A cDNA encoding the human bone sialoprotein was isolated from a λ Zap expression library (made from cultured human bone cell poly(A) RNA) using radiolabeled rat bone sialoprotein cDNA (Oldberg, A., and Heinegard, D. (1988) J. Biol. Chem. 263, 19430–19432) as a probe. A 5'-1-kilobase EcoRI fragment of the purified 3-kilobase clone was sequenced and found to contain the entire protein-encoding region. The deduced protein sequence revealed a 317-amino acid protein (34,982 Da) containing a 16-amino acid hydrophobic signal sequence and a 33,352 Da protein destined to undergo extensive post-translational modifications before being secreted from the cell. A comparison of the human and rat protein sequences showed extensive (>70%) amino acid identities including the Arg-Gly-Asp (RGD) tripeptide thought to confer the cell attachment activity observed previously for this protein. Also conserved were three regions rich in acidic amino acids and three regions rich in tyrosine. While all three tyrosine-rich regions appear to be composed of a nominal repeat structure, only the two carboxyl-terminal regions that flank the RGD sequence fit all three of the requirements for extensive tyrosine sulfation. Interestingly, human bone sialoprotein, whose final secreted product is approximately 50% carbohydrate, contains no cystines. Northern analysis showed that while bone cells are the major source of bone sialoprotein message production, other tissues may contain trace amounts of this message. Southern hybridization of DNA from human-rodent somatic cell hybrids that have segregated human chromosomes indicated that the gene is located on human chromosome 4. The human bone sialoprotein gene is a single copy gene unlikely to exceed 11.1 kilobases in length. No restriction fragment length polymorphisms were observed with 12 different restriction enzymes in 10 normal individuals.

The organic matrix of bone and dentin, like that of most mammalian connective tissues, is made predominantly of type I collagen. For almost 30 years researchers have tried to find bone-derived protein to be given close biochemical scrutiny in this regard was a 23,000-Da, sialic acid-rich protein, bone sialoprotein (BSP) (1). This small glycoprotein was later hypothesized to be a breakdown product of a M, 75,000 sialoprotein containing 12% sialic acid, 7% glucosamine, 6% galactosamine, and ~50% protein (2). Competition enzyme-linked immunosorbert assays showed that BSP constituted about 12% of the noncollagenous proteins of bone and <1% of dentin but was not detectable in a variety of other tissues at an estimated sensitivity of 1/1000th that found in bone (2,3). The BSP (also called BSP II for a short period of time to distinguish it from BSP I until the latter was renamed osteopontin) has also been shown to be a keratan sulfate proteoglycan in rabbit bone (4). Recently, Oldberg and co-workers (5) have determined the primary structure of rat BSP from a cDNA clone (5). Because the deduced sequence contained the cell attachment recognition Arg-Gly-Asp tripeptide (6) generally associated with the integrin-type cell-binding receptors, Oldberg and co-workers looked for and identified a BSP receptor (identical to the vitronectin receptor) in rat osteosarcoma cells (7). Independently, Somerman et al. (8) showed enhanced cell attachment of human fibroblasts in culture using purified BSP (8). Using the rat cDNA as a probe, we have isolated and sequenced a human BSP cDNA and have localized the gene to chromosome 4.

**MATERIALS AND METHODS**

**Screening of a Human Bone-derived Cell Culture cDNA Library**—The λ gt11 Zap library construction, Escherichia coli infection, and growth on agar plates were described earlier (9). Initial screening of 200,000 recombinants with 32P-labeled (nick-translation kit, American Corp.) rat bone sialoprotein cDNA (a 1.5-kb EcoRI fragment, Ref. 5) resulted in many positive plaques. Prehybridization, intermediate stringency hybridization (at 60 °C, rather than the high stringency 68 °C), washing and detection on Kodak X-AR film were as described (10). Five positive colonies were purified and the phage with the largest insert (~3.0 kb, clone B6) was converted into its plasmid form for large scale preparation (9). The rat bone sialoprotein cDNA was a generous gift of Dr. A. Oldberg, Department of Physiological Chemistry, University of Lund, Lund, Sweden.

**DNA Sequencing**—The cDNA insert was released from the plasmid as two fragments (>0.6 and >0.9 kb) by EcoRI restriction enzyme digestion. Agarose gel electrophoresis-purified fragments were subcloned into the EcoRI site of M13mp19 sequencing vectors (11). After transformation into JM101 cells, single stranded DNA from recombinant phage (one member of each strand) was annealed to either the 17-base pair universal primer or the appropriate synthetic oligonucleotide of an internal site and sequenced using the dideoxynucleotide chain termination method (12). The Sequenase kit (United States Biochemicals) using both GTP and TTP nucleotide mixtures and [a-32P]ATP (1110 Ci/mmol, Du Pont-New England Nuclear) were used for the sequencing reactions. Nucleotide sequences were determined by electrophoresis on 6 and 8% polyacrylamide gels followed by exposure of the dried gel to a film.
Human Bone Sialoprotein

**RESULTS**

A puc18 plasmid containing rat bone sialoprotein cDNA was grown in JM101 E. coli and purified through two CsCl density gradients. The cDNA was released from the plasmid by digestion with EcoRI and a 1.8-kb fragment purified by electrophoresis on agarose gels. A λ Zap gt11 expression library made from mRNA of human-bone-derived cells in culture was screened with nick-translated rat BSP cDNA under standard hybridization conditions. Several positive plaques were purified and confirmed to correspond to Bluescript plasmids with R408 helper phage. Release of the cDNA insert of the largest clone (B6) with EcoRI showed both a 1.0- and 2.0-kb fragment. The 1.0-kb EcoRI fragment was separately subcloned into M13 for sequencing. The 1.0-kb fragment contained the complete protein-encoding sequence of the cDNA (Fig. 1). The 16 NH₂-terminal amino acids of the deduced protein sequence are hydrophobic in character, similar to known signal sequences that mark the protein for processing into the endoplasmic reticulum and subsequent secretion from the cell. The underlined protein sequence is identical to that found in the protein isolated from human bone when sequenced by automated Edman degradation (21). The first three Asn-X-Ser/Thr consensus sequences for possible Asn-linked glycosylation (arrowheads, Fig. 1) are similar to those found in the previously published rat bone sialoprotein sequence (5) while the fourth (most COOH-terminal) is found only in the human sequence. The Arg-Gly-Asp (RGD) tripeptide (boxed in Fig. 1) thought to mediate the cell attachment properties of the bone sialoprotein is found in the human protein in approximately the same position as in the rat protein. Indeed, Fig. 2 shows that much of the human bone sialoprotein sequence is identical to that of the rat.

Fig. 3 is a diagram showing several interesting features of the protein. Three regions of the molecule are unusually rich in acid amino acids and these sequences showed regions of strong α-helical tendencies using computer protein secondary structure prediction methods. The same University of Wisconsin Genetics Computer Group program (22) used in the Garnier et al. methods (23), showed that the two tyrosine-rich regions flanking the RGD tripeptide for both rat and human BSP contained a significant number of turn-inducing residues to the extent that the predicted structures were almost exclusively composed of turns. Regions of pure turns have no structural meaning, but because these two RGD-flanking tyrosine regions matched well with current hypotheses on the consensus sequence for tyrosine sulfation (24) including: 1) a large number of turn-inducing acid amino acids; 2) a significant number of acid amino acids; and 3) a lack of disulfide

---

**Human Bone Sialoprotein**

The human and rodent parental cells, fusion procedure, and isolation and characterization of hybrids have been described (17-19). In general, hybrid cells were analyzed for the presence of all human chromosomes (except Y) by standard isoenzyme analysis as well as by Southern analysis with probes from previously localized genes. Hybrids were also frequently checked by cytogenetic analysis.

Southern Analysis—DNA was isolated from human-hamster and human-mouse hybrid cell lines, digested with EcoRI, size-fractionated on 0.7% agarose gel electrophoresis, partially depurinated, and transferred to positively charged membranes in 0.5 M NaOH (20). Membranes were washed 24–48 h at 42 °C with a 32P-labeled (nick translation) 1-kb EcoRI fragment of clone B6, the complete sequence of the largest clone (B6) with EcoRI showed both a 1.0- and 2.0-kb fragment. The 1.0-kb fragment was not completely sequenced. The flanking 3' 2.0-kb fragment was not completely sequenced. The EcoRI fragment (purified and subcloned into a second Bluescript plasmid, clone B6-5g) was labeled by nick translation and used to probe the Northern blots, and hybridized with the 1.0-kb 5' EcoRI fragment of the B6 clone.

Cell Hybrids—The human and rodent parental cells, fusion procedure, and isolation and characterization of hybrids have been described (17–19). In general, hybrid cells were analyzed for the presence of all human chromosomes (except Y) by standard isoenzyme analysis as well as by Southern analysis with probes from previously localized genes. Hybrids were also frequently checked by cytogenetic analysis.

**DNA Restriction Fragment Length Polymorphisms**—DNA was isolated from the peripheral leukocytes of 10 unrelated normal individuals, digested with restriction endonucleases (EcoRI, HindIII, BamHI, XbaI, SacI, TuqI, PvuII, PstI, EcoRV, BglII, MspI, and HpaII), digested with restriction endonucleases (EcoRI, HindIII, BamHI, XbaI, SacI, TuqI, PvuII, PstI, EcoRV, BglII, MspI, and HpaII), and hybridized with 5' EcoRI fragment of clone B6, the complete sequence of the largest clone (B6) with EcoRI showed both a 1.0- and 2.0-kb fragment. The 1.0-kb fragment was not completely sequenced. The flanking 3' 2.0-kb fragment was not completely sequenced. The EcoRI fragment (purified and subcloned into a second Bluescript plasmid, clone B6-5g) was labeled by nick translation and used to probe the Northern blots, and hybridized with the 1.0-kb 5' EcoRI fragment of the B6 clone.

---

The Northern Analysis—3.5 μg portions of total RNA from a variety of cultured cells and tissues from human, cow, and rat (described below) were electrophoresed in 1.2% agarose-formaldehyde gels and transferred to nitrocellulose as described (13). When the entire 3-kb insert was grown in culture was screened by digestion with EcoRI and a 1.8-kb fragment purified by electrophoresis on agarose gels. A λ Zap gt11 expression library made from mRNA of human-bone-derived cells in culture was screened with nick-translated rat BSP cDNA under standard hybridization conditions. Several positive plaques were purified and confirmed to correspond to Bluescript plasmids with R408 helper phage. Release of the cDNA insert of the largest clone (B6) with EcoRI showed both a 1.0- and 2.0-kb fragment with only the 1.0-kb fragment hybridizing with the rat probe on Southern analysis (data not shown). The 1.0- and 2.0-kb EcoRI fragments were separately subcloned into M13 for sequencing. The 1.0-kb fragment contained the complete protein-encoding region of the cDNA (Fig. 1). The 16 NH₂-terminal amino acids of the deduced protein sequence are hydrophobic in character, similar to known signal sequences that mark the protein for processing into the endoplasmic reticulum and subsequent secretion from the cell. The underlined protein sequence is identical to that found in the protein isolated from human bone when sequenced by automated Edman degradation (21). The first three Asn-X-Ser/Thr consensus sequences for possible Asn-linked glycosylation (arrowheads, Figure 1) are similar to those found in the previously published rat bone sialoprotein sequence (5) while the fourth (most COOH-terminal) is found only in the human sequence. The Arg-Gly-Asp (RGD) tripeptide (boxed in Fig. 1) thought to mediate the cell attachment properties of the bone sialoprotein is found in the human protein in approximately the same position as in the rat protein. Indeed, Fig. 2 shows that much of the human bone sialoprotein sequence is identical to that of the rat.

Fig. 3 is a diagram showing several interesting features of the protein. Three regions of the molecule are unusually rich in acid amino acids and these sequences showed regions of strong α-helical tendencies using computer protein secondary structure prediction methods. The same University of Wisconsin Genetics Computer Group program (22) used in the Garnier et al. methods (23), showed that the two tyrosine-rich regions flanking the RGD tripeptide for both rat and human BSP contained a significant number of turn-inducing residues to the extent that the predicted structures were almost exclusively composed of turns. Regions of pure turns have no structural meaning, but because these two RGD-flanking tyrosine regions matched well with current hypotheses on the consensus sequence for tyrosine sulfation (24) including: 1) a large number of turn-inducing acid amino acids; 2) a significant number of acid amino acids; and 3) a lack of disulfide backbone elements.

---

The Northern Analysis—3.5 μg portions of total RNA from a variety of cultured cells and tissues from human, cow, and rat (described below) were electrophoresed in 1.2% agarose-formaldehyde gels and transferred to nitrocellulose as described (13). When the entire 3-kb insert was grown in culture was screened by digestion with EcoRI and a 1.8-kb fragment purified by electrophoresis on agarose gels. A λ Zap gt11 expression library made from mRNA of human-bone-derived cells in culture was screened with nick-translated rat BSP cDNA under standard hybridization conditions. Several positive plaques were purified and confirmed to correspond to Bluescript plasmids with R408 helper phage. Release of the cDNA insert of the largest clone (B6) with EcoRI showed both a 1.0- and 2.0-kb fragment with only the 1.0-kb fragment hybridizing with the rat probe on Southern analysis (data not shown). The 1.0- and 2.0-kb EcoRI fragments were separately subcloned into M13 for sequencing. The 1.0-kb fragment contained the complete protein-encoding region of the cDNA (Fig. 1). The 16 NH₂-terminal amino acids of the deduced protein sequence are hydrophobic in character, similar to known signal sequences that mark the protein for processing into the endoplasmic reticulum and subsequent secretion from the cell. The underlined protein sequence is identical to that found in the protein isolated from human bone when sequenced by automated Edman degradation (21). The first three Asn-X-Ser/Thr consensus sequences for possible Asn-linked glycosylation (arrowheads, Figure 1) are similar to those found in the previously published rat bone sialoprotein sequence (5) while the fourth (most COOH-terminal) is found only in the human sequence. The Arg-Gly-Asp (RGD) tripeptide (boxed in Fig. 1) thought to mediate the cell attachment properties of the bone sialoprotein is found in the human protein in approximately the same position as in the rat protein. Indeed, Fig. 2 shows that much of the human bone sialoprotein sequence is identical to that of the rat.

Fig. 3 is a diagram showing several interesting features of the protein. Three regions of the molecule are unusually rich in acid amino acids and these sequences showed regions of strong α-helical tendencies using computer protein secondary structure prediction methods. The same University of Wisconsin Genetics Computer Group program (22) using the Garnier et al. methods (23), showed that the two tyrosine-rich regions flanking the RGD tripeptide for both rat and human BSP contained a significant number of turn-inducing residues to the extent that the predicted structures were almost exclusively composed of turns. Regions of pure turns have no structural meaning, but because these two RGD-flanking tyrosine regions matched well with current hypotheses on the consensus sequence for tyrosine sulfation (24) including: 1) a large number of turn-inducing acid amino acids; 2) a significant number of acid amino acids; and 3) a lack of disulfide backbone elements.
Human Bone Sialoprotein

FIG. 2. Comparison of human and rat bone sialoprotein nucleotide (A) and amino acid (B) sequences. The nucleic acid sequences were analyzed using the Compare program of the University of Wisconsin Genetics Computer Group using the standard comparison methods of Maizel and Lenk (27). The broken line on the diagonal represents regions of in-register homology between the two cDNA sequences. Sequences off of the diagonal are regions of homology further off-register. Panel B shows that human and rat BSP have several regions of amino acid identity and an overall homology throughout their entire lengths.

FIG. 3. Diagram outlining the features of human bone sialoprotein. Post-translational modifications are not indicated.

FIG. 4. Sequence similarities of the nominal tyrosine repeats in both the human (h) and the rat (r) bone sialoprotein deduced sequences. Numbers on the right are amino acid positions as numbered from the start methionine. The consensus sequence is shown at the bottom.

FIG. 5. Detection of bone sialoprotein mRNA in various tissues and cell cultures. Notice that bone cells and the rat osteosarcoma (UMR-106) have relatively large amounts of message and that decidua (penc) and cartilage all contain small but reproducible amounts of a 2-kb message. Cells and tissues are described under "Materials and Methods."

The human BSP gene was localized to a specific chromosome by Southern analyses of human-hamster and human-mouse somatic cell hybrid DNA with the 1-kb BSP-coding sequence cDNA probe. This probe identified 0.85-, 1.3-, and very weakly hybridizing 16-kb bands in EcoRI digests of human DNA (Fig. 6), whereas 4.6- and 15-kb faint EcoRI mouse bands and no detectable cross-hybridizing hamster bands were detected with this probe. All three EcoRI human bands were either present or absent in the same hybrids indicating that all three sequences were located on a single human chromosome. Analysis of the entire panel of hybrids (Table I) permitted unambiguous assignment of the BSP gene to human chromosome 4, and it segregated discordantly (219%) with all other human chromosomes.

Analyses of the number and size of the restriction fragments hybridizing to the 1-kb cDNA probe (containing the 5′ flanking sequence and the entire protein-encoding sequence) in human DNA digested with 12 different restriction endonucleases indicated that BSP is probably a single copy gene.
and human olacental (H) DNA are also shown. Sizes of the hvbrid-
different hybrid cell DNA; parental Chinese hamster (C), mouse (M),
sequences (+) is indicated above the lanes.

Single 21-kb hybridizing band was found in KpnI digests, and
only two bands were found in HindIII, BumHI, XbaI, SacI, PuuII, PstI, EcoRV, and BlglII digests. These results are
consistent with the presence of a single BSP gene if one
restriction site for each of these enzymes is located within an
intron. Hybridization with subfragments of the cDNA used
as probes would be required to confirm this hypothesis and exclude the presence of two copies of the BSP gene at this
locus, however. Based upon the total sizes of hybridizing
fragments, it could be determined that the total size of the
portion of the gene encoding the 5' flanking and protein-
encoding sequences of human BSP does not exceed 9.1 kb.
Because genomic sequences corresponding to the 3' flanking
sequences of a cDNA usually do not contain any introns, one
can reasonably assume that the total length of the gene is
unlikely to exceed 11.1 kb (9.1 kb plus 2.0 kb of the 3' flanking
cDNA).

No restriction fragment length polymorphisms were de-
tected with the BSP cDNA probe in DNA isolated from 10
normal unrelated individuals digested with 12 different re-
stiction enzymes. These included: EcoRI, HindIII, BamHI, XbaI, SacI, TaqI, PvuII, PstI, EcoRV, BlglII, MspI, and KpnI.

**DISCUSSION**

A cDNA encoding human bone sialoprotein has been cloned
and sequenced. The longest cDNA clone is approximately 3.0
kb, containing an open reading frame of 951 bases within a
5' 1.0-kb EcoRI fragment. The deduced protein sequence
contains a 16-amino acid hydrophobic signal sequence directly
followed by the sequence seen earlier by NH*-terminal protein
sequence (37). BSP is apparently made first as a pre-protein
contains a 16-amino acid hydrophobic signal sequence directly
sequencing (21). BSP is apparently made first as a pre-protein
before being made into the highly modified glycopro-
oligosaccharide attachment sites (one more than seen in the
bone matrix. The deduced protein sequence contains four putative asparagine-linked
oligosaccharide attachment sites (one more than seen in the
rat; Ref. 5) and no disulfide bonds. The protein has three
regions rich in acidic amino acid residues and three tyrosine-
rich sequences. Two of the tryosine-rich regions flank the
two flanking regions completely fit the established consensus
requirements for tyrosine sulfation. Indeed, the RGD-con-
taining region of vitronectin itself also may fit the require-
ments for tyrosine sulfation, thus raising the possibility that
more than one of the proteins that bind to the vitronectin
receptor may have their affinity or secondary affects altered
by sulfation of their tyrosines. (Fibronectin and a variety of
other RGD-containing proteins, including osteopontin and
fibrinogen, appear to contain tyrosines within a short distance
of the RGD tripeptide, but lack the consensus requirements
for sulfation.)

### TABLE I

<table>
<thead>
<tr>
<th>Gene/chromosome</th>
<th>Human chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1/2</td>
</tr>
<tr>
<td></td>
<td>3/4</td>
</tr>
<tr>
<td></td>
<td>5/6</td>
</tr>
<tr>
<td></td>
<td>7/8</td>
</tr>
<tr>
<td></td>
<td>9/10</td>
</tr>
<tr>
<td></td>
<td>11/12</td>
</tr>
<tr>
<td></td>
<td>13/14</td>
</tr>
<tr>
<td></td>
<td>15/16</td>
</tr>
<tr>
<td></td>
<td>17/18</td>
</tr>
<tr>
<td></td>
<td>19/20</td>
</tr>
<tr>
<td></td>
<td>21/22</td>
</tr>
<tr>
<td></td>
<td>23/24</td>
</tr>
<tr>
<td></td>
<td>25/26</td>
</tr>
</tbody>
</table>

**Segregation of the BSP II gene with human chromosome 4**

The BSP gene was detected an 0.85-, 1.3-, and 15-kb bands in
EcoRI-digested human-rodent somatic cell hybrid DNAs after South-
ern hybridization with a 5' 1-kb human BSP cDNA containing the
entire protein-encoding sequence and the 5' flanking sequences. Cross-hybridization with hamster sequences was not observed under
the conditions of hybridization and weakly hybridizing 4.6- and 15-
kb bands were detected in mouse DNA. Detection of human bands is
 correlated with the absence of each human chromosome in the group
of somatic cell hybrids. Discordancy represents the presence of the
gene in the absence of the chromosome (+/-) or presence of the gene
despite the presence of the chromosome (+/+), and the sum of these
differences divided by total hybrids examined (×100) represents percent
discordance. The human-hamster hybrids consisted of 27 primary
clones and 14 subclones (16 positive of 41 total) and the human-
mouse hybrids represented 14 primary clones and 40 subclones (33
positive of 54 total).

**Fig. 6. Hybridization of human-hamster (top) and human-
mouse (bottom) EcoRI-digested somatic cell hybrid DNA with
a 1.0-kb human BSP probe. Each numbered lane contains a
different hybrid cell DNA; parental Chinese hamster (C), mouse (M),
and human placental (H) DNA are also shown. Sizes of the hybrid-
izing human (0.85, 1.3, and 15 kb) and homologous mouse (4.6 and
15 kb) fragments are shown. Presence of the hybridizing human
sequences (+) is indicated above the lanes.
for the sulfation of these groups.) Mouse bone cells in culture have recently been shown to label BSP with $^{35}$SO$_4$ (25) although it was not determined whether the label was on tyrosine. Preliminary results have shown the sulfation of both tyrosine and oligosaccharides of the BSP made by the rat osteosarcoma cell line, UMR 106-01. The human protein sequence is approximately 70% identical to the rat amino acid sequence with significant stretches of identity in the first third of the molecule.

Northern blot analysis of total RNA from a variety of tissues from human, cow, and rat shows that the message for BSP is rich in bone cells and the rat osteosarcoma cell line UMR-106 (but interestingly, barely detectable in the rat osteosarcoma ROS-17/2 and the human osteosarcoma, SAOS-2). RNA from other tissues showed only minor or undetectable levels of BSP mRNA. While earlier work with both protein analysis and Northern analysis suggested that BSP may have been specific to bone, the minor but reproducible positive bands in cartilage and decidua, for example, suggest that this protein may be expressed sparingly in highly localized sites in tissues other than bone. In situ hybridization and immunolocalization studies confirming these observations are currently underway in this laboratory.

Using mouse human and hamster human hybrid cell lines, we have shown that human bone sialoprotein maps unambiguously to chromosome 4. BSP appears to be a single copy gene unlikely to exceed 11.1 kb, although the possibility of a second copy at this locus has not been rigorously excluded. No restriction length polymorphisms were observed in 10 normal individuals using 12 restriction enzymes.

Bone sialoprotein is a highly glycosylated extracellular protein (a proteoglycan in the rabbit (4) and possibly in quail (26)) made by osteoblasts and probably a limited number of other cell types. This cell-adhesion molecule appears to use the integrin system to mediate cell attachment in culture (7). In vivo, the protein may be used by the osteoblasts for attachment to the extracellular matrix. Interestingly, in a single case of bovine osteopetrosis, BSP was significantly decreased in the bone matrix. Because the protein is highly enriched in bone, it may prove to be an excellent marker of bone cell metabolism in bone disease.

REFERENCES


Human bone sialoprotein. Deduced protein sequence and chromosomal localization.
L W Fisher, O W McBride, J D Termine and M F Young