Posttranslational Modifications to Human Bone Sialoprotein Determined by Mass Spectrometry

Joseph Zaia,[‡,§] Raymond Boynton,[‡,||] Dick Heinegård,[⊥] and Frank Barry*[,‡]

Osiris Therapeutics, Inc., Baltimore, Maryland 21231, and Department of Cell and Molecular Biology, Lund University, S-221 00 Lund, Sweden

Received May 1, 2001; Revised Manuscript Received July 23, 2001

ABSTRACT: Bone sialoprotein (BSP) is an acidic 301 amino acid protein expressed by osteoblasts and at a low level by hypertrophic chondrocytes. Its expression is highest during early stages of bone formation, and it is particularly abundant in the cells lining the surface of newly formed trabeculae. BSP contains numerous substituents which are anionic in nature and apparently essential for the function of the protein. Thus, the proposed role of BSP in hydroxyapatite nucleation and growth may depend on such modifying groups. The posttranslational modifications include several acidic oligosaccharides as well as phosphate and sulfate groups. This work combines matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry with selective enzyme treatment of BSP to provide new information on the precise distribution and structure of oligosaccharides, sulfate, and phosphate groups in BSP isolated from human bone. The results provide a high level of detail in the location of these modifying groups toward the end of providing a basis for further understanding the function of BSP in bone nucleation.

The formation and remodeling of bone is closely associated with the expression of bone sialoprotein (BSP)[1] and osteopontin (OPN). These matrix proteins are anionic, containing extended sequences of glutamic acid and aspartic acid, respectively. Extensive posttranslational modifications including serine and threonine phosphate, sialylated N- and O-linked oligosaccharides, and, in the case of BSP, tyrosine sulfate contribute further to the negative charge of these proteins (*1*). Both proteins contain an RGD sequence active in binding of cells via the $\alpha_v\beta_3$ integrin surface receptor (*2*−*5*). OPN is widely expressed in osteoblasts, kidney tubular cells, macrophages, T-lymphocytes, and cells in the brain as well as by a number of transformed cells (*6*). BSP expression is limited to osteoblasts and, at low levels, hypertrophic chondrocytes (*7*, *8*).

Mature human BSP has 301 amino acids that can be divided into 3 domains. The amino (residues 1−45) and carboxy (residues 219−301) domains contain the majority of the bulky aromatic groups in the protein. The amino-terminal domain is comparatively rich in basic residues, while the carboxy domain is comparatively acidic.

The C-terminal domain contains 15 tyrosine residues flanked by acidic residues, meeting the requirements for a tyrosine sulfation motif (*9*, *10*). Such tyrosine residues may be involved in cell attachment (*11*), and it is likely that sulfation modulates this activity.

Three acid-rich sequences are contained in the central domain (residues 46−218), each characterized by polyglutamic acid repeats (residues 52−87, 130−157, and 179−188). These repeats as well as the amino- and carboxy-domain Tyr residues are highly conserved among mammalian BSPs (*12*, *13*). Interestingly, the only sequences where hydrophobic residues cluster to give a localized hydrophobic character (*14*) are 91−96 (LSATTL) and 108−114 (YTG-LAAI), separating the two largest acid-rich sequences. BSP has an extended, flexible structure capable of binding different partners simultaneously (*15*).

Studies by immunolocalization and in situ hybridization show that BSP expression is highest during early stages of bone formation, particularly in the early osteoblast (*7*). Thus, there is a sharp concentration gradient of BSP at osteocartilaginous interfaces, and BSP mRNA is particularly abundant in osteoblasts facing cartilage (*8*, *16*), e.g., in those cells lining the surface of newly formed trabeculae as well as in odontoblasts and cementoblasts (*7*, *8*). Levels decrease with bone growth and mineralization, and imbedded osteocytes display much lower levels of BSP mRNA (*7*, *16*) in mature bone. These lines of evidence indicate an important role for BSP, e.g., in maintaining the cartilage−bone interface.

The presence of BSP in the early osteoid has led to the hypothesis that the protein has a role in hydroxyapatite nucleation and growth, mediated by its acidic groups including the polyglutamic acid repeats (*17*). Studies of calcium phosphate precipitation in vivo have shown that BSP expression increases coincident with mineralization (*8*, *18*). The presence of phosphate groups on BSP and the significant levels of alkaline phosphatase in the bone matrix lead to the speculation that phosphorylation may be a regulating mechanism for BSP activity, including hydroxyapatite nucleation.

* To whom correspondence should be addressed.
‡ Osiris Therapeutics, Inc.
§ Present address: Department of Biochemistry, Boston University School of Medicine, Boston, MA.
|| Present address: Biogen, Inc., Cambridge, MA.
⊥ Lund University.

[1] Abbreviations: BSP, bone sialoprotein; CHCA, α-cyano-4-hydroxycinnamic acid; Gal, galactose; GalNAc, *N*-acetylgalactosamine; Glc, glucose; GlcNAc, *N*-acetylglucosamine; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; MS, mass spectrometry; NeuAc, *N*-acetylneuraminic acid; OPN, osteopontin; SA, sialic acid; TFA, trifluoroacetic acid.

10.1021/bi010887r CCC: $20.00
Published on Web 10/02/2001

Analysis of recombinant human BSP by one-dimensional NMR (*15*) has indicated that BSP in solution is unstructured and acquires appropriate structure upon binding to specific ligands. Posttranslational modifications account for 50% of the mass of BSP, and these have been elucidated in detail in a series of experiments on BSP from a rat osteoblast cell line (*10*, *19*, *24*). While these studies in combination helped to explain much of the structural details of BSP, additional information is needed to account for the precise distribution and structure of oligosaccharide, sulfate, and phosphate substituents in BSP isolated from human bone. In this study, we have examined methods by which matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry may be applied to the analysis of complex and heterogeneous glycoproteins such as BSP. Many of the methods that we developed for the analysis of human BSP will find application in the structural elucidation of other complex glycoproteins and proteoglycans.

EXPERIMENTAL PROCEDURES

Purification of BSP. BSP was purified from adult human bone using a published procedure (*1*).

Amino Acid Analysis. Protein samples were hydrolyzed using a Waters (Milford, MA) Picotag station. The hydrolysis was carried out in 6 M HCl at 105 °C for 24 h. Amino acid analysis was performed using a Hewlett-Packard Aminoquant instrument following the instructions of the manufacturer.

Glycosidase Digestion of BSP. All glycosidases were purchased from Genzyme Diagnostics (Cambridge, MA). BSP (2.5 μg) was digested with *N*-glycanase (75 milliunits) in 0.1 M Tris-HCl, pH 8.5, overnight at 37 °C. The solution pH was adjusted to 7.0 by the addition of HCl, neuraminidase (3.1 milliunits) was added, and the mixture was incubated for 2 h at 37 °C. *O*-Glycanase (75 microunits) was added and the mixture digested overnight at 37 °C.

Endoproteinase Digestion of BSP. Proteases were purchased from Boehringer Mannheim Biochemicals (Indianapolis, IN). The BSP sample (approximate concentration 0.5 μg/μL) was digested with trypsin in 0.1 M Tris-HCl, pH 8.5, or with endoproteinase Asp-N in 0.1 M Tris-HCl, pH 7.5, using 1% protease by weight and incubation for 6 h at 37 °C.

Exoproteinase Digestion of BSP Tryptic Peptides. Tryptic peptide (2 μL of an approximately 10 ng/μL solution in 0.05% TFA) was mixed with 0.1 M Tris-HCl, pH 7.5 (2 μL), and aminopeptidase M (Boehringer, 2 μL, 0.4 milliunit/μL in 0.1 M Tris-HCl, pH 7.5). Aliquots (1.0 μL) were removed at time points and mixed with 5 μL of α-cyano-4-hydroxycinnamic acid (CHCA) solution (see below).

Mass Spectrometry. MALDI-TOF mass spectra were acquired using a Hewlett-Packard (Palo Alto, CA) G2025A mass spectrometer. A sample was diluted to approximately 1 μM in matrix solution and dried on the sample target. Sinapinic acid (SA) was used as the matrix for intact BSP and CHCA (both from Hewlett-Packard) for proteolytic peptides. After drying, the sample−matrix spots were washed by immersing for 3−5 s in 0.05% trifluoroacetic acid (TFA) solution. The samples applied using SA were subsequently treated with 0.7 μL of a formic acid solution (100:100:35 water/acetonitrile/formic acid) and allowed to dry before being placed into the mass spectrometer.

Table 1: Amino Acid Analysis of Human BSP

residue	% observed	% calculated
Asx	13.0	13.1
Glx	21.6	21.73
Ser	7.3	7.67
His	1.6	1.6
Gly	11.1	11.5
Thr	9.2	9.58
Ala	4.4	5.43
Arg	7.3	2.88
Tyr	7.3	7.35
Val	2.2	2.24
Met	0.3	0
Phe	2.2	2.24
Ile	1.6	2.24
Leu	2.5	3.83
Lys	4.4	4.79
Pro	3.8	3.83
total	100.0	100.01

Electrospray liquid chromatography/mass spectrometry was performed using a Hewlett-Packard (Palo Alto, CA) 5989B mass spectrometer equipped with a 59987A API-Electrospray LC/MS Interface. The mass spectrometer was interfaced with a 1090 Series II liquid chromatograph fitted with a 1.7 μL flow cell and a Vydac (Hesperia, CA) protein and peptide C18 column (2.1 mm × 300 mm). Solvents were A (H_2O, 0.05% TFA) and B (30:70 H_2O/ACN, 0.035% TFA). Proteolytic digests were acidified with TFA and injected into the liquid chromatograph. Data were acquired for 80 min using a flow rate of 200 μL/min and the following gradient program: 0 min, 3.5% B; 55 min, 55% B; 75 min, 100% B; 80 min, 100% B. The column effluent was split so that 90% was collected and 10% analyzed by the mass spectrometer. The inhibitory effects of TFA on the electrospray ionization process were overcome by post-split addition of 10 μL/min of a 75:25 propionic acid/2-propanol mixture (*20*). Data analysis was performed using the Hewlett-Packard G1047A LC/MS software.

Monosaccharide Analysis. The monosaccharide composition was determined after hydrolysis in 2 M TFA for 3 h for analysis of neutral sugars and in 4 M HCl for 5 h for analysis of hexosamines. After hydrolysis, samples were taken to dryness, dissolved in water, and separated on a Dionex (Sunnyvale, CA) CarboPac PA-1 column. The column was eluted with 0.05 M NaOH, and pulsed amperometric detection was used (*21*).

Superdex Separation. Fifty micrograms of BSP was deglycosylated as above and digested with 0.5 μg of TPCK-trypsin (Pierce Chemical Co., Rockford, IL). Ten micrograms of this digest mixture in a volume of 25 μL was fractionated using a Superdex column (3.1 mm i.d., Amersham Pharmacia Biotech Products, Piscataway, NJ) equilibrated in 0.05% trifluoroacetic acid at 100 μL/min. The two largest tryptic peptides coeluted in 10−12 min.

RESULTS

BSP was isolated from human bone by guanidine hydrochloride extraction followed by purification to homogeneity by anion exchange chromatography. A single peak was detected by analytical gel filtration chromatography, and amino acid analysis indicated good agreement between calculated and observed residue percentages (Table 1). An

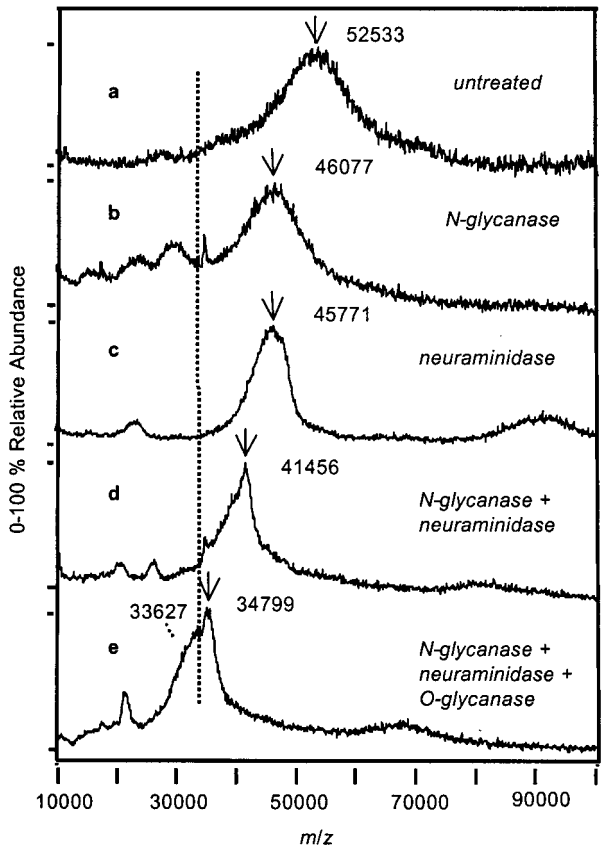

FIGURE 1: Deglycosylation of BSP followed by MALDI-TOF mass spectrometry. Intact BSP (a); BSP after digestion with *N*-glycanase (b); after digestion with neuraminidase (c); after digestion with *N*-glycanase and neuraminidase (d); and after digestion with *N*-glycanase, neuraminidase, and *O*-glycanase (e). The dotted line shows the calculated mass of mature, unmodified BSP core protein. The arrow shows the centroid *m*/*z* value obtained in each spectrum.

amino acid similarity search of the Swissprot database produced an unambiguous match with bone sialoprotein as the top score. Mass spectra of this material were obtained by MALDI-TOF in a series of studies involving selective enzymatic deglycosylation, and the results are shown in Figure 1 and the data are summarized in Table 2. The heterogeneous nature of the substituents on BSP is reflected by the broad peak in Figure 1a, the centroid of which is *m*/*z* 52 533 ± 388 (*22*) and the width 11 850 ± 1900. By comparison, the peak width for bovine serum albumin, a relatively homogeneous unmodified protein in the same mass range, is 600−1000 (result not shown). Treatment of BSP with *N*-glycanase (Figure 1b) decreased the centroid to *m*/*z* 46 077 ± 137, indicating that the total mass of N-linked oligosaccharides was 6456 ± 525 Da. Treatment with neuraminidase (Figure 1c) decreased the centroid to *m*/*z* 45 771 ± 327, corresponding to a total mass of *N*-acetylneuraminic acid substituents of 6762 ± 715, or 23 ± 2 residues. When the neuraminidase-treated BSP was subsequently digested with *N*-glycanase (Figure 1d), the peak centroid was reduced to *m*/*z* 41 459 ± 33, and the width was narrowed significantly. The mass of *N*-glycanase-sensitive oligosaccharide in this case was 4312 ± 1410 Da, corresponding to N-linked oligosaccharide without attached sialic acid. The total mass of sialic acid substituents on the N-linked oligosaccharides was 2144 ± 885 Da, or 7 ± 1 *N*-acetylneuraminic acid residues. The *N*-glycanase and neuraminidase-treated BSP was then treated with *O*-glycanase (Figure 1e) to produce a distinctly bimodal peak. The centroid of the sharper of these peaks was *m*/*z* 34 799 ± 35 (marked with an arrow in Figure 1e). This mass exceeds the calculated core protein mass for protonated BSP (33 543.39 Da, indicated by the vertical dotted line in Figure 1) by 1256 ± 35 Da. This discrepancy can be accounted for by phosphate (*23*) and sulfate (*10*) substitutions on the core protein. The second peak in this spectrum (Figure 1e) was broader and had a centroid *m*/*z* of 33 627 ± 104. Either this may be a variant of BSP that is substituted to a lesser degree or it may result from proteolytic trimming at either the N- or the C-terminus, a process which is known to occur during the isolation of BSP (*24*). The total BSP oligosaccharide mass is therefore 17 734 ± 423 Da.

Table 2: Oligosaccharide Composition of BSP from Human Bone[a]

	mass	peak width[b]
intact BSP	52533 ± 388	11850 ± 1900
N-glycanase	46077 ± 137	9350 ± 200
neuraminidase	45771 ± 327	7200 ± 650
N-glycanase + neuraminidase	41459 ± 33	4950 ± 150
N-glycanase + neuraminidase + *O*-glycanase	34799 ± 35	2700 ± 250[c]

Oligosaccharide Masses from Above Data

	total mass	total − NeuAc	NeuAc mass	no. of NeuAc
N-linked	6456 ± 525	4312 ± 1410	2144 ± 885	7 ± 1
O-linked	11278 ± 172	6660 ± 68	4618 ± 240	16 ± 1
total	17734 ± 423	10972 ± 1138	6762 ± 715	23 ± 2
percent[d]	33.8 ± 1.0	20.9 ± 2.3	12.9 ± 1.4	

[a] MALDI-TOF mass data obtained from human BSP with and without treatment with glycosidases. Three spectra were acquired for each sample, and the values reported are the average ± standard deviation. [b] Measured at half peak height. [c] Value reflects the width of the sharper component of the BSP profile. [d] Percentage of oligosaccharide in the measured BSP mass.

Table 3: Monosaccharide Analysis of BSP from Human Bone

residue	mol of residue/mol of protein	calculated mass (Da)[a]
fucose	1.5 ± 0.4	220 ± 60
galactosamine	16.0 ± 2.3	3250 ± 470
glucosamine	15.1 ± 2.3	3070 ± 470
galactose	21.8 ± 3.8	3530 ± 620
glucose	5.9 ± 2.5	960 ± 400
mannose	6.1 ± 1.7	990 ± 280
		total = 12020 ± 2300

[a] Masses for oligosaccharide residues were used. *N*-Acetylhexosamine residue masses were used for glucosamine and galactosamine.

Samples of BSP were hydrolyzed as described under Experimental Procedures and subjected to monosaccharide and amino acid analysis. The results of these two analyses produce the molar ratio of monosaccharide to core protein. The calculated carbohydrate mass obtained in this analysis was 12 020 ± 2300 Da (Table 3), corresponding to total oligosaccharide mass less sialic acid, which was not determined by this method. This value matches, within experimental error, that determined by mass spectrometry, 10 972 ± 1138 Da.

Human BSP has four putative sites of N-oligosaccharide addition, but the relatively low mannose content (6.1 ± 1.7

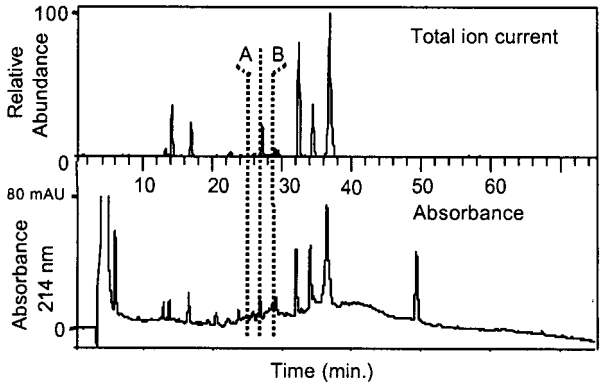

FIGURE 2: Reversed phase HPLC fractionation of the endoproteinase Lys-C digest of deglycosylated BSP. The upper trace shows the mass spectrometric total ion current chromatogram and the lower trace the absorbance chromatogram at 214 nm. The dotted lines refer to the portion of the eluate that was collected for fractions A and B. Further analysis of fractions A and B is discussed in the text and shown in Figures 3−5.

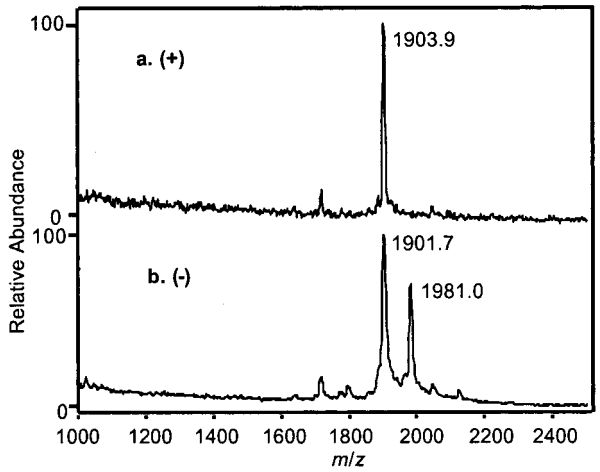

FIGURE 3: Positive (a) and negative (b) mode MALDI-TOF mass spectra of fraction A from the HPLC fractionation of the endoproteinase lysine-C digest of deglycosylated BSP (Figure 2).

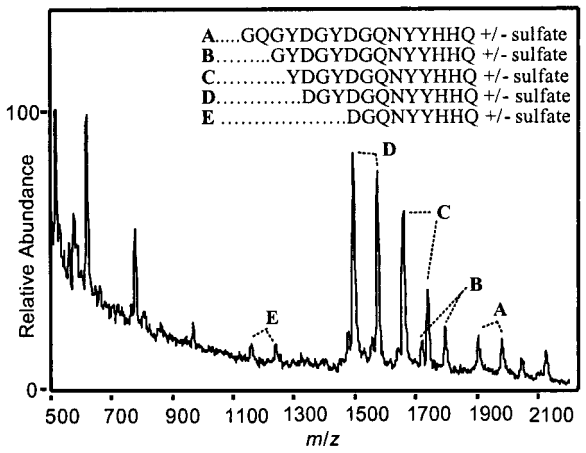

FIGURE 4: Negative mode MALDI-TOF mass spectrum of the aminopeptidase M digestion of fraction A (see Figure 2) acquired after 60 min of digestion. Pairs of ions correspond to sulfated and desulfated peptides containing the C-terminus of various lengths as shown.

mol/mol of protein) suggests that only two are occupied. It is also likely that they are of the complex type, containing no mannose beyond the core structure and with some sialic acid (7 ± 1 residues divided between two structures, Table 2). The remaining sialic acid residues (16 ± 1) are therefore distributed among the O-linked oligosaccharides. The mass of desialylated O-linked oligosaccharides was 6660 ± 68 Da with 16 ± 1 sialic acid residues.

Sulfated tyrosines were identified by MALDI-TOF by taking advantage of the fact that they fragment during ionization in the positive mode to give rise to an ion with mass consistent with that of the unmodified peptide (*25*). In the negative mode, the molecules fragment to a lesser degree to produce a pair of ions separated by 80.06 Da, corresponding to the sulfated and desulfated molecule. Phosphorylated peptides, on the other hand, produce single ions without fragmentation (*26*). However, the presence of phosphate is confirmed by the loss of mass after alkaline phosphatase treatment. Therefore, by acquiring positive and negative MALDI-TOF mass spectra before and after alkaline phosphatase treatment, both phosphate and sulfate modifications can be identified (*27*). This approach was used in the analysis of phosphate and sulfate substituents of BSP. A sample of BSP was deglycosylated by treatment with a mixture of *N*-glycanase, neuraminidase, and *O*-glycanase as described above and then digested with endoproteinase Lys-C. The resulting digest was analyzed by ESI-MS. Figure 2 shows the total ion current (upper trace) and the 214 nm absorbance (lower trace) for the HPLC chromatogram. The absorbance trace is characterized by several sharp peaks superimposed over a peak that elutes broadly between 30 and 50 min and produces no signal in the total ion current trace. An additional sharp peak eluting at 49 min in the absorbance trace also failed to produce a signal in the total ion trace. The eluate from the column was collected in fractions of volume 400 μL. Two fractions in particular, labeled A and B in Figure 2 and eluting between 25 and 29 min, yielded useful information which is described below.

Figure 3 shows the positive (a) and negative (b) mode MALDI-TOF mass spectra of fraction A from Figure 2. The negative mode spectrum shows a pair of ions at *m*/*z* 1901.7 ± 1.0 and 1981.0 ± 1.0. These masses are consistent with unmodified BSP peptide 286−301 (GQGYDGYDGQNYY-HHQ) and that bearing one sulfate group, calculated *m*/*z* values 1900.87 and 1980.93, respectively. The positive mode spectrum shows an ion corresponding to the desulfated peptide. The observed mass difference is 79.3 ± 2 Da (that expected for sulfation is 80.06 Da). Treatment of the same peptide with alkaline phosphatase resulted in no change to the mass spectra (data not shown), confirming the presence of a single sulfate in this sequence. This peptide is derived from the C-terminal region and contains a total of four tyrosine residues as potential sites of sulfation.

To determine the specific tyrosine sulfation site, fraction A was digested with aminopeptidase M (*28*) for 60 min before acquiring MALDI-TOF mass spectra. Figure 4 shows the negative mode MALDI-TOF mass spectrum acquired. The pair of ions labeled **A** corresponds to the sulfated and desulfated full-length Lys-C peptide. The ions labeled **B** correspond to the truncated peptide arising from the loss of the N-terminal two residues and the desulfated partner. The ion-pairs labeled **C** and **D** differ from each other in the loss of a single tyrosine but with the fragmentation pattern still evident, indicating that the loss of this tyrosine did not cause a loss of sulfate. Similarly the ion pair labeled **E** corresponds

Table 4: Mass Data from the Aminopeptidase M Digestion of Fraction **A**, Derived from the Lys-C Digest of Deglycosylated Human BSP

	MALDI-TOF deprotonated molecular ion mass				
	observed		calculated		
label	desulfated	sulfated	desulfated	sulfated	assignment
A	1903.0 ± 0.03	1981.1 ± 0.1	1900.87	1980.93	GQGYDGYDGQNYYHHQ
B	1719.6 ± 0.2	1796.3 ± 0.3	1715.69	1795.75	GYDGYDGQNYYHHQ
C	1658.9 ± 0.6	1738.6 ± 0.6	1658.64	1738.70	YDGYDGQNYYHHQ
D	1495.7 ± 0.2	1575.5 ± 0.1	1495.46	1575.52	DGYDGQNYYHHQ
E	1159.9 ± 1.0	1240.6 ± 1.1	1160.15	1240.21	DGQNYYHHQ

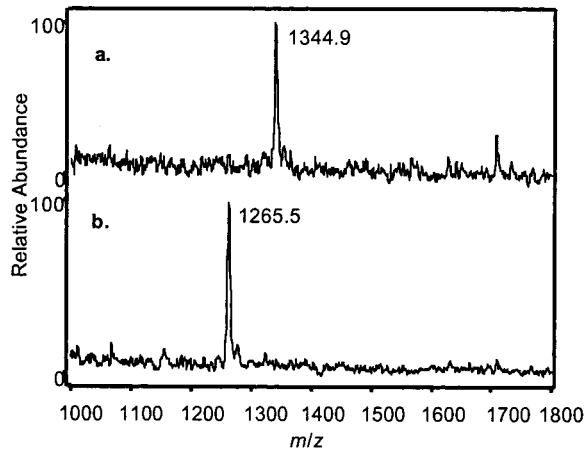

FIGURE 5: Negative mode MALDI-TOF mass spectra of fraction B from the HPLC fractionation of deglycosylated, endoproteinase lysine-C digested BSP (see Figure 2) before (a) and after (b) treatment with alkaline phosphatase.

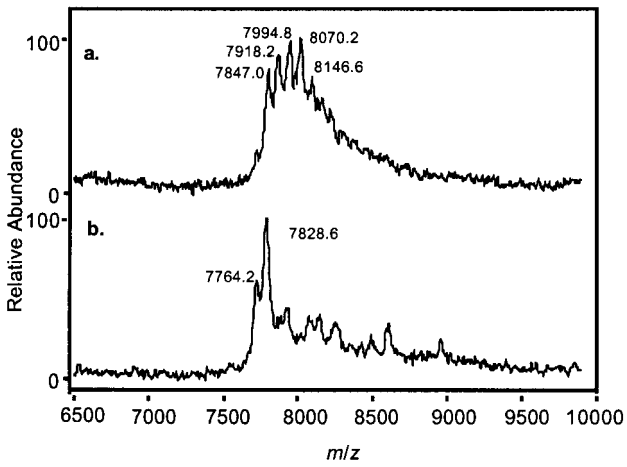

FIGURE 6: Positive mode MALDI-TOF mass spectra of the unfractionated tryptic digest of deglycosylated BSP before (a) and after (b) alkaline phosphatase treatment.

to the peptide DGQNYYHHQ, now with only two tyrosine residues but still carrying the sulfate. This indicates that either tyrosine 297 or tyrosine 298 is sulfated. The aminopeptidase M reaction did not proceed beyond this point, preventing a distinction between these tyrosines. The mass data obtained from this experiment are summarized in Table 4. Ions corresponding to desulfated peptides are produced by fragmentation during the ionization process. This fragmentation results, in some cases, in desulfated peptide mass values with greater than expected discrepancy between calculated and observed *m*/*z* values.

Figure 5 shows the negative mode MALDI-TOF mass spectrum of fraction B from the endoproteinase Lys-C digest of deglycosylated BSP (see Figure 2), before and after treatment with alkaline phosphatase. The *m*/*z* of this deprotonated peptide decreases from 1344.9 ± 0.9 (Figure 5a) to 1265.5 ± 0.3 (Figure 5b), a difference of 79.4 ± 1.2 Da that is consistent with a single phosphorylation site. This mass agrees well with the calculated mass of 1265.32 Da for peptide 12−22 (IEDSEENGVFK). This peptide has only one site of potential phosphorylation, Ser-15.

To locate other sites of phosphorylation, an unfractionated tryptic digest of deglycosylated BSP was used to acquire MALDI-TOF mass spectra before and after alkaline phosphatase treatment. In the 7500−9000 mass region of this spectrum, (Figure 6), a series of five ions was observed with *m*/*z* values of 7847.0 ± 3.4, 7918.2 ± 3.5, 7994.8 ± 3.1, 8070.2 ± 2.2, and 8146.6 ± 6.4. Treatment with alkaline phosphatase caused this series to collapse to an abundant single ion of *m*/*z* 7828.6 ± 2.9, with a shoulder at *m*/*z* 7764.2 ± 2. (Figure 6b). The mass data are shown in Table 5 together with the BSP sequence assignments. The five ions identified in Figure 6a all have masses consistent with that of BSP peptide 130−203 with 0−4 phosphate groups attached. The mass of the abundant ion produced after alkaline phosphatase treatment matches the calculated *m*/*z* of 7830.43 for BSP peptide 130−203. The shoulder of *m*/*z* 7764.2 in Figure 6b probably corresponds to a modified form of peptide 130−203, but the nature of the modification is unknown.

When an unfractionated tryptic digest of BSP, treated with neuraminidase and *O*-glycanase together, was analyzed by MALDI-TOF in the 7000−16000 mass range (Figure 7a), a broad polydisperse peak with little resolution of individual components was obtained. The peak was, however, bimodal, and the centroid *m*/*z* values were 11 133 ± 26 and 12 975 ± 101, respectively. Another tryptic digest, now treated with neuraminidase and *N*-glycanase, gave a spectrum (Figure 7b) consisting of two clusters of signals, the first one between *m*/*z* 8900 and 9200 and the second between 9500 and 12 000 (Table 5). In the first cluster, there is a series of four signals of *m*/*z* values 8939.6 ± 1.3, 9014.7 ± 2.2, 9091.2 ± 2.4, and 9164.8 ± 5.9, strongly suggesting a phosphorylated series. Complete deglycosylation of the tryptic digest by treatment with neuraminidase, *N*-glycanase, and *O*-glycanase sequentially was also carried out, and this digest gave the MALDI-TOF spectrum shown in Figure 7c. (This is the same spectrum shown in Figure 6a but at a broader mass range to allow comparisons between different series of spectra.)

Since the cluster of signals in Figure 7c represents a phosphorylated series of peptide 130−203 (see above), it is possible to extract information about the glycosylated series of fragments (Figure 7b) from which they are derived. It is useful to recall at this point that human BSP has four putative N-glycosylation sites at positions 88, 161, 166, and 174. The

Table 5: MALDI-TOF MS Data of BSP Tryptic Digests[a]

figure	measured mass	calcd mass	identification
6a (7c)	7847.0 ± 3.4	7844.2[b]	(130−203)[c] + 1 phos
deglycosylated	7918.2 ± 3.5	7924.2	(130−203)+ 2 phos
BSP	7994.8 ± 3.1	7990.37	130−203 + 2 phos
	8070.2 ± 2.2	8070.35	130−203 + 3 phos
	8146.6 ± 6.4	8150.33	130−203 + 4 phos
	av 7995.4 ± 3.7		
	8800 ± 50	8306.30	42−119 + modification[d]
6b	7764.2 ± 2.7		(130−203)
alk phos treated	7828.6 ± 2.9	7830.41	130−203
7a: BSP-treated with *O*-glycanase, neuraminidase	11133 ± 26		130−203 + 3138 ± 30 Da
	12975 ± 101		130−203 + 4980 ± 105 Da
7b: BSP-treated with *N*-glycanase, neuraminidase	8939.6 ± 1.3		130−203 + 1092.6 ± 4.7
	9014.7 ± 2.2		130−203 + 1096.5 ± 5.7
	9091.2 ± 2.4		130−203 + 1096.4 ± 5.5
	9164.8 ± 5.9		130−203 + 1094.6 ± 8.1
			av 1095.0 ± 6.0
	10488 ± 38		42−119 + 1688 ± 88

[a] All mass values are given for the protonated peptide molecular ion in daltons. The masses of desialylated N-linked oligosaccharides were obtained by subtracting values obtained from Figure 7c from those obtained in Figure 7a. The masses of desialylated O-linked oligosaccharides were obtained by subtracting values from Figure 7c from those from Figure 7b. [b] This mass was calculated by adding the mass of a phosphate group to the mass of 7764.2 from Figure 6b. [c] This peptide is a probable variant of peptide 130−203. [d] The group(s) modifying peptide 42−119 is (are) not known.

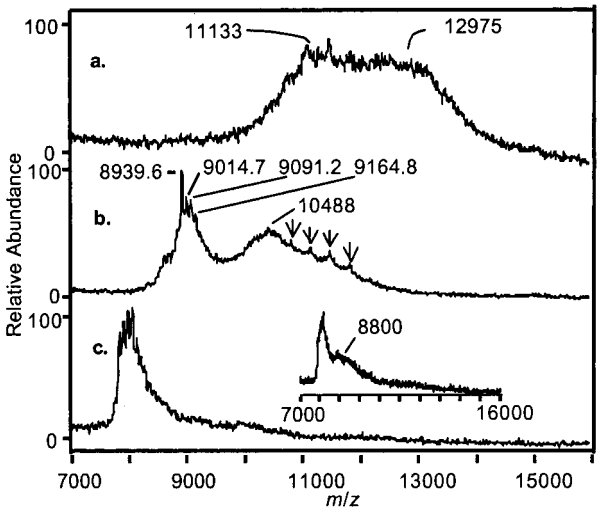

FIGURE 7: Comparison of the 7000−16 000 Da mass range of unfractionated BSP tryptic digests as a function of glycosylation state. The spectra show BSP treated with neuraminidase and *O*-glycanase (N-linked oligosaccharides still attached) (a); BSP treated with *N*-glycanase and neuraminidase (O-linked oligosaccharides still attached) (b); and fully deglycosylated BSP (c). The spectrum of the *O*-glycanase-treated digest obtained in the negative ionization mode is also shown (inset).

latter three represent a cluster of closely spaced sites, all of which are contained within peptide 130−203. The mass difference arising as a result of loss of N-oligosaccharides on this peptide is 3138 ± 30 Da. This represents a large portion of the total desialylated N-oligosaccharide mass of 4312 ± 1410 Da determined previously (Figure 1 and Table 2). The second species evident in Figure 7a with a centroid *m*/*z* of 12 975 ± 101 is a variably N-glycosylated form of peptide 130−203 that carries N-oligosaccharide with a mass of 4980 ± 105 Da. This mass matches to within error the measured mass of desialylated N-linked sugars from the whole BSP molecule (see above). Therefore, all of the N-glycosylation occurs within the cluster of sites on peptide 130−203, and no substitution occurs on the fourth site at position 88.

The cluster of peaks separated by 80 Da in Figure 7b decreases in mass by an average of 1095.0 ± 6.0 Da when treated with *O*-glycanase (Figure 7c), consistent with one or two O-linked sugars on peptide 130−203. Another observation may be made in light of the broad peak seen in Figure 7b with a centroid of *m*/*z* 10 488 ± 38. Since this species disappears upon *O*-glycanase treatment, it represents an O-linked oligosaccharide-substituted peptide. When the MALDI-TOF spectrum of the *O*-glycanase-treated digest was obtained in the negative ionization mode (see Figure 7, inset), an additional shoulder was seen with centroid *m*/*z* 8800 ± 50 Da. This is likely to be the second large tryptic peptide of human BSP (peptide 42−119) of calculated *m*/*z* 8306.30. In this case, the peptide carries substituents of mass approximately 500 Da that are insensitive to all of the enzymes used in this study. These substituents are likely to be O-linked sugars that resist *O*-glycanase and are consistent with the discrepancy between the deglycosylated BSP core protein peak at *m*/*z* 34 799 ± 35 observed in Figure 1e and the mass of the core protein (33 543.39 Da). The mass difference between the *m*/*z* 10 488 ± 38 species (Figure 7b) and the 8800 ± 50 species (Figure 7c) is 1688 ± 88 Da (Table 5). The total mass of desialylated O-linked oligosaccharides on peptides 42−119 and 130−203 is 2783 ± 94 Da, compared to the total desialylated O-linked mass of 6660 ± 68 Da for the entire BSP molecules (Table 2). The additional desialylated O-linked structures of mass 3877 Da are discussed below (see Discussion). The four ions in Figure 7b marked with arrows correspond to *m*/*z* 10 888, 11 229, 11 582, and 11 924, consistent with a series of glycosylated peptides differing by disaccharide units, such as found in a polylactosamine chain, modifying peptide 42−119.

MALDI-TOF spectra of an unfractionated endoproteinase Asp-N digest of BSP were acquired to confirm some of the findings from Figures 6 and 7. Figure 8a shows the negative mode spectrum in which five pairs of ions, numbered **1**−**5**, are separated by 80 Da, indicating phosphate or sulfate modification. In the positive mode spectrum acquired on the same sample (Figure 8b), pairs **1**−**4** collapse to a single ion, indicating that the peptides are sulfated. As the mass assignments in Table 6 show, pairs **1** and **2** correspond to the C-terminal peptides 293−301 [(Y)DGQNYYHHQ] and 290−301 [(Y)DGY...HHQ()], confirming the sulfation site identified in Figure 4. Pairs **3** and **4** correspond to peptides 256−271 [(Y)DNG...EPRG(D)] and 256−278 [(Y)DNG...AYE(D)], respectively, indicating that either tyrosine 259 or tyrosine 262 is sulfated. Pair **5** appears the same in Figure 8a and Figure 8b, and collapses to a single ion after alkaline phosphatase treatment in Figure 8c. This pair corresponds to peptide 122−153 [(G)DIT...EAE(D)] which is present as a mixture of phosphorylated and unphosphorylated forms.

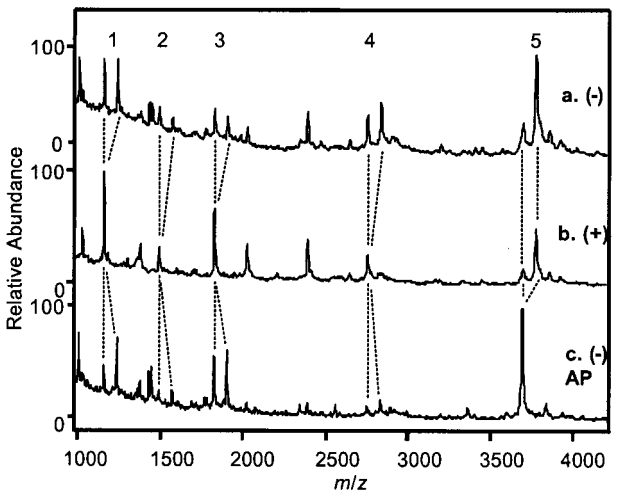

FIGURE 8: Negative mode (top), positive mode (middle), and negative mode after alkaline phosphatase treatment (bottom) MALDI-TOF mass spectra of the unfractionated endoproteinase Asp-N digest of deglycosylated BSP.

Table 6: MALDI-TOF Mass Spectra of an Unfractionated Asp-N Digest of Deglycosylated BSP[a]

	mass (Da)		assignment	
peak	measured	calculated	position	sequence
1	1161.8 ± 0.4	1160.15	293−301	DGQNYYHHQ
	1241.9 ± 0.5	1240.21	sulfated	
2	1498.0 ± 0.7	1495.46	290−301	DGYDGQNYYHHQ
	1578.6 ± 1.2	1575.52	sulfated	
3	1832.5 ± 0.5	1827.81	256−271	DNGYEIYESENGE-PRG
	1911.5 ± 0.1	1907.87	sulfated	
4	2763.2 ± 0.4	2739.74	256−278	DNGYEIYESENG-EPRGDNYRAYE
	2843.5 ± 0.2	2819.80	sulfated	
5	3699.5 ± 1.5	3700.61	122−153	DITNKATKEKES-DEEEEEEEEGNE-NEESEAEV
	3780.4 ± 1.9	3780.59	phosphoryl-ated	

[a] Mass assignments refer to Figure 8a.

Modification of this sequence accounts for one of the phosphorylations observed in Figure 6.

DISCUSSION

The BSP molecular mass measured by MALDI-TOF mass spectrometry was 52 533 ± 388 Da (Table 2), in good agreement with the value of 57 300 Da determined for bovine BSP by equilibrium centrifugation (*1*). Estimates based on electrophoretic mobility were substantially higher, in the range of 70−82 kDa (*10*, *29*, *30*). Both techniques show the heterogeneity of the BSP structure caused by heterogeneous posttranslational modification. From the mass shift caused by enzymatic deglycosylation in conjunction with monosaccharide analysis (Figure 1 and Table 3), it can be seen that two N-linked oligosaccharides modify BSP. Human BSP has four consensus sequences corresponding to putative sites for N-glycosylation (residues 88, 161, 166, 174), three of which are contained in tryptic peptide 130−203. Measurement of the molecular weight of this peptide as a function of glycosylation state (Figure 7 and Table 3) shows that all of the N-linked oligosaccharide mass calculated for the intact protein (Table 2) is found on this peptide, indicating the absence of substitution at position 88. The 7 ± 1 mol of N-linked sialic acid is consistent with one tri- and one tetraantennary N-linked oligosaccharide, assuming one sialic acid per nonreducing terminus. Reasonable compositions for these structures are $(GlcNAc)_5(Man)_3(Gal)_3(NeuAc)_3$ (calculated residue mass 2862.60 Da) and $(GlcNAc)_6(Man)_3(Gal)_4(NeuAc)_4$ (calculated residue mass 3519.20 Da), with a total mass of 6381.80 Da, agreeing with the total N-linked oligosaccharide mass of 6456 ± 525 Da. Since Asn-174 is found only on human BSP and is absent in other known mammalian BSP sequences, Asn-161 and -166 are the most likely sites of glycosylation. While sulfation of these glycans is possible, the mass accuracy of these measurements is not sufficient to propose a sulfation pattern.

Subtracting the composition of N-linked oligosaccharides from the total monosaccharide composition in Table 3, the O-linked oligosaccharides have a composition of $(GalNAc)_{16}(GlcNAc)_4(Gal)_{15}(Glc)_6(NeuAc)_{16}$ with a calculated mass of 12 129.01 Da. This values agrees with the mass of oligosaccharide determined from the monosaccharide analysis (12 020 ± 2300 Da). Monosaccharide analysis (Table 3) indicates a ratio of 16.0 ± 2.3 mol of galactosamine per mole of protein, a value which may define the number of O-linked oligosaccharides per mole of protein. The data are consistent with (a) seven O-linked oligosaccharides of composition (GalNAc)(Gal)(NeuAc), (b) six of (GalNAc)(Glc)(NeuAc), (c) one of $(GalNAc)(Gal)_2(NeuAc)$, and (d) two of $(GalNAc)(GlcNAc)_2(Gal)_3(NeuAc)$. Compositions (a), (b), and (d) were found in a search of known O-linked oligosaccharide structures (*31*). The mass of the O-linked oligosaccharide for peptide 130−203 (an average of 1095.0 ± 6.4 Da, calculated from Table 5) is consistent with the latter composition (calculated mass 1096.01 Da, correcting for neuraminidase treatment). The peaks for peptide 130−203 modified with this oligosaccharide as well as phosphate groups (Figure 7b) are comparatively sharp, indicating that there is a low degree of heterogeneity. A computer algorithm using a database of the activity of *N*-acetylgalactosaminyltranferases on mammalian mucins (*32*) predicts a single O-glycosylation site in BSP peptide 130−203, at Thr-165.

The mass of the O-linked oligosaccharide for peptide 42−119 (1688 ± 88 Da, Table 5) is consistent with modification by compositions (c), $(GalNAc)(Gal)_2(NeuAc)$, and (d), $(GalNAc)(GlcNAc)_2(Gal)_3(NeuAc)$ (calculated total mass 1623.49 Da, correcting for neuraminidase treatment). Composition (d) matches one structure only in the O-linked oligosaccharide database, that of a human milk mucin O-linked polylactosaminoglycan (*33*). It is therefore possible that the heterogeneity of the polylactosamine chain on this structure may account for the series of peaks separated by an average of 346 ± 23 Da marked with arrows in Figure 7b. This mass is consistent with a series of glycosylated peptides differing by disaccharide units, such as found in a polylactosamine chain. In this context, it is relevant to note that keratan sulfate has a polylactosamine repeat attached to a protein via *N*-acetylgalactosamine, similar to the presented structure. Keratan sulfate substituents have been identified in BSP from rabbit (*34*). Thus, it appears that this polylactosamine may be sulfated in some species.

The *N*-acetylgalactosaminyltranferase-based computer algorithm predicts two modification sites in peptide 42−119, at Ser-92 and Thr-95. The corresponding peak in Figure 7b

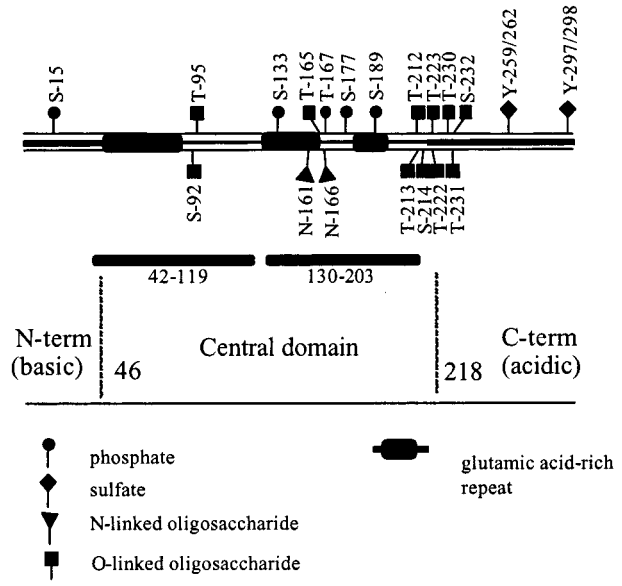

FIGURE 9: Summary of BSP posttranslational modifications. The N-terminal, central, and C-terminal domains are shown along with the glutamic acid-rich repeats and tryptic peptides 42−119 and 130−203.

is broad, indicating heterogeneity of these oligosaccharides. The algorithm also predicts a cluster of eight O-linked sugar modification sites at Thr-212, -213, -222, -223, -230, and -231 and Ser-214 and -232. Assuming these sites are filled by structures of types (a) and (b), this accounts for a total of 11 O-linked oligosaccharides on BSP.

Peptide 130−203 is also modified heterogeneously with one to four phosphate groups, as shown by the peaks separated by 80 Da in Figure 6. This peptide contains seven Ser and seven Thr residues in human BSP, each of which may be sites of modification. Three Ser residues (133, 177, and 189) and one Thr (167) are conserved among the known mammalian sequences and fit the sequence motifs for casein kinase II or protein kinase C (*12*) and are therefore the most likely sites for phosphorylation. The role of these phosphorylation sites in hydroxyapatite binding is an interesting topic. A recent study showed that BSP treated with alkaline phosphatase still bound to hydroxyapatite crystals (*17*). The researchers were not able to determine the extent to which some phosphorylated residues were protected from the enzyme, and thus the question remains open.

HPLC fractionation of an endoproteinase Lys-C digest of deglycosylated BSP with subsequent analysis by ESI and MALDI-TOF MS established that Ser-15 is phosphorylated (Figure 5). Notably, this residue is conserved among the human, bovine, and rat sequences (*12*) and substituted with a Pro residue in the porcine sequence (*13*). MALDI-TOF MS in conjunction with aminopeptidase ladder sequencing established that Tyr-297 or -298 is sulfated (Figures 3 and 4). Sulfation of Tyr-259 or -262 is also present, as shown in Figure 8 and Table 5. These tyrosine residues are conserved among all mammalian BSPs sequenced to date and are located close to the RGD sequence.

The modifications identified on human BSP are summarized in Figure 9. The cluster of phosphate groups in the central domain clearly may influence the hydroxyapatite binding activity of two of the polyglutamic acid repeats. The heterogeneity of phosphorylation may have regulatory roles for hydroxyapatite binding or RGD-dependent cell attachment or both. The polyglutamic acid repeats defined by residues 52−87 and 130−157 in the central domain have the highest α-helix-forming propensity in the BSP sequence (*35*). The latter predicted α-helix is proximal to the phosphorylation sites in the central domain, and the interaction between phosphatases and these elements is a key interest.

Two N-linked oligosaccharides are also present in this region, and these are also observed to be heterogeneous and may influence hydroxyapatite binding. The cluster of O-linked sugars in the region of Thr-213 to Ser-232 is indicative of a mucin-like domain (*19*). The two sulfation sites flank the RGD sequence and thus may modulate integrin binding activity or signaling upon integrin binding. The next phase in this work will be to assess the implications of these modifications as a function of developmental state, distribution, species, and disease state. Such studies will shed light on the role of phosphate groups associated with the polyglutamic acid repeats in HA crystallization, the role of the C-terminal sulfates in cell binding, and the roles of glycosylation in other aspects of BSP function.

REFERENCES

1. Franzén, A., and Heinegård, D. (1985) *Biochem. J. 232*, 715−724.
2. Oldberg, Å., Franzén, A., Heinegård, D., Pierschbacher, M., and Ruoslahti, E. (1988) *J. Biol. Chem. 263*, 19433−19435.
3. Reinholt, F., Hultenby, K., Oldberg, Å., and Heinegård, D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4473−4475.
4. Miyauchi, A., Alvarez, J., Greenfield, E. M., Teti, A., Grano, M., Colucci, S., Zambonin-Zallone, A., Ross, F. P., Teitelbaum, S. L., and Cheresh, D. (1991) *J. Biol. Chem. 266*, 20369−20374.
5. Flores, M., Norgård, M., Heinegård, D., Reinholt, F., and Andersson, G. (1992) *Exp. Cell Res. 201*, 526−530.
6. Nomura, S., Wills, A. J., Edwads, D. R., Heath, J. K., and Hogan, B. L. M. (1988) *J. Cell Biol. 106*, 441−450.
7. Chen, J., Shapiro, H. S., and Sodek, J. (1992) *J. Bone Miner. Res. 7*, 987−996.
8. Shen, Z., Heinegård, D., and Sommarin, Y. (1994) *Matrix Biol. 14*, 773−781.
9. Ecarot-Charrier, B., Bouchard, F., and Delloye, C. (1989) *J. Biol. Chem. 264*, 20049−20053.
10. Midura, R. J., McQuillan, D. J., Benham, K. J., Fisher, L. W., and Hascall, V. C. (1990) *J. Biol. Chem. 265*, 5285−5291.
11. Stubbs, J. T., Mintz, K. P., Eanes, E. D., Torchia, D. A., and Fisher, L. W. (1997) *J. Bone Miner. Res. 12*, 1210−1222.
12. Shapiro, H. S., Chen, J., Wrana, J. L., Zhang, Q., Blum, M., and Sodek, J. (1993) *Matrix 13*, 431−440.
13. Chenu, C., Ibaraki, K., Robey, P. G., Delmas, P. D., and Young, M. F. (1994) *J. Bone Miner. Res. 9*, 417−421.
14. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105−132.
15. Fisher, L. W., Torchia, D. A., Fohr, B., Young, M. F., and Fedarko, N. S. (2001) *Biochem. Biophys. Res. Commun. 280*, 460−465.
16. Hultenby, K., Reinholt, F., Norgård, M., Oldberg, Å., Wendel, M., and Heinegård, D. (1994) *Eur. J. Cell Biol. 63*, 230−239.
17. Hunter, G. K., and Goldberg, H. A. (1994) *Biochem. J. 302*, 175−179.
18. Ibaraki, K., Termine, J. D., Whitson, S. W., and Young, M. F. (1992) *J. Bone Miner. Res. 7*, 743−754.
19. Midura, R. J., and Hascall, V. C. (1996) *Glycobiology 6*, 677−681.
20. Apffel, A., Fischer, S., Goldberg, G., Goodley, P., and Kuhlmann, F. (1995) *J. Chromatogr., A 712*, 177−190.
21. Hardy, M. R. (1989) *Methods Enzymol. 179*, 76−82.

22. Mass spectrometers measure the mass-to-charge ratio (*m*/*z*) of ions rather than determine the mass of molecules directly. Therefore, mass-to-charge ratio will be used to describe ions from mass spectra. When the *m*/*z* value is converted to mass, daltons (Da) will be used.
23. Heinegård, D., and Oldberg, Å. (1989) *FASEB J. 3*, 2042−2051.
24. Mintz, K. P., Grzesik, W. J., Midura, R. J., Robey, P. G., Termine, J. D., and Fisher, L. W. (1993) *J. Bone Miner. Res. 8*, 985−995.
25. Talbo, G., and Roepstorff, P. (1993) *Rapid Commun. Mass Spectrom. 7*, 201−204.
26. This is true for linear MALDI-TOF mass spectrometers such as used here. Instruments equipped with an ion reflector will detect metastable fragmentation of phosphorylated peptides.
27. Annan, R. S., and Carr, S. A. (1996) *Anal. Chem. 68*, 3413−3421.
28. Woods, A. S., Huang, A. Y. C., Cotter, R. J., Pasternack, G. R., Pardoll, D. M., and Jaffee, E. M. (1995) *Anal. Biochem. 226*, 15−25.
29. Fisher, L. W., Whitson, S. W., Avioli, L. V., and Termine, J. D. (1983) *J. Biol. Chem. 258*, 12723−12727.
30. Elk-Rylander, B., Flores, M., Wendel, M., Heinegård, D., and Andersson, G. (1994) *J. Biol. Chem. 269*, 14853−14856.
31. Doubet, S., and Albersheim, P. (1992) *Glycobiology 2*, 505.
32. Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L., and Brunak, S. (1997) *Glycoconjugate J. 15*, 115−130.
33. Hanisch, F. G., Peter-Katalinic, J., Egge, H., Dabrowski, U., and Uhlenbruck, G. (1990) *Glycoconjugate J. 7*, 525−543.
34. Kinne, R. W., and Fisher, L. W. (1987) *J. Biol. Chem. 262*, 10206−10211.
35. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol. 120*, 97−120.
36. MALDI-TOF mass data obtained from human BSP with and without treatment with glycosidases. Three spectra were acquired for each sample, and the values reported are the average ± standard deviation.
37. All mass values are given for the protonated peptide molecular ion in daltons. The masses of desialylated N-linked oligosaccharides were obtained by subtracting values from Figure 7c from those obtained in Figure 7a. The masses of desialylated O-linked oligosaccharides were obtained by subtracting values from Figure 7c from those from Figure 7b.
38. Mass assignments refer to Figure 8a.

BI010887R