COMPLETE TOPOGRAPHICAL DISTRIBUTION OF BOTH THE IN VIVO AND IN VITRO PHOSPHORYLATION SITES OF BONE SIALOPROTEIN AND THEIR BIOLOGICAL IMPLICATIONS

Erdjan Salih* and Rudolf Flückiger
Laboratory for the Study of Skeletal Disorders and Rehabilitation, Department of Orthopedic Surgery, Harvard Medical School and Children’s Hospital, 300 Longwood Avenue, Boston, MA, O2115, U.S.A.

Running title: Variation of Extent of Phosphorylation at Each In Vivo Phosphorylation Site

*Correspondence: Dr Erdjan Salih
Department of Orthopedic Research
Children’s Hospital Boston,
300 Longwood Avenue,
Boston, MA, O2115, U.S.A.
Tel: (617) 247 5181
Fax: (617) 236 7149
E-mail: Erdjan. Salih @ TCH.Harvard. Edu

ABBREVIATIONS USED: [14C]CM-DTT, 1-S-mono-[14C]carboxymethyl dithiothreitol; MALDI-TOF-MS, matrix assisted laser desorption/ionization time of flight mass spectrometry; ATZ, anilinothiazoline; ATP, adenosine triphosphate; CKI, casein kinase I; DTT, dithiothreitol; CH3CN, acetonitrile; OPN, osteopontin; BSP, bone sialoprotein; ECM, extracellular matrix; Ca-P, calcium phosphate; P-Ser, phosphoserine; P-Thr, phosphothreonine; mCKII, microsomal casein kinase II; RP-HPLC, reverse-phase high performance liquid chromatography.
KEY WORDS: Bone sialoprotein, phosphorylation sites; protein kinases; biomineralization; MALDI-TOF-MS; radioactive thiol agent.
Bone sialoprotein (BSP) is a multifunctional highly phosphorylated and glycosylated protein with key roles in biomineralization and tissue remodeling. The present work identifies the complete topographical distribution and precise location of both the in vitro and in vivo phosphorylation sites of bovine BSP by a combination of state-of-the-art techniques and approaches. In vitro phosphorylation of native and deglycosylated BSP by casein kinase II identified seven phosphorylation sites by solid-phase N-terminal peptide sequencing that were within peptides with residues 12-22 (LES$^P$DEENGVFK), 42-62 (FAVQSSDSS$^P$EENGNG- DS$^P$S$^P$EE), 80-91 (EDS$^P$DENDEEES$^P$E) and 135-145 (EDES$^P$DEEEEEEE). The in vivo phosphorylation regions and sites were identified by use of a novel thiol reagent, l-S-[14$^C$]carboxymethyl-dithiothreitol ([14$^C$]CM-DTT). This approach identified all of the phosphopeptides defined by in vitro phosphorylation, but three additional phosphopeptides were defined with residues 135-145 (EDES$^P$DEEEEEEE), 250-264 (DNGYEIYES$^P$ENGDP$^P$) and 282-289 (GYDS$^P$YDGQ). Furthermore, use of native BSP and matrix-assisted laser desorption/ionization time-of-flight-mass spectrometry (MALDI-TOF-MS) identified several of the above peptides including an additional phosphopeptide with residues 125-130 (AGAT$^P$GK), which was not defined in either of the in vitro and in vivo studies above. Overall, seven in vitro and eleven in vivo phosphorylation sites were identified unequivocally, with natural variation in the quantitative extent of phosphorylation at each in vivo phosphorylation site.
Evidence supporting many facets of the observed biological functions of bone sialoprotein (BSP) and osteopontin (OPN) has been accumulating, with a wide range of implications in both mineralizing and non-mineralizing tissues. BSP and OPN are the major non-collagenous extracellular matrix (ECM) phosphoproteins of calcified tissues such as bone and cartilage. Their intimate relationship with biomineralization suggested that they play key roles during the development and maintenance of these tissues (1-8). Studies in vitro using BSP and OPN indicated that while BSP induces Ca-P apatite formation (4-6;9), OPN lacks this property or is inhibitory (3;4;10). The biological functions of BSP and OPN are not limited to mineral deposition but impact cellular behavior such as cell motility, cell adhesion and bone resorption (10-16). Despite extensive studies however, the precise roles of these phosphoproteins remain to be clearly defined. While some studies in vitro indicated that BSP promotes bone resorption and hence participates in bone degradation (17), other studies such as those using glucocorticoids, which increased expression of this protein, suggested its involvement in the anabolic phase of bone remodeling (18). In this laboratory in vivo implants of BSP-collagen composites in the calvarial critical defect bone repair model (19) and during reparative dentinogenesis (7;8;20-22) highlighted the impact of BSP during biomineralization and new bone/dentin formation.
As interest continues and evidence accumulates with regard to the general biological functions of bone phosphoproteins, in particular the covalently-bound phosphate groups, the biochemical factors or processes that can affect the state of phosphorylation of these proteins become significant. Hence, the protein kinases that are involved in the phosphorylation of BSP/OPN prior to secretion into the extracellular matrix (ECM) become important regulators of ECM phosphoprotein functions. Previous studies from this laboratory defined that the microsomal casein kinase type II (mCKII)/casein kinase II were the predominant enzymes involved during the phosphorylation of bone ECM phosphoproteins, OPN and BSP (23-26). Also, the use of a panel of purified protein kinases in \textit{in vitro} phosphorylation of purified bovine bone OPN and BSP demonstrated that the degree of \textit{in vivo} naturally occurring phosphorylation of bovine BSP and OPN were ~65% and ~85%, respectively. Not all of the molecules are phosphorylated on each of the potential phosphorylation sites within a given population of molecules isolated from bone of a particular age. Such heterogeneity in the phosphorylation can be attained by changes in factors such as the rate of phosphoprotein synthesis, the mCKII activity, the available ATP concentrations (within the intracellular compartment where phosphorylation takes place), and some degree of dephosphorylation during their residence in the ECM. The precise sites of phosphorylation on chicken bone OPN clearly defined that the extent of phosphorylation on each site varied between 30-100%, with an average of ~60% total phosphorylation (27). There were 10 phosphorylated sites identified, but the actual quantitative analysis showed 6.3 mols of phosphoamino-acids/mol of chicken OPN. The phosphorylated peptide regions were predominantly recognition sequences for CKII (27), confirming our \textit{in vitro} studies. Further studies which utilized the \textit{in vivo}
repair of calvarial bone defects induced to heal by implants of demineralized bone matrix highlighted
the intimate interrelationship between accumulation and rate of accumulation of Ca-P, OPN, BSP and
mCKII activity as a function of bone development (28). More recently direct evidence was provided
for the first time in support of the concept of a “natural variation” in the extent of phosphorylation of
ECM phosphoproteins during mineralized tissue formation. The extent of phosphorylation of OPN
and BSP in the two anatomically distinct in vivo sites (bony and soft tissue) was substantially different
(varied between ~1 to ~14 mols of P-Ser/mol of OPN or BSP), and varied as a function of time within
both implant sites (29). Furthermore, there was a direct and linear relationship between the “rate” of
Ca-P deposition and the ratio of P-Ser-BSP/P-Ser-OPN for calvarial implants, which clearly
demonstrated hidden and important facets of the factors that control mineral deposition, and the
intimate coupling of the state of phosphorylation of BSP and OPN in this process.

Indeed, the covalently-bound phosphates of BSP and OPN have been also shown to effect
osteoclast cell attachment properties (15;30). More recent conflicting studies have utilized different
forms of OPN and BSP which suggested that while phosphorylation was not important for cell
attachment, it was however, necessary for in vitro osteoclast resorption pit formation in dead bone
slices (31). The differences in the results between these studies, (15;30) vs. (31), were suggested to be
due to the differences in the protein samples and state of phosphorylation. There is no doubt that such
conflicting results will probably continue to arise since these proteins are highly phosphorylated/
glycosylated and as noted above these moieties are subject to inherent “natural variation” (29).
Phosphorylation does not take place to the same extent on each of the potential sites, which leads to
heterogeneity within a given population of the protein isolated from a natural source. *In vitro* phosphorylation can also produce proteins with heterogeneous phosphorylation states or fully phosphorylated forms, but in either case these may not represent precisely the naturally occurring forms.

The phosphorylation sites and regions of chicken bone OPN (27) and bovine milk OPN (32) and the glycosylation sites of human BSP (33;34) have been reported. However, similar details are not available for BSP phosphorylation sites. In the present study a combination of state-of-the-art experimental approaches have been utilized to define for the first time the complete topographical distribution of both the *in vitro* and *in vivo* phosphorylation sites of bovine bone BSP. These included: (i) phosphorylation of native and deglycosylated BSP by CKII, tryptic peptide mapping followed by solid-phase N-terminal peptide sequence analyses to define *in vitro* phosphorylation sites; (ii) use of a novel synthesized radiolabeled thiol agent (\(^ {14} \text{C-CM-DTT} \)) to derivatize phosphoserine tryptic peptides of BSP followed by normal N-terminal peptide sequence analysis to define *in vivo* phosphorylation sites; and (iii) use of MALDI-TOF-MS to rapidly profile BSP proteolytic peptides.
MATERIALS AND METHODS

**In vitro phosphorylation of native BSP by CKII using $^{32}$P[ATP] and Identification of the sites of phosphorylation by solid phase N-terminal peptide sequencing** – Bovine bone BSP was isolated and purified as previously described (26). 200 µg of native BSP was phosphorylated by $^{32}$P-ATP (diluted with cold ATP to give specific activity 250 mCi/mmol) using CKII (200 ng, Upstate Biotechnology, Inc., Waltham, MA) in 0.5 ml of KH$_2$PO$_4$/Na$_2$HPO$_4$ (0.1 M) buffer, pH, 7.4, containing 5 mM MgCl$_2$ and 1 mM EGTA for 1 hr at room temperature (~22°C). The $^{32}$P-labeled native BSP was dialyzed against 50 mM NH$_4$HCO$_3$, pH ~8.0, and digested with trypsin (2% w/w, TPCK treated bovine trypsin, Sigma Co.). The sample was freeze dried, suspended in 0.2 ml of H$_2$O, 0.1% trifluoroacetic acid (v/v) and subjected to RP-HPLC on a Vydac C-18 column (25 x 0.46 cm). The peptides were eluted by a linear gradient from H$_2$O, 0.1% trifluoroacetic acid (v/v) to 60% CH$_3$CN, 0.55% trifluoroacetic acid (v/v) over 90 min at a flow rate of 0.5 ml/min. The absorbance at 219 nm was recorded continuously and fractions of 0.5 ml were collected. Aliquots from each fraction were counted for $^{32}$P radioactivity and the radioactive fractions were then separately pooled for each peak, freeze dried, and rechromatographed. Each purified $^{32}$P-labeled peptide was then sequenced by automated solid-phase amino acid sequencing both to define the sequence and to identify the precise site(s) of phosphorylation, as described previously (27).

**Deglycosylation of bovine bone BSP followed by in vitro phosphorylation using CKII and $[^{32}$P]ATP**
Identification of the sites of phosphorylation – Native bovine bone BSP was deglycosylated by a combination of glycosidases simultaneously. 100 µg of BSP was incubated in 0.5 ml of 20 mM sodium phosphate buffer, pH 7.2, in the presence of O-glycosidase (2.5 mU/10 µg protein), N-glycosidase (0.4 U/10 µg protein) and neuraminidase (2 mU/10 µg protein) (Boehringer Mannheim Biochemica) overnight at 37°C. The deglycosylated BSP was then isolated by RP-HPLC using Vydac C-4 column (25 x 0.46 cm), fractions containing BSP (defined by Western blotting using bovine BSP polyclonal antibody (28;29) were pooled and freeze dried. The sample was then phosphorylated using CKII and [32P]ATP, trypsin digested and peptides separated by RP-HPLC, as described above. The 32P-labeled peptides were subjected to solid-phase N-terminal sequence analyses to define simultaneously the peptide sequence identity and the location of the phosphorylated residue(s), as described previously (27). To evaluate quantitatively, analysis were performed to calculate accurately the degree of absolute phosphorylation in vivo and in vitro at each site and relative to the others within the whole molecule. Since there is an unknown amount of phosphate loss from the P-Ser residues under Edman degradation cycles, it is not possible to quantify P-Ser content from release of 32P alone at that cycle. One way that this was overcome was to calculate the pmols of 32P-Ser released (from 32P counts and the specific activity of the 32P-ATP) plus the pmols of dehydroalanine observed at that cycle. This approximated more closely the actual original P-Ser content of the peptide, since generation of the dehydroalanine from the P-Ser is cumulative during Edman degradation.
Identification of the in vivo phosphorylation regions and sites of BSP by use of a novel synthesized radioactive thiol agent, 1-S-mono-[\(^{14}\text{C}\)]carboxymethyl dithiothreitol (\([^{14}\text{C}]\text{CM-DTT}\)) – To overcome the lack of an efficient and rapid determination of the precise sites of in vivo phosphorylation, a novel radioactive thiol reagent, \([^{14}\text{C}]\text{CM-DTT}\), was designed and synthesized in this laboratory. The synthesis and the utility of such reagent(s) have been recently described in detail for studies related to phosphoserine and phosphothreonine containing proteins using N-terminal peptide sequencing and MALDI-TOF-MS (Applied Biosystems, Inc.) (35;36).

100 µg of bovine bone BSP was deglycosylated and trypsin digested for chemical derivatization by \([^{14}\text{C}]\text{CM-DTT}\). The tryptic peptides were then incubated with 5 mM \([^{14}\text{C}]\text{CM-DTT}\) in 0.33 M NaOH for 1 hr at 50°C to derivatize the phosphoserine containing peptides. The \(^{14}\text{C}\)-labeled peptides were isolated/purified by RP-HPLC and N-terminal sequenced under normal sequencing conditions. N-terminal sequencing was carried out by Edman degradation (37) using the automated protein sequenator model 477A (Applied Biosystems Inc., Foster City, CA) and Biobrene-treated glass filters (27). To define the peptide sequence and the location of the phosphoserine/phosphothreonine containing residues, one third of the ATZ products of each Edman degradation step were converted to phenylthiohydantoin-derivatives and analyzed by on line HPLC (model 120A, Applied Biosystems), and two thirds were collected as ATZ-derivatives for \(^{14}\text{C}\)-counting as described recently (27, 36). Unlike the analysis and quantification of the in vitro \(^{32}\text{P}\)-labeled peptides by solid-phase sequencing above, the quantification of the P-Ser content during
sequencing of the $[^{14}\text{C}]$CM-DTT derivatized peptides was less complicated. This is because the $[^{14}\text{C}]$CM-DTT-Ser moieties during sequence analysis are stable and the released $^{14}\text{C}$ at that cycle reflects the actual original P-Ser content. To quantify yields of sequenced peptides, initial yield ($I_0$) and repetitive yield ($R$) were calculated by linear regression analysis of the observed yield ($M$) at each cycle ($n$): $\log_{10}(M) = n \log_{10}(R) + \log_{10}(I_0)$. The information obtained from such analysis was used in conjunction with the release of $^{14}\text{C}$-radioactivity to quantify the original phosphoserine content per mol of the peptide. The expected theoretical observable yield $M$ for serine quantity at a given cycle was calculated based on the cycle number $n$, initial yield $I_0$ and repetitive yield $R$. For example, the tryptic peptide containing residues 42-122 (FAVQSSSDSSEENGNGD------) was sequenced for 12 cycles with release of $^{14}\text{C}$-radioactivity at cycles 5 and 10, 2020 dpm and 780 dpm, respectively. These counts were for the two thirds of the total Edman degradation products of each cycle, whereas the phenylthiohydantoin-derivatives for the amino acid analyses were for the one third. Hence total expected observable $^{14}\text{C}$-counts if all of the Edman degradation products were counted are 3030 dpm for cycle 5 and 1170 dpm for cycle 10. The expected observable serine content at cycles 5 and 10 if all of the Edman degradation products were converted to phenylthiohydantoin and analyzed are 222 pmols and 132 pmols, respectively. Using the specific activity of $[^{14}\text{C}]$CM-DTT 10 mCi/mmol, 3030 dpm at cycle 5 corresponds to 137 pmols of $[^{14}\text{C}]$CM-DTT incorporated, reflecting 0.62 mols of $^{14}\text{C}$ incorporated/mol of serine, i.e. 62% of this serine was in the form of phosphoserine.
A similar calculation for the serine at position 10 indicated 0.40 mols of $^{14}$C incorporated/mol of serine, i.e. 40% of this serine was phosphorylated.

**Direct identification of native in vivo phosphopeptides by MALDI-TOF-MS without $^{32}$P labeling or chemical modification by $[^{14}C]CM$-DTT** – Native BSP, 5 µg, was trypsin digested and the peptide mixture was subjected directly to peptide profiling using MALDI-TOF-MS. The advantages of this approach are the ability to rapidly peptide profile, pinpoint naturally occurring phosphopeptides, and to sequence and identify each peptide within the mixture without any purification steps. In addition, due to the sensitivity of the instrument a very small amount of the material is required. MALDI-TOF-MS was performed in this laboratory using the Voyager DE Pro™ mass spectrometer (Applied Biosystems Inc.). Aliquots of BSP tryptic peptides (~1 pmol) in 1µl of H$_2$O/0.1% TFA were mixed with 1µl of matrix solution, α-cyano-4-hydroxycinnamic acid (CHCA, Applied Biosystems Inc.), spotted on to a sample target (stainless steel sample plate), allowed to dry and loaded into the mass spectrometer for analysis.

**RESULTS**

**Complete in vitro phosphorylation sites of purified bovine bone BSP using $^{32}$P-ATP and CKII-**

Figure 1A shows tryptic peptide map of native BSP phosphorylated using $^{32}$P-ATP and casein kinase II (CKII). The BSP used in this experiment was native with both glycosylation and phosphorylation.
intact. Under these circumstances $^{32}$P-labeling occurs at residual phosphorylation sites with recognition sequences for CKII. A number of $^{32}$P-labeled peptides were observed within the RP-HPLC profile using a C-18 column. Several of the peptides were subjected to rechromatography using RP-HPLC for clarification of the sequences, Fig. 1A inset. It is worth noting that the absorbance profile of each peak is heightened with the descending limb never declining to baseline, which is obvious both in Fig. 1A and its inset. This phenomenon is a prominent occurrence in the analysis of phosphoproteins with high level of glycosylation, as is the case with BSP in the present work and that previously reported for OPN (27). In order both to simplify the tryptic peptide map and to remove the uncertainty of possible $^{32}$P-incorporation on glycosylated side chains, native BSP was first deglycosylated by a combination of glycosidases followed by phosphorylation and peptide mapping, Fig. 1B. The peptide map of the deglycosylated BSP clearly revealed simplification and sharpening of the absorbance peaks, including baseline approximation of the peak limbs, observed for both Fig. 1B and its inset. Interestingly, however, neither the number of radioactive peaks nor their relative positions and intensities within each profile showed any significant change. This suggested that the peptide sequences and sites for phosphorylation and glycosylation do not overlap, and was supported by the quantitative data relating mols of $^{32}$P-incorporation/mol of native and deglycosylated BSP (~2.5 mol of phosphate/mol of protein in both cases). Sequence analysis of the phosphorylated peptide regions and sites provided direct evidence consistent with this. Each of the radiolabeled fractions from both Figs 1A and 1B were N-terminal sequenced using the solid-phase automated sequencing approach. Three peptides were identified as in vitro $^{32}$P-labeled peptides:
residues 12-22 (LEDS\textsuperscript{P}EENGVFK), 42-122 (FAVQSSDSS\textsuperscript{P}EENGNG DS\textsuperscript{P}S\textsuperscript{P}EEEEEE-
EETSNEEGNNGNED SP\textsuperscript{D}ENEDEE\textsuperscript{P}E-----), and 135-216 (EDES\textsuperscript{P}DEEEE-----) with
specific phosphorylated residues denoted by superscript P. The length of the tryptic peptide with
residues 42-122 was too long for the positions of all of the potential phosphorylated serines beyond
~12-15 residues to be defined clearly. Hence, a portion of this peptide was N-terminal sequenced up
to 12 residues to identify the phosphorylated residues within this section. The remaining peptide was
further proteolytically digested by Asp-N endopeptidase and the shorter radiolabeled peptides isolated
and sequenced. Overall, seven phosphorylated sites were defined within the three peptides. As
expected, all of the phosphorylated peptides contain recognition sequences for CKII (SEE, SXE, ESDE
etc), i.e. the phosphorylated serine residues are flanked with acidic amino acids (E or D). Figure 4
summarizes the topographical distribution of phosphorylation sites within the primary amino-acid
sequence of bovine BSP and Table 1 indicates the quantitative extent of phosphorylation (%
phosphorylation or mols of phosphate/mol of peptide).

*Complete in vivo phosphorylation sites of purified bovine BSP utilizing novel thiol agent [*\textsuperscript{14}C*]CM-DTT* - While defining the *in vitro* phosphorylated peptide regions and sites by a specific protein kinase
(CKII) was important, it was of further major interest to establish the *in vivo* phosphorylated peptides
of native bone BSP. Figure 1C shows the *\textsuperscript{14}C*-radiolabeled tryptic peptides of native deglycosylated
BSP derivatized by [*\textsuperscript{14}C*]CM-DTT. The deglycosylation step was performed prior to base-catalyzed
derivatization to avoid reaction of O-glycosyl sites with [*\textsuperscript{14}C*]CM-DTT, as under base catalysis such
moieties can undergo elimination reaction similar to P-Ser/P-Thr residues. N-terminal sequence analysis of each of the $^{14}$C-radiolabeled tryptic peptides identified five phospho-peptides with residues 12-22 (LEDS$^P$EENGVFK), 42-122 (FAVQS$^P$SSDSSPEENNGD$^P$SPEEEEEETSNEEGNNG GNED $^P$SPDENEDEES$^P$E---), 135-216 (EDES$^P$D----EIYES$^P$ENGDPR), and 272-285 (GYDS$^P$YDQDYSHQ), with the specific phosphorylated residues denoted by superscript $P$. The peptides with residues 42-122 and 235-264 were subjected to additional proteolytic digestion with Asp-N endopeptidase in order to define the precise location of the phosphorylated residue(s). The $[^{14}$C]CM-DTT derivatized peptides were sequenced under normal Edman degradation conditions using Biobrene treated glass filters. Figure 2 shows sequence data for the first 12 cycles of the peptide starting with residue Phe-42 obtained under such conditions. Overall, almost all of the peptides that were found to be phosphorylated \textit{in vitro} by CKII were also found to be \textit{in vivo} phosphorylated. The exceptions were two additional peptides with residues 235-264 (-----EIYES$^P$ENGDPR) and 272-285 (GYDS$^P$YDQDYSHQ), and an additional phosphorylation site at position 46 of the peptide with residues 42-122 were identified. These data indicate that most of the phosphorylation sites of native BSP were partially phosphorylated \textit{in vivo} and two sites on the N-terminal domain were either fully phosphorylated or somewhat not recognized or not accessible by CKII.

\textit{Identification of in vivo phosphorylation sites by direct MALDI-TOF-MS analysis} – Native BSP tryptic peptides were directly analyzed by MALDI-TOF-MS. The phosphorylated peptides were
directly identified from the presence of species with theoretical peptide mass + 80 Da (+1 phosphate group, or multiples of). For example Fig. 3A shows phosphopeptide with residues 12-22 (LEDS\textsuperscript{P}EENGVFK) where three different tryptic cleavage forms are clearly identified, species with a mass of 1347 mU (LEDS\textsuperscript{P}EENGVFK, 1267 Da + 80 Da), 1547 mU (AKLEDS\textsuperscript{P}EENGVFK, 1467 Da + 80 Da), and 1702 mU (RAKLEDS\textsuperscript{P}EENGVFK, 1622 Da + 80 Da). The original spectrum in Fig. 3A is obtained using the linear mode of the mass spectrometer which does not distinguish the naturally occurring isotopic masses of the peptide. The spectra shown as insets in Fig. 3A, however, were obtained in the reflector mode of the instrument where the spectra show distinct naturally occurring isotopic masses of the peptides, each separated by a unit mass (1 Da). Fig. 3B shows a phosphopeptide identified by MALDI-TOF-MS with a mass of 841 mU, residues 124-130 (KAGAT\textsuperscript{P}GKK, 761 Da + 80 Da). Interestingly this \textit{in vivo} phosphorylated peptide was not defined during the evaluation of either the \textit{in vitro} or the \textit{in vivo} phosphorylation sites described above. It is also the only peptide of BSP which is phosphorylated on a threonine residue (all the others are on serines), and the flanking amino acids indicates a recognition sequence for protein kinase C (AT\textsuperscript{P}XK). Hence, it is not surprising that this peptide was not identified during the analysis of \textit{in vitro} phosphorylated BSP by CKII. The \textsuperscript{14}C\textsuperscript{-}CM-DTT derivatization approach also did not identify this peptide most likely due to slower rate of conversion of P-Thr residue to its corresponding dehydrothreonine form with which \textsuperscript{14}C\textsuperscript{-}CM-DTT can react. The location of the phosphorylated serine/threonine residue in peptides such as the two discussed above can be easily defined since there is only one potential phosphorylation site.
(hydroxy-amino acid, Ser or Thr) in each. For peptides containing multiple possible serine/threonine phosphorylation sites, it would be necessary to sequence the peptide using the PSD/CID mode of the MALDI-TOF-MS to perform MS/MS in order to observe precisely the location of the specific phosphorylated residue.

**DISCUSSION**

The present study utilized a combination of multiple state-of-the-art approaches, techniques and reagents to define the complete topographical distribution of both *in vitro* and *in vivo* phosphorylation sites and regions of bovine bone BSP. These sites included the complete *in vitro* phosphorylation regions and specific sites (indicated by superscript P) of peptides with residues 12-22

(LES\(^\text{P}\)DEENGVFK), residues 42-62 (FAV QS\(\text{S}\)SDSS\(\text{P}\)EENGN- GDS\(\text{P}\)S\(\text{P}\)E), residues 80-91

(EDS\(\text{P}\)DENEDEES\(\text{P}\)E), and residues 135-141 (EDES\(\text{P}\)DEE). While these data in their own right are significant, there is no doubt that they may represent only part of the overall possible phosphorylation state/sites of native BSP. This is because *in vitro* phosphorylation was carried out using a single enzyme (CKII), and this process relies on the presence of partially phosphorylated peptides within the native population of BSP molecules. Hence, identification of the complete *in vivo* phosphorylation sites and regions of native bovine BSP was pursued. To accomplish this task, novel reagents developed in this laboratory were used. Thiol agents such as DTT and \([^{3}\text{H}]\)CM-DTT, used as tools to explore the general phenomenon of phosphorylation in conjunction with instrumentation to perform N-terminal sequencing and MALDI-TOF-MS, have been very recently described in detail from this laboratory.
These approaches led to the identification of \textit{in vivo} phosphorylation sites, which included all of the phosphopeptides identified by \textit{in vitro} CKII phosphorylation of BSP (peptides described above). However, three additional phosphopeptides were identified by $^{14}$C-radiolabeling: residues starting with Glu-135 (EDE$^{\text{P}}$DEEEEEE----), Asp-240 (DNGYE$^{\text{P}}$IYES$^{\text{P}}$ENGDP$^{\text{P}}$) and Gly-272 (GYD$^{\text{P}}$YDGQ----). Furthermore, the use of native BSP proteolytic peptides and MALDI-TOF-MS provided confirmation of several identified peptide regions by $^{32}$P-labeling (\textit{in vitro}) and $[^{14}$C]CM-DTT derivatization (\textit{in vivo}), but in addition the peptide with residues 125-130 (AGAT$^{\text{P}}$GK) was identified, which was not defined by either of the above approaches. This phosphopeptide, unlike the others, does not have a recognition sequence for phosphorylation by CKII. Instead, the amino-acids flanking the phosphothreonine constitute a recognition site (XT/SXK) for phosphorylation by PKC, and indeed is the reason why we did not define this site by using CKII in our \textit{in vitro} studies. Importantly, however, our previous \textit{in vitro} quantitative studies using a panel of protein kinases, did highlight the presence of a PKC phosphorylation site in bovine bone BSP (26), whose location now is defined herein. Also, this is the only phosphorylated BSP peptide where the site of phosphorylation is not a serine residue but rather threonine, and was not identified by the thiol agent approach because P-threonine residues undergo base-catalyzed elimination much slower than the P-serine residues. The identification of \textit{in vitro} phosphorylation sites (using $^{32}$P-ATP and CKII) of native bovine BSP were carried out with samples of native BSP with and without glycosylation. Such an approach was of importance for two reasons: (i) to simplify the tryptic peptide maps because presence of glycosylation leads to peak...
broading, and (ii) to eliminate the possibility of $^{32}$P-incorporation on the glycosyl site chains. Figure 1 shows the impact of deglycosylation on the peptide profile with clear peak sharpening. Similarly but for different reason, the identification of \textit{in vivo} phosphorylation sites by $[^{14}$C$]CM$-DTT derivatization were carried out using deglycosylated native BSP. This was an essential step since under base-catalyzed conditions the O-glycosyl sites can undergo elimination reaction similar to P-Ser/P-Thr residues that will generate unsaturated double bond with which $[^{14}$C$]CM$-DTT will react.

Overall, the present work provides the most extensive and detailed study to date of the precise location of 11 \textit{in vivo} and 7 \textit{in vitro} phosphorylation sites. Figure 4 summarizes the overall sites of phosphorylation of BSP both \textit{in vivo} and \textit{in vitro} as defined in the present work and the sites of glycosylation for human BSP, which have been recently reported (33). Interestingly, the glycosylation and phosphorylation sites do not overlap, and the glycosylation sites are concentrated in the mid-portion of the protein, somewhat sandwiched in between the N-terminal phosphorylation sites and those of the C-terminal phosphorylation sites. Clearly, these structural features raise a series of possible biologic functions for such posttranslational modifications and how they may couple to provide specific functional consequence. The role of BSP in biomineralization has been implicated by its temporal deposition into the ECM during development of the mineralizing tissues (1; 38-40), its high affinity for hydroxyapatite (41) and its action as a \textit{de novo} nucleator of hydroxyapatite crystals \textit{in vitro} (4). The poly-Glu regions have been suggested to be involved in hydroxyapatite binding and nucleation of Ca-P crystals (5). Similarly, its glycosylation has been also shown to influence
hydroxyapatite binding (33). Prior to these developments, the participation of the covalently-bound phosphates on ECM phosphoprotein-collagen complexes in Ca-P deposition was described (42). The complete topographical distribution and location of the phosphorylated regions/sites of BSP defined herein presents a plausible synergistic activity between these moieties both in hydroxyapatite binding and Ca-P crystal nucleation. Previous work from this laboratory established the quantitative extent of both in vitro and in vivo phosphorylation of BSP using a panel of purified protein kinases and $^{32}$P-ATP (26). In that study the total degree of phosphorylation was defined and the global domains that contained such modifications, including the predominant protein kinase (CKII) that participates during phosphorylation of BSP. However, the precise locations of the phosphorylation sites and peptide regions were not identified. The location of one phosphoserine in the N-terminal half of BSP, residues 12-22 (LESDEENGVFK), was identified by in vitro $^{32}$P-labeling with CKII and solid-phase N-terminal sequence analysis. A second phosphopeptide was identified consisting of residues 42-122 (FAVQSSDSSEENGNGDSSEE--) containing up to five phosphorylated residues, but the precise locations of these sites within the sequence were not defined at that time (25). The peptide with residues 12-22 was latter identified to be also phosphorylated in vivo (34, 36). In addition, a MALDI-TOF-MS approach was used to predict, based on mass difference, that a peptide with residues 130-203 in human bone BSP contained 0-4 phosphorylated P-Ser or P-Thr residues (34). However, because this peptide contains fourteen potential phosphorylation sites (seven Ser and seven Thr residues) on which the four phosphates may be residing, it was not possible to define the precise location of these phosphate groups. Thus, only one phosphorylation site was defined unequivocally
from previous studies to date (25, 34, 36), compared with eleven in vivo and seven in vitro sites in the present work.

To gain further insights into the biological implications of the phosphorylation state, we have performed extensive quantitative analysis on the extent of phosphorylation at each site both in vivo and in vitro. These data demonstrate the variation in the state of phosphorylation at each site (Table 1). The variable extent of phosphorylation most likely relates to the developmental stage, maturation and age of bone, which in turn reflects regulatory biological processes where at different stages of bone development BSP is synthesized and secreted with differing degrees of phosphorylation (29). Such a process can be the consequence of a combination of intracellular biological variables, including different protein expression levels of both the protein kinases and BSP prior to secretion into the ECM.

Direct experimental evidence for such an occurrence was provided in our previous in vivo studies (28), including most recently the differential extent of phosphorylation of BSP and OPN as a function of time during new bone formation in vivo (29). The coupling of the state of phosphorylation of BSP with that of OPN in regulating the rate of Ca-P deposition presents a unique biological interrelationship between these two proteins. Other studies have indicated that BSP promotes mineral deposition in cell cultures (9,43), and differentiation of bone marrow cells to osteoblasts (44). Further in vivo work using implants of native bone BSP-collagen composites in bone repair (19) and reparative dentinogenesis (7,8,20-22) elaborated on the multifunctional capacity of this protein. Clearly, these studies, combined with the results of the present work, evoke an interesting question as to what extent phosphate groups on BSP influence in vivo bone and dentin formation/repair.
In addition to extensive posttranslational modifications, BSP also contains the integrin receptor binding amino-acid sequence, RGD, which enables interactions with bone cells such as osteoclasts through their \( \alpha_v \beta_3 \) integrin receptor (15,16,45-48). Osteoclasts attach to the extracellular matrix through (particularly, the \( \alpha_v \beta_3 \)) integrin cell receptors (12,49) which mediate not only adhesion, but also signal transduction, and hence the regulation of osteoclast function. Studies with BSP and its non-RGD fragments (46-48) suggest that sequences other than RGD are also involved in cell attachment and modulating osteoclast cell behavior. For example, certain non-RGD sequences from BSP mediate a nuclear calcium response in osteoclasts, first observed with RGD peptides (50). Indeed, in vitro studies indicated that BSP can stimulate bone resorption, and in cell co-cultures inhibits osteoclast formation (51), whereas osteoclast binding can be influenced by the dephosphorylated form of BSP (15). More recently, it was suggested that BSP in its posttranslationally modified form was necessary for in vitro pit formation by osteoclasts seeded on dead bone slices (31). The precise mechanisms of how BSP influences various biological processes in these and other studies including the effects of phosphorylation state still remain to be clarified. Our data from the present study provide a number of intriguing fundamental structural properties with respect to phosphorylation of BSP that open new avenues to study and address various observed biological phenomena.

In conclusion, our extensive chemical and biochemical analyses of BSP defined complete in vivo and in vitro phosphorylation sites unequivocally, which revealed a series of critical information including quantitative natural variation at each phosphorylation site. These results are of major interest to a wide range of biological studies concerning the role of ECM phosphoproteins in biomineralization.
and the influence of the state of phosphorylation on cellular activity/behavior both in normal and pathological developments. Thus, the overall results have relevance to broad areas of biological sciences such as pathological soft tissue calcification, i.e. atherosclerosis, kidney stones, breast and prostate tumors, and bone metastasis in addition to general bone biology.

ACKNOWLEDGEMENTS

This work was supported by NIH grant RO1 AG17969 (E.S.). We thank Dr. James Clifton for his careful proofreading of the manuscript.
REFERENCES


FIGURE LEGENDS

Figure 1. Isolation and purification of tryptic peptides of in vivo and in vitro phosphorylated peptides of bovine BSP.  

A: 200 µg of purified native bovine bone BSP was $^{32}$P-labeled using $^{32}$P-ATP and CKII followed by trypsin digestion. The peptides were separated by reverse-phase HPLC on a Vydac C-18 column (25 x 0.46 cm) using a linear gradient from H$_2$O, 0.1% trifluoroacetic acid (v/v) to 60% CH$_3$CN, 0.055% trifluoroacetic acid (v/v) in 90 min, at a flow rate 0.5 ml/min. Absorbance at 219 nm was recorded continuously (solid line). Fractions of 0.5 ml were collected and aliquots counted for $^{32}$P radioactivity (dashed line). Inset: rechromatography of fraction 3, using same HPLC conditions as above.  

B: 100 µg of deglycosylated BSP was $^{32}$P-labeled using $^{32}$P-ATP and CKII followed by trypsin digestion. The peptides were separated by reverse-phase HPLC on a Vydac C-18 column as described for (A) except the linear gradient was performed over a 60 min, at a flow rate 1.0 ml/min and 1.0 ml fractions were collected. Solid line is the absorbance 219 nm and dashed line $^{32}$P radioactivity. Inset: rechromatography of fraction 3. Under the same HPLC conditions.  

C: 100 µg of deglycosylated BSP was first trypsin digested and the phosphoserine peptides $^{14}$C-labeled using 5 mM [${}^{14}$C]CM-DTT in 0.33 M NaOH at 50°C for 1 hr. The peptides were separated by reverse-phase HPLC on a Vydac C-18 column as described for (B).

Figure 2. N-terminal sequence analysis of purified $^{14}$C-labeled tryptic phosphopeptide of native BSP.  

During N-terminal sequencing one third of Edman degradation products was analyzed as PTH-amino
acid and two thirds were collected and counted for $^{14}$C radioactivity. 2020 dpm and 780 dpm were recovered at cycles 5 and 10 respectively. Initial yield, $I = 250$ pmol; repetitive yield, $R = 90\%$; $= \log_{10}$ (pmol of amino acid) at each cycle.

**Figure 3. Identification by MALDI-TOF-MS of tryptic phosphopeptides derived from native bovine bone BSP.** A: shows three different tryptic cleavage forms of the same phosphopeptide region with sequences, residues 12-22 ($\text{LEDS}^P\text{EENGVFK}$, with 1347.5 mU, 1247Da + 80 Da), residues 10-22 ($\text{AKLEDS}^P\text{EENGVFK}$, with 1546.7 mU, 1467 Da + 80 Da), and residues 9-22 ($\text{RAKLEDS}^P\text{EENGVFK}$, with 1702 mU, 1622 Da + 80 Da); B: a phosphopeptide with residues 125-130 ($\text{AGAT}^P\text{G}$, with 841.2 mU, 741 Da + 80 Da). The observed molecular mass for each peptide includes 80 Da which represents one phosphate group.

**Figure 4. Complete topographical distribution of both in vivo and in vitro phosphorylation regions/sites of bovine bone BSP.** The sites of phosphorylation were all determined in this laboratory, where $^{32}$P = in vitro phosphorylation sites identified by solid-phase N-terminal sequence analysis; $^{14}$C = in vivo phosphorylation sites identified using base catalyzed derivatization by $^{14}$C-CM-DTT and normal N-terminal sequence analysis; and MS = in vivo phosphorylation sites identified by MALDI-TOF-MS method. The complete primary amino acid sequence was from cDNA for bovine BSP (52).

• = glycosylation sites recently reported (33).
Table 1: Quantitative extent of phosphorylation of bovine bone BSP peptide regions/sites both *in vivo* and *in vitro*
defined by [14C]-CM-DTT and 32P-ATP labeling and MALDI-TOF-MS.

<table>
<thead>
<tr>
<th>Residue #</th>
<th>Sequence *</th>
<th>Moles phosphate/mol peptide incorporated <em>in vitro</em> by CKII</th>
<th>Mols phosphate/ Mol peptide <em>in vivo</em> identified by [14C]-CM-DTT</th>
<th>In vivo phosphorylation identified by MALDI-TOF-MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-22</td>
<td>LEDS*EENGVFK</td>
<td>0.37</td>
<td>0.65</td>
<td>LEDS*EENGVFK</td>
</tr>
<tr>
<td>45-62</td>
<td>QS<em>SSDSS</em>EENNGDS<em>S</em>EE</td>
<td>0.0, 0.6, 0.21,0.21=1.02</td>
<td>0.62, 0.4, 0.8,0.8=2.62</td>
<td></td>
</tr>
<tr>
<td>66-74</td>
<td>EEETS*NEEG</td>
<td>0.18</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>80-94</td>
<td>EDSDENEDDEES*EAEN</td>
<td>0.3</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>125-130</td>
<td>AGAT*GK</td>
<td>--</td>
<td>1.0</td>
<td>AGAT*GK</td>
</tr>
<tr>
<td>135-143</td>
<td>EDES*DEEEE</td>
<td>0.25</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>243-256</td>
<td>YEIYES*ENGDPRGD</td>
<td>0.61</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>270-280</td>
<td>GRGYDS*YDGQD</td>
<td>0.19</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>2.92</td>
<td>7.8</td>
<td></td>
</tr>
</tbody>
</table>

Note superscript asterisks (*) following the amino acid denotes phosphorylated residues.
Figure 3
Figure 4