncollagenous proteins in human and bovine bone and is synthesized by skeletal-associated cell types, including hypertrophic chondrocytes, osteoblasts, osteocytes, and osteoclasts (Bianco *et al.*, 1991; Gehron Robey *et al.*, 1992). The only extraskeletal site of BSP synthesis reported is the trophoblast, which, like osteoclasts, forms syncytia (Bianco *et al.*, 1991). The presence of a lower molecular weight product immunologically related to BSP in platelets has been ascribed to the uptake of plasma BSP, presumably released during the process of bone turnover (Chenu and Delmas, 1992). Little is known about the regulation or precise function of BSP. Evidence from *in situ* hybridization and immunolocalization studies indicates that the message is induced in specific populations of mature osteoblasts, particularly those adjacent to the mineralization front (Bianco *et al.*, 1991). This is further supported by the prominent induction of message in bovine osteoblast-like cells when induced to produce a fully mineralized matrix (Ibaraki *et al.*, 1992). The association of BSP with mineralization is not surprising since BSP binds to calcium and hydroxyapatite with extremely high affinity, presumably via its acidic amino acid clusters (Mintz *et al.*, 1991). Rat (Oldberg *et al.*, 1988a) and human (Fisher *et al.*, 1990) BSP have both been shown to contain the cell-attachment motif Arg–Gly–Asp (RGD). The ability of the full-length protein (as well as various BSP sequence-derived polypeptides containing the RGD) to enhance cell attachment of a variety of cell types (Oldberg *et al.*, 1988b; Somerman *et al.*, 1988; Helfrich *et al.*, 1992) is mediated presumably via the vitronectin receptor. The vitronectin receptor has been detected in osteoblast-like cell lines and osteoclasts but not in osteoblasts (Horton and Davies, 1989). Recently, a sequence-derived synthetic BSP polypeptide containing the RGD and flanking amino acids was shown to decrease both intracellular calcium concentration and resorptive activity of chick osteoclasts *in vitro* (Miyauchi *et al.*, 1991). Whether BSP is involved in extracellular signaling *in vivo* remains to be established. To enhance our understanding of the limited tissue expression of this conserved bone-enriched protein, we have isolated and characterized the human IBSP gene. This report describes a detailed analysis of

[1] To whom correspondence should be addressed at BRB, NIDR, NIH, Bldg. 30, Rm. 106, Bethesda, MD 20892. Telephone: (301) 496-8860. Fax: (301) 402-0824.

[2] Present address: Skeletal Diseases Research, Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285.

0888-7543/93 $5.00
Copyright © 1993 by Academic Press, Inc.
All rights of reproduction in any form reserved.

the BSP structure and its localization to chromosome 4q28–q31.

EXPERIMENTAL PROCEDURES

Cloning and sequencing. A lambda fix human genomic library, previously constructed from a partial *Sau*3A/*Mbo*I digest of DNA from a lung fibroblast cell line (WI38), was purchased from Stratagene Cloning Systems (La Jolla, CA). One million plaques, grown on LE392 or SRBP2 *Escherichia coli* lawns, were screened with a 1.0-kb *Eco*RI fragment of human IBSP cDNA (Fisher *et al.*, 1990) labeled with [α-^{32}P]dCTP by nick translation (Amersham). This cDNA fragment contained the 5′ untranslated region and entire coding region of IBSP. Positive clones were isolated, amplified according to published protocols (Sambrook *et al.*, 1989), and purified using Qiagen lambda purification kits according to the manufacturer's directions (Qiagen, Inc., Chatsworth, CA). The cloned genomic DNA was digested with various restriction enzymes for Southern blot analyses, and fragments were subcloned into M13 for single-stranded dideoxy sequencing (Sequenase; U.S. Biochemical Corp., Cleveland, OH). Plaque hybridizations using IBSP-specific, 18-mer deoxyoligonucleotides enabled isolation of clones in specific orientations for sequence analysis. Sequencing primers included M13-specific universal primers or IBSP-specific primers purchased from Synthecell Corp. (Rockville, MD) or The Midland Certified Reagent Co. (Midland, TX). DNA was sequenced in both directions to obtain promoter, exonic, and partial intronic sequence. All sequences generated were analyzed by the Sequence Analysis Software Package (Genetics Computer Group, Inc., Madison, WI).

cDNA cloning and PCR sequencing. A lambda ZAP human cDNA library (made under contract by Stratagene Cloning Systems) was constructed from poly(A)$^+$-selected RNA (Stratagene Poly(A) Quik Kit) isolated from primary cultures of normal adult human bone cells (Gehron Robey and Termine, 1985). The cDNA library was screened as described above using the 1.0-kb *Eco*RI human cDNA insert. Several positive clones were selected and isolated by repeated casting and screening. The final plaques were isolated and the cDNA in the agar plugs was allowed to diffuse into 1 ml SM buffer (50 m*M* Tris–HCl, pH 7.5, 100 m*M* sodium chloride, 8 m*M* magnesium sulfate, 0.01% gelatin) overnight at 4°C (Sambrook *et al.*, 1989). Five-microliter aliquots of the SM buffer containing the cDNA were used as templates for amplification by the polymerase chain reaction using universal primers, T7 or T3, as the 5′ primer and a reagent-grade 20-mer deoxyoligonucleotide "194" (5′-ATTAAAGCAGTCTTCATTTT-3′) as the 3′ primer, which is reverse and complementary to nucleotides 69–88 of the human IBSP cDNA (Fisher *et al.*, 1990). All buffers and reagents were used as described in the Gene-Amp kit (Perkin–Elmer–Cetus, Norwalk, CT). The amplification protocol consisted of denaturation for 1 min at 95°C, annealing for 1 min at 55°C, and elongation for 2 min at 76°C for a total of 30 cycles with a final extension for 7 min at 76°C. The PCR product was purified (Magic PCR Preps DNA Purification System; Promega, Madison, WI), and approximately 1 μg was used as a template for dideoxy PCR sequence analysis using an end-labeled sequencing protocol as described by the manufacturer (*fmol* DNA Sequencing System, Promega). Deoxyoligonucleotide 194 was end-labeled with T4 polynucleotide kinase (Promega) using [γ-^{32}P]-dATP and used as the primer for the sequencing reactions. The extension protocol included an initial denaturation at 95°C for 2 min followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 42°C for 30 s, and extension in the presence of the appropriate dideoxy reagents at 70°C for 1 min. The products were electrophoresed through 6 and 8% denaturing gels and the gels were subjected to autoradiography at room temperature for 24 h.

Primer extension analysis. To determine the transcriptional start point of the human IBSP gene, a primer extension analysis was performed using a method described by Sambrook *et al.* (1989) and deoxyoligonucleotide 194 (end-labeled with [γ-^{32}P]dATP by T4 polynucleotide kinase) as the primer for reverse transcription (AMV reverse transcriptase, Promega). Briefly, 0.5 or 5 μg of poly(A)$^+$-selected RNA (Poly(A) Quik Kit, Stratagene) isolated from primary cultures of normal adult human bone cells or 10 μg of yeast tRNA (as a control) was mixed with 10^5 cpm of the labeled primer, heated to 85°C for 10 min to denature the DNA, and allowed to anneal overnight at 42°C. The mixture was precipitated with ethanol, redissolved in 20 μl of reverse transcriptase buffer (50 m*M* Tris–HCl, pH 7.6, 60 m*M* KCl, 10 m*M* $MgCl_2$, 1 m*M* of each dNTP, 1 m*M* dithiothreitol, 1 unit/μl placental RNase inhibitor, 50 μg/ml actinomycin D), and incubated in 50 units of AMV reverse transcriptase at 37°C for 2 h followed by a second incubation in the presence of 1 μl of 0.5 *M* EDTA (pH 8.0) and 1 μl of DNase-free pancreatic RNase (5 μg/ml) at 37°C for 30 min. The reaction mixture was extracted with phenol/chloroform, precipitated with ethanol, and resuspended in 4 μl of TE (10 m*M* Tris, 0.1 m*M* EDTA) with the addition of 4 μl of stop buffer (Sequenase kit, U.S. Biochemical). Four microliters of the resultant nucleic acid hybrid mix was electrophoresed through an 8% polyacrylamide/7 *M* urea gel. [^{35}S]-dATP-labeled M13 (Sequenase kit, U.S. Biochemical) was used as a marker for nucleotide length.

Chromosomal localization. Human lymphocyte cultures were incubated for 6 h with 20 μg/ml bromodeoxyuridine prior to the addition of colcemid. The chromosomes were prepared according to standard techniques for *in situ* hybridization analysis (Adolph *et al.*, 1987). Briefly, the slides were treated with 70% formamide, 2× SSC (1× SSC is 0.15 *M* sodium chloride, 0.015 *M* sodium citrate, pH 7.4) at 70°C for 2 min to denature the DNA. The slides were then hybridized with a ^{3}H-labeled (nick translation kit, Amersham Corp.), 1-kb human IBSP cDNA fragment in 20 μl of 50% formamide, 10% dextran sulfate, 2× SSC, 100 μg/ml salmon sperm DNA for 16 h at 38°C in a humid atmosphere. The slides were washed three times, 15 min each, with 50% formamide, 2× SSC at 40°C, followed by six washes, 2 min each, with 2× SSC at 40°C with a final wash for several hours in 2× SSC at room temperature. The slides were dipped in Ilford L4 emulsion and developed after 21 days, and the chromosomes were stained as described (Epplen *et al.*, 1975).

RESULTS

Isolation and Sequencing of Genomic Clones Encoding Human IBSP. A lambda fix genomic library made from human lung fibroblast DNA was probed with a 1.0-kb insert of human IBSP cDNA (Fisher *et al.*, 1990). The first clone analyzed, genomic clone 13, contained an insert of ~9.5 kb. Characterization of genomic clone 13 by sequencing and Southern blot analyses indicated that the clone spanned from within an intron preceding the transcriptional start site through the 3′ untranslated region of IBSP. Subsequent rescreening of the library with synthetic 18-mer deoxyoligonucleotides (corresponding to nucleotides 1–18 and 19–36 of the published cDNA sequence) to isolate a genomic clone containing the 5′ untranslated region of IBSP repeatedly failed to yield positive clones. To assess the integrity of the extreme 5′ end sequence of the original IBSP cDNA that had been isolated from a previously made library (Fisher *et al.*, 1990), we proceeded to construct and screen a new normal adult human bone cell cDNA library. Several positive clones were isolated and sequenced by PCR. The results of these sequences indicated that the 5′ sequence obtained from the new cDNA library differed from the 5′ end previously reported (Fisher *et al.*, 1990). We suspect that in the earlier preparation, an artifact may have occurred during cloning to create a nonsense sequence at the 5′ end.

Based on the revised sequence, synthetic deoxyoligonucleotides were generated and used to rescreen the orig-

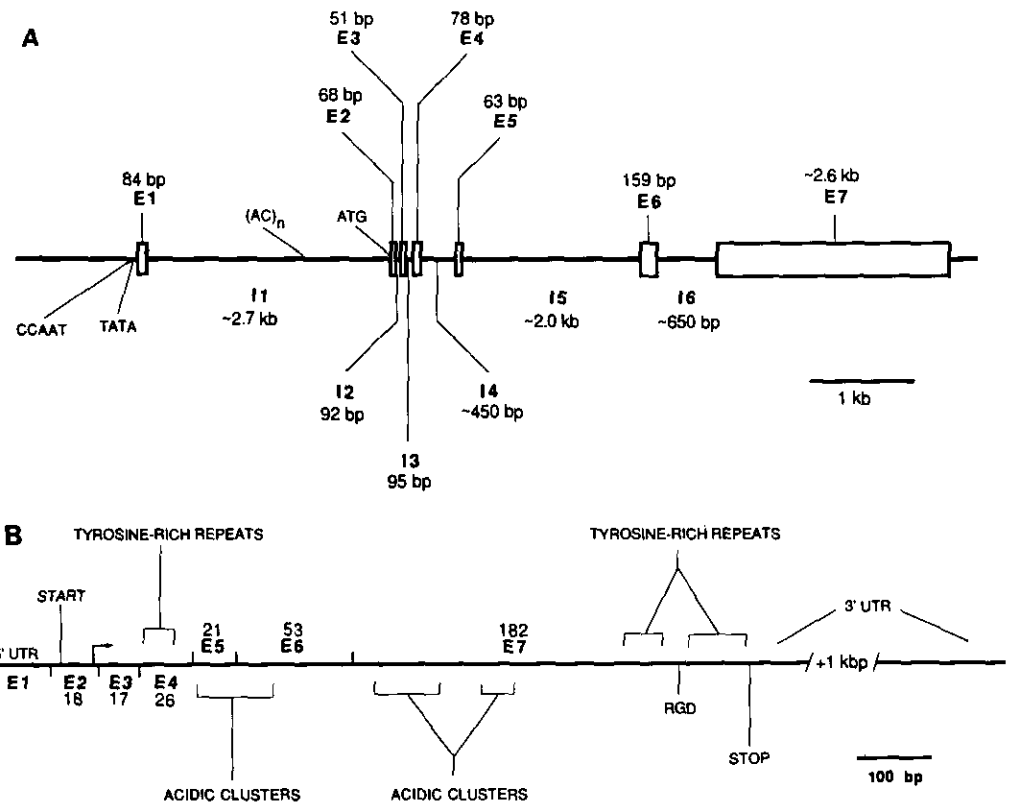

FIG. 1. Schematic of the genomic organization of human IBSP. Relative exon (E) positions are shown in rectangles with the exon number and size identified. Introns (I) are identified by solid vertical lines and their respective sizes are indicated below each line. Intron sizes estimated by PCR analyses using the genomic clones as templates and exonic primers that flanked the introns are designated by an approximation sign. The TATAA and CCAAT consensus sequences are indicated in the promoter region of the gene (**A**). (**B**) Diagram outlining the domains of the deduced amino acid sequence of human BSP. The number of amino acids coded by each exon (E) is denoted by arabic numerals. The 5′ and 3′ untranslated regions (UTR), the start and stop codons, and the cell-binding Arg–Gly–Asp (RGD) polypeptide are shown. The arrow indicates the start of the mature secreted protein and the tyrosine repeats and acidic clusters are located within the bracketed regions.

inal human genomic library. From this screen, clone 19 was isolated and found to contain the 5′ upstream region of the gene and overlapped clone 13 at least through exon 5. Figure 1 shows the genomic organization of IBSP. As indicated in Fig. 1A, IBSP contains six small exons (E1 through E6) ranging in size from 51 to 159 bp and one large exon (E7) containing ~2.6 kb. The size estimate of exon 7 was based on the largest cDNA clone isolated from a human bone cell cDNA library (Fisher *et al.*, 1990). It should be noted that IBSP has two message sizes, a major message size of 2.0 kb and a minor message size of 3.0 kb (Fisher *et al.*, 1990). This makes the last exon (E7) variable, presumably from use of alternative poly(A)$^+$ attachment sites. The six introns range in size from 92 bp to ~2.7 kb.

Figure 1B is a diagram outlining some of the structural features of IBSP and their relationship to the gene. Less than half of the amino acids (135/317 amino acids) are encoded within exons 2 through 6, whereas exon 7 encodes the remainder of the protein (182 amino acids) as well as the 3′ untranslated region. Exon 1 is composed exclusively of untranslated sequence. Exon 2 contains some untranslated sequence, the start methionine and hydrophobic leader sequence, and the first 2 amino acids of the mature protein, as determined by microsequencing (Fisher *et al.*, 1987). The protein contains three tyrosine-rich repeats; the first is located in exon 4 and the remaining two flank the cell-attachment Arg–Gly–Asp (RGD) peptide at the 3′ end of the coding region of exon 7. There are also three acidic clusters in BSP; the first cluster spans exon 5 through the 5′ end of exon 6 and the remaining two clusters are located in the first half of exon 7. There are four potential sites for Asn-linked glycosylation (Asn–X–Ser/Thr), the first of which is located in exon 6 (amino acid 104) with the remainder located in exon 7 (amino acids 177, 182, and 190).

Several discrepancies were noted between the previously published cDNA sequence and the sequences generated from the newly derived cDNA and genomic clones. PCR sequencing of IBSP cDNA clones indicated a difference in 12 of the first 20 nucleotides in exon 1 (denoted by underlines in Fig. 2). Subsequent sequencing data generated from genomic clone 19 confirmed this revised sequence. Three other base revisions were found in exon 7. These changes were found in the coding triplets for amino acids 195 (GGA to GTA), 268 (GTC to GCC), and 270 (GAA to GAT), resulting in changes in the deduced amino acids of Gly to Val (amino acid 195), Val to Ala (amino acid 268), and Glu to Asp (amino acid 270).

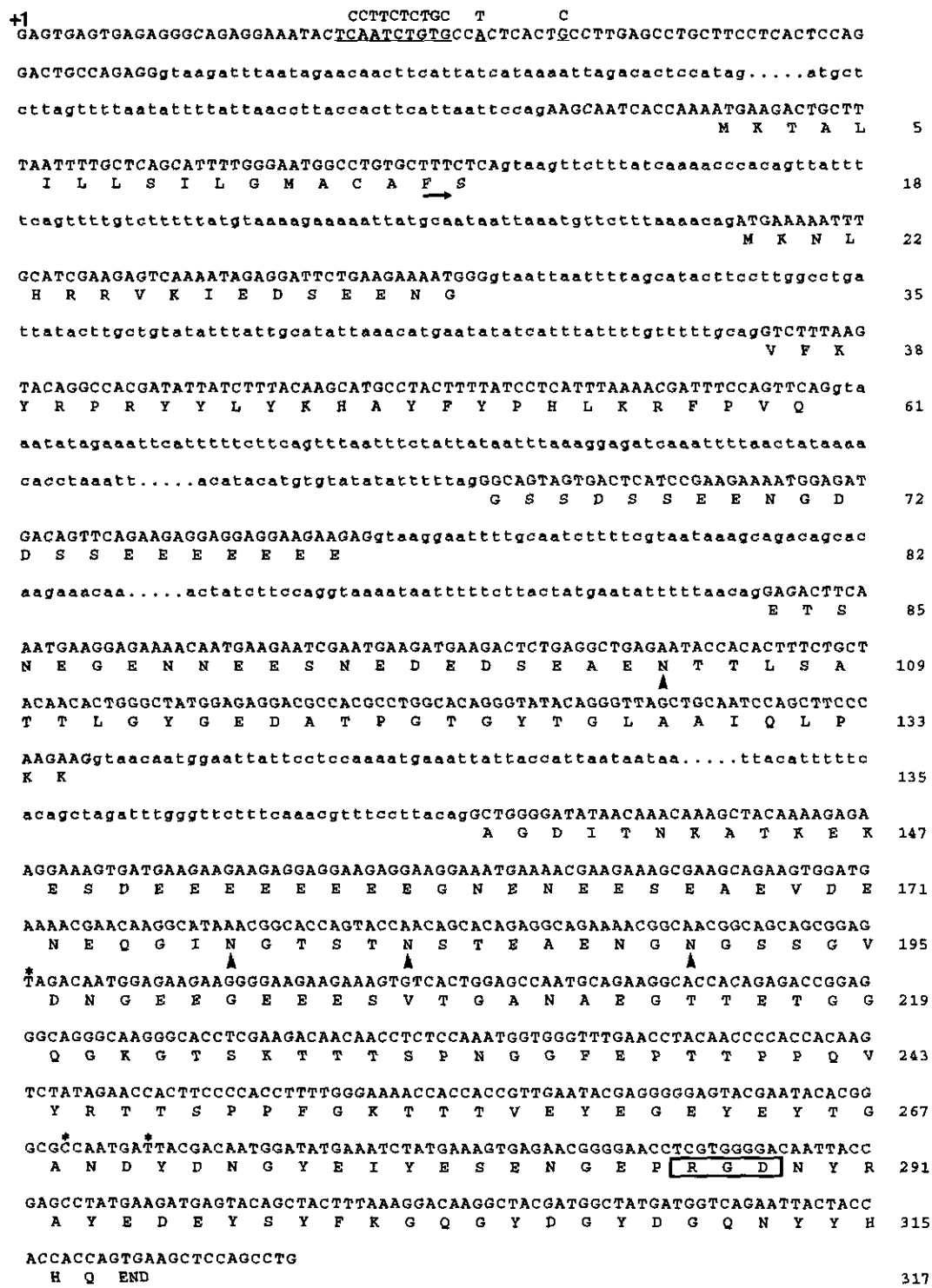

FIG. 2. The exonic, partial intronic, and deduced amino acid sequences of IBSP. The entire sequences for introns 2 and 3 are shown, whereas only partial sequences for introns 1, 4, 5, and 6 (indicated by the ellipses) are included. The remaining intervening sequences generated by this lab can be accessed through GenBank (Accession Nos. L09554, which includes the 5′ upstream, exon 1, and the 5′ end of intron 1 sequences; L09555, which includes the 3′ end of intron 1, exons 2–4, introns 2 and 3, and 5′ end of intron 4 sequences; L09556, which includes the 3′ end of intron 4 and exon 5 and the 5′ end of intron 5 sequences; L09557, which includes the 3′ end of intron 5, exon 6, and the 5′ end of intron 6 sequences; and L09558, which includes the 3′ end of intron 6 and exon 7). The +1 indicates the start of exon 1 as determined by primer extension analysis. The underlined sequence represents the revised human IBSP sequence compared to the 5′ end sequence (denoted above the genomic sequence) of the published cDNA (Fisher *et al.*, 1990). The exonic sequence is shown in uppercase and the intronic sequence in lowercase letters. The horizontal arrow indicates the beginning of the mature, secreted protein and the arrowheads designate potential sites for Asn-linked glycosylation. The asterisks indicate the three base changes in the coding region of the gene compared to the published cDNA, and the cell-attachment polypeptide (RGD) is boxed.

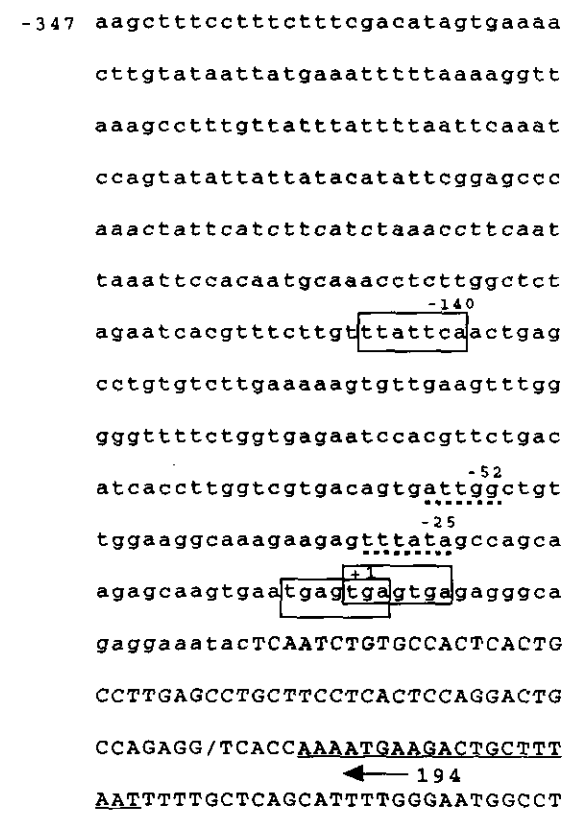

FIG. 3. Primer extension analysis and upstream sequence of human IBSP. Approximately 5 or 0.5 μg of pA^+-selected RNA obtained from primary cultures of adult human bone cells or 5 μg of tRNA (as control) was reverse-transcribed with a 20-mer deoxyoligonucleotide 194. The position of the primer 194 is underlined in exon 2 (the boundary between exons 1 and 2 is designated by a slash). From this analysis, it was determined that the transcriptional start point is located 23 nucleotides from the beginning of the cDNA and is denoted as +1. Exonic sequences are shown in uppercase and 347 nucleotides of upstream sequence are shown in lowercase letters. The CCAAT and TATAA consensus sequences are indicated by a dashed line and the three potential AP-1 sites are boxed.

Structural features of exon 1, intron 1, and the promoter. To determine the transcriptional start point, a primer extension analysis was performed using mRNA obtained from primary cultures of normal adult human bone cells and the primer deoxyoligonucleotide 194, which hybridizes to exon 2 (Fig. 3). From this analysis, exon 1 was found to contain an additional 23 nucleotides from the 5′ end of both the original and the newly derived cDNAs (data not shown). Sequence analysis of the upstream region indicated the presence of TATAA and CCAAT box consensus sequences at nucleotide positions −25 through −30 and −52 through −56, respectively; both motifs were in the reverse orientation (Fig. 3). Three potential AP-1 consensus sequences were observed in the upstream region of the gene. Two of the three motifs (TGAGTGA) overlap and were found at nucleotides −5 to +2 and nucleotides −1 to +6. The third motif (TTATTCA) was located further upstream at nucleotides −140 to −146 (Fig. 3). Intron 1 contained a glucocorticoid response element consensus sequence (AGAWCAGW) and repetitive stretches of pyrimidines, $TC(CT)_3GT(CT)_3GT(CT)_2GT(CT)_2C$, followed by a polypurine–pyrimidine tract, $CC(ACCC)_2AT(AC)_{16}GC$-$ATAT(AC)_8CAA$, approximately 800 nucleotides upstream of exon 2 (data not shown).

Intron/exon boundaries. The intron/exon boundaries of human IBSP are shown in Table 1. All junctions follow the donor and acceptor splice site consensus sequences (Mount, 1982). A striking feature of the five introns that interrupt the coding region is that they are all phase 0; that is, the intron lies between two codons. Because of the position of the flanking introns, the coding exons are considered to be symmetrical and are class 0-0.

Chromosomal localization. A battery of hamster–human cell hybrids was used previously to determine that the IBSP gene is located on chromosome 4 (Fisher *et al.*, 1990). To define more precisely the chromosomal location of IBSP, an *in situ* hybridization of ^3H-labeled human IBSP cDNA to human metaphase chromosomes was performed. For this experiment, a total of 114 metaphase spreads containing 249 silver grains (an average of 2.2 grains per metaphase) were analyzed. Only those metaphase spreads containing a grain on chromosome 4 or 5 were selected for further analysis. Figure 4 shows the relative distribution of grains along the entire chromosome. Of the 249 grains detected, 94 grains (37.8% of total) were distributed on chromosome 4 and 51% of these grains (48/94) were clustered over 4q28–q31.

DISCUSSION

In this report, we have described the isolation and characterization of the human IBSP gene. This gene contains seven exons, the first of which is entirely noncoding. Designation of the start of transcription (+1) was based on primer extension analysis, which indicated that exon 1 contains an additional 23 nucleotides upstream from the beginning of any of the isolated cDNAs. Further upstream sequencing data revealed TATAA and CCAAT boxes (both in the reverse orientation) and AP-1 motifs, all of which have been previously shown to be important in the transcriptional regulation of a number of genes, including two other bone-related genes, murine Spp-1 (osteopontin; Miyazaki *et al.*, 1990; Craig and Denhardt, 1991) and human osteocalcin (Celeste *et al.*, 1986). These three genes are considered to be "late stage" markers of osteoblasts in that induction of these genes is associated with the maturation of osteoblasts *in vivo* (Gehron Robey *et al.*, 1992) and *in vitro* (Stein *et al.* 1990). IBSP also contains, in common with a number of bone matrix genes (osteocalcin, Celeste *et al.*, 1986; murine Spp-1, Craig and Denhardt, 1991; osteonectin, Young *et al.*, 1989; Ibaraki *et al.*, 1992; human collagen, D'Alessio *et al.*, 1988), a repetitive homopurine/homopyrimidine sequence $(CT)_n$. Similar homopurine/homopyrimidine sequences have been shown to be sensitive to S1 nuclease digestion and are considered to adopt a hinged or H-DNA configuration (Bianchi *et al.*, 1990).

TABLE 1

Intron/Exon Boundaries of Human IBSP[a]

Donor sequence AGgt(a/g)agt*	Intron	Acceptor sequence 6(t/c)ncagN*	Intron phase	Exon class
TGCCAGAGGgtaaga	1	ttaattccagAAGCAATCA		
GCTTTCTCAgtaagt	2	tttaaaacagATGAAAAAT	0	0–0
AlaPheSer		MetLysAsn		
GAAAATGGGgtaatt	3	gtttttgcagGTCTTTAAG	0	0–0
GluAsnGly		ValPheLys		
CCAGTTCAGgtaaat	4	atatttttagGGCAGTAGT	0	0–0
ProValGln		GlySerSer		
GAAGAAGAGgtaagg	5	tttttaacagGAGACTTCA	0	0–0
GluGluGlu		GluThrSer		
CCCAAGAAGgtaaca	6	ttccttacagGCTGGGATA	0	0–0
ProLysLys		AlaGlyAsp		

[a] The donor and acceptor sequences follow the gt/ag splice-site rule (Mount, 1982). These consensus sequences are designated with an asterisk (*). Introns are classified as phase 0 since the codons are uninterrupted by the intervening sequences. This configuration results in symmetrical exons which are designated as class 0–0.

Spp-1, osteonectin, and collagen, however, are produced by many nonskeletal tissues. We suspect, therefore, that while this homopurine/homopyrimidine repetitive structure is potentially important for transcription, it is unlikely to contribute to cell specificity or restricted expression in bone. One striking difference between the IBSP gene and the genes for other bone proteins is the presence of a poly(AC)$_n$ tract in intron 1 of IBSP. This sequence appears throughout the genome in numerous species. *In vitro* studies of poly(AC)$_n$ tracts indicate that this sequence can potentially adopt a left-handed or Z-DNA conformation (Haniford and Pulleyblank, 1983; Mclean and Wells, 1988). This element is potentially important in the regulation of IBSP since a similar element located in the promoter region of the rat prolactin gene was shown to inhibit transcription (Naylor and Clark, 1990). BSP has been shown to be induced by dexamethasone and suppressed by 1,25-dihydroxyvitamin D_3 in the rat osteosarcoma line ROS 17/2.8 and in cultured neonatal rat calvaria (Oldberg *et al.*, 1989). Although no vitamin D response element has yet been identified in the available genomic sequence, a glucocorticoid response element consensus sequence has been identified in intron 1.

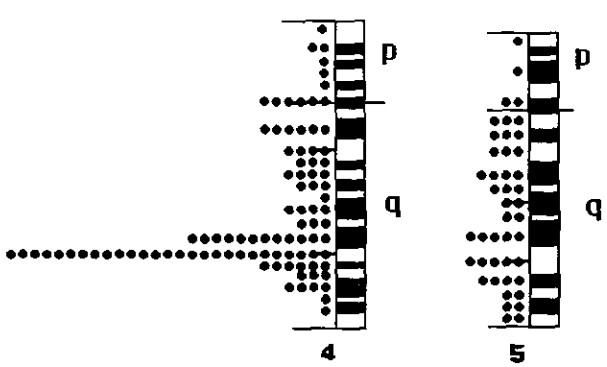

FIG. 4. Subchromosomal localization of the human IBSP gene. The schematic shows the distribution of grains along chromosome 4 with the majority of grains located between 4q28–q31. Chromosome 5 is included for comparison.

The biochemical data and deduced amino acid sequence of IBSP indicate that BSP is an acidic protein that undergoes extensive post-translational modifications (Fisher *et al.*, 1983, 1990). Based on data of the genomic organization of IBSP, limitation of the various structural features of the protein to specific domains is not apparent. Of the three acidic clusters identified in the IBSP cDNA, the first spans exon 5 through the 5′ region of exon 6. The remaining two clusters are located in the first third of exon 7. Approximately 50% of the mass of BSP has been attributed to carbohydrate (Fisher *et al.*, 1983), and four potential glycosylation sites have been identified in the gene (Fisher *et al.*, 1990). Exon 6 contains a potential site for Asn-linked glycosylation; three other motifs of this nature are also found in the first third of exon 7 and this is consistent with the composition of the protein described by Fisher *et al.* (1983). The first of three tyrosine repeats identified in IBSP (Fisher *et al.*, 1983) is within exon 4 and does not follow the consensus for tyrosine sulfation; the remaining two repeats that flank the RGD peptide in exon 7 have the potential for undergoing sulfation (Fisher *et al.*, 1990). Sulfated BSP is synthesized in rat (Midura *et al.*, 1990) and mouse (Ecarot-Charrier *et al.*, 1989) osteoblast-like cell cultures.

Several differences between the genomic sequence and the published cDNA sequence have been noted. The reason for the multiple revisions of bases reported in exon 1 is not clear. However, the repeated difficulty in PCR sequencing of newly derived cDNA and genomic clones suggests secondary structures in this region that may interfere with the cloning and/or subsequent sequencing of double-stranded DNA. A comparison of the newly derived cDNA 5′ sequence to that reported by Oldberg *et al.* (1988) in the rat indicates a greater homology in this region of the gene than previously thought. However, three base revisions in the coding region of the human IBSP gene decrease the similarity between rat and human. These three base revisions, detected in exon 7, are due to compressions (GGA to GTA, aa 195, and

GTC to GCC, aa 268) in the original gels and a data entry error (GAA to GAT, aa 270).

The intron/exon junctions of IBSP are unusual in that the coding triplet is not interrupted by introns. This type of configuration could potentially allow exon shuffling to occur without affecting the downstream open reading frame. The possibility of a deletion of exon 2, the location of the start methionine and hydrophobic leader sequence, would be intriguing since a second methionine is located at the start of exon 3. Since the subsequent sequence is not hydrophobic, use of this methionine could result in the compartmentalization of the BSP molecule in the cytosol rather than the extracellular matrix. There are several lines of evidence indicating that this second potential start site is not utilized, including (1) no reports of spliced mRNA variants, (2) immunolocalization of BSP in the Golgi and extracellular matrix only (Bianco *et al.*, 1991), and (3) nonadherence to the Kozak consensus sequence gcca/gccAUGG, where the −3 intronic underlined nucleotide (A is +1) is a purine 97% of the time and the +4 exonic underlined nucleotide is a guanine 46% of the time (Kozak, 1984) as observed in the IBSP sequence aaacagAUGA (Fig. 2 and Table 1).

From somatic cell hybridization studies (Fisher *et al.*, 1990), human IBSP had been previously assigned to chromosome 4. Based on *in situ* hybridization of metaphase spreads, the localization has been more precisely defined to the long arm of chromosome 4 in the region between q28 and q31. Human Spp-1 is also located on the long arm of chromosome 4, although this gene is found in the region 4q13–q21 (Young *et al.*, 1990). There are few reports of chromosomal abnormalities on the long arm of chromosome 4 (4q). The most common aberration appears to arise from terminal deletions (4q31–qter) and results in clinical manifestations, including developmental, neurological, craniofacial, musculoskeletal, and cardiac defects (Mitchell *et al.*, 1981; Lin *et al.*, 1988). It should be noted that an interstitial deletion at 4q27–4q31.3, which encompasses the region to which IBSP is localized (4q28–q31), results in a phenotype of growth retardation and craniofacial abnormalities (Serville and Broustet, 1977; Mitchell *et al.*, 1981) similar to that observed with the terminal deletion.

In summary, we have isolated and partially sequenced the human IBSP gene. BSP is a conserved protein with an expression restricted predominantly to the skeleton. The isolation and characterization of the human IBSP gene will now enable us to dissect out important *cis*- and *trans*- acting factors that regulate the transient expression and tissue specificity of this protein.

REFERENCES

Adolph, S., Bartram, C. R., and Hameister, H. (1987). Mapping of the oncogenes *myc, sis,* and *int-1* to the distal part of mouse chromosome 15. *Cytogenet. Cell Genet.* **44:** 65–68.

Bianchi, A., Wells, R. D., Heintz, N. G., and Caddle, M. S. (1990). Sequences near the origin of replication of the DHFR locus of Chinese hamster ovary cells adopt left-handed Z-DNA and triplex structures. *J. Biol. Chem.* **265:** 21789–21796.

Bianco, P., Fisher, L. W., Young, M. F., Termine, J. D., and Gehron-Robey, P. (1991). Expression of bone sialoprotein (BSP) in developing tissues. *Calcif. Tissue Int.* **49:** 421–426.

Celeste, A. J., Rosen, V., Buecker, J. L., Kriz, R., Wang, E. A., and Wozney, J. M. (1986). Isolation of the human gene for bone gla protein utilizing mouse and rat cDNA clones. *EMBO J.* **5:** 1885–1890.

Chen, J., Shapiro, H. S., Wrana, J. L., Reimers, S., Heersche, J. N. M., and Sodek, J. (1991). Localization of bone sialoprotein (BSP) expression to sites of mineralized tissue formation in fetal rat tissues by *in situ* hybridization. *Matrix* **11:** 133–143.

Chenu, C., and Delmas, P. D. (1992). Platelets contribute to circulating levels of bone sialoprotein in human. *J. Bone Min. Res.* **7:** 47–54.

Craig, A. M., and Denhardt, D. T. (1991). The murine gene encoding secreted phosphoprotein 1 (osteopontin): Promoter structure, activity, and induction in vivo by estrogen and progesterone. *Gene* **100:** 163–171.

D'Alessio, M., Bernard, M., Pretorious, P. J., de Wet, W., and Ramirez, F. (1988). Complete nucleotide sequence of the region encompassing the first twenty-five exons of the human pro 1(1) collagen gene (COL1A1). *Gene* **67:** 105–115.

Ecarot-Charrier, B., Bouchard, F., and Delloye, C. (1989). Bone sialoprotein II synthesized by cultured osteoblasts contains tyrosine sulfate. *J. Biol. Chem.* **264:** 20049–20053.

Epplen, J. T., Siebers, J. W., and Vogel, W. (1975). DNA replication patterns of human chromosomes from fibroblasts and amniotic fluid cells revealed by a Giemsa staining technique. *Cytogenet. Cell Genet.* **15:** 177–185.

Fisher, L. W., Whitson, S. W., Avioli, L. V., and Termine, J. D. (1983). Matrix sialoprotein of developing bone. *J. Biol. Chem.* **258:** 12723–12727.

Fisher, L. W., Hawkins, G. R., Tuross, N., and Termine, J. D. (1987). Purification and partial characterization of small proteoglycans I and II, bone sialoproteins I and II and osteonectin from the mineral compartment of developing human bone. *J. Biol. Chem.* **262:** 9702–9708.

Fisher, L. W., McBride, O. W., Termine, J. D., and Young, M. F. (1990). Human bone sialoprotein. Deduced sequence and chromosomal localization. *J. Biol. Chem.* **265:** 2347–2351.

Gehron Robey, P., and Termine, J. D. (1985). Human bone cells *in vitro*. *Calcif. Tissue Int.* **37:** 453–460.

Gehron Robey, P., Bianco, P., and Termine, J. D. (1992). The cellular biology and molecular biochemistry of bone formation. *In* "Disorders of Bone and Mineral Metabolism," pp. 241–263, Raven Press, New York.

Haniford, D. B., and Pulleyblank, D. E. (1983). Facile transition of poly[d(TG) · d(CA)] into a left-handed helix in physiological conditions. *Nature* **302:** 632–634.

Helfrich, M. H., Nesbitt, S. A., Dorey, E. L., and Horton, M. A. (1992). Rat osteoclasts adhere to a wide range of RGD (Arg–Gly–Asp) peptide-containing proteins, including the bone sialoproteins and fibronectin via a $\beta 3$ integrin. *J. Bone Min. Res.* **7:** 335–343.

Horton, M. A., and Davies, J. (1989). Perspectives: Adhesion receptors in bone. *J. Bone Min. Res.* **4:** 803–808.

Ibaraki, K., Termine, J. D., Whitson, S. W., and Young, M. F. (1992). Bone matrix mRNA expression in differentiating fetal bovine osteoblasts. *J. Bone Min. Res.* **7:** 743–754.

Kasugai, S., Todescan R., Jr., Nagata, T., Yao, K.-L., Butler, W. T., and Sodek, J. (1991). Expression of bone matrix proteins associated with mineralized tissue formation by adult rat bone marrow cells in vitro: Inductive effects of dexamethasone on the osteoblastic phenotype. *J. Cell. Physiol.* **147:** 111–120.

Kinne, R. W., and Fisher, L. W. (1987). Keratan sulfate proteoglycan in rabbit compact bone is bone sialoprotein II. *J. Biol. Chem.* **262:** 10206–10211.

Kozak, M. (1984). An analysis of vertebrate mRNA sequences: Intimations of translational control. *Nucleic Acids Res.* **12:** 857–872.

Lin, A. E., Garver, K. L., Diggans, G., Clemens, M., Wenger, S. L., Steele, M. W., Jones, M. C., and Israel, J. (1988). Interstitial and terminal deletions of the long arm of chromosome 4: Further delineation of phenotypes. *Am. J. Med. Genet.* **31:** 533–548.

Mclean, M. J., and Wells, R. D. (1988). The role of DNA sequence in the formation of Z-DNA *versus* cruciforms in plasmids. *J. Biol. Chem.* **263:** 7370–7377.

Midura, R. J., McQuillan, D. J., Benham, K. J., Fisher, L. W., and Hascall, V. C. (1990). A rat osteogenic cell line (UMR 106-01) synthesizes a highly sulfated form of bone sialoprotein. *J. Biol. Chem.* **265:** 5285–5291.

Mintz, K., Eanes, E. D., Midura, R. J., Bryan, P., Lewis, M. S., and Fisher, L. W. (1991). Physiochemical analysis of native bone sialoprotein. *J. Bone Min. Res.* **6:** S158.

Mitchell, J. A., Packman, S., Loughman, W. D., Fineman, R. M., Zackai, E., Patil, S. R., Emanuel, B., Bartley, J. A., and Hanson, J. W. (1981). Deletions of different segments of the long arm of chromosome 4. *Am. J. Med. Genet.* **8:** 73–89.

Miyauchi, A., Alvarez, J., Greenfield, E. M., Teti, A., Grano, M., Colucci, S., Zambonin-Zallone, A., Ross, F. P., Teitelbaum, S. L., Cheresh, D., and Hruska, K. A. (1991). Recognition of osteopontin and related peptides by an α v β3 integrin stimulates immediate cell signals in osteoclasts. *J. Biol. Chem.* **266:** 20369–20374.

Miyazaki, Y., Setoguchi, M., Yoshida, S., Higuchi, Y., Akizuki, S., and Yamamoto, S. (1990). The mouse osteopontin gene: Expression in monocytic lineages and complete nucleotide sequence. *J. Biol. Chem.* **265:** 14432–14438.

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* **10:** 459–472.

Nagata, T., Bellows, C. G., Kasugai, S., Butler, W. T., and Sodek, J. (1991). Biosynthesis of bone proteins [SPP-1 (secreted phosphoprotein-1, osteopontin), BSP (bone sialoprotein) and SPARC (osteonectin)] in association with mineralized-tissue formation by fetal-rat calvarial cells in culture. *Biochem. J.* **274:** 513–520.

Nagata, T., Goldberg, H. A., Zhang, Q., Domenicucci, C., and Sodek, J. (1991). Biosynthesis of bone proteins by fetal porcine calvariae *in vitro*. Rapid association of sulfated sialoproteins (secreted phosphoprotein-1 and bone sialoprotein) and chondroitin sulfate proteoglycan (CS-PGIII) with bone mineral. *Matrix* **11:** 86–100.

Naylor, L. H., and Clark, E. M. (1990). d(TG)n · d(CA)n sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription. *Nucleic Acids Res.* **18:** 1595–1601.

Oldberg, A., Franzen, A., and Heinegard, D. (1988a). The primary structure of a cell-binding bone sialoprotein. *J. Biol. Chem.* **263:** 19430–19432.

Oldberg, A., Franzen, A., Heinegard, D., Pierschbacher, M., and Ruoslahti, E. (1988b). Identification of a bone sialoprotein receptor in osteosarcoma cells. *J. Biol. Chem.* **263:** 19433–19436.

Oldberg, A., Jirskog-Hed, B., Axelsson, S., and Heinegard, D. (1989). Regulation of bone sialoprotein mRNA by steroid hormones. *J. Cell. Biol.* **109:** 3183–3186.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Serville, F., and Broustet, A. (1977). Pericentric inversion and partial monosomy 4q associated with congenital anomalies. *Hum. Genet.* **39:** 239–242.

Somerman, M. J., Fisher, L. W., Foster, R. A., and Sauk, J. J. (1988). Human bone sialoprotein I and II enhance fibroblast attachment *in vitro*. *Calcif. Tissue Int.* **43:** 50–53.

Stein, G. S., Lian, J. B., and Owen, T. A. (1990). Relationship of cell growth to the regulation of tissue-specific gene expression during osteoblast differentiation. *FASEB J.* **4:** 3111–3123.

Young, M. F., Findlay, D. M., Dominguez, P., Burbelo, P. D., MaQuilla, C., Kopp, J. B., Robey, P. G., and Termine, J. D. (1989). Osteonectin promoter. DNA sequence analysis and S1 endonuclease site potentially associated with transcriptional control in bone cells. *J. Biol. Chem.* **264:** 450–456.

Young, M. F., Kerr, J. M., Termine, J. D., Wewer, U. M., Ge Wang, M., McBride, O. W., and Fisher, L. W. (1990). cDNA cloning, mRNA distribution and heterogeneity, chromosomal location, and RFLP analysis of human osteopontin (OPN). *Genomics* **7:** 491–502.