Human Bone Sialoprotein

DEDUCED PROTEIN SEQUENCE AND CHROMOSOMAL LOCALIZATION*

(Received for publication, October 5, 1989)

Larry W. Fisher‡, O. Wesley McBride†, John D. Termine‡, and Marian F. Young‡

From the †Bone Research Branch, National Institute of Dental Research and the §Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

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 carboxy-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 hard tissue elements that establish the biophysical conditions necessary for orderly and complete mineralization. The first 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 (1). This small glycoprotein was later hypothesized to be a breakdown product of a M, 75,000 sialoprotein containing 12% sialic acid, 7% glucoseamine, 6% galactosamine, and ~50% protein (2). Competition enzyme-linked immunosorbent 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, Amer sham Corp.) rat bone sialoprotein cDNA (a 1.8-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 (1.0 and 2.0 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 diaryoxy chain termination method (12). The Sequenase kit (United States Biochemicals) using both GTP and ITP nucleotide mixtures and [α-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 expo-

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Room 106, Bldg. 30, BRB/NIDR/NIH, Bethesda, MD 20892.

1 The abbreviations used are: BSP, bone sialoprotein; kb, kilobase pair.
A pucl8 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 gel electrophoresis in 1.2% agarose-formaldehyde gels and transferred to nylon membranes. When the entire 3-kb insert was labeled in a pattern commonly seen for Alu-type repeats. The 1-kb 5' EcoRI fragment (purified and subcloned into a Second-Bluescript plasmid, plasmid B6-5g) containing the entire protein-encoding region did not hybridize with Alu-type repeat probes and was used for all Northern and Southern analyses. Hybrids were also frequently checked by cytogenetic analysis.

RESULTS

A pucl8 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 32P-labeled rat BSP cDNA under intermediate stringency conditions. Several positive plaques were purified and converted into their corresponding Blue-Script plasmids with R408 helper phage. Release of the cDNA was accomplished from the B6-5g clone by digestion with EcoRI and a 1.8-kb fragment purified by gel electrophoresis and transferred to nylon membranes. When the entire 3-kb insert was labeled in a pattern commonly seen for Alu-type repeats. The 1-kb 5' EcoRI fragment (purified and subcloned into a Second-Bluescript plasmid, plasmid B6-5g) containing the entire protein-encoding region did not hybridize with Alu-type repeat probes and was used for all Northern and Southern analyses. 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, XhoI, SacI, TuqI, BglII, MspI, and I[@]), and size-fractionated by gel electrophoresis, transferred to nylon membranes, and hybridized with the 1.0-kb 5' EcoRI fragment of the B6 clone.
Human Bone Sialoprotein

500 1000

500 1000

A

RAT cDNA Sequence

B

Human Bone Sialoprotein

500 1000

Acidic Clusters

pre

Tyr-Consensus Repeats

RGD

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 (pont) and cartilage all contain small but reproducible amounts of a 2-kb message. Cells and tissues are described under “Materials and Methods.”

Bonding or possibly glycosylated amino acids, we have proposed that this region may be sulfated. The most NH2-terminal tyrosine-rich region of the human sequence gave a predicted secondary protein structure of a β-sheet while the same region in the rat predicted a series of turns more reminiscent of the RGD-flanking tyrosine regions. Tyrosine sulfation of these NH2-terminal tyrosine regions appears unlikely in both species because of the lack of negatively charged amino acids and the strong likelihood of the presence of O-linked oligosaccharides in the general region (computer analyses not shown).

The three tyrosine-rich regions of both species can be aligned as in Fig. 4 and this suggests that the three regions may have been derived from a common primordial gene.

Both Northern analysis of the total RNA from a variety of tissues and Southern analysis of human chromosomal DNA showed a large number of broad hybridization bands suggesting that the BSP probe made from the complete plasmid insert contained Alu-like repeats (data not shown). A second probe was made from the 1-kb 5′ EcoRI fragment containing the entire protein-encoding region but little of the 3′ flanking region and this resulted in a much more limited number of sharp bands. This EcoRI-released 1 kb was labeled with 32P by nick translation and used to probe total RNA isolated from a variety of cells and tissues from human, cow, and rat (Fig. 5). Strong bands at 2 kb were seen in bone cell-derived lanes for all three species. The rat osteosarcoma cell line, UMR-106, was the only other lane to show a large amount of hybridization signal. Weak signals of the appropriate molecular weight were seen in the other lanes with the possible exceptions of the human SAOS-2 cells and the fetal bovine and rat skin cells. These results confirm that bone cells are the richest source of mRNA for BSP but also suggest that other cell types may express BSP.

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 (≥19%) 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. A
and human olacental (H) DNA are also shown. Sizes of the hybridizing DNA; parental Chinese hamster (C), mouse (M), human (0.85, 1.3, and 16 kb) and homologous mouse (4.6 and 15 kb) fragments are shown. Presence of the hybridizing human sequences (+) is indicated above the lanes.

A cDNA encoding human bone sialoprotein has been cloned from 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 hybridizing human (0.85, 1.3, and 16 kb) and homologous mouse (4.6 and 15 kb) fragments are shown. Presence of the hybridizing human sequences (+) is indicated above the lanes.

FIG. 6. Hybridization of human-hamster (top) and human-mouse (bottom) EcoRI-digested somatic cell hybrid DNA with a 1.0-kb human BSP cDNA. 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 hybridizing human (0.85, 1.3, and 16 kb) and homologous mouse (4.6 and 15 kb) fragments are shown. Presence of the hybridizing human sequences (+) is indicated above the lanes.

single 21-kb hybridizing band was found in KpnI digests, and only two bands were found in HindIII, XbaI, SacI, PvuII, PstI, EcoRV, and BglII 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 detected with the BSP cDNA probe in DNA isolated from 10 normal unrelated individuals digested with 12 different restriction enzymes. These included: EcoRI, HindIII, BamHI, XbaI, SacI, TaqI, PvuII, PstI, EcoRV, BglII, 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 preceded by the sequence seen earlier by NH2-terminal protein sequencing (21). BSP is apparently made first as a pre-protein containing a 16-amino acid hydrophobic signal sequence directly followed by the sequence seen earlier by NH2-terminal protein sequencing (21). BSP is apparently made first as a pre-protein containing a 16-amino acid hydrophobic signal sequence directly followed by the sequence seen earlier by NH2-terminal protein sequencing (21).

The BSP gene was detected as 0.85-, 1.3-, and 16-kb bands in EcoRI-digested human-rodent somatic cell hybrid DNAs after Southern 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 (−/+). The sum of these numbers divided by total hybrids examined (×100) represents percent discordancy. 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).

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 (−/+). The sum of these numbers divided by total hybrids examined (×100) represents percent discordancy. 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).

The BSP gene was detected as 0.85-, 1.3-, and 16-kb bands in EcoRI-digested human-rodent somatic cell hybrid DNAs after Southern 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 (−/+). The sum of these numbers divided by total hybrids examined (×100) represents percent discordancy. 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).

No restriction fragment length polymorphisms were detected with the BSP cDNA probe in DNA isolated from 10 normal unrelated individuals digested with 12 different restriction enzymes. These included: EcoRI, HindIII, BamHI, XbaI, SacI, PvuII, PstI, EcoRV, BglII, 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 preceded by the sequence seen earlier by NH2-terminal protein sequencing (21). BSP is apparently made first as a pre-protein containing a 16-amino acid hydrophobic signal sequence directly preceded by the sequence seen earlier by NH2-terminal protein sequencing (21).

The contribution of human bone sialoprotein (BSP) to the micro- and macrostructure of extracellular matrix is not yet understood. However, the presence of BSP in the bone matrix suggests a role in skeletal development and repair. The cloned cDNA and the mRNA from human BSP will provide a useful tool for the study of this protein in these processes.

The total DNA sequence of the human BSP gene, which codes for a 22-kDa protein, is approximately 9.1 kb. This is consistent with the size of the cDNA clone (3.0 kb), which contains the complete coding region for the protein. The cDNA sequence is sufficient to predict the amino acid sequence of the protein with a high degree of confidence.

**TABLE I**

<table>
<thead>
<tr>
<th>Human chromosome</th>
<th>Gene/chromosome</th>
<th>Discordancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+/+</td>
<td>-/+</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>34</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>13</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>15</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>16</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>17</td>
<td>32</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>31</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>29</td>
</tr>
<tr>
<td>20</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>21</td>
<td>33</td>
<td>16</td>
</tr>
<tr>
<td>22</td>
<td>20</td>
<td>29</td>
</tr>
<tr>
<td>X</td>
<td>29</td>
<td>20</td>
</tr>
</tbody>
</table>
for the sulfation of these groups.) Mouse bone cells in culture have recently been shown to label BSP with \( ^{35} \text{SO}_4 \) (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.\(^2\) 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.\(^3\) 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


Access the most updated version of this article at http://www.jbc.org/content/265/4/2347

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/4/2347.full.html#ref-list-1