Biochemical and Biophysical Research Communications **280,** 460–465 (2001)
doi:10.1006/bbrc.2000.4146, available online at http://www.idealibrary.com on IDEAL®

Flexible Structures of SIBLING Proteins, Bone Sialoprotein, and Osteopontin

L. W. Fisher,*[,1] D. A. Torchia,† B. Fohr,* M. F. Young,* and N. S. Fedarko*[,‡]
**Craniofacial and Skeletal Diseases Branch and †Molecular Structural Biology Unit, NIDCR, NIH, Bethesda, Maryland, 20892-4320 and ‡Division of Geriatrics, Department of Medicine, Johns Hopkins University, Baltimore, Maryland, 21224*

Received November 22, 2000

Bone sialoprotein (BSP) and osteopontin (OPN) are two members of the SIBLING (Small Integrin-Binding LIgand, N-linked Glycoprotein) family of genetically related proteins that are clustered on human chromosome 4. We present evidence that this entire family is the result of duplication and subsequent divergent evolution of a single ancient gene. The solution structures of these two post-translationally modified recombinant proteins were solved by one dimensional proton NMR and transverse relaxation times. The polypeptide backbones of both free BSP and OPN rapidly sample an ensemble of conformations consistent with them both being completely unstructured in solution. This flexibility appears to enable these relatively small glycoproteins to rapidly associate with a number of different binding partners including other proteins as well as the mineral phase of bones and teeth. These proteins often function by bridging two proteins of fixed structures into a biologically active complex. © 2001 Academic Press

Key Words: **SIBLING; bone sialoprotein; osteopontin; NMR structure; flexible structure.**

Bone sialoprotein (BSP), osteopontin (OPN but sometimes known as SPP1 and Eta-1), dentin matrix protein I (DMPI), dentin sialophosphoprotein (DSPP), and matrix extracellular phosphoglycoprotein (MEPE) are the products of five genes clustered along human chromosome 4 (1–6). Directly comparing the primary amino acid sequences of these proteins does not result in sufficient homology to warrant calling them a related family. For example, Fig. 1A shows the computer alignment of human BSP with that of human OPN and there is little in common between them. But some of the properties of their individual exons are very similar (Fig. 1B). For example, exon 1 is always noncoding, exon 2 is the leader sequence and the first two amino acids of the mature protein, exon 3 usually contains the consensus sequence (SSEE) for casein kinase II (CKII) phosphorylation, exon 4 is usually somewhat proline-rich and among the acid proteins is the only significantly positive-charged domain, exon 5 usually contains a CKII site and is the site of the only two known splice variants (OPN and DMPI, Refs. 2, 3). The majority of each protein is encoded by the last one or two exons and contains the integrin-binding RGD tripeptide. Another point of similarity of all these genes is that all introns always interrupt between codons (type 0), thus leaving open the possibility of splicing together any two exons without causing frame shifts. We have named this family of proteins the SIBLING family for Small Integrin-Binding LIgand, N-linked Glycoprotein based not on current theories of their functions (which are poorly understood) but based on the simple biochemical and genetic features shared by all members.

BSP, OPN, DMPI, and DSPP are somewhat similar in character being secreted, phosphorylated, and sulfated sialoproteins that are acidic in character. BSP is rich in glutamic acids (7, 8). OPN (2) and DSPP (4) are rich in aspartic acid and DMP1 is slightly larger in size and has some regions rich in Glu and others rich in Asp. All of the acidic SIBLINGs are known to bind strongly to hydroxyapatite. MEPE appears to be more distantly related in that it is not acidic in character although it is likely to have a phosphorylated carboxy-terminus (5). In addition to binding to integrins, OPN has been reported to be able to bind to CD44 (9, 10) as has DMP1 (11). Furthermore, at least BSP, OPN and DMP1 share one complex biological function, that of being able to bridge complement Factor H to cell surface receptors and thereby protecting them from lysis by the alternate complement pathway (10, 11).

In normal adults, BSP is relatively specific for the tissues of the skeleton (12). OPN is more widely expressed being found in the skeleton as well as in a number of epithelial cells and in macrophages. During

[1] To whom correspondence should be addressed at Room 228, Building 30, CSDB, NIDCR, NIH, Bethesda, MD 20892-4320. Fax: 301-402-0824. E-mail: lfisher@dir.nidcr.nih.gov.

0006-291X/01 $35.00

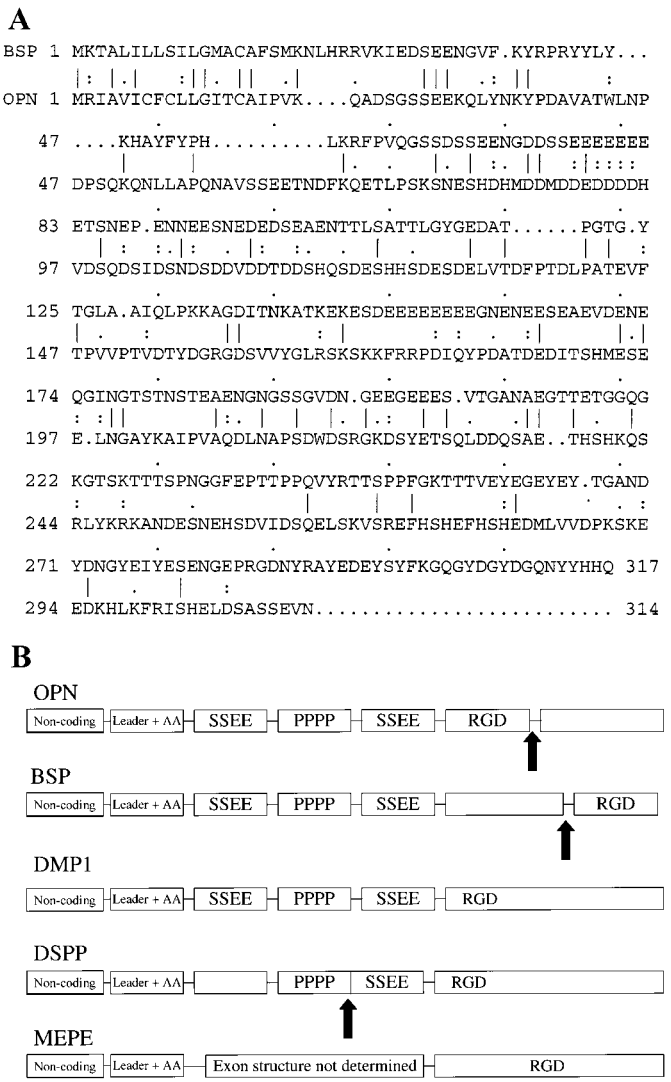

FIG. 1. (A) Direct comparison of the protein sequences of human BSP and OPN to illustrates that the primary amino acid sequences would not normally be considered closely related. (B) Exon–intron similarities of the genes clustered along human chromosome 4 define the SIBLING family. See text for the shared properties of the exons. Some introns have been introduced or lost during the evolution of the family (arrows). Only the first two and the last exon of the more distant MEPE member were discerned by comparing the published cDNA sequence (5) to the human genome project database. The region marked "exon structure not determined" has not yet been published in the human genome project database.

the development of a number of tumors, particularly those that tend to metastasize to bone, BSP and OPN are usually strongly up-regulated. Indeed, the level of BSP expression in sections from tumors from breast (13) and prostate cancers (14) was predictive of the severity of the disease, with patients expressing higher levels tending to have a more severe disease and often having a shorter life span.

Determining the structures of BSP and OPN presents a number of difficulties. They are highly post-translationally modified and would not be expected to make diffracting crystals. Their size, >50,000 daltons (~35,000 Da of protein and probably >15,000 Da of post-translational modifications), has made it unlikely with current technology that NMR could fully assign the locations of all of the atoms. Previously we have shown that the carboxy-terminal 60 amino acid polypeptide of human BSP that included the integrin-binding RGD tripeptide was a random coil in solution (15). In spite of these difficulties, NMR can clearly be used to determine whether or not full length proteins of this size contain significant amounts of structure in solution. In this report, we have produced biologically active recombinant BSP and OPN and subjected them to one dimensional proton NMR analysis to determine if there was evidence for sustained structure in solution.

MATERIALS AND METHODS

Recombinant human BSP and OPN were made as described previously (10). Briefly, the cDNA for human BSP (clone B6-5g, Ref. 8) and OPN (clone OP-10, the splice variant that lacks the 42 bp exon 5, Ref. 2) were subcloned and expressed in replication-deficient adenoviruses. Human bone marrow fibroblasts (third passage) were infected with an optimum dose of adenovirus and cultured for an additional 5–7 days in serum-free media. The BSP and OPN, which represented more than 50% of the protein in the harvested media, were purified using standard non-denaturing HPLC ion exchange chromotography. Both proteins were shown to have biological activity by their abilities to inhibit lysis of cells by the alternate complement pathway through binding to both cell surface receptors ($\alpha_v\beta_3$ integrin and/or CD44) and Factor H as described previously (10).

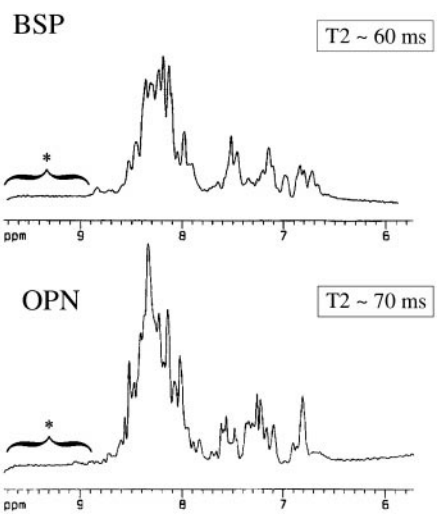

FIG. 2. Standard one dimensional proton NMR of recombinant human BSP (top) and OPN (bottom) using a Bruker DMX spectrometer operating at 500 MHz at 303°K. Proton transverse relaxation times (T2) were obtained using a one–one echo pulse sequence. Brackets with * show the region that would have a significant number of large peaks if the two proteins contained any structure over the time frame sampled by NMR. Clearly both proteins are flexible and unstructured.

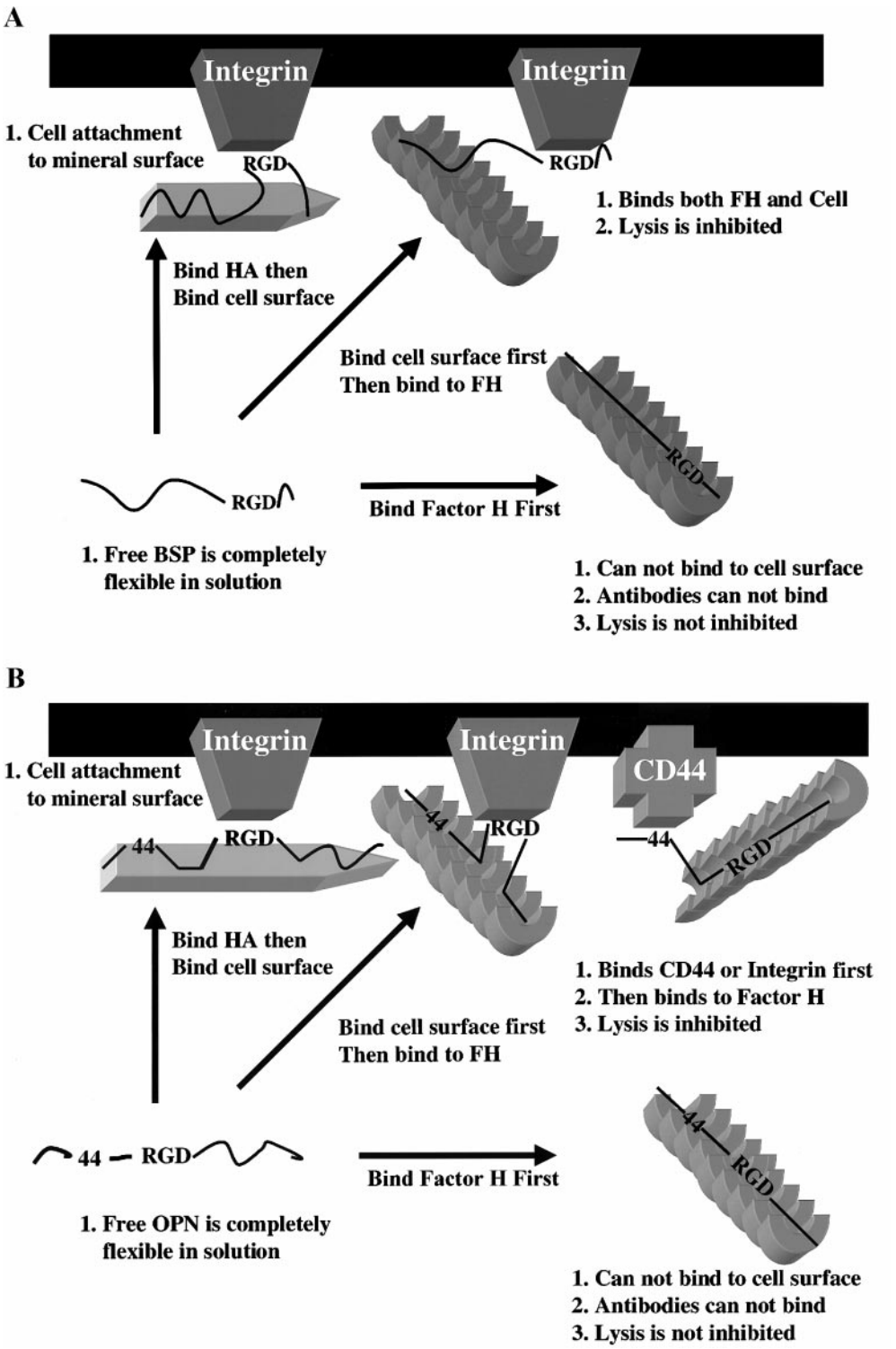

FIG. 3. Diagrams showing the multiple interactions of BSP (A), OPN (B), and DMPI (C). Notice that regions of interactions of the various binding partners share large portions of the SIBLING sequences.

Each protein solution was made to a final concentration of ~15 mg/ml in water (BSP) or dilute (10 mM) sodium phosphate, pH 6.0 (OPN). One dimensional proton NMR spectra were recorded on a Bruker DMX spectrometer operating at 500 MHz and 303°K. Proton transverse relaxation times were obtained using a one–one echo pulse sequence.

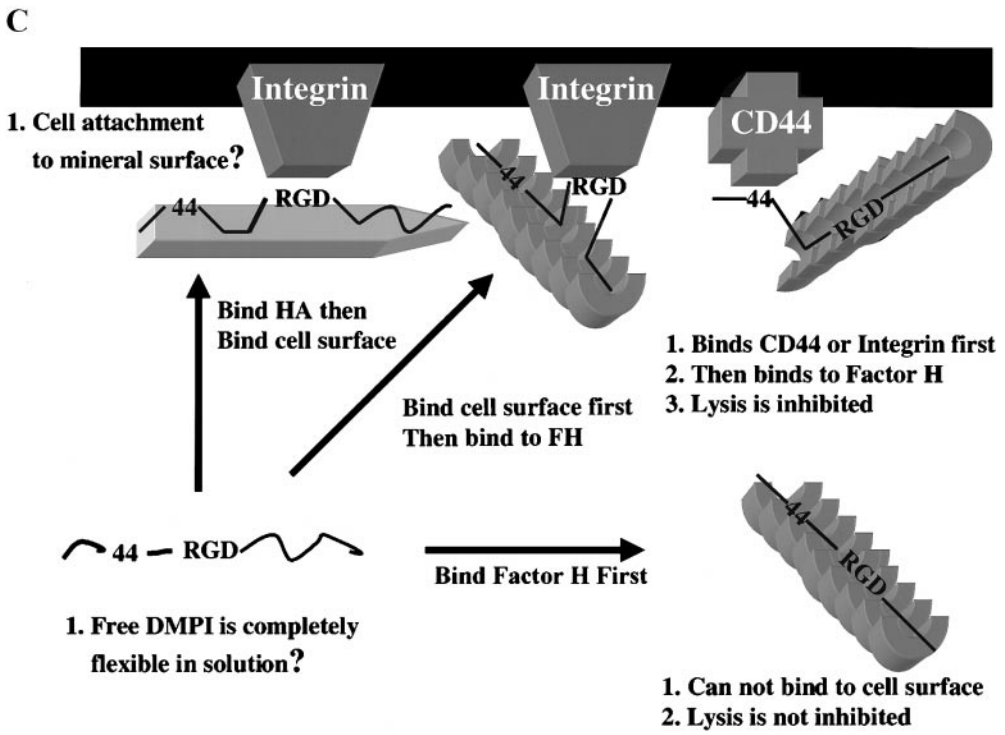

FIG. 3—*Continued*

RESULTS AND DISCUSSION

Recombinant human bone sialoprotein and osteopontin were both made in cultured human marrow fibroblasts by infection with replication-deficient adenoviruses containing the appropriate human cDNA behind a strong eukaryotic promoter. The proteins were made in high yield in serum-free media. The purified proteins were concentrated to ~15 mg/ml and subjected to standard one dimensional proton NMR. Figure 2 shows the spectra of the two proteins. The narrow spread of the peptide backbone amide NH peaks (~7.8–8.6 ppm) and the side chain NH peaks (~6.6–7.4 ppm) are indicative of a lack of ordered structures over the NMR time scale. If either protein had significant amounts of structure in solution, there would have been many significantly sized peaks at ppm > ~8.8. Furthermore, the average proton T2 relaxation times for both of these molecules were on the order of 60 ms. This is fully consistent with the two proteins being extended and flexible in solution. If they had formed ordered compact structures, the T2 relaxation time would have been substantially shorter, in the range of 5–20 ms. Therefore, both BSP and OPN, when purified and isolated in solution, are flexible along their entire length and have no significant regions that persist in a single structural environment for more than a few milliseconds.

That a protein is completely flexible in solution does not mean that it will always remain so. Portions of these protein that strongly interact with other proteins (such as Factor H and the $\alpha_v\beta_3$ or other integrin structures for both BSP and OPN as well as CD44 for OPN) will almost certainly adopt specific structures in relation to their binding partners. Proteins or portions of proteins that lack structure in solution are often involved in bridging two or more proteins or other large molecules such as RNA. Isolated ribosomal proteins (for example, L39e) and portions of many other ribosomal proteins (L2, L4, L3, etc.) are known to have little or no structure. These proteins all, however, do adopt complex structures as they assemble into complete ribosomes (16). Unfortunately, because all of the known binding partners of both BSP and OPN are large (Factor H, for example is ~150,000 Da), such complexes would tumble too slowly to solve their combined ordered structures by NMR.

We are proposing that for a small protein that shares many binding partners, there are advantages of retaining a flexible structure. This property would be particularly important if one or more of the binding partners interacted with the SIBLING over much of its length. For example, we have shown previously that BSP can bind strongly to hydroxyapatite (HA) and does so throughout most of its entire length (15). For BSP to function as an attachment protein for cells to attach to HA (as has been frequently proposed by people in the bone field), the BSP would be bound to HA along its length and yet keep its RGD domain available to bind

an integrin such as $\alpha_v\beta_3$ (Fig. 3A). BSP can also bind first to the $\alpha_v\beta_3$ integrin then it can bind to complement Factor H (Fig. 3A). This four-subunit complex stops the lytic pathway of complement *in vitro* (10). If, however, Factor H binds to BSP before the latter can bind to its integrin, the Factor H binds throughout the length of BSP blocking its ability to subsequently bind to the integrin. The integrin-binding domain of BSP can therefore bind to either the integrin or to specific portions of Factor H. (The Factor H-BSP complex is not only unable to bind to its integrins, but also cannot be bound by a number of different BSP antisera whose epitopes are known to be distributed throughout the protein. This binding effectively inactivates the BSP activity, keeping the functional distance for secreted BSP probably to autocrine or paracrine distances.) Thus, BSP is a protein that can be bound along most of its length separately to at least two different binding partners *in vivo.* The simplest way to do this is to have a completely unstructured protein essentially alternating sequences that can specifically interact with the different binding partners.

OPN has a similar pattern of interactions. Like BSP, it has been shown by many to bind to HA, probably via its many aspartic acid groups spread along its length (Fig. 3B). It has been proposed that osteoclasts use OPN to bridge between the integrins on the cell surface and the mineral phase during resorption of bone matrix (17). We have shown that Factor H also binds to OPN already complexed to either $\alpha_v\beta_3$ integrin or CD44 and this complex can stop the alternate complement lysis pathway (10) (Fig. 3B). Like BSP, if OPN binds first to Factor H in solution, not only is it unable to subsequently bind to its integrin or CD44 but the complex can not function in stopping the lytic pathway. A previously formed Factor H-OPN complex completely masks the integrin-binding domain and the CD44-binding domain (10). Furthermore, our current supply of antisera against OPN also appears to be unable to bind to the OPN when complexed with Factor H (data not shown). This all suggests that essentially the full length of this SIBLING is tightly bound to either Factor H or to HA and that specific portions of OPN can bind to either of two independent cell surface receptors or to additional regions of Factor H.

DMPI, although less well studied, has properties similar to OPN including: HA-binding; Factor H binding; integrin-binding; and CD44-binding domains (Fig. 3C). The same arguments for the sharing of multiple and sometime extensive binding domains can be made for this SIBLING member as for OPN. However, we do not at this time have sufficient amounts of recombinant DMPI to see if it is also entirely flexible in solution when unbound.

In conclusion, we have purified milligram quantities of biologically active BSP and OPN. Both of these members of the SIBLING family, when isolated, are unstructured and highly flexible on the NMR time scale in solution. While the DMPI and DSPP members of the family have not yet been made in sufficient quantities to perform the NMR analysis, we expect from their similar sequences to be mostly if not entirely flexible also. MEPE is substantially different in character and we can not realistically predict if it will be unstructured also. We propose that having the SIBLING protein be flexible allows the same regions to use several interspersed binding sequences for several different partners at different times. That even two members of this same genetic family, BSP and OPN, can have diverged so much at the amino acid level and yet have maintained several binding partners and their full flexibility in solution suggests that this property of flexibility is essential to their functions.

REFERENCES

1. Kerr, J. M., Fisher, L. W., Termine, J. D., Wang, M. G., McBride, O. W., and Young, M. F. (1993) The human bone sialoprotein gene (IBSP): Genomic localization and characterization. *Genomics* **17,** 408–415.
2. Young, M. F., Kerr, J. M., Termine, J. D., Wewer, U. M., Wang, M. G., McBride, O. W., and Fisher, L. W. (1990) cDNA cloning, mRNA distribution and heterogeneity, chromosomal location, and RFLP analysis of human osteopontin (OPN). *Genomics* **7,** 491–502.
3. Hirst, K. L., Simmons, D., Feng, J., Aplin, H., Dixon, M. J., and MacDougall, M. (1997) Elucidation of the sequence and genomic organization of the human dentin matrix acidic phosphoprotein 1 (DMP1) gene: Exclusion of the locus from a causative role in the pathogenesis of dentinogenesis imperfecta type II. *Genomics* **42,** 38–45.
4. MacDougall, M., Simmons, D., Luan, X., Nydegger, J., Feng, J., and Gu, T. T. (1997) Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4. *J. Bio. Chem.* **272,** 835–842.
5. Rowe, P. S. N., de Zoysa, P. A., Dong, R., Wang, H. R., White, K. E., Exons, M. J., and Oudet, C. L. (2000) MEPE, a new gene expressed in bone marrow and tumors causing osteomalcia. *Genomics* **67,** 54–68.
6. Crosby, A. H., Edwards, S. J., Murray, J. C., and Dixon, M. J. (1995) Genomic organization of the human osteopontin gene: exclusion of the locus from a causative role in the pathogenesis of dentinogenesis imperfecta type II. *Genomics* **27,** 155–160.
7. Fisher, L. W., Whitson, S. W., Avioli, L. V., and Termine, J. D. (1983) Matrix sialoprotein of developing bone. *J. Biol. Chem.* **258,** 12723–12727.
8. Fisher, L. W., McBride, O. W., Termine, J. D., and Young, M. F. (1990) Human bone sialoprotein. Deduced protein sequence and chromosomal localization. *J. Biol. Chem.* **265,** 2347–2351.
9. Weber, G. F., Askar, S., Glimcher, M. J., and Cantor, H. (1996) Receptor-ligand interaction between CD44 and osteopontin (Eta-1). *Science* **271,** 509–512.
10. Fedarko, N. S., Fohr, B., Robey, P. G., Young, M. F., and Fisher, L. W. (2000) Factor H binding to bone sialoprotein and osteopontin enables tumor cell evasion of complement-mediated attack. *J. Biol. Chem.* **275,** 16666–16672.

11. Fedarko, N. S., Fohr, B., and Fisher, L. W. (2000) DMP1 inactivates complement by bridging Factor H to cell surface receptors. *J. Dent. Res.* **79,** 339.
12. Bianco, P., Fisher, L. W., Young, M. F., Termine, J. D., and Robey, P. G. (1991) Expression of bone sialoprotein (BSP) in developing human tissues. *Calcif. Tissue Int.* **49,** 421–426.
13. Bellahcene, A., Menard, S., Bufalino, R., Moreau, L., and Castronovo, V. (1996) Expression of bone sialoprotein in primary human breast cancer is associated with poor survival. *Int. J. Cancer.* **69,** 350–353.
14. Waltregny, D., Bellahcene, A., Van Riet, I., Fisher, L. W., Young, M., Fernandez, P., Dewe, W., de Leval, J., and Castronovo, V. (1998) Prognostic value of bone sialoprotein expression in clinically localized human prostate cancer. *J. Natl. Cancer Inst.* **90,** 1000–1008.
15. Stubbs, J. T., III, Mintz, K. P., Eanes, E. D., Torchia, D. A., and Fisher, L. W. (1997) Characterization of native and recombinant bone sialoprotein: Delineation of the mineral-binding and cell adhesion domains and structural analysis of the RGD domain. *J. Bone Miner. Res.* **12,** 1210–1222.
16. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 angstom resolution. *Science* **289,** 905–920.
17. Reinholt, F. P., Hultenby, K., Oldberg, A., and Heinegard, D. (1990) Osteopontin—A possible anchor of osteoclasts to bone. *Proc. Natl. Acad. Sci. USA* **87,** 4473–4475.