Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 2919–2923, May 1988
Biochemistry

Osteonectin cDNA sequence reveals potential binding regions for calcium and hydroxyapatite and shows homologies with both a basement membrane protein (SPARC) and a serine proteinase inhibitor (ovomucoid)

(bone/calcium-binding proteins)

MARK E. BOLANDER*†, MARIAN F. YOUNG‡, LARRY W. FISHER‡, YOSHIHIKO YAMADA*, AND JOHN D. TERMINE‡

*Laboratory of Developmental Biology and Anomalies and ‡Bone Research Branch, National Institute of Dental Research, National Institutes of Health, Bethesda, MD 20892

Communicated by R. H. Wasserman, December 10, 1987 (received for review August 4, 1987)

ABSTRACT **Osteonectin is a prominent noncollagenous protein of developing bone. A 2150-base-pair cDNA coding for osteonectin, isolated from a bovine bone cell λgt11 expression library, was sequenced and identified by comparison with protein sequence data. The nucleotide sequence predicts that osteonectin contains 304 amino acids, including a 17-residue signal peptide. Analysis of the deduced protein sequence suggests that the secreted protein contains at least four distinct structural domains. An acidic region at the amino terminus of the protein appears to be a potential hydroxyapatite-binding site. This is followed by a second domain, rich in cysteine, that shows sequence homology with cysteine-rich domains in turkey ovomucoid and other serine proteinase inhibitors. Two sequences homologous with central calcium-binding loops of "EF hands" and thus having potential to be high-affinity calcium-binding sites are located in two other domains within the carboxyl-terminal half of the protein. Finally, the osteonectin sequence shows near identity (>90%) with another protein, SPARC (*s*ecreted *p*rotein, *a*cidic and *r*ich in *c*ysteine), secreted by mouse parietal endoderm. These data suggest that osteonectin, a protein present in bone and other selected tissues, is a multifunctional protein.**

Osteonectin (M_r 32,000) is one of the principal noncollagenous proteins of bone (1–3). It is not readily extracted unless the tissue is demineralized, suggesting that it is protected from dissolution by the bone mineral (1–3). Immunolocalization studies demonstrated that osteonectin is a normal component of osteoid, the newly deposited matrix of bone (1, 4, 5), and indeed, osteonectin secreted by osteoblasts in culture is incorporated into the extracellular matrix (6, 7). *In vitro* binding studies indicated that osteonectin has high affinity for collagen, calcium, and hydroxyapatite (1, 3, 8), suggesting that it might function in bone by linking the mineral phase to the collagen matrix (1).

Osteonectin is composed of a single polypeptide chain with several intramolecular disulfide bonds (1–3, 9). Recent studies have focused on the calcium- and hydroxyapatite-binding potential of osteonectin and shown the protein to have a high-affinity calcium-binding site ($K_d = 3 \times 10^{-7}$ M) (3). Further, the affinity of osteonectin for hydroxyapatite ($K_d = 8 \times 10^{-8}$ M) is some 6 times stronger than that of osteocalcin (the bone Gla protein), which utilizes γ-carboxyglutamic (Gla) residues for binding calcium and hydroxyapatite mineral (3). While these studies imply important functions for osteonectin in bone, little is known of its structure or biological functions. A cDNA clone to osteonectin has been isolated from a bovine osteoblast library and its characteristics have been described (10). We now report the sequencing of this clone and its entire deduced amino acid sequence.§ These studies permit tentative identification of structural and functional domains in the molecule and comparison of the osteonectin sequence with that of other proteins.

MATERIALS AND METHODS

Isolation of a Bovine cDNA for Osteonectin. Cultured bone cells were prepared from cellular outgrowths of collagenase-treated fetal bovine bone (7). Total cellular RNA was extracted with guanidine hydrochloride (11), and poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose chromatography (type III; Collaborative Research, Waltham, MA). A cDNA library was constructed in λgt11 expression vector (10). An osteonectin clone was isolated by screening with antibody to osteonectin, by a modification of the method of Young and Davis (12), and then purified and characterized as described (10).

cDNA Sequencing. The sequence of the cDNA was determined by the method of Sanger *et al.* (13, 14). Two *Eco*RI restriction fragments, 0.2 and 0.3 kilobase (kb) long (10), were sequenced from both ends. The 1.6-kb *Eco*RI restriction fragment was sequenced on both strands by a shotgun method (15) using a total of 95 clones. Each nucleotide was sequenced on average six times. The nucleotide sequence of certain regions of the cDNA was determined by sequencing restriction fragments subcloned in phage M13 mp18 and mp19 DNA (13, 14). Restriction fragment sequences were aligned by sequencing a 0.6-kb *Kpn* I restriction fragment that overlapped all three *Eco*RI fragments. An ambiguity in the 1.6-kb *Eco*RI fragment was resolved by subcloning and sequencing a *Xho* I–*Kpn* I restriction fragment. An ambiguity in the 5′ 0.2-kb *Eco*RI fragment was resolved by subcloning and sequencing a *Kpn* I–*Eco*RI restriction fragment.

Computer Analysis. The amino acid sequence was analyzed for secondary structure by the Chou–Fasman method (16). Hydrophobicity was predicted using the method of

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

Abbreviations: EF hand, the structure denoting calcium ion-binding sites found in calmodulin and homologous calcium-binding proteins; SPARC, *s*ecreted *p*rotein, *a*cidic and *r*ich in *c*ysteine.

†To whom reprint requests should be sent at the present address: Orthopaedic Research Unit, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892.

§This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03233).

```
5'
1                                                    CGGCCCGGAGCTTGCCTGCTGCCTGCCTGCCTGCCACTGAGGGTTCCCAGCACC

55    ATG AGG GCC TGG ATC TTC TTT CTC CTT TGC CTG GCC GGG AGG GCC TTG GCA GCC CCT CAA
-17   Met Arg Ala Trp Ile Phe Phe Leu Leu Cys Leu Ala Gly Arg Ala Leu Ala Ala Pro Gln
      ------------------------------------ Signal Peptide ------------------------------------

115   CAG GAA GCC TTG CCT GAT GAG ACA GAA GTG GTG GAA GAA ACC GTG GCC GAG GTG GCC GAG
4     Gln Glu Ala Leu Pro Asp Glu Thr Glu Val Val Glu Glu Thr Val Ala Glu Val Ala Glu

175   GTA CCC GTG GGA GCC AAC CCC GTC CAG GTG GAA GTA GGA GAA TTC GAT GAT GGT GCT GAG
24    Val Pro Val Gly Ala Asn Pro Val Gln Val Glu Val Gly Glu Phe Asp Asp Gly Ala Glu

235   GAA ACC GAA GAG GAG GTG GTG GCC GAG AAC CCC TGC CAG AAC CAC CAC TGC AAA CAC GGC
44    Glu Thr Glu Glu Glu Val Val Ala Glu Asn Pro Cys Gln Asn His His Cys Lys His Gly

295   AAG GTG TGC GAA CTG GAC GAG AAC AAC ACC CCC ATG TGT GTG TGC CAG GAC CCC ACC AGC
64    Lys Val Cys Glu Leu Asp Glu Asn Asn Thr Pro Met Cys Val Cys Gln Asp Pro Thr Ser

355   TGC CCT GCC CCC ATC GGC GAG TTT GAG AAG GTG TGC AGC AAC GAC AAC AAG ACC TTC GAC
84    Cys Pro Ala Pro Ile Gly Glu Phe Glu Lys Val Cys Ser Asn Asp Asn Lys Thr Phe Asp

415   TCT TCC TGC CAC TTC TTT GCC ACC AAG TGC ACA CTG GAG GGC ACC AAG AAG GGC CAC AAA
104   Ser Ser Cys His Phe Phe Ala Thr Lys Cys Thr Leu Glu Gly Thr Lys Lys Gly His Lys

475   CTC CAC CTG GAC TAC ATC GGG CCT TGC AAA TAC ATC CCC CCC TGC CTG GAC TCC GAG CTG
124   Leu His Leu Asp Tyr Ile Gly Pro Cys Lys Tyr Ile Pro Pro Cys Leu Asp Ser Glu Leu

535   ACT GAA TTC CCC TTG CGC ATG CGG GAC TGG CTT AAG AAC GTC CTG GTC ACG CTG TAC GAG
144   Thr Glu Phe Pro Leu Arg Met Arg Asp Trp Leu Lys Asn Val Leu Val Thr Leu Tyr Glu

595   AGG GAC GAG GAC AAC AAC CTC CTG ACC GAG AAG CAG AAG CTG CGA GTG AAA AAG ATC CAC
164   Arg Asp Glu Asp Asn Asn Leu Leu Thr Glu Lys Gln Lys Leu Arg Val Lys Lys Ile His
      ------------------------------------ Hydrophilic Region ------------------------------------

655   GAG AAT GAG AAG CGC CTG GAG GCT GGC TGC GAC CAT CCT GTG GAA CTG CTG GCC CGG GAC
184   Glu Asn Glu Lys Arg Leu Glu Ala Gly Cys Asp His Pro Val Glu Leu Leu Ala Arg Asp

715   TTC GAG AAG AAC TAC AAC ATG TAC ATC TTC CCT GTG CAC TGG CAG TTC GGG CAG CTG GAT
204   Phe Glu Lys Asn Tyr Asn Met Tyr Ile Phe Pro Val His Trp Gln Phe Gly Gln Leu Asp

775   CAG CAC CCC ATT GAC GGG TAC CTG TCT CAC ACC GAG CTG GCC CCA CTG CGC GCC CCC CTT
224   Gln His Pro Ile Asp Gly Tyr Leu Ser His Thr Glu Leu Ala Pro Leu Arg Ala Pro Leu

835   ATC CCC ATG GAA CAC TGC ACC ACC CGC TTT TTC GAG ACA TGT GAC CTG GAC AAC GAC AAG
244   Ile Pro Met Glu His Cys Thr Thr Arg Phe Phe Glu Thr Cys Asp Leu Asp Asn Asp Lys

895   TAC ATC GCC CTG GAC GAG TGG GCC GGC TGC TTC GGC ATC AAG GAG AAG GAC ATC GAC AAG
264   Tyr Ile Ala Leu Asp Glu Trp Ala Gly Cys Phe Gly Ile Lys Glu Lys Asp Ile Asp Lys

955   GAC CTC GTG ATC TAA AGCCATGCCTCCTCCTGCAGAACCAGATTCTCTCTCTTTGACCTTCCCCCTCCTGTTTT
284   Asp Leu Val Ile ■

1029  CCCCAATGTTTAAAATGTTTGGATGGTTTGTTGTTCTGCCTGGGGTCAAGGTGCTAATATAGACTTAAACGAATACATT
1108  AACGGTGCTAAACACAGACATTTTAACCCAAGTCACGACATTCTTAGCTGTAACTCAGCTCTCACGGCCTCTTGCTCAC
1187  CCATGAATGGTCCCGTTTTCCTCTTGCCGCGTGCACCTCCACCCATTGTCTTGGTGGCACACGGGTGGAACACTTGATC
1266  TGCTCGAGTCTGCCTTCAACACACATTGCATCTTCAGATTTTCTACTTTTCTGTTCAAAACTAATATTCACCAAGTCAG
1345  ACTTTGTGTTAATTTTATATCAGGGTATTGGCTGCCAGGGGGTCATCCCTAAGTGGCCTGAAGATGGACAAAGGGAAGT
1424  AACAGGCACGTGATGTTGGCAAGGATGCTTCTAGGGCTAGAGGATCAGTGGTGGGAGAGACCTGCAGAATCCACCAGCC
1503  AGAACTGCAGATAACGAATCTTATGGTCAGGGGCTGTGACTGACAGAAGGAAACTGAGGCTGTGTTCTGAAAGTACATA
1582  AACTCTCACATATACCCAGTTCTTCACCATCTCCCCTCCTCACTTTGCAGTGCCATTTCTTTTTGCATTAGGCAATTTG
1661  CTCAGACTTTCCAGAGCCATGGCCCATCCGTTCTCTGGAATCCCCCACACCTCTGAGAGGTGGATCACCATGTCCTGCA
1740  GGGCTGCTCCCCTCCAAACTACCTTTAGGAGATGCAGGACAGGGAGGCTGTTTCAACCAGAAAGACCAAAATCAAGAGC
1819  GAGGAGCAGAACGTGGTAAAACAGAAAAAGGGCAGGTGGCAAATTGGTTTTCTTTTGGGTTTTCTGTTTTCTTTTTTCA
1898  CATCTGGATGGCTGTCACCAGAGATCTTTCAGTCGCTAGCATGTTCCTCCTCTTCTCCCCTCCCCACTTTTTCTTTCTA
1977  TTAATCAAAAGAAATTTCAAAATCAATGGGATGGTCGGATCTCACAGGCTGAGAACTCGTTCACCTCCAAGCATTTCAT
2056  GAAAAAGCTGCTTCTTATTAATCATGCAAACTCTTGCCACGGTGTGAAGAGTTTGACAAATCTTTCAAAATAAAAAGTA
2135  CTGATTTAAAAAAAAAA
3'
```

FIG. 1. Nucleotide sequence of bovine osteonectin cDNA and deduced amino acid sequence. Nucleotides and amino acid residues are numbered at left. Possible polyadenylylation signals in the 3′ untranslated sequence are underlined. The signal peptide (italics) is labeled with negative numbers. The first amino-terminal residue of the bone matrix protein is numbered 1. Two amino-terminal homologous regions are boxed with heavy lines. Potential glycosylation sites are marked by arrowheads. A 30-residue hydrophilic region is underlined. Cysteine residues are boxed. The termination codon is indicated by a black box.

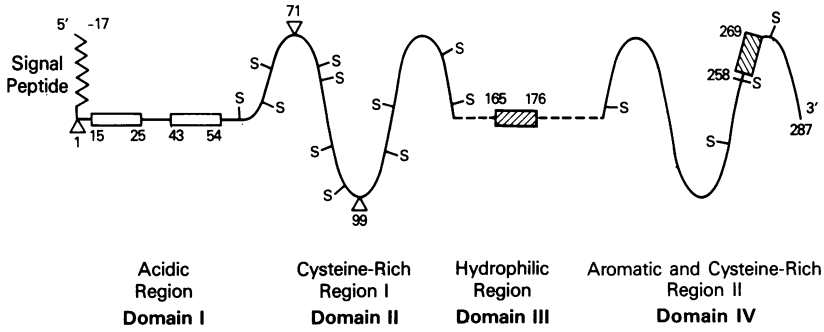

FIG. 2. Model of osteonectin structure. The four domains are labeled I–IV. Approximate positions of the 15 cysteine residues are marked (–S); the precise locations of disulfide bonds are not known. Triangles indicate cleavage site of signal peptide (position 1) as well as potential glycosylation sites (positions 71 and 99); open boxes represent internal homologies; broken lines indicate the hydrophilic region including one of the two "EF-hand" calcium-binding loop structures (hatched boxes).

Kyte and Doolittle (17). The National Biomedical Research Foundation (NBRF) protein data base[¶] was searched for sequence homologies by the method of Lippman and Pearson (18).

RESULTS

Sequencing the cDNA Isolated from the Bovine Library. We previously reported a cDNA clone, 2150 base pairs long, encoding osteonectin. This clone was isolated by antibody screening from a bovine osteoblast (10) cDNA library in the expression vector λgt11. RNA blot analysis showed that it hybridized to a mRNA species of ≈2.0 kb (10). Primer-extension analysis indicated that the 5′ end of the mRNA was ≈45 bases from the 5′ end of the cDNA (data not shown).

The 2150-base sequence obtained by sequencing the cDNA and the predicted amino acid sequence of the protein it encodes are illustrated in Fig. 1. A methionine codon at base 55 is followed by an open reading frame of 304 amino acids extending to a stop codon at base 967. This stop codon is followed by multiple stop codons in all three reading frames.

Analyses of the Nucleotide Sequence. A 9-base sequence surrounding the methionine codon, starting at base 49, GC ACC ATG A, fits well with the consensus sequence for eukaryotic initiation sites (19). The 5′ untranslated sequence is G+C-rich, with four GCCT repeats, extending from base 14 to base 32. A polyadenylylation signal, AAAATAAAAA, begins at position 2121 (20).

Primary Structure of Osteonectin. The deduced amino acid sequence was analyzed by computer predictions of hydrophobicity and secondary structure. Amino acid residues Ala-1 through Val-31 are in agreement with the known amino-terminal sequence of bovine bone osteonectin (10). A 17-residue polypeptide preceding the known amino-terminal sequence of the secreted protein is rich in hydrophobic amino acids, as is characteristic for signal peptides (21). The remainder of the protein, 287 amino acids, can be divided into four distinctive domains (Fig. 2). The first domain (I) at the amino terminus, an acidic region extending from Ala-1 to Pro-54, is rich in glutamic and aspartic acids (18 of 54 residues) and in hydrophobic residues (valine, leucine, and phenylalanine; 13 of 54 residues). Throughout this entire region there are no positively charged amino acid residues. This osteonectin domain contains two homologous internal repeats (66% sequence identity), between the sequences from Glu-15 to Pro-25 and from Glu-43 to Pro-54 (Fig. 1).

A second domain (II) extends approximately from Cys-55 to Cys-138 and contains two potential N-glycosylation sites, at Asn-71 and Asn-99. This region, 84 amino acids in length, contains 11 of the 15 cysteine residues in osteonectin. A third distinctive region (domain III) contains a sequence of hydrophilic amino acids (43 of 56 residues) extending approximately from Leu-139 to Gly-192. Chou–Fasman (16) analysis suggests the presence of α-helical structure (residues Leu-139 to Arg-164 and Arg-178 to Gly-192) in this region, which contains an almost equal number of negatively and positively charged residues. The remaining 95 amino acids of the protein, from approximately Cys-193 to Ile-287, form a fourth domain (IV). This region contains the remaining four cysteine residues of osteonectin and also displays potential for α-helical structure (residues Leu-243 to Cys-257 and Trp-271 to Ile-287). This domain also contains 12 of the total 25 aromatic amino acids in the 304-residue preosteonectin molecule.

Possible Calcium- and Hydroxyapatite-Binding Domains. A search of the NBRF protein data base (18) showed sequence homology between two regions of osteonectin, Asp-165 to Lys-176 in domain III and Asp-258 to Glu-269 in domain IV, and the central calcium-binding loop of "EF hand" structures found in bovine brain calmodulin (22, 23), the calcium-binding protein of muscle (24), and both the α and β chains of bovine S-100 protein (25) (Fig. 3). On the basis of crystallographic data gathered for carp muscle parvalbumin, Kretsinger (23) proposed that calcium binds in this 12 amino acid loop within an "EF hand" structure. Carbonyl-containing residues at the X, Y, Z, −Y, and −Z positions of the loop stabilize calcium binding.

Osteonectin also has a high affinity for hydroxyapatite ($K_d = 8 \times 10^{-8}$ M) and can inhibit hydroxyapatite crystal growth at very low protein concentrations (1.5×10^{-7} M) (3, 8). The amino terminus of osteonectin is completely acidic, with 15 glutamic and 3 aspartic residues in the first 54 residues of the protein. These negatively charged residues also tend to be clustered in this domain of the molecule and thus could effectively interact with the hydroxyapatite lattice (26).

Homologies with Serine Proteinase Inhibitors. The search of the NBRF protein data base also showed partial homology between osteonectin and serine proteinase inhibitors. The homology is strongest with ovomucoids (27, 28), but also includes acrosyn inhibitor (29) and pancreatic trypsin inhib-

	X		Y		Z	G	-Y		-X			-Z
	1	2	3	4	5	6	7	8	9	10	11	12
Osteonectin 165-176	D	E	D	N	N	L	L	T	E	K	Q	K
Osteonectin 258-269	D	L	D	N	D	K	Y	I	A	L	D	E
Bovine Brain Calmodulin: DI	D	K	D	G	N	G	T	I	T	T	K	E
Ca-binding Prot. of muscle: DIII	D	T	D	K	D	R	S	I	D	L	N	E
Bovine S-100 alpha: DI	G	K	E	G	D	K	Y	K	L	S	K	E

FIG. 3. Comparison of osteonectin sequences with EF-hand sequences in domain I of bovine brain calmodulin (22, 23), domain III of the calcium-binding protein of muscle (24), and domain I of bovine brain S-100 α chain (25). X, Y, Z, −X, and −Z denote carbonyl donor residues within the central calcium-binding loop of the EF hand (23). G denotes position of the central glycine residue in calmodulin (23). Homologous residues are boxed. Amino acid residues are represented by standard one-letter symbols.

[¶]Protein Identification Resource (1987) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 12.0.

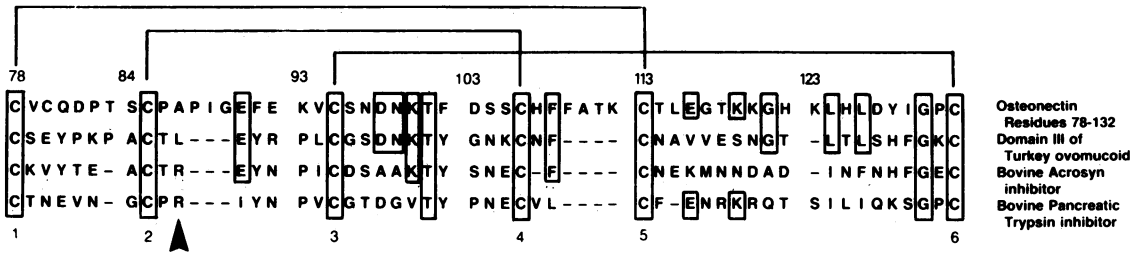

FIG. 4. Comparison of osteonectin sequence with sequences of serine proteinase inhibitors, including domain III of turkey ovomucoid (27), bovine acrosyn inhibitor (29), and bovine pancreatic trypsin inhibitor (30). Standard one-letter amino acid symbols are shown. Homologous residues are boxed. Alignment between cysteine residues was made by inserting spaces into the serine proteinase inhibitor sequence, indicated as dashes. Numbers above the sequence refer to residues in the osteonectin sequence; numbers below refer to positions of cysteine residues in the proteinase inhibitor sequences. Within the serine proteinase inhibitors, disulfide bonds occur between cysteines 1 and 5, 2 and 4, and 3 and 6, as indicated by solid lines. The active site in the serine proteinase inhibitors is marked by an arrowhead.

itor (30), two other members of this protein class (Fig. 4). Ovomucoids are avian egg proteins, each with three repeating domains (27, 28). Osteonectin appears to be homologous with all three domains, but the homology is strongest with the third (16 of 56 residues). Other members of this protein family have between one and seven equivalent domains (acrosyn inhibitor, ovoinhibitor, pancreatic trypsin inhibitor). The homology with osteonectin cysteine residues (6 of 7 residues) is indicated in Fig. 4. The structure of the serine proteinase inhibitors has been elucidated by NMR spectroscopy; the regions homologous to osteonectin are characterized by three disulfide bonds between six cysteine residues, forming a Kringle-type structure (31).

Homologies Between Osteonectin from Bovine and Human Bone and Two Non-bone Proteins. The amino-terminal sequence of human osteonectin, isolated from human stillborn tibia, was determined previously (9). A comparison of this sequence with that of bovine osteonectin (10) shows almost complete sequence identity in the first 36 amino acid residues (Fig. 5). Mason *et al.* (32) published a partial amino acid sequence of a protein secreted by cultured bovine endothelial cells (33, 34). The homology between this protein and osteonectin appears strong (Fig. 5), with sequence identity occurring in the first 10 amino acid residues.

The sequence of a cDNA encoding SPARC (an acronym for *s*ecreted *p*rotein, *a*cidic and *r*ich in *c*ysteine), a protein secreted by mouse parietal endoderm, has been reported (32). As indicated by Mason *et al.* (32), a strong homology exists between SPARC and osteonectin, with 70% identity in the nucleotide sequences and >90% identity in the amino acid sequences. The amino acid sequence differences are concentrated in the amino-terminal region of the molecule (16 of 22 differences are located in domain I, Fig. 2). The most notable of the six amino acid differences between SPARC and osteonectin found outside domain I is the deletion of Cys-193 from the mouse SPARC sequence. Another difference, between Ser-70 (SPARC) and Asn-71 (osteonectin), generates a second potential glycosylation site.

DISCUSSION

The data presented here predict the amino acid sequence of bovine bone osteonectin, a protein of 287 amino acids (M_r = 32,782). The protein appears to have four distinct domains: domain I, a 54-residue amino-terminal sequence rich in acidic and carboxyl-containing amino acid residues; domain II, an 84-residue sequence containing two potential N-glycosylation sites and 11 cysteine residues; domain III, a 54-residue hydrophilic region containing a potential calcium-binding loop sequence; and domain IV, a 95 amino acid sequence at the carboxyl terminus of the protein containing four cysteine residues, almost half of the total protein aromatic amino acid content, and a second calcium-binding loop homologous sequence. Domains III and IV also contain regions of α-helical secondary structure.

Particularly notable in osteonectin are distinct regions likely to have high affinity for binding both calcium and hydroxyapatite. These include two sequences homologous to the central calcium-binding loops in "EF hands" (23) of calmodulin and other intracellular calcium-binding proteins. Osteonectin sequences show partial homology with EF-hand calcium-binding loops at four of five calcium binding coordinates (23), with differences occurring at either the $-X$ or $-Z$ coordinates of the loop (Fig. 3). In addition, there is substitution of either Leu-170 or Lys-263 for the central glycine residue of the loop. These substitutions might suggest different orientation of some residues serving as binding ligands within the loop. However, as seen in the calcium-binding protein of muscle (24) and in bovine S-100 protein (25), such substitutions are still consistent with calcium binding. In calmodulin and other EF-hand structures, calcium-binding loops are stabilized by adjacent α-helices (23). Accordingly, the two homologous sequences in osteonectin are bracketed by regions of potential α-helices. Calcium-binding loops stabilized by structures other than α-helices have been demonstrated recently as well (35). In this regard, Mann *et al.* (36) found that in SPARC, the second calcium-binding loop is stabilized by a disulfide bond between Cys-257 and Cys-273. Finally, osteonectin was reported to have at least one high-affinity calcium-binding site (3). Based on fluorescence data, Romberg *et al.* (3) postulated that calcium binding occurs very near a tyrosine or tryptophan residue. We predict a tryptophan residue (Trp-270) at the carboxyl terminus of one EF-hand-homologous domain (Asp-258 to Glu-269), as well as a tyrosine residue (Tyr-264) in the seventh position of the loop proper.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Osteonectin (Bovine-ON)	Ala	Pro	Gln	Gln	Glu	Ala	Leu	Pro	Asp	Glu	Thr	Glu	Val	Val	Glu	Glu	Thr	Val	Ala	Glu
Osteonectin (Human-ON)	Ala	Pro	Gln	Gln	Glu	Ala	Leu	Pro	Asp	Glu	Thr	Glu	Val	Val	Glu	Glu	Thr	Val	Ala	Glu
Endothelial Cell Culture Shock Protein (Bovine)	Ala	Pro	Gln	Gln	Glu	Ala	Leu	Pro	Asp	Glu	Cys	XXX	Val	Val						
SPARC (Mouse)	Ala	Pro	Gln	Gln	Thr	Glu	Val	Ala	Glu	Glu	Ile	XXX	Val	Glu	Glu	Glu	Thr	Val	Val	Glu

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Bovine-ON	Val	Ala	Glu	Val	Pro	Val	Gly	Ala	Asn	Pro	Val	Gln	Val	Glu	Val	Gly	Glu	Phe	Asp	Asp
Human-ON	Val	Thr	Glu	Val	Pro	Thr	Gly	Ala	Asn	Pro	Val	Gln	Val	Glu	Val	Gly				
SPARC	Glu	Thr	Gly	Val	Pro	Val	Gly	Ala	Asn	Pro	Val	Gln	Val	Glu	Met	Gly	Glu	Phe	Glu	Asp

FIG. 5. Amino-terminal sequence of bovine osteonectin (10), human osteonectin (9), bovine endothelial cell culture shock protein (32), and mouse SPARC (32). Amino acid residues are numbered starting with 1 at the known or presumed amino terminus. Divergent residues are printed in reverse font.

The acidic amino-terminal region of osteonectin could well provide a potential low-affinity calcium-binding site and/or the hydroxyapatite-binding site of osteonectin. The clustering of carboxylate residues in this region provides a suitable environment for apatite binding (26). Moreover, the studies of Romberg *et al.* (3) indicated that the inhibitory activity of osteonectin on hydroxyapatite crystal growth did not correlate with high-affinity calcium binding (3, 8), suggesting that different sites are involved in these two actions of the protein.

Based on our analysis of the amino acid sequence, we propose a possible model for osteonectin structure (Fig. 2). The signal peptide and the four principal domains are labeled, as are the two N-glycosylation sites, the internal homologies, and the EF-hand calcium-binding loop structures. The cysteine-rich regions are indicated, and the approximate positions of the cysteine residues are denoted (–S). Homologies with the positions of cysteine residues in serine proteinase inhibitors suggest potential disulfide bonding between the six cysteine residues indicated in Fig. 4. Positions of other disulfide bonds cannot currently be predicted, although preliminary data may suggest locations of two of them (36). While proteinase inhibitor activity for osteonectin has not been demonstrated (34), this possibility may merit further investigation based on these new data. However, the osteonectin sequence contains inserted residues, not contained in serine proteinase inhibitors, in the region of the active site.

Previous studies have shown that osteonectin is a major noncollagenous protein in developing bone and that it occurs in greater amounts in bone than in other tissues so far examined (1–5, 8, 37, 38). Additionally, since osteonectin has a high affinity for collagen and hydroxyapatite, it was suggested that osteonectin could be involved in linking both matrix and mineral in bone (1). However, osteonectin has been shown to be present in human platelets (39). Other studies have shown that parietal endoderm and related cells produce SPARC, an osteonectin homologue (32, 40), whereas cultured fibroblasts (39) and endothelial cells (33, 34) synthesize the protein *in vitro*. Further, SPARC (40) and osteonectin (10) mRNA appear to be widely distributed. The degree of homology noted above between mouse SPARC and bovine osteonectin sequences is most remarkable (>90% sequence identity). Endodermal and endothelial cells also produce an extracellular matrix that appears *in vitro* to include osteonectin (33, 34, 37). Other studies have suggested that the assembly of basement membranes may require calcium (41) and that a protein similar to osteonectin, called BM40, is present in this *de novo* matrix (42). Preliminary sequence evaluation of BM40 shows strong homology to both osteonectin and SPARC, indicating that these are identical proteins (36). It may be possible, then, that the deposition of many collagenous matrices is calcium-dependent and could involve osteonectin-like proteins to help regulate matrix assembly.

Note. While this paper was under review, a more detailed description of the calcium-binding properties of the mouse protein was published (43).

We thank I. Mason and B. Hogan for information on the nucleotide sequence of SPARC prior to publication. We thank M. Sasaki (National Institute of Dental Research, National Institutes of Health) for assistance with DNA sequencing, M. Kanehisa (Kyoto University, Japan) for the IDEAS program, D. Lippman (National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, National Institutes of Health) for the FAST program, R. Staden (The Medical Research Council, England) for the STADEN-2 program, and S. Taylor (National Institute of Dental Research) for assistance with the computer programs. This work was partially supported by a grant to M.E.B. from the American Academy of Orthopaedic Surgeons, The Journal of Bone and Joint Surgery, and the Orthopaedic Research and Education Foundation.

1. Termine, J. D., Kleinman, H. K., Whitson, S. W., Conn, K. M., McGarvey, M. L. & Martin, G. R. (1981) *Cell* **26,** 99–105.
2. Termine, J. D., Belcourt, A. B., Conn, K. M. & Kleinman, H. K. (1981) *J. Biol. Chem.* **256,** 10403–10408.
3. Romberg, R. W., Werness, P. G., Lollar, P., Riggs, B. L. & Mann, K. G. (1985) *J. Biol. Chem.* **260,** 2728–2736.
4. Bianco, P., Hayashi, Y., Silvestrini, G., Termine, J. D. & Bonucci, E. (1985) *Calcif. Tissue Int.* **37,** 684–686.
5. Jundt, G., Berghauser, K.-H., Termine, J. D. & Schulz, A. (1987) *Cell Tissue Res.* **248,** 409–415.
6. Gehron-Robey, P., Kirshner, J. A., Conn, K. M. & Termine, J. D. (1985) in *Current Advances in Skeletogenesis*, eds. Ornoy, A., Harell, A. & Sela, J. (Elsevier, Amsterdam), pp. 461–466.
7. Gehron-Robey, P. & Termine, J. D. (1985) *Calcif. Tissue Int.* **37,** 453–458.
8. Romberg, R. W., Werness, P. G., Riggs, B. L. & Mann, K. G. (1986) *Biochemistry* **25,** 1176–1180.
9. Fisher, L. W., Hawkins, G. R., Tuross, N. & Termine, J. D. (1987) *J. Biol. Chem.* **262,** 9702–9708.
10. Young, M. F., Bolander, M. E., Day, A. A., Ramis, C. I., Robey, P. G., Yamada, Y. & Termine, J. D. (1986) *Nucleic Acids Res.* **14,** 4483–4497.
11. Adams, S. L., Sobel, M. E., Howard, B. H., Olden, K., Yamada, K. M., de Crombrugghe, B. & Pastan, I. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3399–3403.
12. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1194–1198.
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
14. Sanger, F., Coulson, A. R., Bartell, B. G., Smith, A. J. A. & Roc, B. A. (1980) *J. Mol. Biol.* **143,** 161–178.
15. Deininger, P. G. (1983) *Anal. Biochem.* **129,** 216–223.
16. Chou, P. Y. & Fasman, G. P. (1977) *J. Mol. Biol.* **115,** 135–175.
17. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
18. Lippman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
19. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
20. Proudfoot, N. T. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
21. Von Heijne, G. (1985) *J. Mol. Biol.* **184,** 99–105.
22. Kakiuchi, S. & Yamazaki, R. (1970) *Biochem. Biophys. Res. Commun.* **41,** 1104–1110.
23. Kretsinger, R. H. (1980) *CRC Crit. Rev. Biochem.* **8,** 119–174.
24. Cox, J. A., Wnuk, W. & Stein, E. A. (1976) *Biochemistry* **15,** 2613–2618.
25. Calissano, P., Alema, S. & Fasella, P. (1975) *Proc. FEBS Meet.* **41,** 207–215.
26. Peckauskas, R. A., Termine, J. D. & Pullman, I. (1976) *Biopolymers* **15,** 569–581.
27. Kato, I., Schrode, J., Kohie, W. J. & Laskowski, M., Jr. (1987) *Biochemistry* **26,** 193–201.
28. Laskowski, M., Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M. W., Kohr, W. J., Park, S. J., Parks, K., Schatzley, B. L., Schoenberger, O. L., Tashiro, M., Vichot, G., Whatley, H. E., Wieczorek, A. & Wieczorek, M. (1987) *Biochemistry* **26,** 202–221.
29. Meloun, B., Cechova, D. & Jonakova, V. (1983) *Hoppe-Seylers Z. Physiol. Chem.* **364,** 1665–1670.
30. Greene, L. J. & Bartelt, D. C. (1969) *J. Biol. Chem.* **244,** 2646–2657.
31. Guy, O., Shapanica, R. & Greene, L. J. (1971) *J. Biol. Chem.* **246,** 7740–7747.
32. Mason, I. J., Taylor, A., Williams, J. G., Sage, H. & Hogan, B. L. M. (1986) *EMBO J.* **5,** 1465–1472.
33. Sage, H. (1985) *Biochemistry* **24,** 7430–7440.
34. Sage, H., Johnson, C. & Bornstein, P. (1984) *J. Biol. Chem.* **259,** 3993–4007.
35. Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (1987) *Nature (London)* **327,** 635–638.
36. Mann, K., Deutzmann, R., Paulsson, M. & Timpl, R. (1987) *FEBS Lett.* **218,** 167–172.
37. Gehron-Robey, P., Fisher, L. W., Stubbs, J. T. & Termine, J. D. (1987) in *Development and Diseases of Cartilage and Bone Matrix*, UCLA Symposia on Molecular Biology, New Series, eds. Sen, A. & Thornhill, T. (Liss, New York), Vol. 46, pp. 115–126.
38. Stenner, D. D., Rombert, R. W., Tracy, R. P., Katzman, J. A., Riggs, B. L. & Mann, K. G. (1984) *Proc. Natl. Acad. Sci USA* **81,** 2868–2872.
39. Stenner, D. D., Tracy, R. P., Riggs, B. L. & Mann, K. G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6892–6896.
40. Mason, I. J., Murphy, D., Munke, M., Francke, U., Elliott, R. W. & Hogan, B. L. M. (1986) *EMBO J.* **5,** 1831–1837.
41. Yurchenco, P. O., Tsilibary, E. C., Charonis, A. S. & Furthmayr, H. (1985) *J. Biol. Chem.* **260,** 7636–7644.
42. Dziadek, M., Paulson, M., Aumailley, M. & Timpl, R. (1986) *Eur. J. Biochem.* **161,** 455–464.
43. Engel, J., Taylor, W., Paulsson, M., Sage, H. & Hogan, B. (1987) *Biochemistry* **26,** 6958–6965.